Proximal Point Methods
for
Variational Problems

R. Tichatschke
Preface

The subject of variational problems, in particular variational inequalities, has its origin in the calculus of variations associated with the minimization of functionals in infinite-dimensional spaces. The systematic study of the subject began in the early 1960s with the seminal work of the Italian mathematician Guido Stampacchia and his collaborators, who used variational inequalities (VI’s) as analytic tool for studying free boundary problems defined by nonlinear partial differential operators arising from unilateral problems in elasticity and plasticity theory and in mechanics. Some of the earliest papers in variational inequalities are Lions and Stampacchia [267], Mancino and Stampacchia [283] and Stampacchia [381]. In particular, the first theorem of existence and uniqueness of the solutions of VI’s was proved in [381]. The books by Baiocchi and Capelo [28] and Kinderlehrer and Stampacchia [234] provide a thorough introduction to the application of VI’s in infinite-dimensional function spaces. The book by Glowinski, Lions and Trémolière [135] is among the earliest references to give a detailed numerical treatment of such VI’s. Nowadays there is a huge literature on the subject of infinite-dimensional VI’s and related problems.

The development of mathematical programming and control theory has proceeded almost contemporarily with the systematical investigation of ill-posed problems and their numerical treatment. It was clear from the very beginning that the main classes of extremal problems include ill-posed problems. Among the variational inequalities, having important applications in different fields of physics, there are ill-posed problems, too. Nevertheless, up to now, in the development of numerical methods for finite- and infinite-dimensional extremal problems, ill-posedness has not been a major point of consideration. As a rule, conditions ensuring convergence of a method include assumptions on the problem which warrant its well-posedness in a certain sense. Moreover, in many papers exact input data and exact intermediate calculations are assumed. Therefore, the usage of standard optimization and discretization methods often proves to be unsuccessful for the treatment of ill-posed problems.

In the first methods dealing with ill-posed linear programming and optimal control problems, suggested by Tikhonov [393, 394], the problem under consideration was regularized by means of a sequence of well-posed problems involving a regularized objective and preserving the constraints from the original problem.

Essential progress in the development of solution methods for ill-posed variational inequalities was initiated by a paper of Mosco [296]. Based on Tikhonov’s principle, he investigated a stable scheme for the sequential approximation of variational inequalities, where the regularization is performed simultaneously with an approximation of the objective functional and the feasible set. Later on analogous approaches became known as iterative regularization.

A method using the stabilizing properties of the proximal-mapping (see Moreau [292]) was introduced by Martinet [287] for the unconstrained minimization of a convex functional in a Hilbert space. Rockafellar [351, 352] created the theoretical foundation to the further advances in iterative proximal point regularization for ill-posed VI’s with monotone operators and convex optimization problems. These results attracted the attention of numerical analysts,
and the number of papers in this field was increasing rapidly during the last decades. With regard to the stability of the regularized problems, which have to be solved throughout the iteration process, proximal point methods (PPM) are superior in contrast to other approaches.

In the monograph "Stable Methods for Ill-posed Variational Problems" [215], the purpose was to give a general framework to the investigation of the convergence and the rate of convergence of iterative proximal point regularization (IPR) when applied to ill-posed convex variational problems. The iterative regularization principle was generalized by several schemes and the approaches presented there allow also to analyze discretization and regularization procedures in connection with specific optimization methods, i.e., the behavior of the solution process has been studied as a whole.

Now, I am pleasantly inspired to prepare an extended edition of the monograph, because it has apparently been accepted reasonable well by the optimization community. The objective in putting the present book together is to detail the major ideas in regularizing of different classes of ill-posed variational problems that have evolved in the past twenty years - not only theoretically but also computationally.

While I may have overlooked the importance of some very recent developments, I think that major new breakthroughs on those two fronts that interest me - theory and methods - have occurred mainly in the direction of computational performances since the monograph [215] was published. Now, as a consequence I not only restricted myself to a thorough re-working and face lifting of previous results but have taken also the opportunity to focus on some new topics, like the work on elliptic regularization, regularization of control problems, regularization with non-quadratic distance functions, the consideration of the Auxiliary Problem Principle in context with regularization as well as several numerical experiments in the regularization of certain classes of problems. Most of the results, describe here, are produced in a long-standing fruitful cooperation with ALEXANDER KAPLAN and I am regretting that he feels now unable to work on this project.

The book consists of ten chapters and an appendix. Each chapter provides motivations at the beginning and throughout, and each concludes with notes and comments which furnish credits and references together with historical perspective on how the ideas gradually took shape. Also they give the reader a broader orientation and special hints for further investigation. Sometimes figures and examples are provided giving the reader a clear understanding of the peculiarities of the material. The necessary theoretical background from functional analysis, finite element methods and optimization theory is briefly described in the Appendix.

In the first two chapters several stability concepts in optimization and other fields of mathematics are discussed. Classical approaches to the solution of ill-posed problems are considered and the regularizing properties of different optimization methods are investigated.

Chapter 3 is devoted to IPR for finite-dimensional convex optimization problems, containing some proximal variants of projected gradient methods, penalty methods and augmented Lagrangian methods. In particular the application of proximal point ideas in other fields of numerical analysis and the usage
of IPR in nonconvex programming is discussed.

Chapters 4 and 5 deal with a general framework to the investigation of IPR for infinite-dimensional convex variational problems, including a new class of multi-step regularization methods, regularization on subspaces as well as regularization in weaker norm, which permit a more efficient use of rough approximations of the original problem and therefore may lead to an essential acceleration of the numerical process. Conditions are established concerning the coordinated update of discretization and regularization parameters together with stopping parameters arising in stopping rules for the solution of the auxiliary problems. These conditions provide weak or strong convergence of the methods. Asymptotically exact estimates of the rate of convergence are obtained for a broad class of ill-posed convex variational problems, and linear convergence is proved for a particular class.

In the next four chapters the general framework is applied and modified to the treatment of some specific variational problems.

Subject of Chapter 6 is the description of regularized penalty and barrier methods for convex semi-infinite and convex parametric semi-infinite problems. In particular the numerical behavior of regularized interior point methods for convex semi-infinite problems is studied. In these cases the use of IPR enables us to reduce the dimension of the auxiliary problems substantially by means of special deleting rules for the discretized constraints.

Chapter 7 is devoted to the stable solution of ill-posed control problems. A significant peculiarity of this approach is that regularization is accomplished only with respect to the control variables but not to the state variables. This is a natural application of partial regularization on a subspace, introduced in Chapter 4. For elliptic control problems with mixed boundary conditions, the discretization of the state and control space and the convergence of the inexact multi-step regularization approach is described and its numerical behavior is analyzed for a particular class of such problems – the so called chattering problems.

Chapter 8 reflects the application of the general framework for ill-posed monotone variational inequalities, including contact problems of the elasticity theory and some obstacle problems. Several variants of multi-step methods (with weak regularization and regularization on a subspace) are considered which allow us to use efficiently the monotonicity reserve of the corresponding differential operators, in particular, the structure of their kernels. These regularization procedures are coupled with finite element methods used for the discretization of the VI's.

Chapter 9 is dealing with the solution of VI's with multi-valued maximal monotone operators. Starting with the idea of a relaxation of IPR methods and the use of enlargements of maximal monotone operators, this approach is suggested to solve non-smooth minimization problems. The appearing numerical aspects and computational results are investigated in detail. In this chapter also interior point methods are considered by using non-quadratic distance functions. The main motivation for such proximal methods is not only to guarantee that the iterates stay in the interior of the feasibility set of the VI but also to preserve the main merits of the classical IPR: Good stability of the auxiliary problems and convergence of the whole sequence of iterates to a solution of the original problem. In particular, also Bregman functions for convex and non-polyhedral sets are investigated and the standard requirement of the strict convexity of the
Bregman function is weakened. This leads to an analogy of methods with weak regularization and regularization on a subspace developed on the basis of the classical IPR in Chapter 8.

Finally, Chapter 10 deals with extensions of the auxiliary problem principle in order to solve VI’s with non-symmetric multi-valued operators in Hilbert spaces. The idea to take in the solution scheme an additive component of the main operator of the VI at the sought iterate leads to new methods which link the advantages of the auxiliary problem principle with those of proximal-like methods.

Concerning the bibliographic data given in this book they are not primarily intended to document original publications and sources or to reflect the historical development of the researches at the whole. Also, the list of references makes no attempt at being complete; we have listed mainly the papers which are discussed in the text. Often we refer to more comprehensive texts. Except for the case that reference is given to explicit statements or techniques, it was our aim to draw the readers attention to the basic approaches and fields of investigation.

Both authors have benefitted from the fruitful collaboration with their colleagues and doctoral students, in particular, we mention L. Abbe, E. Huebner, V. Kustova, R. Namm, S. Rotin, H. Schmitt, T. Voetmann. We thank them all. We would like to express our gratitude also to the German Research Foundation, whose consecutively financial support in a series of projects in that field has kept us hopeful that optimization and regularization is still an important subject. We highly appreciate all support and help we got from wonderful colleagues that we have met and worked with at some point or another during the past twenty years.

Trier, Dezember 2010

R. Tichatschke
Contents

Nomenclature ix

1 ILL-POSED PROBLEMS AND THEIR STABILIZATION 1
  1.1 Concepts of Well-Posedness – Examples of Ill-Posed Problems . . 1
  1.2 Well- and Ill-posed Variational Problems . . . . . . . . . . . . . 5
    1.2.1 Notations and definitions . . . . . . . . . . . . . . . . . 5
    1.2.2 Some conditions of well-posedness . . . . . . . . . . . . 9
  1.3 Approaches of Stabilizing Ill-posed Problems . . . . . . . . . . 14
    1.3.1 Stabilizers and quasi-solutions . . . . . . . . . . . . . . 14
    1.3.2 Classical approaches for stabilization of ill-posed problems 15
    1.3.3 Tikhonov regularization . . . . . . . . . . . . . . . . . . 20
    1.3.4 Proximal point regularization . . . . . . . . . . . . . . . 22
  1.4 Comments . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

2 STABLE DISCRETIZATION AND SOLUTION METHODS 29
  2.1 Stabilizing Properties of Numerical Methods . . . . . . . . . . . 29
    2.1.1 Stabilizing properties of gradient type methods . . . . . 29
    2.1.2 Stabilizing properties of penalty methods . . . . . . . . . 32
  2.2 Stabilization of Problems and Methods . . . . . . . . . . . . . . 36
    2.2.1 Structure of regularizing methods . . . . . . . . . . . . . 37
    2.2.2 Iterative one-step regularization . . . . . . . . . . . . . 38
    2.2.3 Iterative multi-step-regularization . . . . . . . . . . . . 38
  2.3 Iterative Tikhonov Regularization . . . . . . . . . . . . . . . . . 39
    2.3.1 Regularized subgradient projection methods . . . . . . . 39
    2.3.2 Regularized penalty methods . . . . . . . . . . . . . . . 42
  2.4 Mosco’s Approximation Scheme . . . . . . . . . . . . . . . . . . . 47
    2.4.1 Approximation of well-posed problems . . . . . . . . . . 48
    2.4.2 Mosco’s scheme for ill-posed problems . . . . . . . . . . 49
  2.5 Comments . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52

3 PPR FOR FINITE-DIMENSIONAL PROBLEMS 55
  3.1 Properties of the Proximal-Point-Mapping . . . . . . . . . . . . . 55
  3.2 PPR for Convex Problems . . . . . . . . . . . . . . . . . . . . . . 60
    3.2.1 Proximal gradient methods . . . . . . . . . . . . . . . . . 60
    3.2.2 Penalty methods with iterative PPR . . . . . . . . . . . . 60
    3.2.3 Augmented Lagrangian methods with iterative PPR . . . . 73
  3.3 PPR in Methods of Numerical Analysis . . . . . . . . . . . . . . 77
    3.3.1 Implicit Euler method . . . . . . . . . . . . . . . . . . . 77
# CONTENTS

## PARTIAL PPR IN CONTROL PROBLEMS 243

7.1 Non-coercive Elliptic Control Problems .......................... 243
7.1.1 Distributed control problems ................................. 245
7.1.2 Regularized penalty methods ................................. 247
7.1.3 A simple example ........................................... 259
7.2 Elliptic Control with Mixed Boundary Conditions .......... 261
7.2.1 Discretization in state and control space .......... 261
7.2.2 Convergence of inexact MSR-methods .............. 268
7.2.3 Numerical results ........................................ 278
7.3 Comments ................................................. 281

## PPR FOR VARIATIONAL INEQUALITIES 283

8.1 Approximation of Variational Inequalities .................. 283
8.2 Contact Problems Without Friction .............................. 288
8.2.1 Formulation of the problems ................................. 289
8.2.2 Finite element approximation of contact problems ...... 294
8.2.3 Three Variants of MSR-Methods ............................. 297
8.2.4 On choice of variants of IPR-methods ............... 306
8.2.5 Special Case: Inner approximation of the set $K$ ...... 308
8.2.6 Exactness of approximations of feasible sets ........ 315
8.2.7 Numerical results ........................................ 320
8.3 Solution of Contact Problems With Friction ................. 328
8.3.1 Signorini problem with friction ............................ 329
8.3.2 Algorithm of alternating iterations ....................... 329
8.3.3 Numerical results ........................................ 333
8.4 Regularization of Well-posed Problems .......................... 335
8.4.1 Linear rate of convergence of MSR-methods ............ 336
8.5 Elliptic Regularization ......................................... 339
8.5.1 A generalized proximal point method ....................... 340
8.5.2 Convergence analysis of the method ....................... 343
8.5.3 Application to minimal surface problems ............... 350
8.5.4 Application to convection-diffusion problems ........ 357
8.6 Comments ................................................. 362

## PPR FOR VI's WITH MULTI-VALUED OPERATORS 365

9.1 PPR with Relaxation ........................................... 365
9.1.1 Relaxed PPR with enlargements of maximal monotone operators .................. 367
9.1.2 Application to non-smooth minimization problems .... 371
9.1.3 Numerical aspects and computational results ........ 375
9.1.4 Choice of relaxation parameters .......................... 379
9.2 Interior PPR on Non-Polyhedral Sets .......................... 387
9.2.1 Bregman function based PPR ................................. 388
9.2.2 Bregman functions with non-polyhedral zones .......... 394
9.2.3 Embedding of original Bregman-function-based methods 401
9.3 PPR with Generalized Distance Functionals .................. 403
9.3.1 Generalized proximal point method ....................... 405
9.3.2 Convergence analysis ..................................... 410
9.4 Comments ................................................. 417
Nomenclature

Mathematical Symbols

\[ \mathbb{N} \] the set of natural numbers
\[ \mathbb{R} \] the set of real numbers
\[ \bar{\mathbb{R}} \] the extended real line, \( \bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty\} \)
\[ \mathbb{R}_+ \] the set of non-negative real numbers
\[ \mathbb{1}_n \] \( n \)-vector of all ones (subscript often omitted)
\[ \lim; \liminf \] the lower (resp. upper) limit
\[ A \Rightarrow B \] \( A \) implies \( B \)
\[ A \iff B \] \( A \) implies \( B \) implies \( A \) and \( B \)
\[ \forall \] for all (universal quantifier)
\[ \exists \] there is (existential quantifier)
\[ \text{iff} \] if and only if

Sets

\[ |S| \] the cardinality of a set \( S \)
\[ \text{meas} S \] the measure of a set \( S \)
\[ \text{bd}(S); \partial S \] the topological boundary of a set \( S \)
\[ \text{int}(S); \text{ri}(S) \] the topological interior (resp. relative interior) of a set \( S \)
\[ \text{cl}(S) \] the topological closure of a set \( S \)
\[ \mathcal{N}_S(x) \] the normal cone of a set \( S \) at \( x \)
\[ \text{aff}(S) \] the affine hull of a set \( S \)
\[ \text{conv}(S) \] the convex hull of a set \( S \)
Contents

\begin{itemize}
  \item \textnormal{cone}(S) \quad \text{the conical hull of a set } S
  \item \textnormal{S}^\perp \quad \text{the orthogonal complement of a set } S
  \item \mathbb{B}_\rho(x) \quad \text{the (closed) ball with radius } \rho \text{ and center } x
  \item S_\rho(x) \quad \text{the sphere with radius } \rho \text{ and center } x
  \item \mathcal{U}(x) \quad \text{the neighborhood of a point } x
  \item 2^S \quad \text{the power set (set of all subsets) of a set } S
  \item \([a, b]; (a, b)\) \quad \text{the closed (resp. open) real interval}
  \item \([a, b); (a, b]\) \quad \text{the half-open real interval}
  \item \rho(x, Q) \quad \text{the distance of the point } x \text{ from the set } Q
  \item \rho(S, Q) \quad \rho(S, Q) := \sup_{s \in S} \rho(s, Q)
  \item \rho_H(S, Q) \quad \text{the Hausdorff distance between } S \text{ and } Q
  \item \Omega \quad \text{an open and connected domain}
  \item \Gamma; \partial \Omega \quad \text{the boundary of a domain } \Omega
  \item \mathcal{B}_h \quad \text{a triangulation of a domain } \Omega
  \item T_k \quad \text{finite } h_k\text{-grid on } T \subset \mathbb{R}^m
\end{itemize}

Functionals and Operators

\begin{itemize}
  \item \(f : X \rightarrow Y\) \quad \text{a function mapping from } X \text{ into } Y
  \item \(G : D \rightarrow R\) \quad \text{a mapping with domain } D \text{ and range } R
  \item \(\mathcal{T} : V \rightarrow V'\) \quad \text{an operator mapping from } V \text{ into } V'
  \item \(\mathcal{T}'\) \quad \text{the conjugate operator to an operator } \mathcal{T}, \mathcal{T}' : V' \rightarrow V
  \item \(\mathcal{T}^{-1}\) \quad \text{the inverse operator to an operator } \mathcal{T}
  \item \text{supp}\_f \quad \text{the support of function } f
  \item \text{lsc; usc} \quad \text{lower (resp. upper) semicontinuous}
  \item \text{wlsc; wusc} \quad \text{weakly lower (resp. weakly upper) semicontinuous function}
  \item \(D(f); D(\mathcal{T})\) \quad \text{definition domain of a functional } f \text{ (resp. operator } \mathcal{T})
  \item \(\text{dom}(f); \text{dom}(\mathcal{T})\) \quad \text{the effective domain of a function } f \text{ (resp. operator } \mathcal{T})
  \item \text{gph}(f); \text{gph}(\mathcal{T}) \quad \text{the graph of a function } f \text{ (resp. operator } \mathcal{T})
  \item \text{epi}(f) \quad \text{the epigraph of a function } f
\end{itemize}
### CONTENTS

- \( f \circ g \) the composition of two functions, \((f \circ g)(u) = f(g(u))\)
- \( \gamma \varphi \) the trace of a function \( \varphi \) on a boundary \( \Gamma \)
- \( u_{I,h} \) the interpolant of a function \( u \) on a finite element \( T_h \)
- \( \text{ind}(\cdot \mid K) \) the indicator functional of the set \( K \)
- \( \ker T \) the kernel of an operator \( T \)
- \( \mathcal{R} \) the kernel of a bilinear form
- \( \text{rank} A \) the rank of a matrix \( A \)
- \( a(u,v) \) a symmetrical bilinear form
- \( \mathcal{N}_K(\cdot) \) the normality operator of a set \( K \)
- \( \Pi_K(u) \) the ortho-projector, projecting \( u \) onto a set \( K \)
- \( \nabla f(u) \) the gradient of a functional \( f \) at the point \( u \)
- \( \nabla^2 f(u) \) the second derivative of a functional \( f \) at the point \( u \), Hessian of \( f \)
- \( \partial f(u) \) the subdifferential of a functional \( f \) at the point \( u \)
- \( G'(u) \) the Fréchet- or Gâteaux derivative of \( G \) at the point \( u \)
- \( G''(u) \) the Fréchet- or Gâteaux derivative of \( G' \) at the point \( u \)
- \( f'(u;v) \) directional derivative of the function \( f \) at the point \( u \) in direction \( v \)
- \( D^\alpha := D_1^{\alpha_1} \ldots D_n^{\alpha_n} \), the derivative in multi-index notation, \( D_i = \frac{\partial}{\partial x_i} \)
- \( \frac{\partial}{\partial \nu} \) the derivative in direction of the outer normal \( \nu \)
- \( \Delta u \) the Laplace operator at the point \( u \)
- \( \epsilon_{kl}(u) \) the tensor field of strains
- \( \tau_{kl}(u) \) the tensor field of stress

### Spaces and Norms

- \( \mathbb{R}^n \) the real \( n \)-dimensional vector space
- \( \mathbb{R}^+_n \) the \( n \)-dimensional nonnegative orthant
- \( \mathbb{R}^+_n \) the \( n \)-dimensional positive orthant
- \( H; V \) a Hilbert space
- \( V' \) the conjugate (or dual) space to the space \( V \)
$V_h$ a finite dimensional approximation of the space $V$

$\langle \cdot, \cdot \rangle_V$ a duality pairing between $V$ and $V'$

$\langle u, v \rangle$ a scalar product in Hilbert spaces

$\{u^k \}; \{u^{k_i} \}$ a sequence (resp. subsequence)

$u^k \to u$ the norm convergence: $\| u^k - u \| \to 0$

$u^k \rightharpoonup u$ the weak convergence: $\forall \ell \in V': \ell(u^k) \to \ell(u)$

$\dim L$ the dimension of a linear subspace $L$

$\| u \|_V$ the norm of an element $u$ in the space $V$

$\| u \|$ the norm of $u$ in a space of vector functions

$\| u \|_V$ the semi-norm of an element $u$ in the space $V$

$\| T \|_V$ the norm of an operator $T$ in the space $V$

$\mathcal{L}(X,Y)$ the space of linear continuous operators from $X$ into $Y$

$C^m(\bar{\Omega})$ the space of $m$-times continuously differentiable functions $f$ on $\bar{\Omega}$

$C^\infty(\bar{\Omega}) : = \cap_{m=0}^\infty C^m(\bar{\Omega})$

$C^m,\lambda(\bar{\Omega})$ the space of functions $C^m(\bar{\Omega})$, whose derivatives of order $m$ satisfy a Hölder condition with constant $\lambda$

$\mathcal{D}(\Omega)$ $\mathcal{D}(\Omega) = \{ v \in C^\infty(\bar{\Omega}) : \text{supp} \ v \text{ is compact in } \Omega \}$

$L_2(\Omega)$ the Lebesque space

$H^m(\Omega)$ $\{ v \in L_2(\Omega) : D^\alpha v \in L_2(\Omega) \text{ for all } \alpha \text{ with } |\alpha| \leq m \}$

$[H^m(\Omega)]^n$ the space of $n$-dimensional vector functions with components in $H^m(\Omega)$

$H_0^m(\Omega)$ the closure of $\mathcal{D}(\Omega)$ in $H^m(\Omega)$

$H^{1/2}(\Omega)$ the space of the traces of functions from $H^1(\Omega)$

$H^{-1/2}(\Omega)$ the dual space to $H^{1/2}(\Omega)$

$\langle u, v \rangle_{m,\Omega}$ the scalar product in $H^m(\Omega)$

$\langle u, v \rangle_{m,\Omega,0}$ the scalar product in $H^m_0(\Omega)$

$\| u \|_{m,\Omega}$ the norm in $H^m(\Omega)$

$\| u \|_{m,\Omega}$ the norm in $[H^m(\Omega)]^n$

$\| u \|_{m,\Omega,0}$ the norm in $H^m_0(\Omega)$

$X \hookrightarrow Y$ $X$ is continuously embedded into $Y$
Minimization and Regularization

- $X \hookrightarrow Y$: $X$ is compactly embedded into $Y$

### Mathematical Symbols and Definitions

- $K; K_k$: the feasible set of a minimizing problem
- $J; J_k$: the objective functional of a minimizing problem
- $\inf_{u \in K} J(u)$: the infimum of the functional $J$ on the set $K$
- $u^*$: $\arg\min_{u \in K} J(u)$, the unique minimizer of $J$ over $K$
- $U^*$: $\arg\min_{u \in K} J(u)$, the set of minimizers of $J$ over $K$
- $J^*$: $J(u^*), u^* \in U^*$, the optimal value of the objective functional of $J$
- $\mathcal{L}; \mathcal{L}_A$: the Lagrangian function; augmented Lagrangian function
- $T(u^*) := \{ t \in T : g(u^*, t) = \max_{\tau \in T} g(u^*, \tau) \}$, the set of active parameters in SIP
- $\text{Prox}_{f,C}(\cdot) := \arg\min_{v \in C}\{f(v) + \frac{\chi}{2} \| v - \cdot \|^2 \}$, the proximal point mapping
- $\eta(\cdot) := \min_{v \in C}\{f(v) + \frac{\chi}{2} \| v - \cdot \|^2 \}$, the Yosida regularization
- $\Psi_{X,u} := f(\cdot) + \frac{\chi}{2} \| \cdot - u \|^2$, a regularized functional
- $\Psi_k := J_k(\cdot) + \| \cdot - u^{k-1} \|^2$
- $\Psi_k := J(\cdot) + \| \cdot - u^{k-1} \|^2$
- $\Psi_{k,i} := J_k(\cdot) + \| \cdot - u^{k,i-1} \|^2$
- $\Psi_{k,i} := J(\cdot) + \| \cdot - u^{k,i-1} \|^2$
- $\phi(\cdot)$: a penalty function
- $F_{k,i} := J_k(\cdot) + \varphi_k(\cdot) + \| \cdot - u^{k,i-1} \|^2$
- $J_{X,Q}(\cdot) := (I + \chi Q)^{-1}$, the resolvent operator
- $\omega(\cdot)$: the stabilizer
- $D_h(x,y)$: the Bregman distance
- $G_u(K)$: the set of feasible directions at $u$ in $K$
- $r_k$: the penalty parameter
- $\chi_k$: the regularization parameter
Problem Classes and Fundamental Objects

APP auxiliary problem principle
BPPA Bregman-function-based proximal point algorithm
IPR iterative proximal point regularization
PPR proximal point regularization
PAP proximal auxiliary problem method
SIP semi infinite programming problem

\((P^k)\) regularized auxiliary problem at step \(k\)

CP\((f,K)\) convex minimization problem of \(f\) on \(K\)

\(\text{IP}(\mathcal{Q}, K)\) inclusion defined by the set \(K\) and the operator \(\mathcal{Q}\)

\(\text{IP}(\mathcal{T}, \Xi, K)\) \(\text{IP}(\mathcal{Q}, K)\) with the operator \(\mathcal{Q} := \mathcal{T} + \Xi\)

HVI\((\mathcal{Q}, f, K)\) hemi-variational inequality defined by the set \(K\), the operator \(\mathcal{Q}\) and the functional \(f\)

VI\((\mathcal{Q}, K)\) variational inequality defined by the set \(K\) and the operator \(\mathcal{Q}\)

\(\text{VI}(\mathcal{F}, \mathcal{P}, K)\) \(\text{VI}(\mathcal{Q}, K)\) with the operator \(\mathcal{Q} := \mathcal{F} + \mathcal{P}\)

\(\text{VI}(\mathcal{T}, \partial f, K)\) \(\text{VI}(\mathcal{Q}, K)\) with the operator \(\mathcal{Q} := \mathcal{T} + \partial f\)

SOL\((\mathcal{Q}, K)\) solution set of \(\text{VI}(\mathcal{Q}, K)\)

Miscellaneous

\(o(\cdot)\) any function such that \(\lim_{t \downarrow 0} \frac{o(t)}{t} = 0\)

\(O(\cdot)\) any function such that \(\limsup_{t \downarrow 0} \frac{|O(t)|}{t} < \infty\)

:= the assignment operator

\(\square\) end of a proof

\(\diamond\) end of a example, remark etc.
Chapter 1

ILL-POISED PROBLEMS
AND THEIR
STABILIZATION

1.1 Concepts of Well-Posedness – Examples of Ill-Posed Problems

We start with the consideration of some ill-posed problems in the classical domain of mathematics, for instance solution of differential and integral equations, computation of the value of an unbounded operator and others, which on the one hand have important applications in natural sciences and on the other hand are frequently used as model-problems in theoretical and numerical studies of ill-posed problems.

At first sight these examples seem far away from the variational problems studied in this book. However, we will see later on that many numerical approaches are based on variational formulations which allow the application of optimization techniques. This is not surprising with regard to the broad reach of variational methods in mathematics. Although we are mainly concerned with problems in Hilbert spaces, in this chapter we will study the problems in Banach spaces to get a more fundamental insight into the class of ill-posed problems. Later on we consider variational problems immediately given in a Hilbert space and discard the examples described in more general spaces.

The following concept of well-posedness of a problem is convenient for a preliminary analysis of different types of problems from a common point of view.

We consider a model describing the dependence of "solution"

\[ u = G(z) \subset R_G \subset U \]

on some "input data" \( z \in D_G \subset Z \), where \( U \) and \( Z \) are Banach spaces with metrics \( \rho_U \) and \( \rho_Z \), respectively. Here \( D_G \) and \( R_G \) denote the domain of definition and the range of the mapping \( G \).

1.1.1 Definition. The problem of computing \( u = G(z) \) is called (conditionally) well-posed on the set \( D \subset D_G \) if
(i) the mapping $G$ is single-valued on $D$ and

(ii) $G$ is continuous on $D$ as a mapping from $Z$ into $U$.

We call $G(D) \subset U$ the set of well-posedness of the considered problem. If at least one of the conditions (i) or (ii) fails to hold, the problem is termed ill-posed on $D$.

1.1.2 Remark. For the time being we do not consider the question of how to select the set $D$ in order to take into account actual information on the problem. However, in applications it is important to know for instance whether the approximate data are contained in $D$. We emphasize that notions of well- and ill-posedness depend on the chosen pair of spaces $U$ and $Z$.

An essentially more general concept will be considered later on for variational problems. In particular, set-valued mappings $G$ will be permitted.

Our first example is a classical ill-posed problem considered by Hadamard. It should be noted that Hadamard [162, 163] has been the first who gave a description of the class of well-posed problems.

1.1.3 Example. (Cauchy Problem for the Laplace equation.)

A solution $u$ of the equation

$$\Delta u(x_1, x_2) = 0$$

is sought on the domain $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 > 0\}$ subject to the boundary conditions

$$u(x_1, 0) = 0, \quad \frac{\partial u}{\partial x_2} \bigg|_{x_2=0} = \varphi(x_1), \quad -\infty < x_1 < \infty,$$

where $\varphi$ is a given function. Here $u = G(\varphi)$ describes the solution in dependence of the boundary function $\varphi$.

For $\varphi(x_1) \equiv \varphi_1(x_1) = a^{-1} \sin ax_1$, $a > 0$, the function

$$u_1(x_1, x_2) = a^{-2} \sin ax_1 \sinh ax_2$$

is the unique solution of the Cauchy problem, whereas in case $\varphi(x_1) \equiv \varphi_2(x_1) = 0$ the solution is

$$u_2(x_1, x_2) = 0.$$

Taking $Z = C(\mathbb{R})$, $U = C(\bar{\Omega})$, we get $\rho_Z(\varphi_1, \varphi_2) = a^{-1}$ and, for arbitrary $x_2$,

$$\rho_U(u_1, u_2) \geq \sup_{x_1 \in \mathbb{R}} |u_1(x_1, x_2) - u_2(x_1, x_2)|$$

$$= \sup_{x_1 \in \mathbb{R}} \left| a^{-2} \sin ax_1 \sinh ax_2 \right| = a^{-2} \sinh ax_2.$$

Obviously, by means of the choice of $a$, the value of $\rho_Z(\varphi_1, \varphi_2)$ can be made arbitrarily small and at the same time $\rho_U(u_1, u_2)$ will be arbitrarily large for a fixed $x_2 > 0$.

This shows that the Cauchy Problem for the Laplace equation considered on the pair of spaces $U = C(\bar{\Omega})$ and $Z = C(\mathbb{R})$ is ill-posed on the set

$$D = \{ \varphi \in Z : \|\varphi\| \leq c_0 \}$$

for any $c_0 > 0$. ◊
1.1. CONCEPTS OF WELL-POSEDNESS – EXAMPLES OF ILL-POSED PROBLEMS

1.1.4 Example. (Differentiation of an approximately given function.)
Let \( u_1 \) be the derivative of a function \( z_1 \) on \([0, 1]\) and \( z_2(t) = z_1(t) + q \sin at \).

For \( Z = C[0, 1] \) we obtain
\[
\rho_Z(z_1, z_2) \leq |q|
\]
and for \( U = C[0, 1] \) and \( u_2 = \frac{dz_2}{dt} \) the distance between \( u_1 \) and \( u_2 \) is
\[
\rho_U(u_1, u_2) = \max_{t \in [0, 1]} |qa \cos at| = |qa|.
\]
Choosing \( q > 0 \) arbitrarily small and tending \( a \to \infty \), we conclude that the problem is ill-posed for the pair of spaces \( U = Z = C[0, 1] \).

It is easy to see that the problem becomes well-posed if we choose \( Z = C^1[0, 1] \) and \( U = C[0, 1] \).

It should be noted that in applications the spaces \( Z \) and \( U \) are in general not of our disposition, but are prescribed by the physical properties of the model involved. For example, the assumption that an approximately given function is close to an unknown exactly given function in the norm of the space \( C^1[0, 1] \) is in general not a natural hypothesis. In models for the geophysical detections of minerals, leading to a Cauchy Problem for the Laplace equation, the solution is of interest just for the pair of spaces considered in Example 1.1.3.

1.1.5 Example. (Fredholm integral equation of the first kind.)
On the pair of spaces \( U = C[0, 1], \ Z = L_2(0, 1) \) the solution of the integral equation
\[
Au(x) \equiv \int_0^1 K(x, s)u(s)ds = z(x)
\]
is sought with \( K \) a kernel continuous on \([0, 1] \times [0, 1]\).

Obviously, the functions \( z \in D_G = R_A \) must be continuous. Let \( u_1 \) be a solution for the fixed function \( z_1 \in D_G \). Then
\[
u_2(x) = u_1(x) + q \sin as
\]
will be the solution of the equation corresponding to the function
\[
z_2(x) = z_1(x) + q \int_0^1 K(x, s) \sin asds.
\]
The distances are
\[
\rho_U(u_1, u_2) = |q| \max_{s \in [0, 1]} |\sin as|,
\]
\[
\rho_Z(z_1, z_2) = |q| \left( \int_0^1 \left[ \int_0^1 K(x, s) \sin asds \right]^2 dx \right)^{1/2}.
\]
For arbitrary numbers \( q \) and \( a \to \infty \) we get
\[
\rho_U(u_1, u_2) \to |q| \quad \text{and} \quad \rho_Z(z_1, z_2) \to 0.
\]
If \( U = L_2(0, 1) \) is chosen, then \( \rho_U(u_1, u_2) \to \frac{|q|}{\sqrt{2}} \), i.e., the problem is ill-posed for pair \( U = C[0, 1], \ Z = L_2(0, 1) \) as well as for pair \( U = Z = L_2(0, 1) \).
In the examples above ill-posedness of the problems is due to the unboundedness of the operator $G$. However, in the latter problem, if $\lambda = 0$ appears to be an eigenvalue of the operator $A$, then the solution will be non-unique if it exists at all.

So-called inverse problems are other prominent examples of ill-posedness. They are, for instance, connected with the reconstruction of the physical environment by means of observation of the generated physical outcome. Important problems belong to this class like the analysis of physical experiments, the interpretation of observational data, tomography models and many others.

Inverse problems described by differential equations usually require to determine unknown operator coefficients on the basis of information about the solution of the equation itself.

We close this introductory section with two simple examples of convex variational problems which show that linear programming problems and linear semi-infinite programming problems are ill-posed in general, too.

1.1.6 Example. (Linear programming problem.)

The function $J(u) = u_2$ has to be minimized subject to

\begin{align*}
0 &\leq u_1 \leq 2, \quad 0 \leq u_2 \leq 2, \\
g_1(u) &= u_1 + u_2 - 1 \leq 0, \\
g_2(u) &= -u_1 - u_2 + 1 \leq 0.
\end{align*}

Obviously, the problem has the unique solution $u^* = (1, 0)$. Now, we consider the dependence of the solution $u \in U = \mathbb{R}^2$ on $g_2 \in Z = C[0, 2]$ near the given function $g_2$.

Replacing the constraint $g_2(u) \leq 0$ by

\[ g_2^\sigma(u) = -(1 + \sigma)u_1 - u_2 + 1 \leq 0, \]

we get

\[ \rho_Z(g_2, g_2^\sigma) = 2|\sigma| \to 0 \quad \text{for} \quad \sigma \to 0. \]

However, for $\sigma \in (-1, 0)$, the perturbed problem has the unique solution $u_2^\sigma = (0, 1)$, hence $\rho_U(u^*, u_2^\sigma) = \sqrt{2}$.

If, instead of $Z = C[0, 2]$ the set $Z = \mathbb{R}^{2 \times 2}$ is considered to describe the perturbations of the coefficients of the functions $g_1, g_2$, the result will be the same independently of the norm chosen in $Z$. ♦

1.1.7 Example. (Linear semi-infinite programming problem.)

Consider the linear semi-infinite problem (LSIP)

\[ \min \{-u_1 : -u_2 \leq 0, \ u_1 - tu_2 \leq 0 \quad \text{for all} \quad t \in T\}, \]

with $T = [0, 1] \subset \mathbb{R}$. Let us investigate its well-posedness subject to the variation of the set $T$, which is estimated in the Hausdorff-metric $\rho_H$, i.e., $Z$ is the space of subsets in $\mathbb{R}$ with this metric.

Obviously, the solutions of the original problems are given by the points $(0, a)$, $a \geq 0$. If $T_\delta \subset (0, 1]$ is a closed set with $\rho_H(T_\delta, T) \leq \delta$, the perturbed problem

\[ \min \{-u_1 : -u_2 \leq 0, \ u_1 - tu_2 \leq 0 \quad \text{for all} \quad t \in T_\delta\}, \]
1.2. WELL- AND ILL-POSED VARIATIONAL PROBLEMS

has no solution:

$$\inf J(u) = -u_1 = -\infty \text{ for arbitrary small } \delta > 0.$$  

♦

The question of well-posedness of convex, and in particular of linear, semi-infinite problems with respect to the variation of the set $T$ is closely related to the question whether they can be approximated in a stable way by finite-dimensional convex problems. Many numerical methods for solving semi-infinite programming problems are based on such approximations.

Examples of ill-posed elliptic variational inequalities will be considered in Sections 8.1 and 8.2.

The investigation of well-posedness in the sense of Definition 1.1.1 requires the transformation of the original problem into a model $u = G(z)$. In the examples considered above this should be done without difficulties. However, for specific classes of problems, using their natural formulations, it is possible to obtain more convenient definitions which widen the notion of well-posed problems for a given class.

In particular, the assumption of single-valuedness of the mapping $G$, usually obtained for linear problems by a suitable factorization, is very restrictive for nonlinear problems because in the case of their non-unique solvability their solution sets are not in general affine manifolds.

Moreover, in Example 1.1.4, we have seen that any problem of the corresponding class is ill-posed, whereas Example 1.1.6 shows that within the class of linear programming problems ill-posedness is more an exceptional case depending on accidental numerical data.

1.2 Well- and Ill-posed Variational Problems

1.2.1 Notations and definitions

Throughout this book within the class of convex functions we will always consider proper convex functions.

In the literature different notions of well-posedness of variational problems can be found (cf. Vasiljev [408], Karmanov [232]). We introduce here a definition which is sufficient flexible for our further considerations.

First we deal with the problem

$$\begin{align*}
\text{minimize} & \quad J(u), \\
\text{subject to} & \quad u \in K = \{ v \in U_0 : B(v) \leq 0 \}.
\end{align*}$$  

(1.2.1)

where $J : U \to \mathbb{R}$ is a convex, continuous functional on a Banach space $U$; $B : U \to Y$ is a convex, continuous mapping into a Banach space $Y$ and $U_0$ is a convex and closed subset of $U$.

Let $U^* = \text{Arg min}\{ J(u) : u \in K \}$ be the set of solutions of Problem (1.2.1) and $J^* = J(u^*)$ with $u^* \in U^*$ the optimal value of the problem.

In case that $U$ is a Hilbert space and $U_0 = U$, we obtain the convex variational problem (A1.5.11), (A1.7.40) (see Appendix) and for $U_0 = U$, $B : U \to \mathbb{R}^m$, a
problem of type (A1.7.35) is given.

For a fixed \( \delta > 0 \) we define the set of variations
\[
\Phi_\delta = \{ \phi_\delta \equiv (J_\delta, B_\delta) : \| J - J_\delta \|_{C(U_0)} \leq \delta, \max_{u \in U_0} \| B(u) - B_\delta(u) \|_Y \leq \delta \},
\]
where \( J_\delta : U \to \mathbb{R} \) and \( B_\delta : U \to Y \) are assumed to be continuous. With an element \( \phi_\delta \in \Phi_\delta \) the perturbed problem
\[
\begin{align*}
\text{minimize} & \quad J_\delta(u), \\
\text{subject to} & \quad u \in U_0, \ B_\delta(u) \leq 0
\end{align*}
\]
can be established. The feasible and optimal sets are denoted by
\[
U(\phi_\delta) = \{ u \in U_0 : B_\delta(u) \leq 0 \}
\]
and
\[
U^*(\phi_\delta) = \{ u^* \in U(\phi_\delta) : J_\delta(u^*) = \inf_{u \in U(\phi_\delta)} J_\delta(u) \},
\]
respectively.

1.2.1 Definition. Problem (1.2.1) is called weakly well-posed if
(i) solution set \( U^* \) is non-empty;
(ii) there exists a constant \( \delta_0 > 0 \) such that for any \( \delta \in (0, \delta_0) \) and any \( \phi_\delta \in \Phi_\delta \) the perturbed solution set \( U^*(\phi_\delta) \) is non-empty;
(iii) \( \lim_{\delta \downarrow 0} \sup_{u \in U^*(\phi_\delta)} \rho(u, U^*) = 0 \) holds for every \( \phi_\delta \in \Phi_\delta \).

Problem (1.2.1) is said to be well-posed if, moreover,
(iv) solution set \( U^* \) is a singleton, i.e. \( U^* = \{ u^* \} \).

If any of these conditions (i)-(iv) fails to hold, then the problem is considered as an ill-posed one. Hence, if condition (iv) is disturbed, weakly well-posed problems also belong to the class of ill-posed problems.

The perturbation class \( \Phi_\delta \) is related in particular to the convex problem
\[
\begin{align*}
\text{minimize} & \quad J(u), \\
\text{subject to} & \quad u \in K,
\end{align*}
\]
where \( J : V \to \mathbb{R} \) is a convex, lower semicontinuous functional, \( K \subset V \) is a convex, closed set and \( K \cap \text{dom} J \neq \emptyset \).

Naturally, the distance between the functions \( J_\delta \) and \( J \) as well as between \( B_\delta \) and \( B \) can be measured also in other metrics.

Defining \( \Phi_\delta \) in (1.2.2) it is often important to specify perturbations of the original information in view of the common methods applied for the resolution of the perturbed problems. For instance, in the case of convex variational problems convexity of the approximate problem is usually assumed, i.e. in the definition of \( \Phi_\delta \) it is supposed that \( J_\delta \) and \( B_\delta \) have to be convex.
1.2. WELL- AND ILL-POSED VARIATIONAL PROBLEMS

1.2.2 Remark. Sometimes it may be convenient to replace $U_0$ by a suitable subset $\tilde{U}_0 \subset U_0$ (with $U^* \cap \tilde{U}_0 \neq \emptyset$), implying an enlargement of $\Phi_\delta$. An analogous effect in other definitions of well-posedness is induced by a relaxation of the inequalities in (1.2.2) (cf. Vasiljev [408], Tikhonov / Arsenin [395]):

$$|J(u) - J_\delta(u)| \leq \delta(1 + \omega(u)), \quad \|B(u) - B_\delta(u)\|_Y \leq \delta(1 + \omega(u)),$$

where $u \in U_0$ and $\omega : U_0 \to \mathbb{R}$ is a non-negative functional such that its level set $\{u : \omega(u) \leq c\}$ is weakly compact for each constant $c$. ♦

When working with experimental data well-posedness of variational problems has to allow an efficient estimation of the propagated error in the solution induced by the error bound of the data.

For problems with exact data the arising perturbations are due to the numerical solution methods. In order to prove convergence, it is usually assumed that all the data are given with arbitrary accuracy.

Now we will investigate well-posedness of convex semi-infinite problems

\[(SIP) \quad \text{minimize} \quad J(u) \]
\[\text{subject to} \quad u \in K \equiv \{v \in \mathbb{R}^n : g(v, t) \leq 0 \ \forall \ t \in T\}, \quad (1.2.6)\]

where $T$ is a compact subset of a normed space; $J, g_{t \in T} : u \to g(u, t)$ are convex, finite-valued functions, $g_u : t \to g(u, t)$ is a continuous function on $T$ for any $u \in K$.

The problem is studied in detail in Appendix (see (A1.7.42)).

Setting either $B : u \to \max_{t \in T} g(u, t)$ or $B : u \to g(u, \cdot)$, we can apply Definition 1.2.1 (cf. also Corollary 1.2.7). However, to consider perturbations of the set $T$ within the whole class $\Phi_\delta$, the verification of the inclusion $\varphi_\delta \in \Phi_\delta$ often causes difficulties. On the other hand, as numerical methods always require a discretization of the set $T$, the study of well-posedness of Problem (1.2.6) subject to an approximation, here considered as a perturbation of the set $T$, is especially important.

Here we investigate well-posedness of convex SIP, when the class of perturbations of the set $T$ is given in the form

$$T_\delta = \{T_\delta \subset T : \rho_H(T, T_\delta) \leq \delta\}, \quad (1.2.7)$$

and everything else is given exactly.

Thus, for an arbitrary $\delta > 0$ and $T_\delta \subset T_\delta$, we consider the problem

\[
\begin{align*}
\text{minimize} & \quad J(u) \\
\text{subject to} & \quad u \in \mathbb{R}^n, \quad g(u, t) \leq 0 \ \forall \ t \in T_\delta.
\end{align*}
\]

Let $U^*(T_\delta)$ be the set of optimal solutions of (1.2.8).

1.2.3 Definition. Problem (1.2.6) is called weakly $T$-well-posed if

(i) $U^* \neq \emptyset$;
(ii) there exists a constant $\delta_0 > 0$ such that for each $\delta \in (0, \delta_0)$ and each $T_\delta \in \mathcal{T}_\delta$ the set $U^*(T_\delta)$ is non-empty;

(iii) $\lim_{\delta \downarrow 0} \sup_{u \in U^*(T_\delta)} \rho(u, U^*) = 0$ holds for arbitrarily chosen $T_\delta \subset \mathcal{T}_\delta$.

The problem is said to be $T$-well-posed if, moreover,

(iv) $U^* = \{u^*\}$.

If any of the conditions (i)-(iv) fails to hold, then Problem (1.2.6) is considered as ill-posed.

1.2.4 Remark. The introduction of these or other weaker definitions of well-posedness for different problems aims at separating an essential part of ill-posed problems. Mostly, if a convex SIP is ill-posed under a possible perturbation of all data, then it is usually ill-posed in the sense of Definition 1.2.3.

1.2.5 Remark. We comment on some alterations in the definition of well-posedness for variational problems given in the literature.

MOSCO [296] has investigated well-posedness of variational inequalities with monotone operators assuming that the sequence of feasible sets $\{K_i\}$ converges to the original feasible set $K$ such that

$\{v : v_i \rightharpoonup v \text{ for } v_i \in K_i\}$

$\quad = \{w : w_j \rightharpoonup w \text{ for } w_j \in K_{ij} \text{ and } \{K_{ij}\} \subset \{K_i\}\} = K$.

LEMAIRE [260], considering the original and the perturbed problems as unconstrained minimization problems, has described the relation between the data of these problems in the framework of the MOREAU-YOSIDA approximation. Indeed, for the problems

$\min \{\varphi(u) : u \in V\}$ and $\min \{\tilde{\varphi}(u) : u \in V\},$

with convex functions $\varphi, \tilde{\varphi} : V \to \mathbb{R}$ on the Hilbert space $V$, as a measure of the distance between $\varphi$ and $\tilde{\varphi}$ the value

$\sup_{\|u\| \leq \rho} |\varphi_\lambda(u) - \tilde{\varphi}_\lambda(u)|$

is introduced by means of the Moreau-Yosida functional

$\eta_\lambda(u) = \inf_{v \in V} \{\frac{1}{2\lambda} \|v - u\|^2 + f(u)\}$

with $\lambda > 0$, $\rho > 0$ suitably chosen.

Instead of the notion "weak well-posedness" in the literature often the notion well-posedness with respect to the argument is used.

Because sometimes the optimal value $J^* = \inf_{u \in K} J(u)$ may be of interest only, the notion well-posedness with respect to the objective is considered, i.e., for any choice of $\varphi_\delta \in \Phi_\delta$ (cf. (1.2.2)),

$\inf J_\delta(u) \to J^*$ as $\delta \downarrow 0$. 
In some papers the notions stable and well-posed are used synonymously. However, when speaking about non-stable problems, one usually has in mind a gap in the mapping $G$ given in Definition 1.1.1.

We shall apply the notion "stable" only in connection with numerical methods.

1.2.2 Some conditions of well-posedness

Now we consider Problem (1.2.1) with $B : U \to C(T) \equiv Y$ given by $B(u) = g(u, \cdot)$. The function $g$ is assumed to be continuous on $U \times T$ and convex with respect to $u$ for every $t \in T$ and $T$ is a compact subset of a Banach space. Depending on the cardinality of $T$ problems with a finite or infinite number of functional constraints are included.

Let us consider variations $B_\delta$, $J_\delta$ of $B$ and $J$, respectively, subject to

$$\max_{u \in U_0} \|B(u) - B_\delta(u)\|_Y \leq \delta,$$

$$|J - J_\delta|_{C(U_0)} \leq \delta.$$  \hspace{1cm} (1.2.9)

If $g_\delta$ is a continuous functional on $U \times T$ satisfying

$$\max_{u \in U_0} \|g(u, \cdot) - g_\delta(u, \cdot)\|_{C(T)} \leq \delta$$

then, for $B_\delta : u \to g_\delta(u, \cdot)$, inequality (1.2.9) holds.

Denote $\bar{g}(u) = \max\{g(u, t) : t \in T\}$.

1.2.6 Theorem. In the problem defined as above, let $U$ be a reflexive Banach space and Slater’s condition be satisfied, i.e., there exists a point $\tilde{u} \in U_0$ such that $\bar{g}(\tilde{u}) < 0$. Assume further that the functionals $J_\delta, B_\delta(\cdot)(t), t \in T$, are weakly lower semi-continuous, $U^*$ is non-empty and bounded and

$$\lim_{\delta \downarrow 0} \sup_{u \in W_\delta} \rho(u, U^*) = 0,$$  \hspace{1cm} (1.2.10)

where $W_\delta = \{u \in U_0 : J(u) \leq J^* + \delta, \bar{g}(u) \leq \delta\}$.

Then the problem is weakly well-posed in the class of variations (1.2.2).

Proof: We first prove that for

$$\bar{J}(\delta) = \inf\{J(u) : u \in U_0, \bar{g}(u) \leq -\delta\}$$

the relation $\lim_{\delta \downarrow 0} \bar{J}(\delta) = J^*$ is valid. This is obviously true if $\bar{g}(u^*) < 0$ for some $u^* \in U^*$.

If $U^* \cap \text{int}K = \emptyset$, then for arbitrarily chosen $u^* \in U^*$ and $u = \tilde{u} + (1 - \alpha)(u^* - \tilde{u})$ with $\alpha \in [0, 1]$, we obtain in view of the convexity of $\bar{g}$ that

$$\bar{g}(u) \leq \bar{g}(u^*) + \alpha(\bar{g}(\tilde{u}) - \bar{g}(u^*)) = \alpha\bar{g}(\tilde{u}).$$

Hence, $\bar{g}(u) \leq -\delta$ if $\alpha = -\frac{\delta}{\bar{g}(\tilde{u})} \leq 1$, and for $\delta \downarrow 0$ the point $u$ tends to $u^*$. The continuity of $J$ ensures that, indeed, $\bar{J}(\delta) \to J^*$. 

Now, we fix $\delta_0 > 0$ such that $W_{\delta_0}$ is bounded (because of (1.2.10) and the boundedness of $U^*$ such a required $\delta_0$ exists) and choose $\delta < \delta_0$ such that
\[
\bar{J}(\delta) - J^* + 3\delta < \delta_0,
\]
and $\delta \in (0, \min_{t \in T} |g(\bar{u}, t)|)$. For $\delta \in (0, \bar{\delta})$ and any $\varphi_\delta \in \Phi_\delta$ the level set $\{u \in U_0 : B_\delta(u) \leq 0\}$ is non-empty (obviously, $B_\delta(\bar{u}) < 0$) and is included in the set $\{u \in U_0 : \bar{g}(u) \leq \delta\}$ on account of inequality (1.2.9).

According to the definition of $\bar{J}(\delta)$ a point $\hat{u} \in U_0$ exists with
\[
J(\hat{u}) \leq \bar{J}(\delta) + \delta \quad \text{and} \quad \bar{g}(\hat{u}) \leq -\delta,
\]
hence
\[
J_\delta(\hat{u}) \leq J(\hat{u}) + \delta \leq \bar{J}(\delta) + 2\delta \quad \text{and} \quad B_\delta(\hat{u}) \leq 0
\]
proving that the system of relations
\[
J_\delta(u) \leq \bar{J}(\delta) + 2\delta, \quad B_\delta(u) \leq 0, \quad u \in U_0
\]
is solvable. Obviously, its solution set is contained in
\[
\bar{U}_\delta = \{u \in U_0 : J(u) \leq \bar{J}(\delta) + 3\delta, \; \bar{g}(u) \leq \delta\},
\]
and $\bar{U}_\delta \subset W_{\delta_0}$ holds in view of (1.2.11) and the inequality $\bar{J}(\delta) + 3\delta \leq \bar{J}(\delta) + 3\delta$ which follows from the definition of $\bar{J}(\delta)$.

Therefore, $\bar{U}_\delta$ is bounded and even weakly compact due to the reflexivity of $U$ and convexity and continuity of the functionals $J$ and $\bar{g}$. Having this, in view of the lower semicontinuity of the functionals $J_\delta$ and $B_\delta(\cdot)(t), t \in T$, one can conclude that the functional $J_\delta$ attains its minimum on the set $\{u \in \bar{U}_\delta : B_\delta(u) \leq 0\}$. Because of the equality
\[
\inf\{J_\delta(u) : u \in U_0, B_\delta(u) \leq 0\} = \min\{J_\delta(u) : u \in \bar{U}_\delta, B_\delta(u) \leq 0\}
\]
and the inclusion $\bar{U}_\delta \subset U_0$ we obtain that the infimum on the left-hand side is attained, proving that $U^*(\varphi_\delta) \neq \emptyset$ and $U^*(\varphi_\delta) \subset \bar{U}_\delta \subset W_{\delta_0}$.

Finally, due to condition (1.2.10) it follows
\[
\lim_{\delta \to 0} \sup_{u \in U^*(\varphi_\delta)} \rho(u, U^*) = 0.
\]

If $U$ is a finite-dimensional space, then condition (1.2.10) is a consequence of the remaining assumptions in Theorem 1.2.6).

1.2.7 Corollary. Assume that for the convex SIP (1.2.6) the following conditions hold:

(i) there exists a Slater point;

(ii) $U^* \neq \emptyset$ and bounded;
1.2. WELL- AND ILL-POSED VARIATIONAL PROBLEMS

(iii) there exist positive constants \( L \) and \( \alpha \) such that

\[
\sup_{u \in U_\alpha} |g(u, t') - g(u, t'')| \leq L \|t' - t''\|_Y \quad \forall t', t'' \in T
\]

holds for \( U_\alpha = \{ u : \rho(u, U^*) \leq \alpha \} \).

Then SIP (1.2.6) is weakly well-posed for the class of variations \( T_\delta \) given in (1.2.7) as well as for the class

\[
\{ \varphi_\delta = (J_\delta, g_\delta, T_\delta) : \|J - J_\delta\|_{C(R^n)} \leq \delta, \max_{t \in T} \|g(\cdot, t) - g_\delta(\cdot, t)\|_{C(R^n)} \leq \delta, T_\delta \subset T_\delta \}
\]

with continuous functionals \( J_\delta \) and \( g_\delta \).

In this context the following lemma is easy to prove.

1.2.8 Lemma. Assume that \( U_0 \) is convex and closed, the functionals \( J \) and \( g(\cdot, t) \) are convex and continuous, Slater’s condition is satisfied and the set \( W_{\delta_0} \) is bounded for some \( \delta_0 > 0 \). Then

\[
\lim_{\delta \downarrow 0} \sup_{u \in W_\delta} \rho(u, K) = 0.
\]

If, moreover, \( \inf_{u \in U_0} J(u) < J^* \), then for \( S = \{ u \in U_0 : J(u) \leq J^* \} \) the relation

\[
\lim_{\delta \downarrow 0} \sup_{u \in W_\delta} \rho(u, S) = 0
\]

holds true.

Nevertheless, the following example in infinite-dimensional spaces shows that

\[
\lim_{\delta \downarrow 0} \sup_{u \in W_\delta} \rho(u, U^*) \neq 0
\]

is possible.

1.2.9 Example. In Problem (1.2.1) let

\[
U = U_0 = \ell_2, \quad J(u) = \sum_{i=1}^{\infty} i^{-1} u_i^2,
\]

\[
B(u) = \max\{g_1(u), g_2(u)\} \quad \text{with} \quad g_1(u) = u_1 + 1, \quad g_2(u) = \sum_{i=1}^{\infty} u_i^2 - 2.
\]

Obviously, the solution set is \( U^* = \{(-1, 0, ..., 0, ...)\} \) and the assumptions of Theorem 1.2.6 are satisfied with exception of (1.2.10). The set \( W_\delta \) is bounded for any \( \delta > 0 \) and \( \inf\{J(u) : u \in U_0\} = 0 < J^* \). However, for the sequence of points \( u^k = (-1, 0, ..., 1, 0, ...) \in W_{\delta_k} \) with \( \delta_k = \frac{1}{k} \) we have

\[
\rho(u^k, K) = 0, \quad \rho(u^k, S) \to 0,
\]

whereas \( \lim_{k \to \infty} \rho(u^k, U^*) = 1 \).\[\diamondsuit\]
It should be noted that the verification of condition (1.2.10) is a common step in the investigation of the convergence of a generalized minimizing sequence (cf. Vainberg [405], Levitin / Polyak [263]).

Now we give a sufficient condition being important for the analysis of weak well-posedness of variational inequalities in infinite-dimensional spaces.

1.2.10 Lemma. Let $V_1$ be a finite-dimensional subspace of a Hilbert space $V$, $\Pi_1 : V \to V_1$ an orthoprojector on $V_1$ (i.e. $\Pi_1 u = \arg\min\{\|u - v\| : v \in V_1\}$, for any $u \in V$), and $\Pi_2 = I - \Pi_1$, where $I$ denotes the identity operator on $V$.

Assume moreover that, for any $u_1, u_2 \in V$ and any $\lambda \in [0, 1]$, the functional $J$ in Problem (1.2.1) (with $U = V$) satisfies the inequality

$$J(\lambda u_1 + (1 - \lambda) u_2) \leq \lambda J(u_1) + (1 - \lambda) J(u_2) - \lambda(1 - \lambda) q(\|\Pi_2 u_1 - \Pi_2 u_2\|),$$

(1.2.12)

where $q : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly increasing function with $q(0) = 0$.

Then relation (1.2.10) is a conclusion of the remaining assumptions in Theorem 1.2.6.

Proof: At first we verify that, for some $\delta_0 > 0$ the set $W_{\delta_0}$ is bounded. Due to the Slater condition there exists a constant $a > 0$ (cf. Remark A3.4.44) such that for $\bar{\psi}(u) = J(u) + a\max\{0, \bar{g}(u)\}$ the relations

$$\text{Arg min}_{u \in V} \bar{\psi}(u) = U^*, \quad \min_{u \in V} \bar{\psi}(u) = J^*$$

(1.2.13)

hold. Denote $\Gamma = \{u : \rho(u, U^*) = c\}$ with $c > 0$ fixed and suppose that

$$\inf_{u \in \Gamma} \bar{\psi}(u) = J^* + c_1, \text{ with } c_1 > 0.$$

Because of the convexity of $\bar{\psi}$ the inclusion

$$\{u : \bar{\psi}(u) \leq J^* + c_1\} \subset \{u : \rho(u, U^*) \leq c\}$$

is satisfied and with $\delta_0 = \frac{c_1}{1 + a}$ it follows that

$$W_{\delta_0} \subset \{u : \bar{\psi}(u) \leq J^* + c_1\}.$$

Now, assume that

$$\inf_{u \in \Gamma} \bar{\psi}(u) = J^*$$

(1.2.14)

would be true. Obviously, inequality (1.2.12) is fulfilled for the functional $\bar{\psi}$ instead of $J$. Hence, for $u^* = \arg\min_{u \in V} \bar{\psi}(u)$, we obtain

$$\bar{\psi}(u) - \bar{\psi}(u^*) \geq q(\|\Pi_2 u - \Pi_2 u^*\|).$$

(1.2.15)

Choosing a weakly convergent sequence $\{v^i\} \subset \Gamma$ with

$$\lim_{i \to \infty} \bar{\psi}(v^i) = J^*,$$

(1.2.16)

then, due to (1.2.15), it follows that

$$\bar{\psi}(v^i) - \bar{\psi}(u^*) \geq q(\|\Pi_2 v^i - \Pi_2 u^*\|),$$
1.2. WELL- AND ILL-POSED VARIATIONAL PROBLEMS

and in view of (1.2.16) and the properties of the functional \( q \) the relation

\[
\lim_{i \to \infty} \| \Pi_2 v^i - \Pi_2 u^* \| = 0
\]

holds. On account of the weak convergence of \( \{ v^i \} \) and the finite dimensionality of \( V_1 \) we conclude that

\[
\lim_{i \to \infty} \| v^i - u^* \| = 0
\]

in contradiction to the chosen sequence \( \{ v^i \} \). Consequently, the case (1.2.14) is impossible, and thus, boundedness of \( W_{\delta_0} \) has been shown.

Let \( \delta_i \downarrow 0 \) and \( u_i \in W_{\delta_i} \). Since \( W_{\delta_i} \) is bounded, the sequence \( \{ u_i \} \) is also bounded and without loss of generality we can assume that \( u_i \rightharpoonup v \). Due to its convexity and continuity, the functional \( \rho(\cdot, K) \) is weakly lower semi-continuous and from

\[
\lim_{\delta \downarrow 0} \sup_{u \in W_{\delta}} \rho(u, K) = 0
\]

we obtain

\[
\lim_{i \to \infty} \rho \left( \frac{u^i + v}{2}, K \right) = \lim_{i \to \infty} \rho \left( u^i, K \right) = \rho(v, K) = 0.
\]

Hence, \( J(v) \geq J^* \). Since \( J \) is a weakly lower semicontinuous functional, we get

\[
\lim_{i \to \infty} J(u^i) \geq J(v), \quad \lim_{i \to \infty} J \left( \frac{u^i + v}{2} \right) \geq J(v).
\]

However, from the definition of \( W_{\delta_i} \) it follows that \( \lim_{i \to \infty} J(u^i) \leq J^* \).

Thus, \( J(v) = J^* \) and from

\[
J \left( \frac{u^i + v}{2} \right) \leq \frac{1}{2} J(u^i) + \frac{1}{2} J(v) - \frac{1}{4} \rho(\| \Pi_2 u^i - \Pi_2 v \|)
\]

we conclude that

\[
\| \Pi_2 u^i - \Pi_2 v \| \to 0,
\]

and finally, in view of the finite dimensionality of \( V_1 \), the relation

\[
\| \Pi_1 u^i - \Pi_1 v \| \to 0
\]

holds. \( \square \)

The verification of the assumptions of Theorem 1.2.6 shows that in Example 1.1.6 the Slater condition fails to hold, whereas in Example 1.1.7 the required boundedness of \( U^* \) is disturbed. Moreover, both problems are not weakly well-posed with respect to variations of the data in the class \( \Phi_\delta \). From the analysis of Example 1.2.9 it is clear that condition (1.2.10) is not satisfied. Indeed, taking \( \delta_k = \frac{1}{k} \) and

\[
J_{\delta_k}(u) = \sum_{i=1}^{\infty} \frac{1}{\delta_k} u^2_i + \max \left\{ \frac{1}{k} u^2_i, \frac{1}{k} \right\},
\]

it is easy to see that this problem is not weakly well-posed, too.
1.3 Approaches of Stabilizing Ill-posed Problems

1.3.1 Stabilizers and quasi-solutions

The idea of stabilizing functionals or stabilizers is a fundamental tool for solving ill-posed problems. Their formal introduction does not require a specification of the problems under consideration. We only need a metric space $U$ in which the solution is sought and some information about the solutions. The content of this information is more or less reflected in the definition of the stabilizer.

In the sequel some standard notations and properties of Sobolev spaces will be used (cf. Section A1.3 in Appendix). The symbols $\langle \cdot, \cdot \rangle_m,\Omega$ (respectively $\langle \cdot, \cdot \rangle_m,\Omega,0$ and $\| \cdot \|_{m,\Omega,0}$) denote the scalar product and the norm of the spaces $H^m(\Omega)$ and $H^m_0(\Omega)$, respectively.

1.3.1 Definition. A functional $\omega$ defined on the set $D_\omega \subset U$ is called a stabilizer of a given problem if

1. $\omega(u) \geq 0$ for all $u \in D_\omega$;
2. set $\omega_c = \{ u \in U : \omega(u) \leq c \}$ is a compact subset of the metric space $U$ for any $c \geq 0$;
3. $U^*_{\omega} \equiv D_\omega \cap U^* \neq \emptyset$, where $U^*$ is the solution set of the problem.

The functional $\omega$ is said to be a weak stabilizer, if condition (ii) is replaced by

(ii)' set $\omega_c$ is a weakly compact subset of the space $U$ for any $c \geq 0$.

It should be noted that in this context the lower semi-continuity of the functional $\omega$ follows from assumption (ii) and weak lower semi-continuity of $\omega$ can be concluded from (ii)'.

A simple example of a stabilizer in the Euclidean space $\mathbb{R}^n$ is given by the functional $\omega(u) = \| u - \bar{u} \|^2_2$, where $\bar{u} \in \mathbb{R}^n$ is an arbitrary element.

In the Banach space $C[a,b]$, assuming that $U^* \cap H^1(a,b) \neq \emptyset$, the functional $\omega(u) = \| u - \bar{u} \|^2_{1,1}(a,b)$ is a stabilizer with any $\bar{u} \in H^1(a,b)$.

In the case $\Omega \subset \mathbb{R}^n$ this follows immediately from the compactness of the embedding of the spaces $H^m(\Omega)$ in $C(\Omega)$ for $n < 2m$ (cf. Appendix, Section A1.3).

Due to the compactness of the embedding of $H^{m+1}(\Omega)$ in $H^m(\Omega)$ (independently of the dimension of $\Omega$), the functional $\omega(u) = \| u - \bar{u} \|_{m+1,\Omega}$ is a stabilizer in the space $H^m(\Omega)$ if $U^* \cap H^{m+1}(\Omega) \neq \emptyset$ and $\bar{u} \in H^{m+1}(\Omega)$.

In a reflexive Banach space $U$, in view of the weak compactness of a sphere, the functional $\omega(u) = \| u - \bar{u} \|_{U}^2$ satisfies the properties of a weak stabilizer.

Analyzing ill-posed equations of the type

$$Au = z_0,$$

one is faced with the difficulty that small perturbations of the function $z_0$ may exceed the range of the operator $A$ such that classical solutions no longer exist for the perturbed equation.

In order to deal with this difficulty, Ivanov [195], Ivanov et Al. [196] has suggested the following generalization of the notion of a solution.
1.3. APPROACHES OF STABILIZING ILL-POSED PROBLEMS

1.3.2 Definition. An element $u$ attaining the minimum of the residue $\rho_Z(Au, z_0)$ on a set $M$ is called a quasi-solution of the equation $Au = z_0$ on $M$.

Obviously, if $z_0 \in AM$, then the corresponding quasi-solution coincides with the solution in the classical sense.

The assumption that $\rho_Z$ attains its minimum on the set $M$ is an essential restriction with respect to the choice of $M$. In some papers the condition that $M$ is the set of well-posedness of the considered problem is part of the definition of the quasi-solution.

1.3.2 Classical approaches for stabilization of ill-posed problems

In order to introduce the basic ideas of the classical methods for solving ill-posed problems we choose as an initial model the operator equation of the first kind

$$Au = z_0,$$  \hspace{1cm} (1.3.1)

with $A$ a continuous operator. The integral equation in Example 1.1.5 is a representative of such a problem.

We assume that the operator $A$ is given exactly and that Problem (1.3.1) is solvable for the exact right-hand side $z_0$. Instead of $z_0$ an approximation $z_\delta$ is considered such that $\rho_Z(z_0, z_\delta) \leq \delta$.

Given $(z_\delta, A, \delta)$ an element $u_\delta$ is sought such that for $\delta \downarrow 0$ we get convergence of $u_\delta$ to the solution $u^*$ (or to the solution set $U^*$) of (1.3.1) in a suitable sense.

The following three variational approaches for constructing such $u_\delta$ are well-known and guarantee the required properties.

**Quasi-solution method:**

Minimize the residue $\rho_Z(Au, z_\delta)$ on a set $M \subset U$ selected such that it contains a solution of Problem (1.3.1). For the moment let $M$ be compact.

If $\omega$ is a stabilizer and $\{u \in U : \omega(u) \leq c\} \cap U^* \neq \emptyset$, a possible choice of $M$ is the set $\omega_c = \{u \in U : \omega(u) \leq c\}$.

**Residual method:**

Minimize the stabilizer $\omega$ on the set

$$\{u \in D_\omega : \rho_Z(Au, z_\delta) \leq \varphi(\delta)\},$$

where $\varphi(\delta) \geq \delta$, $\varphi(\delta) \to 0$ for $\delta \downarrow 0$.

**Tikhonov regularization method:**

Minimize the functional $\rho_Z^2(Au, z_\delta) + \alpha \omega(u)$ on its definition domain for a suitably chosen value $\alpha(\delta) > 0$ of the parameter $\alpha$.

Obviously, in order to get a solution of the original Problem (1.3.1), it is necessary that $\delta$ tends to zero.

Due to the continuity and non-negativity of the function $\rho_Z$ and in view of the compactness of the sets $M$ and $\omega_c$, in each of the three methods above it holds that for any $z_\delta$ and any $c$ there exists an element $u_\delta$. Moreover, if
the operator \( A \) is linear and a weak stabilizer \( \omega \) is used, then the transformed problems have a solution, too.

For the investigation of the convergence of \( u, \delta \downarrow 0 \), the following result is important.

**1.3.3 Proposition.** Let \( U \) and \( Z \) be Banach spaces, \( M \subset D_T \subset U \) be compact, and the continuous operator \( T : U \to Z \) is assumed to be an injection from \( M \) onto \( TM \). Then the inverse operator \( T^{-1} \) is continuous on \( TM \).

**1.3.4 Remark.** If the operator \( T : U \to Z \) is linear, continuous and injective and \( M \) is a composition such that

\[
M = M' + M'' \subset D_T \subset U, \quad M' \cap M'' = \emptyset \quad \text{or} \quad M' \cap M'' = \{0\},
\]

with \( M' \) a compact set and \( M'' \) a finite-dimensional subspace of \( U \), then \( T^{-1} \) is continuous on \( TM \).

In particular, if the operator \( A \) in (1.3.1) satisfies the hypothesizes in Proposition 1.3.3, we obtain immediately the convergence of the quasi-solution method. However, if in addition it is supposed that \( A \) is linear, the set \( M \) is convex, and the space \( Z \) is strongly convex (which is the case if \( Z \) is a Hilbert space), then the problem

\[
\min_{u \in M} \rho_Z(Au, z)
\]

is well-posed for any \( z \).

Now we come back to the variational problem (1.2.1) and the class of variations \( \Psi_{\delta} \) given by (1.2.2). Denote \( \chi(\cdot|Q) \) the indicator functional to \( Q \), i.e.,

\[
\text{ind}(u|Q) = \begin{cases} 
0 & \text{if } u \in Q, \\
+\infty & \text{if } u \notin Q.
\end{cases}
\]

By substituting formally \( J_\delta(u) + \text{ind}(u|U(\varphi_\delta)) \) for \( \rho_Z(Au, z_\delta) \) quasi-solution and regularization methods can be formulated analogously for Problem (1.2.1), whereas, for an analogue to the residual methods, we have to replace \( \rho_Z(Au, z_\delta) \) by

\[
J_\delta(u) + \text{ind}(u|U(\varphi_\delta)) - J^*.
\]

It should be noted that if \( A \) is a linear continuous operator, then the functionals \( \|Au - z\| \) and \( \|Au - z\|^2 \) are convex and continuous and the results obtained below for Problem (1.2.1) can be re-translated to the equation \( Au = z \).

To make use of the generally accepted notion solution methods for ill-posed problems in the approaches considered above, it should be taken into account that, in fact, we deal only with stable approximations of ill-posed problems. This enables us to use numerical methods which are designed for well-posed problems.

**1.3.5 Proposition.** For variational problem (1.2.1) assume that the optimal set \( U^* \) is non-empty, that the constraint \( u \in K \) is given exactly and that the approximation \( J_\delta \) of the objective functional \( J \) satisfies \( \|J - J_\delta\|_{C(U_0)} \leq \delta \). Moreover, supposed that the parameters \( \alpha \) and \( \delta \) in the Tikhonov regularization method are chosen such that \( \frac{\delta}{\alpha} \leq c \), with fixed \( c \) and \( \alpha \downarrow 0 \).

Then
(i) for any stabilizer $\omega$ each of the three methods is defined, the corresponding sequence $\{u_\delta\}$, $\delta \downarrow 0$, belongs to some compact set and any cluster point of this sequence belongs to $U^*$;

(ii) if $\omega$ is a weak stabilizer and $J_\delta$ is a weakly lower semicontinuous functional, then $\{u_\delta\}$ is contained in a weakly compact set and the weak cluster points belong to $U^*$.

Proof: Here we do not use the detailed description of the feasible set $K$ given in (1.2.1). It suffices to assume that $K$ is a convex, closed set and that

$$\|J - J_\delta\|_{C(\hat{K})} \leq \delta,$$

where $\hat{K}$ is some set containing $K$. Let $\omega$ be a stabilizer according to Definition 1.3.1.

For quasi-solution methods, due to the definition of $u_\delta$, one gets $u_\delta \in M$ and $J_\delta(u_\delta) \leq J_\delta(u^\ast)$ with $u^\ast \in M \cap U^\ast$. Hence, since $\|J - J_\delta\|_{C(U_0)} \leq \delta$, it holds

$$J(u_\delta) \leq J^* + 2\delta, \quad u_\delta \in M.$$  
(1.3.2)

For the residual method, in view of the inequalities

$$\delta \leq \varphi(\delta), \quad \|J - J_\delta\|_{C(U_0)} \leq \delta,$$

it follows that

$$J_\delta(u^\ast) \leq J^* + \varphi(\delta) \quad \text{for any} \quad \delta \text{ and} \quad u^\ast \in U^\ast \cap D_\omega.$$  

Consequently, observing the definition of $u_\delta$, all points $u_\delta$ have to belong to the compact set $\omega^* = \{u : \omega(u) \leq \omega(u^\ast)\}$. On account of $J_\delta(u_\delta) \leq J^* + \varphi(\delta)$ we have

$$J(u_\delta) \leq J^* + \varphi(\delta) + \delta, \quad u_\delta \in \omega^*.$$  
(1.3.3)

Finally, consider the case that $u_\delta$ is defined by means of the Tikhonov regularization method. Let $Q = \{u : \omega(u) \leq \omega(u^\ast) + 2\delta\}$. For any point $u \in K \setminus Q$, due to

$$J_\delta(u) - J_\delta(u^\ast) \geq -2\delta \quad \text{for} \quad u \in K,$$  
(1.3.4)

we have

$$J_\delta(u) + \alpha \omega(u) > J_\delta(u^\ast) + \alpha \omega(u^\ast).$$

Hence, $u_\delta \in Q$. Because $2\delta \leq 2\delta$, the sequence $\{u_\delta\}$ belongs to the set

$$\omega^*_2 = \{u : \omega(u) \leq \omega(u^\ast) + 2\delta\}.$$  

Together with the definition of $u_\delta$ this yields

$$J(u_\delta) + \alpha \omega(u_\delta) \leq J(u^\ast) + \alpha \omega(u^\ast) + 2\delta, \quad u_\delta \in \omega^*_2.$$  
(1.3.5)

Now, from the relations (1.3.2), (9.3.15) and (1.3.5) it follows that for each of the methods the sequence $\{u_\delta\}$ is contained in some compact set. Choosing convergent subsequences and taking limits in the inequalities (1.3.2), (9.3.15) and (1.3.5), respectively, the assertion follows.

In the case that $\omega$ is a weak stabilizer, the arguments are analogous.
weak lower semicontinuity of $J_\delta$ yields the existence of $u_\delta$.

Under the assumptions on the stabilizer made so far, it cannot be expected to obtain stronger results on the convergence, since we cannot exclude that the set

$$U^* \cap \{ u : \omega(u) \leq \omega(u^*) \}$$

contains more than one point, and it is even possible that

$$U^* \subset \{ u : \omega(u) \leq \omega(u^*) \}.$$

But, if the hypothesis of Theorem 1.2.6 hold and, additionally, the functional $J_\delta$ is convex and the stabilizer $\omega$ is strongly convex and weakly lower semicontinuous, then, due to Theorem 1.2.6, the existence of a value $\delta_0 > 0$ is ensured such that, for $\delta \in (0, \delta_0]$, the problem of finding an element $u_\delta$, determined by the residual as well as by the regularization method, is well-posed.

The following result shows some relations between the iterates of the considered methods. We emphasize that convexity properties of the original problem are not used and that the result can be proved under rather weak assumptions on the functional $J_\delta$ (cf. Liskovec [275], chapt. 2).

1.3.6 Proposition. Denote

$$u^\alpha_\delta = \arg \min_{u \in U} \{ J_\delta(u) + \ind(u|K) + \alpha \omega(u) \}.$$

Then $u^\alpha_\delta$ is a quasi-solution for $M = \omega_c$ with $c = \omega(u^\alpha_\delta)$.

The same solution $u^\alpha_\delta$ can also be obtained by means of the residual method if $\varphi(\delta) = J_\delta(u^\alpha_\delta) - J^*$ is chosen.

If $\hat{u}_\delta$ is uniquely determined by means of the residual method, then this point is also a quasi-solution for $M = \omega_c$ with $c = \omega(\hat{u}_\delta)$.

1.3.7 Remark. One should be aware that the set $M \cap U^*$ with

$$M = \{ u : \omega(u) \leq \omega(u^\alpha_\delta) \}$$

may be empty. Indeed, in the example with

$$J(u) = (u - 1)^2, \quad K = \mathbb{R}, \quad \omega(u) = u^2, \quad J_\delta = J,$$

we get $u^\alpha_\delta = \frac{1}{1 + \alpha}$, and $M$ does not contain the point $u^* = 1$ which is the unique solution of the original problem.

This example also demonstrates that the point $\hat{u}_\delta$ is not always obtainable by means of the quasi-solution method. However, $u^\alpha_\delta$ and $\hat{u}_\delta$ are quasi-solutions in the sense of Definition 1.3.2.

Therefore, from Proposition 1.3.6 we can conclude that the quasi-solution method is the most general one and this is emphasized by a number of papers.

The question of the practical performance of the described methods is more important in our context. The regularizing method seems to be the simplest one. In the other two methods, in addition to the condition $u \in K$, restrictions of the type $u \in M$ or $J_\delta(u) - J^* \leq \varphi(\delta)$ are imposed.

The choice of a suitable stabilizer $\omega$ or a compact set $M$, which contains a solution point, is usually very difficult in practice. For example, in some elliptic
1.3. APPROACHES OF STABILIZING ILL-POSED PROBLEMS

variational inequalities with a linear differential operator of the second order (cf. Appendix A2.1 and Chapter 8), we have to minimize a quadratic functional

\[ J(u) = \frac{1}{2} \langle \nabla u, \nabla v \rangle_{0, \Omega} - \langle f, u \rangle_{0, \Omega} \]

(or even a more general functional \( J \)) on a closed and convex set \( K \subset H^1(\Omega) \). For such a problem usually a stabilizer \( \omega(u) = \| u - \bar{u} \|^2_{2, \Omega} \) is used and there is no hope that another stabilizer could be chosen which imposes weaker conditions on the smoothness of the function \( u \). In this case the order of the differential operator, corresponding to the functional to be minimized in the regularization or residual method, is higher than the order of the initial differential operator, and a solution has to be sought in the function space \( H^2(\Omega) \) but not in \( H^1(\Omega) \), where the initial problem is given. In particular, this excludes the use of piecewise affine approximations in the finite element method and requires to apply higher order approximations leading to a greater numerical expense.

On should also be aware, however, that due to the lower order of smoothness of the solution of the variational inequality, the use of higher order elements will not lead to an improvement of the asymptotical behavior of the error (cf. the end of Appendix A2.2 and Chapter 8).

Similar difficulties arise in quasi-solution methods if the constraint \( \| u - \bar{u} \|^2_{2, \Omega} \leq c \) is used for the construction of the set \( M \). It should be noted that for some other classes of problems, like integral equations, differentiation of functions, spline approximation etc., some results are known (cf. Lavrentjev / Romanov / Shishatski [259], Hofmann [185]), which enable us to construct stabilizers or compact sets of well-posedness on the basis of additional a priori information. On this way the structure of the auxiliary problems, which have to be solved by applying the above methods, is not essentially more complicated than in the original problem.

In the sequel we analyze Tikhonov’s regularization method based on the use of weak stabilizers. In Section 1.3.3 more restrictive rules for the adaption of the parameters \( \alpha \) and \( \delta \) will be given such that convergence in the metric \( \rho^*_U \) of the minimizing sequence \( \{ u_\delta \} \) to the optimal set \( U^* \) or to a point in \( U^* \) can be proved.

In our opinion, for convex variational problems, the quasi-solution and residual methods are less efficient, a conclusion which may be precipitate, but it is based on the insufficient algorithmical development of these methods for nonlinear problems. As mentioned above, in the quasi-solution method, using a weakly compact set \( M \), we can only expect weak convergence of \( \{ u_\delta \} \). In Vasiljev [408], [chapt. 2, sect. 7] a modification with successive adaption of the set \( M \) is given which enforces convergence in the metric \( \rho^*_U \). However, the numerical procedure is essentially more complicated in comparison with the original version.

In the residual method the value \( J^* \) has to be known a priori or has to be specified during the solution process which leads to an additional expense. However, for the solution of equations and some special variational problems the value \( J^* \) is mostly known.
Then the iterates \( u_i \) are assumed to be weakly lower semicontinuous and (1.3.6) – (1.3.8) fulfilled, where

\[
\|J - J_i\|_{C(D_\omega)} \leq \delta_i, \tag{1.3.6}
\]

\[
\max_{u \in D_\omega} |B(u) - B_i(u)|_1 \leq \sigma_i \leq \delta_i, \tag{1.3.7}
\]

\[
\|\omega - \omega_i\|_{C(D_\omega)} \leq \delta_i, \tag{1.3.8}
\]

where \( |v| = \max_{1 \leq k \leq m} |v_k| \) and \( \{\delta_i\} \) are sequences of non-negative numbers converging to 0 and \( \sup \delta_i < 1 \). Let

\[
K_i = \{ u \in D_\omega : B_i(u) - \tau_i \varnothing \leq 0 \}, \tag{1.3.9}
\]

\[
\theta_i(u) = J_i(u) + \alpha_i \omega_i(u), \tag{1.3.10}
\]

with \( \varnothing = (1, ..., 1)^T \in \mathbb{R}^n \), \( \tau_i = \frac{\sigma_i}{1 + \alpha_i} \), \( \alpha_i > 0 \), \( \alpha_i \downarrow 0 \).

Fixing a sequence \( \{\epsilon_i\} \) with \( \epsilon_i > 0 \), \( \epsilon \downarrow 0 \), the sequence of iterates \( \{u^i\} \) be defined by

\[
u^i \in K_i, \quad \theta_i(u^i) \leq \inf_{u \in K_i} \theta_i(u) + \epsilon_i, \quad i = 1, 2, ... \tag{1.3.11}\]

**1.3.8 Definition.** A point \( u^* \in U_\omega^* \equiv D_\omega \cap U^* \) is called \( \omega \)-normal solution of Problem (1.2.1) if \( \omega(u^* \omega^*) = \inf \{ \omega(u) : u \in U_\omega^* \} \).

The set of \( \omega \)-normal solutions will be denoted by \( U_\omega^* \). Since \( U^* \) is weakly closed, because of its convexity and closedness, and the set

\[
\omega^* = \{ u : \omega(u) \leq \omega(u^*) \}
\]

is weakly compact for a fixed \( u^* \in U^* \), the set \( U^* \cap \omega^* \) is weakly compact. Together with the lower semicontinuity of \( \omega \) this proves that \( U_\omega^* \neq \emptyset \).

**1.3.9 Proposition.** The original assumptions concerning Problem (1.2.1) let be fulfilled, where \( U \) is a Hilbert space and \( Y = \mathbb{R}^n \). Assume further that the Slater condition is satisfied, \( U^* \neq \emptyset \) and that \( \omega \) is strongly convex (with a constant \( \kappa > 0 \)), non-negative and continuous on \( D_\omega = U_0 \). Moreover, the approximations \( J_i, \omega_i, B_i \) are assumed to be weakly lower semicontinuous and (1.3.6) – (1.3.8) hold together with

\[
\lim_{i \to \infty} (\delta_i + \alpha_i + \epsilon_i) = 0, \quad \lim_{i \to \infty} \frac{\epsilon_i + \delta_i}{\alpha_i} = 0. \tag{1.3.12}
\]

Then the iterates \( u^i \), defined by (1.3.9) – (1.3.11), exist for all \( i \) and the sequence \( \{u^i\} \) converges in the norm of the space \( U \) to the unique \( \omega \)-normal solution \( u_\omega^* \) of Problem (1.2.1).
1.3. APPROACHES OF STABILIZING ILL-POSED PROBLEMS

For the sake of simplicity, we will prove this statement only for the case that the feasible set \( K \) is given exactly. For a complete proof of Proposition 1.3.9 the reader is referred to Vasiljev [408], Thm. 2.8.1.

**Proof:** Since \( \omega \) is strongly convex and weakly lower semicontinuous, the set \( \omega_d = \{ u : \omega(u) \leq d \} \) is bounded and weakly closed for all \( d \). Hence, \( \omega \) is a weak stabilizer. In view of \( \alpha_i \equiv 0 \) and (1.3.11) we have

\[
 u^i \in K, \quad \theta_i(u^i) \leq \inf_{u \in K} \theta_i(u) + \epsilon_i, \quad i = 1, 2, \ldots \tag{1.3.13}
\]

Due to the inequalities (1.3.6) – (1.3.8), we obtain for arbitrary \( u \in K \)

\[
 \theta_i(u) \geq J(u) + \alpha_i \omega(u) - \delta_i(1 + \alpha_i) \geq J^* - \delta_i(1 + \alpha_i), \tag{1.3.14}
\]

consequently, the iterates \( u^i \) exist for all \( i \). Since \( \omega \) is a strongly convex functional, the weak normal solution, we conclude that \( \omega(u^*_i) \) exist for all \( i \).

Using the non-negativity of \( \omega \) and the relations (1.3.16), (1.3.18), one gets

\[
 J(u^*_i) = \inf_{u \in K} \{ J(u) + \alpha_i \omega(u) - \delta_i(1 + \alpha_i) \} \leq \frac{1}{\alpha_i} [2 \delta_i(1 + \alpha_i) + \epsilon_i], \quad \omega(u^*_i) =: \omega(u^*_i) + \gamma_i \tag{1.3.15}
\]

Thus,

\[
 \omega(u^i) \leq \omega(u^*_i) + \frac{1}{\alpha_i} [2 \delta_i(1 + \alpha_i) + \epsilon_i] =: \omega(u^*_i) + \gamma_i, \tag{1.3.16}
\]

\[
 J(u^i) \leq J^* + \mu_i, \tag{1.3.17}
\]

with \( \gamma_i = \frac{1}{\alpha_i} [2 \delta_i(1 + \alpha_i) + \epsilon_i] \), \( \mu_i = \alpha_i \omega(u^*_i) + 2 \delta_i(1 + \alpha_i) + \epsilon_i \), and on account of (1.3.12),

\[
 \lim_{i \to \infty} \gamma_i = 0, \quad \lim_{i \to \infty} \mu_i = 0. \tag{1.3.18}
\]

Using the definition of a weak stabilizer and the relations (1.3.16), (1.3.18), boundedness of the sequence \( \{ u^i \} \) can be concluded. Hence, one can select a weakly convergent subsequence \( \{ u^{i_k} \} \), \( u^{i_k} \rightharpoonup u^* \), and in view of the closedness of \( K \), \( u^* \in K \).

Taking limits in the inequalities (1.3.16), (1.3.17) and due to the uniqueness of the normal solution, we conclude that \( u^* = u^*_i \).

Obviously, the sequence \( \frac{1}{2} (u^{i_k} + u^*_i) \) also converges weakly to \( u^*_i \). According to Definition A1.5.20, strong convexity of \( \omega \) leads to

\[
 \frac{1}{4} \kappa \| u^{i_k} - u^*_i \|^2 \leq \frac{1}{2} \left[ \omega(u^{i_k}) + \omega(u^*_i) \right] - \omega \left( \frac{1}{2} (u^{i_k} + u^*_i) \right) \tag{1.3.19}
\]

and, since \( \omega \) is weakly lower semicontinuous,

\[
 \lim_{i \to \infty} \omega \left( \frac{1}{2} (u^{i_k} + u^*_i) \right) \geq \omega(u^*_i), \quad \lim_{i \to \infty} \omega(u^{i_k}) \geq \omega(u^*_i).
\]

The latter inequality together with (1.3.16) implies

\[
 \lim \omega(u^{i_k}) = \omega(u^*_i).
\]
Now, in view of (1.3.19), we obtain
\[
\lim_{i \to \infty} \| u^i - u^*_c \| = 0,
\]
and in order to get
\[
\lim_{i \to \infty} \| u^i - u^*_\omega \| = 0,
\]
it is sufficient to note that \( \{ u^k \} \) was an arbitrarily chosen weakly convergent subsequence. \( \square \)

Even if we assume that the problem data are given exactly, we cannot abstain from the necessity to apply special methods designed for ill-posed problems. Indeed, by means of the examples given in Section 6.1 we will see that methods which are suitable for well-posed problems are not practicable for ill-posed ones, even if we suppose in addition that the computations are performed exactly.

### 1.3.4 Proximal point regularization

Now we turn to the mainly investigated and propagated method in this book for solving ill-posed variational problems, the so-called proximal point method (PPM). This method was suggested by Martinet [287, 288] and became the starting point for a new regularization approach. PPM is based on the proximal point mapping
\[
\text{Prox}_{f,C} u := \arg \min_{v \in C} \{ f(v) + \frac{\chi_k}{2} \| v - u \|^2 \}, \quad u \in V,
\]
where \( C \) is supposed to be a non-empty, closed and convex set in a Hilbert space \( V \), \( f : V \to \mathbb{R}^n \) is a lower semi-continuous functional and
\[
0 < \underline{\chi} \leq \chi_k \leq \bar{\chi}
\]
is a sequence of regularization parameters.

In comparison to the Tikhonov regularization approach the main advantage of PPM consists in the fact that obviously the regularized functional \( \Psi_{\chi,u}(\cdot) := f(\cdot) + \frac{\chi_k}{2} \| \cdot - u^k \|^2 \) is strongly convex for all \( k \) and its regularizing effect does not vanish for \( k \to \infty \). For more properties of this mapping see Section 3.1.

In the following chapters iterative proximal point regularization (PPR) will be studied in more detail for different classes of problems. In this section we restrict our attention to some regularizing properties of these methods, taking the same point of view as in the study of methods in the foregoing sections.

### 1.3.10 Remark.

As long as the regularization parameter \( \chi_k \) stays away from zero the convergence theory of PPM is not inflicted by the value of \( \chi_k \). Therefore, when describing basic properties of the proximal-point approach, we set some times for the sake of simplicity \( \chi_k \equiv 2 \forall k \).

However, as we will see in the numerical examples, a correctly chosen value of \( \chi_k \) – although in general unknown – influences the performance of the method and has to be properly adapted when solving particular problems. \( \diamond \)
1.3. APPROACHES OF STABILIZING ILL-POSED PROBLEMS

We consider again Problem (1.2.1) with $U = V$ being a Hilbert space and the feasible set $K$ being given exactly. Let $\{\delta_i\}, \{\gamma_i\}$ and $\{\epsilon_i\}$ be non-negative sequences converging to 0, $\gamma_0 = \sup \gamma_i$, $\bar{K} = \{u \in V : \rho(u, K) \leq \gamma_0\}$ and $\{J_i\}$ is assumed to be a family of continuous functionals satisfying the condition

$$
\sup_{u \in \bar{K}} |J_i(u) - J(u)| \leq \delta_i, \quad i = 1, 2, ...
$$

(1.3.20)

1.3.11 Proposition. Let $\{\delta_i\}, \{\gamma_i\}$ and $\{\epsilon_i\}$ be non-negative sequences converging to 0, $\gamma_0 = \sup \gamma_i$, $\bar{K} = \bar{K} = \{u \in V : \rho(u, K) \leq \gamma_0\}$ and $\{J_i\}$ is assumed to be a family of continuous functionals satisfying the condition

$$
\sup_{u \in \bar{K}} |J_i(u) - J(u)| \leq \delta_i, \quad i = 1, 2, ...
$$

(1.3.21)

Proximal Point Regularization (PPR):

For arbitrarily chosen $u^0 \in K$ compute $u^i$ according to

$$
\rho(u^i, K) \leq \gamma_i,
$$

(1.3.22)

and let $v^i$ be the nearest point to $u^i$ on $K$, i.e., $v^i = \arg\min_{v \in K} \{\|u^i - v\| : v \in K\}$. Due to (1.3.20), (1.3.22) we obtain

$$
\Psi_i(u^i) = J_i(u) + \|u - u^{i-1}\|^2.
$$

(1.3.23)

Then the sequence $\{u^i\}$ converges weakly to some element $\bar{u} \in U^*$. 

Proof: Denote $\Psi_i(u) = J(u) + \|u - u^{i-1}\|^2$ and let $v^i$ be the nearest point to $u^i$ on $K$, i.e., $v^i = \arg\min_{v \in K} \{\|u^i - v\| : v \in K\}$. Due to (1.3.20), (1.3.22) we obtain

$$
\Psi_i(u^i) \leq \min_{u \in K} \Psi_i(u) + \epsilon_i + 2\delta_i \leq \Psi_i(v^{i-1}) + \epsilon_i + 2\delta_i
$$

(1.3.24)

and because of (1.3.21) the estimate $\|v^i - u^i\| \leq \gamma_i$ holds. Together with the Lipschitz property assumed for $J$, we get

$$
\Psi_i(v^{i-1}) \leq \Psi_i(u^{i-1}) + L\gamma_{i-1} + \gamma_{i-1}^2.
$$

Therefore,

$$
J(v^i) + \|v^i - u^{i-1}\|^2 \leq J(u^{i-1}) + \epsilon_i + 2\delta_i + L\gamma_{i-1} + \gamma_{i-1}^2,
$$

(1.3.25)

and finally

$$
J(v^i) \leq J(u^{i-1}) + \epsilon_i + 2\delta_i + L\gamma_{i-1} + \gamma_{i-1}^2.
$$

Using Lemma A3.1.4 and (1.3.23) we conclude that the sequence $\{J(u^i)\}$ converges and inequality (1.3.25) leads to

$$
\lim_{i \to \infty} \|u^i - u^{i-1}\| = 0.
$$

(1.3.26)

In view of

$$
|J(u^i) + \|u^i - u^{i-1}\|^2 - J(v^i) - \|v^i - u^{i-1}\|^2| \leq L\|u^i - v^i\| + \|u^i - u^{i-1} + v^i - u^{i-1}\| \cdot \|u^i - v^i\|
$$
and the relations (1.3.21), (1.3.23) and (1.3.26), it follows that
\[ \|u^i - u^{i-1} + v^i - u^{i-1}\| \to 0. \]
Hence, for some \( \kappa > 0 \) we have
\[ |\Psi(u^i) - \Psi_i(v^i)| \leq (L + \kappa)\gamma_i \quad \forall \ i \]
and this implies
\[ \Psi_i(v^i) \leq \Psi_i(u^i) + (L + \kappa)\gamma_i \quad \forall \ i. \]
Defining \( \bar{u}^i = \arg\min_{u \in K} \Psi_i(u) \), the latter inequality together with (1.3.20) and (1.3.22) yield
\[ \Psi_i(u^i) \leq \Psi_i(\bar{u}^i) + \epsilon_i + 2\delta_i \] (1.3.27)
and consequently,
\[ \Psi_i(v^i) \leq \Psi_i(\bar{u}^i) + \epsilon_i + 2\delta_i + (L + \kappa)\gamma_i. \] (1.3.28)
In view of the strong convexity of \( \Psi_i \) and the choice of the point \( \bar{u}^i \), we get also the relation
\[ \Psi_i(v^i) - \Psi_i(\bar{u}^i) \geq \langle q, v^i - \bar{u}^i \rangle + \frac{1}{4}\|v^i - \bar{u}^i\|^2 \] (1.3.29)
with \( \langle q, v^i - \bar{u}^i \rangle \geq 0 \) for some element \( q \) of the subdifferential \( \partial \Psi_i(\bar{u}^i) \).
Thus, from (1.3.28) and (1.3.29) the estimate
\[ \|\bar{u}^i - v^i\| \leq 2(\epsilon_i + 2\delta_i + (L + \kappa)\gamma_i)^{1/2} \]
is satisfied and, observing (1.3.21)
\[ \|u^i - u\| \leq 2(\epsilon_i + 2\delta_i + (L + \kappa)\gamma_i)^{1/2} + \gamma_i =: \bar{\gamma}_i. \] (1.3.30)
Now, assumption (1.3.23) provides that the series \( \sum_{i=1}^{\infty} \gamma_i \) converges.
Because the mapping
\[ \text{Prox}_{J,K} : u \to \arg\min_{v \in K} \{J(v) + \|v - u\|^2\} \]
is non-expansive (see Remark 3.1.2) and a fix-point-mapping, i.e.,
\[ \text{Prox}_{J,K} u^* = u^* \text{ for } u^* \in U^* \]
holds, we get
\[ \|\bar{u}^{i+1} - u^*\| \leq \|u^i - u^*\|. \]
On account of (1.3.30) this gives
\[ \|u^{i+1} - u^*\| \leq \|u^i - u^*\| + \bar{\gamma}_{i+1}. \]
Now, applying Polyak’s Lemma A3.1.4 and (1.3.23) we get convergence of the sequence \( \{\|u^i - u^*\|\} \) and thus boundedness of \( \{u^i\} \).
Since \( \bar{u}^{i+1} \) is the minimum point of the functional \( \Psi_{i+1} \) on \( K \), Proposition A1.5.34 (inequality (A1.5.24)) with \( J_2 := J \) and \( J_1(u) := \|u - u^i\|^2 \) implies that for fixed \( u^* \in U^* \) we have
\[ J(u^*) - J(\bar{u}^{i+1}) + 2(\bar{u}^{i+1} - u^*, u^* - \bar{u}^{i+1}) \geq 0. \] (1.3.31)
With (1.3.26) and (1.3.30) it follows that
\[ \|\bar{u}^{i+1} - u^i\| \to 0. \quad (1.3.32) \]
Thus, relation (1.3.31) leads to
\[ J(u^*) \geq \lim_{i \to \infty} J(\bar{u}^i) \geq \lim_{i \to \infty} J(\bar{u}^i) \geq J(u^*), \]
i.e., \( \lim_{i \to \infty} J(\bar{u}^i) = J(u^*) \).
By means of the Lipschitz property of \( J \) and (1.3.32) in addition we conclude that
\[ \lim_{i \to \infty} J(u^i) = J(u^*). \quad (1.3.33) \]
Now, due to the boundedness of \( \{u^i\} \), there exists a subsequence \( \{u^{i_j}\} \) which converges weakly to some point \( \bar{u} \). Moreover, because of the weak closedness of \( K \) and (1.3.33) we obtain \( \bar{u} \in U^* \).
Observing that for any \( u^* \in U^* \) the sequence \( \left\{ \|u^i - u^*\| \right\} \) converges, in view of Opial’s lemma (Proposition A1.1.3) the weak convergence of \( \{u^i\} \) to \( \bar{u} \) is valid immediately .

1.3.12 Remark. Proposition 1.3.11 remains true if the functionals \( J_i \) are convex and Gâteaux-differentiable and instead of (1.3.21) and (1.3.22) the condition
\[ \|\nabla \Psi_i(u^i) - \nabla \Psi_i(\bar{u}^i)\| \leq \epsilon_i \]
on the gradients of \( \Psi_i \) is used, with \( \bar{u}^i = \arg \min_{u \in K_i} \Psi_i(u) \) (see the relations between (A1.5.28) and (A1.5.31)).

1.3.13 Remark. Proposition 1.3.11 can be easily extended to the case where the feasible set \( K \) is bounded and, instead of \( K \), closed and convex sets \( K_i \) are considered with \( \rho_H(K, K_i) \leq \sigma_i \). In this case the conditions (1.3.21) and (1.3.22) in the proximal method above have to be modified to
\[ \rho(u^i, K_i) \leq \gamma_i, \quad (1.3.34) \]
\[ \Psi_i(u^i) \leq \min_{u \in K_i} \Psi_i(u) + \epsilon_i \quad (1.3.35) \]
and, in addition to (1.3.23), it must be assumed that \( \sum_{i=1}^{\infty} \sqrt{\sigma_i} < \infty \).
In order to establish the convergence of this modified method, we show that (1.3.34) and (1.3.35) imply the validity of the inequalities (1.3.21) and (1.3.22) with modified constants \( \gamma_i \) and \( \epsilon_i \). To this end we prove that the estimate
\[ \min_{u \in K_i} \Psi_i(u) - \min_{u \in K} \bar{\Psi}_i(u) \leq (L + 4r)\sigma_i \quad (1.3.36) \]
holds true if
\[ \{ u : \rho(u, K) \leq \sup_{i} (\gamma_i + \sigma_i) \} \subset S_r(0). \]
Indeed, if \( \min_{u \in K_i} \min \Psi_i(u) \geq \min_{u \in K} \Psi_i(u) \), setting
\[ w^i = \arg \min_{u \in K_i} \Psi_i(u), \quad z^i = \arg \min_{v \in K_i} \|v - \bar{u}\|, \]
one can conclude that
\[ \Psi_i(z^i) \geq \Psi_i(w^i) \geq \Psi_i(\bar{u}^i) \]
and, in view of \( \|z^i - \bar{u}^i\| \leq \sigma_i \) and the Lipschitz property, it holds
\[ \Psi_i(z^i) - \Psi_i(\bar{u}^i) \leq (L + 4r)\sigma_i. \]
Hence,
\[ \Psi_i(w^i) - \Psi_i(\bar{u}^i) \leq (L + 4r)\sigma_i. \]
Now, using (1.3.36), we obtain from (1.3.35) that
\[
J_i(u^i) + \|u^i - u^{i-1}\|^2 \leq 
\leq \min_{u \in K} \{J_i(u) + \|u - u^{i-1}\|^2\} + (L + 4r)\sigma_i + 2\delta_i + \epsilon_i,
\]
but, due to (1.3.34) and \( \rho_H(K, K_i) \leq \sigma_i \), it follows immediately that
\[ \rho(u^i, K) \leq \gamma_i + \sigma_i. \]

In Section 4.2 we drop the assumption that the feasible set \( K \) is bounded.

Comparing the Propositions 1.3.9 and 1.3.11, we see with regard to Remark 1.3.13 that for the Tikhonov method a stronger statement on the convergence of the minimizing sequence is obtained than for the PPM. However, the PPM guarantees an essential better conditioning of the auxiliary problems and, as a consequence, a better stability of the numerical process.

In a number of papers the PPM is considered in a slightly more general setting in the sense that in (1.3.22) instead of
\[
J_i(u) + \|u - u^{i-1}\|^2
\]
functionals
\[
J_i(u) + \frac{\chi_i}{2}\|u - u^{i-1}\|^2
\]
are used, where \( \{\chi_i\} \) is a sequence of positive regularization parameters with \( \sup_i \chi_i < \infty \). If, in addition, \( \inf_i \chi_i > 0 \), then Proposition 1.3.11 remains true and the proof requires only some trivial modifications.

1.4 Comments

Among approximately fifty monographs known to us which are dedicated to the theory and solution of ill-posed problems only a minor part is devoted to variational or extremal problems (see, for instance, Tikhonov and Arsenin [395], Liskovec [275], Bakushinski and Goncharski [30]) and only the monographs of Vasiljev [408], Alifanov, Artyukhin and Rumyantsev [7] deal with this subject exclusively. Together with the bibliographical survey by Liskovec [276] they give a sufficiently complete overview about the investigation of the topic with the exception of the proximal point methods.
Section 1.1: An analysis of the notion well-posedness (see Definition 1.1.1) is given in [30]. The first notion of well-posedness was introduced by Hadamard [162, 163]. For the equation of the first kind $Au = z$ this corresponds to $G = A^{-1}$, $DG = D = I$. In fact, Definition 1.1.1 coincides with the concept of conditional well-posedness introduced by Lavrentjev [258]. A comparison of the different notions of well-posedness was performed by Dontchev and Zolezzi [92].

Investigations on the ill-posedness of the Cauchy problem for the Laplace equation can be found in the papers of Pucci [336], Lavrentjev [257], Liskovc [274] and Schaefer [361]. A large number of papers is dedicated to operator equations of first kind, we refer especially to Douglas [93], Phillips [320], Ivanov, Vasin and Tanana [196] and Kress [249].

Various types of problems are analysed by Tikhonov and Arsenin [395], Hoffmann and Sprekels [184], Hofmann [185], Baumeister [34], Romanov [354], Louis [279] and Anger [11]. For the problem of computing values of unbounded operators cf. Morozov [295]. Also recommendable are the proceedings of a conference in Oberwolfach, edited by Hämmerlin and Hoffmann [164], where problems of parameter identification, free boundary and inverse problems for differential equations and also integral equations of first kind are studied, including applications in technical fields and in medicine.

Section 1.2: Well-posedness of finite-dimensional convex optimization problems according to Definition 1.2.1 was investigated by Karmanov [232]. For a notion of well-posedness of constraints see Levitin and Polyak [263].

An analogue of Theorem 1.2.6 for a slightly different definition of weak well-posedness (in particular, convexity of $J_\delta$ and $g_\delta$ is assumed) has been proved by Eremin and Astafjev [105].

Section 1.3: The notion of a stabilizer was created in context with the analysis of the convergence of Tikhonov regularization and residual methods. There, for the first time, the functional $\|u - \bar{u}\|^2$ was applied in order to obtain stability of the solution procedure. Numerous examples of stabilizers and weak stabilizers for different problems (in particular also for control problems) and spaces are given by Vasiljev [408].

The quasi-solution method was introduced by Ivanov [195] for an equation of the first kind with an exactly given continuous operator. Its further development is essentially connected with a generalization of Proposition 1.3.3 (cf. Fuller [124], Liskovc [275], Bakushinski and Gonkharshi [30]). A broad field of applications of quasi-solution methods is described by Ivanov, Vasin and Tanana [196].

The question of the choice of a suitable compact set of well-posedness is of special interest for some problems in order to establish error estimations for the solution under approximately given information. This has motivated research in different areas of applications (cf. John [199], Pucci [336], Lavrentjev, Romanov and Shishatski [259]). An iteration method with a successive enlargement of the compact set of well-posedness by means of a technique from parametric programming was introduced in Tichatschke and Hofmann [387].

The residual method has been applied first by Phillips [320] for solving a Fredholm integral equation of the first kind. A rigorous theoretical foundation
of this method was given by Morosov [293]. The generalization of the residual method to approximately given operators was considered by Gonsharski, Leonov and Jagola [142] and Morozov [295].

One of the most important contributions to the field of ill-posed problems has been the regularization method introduced by Tikhonov [392]. In Tikhonov and Arsenin [395] the method was described for different types of ill-posed problems. Its main two variants differ in the choice of the regularization parameter which is based on an a priori rule or performed according to the residual value and the accuracy of the approximate data. Both approaches have attracted wide publicity and consideration in the literature, cf. Liscovec [275], Groetsch [150], Morozov [295], Hofmann [185], Engl, Kunisch and Neubauer [104].

Systematical investigations of discretized schemes in connection with regularization methods can be found in Liscovec [275], Vainikko [406], Hofmann [185].

Some results modifying Proposition 1.3.5 to different types of ill-posed problems are contained in Liskovec [275], and in Vasiljev [408] its generalization to variational problems subject to inexact constraints is described.
Chapter 2

STABLE DISCRETIZATION AND SOLUTION METHODS

2.1 Stabilizing Properties of Numerical Methods

Using the regularization techniques described in Section 1.3, all the constraints of the original variational problem are included in the regularized problems. Moreover, in the limit of the parameter $\delta \to 0$, according to the family of perturbations in (1.2.3), $(\delta \to 0, \alpha \to 0$ in Tikhonov’s regularization method), we formally return to the original problem, except for some additional restrictions in the cases of the residual- and quasi-solution methods.

It should also be mentioned that, in principle, the controlling parameters (for instance $\epsilon, \gamma, \delta$ in the proximal point method) are chosen \textit{a priori}, and the convergence properties of the considered methods do not depend on the specific algorithms used to solve the regularized problems. Therefore, the regularization techniques can be seen independently of the numerical methods for solving the approximated and regularized problems and can be considered purely as approaches to stabilize the problems.

Now let us study some standard numerical methods which themselves have a regularizing property.

2.1.1 Stabilizing properties of gradient type methods

In this subsection we consider again Problem (1.2.4), i.e.,

\[
\begin{align*}
\text{minimize} & \quad J(u), \\
\text{subject to} & \quad u \in K,
\end{align*}
\]

where $J : V \to \mathbb{R}$ is a convex, lower semicontinuous functional, $K \subset V$ is a convex, closed set and $K \cap \text{dom} J \neq \emptyset$

and assume additionally that the objective functional $J$ is Fréchet-differentiable and convex on the Hilbert space $V$ and that $\nabla J$ satisfies a Lipschitz condition.
on $V$ with constant $L$.

According to Riesz’ Theorem A1.1.1, $\nabla J$ is considered as an element of the space $V$.

We consider the gradient projection method described in Subsection A3.3.2, assuming that the gradients are approximately determined within the iteration procedure.

Thus, starting with some $u^0 \in K$ and choosing $0 < \alpha < \frac{2}{L}$, we calculate $u^{k+1}$ according to

$$u^{k+1} := \Pi_K(u^k - \alpha p^k),$$

where $\Pi_K$ is the orthoprojector onto the set $K$ and $p^k$ is assumed to satisfy the inequality

$$\|p^k - \nabla J(u^k)\| \leq \delta_k, \quad \delta_k \downarrow 0.$$

This corresponds to the case that the set $K$ is given exactly and the error in the approximation of the functional $J$ becomes small in the norm of the space $C^1(V)$.

Now let us investigate the convergence of this gradient projection method.

2.1.1 Proposition. (cf. Proposition A3.3.28)

Suppose that the set $U^* = \text{Arg min}_{u \in K} J(u)$ is non-empty. Then the sequence $\{u^k\}$, generated by the gradient projection method (2.1.2) - (2.1.4) with $\sum_{k=0}^{\infty} \delta_k < \infty$, converges weakly to some $u^* \in U^*$.

Proof: Using inequality (A1.5.22) with $u^1 := u^k$ and $u^2 := u^*$, where $u^* \in U^*$ is chosen arbitrarily, we get

$$\langle \nabla J(u^k) - \nabla J(u^*), u^k - u^* \rangle \geq \frac{1}{L} \|\nabla J(u^k) - \nabla J(u^*)\|^2. \quad (2.1.5)$$

For $u^{k+1} = \Pi_K(u^k - \alpha \nabla J(u^k))$, due to (2.1.4) and the non-expansivity of the projection $\Pi_K$ (cf. Corollary A1.5.36), the estimate

$$\|\hat{u}^{k+1} - u^{k+1}\| = \|\Pi_K(u^k - \alpha \nabla J(u^k)) - \Pi_K(u^k - \alpha p^k)\|$$

$$\leq \alpha \|\nabla J(u^k) - p^k\| \leq \alpha \delta_k \quad (2.1.6)$$

holds true. Due to $u^* \in U^*$ one can show that

$$\Pi_K(u^* - \alpha \nabla J(u^*)) = u^*,$$

and applying the inequalities (2.1.5), (2.1.6), we infer

$$\|u^{k+1} - u^*\| = \|\Pi_K(u^k - \alpha \nabla J(u^k)) - \Pi_K(u^* - \alpha \nabla J(u^*))\|$$

$$\leq \|u^k - u^* - \alpha (\nabla J(u^k) - \nabla J(u^*))\|^2$$

$$\leq \|u^k - u^*\|^2 - 2\alpha (\nabla J(u^k) - \nabla J(u^*), u^k - u^*)$$

$$+ \alpha^2 \|\nabla J(u^k) - \nabla J(u^*)\|^2$$

$$\leq \|u^k - u^*\|^2 - \alpha \left(\frac{2}{L} - \alpha\right) \|\nabla J(u^k) - \nabla J(u^*)\|^2.$$
2.1. STABILIZING PROPERTIES OF NUMERICAL METHODS

Observing that \( \alpha \in (0, \frac{1}{L}) \), the estimate
\[
\|\hat{u}^{k+1} - u^*\| \leq \|u^k - u^*\|
\]
holds, and due to (2.1.6), this implies
\[
\|u^{k+1} - u^*\| \leq \|\hat{u}^{k+1} - u^*\| + \|\tilde{u}^{k+1} - u^{k+1}\| \leq \|u^k - u^*\| + \alpha \delta_k.
\]
(2.1.7)

In view of (2.1.7), assumption \( \sum_{k=0}^{\infty} \delta_k < \infty \) and Lemma A3.1.4, we conclude that the sequence \( \{\|u^k - u^*\|\} \) converges for all \( u^* \in U^* \).

Furthermore, \( u^{k+1} \) is defined by (2.1.3), i.e. it satisfies the variational inequality
\[
\langle u^{k+1} - u^* + \alpha p^k, u - u^{k+1} \rangle \geq 0, \quad \forall u \in K.
\]
Consequently,
\[
\langle p^k, u - u^{k+1} \rangle \geq \frac{1}{\alpha} \langle u^k - u^{k+1}, u - u^{k+1} \rangle
\]
and inserting \( u := u^k \), we have
\[
\langle p^k, u^k - u^{k+1} \rangle \geq \frac{1}{\alpha} \|u^k - u^{k+1}\|^2.
\]
By virtue of (2.1.4) this leads to
\[
\langle \nabla J(u^k), u^k - u^{k+1} \rangle \geq \frac{1}{\alpha} \|u^k - u^{k+1}\|^2 - \langle p^k - \nabla J(u^k), u^k - u^{k+1} \rangle
\]
\[
\geq \frac{1}{\alpha} \|u^k - u^{k+1}\|^2 - \delta_k \|u^k - u^{k+1}\|.
\]
(2.1.9)

But inequality (A1.4.10) and the Lipschitz condition yield for any \( u, v \in V \)
\[
J(u) - J(v) - \langle \nabla J(v), u - v \rangle = \int_0^1 \langle \nabla J(v + t(u - v)) - \nabla J(v), u - v \rangle dt \leq \frac{L}{2} \|u - v\|^2,
\]
such that
\[
J(u^{k+1}) - J(u^k) - \langle \nabla J(u^k), u^{k+1} - u^k \rangle \leq \frac{L}{2} \|u^{k+1} - u^k\|^2.
\]
This together with (2.1.9) ensures that
\[
J(u^k) - J(u^{k+1}) \geq \left( \frac{1}{\alpha} - \frac{L}{2} \right) \|u^k - u^{k+1}\|^2 - \delta_k \|u^k - u^{k+1}\|,
\]
(2.1.10)
with \( \frac{1}{\alpha} - \frac{L}{2} > 0 \) according to the choice of \( \alpha \). Therefore,
\[
J(u^{k+1}) \leq J(u^k) + \delta_k \|u^k - u^{k+1}\|,
\]
and because of the boundedness of the sequence \( \{\|u^k\|\} \) and \( \sum_{k=0}^{\infty} \delta_k < \infty \), we conclude from Lemma A3.1.4 that the sequence \( \{J(u^k)\} \) is convergent. Taking into account relation (2.1.10), it follows that
\[
\|u^k - u^{k+1}\| \rightarrow 0.
\]
(2.1.11)

Again, regarding the boundedness of the sequence \( \{u^k\} \), a subsequence \( \{u^{k_j}\} \) can be chosen converging weakly to some element \( \hat{u} \). The convex set \( K \) is weakly
closed, hence, \( \hat{u} \in K \). In view of the convexity of the functional \( J \) we obtain from (2.1.8) for all \( u \in K \) that

\[
J(u) - J(u^k) \geq \langle \nabla J(u^k), u - u^k \rangle \\
\geq \frac{1}{\alpha} (u^k - u^{k+1}, u - u^{k+1}) \\
+ \langle \nabla J(u^k), u - u^{k+1} - u^k + u^{k+1} \rangle \\
\geq \frac{1}{\alpha} (u^k - u^{k+1}, u - u^{k+1}) - \| \nabla J(u^k) - p_k \| \| u - u^{k+1} \| \\
- \| \nabla J(u^k) \| \| u^{k+1} - u^k \| .
\]

Taking limit in (2.1.12) for \( k = k_j, j \to \infty \), and observing (2.1.4), (2.1.11) as well as the lower semi-continuity of \( J \), we get

\[
J(u) \geq J(\hat{u}), \quad \forall u \in K,
\]

which implies that \( \hat{u} \in U^* \). With regard to Opial’s lemma (cf. Proposition A1.1.3) this shows that \( u^k \rightharpoonup \hat{u} \). \( \square \)

### 2.1.2 Remark. In the case \( V := \mathbb{R}^n \) Proposition 2.1.1 leads immediately to

\[
\lim_{k \to \infty} \| u^k - \hat{u} \| = 0.
\]

For the cases \( K := V \) and \( K := V = \mathbb{R}^n \) corresponding results on the convergence of the resulting gradient method can be obtained, showing that the gradient method has a regularizing behavior, too (cf. Polyak [330] and Gol’stein and Tretyakov [141]). \( \Diamond \)

Alifanov et al. [7] investigated the application of the gradient method in order to solve operator equations \( Au = f \) with \( A \) a linear continuous operator from a Hilbert space \( V \) into a Hilbert space \( F \). Under the conditions that the data are approximately given according to

\[
\| A_{h_k} - A \| \leq h_k, \quad \| f_{h_k} - f \| \leq \delta_k, \quad \delta_k + h_k \to 0,
\]

where \( A_{h_k} \) is linear and continuous, it is proved that for a suitable choice of \( \alpha \) the sequence

\[
u^{k+1} := u^k - \alpha A_{h_k}^* (A_{h_k} u^k - f_{h_k})
\]

converges in the norm of the space \( V \) to a solution of the equation \( Au = f \) having a minimal distance to the starting point \( u^0 \). The technique of the proof uses essentially the linearity of the operators \( A \) and \( A_{h_k} \).

### 2.1.2 Stabilizing properties of penalty methods

Let us consider Problem (A1.7.35) under somewhat stronger conditions

\[
\begin{align*}
\text{minimize} \quad & J(u), \\
\text{subject to} \quad & u \in K, \\
K := \{ u \in V : g_j(u) \leq 0, \quad j = 1, \ldots, m \},
\end{align*}
\]

\( J, g_j : V \to \mathbb{R} \) are convex, continuous functionals on a Hilbert space \( V \), \( U^* = \text{Arg min}_{u \in K} \) is non-empty and bounded.
2.1. STABILIZING PROPERTIES OF NUMERICAL METHODS

We consider a penalty function

\[
\phi_k(u) = \varphi_k(g(u)),
\]
where the family \( \{ \varphi_k \} \) is subject to conditions formulated in the following lemma.

2.1.3 Lemma. Let \( \varphi_k : \mathbb{R}^m \to \mathbb{R} \) be a family of convex functions and

\[
\lim_{k \to \infty} \varphi_k(z) = \begin{cases} 
0 & \text{if } \max_{1 \leq j \leq m} z_j < 0, \\
+\infty & \text{otherwise}.
\end{cases}
\]

Then, for arbitrary \( \alpha_1 < \alpha_2 < 0, \alpha_3 > 0 \) and \( \delta > 0 \), the functions \( \varphi_k \) converge uniformly
to 0 on the set \( \{ z : \alpha_1 \leq z_j \leq \alpha_2, \; j = 1, \ldots, m \} \);
to \( +\infty \) on the set \( \{ z : \max_{1 \leq j \leq m} z_j \geq \alpha_3 \} \cap B_{\delta}(0) \),
with \( B_{\delta}(0) := \{ u \in V : \| u \| \leq \delta \} \).

Proof: The first part of the statement is an immediate consequence of Proposition A1.5.19. In order to verify the second part, we can easily adapt the proof of Theorem A3.4.41 by taking \( n = m, H := \mathbb{R}^m, K := G = \{ z \in \mathbb{R}^m : z \leq 0 \} \), \( \phi_2^k := 0, \phi_1^k := \varphi_k, J := \rho^2(z, B_{\delta}(0)) \) with a suitable \( \delta^* \).

Now, according to (2.1.13) and (2.1.14), we define

\[
\begin{align*}
\tilde{F}_k(u) & := \varphi_k(g(u)), \\
\bar{F}_k(u) & := J(u) + \phi_k(u), \\
\bar{g}(u) & := \sup_{1 \leq j \leq m} g_j(u); \\
W_\tau & := \{ u \in V : J(u) \leq J^* + \tau, \; \bar{g}(u) \leq \tau \}
\end{align*}
\]
and calculate the points \( u^k \) via

\[
\tilde{F}_k(u^k) \leq \inf_{u \in V} \tilde{F}_k(u) + \epsilon_k, \quad k = 1, 2, \ldots,
\]
where \( \tilde{F}_k \) are approximately given functions specified in the following theorem.

2.1.4 Theorem. Assume that \( \lim_{r \to 0} \sup_{u \in W_\tau} \rho(u, U^*) = 0 \), that \( \varphi_k \) satisfy the conditions of Lemma 2.1.3 and

\[
\lim_{k \to \infty} \varphi_k(z) = 0 \quad \text{if} \quad \max_{1 \leq j \leq m} z_j = 0,
\]
and that \( \phi_k \) are convex and non-negative on \( V \).

Moreover, let one of the following conditions be fulfilled:

(i) \( F_k, \tilde{F}_k : V \to \mathbb{R} \) are continuous on \( C(W_\tau) \) for fixed \( \tau > 0 \),

\[
\| \tilde{F}_k - F_k \|_{C(W_\tau)} \leq \mu_k,
\]
and \( \tilde{F}_k \) are convex on \( V \), too;

(ii) the functionals \( F_k, \tilde{F}_k \) are continuous on \( V \) and

\[
\| \tilde{F}_k - F_k \|_{C(V)} \leq \mu_k.
\]
of the Slater condition was not assumed. If it is satisfied, then assumption (2.1.17) is superfluous. In this case, instead of (2.1.18), we require that 

\[ \lim_{k \to \infty} \rho(u^k, U^*) = 0 \]

holds true for all points \( u^k \) satisfying the equation \( \tilde{g}(u) = \tilde{\tau} \).

But for sufficiently large \( k (k \geq k_1 \geq k_0) \) we have \( \frac{1}{2} \tilde{\tau} - 2\mu_k - \epsilon_k > 0 \) and with regard to the convexity of \( \tilde{F}_k \), inequality (2.1.22) is satisfied for all \( u \in V \setminus W_\tau \). Therefore, on account of (2.1.16), it follows that \( u^k \in W_\tau \) for \( k \geq k_1 \).

If condition (ii) is fulfilled, then one can conclude immediately from (2.1.16) and (2.1.19) that

\[ F_k(u^k) \leq \inf_{u \in V} F_k(u) + 2\mu_k + \epsilon_k. \]  

(2.1.23)

With regard to (2.1.20) and (2.1.21) the inequality \( \frac{1}{2} \tilde{\tau} - 2\mu_k - \epsilon_k > 0 \) tells us that \( u^k \in W_\tau \).

The assertion follows now from the fact that \( \tilde{\tau} \in (0, \tilde{\tau}) \) was chosen arbitrarily and that \( \lim_{k \to \infty} \sup_{u \in W} \rho(u, U^*) = 0 \). \( \Box \)

2.1.5 Remark. Note that the validity of the Slater condition was not assumed. If it is satisfied, then assumption (2.1.17) is superfluous. In this case, instead of (2.1.18), we require that

\[ \tilde{F}_k(u) \geq F_k(u) - \mu_k \quad \text{on} \quad \text{dom}\phi_k, \quad \text{dom}\phi_k \supset \text{dom}\tilde{F}_k \]

and

\[ \lim_{k \to \infty} (\tilde{F}_k(u) - F_k(u)) = 0 \quad \text{for any} \quad u \in W_\tau \cap \text{int}K, \]

i.e., condition (i) can be weakened, too. \( \diamond \)
In this way, assuming Slater’s condition, the conclusion of Theorem 2.1.4 remains true for so-called barrier functions, especially for function (A3.4.78), too.

Analyzing Theorem 2.1.4 and Remark 2.1.5 we can conclude that, for a wide class of convex variational problems, including the majority of ill-posed and even not weakly well-posed problems, an appropriate adaption of the parameters (penalty parameter and the parameter of controlling the approximation of the input data) guarantees convergence of the iterates to the solution set of the original problem.

2.1.6 Remark. As an example, consider penalty function (A3.4.77) with \( s = 2 \).

Assume that \( J_{\delta_k} \) and \( g_{\delta_k} \) are convex and

\[
\| J - J_{\delta_k} \|_{C(W)} \leq \delta_k, \quad \| g_j - g_{j,\delta_k} \|_{C(W)} \leq \delta_k.
\]

Denote

\[
\tilde{F}_k(u) := J_{\delta_k}(u) + r_k \sum_{j=1}^m \{\max[0, g_j(u)]\}^2,
\]

and \( c := \sup_{u \in W} \max_{1 \leq j \leq m} |g_j(u)| < \infty \).

Then, in order to satisfy condition (2.1.18) with \( \mu_k \downarrow 0 \), it is sufficient that

\[
\lim_{k \to \infty} r_k \delta_k = 0.
\]

Indeed,

\[
\| F_k - \tilde{F}_k \|_{C(W)} \leq \| J - J_{\delta_k} \|_{C(W)}
\]

\[
+ r_k \sum_{j=1}^m \| \{\max[0, g_j(\cdot)]\}^2 - \{\max[0, g_{j,\delta_k}(\cdot)]\}^2 \|_{C(W)}
\]

\[
\leq \delta_k + r_k (2c + \delta_k) m \delta_k =: \mu_k,
\]

and obviously, \( \mu_k \to 0 \) if \( r_k \delta_k \to 0 \). ♦

Similar rules for the choice of the controlling parameters can be established for other types of penalty functions.

Let us illustrate the application of Theorem 2.1.4 to Example 1.1.6, assuming that it is solved by means of the penalty function (A3.4.77) with \( s = 2 \).

If the data approximation is measured in the norm of the space \( C(\tilde{U}_0) \), where in the context of Remark 1.2.2 the set \( \tilde{U}_0 \) is chosen as

\[
\tilde{U}_0 := \{ u \in \mathbb{R}^2 : |u_1| \leq 2, |u_2| \leq 2 \},
\]

then we get

\[
\max_{u \in \tilde{U}_0} |g_2(u) - g_2^\sigma(u)| = 2|\sigma|,
\]

hence, \( \sigma_k \) must be chosen according to \( 2|\sigma_k| \leq \delta_k \).

Concerning Example 1.1.6, it was mentioned that for any perturbation \( \sigma \in (-1, 0) \) the point \( \sigma_k = (0, 1)^T \) is the unique solution of the perturbed problem, but for the unperturbed problem we have \( U^* = \{(1, 0)\} \).
Now, denote \( \hat{u}^k := \arg\min_{u \in \mathbb{R}^2} \tilde{F}_k(u), \) where \( \tilde{F}_k = J + \phi^\sigma_k \), where \( \phi^\sigma_k \) is the corresponding penalty function for the perturbed problem with \( \sigma = \sigma_k \).

Then we can verify immediately that for \( r_k \delta_k^2 \rightarrow 0 \) and sufficient large \( k \),

\[
\hat{u}_1^k = \frac{(1 + \frac{1}{2r_k})(2 + \delta_k)}{2 + 2\sigma_k + 2\sigma^2_k}, \quad \hat{u}_2^k = 1 - \left(1 + \frac{1}{2r_k}\right) \frac{2 + 2\sigma_k + \sigma^2_k}{2 + 2\sigma_k + 2\sigma^2_k},
\]

(2.1.24)

whereas for \( r_k \sigma^2_k \rightarrow +\infty \) we get

\[
\hat{u}_1^k = \frac{2 + \sigma_k}{2r_k \sigma^2_k}, \quad \hat{u}_2^k = 1 - \frac{1}{4r_k} - \frac{(\sigma_k + 2)^2}{4r_k \sigma^2_k}.
\]

(2.1.25)

In the first case, for large \( k \), the parameter \( r_k \) in (2.1.24) is found to be so small that the permissible perturbation in the component \( u_1 \) (i.e., \( -1 - \sigma_k \) instead of \( -1 \) in the function \( g^\sigma_2 \)) has no observable influence on the solution of the unconstrained (penalized) minimization problem and, therefore, the point \( \hat{u}^k \) differs only slightly from the minimizer

\[
\hat{u} = \left(1 + \frac{1}{2r_k}, -\frac{1}{2r_k}\right)^T
\]

of \( F_k \) on \( \mathbb{R}^2 \).

On the other hand, if \( r_k \sigma^2_k \rightarrow \infty \), then the influence of the perturbation becomes large in (2.1.25) for large \( k \) and the sequence \( \{u^k\} \) converges to the point \( u^* = (0,1)^T \), which is the unique optimal solution for any problem perturbed in the described sense.

It should be noted that under the more restrictive requirement \( r_k \delta_k \rightarrow 0 \) the convergence of \( \{u^k\} \) to the optimal point \( u^* = (1,0)^T \) follows from Theorem 2.1.5 for any perturbation of the initial data which, at the \( k \)-th step, does not exceed \( \delta_k \) in the norm of \( C(\tilde{U}_0) \).

### 2.2 Stabilization of Problems and Methods

In the subsequent analysis of the numerical methods we will consider the errors caused by replacing the given original problem by means of a family of auxiliary problems which are assumed to be solved approximately.

Appearing errors in the initial data can be treated in this context as contributions to the class of errors considered above such that the convergence results obtained for exact initial data can be carried over to those for problems with inexact initial data.

In fact, we frequently have used this approach already. For instance, in the proof of Lemma 2.1.3 convergence of the sequence \( \{u^k\} \), constructed according to (2.1.16) with approximately given data, was established by means of inequality (2.1.23), where \( u^k \) was characterized as an approximate solution of a problem with exact data (see also the proofs of Propositions 1.3.9 and 1.3.11).

Hence, this approach enables us in the sequel, to consider initial data to be given exactly.

Now, we briefly describe three main techniques to construct numerical methods for solving ill-posed convex variational problems. In all of them standard discretization and solution algorithms can be used, assuming that the latter...
2.2. **STABILIZATION OF PROBLEMS AND METHODS**

are suitable for the arising well-posed auxiliary problems. We refer to these solution algorithms as being the basic algorithms (or basic methods) in the whole approach considered.

Let us take any suitable basic algorithm for solving variational problems of type (1.2.1). In case \( U \) and \( Y \) are finite-dimensional spaces, this could be one of the methods studied in Appendix A3.4, whereas in the general case combinations of discretization and minimization techniques have to be implemented on the whole.

In order to illustrate the techniques presented in this section, we consider the finite-dimensional problem (A3.4.56) and, as a basic algorithm for solving the auxiliary problems we chose the penalty method with penalty function (A3.4.77) for \( s = 2 \). We refer to this approach as being the illustrative model.

**2.2.1 Structure of regularizing methods**

Using one of the regularization approaches described in Section 1.3, we have to solve successively the auxiliary problems

\[
\min \{ \Phi_k(u) : u \in K \}, \quad k = 1, 2, ..., \tag{2.2.1}
\]

by means of some basic method, where \( \Phi_k \) stands for the Tikhonov- or proximal point regularization, i.e.,

\[
\Phi_k(u) := J(u) + \alpha_k \omega(u) \quad \text{or} \quad \Phi_k(u) := J(u) + \| u - u^{k-1} \|_2^2,
\]

respectively. In the context of our illustrative model, in order to solve the regularized auxiliary problem (2.2.1) for a fixed \( k \) (\( k \)-th exterior step of the process), we construct a finite sequence of points \( \{ u^{k,i} \} \) \( (i = 1, \ldots, i(k)) \), where \( u^{k,i} \) is an approximate minimizer (at the \( i \)-th interior step) of the function

\[
F_{k,i}(u) := \Phi_k(u) + r_{k,i} \sum_{j=1}^m \{ \max[0, g_j(u)] \}^2,
\]

such that

\[
\| \nabla F_{k,i}(u^{k,i}) \| \leq \epsilon_{k,i}.
\]

For each step \( k \) in this procedure the positive controlling parameters \( \{ r_{k,i} \} \) and \( \{ \epsilon_{k,i} \} \) are given such that

\[
\lim_{i \to \infty} \frac{1}{r_{k,i}} = \lim_{i \to \infty} \epsilon_{k,i} = 0,
\]

and \( i(k) \) is that number of the interior steps \( i \) for which

\[
\Phi_k(u^{k,i}) \leq \min_{u \in K} \Phi_k(u) + \epsilon_k,
\]

where \( \epsilon_k > 0 \) and \( \lim_{k \to \infty} \epsilon_k = 0 \).

Afterwards the exterior iteration continues with step \( k + 1 \) by setting

\[
 u^k := u^{k,i(k)}, \quad k := k + 1, \quad i := 1.
\]

Obviously, in this way we obtain a rather complicated iterative procedure in comparison with the solution of a well-posed problem, which is treated by
means of some standard minimization technique. Moreover, this approach has an essential drawback because the regularized auxiliary problems may have to be solved with high accuracy even in situations where the regularized solutions $\bar{u}^k = \arg \min \{\Phi_k(u) : u \in K\}$ are still far from the solution set of the original problem.

### 2.2.2 Iterative one-step regularization

Currently more attention is devoted to diagonal schemes for solving ill-posed variational problems. They are characterized in the following way: For each external step $k$ in Problem (2.2.1) only one interior step of the chosen basic method is performed followed by an update of the controlling parameters according to an a priori rule or to the outcome of the previous iterations. Concerning Tikhonov regularization, the first methods of this type can be found by Bakushinskij and Polyak [31], whereas the proximal point regularization (PPR) was investigated by Rockafella [351] and Antipin [13].

Iterative one-step regularization applied to our illustrative model reads as follows: For fixed sequences $\{\epsilon_k\}, \{r_k\}$ of positive numbers with $\lim_{k \to \infty} \epsilon_k = 0$, $\lim_{k \to \infty} r_k = \infty$ in the $k$-th external step an approximate minimizer $u_k$ of the functional $F_k(u) := \Phi_k(u) + r_k \sum_{j=1}^{m} \{\max[0, g_j(u)]\}^2$ is calculated such that

$$\|\nabla F_k(u_k)\| \leq \epsilon_k, \quad k = 1, 2, \cdots.$$ 

Thereafter step $k+1$ is executed. It should be noted that the straightforward application of the penalty method to well-posed problems of the type (A3.4.56) yields an iterative scheme of the same structure:

$$\|\nabla \bar{F}_k(u_k)\| \leq \epsilon_k, \quad k = 1, 2, \cdots,$$

with $\bar{F}_k(u) := J(u) + r_k \sum_{j=1}^{m} \{\max[0, g_j(u)]\}^2$.

Diagonal schemes of this type for solving ill-posed problems are usually called methods of iterative regularization. They can be understood as a regularization of a basic method.

### 2.2.3 Iterative multi-step-regularization

Convergence results of iterative regularization methods, see the Sections 2.3.2, 3.2 and 4.2, require an update of the controlling parameters ($\epsilon_k$ and $r_k$ in our illustrative model; in addition $\alpha_k$ if we deal with Tikhonov regularization) such that an essential increase of the expense for one iteration may happen.
2.3. **ITERATIVE TIKHONOV REGULARIZATION**

To overcome this difficulty Kaplan and Tichatschke [213], [216] proposed a multi-step proximal point regularization (PPR), which permits to improve significantly the approximate solutions on a given external iteration level \( k \).

This scheme is carried out as follows: At a given external iteration \( k \), fixing the values of the controlling parameters for the internal iterations, we calculate successively the iterates \( u^{k,1}, u^{k,2}, \ldots, u^{k,i(k)} \), where \( u^{k,i} \) is calculated by performing one step of the respective basic method to the problem

\[
\min \{ J(u) + \| u - u^{k,i-1} \|^2 : u \in K \}.
\]

This internal iteration with respect to \( i \) has to be continued as long as "sufficient progress" is achieved from step \( i \) to step \( i+1 \).

For example, in our illustrative model, at the \( k \)-th external step, we have to compute successively the approximate minimizers \( u^{k,1}, u^{k,2}, \ldots \) of the functions \( \tilde{F}_{k,1}, \tilde{F}_{k,2}, \ldots \), such that

\[
\| \nabla \tilde{F}_{k,i}(u^{k,i}) \| \leq \epsilon_k,
\]

with

\[
\tilde{F}_{k,i}(u) := J(u) + \| u - u^{k,i-1} \|^2 + \rho_k \sum_{j=1}^{m} \left\{ \max\{0, g_j(u)\} \right\}^2.
\]

An appropriate stopping criterion for the internal iterations could be

\[
\| u^{k,i} - u^{k,i-1} \| \leq \delta_k,
\]

(2.2.2)

with \( \{ \delta_k \} \) a given sequence of positive numbers, not necessarily tending to zero. If \( u^{k,1} \) satisfies (2.2.2), we set

\[
i(k) := i, \quad u^{k+1,0} := u^{k,i(k)}
\]

and continue with the next external iteration.

The following chapters deal with the investigation of one-step and multi-step regularization under the usage of different basic algorithms.

**2.3 Iterative Tikhonov Regularization**

In this section we consider the stabilization of numerical methods, in particular, the regularization of gradient projection and penalty methods by means of the Tikhonov approach. Concerning other basic methods combined with this type of regularization we refer to Antipin [12], Vasilyev [408], Bakushinski and Goncharski [30].

**2.3.1 Regularized subgradient projection methods**

We consider Problem (1.2.1) in a Hilbert space \( V \) where the constrained set \( U_0 \equiv V \). Let the optimal set \( U^* \) of this problem be non-empty and the weak stabilizer \( \omega \) be a strongly convex (with constant \( \kappa \)), continuous functional on \( V \). Subgradients of the objective functional \( J \) and \( \omega \) at the point \( u \) are identified
with elements of the space $V$ denoted by $q(u) \in \partial J(u)$ and $s(u) \in \partial \omega(u)$.

Then the regularized subgradient projection method in the sense of Tikhonov for solving Problem (1.2.1) reads as follows:

$$u^{k+1} := \Pi_K \left( u^k - \gamma_k (q(u^k) + \alpha_k s(u^k)) \right), \quad \forall \ k \in \mathbb{N}, \quad (2.3.1)$$

with \( \{ \alpha_k \} \), \( \{ \gamma_k \} \) given sequences of positive numbers satisfying the conditions

$$\lim_{k \to \infty} \alpha_k = 0, \quad \lim_{k \to \infty} \frac{\gamma_k}{\alpha_k} = 0, \quad \lim_{k \to \infty} \frac{\alpha_k - \alpha_{k+1}}{\gamma_k \alpha_k} = 0, \quad \sum_{k=1}^{\infty} \alpha_k \gamma_k = \infty. \quad (2.3.2)$$

Here \( \Pi_K \) denotes the orthogonal projector on the set \( K \) and \( u^0 \in K \) is an arbitrary starting point.

We suppress the question how the projection can be carried out, we suppose only that this can be done efficiently. Hence, at this point we do not assume a specific structure of the feasible set \( K \).

Concerning a suitable choice of the parameters \( \alpha_k \) and \( \gamma_k \) satisfying (2.3.2), for instance one can set \( \alpha_k := (1+k)^{-1/2} \), \( \gamma_k := (1+k)^{-1/3} \).

### 2.3.1 Proposition

Suppose there exists a constant \( L \) such that for arbitrary \( u \in V \) and \( q(u) \in \partial J(u) \), \( s(u) \in \partial \omega(u) \) the estimates

$$\| q(u) \| \leq L(1 + \| u \|), \quad \| s(u) \| \leq L(1 + \| u \|) \quad (2.3.3)$$

hold. Then the sequence \( \{ u^k \} \), generated by Method (2.3.1), (2.3.2), converges to the \( \omega \)-normal solution \( u^*_\omega \in U^* \) of Problem (1.2.1).

#### Proof:

Denote

$$\theta_k(u) := J(u) + \alpha_k \omega(u),$$

$$\xi^k := q(u^k) + \alpha_k s(u^k),$$

$$v^k := \arg \min \{ \theta_k(u) : u \in K \}.$$

Proposition 1.3.9 tells us that \( \lim_{k \to \infty} \| v^k - u^*_\omega \| = 0 \). For the iterate \( u^{k+1} \) being the orthogonal projection of \( u^k - \gamma_k \xi^k \) on \( K \) we get

$$\| u^{k+1} - v^{k+1} \| \leq \| u^k - \gamma_k \xi^k - v^k \| + \| v^k - v^{k+1} \|. \quad (2.3.4)$$

In view of the definition of \( v^k \) the subgradients \( q(v^k) \in \partial J(v^k) \), \( s(v^k) \in \partial \omega(v^k) \) can be chosen such that

$$\langle u^k - v^k, q(v^k) + \alpha_k s(v^k) \rangle \geq 0.$$

Now, using Corollary A1.5.32 (formula (A1.5.21)) the inequality

$$\langle u^k - v^k, \xi^k \rangle \geq \langle u^k - v^k, \xi^k - q(v^k) - \alpha_k s(v^k) \rangle \geq 2\alpha_k \| u^k - v^k \|^2$$

can be established. Therefore, on account of (2.3.3) we get

$$\| u^k - \gamma_k \xi^k - v^k \|^2 \leq \| u^k - v^k \|^2 (1 - 4\alpha_k \gamma_k) + \gamma_k^2 \| \xi^k \|^2$$

$$\leq \| u^k - v^k \|^2 (1 - 4\alpha_k \gamma_k) + \gamma_k^2 (\| q(u^k) \| + \alpha_k \| s(u^k) \|)^2$$

$$\leq \| u^k - v^k \|^2 (1 - 4\alpha_k \gamma_k) + \gamma_k^2 (1 + \alpha_k)^2 L^2 (1 + \| u^k - v^k \| + \| v^k \|)^2$$

$$\leq \| u^k - v^k \|^2 (1 - 4\alpha_k \gamma_k) + 2\gamma_k (1 + \alpha_k)^2 L^2 (\| u^k - v^k \|^2 + (1 + \| v^k \|)^2).$$
Due to \( \| u^k - u^* \| \to 0 \) and (2.3.2), there exist constants \( c_1 \) and \( c_2 \) such that for all \( k \) we have
\[
\| u^k - \gamma_k \xi^k - v^k \|^2 \leq \| u^k - v^k \|^2 (1 - 4 \kappa \alpha_k \gamma_k + c_2 \gamma_k^2) + c_1 \gamma_k^2. \tag{2.3.5}
\]

Now we are looking for an upper bound for \( \| u^k - v^{k+1} \| \).

Let the subgradients \( q(v^k), q(v^{k+1}) \) and \( s(v^k), s(v^{k+1}) \) be chosen such that
\[
\langle q(v^k) + \alpha_k s(v^k), v^{k+1} - v^k \rangle \geq 0,
\]
\[
\langle q(v^{k+1}) + \alpha_{k+1} s(v^{k+1}), v^{k+1} - v^k \rangle \geq 0.
\]

Then summing up both inequalities
\[
2 \kappa \alpha_k \| v^{k+1} - v^k \|^2 \leq \langle q(v^{k+1}) + \alpha_k s(v^{k+1}) - q(v^k) - \alpha_k s(v^k), v^{k+1} - v^k \rangle
\]
\[
\leq \langle q(v^{k+1}) + \alpha_k s(v^{k+1}) - q(v^{k+1}) - \alpha_{k+1} s(v^{k+1}), v^{k+1} - v^k \rangle
\]
\[
\leq \langle \alpha_k - \alpha_{k+1} \rangle s(v^{k+1}) \| v^{k+1} - v^k \| \leq c_3 | \alpha_k - \alpha_{k+1} | \| v^{k+1} - v^k \|,
\]
i.e.,
\[
\| v^{k+1} - v^k \| \leq c_4 \frac{\alpha_k - \alpha_{k+1}}{\alpha_k}, \quad k = 1, 2, \ldots. \tag{2.3.6}
\]

In view of the inequalities (2.3.4), (2.3.5) and (2.3.6) we have
\[
\| u^{k+1} - v^{k+1} \| \leq \left( \| u^k - v^k \|^2 (1 - 4 \kappa \alpha_k \gamma_k + c_2 \gamma_k^2) + c_1 \gamma_k^2 \right)^{1/2} + c_4 \frac{\alpha_k - \alpha_{k+1}}{\alpha_k}. \tag{2.3.7}
\]

Taking into consideration the elementar relation
\[
(a + b)^2 \leq (1 + \kappa \alpha_k \gamma_k) a^2 + (1 + \frac{1}{\kappa \alpha_k \gamma_k}) b^2,
\]
by virtue of (2.3.7) and the boundedness of \( \{ \alpha_k \} \) and \( \{ \gamma_k \} \), we conclude that
\[
\| u^{k+1} - v^{k+1} \|^2 \leq \| u^k - v^k \|^2 (1 - 4 \kappa \alpha_k \gamma_k + c_2 \gamma_k^2) (1 + \kappa \alpha_k \gamma_k)
\]
\[
+ c_1 \gamma_k^2 (1 + \kappa \alpha_k \gamma_k) + c_4^2 \left( 1 + \frac{1}{\kappa \alpha_k \gamma_k} \right) | \alpha_k - \alpha_{k+1}|^2
\]
\[
\leq \| u^k - v^k \|^2 [1 - \kappa \alpha_k \gamma k + 3 \kappa \alpha_k \gamma_k (1 + \kappa \alpha_k \gamma_k) + c_2 \gamma_k^2 (1 + \kappa \alpha_k \gamma_k)]
\]
\[
+ c_1 \gamma_k^2 (1 + \kappa \alpha_k \gamma_k) + c_4^2 \left( 1 + \frac{1}{\kappa \alpha_k \gamma_k} \right) | \alpha_k - \alpha_{k+1}|^2 \tag{2.3.8}
\]
\[
\leq \| u^k - v^k \|^2 (1 - 3 \kappa \alpha_k \gamma_k + c_5 \gamma_k^2) + c_6 \gamma_k^2 + c_4^2 \left( 1 + \frac{1}{\kappa \alpha_k \gamma_k} \right) | \alpha_k - \alpha_{k+1}|^2.
\]

Inequality (2.3.8) has a similar structure as inequality (A3.1.6). Therefore, to finish the proof we can apply Polyak’s Lemma A3.1.8 setting
\[
\mu_k := 3 \kappa \alpha_k \gamma_k - c_5 \gamma_k^2,
\]
\[
\beta_k := c_6 \gamma_k^2 + c_4^2 \left( 1 + \frac{1}{\kappa \alpha_k \gamma_k} \right) | \alpha_k - \alpha_{k+1}|^2.
\]
Therefore, in view of the conditions (2.3.2), one can conclude that
\[ \lim_{k \to \infty} \| u^k - v^k \| = 0, \]
and due to \( \lim_{k \to \infty} \| v^k - u^* \| = 0 \), it follows that \( \lim_{k \to \infty} \| u^k - u^* \| = 0. \)

2.3.2 Remark. If the functionals \( J \) and \( \omega \) are Fréchet-differentiable, one can prove similarly as above the convergence of the method
\[ u^{k+1} := \Pi_K (u^k - \gamma_k (p^k + \alpha_k \nabla \omega(u^k))), \quad k = 1, 2, \ldots, \quad (2.3.9) \]
with \( \| p^k - \nabla J(u^k) \| \leq \delta_k \).
If, in addition to the conditions (2.3.2), (2.3.3), it is required that \( \delta_k \geq 0 \), \( \lim_{k \to \infty} \delta_k \alpha_k = 0 \), then one can infer that \( \| u^k - u^* \| = 0 \). Method (2.3.9) is a regularized gradient projection method with inexactly calculated gradients, however, the conditions for the controlling parameters differ from those for Method (2.1.3).

2.3.2 Regularized penalty methods

As in Section 2.1.2 let us consider Problem (2.1.13) under the additional assumption that the existence of a saddle point of the Lagrangian is guaranteed.
To describe Tikhonov’s regularization approach coupled with penalty methods we chose the penalty function
\[ \phi_k(u) = r_k \sum_{j=1}^{m} \{ \max[0, g_j(u)] \}^2, \quad r_k > 0, \quad \lim_{k \to \infty} r_k = \infty, \]
and positive controlling sequences \( \{ \alpha_k \} \), \( \{ \mu_k \} \) and \( \{ \epsilon_k \} \) with
\[ \lim_{k \to \infty} \alpha_k = \lim_{k \to \infty} \mu_k = \lim_{k \to \infty} \epsilon_k = 0. \]
Again we define
\[ \theta_k := J + \alpha_k \omega, \quad F_k := \theta_k + \phi_k, \]
where \( \omega \) is a strongly convex (with constant \( \kappa \)), continuous and non-negative functional on the Hilbert space \( V \).
Let \( \{ \tilde{F}_k \} \) be a sequence of approximately given continuous functionals on \( V \) with
\[ \| F_k - \tilde{F}_k \|_{C(V)} \leq \mu_k, \quad k = 1, 2, \ldots, \quad (2.3.10) \]
then a regularized penalty method can be considered which generates a sequence of iterates \( \{ u^k \} \) such that
\[ \tilde{F}_k(u^k) \leq \inf_{u \in V} \tilde{F}_k(u) + \epsilon_k, \quad k = 1, 2, \ldots. \quad (2.3.11) \]
Now, we will investigate the convergence of this method.
Let \( (\bar{u}^k, \bar{\lambda}^k) \) be a saddle point of the Lagrangian corresponding to the regularized problem
\[ \min \{ \theta_k(u) : u \in K \}. \quad (2.3.12) \]
Obviously, because of the strong convexity of \( \theta_k \), the first component \( \bar{u}^k \) of the saddle point is the unique solution of (2.3.12).
2.3.3 Proposition. Under the conditions
\[
\lim_{k \to \infty} \frac{\epsilon_k}{\alpha_k} = 0, \quad \lim_{k \to \infty} \frac{\mu_k}{\alpha_k} = 0, \quad \lim_{k \to \infty} \|\bar{\lambda}_k\|^2 = 0, \tag{2.3.13}
\]
the sequence \(\{u_k\}\), generated by Method (2.3.11), converges to the normal solution \(u^*\) of Problem (2.1.13).

**Proof:** Denote for \(k = 1, 2, \cdots\)
\[
\theta^*_k := \min \{\theta_k(u) : u \in K\},
\]
\[
v_k := \arg \min \{F_k(u) : v \in V\}.
\]
In view of (2.3.10) and (2.3.11) the point \(u^k\) satisfies
\[
F_k(u^k) \leq F_k(v^k) + \epsilon_k + 2\mu_k, \quad k = 1, 2, \cdots,
\]
and since the functional \(F_k\) is strongly convex (with constant \(\kappa \alpha_k\)) we have
\[
\|u^k - v^k\| \leq \sqrt{\epsilon_k + 2\mu_k \kappa \alpha_k}. \tag{2.3.14}
\]
Using the definition of a saddle point, it follows for arbitrary \(u \in V\)
\[
F_k(u) - \theta^*_k = \theta_k(u) - \theta^*_k + r_k \sum_{j=1}^{m} \{\max[0, g_j(u)]\}^2
\]
\[
\geq - \sum_{j=1}^{m} \bar{\lambda}_j^k g_j(u) + r_k \sum_{j=1}^{m} \{\max[0, g_j(u)]\}^2
\]
\[
\geq - \sum_{j=1}^{m} \bar{\lambda}_j^k \max[0, g_j(u)] + r_k \sum_{j=1}^{m} \{\max[0, g_j(u)]\}^2
\]
\[
\geq \sum_{j=1}^{m} \min_{t_j \geq 0} (-\bar{\lambda}_j^k t_j + r_k t_j^2)
\]
\[
= - \frac{1}{4r_k} \sum_{j=1}^{m} (\bar{\lambda}_j^k)^2 = - \frac{\|\bar{\lambda}_k\|^2}{4r_k},
\]
i.e., due to \(F_k(\bar{u}^k) = \theta^*_k\), we obtain \(F_k(u) - F_k(\bar{u}^k) \geq - \frac{\|\bar{\lambda}_k\|^2}{4r_k}\).
Hence,
\[
F_k(\bar{u}^k) - F_k(v^k) \leq \frac{\|\bar{\lambda}_k\|^2}{4r_k},
\]
and, with regard to the definition of \(v^k\) and the strong convexity of \(F_k\), we infer
\[
\|\bar{u}^k - v^k\| \leq \frac{\|\bar{\lambda}_k\|}{2\sqrt{r_k \alpha_k}}. \tag{2.3.15}
\]
Owing to (2.3.14) and (2.3.15) it follows that
\[
\|\bar{u}^k - u^k\| \leq \sqrt{\frac{\epsilon_k + 2\mu_k \kappa \alpha_k}{2\sqrt{r_k \alpha_k}}} + \frac{\|\bar{\lambda}_k\|}{2\sqrt{r_k \alpha_k}}.
\]
and, in view of (2.3.13), \( \lim_{k \to \infty} \| \bar{u}^k - u^k \| = 0 \).

But Proposition 1.3.9 says that \( \lim_{k \to \infty} \bar{u}^k = u^*_\omega \), hence, \( \lim_{k \to \infty} u^k = u^*_\omega \). \( \square \)

Because the values \( \| \lambda^k \| \) in (2.3.13) are unknown, it is essential to know whether there exists an upper bound for the sequence \( \{ \lambda^k \} \) of Lagrange multipliers.

To investigate this, we suppose that the feasible set in Problem (2.1.13) has the following structure

\[ K := \{ u \in V : g_j(u) \leq 0, \; j \in I = \{1, \ldots, m\} \}, \quad I = I_1 \cup I_2, \]

where the constraint functions \( g_j \) (\( j \in I_1 \)) are nonlinear, convex and continuous on \( V \) and \( g_j \) (\( j \in I_2 \)) are affine, i.e.

\[ g_j(u) = \langle a^j, u \rangle - b_j, \quad j \in I_2 \subset I \]

with \( a^j \in V, \; b_j \in \mathbb{R} \).

2.3.4 Lemma. Let \( \{ J_i \} \) be a sequence of convex, continuous functionals bounded from above at any point \( u \in V \). Assume that each \( J_i \) attains its minimizer on the set \( K \) and

\[ d_0 = \inf_{u \in K} \min_i J_i(u) > -\infty. \]

Furthermore, assume that the following regularity condition is fulfilled:

there exists a point \( \bar{u} \) such that

\[ g_j(\bar{u}) < 0 \quad \text{if} \quad j \in I_1 = \{1, \ldots, m\} \setminus I_2, \]

\[ g_j(\bar{u}) \leq 0 \quad \text{if} \quad j \in I_2. \]

Then, for each problem

\[ \min \{ J_i(u) : u \in K \} \quad (2.3.16) \]

there exists a Lagrange multiplier vector \( \eta^i \) such that \( \sup_i ||\eta^i|| < \infty \).

**Proof:** The assumptions of the lemma ensure that for each fixed \( i \) there is a saddle point \( (z^i, \eta^i) \) of the Lagrange function corresponding to (2.3.16). Thus, we get in particular

\[ g_j(z^i) \leq 0, \quad \eta^i_j g_j(z^i) = 0, \quad j = 1, \ldots, m. \]

If \( \eta^i_{j_0} = 0 \) for some \( j_0 \in I_2 \), then the point

\[ (z^i, \eta^i_1, \ldots, \eta^i_{j_0-1}, \eta^i_{j_0+1}, \ldots, \eta^i_m) \]

is a saddle point of a problem arising from (2.3.16) by deleting the constraint \( g_{j_0}(u) \leq 0 \). Now, we drop all such constraints and consider the resulting problem with index set \( I^+_2(i) := \{ j \in I_2 : \eta^i_j > 0 \} \) instead of \( I_2 \). Then for the pairs \( (z^i, \eta^i_{j})_{j \in I_1 \cup I^+_2(i)} \) the relation

\[ J_i(u) + \sum_{j \in I_1} \eta^i_j g_j(u) + \sum_{j \in I^+_2(i)} \eta^i_j g_j(u) \geq J_i(z^i) \]

\[ \geq J_i(z^i) + \sum_{j \in I_1} \lambda_j g_j(z^i) + \sum_{j \in I^+_2(i)} \lambda_j g_j(z^i) \quad (2.3.17) \]
is true for any $u \in V$, $\lambda_j \geq 0$ $(j \in I_1 \cup I^*_2 (i))$, moreover,
\[ \eta^*_j > 0, \quad g^*_j(z^*) = 0 \quad \forall j \in I^*_2 (i). \]

If the vectors $a^j \in V$, $j \in I^*_2 (i)$, are linearly dependent, it is possible to reduce this vector system to a system of linearly independent vectors (see for instance the proof of Caratheodory’s theorem in ROCKAFELLAR [348]). Hence, we identify a subsystem of linearly independent vectors $a^j$, $j \in I_2 (i) \subset I^*_2 (i)$, such that
\[ \sum_{j \in I_2^2 (i)} \eta^*_j a^j = \sum_{j \in I_2 (i)} \mu^*_j a^j \]
with $\mu_j > 0$. Furthermore, for any $u \in V$
\[ \sum_{j \in I_2^2 (i)} \eta^*_j g_j (u) = \sum_{j \in I_2^2 (i)} \eta^*_j ([a^j, u] - [a^j, z^*]) \]
\[ = \sum_{j \in I_2 (i)} \mu^*_j ([a^j, u] - [a^j, z^*]) = \sum_{j \in I_2 (i)} \mu^*_j g_j (u), \]
and because of $g_j (z^*) = 0$ for $j \in I^*_2 (i)$, we see that relation (2.3.17) is equivalent to
\[
J_i (u) + \sum_{j \in I_1} \eta^*_j g_j (u) + \sum_{j \in I_2 (i)} \mu^*_j g_j (u) \\
\geq J_i (z^*) + \sum_{j \in I_1} \eta^*_j g_j (z^*) + \sum_{j \in I_2 (i)} \mu^*_j g_j (z^*) \\
\geq J_i (z^*) + \sum_{j \in I_1} \lambda_j g_j (z^*) + \sum_{j \in I_2 (i)} \mu_j g_j (z^*)
\]
for all $\lambda_j \geq 0$ $(j \in I_1)$, $\mu_j \geq 0$ $(j \in I_2 (i))$. Therefore, $(z, \eta^*_j, \mu^*_j)$ is a saddle point of the problem
\[
\min \{ J_i (u) : g_j (u) \leq 0, \quad j \in I_1 \cup I_2 (i) \}. 
\]
Setting $\mu_j := 0$ for $j \in I_2 \setminus I_2 (i)$, then $(z^*, \eta^*_j, \mu^*_j)$ is a saddle point for Problem (2.3.16).

Now, we show the existence of a constant $\gamma_0 < 0$ such that the system of inequalities
\[ g_j (u) \leq \gamma_0, \quad \forall j \in I_1 \cup S, \]
is solvable for all $S \subset \Sigma$, with $\Sigma$ the set of all subsets of $I_2$ for which the corresponding families of vectors $a^j$ are linearly independent. Indeed, let $\text{aff}(a)$ be the affine hull of the points $a^j$ for all $j \in S$. In view of the linear independence of the vectors $a^j$ we have $0 \notin \text{aff}(a)$. Let $\xi \in \text{aff}(a)$ be the closest point to zero, then $[a^j - \xi, \xi] = 0$ for $j \in I_2$, hence $[a^j, -\xi] < 0$.

For $u_S := \tilde{u} - \gamma(S) \xi$, due to the continuity of $g_j$ and the assumed regularity condition, it follows that for a suitable $\gamma(S) > 0$ we get
\[ g_j (u_S) < 0, \quad \forall j \in I_1 \cup S. \]
Now, because $\Sigma$ is a finite set, one can choose
\[ \gamma_0 := \max_{S \subseteq \Sigma} \max_{j \in I_1 \cup S} g_j(u_S). \]
In the remainder of the proof we consider the point $u_S$ to be fixed. In view of
\[ \sum_{j \in I_1} \eta_j^i g_j(z^i) = 0 \quad \text{and} \quad \sum_{j \in I_2(i)} \mu_j^i g_j(z^i) = 0, \]
using the left part of inequality (2.3.18) with $u := u_{S_i}, S_i = I_2(i)$, we get
\[ \sum_{j \in I_1} \eta_j^i + \sum_{j \in I_2} \mu_j^i = \sum_{j \in I_1} \eta_j^i + \sum_{j \in I_2(i)} \mu_j^i \leq \frac{J_i(u_{S_i}) - J_i(z^i)}{-\max_{j \in I_1 \cup S_i} g_j(u_{S_i})} \leq -\frac{1}{\gamma_0} \left( \sup_i J_i(u_{S_i}) - d_0 \right). \]
Hence, for the Lagrange multiplier vector $\bar{\eta}^j = (\eta_j^i | j \in I_1, \mu_j^i | j \in I_2)$ of Problem (2.3.16) the estimate
\[ \|\bar{\eta}^j\| \leq -\frac{1}{\gamma_0} \left( \sup_i J_i(u_{S_i}) - d_0 \right) \]
holds true uniformly for all $i$. To finish the proof it is necessary to note that the set $\Sigma$ is finite and $\sup_i J_i(u_{S_i}) < \infty$, since $\{J_i(u)\}$ was supposed to be bounded from above for any $u \in V$. \hfill \Box

2.3.5 Corollary. If the constraints in Problem (2.1.13) satisfy the regularity condition of Lemma 2.3.4, then there exists a Lagrange multiplier vector $\lambda^k$ for each of the Problems (2.3.12) such that $\sup_k \|\lambda^k\| < \infty$ and
\[ \frac{\|\lambda^k\|}{\sqrt{r_k \alpha_k}} \to 0 \quad \text{for} \quad r_k \alpha_k \to \infty. \]

2.3.6 Remark. If the Slater condition is fulfilled for Problem (2.1.13), then the proof of Lemma 2.3.4 implies that for any Lagrange multiplier vector $\lambda^k$ of Problem (2.3.12) and any $k$ the estimate
\[ \|\lambda^k\| \leq (J(\bar{u}) + \alpha_k \omega(\bar{u}) - J^*) \left( -\max_{1 \leq j \leq m} g_j(\bar{u}) \right)^{-1} \]
holds, with $\alpha_0 = \sup_k \alpha_k$. \hfill \diamond

Under the Slater condition the proof of Proposition 2.3.3 can be extended easily for other penalty functions, for instance, for barrier functions of the type (A3.4.78), too. If the choice of $\phi_k$ guarantees that $v^k \in K$, then, in order to calculate $\|\bar{u}^k - u^k\|$, we can immediately make use of the estimates for the rate of convergence of the penalty method, see for instance, Proposition A3.4.42. Indeed, these are of the form
\[ \theta_k(v^k) - \theta_k^* \leq \eta(\alpha_k, r_k, \epsilon_k) \]
and if $v^k \in K$, due to the strong convexity of $\theta_k$,
\[ \|\bar{u}^k - v^k\| \leq \sqrt{\frac{\eta(\alpha_k, r_k, \epsilon_k)}{r_k \alpha_k}}. \]
We are ready if it is possible to establish that the controlling parameters $\alpha_k, r_k, \epsilon_k$ can be chosen such that
\[
\eta(\alpha_k, r_k, \epsilon_k) \to 0.
\]
This, however, is not difficult. Grossmann and Kaplan [151] show that one can bound $\eta(\alpha_k, r_k, \epsilon_k)$ from above, uniformly with respect to $\alpha_k$, such that
\[
\eta(\alpha_k, r_k, \epsilon_k) \leq \bar{\eta}(r_k, \epsilon_k)
\]
with $\bar{\eta}(r_k, \epsilon_k) \to 0$ for $r_k \to \infty$ and $\epsilon_k \to 0$.

\section*{2.4 Mosco’s Approximation Scheme}

In this section we investigate an approach, suggested by Mosco [296], in order to establish stable sequential approximation methods for ill-posed variational inequalities with monotone, in general non-potential, operators.

This approach, based on a concept of the convergence of sets and functionals, is convenient for the study of standard approximations for elliptic variational inequalities. We start with the description of Mosco’s scheme for the case that Problem (2.1.1) is well-posed with respect to the class of data perturbations, defined in terms of Mosco’s concept. After that we turn to the ill-posed case.

Because this technique will be not used later on, all results are given without proofs. Moreover, some notions and statements are formulated for particular purpose.

Let $V$ be a Hilbert space and $\{C_n\}$ be a sequence of convex, closed subsets in $V$. Denote
\[
\text{Lim}_s C_n : \text{the set of all points } v \in V \text{ for which there is a sequence }
\{v^n\}, v^n \in C_n, \forall n \in \mathbb{N}, \text{ with } v^n \to v;
\]
\[
\text{Lim}_w C_n : \text{the set of all points } v \in V \text{ such that there is a sequence }
\{v^k\}, v^k \in C_{n_k}, k = 1, 2, \cdots, \text{ and } v^k \rightharpoonup v,
\]
where $\{C_{n_k}\}$ is a subsequence of $\{C_n\}$.

\subsection*{2.4.1 Definition.} A sequence $\{C_n\}$ of convex closed sets is called M-convergent to the set $C$ in $V$ if
\[
\text{Lim}_s C_n = \text{Lim}_w C_n = C.
\]
In this case we write $C = \text{Lim}C_n$.

It is obvious that in the limit $C$ is a convex and closed set in $V$.

\subsection*{2.4.2 Definition.} A sequence $\{f_n\}$ of convex and lower-semicontinuous functionals $f_n : V \to \mathbb{R}$ is called M-convergent to the functional $f$ on $V$ if
\[
\text{epi} f = \text{Lim}(\text{epi} f_n) \text{ in } V \otimes \mathbb{R},
\]
where $V \otimes \mathbb{R}$ is the space $V \times \mathbb{R}$ endowed with the scalar product, which is generated by the scalar product in $V$ and induces the norm
\[
\|(v, \beta)\| = \sqrt{\|v\|^2 + |\beta|^2}, \quad (v, \beta) \in V \times \mathbb{R}.
\]
Here we write $f = \text{Lim}f_n$.
It is easy to verify that in the limit the functional \( f \) is convex and lower-semicontinuous.

**2.4.3 Proposition.** The relation \( f = \operatorname{Lim} f_n \) is valid iff for any \( v \in V \)

(i) there exists a sequence \( \{v^n\} \subset V \) with

\[
v^n \to v, \quad \limsup_{n \to \infty} f_n(v^n) \leq f(v),
\]

and

(ii) for any subsequence \( \{f_{n_k}\} \subset \{f_n\} \) and any weakly convergent sequence \( \{v^k\} \) the following implication holds

\[
v^k \rightharpoonup v \implies \liminf_{k \to \infty} f_{n_k} \geq f(v).
\]

In finite-dimensional spaces \( \operatorname{M} \)-convergence of functionals is equivalent to the notion of \( \operatorname{epi} \)-limits, see for instance, Rockafellar and Wets [353].

Now, let us consider Mosco’s approach for the problem mentioned:

\[
\begin{array}{ll}
\text{minimize} & J(u), \\
\text{subject to} & u \in K,
\end{array}
\]

(2.4.1)

where \( J : V \to \mathbb{R} \) is a convex, lower semicontinuous functional, \( K \subset V \) is a convex, closed set.

This problem has to be approximated by a sequence of auxiliary problems

\[
\min \{ J_i(u) : u \in K_i \}, \quad \forall i \in \mathbb{N},
\]

(2.4.2)

where the functionals \( J_i \) and sets \( K_i \) are supposed to satisfy the same conditions as \( J \) and \( K \) in (2.4.1).

### 2.4.1 Approximation of well-posed problems

Suppose that the following assumptions are satisfied for the problems (2.4.1) and (2.4.2), respectively.

**2.4.4 Assumption.**

(i) \( \text{dom} J \supset K, \text{dom} J_i \supset K_i \ \forall i \in \mathbb{N}, \text{and} \ J = \operatorname{Lim} J_i; \)

(ii) \( K = \operatorname{Lim} K_i; \)

(iii) for any \( u \in K \) there exists a continuous, strictly increasing function \( \gamma : \mathbb{R}_+ \to \mathbb{R}_+ \) with \( \gamma(0) = 0 \) and

\[
\gamma(||v - u||) \leq \liminf_{i \to \infty} |J_i(v) - J(u)|,
\]

the latter inequality is supposed to be true uniformly for arbitrary \( v \in W \) of any bounded \( W \subset V \).
(iv) there exists a function $\zeta : \mathbb{R}_+ \to \mathbb{R}_+$ with $\zeta (r) \to \infty$ for $r \to \infty$ such that for any sequence $\{v^i\}, v^i \in K_i$, there is a bounded sequence $\{z^i\}, z^i \in K_i$, with

$$\zeta (\|v^i - z^i\|) \leq \frac{|J_i(v^i) - J_i(z^i)|}{\|v^i - z^i\|};$$

(v) Problems (2.4.2) are solvable for sufficiently large $i$.

Condition (iv) and (v) ensure not only solvability of Problems (2.4.2) for large $i$, but also the uniform boundedness of their optimal sets. Taking this into consideration, due to Assumption 2.4.4 (i) and (ii), the solvability of Problem (2.4.1) can be concluded and in case the solution $u^*$ of (2.4.1) is unique, weak convergence of $\{u^i\}$ to $u^*$ can be proved, with $u^i$ any solution of the $i$-th auxiliary problem (2.4.2). Finally, in view of Assumption 2.4.4 (iii), $\{u^i\} \to u^*$ implies strong convergence $\{u^i\} \to u^*$ in the norm of $V$.

2.4.5 Theorem. Suppose that the Problems (2.4.1) and (2.4.2) satisfy all the conditions required, in particular Assumption 2.4.4 (i)-(v). Then the auxiliary problems (2.4.2) are solvable for sufficiently large $i \geq i_0$ and their optimal sets $U^*_i$ are uniformly bounded. Moreover, Problem (2.4.1) is also solvable and if its solution $u^*$ is unique, then any sequence $\{u^i\}, u^i \in U^*_i, i \geq i_0$, converges to $u^*$ in the norm of the space $V$.

If $J$ is a strictly convex functional on dom $J$, then it immediately follows that Problem (2.4.1) is well-posed with respect to the class of perturbations $\{J_i, K_i\}$ defined by the conditions (i)-(v) in Assumption 2.4.4.

2.4.6 Remark. Formally, the particular consideration of the sets $K$ and $K_i$ is not necessary. Taking

$$\varphi (\cdot) = J (\cdot) + \text{ind}(\cdot | K), \quad \varphi_i (\cdot) = J_i (\cdot) + \text{ind}(\cdot | K_i),$$

with ind the indicator functional of the corresponding sets, the Problems (2.4.1) and (2.4.2) are equivalent to the unconstrained minimization of $\varphi$ and $\varphi_i$, respectively. Besides the functionals $\varphi, \varphi_i$ are of the same class as $J$ and $J_i$ (see Corollary of Theorem B in [296]). However, in view of a numerical treatment, it is more convenient to establish estimates of the approximation for the objective functional as well as for the feasible set separately.

2.4.2 Mosco’s scheme for ill-posed problems

Now, let us suppose that the conditions (iii) - (v) in Assumption 2.4.4 are not satisfied and Problem (2.4.1) may have more than one solution. In this case, in order to obtain stable approximations of Problem (2.4.1), a sequence of regularized auxiliary problems

$$\min \{J_i (u) + \alpha_i \|u\|^2 : u \in K_i\}, \quad \forall i \in \mathbb{N}, \quad (2.4.3)$$

is considered, where $J_i, K_i$ satisfy the assumptions required for $J$ and $K$ in (2.4.1) and $\alpha_i \downarrow 0$.

Here, to obtain convergence for the sequence $\{u^i\}$ of solutions of these auxiliary problems, more restrictive assumptions on the approximations $J_i$ and $K_i$ are required, respectively.
2.4.7 Definition. A sequence \{C_n\} of convex and closed sets \(C_n \in V\) is called convergent of the order \(\geq \sigma\) (\(\sigma \geq 0\) is fixed) to a convex closed set \(C \subset V\) (write \(n^\sigma [C_n - C] \to 0\)), if for any \(v \in C\) there exist \(v_n \in C_n\) such that
\[ n^\sigma (v_n - v) \to 0 \text{ in } V \text{ as } n \to \infty, \]
and for any weakly convergent sequence \{v^k\}, \(v^k \in C_{n_k}\), there exist \(v^{k_j} \in \{v^k\}\) and \(\{w^j\} \subset C\) such that
\[ n^\sigma_{k_j} (v^{k_j} - w^j) \to 0 \text{ in } V \text{ as } j \to \infty, \]
with \(\{C_{n_k}\}\) a subsequence of \(\{C_n\}\).

2.4.8 Definition. For a fixed \(\sigma \geq 0\) a sequence \(\{f_n\}, f_n : V \to \mathbb{R}\), of convex and lower-semicontinuous functionals is called convergent of the order \(\geq \sigma\) to a convex, lsc functional in \(V\) (write \(n^\sigma [f_n - f] \to 0\)), if
\[ n^\sigma [\text{epi } f_n - \text{epi } f] \to 0 \text{ in } V \oplus \mathbb{R} \text{ in the sense of Definition 2.4.7}. \]

2.4.9 Proposition. Functional convergence \(n^\sigma [f_n - f] \to 0\) holds true iff for any \(v \in \text{dom } f\) there is a sequence \(\{v_n\} \subset V\) such that
\[ n^\sigma (v^n - v) \to 0 \text{ in } V \text{ as } n \to \infty, \]
\[ \limsup_{n \to \infty} n^\sigma (f_n(v^n) - f(v)) \leq 0, \]
and from any weakly convergent \(\{v^k\} \subset V\), with \(\limsup_{k \to \infty} f_{n_k}(v^k) < +\infty\) for a chosen \(\{f_{n_k}\}\), a subsequence \(\{v^{k_j}\}\) can be selected such that for some \(\{w^j\} \subset V\) the relations
\[ n^\sigma_{k_j} (v^{k_j} - w^j) \to 0 \text{ in } V \text{ as } j \to \infty, \]
\[ \liminf_{j \to \infty} n^\sigma_{k_j} (f_{n_{k_j}}(v^{k_j}) - f(w^j)) \geq 0 \]
are fulfilled.

Now, suppose that for a fixed \(\sigma > 0\) the approximation of Problem (2.4.1) by means of the sequence of Problems (2.4.3) satisfies the following requirements:

2.4.10 Assumption.
(i) \(\text{dom } J \supset K\), \(\text{dom } J_i \supset K_i \ \forall \ i \in \mathbb{N}\);
(ii) \(i^\sigma [J_i - J] \to 0\);
(iii) \(i^\sigma [K_i - K] \to 0\);
(iv) for any sequence \(\{v^i\}, v^i \in K_i\), with \(\|v^i\| \to \infty\) as \(i \to \infty\) there is a sequence \(\{z^i\} \subset K\) such that
\[ \limsup_{i \to \infty} \frac{i^\sigma (J(z^i) - J_i(v^i))}{\|v^i\|} < \infty. \]
2.4.11 Theorem. Suppose that the Problems (2.4.1) and (2.4.3) satisfy all the conditions required, in particular Assumption 2.4.10 (i)-(v) with some \( \sigma > 0 \). Moreover, the optimal set \( U^* \) of Problem (2.4.1) is supposed to be non-empty and in the regularized problems (2.4.3) the regularization parameter \( \alpha_i := \sigma \) is chosen.

Then the uniquely determined sequence \( \{ u^i \} \) of solutions of Problems (2.4.3) converges strongly to the norm-minimal solution \( u^* \) of Problem (2.4.1), i.e., \( \| u^* \| \leq \| u \| \forall u \in U^* \).

In [296] the approach described above is applied to variational inequalities of the type

\[ u \in K : \quad \langle Qu, v - u \rangle \geq 0, \quad \forall v \in K, \tag{2.4.4} \]

and hemi-variational inequalities of the type

\[ u \in V : \quad \langle Qu, v - u \rangle + f_i(v) - f_i(u) \geq 0, \quad \forall v \in K, \tag{2.4.5} \]

where \( Q : V \to V' \) is a monotone and hemi-continuous operator, \( f \) is a convex, lsc functional on \( V \) and \( K \subset V \) is a convex and closed set. For the definition of a hemi-continuous operator see, for instance, Definition A1.6.38.

In the case of well-posed problems the corresponding auxiliary problems have the form

\[ u^i \in K_i : \quad \langle Q_i u^i, v - u^i \rangle \geq 0, \quad \forall v \in K_i, \tag{2.4.6} \]

\[ u^i \in V : \quad \langle Q_i u^i, v - u^i \rangle + f_i(v) - f_i(u^i) \geq 0, \quad \forall v \in K_i, \tag{2.4.7} \]

respectively, and in the case of ill-posedness

\[ u^i \in K_i : \quad \langle (Q_i + i\sigma T) u^i, v - u^i \rangle \geq 0, \quad \forall v \in K_i, \tag{2.4.8} \]

\[ u^i \in K_i : \quad \langle (Q_i + i\sigma T) u^i, v - u^i \rangle + f_i(v) - f_i(u^i) \geq 0, \quad \forall v \in K_i. \tag{2.4.9} \]

Here \( Q_i \) and \( f_i \) satisfy the same properties as \( Q \) and \( f \), and the operator \( T \) belongs to a sufficient broad class of operators performing a Tikhonov regularization for Problem (2.4.1). For the minimization problem (2.4.3) just \( T u := u \) has to be used.

The proofs of the Theorems B and D in Mosco’s paper [296] and their corollaries about the convergence of the approaches considered can be adapted to our case if we set \( Q = Q_i := 0 \) and \( f := J + \text{ind}(\cdot | K), \ f_i := J_i + \text{ind}(\cdot | K) \) and apply the transition from inequality (7.4) to inequality (7.5) given in Section 0.9 of the paper mentioned.

Mosco’s approach for solving Problem (2.4.1) can be considered, on the one hand, as a regularization of this problem if \( J_i \) and \( K_i \) are interpreted as approximately given data and, on the other hand, as a regularization of the discretization method if, for instance, \( J_i \) and \( K_i \) are derived by some standard discretization of the original problem (2.4.1). Such dual interpretation is also possible in a number of other approaches for solving ill-posed problems, i.e., there is no strict distinction between classical regularization methods and methods of iterative regularization.
2.5 Comments

Section 2.1: For a long time investigations of optimization methods for solving ill-posed convex variational problems were restricted merely in order to proof the convergence towards the optimal set. The first result comprising better properties of convergence was established for the gradient method with a constant step-size. Such a statement, similar to Proposition 2.1.1 (with $K = V$ and $\delta_k \equiv 0$), was formulated without proof by Polyak [328]. Later on a proof was given by Gol’shtein and Tretyakov [139]. In order to establish here the convergence of the gradient method with inexact computation of the gradients we basically were following the proof given in [141].

Proposition 2.1.1 shows the stability of the gradient - and gradient projection method, applied to ill-posed convex variational problems under sufficiently smooth data and small perturbations of the objective functional with respect to the norm in $C^1(V)$. In the infinite setting stability is to be understood in terms of weak convergence.

Lemma A1.5.33, often used to investigate optimization methods, was proved by Gol’shtein and Tretyakov [139]. A proof of Theorem 2.1.4 can be found in Kaplan [205]. Here the rule for adapting the controlling parameters in the penalty methods (see also Remark 2.1.6) corresponds to the idea of regularization by means of a timely termination at each approximation level of the iteration procedure. This idea was exploited for linear problems by many authors, for instance, see Lavrentyev [258], Bakushinski [29], Emelin and Krasnoselski [103] and, in connection with the method of feasible directions for the solution of elliptic variational inequalities of the obstacle problem, by Kirsten and Tichatschke [235]. An approach towards regularization via discretization was made by Natterer [303].

Section 2.2: see the comments to Section 3.2.

Section 2.3: In our description of regularized subgradient projection methods we are going along with the papers of Bakushinski and Polyak [31] and Bakushinski and Goncharski [30]. Tikhonov-regularized penalty methods with various penalty functions are investigated by Vasiljev [407, 408]. In his first paper a statement similar to Proposition 2.3.3 is proved under a weaker regularity condition. In our opinion, Lemma 2.3.4 gives a new result which could be useful in connection with the uniform estimation of the Lagrange multiplier vectors of a family of regularized problems. By the way, a Tikhonov-regularized Newton method for solving variational inequalities has been established in [30], see also Friedrich et al. [121].

Section 2.4: Executable versions of Mosco’s scheme are developed in a number of papers mainly dealing with the discrete approximation of problems in mathematical physics. Discrete approximations of ill-posed problems were investigated by Douglas [93] in connection with the solution of integral equations of the first kind. Among the succeeding papers we mention the monograph by Vainikko [406], where an abstract framework is given for investigating methods based on discrete approximations, including those with iterative regularization. Also the monograph of Hofmann [185] should be mentioned, dealing with applications of regularized discretization methods. We further refer to a number of
publications including the relations between Mosco-approximation and Moreau-Yoshida-approximation, see for instance, Attouch [16], Lemaire [260, 261]. In Tossings [399] Mosco’s scheme is suggested for solving ill-posed problems, where the discretized problems are solved approximately.
CHAPTER 2. STABLE DISCRETIZATION AND SOLUTION METHODS
Chapter 3

PPR FOR
FINITE-DIMENSIONAL
PROBLEMS

3.1 Properties of the Proximal-Point-Mapping

Let $C$ be a non-empty, closed and convex subset of a Hilbert space $V$ and $f : V \to \mathbb{R}$ be a lower semi-continuous convex functional. We recall that only proper convex functionals are considered in this book.

For arbitrary $u \in V$ the proximal-point-mapping is defined by

$$
\text{Prox}_{f,C} u := \arg\min_{v \in C} \{ f(v) + \frac{\chi}{2} \|v - u\|^2 \}, \quad \chi > 0 \text{ fixed.} \tag{3.1.1}
$$

Due to the strong convexity and lower semi-continuity of

$$
\Psi_{\chi,u} : (\cdot) := f(\cdot) + \frac{\chi}{2} \|\cdot - u\|^2,
$$

the proximal-point-mapping (prox-mapping for short) is always single-valued. Obviously, the set of fixed points of this mapping coincides with the set $C^*$ of minimizers of the function $f$ on $C$, i.e.,

$$
u^* \in C^* \iff \text{Prox}_{f,C} u^* = u^*, \quad \forall \ \chi > 0.
$$

Moreover, it holds

$$
\text{Prox}_{f,C} u = \text{Prox}_{f,V} u
$$

for $\bar{f} := f + \text{ind}(\cdot|C)$ (ind(·) the indicator functional).

3.1.1 Proposition. The operator $\text{Prox}_{f,C}$ is firmly non-expansive, i.e., for arbitrary $x, y \in V$ the following inequality is fulfilled:

$$
\| \text{Prox}_{f,C} x - \text{Prox}_{f,C} y \|^2 \leq \| x - y \|^2 - \| x - (\text{Prox}_{f,C} x - \text{Prox}_{f,C} y) \|^2.
$$
Proof: From Proposition A1.5.34 we get for \( v := \text{Prox}_{f,C} x \) and \( w := \text{Prox}_{f,C} y \):

\[
\begin{align*}
f(u) - f(v) + \chi \langle v - x, u - v \rangle & \geq 0 \quad \forall \ u \in C, \\
f(z) - f(w) + \chi \langle w - y, z - w \rangle & \geq 0 \quad \forall \ z \in C,
\end{align*}
\]

(3.1.2)

hence, setting \( u := w \), \( z := v \) it follows

\[
\langle y - x, w - v \rangle \geq \| w - v \|^2
\]

and

\[
\begin{align*}
\|v - w\|^2 & \leq \|x - y\|^2 + \langle y - x, w - v \rangle - \|x - y\|^2 \\
& \leq \|x - y\|^2 + 2\langle y - x, w - v \rangle - \|w - v\|^2 - \|x - y\|^2 \\
& = \|x - y\|^2 - \|x - y - (v - w)\|^2.
\end{align*}
\]

\[ \Box \]

3.1.2 Remark. Proposition 3.1.1 tells us that the prox-mapping is non-expansive, i.e.,

\[
\| \text{Prox}_{f,C} x - \text{Prox}_{f,C} y \| \leq \|x - y\|, \quad \forall \ x, y \in V.
\]

\[ \diamond \]

3.1.3 Proposition. Let \( y \in C \) be an arbitrary point. Then for any \( v^0 \in V \) and \( v^1 := \text{Prox}_{f,C} v^0 \) it holds

\[
\|v^1 - y\|^2 - \|v^0 - y\|^2 \leq -\|v^1 - v^0\|^2 + \frac{2}{\chi} [f(y) - f(v^1)], \quad (3.1.3)
\]

\[
\|v^1 - y\| \leq \|v^0 - y\| + \begin{cases} 0 & \text{if } f(y) \leq f(v^1), \\ \sqrt{\frac{2}{\chi} [f(y) - f(v^1)]} & \text{if } f(y) > f(v^1). \end{cases} \quad (3.1.4)
\]

Proof: Substituting in inequality (3.1.2) \( x := v^0, v := v^1, u := y \), we get

\[
f(y) - f(v^1) + \chi \langle v^1 - v^0, y - v^1 \rangle \geq 0.
\]

Hence,

\[
\begin{align*}
\|v^1 - y\|^2 - \|v^0 - y\|^2 & = -\|v^1 - v^0\|^2 + 2\langle v^1 - v^0, v^1 - y \rangle \\
& \leq -\|v^1 - v^0\|^2 + \frac{2}{\chi} [f(y) - f(v^1)]
\end{align*}
\]

and inequality (3.1.4) can be deduced immediately from here. \[ \Box \]

Without assuming differentiability of the function \( f \), the so called Moreau-Yosida-Regularization

\[
\eta(u) := \min_{v \in C} \{f(v) + \frac{\chi}{2} \|v - u\|^2\}
\]

has the following properties:
3.1. PROPERTIES OF THE PROXIMAL-POINT-MAPPING

3.1.4 Proposition. The functional

$$\eta(u) := \min \{ f(v) + \frac{\chi}{2} \|v - u\|^2 : v \in C\}$$

is convex and continuously Fréchet-differentiable on $V$, and

$$\nabla \eta(u) = \chi (u - \text{Prox}_{f,C} u).$$

Proof: The proof of the convexity of $\eta$ is straightforward.

In order to estimate for any $u, h \in V$ the absolute value of

$$q := \min_{v \in C} \{ f(v) + \frac{\chi}{2} \|v - (u + h)\|^2 \} - \min_{v \in C} \{ f(v) + \frac{\chi}{2} \|v - u\|^2 \} - \chi \langle u - \text{Prox}_{f,C} u, h \rangle,$$

we consider two cases.

(i) $q \geq 0$: Then

$$|q| \leq \frac{\chi}{2} \|u - \text{Prox}_{f,C} u - (u + h)\|^2 - f(\text{Prox}_{f,C} u) - \chi \langle u - \text{Prox}_{f,C} u, h \rangle,$$

hence

$$|q| \leq \langle -\chi h, \text{Prox}_{f,C} u - u - \frac{1}{2} h \rangle - \langle \chi h, u - \text{Prox}_{f,C} u \rangle = \frac{\chi}{2} \|h\|^2. \quad (3.1.5)$$

(ii) $q < 0$: Then we conclude

$$|q| \leq \frac{\chi}{2} \|\text{Prox}_{f,C} (u + h) - (u + h)\|^2 - f(\text{Prox}_{f,C} (u + h)) - \chi \langle u - \text{Prox}_{f,C} u, h \rangle$$

$$= |\langle -\chi h, \text{Prox}_{f,C} (u + h) - u - \frac{1}{2} h \rangle - \chi h, u - \text{Prox}_{f,C} u \rangle|$$

$$= | - \langle \chi h, \text{Prox}_{f,C} (u + h) - \text{Prox}_{f,C} u \rangle + \frac{\chi}{2} \|h\|^2 \rangle.$$

The non-expansivity of the prox-mapping yields

$$|q| \leq -\chi \|h\| \|u + h - u\| + \frac{\chi}{2} \|h\|^2 = \frac{\chi}{2} \|h\|^2. \quad (3.1.6)$$

Now, the differentiability of $\eta$ follows from (3.1.5) and (3.1.6) where as the continuity of its gradient is a conclusion of the non-expansivity of the prox-mapping

$$\|\nabla \eta(u) - \nabla \eta(v)\| = \chi \|u - \text{Prox}_{f,C} u - v + \text{Prox}_{f,C} v\| \leq 2\chi \|u - v\|.$$

Next, let us consider some properties of the exact proximal point algorithm (PPA) applied to Problem (1.2.4).
3.1.5 Algorithm. (Proximal point algorithm)

Data: $u^0 \in K$, $\{\chi_k\} \in [\chi, \infty)$, $(\chi > 0$, $\infty < \infty)$;

S0: Set $k := 0$;

S1: if $u^k$ solves the problem, stop;

S2: compute

$$u^{k+1} := \arg\min_{u \in K} \{J(u) + \frac{\chi_k}{2} \|u - u^k\|^2\}, \quad (3.1.7)$$

set $k := k + 1$ and go to S1.

\[\diamond\]

The following result goes back to GÜLER [153],

3.1.6 Proposition. For any $u \in K$ and iteration $n \geq 1$ in the PPA 3.1.5 it holds

$$J(u^n) - J(u) \leq \frac{1}{2\gamma_n} \left[\|u - u^0\|^2 - \|u - u^n\|^2\right] - \frac{\gamma_n\chi^2}{2} \|u^n - u^{n-1}\|^2, \quad (3.1.8)$$

with $\gamma_n := \sum_{k=1}^n \frac{1}{\chi_k}$.

Proof: First we show that the sequence $\{\chi_k\|u^{k-1} - u^k\|\}$ is non-increasing.

For the functional $J := J + \text{ind}(|K)$, in view of the definition of $u^k$, we get the inclusions

$$\chi_k(u^{k-1} - u^k) \in \partial J(u^k), \quad \chi_{k+1}(u^k - u^{k+1}) \in \partial J(u^{k+1}),$$

and from the monotonicity of the mapping $\partial J$ we conclude that

$$\langle \chi_{k+1}(u^k - u^{k+1}) - \chi_k(u^{k-1} - u^k), u^{k+1} - u^k \rangle \geq 0.$$

Hence,

$$\langle \chi_{k+1}(u^k - u^{k+1}), \chi_{k+1}(u^k - u^{k+1}) \rangle \leq \langle \chi_k(u^{k-1} - u^k), \chi_{k+1}(u^k - u^{k+1}) \rangle$$

and

$$\|\chi_k(u^{k-1} - u^k)\| \geq \|\chi_{k+1}(u^k - u^{k+1})\|.$$

Now, due to the convexity of the functional $J$,

$$\frac{2}{\chi_k} \left( J(u) - J(u^k) \right) \geq 2\langle u^{k-1} - u^k, u - u^k \rangle = \|u^{k-1} - u^k\|^2 + \|u - u^k\|^2 - \|u - u^{k-1}\|^2$$

and, for $u \in K$, it follows

$$\frac{2}{\chi_k} \left( J(u) - J(u^k) \right) \geq \|u^{k-1} - u^k\|^2 + \|u - u^k\|^2 - \|u - u^{k-1}\|^2. \quad (3.1.9)$$

Summing up these inequalities for $k = 1, \ldots, n$, we obtain

$$2\gamma_n J(u) - 2\sum_{k=1}^n \frac{1}{\chi_k} J(u^k) \geq \sum_{k=1}^n \|u^{k-1} - u^k\|^2 + \|u - u^k\|^2 - \|u - u^{k-1}\|^2. \quad (3.1.10)$$
Inequality (3.1.9) with \( u := u^{k-1} \) tells us that
\[
J(u^{k-1}) - f(u^k) \geq \chi_k \|u^{k-1} - u^k\|^2, \quad \forall k \in \mathbb{N},
\]
hence,
\[
\gamma_{k-1} J(u^{k-1}) - \gamma_k J(u^k) + \frac{1}{\chi_k} J(u^k) \geq \gamma_{k-1} \chi_k \|u^{k-1} - u^k\|^2, \quad k = 1, 2, \ldots,
\]
where \( \gamma_0 = 0 \) is taken. This yields
\[
-\gamma_n J(u^n) + \sum_{k=1}^n \frac{1}{\chi_k} f(u^k) \geq \sum_{k=2}^n \gamma_{k-1} \chi_k \|u^{k-1} - u^k\|^2.
\]
The latter inequality together with (3.1.10) leads to
\[
2\gamma_n (J(u) - J(u^n)) \\
\geq \sum_{k=1}^n \|u^{k-1} - u^k\|^2 + \|u - u^n\|^2 - \|u - u^0\|^2 + 2 \sum_{k=2}^n \gamma_{k-1} \chi_k \|u^{k-1} - u^k\|^2
\]
and using the monotonicity of the sequence \( \{\chi_k \|u^k - u^{k-1}\|^2\} \), we get finally
\[
2\gamma_n (J(u) - J(u^n)) \\
\geq \|u - u^n\|^2 - \|u - u^0\|^2 + \left( \sum_{k=1}^n \frac{1}{\chi_k} + 2 \sum_{k=2}^n \frac{\gamma_{k-1}}{\chi_k} \right) \chi_n^2 \|u^{n-1} - u^n\|^2 \\
= \gamma_n^2 \chi_n^2 \|u^{n-1} - u^n\|^2 + \|u - u^n\|^2 - \|u - u^0\|^2,
\]
proving estimate (3.1.8).

3.1.7 **Remark.** Proposition 3.1.6 implies immediately the following estimate
\[
J(u^n) - J(u^0) \leq -\frac{1}{2\gamma_n} \|u^0 - u^n\|^2, \quad \forall n \in \mathbb{N}, \quad u^0 \in K,
\]
and in case \( U^* = \text{Arg min}_{u \in K} J(u) \neq \emptyset \) we get
\[
J(u^n) - \min_{u \in K} J(u) \leq \frac{1}{2\gamma_n} \rho(u^0, U^*), \quad \forall n \in \mathbb{N}, \quad u^0 \in K.
\]

3.1.8 **Proposition.** If \( U^* \neq \emptyset \), then the sequence \( \{u^k\} \), generated by the PPA
3.1.5, converges weakly to some element \( u^* \in U^* \).

This result was established by Martinet [287], which is the first publication dedicated explicitly to proximal-point-methods. It should be mentioned that Proposition 3.1.8 (with \( \chi_k \equiv 2 \)) could have been obtained also as a by-product of Proposition 1.3.11.

As we will see in detail later on, convergence still holds if the prox-point is only approximated with a sufficiently accuracy. This leads to so called *inexact proximal point methods:*
\[
u^{k+1} \approx \text{Prox}_{J,K} u^k.
\]
CHAPTER 3. PPR FOR FINITE-DIMENSIONAL PROBLEMS

Rockafellar [351] gives various conditions for the accuracy of approximation of which the most important in the case of convex minimization is

$$\|u^{k+1} - \text{Prox} \ u^k\| \leq \epsilon_k, \quad \sum_{k=1}^{\infty} \frac{\epsilon_k}{\chi_k} < \infty.$$  

In fact that the regularization parameter $\chi_k$ is not required to tend to zero but only to be confined to an upper bound is an important feature of the PPR in contrast to Tikhonov regularization, possibly slowing down the convergence though but keeping the auxiliary objective functions strongly convex and well-defined.

3.2 PPR for Convex Problems

3.2.1 Proximal gradient methods

In the sequel we make use of the essential fact that the corresponding basic algorithms employed in the proximal point regularization (PPR) are going to approximate successively the original convex minimization problem by means of unconstrained extremal problems. While doing so, the sequence of functions which have to be minimized converges to a function possessing an unconstrained minimizer which is, at the same time, the solution of the original problem.

The finite-dimensional setting of the problems enables us to work with weak requirements on the convergence of the auxiliary functions which have to be minimized converges to a function possessing an unconstrained minimizer which is, at the same time, the solution of the original problem.

3.2.2 Penalty methods with iterative PPR

Let us recall the finite-dimensional Problem (A3.4.56)

\[
\begin{align*}
\text{minimize} & \quad J(u), \\
\text{subject to} & \quad u \in K, \\
K & := \{v \in \mathbb{R}^n : g_j(v) \leq 0, \ j = 1, \ldots, m\},
\end{align*}
\]

where $J, g_j : \mathbb{R}^n \to \mathbb{R}$ are convex, differentiable functions, Slater’s condition is fulfilled and $U^* \neq \emptyset$ and assume additionally that the optimal set $U^*$ is bounded.

Let $\{\phi_k\}, \phi_k \in C^1(\mathbb{R}^n)$, be a system of convex penalty functions, $\{\epsilon_k\}$ and $\{\chi_k\}$ be sequences of positive controlling numbers such that

$$\lim_{k \to \infty} \epsilon_k = 0, \quad 0 < \chi \leq \chi_k < \chi, \quad k = 1, 2, \ldots,$$

and the starting point $u^0 \in \mathbb{R}^n$ be chosen arbitrarily.

In a penalty method, regularized by means of PPR, the iterates are determined via

$$u^k \approx \arg\min_{u \in \mathbb{R}^n} \{J(u) + \phi_k(u) + \frac{\chi_k}{2} \|u - u^{k-1}\|^2\} \quad (3.2.1)$$
such that
\[ \| \nabla J(u^k) + \nabla \phi_k(u^k) + \chi_k(u^k - u^{k-1}) \| \leq \epsilon_k. \tag{3.2.2} \]

To establish convergence for this method we need some technical notions. Let
\[ \tilde{Q} := \{ u \in K : J(u) \leq J^* + \varrho \}, \quad (\varrho > 0 \text{ fixed}) \]
and \( \tilde{u} \in \text{int} \tilde{Q} \). Further, we define
\[ F_k(u) := J(u) + \phi_k(u) + \frac{\chi_k}{2} \| u - u^{k-1} \|^2, \]

\[ v^k := \arg \min_{u \in \mathbb{R}^n} F_k(u), \]

\[ \tau_k := \| w^k - v^k \|, \quad \delta_k := \max_{u \in \tilde{Q}} \phi_k(u), \]

\[ \omega_k := \arg \min_{w \in [\tilde{u}, v^k] \cap \tilde{Q}} \| w - v^k \|^2, \]
i.e., \( w^k \) is a point belonging to the intersection of the interval \([\tilde{u}, v^k]\) and the set \( \tilde{Q} \) and located next to \( v^k \).

From the inequalities (3.2.2), (A1.5.21) and (A1.5.32) it follows for \( k = 1, 2, \ldots \)
\[ \| v^k - u_k \| \leq \frac{\epsilon_k}{\chi_k}, \quad F_k(v^k) - F_k(u) \leq \frac{\epsilon_k^2}{\chi_k} \quad \forall \ u \in \mathbb{R}^n. \]

In the following theorem and in Theorem 3.2.3 the specific structure of the feasible set \( K \) is of no importance.

**3.2.1 Theorem.** Assume that \( \sum_{k=1}^{\infty} \epsilon_k < \infty \) and that the penalty functions \( \phi_k \) satisfy the conditions

1. \( \lim_{k \to \infty} \phi_k(u) = \begin{cases} 0 & \text{if } u \in K, \\ +\infty & \text{if } u \notin K; \end{cases} \)
2. \( \| \nabla \phi_k(z^k) \| \to \infty \) if \( \lim_{k \to \infty} z^k = u \notin K; \)
3. there exists a constant \( c > 0 \) such that for all \( k \in \mathbb{N} \)
\[ \phi_{k+1}(u) \leq c\phi_k(u), \quad \forall \ u \in \mathbb{R}^n. \]

Then it holds

1. the sequence \( \{ u^k \} \), generated by Penalty Method (3.2.1) (3.2.2), is bounded;
2. each of its cluster points belongs to \( K \);
3. \( \lim_{k \to \infty} \| u^k - u^{k-1} \| = 0; \)
4. \( \lim_{k \to \infty} J(u^k) = J^*. \)

If, moreover,

1. \( \sum_{k=1}^{\infty} \sqrt{\delta_k} < \infty, \sum_{k=1}^{\infty} \sqrt{\tau_k} < \infty, \)
then
2. \( \lim_{k \to \infty} u^k = u^* \in U^*. \)
Analogously, if $x^1$, we define for fixed constants $\theta > 0$ and $\delta > 0$ the sets
\[ Q := \{u \in K : J(u) \leq J^* + 2\theta\}, \]
\[ Q_\delta := \{u \in \mathbb{R}^n : \rho(u, Q) \leq \delta\}. \]
Let $Q^\prime \subset \text{int}Q$ be a convex and compact set such that $\text{int}Q^\prime \neq \emptyset$ and
\[ J(u) < J^* + \frac{\theta}{2} \quad \forall u \in Q^\prime. \]

Setting in the proof of Theorem A3.4.41
\[ H := \mathbb{R}^n, \quad G := K, \quad \phi_k^2 \equiv 0, \quad \forall k, \]
we conclude from condition (i) that the inequality
\[ J(q) + \phi_k(q) > J(v) + \phi_k(v) + \theta \quad (3.2.3) \]
is satisfied for sufficiently large $k$ ($k \geq k'$) and all $q \in \partial Q_\delta$, $v \in Q^\prime$.

Now, take an arbitrary $\bar{z} \in Q^\prime$, an index $k_0 > k'$ and define
\[ \rho_0 := \max_{u \in \partial Q_\delta} \| \bar{z} - u \|, \]
\[ \rho := \max \left[ \| \bar{z} - u^{k_0-1} \|, \rho_0 \right] \]
and $B_\rho(\bar{z}) := \{u \in \mathbb{R}^n : \| \bar{z} - u \| \leq \rho \}$.

If $\bar{u} \notin B_\rho(\bar{z})$, with $r := \rho + \rho_0 \frac{\theta}{\chi_k}$, then we chose (see Fig. 3.2.1)
\[ z := \{ \bar{u} + \lambda(\bar{z} - \bar{u}) : 0 \leq \lambda \leq 1 \} \cap \partial Q_\delta, \]
\[ \bar{z} := \{ \bar{u} + \gamma(\bar{z} - \bar{u}) : 0 \leq \gamma \leq 1 \} \cap \partial B_\rho(\bar{z}). \]

Because for a convex function $\varphi$ the inequality
\[ \varphi(x + t(x' - x')) \geq \varphi(x) + t(\varphi(x) - \varphi(x')), \quad t \geq 0, \quad (3.2.4) \]
is true, we get with $\varphi := J + \phi_k, \ x := z, \ x' := \bar{z}, \ t := \frac{\| \bar{z} - z \|}{\| \bar{z} - \bar{z} \|}$ and in view of (3.2.3)
\[ J(\bar{z}) + \phi_k(\bar{z}) = J(z + t(\bar{z} - z)) + \phi_k(z + t(\bar{z} - z)) \]
\[ \geq J(z) + \phi_k(\bar{z}) + \frac{\| \bar{z} - z \|}{\| \bar{z} - \bar{z} \|} (J(z) + \phi_k(z) - J(\bar{z}) - \phi_k(\bar{z})) \]
\[ \geq J(z) + \phi_k(z) + \frac{\rho - \rho_0}{\rho_0} \theta. \]

Analogously, if $x := \bar{z}, \ x' := \bar{z}, \ t := \frac{\| \bar{z} - \bar{u} \|}{\| \bar{z} - \bar{z} \|}$, then in view of (3.2.3) it follows
\[ J(\bar{u}) + \phi_k(\bar{u}) \geq J(\bar{z}) + \phi_k(\bar{z}) + t(J(\bar{z}) + \phi_k(\bar{z}) - J(\bar{z}) - \phi_k(\bar{z})) \]
\[ \geq J(\bar{z}) + \phi_k(\bar{z}) + \frac{\rho_0 \rho}{\theta \chi_k} (J(\bar{z}) + \phi_k(\bar{z}) - J(\bar{z}) - \phi_k(\bar{z}) + \theta) \]
\[ \geq J(\bar{z}) + \phi_k(\bar{z}) + \frac{\rho_0 \rho}{\theta \chi_k} \left( \frac{\rho - \rho_0}{\rho_0} \theta + \theta \right) \]
\[ = J(\bar{z}) + \phi_k(\bar{z}) + \frac{\rho_0^2}{\chi_k}. \]
3.2. PPR FOR CONVEX PROBLEMS

Using the trivial inequality (see Fig 3.2.1)
\[ \| \bar{z} - u_{k_0-1} \| < \| \bar{u} - u_{k_0-1} \|, \]
we can conclude further that
\[ F_{k_0}(\bar{u}) > F_{k_0}(\bar{z}) + \frac{\epsilon^2_{k_0}}{\chi_{k_0}}, \]
and in view of (3.2.2) we get
\[ F_{k_0}(u_{k_0}) \leq F_{k_0}(\bar{z}) + \frac{\epsilon^2_{k_0}}{\chi_{k_0}}. \]

Now it is obvious that \( u_{k_0} \in B_r(\bar{z}) \) for \( r := \rho + \frac{\rho_0}{\beta} \frac{\epsilon^2_{k_0}}{\chi_{k_0}}. \)
Moreover, for \( k \geq k_0 \), the inclusion \( u_k \in B_{r'}(\bar{z}) \) holds with
\[ r' := \rho + \frac{\rho_0}{\beta} \sum_{k \geq k_0} \frac{\epsilon^2_k}{\chi_k}. \]

Hence, due to the assumptions on \( \{ \epsilon_k \} \) and \( \{ \chi_k \} \), the sequence \( \{ u_k \} \) is bounded.
From the relation (3.2.2) it follows that
\[ \nabla \phi_k(u_k) \| \leq \epsilon_k + \chi_k \| u_k - u_{k-1} \| + \| \nabla J(u_k) \|, \]
therefore, \( \limsup_{k \to \infty} \| \nabla \phi_k(u_k) \| < \infty. \) With regard to condition (ii), this proves that \( \lim_{k \to \infty} \rho(u_k, K) = 0. \)

2. Now, let us show (c) and (d). First we prove
\[ \lim_{k \to \infty} \phi_k(u_k) = 0. \quad (3.2.5) \]
Because of the assumptions (i) and (iii) we have $c > 1$ and

$$\phi_k(u) \geq 0 \quad \forall \ u \in \mathbb{R}^n, \ \forall \ k.$$ 

Suppose (3.2.5) is wrong. Then for some $\sigma > 0$ the relation $\phi_k(u^k_i) \geq \sigma$ must be satisfied for a subsequence $\{u^k_i\}$.

In view of the boundedness of $\{u^k\}$ and $\{\chi_k\}$ we can assume without loss of generality that

$$\lim_{i \to \infty} u^k_i = \bar{u}, \quad \lim_{i \to \infty} u^{k-1} = \bar{u}, \quad \lim_{i \to \infty} \chi_k = \chi > 0.$$ 

Taking limit in the relation

$$F_k(\bar{u}) \geq F_k(u^k_i) - \frac{\epsilon^2_k}{\chi_k},$$

for $i \to \infty$, then due to $\bar{u} \in K$ (cf. the first part of the proof), this yields

$$J(\bar{u}) + \frac{\chi}{2} \|\bar{u} - \bar{u}\|^2 \geq J(\bar{u}) + \sigma + \frac{\chi}{2} \|\bar{u} - \bar{u}\|^2.$$ 

But this is impossible. Hence $\lim_{k \to \infty} \phi_k(u^k) = 0$ is true and, in view of the non-negativity of $\phi_k$ and condition (iii), the relation $\lim_{k \to \infty} \phi_{k+1}(u^k) = 0$ holds, too.

Next, we prove that $\limsup_{k \to \infty} J(u^k) = J^*$. Suppose this is wrong. Then

$$\limsup_{k \to \infty} J(u^k) - J^* = \theta_0 > 0.$$ 

Choosing $\tau := \frac{\theta_0}{\alpha}$, $\epsilon > 0$, $\epsilon << \tau$, then there exists a number $k_0$ such that the inequality

$$J(u^{k_0-1}) \geq J^* + \tau$$

is satisfied and

$$\phi_k(u^{k-1}) < \epsilon, \quad \phi_k(u^*) < \epsilon, \quad \frac{\epsilon^2}{\chi_k} < \epsilon$$

for all $k \geq k_0$ and some $u^* \in U^*$.

For $u^k_\alpha := \alpha u^* + (1 - \alpha)u^{k-1}, \ \alpha \in [0, 1]$, we get the inequality

$$J(u^k_\alpha) + \phi_k(u^k_\alpha) + \frac{\chi_k}{2} \|u^k_\alpha - u^{k-1}\|^2$$

$$< \alpha(J(u^*) - J(u^{k-1})) + \frac{\alpha^2 \chi_k}{2} \|u^* - u^{k-1}\|^2 + J(u^{k-1}) + \epsilon =: \eta_k(\alpha).$$

It is easy to show that the minimum of $\eta_k(\cdot)$ on the interval $[0, 1]$ is attained at the point

$$\alpha_k = \min \left\{ \frac{J(u^{k-1}) - J(u^*)}{\chi_k \|u^* - u^{k-1}\|^2}, 1 \right\},$$

hence,

$$\eta_k(\alpha_k) = J(u^{k-1}) - \frac{(J(u^{k-1}) - J(u^*))^2}{2 \chi_k \|u^* - u^{k-1}\|^2} + \epsilon \quad \text{if } \alpha_k < 1,$$
and
\[ \eta_k(\alpha_k) \leq J(u^{k-1}) - \frac{1}{2}(J(u^{k-1}) - J(u^*)) + \epsilon \quad \text{if } \alpha_k = 1. \]
Thus, we get
\[ \eta_k(\alpha_k) \leq \begin{cases} J(u^{k-1}) - \frac{\tau^2}{2} + \epsilon & \text{if } \alpha_k < 1, \\ J(u^{k-1}) - \frac{\tau^2}{2} + \epsilon & \text{if } \alpha_k = 1, \end{cases} \tag{3.2.7} \]
with \( \varsigma = \sup_{k \geq k_0} 2\chi_k \|u^* - u^{k-1}\|^2 < \infty. \)
Regarding the definition of the sequence \( \{u^k\} \), for any \( \alpha \in [0, 1] \) the inequality
\[ J(u^k) + \phi_k(u^k) + \frac{\chi_k}{2}\|u^k - u^{k-1}\|^2 \leq J(u^k_\alpha) + \phi_k(u^k_\alpha) + \frac{\chi_k}{2}\|u^k_\alpha - u^{k-1}\|^2 + \frac{\epsilon_k^2}{\chi_k} \]
holds true, and by means of (3.2.7) and the non-negativity of \( \phi_k \) we get
\[ J(u^k) < \max \left[ J(u^{k-1}) - \frac{\tau^2}{\varsigma}, J(u^{k-1}) - \frac{\tau^2}{2} + \epsilon + \frac{\epsilon_k^2}{\chi_k} \right]. \tag{3.2.8} \]
The latter inequality is also true for \( k_1 > k \) if only \( J(u^{k_1-1}) \geq J^* + \tau. \)
Assuming that \( \epsilon \ll \frac{\tau^2}{\varsigma} \), we conclude from inequality (3.2.8) that after a finite number of steps a point \( u^{k'} \) \((k' > k)\) can be obtained for which
\[ J(u^{k'}) < J^* + \tau. \tag{3.2.9} \]
But from
\[ F_k(u^k) \leq F_k(u^{k-1}) + \frac{\epsilon_k^2}{\chi_k} \tag{3.2.10} \]
which is true for any \( k \in \mathbb{N} \), it follows in view of (3.2.6) that
\[ J(u^{k'+1}) \leq J(u^{k'}) + \epsilon + \frac{\epsilon_{k'+1}^2}{\chi_{k'+1}}, \]
i.e., due to (3.2.9) and \( \epsilon + \frac{\epsilon_{k'+1}^2}{\chi_{k'+1}} < \tau \), we have
\[ J(u^{k'+1}) < J^* + 2\tau. \tag{3.2.11} \]
Now, if \( J(u^{k'+1}) \geq J^* + \tau \), then using (3.2.8) with \( k := k' + 2 \) as well as (3.2.11), again we obtain
\[ J(u^{k'+2}) < J^* + 2\tau. \]
Hence,
\[ J(u^k) < J^* + 2\tau, \quad \forall k \geq k' + 1, \]
which contradicts the assumption
\[ \limsup_{k \to \infty} J(u^k) - J^* = \theta_0 = 3\tau, \]
proving that \( \limsup_{k \to \infty} J(u^k) = J^*. \)
On the other hand, due to condition (i), we have \( \liminf_{k \to \infty} J(u^k) \geq J^* \), hence, \( \lim_{k \to \infty} J(u^k) = J^* \) and the statement
\[ \|u^k - u^{k-1}\| = 0 \]
follows immediately from (3.2.10).

3. Finally we prove (e). Reminding of the notions defined before Theorem 3.2.1 and setting

\[ y^k := \arg\min_{u \in Q} \{J(u) + \frac{\chi_k}{2} \|u - u^{k-1}\|^2\}, \]

we obtain in view of the choice of \(\{v^k\}\), \(\{w^k\}\) and \(\{\delta_k\}\),

\[ F_k(u^k) \leq F_k(y^k) \leq J(y^k) + \frac{\chi_k}{2} \|y^k - u^{k-1}\|^2 + \delta_k, \]

\[ J(y^k) + \frac{\chi_k}{2} \|y^k - u^{k-1}\|^2 \leq J(w^k) + \frac{\chi_k}{2} \|w^k - u^{k-1}\|^2. \]

The last two inequalities, together with \(\phi_k(v^k) \geq 0\), lead to

\[ J(u^k) + \frac{\chi_k}{2} \|u^k - u^{k-1}\|^2 - \delta_k \leq J(y^k) + \frac{\chi_k}{2} \|y^k - u^{k-1}\|^2 \]

\[ \leq J(w^k) + \frac{\chi_k}{2} \|w^k - u^{k-1}\|^2. \] (3.2.12)

In the first and second part of the proof it was shown that the sequences \(\{v^k\}\) and \(\{w^k\}\) are bounded and that their cluster points belong to \(U^*\). Boundedness of the set \(\tilde{Q}\) and consequently also of the sequence \(\{v^k\}\) follows from Proposition A1.7.55. Therefore, for \(k = 1, 2, \cdots\), the inequality

\[ |J(u^k) + \frac{\chi_k}{2} \|u^k - u^{k-1}\|^2 - J(v^k) - \frac{\chi_k}{2} \|v^k - u^{k-1}\|^2| \leq c_1 \|w^k - v^k\| \] (3.2.13)

can be established. From (3.2.12) and (3.2.13) it follows that

\[ J(u^k) + \frac{\chi_k}{2} \|u^k - u^{k-1}\|^2 - J(y^k) - \frac{\chi_k}{2} \|y^k - u^{k-1}\|^2 \leq c_1 \tau_k + \delta_k, \]

and with regard to the strong convexity of \(\Psi_k(\cdot) = J(\cdot) + \frac{\chi_k}{2} \cdot \|u^{k-1}\|^2\) and the choice of \(y^k\) and \(w^k\), one can conclude that

\[ \|u^k - y^k\| \leq \sqrt{\frac{2}{\chi_k} (c_1 \tau_k + \delta_k)}. \]

Non-expansivity of the prox-mapping leads to

\[ \|y^k - \bar{u}\| \leq \|u^{k-1} - \bar{u}\| \] for any \(\bar{u} \in U^*\),

and using the relations \(\|u^k - v^k\| \leq \frac{\epsilon_k}{\chi_k}\) and

\[ \|u^k - \bar{u}\| \leq \|u^k - v^k\| + \|v^k - w^k\| + \|u^k - w^k\| + \|y^k - \bar{u}\|, \]

we obtain

\[ \|u^k - \bar{u}\| \leq \frac{\epsilon_k}{\chi_k} + \tau_k + \sqrt{\frac{2}{\chi_k} (c_1 \tau_k + \delta_k)} + \|u^{k-1} - \bar{u}\|. \]

Now, the assumptions of the theorem ensure the convergence of the series

\[ \sum_{k=1}^{\infty} \left( \frac{\epsilon_k}{\chi_k} + \tau_k + \sqrt{\frac{2}{\chi_k} (c_1 \tau_k + \delta_k)} \right), \]

and due to Lemma A3.1.4 the sequence \(\{\|u^k - \bar{u}\|\}\) converges for any \(\bar{u} \in U^*\). Taking into account that every cluster point of \(\{u^k\}\) belongs to \(U^*\), one can easily seen that \(\{u^k\}\) converges to a point of the optimal set \(U^*\). □
3.2.2 Remark. Let us analyze the condition (iv) in Theorem 3.2.1. Suppose the feasible set \( K \) is given as in Problem (A3.4.56). For sufficiently large \( k \) we conclude from the second part of the proof that \( J(v^k) < J^* + \varrho \), hence, \( w^k = v^k \) (and \( \tau_k = 0 \)) or \( \bar{g}(w^k) = 0 \), with \( \bar{g}(\cdot) = \max_{1 \leq j \leq m} g_j(\cdot) \). In the latter case, from (3.2.4) with

\[
\varphi := \bar{g}, \quad x' := \bar{u}, \quad x := w^k, \quad t := \frac{\|v^k - w^k\|}{\|w^k - \bar{u}\|},
\]

it follows

\[
\bar{g}(v^k) \geq \bar{g}(w^k) + t(\bar{g}(w^k) - \bar{g}(\bar{u})),
\]

\[
\bar{g}(v^k) \geq |\bar{g}(\bar{u})|\frac{\|v^k - w^k\|}{\|w^k - \bar{u}\|} \geq c_2\|w^k - v^k\|,
\]

(3.2.14)

with \( c_2 := |\bar{g}(\bar{u})|\frac{1}{\max_{z \in \partial \tilde{Q}} \|z - \bar{u}\|} \).

For the majority of penalty functions satisfying the conditions (i)-(iii) of Theorem 3.2.1, the inequalities

\[
\frac{\partial \phi_k(u^k)}{\partial g_j} \leq \bar{c}, \quad \text{(hence, } \frac{\partial \phi_k(v^k)}{\partial g_j} \leq \bar{c}), \quad j = 1, \ldots, m, \quad k = 1, 2, \ldots, (3.2.15)
\]

hold (see Proposition 3.2.5 below).

If, for instance,

\[
\phi_k(u) := r_k \sum_{j=1}^{m} \{\max[0, g_j(u)]\}^2, \quad \lim_{k \to \infty} r_k = +\infty,
\]

due to (3.2.15), we conclude immediately that

\[
\max[0, g_j(v^k)] \leq \frac{\bar{c}}{2r_k}.
\]

This together with (3.2.14) gives

\[
\tau_k := \|v^k - w^k\| \leq \frac{\bar{c}}{2c_2r_k}
\]

if \( v^k \neq w^k \) and \( k \) is large enough.

Therefore,

\[
\sum_{k=1}^{\infty} \frac{1}{\sqrt{\tau_k}} < \infty \quad \Rightarrow \quad \sum_{k=1}^{\infty} \sqrt{\tau_k} < \infty.
\]

The condition \( \sum_{k=1}^{\infty} \sqrt{\delta_k} < \infty \) is trivially satisfies because in that case we have \( \delta_k = 0 \) \( \forall k \).

For other penalty functions some alterations are necessary, which are connected with the estimation of \( \bar{g}(v^k) \) on the basis of inequality (3.2.15).

We emphasize that generally, for any penalty function satisfying the conditions (i)-(iii) in Theorem 3.2.1, condition (iv) can be fulfilled by a suitable choice of the penalty parameters.

In order to apply barrier instead of penalty functions, we consider convex, lsc functions \( \phi_k : \mathbb{R}^n \to \mathbb{R} \) with a common effective domain \( D \supset \text{int} K \) and
\( \phi_k \in C^1(D), \forall k \in \mathbb{N} \).

To establish a similar convergence theorem for regularized barrier methods, we choose a sequence \( \{\sigma_k\}, \sigma_k > 0 \forall k \), and define the sets

\[
U_k := \{ u \in \tilde{Q} : \phi_k(u) \leq \sigma_k \}
\]

and the values

\[
\zeta_k := \min_{u \in U_k} \{ J(u) + \frac{\lambda_k}{2} \| u - u^{k-1} \|^2 \} - \min_{u \in \tilde{Q}} \{ J(u) + \frac{\lambda_k}{2} \| u - u^{k-1} \|^2 \}. \tag{3.2.16}
\]

Again we use the notions defined before Theorem 3.2.1.

Now, under previously made assumptions on \( \{\chi_k\} \) and \( \{\epsilon_k\} \) let us consider the convergence of a regularized barrier method (3.2.1) (3.2.2) starting with \( u^0 \in D \).

3.2.3 Theorem. Let \( \epsilon_k \) be chosen such that \( \sum_{k=1}^{\infty} \epsilon_k < \infty \). Assume that the barrier functions \( \phi_k \) have the properties mentioned before and the following conditions are fulfilled:

(i') \( \lim_{k \to \infty} \phi_k(u) = \begin{cases} 0 & \text{if } u \in \text{int } K, \\ +\infty & \text{if } u \notin K; \end{cases} \)

(ii') \( \liminf_{u \in \partial K \cap D} \| \nabla \phi_k(u) \| = \infty \) (or \( \partial K \cap D = \emptyset \));

(iii') for all \( k \in \mathbb{N} \) it holds

\[
\phi_{k+1}(u) \leq \phi_k(u), \quad \forall u \in K \cap D.
\]

Then

(a) the sequence \( \{u^k\} \), generated by Barrier Method (3.2.1) (3.2.2), is bounded;

(b) \( u^k \in \text{int } K \) for sufficiently large \( k \);

(c) \( \lim_{k \to \infty} \| u^k - u^{k-1} \| = 0; \)

(d) \( \lim_{k \to \infty} J(u^k) = J^*; \)

If, moreover,

\( \sum_{k=1}^{\infty} \sqrt{\sigma_k} < \infty, \sum_{k=1}^{\infty} \sqrt{\zeta_k} < \infty, \)

then

(e) \( \lim_{k \to \infty} u^k = u^* \in U^*; \)

Proof: The proof of the statements (a)-(d) is almost identical to that of Theorem 7.1 in GROSSMANN and KAPLAN [145]. In particular, the existence of a number \( k' \) is guaranteed such that for all \( k \geq k' - 1 \)

\[
u^k \in \text{int } \tilde{Q} \quad \text{and} \quad v^k := \arg \min_{u \in \mathbb{R}^n} F_k(u) \in \text{int } \tilde{Q},\]

We prove now the statement (e). Let \( k \geq k' \). Setting again

\[
y^k := \arg \min_{u \in \tilde{Q}} \{ J(u) + \frac{\lambda_k}{2} \| u - u^{k-1} \|^2 \}, \tag{3.2.17}
\]
in view of the obvious inequality
\[ J(y^k) + \frac{\lambda_k}{2} \| y^k - u^{k-1} \|^2 \leq J(v^k) + \frac{\lambda_k}{2} \| v^k - u^{k-1} \|^2 \]
and the choice of \( \zeta_k \), one can conclude that
\[ \min_{u \in U_k} \{ J(u) + \frac{\lambda_k}{2} \| u - u^{k-1} \|^2 \} \leq J(v^k) + \frac{\lambda_k}{2} \| v^k - u^{k-1} \|^2 + \zeta_k. \tag{3.2.18} \]

But due to the conditions (i') and (iii'), the functions \( \phi_k \) are non-negative on \( K \) and, with regard to (3.2.18) and the definition of \( U_k \) and \( \zeta_k \), we obtain for \( x^k := \arg \min \{ F_k(u) : u \in U_k \} \)
\[
F_k(x^k) \leq J(v^k) + \frac{\lambda_k}{2} \| v^k - u^{k-1} \|^2 + \sigma_k + \zeta_k \leq F_k(v^k) + \sigma_k + \zeta_k,
\]
\[
J(x^k) + \frac{\lambda_k}{2} \| x^k - u^{k-1} \|^2 \leq J(y^k) + \frac{\lambda_k}{2} \| y^k - u^{k-1} \|^2 + \sigma_k + \zeta_k.
\]

Now, due to the strong convexity of the functions \( \Psi_k(\cdot) = J(\cdot) + \frac{\lambda_k}{2} \| \cdot - u^{k-1} \|^2 \) and \( F_k \) as well as the choice of \( v^k \) and \( y^k \), the inequalities
\[
\| x^k - v^k \| \leq \sqrt{\frac{2}{\lambda_k} (\sigma_k + \zeta_k)}, \quad \| x^k - y^k \| \leq \sqrt{\frac{2}{\lambda_k} (\sigma_k + \zeta_k)}
\]
hold for \( k \geq k' \). To establish \( u^k \to u^* \in U^* \) we proceed as in the proof of Theorem 3.2.1.

\[ \square \]

3.2.4 Remark. Let us analyze condition (iv') in Theorem 3.2.3 for barrier functions of type
\[
\phi_k(u) = \sum_{j=1}^{m} \exp(r_k g_j(u)), \quad r_{k+1} \geq r_k, \quad \lim_{k \to \infty} r_k = \infty.
\]

For these functions the validity of the conditions (i')-(iii') is evident. If \( \sigma_k > 0 \) are chosen such that \( \sum_{k=1}^{\infty} \sigma_k < \infty \) and \( r_k \) satisfies the condition
\[
\sum_{k=1}^{\infty} \sqrt{\frac{\ln \sigma_k}{r_k}} < \infty,
\]
then one can show that \( \sum_{k=1}^{\infty} \sqrt{\sigma_k} < \infty \). Indeed, suppose that \( u \in \tilde{Q} \) fulfills the inequality
\[
\tilde{g}(u) \leq \frac{\ln(\sigma_k)}{r_k} \tag{3.2.19}
\]
with \( \tilde{g}(\cdot) := \max_{1 \leq j \leq m} g_j(\cdot) \). Then, obviously, \( u \) belongs to \( U_k \). We can assume that \( \sigma_k < 1 \) and \( r_k \) are chosen such that for some fixed \( \tilde{u} \in \text{int} \tilde{Q} \) it holds
\[
\tilde{g}(\tilde{u}) \leq \frac{\ln(\sigma_k)}{r_k}, \quad \forall k.
\]

Then, for arbitrary \( z \in \tilde{Q} \),
\[
\tilde{g}(\lambda \tilde{u} + (1 - \lambda)z) \leq \lambda \tilde{g}(\tilde{u}), \quad \forall \lambda \in [0, 1].
\]
CHAPTER 3. PPR FOR FINITE-DIMENSIONAL PROBLEMS

Setting
\[ \lambda := \frac{\ln \left( \frac{c_k}{r_k} \right)}{r_k g(\tilde{u})}, \]
we get
\[ g(\lambda \tilde{u} + (1 - \lambda)z) \leq \frac{\ln \left( \frac{c_k}{r_k} \right)}{r_k}, \]
proving the inclusion \( \lambda \tilde{u} + (1 - \lambda)z \in U_k \) for all \( \lambda \in [0, 1] \). Hence,
\[ \min_{u \in U_k} \| z - u \| \leq \| z - (\lambda \tilde{u} + (1 - \lambda)z) \| = \lambda \| z - \tilde{u} \| \]
\[ \leq \lambda \max_{u \in \partial Q} \| u - \tilde{u} \| = \frac{\ln \left( \frac{c_k}{r_k} \right)}{r_k g(\tilde{u})} \max_{u \in \partial Q} \| u - \tilde{u} \|. \]

For \( z_k := \arg \min_{v \in U_k} \| v - y_k \| \), with \( y_k \) via (3.2.17), we get
\[ \| z_k - y_k \| \leq c \left( \frac{\ln \left( \frac{c_k}{r_k} \right)}{r_k} \right), \quad \forall \, k. \quad (3.2.20) \]

But
\[ \zeta_k \leq J(z_k) + \frac{\chi_k}{2} \| z_k - u_k \|^2 - J(y_k) - \frac{\chi_k}{2} \| y_k - u_k \|^2 \]
\[ = J(z_k) - J(y_k) + \frac{\chi_k}{2} \| y_k - z_k \|^2 \| z_k + y_k - 2u_k \| \quad (3.2.21) \]
and because the sequences \( \{ u_k \}, \{ z_k \}, \{ y_k \} \) and \( \{ \chi_k \} \) are bounded, the required result follows immediately from (3.2.20) and (3.2.21).

For this type of regularized penalty methods a statement analogously to Proposition A3.4.43 is true.

3.2.5 Proposition. Assume that in Method (3.2.1), (3.2.2) applied to Problem (A3.4.56) the convex penalty functions \( \phi_k : \mathbb{R}^n \rightarrow \mathbb{R} \) have the form
\[ \phi_k := \varphi_k(g(\cdot)), \]
where \( g := (g_1, ..., g_m) \) and \( \varphi_k : \mathbb{R}^m \rightarrow \mathbb{R}, \varphi \in C^1(\text{dom} \varphi_k), \) and that either the conditions (i)-(iii) of Theorem 3.2.1 or (i')-(iii') of Theorem 3.2.3 are fulfilled. Moreover, let the functions
\[ \eta^j_k(z) := \frac{\partial \varphi_k(g(\cdot))}{\partial g_j} \]
be non-negative on \( \mathbb{R}^n \) and for any sequence \( \{ z^k \} \) the following implication holds
\[ \lim_{k \rightarrow \infty} \inf g_j(z^k) > -\infty, \quad \lim_{k \rightarrow \infty} \sup g_j(z^k) < 0 \quad \Rightarrow \quad \lim_{k \rightarrow \infty} \eta^j_k(z^k) = 0. \]

Then
(i) the sequence \( \{ (u^k, \lambda^k) \} \), with \( \lambda^k = (\lambda^k_1, ..., \lambda^k_m) \), \( \lambda^k := \eta^j_k(u^k) \), is bounded,
(ii) each cluster point of the sequence \( \{ \lambda^k \} \) is a solution of the dual problem,
3.2. PPR FOR CONVEX PROBLEMS

(iii) in case $u^k \in K$ the estimate

$$J(u^k) \geq J^* \geq J(u^k) + \langle \lambda^k, g(u^k) \rangle - (\epsilon_k + \chi_k \|u^k - u^{k-1}\|) \rho(u^k, U^*)$$ (3.2.22)

holds true.

Bipartite estimates of type (3.2.22) can be obtained also for infeasible iterates $u^k \notin K$ (cf. (A3.4.83)). In case that upper bounds for $\rho(u^k, U^*)$ are known, inequality (3.2.22) can be used as an efficient stopping rule.

Proposition 3.2.5 immediately follows from Proposition A3.4.43 if $\varphi_k \in C^1(\mathbb{R}^m)$ is chosen. Indeed, in view of inequality (3.2.2) we see that

$$\|\nabla J(u^k) + \nabla \varphi_k(u^k)\| \leq \epsilon_k + \chi_k \|u^k - u^{k-1}\|.$$ 

Remind that the relation $\|u^k - u^{k-1}\| \to 0$ was proved in the Theorems 3.2.1 and 3.2.3 without the usage of the conditions (iv), or (iv'), respectively. Thus, setting

$$\epsilon'_k = \epsilon_k + \chi_k \|u^k - u^{k-1}\|,$$

we obtain $\epsilon'_k \downarrow 0$ and thus, the sequence $\{u^k\}$ satisfies the stopping rule in Proposition A3.4.43.

For convex infinite-dimensional problems Penalty Method (3.2.1) (3.2.2) will be studied in the Sections 6.2 and 6.3 in a general framework choosing penalties of type

$$\phi_k(u) := \frac{1}{2} \sum_{j=1}^{m} \lambda_j^k \left( g_j(u) + \sqrt{g_j^2(u) + \frac{1}{r_k}} \right)$$

with $\lim_{k \to \infty} r_k = +\infty$, $\lim_{k \to \infty} \lambda_j^k = \lambda_j > \lambda_j^*$, $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ a Lagrange multiplier vector.

Moreover, in Section 5.3 estimates of the rate of convergence for Method (3.2.1) (3.2.2) with barrier functions of type (A3.4.78) will be established.

In the remainder of this subsection we present some results obtained by ALART and LEMAIRE [5]. They proposed regularized penalty methods with non-smooth penalty functions which are applicable to more general problems than Problem (A3.4.56). For instance, it might happen that the objective function cannot be extended outside of its effective domain by a finite convex function which is defined on the whole space. In [5] the problem

$$\min \{J(u) : u \in K\},$$

is considered with $K := \{v \in \mathbb{R}^n : g(v) \leq 0\}$ and $J : \mathbb{R}^n \to \overline{\mathbb{R}}$ convex and lower semi-continuous. The function $g$ is supposed to be convex and finite-valued. Moreover, it is assumed that Slater's condition holds and $U^* \neq \emptyset$.

Let $M > \sup_{u \in K} J(u)$. Introducing the functions

$$\tilde{J}(u) := \begin{cases} J(u) & \text{if } u \in K, \\ M & \text{if } u \notin K, \end{cases}$$

and

$$\tilde{J}_k(u) := \max \{\tilde{J}(u), r_k g(u) + M\}, \quad r_k > 0, \quad \lim_{k \to \infty} r_k = +\infty,$$
the iterates $u^k$ are determined by

$$J_k(u^k) + \frac{\chi_k}{2} u^k - u^{k-1} \leq \min_{u \in \mathbb{R}^n} \left( J_k(u) + \frac{\chi_k}{2} u - u^{k-1} \right) + \epsilon_k$$

(3.2.23)

with $\{\chi_k\}, \{\epsilon_k\}$ chosen such that

$$0 < \chi \leq \chi_k \leq \chi < \infty, \epsilon_k \downarrow 0.$$

In [5] it is shown that the functions $\bar{J}_k$ are finite-valued, convex and

$$\lim_{k \to \infty} \bar{J}_k(u) = \begin{cases} J(u) & \text{if } u \in \text{int } K, \\ M & \text{if } u \in \partial K, \\ +\infty & \text{if } u \notin K. \end{cases}$$

3.2.6 Proposition. (cf. [5])

Assume that the set $K$ is bounded, $r_k \leq r_{k+1} \forall k$ and that

$$\sum_{k=1}^{\infty} \sqrt{\epsilon_k} < \infty, \quad \sum_{k=1}^{\infty} \frac{1}{\sqrt{r_k}} < \infty.$$

Then the sequence $\{u^k\}$, defined by Method (3.2.23), converges to some point $u^* \in U^*$ and $\bar{J}_k(u^k) \to J^*$.

This statement was proved by means of the theory of variational convergence, developed by Attouch [16], Attouch and Wets [18] and, concerning proximal-point methods, by Lemaitre [260].

3.2.7 Remark. Proposition 3.2.6 can be derived also from Theorems A3.4.41 and 3.2.3 taking into account the following facts:

Theorem A3.4.41 remains true if, instead of (A3.4.70) and the convexity of $\phi_k$, we assume that $\text{int } K \not= \emptyset$ (i.e. $H = \mathbb{R}^n$ and $\phi_k := \phi_k^1$), that $J + \phi_k$ is convex and

$$\lim_{k \to \infty} (J(u) + \phi_k(u)) = \begin{cases} J(u) & \text{if } u \in \text{int } K, \\ +\infty & \text{if } u \notin K. \end{cases}$$

Theorem 3.2.3 remains valid if, instead of (ii') and the convexity of $\phi_k$, we suppose that $J + \phi_k$ is convex and $\phi_k(u) \geq c > 0$ for all $u \in \partial K$ and all $k$.

In these cases we can also drop the assumption about the smoothness of $J, g$ and $\phi_k$ and replace criterion (3.2.2) by

$$F_k(u^k) - \min_{u \in \mathbb{R}^n} F_k(u) \leq \frac{\epsilon_k^2}{\chi_k}.$$

The alterations required in the proofs of these theorems are insignificant. In order to verify condition (iv') in Theorem 3.2.3, we note that

$$U_k = \{ u \in \tilde{Q} : \max[J(u), r_k g(u) + M] - J(u) \leq \sigma_k \}$$

$$\sup \{ u \in \tilde{Q} : \max[0, r_k g(u) - J^* + M] \leq \sigma_k \}.$$

For arbitrary $z \in \partial \tilde{Q}$ and $\lambda \in [0, 1]$ it follows

$$g(\lambda \tilde{u} + (1 - \lambda)z) \leq \lambda g(\tilde{u}), \quad (\tilde{u} \in \text{int } \tilde{Q} \text{ is fixed}),$$
and hence,
\[
\lambda \tilde{u} + (1 - \lambda)z \in U_k \quad \text{if} \quad r_k \lambda g(\tilde{u}) - J^* + M \leq \sigma_k, \; \forall \; \lambda \in [0, 1].
\]

In particular, if \( k \) is large enough, this inclusion holds for
\[
\lambda = \frac{M - J^* - \sigma_k}{-g(\tilde{u})r_k}.
\]

Thus, similarly to the deduction of (3.2.20), (3.2.21), we obtain \( \sum_{k=1}^{\infty} \sqrt{\sigma_k} < \infty \) if we assume that \( \sum_{k=1}^{\infty} \sqrt{r_k} < \infty \) holds.

\[\diamondsuit\]

Sometimes we take the regularizing parameter \( \chi_k := 2 \forall k \) in order to simplify the description of the PPR. Some practical recommendations with respect to the choice of \( \chi_k \) will be given in this book in particular applications for specific classes of problems as well as in Chapter 7 (see also Rockafellar [352] and Güler [153]).

3.2.3 Augmented Lagrangian methods with iterative PPR

Here the PPR suggested by Rockafellar [351] and Antipin [13] for solving problems of type (A3.4.56) will be described by means of the augmented Lagrangian function (A3.4.85)
\[
L_A(u, \lambda) := J(u) + \frac{1}{2r} \left\{ \| \lambda + rg(u) \|_+^2 - \| \lambda \|_+^2 \right\}, \quad (r > 0).
\]

The establish this method we mainly follow the description in [13].

3.2.8 Algorithm. (Regularized augmented Lagrangian method)

Data: \( u^0 \in K, \lambda^0 \in \mathbb{R}_+^m, \{ \epsilon_k \} \downarrow 0; \)

S0: Set \( k := 0; \)

S1: if \( (u^k, \lambda^k) \) is a saddle point of Problem (A3.4.56), stop;

S2: compute \( u^{k+1} \) such that
\[
\| \nabla F_k(u^{k+1}) \| \leq \epsilon_k, \tag{3.2.24}
\]

with
\[
F_k(u) := L_A(u, \lambda^{k-1}) + \| u - u^k \|^2;
\]

S3: set
\[
\lambda^{k+1} := [\lambda^k + rg(u^{k+1})]_+; \tag{3.2.25}
\]

set \( k := k + 1 \) and go to S1.

\[\diamondsuit\]

Now we are going to study the convergence of Algorithm 3.2.8 applied to Problem (A3.4.56).
\textbf{3.2.9 Theorem.} Assume that the mapping \( u \mapsto g(u) \) on \( \mathbb{R}^n \) is Lipschitz-
continuous with constant \( L \) and the sequence \( \{\epsilon_k\} \) of positive numbers satisfy

\[
\lim_{k \to \infty} \epsilon_k = 0, \quad \sum_{k=1}^{\infty} \epsilon_k < \infty.
\]

Then

\[
\lim_{k \to \infty} u^k = u^* \in U^*, \quad \lim_{k \to \infty} \lambda^k = \lambda^* \in \Lambda^*,
\]

where \( \Lambda^* \) denotes the set of optimal Lagrange multiplier vectors.

\textbf{Proof:} For the pair \((u^{k-1}, \lambda^{k-1})\), computed by Algorithm \textbf{3.2.8}, let \((\bar{u}^k, \bar{\lambda}^k)\) be

the exact calculated iterates, i.e.,

\[
\bar{u}^k = \arg\min_{u \in \mathbb{R}^n} \mathcal{F}_k(u), \quad (3.2.26)
\]

\[
\bar{\lambda}^k = [\lambda^{k-1} + rg(\bar{u}^k)]_+, \quad (3.2.27)
\]

In view of the strong convexity of \( \mathcal{F}_k \) and (3.2.24)-(3.2.27), we obtain

\[
\|\bar{u}^k - u^k\| \leq \frac{\epsilon_k}{2}, \quad (3.2.28)
\]

and

\[
\|\bar{\lambda}^k - \lambda^k\| = \|\lambda^{k-1} + rg(\bar{u}^k)\| \leq \|\lambda^{k-1} + rg(u^k)\| \leq \frac{rL\epsilon_k}{2} =: c_1\epsilon_k. \quad (3.2.29)
\]

Next we estimate the distance of \((\bar{u}^k, \bar{\lambda}^k)\) to an arbitrary but fixed point \((\hat{u}, \hat{\lambda})\) \( \in \)

\( U^* \times \Lambda^* \).

In view of (3.2.26) and Proposition \textbf{A1.5.34} it follows that

\[
2\langle \bar{u}^k - u^{k-1}, u - \bar{u}^k \rangle + (\nabla g(\bar{u}^k))^T [\lambda^{k-1} + rg(\bar{u}^k)]_+, u - \bar{u}^k \rangle + J(u) - J(\bar{u}^k) \geq 0, \quad \forall \ u \in \mathbb{R}^n, \quad (3.2.31)
\]

and

\[
(\nabla g(\bar{u})^T [\hat{\lambda} + rg(\hat{u})]_+, u - \hat{u}) + J(u) - J(\hat{u}) \geq 0, \quad \forall \ u \in \mathbb{R}^n. \quad (3.2.32)
\]

Setting \( u := \hat{u} \) in (3.2.31) and \( u := \bar{u}^k \) in (3.2.32) and combining these inequalities, we obtain

\[
2\langle \bar{u}^k - u^{k-1}, \hat{u} - \bar{u}^k \rangle + (\nabla g(\bar{u}^k))^T [\lambda^{k-1} + rg(\bar{u}^k)]_+ - \nabla g(\bar{u})^T [\hat{\lambda} + rg(\hat{u})]_+, \hat{u} - \bar{u}^k \rangle \geq 0,
\]

and the convexity of \( g \) yields

\[
2\langle \bar{u}^k - u^{k-1}, \hat{u} - \bar{u}^k \rangle + ([\lambda^{k-1} + rg(\bar{u}^k)]_+ - [\hat{\lambda} + \hat{g}(\hat{u})], g(\hat{u}) - g(\bar{u}^k)) \geq 0. \quad (3.2.33)
\]

Now, using the obvious relations

\[
\langle [x]_+, x \rangle = \langle [x]_+, [x]_+ \rangle, \quad \langle [x]_+, y \rangle \leq \langle [x]_+, [y]_+ \rangle,
\]

\[\text{for } y \in \mathbb{R}^n.\]
we conclude that

\[ \langle |\lambda^{k-1} + rg(\hat{u}^k)|_+ - [\hat{\lambda} + rg(\hat{u})]_+, g(\hat{u}^k) - g(\hat{u}) \rangle \]

\[ = \frac{1}{r} \langle |\lambda^{k-1} + rg(\hat{u}^k)|_+ - [\hat{\lambda} + rg(\hat{u})]_+, \]

\[ (\lambda^{k-1} + rg(\hat{u}^k) - \lambda^{k-1}) - (\hat{\lambda} + rg(\hat{u}) - \hat{\lambda}) \rangle \]

\[ \geq \frac{1}{r} \langle |\lambda^{k-1} + rg(\hat{u}^k)|_+ - [\hat{\lambda} + rg(\hat{u})]_+, \]

\[ [\lambda^{k-1} + rg(\hat{u}^k)]_+ - \lambda^{k-1} - ([\hat{\lambda} + rg(\hat{u})]_+ - \hat{\lambda}) \rangle. \]

Hence,

\[ \langle |\lambda^{k-1} + rg(\hat{u}^k)|_+ - [\hat{\lambda} + rg(\hat{u})]_+, g(\hat{u}^k) - g(\hat{u}) \rangle \]

\[ \geq \frac{1}{r} \| ([\lambda^{k-1} + rg(\hat{u}^k)]_+ - \lambda^{k-1}) - ([\hat{\lambda} + rg(\hat{u})]_+ - \hat{\lambda}) \|^2 \]

\[ +\frac{1}{r} (\lambda^{k-1} - \hat{\lambda}, ([\lambda^{k-1} + rg(\hat{u}^k)]_+ - \lambda^{k-1}) - ([\hat{\lambda} + rg(\hat{u})]_+ - \hat{\lambda}) \rangle. \]

On account of (3.2.27) and the property of a saddle point we have

\[ [\hat{\lambda} + rg(\hat{u})]_+ - \hat{\lambda} = 0 \]

and the latter inequality implies

\[ \langle |\lambda^{k-1} + rg(\hat{u}^k)|_+ - [\hat{\lambda} + rg(\hat{u})]_+, g(\hat{u}^k) - g(\hat{u}) \rangle \]

\[ \geq \frac{1}{r} \| \lambda^{k-1} - \lambda^{k-1} \|^2 + \frac{1}{r} (\lambda^{k-1} - \hat{\lambda}, \lambda^{k-1} - \hat{\lambda}) \rangle \]

\[ = \frac{1}{r} (\lambda^{k-1} - \hat{\lambda}, \lambda^{k-1} - \lambda^{k-1}). \]

Now from (3.2.33) and (3.2.34) we conclude that

\[ \langle \hat{u}^k - u^{k-1}, \hat{u} - u^k \rangle \geq \frac{1}{2r} (\lambda^{k-1} - \hat{\lambda}, \lambda^{k-1} - \lambda^{k-1}) \]

and after a short calculation

\[ \| \hat{u}^k - u^k \|^2 + \frac{1}{2r} \| \lambda^{k-1} - \hat{\lambda} \|^2 + \| \hat{u}^k - u^{k-1} \|^2 + \frac{1}{2r} \| \lambda^{k-1} - \lambda^{k-1} \|^2 \]

\[ \leq \| u^{k-1} - \hat{u} \|^2 + \frac{1}{2r} \| \lambda^{k-1} - \hat{\lambda} \|^2. \]

Setting \( z := (u, \frac{1}{\sqrt{2r}}\lambda)^T \), \( z^k := (u^k, \frac{1}{\sqrt{2r}}\lambda^k)^T \) etc., the latter inequality can be rewritten in the form

\[ \| \hat{z}^{k-1} - z \|^2 + \| \hat{z}^{k-1} - z \|^2 \leq \| z^{k-1} - \hat{z} \|^2 \]

(3.2.35)

and consequently,

\[ \| \hat{z}^{k-1} - \hat{z} \| \leq \| z^{k-1} - \hat{z} \|. \]

(3.2.36)

In order to estimate \( \| z^k - z \| \) from below, we infer from (3.2.28) and (3.2.29) that

\[ \| z^k - z \| \leq \sqrt{\frac{1}{4} + \frac{\epsilon_k^2}{2r}} \epsilon_k =: c_2 \epsilon_k, \]
hence,
\[ \|z^k - \hat{z}\|^2 = \|z^k - z^k + z^k - \hat{z}\|^2 \geq \|z^k - \hat{z}\|^2 - 2\|z^k - \hat{z}\|\|z^k - z^k\| \]
\[ \geq \|z^k - \hat{z}\|^2 - 2c_2\epsilon_k\|z^k - \hat{z}\|^2. \]

Together with (3.2.35) this yields
\[ \|z^k - \hat{z}\|^2 + \|z^{k-1} - \hat{z}\|^2 \leq \|z^{k-1} - \hat{z}\|^2 + 2c_2\epsilon_k\|z^k - \hat{z}\|^2. \]  \( (3.2.37) \)

But from (3.2.36) it follows that
\[ \|z^k - \hat{z}\| \leq \|z^k - z^k\| + \|z^k - \hat{z}\| \leq c_2\epsilon_k + \|z^{k-1} - \hat{z}\|. \]

Regarding the structure of the latter inequality and Lemma A3.1.4, we obtain the convergence of the sequence \( \{\|z^k - \hat{z}\|\} \) because \( \sum_{k=1}^{\infty} \epsilon_k < \infty \). In connection with (3.2.37) this leads to \( \|z^k - z^{k-1}\| \to 0 \), hence,
\[ \|\hat{u}^k - u^{k-1}\| \to 0, \quad \|\hat{\lambda}^k - \lambda^{k-1}\| \to 0. \]  \( (3.2.38) \)

It is easy to verify that for \( \lambda \geq 0 \) the equality
\[ [g(u)]_+ = \left[ \frac{1}{r}\lambda + g(u) \right]_+ - \frac{1}{r}\lambda \]
is true. So in view of (3.2.27) and (3.2.38) we get
\[ \|[\hat{u}^k]_+\| \leq \|[\frac{1}{r}\lambda^{k-1} + g(\hat{u}^k)]_+ - \frac{1}{r}\lambda^{k-1}\| = \frac{1}{r}\|\lambda^{k} - \lambda^{k-1}\|. \]  \( (3.2.39) \)

The sequence \( \{ (u^k, \lambda^k) \} \) is bounded. Therefore a convergent subsequence \( \{ (u^{k_i}, \lambda^{k_i}) \} \) can be chosen with \( \lim_{i \to \infty} (u^{k_i}, \lambda^{k_i}) = (u^*, \lambda^*). \) Thus, with (3.2.38) we have
\[ \lim_{i \to \infty} (\hat{u}^{k_i}, \hat{\lambda}^{k_i}) = (u^*, \lambda^*). \]

Observing the continuity of \( [g(\cdot)]_+ \) and relation (3.2.39), we can conclude that \( [g(u^*)]_+ = 0 \), i.e. the point \( u^* \) is feasible. From (3.2.39) it follows also that
\[ \frac{1}{r}\lambda^* + g(u^*)]_+ - \frac{1}{r}\lambda^* = 0, \]  \( (3.2.40) \)

hence
\[ \lambda^* g(u^*) = 0. \]  \( (3.2.41) \)

Inequality (3.2.31) considered for \( k := k_i \), together with (3.2.38), lead to
\[ \langle \nabla J(u^*) + \nabla g(u^*)^T \lambda^* + rg(u^*) \rangle_+ , u - u^* \rangle \geq 0, \quad \forall \ u \in \mathbb{R}^n, \]
and due to (3.2.40) we obtain
\[ \nabla J(u^*) + \nabla g(u^*)^T \lambda^* = 0. \]  \( (3.2.42) \)

Since \( u^* \) is a feasible point, the relations (3.2.41) and (3.2.42) establish that \( (u^*, \lambda^*) \) is a saddle point for Problem A3.4.56.

Finally, from the convergence of \( \{\|z^k - \hat{z}\|\} \) for each \( \hat{z} \in U^* \times \Lambda^* \), we are going
to conclude the statement.

It should be noted that the assumption on the existence of a global Lipschitz constant for $g$ on $\mathbb{R}^n$ can be omitted if we make sharper estimates in the final part of the proof.

In [351] the following estimates have been established additionally:

$$J(u^k) - J^* \leq -\langle \lambda^k, g(u^k) \rangle + (\epsilon_k + 2\|u^k - u^{k-1}\|)\|u^k - u^*\|,$$

$$J(u^k) - J^* \leq 2\|\lambda^*\||u^k - u^{k-1}\|$$

(cf. also inequality (3.2.22) in Proposition 3.2.5).

In [14] a dual method of iterative PPR has been investigated using a Lagrangian function regularized with respect to the dual variables. The iteration proceeds as follows:

$$\lambda^k \approx \arg \max_{\lambda \in \mathbb{R}^m_+} \left\{ -\frac{1}{2} \|\lambda - \lambda^{k-1}\|^2 + D(u^{k-1}, \lambda) \right\},$$

$$u^k \approx u^{k-1} - \alpha_k (\nabla J(u^{k-1}) + \nabla g(u^{k-1})^T \lambda^k),$$

with $D(u, \lambda) := J(u \pm \lambda, g(u)) - \frac{\alpha}{2} \|\nabla J(u) + \nabla g(u)^T \lambda\|^2$, $\alpha > 0$. The step-size $\{\alpha_k\}$ has to be chosen as in Proposition A3.4.39. Conditions imposed on the smoothness of $J$ and $g$, and on the controlling parameters ensure the convergence of $\{u_k, \lambda_k\}$ to some point of $U^* \times \Lambda^*$ even if $\Lambda^*$ is unbounded. The function $D(u^k, \cdot)$ is concave and quadratic. Thus, calculating $\lambda^k$, a simple quadratic programming problem has to be solved.

### 3.3 PPR in Methods of Numerical Analysis

In this section we briefly consider well-known numerical methods which can be described within the framework of proximal-point methods, even though their inventions are motivated by quite different points of view.

#### 3.3.1 Implicit Euler method

The problem of unconstrained minimization of a convex differentiable function $J : \mathbb{R}^n \to \mathbb{R}$ can be formulated as a problem of looking for a stationary point of the differential equation

$$\frac{du}{dt} = -\nabla J(u), \quad u(0) = u^0.$$  

It is obvious that the application of the explicit Euler method

$$\frac{u^{k+1} - u^k}{\tau} = -\nabla J(u^k), \quad k = 1, 2, \cdots,$$

corresponds to the gradient method with step-length parameter $\tau$.

On the other hand, the use of the implicit Euler method

$$\frac{u^{k+1} - u^k}{\tau} = -\nabla J(u^{k+1}), \quad k = 1, 2, \cdots,$$  \hspace{1cm} (3.3.1)
CHAPTER 3. PPR FOR FINITE-DIMENSIONAL PROBLEMS

is equivalent to the minimization of the functional $J$ by means of the iterative proximal-point-method

$$u^{k+1} := \arg \min_{u \in \mathbb{R}^n} \Psi_{k,\chi}(u),$$

with $\Psi_{k,\chi} := J(u) + \frac{\chi}{2} \|u - u^k\|^2$, $\chi := \frac{1}{\tau}$ and the starting point $u^0 := u(0)$. This is true because $\nabla \Psi_{k,\chi}(u) = 0$ is evident with the iteration scheme (3.3.1).

In the theory of finite-difference methods it is well-known that the implicit Euler method provides a better stability. It is absolutely stable, whereas the explicit scheme is only conditional stable (cf. for instance [76, 368]). Of course, one iteration of the proximal method is more expensive than one step of the gradient method. Nevertheless, for stiff differential equations, which usually occur in connection with ill-posed optimization problems, the implicit scheme is more preferable.

Solving in the Hilbert space $V$ the operator equation

$$A u = f$$

with a symmetric, semi-definite operator $A : V \to V$, $f \in V$, Kryanev [250] has investigated the iteration scheme

$$(I + \tau A)u^{k+1} = u^k + \tau f, \quad k = 1, 2, \ldots,$$

with $I$ the identity operator, $\tau > 0$ and $u^0 \in V$ arbitrarily chosen. This method can also be interpreted as a proximal-point regularization

$$u^{k+1} := \arg \min_{u \in V} \left\{ J(u) + \frac{1}{2\tau} \|u - u^k\|^2 \right\}$$

for the unconstrained minimization of the quadratic functional

$$J(u) := \frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle.$$

In this case strong convergence of the PPR is guaranteed, whereas for general convex variational problems only weak convergence can be established.

3.3.2 Levenberg-Marquardt method

The Levenberg-Marquardt method, originally developed for solving non-linear least-squares problems, has been adapted later on in order to minimize functions $J : \mathbb{R}^n \to \mathbb{R}$ which are not necessarily convex. This method generates iterates according to

$$u^{k+1} := u^k - (\nabla^2 J(u^k) + \chi_{k+1} I)^{-1} \nabla J(u^k).$$

Note that for $\chi_k := 0$ we just get the Newton method, whereas for $\chi_k \to \infty$ the descent directions approaches the negative gradient.

It is easily seen that Method (3.3.5) can be rewritten equivalently in the form

$$u^{k+1} := \arg \min_{u \in \mathbb{R}^n} \left\{ J_k(u) + \frac{\chi_{k+1}}{2} \|u - u^k\|^2 \right\}.$$
3.4. PPR FOR NONCONVEX PROBLEMS

with

\[ J_k(u) := J(u^k) + \langle \nabla J(u^k), u - u^k \rangle + \frac{1}{2} \langle \nabla^2 J(u^k)(u - u^k), u - u^k \rangle, \]

i.e., this scheme can be interpreted as an iterative PPR of the classical Newton method.

It should be mentioned that another interpretation of the Levenberg-Marquard method became the starting point for the development of trust-region algorithms which have attracted considerable attention during the last decades (cf. Fletcher [117]). We come back to this point of view in Section 3.4.3.

3.3.3 Linearization method

Now we turn to the linearization method (A3.4.64) sketched briefly in Appendix A3.4.2. This method is also interesting in the context of this book, because it can be interpreted as an iterative PPR for constrained problems. To begin with, consider Problem (A3.4.56), linearized at point \( u^{k-1} \):

\[
\min \{ \langle \nabla J(u^{k-1}), u - u^{k-1} \rangle \} \tag{3.3.6}
\]

\[
st. \quad g_j(u^{k-1}) + \langle \nabla g_j(u^{k-1}), u - u^{k-1} \rangle \leq 0, \quad j \in I_{k-1}(u^{k-1}). \tag{3.3.7}
\]

It should be noted that in the linearized problem some of the initial constraints may be deleted.

In Remark A3.4.40 it was mentioned that the fully linearized problem may not be solvable. Therefore, mere linearization is not suitable even as a standard method for solving well-posed problems.

In order to recapitulate Method (A3.4.64), let \( \bar{u}^k \) be the minimizer of the regularized function

\[
\langle \nabla J(u^{k-1}), u - u^{k-1} \rangle + \frac{\chi}{2} \| u - u^{k-1} \|^2, \quad \chi > 0,
\]

subject to the constraints (3.3.7). Then the next iterate is calculated by

\[
u^k := u^{k-1} + \gamma (\bar{u}^k - u^{k-1}),
\]

with \( \gamma > 0 \) determined by some line search procedure.

This method has the following essential advantage of iterative PPR: Together with a good stability of the auxiliary problems it ensures convergence of a minimizing sequence to a solution of the original problem. Moreover, Proposition A3.4.39 guarantees that the method is stable under small data perturbations in \( C^1(\mathbb{R}^n) \) in the original problem.

3.4 PPR for Nonconvex Problems

In this section we are dealing with the application of proximal point methods for non-convex minimization problems

\[
\min \{ f(u) : u \in K \}
\]

with \( f : \mathbb{R}^n \to \mathbb{R} \) and \( K \subset \mathbb{R}^n \) a closed set.

As far as we know, the first direct application of the proximal point method
for certain nonconvex unconstrained minimization problems was performed by Fukushima and Mine [123].

If we drop the convexity assumptions on the objective function $f$ and/or the feasible set $K$, several problems arise (see [220]). For instance, the prox operator is possibly no more well-defined because it is not clear whether a unique global minimizer of the regularized objective function exists at all. Moreover, in general the prox operator might not be non-expansive anymore even in arbitrary small neighborhoods of local or global minima.

### 3.4.1 Example.
Take $K := \mathbb{R}^2$ and $f(u) := \min[u_1^2, u_2^2]$. Then the optimal set is given by $U^* = \{ u \in \mathbb{R}^2 : u_1 = 0 \lor u_2 = 0 \}$. The prox points of $u^1 = (2\alpha, \alpha)^T$ and $u^2 = (\alpha, 2\alpha)^T$ with an arbitrarily chosen $\alpha > 0$ and $\chi := 2$ are given by

$$\text{Prox}_{f,\mathbb{R}^n} u^1 = \left(2\alpha, \frac{\alpha}{2}\right)^T, \quad \text{Prox}_{f,\mathbb{R}^n} u^2 = \left(\frac{\alpha}{2}, 2\alpha\right)^T,$$

thus

$$\| \text{Prox}_{f,\mathbb{R}^n} u^1 - \text{Prox}_{f,\mathbb{R}^n} u^2 \| = \sqrt{\frac{9}{2}\alpha} > \sqrt{2} = \| u^1 - u^2 \|,$$

showing the destruction of the non-expansivity.

As the following example shows, the prox iteration may terminate in a point which is a fixed point of the prox operator but neither a local minimum nor a stationary point of $f$ even if the regularized objective function is convex.

### 3.4.2 Example.
Consider the feasible set

$$K := \{ u \in \mathbb{R}^2 : u_1 + u_2 \leq 0, u_1 \leq 0, -u_2 \leq 1 \}$$

and the objective function $f(u) := -u_1 - \frac{1}{2}u_2^2$. Starting with $u^0 = (-\frac{1}{4}, \frac{1}{4})^T$ and choosing $\chi := 2$ we obtain

$$\text{Prox}_{f,K} u^0 = (0, 0)^T = \text{Prox}_{f,K} 0,$$

telling us that the point $0 = (0, 0)^T$ is a fixed point of the prox mapping but not a local minimum of the initial problem.

Despite these obstacles there are situations of successful applications of the proximal point algorithm. The following lemma shows that the set of proximal points at a given argument of the prox operator is never empty. Consequently, the proximal mapping is set-valued with $\text{dom} (\text{Prox}_{f,K}) = \mathbb{R}^n$.

### 3.4.3 Proposition.
Let $f$ be continuous and $\inf_{v \in \mathbb{R}^n} f(v) > -\infty$.

(i) Then for all $\chi > 0$ and $u \in \mathbb{R}^n$

$$\text{Arg min}_{v \in \mathbb{R}^n} \Psi_{\chi,\mathbb{R}^n}(v) \neq \emptyset,$$

with

$$\Psi_{\chi,\mathbb{R}^n}(v) = f(v) + \frac{\chi}{2} \| v - u \|^2.$$

(ii) For any $\chi > 0$ and fixed $u \in \mathbb{R}^n$ let $u^*(\chi) = u^*(\chi, u)$ such that

$$u^*(\chi) \in \text{Arg min}_{v \in \mathbb{R}^n} \Psi_{\chi,\mathbb{R}^n}(v).$$

Then $\lim_{\chi \to \infty} u^*(\chi) = u$. 

3.4. PPR FOR NONCONVEX PROBLEMS

Proof: Let \( u \in \mathbb{R}^n \) and \( \chi > 0 \) be arbitrary but fixed. Due to the boundedness of \( f \) we also have \( c := \inf_{v \in \mathbb{R}^n} \Psi_{\chi,R^n}(v) > -\infty \). This together with the continuity of \( f \) guarantees the existence of a bounded sequence \( \{v^k\} \) with \( \Psi_{\chi,R^n}(v^k) \to c \). Unboundedness of \( \{v^k\} \) would be result in the unboundedness of \( \{\Psi_{\chi,R^n}(v^k)\} \), contradicting its convergence. Hence, there exists a convergent subsequence \( \{v^{k_i}\} \) such that

\[
\Psi_{\chi,R^n}(\hat{v}) = \lim_{i \to \infty} \Psi_{\chi,R^n}(v^{k_i}) = c,
\]

and accordingly \( \hat{v} \in \text{Arg min}_{v \in \mathbb{R}^n} \Psi_{\chi,R^n}(v) \).

Again by assumption

\[
f(u^*(\chi)) + \frac{\chi}{2} \|u^*(\chi) - u\|^2 = \Psi_{\chi,R^n}(u^*(\chi)) \leq \Psi_{\chi,R^n}(u) = f(u),
\]

showing that \( f(u^*(\chi)) \leq f(u) \) for any \( \chi > 0 \) and

\[
\|u^*(\chi) - u\|^2 \leq \frac{2}{\chi} [f(u) - f(u^*(\chi))].
\]

Now, taking limit \( \chi \to \infty \), due to the boundedness of \( [f(u) - f(u^*(\chi))] \) we get \( \|u^*(\chi) - u\| \to 0 \). □

The next example was considered by Voetmann [412], who calculated the prox points by a Matlab routine with the highest adjustable accuracy. In this example we are only interested in the prox points themselves.

3.4.4 Example. Consider the Rosenbrock function

\[
f(u) := 100(u_2 - u_1^2)^2 + (1 - u_1)^2,
\]

which is non-convex but has a unique global minimizer at \( u^* = (1,1)^T \) and no further local minima. The regularized objective function \( \Psi_{\chi,R^2}(v) \) also has a unique global minimum for all \( \chi > 0 \) and \( u \in U^0 := \{v \in \mathbb{R}^2 : -2 \leq v_1 \leq 2, -2 \leq v_2 \leq 2\} \). Starting the proximal iterations with arbitrary but bounded \( 0 < \chi_k < \chi \) in \( u^0 = (-1.2, 1)^T \) the resulting sequence always converges to \( u^* \).

<table>
<thead>
<tr>
<th>( k )</th>
<th>( u_1^k )</th>
<th>( u_2^k )</th>
<th>( k )</th>
<th>( u_1^k )</th>
<th>( u_2^k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-1.2</td>
<td>1</td>
<td>8</td>
<td>0.972203</td>
<td>0.945065</td>
</tr>
<tr>
<td>1</td>
<td>0.518094</td>
<td>0.272061</td>
<td>9</td>
<td>0.980335</td>
<td>0.960977</td>
</tr>
<tr>
<td>2</td>
<td>0.720923</td>
<td>0.518498</td>
<td>10</td>
<td>0.986047</td>
<td>0.972232</td>
</tr>
<tr>
<td>3</td>
<td>0.821633</td>
<td>0.674301</td>
<td>11</td>
<td>0.990079</td>
<td>0.980216</td>
</tr>
<tr>
<td>4</td>
<td>0.881008</td>
<td>0.775668</td>
<td>12</td>
<td>0.992935</td>
<td>0.985892</td>
</tr>
<tr>
<td>5</td>
<td>0.918694</td>
<td>0.843659</td>
<td>13</td>
<td>0.994964</td>
<td>0.989934</td>
</tr>
<tr>
<td>6</td>
<td>0.943624</td>
<td>0.890193</td>
<td>14</td>
<td>0.996408</td>
<td>0.992814</td>
</tr>
<tr>
<td>7</td>
<td>0.960538</td>
<td>0.922472</td>
<td>15</td>
<td>0.999999</td>
<td>0.999999</td>
</tr>
</tbody>
</table>

Table 3.4.1: Iteration log of Example 3.4.4

Table 3.4.1 shows the computed iterates for \( \chi_k := 100 \). Obviously convergence occurs and from \( k = 3 \) on the iterates follow the bottom of the steep valley that is characteristic for the Rosenbrock function. ☑
Proposition 3.4.3 and Example 3.4.4 show that one can realistically expect that there are non-convex optimization problems that allow a successful convexification based on an suitable choice of the prox parameter $\chi_k$.

### 3.4.1 Convexification of unconstrained problems

In the sequel, we deal with Problem

$$\min_{u \in \mathbb{R}^n} f(u)$$

(3.4.1)

without convexity of the objective function $f$. However, the application of the exact proximal point method will be studied under the hypothesis that the regularized objective functions

$$\Psi_k(u) := f(u) + \frac{\chi_k}{2}||u - u^{k-1}||^2$$

(3.4.2)

are strongly convex on some convex set $\Omega$, which has to contain the sequence of proximal iterates $\{u^k\}$. Due to the evident non-increase of $\{f(u^k)\}$, $\Omega$ may be any convex set containing $\{u \in \mathbb{R}^n : f(u) \leq f(u^0)\}$. In such situation usually we can handle the regularized auxiliary problem

$$\min_{v \in \Omega} \Psi_k(v)$$

(3.4.3)

as a convex minimization problem, and this is a very essential argument for applying proximal methods. But, of course, the assumption that the function $\Psi_k(\cdot)$ becomes convex under a suitable choice of $\chi_k > 0$ is restrictive (see for instance the problem $f(u) := \min ||u||, ||u + 1||$ on $\mathbb{R}^1$).

Therefore, we start with a simple statement which establishes the possibility of convexification for an important class of non-convex functions.

#### 3.4.5 Proposition. Let $f : \mathbb{R}^n \to \mathbb{R}$ be of the form

$$f(u) := \sup_{\tau \in T} \varphi(u, \tau),$$

with $\tau \in T \subset \mathbb{R}^1$ and $\varphi(\cdot, \tau)$ differentiable for all $t \in T$ on a convex set $\Omega$ with $\Omega \cap \text{dom} f \neq \emptyset$. Suppose that for each $\tau \in T$ the gradient $\nabla_u \varphi(\cdot, \tau)$ is Lipschitz-continuous on $\Omega$ with constant $L_\tau$ and $L = \sup_{\tau \in T} L_\tau < \infty$ and $\sup_{\tau \in T} ||\nabla_u \varphi(u, \tau)|| < \infty$ for some $\hat{u} \in \Omega \cap \text{dom} f$.

Then, for $\chi \geq L$, the function $f(\cdot) + \frac{\chi}{2}||\cdot||^2$ is convex and finite on $\Omega$.

**Proof:** For each $\tau \in T$ and arbitrary $u, v \in \Omega$ one gets

$$\langle \nabla_u \varphi(u, \tau) + \chi u - \nabla_u \varphi(v, \tau) - \chi v, u - v \rangle \geq -L||u - v||^2 + \chi||u - v||^2.$$  

Hence, $\varphi(\cdot, \tau) + \frac{\chi}{2}||\cdot||^2$ is convex on $\Omega$ if $\chi \geq L$.

Due to the properties of $\varphi(\cdot, \tau)$, for any $v, u \in \Omega$, we obtain also

$$\varphi(u, \tau) = \varphi(v, \tau) + \int_0^1 \langle \nabla_u \varphi(v + t(u - v), \tau), u - v \rangle dt$$

$$= \varphi(v, \tau) + \int_0^1 \langle \nabla_u \varphi(v + t(u - v), \tau) - \nabla_u \varphi(v, \tau), u - v \rangle dt$$

$$+ \langle \nabla_u \varphi(v, \tau), u - v \rangle$$

$$\leq \varphi(v, \tau) + \frac{L_\tau}{2}||v - u||^2 + \langle \nabla_u \varphi(v, \tau), u - v \rangle.$$
3.4. PPR FOR NONCONVEX PROBLEMS

Setting in this inequality \( v := \bar{u} \) and taking into account that
\[
\sup_{\tau \in T} \| \nabla_u \varphi(u, \tau) \| =: c < \infty,
\]
one can conclude that
\[
\varphi(u, \tau) \leq \varphi(\bar{u}, \tau) + \frac{L_\tau}{2} \| u - \bar{u} \|^2 + c \| u - \bar{u} \|,
\]
hence,
\[
\sup_{\tau \in T} \varphi(u, \tau) + \frac{\chi}{2} \| u \|^2 = \sup_{\tau \in T} \{ \varphi(u, \tau) + \frac{\chi}{2} \| u \|^2 \} < \infty \ \forall \ u \in \Omega.
\]
In view of the convexity of \( \varphi(\cdot, \tau) + \frac{\chi}{2} \cdot \| \cdot \|^2 \) for each \( \tau \in T \), this ensures that the function \( f(\cdot) + \frac{\chi}{2} \cdot \| \cdot \|^2 \) is convex and finite on \( \Omega \). \( \Box \)

Now, let \( f : \mathbb{R}^n \to \mathbb{R} \) be a given lower semicontinuous function and we suppose that the functions \( \Psi_k(\cdot) \) for \( \chi_k \in (\chi, \bar{\chi}] \), \( \chi > 0 \), are strongly convex on \( \Omega \ni \{ u \in \mathbb{R}^n : f(u) \leq f(\hat{u}) \} \). Then, starting the exact proximal point method
\[
u^k = \arg \min_{u \in \mathbb{R}^n} \{ f(u) + \frac{\chi_k}{2} \| u - u^{k-1} \|^2 \}
\]
with \( u^0 : f(u^0) \leq f(\hat{u}) \), we have the known facts that
\begin{itemize}
  \item \( u^k \) is uniquely defined;
  \item \( 0 \in \partial (f(u^k) + \frac{\chi_k}{2} \| u^k - u^{k-1} \|^2) \);
  \item \( f(u^k) + \frac{\chi_k}{2} \| u^k - u^{k-1} \|^2 \leq f(u^{k-1}) \).
\end{itemize}
If \( \inf_{u \in \mathbb{R}^n} f(u) > -\infty \), then the latter inequality yields
\[ f(u^k) \to \hat{f} > -\infty \text{ and } \| u^k - u^{k-1} \| \to 0. \]
Remember that a nonlinear programming problem
\[
\min \{ f_0(u) : f_j(u) \leq 0, \ j = 1, \ldots, m \},
\]
whose Lagrangian function possesses a saddle point, can be transformed into the unconstrained problem
\[
\min \{ f(u) : u \in \mathbb{R}^n \},
\]
with \( f(u) := \max_{0 \leq j \leq m} \eta_j(u), \ \eta_0 := f_0, \ \eta_j := f_0 + \alpha f_j, \ j = 1, \ldots, m, \) and a sufficiently large \( \alpha > 0 \).
Along with other applications this motivates to consider the proximal point method (3.4.4) for \( f(u) = \max_{i \in I} \varphi_i(u) \), \( |I| < \infty \), with differentiable functions \( \varphi_i \). Concerning \( \varphi_i \), we suppose also that their gradients are Lipschitz-continuous (with constants \( L_i \)) on some convex set \( \Omega \ni \{ u : f(u) \leq f(\hat{u}) \} \). Then, due to the finiteness of the index set \( I \), the conditions of Proposition 3.4.5 are satisfied.

3.4.6 Theorem. Let the hypotheses made before on \( \varphi_i \) be valid and \( \chi_k \) be chosen such that \( \max_{i \in I} L_i < \chi_k \leq \bar{\chi} \). Moreover, let \( \inf_{u \in \mathbb{R}^n} f(u) > -\infty \). Then, starting Method (3.4.4) with \( u^0 : f(u^0) \leq f(\hat{u}) \), we obtain either
(i) $u^{k-1} = u^k$ holds for some $k$, providing that $u^k$ is a stationary point of $f$, or

(ii) any accumulation point of $\{u^k\}$ is a stationary point of $f$.

**Proof** Due to the differentiability of $\varphi_i$ and convexity of $\Psi_k(\cdot) := f(\cdot) + \frac{\lambda_k}{2} \| \cdot - u^{k-1} \|^2$ on $\Omega$ the subdifferential $\partial \Psi_k(u^k)$ is the convex hull of the gradients at $u^k$ of the functions $\varphi_i(\cdot) + \frac{\lambda_k}{2} \| \cdot - u^{k-1} \|^2$, $i \in I(u^k)$, where $I(u^k) = \{ i \in I : \varphi_i(u^k) = f(u^k) \}$.

Hence, for some $\lambda^k_i \geq 0$, $i \in I(u^k)$, such that $\sum_{i \in I(u^k)} \lambda^k_i = 1$, we obtain

$$0 = \sum_{i \in I(u^k)} \lambda^k_i \nabla \varphi_i(u^k) + \chi_k(u^k - u^{k-1}). \quad (3.4.6)$$

If $u^k = u^{k-1}$, then

$$0 = \sum_{i \in I(u^k)} \lambda^k_i \nabla \varphi_i(u^k),$$

thus, the point 0 is included in Clarke’s subdifferential $\partial_{cl} f(u^k)$, proving that $u^k$ is a stationary point of $f$. If $\{u^k\}$ is an infinite sequence and $\bar{u}$ is its accumulation point, then due to the finiteness of $I$ and $0 \leq \lambda^k_i \leq 1 \forall i \in I(u^k), \forall k$, we are able to choose a subsequence $\{k_l\}$ such that

$$I(u^{k_l}) = I(u^{k_2}) = \ldots = \bar{I},$$

and for $l \to \infty$,

$$u^{k_l} \to \bar{u}, \quad \text{and} \quad \lambda^k_l \to \bar{\lambda}_l, \quad \forall i \in \bar{I}.$$ 

Moreover, because of $\inf_{u \in \mathbb{R}^n} f(u) > -\infty$, the relation

$$\lim_{k \to \infty} \|u^k - u^{k-1}\| = 0$$

is valid. Now taking limit in (3.4.6) for $k = k_l$, $l \to \infty$, one gets

$$0 = \sum_{i \in \bar{I}} \lambda_i \nabla \varphi_i(\bar{u}), \quad \text{with} \quad \bar{\lambda}_i \geq 0, \quad \sum_{i \in \bar{I}} \bar{\lambda}_i = 1.$$ 

On account of $I(\bar{u}) \supset \bar{I}$ and the property of Clarke’s subdifferential, this means that $0 \in \partial_{cl} f(\bar{u})$, i.e. $\bar{u}$ is a stationary point of $f$. \hfill \Box

**3.4.7 Remark.** If the set $\{u : f(u) \leq f(u^0)\}$ is unbounded, the existence of accumulation points is not guaranteed in general. Moreover, in the case that Problem (3.4.1) is convex, it is known that $\|u^k\| \to \infty$ if $U^* = \emptyset$. \hfill \Diamond

**3.4.8 Remark.** If $\bar{u} = \lim_{l \to \infty} u^{k_l}$ and $f$ is convex in some neighborhood $\mathcal{B}_r(\bar{u}) := \{u : \|u - \bar{u}\| \leq r\}$ with $r > 0$, then of course, $\bar{u}$ is a local minimum of $f$. In this case, on account of $\|u^k - u^{k-1}\| \to 0$, for sufficient large $l$ one can claim that $\|u^{k_l} - \bar{u}\| \leq \frac{r}{2}$ and

$$\|u^{k_l+s} - u^{k_l+s-1}\| \leq \frac{r}{2}, \quad s = 1, 2, \ldots$$
Thus, \( u^{k+1} \in B_r(u) \) and hence \( u^{k+1} \) is the minimum point of the sum of the convex functions \( f \) and \( \frac{\chi_{k+1}}{2} \cdot -u^k \| u^k \|^2 \) on \( B_r(u) \). From Proposition II.2.2 in [100] we obtain

\[
    f(\bar{u}) - f(u^{k+1}) + \chi_{k+1} \left( u^{k+1} - u^k, \bar{u} - u^{k+1} \right) \geq 0.
\]

Together with the evident identity

\[
    \| u^{k+1} - \bar{u} \|^2 - \| u^k - \bar{u} \|^2 = -\| u^{k+1} - u^k \|^2 + 2 \left( u^{k+1} - u^k, u^{k+1} - \bar{u} \right)
\]

this leads to

\[
    \| u^{k+1} - \bar{u} \|^2 - \| u^k - \bar{u} \|^2 \leq -\| u^{k+1} - u^k \|^2 + \frac{2}{\chi_{k+1}} \left( f(\bar{u}) - f(u^{k+1}) \right)
\]

providing that \( \| u^{k+1} - \bar{u} \| \leq \| u^k - \bar{u} \| \) and, in general,

\[
    \| u^{k+s} - \bar{u} \| \leq \| u^{k+s-1} - \bar{u} \|, \quad s = 1, 2, \ldots
\]

Hence, the sequence \( \left\{ \| u^k - \bar{u} \| \right\} \) converges, and in view of \( \lim_{t \to \infty} u^k = \bar{u} \), one can conclude that \( \lim_{k \to \infty} u^k = \bar{u} \). ♦

It should be noted that we did not suppose that \( \bar{u} \) is the unique minimum in the neighborhood \( B_r(u) \).

The following result is of interest, in particular, for semi-infinite programming problems, which often can be reduced to the unconstrained minimization of functions of the type \( f(u) = \sup_{\tau \in T} \varphi(u, \tau) \) with \( T \) a compact set.

**3.4.9 Theorem.** Let the function \( f \) be defined by \( f(u) := \sup_{\tau \in T} \varphi(u, \tau) \), where \( T \subset \mathbb{R}^l \) is a compact set and \( \varphi \) is continuous on \( \mathbb{R}^n \times T \). Moreover, suppose that \( \inf_{u \in \mathbb{R}^n} f(u) > -\infty \) and that \( \nabla_u \varphi \) is continuous on \( \Omega \times T \), where \( \Omega \) is an open convex set containing \( \{ u : f(u) \leq f(\bar{u}) \} \). Finally, let

\[
    \| \nabla_u \varphi(u', \tau) - \nabla_u \varphi(u'', \tau) \| \leq L \| u' - u'' \| \quad \forall \ u', u'' \in \Omega, \forall \ \tau \in T.
\]

Then the function \( f(\cdot) + \frac{\chi}{2} \cdot \| \cdot \|^2 \) is strongly convex on \( \Omega \) if \( \chi > L \), and for Method (3.4.4) with \( L < \chi_k \leq \chi \) and the starting point \( u^0 : f(u^0) \leq f(\bar{u}) \) the conclusion of Theorem 3.4.6 remains true.

**Proof:** Due to the compactness of \( T \) and the continuity of \( \nabla_u \varphi \) on \( \Omega \times T \), finiteness of \( \sup_{\tau \in T} \| \nabla_u \varphi(u, \tau) \| \) is guaranteed for each \( u \in \Omega \). Therefore, Proposition 3.4.5 implies strong convexity of \( f(\cdot) + \frac{\chi}{2} \cdot \| \cdot \|^2 \) on \( \Omega \) for \( \chi > L \).

Moreover, \( f \) is a locally Lipschitzian function and its upper Clarke’s derivative coincides with the directional derivative (see [89], Example 2.1.4). Therefore, the theorem about Clarke’s subdifferential for the sum of functions (cf. [77]) permits to conclude that, for \( u \in \Omega \),

\[
    \partial D \left( f(u) + \frac{\chi_{k+1}}{2} \| u - u^k \|^2 \right) = \partial D f(u) + \partial D \left( \frac{\chi_{k+1}}{2} \| u - u^k \|^2 \right),
\]

and regarding the convexity of \( f(\cdot) + \frac{\chi_{k+1}}{2} \cdot -u^k \| u - u^k \|^2 \) and \( \frac{\chi_{k+1}}{2} \cdot -u^k \| u - u^k \|^2 \), this leads to

\[
    \partial \left( f(u) + \frac{\chi_{k+1}}{2} \| u - u^k \|^2 \right) = \partial D f(u) + \chi_{k+1} (u - u^k).
\]
Hence,

\[ 0 \in \partial \left( f(u^{k+1}) + \frac{\chi_{k+1}}{2} \| u^{k+1} - u^k \|^2 \right) = \partial_{\chi} f(u^{k+1}) + \chi_{k+1} (u^{k+1} - u^k), \quad (3.4.7) \]

and if \( u^k = u^{k+1} \), then \( u^{k+1} \) is a stationary point of \( f \).

If \( \{ u^k \} \) is an infinite sequence and \( \bar{u} = \lim_{i \to \infty} u^{k_i} \), then the inclusion \( 0 \in \partial_{\chi} f(\bar{u}) \) follows from (3.4.7), relation \( \| u^{k+1} - u^k \| \to 0 \) as well as from the closedness of the mapping \( u \to \partial_{\chi} f(u) \).

In principle, Theorem 3.4.6 could be considered as a particular case of Theorem 3.4.9, however, we preferred to give a direct proof.

Now, let us consider a very special case in non-convex programming where it is possible to guarantee that the proximal point method generates a sequence \( \{ u^k \} \) converging to \( \bar{u} \in \text{Arg min}\{ f(u) : u \in \mathbb{R}^n \} \).

Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a lower semicontinuous function and dom \( f \neq \emptyset \). With a given \( c > \inf_{u \in \mathbb{R}^n} f(u) \), define \( N_c = \{ u : f(u) \leq c \} \). Again we want to solve the problem

\[ \min \{ f(u) : u \in \mathbb{R}^n \}, \]

but now under the following

3.4.10 Assumption.

(i) \( U^* = \{ u : f(u) = \inf_{v \in \mathbb{R}^n} f(v) =: f^* \} \neq \emptyset \);

(ii) \( N_c \) is contained in a convex set \( \Omega \), and for some \( \chi > 0 \) the function \( \Psi_{\chi, \Omega}(u) = f(u) + \frac{\chi}{2} \| u \|^2 \) is convex on \( \Omega \);

(iii) for some \( c_0 \in [f^*, c) \) and \( \chi \) chosen as in (ii) one has

\[ \| y(u) - \chi u \| > d > 0 \quad \forall u \in N_c \setminus N_{c_0}, \quad \forall y(u) \in \Lambda(u), \quad (3.4.8) \]

where \( \Lambda(u) = \partial \Psi_{\chi, \Omega}(u) - \chi u \);

(iv) \( N_{c_0} \) is convex, and \( f \) is convex on \( N_{c_0} \).

3.4.11 Theorem. The proximal iterates in (3.4.4) with \( \chi < \chi_k \leq \bar{\chi} \) and arbitrary \( u^0 \in N_c \) converge to a solution point \( u^* \in U^* \).

Proof: Let \( u^0 \in N_c \setminus N_{c_0} \). Then it is clear that \( u^k \in N_c \) for all proximal steps \( k \). Assume now that \( u^{k+1} \in N_c \setminus N_{c_0} \). Due to

\[ f(u) + \frac{\chi_{k+1}}{2} \| u - u^k \|^2 = \Psi_{\chi_{k+1}}(u) + \frac{\chi_{k+1} - \chi}{2} \| u \|^2 - \chi_{k+1} (u, u^k) + \frac{\chi_{k+1}}{2} \| u^k \|^2 \]

and the definition of \( u^{k+1} \), we obtain

\[ 0 \in \partial \Psi_{\chi_{k+1}}(u^{k+1}) - \chi u^{k+1} + \chi_{k+1} (u^{k+1} - u^k) \]

Observing (3.4.8), this leads to

\[ \| u^{k+1} - u^k \| > \frac{d}{\bar{\chi}}. \]
3.4. PPR FOR NONCONVEX PROBLEMS

Now, taking into account the inequality
\[
f(u_{k+1}) + \frac{\lambda_{k+1}}{2} \|u_{k+1} - u_k\|^2 \leq f(u_k),
\]
one can conclude that
\[
f(u_{k+1}) < f(u_k) - \frac{\lambda}{2\chi^2}d^2. \tag{3.4.9}
\]
Estimate (3.4.9) shows that, after a finite number of steps, the prox-iterates fall into the set \( N_{c_0} \). With regard to Assumption 3.4.10(iv) and the non-increase of \( \{ f(u_k) \} \), the use of the convergence results for convex problems (see, for instance, Proposition 3.1.8) ensures convergence of \( \{ u_k \} \) to a solution \( u^* \in U^* \).

The same happens if \( u^0 \in N_{c_0} \).

\[\Box\]

3.4.12 Example. (Showing the fulfilment of Assumption 3.4.10):
Consider in \( \mathbb{R}^2 \) the following function \( f(u) = u_1^2 + u_2^2 + 15u_1^2u_2^2 \). Obviously, Assumption 3.4.10(i) is valid: \( u^* = (0,0)^T \) is the unique minimum. The function \( f \) is non-convex, but it is strongly convex on the sphere \( \{ u \in \mathbb{R}^2 : \| u \| \leq \frac{1}{\sqrt{15}} \} \). Outside of this sphere we have \( \| \nabla f(u) \| > \frac{2}{\sqrt{15}} \). Therefore, the Assumptions 3.4.10(iii) and (iv) are satisfied, for instance with \( c = 2 \) and \( c_0 = \frac{1}{15} \). As it follows from Proposition 3.4.5, Assumption 3.4.10(ii) can also be satisfied for any convex compact set \( \Omega \) containing \( N_c \) with \( c = 2 \).

\[\Diamond\]

Insignificant modifications in the proofs of the Theorems 3.4.6, 3.4.9 and 3.4.11 permit us to establish analogous statements for the inexact version of the proximal point method.

3.4.2 Convexification of constrained problems

We start this subsection by considering the constrained problem
\[
\min \{ f(u) : u \in K \}, \tag{3.4.10}
\]
where now \( K \subset \mathbb{R}^n \) is supposed to be a convex closed set and \( f \) is twice differentiable on \( K \). Let us introduce the operator
\[
T_u := \begin{cases} 
\nabla f(u) + N_K(u) & \text{if } u \in K, \\
\emptyset & \text{if } u \notin K,
\end{cases}
\]
with \( N_K(u) := \{ v \in \mathbb{R}^n : \langle v, z - u \rangle \leq 0 \; \forall \; z \in K \} \) the normal cone of the set \( K \) at the point \( u \in K \).

3.4.13 Assumption.

(i) For a given \( c > \inf \{ f(u) : u \in K \} \) the set \( Q = \{ u : 0 \in Tu, f(u) \leq c \} \) is non-empty, and if \( \bar{u} \in Q \), then \( \bar{u} \) is a local minimum of \( f \) on \( K \);

(ii) For some \( \bar{d} > 0 \) the set \( \{ u : f(u) \leq c, \inf_{y \in Tu} \| y \| < \bar{d} \} \) is bounded.

\[\Diamond\]
3.4.14 Proposition. Suppose that Assumption 3.4.13 is satisfied for Problem (3.4.10). Then for each \( \delta > 0 \) there exists \( d \in (0, d) \) such that \( u \) belongs to the \( \delta \)-neighborhood \( \mathbb{B}_{\delta}(\bar{u}) \) of some local minimum \( \bar{u} \in Q \), whenever \( u \in K \), \( f(u) \leq c \) and \( \inf_{y \in T} \| y \| < d \).

**Proof:** Suppose that for some \( \delta > 0 \) such a constant \( d \) does not exist. Denoting by \( \mathbb{B}_{\delta} \) the union of \( \delta \)-neighborhoods of all local minima in \( Q \), we choose a sequence \( \{d_k\} \to +0 \), \( d_k < d \) and define \( w^k \in K \setminus \mathbb{B}_{\delta} \), \( y^k \in T w^k \) such that

\[
f(w^k) \leq c, \quad \| y^k \| < d_k.
\]

Due to Assumption 3.4.13(ii), without loss of generality, one can assume that \( \{w^k\} \) converges to a point \( \bar{w} \). For this point we infer that

\[
\bar{w} \in K \quad \text{and} \quad f(\bar{w}) \leq c. \tag{3.4.11}
\]

Because of \( y^k = \nabla f(w^k) + v^k \) for some \( v^k \in \mathcal{N}_K(w^k) \) and \( \nabla f(w^k) \to \nabla f(\bar{w}) \), as well as \( \|y^k\| \to 0 \), one gets \( v^k \to \bar{v} = -\nabla f(\bar{w}) \). But from the definition of the normal cone \( \mathcal{N}_K(w^k) \), the inequality

\[
(v^k, w^k - z) \geq 0, \quad \forall \ z \in K
\]

holds true, and taking limit, we infer that

\[
\langle \bar{v}, \bar{w} - z \rangle \geq 0, \quad \forall \ z \in K,
\]

i.e. \( \bar{v} \in \mathcal{N}_K(\bar{w}) \). Therefore, \( 0 = \nabla f(\bar{w}) + \bar{v} \in T \bar{w} \), and due to (3.4.11) and Assumption 3.4.13(i), \( \bar{w} \) is a local minimum, contradicting the fact that \( \{w^k\} \cap \mathbb{B}_{\delta} = \emptyset \).

Suppose now that for Problem (3.4.10) the set \( \{u \in K : f(u) = \inf_{v \in K} f(v)\} \) is non-empty and that for some \( \chi > 0 \) the function \( f(\cdot) + \frac{\chi}{2} \| \cdot \|^2 \) is convex on \( K \). Due to Proposition 3.4.5, these conditions can be fulfilled, in particular, if \( K \) is a non-empty and bounded set. Then, obviously, the function

\[
\zeta(u) := \begin{cases} f(\cdot) + \frac{\chi}{2} \| \cdot \|^2 & \text{if } u \in K, \\ +\infty & \text{if } u \notin K,
\end{cases}
\]

is convex, lower semicontinuous and \( \partial \zeta(u) - \chi u = T u \). If, moreover, Assumption 3.4.13 is valid, then using Proposition 3.4.14 and the proof of Theorem 3.4.11, one can easily show that, for arbitrary \( \delta > 0 \), the iterates \( w^k \) generated by the exact method (3.4.2) with \( \chi < \chi_k \leq \bar{\chi} \) and the starting point \( u^0 \in K : f(u^0) \leq c \), belong to \( \mathbb{B}_{\delta} \) for sufficiently large \( k \) \( (k \geq k(\delta)) \).

3.4.15 Example. Let \( K := \{u \in \mathbb{R}^2 : -1 \leq u_1 \leq 1, \ -1 \leq u_2 \leq 1\} \) and \( f(u) := (1 - u^2_1)u^2_2 \). Evidently, we have

\[
U^* = \{(u_1, 0)^T : -1 \leq u_1 \leq 1\} \cup \{(1, u_2)^T : -1 \leq u_2 \leq 1\} \\
\cup \{(-1, u_2)^T : -1 \leq u_2 \leq 1\}
\]

and

\[
\{u \in K : \nabla f(u) = 0\} = \{(u_1, 0)^T : -1 \leq u_1 \leq 1\} \subset U^*.
\]

Now, calculating \( \nabla f \) on the boundary of \( K \), it is easy to see that Assumption 3.4.13(i) is fulfilled for any \( c > 0 \), and Assumption 3.4.13(ii) follows from the boundedness of \( K \).

\( \diamond \)
3.4. PPR FOR NONCONVEX PROBLEMS

Of course, it should be emphasized that the verification of the Assumptions 3.4.10(iii)-(iv) and 3.4.13(i) causes difficulties. However, these conditions are not binding for treating proximal point methods. Nevertheless, the complete validity of these assumptions ensures a much more comfortable result, namely, according to the Theorems 3.4.6, 3.4.9 and 3.4.11, convergence to a local or global minimum.

Finally we investigate the following non-convex problem

$$\min \{ f(u) : h(u) = 0 \},$$

with twice differentiable, non-convex functions $f$ and $h_j, h(u) := (h_1(u), ..., h_m(u))^T$.

The augmented Lagrangian function is defined by

$$L_A(u, \lambda, \rho) := L(u, \lambda) + \frac{\rho}{2} \|h(u)\|^2,$$

with $L(u, \lambda) := f(u) + \langle \lambda, h(u) \rangle$.

An often used assumption to ensure convergence of the iterates of the multiplier methods to local saddle points $(u^*, \lambda^*)$ is the positive definiteness of the matrix

$$\nabla_{uu}^2 L_A(u^*, \lambda^*, \rho)$$

for sufficiently large $\rho$. Nash and Sofer \cite{302} gave the following estimate:

**3.4.16 Proposition.** Let $(u^*, \lambda^*)$ be a local solution of Problem (3.4.12), $H$ be the Hessian of $L$ at $(u^*, \lambda^*)$ and $Z$ be a basis matrix of the null-space of $G = \nabla h(u^*)^T$. Assume that the matrix $G$ has full rank, i.e. linear independence constrained qualification is satisfied, and $Z^T \nabla_{uu}^2 L(x^*, \lambda^*) Z$ is positive definite, i.e. strong second order constraint qualification holds.

Then the matrix $\nabla_{uu}^2 L_A(u^*, \lambda^*, \rho)$ is positive definite for

$$\rho > \rho := \frac{1}{\delta} \left( \frac{\beta \gamma^2}{\alpha} + \gamma \right),$$

with

$$\alpha := \ell_{\min}(Z^T H Z), \quad \gamma := \|H\|_2, \quad \beta := \|Z\|_2, \quad \delta := \ell_{\min}(G^T G),$$

where $\ell_{\min}(\cdot)$ denotes the smallest eigenvalue of the given matrix.

It is known that the conditioning of the matrix $\nabla_{uu}^2 L_A(u^*, \lambda^*, \rho)$ usually worsens with growing $\rho$ if $\text{rank}(A) < n$ and although $\rho$ is not required to go to infinity - in contrast to penalty methods - there is obviously a conflict between shifting the eigenvalues of $\nabla_{uu}^2 L_A(u^*, \lambda^*, \rho)$ to the positive axis by large values of $\rho$ and good condition numbers requiring small values of $\rho$.

In order to circumvent this conflict we consider the primal-dual application of the proximal point methods (cf. the dual methods described in Subsection A3.4.4). These are based on the regularized augmented Lagrangian function

$$\Psi_A(v, \lambda, \rho, \chi, u) := L_A(v, \lambda, \rho) + \frac{\chi}{2} \|v - u\|^2$$

and lead to the so-called **regularized method of multipliers**:

$$u^{k+1} := \arg\min_{v \in \mathbb{R}^n} \Psi_A(v, \lambda^k, \rho^k, \chi_k, u^k), \quad \lambda^{k+1} := \lambda^k + \rho^k h(u^k).$$

The counterpart of Proposition 3.4.16 reads now:
3.4.17 Proposition. Let the assumptions of Proposition 3.4.16 hold. Then $\nabla_{uu}^2 \Psi_A(u^*, \lambda^*, \rho, \chi, u)$ is positive definite for all $u \in \mathbb{R}^n$ and

$$\rho > \rho := \frac{1}{\delta} \left( \frac{\beta^2(\gamma + \chi)^2}{\alpha - \chi \beta^2} + \gamma - \chi \right).$$

(3.4.14)

The proof is based on the same principles as the proof of Proposition 3.4.16 simply regarding the proximal term as a part of the objective function.

The local convergence analysis can now be carried out simultaneously to that of the classical method of multipliers with the peculiarity that two parameters, the penalty $\rho$ and the regularization parameter $\chi$, are at hand to control the method.

3.4.3 Proximity control in trust region methods

In the previous subchapters we analyzed the most common used convexification effect of the regularization term. Additionally it can be interpreted as a penalty term for the step length. As this aspect is commented on less frequently in the literature we want to analyze it in this subchapter.

Just as step-size algorithms trust-region methods can be seen as a globalization technique for locally fast convergent methods such as Newton- or Quasi-Newton methods (cf. Subsection A3.2.5). In contrast to those though step-size and descent direction are determined simultaneously. To concentrate on the conceptual ideas we consider the unconstrained minimization problem (3.4.1) such that the prox operator can be simplified to $\text{Prox}_{f, \chi} \equiv \text{Prox}_{f, \chi, \mathbb{R}^n}$.

At a given iterate $u^k$ we solve the so-called trust-region subproblem

$$\min \{ m_k(u) : \| u - u^k \| \leq \Delta_k \},$$

(3.4.15)

where $m_k(\cdot)$ is a model for the objective function $f$ in a neighborhood of the current iterate $u^k$. Generally the model is expected to be a good local approximation only justifying the inclusion of the step-length constraint. Having determined a possibly approximate solution $u^k + d^k$ of (3.4.15) we can assess the quality of the model $m_k$ by relating the actual reduction to the reduction predicted by the model, i.e. we form the following ratio

$$r^k := \frac{f(u^k) - f(u^k + d^k)}{m_k(u^k) - m_k(u^k + d^k)}. $$

(3.4.16)

Based on the value $r^k$ the step $d^k$ is accepted, i.e. $u^{k+1} := u^k + d^k$, and the radius $\Delta_k$ is increased if a sufficient high reduction was achieved (successful step) or the step is rejected and the radius is decreased indicating that the model cannot be trusted and should be improved (unsuccessful step).

The basic scheme still has of course many degrees of freedom allowing for a great variety of trust-region type algorithms that differ mainly in the choice of the model $m_k$, in the choice of the norm used for the step restriction, in the method of determining a global solution of the trust-region subproblem and the adjustment of the trust-radius. For a comprehensive overview of trust-region methods we refer to Conn et al. [82].
3.4. PPR FOR NONCONVEX PROBLEMS

The most advantageous property of this basic trust-region scheme is that
global convergence to at least a stationary point can be shown under mild as-
ssumptions. Furthermore in many algorithms necessary second order conditions
hold in the limit, see for instance [82, 117].

Traditionally the most prominent model functions are quadratic approximations
of \( f \) at \( u^k \):

\[
m_k(u) := f(u^k) + \langle \nabla f(u^k), u - u^k \rangle + \frac{1}{2}(H^k(u - u^k), u - u^k),
\]  

(3.4.17)

with a symmetric positive definite matrix \( H^k \) which is an approximation of the
Hessian of \( f \) at \( u^k \).

Based on these observations we are going to have a look at the relationship
between the trust-region problem

\[
\min \{ m_k(u) : \| u - u^k \|^2 \leq \Delta_k^2 \}
\]  

(3.4.18)

and

\[
\min \{ m_k(u) + \chi_k^2 \| u - u^k \|^2 \}
\]  

(3.4.19)

with varying exact and inexact model functions \( m_k(\cdot) \). Here, instead of Prob-
lem (3.4.15) we formulate an equivalent restricted subproblem (3.4.18) avoiding
peculiarities due to the non-differentiability of the constraints in (3.4.18). Of
course the resulting Lagrange multipliers differ between these problems.

Considering quadratic models of type (3.4.17), the advantage is that when
used in conjunction with a norm of type \( \| x \| = \langle Ax, x \rangle \) with a positive defi-
nite matrix \( A \), the trust-region subproblem (3.4.18) is an all-quadratic problem
whose solutions can be completely characterized.

In the most important case \( A := I \) we have in particular:

3.4.18 Theorem. (see [126])
Consider Problem (3.4.18) with \( \Delta_k > 0 \), Euclidean norm \( \| \cdot \|_2 \) and model func-
tion (3.4.17). Then \( u^* \) is a global solution iff a unique \( \lambda^* \geq 0 \) exists such that
the following holds:

(i) \( (H^k + 2\lambda^* I)(u^* - u^k) = -\nabla f(u^k) \);

(ii) \( \| u^* - u^k \| \leq \Delta_k \) and \( \lambda^*(\| u^* - u^k \| - \Delta_k) = 0 \);

(iii) \( (H^k + 2\lambda^* I) \) is positive semidefinite.

Conditions (i) and (ii) are the Kuhn-Tucker conditions for (3.4.18) but
due to the special structure of the problem they are not only necessary but also
sufficient. Likewise condition (iii) is stronger than the general necessary second
order condition as the latter only ensures positive definiteness of the model
Hessian on the set of feasible directions in \( u^* \). Theorem 3.4.18 shows that the
global minimizer of (3.4.18) lie on the trajectory

\[
\{ u(\lambda) : (H^k + 2\lambda I)(u - u^k) = -\nabla f(u^k), \quad \lambda \in (\max[0, -\frac{\ell_{\min}}{2}], \infty) \},
\]  

(3.4.20)

with \( \ell_{\min} \) smallest eigenvalue of \( (H^k + 2\lambda I) \).

It is furthermore well-known (see [372]) that the Lagrange multiplier is
monotonically decreasing with increasing trust-region radius. The most popular
trust-region methods thus try to find $\lambda^*$ by solving the nonlinear equation

$$\|u(\lambda) - u^k\| = \Delta_k$$

(3.4.21)

through a search along the solution trajectory (3.4.20). This approach can be
motivated by a dual point of view. The dual problem of (3.4.18) is given by

$$\max_{\lambda \geq 0} \{ q(\lambda) : \lambda \in \text{dom}(q) \}$$

with

$$q(\lambda) := \inf_{u \in \mathbb{R}^n} \{ m_k(u) + \lambda(\|u - u^k\|^2 - \Delta_k^2) \}.$$ 

Obviously we have $\lambda \in \text{dom}(q)$ iff $(H_k + 2\lambda I)$ is positive definite. Determining
$u(\lambda)$ by (3.4.20) is then equivalent to evaluating the dual function, and solving
the nonlinear equation (3.4.21) is equivalent to finding the dual optimal $\lambda^*$. Naturally we cannot rely on strong duality results. On the contrary a duality
gap must be expected if the matrix $H_k$ is not positive definite, i.e. $m_k$ is not
convex. But due to the compactness and regularity of the feasible set we can
work with local duality results in the sense of [302].

Observing that the quadratic function $m_k(u) + \frac{\lambda}{2}\|u - u^k\|^2$ is strongly
convex, if $(H_k + 2\lambda I)$ is positive definite, we see that $u^*$ is also a global solution
to (3.4.19), i.e.,

$$u^* \text{ solves Problem (3.4.18)} \iff u^* = \text{Prox}_{m_k,\chi_k}(u^k), \ \chi_k := 2\lambda^*.$$ (3.4.22)

The following statement tells us how the regularization- and the trust-
region parameters are coupled.

3.4.19 Proposition. (see [126])

Let $m_k$ be differentiable and $u^*$ be a global minimum of (3.4.18). Then the unique
Lagrange multiplier $\lambda^*$ is given by

$$\lambda^* := -\frac{1}{2\Delta_k^2} (u^* - u^k, \nabla m_k(u^*)).$$ (3.4.23)

In the non-convex case methods of Levenberg-Marquard type on the other hand
choose a priori $\lambda_k$ such that $(H_k + \lambda_k I)$ is positive definite, then compute
$u(\lambda_k)$ as the respective point on the trajectory (3.4.20) and consequently
determine the trust-region radius implicitly through the chosen $\lambda_k$. In light of (3.4.22)
we can conclude that this approach is essentially the proximal point algorithm
where the exact prox point is approximated by only one (Quasi-)Newton step
yielding the usual convergence properties of trust-region methods without assuming
any convexity on the objective function $f$ (cf. Subsection 3.3.2).

Summing up, the classical trust-region method can be regarded as a prox-
imal point method where the regularization parameter may be chosen a prioriy
(yielding a Levenberg-Marquard type of method) or may be implicitly deter-
dined via the Lagrange multiplier of the trust-region subproblem. The proximal
points are not calculated exactly but approximately by one Newton step.

Concerning more general convex model functions $m_k(\cdot)$ we can show a
complete equivalence between computing a proximal iteration and solving a
trust-region subproblem, because the respective solutions can be fully charac-
terized analogously to Theorem 3.4.18 without the need to distinguish local
and global solutions.

The above approach using (quadratic) convex model functions has been used specifically in the context of bundle methods for non-differentiable minimization problems. Given a non-differentiable, convex function $f$, Schramm and Zowe [366] consider the piece-wise linear model

$$m_k(u) := \max_{j \in J_k} \left\{ f(y^j) + \langle \partial f(y^j), u - y^j \rangle \right\}$$

constructed by means of a bundle $\{y^j : j \in J^k\}$ of previously computed trial points and subgradients $\partial f(y^j)$. A trial step is executed by minimizing the regularized bundle function

$$u^{k+1} := \arg\min_{u \in \mathbb{R}^n} \left\{ m_k(u) + \frac{\chi_k}{2} \|u - u^k\|^2 \right\}.$$

The effect of adding the regularization term to $m_k$ is again that the function to be minimized is strongly convex and the unique minimizer can be found by solving a quadratic auxiliary problem. As shown above the trial step can be regarded as the solution of a trust-region subproblem allowing it to be assessed by the ratio test (3.4.16) yielding a criterion for its acceptance as a new iterate and its admission into the bundle. The regularization parameter $\chi_k$ is decreased if the model is a good approximation of the function $f$ allowing for wider steps, otherwise increased to restrict the steps.

Kiwiel [239] describes a version of this bundle idea for the minimization of differentiable non-convex functions. By defining

$$\alpha(u, y) := \|f(u) - f(y) - (g(y), u - y)\|,$$

$$m_k(u) := \max_{j \in J^k} \left\{ f(u^k) - \alpha(u^k, y^j) + \langle \partial g(y^j), u - u^k \rangle \right\},$$

the model is again a piece-wise linear but now possibly shifted model of the objective function, with $g(y)$ and $g(y^j)$ elements of a generalized gradient (due to the non-convexity it is of course not possible to make use of a subgradient, therefore

$$\partial f(u) := \text{conv}\left\{ \lim_{k \to \infty} \nabla f(u^k) : u^k \to u, \nabla f(u^k) \text{ exists} \forall u^k \right\}$$

denotes a generalized gradient).

In the case of convex $f$ this bundle approximation coincides with (3.4.24). The new search direction is again determined by minimizing the regularized model but this time an additional trust-region constraint is added:

$$u^{k+1} := \arg\min_{\|u - u^k\|^2 \leq \Delta_k^2} \left\{ m_k(u) + \frac{\chi_k}{2} \|u - u^k\|^2 \right\}.$$  

(3.4.25)

This problem is once again uniquely solvable and well defined due to the strong convexity of the regularized model function.

In light of the above statements it seems surprising why the auxiliary problems (3.4.25) are formulated with a regularized objective function and a trust-region constraint. As the approximation function is convex, only one of
these modifications seems to be necessary as either the prox-term or the trust-region constraint captures the desired effect of proximity and the respective other term seems to be redundant. But coupling both terms has the following advantages compensating for the aggravated problem structure.

The iterations are controlled by two parameters assigned to different tasks:

(a) The radius $\Delta_k$ controls the accuracy of $m_k$. If the ratio test demands a reduction of $m_k$ and the discrepancy measured by $\alpha(u^k, y^j)$ is too large, the bundle is reexamined and points $y^j$ distant from the current iterate $u^k$ are eliminated from the bundle.

(b) The regularization parameter $\chi_k$ controls the decrease of $m_k$. After a successful iteration (in view of the ratio test) $\chi_k$ is reduced and vice versa.

Without the regularizing term the objective function of Problem (3.4.25) is piece-wise linear and one has to deal with possibly non-unique minimizers. By including the quadratic regularizing term the subproblem consists of minimizing a strongly convex function subject to a convex constraint such that the solution and the corresponding multiplier are uniquely determined and can be computed efficiently.

Summing up the method can be formally described as a PPR as at the end of a successful step we have:

$$
\hat{u}^{k+1} := \arg\min_{u \in \mathbb{R}^n} \{ m_k(u) + \frac{\chi_k + \lambda_k}{2} \|u - u^k\|^2 \},
$$

$$
u^{k+1} := u^k + t_k(\hat{u}^{k+1} - u^k),
$$

where $\lambda^k$ is the optimal Lagrange multiplier associated with the solution of (3.4.25) and $t_k$ is a suitably chosen step-size. Additionally it can be seen as a trust-region method making use of a regularized model function.

As Toint [398] points out, the choice of the function $f$ itself as an exact model function is possible and leads to globally convergent algorithms. The subproblems to solve then have the form

$$
\min \{ f(u) : \|u - u^k\|^2 \leq \Delta^2_k \}. \tag{3.4.26}
$$

It is not necessary to solve this problem exactly but an approximate solution with respect to a decrease condition is sufficient for convergence. Therefore the problem can be rewritten in

$$
\min \quad f(x) + \frac{\chi_k}{2} \|u - u^k\|^2 - \frac{\chi_k}{2} \|u - u^k\|^2 \\
\quad \text{s.t.} \quad \|u - u^k\|^2 \leq \Delta^2_k,
$$

Now, if $\chi_k$ is chosen such that $f(u) + \frac{\chi_k}{2} \|u - u^k\|^2$ becomes convex on the set $\{u : \|u - u^k\| \leq \Delta_k\}$, the objective function can be seen as a $\text{DC-function}$, that is a difference of two convex functions and one can employ methods of DC-optimization to solve this problem, see Voetmann [412].
3.5 Comments

Section 3.1: The concept of proximal-point mapping was introduced by Moreau [292], including also the proofs of the main properties of this mapping, in particular, non-expansivity of Prox$_{f,C}$ and differentiability of the Moreau-Yosida function $\eta(u) := \min_{v \in C} \{f(v) + \frac{1}{2}\|v - u\|^2\}$ (see Proposition 3.1.4).

Other variants of Proposition 3.1.3 can be found in [211] and [212].

Section 3.2: For the theoretical foundation of proximal point methods we refer to the fundamental paper of Rockafellar [352]. Some results of this paper have been extended by Ha [160] towards the regularization on a subspace of a Hilbert space.

Rockafellar [351] and Antipin [13] comment about earlier applications of the PPR applied to convex optimization methods. In both papers the regularized Lagrangian method (cf. Algorithm 3.2.8) was considered under milder assumptions than required in Theorem 3.2.9: In [351] smoothness of the functionals $J, g$ and $\phi_k$ is not required and in [13] it is permitted that the Lagrange multipliers $\lambda^k$ are calculated approximately. Later on more general regularized multiplier methods have been developed by Antipin [14].

For convex separable problems Spingarn [380] suggested a decomposition scheme, where in the framework of augmented Lagrangian methods iterative PPR is applied to each subproblem generated in this scheme of decomposition.

In the papers of Rockafellar [351] and IBARAKI ET AL. [190] special dual methods for convex optimization problems are considered, where the Lagrangian is regularized with respect to the primal and dual variables. Both approaches differ from each other with respect to the order of performing the min- and max-operation for finding the saddle point of the regularized Lagrangian. While Rockafellar’s approach is sufficiently general, Ibaraki’s is predestinated only for separable problems with linear constraints.

Kaplan investigated in [203, 204] penalty methods in the context with PPR for a wide class of penalty functions. In particular, in [204] a penalty function with a slow increase outside the feasible set is considered. In connection with regularized penalty methods he also proved in [205] their convergence, see the Theorems 3.2.1 and 3.2.3.

The papers of Auslender ET AL. [21], Mouallif [297], Mouallif and Tossings [299] and Alart and Lemaire [5] are mainly concerned with the use of specific classes of penalty functions, in particular those, which are suitable to non-smooth problems.

Section 3.3: Concerning the results mentioned in Subsection 3.3.1 we refer additionally to Morosov [294], who has studied the iteration scheme

$$(A^T A + \alpha I) u^k = A^T f + \alpha u^{k-1}, \quad \alpha > 0,$$

for solving a linear operator equation $Au = f$ with a degenerated operator $A$. This can be interpreted as a PPR for the minimization of the quadratic functional $J(u) = \|Au - f\|^2$.

Section 3.4: A retrospective analysis of some algorithms for non-convex minimization shows their relationship to PPR. This concerns, in particular, a series of linearization methods, see Pshenichnyj [339].
Spingarn [379] has developed the PPR for finding a zero of a maximal strictly hypomonotone operator \( T : \mathbb{R}^n \to 2^{\mathbb{R}^n} \). In this method, proximal iterates are calculated not with the operator \( T \) but with an auxiliary Lipschitz continuous operator constructed in a convex neighborhood \( U \) of a point \( \bar{u} \in T \bar{u} \).

The model, adapted in [379] to the problem of the local minimization of a function \( f = g - h \), with \( g \) lsc, convex and \( h \in C^2 \), turns into the usual proximal point method

\[
 u^{k+1} := \arg \min_{u \in U} \{ f(u) + \frac{\chi_k}{2} \| u - u^k \|^2 \}. 
\]

If \( \chi_k > 0 \) is chosen such that \( f(\cdot) + \frac{\chi_k}{2} \| \cdot \|^2 \) is convex on \( U \), and \( u^0 \) is close enough to a local minimum \( \bar{u} \) of \( f \), and if the mapping \( \partial f^{-1} \) has a monotone derivative at \( (0, \bar{u}) \), then \( \{ u^k \} \) converges to \( \bar{u} \) linearly.

Numerical difficulties arise in the trust-region approach (3.4.18) with quadratic model functions \( m_k \) via (3.4.17) when the Lagrange multiplier \( \lambda \) or the eigenvalues \( \ell_{\min}(H_k + \lambda I) \) become small. The latter happens if \( \nabla f(u^k) \) lies orthogonal to the eigenspace of the smallest eigenvalues of \( H_k \) (the so-called hard case), see Fletcher [117]. Due to these disadvantages methods with direct choice of the radius \( \Delta_k \) are preferred in most practical algorithms. Bonnans and Launay [47] apply this technique of a regularized model function in an SQP-context in order to penalize large steps without including an extra trust region constraint.
Chapter 4

PPR FOR INFINITE-DIMENSIONAL PROBLEMS

In Section 2.2 a preliminary introduction to One-Step (OSR) and Multi Step (MSR) methods of iterative PPR was given. Subject of this chapter is a general framework for investigating methods of PPR.

Although this framework is rather abstract, it is suitable to construct and analyze special numerical methods for ill-posed, infinite as well as semi-infinite convex optimization problems, in particular, also elliptic variational inequalities. The essential property of the related methods is their notable stability at all iteration levels, achieved by a coordinate controlling of the parameters of discretization, regularization as well as those parameters arising in the basic optimization algorithms.

It should be noted that an OSR-method can be interpreted as a special case of a MSR-method. However, in order to demonstrate the peculiarities of OSR-methods, including the update of the controlling parameters and their estimate of the rate of convergence, we will investigated this class of methods separately.

In order to simplify the theoretical study of IPR we assume here and in the following chapter that in the regularized functional

\[ \Psi_k(u) := J_k(u) + \frac{\chi_k}{2} ||u - u^k||^2 \]

and in related expressions the parameter \( \chi_k \) in the proximity term is chosen equal to 2. However, from numerical point of view this might be not a reasonable choice, see also Subsection 4.4.1.

4.1 Preliminary Results

Let us start with the Lipschitz property of some functionals on the ball \( B_r := \{ u \in V : ||u|| \leq r \} \) in the Hilbert space \( V \).
Obviously, for the functional
\[
\varphi_z(u) := \|u - z\|^2
\]
with fixed \(z \in B_r\) it holds
\[
|\varphi_z(u') - \varphi_z(u'')| \leq 4r\|u' - u''\|, \quad \forall u', u'' \in B_r,
\]
whereas for the functional
\[
\delta(u) := \rho^2(u, B_{r_1}), \quad (r_1 < r),
\]
the inequality
\[
|\delta(u') - \delta(u'')| \leq 2r\|u' - u''\|, \quad \forall u', u'' \in B_r
\]
is true. The latter can be easily seen using the representation
\[
\delta(u) = \left\{ \begin{array}{ll}
(||u|| - r_1)^2 & \text{if } ||u|| > r_1, \\
0 & \text{otherwise}
\end{array} \right.
\]
and observing, in view of the symmetry, the following three cases:
\[
||u'|| \leq r_1, \quad ||u''|| \leq r_1,
||u'|| > r_1, \quad ||u''|| \leq r_1,
||u'|| > r_1, \quad ||u''|| > r_1.
\]

Next, we need some technical lemma. Let \(L(r)\) be the Lipschitz constant of the functional \(J\) on the ball \(B_r\) such that
\[
\sup_{u', u'' \in B_r} \frac{|J(u') - J(u'')|}{\|u' - u''\|} \leq L(r) < \infty,
\]
and denote \(\bar{J}(u) := J(u) + \rho^2(u, B_{r_1})\).

4.1.1 Lemma. Let \(\rho > 0\) and \(r_1 > 0\) be fixed, define
\[
r_0 := r_1 + L(r_1) + \sqrt{\rho + L(r_1)(4r_1 + L(r_1))}
\]
and assume that \(J : V \to \mathbb{R}\) is a convex functional such that inequality (4.1.3) is fulfilled for \(r := r_1\). Then the inclusion
\[
\{ u \in V : \bar{J}(u) \leq J(u^0) + \rho \} \subset \mathbb{B}_{r_0}
\]
holds for each \(u^0 \in \mathbb{B}_{r_1}\).

Proof: Due to the convexity of \(J\), the Lipschitz property of \(J\) and \(u^0 \in \mathbb{B}_{r_1}\) (cf. Corollary A1.5.16), we can conclude that for \(J(u) < J(u^0)\) the inequality
\[
J(u) \geq J(u^0) - L(r_1)\|u - u^0\|
\]
is fulfilled. Therefore, for every \(u\) with \(\|u\| \geq r_0\) we obtain
\[
\bar{J}(u) = J(u) + (\|u\| - r_1)^2 \geq J(u^0) - |J(u) - J(u^0)| + (\|u\| - r_1)^2 \geq J(u^0) - L(r_1)\|u - u^0\| + (\|u\| - r_1)^2 \geq J(u^0) - 2L(r_1)\|u\| + (\|u\| - r_1)^2.
\]
The function $-2L(r_1)t + (t-r_1)^2$ increases for $t \geq r_1 + L(r_1)$. Hence, for $\|u\| \geq r_0$, due to $r_0 > r_1 + L(r_1)$, we get
\[
\bar{J}(u) \geq J(u^0) - 2L(r_1)\|u\| + (\|u\| - r_1)^2 \\
\geq J(u^0) - 2L(r_1)r_0 + (r_0 - r_1)^2.
\]
But from (4.1.4) it follows that
\[
\varrho = (r_0 - r_1)^2 - 2L(r_1)r_0 - 2L(r_1)r_1 \\
< (r_0 - r_1)^2 - 2L(r_1)r_0.
\]
This proves that $\bar{J}(u) > J(u^0) + \varrho$, if $\|u\| \geq r_0$. □

4.1.2 Lemma. Assume that $f$ satisfies on $B_r$ the Lipschitz condition (4.1.3) and
\[
\sum_{k=1}^{\infty} \sqrt{\bar{\gamma}_k} < \infty.
\]
Then the sequence $\{\xi^k\}$, generated according to (4.1.5), converges weakly to some element $\xi^* \in G^*$ and $\lim_{k \to \infty} f(\xi^k) = f(\xi^*)$.

This is a particular case of Theorem 1 in Rockafellar [352]. It can also be deduced from the second part of the proof of Proposition 1.3.11 (starting from relation (1.3.30)).

Under the same conditions as above the following method can be investigated:

For an arbitrarily chosen $\xi^0 \in B_r$ a sequence $\{\xi^k\}$ is determined by
\[
\rho(\xi^k, G) \leq \bar{\gamma}_k, \quad \Psi_k(\xi^k) \leq \min_{u \in G} \Psi_k(u) + \bar{\gamma}_k, \quad k = 1, 2, \ldots
\]
with $\Psi_k(u) = f(u) + \|u - \xi^{k-1}\|^2$.

4.1.3 Lemma. Assume that $\{\bar{\gamma}_k\}$ and $\{\bar{\gamma}_k\}$ are non-negative sequences with
\[
\sum_{k=1}^{\infty} \sqrt{\gamma_k} < \infty, \quad \sum_{k=1}^{\infty} \sqrt{\bar{\gamma}_k} < \infty,
\]
and that the functional $f$ satisfies the Lipschitz condition (4.1.3) on a ball $\mathbb{B}_r$ with

$$\tau := r + \sup_k \frac{\gamma_k}{\lambda}$$

(or $\tau := r$ if $\xi^k \in \mathbb{B}_r$ for all $k$).

Then $\{\xi^k\} \to \xi^* \in G^*$ and $\lim_{k \to \infty} f(\xi^k) = f(\xi^*)$.

This can be established by a trivial modification of the proof of Proposition 1.3.11 taking into account Remark 1.3.13.

4.1.4 Lemma. Additional to the assumptions of Lemma 4.1.2 (or Lemma 4.1.3) the following inequality let be satisfied:

$$f(\lambda u' + (1-\lambda)u'') \leq \lambda f(u') + (1-\lambda)f(u'') - \lambda(1-\lambda)\kappa \|\Pi u' - \Pi u''\|^2, \quad \forall \ u', u'' \in V, \ \lambda \in [0,1]$$

(4.1.8)

with some $\kappa > 0$, where $\Pi$ denotes the ortho-projector onto the orthogonal complement $V_1^\perp$ of a finite-dimensional subspace $V_1 \subset V$ (cf. (1.2.12)). Then the sequence $\{\xi^k\}$, generated by (4.1.5) (or (4.1.7)), converges to some $\xi^* \in G^*$ in the norm of $V$.

Proof: Inequality (4.1.8) is equivalent to

$$f(u') - f(u'') \geq (q(u''), u' - u'') + \kappa \|\Pi u' - \Pi u''\|^2, \quad \forall \ u', u'' \in V$$

(4.1.9)

with some subgradient $q(u'') \in \partial f(u'')$ (see Statement 3.4.4 in [313]). Setting $u' := \xi^k, u'' := \xi^* \in (4.1.9)$ and taking limit as $k \to \infty$, then, on account of Lemma 4.1.2 (or Lemma 4.1.3), we conclude that $\|\Pi \xi^k - \Pi \xi^*\| \to 0$. On the other hand, $\|(I - \Pi) \xi^k - (I - \Pi) \xi^*\| \to 0$ follows from the weak convergence of $\{\xi^k\}$ and the finite dimensionality of $V_1$. \qed

The following result is useful for estimating the exactness of the iterates in numerical algorithms for the iterative PPR.

4.1.5 Lemma. If $f$ is a convex, continuous functional, $G \subset \mathbb{B}_r$ is a convex and closed set and for fixed $u^0 \in V$ the inequality

$$\|u^1 - u^0\| \leq \epsilon$$

is fulfilled with $u^1 := \text{Prox}_{f,G} u^0$, then the estimate

$$f(u^1) - \min_{u \in G} f(u) \leq 4\epsilon$$

(4.1.10)

holds.

Proof: Because of the definitions of $u^1$ and $u^* := \arg\min_{u \in G} f(u)$ the relation

$$f(u^1 + \lambda(u^* - u^1)) + \|u^1 + \lambda(u^* - u^1) - u^0\|^2 \geq f(u^1) + \|u^1 - u^0\|^2$$

is satisfied for all $\lambda \in (0,1]$. In view of the convexity of $f$ we get

$$\lambda f(u^*) + (1 - \lambda)f(u^1) + \lambda^2 \|u^* - u^1\|^2 + 2\lambda(u^* - u^1, u^1 - u^0) \geq f(u^1)$$
and
\[ f(u^1) - f(u^*) \leq 2\langle u^1 - u^0, u^* - u^1 \rangle + \lambda \|u^* - u^1\|^2. \]
For \( \lambda \downarrow 0 \), this leads to
\[ f(u^1) - f(u^*) \leq 2\langle u^1 - u^0, u^* - u^1 \rangle \leq 4r\epsilon. \]

To put it briefly, inequality (4.1.10) says: If two successive iterates are close, then the deviation of the current objective value from its optimal value is small, too.

### 4.2 One-Step Regularization Methods

We are going to describe one-step regularization methods (OSR-methods) for the following class of problems (cf. Problem (A1.5.11))

\[
\begin{align*}
\text{minimize} & \quad J(u), \\
\text{subject to} & \quad u \in K,
\end{align*}
\]

where \( J : V \to \bar{\mathbb{R}} \) is a convex, lower semi-continuous functional, \( K \subset V \) is a convex, closed set and \( K \cap \text{dom} J \neq \emptyset \), \( U^* := \text{Arg min}_{u \in K} J(u) \neq \emptyset \).

Let \( \{J_k\}, k = 1, 2, \ldots \), be a sequence of convex, Gâteaux-differentiable functionals on \( V \) and \( \{K_k\}, k = 1, 2, \ldots \), be a sequence of convex closed subsets of \( V \) approximating \( J \) and \( K \), respectively. The question how these approximations can be carried out will be answered in the next chapters when specific instances of this class are considered.

On this way a sequence of approximate problems

\[
\min_{u \in K_k} \{\Psi_k(u) := J_k(u) + \|u - u^{k-1}\|^2\}, \quad k = 1, 2, \cdots, \tag{4.2.2}
\]

can be constructed.

#### 4.2.1 Description of OSR-methods

Given a sequence \( \{\epsilon_k\} \) of non-negative numbers with \( \lim_{k \to \infty} \epsilon_k = 0 \), we analyze the following algorithm

**4.2.1 Method. (OSR-Method)**

Data: \( u^0 \in V, \{\epsilon_k\} \downarrow 0; \)

S0: set \( k := 0; \)

S1: compute in Problem (4.2.2) an approximate solution \( u^k \) such that

\[
\|\nabla \Psi_k(u^k) - \nabla \Psi_k(\tilde{u}^k)\|^2_{V^*} \leq \epsilon_k, \tag{4.2.3}
\]

where \( \tilde{u}^k := \text{Arg min}_{u \in K_k} \Psi_k(u); \tag{4.2.4} \)
S2: set $k := k + 1$ and go to S1.

Due to the strong convexity of $\Psi_k$, the minimizer $\bar{u}_k$ is uniquely defined.

Now, we are interested in conditions which ensure convergence of $\{u_k\}$ to some $u^* \in U^*$ in the weak or strong sense.

Assume that the radius $r^*$ is fixed such that

$$U^* \cap B_{r^*/8} \neq \emptyset,$$

and choose $r \geq r^*$.

Further denote

$$Q := K \cap B_r, \quad Q^* := U^* \cap B_{r^*}, \quad Q_k := K_k \cap B_r.$$

Obviously, we also need some quantitative information about the approximation of $J$ and $K$ mentioned above.

**4.2.2 Assumption.** For given sequences $\{\sigma_k\} \downarrow 0$ and $\{\mu_k\} \downarrow 0$ and for each $k = 1, 2, \cdots$ it holds

$$\sup_{u \in B_r} |J(u) - J_k(u)| \leq \sigma_k, \quad (4.2.5)$$

and

$$\rho(Q_k, Q) \leq \mu_k, \quad \rho(Q^*, Q_k) \leq \mu_k. \quad (4.2.6)$$

Moreover, according to Lemma 4.1.4, we consider

**4.2.3 Assumption.** The inequality

$$f(\lambda u' + (1 - \lambda)u'') \leq \lambda f(u') + (1 - \lambda)f(u'') - \lambda(1 - \lambda)\kappa\|\Pi u' - \Pi u''\|^2,$$

$$\forall u', u'' \in V, \forall \lambda \in [0, 1]$$

let be satisfied with some $\kappa > 0$ and $\Pi$ the ortho-projector onto the orthogonal complement $V_1^\perp$ of some finite-dimensional subspace $V_1 \subset V$.

Now we are able to formulate the convergence result.

**4.2.4 Theorem.** Suppose the radii $r^*, r$ are chosen as before and Assumption 4.2.2 is fulfilled. Moreover, assume that the functional $J$ is Lipschitz on $B_r$ according to (4.1.3), the starting point is chosen such that $u_0 \in B_{r^*/4}$ and the parameters are coordinated such that

$$\sum_{k=1}^{\infty} \left( \sqrt{2L(r)}\mu_k + 2\sigma_k + 2\mu_k + \epsilon_k \right) < \frac{r^*}{2}. \quad (4.2.7)$$

Then for the sequence $\{u_k\}$, generated by OSR-Method 4.2.1, it holds

(i) $\|u_k\| < r^*$;

(ii) $u_k \rightharpoonup u^* \in U^*$.

If, additionally, Assumption 4.2.3 is fulfilled, then
4.2. ONE-STEP REGULARIZATION METHODS

(iii) \( \lim_{k \to \infty} \| u^k - u^* \| = 0. \)

Proof: We deliver the proof in three phases.

Phase 1: Let \( u^{**} \in U^* \cap \mathbb{B}_{r/8} \) be fixed and \( \tilde{u}^k := \arg \min_{u \in Q_k} \Psi_k(u) \). Due to \( u^{**} \in Q^* \) and condition (4.2.6), we can choose \( v^k \in Q_k, \tilde{v}^k \in Q \) such that

\[
\| v^k - u^{**} \| \leq \mu_k, \quad \| \tilde{v}^k - \tilde{u}^k \| \leq \mu_k, \quad k = 1, 2, \ldots . \tag{4.2.8}
\]

In view of (4.1.3) and (4.2.8),

\[
J(v^k) - J(u^{**}) \leq L(r) \mu_k,
\]

hold true, and with regard to \( J(u^{**}) \leq J(\tilde{v}^k) \), we conclude that

\[
J(v^k) - J(\tilde{u}^k) \leq 2\mu(r) \mu_k
\]

and, due to (4.2.5),

\[
J_k(v^k) - J_k(\tilde{u}^k) \leq 2L(r) \mu_k + 2\sigma_k, \quad k = 1, 2, \ldots . \tag{4.2.9}
\]

Using Proposition 3.1.3 with \( v^0 := u^{k-1}, y := v^k, f := J_k, \chi := 2 \) and \( C := Q_k \), we obtain

\[
\| \tilde{u}^k - v^k \| \leq \| u^{k-1} - v^k \| + \sqrt{2L(r) \mu_k + 2\sigma_k} \tag{4.2.10}
\]

and with (4.2.8) this leads to

\[
\| \tilde{u}^k - u^{**} \| \leq \| u^{k-1} - u^{**} \| + \sqrt{2L(r) \mu_k + 2\sigma_k + 2\mu_k}. \tag{4.2.11}
\]

For \( k := 1 \), with regard to (4.2.7), (4.2.10) and the chosen \( u^0 \), the estimate \( \| \tilde{u}^1 \| \leq r^* \) holds. Consequently, \( \tilde{u}^1 \in \text{int} \mathbb{B}_r \), and from the strong convexity of \( \Psi_1 \) and the convexity of \( K_1 \), one can conclude that

\[
\tilde{u}^1 = \bar{u}^1 := \arg \min_{u \in K_1} \Psi_1(u).
\]

But on account of (4.2.3), we get

\[
\| u^k - \bar{u}^k \| \leq \frac{\epsilon_k}{2}, \quad \forall \, k, \tag{4.2.12}
\]

and (4.2.11) yields

\[
\| u^1 - u^{**} \| \leq \| u^0 - u^{**} \| + \sqrt{2L(r) \mu_1 + 2\sigma_1 + 2\mu_1 + \frac{\epsilon_1}{2}},
\]

hence, \( \| u^1 \| \leq r^* \).

Phase 2: Suppose now that the inequalities

\[
\| \tilde{u}^k \| < r^*, \quad \| u^k \| < r^*
\]

are true for \( k = 1, \ldots, s - 1 \). Then, we have \( \tilde{u}^k = u^k \) and from (4.2.8), (4.2.10) and (4.2.11) it follows that for \( k = 1, \ldots, s - 1 \)

\[
\| u^k - u^{**} \| \leq \| u^{k-1} - u^{**} \| + \sqrt{2L(r) \mu_k + 2\sigma_k + 2\mu_k},
\]

\[
\| u^k - u^{**} \| \leq \| u^{k-1} - u^{**} \| + \sqrt{2L(r) \mu_k + 2\sigma_k + 2\mu_k + \frac{\epsilon_k}{2}}. \tag{4.2.13}
\]

\[\]
CHAPTER 4. PPR FOR INFINITE-DIMENSIONAL PROBLEMS

Summing up (4.2.13) for \( k = 1, \ldots, s - 1 \) and taking into account that
\[
\|\tilde{u}^s - u^{**}\| \leq \|u^{s-1} - u^{**}\| + \sqrt{2L(r)\mu_s + 2\sigma_s + 2\mu_s},
\]
then
\[
\|\tilde{u}^s - u^{**}\| \leq \|u^0 - u^{**}\| + \sum_{k=1}^{s} \left(\sqrt{2L(r)\mu_k + 2\sigma_k + 2\mu_k}\right) + \frac{1}{2}\sum_{k=1}^{s-1} \epsilon_k.
\]

Together with (4.2.7) this proves that \( \|\tilde{u}^s\| < r^* \).

Consequently, \( \tilde{u}^s = \bar{u}^s \) and observing (4.2.12), we finally get
\[
\|u^s - u^{**}\| \leq \|u^0 - u^{**}\| + \sum_{k=1}^{s} \left(\sqrt{2L(r)\mu_k + 2\sigma_k + 2\mu_k}\right) + \frac{1}{2}\sum_{k=1}^{s-1} \epsilon_k
\]
and \( \|u^s\| < r^* \).

Phase 3: For an arbitrary element \( w \in Q^* := U^* \cap B_{r^*} \) and elements \( v^k \in Q_k, \bar{v}^k \in Q \), chosen such that
\[
\|v^k - w\| \leq \mu_k, \quad \|\bar{v}^k - \bar{u}^k\| \leq \mu_k, \quad k = 1, 2, \ldots,
\]
we obtain analogously to (4.2.9) the estimate
\[
J_k(v^k) - J_k(\bar{u}^k) \leq 2L(r)\mu_k + 2\sigma_k.
\]

Applying again Proposition 3.1.3, but with the new data
\[
v^0 := u^{k-1}, \quad y := v^k, \quad f := J_k, \quad \chi := 2, \quad C := Q_k,
\]
one can conclude that
\[
\|\bar{u}^k - v^k\|^2 \leq \|u^{k-1} - v^k\|^2 + J_k(v^k) - J_k(\bar{u}^k),
\]
and, similarly to (4.2.11)
\[
\|\bar{u}^k - w\| \leq \|u^{k-1} - w\| + \sqrt{2L(r)\mu_k + 2\sigma_k + 2\mu_k}.
\]
In view of (4.2.12), the latter inequality provides
\[
\|u^k - w\| \leq \|u^{k-1} - w\| + \sqrt{2L(r)\mu_k + 2\sigma_k + 2\mu_k + \frac{\epsilon_k}{2}}, \quad k = 1, 2, \ldots.
\]

Therefore, condition (4.2.7) and Lemma A3.1.4 imply that \( \{\|u^k - w\|\} \) converges for any \( w \in Q^* \) and with regard to (4.2.12), \( \{\|\bar{u}^k - w\|\} \) converges to the same limit.

Due to the choice of the element \( v^k \) the inequalities
\[
\|u^{k-1} - v^k\|^2 \leq \left(\|u^{k-1} - w\| + \mu_k\right)^2
\]
\[
\|\bar{u}^k - w\|^2 \leq \left(\|\bar{u}^k - v^k\| + \mu_k\right)^2
\]
are true. Together with (4.2.15) and the estimates
\[
\|\bar{u}^k\| < r, \quad \|u^{k-1}\| < r, \quad \|v^k\| \leq r, \quad \|w\| \leq r,
\]
this leads to
\[ \|u^{k-1} - w\|^2 - \|u^k - w\|^2 \geq J_k(\tilde{u}^k) - J_k(v^k) - 8r\mu_k - 2\mu_k^2, \]
and because of (4.2.5) and (4.1.3),
\[ \|u^{k-1} - w\|^2 - \|\tilde{u}^k - w\|^2 \geq J(\tilde{u}^k) - J(w) - (L(r) + 8r)\mu_k - 2\mu_k^2 - 2\sigma_k. \quad (4.2.16) \]

As shown in phase 2 of the proof, the sequence \( \{\tilde{u}^k\} \) is bounded. Hence, choosing in (4.2.16) a subsequence \( \{\tilde{u}^{k_j}\} \rightarrow \tilde{u} \), and taking limit (for \( j \to \infty \)), then in view of the weak lower-semicontinuity of the functional \( J \) and the conditions
\[ \lim_{k \to \infty} \sigma_k = \lim_{k \to \infty} \mu_k = 0, \]
we conclude that
\[ J(\tilde{u}) - J(w) \leq 0. \]

But on account of (4.2.6), (4.2.7) and the convexity of \( K \cap B_r \), the limit point \( \tilde{u} \) belongs to \( K \cap B_r \), therefore \( \tilde{u} \in Q^* := U^* \cap B_r \). Setting \( \tilde{u} := u^* \), Opial’s Lemma A1.1.3 shows that \( \{u^k\} \to u^* \), which implies \( u^k \to u^* \), too.

If, additionally, Assumption 4.2.3 is satisfied, then strong convergence of \( \{u^k\} \to u^* \) holds due to Lemma 4.1.4. \( \square \)

4.2.5 Remark. If OSR-Method 4.2.1 is executed with two different starting points \( u^0 \) and \( v^0 \), then in view of the non-expansivity of the prox-mapping the following relation holds between the corresponding limit points \( u^* \) and \( v^* \):
\[ \|u^* - v^*\| \leq \|u^0 - v^0\| + \sum_{k=1}^{\infty} \epsilon_k. \]
\( \diamond \)

4.2.6 Remark. In order to satisfy the assumptions of Theorem 4.2.4 it is sufficient to choose the sequences \( \{\mu_k\} \) and \( \{\sigma_k\} \) such that
\[ \mu_k \leq (c_1 + k)^{-2-c_2}, \quad \sigma_k \leq (c_1 + k)^{-1-c_2}, \quad k = 1, 2, \cdots, \]
where \( c_2 > 0 \) is an arbitrary small number and \( c_1 > 0 \) a constant. For special problems and methods considered in this book later on these conditions for updating the parameters are not too restrictive if \( c \) is not chosen too large. The value \( c \) can be tuned by means of variations of the radius \( r \) of the ball \( B_r \) or suitable scalings of the objective functional \( J \). However, these transformations may lead to an increase of the external iterations.

The choice of the controlling parameters will be discussed in more detail in connection with specific variational problems considered in the following chapters, see in particular Subsections 4.4 and 9.1.3. \( \diamond \)

It should be noted that the statement of Theorem 4.2.4 remains true (with evident modifications in the proof) if one takes \( \sigma_k \equiv 0 \), i.e.,
\[ \Psi_k(\cdot) \equiv \hat{\Psi}_k(\cdot) := J(\cdot) + \|u^k - u^{k-1}\|^2, \quad \forall k \]
and if, moreover, instead of (4.2.3) the condition
\[ \inf_{q(u^k) \in \partial \Psi_k(u^k)} \|q(u^k) - q(\tilde{u}^k)\|_{V'} \leq \epsilon_k \]
is used with some \( q(u^k) \in \partial \Psi_k(u^k), q(\bar{u}^k) \in \partial \Psi_k(\bar{u}^k) \).

However, considering the case that \( J \) is non-differentiable and its approximation cannot be performed sufficiently smooth, the following OSR-method without smoothing may be more convenient:

\[
\text{Generate } \{u^k\} \text{ by }
\begin{align*}
\bar{\Psi}_k(u^k) &\leq \min_{u \in K_k} \bar{\Psi}_k(u) + \epsilon_k, \quad (4.2.17) \\
\rho(u^k, K_k) &\leq \epsilon_k. \quad (4.2.18)
\end{align*}
\]

In that case we maintain the notations used for the description of OSR-Method 4.2.1, but modify only the conditions concerning the choice of the radii \( r^* \) and \( r \) such that

\[
U^* \cap B_{r^*/4} \neq \emptyset, \quad r \geq 2r^* + 2 \sup_k \epsilon_k \quad (4.2.19)
\]

and

\[
(r^*)^2 - (2(L(2r^*) + 8r^*) + 1) \sup_k \epsilon_k > 0, \quad (4.2.20)
\]

with \( L(2r^*) \) the Lipschitz constant of the functional \( J \) on \( B_{2r^*} \).

Further, denote \( u^k_Q := \arg\min_{u \in Q} \bar{\Psi}_k(u) \) and \( \hat{Q} := \bigcup_{k=1}^{\infty} \{u^k_Q\} \).

**4.2.7 Theorem.** Assume that \( J \) fulfils the Lipschitz condition (4.1.3) on \( B_r \), that the estimates (4.2.6) as well as \( \rho(\hat{Q}, Q_k) \leq \mu_k \) are satisfied, and moreover, that \( u^0 \in B_{r^*/4} \) and

\[
\sum_{k=1}^{\infty} \left( \sqrt{2L(r)} \mu_k + 2\mu_k + \sqrt{(1 + L(r) + 4r)\epsilon_k + \epsilon_k} \right) < \frac{r^*}{2}. \quad (4.2.21)
\]

Then the sequence \( \{u^k\} \), generated by (4.2.17)-(4.2.18), converges weakly to some solution \( u^* \) of Problem (4.2.1).

If, additionally, Assumption 4.2.3 holds, then \( \lim_{k \to \infty} \|u^k - u^*\| = 0 \).

**Proof:** Denote

\[
\bar{u}^k := \arg\min_{u \in K_k} \bar{\Psi}_k(u), \quad w^k := \arg\min_{v \in K_k} \|v - u^k\|
\]

and assume that

\[
\|\bar{u}^k\| < r^*, \quad \|u^{k-1}\| < r^*, \quad \forall \ k \leq s.
\]

If \( s = 1 \), the second inequality is obvious and the first one was shown in phase 1 of the proof of Theorem 4.2.4.

If \( u^* \notin B_{r^*} \) with \( r' := 2r^* + \sup_k \epsilon_k \), then due to (4.2.18) we get \( w^s \notin B_{2r^*} \). For the elements

\[
\hat{w}^s \in [\bar{u}^s, w^s] \cap \partial B_{2r^*}, \quad \check{u}^s \in [\check{u}^s, u^s] \cap \partial B_{2r^*}.
\]
the inequality
\[ \| \hat{w}^s - \bar{u}^s \| \leq 2\epsilon_s \]  (4.2.22)
can be concluded by means of elementary geometrical observations, considering
the pair of triangles with vertices \((\bar{u}^s, u^s, w^s)\) and \((\bar{u}^s, \hat{u}^s, \hat{w}^s)\), respectively, and
choosing \(\epsilon_s < \frac{r^*}{16}\), which follows trivially from (4.2.20).

From (4.1.1), (4.1.3) and (4.2.22) one obtains
\[ \bar{\Psi}_s(\hat{u}^s) \geq \bar{\Psi}_s(\hat{w}^s) - 2(L(2r^*) + 8r^*)\epsilon_s. \]  (4.2.23)

But, regarding the strong convexity of \(\bar{\Psi}_s\), the inequality
\[ \bar{\Psi}_s(\hat{w}^s) - \bar{\Psi}_s(\bar{u}^s) \geq \langle q_s(\bar{u}^s), \hat{w}^s - \bar{u}^s \rangle + \| \hat{w}^s - \bar{u}^s \|^2 \]  (4.2.24)
is true for any subgradient \(q_s(\bar{u}^s) \in \partial\bar{\Psi}_s(\bar{u}^s)\). Moreover, by definition of \(\bar{u}^s\), it
holds for some \(q_s(\bar{u}^s) \langle q_s(\bar{u}^s), \hat{w}^s - \bar{u}^s \rangle \geq 0\).

Since \(\hat{w}^s \in \partial B_{2r^*}\) and \(\bar{u}^s \in B_{r^*}\), we infer that
\[ \bar{\Psi}_s(\hat{w}^s) - \bar{\Psi}_s(\bar{u}^s) \geq \| \hat{w}^s - \bar{u}^s \|^2 \geq (r^*)^2. \]

Now, (4.2.23) and (4.2.20) yield
\[ \bar{\Psi}_s(\hat{u}^s) \geq \bar{\Psi}_s(\hat{w}^s) + (r^*)^2 - 2(L(2r^*) + 8r^*)\epsilon_s > \bar{\Psi}_s(\hat{u}^s) + \epsilon_s. \]

Because \(u^s = \hat{u}^s + \lambda(\hat{u}^s - \bar{u}^s)\) for some \(\lambda > 0\), we conclude from the convexity
of \(\bar{\Psi}_s\) and the latter inequality that
\[ \bar{\Psi}_s(u^s) > \bar{\Psi}_s(\hat{u}^s) + \epsilon_s, \]
which contradicts condition (4.2.17).

Hence, \(u^s \in \mathbb{B}_{r^*} \subset \mathbb{B}_r\), and due to (4.2.18), this implies
\[ w^s \in \mathbb{B}_r. \]  (4.2.25)

From (4.2.17) it is obvious that
\[ \bar{\Psi}_k(u^k) \leq \min_{u \in Q_k} \bar{\Psi}_k(u) + \epsilon_k, \quad \forall \ k \in \mathbb{N}. \]  (4.2.26)

Furthermore, the relations (4.2.18), (4.2.25) and \(\| u^k \| < r^*\) (for \(k < s\)) lead to
\[ \rho(u^k, Q_k) \leq \epsilon_k, \quad k = 1, \ldots, s. \]  (4.2.27)

Taking into account (4.1.3), (4.2.26) and \(\| u^k - w^k \| \leq \epsilon_k\), we infer that
\[ \bar{\Psi}_k(u^k) \leq \min_{u \in Q_k} \bar{\Psi}_k(u) + (1 + L(r) + 4r)\epsilon_k. \]

Hence, in view of the strong convexity of \(\bar{\Psi}_k\) and (4.2.27), the inequalities
\[ \| w^k - \bar{u}^k \| \leq \sqrt{(1 + L(r) + 4r)\epsilon_k}, \]
\[ \| u^k - \bar{u}^k \| \leq \sqrt{(1 + L(r) + 4r)\epsilon_k + \epsilon_k} \]  (4.2.28)
hold for \( k = 1, \ldots, s \). Now, formula (4.2.11) can be applied with 
\[
u^{**} \in U^* \cap \mathbb{B}_{r^*/8}, \quad \tilde{u}^k := \arg\min_{u \in Q_k} \Psi(u), \quad \sigma_k := 0
\]
(note that inequality (4.2.11) is true regardless of how \( u^{k-1} \) has been calculated).

According to \( \tilde{u}^k = \bar{u}^k \) for \( k \leq s \), we get
\[
\| \bar{u}^k - u^{**} \| \leq \| u^{k-1} - u^{**} \| + \sqrt{2L(r)\mu_k + 2\mu_k}
\]
and due to (4.2.28),
\[
\| u^k - u^{**} \| \leq \| u^{k-1} - u^{**} \| + \theta_k, \quad \forall k = 1, \ldots, s \tag{4.2.29}
\]
with
\[
\theta_k := \sqrt{2L(r)\mu_k + 2\mu_k + \sqrt{1 - L(r) + 4r}}\epsilon_k + \epsilon_k.
\]
Summing up the latter inequalities, this leads to
\[
\| u^s - u^{**} \| \leq \| u^0 - u^{**} \| + s \sum_{k=1}^s \theta_k. \tag{4.2.30}
\]
Hence, in view of the choice of \( u^0, u^{**} \) and (4.2.21), we conclude that \( \| u^s \| < r^* \).
Moreover, from (4.2.11) (with \( k := s + 1 \)) and (4.2.30) it follows
\[
\| \tilde{u}^{s+1} - u^{**} \| \leq \| u^0 - u^{**} \| + \sum_{k=1}^{s+1} \theta_k,
\]
and as before this yields \( \| \tilde{u}^{s+1} \| < r^* \), i.e.,
\[
\tilde{u}^{s+1} = \bar{u}^{s+1}, \quad \| \tilde{u}^{s+1} \| < r^*.
\]
Therefore, the estimates
\[
\| u^k \| < r^*, \quad \| \tilde{u}^k \| < r^*, \quad \forall k \in \mathbb{N}
\]
are true and we have shown that
\[
\rho(u^k, Q_k) = \rho(u^k, K_k), \quad \forall k.
\]
Now, due to (4.2.26), (4.2.27), (4.2.6) and \( \rho(\hat{Q}, Q_k) \leq \mu_k \), it follows
\[
\Psi_k(u^k) \leq \min_{u \in Q} \Psi_k(u) + (L(r) + 4r)\mu_k + \epsilon_k,
\]
\[
\rho(u^k, Q) \leq \epsilon_k + \mu_k,
\]
and all we have to do is to use Lemma 4.1.3. \( \Box \)

As mentioned before, this variant of an OSR-method may be convenient in case methods of non-smooth optimization solve efficiently Problem (4.2.1).
4.2. ONE-STEP REGULARIZATION METHODS

4.2.2 Modifications of OSR-methods

Now, for Problem (4.2.1) we are going to describe modifications of the OSR-methods which make use of a different approximation of the original objective function $J$ as it was suggested in OSR-Method 4.2.1. This leads to weaker conditions for the choice of the controlling parameters.

Let the sequences $\{J_k\}$ and $\{K_k\}$ satisfy the same assumptions as made at the beginning of Section 4.2.

For fixed $r_1 > 0$ denote

$$\delta(u) := \rho^2(u, \mathbb{B}_{r_1})$$

and

$$\Phi_k(u) := J_k(u) + \delta(u) + \|u - u^{k-1}\|^2.$$

Now, we are dealing with the following family of auxiliary problems

$$\min_{u \in K_k} \Phi_k(u),$$

(4.2.31)

4.2.8 Method. (Modified OSR-Method)

Data: $u^0 \in V$, $\{\epsilon_k\} \downarrow 0$;

S0: Set $k := 0$;

S1: Compute for Problem (4.2.31) an approximate solution $u^k$ such that

$$\|\nabla \Phi_k(u^k) - \nabla \Phi_k(\bar{u}^k)\| \leq \epsilon_k,$$

(4.2.32)

where

$$\bar{u}^k := \arg\min_{u \in K_k} \Phi_k(u).$$

(4.2.33)

S2: set $k := k + 1$ and go to S1.

♦

To prove convergence of this method we need the following assumptions and notations. Suppose that the functional $J$ satisfies the Lipschitz condition (4.1.3) on $\mathbb{B}_{r_1}$. Further, for given $\theta > 0$ choose the radius $r$ such that

$$r \geq 2r_1 + 2L(r_1) + 2\sqrt{\theta} + L(r_1)(4r_1 + L(r_1))$$

(4.2.34)

and denote

$$Q := K \cap \mathbb{B}_r, \quad Q^* := U^* \cap \mathbb{B}_r, \quad Q_k := K_k \cap \mathbb{B}_r, \quad \forall \ k,$$

$$\bar{\Phi}_k(u) := J(u) + \delta(u) + \|u - u^{k-1}\|^2.$$

Again, we need some quantitative information about the exactness of the approximations of $J$ and $K$.

4.2.9 Assumption. For given sequences $\{\sigma_k\} \downarrow 0$ and $\{\mu_k\} \downarrow 0$ and for each $k = 1, 2, \cdots$ it holds

$$\sup_{u \in \mathbb{B}_r} |J(u) - J_k(u)| \leq \sigma_k,$$

(4.2.35)

$$\rho(Q_k, Q_{k+1}) \leq \mu_k, \quad \rho(Q_k, Q) \leq \mu_k, \quad \rho(Q^*, Q_k) \leq \mu_k.$$

(4.2.36)
4.2.10 Theorem. Suppose parameters $r_1$, $r$ and $\theta$ are chosen such that $U^* \cap \mathbb{B}_{r_1} \neq \emptyset$, relation (4.2.34) holds and $J$ satisfies the Lipschitz condition (4.1.3) on $\mathbb{B}_r$. Moreover, let Assumption 4.2.9 be fulfilled and

\[
\sum_{k=1}^{\infty} \sqrt{\mu_k} < \infty, \quad \sum_{k=1}^{\infty} \sqrt{\sigma_k} < \infty, \quad \sum_{k=1}^{\infty} \epsilon_k < \infty, \tag{4.2.37}
\]

\[
\sum_{k=1}^{\infty} \big( (L(r) + 6r)\mu_k + \frac{\epsilon_k^2}{4} + 2\sigma_k \big) < \theta. \tag{4.2.38}
\]

Then sequence $\{u^k\}$, starting with $u^0 \in \text{int}\mathbb{B}_{r_1} \cap K_1$ and generated by Method 4.2.8, converges weakly to some $u^* \in U^*$. If, moreover, Assumption 4.2.3 is satisfied, then $\lim_{k \to \infty} \|u^k - u^*\| = 0$.

Proof: Let $w^k := \arg \min \{\Phi_k(u) : u \in Q_{k-1}\}$ and set $Q_0 := Q_1$, $\mu_0 := \mu_1$. In view of (4.2.36) an element $v^k \in Q_k$ can be chosen with $\|v^k - u^k\| \leq \mu_{k-1}$. The Lipschitz properties of the functionals $J$ and $\delta$ (cf. (4.1.2)) imply

\[
|J(v^k) - J(w^k)| \leq L(r)\mu_{k-1},
\]

\[
|\delta(v^k) - \delta(w^k)| \leq 2r\mu_{k-1},
\]

and if $\|u^{k-1}\| \leq r$, we get in view of (4.1.1)

\[
\|v^k - u^{k-1}\|^2 - \|w^k - u^{k-1}\|^2 \leq 4r\mu_{k-1}.
\]

Hence, if $\|u^{k-1}\| \leq r$, then

\[
|\Phi_k(v^k) - \Phi_k(w^k)| \leq (L(r) + 6r)\mu_{k-1}. \tag{4.2.39}
\]

Now, assume that for $k \leq s$ the inequality

\[
J(\bar{u}^k) + \delta(\bar{u}^k) \leq J(u^0) + (L(r) + 6r)\sum_{j=1}^{k-1} \mu_j + \frac{1}{4} \sum_{j=1}^{k-1} \epsilon_j^2 + 2\sum_{j=1}^{k} \sigma_j \tag{4.2.40}
\]

is satisfied (for $k = 1$ this is trivial because

\[
J(\bar{u}^1) + \delta(\bar{u}^1) \leq \min_{u \in K_1} \{J_1(u) + \delta(u) + \|u - u^0\|^2\} + \sigma_1
\]

\[
\leq J_1(u^0) + \delta(u^0) + \sigma_1 \leq J(u^0) + \delta(u^0) + 2\sigma_1
\]

and, due to the choice of $u^0$, it holds $\delta(u^0) = 0$).

Then, by means of the relations (4.2.38), (4.2.40) and Lemma 4.1.1, we can infer that $\|\bar{u}^s\| < \frac{s}{r}$, meaning that $\bar{u}^s \in Q_s$.

The definition of $w^k$ implies

\[
\Phi_{s+1}(w^{s+1}) \leq J(\bar{u}^s) + \delta(\bar{u}^s) + \|\bar{u}^s - u^*\|^2
\]

and with regard to (4.2.39)

\[
\Phi_{s+1}(v^{s+1}) \leq J(\bar{u}^s) + \delta(\bar{u}^s) + \|\bar{u}^s - u^*\|^2 + (L(r) + 6r)\mu_s. \tag{4.2.41}
\]

But

\[
\|\bar{u}^s - u^*\| \leq \frac{\epsilon_k}{2}, \quad \forall k, \tag{4.2.42}
\]
and taking into account (4.2.40), (4.2.41) and the trivial inequality
\[ \Phi_{s+1}(u^{s+1}) \geq J(u^{s+1}) + \delta(u^{s+1}) - 2\sigma_{s+1}, \]
we get
\[ J(u^{s+1}) + \delta(u^{s+1}) \leq J(u^0) + (L(r) + 6r)\mu_s + 2\sigma_{s+1} \leq J(u^0) + (L(r) + 6r)\mu_s + 2\sigma_{s+1} \]
This induction allows to assert that \( \|u^k\| < \frac{r}{2} \) holds for all \( k \).

Due to (4.2.38), we have \( \frac{\mu_k}{2} < \sqrt{\theta} \), hence, the relations (4.2.34) and (4.2.42) yield that
\[ \epsilon_k < r, \quad \|u^k - u^k\| < \frac{r}{2} \]
and that the estimate \( \|u^k\| < r \) is satisfied for all \( k \in \mathbb{N} \).

Now we make use of Theorem 4.2.4, because the modified OSR-Method 4.2.8 can be interpreted as OSR-Method 4.2.1 applied to the problem
\[ \min\{ J(u) + \delta(u) : u \in K \}. \]

Then in the respective part of the proof of Theorem 4.2.4, instead of relation (4.2.7) only the conditions (4.2.37) have to be used.

In order to find suitable values \( \mu_k \) according to the Theorems 4.2.4, 4.2.7 or Theorem 4.2.10, we realize that a sufficient accuracy of the successive approximation of the set \( Q \) is required, and this is the crucial point in the realization of this approach.

Of course, we do not assume that the minimizers \( u^k_Q := \arg\min_{u \in Q} \Psi_k(u) \) and the sets \( Q^k \) are known. For the elements \( u^* \in U^* \) and \( u^*_Q \) we only have to know some information about their smoothness (or regularity) which enables us to estimate the distance \( \rho(Q, Q_k) \). This will be seen in Chapter 8 when specific variational inequalities in Sobolev spaces are investigated.

In order to generate a sequence \( \{u^k\} \), condition (4.2.3) (but as well the conditions (4.2.17), (4.2.18) or (4.2.32)) has to be translated into a practicable stopping rule according to the properties of the problems and the chosen basic algorithms for solving the discretized and regularized auxiliary problems.

It should be noted that the assumptions of the Theorems 4.2.4 and 4.2.10 do not guarantee solvability of the approximate, non-regularized problems
\[ \min_{u \in K_k} J_k(u) \quad \text{or} \quad \min_{u \in K_k} J(u). \]

This fact will be illustrated in Example 6.2.1 below in context with the approximation of semi-infinite problems. Their corresponding approximate problems have no solution at all regardless of which approximation level \( k \) is used. However, the non-regularized problems
\[ \min_{u \in K_k} \{ J_k(u) + \delta(u) \}, \]
corresponding to the modified OSR-method just before, are solvable in any case.
4.3 Multi-Step Regularization Methods

We describe now a coupled (diagonal) scheme of approximation and iterative PPR applied to Problem (4.2.1), where on the $k$-th approximation level (external loop) the proximal method is applied for solving

$$\min \{ J_k(u) : u \in K_k \}.$$ 

On each external approximation level $k$ the proximal iterations are performed as long as they prove to be sufficiently productive (internal loop), i.e., as long as the distance between the successive iterates are not going below a given threshold.

In comparison with OSR-methods this approach permits one to handle more effectively coarse approximations on the external approximation levels, because one gets on such levels better approximations of the solution of the original problem.

4.3.1 Description of MSR-methods

As before, let $\{ J_k \}, k = 1, 2, \cdots,$ be a sequence of convex and Gâteaux-differentiable functionals on $V$ and $\{ K_k \}, k \in \mathbb{N}$, be a sequence of convex and closed subsets of $V$ which approximate $J$ and $K$, respectively.

In order to obtain a solution of the original Problem (4.2.1) a sequence

$$\{ u^{k,i} \}, \ (i = 0, 1, \ldots, i(k), \ k = 1, 2, \ldots)$$

is generated. The termination index $i(k)$ at the internal loop will be determined within the process and the choice of the starting point $u^{1,0}$ is explained more precisely in Lemma 4.3.5 below.

4.3.1 Method. (Multi-step regularization)

Given $u^0$ and sequences $\{ \delta_k \}, \{ \epsilon_k \}, k = 1, 2, \cdots$, such that

$$\delta_k > 0, \quad \epsilon_k \geq 0, \quad \lim_{k \to \infty} \epsilon_k = 0.$$

Step $k$: Given $u^{k-1}$.

(a) Set $u^{k,0} := u^{k-1}, \ i := 1$.

(b) Given $u^{k,i-1}$ and

$$\Psi_{k,i}(u) := J_k(u) + \| u - u^{k,i-1} \|^2, \quad (4.3.1)$$

compute approximately $u^{k,i}$ according to

$$\| \nabla \Psi_{k,i}(u^{k,i}) - \nabla \Psi_{k,i}(\bar{u}^{k,i}) \|_{V'} \leq \epsilon_k \quad (4.3.2)$$

with

$$\bar{u}^{k,i} := \arg \min \{ \Psi_{k,i}(u) : u \in K_k \}. \quad (4.3.3)$$

(c) If $\| u^{k,i} - u^{k,i-1} \| > \delta_k$, set $i := i + 1$ and repeat (b).

Otherwise, set $u^k := u^{k,i}, \ i(k) := i$, and continue with Step $k := k + 1$. 
Before proving convergence of MSR-Method 4.3.1 we give a simple example warning against a careless choice of the controlling parameters appearing in the method.

4.3.2 Example. Function $J(u) := u_2$ has to be minimized on $K := \{u = (u_1, u_2)^T : 0 \leq u_1 \leq 1, 0 \leq u_2 \leq 1\}$.

Let $J_k := J$ for all $k$ and the sets $K_k$ be defined such that ($s \in \mathbb{Z}_+$):

- for $k = 3s - 2$:
  $K_k := \{u : 0 \leq u_1 \leq 1, u_2 \leq 1, u_1 + \alpha_{s-1} u_2 \geq 1\}$;

- for $k = 3s - 1$:
  $K_k := \{u : 0 \leq u_1 \leq 1, u_2 \leq 1, u_1 - \alpha_{s-1} u_2 \leq \frac{\alpha^2_{s-2}}{\alpha^2_{s-1} + 1} \text{ if } u_1 \geq \frac{1}{2},$
  $u_2 \geq \frac{1}{2\alpha_{s-1}} - \frac{1}{\alpha^2_{s-1} + 1} \text{ if } u_1 \leq \frac{1}{2}\}$;

- for $k = 3s$:
  $K_k := \{u : 0 \leq u_1 \leq 1, \frac{1}{2\alpha_{s-1}} \leq u_2 \leq 1\}$;

where $\alpha_0 > 1$ is fixed and $\alpha_s$ satisfies for $s \geq 1$ the condition

$$\alpha_s \geq \frac{\alpha^2_{s-1} + 1}{(\alpha_{s-1} - 1)^2}.$$

In Figur 4.3.1 the first three sets $K_k$ are drawn.

![Figure 4.3.1](image)

It is easily seen that, increasing $\alpha_s$ suitably, the approximation of $K$ by means of $\{K_k\}$ can be performed arbitrarily fast, but also extremely slow, if

$$\alpha_s = \alpha_{s-1} \frac{\alpha^2_{s-1} + 1}{(\alpha_{s-1} - 1)^2}.$$

In each case $\lim_{s \to \infty} \alpha_s = \infty$, hence, $\rho(K, K_k) \to 0$ if $k \to \infty$.

Assume now that $\epsilon_k = 0 \forall k$, i.e., the regularized auxiliary problems

$$\min \{\Psi_k(u) : u \in K_k\}$$
will be solved exactly. Then a straightforward calculation shows the following:

Starting with \( u^0 := (\alpha, \frac{1}{\alpha})^T \) and sequence \( \{\delta_k\} \) chosen such that

\[
\delta_{3s-2} \geq \sqrt{\frac{\alpha_1^2}{4(\alpha_{s-1}^2 + 1)^2} + \frac{1}{4(\alpha_{s-1}^2 + 1)^2}} := \bar{\delta},
\]
\[
\delta_{3s-1} > 0 \text{ (arbitrarily)},
\]
\[
\delta_{3s} > 0 \text{ (arbitrarily)},
\]
we get \( i(3s-2) = 1, i(3s-1) = 1 \) or \( 2, i(3s) = 1 \) or \( 2 \), and

\[
\begin{align*}
\delta_{3s-2} & \geq \frac{1}{2} + \frac{\alpha_{s-1}}{2(\alpha_{s-1}^2 + 1)}, \quad \delta_{3s-1} > 0, \\
\delta_{3s} & > 0.
\end{align*}
\]

\[
\begin{align*}
\{u^0, u^1, \ldots\} & \text{ fulfilling Assumption 4.2.2.}
\end{align*}
\]

Next we show under which conditions the internal loop of a MSR-method \( \{u^{k,i}\} \) has two limit points \( (\frac{1}{2}, 0)^T \) and \( (1, 0)^T \).

\[ u^{3s-2,1} = \left( 1 + \frac{1}{2} + \frac{1}{2\alpha_{s-1}} - \frac{1}{\alpha_{s-1}^2 + 1} \right)^T, \]
\[
\begin{align*}
u^{3s-1,i(3s-1)} & = \left( \frac{1}{2} + \frac{1}{2\alpha_{s-1}} - \frac{1}{\alpha_{s-1}^2 + 1} \right)^T, \\
\{u^{3s,i(3s)} = \left( \frac{1}{2} + \frac{1}{2\alpha_{s-1}} \right)^T, \text{ for } i(3s) = 1 \text{ or } 2 \).
\end{align*}
\]

Hence, \( \{u^{k,i}\} \) converges to the point \( u^* = (\frac{1}{2}, 0) \in U^* \).

But, if \( \max[\delta_{3s-2}, \delta_{3s-1}] < \bar{\delta} \) and \( \delta_{3s} > 0 \) is chosen arbitrarily, then

\[
\begin{align*}
u^{3s-2,1} & = (1, 0)^T, \\
\{u^{3s-1,i(3s-1)} & = \left( \frac{1}{2} + \frac{1}{2\alpha_{s-1}} - \frac{1}{\alpha_{s-1}^2 + 1} \right)^T, \\
\{u^{3s,i(3s)} = \left( \frac{1}{2} + \frac{1}{2\alpha_{s-1}} \right)^T, \text{ for } i(3s) = 1 \text{ or } 2 \).
\end{align*}
\]

\[ \Box \]

### 4.3.2 Convergence of the MSR-scheme

Now, we are going to prove convergence of MSR-Method 4.3.1 applied to Problem (4.2.1). Again at first we start with some notations.

For a fixed radius \( r^* > 0 \) and \( r \geq r^* \) let

\[
\begin{align*}
Q & := K \cap B_r, \\
Q_k & := K \cap B_{r^*}, \forall \ k \in \mathbb{N}; \\
\bar{\Psi}_{k,i}(u) & := J(u) + \|u - u^{k,i-1}\|^2, \\
u^{k,i}_{Q} := \arg\min_{u \in Q} \bar{\Psi}_{k,i}(u); \\
\hat{Q} & := \bigcup_{k=1}^{\infty} \bigcup_{i=1}^{\infty} \{u^{k,i}_{Q}\}, \\
Q^* & := (U^* \cap B_r) \cup (\hat{Q} \cap B_r).
\end{align*}
\]

Working with Assumption 4.2.2 in the case of MSR-methods, we suppose that \( Q^* \) appearing there is defined by (4.3.4).

Next we show under which conditions the internal loop of a MSR-method terminates after a finite number of steps.

### 4.3.3 Lemma. Suppose \( r^* \) is fixed such that \( U^* \cap B_r \neq \emptyset \), \( r \geq r^* \), functional \( J \) is Lipschitz on \( B_r \) according to (4.1.3), and Assumption 4.2.2 is fulfilled.

Moreover, controlling parameters \( \epsilon_k, \delta_k, \mu_k \) and \( \sigma_k \) let be chosen such that

\[ \frac{1}{4r} \left( 2L(r) \mu_k + 2\sigma_k - (\delta_k - \frac{\epsilon_k}{2})^2 \right) + \frac{\epsilon_k}{2} < 0, \]
\[ \tilde{\epsilon}_k := \delta_k - \frac{\epsilon_k}{2} > 0. \]
If, for fixed $k_0$, the iterates of MSR-Method 4.3.1 satisfy the inequalities

$$
\|u_{k_0,i}\| < r^*, \quad \|\bar{u}_{k_0,i}\| < r^* \quad \forall \ i,
$$

then for the termination index it holds $i(k_0) < \infty$.

**Proof:** Let $u^{**} \in U^* \cap \mathbb{B}_r$ be fixed. Due to $u^{**} \in Q^*$, $\bar{u}_{k_0,i} \in Q_{k_0}$ and condition (4.2.6), one can choose $\nu_{k_0} \in Q_{k_0}$ and $\bar{\nu}_{k_0,i} \in Q$ such that

$$
\|\nu_{k_0} - u^{**}\| \leq \mu_{k_0}, \quad \|\bar{u}_{k_0,i} - \bar{\nu}_{k_0,i}\| \leq \mu_{k_0}, \quad \forall \ i = 1, ..., i(k_0). \tag{4.3.6}
$$

On account of (4.1.3) and (4.3.6) the estimates

$$
J(\bar{u}_{k_0,1}) - J(\bar{u}_{k_0,i}) \leq L(r)\mu_{k_0},
$$

$$
J(\nu_{k_0}) - J(u^{**}) \leq L(r)\mu_{k_0}
$$

are true, and with regard to $J(u^{**}) \leq J(\bar{u}_{k_0,1})$, we have

$$
J(\nu_{k_0}) - J(\bar{u}_{k_0,1}) \leq 2L(r)\mu_{k_0}.
$$

On account of (4.2.5),

$$
J_{k_0}(\nu_{k_0}) - J_{k_0}(\bar{u}_{k_0,1}) \leq 2L(r)\mu_{k_0} + 2\sigma_{k_0} =: \tau_{k_0} \tag{4.3.7}
$$

holds true for all $i = 1, ..., i(k_0)$.

Using Proposition 3.1.3 with

$$
C := Q_{k_0}, \quad f := J_{k_0}, \quad v^0 := u_{k_0,1}^0, \quad y := \nu_{k_0},
$$

we obtain from (3.1.3)

$$
\|\bar{u}_{k_0,i} - \nu_{k_0}\|^2 - \|u_{k_0,i} - v_{k_0}\|^2 \leq -\|u_{k_0,i}^0 - u_{k_0,i}^{n-1}\|^2 + \tau_{k_0}. \tag{4.3.8}
$$

But in view of stopping rule (4.3.2), for each $k$ and $i$ the estimate

$$
\|u_{k,i}^0 - \bar{u}_{k,i}^0\| \leq \frac{\epsilon_k}{2} \tag{4.3.9}
$$

is true. For $1 \leq i < i(k_0)$, due to

$$
\|u_{k,i}^0 - u_{k,i-1}^0\| > \delta_{k_0} > \frac{\epsilon_{k_0}}{2}
$$

and (4.3.8), (4.3.9), it follows that

$$
\|u_{k,i}^0 - v_{k_0}\|^2 - \|u_{k,i-1}^0 - v_{k_0}\|^2 \leq -\left(\|u_{k,i}^0 - u_{k,i-1}^0\| - \frac{\epsilon_{k_0}}{2}\right)^2 + \tau_{k_0},
$$

$$
\|u_{k,i}^0 - v_{k_0}\|^2 - \|u_{k,i-1}^0 - v_{k_0}\|^2 \leq -\epsilon_{k_0}^2 + \tau_{k_0}
$$

and from (4.3.5) we have $\tau_{k_0} < \epsilon_{k_0}^2$. Hence, for $1 \leq i < i(k_0)$,

$$
\|\bar{u}_{k_0,i}^0 - u_{k_0}\| - \|u_{k_0,i-1}^0 - v_{k_0}\| \leq \frac{\epsilon_{k_0}^2}{2\|u_{k,i} - v_{k_0}\| + \|u_{k,i-1} - v_{k_0}\|}
$$

$$
< \frac{-\epsilon_{k_0}^2 + \tau_{k_0}}{2\|u_{k,i} - v_{k_0}\| + \|u_{k,i-1} - v_{k_0}\|}. \tag{4.3.10}
$$
With regard to $\|v^{0}\| \leq r$ and $\|u^{k,i}\| < r \ \forall \ i$, the inequalities (4.3.9) and (4.3.10) lead to

$$
\|u^{0} - v^{0}\| < \|u^{0} - v^{0}\| + \frac{1}{4r}(-\tilde{\epsilon}^{2} + \tau) + \frac{\epsilon_{k}}{2}, \quad 1 \leq i \leq i(k_{0}).
$$

Summing up these inequalities for $i = 1, \ldots, i_{\max}$, with $i_{\max} \leq i(k_{0}) - 1$ arbitrarily chosen, we get

$$
\|u^{0, i_{\max}} - v^{0}\| < \|u^{0,0} - v^{0}\| + i_{\max}\left(\frac{1}{4r}(-\tilde{\epsilon}^{2} + \tau) + \frac{\epsilon_{k}}{2}\right),
$$

where, because of (4.3.5),

$$
\frac{1}{4r}(-\tilde{\epsilon}^{2} + \tau) + \frac{\epsilon_{k}}{2} < 0.
$$

Hence,

$$
i_{\max} < \frac{\|u^{0,0} - v^{0}\|}{\frac{1}{4r}(-\tilde{\epsilon}^{2} + \tau) + \frac{\epsilon_{k}}{2}},
$$

showing that $i(k_{0}) < \infty$.

It should be emphasized that the relations (4.2.5), (4.2.6) and (4.3.5) are used only for a fixed iteration index $k_{0}$.

4.3.4 Corollary. If the assumptions of Lemma 4.3.3 are fulfilled and for $u^{k,i}$, $\bar{u}^{k,i}$, generated according to MSR-Method 4.3.1, the inequalities

$$
\|u^{k,i}\| < r^{*}, \quad \|\bar{u}^{k,i}\| < r^{*}, \quad \forall \ k, \forall \ i
$$

are satisfied, then $i(k) < \infty$ holds for any external iteration $k$, i.e., the MSR-Method is well defined.

4.3.5 Lemma. Suppose that for given $r^{*}$ and $r \geq r^{*}$ the functional $J$ is Lipschitz on $B_{r}$ according to (4.1.3) and Assumption 4.2.2 is fulfilled. Moreover, assume that

$$
U^{*} \cap B_{r^{*}} / 4 \neq \emptyset, \quad u^{1.0} := u^{0} \in B_{r^{*}} / 4, \quad (4.3.12)
$$

and controlling parameters $\epsilon_{k}$, $\delta_{k}$, $\mu_{k}$ and $\sigma_{k}$ are chosen such that

$$
\frac{1}{4r}\left(2L(r)\mu_{k} + 2\sigma_{k} - (\delta_{k} - \frac{\epsilon_{k}}{2})^{2}\right) + \frac{\epsilon_{k}}{2} < 0, \quad (4.3.13)
$$

$$
\sum_{k=1}^{\infty}\left(\sqrt{2L(r)}\mu_{k} + \frac{\epsilon_{k}}{2} + 2\sigma_{k}\right) < \frac{r^{*}}{2}. \quad (4.3.14)
$$

Then for MSR-Method 4.3.1 it holds:

(i) $i(k) < \infty \ \forall \ k$;

(ii) $\|u^{k,i}\| < r^{*}, \quad \|\bar{u}^{k,i}\| < r^{*}, \quad \forall \ k, \forall \ i$. 
4.3. MULTI-STEP REGULARIZATION METHODS

**Proof:** In view of (4.3.13), (4.3.14) we have

$$2r\epsilon_k < (\delta_k - \frac{\epsilon_k}{2})^2, \quad \epsilon_k < r^*,$$

consequently

$$2\epsilon_k^2 < (\delta_k - \frac{\epsilon_k}{2})^2, \quad \delta_k > \frac{\epsilon_k}{2}.$$ 

Therefore,

$$\tilde{\epsilon}_k := \delta_k - \frac{\epsilon_k}{2} > \sqrt{r\epsilon_k} > 0. \quad (4.3.15)$$

Now, let $u^{**} \in U^* \cap B_{r^*}$ be fixed. For proof by induction we make the following assumptions:

Suppose that $k_0$ and $i_0$ are fixed, $0 \leq i_0 < i(k_0)$ and

(a) $i(k) < \infty \quad \forall \ k < k_0$;

(b) $\|u^{k,i}\| < r^*$, $\|\tilde{u}^{k,i}\| < r^*$, $\forall (k, i) \in I_0$,

with $I_0 := \{(k', i') : k' < k_0, 0 < i' \leq i(k') \text{ and } k' = k_0, 0 < i' \leq i_0\}$.

Denote

$$\hat{u}^{k,i} := \arg \min \{\Psi_{k,i}(u) : u \in Q_k\}.$$ 

For $(k, i) \in I_0$ it holds obviously $\tilde{u}^{k,i} = \hat{u}^{k,i}$.

Now we choose $v^k \in Q_k$ such that

$$\|v^k - u^{**}\| \leq \mu^k, \quad (4.3.16)$$

which is possible due to Assumption 4.2.2.

Similarly as in the proof of the previous lemma we conclude that for $(k, i) \in I_0$

$$J_k(v^k) - J_k(\hat{u}^{k,i}) \leq 2L(r)\mu_k + 2\sigma_k =: \tau_k. \quad (4.3.17)$$

Now, if $k < k_0$, $0 \leq i < i(k) - 1$ or $k = k_0$, $0 \leq i < i_0$, then using Proposition 3.1.3 with

$$C := Q_k, \quad f := J_k, \quad u^0 := u^{k,i}, \quad y := v^k,$$

we get (cf. inequality (4.3.10)) in view of $\hat{u}^{k,i+1} = \hat{u}^{k,i+1}$

$$\|\tilde{u}^{k,i+1} - v^k\| - \|u^{k,i} - v^k\| \leq \frac{\tilde{\epsilon}_k^2 + \tau_k}{2\|u^{k,i} - v^k\|} \leq \frac{1}{4r} \left(-\tilde{\epsilon}_k^2 + \tau_k\right), \quad (4.3.18)$$

and on account of (4.3.9) and (4.3.13)

$$\|u^{k,i+1} - v^k\| - \|u^{k,i} - v^k\| \leq \frac{1}{4r} \left(-\tilde{\epsilon}_k^2 + \tau_k\right) + \frac{\epsilon_k}{2} < 0. \quad (4.3.19)$$

Analogously, for $k < k_0$ and $i = i(k) - 1$,

$$\|u^{k,i(k)} - v^k\| - \|u^{k,i(k)-1} - v^k\| \leq \sqrt{\tau_k} + \frac{\epsilon_k}{2}. \quad (4.3.20)$$

Summing up the inequalities (4.3.19), (4.3.20) for a fixed $k < k_0$ and $0 \leq i \leq i(k) - 1$, we obtain

$$\|u^{k,i(k)} - v^k\| - \|u^{k,0} - v^k\| \leq \sqrt{\tau_k} + \frac{\epsilon_k}{2}, \quad \forall \ k < k_0,$$
hence, due to (4.3.16)
\[ \|u^{k,i(k)} - u^{**}\| - \|u^{k,0} - u^{**}\| \leq \sqrt{\tau_k} + \frac{\epsilon_k}{2} + 2\mu_k. \] (4.3.21)

Because of
\[ J_{k_0}(u^{k_0}) - J_{k_0}(\hat{u}^{k_0,i_0+1}) \leq \gamma_{k_0} \]
and Proposition 3.1.3 it follows that
\[ \|\hat{u}^{k_0,i_0+1} - v^{k_0}\| \leq \|u^{k_0,0} - v^{k_0}\| + \sqrt{\tau_{k_0}}, \] (4.3.22)
and summing up the inequalities (4.3.19) for \( k = k_0, 0 \leq i < i_0 \) and (4.3.22), this leads to
\[ \|\hat{u}^{k_0,i_0+1} - v^{k_0}\| \leq \|u^{k_0,0} - v^{k_0}\| + \sqrt{\tau_{k_0}}, \]
consequently,
\[ \|\hat{u}^{k_0,i_0+1} - u^{**}\| \leq \|u^{k_0,0} - u^{**}\| + \sqrt{\tau_{k_0}} + 2\mu_{k_0}. \] (4.3.23)

Due to (4.3.12), the chosen element \( u^{**} \) and (4.3.14), we obtain
\[ \|\hat{u}^{k_0,i_0+1}\| < r^* - \frac{\epsilon_{k_0}}{2}, \]

\[ \|\hat{u}^{k_0,i_0+1}\| < r^* - \frac{\epsilon_{k_0}}{2}. \] (4.3.25)
Inequality (4.3.25) together with (4.3.9) provides
\[ \|u^{k_0,i_0+1}\| < r^*, \] (4.3.26)
This enables us to conclude from condition (b) that
\[ \|\hat{u}^{k_0,i}\| < r^*, \quad \|u^{k_0,i}\| < r^*, \quad \forall i \] (4.3.27)
and Lemma 4.3.3 guarantees that \( i(k_0) < \infty \).
In a similar way, by means of Proposition 3.1.3, the inequalities
\[ \|u^{1,i}\| < r^*, \quad \|\hat{u}^{1,i}\| < r^* \]
can be verified. Hence, the estimates
\[ \|u^{1,i}\| < r^*, \quad \|\hat{u}^{1,i}\| < r^*, \quad \forall 0 < i \leq i(1) \]
are valid and again the usage of Lemma 4.3.3 implies that \( i(1) < \infty \). Thus, induction is complete. \( \square \)

Having the previous lemmata proved, we are now able to state convergence of the MSR-methods.
4.3. MULTI-STEP REGULARIZATION METHODS

4.3.6 Theorem. Suppose the assumptions of Lemma 4.3.3 are fulfilled and \( u^{k,i}, \tilde{u}^{k,i} \), generated by MSR-Method 4.3.1, satisfy the inequalities

\[
\|u^{k,i}\| < r^*, \quad \|\tilde{u}^{k,i}\| < r^*, \quad \forall \ k, \forall \ i. \tag{4.3.28}
\]

Moreover, let

\[
\sum_{k=1}^{\infty} \sqrt{\mu_k} < \infty, \quad \sum_{k=1}^{\infty} \sqrt{\sigma_k} < \infty, \quad \sum_{k=1}^{\infty} \epsilon_k < \infty. \tag{4.3.29}
\]

Then \( \{u^{k,i}\} \rightarrow u^* \in U^* \).

If, in addition, Assumption 4.2.3 holds, then \( \{u^{k,i}\} \rightarrow u^* \in U^* \) in the norm of the space \( V \).

Proof: Well-definedness of MSR-Method 4.3.1, i.e. \( i(k) < \infty \ \forall k \), has been noted in Corollary 4.3.4.

Again, let \( u^{*,*} \in U^* \cap B_r \) be arbitrarily chosen but fixed and

\[
u^{k,i+1}_Q := \arg \min_{u \in Q} \{|\Psi_k u^{k+1}(u)|\}.
\]

For \( i \leq i(k) - 1 \) inequality

\[
0 \leq 2(|u^{k,i+1}_Q - u^{k,i} - u^{*,*} - u^{k,i+1}_Q|) + J(u^{*,*}) - J(u^{k,i+1}_Q)
\]

follows from Proposition A1.5.34.

For \( i := i(k) \) all considerations with respect to \( u^{k,i+1}_Q \) can be performed completely analogous.

Consequently, due to (3.1.3),

\[
\|u^{k,i} - u^{*,*}\|^2 - \|u^{k,i+1}_Q - u^{*,*}\|^2 \geq J(u^{k,i+1}_Q) - J(u^{*,*}) + \|u^{k,i+1}_Q - u^{k,i}\|^2. \tag{4.3.30}
\]

From (4.2.5), (4.2.6) and (4.1.3) the inequalities

\[
\min_{u \in Q_k} \Psi^k u^{k,i+1}(u) - \min_{u \in Q_k} \Psi' u^{k,i+1}(u) \leq (L(r) + 4r)\mu_k,
\]

\[
\min_{u \in Q_k} \Psi^k u^{k,i+1}(u) - \min_{u \in Q_k} \Psi' u^{k,i+1}(u) \leq (L(r) + 4r)\mu_k + \sigma_k
\]

can be deduced, hence

\[
\Psi^k u^{k,i+1}_Q - \Psi' u^{k,i+1}_Q \leq (L(r) + 4r)\mu_k + \sigma_k. \tag{4.3.31}
\]

Denote \( \tilde{u}^{k,i} := \arg \min_{v \in Q} \|v - u^{k,i}\| \).

Due to (4.2.6), the Lipschitz property of \( \Psi^k u^{k,i+1}_Q \) and

\[
|\Psi^k u^{k,i+1}_Q(\tilde{u}^{k,i+1}_Q) - \Psi^k u^{k,i+1}_Q(\tilde{u}^{k,i+1}_Q)| \leq |\Psi^k u^{k,i+1}_Q(\tilde{u}^{k,i+1}_Q) - \Psi^k u^{k,i+1}_Q(\tilde{u}^{k,i+1}_Q)| + 2\sigma_k,
\]

we conclude that

\[
\Psi^k u^{k,i+1}_Q(\tilde{u}^{k,i+1}_Q) \leq \Psi^k u^{k,i+1}_Q(\tilde{u}^{k,i+1}_Q) + (L(r) + 4r)\mu_k + 2\sigma_k,
\]

which together with (4.3.31) leads to

\[
\Psi^k u^{k,i+1}_Q(\tilde{u}^{k,i+1}_Q) - \Psi^k u^{k,i+1}_Q(\tilde{u}^{k,i+1}_Q) \leq 2(L(r) + 4r)\mu_k + 3\sigma_k,
\]

\[
\Psi^k u^{k,i+1}_Q(\tilde{u}^{k,i+1}_Q) - \Psi^k u^{k,i+1}_Q(\tilde{u}^{k,i+1}_Q) \leq 2(L(r) + 4r)\mu_k + 4\sigma_k.
\]
By virtue of the strong convexity of $\tilde{\Psi}_{k,i+1}$, we have
\[ \|\tilde{u}_{k,i+1}^{k+1} - u_{Q}^{k+1}\| \leq \sqrt{2(L(r) + 4r)\mu_k + 4\sigma_k}. \]

Finally, because of (4.2.6) and (4.3.9), the estimate
\[ \|u^{k,i+1} - u_{Q}^{k+1}\| \leq \sqrt{2(L(r) + 4r)\mu_k + \mu_k + 4\sigma_k} \frac{\epsilon_k}{2} := \theta_k \quad (4.3.32) \]
is true. From (4.3.28) and the choices $u^{*}$ and $u_{Q}^{k,i}$, the above result yields
\[ \|u^{k,i+1} - u^{*}\|^2 - \|u_{Q}^{k+1} - u^{*}\|^2 \leq 4r\theta_k, \]
and in view of (4.3.30),
\[ \|u^{k,i} - u^{*}\|^2 - \|u_{Q}^{k,i+1} - u^{*}\|^2 \geq \|u^{k,i} - u_{Q}^{k+1,0}\|^2 + J(u_{Q}^{k+1,0}) - J(u^{*}) - 4r\theta_k \]
holds true. But for $i < i(k)$ and
\[ v^k \in Q_k : \|v^k - u^{*}\| \leq \mu_k, \quad (4.3.34) \]
following the proof of inequality (4.3.11), we get
\[ \|u^{k,i} - v^k\| \leq \|u^{k,i-1} - v^k\|. \quad (4.3.35) \]
Hence,
\[ \|u^{k,i} - v^k\| < \|u^{k,i-1} - v^k\|, \]
\[ \|u^{k,i} - u^{*}\| < \|u^{k,0} - u^{*}\| + 2\mu_k. \]
Now, taking into account that $\|u^{k,0}\| < r$, $\|u^{*}\| \leq r^{*}$, it follows for $i := i(k) - 1$
\[ \|u^{k,i} - u^{*}\|^2 < \|u^{k,0} - u^{*}\|^2 + 8r\mu_k + 4\mu_k^2. \]
Due to (4.3.33) and the definitions of $u_{Q}^{k,i}$ and $u^{*}$, the latter inequality provides
\[ \|u^{k,0} - u^{*}\|^2 - \|u^{k+1,0} - u^{*}\|^2 \]
\[ > \|u^{k,i(k)-1} - u^{k+1,0}\|^2 + J(u_{Q}^{k+1,0}) - J(u^{*}) - 4r(\theta_k + 2\mu_k) - 4\mu_k^2, \]
therefore,
\[ \|u^{k+1,0} - u^{*}\|^2 < \|u^{k,0} - u^{*}\|^2 + 4r(\theta_k + 2\mu_k) + 4\mu_k^2. \]
But in virtue of (4.3.29) the relations
\[ \sum_{k=1}^{\infty} \mu_k < \infty, \quad \sum_{k=1}^{\infty} \theta_k < \infty. \]
are fulfilled, therefore Lemma A3.1.4 ensures convergence of the sequence
\[ \{\|u^{k,0} - u^{**}\|\} \quad \text{for any } u^{**} \in U^* \cap B_{r^*}. \]
Due to Proposition 3.1.3 it holds
\[ \|\bar{u}^{k,i(k)} - v^k\| \leq \|u^{k,i(k)-1} - v^k\| + \sqrt{\tau_k}, \]
hence,
\[ \|u^{k,i(k)} - u^k\| \leq \|u^{k,i(k)-1} - v^k\| + \sqrt{\tau_k} + \epsilon_k \]  \hspace{1cm} \text{(4.3.36)}
and together with (4.3.34) and (4.3.35), this leads to
\[ \|u^{k,i} - u^{**}\| \geq \|u^{k+1,0} - u^{**}\| - \sqrt{\tau_k} - \frac{\epsilon_k}{2} - 2\mu_k. \]  \hspace{1cm} \text{(4.3.37)}
Using (4.3.35), (4.3.37) and (4.3.29), one can conclude that the sequences
\[ \{\|u^{k,i} - u^{**}\|\}, \quad \{\|u_Q^{k,i} - u^{**}\|\} \]
converge to the same limit as \(\{\|u^{k,0} - u^{**}\|\}\) it does (the order of the elements is defined by the method).
Thus, (4.3.30) and (4.3.32) yield
\[ \lim_{k \to \infty} \max_{1 \leq i \leq n(k)} |J(u_Q^{k,i}) - J(u^{**})| = 0, \]
\[ \lim_{k \to \infty} \max_{1 \leq i \leq n(k)} |J(u^{k,i}) - J(u^{**})| = 0. \]
Now, in order to finish this proof, we can follow the final part of the proof of Proposition 1.3.11 and the proof of Lemma 4.1.4. □

4.3.7 Corollary. Theorem 4.3.6 remains true if instead of condition (4.3.28) it is assumed that
\[ K_k \in \text{int} B_{r_k} \quad \text{with } r_k := r^* - \frac{\epsilon_k}{2}, \quad k = 1, 2, \cdots. \]  \hspace{1cm} \text{(4.3.38)}

Applying Lemma 4.3.5 and Theorem 4.3.6 we conclude immediately convergence of the MSR-methods.

4.3.8 Theorem. Suppose the assumptions of Lemma 4.3.5 are fulfilled. Then, MSR-Method 4.3.1 applied to Problem (4.2.1) is well-defined, i.e. \(i(k) < \infty \) \(k = 1, 2, \cdots, \) and \(\{u^{k,i}\} \to u^* \in U^*\).
If, in addition, Assumption 4.2.3 holds true, then \(\|u^{k,i} - u^*\|_V \to 0.\)
A comparison between Theorem 4.3.8 on the one hand and Theorem 4.3.6 together with Corollary 4.3.7 on the other hand shows: If the sets \(K_k\) are uniformly bounded, then the requirements for choosing the controlling parameters in order to ensure convergence of the MSR-Method b4.3.1 are essentially weaker than in Theorem 4.3.8. In fact, condition (4.3.14) may be replaced by (4.3.29). In particular, this enables us to start the iteration with a coarser approximation of the original problem.
Knowing a radius \(r^*\) with \(U^* \cap B_{r^*} \neq \emptyset\) allows us to add a constraint
CHAPTER 4. PPR FOR INFINITE-DIMENSIONAL PROBLEMS

\[ g_0(u) \leq r_0 \text{ to the original problem. The choice of the functional } g_0 \text{ and constant } r_0 \text{ (for example } g_0(\cdot) := \| \cdot \| \text{ and } r_0 := r^* \text{) has to guarantee that the set } \{ u : g_0(u) \leq r_0 \} \text{ is bounded and that} \]

\[ U^* \cap \{ u : g_0(u) \leq r_0 \} \neq \emptyset. \]

Obviously, in this way approximations of the transformed problem can be constructed such that the family \( \{ K_k \} \) is bounded. This artificial approach may be efficient, for instance, for solving convex semi-infinite problems with non-linear constraints. However, for specific variational inequalities, as considered in Section 8.2, this may lead to essential more complicated approximate problems on each discretization level. Moreover, the estimate \( \rho(Q^*, Q_k) \), which is necessary for controlling the discretization procedure, becomes more difficult.

4.3.9 Remark. If the assumptions of Lemma 4.3.3 or Lemma 4.3.5 are fulfilled for a given sequence \( \{ \delta_k \} \), then they are also satisfied for a sequence \( \{ \bar{\delta}_k \} \) with \( \bar{\delta}_k \geq \delta_k \), whereas the other parameters remain unchanged. Particularly, for \( \bar{\delta}_k > 2r^* \forall k \in \mathbb{N} \), MSR-Method 4.3.1 turns into an OSR-method, i.e., for the termination index yields \( i(k) = 1 \forall k \). In this case condition (4.3.13) is a conclusion of (4.3.14). Thus, Lemma 4.3.5 and Theorem 4.3.8 provide a convergence result for an OSR-method similarly to Theorem 4.2.4. An analogous investigation tells us that Theorem 4.3.6 remains also true for the modified OSR-Method 4.2.8. In this situation condition (4.3.5) is superfluous.

On the other hand, one can formally consider MSR-Method 4.3.1 as an OSR-method if one takes

\[ K_{k,i} := K_k, \quad J_{k,i} := J_k, \quad \text{for } 1 \leq i \leq i(k) \]

with \textit{a priori} unknown \( i(k) \). However, a straightforward application of Theorem 4.2.4 leads to essentially stronger sufficient convergence conditions than in Theorem 4.3.8. Indeed, instead of (4.3.14) the relation

\[ \sum_{k=1}^{\infty} i(k) \left( \sqrt{2L(r)}\mu_k + 2\sigma_k + \frac{\epsilon_k}{2} + 2\mu_k \right) < \frac{r^*}{2} \]

has to be required. \( \diamond \)

4.3.10 Theorem. Under the assumptions of Lemma 4.3.5 the following estimates are true for each exterior step \( k \) in MSR-Method 4.3.1:

\[ J_k(u^{k,i(k)}) - \inf_{u \in Q_k} J_k(u) < 4r(\delta_k + \frac{\epsilon_k}{2}) + \frac{1}{2} L(r)\epsilon_k + 2\sigma_k, \quad (4.3.39) \]

\[ J(u^{k,i(k)}) - \inf_{u \in K} J(u) < 4r\delta_k + \frac{1}{2}(4r + L(r))\epsilon_k + L(r)\mu_k + 2\sigma_k. \quad (4.3.40) \]
Proof: Since $J_k$ is supposed to be weakly lower semi-continuous at $Q_k$, there exists

$$\hat{v}^k := \arg \min_{v \in Q_k} J_k(v).$$

By definition of $\hat{u}^{k,i(k)}$, for any $\lambda \in (0, 1]$ it holds

$$J_k(\hat{u}^{k,i(k)} + \lambda(\hat{v}^k - \hat{u}^{k,i(k)})) + \frac{1}{2} \left\| \hat{u}^{k,i(k)} + \lambda(\hat{v}^k - \hat{u}^{k,i(k)}) - u^{k,i(k)} \right\|^2 \geq J_k(\hat{u}^{k,i(k)}) + \left\| \hat{u}^{k,i(k)} - u^{k,i(k)} \right\|^2,$$

and in view of convexity of $J_k$ and positivity of $\lambda$, we obtain

$$J_k(\hat{u}^{k,i(k)}) \leq \lambda J_k(\hat{v}^k) + (1 - \lambda)J_k(\hat{u}^{k,i(k)}) + \frac{1}{2} \lambda^2 \left\| \hat{v}^k - \hat{u}^{k,i(k)} \right\|^2 + 2\lambda \langle \hat{v}^k - \hat{u}^{k,i(k)}, \hat{u}^{k,i(k)} - u^{k,i(k)} \rangle,$$

hence,

$$J_k(\hat{u}^{k,i(k)}) - J_k(\hat{v}^k) \leq 2\langle \hat{u}^{k,i(k)} - u^{k,i(k)} - 1, \hat{v}^k - \hat{u}^{k,i(k)} \rangle + \lambda \left\| \hat{v}^k - \hat{u}^{k,i(k)} \right\|^2.$$  

For $\lambda \downarrow 0$ we get

$$J_k(\hat{u}^{k,i(k)}) - J_k(\hat{v}^k) \leq 2\left\| \hat{u}^{k,i(k)} - u^{k,i(k)} - 1 \right\| \left( \left\| \hat{u}^{k,i(k)} \right\| + \left\| u^{k,i(k)} - u^{k,i(k)} \right\| \right),$$

and, due to the stopping rule for the interior loop, one has

$$J_k(\hat{u}^{k,i(k)}) - J_k(\hat{v}^k) \leq 4r (\delta_k + \frac{\epsilon_k}{2}).$$  

(4.3.41)

Because of

$$|J_k(u^{k,i}) - J_k(\hat{u}^{k,i})| \leq |J(u^{k,i}) - J(\hat{u}^{k,i})| + 2\sigma_k \leq \frac{1}{2} L(r)\epsilon_k + 2\sigma_k$$

for $1 \leq i \leq i(k)$, it yields (4.3.39).

For a fixed $u^* \in U^* \cap B_r$, and $w^k := \arg \min_{v \in Q_k} \| u^* - v \|$ we conclude

$$J(u^{k,i(k)}) - J(u^*) = J(\hat{u}^{k,i(k)}) - J(u^*) + L(r)(\frac{\epsilon_k}{2} + \mu_k) \leq J(\hat{u}^{k,i(k)}) - J(w^k) + L(r)(\frac{\epsilon_k}{2} + \mu_k) + 2\sigma_k \leq J(\hat{u}^{k,i(k)}) - J(\hat{v}^k) + L(r)(\frac{\epsilon_k}{2} + \mu_k) + 2\sigma_k$$

and, due to (4.3.41), estimate (4.3.40) is true.

\[\Box\]

4.3.11 Remark. Using Assumption 4.2.2, Lemma 4.3.5 and the inequalities (4.1.3) and (4.3.9), we immediately obtain the following lower estimates for each $k$ and $1 \leq i \leq i(k)$:

$$J_k(u^{k,i}) - J_k(\hat{v}^k) \geq -L(r)(\frac{\epsilon_k}{2} + \mu_k)$$

and

$$J(u^{k,i}) - \inf_{u \in K} J(u) \geq -L(r)(\frac{\epsilon_k}{2} + \mu_k),$$

(4.3.42)

with $\hat{v}^k := \arg \min_{v \in Q_k} J_k(v)$.

\[\Box\]
4.3.3 MSR-method with regularization on subspaces

Let $V$ be divided into two subspaces such that $V_1 + V_2 = V$, $V_1 \cap V_2 = \emptyset$ and $\Pi := \Pi_2$ is an orthogonal projector onto the subspace $V_2$, whereas $\Pi_1 := I - \Pi$ projects $V$ onto the subspace $V_1$.

Suppose now that the functional $J$ in Problem (4.2.1) has the property
\[
J(u) - J(v) \geq \langle q(v), u - v \rangle + \kappa \|\Pi u - \Pi v\|^2, \quad \forall \, u, v \in V,
\]
(4.3.43)
where $\kappa > 0$, $q(v) \in \partial J(v)$.

Concerning the approximation of Problem (4.2.1), we assume that the requirements made with respect to $\{K_k\}$, $\{J_k\}$ and the parameters $\{\delta_k\}$ and $\{\epsilon_k\}$ are the same as before.

Additionally we assume that condition (4.3.43) holds for the family $\{J_k\}$ with the common constant $\kappa > 0$.

Now we consider the following regularized functional
\[
\Psi_{k,i}(u) := J_k(u) + \|\Pi_1 u - \Pi_1 u^{k,i-1}\|^2 + \alpha \|\Pi u - \Pi u^{k,i-1}\|^2,
\]
(4.3.44)
with $\alpha \geq 0$ a given constant.

In this case we suggest a MSR-method which carries out regularization on subspaces.

4.3.12 Method. (MSR-method on subspaces)

Given $u^0$ and sequences $\{\delta_k\}$, $\{\epsilon_k\}$, $k = 1, 2, \cdots$, such that
\[
\delta_k > 0, \quad \epsilon_k \geq 0, \quad \lim_{k \to \infty} \epsilon_k = 0.
\]

Step $k$: Given $u^{k-1}$.

(a) Set $u^{k,0} := u^{k-1}$, $i := 1$.

(b) Given $u^{k,i-1}$, compute approximately $u^{k,i}$ according to
\[
\|\nabla \Psi_{k,i}(u^{k,i}) - \nabla \Psi_{k,i}(\bar{u}^{k,i})\|_{V^*} \leq \epsilon_k
\]
(4.3.45)
with
\[
\bar{u}^{k,i} := \arg \min \{\Psi_{k,i}(u) : u \in K_k\},
\]
\[
\epsilon_k' \leq \epsilon_k \frac{\min[1, \alpha + \kappa]}{\sqrt{\max[1, \alpha + \kappa]}},
\]
(4.3.46)

(c) If $\|\Pi_1 u^{k,i} - \Pi_1 u^{k,i-1}\| > \delta_k$, set $i := i + 1$ and repeat (b).

Otherwise, set $u^k := u^{k,i}$, $i(k) := i$, and continue with Step $k := k + 1$.

We emphasize that, in comparison with Assumption 4.2.3, finite-dimensionality of $V_1 := (I - P)V$ has been not assumed here.

Setting $\alpha := 0$ in (4.3.44), then an essentially different MSR-method arises,
whereas in the case $\alpha := 1$ we deal with the previously described MSR-method.

In order to obtain sufficient convergence conditions of MSR-Method 4.3.12 we are going to adapt Lemma 4.3.5 and Theorem 4.3.8 to the new situation.

Denote $\| \cdot \|$ the norm induced by the scalar product

$$\langle u,v \rangle := (\Pi_1 u, v) + (\alpha + \kappa) (\Pi u, v),$$

where $\langle \cdot, \cdot \rangle$ is the duality pairing between $V$ and its dual space $V^\prime$.

Let us prove the following lemma which is similar to Proposition 3.1.3:

**Lemma.** Assume the functional $J$ in Problem (4.2.1) satisfies condition (4.3.43) and $v^0 \in V$ is arbitrarily chosen and

$$v^1 := \arg\min_{u \in K} \{ J(u) + \|\Pi_1 u - \Pi_1 v^0\|^2 + \alpha \|\Pi u - \Pi v^0\|^2 \} \quad (\alpha \geq 0 \text{ fixed}).$$

Then, the following estimates hold for each $u \in K$:

$$\|v^1 - u\|^2 \leq \|v^0 - u\|^2 - \|\Pi_1 v^0 - \Pi_1 v^1\|^2 - \alpha\|\Pi v^0 - \Pi v^1\|^2 - \kappa\|\Pi v^0 - \Pi u\|^2 + J(u) - J(v^1),$$

$$\|\Pi v^1 - u\| \leq \|\Pi v^0 - u\| + \sqrt{J(u) - J(v^1)}. \quad (4.3.47)$$

If, moreover, $J(u) - J(v^1) - \|\Pi_1 u^0 - \Pi_1 v^1\|^2 < 0$, then

$$\|v^1 - u\| - \|v^0 - u\| < \frac{J(u) - J(v^1) - \|\Pi_1 u^0 - \Pi_1 v^1\|^2}{2\|v^0 - u\|}. \quad (4.3.48)$$

**Proof:** Because of the definition of $v^1$ inequality

$$\langle q, u - v^1 \rangle + 2(\Pi_1 v_1 - \Pi_1 v^0 + \alpha(\Pi v^1 - \Pi v^0), u - v^1 \rangle \geq 0$$

is fulfilled for some $q \in \partial J(v^1)$ and all $u \in K$.

From the latter inequality together with (4.3.43) one can conclude

$$0 \leq 2(\Pi_1 v^1 - \Pi_1 v^0 + \alpha(\Pi v^1 - \Pi v^0), u - v^1) + J(u) - J(v^1) - \kappa\|\Pi u - \Pi v^1\|^2$$

and because $\Pi$ and $\Pi_1 := I - \Pi$ are self-adjoint operators and $\Pi^2 = \Pi$, $\Pi^2_1 = \Pi_1$,

$$\|v^1 - u\|^2 - \|v^0 - u\|^2 = -\|\Pi_1 v^0 - \Pi_1 v^1\|^2 - \alpha\|\Pi v^0 - \Pi v^1\|^2$$

$$+ 2(\Pi_1 v^1 - \Pi_1 v^0, v^1 - u) + 2\alpha(\Pi v^1 - \Pi v^0, v^1 - u)$$

$$- \kappa(\|\Pi v^0\|^2 - \|v^1\|^2 - 2(\Pi v^0, u) + 2(\Pi v^1, u))$$

$$\leq -\|\Pi_1 v^0 - \Pi_1 v^1\|^2 - \alpha\|\Pi v^0 - \Pi v^1\|^2 - \kappa\|\Pi v^0 - \Pi u\|^2 + J(u) - J(v^1).$$

Thus, relation (4.3.47) and

$$\|v^1 - u\|^2 - \|v^0 - u\|^2 \leq -\|\Pi_1 v^0 - \Pi_1 v^1\|^2 + J(u) - J(v^1)$$

hold true. If

$$J(u) - J(v^1) - \|\Pi_1 u^0 - \Pi_1 v^1\|^2 < 0,$$

then inequality (4.3.48) follows immediately. \qed
Note that the norms $\| \cdot \|$ and $||| \cdot |||$ are equivalent, i.e.
\[
\min[1, \alpha + \kappa]\|u\|^2 \leq |||u|||^2 \leq \max[1, \alpha + \kappa]\|u\|^2.
\]
Since functional $\Psi_{k,i}$ via (4.3.44) is strongly convex with constant $\bar{\kappa} := \min[1, \alpha + \kappa]$ (with respect to the norm $\| \cdot \|$), the latter relation together with (4.3.45) and (4.3.46) lead to
\[
|||u^{k,i} - \bar{u}^{k,i}||| \leq \frac{\kappa}{2}.
\]
Now, let $\mathcal{B}_r := \{ u \in V : |||u||| \leq r \}$ and
\[
\bar{\Psi}_{k,s}(u) := J(u) + \| \Pi_1 u - \Pi_1 u^{k,i-1} \|^2 + \alpha \| \Pi u - \Pi u^{k,i-1} \|^2,
\]
and let the terms $L(r), r^*, Q, Q_k, Q^*$ and $\sigma_k$ be defined as before but assigned now to the new norm $||| \cdot |||$ and new functionals $\Psi_{k,i}$ and $\bar{\Psi}_{k,i}$.

With these substitutions Assumption 4.2.2, Lemma 4.3.5 and Theorem 4.3.8 can be maintained also for MSR-Method 4.3.12 with only one modification: $\| \cdot \|$ has to be replaced by $||| \cdot |||$. Now, using Lemma 4.3.13, the proofs of the statements in Section 4.3.2 can be repeated for this case with evident corrections.

If $\alpha + \kappa = 1$, then $|||u||| = \|u|||$, hence, the terms $L(r), r^*, Q, Q_k, \bar{Q}_k$ and $\sigma_k$ coincide with the former ones. In principle, the choice of $\alpha + \kappa = 1$ can always be guaranteed by means of a suitable scaling of the objective functional.

All results described above concerning convergence of OSR- and MSR-methods, except for Theorem 4.2.7, permit us to choose $r^* = r$. However, as we will see in Section 8.2, this choice may be inappropriate for specific problems.

Now we are going to present two simple examples. The first one illustrates the advantages that are possible in applying MSR-methods in comparison with OSR-methods. The second one gives a look at the possible gain of the method with regularization on a subspace in comparison with the main variants of OSR- and MSR-methods.

Concerning the latter issue we refer also to Chapter ??, were the idea of regularization on a subspace is applied in a natural manner to control problems, i.e. regularization is performed only with respect to the control- but not to the space variables.

4.3.14 Example. Function $J(u) := \frac{1}{10}(u_2 + u_3 - 1)^2 + \frac{1}{10}u_4$ has to be minimized on
\[
K := \{ u \in U_0 : u_1 \sin \frac{\pi}{2} t + u_2 \cos \frac{\pi}{2} t + u_3 \leq 1 \text{ and } t \in [0, 1] \},
\]
with
\[
U_0 := \{ u \in \mathbb{R}^4 : u_1 \geq 0, u_2 \geq 0, 0 \leq u_3 \leq 1, -1 \leq u_4 \leq 0 \}.
\]

Obviously,
\[
U^* := \{ u \in \mathbb{R}^4 : u_1 = 0, u_2 + u_3 = 1, 0 \leq u_3 \leq 1, u_4 = -1 \}
\]
is the optimal set.

In the methods considered we take $\epsilon_k := 0, J_k := J, \forall k$, hence $\sigma_k := 0 \forall k$.

Approximating set $T = [0, 1]$ by means of a sequence of grids
\[
T_k := \{ 0, h_k, 2h_k, \ldots, (n(k) - 1)h_k, 1 \},
\]
4.3. MULTI-STEP REGULARIZATION METHODS

with \( n(k) > 0 \) integer, \( \lim_{k \to \infty} n(k) = \infty \) and \( h_k := \frac{1}{n(k)} \), we obtain the following sequence of sets

\[
K_k := \{ u \in U_0 : u_1 \sin \frac{\pi}{2} t + u_2 \cos \frac{\pi}{2} t + u_3 \leq 1 \quad \forall \ t \in T_k \}
\]

approximating the original feasible set \( K \).

It is easy to prove that \( \rho(K_i, K) \leq h_k^2 \) and that for \( r := \frac{3}{2} \) and \( h_k \leq \frac{1}{4} \) the inclusions \( K \subset \mathbb{B}_r, K_k \subset \mathbb{B}_r \) are valid, i.e., \( Q := K \) and \( Q_k := K_k \).

This allows us to choose the controlling parameters of the OSR- and MSR-methods according to Theorem 4.3.6 (see Remark 4.3.9).

The function \( J \) satisfies the Lipschitz condition (4.1.3) on \( \mathbb{B}_r \) with constant \( L(r) := L(\frac{3}{2}) = \frac{1}{3} \), thus, the assumptions of Theorem 4.3.6 are fulfilled for MSR-Method 4.3.1 if we set

\[
\mu_k := h_k^2, \quad \delta_k^2 > \frac{8}{3} \mu_k, \quad \sum_{k=1}^{\infty} h_k < \infty.
\]

Convergence of the OSR-Method 4.2.1 is guaranteed if \( \sum_{k=1}^{\infty} h_k < \infty \).

Now we choose the controlling parameters as follows:

- in the MSR-method:
  \[
  \delta_k := 0.8 \frac{1}{\sqrt{k}} \quad \text{for} \quad k \geq 1, \quad \text{i.e.,} \quad \delta_1 := 0.8, \quad \delta_2 := 0.6, \quad \delta_3 := 0.48, \ldots,
  \]
  \[
  h_k := \frac{1}{k(k-1)+1} \quad \text{for} \quad k \geq 1, \quad \text{i.e.,} \quad h_1 := \frac{1}{4}, \quad h_2 := \frac{1}{6}, \quad h_3 := \frac{1}{10}, \ldots;
  \]

- in the OSR-method:
  \[
  \tilde{h}_s := \frac{1}{s(s+1)+4} \quad \text{for} \quad s \geq 1,
  \]
  (in order to compare both methods we use the notation \( s \) and \( \tilde{h}_s \) instead of \( k \) and \( h_k \)).

From the discretization point of view this updating of \( h_k \) and \( \tilde{h}_s \) is not fast, compared with implementations in practice.

Starting with \( u^0 := (0, 0, 0, 0)^T \) one can verify that the iterates \( u^k \) of both methods satisfy the relations

\[
\sup_{t \in T} \{ u_1^k \sin \frac{\pi}{2} t + u_2^k \cos \frac{\pi}{2} t + u_3^k \} < 1, \quad u_2^k > 0, \quad 0 < u_3^k < 1, \tag{4.3.50}
\]

and that \( u_4^k = 0 \) holds for all \( k \).

Therefore, in the OSR-method we obtain for all \( s \geq 1 \):

\[
\frac{1}{2} \frac{\partial \Psi_s(u^s)}{\partial u_2} = \frac{1}{10} (u_2^s + u_3^s - 1) + (u_2^s - u_2^{s-1}) = 0, \tag{4.3.51}
\]

\[
\frac{1}{2} \frac{\partial \Psi_s(u^s)}{\partial u_3} = \frac{1}{10} (u_2^s + u_3^s - 1) + (u_3^s - u_3^{s-1}) = 0, \tag{4.3.52}
\]

\[
\frac{\partial \Psi_s(u^s)}{\partial u_4} = \frac{1}{12} + 2(u_4^s - u_4^{s-1}) = 0 \quad \text{if} \quad u_4^s \geq -1. \tag{4.3.53}
\]

Due to (4.3.51) and (4.3.52) we get

\[
u_2^s - u_2^{s-1} = u_2^{s-1} - u_3^{s-1}
\]

\[
u_3^s - u_3^{s-1} = u_2^{s-1} - u_3^{s-1}
\]

\[
u_4^s - u_4^{s-1} = u_4^{s-1} - u_4^{s-1}
\]
and from $u_2^0 = u_3^0 = 0$ it follows $u_2^s = u_3^s$ with
\[
\begin{align*}
    u_2^s &= \frac{5}{6} u_{s-1}^s + \frac{1}{12}, \\
    u_3^s &= \frac{5}{6} u_{s-1}^s + \frac{1}{12}.
\end{align*}
\]
Relation (4.3.53) implies now
\[u_4^s = \max[-1, -\frac{s}{24}].\]
Hence,
\[
\begin{align*}
    u_2^{s+1} - u_2^s &= \frac{5}{6} (u_2^{s-1} - u_2^{s-2}) = \left(\frac{5}{6}\right)^s (u_2^1 - u_2^0) = \frac{1}{12} \left(\frac{5}{6}\right)^s, \\
    u_3^{s+1} - u_3^s &= \frac{1}{12} \left(\frac{5}{6}\right)^s, \\
    u_4^{s+1} - u_4^s &= \begin{cases} 
        -\frac{1}{12} & \text{if } s \leq 23, \\
        0 & \text{if } s \geq 24,
\end{cases}
\end{align*}
\]
and
\[
\|u^{s+1} - u^s\| = \begin{cases} 
        \sqrt{\frac{1}{12} (\frac{5}{6})^{2s} + \frac{1}{12}} & \text{if } s \leq 23, \\
        \frac{1}{12} (\frac{5}{6})^s & \text{if } s \geq 24.
\end{cases} \tag{4.3.54}
\]
By an MSR-method a sequence $\{u^{k,i}\}$ is generated which corresponds to the sequence $\{u^s\}$ such that
\[u^{k,i} = u^s \quad \text{with} \quad s = \sum_{j=1}^{k-1} i(j) + i, \quad 1 \leq i \leq i(k).\]
The first iterations steps in the MSR-method lead to
\[i(1) := 3, \quad i(2) := 3, \quad i(3) := 3, \quad i(4) := 15\]
and the iterate
\[u^{4,15} := (0, 0.4937, 0.4937, -1)^T\]
corresponds to the discretization parameter $h_4 := \frac{1}{12}$.
The unique cluster point of these iterates is
\[u^* = (0, 0.5, 0.5, -1)^T \in U^*.\]
In an OSR-method the same point $(0, 0.4937, 0.4937, -1)^T$ will be obtained after 24 iterations and the corresponding discretization parameter is $\bar{h}_{24} := \frac{1}{556}$.
Thus, in a MSR-method semi-infinite constraint (4.3.49) generates 16 constraints (i.e. 16 exterior steps) whereas the OSR-method needs 556 constraints in order to get the same level of accuracy for the termination point.

4.3.15 Example. Function
\[J(u) := u_1^2 + u_2 + u_3 \tag{4.3.55}\]
4.3. MULTI-STEP REGULARIZATION METHODS

has to be minimized on
\[ K := \{ u \in \mathbb{R}^3 : u_2 + u_3 \geq 0 \}. \]

Obviously, the optimal set is
\[ U^* := \{ u \in \mathbb{R}^3 : u_1 = 0, u_2 + u_3 = 0 \}. \]

Now we choose
\[ J_k := J, \quad K_k := K; \quad \epsilon_k = \mu_k = \sigma_k := 0; \quad \forall \ k \in \mathbb{N} \]
and, in the case of the MSR-method, \( \delta_k > 0 \) is chosen arbitrarily. Then sequences, generated according to the exact Proximal Point Algorithm 3.1.5 (with \( \chi_k := 2 \)), coincide for both methods, OSR- as well as MSR-method, if they start with the same point \( u^0 \). Their convergence to some \( u^* \in U^* \) follows in particular from Proposition 3.1.8 or Theorem 4.2.4.

Let us focus on the sequence \( \{ u^k \} \), obtained by means of OSR-method, when starting with some \( u^0 > 0 \):
\[
\begin{align*}
u_1^{k+1} &:= \frac{1}{2} u_1^k, \quad k = 1, 2, \ldots; \\
u_2^{k+1} &:= \frac{1}{2} u_2^k - \frac{1}{2}, \quad \nu_3^{k+1} := \frac{1}{2} - \frac{1}{2} \quad \text{while } u_2^k + u_3^k - 1 > 0; \quad (4.3.56) \\
u_2^{k+j} &:= \frac{1}{2} (u_2^k - u_3^k), \quad \nu_3^{k+j} := \frac{1}{2} (u_3^k - u_2^k), \quad \text{if } u_2^k + u_3^k - 1 \leq 0, \quad (4.3.57)
\end{align*}
\]

for \( j := 1, 2, \ldots \). Consequently,
\[
\begin{align*}
u_2^k = u_2^{k+1} &= \ldots = u_2^* \\
u_3^k = u_3^{k+1} &= \ldots = u_3^* \quad \text{for } k > u_2^0 + u_3^0.
\end{align*}
\]

Starting from index \( k := k_0 > u_2^0 + u_3^0 \), linear convergence of the sequence holds, i.e.,
\[
\| u^{k+j} - u^* \| = \left( \frac{1}{2} \right)^j \| u^k - u^* \|, \quad \forall j \in \mathbb{N}.
\]

Now, let us turn to the regularization on subspace applied to the same problem. For function \( J \) condition (4.3.43) is true with the ortho-projector
\[
\Pi : u \mapsto (u_1, 0, 0)^T \quad \text{with } \kappa := 1,
\]

hence \( \Pi_1 = \mathcal{I} - \Pi : u \mapsto (0, u_2, u_3)^T \).

Taking in MSR-Method 4.3.12 \( \alpha := 0 \), we have to minimize the functions
\[
\Psi_k(u) := u_1^2 + u_2 + u_3 + (u_2 - u_2^{k-1})^2 + (u_3 - u_3^{k-1})^2
\]
successively on \( K \). In this situation the behavior of the method is again independent of the choice of \( \{ \delta_k \} \), therefore the iteration procedure can be interpreted as OSR-method.

Obviously, we get the sequence \( \{ u^k \} \), with \( u_1^k := 0 \) for \( k > 1 \) and \( u_2^k, u_3^k \) according to (4.3.56) and (4.3.57). Hence, this method generates the solution \( u^* \)
of the example in a finite number of steps.

If, instead of (4.3.55), the function

\[ J(u) := u_1^2 + a\sqrt{(u_2 + u_3)^3} \]

has to be minimized (with arbitrary \( a > 0 \) and \( u^0 > 0 \)), then MSR-Method 4.3.12, applied with the previously given ortho-projector \( \Pi_1 \) and \( \alpha := 0 \), converges also faster than an usual MSR-method, but not in a finite number of steps. In this case we get quadratic convergence, whereas the basic variants of OSR- and MSR-methods provide only linear convergence.

\[ \diamond \]

Of course, these examples show only some possible effects of particularly chosen regularization methods. In the following chapters we will find more arguments answering the question which PPR-method should be chosen for which specific problem.

Condition (4.3.13), responsible for a slow decrease of \( \delta_k \) (in comparison with the other parameters), could give rise to the expectation that MSR-Method 4.3.1 turns into an OSR-method after some iterations. The next example shows that this expectation may be wrong.

**4.3.16 Example.** Function \( J(u) := u_2^2 + u_3 \) has to be minimized on

\[ K := \{ u \in \mathbb{R}^3 : 0 \leq u_j \leq 1, j = 1, 2, 3 \} \]

Consider the approximations

\[ K_k := \{ u \in \mathbb{R}^3 : 0 \leq u_1 \leq 1, \frac{1}{2^k(k+1)+5} \leq u_2 \leq 1, 0 \leq u_3 \leq 1 \} \]

with

\[ J_k := J, \quad \epsilon_k := 0 \quad k = 1, 2, \ldots ; \quad r = r^* := 2 \]

In this situation we get

\[ Q := K, \quad Q_k := K_k, \quad \forall k \in \mathbb{N} ; \quad L(r) := 5. \]

Assumption 4.2.2 is fulfilled with \( \mu_k := \frac{1}{2^k(k+1)+5} \) and the conditions in Lemma 4.3.5 are satisfied if

\[ \delta_k := \frac{1}{\sqrt{2^k(k+1)+1}}, \quad u^{1,0} := (0, \frac{1}{2}, 0)^T. \]

We observe immediately that, applying MSR-methods, the resulting minimizing sequence is

\[ \{(0, \frac{1}{2}, 0)^T, (0, \frac{1}{4}, 0)^T, (0, \frac{1}{8}, 0)^T, (0, \frac{1}{16}, 0)^T, \ldots \} \]

and for the termination index of the interior loop it holds \( i(k) := k, \ k = 1, 2, \ldots \) (see Figur 4.3.2).

Indeed, the distance between successive iterates decreases and

\[ \|u^{k,i(k)-1} - u^{k,i(k)-2}\| = \frac{1}{\sqrt{2^k(k+1)+1}} > \delta_k, \]

\[ \|u^{k,i(k)} - u^{k,i(k)-1}\| = \frac{1}{2\sqrt{2^k(k+1)}} < \delta_k. \]
4.4 Choice of Controlling Parameters

We do not intend at giving in each case a strict formalization and argumentation of the recommendations contained in this section. Sometimes we will only sketch the way how an improvement of the numerical process can be obtained.

4.4.1 On adaptive controlling procedures for parameters

Let

$$
\theta_k := \sum_{s=1}^{k} \left( \sqrt{2L(r)} \mu_s + 2\sigma_s + 2\mu_s + \frac{\epsilon_s}{2} \right). 
$$

(4.4.1)

From the proof of Theorem 4.2.4 we see that the conclusion of this theorem remain to be true if

(i) for arbitrarily chosen $k$ in the $(k+1)$-th iteration, instead of the iterate $u^k$, an arbitrary point $u^k \in B_r^*/4+\theta_k$ is used for the construction of the function $\Psi_{k+1}$;

(ii) thereafter the iteration process is carried on under the condition

$$
\sum_{s=k+1}^{\infty} \left( \sqrt{2L(r)} \mu_s + 2\sigma_s + 2\mu_s + \frac{\epsilon_s}{2} \right) < \frac{r^*}{2} - \theta_k. 
$$

(4.4.2)

This observation enables us to ensure convergence of the OSR-methods by controlling the parameters in the following way (cf. (4.2.2), (4.2.5) and (4.2.6)).

Let the sequences $\{\bar{\mu}_k\}$, $\{\bar{\sigma}_k\}$ and $\{\bar{\epsilon}_k\}$ be chosen according to Theorem 4.2.4. Additionally, let positive sequences

$$
\{\bar{\mu}_k\}, \quad \{\bar{\sigma}_k\}, \quad \{\bar{\epsilon}_k\} 
$$

(4.4.3)
be given such that
\[
\sum_{k=1}^{\infty} \sqrt{\mu_k} < \infty, \quad \sum_{k=1}^{\infty} \sqrt{\sigma_k} < \infty, \quad \sum_{k=1}^{\infty} \sqrt{\epsilon_k} < \infty, \tag{4.4.4}
\]
and set
\[
\varrho_k := \sqrt{2L(r)\bar{\mu}_k + 2\bar{\sigma}_k + 2\bar{\mu}_k + \frac{\epsilon_k}{2}}.
\]
Starting with \(u^0 \in \mathbb{B}_{r^*} / 8\), in step \((k+1)\) of the OSR-method \textbf{4.2.1}, the inequality
\[
\|u^k\| \leq r^* / 4 + \theta_{k+1} - \varrho_{k+1} \tag{4.4.5}
\]
has to be observed. If it is wrong, we make a step under the usage of the parameters \(\bar{\mu}_{k+1}, \bar{\sigma}_{k+1}\) and \(\bar{\epsilon}_{k+1}\) in order to compute \(u^{k+1}\).

If inequality (4.4.5) is true, \(u^{k+1}\) has to be computed by using the parameters \(\bar{\mu}_{k+1}, \bar{\sigma}_{k+1}\) and \(\bar{\epsilon}_{k+1}\). In particular, this means that
\[
\sup_{u \in \mathbb{B}_r} |J(u) - J_{k+1}(u)| \leq \bar{\sigma}_{k+1},
\]
\[
\rho(Q_{k+1}, Q) \leq \bar{\mu}_{k+1}, \quad \rho(Q^*, Q_{k+1}) \leq \bar{\mu}_{k+1}
\]
have to be satisfied for a given \(k\).

For efficiency reasons the sequences \(\mu_k, \sigma_k, \epsilon_k\) and \(\bar{\mu}_k, \bar{\sigma}_k, \bar{\epsilon}_k\) should be coordinated. Otherwise the numerical expense in solving the next auxiliary problem may essentially increase.

Two substantial peculiarities of the procedure above should be noted. Firstly, its application will not disturb convergence of the OSR-method. If the conditions (4.4.4) and (4.4.4) are valid, then the corresponding sequence \(\{u^k\}\) belongs to \(\mathbb{B}_{r^*}\) and converges weakly to a solution of \(u^* \in U^*\) of the problem under consideration. Moreover, \(\lim_{k \to \infty} \|u^k - u^*\| = 0\) holds true if the functional \(J\) satisfies Assumption \textbf{4.2.3}. Secondly, an analysis of the properties of the proximal mapping and numerical experiments indicate that (4.4.5) can be maintained for the majority of real-life problems if the sequence \(\{\bar{\mu}_k\}\) harmonizes with the discretization strategies applied to semi-infinite problems and variational inequalities and if, moreover, the sequences \(\{\bar{\sigma}_k\}\) and \(\{\bar{\epsilon}_k\}\) are chosen according to the accuracy of the discretization. In that case the controlling of the method can be performed by means of the sequences (4.4.3) only.

\textbf{4.4.1 Remark.} In order to ensure a start with \(\bar{\mu}_1, \bar{\sigma}_1, \bar{\epsilon}_1\) and a substantial progress in the initial iterations, where the values of the controlling parameters are still rather large, it would be useful to require that in addition
\[
\frac{r^*}{8} + \theta_1 - \varrho_1 > r^0
\]
with some \(r^0 > 0\).

We assume that the reader is familiar with the basic futures of finite element methods, including the efficient handling of grid sequences. Probably, in the successive approximation of variational inequalities it makes sense to decrease the triangulation parameter slower than in multi-grid methods. The reason is that, on the one hand, the favorable properties of multi-grid methods
for linear problems are not guaranteed for variational inequalities in general, and, on the other hand, it is important to keep the dimension of the discretized and regularized auxiliary problems low as long as the iterates are far from the solution set $U^*$. This is especially important for OSR-methods.

An *a priori* choice of the sequences (4.4.3) is not obligatory. They can be corrected during execution of the OSR-method, particularly, in dependence of the values $\|u^k - u^{k-1}\|$ and $r^*/4 + \theta_{k+1} - \|u^k\|$. It may be helpful to choose $\mu_{k+1}, \sigma_{k+1}$ and $\epsilon_{k+1}$ close to $\mu_k, \sigma_k$ and $\epsilon_k$ if the distance $\|u^k - u^{k-1}\|$ is large (of order $O(\sqrt{2L(r)\mu_k + 2\sigma_k})$) and to accelerate the change of the parameters in case that $r^*/4 + \theta_{k+1} - \|u^k\|$ is small (close to $\theta_k$).

This consideration can be extended to MSR-methods if in the determination of the next parameter $\delta_k$ according to (4.3.13) the triple $(\mu_k, \sigma_k, \epsilon_k)$ (respectively $(\bar{\mu}_k, \bar{\sigma}_k, \bar{\epsilon}_k)$) is used which belongs to the $k$-th exterior level. Here the choice of the parameters (4.4.3) requires in the $k$-th level only to consider points $u^{k,0}$, in particular, (4.4.5) takes the form

$$\|u^{k,0}\| \leq r^*/4 + \theta_{k+1} - \theta_{k+1}.$$  

This conclusion is obvious if we use inequality (4.3.9) and the relation

$$\|u^{k,i} - u^{**}\| \leq \|u^{1,0} - u^{**}\| + \sum_{s=1}^{k-1} \left( \sqrt{2L(r)\mu_s + 2\sigma_s + 2\mu_s + \epsilon_s} \right), \quad (4.4.6)$$

which follows from the proof of Lemma 4.3.5 for $u^{**} \in U^* \cap B_{r^*/8}$ and all $k, i < i(k)$.

### 4.4.1.1 Controlling in case of unknown Lipschitz constants

If the calculation of a Lipschitz constant $L(r)$ is difficult, it is recommendable to choose the sequences $\{\mu_k\}, \{\sigma_k\}$ and $\{\epsilon_k\}$ in Method 4.2.1 such that

$$\sum_{k=1}^{\infty} \sqrt{\mu_k} < \infty, \quad \sum_{k=1}^{\infty} \sqrt{\sigma_k} < \infty, \quad \sum_{k=1}^{\infty} \sqrt{\epsilon_k} < \infty, \quad (4.4.7)$$

and to verify the inclusion $u^k \in B_r$. In case

$$u^k \in B_r \quad \forall k \quad (4.4.9)$$

convergence is guaranteed due to phase 3 in the proof of Theorem 4.2.4. We emphasize that in this part of the proof the real value of $L(r)$ is not important. It is only essential that the functional $J$ is Lipschitz continuous. Nevertheless, the validity of relation (4.4.9) depends on the value $L(r)$.

Another possibility is to treat the equivalent problem

$$\min\{z : J(u) - z \leq 0, \ u \in K\} \quad (4.4.10)$$

in the space $V \times \mathbb{R}$. In this case, the new objective functional satisfies the Lipschitz condition with $L \equiv 1$ on the whole space $V \times \mathbb{R}$.
CHAPTER 4. PPR FOR INFINITE-DIMENSIONAL PROBLEMS

However, from several reasons, this transformation may not be appropriate. If we deal, for instance, with minimization of a quadratic functional subject to linear constraints, the regularized problems belong to the same class and can be solved by special algorithms, including finite ones. However, the application of iterative regularization methods to Problem (4.4.10) needs universal algorithms of convex optimization. Concerning other transformations of the original problem see Remark 4.2.6.

4.4.1.2 Effective use of rough approximations

As mentioned in Section 2.2, in order to apply MSR-methods it is always important to perform the iterations on rough approximation levels efficiently. The this end the following procedures may be successful:

At a fixed exterior iteration level $k$, if $i(k) > 1$, compute

$$b_k := - \sum_{i=1}^{\hat{i}(k)-1} \left( \|u^{k,i} - u^{k,i-1}\| - \frac{\epsilon_k}{2} \right)^2 + 2(i(k) - 1) (L(r)\mu_k + \sigma_k + r\epsilon_k).$$

This value is negative due to (4.3.13). Then the iterations in the inner loop $i$ have to be continued while the inequalities

$$\|u^{k,i} - u^{k,i-1}\| > \frac{\epsilon_k}{2},$$

$$J(u^{k,i}) < J(u^{k,i-1}) - \frac{\epsilon_k^2}{4}$$

and

$$\sum_{s=i(k)}^{\hat{i}(k)} \left( \|u^{k,s} - u^{k,s-1}\| - \frac{\epsilon_k}{2} \right)^2 - 2(i - i(k) + 1) (L(r)\mu_k + \sigma_k + r\epsilon_k) > \theta b_k$$

are satisfied for a chosen $\theta \in (0,1)$.

This modification of the MSR-method will not disturb convergence. Indeed, following the proof of Lemma 4.3.5, relation (4.3.27) can be established by induction for $i \leq i(k_0)$. Then, Lemma 4.3.3 immediately gives $i(k_0) < \infty$ because, instead of (4.3.19), the improved relations

$$\|u^{k_0,i+1} - v^{k_0}\| - \|u^{k_0,i} - v^{k_0}\|$$

$$\leq \frac{1}{4r} \left( \left( \|u^{k_0,i+1} - u^{k_0,i}\| - \frac{\epsilon_{k_0}}{2} \right)^2 + 2L(r)\mu_{k_0} + 2\sigma_{k_0} + \frac{\epsilon_{k_0}}{2} \right)$$

have to be used for $i < i(k_0) - 1$. Summing up these inequalities for $i = 0, ..., i(k_0) - 1$, we obtain

$$\|u^{k_0,i(k_0)-1} - u^{**}\| \leq \|u^{k_0,0} - u^{**}\| + \frac{b_{k_0}}{4r}.$$  \hspace{1cm} (4.4.13)

In view of (4.4.12), (4.4.13) and the first inequality in (4.4.11), the estimates (4.3.27) can be obtained for $i = i(k_0) + 1, i(k_0) + 2, ..., \hat{i}(k_0)$, while the conditions (4.4.11) and (4.4.12) hold true. Finiteness of the inner loop $i$ for a fixed exterior level $k$ results immediately from (4.3.27) and the second inequality in (4.4.11). The remaining modifications in the proofs of Lemma 4.3.5 and Theorem 4.3.6 are trivial.
4.4. CHOICE OF CONTROLLING PARAMETERS

4.4.1.3 Choice of regularization parameters

In order to simplify the study of IPR we have assumed beginning with Section 4.1 that in the regularized functional

\[ \Psi_k(u) := J_k(u) + \frac{\chi_k}{2} \|u - u^k\|^2 \]

and in related expressions \( \chi_k := 2 \forall k \).

However, the efficiency of the methods depends substantially on the chosen values of \( \chi_k \).

The theoretical requirements open a great variety for that choice. For the exact proximal algorithm 3.1.5 weak convergence of the iterates to an element of the optimal solution set has been established by Brezis and Lions [52] in case

\[ \chi_k > 0, \quad \sum_{k=1}^{\infty} \frac{1}{\chi_k} = \infty, \]

whereas G"uler [153] has proved this convergence under the assumption

\[ \sum_{k=1}^{\infty} \frac{1}{\sqrt{\chi_k}} = \infty. \]

In simple situations, when OSR-methods are applied with \( K_k \equiv K, J_k \equiv J \) but also \( \chi_k \equiv \chi \), one can expect the following behavior based on estimates in [352] and confirmed numerically: Usually, a decrease of \( \chi \) will lead to increasing ill-conditioning of the regularized auxiliary problems and as a result we get a higher numerical expense for the outer loop. But in order to obtain a sufficiently exact solution of the original problem a relatively small number of outer iterations is required. Ibaraki et al. [190] have done numerical experiments with their own method and with a primal-dual proximal point algorithm suggested by Rockafellar [351]. They solved a number of network flow problems ( # of nodes \( \leq 400 \) and # of arcs \( \leq 4000 \)) with quadratic cost functionals including well-conditioned, ill-conditioned and ill-posed problems. Resulting from this analysis the following strategy for the choice of \( \{\chi_k\} \) proved to be efficient:

\[ \chi_1 := 1, \quad \chi_k := \begin{cases} \frac{\chi_{k-1}}{2} & \text{if } \chi_{k-1} > 0.01 \\ 0.01 & \text{otherwise} \end{cases} \]

In [165] a class of self-adaptive proximal point methods is proposed by Hager and Zhang. If the regularization parameter has the form \( \chi(u) := \beta \|\nabla J(u)\|^p \) where \( \eta \in [0, 2) \) and \( \beta > 0 \) is a constant, they obtain convergence to the set of minimizers that is linear for \( \eta := 0 \) and \( \beta \) sufficiently small, superlinear for \( \eta \in (0, 1) \), and at least quadratic for \( \eta \in [1, 2) \).

Using IPR for solving variational inequalities, it may be important to coordinate \( \chi_k \) with \( \kappa \), the value of the modulo of strong convexity of the energy functional \( J \) on the orthogonal complement of the kernel of the corresponding bilinear form. The goal is to make the regularized functional \( \Psi_k \) strongly convex on the whole space \( V \) with a constant which is close to \( \kappa \).

Variation of \( \chi_k \) makes sense, since the value \( \kappa \) is usually calculated by means of rough estimates. In the numerical experiments in Chapters 6 - 7, carried out for SIP, for contact problems in elasticity theory but also for control
problems, the value of $\chi_k$ are changing suitably and the ratio of maximal and minimal value of $\chi_k$ is comparatively big. In this sense OSR-and MSR-methods prove to be rather sensitive with respect to the choice of $\chi_k$.

4.4.1.4 Choice of penalty parameters

Concerning the penalty methods described in the Chapters 3 and 4 the choice of a starting value $r_1$ of the penalty parameter should be related to the starting point $u^0$ or $u^{1,0}$ in MSR-methods, respectively. otherwise it could be that in the initial iterations we move into "irregular directions" with an substantial increase of the values of the objective function.

This behavior can be demonstrated, for example, for the penalty method (3.2.1) applied with parameters $\chi_k := 2$, $\epsilon_k := 0$, for all $k$, and data

$$
J(u) := -10u, \\
\phi_k(u) := r_k(\max[0,u])^2
$$

(corresponding to the restriction $u \leq 0$),

$$
r_1 := 1, \quad r_k := 2r_{k-1}, \quad k = 2, 3, \ldots, \quad u^0 = 0.1.
$$

In fact, for infinite and semi-infinite problems these undesirable iterations are leading to an increase of the dimension of the approximate auxiliary problems. The same happens in the case of MSR-methods.

In the sequel we describe a procedure for choosing a start value for the penalty parameter suggested by GROSSMANN AND KAPLAN [151].

If, for instance, Problem (A3.4.56) is solved by the penalty method (3.2.1) with $\chi_k := 2$ and penalty function (A3.4.78), then the point $u^1$ can be determined by an non-exact minimization of the function

$$
d(u) := -\frac{1}{J(u) + \|u - u^0\|^2} - \sum_{j=1}^{m} \frac{1}{g_j(u)}
$$

on the open set

$$
\{ u \in \text{int} K : J(u) + \|u - u^0\|^2 < t \},
$$

where $t > J(u^0)$ is chosen close to $J(u^0)$.

Obviously, this point can be computed by means of the same algorithm for unconstrained minimization which is also chosen to minimize the auxiliary problems in the penalty method. Computing on this way $u^1$, as a stopping criterion the condition

$$
\|\nabla d(u)\| \leq \frac{\epsilon_1}{c_0},
$$

with $c_0$ a lower bound of $\min_{u \in \Pi} \{ J(u) + \|u - u^0\|^2 \} - t$, can be used. Then the following relations are satisfies:

$$
J(u^1) + \|u^1 - u^0\|^2 < t
$$

and

$$
\|\nabla J(u^1) + 2(u^1 - u^0) + (J(u^1) + \|u^1 - u^0\|^2 - t)g_j(u^1)\| \leq \epsilon_1.
$$
Therefore, \( u^1 \) is an approximate minimum of the function

\[
F_1(u) := J(u) + \phi_1(u) + \|u - u^0\|^2
\]

and satisfies the condition

\[
\|\nabla F_1(u^1)\| \leq \epsilon_1,
\]

with

\[
\phi_1(u) := \begin{cases} 
- \frac{1}{r_1} \sum_{j=1}^{m} \frac{1}{g_j(u)} & \text{if } u \in \text{int} K \\
+\infty & \text{if } u \notin K
\end{cases}
\]

a penalty function of type (A3.4.78) and

\[
\epsilon_1 := \frac{1}{(J(u^1) + \|u^1 - u^0\|^2 - t)^2}.
\]

Now, choosing \( r_2 \geq r_1 \), we can continue with method (3.2.1).

### 4.4.2 Stopping rules for inner and outer loops

Simple examples show that, in general, when the exact proximal point method 3.1.5 terminates at a given distance \( \|u^k - u^{k-1}\| \) of two consecutive iterates or a given distance \( |J(u^k) - J^*| \) of the current function value and \( J^* \) (see Lemma 4.1.5), then we cannot conclude that the iterates \( u^k \) are close to the solution set in the metric of the initial space. This situation is typical for certain methods applied to ill-posed problems (see also POLYAK [331], chapt. 1 and 6).

Nevertheless we obtain in IPR-methods a stable sequence of iterates toward some solution point by using a stopping rule which applies the distance between function values. This is often decisive for the quality of the approximate solutions.

If Assumption 5.1.4 below (or 5.1.6) is fulfilled and if the controlling parameters decrease with geometrical progression, one may expect a rate of convergence like \( O\left(\frac{1}{k^2}\right) \) or even a linear rate with respect to the iterates \( u^k \). More precisely, the stopping rule for the outer loop in MSR-method 4.3.1 is based on Theorem 4.3.10 and Remark 4.3.11, and the algorithm terminates at a point \( u^{k,i(k)} \) where

\[
4r\delta_k + \frac{4r + L(r)}{2} \epsilon_k + L(r) \mu_k + 2\delta_k < \bar{\epsilon},    \tag{4.4.14}
\]

with \( \bar{\epsilon} \) suitably chosen. In that case, due to (4.3.40) and (4.3.42), the estimate

\[
|J(u^{k,i(k)}) - J^*| < \bar{\epsilon}
\]

is true.

For OSR-methods, taking into consideration (4.3.42) and the relation

\[
J(u^k) - J^* \leq 4r\|u^k - u^{k-1}\| + \frac{4r + L(r)}{2} \epsilon_k + L(r) \mu_k + 2\sigma_k,
\]

the stopping condition is

\[
4r\|u^k - u^{k-1}\| + \frac{4r + L(r)}{2} \epsilon_k + L(r) \mu_k + 2\sigma_k < \bar{\epsilon}. \tag{4.4.15}
\]
Usually, in order to solve well-posed problems, as a rule the accuracy in solving approximate problems has to be chosen in the same order as the accuracy of the approximation of the original problem. But, solving ill-posed problems by means of IPR, such a strong rule is superfluous up to a decisive moment. Indeed, as mentioned, the influence of the regularization at the beginning of the solution process may drive us far away from the sought solution even if the approximation of the data is sufficiently precise.

Hence, at the beginning of the solution process the choice of \( \epsilon_k \) should be basically oriented on the conditions of the convergence of the method considered. However, we have to regard that the choice of \( \epsilon_k \) influences the values of \( \delta_k \) in MSR-methods (see condition (4.3.13)), and with an increase of \( \epsilon_k \) the number of iterations in the interior loop cuts down at the \( k \)-th level. This reduces the efficiency of the method. In the final stage of the MSR-algorithm we recommend to choose \( \epsilon_k \) in the same order as \( \mu_k \) and \( \sigma_k \).

Concerning the solution of variational inequalities, in the starting phase, especially for MSR-methods, an opposite strategy may be profitable: If the discretization initially on rough grids generates a quadratic programming problem with simple constraints, often the application of exact methods is more efficient. Moreover, the choice of small or zero values for \( \epsilon_k \) enables us to use smaller values for \( \delta_k \) and, hence, we obtain better approximate solutions of the original problem on these rough grids.

### 4.5 Comments

**Section 4.2:** The results of this section are based on the papers of Kaplan and Tichatschke [208, 209, 211, 213]. Theorem 4.2.7 is new. Lemaire [261] used another framework for investigating iterative PPM’s. He considered unconstrained problems

\[
\min_{u \in V} J(u), \quad \min_{u \in V} \{J_k(u) + \|u - u^{k-1}\|^2\}
\]

and the distance between the convex, lsc functionals \( J \) and \( J_k \) are measured in terms of Moreau-Yoshida-approximations. In his paper the assumptions on the accuracy of the approximation are not predestinated when finite element methods are applied to variational inequalities with differential operators.

**Section 4.3:** Apparently, the first results concerning MSR-methods are given in [213], where two variants of MSR-methods are suggested for solving convex, semi-infinite problems and the penalty method is used as a basic algorithm. Most of the results of this section are proved for the first time in [214], except for MSR-methods with regularization on a subspace, which have been not published before.
Chapter 5

RATE OF CONVERGENCE FOR PPR

In this chapter we analyze the rate of convergence of MSR-Method 4.3.1. It turns out that for a wide class of ill-posed convex variational problems a linear rate of convergence can be established for functional values as well as for iterates. Moreover, for Problem (4.2.1), which is quite general, estimates of the objective values will be obtained that cannot be improved substantially.

In Section 5.2 this analysis will be adapted to OSR-Method 4.2.8.

In Section 5.3 by means of the regularized penalty method (3.2.2) it is shown how the approach developed in Section 4.1 may be used in situations, where the conditions on approximation (cf. Assumption 4.2.2) are disturbed.

5.1 Rate of Convergence for MSR-Methods

5.1.1 Convergence results for the general case

To start with we take in MSR-Method 4.3.1

\[ \sigma_0 := \sigma_1, \quad \mu_0 := \mu_1, \quad \epsilon_0 := \epsilon_1, \]

\[ \bar{u}^{k,0} := \bar{u}^{k-1,i(k-1)} \quad \text{(for } k > 1). \]

For readers convenience, in accordance with the notation introduced in Subsection 4.3.2, we use the following abbreviations:

\[ \xi_1^k := \frac{1}{2}(L(r) + 4r)\epsilon_k + 2\sigma_k, \]

\[ \xi_2^k := L(r)(\mu_k + \frac{\epsilon_k}{2}), \]

139
\( \zeta'_k := 6 \sigma_k + L(r) \left( 3 \mu_k + \frac{\epsilon_k}{2} \right) + r \left( 2 \mu_k + \frac{\epsilon_k}{2} \right)^2 + \xi_k^2 + \xi_1^k, \)

\( \zeta''_k := \frac{1}{8r^2} \left( 2 \sigma_k + L(r) \mu_k + \xi_k^k \right) (2rL(r) + \xi_k^k) + 4 \sigma_k + (L(r) + r) \left( 2 \mu_k + \frac{\epsilon_k}{2} \right)^2 + \xi_k^2 + \xi_1^k, \)

\( \zeta_k := \max[\zeta'_k, \zeta''_k], \)

\( \varrho_{k,0} := J(u^{k,0}) - J^* + \xi_{k-1}^k, \)  \( (5.1.1) \)

\( \varrho_{k,i} := J(u^{k,i}) - J^* + \xi_k^k, \quad 0 \leq i \leq i(k), \)  \( (5.1.2) \)

\( \hat{i}(k) := i(k) + \max[0, i(k) - 3]. \)

Note that the notation \( \sigma_0, \mu_0 \) and \( \epsilon_0 \) is used to define formally for \( k := 0 \) the constant \( \xi_0^k \) in (5.1.1) and some other values which will be introduced later on.

Concerning the rate of convergence of the objective values of the iterates in MSR-Method 4.3.1 the following result can be proved.

5.1.1 Theorem. Suppose the assumptions of Lemma 4.3.5 are fulfilled, \( u^0 := u^{1,0} \in K \) and, moreover,

(i) \( \sigma_{k+1} \leq \sigma_k, \quad \epsilon_{k+1} \leq \epsilon_k, \quad \mu_{k+1} \leq \mu_k, \quad \rho(Q_k, Q_{k+1}) \leq 2 \mu_k, \quad k = 1, 2, \ldots, \)

\( \epsilon_1 < 1, \quad \mu_1 < 1, \quad J(u^{1,0}) > J^*; \)

(ii) constant \( \beta \) is fixed such that

\[ 0 < \beta < \frac{1}{4} \min \left[ \frac{1}{L(r) (2r + \mu_1 + \frac{\epsilon_1}{2})}, \frac{1}{8r^2} \right]; \]  \( (5.1.3) \)

(iii) constants \( \zeta_k \) satisfy

\[ \zeta_0 = \zeta_1 \leq \beta \left( \frac{\varrho_{1,0}}{1 + 3 \beta \varrho_{1,0}} \right)^2, \]

\[ \zeta_{k-1} + 2 \zeta_k \leq \left( \delta_k - \frac{\epsilon_k}{2} \right)^2 =: \hat{\epsilon}_k^2, \]  \( (5.1.4) \)

\[ \zeta_k \leq \frac{\beta}{4} \left( \frac{\varrho_{1,0}}{1 + \beta \left( \sum_{j=1}^{k-1} \hat{i}(j) + 3 \right) \varrho_{1,0}} \right)^2 \text{ for } k \geq 2. \]

Then it holds for MSR-Method 4.3.1

\[ \varrho_{k,1} < \frac{\varrho_{1,0}}{1 + \beta \left( \sum_{j=1}^{k-1} \hat{i}(j) + 1 \right) \varrho_{1,0}}, \quad k = 1, 2, \ldots. \]  \( (5.1.5) \)

Proof: Due to (4.1.3), (4.2.5), (4.3.9), Lemma 4.3.5 and the strong convexity of the functionals \( \Psi_{k,i} \), we get for all \( k \) and \( 0 \leq i \leq i(k) - 1, \)

\[ \Psi_{k,i+1}(u^{k,i+1}) - \Psi_{k,i+1}(\bar{u}^{k,i+1}) \leq \xi_1^k, \]  \( (5.1.6) \)
and from (4.3.42)
\[ J(u^{k,i+1}) - J^* \geq -\xi^k_2. \] (5.1.7)

Let
\[
\begin{align*}
\theta^{k,i} &:= \arg\min\{\|u^{k,i} - v\| : v \in Q^*\}, \quad Q^* := U^* \cap B_{r^*}, \\
\tilde{\theta}^{k,i} &:= \arg\min\{\|u^{k,i} - v\| : v \in Q_k\}, \quad Q_k := K_k \cap B_r, \\
\omega^{1,0} &:= \arg\min\{\|u^{1,0} - w\| : w \in Q_1\}, \\
\omega^{k,0} &:= \arg\min\{\|u^{k,0} - w\| : w \in Q_k\}, \quad \text{for } k > 1, \\
\eta_{i,k}(\lambda) &:= \lambda J_k(\tilde{\theta}^{k,i}) + (1 - \lambda)J_k(u^{k,i}) + \lambda^2\|\hat{u}^{k,i} - u^{k,i}\|^2.
\end{align*}
\]

Due to (5.1.6), the convexity of $J_k$, $Q_k$ and the definition of $\tilde{\theta}^{k,i}$ in (4.3.3), we have for $1 \leq i < i(k)$ and $\lambda \in [0, 1]$
\[
\begin{align*}
J_k(\theta^{k,i+1}) + \|u^{k,i+1} - u^{k,i}\|^2 &
\leq \lambda J_k(\theta^{k,i}) + (1 - \lambda)J_k(\tilde{\theta}^{k,i}) + \|\lambda \hat{u}^{k,i} + (1 - \lambda)u^{k,i} - u^{k,i}\|^2 + \xi_1^k \\
&\leq \lambda J_k(\theta^{k,i}) + (1 - \lambda)J_k(u^{k,i}) + (1 - \lambda)(J_k(\tilde{\theta}^{k,i}) - J_k(u^{k,i})) \\
&\quad + \lambda^2\|\hat{u}^{k,i} - u^{k,i}\|^2 + (1 - \lambda)^2\|\hat{u}^{k,i} - u^{k,i}\|^2 \\
&\quad + 2\lambda(1 - \lambda)\|\hat{u}^{k,i} - u^{k,i}\||\hat{u}^{k,i} - u^{k,i}\| + \xi_1^k.
\end{align*}
\]

With regard to $\lambda(1 - \lambda) \leq \frac{1}{4}$ and the relations (4.1.3), (4.2.5), (4.3.9), the estimation above can be continued such that
\[
\begin{align*}
J_k(u^{k,i+1}) + \|u^{k,i+1} - u^{k,i}\|^2 &
\leq \lambda J_k(\theta^{k,i}) + (1 - \lambda)J_k(u^{k,i}) + \lambda^2\|\hat{u}^{k,i} - u^{k,i}\|^2 + |J_k(u^{k,i}) - J_k(\tilde{\theta}^{k,i})| \\
&\quad + \frac{1}{2}\|\hat{u}^{k,i} - u^{k,i}\||\hat{u}^{k,i} - u^{k,i}\| + \|\tilde{\theta}^{k,i} - u^{k,i}\|^2 + \xi_1^k \\
&\leq \eta_{i,k}(\lambda) + \xi_3^k, \quad (5.1.8)
\end{align*}
\]

with
\[
\xi_3^k := 2\sigma_k + \frac{1}{2}(L(r) + r)\epsilon_k + \frac{\epsilon_3^k}{4} + \xi_1^k.
\]

For $k > 1$, $i = 0$ we notice that $\tilde{\theta}^{k,0} \in Q_{k-1}$, but $\tilde{\theta}^{k,1} \in Q_k$.

Using (4.1.3), (4.3.9) and Assumption 4.2.2, this leads to
\[
\begin{align*}
J_k(u^{k,1}) + \|u^{k,1} - u^{k,0}\|^2 &
\leq \lambda J_k(\theta^{k,0}) + (1 - \lambda)J_k(u^{k,0}) + \|\lambda \hat{u}^{k,0} + (1 - \lambda)u^{k,0} - u^{k,0}\|^2 + \xi_1^k \\
&\leq \lambda J_k(\theta^{k,0}) + (1 - \lambda)J_k(u^{k,0}) + \lambda^2\|\hat{u}^{k,0} - u^{k,0}\|^2 + |J_k(u^{k,0}) - J_k(u^{k,0})| \\
&\quad + \frac{1}{2}\|\hat{u}^{k,0} - u^{k,0}\||\hat{u}^{k,0} - u^{k,0}\| + \|u^{k,0} - u^{k,0}\|^2 + \xi_1^k \\
&\leq \eta_{k,0}(\lambda) + \xi_4^k, \quad (5.1.9)
\end{align*}
\]

with
\[
\xi_4^k := 2\sigma_k + (L(r) + r)\left(2\mu_{k-1} + \frac{\epsilon_{k-1}}{2}\right) + \left(2\mu_{k-1} + \frac{\epsilon_{k-1}}{2}\right)^2 + \xi_1^k.
\]
Finally, for $k = 1$, $i = 0$ we get

$$J_1(u^{1,1}) + ||u^{1,1} - u^{1,0}||^2$$

$$
\leq \lambda J_1(\bar{v}^{1,0}) + (1 - \lambda) J_1(w^{1,0}) + \|\lambda \bar{v}^{1,0} + (1 - \lambda) w^{1,0} - u^{1,0}\| + \xi_1^4
$$

$$
\leq \eta_1(\lambda) + |J_1(u^{1,0}) - J_1(w^{1,0})| + \frac{1}{2} ||\bar{v}^{1,0} - u^{1,0}|| ||u^{1,0} - u^{1,0}||
$$

$$
+ ||u^{1,0} - u^{1,0}||^2 + \xi_1^4
$$

$$
\leq \eta_1(\lambda) + 2\sigma_1 + L(r)\mu_0 + r\mu_0 + \mu^2_0 + \xi_1^4
$$

$$\leq \eta_1(\lambda) + \xi_1^4.
$$

In this way, for each $k$ and $0 \leq i < i(k)$, we conclude that

$$J_k(u^{k,i+1}) + ||u^{k,i+1} - u^{k,i}||^2 \leq \eta_{k,i}(\lambda) + \xi_{0,k}^4,$$

(5.1.10)

with $\xi_{0,k}^4 := \xi_{k}^4$ for $i > 0$ and $\xi_{0,0}^4 := \xi_{1}^4$ for $i = 0$.

Minimizing the quadratic function $\eta_{k,i}$ on the interval $[0, 1]$, the following three cases have to be analyzed:

(a) $\lambda_{k,i} = \arg\min \eta_{k,i}(\lambda)$.

Hence, $\eta_{k,i}(\lambda_{k,i}) = J_k(u^{k,i})$ and $J_k(u^{k,i}) \leq J_k(\bar{v}^{k,i})$.

Therefore, the following estimates are true:

$$J_k(u^{k,i+1}) + ||u^{k,i+1} - u^{k,i}||^2 \leq J_k(u^{k,i}) + \xi_{0,k}^4 \leq J_k(\bar{v}^{k,i}) + \xi_{k}^4,$$

$$J_1(u^{k,i+1}) + ||u^{k,i+1} - u^{k,i}||^2 \leq J_1(\bar{v}^{k,i}) + 2\sigma_k + \xi_{0,k}^4,$$

and

$$J(u^{k,i+1}) - J^* + \xi_{2}^4$$

$$\leq J(\bar{v}^{k,i}) - J^* + \xi_{0,k}^4 + \xi_{2}^4 + 2\sigma_k - ||u^{k,i+1} - u^{k,i}||^2$$

$$= J(\bar{v}^{k,i}) - J(\bar{v}^{k,i}) + \xi_{0,k}^4 + \xi_{2}^4 + 2\sigma_k - ||u^{k,i+1} - u^{k,i}||^2.$$  

Due to (4.2.6), (5.1.1) and (5.1.2), this leads to

$$\varrho_{k,i+1} \leq \xi_{2}^4 + \xi_{0,k}^4 + L(r)\mu_k + 2\sigma_k - ||u^{k,i+1} - u^{k,i}||^2.$$

(5.1.11)

(b) $\lambda_{k,i} = 1$ implies $\eta_{k,i}(\lambda_{k,i}) = J_k(\bar{v}^{k,i}) + ||\bar{v}^{k,i} - u^{k,i}||^2$ and

$$\frac{d}{d\lambda} \eta_{k,i}(\lambda)|_{\lambda=1} \leq 0$$

is obvious, i.e.,

$$\frac{J_k(u^{k,i}) - J_k(\bar{v}^{k,i})}{2||\bar{v}^{k,i} - u^{k,i}||^2} \geq 1 = \lambda_{k,i}.$$

Thus

$$\eta_{k,i}(\lambda_{k,i}) \leq \frac{1}{2} (J_k(u^{k,i}) + J_k(\bar{v}^{k,i}))$$

is valid and with regard to $J(\bar{v}^{k,i}) \leq J^* + L(r)\mu_k$ we conclude that

$$\varrho_{k,i+1} \leq \frac{1}{2} (J_k(u^{k,i}) + J^* + L(r)\mu_k) - J^* + \xi_{0,k}^4 + \xi_{2}^4 + 2\sigma_k - ||u^{k,i+1} - u^{k,i}||^2.$$
5.1. RATE OF CONVERGENCE FOR MSR-METHODS

In order to combine the cases \( i = 0 \) and \( i > 0 \), we utilize that \( \xi^k_2 \leq \xi^{k-1}_2 \), hence for \( 0 \leq i < i(k) \) the inequalities

\[
\varrho_{k,i} \geq J(u^{k,i}) - J^* + \xi^k_2
\]

and

\[
\varrho_{k,i+1} \leq \frac{1}{2} \varrho_{k,i} + \frac{1}{2} L(r) \mu_k + \frac{1}{2} \xi^k_2 + \xi_0^k + 2 \sigma_k - \|u^{k,i+1} - u^{k,i}\|^2 \quad (5.1.12)
\]

are true.

(c) If \( 0 < \lambda_{k,i} < 1 \) then

\[
\lambda_{k,i} = \frac{J_k(u^{k,i}) - J_k(\tilde{v}^{k,i})}{2\|\tilde{v}^{k,i} - u^{k,i}\|^2}
\]

and in view of (5.1.10) we have

\[
J_k(u^{k,i+1}) \leq J_k(u^{k,i}) - \frac{(J_k(u^{k,i}) - J_k(\tilde{v}^{k,i}))^2}{4\|\tilde{v}^{k,i} - u^{k,i}\|^2} + \xi_0^k - \|u^{k,i+1} - u^{k,i}\|^2. \quad (5.1.13)
\]

But

\[
J_k(u^{k,i}) - J_k(\tilde{v}^{k,i}) \geq J(u^{k,i}) - J(v^{k,i}) - 2 \sigma_k - L(r) \mu_k
\]

and we have to distinguish two cases:

(o) If \( J(u^{k,i}) - J(v^{k,i}) > 2 \sigma_k + L(r) \mu_k \), then

\[
(J_k(u^{k,i}) - J_k(\tilde{v}^{k,i}))^2 \geq \left(J_k(u^{k,i}) - J_k(\tilde{v}^{k,i}) + \hat{\xi}^k\right)^2 - 2 \left(2 \sigma_k + L(r) \mu_k + \hat{\xi}^k\right) \left(J_k(u^{k,i}) - J_k(\tilde{v}^{k,i}) + \hat{\xi}^k\right) + \left(2 \sigma_k + L(r) \mu_k + \hat{\xi}^k\right)^2 > 0
\]

holds with \( \hat{\xi}^k := \xi^k_2 \) for \( i > 0 \) and \( \hat{\xi}^k := \xi^{k-1}_2 \) for \( i = 0 \).

Due to (4.1.3) and Lemma 4.3.5, the latter inequality leads to

\[
\frac{(J_k(u^{k,i}) - J_k(\tilde{v}^{k,i}))^2}{4\|\tilde{v}^{k,i} - u^{k,i}\|^2} > \frac{1}{16 \sigma^2} \left[ g_{k,i}^2 - 2 \left(2 \sigma_k + L(r) \mu_k + \hat{\xi}^k\right) \left(J(u^{k,i}) - J(v^{k,i}) + \hat{\xi}^k\right) + \left(2 \sigma_k + L(r) \mu_k + \hat{\xi}^k\right)^2 \right]
\]

\[
> \frac{1}{16 \sigma^2} \left[ g_{k,i}^2 - 2 \left(2 \sigma_k + L(r) \mu_k + \hat{\xi}^k\right) \left(2L(r)r + \hat{\xi}^k\right) \right]. \quad (5.1.14)
\]

However, (5.1.13) implies

\[
J(u^{k,i+1}) - J^* + \xi^k_2 \leq J(u^{k,i}) - J^* + \xi^k_2
\]

\[
- \frac{(J_k(u^{k,i}) - J_k(\tilde{v}^{k,i}))^2}{4\|\tilde{v}^{k,i} - u^{k,i}\|^2} + \xi_0^k + 2 \sigma_k - \|u^{k,i+1} - u^{k,i}\|^2,
\]

\[
\frac{1}{16 \sigma^2} \left[ g_{k,i}^2 - 2 \left(2 \sigma_k + L(r) \mu_k + \hat{\xi}^k\right) \left(2L(r)r + \hat{\xi}^k\right) \right].
\]

(5.1.14)
i.e.,
\[ \varrho_{k,i+1} \leq \varrho_{k,i} - \frac{(J_k(u_{k,i}^i) - J_k(u_{k,i}^{-i}))^2}{4\|u_{k,i}^i - u_{k,i}^{-i}\|^2} + \xi_0^k + 2\sigma_k - \|u_{k,i+1}^{k,i} - u_{k,i}^{k,i}\|^2 \]
and with regard to (5.1.14)
\[ \varrho_{k,i+1} < \varrho_{k,i} - \frac{2}{16r^2} (2\sigma_k + L(r)\mu_k + \xi_k^k) \left( 2L(r)r + \xi_k^k \right) + 2\sigma_k + \xi_0^k - \|u_{k,i}^{k,i+1} - u_{k,i}^{k,i}\|^2. \]  

(\circ) If \( J(u_{k,i}^{k,i}) - J(v_{k,i}^{k,i}) \leq 2\sigma_k + L(r)\mu_k \), then using inequality (5.1.10) for \( \lambda = 0 \), we get
\[ J(u_{k,i}^{k,i+1}) + \|u_{k,i}^{k,i+1} - u_{k,i}^{k,i}\|^2 \leq J(v_{k,i}^{k,i}) + 4\sigma_k + L(r)\mu_k + \xi_0^k, \]
which leads immediately to
\[ \varrho_{k,i+1} < 4\sigma_k + L(r)\mu_k + \xi_0^k + \xi_0^k - \|u_{k,i}^{k,i+1} - u_{k,i}^{k,i}\|^2. \]  

Because of \( \xi_0^k \leq \xi_2^{k-1} \) and \( \xi_2^0 = \xi_2^1 \), the inequality
\[ \varrho_{k,i} < L(r) \left( 2r + \mu_{k-1} + \frac{\xi_{k-1}}{2} \right) \]
is obvious and, due to the monotonicity assumed for \( \{\mu_k\} \) and \( \{\epsilon_k\} \), we obtain
\[ \varrho_{k,i} < L(r) \left( 2r + \mu_1 + \frac{\epsilon_1}{2} \right), \quad \forall \ k, \forall \ i. \]
This leads together with (5.1.3) to
\[ 2\beta \varrho_{k,i} < \frac{\varrho_{k,i}}{2L(r)(2r + \mu_1 + \frac{\epsilon_1}{2})} < \frac{1}{2}, \]
hence,
\[ \frac{1}{2} \varrho_{k,i} - 2\beta \varrho_{k,i}^2 \geq 0. \]  

Taking into account the definition of \( \zeta_k \), it can be easily seen that for corresponding pairs \((k,i)\) the inequalities
\[ \varrho_{k,i+1} \leq \varrho_{k,i} - 2\beta \varrho_{k,i}^2 + \zeta_k - \|u_{k,i+1}^{k,i} - u_{k,i}^{k,i}\|^2, \quad (i > 0) \]  
and
\[ \varrho_{k,1} \leq \varrho_{k,0} - 2\beta \varrho_{k,0}^2 + \zeta_{k-1} - \|u_{k,1}^{k,1} - u_{k,0}^{k,0}\|^2 \]
are conclusions from any of the inequalities (5.1.11), (5.1.12), (5.1.15) and (5.1.16) above. Thus, estimate (5.1.18) holds for each \( k \) and \( 0 < i \leq i(k) - 1 \), and estimate (5.1.19) is satisfied for each \( k \).

It should be noted that the inequality \( \zeta_k \geq \zeta_k' \) was used for the transition
from (5.1.11), (5.1.12) to (5.1.18), (5.1.19), whereas for the transition of the estimates (5.1.15), (5.1.16) the inequality $\zeta_k \geq \zeta_k'$ was applied.

Due to assumption (iii) and $J(u^{1,0}) > J^*$, we get from (5.1.19)

\[ \varrho_{k+1} \leq \varrho_{k+1,0} - 2\beta \varrho_{k+1,0}^2 + \frac{\beta}{4} \left( \frac{\varrho_{k+1,0}}{1 + 3\beta \varrho_{k+1,0}} \right)^2 - \|u^{k+1,1} - u^{k+1,0}\|^2 \]

\[ < \varrho_{k+1,0}(1 - \beta \varrho_{k+1,0}) - \beta \varrho_{k+1,0}^2 + \frac{\beta}{4} \varrho_{k+1,0}^2 - \|u^{k+1,1} - u^{k+1,0}\|^2 \]

\[ < \frac{\varrho_{k+1,0}}{1 + \beta \varrho_{k+1,0}} - \|u^{k+1,1} - u^{k+1,0}\|^2. \]

Continuing the proof we assume that, for fixed $k$ and an integer $\gamma > 0$, the inequalities

\[ \varrho_{k,0} \leq \varrho_{k,0}^1 \quad \text{and} \quad \varrho_{k,1} < \varrho_{k,1}^1 \]

hold and, moreover, that

\[ \zeta_k \leq \beta \left( \frac{\varrho_{k,1}}{1 + \beta (\gamma + 2) \varrho_{k,1}} \right) \].

(1) In case $i(k) = 1$ we have $\varrho_{k+1,0} = \varrho_{k,1}(k) = \varrho_{k,1}$. Then

\[ \varrho_{k+1,1} < \frac{\varrho_{k,1}}{1 + \beta (\gamma + 1) \varrho_{k,1}} - \|u^{k+1,1} - u^{k+1,0}\|^2 \]

is true. Indeed, if $\zeta_k \geq \beta \varrho_{k+1,0}^2$, i.e.,

\[ \frac{\beta}{4} \left( \frac{\varrho_{k,1}}{1 + \beta (\gamma + 2) \varrho_{k,1}} \right)^2 \geq \beta \varrho_{k+1,0}^2 \]

then

\[ \varrho_{k+1,0} \leq \frac{\varrho_{k,1}}{2(1 + \beta (\gamma + 2) \varrho_{k,1})} \]

holds and, due to (5.1.19),

\[ \varrho_{k+1,1} \leq \varrho_{k+1,0} + \zeta_k - \|u^{k+1,1} - u^{k+1,0}\|^2 \]

\[ < \frac{\varrho_{k,1}}{2(1 + \beta (\gamma + 2) \varrho_{k,1})} + \frac{\beta}{4} \left( \frac{\varrho_{k,1}}{1 + \beta (\gamma + 2) \varrho_{k,1}} \right)^2 - \|u^{k+1,1} - u^{k+1,0}\|^2 \]

\[ < \frac{\varrho_{k,1}}{2(1 + \beta (\gamma + 2) \varrho_{k,1})} + \frac{1}{4(\gamma + 2)} \frac{\varrho_{k,1}}{1 + \beta (\gamma + 2) \varrho_{k,1}} - \|u^{k+1,1} - u^{k+1,0}\|^2 \]

\[ < \frac{\varrho_{k,1}}{1 + \beta (\gamma + 1) \varrho_{k,1}} - \|u^{k+1,1} - u^{k+1,0}\|^2. \]
If $\zeta_k < \beta \varrho_{k+1,0}$, then with regard to (5.1.19)

$$g_{k+1,1} < g_{k+1,0} - \beta \varrho_{k+1,0} - \|u^{k+1,1} - u^{k+1,0}\|^2.$$ 

For $0 \leq t \leq \frac{1}{2}$ function $t - t^2$ increases and $\beta g_{k,0} \leq \frac{1}{4}$ holds according to the chosen $\beta$ in (5.1.3). Therefore, in view of $g_{k+1,0} = g_{k,1}$ and (5.1.21),

$$g_{k+1,1} < \frac{g_{1,0}}{1 + \beta \gamma g_{1,0}} - \beta \left( \frac{g_{1,0}}{1 + \beta \gamma g_{1,0}} \right)^2 - \|u^{k+1,1} - u^{k+1,0}\|^2$$

$$< \frac{g_{1,0}}{1 + \beta \gamma g_{1,0}} \left( 1 - \frac{\beta g_{1,0}}{1 + \beta (\gamma + 1) g_{1,0}} \right) - \|u^{k+1,1} - u^{k+1,0}\|^2$$

$$= \frac{g_{1,0}}{1 + \beta (\gamma + 1) g_{1,0}} - \|u^{k+1,1} - u^{k+1,0}\|^2.$$

(2) If $i(k) = 2$, inequality (5.1.18) implies

$$g_{k+1,0} \equiv g_{k,2} \leq g_{k,1} - 2\beta \varrho_{k,1}^2 + \zeta_k.$$

Proceeding analogously to case (1), we obtain at first

$$g_{k+1,0} < \frac{g_{1,0}}{1 + \beta (\gamma + 1) g_{1,0}}$$

and afterwards, according to the restriction for $\zeta_k$ and the validity of

$$g_{k+1,1} \leq g_{k+1,0} - 2\beta \varrho_{k+1,0}^2 + \zeta_k - \|u^{k+1,1} - u^{k+1,0}\|^2,$$

we infer

$$g_{k+1,1} < \frac{g_{1,0}}{1 + \beta (\gamma + 2) g_{1,0}} - \|u^{k+1,1} - u^{k+1,0}\|^2.$$ 

(3) But if $i(k) > 2$, the following inequalities are true:

$$g_{k,1} \leq g_{k,0} - 2\beta \varrho_{k,0}^2 + \zeta_{k-1} - \|u^{k,1} - u^{k,0}\|^2,$$

$$g_{k,2} \leq g_{k,1} - 2\beta \varrho_{k,1}^2 + \zeta_k - \|u^{k,2} - u^{k,1}\|^2,$$

$$\vdots$$

$$g_{k,i(k)-1} \leq g_{k,i(k)-2} - 2\beta \varrho_{k,i(k)-2}^2 + \zeta_{k-1} - \|u^{k,i(k)-1} - u^{k,i(k)-2}\|^2,$$

$$g_{k+1,0} \equiv g_{k,i(k)} \leq g_{k,i(k)-1} - 2\beta \varrho_{k,i(k)-1}^2 + \zeta_{k-1},$$

$$g_{k+1,1} \leq g_{k+1,0} - 2\beta \varrho_{k+1,0}^2 + \zeta_k - \|u^{k+1,1} - u^{k+1,0}\|^2.$$ (5.1.22)

Using the first inequality of this row together with

$$\zeta_{k-1} - \|u^{k,1} - u^{k,0}\|^2 < 0,$$

which follows from

$$\zeta_{k-1} + 2\zeta_k \leq \tau_k^2, \quad i(k) > 2,$$
5.1. **Rate of Convergence for MSR-Methods**

Finally, in each of the cases (1)-(3) the estimate

\[ q - 2\beta q^2 < \frac{q}{1 + 2\beta q}, \quad (q > 0, \ \beta > 0), \]  

(5.1.25)

we conclude that

\[ \varrho_{k,1} > \frac{\varrho_{1,0}}{1 + \beta(\gamma + 1)\varrho_{1,0}}. \]

Due to (5.1.4) and \(\|u^{k,i} - u^{k,i-1}\| > \delta_k\) for \(i < i(k)\) (cf. substep (c) in MSR-Method 4.3.1), we infer

\[ \varrho_{k,i(k)-2} < \frac{\varrho_{1,0}}{1 + \beta(\gamma + 1 + 2(i(k) - 3))\varrho_{1,0}} \]  

(5.1.26)

and

\[ \varrho_{k,i(k)-1} < \frac{\varrho_{1,0}}{1 + \beta(\gamma + 1 + 2(i(k) - 2))\varrho_{1,0}}. \]

But taking into account the relations (5.1.22), (5.1.23), one gets

\[ \varrho_{k+1,0} \equiv \varrho_{k,i(k)} \leq \varrho_{k,i(k)-2} - 2\beta\hat{\varrho}_{k,i(k)-2} - 2\beta\hat{\varrho}_{k,i(k)-1} + 2\zeta_k - \|u^{k,i(k)-1} - u^{k,i(k)-2}\|^2. \]

Because of (5.1.4) we have

\[ 2\zeta_k - \|u^{k,i(k)-1} - u^{k,i(k)-2}\|^2 < 0, \]

hence,

\[ \varrho_{k+1,0} < \varrho_{k,i(k)-2} - 2\beta\hat{\varrho}_{k,i(k)-2}. \]

As beforehand, using (5.1.25), inequality (5.1.26) leads to

\[ \varrho_{k+1,0} < \frac{\varrho_{1,0}}{1 + \beta(\gamma + 1 + 2(i(k) - 2))\varrho_{1,0}} < \frac{\varrho_{1,0}}{1 + \beta(\gamma + 2(i(k) - 2))\varrho_{1,0}}. \]

Analogously, from (5.1.4) and (5.1.22)-(5.1.24) we infer

\[ \varrho_{k+1,1} \leq \varrho_{k,i(k)-1} - 2\beta\hat{\varrho}_{k,i(k)-1} - 2\beta\hat{\varrho}_{k+1,0} + 2\zeta_k - \|u^{k+1,1} - u^{k+1,0}\|^2 \]

\[ \leq \varrho_{k,i(k)-2} - 2\beta\hat{\varrho}_{k,i(k)-2} - 2\beta\hat{\varrho}_{k,i(k)-1} - 2\beta\hat{\varrho}_{k+1,0} + 3\zeta_k - \|u^{k,i(k)-1} - u^{k,i(k)-2}\|^2 - \|u^{k+1,1} - u^{k+1,0}\|^2. \]

Thus,

\[ \varrho_{k+1,1} < \varrho_{k,i(k)-2} - 2\beta\hat{\varrho}_{k,i(k)-2} - \|u^{k+1,1} - u^{k+1,0}\|^2 \]

and

\[ \varrho_{k+1,1} < \frac{\varrho_{1,0}}{1 + \beta(\gamma + 1 + 2(i(k) - 2))\varrho_{1,0}} - \|u^{k+1,1} - u^{k+1,0}\|^2. \]

Finally, in each of the cases (1)-(3) the estimate

\[ \varrho_{k+1,1} < \frac{\varrho_{1,0}}{1 + \beta(\gamma + i(k))\varrho_{1,0}} \]
is true. Hence, if for $2 \leq n \leq k$ the values $ζ_n$ fulfil the condition

$$ζ_n \leq \frac{β}{4} \left( 1 + \frac{ϑ_{1,0}}{1 + β \left( \sum_{j=1}^{n-1} i(j) + 3 \right) ϑ_{1,0}} \right)^2,$$

then we derive

$$ϑ_{k,1} < \frac{ϑ_{1,0}}{1 + β \left( \sum_{j=1}^{k-1} i(j) + 1 \right) ϑ_{1,0}}.$$

□

### 5.1.2 Remark

The proof of Theorem 5.1.1 implies that the values $ϑ_{k,i}$ decrease monotonously for fixed $k$ and $0 \leq i < i(k)$ and, obviously, $ϑ_{k,i} < ϑ_{1,0}$ holds if $(k, i) \neq (1, 0)$.

In order to simplify the procedure of choosing the parameters, we note that estimate (5.1.5) is also true if the term $ζ_k$ is replaced by $τ'_k := c(σ_k + μ_k + ϵ_k)$ everywhere in Theorem 5.1.1, with $τ'_k \geq ζ_k$. The real value of the constant $c$ can easily be calculated. In particular, if $r > 1$, then it is possible to choose

$$c := 8 + (r + L(r)) \left( \frac{L(r)}{r} + 4 \right).$$

♦

### 5.1.2 Asymptotic properties of MSR-methods

Now we establish an asymptotic rate of convergence for the sequence $\{ϑ_{k,1}\}$ studied in the previous subsection. To do so we assume in addition that

$$\lim_{k \to \infty} \max_{1 \leq i \leq i(k)} ρ(υ^{k,i}, U^*) = 0.$$

It should be noted that the inequalities (5.1.14) and (5.1.15) remain true if, in each step $k$, term $\frac{1}{r}$ is replaced by

$$\frac{4}{\left( \max_{n \geq k} \max_{0 \leq i \leq i(n)-1} \|υ^{n,i} - u^{n,i}\| \right)^2},$$

with $υ^{k,i} := \arg \min_{v \in Q_k} \|υ^{k,i} - v\|$ as beforehand. Provided with these replacements, both inequalities will be denoted from here on as modified inequalities (5.1.14)', (5.1.15').

The same replacement will be used in the definition of $ζ'_k$ and we re-denote in this case $ζ_k := \max[ζ'_k, ζ''_k]$ by $ζ(k)$.

### 5.1.3 Theorem

Suppose the assumptions of Lemma 4.3.5 and Assumption (i) of Theorem 5.1.1 are fulfilled, parameters $σ_k, μ_k$ and $ϵ_k$ decrease sufficiently fast and $δ_k$ is chosen such that

$$\left( δ_k - \frac{ϵ_k}{2} \right)^2 \geq ζ(k - 1) + 2ζ(k).$$
Moreover let
\[ \lim_{k \to \infty} \max_{1 \leq i \leq i(k)} \rho(u^{k,i}, U^*) = 0. \]
Then it holds
\[ \lim_{k \to \infty} \rho_k, i \left( \sum_{j=1}^{k-1} \hat{i}(k) \right) = 0. \]

**Proof:** Indeed, for \( 1 \leq i \leq i(k) \) we make use of
\[ \rho_k, i = J(u^{k,i}) - J^* + L(r) \left( \mu_k + \frac{\epsilon_k}{2} \right) \]
instead of
\[ \rho_k, i \leq L(r) \left( 2r + \mu_1 + \frac{\epsilon_1}{2} \right) \]
and inequality
\[ \frac{(J_k(u^{k,i}) - J_k(\bar{v}^{k,i}))^2}{4\|\bar{v}^{k,i} - u^{k,i}\|^2} > \frac{\sigma_k^2 - 2(2\sigma_k + L(r)\mu_k + \hat{\zeta}_k)(2L(r)r + \hat{\xi}_k^2)}{4\|\bar{v}^{k,i} - u^{k,i}\|^2} \]
instead of inequality (5.1.14).

Proceeding analogously as in the proof of Theorem 5.1.1 (cf. (5.1.17)), the following conclusion is easy to verify:
If \( 0 < \beta_k \leq q(k) \), with
\[ q(k) := \frac{1}{4} \min \left[ \max_{n \geq k} \frac{1}{\max_{1 \leq i \leq i(n)}} \left( J(u^{n,i}) - J^* + L(r)(\mu_n + \frac{\epsilon_n}{2}) \right) \right], \]
then
\[ \frac{1}{2} \sigma_k, i - 2\beta_k \sigma_k^2, i \geq 0, \quad 1 \leq i \leq i(k). \]
Hence, for the corresponding steps \( k \) and \( i \) the inequalities
\[ \sigma_{k,i+1} \leq \sigma_{k,i} - 2\beta_k \sigma_{k,i}^2 + \zeta(k) - \|u^{k,i+1} - u^{k,i}\|^2 \quad (i > 0) \]
\[ \sigma_{k,1} \leq \sigma_{k,0} - 2\beta_k \sigma_{k,0}^2 + \zeta(k) - 1 - \|u^{k,1} - u^{k,0}\|^2 \]
are conclusions from each of the inequalities (5.1.11), (5.1.12), (5.1.16) and of the modified inequality (5.1.15).
As beforehand, we assume that \( \sigma_0 = \sigma_1, \mu_0 = \mu_1, \epsilon_0 = \epsilon_1 \) and that \( \beta \) satisfies (5.1.3). Let \( \beta_0 = \beta_1 = \beta \) and
\[ \zeta(k) \leq \frac{\beta_{k-1}^2 \sigma_{k-1}^2}{4 \left( 1 + \left( \beta_0 + \sum_{j=1}^{k-1} \beta_j \hat{i}(j) + 2\beta_{k-1} \right) \sigma_{1,0} \right)^2}, \quad (5.1.27) \]
with $0 < \beta_j \leq q(j)$ for $j > 1$.

Then, according to the proof of estimate (5.1.5), we obtain

$$
\varrho_{k,1} < \frac{\varrho_{1,0}}{1 + \left( \beta_0 + \sum_{j=1}^{k-1} \beta_j \hat{i}(j) \right) \varrho_{1,0}}.
$$

In particular, for $\beta_k \geq \frac{1}{2} q(k)$,

$$
\lim_{k \to \infty} \frac{\sum_{j=1}^{k-1} \beta_j \hat{i}(j)}{1 + \left( \beta_0 + \sum_{j=1}^{k-1} \beta_j \hat{i}(j) \right) \varrho_{1,0}} = 0,
$$

which shows an asymptotic rate of the estimate. In order to verify this it suffices to prove that

$$
\lim_{k \to \infty} \frac{\sum_{j=1}^{k} \beta_j \hat{i}(j)}{\sum_{j=1}^{k} \hat{i}(j)} = \infty
$$
or, because of $i(k) \leq \hat{i}(k) < 2i(k)$,

$$
\lim_{k \to \infty} \frac{\sum_{j=1}^{k} \beta_j i(j)}{\sum_{j=1}^{k} i(j)} = \infty.
$$

Let $\hat{j}(k)$ be the maximal index for which

$$
\sum_{j=1}^{\hat{j}(k)} i(j) \leq \sum_{j=\hat{j}(k)+1}^{k} i(j).
$$

It is obvious that $\hat{j}(k)$ may be equal to zero at the beginning. But there exists a number $k_0$ such that $\hat{j}(k) > 0 \ \forall k > k_0$, moreover, $\hat{j}(k) \to \infty$ as $k \to \infty$.

Moreover, due to the definition of $q(k)$, it follows that

$$
q(k + 1) \geq q(k), \ \forall k, \ \lim_{k \to \infty} q(k) = \infty,
$$
hence,

$$
\lim_{k \to \infty} q(\hat{j}(k)) = \infty.
$$

Now, in view of

$$
\beta_k \geq \frac{1}{2} q(k) \geq \frac{1}{2} q(\hat{j}(k) + 1)
$$

and the definition of $\hat{j}(k)$, we obtain for $k > k_0$

$$
\frac{\sum_{j=1}^{k} \beta_j \hat{i}(j)}{\sum_{j=1}^{k} \hat{i}(j)} \geq \frac{\sum_{j=1}^{\hat{j}(k)} \beta_j \hat{i}(j) + \sum_{j=\hat{j}(k)+1}^{k} \beta_j i(j)}{2 \sum_{j=\hat{j}(k)+1}^{k} \hat{i}(j)} \geq \frac{1}{4} q(\hat{j}(k) + 1)
$$

and the use of the fact that $\lim_{k \to \infty} q(\hat{j}(k)) = \infty$ completes the proof. \hfill \Box

We emphasize that the assumptions made here on the control parameters do not contradict the assumptions of Theorem 5.1.1.

In case of ill-posed problems, however, an suitable estimate of $q(k)$ from
below cannot be expected and, consequently, we do not have an appropriate advice for choosing the parameters $\sigma_k$, $\mu_k$ and $\epsilon_k$ in correspondence with condition (5.1.27).

In view of $i(k) \leq \hat{i}(k) < 2i(k)$ the asymptotic estimate indicates that

$$\varrho_{k,1} = o\left(\frac{1}{s_k}\right), \quad s_k := \sum_{j=1}^{k-1} i(k) + 1,$$

with $s_k$ the number of necessary prox-iterations by transition from iterate $u_{1,0}$ to $u_{k,1}$. Due to (5.1.5), the estimate

$$\varrho_{k,1} \leq \frac{\varrho_{1,0}}{1 + \beta \left(\sum_{j=1}^{k-1} \hat{i}(j) + 1\right)} \varrho_{1,0} < \frac{1}{\beta s_k}$$

is also true.

Now, we are going to establish that, on the class of problems considered, both estimates are sharp in the sense that it is impossible to replace $s_k$ by $s_k + \tau$ with $\tau > 0$ (in the inequality above). It is also impossible even if $\tau > 0$ is chosen arbitrarily small and $\beta$ is replaced by an arbitrary large constant, whereas the controlling parameters of the method are decreased as fast as we want.

In order to show this, we consider MSR-Method 4.3.1 with $\sigma_k := 0$, $\epsilon_k := 0$ and $K_k := K$ for all $k$. In that case, obviously, the method turns into the exact proximal point algorithm

$$u^k = \text{Prox}_{J,K} u^{k-1}, \quad (5.1.28)$$

regardless of how the sequence $\{\delta_k\}$ is chosen. For any $\tau > 0$ we construct a problem for which the inequality

$$J(u^k) - J^* \leq \frac{c}{k^{1+\tau}}$$

cannot be fulfilled for each $k$, even if $c$ is chosen arbitrarily large. Indeed, fixing $\tau > 0$, we take such an even number $m$ that $m > 4^{1+\tau}$ and consider the problem

$$\min_{u \in \mathbb{R}^1} J(u), \quad J(u) := 2u^m.$$  

Starting with $u^0 > 0$, iteration (5.1.28) leads to

$$m(u^k)^{m-1} + u^k - u^{k-1} = 0,$$

i.e.,

$$u^k := \frac{u^{k-1}}{m(u^k)^{m-2} + 1}, \quad \forall k.$$  

Iterating the latter recursions for $k = 1, ..., n$, we get

$$u^n = \frac{u^0}{(m(u^1)^{m-2} + 1) \cdot \ldots \cdot (m(u^n)^{m-2} + 1)},$$

and because $u^n \to 0$, we have

$$\lim_{n \to \infty} \prod_{k=1}^{n} (m(u^k)^{m-2} + 1) = \infty.$$
hence
\[ \lim_{n \to \infty} \sum_{k=1}^{n} (u^k)^{m-1} = \infty. \]

Suppose that inequality
\[ J(u^k) - J^* = J(u^k) \leq \frac{c}{k^{1+\tau}} \]
is true for some \( c > 0 \) and all \( k \). Then, obviously,
\[ (u^k)^m \leq \frac{c}{2k^{1+\tau}}, \]
\[ (u^k)^{m-2} \leq \frac{(\frac{c}{2})^{m-2}}{k^{(1+\tau)(m-2)}}. \]

With regard to the choice of \( m \) we have
\[ (1 + \tau)^{m-2} > 1 + \frac{\tau}{2}, \]
thus
\[ (u^k)^{m-2} \leq \frac{(\frac{c}{2})^{m-2}}{k^{(1+\tau)^{m-2}}}, \quad \forall \ k. \]

But \( \sum_{k=1}^{\infty} k^{-(1+\frac{\tau}{2})} \) converges and this contradicts the fact that
\[ \lim_{n \to \infty} \sum_{k=1}^{n} (u^k)^{m-2} = \infty. \]

### 5.1.3 Linear convergence of MSR-methods

In order to continue the analysis of convergence rates with respect to the values of the objective functional, we make use of the following assumption.

**5.1.4 Assumption.** There exists a constant \( \theta > 0 \) such that, for each \( c \in (0, \mu_1 + \frac{\epsilon_1}{2}) \),
\[ \inf_{u \in \Omega_c} \frac{J(u) - J^* + L(r)c}{\rho^2(u, U^* \cap B_{r*})} \geq \theta, \quad (5.1.29) \]
with
\[ \Omega_c := \{ u \in B_r : \rho(u, Q) \leq c, \ J(u) > J^* \}. \]

This assumption can be understood as a certain growth condition for the objective functional \( f \) around \( U^* \cap B_{r*} \).

We choose a constant \( \zeta_k \) such that
\[ \zeta_k \geq 6\sigma_k + (L(r) + r) \left( 2\mu_k + \frac{\epsilon_k}{2} \right) + \left( 2\mu_k + \frac{\epsilon_k}{2} \right)^2 + \xi_1 \]
\[ + \max \left[ L(r)\mu_k + \xi_k^L, \max \left( L(r), \frac{\rho_{1,0}}{2} \right) \left( \left( 2\sigma_k + 2L(r)\mu_k + L(r)\xi_k^L \right)^{1/3} + \mu_k \right) \right], \quad (5.1.30) \]
with \( \xi_1 \) and \( \xi_k^L \) defined as at the beginning of Section 5.1.1.
5.1.5 Theorem. Suppose the following conditions are fulfilled:

(i) Assumption 5.1.4;

(ii) assumptions of Lemma 4.3.5 and assumption (i) in Theorem 5.1.1.

Moreover, \( \xi_k \) and \( \delta_k \) let be chosen such that

\[
\xi_{k-1} + 2 \xi_k \leq \left( \delta_k - \frac{\xi_k}{2} \right)^2 =: \xi_k^2, \\
\xi_k \leq \left( 1 - \frac{\theta'}{16} \right)^2 \xi_{k-1}, \quad \xi_k < \frac{\theta'}{32} \varrho_{1,0} q^{s_k + 2},
\]

with

\[
\theta' := \min[\theta, 8], \quad s_k := \sum_{j=1}^{k-1} i(j), \quad q \in \left[ 1 - \frac{\theta'}{32}, 1 \right).
\]

Then it holds

\[
\varrho_{k,i} \leq \varrho_{1,0} q^{s_k + i}, \quad k = 1, 2, \ldots, 0 \leq i \leq i(k).
\]

**Proof:** Let us return to case (c) in the proof of Theorem 5.1.1 and assume that

\[
J(u^{k,i}) - J(v^{k,i}) > 2\sigma_k + L(r)\mu_k.
\]

If

\[
\|u^{k,i} - v^{k,i}\| \leq \left( 2\sigma_k + L(r)\mu_k + \xi_k \right)^{1/3} + \mu_k,
\]

then we obtain immediately from (5.1.13), (4.1.3) and (4.2.5)

\[
\varrho_{k,i+1} \leq L(r) \left( 2\sigma_k + L(r)\mu_k + \xi_k \right)^{1/3} + \mu_k.
\]

Now, let

\[
\|u^{k,i} - v^{k,i}\| > \left( 2\sigma_k + L(r)\mu_k + \xi_k \right)^{1/3} + \mu_k.
\]

Then, obviously \( \|u^{k,i} - v^{k,i}\| > \|v^{k,i} - v^{k,i}\| \), consequently,

\[
\|v^{k,i} - u^{k,i}\| \leq \frac{2 \left( \|v^{k,i} - v^{k,i}\|^2 + \|v^{k,i} - u^{k,i}\|^2 \right)}{4 \|v^{k,i} - u^{k,i}\|^2} < 4 \|v^{k,i} - u^{k,i}\|^2
\]

is true, and in view of (5.1.29) and \( \rho(u^{k,i}, Q_k) \leq \frac{\xi_k}{L(r)} \) it follows

\[
\frac{J(u^{k,i}) - J(v^{k,i}) + \xi_k}{4 \|v^{k,i} - u^{k,i}\|^2} > \frac{1}{16} \cdot \frac{J(u^{k,i}) - J(v^{k,i}) + \xi_k}{4 \|v^{k,i} - u^{k,i}\|^2} \geq \frac{\theta'}{16}.
\]

Furthermore,

\[
\|u^{k,i} - v^{k,i}\| \geq \|u^{k,i} - v^{k,i}\| - \|v^{k,i} - v^{k,i}\|,
\]

and because \( \|v^{k,i} - v^{k,i}\| \leq \mu_k \), we obtain in this case

\[
\|u^{k,i} - v^{k,i}\| > \left( 2\sigma_k + L(r)\mu_k + \xi_k \right)^{1/3}.
\]
Inequality (5.1.13) leads to

\[
J_k(u^{k,i+1}) \leq J_k(u^{k,i}) + \xi_0 - \|u^{k,i+1} - u^{k,i}\|^2 \\
- \frac{1}{4} \left( J(u^{k,i}) - J(u^{k,i}) + \xi_k \right) + \frac{1}{4} \left( J(u^{k,i}) - J(u^{k,i}) + \xi_k \right)^2 \\
+ \frac{1}{2} \left( J(u^{k,i}) - J(u^{k,i}) + \xi_k \right) \times \\
\left| J_k(u^{k,i}) - J(u^{k,i}) \right| \left( J(u^{k,i}) - J(u^{k,i}) \right) - \xi_k |.
\]

Using (5.1.35) and (5.1.36), we continue with

\[
J_k(u^{k,i+1}) < J_k(u^{k,i}) + \xi_0 - \|u^{k,i+1} - u^{k,i}\|^2 \\
- \frac{1}{16} \left( J(u^{k,i}) - J(u^{k,i}) + \xi_k \right) \left( 2\sigma_k + L(r)\mu_k + \xi_k \right) \\
< J_k(u^{k,i}) + \xi_0 - \|u^{k,i+1} - u^{k,i}\|^2 \\
- \frac{1}{16} \left( J(u^{k,i}) - J(u^{k,i}) + \xi_k \right) \left( 2\sigma_k + L(r)\mu_k + \xi_k \right)^{1/3}.
\]

Therefore, regarding Remark 5.1.2,

\[
\theta_{k,i+1} < \left( 1 - \frac{\theta'}{16} \right) \theta_{k,i} + 2\sigma_k + \xi_0 - \|u^{k,i+1} - u^{k,i}\|^2 \\
+ \frac{1}{2} \left( J(u^{k,i}) - J(u^{k,i}) + \xi_k \right) \left( 2\sigma_k + L(r)\mu_k + \xi_k \right)^{1/3} \\
< \left( 1 - \frac{\theta'}{16} \right) \theta_{k,i} + 2\sigma_k + \xi_0 - \|u^{k,i+1} - u^{k,i}\|^2 \\
+ \frac{1}{2} \theta_0 \left( 2\sigma_k + L(r)\mu_k + \xi_k \right)^{1/3}.
\]

From the definition of \( \bar{\zeta}_k \) and \( \theta' \leq 8 \) we conclude that for corresponding \( k \) and \( i \) the estimates

\[
\theta_{k,i+1} \leq \theta_{k,i} \left( 1 - \frac{\theta'}{16} \right) + \bar{\zeta}_k - \|u^{k,i+1} - u^{k,i}\|^2, \quad (5.1.38)
\]

\[
\theta_{k,1} \leq \theta_{k,0} \left( 1 - \frac{\theta'}{16} \right) + \bar{\zeta}_{k-1} - \|u^{k,1} - u^{k,0}\|^2, \quad (5.1.39)
\]

result from any of the inequalities (5.1.11), (5.1.12), (5.1.16), (5.1.34) and (5.1.37).

Assume now that \( k \) and \( i \) are fixed, \( i < i(k) \) and that inequality (5.1.33) is fulfilled with some \( q \in \left( 1 - \frac{\theta'}{32}, 1 \right) \) for \( j < k, 0 \leq i' \leq i(j) \) and for \( j = k, i' \leq i \).

1. If \( i+1 < i(k) \), then due to

\[
\theta_{k,i+1} \leq \theta_{k,i} \left( 1 - \frac{\theta'}{16} \right) + \bar{\zeta}_{k-1} - \|u^{k,i+1} - u^{k,i}\|^2,
\]

which follows from (5.1.39), (5.1.39) and \( \bar{\zeta}_{k-1} \geq \bar{\zeta}_k \), and on account of

\[\xi_k^2 < \|u^{k,i+1} - u^{k,i}\|^2, \quad \xi_k^2 \geq \bar{\zeta}_{k-1} + 2\bar{\zeta}_k,\]
it holds
\[ \varrho_{k,i+1} < \varrho_{k,i} \left(1 - \frac{\theta'}{16}\right). \]

Consequently,
\[ \varrho_{k,i+1} < \varrho_{1,0q^{k,i+1}}. \]

(2) Let \( i > 0, i + 1 = i(k) \). In this case, from
\[
\begin{align*}
\varrho_{k,i+1} & \leq \varrho_{k,i} \left(1 - \frac{\theta'}{16}\right) + \bar{\zeta}_k - \|u^{k,i+1} - u^{k,i}\|^2, \\
\varrho_{k,i} & \leq \varrho_{k,i-1} \left(1 - \frac{\theta'}{16}\right) + \bar{\zeta}_{k-1} - \|u^{k,i} - u^{k,i-1}\|^2,
\end{align*}
\]
and \( \|u^{k,i} - u^{k,i-1}\|^2 > \tilde{c}_k^2 \), we conclude that
\[ \varrho_{k,i+1} < \varrho_{k,i-1} \left(1 - \frac{\theta'}{16}\right)^2 + \bar{\zeta}_{k-1} \left(1 - \frac{\theta'}{16}\right) + \bar{\zeta}_k - \left(1 - \frac{\theta'}{16}\right) \tilde{c}_k^2. \]

With regard to \( \tilde{c}_k^2 \geq \tilde{\zeta}_{k-1} + 2\tilde{\zeta}_k \) and \( \theta' \leq 8 \),
\[ \varrho_{k,i+1} < \varrho_{k,i-1} \left(1 - \frac{\theta'}{16}\right)^2, \]
thus, (5.1.33) is true.

(3) If \( i = 0, i(k) = 1 \) and \( i(k-1) = 1 \) or \( i(k-1) = 2 \), then, due to \( s_k \leq s_{k-1} + 2 \), (5.1.39) and the second inequality in (5.1.32), we get
\[
\begin{align*}
\varrho_{k,1} & \leq \varrho_{1,0q^{k}} \left(1 - \frac{\theta'}{16}\right) + \frac{\theta'}{32} \varrho_{1,0q^{k-1}+2} \\
& \leq \varrho_{1,0q^{k}} \left(1 - \frac{\theta'}{16}\right) \\
& \leq \varrho_{1,0q^{k}+1}.
\end{align*}
\]

(4) Finally, let \( i = 0, i(k) = 1 \) and \( i(k-1) > 2 \). Multiplying the two inequalities
\[
\begin{align*}
\varrho_{k-1,i(k-1)-1} & \leq \varrho_{k-1,i(k-1)-2} \left(1 - \frac{\theta'}{16}\right) + \bar{\zeta}_{k-1} - \|u^{k-1,i(k-1)-1} - u^{k-1,i(k-1)-2}\|^2, \\
\varrho_{k-1,i(k-1)} & \leq \varrho_{k-1,i(k-1)-1} \left(1 - \frac{\theta'}{16}\right) + \bar{\zeta}_{k-1} - \|u^{k-1,i(k-1)} - u^{k-1,i(k-1)-1}\|^2,
\end{align*}
\]
by \( \left(1 - \frac{\theta'}{16}\right)^2 \) and \( \left(1 - \frac{\theta'}{16}\right) \), respectively, and summing up the results together with
\[ \varrho_{k,1} \leq \varrho_{k-1,i(k-1)} \left(1 - \frac{\theta'}{16}\right) + \bar{\zeta}_{k-1} - \|u^{k,1} - u^{k-1,i(k-1)}\|^2, \]
CHAPTER 5. RATE OF CONVERGENCE FOR PPR

then, due to
\[ \|u^{k-1,i(k-1)-1} - u^{k-1,i(k-1)-2}\|^2 > \epsilon_{k-1}^2, \]
we obtain
\[ \varrho_{k,1} < \varrho_{k-1,i(k-1)-2} \left( 1 - \frac{\theta'}{16} \right)^3 \]
\[ + \tilde{\epsilon}_{k-1} \left( 1 + \left( 1 - \frac{\theta'}{16} \right) + \left( 1 - \frac{\theta'}{16} \right)^2 \right) - \tilde{\epsilon}_{k-1}^2 \left( 1 - \frac{\theta'}{16} \right)^2. \]

Now, because of (5.1.32) and \( \tilde{\epsilon}_{k-1}^2 \geq \bar{\zeta}_{k-2} + 2\bar{\zeta}_{k-1} \), the estimate
\[ \varrho_{k,1} < \varrho_{k-1,i(k-1)-2} \left( 1 - \frac{\theta'}{16} \right)^3 \]
holds and consequently, estimate (5.1.33) is also true. \( \square \)

Note that, deducing the estimates (5.1.5) and (5.1.33), we do not require the validity of Assumption 4.2.3, i.e., these estimates are true even if the strong convergence of \( \{u^{k,i}\} \) is not guaranteed.

Now we estimate \( \|u^{k,i} - u^*\| \), with \( u^* \) the strong limit of \( \{u^{k,i}\} \). To do so, we need the following supposition.

5.1.6 Assumption. Assume 5.1.4 is fulfilled with
\[ \Omega'_{c} := \{ u \in B_r : \rho(u, Q) \leq c \} \]

instead of \( \Omega_{c}. \)

Let \( \hat{i}(k) := \frac{16\sigma^2}{\delta_{k}^2} + 1 \) and
\[ \gamma_k := \sum_{j=k}^{\infty} \left( \sqrt{\tau_j} + \frac{\sigma_j}{2} + 2\mu_j \right), \]
with \( \tau_j := 2L(r)\mu_j + 2\sigma_j \). Furthermore, let us keep the former definitions of \( v^{k,i}, s_k, \theta' \) and \( q \).

5.1.7 Theorem. Suppose the Assumptions 4.2.3 and 5.1.6 and hypotheses of Theorem 5.1.5, except for (i), are satisfied and
\[ \frac{1}{4r} \left( \tau_k - \left( \frac{\epsilon_k}{2} \right)^2 \right) + \frac{\epsilon_k}{2} + \sqrt{\tau_k} + 2\mu_k < -\frac{\delta_k^2}{8r}, \quad \forall \ k \in \mathbb{N}; \]
\[ \gamma_k \leq \sqrt{\frac{\vartheta_{1,0}}{\vartheta'}} \sqrt{q^{(s_k + \hat{i}(k)-1)}}, \quad \forall \ k \in \mathbb{N}. \]

Then it holds
\[ \|u^{k,i} - u^*\| \leq 3\sqrt{\frac{\vartheta_{1,0}}{\vartheta'}} \sqrt{q^{(s_k + \hat{i})}}, \quad \forall \ k \in \mathbb{N}, \ 0 \leq i \leq i(k) - 1, \]
with \( u^* = \lim_{k \to \infty} u^{k,1} \).
5.1. RATE OF CONVERGENCE FOR MSR-METHODS

**Proof:** Convergence of \( \{u^{k,i}\} \) to \( u^* \in U^* \) holds in view of Theorem 4.3.8. For \( 0 < i \leq i(k) \) and \( c := a_1 + \frac{1}{r} \), due to \( u^{k,i} \in \Omega^c \) and Assumption 5.1.6, we have

\[
\|u^{k,i} - v^{k,i}\| \leq \sqrt{\frac{\varrho_k}{r}} \tag{5.1.44}
\]

and, because of \( u^{k,0} = u^{k-1,i(k-1)} \), inequality

\[
\|u^{k,0} - v^{k,0}\| \leq \sqrt{\frac{\varrho_k}{r}} \tag{5.1.45}
\]

is also true.

Let \((k_0, i_0)\) be fixed with \( 0 \leq i_0 < i(k_0) \) and \( k' > k_0 + 1 \). Now we estimate \( \|u^{k',0} - v^{k_0,i_0}\| \).

The inequalities (4.3.35) and (4.3.36), established in the proof of Theorem 5.1.5, are obviously true under the hypotheses of Lemma 4.3.5 and, regarding the definition of \( v^k \) in (4.3.6), we get

\[
\|u^{k+1,0} - v^{k,i_0}\| \leq \|u^{k,0} - v^{k,i_0}\| + \sqrt{r_k} + \frac{\epsilon_k}{2} + 2\mu_k,
\]

with \( u^{k,i_0} \) an arbitrary element of \( U^* \cap B_{r^k} \).

For \( k := j, u^{**} := v^{k_0,i_0} \) the latter inequality reads as

\[
\|u^{j+1,0} - v^{k_0,i_0}\| \leq \|u^{j,0} - v^{k_0,i_0}\| + \sqrt{r_j} + \frac{\epsilon_j}{2} + 2\mu_j,
\]

and an iteration of this inequality for \( j = k_0 + 1, \ldots, k' - 1 \) implies

\[
\|u^{k',0} - v^{k_0,i_0}\| \leq \|u^{k_0+1,0} - v^{k_0,i_0}\| + \sum_{j=k_0+1}^{k'-1} \left( \sqrt{r_j} + \frac{\epsilon_j}{2} + 2\mu_j \right). \tag{5.1.46}
\]

Due to (4.3.36) and the inequalities (4.3.11), which are also true in that case, we obtain for \( 0 \leq i_0 < i(k_0) - 1 \)

\[
\|v^{k_0+1,0} - v^{k_0}\| \leq \|v^{k_0,i_0} - v^{k_0}\| + \frac{1}{4r} \left( -r_k^2 + \tau_k + \frac{\epsilon_k}{2} \right),
\]

with \( v^{k_0} \) chosen according to (4.3.6), and

\[
\|u^{k_0+1,0} - v^{k_0}\| \leq \|u^{k_0,i(k_0)-1} - v^{k_0}\| + \sqrt{r_k} + \frac{\epsilon_k}{2}.
\]

Setting \( u^{**} = v^{k_0,i_0} \) in (4.3.34), these inequalities lead to

\[
\|u^{k_0+1,0} - v^{k_0,i_0}\| < \|u^{k_0,i_0} - v^{k_0,i_0}\| + \frac{1}{4r} \left( -r_k^2 + \tau_k + \frac{\epsilon_k}{2} \right) + \sqrt{r_k} + \frac{\epsilon_k}{2} + 2\mu_{k_0}
\]

if \( 0 \leq i_0 < i(k_0) - 1 \), and

\[
\|u^{k_0+1,0} - v^{k_0,i(k_0)-1}\| < \|u^{k_0,i(k_0)-1} - v^{k_0,i(k_0)-1}\| + \sqrt{r_k} + \frac{\epsilon_k}{2} + 2\mu_{k_0}.
\]

Now, (5.1.41), (5.1.46) and the latter inequalities give

\[
\|u^{k',0} - v^{k_0,i_0}\| \leq \|u^{k_0,i_0} - v^{k_0,i_0}\| + \sum_{j=k_0}^{k'-1} \left( \sqrt{r_j} + \frac{\epsilon_j}{2} + 2\mu_j \right), \quad 0 \leq i_0 < i(k_0). \tag{5.1.47}
\]
Passing to the limit $k' \to \infty$ in (5.1.47), we conclude that
\[
\|u^* - v^{k_0,i_0}\| \leq \|u^{k_0,i_0} - v^{k_0,i_0}\| + \sum_{j=k_0}^{\infty} \left( \sqrt{\tau_j} + \frac{\epsilon_j}{2} + 2\mu_j \right).
\]
Hence,
\[
\|u^{k_0,i_0} - u^*\| \leq \|u^{k_0,i_0} - v^{k_0,i_0}\| + \|v^{k_0,i_0} - u^*\| 
\leq 2\|u^{k_0,i_0} - v^{k_0,i_0}\| + \gamma_{k_0},
\]
and, due to (5.1.44) and (5.1.45),
\[
\|u^{k_0,i_0} - u^*\| \leq 2\sqrt{\frac{\theta_{k_0,i_0}}{\theta^r}} + \gamma_{k_0}.
\]
(5.1.48)
Taking into account the last inequality in the proof of Lemma 4.3.3, then it follows immediately
\[
i(k_0) < 2r \left( \frac{(\delta_{k_0} - \epsilon_{k_0})^2 - \tau_{k_0} - \frac{\epsilon_{k_0}}{2}}{4r} \right)^{-1} + 1
\]
and because of (5.1.41) we get $i(k_0) < i(k_0)$ and
\[
\gamma_{k_0} \leq \sqrt{\frac{\theta_{k_0,i_0}}{\theta^r}} \sqrt{q^{(s_{k_0,i_0})}}.
\]
(5.1.49)
Now, (5.1.48) together with (5.1.33) and (5.1.49) implies
\[
\|u^{k_0,i_0} - u^*\| \leq 3\sqrt{\frac{\theta_{k_0,i_0}}{\theta^r}} \sqrt{q^{(s_{k_0,i_0})}}.
\]
□

5.1.8 Remark. The conditions controlling the MSR-methods described in Theorem 4.3.6 and Theorem 4.3.8 enable us to choose a priorily the parameters. However, in the Theorems 5.1.1 and 5.1.5 on the rate of convergence, the sequences of controlling parameters can be generated within the methods. Regarding Remark 5.1.2, this procedure is not complicated. Indeed, at the beginning of the $k$-th exterior step, the values $\sigma_k$, $\epsilon_k$, $\mu_k$ and $\delta_k$ are known and they can be kept constant for all interior steps $i$. After termination of the interior loop the value of $i(k)$ is known, which allows us to calculate the necessary bounds for $\zeta_{k+1}$ of $\hat{\zeta}_{k+1}$ (cf. the last inequalities in hypothesis (iii) of Theorem 5.1.1 or (5.1.32), all other conditions can be satisfies a priorily). Furthermore, in order to determine $\sigma_{k+1}$, $\epsilon_{k+1}$ and $\mu_{k+1}$, we have only to observe upper bounds, including (4.3.14), and afterwards one can calculate $\delta_{k+1}$ in view of (4.3.13), (5.1.4) or (4.3.13), (5.1.31).

Questions on how to control the method efficiently will be considered in Section ??, too.

The updating rule for the controlling parameters ensuring linear convergence of $\{u^{k,i}\}$ to $u^*$ (cf. Theorem 5.1.7), is rather complicated. In fact, together with the choice of $\sigma_{k_0}$, $\epsilon_{k_0}$ and $\mu_{k_0}$ we have to satisfy the inequalities (5.1.42) for
all \( k \leq k_0 - 1 \). From the practical point of view we cannot exclude that the discretization parameter has to be decreased inadmissible fast. However, it should be taken into account that for ill-posed problems a qualified convergence of the methods is usually an important advantage, even if corresponding assumptions on the controlling parameters and the accuracy of data approximation are idealized. Condition (5.1.42), which seems to be strange at the first glance, can be explained as follows: In general, the original problem has a non-unique solution and to which element \( u^* \) the method is converging depends on the chosen sequences \( \{\mu_k\}, \{\sigma_k\}, \{\epsilon_k\} \) and \( \{\delta_k\} \).

Finally, we note that the rate of convergence of MSR-Method 4.3.12 can be investigated in the same way.

### 5.2 Rate of Convergence for OSR-Methods

Now we are going to investigate the rate of convergence for OSR-Method 4.2.1. This will be done by applying techniques used in Section 5.1 for MSR-methods.

As in Subsection 5.1.1, we take \( \sigma_0 = \sigma_1, \epsilon_0 = \epsilon_1, \mu_0 = \mu_1 \) and use the former definitions of \( \xi_1^k, \xi_2^k, \zeta_1', \zeta_2', \zeta_2'' \) and \( \zeta_2 \) given there.

Define

\[
\varrho_k := J(u^k) - J^* + \xi_2^k, \quad k = 1, 2, \ldots
\]

#### 5.2.1 Theorem

Suppose the hypotheses of Theorem 4.2.4 are fulfilled, \( u^0 \in K \) is chosen and, moreover,

(i) \( \sigma_k+1 \leq \sigma_k, \epsilon_k+1 \leq \epsilon_k, \mu_k+1 \leq \mu_k, \rho(Q_k, Q_{k+1}) \leq 2\mu_k, \forall k, \epsilon_1 < 1, \mu_1 < 1 \) and \( J(u^0) > J^* \);

(ii) constant \( \beta \) is fixed such that

\[
0 < \beta < \frac{1}{4} \min \left[ \frac{1}{L(r)(2r + \mu_1 + \frac{\epsilon_1}{2})}, \frac{1}{8r^2} \right];
\]

(iii) controlling parameters are chosen according to

\[
\zeta_k \leq \beta \left( \frac{\varrho_0}{1 + (k+1)\beta \varrho_0} \right)^2, \quad k = 1, 2, \ldots
\]

Then, for OSR-Method 4.2.1 it holds

\[
\varrho_k < \frac{\varrho_0}{1 + \beta \varrho_0}. \tag{5.2.2}
\]

**Proof:** Obviously, \( \varrho_k \geq 0 \) holds for all \( k \).

Following completely the proof of relation (5.1.19) in Theorem 5.1.1, we establish

\[
\varrho_k \leq \varrho_{k-1} - 2\beta \varrho_{k-1}^2 + \zeta_{k-1}, \quad k = 1, 2, \ldots
\]

With regard to \( \zeta_0 = \zeta_1 \), (5.2.1) and (5.2.3), this leads to

\[
\varrho_1 < \frac{\varrho_0}{1 + \beta \varrho_0}.
\]
Due to assumptions (i) and (ii), \( \varrho_k \leq \frac{1}{4} \) is true for all \( k \). Because the function \( t - 2\beta t^2 \) increases on the interval \([0, \frac{1}{4\beta}]\), the inequalities (5.2.3) and (5.2.4) imply immediately

\[
\varrho_2 \leq \frac{\varrho_0}{1 + \beta \varrho_0} - 2\beta \left( \frac{\varrho_0}{1 + \beta \varrho_0} \right)^2 + \frac{\beta}{4} \left( \frac{\varrho_0}{1 + 2\beta \varrho_0} \right)^2
\]

and

\[
\varrho_2 < \frac{\varrho_0}{1 + 2\beta \varrho_0}.
\]

In the sequel we follow part (1) of the proof of Theorem 5.1.1. Suppose that, for fixed \( k \), inequality

\[
\varrho_k < \frac{\varrho_0}{1 + \beta k \varrho_0}
\]

holds true. If \( \zeta_k \geq \beta \varrho_k^2 \), then according to (5.2.1),

\[
\varrho_k \leq \frac{1}{2} \cdot \frac{\varrho_0}{1 + \beta(k + 1) \varrho_0}
\]

and using (5.2.3), we get

\[
\varrho_{k+1} \leq \varrho_k + \zeta_k \\
\leq \frac{1}{2} \left( \frac{\varrho_0}{1 + \beta(k + 1) \varrho_0} \right) + \frac{\beta}{4} \left( \frac{\varrho_0}{1 + \beta(k + 1) \varrho_0} \right)^2 \\
< \frac{\varrho_0}{1 + \beta(k + 1) \varrho_0}.
\]

But if \( \zeta_k < \beta \varrho_k^2 \), then (5.2.3) implies

\[
\varrho_{k+1} \leq \varrho_k - \beta \varrho_k^2,
\]

and again, in view of the monotonicity of the function \( t - 2\beta t^2 \) on \([0, \frac{1}{4\beta}]\), one can conclude that

\[
\varrho_{k+1} \leq \frac{\varrho_0}{1 + \beta k \varrho_0} - \beta \left( \frac{\varrho_0}{1 + \beta k \varrho_0} \right)^2 \\
< \frac{\varrho_0}{1 + \beta(k + 1) \varrho_0} \cdot \frac{1 - \beta \varrho_0}{1 + \beta(k + 1) \varrho_0} = \frac{\varrho_0}{1 + \beta(k + 1) \varrho_0}.
\]

\( \square \)

Now, for OSR-methods linear convergence of the objective function values as well as of the iterates can be proved analogously to the statements in the Theorems 5.1.5 and 5.1.7.

5.2.2 Theorem. Let the following hypotheses be fulfilled:

(i) Assumption 5.1.4;

(ii) all assumptions of Theorem 4.2.4 and assumption (i) of Theorem 5.2.1;
choose a constant $\bar{\zeta}_k$ such that

$$
\bar{\zeta}_k < \frac{\theta'}{32} q^k
$$

with $\theta' := \min[\theta, 8]$ and $q \in \left(1 - \frac{\theta'}{32}, 1\right)$.

Then, for OSR-Method 4.2.1 it holds

$$
\varrho_k \leq \varrho_0 q^k, \quad k = 1, 2, \ldots
$$

(5.2.5)

Proof: An OSR-method corresponds in the context of an MSR-method to part (3) of the proof of Theorem 5.1.5, i.e. in all exterior steps $k$ only one interior step is created. Thus $i(k) = 1$ for all $k$.

Therefore, following the proofs of the Theorems 5.1.1 and 5.1.5, we obtain

$$
\varrho_k \leq \left(1 - \frac{\theta'}{16}\right) \varrho_{k-1} + \bar{\zeta}_{k-1}, \quad \forall \ k > 0,
$$

(5.2.6)

which corresponds to (5.1.39). Now assume that

$$
\varrho_k \leq \varrho_0 q^k
$$

(5.2.7)

holds for a fixed $k$. For $k = 1$ this can be verified trivially. Then, due to the assumption (iii) and (5.2.6), (5.2.7), we conclude

$$
\varrho_{k+1} < \left(1 - \frac{\theta'}{16}\right) \varrho_0 q^k + \frac{\theta'}{32} \varrho_0 q^k
$$

$$
= \left(1 - \frac{\theta'}{32}\right) \varrho_0 q^k \leq \varrho_0 q^{k+1}.
$$

□

5.2.3 Theorem. Suppose that Assumptions 4.2.3 and 5.1.6 and the hypotheses of Theorem 5.2.2, except for (i), are satisfied and that

$$
\sum_{j=k+1}^{\infty} \left(\sqrt{\tau_j} + \frac{\epsilon_j}{2} + 2 \mu_j\right) \leq \sqrt{\frac{\varrho_0}{\theta'}} \sqrt{q^k}, \quad \forall \ k.
$$

(5.2.8)

Then it holds

$$
\|u^k - u^*\| \leq 3 \sqrt{\frac{\varrho_0}{\theta'}} \sqrt{q^k}, \quad (5.2.9)
$$

with $u^* = \lim_{k \to \infty} u^k$.

Proof: Convergence $u^k \to u^* \in U^*$ follows immediately from Theorem 4.2.4. Using Assumption 5.1.6 with $c := \mu_1 + 2 \mu$, and having in mind that $u^k \in \Omega_c$, we get

$$
\|u^k - w^k\| \leq \sqrt{\frac{\varrho_k}{\theta'}}, \quad (5.2.10)
$$

with $w^k := \arg\min\{\|u^k - w\| : w \in U^* \cap B_r\}$.

From (4.2.13), considered with fixed $k_0$ and $u^{**} := w^{k_0}$, it follows

$$
\|u^k - w^{k_0}\| \leq \|u^{k-1} - w^{k_0}\| + \sqrt{\tau_k} + \frac{\epsilon_k}{2} + 2 \mu_k, \quad \forall \ k,
$$
and iterating these inequalities for \( k = k_0 + 1, \ldots, k_0 + j \), one gets
\[
\|u^{k_0+j} - w^{k_0}\| \leq \|u^{k_0} - w^{k_0}\| + \sum_{k=k_0+1}^{k_0+j} \left( \sqrt{\tau_k} + \epsilon_k + 2\mu_k \right).
\]
Passing \( j \to \infty \), we have
\[
\|u^* - w^{k_0}\| \leq \|u^{k_0} - w^{k_0}\| + \sum_{k=k_0+1}^{\infty} \left( \sqrt{\tau_k} + \epsilon_k + 2\mu_k \right).
\]
Thus,
\[
\|u^* - u^{k_0}\| \leq 2\|u^{k_0} - w^{k_0}\| + \sum_{k=k_0+1}^{\infty} \left( \sqrt{\tau_k} + \epsilon_k + 2\mu_k \right),
\]
and in view of (5.2.10),
\[
\|u^* - u^{k_0}\| \leq 2 \sqrt{\frac{\theta_{k_0}}{\varrho'}} + \sum_{k=k_0+1}^{\infty} \left( \sqrt{\tau_k} + \epsilon_k + 2\mu_k \right).
\]
The latter inequality together with (5.2.7) and (5.2.8) allows us to conclude that
\[
\|u^* - u^{k_0}\| \leq 3 \sqrt{\frac{\theta_{k_0}}{\varrho'}} \sqrt{q_k}.
\]

It should be noted that the estimates (5.2.2) and (5.2.5) follow immediately as particular cases from the Theorems 5.1.1 and 5.1.5 which correspond to MSR-methods. However, deducing this way, it requires stronger conditions on the controlling parameters.

5.3 Rate of Convergence for Regularized Penalty Methods

In this section the approaches, developed in the previous Sections 5.1 and 5.2, are applied in order to investigate the rate of convergence for penalty methods studied in Section 3.2.

Considering Problem (A3.4.56) under the additional assumption that the optimal set \( \mathcal{U}^* \) is bounded, we analyze Method (3.2.2) with \( \chi_k = 2 \) \( \forall \ k \) and penalty function (A3.4.78):
\[
\phi_k(u) := \begin{cases} 
-\frac{1}{r_k} \sum_{j=1}^{m} \frac{1}{g_j(u)} & \text{if } u \in \text{int} K, \\
+\infty & \text{if } u \notin \text{int} K,
\end{cases}
\]
with \( u^0 \in \text{int} K, \ r_1 > 0, \ r_{k+1} \geq r_k, \ \forall \ k \) and \( \lim_{k \to \infty} r_k = \infty. \)

An immediate application of the abstract scheme of the investigation of OSR- and MSR-methods is impossible here, because in the normal case that \( \inf_{u \in \mathbb{R}^n} J(u) < \inf_{u \in K} J(u) \), condition (4.2.5) cannot be fulfilled for
\[
J_k(u) = J(u) + \phi_k(u)
\]
5.3. RATE OF CONVERGENCE FOR REGULARIZED PENALTY METHODS

in general, because on a subset of any neighborhood of the optimal set it holds
\[ \lim_{k \to \infty} \phi_k(u) = \infty. \]

The same situation happens for other penalty functions satisfying the conditions of the Theorems 3.2.1 and 3.2.3.

Due to Theorem 3.2.3, sequence \( \{u^k\} \) generated by Method (3.2.2) with penalty function (5.3.1) is bounded if \( \sum_{k=1}^{\infty} \epsilon_k < \infty \). Moreover,
\[ \lim_{k \to \infty} \|u^k - u^{k-1}\| = 0, \quad \lim_{k \to \infty} J(u^k) = J^*. \]
Relation (3.2.2) implies
\[ \|\nabla J_k(u^k)\| \leq \epsilon_k + 2\|u^k - u^{k-1}\| =: \bar{\epsilon}_k \] (5.3.2)
and, in view of \( \lim_{k \to \infty} \|u^k - u^{k-1}\| = 0 \), we get \( \lim_{k \to \infty} \bar{\epsilon}_k = 0 \). Choosing an arbitrary Slater point \( \tilde{u} \), we obtain from (5.3.2)
\[ \langle \nabla J(u^k) + \frac{1}{r_k} \sum_{j=1}^{m} \frac{1}{g_j^2(u^k)} \nabla g_j(u^k), u^k - \tilde{u} \rangle \leq \bar{\epsilon}_k \|u^k - \tilde{u}\| \]
and, with regard to the convexity of \( J \) and \( g_j \),
\[ \frac{1}{r_k} \sum_{j=1}^{m} \frac{1}{g_j^2(u^k)} \left[ g_j(u^k) - g_j(\tilde{u}) \right] \leq \bar{\epsilon}_k \|u^k - \tilde{u}\| + J(\tilde{u}) - J(u^k) \]
\[ \leq \bar{\epsilon}_k \|u^k - \tilde{u}\| + J(\tilde{u}) - J^*. \] (5.3.3)
Choosing a convergent subsequence \( \{u^{k_i}\} \), then \( \tilde{u} = \lim_{i \to \infty} u^{k_i} \in U^* \).

Denote
\[ I_0(\tilde{u}) := \{ j : g_j(\tilde{u}) = 0 \}, \quad I_-(\tilde{u}) := \{ j : g_j(\tilde{u}) < 0 \}, \]
\[ \gamma := \max \left[ \max_{j \in I_-(\tilde{u})} g_j(\tilde{u}), \max_{j \in I_0(\tilde{u})} g_j(\tilde{u}) \right]. \]
Due to the continuity of the functions \( g_j \) and the definition of the index sets \( I_0(\tilde{u}) \) and \( I_-(\tilde{u}) \), we can conclude that, beginning with some \( i_0 \) and \( i \geq i_0 \),
\[ g_j(u^{k_i}) < \frac{\gamma}{2} < 0 \quad \text{if} \quad j \in I_-(\tilde{u}), \]
\[ g_j(u^{k_i}) - g_j(\tilde{u}) > -\frac{\gamma}{2} \quad \text{if} \quad j \in I_0(\tilde{u}). \]
Consequently, if \( j \in I_-(\tilde{u}) \),
\[ \frac{1}{r_{k_i} g_j^2(u^{k_i})} \leq \frac{4}{\gamma^2 r_{k_i}}, \quad i \geq i_0, \] (5.3.4)
and, according to (5.3.3), this implies
\[ -\frac{\gamma}{2r_{k_i}} \sum_{j \in I_0(\tilde{u})} \frac{1}{g_j^2(u^{k_i})} \leq \epsilon_k \|u^{k_i} - \tilde{u}\| + J(\tilde{u}) - J^* - \frac{1}{r_{k_i}} \sum_{j \in I_0(\tilde{u})} \frac{1}{g_j(u^{k_i})}. \] (5.3.5)
Estimates (5.3.4) and (5.3.5) ensures that
\[ \frac{1}{r_{k_i} g_j^2(u^{k_i})} \leq c, \quad \forall i \in \mathbb{N} \]
CHAPTER 5. RATE OF CONVERGENCE FOR PPR

holds with some \( c \) for all functions \( g_j \).

Proving by contradiction, it is not difficult to verify that

\[
\frac{1}{r_k g_j^2(u^k)} \leq c', \quad k = 1, 2, \ldots
\]

is satisfied for some \( c' \) and any \( j = 1, \ldots, m \). Thus,

\[
\phi_k(u^k) = -\frac{1}{r_k} \sum_{j=1}^{m} g_j(u^k) \leq m \sqrt{\frac{c'}{r_k}} =: c_1 \sqrt{r_k}, \quad (5.3.6)
\]

With \( \bar{v}^k := \arg \min \{ \| u^k - v \| : v \in U^* \} \), in view of (A1.5.32), we have for any \( \lambda \in [0, 1] \):

\[
J(u^{k+1}) + \phi_{k+1}(u^{k+1}) + \| u^{k+1} - u^k \|^2 \\
\leq \lambda J(\bar{v}^k) + (1 - \lambda) J(u^k) + \phi_{k+1}(\lambda \bar{v}^k + (1 - \lambda) u^k) + \lambda^2 \| u^k - \bar{v}^k \|^2 + \frac{\epsilon_{k+1}^2}{2}.
\]

But, due to

\[
g_j(u^k + \lambda(\bar{v}^k - u^k)) \leq \lambda g_j(\bar{v}^k) + (1 - \lambda) g_j(u^k) \\
\leq (1 - \lambda) g_j(u^k) < 0
\]

and \( r_{k+1} \geq r_k \), we obtain

\[
\phi_{k+1}(\lambda \bar{v}^k + (1 - \lambda) u^k) \leq \frac{1}{1 - \lambda} \phi_{k+1}(u^k) \\
\leq \frac{1}{1 - \lambda} \phi_k(u^k) \leq \frac{c_1}{(1 - \lambda) \sqrt{r_k}}
\]

and hence, the relations

\[
J(u^{k+1}) \leq \lambda J(\bar{v}^k) + (1 - \lambda) J(u^k) \\
+ \lambda^2 \| u^k - \bar{v}^k \|^2 + \frac{\epsilon_{k+1}^2}{2} + \frac{c_1}{(1 - \lambda) \sqrt{r_k}}
\]

and

\[
J(u^{k+1}) - J^* \leq (1 - \lambda) (J(\bar{v}^k) - J^*) \\
+ \lambda^2 \| u^k - \bar{v}^k \|^2 + \frac{\epsilon_{k+1}^2}{2} + \frac{c_1}{(1 - \lambda) \sqrt{r_k}} \quad (5.3.7)
\]

are true for any \( \lambda \in [0, 1] \). Now we choose

\[
\lambda = \lambda_k := \min \left[ \frac{J(u^k) - J^*}{2\| u^k - \bar{v}^k \|^2}, \frac{7}{8} \right].
\]

If \( \frac{J(u^k) - J^*}{2\| u^k - \bar{v}^k \|^2} \geq \frac{7}{8} \), then (5.3.7) leads to

\[
\bar{v}_{k+1} \leq \frac{9}{16} \bar{v}_k + \frac{8c_1}{\sqrt{r_k}} + \frac{\epsilon_{k+1}^2}{2}, \quad (5.3.8)
\]
with \( \tilde{\varrho}_k := J(u^k) - J^* \). But, if \( \frac{J(u^k) - J^*}{2\|u^k - \bar{v}_k\|^2} < \frac{7}{8} \), we get

\[ \lambda_k := \frac{J(u^k) - J^*}{2\|u^k - \bar{v}_k\|^2}, \]

and (5.3.7) implies

\[ \tilde{\varrho}_{k+1} \leq \tilde{\varrho}_k - \frac{1}{4\|u^k - \bar{v}_k\|^2} \tilde{\varrho}_k^2 + \frac{8c_1}{\sqrt{\tau_k}} \epsilon_k + \epsilon_{k+1}^2, \quad (5.3.9) \]

Comparing the relations (5.3.8), (5.3.9) with (5.1.12), (5.1.15) and acting as in the proof of Theorem 5.1.1 (cf. the transition from the inequalities (5.1.12), (5.1.15) to (5.1.18), (5.1.19), we can establish that for a sufficiently fast decrease of the controlling parameters the estimate

\[ \tilde{\varrho}_{k+1} \leq \tilde{\varrho}_k - 2\tilde{\beta} \tilde{\varrho}_k^2 + \tilde{\epsilon} \left( \frac{1}{\sqrt{\tau_k}} + \epsilon_{k+1}^2 \right), \quad \tilde{\beta} > 0 \]

is a conclusion of (5.3.8) and (5.3.9).

Furthermore, using the techniques for proving the rate of convergence for OSR-methods, analogue statements as in Theorems 5.2.1 - 5.2.3 can be obtained. However, these statements have a non-constructive character if we do not know upper bounds for the values \( c_1 \) in (5.3.6) and \( \sup_k \|u^k - \bar{v}_k\| \).

5.4 Comments

Section 5.1: In ROCKAFELLAR [349] an OSR-method for solving variational inequalities with maximal monotone operators \( T \) has been described. Furthermore, under the assumption that \( T^{-1} \) is Lipschitz continuous at zero-point, a linear convergence rate has been proved (cf. Theorem 2). This assumption implies uniqueness of the solution \( u^* \). Moreover, solving Problem (4.2.1) in this context (with \( T := \partial J \)), a growth condition

\[ \liminf_{u \to u^*} \frac{J(u) - J(u^*)}{\|u - u^*\|^2} > 0 \]

is required, which is stronger than (5.1.29). In LUQUE [281] this investigation has been continued and the Lipschitz continuity of \( T^{-1} \) has been dropped. In both papers referred an approximation of the problem data has been not taken into account, i.e., \( J_k := J \), \( K_k := K \), \( \forall k \).

For the case \( K_k := K = V \), KAPLAN [206] has investigated the convergence rate for an OSR-method and assumed, for the approximation of \( J \), a condition like (4.2.5).

GÜLER [153] has studied an asymptotic rate of convergence for the exact iteration

\[ u^{k+1} := \arg \min_{u \in K} \left\{ J(u) + \frac{1}{2\lambda_k} \|u - u^k\|^2 \right\}, \quad \lambda_k > 0, \sum_{k=1}^{\infty} \lambda_k = \infty. \]

For \( \lambda_k = \text{const} \) this represents a particular case of an OSR-method.

Convergence rate estimates for MSR-methods in the framework of convex
semi-infinite problems have been established in [214] and our presentation follows this paper.

Section 5.2: The results in this section are not yet published.

Section 5.3: These results can be found in [206].
Chapter 6

PPR FOR CONVEX SEMI-INFINITE PROBLEMS

6.1 Introduction to Numerical Analysis of SIP

In the space $V = \mathbb{R}^n$ we consider the following convex problem

$$\min \{ J(u) : u \in K \},$$

with

$$K = \{ u \in U_0 : g(u, t) \leq 0, \quad \forall t \in T \}. \quad (6.1.2)$$

Here set $T$ is a compact subset of the space $\mathbb{R}^m$; $U_0$ is a closed convex subset of $\mathbb{R}^n$; $J$ and $g_t : u \mapsto g(u, t)$ are convex functions from $\mathbb{R}^n$ into $\mathbb{R}$ (see also Problem (A1.7.42)).

Theoretical and practical manifestations of this model are abundant and an increasing number of papers show a growing interest in semi-infinite programming problems during the last two decades. The notion semi-infinite programming (SIP) stems from the property that set $K$ is finite-dimensional whereas $T$ is an infinite set, hence we are dealing with a continuum of inequality constraints.

Concerning theory and properties of SIP we refer to the monographs Blum and Oettli [46], Krabs [248], Hettich and Zenke [178], Glashoff and Gustafson [132], Goerretta and Lopez [137] and to survey papers Gustafson and Kortanek [155], Anderson and Philpott [10], Hettich and Kortanek [177] and Reemtsen and Rückmann [344].

Not being able to deal with infinitely many constraints in a finite algorithm, certain approximations of the problem are always required. Algorithms to be considered generate finitely constrained auxiliary problems and, following [177], one can roughly distinguish three conceptual approaches: Exchange methods, including cutting plane methods, discretization methods and methods based on local reductions.

In order to sketch briefly the respective ideas behind these approaches, we
consider the following finitely constrained auxiliary problems

\[ P(T_k) : \min \{ J(u) : u \in K_k \}, \quad (6.1.3) \]

with

\[ K_k := \{ u \in U_0 : g(u, t) \leq 0, \quad \forall \ t \in T_k \}. \quad (6.1.4) \]

If \( U_0 \) is compact and \( K \neq \emptyset \), the approximate problem (6.1.3) has a solution for every finite set \( T_k \subset T \). Sometimes compactness of \( U_0 \) is artificially introduced in order to secure solvability of these problems.

Omitting for a moment precise assumptions on the problems and methods involved, we emphasize in the sequel the main ideas of the numerical approaches mentioned above.

### 6.1.1 Conceptual exchange approach

The notion exchange methods refers to the fact that in step \( k \) the set \( K_{k+1} \) is generated from \( K_k \) by adding at least one new constraint and (in some algorithms) deleting some of the constraints of \( K_k \), i.e., an exchange of constraints takes place.

#### 6.1.1 Algorithm. (Exchange algorithm)

**Step k**: Given \( T_k \subset T, |T_k| < \infty \);

(i) compute approximately a solution \( u^k \) of \( P(T_k) \) and some or all local solutions \( t^k_1, \ldots, t^k_{q_k} \) of the subproblem

\[ Q(u^k) : \max \{ g(u^k, t) : t \in T \}; \quad (6.1.5) \]

(ii) stop if \( g(u^k, t^k_j) \leq 0, \quad j = 1, \ldots, q_k \); Otherwise choose \( T_{k+1} \) according to a rule guaranteeing the following inclusions

\[ T_{k+1} \subset T_k \cup \{ t^k_1, \ldots, t^k_{q_k} \} \quad \text{and} \quad T_{k+1} \cap \{ t^k_1, \ldots, t^k_{q_k} \} \neq \emptyset. \]

#### 6.1.2 Remark.

Detailed algorithms differ mainly in the choice of \( T_{k+1} \). Determining a global solution of the (in general) non-convex problem (6.1.5) is very costly particularly if \( \dim T \geq 2 \). One can find algorithms, cf. [147, 173, 389], which have important additional features. For instance, the ability to add several constraints \( g(u, t^b_k) \leq 0 \), where \( t^b_k \) are some points with \( g(u^k, t^b_k) > 0 \), or algorithms with efficient rules for deleting constraints. In [247] an algorithm with an automatic grid refinement is suggested, which ensures a linear rate of convergence with respect to the values of the objective functional.

### 6.1.2 Conceptual discretization approach

Algorithms of this type compute a solution of Problem (6.1.1) by solving a sequence of Problems \( P(T_k) \), where \( T_k \) is a grid of mesh-size \( h_k \), i.e. a finite set \( T_k \subset T \) such that

\[ \sup_{t \in T} \rho(t, T_k) \leq h_k. \]
In the $k$-th step of the algorithm a fixed (usually uniform) grid $T_k$ is considered, generated by $h_k := \gamma_k h_{k-1}$, with $\gamma_k \in (0, 1)$ chosen \textit{a priori} or defined within the solution procedure. In step $k$ only subsets $\bar{T}_k$ of $T_k$ are used and in some algorithms one has

$$\gamma_k = \frac{1}{n_{k-1}}, \quad n_{k-1} \in \mathbb{N}, \quad n_{k-1} \geq 2, \quad T_{k-1} \subset T_k \forall k.$$

6.1.3 Algorithm. (Discretization algorithm)

\textit{Step k:} Given $h_{k-1}, \bar{T}_{k-1} \subset T_{k-1}$ and a solution $u^{k-1}$ of $P(T_{k-1})$;

(i) choose $h_k := \gamma_k h_{k-1}$ and generate $T_k$;

(ii) select $\bar{T}_k \subset T_k$;

(iii) compute a solution $\bar{u}$ of $P(\bar{T}_k)$.

If $\bar{u}$ is feasible for $P(T_k)$ within a given accuracy, set $u^k := \bar{u}$, select $\gamma_{k+1}$ and continue with Step (k+1).

Otherwise repeat (ii) for a new choice of $\bar{T}_k$.

6.1.4 Remark. Concerning efficiency, it is important to use as much information as possible from former grids for solving $P(\bar{T}_k)$. Substantial differences among the algorithms of this type lie in the choice of $\bar{T}_k$. An obvious criterion in selecting $\bar{T}_k$ is to ensure that $\bar{T}_k \supset T_\alpha := \{ t \in T_k : g(\bar{u}, t) \geq -\alpha \}$, with $\alpha > 0$ some chosen threshold and $\bar{u}$ the foregoing inner iterate according to Step k(iii) of Algorithm 6.1.3.

The choice of $\alpha$ is crucial: A large value of $\alpha$ leads to many constraints in $P(\bar{T}_k)$ and choosing $\alpha$ too small may have the same effect in subsequent problems, because parts of $T$ may be overlooked (cf. [174, 343]). Applying methods of feasible directions for solving SIP (6.1.1), in [390, 391] some rules are described in order to coordinate the grids and search directions to preserve feasibility of the iterates. This requires information about the last $\bar{T}_{k-1}$ and $u^{k-1}$ in order to choose $\gamma_k$ and the initial set $\bar{T}_k$ at Step $k$. In [325] a discretization is suggested, preserving the rate of convergence of the basic optimization algorithm for the whole solution procedure. In some publications, see for instance [325, 314] and the references quoted there, the stopping rule in Step $k$(iii) of Algorithm 6.1.3 is weakened. In some of these methods (cf. [326, 143]) the convergence results require a sufficient dense initial grid.

The methods sketched above include some ideas which are also important for our abstract regularization scheme and for the algorithms considered in Section 6.2. Particularly this effects the following questions: how to guarantee stability of the solutions of the auxiliary problems by means of stable approximations of the objective function, how to coordinate data (accuracy of approximations and solutions) belonging to former steps together with new grid parameters, or to make use of results of several previous iterations for constructing the next auxiliary problem.
6.1.3 Conceptual reduction approach

These methods are especially recommendable in the final stage of an iteration procedure in order to obtain a high accuracy of the solution. They are constructed under some assumptions allowing a local reduction of SIP which insure, in particular, that for each point \( \bar{u} \), generated by the method, the set

\[
T(\bar{u}) := \{ \tau \in T : g(\bar{u}, \tau) = \max_{t \in T} g(\bar{u}, t) \}
\]

is finite, i.e., \( T(\bar{u}) = \{ \ell_1, ..., \ell_q \} \).

Moreover these assumptions imply that, for \( j = 1, ..., q \), continuous mappings

\[
t_j : U(\bar{u}) \rightarrow U(\ell_j) \cap T,
\]

exist on pairs of neighborhoods \( U(\bar{u}) \), \( U(\ell_j) \cap T \) such that on \( U(\bar{u}) \) the infinitely many original constraints can be equivalently replaced by a finite number

\[
G_j(u) := g(u, t_j(u)) \leq 0, \quad j = 1, ..., q.
\]

6.1.5 Algorithm. (Reduction algorithm) Step k: Given \( u^{k-1} \) (not necessarily feasible);

(i) compute all local maxima \( t_1^{k-1}, ..., t_q^{k-1} \) of \( Q(u^{k-1}) \) (see (6.1.5)) for which \( g(u^{k-1}, t_j^{k-1}) \) are close to \( \max_{t \in T} g(u^{k-1}, t) \);

(ii) execute some steps of a basic algorithm to the reduced finite-dimensional problem

\[
\min \{ J(u) : G_j(u) \leq 0, \quad j = 1, ..., q_{k-1} \},
\]

where \( G_j(u) := g(u, t_j(u)) \) and \( t_j(u) \) are defined according to \( \bar{u} := u^{k-1} \) (\( u^{k-1} \) be the last of these iterates);

(iii) set \( u^k := \bar{u}^k \) and continue with Step \( k+1 \).

6.1.6 Remark. It should be noted that the violation of the assumptions required for the application of the reduction method, in particular finiteness of the set \( T(\bar{u}) \), can be considered as a degenerate situation (cf. Jongen, Jonker and Twilt [200]). The existence of suitable functions \( t_j(\cdot) \) in a neighborhood of a fixed \( \bar{u} \) is closely related with the possibility to apply the implicit function theorem to the Karush-Kuhn-Tucker system of Problem \( Q(\bar{u}) \) (see Hettich and Jongen [175]).

As mentioned in Remark 6.1.2, substep (i) in Algorithm 6.1.5 is very expensive and it is desirable to reduce the number of its executions.

If the functions \( G_j \) are sufficiently smooth, in substep (ii) methods with super-linear rate of convergence should be applied for solving the arising nonlinear programming problems. Sequential quadratic programming (SQP-methods) have been used efficiently in this context (cf. [327, 383, 146]).

In the first instance SQP-methods are locally convergent. Therefore, hybrid techniques combining robust globally convergent descent algorithms with SQP-methods have been developed (cf. [155, 189]). Globalization techniques for SQP-methods were introduced by Han [167], who used for controlling the step-size an exact penalty function of type

\[
\phi(u) = d \sum_{j=1}^{q_k} \max[0, G_j(u)].
\]
6.2 Stable Methods for Ill-posed Convex SIP

To start with some motivation for applying regularization methods to SIP, we illustrate some peculiarities of ill-posed convex SIP under discretization.

6.2.1 Behavior of ill-posed SIP under discretization

Let us consider the following simple examples explaining the typical behavior of SIP under discretization.

6.2.1 Example. Consider the linear SIP

\[ J(u) := -u_1 \rightarrow \min \]
\[ \text{s.t. } u \in U_0 := \{(v_1, v_2) : v_2 \geq 0\}, \]
\[ g(u, t) := u_1 - \left( t - \frac{1}{\sqrt{2}} \right)^2 u_2 \leq 0, \quad \forall t \in T := [0, 1]. \]

Obviously, solutions of this problem are the points \( u^* = (0, a)^T \) for all \( a \geq 0 \).

Let \( T_k \) be a finite \( h_k \)-grid on \([0, 1]\) with

\[ t^*_k := \arg\min_{t \in T_k} |t - \frac{1}{\sqrt{2}}|, \quad \text{but } t^*_k \neq \frac{1}{\sqrt{2}}. \]

For the approximate problems with \( T_k \) instead of \( T \) the whole ray

\[ \{ u \in \mathbb{R}^2 : u_1 = \left( t^*_k - \frac{1}{\sqrt{2}} \right)^2 u_2, u_2 \geq 0 \} \]
is feasible and on this ray $J(u) \to -\infty$.

Choosing a suitable ball $B_r(0)$, we take

$$Q := \{ u \in U_0 \cap B_r(0) : u_1 - \left( t - \frac{1}{\sqrt{2}} \right)^2 u_2 \leq 0 \ \ \forall \ t \in T \},$$

$$Q_k := \{ u \in U_0 \cap B_r(0) : u_1 - \left( t - \frac{1}{\sqrt{2}} \right)^2 u_2 \leq 0 \ \ \forall \ t \in T_k \}$$

and obviously

$$Q := \{ u \in B_r(0) : u_1 \leq 0, u_2 \geq 0 \},$$

$$Q_k := \{ u \in B_r(0) : u_1 \leq \left( t_0^k - \frac{1}{\sqrt{2}} \right)^2 u_2, u_2 \geq 0 \}.$$

For arbitrary $u' \in Q_k \setminus Q$ one has $w' := (\min[0, u_1], u_2)^T \in Q$ and

$$\|u' - w'\| \leq \left( t_0^k - \frac{1}{\sqrt{2}} \right)^2 r,$$

hence, $\rho_H(Q, Q_k) \leq rh^2$.

6.2.2 Example. Consider the linear SIP

$$J(u) := -u_1 \to \min$$

s.t. $u \in U_0 := \{(v_1, v_2) : v_2 \geq 0 \},$

$$g(u, t) \leq 0 \ \ \forall \ t \in T := [0, 1],$$

with $g(u, t) := \max \left[ u_1 - \left( t - \frac{1}{\sqrt{2}} \right)^2 u_2, u_1 - |t - \frac{1}{\sqrt{2}}| \right].$

As before the optimal solutions are $u^* = (0, a)^T$ for all $a \geq 0$.

Choosing $T_k$ and $t_0^k$ as in Example 6.2.1, then

$$u^{*, k} := \left( |t_0^k - \frac{1}{\sqrt{2}}|, b_k \right) \ \ \text{with} \ b_k \geq \frac{1}{|t_0^k - \frac{1}{\sqrt{2}}|}$$

are solutions of the corresponding approximate problems, thus

$$\lim_{k \to \infty} \|u^{*, k}\| = +\infty.$$ 

For the analogous choice of $Q$ and $Q_k$ it is easy to show that

$$\rho_H(Q, Q_k) \leq \min[h_k, rh_k^2].$$

Hence, if for the approximate problems considered in both examples we choose $rh_1 < 1$ and

$$J_k := J,$$

$$K_k := \{ u \in U_0 : g(u, t) \leq 0 \ \ \forall \ t \in T_k \},$$
then Assumption 4.2.2 is fulfilled with $\sigma_k := 0$, $\mu_k := rh_k^2$. Therefore, the Theorems 4.2.4 and 4.3.8 deliver conditions for the controlling parameters which ensure convergence of the OSR- as well as the MSR-method when applied to these problems.

6.2.3 Example. Consider the linear SIP

$$J(u) := -u_1 - u_2 \rightarrow \min$$

s.t. $u \in U_0 := \{(v_1, v_2) : v_1 \geq 0.1, \; v_2 \geq 0.1\}$,

$$g(u, t) \leq 0 \; \forall \; t \in T := [0, 1] \times [0, 1],$$

with

$$g(u, t) := \left(1 - \left(\frac{1}{\sqrt{2}} - t_1\right)^2\right) u_1 + \left(1 - \left(\frac{1}{\sqrt{2}} - t_2\right)^2\right) u_2 - 1.$$ 

Obviously,

$$K := \{u \in \mathbb{R}^2 : u_1 + u_2 \leq 1, \; u_1 \geq 0.1, \; u_2 \geq 0.1\}$$

and the constraint $u_1 + u_2 \leq 1$ corresponds to the unique parameter value $t = (1/\sqrt{2}, 1/\sqrt{2})^T$. The optimal set of this problem is

$$U^* := \left\{ u = \alpha \begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix} + (1 - \alpha) \begin{pmatrix} 0.1 \\ 0.9 \end{pmatrix} : 0 \leq \alpha \leq 1 \right\}.$$ 

Now, we choose an arbitrary sequence of grids $T_k$:

$$0 \leq t_1^1 < t_1^2 < \ldots < t_1^{q(k)} \leq 1,$$

$$0 \leq t_2^1 < t_2^2 < \ldots < t_2^{q(k)} \leq 1,$$

with rational numbers $t_1^1, t_2^1$. Then it is easy to verify that for

$$a_1 := \min_{1 \leq s \leq q(k)} \left| \frac{1}{\sqrt{2}} - t_1^s \right|, \quad a_2 := \min_{1 \leq j \leq q(k)} \left| \frac{1}{\sqrt{2}} - t_2^j \right|$$

the approximate problem has the unique solution:

$$\left(1 - 0.1(1 - a_2^2), 0.1\right) \text{ if } a_1 < a_2$$

and

$$\left(0.1, 1 - 0.1(1 - a_1^2)\right) \text{ if } a_2 < a_1.$$ 

Of course the cases $a_1 < a_2$ and $a_2 < a_1$ have the same probability and, if the discretization parameter $h_k \rightarrow 0$, in general two cluster points arise: $(0.9, 0.1)^T$ and $(0.1, 0.9)^T$. Thus, usually, discretization causes instability for this problem.

To apply MSR-methods described in Section 4.3, we consider Theorem 4.3.6 and Corollary 4.3.7, because for the problem above the feasible set is bounded. Denoting again by $K_k$ the feasible set of the approximate problem with grid parameter $h_k$, we get

$$\rho_H(K_k, K) \leq \frac{3}{4} h_k^2 \quad \text{if } h_k \leq \frac{1}{2}.$$
Indeed, 

$$K_k \subset K' := \{ u \in \mathbb{R}^2 : (1 - h_k^2)u_1 + (1 - h_k^2)u_2 \leq 1, u_1 \geq 0.1, u_2 \geq 0.1 \}$$

and for each \( u \in K' \setminus K \) the point \( (1 - h_k^2)u \) belongs to \( K \). But, due to \( u \in K' \), we have \( \|u\| \leq \frac{1}{1 - h_k^2} \) and, hence,

$$\|u - (1 - h_k^2)u\| \leq \frac{h_k^2}{1 - h_k^2}.$$ 

Thus, in case \( r := 2, \epsilon_k \leq \frac{1}{2}, h_k \leq \frac{1}{2} \) \( \forall k \), relation (4.3.38) is fulfilled. Therefore, Corollary 4.3.7 ensures convergence of the MSR-method if

$$\sum_{k=1}^{\infty} h_k < \infty, \quad \sum_{k=1}^{\infty} \epsilon_k < \infty, \quad \sigma_k := 0$$

and \( \delta_k \) satisfies condition (4.3.5). \( \diamond \)

Hence, under relatively weak assumptions on the choice of the controlling parameters the IPR-schemes, described in Sections 4.2 and 4.3, lead to stable solution procedures for the three examples above.

6.2.4 Remark. The application of the exchange and discretization methods to Example 6.2.1 requires additional constraints ensuring boundedness of the feasible set. Otherwise the iterates would not be defined for any discretization \( T_k \) with \( 1/\sqrt{2} \notin T_k \), because the resulting discretized problem is unsolvable.

Solutions obtained by discretization methods, in general, depend on these additional constraints. The same dependence may also arise in exchange methods because of the inexact computation of \( \arg \max_{t \in T} g(u^k, t) \).

In the case of Example 6.2.2 discretization methods will generate unbounded minimizing sequences and, in Examples 6.2.1 and 6.2.2, the assumptions required for reduction methods are violated. In particular, for the optimal point \( u^* = (0, 0)^T \) we have \( T(u^*) = T \), i.e. \( T(u^*) \) contains infinitely many points.

In the case of Example 6.2.3 all the three approaches are applicable. Nevertheless discretization as well as exchange methods prove to be instable with respect to the argument. For discretization methods this follows immediately from the previous analysis. In exchange methods if

$$\tau \approx \arg \min_{t \in T} g(u^k, t)$$

is computed such that

$$g(u^k, \tau) > 0 \quad \text{and} \quad |\tau_1 - 1/\sqrt{2}| < |\tau_2 - 1/\sqrt{2}|,$$

and if we insert this point \( \tau \) into \( T_{k+1} \) (assuming that \( \tau \) is closer to \( 1/\sqrt{2}, 1/\sqrt{2} \) than any other point in \( T_{k+1} \)), then the next iterate \( u^{k+1} \) will be close to \( (0, 0, 1)^T \) whereas in the opposite case of \( |\tau_1 - 1/\sqrt{2}| > |\tau_2 - 1/\sqrt{2}| \) the point \( u^{k+1} \) is located near \( (0, 1, 0.9)^T \).

For a similar problem obtained by substituting the constraints

$$u_1 \geq 0, u_2 \geq 0 \quad \text{for} \quad u_1 \geq 0.1, u_2 \geq 0.1$$

the assumptions required for reduction methods are violated in the optimal points \( (0, 1)^T \) and \( (1, 0)^T \). \( \diamond \)
6.2. STABLE METHODS FOR ILL-POSED CONVEX SIP

Although the examples considered above are quite simple, the behavior of more complicated SIP under discretization is similar and one could expect that regularization methods are quite natural to solve them on a stable way.

6.2.2 Preliminary results

In the sequel we consider two types of programs with infinitely many constraints: convex SIP and fully parameterized convex SIP.

The first one is given by

\[
\begin{align*}
\text{minimize} & \quad J(u), \\
\text{subject to} & \quad u \in K; \\
K & := \{u \in \mathbb{R}^n : g(u,t) \leq 0, \ \forall \ t \in T\},
\end{align*}
\]

(6.2.1)

(6.2.2)

where

\[
T \text{ is a compact subset of the space } \mathbb{R}^m; \\
J \text{ and } g_u : u \mapsto g(u,t) \text{ are convex, differentiable functions on } \mathbb{R}^n; \\
U^* \neq \emptyset \text{ and the Slater condition is fulfilled, i.e.,}
\]

\[
\exists \ \bar{u} \in \mathbb{R}^n : \sup_{t \in T} g(\bar{u},t) < 0.
\]

(6.2.3)

The second problem can be described in the following way:

\[
\begin{align*}
P(\gamma) \quad \text{minimize} & \quad J(u,\gamma), \\
\text{subject to} & \quad g(u,t,\gamma) \leq 0 \ \forall \ t \in T; \\
\end{align*}
\]

(6.2.4)

where

\[
\gamma \in \Gamma, \Gamma \subset \mathbb{R}^p \text{ is a connected and bounded set;} \\
T \subset \mathbb{R}^m \text{ is a compact set;} \\
J : \mathbb{R}^n \times \Gamma \rightarrow \mathbb{R} \text{ and } g : \mathbb{R}^n \times T \times \Gamma \rightarrow \mathbb{R} \text{ are convex, differentiable functions with respect to } u \text{ for each } t \in T \text{ and } \gamma \in \Gamma, \text{ and continuous with respect to } (t,\gamma) \in T \times \Gamma \text{ for each } u \in \mathbb{R}^n.
\]

We assume that for some \(\gamma_0 \in \Gamma\) an approximate solution of \(P(\gamma_0)\) is known. Then, for a given \(\bar{\gamma} \in \Gamma\), or for several values \(\gamma_i \in \Gamma \ (i = 1, \ldots, s)\), an optimal solution of \(P(\bar{\gamma}_i)\), respectively \(P(\gamma_1), P(\gamma_2), \ldots, P(\gamma_s)\), is sought.

Of course, Problem \(P(\bar{\gamma})\) could be attacked at ones. However, we expect that it would be more efficient to determine the solution of \(P(\bar{\gamma})\) by following a certain path from \(\gamma_0\) to \(\bar{\gamma}\). As a by-product we will see that one can obtain approximate solutions of the problems \(P(\gamma)\) at the chosen intermediate values \(\gamma\) together with two-sided error bounds of these solutions.

Sometimes it may be useful to introduce artificially a family of Problems \(P(\gamma)\) in order to solve a SIP of type (6.2.1). In this case the family \(P(\gamma)\) has to be constructed such that Problem (6.2.1) corresponds to the parameter \(\gamma := \bar{\gamma}\). This approach makes sense if the starting problem \(P(\gamma_0)\) can be solved easily and the path-following procedure is efficient. Path-following or continuation strategies were used in finite-dimensional non-linear programming, for instance, by Gfrerer et al. [127], Guddat et al. [152] and Gollmer et al. [138]. Of course, applying fast convergent methods, the choice of the parameters
\(\gamma_1, \gamma_2, \ldots\) and the successive solution of the arising problems should be performed such that the last iterate of the previous problem serves as a suitable starting point for the next problem. The structural dependence of feasible and optimal sets with respect to the parameters, which have to be taken into account along the path, can be classified according to the structural theory developed by Jongen et al. [200], Kojima [243] and Kojima and Hirabayashi [244].

Concerning parametric SIP a similar strategy has been studied by Rupp [358]. As in finite optimization the path-following idea is applied to the corresponding Karush-Kuhn-Tucker system, which changes correspondingly with the active set. A theoretical analysis of parametric SIP can be found by Brosowski [55, 56].

The initial situation, when a family of Problems (6.2.4) is given, may be substantial for analyzing the parametric problem. Indeed, often the value of the final parameter \(\bar{\gamma}\), in which we are interested, is a priori unknown and only tentatively determined. Therefore, bounds of solutions of the type (4.3.39), (4.3.40), given for the intermediate problems \(P(\gamma_i)\), enable us to specify the sought value of \(\bar{\gamma}\) (cf. Theorem 6.3.7 below).

Naturally, in order to study IPR for path-following methods we need some additional requirements on the family \(P(\gamma)\) which will be formulated below.

Next we prove some auxiliary statements for SIP under weaker assumptions on the smoothness of the functions involved than those required in (6.2.1) - (6.2.4).

**6.2.5 Lemma.** Assume that \(f : \mathbb{R}^n \to \bar{\mathbb{R}}\) is a convex lsc. function and \(\text{dom } f \supset \mathbb{B}_r\) with \(r > 0\) fixed. Then for any \(\bar{v} \in \mathbb{B}_r\) the function

\[
f(u) + \|u - \bar{v}\|^2
\]

attains its unconstrained minimum on the set \(\mathbb{B}_\rho\) for

\[
\rho := 2r + \frac{1}{r} (\beta - f(0)), \quad \beta := \max_{u \in \mathbb{B}_r} f(u).
\]

**Proof:** For any subgradient \(q \in \partial f(0), q \neq 0\), it holds

\[
f(u) - f(0) \geq \langle q, u \rangle
\]

and for \(u := r \frac{q}{\|q\|}\) we have

\[
r\|q\| \leq f(r \frac{q}{\|q\|}) - f(0) \leq \beta - f(0).
\]

We verify that

\[
f(u) + \|u - \bar{v}\|^2 > f(0) + \|\bar{v}\|^2, \quad \forall \bar{v} \in \mathbb{B}_r, u \notin \mathbb{B}_\rho.
\]

Indeed, in view of (6.2.5) this inequality is obvious if

\[
\langle q, u \rangle + \|u - \bar{v}\|^2 > \|\bar{v}\|^2.
\]

Rewriting (6.2.7) in the form

\[
\|u - \bar{v} + \frac{1}{2} q\|^2 - \|\bar{v} - \frac{1}{2} q\|^2 > 0,
\]
we conclude that (6.2.7) is fulfilled if \( \| u - \bar{v} + \frac{1}{2}q \| > \| \bar{v} - \frac{1}{2}q \| \times 2 \) and, hence, if \( \| u \| > 2\| \bar{v} - \frac{1}{2}q \| \) is satisfied. But, due to the choice of the radius \( \rho \), the point \( \bar{v} \) and relation (6.2.6), the latter inequality holds for \( u \notin B_\rho \).

The case \( \partial f(0) = \{ 0 \} \) is trivial. □

6.2.6 Corollary. Let the hypotheses of Lemma 6.2.5 be satisfied and assume that \( \tilde{f} : \mathbb{R}^n \to \mathbb{R} \) is a convex lsc. function such that

\[ \sup_{u \in B_r} |f(u) - \tilde{f}(u)| \leq \sigma < \infty. \] (6.2.8)

Then, for any \( \tilde{v} \in B_r \), the function

\[ \tilde{f}(u) + \| u - \tilde{v} \|^2 \]

attains its unconstrained minimum on the set \( B_\tilde{\rho} \) for

\[ \tilde{\rho} := 2r + \frac{1}{r}(\beta - f(0) + 2\sigma) \].

Indeed, it suffices to remark that for any \( q \in \partial \tilde{f}(0) \) with regard to (6.2.8) the following estimate holds

\[ r \| q \| \leq \tilde{f}(r \frac{q}{\| q \|}) - \tilde{f}(0) \leq \beta - f(0) + 2\sigma. \]

Now, for convex, lsc. functions \( f, g, \tilde{f}, \tilde{g} \), mapping from \( \mathbb{R}^n \) into \( \mathbb{R} \), we consider the problems

\[ \min \{ f(u) + \| u - \tilde{v} \|^2 : g(u) \leq 0 \} \] (6.2.9)

and

\[ \min \{ \tilde{f}(u) + \| u - \tilde{v} \|^2 : \tilde{g}(u) \leq 0 \}. \] (6.2.10)

If condition (6.2.14) below is true, then both problems are solvable for any \( \tilde{v} \).

The following estimates for the corresponding Lagrange multipliers \( \lambda(\tilde{v}) \) and \( \tilde{\lambda}(\tilde{v}) \) of these problems can be established.

6.2.7 Lemma. Suppose that for the functions involved in the Problems (6.2.9) and (6.2.10) the following relations are fulfilled for fixed radius \( r > 0 \) and \( \beta := \max_{u \in \partial B_r} f(u) \):

\[ \text{dom} f \supset B_{\tilde{\rho}}, \quad \text{with} \quad \tilde{\rho} := 2r + \frac{1}{r}(\beta - f(0) + 2\sigma), \] (6.2.11)

\[ \sup_{u \in B_r} |f(u) - \bar{f}(u)| \leq \sigma, \] (6.2.12)

\[ \sup_{u \in B_\tilde{\rho}} |f(u) - \bar{f}(u)| \leq \tilde{\sigma}, \] (6.2.13)

\[ \max \left[ \min_{u \in B_r} g(u), \min_{u \in B_\tilde{\rho}} \tilde{g}(u) \right] < -\alpha < 0. \] (6.2.14)
CHAPTER 6. PPR FOR CONVEX SEMI-INFINITE PROBLEMS

Then, for each \( \bar{v} \in \mathbb{B}_r \), the Lagrange multiplier of Problem (6.2.10) can be estimated by

\[
\bar{\lambda}(\bar{v}) \leq \frac{1}{\alpha} (f(\bar{u}) + (\|\bar{u}\| + r)^2 + \sigma + \bar{\sigma} + r^2 - c) =: \theta \tag{6.2.15}
\]

holds with

\[
c := \min_{v \in \mathbb{R}^n} \{f(v) + \frac{1}{2} \|v\|^2\}, \quad \bar{u} \in \mathbb{B}_r, \quad \bar{g}(\bar{u}) < -\alpha.
\]

Proof: Let \( \bar{u}^*(\bar{v}) \) be an optimal solution of Problem (6.2.10). Due to the saddle point inequality

\[
f(\bar{u}) + \|u - \bar{v}\|^2 + \bar{\lambda}(\bar{v})\bar{g}(u) \geq \bar{f}(\bar{u}^*(\bar{v})) + \|\bar{u}^*(\bar{v}) - \bar{v}\|^2, \quad \forall \ u \in \mathbb{R}^n, \tag{6.2.16}
\]

we get from (6.2.14) and (6.2.16)

\[
f(\bar{u}) + \|\bar{u} - \bar{v}\|^2 - \bar{\lambda}(\bar{v})\alpha \geq \bar{f}(w^*(\bar{v})) + \|w^*(\bar{v}) - \bar{v}\|^2,
\]

with \( w^*(\bar{v}) := \arg \min_{u \in \mathbb{R}^n} \{\bar{f}(u) + \|u - \bar{v}\|^2\} \).

In view of (6.2.12) and Corollary 6.2.6 the inclusion \( w^*(\bar{v}) \in \mathbb{B}_\tilde{\beta} \) is true and, with regard to (6.2.12) and (6.2.13),

\[
f(\bar{u}) + (\|\bar{u}\| + r)^2 - \bar{\lambda}(\bar{v})\alpha \geq f(w^*(\bar{v})) + \|w^*(\bar{v}) - \bar{v}\|^2 - \tilde{\sigma}
\]

Taking into account that the quadratic function

\[
\frac{1}{2} t^2 - 2t\|\bar{v}\| + \|\bar{v}\|^2
\]

attains its minimum at the point \( t := 2\|\bar{v}\| \), we get

\[
\|v - \bar{v}\|^2 \geq \frac{1}{2} \|v\|^2 - 2\|v\|\|\bar{v}\| + \|\bar{v}\|^2 + \frac{1}{2} \|v\|^2
\]

\[
\geq -\|\bar{v}\|^2 + \frac{1}{2} \|v\|^2 \geq \frac{1}{2} \|v\|^2 - r^2.
\]

This inequality together with (6.2.17) leads to

\[
f(\bar{u}) + (\|\bar{u}\| + r^2) - \bar{\lambda}(\bar{v})\alpha + \sigma \geq \min_{v \in \mathbb{R}^n} \{f(v) + \frac{1}{2} \|v\|^2\} - r^2 - \tilde{\sigma},
\]

and the statement of the lemma follows immediately. \( \square \)

6.2.8 Remark. It should be noted that the right-hand side in (6.2.15) does not depend on \( \bar{v} \). Moreover, it is obvious that for the multiplier of Problem (6.2.9) it holds

\[
\lambda(\bar{v}) \leq \frac{1}{\alpha} (f(\bar{u}) + (\|\bar{u}\| + r)^2 + r^2 - c).
\]

This estimate follows formally from Lemma 6.2.7 too if we take

\[
\tilde{f} := f \quad \text{and} \quad \sigma = \tilde{\sigma} := 0.
\]

\( \diamond \)
6.2. STABLE METHODS FOR ILL-POSED CONVEX SIP

6.2.9 Remark. If in Lemma 6.2.7

\[ \tilde{g}(\cdot) := \sup_{1 \leq j \leq m} g_j(\cdot), \]

with \( g_j \) convex, lsc. functions, then estimate (6.2.15) is true for each Lagrange multiplier \( \tilde{\lambda}_j(\bar{v}) \) of the problem

\[
\min \{ \tilde{f}(u) + \| u - \bar{v} \|^2 : g_j(u) \leq 0, \ j = 1, \ldots, m \},
\]

moreover, \( \sum_{j=1}^m \tilde{\lambda}_j(\bar{v}) \leq \theta \) in (6.2.15).

Indeed, it suffices to use instead of (6.2.16) the inequality

\[
\tilde{f}(u) + \| u - \bar{v} \|^2 + \tilde{g}(u) \sum_{j=1}^m \tilde{\lambda}_j(\bar{v}) \geq \tilde{f}(\tilde{u}^*(\bar{v})) + \| \tilde{u}^*(\bar{v}) - \bar{v} \|^2.
\]

\[ \diamond \]

For further description denote \( \| \cdot \|_p \) the Euclidean norm in the space \( \mathbb{R}^p \) (for \( p = n \) we write as before \( \| \cdot \| \)).

Now, let us return to the parameterized SIP (6.2.4) and estimate the Hausdorff distance between the feasible sets of the original and the discretized problem, where the latter is approximated by means of a finite grid \( T_h \).

Suppose that positive numbers \( h, \sigma, r \) are given, \( T_h \) is a finite \( h \)-grid and the vector \( \gamma_{\sigma} \in \Gamma \) is chosen such that \( \| \gamma_{\sigma} - \bar{\gamma} \|_p \leq \sigma \).

6.2.10 Lemma. Suppose that for each \( t \in T, \gamma \in \Gamma \) the function \( g(\cdot, t, \gamma) : \mathbb{R}^n \to \mathbb{R} \) is convex and lsc. Assume further that \( \text{dom} g(\cdot, t, \gamma) \supset \mathbb{B}_r \) and that for arbitrary \( t, t' \in T, \gamma' \in \Gamma \) and some constant \( L > 0 \) the inequality

\[
\sup_{u \in \mathbb{B}_r} | g(u, t', \gamma') - g(u, t, \bar{\gamma}) | \leq L(\| t - t' \|_m + \| \gamma' - \bar{\gamma} \|_p)
\]

is fulfilled, and there exists a point \( \tilde{u} \in \mathbb{B}_r \) with

\[
\sup_{t \in T} g(\tilde{u}, t, \bar{\gamma}) \leq -\alpha < 0.
\]

Then, for \( h + \sigma < \frac{\alpha}{2L} \), the estimate

\[
\rho_H(Q, Q_h) \leq \frac{4}{\alpha} L(h + \sigma)r
\]

holds, where

\[
Q := \{ u \in \mathbb{B}_r : \sup_{t \in T} g(u, t, \bar{\gamma}) \leq 0 \},
\]

\[
Q_h := \{ u \in \mathbb{B}_r : \sup_{t \in T_h} g(u, t, \gamma_{\sigma}) \leq 0 \},
\]

with \( \| \gamma_{\sigma} - \bar{\gamma} \|_p \leq \sigma \).
Proof: Denote

\[
\begin{align*}
\bar{g}(u) & := \max \left\{ \sup_{t \in T} g(u, t, \gamma), \sup_{t \in T_h} g(u, t, \gamma^\alpha) \right\}, \\
\tilde{g}(u) & := \sup_{t \in T_h} g(u, t, \gamma^\alpha), \\
Q & := \{ u \in B_r : \bar{g}(u) \leq 0 \}.
\end{align*}
\]

In this proof we always assume that \( u \in B_r \).

Because \( \bar{g}(u) \geq \tilde{g}(u) \ \forall \ u \), inclusion \( Q \subset Q_h \) holds and it is obvious that

\[
\rho_H(Q, Q_h) \leq \max \left[ \rho_H(Q, \bar{Q}), \rho_H(Q_h, \bar{Q}) \right].
\]

Now we are going to estimate both distances \( \rho_H(Q, \bar{Q}) \) and \( \rho_H(Q_h, \bar{Q}) \).

Denote \( \sum := (T \times \{ \gamma \}) \cup (T_h \times \{ \gamma^\alpha \}) \), and for arbitrary \( u \in B_r \), we choose a point \((\tau(u), \gamma(u)) \in \sum \) such that \( \bar{g}(u) = g(u, \tau(u), \gamma(u)) \) and take

\[
\tau_h(u) := \arg \min_{t \in T_h} \| t - \tau(u) \|_m.
\]

With regard to the definition of \( \tilde{g} \) the inequality \( \bar{g}(u) \geq g(u, \tau_h(u), \gamma^\alpha) \) is satisfied and, due to (6.2.18),

\[
\bar{g}(u) - \tilde{g}(u) \leq \sup_{t \in T_h} g(\bar{u}, t, \gamma^\alpha) - \sup_{t \in T} g(\bar{u}, t, \gamma) < L \sigma
\]

hence,

\[
\sup_{t \in T_h} g(\bar{u}, t, \gamma^\alpha) - \sup_{t \in T} g(\bar{u}, t, \gamma) < L \sigma
\]

and

\[
\tilde{g}(\bar{u}) = \sup_{t \in T_h} g(\bar{u}, t, \gamma^\alpha) < -\alpha + L \sigma < -\frac{\alpha}{2}.
\]

If \( \bar{g}(u) \leq -L(h + \sigma) \), then \( \bar{g}(u) \leq 0 \) holds in view of (6.2.23), consequently \( u \in Q \).

The case \( 0 \geq \bar{g}(u) > -L(h + \sigma) \) leads to \( \bar{g}(u) \geq -L(h + \sigma) \) and, because of \( h + \sigma < \frac{\alpha}{2} \), inequality \( \bar{g}(u) \leq \tilde{g}(u) \) is fulfilled.

Now, due to the convexity of \( \tilde{g} \), for a point

\[
z := u + \frac{\alpha - 2L(h + \sigma)}{\alpha}(u - \bar{u}),
\]

we have

\[
\bar{g}(z) \leq \frac{2L(h + \sigma)}{\alpha} \bar{g}(\bar{u}) + \frac{\alpha - 2L(h + \sigma)}{\alpha} \tilde{g}(u) \leq -L(h + \sigma). \tag{6.2.25}
\]

Because the points \( \bar{u} \) and \( u \) belong to \( B_r \), also \( z \in B_r \) and the inequalities (6.2.23), (6.2.25) imply \( \bar{g}(z) \leq 0 \), thus \( z \in \bar{Q} \).

But \( u \in Q_h \) and, due to (6.2.24),

\[
\| z - u \| = \frac{2L(h + \sigma)}{\alpha} \| \bar{u} - u \| \leq \frac{4L(h + \sigma)}{\alpha} r
\]
holds true, consequently,
\[ \rho_H(Q_h, \bar{Q}) \leq \frac{4L(h + \sigma)}{\alpha} r. \]

Using instead of \( \bar{g} \) the function \( \max_{t \in T} g(u, t, \bar{\gamma}) \), we obtain analogously
\[ \rho_H(Q, \bar{Q}) \leq \frac{4L(h + \sigma)}{\alpha} r. \]

\[ \square \]

6.2.11 Corollary. If \( g \) is independent of \( \gamma \) and \( h < \frac{\alpha}{L} \), then
\[ \rho_H(Q, Q_h) \leq \frac{2Lh}{\alpha} r \]
holds and the inclusion \( Q \subset Q_h \) is obvious.

This result follows immediately from the proof above if we take \( \sigma := 0 \) and \( z := \tilde{u} + \frac{\alpha - Lh}{\alpha}(u - \tilde{u}) \).

6.2.3 Regularized penalty methods for SIP

According to the convergence results of MSR-methods in Section 4.3 in the methods considered in the sequel a successive approximation of SIP and a regularization of the approximate auxiliary problems is performed. However, in comparison with the abstract scheme, where it is supposed that the solution of each regularized auxiliary problem is computed up to a given accuracy, here only one step of a certain penalty method is carried out. This requires a coordinated choice between the penalty and controlling parameters used in the general scheme of MSR-methods.

The first two algorithms described below are suggested for solving convex SIP (6.2.1) and the third algorithm solves parametric SIP of type \( P(\gamma) \) (cf. (6.2.4)).

Suppose function \( g \) in Problem (6.2.1) satisfies the following Lipschitz condition
\[ \sup_{u \in B_r} |g(u, t) - g(u, t')| \leq L_t(r) \| t - t' \|_m \] (6.2.26)
with some \( L_t(r) < \infty \) (for arbitrary \( r > 0 \)) and \( t, t' \in T \). In the sequel, we usually omit the radius \( r \) in the notation of the Lipschitz constant if it is clear from the context.

Choosing any finite subset \( T' \subset T \) and any \( \bar{v} \in B_r \), due to the Remarks 6.2.8 and 6.2.9, the following estimate of the Lagrange multipliers \( \lambda(\bar{v}, t)_{t \in T'} \) of the problem
\[ \min \{ J(u) + \| u - \bar{v} \|^2 : g(u, t) \leq 0 \ \forall \ t \in T' \} \] (6.2.27)
holds:
\[ \max_{t \in T'} \lambda(\bar{v}, t) \leq \frac{1}{\alpha} \left( J(\bar{u}) + (\| \bar{u} \| + r)^2 + r^2 - c \right) =: c(r) \] (6.2.28)
with \( \bar{u} \) a Slater point of Problem (6.2.1),
\[ -\alpha \geq \sup_{t \in T} g(\bar{u}, t) \] (6.2.29)
and $c := \min_{v \in \mathbb{R}^n} \{ J(v) + \frac{1}{2} \|v\|^2 \}$. Because the right-hand side in (6.2.28) does not depend on $\bar{v}$ and $T'$, this estimate is uniform with respect to $\bar{v} \in B_r$ and $T' \subset T$ for finite $T'$.

Regarding Remark A3.4.44, if for a penalty parameter $d$ it holds $d > c(r)$, then Problem (6.2.27) is equivalent to the unconstrained minimization problem

$$\min_{u \in \mathbb{R}^n} \left\{ J(u) + \frac{d}{2} \sum_{t \in T'} (g(u, t) + |g(u, t)|) + \|u - \bar{v}\|^2 \right\}. \quad (6.2.30)$$

Now we turn to the description of solution methods for Problem (6.2.1). In order to control such methods, we need several parameters: a discretization parameter $h_k$ for defining the grid $T_k$, a termination parameter $\delta_k$ for stopping the interior loop of the MSR-method, an accuracy parameter $\epsilon_k$ for the inexact solving of the regularized auxiliary problems and an smoothing parameter $\tau_k$ for approximating the exact penalty function in (6.2.30).

We take sequences $\{h_k\}$, $\{\delta_k\}$, $\{\epsilon_k\}$ and $\{\tau_k\}$ of positive numbers satisfying

$$\lim_{k \to \infty} h_k = \lim_{k \to \infty} \epsilon_k = \lim_{k \to \infty} \tau_k = 0, \quad h_k < \frac{\alpha}{L_t},$$

and choose some radius $r > 0$ such that

$$U^* \cap B_{r/8} \neq \emptyset. \quad (6.2.31)$$

Solving SIP (6.2.1), in each of the following methods a minimizing sequence $\{u^{k,i}\}, \ (k, i) : \ i = 0, ..., i(k), \ k = 1, 2, ...$, has to be generated and, as usual for MSR-methods, the termination values $i(k)$ of the interior loops are defined within the iterations.

For a chosen family $\{T_k\}$ of finite $h_k$-grids on $T$ we consider the approximations

$$K_k := \{ u \in \mathbb{R}^n : g(u, t) \leq 0 \ \forall \ t \in T_k \}, \ k = 1, 2, ..., \quad (6.2.32)$$

of the feasible set (6.2.2) and introduce the functions

$$\varphi_k(u) := \frac{d}{2} \sum_{t \in T_k} \left( g(u, t) + \sqrt{g^2(u, t) + \tau_k} \right), \quad (6.2.33)$$

$$F_{k,i}(u) := J(u) + \varphi_k(u) + \|u - u^{k,i-1}\|^2, \quad (6.2.34)$$

with $d > c(r)$ and $c(r)$ defined by (6.2.28).

6.2.12 Algorithm.

S0: Set $k := 1$, $i := 0$, $i(1) := 1$ and choose $u^{1,0} \in K_1 \cap B_{r/4}$.

S1: If $i < i(k)$, set $i := i + 1$,

if $i := i(k)$, set $u^{k+1,0} := u^{k,i(k)}$, $k := k + 1$, $i := 1$.

S2: Compute $u^{k,i}$ such that $\|\nabla F_{k,i}(u^{k,i})\| \leq \epsilon_k$.

S3: If $\|u^{k,i} - u^{k,i-1}\| > \delta_k$, set $i(k) := i + 1$ and go to S1;

otherwise set $i(k) := i$ and go to S1.
In this method each iterate is computed at the \( k \)-th approximation level with regard to all constraints describing the set \( K_k \).

Now we are going to describe a modified algorithm possessing an adaptive deleting procedure for constraints describing the set \( K_k \). For this modification we need additionally a Lipschitz constant \( L_u \) such that

\[
\sup_{t \in T} |g(u, t) - g(u', t)| \leq L_u \|u - u'\|, \quad \forall \ u, u' \in \mathbb{B}_r. \tag{6.2.35}
\]

Let \( q(r) := \frac{2}{d} L_t r \) with \( L_t \) via (6.2.26), and just as in the general scheme (cf. (4.1.3)) we assume that

\[
\sup_{u \in \mathbb{B}_r} \|\nabla J(u)\| \leq L(r) \tag{6.2.36}
\]

with a given non-decreasing function \( L(r) : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \).

For subsets \( T_{k,i} \subset T_k \) specified within the following algorithm, we define the sets

\[
K_{k,i} := \{ u \in \mathbb{R}^n : g(u, t) \leq 0 \quad \forall \ t \in T_{k,i} \}, \quad i = 1, \ldots, i(k), \quad k = 1, 2, \ldots \tag{6.2.37}
\]

and functions

\[
\varphi_{k,i} := \frac{d}{2} \sum_{t \in T_{k,i}} \left( g(u, t) + \sqrt{g^2(u, t) + 2 \tau_k} \right), \quad (d > c(r)),
\]

\[
\mathcal{F}_{k,i} := J(u) + \varphi_{k,i}(u) + \|u - u^{k,i}\|^2.
\]

### 6.2.13 Algorithm.

**S0:** Set \( k := 1, T_{1,1} := T_1 \) and choose \( u^{1,0} \in K_1 \cap \mathbb{B}_{r/4} \).

Compute \( u^{1,1} \) such that \( \|\nabla \mathcal{F}_{1,1}(u^{1,1})\| \leq \epsilon_1 \) and go to S3.

**S1:** Determine \( \hat{\epsilon}_{k,i} := \frac{\sqrt{d}}{2} \sqrt{T_{k,i}} \).

If \( i < i(k) \) for a given \( k \), generate

\[
T_{k,i+1} := \{ t \in T_k : g(u^{k,i}, t) + L_u(\|u^{k,i} - u^{k,i-1}\| + \hat{\epsilon}_{k,i}) \geq 0 \},
\]

set \( i := i + 1 \);

if \( i = i(k) \), set

\[
\begin{align*}
\hat{u}^{k+1,0} &:= u^{k,i(k)}, \\
\hat{\epsilon}_{k+1,0} &:= \sqrt{(L(r) + 4r)q(r)h_k + q(r)h_k + \hat{\epsilon}_{k,i(k)}}, \\
T_{k+1,1} &:= \{ t \in T_{k+1} : g(u^{k+1,0}, t) + L_u(\|u^{k+1,0} - u^{k,i(k)-1}\| + \hat{\epsilon}_{k+1,0}) \geq 0 \};
\end{align*}
\]

\( k := k + 1, i := 1 \).

**S2:** Compute \( u^{k,i} \) such that \( \|\mathcal{F}_{k,i}(u^{k,i})\| \leq \epsilon_k \).

**S3:** If \( \|u^{k,i} - u^{k,i-1}\| > \delta_k \), set \( i(k) := i + 1 \) and go to S1;

otherwise set \( i(k) := i \) and go to S1.
The constants \( \hat{\epsilon}_{k,i} \) and \( \hat{\epsilon}_{k+1,0} \) serve as deleting levels for the interior and exterior loops, respectively.

According to the general scheme in Subsection 4.3.2 denote
\[
Q := K \cap \mathbb{B}_r, \quad Q_k := K_k \cap \mathbb{B}_r, \quad k = 1, 2, \ldots
\]

**6.2.14 Remark.** In view of \( T_k \subset T \) we have an outer approximation \( Q \subset Q_k \) for all \( k \) and Corollary 6.2.11 provides for \( k = 1, 2, \ldots \)
\[
\rho_H(Q, Q_k) \leq q(r) h_k,
\]
\[
\rho_H(Q, Q_{k+1}) \leq q(r) \max\{h_k, h_{k+1}\}.
\]

Although a similar estimate \( \rho_H(Q, K_{k,i} \cap \mathbb{B}_r) \) is not true, moreover, sequence \( \{K_{k,i} \cap \mathbb{B}_r\} \) does not converge to \( Q \) in general.

**6.2.15 Assumption.**

(i) Functions \( J(\cdot, \gamma) \) and \( g(\cdot, t, \gamma) \) are convex and belong to \( C^1(\mathbb{R}^n) \) for each \( \gamma \in \Gamma \) and each \( t \in T \);

Starting with some \( r_0 > 0 \), for \( r \geq r_0 \)

(ii) some solutions of \( P(\gamma_0) \) and \( P(\bar{\gamma}) \) belong to \( \mathbb{B}_r \) and

(iii) there exists \( \bar{u} \in \mathbb{B}_r \) and a constant \( a > 0 \) with \( \sup_{t \in T} g(\bar{u}, t, \bar{\gamma}) \leq -a \).

The following conditions are fulfilled for each \( r > 0 \):

(iv) \( \sup_{u \in \mathbb{B}_r} \|\nabla J(u, \gamma)\| \leq L(r) \) for a given function \( L(r) : \mathbb{R}_+ \to \mathbb{R}_+ \);

(v) there is a constant \( L^f_r \) with
\[
\sup_{u \in \mathbb{B}_r} |J(u, \gamma) - J(u, \gamma)| \leq L^f_r \|\gamma - \bar{\gamma}\| \quad \forall \gamma \in \Gamma;
\]

(vi) there is a constant \( L^g_r \) such that for all \( t, t' \in T, \gamma \in \Gamma \)
\[
\sup_{u \in \mathbb{B}_r} |g(u, t, \gamma) - g(u, t', \gamma)| \leq L^g_r (\|t - t'\|_m + \|\gamma - \bar{\gamma}\|_p);
\]

(vii) for every \( u, u' \in \mathbb{B}_r \) and some \( \bar{L}^g_r \)
\[
\sup_{t \in T, \gamma \in \Gamma} \sup_{u \in \mathbb{B}_r} |g(u, t, \gamma) - g(u', t, \gamma)| \leq \bar{L}^g_r \|u - u'\|.
\]

Now, let \( r \geq r_0 \) be chosen such that
\[
U^*(\bar{\gamma}) \cap \mathbb{B}_{r/8} \neq \emptyset, \quad u^{1,0} \in \mathbb{B}_{r/4}, \quad (6.2.38)
\]
with \( U^*(\bar{\gamma}) \) the optimal set of \( P(\bar{\gamma}) \) and \( u^{1,0} := u^*(\gamma_0) \) a known solution of \( P(\gamma_0) \).
6.2. STABLE METHODS FOR ILL-POSED CONVEX SIP

In order to control the iteration procedure, we choose sequences \( \{h_k\}, \{\delta_k\}, \{\epsilon_k\} \) and \( \{\tau_k\} \) of positive numbers and a non-negative sequences \( \{\sigma_k\} \) such that

\[
\lim_{k \to \infty} h_k = \lim_{k \to \infty} \sigma_k = \lim_{k \to \infty} \epsilon_k = \lim_{k \to \infty} \tau_k = 0,
\]

\[
h_k + \sigma_k < \frac{a}{2L_k^p}, \quad k = 1, 2, ..., \tag{6.2.39}
\]

where \( L_k^p \) is given according to Assumption 6.2.15(vi).

For an arbitrary finite subset \( T' \subset T \) and \( \gamma_k \in \Gamma \) such that \( \|\gamma_k - \bar{\gamma}\|_p \leq \sigma_k \), the estimate

\[
\sup_{t \in T'} g(\bar{u}, t, \gamma_k) \leq -a + L_k^s \sigma_k < -\frac{a}{2}
\]

was established in the proof of Lemma 6.2.10. Thus, if \( \bar{v} \in B_r \), the assumptions of Lemma 6.2.7 are satisfied for the pair of problems

\[
\min \{J(u, \bar{\gamma}) + \|u - \bar{v}\|^2 : g(u, t, \bar{\gamma}) \leq 0 \quad \forall \ t \in T' \}
\]

and

\[
\min \{J(u, \gamma_k) + \|u - \bar{v}\|^2 : g(u, t, \gamma_k) \leq 0 \quad \forall \ t \in T' \} \tag{6.2.40}
\]

with

\[
f := J(\cdot, \bar{\gamma}), \quad \bar{f} := J(\cdot, \gamma_k), \quad g := \sup_{t \in T} g(\cdot, t, \bar{g}), \quad \bar{g} := \sup_{t \in T'} g(\cdot, t, \gamma_k), \]

\[
\alpha := \frac{a}{2}, \quad \sigma := L_k^s \sup \sigma_k, \quad \bar{\sigma} := L_k^s \sup \sigma_k, \tag{6.2.41}
\]

and \( \bar{\rho} \) defined by (6.2.11).

6.2.16 Remark. Taking into account Remark 6.2.9, Lagrange multipliers which correspond to Problem (6.2.40) can be estimated by (6.2.15) uniformly with respect to \( \bar{v} \in B_r \), as well as the iterations steps \( k \) and \( T' \subset T \). \( \diamond \)

Now we are going to describe a MSR-method for Problem \( P(\bar{\gamma}) \) working with a deleting rule for the discretized constraint set. In the sequel the penalty parameter \( \delta \) is chosen such that \( \delta > \theta \), where \( \theta \) is calculated according to (6.2.11) - (6.2.15) with data from (6.2.41).

In the \( k \)-th step we fix a finite \( h_k \)-grid \( T_k \) on the compact set \( T \) and a parameter \( \gamma_k \in \Gamma \) such that \( \|\gamma_k - \bar{\gamma}\|_p \leq \sigma_k \). Then the sets

\[
K_k := \{u \in \mathbb{R}^n : g(u, t, \gamma_k) \leq 0 \quad \forall \ t \in T_k\}, \quad k = 1, 2, ..., \tag{6.2.42}
\]

can be constructed. Analogously to Algorithm 6.2.13, for a subset \( T_{k,j} \subset T_k \) described below, we define the set

\[
K_{k,j} := \{u \in \mathbb{R}^n : g(u, t, \gamma_k) \leq 0 \quad \forall \ t \in T_{k,j}\} \tag{6.2.43}
\]

and the functions

\[
\varphi_{k,j}(u) := \frac{d}{2} \sum_{t \in T_{k,j}} \left( g(u, t, \gamma_k) + \sqrt{g^2(u, t, \gamma_k) + \tau_k} \right), \tag{6.2.44}
\]

\[
F_{k,j}(u) := J(u, \gamma_k) + \varphi_{k,j}(u) + \|u - u^{k,j-1}\|^2 \tag{6.2.45}
\]
and set
\[ \bar{q}(r) := \frac{4}{a}L^a g r. \] (6.2.46)

Now starting with a known solution \( u^*(\gamma_0) \) of \( P(\gamma_0) \), a minimizing sequence
\[ \{u^{k,i}\}, \quad i = 0, ..., i(k), \quad k = 1, 2, ... \]
will be generated which solves \( P(\bar{\gamma}) \).

6.2.17 Algorithm.

S0: Set \( k := 1, \ i := 1, \ T_{1,1} := T_1, \ u^{1,0} := u^*(\gamma_0) \).

Compute \( u^{1,1} \) such that \( \|\nabla F_{1,1}(u^{1,1})\| \leq \epsilon_1 \) and go to S3.

S1: Determine \( \hat{\epsilon}_{k,i} := \frac{\epsilon_1}{2} + \sqrt{\frac{\epsilon_1}{2} |T_{k,i}|} \).

If \( i < i(k) \) for a given \( k \), generate
\[ T_{k,i+1} := \{ t \in T_k : g(u^{k,i}, t, \gamma_k) + L^g \left( \|u^{k,i} - u^{k,i-1}\| + 2\sqrt{L^f \sigma_k + \hat{\epsilon}_{k,i}} \right) \geq 0 \}, \]
set \( i := i + 1; \)
if \( i = i(k) \), take
\[ u^{k+1,0} := u^{k,i(k)}, \]
\[ \tilde{\epsilon}_{k+1,0} := \sqrt{2L^g} \tilde{q}(r)(h_k + \sigma_k) + 2\tilde{q}(r)(h_k + \sigma_k) + 4\sqrt{L^f \sigma_k + \hat{\epsilon}_{k,i(k)}}, \]
\[ T_{k+1,1} := \{ t \in T_{k+1} : g(u^{k+1,0}, t, \gamma_{k+1}) + \tilde{L}^g \left( \|u^{k+1,0} - u^{k,i(k)}\| + \tilde{\epsilon}_{k+1,0} \right) \geq 0 \}, \]
set \( k := k + 1, \ i := 1. \)

S2: Compute \( u^{k,i} \) such that \( \|\nabla F_{k,i}(u^{k,i})\| \leq \epsilon_k. \)

S3: If \( \|u^{k,i} - u^{k,i-1}\| > \delta_k \), set \( i(k) := i + 1 \) and go to S1, otherwise set \( i(k) := i \) and go to S1.

As we will see in the next section, only in the convergence proof it is required that \( u^{1,0} \) is a feasible point of \( P(\gamma_0) \). However, for the efficiency of the method it is essential that \( u^{1,0} \) is a good approximation of the solution \( u^*(\gamma_0) \).

Algorithms 6.2.12 - 6.2.17 and the results of convergence described in the next section can be easily generalized for problems having a finite number of semi-infinite constraints.

6.3 Convergence of Regularized Penalty Methods for SIP

6.3.1 Basic theorems on convergence

In principle, in case the functions \( J \) and \( g \) do not depend on the parameter \( \gamma \), Algorithm 6.2.13 can be considered as a particular version of Algorithm 6.2.17. However, the iterates of both algorithms coincide only under a special
6.3. CONVERGENCE OF REGULARIZED PENALTY METHODS FOR SIP

coordination of the Lipschitz constants and the controlling parameters and such an artificially forced coordination makes the iteration procedure more expensive in general.

Investigations of the convergence of Algorithm 6.2.17, carried out below, enables us also to establish convergence of Algorithm 6.2.13 by means of simple modifications. For the study of Algorithm 6.2.17 we need the following auxiliary statements.

Suppose Assumption 6.2.15 is fulfilled and the radius \( r \geq r_0 \) is fixed. Denote

\[
\bar{J}(u) := J(u, \bar{\gamma}), \quad J_k(u) := J(u, \gamma_k),
\]

and

\[
Q := \{ u \in B_r : \sup_{t \in T} g(u, t, \bar{\gamma}) \leq 0 \}, \\
Q_k := \{ u \in B_r : \sup_{t \in T_k} g(u, t, \gamma_k) \leq 0 \}, \quad k = 1, 2, \ldots.
\]

The inequalities (6.2.39) and Lemma 6.2.10 ensure that

\[
\rho_H(Q, Q_k) \leq \bar{q}(r)(h_k + \sigma_k)
\]

and if we have a monoton decreasing \( h_{k+1} \leq h_k, \sigma_{k+1} \leq \sigma_k \), then

\[
\rho_H(Q_{k+1}, Q_k) \leq 2\bar{q}(r)(h_k + \sigma_k).
\]

6.3.1 Lemma. Suppose that Assumption 6.2.15 is satisfied and

\( h_{k+1} \leq h_k, \sigma_{k+1} \leq \sigma_k \), for a given \( k \).

Moreover, assume that for fixed \( v^0 \in \mathbb{R}^n, \beta_k > 0 \) the points \( \bar{v}^1, v^1, \bar{v}^2, \bar{v}^2 \) are chosen such that

\[
\bar{v}^1 := \text{Prox}_{J_k, Q_k} v^0, \quad \| v^1 - \bar{v}^1 \| \leq \beta_k, \quad v^1 \in B_r, \\
\bar{v}^2 := \text{Prox}_{J_{k+1}, Q_k} v^1, \\
\bar{v}^2 := \text{Prox}_{J_{k+1}, Q_{k+1}} v^1.
\]

Then the following estimates hold true:

\[
\| \bar{v}^2 - v^1 \| \leq \| v^1 - v^0 \| + 2\sqrt{L^j \sigma_k + \beta_k},
\]

\[
\| \bar{v}^2 - v^1 \| < \| v^1 - v^0 \| + \beta_k^*,
\]

with

\[
\beta_k^* := \sqrt{2(L(r) + 4r)\bar{q}(r)(h_k + \sigma_k) + 2\bar{q}(r)(h_k + \sigma_k) + 4L^j \sigma_k + \beta_k}.
\]

Proof: Let \( w^1 := \text{Prox}_{J_{k+1}, Q_k} v^0 \). In view of the non-expansivity of the prox-mapping we get \( \| \bar{v}^2 - w^1 \| \leq \| v^1 - v^0 \| \). With regard to \( \sigma_{k+1} \leq \sigma_k \) and Assumption 6.2.15(v),

\[
-2L^j \sigma_k + J_k(w^1) + \| w^1 - v^0 \|^2 \leq J_{k+1}(w^1) + \| w^1 - v^0 \|^2 \\
\leq J_{k+1}(\bar{v}^1) + \| \bar{v}^1 - v^0 \|^2 \\
\leq J_{k}(\bar{v}^1) + \| \bar{v}^1 - v^0 \|^2 + 2L^j \sigma_k,
\]
Due to (6.3.4), (6.3.5) and (6.3.8) we get finally
\[ J_k(w^1) + \|w^1 - v^0\|^2 \leq J_k(\tilde{v}^1) + \|\tilde{v}^1 - v^0\|^2 + 4L'_r\sigma_k. \]
Due to the strong convexity of \( J_k(\cdot) + \cdot - v^0 \) and the choice of \( \tilde{v}^1 \) it follows immediately
\[ \|w^1 - \tilde{v}^1\| \leq 2\sqrt{L'_r\sigma_k}. \]
Consequently, using the triangle inequality
\[ \|\tilde{v}^2 - v^1\| \leq \|\tilde{v}^2 - w^1\| + \|w^1 - v^1\| + \|v^1 - \tilde{v}^1\|, \]
we obtain estimate (6.3.5).

Denote \( \Psi_{k+1}(u) := J_{k+1}(u) + \|u - v^1\|^2 \). Now, if \( \Psi_{k+1}(\tilde{v}^2) \geq \Psi_{k+1}(\tilde{v}^1) \),
then
\[ \Psi_{k+1}(w^0) \geq \Psi_{k+1}(\tilde{v}^2) \geq \Psi_{k+1}(\tilde{v}^1), \quad (6.3.7) \]
with
\[ w^0 := \arg\min_{u \in Q_k} \|v - \tilde{v}^2\|. \]
Furthermore denote \( \Psi(u) := J(u) + \|u - v^1\|^2 \) and \( c_0 := 2(L(r) + 4r)\tilde{q}(r) \).
Because of \( \sigma_{k+1} \leq \sigma_k \) as well as Assumption 6.2.15(iv),(v) and (6.3.4) we conclude that
\[ \Psi_{k+1}(v^0) - \Psi_{k+1}(\tilde{v}^2) \leq c_0(h_k + \sigma_k) + 2L'_r\sigma_k, \]
This inequality together with (6.3.7) leads to
\[ \Psi_{k+1}(v^0) - \Psi_{k+1}(\tilde{v}^2) \leq \|w^0 - \tilde{v}^2\| \leq \sqrt{c_0(h_k + \sigma_k) + 2L'_r\sigma_k}. \]

Due to (6.3.4), (6.3.5) and (6.3.8) we get finally
\[ \|\tilde{v}^2 - v^1\| \leq \|\tilde{v}^2 - w^0\| + \|w^0 - \tilde{v}^2\| + \|\tilde{v}^2 - v^1\| \]
\[ \leq 2\tilde{q}(r)(h_k + \sigma_k) + \sqrt{c_0(h_k + \sigma_k) + 2L'_r\sigma_k} + 2L'_r\sigma_k \]
\[ + \|v^1 - v^0\| \]
\[ \leq 2\tilde{q}(r)(h_k + \sigma_k) + \sqrt{c_0(h_k + \sigma_k) + 4L'_r\sigma_k} + \beta_k + \|v^1 - v^0\|, \]
i.e., estimate (6.3.6) is satisfied.
But if the reverse inequality \( \Psi_{k+1}(\tilde{v}^2) < \Psi_{k+1}(\tilde{v}^1) \) holds, then we get
\[ \Psi_{k+1}(\tilde{v}^2) < \Psi_{k+1}(\tilde{v}^1) \leq \Psi_{k+1}(\tilde{w}^0), \]
with \( \tilde{w}^0 := \arg\min_{u \in Q_{k+1}} \|u - \tilde{v}^2\| \), and acting as in the previous case, again we obtain (6.3.6). \( \square \)

Now we consider the following functions
\[ F_{k,i}(u) := J_k(u) + \frac{d}{2} \sum_{t \in T_k,i} (g(u, t, \gamma_k) + |g(u, t, \gamma_k)|) + \|u - u_{k,i}^{-1}\|^2, \]
with \( d \) the same penalty parameter as in (6.2.44), and denote by
\[
\bar{u}^{k,i} := \arg \min_{u \in \mathbb{R}^n} \bar{F}_{k,i}(u).
\]
If \( u^{k,i-1} \in \mathbb{B}_{\tau} \) and Assumption 6.2.15 is fulfilled, then due to the Remarks 6.2.16 and A3.4.44, the identity
\[
\bar{u}^{k,i} = \arg \min_{u \in K_{k,i}} \Psi_{k,i}
\]
holds with
\[
\Psi_{k,i}(u) := J_k(u) + \|u - u^{k,i-1}\|^2.
\]
Moreover, for each \( u \in \mathbb{R}^n \) and pair of indices \((k, i)\) we get
\[
0 \leq F_{k,i}(u) - \bar{F}_{k,i}(u) \leq d \sum_{t \in T_{k,i}} \left( \sqrt{g^2(u, t, \gamma_k) + \tau_k} + |g(u, t, \gamma_k)| \right)
\]
\[
= \frac{d}{2} \sum_{t \in T_{k,i}} \left( \sqrt{\sqrt{\gamma_k^2(u, t, \gamma_k)} + \tau_k} + |g(u, t, \gamma_k)| \right)
\]
\[
\leq \frac{d}{2} \sqrt{\tau_k |T_{k,i}|}.
\]
In view of the choice of \( u^{k,i} \) and the relations (6.3.9) and (6.3.11) it holds
\[
\|u^{k,i} - \bar{u}^{k,i}\| < \epsilon_{k,i}
\]
if \( u^{k,i} \in \mathbb{B}_{\tau} \). Indeed, in view of (6.3.9) we have \( \bar{u}^{k,i} \in K_{k,i} \), consequently,
\[
\bar{F}_{k,i}(\bar{u}^{k,i}) = \Psi_{k,i}(\bar{u}^{k,i}).
\]
For \( w^{k,i} := \arg \min_{u \in \mathbb{R}^n} F_{k,i}(u) \), using (6.3.11) and the obvious relations
\[
\bar{F}_{k,i}(\bar{u}^{k,i}) \leq \bar{F}_{k,i}(w^{k,i}) < F_{k,i}(w^{k,i}),
\]
we obtain
\[
F_{k,i}(\bar{u}^{k,i}) - F_{k,i}(w^{k,i}) \leq \bar{F}_{k,i}(w^{k,i}) - F_{k,i}(w^{k,i}) + \frac{d}{2} \sqrt{\tau_k |T_{k,i}|}
\]
\[
< \frac{d}{2} \sqrt{\tau_k |T_{k,i}|}.
\]
Taking into account the strong convexity of \( F_{k,i}(\cdot) \) and the definition of \( w^{k,i} \), the latter inequality leads to
\[
\|u^{k,i} - w^{k,i}\| < \sqrt{\tau_k} \sqrt{\frac{d}{2} |T_{k,i}|},
\]
and (A1.5.30) implies
\[
\|u^{k,i} - w^{k,i}\| \leq \frac{\epsilon_k}{2},
\]
proving (6.3.12).

Now, we are able to formulate the main result on convergence of Algorithm 6.2.17.
6.3.2 Theorem. Let Assumption 6.2.15 and the relations (6.2.38), (6.2.39) be satisfied and suppose that
\[ h_{k+1} \leq h_k, \quad \sigma_{k+1} \leq \sigma_k, \quad \text{for each } k. \]
Moreover, assume that
\[
\frac{1}{4r} \left( 2L(r)q'(r)(h_k + \sigma_k) + 2L\sigma_k - (\delta_k - \tilde{\epsilon}_k)^2 \right) + \tilde{\epsilon}_k < 0, \quad (6.3.13)
\]
and
\[
\sum_{k=1}^{\infty} \left( \sqrt{2L(r)q'(r)(h_k + \sigma_k)} + 2L(\sigma_k^2) + 2q'(r)(h_k + \sigma_k) + \tilde{\epsilon}_k \right) < \frac{r}{2} \quad (6.3.14)
\]
are fulfilled with \( \tilde{\epsilon}_k := \frac{\sqrt{\tau_k}}{\sqrt{d^2 |T_{k,i}|}} + \hat{\epsilon}_k \).

Then it holds
\[
(i) \quad i(k) < \infty \quad \forall \ k,
\]
\[
(ii) \quad \|u^{k,i}\| < r \quad \forall \ (k, i),
\]
\[
(iii) \quad u^{k,i} \to u^* \in U^*(\bar{\gamma}).
\]

Proof: Suppose that \( k_0 \) and \( i_0 \) are fixed with \( 1 \leq i_0 \leq i(k_0) \) and that
(1) \( i(k) < \infty \) for \( k < k_0 \);
(2) the relations
\[
\hat{u}^{k,i} := \text{Prox}_{J_kQ_k} u^{k,i-1}, \quad \|u^{k,i}\| < r, \quad \|\hat{u}^{k,i}\| < r
\]
are true for the pairs of indices
\[
(k, i) \in \Theta_0 := \{(k', i') : k' < k_0, 1 \leq i' \leq i(k') \land k' = k_0, 1 \leq i' < i_0 \}.
\]
In order to avoid repetitions, conditions for the starting position for \( \hat{u}^{1,1} \), \( u^{1,1} \) and \( i(1) \) will be verified simultaneously within the main part of the proof by induction.
Obviously \( \hat{\epsilon}_{k,i} \leq \hat{\epsilon}_k \) holds for all \( (k, i) \) and (6.3.13) and (6.3.14) imply
\[
\delta_k - \hat{\epsilon}_k > \sqrt{r\tilde{\epsilon}_k} > 0.
\]
Now, let \( u^* \) be an arbitrary point of the set \( U^*(\bar{\gamma}) \cap B_r \) and
\[
\hat{u}^{k,i} := \arg\min_{u \in Q_k} \Psi_{k,i}(u) \quad \text{for } 0 < i \leq i(k), k = 1, 2, ...
\]
(formally, at the moment we are not convinced of \( i(k-1) < \infty \)).
Due to inequality (6.3.3), a point \( \hat{u}^{k,i} \in Q \) can be chosen such that
\[
\|\hat{u}^{k,i} - \hat{u}^{k,i}\| \leq \hat{q}(r)(h_k + \sigma_k), \quad \forall (k, i) \in \Theta_0. \quad (6.3.15)
\]
From Assumption 6.2.15(iv) we conclude that
\[
\bar{J}(\hat{u}^{k,i}) - \bar{J}(\hat{u}^{k,i}) \leq L(r)\hat{q}(r)(h_k + \sigma_k)
\]
6.3. CONVERGENCE OF REGULARIZED PENALTY METHODS FOR SIP

and with regard to \( J(u^*) \leq J(\hat{u}^{k,i}) \),

\[
J(u^*) - J(\hat{u}^{k,i}) \leq L(r)\bar{q}(r)(h_k + \sigma_k). \tag{6.3.16}
\]

Let

\[
v^k := \arg \min_{v \in Q_k} \|v - u^*\|. \tag{6.3.17}
\]

Then

\[
J(v^k) - J(u^*) \leq L(r)\bar{q}(r)(h_k + \sigma_k)
\]

holds, hence,

\[
J(v^k) - J(\hat{u}^{k,i}) \leq 2L(r)\bar{q}(r)(h_k + \sigma_k)
\]

and

\[
J_k(v^k) - J_k(\hat{u}^{k,i}) \leq 2L(r)\bar{q}(r)(h_k + \sigma_k) + 2L_v J_{\sigma_k} := \eta_k.
\]

Now, exploiting Proposition 3.1.3 with \( C := Q_k, f := J_k, v^0 := u^{k,i-1}, y := v^k \), we obtain

\[
\|\hat{u}^{k,i} - v^k\| \leq \|u^{k,i-1} - v^k\| + \sqrt{\eta_k}, \tag{6.3.18}
\]

therefore,

\[
\|\hat{u}^{k,i} - u^*\| \leq \|u^{k,i-1} - u^*\| + \sqrt{\eta_k} + 2\bar{q}(r)(h_k + \sigma_k), \forall (k, i) \in \Theta_0. \tag{6.3.19}
\]

Further, in view of assumption (2) and (6.3.12), the points \( \bar{u}^{k,i} \) and \( \hat{u}^{k,i} \) coincides and

\[
\|\hat{u}^{k,i} - u^{k,i-1}\| > \|u^{k,i} - u^{k,i-1}\| - \hat{\epsilon}_{k,i} \geq \|u^{k,i} - u^{k,i-1}\| - \bar{\epsilon}_k, \forall (k, i) \in \Theta_0.
\]

From (6.3.13) it is obvious that \( \eta_k < (\delta_k - \bar{\epsilon}_k)^2 \), hence, Proposition 3.1.3 applied with the same data leads to

\[
\|\hat{u}^{k,i} - v^k\| < \|u^{k,i-1} - v^k\| + \frac{1}{4r} (\eta_k - (\delta_k - \bar{\epsilon}_k)^2), (k, i) \in \Theta_0, 1 \leq i < i(k)
\]

and if \( k < k_0 \)

\[
\|\hat{u}^{i(k)} - v^k\| \leq \|u^{i(k)-1} - v^k\| + \sqrt{\eta_k}.
\]

Consequently,

\[
\|u^{k,i} - v^k\| \leq \|u^{k,i-1} - v^k\| + \frac{1}{4r} (\eta_k - (\delta_k - \bar{\epsilon}_k)^2) + \bar{\epsilon}_k
\]

\[
< \|u^{k,i-1} - v^k\|, \forall (k, i) \in \Theta_0, i < i(k), \tag{6.3.20}
\]

but in case \( k < k_0, i = i(k) \) the estimate

\[
\|u^{i(k)} - v^k\| \leq \|u^{i(k)-1} - v^k\| + \sqrt{\eta_k} + \bar{\epsilon}_k \tag{6.3.21}
\]

is satisfied.

Summing up the inequalities (6.3.20) and (6.3.21) for fixed \( k < k_0 \) with respect to \( i = 1, ..., i(k) \), we obtain

\[
\|u^{k+1,0} - \hat{v}^k\| = \|u^{k,i(k)} - \hat{v}^k\| \leq \|u^{k,0} - v^k\| + \sqrt{\eta_k} + \bar{\epsilon}_k.
\]
Hence, if $k \leq k_0$, the relation
\[ \|u^{k, 0} - u^*\| \leq \|u^{1, 0} - u^*\| + \sum_{s=1}^{k-1} \omega_s \]  
(6.3.22)
is fulfilled with $\omega_s := \sqrt{\eta_k} + \bar{\epsilon}_s + 2\bar{q}(r)(h_s + \sigma_s)$. Using (6.3.20) with $k = k_0$, we conclude that
\[ \|u^{k_0,i_0 - 1} - v^{k_0}\| \leq \|u^{k_0,0} - v^{k_0}\| \]
and this together with (6.3.18) leads to
\[ \|\tilde{u}^{k_0, i_0} - v^{k_0}\| \leq \|u^{k_0, 0} - v^{k_0}\| + \sqrt{\eta_{k_0}}, \]
thus
\[ \|\tilde{u}^{k_0, i_0} - u^*\| \leq \|u^{k_0, 0} - u^*\| + \sqrt{\eta_{k_0}} + 2\bar{q}(r)(h_{k_0} + \sigma_{k_0}). \]  
(6.3.23)
Now, (6.3.22) and (6.3.23) provide
\[ \|\tilde{u}^{k_0, i_0} - u^*\| \leq \|u^{1, 0} - u^*\| + \sum_{s=1}^{k_0-1} \omega_s + \sqrt{\eta_{k_0}} + 2\bar{q}(r)(h_{k_0} + \sigma_{k_0}). \]  
(6.3.24)
We recall that $u^*$ is an arbitrarily chosen point in $U^\star(\bar{\gamma}) \cap \mathbb{B}_r$. In view of (6.2.38), (6.3.14) and (6.3.24) and taking for the moment $u^* \in U^\star(\bar{\gamma}) \cap \mathbb{B}_r / \mathbb{S}$, we obtain $\|\tilde{u}^{k_0, i_0}\| < r - \bar{\epsilon}_{k_0}$.
In the sequel the cases $i_0 \geq 2$ and $i_0 = 1$ will be investigated separately. If $i_0 \geq 2$, the first statement of Lemma 6.3.1, used with the data
\[ k := k_0, \quad v^0 := u^{k_0, i_0 - 2}, \quad \bar{v}^1 := \tilde{u}^{k_0, i_0 - 1}, \]
\[ v^1 := u^{k_0, i_0 - 1}, \quad \beta_{k_0} := \bar{\epsilon}_{k_0, i_0 - 1}, \quad \bar{v}^2 := \tilde{u}^{k_0, i_0}, \]
implies
\[ \|u^{k_0, i_0} - u^{k_0, i_0 - 1}\| \leq \|u^{k_0, i_0 - 1} - u^{k_0, i_0 - 2}\| + 2\sqrt{L_t^2 \sigma_{k_0} + \bar{\epsilon}_{k_0, i_0 - 1}}. \]
With regard of Assumption 6.2.15(vii) it holds
\[ g(\tilde{u}^{k_0, i_0}, t, \gamma_{k_0}) \leq g(u^{k_0, i_0 - 1}, t, \gamma_{k_0}) \]
\[ + L_t g \left( \|u^{k_0, i_0 - 1} - u^{k_0, i_0 - 2}\| + 2\sqrt{L_t^2 \sigma_{k_0} + \bar{\epsilon}_{k_0, i_0 - 1}} \right), \quad \forall \ t \in T, \]
consequently, from the definition of $T_{k, t}$ in Algorithm 6.2.17 we get
\[ g(\tilde{u}^{k_0, i_0}, t, \gamma_{k_0}) < 0 \quad \forall \ t \in T_{k_0} \setminus T_{k_0, i_0}. \]
This implies that
\[ \tilde{u}^{k_0, i_0} = \tilde{u}^{k_0, i_0}, \quad \|\tilde{u}^{k_0, i_0}\| < r - \bar{\epsilon}_{k_0} \]
and (6.3.12) yields $\|u^{k_0, i_0}\| < r$.
The case $i = 1$ can be analyzed analogously by means of the inequality (6.3.6) if we take in Lemma 6.3.1
\[ k := k_0 - 1, \quad v^0 := u^{k_0 - 1, i(k_0 - 1) - 1}, \quad \bar{v}^1 := \tilde{u}^{k_0 - 1, i(k_0 - 1)}, \]
\[ v^1 := u^{k_0 - 1, i(k_0 - 1)}, \quad \beta_{k_0 - 1} := \bar{\epsilon}_{k_0, 0}, \quad \bar{v}^2 := \tilde{u}^{k_0, 1}. \]
6.3. CONVERGENCE OF REGULARIZED PENALTY METHODS FOR SIP

Thus, condition (2) can be continued for \( k := k_0 \) and each \( i \).

Now it is not difficult to verify that \( i(k_0) < \infty \). Indeed, using (6.3.20), we obtain for \( k = k_0 \) and \( i < i(k_0) \)

\[
\| u^{k,0} - v^{k,0} \| \leq \| u^{k,0} - v^{k,0} \| + i \left( \frac{1}{4r} \left( \eta_{k_0} - (\delta_{k_0} - \bar{\epsilon}_{k_0})^2 \right) + \bar{\epsilon}_{k_0} \right),
\]

hence

\[
i(k_0) \leq \frac{\| u^{k,0} - v^{k,0} \|}{\frac{1}{4r} \left( \eta_{k_0} - (\delta_{k_0} - \bar{\epsilon}_{k_0})^2 \right) + \bar{\epsilon}_{k_0}} + 1.
\]

In order to complete the induction, we note that due to (6.3.14) the estimate \( \| \hat{u}_1 \| < r - \bar{\epsilon}_1 \) is satisfied for the starting position and the inequality (6.3.19) with \( k = 1, i = 1 \), too. Consequently, the identity

\[
\hat{u}_1 = \arg \min_{u \in K_1} \{ J_1(u) + \| u - u^{1,0} \|^2 \}
\]

holds and, because of (6.3.9) and \( K_{1,1} = K_1 \), this ensures \( \hat{u}_1 = \bar{u}_1 \). Finally, from \( \| u^1 - \hat{u}_1 \| \leq \bar{\epsilon}_{1,1} \) we conclude that \( \| u^1 \| < r \).

Therefore, \( \| u^1 \| < r \) holds for all \( i \), and finiteness of \( i(1) \) can be established as in the general scheme of the MSR-method in Section 4.3.1.

Now, let us prove convergence of the sequence \( \{ u^k \} \) (cf. also the proof of Theorem 4.3.6).

For \( u^k_{Q,i} := \arg \min_{u \in Q} \Psi_{k,i}(u) \) with

\[
\Psi_{k,i}(u) := J(u) + \| u - u^{k,i-1} \|^2,
\]

and \( J(\cdot) \) given in (6.3.1), the inequality

\[
\bar{J}(u_{Q,i}^{k,i+1}) - \bar{J}(u^*) \leq 2(\| u_Q^{k,i+1} - u^{k,i} - u_{Q}^{k,i+1} \|)
\]

follows from Proposition A1.5.34, hence,

\[
\| u^{k,i} - u^* \|^2 - ||u_Q^{k,i+1} - u^*||^2 \geq \bar{J}(u_{Q,i}^{k,i+1}) - \bar{J}(u^*) + \| u_{Q,i}^{k,i+1} - u_{Q,i}^{k,i+1} \|^2. \quad (6.3.25)
\]

But, due to Assumption 6.2.15(iv),(v) and estimate (6.3.3), the inequalities

\[
\min_{u \in Q_k} \Psi_{k,i+1}(u) - \min_{u \in Q} \Psi(u) \leq (L(r) + 4r)\bar{q}(r)(h_k + \sigma_k)
\]

and (cf. (6.3.10))

\[
\Psi_{k,i+1}(\tilde{u}_{Q,i+1}^{k,i+1}) - \Psi_{k,i+1}(u_Q^{k,i+1}) \leq (L(r) + 4r)\bar{q}(r)(h_k + \sigma_k) + L_{r}^I \sigma_k \quad (6.3.26)
\]

hold. For \( \hat{u}_{Q,i+1} := \arg \min_{v \in Q} \| v - \tilde{u}_{Q,i+1} \| \) we have

\[
|\Psi_{k,i+1}(\hat{u}_{Q,i+1}^{k,i+1}) - \Psi_{k,i+1}(u_Q^{k,i+1})| \leq |\Psi_{k,i+1}(\hat{u}_{Q,i+1}^{k,i+1}) - \Psi_{k,i+1}(\tilde{u}_{Q,i+1}^{k,i+1})| + L_{r}^I \sigma_k,
\]

therefore,

\[
\bar{\Psi}_{k,i+1}(\hat{u}_{Q,i+1}^{k,i+1}) - \bar{\Psi}_{k,i+1}(u_Q^{k,i+1}) \leq (L(r) + 4r)\bar{q}(r)(h_k + \sigma_k) + L_{r}^I \sigma_k.
\]

The latter inequality together with (6.3.26) leads to

\[
\bar{\Psi}_{k,i+1}(\hat{u}_{Q,i+1}^{k,i+1}) - \bar{\Psi}_{k,i+1}(u_Q^{k,i+1}) \leq 2(L(r) + 4r)\bar{q}(r)(h_k + \sigma_k) + 2L_{r}^I \sigma_k.
\]
Strong convexity of $\Psi_{k,i}(\cdot)$ implies
\[
\|\bar{u}^{k,i+1} - u^{k,i+1}_Q\| \leq \sqrt{2(L(r) + 4r)\bar{q}(r)(h_k + \sigma_k) + 2L^f r \sigma_k} \tag{6.3.27}
\]
and
\[
\|u^{k,i+1} - u^{k,i+1}_Q\| \leq \sqrt{2(L(r) + 4r)\bar{q}(r)(h_k + \sigma_k) + 2L^f r \sigma_k}
+ \bar{q}(r)(h_k + \sigma_k) + \bar{\epsilon}_k := \theta_k. \tag{6.3.28}
\]
Because $\|u^{k,i}\| < r$ for all $k$ and $i \leq i(k)$, we obtain
\[
\|u^{k,i+1} - u^*\|^2 - \|u^{k,i+1}_Q - u^*\|^2 \leq 4r\theta_k,
\]
and together with (6.3.25),
\[
\|u^{k,i} - u^*\|^2 - \|u^{k,i+1} - u^*\|^2 \geq \|u^{k,i+1}_Q - u^{k,i}\|^2 + \bar{J}(u^{k,i+1}_Q) - \bar{J}(u^*) - 4r\theta_k.
\]
In particular, for $i = i(k) - 1$ the inequality
\[
\|u^{k,i(k)-1} - u^*\|^2 - \|u^{k+1,0} - u^*\|^2 \geq \|u^{k+1,0}_Q - u^{k,i(k)-1}\|^2 + \bar{J}(u^{k+1,0}_Q) - \bar{J}(u^*) - 4r\theta_k \tag{6.3.29}
\]
is satisfied.

But, for $0 < i < i(k)$, estimate (6.3.20) and the choice of $v^k$ in (6.3.17) imply
\[
\|u^{k,i} - u^k\| < \|u^{k,0} - v^k\|.
\]
Thus, including the case $i(k) = 1$, inequality
\[
\|u^{k,i(k)-1} - u^*\| < \|u^{k,0} - u^*\| + 2\bar{q}(r)(h_k + \sigma_k)
\]
and the facts that $\|u^{k,0}\| < r$ and $\|u^*\| \leq r$, we infer that
\[
\|u^{k,i(k)-1} - u^*\|^2 < \|u^{k,0} - u^*\|^2 + 8r\bar{q}(r)(h_k + \sigma_k) + 4(\bar{q}(r)(h_k + \sigma_k))^2.
\]

The latter inequality together with (6.3.29) leads to
\[
\|u^{k+1,0} - u^*\|^2 < \|u^{k,0} - u^*\|^2 - \|u^{k,i(k)-1} - u^{k+1,0}_Q\|^2 - \left(\bar{J}(u^{k+1,0}_Q) - \bar{J}(u^*)\right)
+ 4r\theta_k + 8r\bar{q}(r)(h_k + \sigma_k) + 4(\bar{q}(r)(h_k + \sigma_k))^2
\leq \|u^{k,0} - u^*\|^2 + 4r(\theta_k + 2\bar{q}(r)(h_k + \sigma_k)) + 4(\bar{q}(r)(h_k + \sigma_k))^2.
\]

From (6.3.14) we have $\sum_{k=1}^\infty \theta_k < \infty$, consequently,
\[
\sum_{k=1}^\infty h_k < \infty \quad \text{and} \quad \sum_{k=1}^\infty \sigma_k < \infty.
\]

Due to Lemma A3.1.4, $\{\|u^{k,0} - u^*\|\}$ converges for each $u^* \in U^*(\bar{q}) \cap B_r$. Hence, $\bar{J}(u^{k,0}_Q) \to \bar{J}(u^*)$, and with regard to (6.3.28)
\[
\lim_{k \to \infty} \bar{J}(u^{k,0}) = \bar{J}(u^*).
\]
Now, for a fixed cluster point \( u^{**} \) of the sequence \( \{u^{k,0}\} \), we conclude that
\[
\lim_{k \to \infty} \|u^{k,0} - u^{**}\| = 0.
\]
Because
\[
\|u^{k,i} - u^*\| < \|u^{k,0} - u^*\| + 2\bar{q}(r)(h_k + \sigma_k)
\]
holds for any \( u^* \in U^*(\bar{\gamma}) \cap \mathbb{B}_r \), the relation
\[
\lim_{k \to \infty} \max_{0 \leq i \leq i(k)} \|u^{k,i} - u^{**}\| = 0
\]
is also true.

\[\Box\]

6.3.3 Remark. The controlling parameters \( h_k, \sigma_k, \epsilon_k, \delta_k \) and \( \tau_k \) are mainly defined by the conditions (6.3.13), (6.3.14), which are not restrictive and enable us to change the parameters slowly, for instance like
\[
h_k \approx \sigma_k \approx \frac{1}{(c_1 + k)^{2+c_2}}, \quad \sqrt{\tau_k} \approx \frac{1}{|T_k| (c_1 + k)^{2+c_2}}, \quad \epsilon_k \approx \frac{1}{(c_1 + k)^{1+c_2}},
\]
with fixed \( c_1 > 0 \) and arbitrary small \( c_2 > 0 \). In practice the corresponding parameters in optimization and discretization methods decrease usually with geometrical progression. However, if the value
\[
\max[L(r)\bar{q}(r), L^2_r, L^g_r, \bar{L}^g_r]
\]
is large, then it is dangerous of \( c_1 \) has to be chosen large in order to satisfy (6.3.14). In this connection it is important to note that if the set
\[
K := \{ u \in \mathbb{R}^n : g(u, t, \bar{\gamma}) \leq 0 \quad \forall \ t \in T \}
\]
is bounded then we are able to make use of Lemma 4.3.3 and Theorem 4.3.6 instead of Lemma 4.3.5 and Theorem 4.3.8. Choosing \( r \) such that \( K \subset \mathbb{B}_{r/2} \), requirement (6.3.14) can be replaced by
\[
\bar{q}(r)(h_k + \sigma_k) + \bar{\epsilon}_k < \frac{r}{2},
\]
\[
\sum_{k=1}^{\infty} \sqrt{h_k} < \infty, \quad \sum_{k=1}^{\infty} \sqrt{\sigma_k} < \infty, \quad \sum_{k=1}^{\infty} \bar{\epsilon}_k < \infty.
\]
These conditions are weaker for the choice of \( c_1 \) (see also Remark 4.2.6).  \[\Box\]

On the basis of the given proof of convergence for Algorithm 6.2.17 we are going now to analyze convergence of the Algorithms 6.2.12 and 6.2.13.

6.3.4 Theorem. Suppose that radius \( r \) is chosen such that the relations (6.2.31) and Lipschitz condition (6.2.26) are fulfilled. Moreover, assume that
\[
h_{k+1} \leq h_k \ \forall \ k; \quad h_1 < \frac{\alpha}{L_t},
\]
that the inequalities

\[
\frac{1}{4r} \left( L(r)q(r)h_k - (\delta_k - \tilde{\epsilon}_k)^2 \right) + \tilde{\epsilon}_k < 0, \tag{6.3.30}
\]

\[
\sum_{k=1}^{\infty} \left( \sqrt{L(r)q(r)h_k + \tilde{\epsilon}_k + q(r)h_k} \right) < \frac{r}{2} \tag{6.3.31}
\]

hold and, in addition, that for Algorithm 6.2.13 condition (6.2.35) is true.

Then for both Algorithms 6.2.12 and 6.2.13 it holds:

(i) \( i(k) < \infty \quad \forall \ k, \)

(ii) \( \| u^{k,i} \| < r \quad \forall \ (k,i), \)

(iii) \( u^{k,i} \rightarrow u^* \in U^*. \)

Sketch of Proof: In order to apply Lemma 6.3.1 and the proof of Theorem 6.3.2 to this situation, we use some stronger estimates than those following from the analysis of Algorithm 6.2.17 for \( \sigma_k := 0, J_k := \bar{J} = J \quad \forall \ k. \)

Better estimates can be obtained if we investigate the approximation of Problem 6.2.1 directly. To this end we assume that \( \alpha := a \) in (6.2.29), with \( a \) from Assumption 6.2.15(iii).

First, for both Algorithms 6.2.12 and 6.2.13 the inclusion \( K \subset K_k \quad \forall \ k \) is true, therefore due to Corollary 6.2.11

\[ \rho \eta(Q_k, Q_{k+1}) \leq q(r)h_k, \quad k = 1, 2, \ldots, \]

if \( h_k < \frac{\alpha}{T}. \) Thus estimate (6.3.6) in Lemma 6.3.1 is true with

\[ \beta^*_k = \sqrt{(L(r) + 4r)q(r)h_k} + q(r)h_k + \beta_k. \]

Second, between the constants \( q(r) \) and \( \bar{q}(r) \) the relation \( \bar{q}(r) = 2q(r) \) holds, and formula (6.2.28) provides a more precise upper bound \( c(r) \) for the Lagrange multiplier of the auxiliary problems than formula (6.2.15), namely \( c(r) = \frac{\bar{q}}{2}. \)

In order to exploit the proof of Theorem 6.3.2 to Algorithm 6.2.13, we take \( \sigma_k := 0, J_k := J = J \quad \forall \ k \) and preserve all the other notations. Then, instead of inequality (6.3.16), we obtain

\[ J(u^*) - J(\hat{u}^{k,i}) \leq L(r)q(r)h_k, \]

and because of \( Q \subset Q_k, \) Proposition 3.1.3 can be used with

\[ C := Q_k, \quad f := J, \quad v^0 := u^{k,i-1}, \quad y := u^*. \]

For \( (k,i) \in \Theta_0, \) \( 1 \leq i < i(k) \) this leads to

\[
\| \hat{u}^{k,i} - u^* \| \leq \| u^{k,i-1} - u^* \| + \frac{1}{4r} \left( L(r)q(r)h_k - (\delta_k - \tilde{\epsilon}_k)^2 \right), \tag{6.3.32}
\]

and if \( k < k_0, \)

\[
\| u^{k,i(k)} - u^* \| \leq \| u^{k,i(k)-1} - u^* \| + \sqrt{L(r)q(r)h_k}. \tag{6.3.33}
\]
Moreover, the assumption of induction in the proof of Theorem 6.3.2 gives $\hat{u}^{k,i} = \bar{u}^{k,i}$ for $(k,i) \in \Theta_0$, and due to (6.3.30), (6.3.31),
$$
\|\hat{u}^{k,i} - u^{k,i-1}\| \geq \|u^{k,i} - u^{k,i-1}\| - \tilde{\epsilon}_k.
$$
Hence, taking into account (6.3.32), (6.3.33) and (6.3.13), we get
$$
\|u^{k,i} - u^*\| < \|u^{k,i-1} - u^*\|, \quad \forall (k,i) \in \Theta_0, \ 1 \leq i < i(k)
$$
(6.3.34)
and if $k < k_0$,
$$
\|u^{k,i(k)} - u^*\| < \|u^{k,i(k)-1} - u^*\| + \sqrt{L(r)q(r)h_k} + \tilde{\epsilon}_k.
$$
(6.3.35)
Finally, for $k = k_0$ and $i = i_0$ Proposition 3.1.3 enables us to conclude that
$$
\|\hat{u}^{k_0,i_0} - u^*\| < \|u^{k_0,i_0-1} - u^*\| + \sqrt{L(r)q(r)h_{k_0}}
$$
(6.3.36)
and adding up the inequalities (6.3.34)–(6.3.36) yields
$$
\|\hat{u}^{k_0,i_0} - u^*\| < |u^{1.0} - u^*| + \sum_{s=1}^{k_0} \left( \sqrt{L(r)q(r)h_s} + \tilde{\epsilon}_s \right).
$$
The further analysis requires only trivial modifications in the proof of Theorem 6.3.2.

It should be noted that Algorithm 6.2.12 is not a particular case of the Algorithms 6.2.13 or 6.2.17. Indeed, even if choosing a constant $\bar{L}$ large enough, we are not sure that $T_{k,i} = T_k \forall (k,i)$. Nevertheless the proof idea of Theorem 6.3.4 for the case of Algorithm 6.2.12 is similar, and the complete proof of this theorem for both Algorithms 6.2.12 and 6.2.13 is published in Kaplan and Tichatschke [214].

6.3.2 Efficiency of the deleting rule

The efficiency of the deleting rule applied in the Algorithms 6.2.13 and 6.2.17 is shown in the following statement.

6.3.5 Theorem. Suppose that the hypotheses of Theorem 6.3.4 are fulfilled and for a point $u^*$ with
$$
\lim_{k \to \infty} \sup_{1 \leq i \leq i(k)} \|u^{k,i} - u^*\| = 0
$$
the set
$$
T(u^*) := \{ t \in T : g(u^*, t) = \max_{\tau \in T} g(u^*, \tau) \}
$$
(6.3.37)
consists only of a finite number of elements. Then, for Algorithm 6.2.13 it holds
$$
\sup_{1 \leq i \leq i(k)} \frac{|T_{k,i}|}{|T_k|} \to 0 \quad \text{as } k \to \infty.
$$
CHAPTER 6.  PPR FOR CONVEX SEMI-INFINITE PROBLEMS

Proof: Let \( T(u^*) = \{ t^1, ..., t^q \} \) and \( B_\rho(u^*) := \{ u \in \mathbb{R}^n : \| u - u^* \| \leq \rho \} \) with arbitrarily chosen \( \rho > 0 \). We fix a small \( \delta > 0 \) and define

\[
T_\delta := \bigcup_{\nu=1}^{\rho} \{ t \in T : \| t - t^{\nu} \|_m < \delta \}.
\]

Due to the continuity of \( g(u^*, \cdot) \) on the compact set \( T \setminus T_\delta \), the inequality \( g(u^*, t) < -\alpha(\delta) \) is fulfilled with some \( \alpha(\delta) > 0 \) for all \( t \in T \setminus T_\delta \). In view of the convergence \( \{ u^{k,i} \} \to u^* \) and the uniform continuity of \( g \) on \( B_\rho(u^*) \times (T \setminus T_\delta) \), we obtain for sufficiently large \( k_0(\delta) \) and \( k \geq k_0(\delta) \)

\[
g(u^{k,i}, t) < -\frac{\alpha(\delta)}{2}, \quad \forall \ i = 1, ..., i(k), \ \forall \ t \in T \setminus T_\delta. \tag{6.3.38}
\]

But because of

\[
\lim_{k \to \infty} \sup_{1 \leq i \leq i(k)} \| u^{k,i} - u^* \| = 0
\]

and the conditions on the controlling parameters we conclude that

\[
\lim_{k \to \infty} \sup_{1 \leq i \leq i(k) - 1} L_u \left( \| u^{k,i} - u^{k,i-1} \| + \hat{\epsilon}_{k,i} \right) = 0,
\]

\[
\lim_{k \to \infty} L_u \left( \| u^{k+1,0} - u^{k,i(k)-1} \| + \hat{\epsilon}_{k+1,0} \right) = 0.
\]

These relations together with (6.3.38) imply

\[
g(u^{k,i}, t) + L_u \left( \| u^{k,i} - u^{k,i-1} \| + \hat{\epsilon}_{k,i} \right) < 0, \quad i = 1, ..., i(k),
\]

and

\[
g(u^{k+1,0}, t) + L_u \left( \| u^{k+1,0} - u^{k,i(k)-1} \| + \hat{\epsilon}_{k+1,0} \right) < 0
\]

for all \( t \in T \setminus T_\delta \) and sufficiently large \( k \).

Hence, for such \( k \), all points \( t \in T_k \) corresponding to the remaining constraints are contained in \( T_\delta \). But \( \delta \) is arbitrarily chosen and

\[
\text{meas} T_\delta \to 0 \quad \text{as} \ \delta \to 0.
\]

Reduction theorems (cf. Proposition A1.7.54, for details see also [178]) state the existence of a finite subset \( T \) characterizing the solution of SIP. Finiteness of \( T(u^*) \) for each \( u^+ \in U^* \) can be considered as a generic case for SIP. This was already remarked in Section 6.1.

For Algorithm 6.2.17 in a similar way a result can be established which is analogous to Theorem 6.3.5. In this case, instead of (6.3.37), one has to deal with the set

\[
T(u^*(\gamma)) := \{ t \in T : g(u^*(\gamma), t, \gamma) = \max_{\tau \in T} g(u^*(\gamma), \tau, \gamma) \},
\]

with \( u^*(\gamma) \) such that \( \lim_{k \to \infty} \sup_{1 \leq i \leq i(k)} \| u^{k,i} - u^*(\gamma) \| = 0 \).

Theorem 6.3.5 says that if we know the Lipschitz constant \( L_u \), then Algorithm 6.2.13 is preferable in comparison with Algorithm 6.2.12. However, the determination of such a constant may be connected with serious difficulties.

For parametric SIP an analogue of Algorithm 6.2.12 can be investigated in the same manner. Taking into account Remark 4.3.9, convergence of one-step variants of these methods follow from Theorems 6.3.2 and 6.3.4. In this context the conditions (6.3.13) and (6.3.30) can be omitted.
6.3. CONVERGENCE OF REGULARIZED PENALTY METHODS FOR SIP

6.3.6 Remark. For the sequences \( \{u^{k,i}\} \) and \( \{\hat{u}^{k,i}\} \), generated by the Algorithms 6.2.12, 6.2.13 and 6.2.17, the relations

\[
\hat{u}^{k,i} = \text{Prox}_{Q_k} u^{k,i-1}, \quad \|u^{k,i} - \hat{u}^{k,i}\| \leq \varepsilon_{k,i} \leq \varepsilon_k,
\]

\[
\|\hat{u}^{k,i}\| < r, \quad \|u^{k,i}\| < r
\]

are true. Therefore, all results concerning the rate of convergence established in Section 5.1 can be preserved for these algorithms, if we replace in the Theorems 5.1.1, 5.1.5 and 5.1.7 the triple

\[
\frac{\varepsilon_k}{2}, \quad \mu_k, \quad \sigma_k
\]

by \( \varepsilon_k, q(r)h_k, 0 \) (in the case of Algorithms 6.2.12 and 6.2.13) and by \( \varepsilon_k, q'(r)(h_k + \sigma_k), L'J\sigma_k \) (for Algorithm 6.2.17), respectively.

6.3.7 Theorem. Under the assumptions of Theorem 6.3.2 the following error bounds are true for each iteration \( k \) of Algorithm 6.2.17:

\[
-L(r)\varepsilon_{k,i(k)} - 2L'_J\sigma_k \leq J_k(u^{k,i(k)}) - \min_{u \in Q_k} J_k(u) \leq 4r\delta_k + (4r + L(r))\varepsilon_{k,i(k)} + 2L'_J\sigma_k,
\]

(6.3.39)

\[
-L(r)\left(\bar{q}(r)(h_k + \sigma_k) + \varepsilon_{k,i(k)}\right) \leq \bar{J}(u^{k,i(k)}) - \min_{u \in K} \bar{J}(u) \leq 4r\delta_k + (4r + L(r))\varepsilon_{k,i(k)} + L(r)\bar{q}(r)(h_k + \sigma_k) + 2L'_J\sigma_k,
\]

(6.3.40)

\[
\inf_{v \in Q_k} \|v - u^{k,i(k)}\| \leq \varepsilon_{k,i(k)},
\]

\[
\inf_{v \in K} \|v - u^{k,i(k)}\| \leq \varepsilon_{k,i(k)} + \bar{q}(r)(h_k + \sigma_k).
\]

The first two estimates are corollaries of Theorem 4.3.10 and Remark 4.3.11, whereas the latter two can be obtained immediately due to (6.3.12), \( \hat{u}^{k,i} \in Q_k \) and (6.3.3). Estimate (6.3.39) defines the deviation between the computed values \( J_k(u^{k,i(k)}) \) and the optimal value of the objective function of Problem \( P(\gamma_k) \) if some solution of this problem belongs to the ball \( B_r \).

As we noted at the beginning of Section 6.2, such intermediate bounds are important for applied parametric problems in case an exact value of the parameter \( \gamma \), which is of interest, is unknown beforehand. Indeed, with these bounds a more suitable value of \( \gamma \) can be defined or its bounds can be improved.

The path-following from \( \gamma_0 \) to \( \gamma \) enables us to solve the Problem \( P(\gamma) \) more efficiently. For instance, taking into account the fast convergence of conjugate direction methods and Quasi-Newton methods in a neighborhood of the sought solution, we may expect a better efficiency of these methods as if Problem \( P(\gamma) \) is tackled directly.

If the distance between \( \gamma_0 \) and \( \gamma \) is large, then in order to accelerate the solution procedure, we recommend to choose some intermediate values \( \gamma_1, \ldots, \gamma_k \) and to compute subsequently solutions for the problems \( P(\gamma_1), \ldots, P(\gamma_k) \) with an increasing accuracy. This can be done by means of the described procedure on each subinterval \( [\gamma_0, \gamma_1], \ldots, [\gamma_k, \gamma] \).
6.3.8 Remark. The conditions (6.3.13) and (6.3.14) in Theorem 6.3.2 can be weakened if for each $k$ and $i$ instead of $\delta_k, \epsilon_k$ and $\tilde{\epsilon}_k$ the values

$$\delta_{k,i}, \epsilon_{k,i}, \tilde{\epsilon}_{k,i} := \frac{\epsilon_{k,i}}{2} + \sqrt{\tau_k \frac{d}{2}} |T_{k,i}|$$

are used, respectively. More precisely, condition (6.3.13) has to be replaced by

$$\delta_{k,i} > \tilde{\epsilon}_{k,i}, \quad \frac{1}{4r} (2L(r)\hat{q}(r)(h_k + \sigma_k) + 2L_2^d \sigma_k - (\delta_{k,i} - \tilde{\epsilon}_{k,i})^2) + \tilde{\epsilon}_{k,i} < 0$$

(6.3.41)

for all $k$ and $i$, $1 \leq i \leq i(k)$, and in (6.3.14) the value $\tilde{\epsilon}_{k,i(k)}$ should be used instead of $\tilde{\epsilon}_k$. In order to guarantee that $i(k) < \infty$ it is sufficient, for instance, to ensure that

$$\frac{1}{4r} \left( \eta_k - (\delta_{k,i} - \tilde{\epsilon}_{k,i})^2 \right) + \tilde{\epsilon}_{k,i} < -\frac{\alpha_k}{i},$$

with $\alpha_k > 0$ arbitrarily chosen. Using these modifications the proof of Theorem 6.3.2 is completely analogous. Only in the case $i_0 < i(k_0)$, in order to estimate $\| \tilde{u}^{k_0,i_0} \|$, we have to strengthen the estimate (6.3.23) and (6.3.24). This is possible if, instead of inequality (6.3.18), the stronger one

$$\| \tilde{u}^{k_0,i_0} - u^{k_0} \| < \| u^{k_0} - 1 - u^{k_0} \| - \tilde{\epsilon}_{k_0,i_0}$$

is taken. The latter inequality can be obtained by means of the second statement in Proposition 3.1.3 and the modified relation (6.3.13).

Estimate (6.3.24) turns into

$$\| \tilde{u}^{k_0,i_0} - u^* \| \leq \| u^{1,0} - u^* \| + \sum_{s=1}^{k_{0,1}} (\sqrt{\eta_s} \tilde{\epsilon}_{s,i(s)} + 2\hat{q}(r)(h_s + \sigma_s))$$

$$+ 2\hat{q}(r)(h_{k_0} + \sigma_{k_0}) - \tilde{\epsilon}_{k_0,i_0}$$

and this allows us to conclude that $\| \tilde{u}^{k_0,i_0} \| < r - \tilde{\epsilon}_{k_0,i_0}$.

Of course, this is not an a priori choice of the controlling parameters, however it can be done automatically within the solution process.

Due to the relations

$$\tilde{u}^{k,i} = \hat{u}^{k,i}, \quad \| u^{k,i} \| < r, \quad \| \hat{u}^{k,i} \| < r,$$

established in Theorem 6.3.2, Assumption 6.2.15 can be reformulated for $r$ chosen on account of (6.2.38). In particular, instead of condition (i) in Assumption 6.2.15 it is sufficient to assume that $J(\cdot, \gamma) : \mathbb{R}^n \to \mathbb{R}$, $g(\cdot, t, \gamma) : \mathbb{R}^n \to \mathbb{R}$ are convex differentiable functions on $B_r$.

In order to choose the penalty parameter $d$, condition (6.2.15) can be replaced also by a weaker one, namely

$$d > \frac{2}{a} \left( \tilde{J}(\tilde{u}) + (\| \tilde{u} \| + r)^2 + 2\sigma_0 + r^2 - c_0 \right)$$

with $\sigma_0 := L_2^d \sup_B \sigma_k$, $c_0 := \min_{v \in \mathbb{R}^n} \{ \tilde{J}(v) + \frac{1}{2} \| v \| ^2 \}$ and $a$ defined according to Assumption 6.2.15(iii). Consequently, there is no need to evaluate $L_2^d$ with $\tilde{\rho}$ defined via (6.2.15) and (6.2.13). $\diamondsuit$
6.3. CONVERGENCE OF REGULARIZED PENALTY METHODS FOR SIP

6.3.9 Remark. Now we consider the case where Slater’s condition fails to be hold for Problem (6.2.1). Assume that for a chosen discretization and given radius $r$ it is possible to estimate

$$\rho_H(Q_k, Q) \leq \mu(h_k), \quad k = 1, 2, \ldots,$$

by means of a monotonously increasing function $\mu : \mathbb{R}_+ \to \mathbb{R}_+, \mu(0) = 0$. Moreover, we assume that for each problem

$$\min\{J(u) + \|u - \bar{v}\|^2 : g(u, t) \leq 0, \quad \forall t \in T_k\}$$

with $\bar{v} \in B_r$, a constant $d(k)$ and a Lagrange multiplier vector $\lambda(\bar{v}; k) |_{t \in T_k}$ exists such that

$$\lambda(\bar{v}; k) < d(k), \quad t \in T_k,$$

holds uniformly with respect to $\bar{v}$.

Then Algorithm 6.2.12 is well defined substituting penalty parameter $d(k)$ for $d$ in function (6.2.33). The corresponding statement on convergence is similar to Theorem 6.3.4. It requires only to replace constant $d$ by $d(k)$ in the term $\tilde{\epsilon}_k$ as well as $q(r)h_k$ by $\mu(h_k)$ in the formulas (6.3.30) and (6.3.31).

With analogous substitutions Algorithm 6.2.13 is also practicable if the estimate (6.3.43) can be maintained for problem

$$\min\{J(u) + \|u - \bar{v}\|^2 : g(u, t) \leq 0, \quad \forall t \in T_{k,i}\}$$

with $T_{k,i} \subset T_k$, obtained from $T_k$ by means of the deleting rule in S1. The alterations in the proofs of Lemma 6.3.1 and Theorem 6.3.4 are obvious in both cases.

Considering an example introduced by CHARNES, COOPER AND KORTANEN [72], we illustrate how problems can be solved by means of the described methods, when Slater’s condition fails to hold.

6.3.10 Example. We consider the problem

$$J(u) := -u_1 \to \min$$

s.t. $g(u, t) := u_1t - u_2t^2 \leq 0, \quad \forall t \in T := [0, 1].$ (6.3.44)

Obviously, constraint (6.3.44) does not satisfy the Slater condition. It is easily seen that

$$K := \{u \in \mathbb{R}^2 : u_1 \leq 0, \ u_1 - u_2 \leq 0\},$$

$$U^* := \{u \in \mathbb{R}^2 : u = (0, a)^T : a \geq 0\}.$$ 

Let $r > 0$ and $T_k := \{t_j = \frac{j}{k} : j = 1, \ldots, k\}$, then we get

$$K_k := \{u \in \mathbb{R}^2 : \left(\frac{j}{k}\right)u_1 - \left(\frac{j}{k}\right)^2 u_2 \leq 0, \quad j = 1, \ldots, k\},$$

i.e., $K_k$ can be represented as

$$K_k = \{u \in \mathbb{R}^2 : u_1 - \left(\frac{1}{k}\right)^2 u_2 \leq 0, \ u_1 - u_2 \leq 0\}.$$
For an arbitrary point $\bar{u} \in Q_k \setminus Q$ we choose $u_1 := \min[\bar{u}_1, 0]$, $u_2 := \bar{u}_2$, and obtain
\[
\rho_H(Q_k, Q) \leq rh_k, \quad h_k = \frac{1}{k}.
\]
Now, under this discretization a Slater point $\tilde{u}$ exists. In order to estimate the Lagrange multipliers $\lambda(\bar{v}; k, t)$ of the problems (6.3.42) with $\bar{v} \in B_r$, we observe that
\[
\max_{t \in T_k} \lambda(\bar{v}; k, t) \leq \frac{J(\tilde{u}) + (\|\tilde{u}\| + r)^2 - c_0}{\max_{t \in T_k} g(\tilde{u}, t)}.
\] (6.3.45)
This inequality follows from Lemma 6.2.7 and Remark 6.2.9, if we apply them to Problem 6.2.9 with $f := J$ and $g(\cdot) := \max_{t \in T_k} g(\cdot, t)$.

For $\tilde{u} := (-1, 0)^T$ we obtain
\[
\max_{t \in T_k} \lambda(\bar{v}; k, t) \leq (1 + (1 + r)^2 - c_0) k =: c(r) k.
\] (6.3.46)
Moreover,
\[
u_1 := \frac{1}{2(1 + k^2)}, \quad \nu_2 := \frac{k}{2(1 + k^2)},
\]
is the solution of Problem (6.3.42) with $\bar{v} := 0$ and
\[
\lambda(0; \frac{1}{k}) := \frac{k^3}{1 + k^2}
\]
is the unique non-zero Lagrange multiplier.

Hence, estimate (6.3.46) is exact with respect to the order. Obviously, this estimate can be maintained also if a deleting rule is applied.

Therefore, if the controlling parameters are chosen according to Theorem 6.3.4 with the above mentioned replacements of $q(r)h_k$ by $\frac{r}{k}$ and $d$ by $d(k) := 2c(r)k$, then convergence of the Algorithms 6.2.12 and 6.2.13 is ensured for the problem under consideration.

If the constrained in the example above is replaced by
\[
g(u, t) := u_1 |t - \frac{1}{\sqrt{2}}| - u_2 |t - \frac{1}{\sqrt{2}}|^2 \leq 0, \quad \forall t \in [0, 1]
\]
the same effect with a different function $d(k)$ occurs for the same uniform discretization $T_k$.

### 6.3.3 MSR-approach revisited: Deleting rules for constraints

Now, according to the general scheme of MST-methods and the assumptions of Theorem 4.3.8, the solution of SIP (6.2.1) under the validity of the Lipschitz property (6.2.26) will be sketched. We assume that
\[
h_{k+1} \leq h_k, \quad h_1 \leq \frac{\alpha}{L^t}, \quad \sigma_k \equiv 0
\]
with $\alpha$ given in (6.2.29). Then, due to Lemma 6.3.1, inequality (4.3.9) and Remark 6.2.14, we can conclude immediately that for each $k$
\[
\|u^{k,i} - u^{k,i-1}\| \leq \|u^{k,i-1} - u^{k,i-2}\| + \frac{\epsilon_k^t}{2}, \quad \text{if } 2 \leq i \leq i(k),
\] (6.3.47)
\[
\|u^{k+1,1} - u^{k+1,0}\| \leq \|u^{k+1,0} - u^{k,i(k)-1}\| + \epsilon^t_k,
\] (6.3.48)
with $\epsilon_k' := \sqrt{(L(r) + 4r)q(r)h_k + q(r)h_k + \epsilon_k}.$

Thus, if the function $g$ satisfies the Lipschitz condition (6.2.35), then the inequalities (6.3.47) and (6.3.48) lead to

\[
g(\tilde{u}^{k,i}, t) \leq g(u^{k,i-1}, t) + L_u \left( \|u^{k,i-1} - u^{k,i-2}\| + \frac{\epsilon_k}{2} \right), \quad \text{for } 2 \leq i \leq i(k)
\]

and

\[
g(\tilde{u}^{k+1,1}, t) \leq g(u^{k+1,0}, t) + L_u \left( \|u^{k+1,0} - u^{k,i(k)-1}\| + \epsilon_k \right).
\]

Hence, in the framework of general MSR-methods, instead of the auxiliary problems

\[
\min_{u \in K_k} \Psi_{k,i}(u), \quad i = 2, \ldots, i(k) \quad \text{and} \quad \min_{u \in K_{k+1}} \Psi_{k+1,1}(u)
\]

we solve (for each $k$) the problems

\[
\begin{align*}
\min \{ \Psi_{k,i}(u) : g(u, t) & \leq 0 \quad \forall \ t \in T_{k,i} \}, \quad i = 2, \ldots, i(k), \quad (6.3.49) \\
\min \{ \Psi_{k+1,1}(u) : g(u, t) & \leq 0 \quad \forall \ t \in T_{k+1,1} \}, \quad (6.3.50)
\end{align*}
\]

respectively, with

\[
T_{k,i} := \{ t \in T_k : g(u^{k,i-1}, t) + L_u \left( \|u^{k,i-1} - u^{k,i-2}\| + \frac{\epsilon_k}{2} \right) \geq 0 \} \text{ for } 2 \leq i \leq i(k)
\]

and

\[
T_{k+1,1} := \{ t \in T_{k+1} : g(u^{k+1,0}, t) + L_u \left( \|u^{k+1,0} - u^{k,i(k)-1}\| + \epsilon_k \right) \geq 0 \}.
\]

In accordance with the starting phase of the method, we have to calculate

\[
u^{1,1} \approx \arg \min_{u \in K_1} J(u).
\]

The described transition from $T_k$ to $T_{k,i}$ should be considered as the first stage of the cleaning of a given grid. Thereafter, in order to solve Problem (6.3.49) or (6.3.50), those deleting rules can be applied which are included in the solvers for these problems.

For instance, if a SIP is solved by means of a discretization approach (see Algorithm 6.1.3), then we replace $T_k$ by $T_{k,i}$ at the beginning of Step (ii). In step (iii), the feasibility of the resulting point $\tilde{u}$ has to be verified for Problem $P(T_{k,i})$, but not for $P(T_k)$. Nevertheless, at the end of step $k$ a point $u^k$, which is also feasible for Problem $P(T_k)$, is obtained.

We emphasize that, due to the application of IPR, the objective functions in the auxiliary problems $P(T_{k,i})$ are strongly convex. This is essential for the use of efficient deleting rules.

In the same way it is possible to combine the described cleaning procedure with deleting rules which are part of exchange methods, like cutting plane methods with a deletion of constraints.

### 6.4 Regularized Interior Point Methods

In this subsection we describe mainly results which we have obtained together with Abbé [2]. Starting with the introduction of logarithmic barrier methods and its regularization, the subsequent subsections contain corresponding convergence results and some numerical experiences with MSR-methods for SIP.
6.4.1 Regularized logarithmic barrier methods

In the sequel we want to apply the classical logarithmic barrier method to convex semi-infinite problems of the form

\[
\begin{align*}
\text{minimize} & \quad f(x) \quad \text{s.t.} \quad x \in \mathbb{R}^n, \quad g(x,t) \leq 0 \quad (t \in T).
\end{align*}
\] (6.4.1)

Without specifying any assumptions at this point it turns out that the most difficult question for this transfer is how can we choose a suitable barrier function. This is caused by the appearance of infinitely many constraints.

Sonnevend [377] and Schättler [362] suggest to use the following integral barrier function

\[
- \int_T \ln(-g(x,t)) \, dt.
\] (6.4.2)

Of course, \(\text{meas} T > 0\) is assumed in this case so that especially finite sets \(T\) are excluded. Nevertheless, let us have a closer look at some important details of the arising method.

Due to the smoothing effect of the integral, (6.4.2) does not have to possess the barrier property at all. That means it is possible that (6.4.2) is bounded from above if one approaches the boundary of the feasible region. This fact is illustrated by the following example.

6.4.1 Example. Let \(T := [0, 1]\) and consider the linearly restricted feasible set

\[
K := \left\{ x \in \mathbb{R}^2 : g(x,t) := - \left( t - \frac{1}{2} \right)^2 x_1 - x_2 \leq 0 \quad \forall \ t \in T \right\}.
\]

Choosing \(x = (1, 0)^T\), we have \(g(x,t) = -(t - \frac{1}{2})^2 \leq 0\) for all \(t \in [0, 1]\). Thus \(x \in K\), but \(g(x,\hat{t}) = 0\) for \(\hat{t} = \frac{1}{2}\) implies \(x \notin \text{int}K\). Furthermore, we conclude

\[
- \int_T \ln(-g(x,t)) \, dt = - \int_0^1 \ln \left( \left( t - \frac{1}{2} \right)^2 \right) \, dt = -4 \int_{1/2}^1 \ln \left( t - \frac{1}{2} \right) \, dt = 2 \ln 2 < \infty.
\]

\[\diamondsuit\]

Finally, let us have a look at (6.4.2) from the numerical point of view. Here we have the task to evaluate integrals of the form (6.4.2) at different points \(x\). If \(x\) is not located near the boundary of the feasible region this could be done with standard formulas for numerical integration. But if we evaluate this integral for a point near the boundary of the feasible region the logarithm will have large absolute values for certain \(t\). Consequently standard formulas for numerical integration do not work very well in this area. Nevertheless, we have to be able to evaluate the barrier function (and therefore also the integrals) near the boundary of the feasible region, because optimal solutions are typically located on the boundary. Due to these problems Schättler refers to special integration rules for evaluating these integrals. Lin et al. [266] suggest to use Simpson’s method to compute similar integrals. In order to achieve a suitable accuracy they decomposed in some cases the interval \([0, 1]\) into 400000 small parts. Thus, independent of the formulae, evaluating such integrals requires a high computational effort.
In our approach we consider the following reformulation of SIP (6.1.1), (6.1.2) with \( U_0 = \mathbb{R}^n \):

\[
\text{minimize } f(x) \quad \text{s.t. } x \in \mathbb{R}^n, \quad \max_{t \in T} g(x, t) \leq 0. \tag{6.4.3}
\]

The theoretical properties as well as practical applications of this approach are extensively studied by Polak [323, 324]. The main advantage of the reformulation (6.4.3) is that we are now dealing with a single constraint. Consequently we get the barrier function

\[
f(x) - \mu \ln \left( -\max_{t \in T} g(x, t) \right). \tag{6.4.4}
\]

But now we face a non-differentiable barrier function due to the involved \( \max \)-term and we have to solve a global auxiliary optimization problem

\[
\text{maximize } g(x, t) \quad \text{s.t. } t \in T \tag{6.4.5}
\]

in order to evaluate the barrier function at a given point \( x \), which is in general a very hard task. Thus except for special cases we cannot suppose that (6.4.5) is exactly solvable for any given \( x \) with acceptable computational effort. Accordingly there is only an approximate maximizer of Problem (6.4.5) available, hence the barrier function is only approximately computable. Consequently we have to use a method for minimizing (6.4.4) which requires an approximately computable objective function.

Such a method, deduced from a bundle method suggested by (cf. Kiwiel [238]), will be presented in the next subchapter. This method requires compact feasible sets. Nevertheless, having in mind that the barrier function (6.4.4) has to be minimized on open sets of the form

\[
\{ x \in \mathbb{R}^n : g(x, t) < 0 (t \in T) \},
\]

in order to use the method proposed we will approximately minimize the convex barrier function successively on compact sets like closed boxes or balls. Altogether we obtain the following conceptual algorithm for solving (6.4.3).

### 6.4.2 Algorithm.

Given \( \mu_1 > 0 \).

For \( k = 1, 2, \ldots \):

- For \( i = 1, 2, \ldots \):
  - Determine a nonempty compact set
    \[
    S^{k,i} \subset \{ x \in \mathbb{R}^n : \max_{t \in T} g(x, t) < 0 \}.
    \]
  - Compute an approximate minimizer \( x^{k,i} \) of (6.4.4) on \( S^{k,i} \).
  - If \( x^{k,i} \) is an approximate unconstrained minimizer of a certain accuracy of (6.4.4) set \( x^k := x^{k,i} \) and leave the inner loop.
- Choose \( \mu_{i+1} \in (0, \mu_i) \).

In the following subchapter we present a numerical method for minimizing the nondifferentiable barrier function (6.4.4) on \( S^{k,i} \) and give all necessary details to put this conceptual algorithm into an implementable form.
6.4.1.1 An implementable barrier algorithm for SIP

We make use of the following general assumptions under which the method is considered.

6.4.3 Assumption. Suppose

1. \( f : \mathbb{R}^n \to \mathbb{R} \) is a convex function;
2. \( T \subset \mathbb{R}^p \) is a compact set;
3. \( g(\cdot, t) \) is convex on \( \mathbb{R}^n \) for any \( t \in T \);
4. \( g(x, \cdot) \) is continuous on \( T \) for any \( x \in \mathbb{R}^n \);
5. set \( \mathcal{M}_0 := \{ x \in \mathcal{M} : \max_{t \in T} g(x, t) < 0 \} \) is nonempty;
6. set of optimal solutions \( \mathcal{M}_{\text{opt}} := \{ x \in \mathcal{M} : f(x) = f^* \} \) of Problem (6.4.3) is nonempty;
7. set \( T_h \) (\( h > 0 \)) is a finite \( h \)-grid on \( T \) (i.e. for each \( t \in T \) there exists \( t_h \in T_h \) with \( \| t - t_h \|_2 \leq h \));
   (in case \( h = 0 \) the sets \( T_h, T \) coincide);
8. for each compact set \( S \subset \mathbb{R}^n \) there exists a constant \( L_S \) with
   \[ |g(x, t_1) - g(x, t_2)| \leq L_S^2 \| t_1 - t_2 \|_2 \] \hspace{1cm} (6.4.6)
   for all \( x \in S \) and all \( t_1, t_2 \in T \);
9. for each compact set \( S \subset \mathcal{M}_0 \) a constant \( C_S < \infty \) with
   \[ C_S \geq \max_{x \in S} \left| \frac{1}{\max_{t \in T} g(x, t)} \right| \] \hspace{1cm} (6.4.7)
   can be computed such that \( S' \subset S \subset \mathcal{M}_0 \) implies \( C_{S'} \leq C_S \);
10. for each \( x \in \mathbb{R}^n \) and each \( t \in T \) an element of the subdifferential of \( f \) in \( x \) and an element of the subdifferential of \( g(\cdot, t) \) in \( x \) can be computed.

Regarding Assumptions (1) and (3) it is ensured that we deal with convex SIP of type (6.4.3). Furthermore, due to (2) and (4) the maximization problems (6.4.5) are solvable and consequently the barrier functions (6.4.4) are computable at least from the theoretical point of view. Moreover, (5) and (6) are motivated by theoretical results below, ensuring the existence of a minimizer of the barrier function (6.4.4) for any given \( \mu > 0 \). Furthermore, the classical logarithmic barrier method with exact minimizers \( x^k \) leads to an optimal solution of SIP (6.4.3) in the sense that each accumulation point of \( \{ x^k \} \) is an optimal solution and \( f(x^k) \to f^* \) (cf. Fiacco [112]). Because we cannot expect that the maximization problems (6.4.5) are exactly solvable, therefore we consider the next assumptions: (7) allows to compute inexact maximizer while their accuracy can be controlled with (8). The necessity of Assumption (9) will become clear in the further course, while by (10) we want to point out that indeed the computation of the subgradients are necessary in the implementation of the method.
6.4. REGULARIZED INTERIOR POINT METHODS

From the classical logarithmic barrier approach it is known that the barrier parameter has to converge to zero, e.g. by reducing it from step to step. But, due to the fact that the conditioning of the barrier problems is getting worse with decreasing the barrier parameter, it makes sense to keep this parameter fixed for a couple of steps.

Again the idea of MSR-methods from Section 4.3 comes into the play (see also [222]). In order to permit a dynamical control, the choice of the barrier parameter has to depend on the progress of the iterates. To avoid side effects which can influence this choice we keep the prox-parameter \( \chi \) constant as long as the barrier parameter is not changed. Merely the proximal point is updated more frequently.

6.4.4 Algorithm.

Given \( \mu_1 > 0, x^0 \in \mathcal{M}_0, \sigma_1 > 0 \) and \( \chi_1 \) with \( 0 < \chi \leq \chi_1 \leq \overline{\chi} \).

For \( k := 1, 2, \ldots \):

- Set \( x^{k,0} := x^{k-1} \).

- For \( i := 1, 2, \ldots \):
  - Set \( x^{k,i,0} := x^{k,i-1} \), select \( \varepsilon_{k,i} > 0 \) and define \( F_{k,i} : \mathcal{M}_0 \rightarrow \mathbb{R} \) by
    \[
    F_{k,i}(x) := f(x) - \mu_k \ln \left( -\max_{t \in T} g(x,t) \right) + \frac{\chi_k}{2} \| x - x^{k,i-1} \|^2.
    \]  
    (6.4.8)
  - For \( s := 1, 2, \ldots \):
    (a) Select \( r_{k,i,s} > 0 \) such that
        \[
        S^{k,i,s} := \{ x \in \mathbb{R}^n : \| x - x^{k,i,s-1} \|_\infty \leq r_{k,i,s} \} \subset \mathcal{M}_0.
        \]
    (b) Select \( h_{k,i,s} \geq 0 \) and define \( \tilde{F}_{k,i,s} : \mathcal{M}_0 \rightarrow \mathbb{R} \) by
        \[
        \tilde{F}_{k,i,s}(x) := f(x) - \mu_k \ln \left( -\max_{t \in T_{h_{k,i,s}}} g(x,t) \right) + \frac{\chi_k}{2} \| x - x^{k,i-1} \|^2
        \]
        where \( T_{h_{k,i,s}} \) fulfills Assumption 6.4.3(7).
    (c) Select \( \beta_{k,i,s} \geq 0 \) and compute an approximate solution \( x^{k,i,s} \) of
        \[
        \text{minimize } F_{k,i}(x) \quad \text{s.t. } x \in S^{k,i,s}
        \]  
        such that
        \[
        \tilde{F}_{k,i,s}(x^{k,i,s}) - \min_{x \in S^{k,i,s}} F_{k,i}(x) \leq \frac{\varepsilon_{k,i}}{2} + \beta_{k,i,s}
        \]  
        (6.4.9)
        and \( \tilde{F}_{k,i,s}(x^{k,i,s}) \leq \tilde{F}_{k,i,s}(x^{k,i,s-1}) \) are true.
    (d) If
        \[
        \tilde{F}_{k,i,s}(x^{k,i,s-1}) - \tilde{F}_{k,i,s}(x^{k,i,s}) \leq \frac{\varepsilon_{k,i}}{2}
        \]  
        (6.4.10)
        then set \( x^{k,i} := x^{k,i,s-1}, S^{k,i} := S^{k,i,s}, r_{k,i} := r_{k,i,s} \) and stop the loop in \( s \), otherwise continue with the loop in \( s \).
CHAPTER 6. PPR FOR CONVEX SEMI-INFINITE PROBLEMS

- If \( \|x^{k,i} - x^{k,i-1}\| \leq \sigma_k \) then set \( x^k := x^{k,i}, r_k := r_{k,i}, i(k) := i \) and stop the loop in \( i \), otherwise continue with the loop in \( i \).

- Select \( 0 < \mu_{k+1} < \mu_k, 0 < \chi \leq \chi_{k+1} \leq \chi \) and \( \sigma_{k+1} > 0 \).

Aspects concerning the numerical realization of the inner loops will be discussed in Subsection 6.4.3.

Let us give some explanations for each particular step. In (a) we specify the compact set \( S^{k,i,s} \) as a linearly bounded set. This decision is caused by the fact that linearly bounded sets are normally the simplest bounded structures on that minimization can be done. Consequently minimizing the barrier function on the chosen compact set is normally easier than minimizing it on more complex structures like quadratically bounded sets such as balls or ellipsoids. Additionally, having a bundle method in mind, we point out that the decision to choose linearly bounded sets is important because the auxiliary problems of the bundle method are linear and quadratic problems. Thus each of them should be efficiently solvable by standard algorithms. Furthermore, since \( \mathcal{M}_0 \) is an open set, step (a) is realizable.

In (b) we define the approximation of the regularized barrier function while in (c) the approximate minimization of the regularized barrier function is done with a certain accuracy. The condition (6.4.10) is stimulated by Corollary 3.6 in [238], when solving (6.4.9) with the bundle method. Finally, in (d) the stopping criterion of the inner loop is given. It is divided into three parts but mainly only two inequalities occur. The first one checks whether there is achieved a sufficient improvement of the approximate solution on the current box with the selected accuracy. The second part of the criterion checks whether the accuracy parameter and the barrier parameter are in an appropriate order. If this is not the case then the accuracy parameter is readjusted.

The critical point for a realization of the presented method is the question whether there exist approximate solutions of (6.4.9) which fulfil the demanded criterions. As stated above (6.4.10) is initiated by a bundle method. Therefore we want to show that in fact we can use a bundle method for solving (6.4.9). We first notice that the sets \( S^{k,i,s} \subset \mathcal{M}_0 \) are convex and compact by construction (for any given \( r_{k,i,s} > 0 \)). Additionally, they are also nonempty, because \( x^{k,i,s-1} \in \mathcal{M}_0 \) holds by construction and due to the open structure of \( \mathcal{M}_0 \) there exists a radius \( r_{k,i,s} > 0 \) such that \( S^{k,i,s} \subset \mathcal{M}_0 \) is fulfilled. Moreover, we have already remarked that the functions \( f_k \) are continuous on \(\text{dom} f_k = \mathcal{M}_0 = \text{int} \text{dom} f_k \) in combination with Theorem 2.35 in [353]). Consequently in (6.4.9) we have to minimize a continuous function on a compact set, which is obviously solvable.

In [1], Lemma 5.3, it is shown that under Assumption 6.4.3 Kiwiel’s proximal level bundle method in [238] can be applied to solve the auxiliary problems arising in Algorithm 6.4.4. Using the constants \( L_S \) and \( C_S \) from Assumption 6.4.3, the value \( \beta_{k,i,s} := \mu_k C_S h_{k,i,s} \) with predefined \( h_{k,i,s} \) can be used as error level when applying a bundle method.
6.4.2 Convergence analysis

In this subsection we want to show that Algorithm 6.4.4 leads to an optimal solution of SIP (6.4.3) under appropriate assumptions. Moreover we investigate the rate of convergence of this method.

We start with a statement about the finiteness of the loop in \( s \).

6.4.5 Lemma. Let Assumption 6.4.3 be fulfilled. Furthermore, let \( k, i \) be fixed, \( \delta_{k,i} > 0 \) and \( q_{k,i} \in (0, 1) \) be given. If

\[
\mu_k L_{S_{k,i,s}}^{1} C_{S_{k,i,s}} h_{k,i,s} \leq \beta_{k,i,s} \leq q_{k,i} \delta_{k,i}, \quad \forall \ s,
\]

then the loop in \( s \) of Algorithm 6.4.4 terminates after a finite number of steps.

Proof: see Proposition 4.6 in [1]. \( \square \)

At this point we want to analyze the consequences of the stopping criterion of the loop in \( s \).

6.4.6 Lemma. Let Assumption 6.4.3 be fulfilled. Furthermore, let \( k, i \) be fixed and \( \hat{x} \) be the unique optimal solution of

\[
\text{minimize } F_{k,i}(x) \quad \text{s.t. } x \in \mathcal{M}_0.
\]

Moreover, let \( x^{k,i,s-1}, x^{k,i,s} \) be generated by Algorithm 6.4.4 and \( \beta_{k,i,s} \geq \mu_k L_{S_{k,i,s}}^{1} C_{S_{k,i,s}} h_{k,i,s} \) be valid. If inequality (6.4.11) is true, then

\[
0 \leq F_{k,i}(x^{k,i,s-1}) - F_{k,i}(\hat{x}) \leq \max \left[ 1, \frac{\|x^{k,i,s-1} - \hat{x}\|_{\infty}}{r_{k,i,s}} \right] (\varepsilon_{k,i} + 2\beta_{k,i,s}) (6.4.14)
\]

and

\[
\|x^{k,i,s-1} - \hat{x}\|_{\infty} \leq \|x^{k,i,s-1} - \hat{x}\| \leq \max \left[ \frac{2(\varepsilon_{k,i} + 2\beta_{k,i,s})}{\chi_k}, \frac{2(\varepsilon_{k,i} + 2\beta_{k,i,s})}{r_{k,i,s}} \right] (6.4.15)
\]

hold.

Proof: First, let us remark that Problem (6.4.13) is uniquely solvable. Inequality (6.4.14) can be shown analogously as in the proof of Lemma 4.5 in [1]. Thus, only the second inequality needs to be proven.

Due to the strong convexity of \( F_{k,i} \) with modulus \( \chi_k/2 \) (in the sense of Definition A1.5.20) we have

\[
F_{k,i}(\lambda x + (1 - \lambda)y) \leq \lambda F_{k,i}(x) + (1 - \lambda)F_{k,i}(y) - \frac{\chi_k}{2} \lambda(1 - \lambda)\|x - y\|^2
\]

for all \( \lambda \in [0, 1] \) and all \( x, y \in \mathcal{M}_0 \). Taking into account that

\[
F_{k,i}(\hat{x}) = \inf_{z \in \mathcal{M}_0} F_{k,i}(z) \leq F_{k,i}(\lambda x + (1 - \lambda)y)
\]

for all \( \lambda \in [0, 1] \) and all \( x, y \in \mathcal{M}_0 \), it follows that

\[
F_{k,i}(\hat{x}) \leq \lambda F_{k,i}(\hat{x}) + (1 - \lambda)F_{k,i}(x^{k,i,s-1}) - \frac{\chi_k}{2} (1 - \lambda)\|\hat{x} - x^{k,i,s-1}\|^2
\]
and

\[(1 - \lambda)F_{k,i}(\tilde{x}) \leq (1 - \lambda)F_{k,i}(x^{k,i,s-1}) - \frac{\lambda}{2} \|\tilde{x} - x^{k,i,s-1}\|^2\]

are true for all \(\lambda \in [0,1]\). Hence,

\[F_{k,i}(\tilde{x}) \leq F_{k,i}(x^{k,i,s-1}) - \frac{\lambda}{2} \|\tilde{x} - x^{k,i,s-1}\|^2\]

for all \(\lambda \in [0,1]\) so that for \(\lambda \uparrow 1\)

\[\frac{\lambda}{2} \|\tilde{x} - x^{k,i,s-1}\|^2 \leq F_{k,i}(x^{k,i,s-1}) - F_{k,i}(\tilde{x}).\]  \(\text{(6.4.16)}\)

Using (6.4.14) one obtains

\[\frac{\lambda}{2} \|\tilde{x} - x^{k,i,s-1}\|^2 \leq \max \left[1, \frac{\|x^{k,i,s-1} - \tilde{x}\|_\infty}{r_{k,i,s}} (\varepsilon_{k,i} + 2\beta_{k,i,s}) \right].\]

At this point we distinguish two cases. We first suppose that \(1 \geq \|x^{k,i,s-1} - \tilde{x}\|_\infty/r_{k,i,s}\). Then it holds

\[\frac{\lambda}{2} \|\tilde{x} - x^{k,i,s-1}\|^2 \leq (\varepsilon_{k,i} + 2\beta_{k,i,s})\]

and

\[\|x^{k,i,s-1} - \tilde{x}\|_\infty \leq \|\tilde{x} - x^{k,i,s-1}\| \leq \sqrt{\frac{2(\varepsilon_{k,i} + 2\beta_{k,i,s})}{\lambda k}}.\]  \(\text{(6.4.17)}\)

In the second case the inequality \(1 < \|x^{k,i,s-1} - \tilde{x}\|_\infty/r_{k,i,s}\) is supposed to be true. Then

\[\frac{\lambda}{2} \|\tilde{x} - x^{k,i,s-1}\|^2 \leq \frac{\|x^{k,i,s-1} - \tilde{x}\|_\infty}{r_{k,i,s}} (\varepsilon_{k,i} + 2\beta_{k,i,s})\]

\[\leq \frac{\|x^{k,i,s-1} - \tilde{x}\|}{r_{k,i,s}} (\varepsilon_{k,i} + 2\beta_{k,i,s})\]

is valid. We conclude that

\[\|\tilde{x} - x^{k,i,s-1}\|_\infty \leq \|\tilde{x} - x^{k,i,s-1}\| \leq \frac{2(\varepsilon_{k,i} + 2\beta_{k,i,s})}{\lambda k r_{k,i,s}}.\]  \(\text{(6.4.18)}\)

Combining (6.4.17) and (6.4.18) completes the proof. \(\square\)

In the following we denote the Euclidean ball with radius \(\tau > 0\) around \(x_c \in \mathbb{R}^n\) by \(B_{\tau}(x_c)\), i.e.

\[\mathbb{B}_{\tau}(x_c) := \{ x \in \mathbb{R}^n : \|x - x_c\| \leq \tau \} .\]

\(\text{6.4.7 Theorem.}\) Let Assumption 6.4.3 be fulfilled. Furthermore, let \(\tau \geq 1\) and \(x_c \in \mathbb{R}^n\) be chosen such that \(\mathcal{M}_{opt} \cap B_{\tau/8}(x_c) \neq \emptyset\). Let \(x^* \in \mathcal{M}_{opt} \cap B_{\tau/8}(x_c)\), \(\tilde{x} \in \mathcal{M}_0 \cap B_{\tau/4}(x_c)\) and \(\tilde{x}^0 \in \mathcal{M}_0 \cap B_{\tau/4}(x_c)\) be fixed and \(\delta_{k,i} > 0\), \(q_{k,i} \in (0,1)\), \(\alpha_k > 0\), \(i \in T(\tilde{x})\), \(\tilde{v} \in \partial g(\tilde{x}, \tilde{v})\) as well as

\[\varepsilon \geq \|\tilde{x} - x^*\| \quad \text{and} \quad c_0 := f(\tilde{x}) - f_- + c_0 + c_1\]
6.4. REGULARIZED INTERIOR POINT METHODS

with

\[ f_\leq = \min_{x \in M^*} f(x), \quad c_0 := \ln \left( -\max_{t \in T} g(\bar{x}, t) \right), \quad c_1 := \ln \left( -\max_{t \in T} g(\bar{x}, t) + 2\|v\| \right) \]

be given. Moreover, assume that (6.4.12) is true for \( k = 1, 2, \ldots, 1 \leq i \leq i(k) \) and all \( s \) occurring in the outer loop \((k, i)\) and that the controlling parameters of Algorithm 6.4.4 satisfy the following conditions:

\[ \max \left[ \frac{2(\varepsilon_{k,i} + 2\delta_{k,i})}{\lambda_k}, \frac{2(\varepsilon_{k,i} + 2\delta_{k,i})}{\sigma_{k,i} \lambda_k} \right] \leq \alpha_i, \]  
\[ 0 < \mu_{k+1} \leq \mu_k < 1 \quad \text{for} \quad k = 1, 2, \ldots, \quad \mu_1 \leq e^{-c_0}, \]  
\[ \sum_{k=1}^{\infty} \left[ \frac{2\mu_k}{\lambda_k} (2|\ln \mu_k| + \ln \tau) + 2\sigma_\tau \mu_k + \alpha_k \right] < \tau \]

and

\[ \sigma_k > \frac{2\mu_k}{\lambda_k} (2|\ln \mu_k| + \ln \tau) + 4\tau\alpha_k + \alpha_k. \]

Then it holds

(1) the loop in \( s \) is finite for each \((k, i)\);

(2) the loop in \( i \) is finite for each \( k \), i.e. \( i(k) < \infty \);

(3) \( \|x^{k,i} - x^*\| < \tau \) for all pairs \((k, i)\) with \( 0 \leq i \leq i(k) \);

(4) the sequence \( \{x^{k,i}\} = \{x^{1,0}, \ldots, x^{1,i(1)}, x^{2,0}, \ldots, x^{2,i(2)}, x^{3,0}, \ldots\} \) converges to an element \( x^* \in M_{\text{opt}} \cap B_\tau(x_c). \)

**Proof:** The first statement follows immediately from Lemma 6.4.5. The other statements can be proven similarly to the proof of Theorem 1 in [222].

We first define \( z^k = \mu_k \bar{x} + (1 - \mu_k)x^* \). Due to \( 0 < \mu_k < 1 \), \( x \in M_0 = \text{int} M \), \( x^* \in M \) and the fact that \( M \) is convex one can infer \( z^i \in M_0 \) by means of Theorem 6.1 in [348]. Then it follows

\[ -\mu_k \ln \left( -\max_{t \in T} g(z^k, t) \right) = -\mu_k \ln \left( -\max_{t \in T} g(\mu_k \bar{x} + (1 - \mu_k)x^*, t) \right) \]
\[ \leq -\mu_k \ln \left( -\mu_k \max_{t \in T} g(\bar{x}, t) - (1 - \mu_k) \max_{t \in T} g(x^*, t) \right) \]
\[ \leq -\mu_k \ln \left( -\mu_k \max_{t \in T} g(\bar{x}, t) \right) \]
\[ = -\mu_k \left( \ln \mu_k + \ln \left( -\max_{t \in T} g(\bar{x}, t) \right) \right) \leq \mu_k (|\ln \mu_k| + c_0), \]  

because the max–function is convex and the logarithm increases monotonically. Furthermore, the estimates

\[ \|z^k - x^*\| = \mu_k \|\bar{x} - x^*\| \leq \tau \mu_k \]  

and

\[ f(z^k) \leq \mu_k f(\bar{x}) + (1 - \mu_k)f(x^*) \leq f(x^*) + \mu_k (f(\bar{x}) - f_-) \]  

(6.4.25)
are obviously true. Additionally, this yields
\[ \max_{t \in T} g(x, t) \geq g(x, \tilde{t}) \geq g(\tilde{x}, \tilde{t}) + \langle \tilde{v}, x - \tilde{x} \rangle = \max_{t \in T} g(\tilde{x}, t) + \langle \tilde{v}, x - \tilde{x} \rangle \]
for all \( x \in \mathbb{R}^n \) since \( \tilde{v} \in \partial g(\tilde{x}, \tilde{t}), \tilde{t} \in T(\tilde{x}) \). Consequently, using the Cauchy-Schwarz inequality and \( \tilde{x} \in B_r(x_c) \), we obtain
\[ 0 < -\max_{t \in T} g(x, t) \leq -\max_{t \in T} g(\tilde{x}, t) + 2\tau \| \tilde{v} \|, \quad \forall \ x \in B_r(x_c) \cap M_0. \]
Regarding the monotonicity of the logarithm function and \( \tau \geq 1 \), this leads to
\[ \inf \left\{ -\mu_k \ln \left( -\max_{t \in T} g(x, t) \right) : x \in B_r(x_c) \cap M_0 \right\} \geq -\mu_k (c_1 + \ln \tau). \quad (6.4.26) \]
For \( x \in M_0 \) we introduce
\[ f_k(x) := f(x) - \mu_k \ln \left( -\max_{t \in T} g(x, t) \right). \]
The inequalities (6.4.23), (6.4.25), (6.4.26) and the optimality of \( x^* \) show that for all \( x \in M_0 \cap B_r(x_c) \)
\[ f_k(x^k) \leq f_k(x) + \mu_k (2|\ln \mu_k| + \ln \tau) \quad \forall \ x \in M_0 \cap B_r(x_c). \quad (6.4.28) \]
Using the results above we can prove the second and third statement of the theorem by induction. To do so we assume:
(i) \( k_0, i_0 \) are kept fixed with \( 0 \leq i_0 < i(k_0) \),
(ii) \( i(k) < \infty \) if \( k < k_0 \),
(iii) denoting
\[ \bar{x}^{k,i} := \arg \min_{x \in M_0} F_{k,i}(x), \]
\[ \bar{x}^{k,i} := \arg \min_{x \in M} \left\{ f(x) + \frac{\lambda_k}{2} \| x - x^{k,i-1} \|^2 \right\}, \]
the following relations hold
\[ \| x^{k,i} - x_c \| < \tau, \quad \| x^{k,i} - x_c \| < \| \bar{x}^{k,i} - x_c \| < \tau, \quad (6.4.29) \]
hold for all pairs of indices
\( (k, i) \in Q_0 := \{(k', i') : k' < k_0, 0 < i' \leq i(k') \} \cup \{k' = k_0, 0 < i' \leq i_0 \} \).
Let us remark that \( \bar{x}_{k,i} \in \mathcal{M}_0 \) as minimizer of \( F_{k,i} \) exists. The existence of \( \bar{x}_{k,i} \in \mathcal{M} \) as minimizer of \( f(x) + \frac{\lambda_k}{2} \| x - x^{k,i-1} \|^2 \) is ensured by the strong convexity and continuity of this function on the nonempty and closed set \( \mathcal{M} \).

At this point we have to check (i)-(iii) for the starting values \( k_0 = 1, i_0 = 0 \), but this is easy: By construction \( i(1) > 0 \) so that \( k_0 = 1, i_0 = 0 \) fulfill the first assumption. The other two assumptions are also obvious by construction.

Using the stopping criterion of the loop in \( s \) of Algorithm 6.4.4, (6.4.12), (6.4.15), (6.4.19) as well as the definition of \( \bar{x}_{k,i} \) we deduce
\[
\| \bar{x}_{k,i} - x_{k,i} \| \leq \alpha_k. \tag{6.4.30}
\]
Furthermore, taking the definitions of \( \bar{x}_{k,i} \) and \( \bar{x}_{k,i} \) into account we can conclude
\[
f(\bar{x}_{k,i}) + \frac{\lambda_k}{2} \| x_{k,i} - x_{k,i-1} - \frac{\lambda_k}{2} \| \| \bar{x}_{k,i} - x_{k,i-1} \|^2 \leq \mu_k. \tag{6.4.31}
\]
Additionally one can establish
\[
\frac{\lambda_k}{2} \| \bar{x}_{k,i} - x_{k,i} \|^2 \leq f(\bar{x}_{k,i}) + \frac{\lambda_k}{2} \| x_{k,i} - x_{k,i-1} \|^2 - f(\bar{x}_{k,i}) - \frac{\lambda_k}{2} \| \bar{x}_{k,i} - x_{k,i-1} \|^2
\]
in the same manner as (6.4.16). Combining (6.4.31) and (6.4.32) we see
\[
\frac{\lambda_k}{2} \| \bar{x}_{k,i} - x_{k,i} \|^2 \leq \mu_k,
\]
so that
\[
\| \bar{x}_{k,i} - x_{k,i} \|^2 \leq \frac{2\mu_k}{\lambda_k}. \tag{6.4.33}
\]
Using this and (6.4.30) we obtain
\[
\| \bar{x}_{k,i} - x_{k,i} \| \leq \frac{\sqrt{2\mu_k}}{\lambda_k}. \tag{6.4.34}
\]
Due to \( \bar{x}_{k,i} \in \mathcal{M}_0 \cap B_{\tau}(x_c) \) estimate (6.4.28) implies
\[
f_k(z_k) \leq f_k(\bar{x}_{k,i}) + \mu_k \left( 2 \ln \mu_k + \ln \tau \right), \quad \forall (k, i) \in Q_0. \tag{6.4.35}
\]
In the sequel we distinguish the following cases:

(a) \( k < k_0, 0 \leq i < i(k) - 1 \) or \( k = k_0, 0 \leq i < i_0 \),

(b) \( k < k_0, i = i(k) - 1 \) and

(c) \( k = k_0, i = i_0 + 1 \).

ad (a): In this case we obtain
\[
\| x_{k,i+1} - z_k \|^2 - \| x_{k,i} - z_k \|^2 \leq - \| x_{k,i+1} - x_{k,i} \|^2 + \frac{2 \mu_k}{\lambda_k} \left( f_k(z_k) - f_k(\bar{x}_{k,i+1}) \right)
\]
\[
\leq - \| x_{k,i+1} - x_{k,i} \|^2 + \frac{2 \mu_k}{\lambda_k} \left( 2 \ln \mu_k + \ln \tau \right)
\]
\[
6.4. \quad \text{REGULARIZED INTERIOR POINT METHODS} \quad 213
\]
by using Proposition 3.1.3 and (6.4.35). Taking into account (6.4.22), (6.4.30) and the stopping criterion of the loop in $i$ of Algorithm 6.4.4, we conclude

$$
\|x^{k,i+1} - x^{k,i}\| \geq \|x^{k,i+1} - x^{k,i}\| - \|x^{k,i+1} - x^{k,i}\| > \sigma_k - \alpha_k > 0
$$

(6.4.37)

and we have

$$
\|x^{k,i+1} - z^k\|^2_2 - \|x^{k,i} - z^k\|^2_2 \leq -\tilde{\varepsilon}_k^2 + \gamma_k < 0
$$

(6.4.38)

with $\tilde{\varepsilon}_k = \sigma_k - \alpha_k$ and $\gamma_k = 2\mu_k(2 \ln \mu_k + \ln \tau)/\chi_k$.

Moreover, regarding $\|x^{k,i} - x_c\| < \tau$ and $\|z^k - x_c\| < \tau$, the estimate

$$
\|x^{k,i+1} - z^k\| - \|x^{k,i} - z^k\| < \frac{1}{4\tau} (-\tilde{\varepsilon}_k^2 + \gamma_k)
$$

(6.4.39)

holds. Together with (6.4.22) and (6.3.60) we obtain

$$
\|x^{k,i+1} - z^k\| - \|x^{k,i} - z^k\| < \frac{1}{4\tau} (-\tilde{\varepsilon}_k^2 + \gamma_k) + \alpha_k < 0.
$$

(6.4.40)

ad (b): Now we assume $k < k_0$, $i = i(k) - 1$.

In this case we can combine (6.4.28), (6.4.30) and the implications of Proposition 3.1.3 to see that

$$
\|x^{k,i(k)} - z^k\| - \|x^{k,i(k)-1} - z^k\| \leq \frac{2}{\chi_k} (f_k(z^k) - f_k(x^{k,i(k)})) + \alpha_k \leq \frac{2\mu_k}{\chi_k} (2 \ln \mu_k + \ln \tau) + \alpha_k
$$

(6.4.41)

holds. Summing up the inequalities (6.4.40) w.r.t. $i = 0, 1, \ldots, i(k) - 2$ for a fixed $k < k_0$ and adding (6.4.41) leads to

$$
\|x^{k,i(k)} - z^k\| - \|x^{k,0} - z^k\| \leq \sqrt{\gamma_k} + \alpha_k,
$$

(6.4.42)

and together with (6.4.24) one has

$$
\|x^{k,i(k)} - x^*\| - \|x^{k,0} - x^*\| \leq \sqrt{\gamma_k} + \alpha_k + 2\sqrt{\gamma_k}.
$$

(6.4.43)

ad (c): Now we assume $k = k_0$, $i = i_0 + 1$. In this case we consider

$$
\hat{x}^{k_0,i_0+1} := \arg \min_{x \in \mathcal{M}^{\mathbb{R}_+}(x_o)} \left\{ f(x) + \frac{\chi_k}{2} \|x - x^{k_0,i_0}\|_2^2 \right\}.
$$

The non-expansivity of the prox-mapping (see Remark 3.1.2) yields

$$
\|\hat{x}^{k_0,i_0+1} - x^*\| \leq \|x^{k_0,i_0} - x^*\|.
$$

(6.4.44)

Using this, (6.4.24) and (6.4.40) for $k = k_0$, $0 \leq i < i_0$ we obtain

$$
\|\hat{x}^{k_0,i_0+1} - x^*\| \leq \|x^{k_0,i_0} - z^{k_0}\| + \tilde{\varepsilon}_k \mu_k
$$

$$
\leq \|x^{k_0,0} - z^{k_0}\| + \tilde{\varepsilon}_k \mu_k
$$

$$
\leq \|x^{k_0,0} - x^*\| + 2\tilde{\varepsilon}_k \mu_k.
$$
If \( k_0 > 1 \) this leads to
\[
\| \tilde{x}^{k_0,0} - z^* \| \leq \| x^{k_0-1,i(k_0-1)} - x^* \| + 2\epsilon \mu_{k_0}.
\]

Now a successive application of (6.4.43) gives
\[
\| \tilde{x}^{k_0,i+1} - x^* \| \leq \| x^{1.0} - x^* \| + \sum_{j=1}^{k_0-1} (\sqrt{\gamma_j} + \alpha_j + 2\epsilon \mu_j) + 2\epsilon \mu_{k_0}. \tag{6.4.45}
\]

Due to the assumptions \( \| x^* - x_c \| \leq \frac{T}{T} \) and \( \| x^{1.0} - x_c \| < \frac{T}{T} \) we can now assemble (6.4.20), (6.4.21) and (6.4.45) to get
\[
\| \tilde{x}^{k_0,i+1} - x_c \| < \tau - \alpha_{k_0} - \sqrt{\gamma_{k_0}} < \tau - \alpha_{k_0} - \sqrt{\frac{2\mu_{k_0}}{\chi_{k_0}}} \tag{6.4.46}
\]

We see that \( \| \tilde{x}^{k_0,i+1} - x_c \| < \tau \) and due to the strong convexity of \( f + \frac{\chi_{k_0}}{2} \| x - x_{opt} \|^2 \) we can deduce that \( \tilde{x}^{k_0,i+1} \) must actually be identical to \( \tilde{x}^{k_0,i+1} \). But then the estimates (6.4.33) and (6.4.44) imply \( \| \tilde{x}^{k_0,i+1} - x_c \| < \tau \) as well.

So far we have proven that Assumption (iii) is also true for \( k = k_0, i = i_0 + 1 \). It still remains to prove that \( i(k_0) < \infty \) holds. In order to do so we sum up the inequalities (6.4.40) to an arbitrary \( i \leq i(k_0) - 1 \) and obtain
\[
\| \tilde{x}^{k_0,i} - z^k \| < \| x^{k_0,0} - z^k \| + \gamma \left( \frac{1}{4\tau} (-\tilde{\varepsilon}_k^2 + \gamma_k) + \alpha_k \right),
\]
giving an upper bound for \( \tilde{i} \):
\[
\tilde{i} < \frac{-\| x^{k_0,0} - z^k \|}{\frac{1}{4\tau} (-\tilde{\varepsilon}_k^2 + \gamma_k) + \alpha_k} < \infty. \tag{6.4.47}
\]

Thus we have shown the induction statements to hold for \( k_0, i_0 + 1 \) if \( i_0 < i(k_0) \). But the case \( i_0 = i(k_0) \) is equivalent to the case \( (k_0+1, 0) \) and so the induction holds for all possible indices \((k_0, i_0)\). As a consequence of this the second and third statement of the theorem are proven.

It remains to prove the convergence of the generated sequence \\{x^{k,i}\} to an optimal solution of SIP (6.4.3). Let \( \bar{x} \) be an arbitrary element of \( \mathcal{M}_{opt} \cap B_\tau(x_c) \). Defining
\[
\tilde{z}^k := \tilde{x} + (1 - \mu_k)(\bar{x} - \tilde{x}),
\]
we can show \( \| \tilde{z}^k - \bar{x} \| \leq 2\tau \mu_k \) similar to (6.4.24) and analogous results to (6.4.42) and (6.4.43) with \( \tilde{z}^k \) instead of \( z^k \) and \( \bar{x} \) instead of \( x^* \). Additionally, we obtain from (6.4.21)
\[
\sum_{k=1}^{\infty} \sqrt{\gamma_k} < \infty, \quad \sum_{k=1}^{\infty} \mu_k < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \alpha_k < \infty
\]
and the convergence of \( \{ \| x^{k,0} - \bar{x} \| \} \) is ensured by Lemma A3.1.7. Moreover, the results (6.4.27), (6.4.36), (6.4.40) and (6.4.41) remain true if we use \( \tilde{z}^k \) instead of \( z^k \) and
\[
\| x^{k,i} - \tilde{z}^k \| < \| x^{k,0} - \tilde{z}^k \|, \quad \| x^{k+1,0} - \tilde{z}^k \| \leq \sqrt{\gamma_k} + \alpha_k + \| x^{k,i} - \tilde{z}^k \|
\]
for all $k$ and $0 < i < i(k)$. Since $\bar{z}^k, \bar{x} \in B_\tau(x_c)$ this leads to
\[
\|x^{k+1,0} - \bar{x}\| - \sqrt{\gamma_k} - \alpha_k - 4\tau \mu_k \leq \|x^{k,i} - \bar{x}\| < \|x^{k,0} - \bar{x}\| + 4\tau \mu_k,
\]
hence the sequence $\{\|x^{k,i} - \bar{x}\|\}$ converges. Furthermore, regarding (6.4.30) and
\[
\lim_{k \to \infty} \alpha_k = 0 \text{ which is enforced by (6.4.21), it is clear that } \{\|\bar{x}^{k,i} - \bar{x}\|\}
\]
converges to the same limit point.

Due to $\bar{x}, \bar{x} \in B_\tau(x_c)$ and $0 < \mu_k < 1$ for $k = 1, 2, \cdots$ we have $\bar{z}^k \in B_\tau(x_c)$
for all $k \in \mathbb{N}$ as well as
\[
\|x^{k,i-1} - \bar{z}^k\| \leq \|x^{k,i-1} - \bar{x}\| + 2\tau \mu_k
\]
for all pairs $(k, i)$ with $1 \leq i \leq i(k)$ and
\[
\|\bar{x}^{k,i} - \bar{z}^k\| \geq \|\bar{x}^{k,i} - \bar{x}\| - 2\tau \mu_k
\]
for all pairs $(k, i)$ with $0 \leq i \leq i(k)$. Consequently we obtain
\[
\|x^{k,i-1} - \bar{z}^k\|^2 \leq \|x^{k,i-1} - \bar{x}\|^2 + 8\tau^2 \mu_k + 4\tau^2 \mu_k^2
\]
for all pairs $(k, i)$ with $0 \leq i < i(k)$ and
\[
\|\bar{x}^{k,i} - \bar{z}^k\|^2 \geq \|\bar{x}^{k,i} - \bar{x}\|^2 - 8\tau^2 \mu_k - 4\tau^2 \mu_k^2
\]
for all pairs $(k, i)$ with $0 \leq i < i(k)$. Additionally the modified estimates (6.4.27)
\[
f_k(z^k) \leq f_k(x) + \mu_k(c_3 + \ln \tau + |\ln \mu_k|)
\]
for all $x \in M_0 \cap B_\tau(x_c)$ and (6.4.36)
\[
\|\bar{x}^k - \bar{z}^k\|^2 - \|x^{k,i-1} - \bar{z}^k\|^2 \leq \frac{2}{\chi_k} (f_k(z^k) - f_k(\bar{x}^{k,i}))
\]
allow to infer
\[
\|x^{k,i-1} - \bar{x}\|^2 - \|x^{k,i} - \bar{x}\|^2 \geq 
\]
\[
\geq \frac{2}{\chi_k} (f_k(\bar{x}^{k,i}) - f_k(x) - \mu_k(c_3 + \ln \tau + |\ln \mu_k|)) - 16\tau^2 \mu_k - 8\tau^2 \mu_k^2
\]
for all $x \in M_0 \cap B_\tau(x_c)$. Then, regarding $\bar{x}^{k,i} \in M_0 \cap B_\tau(x_c)$ and estimate (6.4.26), we obtain
\[
\|x^{k,i-1} - \bar{x}\|^2 - \|x^{k,i} - \bar{x}\|^2 \geq 
\]
\[
\geq \frac{2}{\chi_k} (f(\bar{x}^k) - f_k(x) - \mu_k(c_3 + \ln \tau + |\ln \mu_k|)) - 8\tau^2 \mu_k(2 + \mu_k).
\]
Furthermore, we have $\lim_{k \to \infty} f_k(x) = f(x)$ for each fixed $x \in M_0$, thus $\mu_k \to 0$
and $\chi_k \leq \chi$ give
\[
\limsup_{k \to \infty} \left( \max_{1 \leq i \leq i(k)} (f(\bar{x}^{k,i}) - f(x)) \right) \leq 0
\]
for each fixed \( x \in M_0 \cap \bar{B}_\tau(x_c) \).

Now let \( x^{**} \) be an accumulation point of the sequence \( \{x^{k,i}\} \). Such an accumulation point exists because \( x^{k,i} \in \bar{B}_\tau(x_c) \cap M \) for all pairs \((k,i)\). Regarding (6.4.30) and \( \lim_{k \to \infty} \alpha_k = 0 \) it follows that \( x^{**} \) is also an accumulation point of the sequence \( \{x^{k,i}\} \). Furthermore we obtain \( x^{**} \in M \cap \bar{B}_\tau(x_c) \) because the sets \( M \) and \( \bar{B}_\tau(x_c) \) are closed. For each \( x \in M_0 \cap \bar{B}_\tau(x_c) \) estimate (6.4.49) establishes \( f(x) \) as an upper bound for \( f(x^{**}) \) so that we deduce

\[
f(x^{**}) \leq \inf \{ f(x) : x \in M_0 \cap \bar{B}_\tau(x_c) \}. \tag{6.4.50}
\]

Obviously \( M \cap \bar{B}_\tau(x_c) \) is the closure of \( M_0 \cap \bar{B}_\tau(x_c) \). Furthermore \( x^* \in M_{\text{opt}} \cap \bar{B}_\tau(x_c) \) such that (6.4.50) implies \( f(x^{**}) \leq f(x^*) \), respectively \( x^{**} \in M_{\text{opt}} \cap \bar{B}_\tau(x_c) \). Consequently, regarding that \( \|x^{k,i} - \bar{x}\| \) converges for each \( \bar{x} \in M_{\text{opt}} \cap \bar{B}_\tau(x_c) \), the sequence \( \{\|x^{k,i} - x^{**}\|\} \) converges to zero. Thus the sequence \( \{x^{k,i}\} \) converges to \( x^{**} \in M_{\text{opt}} \).

6.4.8 Remark. If \( \max_{t \in T} g(\cdot, t) \) is bounded below on the feasible set \( M \) of SIP (6.4.3), i.e. there exists a (non-positive) constant \( d > -\infty \) with \( d \leq \max_{t \in T} g(x,t) \) for all \( x \in M \), one obtains

\[
\inf \left\{ -\mu_k \ln \left( -\max_{t \in T} g(x, t) \right) : x \in \bar{B}_\tau(x_c) \right\} \geq -\mu_k \ln(-d).
\]

Consequently, using this estimate instead of (6.4.26) in the proof above, the conditions on the parameter of Algorithm 6.4.4 in Theorem 6.4.7 can be simplified in the considered case. In particular, \( c_3 \) can be changed into

\[
c_3 := f(\bar{x}) - f_* - c_0 - \ln(-d)
\]

and in (6.4.21) and (6.4.22) the term \( \ln \tau \) can be dropped. Thus the left-hand side of the modified estimate (6.4.21) does not depend on (the unknown) \( \tau \). Therefore one could choose the value of \( \tau \) after determining \( \{\mu_k\} \), \( \{\alpha_k\} \) and \( \{\chi_k\} \). Finally, the value of \( \sigma_k \) can be fixed such that (6.4.22) holds. Altogether the described procedure is much easier than the simultaneous determination of all parameters in the general case.

6.4.9 Remark. The conditions on the parameters of the method require their separate adjustment for each particular example, which can be a very fragile task when applying a MSR-method. In the case of an OSR-method parameters according to Theorem 6.4.7 can be easily chosen. As we already know, an OSR-method procedure is given if \( i(k) = 1 \) for each \( k \), which can be ensured by choosing \( \sigma_k \) sufficiently large. Of course, applying a MSR-method large values of \( \sigma_k \) must be avoided. Then (6.4.22) is automatically satisfied for each fixed \( \tau \geq 1 \). Furthermore, (6.4.21) holds for all sufficiently large values of \( \tau \) if one guarantees that

\[
\sum_{k=1}^{\infty} \sqrt{\frac{\mu_k |\ln \mu_k|}{\chi_k}} < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \alpha_k < \infty.
\]

Consequently, (6.4.21) and (6.4.22) can be replaced by the given conditions above and \( \tau, \sigma_k \) need not to be specified explicitly.
At the end of this subsection an estimate of the difference between the current value of the objective function \( f \) at each exterior step and its minimal value \( f^* \) on \( \mathcal{M} \) is established (cf. also Theorem 4.3.10 and Theorem 6.3.7).

**6.4.10 Lemma.** Let the assumptions of Theorem 6.4.7 be satisfied and let \( f \) be Lipschitz continuous with modulus \( L \) on \( \mathbb{B}_r(x_c) \). Then it holds

\[
f(x^k) - f^* \leq \left( \frac{9}{8} \nu_k + L \right) \left( \sqrt{\frac{2\mu_k}{\chi_k}} + \alpha_k \right) + \frac{9}{8} \tau \chi_k \sigma_k, \quad k = 1, 2, \ldots.
\]

**Proof:** Let \( k \) be fixed and \( 0 \leq i < i(k) \) be arbitrarily given. In the proof of Theorem 6.4.7 we defined already

\[
x(k,i+1) := \arg \min_{x \in \mathcal{M}} \left\{ f(x) + \frac{\nu_k}{2} \| x - x(k,i) \|^2 \right\}.
\]

The inclusion \( x(k,i+1) \in \mathbb{B}_r(x_c) \) has already been shown. Because \( \mathcal{M} \cap \mathbb{B}_r(x_c) \) is convex, \( (1 - \lambda)x^{k,i+1} + \lambda x^* \in \mathcal{M} \cap \mathbb{B}_r(x_c) \) for all \( \lambda \in [0,1] \). The optimality of \( x(k,i+1) \) gives

\[
f((1 - \lambda)x^{k,i+1} + \lambda x^*) + \frac{\nu_k}{2} \| (1 - \lambda)x^{k,i+1} + \lambda x^* - x(k,i) \|^2
\]

\[
\geq f(x(k,i+1)) + \frac{\nu_k}{2} \| x^{k,i+1} - x(k,i) \|^2,
\]

so that, regarding the convexity of \( f \), for all \( \lambda \in (0,1] \)

\[
0 \leq \lambda f(x^*) - \lambda f(x(k,i+1)) + \frac{\nu_k}{2} \| x^* - x(k,i+1) \|^2 + \lambda \chi_k \langle x^* - x(k,i+1), x^{k,i+1} - x(k,i) \rangle
\]

and

\[
f(x(k,i+1)) - f(x^*) \leq \nu_k \langle x^* - x(k,i+1), x^{k,i+1} - x(k,i) \rangle + \frac{\nu_k}{2} \| x^* - x(k,i+1) \|^2.
\]

Passing to the limit \( \lambda \downarrow 0 \) combined with the Cauchy-Schwarz inequality leads to

\[
f(x(k,i+1)) - f(x^*) \leq \nu_k \| x^* - x(k,i+1) \| \left( \| x^{k,i+1} \| + \| x^{k,i+1} - x(k,i) \| \right)
\]

and using the Lipschitz continuity of \( f \)

\[
f(x(k,i+1)) - f(x^*) \leq \| x^* - x(k,i+1) \| \left( L + \nu_k \| x^* - x(k,i+1) \| \right)
\]

\[
+ \nu_k \| x^* - x(k,i+1) \| \| x^{k,i+1} - x(k,i) \|.
\]

In view of (6.4.34) and

\[
\| x^* - x(k,i+1) \| \leq \| x^* - x_c \| + \| x^{k,i+1} - x_c \| \leq \frac{\tau}{8} + \tau = \frac{9}{8} \tau,
\]

we obtain

\[
f(x(k,i+1)) - f(x^*) \leq \left( \frac{9}{8} \nu_k + L \right) \left( \sqrt{\frac{2\mu_k}{\chi_k}} + \alpha_k \right) + \frac{9}{8} \tau \chi_k \| x^{k,i+1} - x(k,i) \|,
\]

so that the statement follows by means of \( \| x^{k,i(k)} - x^{k,i(k)-1} \| \leq \sigma_k, \| x^k - x^{k,i(k)} \| \) and \( f(x^k) = f^* \). \( \square \)
6.4.2.1 Rate of convergence

In the following subsections we analyze further convergence properties of Algorithm 6.4.4 with regard to the rate of convergence based on results of Kaplan and Tichatschke [222]. In doing so we consider the sequence \( \{ \bar{x}^{k,i} \} \) instead of the generated sequence \( \{ x^{k,i} \} \) defined as in the proof of Theorem 6.4.7. That means we consider the sequence of the exact minimizer of \( F^{k,i} \) instead of the computed approximate minimizer.

However, based on results for the exact minima we can also achieve results for the approximate minima, e.g. by using (6.4.30).

First the value of \( \max_{t \in T} g(\bar{x}^{k,i}, t) \) is estimated.

6.4.11 Lemma. Let the assumptions of Theorem 6.4.7 be satisfied. Then for each \( k \) and \( 1 \leq i \leq i(k) \)

\[
- \max_{t \in T} g(\bar{x}^{k,i}) \geq c_4 \mu_k
\]

holds with

\[
c_4 := - \frac{\max_{t \in T} g(\bar{x}, t)}{\mu_1 + f(\bar{x}) - f^* + 2\tau^2 \chi}
\]

and \( \bar{x} \) defined in Theorem 6.4.7.

Proof: Let \( k, i \) be arbitrarily given with \( 1 \leq i \leq i(k) \). Due to \( \bar{x}^{k,i} \in \mathcal{M}_0 = \text{dom} F^{k,i} \), Theorem 23.1 in Rockafellar [348] ensures the existence of the directional derivative \( F^{k,i}_d(\bar{x}^{k,i}; d) \) for each \( d \in \mathbb{R}^n \). Since \( F^{k,i} \) attains its minimal value at \( \bar{x}^{k,i} \) and since \( \mathcal{M}_0 \) is open we obtain \( F^{k,i}_d(\bar{x}^{k,i}; d) \geq 0 \) for each \( d \in \mathbb{R}^n \) so that \( 0 \in \partial F^{k,i}(\bar{x}^{k,i}) \) follows from Theorem 23.2 in [348]. From subdifferential calculus we already know

\[
\partial F^{k,i}(\bar{x}^{k,i}) \supset \partial f^{k,i}(\bar{x}^{k,i}) + \{ \chi_k (\bar{x}^{k,i} - x^{k,i-1}) \}.
\]

Regarding

\[
\bar{x}^{k,i} \in \mathcal{M}_0 = \text{ridom}(f_k) \cap \text{ridom} \left( \frac{\chi_k}{2} \| . - x^{k,i-1} \|^2 \right),
\]

Theorem 23.8 in [348] even leads to

\[
\partial F^{k,i}_d(\bar{x}^{k,i}) = \partial f^{k,i}_d(\bar{x}^{k,i}) + \{ \chi_k (\bar{x}^{k,i} - x^{k,i-1}) \}.
\]

Moreover, one obtains

\[
\partial f^{k,i}(\bar{x}^{k,i}) = \partial f(\bar{x}^{k,i}) + \mu_k \frac{1}{- \max_{t \in T} g(\bar{x}^{k,i}, t)} \partial \left( \max_{t \in T} g(\bar{x}^{k,i}, t) \right),
\]

hence

\[
\partial F^{k,i}_d(\bar{x}^{k,i}) =
\]

\[
= \partial f(\bar{x}^{k,i}) + \mu_k \frac{1}{- \max_{t \in T} g(\bar{x}^{k,i}, t)} \partial \left( \max_{t \in T} g(\bar{x}^{k,i}, t) \right) + \{ \chi_k (\bar{x}^{k,i} - x^{k,i-1}) \},
\]
i.e. there exist \( u_f \in \partial f(\bar{x}^{k,i}) \) and \( u_g \in \partial \left( \max_{t \in T} g(\bar{x}^{k,i}, t) \right) \) with

\[
u_f - \frac{\mu_k}{\max_{t \in T} g(\bar{x}^{k,i}, t)} u_g + \chi_k (\bar{x}^{k,i} - x^{k,i-1}) = 0
\]

and multiplication with \((\bar{x}^{k,i} - \bar{x}) \) leads to

\[
\langle u_f, \bar{x}^{k,i} - \bar{x} \rangle + \frac{\mu_k}{\max_{t \in T} g(\bar{x}^{k,i}, t)} \langle u_g, \bar{x}^{k,i} - \bar{x} \rangle + \chi_k (\bar{x}^{k,i} - x^{k,i-1}, \bar{x}^{k,i} - \bar{x}) = 0.
\]

Using the properties of the subgradients \( u_f \) and \( u_g \) as well as the convexity of the norm we obtain

\[
0 \geq f(\bar{x}^{k,i}) - f(\bar{x}) + \frac{\mu_k}{\max_{t \in T} g(\bar{x}^{k,i}, t)} \left( \max_{t \in T} g(\bar{x}^{k,i}, t) - \max_{t \in T} g(\bar{x}, t) \right)
+ \frac{\chi_k}{2} \left( \|\bar{x}^{k,i} - x^{k,i-1}\|^2 - \|\bar{x} - x^{k,i-1}\|^2 \right),
\]

so that we can conclude

\[
\frac{\mu_k}{\max_{t \in T} g(\bar{x}^{k,i}, t)} \left( \max_{t \in T} g(\bar{x}^{k,i}, t) - \max_{t \in T} g(\bar{x}, t) \right) \leq f(\bar{x}) - f^* + \frac{\chi_k}{2} \|\bar{x} - x^{k,i-1}\|^2
\]

and using \( \bar{x}, x^{k,i-1} \in \mathcal{B}_r(x_c) \) gives

\[
\frac{\mu_k}{\max_{t \in T} g(\bar{x}^{k,i}, t)} \left( \max_{t \in T} g(\bar{x}^{k,i}, t) - \max_{t \in T} g(\bar{x}, t) \right) \leq f(\bar{x}) - f^* + 2\tau^2 \chi.
\]

Now, regarding \( \mu_k \leq \mu_1 \), it is obvious that

\[
- \max_{t \in T} g(\bar{x}^{k,i}, t) \geq \mu_k \frac{- \max_{t \in T} g(\bar{x}, t)}{\mu_1 + f(\bar{x}) - f^* + 2\tau^2 \chi}
\]

holds and the proof is complete. \(\square\)

Denote

\[
\Delta_{k,i} := f(\bar{x}^{k,i}) - f^*
\]

for each \( k \) and \( 1 \leq i \leq i(k) \). In order to complete this definition for \( i = 0 \) we set \( \bar{x}^{k+1,0} := \bar{x}^{k,i(k)} \) for each \( k = 1, 2, \ldots \) as well as \( \bar{x}^{1,0} := x^{1,0} = x^0 \) such that \( \Delta_{k,0} \) can be defined as \( \Delta_{k,i} \) above.

**6.4.12 Theorem.** Let the assumptions of Theorem 6.4.7 be satisfied. Moreover, assume that

\[
\mu_1 \leq -\frac{1}{c_4} \max_{t \in T} g(x^0)
\]

with \( c_4 \) given as in Lemma 6.4.11. Additionally let be given a positive constant \( \alpha \) with

\[
\alpha \leq \frac{1}{16 \chi^2 \tau^2}, \quad \alpha \sup_{k,i} \Delta_{k,i} \leq \frac{7}{32}
\]

and assume that for each \( k \) constant

\[
\zeta_k := \mu_k (c_1 + \ln \tau) - \mu_k \ln \left( \frac{1}{8 c_4 \mu_k} \right) + \frac{\chi_k}{2} \alpha \tau^2 + \frac{7}{4} \chi \alpha \tau^2
\]

220 \text{ CHAPTER 6. PPR FOR CONVEX SEMI-INFINITE PROBLEMS}
satisfies
\[ \zeta_k \leq \alpha \left( \frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{k-1} i^1(j) \Delta_{1,0}} \right)^2, \]
(6.4.54)
\[ 2\zeta_k \leq \frac{\chi_k}{2} (\sigma_k - \alpha_k)^2 \]
(6.4.55)

with \( i^1(j) := \max\{1, 2i(j) - 2\} \).

Then the estimate
\[ \Delta_{k,i} \leq \frac{\Delta_{1,0}}{1 + \alpha \left( 2i + \sum_{j=1}^{k-1} i^1(j) \right) \Delta_{1,0}} \]
(6.4.56)
is true for \( k = 1, 2, \ldots, 0 \leq i < i(k) \) if \( \Delta_{1,0} > 0 \) or \( x^0 \notin \mathcal{M}_{\text{opt}} \).

**Proof:** Denote \( \tilde{z}^{k,i} := \arg\min_{z \in \mathcal{M}_{\text{opt}} \cap \mathbb{B}_r(x_c)} \| z^{k,i} - z \| \) for each \( k \) and \( 0 \leq i \leq i(k) \). Then \( \tilde{z}^{k,i}(\lambda) := \lambda z^{k,i} + (1 - \lambda) z^{k,i}_{\text{opt}} \in \mathcal{M}_0 \cap \mathbb{B}_r(x_c) \) for all \( \lambda \in [0, 1) \) and \( k, 0 \leq i \leq i(k) \) and we have
\[ F_{k,i+1}(\tilde{z}^{k,i+1}) \leq F_{k,i+1}(\tilde{z}^{k,i}(\lambda)) \]
for \( \lambda \in [0, 1) \) if \( i < i(k) \). Taking into account the convexity of \( f \) as well as (6.4.30) we obtain
\[ F_{k,i+1}(\tilde{z}^{k,i+1}) = f(\tilde{z}^{k,i+1}) - \mu_k \ln \left( -\max_{t \in T} g(\tilde{z}^{k,i+1}, t) \right) + \frac{\chi_k}{2} \| \tilde{z}^{k,i+1} - x^{k,i} \|^2 \]
\[ \leq F_{k,i+1}(\tilde{z}^{k,i}(\lambda)) \]
\[ \leq \lambda f(\tilde{z}^{k,i}) + (1 - \lambda)f(\tilde{z}^{k,i}) - \mu_k \ln \left( -\max_{t \in T} g(\tilde{z}^{k,i}(\lambda), t) \right) \]
\[ + \frac{\chi_k}{2} (\lambda \| \tilde{z}^{k,i} - x^{k,i} \| + \alpha_k)^2. \]
(6.4.57)

Using the convexity of \( \max_{t \in T} g(\cdot, t) \) and the inclusion \( \tilde{z}^{k,i} \in \mathcal{M} \), we have
\[ \max_{t \in T} g(\tilde{z}^{k,i}(\lambda), t) \leq (1 - \lambda) \max_{t \in T} g(\tilde{z}^{k,i}, t) \]
such that we infer
\[ -\mu_k \ln \left( -\max_{t \in T} g(\tilde{z}^{k,i}(\lambda), t) \right) \leq -\mu_k \ln \left( -(1 - \lambda) \max_{t \in T} g(\tilde{z}^{k,i}, t) \right) \]
for all pairs \((k, i)\) with \( 1 \leq i \leq i(k) \) and \( \lambda \in [0, 1) \). Applying Lemma 6.4.11 it follows
\[ -\mu_k \ln \left( -\max_{t \in T} g(\tilde{z}^{k,i}(\lambda), t) \right) \leq -\mu_k \ln ((1 - \lambda)c_4 \mu_k) \]
(6.4.58)
for all pairs \((k, i)\) with \( 1 \leq i \leq i(k) \). But \( x^{k,0} = x^{k-1,i(k-1)} \) and \( i(k-1) > 0 \) for each \( k > 1 \) so that
\[ -\mu_k \ln \left( -\max_{t \in T} g(\tilde{z}^{i,0}(\lambda), t) \right) \leq -\mu_k \ln ((1 - \lambda)c_4 \mu_{i-1}) \]
follows by Lemma 6.4.11. The monotonic decrease of \( \{\mu_k\} \) leads to (6.4.58) again and regarding (6.4.51) the estimate now holds for all \( k \) and \( 0 \leq i \leq i(k) \).

From the proof of Theorem 6.4.7 we know that \( \bar{x}^{k,i+1} \in \mathcal{M}_0 \cap B_r(x_c) \) for all \( k \) and \( 0 \leq i < i(k) \). Using (6.4.26) we therefore conclude

\[
-\mu_k \ln \left( -\max_{t \in \mathcal{T}} g(\bar{x}^{k,i+1}, t) \right) \geq -\mu_k(c_1 + \ln \tau).
\]

This together with (6.4.57), (6.4.58) and \( \bar{x}^{k,i} \in \mathcal{M}_{opt} \) yields

\[
\Delta_{k,i+1} = f(\bar{x}^{k,i+1}) - f^*
\]

\[
\leq \lambda(f(\bar{x}^{k,i}) - f^*) + (1 - \lambda)(f(\bar{x}^{k,i}) - f^*)
\]

\[
- \mu_k \ln \left( -\max_{t \in \mathcal{T}} g(\bar{x}^{k,i}, t) \right) + \mu_k \ln \left( -\max_{t \in \mathcal{T}} g(\bar{x}^{k,i+1}, t) \right)
\]

\[
+ \frac{\chi_k}{2} (\lambda \|\bar{x}^{k,i} - \bar{x}^{k,i}\| + \alpha_k)^2 - \frac{\chi_k}{2} \|\bar{x}^{k,i+1} - \bar{x}^{k,i}\|^2
\]

\[
\leq (1 - \lambda) \Delta_{k,i} - \mu_k \ln ((1 - \lambda)c_4 \mu_k) + \mu_k(c_1 + \ln \tau)
\]

\[
+ \frac{\chi_k}{2} \lambda^2 \|\bar{x}^{k,i} - \bar{x}^{k,i}\|^2 + \chi_k \alpha_k \lambda \|\bar{x}^{k,i} - \bar{x}^{k,i}\| + \frac{\chi_k}{2} \alpha_k^2
\]

\[
- \frac{\chi_k}{2} \|\bar{x}^{k,i+1} - \bar{x}^{k,i}\|^2
\]

for all \( k \) and \( 0 \leq i < i(k) \). In view of \( \bar{x}^{k,i}, \bar{x}^{k,i} \in B_r(x_c) \) as well as the definition of \( \zeta_k \) we obtain

\[
0 \leq \Delta_{k,i+1} \leq (1 - \lambda) \Delta_{k,i} + \zeta_k + \frac{\chi_k}{2} \lambda^2 \|\bar{x}^{k,i} - \bar{x}^{k,i}\|^2 - \frac{\chi_k}{2} \|\bar{x}^{k,i+1} - \bar{x}^{k,i}\|^2,
\]

(6.4.60)

If \( \lambda \in [0, 7/8] \), which can be enforced by setting

\[
\lambda := \lambda_{k,i} = \min \left[ \frac{-\Delta_{k,i}}{\chi_k \|\bar{x}^{k,i} - x_c\|^2}, \frac{7}{8} \right].
\]

(6.4.61)

If \( \lambda_{k,i} \) equals 7/8 (6.4.61) immediately leads to

\[
\chi_k \|\bar{x}^{k,i} - \bar{x}^{k,i}\|^2 \leq \frac{8}{7} \Delta_{k,i}
\]

and we can infer from (6.4.60)

\[
\Delta_{k,i+1} \leq \frac{\Delta_{k,i}}{8} + \zeta_k + \frac{7}{16} \Delta_{k,i} - \frac{\chi_k}{2} \|\bar{x}^{k,i+1} - \bar{x}^{k,i}\|^2
\]

\[
= \frac{9}{16} \Delta_{k,i} + \zeta_k - \frac{\chi_k}{2} \|\bar{x}^{k,i+1} - \bar{x}^{k,i}\|^2,
\]

(6.4.62)

Due to the second part of (6.4.52) this allows us to conclude

\[
\Delta_{k,i+1} \leq \Delta_{k,i} - 2\alpha \Delta_{k,i}^2 + \zeta_k - \frac{\chi_k}{2} \|\bar{x}^{k,i+1} - \bar{x}^{k,i}\|^2.
\]

(6.4.63)

But if \( \lambda_{k,i} < 7/8 \) holds we obtain

\[
\Delta_{k,i+1} \leq \Delta_{k,i} - \frac{\Delta_{k,i}^2}{2\chi_k \|\bar{x}^{k,i} - x_c\|^2} + \zeta_k - \frac{\chi_k}{2} \|\bar{x}^{k,i+1} - \bar{x}^{k,i}\|^2
\]

(6.4.64)
and now using the first part of (6.4.52) inequality (6.4.63) follows again. Consequently this estimate holds for all pairs \((k,i)\) with \(0 \leq i < i(k)\) making it the basis of the following induction proof.

Let us assume that (6.4.56) holds for a fixed pair \(k,i\) with \(i < i(k)\). This is obvious for the starting indices \(k = 1, i = 0\). Now we distinguish three cases.

(a) We first consider \(0 \leq i < i(k) - 1\). Then \(i + 1 < i(k)\) and due to (6.4.22) as well as (6.4.37) we have

\[
\chi_k \frac{\|x^{k,i+1} - x^{k,i}\|_2}{2} > \frac{\chi_k}{2} (\sigma_k - \alpha_k)^2
\]

such that (6.4.55) leads to

\[
\zeta_k - \chi_k \frac{\|x^{k,i+1} - x^{k,i}\|_2}{2} < 0.
\]

Consequently, with (6.4.63) we obtain

\[
\Delta_{k,i+1} \leq \Delta_{k,i} - 2\alpha \Delta_{k,i}^2.
\]

Moreover, the trivial inequality

\[
y - \vartheta y^2 \leq \frac{y}{1 + \vartheta y}
\]

is true for all \(y \geq 0\) with fixed \(\vartheta > 0\) and the function \(\frac{y}{1 + \vartheta y}\) increases monotonically for nonnegative \(y\) such that one can set \(\vartheta := 2\alpha, y := \Delta_{k,i}\). Then, with regard to the induction assumption, we infer

\[
\Delta_{k,i+1} \leq \frac{\Delta_{k,i}}{1 + 2\alpha \Delta_{k,i}} \leq \frac{\Delta_{1,0}}{1 + \alpha \left(2i + \sum_{j=1}^{k-1} i^1(j)\right) \Delta_{1,0}} \leq \frac{\Delta_{1,0}}{1 + \alpha \left(2i + 2 + \sum_{j=1}^{k-1} i^1(j)\right) \Delta_{1,0}}
\]

and the induction statement holds.

(b) In case \(i = 0, i(k) = 1\) we have \(i^1(k) = 1\) and (6.4.63) leads to

\[
\Delta_{k+1,0} = \Delta_{k,1} \leq \Delta_{k,0} - 2\alpha \Delta_{k,0}^2 + \zeta_k.
\]

Furthermore, regarding the second part of (6.4.52), it holds

\[
\frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{k-1} i^1(j) \Delta_{1,0}} \leq \Delta_{1,0} \leq \frac{7}{32\alpha} < \frac{1}{4\alpha}.
\]

Because the function \(y - 2\alpha y^2\) increases monotonically if \(y < \frac{1}{4\alpha}\), the induction assumption allows to conclude

\[
\Delta_{i+1,0} \leq \frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{k-1} i^1(j) \Delta_{1,0}} - 2\alpha \left(\frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{k-1} i^1(j) \Delta_{1,0}}\right)^2 + \zeta_k
\]
such that
\[ \Delta_{i+1,0} \leq \frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{i} i^1(j) \Delta_{1,0}} - \alpha \left( \frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{i} i^1(j) \Delta_{1,0}} \right)^2 \] (6.4.67)

follows from (6.4.54). Using (6.4.66) again - this time with \( \theta := \alpha \), \( y := \frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{i} i^1(j) \Delta_{1,0}} \) - the combination with (6.4.67) and \( i^1(k) = 1 \) leads to
\[
\Delta_{k+1,0} \leq \frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{k-1} i^1(j) \Delta_{1,0}} \left( 1 + \alpha \frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{k-1} i^1(j) \Delta_{1,0}} \right)
\]

\[
= \frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{k-1} i^1(j) \Delta_{1,0}} + \alpha \Delta_{1,0}
\]

\[
= \frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{k} i^1(j) \Delta_{1,0}}
\]
such that the induction statement holds.

(c) Finally let us consider the case \( i = i(k) - 1 \) with \( i(k) > 1 \). Then from (6.4.63) we have the inequalities
\[
\Delta_{k,i(k)-1} \leq \Delta_{k,i(k)-2} - 2\alpha \Delta_{k,i(k)-2}^2 + \zeta_k - \frac{\chi_k}{2} \| x^{k,i(k)-1} - x^{k,i(k)-2} \|^2_2,
\]
\[
\Delta_{k+1,0} = \Delta_{k,i(k)} \leq \Delta_{k,i(k)-1} - 2\alpha \Delta_{k,i(k)-1}^2 + \zeta_k.
\]

Coupling these we infer
\[
\Delta_{k+1,0} \leq \Delta_{k,i(k)-2} - 2\alpha \Delta_{k,i(k)-2}^2 - 2\alpha \Delta_{k,i(k)-1}^2 + 2\zeta_k - \frac{\chi_k}{2} \| x^{k,i(k)-1} - x^{k,i(k)-2} \|^2_2,
\]
so that together with (6.4.55), (6.4.65) and \( \alpha \Delta_{k,i(k)-1}^2 \geq 0 \) the estimate
\[
\Delta_{k,i(k)} < \Delta_{k,i(k)-2} - 2\alpha \Delta_{k,i(k)-2}^2
\]
holds. Thus we can conclude analogously to case (a) that
\[
\Delta_{k+1,0} < \frac{\Delta_{1,0}}{1 + \alpha \left( 2\alpha k - 2 + \alpha \sum_{j=1}^{k-1} i^1(j) \right) \Delta_{1,0}} \leq \frac{\Delta_{1,0}}{1 + \alpha \sum_{j=1}^{k} i^1(j) \Delta_{1,0}}
\]
since \( 2\alpha k - 2 > 1 \) and the induction is complete. \( \square \)

6.4.2.2 Linear convergence

Theorem 6.4.12 establishes the important estimate (6.4.56) which holds for any SIP satisfying Assumption 6.4.3. If we consider problems adhering to tighter assumptions it is possible to prove linear convergence for the iterates as well as the values of the objective function. The condition which is to use in this case is the following growth condition
\[
\inf_{x \in \mathcal{M}} \frac{f(x) - f^*}{\rho^2(x, \mathcal{M}_{opt})} \geq \theta > 0
\] (6.4.68)
with
\[ \tilde{M} := (M_0 \cap B_\tau(x_c)) \setminus M_{\text{opt}}, \quad \rho(x, M_{\text{opt}}) := \min_{z \in M_{\text{opt}} \cap B_\tau(x_c)} \|x - z\|. \]

This growth condition generalizes that of Rockafellar [352] which occurs in the context for proving linear convergence of the iterates of an inexact proximal point method.

If \( \frac{\theta}{\lambda_k} < \frac{7}{8} \) is true, (6.4.61) admits \( \lambda_{k,i} = \frac{7}{8} \) as well as \( \lambda_{k,i} < \frac{7}{8} \) for all \( i \) with \( 0 \leq i < i(k) \). If \( \lambda_{k,i} = \frac{7}{8} \), inequality (6.4.62) is true, while in the case \( \lambda_{k,i} < \frac{7}{8} \) estimate (6.4.64) follows. Then we have
\[ \lambda_{k,i} = \frac{\Delta_{k,i}}{\chi_k \| z^{k,i} - x^{k,i} \|^2} \]
and one can conclude
\[ \Delta_{k,i+1} \leq \left( 1 - \frac{\theta}{2\chi_k} \right) \Delta_{k,i} + \zeta_k - \frac{\chi_k}{2} \left\| z^{k,i+1} - x^{k,i} \right\|^2 \]
with regard to (6.4.68) and \( \chi_k \leq \tilde{\chi} \).

But if \( \frac{\theta}{\lambda_k} \geq \frac{7}{8} \) is true, (6.4.61) only admits \( \lambda_{k,i} = \frac{7}{8} \) for all \( i \) with \( 0 \leq i < i(k) \) which immediately leads to (6.4.62).

Thus \( \Delta_{k,i}, \Delta_{k,i+1} \) always fulfill
\[ 0 \leq \Delta_{k,i+1} \leq (1 - \theta_1) \Delta_{k,i} + \zeta_k - \frac{\chi_k}{2} \left\| z^{k,i+1} - x^{k,i} \right\|^2, \tag{6.4.69} \]
if \( i < i(k) \), where \( \theta_1 = \min \left\{ \frac{7}{16}, \frac{\theta}{2\chi_k} \right\} \).

Using these preliminary remarks a linear convergence of the sequence \( \{\Delta_{k,i}\} \) can be established under the given growth condition.

**6.4.13 Theorem.** Let the assumptions of Theorem 6.4.7 be satisfied. Moreover, let (6.4.51) as well as (6.4.68) be satisfied. Additionally assume that
\[ \zeta_k \leq \frac{\chi_k(1 - \theta_1)}{2(2 - \theta_1)} \left( \sigma_k - \alpha_k \right)^2, \quad \zeta_k \leq \frac{\theta_1}{2} \Delta_{1,0} q^s \]
with \( s_k = \sum_{j=1}^{k-1} i(j) \), \( q \in [1 - \frac{\theta_1}{2}, 1) \). Then
\[ \Delta_{k,i} \leq \Delta_{1,0} q^{s_k+i} \]
holds.

**Proof:** The proof is by induction again. The statement is for \( k = 1, i = 0 \).

Thus we suppose that a fixed \( k \) and \( i < i(k) \) are given such that
\[ \Delta_{\ell,i} \leq \Delta_{1,0} q^{s_k+i} \]
holds for all \( \ell < k, 0 \leq i' < i(\ell) \) and \( \ell = k, 0 \leq i' \leq i \). The condition \( i < i(k) \) is not a restriction because in case \( i = i(k) \) we consider the equivalent pair \( k+1, i = 0 < i(k+1) \).

We distinguish three cases.
(a) Suppose $i + 1 < i(k)$. Combining $\|\bar{x}^{k,i+1} - x^{k,i}\| > \sigma_k - \alpha_k$, the first inequality in (6.4.70) and (6.4.69), we obtain

$$\Delta_{k,i+1} < (1 - \theta_1)\Delta_{k,i}$$

and along (6.4.72) this implies

$$\Delta_{k,i+1} < \Delta_{1,0}q^{\bar{i} + i + 1}.$$

b) Suppose $i > 0, i + 1 = i(k)$. Then the inequalities

$$\Delta_{k,i+1} \leq \Delta_{k,i}(1 - \theta_1) + \zeta_k - \frac{\chi_k}{2} \|\bar{x}^{k,i+1} - x^{k,i}\|^2,$$

$$\Delta_{k,i} \leq \Delta_{k,i-1}(1 - \theta_1) + \zeta_k - \frac{\chi_k}{2} \|\bar{x}^{k,i} - x^{k,i-1}\|^2$$

and $\|\bar{x}^{k,i} - x^{k,i-1}\| > \sigma_k - \alpha_k$ hold. Substituting the second inequality into the first one gives

$$\Delta_{k,i+1} < \Delta_{k,i-1}(1 - \theta_1)^2 + \zeta_k(2 - \theta_1) - (1 - \theta_1)\frac{\chi_k}{2}(\sigma_k - \alpha_k)^2,$$

leading to

$$\Delta_{k,i+1} < (1 - \theta_1)^2\Delta_{k,i-1}$$

if we consider the first inequality in (6.4.70). Hence,

$$\Delta_{k,i+1} < \Delta_{1,0}q^{\bar{i} + i + 1}$$

and the induction statement holds.

c) Suppose $i = 0, i(k) = 1$. Taking (6.4.72) and the second inequality in (6.4.70) into account we obtain

$$\Delta_{k,1} \leq \Delta_{1,0}q^{\bar{i}}(1 - \theta_1) + \frac{\theta_1}{2}\Delta_{1,0}q^{\bar{i}} < \Delta_{1,0}q^{\bar{i} + 1}$$

from (6.4.69) and the proof is complete. □

If the considered problem fulfills the growth condition (6.4.68) we can additionally prove the linear convergence of the sequence $\{\bar{x}^{k,i}\}$ to an element of $M_{\text{opt}}$.

For this purpose we define $\bar{i}(k) := \frac{16\tau^2}{(\sigma_k - \alpha_k)^2} + 1$ and

$$\varsigma_k := \sum_{j=k}^{\infty} \left(\sqrt{\gamma_j} + \alpha_j + 4\tau\mu_j\right), \quad \forall k,$$

(6.4.73)

with $\gamma_j := \frac{2\mu_j}{\chi_j}(2 |\ln \mu_j| + |\ln \tau|)$ as in the proof of Theorem 6.4.7.

6.4.14 Theorem. Let the assumptions of Theorem 6.4.13 be satisfied. Moreover, assume that

$$\frac{1}{4\tau} \gamma_k + \alpha_k < \frac{1}{8\tau}(\sigma_k - \alpha_k)^2,$$

(6.4.74)

$$\varsigma_k \leq \sqrt{\frac{\Delta_{1,0}}{\theta}} \sqrt{q^{\bar{i} + \bar{i}(k)}}, \quad \forall k.$$

(6.4.75)
Then it holds
\[
\|x^{k,i} - x^*\| \leq 3 \sqrt{\frac{\Delta_{1.0}}{d}} \sqrt{q_{k+1}}, \quad \forall \ k, \ 0 \leq i < i(k),
\] (6.4.76)
where \(x^* := \lim_{k \to \infty} x^{k,0}\) is an optimal solution of SIP (6.4.3).

**Proof:** Let \(z^{k,i} := \arg \min_{z \in \mathcal{M}_{opt} \cap B(x_c)} \|x^{k,i} - z\|\) be given as in the proof of Theorem 6.4.12.

The inequality
\[
\|x^{k,i} - z^{k,i}\| \leq \sqrt{\frac{\Delta_{k,i}}{\theta}}
\] (6.4.77)
is obviously true if \(z^{k,i} \in \mathcal{M}_{opt}\), otherwise (6.4.77) holds due to (6.4.68).

In the sequel let \(k_0, i_0\) be fixed with \(0 \leq i_0 < i(k_0)\). From the proof of Theorem 6.4.7 inequality (6.4.38) is known, i.e.
\[
\|x^{k,i+1} - z^k\|^2 - \|x^{k,i} - z^k\|^2 \leq -((\sigma_k - \alpha_k)^2 + \gamma_k)
\]
holds for all \(k, 0 \leq i < i(k) - 1\) with \(z^k := \bar{x} + (1 - \mu_k)(x^* - \bar{x})\). If we use \(\tilde{z}^k := \bar{x} + (1 - \mu_k)(z^{k_0,k_0} - \bar{x})\) instead of \(z^k\) we can conclude
\[
\|x^{k,i+1} - \tilde{z}^k\|^2 - \|x^{k,i} - \tilde{z}^k\|^2 \leq -((\sigma_k - \alpha_k)^2 + \gamma_k)
\]
analogously for all \(k\) and \(0 \leq i < i(k) - 1\). Furthermore it yields (cf. (6.4.40))
\[
\|x^{k,i+1} - z^k\| - \|x^{k,i} - z^k\| < \frac{1}{4\tau} ((\sigma_k - \alpha_k)^2 + \gamma_k) + \alpha_k < 0
\] (6.4.78)
and (cf. (6.4.41))
\[
\|x^{k,i(k)} - z^k\| - \|x^{k,i(k)-1} - z^k\| \leq \sqrt{\gamma_k + \alpha_k}
\] (6.4.79)
for all \(k\) and \(0 \leq i < i(k) - 1\). Summing up these inequalities we obtain
\[
\|x^{k,i(k)} - \tilde{z}^k\| - \|x^{k,0} - \tilde{z}^k\| \leq \sqrt{\gamma_k + \alpha_k},
\]
and together with \(\|z^k - z^{k_0,i_0}\| \leq 2\mu_k \tau\) and \(x^{k+1,0} = x^{k,i(k)}\) the estimate
\[
\|x^{k+1,0} - z^{k_0,i_0}\| - \|x^{k,0} - z^{k_0,i_0}\| \leq \sqrt{\gamma_k + \alpha_k} + 4\mu_k \tau
\]
follows. Summing these inequalities for \(j = k_0 + 1, \ldots, k' - 1\) with \(k' > k_0 + 1\) we get
\[
\|x^{k',0} - z^{k_0,i_0}\| \leq \|x^{k_0,i_0} - z^{k_0,i_0}\| + \sum_{j=k_0+1}^{k'-1} (\sqrt{\gamma_j} + \alpha_j + 4\mu_j \tau).
\]
In combination with (6.4.78) and (6.4.79) this leads to
\[
\|x^{k',0} - z^{k_0,i_0}\| \leq \|x^{k_0,i_0} - z^{k_0,i_0}\| + \sum_{j=k_0}^{k'-1} (\sqrt{\gamma_j} + \alpha_j + 4\mu_j \tau).
\]
Moreover, \( \lim_{k \to \infty} x^{k,0} = \lim_{k \to \infty} \bar{x}^{k,0} \) follows from (6.4.30) and \( \lim_{k \to \infty} \alpha_k = 0 \) is enforced by (6.4.21). Thus, taking limit \( k' \to \infty \), it allows us to conclude

\[
\|x^{**} - z_{k_0,i_0}\| \leq \|\bar{x}^{k_0,i_0} - z_{k_0,i_0}\| + \sum_{j=k_0}^{\infty} (\sqrt{\gamma_j} + \alpha_j + 4\mu_j \tau) = \|\bar{x}^{k_0,i_0} - z_{k_0,i_0}\| + \varsigma_{k_0},
\]

leading to

\[
\|x^{k_0,i_0} - x^{**}\| \leq \|\bar{x}^{k_0,i_0} - z_{k_0,i_0}\| + \|z_{k_0,i_0} - x^{**}\| \leq 2 \|\bar{x}^{k_0,i_0} - z_{k_0,i_0}\| + \varsigma_{k_0}
\]

and we obtain

\[
\|x^{k_0,i_0} - x^{**}\| \leq 2 \sqrt{\bar{\Delta}_{k_0,i_0} \theta} + \varsigma_{k_0} \tag{6.4.80}
\]

in combination with (6.4.77). We remember that estimate (6.4.47) gives

\[
i(k_0) < \frac{2r}{r^2 ((\sigma_{k_0} - \alpha_{k_0})^2 - \gamma_{k_0}) - \alpha_{k_0}} + 1
\]

so that \( i(k_0) < \bar{i}(k_0) \) follows in view of (6.4.74). Then relation (6.4.75) gives

\[
\varsigma_{k_0} \leq \sqrt{\frac{\Delta_{1,0}}{\theta} \sqrt{q^{k_{i_0} + i(k_0)}}}
\]

and along with (6.4.71) and (6.4.80) we can deduce

\[
\|x^{k_0,i_0} - x^{**}\| \leq 3 \sqrt{\frac{\Delta_{1,0}}{\theta} \sqrt{q^{k_{i_0} + i(k_0)}}}.
\]

Because \( k_0 \) and \( i_0 \) were chosen arbitrarily the statement is proved. \( \square \)

6.4.15 Corollary. Under the assumptions of Theorem 6.4.14 it holds

\[
\|x^{k,i} - x^{**}\| \leq 4 \sqrt{\frac{\Delta_{1,0}}{\theta} \sqrt{q^{k+i}}}, \quad \forall k, 1 \leq i \leq i(k),
\]

where \( x^{**} = \lim_{k \to \infty} x^{k,0} \) is an optimal solution of SIP (6.4.3).

Proof: Regarding (6.4.30) and (6.4.76) we see that

\[
\|x^{k,i} - x^{**}\| \leq 3 \sqrt{\frac{\Delta_{1,0}}{\theta} \sqrt{q^{k+i}}} + \alpha_k
\]

holds for all \( k \) and \( 1 \leq i < i(k) \). Moreover, using the definition of \( \varsigma_k \) in (6.4.73), (6.4.75) and \( 0 < q < 1 \), we infer

\[
\alpha_k \leq \sqrt{\frac{\Delta_{1,0}}{\theta} \sqrt{q^{k+i}}}, \quad \forall k, 1 \leq i \leq i(k).
\]

Combining both estimates then the proof is complete. \( \square \)
6.4. REGULARIZED INTERIOR POINT METHODS

6.4.3 Application to model examples

We present some numerical results computed by the proposed interior point methods. For that purpose the algorithms were implemented in the programming language C by using version 2.7.2.3 of the gcc-compiler on a Pentium III/800-computer with the operating system Suse Linux 6.2. The included linear programs are solved by the Simplex-method while the quadratic problems are solved by a finite algorithm of FLETCHER [115].

Before we have a closer look at the examples some general numerical considerations are necessary. The first question which raises in the loops in $s$ of the Algorithms 6.4.4 is how can we determine the positive radius $r_{i,j,k}$ such that the box $S_{k,i,s}^e$ is completely contained in $M_0$. The simplest way to find such a radius is a trial-and-error strategy, whereby only the edges of the considered box have to be checked for their feasibility. In particular this fact requires that we can decide whether a given point $x$ fulfills $g(x,t) < 0$ for all $t \in T$. Such a decision procedure can be very costly, especially if the exact evaluation of the constraint values is not possible. Therefore we offer another method in the following lemma.

In order to formulate it let $L_S^r$ denote a constant for a given nonempty set $S \subset \mathbb{R}^n$ with

$$\sup_{z \in S} \sup_{t \in T} \sup_{v \in \partial g(z,t)} \|v\|_1 \leq L_S^r$$

(6.4.81)

and the additional property that $S' \subset S$ implies $L^r_{S'} \leq L^r_S$. Furthermore, let us define

$$B_r(M) := \left\{ z \in \mathbb{R}^n : \min_{v \in M} \|z - v\|_\infty \leq r \right\}$$

for $r > 0$ and nonempty compact sets $M \subset \mathbb{R}^n$.

6.4.16 Lemma. Let $\hat{x} \in M_0$ and $\hat{r} > 0$ be given. Moreover, let $h \geq 0$ be given such that

$$-\max_{t \in T_h} g(\hat{x},t) - L_{\{\hat{x}\}}^1 h > 0$$

holds with $L_{\{\hat{x}\}}^1$ fulfilling (6.4.6), i.e.,

$$|g(\hat{x},t_1) - g(\hat{x},t_2)| \leq L_{\{\hat{x}\}}^1 \|t_1 - t_2\|, \quad \forall t_1, t_2 \in T.$$

Then the inclusion $B_r(\{\hat{x}\}) \subset M_0$ is valid if

$$0 < r < \min \left[ \hat{r}, \frac{\max_{t \in T_h} g(\hat{x},t) - L_{\{\hat{x}\}}^1 h}{L_{B_r(\{\hat{x}\})}} \right].$$

(6.4.82)

Proof: Let $z \notin M_0$ be given. One has to show $\|z - \hat{x}\|_\infty > r$. If $\|z - \hat{x}\|_\infty \geq \hat{r}$ this follows immediately. Thus in the sequel we assume that $\|z - \hat{x}\|_\infty < \hat{r}$ holds.

Let $t^* \in T(z)$ be given such that $g(z,t^*) = \max_{t \in T} g(z,t) \geq 0$. Then one can conclude

$$0 > \max_{t \in T_h} g(\hat{x},t) + L_{\{\hat{x}\}}^1 h \geq \max_{t \in T} g(\hat{x},t) \geq g(\hat{x},t^*) - g(z,t^*) \geq \langle v, \hat{x} - z \rangle$$

with $v \in \partial g(z,t^*)$. Due to $\|z - \hat{x}\|_\infty < \hat{r}$ the estimate

$$-\max_{t \in T_h} g(\hat{x},t) - L_{\{\hat{x}\}}^1 h \leq |v^T(\hat{x} - z)| \leq L_{B_r(\{\hat{x}\})}^r \|\hat{x} - z\|_\infty$$
follows. Hence, we obtain \( \| z - \hat{x} \|_\infty > r \) and the proof is complete. \( \square \)

6.4.17 Remark. In case of more than one inequality constraint (i.e. \( l > 1 \) in (6.4.3)) or in case of occurring linear equality constraints one has to replace \( M_0 \) by \( \{ x \in \mathbb{R}^n : g_\nu(x,t) < 0 \ (t \in T^\nu) \} \) for each inequality constraint in the proposition of the lemma. In this way a feasible radius for each inequality constraint can be separately determined by Lemma 6.4.16. Then the smallest value of these radii can be used for fixing the box. \( \diamond \)

After determination of the boxes we have to select values for \( h_{i,j,k} \). Normally they influence directly the costs of the maximization processes such that we want to choose them as large as possible. Upper bounds for \( h_{k,i,s} \) are given by (6.4.12). But in order to use these upper bounds the constant \( C_S \) fulfilling part (9) of the Assumption 6.4.3 is needed.

6.4.18 Lemma. Let the assumptions of Lemma 6.4.16 be fulfilled. Furthermore, let \( r > 0 \) be given such that (6.4.82) is valid. Then

\[
C_S := 1 - \max_{t \in T} g(\hat{x}, t) - L^T_{\{x\}} h - L^S r \quad (6.4.83)
\]

fulfills (6.4.7) with \( S := B_r(\{\hat{x}\}) \).

Proof: From Lemma 6.4.16 it follows that \( S \subset M_0 \). Let \( x \in S \), \( t^* \in T(x) \) and \( v(x,t^*) \in \partial g(x,t^*) \) be arbitrarily given. Then we infer due to the Lipschitz property of function \( g \)

\[
- \max_{t \in T} g(x,t) = -g(x,t^*) \\
\geq -g(\hat{x}, t^*) + \langle v(x,t^*), \hat{x} - x \rangle \\
\geq - \max_{t \in T} g(\hat{x}, t) - L^T_{\{x\}} h - L^S r \\
\geq - \max_{t \in T} g(\hat{x}, t) - L^T_{\{x\}} h - L^S r.
\]

Moreover, using (6.4.82), we deduce

\[- \max_{t \in T} g(\hat{x}, t) - L^T_{\{x\}} h - L^S r > 0\]

so that we have

\[
\left| \frac{1}{\max_{t \in T} g(x,t)} \right| = \frac{1}{- \max_{t \in T} g(x,t)} \\
\leq \frac{1}{- \max_{t \in T} g(\hat{x}, t) - L^T_{\{x\}} h - L^S r} = C_S.
\]

Consequently (6.4.7) holds since \( x \in S \) was chosen arbitrarily. \( \square \)

6.4.19 Remark. The monotonicity property of \( C_S \) is automatically given if \( L^S \) is chosen as in (6.4.81) and one computes \( C_S \) by (6.4.83). \( \diamond \)
6.4. REGULARIZED INTERIOR POINT METHODS

Now the sets $T_h \subset T$ are always determined as equidistant discretization of $T$ with step size $2h$. Furthermore, the radii of the considered boxes are always computed as $9/10$ of the maximal value allowed by the previous Lemma 6.4.16. But the values of $\hat{r}$ in the formula (6.4.82) have to be adapted to each example. Particularly they are adapted to each step of the chosen algorithm. Finally, the values of $C_S$ are always computed as suggested in (6.4.83).

Let us finish our general statements with a remark on the application of several convergence results stated before. Each of them says that the algorithms generate sequences which converge to an optimal solution (in case of Algorithm 6.4.4 or its generalization). Moreover, in each presented convergence theorem it is required that some positive sequences converge to zero. But, caused by the fact that we cannot generate complete sequences, in practice it is impossible to check these assumptions. Nevertheless, they are the basis of the practical parameter setting in the following sense: We choose the occurring finite values of each sequence which has to converge in such a way that they fulfil a geometric decrease condition.

Now let us start with a small academic example in two variables with unbounded solution set.

6.4.20 Example. Consider the problem

$$\text{minimize } f(x) := (x_1 - x_2)^2$$

$$\text{s.t. } g(x, t) := x_1 \cos t + x_2 \sin t - 1 \leq 0 \text{ for all } t \in T := [0, 1].$$

The feasible (filled area) and the solution set (dotted line) of this problem are illustrated in Figure 6.4.1 (both sets are shrunk to the presented clipped area).

The solution set is given by

$$\mathcal{M}_{opt} = \left\{ x \in \mathbb{R}^2 : x_1 = x_2 \leq \frac{1}{2} \sqrt{2} \right\}.$$

Thus we deal with an unbounded feasible and an unbounded solution set so that a non-regularized Algorithm cannot be used to solve the problem. Therefore we want to show that Assumption 6.4.3 is fulfilled except for (8). This is obvious if we regard our introductory remarks for (7) and (9). Part (8) is fulfilled with...
constants $L_S^t$ defined by

$$\max_{x \in S} \max_{t \in [0,1]} \left| \frac{\partial g}{\partial t}(x,t) \right| = \max \max_{x \in S} \max_{t \in [0,1]} \left| -x_1 \sin t + x_2 \cos t \right|$$

$$\leq \max_{x \in S} \|x\|_{\infty} \max_{t \in [0,1]} \left| \cos t - \sin t \right| = \max_{x \in S} \|x\|_{\infty} =: L_S^t.$$  

Furthermore, for computing the values $C_S$ and the radii, the constants $L_S^x$ are needed and can be given by

$$L_S^x := \max_{t \in [0,1]} \left\| \left( \cos t \sin t \right) \right\|_{1} = \sqrt{2}.$$  

Thus, Assumption 6.4.3 is completely fulfilled and we can use the regularized method for solving the problem.

Now we use Algorithm 6.4.4 for solving the given problem. Having in mind Theorem 6.4.7 we must specify a few constants. First we use the standard parameters contained in Table 6.4.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Start Value</th>
<th>Decreasing Factor</th>
<th>Lower Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon_k$</td>
<td>0.001</td>
<td>0.06</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_k$</td>
<td>0.001</td>
<td>0.06</td>
<td>-</td>
</tr>
<tr>
<td>$q_k$</td>
<td>0.999</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 6.4.1: Example 6.4.20 - standard parameter

Furthermore, we set $\chi_1 := 1$, $\chi_{k+1} := \max\{0.01, 0.2 \chi_k\}$, $\tau := 12$, $x_c := (-3.3, -1.7)$, $x^* := (-2.5, 2.5)$, $\tilde{x} := (0, 0)$ and $x^0 := (-5, 0)$ in order to fulfill $x^* \in M_{\text{opt}} \cap B_{\tau/8}(x_c)$, $\tilde{x} \in M_{0} \cap B_{\tau}(x_c)$ and $x^0 \in M_{0} \cap B_{\tau/4}(x_c)$. This leads to $\bar{c} = \|\tilde{x} - x^*\| = \sqrt{12.5}$, $f(\tilde{x}) = 0$, $f_- = 0$, $c_0 = |\ln(1)| = 0$, $\tilde{t} = 0$, $\tilde{v} = (1, 0)$ and $c_1 = \ln(1 + 2) = \ln(3)$ so that we have $c_3 = \ln(3)$ and $\mu_1 = 0.1 \leq e^{-c_3}$ can be used. Setting the lower bound of the barrier parameter to $10^{-6}$ and $\sigma_k$ as small as possible by (6.4.22) all assumptions of Theorem 6.4.7 are fulfilled and we obtained the iteration process given in Table 6.4.2.

<table>
<thead>
<tr>
<th>k,i</th>
<th>$x_k^1$</th>
<th>$x_k^2$</th>
<th>$d_2(x^k, M_{\text{opt}})$</th>
<th>#LP</th>
<th>#QP</th>
<th>#BP</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,1</td>
<td>-3.013535</td>
<td>-2.008399</td>
<td>7.11E-01</td>
<td>15</td>
<td>31</td>
<td>4</td>
<td>0.01</td>
</tr>
<tr>
<td>1,2</td>
<td>-2.622608</td>
<td>-2.414308</td>
<td>1.47E-01</td>
<td>11</td>
<td>33</td>
<td>2</td>
<td>0.02</td>
</tr>
<tr>
<td>2,1</td>
<td>-2.531565</td>
<td>-2.519300</td>
<td>8.67E-03</td>
<td>13</td>
<td>36</td>
<td>2</td>
<td>0.03</td>
</tr>
<tr>
<td>3,1</td>
<td>-2.529324</td>
<td>-2.528754</td>
<td>4.03E-04</td>
<td>12</td>
<td>24</td>
<td>2</td>
<td>0.04</td>
</tr>
<tr>
<td>4,1</td>
<td>-2.529324</td>
<td>-2.528754</td>
<td>4.03E-04</td>
<td>6</td>
<td>10</td>
<td>2</td>
<td>0.06</td>
</tr>
<tr>
<td>5,1</td>
<td>-2.525245</td>
<td>-2.524960</td>
<td>2.01E-04</td>
<td>10</td>
<td>33</td>
<td>2</td>
<td>0.12</td>
</tr>
<tr>
<td>6,1</td>
<td>-2.525539</td>
<td>-2.525535</td>
<td>2.69E-06</td>
<td>14</td>
<td>40</td>
<td>2</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Table 6.4.2: Example 6.4.20 - iterates computed by Algorithm 6.4.4

Remarkably there was not required any restart procedure for adapting the accuracy parameter since it turned out that the radius was always equal the maximal
possible value 0.9. Furthermore, the multi-step technique was in fact used in the first outer step.

\[\text{6.4.21 Remark. In ABBE [1] the same example was used in order to show that the non-regularized MSR-algorithm does not converge in the case of an unbounded solution set.}\]

\[\text{6.4.3.1 An application to the financial market}\]

We present the application of our algorithms to a problem occurring in the field of finance. Our goal is to approximate the yield curve of an underlying asset. For that purpose we follow the considerations by Tichatschke et al. [388] which are based on the model of VASICEK [409]. Let \(y(t)\) be a given yield curve fulfilling the stochastic differential equation
\[
dy = (\alpha + \beta y)dt + \sigma dZ,
\]
with a Brownian motion \(Z\) and parameters \(\alpha, \beta, \sigma\). Now this yield curve should be approximated on an interval \(T = [0, \bar{t}]\) by a function \(r\). Then \(r\) has to fulfil the initial value problem
\[
\dot{r} = \beta r + \alpha + \sigma w(t), \quad r(0) = r_0 \in [r, \bar{r}],
\]
\[
w \leq w(t) \leq \bar{w}, \quad t \in T,
\]
which can be derived from the stochastic differential equation stated above. Therein the Brownian motion \(Z\) is modeled by a piecewise continuous function \(w\) with bounds \(w, \bar{w}\). Additionally the initial value \(r_0\) is variable so far. The solution of (6.4.84) is given by
\[
r(t) = -\frac{\alpha}{\beta}(1 - e^{\beta t}) + r_0 e^{\beta t} + \sigma \int_0^t e^{\beta(t-\tau)}w(\tau)d\tau
\]
so that the approximation error can be minimized by solving the problem
\[
\min \max_{t \in T} |y(t) - r(t)|
\]
\[
s.t. \quad r_0 \in [r, \bar{r}],
\]
\[
w \leq w(t) \leq \bar{w}, \quad \text{for all } t \in T.
\]
Since the feasible set is mainly described by the space of piece-wise continuous functions
\[
\{w : w \leq w(t) \leq \bar{w}, t \in T\},
\]
we deal with an infinite problem. In order to simplify the model the available functions of \(w\) are chosen in the class of piecewise constant functions, i.e. we set \(w(t) := w_i\) for all \(t \in T_i, \quad i = 1, \ldots, N\), with \(0 := t_0 < t_1 < \ldots < t_{N-1} < t_N := \bar{t}\), \(T_i := [t_{i-1}, t_i]\) for \(i \in \{1, \ldots, N - 1\}\) and \(T_N := [t_{N-1}, t_N]\). Then the integral term in (6.4.85) becomes
\[
\sigma \int_0^t e^{\beta(t-\tau)}w(\tau)d\tau = -\sum_{j=1}^{i-1} B_j e^{\beta t_j} w_j - \frac{\sigma}{\beta} \left(1 - e^{\beta(t_{i-1})}\right) w_i
\]
for all \( t \in T_i \) with
\[
B_j := \frac{\sigma}{\beta} \left( e^{-\beta t_j} - e^{-\beta t_{j-1}} \right), \quad \forall j = 1, \ldots, N - 1.
\]

Consequently, using
\[
f_k(r_0, w, t) := -\frac{\alpha}{\beta} (1 - e^{\beta t}) + r_0 e^{\beta t} - \sum_{j=1}^{i-1} B_j e^{\beta t_j} w_j - \frac{\sigma}{\beta} \left( 1 - e^{\beta(t-t_{i-1})} \right) w_i
\]
for all \( t \in T_i \), (6.4.86) can be rewritten as
\[
\begin{align*}
\text{minimize} & \quad f(r_0, w, \vartheta) := \vartheta \\
\text{subject to} & \quad g_i(r_0, w, \vartheta, t) := |\hat{y}(t) - f_k(r_0, w, t)| - \vartheta \leq 0 \quad \text{for all } t \in T^i \ (i = 1, \ldots, N) \\
& \quad g_{N+1}(r_0, w, \vartheta) := \max \left\{ \max_{i=1, \ldots, N} \{ w_i - \bar{w}, \bar{w} - w_i \}, r_0 - \bar{r}, \bar{r} - r_0 \right\} \leq 0
\end{align*}
\]
with an approximation \( \hat{y} \) of \( y \) constructed by observable values. Thus we now deal with a linear SIP with \( N + 1 \) constraints. The number of constraints in (6.4.87) is much smaller than in the formulation in [388] which is caused by the fact that we can treat non-differentiable constraint functions. Consequently, we can use larger barrier parameter in order to expect similar accuracies.

We want to show that (a non-regularized variant of) Algorithm 6.4.4 can be used for solving (6.4.87) approximately. For that purpose we have to check the corresponding parts of Assumption 6.4.3. We first notice that some parts of this assumptions does not have to hold for the last constraint \( g_{N+1} \) since it does not depend on \( t \). However, in practice \( g_{N+1} \) is treated as constraint of type \( g(x, t) \leq 0, \ t \in T \) with single-valued \( T \). Additionally we have to know something more about \( \hat{y} \) if we want to show some parts of Assumption 6.4.3. Therefore we only consider the special case \( \hat{y} = y_i \) is constant on each interval \( T_i \).

We observe that the assumptions of the convexity of \( f \) and all \( g_i \) are fulfilled since \( r_0, w, \vartheta \) occur at most linearly in each constraint and the absolute values of linear functions as well as the maxima of finitely many linear functions are convex. Part (2) of Assumption 6.4.3 is not fulfilled since \( T^1, \ldots, T^{N-1} \) are not closed. But it is possible to consider the closures of all sets \( T^i \) and the continuous extensions of all \( g_i \) in (6.4.87) without changing the feasible set. Then the continuity of all constraints w.r.t. \( t \) is obvious. Furthermore, part (5) of Assumption 6.4.3 is fulfilled if \( w < \bar{w} \) and \( \vartheta < \bar{\vartheta} \) are true. Moreover, we observe that lower level sets of our considered semi-infinite problem are bounded since \( w, r_0 \) are bounded by the \( (N+1) \)-th constraint and \( \vartheta \) is bounded below by 0 and bounded above by the given level. Consequently, since we only deal with continuous functions, these level sets are compact. Thus, regarding that the given problem is feasible, the solution set \( M_{opt} \) has to be nonempty and compact. Part (7) is simply fulfilled if we choose equidistant finite grids for each constraint as it is done in the chapter before. Regarding Lemma 6.4.18 part (9) of Assumption 6.4.3 can be fulfilled while subgradients of \( f \) and \( g_i(\cdot, t) \) can be easily given if one regards that, excluding the absolute value or the maxima, the functions therein are differentiable. Thus it remains to determine the constants \( L^i_{\varphi, S} \) enforced by part (8) of our assumption and \( L^k_S \) for determining \( C_{i, S} \) and the
Regarding the differentiability of the function inside the absolute value in $g_i$, we can set

$$L_{t,S}^i := \max_{(r_0,w,\theta) \in S} \sup_{t \in T} \left| \alpha e^{\beta t} + r_0 \beta e^{\beta t} - \sum_{j=1}^{i-1} B_j \beta e^{\beta t} w_j + \sigma e^{\beta (t-t_{i-1})} u_i \right|$$

for all $i = 1, \ldots, N$. In addition to this

$$L_{t,S}^i := \max_{t \in T} \left( e^{\beta t} + \sum_{j=1}^{i-1} |B_j| e^{\beta t} + \frac{\sigma}{\beta} \left( 1 - e^{\beta (t-t_{i-1})} \right) \right) + 1$$

for all $i = 1, \ldots, N$ and $L_{N+1,S}^i := 1$ are used. Consequently Assumption 6.4.3 is completely fulfilled so that Algorithm 6.4.4 can be used for solving (6.4.87).

Demonstrating this we want to approximate the German stock index DAX in two time periods of each 30 days. The required data, consisting of the daily opening DAX prices, are given in Table 6.4.3. The first period represents a quite stable but slowly growing DAX while the second period covers a big fluctuation in a short time interval. In addition there are needed values for $\alpha, \beta$ and $\sigma$. We used the setting

$$\alpha := 0.0154, \quad \beta := -0.1779 \quad \text{and} \quad \sigma := 0.02,$$

which is derived from the observation of US interests for government bonds within the years 1964 to 1989, because there was no similar investigation of the German stock exchange. Moreover, the German stock exchange tends to follow the US stock exchange so that this choice is not too bad.

Due to the fact that we only consider two different scenarios we do not give a standard parameter setting. Rather we have a separate look at both situations. Nevertheless, there were some common settings. So in both situations the considered time period was uniformly mapped to the interval $[0,1]$ which implies that each trading day was represented by a subinterval of length $1/30$ of $[0,1]$. Additionally we set $q_k := 0.999$ in each case and, regarding Remark 6.4.17, the radii were computed by Lemma 6.4.16. In fact we used

$$r_{i,k} = 0.9 \min \left\{ \hat{r}, -\max_{t \in T_h} g(x^{i,k-1}, t) - L_{\{x^{i,k-1}\}}^i \right\}$$

(6.4.88)

to determine a radius for each constraint with $\hat{r} := \min\{1000, 2r_{k,i-1}\}$ if $k > 1$ and $\hat{r} := 1000$ if $k = 1$ for all constraints and $h := h_{\nu,k,i-1}$ if $k > 1$, $h := 0.001$ if $k = 1$ for $\nu = 1, \ldots, 30$. In consequence of this it was possible to compute $C_{\nu,S}^i$, by Lemma 6.4.18 for each constraint.
### Example DAX1

<table>
<thead>
<tr>
<th>Bounds</th>
<th>Trading Day</th>
<th>Price $y_1^t$</th>
<th>Trading Day</th>
<th>Price $y_1^t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w = -10^5$</td>
<td>04.01.1993</td>
<td>1533.06</td>
<td>05.03.1998</td>
<td>4642.79</td>
</tr>
<tr>
<td></td>
<td>05.01.1993</td>
<td>1547.99</td>
<td>06.03.1998</td>
<td>4686.24</td>
</tr>
<tr>
<td></td>
<td>06.01.1993</td>
<td>1560.27</td>
<td>09.03.1998</td>
<td>4775.83</td>
</tr>
<tr>
<td></td>
<td>07.01.1993</td>
<td>1546.33</td>
<td>10.03.1998</td>
<td>4807.92</td>
</tr>
<tr>
<td></td>
<td>08.01.1993</td>
<td>1540.56</td>
<td>11.03.1998</td>
<td>4855.22</td>
</tr>
<tr>
<td></td>
<td>11.01.1993</td>
<td>1526.66</td>
<td>12.03.1998</td>
<td>4822.78</td>
</tr>
<tr>
<td></td>
<td>12.01.1993</td>
<td>1527.33</td>
<td>13.03.1998</td>
<td>4863.44</td>
</tr>
<tr>
<td></td>
<td>13.01.1993</td>
<td>1529.61</td>
<td>16.03.1998</td>
<td>4891.85</td>
</tr>
<tr>
<td></td>
<td>14.01.1993</td>
<td>1521.03</td>
<td>17.03.1998</td>
<td>4932.42</td>
</tr>
<tr>
<td></td>
<td>15.01.1993</td>
<td>1542.91</td>
<td>18.03.1998</td>
<td>4936.17</td>
</tr>
<tr>
<td>$w = 10^5$</td>
<td>18.01.1993</td>
<td>1559.83</td>
<td>19.03.1998</td>
<td>4923.51</td>
</tr>
<tr>
<td></td>
<td>19.01.1993</td>
<td>1576.13</td>
<td>20.03.1998</td>
<td>4993.53</td>
</tr>
<tr>
<td></td>
<td>20.01.1993</td>
<td>1586.94</td>
<td>23.03.1998</td>
<td>5017.48</td>
</tr>
<tr>
<td></td>
<td>21.01.1993</td>
<td>1577.62</td>
<td>24.03.1998</td>
<td>5014.62</td>
</tr>
<tr>
<td>$r = 1000$</td>
<td>22.01.1993</td>
<td>1587.95</td>
<td>25.03.1998</td>
<td>5058.54</td>
</tr>
<tr>
<td></td>
<td>25.01.1993</td>
<td>1582.21</td>
<td>26.03.1998</td>
<td>5093.52</td>
</tr>
<tr>
<td>$r = 2000$</td>
<td>26.01.1993</td>
<td>1566.83</td>
<td>27.03.1998</td>
<td>5041.84</td>
</tr>
<tr>
<td></td>
<td>27.01.1993</td>
<td>1570.96</td>
<td>30.03.1998</td>
<td>5069.98</td>
</tr>
<tr>
<td></td>
<td>28.01.1993</td>
<td>1561.02</td>
<td>31.03.1998</td>
<td>5070.81</td>
</tr>
<tr>
<td></td>
<td>29.01.1993</td>
<td>1571.28</td>
<td>01.04.1998</td>
<td>5093.52</td>
</tr>
<tr>
<td></td>
<td>01.02.1993</td>
<td>1582.35</td>
<td>02.04.1998</td>
<td>5163.11</td>
</tr>
<tr>
<td></td>
<td>02.02.1993</td>
<td>1587.20</td>
<td>03.04.1998</td>
<td>5203.58</td>
</tr>
<tr>
<td></td>
<td>03.02.1993</td>
<td>1595.08</td>
<td>06.04.1998</td>
<td>5256.69</td>
</tr>
<tr>
<td></td>
<td>04.02.1993</td>
<td>1605.07</td>
<td>07.04.1998</td>
<td>5276.79</td>
</tr>
<tr>
<td></td>
<td>05.02.1993</td>
<td>1635.67</td>
<td>08.04.1998</td>
<td>5282.94</td>
</tr>
<tr>
<td></td>
<td>08.02.1993</td>
<td>1643.83</td>
<td>09.04.1998</td>
<td>5270.35</td>
</tr>
<tr>
<td></td>
<td>09.02.1993</td>
<td>1642.32</td>
<td>14.04.1998</td>
<td>5378.91</td>
</tr>
<tr>
<td></td>
<td>10.02.1993</td>
<td>1649.79</td>
<td>15.04.1998</td>
<td>5379.99</td>
</tr>
<tr>
<td></td>
<td>11.02.1993</td>
<td>1651.22</td>
<td>16.04.1998</td>
<td>5362.26</td>
</tr>
<tr>
<td></td>
<td>12.02.1993</td>
<td>1655.13</td>
<td>17.04.1998</td>
<td>5266.34</td>
</tr>
</tbody>
</table>

### Example DAX2

<table>
<thead>
<tr>
<th>Bounds</th>
<th>Trading Day</th>
<th>Price $y_2^t$</th>
<th>Trading Day</th>
<th>Price $y_2^t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w = -10^6$</td>
<td>04.01.1993</td>
<td>1533.06</td>
<td>05.03.1998</td>
<td>4642.79</td>
</tr>
<tr>
<td></td>
<td>05.01.1993</td>
<td>1547.99</td>
<td>06.03.1998</td>
<td>4686.24</td>
</tr>
<tr>
<td></td>
<td>06.01.1993</td>
<td>1560.27</td>
<td>09.03.1998</td>
<td>4775.83</td>
</tr>
<tr>
<td></td>
<td>07.01.1993</td>
<td>1546.33</td>
<td>10.03.1998</td>
<td>4807.92</td>
</tr>
<tr>
<td></td>
<td>08.01.1993</td>
<td>1540.56</td>
<td>11.03.1998</td>
<td>4855.22</td>
</tr>
<tr>
<td></td>
<td>11.01.1993</td>
<td>1526.66</td>
<td>12.03.1998</td>
<td>4822.78</td>
</tr>
<tr>
<td></td>
<td>12.01.1993</td>
<td>1527.33</td>
<td>13.03.1998</td>
<td>4863.44</td>
</tr>
<tr>
<td></td>
<td>13.01.1993</td>
<td>1529.61</td>
<td>16.03.1998</td>
<td>4891.85</td>
</tr>
<tr>
<td></td>
<td>14.01.1993</td>
<td>1521.03</td>
<td>17.03.1998</td>
<td>4932.42</td>
</tr>
<tr>
<td></td>
<td>15.01.1993</td>
<td>1542.91</td>
<td>18.03.1998</td>
<td>4936.17</td>
</tr>
<tr>
<td>$w = 10^6$</td>
<td>18.01.1993</td>
<td>1559.83</td>
<td>19.03.1998</td>
<td>4923.51</td>
</tr>
<tr>
<td></td>
<td>19.01.1993</td>
<td>1576.13</td>
<td>20.03.1998</td>
<td>4993.53</td>
</tr>
<tr>
<td></td>
<td>20.01.1993</td>
<td>1586.94</td>
<td>23.03.1998</td>
<td>5017.48</td>
</tr>
<tr>
<td></td>
<td>21.01.1993</td>
<td>1577.62</td>
<td>24.03.1998</td>
<td>5014.62</td>
</tr>
<tr>
<td>$r = 1000$</td>
<td>22.01.1993</td>
<td>1587.95</td>
<td>25.03.1998</td>
<td>5058.54</td>
</tr>
<tr>
<td></td>
<td>25.01.1993</td>
<td>1582.21</td>
<td>26.03.1998</td>
<td>5093.52</td>
</tr>
<tr>
<td>$r = 4000$</td>
<td>26.01.1993</td>
<td>1566.83</td>
<td>27.03.1998</td>
<td>5041.84</td>
</tr>
<tr>
<td></td>
<td>27.01.1993</td>
<td>1570.96</td>
<td>30.03.1998</td>
<td>5069.98</td>
</tr>
<tr>
<td></td>
<td>28.01.1993</td>
<td>1561.02</td>
<td>31.03.1998</td>
<td>5070.81</td>
</tr>
<tr>
<td></td>
<td>29.01.1993</td>
<td>1571.28</td>
<td>01.04.1998</td>
<td>5093.52</td>
</tr>
<tr>
<td></td>
<td>01.02.1993</td>
<td>1582.35</td>
<td>02.04.1998</td>
<td>5163.11</td>
</tr>
<tr>
<td></td>
<td>02.02.1993</td>
<td>1587.20</td>
<td>03.04.1998</td>
<td>5203.58</td>
</tr>
<tr>
<td></td>
<td>03.02.1993</td>
<td>1595.08</td>
<td>06.04.1998</td>
<td>5256.69</td>
</tr>
<tr>
<td></td>
<td>04.02.1993</td>
<td>1605.07</td>
<td>07.04.1998</td>
<td>5276.79</td>
</tr>
<tr>
<td></td>
<td>05.02.1993</td>
<td>1635.67</td>
<td>08.04.1998</td>
<td>5282.94</td>
</tr>
<tr>
<td></td>
<td>08.02.1993</td>
<td>1643.83</td>
<td>09.04.1998</td>
<td>5270.35</td>
</tr>
<tr>
<td></td>
<td>09.02.1993</td>
<td>1642.32</td>
<td>14.04.1998</td>
<td>5378.91</td>
</tr>
<tr>
<td></td>
<td>10.02.1993</td>
<td>1649.79</td>
<td>15.04.1998</td>
<td>5379.99</td>
</tr>
<tr>
<td></td>
<td>11.02.1993</td>
<td>1651.22</td>
<td>16.04.1998</td>
<td>5362.26</td>
</tr>
<tr>
<td></td>
<td>12.02.1993</td>
<td>1655.13</td>
<td>17.04.1998</td>
<td>5266.34</td>
</tr>
</tbody>
</table>

Table 6.4.3: DAX data
Figure 6.4.2: Example DAX1 - trajectory of the regularized solution

Figure 6.4.3: Example DAX1 - trajectory of the unregularized solution

Figure 6.4.4: Example DAX2 - trajectory of the regularized solution
Now considering the first example data DAX1 the starting point 
\[ r_0 := 1600, \quad w_1 = \ldots = w_{30} := 0, \quad \vartheta := 3000 \]
was used and we set 
\[
\epsilon_{1,0} := 0.001, \quad \epsilon_{k+1,0} := 0.7\epsilon_k, \\
\delta_1 := 10, \quad \delta_{k+1} := 0.7\delta_k, \\
\mu_1 := 10, \quad \mu_{k+1} := 0.8\mu_k.
\]
But as in all examples before restart procedures were applied to adapt automatically the accuracy parameter. Then the algorithm was stopped when the barrier parameter fell below 0.1. Although this stopping criterion seems to be very bad, Figure 6.4.2 shows that the tendency of the DAX curve is correctly reconstructed by our final approximate solution which can be found in Table 6.4.4. Furthermore the final approximation error of 15.65 is better than 16.78 achieved in [411] and close to the correct minimal value 15.30, which is, due to the theory of continuous Chebyshev approximation, the half of the maximal gap between two successive observed DAX values. As a comparison in Figure 6.4.3 the unregularized solution for DAX1 is depicted.

For the second example data DAX2 we used 
\[ r_0 := 5000, \quad w_1 = \ldots = w_{30} := 30000, \quad \vartheta := 5000 \]
as starting point, 
\[
\epsilon_{1,0} := 0.005, \quad \epsilon_{k+1,0} := 0.6\epsilon_k, \\
\delta_1 := 50, \quad \delta_{k+1} := 0.6\delta_k, \\
\mu_1 := 100, \quad \mu_{k+1} := 0.7\mu_k
\]
and, again, the restart procedure. The stopping criterion was now fulfilled if the barrier parameter reached 0.01. The resulting approximate solution is stated in Table 6.4.7, while our final approximation error given by 54.30 is comparable with the result achieved in [411]. The optimal value is 54.28 so that the more accurate stopping criterion leads also to a more accurate final solution in comparison to the first situation. Figure 6.4.4 shows again that the complete curve is correctly reconstructed, but it can be observed as in the first case that there is not detected each particular fluctuation.

In the following tables more technical details about the computations are listed.
239

Table 6.4.4: Example DAX1: Computed optimal solution

<table>
<thead>
<tr>
<th>n</th>
<th>start vector</th>
<th>approximate solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>$r_0 = 1600$</td>
<td>$w_1 = 36242.75$</td>
</tr>
<tr>
<td></td>
<td>$w_1 = \ldots = w_{30} = 0$</td>
<td>$w_2 = 32815.99$</td>
</tr>
<tr>
<td></td>
<td>$\delta = 3000$</td>
<td>$w_3 = 13815.77$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_4 = 4020.24$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_5 = -6069.15$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_6 = 4758.85$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_7 = 14737.18$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_8 = 14112.17$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_9 = 18346.01$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{10} = 46544.21$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{11} = 33868.20$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{12} = 40881.96$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{13} = 14098.55$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{14} = 14098.55$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{15} = 15506.86$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{16} = -3195.49$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{17} = 7344.09$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{18} = 7793.20$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{19} = 13793.60$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{20} = 30348.69$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{21} = 30674.80$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{22} = 19024.95$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{23} = 26004.27$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{24} = 46345.79$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{25} = 49706.85$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{26} = 14621.09$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{27} = 17968.50$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{28} = 20163.07$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{29} = 19863.05$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$w_{30} = 14764.52$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\delta = 15.65$</td>
</tr>
</tbody>
</table>

Table 6.4.5: Example DAX1: Numerical effort

<table>
<thead>
<tr>
<th># of restarts of inner loops</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td># of solved linear programs</td>
<td>946</td>
</tr>
<tr>
<td># of solved quadratic programs</td>
<td>39337</td>
</tr>
<tr>
<td># of constrained boxes</td>
<td>7508</td>
</tr>
<tr>
<td>$t_{LP}$: time in seconds for solving all LP</td>
<td>633.58</td>
</tr>
<tr>
<td>$t_{QP}$: time in seconds for solving all QP</td>
<td>137.58</td>
</tr>
<tr>
<td>$T_{tot}$: total time in seconds for complete iteration process</td>
<td>1429.63</td>
</tr>
</tbody>
</table>

Table 6.4.6: Example DAX1: Final parameter values

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>1.15E-01</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>1.59E-01</td>
</tr>
<tr>
<td>$h_{min}$</td>
<td>2.13E-05</td>
</tr>
<tr>
<td>$</td>
<td>\hat{T}_h</td>
</tr>
</tbody>
</table>
CHAPTER 6. PPR FOR CONVEX SEMI-INFINITE PROBLEMS

Table 6.4.7: Example DAX2: Computed optimal solution

<table>
<thead>
<tr>
<th>n</th>
<th>start vector</th>
<th>approximate solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>( r_0 = 5000 ) &lt;br&gt; ( w_1 = \cdots = w_{30} = 10000 ) &lt;br&gt; ( \delta = 5000 )</td>
<td>( r_0 = 4642.79 ) &lt;br&gt; ( w_2 = 174057.85 ) &lt;br&gt; ( w_4 = 113992.06 ) &lt;br&gt; ( w_6 = 51645.83 ) &lt;br&gt; ( w_8 = 103728.91 ) &lt;br&gt; ( w_{10} = 43906.46 ) &lt;br&gt; ( w_{12} = 132807.74 ) &lt;br&gt; ( w_{14} = 73364.74 ) &lt;br&gt; ( w_{16} = 43553.69 ) &lt;br&gt; ( w_{18} = 47586.70 ) &lt;br&gt; ( w_{20} = 131617.59 ) &lt;br&gt; ( w_{22} = 120279.34 ) &lt;br&gt; ( w_{24} = 52892.84 ) &lt;br&gt; ( w_{26} = 110867.23 ) &lt;br&gt; ( w_{28} = 47854.24 ) &lt;br&gt; ( w_{30} = 46863.05 )</td>
</tr>
</tbody>
</table>

Table 6.4.8: Example DAX2: Numerical effort

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># of restarts of inner loops</td>
<td>7</td>
</tr>
<tr>
<td># of solved linear programs</td>
<td>1643</td>
</tr>
<tr>
<td># of solved quadratic programs</td>
<td>35399</td>
</tr>
<tr>
<td># of constrained boxes</td>
<td>3366</td>
</tr>
<tr>
<td>( t_{LP} ): time in seconds for solving all LP</td>
<td>848.61</td>
</tr>
<tr>
<td>( t_{QP} ): time in seconds for solving all QP</td>
<td>153.91</td>
</tr>
<tr>
<td>( T_{tot} ): total time in seconds for complete iteration process</td>
<td>1711.77</td>
</tr>
</tbody>
</table>

Table 6.4.9: Example DAX2: Final parameter values

\[
\begin{array}{|c|}
\hline
\mu & 1.34E-02 \\
\hline
r & 1.17E-02 \\
\hline
h_{min} & 1.01E-07 \\
\hline
|T_b|/|T_{\beta}| & 0.78 \\
\hline
\end{array}
\]

Table 6.4.9: Example DAX2: Final parameter values
6.5 Comments

Section 6.1: This part reviews different approaches for the numerical solution of SIP. Apart from the papers mentioned in the text we refer to a number of publications that gave major contributions to duality theory (cf. Charnes, Cooper and Kortanek [73], Ben-Tal, Ben-Israel and Rosinger [38]), optimality conditions (see Ben-Tal, Teboulle and Zowe [39], Ioffe [191], Jongen, Wetterling and Zwier [201]) and also numerical methods, in particular, exchange algorithms Gustafson and Kortanek [156], Gustafson [154] or reduction methods (cf. Watson [416], Coope and Watson [83]).

Section 6.2–6.3: Path-following methods for solving parameter-dependent equations (see, for instance, Allgower and Georg [8]) have been applied to optimization problems in Guddat et al. [152] and Golmer et al. [138].

The Algorithms 6.2.12–6.2.17 and their theoretical foundation are contained in the papers of Kaplan and Tichatschke [210, 212, 214, 218].

Section 6.4: The first algorithm for SIP in the context of interior point strategies was an extension of an affine-scaling algorithm to linear SIP suggested by Ferris and Philpott [110, 111]. But it is not easily possible to extend each interior-point approach to SIP. For instance Powell [334] showed that the application of Karmarkar’s algorithm to linear SIP does not have to work. Additionally, a survey of interior-point approaches which can naturally be extended to SIP is given by Todd [396] and Tunçel and Todd [403]. A further approach originates from the method of analytic centers which was introduced by Sonnevend [377] and extensively studied by Jarre [197] for finite convex problems. In order to tackle the SIP directly, Sonnevend [377, 378] and Schättler [362, 367] extended this approach to convex SIP by introducing an integral form of the logarithmic barrier. But, unfortunately the barrier property may be lost due to the smoothing effect of the integral (cf., e.g., [403, 198]). Usually boundedness (or in fact compactness) of the feasible set or at least of the solution set of the given problem is assumed in all interior point approaches for SIP mentioned above. Dropping this restrictive assumption Kaplan and Tichatschke [226] suggested a combination of the logarithmic barrier method with a discretization procedure for the constrained set and the proximal point method. Furthermore, due to the regularization, this approach allows to treat ill-posed SIP. A further advantage of the method proposed in [226] is given by the fact that convergence of the iterates can be established. This is not always clear if one applies pure interior-point methods for convex problems except for linear and quadratic ones.
Chapter 7

PARTIAL PPR IN CONTROL PROBLEMS

A large number of interesting physical and technical problems give rise to optimal control models where the state of the system is governed by partial differential equations. The fundamental monograph of Lions [269] gives an excellent introduction into the mathematics of these models for various types of differential equations, boundary conditions, and control. In this chapter we deal with problems whose states are described by second order, elliptic equations.

7.1 Non-coercive Elliptic Control Problems

Let \( \Omega \subset \mathbb{R}^n \) be an open domain with a boundary \( \Gamma \) of the class \( C^2, \Omega = \Omega \cup \Gamma \). Then, with coefficients

\[
a_{ij} \in C^2(\bar{\Omega}), \quad i, j = 1, \ldots, n, \quad a_0 \in C^2(\bar{\Omega})
\]

such that for all \( x \in \bar{\Omega}, \xi \in \mathbb{R}^n \) and a constant \( c > 0 \)

\[
a_0(x) > 0 \quad \text{and} \quad \sum_{i,j=1}^{n} a_{ij}(x)\xi_i \xi_j \geq c \sum_{i=1}^{n} \xi_i^2,
\]

we consider the elliptic second order differential operator \( A \) defined by

\[
A_y := -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_i} \left( a_{ij} \frac{\partial y}{\partial x_j} \right) + a_0 y.
\]

Further let \( H \) denote a Hilbert space and

\[
U_{ad} \subset H \text{ a non-empty, closed, convex set},
\]

the so-called set of admissible controls.

In the case of distributed control, the state of the system is defined by

\[
\begin{align*}
A_y &= f + D_0 u \quad \text{in } \Omega, \\
B y &= 0 \quad \text{on } \Gamma,
\end{align*}
\]
with \( u \in U_{ad} \subset H = L_2(\Omega), f \in L_2(\Omega), \mathcal{D}_0 \in \mathcal{L}(L_2(\Omega), L_2(\Omega)) \) a linear continuous operator, and \( \mathcal{B} \) a boundary operator (of Dirichlet or Neumann type, for instance).

In case of boundary control, the state of the system is

\[
\begin{align*}
Ay &= f \quad \text{in } \Omega, \\
By &= g + \mathcal{D}_1 u \quad \text{on } \Gamma,
\end{align*}
\]

with \( f \in L_2(\Omega), g \in L_2(\Gamma), u \in U_{ad} \subset H = L_2(\Gamma), \mathcal{D}_1 \in \mathcal{L}(L_2(\Gamma), L_2(\Gamma)), \) and \( \mathcal{B} \) as above.

Let us assume now that, for \( u \in U_{ad} \), \( y(u) \) is uniquely determined by (7.1.5) or (7.1.6) respectively and that \( y(u) \in V, V \) a Hilbert space. For instance, in case of (7.1.5) with \( \mathcal{B} \) the trace operator, and \( \Gamma \) sufficiently smooth, an appropriate choice could be \( V = H_0^1(\Omega) \).

Given now \( U_{ad} \) and a state equation as above, the problem is to minimize a functional

\[
J(u) = ||C y(u) - \kappa_d||_H^2 + ((N u, u))_H
\]

subject to \( u \in U_{ad} \), where \( C \in \mathcal{L}(V, \mathcal{H}), \mathcal{H} \) a Hilbert space, \( \kappa_d \in \mathcal{H} \) a given element, and \( N \in \mathcal{L}(\mathcal{H}, \mathcal{H}) \) a positive semi-definite operator.

Throughout this chapter, \( ||\cdot||_S, (\cdot, \cdot)_S \) denote norm and scalar product in space \( S \), the subscript being omitted, i.e.\( ||\cdot||, (\cdot, \cdot) \), in case of \( S = L_2(\Omega) \).

For further reference we define:

### 7.1.1 Problem

With a given admissible control set (7.1.4), a boundary problem (7.1.5) or (7.1.6) and an objective \( J \) via (7.1.7) consider the control problem

\[
\min \{ J(u) : u \in U_{ad} \}.
\]

In addition, a state constraint

\[
y(u) \in G \subset V, G \text{ closed, convex}
\]

may be considered (cf. [41], [68]).

We note that the assumptions on the smoothness of the data and the type of the boundary conditions may differ essentially from the above (for instance \( a_{ij}, a_0 \in L^\infty(\Omega) \) instead of (7.1.1), or a non-homogeneous boundary condition in (7.1.5) etc.).

Particularly in numerical contributions it is common to assume the operator \( N \) to be positive definite, implying Problem 7.1.1 to be well-posed. A frequent choice is \( N : = \lambda I, \lambda > 0 \). However, from the practical point of view, the term \( ((N u, u))_H \), usually considered as a cost of the control, is not evident. On the one hand, it is not clear why this cost should depend quadratically on \( u \) and, on the other hand, (7.1.7) appears to be a rather arbitrary scalarization of a two-objective model. In the sequel we will deal with the case \( N \equiv 0 \). We believe that in many practical cases this is a more natural and relevant model. It concentrates on the primary aim of the process expressed by the first term in the objective (7.1.7) and leaves restrictions on the cost of control to the constraints defining \( U_{ad} \).

For \( N \equiv 0 \), however, the problem is likely to become ill-posed and more
difficult to handle. For bounded sets $U_{ad}$, a solution still exists (cf. [269]) but it may be non-unique unless rather unreasonable assumptions are imposed (e.g. that $B$ and $C$ be injective). In case of an unbounded $U_{ad}$ it may occur that the set of optimal controls is empty or unbounded.

In this chapter, to deal with possibly ill-posed control problems in infinite-dimensional spaces, we consider penalty methods stabilized by means of iterative PPR (see also Hettich, Kaplan and Tichatschke [176]). To be applicable to our optimal control Problem 7.1.1 a number of substantial modifications and supplements are necessary due to the following circumstances:

- There are serious difficulties in estimating the closeness between the solutions of the original and discretized problems.
- The objective $J$ depends in an implicit way on the control $u$.
- In general, it is impossible to uniformly estimate the Lagrange multipliers of the discretized problems as there are no suitable regularity conditions available.

The next Subsection 7.1.1 extensively and exemplarily deals with the important case of distributed control, Dirichlet boundary conditions, and bounded $U_{ad}$.

In [176], Part II, the results are extended to more general problems admitting distributed and boundary control with unbounded control sets and state constraints.

A significant peculiarity of our approach is that regularization is accomplished only with respect to the control variable but not the state variable. Therefore, we deal with a natural application of a partial regularization – in the subspace of control variables – which has been considered formally already in Subsection 4.3.3.

### 7.1.1 Distributed control problems

In the whole subsection we will deal with the following instance of Problem 7.1.1 (cf. also [42, 68, 118]):

With $\Omega \subset \mathbb{R}^n$ an open and bounded domain, $\Gamma$ its boundary, and $\bar{\Omega} := \Omega \cup \Gamma$, let

$$H := \mathcal{H} = L_2(\Omega), \quad V := H_0^1(\Omega) \quad (7.1.10)$$

and $\mathcal{A}$ a second order, elliptic operator (7.1.3) with coefficients $a_0, a_{ij}$ satisfying (7.1.1) and (7.1.2). The inequality $a_0(x) > 0$ may be relaxed to $a_0(x) \geq 0$ on $\bar{\Omega}$.

For $u \in U_{ad}$ we consider the Dirichlet problem

$$S_1(u) : \begin{align*}
\mathcal{A}y &= f + u \quad \text{in } \Omega, \\
y &= 0 \quad \text{on } \Gamma,
\end{align*} \quad (7.1.11)$$

with given $f \in L_2(\Omega)$ and $U_{ad}$ a bounded, closed, and convex subset of $L_2(\Omega)$.

Assuming $\Gamma$ belongs to the class $C^2$, it is well-known (cf. [19, Theorem 7.11]) that, given $u \in L_2(\Omega)$, $S_1(u)$ has a unique solution

$$y(u) \in H^2(\Omega) \cap H_0^1(\Omega) \quad (7.1.12)$$
such that the mapping
\[ T u := y(u), \quad (T : L^2(\Omega) \to H^2(\Omega) \cap H^1_0(\Omega)) \] (7.1.13)
is well defined. By \( Y \) we denote the space of functions
\[ Y := \{ y : y \in H^1_0(\Omega), \ A y \in L^2(\Omega) \}. \] (7.1.14)
Employing the Cauchy-Schwarz-Inequality and (7.1.1), it gets immediate that
\[ ((y, z))_Y := ((A y, A z)), \quad ||y||_Y := ||A y|| \] (7.1.15)
a scalar product and a norm are given on \( Y \) (recall that \((\cdot, \cdot))\) and \(||\cdot||\) denote
the scalar product and the norm in \( L^2(\Omega) \), respectively. Moreover we have

7.1.2 Proposition. \( Y \) is a Hilbert space with scalar product and norm given by
(7.1.15). Algebraically, \( Y \) coincides with the set of functions \( H^2(\Omega) \cap H^1_0(\Omega) \).

To establish the second part of Proposition 7.1.2, let \( y \in Y \). Thus, \( f := A y \in L^2(\Omega) \), and of course, \( y \) solves
\[ A y = f \text{ in } \Omega, \quad y = 0 \text{ on } \Gamma, \]
implies \( y \in H^2(\Omega) \cap H^1_0(\Omega) \). The opposite implication is obvious.

Further, we define
\[ X := Y \times L^2(\Omega) \] (7.1.16)
with norm
\[ ||(y, u)||_X = (||y||_Y^2 + ||u||^2)^{1/2} = (||A y||^2 + ||u||^2)^{1/2}. \] (7.1.17)
Using for the dual space \( (L^2(\Omega))' \) we obtain \( X' = Y' \times L^2(\Omega) \).

7.1.3 Proposition. \[269]\]
Under the assumptions on \( S_1(u) \), every \( u \in U_{ad} \) uniquely defines a process
\( (y(u), u) \in X \).

Finally, to specify the objective functional (see (7.1.10)) let
\[ C \in \mathcal{L}(H^1_0(\Omega), L^2(\Omega)) = \mathcal{L}(V, \mathcal{H}), \quad \kappa_d \in L^2(\Omega). \] (7.1.18)
Then Control Problem 7.1.1 can be reconsidered as

\[
\begin{align*}
\min & \quad \{ J(u) := \int_{\Omega} (Cy(u) - \kappa_d)^2 d\Omega \}, \\
\text{s.t.} & \quad u \in U_{ad}, y(u) \in G \subset V, \ G \text{ closed, convex}.
\end{align*}
\] (7.1.19)

We assume additionally that there exists a \( \tilde{u} \in U_{ad} \) such that \( y(\tilde{u}) \in \text{int}G \) (in \( Y \)).

Recall that \( y(u) \) is the unique solution of system (7.1.11).

7.1.4 Proposition. \( \text{cf. [269]} \)
In case \( U_{ad} \) is bounded, the set \( U^* \) of optimal controls for Problem (7.1.19) is a
non-empty, closed, convex subset of \( U_{ad} \).
7.1. NON-COERCIVE ELLIPTIC CONTROL PROBLEMS

7.1.2 Regularized penalty methods

Here we are going to formulate two methods for solving Problem (7.1.19) and state their main properties. According to the Subsection 2.2.3 (see also Subsection 4.3.1) we consider in this chapter one-step as well as multi-step regularization methods, performed by means of penalty algorithms. In the first method the penalty and regularization term are updated synchronously (one-step regularization) whereas in the second method the proximal iterations continue with fixed penalty term as long as reasonable progress is achieved (multi-step regularization). As mentioned, a further significant peculiarity is that we deal with a partial regularization carried out only with respect to the control variables, but not to the state variables. In this sense we deal with a particular case of regularization on a subspace.

In the spirit of [269], a penalty method for solving Problem (7.1.19) could proceed as follows:

Given a sequence \( \{r_k\}, r_k > 0, \lim_{k \to \infty} r_k = 0 \)

\[
\begin{align*}
J_k(y,u) & := \int_{\Omega} (Cy - \kappa_d)^2 d\Omega + \frac{1}{r_k} \int_{\Omega} (Ay - f - u)^2 d\Omega, \quad k = 1, 2, \ldots, \\
& \text{(7.1.20)}
\end{align*}
\]

and compute a sequence of minimal points \( z^k = (y^k, u^k) \) of \( J_k \) s.t. \( Y \times U_{ad} \).

Then the question is whether \( z^k \) converges to an optimal process for Problem (7.1.19). To simplify the convergence analysis, we do not include a penalization of the state constraints as it is done by Neitaanmäki and Tiba [305].

In the case of strictly convex functionals \( J \) of type (7.1.7) with \( H = H = L^2(\Omega), N := I \) and \( \mathcal{C} : V \to H \) an embedding operator, a positive answer to this question is given in [42]. For possibly ill-posed problems similar results cannot be expected. Therefore, to enforce strict convexity of the auxiliary problems, we use IPR employing an regularization term w.r.t. variable \( u \), i.e., regularization on the subspace of control variables is carried out. Note that in (7.1.20), \( y \) and \( u \) are considered to be independent variables.

7.1.5 Method. (One-step regularization)

Choose positive sequences \( \{r_k\}, \{\epsilon_k\} \), with \( \lim_{k \to \infty} r_k = \lim_{k \to \infty} \epsilon_k = 0, \sup_k r_k < 1, \sup_k \epsilon_k < 1 \), and \( u^0 \in U_{ad} \).

**Step k:** Given \( u^{k-1} \in U_{ad} \). With \( J_k \) defined by (7.1.20), let

\[
\Psi_k(y,u) := J_k(y,u) + \int_{\Omega} (u - u^{k-1})^2 d\Omega \quad (7.1.21)
\]

and

\[
(\bar{y}^k, \bar{u}^k) := \arg\min\{\Psi_k(y,u) : (y,u) \in G \times U_{ad}\} \quad (7.1.22)
\]

Compute an approximate \( (y^k, u^k) \in G \times U_{ad} \) of \( (\bar{y}^k, \bar{u}^k) \) such that

\[
||\nabla \Psi_k(y^k, u^k) - \nabla \Psi_k(\bar{y}^k, \bar{u}^k)||_{X'} \leq \epsilon_k. \quad (7.1.23)
\]

\( \diamond \)

Here, \( \nabla \Psi_k \in X' \) denotes the Gâteaux-derivative of \( \Psi_k \) which is easily seen to be given by

\[
\nabla \Psi_k(y,u)(\eta, \mu) = 2 \int_{\Omega} (Cy - \kappa_d)C\eta \\
+ \frac{1}{r_k} (Ay - f - u)(A\eta - \mu) + (u - u^{k-1})\mu d\Omega \quad (7.1.24)
\]
CHAPTER 7. PARTIAL PPR IN CONTROL PROBLEMS

for \((\eta, \mu) \in X\).

Note that for \((\bar{y}^k, \bar{u}^k)\) being a minimum according to (7.1.22) it is necessary that
\[
\nabla \Psi_k(\bar{y}^k, \bar{u}^k)(y - \bar{y}^k, u - \bar{u}^k) \geq 0, \quad \forall (y, u) \in G \times U_{ad}.
\] (7.1.25)

7.1.6 Method. (Multi-step regularization)
Let \(\{\tau_k\}, \{\epsilon_k\}, u^0 \in U_{ad}\) be as in Method 7.1.5, and \(\{\delta_k\}\) a third positive sequence (not necessarily tending to 0).

**Step k:** Given \(u^{k-1} \in U_{ad}\).

(a) Set \(u^{k,0} := u^{k-1}, i := 1\).

(b) Given \(u^{k,i-1}\), let (with \(J_k\) defined by (7.1.20))
\[
\Psi_{k,i}(y, u) := J_k(y, u) + \int_{\Omega} (u - u^{k,i-1})^2 d\Omega
\] (7.1.26)

and
\[
(\bar{y}^{k,i}, \bar{u}^{k,i}) := \arg \min \{\Psi_{k,i}(y, u) : (y, u) \in G \times U_{ad}\}. \quad (7.1.27)
\]

Compute an approximate \((y^{k,i}, u^{k,i}) \in G \times U_{ad}\) of \((\bar{y}^{k,i}, \bar{u}^{k,i})\) such that
\[
\|\nabla \Psi_{k,i}(y^{k,i}, u^{k,i}) - \nabla \Psi_{k,i}(\bar{y}^{k,i}, \bar{u}^{k,i})\|_X \leq \epsilon_k.
\] (7.1.28)

(c) If \(\|u^{k,i} - u^{k,i-1}\| > \delta_k\), set \(i := i + 1\) and repeat (b).
Otherwise, set \(u^k := u^{k,i}, i(k) := i\), and continue with Step \((k+1)\).

\[\diamondsuit\]

In order to prove convergence of both, the one-step and multi-step method, we have to take into account whether the problems possess state constraints or not.

1. The case of \(G = Y\), i.e. no state constraints.

For \(\nabla \Psi_{k,i}\) it holds relation (7.1.24) with \(u^{k-1,i}\) instead of \(u^{k-1}\) and the necessary condition (7.1.25) applies accordingly.

7.1.7 Proposition. The functionals \(\Psi_k\) in (7.1.20) as well as \(\Psi_{k,i}\) in (7.1.26) are quadratic, continuous on \(X\), and strongly convex, i.e., for some \(\alpha > 0\) we have for all \(z = (y, u), \bar{z} = (\bar{y}, \bar{u}) \in X\)
\[
\Psi_{k,i}(z) - \Psi_{k,i}(\bar{z}) \geq \langle \nabla \Psi_{k,i}(\bar{z}), z - \bar{z} \rangle + \alpha\|z - \bar{z}\|_X^2.
\]

**Proof:** The fact that \(\Psi_{k,i}\) are continuous, quadratic functionals on \(X\) is obvious.
To show that \(\Psi_{k,i}\) is strongly convex (the proof for \(\Psi_k\) is analogous) it suffices to show that its quadratic part is a positive definite quadratic form.
An elementary calculation gives
\[
\Psi_{k,i}(y, u) = Q_1(y, u) + Q_2(y, u) + \ell(y, u)
\]
with an affine linear part $\ell$ and
\[
Q_1(y, u) = \int_\Omega (Cy)^2 d\Omega + \left(\frac{1}{r_i} - 1\right) \int_\Omega (Ay - u)^2 d\Omega
\]
a positive semi-definite quadratic form (recall that $r_k < 1$) and
\[
Q_2(y, u) = \int_\Omega (Ay - u)^2 d\Omega + \int_\Omega u^2 d\Omega.
\]
We are done, if we can show that $Q_2$ is positive definite. We calculate (cf. (7.1.17))
\[
Q_2(y, u) = \int_\Omega \left(\sqrt{\frac{2}{3}} Ay - \sqrt{\frac{3}{2}} u\right)^2 d\Omega + \frac{1}{3} (\|Ay\|^2 + \|u\|^2) + \frac{1}{6} \|u\|^2 \geq \frac{1}{3} \|(y, u)\|_{X}^2.
\]
(7.1.29)
Thus, Proposition 7.1.7 is proved. □

From the proof above it follows, that $Q_2$ defines another norm $|\cdot|$ on $X$:
\[
|Q(y, u)|^2 = Q_2(y, u) = \|Ay - u\|^2 + \|u\|^2.
\]
(7.1.30)
From (7.1.29) and a simple estimation we have
\[
\frac{1}{3} \|(y, u)\|_{X}^2 \leq |Q(y, u)|^2 \leq 3 \|(y, u)\|_{X}^2.
\]
(7.1.31)
For shortness, in the sequel, we will use the abbreviations
\[
z = (y, u), \quad z^* = (y^*, u^*), \quad z^{k,i} = (\tilde{y}^{k,i}, \tilde{u}^{k,i}) \quad \text{etc.}
\]
(7.1.32)
for elements in $X = Y \times L^2(\Omega)$.

Proposition 7.1.7 ensures that the optimization problems (7.1.22), (7.1.27) become well-posed such that their respective unique solutions $(\tilde{y}^i, \tilde{u}^i)$ and $(\tilde{y}^{k,i}, \tilde{u}^{k,i})$ can be approximated by common finite element methods. Note however, that Methods 7.1.5 and 7.1.6 are conceptual because
- no stopping rule is defined for the outer iteration,
- the stopping rules (7.1.23), (7.1.28) are not yet practicable, as we don’t know the exact minimizers,
- the method for minimizing (approximately) $\Psi_k$ and $\Psi_{k,i}$ respectively is not specified.

Due to the strong convexity of $\Psi_k$ and $\Psi_{k,i}$, one can use – in order to satisfy (7.1.23) or (7.1.28) – any basic method which enables to define a point $(\hat{y}, \hat{u}) \in G \times U_{ad}$ such that
\[
\Psi_{k,i}(\hat{y}, \hat{u}) \leq \inf_{(y, u) \in G \times U_{ad}} \Psi_{k,i}(y, u) + \mu
\]
with given $\mu > 0$ (analogously, for $\Psi_{k}$). For instance, if $U_{ad} \subset L_\infty(\Omega)$ and $U_{ad}$ and $G$ are given by means of point-wise constraints, the usual discretization approach, like FEM, can be combined with (finite) conjugate direction methods or simple gradient projection methods.
7.1.8 Remark. Note, that Method 7.1.5 can be considered as a special case of Method 7.1.6 by taking \( \delta_k \) "sufficiently large", for instance
\[
\delta_k > d_0 := \sup \{ \| u - v \| : u, v \in U_{ad} \}. \tag{7.1.33}
\]
Method 7.1.6 allows for a fixed \( J_k \) a more exact minimization. This gives rise to the hope that in general, to obtain a certain accuracy, the value of the penalty parameter \( \frac{1}{r_k} \) can be kept smaller than in Method 7.1.5. Therefore, the numerical behavior of Method 7.1.6 can be expected to be much better.

To prove the next Proposition 7.1.10 we need some auxiliary estimate:

7.1.9 Lemma. : Let \( z^{k,i}, \bar{z}^{k,i} \) be as in Method 7.1.6, i.e. (cf. (7.1.28))
\[
\| \nabla \Psi_{k,i}(z^{k,i}) - \nabla \Psi_{k,i}(\bar{z}^{k,i}) \|_{X'} \leq \varepsilon_k. \tag{7.1.34}
\]
Then,
\[
\| z^{k,i} - \bar{z}^{k,i} \|_X \leq \frac{3}{2} \varepsilon_k. \tag{7.1.34}
\]

The proof is immediate from (7.1.29) and the fact that \( (\Psi_{k,i} - Q_2) \) is a convex, quadratic functional.

It should be emphasized that Lemma 9.2.5 is a consequence only of the properties of \( \Psi_{k,i} \) and (7.1.28) and otherwise independent of the method.

Now, we start with two auxiliary estimates which are in essence consequences of the properties of the operator \( A \) and the boundedness of \( U_{ad} \).

In this description, we do not focus on the explicit calculation of various constants \( c_j \), especially those which are connected with the estimation of solutions of boundary value problems and the norm of certain operators.

7.1.10 Proposition. Let \( (y^*, u^*) \) be an optimal process of Problem (7.1.19). With an arbitrarily chosen \( u^{k,i-1} \in U_{ad} \) let \( (y^{k,i}, u^{k,i}) \) and \( (\bar{y}^{k,i}, \bar{u}^{k,i}) \) be as in Substep (b) of Method 7.1.6. Then there exist constants \( d_1 \) and \( d_2 \), independent of \( u^{k,i-1}, \{ \varepsilon_k \}, \{ r_k \}, k, \) and \( i \geq 1 \), such that
\[
J_k(y^*, u^*) - J_k(\bar{y}^{k,i}, \bar{u}^{k,i}) < d_1 r_k \tag{7.1.35}
\]
and
\[
\sup_{(k,i)} \{ | (y^{k,i}, u^{k,i}) - (y^*, u^*) | \} \leq d_2, \tag{7.1.36}
\]
where \( | \cdot | \) is a norm on \( X = Y \times L_2(\Omega) \) defined by (7.1.30).

Proof: Let \( \tilde{y}^{k,i} = (\tilde{y}^{k,i}, \tilde{u}^{k,i}) \) be the unique minimal point of \( \Psi_{k,i} \) on \( Y \times U_{ad} \) and
\[
\tilde{q}^{k,i} := \frac{1}{r_k}(A\tilde{y}^{k,i} - f - \tilde{u}^{k,i}). \tag{7.1.37}
\]
Using (7.1.24) and the optimality condition (7.1.25) for \( \Psi_{k,i} \) instead of \( \Psi_k \), we obtain due to the independence of \( y, u \) that
\[
\int (C\tilde{y}^{k,i} - \kappa_d)(C\tilde{y} - C\tilde{y}^{k,i})d\Omega + \int \tilde{q}^{k,i}(A\tilde{y} - A\tilde{y}^{k,i})d\Omega \geq 0 \tag{7.1.38}
\]
for all \( y \in Y \) and
\[
\int (\tilde{y}^{k,i} - u^{k,i-1})(u - \tilde{u}^{k,i})d\Omega - \int \tilde{q}^{k,i}(u - \tilde{u}^{k,i})d\Omega \geq 0 \tag{7.1.39}
\]
for all $u \in U_{ad}$.

Furthermore, with
\begin{equation}
F_{k,i}(z) := \int_{\Omega} (C y - \kappa_d)^2 d\Omega + \int_{\Omega} (u - u^{k,i-1})^2 d\Omega, \tag{7.1.40}
\end{equation}
in view of the gradient inequality for convex functions, we obtain for all $y \in Y$, $u \in L_2(\Omega)$ the inequality
\begin{equation}
F_{k,i}(z) - F_{k,i}(\bar{z}^{k,i}) \geq 2 \int_{\Omega} (C y^{k,i} - \kappa_d)(C y - C y^{k,i}) d\Omega + 2 \int_{\Omega} (u^{k,i} - u^{k,i-1})(u - u^{k,i}) d\Omega. \tag{7.1.41}
\end{equation}

Taking $y := y^*, u := u^*$ in (7.1.38)-(7.1.41) and observing $A y^* - f - u^* = 0$, we find
\begin{equation}
F_{k,i}(z^*) \geq \frac{2}{r_k} \int_{\Omega} (A y^{k,i} - f - u^{k,i})^2 d\Omega. \tag{7.1.42}
\end{equation}
Thus
\begin{equation}
\left( \frac{1}{2} F_{k,i}(z^*) \right)^{1/2} \geq \left( \frac{r_k}{2} F_{k,i}(z^*) \right)^{1/2} \geq \|A y^{k,i} - f\| - \|u^{k,i}\|. \tag{7.1.43}
\end{equation}
Together with the boundedness of $U_{ad}$, $\sup_k r_k < 1$, $\sup_k \epsilon_k < 1$, and (7.1.34) this shows that there exists a constant $c_2$ such that
\begin{equation}
\|y^{k,i}\|_{Y} < c_2, \quad \|y^{k,i}\|_{Y} < c_1. \tag{7.1.44}
\end{equation}

Now, $\tilde{y}^{k,i}$ and $y^{k,i}$ solve the boundary value problems
\[ A y = A y^{k,i}, \quad y|_{\Gamma} = 0 \quad \text{and} \quad A y = A y^{k,i}, \quad y|_{\Gamma} = 0. \]

Therefore, a standard result from the theory of elliptic operators (see, for instance, [19]) gives the estimate $\|y^{k,i}\|_{H^2(\Omega)} \leq \text{const.} \cdot \|A y^{k,i}\|$, or due to (7.1.44), the existence of $c_2$ such that
\begin{equation}
\|y^{k,i}\|_{H^2(\Omega)} < c_2 \text{ and, analogously, } \|y^{k,i}\|_{H^2(\Omega)} < c_2. \tag{7.1.45}
\end{equation}

Let $\tilde{z}^{k,i} = (\tilde{y}^{k,i}, \tilde{u}^{k,i})$ be a feasible point (i.e. $\tilde{u}^{k,i} \in U_{ad}$, $\tilde{y}^{k,i} = \mathcal{T} \tilde{u}^{k,i}$) with minimal distance (with regard to $\|\cdot\|_{X}$) to $z^{k,i}$, i.e.
\begin{equation}
\tilde{z}^{k,i} = \arg\min \{ \|z^{k,i} - z\|_{X} : z \text{ feasible} \}. \tag{7.1.46}
\end{equation}

Now we estimate $\|\tilde{z}^{k,i} - z^{k,i}\|_{X}$. Inequality (7.1.42) shows that, due to the boundedness of $U_{ad}$, there exists a constant $c_3$ such that
\begin{equation}
\|r_k q^{k,i}\| = \|A y^{k,i} - f - u^{k,i}\| < c_3 \sqrt{r_k}. \tag{7.1.47}
\end{equation}

Let $\tilde{y}^{k,i}$ be the solution of the boundary value problem $A y = f + \tilde{u}^{k,i}, \quad y|_{\Gamma} = 0$. Then
\begin{equation}
\|A y^{k,i} - A \tilde{y}^{k,i}\| = \|A \tilde{y}^{k,i} - f - \tilde{u}^{k,i}\| < c_3 \sqrt{r_k}
\end{equation}
and hence
\begin{equation}
\|\tilde{y}^{k,i} - u^{k,i}) - (\bar{y}^{k,i}, \bar{u}^{k,i})\|_{X} < c_3 \sqrt{r_k}. \tag{7.1.49}
\end{equation}
CHAPTER 7. PARTIAL PPR IN CONTROL PROBLEMS

By the definition (7.1.46) of $\bar{z}^{k,i}$ this yields

$$\|\bar{z}^{k,i} - \bar{z}^{k,i}\|_X < c_3\sqrt{r_k}.$$  (7.1.48)

Next, with $A^*$ and $C^*$ the adjoint operators to $A$ and $C$, let $\bar{p}^{k,i}$ be a solution of the adjoint problem

$$A^* \bar{p}^{k,i} = C^*(Cy^{k,i} - \kappa_d), \quad \bar{p}^{k,i}|\Gamma = 0.$$  (7.1.49)

Because $A^*$ is again an elliptic operator of second order with coefficients in $C^2(\bar{\Omega})$, we have $\bar{p}^{k,i} \in H^2(\Omega) \cap H^1_0(\Omega)$ and, due to (7.1.45), the existence of $c_4$ such that

$$\|\bar{p}^{k,i}\|_{H^2(\Omega)} < c_4.$$  (7.1.50)

From

$$\int_{\Omega} \bar{p}^{k,i} A(y - \bar{y}^{k,i}) d\Omega = \int_{\Omega} (y - \bar{y}^{k,i}) A^* \bar{p}^{k,i} d\Omega = \int_{\Omega} (y - \bar{y}^{k,i}) C^*(Cy^{k,i} - \kappa_d) d\Omega = \int_{\Omega} (Cy^{k,i} - \kappa_d)(Cy - Cy^{k,i}) d\Omega$$

and (7.1.38), (7.1.39) we get for $(y,u) \in Y \times U_{ad}$

$$\int_{\Omega} (\bar{p}^{k,i} + \bar{q}^{k,i})(Ay - Ay^{k,i}) d\Omega + \int_{\Omega} (u - \bar{u}^{k,i})(\bar{u}^{k,i} - u^{k,i-1} - \bar{q}^{k,i}) d\Omega \geq 0.$$  (7.1.51)

Choosing $u := \bar{u}^{k,i}, \ y := T \left(\bar{u}^{k,i} - \frac{\bar{q}^{k,i}}{\|\bar{q}^{k,i}\|}\right), T$ the operator defined by (7.1.13), (thus $Ay = \bar{u}^{k,i} - \frac{\bar{q}^{k,i}}{\|\bar{q}^{k,i}\|} + f$), this gives

$$- \int_{\Omega} (\bar{p}^{k,i} + \bar{q}^{k,i}) \left(\frac{\bar{q}^{k,i}}{\|\bar{q}^{k,i}\|} + r_k \bar{q}^{k,i}\right) d\Omega \geq 0.$$  (7.1.52)

Together with (7.1.47), (7.1.50), and $r_k < 1$, we obtain with the aid of the Cauchy-Schwarz-Inequality

$$\|\bar{q}^{k,i}\| \leq - \int_{\Omega} \bar{p}^{k,i} \bar{q}^{k,i} \left(r_k + \frac{1}{\|\bar{q}^{k,i}\|}\right) d\Omega - r_k \|\bar{q}^{k,i}\|^2 \leq \|\bar{p}^{k,i}\| \|\bar{q}^{k,i}\| \left(r_k + \frac{1}{\|\bar{q}^{k,i}\|}\right) = \|\bar{p}^{k,i}\| + \sqrt{r_k} \|\bar{p}^{k,i}\| \cdot \sqrt{r_k} \|\bar{q}^{k,i}\| < c_4 + c_4c_3 =: c_5.$$  (7.1.53)

Therefore, we get the improved bounds

$$\|Ay^{k,i} - f - \bar{u}^{k,i}\| < c_5r_k$$  (7.1.54)

and, from here, according to the derivation of (7.1.48),

$$\|\bar{z}^{k,i} - \bar{z}^{k,i}\|_X < c_5r_k.$$  (7.1.55)

With an argument analogous to that used to derive (7.1.45), we estimate with constant $c_6$

$$\|y\| \leq c_6 \|y\|_Y \forall y \in Y.$$  (7.1.56)
Therefore, (7.1.55) gives
\[ \| \bar{y}^{k,i} - \bar{y}^{k,i} \| < c_6 c_5 r_k. \] (7.1.57)
Furthermore, using (with \( J_k \) according to (7.1.20))
\[ J_k(z^*) = \int_{\Omega} (Cy^* - \kappa_d)^2 d\Omega \leq \int_{\Omega} (C\bar{y}^{k,i} - \kappa_d)^2 d\Omega = J_k(\bar{z}^{k,i}) \]
we obtain with (7.1.44) and (7.1.57)
\[ J_k(z^*) - J_k(\bar{z}^{k,i}) = J_k(z^*) - J_k(\bar{z}^{k,i}) + J_k(\bar{z}^{k,i}) - J_k(z^{k,i}) \leq \int_{\Omega} (C\bar{y}^{k,i} - \kappa_d)^2 (C\bar{y}^{k,i} + C\bar{y}^{k,i} - 2\kappa_d) d\Omega \]
proving (7.1.35).

The existence of a \( d_2 \) such that (7.1.36) holds, is an easy consequence of (7.1.31), (7.1.44) and the boundedness of \( U_{ad} \). Thus the proposition is proved. □

To prove the convergence of the methods considered, we need one more lemma. Its formulation requires some further notations.

Let \( Z \) be a Hilbert space, \( Z_1 \) a subspace and \( \Pi : Z \rightarrow Z_1 \) the orthogonal projection operator. Let \( a(\cdot, \cdot) \) be a continuous, symmetric, positive semi-definite bilinear form on \( Z \times Z \) and \( \ell \) a linear, continuous functional on \( Z \).

With \( K \subset Z \) convex and closed, consider the problem
\[ \min_{z \in K} \{ \varphi(z) := a(z, z) - \ell(z) \}. \] (7.1.58)
Let \( b(\cdot, \cdot) \) be an other symmetric bilinear form on \( Z \times Z \) such that
\[ 0 \leq b(z, z) \leq a(z, z) \quad \forall z \in Z \] (7.1.59)
and
\[ b(z, z) + \| \Pi z \|^2 \geq \beta \| z \|^2 \quad \forall z \in Z \] (7.1.60)
with some \( \beta > 0 \). By
\[ |z|^2 = b(z, z) + \| \Pi z \|^2 \] (7.1.61)
another norm is defined on \( Z \) equivalent to \( \| \cdot \|_Z \) according to the obvious relation
\[ (M + 1)\| z \|^2 \geq |z|^2 \geq \beta \| z \|^2 \] (7.1.62)
with \( M \geq \sup_{z \neq 0} \frac{b(z, z)}{\| \Pi z \|^2} \).

7.1.11 Lemma. For each \( v^0 \in Z \) and
\[ v^1 := \arg \min_{z \in K} \{ \varphi(z) + \| \Pi z - \Pi v^0 \|^2 \} \]
we have for all \( z \in K \) the inequalities

\[
|v^1 - z|^2 - |v^0 - z|^2 \leq -\|\Pi v^1 - \Pi v^0\|^2_Z + \phi(z) - \phi(v^1)
\]  
(7.1.63)

and

\[
|v^1 - z| \leq |v^0 - z| + \eta(z),
\]  
(7.1.64)

where

\[
\eta(z) = \begin{cases} 
(\phi(z) - \phi(v^1))^{1/2} & \text{if } \phi(z) > \phi(v^1) \\
0 & \text{otherwise}
\end{cases}
\]

If, moreover, \( \|\Pi v^1 - \Pi v^0\|_Z \geq \delta \geq \eta(z) \), then

\[
|v^1 - z| \leq |v^0 - z| + \frac{\eta^2(z) - \delta^2}{2|v^0 - z|}.
\]  
(7.1.65)

**Proof.** \( \|\Pi z\|_Z \leq \|z\|_Z \) shows the boundedness of the bilinear form \((\Pi z, \Pi \omega)_Z\) on the space \( Z \times Z \). Due to the optimality of \( v^1 \) we have for all \( z \in K \)

\[
2a(v^1, z - v^1) - \ell(z - v^1) + 2((\Pi v^1 - \Pi v^0, \Pi z - \Pi v^1))_Z \geq 0.
\]  
(7.1.66)

Taking account of (7.1.61) a simple calculation shows

\[
|v^1 - z|^2 - |v^0 - z|^2 = b(v^1, v^1) - 2b(v^1, z) + 2b(v^0, z) - b(v^0, v^0) - \|\Pi v^1 - \Pi v^0\|^2_Z + 2((\Pi v^1 - \Pi v^0, \Pi v^1 - \Pi z))_Z.
\]

Utilizing (7.1.66), (7.1.59) and the simple inequality

\[
2a(v^1, z - v^1) \leq a(z, z) - a(v^1, v^1)
\]

a straightforward calculation leads to (7.1.63) and (7.1.64) follows immediately.

In case \( \|\Pi v^1 - \Pi v^0\|_Z \geq \delta \geq \eta(z) \), estimate (7.1.63) gives

\[
|v^1 - z|^2 - |v^0 - z|^2 \leq -\delta^2 + \eta^2(z) \leq 0,
\]

hence

\[
|v^1 - z| \leq |v^0 - z|
\]

and

\[
0 \geq -\delta^2 + \eta^2(z) \geq (|v^1 - z| + |v^0 - z|)(|v^1 - z| - |v^0 - z|)
\]

\[
\geq 2|v^0 - z|(|v^1 - z| - |v^0 - z|)
\]

proving (7.1.65). \( \square \)

We note that in Lemma 9.2.6 closedness of \( K \) is only required to ensure the existence of \( v^1 \) and could be replaced by the requirement that \( v^1 \) exists.

7.1.12 Theorem. Assume that the sequences \( \{r_k\}, \{\epsilon_k\}, \text{and} \{\delta_k\} \) in Method 7.1.6 satisfy the following conditions

\[
\sup_k \epsilon_k < 1, \sup_k r_k < 1, \sum_{k=1}^{\infty} \sqrt{r_k} < \infty, \sum_{k=1}^{\infty} \epsilon_k < \infty
\]  
(7.1.67)
7.1. NON-COERCIVE ELLIPTIC CONTROL PROBLEMS

and, with \(d_1, d_2\) from (7.1.35), (7.1.36),

\[
\frac{1}{2d_2}(d_1 r_k - (\delta_k - \frac{3}{2} \epsilon_k)^2) + \frac{3\sqrt{3}}{2} \epsilon_k < 0, \quad \delta_k > \frac{3}{2} \epsilon_k. \tag{7.1.68}
\]

Then, for arbitrary \(u^0 \in U_{ad}\), Method 7.1.6 is well-defined, especially \(i(k) < \infty \) for every \(k\), and \(\{u^{k,i}\}, \{y^{k,i}\}\) converge weakly in \(L_2(\Omega)\), \(Y\) to \(\tilde{u}, \tilde{y}\), respectively, with \((\tilde{y}, \tilde{u})\) being an optimal process for Problem (7.1.19).

For Method 7.1.5 already (7.1.67) is sufficient for weak convergence of \(\{(u^k, y^k)\}\) to \((\tilde{u}, \tilde{y})\).

**Proof:** From (7.1.68), the definition of \(i(\cdot)\) in sub-step (c) of Method 7.1.6, and Lemma 9.2.5 we conclude

\[
\|u^{k,i} - u^{k,i-1}\| \geq \|u^{k,i} - u^{k,i-1}\| - \|u^{k,i} - \tilde{u}\| > \delta_k - \frac{3}{2} \epsilon_k > 0
\]

for all \(1 \leq i < i(k)\).

Together with inequality (7.1.35) and \(d_1 r_k < (\delta_k - \frac{3}{2} \epsilon_k)^2\), due to the first inequality in (7.1.68), this implies

\[
\|u^{k,i} - u^{k,i-1}\|^2 > J_k(z^*) - J_k(z^{k,i}).
\]

Let \(z^{1,0} = (T u^{1,0}, u^{1,0})\). Applying Lemma 9.2.6 (inequality (7.1.65)) with

\[
Z := X, \quad Z_1 := \{z = (y, u) \in X : y = 0\},
\]

\[
\phi := J_k \text{ given by (7.1.20),}
\]

\[
a(z, \tilde{z}) := \int_{\Omega} Cy\tilde{y}d\Omega + \frac{1}{r_k} b(z, \tilde{z}), \quad \forall z, \tilde{z} \in X,
\]

\[
b(z, \tilde{z}) := \int_{\Omega} (Ay - u)(A\tilde{y} - \tilde{u})d\Omega,
\]

\[
\ell(z) := -J_k(z) + a(z, z),
\]

\[
K := Y \times U_{ad} \quad \delta := \delta_k - \frac{3}{2} \epsilon_k, \quad v^0 := z^{k,i-1},
\]

together with Proposition 7.1.10, we get for \(1 \leq i < i(k)\)

\[
|z^{k,i} - z^*| < |z^{k,i-1} - z^*| + \frac{1}{2d_2}(d_1 r_k - (\delta_k - \frac{3}{2} \epsilon_k)^2)
\]

and (cf. (7.1.64))

\[
|z^{k,i(k)} - z^*| < |z^{k,i(k)-1} - z^*| + \sqrt{d_1 r_k}.
\]

Utilizing (7.1.31), (7.1.34), and (7.1.68), one obtains for \(1 \leq i < i(k)\)

\[
|z^{k,i} - z^*| - |z^{k,i-1} - z^*| < \frac{1}{2d_2}(d_1 r_k - (\delta_k - \frac{3}{2} \epsilon_k)^2) + \frac{3\sqrt{3}}{2} \epsilon_k < 0 \tag{7.1.69}
\]

and

\[
|z^{k,i(k)} - z^*| - |z^{k,i(k)-1} - z^*| < \sqrt{d_1 r_k} + \frac{3\sqrt{3}}{2} \epsilon_k. \tag{7.1.70}
\]

Inequality (7.1.69) proves \(i(k) < \infty\), because, as long as \(\|u^{k,i} - u^{k,i-1}\| > \delta_k\), the reduction in \(|z^{k,i} - z^*|\) is better than a nonzero amount independent of \(i\).
POLYAK’s Lemma A3.1.8 enables us to state that the sequence \( \{ |z^{k,i} - z^*| \} \) converges if \( \sum_{k=1}^{\infty} \epsilon_k < \infty \) and \( \sum_{k=1}^{\infty} \sqrt{r_k} < \infty \). Due to (7.1.31) and (7.1.34), \( \{ |\bar{z}^{k,i} - z^*| \} \) converges to the same limit.

Let \( \{ \bar{z}^{k,i,j} \} \), with \( i_j > 0 \) for each \( j \), be a weakly convergent subsequence in \( X \), and \( \bar{e} = (\bar{y}, \bar{u}) \) its weak limit. Observing (7.1.46), (7.1.48) and the convexity and closedness of \( \{(y,u) : u \in U_{ad}, y = T u\} \) (\( T \) given by (7.1.13)) in \( X \) we conclude that \( \bar{e} \) is feasible.

By the definition of \( J_k \) and \( J_k(z^*) = J(u^*) \)

\[
|z^{k,i} - z^*|^2 - |z^{k,i-1} - z^*|^2 \leq J_k(z^*) - J_k(z^{k,i})
\]

yields

\[
|\bar{z}^{k,i} - z^*|^2 - |\bar{z}^{k,i-1} - z^*|^2 \leq J(u^*) - \int_{\Omega} (\tilde{C}\bar{y}^{k,i} - \kappa_d)^2 \, d\Omega.
\]  

(7.1.71)

Taking limit in (7.1.71) for the subsequence \( \{ \bar{z}^{k,i,j} \} \), the weak lower semi-continuity of \( \int_{\Omega} (\tilde{C}\bar{y} - \kappa_d)^2 \, d\Omega \) in \( Y \) leads to

\[
J(u^*) \geq \int_{\Omega} (\tilde{C}\bar{y} - \kappa_d)^2 \, d\Omega.
\]

Since \( \bar{e} \) is feasible, this proves that \( \bar{u} \) is optimal for Problem (7.1.19). Finally, Opial’s Lemma A1.1.3 proves the weak convergence of both \( \{ \bar{z}^{k,i} \} \) and \( \{ z^{k,i} \} \) to \( \bar{e} \) in \( X \). □

According to Remark 9.2.3, Theorem 9.2.11 also proves the convergence of Method 7.1.5 even without condition (7.1.68), which is fulfilled automatically for sufficiently large \( \delta_k \).

2. The case of state constraints

Now let a state constraint (7.1.9)

\[
y(u) \in G \subset Y
\]

be given, with \( G \) convex, closed and such that for some \( \bar{u} \in U_{ad} \)

\[
y(\bar{u}) \in \text{int } G,
\]

(7.1.72)

as assumed in Problem (7.1.19).

The only relevant modification is in the proof of Proposition 7.1.10. We show that inequalities (7.1.48), (7.1.54), and (7.1.55) remain true with modified constants \( c_j \). Then the other parts of the proofs of Proposition 7.1.10 and Theorem 9.2.11 remain unchanged.

In the sequel we give the necessary modifications of the proof of Proposition 7.1.10 in between relations (7.1.47) and (7.1.55).

Even though \( \tilde{y}^{k,i} \) may not satisfy the state constraints we still have, with \( \zeta^{k,i} = (\tilde{y}^{k,i}, \tilde{u}^{k,i}) \), that

\[
\|\bar{z}^{k,i} - \zeta^{k,i}\|_X < c_3 \sqrt{T_k}.
\]

(7.1.73)

As \( \bar{z}^{k,i} \in G \times U_{ad} \), this shows that

\[
\min_{v \in G \times U_{ad}} \|\zeta^{k,i} - v\|_X < c_3 \sqrt{T_k}.
\]

(7.1.74)
With the abbreviation \( z(u) := (y(u), u) \), let with \( \tilde{u} \) according to (7.1.72),

\[
\tau_{\min} := \min_{w \in \partial G} \| T \tilde{u} - w \|_Y,
\]

\[
\tau_{\max} := \max_{u \in U_{ad}} \| z(\tilde{u}) - z(u) \|_X,
\]

\[
\omega^{k,i} := \arg \min_{v \in G \times U_{ad}} \| \zeta^{k,i} - v \|_X.
\]

In case of \( \hat{y}^{k,i} \not\in G \), \( \omega^{k,i} \not\in \{ z(\tilde{u}) + \lambda (\zeta^{k,i} - z(\tilde{u})) : \lambda \geq 0 \} \), let

\[
h^{k,i} \in \{ z(\tilde{u}) + \lambda (\zeta^{k,i} - z(\tilde{u})) : \lambda \geq 0 \} \cap (\partial G \times U_{ad}),
\]

\[
b^{k,i} \in \{ z(\tilde{u}) + \lambda (\zeta^{k,i} - \omega^{k,i}) : \lambda \geq 0 \} \cap \{ h^{k,i} + \mu (h^{k,i} - \omega^{k,i}) : \mu \geq 0 \}
\]

\((h^{k,i}, b^{k,i})\) are uniquely defined by these relations, see Fig. 7.1.1.

\[
\begin{align*}
\| \zeta^{k,i} - \omega^{k,i} \|_X &\leq \| \zeta^{k,i} - h^{k,i} \|_X \\
&\leq \| \zeta^{k,i} - \omega^{k,i} \|_X + \| \zeta^{k,i} - h^{k,i} \|_X
\end{align*}
\]

(7.1.75)

If \( \hat{y}^{k,i} \not\in G \), but \( \omega^{k,i} \in \{ z(\tilde{u}) + \lambda (\zeta^{k,i} - z(\tilde{u})) : \lambda \geq 0 \} \), this estimate holds true, too.

Let \( \bar{z}^{k,i} \) be again a feasible point closest to \( \bar{z}^{k,i} \). Then because \( h^{k,i} \) is feasible, we get from (7.1.73) and (7.1.75)

\[
\| \zeta^{k,i} - h^{k,i} \|_X \leq \left( \frac{\tau_{\max}}{\tau_{\min}} + 1 \right) c_3 \sqrt{k}.
\]

(7.1.76)
corresponding to (7.1.48).

With \( \hat{q}^{k,i} \) and \( \hat{p}^{k,i} \) given by (7.1.37), (7.1.49) we again have for all \((y,u) \in G \times U_{ad} \) the inequality

\[
\int_{\Omega} (\hat{p}^{k,i} + \hat{q}^{k,i})(Ay - A\hat{y}^{k,i})d\Omega + \int_{\Omega} (u - \hat{u}^{k,i})(\hat{u}^{k,i} - u^{k,i-1} - \hat{q}^{k,i})d\Omega \geq 0. \tag{7.1.77}
\]

With \( \tilde{u} \) according to (7.1.72), let \( \tilde{w}^{k,i} \) be given by

\[
\tilde{w}^{k,i} = T \left( \tilde{u} - \gamma_0 \frac{q^{k,i}}{q^{k,i}} \right), \tag{7.1.78}
\]

where \( \gamma_0 > 0 \) is chosen to be a small number such that \( \tilde{w}^{k,i} \in G \) for all \((k,i) \).

Such a \( \gamma_0 \) exists because the solution of \( S_1(u) \) (cf. (7.1.11)) depends continuously on \( u \in L^2(\Omega) \) (with regard to \( \| \cdot \| \)).

Setting \( u := \tilde{u} \) and \( y := \tilde{w}^{k,i} \), inequality (7.1.77) leads to

\[
\left\| q^{k,i} \right\| = \frac{1}{\gamma_0} \left( \left\| \hat{p}^{k,i} \right\| \left( \left\| \tilde{u} - \hat{u}^{k,i} \right\| + \gamma_0 \right) + \sqrt{r_k} \left\| \hat{q}^{k,i} \right\| \right) + \left\| \tilde{u} - \hat{u}^{k,i} \right\| \left\| \hat{w}^{k,i} - u^{k,i-1} \right\|, \tag{7.1.79}
\]

and, using (7.1.47), which is true in this case too, we obtain

\[
\left\| Ay^{k,i} - f - \hat{u}^{k,i} \right\| < \tilde{c}_s r_k \]

\[
\left\| z^{k,i} - z_{k,i} \right\|_X < \frac{T_{\max}}{T_{\min} + 1} \tilde{c}_s r_k
\]

analogously to (7.1.54), (7.1.55).

\[\square\]

7.1.13 Remark. \((On \ strong \ convergence)\)

A thorough analysis of the previous results on the convergence of Methods 7.1.5 and 7.1.6 enables us to make some additional statements on the convergence of the sequences \( \{y^i\} \) and \( \{y^{k,i}\} \). For Problem (7.1.19), because the operator \( \mathcal{A} \) is an isomorphism from \( H^2(\Omega) \cap H^1_0(\Omega) \) onto \( L^2(\Omega) \), weak convergence of \( \{\hat{u}^{k,i}\} \) in \( L^2(\Omega) \) implies the same for \( \{\hat{y}^{k,i}\} \) in \( H^2(\Omega) \). Under the assumptions of Theorem 9.2.11, due to the relations (7.1.48) and (7.1.34), this immediately proves convergence of the sequence \( \{y^{k,i}\} \) to a solution of Problem (7.1.19) in the norm of the space \( H^1(\Omega) \).

This remains true in case that there are state constraints with (7.1.72) satisfied for some \( \tilde{u} \in U_{ad} \).

Now, let \( Q(\cdot) := T(\cdot) - T(0) \), with \( T \) defined by (7.1.13). Obviously, \( Q \) is a linear operator. Let us assume that the operator \( CQ \) has a finite dimensional kernel \( K \) in \( L^2(\Omega) \) and that there exists a constant \( \alpha > 0 \) such that for each \( u \in L^2(\Omega) \)

\[
\|CQ u\| \geq \alpha \|\Pi\|, \tag{7.1.80}
\]

with \( \Pi \) the operator of orthogonal projection onto the orthogonal complement of \( K \). Then, under the assumptions of Theorem 9.2.11, the sequence \( \{z^{k,i}\} \) (resp. \( \{z^k\} \)) converges strongly in the space \( X \).
Indeed, we have
\[
\|Cy(u) - \kappa_d\|^2 = \|Cy(u) - Cy(0)\|^2 + \|Cy(0) - \kappa_d\|^2 + \\
+ 2 \int_\Omega (Cy(u) - Cy(0))(Cy(0) - \kappa_d) d\Omega \\
= \|CQu\|^2 + 2 \int_\Omega (Cy(0) - \kappa_d)CQu d\Omega + \|Cy(0) - \kappa_d\|^2.
\]
Now, observing assumption (7.1.80), strong convergence of \{\bar{u}_{k,i}\} in \(L^2(\Omega)\) follows immediately from Lemma 4.1.4. Therefore, the estimates for \(\|A\bar{y}_{k,i} - f - \bar{u}_{k,i}\|\) ensure strong convergence of \{\bar{y}_{k,i}\} in \(Y\). Finally, strong convergence of \{\bar{z}_{k,i}\} in \(X\) follows from (7.1.34).

We note that penalty methods in combination with PPR (with respect to the full space) have been applied to convex variational problems for instance in (cf. [5, 21, 218, 261]). In contrast with the consideration here, in the mentioned contributions Slater’s condition with regard to the penalized constraints appears to be substantial.

### 7.1.3 A simple example

To demonstrate the effect of regularization, let us consider the following example:

\[
\min \int_{-1}^{1} (y(0) - 1)^2 dx \\
\text{s.t. } -y'' = u, \quad y(-1) = y(1) = 0
\]

with
\[
u \in U_{ad} := \{ v \in L^2(-1, 1) : v(x) \leq 0 \text{ a.e. on } (0, 1), \quad \int_{-1}^{0} \int_{-1}^{x} v(t) dt dx \geq 0, \quad \int_{-1}^{0} v(t) dt = 0 \}.
\]
The set \(U_{ad}\) is a convex and closed subset of \(L^2(-1, 1)\) and the above objective corresponds to the choice \(Cy = y(0), \ \kappa_d \equiv 1\). The operator \(C\) maps \(H^1_0(-1, 1)\) into \(L^2(-1, 1)\) and its boundedness is a consequence of the continuous embedding of \(H^1_0(-1, 1)\) into \([C[-1, 1]]\).

For a given \(u \in U_{ad}\) the function
\[
y'(x) = y'(-1) - \int_{-1}^{x} u(t) dt
\]
is absolutely continuous and
\[
y(x) = y'(-1)(x + 1) - \int_{-1}^{x} \int_{-1}^{\xi} u(t) dtd\xi
\]
solves the boundary value problem. Observing the latter formula, due to the boundary conditions, we conclude for \(x = 1\) that
\[
y'(-1) = \frac{1}{2} \int_{-1}^{1} \int_{-1}^{\xi} u(t) dtd\xi
\]
CHAPTER 7. PARTIAL PPR IN CONTROL PROBLEMS

and for $x = 0$

$$y(0) = y'(-1) - \int_{-1}^{0} \int_{-1}^{\xi} u(t) dt d\xi.$$  

Taking into account that $u \in U_{ad}$, we get further

$$y'(-1) = \frac{1}{2} \int_{-1}^{0} \int_{-1}^{\xi} u(t) dt d\xi + \frac{1}{2} \int_{0}^{1} \int_{-1}^{\xi} u(t) dt d\xi = \frac{1}{2} \int_{-1}^{0} \int_{-1}^{\xi} u(t) dt d\xi + \frac{1}{2} \int_{0}^{1} \int_{-1}^{\xi} u(t) dt d\xi \leq \frac{1}{2} \int_{-1}^{0} \int_{-1}^{\xi} u(t) dt d\xi.$$  

Hence, for every $u \in U_{ad}$, we infer

$$y(0) = y'(-1) - \int_{-1}^{0} \int_{-1}^{\xi} u(t) dt d\xi \leq -\frac{1}{2} \int_{-1}^{0} \int_{-1}^{\xi} u(t) dt d\xi \leq 0,$$

and

$$\int_{-1}^{1} (y(0) - 1)^2 dx \geq 2.$$  

However, it is obvious that the process $(y^{(1)}, u^{(1)}) \equiv (0, 0)$ is feasible for the control problem and has the objective value $\int_{-1}^{1} (y^{(1)}(0) - 1)^2 dx = 2$, i.e. $(y^{(1)}, u^{(1)})$ is optimal for the problem under consideration.

It is easily seen that there are other solutions, in particular,

$$y^{(2)}(x) := \begin{cases} \frac{1}{4\pi^2} (\cos 2\pi x - 1) & \text{for } x \in [-1, 0] \\ 0 & \text{for } x \in (0, 1] \end{cases},$$

$$u^{(2)}(x) := \begin{cases} \cos 2\pi x & \text{for } x \in [-1, 0] \\ 0 & \text{for } x \in (0, 1] \end{cases},$$

and

$$y^{(3)}(x) := \begin{cases} -\frac{x^2}{2} - x - \frac{1}{2} & \text{for } x \in [-1, -\frac{1}{2}] \\ \frac{x^2}{2} + \frac{1}{16} & \text{for } x \in (-\frac{1}{2}, -\frac{1}{4}) \\ -\frac{x^2}{2} & \text{for } x \in (-\frac{1}{4}, 0] \\ 0 & \text{for } x \in (0, 1] \end{cases},$$

$$u^{(3)}(x) := \begin{cases} +1 & \text{for } x \in [-1, -\frac{1}{2}] \\ -1 & \text{for } x \in (-\frac{1}{2}, -\frac{1}{4}) \\ +1 & \text{for } x \in (-\frac{1}{4}, 0] \\ 0 & \text{for } x \in (0, 1] \end{cases}.$$  

Moreover, it is not difficult to verify that for arbitrary $a \in \mathbb{R}$

$$(ay^{(2)}, au^{(2)}) \text{ and } (ay^{(3)}, au^{(3)})$$

are again solutions, i.e. the optimal set $U^*$ is unbounded.

Now, let us consider the penalized problem: Minimize, with fixed $r_k > 0,$

$$J_k(y, u) = \int_{-1}^{1} (y(0) - 1)^2 dx + \frac{1}{r_k} \int_{-1}^{1} (y'' + u)^2 dx \quad \text{s.t. } y \in Y, \ u \in U_{ad},$$

where, according to (7.1.14), $Y = H^0_0(-1, 1) \cap H^2(-1, 1).$

No solution of this problem (if it is solvable) can satisfy the differential equation $-y'' = u.$
Indeed, the first variation of the functional $J_k$ in $(y,u) \in Y \times U_{ad}$ gives for $\eta \in Y$, $\nu \in U_{ad}$

\[
\frac{1}{\alpha} \left[ \int_{-1}^{1} \{(y(0) + \alpha \eta(0) - 1)^2 - (y(0) - 1)^2\} dx 
+ \frac{1}{r_k} \int_{-1}^{1} ((y'' + \alpha u' + u + \alpha \nu) dy - (y'' + u) dy) dx \right] |_{\alpha = 0} = 2 \left[ \int_{-1}^{1} (y(0) - 1) \eta(0) dx + \frac{1}{r_k} \int_{-1}^{1} (y'' + u)(\eta'' + \nu) dx \right].
\]

If $y'' + u = 0$, then $(y,u)$ is a feasible control, hence $y(0) - 1 \leq -1$, and choosing for instance $\eta(x) = x^2 - 1$ and $\nu(x) \equiv 0$, we obtain a non-zero value of the first variation.

However, if $(\hat{y}, \hat{u})$ is a solution of the penalized problem, then it is easily seen that for arbitrary $a$ the pairs

\[(\hat{y} + ay^{(2)}, \hat{u} + au^{(2)}) \quad \text{or} \quad (\hat{y} + ay^{(3)}, \hat{u} + au^{(3)})\]

are also solutions.

Moreover, approximating $u$ by piecewise constant functions and $y$ by Hermite cubic splines, one can verify that the approximate penalized problem is in general not uniquely solvable. For instance, if on the interval $[-1,1]$ a grid with step-size $h = \frac{1}{4}$ is chosen, then, with $(\hat{y}_h, \hat{u}_h)$ a solution of the approximate problem, the pair $(\hat{y}_h + ay^{(3)}, \hat{u}_h + au^{(3)})$ is also a solution for arbitrary $a$ (this is true because the Hermite approximation of the function $y^{(3)}$ coincides with $y^{(3)}$).

Hence, the Hessian of the approximate penalty function is not regular.

In Method 7.1.6 we deal with regularized penalty problems:

\[
\min \Psi_{k,i}(y,u) = \int_{-1}^{1} (y(0) - 1)^2 dx + \frac{1}{r_k} \int_{-1}^{1} (y'' + u)^2 dx + \|u - u^{k-1}\|_{L_2(-1,1)}^2
\]

s.t. $y \in Y$, $u \in U_{ad}$.

where $\Psi_{k,i}$ are strongly convex functions in $X$ (cf. (7.1.16),(7.1.17)). Approximating these problems as mentioned above, we obtain quadratic programming problems with strongly convex objective functions in corresponding finite-dimensional spaces. Hence, fast convergent methods can be applied for solving these finite-dimensional problems.

### 7.2 Elliptic Control with Mixed Boundary Conditions

#### 7.2.1 Discretization in state and control space

As considered in the previous Chapter 7.1, penalization of the state equation permits to handle the state variable $y$ and control variable $u$ as independent. On the one side this is an advantage on the other side it complicates the discretization process. For instance, applying finite element methods, one has to use elements with order higher than one (cf. LASIECKA [256]).

In the following chapter we renounce the use of penalty techniques. Considering the elliptic boundary value problem as a functional restriction in the process
space, the auxiliary problems in the MSR-method becomes simpler and the class of elliptic state equations can be widened.

In the previous chapter we did not focus on the question about discretization and numerical realization. In particular, we did not estimate the discretization and computational errors and the question leaves open why the regularizing parameter $\chi_k$ cannot be arbitrarily decreased in the course of switching from one discretization level to the next. Moreover, the stopping parameter of the inner prox-cycle $\delta_k$ in Algorithm 7.1.6 could not be computed because of the unknown constants $d_1$ and $d_2$. Now, these questions will be answered in this chapter more thoroughly.

Let us consider a convex, two-dimensional polygonal domain $\Omega$ with edges $S_m, m \in \mathcal{M} = \{1,...,M\}$ (cf. Figure 7.2.2). It is supposed that the boundary of this domain can be partitioned into two parts: $\partial \Omega = \Gamma_D \cup \Gamma_N$. Each part can be described as a union of a certain number of facets $\Gamma_m, m \in \mathcal{M}$, of the polygon:

$$\Gamma_D = \bigcup_{m \in \mathcal{M}_D} \Gamma_m, \quad \Gamma_N = \bigcup_{m \in \mathcal{M}_N} \Gamma_m,$$

with $\mathcal{M}_D \cup \mathcal{M}_N = \mathcal{M}, \mathcal{M}_D \neq \emptyset$. The facets and edges of the polygon are numbered corresponding to a positive orientation of the boundary (counterclockwise). Hereby let $S_m$ be the last edge of $\Gamma_m$. Each edge is characterized by the modul of its angle $\omega_m \in (0, \pi], m \in \mathcal{M}$.

On such polygonal domain we study the boundary value problem

$$-\Delta y(x) = f(x) \quad \text{in } \Omega,$$

$$\frac{\partial y}{\partial \nu_m}(x) = g_m(x) \quad \text{on } \Gamma_m, \quad m \in \mathcal{M}_N,$$

$$y(x) = 0 \quad \text{on } \Gamma_m, \quad m \in \mathcal{M}_D,$$

(7.2.1)

where $f \in L_2(\Omega), g_m \in L_2(\Gamma_m)$.

With $s = k + \lambda > 0, k \in \mathbb{Z}_+, \lambda \in [0,1)$ and $\mathcal{M} \subset \mathbb{R}^n$ a measurable set, $\text{meas} \mathcal{M} > 0$, we introduce the following norms (cf. ([148] and [108]):

\begin{align*}
\|u\|_{\mathcal{W}^1,s}^2 = & \sum_{m \in \mathcal{M}} \left( \sum_{\omega_m \in (0,\pi]} \frac{1}{s^2 - \omega_m^2} \int_{\Gamma_m} \left( \left| \frac{\partial u}{\partial \nu_m} \right|^2 + \lambda |\nabla u|^2 \right) \, ds \right), \\
\|u\|_{\mathcal{W}^2,s}^2 = & \sum_{m \in \mathcal{M}} \left( \sum_{\omega_m \in (0,\pi]} \frac{1}{s^2 - \omega_m^2} \int_{\Gamma_m} \left( \left| \frac{\partial^2 u}{\partial \nu_m^2} \right|^2 + \lambda |\nabla^2 u|^2 \right) \, ds \right), \\
\|u\|_{\mathcal{W}^3,s}^2 = & \sum_{m \in \mathcal{M}} \left( \sum_{\omega_m \in (0,\pi]} \frac{1}{s^2 - \omega_m^2} \int_{\Gamma_m} \left( \left| \frac{\partial^3 u}{\partial \nu_m^3} \right|^2 + \lambda |\nabla^3 u|^2 \right) \, ds \right),
\end{align*}
7.2. ELLIPTIC CONTROL WITH MIXED BOUNDARY CONDITIONS

\[ \|u\|_{s,M} = \|u\|_{W^k(M)} = \left( \sum_{|\alpha| \leq k} \int_M |D^\alpha u(x)|^2 dx \right)^{1/2}, \quad \text{if } \lambda = 0, \]
otherwise

\[ \|u\|_{s,M} = \left( \|u\|^2_{W^k(M)} + \sum_{|\alpha| = k} \int\int_M \frac{|D^\alpha u(x) - D^\alpha u(y)|^2}{|x-y|^{n+2\lambda}} dx dy \right)^{1/2}, \]
and

\[ \|f\|_{s,M} = \sup_{\phi \in C^\infty(M) \setminus \{0\}} \frac{\langle f, \phi \rangle}{\|\phi\|_{s,M}}. \]

In order to allow a weak formulation of Problem (7.2.1) we introduce the following Hilbert space:

\[ \mathcal{Y} = \{ y \in W^1(\Omega) : \gamma_0^{(m)} y = 0 \, \text{ in } H^{1/2}(\Gamma_m), \, m \in \mathcal{M}_D \}, \quad (7.2.2) \]
which is considered in the sequel as the state space. The trace operators \( \gamma_0^{(m)} \) in (7.2.2) correspond to those facets of the polygon which are associated with the homogenous boundary conditions. Now, the variational problem consists in finding an element \( y \in \mathcal{Y} \) which satisfies for arbitrary \( \tilde{y} \in \mathcal{Y} \) the following equation:

\[ \sum_{j=1,2} \langle D_{x_j} y, D_{x_j} \tilde{y} \rangle_\Omega = \langle f, \tilde{y} \rangle_\Omega + \sum_{m \in \mathcal{M}_N} \langle g_m, \gamma_0^{(m)} \tilde{y} \rangle_{\Gamma_m}. \quad (7.2.3) \]
Because Green’s formula holds for bounded Lipschitzian domains [148], each “classical” solution of (7.2.1) satisfies the variational equation (7.2.3). Equation (7.2.3) can be equivalently reformulated in operator form

\[ -\Delta y = f + \sum_{m \in \mathcal{M}_N} g_m \delta_{\Gamma_m} \quad \text{in } \mathcal{M}', \quad (7.2.4) \]

defining a generalized Laplace operator and the so-called Delta layers [19]. It should be mentioned that the operator form of the elliptic equation (7.2.4) is not equivalent to the classical boundary value problem (7.2.1). In (7.2.1) it is supposed that all functions involved are measurable, whereas in (7.2.4) one deals with generalized distributions. The elements \( f \) and \( g_m \) are now permitted to be located outside of the space \( L_2 \). Thus, one has to find additional regularity conditions for the elements \( g_m, m \in \mathcal{M}_N \), which guarantee the existence of a classical solution \( y \in \mathcal{Y} \cap W^2(\Omega) \) of (7.2.4) (cf. [148, Kap. 4]).

To each edge \( S_m \) on the oriented boundary \( \partial \Omega \) a periodical vector function \( x_m \in [C^{0,1}(\infty, \infty)]^2 \) is assigned. Hereby it is agreed that the vector \( x_m(s) \) circulates infinitely often on the boundary \( \partial \Omega \), whereas the parameter \( s \), defining the sign-oriented distance from the point \( S_m \) on the curve \( \partial \Omega \), is changing from \( -\infty \) to \( \infty \). Now, for each edge \( S_m \) we construct the following (bi)quadratic functionals on \( L_2(\Gamma_m) \times L_2(\Gamma_{m+1}) \):
CHAPTER 7. PARTIAL PPR IN CONTROL PROBLEMS

(A) If \( m \in \mathcal{M}_N \) and \( (m + 1) \in \mathcal{M}_D \), then

\[
T_m(g_m) = \int_0^{c_m} \frac{|g_m(x_m(-s))|^2}{s} \, ds;
\]

(B) if \( m \in \mathcal{M}_D \) and \( (m + 1) \in \mathcal{M}_N \), then

\[
T_m(g_{m+1}) = \int_0^{c_m} \frac{|g_{m+1}(x_m(s))|^2}{s} \, ds;
\]

(C) if \( m \in \mathcal{M}_N \) and \( (m + 1) \in \mathcal{M}_N \), then

\[
T_m(g_m, g_{m+1}) = \int_0^{c_m} \frac{|g_{m+1}(x_m(s)) - g_m(x_m(-s))|^2}{s} \, ds,
\]

with \( c_m = \min \{ \text{meas } \Gamma_m, \text{meas } \Gamma_{m+1} \} \). (Index \( M + 1 \) is identified with index 1, where \( M \) denotes the number of facets of the polygon). According to the introduced classification of the edges above we can determine corresponding subsets of index sets \( \mathcal{M}_A \subset \mathcal{M} \), \( \mathcal{M}_B \subset \mathcal{M} \) and \( \mathcal{M}_C \subset \mathcal{M} \).

7.2.1 Theorem. Assume the system of elements \( g_m, m \in \mathcal{M}_N \), in (7.2.4) satisfies the following conditions:

- \( g_m \in H^{1/2}(\Gamma_m), m \in \mathcal{M}_N \);
- \( T_m(g_m) < \infty \), if \( \omega_m = \pi/2 \) and \( m \in \mathcal{M}_A \);
- \( T_m(g_{m+1}) < \infty \), if \( \omega_m = \pi/2 \) and \( m \in \mathcal{M}_B \);
- \( T_m(g_m, g_{m+1}) < \infty \), \( m \in \mathcal{M}_C \)

and \( \omega_m \leq \pi/2 \), \( m \in \mathcal{M}_A \cup \mathcal{M}_B \).

Then the solution of the state equation (7.2.4) belongs to the space \( W^2(\Omega) \).

The next two theorems describe some constants \( c_F \) and \( c_\gamma \) which are important for the estimates needed afterwards.

7.2.2 Theorem. For each element \( y \in \mathcal{M} \) with \( \mathcal{M}_D \neq \emptyset \) there exists a positive constant \( c_F \) (independent from \( y \)) such that

\[
\|y\|_\Omega \leq c_F \sqrt{\sum_{j=1,2} \|D_x y\|^2_\Omega}.
\]  

(7.2.5)

7.2.3 Theorem. For each element \( y \in W^1(\Omega) \) there exists a positive constant \( c_\gamma \) (independent from \( y \)) such that

\[
\|\gamma y\|_{\partial \Omega} \leq c_\gamma \|y\|_{1,\Omega}.
\]  

(7.2.6)
7.2. ELLIPTIC CONTROL WITH MIXED BOUNDARY CONDITIONS

The formulas for the determination of $c_F$ and $c_\gamma$ by means of the geometric parameters of the polygon can be found in [357]. The corresponding proofs of Theorems 7.2.2 and 7.2.3 can be transferred to the $n$-dimensional case.

Very rarely one succeeds in obtaining a closed form of the solution of (7.2.4). Hence, in the most cases approximate solutions are the only possibility to get information about the exact solutions.

According to the finite-element-technique (see, for instance, [286], [291], [382]), we start with a triangulation of the polygon. In fact we consider a sequence of nested triangulations. Each grid gets a number $k = 1, 2, ...$ and we determine the sets of all edges $\{K^k_j, j \in I^k\}$ and of all sub-domains $\{\Omega^k_j, j \in T^k\}$ on the $k$-th level of discretization. The parameter $h_k$ of the discretization of the grid is then determined by

$$h_k = \max_{j \in I^k} \text{diam}(\Omega^k_j),$$

characterizing the coarseness of the triangulation. We single out a subset of indices $J^k_0 \subset J^k$ corresponding to those edges, which are not located on the Dirichlet boundary $\Gamma_D$. Dealing with a fixed level of discretization, we will omit the index $i$. To each $K_j, j \in J_0$ a "hat"-function $\varphi_j(x)$ is assigned (cf. [135], [286], [382]) and the space of inner approximations can be defined (cf. [135]):

$$Y^h = \{ y_h \in Y : y_h(x) = \sum_{j \in J_0} y_j \varphi_j(x), y_j \in \mathbb{R} \}.$$

On that way we get the discretized minimization problem

$$\min \{ F(y_h) : y_h \in Y^h \}, \quad (7.2.7)$$

with

$$F(y_h) = \frac{1}{2} \sum_{j=1,2} \| D_{x_j} y_h \|_{\Omega}^2 - \langle f, y_h \rangle_{\Omega} - \sum_{m \in M_N} \langle g_m, \gamma_0(\Omega) y_h \rangle_{\Gamma_m}.$$

Naturally the question arises how good approximates the solution of Problem (7.2.7) the corresponding exact solution of Problem (7.2.4). The known results in [286, 382] correspond only to boundary conditions of the first, second and third type but not to mixed boundary problems. So we are going to estimate these solution in our case.

Let $y \in Y$ be the exact solution of Problem (7.2.4) and $y_h \in Y^h$ the exact solution of the discretized Problem (7.2.7), then it holds

$$\| y - y_h \|_{\Omega} \leq \tilde{C}_0 h^2 \left( \| f \|_{\Omega} + \sum_{m \in M_N} \| g_m \|_{1/2, \Gamma_m} \right) + \sum_{m \in M_A} T_m^{1/2} (g_m) + \sum_{m \in M_B} T_m^{1/2} (g_{m+1}) + \sum_{m \in M_C} T_m^{1/2} (g_m, g_{m+1}) \quad (7.2.8)$$

To study the behavior of the MSR-method we consider the following special case:
\[ \min \{ J(y, u) := \frac{1}{2} \| y \|_{L^2}^2 : (y, u) \in \mathcal{Y} \times \mathcal{U}_{ad} \} \]

s.t.
\[ -\Delta y = u + f + \sum_{m \in M_N} g_m \delta_{\Gamma_m} \text{ in } \mathcal{Y}', \]

where the state space \( \mathcal{Y} \) is defined by (7.2.2). For the control space and the set of admissible controls we choose
\[ \mathcal{U} = L^2(\Omega), \quad \mathcal{U}_{ad} = \{ u \in \mathcal{U} : \xi_0(x) \leq u(x) \leq \xi_1(x) \text{ a.a in } \Omega \}, \]

with \( \xi_0(x) \in W^1(\Omega) \cap L^\infty(\Omega), \quad \xi_1(x) \in W^1(\Omega) \cap L^\infty(\Omega) \).

The domain \( \Omega' \) can be chosen as an open measurable sub-domain of \( \Omega \), knowing that the smaller the sub-domain \( \Omega' \) the more restricted is the space of observations \( \mathcal{K} = L^2(\Omega') \). From the previous chapter we learned that the absence of quadratic control costs in the objective functional is going to cause the ill-posedness of the Problem (7.2.9).

In the framework of this chapter we investigate the well-known chattering regimes [419, 420]. Our goal is to observe an ill-posed behavior for problems with distributed controls on a two-dimensional polygon. Such problems correlate with problems considered in [176], Part II, where inhomogeneous boundary conditions of the Neumann type are considered. Because the behavior of the solution is complicated anyway, we restrict ourselves on the study of the case where in (7.2.9) no state constraints are given: \( \mathcal{Z}_{ad} = \mathcal{Y} \times \mathcal{U}_{ad} \). In [356] the existence of a solution for a general problem is proved.

Let us consider the solution set \( \mathcal{Z}^* \) in the space \( \mathcal{Z} = \mathcal{Y} \times \mathcal{U} \), equipped with the norm
\[ \| z \|_Z = \sqrt{\| y \|_{L^2(\Omega)}^2 + \| u \|_{L^2(\Omega)}^2}. \] (7.2.10)

Because of the smallness of the observation space one cannot expect a unique solution of (7.2.9).

In the spirit of Subsection 2.2.3 we construct the MSR-algorithm for solving Problem (7.2.9) in such a way that a small deviation of the approximate solution \( z^{k,i} \) from the exact solution \( \tilde{z}^{k,i} \) of the auxiliary problem
\[ \min \{ \Psi_{k,i}(y, u) = \frac{1}{2} \| y \|_{L^2}^2 + \frac{\lambda_k}{2} \| u - u^{k,i-1} \|_{L^2(\Omega)}^2 : (y, u) \in \mathcal{Z}_{ad} \}, \]

s.t.
\[ -\Delta y = u + f + \sum_{m \in M_N} g_m \delta_{\Gamma_m} \text{ in } \mathcal{Y}' \] (7.2.11)

is permitted in the weaker norm (7.2.10), i.e., we are looking for an approximate solution of the regularized auxiliary Problem 7.2.11.

Next we explain in detail the components of the error appearing by solving the auxiliary Problem (7.2.11). All together we are dealing with two components. We start with the investigation of the discretization error. To this end we need an inner approximation of \( \mathcal{U} \) and \( \mathcal{U}_{ad} \) analogously to the approximation of the state space \( \mathcal{Y} \). To each elementary triangle \( \Omega_j \), a piece-wise constant basis function

\[ \phi_i \]
\[ \psi_j(x), \ j \in \mathcal{I}, \ \text{is assigned (cf. [108], [286], [382]). Then the inner approximation is defined by:} \]

\[ \mathcal{U}^h = \{ u_h \in L_2(\Omega) : u_h(x) = \sum_{j \in \mathcal{I}} u_j \psi_j(x), \ u_j \in \mathbb{R} \}. \]

To describe the inner approximation of the convex set \( \mathcal{U}_{ad} \) via [135], we introduce the following notation for the mean value of an arbitrary function \( \psi \in L^2(\Omega) \) on the elementary triangle \( \Omega_j \):

\[ M_j(\psi) = \frac{1}{\text{meas } \Omega_j} \int_{\Omega_j} \psi \, dx. \] (7.2.12)

Then the set of inner approximation reads as:

\[ \mathcal{U}_{ad}^h = \{ u_h \in \mathcal{U}^h : M_j(\xi_0) \leq u_h(x) \leq M_j(\xi_1) \ \text{a.a. in } \Omega_j, \ \forall \ j \in \mathcal{I} \}. \]

It should be noted that to simplify the notation all approximated sets \( \mathcal{Y}^h, \mathcal{U}^h, \mathcal{U}_{ad}^h \) are constructed with the same parameter of discretization \( h \), whereas in [108] two different parameters – one for the state and one for the control functions – are used.

Let \( z_h^{k,i} = (y_h^{k,i}, u_h^{k,i}) \) be the exact solution of the discretized auxiliary problem

\[
\min \{ \bar{\Psi}_{k,i}(y_h, u_h) : (y_h, u_h) \in \mathcal{Y}^h \times \mathcal{U}_{ad}^h \},
\]

s.t.

\[ \sum_{j=1,2} \langle D_{x_j}y_h, D_{x_j}\tilde{y}_h \rangle_\Omega = \langle u_h + f, \tilde{y}_h \rangle_\Omega + \sum_{m \in \mathcal{M}_N} \langle g_m, \gamma_0^{(m)} \tilde{y}_h \rangle_{\Gamma_m} \forall \ \tilde{y}_h \in \mathcal{Y}^h, \] (7.2.13)

where the state equation is described in variational form. Problem (7.2.13) is a finite-dimensional quadratic minimization problem, which can be solved by standard optimization algorithms (cf. Section A3.4). These algorithms have also an error such that the solution \( z_h^{k,i} \) can never be computed exactly. Thus, we have to deal with an inexact solution \( (y_h^{k,i}, u_h^{k,i}) \) of the auxiliary Problem (7.2.13). The MSR-method now can be described as follows:

**7.2.4 Algorithm. (MSR-method)**

Data: \( u^{0,0} \in \mathcal{U}, \{ \eta_k \} \downarrow 0, \{ \delta_k \} \downarrow \delta_0 \).

S0: Set \( i(0) := 0, \ k := 1 \);

S1: (a) set \( u^{k,0} := u^{k-1,i(k-1)}, \ i := 0 \);

(b) given \( u^{k,i} \in \mathcal{U}, \ i := i + 1 \);

compute a approximate \( z_{k,i} = (y_h^{k,i}, u_h^{k,i}) \in \mathcal{Y}^h \times \mathcal{U}_{ad}^h \)

of the exact solution \( z_h^{k,i} = (y_h^{k,i}, u_h^{k,i}) \in \mathcal{Y}^h \times \mathcal{U}_{ad}^h \)

of the discretized auxiliary Problem (7.2.13), such that

\[ \bar{\Psi}_{k,i}(y_h^{k,i}, u_h^{k,i}) - \bar{\Psi}_{k,i}(y_h^{k,i}, u_h^{k,i}) \leq \eta_k, \] (7.2.14)

and

\[ \sum_{j=1,2} \langle D_{x_j}y_h^{k,i}, D_{x_j}\tilde{y}_h \rangle_\Omega = \langle u_h^{k,i} + f, \tilde{y}_h \rangle_\Omega + \sum_{m \in \mathcal{M}_N} \langle g_m, \gamma_0^{(m)} \tilde{y}_h \rangle_{\Gamma_m} \forall \ \tilde{y}_h \in \mathcal{Y}^h; \]
(c) if \( \|y^{k,i} - a^{k,i-1}\|_\Omega \geq \delta_k \), go to (b),
otherwise set \( i(k) := i, k := k + 1 \) go to (a).

\[ \diamond \]

The control parameter \( \eta_k \) characterizes the computational error. The assumption that the discretized state equation can be solved exactly is agreeable because the state equation itself is well-posed and therefore it is solvable with high accuracy.

### 7.2.2 Convergence of inexact MSR-methods

The next goal is to prove the convergence of MSR-method 7.2.4 (see Theorem 7.2.12 below). We start with the estimation of the computational error by means of the controlling parameters. In [357] the following lemma is proved:

#### 7.2.5 Lemma

For the approximate solutions introduced in (7.2.14) and (7.2.13), respectively, it holds:

\[
\|y^{k,i} - a\|_{1,\Omega} \leq \sqrt{1 + \varepsilon_F^2} \sqrt{\frac{2\eta_k}{\lambda_k}} \|u^{k,i} - a\|_{\Omega} \leq \sqrt{\frac{2\eta_k}{\lambda_k}} \quad (7.2.15)
\]

The next question which we have to answer is how good the auxiliary Problem (7.2.13) approximates the Problem (7.2.11)? We fix the indices \( k \) and \( i \) in the MSR-method and will drop these sometimes if convenient. It should be noted that the proof techniques are similar to those in [108]. The state equations in (7.2.11) and (7.2.13) correspond to two affine mappings \( \bar{y}(\bar{u}) \) and \( y_h(u_h) \), satisfying (due to the optimality conditions of first order for Problem (7.2.11) and (7.2.13)) the following variational inequalities, correspondingly:

\[
\langle \bar{y}(\bar{u}), y(v) - \bar{y}(\bar{u}) \rangle_{\Omega} + \lambda_k \langle \bar{u} - u^{k,i-1}, v - \bar{u} \rangle_{\Omega} \geq 0, \quad \forall v \in U_{ad}, \tag{7.2.16}
\]

\[
\langle y_h(u_h), y_h(v_h) - y_h(u_h) \rangle_{\Omega} + \lambda_k \langle u_h - u^{k,i-1}, v_h - u_h \rangle_{\Omega} \geq 0, \quad \forall u_h \in U_{ad}^{h}. \tag{7.2.17}
\]

Adding up (7.2.16) and (7.2.17) we get

\[
a_u(v, v_h) + \lambda_k a_u(v, v_h) \geq 0, \quad \forall v \in U_{ad}, \quad \forall u_h \in U_{ad}^{h} \tag{7.2.18}
\]

with

\[
a_u = \langle \bar{u} - u^{k,i-1}, v - \bar{u} \rangle_{\Omega} + \langle u_h - u^{k,i-1}, v_h - u_h \rangle_{\Omega}
\]

\[
= -\|\bar{u} - u_h\|_{\Omega}^2 + \langle \bar{u}, v - u_h \rangle_{\Omega} + \langle \bar{u}, v_h - \bar{u} \rangle_{\Omega}
\]

\[
+ \langle u_h - \bar{u}, v_h - \bar{u} \rangle_{\Omega} - \langle u^{k,i-1}(v - u_h) + \langle v_h - \bar{u} \rangle_{\Omega} \rangle \tag{7.2.19}
\]

and

\[
a_y = \langle \bar{y}(\bar{u}), y(v) - \bar{y}(\bar{u}) \rangle_{\Omega} + \langle y_h(u_h), y_h(v_h) - y_h(u_h) \rangle_{\Omega}
\]

\[
= -\|\bar{y}(\bar{u}) - y_h(u_h)\|_{\Omega}^2 + \langle \bar{y}(\bar{u}), (y(v) - \bar{y}(u_h)) + \langle y(v_h) - \bar{y}(\bar{u}) \rangle_{\Omega}
\]

\[
+ \langle \bar{y}(\bar{u}), (y_h(u_h) - y_h(u_h)) + (y_h(v_h) - \bar{y}(\bar{u})) \rangle_{\Omega}
\]

\[
+ \langle (y_h(u_h) - \bar{y}(\bar{u})), (y_h(v_h) - \bar{y}(\bar{v})) + (\bar{y}(v_h) - \bar{y}(\bar{u})) \rangle_{\Omega}. \tag{7.2.20}
\]
7.2. ELLIPTIC CONTROL WITH MIXED BOUNDARY CONDITIONS

From (7.2.18) we infer

$$
\chi_k \| \bar{u} - u_h \|^2_{\Omega} + \| \bar{y}(u) - y_h(u_h) \|^2_{\Omega'} \\
\leq \chi_k \left( \langle u^{k,i-1}, (v - u_h) + (v_h - \bar{u}) \rangle_{\Omega} \\
+ \langle \bar{y}(\bar{u}), (\bar{y}(v) - \bar{y}(u_h)) + (\bar{y}(v_h) - \bar{y}(u)) \rangle_{\Omega'} \\
+ \langle y_h(u_h) - y_h(u_h), (y_h(v_h) - y_h(v)) \rangle_{\Omega'} \right) \\
+ \langle (y_h(u_h) - \bar{y}(\bar{u})), (y_h(v_h) - \bar{y}(v_h)) \rangle_{\Omega'} \ .
$$

(7.2.21)

In Lemma 7.2.11 inequality (7.2.21) is basic for estimating the discretization error in the norm (7.2.10). Moreover, the terms in (7.2.21) require a computation of the approximation error of the control variables in the $L^2$-norm, which is only possible if an upper bound for $\| \bar{u} \|_{1,\Omega_j}$, $j \in I$ is known (cf. [286, 382]). These computations can be done via Lemma 7.2.6, which is a generalization of Lemma 2 in [108] for Problem (7.2.11). In particular this requires that the prox-point $u^{k,i-1}$ is discretized on the same or on the previous coarser grid. Hence, it is important that the grids are exactly nested one into the other, see Figure 7.2.3.

![Nested triangulations](image)

Figure 7.2.3: Nested triangulations

7.2.6 Lemma. If the function $u^{k,i-1}$ belongs to the space $\bigoplus_{j \in I} W^1(\Omega_j)$, then it holds with (7.2.24) \footnote{The sets $W^1(\Omega_1), W^1(\Omega_2)$, can be considered as subspaces of $U$ if the corresponding functions are extended with values zero on the whole domain $\Omega$.}

$$
\sum_{j \in I} \| \bar{u} \|^2_{1,\Omega_j} \leq \frac{2c^2(1 + c_f^2)}{\chi_k^2} + 2\| u^{k,i-1} \|^2_{\Omega} + \| \xi_0 \|^2_{1,\Omega} + \| \xi_1 \|^2_{1,\Omega} .
$$

Proof: Regularity of the solution $\bar{u}$ of the auxiliary Problem (7.2.11) can be shown by standard techniques (see Lions [269]). Function $\bar{u}$ is continuous and
its graph consists of tree parts:

\[
\bar{u}(x) = \begin{cases} 
  u^{k,i-1}(x) - \frac{\bar{p}(x)}{\lambda_k} & \text{if } \xi_0(x) \leq u^{k,i-1}(x) - \frac{\bar{p}(x)}{\lambda_k} \leq \xi_1(x), \\
  \xi_0(x) & \text{if } u^{k,i-1}(x) - \frac{\bar{p}(x)}{\lambda_k} < \xi_0(x), \\
  \xi_1(x) & \text{if } u^{k,i-1}(x) - \frac{\bar{p}(x)}{\lambda_k} > \xi_1(x),
\end{cases}
\]

where \( \bar{p}(x) \) denotes the solution of the adjoint state equation. Because for Hilbert spaces it holds that the double-dual space is the original space, i.e. \( Y'' = Y \), and the generalized Laplace operator is self-adjoint, i.e. \( (-\Delta)^* = -\Delta \) in \( L(Y,Y') \), the function \( \bar{p}(x) \) can be considered as classical solution of the Poisson problem with mixed boundary conditions

\[
\Delta \bar{p}(x) = \bar{y}(x) \quad \text{in } \Omega, \\
\frac{\partial \bar{p}}{\partial n_m}(x) = 0 \quad \text{on } \Gamma_m, \ m \in M_N, \\
\bar{p}(x) = 0 \quad \text{on } \Gamma_m, \ m \in M_D.
\]

Obviously with Theorem 7.2.1 it holds that \( \bar{p} \in Y \cap W^{2,1}(\Omega) \). Analogous to [108] the domain \( \Omega \) can be decomposed according to (7.2.22) into tree parts. We carry out this decomposition locally on each \( \Omega_j, \ j \in I \):

\[
\|\bar{u}\|^2_{1,\Omega_j} \leq \left( u^{k,i-1} - \frac{\bar{p}}{\lambda_k} \right)^2_{1,\Omega_j} + \|\xi_0\|^2_{1,\Omega_j} + \|\xi_1\|^2_{1,\Omega_j}.
\]

Dealing with nested grids, the values \( \|u^{k,i-1}\|_{1,\Omega_j}, \ j \in I \), exists and because the distributed control is discretized by means of piece-wise constant basic functions it holds \( \|u^{k,i-1}\|_{1,\Omega_j} = \|u^{k,i-1}\|_{1,\Omega_j} \).

The estimate for the adjoint state function \( \bar{p}(x) \) can be picked up from Lemma 7.2.7 below, hence the proof is complete.

\[
7.2.7 \text{ Lemma. } [357]
\]

The functions \( \bar{y}(x) \) and \( \bar{p}(x) \) are bounded and it holds

\[
\|\bar{y}\|_\Omega \leq \bar{c}_y, \quad \|\bar{p}\|_{1,\Omega} \leq \bar{c}_y \sqrt{1 + \bar{c}_p^2},
\]

with

\[
\bar{c}_y = c_F^2 \left( \|f\|_\Omega + \ sup_{v \in U_{ad}} \|v\|_\Omega + c_\gamma \sqrt{1 + c_F^2} \sum_{m \in M^N} \|g_m\|_{\Gamma_m} \right).
\]

The next tree lemmata correspond to Lemmata 3, 4, and 5 in [108]. These results are classical tools for analyzing inner approximations of convex sets in \( L_2 \)-spaces. Because one deals only with the existence of some constants a majorizing constant \( C \) is used.

\[
7.2.8 \text{ Lemma. } \text{Let } \bar{u} \text{ be the solution of Problem (7.2.11) and } v_h \text{ let be given by}
\]

\[
v_h(x) = \sum_{j \in I} M_j(\bar{u}) \psi_j(x),
\]
7.2. ELLIPTIC CONTROL WITH MIXED BOUNDARY CONDITIONS

with $M_j(\cdot)$ via (7.2.12).

Then it holds $v_h \in \mathcal{U}_{ad}$ and there exists a constant $C$, independent from $\bar{u}$ and $h$, such that

$$\|\bar{u} - v_h\|_\Omega \leq Ch \sqrt{\sum_{j \in I} \|\bar{u}\|_{1,\Omega_j}^2},$$  \hspace{1cm} (7.2.25)

$$\|\bar{u} - v_h\|_{-1,\Omega} \leq Ch^2 \sqrt{\sum_{j \in I} \|\bar{u}\|_{1,\Omega_j}^2},$$  \hspace{1cm} (7.2.26)

$$\sqrt{\sum_{j \in I} \|\bar{u} - v_h\|_{2,\Omega_j}^2} \leq Ch^2 \sqrt{\sum_{j \in I} \|\bar{u}\|_{1,\Omega_j}^2}.  \hspace{1cm} (7.2.27)$$

7.2.9 Lemma. Let $v_h$ be an arbitrary element belonging to $\mathcal{U}_{ad}$, then there exists a function $v^* \in \mathcal{U}_{ad} \cap \bigoplus_{j \in I} W^1(\Omega_j)$, such that

$$M_j(v^*) = M_j(v_h), \quad \forall j \in I$$

and

$$\|v^*\|_{1,\Omega_j}^2 \leq \|\xi_0\|_{1,\Omega_j}^2 + \|\xi_1\|_{1,\Omega_j}^2, \quad \forall j \in I.$$  \hspace{1cm} (7.2.28)

7.2.10 Lemma. Let $u_h$ be the solution of Problem (7.2.13) and $v$ corresponds to $v^*$ in Lemma 7.2.9. Then there exists a constant $C$, independent from $\xi_0$, $\xi_1$ and $h$, such that

$$\|u_h - v\|_{-1,\Omega} \leq Ch \sqrt{\|\xi_0\|_{1,\Omega}^2 + \|\xi_1\|_{1,\Omega}^2},$$  \hspace{1cm} (7.2.29)

$$\sqrt{\sum_{j \in I} \|u_h - v\|_{2,\Omega_j}^2} \leq Ch^2 \sqrt{\|\xi_0\|_{1,\Omega}^2 + \|\xi_1\|_{1,\Omega}^2}.  \hspace{1cm} (7.2.30)$$

The following lemma makes use of the previous Lemmata 7.2.8, 7.2.9 and 7.2.10.

7.2.11 Lemma. On each iteration of the MSR-method 7.2.4 the following relations are true

$$\|u^{k,i} - \bar{u}^{k,i}\|_\Omega \leq \epsilon_k,$$  \hspace{1cm} (7.2.31)

with

$$\epsilon_k = \sqrt{\frac{2\eta_k}{\chi_k}} + \sqrt{\frac{2\eta_k}{\chi_k}}.  \hspace{1cm} (7.2.31)$$
and

\[ o_k = \max_{1 \leq i \leq \xi(k)} \left\{ C \chi_k h_k^2 \left( \frac{C}{\chi_k} + \|u^{k,i-1}\|^2_{\Omega} + C_1 \right) \right\} \]

\[ + C \chi_k h_k^2 \left( \|u^{k,i-1}\|^2_{\Omega} + \sqrt{\frac{C}{\chi_k}} + \|u^{k,i-1}\|^2_{\Omega} + C_1 + C_2 \right) \]

\[ \times \left( \sqrt{C_1} + \sqrt{\frac{C}{\chi_k}} + \|u^{k,i-1}\|^2_{\Omega} + C_1 \right) \]

\[ + C h_k^4 \left( \frac{C}{\chi_k} + \|u^{k,i-1}\|^2_{\Omega} + C_1 \right) + (C_3 h_k^2)^2 + C_4 h_k^2 \}

(7.2.32)

The constants \( C, C_1, \ldots, C_4 \) are independent from \( \|u^{k,i-1}\|_{\Omega}, \chi_k, \eta_k \) and \( h_k \).

**Proof:** We have to analyze the amount of the error (7.2.30) in dependence of the given approximate \( u^{k,i-1} \) and of the controlling parameters. Therefore, we continue with the investigation of inequality (7.2.21). Following the proof technique in [108], the first term on the right-hand side in (7.2.21) can be estimated via the generalized Cauchy–Schwarz inequality. It holds for instance for any \( j \in \mathcal{I} \)

\[ \langle u^{k,i-1}, v - u_h \rangle_{\Omega_j} \leq \|u^{k,i-1}\|_{1,\Omega_j} \|v - u_h\|_{-1,\Omega_j} \]

Adding up such inequalities we get from (7.2.21) the following relation

\[ \chi_k \| \bar{u} - u_h \|^2_{\Omega_j} + \| \bar{g}(\bar{u}) - g_h(u_h) \|^2_{\Omega_j} \]

\[ \leq \chi_k \sum_{j \in \mathcal{I}} \left( \|u^{k,i-1}\|_{1,\Omega_j} \|v - u_h\|_{-1,\Omega_j} + \|u^{k,i-1}\|_{1,\Omega_j} \|v_h - \bar{u}\|_{-1,\Omega_j} \right) \]

\[ + \chi_k \left( \langle \bar{u}, v - u_h \rangle_{\Omega} + \langle \bar{u}, v_h - \bar{u} \rangle_{\Omega} + \langle u_h - \bar{u}, v_h - \bar{u} \rangle_{\Omega} \right) \]

\[ + \langle \bar{g}(\bar{u}), (\bar{g}(v) - g_h(u_h)) \rangle_{\Omega} + \langle g_h(v_h) - \bar{g}(\bar{u}) \rangle_{\Omega} \]

\[ + \langle g_h(v_h) - \bar{g}(\bar{u}), (g_h(v_h) - \bar{g}(v_h)) \rangle_{\Omega} \]

(7.2.33)

The following estimates hold for an affine mapping \( \bar{g}(\cdot) : \mathcal{Y}' \rightarrow \mathcal{Y} \) for all \( v \in \mathcal{U} \) and all \( u_h \in \mathcal{U}^h \), (cf. [108], page 44)

\[ \| \bar{g}(v) - g_h(u_h) \|^2_{1,\Omega} \leq (1 + c_F^2) \| v - u_h \|_{-1,\Omega} \]

\[ \| \bar{g}(v_h) - \bar{g}(\bar{u}) \|^2_{1,\Omega} \leq (1 + c_F^2) \| v_h - \bar{u} \|_{-1,\Omega} \]

Using the inequalities

\[ 2 \langle u_h - \bar{u}, v_h - \bar{u} \rangle_{\Omega} \leq \|u_h - \bar{u}\|^2_{\Omega} + \|v_h - \bar{u}\|^2_{\Omega} \]

and

\[ 2 \langle y_h(u_h) - \bar{g}(\bar{u}), (y_h(v_h) - \bar{g}(v_h)) + (\bar{g}(v_h) - \bar{g}(\bar{u})) \rangle_{\Omega'} \]

\[ \leq 2 \|y_h(u_h) - \bar{g}(\bar{u})\|^2_{\Omega'} + \|y_h(v_h) - \bar{g}(v_h)\|^2_{\Omega'} + \|\bar{g}(v_h) - \bar{g}(\bar{u})\|^2_{\Omega'} \]
we obtain in (7.2.33)

\[ \chi_k \| \bar{u} - u_h \|_{\Omega}^2 + \| \bar{y}(\bar{u}) - y_h(u_h) \|_{\Omega}^2 \]

\[ \leq \chi_k \sum_{j \in I} \left( \| v - u_h \|_{-1, \Omega_i} + \| v_h - \bar{u} \|_{-1, \Omega_i} \right) + \chi_k \sum_{j \in I} \| \bar{u} \|_{1, \Omega_i} \left( \| v - u_h \|_{-1, \Omega_i} + \| v_h - \bar{u} \|_{-1, \Omega_i} \right) + \frac{\chi_k}{2} \left( \| v_h - \bar{u} \|_{\Omega}^2 + \| v_h - \bar{u} \|_{\Omega}^2 \right)
\]

\[ + \left( \| \bar{y}(\bar{u}) \|_{\Omega} \left( \| v - u_h \|_{-1, \Omega_i} + \| v_h - \bar{u} \|_{-1, \Omega_i} \right) + \| y_h(u_h) - \bar{y}(u_h) \|_{\Omega} \right) + \| y_h(u_h) - \bar{y}(u_h) \|_{\Omega} \right) \]

(7.2.34)

Now we have to bear in mind that the values \( \| u^{k,i-1} \|_{\Omega_j} \) and \( \| u^{k,i-1} \|_{1, \Omega_j} \) coincide. The first and second term on the right-hand side in (7.2.34) allows us to apply the Cauchy–Schwarz inequality for finite sums. In the fourth term we observe that the \( L_2 \)-norms can be majorized by \( W^1 \)-norms. After all transformations we have for all \( v \in U_{ad} \) and \( v_h \in U_{ad}^{h} \)

\[ \frac{\chi_k}{2} \| \bar{u} - u_h \|_{\Omega}^2 \leq \]

\[ \leq \frac{\chi_k}{2} \| v_h - \bar{u} \|_{\Omega}^2 + \left( \chi_k \| u^{k,i-1} \|_{\Omega} + \chi_k \sum_{j \in I} \| \bar{u} \|_{\Omega_j}^2 \right) T_\Delta \]

\[ + C \| \bar{y}(\bar{u}) \|_{\Omega} \left( \| v - u_h \|_{-1, \Omega_i} + \| v_h - \bar{u} \|_{-1, \Omega_i} \right) + \frac{1}{2} \| y_h(v_h) - \bar{y}(v_h) \|_{\Omega}^2 + \frac{C^2}{2} \| v_h - \bar{u} \|_{-1, \Omega_i}^2 \]

(7.2.35)

with

\[ T_\Delta = \sqrt{\sum_{j \in I} \| v - u_h \|_{-1, \Omega_i}^2} + \sqrt{\sum_{j \in I} \| v_h - \bar{u} \|_{-1, \Omega_i}^2} \]

In (7.2.35) we set due to Lemma 7.2.10 \( v := v^*(u_h) \in U_{ad} \) and in view of Lemma 7.2.8

\[ v_h := \sum_{j \in I} M_j(\bar{u}) \psi_j \in U_{ad}^{h} \]

Applying to the last terms in (7.2.35) the estimates of the discretization error
of the state equation (7.2.8), we obtain
\[
\frac{\chi_k}{2} \| \bar{u} - u_h \|^2_{\Omega} \leq \frac{\chi_k}{2} C h_k^2 \sum_{j \in I} \| \bar{u} \|^2_{1, \Omega_j} \\
+ \left( \chi_k \| u^{k,i-1} \|_{\Omega} + \chi_k \sqrt{\sum_{j \in I} \| \bar{u} \|^2_{1, \Omega_j} + C \| \bar{y} \|_{\Omega}} \right) \\
\times \left( C h_k^2 \sqrt{\| \xi_0 \|^2_{I, \Omega} + \| \xi_1 \|^2_{I, \Omega} + C h_k^2 \sqrt{\sum_{j \in I} \| \bar{u} \|^2_{1, \Omega_j}} \right) \\
+ C h_k \sum_{j \in I} \| \bar{u} \|^2_{1, \Omega_j} + (Ch_k^2 \| f + v_h \|_{\Omega} + Ch_k^2 T_{\Sigma})^2 \\
+ \| \bar{y} \|_{\Omega} C h_k^2 (\| f + v_h \|_{\Omega} + \| f + u_h \|_{\Omega} + 2T_{\Sigma})
\]

with
\[
T_{\Sigma} = \sum_{m \in M_N} \| g_m \|_{1/2, r_m} + \\
+ \sum_{m \in M_A} T_{m}^{1/2}(g_m) + \sum_{m \in M_B} T_{m}^{1/2}(g_{m+1}) + \sum_{m \in M_C} T_{m}^{1/2}(g_m, g_{m+1}).
\]

Analogous to [108], page 45, we get the estimates
\[
\| u_h \|_{\Omega} \leq \sqrt{\| \xi_0 \|^2_{I, \Omega} + \| \xi_1 \|^2_{I, \Omega}}, \quad \| v_h \|_{\Omega} \leq \sqrt{\| \xi_0 \|^2_{I, \Omega} + \| \xi_1 \|^2_{I, \Omega}}.
\]
The terms \( \| \bar{u} \|_{1, \Omega_j} \) in (7.2.36) can be estimated via Lemma 7.2.6 and \( \| \bar{y} \|_{\Omega} \) via (7.2.24), respectively. Finally, with new constants \( C, C', C \) we obtain
\[
\frac{\chi_k}{2} \| \bar{u} - u_h \|^2_{\Omega} \leq C \chi_k h_k^2 \left( \frac{C}{\chi_k} + \| u^{k,i-1} \|^2_{\Omega} + C \right) \\
+ C h_k^2 \left( \chi_k \| u^{k,i-1} \|_{\Omega} + \chi_k \sqrt{\frac{C}{\chi_k} + \| u^{k,i-1} \|^2_{\Omega} + C + C \bar{y}} \right) \\
\times \left( \sqrt{C} + \sqrt{\frac{C}{\chi_k} + \| u^{k,i-1} \|^2_{\Omega} + C} \right) \\
+ C h_k \left( \frac{C}{\chi_k} + \| u^{k,i-1} \|^2_{\Omega} + C \right) + (C' h_k^2)^2 + \bar{c}_p C' h_k^2.
\] (7.2.37)

It is not difficult to see that the definition of \( o_k \) in (7.2.32) is based on (7.2.37). Term (7.2.31) for the error \( \epsilon_k \) in the auxiliary problem of the MSR-method can be calculated via Lemma 7.2.5 by means of the triangle inequality. From (7.2.37) it follows
\[
\epsilon_k = \sqrt{\frac{2 o_k}{\chi_k}} + \sqrt{\frac{2 \eta_k}{\chi_k}} \geq \| u^{k,i} - u^{k,i}_h \|_{\Omega} + \| u^{k,i} - u^{k,i}_h \|_{\Omega} \geq \| u^{k,i} - u^{k,i}_h \|_{\Omega}.
\]

\[\square\]

Now all preliminary results are provided and we are going to formulate the convergence result of the MSR-method.
7.2.12 Theorem. The sequence of iterates \( z^{k,i} = (y^{k,i}, u^{k,i}) \) in MSR-Algorithm 7.2.4 has always a finite number of elements in the internal loop \( i \), i.e.,

\[
i(k) \leq 2d_2/\vartheta_k
\]

(7.2.38)

and \( \{z^{k,i}\} \) converges weakly in the space \( W^1(\Omega) \times L^2(\Omega) \) to some element of the solution set \( Z^* \), if

\[
0 \leq \chi < \chi_k \leq \bar{\chi}, \quad 0 < \vartheta_k \leq \bar{\vartheta},
\]

\[
\sum_{k=1}^{\infty} \frac{\eta_k}{\chi_k} < \infty, \quad \sum_{k=1}^{\infty} \frac{h_k}{\chi_k} < \infty,
\]

(7.2.39)

\[
\delta_k \geq \epsilon_k + \sqrt{2(d_2 + \epsilon_k)(\epsilon_k + \vartheta_k)},
\]

(7.2.40)

where \( \epsilon_k \) is defined via Lemma 7.2.11 and

\[
d_2 = 2 \sup_{v \in \mathcal{U}_d} \|v\|_\Omega.
\]

(7.2.41)

Proof: Choosing the parameter \( \delta_k \) in MSR-algorithm 7.2.4 sufficiently large, then

\[
\|\tilde{z}^{k,i} - u^{k,i-1}\|_\Omega \geq \|u^{k,i} - \tilde{u}^{k,i}\|_\Omega - \|u^{k,i} - \tilde{u}^{k,i}\|_\Omega \geq (\delta_k - \epsilon_k) > 0.
\]

(7.2.42)

We apply Lemma 8 in [357] to the \( k \)-th iteration of the MSR-method. As space \( \mathcal{X} \) we chose the process space \( Z \), i.e. \( \|\cdot\|_\mathcal{X} := \|\cdot\|_Z \). Moreover, the following substitutions are necessary:

\[
\begin{align*}
X_1 & := \{z = (y, u) \in Z : y = 0 \text{ in } \mathcal{Y}\}, & \Phi(z) & := 2J(z), \\
b(z_1, z_2) & := 0, & a(z_1, z_2) & := \langle y_1, y_2 \rangle_{\mathcal{Y}'}, & \ell(z) & := 0, \\
G & := \{(y, u) \in \mathcal{Y} \times \mathcal{U}_{ad} : -\Delta y = u + f + \sum_{m \in \mathcal{M}_N} g_m \delta_{m \tau_m} \text{ in } \mathcal{Y}'\}, \\
a^0 & := \tilde{z}^{k,i-1}, & a^1 & := z^{k,i}, & x & := z^*, & \delta & := \delta_k - \epsilon_k, & \chi & := \chi_k.
\end{align*}
\]

The application of this Lemma needs to show the closedness and convexity of \( G \) in space \( \mathcal{X} = \mathcal{Z} = \mathcal{Y} \times \mathcal{U} \) has to be shown. By means of the state equation in \( G \) an affine closed manifold in \( \mathcal{Y} \times \mathcal{Y} \) is generated, because the generalized Laplace operator defines an isometric isomorphism between \( \mathcal{Y} \) and \( \mathcal{Y}' \). Therefore, the intersection of this manifold with the topological stronger space \( Z \) in \( \mathcal{X} \) is closed. The closedness of the whole set \( G \) holds due to the additional intersection of the affine set with the closed (in \( \mathcal{X} \)) set \( \mathcal{Y} \times \mathcal{U}_{ad} \). Moreover, these two sets are convex, hence \( G \) is convex. Due to [357], formula (2.124), we get a norm-equivalence:

\[
|\langle y, u \rangle|_{\mathcal{X}} = \|u\|_\Omega.
\]

(7.2.43)

In this case the required equivalence between \( |\cdot|_{\mathcal{X}} \) and \( \|\cdot\|_{\mathcal{X}} \) (cf. the previous chapter) is not necessary, because the convergence proof is carried out directly for the half-norm.
Since $\Phi(x) = 2J(z^*) \leq 2J(\hat{z}^{k,i}) = \Phi(a^1)$, we have $\eta(x) = 0$ in Lemma 8 in [357]. In view of [357], formula (2.128), we have

$$|z^{k,i} - z^*|_X \leq |z^{k,i-1} - z^*|_X - \frac{(\delta_k - \epsilon_k)^2}{2(d_2 + \epsilon_k)}$$ \hfill (7.2.44) 

The denominator in (7.2.44) can be calculated due to (7.2.41) in the following way

$$|z^{k,i-1} - z^*|_X \leq |z^{k,i-1} - z^{k,i-1}|_X + |z^{k,i-1} - z^*|_X \leq \epsilon_k + \sup_{(v_1, v_2) \in U_2} \|v_1 - v_2\|_\Omega = \epsilon_k + d_2.$$ 

With (7.2.43) and (7.2.44) one get for $i = 1, \ldots, i(k) - 1$

$$|z^{k,i} - z^*|_X - |z^{k,i-1} - z^*|_X \leq |\hat{z}^{k,i} - z^{k,i}|_X - \frac{(\delta_k - \epsilon_k)^2}{2(d_2 + \epsilon_k)} \leq \epsilon_k - \frac{(\delta_k - \epsilon_k)^2}{2(d_2 + \epsilon_k)} \leq \eta_k < 0.$$ \hfill (7.2.45) 

Hence, the values of $|z^{k,i} - z^*|_X$ can be majorized by an decreasing arithmetic series (in $i$). Because $\eta_k > 0$ and $|z^{k,i} - z^*|_X \leq d_2 + \epsilon_k$, we have a finite number of $i$-steps and estimate (7.2.38) holds true. At the same time we get via (7.2.45) a lower bound for $\delta_k$, namely (7.2.40). In view of [357], formula (2.127), with $\eta(x) = 0$ it holds

$$|z^{k,i} - z^*|_X - |z^{k,i-1} - z^*|_X \leq \epsilon_k.$$ \hfill (7.2.46) 

Due to (7.2.45) $\{|z^{k,i} - z^*|_X\}$ is monotonously decreasing for $i = 0, \ldots, i(k) - 1$ and in (7.2.46) we can carry on for all $i = 1, \ldots, i(k)$. Because of the identity $z^{k+1,0} = z^{k,i(k)}$ for all $k$, we get the following estimate which is independent from $i$

$$|z^{k+1,0} - z^*|_X - |z^{k,0} - z^*|_X \leq \epsilon_k.$$ \hfill (7.2.47) 

If $\sum_{k=1}^\infty \epsilon_k < \infty$, then with POLYAK’S Lemma A3.1.7 we get the convergence (in $k$) of $|z^{k,0} - z^*|_X$ towards a finite positive value. The sequence of exact solutions $\hat{z}^{k,0} = z^{k,i(k)}$ for all $k$, we get the following estimate which is independent from $i$

$$|z^{k,i(k)} - z^*|_X - |z^{k,i(k)-1} - z^*|_X \leq \frac{4}{\chi_k} \left( J(z^*) - J(z^{k,i(k)}) \right).$$ \hfill (7.2.48) 

and after backward tracking of the index $i$ as in (7.2.46), and $z^{k,0} = z^{k-1,i(k-1)},$

\footnote{Here it is important that the set $G$ is bounded in the space $Z$.}
we get the following relations between negative values
\[ J(z^*) - J(z^{k+1.0}) = J(z^*) - J(z^{k,i(k)}) \]
\[ \geq \frac{\chi^k}{4} \left( |z^{k,i(k)}| - z^*| \chi + |z^{k,i(k)} - 1 - z^*| \chi \right) \left( |z^{k,i(k)} - z^*| \chi - |z^{k,i(k)} - 1 - z^*| \chi \right) \]
\[ \geq \frac{\chi}{4} \left( |z^{k,i(k)} - z^*| \chi + |z^{k,i(k)} - 1 - z^*| \chi \right) \left( |z^{k+1.0} - z^*| \chi - |z^{k,0} - z^*| \chi \right) \]
\[ \geq \frac{\chi}{4} \left( d_2 + d_2 + \epsilon_k \right) \left( |z^{k+1.0} - z^*| \chi - |z^{k,0} - z^*| \chi \right) . \]

The terms in the last brackets are equal in the limit and with \( k_\epsilon \to \infty \) we get
\[ J(z^*) \geq J(z_3) \], because \( J \) is weakly lsc. Hence \( z_3 = (y_3, u_3) \in Z^* = Y^* \times U^* \). The subsequence \( \{z^{k,0}\} \) converges weakly towards the \( y_3 \), because of \( \epsilon_k \to 0 \).

Finally we apply Opial’s Lemma A1.1.3 to the sequence \( u^{k,0} \) and the set \( U^* \), which gives the weak convergence of the control iterate in the MSR-algorithm towards \( u_3 \).

7.2.2.1 Strong convergence of the state component

Now we can prove strong convergence of the sequence \( \{y^{k,0}\} \) to \( y_3 \) in the \( W^1 \)-norm. The triangle inequality gives us
\[ \|y^{k,0} - y_3\|_{1,\Omega} \leq \|y^{k,0} - y^{k,0}\|_{1,\Omega} + \|\tilde{y}(\tilde{u}^{k,0}) - \tilde{y}(u_3)\|_{1,\Omega} \]
\[ \leq \|\tilde{y}(\tilde{u}) - \tilde{y}(u_h)\|_{1,\Omega} + \|\tilde{y}(u_h) - \tilde{y}(u_3)\|_{1,\Omega} + \|y^{k,i} - y_h^{k,i}\|_{1,\Omega} + \|\tilde{y}(\tilde{u}) - \tilde{y}(u_3)\|_{1,\Omega} \]
\[ \leq \sqrt{1 + c_F^2} \|\tilde{u} - u_h\|_{-1,\Omega} + C_5 h_k^2 \]
\[ \leq C_6 \sqrt{\frac{2 \eta_k}{\lambda_k}} + C_5 h_k^2 + \sqrt{1 + c_F^2} \sqrt{\frac{2 \eta_k}{\lambda_k}} \]
\[ + (1 + c_F^2)^{3/2} \|\tilde{u}^{k,0} - u^{k,0}\|_{1,\Omega} + (1 + c_F^2)^{3/2} \|\tilde{u}^{k,0} - u_3\|_{-1,\Omega} . \]

Term \( C_5 h_k^2 \) yields due to (7.2.8).

It is sufficient to show that the last norm-term in the chain of inequalities above tend to zero if \( k \to \infty \). To do that we define new elements \( y_3^k \) in \( W^1(\Omega) \) by means of the variational inequality
\[ \langle y_3^k, \phi \rangle_{1,\Omega} = \langle u^{k,0} - u_3, \phi \rangle_{\Omega}, \quad \forall \phi \in W^1(\Omega). \]

Setting \( \phi := y_3^k \) and taking limit as \( k \to \infty \), we obtain
\[ \|y_3^k\|_{1,\Omega}^2 = \langle u^{k,0} - u_3, y_3^k \rangle_{\Omega} \to 0, \]
CHAPTER 7. PARTIAL PPR IN CONTROL PROBLEMS

because of \( u^{k,0} \rightarrow u_3 \). This means

\[
\|u^{k,0} - u_3\|_{-1,\Omega} = \sup_{\phi \in C^\infty(\overline{\Omega}) \setminus \{0\}} \frac{\langle u^{k,0} - u_3, \phi \rangle_\Omega}{\|\phi\|_{1,\Omega}} \leq \|y_3^k\|_{1,\Omega} \rightarrow 0.
\]

To obtain the convergence conditions (7.2.39) we have to take the limit \( h_k \rightarrow 0 \) in (7.2.31), (7.2.32). The weakest conditions yield under the assumption \( \chi = 0 \).

With Hardy’s \( O \)-notation let us study the terms in (7.2.31) and (7.2.32):

\[
\varepsilon_k \leq O \left( \sqrt{\frac{\eta_k}{\lambda_k}} \right) + O \left( \frac{\eta_k}{\chi_k} \right)
\]

\[
= O \left( \sqrt{\frac{\eta_k}{\chi_k}} \right) + \frac{1}{\sqrt{\chi_k}} \left[ O \left( \frac{h_k^2}{\lambda_k} \right) + O \left( \frac{\chi_k h_k^2}{\lambda_k} \|u^{k,i-1}\|_\Omega^2 \right) + O \left( \frac{1}{\chi_k} \right) \right]
\]

\[
+ O \left( \frac{h_k^2}{\lambda_k} \right) + O \left( \frac{h_k^4}{\lambda_k^2} \right) \bigg[ O \left( \frac{1}{\chi_k} \right) + O \left( \frac{\|u^{k,i-1}\|_\Omega^2}{\chi_k} \right) \bigg]^{1/2}.
\]

It can easily be seen that the underlined terms are majorized and that from (7.2.39) it follows convergence of \( \sum_{k=0}^{\infty} \varepsilon_k \).

Theorem 7.2.12 gives an \textit{a priori} criteria for the convergence of the inner approximation of the MSR-method with respect to the control space. For the first glance this result is common knowledge and for the treatment of ill-posed control problems very important, because in that case there are no stopping criteria or other indicators at hand to prove the convergence. Mostly the solution set is quite instable. It may happen that the sequence of approximations is going to miss the right landing or there is no convergence towards some element at all. But if the conditions of Theorem 7.2.12 hold, both disadvantages cannot happen and, therefore, one could apply traditional criteria of strong convergence, like the boundedness of the sum of distances between neighbored iterates in the norm of the process space. This would give also an \textit{a posteriori} criteria for the strong convergence.

For the weakest assumption under which it is possible to formulate an \textit{a priori} criteria for the strong convergence we refer to [217].

7.2.3 Numerical results

The numerical investigation of the MSR-algorithms 7.2.4 is performed with a simple model of type (7.2.9):

\[
\min_{(y,u) \in \mathcal{Y} \times \mathcal{U}_{ad}} \{ J(y, u) = \frac{1}{2} \|y\|_{\mathcal{Y}}^2 \} \quad (7.2.49)
\]

\[-\Delta y = u + g\delta \Gamma \quad \text{in} \ \mathcal{Y} \]
with
\[ \mathcal{Y} = \left\{ y \in W^1(\Omega) : \gamma_0 y = 0 \text{ in } H^{1/2}(\Gamma_D) \right\}, \]
\[ \mathcal{K} = \mathcal{U} = L^2(\Omega), \quad \mathcal{U}_{ad} = \left\{ u \in \mathcal{U} : -1 \leq u(x) \leq 1 \text{ a.a. in } \Omega \right\}, \]
where \( \Omega \) is a square domain and \( \Gamma = \partial \Omega \setminus \Gamma_D \) is one side of the square. The sides with Dirichlet boundary conditions are drawn in Figure 7.2.4 with double lines.

Here we are dealing with an "optimal reflection" of a delta layer due to a distributed control \( u(x) \), whose amplitude is bounded from above and below. Surely there exist different applications of this formulation. The most simple physical meaning consist in the observation of a homogeneous squared element \( \Omega \) which is heated from one side \( \Gamma = \partial \Omega \setminus \Gamma_D \) by a fixed source of heat. The temperature at the other sides \( \Gamma_D \) is fixed to the zero-level. In this situation one is looking for a cooling regime on the element \( \Omega \) such that the deviation of the potential of the temperature on the square \( \Omega \) to the zero-level on the boundary \( \Gamma_D \) is as small as possible. The construction of the cooling system is determined only by the form of the switching lines of the optimal control function \( u^*(t) \).

Hence, Problem (7.2.49) describes a two-dimensional analogue of the classical Fuller problem and one can expect that there is an infinite number of switching lines in the optimal solution. Although this problem has a unique solution, it belongs to the class of ill-posed problems.

The numerical reconstruction of such a solution becomes very difficult because the smallest bang-bang formations in \( u^*(x) \) between the neighboring switches becomes extremely instable. An example for such a solution for \( g(x_2) = 0.7 \cos \pi x_2 \) is shown in Figure 7.2.5.

The discrete auxiliary problems of type (7.2.13) can be formulated as quadratic minimization problems in finite-dimensional spaces. The variables which have to be optimized are the weights of the finite basic functions. The...
Figure 7.2.5: "Chattering regime" on a square for an optimal $u^*(x_1, x_2)$.

Figure 7.2.6: Approximation of $u^{k,i(k)}(x_1, x_2)$ on a sequence of grids: $h = 2^{-6}$,
$h = 2^{-7}$, $h = 2^{-8}$
optimal weights determine the sought approximate solutions $y^{k,i}$ and $u^{k,i}$.

The main idea of this approach is to show the interaction of discretization and regularization. This can be observed clearly on Figure 7.2.6, where it is shown how the fourth instable bang-bang formation can be reproduced on a sequence of grids. Hereby the parameter $\chi_k$ has to be slowly decreased, following the updating rule (7.2.39).

In the one-dimensional Figures 7.2.7 and 7.2.8 another phenomenon can be observed for Problem (7.2.49): On the same discretization level and under the same updating rule of the regularization parameter the MSR-method delivers better results than Tichonov regularization.

7.3 Comments

In the customary ideology to handle ill-posed optimal control problems mostly the object of the control is an evolution system and the non-uniqueness of the quasi-optimal control is believed to be an advantage. In contrast to that, the here described approaches require a careful analysis of convergence not only for the objective values but also for the state and control components of the solutions.

Usually the convergence of numerical algorithms for control problems is studied under the additional assumption that the objective functional is strong convex with respect to the control variable and that the control function possesses a classic bang-bang behavior (cf., for instance, Alt and Mackenroth...
The main result hereby is the proof of the convergence of the approximate solution to the unique solution of the original problem in some given Banach spaces. However, an analogous result for ill-posed control problems is approachable only by using special regularization techniques.

Difficult is also the question of the existence of a solution, if a nonlinear state equation comes into the play and non-quadratic control costs arises. This is the reason why we restricted our investigation to linear state equations only. It should be mentioned that for nonlinear problems a number of results can be found, for instance, by Kelley and Sachs, Heinkenschloss and Sachs, and Kupfer and Sachs. In this case the validity of sufficient optimality conditions of second order are substantial.

The first regularized method for ill-posed linear control problems was suggested by Tikhonov. Penalty techniques applied to control problems can be found first by Lions and Balakrishnan. Embedding the state equation into the penalty term, the control and state variables can be considered as independent, hence, the control problem can be reformulated as a variational inequality. Bergounioux applied this penalty technique in order to prove under weak regularity conditions the existence of the Lagrange multipliers for elliptic and parabolic control problems with state constraints. In a regularized penalty method is described for solving ill-posed parabolic control problems.
Chapter 8

PPR FOR VARIATIONAL INEQUALITIES

8.1 Approximation of Variational Inequalities

The subsequent introduction is dispensable for understanding the main results of this chapter. Nevertheless, it makes the reader acquainted with some peculiarities of the finite-dimensional approximation of monotone (elliptic) variational inequalities stated in Hilbert spaces, which are not common in linear problems in mathematical physics.

Finite-dimensional approximations of elliptic variational inequalities are mainly performed by the usage of finite element technique. In this direction one can distinguish three approaches: approximation of the original (primal) or the dual problem, or the mixed variational formulation of the problem.

We are going to illustrate briefly these approaches for the model problem (A2.1.6), known as problem with an obstacle on the boundary. This problem will be considered also later on in detail.

For an introduction into Hilbert spaces and various notions of monotone variational inequalities considered in these spaces we refer to the Appendix-Sections A1 and A2 in the Appendix. Later on the symbols \((\cdot, \cdot)\) and \(\|\cdot\|\) provided with corresponding indices are used to denote the scalar product and the norm of a given space of vector functions.

On a given open and bounded domain \(\Omega \subset \mathbb{R}^n\) with Lipschitz continuous boundary \(\Gamma\) the space \(H^{\frac{1}{2}}(\Gamma)\) can be defined as the image of the space \(H^1(\Omega)\) under the trace mapping \(\gamma:\)

\[
H^{\frac{1}{2}}(\Gamma) := \gamma \left( H^1(\Omega) \right)
\]

endowed with the norm

\[
\|u\|_{\frac{1}{2}, \Gamma} := \inf \{\|v\|_{1, \Omega} : v \in H^1(\Omega), \gamma v = u\}.
\]

Here, \(\gamma v\) denotes the trace of the function \(v\) on \(\Gamma\). The space \(H^{\frac{1}{2}}(\Gamma)\) is a subspace of \(L_2(\Gamma)\) and \(H^{-\frac{1}{2}}(\Gamma)\) is the space of linear (continuous) functionals on \(H^{\frac{1}{2}}(\Gamma)\).
CHAPTER 8. PPR FOR VARIATIONAL INEQUALITIES

Model Problem: (cf. (A2.1.6))

\[
\begin{align*}
\text{minimize} & \quad J(u), \\
\text{subject to} & \quad u \in K,
\end{align*}
\]

with

\[
J(u) := \frac{1}{2} \|\nabla u\|_{0,\Omega}^2 - (f, u)_{0,\Omega},
\]

\[
K := \{ u \in H^1(\Omega) : \gamma u \geq g \text{ on } \Gamma \},
\]

\(\Omega \subset \mathbb{R}^n\) an open domain with sufficiently regular boundary (at least Lipschitz-continuous),

\(f \in L_2(\Omega)\) and \(g\) the trace of function \(g_0 \in H^2(\Omega)\) on \(\Gamma\).

In case \(n = 3\), under certain symmetry conditions, it is possible to reduce the problem to \(n = 2\). Physically the problem describes a stationary fluid stream in the domain \(\Omega\) which is bounded by a semi-permeable membrane \(\Gamma\). The given functions \(g\) and \(f\) define the fluid pressure on the boundary \(\Gamma\) and the fluid stream in \(\Omega\), respectively. The solution \(u\) represents the distribution of the pressure in \(\Omega\). Due to the semi-permeability of the boundary, on those parts of \(\Gamma\) where \(u(x) \leq g(x)\), the fluid can flow only into \(\Omega\), but it is impossible to flow out of \(\Omega\).

It is easy to see that, on the space \(H^1(\Omega)\), the kernel of the bilinear form

\[
a(u, v) := ((\nabla u, \nabla v))_{0,\Omega},
\]

corresponding to the functional (8.1.1), consists only of elements \(z = \text{const}\).

Solvability of Problem (8.1.2) is guaranteed under the condition

\[\langle f, 1 \rangle_{0,\Omega} \leq 0.\]

In case

\[\langle f, 1 \rangle_{0,\Omega} < 0,\]

the problem has a unique solution (cf. Glowinski, Lions and Trémoi-lières [135]). Condition (8.1.4) is obvious. Indeed, if \(\langle f, 1 \rangle_{0,\Omega} > 0\), then the volume of the fluid increases. Consequently, an equilibrium of the distribution of the pressure is impossible.

However, if \(\langle f, 1 \rangle_{0,\Omega} = 0\) and Problem (8.1.1) has a solution \(u^*\), then its solution set has the structure

\[U^* := \{ u \in H^2(\Omega) : u = u^* + c, \gamma u^* + c \geq g \text{ on } \Gamma \}.\]

In order to describe the above mentioned approximation approaches, for the seek of simplicity we restrict ourselves to the case where \(\Omega\) is an open polyhedral set in \(\mathbb{R}^3\). In this situation straightforward approximations of variational problems in \(H^1(\Omega)\) by means of finite elements are presented in Appendix A2.4.

In order to formulate the dual problem to (8.1.1), we need the space

\[H(\text{div}; \Omega) := \{ q \in [L_2(\Omega)]^n : \text{div } q \in L_2(\Omega) \}.\]
8.1. APPROXIMATION OF VARIATIONAL INEQUALITIES

with the divergence operator used in the sense of distributions

\[(q, \nabla \varphi)_{0, \Omega} := -\langle \varphi, \text{div } q \rangle_{0, \Omega} \quad \forall \varphi \in \mathcal{D}(\Omega).\]

In \(H(\text{div}; \Omega)\) the norm is given by

\[\|q\|^2_{H(\text{div}; \Omega)} := \|q\|^2_{0, \Omega} + \|\text{div } q\|^2_{0, \Omega}.\]

For each \(q \in H(\text{div}; \Omega)\) there is a unique linear functional \(qv \in H^{-\frac{1}{2}}(\Gamma)\) such that

\[\langle qv, w \rangle := ((q, \nabla v))_{0, \Omega} + \langle \text{div } q, v \rangle_{0, \Omega}, \quad (8.1.5)\]

where \(\langle \cdot, \cdot \rangle\) denotes the duality pairing between \(H^{\frac{1}{2}}(\Gamma)\) and \(H^{-\frac{1}{2}}(\Gamma)\) (see Lions and Magenes [272]) and the elements \(w \in H^{\frac{1}{2}}(\Gamma)\) and \(v \in H^1(\Omega)\) are related according to the trace mapping \(\gamma v = w\).

Non-negativity \(qv \geq 0\) on \(\Gamma\) means that

\[\langle qv, \gamma v \rangle \geq 0 \quad \forall v \in H^1(\Omega) \text{ with } \gamma v \geq 0.\]

Now, the dual problem to (8.1.1) can be formulated as

\[\min \{J(q) : q \in K\} \quad (8.1.6)\]

with

\[J(q) := \frac{1}{2} \|q\|^2_{0, \Omega} - \langle qv, g \rangle, \quad (8.1.7)\]

\[K := \{q \in H(\text{div}; \Omega) : \text{div } q + f = 0 \text{ in } \Omega, \ qv \geq 0 \text{ on } \Gamma\}. \quad (8.1.8)\]

The relation between the primal problem (8.1.1) and its dual (8.1.6) is clarified in the following statement.

8.1.1 Theorem. If the primal problem (8.1.1) is solvable, then there exists a unique solution \(q^*\) of the dual problem (8.1.6), and the equations

\[q^* = \nabla u^*, \quad J(u^*) + J(q^*) = 0 \quad (8.1.9)\]

are true for every solution of Problem (8.1.1).

This theorem holds also true for more general variational problems with convex quadratic objective functionals and constraints of the form \(Bu \leq b\), with \(B\) a linear operator in a space of functions on \(\Omega\) or \(\Gamma\).

Let us sketch its proof. Introducing new functions

\[\zeta_k := \frac{\partial u}{\partial x_k}, \quad (8.1.10)\]

Problem (8.1.1) can be reformulated as

\[\min \left\{ \frac{1}{2} \|\zeta\|^2_{0, \Omega} - \langle f, u \rangle_{0, \Omega} : \zeta_k = \frac{\partial u}{\partial x_k}, \ k = 1, \ldots, n; \ u \in K \right\}. \]

By means of the Lagrangian function

\[\mathcal{L}(u, \zeta, q) := \frac{1}{2} \|\zeta\|^2_{0, \Omega} - \langle f, u \rangle_{0, \Omega} + \sum_{k=1}^{n} \langle q_k, -\zeta_k + \frac{\partial u}{\partial x_k} \rangle_{0, \Omega} \]

8.1. APPROXIMATION OF VARIATIONAL INEQUALITIES

285

In \(H(\text{div}; \Omega)\) the norm is given by

\[\|q\|^2_{H(\text{div}; \Omega)} := \|q\|^2_{0, \Omega} + \|\text{div } q\|^2_{0, \Omega}.\]

For each \(q \in H(\text{div}; \Omega)\) there is a unique linear functional \(qv \in H^{-\frac{1}{2}}(\Gamma)\) such that

\[\langle qv, w \rangle := ((q, \nabla v))_{0, \Omega} + \langle \text{div } q, v \rangle_{0, \Omega}, \quad (8.1.5)\]

where \(\langle \cdot, \cdot \rangle\) denotes the duality pairing between \(H^{\frac{1}{2}}(\Gamma)\) and \(H^{-\frac{1}{2}}(\Gamma)\) (see Lions and Magenes [272]) and the elements \(w \in H^{\frac{1}{2}}(\Gamma)\) and \(v \in H^1(\Omega)\) are related according to the trace mapping \(\gamma v = w\).

Non-negativity \(qv \geq 0\) on \(\Gamma\) means that

\[\langle qv, \gamma v \rangle \geq 0 \quad \forall v \in H^1(\Omega) \text{ with } \gamma v \geq 0.\]

Now, the dual problem to (8.1.1) can be formulated as

\[\min \{J(q) : q \in K\} \quad (8.1.6)\]

with

\[J(q) := \frac{1}{2} \|q\|^2_{0, \Omega} - \langle qv, g \rangle, \quad (8.1.7)\]

\[K := \{q \in H(\text{div}; \Omega) : \text{div } q + f = 0 \text{ in } \Omega, \ qv \geq 0 \text{ on } \Gamma\}. \quad (8.1.8)\]

The relation between the primal problem (8.1.1) and its dual (8.1.6) is clarified in the following statement.

8.1.1 Theorem. If the primal problem (8.1.1) is solvable, then there exists a unique solution \(q^*\) of the dual problem (8.1.6), and the equations

\[q^* = \nabla u^*, \quad J(u^*) + J(q^*) = 0 \quad (8.1.9)\]

are true for every solution of Problem (8.1.1).

This theorem holds also true for more general variational problems with convex quadratic objective functionals and constraints of the form \(Bu \leq b\), with \(B\) a linear operator in a space of functions on \(\Omega\) or \(\Gamma\).

Let us sketch its proof. Introducing new functions

\[\zeta_k := \frac{\partial u}{\partial x_k}, \quad (8.1.10)\]

Problem (8.1.1) can be reformulated as

\[\min \left\{ \frac{1}{2} \|\zeta\|^2_{0, \Omega} - \langle f, u \rangle_{0, \Omega} : \zeta_k = \frac{\partial u}{\partial x_k}, \ k = 1, \ldots, n; \ u \in K \right\}. \]
CHAPTER 8. PPR FOR VARIATIONAL INEQUALITIES

defined on the set $K \times [L_2(\Omega)]^n \times [L_2(\Omega)]^n$ we express the dual problem as

$$\sup_{q \in [L_2(\Omega)]^n} \inf_{(u, \zeta) \in K \times [L_2(\Omega)]^n} \bar{L}(u, \zeta, q).$$

This can be put in the form (8.1.6) by an explicit calculation of

$$\psi(q) := \inf_{(u, \zeta) \in K \times [L_2(\Omega)]^n} \bar{L}(u, \zeta, q).$$

Indeed, we can decompose $\psi(q) = \psi_1(q) + \psi_2(q)$, where

$$\psi_1(q) := \inf_{\zeta \in [L_2(\Omega)]^n} \left\{ \frac{1}{2} \|\zeta\|^2_{0, \Omega} - \sum_{k=1}^n \langle q_k, \zeta_k \rangle_{0, \Omega} \right\},$$

$$\psi_2(q) := \inf_{u \in K} \left\{ \sum_{k=1}^n \langle q_k, \frac{\partial u}{\partial x_k} \rangle_{0, \Omega} - \langle f, u \rangle_{0, \Omega} \right\}.$$

A short calculation gives for $q \in K$

$$\psi_1(q) = -\frac{1}{2} \|q\|^2_{0, \Omega}$$

and, due to (8.1.8), (8.1.5),

$$\psi_2(q) = \langle qv, g \rangle.$$

Moreover, it can be easily seen that $\psi_2(q) = -\infty$ for $q \notin K$.

As $H(\text{div}; \Omega)$ is a Hilbert space and the set $K$ is convex and closed, existence and uniqueness of the solution of the dual problem (8.1.6) follows from the strong convexity of the functional $\frac{1}{2} \|q\|^2_{0, \Omega}$ on $[L_2(\Omega)]^n$.

Concerning the duality relation (8.1.9), it should be noted that for Problem (8.1.1) as well as for many other convex variational problems, relations of that type are established according to a general framework developed by EKELAND and TEMAM [100].

Finally, the relation $q^* = \nabla u^*$ can be concluded from the uniqueness of $q^*$ and the inclusion $\nabla u^* \in K$.

Essentially for the application of finite element methods to the dual problem (8.1.6) is that the equation

$$\text{div } q + f = 0 \quad \text{in } \Omega \tag{8.1.11}$$

is satisfied on each approximation level. This can be done by means of a suitable substitution of the unknown function $q$, which makes equation (8.1.11) homogeneous, and by the use of special solenoidal linear finite elements $\varphi_i$ (see [87]). These elements satisfy the condition $\text{div } \varphi_i = 0$ on each triangle. Note that, for $n = 2$, we actually deal with two functions $q_1$ and $q_2$ related by (8.1.11). Consequently, the finite element $\varphi_i$ is a two-dimensional vector-function. Concerning details of this approach, we refer to HLAVÁČEK ET AL. [182].

The construction of finite-dimensional approximations of the dual problem of an elliptic variational inequality is more complicated. Moreover, in order to prove the convergence of the approximate solutions of these auxiliary problems and also some rate of convergence, higher smoothness of the solution of the
variational inequality is required. Nevertheless, this approach has the advantage that, in view of the strong convexity of the objective function, strong convergence of the minimizing sequence and an acceptable stability of the auxiliary problems can be established.

Note that for the model problem (8.1.1) the question of non-uniqueness of the solution is of minor importance. In the indefinite case $\langle f, 1 \rangle_{0, \Omega} = 0$ solutions of the problem can be obtained, provided they exist, by solving the corresponding Neumann problem.

However, for problems in elasticity theory, considered in the sequel, possible non-uniqueness of the solution originates from meaningful physical situations and may be a serious obstacle for the numerical treatment of such problems.

Problems in a mixed variational formulation are obtained in form of saddle point problems for a Lagrangian function which is different from the one we used before. Such a function is introduced in order to exclude unilateral constraints, i.e. inequality constraints like $\gamma u \geq g$ on $\Gamma$ as considered in (8.1.3). For details of this notion see Fichera [114] and Panagiotopoulos [316].

The saddle point problem related to (8.1.1) is given as

$$\text{find a point } (u^*, \lambda^*) \in H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma) \text{ such that}$$

$$\mathcal{L}(u^*, \lambda) \leq \mathcal{L}(u, \lambda^*) \leq \mathcal{L}(u, \lambda) \quad \forall (u, \lambda) \in H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma), \quad (8.1.12)$$

with

$$\mathcal{L}(u, \lambda) := J(u) + \langle \lambda, -\gamma u + g \rangle. \quad (8.1.13)$$

Assuming solvability of the original problem (8.1.1), its equivalence (in the sense of Proposition A1.7.53) with Problem (8.1.12) can be established by arguments analogous to those used in the proofs of Kuhn-Tucker type theorems under Slater’s condition.

**8.1.2 Theorem.** The point $(u^*, \lambda^*) \in H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$ is a saddle point of the Lagrangian function (8.1.13) on the set $H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$ iff $u^*$ is a solution of Problem (8.1.1) and $\lambda^* = \frac{\partial u^*}{\partial \nu}$, with $\nu$ a unit vector of the outward normal to $\Gamma$.

The proof of this statement is almost literally the same as that in Cea [69] for the problem

$$\frac{1}{2} \|u\|^2_{1, \Omega} - \langle f, u \rangle_{0, \Omega} \to \min$$

$$\text{s.t. } K := \{u \in H^1(\Omega) : \gamma u \geq 0 \text{ on } \Gamma\}.$$

In order to discretize problems of the mixed variational formulation, two systems of finite elements have to be introduced to approximate both the functions of the initial space and those of the space of the Lagrange multiplier vectors. For the model problem (8.1.1) functions in the initial space $H^1(\Omega)$ can be approximated by piece-wise linear elements on a standard quasi-uniform triangulation of the domain $\Omega$, see Definition A2.2.2. Functions from $H^{-\frac{1}{2}}(\Gamma)$ have to be approximated by piece-wise constant elements on $\Gamma$. Edges of the partition of $\Gamma$ are usually those vertices of the corresponding triangulation of the domain $\Omega$ which are located on the boundary. However, such a close connection between
the triangulations is not obligatory, see Haslinger and Lovišek [170].

In this way, for the saddle point problem (8.1.12) a sequence of approximate saddle point problems is constructed on \( V_h \times \Lambda_{h}^{+} \), where \( V_h \) and \( \Lambda_{h}^{+} \) are finite element spaces corresponding to a given triangulation \( T_h \) on \( \Omega \), and \( \Lambda_{h}^{+} \) denotes the subset of non-negative functions on \( \Lambda_h \).

Concerning the approach of discretization of the original problem (8.1.1) it should be noted that, in general, we do not have the inclusion \( K_h \subset K \) for the approximated auxiliary problems. This fact complicates the investigation of convergence and rate of convergence of the solution methods. The use of the mixed variational formulation of Problem (8.1.1) overcomes this disadvantage because \( V_h \subset V \) and \( \Lambda_{h}^{+} \subset H_{+}^{-\frac{1}{2}}(\gamma) \). This situation is typical for variational inequalities in real life problems.

The advantage of the latter approach consists in a simultaneous computation of an approximate solution \( u_{h}^{*} \) and an approximate multiplier \( \lambda_{h}^{*} \). For problems considered in this chapter the functionals \( \lambda^{*} \) have an explicit physical interpretation which may be of interest in itself. Moreover, these functionals, computed by the problem in mixed variational formulation, are usually more accurate that those obtained by numerical differentiation of the approximate solutions of the original problem.

In some cases the mixed variational formulation enables us to replace the minimization of a non-smooth functional by solving a saddle point problem for a functional which is differentiable in both variables \( u \) and \( \lambda \), see Glowinski et al. [135] for a model problem with friction.

However, the use of the mixed formulation does not eliminate difficulties which are connected with the non-uniqueness of the solution of the original problem, because in that case the Lagrangian function has more than one saddle point. Moreover, the selection of particular algorithms for solving the corresponding approximate auxiliary problems is more restricted. On the other hand, for a straightforward numerical solution of the approximated auxiliary problems of the original problem the typical simple structure of the feasible sets \( K \) allows us to apply fast-convergent methods.

### 8.2 Contact Problems Without Friction

In this chapter the abstract MSR-method described in Section 4.3 will be specified in order to get stable solution algorithms for some elliptic variational inequalities in elasticity theory. To this end a straightforward finite element approximation of the original problems will be used.

Our attention is mainly devoted to contact problems of two elastic bodies (two-body contact problem) and of the equilibrium of an elastic body supported by an absolutely rigid surface (Signorini problem). Both problems will be considered in the plane and without friction. These plane problems have been investigated intensively for instance by Fichera [114] and Hlaváček et al. [182] (see also the references therein). These problems are ill-posed as it will be seen in the sequel.

In Section 8.3, in connection with an algorithm of alternating iterations suggested by Panagiotopoulos [315] for contact problems with friction, we shall also investigate the application of IPR for elliptic variational inequalities.
corresponding to the minimization of non-smooth energy functionals.

For notions in the elasticity theory the reader is referred to the monographs of Hahn [166] and Gladwell [131].

We shortly review some results, given for instance in Duvaut and Lions [95], Panagiotopoulos [316] and Hlaváček and Lovišek [183], on the existence, uniqueness and characterization of the solutions. We mainly make use of the notations and terminology introduced in [182].

8.2.1 Formulation of the problems

All the problems studied here are considered in the framework of the theory of small deformations under the linear Hooke’s law for anisotropic materials and a constant field of temperature. Moreover, it is assumed that the initial stresses and strains are equal to zero.

**Contact problem:** We first formulate the boundary value problem describing the contact of two elastic bodies, supposing that these bodies occupy bounded domains \( \Omega' \subset \mathbb{R}^2 \) and \( \Omega'' \subset \mathbb{R}^2 \) with Lipschitz-continuous boundaries \( \Gamma := \partial \Omega' \) and \( \Gamma'' := \partial \Omega'' \) in the initial situation. Everywhere the upper strings ‘ and ” in the notations correspond to the bodies \( \bar{\Omega}' \) and \( \bar{\Omega}'' \), respectively. The bodies are in contact along a part \( \Gamma_c := \partial \Omega' \cap \partial \Omega'' \) of the boundary, the so-called contact zone, and their forms and mutual positions exclude an enlargement of the contact zone \( \Gamma_c \) within the deformation. The partitions of the boundaries

\[
\partial \Omega' := \Gamma_u \cup \Gamma'_u \cup \Gamma_c \\
\partial \Omega'' := \Gamma_0 \cup \Gamma''_0 \cup \Gamma_c,
\]

with \( \text{meas } \Gamma_u > 0 \) and \( \text{meas } \Gamma_c > 0 \), are assumed to be known.

---

Figure 8.2.1:

The body \( \Omega' \) is fixed on \( \Gamma_u \) and external forces \( P' \) and \( P'' \) are given on \( \Gamma'_u \) and \( \Gamma''_u \), see Figure 8.2.1.

On the boundary \( \Gamma_0 \) a zero displacement \( u_\nu \) in the direction of the outer
normal $\nu$ and a zero tangential stress $T_t$ are assumed. The case of $\Gamma_0 = \emptyset$ is permitted. Typically, conditions of this type are required on the symmetry axis of the system and can be reformulated as boundary conditions if an equivalent problem for the "half-system" is considered.

Under the above assumptions, taking into account the influence of body forces $F$, we consider a vector field of displacements $u = (u_1(x), u_2(x))$ of points $x \in \Omega' \cup \Omega''$ which correspond to the equilibrium state, as well as the tensor field of strains

$$\epsilon_{kl}(u) := \frac{1}{2} \left( \frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k} \right), \quad k, l = 1, 2. \quad (8.2.1)$$

The components of the stress tensor $\tau_{kl}$ are given according to Hooke’s law:

$$\tau_{kl} := a_{klpm} \epsilon_{pm}, \quad k, l = 1, 2,$$

with $a_{klpm}$ the coefficients of elasticity. As usually in elasticity theory, here and in the sequel we follow Einstein’s summation convention, i.e., the summation is performed over terms with repeating indices.

The functions $a_{klpm}$ are assumed to be measurable and bounded on $\Omega = \Omega' \cup \Omega''$. Moreover, symmetry

$$a_{klpm} = a_{lkpm} = a_{pmkl} \quad (8.2.2)$$

as well as the ellipticity property has to be satisfied, i.e., there exists a positive constant $c_0$ such that

$$a_{klpm}(x) \sigma_{kl} \sigma_{pm} \geq c_0 \sigma_{kl} \sigma_{pm} \quad (8.2.3)$$

for all symmetric matrices $\{\sigma_{kl}\}_{k,l=1,2}$ and almost every $x \in \Omega$.

According to the linear elasticity theory the stress tensor has to fulfil the equilibrium conditions

$$\frac{\partial \tau_{kl}}{\partial x_l} + F_k = 0 \quad \text{in } \Omega, \ k = 1, 2. \quad (8.2.4)$$

Body $\Omega'$ being fixed on $\Gamma_u$ means that

$$u = 0 \quad \text{on } \Gamma_u \quad (8.2.5)$$

and the two-sided contact conditions on $\Gamma_0$ can be described by

$$u_{\nu} = 0, \quad T_t = 0 \quad \text{on } \Gamma_0, \quad (8.2.6)$$

with $u_{\nu} := u_k \nu_k$, $T_t := \tau_{kl} \nu_k t_l$, $t := (-\nu_2, \nu_1)$ or $t := (\nu_2, -\nu_1)$ and $\nu$ a unit vector of an outward normal to $\Gamma_0$.

On $\Gamma'''$ and $\Gamma''''$ the conditions

$$\tau'_{kl} \nu'_l = P'_k, \quad \tau''''_{kl} \nu''''_l = P''''_k \quad (8.2.7)$$

have to be satisfied for $k = 1, 2$.

Finally, the following conditions are valid on $\Gamma_c$:

$$u'_{\nu'} + u''''_{\nu''} \leq 0, \quad (8.2.8)$$
meaning that penetration of the points from one body into the other is impossible.

From Newton’s third law (actio=reactio) and the absence of friction and stretching of the bodies, we get the conditions

\[ T_1' = T_1'' = 0, \]
\[ T_\nu' = T_\nu'' \leq 0, \]
\[ (u_\nu' + u_\nu'')T_\nu'' = 0, \]

where \( T_\nu \) denotes the normal component of the force \( T \).

Condition (8.2.8) holds for the contact problem under the standard assumption that the radius of the curvature of the arc \( \Gamma_c \) is large in comparison with its length.

**Signorini problem:** This problem can be represented as a special case of the two-body contact problem, where \( \Omega' \) is supposed to be an absolutely rigid body. The transition from the two-body contact problem to its formulation is not difficult, we have only to reformulate the conditions on the boundary \( \Gamma_c \):

\[ u_\nu'' \leq 0, \quad T_1'' = 0, \quad T_\nu'' \leq 0, \quad u_\nu'', T_\nu'' = 0 \quad \text{on} \quad \Gamma_c. \]  

(8.2.10)

In future, we formulate the Signorini problem so that the elastic body is denoted by \( \Omega \) and \( \Gamma_c \) denotes the support part of the boundary. The strings “” will be omitted.

A vector-function \( u \) satisfying the equilibrium system (8.2.4) and the boundary conditions (8.2.5)–(8.2.9) can be interpreted as a classical solution of the two-body contact problem. We use this notion analogously for the Signorini problem.

### 8.2.1.1 Variational formulations of the problems

The variational formulation of these problems corresponds to the principle of minimal potential energy.

Concerning the two-body contact problem, we introduce the space of virtual displacements

\[ V := \{ u = (u', u'') \in [H^1(\Omega')]^2 \times [H^1(\Omega'')]^2 : u' = 0 \quad \text{on} \quad \Gamma_u, \quad u'' = 0 \quad \text{on} \quad \Gamma_0 \}, \]  

(8.2.11)

with the norm

\[ \|u\|_{1, \Omega} := \sqrt{\|u'\|_{1, \Omega'}^2 + \|u''\|_{1, \Omega''}^2} \]

and the set of admissible displacements

\[ K := \{ u \in V : u_\nu' + u_\nu'' \leq 0 \quad \text{on} \quad \Gamma_c \}. \]  

(8.2.12)

**Variational formulation of the two-body contact problem:**
Minimize the functional of potential energy

\[ J(u) := \frac{1}{2} a(u, u) - \ell(u) \]

\[ s.t. \quad u \in K := \{ v \in V : v'_\nu + v''_{\nu} \leq 0 \quad \text{on} \quad \Gamma_c \}, \]  

(8.2.13)

where

\[ a(u, v) := \int_{\Omega} a_{klpm} \epsilon_{kl}(u) \epsilon_{pm}(v) d\Omega, \]  

(8.2.14)

\[ \ell(v) := \int_{\Omega} F_k v_k d\Omega + \int_{\Gamma_{\tau}} P_k v_k d\Gamma, \quad \Gamma_{\tau} := \Gamma'_{\tau} \cup \Gamma''_{\tau}. \]  

(8.2.15)

A minimizer \( u^* \) of Problem (8.2.13) is called a weak solution of the two-body contact problem.

**8.2.1 Theorem.** Each classical solution of the two-body contact problem is a weak solution. If a weak solution is sufficiently smooth, then it is a solution in the classical sense.

Concerning the Signorini problem, we suppose that the body occupies in its initial state a bounded domain \( \Omega \) with a fixed partition of the Lipschitz-continuous boundary \( \Gamma := \partial\Omega \):

\[ \Gamma = \Gamma_0 \cup \Gamma_c \cup \Gamma_{\tau}. \]

Vector-functions as body forces \( F \in [L_2(\Omega)]^2 \) and external forces \( P \in [L_2(\Gamma_{\tau})]^2 \) (acting on the boundary \( \Gamma_{\tau} \)) are given. Under the previously made assumptions for the elasticity coefficients \( a_{klpm} \), we consider the following problem:

**Variational formulation of the Signorini problem:**

Minimize the functional of potential energy

\[ J(u) := \frac{1}{2} a(u, u) - \ell(u) \]

\[ s.t. \quad u \in K := \{ v \in [H^1(\Omega)]^2 : v_\nu = 0 \quad \text{on} \quad \Gamma_0, \quad v'_\nu \leq 0 \quad \text{on} \quad \Gamma_c \}, \]

(8.2.16)

with

\[ a(u, v) := \int_{\Omega} a_{klpm} \epsilon_{kl}(u) \epsilon_{pm}(v) d\Omega, \]  

(8.2.17)

\[ \ell(v) := \int_{\Omega} F_k v_k d\Omega + \int_{\Gamma_{\tau}} P_k v_k d\Gamma. \]  

(8.2.18)

The meaning of \( \epsilon_{kl} \) and \( v_\nu \) is the same as in Problem (8.2.13). The connection between the solutions of this problem and the classical one is analogous to Theorem 8.2.1.

**8.2.1.2 Solvability of the problems**

Now, we turn to the question of solvability and characterization of the optimal sets of the problems under consideration.
8.2. CONTACT PROBLEMS WITHOUT FRICTION

In the case of the two-body contact problem the kernel $K$ of the bilinear form (8.2.14) on the space $Z := [H^1(\Omega')]^2 \times [H^1(\Omega'')]^2$ consists of elements $z = (z', z'')$, where $z'$ and $z''$ are vector-functions with components

\[ z'_1(x) := a'_1 - b'_x_2, \quad z'_2(x) := a'_2 + b'_x_1, \]
\[ z''_1(x) := a''_1 - b''x_2, \quad z''_2(x) := a''_2 + b''x_1, \]

with arbitrary coefficients $a'_1, a'_2, b'$ and $a''_1, a''_2, b''$.

Since the body $\Omega'$ is fixed on a part of the boundary $\Gamma_u$, its rigid displacement is impossible.\(^1\)

For $\ell$ defined according to (8.2.15) it is easy to see that

\[ \ell(y) \leq 0 \quad \forall \ y \in K \cap \mathcal{R}. \]

This is a necessary condition for the existence of a solution of Problem (8.2.13).

**8.2.2 Theorem.** If the conditions

\[ \ell(y) \leq 0 \quad \forall \ y \in K \cap \mathcal{R}, \]
\[ \ell(y) < 0 \quad \forall \ y \in K \cap \mathcal{R} \quad \text{with} \quad \inf_{x \in \Gamma_c} \{ y'_\nu(x) + y''_\nu(x) \} < 0 \]

hold, then Problem (8.2.13) has at least one solution $u^*$.

Moreover, the solution set has the structure

\[ U^* = \{ u^* + y : \ y \in V \cap \mathcal{R}, \ u^* + y \in K, \ \ell(y) = 0 \}, \]

with $u^*$ fixed.

If the boundary part $\Gamma_0 = \emptyset$, then the dimension of the subspace $V \cap \mathcal{R}$ of virtual, rigid displacements of the body $\Omega''$ is equal to three. In case $\text{meas} \Gamma_0 > 0$ we have $\dim(V \cap \mathcal{R}) \leq 1$.

For the Signorini problem the kernel $\mathcal{R}$ of the bilinear form (8.2.17) on the space $[H^1(\Omega)]^2$ is a three-dimensional subspace consisting of vector-functions

\[ z_1(x) = a_1 - bx_2, \quad z_2(x) = a_2 + bx_1. \]

If there exists a solution of Problem (8.2.16), then the functional (8.2.18) has to be non-positive on the set $K \cap \mathcal{R}$.

**8.2.3 Theorem.** Let for Problem (8.2.16) the following conditions be fulfilled:

\[ \ell(y) \leq 0 \quad \forall \ y \in K \cap \mathcal{R}, \]
\[ \ell(y) < 0 \quad \forall \ y \in K \cap \mathcal{R} \quad \text{satisfying} \quad \inf_{x \in \Gamma_c} y_\nu(x) < 0. \]

Then Problem (8.2.16) is solvable and, with a fixed solution $u^*$, the optimal set can be expressed by

\[ U^* = \{ u^* + y : \ y \in V \cap \mathcal{R}, \ u^* + y \in K, \ \ell(y) = 0 \}, \]

with

\[ V := \{ u \in [H^1(\Omega)]^2 : u_\nu = 0 \quad \text{on} \ \Gamma_0 \}. \]

\(^1\)A displacement is called rigid if the distance between any point of the body is not changing.
As in the two-body contact problem, \( \dim(V \cap \mathcal{R}) = 3 \) for \( \Gamma_0 = \emptyset \) and equal to zero or one in case \( \text{meas } \Gamma_0 > 0 \).

Applying the finite element method to the variational formulations of the both problems considered, convergence of the corresponding minimizing sequences in the norm of the space \( V \) can be established only in the following particular cases (i)–(ii):

(i) \( \dim(V \cap \mathcal{R}) \leq 1 \),

(ii) either Korn's inequality (see (8.2.55)) is fulfilled on \( V \), or a subspace \( V^1 \subset V \) can be chosen such that \( U^* \cap (K \cap V^1) \neq \emptyset \) and an analogue of Korn's inequality holds on \( V^1 \) (cf. [182], [158]).

In case \( \dim(V \cap \mathcal{R}) = 3 \) finite element approximations of the dual problem are commonly used. The corresponding Lagrange multipliers represent the components of the uniquely defined stress tensor.

However, the description of the feasible set of the dual problem is considerably complicated such that sophisticated approximation techniques, based on equilibrium models for the finite elements have to be applied.

### 8.2.2 Finite element approximation of contact problems

Now, we consider the usage of finite element methods to the Problems (8.2.13) and (8.2.16). For the seek of simplicity, let us assume that the domains \( \Omega' \) and \( \Omega'' \) (or \( \Omega \) in the Signorini problem) are bounded and polyhedral, such that we can apply the FEM-scheme given in Appendix A2.2.

In this situation the contact boundary \( \Gamma_c \) of the two-body contact problem can be described as

\[
\Gamma_c = \bigcup_{j=1}^{m} \Gamma_{c,j},
\]

with \( \Gamma_{c,j} \) denoting the closed straight-line intervals. Points, at which the type of the boundary condition changes and vertices belonging to the polyhedrons \( \Omega' \) and \( \Omega'' \), have to be included in the vertex set of the corresponding triangulations \( \mathcal{T}'_h \) or \( \mathcal{T}''_h \) for each triangulation parameter \( h \). End-points of the intervals \( \Gamma_{c,j} \) have to be common vertices for both triangulations. Under these conditions we call a system of triangulation of each domain consistent with the partition of the boundary.

Moreover, vertices located on \( \Gamma_c \) are common vertices of \( \mathcal{T}'_h \) and \( \mathcal{T}''_h \) and the quasi-uniformity of the triangulation sequences \( \{\mathcal{T}'_h\} \) and \( \{\mathcal{T}''_h\} \) must be ensured according to Definition A2.2.2. It is easy to see that in this case the sequence \( \{\mathcal{T}_h\} \) with \( \mathcal{T}_h = \mathcal{T}'_h \cup \mathcal{T}''_h \) forms a quasi-uniform triangulation of the domain \( \Omega = \Omega' \cup \Omega'' \).

Vertices belonging to \( \mathcal{T}'_h \) and \( \mathcal{T}''_h \) we shall denote by \( \pi_k \). The index sets \( I'_h, I''_h, I_u^0, I_0^0 \) and \( I_{c,j}^0 \) indicate that the points \( \pi_k \) belong to the sets \( \Omega', \Omega'', \Gamma_u, \Gamma_0^0 \) and \( \Gamma_{c,j}^0 \), respectively.

In order to simplify the description, we assume that \( \Gamma_0 = \emptyset \) or that \( \Gamma_0 \) is an interval. In the latter case \( \nu^0 \) denotes a unit outward normal to \( \Gamma_0 \), and \( \nu^j \) is a unit normal to \( \Gamma_{c,j} \), pointed to the exterior of \( \Omega' \).

For a fixed pair of triangulation \( \mathcal{T}'_h \) and \( \mathcal{T}''_h \) the finite element space \( V_h \) of vector-functions

\[
v_h = (v'_h, v''_h) \in \left( [C(\Omega')]^2 \times [C(\Omega'')]^2 \right) \cap V
\]
The set $K_{\beta}$ due to (8.2.12), the coefficients $\beta^k$ have to satisfy

$$\langle \nu^0, \beta^k \rangle_{R^2} = 0, \quad k \in I_h^0. \quad (8.2.23)$$

The set $K_h$ approximating $K$ on the space

$$V_h := \{ v_h = (v'_h, v''_h) : \langle \nu^0, \beta^k \rangle_{R^2} = 0, \quad k \in I_h^0 \}, \quad (8.2.24)$$

can be expressed by

$$K_h := \{ v_h \in V_h : \langle \nu^0, \alpha^k - \beta_k \rangle_{R^2} \leq 0, \quad k \in I_h^{c,j}, j = 1, ..., m \} \quad (8.2.25)$$

or

$$K_h := \{ v_h \in V_h : \langle \nu^0, \alpha^k - \beta_k \rangle_{R^2} \leq 0, \quad k \in I_h^{c,j}, j = 1, ..., m \}. \quad (8.2.26)$$

In this case the inclusion $K_h \subset K$ is established. To prove this fact it is sufficient to remark that, on the one hand, $v'_h, v''_h$ as well as the functions describing the boundary conditions of Problem (8.2.13) on $\Gamma_0, \Gamma_0, \Gamma_c$ are linear on each interval which connects neighboring vertices on the boundaries, and on the other hand, that the functions $v'_h, v''_h$ satisfy the boundary conditions at the vertices of the triangulation.

The approximate problem in the space $V_h$ can be formulated simply as

$$\min \{ J(u_h) : u_h \in K_h \}, \quad (8.2.27)$$

where $J$ is defined according to (8.2.13), (8.2.14), (8.2.15) and the feasible set $K_h$ is given by relation (8.2.25).

Concerning the Signorini problem, we use the same representation of the contact zone $\Gamma_c$ as in (8.2.20) and assume that $\Gamma_0 = 0$ or that $\Gamma_0$ is an interval. Again $\nu^0$ and $\nu^j$ are unit outward normals to $\Gamma_0$ and $\Gamma_{c,j}$, respectively. Again a sequence of quasi-uniform triangulations of the domain $\Omega$ has to be performed which is consistent with the partition of the boundary. For fixed $\mathcal{T}_h$ the vertices of the triangulation are denoted by $\pi_k$, and the sets $I_h, I_h^0, I_h^{c,j}$ represent the index sets of the vertices located on $\Omega, \Gamma_0$ and $\Gamma_{c,j}$, respectively. On $\mathcal{T}_h$ a system of finite elements $\{ \varphi_k \}$ is introduced as above, and with regard to (8.2.16) and (8.2.19) the space $V_h$ is defined by

$$V_h := \{ v_h = \sum_{k \in I_h} \alpha^k \varphi_k(x) : \alpha^k \in R^2, \langle \nu^0, \alpha^k \rangle_{R^2} = 0, \quad k \in I_h^0 \} \quad (8.2.28)$$
CHAPTER 8. PPR FOR VARIATIONAL INEQUALITIES

and the set $K_h$ by

$$K_h := \{ v_h \in V_h : \langle \nu^j, v_h(\pi_k) \rangle_{\mathbb{R}^2} \leq 0, \ k \in I^c_h, j = 1, \ldots, m \}, \quad (8.2.29)$$

with

$$\langle \nu^j, v_h(\pi_k) \rangle_{\mathbb{R}^2} = \langle \nu^j, \alpha^k \rangle_{\mathbb{R}^2}.$$  

On this way we get an approximate problem of the form (8.2.27), where functional $J$ is given by (8.2.16), (8.2.17), (8.2.18) and set $K_h$ by (8.2.29).

Using analogous arguments as above, one can show that again the inclusion $K_h \subset K$ holds.

For a more detailed description of the approximate problems we refer to the finite element analysis of the Problems (A2.1.3) and (A2.1.6) in Appendix A2.4.

If the above described sufficient conditions of solvability of the problems hold, cf. Theorems 8.2.2 and 8.2.3, then solvability of the approximate problems (8.2.27) follows immediately from the property $K_h \subset K$.

The case where the contact boundary is described by a convex, continuous function has been investigated in [182]. Here the triangulations of $\Omega'$ and $\Omega''$ are described by curved triangles $T_j$ along the contact boundary such that

$$\bigcup_{T_j \in \mathcal{T}'_h} \bar{T}_j = \bar{\Omega}', \quad \bigcup_{T_j \in \mathcal{T}''_h} \bar{T}_j = \bar{\Omega}''$$

and the vertices $\pi_k$ on this boundary ($k \in I^c_h$) are common vertices of both triangulation systems $\mathcal{T}'_h$ and $\mathcal{T}''_h$ which are assumed to be quasi-uniform.

The space $V_h$ can be defined according to (8.2.24) and

$$K_h := \{ v_h \in V_h : \langle \nu(\pi_k), v'_h(\pi_k) - v''_h(\pi_k) \rangle_{\mathbb{R}^2} \leq 0, \ k \in I^c_h \},$$

with $\nu(\pi_k)$ a unit outward normal to $\Omega'$.

Note that in this case the inclusion $K_h \subset K$ is not valid in general.

Let us briefly consider the numerical solution methods which can be suggested for solving the approximate problems of type (8.2.27). We confine ourselves to the case of a contact problem with two polyhedral bodies.

The approximate problem is a quadratic programming problem with a positive semi-definite Hessian of the objective functional and linear, homogeneous inequalities (for $\pi_k \in \Gamma_c$) and equality constraints (if $\pi_k \in \Gamma_0$). The coefficient matrix of the constraints has a special sparsity structure, i.e., the maximal number of non-zero elements does not exceed 4 in the rows and 2 in columns corresponding to break-points of $\Gamma_c$ and entries 1 in the remaining columns.

The insertion of slack variables into the inequality constraints on $\Gamma_c$ and a successive elimination of a part of the variables enables us to reformulate the problem into a convex quadratic program with simple constraints such that finally only non-negativity of the slack variables is required.

Hence, either special numerical methods as described in Appendix A3.3 can be applied to the transformed problems with simple constraints, or standard quadratic programming algorithms to both the transformed and non-transformed problems are applicable.
However, the numerical realization of these methods is connected with serious difficulties due to the non-exact computation of the Hessian which may lead to non-convex auxiliary problems. The influence of errors may be significant, in particular, because in practice the discretization is often performed with the use of considerably denser triangulations along the contact zone than in the other parts of the domain, i.e., one has to work with elements of very different size.

For solving such problems in a stable way with an accuracy adjusted to the accuracy of the approximation, special quasi-Newton methods may be particularly suitable. In each iteration a (not strongly) convex objective functional has to be approximated by a quadratic functional with positive definite Hessian. Positive definiteness can be obtained by small variations of the diagonal elements in the framework of usual quasi-Newton procedures.

However, this type of regularization immanent in quasi-Newton methods applied to each approximate problem (8.2.27) \((h \text{ is fixed})\) seems less efficient than the use of an iterative regularization of the external iteration process \((h \downarrow 0)\) ensuring strong convexity of the objective functional in the approximate problems, too.

### 8.2.3 Three Variants of MSR-Methods

In this subsection we apply the MSR-approach, described in Sections 4.3 and 5.1, to the solution of contact problems.

The family of sets \(\{K_k\}\) is assumed to be constructed by means of a finite element approximation of the feasible set \(K\) according to Subsection 8.2.2. For fixed triangulations \(\mathcal{T}_h\) and \(\mathcal{T}_h\) (or \(\mathcal{T}_h\) in the case of Signorini problems) we take \(K_k := K_{h_k}\) with a triangulation parameter \(h_k\).

#### Standard MSR-scheme:

According to the MSR-scheme in Section 4.3 we choose \(J_k := J\), because the functionals (8.2.13) and (8.2.16) are differentiable. With regard to the expression of the approximate problems (8.2.27) their regularization leads to the problems

\[
\min\{J(u) + \|u - u^{k,i-1}\|_V^2 : u \in K_k\}, \tag{8.2.30}
\]

In addition to this standard MSR-method the following two particular methods will be considered:

**Regularization on the subspace:**

\[
V_1 := \{u = (u', u'') \in Z : u' \equiv 0 \text{ in } \Omega', \\
u_1'(x) = a_1 - bx_2, \; u_2'(x) = a_2 + bx_1 \text{ in } \Omega'; \tag{8.2.31}
\]

(in the case of two-body contact problems) or

\[
V_1 := \{u \in [H^1(\Omega)]^2 : u_1(x) = a_1 - bx_2, u_2(x) = a_2 + bx_1 \text{ in } \Omega}; \tag{8.2.32}
\]

(in the case of Signorini problems),

where the auxiliary problems are of the form

\[
\min\{J(u) + \|\hat{\Pi}_1 u - \hat{\Pi}_1 u^{k,i-1}\|_V^2 : u \in K_k\}, \tag{8.2.33}
\]

with \(\hat{\Pi}_1 : V \to V_1\) an ortho-projector according to the norm \(\|\cdot\|_{0,\Omega}\).
Regularization with weaker norm:
instead of Problem (8.2.30) we have to solve
\[
\min \{ J(u) + \| \bar{\Pi}_1 u - \bar{\Pi}_1 u^{k,i-1} \|_{L^2_0}^2 : u \in K_k \},
\]
(8.2.34)
with
\[
\bar{\Pi}_1(u(x)) = \begin{cases} 
0 & \text{for } x \in \Omega', \\
\bar{u}''(x) & \text{for } x \in \Omega'',
\end{cases}
\]
in the case of two-body contact problems and
\[
\bar{\Pi}_1 u = u
\]
for Signorini problems.

8.2.3.1 Convergence analysis
The three methods above can be studied as particular realizations of the following basic approach.

8.2.4 Assumption. Let \( Y \) and \( H \) be two Hilbert spaces and \( Y \) be continuously imbedded into \( H \) (see Definition A1.3.7). Furthermore, we assume that
\begin{enumerate}
\item \( a(\cdot, \cdot) \) and \( b(\cdot, \cdot) \) are continuous, symmetric bilinear forms on \( Y \times Y \) and
\[
a(u, u) \geq b(u, u) \geq 0 \quad \text{on } Y;
\]
\item the objective functional \( J \) has the form
\[
J(u) := \frac{1}{2} a(u, u) - \langle f, u \rangle, \quad (f \in Y') \text{ fixed};
\]
\item \( \Pi_1 : Y \to Y_1 \) is an ortho-projector in the norm of \( Y \) or \( H \) with \( Y_1 \) a subspace of \( Y \);
\item for some \( \beta > 0 \) it holds
\[
\frac{1}{2} b(u, u) + \| \Pi_1 u \|_H^2 \geq \beta \| u \|_Y^2, \quad \forall u \in Y.
\]
\end{enumerate}
(8.2.35)
\[\diamond \]
Under this assumption we introduce on \( Y \) another norm \( | \cdot | \), defined by
\[
|u|^2 := \frac{1}{2} b(u, u) + \| \Pi_1 u \|_H^2.
\]
(8.2.36)
The space \( Y \) with this norm will be denoted by \( Y \) and its conjugate by \( Y' \).

The norms \( \| \cdot \|_Y \) and \( | \cdot | \) are equivalent. Indeed, there exist constants \( c \) and \( M \) such that
\[
\| u \|_H \leq c \| u \|_Y, \quad |b(u, v)| \leq M \| u \|_Y \| v \|_Y,
\]
and for the ortho-projector $\Pi_1$, defined according to the norm $\| \cdot \|_Y$ or $\| \cdot \|_H$, we have
\[ \| \Pi_1 u \|_H \leq c \| u \|_Y. \] (8.2.37)

Finally, using inequality (8.2.35), it holds
\[ \beta \| u \|_Y^2 \leq |u|^2 \leq \left( \frac{M}{2} + c^2 \right) \| u \|_Y^2, \quad \forall u \in Y. \] (8.2.38)

For a functional $J$, defined according to Assumption 8.2.4(i),(ii) we consider the generalized problem
\[ \min \{ J(u) : u \in K \}, \] (8.2.39)
with $K$ a convex, closed subset of $Y$.

The corresponding family of regularized and approximate auxiliary problems
\[ \Psi_{k,i}(u) := J(u) + \| \Pi_1 u - \Pi_1 u^{k,i-1} \|_H^2 \to \min, \quad u \in K_k, \] (8.2.40)
\[ \bar{u}_{k,i} := \arg \min_{u \in K_k} \Psi_{k,i}(u). \] (8.2.41)

is constructed with convex, closed sets $K_k \subset Y$ and with an ortho-projector $\Pi_1$ satisfying Assumption 8.2.4(iii),(iv).

Now, we modify the MSR-method for Problem (8.2.39) as follows:

8.2.5 Method. (inexact MSR-method)
Given sequences $\{ \Psi_{k,i}(\cdot) \}$ and $\{ K_k \}$ via (8.2.40) and (8.2.41), respectively and $\{ \delta_k \}$ and $\{ \epsilon_k \}$ with
\[ \delta_k > 0, \quad \epsilon_k \geq 0, \quad \lim_{k \to \infty} \epsilon_k = 0. \]

For fixed $k$ and $i > 0$ the points $u^{k,i}$ are generated by
\[ \| \nabla \Psi_{k,i}(u^{k,i}) - \nabla \Psi_{k,i}(\bar{u}^{k,i}) \|_Y \leq \epsilon'_k, \] (8.2.42)
with
\[ \epsilon'_k \leq \frac{\beta}{\sqrt{M^2 + c^2}} \epsilon_k \] (8.2.43)
and
\[ \bar{u}^{k,i} := \arg \min_{u \in K_k} \Psi_{k,i}(u). \] (8.2.44)

If $\| \Pi_1 u^{k,i} - \Pi_1 u^{k,i-1} \|_H > \delta_k$, set $i(k) := i + 1$ and compute $u^{k,i+1}$, otherwise set $i(k) := i$, $u^{k+1,0} := u^{k,i}$ and compute $u^{k+1,1}$.

Further on in the applications the dependence between $\epsilon'_k$ and $\epsilon_k$ has to be specified.

8.2.6 Remark. Identifying Problem (8.2.39) with the two-body contact problem (8.2.13) or the Signorini problem (8.2.16) and supposing that the sets $K_k$ are results of a finite element approximation, we can specify Method 8.2.5 in order to apply it to these model problems. Recall that for the contact problems the space $V$ is defined according to (8.2.11) or (8.2.19).
Identifying \( Y = V, Y_1 = V, H = V, \Pi_1 = I \) (identity operator in \( V \)), Method 8.2.5 turns into the standard MSR-method 4.3.1. Moreover, taking \( b(u,v) := 0 \), we deal with \( \epsilon_k' = \epsilon_k \) in (8.2.43).

The choice \( Y = V, H = V, \Pi_1 = \hat{\Pi}_1, Y_1 = V_1 \), with \( V_1 \) defined according to (8.2.31) or (8.2.32), leads to a particular example of the method with regularization on the subspace (cf. Section 4.3.3), where the auxiliary problems are of the form (8.2.33). Here we have to suppose that

\[
b(u,v) := c_0 \int_{\Omega} \epsilon_{kl}(u)\epsilon_{kl}(v) d\Omega
\]

(concerning constant \( c_0 \) see (8.2.3)).

Finally, in the case of \( Y = V, Y_1 = V, H = [L^2(\Omega)]^2, \Pi_1 = \bar{\Pi}_1 \), Method 8.2.5 corresponds to the method of weak regularization, in which the iterates are computed by means of Problem (8.2.34). Again we have to take

\[
b(u,v) := c_0 \int_{\Omega} \epsilon_{kl}(u)\epsilon_{kl}(v) d\Omega,
\]

and the orthoprojector \( \bar{\Pi}_1 \) has to be considered as an operator mapping from \( V \) into \([L^2(\Omega)]^2\).

However, the validity of condition (8.2.35) has to be verified in each of the latter two cases.

Let us return to Problem (8.2.39). In order to prove the convergence of Method 8.2.5, we need the following statement, which turns out to be a generalization of Lemma 4.3.13.

8.2.7 Lemma. Suppose that Assumption 8.2.4 is fulfilled, that \( C \) is a convex, closed subset of the space \( Y \) and that \( z^0 \in Y \) is an arbitrarily chosen point. Let

\[
z^1 := \arg \min_{u \in C} \{ J^1(u) + \| \Pi_1 u - \Pi_1 z^0 \|_H^2 \}, \tag{8.2.45}
\]

with \( J^1(u) := J(u) + j(u) \) and \( j(\cdot) \) a convex lsc functional on \( Y \). Then for each \( u \in C \) the estimates

\[
|z^1 - u|^2 - |z^0 - u|^2 \leq -\|\Pi_1 z^1 - \Pi_1 z^0\|_H^2 + J^1(u) - J^1(z^1) \tag{8.2.46}
\]

and

\[
|z^1 - u| \leq |z^0 - u| + \eta(u) \tag{8.2.47}
\]

hold, where the norm \( |\cdot| \) is defined by (8.2.36) and

\[
\eta(u) = \begin{cases} 0 & \text{if } J^1(u) \leq J^1(z^1), \\ \sqrt{J^1(u) - J^1(z^1)} & \text{otherwise}. \end{cases}
\]

If moreover, \( \|\Pi_1 z^1 - \Pi_1 z^0\|_H \geq \delta \) and \( \delta \geq \eta(u) \), then

\[
|z^1 - u| \leq |z^0 - u| + \frac{\eta^2(u) - \delta^2}{2|z^0 - u|}. \tag{8.2.48}
\]
Using the definition of the norm $Y$, with regard to (8.2.37) the bilinear form
$$j(u) - j(z^1) + a(z^1, u - z^1) - \langle f, u - z^1 \rangle + 2\langle \Pi_1 u - \Pi_1 z^1, \Pi_1 z^1 - \Pi_1 z^0 \rangle_H \geq 0, \quad \forall u \in C.$$  
(8.2.49)

Using the definition of the norm $| \cdot |$, 
$$|z^1 - u|^2 - |z^0 - u|^2 =$$

$$= \frac{1}{2} b(z^1, z^1) - b(z^1, u) + b(z^0, u) - \frac{1}{2} b(z^0, z^0)$$
$$+ \|\Pi_1 z^1\|_H^2 - \|\Pi_1 z^0\|_H^2 - 2\langle \Pi_1 z^1, \Pi_1 u \rangle_H + 2\langle \Pi_1 z^0, \Pi_1 u \rangle_H$$

$$= \frac{1}{2} b(z^1, z^1) - b(z^1, u) + b(z^0, u) - \frac{1}{2} b(z^0, z^0)$$
$$- \|\Pi_1 z^1 - \Pi_1 z^0\|_H^2 + 2\langle \Pi_1 z^1 - \Pi_1 z^0, \Pi_1 z^1 - \Pi_1 u \rangle_H,$$

and in view of (8.2.49) and 
$$a(u, u) \geq b(u, u) \geq 0, \quad \forall u \in Y,$$

one can conclude that 

$$|z^1 - u|^2 - |z^0 - u|^2 \leq$$

$$\leq \frac{1}{2} b(z^1, z^1) - b(z^1, u) + b(z^0, u) - \frac{1}{2} b(z^0, z^0)$$
$$+ \frac{1}{2} b(u, u) - \frac{1}{2} b(u, u) + j(u) - j(z^1) + a(z^1, u) - a(z^1, z^1)$$
$$- \langle f, u - z^1 \rangle - \|\Pi_1 z^1 - \Pi_1 z^0\|_H^2$$

$$= \frac{1}{2} b(z^1 - u, z^1 - u) - \frac{1}{2} b(z^0 - u, z^0 - u) + j(u) - j(z^1) + a(z^1, u) - a(z^1, z^1)$$
$$- \langle f, u - z^1 \rangle - \|\Pi_1 z^1 - \Pi_1 z^0\|_H^2$$

$$\leq \frac{1}{2} a(z^1 - u, z^1 - u) - \frac{1}{2} a(z^0 - u, z^0 - u) + j(u) - j(z^1) + a(z^1, u) - a(z^1, z^1)$$
$$- \langle f, u - z^1 \rangle - \|\Pi_1 z^1 - \Pi_1 z^0\|_H^2$$

$$\leq \frac{1}{2} a(z^1, z^1) + \frac{1}{2} a(u, u) + j(u) - j(z^1) - \langle f, u - z^1 \rangle - \|\Pi_1 z^1 - \Pi_1 z^0\|_H^2$$

$$= J^1(u) - J^1(z^1) - \|\Pi_1 z^1 - \Pi_1 z^0\|_H^2.$$  

Now, the inequalities (8.2.47) and (8.2.48) follow immediately from the latter relation. \(\square\)

Similarly as in the method with regularization on the subspace described in Section 4.3.3 the basic statements about convergence of MSR-method 4.3.1 can be extended to Method 8.2.5.

Due to the proofs of the Lemmata 4.3.3, 4.3.5 and Theorem 4.3.6, we get the following result by means of Lemma 8.2.7.

8.2.8 Corollary. Assume that 

$$\Psi_{k,i} = \Psi_{k,i}(u) = J(u) + \|\Pi_1 u - \Pi_1 u^{k,i-1}\|_H^2,$$
and $\mathbb{B}_r := \{u \in Y : |u| \leq r\}$. The terms $L(r)$, $r^*$, $Q$, $Q_k$ and $Q^*$, appearing in (4.3.4), let be defined according to $\mathbb{B}_r$, $\Psi_{k,i}$, $\overline{\Psi}_{k,i}$ and $|\cdot|$. Moreover, the distances between points and sets in Assumption 4.2.2 let also be measured in the norm $|\cdot|$.

Then the statements of the Lemmata 4.3.3, 4.3.5 and of the Theorems 4.3.6 and 4.3.8 remain true for Method 8.2.5 if the norm $\|\cdot\|$ is replaced by $|\cdot|$ everywhere. Strong convergence of $\{u^{k,i}\}$ holds if $\dim Y_1 < \infty$.

**Comments on the proof:** The only substantial modification in the proofs of the statements mentioned above consists in the use of Lemma 8.2.7 (with $j(\cdot) \equiv 0$ wherever Proposition 3.1.3 is applied before).

Now we make some comments on those parts of the proofs whose adaption might be connected with some difficulties.

1. In (4.3.6), (4.3.16) and (4.3.34) the norm $\|\cdot\|$ has to be replaced by $|\cdot|$ and the points $\tilde{u}^{k,i}$ (see the proof of Theorem 4.3.6) have to be defined by

$$\tilde{u}^{k,i} := \arg \min_{v \in Q} |v - \overline{u}^{k,i}|.$$

2. Instead of inequality (4.3.8) we now obtain

$$|\overline{u}^{k_0,i} - v^{k_0}|^2 - |u^{k_0,i-1} - v^{k_0}|^2 \leq -\|\Pi_1 \tilde{u}^{k_0,i} - \Pi_1 u^{k_0,i-1}\|^2_H + \tau_{k_0}, \quad (8.2.50)$$

and, with regard to $\sigma_{k_0} = 0$, the value $\tau_{k_0}$ coincides with $2L(r)\mu_{k_0}$.

Because of (8.2.42) and the strong convexity of $\Psi_{k,i}(\cdot)$ (with constant $\beta$) on $Y$, the estimate

$$\|\tilde{u}^{k,i} - u^{k,i}\|_Y \leq \frac{1}{2\beta} \epsilon'_{k}$$

can easily be established.

On this way, due to (8.2.37) and (8.2.43), for each index pair $(k,i)$ we obtain

$$\|\Pi_1 \tilde{u}^{k,i} - \Pi_1 u^{k,i}\|_H \leq \frac{c}{2\beta} \epsilon'_{k} = \frac{c}{2 \sqrt{\frac{M}{2} + c^2}} \epsilon_{k} \leq \frac{1}{2} \epsilon_{k}.$$  \quad (8.2.51)

For $1 \leq i < i(k_0)$, due to

$$\|\Pi_1 u^{k_0,i} - \Pi_1 u^{k_0,i-1}\|_H > \delta_{k_0} > \frac{1}{2} \epsilon_{k_0}$$

and (8.2.50) and (8.2.51), it follows

$$|\tilde{u}^{k_0,i} - v^{k_0}|^2 - |u^{k_0,i-1} - v^{k_0}|^2 \leq -\left(\|\Pi_1 u^{k_0,i} - \Pi_1 u^{k_0,i-1}\|_H - \frac{\epsilon_{k_0}}{2}\right)^2 + \tau_{k_0}$$

and, instead of (4.3.10), it holds

$$|\tilde{u}^{k_0,i} - v^{k_0}| - |u^{k_0,i-1} - v^{k_0}| \leq \frac{-\epsilon_{k_0}^2 + \tau_{k_0}}{2|u^{k_0,i-1} - v^{k_0}|}. \quad (8.2.52)$$

Formula (4.3.18) has to be transformed analogously.

In the other places, where (4.3.9) was formerly used, we should now apply the inequality

$$|u^{k,i} - \tilde{u}^{k,i}| \leq \frac{1}{2} \epsilon_{k},$$

which is a conclusion of the relations (8.2.38), (8.2.43) and
\[ \| u^{k,i} - u^{k,i} \|_Y \leq \frac{1}{2\beta} \epsilon_k. \]

(3) Instead of the inequality (4.3.30) we have to use
\[ |u^{k,i} - u^{**}|^2 - |u^{k,i+1} - u^{**}|^2 \geq J(u^{k,i+1}) - J(u^{**}) + \| \Pi_1 u^{k,i+1} - \Pi_1 u^{k,i} \|_H^2, \]
which follows from Lemma 8.2.7.

8.2.3.2 Verification of condition (8.2.35)
The validity of inequality (8.2.35) is obvious for the standard MSR-method. In the other two methods, the method with regularization on a subspace and method of weak regularization (see Remark 8.2.6), in the case of the two-body contact problem we make use of the relation
\[ \int_{\Omega'} \epsilon_{kl}(u) \epsilon_{kl}(u) \, d\Omega \geq c_1 \| u \|^2_{1,\Omega'}, \quad (c_1 > 0), \]
which reflects the equivalence between the seminorm
\[ [u] := \sqrt{\int_{\Omega'} \epsilon_{kl}(u) \epsilon_{kl}(u) \, d\Omega} \]
and the norm \( \| u \|^2_{1,\Omega'} \) in the space \( \{ u \in [H^1(\Omega')]^2 : u|_{\Gamma_u} = 0 \} \) if \( \text{meas } \Gamma_u > 0 \) (cf. Duvaut and Lions [95], chapt.3, and Ciarlet [75], sect.1.2).

Together with the second Korn inequality used on the set \( \Omega'' \)
\[ \int_{\Omega''} \epsilon_{kl}(u) \epsilon_{kl}(u) \, d\Omega + \int_{\Omega''} u_k u_k \, d\Omega \geq c_2 \| u \|^2_{1,\Omega''}, \quad (c_2 > 0), \]
see, for instance, Fichera [113], chapt.1, and [95], chapt.3, relation (8.2.54) enables us immediately to establish the validity of condition (8.2.35) with
\[ \beta := \min[c_1, c_2] \min[\frac{c_0}{2}, 1] \]
in case the method with weak regularization is considered. Indeed, due to (8.2.54) and (8.2.55) we get
\[ \frac{1}{2} b(u, u) + \| \Pi_1 u \|^2_H = \frac{c_0}{2} \int_{\Omega} \epsilon_{kl}(u) \epsilon_{kl}(u) \, d\Omega + \int_{\Omega'} u_k u_k \, d\Omega \]
\[ = \frac{c_0}{2} \int_{\Omega'} \epsilon_{kl}(u) \epsilon_{kl}(u) \, d\Omega + \frac{c_0}{2} \int_{\Omega''} \epsilon_{kl}(u) \epsilon_{kl}(u) \, d\Omega + \int_{\Omega''} u_k u_k \, d\Omega \]
\[ \geq \frac{c_0 c_1}{2} \| u \|^2_{1,\Omega'} + \min[\frac{c_0}{2}, 1] c_2 \| u \|^2_{1,\Omega''}. \]

The corresponding analysis for the method with weak regularization on the subspace, which we are going to establish now, proves to be more complicated.
Let $\Theta_j$ be the ortho-projector mapping from \([L_2(\Omega')]^2\) into the linear set of kernel functions

\[
\mathcal{K}' := \{ z : z_1(x) = a_1 - bx_2, \ z_2(x) = a_2 + bx_1 \text{ on } \Omega' \},
\]

with $a_1, a_2, b$ arbitrary real numbers and $\Theta := \mathcal{I} - \Theta_j$. Here $\mathcal{I}$ is the identity operator on \([L_2(\Omega')]^2\).

Due to the projection, we get

\[
\int_{\Omega'} \epsilon_{kl}(u)\epsilon_{kl}(u) d\Omega = \int_{\Omega'} \epsilon_{kl}(\Theta u)\epsilon_{kl}(\Theta u) d\Omega, \quad \forall u \in [H^1(\Omega')]^2. \tag{8.2.56}
\]

Now, let us show that for some $c_3 > 0$ the following inequality is satisfied

\[
\int_{\Omega'} \epsilon_{kl}(u)\epsilon_{kl}(u) d\Omega \geq c_3\|\Theta u\|_{1,\Omega'}^2 \quad \forall u \in [H^1(\Omega')]^2. \tag{8.2.57}
\]

Supposing this is wrong, then a sequence \(\{u^j\} \subset [H^1(\Omega')]^2\) exists such that for $u^j := \Theta u^j$, due to (8.2.56), the relations

\[
\|u^j\|_{2,\Omega'} = 1 \tag{8.2.58}
\]

\[
\lim_{j \to \infty} \int_{\Omega'} \epsilon_{kl}(u^j)\epsilon_{kl}(u^j) d\Omega = 0 \tag{8.2.59}
\]

\[
\lim_{j \to \infty} \int_{\Omega'} \epsilon_{kl}(u^j)\epsilon_{kl}(u^j) d\Omega = 0
\]

are true. Without loss of generality, we assume here that the sequence \(\{u^j\}\) converges weakly in \([H^1(\Omega')]^2\).

Then, with regard to the compact embedding \([H^1(\Omega')]^2 \hookrightarrow [L_2(\Omega')]^2\), the sequence \(\{u^j\}\) converges in the norm of the space \([L_2(\Omega')]^2\) to some element $\bar{v} \in [H^1(\Omega')]^2$.

On account of the second Korn inequality (8.2.55), the estimate

\[
\|u^{j+p} - v^j\|_{1,\Omega'}^2 \leq \frac{1}{c_2} \left( \int_{\Omega'} \epsilon_{kl}(u^{j+p} - v^j)\epsilon_{kl}(u^{j+p} - v^j) d\Omega + \|u^{j+p} - v^j\|_{0,\Omega'}^2 \right) \tag{8.2.60}
\]

is fulfilled for all $j$ and $p$.

But, due to (8.2.59) and the strong convergence of \(\{u^j\}\) in \([L_2(\Omega')]^2\), inequality (8.2.60) implies that \(\{u^j\}\) converges to $\bar{v}$ in the norm of the space \([H^1(\Omega')]^2\).

Hence, according to (8.2.58), we get

\[
\|\bar{v}\|_{1,\Omega'}^2 = 1. \tag{8.2.61}
\]

On the one hand, the relation

\[
\int_{\Omega'} \epsilon_{kl}(\bar{v})\epsilon_{kl}(\bar{v}) d\Omega = 0,
\]

following from (8.2.59) and $\lim_{j \to \infty} \|v^j - \bar{v}\|_{1,\Omega'} = 0$ means that $\bar{v} \in \mathcal{R}'$ (cf. NeČAS and HLAVÁČEK [307]).

On the other hand, since $\lim_{j \to \infty} \|v^j - \bar{v}\|_{0,\Omega'} = 0$ and

\[
((v^j, z))_{0,\Omega'} = 0 \quad \forall z \in \mathcal{R}',
\]

we conclude that $\bar{v} \in \mathcal{R}'$. Thus, the inequality (8.2.57) is satisfied for all $u \in [H^1(\Omega')]^2$.
we obtain
\[(\bar{v},z)_{0,\Omega''} = 0 \quad \forall \ z \in \mathbb{R}''\].

Hence, \(\bar{v} = 0\), but this contradicts (8.2.61).

Due to the definition of the projectors \(\hat{\Pi}_1\) and \(\Theta_1\),
\[
\hat{\Pi}_1 u|_{\Omega''} = \Theta_1 u'', \quad \hat{\Pi}_1 u|_{\Omega'} = 0,
\]
and with regard to (8.2.57), for each \(u \in [H^1(\Omega'')]^2\) the inequality
\[
\frac{c_0}{2} \int_{\Omega''} \epsilon_{kl}(u)\epsilon_{kl}(u) d\Omega + \|\hat{\Pi}_1 u\|^2_{1,\Omega''} \geq \frac{c_0c_3}{4} \|\Theta_1 u\|^2_{1,\Omega''} + \|\hat{\Pi}_1 u\|^2_{1,\Omega''}
\]
holds true. Therefore,
\[
\frac{c_0}{2} \int_{\Omega''} \epsilon_{kl}(u)\epsilon_{kl}(u) d\Omega + \|\hat{\Pi}_1 u\|^2_{1,\Omega''} \geq \frac{c_0c_3}{4} \|\Theta_1 u\|^2_{1,\Omega''} + \|\hat{\Pi}_1 u\|^2_{1,\Omega''} \geq \min \left[ \frac{c_0c_3}{4}, 1 \right] \|u\|^2_{0,\Omega''}
\]
and in view of (8.2.55), we finally obtain
\[
\frac{c_0}{2} \int_{\Omega''} \epsilon_{kl}(u)\epsilon_{kl}(u) d\Omega + \|\hat{\Pi}_1 u\|^2_{1,\Omega''} \geq c_4 \|u\|^2_{1,\Omega''}, \quad (c_4 > 0).
\]

Now, relation (8.2.35) follows immediately from the latter inequality together with (8.2.54).

The same arguments enable us also to conclude that inequality (8.2.35) holds true if the considered regularization methods are applied to the Signorini problem.

Let us turn to the question about the rate of convergence of Method 8.2.5. Regarding the proofs of the Theorems 5.1.1, 5.1.3, 5.1.5 and 5.1.7, we can establish the following result.

8.2.9 Corollary. The statements of the rate of convergence of the MSR-method 4.3.1 remain true also for Method 8.2.5 if

(i) the alterations made in Corollary 8.2.8 are maintained;

(ii) in the Assumptions 5.1.4, 5.1.6 and in condition (i) of Theorem 5.1.1 the distance function \(\rho\) is defined according to the norm of the space \(Y\) (cf. (8.2.36)).

Comments on the proof: We start with the beginning of the proof of Theorem 5.1.1. Let the points \(v^{k,i}, \bar{v}^{k,i}, w^{k,i}\) be determined with respect to the norm \(|\cdot|\), defined by (8.2.36), and the function \(\eta_{k,i}\) be expressed by
\[
\eta_{k,i}(\lambda) = \lambda J(\bar{v}^{k,i}) + (1 -\lambda) J(w^{k,i}) + \lambda^2 \|\Pi_1 \bar{v}^{k,i} - \Pi_1 w^{k,i}\|^2_H.
\]

Then using the relations
\[
|\bar{v}^{k,i} - v^{k,i}| \leq \frac{\epsilon_k}{2} \quad \text{and} \quad \|\Pi_1 u\|_H \leq |u|,
\]
we can conclude (as in the previous argumentation) that the inequalities (5.1.11), (5.1.12), (5.1.15), (5.1.16), (5.1.34) and (5.1.37) remain true if the term $\|u^{k,i+1} - u^{k,i}\|^2$ is replaced by $\|\Pi_1 u^{k,i+1} - \Pi_1 u^{k,i}\|^2_{H_0}$.

Now the continuation of the proof requires only formal corrections which are based on these replacements.

Assumption 5.1.6 and especially Assumption 5.1.4 seem to be natural for variational inequalities. However, their verification causes serious difficulties. We shall consider this question in Section 8.4.

8.2.4 On choice of variants of IPR-methods

In Section 4.3, by means of Example 4.3.15, it was shown that one can accelerate OSR- or MSR-processes if a more suitable variant as the standard regularization is used. In this example regularization on a subspace, coinciding with the kernel of the bilinear form of the objective functional, proves to be more efficient.

Now, we analyze this question for variational inequalities, considering the three variants of MSR-methods described in Section 8.2.3.

First, we notice that the order of the condition number of the Hessians in the regularized and discretized problems is of the same magnitude in all the three variants if the approximation is performed by finite element methods with quasi-uniform triangulation.

Applying regularization on a subspace (see Method 8.2.5, with $\Pi_1$ an ortho-projector on the kernel of the bilinear form of the objective functional $J$), on rough triangulation grids, the sequence $\{(I - \Pi_1)u^{k,i}\}$ draws close substantially faster to $(I - \Pi_1)u^*, u^* \in U^*$ than the corresponding projections of solutions obtained by the other two methods. Moreover, the values of the objective functional decreases faster, too.

8.2.10 Remark. In order to understand the advantages of the regularization on the subspace, we suggest to analyze all the three variants for solving the simple model problem

$$\min_{u \in H^1(\Omega)} \{J(u) := \frac{1}{2} \|\nabla v\|^2_{0,\Omega} - \langle f, u \rangle_{0,\Omega}\},$$

assuming that $\langle f, u \rangle_{0,\Omega} = 0$, $K_k = K \forall k$ and $\Pi_1$ is the ortho-projector on the kernel in the space $L_2(\Omega)$.

Supposing that $U^* \neq 0$, this example should be considered for the following tree cases:

(i) $\|\Pi_1 u^* - \Pi_1 u^{1,0}\|_{0,\Omega}$ is large;

(ii) $\|\Pi_1 u^* - \Pi_1 u^{1,0}\|_{0,\Omega} \approx 0$, but $\|u^* - u^{1,0}\|_{0,\Omega}$ is large;

(iii) $\|\Pi_1 u^* - \Pi_1 u^{1,0}\|_{0,\Omega} = 0$.

We emphasize that additional numerical expense connected with the projection on the kernel proves to be insignificant for contact problems, because due to the structure of the kernel $K$, each projection operation requires only to solve a system of three linear equations with respect to $a_1$, $a_2$ and $b$. 

\[ \boxdot \]
However, at the final stage of MSR-methods, especially, if fast convergent algorithms for the solution of the auxiliary problems are used, the other two variants may be found more efficient. In case $K_k \subset K$, the transition from one regularization variant to any other during the solution procedure does not require a correction of the triangulation parameter $h_k$ and accuracy parameter $\epsilon_k$, but $\delta_k$ has to satisfy condition (8.2.76) with the corresponding new values $L(r)$ and $\bar{\mu}_k$. Recall that $L(r)$ and $\bar{\mu}_k$ depend on the new norm in the space $V$, introduced according to the chosen regularization approach (see Section 8.2.3.1).

If we will see in several numerical examples described later on in Subsections 8.2.7 and 8.3.3, experiments with MSR- and OSR-methods for solving special classes of Problem (8.1.1) show that weak regularization is more preferable than the standard variant. This can be expected taking into account the structure of the kernel of the bilinear form.

The following pure analytical investigation is going to emphasize this advantage.

8.2.11 Example. Let $\Omega := \{(x,y) : -\frac{\pi}{2} < x,y < \frac{\pi}{2}\}$, $V := H^1(\Omega)$. Consider the variational inequality

$$\text{find } u \in K : \quad \langle Q(u), u - v \rangle \geq 0 \quad \forall v \in K,$$

with

$$K := V, \quad Q : u \mapsto -\Delta u - f,$$

$f$ is given by $f(x,y) := 2 \sin x \sin y$.

Obviously this is equivalent to the Neumann problem

$$-\Delta u = 2 \sin x \sin y, \quad \frac{\partial u}{\partial n} |_{\partial \Omega} = 0,$$

whose solution set is

$$U^* := \{\sin x \sin y + d : d \in \mathbb{R}\}.$$

Applying the exact proximal point method 4.2.1 with $u^1 := 0$, $\chi_k \equiv 1$, we obtain a sequence $\{u^k\} \subset H^2(\Omega)$, where $u^{k+1}$ is the unique solution of the boundary value problem

$$-2\Delta u + u = 2 \sin x \sin y - \Delta u^k + u^k, \quad \frac{\partial u}{\partial n} |_{\partial \Omega} = 0. \quad (8.2.62)$$

It is not difficult to verify (by means of an immediate substitution) that

$$u^k = a_k \sin x \sin y,$$

where $(1 - a_k) = \frac{3}{5}(1 - a_{k-1})$, $a_1 = 0$.

With $u^* = \sin x \sin y \in U^*$ one gets

$$u^* - u^k = \left(\frac{3}{5}\right)^{k-1} \sin x \sin y.$$

But, replacing in this method the classical regularizing functional

$$h : u \mapsto \frac{1}{2} \|u\|_{H^1(\Omega)}^2 \quad \text{by} \quad \hat{h} : u \mapsto \frac{1}{2} \|u\|_{L^2(\Omega)}^2,$$
a sequence \( \{v^k\} \subset H^2(\Omega) \), \( v^1 := u^1 = 0 \), is generated, where \( v^{k+1} \) is the unique solution of the problem

\[
-\Delta v + v = 2 \sin x \sin y + v^k, \quad \frac{\partial v}{\partial n} |_{\partial \Omega} = 0.
\]

(8.2.63)

Here

\( v^k = b_k \sin x \sin y \),

holds, with \((1 - b_k) = \frac{1}{3}(1 - b_{k-1}) \), \( b_1 := a_1 = 0 \). Hence,

\[
u^* - v^k = \left( \frac{1}{3} \right)^{k-1} \sin x \sin y.
\]

To compare numerically these solutions, for instance, we have

\[
\|u^* - v^6\| \approx 0.053, \quad \|u^* - u^{11}\| \approx 0.0028, \quad \text{etc.}
\]

Moreover, under identical finite element approximations in problems (8.2.62) and (8.2.63), the conditioning of the discretized problems for (8.2.63) is, at least, not worse than in (8.2.62).

\( \Box \)

**8.2.5 Special Case: Inner approximation of the set \( K \)**

Throughout this section we suppose that Assumption 8.2.4 is fulfilled and Problem (8.2.39) is considered in the space \( Y \) (cf. (8.2.36)), i.e., statements on convergence of the methods are studied in this space. However, from the numerical point of view it makes sense to preserve relation (8.2.42) in its original form, i.e., in (8.2.42) the norm of the dual space \( Y' \) is used.

If the sets \( K_k \) in the auxiliary problems (8.2.40), (8.2.41) have the property \( K_k \subset K, \forall i \), convergence of Method 8.2.5 can be established under essential weaker conditions with respect to the controlling parameters than those in the modified Theorems 4.3.6 and 4.3.8 (cf. also Corollary 8.2.8).

**8.2.12 Assumption.** Fix a solution \( u^{**} \in U^* \) of Problem (8.2.39) and

(i) for each \( k = 1, 2, \ldots \) the inclusion \( K_k \subset K \) and the inequality

\[
\rho(u^{**}, K_k) \leq \mu_k
\]

(8.2.64)

hold, with \( \{\mu_k\} \downarrow 0 \);

(ii) for a given \( r_0 \geq \sup_k \mu_k \) the functional \( J \) satisfies a Lipschitz condition with the constant \( L_0 \) on \( B_{r_0}(u^{**}) \);

(iii) \( \sum_{k=1}^{\infty} \sqrt{\mu_k} < \infty, \sum_{k=1}^{\infty} \epsilon_k < \infty \) hold true, where \( \{\epsilon_k\} \) is defined according to Method 8.2.5.

\( \Box \)

**8.2.13 Lemma.** Let Assumption 8.2.12 be fulfilled, radius \( r^* \) be chosen such that \( r^* \geq 8r_0 \) and

\[
\sum_{k=1}^{\infty} \left( \sqrt{L_0 \mu_k} + \frac{\epsilon_k}{2} + 2\mu_k \right) < \frac{r^*}{2}.
\]

(8.2.65)
Moreover, assume that parameter \( \delta_k \) satisfies the condition
\[
\frac{1}{2r^*} \left( L_0 \mu_k - (\delta_k - \frac{\epsilon_k}{2})^2 \right) + \frac{\epsilon_k}{2} < 0, \quad k = 1, 2, \ldots
\] (8.2.66)

Then, for Method 8.2.5, starting with \( u^{1,0} \in B_{r^*/4}(u^{**}) \), the interior loop is finite, i.e., \( i(k) < \infty \) for each \( k \), and the inclusions
\[
\bar{u}^{k,i} \in \text{int} B_r(u^{**}), \quad u^{k,i} \in \text{int} B_{r^*/4}(u^{**})
\]
are valid for all pairs \((k, i)\).

**Proof:** Consider a fixed pair \((k, i)\) with \( i \geq 1 \). In view of \( K_k \subset K \) and (8.2.44) the relation \( J(u^{**}) \leq J(\bar{u}^{k,i}) \) holds. Choosing \( v^k \in K_k \) such that
\[
|v^k - u^{**}| \leq \mu_k,
\]
Assumption 8.2.12(ii) provides that
\[
J(v^k) \leq J(u^{**}) + L_0 \mu_k,
\]
and
\[
J(v^k) \leq J(\bar{u}^{k,i}) + L_0 \mu_k.
\] (8.2.67)

Following the proof of Lemma 4.3.5, one can establish
\[
\delta_k - \frac{\epsilon_k}{2} > \sqrt{L_0 \mu_k}.
\]

In case \( i < i(k) \), with regard to (8.2.51) (see the comment on the proof of Corollary 8.2.8), it is possible to conclude that
\[
\|\Pi_1 \bar{u}^{k,i} - \Pi_1 u^{k,i-1}\|_H > \|\Pi_1 v^k - \Pi_1 u^{k,i} - \Pi_1 \bar{u}^{k,i} - \Pi_1 v^k\|_H > \sqrt{L_0 \mu_k}.
\]

Taking into account the inequalities (8.2.67) and (8.2.48), the use of Lemma 8.2.7 with \( C := K_k \), \( z^0 := u^{k,i-1} \) and \( u := v^k \) leads to
\[
|\bar{u}^{k,i} - v^k| \leq |u^{k,i-1} - v^k| + \frac{L_0 \mu_k - (\delta_k - \frac{\epsilon_k}{2})^2}{2|u^{k,i-1} - v^k|}, \quad \text{for } 1 \leq i < i(k). \] (8.2.68)

If \( i = i(k) \), Lemma 8.2.7 gives immediately
\[
|\bar{u}^{k,i(k)} - v^k| < |u^{k,i(k)-1} - v^k| + \sqrt{L_0 \mu_k}.
\] (8.2.69)

Hence, in view of
\[
|u^{k,i} - \bar{u}^{k,i}| \leq \frac{\epsilon_k}{2},
\]
the inequalities
\[
|u^{k,i} - v^k| \leq |u^{k,i-1} - v^k| + \frac{L_0 \mu_k - (\delta_k - \frac{\epsilon_k}{2})^2}{2|u^{k,i-1} - v^k|} + \frac{\epsilon_k}{2} \quad \text{for } 1 \leq i < i(k) \] (8.2.70)

and
\[
|u^{k,i(k)} - v^k| \leq |u^{k,i(k)-1} - v^k| + \sqrt{L_0 \mu_k} + \frac{\epsilon_k}{2} \] (8.2.71)
CHAPTER 8. PPR FOR VARIATIONAL INEQUALITIES

hold. Using the assumptions that \( u^{1,0} \in \mathbb{B}_{r^*/4}(u^*) \) and \( r^* \geq 8r \), in case \( i(1) > 1 \) we obtain from (8.2.68), (8.2.70) and (8.2.66) the estimate

\[
\max [ |u^{1,1} - v^1|, |u^{1,1} - v^1| ] \leq |u^{1,0} - v^1| + \frac{1}{2r^*} \left( L_0 \mu_1 - (\delta_1 - \frac{\epsilon_1}{2})^2 \right) + \frac{\epsilon_1}{2}
\]

\[
< |u^{1,0} - v^1| < \frac{r^*}{2}, \quad (8.2.72)
\]

and in case \( i(1) = 1 \), due to (8.2.69), (8.2.71) and (8.2.65)

\[
\max [ |\bar{u}^{1,1} - v^1|, |u^{1,1} - v^1| ] \leq |u^{1,0} - v^1| + \sqrt{L_0 \mu_1} + \frac{\epsilon_1}{2} < r^*. \quad (8.2.73)
\]

Analogously to (8.2.72) we get

\[
|u^{1,i} - v^1| < |u^{1,0} - v^1| < \frac{r^*}{2} \quad \text{for } i < i(1), \quad (8.2.74)
\]

and the finiteness of \( i(1) \) can be established as in Lemma 4.3.3.

In view of \( |v^k - u^*| \leq \mu_k \), the relations (8.2.72) and (8.2.73) lead to

\[
\max [ |\bar{u}^{1,i} - u^*|, |u^{1,i} - u^*| ] \leq |u^{1,0} - u^*| + \sqrt{L_0 \mu_1} + \frac{\epsilon_1}{2} + 2\mu_1
\]

\[
< r^* \quad \text{for } 1 \leq i \leq i(1).
\]

Now, for the starting point \( u^{2,0} \) on the level \( k = 2 \) the following inequality is valid

\[
|u^{2,0} - v^2| \leq |u^{1,0} - u^*| + \sqrt{L_0 \mu_1} + \frac{\epsilon_1}{2} + 2\mu_1 + \mu_2 < r^*. \quad (8.2.75)
\]

Continuation of this procedure for \( k = 2, 3, \ldots \) provides step by step the following estimates:

\[ \diamond \quad i(k) < \infty; \]

\[ \diamond \quad \text{for } 1 \leq i < i(k); \]

\[
\max [ |u^{k,i} - v^k|, |v^{k,i} - v^k| ] \leq |u^{k-1,i} - v^k| + \frac{1}{2r^*} \left( L_0 \mu_k - (\delta_k - \frac{\epsilon_k}{2})^2 \right) + \frac{\epsilon_k}{2};
\]

\[ \diamond \quad \text{for } 1 \leq i \leq i(k); \]

\[
\max [ |u^{k,i} - v^k|, |v^{k,i} - v^k| ] \leq |u^{1,0} - u^*| + \sum_{s=1}^{k-1} \left( \sqrt{L_0 \mu_s} + \frac{\epsilon_s}{2} + 2\mu_s \right) + \sqrt{L_0 \mu_k} + \frac{\epsilon_k}{2} + \mu_k;
\]

\[ \diamond \quad \text{for } 1 \leq i \leq i(k); \]

\[
\max [ |u^{k,i} - u^*|, |v^{k,i} - u^*| ] \leq |u^{1,0} - u^*| + \sum_{s=1}^{k} \left( \sqrt{L_0 \mu_s} + \frac{\epsilon_s}{2} + 2\mu_s \right) < r^*\]

\[ \diamond \]

\[
\max [ |u^{k+1,0} - v^{k+1}|, |v^{k+1,0} - v^{k+1}| ] \leq |u^{1,0} - u^*| + \sum_{s=1}^{k} \left( \sqrt{L_0 \mu_s} + \frac{\epsilon_s}{2} + 2\mu_s \right) + \mu_{k+1} < r^*.
\]
8.2. CONTACT PROBLEMS WITHOUT FRICTION

Hence, we can conclude that
\[ \bar{u}^{k,i} \in \text{int} B_r(u^{**}), \quad u^{k,i} \in \text{int} B_r(u^{**}) \quad \text{for all } k,i. \]

\[ \square \]

8.2.14 Theorem. Let \( r \geq r^* \) be fixed and assume that the following conditions are fulfilled:

(i) the hypotheses of Lemma 8.2.13;

(ii) \( \rho(Q^*, Q_k) \leq \bar{\mu}_k, \) \( k = 1, 2, \ldots \), where \( \bar{\mu}_k \leq c_0 \mu_k \) (with some constant \( c_0 \)), \( Q^* = U^* \cap B_r(u^{**}) \) and \( Q_k = K_k \cap B_r(u^{**}) \);

(iii) functional \( J \) satisfies a Lipschitz-condition with constant \( L(r) \) on \( B_r(u^{**}) \);

(iv) sequence \( \{ \delta_k \} \) is chosen such that
\[
\frac{1}{4r} \left( L(r) \bar{\mu}_k - \left( \delta_k - \frac{\epsilon_k}{2} \right)^2 \right) + \frac{\epsilon_k}{2} < 0. \quad (8.2.76)
\]

Then the sequence \( \{ u^{k,i} \} \), generated by Method 8.2.5 with a starting point \( u^{1,0} \in B_{r/4}(u^{**}) \), converges weakly to some solution \( u^* \) of Problem (8.2.39).

Proof: Under the obvious assumption \( L_0 \leq L(r) \) condition (8.2.66) is an evident consequence of (8.2.76).

Let be chosen \( w \in U^* \cap B_r(u^{**}) \) arbitrarily and the point \( v^k \in Q_k \) be defined such that
\[
|v^k - w| \leq \bar{\mu}_k. \quad (8.2.77)
\]

Then, due to hypothesis (iii),
\[
J(v^k) \leq J(\bar{u}^{k,i}) + L(r) \bar{\mu}_k.
\]

Lemma 8.2.13 ensures that \( u^{k,i} \in \text{int} B_r(u^{**}) \) for all pairs \( (k,i) \), consequently,
\[
|u^{k,i} - v^k| \leq |u^{k,i} - u^{**}| + |v^k - u^{**}| < 2r.
\]

For a fixed index \( k \), using (8.2.76) and Lemma 8.2.7 with \( C := Q_k, u := v^k, z^0 := u^{k,i-1} \), we obtain similarly as at the beginning of the proof of Lemma 8.2.13
\[
|\bar{u}^{k,i} - v^k| \leq |u^{k,i-1} - v^k| + \frac{1}{4r} \left( L(r) \bar{\mu}_k - \left( \delta_k - \frac{\epsilon_k}{2} \right)^2 \right), \quad 1 \leq i < i(k),
\]
and
\[
|\bar{u}^{k,i(k)} - v^k| \leq |u^{k,i(k)-1} - v^k| + \sqrt{L(r) \bar{\mu}_k}.
\]

This leads, in view of \( |\bar{u}^{k,i} - u^{k,i}| \leq \frac{\epsilon}{r} \) and (8.2.76), to
\[
|u^{k,i} - v^k| < |u^{k,i-1} - v^k|, \quad 1 \leq i < i(k), \quad (8.2.78)
\]
and
\[
|u^{k,i(k)} - v^k| \leq |u^{k,i(k)-1} - v^k| + \sqrt{L(r) \bar{\mu}_k} + \frac{\epsilon_k}{2}. \quad (8.2.79)
\]
Therefore,
\[ |u^{k+1,0} - v^k| \leq |u^{k,0} - v^k| + \sqrt{L(r)\bar{\mu}_k} + \frac{\epsilon_k}{2}, \]
and, due to (8.2.77), the estimate
\[ |u^{k+1,0} - w| \leq |u^{k,0} - w| + \sqrt{L(r)\bar{\mu}_k} + \frac{\epsilon_k}{2} + 2\bar{\mu}_k \]
is satisfied. Because of \( \sum_{k=1}^{\infty} \sqrt{\bar{\mu}_k} < \infty \) and \( \sum_{k=1}^{\infty} \epsilon_k < \infty \), Lemma A3.1.4 ensures convergence of \( \{ |u^{k,0} - w| \} \) for each \( w \in U^* \cap B_r^+ (u^{**}) \).

Now, from (8.2.78) and (8.2.79) we can conclude that
\[ -\frac{\epsilon_k}{2} - \sqrt{L(r)\bar{\mu}_k} + |u^{k+1,0} - v^k| \leq |u^{k,i} - v^k| \leq |u^{k,0} - v^k| \]
and hence,
\[ -\frac{\epsilon_k}{2} - \sqrt{L(r)\bar{\mu}_k} - 2\bar{\mu}_k + |u^{k+1,0} - w| < |u^{k,i} - w| < |u^{k,0} - w| + 2\bar{\mu}_k. \]
Thus, \( \{ |u^{k,i} - w| \} \) converges for each \( w \in U^* \cap B_r^+ (u^{**}) \) and it is obvious that \( \{ |u^{k,i} - w| \} \) converges to the same limit.

Lemma 8.2.7 applied with the former data gives also (cf. (8.2.46))
\[ |u^{k,i-1} - v^k|^2 - |\bar{u}^{k,i} - v^k|^2 \geq J(\bar{u}^{k,i}) - J(v^k), \quad (8.2.80) \]
and using the inequalities
\[ |u^{k,i-1} - v^k| \leq |u^{k,i-1} - w| + \bar{\mu}_k, \]
\[ |\bar{u}^{k,i} - w| \leq |\bar{u}^{k,i} - v^k| + \bar{\mu}_k, \]
we obtain from (8.2.80)
\[ |u^{k,i-1} - w|^2 - |\bar{u}^{k,i} - w|^2 \geq J(\bar{u}^{k,i}) - J(w) - L(r)\bar{\mu}_k - 8r\bar{\mu}_k - 2\bar{\mu}_k^2. \quad (8.2.81) \]
Because the limits of \( \{ |u^{k,i} - w| \} \) and \( \{ |\bar{u}^{k,i} - w| \} \) coincide and \( J(\bar{u}^{k,i}) \geq J(w) \), inequality (8.2.81) ensures that
\[ \lim_{k \to \infty} \sup_{1 \leq i \leq \bar{i}(k)} (J(\bar{u}^{k,i}) - J(w)) = 0. \]
Moreover, in a standard way we can establish that any weak cluster point of the sequence \( \{ \bar{u}^{k,i} \} \) belongs to \( K \cap B_r^+ (u^{**}) \). Hence, Opial’s Lemma A1.1.3 yields that \( \{ u^{k,i} \} \) and \( \{ \bar{u}^{k,i} \} \) converge weakly to some \( u^* \in U^* \).

8.2.15 Remark. In view of the finite dimensionality of the subspace (8.2.31) and the inequalities (8.2.54) and (8.2.57), which imply strong convexity of the objective functional on the orthogonal complement of this subspace, Theorem 8.2.14 guarantees strong convergence of the sequence \( \{ u^{k,i} \} \) to some solution of the two-body contact problem. FICHERA [114] has verified the validity of Assumption 4.2.3 for the Signorini problem with \( V_1 \) defined by (8.2.32).
### 8.2. CONTACT PROBLEMS WITHOUT FRICTION

#### 8.2.16 Remark

Comparing the results of this section with the general statement on convergence of MSR-methods, we emphasize that the sets \( Q_k, Q \) and the values \( \mu_k \) are defined here somewhat differently than in Section 4.3.2. In comparison with the choice of the controlling parameters in (4.3.14), here we have essentially weaker requirements for \( \{\mu_k\} \) and \( \{\epsilon_k\} \): In principle, it is sufficient to guarantee only convergence of the series \( \sum_{k=1}^{\infty} \sqrt{\mu_k} \) and \( \sum_{k=1}^{\infty} \epsilon_k \). After choosing \( \{\mu_k\} \) and \( \{\epsilon_k\} \) the Lipschitz constant \( L_0 \) has to be determined and \( r^* \), \( r \geq r^* \) can be chosen (in contrast to (4.3.14)) such that

\[
\max \left[ 2 \sum_{k=1}^{\infty} \left( \sqrt{L_0 \mu_k + \frac{\epsilon_k}{2}} + 2 \mu_k \right), 8r_0 \right] < r^*.
\]

Note that the left part of this inequality does not depend on \( r \). Thereafter \( \{\delta_k\} \) has to be determined according to inequality (8.2.76).

In Theorem 8.2.14 set \( Q^* \) coincides with a part of the optimal set \( U^* \), whereas due to (4.3.4) \( Q^* \) includes also points

\[
u_{k,i} := \arg \min_{u \in Q} \Psi_{k,i}(u).
\]

Therefore, in general, more precise estimates in approximating \( Q^* \) by \( Q_k \) or by \( K_k \) can be obtained from hypothesis (ii) in Theorem 8.2.14 than from Assumption 4.2.2 (cf. condition (4.2.6)).

The choice of the sequences \( \{\mu_k\} \) and \( \{\hat{\mu}_k\} \) according to

\[
\rho(u^{**}, K_k) \leq \mu_k \quad \text{and} \quad \rho(Q^*, Q_k) \leq \hat{\mu}_k \quad \text{or} \quad \rho(Q^*, K_k) \leq \hat{\mu}_k
\]

depends on the particular data of the problem and on the chosen metric in the space \( \mathcal{Y} \).

We study this question for two-body contact problems, restricting our consideration to the standard MSR-method 4.3.1. The corresponding analysis for the other two variants of MSR-methods and also for the Signorini problem can be carried out analogously.

According to Remark 8.2.6, in this case one has to take \( \mathcal{Y} := V \) with \( V \) defined by (8.2.11). The norm of \( u = (u', u'') \) in the space \([H^s(\Omega)]^2 \times [H^s(\Omega')]^2\) \((s > 0\) integer) is defined by

\[
\|u\|_{s,\Omega} := \sqrt{\|u'\|^2_{s,\Omega'} + \|u''\|^2_{s,\Omega'}}.
\]

Assuming that a solution \( \bar{u} \) of the two-body contact problem belongs to the space \([H^2(\Omega)]^2 \times [H^2(\Omega')]^2\) and taking into account the structure of the solution set (cf. Theorem 8.2.2), any other solution \( \bar{u} \) is also contained in this space, moreover,

\[
\|\bar{u} - \bar{u}\|_{1,\Omega} = \|\bar{u} - \bar{u}\|_{2,\Omega}.
\]

Therefore, for any radius \( r_1 \) the set \( U^* \cap B_{r_1}(u^{**}) \) is also bounded in the space \([H^2(\Omega)]^2 \times [H^2(\Omega')]^2\), i.e.,

\[
\|u\|_{2,\Omega} \leq \|u^{**}\|_{2,\Omega} + r_1, \quad \forall \ u \in U^* \cap B_{r_1}(u^{**}).
\]

Hence, on account of Theorem A2.2.5, the interpolant \( u_{I,h} \) (see Definition A2.2.4) of each function \( u \in U^* \cap B_{r_1}(u^{**}) \) on the triangulation \( T_h \) yields

\[
\|u - u_{I,h}\|_{1,\Omega} \leq c(\|u^{**}\|_{2,\Omega} + r_1) h,
\]

where \( c \) is a positive constant depending only on \( \mathcal{Y} \).

#### 4.2.2 Theory of Approximation

### A2.2.5 Theorem

Theorem 4.2.2 ensures that the norm of the interpolant \( u_{I,h} \) in the space \( \mathcal{Y} \) is bounded by the norm of the function \( u \in U^* \cap B_{r_1}(u^{**}) \) on the triangulation \( T_h \), where \( c \) is a positive constant depending only on \( \mathcal{Y} \).

\[
\|u - u_{I,h}\|_{s,\Omega} \leq c(\|u^{**}\|_{s,\Omega} + r_1) h,
\]

where \( s \) is an integer satisfying the requirements (4.2.6) for the choice of the sequences \( \{\mu_k\} \) and \( \{\epsilon_k\} \), and \( r_1 \) is the radius of the ball on the triangulation \( T_h \).
with $c$ independent of $u$ and $r_1$.

But for $h := h_k$, due to (8.2.12), (8.2.25) and the construction of $\mathcal{B}_k$, one can conclude that $u_{I,h_k} \in K_k$. This enables us to define immediately $\{\mu_k\}$ and $\{\bar{\mu}_k\}$ according to Theorem 8.2.14.

Indeed, if

$$\sup_k h_k \leq \frac{r_1}{c(c_1 + r_1)},$$

with arbitrarily chosen radius $r_1$ and an upper bound $c_1$ for $\|u^{**}\|_{2,\Omega}$, the sequence

$$\mu_k := c(c_1 + r_1)h_k, \quad k = 1, 2, ...$$

(8.2.85)

satisfies the inequality $\mu_k \leq r_1$. Thus, we can identify the radius $r_0$ with $r_1$ (see Lemma 8.2.13), and $\sum_{k=1}^{\infty} \sqrt{h_k} < \infty$ ensures that $\sum_{k=1}^{\infty} \sqrt{\mu_k} < \infty$.

Now, we determine $r^*$ according to (8.2.65), $r^* > 8r_0$ and take

$$\bar{\mu}_k := c(c_1 + r^*)h_k, \quad k = 1, 2, ...$$

(8.2.86)

Using (8.2.84) with $r_1 := r^*$, $h := h_k$, we obtain

$$\|u - u_{I,h_k}\|_{1,\Omega} \leq \bar{\mu}_k$$

(8.2.87)

and, because of $u_{I,h_k} \in K_k$, the estimate $\rho(u,K_k) \leq \bar{\mu}_k$ is valid for any $u \in U^* \cap \mathbb{B}_{r^*}(u^{**})$.

Resuming this analysis, we have to choose the sequences $\{h_k\}, \{\varepsilon_k\}$ such that $\sum_{k=1}^{\infty} \sqrt{h_k} < \infty$, $\sum_{k=1}^{\infty} \varepsilon_k < \infty$, and thereafter $\{\delta_k\}$ has to be defined by (8.2.76), with $r := r^* + \bar{r}$ and $\bar{r} \geq \sup_k \mu_k$. On this way the sequence $\{u_{I,h_k}\}$, calculated by MSR-method 4.3.1, converges in the norm of the space $V$ to some solution of the two-body contact problem.

The crucial point is relation (8.2.83). It makes sense to extend this consideration to other monotone variational inequalities provided their solution sets fulfil a property similar to (8.2.83).

However, it should be remarked that the verification of the conditions ensuring the previously made assumption

$$U^* \subset [H^2(\Omega')]^2 \times [H^2(\Omega'')]^2,$$

is very complicated (cf. FICHERA [114] about smoothness of the solutions of Signorini problems).

KUSTOVA [252] has investigated an OSR-method for two-body contact problems under the assumption that the sought solutions have the form $\sum \gamma_p \eta_p + \omega$, where $\eta_p$ are known functions belonging to $[H^1(\Omega')]^2 \times [H^1(\Omega'')]^2$ with compact supports in the neighborhoods of break-points of the boundary and points, in which the type of the boundary condition changes; $\omega$ is supposed to belong to $[H^2(\Omega')]^2 \times [H^2(\Omega'')]^2$ (concerning such assumptions for problems in elasticity theory, see VOROVICH AT AL. [414]). In this case the values $\rho(Q^*, Q_k)$ and $\rho(Q_k, Q)$ are of order $\sqrt{h_k}$. 
8.2. CONTACT PROBLEMS WITHOUT FRICTION

8.2.6 Exactness of approximations of feasible sets

If the inclusion $K_k \subset K$ cannot be guaranteed, the application of MSR-method 4.3.1 to variational inequalities is more complicated. In particular, this concerns the changing rule of the triangulation parameter in order to satisfy the conditions (4.2.6) and (4.3.14). Recall that for MSR-methods the set $Q^*$ in (4.2.6) is defined according to (4.3.4).

For certain variational inequalities disturbance of the inclusion $K_k \subset K$ can be caused not only by the geometry of the domain $\bar{\Omega}$ and the chosen triangulation. For instance, in our Model Problem (8.1.1) this inclusion can be disturbed, even if $\bar{\Omega}$ is a rectangle and a uniform triangulation sequence is used. In order to be convinced of that it suffices to consider an example where the function $g$ in the boundary condition (8.1.3) is assumed to be strictly concave on one side of $\Omega$.

8.2.6.1 On estimation of the value $\rho(K_k, K)$

Let us consider in Problem (8.1.1) the feasible set (8.1.3), where $g$ is a trace of a function in the space $H^2(\Omega)$. We assume that for the special case $g = 0$ the inclusion $K_k \subset K$ holds. This always occurs if

$$\Omega \text{ is polyedral and } \bigcup_{T \in \mathcal{T}_k} \bar{T} = \bar{\Omega}, \quad \forall \, k.$$  

Sometimes this inclusion holds also for domains with a more complicated geometry, for instance, if curved triangles near the boundary of $\Omega$ and iso-parametric finite elements have to be used (for iso-parametric elements see Ciarlet [75]). Under the above assumption, for an arbitrarily chosen function $g_0 \in H^2(\Omega)$ and $g := \gamma g_0$ ($\gamma$ trace operator) the estimate

$$\rho(K_k, K) \leq c \|g_0\|_{2,\Omega} h_k$$

can be proved for corresponding sets $K_k$ and $K$ with $c$ independent of $g_0$ and $h_k$. Indeed, for each element $v_h \in K_h$ the following relation with respect to the interpolate of $g_0$ on the triangulation $\mathcal{T}_h$ holds true

$$v_h \geq (g_0)_{I,h} \quad \text{on } \Gamma,$$

hence,

$$v_h - (g_0)_{I,h} \in K - g_0.$$  

But, due to Theorem A2.2.5,

$$\|g_0 - (g_0)_{I,h}\|_{1,\Omega} \leq c \|g_0\|_{2,\Omega} h$$

is valid.

As mentioned before, in the case of two-body contact problems with a curved boundary the set $K_h$ has the form

$$K_h = \{v_h \in V_h : \langle \nu(\pi_k), v'_h(\pi_k) - v''_h(\pi_k) \rangle_{\mathbb{R}^2} \leq 0, \quad k \in I^c_h \},$$

with $\nu(x)$ a unit outward normal to $\Omega'$ in $x \in \Gamma_c$.

If $\nu$ is a trace on $\Gamma_c$ of a three-times differentiable vector-function defined on
one can prove that there exists a constant $h_0 > 0$ such that for $h \leq h_0$ and each $v_h \in K_h$ the function

$$\tilde{v}_h := (v'_h, v''_h + h^2^\frac{3}{4} \nu''_{I,h})$$

satisfies the inequality

$$\langle \nu(x), \tilde{v}_h'(\pi_k) - \tilde{v}_h''(\pi_k) \rangle_{\mathbb{R}^2} \leq 0, \quad \forall k \in I_h, \quad \forall x \in \Gamma_c.$$

This means that for $h_k \leq h_0$

$$\rho(K_h, K) \leq \tilde{c} h^2 h_k\frac{3}{4}.$$

8.2.17 Assumption.

(i) The optimal set $U^*$ of Problem (8.2.88) is non-empty and belongs to $[H^2(\Omega)]^m$, moreover, for each element $u \in U^*$ the estimate

$$\|u\|_{2,\Omega}^2 \leq c_2 \left( \|f\|_{0,\Omega}^2 + \|u\|_{V}^2 \right)$$

holds, with $c_2$ independent of $u$.

(ii) For every $\bar{f} \in [L_2(\Omega)]^m$ the function

$$\bar{u} := \arg \min_{u \in K} \left\{ \frac{1}{2} a(u, u) + \|u\|_{V}^2 - ((\bar{f}, u))_{0,\Omega} \right\}$$

belongs to $[H^2(\Omega)]^m$ and

$$\|\bar{u}\|_{2,\Omega}^2 \leq c_2 \left( \|\bar{f}\|_{0,\Omega}^2 + \|\bar{u}\|_{V}^2 \right).$$

♦
8.2. CONTACT PROBLEMS WITHOUT FRICTION

In order to solve Problem (8.2.88) we suppose that MSR-method 4.3.1 is applied, and the approximation is performed by means of finite elements with piece-wise affine basis functions on a quasi-uniform system of triangulations. Taking

\[ M \geq \sup_{\|u\|_V \neq 0} \frac{a(u,u)}{\|u\|_V^2}, \]

we choose the radii \( r \) and \( r^* \) such that

\[ r > 4\|f\|_{0,\Omega}, \quad r^* < \frac{4r}{33 + \sqrt{M + 34}}, \quad U^* \cap B_{r^*/8} \neq \emptyset, \quad (8.2.91) \]

with \( B_r := \{ u : \|u\|_V \leq r \} \). In the sequel the relation \( r^* < \frac{r}{8} \), being an obvious conclusion of the relation in (8.2.91), will be used.

Let \( K_k := K_{h_k} \), the choice of sequence \( \{h_k\} \) will be explained below. We make use of the sets \( Q, Q_k \) and \( Q^* \) according to the general framework of the investigation of MSR-methods described at the beginning of Section 4.3.2.

8.2.18 Assumption.

(i) For each \( k \) the interpolant \( u_{I,h_k} \) of an arbitrary function \( u \in K \) belongs to \( K_k \).

(ii) \( \rho(Q_k, Q) \leq c_1 h_k^\alpha \), \( k = 1, 2, ..., \) with fixed constants \( c_1 > 0 \) and \( \alpha > 0 \).

\[ \diamond \]

On account of Theorem A2.2.5, for a function \( v \in [H^2(\Omega)]^m \) we have

\[ \|v - v_{I,h_k}\|_V \leq c\|v\|_{2,0,h_k}, \quad (8.2.92) \]

and taking

\[ \mu_k := \mu(h_k) = \max \left[ c_1 h_k^\alpha, \frac{c\sqrt{2}}{\sqrt{2}} h_k \right], \quad k = 1, 2, ..., \]

\[ L(r) := Mr + \|f\|_{0,\Omega} \quad (\text{Lipschitz constant of } J \text{ on } B_r), \]

the following result can be obtained.

8.2.19 Theorem. Suppose that the Assumptions 8.2.17, 8.2.18 and condition (8.2.91) are fulfilled, moreover, let \( u^{1.0} \in B_{r/4} \). With \( \sigma_k := 0 \forall k \) let the controlling parameters \( \mu_k, \delta_k \) and \( \epsilon_k \) of MSR-method 4.3.1 be chosen according to (4.3.13) and (4.3.14).

Then it holds

\[ \rho(Q^*, Q_k) \leq \mu_k, \quad k = 1, 2, ... \]

Proof: Note that in the proof of Lemma 4.3.5 instead of the requirement \( \rho(Q^*, Q_k) \leq \mu_k \) only \( \rho(u^{**}, Q_k) \leq \mu_k \) was used with \( u^{**} \in U^* \cap B_{r^*/8} \). Now we are going to verify the latter estimate.

In view of Assumption 8.2.17(i) the inequality

\[ \|u^{**}\|_{2,\Omega} \leq \sqrt{c_2\|f\|_{0,\Omega}^2 + \left( \frac{r^*}{8} \right)^2} < \frac{\sqrt{c_2}}{2} \]
holds, and applying (8.2.92), one can conclude that

$$\|u^{**} - u_{I,h,k}^{**}\|_V < \frac{\sqrt{c_2}}{2} rh_k < \mu(h_k) =: \mu_k.$$ 

Relation $\mu_k < \frac{r^*}{4}$ follows from (4.3.14), consequently, $\|u_{I,h,k}^{**}\|_V < r$. But Assumption 8.2.18(i) ensures that $u_{I,h,k}^{**} \in Q_k$, hence,

$$u_{I,h,k}^{**} \in Q_k, \quad \rho(u^{**}, Q_k) < \mu_k.$$

Thus, in view of Lemma 4.3.5, the estimates

$$\|u^{k,i}\|_V < r^*, \quad \|\bar{u}^{k,i}\|_V < r^* \quad \forall (k,i)$$

holds true. Now, due to

$$J(u) + \|u - u^{k,i-1}\|_V^2 \leq \left(\frac{M}{2} + 1\right) \|u\|_V^2 + \|f\|_{0,\Omega} \|u\|_V + \|u^{k,i-1}\|_V^2 + 2\|u\|_V \|u^{k,i-1}\|_V,$$

and the facts that $u^{**} \in K \cap \mathbb{B}_{r^*/8}$ and $\|u^{k,i-1}\|_V < r^*$, we can conclude that

$$\min_{u \in Q} \{J(u) + \|u - u^{k,i-1}\|_V^2\} <$$

$$< \left(\frac{M}{2} + 1\right) \frac{(r^*)^2}{64} + \frac{r^*}{8} \|f\|_{0,\Omega} + (r^*)^2 + \frac{(r^*)^2}{4} \quad (8.2.93)$$

On the other hand, for $\|u\|_V \geq \frac{r^*}{2} > 4r^*$ the inequality

$$J(u) + \|u - u^{k,i-1}\|_V^2 \geq -\|f\|_{0,\Omega} \|u\|_V + \left(\|u\|_V - \|u^{k,i-1}\|_V\right)^2$$

$$\geq -\|f\|_{0,\Omega} \|u\|_V + (\|u\|_V - r^*)^2 \quad (8.2.94)$$

is satisfied. In order to estimate the right-hand part in (8.2.94), keeping in mind (8.2.91) and $\|u\|_V \geq \frac{r^*}{2}$, we have to minimize the function

$$\xi(t) := (t - r^*)^2 - \|f\|_{0,\Omega} t$$

subject to (8.2.91) and $t \geq \frac{r^*}{2}$.

Obviously, $\xi'(t) = 2(t - r^*) - \|f\|_{0,\Omega} > 0$ for $t \geq \frac{r^*}{2}$, hence, this minimum is attained at $t = \frac{r^*}{2}$.

Therefore, continuation of estimate (8.2.94) leads with $\|u\|_V \geq \frac{r^*}{2}$ to

$$J(u) + \|u - u^{k,i-1}\|_V^2 \geq \frac{r^2}{4} - rr^* - \frac{1}{2} \|f\|_{0,\Omega} r + (r^*)^2$$

$$\geq \frac{r^2}{8} - rr^* + (r^*)^2. \quad (8.2.95)$$

A comparison of the estimates (8.2.93) and (8.2.95) shows: If the choices of $r$ and $r^*$ obey the condition

$$\frac{r^2}{8} - rr^* + (r^*)^2 > \left(\frac{M}{128} + \frac{81}{64}\right) (r^*)^2 + \frac{rr^*}{32}, \quad (8.2.96)$$
8.2. CONTACT PROBLEMS WITHOUT FRICTION

then \( u^{k,i}_Q := \arg \min \{ J(u) + \| u - u^{k,i-1} \|^2_V : u \in Q \} \) belongs to \( \interior B_{r/2} \).

But inequality (8.2.96) is just obtained from

\[
r^* < \frac{4r}{33 + \sqrt{M + 34}}.
\]

Because \( u^{k,i}_Q \in \interior B_{r/2} \), we get \( u^{k,i}_Q = \arg \min \{ \Psi_{k,i}(u) : u \in K \} \).

In the sequel, using the expression

\[
\Psi_{k,i}(u) := \frac{1}{2} a(u,u) + \| u \|^2_V - ((2u^{k,i-1} + f,u))_{0,\Omega} + \| u^{k,i-1} \|^2_V,
\]

we can apply Assumption 8.2.17(ii) with \( \bar{f} := f + 2u^{k,i-1} \).

Hence, with regard to \( \| u^{k,i-1} \| < r^* \) and (8.2.90), we get

\[
\| \| u^{k,i}_Q \| \|_2,\Omega \leq \sqrt{c^2 \left( \| f \|_0,\Omega + 2r^* \right)^2 + \frac{r^2}{4}}
\]

and, due to (8.2.91),

\[
\| \| u^{k,i}_Q \| \|_2,\Omega \leq \sqrt{\frac{c^2}{2} r}.
\]

Now, (8.2.92) leads to

\[
\| u^{k,i}_Q - (u^{k,i}_{I,h})_{I,h} \|_V < c \sqrt{\frac{c^2}{2} r h_k} \leq \mu_k \quad (8.2.97)
\]

and taking into account that \( \mu_k < \frac{r}{4} \), one can conclude that

\[
\| (u^{k,i}_{I,h})_{I,h} \|_V < r.
\]

Together with Assumption 8.2.17(i) this ensures that \( (u^{k,i}_{Q})_{I,h} \in Q_k \). Furthermore, in view of (8.2.89) and (8.2.92), we verify analogously for functions \( u \in U \cap \mathbb{B}_r \) that

\[
\| u - u_{I,h} \|_V < \mu_k \quad (8.2.98)
\]

Because \( \mu_k < \frac{r}{4} \), the relations

\[
\| u_{I,h} \|_V < \frac{r}{4}, \quad u_{I,h} \in Q_k
\]

are true. Now, the inequalities (8.2.97) and (8.2.98) imply that

\[
\rho(Q^*, Q_k) \leq \mu_k, \quad \forall k.
\]

□

The same analysis can be performed for the two-body contact problem if, instead of \( U^* \subset [H^2(\Omega)]^m \), the inclusion \( U^* \subset [H^2(\Omega')]^2 \times [H^2(\Omega'')]^2 \) is used.

An analogous theorem is valid if the inequalities (8.2.89) and (8.2.90) are replaced by estimates of the type

\[
\| u \|^2_{2,\Omega} \leq c_2 \| f \|^2_{0,\Omega} + c_0,
\]

\[
\| \bar{u} \|^2_{2,\Omega} \leq c_2 \| f \|^2_{0,\Omega} + c_0.
\]
8.2.20 Remark. The usage of the radii \( r \) and \( r^* \) in the statements on convergence of OSR- and MSR-methods is connected first of all with the technique of estimating \( \rho(Q^*, Q_k) \) for variational inequalities. Usually, upper bounds for \( \rho(U^*, K_k) \) can be obtained by means of an estimation of the distance between \( u \in U^* \) and its interpolant.

However, the norm of such an interpolant in \( V \) may be larger than the norm of the interpolated function, and in case \( r = r^* \) we cannot guarantee that the interpolant of an arbitrary function \( u \in U^* \cap B_r \) belongs to \( Q_k \), even if \( K_k \subset K \).

A suitable harmonization of \( r^* \) and \( r \) ensures that for \( u \in U^* \cap B_r^* \) the inclusion \( u_{I,h_k} \in B_r \) holds for all \( k \) (cf. Theorem 8.2.19).

From the numerical point of view it is important to specify the choice of the parameters \( r, r^* \) and \( \{h_k\} \) for each particular problem. However, Theorem 8.2.19 gives only a qualitative characterization of the relations among these parameters.

Obviously, in the case of semi-infinite problems, such questions do not arise, because the choice \( r = r^* \) is the best one.

8.2.7 Numerical results

The numerical examples described in this Subsection have been investigated together with Voetmann [413].

8.2.21 Example. (Well-posed Signorini boundary obstacle problem)

We consider the two-dimensional Signorini boundary obstacle problem (A2.1.6) with homogeneous boundary values

\[
\begin{align*}
-\Delta u &= f \quad \text{in } \Omega \\
u &= 0 \quad \text{on } \Gamma_D \\
\frac{\partial u}{\partial \nu} &= 0 \quad \text{on } \Gamma_N \\
u \geq 0, \quad \frac{\partial u}{\partial \nu} \geq 0, \quad u \frac{\partial u}{\partial \nu} &= 0 \quad \text{on } \Gamma_S
\end{align*}
\]

with data

\[
\begin{align*}
\Omega &= (0,1) \times (0,1), \\
\Gamma_S &= \{(x,y) \in \Gamma : y = 1\}, \quad \Gamma_D = \Gamma \setminus \Gamma_S, \quad \Gamma_N = \emptyset, \\
f(x,y) &= 4\pi^2 \sin(2\pi y).
\end{align*}
\]

The domain \( \Omega \), the boundary conditions and the initial triangulation are shown in Figure 8.2.2. The triangulation is refined globally once at the start of each exterior step.
8.2. CONTACT PROBLEMS WITHOUT FRICTION

We note that the problem has a unique solution, since \( \text{meas}(\Gamma_D) > 0 \), although \( \langle f, 1 \rangle_{L^2(\Omega)} = 0 \).

We apply the MSR-method 4.3.1 with strong regularization, i.e. regularization in the space \( V \). Here the multi-step scheme is somewhat simpler performed, because the interior cycle is done by setting a maximum number of inner iterations \( i_{\text{max}} \). Inner steps are meant to be taken as long as a sufficiently large progress of the iterates indicates movement towards the solution on the given approximation level. They are stopped, if the iterations indicate that no more progress on the current approximation level can be expected. This motivated us to set

\[
\delta_{k,i} := \delta_k \|u^{k,0}\|, \quad \delta_k \downarrow 0,
\]

and the numerical results below show the effectiveness of even this simple approach.

The controlling parameters are chosen such that

\[
\chi_k := \chi_0 \cdot \chi^k, \quad \epsilon_k := \epsilon_0 \cdot \epsilon^k, \quad \delta_k := \delta_0 \cdot \delta^k,
\]

with \( \chi_0 := 1.0, \chi := 0.9, \epsilon_0 := 0.001, \epsilon := 0.5 \) and \( \delta_0 := 0.01, \delta := 0.8 \). The arising regularized finite dimensional auxiliary problems in each interior step are solved by a Newton method from the Matlab package.

Table 8.2.1 shows the iteration log of the MSR-method with strong regularization and Figure 8.2.3 shows the computed solution.

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \text{dim} )</th>
<th>#reg</th>
<th>#nwt</th>
<th>( \text{err} )</th>
<th>EOC</th>
<th>cpu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>5</td>
<td>3.4</td>
<td>1.24e-0</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>41</td>
<td>8</td>
<td>5</td>
<td>7.14e-1</td>
<td>0.81</td>
<td>0.10</td>
</tr>
<tr>
<td>3</td>
<td>145</td>
<td>8</td>
<td>5</td>
<td>3.82e-1</td>
<td>0.90</td>
<td>0.30</td>
</tr>
<tr>
<td>4</td>
<td>545</td>
<td>7</td>
<td>4</td>
<td>1.98e-1</td>
<td>0.95</td>
<td>1.09</td>
</tr>
<tr>
<td>5</td>
<td>2113</td>
<td>5</td>
<td>3</td>
<td>1.02e-1</td>
<td>0.95</td>
<td>4.91</td>
</tr>
<tr>
<td>6</td>
<td>8321</td>
<td>4</td>
<td>3</td>
<td>5.38e-2</td>
<td>0.92</td>
<td>33.26</td>
</tr>
<tr>
<td>7</td>
<td>33025</td>
<td>4</td>
<td>3</td>
<td>2.94e-2</td>
<td>0.87</td>
<td>216.27</td>
</tr>
</tbody>
</table>

Table 8.2.1: Iteration log of Example 8.2.21 with strong regularization
The first column shows the number of global refinements of the triangulation mesh; in the second it can be seen the number of resulting degrees of freedom, i.e. number of variables to be optimized. The next three columns indicate the number of interior steps taken with the respective regularization parameter (fixed on each refinement level), the number of Newton steps and the relative error distance of the computed iterate to the iterate on the preceding mesh interpolate after refining the mesh. The norm used to measure this error distance is the $H^1$-norm as this is the underlying space. Finally, an estimated order of convergence and cpu-time for the overall exterior step are given. The estimated order of convergence is based on a simple residual type error estimator and provided by the used finite element code ALBERT.

It is seen that the partition of the Signorini boundary into the parts with Dirichlet and Neumann boundary conditions is accurately traced.

As a comparison we now present in Table 8.2.2 and Figure 8.2.4 the results obtained with the same algorithm and same controlling parameters but using weak regularization, i.e. with $L_2$-norm.

<table>
<thead>
<tr>
<th>k</th>
<th>dim</th>
<th>#reg</th>
<th>#nwt</th>
<th>err</th>
<th>EOC</th>
<th>cpu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>2</td>
<td>3.5</td>
<td>4.45e-1</td>
<td>1.81</td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
<td>4</td>
<td>2</td>
<td>1.26e-1</td>
<td>1.81</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>145</td>
<td>2</td>
<td>4.5</td>
<td>3.37e-1</td>
<td>1.90</td>
<td>0.08</td>
</tr>
<tr>
<td>4</td>
<td>545</td>
<td>2</td>
<td>3</td>
<td>8.71e-3</td>
<td>1.95</td>
<td>0.26</td>
</tr>
<tr>
<td>5</td>
<td>2113</td>
<td>2</td>
<td>3</td>
<td>2.25e-3</td>
<td>1.95</td>
<td>1.87</td>
</tr>
<tr>
<td>6</td>
<td>8321</td>
<td>2</td>
<td>2</td>
<td>5.95e-4</td>
<td>1.92</td>
<td>13.04</td>
</tr>
<tr>
<td>7</td>
<td>33025</td>
<td>1</td>
<td>2</td>
<td>1.63e-4</td>
<td>1.87</td>
<td>51.03</td>
</tr>
</tbody>
</table>

Table 8.2.2: Iteration log of Example 8.2.21 with weak regularization
We see that the number of interior iterations is significantly decreased. This is naturally reflected in the cpu-time, although the number of Newton steps per interior iteration does not change notably. Our primary motivation for introducing the weak regularization is therefore satisfied. We also observe that as a price the boundary partition is not as cleanly traced as using strong regularization.

The multi-step scheme was motivated by the desire not to let the mesh be refined too fast. An alternative approach is to use a one-step scheme without refining the mesh globally in each iteration. Based on an error estimator provided by the finite element code ALBERT we present next the results of such an approach. After each exterior step the mesh is refined by the following simple maximum strategy.

**Mesh refinement algorithm** (Maximum strategy)

At each iteration $k$, given

- a triangulation $T^k$ and

- a local error estimate $\eta_T$ for all $T \in T^k$.

Choose a maximum threshold $\gamma \in (0, 1)$ and refine all elements $S \in T^k$ for which

$$\eta_S > \gamma \max_{T \in T^k} \eta_T.$$

Typically, a threshold $\gamma := 0.5$ is chosen.

Using this approach the resulting iterations and the solution are displayed in Table 8.2.3 as well in Figure 8.2.5 with the corresponding meshes depicted in Figures 8.2.6 and 8.2.7.

<table>
<thead>
<tr>
<th>$k$</th>
<th>dim</th>
<th>#reg</th>
<th>#nwt</th>
<th>err</th>
<th>EOC</th>
<th>cpu</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>179</td>
<td>1</td>
<td>3</td>
<td>3.01e-2</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>20</td>
<td>576</td>
<td>1</td>
<td>3</td>
<td>7.76e-3</td>
<td>1.95</td>
<td>0.37</td>
</tr>
<tr>
<td>30</td>
<td>1399</td>
<td>1</td>
<td>3</td>
<td>3.58e-3</td>
<td>1.12</td>
<td>0.37</td>
</tr>
<tr>
<td>40</td>
<td>2407</td>
<td>1</td>
<td>2</td>
<td>2.04e-3</td>
<td>0.81</td>
<td>0.97</td>
</tr>
<tr>
<td>50</td>
<td>5183</td>
<td>1</td>
<td>2</td>
<td>9.74e-4</td>
<td>1.07</td>
<td>3.16</td>
</tr>
</tbody>
</table>

Table 8.2.3: Iteration log of Example 8.2.21 with weak regularization and adaptive mesh refinement
Figure 8.2.5: Computed solution of Example 8.2.21 with weak regularization and adaptive mesh refinement

Figure 8.2.6: Adaptively refined meshes of Example 8.2.21: \( k = 10 \) and \( k = 20 \)

Figure 8.2.7: Adaptively refined meshes of Example 8.2.21: \( k = 30 \) and \( k = 40 \)

We see that the desired effect is taking place. In each iteration, only a small proportion of the mesh is refined. Due to the slowly increasing dimensions of the problems the Newton iterations are very cheap with respect to the cpu-time. Most important, the tracing of the partition is even sharper than by using the strong regularization.
8.2. CONTACT PROBLEMS WITHOUT FRICTION

8.2.22 Example. (Frictionless contact with a rigid foundation)

Now, let the boundary be divided into three parts \( \Gamma := \Gamma_D \cup \Gamma_F \cup \Gamma_C \), where \( \Gamma_D \neq \emptyset \) and \( \Gamma_C \neq \emptyset \) and the three sets are open and mutually disjoint. \( \Gamma_D \) is the boundary part with prescribed displacements and \( \Gamma_F \) is the boundary part where boundary forces apply. Additionally we consider the existence of a rigid foundation and a parametrization \( \alpha \) of the distance of any boundary point on \( \Gamma_C \) to the rigid foundation in \( x_2 \)-direction, i.e., \( \alpha : I \rightarrow \mathbb{R} \) is a smooth real function from a bounded interval \( I \subset \mathbb{R} \) such that for any \( x_1 \in I \) the point \((x_1, \alpha(x_1))\) is a point on the rigid foundation (see Figure 8.2.8).

Denote \( u(x_1,x_2) \) the displacement vector of a body-particle \( x = (x_1, x_2)^T \),

\[
\epsilon_{ij} := \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right), \quad i, j = 1, 2
\]

the linearized strain tensor,

\[
\tau_{ij} := a_{ijkl} \epsilon_{kl}, \quad i, j, k, l = 1, 2
\]

the linearized stress tensor with \( a_{ijkl} \) the elasticity coefficients. Moreover, denote \( T_\nu = \tau(u) \cdot \nu \) the stress of the displacements in direction of the normal vector \( \nu \), \( T_t = \tau(u) \cdot t \) the stress of the displacements in tangential direction, respectively. Then the following simple one-sided contact problem can be imposed:

\[
-\text{div} \tau(u) = f \quad \text{in} \ \Omega, \quad (8.2.100)
\]

\[
uu = d \quad \text{on} \ \Gamma_D, \quad (8.2.101)
\]

\[
\tau(u) \cdot \nu = g \quad \text{on} \ \Gamma_F, \quad (8.2.102)
\]

\[
uu \leq \alpha \quad \text{on} \ \Gamma_C, \quad (8.2.103)
\]

\[
T_\nu \leq 0 \quad \text{on} \ \Gamma_C, \quad (8.2.104)
\]

\[
(u_\nu - \alpha)T_\nu = 0 \quad \text{on} \ \Gamma_C, \quad (8.2.105)
\]

\[
T_t = 0 \quad \text{on} \ \Gamma_C. \quad (8.2.106)
\]
The first three equations are equilibrium conditions, in particular, equation (8.2.100) describes the equilibrium of the stress of the displacements inside the body, (8.2.101) requires an equilibrium of the displacements on the Dirichlet boundary, whereas (8.2.102) is due to the equilibrium of the stress of displacements in direction of the normal vector on \( \Gamma_F \). Condition (8.2.103) prevents penetration into the rigid foundation, conditions (8.2.104) says that the stress of the displacements on the contact zone is acting towards the body and (8.2.105) makes sure that if no contact then there is no stress of the displacements on \( \Gamma_C \). The last condition says that no friction appears in this simplified model.

The weak formulation of this model is described by

\[
K := \{ u \in [H^1(\Omega)]^2 : \gamma u|_{\Gamma_D} = d, \, \gamma u|_{\Gamma_C} \leq \alpha \},
\]

and

\[
a(u, v) := \int_{\Omega} a_{ijkl} \epsilon_{kl} d\Omega,
\]

\[
\ell(u) := \int_{\Omega} f_i u_i d\Omega + \int_{\Gamma_F} g_i u_i d\Gamma.
\]

The energy functional \( J(u) = \frac{1}{2} a(u, u) - \ell(u) \) which is to minimize is not coercive on \( V \), but the second Korn inequality allows us to employ weak regularization.

In the sequel we show two simple examples of such an one-sided contact problem which differ only in the elasticity of the material, expressed by the corresponding Yung’s modulus.

**Deflection of a steel bar:** (cf. Figure 8.2.9)

Data: Yung’s modulus \( E := 2.15 \cdot 10^{11} \), Poisson’s coefficient \( \pi := 0.29 \) and \( g_2(x) := -5.75 \cdot 10^8 \, \text{N} \cdot \text{m}^{-2} \).

![Steel bar fastened on the left side of the rigid body](image)

In Figure 8.2.10 and Table 8.2.4 the graph of displacements and the iterations are fixed.
8.2. CONTACT PROBLEMS WITHOUT FRICTION

Figure 8.2.10: Displacements of the steel bar

<table>
<thead>
<tr>
<th>k</th>
<th>dim</th>
<th>#reg</th>
<th>#nwt</th>
<th>cpu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54</td>
<td>3</td>
<td>6.6</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>170</td>
<td>1</td>
<td>8</td>
<td>0.21</td>
</tr>
<tr>
<td>3</td>
<td>594</td>
<td>1</td>
<td>6</td>
<td>1.01</td>
</tr>
<tr>
<td>4</td>
<td>2210</td>
<td>1</td>
<td>5</td>
<td>10.81</td>
</tr>
<tr>
<td>5</td>
<td>8514</td>
<td>1</td>
<td>5</td>
<td>75.28</td>
</tr>
<tr>
<td>6</td>
<td>33410</td>
<td>1</td>
<td>5</td>
<td>332.52</td>
</tr>
</tbody>
</table>

Table 8.2.4: Iteration log: deflection of a steel bar

Deflection of an aluminium bar: (cf. Figure 8.2.11)

Data: Young’s modulus $E := 2.15$, Poisson’s coefficient $\pi := 0.29$ and $g_2(x) := -5.75 \cdot 10^8 \text{Nm}^{-2}$.

Figure 8.2.11: Aluminium bar fastened on the left side of the rigid body
In Figure 8.2.12 the graph of displacements is depicted and Table 8.2.5 gives some information about the iterations.

<table>
<thead>
<tr>
<th>k</th>
<th>dim</th>
<th>#reg</th>
<th>#nwt</th>
<th>cpu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54</td>
<td>10</td>
<td>3</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>170</td>
<td>1</td>
<td>3</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>594</td>
<td>1</td>
<td>2</td>
<td>0.23</td>
</tr>
<tr>
<td>4</td>
<td>2210</td>
<td>1</td>
<td>2</td>
<td>1.75</td>
</tr>
<tr>
<td>5</td>
<td>8514</td>
<td>1</td>
<td>2</td>
<td>14.26</td>
</tr>
</tbody>
</table>

Table 8.2.5: Iteration log: deflection of an aluminium bar

8.3 Solution of Contact Problems With Friction

Our study of this class of problems is based on an algorithm of alternating iterations, suggested by Panagiotopoulos [315]. This algorithm, dealing with the alternate solution of variational inequalities of two types, does not have a strong theoretical foundation in the case of non-coercive operators. However, it is common use in practice.

To make its description simple, we analyze this algorithm only for the Signorini problem with friction. Its formal extension to the two-body contact problem is possible without difficulties.

Concerning solvability of contact problems with friction only a result of Nečas, Jarušek and Haslinger [308] is known for the Signorini problem in case the friction coefficient is small and the body is fastened on a part of its boundary.

The main idea of the algorithm is as follows: On each iteration two convex auxiliary problems have to be solved, one of which is an analogue of the classical Signorini problem (without friction) and the other one consists in the unconstrained minimization of a non-smooth, convex and continuous functional.

In the sequel we construct by means of the PPR stable solution procedures for these auxiliary problems, which are ill-posed in general.
8.3. Solution of Contact Problems with Friction

8.3.1 Signorini problem with friction

Preserving the notations and assumptions made previously in Subsection 8.2.1 for the classical Signorini problem, we take into account the friction arising on the contact boundary.

Let the friction coefficient $\rho$ belong to $C^2(\Gamma_c)$ and denote by $T_{\nu}$ the normal component of the reaction of the boundary $\Gamma_c$ (normal force). Considering $T_{\nu} \leq 0$ as a fixed parameter, the following variational problem can be formulated

$$\min_{u \in K} \left\{ \frac{1}{2} a(u, u) - \int_\Omega F_k u_k d\Omega - \int_{\Gamma_r} P_k u_k d\Gamma + j(u; T_{\nu}) \right\}, \quad (8.3.1)$$

where

$$K := \{ v \in [H^1(\Omega)]^2 : v_\nu = 0 \text{ on } \Gamma_0, \; v_\nu \leq 0 \text{ on } \Gamma_c \}, \quad (8.3.2)$$

with $j(u; T_{\nu}) := \int_{\Gamma_c} \rho |T_{\nu}| |u_t| d\Gamma$, and $u_t$ denotes the tangential component of the displacement vector $u$.

For fixed $T_{\nu}$ Problem (8.3.1) is called Signorini problem with friction under given normal forces.

In realistic situations it is not natural that the normal forces on the contact boundary are known. The equilibrium state answers to the relation

$$T_{\nu} = \tau_{kl}(u(T_{\nu})) v_{kl}, \quad (8.3.3)$$

understanding in a generalized sense (see Duvaut and Lions [95], chapt. 3). Here $u(T_{\nu})$ describes the displacement of the points on $\Gamma_c$ according to Problem (8.3.1) and $\tau_{kl}$ are coefficients of the corresponding stress tensor. Problem (8.3.1)–(8.3.3) is called Signorini problem with friction. Apparently the question about its unique solvability is still open.

8.3.2 Algorithm of alternating iterations

On each step of this method the following two problems have to be solved one after the other:

**Problem I:** Given tangential forces $T_t$ on the boundary $\Gamma_c$, minimize

$$\frac{1}{2} a(u, u) - \ell_t(u),$$

subject to $K$, defined by (8.3.2), where

$$\ell_t(u) := \int_\Omega F_k u_k d\Omega + \int_{\Gamma_r} P_k u_k d\Gamma + \int_{\Gamma_c} T_t u_t d\Gamma.$$

**Problem II:** Given normal forces $T_{\nu}$ on the boundary $\Gamma_c$, minimize

$$\frac{1}{2} a(u, u) - \ell_{\nu}(u) + j(u; T_{\nu}),$$

subject to $V := \{ u \in [H^1(\Omega)]^2 : u_\nu = 0 \text{ on } \Gamma_0 \}$, where

$$\ell_{\nu}(u) := \int_\Omega F_k u_k d\Omega + \int_{\Gamma_r} P_k u_k d\Gamma + \int_{\Gamma_c} T_{\nu} u_{\nu} d\Gamma.$$
Starting, for instance, with $T^0_t := 0$, at the beginning of the $i$-th step an approximate value $T^{i-1}_t$ of the tangential forces has to be known. Then, at the $i$-th step, initially Problem I has to be solved with $T^i_t := T^{i-1}_t$. This results in a vector $u^{i-1/2}$, characterizing the corresponding displacement. Now, we are in the position to calculate an approximation of the normal forces via (8.3.3) by

$$T^i_\nu := T_\nu(u^{i-1/2}) = \tau_{kl}(u^{i-1/2}) \nu_l \nu_k.$$ 

Thereafter, solving Problem II subject to the normal forces $T^i_\nu$, we find the vector $u^i$ and compute the tangential forces:

$$(T^i_t)_k = \tau_{kl}(u^i) \nu_l - (\tau_{kl}(u^i) \nu_l \nu_k) \nu_k, \quad k = 1, 2.$$ 

Symbol $T_t$ is used for the vector $T_t$ as well as for the value of its tangential component. The iteration procedure terminates if the distances between the normal forces as well as between the tangential forces in two successive iterations become small.

In the following we analyze the Problems I and II under the assumptions that

$$P_k \in L_\infty(\Gamma_\tau), \quad k = 1, 2; \quad T_t \in L_\infty(\Gamma_c), \quad T_\nu \in L_\infty(\Gamma_c),$$

maintaining the other suppositions made for Problem (8.2.16).

Solvability of Problem I can be concluded from Theorem 8.2.3, by replacing the linear functional $\ell$ in (8.2.16) by $\ell_t$ (cf. [182]). This result says: If Problem I is solvable, then there exists no unique solution in general.

Concerning Problem II, in [95], chapt. 3, the following result is established.

8.3.1 Theorem. Denote $\mathcal{K}$ the kernel of the bilinear form (8.2.17). For solvability of Problem II it is necessary that

$$|\ell_\nu(y)| \leq j(y; T_\nu), \quad \text{for any } y \in \mathcal{K} \cap V.$$ 

A solution exists if for any non-zero vector $y \in \mathcal{K} \cap V$

$$|\ell_\nu(y)| < j(y; T_\nu).$$

Uniqueness of the solution of Problem II is not known in the general case. Nevertheless, it is obvious that if there exists more than one solution, then their differences belong to $\mathcal{K}$.

In the sequel we assume that the algorithm of alternating iterations is well-defined, i.e., at each step the arising auxiliary problems are solvable. We also suppose that the functions $T^i_\nu$ and $T^i_t$ are necessarily smooth.

We will leave open the questions about verification of these assumptions and the study of the smoothness of solutions of these auxiliary problems as well, because our main concern is to construct a stable solution process for these problems.

Problem I can be solved by any of the three MSR-methods described in Section 8.2.3, and the changing rule for the controlling parameters is completely analogous to those guaranteeing convergence of the corresponding method for
the classical Signorini problem. In particular, if the discretization secures the inclusion $K_k \subset K$, then the controlling parameters are determined by Theorem 8.2.14 and Remark 8.2.16, but in more general situations the results of Theorem 8.2.19 can be used.

However, in the framework of the algorithm of alternating iterates these rules are not strictly substantiated. Nevertheless, they ensure convergence of the iterates to some solution of Problem I and Problem II, respectively, in case the solutions of these problems are sufficiently smooth, say they belong to $H^2(\Omega)$.

Now we consider the numerical treatment of Problem II. Difficulties in minimizing the functional

$$ J(u) := \frac{1}{2} a(u, u) - \ell_\nu(u) + j(u; T_\nu) $$

(8.3.4)

on the space $V$ under fixed $T_\nu$ are related in the first line with the non-differentiability of $j(u; T_\nu)$. In order to use one of the MSR-methods mentioned above for solving this problem, a smoothing procedure for the objective functional (8.3.4) has to be included in this method. Therefore, we shall approximate $j(u; T_\nu)$ by a sequence of functionals

$$ j_k(u; T_\nu) = \int_{\Gamma_\nu} \rho |T_\nu| \sqrt{|u|^2 + \tau_k} d\Gamma, \quad \tau_k \downarrow 0. $$

(8.3.5)

In the sequel our consideration is concentrated on the method with weak regularization, which generates the following sequence of auxiliary problems:

$$ \min_{u \in V_k} \{ J_k(u) + \|u - u^{k,i-1}\|_{0,\Omega}^2, \quad i = 1, \ldots, i(k); \quad k = 1, 2, \ldots \} $$

(8.3.6)

with

$$ J_k(u) := \frac{1}{2} a(u, u) - \ell_\nu(u) + j_k(u; T_\nu), $$

and $V_k := V_h_k$ the finite element space of $V$, obtained by means of the triangulation $T_h_k$. Thus, in Method 8.2.5 we take

$$ \Psi_{k,i}(u) := J_k(u) + \|u - u^{k,i-1}\|_{0,\Omega}^2. $$

Obviously, the following estimate holds:

$$ J_k(u) - J(u) = \int_{\Gamma_\nu} \rho |T_\nu| \left( \sqrt{|u|^2 + \tau_k} - |u| \right) d\Gamma $$

$$ = \int_{\Gamma_\nu} \sqrt{|u|^2 + \tau_k + |u|} d\Gamma $$

$$ \leq c_3 \sqrt{\tau_k}, \quad c_3 := \int_{\Gamma_\nu} \rho |T_\nu| d\Gamma. $$

Hence, for arbitrary radius $r$ condition (4.2.5) in Assumption 4.2.2 is fulfilled with $\sigma_k := c_3 \sqrt{\tau_k}$.

8.3.2 Remark. If instead of $\|u - u^{k,i-1}\|_{0,\Omega}^2$ the regularization term $\|u - u^{k,i-1}\|_V^2$ is used, then convergence follows immediately from the Theorems 4.3.6 and 4.3.8. However, numerical experiments show that the method with weak regularization is more successful for such problems (see [365]).
Now we study convergence of the method with weak regularization applied to Problem II. This can be done similarly to the proofs of Lemma 8.2.13 and Theorem 8.2.14. It should be noted that in the given situation the relations $K_k = V_k$ and $K = V$ are true, i.e. inclusion $V_k \subset V$ means that $K_k \subset K$.

Assume that the solution set $U^*$ of Problem II is non-empty and belongs to the space $[H^2(\Omega)]^2$. On the space $V$ we introduce the new norm (cf. with (8.2.36))

$$|u|_\Omega^2 := \frac{1}{2} b(u, u) + \|u\|_{\delta, \Omega}^2,$$

with $b(u, v) := c_0 \int_\Omega \epsilon_{kl}(u) \epsilon_{kl}(v) d\Omega$ and $c_0 > 0$ chosen according to (8.2.3).

The space equipped with this norm will be denoted by $V$. Equivalence between the norms $\|u\|_{1, \Omega}$ and $|u|_\Omega$ can be established in the same way as it was done for the two-body contact problem in Section 8.2.3.

In the sequel we consider Problem II in the space $V$. If $\bar{u}$ and $\bar{u}$ are two solutions of this problem, then relation (8.2.83) holds true. Hence, for some fixed $u^{**} \in U^*$ and any $u \in U^* \cap B_{r_1}(u^{**})$, the estimate

$$\|u\|_{2, \Omega} \leq \|u^{**}\|_{2, \Omega} + c_4 r_1,$$

is satisfied, with $c_4$ independent of $r_1$ and $B_{r_1}(u^{**}) \subset V$.

We make use of the same finite element approximation as in the case of the classical Signorini problem and assume that $h_k \geq h_{k+1}$ $\forall k$. Due to the known estimate

$$\|v - v_{I, h}\|_{1, \Omega} \leq c_5 \|v\|_{2, \Omega}.$$

Equivalence between $\|v\|_{1, \Omega}$ and $|v|_\Omega$ yields

$$|v - v_{I, h}|_\Omega \leq \bar{c}\|v\|_{2, \Omega} h.$$ (8.3.7)

Choosing

$$M \geq \sup_{u \neq 0} \frac{\alpha(u, u)}{|u|_\Omega^2},$$

and denoting by $M_1(T_\nu)$ the Lipschitz constant of the functional

$$-\ell_\nu(u) + \int_{T_\nu} \rho|T_\nu| u_t d\Gamma$$

on $V$, then

$$L(r_1) := M (|u^{**}|_\Omega + r_1) + M_1(T_\nu)$$

is the Lipschitz constant of the functional (8.3.4) on $B_{r_1}(u^{**})$.

Due to (8.3.7),

$$|u^{**} - u_{I, h_k}^{**}|_\Omega \leq \bar{c}\|u^{**}\|_{2, \Omega} h.$$ (8.3.8)

Now we take

$$\mu_k := c_5 h_k, \quad \bar{\mu}_k := c_6 h_k, \quad k = 1, 2, \ldots,$$

$$r_0 := c_0 h_1, \quad L_0 := M (|u^{**}|_\Omega + r_0) + M_1(T_\nu),$$

where $c_5 \geq \bar{c}\|u^{**}\|_{2, \Omega}$ and $c_6 \geq \bar{c}(\|u^{**}\|_{2, \Omega} + c_4 r_*)$ are fixed, and the choice of $r_*$ will be specified below.
8.3.3 Theorem. Suppose that
\[ \sum_{k=1}^{\infty} \sqrt{\mu_k} < \infty, \quad \sum_{k=1}^{\infty} \sqrt{\sigma_k} < \infty, \quad \sum_{k=1}^{\infty} \epsilon_k < \infty, \]
and assume further that \( r^* \) and \( r \) satisfy the inequalities
\[ r^* > \max \left[ 8r_0, \sum_{k=1}^{\infty} \left( \sqrt{L_0 \mu_k + 2\sigma_k + \epsilon_k^2 + 2\mu_k} \right) \right], \quad (8.3.9) \]
\[ r \geq r^* + \bar{c} \left( \|u^{**}\|_{2,\Omega} + cr^* \right) h_1, \quad (8.3.10) \]
and the parameters \( \{ \delta_k \} \) are chosen such that
\[ \frac{1}{4r} \left( L(r)\bar{\mu}_k + 2\sigma_k - (\delta_k + \frac{\epsilon_k}{2})^2 \right) + \frac{\epsilon_k}{2} < 0, \quad k = 1, 2, \ldots \]
Then, starting with \( u^{1,0} \in B_{r^*/4}(u^{**}) \), the method with weak regularization, with auxiliary problems via (8.3.6), converges to some solution of Problem II.

Using (8.3.8) together with \( \sigma_k := c_3 \sqrt{\tau_k} \), one can reformulate this statement in terms of the original controlling parameters \( h_k, \tau_k, \epsilon_k \) and \( \delta_k \).

Comments on the proof: Finiteness of \( k(i) \) and the validity of the inclusions \( u^{k,i} \in \text{int}B_{r^*}(u^{**}), \bar{u}^{k,i} \in \text{int}B_{r^*}(u^{**}) \) follow from the proof of Lemma 8.2.65 if we use, instead of (8.2.67) the estimate
\[ J_k(v^k) \leq J_k(\bar{u}^{k,i}) + L_0 \mu_k + 2\sigma_k \]
and thereafter we apply Lemma 8.2.7 with \( J^1 := J_k, \quad C := K_k, \quad z^0 := u^{k,i-1}, \quad u := v^k \).

With analogous modifications in the proof of Theorem 8.2.14 we obtain, instead of (8.2.79), the inequality
\[ |u^{k,i(k)}(\cdot) - v^k|_{\Omega} \leq |u^{k,i(k)-1} - v^k|_{\Omega} + \sqrt{L(r)\bar{\mu}_k + 2\sigma_k + \frac{\epsilon_k}{2}}, \]
which should be used further together with (8.2.77) in order to prove convergence of the sequence \( \{u^{k,0} - w|_{\Omega}\} \) for each \( w \in U^* \cap B_{r^*}(u^{**}) \). The remaining modifications are obvious. \( \square \)

We emphasize that the successive approximation of the objective functional \( J \) in (8.3.4) by means of a family of convex functionals \( J_k \) is an eminent part of the general MSR-scheme. Here this approximation is carried out by replacing a non-smooth problem by a sequence of smooth problems, whereas in Section 6.2 this possibility was used to construct path-following strategies for solving parametric SIP.

8.3.3 Numerical results
Let us consider the following simple model with boundary friction:
8.3.4 Example. (Simplified Signorini problem with boundary friction)

\[ u \in K : \quad a(u, v - u) + j(v) - j(u) \leq \langle \ell, v - u \rangle \quad \forall \ v \in K, \]

where \( \Gamma = \partial \Omega = \Gamma_D + \Gamma_F \), \( f : \Omega \to \mathbb{R} \) and

\[ K := \{ u \in H^1(\Omega) : \gamma(u) = 0 \text{ on } \Gamma_D \}, \]

with

\[ a(u, v) := \mu \int_\Omega uv dx + \int_\Omega \nabla u \cdot \nabla v dx, \]
\[ \langle \ell, u \rangle := \int_\Omega f u dx, \]
\[ j(u) := \rho \int_{\Gamma_F} |u| d\Gamma. \]

The geometrical configuration, data and solution for Example 8.3.4 with friction coefficient \( \rho := 1.0 \) can be seen in Figure 8.3.13. One should notice that the irregularity in the displacements along \( \Gamma_F \) is due to the influence of the friction on the one-dimensional friction domain \( \Gamma_F \).

![Figure 8.3.13: Simplified Signorini problem with boundary friction (\( \rho := 1.0 \))]
### Table 8.3.7: Iteration log: strong regularization

<table>
<thead>
<tr>
<th>dim</th>
<th>#reg</th>
<th>err</th>
<th>cpu</th>
</tr>
</thead>
<tbody>
<tr>
<td>386</td>
<td>10</td>
<td>1.284012e+00</td>
<td>5.52e+01</td>
</tr>
<tr>
<td>1545</td>
<td>4</td>
<td>9.199661e-02</td>
<td>1.74e+01</td>
</tr>
<tr>
<td>6179</td>
<td>3</td>
<td>4.841953e-02</td>
<td>7.61e+01</td>
</tr>
<tr>
<td>24711</td>
<td>3</td>
<td>2.464253e-02</td>
<td>9.58e+02</td>
</tr>
<tr>
<td>98831</td>
<td>3</td>
<td>1.246926e-02</td>
<td>1.27e+04</td>
</tr>
</tbody>
</table>

8.4 Regularization of Well-posed Problems

Let us return to Problem (8.1.1) and assume that $\langle f, 1 \rangle_{0, \Omega} < 0$. Then, as mentioned in (8.1.4), the problem has a unique solution $u^*$. If for each function $u \in H^1(\Omega)$ the Poincaré inequality

$$\|\nabla u\|_{0, \Omega}^2 + \left(\int_{\Omega} ud\Omega\right)^2 \geq A \|u\|_{1, \Omega}^2 \quad (A > 0 \text{ independent of } u)$$

holds on the domain $\Omega$, then each generalized minimizing sequence converges in the norm of the space $H^1(\Omega)$ to $u^*$ (cf. Kaplan and Namm [207]). Thus, regularization is first of all introduced in order to reduce the influence of errors arising in the construction of discrete approximations of the problems (especially if the boundary is curved) and also to guarantee a stable solution process of the approximate problems as well as a more efficient application of fast algorithms. For the case that a quasi-uniform triangulation sequence can be constructed.
with
\[ K_k \subset K_{k+1} \subset K, \]
which is possible, for instance, if \( \Omega \) is a polyhedral set and the function \( g \) in (8.1.3) is convex on each face of \( \Omega \), an OSR-method is considered in [207], where \( \{ u^k \} \) is generated by
\[ u^k \in K_k, \quad \| u^k - \bar{u}^k \|_{1, \Omega} \leq \epsilon_k, \]
with \( \bar{u}^k := \arg \min_{u \in K_k} \{ J(u) + \| u - u^{k-1} \|_{0, \Omega}^2 \} \).
If the solution of the original problem belongs to the space \( H^2(\Omega) \), then convergence of \( \{ u^k \} \) to \( u^* \) in the norm of \( H^1(\Omega) \) is shown under the conditions that
\[ \epsilon_k \leq \sqrt{h_k^2}, \quad \sum_{k=1}^{\infty} \sqrt{h_k^3} < \infty, \]
which are in general weaker than those results from Theorem 8.2.14 for OSR-methods. Assumption \( u \in H^2(\Omega) \) seems to be natural, taking into account the results of LIONS [271] and BREZIS [51] for the similar problem with
\[ J(u) := \frac{1}{2} \| u \|_{2, \Omega}^2 - \langle f, 1 \rangle_{0, \Omega} \]
instead of (8.1.2).
These conditions for \( \epsilon_k \) and \( h_k \) are not surprising, because the problem is well-posed with respect to small perturbations of both, the objective function in the norm of the space \( C(V) \) and the feasible set in the metric of \( H^1(\Omega) \).
Numerical experiments for solving the regularized auxiliary problems with relaxation methods show a good stability of the solution process with respect to discretization errors (cf. Namm [300]).

8.4.1 Linear rate of convergence of MSR-methods
Now we concentrate our investigation on the verification of the Assumptions 5.1.4 and 5.1.6, which guarantee a linear rate of convergence of MSR-methods by changing the controlling parameters in the framework of the Theorems 5.1.5, 5.1.7 and Corollary 8.2.9.
We assume that finite element approximation of the considered variational inequality secures the inclusions \( K_k \subset K \) and that the chosen method generates iterates \( u^{k,i} \in K \). Then the Assumptions 5.1.4 and 5.1.6 can be transformed into a formally simpler condition if we take \( c := 0 \) in (5.1.29).
For Problem (8.1.1) and a number of other variational problems, where the kernel of the bilinear form on the space \( V \) is one-dimensional, the validity of the modified assumption follows from a result in [301] which we describe now.
We consider the problem
\[ \min_{u \in K} \left\{ \frac{1}{2} a(u, u) - \ell(u) \right\}, \]
with \( K \) a convex, closed subset of a Hilbert space \( V \), \( a(\cdot, \cdot) \) a symmetric, continuous bilinear form on \( V \times V \) and \( \ell : V \to \mathbb{R} \) a bounded linear functional.

8.4.1 Assumption.
(i) Kernel $\mathfrak{K}$ of the bilinear form $a(\cdot, \cdot)$ on $V$ is one-dimensional; 
(ii) $a(u, u) \geq c_0 \|\Pi u\|^2$, $\forall u \in V$, ($c_0 > 0$), 
\Pi an ortho-projector on the orthogonal complement $\mathfrak{K}^\perp$ of the kernel $\mathfrak{K}$; 
(iii) $\ell(z) \neq 0$, $\forall z \in \mathfrak{K}$, $z \neq 0$; 
(iv) Problem (8.4.1) is solvable. \hfill $\Diamond$

This assumption ensures that the solution $u^*$ of Problem (8.4.1) is unique. We show now that for Problem (8.1.1) the Assumption 8.4.1(ii) is fulfilled under the validity of the Poincaré inequality.

Indeed, for each $u \in V := H^1(\Omega)$ the element $v := \frac{1}{\text{meas} \Omega} \int_\Omega u \, d\Omega$ belongs to $\mathfrak{K}$ and for $u^1 := u - v$ one gets

$$
a(u, u) = (\nabla u, \nabla u)_{0, \Omega} = ((\nabla u^1 + v, \nabla (u^1 + v)))_{0, \Omega}$$

$$= (\nabla u^1, \nabla u^1)_{0, \Omega} = ((\nabla u^1, \nabla u^1))_{0, \Omega} + \left( \int_\Omega u^1 \, d\Omega \right)^2$$

$$\geq A\|u^1\|_{1, \Omega}^2 \geq A\|\Pi u\|^2_{1, \Omega} = A\|u\|^2_{1, \Omega}.$$ 

Using the analysis given in [307] for problems in elasticity theory, we can also state that Assumption 8.4.1 holds true for two-body contact problems.

**8.4.2 Theorem.** Let Assumption 8.4.1 be fulfilled for Problem (8.4.1). Then there exists a constant $c_1 > 0$ such that

$$J(u) - J(u^*) \geq c_1 \min \{ \|u - u^*\|, \|u - u^*\|^2 \}, \forall u \in K,$$

with $J(u) := \frac{1}{2} a(u, u) - \ell(u)$.

**Proof:** For arbitrarily chosen $u \in V$ the following relation is obvious

$$J(u) - J(u^*) = a(u^*, u - u^*) - \ell(u - u^*) + \frac{1}{2} a(u - u^*, u - u^*).$$

We decompose the vector $(u - u^*)$ into the direct sum

$$u - u^* = \lambda\|u - u^*\|v' + \mu\|u - u^*\|v'', \quad \lambda^2 + \mu^2 = 1,$$

where $v' \in \mathfrak{K}$ and $v'' \in \mathfrak{K}^\perp$ are unit vectors, with $\ell(v') < 0$. Then

$$J(u) - J(u^*) = a(u^*, u - u^*) - \ell(\lambda\|u - u^*\|v' + \mu\|u - u^*\|v'')$$

$$+ \frac{1}{2} a(\mu\|u - u^*\|v', \mu\|u - u^*\|v'')$$

$$\geq \|u - u^*\| (-\lambda\ell(v') - \mu\ell(v'') + \mu a(u^*, v''))$$

$$+ \frac{\mu^2}{2}\|u - u^*\|^2 a(v'', v''). \quad (8.4.2)$$

The variational inequality

$$a(u^*, u - u^*) - \ell(u - u^*) \geq 0 \quad \forall u \in K,$$

which corresponds to Problem (8.4.1), can be described as

$$-\lambda\ell(v') - \mu (a(u^*, v'') - \ell(v'')) \geq 0. \quad (8.4.3)$$
If $\lambda < 0$, then we obtain
\[
-\sqrt{1 - \mu^2} \lambda (v') + \mu (a(u^*, v'') - \ell(v'')) \geq 0
\] (8.4.4)
and because of $\ell(-v') > 0$, relation (8.4.4) leads to
\[
\mu^2 \geq \frac{(\ell(-v'))^2}{(a(u^*, v'') - \ell(v''))^2 + (\ell(-v'))^2}.
\]
Choosing
\[
M \geq \sup_{u \neq 0} \frac{a(u, u)}{\|u\|^2}, \quad \ell_0 := \|\ell\|_{V'},
\]
the inequality
\[
\mu^2 \geq \frac{(\ell(-v'))^2}{(M\|u^*\| + \ell_0)^2 + (\ell(-v'))^2} =: c_2
\] (8.4.5)
is true. Assumption 8.4.1(ii) together with (8.4.2), (8.4.3) and (8.4.5) yield
\[
J(u) - J(u^*) \geq \frac{c_2 c_0}{2} \|u - u^*\|^2.
\] (8.4.6)
If $\lambda \geq 0$, then we have to face two possibilities:
\[
|\mu| > c_3 > 0 \quad \text{and} \quad |\mu| \leq c_3,
\]
with
\[
c_3 := \frac{-c_4 c_5 + c_5 \sqrt{4c_4^2 + 3c_5^2}}{2(c_4^2 + c_5^2)} < 1
\]
and $c_4 := M\|u^*\| + \ell_0$, $c_5 := \ell(-v')$.
In the first case the estimate
\[
J(u) - J(u^*) \geq \frac{c_5 c_0}{2} \|u - u^*\|^2
\] (8.4.7)
follows immediately from (8.4.2), (8.4.3) and Assumption 8.4.1(ii).
In the second case the inequality
\[
\sqrt{1 - \mu^2} c_5 - |\mu| c_4 \geq \frac{c_5}{2}
\]
can be easily verified. Thus
\[
\lambda \ell(-v') + \mu (a(u^*, v'') - \ell(v'')) \geq \sqrt{1 - \mu^2} c_5 - |\mu| c_4 \geq \frac{c_5}{2}
\]
holds true and, due to (8.4.2), we obtain
\[
J(u) - J(u^*) \geq \frac{c_5}{2} \|u - u^*\|.
\] (8.4.8)
Combining the inequalities (8.4.6), (8.4.7) and (8.4.8), we get the statement of Theorem 8.4.2 with
\[
c_1 := \min \left[ \frac{c_0 c_2}{2}, \frac{c_0 c_3^2}{2}, \frac{c_5}{2} \right].
\]
Now, setting $\theta := \frac{c_1}{2}$, the validity of Assumption \ref{assumption:5.1.4} with $c := 0$ follows immediately from this statement.

It seems to be important to study proximal point methods for elliptic variational inequalities with non-linear operators, in particular, for the generalized Problem (A2.1.6) (see Gwinner \cite{157}, \cite{159}), where, instead of the operator $-\Delta u$, the $p$-harmonic operator $-\text{div} \left( \|\nabla u\|^{p-2} \nabla u \right)$ with $p > 2$ is considered. In case the differential operator is not strongly monotone, Lemma 3.1 in \cite{157} gives a useful criterion for the well-posedness of the variational inequality.

\section{Elliptic Regularization}

This chapter is mainly focused upon the applications of proximal-like methods for solving variational inequalities with degenerate or singularly perturbed elliptic operators. The conditions on the data approximation in the IPR-method (cf. Assumptions \ref{assumption:8.5.3}(iii)-(v) below) are weaker than those arising from the theory of variational convergence for ill-posed problems, and at the same time they are well-coordinated with the estimates of finite element interpolation in Sobolev spaces (see the analysis of these conditions in Subsection \ref{subsection:8.5.3} below).

For variational inequalities related to minimal surface problems and to convection-diffusion problems, considered in Subsection \ref{subsection:8.5.3}, the studied general framework covers a new (proximal based) elliptic regularization method, in which a successive approximation of the set $K$ is performed by means of the finite element method on a sequence of triangulations. The idea of elliptic regularization, proposed by Lions \cite{268} and Olejnik \cite{311}, consists in an approximation of a degenerate elliptic problem (parabolic problems are treated as a special case) by a sequence of non-degenerate elliptic problems. In the classical scheme it is carried out by adding the term $\epsilon T (T$ an appropriately chosen operator) to the operator of the original problem and by considering the sequence of solutions of the regularized problems for $\epsilon \downarrow 0$.

Elliptic regularization is an efficient tool for the theoretical analysis of degenerate problems (cf. \cite{234}, \cite{268}, \cite{271}) and serves as a basis for some numerical methods (see \cite{106} and references therein). However, the necessity to reduce the parameter $\epsilon$ to 0 enforces hard requirements on the exactness of the discretization and causes ill-conditioning of the discretized problems.

Applying a proximal elliptic regularization, one can choose the regularization parameter apart from 0, and one obtains a sequence of uniformly elliptic auxiliary problems with a common constant of ellipticity.

Noteworthy is that for the minimal surface problems (with and without obstacles) considered in a new space $H^L(\Omega)$, the special analysis based on Theorem \ref{theorem:8.5.13} establishes a new result on the convergence of the proximal elliptic regularization method in the space $W^{1,1}(\Omega)$, whereas the general theory of proximal point methods guarantees weak convergence in $H^1(\Omega)$ only (cf. Lemma \ref{lemma:4.1.2}). The proved strong convergence for the convection-diffusion problem is not so surprising, because the operator of the problem is strongly monotone although singularly perturbed, too.

In both cases regularization permits to obtain "well-conditioned" discretized problems if the successive approximation of $K$ is performed by means of stan-
CHAPTER 8. PPR FOR VARIATIONAL INEQUALITIES

dard finite element techniques.

8.5.1 A generalized proximal point method

Let \((V, \| \cdot \|)\) be a Hilbert space with the topological dual \(V'\). We consider the variational inequality

\[
\text{VI}(Q, \varphi, K) \quad \text{find } u \in K \text{ and } p \in \partial \varphi(u) : \langle Q(u) + p, v - u \rangle \geq 0, \quad \forall \ v \in K,
\]

where \(K\) is a convex, closed subset of \(V\); \(Q : V \to V'\) is a single-valued, monotone operator, its domain \(D(Q)\) contains \(K\) and \(Q\) is hemicontinuous on \(K\) (see Definition A1.6.38); \(\varphi : V \to \bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}\) is a convex, proper and lower-semicontinuous functional, \(\partial \varphi\) denotes its subdifferential and \(D(\partial \varphi) \supset K\).

In this subsection we continue our general algorithmic framework of IPR, now for solving \(\text{VI}(Q, \varphi, K)\). As before, it joins iterative proximal regularization and data approximation in a manner of a diagonal process, i.e. approximation of \(K\) and \(\partial \varphi\) is improved after each proximal iteration. If the operator \(Q\) possesses a certain reserve of monotonicity as described in the Assumptions 8.5.1 and 8.5.3(ii) below, conditions on the choice of the regularizing functional admit the application of \textit{weak proximal regularization} as well as of \textit{regularization on a subspace}, introduced in Section 8.2.3.

In the sequel the following assumption concerning \(\text{VI}(Q, \varphi, K)\) will be used.

8.5.1 Assumption.

(i) \(K \cap \text{int} D(\partial \varphi) \neq \emptyset\);

(ii) for a given linear continuous and monotone operator \(B : V \to V'\) with symmetry property \(\langle Bu, v \rangle = \langle Bv, u \rangle\), the inequality

\[
\langle Q(u) - Q(v), u - v \rangle \geq \langle B(u - v), u - v \rangle, \quad \forall \ u, v \in D(Q)
\]

is valid;

(iii) \(\text{VI}(Q, \varphi, K)\) is solvable. ♦

We denote the solution set of \(\text{VI}(Q, \varphi, K)\) by \(\text{SOL}(Q, \varphi, K)\) and write \(\text{VI}(Q, K)\) and \(\text{SOL}(Q, K)\) in case \(\varphi \equiv 0\).

With the normality operator

\[
\mathcal{N}_K : u \mapsto \begin{cases} \{ z \in V' : \langle z, u - v \rangle \geq 0 \quad \forall \ v \in K \} & \text{if } u \in K, \\ \emptyset & \text{otherwise,} \end{cases}
\]

the maximal monotonicity of \(Q + \mathcal{N}_K\) follows from Theorem 1 in [350]. Hence, Assumption 8.5.1(i) provides that the operator \(Q + \mathcal{N}_K + \partial \varphi\) is maximal monotone, too (see [350], Theorem 3). The IPR-framework studied here includes successive approximation of the set \(K\) and of the functional \(\varphi\) by means of a sequence \(\{K^k\}\), \(K^k \subseteq K\), of convex closed sets and a sequence \(\{\varphi_k\}\), \(\varphi_k : V \to \bar{\mathbb{R}}\), of convex functionals, respectively. Moreover, we suppose that \(\varphi_k\) is Gâteaux-differentiable on \(K\) and \(\nabla \varphi_k\) is hemicontinuous on \(K\).
Let $\varrho : V \to \mathbb{R}$ be a convex Gâteaux-differentiable functional such that
\[ v \mapsto \varrho(v) + \langle Bv, v \rangle \]
is strongly convex on $D(Q)$, where $B$ satisfies Assumption 8.4.1(ii).

As usual, we make also use of the controlling sequences $\{\delta_k\}$ and $\{\chi_k\}$ such that
\[ \delta_k \geq 0, \quad \lim_{k \to \infty} \delta_k = 0, \quad 0 < \chi_k \leq \bar{\chi} < \infty. \quad (8.5.1) \]
The choice of $\{K_k\}$, $\{\varphi_k\}$, $\varrho$ as well as of the sequences $\{\delta_k\}$ and $\{\chi_k\}$ will be specified in Assumption 8.5.3.

8.5.2 Method. (Generalized Proximal Point Method)
Data: $x^0 \in K$; $\chi_0 > 0$, $\delta_0 \geq 0$.

S1: Set $k := 0$.

S2: If $u^k$ is a solution of $\text{VI}(Q, \varphi, K)$, stop.

S3: Find $u^{k+1} \in K_k$ such that
\[ (P^k) \quad \langle Q(u^{k+1}) + \nabla \varphi_k(u^{k+1}) + \chi_k(\nabla \varrho(u^{k+1}) - \nabla \varrho(u^k)), v - u^{k+1} \rangle \geq -\delta_k \|v - u^{k+1}\|, \quad \forall v \in K_k. \]

S4: Select $\chi_{k+1}, \delta_{k+1}$, set $k := k + 1$, go to S2.

8.5.3 Assumption.

(i) $\varrho : V \to \mathbb{R}$ is a convex functional and the mapping $\nabla \varrho$ is Lipschitz continuous on $D(Q)$;

(ii) with given constants $\bar{\chi} \geq 0$, $m > 0$ and the operator $B$ satisfying Assumption 8.5.1(ii), the inequality
\[ \frac{\bar{\chi}}{2} \langle B(u - v), u - v \rangle + \varrho(u) - \varrho(v) - \langle \nabla \varrho(v), u - v \rangle \geq m \|u - v\|^2 \]
holds for all $u, v \in D(Q)$, and the choice of $\bar{\chi}$ in (9.3.1) provides $2\bar{\chi}\bar{\chi} < 1$;

(iii) for each $w \in K$, there exists a sequence $\{w^k\}, w^k \in K_k$, such that
\[ w^k \rightharpoonup w, \quad Q(w^k) \to Q(w), \quad \lim_{k \to \infty} \varphi_k(w^k) = \varphi(w); \]

(iv) $\varphi_k(v) \geq \varphi(v), \quad \forall v \in K, \forall k$;

(v) with given nonnegative constants $c_1, c_2, c$ and sequences $\{h_k\}$, $\{\sigma_k\}$ satisfying
\[ \sum_{k=1}^{\infty} \frac{h_k}{\chi_k} < \infty, \quad \sum_{k=1}^{\infty} \frac{\sigma_k}{\chi_k} < \infty, \quad (8.5.2) \]
there exists a sequence \( \{w^k\}, w^k \in K_k \), such that

\[
\begin{align*}
\|w^k - u^*\| & \leq c_1 h_k, \\
\|Q(w^k) - Q(u^*)\|_{V'} & \leq c_2 \sigma_k, \\
\varphi_k(w^k) - \varphi(u^*) & \leq c \sigma_k 
\end{align*}
\]  \hspace{1cm} (8.5.3)

for some \( u^* \in SOL(Q, \varphi, K) \).

For instance, for contact problems (with given friction), taking into account relation (8.2.3), the operator

\[ B : \langle Bu, v \rangle := c_0 \int_\Omega \epsilon_{kl}(u)\epsilon_{kl}dx, \quad \forall u, v \in V \]

satisfies Assumption 8.5.1(ii), and the second Korn inequality (see (8.2.55)) allows one to guarantee the validity of Assumption 8.5.3(ii) with the regularizing functional

\[ \varrho : u \mapsto \|u\|^2_{L^2(\Omega)}, \]  \hspace{1cm} (8.5.4)

which meets also Assumption 8.5.3(i). Of course, \( \varrho : u \mapsto \|u\|^2 \) is also a possible choice, but (8.5.4) is more preferable from the numerical point of view.

8.5.4 Remark. In Subsection 8.5.3.1 we meet a situation when the assumption \( K_k \subset K \) can be rather restrictive, especially if the sets \( K_k \) are constructed by means of standard finite element techniques. In the problems considered there, however, we have \( \varphi \equiv 0 \) and the operator \( Q \) possesses good additional properties like strict monotonicity and Lipschitz continuity on the whole space \( V \). This permits one to replace the assumption \( K_k \subset K \) by a weaker one:

(a) with a given \( c_3 > 0 \) and \( \{h_k\} \) as in (9.3.2), for any sequence \( \{v^k\}, v^k \in K_k \), there exists a sequence \( \{z^k\} \subset K \) such that

\[ \|z^k - v^k\| \leq c_3(\|v^k - u^*\|^2 + 1)h_k, \quad \forall k, \]

or

(b) with a given \( c_3 > 0 \) and \( \{h_k\}, \{v^k\} \) as in (a), there exists a sequence \( \{z^k\} \subset K \) such that

\[ \langle Q(u^*), z^k - v^k \rangle \leq c_3(\|v^k - u^*\|^2 + 1)h_k, \quad \forall k; \]

moreover, each weak limit point of \( \{v^k\} \) belongs to \( K \).

All statements of Subsection 8.5.2 (except for Theorem 8.5.13 if the weaker assumption (b) is used) remain true, the technical modifications in the proofs can be carried out on the base of the convergence analysis in [225].

8.5.5 Remark. If the Assumptions 8.5.1(ii), 8.5.3(i) and (ii) are valid, then for each \( k \), the operator

\[ v \mapsto Q(v) + \nabla \varphi_k(v) + \chi_k(\nabla \varrho(v) - \nabla \varrho(v^k)) + N_{K_k}(v) \]

is maximal monotone ([350], Theorem 3). Hence, the exact problems \( (P^k) \) (with \( \delta_k = 0 \) has a unique solution, and the solvability of the inexact ones (with \( \delta_k > 0 \) is guaranteed. 

\[ \diamond \]
8.5.2 Convergence analysis of the method

In the sequel we need the following modification of Minty’s lemma [290].

8.5.6 Lemma. Let Assumption 8.5.1(i) be valid, and for some \( u \in K \) and any \( v \in K \) there exists \( p(v) \in \partial \varphi(v) \) such that

\[
\langle Q(v) + p(v), v - u \rangle \geq 0.
\]  

(8.5.5)

Then, with some \( p \in \partial \varphi(u) \) the inequality

\[
\langle Q(u) + p, v - u \rangle \geq 0
\]  

(8.5.6)

holds for all \( v \in K \).

Proof: Denote

\[ G : v \mapsto Q(v) + \partial \varphi(v) + \mathcal{I}(v - u), \]

where \( \mathcal{I} : V \to V' \) is the canonical isometry operator. The operator \( \mathcal{G} := G + \mathcal{N}_K \) is maximal monotone and strongly monotone. Therefore, there exists \( w \in K \), such that \( 0 \in \mathcal{G}(w) \), and using the definition of \( \mathcal{N}_K \), we infer from here that

\[
\langle g(w), v - w \rangle \geq 0, \quad \forall \, v \in K
\]  

(8.5.7)

holds with some \( g(w) \in \mathcal{G}(w) \). In view of the convexity of the functional \( \varphi \), the last inequality implies

\[
\langle Q(w) + \mathcal{I}(w - u), v - w \rangle + \varphi(v) - \varphi(w) \geq 0, \quad \forall \, v \in K.
\]  

(8.5.8)

If \( w = u \), then of course \( g(w) \in Q(w) + \partial \varphi(w) \) and the conclusion of the lemma is obvious.

Suppose now that \( w \neq u \). We make use of the relation

\[
\langle \tilde{g}(v), v - u \rangle \geq 0, \quad \forall \, v \in K
\]  

(8.5.9)

which follows from (9.3.5) for \( \tilde{g}(v) := g(v) + \mathcal{I}(v - u) \) with an appropriate \( g(v) \in Q(v) + \partial \varphi(v) \). Take \( w_\lambda = u + \lambda(w - u) \) for \( \lambda \in [0, 1) \). Obviously, \( w_\lambda \in K \), and according to (9.3.9), for each \( \lambda \) there exists \( \tilde{g}(w_\lambda) \in \mathcal{G}(w_\lambda) \) satisfying

\[
\langle \tilde{g}(w_\lambda), w - u \rangle \geq 0.
\]  

(8.5.10)

Using again the convexity of \( \varphi \), one can immediately conclude from (9.3.10) that

\[
\langle Q(w_\lambda) + \mathcal{I}(w_\lambda - u), w - u \rangle + \frac{1}{1 - \lambda} [\varphi(w) - \varphi(u + \lambda(w - u))] \geq 0. \]  

(8.5.11)

Passing to the limit in (9.3.11) for \( \lambda \downarrow 0 \) and observing that the operator \( Q \) is hemicontinuous on \( K \) and the functional \( \varphi \) is lsc, we get

\[
\langle Q(u), w - u \rangle + \varphi(w) - \varphi(u) \geq 0.
\]  

(8.5.12)

Inequality (9.3.8) (given with \( v = u \)) together with (9.3.12) leads to

\[
\langle Q(u) - Q(w), u - w \rangle + \langle \mathcal{I}(u - w), u - w \rangle \leq 0,
\]

but this contradicts the monotonicity of \( Q \). □
8.5.7 Remark.

(a) The reverse conclusion that (9.3.6) implies (9.3.5) (with any \( p(v) \in \partial \varphi(v) \)) follows immediately from the monotonicity of the operator \( Q + \partial \varphi \). Using this fact and Lemma 8.5.6, one can easily show that SOL(\( Q, \varphi, K \)) is a convex closed set.

(b) Under Assumption 8.5.1(i) the following statements are equivalent:

(b1) \( u \in K \) and \( (Q(v), v - u) + \varphi(v) - \varphi(u) \geq 0, \forall v \in K \);

(b2) \( u \in K \) and \( \exists p \in \partial \varphi(u): (Q(u) + p, v - u) \geq 0, \forall v \in K \).

Indeed, the implication (b2) \( \Rightarrow \) (b1) is evident. But if (b1) is fulfilled, then the monotonicity of \( Q \) and convexity of \( \varphi \) yield

\[
(Q(v), v - u) + (p(v), v - u) \geq 0, \forall v \in K, \forall p(v) \in \partial \varphi(v),
\]

and Lemma 8.5.6 provides the validity of (b2). ♦

Now, with \( \tilde{\chi} \) from Assumption 8.5.3(ii) we introduce the function

\[
\Gamma : (u, v) \mapsto \tilde{\chi} (B(v - u), v - u) + \varrho(u) - \varrho(v) - (\nabla \varrho(v), u - v). \tag{8.5.13}
\]

For \( u^* \) chosen as in Assumption 8.5.3(v), \( \Gamma(u^*, \cdot) \) plays the role of a Lyapunov function in this convergence analysis.

8.5.8 Lemma. Let the Assumptions 8.5.1(ii), (iii), 8.5.3(i), (ii), (iv) and (v) be satisfied and \( \sum_{k=1}^{\infty} \frac{\delta_k}{\chi_k} < \infty \). Then it holds

(a) the sequence \( \{u^k\} \) generated by Method 8.5.2 is bounded;

(b) \( \lim_{k \to \infty} \Gamma(u^{k+1}, u^k) = 0 \);

(c) \( \lim_{k \to \infty} ||u^{k+1} - u_k|| = 0 \);

(d) sequence \( \{\Gamma(u^*, u^k)\} \) converges.

Sketch of the proof: Let \( \{w^k\} \) be chosen as in Assumption 8.5.3(v). Taking into account Assumption 8.5.3(iv), (v) and the convexity of \( \varphi_k \), we obtain

\[
(\nabla \varphi_k(u^{k+1}), w^k - u^{k+1}) \leq \varphi_k(w^k) - \varphi_k(u^{k+1})
\]

\[
= [\varphi_k(w^k) - \varphi(u^*)] + [\varphi(u^*) - \varphi_k(u^{k+1})]
\]

\[
\leq \sigma + \varphi(u^*) - \varphi(u^{k+1}). \tag{8.5.14}
\]

Because \( u^* \in SOL(Q, \varphi, K) \) and \( u^{k+1} \in K_k \subset K \), the inequality

\[
(Q(u^*), u^{k+1} - u^*) + \varphi(u^{k+1}) - \varphi(u^*) \geq 0 \tag{8.5.15}
\]

is valid.

Applying \( (P^k) \), (9.3.17) and (9.3.18) for the estimation of

\[
(\nabla \varrho(u^k) - \nabla \varrho(u^{k+1}), w^k - u^{k+1}),
\]

then the rest of the proof is analogous to those of Lemma 2 in [225]. □
8.5. **ELLIPIC REGULARIZATION**

345

8.5.9 Lemma. Let the Assumptions 8.5.1(i), 8.5.3(i), (iii) and (iv) be fulfilled. Moreover, suppose that the sequence \{u^k\} generated by Method 8.5.2 is bounded and

\[
\lim_{k \to \infty} \|u^{k+1} - u^k\| = 0.
\]

Then each weak limit point of \{u^k\} is a solution of VI(Q, \varphi, K).

**Proof:** Let the subsequence \{u^k\}_{k \in \mathbb{N}} converge weakly to \bar{u}. Because \(K_k \subset K \forall k \) and \(K\) is a closed convex set, one gets \(\bar{u} \in K\), whereas \(\lim_{k \to \infty} \|u^{k+1} - u^k\| = 0\) implies \(\bar{u}^k \to \bar{u}\) if \(k \in \mathbb{S}, k \to \infty\).

According to Assumption 8.5.3(ii), for each \(v \in K\), one can choose a sequence \{v^k\}, \(v^k \in K_k\), such that \(v^k \rightharpoonup v\) for \(k \to \infty\) and

\[
\lim_{k \to \infty} \|Q(v^k) - Q(v)\|_V = 0, \quad \lim_{k \to \infty} \varphi_k(v^k) = \varphi(v). \tag{8.5.16}
\]

By the definition of \{u^k\} (see \(P^k\)) and the inclusion \(v^k \in K_k\), the inequality

\[
(Q(u^{k+1}) + \nabla \varphi_k(u^{k+1}) + \chi_k (\nabla g(u^{k+1}) - \nabla g(u^k)), v^k - u^{k+1}) 
\geq -\delta_k \|v^k - u^{k+1}\|
\]

holds for all \(k\). Due to the monotonicity of \(Q\), the convexity of \(\varphi_k\) and Assumption 8.5.3(iv), this leads to

\[
(Q(v^k) + \chi_k (\nabla g(u^{k+1}) - \nabla g(u^k)), v^k - u^{k+1}) + \varphi_k(v^k) - \varphi(u^{k+1}) 
\geq -\delta_k \|v^k - u^{k+1}\|
\]

Now, passing to the limit in the latter inequality for \(k \in \mathbb{S}, k \to \infty\), we obtain from (9.3.1), (9.3.19), and Assumption 8.5.3(i) that

\[
\lim_{k \to \infty} \|u^{k+1} - u^k\| = 0, \quad v^k \rightharpoonup v, \quad u^{k+1} \to \bar{u}
\]

and the lower semicontinuity of \(\varphi\) that

\[
(Q(v), v - \bar{u}) + \varphi(v) - \varphi(\bar{u}) \geq 0, \quad \forall v \in K.
\]

Finally, Lemma 8.5.6 and Remark 8.5.7(b) enable us to conclude that \(\bar{u} \in SOL(Q, \varphi, K)\). \(\square\)

8.5.10 Theorem. Let the Assumptions 8.5.1 and 8.5.3 be fulfilled and

\[
\sum_{k=1}^{\infty} \frac{\delta_k}{\chi_k} < \infty.
\]

Then it holds

(i) Problem \((P^k)\) is solvable for each \(k\), the sequence \{u^k\} generated by Method 8.5.2 is bounded and each weak limit point of \{u^k\} is a solution of VI(Q, \varphi, K).

(ii) If, in addition, Assumption 8.5.3(v) holds for each \(u \in SOL(Q, \varphi, K)\) (the constants \(c_1, c_2, c\) may depend on \(u\)) and

\[
v^k \rightharpoonup v \text{ in } V, \quad v^k \in K_k \implies \nabla g(v^k) \rightharpoonup \nabla g(v) \text{ in } V', \tag{8.5.17}
\]

then the whole sequence \{u^k\} converges weakly to \(u^* \in SOL(Q, \varphi, K)\).
(iii) If, moreover, there exists a linear compact operator \( \hat{B} : V \to V' \) such that \( B + \hat{B} \) is strongly monotone, then \( \{u^k\} \) converges strongly to \( u^* \in SOL(Q, \varphi, K) \).

**Proof:** Conclusion (i) follows immediately from Remark 8.5.5 and the Lemmata 8.5.8 and 8.5.9.

The proof of conclusion (ii) is the same as in [225], Theorem 1.

Thus, it remains to prove (iii) only. Let \( u^* \) be the weak limit of \( \{u^k\} \). Choosing \( \{u^k\} \) according to Assumption 8.5.3(v) we obtain from Assumption 8.5.1(ii)

\[
\langle B(u^{k+1} - u^*), u^{k+1} - u^* \rangle 
= \langle B(u^{k+1} - u^k), u^{k+1} - u^k \rangle 
- \langle B(u^k - u^{k+1}), u^{k+1} - u^* \rangle 
\leq \langle Q(u^{k+1}) - Q(u^k), u^{k+1} - u^k \rangle 
- \langle B(u^{k+1} - u^k), u^* - u^{k+1} \rangle.
\]

Now, we estimate the term \( \langle Q(u^{k+1}), u^{k+1} - u^k \rangle \) by setting \( v = u^{k+1} \) in \( (P^k) \), and then insert this estimate in (9.3.21). Together with Assumption 8.5.3(iv) this yields

\[
\langle B(u^{k+1} - u^*), u^{k+1} - u^* \rangle 
\leq \langle Q(u^k), w^k - u^{k+1} \rangle + \langle \nabla \varphi_k(u^{k+1}), w^k - u^{k+1} \rangle 
\quad + \chi_k (\nabla \varphi(u^k) - \nabla \varphi(u^{k+1}), w^k - u^{k+1}) + \delta_k \|w^k - u^{k+1}\| 
\quad + \langle Bu^* + Bu^k - 2B(u^{k+1}), u^* - w^k \rangle 
\leq \langle Q(u^k) - Q(u^*), w^k - u^{k+1} \rangle + \langle Q(u^*), u^k - u^* \rangle + \langle Q(u^*), u^* - u^{k+1} \rangle 
\quad + \chi_k (\nabla \varphi(u^k) - \nabla \varphi(u^{k+1}), w^k - u^{k+1}) + \delta_k \|w^k - u^{k+1}\| 
\quad + \langle Bu^* + Bu^k - 2B(u^{k+1}), u^* - w^k \rangle + (\varphi_k(w^k) - \varphi_k(u^{k+1})). 
\]

Taking into account that the sequences \( \{u^k\} \) and \( \{u^{k+1}\} \) are bounded, one can conclude that all terms in the right hand side of (9.3.22) tend to zero for \( k \to \infty \).

Indeed, it vanishes

- the first, second and sixth term in view of Assumption 8.5.3(v);
- the third term because \( u^k \to u^* \);
- the fourth term due to \( \|u^{k+1} - u^k\| \to 0 \) and Assumption 8.5.3(i);
- the fifth term owing to \( \delta_k \to 0 \);
- the last term in view of Assumption 8.5.3(v), \( u^{k+1} \to u^* \) and the lower semicontinuity of \( \varphi \).

Thus, (9.3.22) implies

\[
\lim_{k \to \infty} \langle B(u^k - u^*), u^k - u^* \rangle = 0. \quad (8.5.20)
\]

At the same time

\[
\lim_{k \to \infty} \langle \hat{B}(u^k - u^*), u^k - u^* \rangle = 0 \quad (8.5.21)
\]

follows from \( u^k \to u^* \) and the compactness of \( \hat{B} \). Adding (9.3.23), (9.3.24) and obeying the strong monotonicity of \( \hat{B} + B \) we conclude finally that \( \{u^k\} \) converges to \( u^* \) strongly in \( V \). \( \square \)
8.5.11 Remark. From the compactness of the canonical injection $I : H^1(\Omega) \to L^2(\Omega)$ ($\Omega$ is here an open domain in $\mathbb{R}^2$ with a Lipschitz continuous boundary) and the second Korn inequality, the existence of an operator $\mathcal{B}$ satisfying condition (iii) of Theorem 8.5.10 can be shown, in particular, for ill-posed elliptic variational inequalities, which describe the problem of linear elasticity with given friction and the two-body contact problem. We dealt with these problems in Sections 8.2 and 8.3.

For the problem of linear elasticity, for example, the operator $\mathcal{B}$ defined by

$$
\langle \mathcal{B}u, v \rangle = \int_{\Omega} (u_1 v_1 + u_2 v_2) \, dx, \quad \forall \, u, v \in V := [H^1(\Omega)]^2
$$

is appropriate.

The following assumption serves to establish a more qualitative convergence of Method 8.5.2 in those situations when the conditions (ii) and (iii) of Theorem 8.5.10 are not guaranteed.

Let $Y \supset V$ be a Banach space with the norm $\| \cdot \|_Y$, and

$$
dist_Y (y, A) := \inf_{z \in A} \| y - z \|_Y.
$$

According to Lemma 8.5.8, there exists $\rho > 0$ such that

$$
\{ u^k \} \subset B_{\rho}, \quad \text{and} \quad B_{\rho} \cap SOL(Q, \varphi, K) \neq \emptyset.
$$

Denote $S^* := SOL(Q, \varphi, K) \cap B_{\rho}$.

8.5.12 Assumption. There exists a continuous function $\tau : [0, \infty) \to [0, \infty)$, $\tau(0) = 0$, $\tau(s) > 0 \forall \, s > 0$, such that

$$
(Q(v), v - u) + \varphi(v) - \varphi(u) \geq \tau(\text{dist}_Y(v, S^*)) , \quad \forall \, u \in S^*, \forall \, v \in K \cap B_{\rho}.
$$

For a fixed $u \in K \cap B_{\rho}$, the function

$$
\xi(\cdot, u) : v \to (Qv, v - u) + \varphi(v) - \varphi(u)
$$

possesses the properties

$$
\xi(v, u) \geq 0 \quad \forall \, v \in K \quad \Leftrightarrow \quad u \in S^*,
$$

$$
\xi(v, u) = 0 \quad \text{if} \quad v \in SOL(Q, \varphi, K), \quad u \in S^*.
$$

Assumption 8.5.12 describes a growth condition for the function $\inf_{u \in S^*} \xi(\cdot, u)$ on the set $(K \cap B_{\rho}) \setminus S^*$.

In case $Q = 0$, $K = V$, $Y = V$ and $\tau(s) = cs^2$, Assumption 8.5.12 is closely related to a growth condition used by Kort and Bertsekas [246] for the quadratic method of multipliers in convex programming, which in fact is a proximal point method applied to the dual program.

8.5.13 Theorem. Let the conditions of Lemma 8.5.8 and Assumption 8.5.12 be fulfilled. Moreover, suppose that Assumption 8.5.3 is valid for each $u^* \in S^*$ and the operator $Q$ is bounded on $K \cap B_{\rho}$. Then, for the sequence $\{ u^k \}$ generated by Method 8.5.2 it holds

$$
\lim_{k \to \infty} \text{dist}_Y (u^k, S^*) = 0. \quad (8.5.22)
$$
\(8.5.12\) Assumption

Now the properties of \(Q\) and applying \((P)\) in a

\(8.5.8\) obtain

\[\Gamma(u^*, u^{k+1}) - \Gamma(u^*, u^k)\]

\[= -\Gamma(u^{k+1}, u^k) + \langle \nabla \varphi(u^k) - \nabla \varphi(u^{k+1}), u^* - w^k + w^k - u^{k+1} \rangle\]

\[+ 2\bar{\chi}(B(u^k - u^{k+1}), u^* - u^{k+1}),\]

and applying \((P^k)\) to estimate the term \(\langle \nabla \varphi(u^k) - \nabla \varphi(u^{k+1}), w^k - u^{k+1} \rangle\) we obtain

\[\Gamma(u^*, u^{k+1}) - \Gamma(u^*, u^k)\]

\[\leq 2\bar{\chi}(B(u^k - u^{k+1}), u^* - u^{k+1})\]

\[+ \langle \nabla \varphi(u^k) - \nabla \varphi(u^{k+1}), u^* - w^k \rangle + \frac{\delta_k}{\chi_k} \|w^k - u^{k+1}\|\]

\[+ \frac{1}{\chi_k}(Q(u^{k+1}), w^k - u^{k+1}) + \frac{1}{\chi_k}(\nabla \varphi_k(u^{k+1}), w^k - u^{k+1}).\tag{8.5.23}\]

Now, taking into account the convexity of \(\varphi_k\) and Assumption \(8.5.3\)(iv), inequality (9.3.26) yields

\[\Gamma(u^*, u^{k+1}) - \Gamma(u^*, u^k)\]

\[\leq 2\bar{\chi}(B(u^k - u^{k+1}), u^* - u^{k+1})\]

\[+ \langle \nabla \varphi(u^k) - \nabla \varphi(u^{k+1}), u^* - w^k \rangle + \frac{\delta_k}{\chi_k} \|w^k - u^{k+1}\|\]

\[+ \frac{1}{\chi_k}(Q(u^{k+1}), w^k - u^*) + \frac{1}{\chi_k} (\varphi_k(w^k) - \varphi(u^*))\]

\[+ \frac{1}{\chi_k} [Q(u^{k+1}), u^* - u^{k+1} + \varphi(u^*) - \varphi(w^k)].\]

In view of Assumption \(8.5.12\) and \(\chi_k \leq \bar{\chi}\), the last term can be replaced by

\[-\frac{1}{\bar{\chi}} \tau \left( \text{dist}_Y(u^{k+1}, S^*) \right).

Finally, passing to the limit in the so modified inequality, and owing to Lemma \(8.5.8\), the assumptions Assumption \(8.5.3\)(i), (v), the boundedness of the operator \(Q\) on \(K \cap \mathbb{B}_\rho\), and \(\sum_{k=1}^{\infty} \frac{\delta_k}{\chi_k} < \infty\), we immediately obtain

\[\lim_{\tau \to \infty} \tau(\text{dist}_Y(u^{k+1}, S^*)) = 0.

Now the properties of \(\tau\) imply the validity of (9.3.25). \(\square\)

We conclude this section with a statement which can be useful for checking Assumption \(8.5.12\). It allows us to analyze a growth property of the function

\[v \mapsto \inf_{u \in S} \{ \langle Q(v), v - u \rangle + \varphi(v) - \varphi(u) \}\]

in a \(Y\)-neighborhood of \(S^*\) only. With \(\rho\) as above, a constant \(c\) such that

\[\|v\|_Y \leq c\|v\| \forall v \in V,\]

and a given \(\delta \in (0, 2c\rho)\), we consider the set

\[K_\delta := \{ v \in K \cap \mathbb{B}_\rho : \text{dist}_Y(v, S^*) \leq \delta \}.\]
8.5.14 Lemma. Suppose that \( \tau : [0, \infty) \rightarrow [0, \infty) \) is a nondecreasing function such that
\[
(Q(v), v - u) + \varphi(v) - \varphi(u) \geq \tau(\text{dist}_Y(v, S^*)) \tag{8.5.24}
\]
is valid for any \( v \in K_\delta \), \( u \in S^* \). Then the inequality
\[
\inf_{w \in S^*} [(Q(v), v - u) + \varphi(v) - \varphi(u)] \geq \tau \left( \frac{\delta}{2c_\rho} \text{dist}_Y(v, S^*) \right) \tag{8.5.25}
\]
holds for any \( v \in K \cap \mathbb{B}_\rho \).

Proof: Obviously, we need to check (8.5.25) for \( v \in (K \cap \mathbb{B}_\rho) \setminus K_\delta \) only. Take an arbitrary \( u \in S^* \) and introduce the function
\[
\psi(\lambda) := \text{dist}_Y(\lambda v + (1 - \lambda)u, S^*) \quad \text{for } 0 < \lambda < 1.
\]
In view of the convexity of \( S^* \) (see Remark 8.5.7(a)), the function \( \text{dist}_Y(\cdot, S^*) \) is convex, and this implies the convexity of \( \psi \).
In turn, because \( \text{dom} \psi = (0, 1) \), \( \psi \) is continuous on \((0, 1)\), and taking into account that \( u \in S^* \subset K_\delta \), \( v \notin K_\delta \), we conclude that there exists \( \lambda \in (0, 1) \) such that
\[
\bar{v} := \lambda v + (1 - \lambda)u \in K_\delta \quad \text{and} \quad \text{dist}_Y(\bar{v}, S^*) = \delta.
\]
Now using
\[
\bar{v} - u = \lambda(v - u), \quad \frac{1 - \lambda}{\lambda}(\bar{v} - u) = v - \bar{v},
\]
from the monotonicity of \( Q \) and the convexity of \( \varphi \) we obtain
\[
\frac{1 - \lambda}{\lambda} (Q(v) - Q(\bar{v}), \bar{v} - u) = \langle Q(v) - Q(\bar{v}), v - \bar{v} \rangle \geq 0
\]
and
\[
\varphi(v) - \varphi(u) \geq \frac{1}{\lambda} (\varphi(\bar{v}) - \varphi(u)).
\]
Therefore,
\[
(Q(v), v - u) = \frac{1}{\lambda} (Q(v), \bar{v} - u) \geq \frac{1}{\lambda} (Q(\bar{v}), \bar{v} - u)
\]
and
\[
(Q(v), v - u) + \varphi(v) - \varphi(u) \geq \frac{1}{\lambda} [(Q(\bar{v}), \bar{v} - u) + \varphi(\bar{v}) - \varphi(u)]
\]
hold, and inequality (9.3.27) yields
\[
(Q(v), v - u) + \varphi(v) - \varphi(u) \geq \frac{1}{\lambda} \tau(\text{dist}_Y(\bar{v}, S^*))
\]
\[
\geq \tau(\text{dist}_Y(\bar{v}, S^*)).
\]
But, \( \text{dist}_Y(v, S^*) \leq 2c_\rho \) follows from \( v \in \mathbb{B}_\rho, \|v\|_Y \leq c\|v\| \), and \( S^* \subset \mathbb{B}_\rho \). Hence
\[
\text{dist}_Y(\bar{v}, S^*) \geq \frac{\delta}{2c_\rho} \text{dist}_Y(v, S^*),
\]
and taking into account the nondecreasing of \( \tau \), we conclude that
\[
(Q(v), v - u) + \varphi(v) - \varphi(u) \geq \tau \left( \frac{\delta}{2c_\rho} \text{dist}_Y(v, S^*) \right).
\]
Because \( u \in S^* \) is arbitrarily chosen, this leads to (8.5.25).

A growth condition like (9.3.27) with \( \tau(s) = cs^2 \) and \( \tau(s) = cs \) was introduced in [224] to investigate the rate of convergence of multi-step proximal regularization methods.

The result above will be applied in the next Subsection to show \( W^{1,1} \)-convergence of the iterates of Method 8.5.2 for the minimal surface problem and related variational inequalities considered in the \( H^1 \)-space.

### 8.5.3 Application to minimal surface problems

In the next two subsections we deal with elliptic variational problems in the space \( V = H^1(\Omega) \), where \( \Omega \) is an open domain in \( \mathbb{R}^2 \) with a Lipschitz continuous boundary \( \Gamma \). In this context the convex closed set \( K \) is defined as

\[
K = \{ v \in V : v = g \text{ on } \Gamma_1 \} \quad (8.5.26)
\]

or

\[
K = \{ v \in V : v = g \text{ on } \Gamma, \ v \geq \psi \text{ a.e. on } \Omega \}, \quad (8.5.27)
\]

where \( \Gamma_1 \subseteq \Gamma \), \( \mathrm{meas} \ \Gamma_1 > 0 \); \( g \) and \( \psi \) are sufficiently smooth functions on \( \Omega := \Omega \cup \Gamma \) and \( \psi \leq g \) on \( \Gamma \).

Applying Method 8.5.2 to these problems, a successive approximation of \( K \) by means of the finite element method on a sequence of triangulations \( \{ T_k \} \) is performed.

In accordance with the conditions on approximation, formulated in Assumption 8.5.3, we will suppose that

- \( \Omega \) is a polygonal domain;
- the solution of the problem is sufficiently smooth;
- a standard finite element method with piece-wise linear basis functions on the regular sequence of triangulations \( \{ T_k \} \) of \( \Omega \) is applied.

Analogously to Section 8.1, denote \( h_k \) the characteristic triangulation parameter of \( T_k \), i.e. \( h_k \) is the length of the largest edge of the triangles \( T \) in \( T_k \); \( \Sigma_k \) indicates the set of vertices of all triangles in \( T_k \); \( \Sigma_k(\Gamma_1) \), \( \Sigma_k(\Gamma) \) are the sets of all vertices lying on \( \Gamma_1 \) and \( \Gamma \), respectively; \( P_1 \) denotes the space of polynomials in two variables of degree \( \leq 1 \).

Then, on the functional space

\[
V^k := \{ v \in C(\Omega) : v|_T \in \mathcal{P}_1(T) \ \forall \ T \in T_k \} \quad (8.5.28)
\]

the sets (8.5.26) and (8.5.27) are approximated by

\[
K_k := \{ v \in V^k : v(a_i) = g(a_i) \ \forall \ a_i \in \Sigma_k(\Gamma_1) \} \quad (8.5.29)
\]

and

\[
K_k := \{ v \in V^k : v(a_i) = g(a_i) \ \forall \ a_i \in \Sigma_k(\Gamma), \quad v(a_i) \geq \psi(a_i) \ \forall \ a_i \in \Sigma_k \}, \quad (8.5.30)
\]

respectively.
8.5. ELLIPTIC REGULARIZATION

8.5.3.1 Formulations and properties of minimal surface problems

The classical (non-parametric) minimal surface problem, considered in the space $V := H^1(\Omega)$, can be formulated as follows:

$$\min \{ J(u) : u \in g + H^1_0(\Omega) \}, \quad J(u) := \int_\Omega \sqrt{1 + |\nabla u|^2} \, dx$$

(8.5.31)

$(g \in V$ is given), i.e. among all functions $u \in H^1(\Omega)$, $u = g$ on $\Gamma$, we are looking for a function, which defines a surface $z = u(x_1, x_2)$ with the smallest area.

Introducing the operator $Q : V \to V'$ defined by

$$\langle Q(u), v \rangle = \int_\Omega \frac{\nabla u \cdot \nabla v}{\sqrt{1 + |\nabla u|^2}} \, dx \quad \forall \, u, v \in V$$

and the affine set

$$K := \{ v \in V : v - g \in H^1_0(\Omega) \}$$

problem (8.5.31) can be rewritten as variational equality

$$\text{find } u \in K : \quad \langle Q(u), v \rangle = 0 \quad \forall \, v \in H^1_0(\Omega),$$

which in turn represents a weak formulation of the boundary value problem

$$-\text{div} \left( \frac{\nabla u}{\sqrt{1 + |\nabla u|^2}} \right) = 0 \quad \text{on } \Omega,$nabla u = g \quad \text{on } \Gamma.$$nabla u (8.5.34)

Equation (8.5.34) is nothing else but the Euler equation for the classical minimal surface problem.

The non-homogeneous problem

$$-\text{div} \left( \frac{\nabla u}{\sqrt{1 + |\nabla u|^2}} \right) = p \quad \text{on } \Omega,$nabla u = g \quad \text{on } \Gamma$$

is known as the Dirichlet problem for the equation of prescribed mean curvature. For a long history and survey of numerous investigations connected with these two problems we refer to the monographs [309] and [130].

The variational inequality VI($Q, K$) with $K$ given by (8.5.27) corresponds to the minimal surface problem with an obstacle. Problems of such type were mainly investigated in non-reflexive Banach spaces, where the functional

$$J(u) = \int_\Omega \sqrt{1 + |\nabla u|^2} \, dx$$

possesses better coercivity properties (see [234], Chapt. 3.4).

The problems considered here are not uniformly elliptic (see [253], Chapt. VI for the corresponding definition). Indeed, using the identity

$$\sum_{i,j=1}^2 \frac{\partial^2 f(t)}{\partial t_i \partial t_j} \xi_i \xi_j = \frac{\|\xi\|^2 + (t_2 \xi_1 - t_1 \xi_2)^2}{(1 + |t|^2)^{3/2}}$$


Symbols $a \cdot b$ and $|a|$ stand for the inner product and the Euclidean norm of vectors in $\mathbb{R}^2$, respectively.
for \( f(t) = \sqrt{1 + t_1^2 + t_2^2} \), we obtain

\[
\beta(u)|\xi|^2 \leq \sum_{i,j} \partial^2 f(t) \bigg|_{t = \nabla u} \xi_i \xi_j \leq \frac{1}{\sqrt{1 + \|\nabla u\|^2}} |\xi|^2, \quad \forall \xi \in \mathbb{R}^2,
\]

where \( \beta(u) > 0 \); for instance \( \beta(u) = (1 + \|\nabla u\|^2)^{-\frac{3}{2}} \) is appropriate. But the right inequality shows that \( \beta(u) \) cannot be separated from 0 uniformly in \( u \).

The violation of the uniform ellipticity causes serious difficulties in the theoretical and numerical analysis of these problems, including the investigation of their solvability.

8.5.15 Remark. The following facts point implicitly to the nature of these difficulties:

\( \circ \) For the minimal surface problem with

\[
\Omega := \{ x \in \mathbb{R}^2 : 1 < \|x\| < 2 \}, \quad g = \begin{cases} 
0 & \text{if } \|x\| = 2 \\
\gamma & \text{if } \|x\| = 1
\end{cases}
\]

there exists \( \gamma^* \) such that the problem is solvable for \( \gamma \in [0, \gamma^*] \) and has no solution if \( \gamma > \gamma^* \) (for this well-known example see, for instance, [100], Chapt. V);

\( \circ \) A necessary condition for the solvability of the Dirichlet problem for the equation of prescribed mean curvature is that

\[
\left| \int_\omega p(x)dx \right| < \text{mes } \partial \omega
\]

holds for all proper subsets \( \omega \subset \Omega \) with Lipschitz continuous boundaries, \( (\text{mes } \partial \omega \text{ denotes the perimeter of } \omega) \), cf. [130].

The existence of a classical solution of problem (8.5.34) with continuous data was proved by Radó [342] in the case that \( \Omega \) is a convex set. Conditions ensuring that the solution of (8.5.34) belongs to \( C^{2,1}(\bar{\Omega}) \) can be found in [253], Theorem IV.10.9.

For the minimal surface problem with an obstacle, but in the space \( V = H^1_0(\Omega) \), Lewy and Stampacchia [265] have shown that the solution is in \( W^{2,s}(\Omega) \cap C^1(\Omega) \), \( 1 \leq s < \infty \), if \( \psi \in C^2(\Omega) \) and \( \Omega \) is a convex set with a smooth boundary.

The uniqueness of a solution (if it exists) in case \( K \) is given by (8.5.33) or (8.5.27) is a rather evident corollary of the strict convexity of the functional (8.5.35). In turn, the strict convexity of \( J \) on \( K \) can be concluded by integration (over \( \Omega \)) of the left inequality in (8.5.37) below given with \( a = \|\nabla u\|, \ b = \|\nabla v\| \), where \( u,v \in K \) (hence, \( u - v \in H^1_0(\Omega) \)).

8.5.16 Proposition. The operator \( Q \) in (8.5.32) is Lipschitz continuous on \( V \).
8.5. **ELLIPTIC REGULARIZATION**

353

**Proof:** For any \(u, v, w \in V\) one gets

\[
\left| \int_\Omega \left( \frac{1}{\sqrt{1 + |\nabla u|^2}} - \frac{1}{\sqrt{1 + |\nabla v|^2}} \right) \nabla u \cdot \nabla w \, dx \right|
\]

\[
\leq \int_\Omega \left| \frac{(\nabla u - \nabla v) \cdot (\nabla u + \nabla v)}{\sqrt{1 + |\nabla u|^2} \sqrt{1 + |\nabla v|^2}} \left( \sqrt{1 + |\nabla u|^2} + \sqrt{1 + |\nabla v|^2} \right) \right| \nabla u \cdot \nabla w \, dx
\]

\[
\leq \int_\Omega |\nabla u - \nabla v| |\nabla w| \, dx \leq ||u - v|| ||w||,
\]

whereas the Cauchy-Schwartz inequality implies

\[
\left| \int_\Omega \frac{\nabla u - \nabla v}{\sqrt{1 + |\nabla v|^2}} \cdot \nabla w \, dx \right| \leq \left( \int_\Omega |\nabla u - \nabla v|^2 \, dx \right)^{1/2} \|w\| \leq ||u - v|| ||w||.
\]

Thus, we have

\[
|\langle Q(u) - Q(v), w \rangle| \leq 2||u - v|| ||w||, \quad \forall w \in V,
\]

hence

\[
\|Q(u) - Q(v)\|_{V'} \leq 2||u - v||.
\]

\[\square\]

8.5.17 Proposition. Suppose that a solution \(u\) of \(\text{VI}(Q, K)\), with \(Q\) in (8.5.32) and \(K\) in (8.5.33) or (8.5.27), belongs to \(W^{1, \infty}(\Omega)\). Then Assumption 8.5.12 is valid with \(Y = W^{1,1}(\Omega)\), arbitrary \(\rho > ||u||\) and \(\tau(s) := c(\rho)s^2\).

**Proof:** Let us recall that \(u\) is the unique solution of \(\text{VI}(Q, K)\), hence \(S^* = \{u\}\).

The convexity of \(J\) implies

\[
J(v) - J(u) \leq \langle Q(v), v - u \rangle, \quad \forall v \in V
\]

and because \(\langle Q(u), v - u \rangle \geq 0\) holds true for \(v \in K\), we have for all \(v \in K\)

\[
\langle Q(v), v - u \rangle \geq J(v) - J(u) - \langle Q(u), v - u \rangle
\]

\[
= \int_\Omega \left[ \frac{1}{\sqrt{1 + |\nabla v|^2}} - \frac{1}{\sqrt{1 + |\nabla u|^2}} - \frac{\nabla u \cdot (\nabla v - \nabla u)}{\sqrt{1 + |\nabla u|^2}} \right] \, dx. \quad (8.5.36)
\]

With \(a \in \mathbb{R}^2, b \in \mathbb{R}^2\) the identity

\[
\sqrt{1 + |a|^2} - \sqrt{1 + |b|^2} = \frac{b \cdot (a - b)}{\sqrt{1 + |b|^2}} = \frac{|a - b|^2}{\sqrt{1 + |b|^2} \left( \sqrt{1 + |a|^2} \sqrt{1 + |b|^2} + 1 + b \cdot a \right)}
\]

is evident, and using the inequality

\[
\sqrt{1 + |a|^2} \sqrt{1 + |b|^2} \geq 1 + b \cdot a,
\]
this yields
\[ \sqrt{1 + |a|^2} - \sqrt{1 + |b|^2} - \frac{b \cdot (a - b)}{\sqrt{1 + |b|^2}} \geq \frac{|a - b|^2}{2 (1 + |b|^2) \sqrt{1 + |a|^2}} \]
\[ \geq \frac{1}{2(1 + |b|^2)} \frac{|a - b|^2}{1 + |a|^2} . \]  \hspace{1cm} (8.5.37)

From (8.5.37), given with \( a := \nabla v, b := \nabla u \), and (8.5.36) we conclude that
\[ \langle Q(v), v - u \rangle \geq \frac{1}{2(1 + M^2)} \int_{\Omega} \frac{|
abla v - \nabla u|^2}{1 + |
abla v|^2} \, d\Omega, \quad \forall \, v \in K, \] \hspace{1cm} (8.5.38)
where \( M := \|u\|_{W^{1,\infty}(\Omega)} \). But the Cauchy-Schwarz inequality
\[ \left| \int_{\Omega} zw \, d\Omega \right| \leq \left( \int_{\Omega} z^2 \, d\Omega \right)^{1/2} \left( \int_{\Omega} w^2 \, d\Omega \right)^{1/2} , \]
applied with \( z := \frac{|\nabla u - \nabla v|}{\sqrt{1 + |
abla v|^2}} \), \( w := \sqrt{1 + |
abla v|^2} \), implies
\[ \left( \int_{\Omega} |
abla u - \nabla v| \, d\Omega \right)^2 \leq \int_{\Omega} \frac{|
abla u - \nabla v|^2}{1 + |
abla v|^2} \, d\Omega \cdot \int_{\Omega} (1 + |
abla v|^2) \, d\Omega . \]
Together with (8.5.38), the latter inequality leads to
\[ \langle Q(v), v - u \rangle \geq \frac{1}{2(1 + M^2)} \cdot \frac{\left( \int_{\Omega} |
abla u - \nabla v| \, d\Omega \right)^2}{\int_{\Omega} (1 + |
abla v|^2) \, d\Omega} , \quad \forall \, v \in K. \] \hspace{1cm} (8.5.39)
For \( v \in K, \|v\| \leq \rho \), inequality (8.5.39) gives
\[ \langle Q(v), v - u \rangle \geq \frac{1}{2(1 + M^2)(\text{meas } \Omega + \rho^2)} \left( \int_{\Omega} |
abla u - \nabla v| \, d\Omega \right)^2 . \] \hspace{1cm} (8.5.40)
Now, we use the fact that the standard norm and seminorm of the space \( W^{1,1}(\Omega) \) are equivalent on the subspace \( W_0^{1,1}(\Omega) \). Because \( (u - v)|_\Gamma = 0 \) holds for any \( v \in K \), this implies
\[ \exists \, c > 0 : \int_{\Omega} |
abla u - \nabla v| \, d\Omega \geq c\|u - v\|_{W^{1,1}(\Omega)} , \quad \forall \, v \in K, \] \hspace{1cm} (8.5.41)
and the conclusion of Proposition 8.5.17 follows from (8.5.40) and (8.5.41) with
\[ c(\rho) = \frac{c^2}{2(1 + M^2)(\text{meas } \Omega + \rho^2)} . \]
\[ \square \]

It should be noted that the embedding \( H^1 \subset W^{1,1} \) is not a compact one.
8.5.3.2 Application of general proximal-point-method to minimal surface problems

Considering the application of Method 8.5.2 to VI(Q,K) (with Q in (8.5.32) and K in (8.5.33) or (8.5.27)), we suppose, as already mentioned, that Ω is a convex polygonal domain, VI(Q,K) is solvable and its solution \( u^* \) belongs to \( H^2(\Omega) \).

As a regularizing functional the functional \( \varrho : v \mapsto \| v \|_2 \) can be used (in case \( K_k \subset K \ \forall \ k \), the choice \( \varrho(v) = \| v \|_{H^1_0(\Omega)} \) may be preferable). Obviously, these functionals possess the property (9.3.20).

One can easily show that the operators \( v \mapsto Q(v) + \chi_k(\nabla \varrho(v) - \nabla \varrho(u^k)) \)

in the subproblems \( (P^k) \) of Method 8.5.2 are uniformly elliptic, in distinction to the operator \( Q \) itself. This is the reason to speak about an elliptic regularization. Moreover, choosing a positive sequence \( \{\chi_k\} \) apart from 0 (this is allowed by the conditions on the regularization parameter), the uniform ellipticity of these operators with a common constant of ellipticity is guaranteed. This is an important advantage in comparison with the classical elliptic regularization approach.

Applying the finite element method as described at the beginning of Subsection 8.5.3, we deal here with sets \( K_k \) given by (8.5.29) (but with \( \Gamma = \Gamma_1 \)) or (8.5.30). The inclusion \( K_k \subset K \) is not very realistic in this case, therefore we have to check Assumption 8.5.3 modified as described in Remark 8.5.4.

The validity of Assumption 8.5.3(i) and (ii) (with \( B = 0, \tilde{\chi} = 0 \)) is obvious. To show Assumption 8.5.3(iii) for an arbitrary \( w \in K \), if \( K \) is given by (8.5.27), one can rewrite \( K \) in the form

\[
K = g + \{ v \in H^1_0(\Omega) : v \geq \psi - g \}
\]

and then follow the proof of Theorem 3.2 in [182], Sect. 1.2. This provides

\[
\exists \ w^k \in K_k : \lim_{k \to \infty} \| w^k - w \| = 0. \tag{8.5.42}
\]

If \( K \) is given by (8.5.33), the relation (8.5.42) is well-known. The application of Proposition 8.5.16 and (8.5.42) yields

\[
\lim_{k \to \infty} \| Q(w^k) - Q(w) \|_{V'} = 0.
\]

Analogously, taking into account that \( u^* \in H^2(\Omega) \), the validity of Assumption 8.5.3(v) with \( \sigma_k := h_k \) follows from the properties of finite element approximation, see Theorem 3.2.1 in [75] and Proposition 8.5.16.

Now we check the fulfillment of the conditions in Remark 8.5.4. Denote \( \phi_{1_k} \) the linear interpolant of a function \( \phi \) on the triangulation \( T_k \).

At first, let \( K \) be given by (8.5.33). For an arbitrary \( v^k \in K_k \) take \( z^k(v^k) := v^k + g - g^k \), where \( g^k := g_{1_k} \). Obviously, \( z^k(v^k) \in K \). From Theorem 3.2.1 in [75], already for \( g \in H^2(\Omega) \), the estimate

\[
\| g - g^k \| \leq \tilde{c}\| g \|_{H^2(\Omega)} h_k
\]
356  CHAPTER 8. PPR FOR VARIATIONAL INEQUALITIES

holds with \( \overline{c} \) independent of \( g, h, k \) and \( T_k \). Hence,

\[
\| z_k(v_k) - v_k \| \leq \overline{c} \| g \|_{H^2(\Omega)} h_k,
\]

i.e. condition (a) in Remark 8.5.4 is guaranteed with \( c_3 = \overline{c} \| g \|_{H^2(\Omega)} \).

Now, let \( K \) be given by (8.5.27). In this case, for an arbitrary \( v_k \in K_k \), take

\[
z_k(v_k) := \max \left[ v_k + g - g_k, \psi \right].
\]

Then \( z_k(v_k) \in K \), and with \( g_k := g_{I_k}, \psi_k := \psi_{I_k} \) the relation

\[
g - g_k \leq z_k(v_k) - v_k \leq \max \left[ g - g_k, \psi - \psi_k \right]
\]

holds on \( \overline{\Omega} \). If \( g - g_k \geq \psi - \psi_k \) on \( \Omega \) (in particular, if \( g - \psi \) is a concave function), then again \( z_k(v_k) := v_k + g - g_k \) holds true and condition (a) is also fulfilled.

Otherwise, we are able to prove only the weaker condition (b) in Remark 8.5.4. Indeed, Green’s formula yields

\[
\langle Q(u^*), z_k(v_k) - v_k \rangle = - \int_\Omega \text{div} \frac{\nabla u^*}{\sqrt{1 + |\nabla u^*|^2}} (z_k(v_k) - v_k) dx + \int_{\Gamma} \frac{\partial u^*}{\partial n} \frac{1}{\sqrt{1 + |\nabla u^*|^2}} (z_k(v_k) - v_k) d\Gamma,
\]

where \( \frac{\partial}{\partial n} \) denotes the normal derivative on \( \Gamma \).

Assuming that \( g \in C^2(\overline{\Omega}), \psi \in C^2(\overline{\Omega}) \), Theorem 3.1 in [382] provides the estimates

\[
\| g - g_k \|_{C(\overline{\Omega})} \leq c(g) h_k^2, \quad \| \psi - \psi_k \|_{C(\overline{\Omega})} \leq c(\psi) h_k^2.
\]

From (8.5.43)-(8.5.45) we conclude the fulfillment of the first part of condition (b) in Remark 8.5.4. The second part follows from the proof of Theorem II.2.3 in [134].

Now, we are ready to apply the convergence results from Section 8.5.2. Let us recall that \( \Omega \) is a convex polygonal domain, the sets \( K_k \) are described by (8.5.29) (with \( \Gamma_1 = \Gamma \)) or (8.5.30) and the functions \( g, \psi \) are sufficiently smooth. The following statement is an immediate corollary of the Theorems 8.5.10 and 8.5.13.

8.5.18 Theorem.

(i) The problems \{\( (P_k) \)\}, corresponding to \( \text{VI}(Q, K) \) (with \( Q \) defined by (8.5.32) and \( K \) by (8.5.33) or (8.5.27)) are solvable.

(ii) Let the controlling sequences of Method 8.5.2 satisfy the conditions (9.3.1), (9.3.2) and \( \sum_{k=1}^{\infty} \frac{\delta_k}{\chi_k} < \infty \).

a) If the solution \( u^* \) belongs to \( H^2(\Omega) \), then \( u_k \rightarrow u^* \) in \( V \).

b) If assumption (a) from Remark 8.5.4 is valid\(^3\) and \( u^* \) belongs to \( H^2(\Omega) \cap W^{1,\infty}(\Omega) \), then \( \lim_{k \rightarrow \infty} \| u_k - u^* \|_{W^{1,\infty}(\Omega)} = 0 \).

\(^3\)For instance if the set \( K \) is given by (8.5.33).
8.5. ELLIPTIC REGULARIZATION

8.5.19 Remark. A quite similar analysis can be performed for Method 8.5.2 applied to the variational formulation of the Dirichlet problem for the equation of prescribed mean curvature. Here the operator $Q$ is defined by

$$\langle Q(u,v) \rangle = \int_{\Omega} \left( \frac{\nabla u \cdot \nabla v}{\sqrt{1 + |\nabla u|^2}} - pv \right) dx, \quad \forall u, v \in V$$

(cf. with (8.5.32)). Under the assumption that $u^* \in H^2(\Omega) \cap W^{1,\infty}(\Omega)$, we also obtain

$$\lim_{k \to \infty} \|u^k - u^*\|_{W^{1,1}(\Omega)} = 0.$$  

8.5.4 Application to convection-diffusion problems

These problems arise in many areas such as the transport and diffusion of pollutants, simulation of oil extraction from underground reservoirs, heat transport problems in the convection-dominated case, etc.

8.5.4.1 Formulations and properties of the problem

Again, let $\Omega$ be an open domain in $\mathbb{R}^2$ with a Lipschitz continuous boundary $\Gamma$, which now is divided into disjoint connected pieces $\Gamma_1$ and $\Gamma_2$, meas $\Gamma_1 > 0$ ($\Gamma_2 = \emptyset$ is not excluded). We consider the convection-diffusion problem

$$-\epsilon \Delta u + b \cdot \nabla u + cu = f \quad \text{on } \Omega$$  

(8.5.46)

with boundary conditions

$$u = g_1 \quad \text{on } \Gamma_1, \quad \frac{\partial u}{\partial n} = g_2 \quad \text{on } \Gamma_2.$$  

(8.5.47)

The functions $b = (b_1, b_2)$, $c$ and $f$ are supposed to be sufficiently smooth on $\Omega$, $g_1, g_2 \in H^2(\Omega)$; $c \geq 0$ holds on $\Omega$ and $\epsilon$ is a small positive constant such that

$$0 < \epsilon << \|b\|_{L^\infty(\Omega)}^2.$$  

(8.5.48)

The unknown function $u$ may represent the concentration of a pollutant being transported along a stream moving at velocity $b$ and also subject to diffusive effects. Alternatively, $u$ may represent the temperature of a fluid moving along a heated wall. Relation (9.1.17) corresponds to the situation that the diffusion is a less significant physical effect than the convection. For instance, on a windy day a pollutant moves fast in the direction of the wind, whereas its spreading due to molecular diffusion remains small.

Condition (9.1.17) causes a so-called boundary layer: a fast variation of the gradient of the solution near a part of the boundary. Such problems are called singularly perturbed. This peculiarity is illustrated in the following slightly modified example from [102].

8.5.20 Example. The equation

$$-\epsilon \Delta u + \frac{\partial u}{\partial x_2} = 0 \quad \text{on } \Omega := (0,1) \times (-1,1)$$

on
is considered subject to Dirichlet boundary conditions
\[ u(x_1, -1) = x_1, \quad u(x_1, 1) = 0, \]
\[ u(0, x_2) = 0, \quad u(1, x_2) = \frac{1 - \exp((x_2 - 1)/\epsilon)}{1 - \exp(-2/\epsilon)}, \]
(i.e. \( u(1, x_2) \in [0, 1] \forall x_2 \in (-1, 1) \)).

The unique solution of this problem is
\[ u(x_1, x_2) = x_1 \frac{1 - \exp((x_2 - 1)/\epsilon)}{1 - \exp(-2/\epsilon)}. \] (8.5.49)

![Figure 8.5.15: Graph of solution (8.5.49) for \( \epsilon = 0.5 \) and \( \epsilon = 0.005 \)]

One can easily see that, for small \( \epsilon \), this solution is very close to the function \( x_1 \) except near the boundary part \( x_2 = 1 \). But,
\[ \frac{\partial u(x)}{\partial x_2} = -x_1 \frac{\exp((x_2 - 1)/\epsilon)}{1 - \exp(-2/\epsilon)} \frac{1}{\epsilon}, \]
and for any \( x_1 > 0, \ d > 0 \)
\[ \lim_{\epsilon \to 0} \frac{\partial u(x)}{\partial x_2} = -\infty \text{ if } x_2 = 1 - d \epsilon. \]

In general, in the most part of the domain the solution of problem (8.5.46), (8.5.47) is close to the solution of the reduced (hyperbolic) equation
\[ b \cdot \nabla u + cu = f \] (8.5.50)
with appropriate boundary conditions. If \( \Gamma_- \subset \Gamma_1 \), where \( \Gamma_- = \{x \in \Gamma : b \cdot n < 0\} \) is a so-called inflow boundary (\( n \) denotes the outward-pointed unit vector normal to \( \Gamma \)), then this boundary condition is
\[ u = g_1 \text{ on } \Gamma_- . \] (8.5.51)

Boundary layers arise near an outflow boundary \( \Gamma_+ = \{x \in \Gamma : b \cdot n > 0\} \) and a characteristic boundary \( \Gamma_0 = \{x \in \Gamma : b \cdot n = 0\} \), where the solutions of the
problems (8.5.46), (8.5.47) and (8.5.50), (8.5.51) can differ significantly, and the boundary layer functions (this is $x_1 \exp((x_2 - 1)/\epsilon)$ in our example) characterize approximately the difference between these solutions.

The presence of boundary layers causes serious difficulties for the applications of discretization techniques (finite-difference- and finite element methods) to convection-diffusion problems. There are numerous publications dealing with special discretization procedures and special algorithms for solving discretized convection-diffusion problems (see [37], [102], [355], and references therein).

Introducing the space $V = \{ v \in H^1(\Omega) : u|_{\Gamma_1} = 0 \}$ with the norm $\| v \| := \| \nabla v \|_{L^2(\Omega)^2}$ (see [75], Theorem 1.2.1 concerning the equivalence of this norm and the standard norm of $H^1(\Omega)$ in case $\text{meas } \Gamma_1 > 0$), one can describe a weak formulation of problem (8.5.46), (8.5.47), i.e., a variational equation of the convection-diffusion problem as follows:

$$\text{find } u \in V \text{ such that } \quad \epsilon \int_{\Omega} \nabla u \cdot \nabla v dx + \int_{\Omega} \left[ (b \cdot \nabla u)v + cuv \right] dx = \int_{\Omega} f v dx + \epsilon \int_{\Gamma_2} g \nu v d\Gamma, \quad \forall \ v \in V,$$

where $\bar{f} := f + \epsilon \Delta g_1 - b \cdot \nabla g_1 - cg_1$, $\bar{g} := g_2 - \frac{\partial g_1}{\partial n}$.

Applying the trace inequality $\| v \|_{L^2(\Gamma)} \leq c_1 \| v \|$, $\forall v \in H^1(\Omega)$ (8.5.53) (see, for instance, [75], Section 1.2) to estimate the term $\epsilon \int_{\Gamma_2} \bar{g} \nu v d\Gamma$, the continuity of the functional

$v \mapsto \int_{\Omega} f v dx + \epsilon \int_{\Gamma_2} \bar{g} \nu v d\Gamma$

in the space $V$ can be easily concluded, hence

$$\exists \ l \in V' : \quad (l, v) = \int_{\Omega} f v dx + \epsilon \int_{\Gamma_2} \bar{g} \nu v d\Gamma, \quad \forall \ v \in V. \quad (8.5.54)$$

The estimate

$$\left| \int_{\Omega} (b \cdot \nabla u)v dx \right| \leq d \sup_{x \in \Omega} |b(x)| \| u \||v||$$

(with $d : \| v \|_{L^2(\Omega)} \leq d \| v \| \forall \ v \in V$) is proved by using twice the Cauchy-Schwarz inequality. Now the continuity of the bilinear form in (8.5.52) follows in a standard way. Thus, according to the Riesz theorem A1.1.1, there exists an operator $A \subset L(V, V')$ such that

$$\langle Au, v \rangle = \epsilon \int_{\Omega} \nabla u \cdot \nabla v dx + \int_{\Omega} (b \cdot \nabla u)v dx + \int_{\Omega} cuv dx. \quad (8.5.55)$$

Its strong monotonicity can be shown under the additional assumption that

$$c - \frac{1}{2} \text{div } b \geq 0 \text{ on } \Omega, \quad \text{and } \Gamma_2 \subset \Gamma_+ ,$$
which we will suppose in the sequel. Indeed, applying Green’s formula to
\[ \alpha(u, v) := \int_{\Omega} (b \cdot \nabla u)v \, dx \]
we obtain
\[ \alpha(u, v) = -\int_{\Omega} u \text{div}(vb) \, dx + \int_{\Gamma_2} uv(b \cdot n) \, d\Gamma \]
\[ = -\int_{\Omega} uv \text{div}bdx - \int_{\Omega} (b \cdot \nabla v)udx + \int_{\Gamma_2} uv(b \cdot n) \, d\Gamma \]
Thus
\[ \alpha(u, u) = -\frac{1}{2} \int_{\Omega} u^2 \text{div}bdx + \frac{1}{2} \int_{\Gamma_2} u^2 (b \cdot n) \, d\Gamma, \]
and \( b \cdot n > 0 \) in \( \Gamma_2 \) holds because of \( \Gamma_2 \subset \Gamma_+ \). Now
\[ \alpha(u, u) + \int_{\Omega} cu^2 \, dx \geq \int_{\Omega} \left( c - \frac{1}{2} \text{div}b \right) u^2 \, dx \geq 0, \]
and according to (8.5.55)
\[ \langle Au, u \rangle \geq \epsilon \|u\|^2, \quad \forall \ u \in V. \quad (8.5.56) \]

From the monotonicity and continuity of the operator \( A \) it follows that \( A \)

is maximal monotone, and together with (8.5.56) this guarantees the existence of a unique \( u^* \in V \) such that
\[ \langle Au^* - l, v \rangle = 0, \quad \forall \ v \in V. \]

Because \( \alpha(u, v) \neq \alpha(v, u) \), the operator is not symmetric, hence problem (8.5.46),
(8.5.47) cannot be transformed - at least not in a natural way - into an optimization problem.

Conditions on the data of problem (8.5.46), (8.5.47), which provide \( u^* \in H^2(\Omega) \), can be found in [148], [253]. In particular, \( u^* \in H^2(\Omega) \) holds in the case that \( \Omega \) is a convex polygonal domain, \( \Gamma_1 = \Gamma \), functions \( b, c, g_1 \) are sufficiently smooth, and \( f \in L^2(\Omega) \) (see [253], Theorem III.9.1 and Remark III.9.4).

8.5.4.2 Application of generalized IPR-method to convection-diffusion problems

Applying Method 8.5.2 with \( \rho : v \mapsto \|v\|^2 \) and an appropriate parameter sequence \( \{\chi_k\} \), we approximate the singularly perturbed elliptic problem (8.5.46),
(8.5.47) by a sequence of problems with unperturbed elliptic operators. Remind,
this cannot be carried out by means of the classical approach of elliptic regulariza-
tion. On this way, the boundary layers (also inner layers if exist; see [355] for this notion) will be accumulated gradually, because of the term \( -\chi_k \nabla g(u^k) \)
in the operator of problem \((P^k)\).
In particular, for problem (8.5.46), (8.5.47) with \( \Gamma_1 = \Gamma, g_1 \equiv 0, \) the exact problem \((P^k)\) (with \( \delta_k = 0 \)) consists in the finding of a weak solution of the equation

\[-(\epsilon + 2\chi_k)\Delta u + b \cdot \nabla u + cu = f + 2\chi_k\Delta u^k, \quad \text{in } H^1_0(\Omega).\]

The gradual accumulation of boundary- and inner layers allows us a more successful application of standard finite element methods, and with \( \epsilon + 2\chi_k \) in place of \( \epsilon \), we obtain a better stability and conditioning of the discretized problems.

8.5.21 Remark. In [355], authors analyze situations where boundary- and inner layer functions can be defined \textit{a priori} - sometimes in explicit form - by using a standard technique from the singular perturbation theory. If such functions are known (exact or approximately), they can be used to choose a starting point in Method 8.5.2 or to correct an approximate solution after certain number of iterations.

Now, we examine the application of the convergence results from Section 8.5.2 to Method 8.5.2 for solving the convection-diffusion problem in the variational formulation (8.5.52). As in the previous case, it is supposed that \( \Omega \) is a convex polygonal set and that the solution \( u^* \) belongs to \( H^2(\Omega) \). So, we deal with auxiliary problems \((P^k)\) in the space \( V = \{v \in H^1(\Omega) : u|_{\Gamma_1} = 0\} \), in which

\[ Q : v \mapsto Av - l, \quad \varphi_k \equiv 0, \quad \rho : v \mapsto \|v\|^2 \]

and \( K_k \) are given by (8.5.29).

Obviously, in this case \( K_k \subset K := V \), Assumption 8.5.3(i) and (9.3.20) are satisfied. Assumption 8.5.1(ii) is valid with the operator \( B := -\epsilon \Delta \), and in Assumption 8.5.3(ii) one can take \( \tilde{\chi} := 0 \).

The relation

\[ \forall w \in K, \quad \forall k, \exists w^k \in K_k : \lim_{k \to \infty} \|w - w^k\| = 0 \]

follows immediately from the proof of Theorem 3.3 in [182], Section 1.1., and therefore, the continuity of the operator \( A \) implies the validity of Assumption 8.5.3(iii).

Next, because \( u^* \in H^2(\Omega) \), the estimate

\[ \|u^* - u^*_{I_k}\| \leq c\|u^*\|_{H^2(\Omega)}h_k, \quad \forall k \]

is known, and taking into account that \( u^*_{I_k} \in K_k \) and

\[ \|Q(u^*) - Q(u^*_{I_k})\| \leq \|A\|_{V'}\|u^* - u^*_{I_k}\|, \]

Assumption 8.5.3(v) is valid with \( \sigma_k := h_k \) if \( \sum_{k=1}^{\infty} \frac{h_k}{\chi_k} < \infty \).

Finally, since the operator \( B = -\epsilon \Delta \) is strongly monotone on \( V \), condition (iii) in Theorem 8.5.10 holds with \( \tilde{B} := 0 \).

Therefore, choosing the controlling parameters according to (9.3.1), (9.3.2) and \( \sum_{k=1}^{\infty} \frac{\delta_k}{\chi_k} < \infty \), one can use Theorem 8.5.10, which guarantees that the iterates \( u^k \) of Method 8.5.2 converge to \( u^* \) strongly in \( V \).
8.6 Comments

Section 8.1: Various approaches to the solution of the problem with an obstacle on the boundary, including results on the convergence and the rate of convergence of finite element approximations, can be found in the extended edition of the monograph Glowinski, Lions and Trémolières [135] and in the papers of Haslinger [168], Hlaváček [180, 181] and Scarpini and Vivaldi [360].

For coercive problems of this type see, for instance, Lions [271], Cea [69], Hlaváček et al. [182].

Numerical methods for elliptic variational inequalities with finite element approximations of the dual problem are considered in [135] and also in the papers of Fremond [119] and Haslinger [169].

Mixed variational formulations are investigated in the framework of the duality theory for variational inequalities, developed by Cea [69], and Ekeland and Temam [100]. The general approach to mixed methods (see Brezzi and Fortin [54]) is based, in particular, on investigations of Aubin [19], Babuška and Gatica [27], Lapin [254] and Wang and Yang [415].

Section 8.2: For a more detailed analysis of the two-body contact problem see Hlaváček et al. [182]. The Signorini problem was studied by Fichera [114]. Concerning the finite element approximation of the Signorini Problem in case of a curved boundary, we refer to Haslinger [169].

The results in the Subsections 8.2.3-8.2.5 are now published in [219, 224]. For the analysis of the coercivity of the auxiliary problems in the methods with regularization on the subspace a technique is used, which has been developed by Fichera [114] and Panagiotopoulos [316].

Theorem 8.2.19 improves a corresponding statement of the paper [212].

Section 8.3: The basic algorithm of alternating iterations has been studied by Panagiotopoulos [315] for the discrete version of the Signorini problem with friction. In our description of the algorithm we follow Hlaváček et al. [182]. The main results about the convergence of the MSR-method with weak regularization for the non-smooth Problem II (Theorem 8.3.3) is new.

Section 8.4: In connection with the application of IPR-methods to well-posed problems we emphasize that in Rockafellar’s paper [352] the rate of convergence of the proximal-point-algorithm for an abstract variational inequality is studied under conditions, which guarantee the well-posedness of the problem in a special sense (see the comments to Section 5.2).

Section 8.5: Typically, standard discretization methods in mathematical physics are not efficient when applied to degenerate and singularly perturbed elliptic variational inequalities, and there are numerous investigations addressed to the creation of special discretization procedures, for instance finite element methods with upwinding, streamline diffusion finite element methods, etc. [37, 102, 355]. Also special algorithms for solving the arising discretized problems are needed.

In [231], we develop a quite different idea, which may be presented as follows: Using the proximal regularization, the original variational inequality is approximated by a sequence of uniformly elliptic problems, which can be treated with standard finite element techniques and standard solvers. Moreover, only a
single discretization is used for each regularized problem, with a mild rule for decreasing of the triangulation parameter in the outer process.

In various schemes of proximal point methods the conditions on data approximation are of Mosco’s type with order $\sigma > 0$ (see Subsection 2.4.2 as well as [260, 5]), or outer approximations of the set $K$ are used [226]. These conditions are certainly not suitable if we deal with problems in mathematical physics and use finite element or finite-difference methods. Indeed, in this case $K^k \not\supset K$, and for an arbitrary element $u \in K$ at best the relation

$$\lim_{k \to \infty} \min_{v \in K^k} \|u - v\| = 0$$

can be concluded (without any estimate for the rate of convergence), i.e., the Mosco convergence with order $\sigma > 0$ cannot be guaranteed.

Abstract assumptions admitting the described peculiarity were introduced first in [224]. Here we deal with a weaker form of these assumptions.

Growth conditions used so far to obtain a more qualitative convergence of proximal point methods (see for instance [352, 281, 225]) are not fulfilled in the case of the operator $Q$ in the minimal surface problem. Therefore, the known convergence results provide only weak convergence in $H^1(\Omega)$ of the proximal point methods when applied to this problem or related variational inequalities.

We conclude with the remark that an application of the elliptic proximal regularization method (without discretization) for solving parabolic variational inequalities was studied in [225].
Chapter 9

PPR FOR VI’s WITH MULTI-VALUED OPERATORS

For a given set-valued operator $Q : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ and a closed convex subset $K$ of $\mathbb{R}^n$ the following variational inequality is considered:

$$\text{VI}(Q, K) \quad \text{find a pair } x^* \in K \text{ and } q^* \in Q(x^*) \text{ such that}$$

$$\langle q^*, x - x^* \rangle \geq 0 \quad \forall \ x \in K.$$

As before we make use of the notations and definitions commonly used in convex analysis, see for instance Rockafellar [353]. Concerning properties of VI’s with set-valued maximal monotone operators we also refer to Subsection A1.6.

It is well-known that $\text{VI}(Q, K)$ is closely related to the problem of finding a zero of a maximal monotone operator. Indeed, if the operator $Q$ is maximal monotone and $\text{ri}(\text{dom}(Q)) \cap \text{ri}(K) \neq \emptyset$, then $\text{VI}(Q, K)$ is equivalent to the inclusion

$$\text{IP}(T) \quad 0 \in T(x^*),$$

where $T = Q + N_K$ and $N_K$ denotes the normality operator of $K$.

9.1 PPR with Relaxation

The exact PPR for $\text{IP}(T)$ (with an arbitrary maximal monotone $T$) is a fixed point iteration as follows: For any $x^0 \in \mathbb{R}^n$ the sequence $\{x^k\}$ is generated by the iteration

$$x^{k+1} := J_{\chi_k T}(x^k), \quad k = 0, 1, 2, \ldots,$$

where $\{\chi_k\} \subset (0, +\infty)$ is a sequence of regularization parameters and

$$J_{\chi_k T} = (I + \chi_k T)^{-1}$$

is the resolvent operator associated with $T$. Due to Minty’s theorem [290], this operator is single-valued.
CHAPTER 9. PPR FOR VI’S WITH MULTI-VALUED OPERATORS

Gol’shtein and Tret’yakov [140] introduced a sequence of relaxation parameters \( \{ \rho_k \} \subset (0, 2) \) and proved the convergence of the relaxed proximal point algorithm:

\[
x^{k+1} := (1 - \rho_k)x^k + \rho_k J_{\chi_k}(x^k), \quad k = 0, 1, 2, \ldots,
\]

for \( \chi_k \equiv \chi > 0 \). Considering the same scheme, Eckstein and Bertsekas [98] allow the regularization parameter to vary with \( k \). In the terminology of relaxation algorithms the choices \( \rho_k \in (0, 1) \) are usually referred to as under-relaxation and \( \rho_k \in (1, 2) \) as over-relaxation, respectively.

Usually, the main goal for introducing a relaxation step in PPR is the acceleration of the convergence. However, in general one cannot expect that a relaxation step leads always to a reduction of the iteration number as the following example shows.

9.1.1 Example. For the rotation operator \( \mathcal{T} : \mathbb{R}^2 \to \mathbb{R}^2 \),

\[
\mathcal{T}(x_1, x_2) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},
\]

which is monotone, continuous (and therefore maximal monotone), the relaxed proximal point algorithm generates the following sequence \( \{x^k\} \):

\[
\begin{bmatrix} x_1^{k+1} \\ x_2^{k+1} \end{bmatrix} := \frac{1}{1 + \chi_k} \begin{bmatrix} (1 - \rho_k)\chi_k^2 & \chi_k \rho_k \\ -\chi_k \rho_k & (1 - \rho_k)\chi_k^2 \end{bmatrix} \begin{bmatrix} x_1^k \\ x_2^k \end{bmatrix}, \quad k = 0, 1, 2, \ldots.
\]

The eigenvalues of the iteration matrix depend on the relaxation parameter \( \rho_k \) and can be evaluated explicitly:

\[
\begin{align*}
\mu_1(\rho_k) &= 1 - \frac{\rho_k \chi_k^2}{1 + \chi_k} - \frac{\rho_k \chi_k}{1 + \chi_k} \sqrt{-1}, \\
\mu_2(\rho_k) &= 1 - \frac{\rho_k \chi_k^2}{1 + \chi_k} + \frac{\rho_k \chi_k}{1 + \chi_k} \sqrt{-1},
\end{align*}
\]

with

\[
|\mu_1(\rho_k)|^2 = |\mu_2(\rho_k)|^2 = 1 - \frac{\rho_k \chi_k^2 (2 - \rho_k)}{1 + \chi_k^2}.
\]

Minimizing the norm of the eigenvalues with respect to the relaxation parameter, one obtains

\[
\rho_k^* = \arg \min_{\rho_k} |\mu_1(\rho_k)|^2 = \arg \min_{\rho_k} |\mu_2(\rho_k)|^2 = 1.
\]

The last two equations indicate two interesting facts. The first equation shows that the lower and upper bounds for the relaxation parameter \( \rho_k \) are sharp. The choice \( \rho_k = 0 \) or \( \rho_k = 2 \) leads to \( |\mu_1(\rho_k)| = |\mu_2(\rho_k)| = 1 \) and hence to the divergence of the relaxed PPR. The second equation says that the best choice of \( \rho_k \) is \( \rho_k^* \equiv 1 \). The case \( \rho_k \neq 1 \) leads to an increasing of iteration steps in both cases - under-relaxation as well as over-relaxation.

Hence, there is no universally valid recommendation for the choice of the sequence of relaxation parameters \( \{\rho_k\} \).
The evaluation of the resolvent operator $J_{\chi T}$ requires the most numerical effort in one iteration of PPA. The exact evaluation of $J_{\chi T}$ is equivalent to the solution of a nontrivial auxiliary problem and might be very expensive in practice. Eckstein and Bertsekas [98] suggest to use Rockafellar’s error tolerance criterion which permits to work with approximate solutions of the auxiliary problems in the following manner:

$$x^{k+1} := (1 - \rho_k)x^k + \rho_k y^k, \quad k = 0, 1, 2, \ldots,$$

where

$$\|y^k - J_{\chi T}(x^k)\| \leq \delta_k$$

and $\{\delta_k\} \subset [0, +\infty)$ is a sequence of error tolerance parameters satisfying

$$\sum_{k=0}^{\infty} \delta_k < \infty.$$  

Some other error tolerance criteria for PPA can be found in [225] and [375]. However, from the numerical point of view the evaluation of the resolvent operator $J_{\chi T}$ is always a difficult task. It requires to invert the operator $I + \chi T$, what depends on the nature of $T$ and implies information about the whole image of the operator.

### 9.1.1 Relaxed PPR with enlargements of maximal monotone operators

Burachik et. al. [61] try to overcome this difficulty and introduce an outer approximation, the so-called $\epsilon$-enlargement $T^\epsilon$ of the operator $T$.

To start, we repeat briefly the definition and the main properties of the $\epsilon$-enlargements of monotone operators. For a given monotone operator $F: \mathbb{R}^n \to 2^{\mathbb{R}^n}$ and $\epsilon \geq 0$ the $\epsilon$-enlargement $F^\epsilon: \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is defined as follows

$$F^\epsilon(x) = \{u \in \mathbb{R}^n : \langle u - v, x - y \rangle \geq -\epsilon \quad \forall \ (y, v) \in \text{gph}(F)\}.$$

**9.1.2 Proposition.** [61], (Propositions 1, 2)

Let $F: \mathbb{R}^n \to 2^{\mathbb{R}^n}$ be a monotone operator, then

(a) $\text{gph}(F) \subseteq \text{gph}(F^\epsilon_1) \subseteq \text{gph}(F^\epsilon_2)$ for any $\epsilon_2 \geq \epsilon_1 \geq 0$.

If additionally $F$ is maximal monotone, then

(b) $\text{gph}(F) = \text{gph}(F^0)$.

(c) If $\text{dom}(F)$ is closed, then $\text{dom}(F) = \text{dom}(F^\epsilon)$ for all $\epsilon \geq 0$.

Using the error tolerance criterion introduced in [225], we suggest the following version of the relaxed proximal point algorithm for solving VI($Q, K$):

### 9.1.3 Algorithm. (Relaxed Proximal Point Algorithm) (RPPA)

**S0.** Choose any $x^0 \in \mathbb{R}^n$. Set $k := 0$.

**S1.** If $x^k$ is a solution of VI($Q, K$): STOP.

**S2.** Choose some $\chi_k > 0$, $\delta_k \geq 0$, $\epsilon_k \geq 0$ and find $y^{k+1} \in K$, $q^{k+1} \in Q^\epsilon(x^{k+1})$, such that

$$\langle \chi_k q^{k+1} + y^{k+1} - x^k, y - y^{k+1} \rangle \geq -\delta_k \|y - y^{k+1}\| \quad \forall \ y \in K. \quad (9.1.1)$$
Choose \( \rho_k \in \left( 0, \frac{2}{1 + 2\delta_k} \right) \) and set
\[
x^{k+1} := (1 - \rho_k)x^k + \rho_k y^{k+1}.
\]
(9.1.2)

Set \( k := k + 1 \) and go to S1.

It should be noted that a solution of the auxiliary problem (9.1.1) can be found similarly to Eckstein [97]: Find \( y^{k+1} \in \mathbb{R}^n \) and an error vector \( e^{k+1} \), such that
\[
e^{k+1} \in \chi_k Q^s(y^{k+1}) + y^{k+1} - x^k + N_K(y^{k+1}),
\]
(9.1.3)
\[\|e^{k+1}\| \leq \delta_k.\]
(9.1.4)

Indeed, if \( y^{k+1} \) and \( e^{k+1} \) satisfy (9.1.3)-(9.1.4), then \( y^{k+1} \in K \) and the definition of the normality operator \( N_K \) yields
\[
\langle \chi_k q^{k+1} + y^{k+1} - x^k, y - y^{k+1} \rangle \geq -\langle e^{k+1}, y - y^{k+1} \rangle \quad \forall y \in K
\]
for some \( q^{k+1} \in Q^s(y^{k+1}) \). Owing to the Cauchy–Schwarz inequality we can conclude that \( (y^{k+1}, q^{k+1}) \) satisfy (9.1.1).

9.1.4 Proposition. Let \( K \subseteq \mathbb{R}^n \) be a closed, convex set, \( Q : \mathbb{R}^n \to 2^{\mathbb{R}^n} \) a maximal monotone operator with closed effective domain and \( \text{ri}(\text{dom}(Q)) \cap \text{ri}(K) \neq \emptyset \). Then Algorithm 9.1.3 is well defined, i.e. for each \( k \in \mathbb{N} \) and any \( \chi_k > 0, \epsilon_k \geq 0, \delta_k \geq 0 \) there exists a solution of the auxiliary problem (9.1.1).

Proof: Because the restriction set \( K \) is closed and convex, the normality operator \( N_K \) is maximal monotone. Besides, due to [350], Theorem 2, the operator
\[
y \mapsto \chi_k Q(y) + y - x^k + N_K(y)
\]
is maximal monotone and strongly monotone for \( \chi_k > 0 \). Using [353], Proposition 12.54, one can infer that the auxiliary problem (9.1.1) has a unique solution for \( \epsilon_k = \delta_k = 0 \), i.e. there exist \( \bar{y} \in \mathbb{R}^n \), such that
\[
0 \in \chi_k Q(\bar{y}) + \bar{y} - x^k + N_K(\bar{y}).
\]
Due to Proposition 9.1.2, the last relation yields immediately for any \( \epsilon_k \geq 0 \) that is
\[
0 \in \chi_k Q^s(\bar{y}) + \bar{y} - x^k + N_K(\bar{y}),
\]
and thus \( y^{k+1} := \bar{y} \) satisfies (9.1.1) for any \( \delta_k \geq 0 \).

9.1.5 Proposition. Assume that the solution set of VI(\( Q, K \)) is nonempty and the sequences of parameters in Algorithm 9.1.3 are chosen such that
\[
0 < \chi_k \leq \bar{\chi} < \infty, \quad 0 < \rho_k \leq \rho \leq \frac{2}{1 + 2\delta_k} \quad \forall k \in \mathbb{N},
\]
\[
\delta_k \geq 0, \quad \epsilon_k \geq 0 \quad \forall k \in \mathbb{N}, \quad \sum_{k=0}^{\infty} \delta_k < \infty, \quad \sum_{k=0}^{\infty} \epsilon_k < \infty,
\]
for some \( \sigma \in (0, 1) \). Then it holds:
(i) sequence \( \{x^k\} \) generated by Algorithm 9.1.3 is bounded;
(ii) sequence \( \{\|x^k - x^*\|\} \) converges for any solution \( x^* \) of \( \text{VI}(Q,K) \)
(iii) and it holds
\[
\lim_{k \to \infty} \|x^{k+1} - x^k\| = 0. \tag{9.1.5}
\]

**Proof:** Let \( x^* \in K \) be a solution of \( \text{VI}(Q,K) \). Then there exists \( q^* \in Q(x^*) \), such that
\[
\langle q^*, y - x^* \rangle \geq 0 \quad \forall y \in K. \tag{9.1.6}
\]
For \( y = y^{k+1} \) in (9.1.6) and \( y = x^* \) in (9.1.1) one gets
\[
\frac{1}{2}\|x^k - x^*\|^2 - \frac{1}{2}\|x^{k+1} - x^*\|^2 = \frac{1}{2}\|x^{k+1} - x^k\|^2 + \langle x^{k+1} - x^k, x^* - x^{k+1} \rangle
\]
\[
\geq \frac{1}{2}\|x^{k+1} - x^k\|^2 + \langle x^{k+1} - x^k, x^* - x^{k+1} \rangle
\]
\[
+ \rho_k \langle \chi_k q^{k+1} + y^{k+1} - x^k, y^{k+1} - x^* \rangle
\]
\[
- \delta_k \|x^* - y^{k+1}\| \rangle
\]
\[
- \rho_k \chi_k (q^*, y^{k+1} - x^*).\]

Relation (9.1.2) leads to
\[
y^{k+1} = x^k + \frac{1}{\rho_k} (x^{k+1} - x^k)
\]
and, due to \( q^{k+1} \in Q(x^k) (y^{k+1}) \), one can continue and conclude that
\[
\frac{1}{2}\|x^k - x^*\|^2 - \frac{1}{2}\|x^{k+1} - x^*\|^2
\]
\[
\geq \frac{1}{2}\|x^{k+1} - x^k\|^2 + \langle x^{k+1} - x^k, x^* - x^{k+1} \rangle
\]
\[
+ \rho_k \chi_k (q^{k+1} - q^*, y^{k+1} - x^*)
\]
\[
+ \langle x^{k+1} - x^k, x^k + \frac{1}{\rho_k} (x^{k+1} - x^k) - x^* \rangle
\]
\[
- \rho_k \delta_k \|x^* - x^k - \frac{1}{\rho_k} (x^{k+1} - x^k)\|
\]
\[
\geq \left( \frac{1}{\rho_k} - \frac{1}{2} \right) \|x^{k+1} - x^k\|^2 - \rho_k \chi_k \epsilon_k
\]
\[
- \rho_k \delta_k \left( \|x^* - x^k\| + \frac{1}{\rho_k} \|x^{k+1} - x^k\| \right)
\]
\[
\geq \left( \frac{1}{\rho_k} - \frac{1}{2} \right) \|x^{k+1} - x^k\|^2 - \rho_k \chi_k \epsilon_k
\]
\[
- \rho_k \delta_k \left( \|x^* - x^k\|^2 + \frac{1}{\rho_k} \|x^{k+1} - x^k\|^2 + 1 + \frac{1}{\rho_k} \right).
\]
The last inequality follows from \( r \leq r^2 + 1 \ \forall \ r \in \mathbb{R} \). Rearranging the terms it yields
\[
\frac{1}{2}\|x^{k+1} - x^*\|^2 \leq (1 + 2\delta_k \rho_k) \frac{1}{2}\|x^k - x^*\|^2
\]
\[
- \left( \frac{1}{\rho_k} - \frac{1}{2} - \delta_k \right) \|x^{k+1} - x^k\|^2 + \rho_k \chi_k \epsilon_k + \delta_k (\rho_k + 1).
\]
If \( \rho_k \leq \frac{2\sigma}{1+2\delta_k} \) \( \forall k \), then
\[
\frac{1}{2} \|x^{k+1} - x^*\|^2 \leq \left( 1 + \frac{4\delta_k \sigma}{1 + 2\delta_k} \right) \frac{1}{2} \|x^k - x^*\|^2 - \frac{(1 + 2\delta_k)(1 - \sigma)}{2\sigma} \|x^{k+1} - x^k\|^2 + \frac{2\sigma \chi_k \epsilon_k}{1 + 2\delta_k} + \delta_k 3 + 2\delta_k. \tag{9.1.7}
\]

According to \( \delta_k \geq 0 \) for all \( k \in \mathbb{N} \) and \( \sigma \in (0,1) \), one can finally conclude that
\[
\|x^{k+1} - x^*\|^2 \leq (1 + 4\delta_k) \|x^k - x^*\|^2 + 4\chi_k \epsilon_k + 2\delta_k (3 + 2\delta_k).
\]

Due to the assumptions on the sequences \( \{\chi_k\}, \{\epsilon_k\} \) and \( \{\delta_k\} \) together with Lemma A3.1.7, the convergence of \( \{\|x^k - x^*\|\} \) is established. Hence, the sequence \( \{x^k\} \) is bounded, moreover in view of (9.1.7) it follows
\[
\|x^{k+1} - x^k\| \to 0 \quad (k \to \infty).
\]

In order to prove convergence of the sequence \( \{x^k\} \) to a solution of \( \text{VI}(Q,K) \), we make use of a gap function, introduced in [59], Lemma 3, Lemma 4: The function \( \psi : \text{dom}(Q) \cap K \to \mathbb{R} \cup \{+\infty\}, \)
\[
\psi(x) = \sup_{y \in K} \sup_{(v,v) \in \text{gph}(Q)} \langle v, x - y \rangle \tag{9.1.8}
\]
is a convex gap function for \( \text{VI}(Q,K) \), i.e. \( \psi(x) \geq 0 \) for all \( x \in \text{dom}(Q) \cap K \) and \( \psi(x^*) = 0 \) if and only if \( x^* \) is a solution of \( \text{VI}(Q,K) \).

**9.1.6 Theorem.** Suppose the assumptions of Propositions 9.1.4 and 9.1.5 are fulfilled, then the sequence \( \{x^k\} \) generated by Algorithm 9.1.3 converges to a solution of \( \text{VI}(Q,K) \).

**Proof:** Proposition 9.1.5, in particular (9.1.5) as well as (9.1.2) together with the assumption on the sequence \( \{\rho_k\} \) guarantee that
\[
\|y^{k+1} - x^k\| \to 0 \quad (k \to \infty). \tag{9.1.9}
\]

According to Proposition 9.1.5, the sequence \( \{x^k\} \) is bounded and therefore there exists a subsequence \( \{x^{k_l}\} \) of \( \{x^k\} \) having a limit point \( \bar{x} \). Hence, from (9.1.9) it follows immediately that
\[
y^{k_l+1} \to \bar{x} \quad (l \to \infty).
\]

Inequality (9.1.1) guarantees that \( y^{k_l+1} \in \text{dom}(Q) \cap K \) for all \( l \in \mathbb{N} \). The effective domain of \( Q \) as well as the restriction set \( K \) are closed and we get \( \bar{x} \in \text{dom}(Q) \cap K \). Now, for any \( l \in \mathbb{N} \) one can conclude from (9.1.1) that
\[
\langle \chi_k q^{k+1} + y^{k+1} - x^{k_l}, y - y^{k+1} \rangle \geq -\delta_k \|y - y^{k+1}\| \quad \forall y \in K,
\]
where \( q^{k+1} \in Q^{*h}((y^{k+1}) \). The definition of the \( \epsilon \)-enlargement yields for any \( y \in K \) and any \( q \in Q(y) \)
\[
\langle \chi_k q + y^{k+1} - x^{k_l}, y - y^{k+1} \rangle \geq -\epsilon_k \chi_k - \delta_k \|y - y^{k+1}\|.
\]
Passing to the limit for \( l \to \infty \) in the last inequality, then (9.1.9) and the assumptions on the sequences \( \{\chi_k\}, \{\epsilon_k\}, \{\delta_k\} \) lead to
\[
\langle q, y - \bar{x} \rangle \geq 0 \quad \forall (y, q) \in \text{gph}(Q), \forall y \in K.
\] (9.1.10)
Using the definition of the gap function \( \psi \), from (9.1.10) one can conclude that
\[
\psi(\bar{x}) = 0,
\]
i.e., \( \bar{x} \) is a solution of \( \text{VI}(Q, K) \). It remains to show that the whole sequence \( \{x^k\} \) converges to \( \bar{x} \). In view of Proposition 9.1.5, the sequence \( \{\|x^* - x^k\|\} \) converges for an arbitrary solution \( x^* \) of \( \text{VI}(Q, K) \). For \( \bar{x} \) in place of \( x^* \) the sequence \( \{\|\bar{x} - x^k\|\} \) converges, too. Moreover, for the subsequence \( \{x^{k_l}\} \) it holds \( \|\bar{x} - x^{k_l}\| \to 0 \) as \( l \to \infty \). Thus, the whole sequence \( \{x^k\} \) converges to the limit point \( \bar{x} \).

9.1.7 Remark. The parameter \( \sigma \), introduced in Proposition 9.1.5, is a result of the proof and it is used to obtain the convergence \( \|x^{k+1} - x^k\| \to 0 \). Parameter \( \sigma \) can be selected arbitrarily from the interval \( (0, 1) \). However, from the numerical point of view it should be chosen closely to 1, which allows us to perform larger steps in the relaxation step (9.1.2) of Algorithm 9.1.3.

The error tolerance parameter \( \delta_k \) has also some influence on the step size in the relaxation step (9.1.2). If the inexact solution \( y^{k+1} \) of auxiliary problem (9.1.1) is coarse, which is usual for iterates far from the solution set, then one has to perform shorter steps. For \( \delta_k = 0 \), i.e. the auxiliary problem is solved exactly, the relaxation parameter \( \rho_k \) can be selected from the interval \( (0, 2) \). This selection corresponds to the classical result of Eckstein and Bertsekas [98]. It should be noted that the value of the error tolerance parameter in the scheme of Eckstein and Bertsekas has no consequences for the relaxation steps.

Although we use an outer approximation of the set-valued operator \( Q \), the auxiliary problem (9.1.1) is still a general variational inequality and not easier to solve than the original problem. In the next subsection we will show how the auxiliary problems in S2 of Algorithm 9.1.3 can be solved numerically for specific operators \( Q \).

9.1.2 Application to non-smooth minimization problems

Here the consideration of convex non-smooth optimization problems serves as an application of Algorithm 9.1.3, because in this case the \( \epsilon \)-enlargements can be computed by bundle methods relatively easy. As far as we know, the numerical treatment of \( \epsilon \)-enlargements for more general multi-valued operators is not an easy task (cf., for instance, [62]).

Suppose that \( K \subset \mathbb{R}^n \) is a closed, convex set and \( f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) is a proper, convex, lower semicontinuous function with \( \text{ri}(\text{dom}(f)) \cap \text{ri}(K) \neq \emptyset \). Then the optimization problem
\[
\text{CP}(f, K) \quad \min_{x \in K} f(x)
\]
is equivalent to the variation inequality \( \text{VI}(\partial f, K) \).

In the sequel we discuss a possible implementation of the Algorithm 9.1.3
for CP\((f, K)\). In particular, we are interested in an algorithmic approach for Step 1 and Step 2.

Using the error tolerance criterion (9.1.3)–(9.1.4) to solve the auxiliary problem (9.1.1) for \(Q = \partial f\), (9.1.1) reduces to:

For given \(\chi_k > 0\), \(\delta_k \geq 0\) and \(\epsilon_k \geq 0\) determine \(y^{k+1} \in K\) and \(e^{k+1} \in \mathbb{R}^n\), such that

\[
e^{k+1} \in \chi_k(\partial f)^e(y^{k+1}) + y^{k+1} - x^k + N_K(y^{k+1}), \tag{9.1.11}
\]

where \(\|e^{k+1}\| \leq \delta_k\).

The most numerical effort in the algorithm is the solving of the auxiliary problem (9.1.11). To choose a suitable solver let us analyze the structure of (9.1.11). Note that for \(\epsilon_k = \delta_k = 0\) the inclusion (9.1.11) is equivalent to the minimization problem

\[
\min_{y \in K} \{f(y) + \frac{1}{2\chi_k}\|y - x^k\|^2\}, \tag{9.1.12}
\]

where the objective function is obviously strongly convex. However, for a non-smooth function \(f\), the latter problem may be hard to solve. For this reason we make use of a bundle method in order to find an approximate solution of (9.1.12) in accordance with (9.1.11).

The suggested bundle strategy consists of two steps: At first replace \(f\) by some convex function such that the modified problem can be treated easier than (9.1.12). Then show that the solution of the resulting problem can be considered as an approximate solution in the sense of (9.1.11).

Assume that, for some fixed \(k \in \mathbb{N}\), a convex function \(f^k_{i+1}\) is constructed for \(i \geq 0\), approximating \(f\) from below, i.e.

\[
f^k_{i+1}(y) \leq f(y) \quad \forall y \in \text{dom}(f). \tag{9.1.13}
\]

The rules defining \(f^k_{i+1}\) will be discussed below. Replacing \(f\) by \(f^k_{i+1}\) in (9.1.12), one gets the convex optimization problem

\[
\min_{y \in K} \{f^k_{i+1}(y) + \frac{1}{2\chi_k}\|y - x^k\|^2\}. \tag{9.1.14}
\]

Let \(z^k_{i+1}\) be an approximate solution of (9.1.14), i.e. according to (9.1.11) there exists \(e^k_{i+1} \in \mathbb{R}^n\) with a sufficiently small norm such that

\[
e^k_{i+1} \in \partial f^k_{i+1}(z^k_{i+1}) + \frac{1}{\chi_k}(z^k_{i+1} - x^k) + N_K(z^k_{i+1}). \tag{9.1.15}
\]

For \(e^k_{i+1} = 0\) the point \(z^k_{i+1}\) solves (9.1.14) exactly.

Now, we show that, under appropriate conditions on \(f^k_{i+1}\) and \(e^k_{i+1}\), the iterate \(z^k_{i+1}\) solves (9.1.11). Indeed, by the definition of the subdifferential we get immediately from the last inclusion

\[
f^k_{i+1}(y) \geq f^k_{i+1}(z^k_{i+1}) + \langle e^k_{i+1} - \frac{1}{\chi_k}(z^k_{i+1} - x^k) - w^k_{i+1}, y - z^k_{i+1} \rangle \quad \forall y \in \mathbb{R}^n,
\]

with \(w^k_{i+1} \in N_K(z^k_{i+1})\). Further, due to (9.1.13) and the property of \(z^k_{i+1}\), the latter inequality leads to

\[
f(y) \geq f(z^k_{i+1}) + \langle e^k_{i+1} - \frac{1}{\chi_k}(z^k_{i+1} - x^k) - w^k_{i+1}, y - z^k_{i+1} \rangle - \epsilon_{k,i+1} \quad \forall y \in \mathbb{R}^n. \tag{9.1.16}
\]
9.1. PPR WITH RELAXATION

with
\[ \epsilon_{k,i+1} = f(z_{i+1}^k) - f_{i+1}(z_{i+1}^k). \]  
(9.1.17)

Due to (9.1.13) it holds \( \epsilon_{k,i+1} \geq 0 \). In view of the definition of the \( \epsilon \)-subdifferential, (9.1.16) is equivalent to
\[ \epsilon_{i+1}^k \in \partial \epsilon_{k,i+1} f(z_{i+1}^k) + \frac{1}{\chi_k} (z_{i+1}^k - x^k) + N_K(z_{i+1}^k). \]  
(9.1.18)

According to \( \partial \epsilon f \subseteq (\partial f)' \) for all \( \epsilon \geq 0 \) (cf. for instance, [61], Proposition 3) and observing that \( \chi_k N_K = N_K \) for any \( \chi_k > 0 \), one obtains finally that \( z_{i+1}^k \) satisfies the inclusion
\[ \chi_k \epsilon_{i+1}^k \in \chi_k (\partial f)' \chi_{i+1} (z_{i+1}^k) + z_{i+1}^k - x^k + N_K(z_{i+1}^k). \]  
(9.1.19)

If \( \epsilon_{k,i+1} \leq \epsilon_k \) and \( \|\epsilon_{k,i+1}\| \leq \delta_k/\chi_k \), then \( z_{i+1}^k \) solves subproblem (9.1.11) and we can take \( y_{i+1}^k := z_{i+1}^k \). If \( \|\epsilon_{k,i+1}\| > \delta_k/\chi_k \), then we should try to solve (9.1.14) more precisely. If \( \|\epsilon_{k,i+1}\| \leq \delta_k/\chi_k \), but \( \epsilon_{k,i+1} > \epsilon_k \), one has to enhance function \( f_{i+1}^k \) by some \( f_{i+2}^k \). After replacing \( f_{i+1}^k \) by \( f_{i+2}^k \) in (9.1.14) and solving the problem again, the resulting enlargement parameter \( \epsilon_{k,i+2} \) is less than \( \epsilon_{k,i+1} \).

The choice of the sequence \( \{f_{i+1}^k\}_{i \geq 1} \) is crucial for the determination of an approximate solution of (9.1.11). According to Correa and Lemaréchal [84], there are three conditions defining \( f_{i+1}^k \). For our purpose we adapt condition (4.8) in [84] to allow an inexact solving of (9.1.14) and require for \( \{f_{i+1}^k\}_{i \geq 1} \):

(CP1) \( f_i^k(y) \leq f(y) \forall y \in \text{dom}(f), \forall i \geq 1, \)

(CP2) \( f_{i+1}^k(y) \geq f_i^k(z_i^k) + (e_i^k - \frac{1}{\chi_k} (z_i^k - x_i^k), y - z_i^k), \quad \forall y \in K, \forall i \geq 1, \)

(CP3) \( f_{i+1}^k(y) \geq f_i^k(z_i^k) + (s_i^k, y - z_i^k) \quad \forall y \in \mathbb{R}^n, \forall i \geq 0, \)

where the vectors \( s_i^k \) and \( e_i^k \) satisfy the inclusion (9.1.15) for all \( i \geq 1 \) and \( s_i^k \in \partial f_i^k(z_i^k) \forall i \geq 0 \) are arbitrarily chosen. For \( i := 0 \) set \( z_0^k := x_i^k \) in (CP3).

There are at least three possibilities to choose a sequence \( \{f_{i+1}^k\}_{i \geq 1} \), satisfying (CP1)-(CP3). The first one is a choice of “maximal size”
\[ f_{i+1}^k(y) := \max_{j \in J_i^k} f(z_j^k) + (s_j^k, y - z_j^k), \]  
(9.1.20)

where \( J_i^k := \{0, \ldots, i\} \). Due to \( f_i^k(y) \leq f_{i+1}^k(y) \) for any \( y \) and all \( i \geq 1 \), the function \( f_{i+1}^k \) given by (9.1.20) satisfies (CP1)-(CP3). The second one is a choice of “minimal size”
\[ f_{i+1}^k(y) := \max\{l_i^k(y), f(z_i^k) + (s_i^k, y - z_i^k)\}, \quad \forall i \geq 1, \]  
(9.1.21)

with
\[ l_i^k(y) := f_i^k(z_i^k) + (e_i^k - \frac{1}{\chi_k} (z_i^k - x_i^k), y - z_i^k). \]

For \( i := 0 \) we set \( f_i^k(y) := f(z_0^k) + (s_0^k, y - z_0^k) \) and \( z_0^k := x_i^k \). The validity of (CP2), (CP3) follows immediately from the definition. (CP1) is satisfied, since the two linear functions in (9.1.21) estimate \( f \) from below. The third possibility is a combination of (9.1.20) and (9.1.21)
\[ f_{i+1}^k(y) := \max\{l_i^k(y), f(z_i^k) + (s_j^k, y - z_j^k) : j \in J_i^k\}, \]  
(9.1.22)
where \( \hat{J}^k_i \subset J^k_i \) is an arbitrary set with \( i \in \hat{J}^k_i \).

The following proposition serves as a basis for the application of inexact solution of the bundle subproblems.

9.1.8 Proposition. (Lemma 5.3.1 [187])

Let \( K \) be a convex and closed set. Assume that \( f \) is a convex, lower semicontinuous function with \( K \subset \text{dom}(f) \) and the subdifferential \( \partial f \) is bounded on bounded subsets of \( K \). Further suppose that for any fixed \( k \in \mathbb{N} \) the sequence \( \{f^k\}_{i \geq 1} \) satisfies the conditions (CP1)-(CP3). If the solution set of \( \text{CP}(f,K) \) is nonempty and

\[
\sum_{i=1}^{\infty} \|e^k_{i+1}\| < +\infty,
\]

then

(a) \( \lim_{i \to \infty} \epsilon_{k,i+1} = \lim_{i \to \infty} f(z^k_{i+1}) - f^k_{i+1}(z^k_{i+1}) = 0. \)

(b) \( \{z^k\}_{i \in \mathbb{N}} \) converges to the unique solution of (9.1.12).

It should be noted that \( \partial f \) is bounded on bounded subsets of \( K \) if \( K \subset \text{int}(\text{dom}(f)) = \text{int}(\text{dom}(\partial f)) \).

Now we are looking for some numerical realization of Step 1 in Algorithm 9.1.3 in order to decide whether the current iterate \( x^k \) is a solution of \( \text{CP}(f,K) \). First of all one should be aware that \( x^k \) may not always be feasible for \( \text{CP}(f,K) \). In the case of over-relaxation, in Step 3 it can happen that \( x^k \) lies outside of \( K \). On the other hand \( y^k_{i+1} \) is always feasible. Therefore, we try to test \( y^k_{i+1} \) for optimality. If the conditions \( e^k_{i+1} \leq \delta_k / \chi_k \) and \( \epsilon_{k,i+1} \leq \epsilon_k \) hold for some \( i_k \geq 0 \), then according to (9.1.18) \( y^k_{i+1} := z^k_{i+1} \) satisfies the following condition

\[
\epsilon^k_{i_k+1} \in \partial e_{i_k,i+1} f(y^{k+1}) + \frac{1}{\chi_k} (y^{k+1} - y^k) + N_K(y^{k+1}),
\]

or

\[
f(y) \geq f(y^{k+1}) + \langle e^k_{i_k+1} - \frac{1}{\chi_k} (y^{k+1} - x^k), y - y^{k+1} \rangle - \epsilon_{k,i_k+1} \quad \forall \ y \in K.
\]

If for some \( \theta \geq 0 \)

\[
\|e^k_{i_k+1}\| + \frac{1}{\chi_k} \|y^{k+1} - x^k\| \leq \theta \quad \text{and} \quad \epsilon_{k,i_k+1} \leq \theta,
\]

(9.1.23)
is true, then it follows immediately for \( y^{k+1} \) that

\[
f(y) \geq f(y^{k+1}) - \theta \|y - y^{k+1}\| - \theta \quad \forall \ y \in K.
\]

The point \( y^{k+1} \) is called \( \theta \)-optimal. Obviously if \( \theta := 0 \) then \( y^{k+1} \) is a solution of \( \text{CP}(f,K) \).

Now we are able to formulate an implementable method for solving \( \text{CP}(f,K) \) based on Algorithm 9.1.3.
9.1.9 Algorithm. (Bundle RPPA)

S0. Choose any \( x^0 \in \mathbb{R}^n \) and \( \theta > 0 \). Set \( k := 0 \).
S1. Choose some \( \chi_k > 0 \), \( \epsilon_k > 0 \), \( \delta_k > 0 \). Set \( \epsilon_0 := x^k \) and \( i := -1 \).
S1.1 Set \( i := i + 1 \) and construct \( f_{k+1}^i \) satisfying (CP1)-(CP3).
S1.2 Find \( (z_{i+1}^k, \epsilon_{i+1}^k) \) as a solution of (9.1.15) with \( \| \epsilon_{i+1}^k \| \leq \delta_k / \chi_k \).
S1.3 Evaluate \( \epsilon_{k,i+1} := f(z_{i+1}^k) - f_{i+1}^k(z_{i+1}^k) \).
S1.4 If \( \epsilon_{k,i+1} \leq \epsilon_k \), then make a Serious Step:
set \( y^{k+1} := z_{i+1}^k \) and go to S2.
S1.5 If \( \epsilon_{k,i+1} > \epsilon_k \), then make a Null Step:
go to S1.1.
S2. If \( \| \epsilon_{i+1}^k \| + \frac{1}{\chi_k} \| y^{k+1} - x^k \| \leq \theta \) and \( \epsilon_{k,i+1} \leq \theta \): STOP,
y^{k+1} is a \( \theta \)-optimal point.
S3. Select \( \rho_k \in (0, 2/(1 + 2\delta_k)) \) and set
\[ x^{k+1} := (1 - \rho_k)x^k + \rho_k y^{k+1}. \]
S4. Set \( k := k + 1 \) and go to S1.

9.1.10 Theorem. Let \( K \subset \mathbb{R}^n \) be a closed, convex set and \( f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) be a proper convex, lower semi-continuous function with \( K \subset \text{dom}(f) \). Further let the subdifferential of \( f \) be bounded on bounded subsets of \( K \). If \( CP(f,K) \) is solvable and the sequences \( \{ \epsilon_k \} \subset (0, +\infty) \), \( \{ \delta_k \} \subset [0, +\infty) \), \( \{ \chi_k \} \) and \( \{ \rho_k \} \) satisfy
\[
\sum_{k=0}^{\infty} \epsilon_k < \infty, \quad \sum_{k=0}^{\infty} \delta_k < \infty, \quad 0 < \chi_k \leq \chi \leq \infty,
\]
\[
0 < \rho \leq \rho_k \leq \frac{2\sigma}{1 + 2\delta_k} \quad \forall k \in \mathbb{N}, \text{ for some } \sigma \in (0,1),
\]
then Algorithm 9.1.9 generates a \( \theta \)-optimal point of \( CP(f,K) \) after a finite number of iterations.

Proof: According to the assumption \( \epsilon_k > 0 \) and in view of Proposition 9.1.8, for each \( k \in \mathbb{N} \) there exists an \( i_k \in \mathbb{N} \) such that \( \epsilon_{k,i_k+1} \leq \epsilon_k \). Hence, after finitely many iterations, the inner iteration of Algorithm 9.1.9 – Step 1 will be done, i.e. this algorithm is well defined.
After Step 1 the solution \( z_{i+1}^k \) of the inclusion (9.1.19) is an approximative solution of the auxiliary problem (9.1.11). Therefore, Theorem 9.1.6 guarantees the convergence of the sequence \( \{ x^k \} \) to a solution of \( CP(f,K) \).
Further we can conclude via Theorem 9.1.6 that
\[
\lim_{k \to \infty} \| y^{k+1} - x^k \| = 0, \quad \lim_{k \to \infty} \epsilon_{k,i_k+1} \leq \lim_{k \to \infty} \epsilon_k = 0.
\]
Hence, the stopping criteria in Step 2 of Algorithm 9.1.9 will be satisfied for any \( \theta > 0 \) if \( k \) is large enough.

9.1.3 Numerical aspects and computational results

The Bundle RPPA algorithm has been implemented and tested on a set of well-known problems in convex non-smooth optimization. All benchmark examples are of the following form
\[
\min_{x \in K} f(x), \quad (9.1.24)
\]
where
\[ K = \{ x \in \mathbb{R}^n : g_1(x) \leq 0, \ldots, g_m(x) \leq 0 \} . \]

The functions \( f \) and \( g_j \) \((j = 1, \ldots, m)\) are assumed to be convex and \( K \) has to satisfy a constraint qualification, for instance, Slater’s condition. The detailed structure of the problems is described by means of the test examples in Subsection 9.1.4.1.

The assumptions of Theorem 9.1.10 are satisfied for these examples, because in all cases \( \text{dom}(f) = \mathbb{R}^n \). Consequently, the objective function is continuous on the whole space and the subdifferential is bounded on any bounded set (cf. [348], Theorem 10.1, Theorem 23.4).

In the sequel, we discuss some aspects of the practical implementation of Algorithm 9.1.9. For key relations of the bundle methods we refer to [237] and [366].

9.1.3.1 The quadratic subproblem and the sequence \( \{ \delta_k \} \)

The crucial point in the algorithm is the computation of the solution of the quadratic problem (9.1.14). For each fixed \( k \) we have to construct a sequence \( \{ f_{k+1} \} \) satisfying (CP1)-(CP3). As described above, there are several possibilities to do this. We prefer (9.1.22), because it allows to control the number of elements included in the bundle. For \( i \geq 1 \) let \( \hat{J}^k_i \) be a subset of \( \{ 0, \ldots, i \} \) with \( i \in \hat{J}^k_i \) (for \( i := 0 \) the choice of \( f_{k+1} \) is obvious). With an additional variable \( w \in \mathbb{R} \) Problem (9.1.14) can be written equivalently as

\[
\begin{align*}
\min_{(w,y) \in \mathbb{R}^{n+1}} & \quad w + \frac{1}{2\lambda_k} \| y - x^k \|^2 \\
\text{s.t.} & \quad f(z^k_j) + \langle s^k_j, y - z^k_j \rangle \leq w \quad \forall j \in \hat{J}^k_i, \\
& \quad f^k_i(z^k_i) + \langle \tilde{s}^k_i, y - z^k_i \rangle \leq w, \\
& \quad y \in K, \\
\end{align*}
\]

or

\[
\begin{align*}
\min_{(v,y) \in \mathbb{R}^{n+1}} & \quad v + \frac{1}{2\lambda_k} \| y - x^k \|^2 \\
\text{s.t.} & \quad -\alpha_{k,j} + \langle s^k_j, y - x^k \rangle \leq v \quad \forall j \in \hat{J}^k_i, \\
& \quad -\tilde{\alpha}_{k,i} + \langle \tilde{s}^k_i, y - x^k \rangle \leq v, \\
& \quad y \in K, \\
\end{align*}
\]

with
\[
\begin{align*}
v & := w - f(x^k), \\
\tilde{s}^k_i & := -\frac{1}{\lambda_k} (z^k_i - x^k), \\
\alpha_{k,j} & := f(x^k) - (f(z^k_j) + \langle s^k_j, x^k - z^k_j \rangle) \quad \forall j \in \hat{J}^k_i, \\
\tilde{\alpha}_{k,i} & := f(x^k) - (f^k_i(z^k_i) + \langle \tilde{s}^k_i, x^k - z^k_i \rangle).
\end{align*}
\]

Due to the convexity of \( f \), the linearization errors \( \alpha_{k,j} \) \((j \in \hat{J}^k_i)\) and \( \tilde{\alpha}_{k,i} \) are non-negative.

If \( K \) is polyhedral, then (9.1.26) is a convex quadratic programming problem with linear constraints for which efficient solvers are known. If some of the functions \( g_j \) are nonlinear, then the choice of a suitable solution method depends
on further properties of $g_j$. In order to design an universal solver, we suggest to eliminate the restriction set $K$ by using an exact penalty function. Therefore, if $m > 0$ we consider

$$F(x) := f(x) + c \sum_{i=1}^{m} \max\{g_i(x), 0\},$$

with some constant penalty parameter $c > 0$ and, instead of (9.1.24), we solve the unrestricted non-smooth problem

$$\min_{x \in \mathbb{R}^n} F(x).$$

Since we are dealing in (9.1.24) with a convex optimization problem satisfying a constraint qualification, an a priori choice of the penalty parameter $c$ is possible (see, for instance [179], Chapter VII, Corollary 3.2.3). The modifications in the problem data of (9.1.26) are marginal: $f$ must be replaced by $F$ and consequently the function as well as the subgradient evaluations have to be changed.

The transfer to the case $K := \mathbb{R}^n$ has a further advantage. The dual of (9.1.26) can be evaluated explicitly. Denoting by $(\nu_j^k : j \in \hat{J}_k^i; \tilde{\nu}_i^k)$ the Lagrange multipliers of the linear constraints, the dual problem has the form

$$\min_{\nu \in \mathbb{R}^{|\hat{J}_k^i|+1}} \left\| \sum_{j \in \hat{J}_k^i} \nu_j^k s_j^k + \tilde{\nu}_i^k \tilde{s}_i^k \right\|^2 + \frac{1}{\chi_k} \left( \sum_{j \in \hat{J}_k^i} \nu_j^k \alpha_{k,j} + \tilde{\nu}_i^k \tilde{\alpha}_{k,i} \right)$$

$$\text{s.t.} \quad \sum_{j \in \hat{J}_k^i} \nu_j + \tilde{\nu}_i = 1, \quad \nu_j \geq 0 \quad \forall j \in \hat{J}_k^i, \quad \tilde{\nu}_i \geq 0. \quad (9.1.27)$$

Problem (9.1.27) is a quadratic programming problem. It has $|\hat{J}_k^i| + 1$ variables and $|\hat{J}_k^i| + 2$ constraints in comparison to $n+1$ variables and $|\hat{J}_k^i| + 1$ constraints in problem (9.1.26). Moreover, the constraints of (9.1.27) are structurally simpler than the constraints of (9.1.26). Further, the number of the bundle elements $|\hat{J}_k^i| + 1$ can be chosen arbitrarily between 2 and $i+1$, i.e. we can decide on the size of problem (9.1.27) in dependence of the numerical effort and the quality of the approximation of $f$ by $f_{i+1}^k$.

Now, if the solution $(\nu_j^k : j \in \hat{J}_k^i; \tilde{\nu}_i^k)$ of the dual problem (9.1.27) is calculated, the solution $(v_{i+1}^k, z_{i+1}^k)$ of (9.1.26) is given by

$$z_{i+1}^k := x^k - \chi_k \left( \sum_{j \in \hat{J}_k^i} \nu_j^k s_j^k + \tilde{\nu}_i^k \tilde{s}_i^k \right), \quad (9.1.28)$$

$$v_{i+1}^k := -\frac{1}{\chi_k} z_{i+1}^k - x^k - \left( \sum_{j \in \hat{J}_k^i} \nu_j^k \alpha_{k,j} + \tilde{\nu}_i^k \tilde{\alpha}_{k,i} \right). \quad (9.1.29)$$

Observe that at the solution $(w_{i+1}^k, z_{i+1}^k)$ of (9.1.25) the relation $f_{i+1}^k(z_{i+1}^k) = w_{i+1}^k$ is true. Therefore, it holds

$$f_{i+1}^k(z_{i+1}^k) = v_{i+1}^k + f(x^k), \quad (9.1.30)$$

and due to (9.1.17)

$$\epsilon_{k,i+1} = f(z_{i+1}^k) - v_{i+1}^k - f(x^k). \quad (9.1.31)$$
With \( \epsilon_{k,i+1} \) given by (9.1.31) the validity of the stopping criteria for the inner iteration in Algorithm 9.1.9 (see Step 1.4 and Step 1.5) can be proved. If the stopping criteria is not satisfied, then one has to update \( \tilde{s}^k_{i+1} \) and \( \tilde{\alpha}^k_{i+1} \) by

\[
\tilde{s}^k_{i+1} := \sum_{j \in \hat{J}^k} \nu_j^k s^k_j + \tilde{\nu}_i^k \tilde{s}^k_i,
\]

\[
\tilde{\alpha}^k_{i+1} := \sum_{j \in \hat{J}^k} \nu_j^k \alpha^k_{j,i} + \tilde{\nu}_i^k \tilde{\alpha}^k_{i,i},
\]

and to solve the dual problem (9.1.27) for \( i + 1 \) again.

For the sake of simplicity we take here the accuracy parameter \( \delta_k \equiv 0 \) and assume \( e^k_i = 0 \) for all \( i \in \mathbb{N} \) and all \( k \in \mathbb{N} \) in (CP2). It is possible to carry out a deeper analysis and to consider an inexact solution of the dual problem (9.1.27). Nevertheless, for computational purposes this is not crucial and can be neglected.

### 9.1.3.2 Choice of sequences \( \{ \chi_k \} \) and \( \{ \epsilon_k \} \)

By Theorem 9.1.10 the sequence \( \{ \chi_k \} \) must be bounded. Numerical experiences carried out by Kiwiel [237] and Schramm and Zowe [366] show that the choice of \( \{ \chi_k \} \) has a great influence on the convergence of the algorithm. Our numerical computations are executed with two alternative selections of this sequence: Firstly, \( \chi_k \equiv \chi > 0 \) and secondly \( \chi_k \) is updated by the weighting technique which can be found in [237].

The choice of \( \{ \epsilon_k \} \) in Algorithm 9.1.9 determines the number of inner iterations of the algorithm and therefore it is directly responsible for the numerical progress. Due to Theorem 9.1.10 there is only one requirement: the sequence \( \{ \epsilon_k \} \) has to be summable. In order to avoid an a priori setting and to make its choice dependent on the progress of the algorithm, we suggest the following strategy:

\[
\epsilon_k := -\gamma_1 v^k_{i_{k+1}} \quad \forall \; k \in \mathbb{N},
\]

(9.1.32)

where \( \gamma_1 \in (0,1) \) is some constant and \( i_k \in \mathbb{N} \) is the first index for which it holds

\[
\epsilon_{k,i_{k+1}} \leq -\gamma_1 v^k_{i_{k+1}}.
\]

(9.1.33)

To show the existence of such an index \( i_k \) in (9.1.33), we assume the opposite, i.e. let \( \epsilon_{k,i+1} > -\gamma_1 v^k_{i+1} \) for all \( i \geq 0 \). In view of (9.1.30) and (9.1.31) the latter inequality is equivalent to

\[
\epsilon_{k,i+1} > -\gamma_1 (f(x^k) - f(z^k_{i+1})) = -\gamma_1 (f(x^k) - f(z^k_{i+1}) + \epsilon_{k,i+1}) \quad \forall \; i \geq 0.
\]

With Proposition 9.1.8 and the continuity of \( f \) we get for \( i \to \infty \)

\[ 0 \geq \gamma_1 (f(x^k) - f(z^k)). \]

Hence, due to \( \gamma_1 > 0 \), we infer \( f(x^k) \leq f(z^k) \). The point \( z^k \) is optimal for Problem (9.1.12), therefore

\[
f(x^k) \geq f(z^k) + \frac{1}{\chi_k}\|z^k - x^k\|^2.
\]
Combining the last two inequalities we obtain \( z^k = x^k \). Finally, the optimality condition for \( z^k \) says that \( 0 \in \partial f(x^k) \), i.e. the point \( x^k \) solves Problem \( CP(f, K) \).

Let us summarize: If (9.1.33) is false for infinitely many steps, then \( x^k \) solves the original problem. If \( x^k \) is not a solution of \( CP(f, K) \), then (9.1.33) is true after finitely many steps.

Assume now that the current iteration point is not a solution of \( CP(f, K) \). It remains to show if \( \{ \epsilon_k \} \) is defined by (9.1.32), then it is summable. Indeed, after \( i_k \) steps of the inner iteration (see Step 1.4)

\[
f(z_{i_{k+1}}^k) \leq f(x^k) + (1 - \gamma_1)v_{i_{k+1}}^k
\]

is true. If the stopping criteria is valid, then set \( y_{k+1} := z_{i_{k+1}}^k \) and one gets

\[
f(y_{k+1}) \leq f(x^k) + (1 - \gamma_1)v_{i_{k+1}}^k. \quad (9.1.34)
\]

If after the relaxation step (Step 3) the relation

\[
f(x^{k+1}) \leq f(y_{k+1}) - \gamma_2 v_{i_{k+1}}^k
\]

holds true for some \(-\infty < \gamma_2 < 1 - \gamma_1\) (see the discussion in Subsection 9.1.4 below), then (9.1.34) and (9.1.35) lead to

\[
f(x^{k+1}) \leq f(x^k) + (1 - \gamma_1 - \gamma_2)v_{i_{k+1}}^k. \quad (9.1.36)
\]

With (9.1.36) and (9.1.32) we conclude that for any \( N \in \mathbb{N} \)

\[
\sum_{k=0}^{N} \epsilon_k = -\gamma_1 \sum_{k=0}^{N} v_{i_{k+1}}^k \leq \frac{\gamma_1(f(x^0) - f(x^N))}{1 - \gamma_1 - \gamma_2} \leq \frac{\gamma_1(f(x^0) - f(x^*))}{1 - \gamma_1 - \gamma_2},
\]

where \( x^* \) is an arbitrary solution of \( CP(f, K) \). Taking limit as \( N \to \infty \) it yields immediately \( \sum_{k=0}^{\infty} \epsilon_k < \infty \).

9.1.11 Remark. Inequality (9.1.33) serves as a measure for the quality of the relaxation step. With the help of the parameter \( \gamma_2 \) the selection of the points \( x^{k+1} \) in question can be sharpened. If \( \gamma_2 < 0 \) then for the next iterate \( x^{k+1} \) it holds \( \varphi(x^{k+1}) < \varphi(y^{k+1}) \). However, it is possible to have \( \gamma_2 > 0 \). In this case \( \varphi(x^{k+1}) < \varphi(y^{k+1}) \) is also feasible after the relaxation step. Hence, a trivial choice of the parameter \( \gamma_2 \) could be \( \gamma_2 \in [0, 1 - \gamma_1] \). In this case various relaxation parameters can be chosen, in particular \( \rho_k := 1 \) is possible.

Note that, according to the choice of \( \{ \epsilon_k \} \) in (9.1.32), the stopping condition for a Serious Step in Algorithm 9.1.9 coincides with the typical condition for a serious step in minimization methods for convex non-differentiable functions (see [237], [262], and [366]).

9.1.4 Choice of relaxation parameters

By Theorem 9.1.10 the relaxation parameter \( \rho_k \) can be selected arbitrarily from the interval \((0, 2/(1 + 2\delta_k))\). To guarantee the summability of \( \{ \epsilon_k \} \), in view of (9.1.35), a further condition on \( \rho_k \) is needed. But it turns out that (9.1.35) is not a strong restriction. For example it can be trivially satisfied for \( \gamma_2 = 0 \) if \( \rho_k := 1 \) is chosen. For \( \gamma_2 \in (0, 1 - \gamma_1) \) we are allowed to determinate \( \rho_k \) such that \( f(x^{k+1}) > f(y^{k+1}) \). In any cases if (9.1.35) is true after the relaxation step,
then \( f(x^{k+1}) < f(x^k) \) holds for all \( k \in \mathbb{N} \).

Certainly there are many possibilities to execute the relaxation step. It could be a line search procedure, where (9.1.35) plays the role of a stopping criterion. Line search requires usually a big number of function evaluations. Our goal is to implement Step 3 such that an adaptive choice of \( \rho_k \) is possible and the number of function evaluations becomes minimal. For instance, it can be done by taking \( \rho_k \) as the minimizer of a quadratic polynomial \( p_k \) approximating \( f(x^k + t(y^{k+1} - x^k)) \) on \([0, 2] \). In this case the detailed description of Step 3 in Algorithm 9.1.9 reads as follows:

S3. Initialization: \( relaxation := true, i_r := 0, \gamma_2 \in (-\infty, 1 - \gamma_1) \).

S3.0 If \( relaxation = false \), then set \( x^{k+1} := y^{k+1} \) and go to S4.

S3.1 Approximate \( f(x^k + t(y^{k+1} - x^k)) \) on \([0, 2] \) by
\[
p_k(t) := a_0 + a_1 t + a_2 t^2
\]
with
\[
a_0 := f(x^k),
a_1 := -\frac{1}{2} f(x^k) + 2 f(y^{k+1}) - \frac{1}{2} f(x^k + 2(y^{k+1} - x^k)),
a_2 := \frac{1}{2} f(x^k) - f(y^{k+1}) + \frac{1}{2} f(x^k + 2(y^{k+1} - x^k)).
\]

S3.2 Set \( \rho_k := \min\{-\frac{a_1}{2a_2}, 1.99\} \) and compute
\[
x^{k+1} := (1 - \rho_k)x^k + \rho_k y^{k+1}.
\]

S3.3 If \( f(x^{k+1}) > f(y^{k+1}) - \gamma_3 v_{k+1} \), set \( x^{k+1} := y^{k+1} \), go to S4.

S3.4 If \(|1 - \rho_k| \leq 0.1 \), then set \( i_r := i_r + 1 \).

S3.5 If \( i_r > 3 \), then set \( relaxation := false \), go to S4.

It can be easily verified that the polynomial \( p_k(\cdot) \) takes its minimum at \( -\frac{a_1}{2a_2} \).

If \( a_2 := 0 \) or \( -\frac{a_1}{2a_2} \geq 2 \), we set \( \rho_k := 1.99 \). In Step 3.4 it is checked whether the new iterate \( x^{k+1} \) is far from \( y^{k+1} \). If \( ||x^{k+1} - y^{k+1}|| \leq 0.1 ||y^{k+1} - x^k|| \), then the counter \( i_r \) has to be increased. If \( i_r \) is greater than some positive integer, e.g. \( i_r > 3 \), then the variable \( relaxation \) is set to \( false \) and no further relaxation steps have to be executed. With this strategy one can identify whether the relaxation step accelerates the convergence of the iteration process.

9.1.4.1 Numerical tests

Most of the benchmark examples listed here were taken from [237], [262] or [366] (see there for more details). First we describe briefly the problem structure, in-
9.1. PPR WITH RELAXATION

Test Example 3.
\[ f(x) := \max \{ x_1^2 + x_2^2 - \frac{16}{25} x_1^3 \}, \]
\[ n := 2, m := 0, x^0 := (0, 0, 0.6)^T, f(x^0) = -0.8, x^* = (1, 0)^T, f(x^*) = -1. \]

Test Example 4. (Rosen-Suzuki)
\[ f(x) := x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 12x_3 + 7x_4, \]
\[ g_1(x) := x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8, \]
\[ g_2(x) := x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10, \]
\[ g_3(x) := x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5, \]
\[ n := 4, m := 3, x^0 := 0, f(x^0) = 0, x^* = (0, 1, 2, -1)^T, f(x^*) = -44. \]

Test Example 5. (Minimax location problem)
\[ f(x) := \max \{ w_1\|a^i - (x_1, x_2)\|_{p_1}, w_2\|a^i - (x_3, x_4)\|_{p_2}, \|((x_1, x_2) - (x_3, x_4))\| \}, \]
\[ g_1(x) := \|((x_1, x_2) - a^1)^2 - 144, \]
\[ g_2(x) := \|((x_1, x_2) - a^i)^2 - 128, \]
\[ g_3(x) := \|((x_3, x_4) - a^i)^2 - 225, \]
\[ g_4(x) := \|((x_3, x_4) - a^i)^2 - 144, \]
\[ n := 4, m := 4, a^i \in \mathbb{R}^2 \forall i, x^0 := (15, 22, 27, 11), f(x^0) = 218.3607415, \]
\[ f(x^*) = 23.886767. \]

Test Example 6. (Streit’s problem No. 1)
\[ f(z) := \|Az - b\|_\infty, \]
\[ g_i(z) := (z - c_i)| - t_i, \quad i = 1, \ldots, d, \]
\[ g_{d+1}(z) := (Bz - g)| - c_i, \quad i = 1, \ldots, r, \]
where \( A \in \mathbb{C}^{n \times d}, b \in \mathbb{C}^n, c \in \mathbb{C}^d, B \in \mathbb{C}^{r \times d}, g \in \mathbb{C}^r. \)
Setting \( z_1 := x_2i - 1 + \sqrt{2} x_2i \) we get an optimization problem in \( \mathbb{R}^d, \) considered for
\( d := 2, s := 5, r := 2, n := 4, m := 4, z^0 := 0, f(z^0) = \sqrt{2}, z^* = (-\frac{1 + \sqrt{2}}{2}, 0)^T, \)
\( f(z^*) = -\sqrt{2}. \)

Test Example 7. (Streit’s problem No. 2)
Same form as Streit’s problem No. 1 with \( d := 2, s := 5, r := 2, n := 4, m := 4, z^0 := 0, f(z^0) = \sqrt{2}, f(z^*) = 1.0142136. \)

Test Example 8. (Streit’s problem No. 3)
Same form as Streit’s problem No. 1 with \( d := 3, s := 101, r := 0, n := 6, m := 6, z^0 := 0, f(z^0) = 1, f(z^*) = 0.01470631. \)

Test Example 9. (Shor)
\[ f(x) := \max_{1 \leq r \leq 10} b_i \sum_{j=1}^{5} (x_j - a_{ij})^2, \]
\[ n := 5, m := 0, x^0 := (1, \ldots, 1)^T, f(x^0) = 80, f(x^*) = 22.60016. \]
Test Example 10. (Colville)

\[ f(x) := \sum_{i=1}^{5} \sum_{j=1}^{5} c_{ij} x_i x_j + \sum_{i=1}^{5} d_i x_i + \sum_{i=1}^{5} e_i x_i, \]
\[ g_i(x) := b_i - \sum_{j=1}^{5} a_{ij} x_j, \quad i = 1, \ldots, 10, \]
\[ g_{10+i}(x) := -x_i, \quad i = 1, \ldots, 10, \]
\[ n := 5, \quad m := 20, \quad x^0 := (0,0,0,0,1)^T, \quad f(x^0) = 20, \quad f(x^*) = -23.0448869. \]

Test Example 11. (Love’s minimum location problem)

\[ f(x) := \frac{3}{5} \sum_{i,j=1}^{5} w_{ij} \left\| (x_{2i-1}, x_{2i}) - a^i \right\|_p + \sum_{1 \leq i < j \leq 5} \left\| (x_{2i-1}, x_{2i}) - (x_{2j-1}, x_{2j}) \right\|_p, \]
\[ g_1(x) := x_5 + x_6 - 3.0, \]
\[ n := 6, \quad m := 1, \quad x^0 := 0, \quad f(x^0) = 177.31, \quad f(x^*) = 68.8256. \]

Test Example 12. (Lemarechal’s MaxQuad problem)

\[ f(x) := \max_{1 \leq i \leq 5} \langle A^i x, x \rangle - \langle b^i, x \rangle, \]
\[ n := 10, \quad m := 0, \quad x^0 := (1, \ldots, 1)^T \in \mathbb{R}^{10}, \quad f(x^0) = 5337.07, \quad f(x^*) = -0.8414084. \]

Test Example 13. (Constrained MaxQuad)

\[ f(x) := \max_{1 \leq i \leq 5} \langle A^i x, x \rangle - \langle b^i, x \rangle, \]
\[ g_i(x) := |x_i| - 0.05 \quad i = 1, \ldots, 10, \]
\[ g_{11}(x) := \sum_{i=1}^{10} x_i - 0.05, \]
\[ n := 10, \quad m := 11, \quad x^0 := 0, \quad f(x^0) = 0, \quad f(x^*) = -0.36816644175. \]

Test Example 14. (Ill-conditioned LP)

\[ f(x) := \sum_{i=1}^{30} c_i (x_i - 1), \]
\[ g_i(x) := \langle a^i, x \rangle - b_i \quad i = 1, \ldots, 30, \]
\[ \text{where } a^i_j := \frac{1}{i+j}, \quad b_i := \sum_{j=1}^{30} a^i_j, \quad c_i := -b_i - \frac{1}{i^2}, \]
\[ \text{considered for } n := 30, \quad m := 30, \quad x^0 := 0, \quad f(x^0) = 40.81, \quad x^* = (1, \ldots, 1)^T, \quad f(x^*) = 0. \]

Test Example 15. (Lemarechal’s problem TR48)

\[ f(x) := \sum_{j=1}^{48} d_j \max_{1 \leq i \leq 48} \left( x_i - a_{ij} \right) - \sum_{i=1}^{48} b_i x_i, \]
\[ n := 48, \quad m := 0, \quad x^0 := 0, \quad f(x^0) = -464816, \quad f(x^*) = -638565. \]

Test Example 16. (Goffin)

\[ f(x) := 50 \max_{1 \leq i \leq 50} x_i - \sum_{i=1}^{50} x_i, \]
\[ n := 50, \quad m := 0, \quad x^0_i := i - \frac{51}{2} \quad (i = 1, \ldots, 50), \quad f(x^0) = 1225, \quad x^* = 0, \quad f(x^*) = 0. \]
9.1. PPR WITH RELAXATION

Test Example 17. (Polyhedral Problem)

\[ f(x) := \sum_{i=1}^{50} \left| \sum_{j=1}^{50} x_j - 1 \right|, \]

\( n := 50, m := 0, x^0 := 0, f(x^0) = 68.82, x^* = (1, \ldots, 1)^T, f(x^*) = 0. \)

The aim of the numerical tests is to study the influence of the relaxation step in the framework of proximal point algorithms. With respect to each test example Table 9.1.1 and Table 9.1.2 indicate the number \( k \) of iterations (number of serious steps), the total number \( \#_{\text{tot}} \) of steps (sum of serious steps and null steps), function values \( f(x^k) \) at the found minimal points \( x^k \) as well as the number of function evaluations \( \# f \). Mainly, we are interested in a comparison of the behavior of Algorithm 9.1.9 under the following three choices of \( \rho_k \): \( \rho_k \equiv 1 \) (no relaxation), \( \rho_k \equiv 1.2 \) (over-relaxation) and an adaptive variation of \( \rho_k \) according to Step 3.

As mentioned above, an adaptive choice of the relaxation parameter can be understood as a correction of the step length. Due to Step 1 in Algorithm 9.1.9 and in particular to (9.1.28), we observe that the regularization parameter \( \chi_k \) and therefore the length of the iteration (serious) step is fixed before the descent direction is found. In case the direction leads to a significant decrease of the objective function, it is natural to execute a larger step. Inequality (9.1.35), given with some negative parameter \( \gamma_2 \), is a measure for the quality of the step.

To make the influence of the relaxation parameter more clearly, we distinguish between an algorithm with a fixed regularization parameter (cf. Table 9.1.1) and one with a regularization parameter adopted by rules given in [237] (cf. Table 9.1.2).

Algorithm 9.1.9 was implemented in programming language C. In order to solve the quadratic problem (9.1.27) SCHITTKOWSKI’s QL-solver [364] was used, which is translated from FORTRAN to C by f2c. All results, summarized in Table 9.1.1 and Table 9.1.2, are calculated with the parameter settings \( \gamma_1 := 0.9, \gamma_2 := -10^{-2} \) and \( \theta := 10^{-4} \). The accuracy of the QL-solver is set to \( 10^{-10} \). For the test examples 4 – 8 and 11 – 14 the penalty coefficient \( c := 10 \) is chosen whereas \( c := 50 \) is set for the test example 10.

The first numerical experiences indicate that the relaxation steps with an adapted relaxation parameter contribute to the acceleration of the convergence. In particular in Table 9.1.1, where the regularization parameter \( \chi_k \) is fixed, one recognizes the positive influence of the relaxation parameter on the iteration process. The number of iterations (denoted by \( k \)) with an adapted \( \rho_k \) is in more than 60% of test cases smaller or equal in comparison with the number of iterations with \( \rho_k \equiv 1 \) and analogously, in more than 70% smaller, compared to the algorithm with constant over-relaxation. Comparing the function evaluations required for one relaxation step, it should be noted that the algorithm with constant over-relaxation needs only one additional function evaluation opposite to two additional evaluations for the adaptive relaxation. Nevertheless, the number of function evaluations for the adapted \( \rho_k \) is in more than 70% of the test cases smaller.

\[^1\text{Meanwhile the code "bundle method" corresponding to this approach is available as an extension of the GSL library (GNU Scientific Library, www.gsl.org/software/gsl).} \]
Choosing an adaptive relaxation parameter, the tuning of an optimal balance between the parameters $\chi_k$ and $\rho_k$ is more complicated. However, we can observe that the total number of steps (denoted by $\#_{\text{tot}}$) (and therefore the numerical effort for solving the quadratic problem (9.1.27)) is in more than 50% of the test cases smaller for adaptive $\rho_k$ than for others choices. An adaptive choice of the relaxation parameter for each test problem separately can certainly still more enhance the results.
<table>
<thead>
<tr>
<th>Nr.</th>
<th>( k )</th>
<th>( \rho_k \equiv 1 )</th>
<th>( f(x^k) )</th>
<th>( #f )</th>
<th>( k )</th>
<th>( \rho_k \equiv 1.2 )</th>
<th>( f(x^k) )</th>
<th>( #f )</th>
<th>( k )</th>
<th>( \rho_k )</th>
<th>( f(x^k) )</th>
<th>( #f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>9</td>
<td>-3.000000</td>
<td>10</td>
<td>11</td>
<td>13</td>
<td>-2.999994</td>
<td>25</td>
<td>5</td>
<td>9</td>
<td>-3.000000</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>37</td>
<td>-1.000000</td>
<td>38</td>
<td>7</td>
<td>20</td>
<td>-0.999998</td>
<td>28</td>
<td>14</td>
<td>46</td>
<td>-1.000000</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>8</td>
<td>0.000000</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>0.000000</td>
<td>13</td>
<td>1</td>
<td>2</td>
<td>0.000000</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>72</td>
<td>-44.000000</td>
<td>73</td>
<td>25</td>
<td>73</td>
<td>-44.000000</td>
<td>99</td>
<td>18</td>
<td>58</td>
<td>-44.000000</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>45</td>
<td>23.886767</td>
<td>46</td>
<td>26</td>
<td>40</td>
<td>23.886802</td>
<td>67</td>
<td>22</td>
<td>46</td>
<td>23.886767</td>
<td>57</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>33</td>
<td>0.707107</td>
<td>34</td>
<td>16</td>
<td>23</td>
<td>0.707108</td>
<td>40</td>
<td>16</td>
<td>36</td>
<td>0.707107</td>
<td>49</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>18</td>
<td>1.014214</td>
<td>19</td>
<td>14</td>
<td>29</td>
<td>1.014214</td>
<td>44</td>
<td>8</td>
<td>17</td>
<td>1.014214</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>254</td>
<td>260</td>
<td>0.014706</td>
<td>261</td>
<td>212</td>
<td>221</td>
<td>0.014706</td>
<td>434</td>
<td>208</td>
<td>222</td>
<td>0.014706</td>
<td>335</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
<td>56</td>
<td>22.600162</td>
<td>57</td>
<td>27</td>
<td>58</td>
<td>22.600166</td>
<td>86</td>
<td>19</td>
<td>62</td>
<td>22.600163</td>
<td>79</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>88</td>
<td>-23.044862</td>
<td>89</td>
<td>28</td>
<td>1029</td>
<td>-23.044884</td>
<td>1058</td>
<td>10</td>
<td>936</td>
<td>-23.044842</td>
<td>947</td>
</tr>
<tr>
<td>11</td>
<td>18</td>
<td>43</td>
<td>68.829556</td>
<td>44</td>
<td>39</td>
<td>66</td>
<td>68.829556</td>
<td>106</td>
<td>15</td>
<td>41</td>
<td>68.829557</td>
<td>58</td>
</tr>
<tr>
<td>12</td>
<td>34</td>
<td>247</td>
<td>-0.841408</td>
<td>248</td>
<td>84</td>
<td>274</td>
<td>-0.841408</td>
<td>359</td>
<td>28</td>
<td>221</td>
<td>-0.841408</td>
<td>258</td>
</tr>
<tr>
<td>13</td>
<td>24</td>
<td>252</td>
<td>-0.368153</td>
<td>253</td>
<td>28</td>
<td>229</td>
<td>-0.368155</td>
<td>258</td>
<td>20</td>
<td>244</td>
<td>-0.368138</td>
<td>261</td>
</tr>
<tr>
<td>14</td>
<td>136</td>
<td>155</td>
<td>0.000000</td>
<td>156</td>
<td>118</td>
<td>139</td>
<td>0.000000</td>
<td>258</td>
<td>136</td>
<td>155</td>
<td>0.000000</td>
<td>176</td>
</tr>
<tr>
<td>15</td>
<td>457</td>
<td>636</td>
<td>-638565.00</td>
<td>637</td>
<td>380</td>
<td>514</td>
<td>-638564.99</td>
<td>895</td>
<td>458</td>
<td>616</td>
<td>-638565.00</td>
<td>631</td>
</tr>
<tr>
<td>16</td>
<td>25</td>
<td>56</td>
<td>0.000000</td>
<td>57</td>
<td>29</td>
<td>92</td>
<td>0.000000</td>
<td>122</td>
<td>25</td>
<td>56</td>
<td>0.000000</td>
<td>65</td>
</tr>
<tr>
<td>17</td>
<td>171</td>
<td>214</td>
<td>0.000001</td>
<td>215</td>
<td>149</td>
<td>182</td>
<td>0.000001</td>
<td>332</td>
<td>173</td>
<td>211</td>
<td>0.000001</td>
<td>224</td>
</tr>
</tbody>
</table>

Table 9.1.1: Algorithm 9.1.9 with \( \chi_k \equiv 1 \).
Table 9.1.2: Algorithm 9.1.9 with variable $\chi_k$.

<table>
<thead>
<tr>
<th>$f#$</th>
<th>$(\chi^*)f#_{\mathcal{F}_i}$</th>
<th>$\mathfrak{g}$</th>
<th>$f#_{\mathcal{F}_i}$</th>
<th>$\mathfrak{g}$</th>
<th>$f#_{\mathcal{F}_i}$</th>
<th>$\mathfrak{g}$</th>
<th>$f#_{\mathcal{F}_i}$</th>
<th>$\mathfrak{g}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000000.0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.000000</td>
<td>0.0</td>
<td>9</td>
<td>0.0</td>
<td>1</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1.0</td>
<td>10</td>
<td>1.0</td>
<td>1</td>
<td>1.0</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>1000000.0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0.000000</td>
<td>0.0</td>
<td>9</td>
<td>0.0</td>
<td>1</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1.0</td>
<td>10</td>
<td>1.0</td>
<td>1</td>
<td>1.0</td>
<td>0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

N, PPR FOR VIS WITH MULTI-VALUED OPERATORS
9.2 Interior PPR on Non-Polyhedral Sets

As in Section 9.1 we deal here with the variational inequality
\[ \text{VI}(Q, K) \quad \text{find } x \in K, \; q \in Q(x) : \quad \langle q, y - x \rangle \geq 0, \quad \forall y \in K, \]
where \( K \) is a convex, closed subset of \( \mathbb{R}^n \) but the operator \( Q \) is composed such that
\[ Q := \partial f + P, \]
where \( \partial f \) is the subdifferential of a proper convex lower semi-continuous function \( f \) and \( P : \mathbb{R}^n \to 2^{\mathbb{R}^n} \) is a maximal monotone operator. The application of the exact PPR can be described as follows.

Given \( v^1 \in K \) and a sequence \( \{\chi_k\}, \; 0 < \chi_k \leq \bar{\chi} < \infty \). With \( v^k \in K \) obtained in the previous step, define \( v^{k+1} \in K, \; q^{k+1} \in Q(v^{k+1}) \) such that
\[ \langle q^{k+1} + \chi_k \nabla_1 D(v^{k+1}, v^k), v - v^{k+1} \rangle \geq 0 \quad \forall v \in K. \]
Here \( D : (x, y) \mapsto \|x - y\|^2 \) and \( \nabla_1 D \) denotes the partial gradient of \( D \) with respect to the first vector argument.

For different modifications of the PPR, also with other quadratic distance functions \( D \), we refer to [215, 224, 240] and the bibliographies therein.

In the last decade, a new branch in PPR’s is intensively studied dealing with the use of non-quadratic distance functions. The main motivation for this type of proximal methods is the following: for certain classes of problems an appropriate choice of non-quadratic distance functions permits one to preserve the regularizing properties of the original version of the PPR and at the same time, this choice guarantees that the iterates stay in the interior of the set \( K \), i.e., with certain precaution, the regularized problems can be treated as unconstrained ones.

Usually, a Bregman distance, an entropic \( \varphi \)-divergence or a logarithmic-quadratic distance are applied to construct such \textit{interior proximal methods} (see references in [64, 240, 386] and the recent papers [23, 25, 229]). However, up to now, distance functions providing an interior point effect have been created only for the case that \( K \) is a polyhedral convex set or a ball.

Using a slightly modified definition of the class of Bregman functions (see Remark 9.2.3), we extend the Bregman-function-based interior proximal methods to solve \( \text{VI}(Q, K) \) on a non-polyhedral set
\[ K := \{ x \in \mathbb{R}^n : g_i(x) \leq 0, \; i \in I_1 \cup I_2 \}, \quad (9.2.1) \]
where the Slater condition is supposed to be valid and
\[ \begin{align*}
  g_i & (i \in I_1) \text{ are affine functions,} \\
  g_i & (i \in I_2) \text{ are convex and continuously differentiable functions,} \\
  \max_{i \in I_2} g_i & \text{ is strictly convex on } K.
\end{align*} \quad (9.2.2, 9.2.3) \]
Later on in Subsection 9.2.2.1 we will show that condition (9.2.3) can be weakened substantially.

The modification in the definition of Bregman functions consists mainly in
388 CHAPTER 9. PPR FOR VI’S WITH MULTI-VALUED OPERATORS

a relaxation of the standard convergence sensing condition (see Stand-(B4) in Remark 9.2.3) which proves to be restrictive already in case $K$ is a ball.

The convergence analysis of the method studied admits a successive approximation of the operator $Q$ by means of the $\epsilon$-enlargement concept, as well as an inexact solution of the subproblems under a criterion of the summability of the error vectors.

In Section 9.2.2 we check the validity of the modified requirements on Bregman functions for some particular functions and discuss the convergence of earlier developed Bregman-function-based methods under the modification mentioned.

9.2.1 Bregman function based PPR

Variational inequality VI($Q, K$) is considered under the following basic assumptions:

9.2.1 Assumption.

(A1) $K \subset \mathbb{R}^n$ is a convex closed set;

(A2) $\text{dom} Q \cap \text{int} K \neq \emptyset$;

(A3) $\text{SOL}(Q, K)$ of VI($Q, K$) is non-empty;

(A4) (a) $\mathcal{P} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is a maximal monotone operator,

(b) $\lim_{k \to \infty} y^k = \bar{y} \in K$, $p^k \in \mathcal{P}(y^k)$ implies that $\{p^k\}$ is a bounded sequence,

(c) $\partial f + \mathcal{P}$ is a paramonotone operator, see Definition A1.6.43.

Assumptions (A1), (A2) and (A4)(a) provide the maximal monotonicity of $Q + N_K$, where $N_K$ is the normality operator of $K$ (cf. [350], Theorem 1).

For some properties of paramonotone operators see [194]. In particular, because $\partial f$ is paramonotone we have paramonotonicity of $Q$. Paramonotonicity is a rather standard assumption in Bregman-function-based proximal methods.

Such kind of methods for variational inequalities with non-paramonotone operators was studied in [228].

The composed operators $Q = \partial f + \mathcal{P}$, where $\mathcal{P}$ is linear, paramonotone and non-symmetric and $\partial f$ is the subdifferential of a proper convex, lower semi-continuous function $f$, such that $K \setminus \text{dom} f \neq \emptyset$, form a wide class of operators satisfying (A). Assumption (A4)(b) is weaker than an assumption on the boundedness of the whole operator $Q$. Indeed, because the operator $\partial f$ is maximal monotone, the assumption $K \setminus \text{dom} f \neq \emptyset$ means that the restriction of $\partial f$ on $K$ is an unbounded operator, and hence, an assumption on the boundedness of the whole operator $Q$ is stronger than required here.

Now, let $h$ be a Bregman-type function with zone $\text{int} K$. According to the terminology of Bregman functions, under a zone of $h$ one understands an open convex set $S := \text{int}(\text{dom} h)$. More precisely, here we suppose:

9.2.2 Assumption.

(B1) $\text{dom} h = K$ and $h$ is continuous and strictly convex on $K$;
9.2. INTERIOR PPR ON NON-POLYHEDRAL SETS

(B2) \( h \) is continuously differentiable on \( \text{int} K \);

(B3) With a distance function

\[
D_h : (x, y) \in K \times \text{int} K \mapsto h(x) - h(y) - \langle \nabla h(y), x - y \rangle,
\]

for each \( x \in K \) there exist constants \( \alpha(x) > 0, c(x) \) such that

\[
D_h(x, y) + c(x) \geq \alpha(x) \| x - y \|, \quad \forall y \in \text{int} K;
\]

(B4) If \( \{ z^k \} \subset \text{int} K \) converges to \( z \), then at least one of the following properties is valid:

(i) \( \lim_{k \to \infty} D_h(z, z^k) = 0 \)

(ii) \( \lim_{k \to \infty} D_h(\bar{z}, z^k) = +\infty \) if \( \bar{z} \neq z, \bar{z} \in \text{bd} K \);

(B5) (zone coercivity) \( \nabla h(\text{int} K) = \mathbb{R}^n \).

9.2.3 Remark. Assumption (B4) is evidently weaker than the standard convergence sensing condition:

Stand-(B4) If \( \{ z^k \} \subset \text{int} K \) converges to \( z \), then \( \lim_{k \to \infty} D_h(z, z^k) = 0 \),

and it is closely related to condition (iv) in the definition of a generalized Bregman function introduced by Kiwiel, see [240], Definition 2.4.

Another standard condition

Stand-(B3) For any \( x \in K \) and any constant \( \alpha \) the set

\[
L^\alpha(x) := \{ y \in \text{int} K : D_h(x, y) \leq \alpha \}
\]

is bounded, if \( K \) is a bounded set.

Referring further on to standard conditions on Bregman functions, we have in mind the fulfillment of (B1), (B2), Stand-(B3), Stand-(B4), and one of the conditions (B5) or

(B6) (boundary coercivity)

If \( \{ z^k \} \subset \text{int} K \), \( \lim_{k \to \infty} z^k = z \in \text{bd} K \), then it holds for each \( x \in \text{int} K \)

\[
\lim_{k \to \infty} \langle \nabla f(z^k), x - z^k \rangle = -\infty.
\]

9.2.4 Remark. To our knowledge, among the Bregman functions (under standard conditions), considered in the literature (see [70], [71], [193] and references therein), only the function

\[
\sum_{j=1}^{n} (x_j - x_j^\beta), \quad 0 < \beta < 1
\]

on \( K := \mathbb{R}^n_+ \) does not satisfy (B3). Assumption (B3) was introduced in [229] in order to weaken the stopping criteria in generalized proximal methods.
CHAPTER 9. PPR FOR VI’S WITH MULTI-VALUED OPERATORS

In the sequel we make use of

9.2.5 Lemma. ([376], Theorem 2.4)
Let \( h \) satisfy the Assumptions 9.2.2 (B1) and (B2). If
\[
\{z^k\} \subset K, \quad \{y^k\} \subset \text{int}K, \quad \lim_{k \to \infty} D_h(z^k, y^k) = 0
\]
and one of these sequences converges, then the other one converges to the same limit, too.

9.2.6 Lemma. Let the Assumptions 9.2.2 (B1), (B2), Stand-(B3) and (B4) be valid and \( \{v^k\} \subset \text{int} K \). Moreover, suppose that \( C \) is a non-empty subset of \( K \), \( \{D_h(x, v^k)\} \) converges for each \( x \in C \), and each cluster point of \( \{v^k\} \) belongs to \( C \).
Then \( \{v^k\} \) converges to some \( v \in C \).

Proof: Because \( \{D_h(x, v^k)\} \) converges for \( x \in C \) and Stand-(B3) is valid, the sequence \( \{v^k\} \) is bounded. Take two subsequences \( \{v^{l_k}\} \) and \( \{v^{n_k}\} \) of \( \{v^k\} \) with
\[
\lim_{k \to \infty} v^{l_k} = \bar{v}, \quad \lim_{k \to \infty} v^{n_k} = \bar{v}.
\]
If \( \bar{v} \in \text{int} K \), then \( \lim_{k \to \infty} D_h(\bar{v}, v^{n_k}) = 0 \) follows from (B2), and since the whole sequence \( \{D_h(\bar{v}, v^k)\} \) converges, one gets
\[
\lim_{k \to \infty} D_h(\bar{v}, v^k) = 0.
\]
In turn, applying Lemma 9.2.5 with \( z^k := \bar{v}, \ y^k := v^{l_k} \), we conclude \( v^* = \bar{v} \).
Now let \( \bar{v} \in \text{bd} K \). We make use of (B4) setting \( z := v^*, \ z^k := v^{l_k} \) and \( \bar{z} := \bar{v} \).
If (B4)(i) is valid, then \( \lim_{k \to \infty} D_h(v^*, v^{l_k}) = 0 \), hence
\[
\lim_{k \to \infty} D_h(v^*, v^k) = 0.
\]
Again, Lemma 9.2.5 with \( z^k := v^*, \ y^k := v^{n_k} \), yields \( v^* = \bar{v} \).
But if (B4)(ii) holds, then \( \bar{v} \neq v^* \) is also impossible taking into account that \( \{D_h(\bar{v}, v^k)\} \) is a convergent sequence.

Let us remind on the \( \epsilon \)-enlargement of a maximal monotone operator \( Q \):
\[
Q_{\epsilon}(x) = \{u \in \mathbb{R}^n : \langle u - v, x - y \rangle \geq -\epsilon \quad \forall \ y \in \text{dom}Q, \ v \in Q(y)\}.
\]
For properties of the \( \epsilon \)-enlargement, see [61], [63].

Now we describe the method under consideration.

9.2.7 Method. (Bregman-function-based PPR)
Starting with an arbitrary \( x^1 \in \text{int} K \), the method generates two sequences \( \{x^k\} \subset \mathbb{R}^n \) and \( \{e^k\} \subset \mathbb{R}^n \) conforming to the recursion
\[
e^{k+1} \in Q_{\epsilon}(x^{k+1}) + \chi_k \nabla_1 D_h(x^{k+1}, x^k).
\]
\[
0 < \chi_k < \bar{\chi} \quad (\bar{\chi} > 0\ \text{arbitrary}), \quad \epsilon_k \geq 0, \quad \sum_{k=1}^{\infty} \frac{\epsilon_k}{\chi_k} < \infty, \quad \sum_{k=1}^{\infty} \frac{\|e^k\|}{\chi_k} < \infty. \tag{9.2.5}
\]
\[}
Here \( h \) is a Bregman function satisfying (B1)-(B5) and \( Q^k \) is an approximation of \( Q \) such that
\[
Q \subset Q^k \subset \partial_\epsilon f + P_\epsilon f,
\]
where \( \partial_\epsilon f \) denotes the \( \epsilon \)-subdifferential of \( f \).

Considering this method with \( e^k \equiv 0 \) instead of the last condition in (9.2.5), assumption (B3) can be replaced by Stand-(B3), with minor and evident modifications in the convergence analysis below.

According to [59], Lemma 1, the conditions (B1), (B2) and (B5) ensure that for each \( y \in \text{int}K \)
\[
\text{dom } \partial_1 D_h(\cdot, y) = \text{int}K, \tag{9.2.6}
\]
and
\[
\partial_1 D_h(x, y) = \begin{cases} \nabla h(x) - \nabla h(y) & \text{if } x \in \text{int}K \\ 0 & \text{otherwise} \end{cases} \tag{9.2.7}
\]
(\( \partial_1 D_h \) denotes the partial subdifferential of \( D_h \)).

Moreover, the conditions (A1), (A2), (B1), (B2) and (B5) imply that, for any \( e \in \mathbb{R}^n \), \( y \in \text{int}K \) and \( \chi > 0 \), the inclusion
\[
e \in Q(x) + \chi \partial_1 D_h(x, y)
\]
has a unique solution belonging to \( \text{int}K \) (see [59], Theorem 1).

Thus, the mentioned assumptions guarantee that for any sequences \( \{e^k\} \subset \mathbb{R}^n \) and \( \{\chi_k\}, \chi_k > 0 \), there exists a sequence \( \{x^k\} \) satisfying (9.2.4), and \( \{x^k\} \subset \text{int}K \) is a straightforward corollary of (9.2.6).

**9.2.8 Lemma.** Suppose that the sequence \( \{(x^k, e^k)\} \) fulfils recursion (9.2.4), that \( \{x^k\} \subset \text{int}K \) and that the Assumptions 9.2.1 (A1)-(A3), Assumptions 9.2.2 (B1)-(B3) and (9.2.5) are valid. Then

(i) \( \{D_h(x^*, x^k)\} \) is convergent for each \( x^* \in \text{SOL}(Q, K) \);

(ii) \( \{x^k\} \) is bounded;

(iii) \( \lim_{k \to \infty} D_h(x^{k+1}, x^k) = 0 \).

**Proof:** According to (9.2.4) there exists \( q^{k+1} \in Q^k(x^{k+1}) \) such that
\[
e^{k+1} = q^{k+1} + \chi_k \nabla_1 D_h(x^{k+1}, x^k).
\]
From this equality and (9.2.7) we conclude that
\[
\langle q^{k+1} + \chi_k (\nabla h(x^{k+1}) - \nabla h(x^k)), x - x^{k+1} \rangle \geq -\|e^{k+1}\|\|x - x^{k+1}\| \tag{9.2.8}
\]
holds for all \( x \in K \). Together with the obvious identity
\[
D_h(x, x^{k+1}) - D_h(x, x^k) = -D_h(x^{k+1}, x^k) + \langle \nabla h(x^k) - \nabla h(x^{k+1}), x - x^{k+1} \rangle \tag{9.2.9}
\]
this yields

\[
D_h(x, x^{k+1}) - D_h(x, x^k) \leq - D_h(x^{k+1}, x^k)
\]

\[+
\frac{1}{\lambda_k} \langle q^{k+1}, x - x^{k+1} \rangle + \frac{\|e^{k+1}\|}{\lambda_k} \|x - x^{k+1}\|, \forall x \in K.
\] (9.2.10)

Choose \(x^* \in \text{SOL} Q, K\) and \(q^* \in Q(x^*)\) satisfying

\[\langle q^*, x - x^* \rangle \geq 0, \ \forall x \in K.
\]

From the definition of \(P_\epsilon, \partial_\epsilon f \subset (\partial f)\) the inclusions

\[Q \subset Q^k \subset \partial_\epsilon^k f \subset Q^\prime \epsilon \subset Q^k.
\]

are true and one gets

\[\langle q^{k+1} - q^*, x^* - x^{k+1} \rangle \leq \epsilon_k.
\] (9.2.11)

Now, we take (9.2.10) with \(x := x^*\) and insert there the following two estimates

\[\langle q^{k+1}, x^* - x^{k+1} \rangle \leq \langle q^*, x^* - x^{k+1} \rangle + \epsilon_k \leq \epsilon_k,
\]

\[\|x^* - x^{k+1}\| \leq \frac{1}{\alpha(x^*)} [D_h(x^*, x^{k+1}) + c(x^*)].
\]

The first one is true because of (9.2.11) and \(x^{k+1} \in K\), and the second one is a consequence of (B3). These insertions lead to

\[
D_h(x^*, x^{k+1}) - D_h(x^*, x^k)
\]

\[\leq - D_h(x^{k+1}, x^k) + \frac{\delta_k}{\alpha(x^*)} D_h(x^*, x^{k+1}) + \delta_k \left(1 + \frac{c(x^*)}{\alpha(x^*)}\right),
\] (9.2.12)

where \(\delta_k := \chi_k^{-1} \max\{\|e^{k+1}\|, \epsilon_k\}\).

Conditions (9.2.5) provide that \(\frac{\delta_k}{\alpha(x^*)} < \frac{1}{2}\) for \(k \geq k_0, k_0\) sufficiently large. Therefore,

\[1 \leq \left(1 - \frac{\delta_k}{\alpha(x^*)}\right)^{-1} \leq \left(1 + \frac{2\delta_k}{\alpha(x^*)}\right) < 2, \ \forall k \geq k_0,
\]

and (9.2.12) results in

\[
D_h(x^*, x^{k+1})
\]

\[\leq \left(1 + \frac{2\delta_k}{\alpha(x^*)}\right) D_h(x^*, x^k) - D_h(x^{k+1}, x^k) + 2\delta_k \left(1 + \frac{c(x^*)}{\alpha(x^*)}\right).
\] (9.2.13)

Taking into account that \(D_h\) is a non-negative function and \(\sum_{k=1}^\infty \delta_k < \infty\), Lemma A3.1.7, applied to (9.2.13), guarantees that \(\{D_h(x^*, x^k)\}\) converges and

\[\lim_{k \to \infty} D_h(x^{k+1}, x^k) = 0.
\]

Now, condition (B3) implies that the sequence \(\{x^k\}\) is bounded. \qed
9.2.9 Lemma. Let Assumption 9.2.1 be valid, \( x^* \in \text{SOL}(Q, K), \bar{x} \in K \) and
\[
\langle \bar{p}, \bar{x} - x^* \rangle + f(\bar{x}) - f(x^*) \leq 0
\]
holds for some \( \bar{p} \in \mathcal{P}(\bar{x}) \).

Then \( \bar{x} \in \text{SOL}(Q, K) \).

Proof: Because \( x^* \in \text{SOL}(Q, K) \), there are \( \ell^* \in \partial f(x^*) \) and \( p^* \in \mathcal{P}(x^*) \) satisfying
\[
\langle \ell^* + p^*, x - x^* \rangle \geq 0 \quad \forall \; x \in K.
\]
In view of the monotonicity of the operator \( \mathcal{P} \), (9.2.15) implies
\[
\langle \bar{p} + \ell^*, \bar{x} - x^* \rangle \geq 0
\]
and using (9.2.14) we obtain
\[
f(\bar{x}) - f(x^*) \leq \langle \ell^*, \bar{x} - x^* \rangle.
\]
Thus, for any \( x \in \mathbb{R}^n \)
\[
f(x) - f(\bar{x}) = f(x) - f(x^*) - f(\bar{x}) + f(x^*) \geq f(x) - f(x^*) - \langle \ell^*, x - x^* \rangle - \langle \ell^*, \bar{x} - x \rangle \geq \langle \ell^*, x - \bar{x} \rangle.
\]
This indicates that \( \ell^* \in \partial f(\bar{x}) \), whereas (9.3.19) and the obvious inequality
\[
f(\bar{x}) - f(x^*) \geq \langle \ell^*, \bar{x} - x^* \rangle
\]
yield
\[
\langle \bar{p} + \ell^*, \bar{x} - x^* \rangle \leq 0.
\]
Now, the paramonotony of \( \partial f + \mathcal{P} \) provides that \( \bar{x} \in \text{SOL}(Q, K) \). \( \square \)

Assuming (A4), the following property (\( \star \)) of a paramonotone operator \( A \) on a convex closed set \( C \) is decisive (cf. [194]):

(\( \star \)) If \( u^* \) solves the variational inequality
\[
\text{find } u \in C, q \in A(u) : \langle q, v - u \rangle \geq 0, \; \forall \; v \in C,
\]
and for some \( \bar{u} \in C \) there exists \( \bar{z} \in A(\bar{u}) \) with
\[
\langle \bar{z}, u^* - \bar{u} \rangle \geq 0,
\]
then \( \bar{u} \) is also a solution of (9.2.16).

9.2.10 Lemma. Let Assumption 9.2.1 and the Assumptions 9.2.2 (B1)-(B3), (B5) be satisfied. Then each cluster point of the sequence \( \{x^k\} \), generated by Method 9.2.7, belongs to \( \text{SOL}(Q, K) \).

Proof: According to Lemma 9.2.8, the sequence \( \{x^k\} \) is bounded, hence there exists a convergent subsequence \( \{x^{j_k}\} \) with \( \lim_{k \to \infty} x^{j_k} = \bar{x} \). Because \( K \) is a closed set and \( x^k \in \text{int}K \) one gets \( \bar{x} \in K \). Taking into account conclusion (iii) of Lemma 9.2.8, the application of Lemma 9.2.5 with \( z^k := x^{j_k+1}, y^k := x^{j_k} \), yields \( \lim_{k \to \infty} x^{j_k+1} = \bar{x} \). Moreover, using the identity (9.2.9) with \( x := x^* \in \text{SOL}(Q, K) \), the relation
\[
\lim_{k \to \infty} \chi_{j_k} (\nabla f(x^{j_k+1}) - \nabla f(x^{j_k}), x^* - x^{j_k+1}) = 0
\]
(9.2.17)
follows immediately from Lemma 9.2.8 and $\chi_k \in (0, \bar{\chi}]$ for all $k$.

Let us at first suppose that (A4) is valid with $\mathcal{P} \equiv 0$. Due to the convexity of the function $f$ and $q^{k+1} \in \mathcal{Q}^k(x^{k+1}) \subset \partial f(x^{k+1})$, relation (9.2.8) considered for $x := x^*$ and $k := j_k$ implies that

$$-f(x^{k+1}) + f(x^*) + \chi_{j_k} \langle \nabla h(x^{j_k+1}) - \nabla h(x^{j_k}), x^* - x^{j_k+1} \rangle \geq -\|e^{j_k+1}\|\|x^* - x^{j_k+1}\| - \epsilon_{j_k}. \quad (9.2.18)$$

Passing to the limit in (9.2.18) for $k \to \infty$, in view of (9.2.17), (9.2.5), $\chi_k \in (0, \bar{\chi}]$ and the lower semi-continuity of $f$, we obtain $f(\bar{x}) \leq f(x^*)$. Together with $\bar{x} \in K$, $x^* \in \text{SOL}(\mathcal{Q}, K)$, this yields

$$0 \in \partial (f(\bar{x}) + \delta(\bar{x}|K)),$$

where $\delta(\cdot|K)$ is the indicator function of $K$. In view of assumption (A2), Theorem 23.8 in [348] provides $\bar{x} \in \text{SOL}(\mathcal{Q}, K)$.

Now, let us turn to the case that the operator $\mathcal{Q} \equiv \mathcal{P}$ possesses property (A4). Owing to the Brønsted-Rockafellar property of the $\epsilon$-enlargement (cf. [63]) and the relation $\mathcal{Q} \subset \mathcal{Q}^k \subset \mathcal{P}_{\epsilon_k}$, there exist $\bar{x}^{j_k+1}$ and $q(\bar{x}^{j_k+1}) \in \mathcal{P}(\bar{x}^{j_k+1})$ such that

$$\|x^{j_k+1} - \bar{x}^{j_k+1}\| \leq \sqrt{\epsilon_{j_k}} \quad \text{and} \quad \|q^{j_k+1} - q(\bar{x}^{j_k+1})\| \leq \sqrt{\epsilon_{j_k}}, \quad \forall k. \quad (9.2.19)$$

Hence, $\lim_{k \to \infty} \bar{x}^{j_k+1} = \bar{x}$, and taking into account (A4)(b) with $y^k := \bar{x}^{j_k+1}$, both sequences $\{q(\bar{x}^{j_k+1})\}$ and $\{q^{j_k+1}\}$ are bounded. Together with the second inequality in (9.2.19), this allows us to conclude, without loss of generality, that

$$\lim_{k \to \infty} q^{j_k+1} = \bar{q} \quad \text{and} \quad \lim_{k \to \infty} q(\bar{x}^{j_k+1}) = \bar{q}.$$

In turn, the maximal monotonicity of $\mathcal{P}$ ensures $\bar{q} \in \mathcal{P}(\bar{x})$.

Now we assemble both cases and observe Lemma 9.2.9. Inserting $x := x^* \in \text{SOL}(\mathcal{Q}, K)$ into (9.2.8) and passing to the limit for $k := j_k$, $k \to \infty$, we infer from (9.2.17) and (9.2.5) that

$$\langle \bar{q}, x^* - \bar{x} \rangle \geq 0.$$

Finally, property $(\star)$ of the paramonotone $\mathcal{Q}$ provides $\bar{x} \in \text{SOL}(\mathcal{Q}, K)$. \hfill $\square$

Now, from $\{x^k\} \subset \text{int}K$ and Lemmata 9.2.6, 9.2.10, the main convergence result follows immediately.

9.2.11 Theorem. Let Assumptions 9.2.1 and Assumption 9.2.2 be valid. Then the sequence $\{x^k\}$, generated by Method 9.2.7, belongs to $\text{int}K$ and converges to a solution of VI$(\mathcal{Q}, K)$.

9.2.2 Bregman functions with non-polyhedral zones

Bregman functions with zone $S := \text{int}K$ are of main interest in applications. Exactly in this case we deal with interior proximal methods. However, as it was already mentioned, up to now such Bregman functions have been constructed only for linearly constrained sets $K$ or in the case that $K$ is a ball.
9.2. INTERIOR PPR ON NON-POLYHEDRAL SETS

Usually, for linearly constrained $K$ condition Stand-(B4) (i.e. (B4)(i)) is supposed. In case that $K$ is a ball, instead of Stand-(B4) the condition (B4)(ii) was introduced in [70].

These convergence results are not applicable, for instance, if $K$ is the intersection of a half-space and a ball. The reason is that neither condition (B4)(i) nor (B4)(ii) considered alone for itself is guaranteed in this case. This will be shown in Example 9.2.16 below.

In this Section Method 9.2.7 is studied for the case

$$
K := \{ x \in \mathbb{R}^n : g_i(x) \leq 0, \ i \in I \}, \quad I = I_1 \cup I_2,
$$

where $I$ is a finite index set, $g_i$ $(i \in I_1)$ are affine functions, $g_i$ $(i \in I_2)$ are convex continuously differentiable functions and $\max_{i \in I_2} g_i$ is supposed to be strictly convex on $K$. Further Slater’s condition

$$
\exists \tilde{x} : g_i(\tilde{x}) < 0 \quad \forall \ i \in I
$$

is supposed to be valid.

The following statement clarifies a choice of Bregman-like functions with zone int$K$.

9.2.12 Theorem. Let $\varphi$ be a strictly convex, continuous and increasing function with $\text{dom}\varphi = (-\infty, 0]$, and $\varphi$ be continuously differentiable on $(-\infty, 0)$. Moreover, let

$$
\lim_{t \to 0} \varphi'(t) t = 0,
$$

$$
\lim_{t \to 0} \varphi'(t) = +\infty.
$$

Then the function

$$
h(x) := \sum_{i \in I} \varphi(g_i(x)) + \theta \sum_{j=1}^n |x_j|^\gamma, \quad \gamma > 1 \text{ is fixed},
$$

with $\theta := 0$ if $K$ is bounded and $\theta := 1$ if boundedness of $K$ is unknown, satisfies Assumption 9.2.2.

As it will be clear from the proof of this theorem, instead of $\sum_{i=1}^n |x_j|^\gamma$ in (9.2.24) one can take any strictly convex and continuously differentiable function guaranteeing the fulfilment of (B3) for $h$.

Proof: Step by step we check whether the assumptions (B1)-(B5) are satisfied for the function $h$ in (9.2.24) to be a Bregman function.

- Because $\varphi$ is a convex increasing function, the convexity of the composition $\varphi \circ g_i$ (and hence, of $h$) on the set $K$ is guaranteed. Thus, if $\theta := 1$, the strict convexity of $h$ is evident.

But, if $\theta := 0$ (i.e. $K$ is bounded), we have to consider two cases.

(a) $I_2 \neq \emptyset$. Assume that $\sum_{i \in I_2} \varphi \circ g_i$ is not strictly convex on $K$. Then one can choose points $x^1, x^2 \in K$, $x^1 \neq x^2$, and $\lambda \in (0, 1)$ such that

$$
\sum_{i \in I_2} \varphi(g_i(\lambda x^1 + (1 - \lambda)x^2)) = \lambda \sum_{i \in I_2} \varphi(g_i(x^1)) + (1 - \lambda) \sum_{i \in I_2} \varphi(g_i(x^2)).
$$
Because each function $\varphi \circ g_i$ is convex, the last equality means that, for each $i \in I_2$,

$$\varphi(g_i(\lambda x^1 + (1 - \lambda)x^2)) = \lambda \varphi(g_i(x^1)) + (1 - \lambda)\varphi(g_i(x^2))$$  \hspace{1cm} (9.2.25)

is valid. Using that $g_i$ is convex, $\varphi$ is increasing and $x^1, x^2 \in K$, we obtain

$$\varphi(\lambda g_i(x^1) + (1 - \lambda)g_i(x^2)) \geq \lambda \varphi(g_i(x^1)) + (1 - \lambda)\varphi(g_i(x^2)), \quad i \in I_2.$$  \hspace{1cm} (9.2.26)

The strict convexity of $\varphi$ implies, that only the equality is possible in (9.2.26), and $g_i(x^1) = g_i(x^2)$ holds for each $i \in I_2$. Finally, in view of (9.2.25) and the increase of $\varphi$ one gets for $i \in I_2$

$$g_i(\lambda x^1 + (1 - \lambda)x^2) = \lambda g_i(x^1) + (1 - \lambda)g_i(x^2),$$

which contradicts the strict convexity of $\max_{i \in I_2} g_i$.

(b) $I_2 = \emptyset$, i.e. all functions $g_i$ are affine:

$$g_i(x) := \langle a^i, x \rangle - b_i, \quad i \in I_1 = I.$$

Then, due to Slater’s condition and the boundedness of $K$, the rank of the matrix $A = \{a^i\}_{i \in I}$ equals $n$. Using this fact, the strict convexity of the function $h$ can be easily established following [289].

Thus, taking into account also the differentiability properties of $g_i$ and $\varphi$, assumptions (B1) and (B2) for the function $h$ are always valid.

• Of course, assumption (B3) is guaranteed if $K$ is a bounded set. In case of an unbounded $K$, (B3) is evident if $\gamma := 2$ in (9.2.24); for arbitrary $\gamma > 1$ condition (B3) follows from Proposition 2 in [229] and

$$D_h(x, y) \geq h_0(x) - h_0(y) - \langle \nabla h_0(y), x - y \rangle,$$

(where $h_0(x) := \sum_{i=1}^n |x_i|^\gamma$).

• Now, we check the validity of assumption (B4). Denote

$$I_<(y) := \{i : g_i(y) < 0\}, \quad I_=(y) := \{i : g_i(y) = 0\},$$

and let $\{z_k\} \subset \text{int}K$, $\lim_{k \to \infty} z_k = z$.

(a) At first, suppose that $I_<(z) \supset I_2$. For $i \in I_<(z)$, one gets

$$\lim_{k \to \infty} g_i(z_k) = g_i(z) < 0 \quad \text{and} \quad \lim_{k \to \infty} \langle \nabla g_i(z_k), z - z_k \rangle = 0,$$

whereas for $i \in I_=(z)$ we have $\lim_{k \to \infty} g_i(z_k) = g_i(z) = 0$, $b_i = \langle a^i, z \rangle$. Thus, for $i \in I_=(z)$,

$$\lim_{k \to \infty} \varphi'(\langle a^i, z_k \rangle - b_i)\langle a^i, z - z_k \rangle = \lim_{k \to \infty} \varphi'(\langle a^i, z_k \rangle - z)\langle a^i, z - z_k \rangle = 0$$

follows from $\langle a^i, z_k - z \rangle \to 0$ and (9.2.22). Using these relations and the identity

$$D_h(x, y) = h(x) - h(y) - \sum_{i \in I_=(y)} \varphi'(\langle a^i, y \rangle - b_i)\langle a^i, x - y \rangle$$

$$- \sum_{i \in I_< (y)} \varphi'(g_i(y))\langle \nabla g_i(y), x - y \rangle - \theta(\nabla h_0(y), x - y),$$
one can easily conclude that \( \lim_{k \to \infty} D_h(z, z^k) = 0 \).

(b) Now, let \( g_i(z) = 0 \) be valid for some \( i_0 \in I_2 \). Denote

\[
I_2(y) := \{ i \in I_2 : g_i(y) = \max_{j \in I_2} g_j(y) \}
\]

and take \( \tilde{z} \in K, \tilde{z} \neq z \). In view of the convexity of the functions \( \varphi \circ g_i \) and \( h_0 \) it holds

\[
D_h(\tilde{z}, z^k) \geq \varphi(g_i(\tilde{z})) - \varphi(g_i(z^k)) - \varphi'(g_i(z^k)) \langle \nabla g_i(z^k), \tilde{z} - z^k \rangle.
\] (9.2.27)

Obviously, relation (9.2.23) implies

\[
\lim_{k \to \infty} \varphi'(g_i(z^k)) = +\infty,
\] (9.2.28)

whereas

\[
\lim_{k \to \infty} \langle \nabla g_i(z^k), \tilde{z} - z^k \rangle = \langle \nabla g_i(z), \tilde{z} - z \rangle.
\] (9.2.29)

But, using the structure of the subdifferential of a max-function and the strict convexity of \( \max_{i \in I_2} g_i \), we obtain

\[
\max_{i \in I_2} g_i(\tilde{z}) - \max_{i \in I_2} g_i(z) > \langle \nabla g_j(z), \tilde{z} - z \rangle, \quad \forall j \in I_2(z).
\]

Because \( \tilde{z}, z \in K \) and \( g_{i_0}(z) = 0 \), the last inequality yields

\[
\langle \nabla g_j(z), \tilde{z} - z \rangle < 0, \quad j \in I_2(z).
\] (9.2.30)

The relations (9.2.27)-(9.2.30) and the continuity of \( \varphi \circ g_i \) ensure that

\[
\lim_{k \to \infty} D_h(\tilde{z}, z^k) = +\infty.
\] (9.2.31)

In fact, we have proved a stronger property than (B4)(ii): relation (9.2.31) holds for each \( \tilde{z} \in K, \tilde{z} \neq z \), whereas (B4)(ii) supposes

\[
\lim_{k \to \infty} D_h(\tilde{z}, z^k) = +\infty \quad \text{if} \quad \tilde{z} \in \text{bd}K, \tilde{z} \neq z.
\]

To prove the fulfillment of assumption (B5), we use Theorem 4.5 in [35], which states, in particular, that

For a function \( f \) with properties (B1) and (B2), boundary coercivity implies zone coercivity if \( K \) is bounded or the super-coercivity condition

\[
\lim_{x \in K, \|x\| \to \infty} f(x) = +\infty
\]
is valid.

Let \( z \in \text{bd}K, \{z^k\} \in \text{int}K, \lim_{k \to \infty} z^k = z \) and \( x \in \text{int}K \). Then

\[
\lim_{k \to \infty} \langle \nabla g_i(z^k), x - z^k \rangle = \langle \nabla g_i(z), x - z \rangle,
\]
and for \( i \in I_2(z) \) the relations

\[
\langle \nabla g_i(z), x - z \rangle < 0, \quad \lim_{k \to \infty} \varphi'(g_i(z^k)) = +\infty
\]
are obvious. Hence, if $i \in I(z)$, then
\[ \lim_{k \to \infty} \varphi'(g_i(z^k)) \langle \nabla g_i(z^k), x - z^k \rangle = -\infty, \] (9.2.32)
whereas for $i \in J(z)$ it holds
\[ \lim_{k \to \infty} \varphi'(g_i(z^k)) \langle \nabla g_i(z^k), x - z^k \rangle = \varphi'(g_i(z)) \langle \nabla g_i(z), x - z \rangle. \] (9.2.33)

From (9.2.32) and (9.2.33) and the continuous differentiability of the function $h_0$, we immediately conclude that the function $h$ is boundary coercive, hence according to Theorem 4.5 in [35], $h$ satisfies assumption (B5). □

Particular functions satisfying the conditions of Theorem 9.2.12 are
\[ \varphi(t) := -(t)^p, \quad p \in (0, 1) \text{ arbitrarily chosen}, \] (9.2.34)
\[ \varphi(t) := \begin{cases} -t \ln(-t) + t & \text{if } -\frac{1}{2} \leq t \leq 0, \\ -2 \ln(2) - \frac{1}{2} \ln 2 - \frac{1}{2} t & \text{if } t < -\frac{1}{2} \end{cases}, \] (9.2.35)
\[ \varphi(t) := -t \ln(-t) + t \ln(-t + 1) + t, \] (9.2.36)
where by convention $\varphi(0) = 0$.

The principal idea for constructing function (9.2.36) consists in the following: Constraints, which turn out to be active at the limit point, are handled mainly by the first term in $\varphi$, whereas those which become inactive at the limit point, are observed by both terms of $\varphi$.

Combining Theorems 9.2.11 and 9.2.12 we immediately obtain the following result.

9.2.13 Theorem. Let the set $K$ be described by (9.2.20), (9.2.21) and the function $h$ be defined as in Theorem 9.2.12. Moreover, suppose that $VI(Q, K)$ satisfies Assumption 9.2.1. Then, the sequence $\{x^k\}$, generated by Method 9.2.7, belongs to $\text{int}K$ and converges to $x \in \text{SOL}(Q, K)$.

9.2.14 Corollary. Suppose that the hypothesis of Theorem 9.2.13 are fulfilled. If $VI(Q, K)$ has more than one solution, then the sequence $\{x^k\}$, generated by Method 9.2.7 with Bregman function (9.2.24), converges to a solution $x$ such that $g_i(x) < 0$, $i \in I_2$.

Indeed, the opposite assumption that $g_i(x) = 0$ holds for some $i \in I_2$ leads immediately to the contradiction between the statement (i) of Lemma 9.2.8 and relation (9.2.31) given with $z^k := x^k$ and $\bar{z} \in \text{SOL}(Q, K)$, $\bar{z} \neq x$.

9.2.15 Remark. If functions $g_i$ in (9.2.20) satisfy
\[ g_i(x) \geq -1 \quad \forall x \in K, \forall i \in I_2, \]
one can quite similarly prove that the function $h_l : K \to \mathbb{R}$, defined by
\[ h_l(x) := \sum_{i \in I} \lbrack (-g_i(x)) \ln(-g_i(x)) + g_i(x) \rbrack + \theta \sum_{j=1}^{n} |x_j|^\gamma \] (9.2.37)
(with $\theta, \gamma$ as in (9.2.24)), possesses the properties (B1)-(B5), too. ♦
The following example indicates that assumption (B4) may be violated for function (9.2.24) if the function \( \max_{i \in I_2} g_i \) is not strictly convex.

**Example.** Let 
\[
K := \{ x \in \mathbb{R}^4 : x_1^2 + x_2^2 + x_3 + x_4 \leq 1 \}.
\]
With \( g(x) := x_1^2 + x_2^2 + x_3 + x_4 - 1 \), take (9.2.24), (9.2.35), i.e.,
\[
h(x) := -(g(x))^p + \sum_{j=1}^4 x_j^2, \quad p \in (0, 1),
\]
and \( \{z^k\} \) such that
\[
\begin{align*}
\text{if } k := 2l - 1: & \quad z_1^k = \sqrt{1 - \sigma_{k-1}^p - (1 - \sigma_{k-1}^p - \sigma_k)^2}, \\
& \quad z_2^k = 1 - \sigma_{k-1}^p - \sigma_k, \ z_3^k = z_4^k = 0;
\end{align*}
\]
\[
\begin{align*}
\text{if } k := 2l: & \quad z_1^k = 0, \ z_2^k = 1, \ z_3^k = -\sigma_k, \ z_4^k = 0, \quad \sigma_k > 0, \lim_{k \to \infty} \sigma_k = 0.
\end{align*}
\]
Then \( z = \lim_{k \to \infty} z^k = (0, 1, 0, 0)^T \) and for large \( l \)
\[
(-g(z^k))^{p-1} \langle \nabla g(z^k), z - z^k \rangle = -2, \quad \text{if } k = 2l - 1.
\]
But, choosing \( \bar{z} := (0, 1, -1, 1)^T \), one gets
\[
(-g(z^k))^{p-1} \langle \nabla g(z^k), \bar{z} - z^k \rangle = \sigma_k^p, \quad \text{if } k = 2l.
\]
Thus, neither (B4)(i) nor (B4)(ii) is valid. \( \diamond \)

**9.2.16 On the weakening of condition (9.2.20)**

Now instead of (9.2.20) we are going to consider a condition saying that the boundary of the feasible set which is described by nonlinear constraints does not contain a line segment. More precisely we assume that

\[ g_i (i \in I_2) \text{ are convex, continuously differentiable functions, and} \]
\[ K_{bd} := \{ y \in K : \max_{i \in I_2} g_i(y) = 0 \}, \quad (9.2.38) \]
does not contain any line segment.

As we have seen in the proof of Theorem 9.2.13 the Assumptions 9.2.2 (B1), (B2), (B3) and (B5) are fulfilled. This proof establishes also that the standard convergence sensing condition (B4)(i) is certainly valid if \( g_i(z) < 0 \ \forall \ i \in I_2 \), i.e., if \( z \in K \setminus K_{bd} \). In order to check the convergence sensing conditions (B4)(ii) in case \( \lim_{k \to \infty} z^k = z \in K_{bd} \), we need the following statement.

**9.2.17 Lemma.** The following conclusions are equivalent:

(i) assumption (9.2.38) is valid;

(ii) \( z \in K_{bd} \) implies

\[
\min_{i \in I_2(z)} \langle \nabla g_i(z), x - z \rangle < 0, \quad \forall \ x \in \text{bd}K, \ x \neq z,
\]

where \( I_2(z) := \{ i \in I_2 : \ g_i(z) = 0 \} \);
(iii) $z \in \bar{K}_{bd}$ implies

$$
(\nabla g_i(z), x - z) < 0, \ \forall \ x \in \text{bd}K, \ x \neq z, \ \forall \ i \in I_2(z).
$$

**Proof:** (iii) $\Rightarrow$ (ii) is obvious.

(ii) $\Rightarrow$ (i): Assume that $\bar{K}_{bd}$ contains a line segment $[a, b]$. Then for $z := \frac{1}{2}(a + b) \in \bar{K}_{bd}$, due to $b \in \text{bd}K$, $b \neq z$ and (ii), we obtain

$$
\exists \ i_0 \in I_2(z) : \ (\nabla g_{i_0}(z), b - z) < 0.
$$

This implies

$$
0 < - (\nabla g_{i_0}(z), b - z) = (\nabla g_{i_0}(z), \frac{a - b}{2}) = (\nabla g_{i_0}(z), a - z),
$$

hence

$$
g_{i_0}(a) - g_{i_0}(z) > 0,
$$

and $g_{i_0}(z) = 0$ leads to $g_{i_0}(a) > 0$, which is a contradiction to $a \in \bar{K}_{bd}$.

(i) $\Rightarrow$ (iii): Let be $z \in \bar{K}_{bd}$. If $x \in \bar{K}_{bd}$, $x \neq z$, then $v := \frac{1}{2}(z + x) \in K \setminus \bar{K}_{bd}$ and for arbitrary $i \in I_2(z)$ one gets

$$
g_i(v) - g_i(z) \geq \frac{1}{2}(\nabla g_i(z), x - z).
$$

Taking into account that $g_i(v) < 0$ and $g_i(z) = 0$, this implies

$$
(\nabla g_i(z), x - z) < 0.
$$

But, if $x \in \text{bd}K \setminus \bar{K}_{bd}$, then $g_i(x) < 0$ holds for all $i \in I_2$, and the inequality

$$
g_i(x) - g_i(z) \geq (\nabla g_i(z), x - z)
$$

also guarantees that

$$
(\nabla g_i(z), x - z) < 0, \ \forall \ i \in I_2(z).
$$

\[ \square \]

Now, let the sequence $\{z^k\} \subset \text{int}K$ converge to $z \in \bar{K}_{bd}$ and assume that condition (9.2.38) is valid. We show that

$$
\lim_{k \to \infty} D_h(\bar{z}, z^k) = +\infty \quad (9.2.39)
$$

holds if $\bar{z} \neq z$ and $\bar{z} \in \text{bd}K$. This will immediately imply the fulfillment of the modified convergence sensing conditions (B4).

In view of $z \in \bar{K}_{bd}$, the equality $g_{i_0}(z) = 0$ is valid for some $i_0 \in I_2$. From the convexity of the functions $\varphi \circ g_i$ and $x \to \sum |x_i|^\gamma$ it follows

$$
D_h(\bar{z}, z^k) \geq \varphi(g_{i_0}(\bar{z})) - \varphi(g_{i_0}(z^k)) - \varphi'(g_{i_0}(z^k))(\nabla g_{i_0}(z^k), \bar{z} - z^k). \quad (9.2.40)
$$

Obviously, the relation $\lim_{t \to 0} \varphi'(t) = +\infty$ provides

$$
\lim_{k \to \infty} \varphi'(g_{i_0}(z^k)) = +\infty, \quad (9.2.41)
$$
whereas
\[ \lim_{k \to 0} \varphi(g_{i_0}(z^k)) = 0, \]
\[ \lim_{k \to \infty} \langle \nabla g_{i_0}(z^k), \bar{z} - z^k \rangle = \langle \nabla g_{i_0}(z), \bar{z} - z \rangle. \]  
(9.2.42)

But, according to Lemma 9.2.17,
\[ \langle \nabla g_{i_0}(z), \bar{z} - z \rangle < 0, \]  
(9.2.43)
and (9.2.40)-(9.2.43) yields immediately the fulfillment of (9.2.39).

Therefore, all assumptions on Bregman functions made are valid. Hence, the convergence results proved there remain true under the use of the weaker assumption (9.2.38) on functions \( g_i \).

The following simple example shows that the conditions (9.2.38) on the functions \( g_i \) (\( i \in I_2 \)) are indeed essentially weaker than the conditions assumed in (9.2.20).

**9.2.18 Example.** The set \( K = \{ x \in \mathbb{R}^2 : g_i(x) \leq 0 \} \), \( g(x) = -x_1 + x_2^2 \), satisfies Assumption (9.2.38), whereas the related condition (9.2.20) is evidently violated. Moreover, a comparison with Example 9.2.16 points to the fact that there are hardly any chances for a further weakening of the conditions (9.2.38).

Assumption (9.2.38) does not entail any geometrical peculiarities of the function \( \max_{i \in I_2} g_i \) (cf. (9.2.3)) in \( \text{int} K \). In particular, considering
\[ K = \{ x \in \mathbb{R}^2 : g_i(x) \leq 0, i = 1, 2 \} \]
with
\[ g_1(x) = x_1^2 + x_2^2 - 1, \quad g_2(x) = e^{x_1} - 1, \]
we meet the situation that Assumption (9.2.38) is valid, but for arbitrary \( x^0 \in \text{int} K \) and
\[ \ell : \quad \max\{g_1(x^0), g_2(x^0)\} < \ell < 0 \]
the boundary of the level set
\[ \{ x \in \mathbb{R}^2 : \max\{g_1(x), g_2(x)\} = \ell \} \]
contains a line segment.

\[ \diamond \]

**9.2.3 Embedding of original Bregman-function-based methods**

Let us analyze the original proof technique of Bregman-function-based proximal methods destined to variational inequalities on polyhedral sets. Denote by \( D \) a distance function and by \( v^f \) an iterate of such a method. To our knowledge, original convergence results for these methods establish convergence of the sequence \( \{D(x, v^f)\} \) for each solution \( x \) without the use of assumption Stand-(B4).

Hence, for interior proximal methods involving *standard* requirements on Bregman functions (see, in particular, [61, 97, 376]), the original convergence analysis can be preserved (with minor alterations in the final stage only) if we replace
Stand-(B4) by (B4) with $\bar{z} \in K$ instead of $\bar{z} \in \text{bd} K$ in (B4)(ii). Let us recall that, for $K$ given by (9.2.20), (9.2.21), the conditions of Theorem 9.2.12 ensure the fulfillment of the so modified assumption (B4).

Indeed, let a subsequence $\{v^{\ell_k}\}$ of $\{v^\ell\}$ with $\lim_{k \to \infty} v^{\ell_k} = v$ be chosen such that there exists $x \in \text{SOL}(Q, K)$, $x \neq v$ (if this is impossible, clearly $\lim_{\ell \to \infty} v^\ell = v$ and $\text{SOL}(Q, K) = \{v\}$). The convergence of $\{D(x, v^\ell)\}$ implies that condition (B4)(ii) with $\bar{z} := x$, $z := v$, $z^k := v^{\ell_k}$ is violated. Hence, (B4)(i) is valid, i.e. $\lim_{k \to \infty} D(v, v^{\ell_k}) = 0$, and (B4) turns into Stand-(B4). Referring now to the original convergence results, we immediately conclude that $v \in \text{SOL}(Q, K)$. Then, convergence of $\{D(v, v^\ell)\}$ implies $\lim_{\ell \to \infty} D(v, v^\ell) = 0$ and Theorem 2.4 in [376] yields $\lim_{\ell \to \infty} v^\ell = v$.

On this way, inserting the function (9.2.24) (for instance, with $\varphi$ defined by (9.2.35) or (9.2.36)) in methods where Bregman functions are described exactly by standard conditions, one can extend these methods to solve VI($Q, K$) on non-polyhedral sets $K$ given by (9.2.35)–(9.2.36).

Return to the convergence analysis in Section 9.2.2. For Method 9.2.7 without an approximation of the operator $Q$, i.e. with $Q^k \equiv Q$, the same arguments as given just above permit us to weaken Assumption 9.2.1 (A4), replacing (A4)(b) by the condition that $Q$ is a pseudo-monotone operator in the sense of Brézis-Lions [270], here boundedness of the operator is included in this notion. With this alteration, the proof of Lemma 9.2.10 can be performed quite similarly to the proof of the case (c) in Lemma 9.3.10 in Subsection 9.3.2.

Moreover, for the exact version of Method 9.2.7, with $Q^k \equiv Q$, $v^k \equiv 0$, assumption (A4)(b) can be omitted at all according to the analysis of Solodov and Svaiter [376].

9.2.3.1 On entropy-like and logarithmic-quadratic PPR

We have tried to extend in a similar manner entropy-like and logarithmic-quadratic proximal methods (see [22, 25, 26, 386]). Applications of these methods to the dual of linearly constrained programs or to variational inequalities on polyhedral convex sets provide attractive properties of subproblems (for instance, $C^\infty$ multiplier methods with bounded Hessians are constructed on this way in [25]). However, the basic requirements on the kernel functions in these methods seem to be not appropriate for such an extension. Indeed, in case $K := \mathbb{R}^m_+$ the corresponding distance functions have the form

$$d(u, v) = \sum_{i=1}^m v_i^\alpha \left[ \varphi \left( \frac{u_i}{v_i} \right) + \beta \left( \frac{u_i}{v_i} - 1 \right)^2 \right]$$

(9.2.44)

($\alpha := 1$, $\beta := 0$ in entropy-like methods and $\alpha := 2$, $\beta > 0$ in logarithmic-quadratic methods). If

$$K := \{ x \in \mathbb{R}^n : g_i(x) := \langle a_i, x \rangle - b_i \leq 0, \ i = 1, \ldots, m \},$$

the distance $D$ is given by

$$D(x, y) := d(-g(x), -g(y)), \ g = (g_1, \ldots, g_m)^T.$$  

(9.2.45)
Concerning the kernel function $\phi$, it is supposed, in particular, that $\text{dom}\phi \subseteq [0, +\infty)$, $\phi$ is twice continuously differentiable on $(0, +\infty)$ and strictly convex on its domain, and

$$\phi(1) = \phi'(1) = 0, \quad \phi''(1) > 0.$$  \hfill (9.2.46)

Already these conditions enforce that the function $D(\cdot, y)$ defined by (9.2.45) is, in general, non-convex for some $y \in \text{int}K$, if at least one function $g_i$ is not affine as the following example shows.

9.2.19 Example. Let the kernel function $\phi$ with the mentioned properties be fixed, and $K := \{x \in \mathbb{R}^1 : g(x) \leq 0\}$, where the choice of the convex and sufficiently smooth function $g : \mathbb{R}^1 \to \mathbb{R}^1$ will be specified below. According to (9.2.44), (9.2.45)

$$D(x, y) = (-g(y))^\alpha \left[ \phi \left( \frac{g(x)}{g(y)} \right) + \frac{\beta}{2} \left( \frac{g(x)}{g(y)} - 1 \right)^2 \right].$$

Assume first that $\phi(\frac{1}{2}) \leq \phi(\frac{3}{2})$ and take $g$ with values

$$g(x^1) := -\frac{1}{2}, \quad g(1) := -\frac{3}{2}, \quad g(x^2) := -1,$$

where $x^1, x^2$ are given points, such that $(x^1 - 1)(x^2 - 1) < 0$. Then, for $y := x^2$ one gets

$$D(x^1, y) = \phi(\frac{1}{2}) + \frac{\beta}{8}, \quad D(1, y) = \phi(\frac{3}{2}) + \frac{\beta}{8}, \quad D(x^2, y) = 0.$$

Because of $\phi(\frac{1}{2}) \leq \phi(\frac{3}{2})$ and $\phi(t) \geq 0 \forall t \in \text{dom}\phi$ (the last inequality follows from the convexity of $\phi$ and (9.2.46)), we conclude that the functions $D(\cdot, y)$ and $\phi \circ (-g)$ are not convex.

But, if $\phi(\frac{1}{2}) > \phi(\frac{1}{2})$, the same conclusion holds true with $y := x^1$ and

$$g : \quad g(x^1) = -1, \quad g(1) = -\frac{1}{2}, \quad g(x^2) = -\frac{3}{2}.$$  \hfill ♦

9.3 PPR with Generalized Distance Functionals

The purpose of this subsection is a uniform approach for analyzing convergence of proximal-like methods for solving variational inequalities in Hilbert spaces with non-quadratic distance functionals. In comparison with preceding publications dealing with such non-quadratic regularization terms, here the standard requirement of the strict monotonicity of the operator $\nabla_1 D(\cdot, y)$ (usually formulated as the strict convexity of Bregman’s or an other function generating the distance $D$) is weakened. This leads to an analogy of methods with weak regularization and regularization on a subspace developed on the basis of the classical PPR in Section 8.2.3. Besides a successive approximation of the set $K$ is included.
Let \((V, \|\cdot\|)\) be a Hilbert space with the topological dual \(V'\) and the duality pairing \(\langle \cdot, \cdot \rangle\) between \(V\) and \(V'\). We consider the variational inequality

\[
\text{VI}(Q, K) \quad \text{Find a pair } x^* \in K \text{ and } q^* \in Q(x^*) \text{ such that } \langle q^*, x - x^* \rangle \geq 0 \quad \forall \, x \in K,
\]

where \(K \subset V\) is a convex closed set and \(Q : V \to 2^{V'}\) is a maximal monotone operator.

We generally suppose that \(\text{VI}(Q, K)\) is solvable and denote by \(\text{SOL}(Q, K)\) its solution set.

The exact proximal point method, applied to the variational inequality \(\text{VI}(Q, K)\), can be described as follows:

\[
\text{Given } x^0 \in K \text{ and a sequence } \{\chi_k\}, \, 0 < \chi_k \leq \bar{\chi} < \infty; \quad x^{k+1} \in K \text{ is defined such that } \\
\exists \, q(x^{k+1}) \in Q(x^{k+1}) : \\
\langle q(x^{k+1}) + \chi_k \nabla_1 D(x^{k+1}, x^k), y - x^{k+1} \rangle \geq 0 \quad \forall \, y \in K,
\]

where \(D(x, y) = \frac{1}{2}\|x - y\|^2\) and \(\nabla_1\) is the partial gradient w.r.t. \(x\).

For different modifications of the PPR, also with other quadratic functionals \(D\), we address the reader to [224] and [241], where numerous references can be found.

The scheme studied here can be described as follows:

Taking a linear monotone operator \(B : V \to V'\) such that \(Q - B\) is still monotone, we choose a convex continuous functional \(h : \bar{S} \to \mathbb{R}\) so that \(x \mapsto \frac{1}{2}\langle Bx, x \rangle + h(x)\) possesses properties like usually required for Bregman functions (with zone \(S\)).

At the \(k\)-th step, with \(x^k \in K_{k-1} \cap \bar{S}\) obtained at the previous iteration, the iterate \(x^{k+1} \in K_k \cap \bar{S}\) is defined such that

\[
\exists \, q(x^{k+1}) \in Q(x^{k+1}) : \\
\langle q(x^{k+1}) + \chi_k (\nabla h(x^{k+1}) - \nabla h(x^k)), x - x^{k+1} \rangle \\
\geq -\delta_k \sqrt{\Gamma_1(x, x^{k+1})} \quad \forall \, x \in K_k \cap \bar{S}.
\]

(9.3.1)

Here, \(\{K_k\}\) is a sequence of convex sets approaching \(K\), \(\{\chi_k\}\) chosen as above, \(\{\delta_k\} \downarrow 0\) is a given non-negative sequence and

\[
\Gamma_1(x, y) := \min\{\alpha \|x - y\|^2, \Gamma(x, y) + 1\}, \quad \alpha > 0 - \text{const.},
\]

(9.3.2)

with

\[
\Gamma(x, y) := \frac{1}{2}\langle B(x - y), x - y \rangle + h(x) - h(y) - \langle \nabla h(y), x - y \rangle
\]

(9.3.3)

considered on \(\text{dom} \Gamma = \bar{S} \times D(\nabla h)\) and used below as a Lyapunov function.
This scheme and the required conditions on \( h \) in the following Subsection 9.3.1 do not exclude the use of quadratic functionals \( h \), in particular, the choice \( h(x) = \frac{1}{2}\|x\|^2 \) leads to a perturbed version of the classical proximal point method (for this version with more general assumptions w.r.t. data approximation see [225]). Therefore, in the sequel the notion “non-quadratic” indicates the predominant aspect of this investigation.

9.3.1 Generalized proximal point method

In the sequel we make use of the following notations: \( S \subset V \) is an open convex set, its closure is denoted by \( \text{cl}S \); \( \{K_k\} \subset V \), \( K_k \supset K \), is a family of convex closed sets approximating \( K \);

\[
\mathcal{N}_K : y \rightarrow \begin{cases} \{ z \in V' : \langle z, y - x \rangle \geq 0 \forall x \in K \} & \text{if } y \in K \\ \emptyset & \text{otherwise} \end{cases}
\]

is the normality operator for \( K \). With \( B \) and \( h : \text{cl}S \rightarrow \mathbb{R} \) as introduced, we define the functional

\[
\eta(x) = \begin{cases} \frac{1}{2}\langle Bx, x \rangle + h(x) & \text{if } x \in \text{cl}S \\ +\infty & \text{otherwise.} \end{cases} \quad (9.3.4)
\]

Now the following basic assumptions on the successive approximation of \( \text{VI}(Q, K) \) and the choice of the controlling parameters are considered.

9.3.1 Assumption.

(A1) For each \( k \), the operator \( Q + \mathcal{N}_{K_k} \) is maximal monotone;

(A2) \( S \cap D(Q) \cap K_k \neq \emptyset \forall k \);

(A3) For each \( k \) it holds

\[
\langle q(x) - q(y), x - y \rangle \geq \langle B(x - y), x - y \rangle \forall x, y \in D(Q) \cap K_k, \forall q(\cdot) \in Q(\cdot),
\]

where \( B : V \rightarrow V' \) is a given linear continuous and monotone operator with symmetry property \( \langle Bx, y \rangle = \langle By, x \rangle \);

(A4) Any weak limit point of an arbitrary sequence \( \{v^k\} \), \( v^k \in S \cap D(Q) \cap K_k \), belongs to \( K \cap D(Q) \);

(A5) The non-negative sequences \( \{\varphi_k\} \) (accuracy of approximation), \( \{\chi_k\} \) (regularization parameter) and \( \{\delta_k\} \) (exactness for solving the auxiliary problems) satisfy

\[
0 < \chi_k \leq 1, \quad \sum_{k=1}^{\infty} \frac{\varphi_k}{\chi_k} < \infty, \quad \sum_{k=1}^{\infty} \frac{\delta_k}{\chi_k} < \infty;
\]

(A6) For some \( x^* \in \text{SOL}(Q, K) \cap \text{cl}S \) and \( q^*(x^*) \in Q(x^*) \) satisfying

\[
\langle q^*(x^*), y - x^* \rangle \geq 0 \forall y \in K,
\]

and for an arbitrary sequence \( \{v^k\} \), \( v^k \in S \cap D(Q) \cap K_k \), there exists a sequence \( \{w^k(v^k)\} \subset K \cap S \) such that

\[
\langle q^*(x^*), w^k(v^k) - v^k \rangle \leq c \left( \Gamma(x^*, v^k) + 1 \right) \varphi_k \quad (c \geq 0 - \text{const.}) \quad (9.3.5)
\]
Condition (A6) seems to be rather artificial, especially, due to the unknown element $q^*(x^*)$. However, for a series of variational inequalities in mechanics and physics, we have a helpful \textit{a priori} information about $q^*(x^*)$. For instance, for the problem on a steady movement of a fluid in a domain $\Omega$ bounded by a semi-permeable membrane (see [135], Sect.1) $q^*(x^*) = 0$ has to be.

In the general situation, one can replace (9.3.5) by

$$\|w^k(v^k) - v^k\| \leq c_1 \varphi_k.$$  

Because in (9.3.5) $c$ is an arbitrary (non-negative) constant, this causes no alterations in the analysis below.

9.3.2 Remark. In [224] and [225], using the functional $h$ such that

$$\exists m > 0 : \Gamma(x,y) \geq m\|x - y\|^2, \forall x, y,$$

more general approximations have been considered ($K_k \supset K$ is not supposed), mainly inspired by finite element methods in mathematical physics. Here we renounce it in order to avoid too much technicalities.

Now we are going to define the \textit{generalized distance} functional $\Gamma$.

9.3.3 Assumption.

(B1) $h : \text{cl}S \to \mathbb{R}$ is a convex and continuous functional;

(B2) $h$ is Gâteaux-differentiable on $S$;

(B3) The functional $\eta$ in (9.3.4) is strictly convex on $\text{cl}S$;

(B4) $\text{SOL}(Q, K) \cap \text{cl}S \neq \emptyset$;

(B5) $L_1(x, \delta) = \{y \in S : \Gamma(x,y) \leq \delta\}$ is bounded for each $x \in \text{cl}S$ and each $\delta$;

(B6) If $\{v^k\} \subset S$, $\{y^k\} \subset S$ converge weakly to $v$ and $\lim_{k \to \infty} \Gamma(v^k, y^k) = 0$, then

$$\lim_{k \to \infty} \left[ \Gamma(v, v^k) - \Gamma(v, y^k) \right] = 0;$$

(B7) If $\{v^k\} \subset S$ is bounded, $\{y^k\} \subset S$, $y^k \rightharpoonup y$ and $\lim_{k \to \infty} \Gamma(v^k, y^k) = 0$, then

$$\lim_{k \to \infty} \|v^k - y^k\| = 0;$$

(B8) If $\{v^k\} \subset S$, $\{y^k\} \subset S$, $v^k \rightharpoonup v$, $y^k \rightharpoonup y$ and $v \neq y$, then

$$\lim_{k \to \infty} \left| \langle \nabla h(v^k) + Bv^k - \nabla h(y^k) - By^k, v - y \rangle \right| > 0;$$

(B9) $\forall z \in V' \exists x \in S : \nabla h(x) + Bx = z.$
In [219], Sect.5, for two problems in elasticity theory the chosen regularizing functionals satisfy the Assumptions 9.3.3 and Assumption 9.3.1 (A3), but they are not strictly convex (see, also Method 8.2.5 (regularization on the kernel)).

As it can be concluded from [65], Sect.2.1.2, Assumption 9.3.3 (B7) is equivalent to the following sequential consistency property for the functional $\eta$ on $S$:

$$\inf_{x \in E} \inf_{y \in \text{cl} S, \|y - x\| = t} \left\{ \eta(y) - \eta(x) - (\nabla \eta(x), y - x) : y \in \text{cl} S, \|y - x\| = t \right\} > 0.$$ 

For the case $B = 0$, the totality of Assumptions 9.3.3 (B1)-(B9) is similar to the system of hypothesizes for Bregman functions considered Burachik and Iusem [59], only (B7) here is stronger than the corresponding assumption in the paper mentioned, where $y_k \rightharpoonup \bar{y}$ stands in place of $\lim_{k \to \infty} \|v_k - y_k\| = 0$. In the cases (b) and (c) (see Lemma 9.3.10 and Theorem 9.3.11 below), this assumption from [59] suffices if $B$ is a compact operator. At the same time, the use of (B7) permits us, in particular, to extend the class of operators $Q$ by including the case (a) (again, see Lemma 9.3.10 and Theorem 9.3.11 below).

Let us give a simple example illustrating the choice of the functional $h$.

Let $V := \mathbb{R}^n$, $K := \{ x \in \mathbb{R}^n : x_j \geq 0, j = 1, ..., n_1; \sum_{j=n_1+1}^{n_2} j|x_j| \leq 1 \}$

with $0 < n_1 < n_2 < n$,

$$Q : x \to (A(x_1, ..., x_{n_1}), x_{n_1+1} - 1, ..., x_n - 1),$$

where $A : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ is an arbitrary continuous and monotone operator such that the corresponding VI$(Q, K)$ is solvable. Then, considering the approximation

$$K_k := \{ x \in \mathbb{R}^n : x_j \geq 0, j = 1, ..., n_1; \sum_{j=n_1+1}^{n_2} j \sqrt{x_j^2 + \tau_k} \leq 1 + \sqrt{\tau_k} \sum_{j=n_1+1}^{n_2} j \},$$

where $\tau_k \downarrow 0$, take

$$B : x \to (0, ..., 0, x_{n_1+1}, ..., x_n).$$

Since

$$K_k \subset \{ x \in \mathbb{R}^n : x_j \geq 0, j = 1, ..., n_1; \sum_{j=n_1+1}^{n_2} j|x_j| \leq 1 + \sqrt{\tau_k} \sum_{j=n_1+1}^{n_2} j \},$$
the estimate
\[ \min_{z \in K} \|z - v^k\| \leq c_1 \sqrt{\tau_k} \quad \forall \{v^k\}, \quad v^k \in K_k, \]
is evident.

Now, it is easy to verify that the choices of \{K_k\},
\[ h(x) := \sum_{j=1}^{n_1} x_j \ln x_j - x_j \quad \text{(with } 0 \times \ln 0 = 0 \text{ by convention}) \]
and
\[ S := \{x \in \mathbb{R}^n : x_j > 0, j = 1, \ldots, n_1\} \]
satisfy the Assumptions 9.3.1 (A1)-(A4), Assumptions 9.3.3(B1)-(B9). Assumption 9.3.1 (A6) is fulfilled with \[\phi_k = \sqrt{\tau_k}\], whereas the second condition in (A5) forces that \[\sum_{k=1}^{\infty} \frac{\sqrt{\tau_k}}{\chi_k} < \infty\].

Now, let us recall the method under consideration.

9.3.4 Method. (Generalized proximal point method) (GPPM):

Given \(x^1 \in S\) and \(x^k \in K_{k-1} \cap S\) (at the \((k-1)\)-th step). At the \(k\)-th step solve

\[ (P_{0k}^k) \quad \text{find} \quad x^{k+1} \in K_k \cap \text{cl} S \quad \text{such that} \]
\[ \exists \ g(x^{k+1}) \in Q(x^{k+1}) : \]
\[ \langle g(x^{k+1}) + \chi_k (\nabla h(x^{k+1}) - \nabla h(x^k)) , x - x^{k+1} \rangle \geq -\delta_k \sqrt{\Gamma_1(x, x^{k+1})} \forall x \in K_k \cap \text{cl} S. \] (9.3.6)

We denote by \((P_0^k)\) Problem (9.3.6) with \(\delta_k = 0\) and by \(\bar{x}^{k+1}\) its solution.

However, the criterion for the approximate calculation of \(x^{k+1}\) in \((P_{0k}^k)\) is not suitable for a straightforward use, but it permits one to extend the convergence results, obtained here, to related algorithms with more reasonable criteria.

Eckstein [97] has analyzed different accuracy conditions of Bregman-function-based methods for the inclusion

\[ \text{find} \quad z \in \mathbb{R}^n : \quad 0 \in Tz, \] (9.3.7)

with \(T : \mathbb{R}^n \to 2^{\mathbb{R}^n}\) a maximal monotone operator.

The method, studied in [97] under rather standard assumptions on a Bregman function \(g\), has the form

\[ 0 \in x^{-1}_k T(x^{k+1}) + \nabla g(x^{k+1}) - \nabla g(x^k) + e^{k+1}, \] (9.3.8)

and solves VI\((Q, K)\) if \(T := Q + N_K\) and \(Q + N_K\) is maximal monotone (here again, \(e^{k+1}\) is an error vector). A relaxation of the exact inclusion ((9.3.8) given with \(e^{k+1} = 0\)) is considered in [97] as to be preferable for numerical implementations. Convergence of the iterates \(x^k\) generated in (9.3.8) to a solution of (9.3.7) is established under the following conditions on \(\{e^k\}\):

\[ \sum_{k=1}^{\infty} \|e^k\| < \infty \] (9.3.9)
and

\[ \sum_{k=1}^{\infty} \langle e^k, x^k \rangle < \infty. \]  

(9.3.10)

Considering in the sequel method (9.3.8) with \( T := Q + N_K \) on a pair \((V, V')\), we have to take \( \|\cdot\|_{V'} \) in (9.3.9). Now we show that condition (9.3.9) suffices to conclude that (9.3.6) is valid if \( \sum_{k=1}^{\infty} \chi_k^{-1} \delta_k < \infty \).

If \( K \cap \text{cl}S \) is a bounded set, then (9.3.10) follows immediately from (9.3.9). In this case, one can easily see that (9.3.8) implies the validity of (9.3.6) given with an appropriate \( \alpha \) in (9.3.2) and \( \delta_k := \alpha^{-1/2} \chi_k \| e^{k+1} \|_{V'} \), such that (9.3.9) provides the condition \( \sum_{k=1}^{\infty} \chi_k^{-1} \delta_k < \infty \) in (A5) (naturally, for this comparison, we assume that \( \mathcal{B} := 0, K_k \equiv K \) and \( g := h \); but the same arguments suit if \( g := h \) in (9.3.8) satisfies Assumption 9.3.1 (A3), Assumption 9.3.3 (B1)-(B3) with \( B \neq 0 \).

Indeed, taking \( \alpha < (2 \text{ diam}(K \cap \text{cl}S))^{-2} \), the equality

\[ \Gamma_1(x, y) = \alpha \|x - y\|^2 \quad \forall x \in K \cap \text{cl}S, \quad y \in K \cap S \]

follows from the definition of \( \Gamma_1 \).

Due to the definition of \( N_K \), one can rewrite (9.3.8) as

\[ \langle \chi_k e^{k+1} + q(x^{k+1}) + \chi_k (\nabla h(x^{k+1}) - \nabla h(x^k)), x - x^{k+1} \rangle \geq 0 \quad \forall x \in K, \]

with \( q(x^{k+1}) \in Q(x^{k+1}) \), hence

\[ \langle q(x^{k+1}) + \chi_k (\nabla h(x^{k+1}) - \nabla h(x^k)), x - x^{k+1} \rangle \geq -\chi_k \| e^{k+1} \|_{V'} \| x - x^{k+1} \| \quad \forall x \in K. \]

In view of the assumption \( \{ x^k \} \subset S \) made in [97] (see also Subsection 9.3.2 below), \( \Gamma_1(x, x^{k+1}) = \alpha \|x - x^{k+1}\|^2 \) holds for \( x \in K \cap \text{cl}S \), and together with the last inequality this yields

\[ \langle q(x^{k+1}) + \chi_k (\nabla h(x^{k+1}) - \nabla h(x^k)), x - x^{k+1} \rangle \geq -\alpha^{-1/2} \chi_k \| e^{k+1} \|_{V'} \| \Gamma_1(x, x^{k+1}) \| \quad \forall x \in K \cap \text{cl}S. \]

(9.3.11)

Therefore, the claim above follows immediately.

Now, let us trace the situation when \( K \) is not bounded, nevertheless condition (9.3.9) ensures (9.3.6) with \( \sum_{k=1}^{\infty} \chi_k^{-1} \delta_k < \infty \) (hence, condition (9.3.10) is superfluous). In this case, however, the use of a suitable operator \( B \neq 0 \) is in essence.

We suppose that \( V := H^1(\Omega) \) (where \( \Omega \) is an open domain in \( \mathbb{R}^n \)), that \( K \subset \{ x \in V : \|x\|_{L^2(\Omega)} \leq 1 \} \) is an unbounded set in \( V \) and \( B : V \to V' \) is given by \( \langle Bx, x \rangle := \| \nabla x \|_{L^2(\Omega)}^2 \). A similar choice of \( B \) is quite realistic for elliptic problems. In this situation, for any functional \( h \) satisfying Assumption 9.3.3 (B1)-(B3),

\[ \Gamma(x, y) + 1 \geq \frac{1}{4} \|x - y\|^2_{L^2(\Omega)} + B(x - y), x - y \]

\[ \geq \frac{1}{4} \|x - y\|^2_{H^1(\Omega)} \quad \forall x \in K \cap \text{cl}S, \quad y \in K \cap S, \]

is valid, hence setting \( \alpha \leq 1/4 \) in (9.3.2), one gets

\[ \Gamma_1(x, y) = \alpha \|x - y\|^2_{H^1(\Omega)} \quad \forall x \in K \cap \text{cl}S, \quad y \in K \cap S. \]
Now, the same arguments as in the case of a bounded set $K$ enable us to conclude that (9.3.8) implies (9.3.6) with $\delta_k := \alpha^{-1/2} \chi_k \|e^{k+1}\|_{L^2}$. Return to the general process, assuming that $g := h$ and

$$\Gamma(x, y) \geq m \|x - y\|^2 \quad \forall x \in \text{cl}S, \ y \in S \quad (9.3.12)$$

is valid with $m > 0$. Here, boundedness of $K$ is not supposed. However, $\alpha \leq m$ ensures

$$\Gamma_1(x, y) = \alpha \|x - y\|^2, \quad (9.3.13)$$

and relation (9.3.11) follows as above. Hence, condition (9.3.9) suffices to conclude that (9.3.6) is valid with $\sum_{k=1}^{\infty} \chi_k^{-1} \delta_k < \infty$.

Note that in the particular case $h(x) := \frac{1}{2} \|x\|^2$ we deal with the classical proximal point method, and (9.3.9) is nothing else as the known criterion (A') in [352]. Relation (9.3.13) holds also true if the functional $h$ is chosen as in methods with weak regularization or regularization on a subspace (see Subsection 8.2.3). Thus, the convergence results established below can be applied to these methods in the form (9.3.8), (9.3.9).

9.3.5 Remark. Eckstein explains the discrepancy in the convergence conditions (9.3.9) (for the classical method) and (9.3.9),(9.3.10) (for non-quadratic proximal methods) with the fact that "no triangle inequality applies" to Bregman distances (9.3.9) (for the classical method) and (9.3.9),(9.3.10) (for non-quadratic methods) with non-quadratic $h$. However, the analysis above shows that the "sufficiency" of criterion (9.3.9) depends more on the fulfillment of relation (9.3.13) for some $\alpha > 0$.

9.3.2 Convergence analysis

At first we show existence and uniqueness of a solution of Problem $(P^k_0)$, and the inclusion $x^{k+1} \in S$ for a solution of $(I^k_0)$.

According to Assumption 9.3.3 (B1), the subdifferential operator $\partial \eta$ is maximal monotone. The Assumptions 9.3.3 (B1)-(B3) and (B9) provide that $D(\partial \eta) = S$. Indeed, the inclusion $D(\partial \eta) \supset S$ follows from (B2), and assuming that $\partial \eta(x) \neq \emptyset$ holds for some $x \in \text{cl}S \setminus S$, in view of (B3) we obtain

$$(\nabla \eta(y) - \xi(x), y - x) > 0 \quad \forall y \in S, \ \xi(x) \in \partial \eta(x).$$

But, for a fixed $\xi(x) \in \partial \eta(x)$, due to (B9), there exists $g \in S$ such that $\nabla \eta(y) = \xi(x)$, in contradiction to the last inequality.

In turns, the conclusion $D(\partial \eta) = S$ means that $D(\nabla h) = S$, and the both operators $\nabla \eta$ and $\nabla h$ are maximal monotone.

Thus, if Problem $(P^k_0)$ is solvable, then it has a unique solution, here denoted by $x^{k+1}$ (the strict monotonicity of $Q + \chi_k \nabla h$ on $S \cap K_k \cap D(Q)$ follows immediately from Assumption 9.3.1 (A3) and Assumption 9.3.3 (B3)), and $x^{k+1} \in S$. Then, of course, the solution $x^{k+1}$ of Problem $(P^k_{\alpha_k})$ exists, and $D(\nabla h) = S$ provides $x^{k+1} \in S$.

Because the operator $\nabla h$ is maximal monotone and $S$ is an open set, the maximal monotonicity of the operators $Q + \chi_k \nabla h + N_{K_k}$ and $x \rightarrow Q(x) + \chi_k \nabla h(x) + N_{K_k}(x) - \chi_k \nabla h(x^k)$ follows from the Assumptions 9.3.1 (A1), (A2)
9.3. **PPR WITH GENERALIZED DISTANCE FUNCTIONALS**

and (A5) according to Theorem 1 in [350].

Since \( K_k \cap S \neq \emptyset \), the Moreau-Rockafellar theorem yields

\[
N_{K_k \cap S} = N_{K_k} + N_{cl S}.
\]

Taking into account that \( \bar{x}^{k+1} \) (if it exists) belongs to \( S \), this permits to transform Problem (\( P_{k_0}^k \)) into the inclusion

\[
0 \in Q(x) + \chi_k \nabla h(x) + N_{K_k}(x) - \chi_k \nabla h(x^k)
\]

\[
= Q(x) - \chi_k B x + N_{K_k}(x) + \chi_k B x^k + \chi_k (\nabla \eta(x) - \nabla \eta(x^k)),
\]

and with regard to Assumption 9.3.1 (A1), (A3) and \( 0 < \chi_k \leq 1 \), the operator

\[
x \rightarrow Q(x) - \chi_k B x + N_{K_k}(x) + \chi_k B x^k
\]

is maximal monotone (see Proposition 2.6 in [345]). Now, applying Lemma 5 in [59], one can conclude solvability of Problem (\( P_{k_0}^k \)). So, the following statement is proved.

**9.3.6 Theorem.** Let the Assumptions 9.3.1 (A1)-(A3), (A5) and Assumptions 9.3.3 (B1)-(B3), (B9) be valid. Then Problem (\( P_{k_0}^k \)) is uniquely solvable (for each \( k \)), the sequence \( \{x^k\} \) generated by Method 9.3.4 is well defined and contained in \( S \).

**9.3.7 Remark.** Using instead of Assumption 9.3.3 (B9) the condition (see [193])

\[
\{v^k\} \subset S, \ v^k \rightharpoonup v \in cl S \setminus S \implies \lim_{k \to \infty} \langle \nabla h(v^k), y - v^k \rangle = -\infty \forall y \in S,
\]

the conclusion \( D(\partial \eta) = S \) can be obtained from Lemma 1 in [59], and a result on solvability like Theorem 2 in [59] holds also true. ♦

In the sequel, we investigate the convergence of the method and need the following assertion proved first in [223].

**9.3.8 Lemma.** Let \( C \subset V \) be a convex closed set, the operators \( A : V \to 2^{V^*} \), \( A + N_C \) be maximal monotone and \( D(A) \cap C \) be a convex set. Moreover, assume that the operator

\[
A_C : v \to \begin{cases} A(v) & \text{if } v \in C \\ \emptyset & \text{otherwise} \end{cases}
\]

is locally hemi-bounded at each point \( v \in D(A) \cap C \) (see Definition A1.6.38) and that, for some \( u \in D(A) \cap C \) and each \( v \in D(A) \cap C \), there exists \( \zeta(v) \in A(v) \) satisfying

\[
\langle \zeta(v), v - u \rangle \geq 0. \tag{9.3.14}
\]

Then, with some \( \zeta \in A(u) \), the inequality

\[
\langle \zeta, v - u \rangle \geq 0
\]

holds for all \( v \in C \).
Therefore, there exists injection) are also maximal monotone. Moreover, they are strongly monotone. In view of the maximal monotonicity of $A$, the definition of the normality operator, this yields

$$\langle \zeta(w), v - w \rangle \geq 0 \quad \forall v \in C,$$  \hspace{1cm} (9.3.15)

with some $\zeta(w) \in A(w)$. If $w = u$, then, of course, $\langle \zeta(w), \zeta(w) \rangle \leq 0$, hence, the conclusion of the lemma is valid. Otherwise, we use the relation

$$\langle \zeta(v), v - u \rangle \geq 0 \quad \forall v \in D(A) \cap C,$$  \hspace{1cm} (9.3.16)

which follows from (9.3.14) taking $\zeta(v) = \zeta(v) + I(v - u) \in A(v)$.

Let $w_\lambda = u + \lambda(w - u)$ for $\lambda \in (0, 1]$. Obviously, $\lambda \in D(A) \cap C$, and according to (9.3.16) there exists $\zeta(w_\lambda) \in A(w_\lambda)$ ensuring

$$\langle \zeta(w_\lambda), w - u \rangle \geq 0.$$

Because the operator $A_C$ is locally hemi-bounded at $u$, the set $\{\zeta(w_\lambda) : \lambda \in (0, \lambda_0]\}$ is bounded in $V'$ for a sufficiently small $\lambda_0 > 0$. Hence, if $\lambda$ tends to 0 in an appropriate manner, the corresponding sequence $\{\zeta(w_\lambda)\}$ converges weakly in $V'$ to some $\bar{\zeta}$. Taking into account that $\lim_{\lambda \to 0} \|w_\lambda - u\| = 0$ and that $A$ is maximal monotone, one can conclude that $\bar{\zeta} \in A(u)$ and

$$0 \leq \lim_{\lambda \to 0} \langle \zeta(w_\lambda), w - u \rangle = \langle \bar{\zeta}, w - u \rangle.$$

Combining this inequality and inequality (9.3.15) given with $v = u$, we obtain

$$\langle \bar{\zeta} - \zeta(w), u - w \rangle \leq 0,$$

but that contradicts the strong monotonicity of $A$. \hfill \Box

9.3.9 Lemma. Let the sequence $\{x^k\}$, generated by Method 9.3.4, belong to $S$ and assume that, for some $x^* \in \text{SOL}(Q, K) \cap \text{clS}$, Assumption 9.3.1 (A6) is valid. Moreover, let the Assumptions 9.3.1 (A3), (A5) and Assumptions 9.3.3 (B1), (B2), (B5) be fulfilled. Then

(i) $\{\Gamma(x^*, x^k)\}$ is convergent,

(ii) $\{x^k\}$ is bounded, and

(iii) $\lim_{k \to \infty} \Gamma(x^{k+1}, x^k) = 0$.

Proof: We rewrite

$$\Gamma(x^*, x^{k+1}) - \Gamma(x^*, x^k) = s_1 + \chi_k^{-1}s_2 + s_3,$$

with

\[
\begin{align*}
    s_1 & := h(x^k) - h(x^{k+1}) + (\nabla h(x^k), x^{k+1} - x^k), \\
    s_2 & := \chi_k(\nabla h(x^k) - \nabla h(x^{k+1}), x^* - x^{k+1}), \\
    s_3 & := \frac{1}{2}\langle B(x^{k+1} - x^*), x^{k+1} - x^* \rangle - \frac{1}{2}\langle B(x^{k} - x^*), x^{k} - x^* \rangle.
\end{align*}
\]
Hence, and (9.3.17), (9.3.18) and (A6) permit us to conclude that $x$ follows immediately from (9.3.6). Together with (A3), this yields

$$s_2 \leq \langle q(x^{k+1}), x^* - x^{k+1} \rangle + \delta_k \sqrt{\Gamma(x^*, x^{k+1})} + 1$$

$$\leq \langle q^*(x^*), x^* - x^{k+1} \rangle - \langle B(x^* - x^{k+1}), x^* - x^{k+1} \rangle + \delta_k \sqrt{\Gamma(x^*, x^{k+1})} + 1$$

$$\leq \langle q^*(x^*), w^{k+1} - x^{k+1} \rangle + \langle q^*(x^*), x^* - w^{k+1} \rangle - \langle B(x^* - x^{k+1}), x^* - x^{k+1} \rangle + \delta_k \sqrt{\Gamma(x^*, x^{k+1})} + 1,$$  (9.3.17)

where we take $q^*(x^*)$ and $w^{k+1} := w^{k+1}(x^{k+1})$ according to (A6).

From the definition of $x^*$ and $q^*(x^*)$, one gets

$$\langle q^*(x^*), x^* - w^{k+1} \rangle \leq 0,$$  (9.3.18)

and (9.3.17), (9.3.18) and (A6) permit us to conclude that

$$s_2 \leq c(\Gamma(x^*, x^{k+1}) + 1) \varphi_k - \langle B(x^* - x^{k+1}), x^* - x^{k+1} \rangle + (1 + \Gamma(x^*, x^{k+1})) \delta_k.$$  

With regard to $0 < \chi_k \leq 1$, this leads to

$$\chi_k^{-1} s_2 + s_3 \leq \Gamma(x^*, x^{k+1}) \left( \frac{c \varphi_k}{\chi_k} + \frac{\delta_k}{\chi_k} \right) + \frac{c \varphi_k}{\chi_k} + \frac{\delta_k}{\chi_k} - \frac{1}{2} \langle B(x^* - x^{k+1}), x^* - x^{k+1} \rangle - \frac{1}{2} \langle B(x^* - x^k), x^* - x^k \rangle.$$

But

$$\frac{1}{2} \langle B(x^{k+1} - x^k), x^{k+1} - x^k \rangle$$

$$\leq \langle B(x^* - x^{k+1}), x^* - x^{k+1} \rangle + \langle B(x^* - x^k), x^* - x^k \rangle$$

is obvious, and taking into account the definition of $\Gamma$ and the convexity of $h,$

$$s_1 + \chi_k^{-1} s_2 + s_3 \leq \Gamma(x^*, x^{k+1}) \left( \frac{c \varphi_k}{\chi_k} + \frac{\delta_k}{\chi_k} \right) + \frac{c \varphi_k}{\chi_k} + \frac{\delta_k}{\chi_k}$$

$$- \frac{1}{4} \langle B(x^{k+1} - x^k), x^{k+1} - x^k \rangle$$

$$- [h(x^{k+1}) - h(x^k) - \langle \nabla h(x^k), x^{k+1} - x^k \rangle]$$

$$\leq \Gamma(x^*, x^{k+1}) \left( \frac{c \varphi_k}{\chi_k} + \frac{\delta_k}{\chi_k} \right) + \frac{c \varphi_k}{\chi_k} + \frac{\delta_k}{\chi_k} - \frac{1}{2} \Gamma(x^{k+1}, x^k).$$

Hence,

$$\Gamma(x^*, x^{k+1}) \left( 1 - \frac{c \varphi_k}{\chi_k} - \frac{\delta_k}{\chi_k} \right) \leq \Gamma(x^*, x^k) + \frac{c \varphi_k}{\chi_k} + \frac{\delta_k}{\chi_k} - \frac{1}{2} \Gamma(x^{k+1}, x^k)$$  (9.3.19)
is valid. But, condition (A5) implies the existence of an index $k_0$ such that
\[ \frac{c_{\phi k}}{\chi_k} + \frac{\delta_k}{\chi_k} \leq \frac{1}{2} \text{ for } k \geq k_0, \]
i.e.,
\[ 1 \leq \left( 1 - \frac{c_{\phi k}}{\chi_k} - \frac{\delta_k}{\chi_k} \right)^{-1} \leq 1 + 2 \left( \frac{c_{\phi k}}{\chi_k} + \frac{\delta_k}{\chi_k} \right) \leq 2. \]
Thus, for $k \geq k_0$, we obtain from (9.3.19) and $\Gamma(x^{k+1}, x^k) \geq 0$ that
\[ \Gamma(x^*, x^{k+1}) \leq \left[ 1 + 2 \left( \frac{c_{\phi k}}{\chi_k} + \frac{\delta_k}{\chi_k} \right) \right] \Gamma(x^*, x^k) \]
\[ + 2 \left( \frac{c_{\phi k}}{\chi_k} + \frac{\delta_k}{\chi_k} \right) - \frac{1}{2} \Gamma(x^{k+1}, x^k), \quad (9.3.20) \]
and, in view of condition (A5), Lemma A3.1.7 provides that the sequence $\{\Gamma(x^*, x^k)\}$ is convergent. Now, boundedness of $\{x^k\}$ follows from (B5), and using again (9.3.20) one gets $\lim_{k \to \infty} \Gamma(x^{k+1}, x^k) = 0$. □

In the sequel, we deal with the case that, besides the usual property of maximal monotonicity, the operator $Q$ is pseudo-monotone and paramonotone (see the Definitions A1.6.42 and A1.6.43).

In particular, we will use the following property of a paramonotone operator $A$ in $C$ (cf. [59]):

\[ (\star) \text{ If } u^* \text{ solves the variational inequality } \langle A(u), v - u \rangle \geq 0 \ \forall v \in C \quad (9.3.21) \]
and for $\bar{u} \in C$ there exist $\bar{z} \in A(\bar{u})$ with $\langle \bar{z}, u^* - \bar{u} \rangle \geq 0$, then $\bar{u}$ is also a solution of (9.3.21).

9.3.10 Lemma. Let the assumptions of Lemma 9.3.9 as well as the Assumptions 9.3.1 (A1), (A4) and Assumptions 9.3.3 (B6), (B7) be valid. Moreover, suppose that one of the following assumptions is fulfilled:

(a) $S \supset \text{cl}(D(Q))$, $\nabla h$ is Lipschitz continuous on closed and bounded subsets of $S$ and the hypotheses of Lemma 9.3.8 hold with $A := Q$, $C := K \cap \text{cl}S$;
(b) $Q := \partial f$, with $f$ a proper convex lower semicontinuous functional which is continuous at some $x \in K$;
(c) $Q : D(Q) \to 2^{V'}$ is pseudo-monotone and paramonotone on $\text{cl}S$.

Then the sequence $\{x^k\}$, generated by Method 9.3.4, is bounded and each weak limit point is a solution of VI($Q, K$).

For a motivation of the inclusion $S \supset \text{cl}(D(Q))$, which excludes the usual choice of a function $h$ leading to interior point methods, see [96].

In the case (b), condition (A4) can be weakened assuming that each limit point of $\{v^k\}$ belongs to $K$ (in place of $K \cap D(Q)$).
9.3. PPR WITH GENERALIZED DISTANCE FUNCTIONALS

Proof: According to Lemma 9.3.9, the sequence \( \{x^k\} \) is bounded, hence, there exists a weakly convergent subsequence \( \{x^{j_k}\}, x^{j_k} \to_{k \to \infty} \bar{x}. \) In view of \( \{x^k\} \subset S, \) (A4) and the convexity of \( S, \) the inclusion \( \bar{x} \in \text{cl} S \cap K \cap D(\mathcal{Q}) \) (respectively \( \bar{x} \in \text{cl} S \cap K \) in the case (b)) is valid.

Due to \( \lim_{k \to \infty} \Gamma(x^{k+1}, x^k) = 0, \) one can use condition (B7) with \( y^k := x^{j_k+1}, \) \( y^k := x^{j_k}. \) This leads to

\[
\lim_{k \to \infty} \|x^{j_k+1} - x^{j_k}\| = 0. \tag{9.3.22}
\]

If (a) holds, then with regard to the boundedness of \( \{x^k\} \), \( \{x^k\} \subset D(\mathcal{Q}), \) (A5) and (9.3.22), the relation

\[
\lim_{k \to \infty} \chi_j \langle \nabla h(x^{j_k+1}) - \nabla h(x^{j_k}), x - x^{j_k+1} \rangle = 0 \quad \forall x \in V \tag{9.3.23}
\]

follows immediately.

Now, take (9.3.6) with an arbitrary \( x \in K \cap \text{cl} S \) and replace \( q(x^{k+1}) \) by \( q(x) \in \mathcal{Q}(x) \) (this is possible in view of the monotonicity of \( \mathcal{Q} \)). Then, passing to the limit in the obtained inequality with \( k := j_k, \) \( k \to \infty, \) due to the boundedness of \( \{x^k\}, \) the definition of \( \Gamma_1, \) (A5) and (9.3.23), we obtain

\[
\langle q(x), x - \bar{x} \rangle \geq 0 \quad \forall x \in K \cap \text{cl} S.
\]

The conditions (A1) and \( S \supset \text{cl}(D(\mathcal{Q})) \) guarantee the maximal monotonicity of the operator \( \mathcal{Q} + \mathcal{N}_{K \cap \text{cl} S} \) (in fact, \( \mathcal{Q} + \mathcal{N}_{K \cap \text{cl} S} \) coincides with \( \mathcal{Q} + \mathcal{N}_K \)). Thus, we are able to apply Lemma 9.3.8 with \( C := K \cap \text{cl} S, \) \( A := \mathcal{Q}, \) which ensures that

\[
\exists q(\bar{x}) \in \mathcal{Q}(\bar{x}): \langle q(\bar{x}), y - \bar{x} \rangle \geq 0 \quad \forall y \in K \cap \text{cl} S.
\]

This yields

\[
0 \in q(\bar{x}) + \mathcal{N}_{K \cap \text{cl} S}(\bar{x}) \subset \mathcal{Q}(\bar{x}) + \mathcal{N}_{K \cap \text{cl} S}(\bar{x}),
\]

hence \( 0 \in \mathcal{Q}(\bar{x}) + \mathcal{N}_K(\bar{x}) \) holds, proving \( \bar{x} \in \text{SOL}(\mathcal{Q}, K). \)

Suppose now that (b) is valid and take \( x^* \) as in (A6). With regard to the symmetry of \( B \) a straightforward calculation gives

\[
-\langle \nabla h(x^k) - \nabla h(x^{k+1}), x^* - x^{k+1} \rangle = \Gamma(x^*, x^k) - \Gamma(x^*, x^{k+1})
- \Gamma(x^{k+1}, x^k) - \langle B(x^{k+1} - x^k), x^* - x^{k+1} \rangle. \tag{9.3.24}
\]

Using Lemma 9.3.9, (9.3.22) and \( 0 < \chi_k \leq 1, \) we infer from (9.3.24) that

\[
\lim_{k \to \infty} \chi_j \langle \nabla h(x^{j_k+1}) - \nabla h(x^{j_k}), x^* - x^{j_k+1} \rangle = 0. \tag{9.3.25}
\]

But, relation (9.3.6) given with \( x = x^* \) and \( k := j_k \) implies, due to the convexity of \( f, \) that

\[
\delta_{j_k} \sqrt{\Gamma_1(x^*, x^{j_k+1})} + \chi_j \langle \nabla h(x^{j_k+1}) - \nabla h(x^{j_k}), x^* - x^{j_k+1} \rangle \geq f(x^{j_k+1}) - f(x^*). \tag{9.3.26}
\]

Taking the limit in (9.3.26) as \( k \to \infty, \) due to Lemma 9.3.9, (9.3.25), (A5) and the lower semicontinuity of \( f, \) one gets \( f(\bar{x}) \leq f(x^*). \) Thus, \( 0 \in \partial (f(\bar{x}) + \delta(\bar{x}|K)), \)
\[ \delta(|K|) \] is the indicator functional of \( K \), and the Moreau-Rockafellar theorem provides \( \bar{x} \in \text{SOL}(Q, K) \).

Finally, let us consider the case (c). Using equality (9.3.24) with \( k := j_k \) and \( \bar{x} \) in place of \( x^* \), from (9.3.22), Lemma 9.3.9 and condition (B6) for \( v^k := x^{j_k+1}, y^k := x^{j_k} \), we conclude that
\[ \lim_{k \to \infty} \langle \nabla h(x^{j_k+1}) - \nabla h(x^{j_k}), \bar{x} - x^{j_k+1} \rangle = 0. \]
Thus, (9.3.6) taken with \( x := \bar{x} \) implies
\[ \lim_{k \to \infty} \langle q(x^{j_k+1}), x^{j_k+1} - \bar{x} \rangle \leq 0, \]
and the pseudo-monotonicity of \( Q \) provides the existence of \( q(\bar{x}) \in Q(\bar{x}) \) such that
\[ \langle q(\bar{x}), \bar{x} - x^* \rangle \leq \lim_{k \to \infty} \langle q(x^{j_k+1}), x^{j_k+1} - x^* \rangle. \]
Now, from (9.3.6) and relation (9.3.25), which is true also in this case, one gets
\[ \langle q(\bar{x}), \bar{x} - x^* \rangle \leq 0. \]
Therefore, the above mentioned property (★) of paramonotonicity permits to conclude that \( \bar{x} \in \text{SOL}(Q, K) \).

\[ \square \]

**Theorem 9.3.11.** Let the Assumptions 9.3.1 (A1)-(A5) and Assumptions 9.3.3 (B1)-(B9) be valid, and Assumption 9.3.1 (A6) holds for each \( x \in \text{SOL}(Q, K) \cap \text{clS} \) (constant \( c \) in (A6) may depend on \( x \)). Moreover, let the operator \( Q \) possess one of the properties (a), (b) or (c) in Lemma 9.3.10. Then the sequence \( \{x^k\} \), generated by Method 9.3.4, converges weakly to a solution of \( \text{VI}(Q, K) \).

**Proof:** The existence of the sequence \( \{x^k\} \) and the inclusion \( \{x^k\} \subset S \) are guaranteed by Theorem 9.3.6. Denote
\[ d_k(x) := \Gamma(x, x^k) - \frac{1}{2} \langle Bx, x \rangle - h(x). \]
According to Lemma 9.3.9, \( \{\Gamma(x, x^k)\} \) converges for each \( x \in \text{SOL}(Q, K) \cap \text{clS} \), hence, the sequence \( \{d_k(x)\} \) possesses the same property.

Boundedness of \( \{x^k\} \) was proved in Lemma 9.3.9, and Lemma 9.3.10 yields that each weak limit point of \( \{x^k\} \) belongs to \( \text{SOL}(Q, K) \cap \text{clS} \).

Assume that \( \{x^{j_k}\} \) and \( \{x^{i_k}\} \) converge weakly to \( \bar{x} \), \( \tilde{x} \), respectively. Then it holds \( \bar{x}, \tilde{x} \in \text{SOL}(Q, K) \cap \text{clS} \). Let
\[ l_1 := \lim_{k \to \infty} d_k(\bar{x}), \quad l_2 := \lim_{k \to \infty} d_k(\tilde{x}). \]
Obviously,
\[ l_1 - l_2 = \lim_{k \to \infty} (d_k(\bar{x}) - d_k(\tilde{x})) = \lim_{k \to \infty} \langle \nabla h(x^k) + Bx^k, \bar{x} - \tilde{x} \rangle. \]
Considering the latter equality now for the subsequences \( \{x^{j_k}\} \) and \( \{x^{i_k}\} \), one can conclude that
\[ \lim_{k \to \infty} \langle \nabla h(x^{j_k}) + Bx^{j_k} - \nabla h(x^{i_k}) - Bx^{i_k}, \bar{x} - \tilde{x} \rangle = 0. \quad (9.3.27) \]
A comparison of (9.3.27) and (B8) (given with \( v^k := x^{j_k}, y^k := x^{i_k} \)) indicates \( \bar{x} = \tilde{x} \), proving uniqueness of the weak limit point of \( \{x^k\} \). \( \square \)
9.3.12 Remark. Theorem 9.3.6 remains true if Assumption 9.3.3 (B9) is replaced by any other condition guaranteeing that \( \{x^k\} \) is well defined and \( \{x^k\} \subset S \) (see, in particular, Remark 9.3.7).

If we replace \( S \supset \text{cl}(D(Q)) \) by the weaker requirement that
\[
S \supset \text{cl}(D(Q) \cap (\cup_{k \geq k_0} K_k))
\]
is valid for an arbitrary large \( k_0 \), then – with evident technical alterations – the proofs of Lemma 9.3.10 and Theorem 9.3.11 hold true.

9.4 Comments

Section 9.1: Using a relaxation factor \( \rho_k > 1 \), Bertsekas [44] reports about the improvement of the convergence rate of the multiplier method for constrained convex programming, which is – as we already mentioned – a well-known application of the proximal point algorithm.

Facchinei and Pang [107] have shown that for a strongly monotone (with constant \( m \)) and Lipschitz continuous (with constant \( L > m \)) operator \( Q \) the sequence \( \{x^k\} \) generated by relaxed PPR with \( \rho_k = 1 + \frac{m}{\chi_k L^2} > 1 \) satisfies
\[
\|J_{\chi_k Q}(x^k) - x^*\| \leq (1 + \chi_k m)^{-1} \|x^k - x^*\|
\]
\[
\|x^{k+1} - x^*\| \leq \left(1 - \frac{m^2}{L^2}\right)^{1/2} (1 + \chi_k m)^{-1} \|x^k - x^*\|
\]
where \( x^* \) is the unique solution of the equation \( 0 = Q(x) \). Due to \( \left(1 - \frac{m^2}{L^2}\right) < 1 \) in the right-hand side of the last inequality, after one step of the relaxed proximal point algorithm the distance \( \|x^{k+1} - x^*\| \) is less than the one in the classical PPR.

Eckstein and Bertsekas [98] observed that only in the case of over-relaxation in PPR it should be possible to accelerate the convergence.

Based on Solodov and Svaiter’s investigation in [374], Hue and Strodiot [186], considered the following version of the relaxed PPR: For some fixed relaxation parameter \( \rho \in (0, 2) \) and an error tolerance parameter \( \delta \in [0, 1) \) find a triplet \( (y^k, q^k, \epsilon_k) \), \( \epsilon_k \geq 0 \), such that
\[
q^k \in Q^\epsilon_k(y^k), \quad \chi_k q^k + y^k - x^k = r^k,
\]
and
\[
\gamma\|r^k\|^2 + 2\gamma\epsilon_k \chi_k + 2|1 - \rho\|r^k\||y^k - x^k| \leq \delta^2(2 - \rho)\|y^k - x^k\|^2.
\]
where \( \gamma := \max\{1, \rho^2\} \). Afterwards the authors execute the extra-gradient step
\[
x^{k+1} := x^k - \rho\chi_k q^k.
\]
For \( \delta := 0 \) the algorithm of Hue and Strodiot coincides with the exact version of the relaxed proximal point algorithm of Eckstein, Bertsekas with \( \rho_k \equiv \rho \).

Hue and Strodiot report about first numerical experiences and obtain better convergence results in the case of over-relaxation, i.e. for \( \rho > 1 \). But they mention also, see [186] (Sect. 5), that the behavior of the algorithm for small
and large values of $\rho$ “becomes worse: the number of steps is increasing”. This fact is not surprising because the proximal point algorithm is just a fixed point iteration for the resolvent operator $J_{\chi_k}\tau$, with $\tau = Q + \mathcal{N}_K$. A solution of the inclusion $0 \in \tau(x)$ is found if $x^{k+1} := J_{\chi_k}\hat{Q}(x^k) = x^k$. Therefore, if the current iterate $x^k$ is far from the solution set, then one can execute over-relaxation steps with the hope to accelerate the convergence. But, if $x^k$ is in a neighborhood of the solution set, then it is natural to keep the relaxation parameter $\rho_k \approx 1$. From this point of view it is important to vary the relaxation parameter $\rho_k$ during the iteration process.

The results of this chapter are first published by Huebner and Tichatschke [188], see also the PhD-thesis of Huebner [187].

**Section 9.2:** Bregman functions have been introduced by Bregman [48] to find common points in the intersection of convex sets. Later on they were used by several authors (cf. Butnariu and Iusem [59], Chen and Teboulle [74], Eckstein [97], Solodov and Svaiter [376]) in order to generalize proximal point methods for problems with polyhedral feasibility sets.

The main motivation for such proximal methods is not only to guarantee that the iterates stay in the interior of the set $K$ and to preserve the main merits of the classical PPR (good stability of the auxiliary problems and convergence of the whole sequence of iterates to a solution of the original problem), at the same time, the application of non-quadratic proximal techniques (as in Auslender, Teboulle and Ben-Tiba [25], Teboulle [385], Tseng and Bertsekas [402]) to the dual of a smooth convex program leads to multiplier methods with twice or higher differentiable augmented Lagrangian functionals. Moreover, in [25] the Hessians of these functionals are bounded.

The implementation of a Bregman-function-based proximal Algorithm considered in this Section is described in [187]. There also numerical examples can be find showing the performance of the algorithm.

**Section 9.3:** More motivation of non-quadratic proximal methods can be found by Auslender and Haddou [22], Ben-Tal and Zibulevski [40], Eckstein [96], Polyak and Teboulle [332]. For infinite-dimensional convex optimization problems such methods have been studied by Alber, Burachik and Iusem [6], Butnariu and Iusem [65, 66], and for variational inequalities in Hilbert spaces see Burachik and Iusem [59].
Chapter 10

THE AUXILIARY PROBLEM PRINCIPLE

The auxiliary problem principle (APP), originally introduced by Cohen [78, 79] as a general framework to analyze optimization algorithms of gradient and subgradient types as well as decomposition algorithms, was extended later to different numerical methods for solving variational inequalities. In the majority of the papers dedicated to such extensions, the $(k+1)$-th auxiliary problem is constructed by applying the operator of the variational inequality (here called main operator) to the $k$-th iterate. The idea to take an additive component of this operator at a variable point leads to a scheme which appears to be a generalization of proximal-like methods, too.

In order to illustrate this, let us consider the inclusion problem

\[ \text{IP}(\mathcal{T}, X) \quad \text{find} \quad x \in X : \quad 0 \in \mathcal{T}(x), \quad (10.0.1) \]

with $\mathcal{T}$ a given multi-valued maximal monotone operator in a Hilbert space $X$. The APP is given in the form

\[ \text{find} \quad x^{k+1} \in X : \quad 0 \in \frac{1}{\chi_k} \left[ \Xi(x^{k+1}) - \Xi(x^k) \right] + \mathcal{T}(x^k), \quad (10.0.2) \]

with $\Xi : X \to X$ an auxiliary operator, $\chi_k > 0$, and this iterative scheme can be modified as follows:

\[ \text{find} \quad x^{k+1} \in X : \quad 0 \in \frac{1}{\chi} \left[ \Xi(x^{k+1}) - \Xi(x^k) \right] + \mathcal{T}(x^{k+1}), \quad (10.0.3) \]

or

\[ \text{find} \quad x^{k+1} \in X : \quad 0 \in \frac{1}{\chi} \left[ \Xi(x^{k+1}) - \Xi(x^k) \right] + \mathcal{T}_1(x^k) + \mathcal{T}_2(x^{k+1}), \quad (10.0.4) \]

where $\chi > 0$ and $\mathcal{T}$ is decomposed into a sum of a single-valued monotone operator $\mathcal{T}_1$ and a maximal monotone operator $\mathcal{T}_2$.

Then, in case $\Xi := I$ ($I$ - identity operator), the following well-known methods arise:
inclusion (10.0.2) leads to
\[
x^{k+1} \in x^k - \chi_k T(x^k),
\]
which is an analogue of the subgradient method;

- on using (10.0.3) we obtain
\[
x^{k+1} = (I + \chi_k T)^{-1}(x^k),
\]
the proximal point method;

- finally, (10.0.4) produces
\[
x^{k+1} = (I + \chi T_2)^{-1}(I - \chi T_1)(x^k), \tag{10.0.5}
\]
a splitting algorithm, suggested by Lions and Mercier [273] (see also Gabay [125] and Passty [319]).

Obviously, the latter algorithm can be represented as
\[
z^k = x^k - \chi T_1(x^k), \quad x^{k+1} = (I + \chi T_2)^{-1}(z^k),
\]
where \( x^{k+1} \) is calculated from \( z^k \) by means of the resolvent of the proximal mapping. Tseng [400, 401] has used method (10.0.5) (with variable \( \chi \)) as a basic process to investigate convergence of several known but also new splitting methods for variational inequalities with separability properties (including linear complementarity problems) as well as for related convex optimization problems with linear constraints.

10.1 Extended Auxiliary Problem Principle

10.1.1 On merging of APP and IPR

Throughout this chapter the APP is studied for variational inequalities of the form
\[
\text{VI}(F, Q, K) \quad \text{find } x^* \in K \text{ and } q^* \in Q(x^*) : \\
\langle F(x^*) + q^*, x - x^* \rangle \geq 0 \quad \forall x \in K,
\]
with \( K \) a convex closed subset of a Hilbert space \((X, \| \cdot \|)\), \( F \) a single-valued operator from \( X \) into the dual space \( X' \) and \( Q : X \to 2^{X'} \) a maximal monotone (in general, non-symmetric) operator. For a monotone operator, the notion symmetric means that the operator is the subdifferential of a convex functional.

The suggested scheme includes successive approximation of \( K \) by means of a sequence \( \{K_k\} \) of convex closed sets. Taking a set
\[
\hat{K} \supset K \cup (\cup_{k=1}^\infty K_k), \quad \hat{K} \subset X,
\]
auxiliary problems of the form
\[
(P_k) \quad \text{find } \bar{x}^{k+1} \in K_k, \quad q^k(\bar{x}^{k+1}) \in Q^k(\bar{x}^{k+1}) : \\
\langle F(\bar{x}^k) + q^k(\bar{x}^{k+1}) + L^k(\bar{x}^{k+1}) - L^k(x^k) + \chi_k(\nabla h(\bar{x}^{k+1}) - \nabla h(x^k)), x - \bar{x}^{k+1} \rangle \geq 0 \quad \forall x \in K_k,
\]
are considered. Here \( h : X \to (-\infty, +\infty] \) is a convex functional, Gâteaux-differentiable on \( \hat{K} \); \( Q^k \) is a monotone operator approximating \( Q \); \( L^k : X \to X' \) is a monotone operator linked with \( F \) by a kind of pseudo Dumi property. \( D(L^k) \supset \hat{K} \); \( x^k \) is an approximate solution of the previous problem and \( \chi_k \) is a positive scalar. The sum \( L^k + \chi_k \nabla h \) corresponds to the customary notion of an auxiliary operator \( \Xi^k \).

For practical implementation, the cases when \( Q^k \equiv Q \) (i.e. there is no approximation of \( Q \)) or if \( Q^k \) is a single-valued approximation of the multi-valued \( Q \) are of most interest, and the scheme under consideration covers these cases.

To make the APP numerically tractable we will investigate in this section the iteration process

\[
\begin{align*}
\text{(P}^k\text{)} \quad & \text{find } x^{k+1} \in K_k, \quad q^k(x^{k+1}) \in Q^k(x^{k+1}) : \\
& (F(x^k) + q^k(x^{k+1}) + L^k(x^{k+1}) - L^k(x^k)) \\
& + \chi_k(\nabla h(x^{k+1}) - \nabla h(x^k)), x - x^{k+1} \geq -\epsilon_k \| x - x^{k+1} \| \quad \forall x \in K_k,
\end{align*}
\]

where \( \epsilon_k \downarrow 0 \) is a given sequence and \( x^1 \in K \) is an arbitrary starting point. We call this process the Proximal Auxiliary Problem Method (PAP-method).

It is easy to see, that replacing \( Q^k \) by \( F + Q^k \), \( F \) by 0 and taking \( L^k \equiv 0 \), proximal-like methods considered in Section 9.1 can be derived from \( (P^k) \).

The conditions concerning an approximation of \( Q \) by \( \{Q^k\} \) and \( K \) by \( \{K_k\} \) are of the type of Mosco-convergence (cf. Section 2.4 and [296]). For the exact conditions which are supposed to be valid for the both problems VI\( (F, Q, K) \) and \( (P^k) \) see the Assumptions 10.1.1 and 10.1.4 below. They allow the solution set of VI\( (F, Q, K) \) to be infinite-dimensional and unbounded, i.e. we deal with essentially ill-posed problems.

Coming back to the sources of the PAP-method, we begin with a version of the APP for convex optimization problems, where a linearization of the Gâteaux-differentiable term \( J_1 \) of an objective functional \( J = J_1 + J_2 \) is used. This was already studied by Cohen in [78]. Taking this partial linearization, the term \( \langle \nabla J_1(x^k), \cdot - x^k \rangle + J_2(\cdot) \) is inserted into the objective functional of the \((k+1)\)-th auxiliary problem. Such an approach is of special interest for constructing decomposition methods. In fact, if the problem \( \min \{J_2(x) : x \in K \} \) splits up into independent subproblems, then the mentioned linearization permits one to provide the same splitting in the framework of the APP for the original problem \( \min \{J(x) : x \in K \} \), i.e. the corresponding auxiliary problems can be split up, too.

The general scheme oriented to decomposition methods for hemi-variational inequalities of the form

\[
\text{HVI}(T, f, K) \quad \text{find } x^* \in X : \\
\langle T(x^*), x - x^* \rangle + f(x) - f(x^*) \geq 0, \quad \forall x \in X,
\]

where \( T \) is a single-valued monotone operator and the functional \( f \) is convex, lower semicontinuous (lsc) and additive with respect to a Cartesian factorization of the space \( X \), has been developed by Makler-Scheimberg, Nguyen and Strodiot in [282]. An extension of this scheme (in case \( X := \mathbb{R}^n \) and without
accentuating decomposition methods), described by Salmon, Nguyen and Strodiot in [359], is connected with relaxation of the monotonicity condition for $T$ and with the use of a wider class of auxiliary operators (as distinct from [282], these operators may be non-symmetric). The auxiliary problems in [359] can be written as

\[
\begin{align*}
\text{find } x^{k+1} \in X : \\
\langle T(x^k) + L^k(x^{k+1}) - \mathcal{L}^k(x^k) + \chi_k(\nabla h^k(x^{k+1}) - \nabla h^k(x^k)), x - x^{k+1} \rangle + f^k(x) - f^k(x^{k+1}) & \geq 0 \quad \forall x \in X, \\
\end{align*}
\]

(10.1.1)

and the conditions concerning $L^k$, $h^k$ and $f^k$ made there permit one to cover a lot of earlier versions of APP and special algorithms.

Formally, $\text{HVI}(T, f, K)$ and (10.1.1) correspond to $\text{VI}(F, Q, K)$ and $(P^k)$, respectively, with

\[
T := F + Q, \quad f := \text{ind}(\cdot|K)
\]

in $\text{HVI}(T, f, K)$ and

\[
L^k := L^k + Q, \quad h^k := h, \quad f^k := \text{ind}(\cdot|K_k)
\]

in (10.1.1). However, the conditions on $T$, $L^k$ and $f^k$ in [359] enforce that $Q$ is a single-valued, continuous operator and $K_k \subset K$.

In Section 10.2 we will modify the PAP-method for hemi-variational inequalities, but with a multi-valued operator $T$. In some applications, this leads to more convenient conditions on data approximation, than applying the basic process $(P^k)$ to the equivalent $\text{VI}(F, Q, K)$.

Our PAP-method described here can be considered as a generalization of the method studied by Zhu and Marcotte in [422], Sect. 4.2. The auxiliary problems in [422] correspond to $(P^k)$ with $X := \mathbb{R}^n$, $Q$ - the sum of a single-valued, monotone operator and a symmetric, monotone operator, $K_k := K$, $Q^k := Q$, $L^k := 0$, $\chi_k := \chi$, and with stronger assumptions on $h$ and $F$ (it is supposed that $h$ is strongly convex and $F$ possesses the Dunn property).

Paper [345] of Renaud and Cohen should also be mentioned, in which the APP is studied for the inclusion problem $\text{IP}(T, X)$ where $T : X \to 2^X$ is a maximal monotone operator. The corresponding auxiliary problems have the form (10.0.3) with a single-valued (in general, non-symmetric) auxiliary operator $\Xi = \Xi_1 + c \Xi_2$. Here, $\Xi_1$ is supposed to be the gradient of a strongly convex functional, and supposing the following relation between $T$ and $\Xi_2$:

\[
\exists \gamma_0 > 0 : \quad \langle t(x) - t(y), x - y \rangle \geq \gamma_0 \|\Xi_2(x) - \Xi_2(y)\|^2, \quad \forall t(x) \in T(x), \ t(y) \in T(y), \ \forall x, y \in \text{dom} T, \quad (10.1.2)
\]

the operator $\Xi_2$ is assumed to be hemicontinuous only.

In [345] the general convergence results for method (10.0.3) have been also adapted to prove convergence of a new algorithm for solving saddle-point problems with convex-concave functions on the product of convex sets. Depending on the decomposition of the related maximal monotone operator, the Arrow-Hurwicz algorithm and the proximal point method can be obtained as particular
The PAP-method studied in this section has the following distinguishing features: In comparison with \[282, 359\], the operator $Q$ is not supposed to be symmetric; and as distinct from \[345\], the main operator $F + Q$ is not necessarily monotone. Approximations of $Q$ and $K$ are also included, and the auxiliary operator is allowed to vary after each step.

Besides, we weaken the standard (for the APP) assumption on strong convexity of the auxiliary function $h$ in the Problems $(P_k^k)$: $h$ is supposed to be convex and the operators $Q^k + \nabla h$ have to be strongly monotone with a common modulus for all $k$.

Note that the conditions, which link the main and auxiliary operators in our scheme (see Assumption 10.1.1(v) below) are similar but not the same as in \[282, 345, 359\].

10.1.2 PAP-method and its convergence analysis

In the sequel, for each $x \in \text{SOL}(F, Q, K)$, the elements $q^*$ or $q^*(x)$ belong to $Q(x)$ and satisfy $VI(F, Q, K)$.

We consider the variational inequality $VI(F, Q, K)$ under the following basic assumptions.

10.1.1 Assumption. (Basic assumptions)

(i) $K \subset X$ is a convex, closed set;

(ii) $Q : X \to 2^{X'}$ is a maximal monotone operator, $D(Q) \cap K$ is a non-empty, convex set, and

$$Q_K : y \mapsto \begin{cases} Q(y) & \text{if } y \in K \\ \emptyset & \text{otherwise} \end{cases}$$

is locally hemi-bounded at each point of $D(Q) \cap K$;

(iii) the operator $Q + N_K$ is maximal monotone;

(iv) the operator $F : X \to X'$ is single-valued and weakly continuous on $K$ and the functional $x \mapsto \langle F(x), x \rangle$ is weakly lsc on $K$;

(v) given a family $\{L_y\}, L_y : X \to X'$, of monotone operators on $K$ parameterized by $y \in K$; there exists $\gamma > 0$ (independent of $x, y$) such that, for $x \in \text{SOL}(F, Q, K)$, $y \in K$, the inequality

$$\langle F(y) - F(x) - L_y(y) + L_y(x), y - x \rangle + \langle F(x) + q^*(x), z(y) - x \rangle \geq \gamma \|F(y) - L_y(y) - F(x) + L_y(x)\|_X^2,$$

is valid (here and in the sequel, $z(\cdot) = \arg\min_{v \in K} \|\cdot - v\|$);

(vi) $VI(F, Q, K)$ is solvable.

Let us discuss some notions and conditions given above.
CHAPTER 10. THE AUXILIARY PROBLEM PRINCIPLE

- Local hemi-boundedness of $Q$ (see Definition A1.6.38) is used to provide (by means of Lemma 9.3.8) the following implication with a fixed $\bar{x} \in K \cap D(Q)$: if
  \[ \forall x \in K \cap D(Q), \exists q(x) \in Q(x) : \langle F(\bar{x}) + q(x), x - \bar{x} \rangle \geq 0, \]

then $\bar{x}$ is a solution of $VI(F, Q, K)$.

This implication is crucial for proving that each weak limit point of the sequence $\{x^k\}$ generated by the APP-method is a solution of $VI(F, Q, K)$ (see Lemma 10.1.8).

- With $Q$ a maximal monotone operator and $K$ a convex closed set, the operator $Q + N_K$ is maximal monotone if, for instance, $\text{int} D(Q) \cap K \neq \emptyset$ or $Q$ is locally bounded at some $x \in K \cap \text{cl} D(Q)$ (see [350]).

Local boundedness of $Q$ at $x$ means that $Q$ carries some neighborhood of $x$ into a bounded set.

- By definition, $F$ is weakly continuous on $\hat{K}$ if $v_n \rightharpoonup v$ in $X$, $v_n \in \hat{K}$, $v \in \hat{K}$ imply $F(v_n) \rightharpoonup F(v)$ in $X'$.

Assumption 10.1.1(iv) is valid, in particular, if $F$ is a compact operator. Also the second part of (iv) follows from the first part if $F$ is a monotone operator. Indeed, let $v_n \rightharpoonup v$, $v_n \in \hat{K}$, $v \in \hat{K}$. In case $F$ is compact, this ensures that $\{F(v_n)\}$ converges to $F(v)$ strongly in $X'$. Then, using the identity
  \[ \langle F(v_n), v_n \rangle = \langle F(v_n) - F(v), v_n \rangle + \langle F(v), v_n \rangle, \]

we obtain immediately
  \[ \lim_{n \to \infty} \langle F(v_n), v_n \rangle = \langle F(v), v \rangle. \]

But also in the case if $F$ is monotone, the relation
  \[ \langle F(v_n), v_n \rangle = \langle F(v_n), v_n - v \rangle + \langle F(v_n), v \rangle \geq \langle F(v), v_n - v \rangle + \langle F(v_n), v \rangle \]

together with $v_n \rightharpoonup v$, $F(v_n) \rightharpoonup F(v)$ yields
  \[ \lim_{n \to \infty} \langle F(v_n), v_n \rangle \geq \langle F(v), v \rangle. \]

Instead of Assumption 10.1.1(iv) one can assume that $F$ is a single-valued, monotone, hemicontinuous and bounded operator on $X$, $D(F) = X$ (see [223]).

- Assumption 10.1.1(v) is not so strange as it seems at the first glance. For instance, it is certainly fulfilled if the operators $F - L_y$ possess the Dunn property (are co-coercive) with a common modulus $\gamma > 0$, i.e.
  \[ \langle F(x) - L_y(x) - F(v) + L_y(v), x - v \rangle \]
  \[ \geq \gamma \|F(x) - L_y(x) - F(v) + L_y(v)\|_{X'}^2, \forall x, v \in \hat{K}, \forall y \in \hat{K}, \]

which is a rather standard hypothesis for the APP.

On the other hand, let us consider the simple
10.1.2 Example.

\[ X = \mathbb{R}, \quad Q(x) = x + 4, \ K = [-1, 1], \ \hat{K} = [-2, 2], \ L_y \equiv 0 \]

and

\[ F(x) = \begin{cases} x + 2 & \text{if } x < -1 \\ x^2 & \text{if } x \geq -1 \end{cases} \]

It illustrates the situation that Assumption 10.1.1 (v) is fulfilled (with \( \gamma = 1 \)), although the operator \( F \) does not possess the Dunn property and even is not pseudo-monotone. For the definition of pseudo-monotonicity see 10.1.6.12.

It is also worth to note that in this example the operator \( F + Q \) is not monotone on \( K \).

Comparing Assumption 10.1.1 (v) with the corresponding condition in [345] (see (10.1.2)), we have to take

\[ X' := X, \ K_k := K, \ \hat{K} := K, \ L_y := L \]

in Assumption 10.1.1 (v) and to consider (10.0.4) and (10.1.2) with \( T = F + Q + N_K, \ \Xi = \nabla h + \epsilon L - \epsilon F \). Then the condition in [345] can be rewritten in the form

\[ \exists \gamma_0 > 0 : \]

\[
(\langle F(y) + q_K(y) - F(x) - q_K(x), y - x \rangle \geq \gamma_0 \| F(y) - L(y) - F(x) + L(x) \|^2
\]

\[ \forall x, y \in K \cap D(Q), \ q_K(\cdot) \in Q(\cdot) + N_K(\cdot). \quad (10.1.3) \]

From here it can be seen that the operator \( F + Q \) must be monotone on \( K \), and hence, condition (10.1.3) is not valid for Example 10.1.2. However, one can give an "opposite" example, where (10.1.3) is satisfied, but our Assumption 10.1.1 (v) is violated. Also, it should be mentioned that like (v), condition (10.1.3) is valid if \( F - L \) possesses the Dunn property and \( L \) is monotone.

Let us underline that the auxiliary operator in [345] cannot vary during the iteration process. This excludes some traditional applications of the APP, in particular, Newton - and Quasi-Newton methods.

Now, the method suggested reads as follows:

10.1.3 Method. Proximal Auxiliary Problem method (PAP-method)

Starting with \( x^1 \in \hat{K} \), the sequence \( \{x^k\} \) is defined by solving successively the auxiliary problems

\[ (P^k_x) \quad \text{find } x^{k+1} \in K_k, \ q^k(x^{k+1}) \in Q^k(x^{k+1}) : \]

\[ (F(x^k) + q^k(x^{k+1}) + L^k(x^{k+1}) - L^k(x^k) + \lambda_k (\nabla h(x^{k+1}) - \nabla h(x^k)), x - x^{k+1}) \geq -\epsilon_k \| x - x^{k+1} \| \quad \forall x \in K_k, \]

where \( \epsilon_k \downarrow 0 \) and \( L^k = L_y |_{y=x^k} \) are given for \( k = 1, 2, \ldots \).

In the sequel we are going to study the convergence of the PAP-method using Assumption 10.1.1 and the following conditions on the data of the auxiliary problems \( (P^k_x) \).
10.1.4 Assumption. \textit{(Assumptions on the auxiliary problems)}

(i) The operators $\mathcal{L}_y, y \in \hat{K}$, are monotone and Lipschitz continuous on $\hat{K}$, with common Lipschitz constant $l_L$;

(ii) $h$ is a convex functional on $X$ and the mapping $\nabla h$ is Lipschitz continuous on $\hat{K}$, with Lipschitz constant $l_h$;

(iii) for each $k$, it holds $K_k \cap D(Q^k) \neq \emptyset$ and

$$q^k(x) - q^k(y), x - y \geq (B(x - y), x - y) \forall x, y \in K_k \cap D(Q^k), \forall q^k(\cdot) \in Q^k(\cdot),$$

where $B : X \to X'$ is a given linear continuous and monotone operator with the symmetry property $\langle B(x), y \rangle = \langle B(y), x \rangle$;

(iv) with given constants $\tilde{\chi} > 0$, $m > 0$, the inequality

$$\frac{1}{2} \tilde{\chi} \langle B(x - y), x - y \rangle + h(x) - h(y) - \langle \nabla h(y), x - y \rangle \geq m \|x - y\|^2$$

is valid for all $x, y \in \hat{K}$;

(v) the regularization parameters satisfy $0 < \chi_k \leq \chi_{k+1} \leq \tilde{\chi} < \infty \ \forall \ k$;

(vi) for all $k$ and $y \in \hat{K}$, the operators $Q^k + \mathcal{L}_y + \mathcal{N}_{K_k} + \chi_k \nabla h$ are maximal monotone.

Assumptions 10.1.1 (iii)-(vi) and the monotonicity of $\mathcal{L}_y$ provide the unique solvability of the auxiliary problems ($P^k$). Moreover, the usage of the functional $h$ with properties (iii), (iv) corresponds to weak regularization or regularization on a subspace in proximal methods (see Section 8.2.3) and ensures the same kind of convergence as in the case of a strongly convex functional $h(x) = \|x\|^2$.

The next group of conditions defines the successive approximation of the pair $(Q, K)$ in the auxiliary problems ($P^k$).

10.1.5 Assumption. \textit{(Assumptions on the approximations)}

(i) for each $w \in D(Q) \cap K$, there exists a sequence $\{w^k\}$, $w^k \in D(Q^k) \cap K_k$, such that

$$\lim_{k \to \infty} \|w^k - w\| = 0, \quad \lim_{k \to \infty} \inf_{\zeta \in Q^k(w^k)} \|\zeta - q(\zeta)\|_{X'} = 0,$$

with $q(w) \in Q(w)$ (in general, $q(w)$ depends on $\{w^k\}$);

(ii) for some pair $x^*$ and $q^*$ solving $\text{VI}(F, Q, K)$, there exist a constant $\alpha > 1$ and a sequence $\{w^k\}$, $w^k \in D(Q^k) \cap K_k$, such that

$$\lim_{k \to \infty} k^\alpha \|w^k - x^*\| = 0, \quad \lim_{k \to \infty} k^\alpha \left[ \inf_{\zeta \in Q^k(w^k)} \|\zeta - q^*\|_{X'} \right] = 0;$$
(iii) with $x^*$, $q^*$ and $\alpha$ as in (ii), for any sequence $\{v^k\}$, $v^k \in K_k \cap D(Q^k)$, the relation
\[
\lim_{k \to \infty} k^{\alpha} \max \{0, \langle F(x^*) + q^*, z(v^k) - v^k \rangle \} = 0
\]
is valid;

(iv) each weak limit point of an arbitrary sequence $\{v^k\}$, $v^k \in K_k \cap D(Q^k)$, belongs to $K \cap D(Q)$.

If $\chi_k \equiv \chi \in [\chi, \bar{\chi}]$ and $G$ is a symmetric monotone operator such that $L_y - \chi G$ (taken as $L_y$) satisfies Assumption 10.1.4(i), we have a chance to weaken the Assumptions 10.1.1 and 10.1.4 considering $L_y - \chi G$ as $L_y$ and $h + g$ as $h$, where $G = \nabla g$.

Assumption 10.1.4(iii) is valid under the Assumptions 10.1.4(i) and (ii) if each operator $Q_k$ is maximal monotone and locally bounded at some $x \in clD(Q^k) \cap K_k$. Regarding also Assumption 10.1.4(iii), Assumption 10.1.4(iii) is also valid if $D(Q^k) \supset K_k$ and $Q^k$ is hemicontinuous on $K_k$ (this follows from the Theorems 1 and 3 in [350]). Other conditions ensuring Assumption 10.1.4(iii) can be derived from the results about the maximality of the sum of two monotone operators in [17, 57, 350].

In case $F$ is a compact operator, instead of the strong convergence of $\{w^k\}$ to $w$ in Assumption 10.1.5(i) one can require that $w^k \rightharpoonup w$. Of course, the both Assumptions 10.1.5(iii) and (iv) are fulfilled if
\[
k^{\alpha}(z(v^k) - v^k) \to 0
\]
holds for any sequence $\{v^k\}$, $v^k \in K_k$.

Assumptions 10.1.5(i) - (iv) about the simultaneous approximation of $Q$ and $K$ are mainly inspired by the error estimation technique in finite element methods for elliptic variational inequalities (see Appendix A2).

Constructing a sequence $\{K_k\}$, $K_k = K_{h_k}$, by means of a finite element method on a sequence of triangulations with a triangulation parameter $h_k \to 0$, we meet the following rather typical situation (see [75, 134] for detailed results):

- with an arbitrary element $v \in K$ and $v^k := \arg \min_{z \in K_k} \|v - z\|$, the asymptotic $\lim_{k \to \infty} \|v - v^k\| = 0$ is known without any practicable estimation for the rate of convergence;

- according to theorems on regularity of the solutions of elliptic problems, for an important class of variational inequalities the solutions possess a better regularity than arbitrary elements from $K$, and then for $v \in \text{SOL}(Q, K)$
  \[
  \|v - v^k\| \leq c(v)h_k^{\beta_1}
\]
is guaranteed with some $\beta_1 > 0$;

- for an arbitrary sequence $\{w^k\}$, $w^k \in K_k$, the estimate
  \[
  \min_{v \in K} \|v - w^k\| \leq c h_k^{\beta_2}
\]
holds with some $\beta_2 > 0$. 

If the operator $Q$ is Lipschitz-continuous on $\hat{K}$ and $Q^k := Q \forall k$ is taken, Assumptions 10.1.5(i) - (iv) follow from here immediately (under an appropriate variation of the triangulation parameter $h_k$). The maintenance of these assumptions in a more general case, especially with multi-valued operators $Q$, depends essentially on the special structure of the operator (see also Remark 10.2.1 b) below).

As previously mentioned, the case $Q^k := Q \forall k$ and the one, where $Q^k$ is a single-valued approximation of $Q$, are of most interest for applications. In Appendix II of [225], we analyze the fulfillment of conditions on the simultaneous approximation of a (multi-valued non-symmetric) operator $Q$ and a set $K$ for particular examples.

10.1.6 Remark. The Assumptions 10.1.4(ii) - (v) provide strong monotonicity of the operator $Q^k + \chi_k \nabla h$ on $K_k \cap D(Q^k)$, and together with Assumption 10.1.1(i) and Assumption 10.1.4(i), this yields strong monotonicity of $Q^k + L_k + N_{K_k} + \chi_k \nabla h$.

Moreover, according to Assumption 10.1.4(vi), this operator is maximal monotone. Hence, for each $k$, Problem $(P_k)$ is uniquely solvable. ♦

With $x^*$ and $q^*$ solving VI($F$, $Q$, $K$), we define

$$\Gamma^k(x^*, x) := \chi_k \langle B(x - x^*), x - x^* \rangle + h(x^*) - h(x) - \langle \nabla h(x), x^* - x \rangle + \frac{1}{\chi_k} (F(x^*) + q^*, z(x) - x^*),$$

where $z(x) := \arg \min_{v \in K} \|x - v\|$. For $x \in \hat{K}$, under Assumptions 10.1.4(iv) and (v) it holds

$$\Gamma^k(x^*, x) \geq m \|x^* - x\|^2 \quad \text{and} \quad \Gamma^{k+1}(x^*, x) \leq \Gamma^k(x^*, x).$$

(10.1.5)

Again, the sequence $\{\Gamma^k\}$ plays the role of a Ljapunov function in the further analysis.

10.1.7 Lemma. Let Assumptions 10.1.1(i), (v), (vi), Assumptions 10.1.4(ii) - (vi) as well as Assumptions 10.1.5 (ii), (iii) be fulfilled. Moreover, let $\sum_{k=1}^{\infty} \epsilon_k < \infty$ and

$$\frac{1}{4\gamma m} < \chi, \quad 2\bar{\chi} < 1.$$

(10.1.6)

Then sequence $\{x^k\}$, generated by Method 10.1.3, is bounded, $\lim_{k \to \infty} \|x^{k+1} - x^k\| = 0$ and sequence $\{\Gamma^k(x^*, x^k)\}$ converges.

This statement can be proved by modifying the proof of Theorem 2.1 in [359] in the following way.

Proof: In the sequel, we make use of the following inequalities, which are valid for arbitrary $a, b, x \in X$, $p \in X'$ and $\epsilon > 0$:

$$\langle p, a \rangle \leq \frac{1}{2\epsilon} \|p\|^2_{X'} + \frac{\epsilon}{2} \|a\|^2,$$

(10.1.7)
\[ \langle B(a - b), a - b \rangle \leq (1 + \epsilon)\langle B(a - x), a - x \rangle + \frac{1 + \epsilon}{\epsilon} \langle B(b - x), b - x \rangle. \] (10.1.8)

In order to estimate \( \Gamma^{k+1}(x^*, x^{k+1}) - \Gamma^k(x^*, x^k) \), where \( \Gamma^k \) is defined by (10.1.4), we obtain from (10.1.5) that

\[ \Gamma^{k+1}(x^*, x^{k+1}) - \Gamma^k(x^*, x^k) \leq \Gamma^k(x^*, x^{k+1}) - \Gamma^k(x^*, x^k) \]

(the existence of \( x^* \) and \( x^k \) for all \( k \) is guaranteed by Assumption 10.1.1(vi) and Remark 10.1.6). Denoting \( z^l := \arg \min_{v \in \mathbb{K}} ||x^l - v|| \), the right-hand side of this inequality can be decomposed as follows:

\[ \Gamma^k(x^*, x^{k+1}) - \Gamma^k(x^*, x^k) = s_1 + s_2 + s_3 + s_4, \]

with

\[ s_1 := h(x^k) - h(x^{k+1}) + \langle \nabla h(x^k), x^{k+1} - x^k \rangle, \]

\[ s_2 := \langle \nabla h(x^k) - \nabla h(x^{k+1}), w^k - x^{k+1} \rangle + \frac{1}{\lambda_k} \langle F(x^*) + q^*, z^{k+1} - z^k \rangle, \]

\[ s_3 := \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^* - w^k \rangle \]

and

\[ s_4 := \chi \langle B(x^{k+1} - x^*), x^{k+1} - x^* \rangle - \chi \langle B(x^k - x^*), x^k - x^* \rangle. \] (10.1.10)

We suppose that the sequence \( \{w^k\} \) satisfies Assumption 10.1.5(ii). Then Assumption 10.1.4(ii) and relation (10.1.7) yield

\[ s_3 \leq \frac{\tau}{2} ||x^{k+1} - x^k||^2 + \frac{1}{2\tau} \frac{\tau}{\lambda_k} ||x^* - w^k||^2, \] (10.1.11)

with an arbitrary \( \tau > 0 \). Setting \( x := w^k \) in Problem \( (P^k_x) \), we obtain

\[ s_2 \leq \frac{1}{\lambda_k} \langle F(x^k) + q^k(x^{k+1}) + L^k(x^{k+1}) - L^k(x^k), w^k - x^{k+1} \rangle \]

\[ + \frac{1}{\lambda_k} \langle F(x^*) + q^*, z^{k+1} - z^k \rangle + \frac{\epsilon_k}{\lambda_k} ||w^k - x^{k+1}||, \]

and due to Assumption 10.1.4(iii), for an arbitrarily chosen \( q^k(w^k) \in Q^k(w^k) \), it holds

\[ \chi_ks_2 \leq \langle F(x^k) + q^k(w^k) + L^k(x^{k+1}) - L^k(x^k), w^k - x^{k+1} \rangle \]

\[ - \langle B(w^k - x^{k+1}), w^k - x^{k+1} \rangle + \langle F(x^*) + q^*, z^{k+1} - z^k \rangle \]

\[ + \epsilon_k ||w^k - x^{k+1}|| \]

\[ = \langle F(x^k) - F(x^*) + L^k(x^*) + L^k(x^k), x^* - x^k \rangle \]

\[ + \langle F(x^*) + q^*, x^* - z^k \rangle \]

\[ + \langle F(x^k) - F(x^*) + L^k(x^k), w^k - x^* + x^k - x^{k+1} \rangle \]

\[ + \langle q^k(w^k) - q^*, x^* - z^{k+1} + w^k - x^* \rangle + \langle F(x^*) + q^*, w^k - x^* \rangle \]

\[ + \langle F(x^k) + q^*, z^{k+1} - x^{k+1} \rangle - \langle B(w^k - x^{k+1}), w^k - x^{k+1} \rangle \]

\[ + \langle L^k(x^*) - L^k(x^{k+1}), x^{k+1} - w^k \rangle + \epsilon_k ||w^k - x^{k+1}||. \] (10.1.12)
With Assumption 10.1.1(v) one can continue:

\[ \chi_k s_2 \leq -\langle B(w^k - x^{k+1}), w^k - x^{k+1} \rangle \]
\[ - \gamma \| F(x^k) - L^k(x^k) - F(x^*) + L^k(x^*) \|^2_{\chi}, \]
\[ + \langle q^*(w^k) - q^*, x^* - x^{k+1} + w^k - x^* \rangle \]
\[ + \langle F(x^k) - F(x^*) - L^k(x^k) + L^k(x^*), w^k - x^k - x^{k+1} \rangle \]
\[ + \langle F(x^*) + q^*, w^k - x^* \rangle \]
\[ + \max \{ 0, \langle F(x^*) + q^*, z^{k+1} - x^{k+1} \rangle \} \]
\[ + \langle L^k(x^*) - L^k(x^{k+1}), x^{k+1} - w^k \rangle + \epsilon_k \| w^k - x^{k+1} \|. \quad (10.1.13) \]

Due to the monotonicity of \( L^k \),

\[ \langle L^k(x^*) - L^k(x^{k+1}), x^{k+1} - w^k \rangle \leq \langle L^k(x^{k+1}) - L^k(x^*), w^k - x^* \rangle \]

is valid. Now, using the inequalities (10.1.7), (10.1.8) together with Assumption 10.1.4(i) (in order to estimate the right-hand side of (10.1.13) as well as the term \( \langle L^k(x^{k+1}) - L^k(x^*), w^k - x^* \rangle \)), we obtain

\[ \chi_k s_2 \leq \frac{\mu + \eta - 2\gamma}{2} \| F(x^k) - L^k(x^k) - F(x^*) + L^k(x^*) \|^2_{\chi}, \]
\[ + \frac{1}{2\mu} \| x^k - x^{k+1} \|^2 + \frac{1}{2\eta} \| w^k - x^* \|^2 \]
\[ - \frac{1}{\theta_k} \| B(x^{k+1} - x^*), x^{k+1} - x^* \rangle + \frac{1}{\epsilon} \langle B(w^k - x^*), w^k - x^* \rangle \]
\[ + \| F(x^*) + q^* \|_{\chi} \| w^k - x^* \| + \frac{l_c}{\theta_k} \| w^k - x^* \|^2 \]
\[ + \max \{ 0, \langle F(x^*) + q^*, z^{k+1} - x^{k+1} \rangle \} \]
\[ + \epsilon_k \| w^k - x^* \| + \epsilon_k \left( \frac{1}{4\lambda} + \lambda \| x^{k+1} - x^* \|^2 \right) \]
\[ + l_c \theta_k \| x^{k+1} - x^* \|^2 + \left( \frac{1}{2l_c \theta_k} + \frac{\theta_k}{2l_c} \right) \| q^k(w^k) - q^* \|^2_{\chi}, \quad (10.1.14) \]

with arbitrary positive \( \mu, \eta, \theta_k, \lambda \) and \( \epsilon \). Choosing

\[ \epsilon := \frac{1}{2\lambda} - 1, \quad \mu \in \left( \frac{1}{2\lambda m}, 2\gamma \right), \quad \tau := m - \frac{1}{2\lambda m}, \quad \eta \in (0, 2\gamma - \mu), \quad (10.1.15) \]

one can conclude from Assumption 10.1.4(iv), (v) and relations (10.1.6) and
(10.1.8) that

\[ \left( \tilde{\chi} - \frac{1}{\lambda_k} \frac{1}{1 + \epsilon} \right) \langle B(x^{k+1} - x^*), x^{k+1} - x^* \rangle - \tilde{\chi} \langle B(x^k - x^*), x^k - x^* \rangle \]

\[ - h(x^{k+1}) + h(x^k) + \langle \nabla h(x^k), x^{k+1} - x^k \rangle + \left( \frac{1}{2 \lambda_k \mu} + \frac{\gamma}{2} \right) \| x^{k+1} - x^k \|^2 \]

\[ \leq - \tilde{\chi} \left[ \langle B(x^{k+1} - x^*), x^{k+1} - x^* \rangle + \langle B(x^k - x^*), x^k - x^* \rangle \right] \]

\[ - h(x^{k+1}) + h(x^k) + \langle \nabla h(x^k), x^{k+1} - x^k \rangle \]

\[ + \left( \frac{1}{2 \lambda_k \mu} + \frac{m}{2} - \frac{1}{4 \mu} \right) \| x^{k+1} - x^k \|^2 \]

\[ \leq - \frac{\tilde{\chi}}{2} \langle B(x^{k+1} - x^*), x^{k+1} - x^* \rangle - h(x^{k+1}) + h(x^k) + \langle \nabla h(x^k), x^{k+1} - x^k \rangle \]

\[ + \left( \frac{m}{2} + \frac{1}{4 \mu} \right) \| x^{k+1} - x^k \|^2 \leq \left( - \frac{m}{2} + \frac{1}{4 \mu} \right) \| x^{k+1} - x^k \|^2, \quad (10.1.16) \]

and that \(- \frac{m}{2} + \frac{1}{4 \mu} < 0\). Now, we sum up (10.1.9), (10.1.10) and the estimates for \( s_2, s_3 \) in (10.1.11), (10.1.14), and insert (10.1.15) and (10.1.16) in this sum. This yields

\[ \Gamma^{k+1}(x^*, x^{k+1}) - \Gamma^k(x^*, x^k) \leq \left( - \frac{m}{2} + \frac{1}{4 \mu} \right) \| x^{k+1} - x^k \|^2 \]

\[ + \frac{\lambda}{\lambda_k} \left[ \frac{1}{2 \eta} \| F(x^k) - L^k(x^k) - F(x^*) + L^k(x^*) \|^2 \right] \]

\[ + \frac{\lambda}{\lambda_k} \left[ \left( \frac{1}{2 \eta} \| \theta_k \theta_k \| + \frac{h_k^2}{2 \theta} \right) \| x^* - w^k \|^2 + l C \| x^{k+1} - x^* \|^2 \right] \]

\[ + \left( \frac{1}{2 \theta k} + \frac{\theta_k}{\theta k} \right) \| q^k(w^k) - q^* \|^2 \]

\[ + \frac{1}{\epsilon} \langle B(w^k - x^*), w^k - x^* \rangle + \| F(x^*) + q^* \| \| w^k - x^* \| \]

\[ + \max \{ 0, \langle F(x^*) + q^* , z^{k+1} - x^{k+1} \rangle \} \]

\[ + \frac{\lambda_k}{\lambda_k} \| w^k - x^* \| + \frac{\lambda_k}{\lambda_k} \left( \frac{1}{4 \lambda} + \lambda \| x^{k+1} - x^* \|^2 \right). \quad (10.1.17) \]

But, from the first inequality in (10.1.5) one gets

\[ \| x^{k+1} - x^* \|^2 \leq \frac{1}{m} \Gamma^{k+1}(x^*, x^{k+1}) \]
and (10.1.17) leads to

\[
\left( 1 - \frac{l_c \theta_k}{\lambda m} - \frac{\lambda \epsilon_k}{\lambda m} \right) \Gamma^{k+1}(x^*, x^{k+1}) \\
\leq \Gamma^k(x^*, x^k) + \frac{1}{\lambda} \left[ \left( \frac{l_c}{2\eta} + \frac{\lambda \epsilon_k}{\lambda m} \right) \|w^k - x^*\|^2 \\
+ \left( \frac{1}{2l_c \theta_k} + \frac{\theta_k}{2l_c} \right) \|q^k(w^k) - q^*\|_X^2, \\
\right.
\]

\[\left. + \frac{1}{\epsilon} \langle B(w^k - x^*), w^k - x^* \rangle + \|F(x^*) + q^*\|_X \|w^k - x^*\| \\
+ \max [0, \langle F(x^*) + q^*, z^{k+1} - x^{k+1} \rangle] \right] \\
+ \left( - \frac{m}{2} + \frac{1}{4\lambda \mu} \right) \|x^{k+1} - x^k\|^2 + \frac{\epsilon_k}{\lambda} \|w^k - x^*\| + \frac{\epsilon_k}{4\lambda}. \tag{10.1.18} \]

Choose \( \lambda \in \left(0, \frac{m}{\sup_{x \in X} \epsilon(x^*)} \right) \) and set \( \theta_k := \theta k^{-\alpha} \), where \( \theta \in \left(0, \frac{m}{2\epsilon_0} \right) \) and \( \alpha > 1 \) is the same as in Assumption 10.1.5(ii). Then, we obtain with \( d_0 := \frac{\lambda \epsilon_0}{\lambda m} \) that \( 1 - d_0 k^{-\alpha} - d_1 \epsilon_k > \frac{1}{2} - d_0 > 0 \), and with regard to \(-\frac{m}{2} + \frac{1}{4\lambda \mu} < 0 \), one can conclude for \( k = 1, 2, \ldots \) that

\[
\Gamma^{k+1}(x^*, x^{k+1}) \leq \frac{1}{1 - d_0 k^{-\alpha} - d_1 \epsilon_k} \Gamma^k(x^*, x^k) \\
+ \frac{1}{\lambda (1/2 - d_0)} \left[ \left( \frac{l_c}{2\eta} + \frac{\lambda \epsilon_k}{\lambda m} \right) \|w^k - x^*\|^2 \\
+ \left( \frac{k^\alpha}{2l_c \theta} + \frac{\theta}{2l_c} \right) \|q^k(w^k) - q^*\|_X^2, \\
\right.
\]

\[\left. + \frac{1}{\epsilon} \langle B(w^k - x^*), w^k - x^* \rangle + \|F(x^*) + q^*\|_X \|w^k - x^*\| \\
+ \max [0, \langle F(x^*) + q^*, z^{k+1} - x^{k+1} \rangle] \right] + \epsilon_k \|w^k - x^*\| + \frac{\epsilon_k}{4\lambda}. \tag{10.1.19} \]

Taking into account that \( q^k(w^k) \) is an arbitrary element of \( Q^k(w^k) \), in view of

\[
\frac{1}{1 - d_0 k^{-\alpha} - d_1 \epsilon_k} \leq 1 + \frac{d_0 k^{-\alpha} + d_1 \epsilon_k}{1/2 - d_0}, \quad \sum_{k=1}^{\infty} \epsilon_k < \infty,
\]

and the Assumptions 10.1.5(ii) and (iii), convergence of the sequence \( \{\Gamma^k(x^*, x^k)\} \) follows from Lemma A3.1.7, and the first inequality in (10.1.5) ensures the boundedness of the sequence \( \{x^k\} \). Passing to the limit in (10.1.18), one can immediately conclude that \( \lim_{k \to \infty} \|x^{k+1} - x^k\| = 0. \)

10.1.8 Lemma. Let Assumptions 10.1.1(i) - (iv), Assumptions 10.1.4(i) - (iii) and Assumptions 10.1.5(i), (iv) be fulfilled, and \( 0 < \chi_k \leq \tilde{\chi} \) holds for all \( k \).
Moreover, let the sequence \( \{x^k\} \) generated by Method 10.1.3 be bounded, and \( \lim_{k \to \infty} \|x^{k+1} - x^k\| = 0. \)

Then each weak limit point of \( \{x^k\} \) is a solution of VI\( (F, Q, K) \).

Proof: Let \( S \subseteq \{1, 2, \ldots \} \) and \( \bar{x} \) be an arbitrary weak limit point of \( \{x^k\} \) and let \( \{x^k\}_{k \in S} \) converge weakly to \( \bar{x} \). Due to Assumption 10.1.5(iv), \( \bar{x} \) belongs to
10.1. EXTENDED AUXILIARY PROBLEM PRINCIPLE

$K \cap D(Q)$. Since $\lim_{k \to \infty} \|x^{k+1} - x^k\| = 0$, one gets $x^{k+1} \to \bar{x}$ if $k \in \mathcal{S}$, $k \to \infty$. According to Assumption 10.1.5(i), for each $y \in D(Q) \cap K$ one can choose a sequence $\{y^k\}$, $y^k \in D(Q^k) \cap K_k$, such that $\lim_{k \to \infty} \|y^k - y\| = 0$, and

$$
\lim_{k \to \infty} \|q^k(y^k) - q(y)\|_{x^k} = 0
$$

(10.1.20)

holds true with some $q^k(y^k) \in Q^k(y^k)$ and $q(y) \in Q(y)$. By definition of $x^{k+1}$, for suitably chosen $q^k(x^{k+1}) \in Q^k(x^{k+1})$

$$
\langle F(x^k) + q^k(x^{k+1}) + \mathcal{L}^k(x^{k+1}) - \mathcal{L}^k(x^k)
\quad + \chi_k (\nabla h(x^{k+1}) - \nabla h(x^k)) , y^k - x^{k+1} \rangle \geq -\epsilon_k \|y^k - x^{k+1}\|
$$

is valid, and the monotonicity of $Q^k$ (see Assumption 10.1.4(iii)) leads to

$$
\langle F(x^k) + q^k(y^k) + \mathcal{L}^k(x^{k+1}) - \mathcal{L}^k(x^k)
\quad + \chi_k (\nabla h(x^{k+1}) - \nabla h(x^k)) , y^k - x^{k+1} \rangle \geq -\epsilon_k \|y^k - x^{k+1}\|. \quad (10.1.21)
$$

From the identity

$$
\langle F(x^k), y^k - x^{k+1} \rangle = \langle F(x^k), y^k - y \rangle + \langle F(x^k), x^k - x^{k+1} \rangle + \langle F(x^k), y \rangle - \langle F(x^k), x^k \rangle,
$$

using the relations

$$
\lim_{k \to \infty} \|y^k - y\| = 0, \quad x^k \to \bar{x} (k \in \mathcal{S}), \quad \lim_{k \to \infty} \|x^{k+1} - x^k\| = 0
$$

(10.1.22)

and Assumption 10.1.1 (iv), we obtain that

$$
\lim_{k \to \infty} \langle F(x^k), y^k - x^{k+1} \rangle \leq \langle F(\bar{x}), y \rangle + \lim_{k \to \infty} \langle F(x^k), -x^k \rangle
= \langle F(\bar{x}), y \rangle - \lim_{k \to \infty} \langle F(x^k), x^k \rangle \leq \langle F(\bar{x}), y - \bar{x} \rangle. \quad (10.1.23)
$$

Now, passing to the limit for $k \in \mathcal{S}$ in (10.1.21), the relations (10.1.20), (10.1.23), $0 < \chi_k \leq \bar{\chi}$, $\lim_{k \to \infty} \epsilon_k = 0$ and the Assumptions 10.1.4(i), (ii) imply

$$
\langle F(\bar{x}) + q(y), y - \bar{x} \rangle \geq 0.
$$

Moreover, due to Assumptions 10.1.1(ii), (iii), the operators $Q_0 : y \to F(\bar{x}) + Q(y)$ and $Q_0 + N_K$ are maximal monotone, and the operator

$$
Q_K : y \to \begin{cases} Q_0(y) & \text{if } v \in K \\ \emptyset & \text{otherwise} \end{cases}
$$

is locally hemi-bounded at each point of $K$. Thus, the application of Lemma 9.3.8 with $u = \bar{x}$, $C = K$ and $A_0 = Q_0$ yields

$$
\langle F(\bar{x}) + q(\bar{x}), y - \bar{x} \rangle \geq 0 \quad \forall y \in K,
$$

where $q(\bar{x}) \in Q(\bar{x})$. Hence, $\bar{x}$ is a solution of VI$(F, Q, K)$. \qed
10.1.9 Lemma. Let $\mathcal{S} \subset \{1, 2, \ldots\}$, assume that the sequence $\{v^k\}_{k \in \mathcal{S}}$, $v^k \in K_k$, converges weakly to some $x^* \in \text{SOL}(F, Q, K)$ and
\[
\lim_{k \to \infty} \max_{\mathcal{S}} \{0, \langle F(x^*) + q^*, z(v^k) - v^k \rangle\} = 0.
\tag{10.1.24}
\]
Then
\[
\lim_{k \to \infty} \langle F(x^*) + q^*, z(v^k) - x^* \rangle = 0
\tag{10.1.25}
\]
holds true.

**Proof:** Due to $x^* \in \text{SOL}(F, Q, K)$ and $z(v^k) \in K$, one gets
\[
\langle F(x^*) + q^*, z(v^k) - x^* \rangle \geq 0 \quad \forall k \in \mathcal{S}.
\tag{10.1.26}
\]
The sequence $\{v^k\}_{k \in \mathcal{S}}$, being weakly convergent, is bounded, and hence the both sequences $\{z(v^k)\}_{k \in \mathcal{S}}$ and $\{z(v^k) - v^k\}_{k \in \mathcal{S}}$ are bounded, too. Taking $\mathcal{S}_0 \subset \mathcal{S}$ such that the subsequence $\{z(v^k)\}_{k \in \mathcal{S}_0}$ is weakly convergent, we obtain from (10.1.26)
\[
0 \leq \lim_{k \to \infty} \langle F(x^*) + q^*, z(v^k) - v^k \rangle + \lim_{k \to \infty} \langle F(x^*) + q^*, v^k - v^k \rangle
\]
\[
= \lim_{k \to \infty} \langle F(x^*) + q^*, z(v^k) - v^k \rangle.
\]
If
\[
\lim_{k \to \infty} \langle F(x^*) + q^*, z(v^k) - v^k \rangle > 0,
\]
then obviously,
\[
0 < \lim_{k \to \infty} \max_{\mathcal{S}_0} \{0, \langle F(x^*) + q^*, z(v^k) - v^k \rangle\},
\]
but this contradicts (10.1.24). Hence
\[
\lim_{k \to \infty} \langle F(x^*) + q^*, z(v^k) - v^k \rangle = 0
\]
has to be, and the relations
\[
\lim_{k \to \infty} \langle F(x^*) + q^*, z(v^k) - v^k \rangle = 0
\]
and (10.1.25) follow immediately. \(\square\)

Condition (10.1.24) is obviously guaranteed if Assumption 10.1.5(iii) is fulfilled.

10.1.10 Theorem. Let the Assumptions 10.1.1 and 10.1.4 and condition (10.1.6) be fulfilled and let $\sum_{k=1}^{\infty} \epsilon_k < \infty$. Then the following conclusions are true:

(i) Problem (P) is uniquely solvable for each $k$, the sequence $\{x^k\}$ generated by Method 10.1.3 is bounded, and each weak limit point of $\{x^k\}$ is a solution of $\text{VI}(F, Q, K)$;

(ii) if, in addition, the Assumptions 10.1.5(ii), (iii) (with common $\alpha > 1$) are valid for all $x \in \text{SOL}(F, Q, K)$ and
\[
z^k \rightharpoonup z \text{ in } X, \quad z^k \in K_k \implies \nabla h(z^k) \rightharpoonup \nabla h(z) \text{ in } X',
\tag{10.1.27}
\]
then the whole sequence $\{x^k\}$ converges weakly to a solution of $\text{VI}(F, Q, K)$;
Due to subsequences converging weakly to $\bar{x}$, holds true, whereas the monotonicity of $\{x^k\}_{k \in \mathbb{N}}$, $\{\bar{x}^k\}_{k \in \mathbb{N}}$ are two subsequences converging weakly to $\bar{x}$, $\tilde{x}$, respectively. Then, according to (i), $\bar{x}$ and $\tilde{x}$ belong to $\text{SOL}(\mathcal{F}, Q, K)$, and because the Assumptions 10.1.5(ii) and (iii) are valid for each $x \in \text{SOL}(\mathcal{F}, Q, K)$, Lemma 10.1.7 ensures that the sequences $\{\Gamma^k(\bar{x}, x^k)\}_{k=1}^{\infty}$, $\{\Gamma^k(\tilde{x}, x^k)\}_{k=1}^{\infty}$ are convergent (let us remind that the choice of $q^*(\cdot)$ in $\Gamma^k$ satisfies Assumption 10.1.5 (ii)).

Due to $\bar{x} \in \text{SOL}(\mathcal{F}, Q, K)$, $z(x) \in K$ and the definition of $q^*(\tilde{x})$, the inequality

$$\langle \mathcal{F}(\bar{x}) + q^*(\tilde{x}), z(x) - \bar{x} \rangle \geq 0$$

holds true, whereas the monotonicity of $\mathcal{B}$ and Assumption 10.1.4 (iv) provide that

$$h(\bar{x}) - h(\tilde{x}) - \langle \nabla h(\tilde{x}), \bar{x} - \tilde{x} \rangle + \tilde{\chi}(\mathcal{B}(\bar{x} - \tilde{x}), \bar{x} - \tilde{x}) \geq m\|\bar{x} - \tilde{x}\|^2.$$

From these inequalities and the symmetry property of the operator $\mathcal{B}$, we obtain for $x \in K$

$$\Gamma^k(\bar{x}, x) - \Gamma^k(\tilde{x}, x)$$

$$= \langle h(\bar{x}) - h(\tilde{x}) - \langle \nabla h(\tilde{x}), \bar{x} - \tilde{x} \rangle, \nabla h(x) - \nabla h(x) + \bar{x} - \tilde{x} \rangle - \frac{1}{\lambda_k} \langle \mathcal{F}(\bar{x}) + q^*(\tilde{x}), z(x) - \tilde{x} \rangle$$

$$+ \tilde{\chi}(\mathcal{B}(\bar{x} - \tilde{x}), \bar{x} - \tilde{x}) + 2\chi(\mathcal{B}(\bar{x} - \tilde{x}), \bar{x} - x)$$

$$\geq m\|\bar{x} - \tilde{x}\|^2 + \langle \nabla h(\tilde{x}) - \nabla h(x), \bar{x} - \tilde{x} \rangle$$

$$- \frac{1}{\lambda_k} \langle \mathcal{F}(\bar{x}) + q^*(\tilde{x}), z(x) - \tilde{x} \rangle + 2\chi(\mathcal{B}(\bar{x} - \tilde{x}), \bar{x} - x).$$

Inserting $x := x^k$ in (10.1.29) and passing to the limit for $k \in \mathcal{G}_2$, one can conclude from (10.1.27), Assumption 10.1.4(v), Assumption 10.1.5(iii) and Lemma 10.1.9 that

$$\bar{\gamma} - \tilde{\gamma} \geq m\|\bar{x} - \tilde{x}\|^2,$$

where $\bar{\gamma} := \lim_{k \to \infty} \Gamma^k(\bar{x}, x^k)$, $\tilde{\gamma} := \lim_{k \to \infty} \Gamma^k(\tilde{x}, x^k)$.

Obviously, in the same way the "symmetric" inequality

$$\bar{\gamma} - \tilde{\gamma} \geq m\|\bar{x} - \tilde{x}\|^2$$

can be concluded, and therefore $\bar{x} = \tilde{x}$ is valid, proving the uniqueness of the weak limit point of $\{x^k\}$.

Denoting this limit point by $x^*$, now we suppose additionally that relation (10.1.28) is fulfilled. Choosing $\{w^k\}$ according to Assumption 10.1.5(ii), $q^k(x^{k+1})$ as in Problem $(P^k_w)$ and an arbitrary $q^k(w^k) \in Q^k(w^k)$, then Assumption
10.1.4 (iii) yields
\[
\langle B(x^{k+1} - x^*), x^{k+1} - x^* \rangle \\
= \langle B(x^{k+1} - w^k), x^{k+1} - w^k \rangle - \langle B(x^* - w^k), x^{k+1} - x^* \rangle \\
- \langle B(x^{k+1} - w^k), x^* - w^k \rangle \\
\leq \langle q^k(x^{k+1}) - q^k(w^k), x^{k+1} - w^k \rangle - \langle B(x^{k+1} - x^*), x^* - w^k \rangle \\
- \langle B(x^{k+1} - w^k), x^* - w^k \rangle.
\] (10.1.30)

To estimate the term \(\langle q^k(x^{k+1}), x^{k+1} - w^k \rangle\), we use Problem \((P^k)\). Together with (10.1.30) this gives
\[
\langle B(x^{k+1} - x^*), x^{k+1} - x^* \rangle \leq \langle F(x^k) - F(x^*), w^k - x^{k+1} \rangle \\
+ \chi_k (\nabla h(x^{k+1}) - \nabla h(x^k), w^k - x^{k+1}) + (L^k(x^{k+1}) - L^k(x^k), w^k - x^{k+1}) \\
+ \langle F(x^*), w^k - x^* \rangle + \langle F(x^*), x^* - x^{k+1} \rangle + \langle q^k(w^k) - q^k(x^*), w^k - x^{k+1} \rangle \\
+ \langle Bx^* + Bw^k - 2Bx^{k+1}, x^* - w^k \rangle + \epsilon_k \|w^k - x^{k+1}\|. \tag{10.1.31}
\]

Taking into account that the sequences \(\{w^k\}\) and \(\{x^k\}\) are bounded, one can derive that in the right part of inequality (10.1.31) all terms except for the first one tend to 0 as \(k \to \infty\).

Indeed, term 2 in view of Assumptions 10.1.4(ii), (v) and \(\|x^{k+1} - x^k\| \to 0\); term 3 because of Assumption 10.1.4(i) and \(\|x^{k+1} - x^k\| \to 0\); the terms 4, 6 and 7 due to Assumption 10.1.5(ii); the terms 5 and 8 owing to \(x^k \to x^*\); term 9 on account of Assumption 10.1.5(iii) and the boundedness of \(B\), and finally term 10 by \(\epsilon_k \to 0\).

But, due to Assumption 10.1.1(iv),
\[
\lim_{k \to \infty} \langle F(x^k) - F(x^*), x^* - x^k \rangle \\
= \lim_{k \to \infty} \langle F(x^k), x^* \rangle - \langle F(x^*), x^k \rangle + \lim_{k \to \infty} \langle F(x^*), x^k \rangle \\
+ \lim_{k \to \infty} \langle F(x^k), -x^k \rangle \\
= \langle F(x^*), x^* \rangle - \lim_{k \to \infty} \langle F(x^k), x^k \rangle \leq 0, \tag{10.1.32}
\]

and using the identity
\[
\langle F(x^k) - F(x^*), w^k - x^{k+1} \rangle \\
= \langle F(x^k) - F(x^*), w^k - x^k + x^k - x^{k+1} \rangle + \langle F(x^k) - F(x^*), x^k - x^k \rangle,
\]
the inequality
\[
\lim_{k \to \infty} \langle F(x^k) - F(x^*), w^k - x^{k+1} \rangle \leq 0 \tag{10.1.33}
\]
follows immediately from \(x^k \to x^*\), \(\|x^k - x^{k+1}\| \to 0\), (10.1.32), Assumption 10.1.1(iv) and Assumption 10.1.5(ii).

Now, the relation
\[
\lim_{k \to \infty} \langle B(x^{k+1} - x^*), x^{k+1} - x^* \rangle = 0
\]
can be deduced from (10.1.31). Together with (10.1.28) and Assumption 10.1.4(iv) this provides conclusion (iii). \(\Box\)
10.2. Modifications and Applications of the PAP-Method

10.1.11 Remark. Obviously, the conditions (10.1.6) used in Lemma 10.1.7 and in Theorem 10.1.10 are compatible if and only if \(2\gamma m > \tilde{\chi}\). But they are certainly compatible, for instance, if the regularizing functional \(h\) is strongly convex. In this case, assuming \(m\) is the modulus of the strong convexity of \(h\), an arbitrary small \(\tilde{\chi} > 0\) is appropriate in Assumption 10.1.4(iv). Also, the inequality \(2m\gamma > \tilde{\chi}\) can be always satisfied if \(F\) is a monotone operator and we deal with proximal-like methods, which correspond formally to the PAP-method by setting \(Q^k := Q^k + F, \ F := 0, \ L^k := 0\) (of course, \(h\) has to obey Assumption 10.1.4(iv)). In this case, Assumption 10.1.1(v) is valid for arbitrary large \(\gamma\).

Working with conditions which are similar to Assumption 10.1.4(iii), (iv), proximal-like methods with weak regularization and regularization on a subspace have been developed in Chapters 8 and 7 for solving problems in elasticity theory and optimal control. In general, one should choose the constants \(\tilde{\chi}\) and \(m\) such that Assumption 10.1.4(iv) is satisfied and the ratio \(\frac{\tilde{\chi}}{m}\) is as small as possible.

10.1.12 Remark. If \(x^{k+1} \in K_k, \ q^k(x^{k+1}) \in Q^k(x^{k+1})\) and a residual vector \(r^k \in X'\) satisfy

\[
F(x^k) + q^k(x^{k+1}) + L^k(x^{k+1}) - L^k(x^k) + \chi_k (\nabla h(x^{k+1}) - \nabla h(x^k)) + \mathcal{N}_{K_k}(x^{k+1}) \ni r^k,
\]

with \(\|r^k\|_{X'} \leq \varepsilon_k\), then the straightforward use of the definition of the normality operator shows that \(x^{k+1}\) is a solution of Problem \((P^k)\).

This permits us to apply the convergence analysis of this section to methods working with summable error criteria (10.1.34) of Eckstein’s type (cf. [97]).

10.2 Modifications and Applications of the PAP-method

We are going to observe extensions of the PAP-method 10.1.3 to different types of variational inequalities.

10.2.1 Extension of PAP-method to different types of VI’s

For the hemi-variational inequality

\[
\text{HVI}(F, Q, f, X) \quad \text{find } x^* \in X, \ q^*(x^*) \in Q(x^*) : \ (F(x^*) + q^*(x^*), x - x^*) + f(x) - f(x^*) \geq 0, \ \forall \ x \in X,
\]

where \(f : X \to \overline{\mathbb{R}} \equiv \mathbb{R} \cup \{+\infty\}\) is a convex lsc functional and the operators \(Q + \partial f\) (instead of \(Q\)) and \(F\) satisfy Assumption 10.1.1 with \(K = \text{dom } f\), the studied scheme can be directly applied if we construct the auxiliary problems \((P^k)\) being caused from the equivalent variational inequality:

\[
\text{VI}(F, Q, \text{dom } f) \quad \text{find } x^* \in \text{dom } f, \ s^*(x^*) \in Q(x^*) + \partial f(x^*) : \ (F(x^*) + s^*(x^*), x - x^*) \geq 0, \ \forall \ x \in \text{dom } f.
\]
However, one can handle more convenient requirements for the approximation of \( f \) if the auxiliary problems have the form

\[
\begin{align*}
(P^k_f) & \quad \text{find } x^{k+1} \in X, \ q^k(x^{k+1}) \in Q^k(x^{k+1}); \\
& \quad \langle F(x^k) + q^k(x^{k+1}) + L^k(x^{k+1}) - L^k(x^k) \\
& \quad + \chi_k \left( \nabla h(x^{k+1}) - \nabla h(x^k) \right), x - x^{k+1} \\
& \quad + f^k(x) - f^k(x^{k+1}) \geq -\epsilon_k \|x - x^{k+1}\| \quad \forall \ x \in X,
\end{align*}
\]

where \( f^k : X \to \mathbb{R} \) is a convex lsc functional. In this case, merging the convergence analysis from SALMON, NGUYEN AND STRODIOT in [359] and in Subsection 10.1.2, Theorem 10.1.10 can be proved under the following modifications of the Assumptions 10.1.1, 10.1.4 and 10.1.5:

- in the Assumptions 10.1.1 and 10.1.4 (i), (ii), (iv) replace \( K \) and \( \hat{K} \) by \( \text{dom} f \), and \( N_K \) by \( \partial f \), respectively;

- Assumption 10.1.1(iii)’:

\[
\langle F(y) - L_y(y) + L_y(x) + q(x), y - x \rangle + f(y) - f(x) \\
\geq \gamma \|F(y) - L_y(y) - F(x) + L_y(x)\|_Y, \quad (\gamma > 0 \text{ const.})
\]

holds true, whenever \( x \in D(Q) \cap \text{dom} f \) and

\[
\langle F(x) + q(x), v - x \rangle + f(v) - f(x) \geq 0 \quad \forall \ v \in \text{dom} f;
\]

- Assumption 10.1.4(iii)’:

\[
\langle g^k(x) - q^k(y), x - y \rangle + f^k(x) - f^k(y) - \langle g^k(y), x - y \rangle \geq \langle B(x - y), x - y \rangle, \\
\forall \ x, y \in D(Q) \cap D(\partial f), \quad \forall \ q^k(y) \in Q^k(y), \quad \forall \ g^k(y) \in \partial f^k(y),
\]

with \( B \) as in the former Assumption 10.1.4(iii);

- Assumption 10.1.4(vi)’:

for all \( k \) and \( y \) in \( \text{dom} f \), the operators \( Q^k + L_y + \partial f^k + \chi_k \nabla h \) are maximal monotone;

- Assumption 10.1.5(i)’:

\[
f^k \geq f, \ D(Q^k) \cap \text{dom} f^k \subset D(Q) \cap \text{dom} f, \quad \text{and for each } w \in D(Q) \cap \text{dom} f
\]
there exists a sequence \( \{w^k\}, \ w^k \in D(Q^k) \cap \text{dom} f^k \), such that

\[
\lim_{k \to \infty} \|w^k - w\| = 0,
\]

\[
\lim_{k \to \infty} f^k(w^k) = f(w),
\]

\[
\lim_{k \to \infty} \inf_{\zeta \in Q^k(w^k)} \|\zeta - q(w)\|_{X'} = 0,
\]

with \( q(w) \in Q(w) \);
10.2. MODIFICATIONS AND APPLICATIONS OF THE PAP-METHOD

- Assumption 10.1.5(ii)′:
  for some solution $x^*$ of $HVI(F, Q, X)$ there exist a constant $\alpha > 1$ and a sequence $\{w^k\}$, $w^k \in D(Q^k) \cap \text{dom} f^k$, such that
  \[
  \lim_{k \to \infty} k^\alpha \|w^k - x^*\| = 0, \\
  \lim_{k \to \infty} k^\alpha \inf_{\zeta \in \Phi^{(w^k)}} \|\zeta - \hat{q}(x^*)\|_{X'} = 0 \\
  \lim_{k \to \infty} k^\alpha \max\{f^k(w^k) - f(x^*), 0\} = 0.
  \]
  Here $\hat{q}(x^*) \in Q(x^*)$ has to satisfy
  \[
  (F(x^*) + \hat{q}(x^*, x - x^*) + f(x) - f(x^*)) \geq 0 \quad \forall x \in \text{dom} f;
  \]

- the Assumptions 10.1.5 (iii) and (iv) must be skipped.

10.2.1 Remark.

a) The version (10.1.1) of the APP-method developed in [359] deals with a finite-dimensional $HVI(F, Q, f, X)$ where $Q = 0$. In that case Assumption 10.1.1 (v)' coincides with the condition linking $F$, $L$ and $f$ in the mentioned paper, see Theorem 2.1 there. Moreover, our convergence result for process $(P_k)$ covers up this theorem for method (10.1.1) with $h_k \equiv h$.

b) Of course, real possibilities to satisfy the conditions 10.1.5(i), (ii), (ii)', respectively, are connected with mutual properties of $Q$ and $K$ or $Q$ and $f$. The general way based on the Moreau-Yosida approximation is very expensive. However, the use of this approximation within a special algorithm of the APP in [101] and the proximal method in [298] seems to be promising. If $Q^k \equiv Q$ and $K_k \supset K$, the Assumptions 10.1.5 (i), (ii) are automatically fulfilled.

A couple of problems in mathematical physics, for instance the problem of linear elasticity with friction, the Bingham problem (cf. [95, 134]) etc., takes the form $HVI(F, Q, f, X)$ with a single-valued operator $Q$. In this case a uniform approximation of $f$ by a sequence of differentiable functionals is possible (see [134, 219, 316]) and there are no serious difficulties in satisfying Assumptions 10.1.5(i)', (ii)'.

A simultaneous approximation of $X$ and $f$ by a sequence of subspaces $\{X^k\}$ and a sequence of functionals $\{\hat{f}^k\}$ (in particular, with better smoothness properties) can be inserted in the scheme above setting $f^k = \hat{f}^k + \text{ind}(\cdot|X^k)$ in $(\tilde{P}_k^*)$.

It should also be noted that $VI(F, Q, X)$ in Section 10.1.1 takes the form of $HVI(F, Q, f, X)$ with $f = \text{ind}(\cdot|K)$. The auxiliary problems $(P_k)$ are obtained from $(\tilde{P}_k^*)$ by setting $f^k = \text{ind}(\cdot|K_k)$. But, the condition $f^k \geq f$ in Assumption 10.1.5(i)' allows us to consider only the case $K_k \subset K$. ◊
10.2.2 Extension of PAP-method to inclusions

Theorem 10.1.10 can be extended to methods treating IP($\mathcal{T}, X$) by the use of the relationship

$$0 \in \mathcal{T}(x) \iff x \in K \text{ and } \exists t(x) \in \mathcal{T}(x) : \langle t(x), y - x \rangle \geq 0 \forall y \in K,$$

where $K$ is an arbitrary convex closed set such that $K \supset D(\mathcal{T})$.

So, scheme (10.0.3) developed for this inclusion by Renaud and Cohen in [345] can be considered as a particular realization of Method 10.1.3 if we split the operator $\mathcal{T}$ into a sum $\mathcal{F} + \mathcal{Q}$ maintaining the conditions for the operators $\mathcal{F}$ and $\mathcal{Q}$ in Assumption 10.1.1 (with a convex set $K \supset D(\mathcal{Q})$ and $\hat{K} = K$). Indeed, (10.0.3) corresponds to $(P_k^\varepsilon)$ if one takes

$$X' := X, \quad K_k := K, \quad Q_k := \mathcal{Q},$$

$$\nabla h := \Xi_1, \quad L_k := \Xi_2 + \mathcal{F},$$

$$\chi_k := \frac{1}{\varepsilon}, \quad \epsilon_k := 0.$$

A partial case of scheme (10.0.3) with $\Xi_1 = \mathcal{I}$ and $\Xi_2 = -\mathcal{F}$ is known as the Lions-Mercier splitting algorithm (cf. [273]):

$$x^{k+1} = (I + \varepsilon \mathcal{Q})^{-1}(I - \varepsilon \mathcal{F})(x_k). \quad (10.2.1)$$

Its convergence analysis can be found in [125, 319]. A straightforward adaptation of Theorem 10.1.10 yields new convergence results for the method (10.0.3) as well as for the Lions-Mercier algorithm.

On the basis of Remark 10.1.12, Theorem 10.1.10 can be extended also to the inexact version of method (10.0.3):

$$\text{find } x^{k+1} \in X : \mathcal{T}(x^{k+1}) + \frac{1}{\varepsilon} [\Xi(x^{k+1}) - \Xi(x_k)] \ni r_k,$$

with residual vectors $r^k$ such that $\sum_{k=1}^{\infty} \|r^k\| < \infty$. In [345] convergence has been studied for the exact method only.

10.2.3 Decomposition properties of PAP-method

Due to the splitting of the main operator in VI($\mathcal{F}, \mathcal{Q}, K$) into a sum $\mathcal{F} + \mathcal{Q}$, certain decomposition properties are already inherent in the auxiliary problem ($P_k^\varepsilon$), where $x^{k+1}$ is calculated with $\mathcal{F}$ fixed at the point $x^k$. Such a splitting can be caused by several reasons, in particular,

- traditional APP-schemes assume usually that the operator in the variational inequality is single-valued, whereas - applying proximal methods - the operator is supposed to be monotone, but not necessary single-valued (concerning relaxations of these conditions see [79], Sect. 3 and [220, 379]).

The class of problems admitting the use of Method 10.1.3 is wider, including variational inequalities with an operator $\mathcal{F} + \mathcal{Q}$, whose "geometrical" properties are defined by Assumptions 10.1.1(ii), (v);
for some variational inequalities, the problems \((P_k^k)\) can be solved easier than the auxiliary problems arising in proximal methods. This fact was motivation for the Lions-Mercier algorithm (10.2.1) for Problem (10.0.3), although the conditions assumed for the operator \(T\) in [273, 125] permit a straightforward use of the proximal point method, too.

Concerning traditional applications of the APP to decomposition and linear approximation methods we point out the papers [78, 223, 282] for general concepts and [101, 317, 400, 401] for special algorithms which can be incorporated into the APP.

10.3 Comments

Let us come back to the weakening of the standard (for the APP) assumption on strong convexity of the auxiliary function \(h\) in the Problems \((P_k^k)\) in the sense that we supposed only \(h\) to be convex and the operators \(Q_k + \nabla h\) have to be strongly monotone with a common modulus for all \(k\). This enables us to extend the APP-scheme to proximal-like methods with weak regularization and regularization on a subspace (cf. Section 8.2.3), as well as to methods using Bregman functions (see Section 9.3), if the chosen Bregman function ensures the mentioned property of the strong monotonicity for \(Q_k + \nabla h\). In the latter case, however, substantial modifications in the basic assumptions as well as in the convergence analysis are needed. For instance, if \(K_k := K, \hat{K} := K\) and \(h\) is a Bregman function with a zone \(S = \text{int}K\), complications emerge due to the non-differentiability of \(h\) on the boundary of \(K\). Such kind of convergence analysis is carried out in the papers [227, 230].
Chapter 11

APPENDIX

Some results from the theory of Hilbert spaces and nonlinear, in particular, convex analysis are presented here, predominantly without proofs. Some of them hold true in more general spaces, but we will not focus on this.

The following references may be useful for a more detailed study of the basic facts: as in functional analysis – Kantorovich and Akilov [202], Edwards [99], Schwarz [370]; as in Sobolev spaces – Adams [3], Schwarz [369]; in finite element methods and variational inequalities – Ciarlet [75], Glowinski, Lions and Tremolieres [135]; in convex analysis – Ekeland and Temam [100], Rockafellar [348], Rockafellar and Wets [353], Ioffe and Tichomirov [192]; in optimization methods – Fletcher [117], Polak [322], Polyak [330], BAZARAA, SHERALY and SHETTY [36], Dennis and Schnabel [90], Golshtein and Tretyakov [141] and Mangasarian [285].

A1 Some Preliminary Results in Functional Analysis

A1.1 Norm and convergence in Hilbert spaces

Let $V$ be a real Hilbert space. A scalar product of two elements $u$ and $v$ in $V$ is denoted by $\langle u, v \rangle$ or $\langle u, v \rangle_V$ if it is convenient to emphasize the connection with a given space. A norm of an element $u \in V$ is denoted by $\|u\|$ or $\|u\|_V$. Unless stated otherwise, the norm is endowed by a scalar product, i.e., $\|u\| = \sqrt{\langle u, u \rangle}$.

Let us recall the Cauchy-Schwarz inequality

$$|\langle u, v \rangle| \leq \|u\| \cdot \|v\| \quad \forall u, v \in V.$$ 

A functional $|\cdot| : V \to \mathbb{R}$ is called a seminorm if it has all the properties of a norm except for $|u| > 0 \quad \forall u \neq 0$.

The set of linear continuous functionals on $V$ forms a conjugate space $V'$, in which the norm of an element $\ell \in V'$ is defined by

$$\|\ell\|_{V'} = \sup_{u \neq 0} \frac{\|\ell(u)\|_V}{\|u\|_V}.$$ 

For the value of a linear functional $\ell$ at $u$ we shall also use the notation $\langle \ell, u \rangle$, in particular, if the mapping $\ell : u \mapsto \langle \ell, u \rangle$ is understood as a duality pairing.
between $V$ and $V'$.

**A1.1.1 Theorem. (Riesz theorem)**

For any functional $\ell \in V'$ there exists a unique element $v \in V$ such that

$$
\ell(u) = \langle v, u \rangle \quad \forall \ u \in V.
$$

Due to the Riesz theorem, the space $V'$ can be identified with $V$ and it is easily seen that a Hilbert space $V$ is reflexive, i.e., $V'' = V$.

Additionally to the usual (strong) convergence of a sequence $\{v_n\}$ to $v$, denoted by $\|v_n - v\| \to 0$, or $v_n \to v$, we deal also with weakly convergent sequences.

**A1.1.2 Definition.**

A sequence $\{v_n\} \subset V$ is called weakly convergent to $v \in V$ (denoted by $v_n \rightharpoonup v$), if $\lim_{n \to \infty} \ell(v_n) = \ell(v)$ for each $\ell \in V'$. The point $v$ is said to be a weak limit of $\{v_n\}$. ♦

Uniqueness of a weak limit for such sequences can be shown easily. This result is a corollary of Lemma 1 in Opial [312].

**A1.1.3 Proposition. (Opial’s lemma)**

Assume that a subset $A$ of $V$ and a sequence $\{u_n\} \subset V$ are given such that

(i) $\|u_n - u\|$ converges for each $u \in A$;

(ii) if $u_n \rightharpoonup v$, then $v \in A$.

Then the sequence $\{u_n\}$ converges weakly.

**Proof:** In view of (i) the sequence $\{u_n\}$ is bounded. Now, let $u_{n_k} \to v$, $u_{n_j} \to v'$, $v' \neq v$. Due to (ii), the inclusions $v \in A$ and $v' \in A$ hold. From

$$
\|u_{n_k} - v'\|^2 = \|u_{n_k} - v\|^2 + \|v' - v\|^2 + 2\langle u_{n_k} - v, v - v' \rangle
$$

and

$$
\|u_{n_j} - v\|^2 = \|u_{n_j} - v'\|^2 + \|v' - v\|^2 + 2\langle u_{n_j} - v', v' - v \rangle,
$$

we obtain

$$
\lim_{k \to \infty} \|u_{n_k} - v'\|^2 = \lim_{k \to \infty} \|u_{n_k} - v\|^2 + \|v' - v\|^2,
$$

and

$$
\lim_{j \to \infty} \|u_{n_j} - v\|^2 = \lim_{j \to \infty} \|u_{n_j} - v'\|^2 + \|v' - v\|^2,
$$

which contradicts assumption (i).

Obviously, if $\{v_n\}$ converges to $v$, then $v$ is also a weak limit of $\{v_n\}$.

If $v_n \to v$ and $\|v_n\| \to \|v\|$, then strong convergence $v_n \to v$ holds true.

In finite-dimensional spaces weak convergence ensures convergence.

**A1.1.4 Definition.** A set $B$ is called the closure (resp. weak closure) of a set $A$ if each point of $B$ is a limit (resp. weak limit) of some sequence in $A$.

Correspondingly, a set $A$ is called closed (resp. weakly closed) if it coincides with its closure (resp. weak closure). ♦
The closure of a set $A$ is denoted by $\overline{A}$ or $\text{cl}A$ and a set $A$ is called \textit{dense} in $B$ if $A \subset B$ and $\text{cl}A \supseteq B$.

A subset $A$ of $V$ is said to be \textit{compact} (resp. \textit{weakly compact}) providing that each infinite sequence $\{u_n\} \subset A$ has a subsequence which converges (resp. weakly converges) to some element of $A$.

It is known that a bounded, weakly compact subset of a reflexive space, hence of a Hilbert space, is weakly compact.

### A1.2 Functionals in Hilbert spaces

Later on real-valued functionals are considered which can attain the value $+\infty$.

The set $\text{dom}f = \{u \in V : f(u) < \infty\}$ is called \textit{effective domain} of a functional $f$ and we always suppose that $\text{dom} f \neq \emptyset$.

**A1.2.5 Definition.** A functional $f : V \to \mathbb{R} := \mathbb{R} \cup \{+\infty\}$ is called \textit{lower semicontinuous} (resp. \textit{weakly lower semicontinuous}) if for each $u \in V$ and $u_n \to u$ (resp. $u_n \rightharpoonup u$)

$$\lim_{n \to \infty} f(u_n) \geq f(u).$$

Accordingly, $f$ is called \textit{upper semicontinuous} (resp. \textit{weakly upper semicontinuous}) if for each $u \in V$ and $u_n \to u$ (resp. $u_n \rightharpoonup u$)

$$\liminf_{n \to \infty} f(u_n) \leq f(u).$$

Sometimes we use the abbreviation lsc functional, wlsc functional, etc..

A functional $f$ is \textit{continuous} (resp. \textit{weakly continuous}) if it is lsc and usc (resp. wlsc and wusc) simultaneously.

**A1.2.6 Theorem. (Generalized Weierstrass theorem)**

A weakly lower semicontinuous functional attains its infimum on a weakly compact subset of a Hilbert space $V$.

We recall that a functional $f$ is \textit{Hölder continuous} on a set $G \subset V$ if

$$|f(u) - f(v)| \leq L \|u - v\|^\lambda$$

for some $L, \lambda > 0$ and all $u, v \in G$.

If $\lambda = 1$ then $f$ is called \textit{Lipschitz continuous} (Lipschitz for short) (with constant $L$).

We will consider linear (continuous) operators from a Hilbert space $V$ into another normed space $W$. The \textit{norm} of such an operator $T$ is defined by

$$\|T\| = \sup_{u \neq 0} \frac{\|Tu\|_W}{\|u\|_V}.$$ 

A mapping $a : V \times V \to \mathbb{R}$ is a \textit{bilinear form} if both

$$a(u, \cdot) : V \to \mathbb{R} \quad \text{and} \quad a(\cdot, u) : V \to \mathbb{R}$$

are linear functionals for each $u \in V$.

In the sequel \textit{symmetric and continuous} bilinear forms are considered only, i.e.,

$$a(u, v) = a(v, u) \quad \forall \ u, v \in V$$
and the existence of a constant $M$ is supposed such that

$$|a(u, v)| \leq M ||u|| ||v|| \quad \forall \ u, v \in V.$$  \hfill (A1.2.1)

Since for each $u \in V$ the mapping $v \mapsto a(u, v)$ is a linear functional, there exists a unique element $\Lambda u \in V'$ such that

$$a(u, v) = (\Lambda u)(v) \quad \forall \ v \in V$$  \hfill (A1.2.2)

and

$$||\Lambda u||_{V'} \leq M ||u||_V.$$

Hence, $\Lambda : u \mapsto \Lambda u$ is a linear operator from $V$ into $V'$ and $||\Lambda|| \leq M$.

Of course, according to the Riesz theorem for each $u$ the mapping $v \mapsto a(u, v)$ can be identified with a unique element $w \in V$ and we can understand $\Lambda$ as a linear operator from $V$ to $V'$ defined by $\Lambda u = w$. However, this identification is not convenient if the spaces $V$ and $V'$ are considered in the chain of inclusions, so-called Gelfand triple

$$V \subset H \subseteq H' \subset V',$$

where $H$ is another Hilbert space, especially if $H$ is identified with its conjugate $H'$. In the latter case, if $V$ is a dense set in $H$ and the embedding $V$ into $H$ is continuous\(^1\), then $V$ and $H$ can be identified with dense subspaces of $V'$.

### A1.3 Some Hilbert spaces and Sobolev spaces

Let $\Omega \subset \mathbb{R}^n$ be an open set with a boundary $\Gamma$. In order to classify the boundary $\Gamma$, following Adams [3], Nečas [306], we suppose that there are positive constants $\beta$ and $\delta$, a finite number of local coordinate systems and local mappings $a_r$, $1 \leq r \leq R$, such that

$$\Gamma := \bigcup_{r=1}^{R} \{(x^r, \hat{x}^r) : x^r_1 = a_r(\hat{x}^r), \ |\hat{x}^r| < \delta \},$$

$$\{(x^r_1, \hat{x}^r) : a_r(\hat{x}^r) < x^r_1 < a_r(\hat{x}^r) + \beta, \ |\hat{x}^r| < \delta \} \subset \Omega, \quad 1 \leq r \leq R,$$

$$\{(x^r_1, \hat{x}^r) : a_r(\hat{x}^r) - \beta < x^r_1 < a_r(\hat{x}^r), \ |\hat{x}^r| < \delta \} \cap \hat{\Omega} = \emptyset, \quad 1 \leq r \leq R,$$

with $\hat{x}^r = (x^r_2, ..., x^r_n)$ and $|\hat{x}^r| < \delta$ $\Leftrightarrow$ $|x^r_i| < \delta$, $2 \leq i \leq n$.

If the functions $a_r$ are Lipschitz (or belong to $C^m$) in their definition domains $\{\hat{x}^r : |\hat{x}^r| \leq \delta \}$, then the boundary is called Lipschitz-continuous (Lipschitz for short) (or boundary of class $C^m$).

For a function $\varphi : \Omega \rightarrow \mathbb{R}$ the closure of the set $\{x \in \Omega : \varphi(x) \neq 0\}$ is called support of $\varphi$.

Let $\mathcal{D}(\Omega)$ be a space of infinitely often differentiable functions with compact supports on $\Omega$. For a multi-index $\alpha$ with $|\alpha| := \alpha_1 + ... + \alpha_n$ ($\alpha_i$ non-negative integers) let a differential operator

$$D^\alpha := \frac{\partial^{|\alpha|}}{\partial x^\alpha_1 ... \partial x^\alpha_n}$$

---

\(^1\)Notion of a continuous embedding cf. Definition A1.3.7
be defined on the space $\mathcal{D}(\Omega)$. If for functions $f, g \in L^2(\Omega)$ and arbitrary $\varphi \in \mathcal{D}(\Omega)$ the relation
\[
\int_{\Omega} g \varphi d\Omega = (-1)^{|\alpha|} \int_{\Omega} f^{D^\alpha} \varphi d\Omega
\]
is fulfilled, then $g$ is called a generalized derivative of the function $f$
\[g := D^\alpha f.\]

A set of functions $u \in L^2(\Omega)$ with $D^\alpha u \in L^2(\Omega)$ for all $|\alpha| \leq m$ ($m$ integer) forms a Sobolev space $H^m(\Omega)$. It turns into a Hilbert space if the scalar product is defined by
\[
\langle u, v \rangle_{m, \Omega} = \sum_{0 \leq |\alpha| \leq m} \langle D^\alpha u, D^\alpha v \rangle_{L^2(\Omega)}. \tag{A1.3.3}
\]
We consider the space $H^m(\Omega)$ to be a Hilbert space with this scalar product and the corresponding norm
\[
\|u\|_{m, \Omega} = \left( \sum_{0 \leq |\alpha| \leq m} \|D^\alpha u\|_{L^2(\Omega)}^2 \right)^{1/2}. \tag{A1.3.4}
\]
An analogous notation is applied for $H^0(\Omega) = L^2(\Omega)$, too.

If the smoothness of the boundary $\Gamma$ is not too bad, for example if $\Gamma$ is Lipschitz as we always suppose, then $H^m(\Omega)$ coincides with the closure of the set $C^\infty(\bar{\Omega})$ in the norm (A1.3.4).

We also need the Sobolev space $H^m_0(\Omega)$, which is obtained as the closure of the set $\mathcal{D}(\Omega)$ in the norm (A1.3.4). The scalar product in $H^m_0(\Omega)$ is introduced by (A1.3.3) or
\[
\langle u, v \rangle_{m, \Omega, 0} = \sum_{|\alpha| = m} \langle D^\alpha u, D^\alpha v \rangle_{L^2(\Omega)}, \tag{A1.3.5}
\]
moreover, due to Friedrichs’ inequality
\[
c \|u\|_{0, \Omega} \leq \sum_{|\alpha| = 1} \|D^\alpha u\|_{0, \Omega} \quad \forall u \in H^1_0(\Omega) \tag{A1.3.6}
\]
(with $c > 0$ independent of $u$), the norms (A1.3.4) and
\[
\|u\|_{m, \Omega, 0} = \left( \sum_{|\alpha| = m} \|D^\alpha u\|_{0, \Omega}^2 \right)^{1/2}. \tag{A1.3.7}
\]
are equivalent.

The definitions observed can easily be adapted to vector functions
\[u = (u_1, \ldots, u_n) : u \in [H^m(\Omega)]^n \text{ if } u_i \in H^m(\Omega), \ i = 1, \ldots, n,
\]
and
\[
\|u\|_{m, \Omega} = \left( \sum_{i=1}^n \|u_i\|_{m, \Omega}^2 \right)^{1/2}.
\]
Similarly, the spaces $[H^m_0(\Omega)]^n$ and norms $\|u\|_{m, \Omega, 0}$ are defined if $u_i \in H^m_0(\Omega)$.

Now, we describe some facts of the embedding theory for Sobolev spaces.
A normed space $X$ is called \textit{continuously embedded} into a normed space $Y$ if $X \subset Y$ holds in the algebraic sense and there exists a fixed constant $c > 0$ such that

$$
\|x\|_Y \leq c\|x\|_X \quad \forall x \in X.
$$

Obviously, an operator $T$, which performs this embedding, is linear (continuous).

An embedding is said to be \textit{compact} if an operator $T$ maps each bounded subset $A$ of $X$ into a relatively compact subset $TA$ of $Y$, i.e. $\text{cl}(TA)$ is compact in $Y$.

The notation $X \hookrightarrow Y$ is used for a continuous embedding, whereas $X \overset{c}{\hookrightarrow} Y$ stands for a compact one.

A set of functions in $C^m(\overline{\Omega})$ for which the derivatives of order $m$ satisfy the H"{o}lder condition with constant $\lambda$ and for which the norm is defined by

$$
\|u\|_{C^m,\lambda}(\overline{\Omega}) = \|u\|_{C^m(\overline{\Omega})} + \max_{|\alpha|=m} \sup_{x,y \in \Omega, x \neq y} \frac{|D^\alpha u(x) - D^\alpha u(y)|}{\|x-y\|_\Lambda^{n/2}}
$$

build the Banach space $C^{m,\lambda}(\overline{\Omega})$.

According to the embedding theorems, in particular, the following results hold true with an arbitrary integer $k \geq 0$:

- $H^{m+k}(\Omega) \hookrightarrow C^k,\lambda(\overline{\Omega})$ if $0 < \lambda \leq m - \frac{n}{2}$ and $2(m-1) < n < 2m$;
- $H^{m+k}(\Omega) \hookrightarrow C^k,\lambda(\overline{\Omega})$ if $0 < \lambda < 1$ and $n = 2m - 2$;
- $H^{m+k}(\Omega) \hookrightarrow C^{k,1}(\Omega)$ if $n < 2m - 2$.

The trivial continuous embedding $H^{m+k}(\Omega) \hookrightarrow H^k(\Omega)$ is compact except for the case $m = 0$. Moreover, the relations

- $H^{m+k}(\Omega) \overset{c}{\hookrightarrow} C^k(\overline{\Omega})$ if $n < 2m$;
- $H^{m+k}(\Omega) \overset{c}{\hookrightarrow} C^{k,\lambda}(\overline{\Omega})$ if $2(m-1) \leq n < 2m$ and $0 < \lambda < m - \frac{n}{2}$

are satisfied.

All these embeddings remain true for the corresponding spaces $H^m_0(\Omega)$ with the norms (A1.3.4) or (A1.3.7) without any assumption about smoothness of the boundary $\Gamma$.

We briefly deal with the \textit{trace} of a function of the class $H^m(\Omega)$. As the boundary $\Gamma$ is Lipschitz, a surface measure $d_\gamma$ can be defined on this boundary, and one can consider the space $L_2(\Gamma)$ with the norm $\|\cdot\|_{L_2(\Gamma)}$. For functions in $C^\infty(\overline{\Omega})$ the inequality

$$
\|u\|_{L_2(\Gamma)} \leq c\|u\|_{1,\Omega}
$$

is known, where $c$ depends only on $\Omega$. Taking into account that $H^1(\Omega)$ is the closure of $C^\infty(\Omega)$ in the norm $\|\cdot\|_{1,\Omega}$, there exists a linear (continuous) mapping $\gamma : u \in H^1(\Omega) \mapsto \gamma u \in L_2(\Gamma)$ such that $\gamma u = u|_\Gamma$ for $u \in C^\infty(\Omega)$. The factor space

$$
\{w \in L_2(\Gamma) : \exists v \in H^1(\Omega), \gamma v = w\}
$$

is a Hilbert space $H^\frac{1}{2}(\Gamma)$ with the norm

$$
\|w\|_{H^\frac{1}{2}(\Gamma)} = \inf\{\|v\|_{1,\Omega} : v \in H^1(\Omega), \gamma v = w \text{ on } \Gamma\}.
$$
A1. SOME PRELIMINARY RESULTS IN FUNCTIONAL ANALYSIS

We note that

$$H^1_0(\Omega) = \{ v \in H^1(\Omega) : \gamma u = 0 \}.$$ 

In view of the properties of $\Gamma$ an outward normal $\nu = (\nu_1, ..., \nu_n)$ exists almost everywhere on $\Gamma$, and assuming that $\|\nu\|_{\mathbb{R}^n} = 1$, the normal derivative operator

$$\frac{\partial}{\partial \nu} = \sum_{i=1}^{n} \nu_i \frac{\partial}{\partial x_i}$$

can be extended on all functions in $H^2(\Omega)$. In particular, we get

$$H^2_0(\Omega) = \{ u \in H^2(\Omega) : \gamma u = 0, \sum_{i=1}^{n} \nu_i \gamma \frac{\partial}{\partial x_i} = 0 \text{ on } \Gamma \}$$

with $\frac{\partial u}{\partial x_i}$ a generalized derivative.

Sometimes we also write $u|_{\Gamma}$ instead of $\gamma u$ for a function $u \in H^1(\Omega)$.

A1.4 Differentiation of operators and functionals

The facts on the differential calculus in functional spaces presented here will be replenished in Section A1.5, where convex functionals are considered.

Assume that $X$ and $Y$ are Banach spaces, that $U(x) \subset X$ is a neighborhood of an arbitrary point $x \in X$ and that $F : U(x) \to Y$ is a given mapping. The symbol $\mathcal{L}(X,Y)$ indicates the space of all linear (continuous) operators mapping from $X$ into $Y$.

A1.4.8 Definition. An operator $T(x) \in \mathcal{L}(X,Y)$ is called a Gâteaux-derivative of a Mapping $F$ at a point $x$ if for each $v \in U(x)$

$$\lim_{t \downarrow 0} \frac{\| F(x + tv) - F(x) - T(x)(v) \|_Y}{t} = 0. \tag{A1.4.8}$$

If the relation

$$\lim_{\|v\|_X \to 0} \frac{\|F(x + v) - F(x) - T(x)(v)\|_Y}{\|v\|_X} = 0 \tag{A1.4.9}$$

is fulfilled, then $T(x)$ is said to be a Fréchet-derivative of $F$ at $x$.

Accordingly, an operator $F$ is called Gâteaux- or Fréchet-differentiable at $x$, and in both cases we write $T(x) = F'(x)$. \hfill \Box$

If $F$ has a Gâteaux- or Fréchet-derivative at every point of the set $U(x)$, then $F'$ can be considered as a mapping from $U(x)$ into $\mathcal{L}(X,Y)$. If the derivative $F'$ exists in the mentioned sense, then $F''(x)$ is an element of the space $\mathcal{L}(X, \mathcal{L}(X,Y))$.

In the case $X = V$ and $F := J : V \to \mathbb{R}$ an adaption of the definitions of Gâteaux- and Fréchet-derivatives is trivial: Instead of the norms $\| \cdot \|_Y, \| \cdot \|_X$ in (A1.4.8) and (A1.4.9) we write $| \cdot |, \| \cdot \|_V$, respectively, and in that case

$$T(x) = J'(x) \in V', \quad J''(x) \in \mathcal{L}(V,V').$$
Sometimes the Fréchet-derivative $J'(x)$ is denoted by $\nabla J(x)$.

If $V = \mathbb{R}^n$, using the Riesz theorem A1.1.1, we obtain

$$\nabla J(x) = \left( \frac{\partial J}{\partial x_1}(x), \ldots, \frac{\partial J}{\partial x_n}(x) \right)$$

and instead of "Fréchet-differentiable" and "Fréchet-derivative" the notions differentiable and derivative, respectively, are used.

Relations between Gâteaux- and Fréchet-derivatives are described in the following statement.

A1.4.9 Proposition.

(i) If an operator $F$ has a Fréchet-derivative at a point $x$, then $F$ is continuous and Gâteaux-differentiable at $x$ and both derivatives coincide.

(ii) If a Gâteaux-derivative of $F$ is continuous in some neighborhood of a point $x$, then the operator $F$ has a Fréchet-derivative at $x$.

Usually, the notations $Fx$ and $F(x)$ are used for the values of linear and nonlinear operators at the point $x$, respectively.

If an operator $F$ is Fréchet-differentiable on an interval $[u, u + v]$, then an integral representation

$$F(u + v) = F(u) + \int_0^1 F'(u + \tau v)(v)d\tau \quad \text{(A1.4.10)}$$

holds true.

Concerning a Taylor formula for operators, in particular, we refer to VAINBERG[405]. For a Fréchet-differentiable functional $J : V \rightarrow \mathbb{R}$ the mean value theorem

$$J(u + v) = J(u) + \langle J'(u + \tau v), v \rangle \quad \text{for some } \tau \in [0, 1]$$

holds, and if $J$ is twice Fréchet-differentiable on $[u, u + v]$, then the relations

$$J'(u + v) - J'(u) = J''(u)(v) + o(\|v\|)$$

$$J(u + v) = J(u) + \langle J'(u), v \rangle + \frac{1}{2} \langle J''(u + \theta v)(v), v \rangle \quad \text{for some } \theta \in [0, 1]$$

are true.

If $a(x, y)$ is a symmetric, continuous bilinear form, using the presentation $a(u, v) = (\Lambda u)(v)$ (see (A1.2.2)) and

$$a(u + \tau v, u + \tau v) - a(u, u) = 2\tau a(u, v) + \tau^2 a(v, v),$$

we obtain $J'(u) = 2\Lambda u$ and $J''(u) = 2\Lambda$.

**A1.5 Convexity of sets and functionals**

We assume that the definitions of affine and convex sets, affine and convex hulls and convex cones as well as some of their simple properties are well-known. For the reader which is not familiar with these notions we refer to ROCKAFELLAR and WETS [353] and HIRIART-URRUTY and LEMARÉCHAL [179]. Concerning the convexity of particular sets, usually we shall not verify this fact in simple cases.

Now, we consider sets in a Hilbert space $V$. 
A1. SOME PRELIMINARY RESULTS IN FUNCTIONAL ANALYSIS

A1.5.10 Definition. The set of all interior points of a convex set \( C \) is called the **interior** of \( C \) and denoted by \( \text{int} C \). A point \( x \) of a convex set \( C \) is called **relatively interior** if

\[
U(x) \cap \text{aff} C \subset C,
\]

with \( \text{aff} C \) the affine hull of \( C \) and \( U(x) \) some neighborhood of the point \( x \). The set of all relatively interior points of a set \( C \) is denoted by \( \text{ri} C \) and forms the **relative interior** of \( C \). ♦

A1.5.11 Proposition. For a convex set \( C \) the sets \( \text{int} C \) and \( \text{ri} C \) are also convex and

\[
\text{cl int} C = \text{cl} C.
\]

If \( u \in \text{int} C, v \in \text{cl} C, \) then \( [u, v) \subset \text{int} C \), with

\[
[u, v) := \{ u + \lambda(v - u) : \lambda \in [0, 1) \}.
\]

For a convex set its closure and weak closure coincide, hence, a closed convex set is also weakly closed.

Given a functional \( J : M \subset V \to \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \), the set

\[
\text{epi} J = \{ (\alpha, u) \in \mathbb{R} \times M : J(u) \leq \alpha \}
\]

is said to be the **epigraph** of \( J \).

A functional \( J : V \to \bar{\mathbb{R}} \) is lower semi-continuous iff \( \text{epi} J \) is a closed set.

A1.5.12 Definition. A functional \( J : M \to \bar{\mathbb{R}} \) is called **convex** on a convex set \( C \) if the set \( \text{epi} J \) is convex.

Usually, for a convex functional \( J : V \to \bar{\mathbb{R}} \), with

\[
\text{dom} J := \{ u \in V : J(u) < \infty \} \neq \emptyset,
\]

the notion of a **proper convex** functional is used. We omit ”proper”, because only such convex functionals will be considered in this book.

Obviously, \( J : C \to \bar{\mathbb{R}} \) is convex on \( C \) iff the functional

\[
\hat{J}(u) := \begin{cases} J(u) & \text{if } u \in C \\ +\infty & \text{if } u \in V \setminus C \end{cases}
\]

is convex on \( V \). In this case \( \text{dom} J = \text{dom} \hat{J} \) holds.

We can immediately conclude that the epigraph of a convex, lower semi-continuous functional \( J : V \to \bar{\mathbb{R}} \) is a convex, closed set, hence also weakly closed. Moreover, \( J \) is a weakly lower semicontinuous functional.

Often, the following property is considered as an equivalent definition for the convexity of the function \( J \) on a convex set \( C \): For arbitrary two points of a convex set \( C \subset V \) and a scalar \( \lambda \in [0, 1] \) the inequality

\[
J(\lambda u + (1 - \lambda)v) \leq \lambda J(u) + (1 - \lambda)J(v)
\]

holds. (We assume that the inequality \( f(x) \leq g(y) \) is always fulfilled if \( g(y) = +\infty \).

A functional \( J \) is called **concave** if \( -J \) is convex.
Convexity of a mapping $F$, acting from a convex set $C \subset V$ into a vector space $Y$, is defined analogously: For arbitrary $u, v \in C$ and $\lambda \in [0, 1]$

$$F(\lambda u + (1 - \lambda)v) - \lambda F(u) - (1 - \lambda)F(v) \leq 0$$

holds. Here the inequality $a \leq 0$ on $Y$ means that $-a \in Y_+$, with $Y_+$ a non-negative cone in $Y$.

It is easy to verify that a non-negative linear combination of a finite number of convex functionals and the supremum of any number of convex functionals are also convex functionals.

Now, we formulate some properties of convex functionals.

**A1.5.13 Proposition.** Let $J : V \to \bar{\mathbb{R}}$ be a convex functional.

(i) If $(\alpha, u) \in \text{epi}J$, then $u \in \text{dom}J$, hence, $\text{dom}J$ is a convex set.

(ii) For arbitrary points $u^1, \ldots, u^n$ from $V$ and scalars $\lambda_i \geq 0$ ($i = 1, \ldots, n$), with $\sum_{i=1}^{n} \lambda_i = 1$, Jensen’s inequality holds true:

$$J\left(\sum_{i=1}^{n} \lambda_i u^i\right) \leq \sum_{i=1}^{n} \lambda_i J(u^i).$$

(iii) For any constant $c$ the level sets

$$\{u \in V : J(u) \leq c\}, \quad \{u \in V : J(u) < c\}$$

are convex, and if additionally $J$ is lower semicontinuous, then the level set

$$\{u \in V : J(u) \leq c\}$$

is closed.

(iv) The function $\eta(\lambda) = J(\lambda u + (1 - \lambda)v)$ is convex on $\mathbb{R}$ for arbitrary $u, v \in V$ and $\lambda \in [0, 1]$.

(v) If $\Lambda \in \mathcal{L}(Y, V)$, with $Y$ a Banach space, then $\psi(y) = J(\Lambda y)$ is a convex functional on $Y$.

**A1.5.14 Proposition.** For convex functionals $J : V \to \bar{\mathbb{R}}$ the following statements are equivalent:

(i) $J$ is bounded from above in a neighborhood of some point $u \in V$;

(ii) $J$ is continuous at $u \in V$;

(iii) $\text{int}(\text{epi}J) \neq \emptyset$;

(iv) $\text{int}(\text{dom}J) \neq \emptyset$ and $J$ is continuous on $\text{int}(\text{dom}J)$.

This proposition implies a number of non-trivial corollaries for convex functionals $J : V \to \bar{\mathbb{R}}$.

**A1.5.15 Corollary.** In a finite-dimensional space a convex functional $J$ is continuous on $\text{int}(\text{dom}J)$. 
A1.5.16 Corollary. If \( J \) is convex and bounded from above on an open, convex set \( C \), then \( J \) is locally Lipschitz-continuous on \( \text{int}(\text{dom} J) \), i.e., Lipschitz-continuous in a neighborhood of each point of \( \text{int}(\text{dom} J) \).

A1.5.17 Corollary. If \( J \) is a convex, lower semicontinuous functional, then it is continuous on \( \text{int}(\text{dom} J) \).

For convex finite-valued functionals \( J : C \to \mathbb{R} \) the straightforward transitions of the properties formulated above is obvious.

We remind also on the superposition of convex functionals.

A1.5.18 Proposition. Let \( M \subset V \) be a convex set. If \( J : M \to \mathbb{C} \subset \mathbb{R} \) and \( \varphi : C \to \mathbb{R} \) are convex functionals and \( \varphi \) does not decrease on \( C \), then \( \varphi : u \mapsto \varphi(J(u)) \) is a convex functional on \( M \).

The following statement uses essentially the finite dimensionality of the space.

A1.5.19 Proposition. Let \( C \subset \mathbb{R}^n \) be a relatively open convex set, and \( \hat{C} \) is supposed to be dense in \( C \). If a sequence of finite-valued convex functions \( f_i : C \to \mathbb{R} \) converges in each point of \( \hat{C} \), then there exists a finite-valued convex function \( f \) such that
\[
\lim_{i \to \infty} f_i(x) = f(x) \quad \forall x \in C,
\]
moreover, the sequence \( \{f_i\} \) converges uniformly to \( f \) on each compact subset of \( C \).

Note that if \( C \subset \mathbb{R}^n \) is open and convex and \( \varphi : C \to \mathbb{R} \) is convex and differentiable on \( C \), then \( \varphi \) is continuously differentiable on \( C \).

In this book we deal with particular cases of the following convex variational problem:

\[
\begin{align*}
\text{minimize} & \quad J(u), \\
\text{subject to} & \quad u \in K,
\end{align*}
\]

where \( J : V \to \mathbb{R} \) is a convex, lower semicontinuous functional, \( K \subset V \) is a convex, closed set and \( K \cap \text{dom} J \neq \emptyset \).

The set \( K \) is called feasible set and
\[
U^* : \{u \in K : J(u) = \inf_{v \in K} J(v) \equiv J^* \}
\]
is said to be solution set or optimal set of the problem.

The symbols \( \text{Arg min}_{u \in G} f(u) \) and \( \text{arg min}_{u \in G} f(u) \) denote the optimal set or optimal point of the problem \( \min\{f(u) : u \in G\} \), respectively.

It is easy to verify that for convex variational problems the optimal set \( U^* \) is closed and convex, however, \( U^* = \emptyset \) is not excluded. Moreover, any local minimizer of \( J \) on \( K \) is at the same time a (global) solution of Problem (A1.5.11).

Obviously, Problem (A1.5.11) can be reformulated as
\[
\min\{J(u) : u \in K \cap \text{dom} J\}.
\]
and also as
\[ \min\{ J|_{\text{dom}J}(u) : u \in K \cap \text{dom}J\}, \quad (A1.5.12) \]
with \( J|_D : D \rightarrow \bar{\mathbb{R}} \), \( J|_D(u) = J(u) \forall u \in D \).

If \( K \cap \text{dom}J \) is bounded, then due to the generalized Weierstrass Theorem A1.2.6, \( U^* \neq \emptyset \), i.e. Problem (A1.5.11) is solvable. Otherwise, in order to secure solvability of Problem (A1.5.11), we need some additional assumptions, for instance coercivity of \( J \) on \( K \), i.e.,
\[ \lim_{\|u\| \rightarrow \infty} J(u) = +\infty. \]

Unique solvability of Problem (A1.5.11) can be established by means of stronger convexity conditions for the functional \( J \). Usually, such conditions are formulated for finite-valued functionals on a convex subset of the space \( V \). Taking into account the equivalence between the Problems (A1.5.11) and (A1.5.12), these conditions can be applied quite simply.

**A1.5.20 Definition.** A functional \( J : C \rightarrow \mathbb{R} \) on a convex set \( C \subset V \) is said to be

(i) **strictly convex** if for any \( u,v \in C \), \( u \neq v \) and \( 0 < \lambda < 1 \)
\[ J(\lambda u + (1 - \lambda)v) < \lambda J(u) + (1 - \lambda)J(v); \]

(ii) **uniformly convex** if for any \( u,v \in C \) and \( 0 \leq \lambda \leq 1 \)
\[ J(\lambda u + (1 - \lambda)v) \leq \lambda J(u) + (1 - \lambda)J(v) - \lambda(1 - \lambda)\delta(\|u - v\|), \]
with \( \delta : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) a strictly increasing function, \( \delta(0) = 0; \)

(iii) **strongly convex** if \( J \) is uniformly convex with \( \delta(t) = \kappa t^2 \), \( \kappa > 0 \).

\[ \diamond \]

Constant \( \kappa \) is called constant of strong convexity. Obviously, a uniformly or strongly convex functional is strictly convex.

**A1.5.21 Proposition.**

(i) If \( J \) is a strictly convex functional on \( \text{dom}J \), then Problem (A1.5.11) has at most one solution.

(ii) If \( J \) is uniformly convex and lower semicontinuous on \( \text{dom}J \), then there exists a unique solution of Problem (A1.5.11).

For the investigation of variational inequalities the coercivity of bilinear forms is essential, i.e., the bilinear form \( a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R} \) is said to be **coercive** if there exists a constant \( \alpha > 0 \) such that
\[ a(u, u) \geq \alpha \|u\|^2 \quad \forall u \in V. \]
A1.5.22 Corollary. Let \( a(\cdot, \cdot) \) be a continuous, symmetric and coercive bilinear form on \( V \times V \) and \( f \in V' \) be a given element. Then Problem (A1.5.11) with the quadratic functional

\[
J(u) = \frac{1}{2} a(u,u) - \langle f, u \rangle
\]

has a unique solution.

For non-smooth convex problems we need the technique of directional derivatives and subdifferentials.

A1.5.23 Definition. Let \( J : V \to \overline{\mathbb{R}} \) be an arbitrary functional. If for fixed \( u \in \text{dom} J \) and \( d \neq 0 \) the limit

\[
\lim_{t \downarrow 0} \frac{J(u + td) - J(u)}{t} =: J'(u; d)
\]

exists (\( J'(u; d) = \pm \infty \) is possible), then \( J'(u; d) \) is called a directional derivative of \( J \) at the point \( u \) in the direction \( d \).

A1.5.24 Proposition. A convex functional \( J : V \to \overline{\mathbb{R}} \) has a directional derivative at each point \( u \in \text{dom} J \) in any direction.

If \( u, v \in \text{ri}(\text{dom} J) \), \( v \neq u \), then the directional derivative of \( J \) at \( u \) in direction \( v - u \) is finite.

For two points \( u, u + v \in \text{dom} J \), due to the convexity of \( J \), the function

\[
\varphi(t) = \frac{J(u + tv) - J(u)}{t}
\]

increases on \((0, 1]\). Hence,

\[
J(u + v) - J(u) \geq J'(u; v) \quad \forall u \in \text{dom} J, \ v \in V.
\]

(A1.5.13)

A1.5.25 Definition. Let \( J : V \to \overline{\mathbb{R}} \) be a convex functional. An element \( q \in V' \) is called a subgradient of \( J \) at the point \( u \) if

\[
J(v) - J(u) \geq \langle q, v - u \rangle \quad \forall \ v \in V.
\]

The set of all subgradients of \( J \) at the point \( u \) is called the subdifferential of \( J \) at the point \( u \) and denoted by \( \partial J(u) \), i.e.,

\[
\partial J(u) = \{ q \in V' : J(v) - J(u) \geq \langle q, v - u \rangle \quad \forall \ v \in V \}.
\]

The subdifferential \( \partial J(u) \) (possibly empty) is a convex and weakly closed subset in \( V' \). If \( \partial J(u) \neq \emptyset \), we say that \( J \) is subdifferentiable at the point \( u \). Note that the relation

\[
J'(u; v) \geq \langle q, v \rangle
\]

(A1.5.14)

holds always true.

In particular, the following connection between directional derivatives and sub-differentials are true.
A1.5.26 Proposition. Assume that the convex functional $J : V \to \bar{\mathbb{R}}$ is continuous at $u \in \text{dom} J$. Then

$$\partial J(u) \neq \emptyset \quad \forall \ w \in \text{int}(\text{dom} J)$$

and

$$\sup_{q \in \partial J(u)} \langle q, v \rangle = J'(u; v) < \infty. \quad (A1.5.15)$$

If $V = \mathbb{R}^n$, then $\partial J(v) \neq \emptyset \quad \forall \ v \in \text{ri}(\text{dom} J)$.

We remind of some rules of the subdifferential calculus which follow immediately from the definition of the subdifferential:

if $J_1(u) = \lambda J(u)$ with $\lambda > 0$, then $\partial J_1(u) = \lambda \partial J(u)$;

if $J_1(u) = J(\lambda u)$ with $\lambda > 0$, then $\partial J_1(u) = \lambda \partial J(\lambda u)$.

A1.5.27 Theorem. (Moreau-Rockafellar) Let $J_1, J_2 : V \to \bar{\mathbb{R}}$ be convex functionals. Then

$$\partial (J_1 + J_2)(u) \supset \partial J_1(u) + \partial J_2(u)$$

and if, moreover, $J_1$ is continuous at $\tilde{u} \in \text{dom} J_1 \cap \text{dom} J_2$, then

$$\partial (J_1 + J_2)(u) = \partial J_1(u) + \partial J_2(u).$$

For fixed $\Lambda \in \mathcal{L}(X, V)$, with $X$ a Banach space, and for convex functionals $J : V \to \bar{\mathbb{R}}$ the superposition $J \circ \Lambda$ is always a convex functionals on $X$.

A1.5.28 Proposition. If $J$ is continuous at some point $\tilde{u} = \Lambda x$, then

$$\partial (J \circ \Lambda)(x) = \Lambda' \partial J(\Lambda x) \quad \forall \ x \in X,$$

with $\Lambda' : V' \to X'$ the conjugate operator of $\Lambda$, i.e.,

$$\langle \Lambda' f, x \rangle = \langle f, \Lambda x \rangle \quad \forall \ f \in V', \ x \in X.$$

The following connection between subdifferentials and Gâteaux-derivatives can easily be verified.

A1.5.29 Proposition. If a convex functional $J : V \to \bar{\mathbb{R}}$ is Gâteaux-differentiable at $u \in V$, then $J$ is subdifferentiable at this point and

$$\partial J(u) = \{J'(u)\}.$$ 

On the other hand, if a functional $J$ is continuous at $u$ and has a unique subgradient, then $J$ is Gâteaux-differentiable at $u$.

Due to Definition A1.5.23 and the Propositions A1.5.26 and A1.5.29 it follows immediately that the functional $J(u + \lambda d)$ decreases for sufficiently small $\lambda > 0$ if $\langle J'(u), d \rangle < 0$.

A1.5.30 Proposition. If $C \subset V$ is a convex set and $J : C \to \bar{\mathbb{R}}$ is Gâteaux-differentiable on $C$, then convexity of $J$ on $C$ is equivalent to

$$J(v) \geq J(u) + \langle J'(u), v - u \rangle \quad \forall \ u, v \in C.$$ \quad (A1.5.16)
A1.5.31 Proposition. If $C \subset V$ is a convex set and the convex functional $J : C \to \mathbb{R}$ is subdifferentiable on $C$, then strict (resp. strong) convexity of $J$ guarantees that

$$J(v) - J(u) > (q, v - u) \quad \forall \ q \in \partial J(u), \forall \ u, v \in C, \ u \neq v,$$

(resp. $J(v) - J(u) \geq (q, v - u) + \kappa \|u - v\|^2$, $\kappa > 0$). (A1.5.17) (A1.5.18)

On the other hand, if inequality (A1.5.17) (resp. (A1.5.18)) is fulfilled for some $q \in \partial J(u)$ and each $u, v \in C$, $u \neq v$, then $J$ is strictly (resp. strongly) convex.

By means of the Propositions A1.5.29 and A1.5.30, Proposition A1.5.31 can easily be reformulated for Gâteaux-differentiable functionals.

A1.5.32 Corollary. Under the hypotheses of Proposition A1.5.30 or A1.5.31 for the functional $J$ the following necessary and sufficient conditions of convexity, strict and strong convexity hold, respectively:

$$\langle q(u) - q(v), u - v \rangle \geq 0 \quad \forall \ u, v \in C;$$

$$\langle q(u) - q(v), u - v \rangle > 0 \quad \forall \ u, v \in C, u \neq v;$$

$$\langle q(u) - q(v), u - v \rangle \geq 2\kappa \|u - v\|^2 \quad \forall \ u, v \in C;$$

with $q(u) \in \partial J(u)$ if $J$ is subdifferentiable and $q(u) = J'(u)$ if $J$ is Gâteaux-differentiable.

Inequality (A1.5.19) characterizes the subdifferentials and Gâteaux-derivatives of convex functionals as monotone mappings.

A1.5.33 Lemma. Let $J : V \to \mathbb{R}$ be a convex, Fréchet-differentiable functional on the Hilbert space $V$, and assume that its gradient satisfies a Lipschitz condition on $V$ with constant $L$. Then it holds

$$\langle \nabla J(u^1) - \nabla J(u^2), u^1 - u^2 \rangle \geq \frac{1}{L} \|\nabla J(u^1) - \nabla J(u^2)\|^2, \quad \forall \ u^1, u^2 \in V. \quad (A1.5.22)$$

Proof: Introducing the functional

$$\tilde{J}(u) := J(u) - J(u^2) - \langle \nabla J(u^2), u - u^2 \rangle,$$

it is obvious that this functional satisfies the assumptions of the lemma with the same constant $L$. Using formula (A1.4.10) and the Lipschitz property, we get for any $u \in V$

$$\tilde{J}(u) = \tilde{J}(u^1) + \langle \nabla \tilde{J}(u^1), u - u^1 \rangle$$

$$+ \int_0^1 \langle \nabla \tilde{J}(u^1 + t(u - u^1)) - \nabla \tilde{J}(u^1), u - u^1 \rangle dt$$

$$\leq \tilde{J}(u^1) + \langle \nabla \tilde{J}(u^1), u - u^1 \rangle + \frac{L}{2} \|u - u^1\|^2.$$

Inserting $u := u^1 - \frac{1}{L} \nabla \tilde{J}(u^1)$, we get

$$\tilde{J}(u^1 - \frac{1}{L} \nabla \tilde{J}(u^1)) \leq \tilde{J}(u^1) - \frac{1}{L} \|\nabla \tilde{J}(u^1)\|^2 + \frac{1}{2L} \|\nabla \tilde{J}(u^1)\|^2$$

$$= \tilde{J}(u^1) - \frac{1}{2L} \|\nabla \tilde{J}(u^1)\|^2.$$
Since \( u^2 \) is a minimizer of the functional \( \tilde{J} \) with \( \tilde{J}(u^2) = 0 \), we conclude that

\[
\tilde{J}(u^1 - \frac{1}{L} \nabla \tilde{J}(u^1)) \geq 0 \quad \text{and} \quad \tilde{J}(u^1) \geq \frac{1}{2L} \|\nabla \tilde{J}(u^1)\|^2.
\]

Using the definition of \( \tilde{J} \) this leads to

\[
J(u^1) - J(u^2) - \langle \nabla J(u^2), u^1 - u^2 \rangle \geq \frac{1}{2L} \|\nabla J(u^1) - \nabla J(u^2)\|^2.
\]

Exchanging the positions of \( u^1 \) and \( u^2 \) in the latter inequality, we also have

\[
J(u^2) - J(u^1) - \langle \nabla J(u^1), u^2 - u^1 \rangle \geq \frac{1}{2L} \|\nabla J(u^2) - \nabla J(u^1)\|^2
\]

and inequality (A1.5.22) is obtained by summing up the latter two inequalities. \( \square \)

Now we return to Problem (A1.5.11) and consider the set of feasible directions on \( K \) at the point \( u \), i.e.,

\[
C_u(K) = \{ v \neq 0 : u + \lambda v \in K \text{ for some } \lambda > 0 \}. \tag{A1.5.23}
\]

Due to Definition A1.5.23 and inequality (A1.5.13), the point \( u \in K \cap \text{dom} J \) is a solution of Problem (A1.5.11) if

\[
\tilde{J}'(u; v) \geq 0 \quad \forall v \in V,
\]

or

\[
J'(u; v) \geq 0 \quad \forall v \in C_u(K).
\]

Immediately from Definition A1.5.25 of a subdifferential we imply that \( u \in U^* \) iff \( 0 \in \partial \tilde{J}(u) \).

A1.5.34 Proposition. Let \( J = J_1 + J_2 \), with \( J_1, J_2 : V \rightarrow \bar{R} \) convex, lsc functionals and \( J_1 \) Gâteaux-differentiable on \( V \). Then, concerning Problem (A1.5.11), the following statements are equivalent for some \( u \in K \):

(i) \( u \) is a solution of Problem (A1.5.11);

(ii) the following inequalities holds true

\[
\langle J'_1(u), v - u \rangle + J_2(v) - J_2(u) \geq 0 \quad \forall v \in K;
\]

\[
\langle J'_1(v), v - u \rangle + J_2(v) - J_2(u) \geq 0 \quad \forall v \in K. \tag{A1.5.24}
\]

Proof: Let \( u \) be a solution of Problem (A1.5.11). Then, for \( \lambda \in (0, 1) \) and \( u^1 = (1 - \lambda)u + \lambda v \), due to the convexity of \( J_2 \),

\[
J(u) \leq J(u^1) \leq J_1(u^1) + (1 - \lambda)J_2(u) + \lambda J_2(v)
\]

is true for all \( v \in K \), i.e.,

\[
\frac{1}{\lambda} [J_1((1 - \lambda)u + \lambda v) - J_1(u)] + J_2(v) - J_2(u) \geq 0. \tag{A1.5.25}
\]
Taking $\lambda \downarrow 0$, the first relation in (ii) is proved. Conversely, using (A1.5.16) for $J'_1$, from the first relation in (ii) we just obtain

$$J(v) \geq J(u) \quad \forall \ v \in K.$$ 

Now, if the first inequality in (ii) holds, then the second is a conclusion of the monotonicity of $J'_1$

$$\langle J'_1(v) - J'_1(u), v - u \rangle \geq 0,$$

which follows from (A1.5.19) for $q = J'_1$.

Finally, let the second inequality in (ii) be fulfilled. For $v = (1 - \lambda)u + \lambda w$ with $w \in K$ and $\lambda \in (0, 1]$ we get

$$\langle J'_1((1 - \lambda)u + \lambda w), w - u \rangle + J_2(w) - J_2(u) \geq 0.$$ 

The functional

$$\eta(\lambda) := \eta_w(\lambda) = J_1((1 - \lambda)u + \lambda w) + \lambda (J_2(w) - J_2(u))$$

is convex and differentiable and, obviously,

$$\eta'(\lambda) = \langle J'_1((1 - \lambda)u + \lambda w), w - u \rangle + J_2(w) - J_2(u).$$

Consequently, $\eta'(\lambda) \geq 0$ for all $w \in K$ and $\lambda \in (0, 1]$. This leads to

$$\eta(1) \geq \lim_{\lambda \downarrow 0} \eta(\lambda) \geq \eta(0),$$

and the inequality $J(w) \geq J(u) \ \forall \ w \in K$ holds as well as the first inequality in (ii) is true. □

In case $J_2 \equiv 0$, we conclude from (A1.5.24) that

$$\langle J'(u), v - u \rangle \geq 0 \quad \forall \ v \in K. \quad (A1.5.26)$$

A1.5.35 Remark. If the convex functional $J$ is only subdifferentiable on the set $K$, then a solution $u$ of Problem (A1.5.11) is defined by

$$\langle q, v - u \rangle \geq 0 \quad \text{for some } q \in \partial J(u), \ \forall \ v \in K. \quad (A1.5.27)$$

Relations of type (A1.5.24), (A1.5.26) or (A1.5.27) are called variational inequalities. Here they have been obtained as solution criterions for convex extremal problems. There exist a number of applied problems which can be described by more general variational inequalities, so called hemi-variational inequalities,

$$\langle Q(u), v - u \rangle + \varphi(v) + \varphi(u) \geq 0 \quad \forall \ v \in K,$$

with $Q : V \rightarrow V'$ a monotone (not necessarily potential) operator and $\varphi : V \rightarrow \mathbb{R}$ a convex functional.

From Proposition A1.5.34 the following result can be immediately obtained (cf. KINDERLEHRER and STAMPACCHIA [234], chap. 1).
A1.5.36 Corollary. The operator
\[ \Pi_K(u) = \underset{z \in K}{\text{arg min}} \| u - z \| \]
of the orthogonal projection onto the convex, closed subset \( K \subset V \) is non-expansive, i.e.,
\[ \| \Pi_K u - \Pi_K v \| \leq \| u - v \| \quad \forall \ u, v \in V. \]

In order to investigate the variational problem (A1.5.11) from the numerical point of view, often the following criterion of an approximate solution \( v \) is used:
\[ \inf_{q(v) \in \partial J(v)} \| q(v) - q(\bar{u}) \|_{V'} \leq \epsilon \quad \text{for some } q(v) \in \partial J(v), \quad (A1.5.28) \]
with \( \bar{u} \) a minimum point of \( J \) on \( K \) and \( \epsilon > 0 \) a given tolerance.

In case \( J \) is Gâteaux-differentiable on \( V \) (A1.5.28) turns into
\[ \| J'(v) - J'(\bar{u}) \|_{V'} \leq \epsilon. \]

If \( J \) is a strongly convex functional with the constant \( \kappa > 0 \) in some neighborhood \( U(\bar{u}) \), then from (A1.5.28) we can estimate the distance \( \| v - \bar{u} \| \) for some \( v \in U(\bar{u}) \). Indeed, let \( \delta > 0 \) be an arbitrarily small number and \( q(\bar{u}) \) be an element of the subdifferential \( \partial J(\bar{u}) \) such that
\[ \| q(v) - q(\bar{u}) \|_{V'} \leq \epsilon + \delta. \quad (A1.5.29) \]
Then, using (A1.5.21), we get
\[ \| q(v) - q(\bar{u}) \|_{V'} \| v - \bar{u} \| \geq 2\kappa \| v - \bar{u} \|^2, \]
and because of (A1.5.29),
\[ \| v - \bar{u} \| \leq \frac{\epsilon + \delta}{2\kappa}. \]
But \( \delta > 0 \) is arbitrarily chosen, hence,
\[ \| v - \bar{u} \| \leq \frac{\epsilon}{2\kappa}. \quad (A1.5.30) \]
Now, it becomes clear that it suffices to require strong convexity of \( J \) only on the set
\[ \{ u : \| v - \bar{u} \| \leq \frac{\epsilon}{2\kappa} + \delta_0 \} \quad \text{with a small } \delta_0 > 0. \]

If, moreover, \( J \) is Lipschitz-continuous with the constant \( L \), the trivial estimate
\[ |J(v) - J(\bar{u})| \leq \frac{L\epsilon}{2\kappa} \quad (A1.5.31) \]
results from (A1.5.30).

In the case \( K \equiv V \), considering instead of (A1.5.28) the criterion
\[ \| q(v) \|_{V'} \leq \epsilon \quad \text{for some } q(v) \in \partial J(v), \]
we obtain (without using the Lipschitz-continuity of \( J \))
\[ J(v) - J(\bar{u}) \leq \frac{\epsilon^2}{2\kappa}. \quad (A1.5.32) \]
Indeed, the convexity of \( J \) provides
\[ J(v) - J(\bar{u}) \leq \langle q(v), v - \bar{u} \rangle \leq \| q(v) \|_{V'} \| v - \bar{u} \|_V \]
and there is nothing to do but to use inequality (A1.5.30).
A1.6 Monotone operators and monotone variational inequalities

For properties of monotone operators we refer, for instance, to Vainberg [405], Aubin and Ekeland [20], Burachik and Iusem [60] and Zeidler [418].

A1.6.37 Definition. A mapping $T : V \to 2^{V'}$ in a Hilbert space $V$ is called a multi-valued operator, i.e. $T$ maps a point $x \in V$ to a set $T(x) \subset V$. An operator $T$ is called single-valued if the image $T(x)$ contains at most one element. ♦

A single-valued operator $T : V \to V'$ is called hemi-continuous if its restriction to segments or lines are continuous, i.e.,

$$\forall u, v \in V : \quad T(u + \alpha(v - u)) \to T(v) \quad \text{as} \quad \alpha \to 1.$$  

For multi-valued operators one defines the effective domain, the range and the graph, respectively, as

\[
\begin{align*}
\text{dom}(T) & := \{x \in V : T(x) \neq \emptyset\}, \\
\text{rg}(T) & := \bigcup_{x \in \text{dom}(T)} T(x), \\
\text{gr}(T) & := \{(x, u) \in V \times V' : u \in T(x), x \in \text{dom}(T)\}.
\end{align*}
\]

The inverse operator $T^{-1}$ of $T$ is the multi-valued operator defined by the equivalence

$$x \in T^{-1}(y) \iff y \in T(x).$$

For two operators $T_1, T_2 : V \to 2^{V'}$ and two scalars $\alpha, \beta \in \mathbb{R}$ one defines the sum $\alpha T_1 + \beta T_2$ by

$$\alpha T_1 + \beta T_2)(x) = \begin{cases} 
\alpha T_1(x) + \beta T_2(x) & \text{if } x \in \text{dom}(T_1) \cap \text{dom}(T_2), \\
\emptyset & \text{otherwise}
\end{cases}$$

A1.6.38 Definition. (Vainberg [405])

A multi-valued operator $T : V \to 2^{V'}$ is called locally bounded if for any $x \in \text{int}(\text{dom}(T))$ there exists a neighborhood $U(x)$ such that the set $\cup_{u \in U(x)} T(u)$ is bounded. It is called locally hemi-bounded at a point $x^0 \in D(T) \subset V$ if and only if for each $x \in D(T)$, $x \neq x^0$, there exists a number $\ell_0(x^0, x) > 0$ such that $x^0 + \ell(x - x^0) \in D(T)$ holds for $0 \leq \ell \leq \ell_0(x^0, x)$ and the set

$$\bigcup_{0 < \ell \leq \ell_0(x^0, x)} T(x^0 + \ell(x - x^0))$$

is bounded in $V'$.

Here we use a weakened notion of local hemi-boundedness: the standard notion supposes boundedness of $\bigcup_{0 \leq \ell \leq \ell_0(x^0, x)} T(x^0 + \ell(x - x^0))$.

Different monotonicity assumptions can be found when studying solution methods for variational inequalities.

A1.6.39 Definition. The multi-valued operator $T : C \to 2^{V'}$, with $C \subset V$ is called monotone if

$$\langle u - v, x - y \rangle \geq 0 \quad \forall \ x, y \in C, \forall \ u \in T(x), v \in T(y);$$
strictly monotone, if
\[ \langle u - v, x - y \rangle > 0 \quad \forall \ x, y \in C, x \neq y, \ \forall \ u \in T(x), v \in T(y); \]
and strongly monotone if
\[ \langle u - v, x - y \rangle \geq \kappa \| x - y \|^2 \quad \forall \ x, y \in C, \ \forall \ u \in T(x), v \in T(y), \]
with a constant \( \kappa > 0 \).

The operator \( T \) is called maximal monotone if its graph is not properly contained in the graph of any other monotone operator \( T' \). ♦

If \( T_1, T_2 \) are monotone, then \( T_1^{-1}, \lambda T_1 \ (\lambda \geq 0) \) and \( T_1 + T_2 \) are monotone, too. A strongly monotone operator is obviously strictly monotone. For instance, the subdifferential of a proper lsc and strictly (strongly) convex functional is strictly (strongly) monotone.

If \( T \) is maximal monotone, then \( T^{-1} \) and \( \lambda T \ (\lambda > 0) \) are maximal monotone. Further properties of a maximal monotone operator \( T \) are:

- \( \text{gph}(T) \) is a convex and closed set for all \( x \in \text{dom}(T) \) (cf. [418], Prop. 32.6);
- \( \text{gph}(T) \) is closed;
- \( \text{rge}(T) = V \) of \( \text{dom}(T) \) is bounded (cf. [418], Prop. 32.35);
- \( \text{cl}(\text{dom}(T)), \text{ri}(\text{dom}(T)), \text{cl}(\text{rge}(T)) \) and \( \text{ri}(\text{rge}(T)) \) are convex sets (cf. [24], Prop. 6.4.1);
- \( T \) is locally bounded in \( \text{int}(\text{dom}(T)) \) (cf. [347], Thm.1);
- \( T \) is upper semi-continuous in \( \text{int}(\text{dom}(T)) \) (cf. [24], Prop. 6.6.8), i.e., for any \( x \in \text{int}(\text{dom}(T)) \) and any open set \( \mathcal{O} \supset T(x) \) there exists a neighborhood \( U(x) \) such that \( T(y) \subseteq \mathcal{O} \) for all \( y \in U(x) \). Thus, a single-valued, maximal monotone operator is continuous in the interior of its domain.

Important examples of maximal monotone operators is are

(i) For a proper, lsc convex functional \( f : V \rightarrow \mathbb{R} \cup \{+\infty\} \) the subdifferential \( \partial f : V \rightarrow 2^{V'} \), defined as
\[ \partial f(x) := \{ s \in V : f(y) \geq f(x) + \langle s, y - x \rangle \ \forall \ y \in V \} \]
is a maximal monotone operator (cf. [347], Thm. 12.17).

(ii) Single-valued operators that are monotone and continuous are maximal monotone (cf. [418], Prop. 32.7).

(iii) An affine operator \( T(x) = Ax + b \) with \( A \in \mathbb{R}^{n \times n} \) and \( b \in \mathbb{R}^n \) is maximal monotone if \( \frac{1}{2}(A + A^T) \) is positive semi-definite (cf. [353], Example 12.2).

(iv) The normality operator at a point \( u \) on the convex closed set \( K \):
\[ N_K(u) := \{ d \in V' : \langle d, v - u \rangle \leq 0 \ \forall \ v \in K \}. \]
is maximal monotone (cf. [353], Corollary 12.18).
The sum of maximal monotone operators is not necessarily maximal monotone.

**A1.6.40 Theorem.** ([350], Thm. 1)  
Let $T_1$ and $T_2$ be maximal monotone operators on $V$ such that  
$$\text{dom}(T_1) \cap \text{int}(\text{dom}(T_2)) \neq \emptyset.$$  
Then $T_1 + T_2$ is maximal monotone.

**A1.6.41 Theorem.** ([350], Thm. 3)  
Let $K \subset V$ be a nonempty, closed and convex subset and $T$ be a single-valued monotone operator (not necessarily maximal) such that $K \subset \text{dom}(T)$ and $T$ is hemi-continuous on $K$. Then $T + N_K$ is maximal monotone.

**A1.6.42 Definition.** (Lions [270])  
The multi-valued operator $Q : V \to 2^{V'}$ is called pseudo-monotone if the following implication holds: If $\{v^k\} \subset D(Q)$ converges weakly to $v \in D(Q)$ and  
$$\lim_{k \to \infty} \langle w^k, v^k - v \rangle \leq 0$$  
holds with $w^k \in Q(v^k)$, then for each $y \in D(Q)$ there exists $w \in Q(v)$ such that  
$$\lim_{k \to \infty} \langle w^k, v^k - y \rangle \geq \langle w, v - y \rangle.$$  

There are several notions of pseudo-monotonicity. This notion was introduced by Brézis [50] and [270] for single-valued operators. It should not be mixed up with those used in a couple of recent papers on variational inequalities (see, for instance Crouzeix [85]). For multi-valued operators see for instance [318].

**A1.6.43 Definition.** (Censor et al. [70])  
A multi-valued operator $Q : V \to 2^{V'}$ is called paramonotone on the set $C \subset V$ if it is monotone and  
$$\langle z - z', v - v' \rangle = 0 \text{ with } v, v' \in C, \ z \in Q(v), \ z' \in Q(v')$$  
implies  
$$z \in Q(v'), \ z' \in Q(v).$$  

**A1.6.44 Remark.** The following property of paramonotone operators is crucial:  

\begin{itemize}  
  \item[(*]) If $v^*$ solves $\text{VI}(Q, K)$ with a paramonotone $Q$ and for $\bar{v} \in K$ there exists $\bar{q} \in Q(\bar{v})$:  
    $$\langle \bar{q}, v^* - \bar{v} \rangle \geq 0,$$  
then $\bar{v}$ is a solution of $\text{VI}(Q, K)$, too.  
\end{itemize}
Paramonotonicity is a property lying in between monotonicity and strict monotonicity, i.e., strictly monotone operators are paramonotone. If $\mathcal{T}$ is the subdifferential of a proper, lsc and convex functional then $\mathcal{T}$ is paramonotone. The sum of two paramonotone operators is paramonotone, too (cf. IUSEM [194]).

By the way, the maximal monotone operator associated with a Lagrangian of a smooth convex programming problem, see (A1.7.37) below, is paramonotone only in the very special situation that all constraints are not active.

A1.6.45 **Definition.** The operator $\mathcal{T} : V \rightarrow V'$ is called co-coercive on a set $K \subset V$ with modulus $\gamma > 0$ if

$$\langle \mathcal{T}(x) - \mathcal{T}(y), x - y \rangle \geq \gamma \| \mathcal{T}(x) - \mathcal{T}(y) \|^2 \quad \forall \ x, y \in K.$$

\[\Diamond\]

Sometimes the latter relation is considered as Dunn property (cf. [94]).

Co-coercivity is a concept of generalized monotonicity for single-valued operators that lies strictly between simple and strict monotonicity. A co-coercive operator $\mathcal{T}$ with modulus $\gamma$ is monotone and Lipschitz-continuous with constant $1/\gamma$. Co-coercive operators are paramonotone but not necessarily strongly monotone.

A sufficient condition for an operator to be co-coercive is: Let $\mathcal{T} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be Lipschitz-continuous with constant $L$ and strongly monotone with modulus $\kappa$. Then, $\mathcal{T}$ is co-coercive with modulus $\gamma = \frac{\kappa}{L^2}$ (cf. [421], Prop. 3.1).

For convex and differentiable functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$, Lipschitz-continuity of $\nabla f$ with constant $L$ is equivalent to $\nabla f$ being co-coercive with modulus $L^{-1}$ (cf. [421], Prop. 3.5).

Further, co-coercivity of an operator $\mathcal{T}$ is equivalent to the strong monotonicity of the possibly multi-valued operator $\mathcal{T}^{-1}$ (cf. [421], Prop. 3.3).

Notions that describe the behaviour of a monotone operator at infinity are useful to get statements about solvability of variational inequalities.

A1.6.46 **Definition.** (ZEIDLER [418]) A multi-valued operator $\mathcal{T} : V \rightarrow 2^V$ is said to be coercive if

$$\lim_{\|x\| \to \infty} \frac{\inf_{u \in \mathcal{T}(x)} \langle u, x \rangle}{\|x\|} = +\infty$$

and weakly coercive if

$$\lim_{\|x\| \to \infty} \inf_{u \in \mathcal{T}(x)} \|u\| = +\infty,$$

with $\inf \emptyset = +\infty$.

\[\Diamond\]

It is clear that coercive operators are weakly coercive, and that a bounded effective domain of $\mathcal{T}$ is sufficient for $\mathcal{T}$ to be coercive. Further, strongly monotone operators are weakly coercive.

The connection between continuous differentiable mappings and their Jacobians describes the following
A1.6.47 Lemma. Let \( F : D \subset \mathbb{R}^n \to \mathbb{R}^n \) be continuous differentiable on the open set \( D \). Then it holds:

(a) \( F \) is monotone on \( D \) ⇔ the Jacobian \( JF(x) \) is positive semi-definite ∀ \( x \in D \);

(b) \( F \) is strictly monotone on \( D \) if \( JF(x) \) is positive definite ∀ \( x \in D \);

(c) \( F \) is strongly monotone on \( D \) ⇔ \( JF(x) \) is uniformly positive definite ∀ \( x \in D \), i.e., there exists \( c > 0 \) with

\[
\langle y, JF(x)y \rangle \geq c\|y\|^2, \quad \forall \ y \in \mathbb{R}^n, \ \forall \ x \in D.
\]

Now we describe some basic results about existence and uniqueness of solutions for variational inequalities needed to prove well-definedness of solution methods.

Let be given a set-valued operator \( Q : V \to 2V' \) and a closed convex subset \( K \) of \( V \). Since the variational inequality

\[
\text{VI}(Q,K) \quad \text{find a pair } x^* \in K \text{ and } q^* \in Q(x^*) \text{ such that}
\]

\[
\langle q^*, x - x^* \rangle \geq 0 \quad \forall \ x \in K.
\]

is equivalent to the inclusion problem \( \text{IP}(T,K) \)

\[
\text{IP}(T,V) \quad \text{find } x \in V : \ 0 \in T(x),
\]

with operator \( T := Q + N_K \), existence and uniqueness results for inclusion problems can easily be transferred to VI’s.

Obviously an inclusion problem \( \text{IP}(T,K) \) is solvable if and only if

\[
0 \in \text{rge}(T).
\]

So, conditions on \( T \) guaranteeing that \( \text{rge}(T) = V \) are sufficient for the existence of a solution.

A1.6.48 Theorem. ([418], Corollary 32.35)

Let \( T : V \to 2V' \) be a maximal monotone and weakly coercive operator. Then \( \text{rge}(T) = V \).

Of particular interest for us are inclusion problems with operators of special structure, like \( T + \partial f \).

A1.6.49 Theorem. ([59], Prop. 3)

Let \( T : V \to 2V' \) be a monotone operator and \( f : V \to \mathbb{R} \cup \{+\infty\} \) a proper, lsc and convex functional. Suppose further that the following conditions are satisfied:

(i) \( \text{dom}(T) \cap \text{dom}(\partial f) \neq \emptyset \) and \( \text{rge}(\partial f) = V \);

(ii) \( T + \partial f \) is maximal monotone.

Then \( \text{rge}(T + \partial f) = V \).

This theorem does not ensure uniqueness of the solution. Uniqueness is guaranteed if the operator \( T + \partial f \) is strictly monotone as follows immediately from the definition. Strong monotone operators are strictly monotone and weakly coercive and therefore ensure both existence and uniqueness.

A1.6.50 Theorem. ([353], Thm. 12.54)

For a maximal monotone and strongly monotone operator \( T : V \to 2V' \) the inclusion \( \text{IP}(T,V) \) has a unique solution.
A1.7 Convex minimization problems in Hilbert spaces

Now we consider the problem:

\[
\begin{align*}
\text{minimize} & \quad J(u), \\
\text{subject to} & \quad u \in K,
\end{align*}
\]

\[K := \{ v \in V : g_j(v) \leq 0, \quad j = 1, \ldots, m \},\]

\[(A1.7.35)\]

\(J, g_j : V \to \mathbb{R}\) are convex, lower semi-continuous functionals.

Due to Proposition A1.5.13(iii) the set \(K\) is convex and closed, hence, Problem (A1.7.35) is a particular case of Problem (A1.5.11).

Together with this problem we consider the saddle point problem:

\[
\text{find } (\bar{u}, \bar{\lambda}) \in V \times \mathbb{R}_+^m \text{ such that }
\]

\[
L(u, \bar{\lambda}) \geq L(\bar{u}, \bar{\lambda}) \geq L(\bar{u}, \lambda) \quad \forall \ u \in V, \quad \forall \ \lambda \in \mathbb{R}_+^m,
\]

\[(A1.7.36)\]

with

\[
L(u, \lambda) = J(u) + \sum_{j=1}^m \lambda_j g_j(u).
\]

\[(A1.7.37)\]

Function \(L\) is called Lagrangian function of Problem (A1.7.35) and the pair \((\bar{u}, \bar{\lambda})\), satisfying (A1.7.36), is called a saddle point of \(L\) with \(\bar{\lambda}\) the Lagrange multiplier vector corresponding to \(\bar{u}\).

For \(\overline{Z}(u) = \sup_{\lambda \geq 0} L(u, \lambda)\) we obtain

\[
\overline{Z}(u) = \begin{cases} 
J(u) & \text{if } u \in K, \\
+\infty & \text{otherwise}.
\end{cases}
\]

Hence, one can reformulate Problem (A1.7.35) in the form

\[
\min \{ \overline{Z}(u) : u \in V \}.
\]

The problem

\[
\min \{ L(\lambda) : \lambda \in \mathbb{R}_+^m \},
\]

\[(A1.7.38)\]

with \(L(\lambda) = \inf_{u \in V} L(u, \lambda)\) is called the dual problem to Problem (A1.7.35).

One can easily show that for any \(u \in V\) and \(\lambda \in \mathbb{R}_+^m\)

\[
\overline{Z}(u) \geq L(\lambda),
\]

and equality in the latter relation means that \((u, \lambda)\) is a saddle point of the Lagrangian function. Hence, if \((\bar{u}, \bar{\lambda})\) is a saddle point of (A1.7.36), then \(\bar{u}\) is a solution of Problem (A1.7.35).

Later on, we need the following constraint qualification for the constraints in Problem (A1.7.35):

**Slater condition**: there exists a point \(\tilde{u}\) such that

\[g_j(\tilde{u}) > 0 \quad \text{for } j = 1, \ldots, m.\]
\section*{A1. SOME PRELIMINARY RESULTS IN FUNCTIONAL ANALYSIS}

\subsection*{A1.7.51 Theorem. (Kuhn-Tucker theorem)}
Assume the constraints of Problem (A1.7.35) fulfill the Slater condition and \( \bar{u} \) is a solution of this Problem. Then there exists a vector \( \bar{\lambda} \in \mathbb{R}^m_+ \) such that \( \langle \bar{u}, \bar{\lambda} \rangle \) is a saddle point of Problem (A1.7.36).

Statements of this type under different constraint qualifications are the foundation of the theory of convex programming.

\subsection*{A1.7.52 Proposition.}
Assume the Slater condition is fulfilled for Problem (A1.7.35). Then \( \bar{u} \in K \) is a solution of this problem iff there exists a vector \( \bar{\lambda} \geq 0 \) such that
\[
0 \in \partial J(\bar{u}) + \sum_{i=1}^m \bar{\lambda}_i \partial g_j(\bar{u}),
\]
(A1.7.39)
\[\lambda_j g_j(\bar{u}) = 0, \ j = 1, \ldots, m.\]

If in Problem (A1.5.11) the set \( K \) is defined by
\[
K = \{ u \in V : B(u) \leq 0 \},
\]
(A1.7.40)
with \( B \) a convex mapping from \( V \) into a Banach space \( X \), then the following analogue of the Kuhn-Tucker theorem holds (see Krabs \[248\]). Denote \( Y' \) a non-negative cone in the dual space \( Y' \).

\subsection*{A1.7.53 Proposition.}
Assume that \( \text{int} Y_+ \neq \emptyset \) and that for each \( \lambda \in Y'_+ \), \( \lambda \neq 0 \), a point \( \tilde{u} \in V \) exists such that
\[
\langle \lambda, B(\tilde{u}) \rangle_Y < 0,
\]
(A1.7.41)
Then \( \bar{u} \) is a solution of Problem (A1.5.11), (A1.7.40) iff there exists an element \( \bar{\lambda} \in Y'_+ \) such that for all \( u \in V \) and \( \lambda \in Y'_+ \)
\[
J(u) + \langle \bar{\lambda}, B(u) \rangle_Y \geq J(\bar{u}) + \langle \bar{\lambda}, B(\bar{u}) \rangle_Y \geq J(\bar{u}) + \langle \lambda, B(\bar{u}) \rangle_Y.
\]

Now, let us consider the following convex semi-infinite problem
\[
\begin{align*}
(\text{SIP}) \quad & \text{minimize} \quad J(u) \\
\text{subject to} \quad & u \in K \equiv \{ v \in \mathbb{R}^n : g(v, t) \leq 0 \quad \forall \ t \in T \},
\end{align*}
\]
(A1.7.42)
with \( T \) a compact subset of a normed space \( Y \); \( J, g_{t \in T} : u \to g(u, t) \) convex, finite-valued functions, \( g_u : t \to g(u, t) \) a continuous function on \( T \) for any \( u \in \mathbb{R}^n \),
\[
\exists \bar{u} : \sup_{t \in T} g(\bar{u}, t) < 0.
\]
(A1.7.43)
Setting \( Y = C(T) \) and \( B : u \to g(u, \cdot) \), Problem (A1.7.42) can be considered as a special case of Problem (A1.5.11), (A1.7.40). Since condition (A1.7.41) is a conclusion of (A1.7.43), Proposition A1.7.53 is also true for Problem (A1.7.42).

However, in that case the space \( Y' \) is the space of all Diraque-measures and very inconvenient with respect to an approximation.
The following statement characterizes the connections between convex semi-infinite and finite-dimensional problems (cf. Pšenčínyj [338], for other variants of reduction theorems see Hettich and Jongen [175]).

Let $T(u) = \{ t \in T : g(u, t) = \max_{\tau \in T} g(u, \tau) \}$. 

**A1.7.54 Proposition.** (Reduction theorem)

Assume that for Problem (A1.7.42) the Slater condition is fulfilled. Then $\bar{u}$ is a minimizer of the problem considered iff there exist $r \leq n + 1$ points $t_i \in T(\bar{u})$ such that

$$
\min \{ J(u) : g(u, t_i) \leq 0, \ i = 1, ..., r \}. \tag{A1.7.44}
$$

**Proof:** Without loss of generality we take $\rho_o = \sup\{ ||u - \bar{u}|| : u \in U_0 \}$. For arbitrary $u \in \partial U_0$ ($\partial U_0$ is the boundary of $U_0$) we define the point $z = \bar{u} + \lambda(\bar{u} - u)$, with $\lambda = \rho_o \delta(\psi(\bar{u})) / ||\bar{u} - u||^{-1}$. Then,

$$
\bar{u} = \frac{1}{1 + \lambda} z + \frac{\lambda}{1 + \lambda} \bar{u}.
$$
and, due to $\psi(\tilde{u}) = 0$ and the convexity of $\psi$,

$$0 = \psi(\tilde{u}) \leq \frac{1}{1 + \lambda} \psi(z) + \frac{\lambda}{1 + \lambda} \psi(\tilde{u})$$

holds true, hence,

$$\psi(z) \geq \lambda |\psi(\tilde{u})| = \rho_0 \delta(\|\tilde{u} - \tilde{u}\|)^{-1} \geq \delta.$$

But $z - \tilde{u} = (1 + \lambda)(\tilde{u} - \tilde{u})$, therefore,

$$\|z - \tilde{u}\| \leq (1 + \rho_0 \delta(\|\tilde{u} - \tilde{u}\|)^{-1}) \|\tilde{u} - \tilde{u}\| \leq \rho_0 + \rho_0 \delta(\|\tilde{u}\|)^{-1} = r_0.$$

Consequently, $U_\delta \subset B_{r_0}(\tilde{u})$, with $B_{r_0}(\tilde{u}) = \{u \in V : \|u - \tilde{u}\| \leq r_0\}$.

Now, let $d_j \downarrow 0$, $j = 1, \ldots, m$. This leads to $\delta \downarrow 0$ and assuming that

$$\lim_{\delta \downarrow 0} \rho(U_\delta, U_0) = \alpha > 0,$$

we have

$$\lim_{\delta \downarrow 0} \rho(U_\delta, U_0) = \alpha > 0$$

for a monotone subsequence $\delta \downarrow 0$ which is used in the sequel instead of the initial sequence. Then

$$\rho(\cap_{\delta > 0} U_\delta, U_0) = \alpha > 0$$

and one can choose points $w$ and $\tilde{w}$ such that

$$w \in (\cap_{\delta > 0} U_\delta) \setminus U_0, \quad \tilde{w} \in [\tilde{u}, w] \cap \partial U_0.$$

Due to the convexity of $\psi$ on the interval $[\tilde{u}, w]$ and $\psi(\tilde{w}) < 0, \psi(\tilde{w}) = 0$, we obtain $\psi(w) = \text{const} > 0$. But this is impossible, because $w \in U_\delta$ for all $\delta > 0$ belonging to the considered monotone subsequence.

For a finite-dimensional space this result remains true without the assumption about the existence of a Slater point for the set $U_\delta$. Indeed, as before we can suppose that $c_j = 0$, $j = 1, \ldots, m$, and because $U_0 \subset B_{\rho'}(v)$ with some $v \in U_0$ and $\rho'$, the relation

$$\min_{w \in \partial B_{\rho'}(v)} \psi(w) = c_0 > 0$$

is true. Hence, $U_{c_0} \subset B_{2\rho'}(v)$ and $\psi_j(v) < c_0$ for all $j$. Now, from Proposition A1.7.55 we conclude that $U_d$ is bounded for any $d$.

Moreover, we suppose that $d_j \downarrow 0$, $j = 1, \ldots, m$, and that

$$\lim_{\delta \downarrow 0} \rho(U_\delta, U_0) > 0$$

for a monotone subsequence $\delta \downarrow 0$ (as before, $\delta = \max_{1 \leq j \leq m} d_j$). Then, choosing $w \in (\cap_{\delta > 0} U_\delta) \setminus U_0$ and observing the continuity of $\psi$ and the closedness of $U_0$, we obtain $\psi(w) > 0$, but this is also impossible.

The following simple example shows that the Slater condition is essential in the infinite-dimensional case. Let $V = \ell_2$, $m = 1$ and

$$\psi(u) = \sum_{i=1}^{\infty} i^{-1} u_i^2.$$

Obviously, $\{u : \psi(u) \leq 0\} = \{0\}$, and for any $d > 0$ the set $U_d = \{u : \psi(u) \leq d\}$ contains the points $u = (0, \ldots, 0, \sqrt{d}, 0, \ldots)$, with $\sqrt{d}$ the value of the $i$-th component. Consequently, $U_d$ is bounded.
A2 Approximation Techniques and Estimates of Solutions

Our investigation of variational problems is mainly concerned with monotone (elliptic) variational inequalities. To provide a finite-dimensional approximation of such problems finite difference methods (FDM), wavelets and finite element methods (FEM) are applicable. The latter are most popular and we will use them here.

There are numerous publications, including a number of monographs, devoted to finite element approximations for linear as well as nonlinear problems of mathematical physics (see, in particular, Strang and Fix [382], Ciarlet [75], Glowinski, Lions and Tremolières [135], Neittaanmaki, Sprekels and Tiba [304]).

In this section finite element methods and some estimates of finite element approximations are briefly described.

A2.1 Some examples of monotone variational inequalities

In order to illustrate the peculiarity of the approximation of variational inequalities, we consider three examples. For simplicity, everywhere in this section the set Ω is supposed to be a polygonal domain in \( \mathbb{R}^2 \) with boundary \( \Gamma \).

**Dirichlet Problem:**

\[
-\Delta u = f \quad \text{in } \Omega, \\
\quad u = 0 \quad \text{on } \Gamma.
\]  

(A2.1.1)

Here and in the examples below the equalities or inequalities with respect to functions are to be understood in the corresponding functional spaces (see Kinderlehrer and Stampacchia [234], Chapt.2). They can be considered formally until there are not made suitable assumptions or statements about regularity of the solutions.

For fixed \( f \in L^2(\Omega) \) the Dirichlet problem has a unique weak solution \( u \in V \equiv H^1_0(\Omega) \) defined by

\[
\langle \nabla u, \nabla v \rangle_{0, \Omega} = \langle f, v \rangle_{0, \Omega} \quad \forall \ v \in V.
\]  

(A2.1.2)

In particular, if \( \Omega \) is a convex polygon, then \( u \in H^2_0(\Omega) \).

**Obstacle Problem:**

\[
\begin{align*}
-\Delta u & \geq f \\
\quad u & \geq \varphi \\
(-\Delta u - f)(u - \varphi) & = 0 \\
\quad u & = 0 \quad \text{on } \Gamma, \\
\quad u & = \varphi, \quad \frac{\partial u}{\partial \nu} = \frac{\partial \varphi}{\partial \nu} \quad \text{on } \Gamma^*,
\end{align*}
\]

(A2.1.3)

with \( f \in L^2(\Omega), \varphi \in H^2(\Omega), \varphi|_{\Gamma} \leq 0, \Gamma^* \) the unknown "boundary" between

\[
\{ u \in \Omega : u(x) = \varphi(x) \} \quad \text{and} \quad \{ u \in \Omega : u(x) > \varphi(x) \}
\]
A2. APPROXIMATION TECHNIQUES AND ESTIMATES OF SOLUTIONS

and \( \nu \) a unit normal to \( \Gamma^* \).

The relations (A2.1.3) characterize the unique solution \( u \) of the variational inequality

\[
\langle \nabla u, \nabla(v - u) \rangle_{0,\Omega} \geq \langle f, v - u \rangle_{0,\Omega} \quad \forall \ v \in K, \tag{A2.1.4}
\]

\[
K = \{ w \in V : w \geq \varphi \text{ on } \Omega \}, \quad V = H^1_0(\Omega). \tag{A2.1.5}
\]

Conditions, ensuring that \( u \in H^1_0(\Omega) \cap H^2(\Omega) \) are described in Brezis and Stampacchia [53], Lewy and Stampacchia [264]. For stronger foundations of the regularity of the solutions see Friedman [120].

**Problem with Obstacle on the Boundary:**
(for short: Boundary-obstacle Problem)

\[
-\Delta u = f \quad \text{in } \Omega,
\]

\[
u a \text{ unit normal to } \Gamma^*.
\]

\[
u a \text{ unit normal to } \Gamma^*.
\]

\[
\begin{align*}
-\Delta u &= f \quad \text{in } \Omega, \\
\quad u - \varphi_0 &\geq 0, \quad \frac{\partial u}{\partial \nu} \geq 0, \quad (u - \varphi_0) \frac{\partial u}{\partial \nu} = 0 \quad \text{on } \Gamma,
\end{align*}
\]

\[
K = \{ w \in V : w \geq \varphi_0 \text{ on } \Gamma \}, \quad V = H^1(\Omega). \tag{A2.1.8}
\]

Under the conditions described all these problems can be reformulated as convex variational problems:

\[
\min_{u \in K} J(u), \quad J(u) = \frac{1}{2} a(u, u) - \langle f, u \rangle, \tag{A2.1.9}
\]

with \( a(u, v) = (\nabla u, \nabla v)_{0,\Omega} \). For the Dirichlet Problem (A2.1.1) the set \( K \) has to be chosen as \( K = V \), whereas in the Obstacle Problem (A2.1.3) and Boundary-obstacle Problem (A2.1.6) the set \( K \) is defined by (A2.1.5) and (A2.1.8), respectively. Here \( \langle \cdot, \cdot \rangle \) is the duality connection between \( V \) and \( V' \) and \( \langle f, u \rangle \) coincides with the scalar product \( \langle f, u \rangle_{0,\Omega} \) for \( f \in L^2(\Omega) \).

The statements above on the equivalence between boundary value problems and variational inequalities remain also true for more general domains (see Friedman [120], Hlaváček et al. [182]). The Obstacle Problems (A2.1.3) and (A2.1.6) have a unique solution. Moreover, in (A2.1.3) the corresponding bilinear form \( a(\cdot, \cdot) \) is coercive.

A2.2 Finite element methods

**A2.2.1 Definition.** A family \( \{ V_h \} \) of finite-dimensional vector spaces is called an inner approximation of a Hilbert space \( V \) if

\( i \) \( V_h \subset V \) for all \( h \), and

\( ii \) for each \( v \in V \) there exist elements \( v_h \in V_h \) such that \( ||v_h - v||_V \to 0 \) if \( h \downarrow 0 \).
Sometimes it is more convenient to reformulate condition (ii) into

\[(ii)' \quad \text{for every } v \in \tilde{V}, \text{ with } \tilde{V} \text{ a dense set in } V, \text{ there exist elements } v_h \in V_h \text{ such that} \]

\[\|v_h - v\| \to 0 \text{ if } h \downarrow 0.\]

We first describe a finite element method for polygonal domains \(\Omega \subset \mathbb{R}^2\) and show afterwards its modification for more general domains.

**A2.2.2 Definition.** For a polygonal domain \(\Omega\) a sequence of triangulations \(\{T_h\}\) with parameter \(h \downarrow 0\) is called *quasi-uniform* if for fixed \(h\) the set \(T_h\) consists of open triangles \(T_j\) such that:

1. \(\Omega = \text{int}(\cup T_j) =: \Omega_h\), (\(\bar{T}_j\) the closure of \(T_j\));
2. \(\bar{T}_i \cap \bar{T}_j = \emptyset\), otherwise \(\bar{T}_i\) and \(\bar{T}_j\) have only a common vertex or a common side;
3. the maximal length of a side of each triangle \(\bar{T}_j\) is not larger than \(h\);
4. the area of each triangle \(\bar{T}_j \in \bar{T}_h\) is bounded from below by \(ch^2\), with \(c > 0\) independent of \(h\) and \(T_j\).

Now we consider some special finite element spaces which are subspaces of \(H^1(\Omega)\) or \(H^1_0(\Omega)\). Usually, solving boundary value problems with (essentially nonlinear) differential operators of second order on polygonal domains, approximations by elements of these subspaces are performed.

We further introduce the following notations:

\(M_h\) is the set of all vertices of the triangulation,

\[M_h^0 := M_h \cap \Omega_h, \quad \Gamma_h := \bar{\Omega}_h \setminus \Omega_h, \quad N_h := M_h \cap \Gamma_h,\]

\(I_h, \quad I_0^0, \quad I'_h\) are index sets of the sets \(M_h, M_0^0, N_h\), respectively. (A2.2.10)

Let \(P_i, i \in I_h\), be the vertices of the triangles in a fixed triangulation \(\mathcal{T}_h\). For each \(i \in I_h\) we define a function \(\varphi_i \in C(\bar{\Omega})\) that is affine on each triangle \(\bar{T} \in \mathcal{T}_h\) and

\[\varphi_i(P_j) = 1, \quad \varphi_i(P_j) = 0 \quad \text{for } j \neq i.\]

Using \(\{\varphi_i\}\) as a basis, we construct the *finite element spaces*

\[V_h := \{v_h(x) = \sum_{i \in I_h} \alpha_i \varphi_i(x) : \alpha_i \in \mathbb{R}\}\]

and

\[V_0^0 := \{v_h(x) = \sum_{i \in I_0^0} \alpha_i \varphi_i(x) : \alpha_i \in \mathbb{R}\}.\]

The families of spaces \(\{V_h\}\) and \(\{V_0^0\}\) build an inner approximation of the spaces \(V = H^1(\Omega)\) and \(V = H^1_0(\Omega)\), respectively. In order to verify this we recall that \(C^\infty(\Omega)\) and \(\mathcal{D}(\Omega)\) are dense in \(H^1(\Omega)\) and \(H^1_0(\Omega)\), respectively.
A2. APPROXIMATION TECHNIQUES AND ESTIMATES OF SOLUTIONS

A2.2.3 Remark. We make use of an affine mapping of an arbitrary triangle $T \subset \mathcal{T}_h$ onto a standard triangle $\Delta \subset \mathbb{R}^2$ (with vertices $(0,0), (1,0), (0,1)$). Such a mapping can be represented as a successive transformation of $T$ onto the triangle $\Delta_h$ (with vertices $(0,0), (h,0), (0,h)$) and of $\Delta_h$ onto $\Delta$. Due to quasi-uniformity of the triangulation, the affine mapping $Q : T \rightarrow \Delta_h$ and its inverse have Jacobians uniformly bounded away from 0 (with respect to $T$). Consequently, for functions $v \in H^1(T)$ and $v_Q = v \circ Q^{-1} \in H^1(\Delta_h)$ we get

$$c_1 \|v\|_{s,T} \leq \|v_Q\|_{s,\Delta_h} \leq c_2 \|v\|_{s,T}$$  \hspace{1cm} (A2.2.11)

where $s = 0, 1$, and $c_1 > 0$ and $c_2$ are independent of $h$ and $T$. \hfill \Diamond

A2.2.4 Definition. A function $u_I \in V_h(\Omega)$ is called interpolant of a function $u \in C(\Omega)$ (on the triangulation $\mathcal{T}_h$) if

$$u_I(P_i) = u(P_i) \quad \forall i \in I_h.$$ \hfill \Diamond

In the case $\Omega \subset \mathbb{R}^2$, regarding the continuous embedding $H^2(\Omega) \hookrightarrow C(\Omega)$, the notion of an interpolant can be extended to functions $u \in H^2(\Omega)$, too.

A2.2.5 Theorem. If $u \in H^2(\Omega)$ (resp., $u \in H^2(\Omega) \cap H^1_0(\Omega)$), then the interpolant $u_I \in V_h(\Omega)$ (resp. $u_I \in V_h^0(\Omega)$) satisfies the inequality

$$\|u - u_I\|_{0,\Omega} + h\|\nabla(u - u_I)\|_{0,\Omega} \leq c h^2 |u|_{2,\Omega},$$  \hspace{1cm} (A2.2.12)

with $|u|_{2,\Omega} = (\sum_{|\alpha| = 2} \|D^\alpha u\|^2)^{1/2}$.  

Proof: We will denote by $c$ different constants which are independent of $u$, $h$ and $\mathcal{T}_h$. Firstly, estimate (A2.2.12) can be established on $\Delta_h$. Put

$$y_1 := \frac{x_1}{h}, \quad y_2 := \frac{x_2}{h} \quad \text{and} \quad \tilde{u}(y_1, y_2) = u(x_1, x_2), \quad \tilde{u}_I(y_1, y_2) = u_I(x_1, x_2).$$

Then, obviously,

$$\tilde{u}_I(y_1, y_2) = \tilde{u}(0, 0) + \tilde{u}(1, 0) y_1 + \tilde{u}(0, 1) y_2$$

and, consequently,

$$\|\tilde{u}_I\|_{0,\Delta} \leq 5 \|\tilde{u}\|_{C(\Delta)}.$$  

Furthermore,

$$\|\tilde{u} - \tilde{u}_I\|_{0,\Delta} \leq \|\tilde{u}\|_{0,\Delta} + \|\tilde{u}_I\|_{0,\Delta} \leq \|\tilde{u}\|_{2,\Delta} + 5 \|\tilde{u}\|_{C(\Delta)},$$

and using the continuous embedding of $H^2(\Delta)$ into $C(\Delta)$, we get

$$\|\tilde{u} - \tilde{u}_I\|_{0,\Delta} \leq c \|\tilde{u}\|_{2,\Delta}.$$  \hspace{1cm} (A2.2.13)

Since linear interpolants reproduce linear functions, it follows that

$$\tilde{u}_I - \tilde{u}_I = \tilde{w} - \tilde{w} \quad \text{for any} \quad \tilde{w}(x_1, x_2) = \alpha + \beta x_1 + \gamma x_2.$$
Hence, in view of $\text{(A2.2.13)}$,

$$
\| \tilde{u} - \tilde{u}_I \|_{0,\Delta} = \| \tilde{u} - \tilde{w} - (\tilde{u}_I - \tilde{w}_I) \|_{0,\Delta} \leq c \| \tilde{u} - \tilde{w} \|_{2,\Delta},
$$

and because $w$ is arbitrarily chosen,

$$
\| \tilde{u} - \tilde{u}_I \|_{0,\Delta} \leq \inf_{z \in \mathcal{P}_1(\Delta)} \| \tilde{u} - z \|_{2,\Delta}, \quad \text{(A2.2.14)}
$$

with $\mathcal{P}_1(\Delta)$ a set of affine functions on $\Delta$. Similarly we can obtain

$$
\| \nabla(\tilde{u} - \tilde{u}_I) \|_{0,\Delta} \leq \inf_{z \in \mathcal{P}_1(\Delta)} \| \tilde{u} - z \|_{2,\Delta}, \quad \text{(A2.2.15)}
$$

Now, it is easy to show that

$$
\inf_{z \in \mathcal{P}_1(\Delta)} \| \tilde{u} - z \|_{2,\Delta} \leq c |\tilde{u}|_{2,\Delta},
$$

and thus,

$$
\| \tilde{u} - \tilde{u}_I \|_{0,\Delta} \leq c |\tilde{u}|_{2,\Delta}, \quad \text{(A2.2.16)}
$$

Using the transformation $x_1 = y_1 h$, $x_2 = y_2 h$, we get

$$
\| \tilde{u} - \tilde{u}_I \|_{0,\Delta} = h^{-1} \| u - u_I \|_{0,\Delta_h},
$$

$$
\| \nabla(\tilde{u} - \tilde{u}_I) \|_{0,\Delta} = \| \nabla(u - u_I) \|_{0,\Delta_h},
$$

$$
|\tilde{u}|_{2,\Delta} = h |u|_{2,\Delta_h}
$$

and due to $\text{(A2.2.15)}$, $\text{(A2.2.16)}$,

$$
\| u - u_I \|_{0,\Delta_h} + h \| \nabla(u - u_I) \|_{0,\Delta_h} \leq ch^2 |u|_{2,\Delta_h}. \quad \text{(A2.2.17)}
$$

From $\text{(A2.2.11)}$ and $\text{(A2.2.17)}$ it follows that for any $T \in \mathcal{H}_h$

$$
\| u - u_I \|_{0,T} + h \| \nabla(u - u_I) \|_{0,T} \leq ch^2 |u|_{2,T}. \quad \text{(A2.2.18)}
$$

Now, in order to obtain $\text{(A2.2.12)}$, we square both sides of inequality $\text{(A2.2.18)}$ and summing up all triangles $T \in \mathcal{H}_h$. □

**A2.2.6 Definition.** For a domain $\Omega \subset \mathbb{R}^2$, not necessarily polygonal, a system of triangulations $\{\mathcal{T}_h\}$ is called uniform if it is constructed as follows (see Fig. A2.2.1 below):

(i) $\Omega \subset \mathbb{R}^2$ is covered by a uniform grid (with step lengths $h_x$ and $h_y$ in direction of the corresponding axes);

(ii) the arising rectangles are split into triangles by diagonals having pointed angles with the $x$-axis;

(iii) $\mathcal{T}_h$ is the set of all open triangles, belonging to $\Omega$;

(iv) the ratio $\frac{1}{h} \min\{h_x, h_y\}$ is uniformly bounded from below by a constant $c > 0$, with $h := \sqrt{h_x^2 + h_y^2}$, $h \downarrow 0$. ◊
Obviously, for rectangular domains this triangulation $\mathcal{T}_h$ satisfies the condition $\Omega_h \equiv \Omega$ and enables us in this case to approximate $\Omega$ with high accuracy.

For a uniform triangulation basis functions $\varphi_i$ can be constructed using the standard function

$$
\varphi(s,t) := \begin{cases}
1 - \frac{1}{2}(|s| + |t| + |s-t|) & \text{for } (s,t) \in Q, \\
0 & \text{otherwise},
\end{cases}
$$

with $Q = \{(s,t) : -1 \leq s \leq 1, |s-t| \leq 1\}$ (see Fig. A2.2.2 below).

Indeed, for a vertex $P_i = (kh_x, lh_y)$ ($k, l$ integers), we get

$$
\varphi_i(x,y) = \varphi\left(\frac{x}{h_x} - k, \frac{y}{h_y} - l\right).
$$

This representation simplifies essentially the discretization.

However, in general situations a uniform triangulation of the whole domain $\Omega$ leads to a distance of order $h$ between the boundary $\Gamma$ and the discrete
boundary $\Gamma_h$ of the grid domain, i.e., large numerical errors may occur.

Now, let us consider a more effective approach of the triangulation, which guarantees a good accuracy of the approximation of the domain $\Omega$, supposing that $\Gamma$ is a piece-wise curve of class $C^2$.

We construct a closed broken line $\tilde{\Gamma}_h$ with the following properties:

(i) the domain $\tilde{\Omega}_h$, restricted by $\tilde{\Gamma}_h$, is contained in $\Omega$;

(ii) for the Hausdorff distance it holds $\rho_H(\Gamma, \tilde{\Gamma}_h) \leq \delta h^2$, with $\delta$ independent of $h$;

(iii) the lengths of the segments in $\tilde{\Gamma}_h$ are bounded from below by $c h$, with $c > 0$ independent of $h$;

(iv) the triangulation of $\tilde{\Omega}_h$ is carried out in the same manner as for polygonal domains, preserving the quasi-uniformity for $h \downarrow 0$ (cf. Definition A2.2.2).

A triangulation with these properties can always be performed, moreover, it can be done uniformly on the majority of the entire area of the domain $\Omega$, except near the boundary $\Gamma_h$.

A2.2.7 Remark. Sometimes it is more convenient to build a domain $\Omega_h$ with the property $\Omega_h \supset \Omega$ maintaining the conditions (ii)-(iv) above.

The spaces $V_h$, which form an inner approximation of $V$, can be obtained by a suitable continuation of the functions

$$v_h(x) := \sum_{i \in I_h} \alpha_i \varphi_i(x) \quad \text{on } \Omega \setminus \bar{\Omega}_h$$

or

$$v_h(x) := \sum_{i \in I^0_h} \alpha_i \varphi_i(x) \quad \text{on } \Omega \setminus \bar{\Omega}_h.$$ 

For $V := H_0^1(\Omega)$ the space $V^0_h$ can be constructed by means of the following continuation of the functions $v_h$:

$$v_h(x) := \begin{cases} \sum_{i \in I_h} \alpha_i \varphi_i(x) & \text{on } \bar{\Omega}_h, \\ 0 & \text{on } \Omega \setminus \bar{\Omega}_h. \end{cases}$$

In more general cases these constructions can face serious difficulties. For different approaches see Skarpini and Vivaldy [360], Marchuk and Agoshkov [286] and the references quoted there.

A2.2.8 Proposition. For a boundary of class $C^2$ estimate (A2.2.12) remains true for $u \in H^2(\Omega) \cap H_0^1(\Omega)$ if the spaces $V = H^1(\Omega)$ and $V^0_h$ are constructed as above.

A2.2.9 Remark. Estimate (A2.2.12) holds also true for some other techniques of the continuation or restriction of $v_h$ from $\Omega_h$ to $\Omega$, including the case $\Omega_h \supset \Omega$.

Due to (A2.2.12) we obtain immediately that

$$\|u - u_I\|_{0, \Omega} \leq c h^2 |u|_{2, \Omega}$$

and in case $\Omega_h \supset \Omega$

$$\|u - u_I\|_{1, \Omega} \leq \tilde{c} h^2 |u|_{2, \Omega}.$$
A2. APPROXIMATION TECHNIQUES AND ESTIMATES OF SOLUTIONS

Usually, the solutions of variational inequalities arising from problems with differential operators of second order (for short: variational inequalities of second order) do not belong to the space \( C^2(\bar{\Omega}) \) (and sometimes even not to \( H^2(\Omega) \)) whatever a smoothness of data we have. Therefore, in order to solve such variational inequalities, it makes sense to restrict ourselves to the implementation of piece-wise affine basis functions, because in that case the use of smoother functions does not improve the estimates of the approximation, but it increases the expense of calculations.

For linear variational inequalities of second order, when the solution has a better smoothness, and especially for linear equations and variational inequalities with differential operators of higher order, smoother functions can be constructed. Here, curved "triangular" elements are implemented in order to obtain a sufficiently exact approximation along the boundary (cf. CIARLET [75], ODEN [310], KORNEEV [245]).

A2.3 FEM-approximation of feasible sets

Let us return to the variational problem (A1.5.11) and consider an example-setting finite element approximation for the feasible set of this problem.

A2.3.10 Definition. The sequence \( \{K_h\} \) of closed convex sets is called an inner approximation of the feasible set \( K \) if

(i) \( K_h \subset V_h \) and for every \( v \in K \) there exist elements \( v_h \in K_h \) such that

\[ \|v_h - v\|_V \to 0 \quad \text{for} \quad h \downarrow 0; \]

(ii) \( u_h \in K_h \) and \( u_h \rightharpoonup u \) in \( V \) imply that \( u \in K \).

The notion "inner" does not mean \( K_h \subset K \). Similarly to the inner approximation of the space \( V \), it is sufficient to verify condition (i) for elements \( v \in K \cap \tilde{V} \), with \( \tilde{V} \) a dense set in \( V \).

The conditions in Definition A2.3.10 are fulfilled for the Problems (A2.1.3) and (A2.1.6) with

\[ K_h := \left\{ u_h = \sum_{i \in I_0^h} \alpha_i \varphi_i(x) : \alpha_i \geq \varphi_i(P_i) \quad \forall \, i \in I_0^h \right\} \quad \text{(A2.3.21)} \]

and

\[ K_h := \left\{ u_h = \sum_{i \in I_1^h} \alpha_i \varphi_i(x) : \alpha_i \geq \varphi_0(P_i) \quad \forall \, i \in I_1^h \right\}, \quad \text{(A2.3.22)} \]

(\( I_0^h \) and \( I_1^h \) according to (A2.2.10)) respectively, if the system of triangulations is quasi-uniform. Indeed, for each \( v \in K \cap \tilde{V} \), with \( \tilde{V} = \mathcal{D}(\bar{\Omega}) \) in the case of Problem (A2.1.3) and \( \tilde{V} = \mathcal{C}\mathcal{C}^\infty(\bar{\Omega}) \) for Problem (A2.1.6), the interpolant \( v_{ih} \in V_h \) belongs to \( K_h \). Due to Theorem A2.2.5 it holds \( \|v_{ih} - v\| \to 0 \) for \( h \downarrow 0 \), i.e., condition (i) in Definition A2.3.10 is fulfilled.

To verify condition (ii) in Definition A2.3.10 for Problem (A2.1.3) we note that the inclusion \( u_h \in K_h \) implies \( u_h \geq \varphi_{ih} \) and, consequently,

\[ \langle \ell, u_h - \varphi_{ih} \rangle \geq 0 \quad \text{for any} \quad \ell \in V'_+, \quad \text{(A2.3.23)} \]
Since \( \| \varphi_{Ih} - \varphi \|_V \to 0 \) and \( u_h \to u \), inequality (A2.3.23) leads to
\[
\langle \ell, u - \varphi \rangle \geq 0 \quad \text{for any } \ell \in V_1',
\]
i.e., \( u \in K \).

Now, we consider the situation for Problem (A2.1.6). Regarding that \( \varphi_0 \) is the trace of \( \varphi \in H^2(\Omega) \hookrightarrow C(\bar{\Omega}) \) and \( \| \varphi - \varphi_{Ih} \|_{C(\bar{\Omega})} \to 0 \), there exists a sequence of numbers \( c_h \downarrow 0 \) such that \( u_h + c_h \in K \). Since \( u_h \to u \), we obtain \( u_h + c_h \to u \) and, due to the weak closedness of \( K \), the inclusion \( u \in K \) holds.

It would be desirable to obtain approximations of \( K \) satisfying simultaneously \( K_{Ih} \subset K \). Unfortunately, this cannot be performed effectively. For instance, in Problem (A2.1.3) we should take into account the constraints
\[
v_h(x) = \sum_{i \in I_0^h} \alpha_i \varphi_i(x) \geq \varphi(x) \quad \text{a.e. in } \Omega,
\]
i.e., each set \( K_h \) has to be described by an infinite number of restrictions.

### A2.4 FEM-approximation of variational problems

Let us suppose that the objective functional \( J \) for the convex variational problem (A1.5.11) is continuous on \( V = H^1(\Omega) \) or \( V = H^1_0(\Omega) \), respectively. Implementing a finite element method, we are looking for an approximate solution of Problem (A1.5.11) in the form
\[
u_h(x) := \sum_{i} \alpha_i \varphi_i(x), \quad \text{(A2.4.24)}
\]
summing up for \( i \in I_h \) or \( i \in I_0^h \), in dependence of the choice of the space \( V \).

Hence, for \( h \downarrow 0 \) a sequence of approximate problems
\[
\min \{ J(u_h) : u_h \in K_h \} \quad \text{(A2.4.25)}
\]
is obtained.

Assuming that \( J = J_1 + J_2 \), with \( J_1, J_2 \) convex and \( J_1 \) Gâteaux-differentiable on \( V \), Problem (A2.4.25) corresponds to the variational inequality
\[
u_h \in K_h : \quad \langle J_1'(u_h), v_h - u_h \rangle + J_2(v_h) - J_2(u_h) \geq 0 \quad \forall v_h \in K_h. \quad \text{(A2.4.26)}
\]

Now, we prove convergence of a finite element method for variational inequalities with strongly monotone operators, i.e., the objective functional \( J \) in (A2.4.25) is supposed to be strongly convex. Here we assume that the Problems (A2.4.25) are solved approximately.

#### A2.4.11 Theorem

Suppose that the family of spaces \( \{ V_h \} \) forms an inner approximation of the space \( V \), and \( \{ K_h \} \) approximates \( K \) in the sense of Definition A2.3.10. Moreover, let \( J \) be a continuous and strongly convex functional on \( V \) (with constant \( \kappa \)), and an element \( u_h' \) let be computed such that
\[
\rho(q(u_h'), \partial J(\bar{u}_h)) \leq \epsilon_h, \quad \epsilon_h \downarrow 0, \quad \text{(A2.4.27)}
\]
with
\[
\bar{u}_h := \arg \min_{u_h \in K_h} J(u_h) \quad \text{and some } q(u_h') \in \partial J(u_h').
\]
Then
\[ \lim_{h \downarrow 0} \| u_h^\varepsilon - \bar{u} \| = 0, \]
with \( \bar{u} = \arg \min_{u \in K} J(u) \).

**Proof:** Using estimate (A1.5.30), we conclude from (A2.4.27)
\[ \| u_h^\varepsilon - \bar{u}_h \| \leq \frac{\epsilon h}{2\kappa}. \]
Hence, it suffices to show that \( \lim_{h \downarrow 0} \| u_h^\varepsilon - \bar{u} \| = 0 \).

For arbitrary \( v \in K \), due to condition (i) in Definition A2.3.10, one can choose points \( v_h \in K_h \) such that \( \lim_{h \downarrow 0} \| v_h - v \| = 0 \).

On account of Proposition A1.5.31 the inequality
\[ J(v_h) - J(\bar{u}_h) \geq (q(\bar{u}_h), v_h - \bar{u}_h) + \kappa \| v_h - \bar{u}_h \|^2 \]
is fulfilled for all \( q(\bar{u}_h) \in \partial J(\bar{u}_h) \) and with regard to Remark A1.5.35
\[ (q(\bar{u}_h), v_h - \bar{u}_h) \geq 0 \]
is valid for some \( q(\bar{u}_h) \in \partial J(\bar{u}_h) \). Consequently,
\[ \kappa \| v_h - \bar{u}_h \|^2 \leq J(v_h) - J(\bar{u}_h) \leq J(v_h) - \min_{v \in V} J(v), \quad (A2.4.28) \]
\[ \sqrt{\kappa} \| u_h^\varepsilon - v \| \leq \left( J(v_h) - \min_{v \in V} J(v) \right)^{1/2} + \sqrt{\kappa} \| v_h - v \|. \]
Hence, in view of \( \lim_{h \downarrow 0} \| v_h - v \| = 0 \) and the continuity of \( J \), the sequence \( \{ u_h^\varepsilon \} \) is bounded and we can choose a subsequence that converges weakly to some \( w \in V \). Because \( J \) is convex and continuous it is also a wsc functional on \( V \) and, with regard to \( J(\bar{u}_h) \leq J(v_h) \), we obtain \( J(w) \leq J(v) \).

But condition (ii) in Definition A2.3.10 implies that \( w \in K \). Since \( v \) is an arbitrary element of \( K \), one can conclude that \( w \) is a solution of the original problem. Hence, \( w = \bar{u} \), and the sequence \( \{ u_h^\varepsilon \} \) converges weakly to \( \bar{u} \).

Using condition (i) in Definition A2.3.10, we choose a sequence \( \{ v_h \} \), \( v_h \in K_h \), with \( \lim_{h \downarrow 0} \| v_h - \bar{u} \| = 0 \). Then, (A2.4.28) leads to \( \lim_{h \downarrow 0} \| u_h^\varepsilon - v_h \| = 0 \), consequently, \( \lim_{h \downarrow 0} \| u_h^\varepsilon - \bar{u} \| = 0 \). \( \square \)

On the next pages, we explain the discretization procedure for the above introduced variational inequalities: Dirichlet Problem (A2.1.1), Obstacle Problem (A2.1.3) as well as the Boundary-obstacle Problem (A2.1.6) in more detail. Here the approximate solutions of the first two problems are sought in the form
\[ u_h(x) := \sum_{i \in J_h^0} \alpha_i \varphi_i(x) \]
and this leads to the finite-dimensional problems
\[ \phi_h(\alpha) := \frac{1}{2} \alpha^T A_h \alpha - f_h^T \alpha \to \min, \quad \alpha \in K_h. \quad (A2.4.29) \]

In case of the Dirichlet Problem (A2.1.1) we have
\[ K_h := \mathbb{R}^n, \quad (A2.4.30) \]
and of the Obstacle Problem (A2.1.3)

\[ K_h := \{ \alpha \in \mathbb{R}^n : \alpha_i \geq \varphi(P_i), \quad i \in I^0_h \}, \quad (A2.4.31) \]

\( (n = |I^0_h| \) is the cardinality of the set \( I^0_h \), described in (A2.2.10)).

For convenience of the description we take \( I^0_h = \{ 1, \ldots, n \} \) and denote

\[ A_h := (a_{ij})_{i,j=1}^n, \quad a_{ij} = a(\varphi_i, \varphi_j); \]
\[ f_h := (f_1, \ldots, f_n)^T, \quad f_i = \langle f, \varphi_i \rangle. \]

For the Boundary-obstacle Problem (A2.1.6) the approximate solutions have the form

\[ u_h = \sum_{i \in I_h} \alpha_i \varphi_i(x), \]

thus, we obtain the finite-dimensional minimization problems

\[ \phi_h(x) := \frac{1}{2} \alpha^T A_h \alpha - \int_h^T \alpha \to \min, \quad \alpha \in K_h, \quad (A2.4.32) \]

with

\[ K_h := \{ \alpha \in \mathbb{R}^n : \alpha_i \geq \varphi_0(P_i), \quad i \in I^0_h \}, \]
\[ I^0_h := \{ 1, \ldots, n \}, \quad A_h := (a_{ij})_{i,j=1}^n, \quad f_h := (f_1, \ldots, f_n)^T \quad (A2.4.33) \]

and \( a_{ij}, f_i \) defined as above.

Let \( P_{T_k} := (x_{T_k}, y_{T_k}) \) \((k = 1, 2, 3)\) be the vertices of some triangle \( T \in \mathcal{T}_h \).

Then an affine function \( \eta_T(x, y) \) having the value \( \alpha_{T_k} \) at \( P_{T_k} \) can be described by

\[ \eta_T(x, y) := \frac{1}{\sigma_T} \sum_{k=1}^3 (\beta_{T_k} x + \gamma_{T_k} y + \delta_{T_k}) \alpha_{T_k}, \quad (A2.4.34) \]

with

\[ \beta_{T_1} = y_{T_2} - y_{T_3}, \quad \gamma_{T_1} = x_{T_3} - x_{T_2}, \quad \delta_{T_1} = x_{T_2} y_{T_3} - x_{T_3} y_{T_2}, \]
\[ \beta_{T_2} = y_{T_3} - y_{T_1}, \quad \gamma_{T_2} = x_{T_2} - x_{T_3}, \quad \delta_{T_2} = x_{T_3} y_{T_1} - x_{T_1} y_{T_3}, (A2.4.35) \]
\[ \beta_{T_3} = y_{T_1} - y_{T_2}, \quad \gamma_{T_3} = x_{T_2} - x_{T_1}, \quad \delta_{T_3} = x_{T_1} y_{T_2} - x_{T_2} y_{T_1}, \]
\[ \sigma_T = \sum_{k=1}^3 \delta_{T_k}, \text{ i.e., } |\sigma_T| = 2m(T), \text{ with } m(T) \text{ the area of the triangle } T. \]

Taking

\[ \theta_T(x, y) = \begin{cases} 1 & \text{if } (x, y) \in T, \\ 0 & \text{if } (x, y) \notin T, \end{cases} \]

we obtain the function

\[ \tilde{u}_h(x, y) := \sum_{T \in \mathcal{T}_h} \frac{1}{\sigma_T} \left( \sum_{k=1}^3 (\beta_{T_k} x + \gamma_{T_k} y + \delta_{T_k}) \alpha_{T_k} \right) \theta_T(x, y), \]

which coincides with \( u_h(x, y) \) on the domain \( \Omega \), except for the set \( \tilde{\Gamma} \) of sides of all triangles in \( \mathcal{T}_h \).
In order to describe the objective function of the approximate Problems (A2.1.1) and (A2.1.3), we have to calculate for any \((x, y) \in \tilde{\Gamma}\)
\[
\frac{\partial u_h}{\partial x}(x, y) = \sum_{T \in \mathcal{T}_h} \frac{1}{\sigma_T} \left( \sum_{k=1}^{3} \beta_{T_k} \alpha_{T_k} \right) \theta_T(x, y),
\]
\[
\frac{\partial u_h}{\partial y}(x, y) = \sum_{T \in \mathcal{T}_h} \frac{1}{\sigma_T} \left( \sum_{k=1}^{3} \gamma_{T_k} \alpha_{T_k} \right) \theta_T(x, y).
\]
(A2.4.36)

Therefore, in that case the objective functions have the form
\[
\phi_h(\alpha) := \frac{1}{8} \sum_{T \in \mathcal{T}_h} \frac{1}{m(T)} \left[ \left( \sum_{k=1}^{3} \beta_{T_k} \alpha_{T_k} \right)^2 + \left( \sum_{k=1}^{3} \gamma_{T_k} \alpha_{T_k} \right)^2 \right]
- \sum_{i \in \mathcal{I}_h} \alpha_i \int_{\text{supp} \varphi_i} f(x, y) \varphi_i(x, y) dx dy,
\]
(A2.4.37)

\((\alpha_i = 0 \text{ if } i \in \mathcal{I}_h'\), and the feasible sets \(K_h\) are defined by (A2.4.30) and (A2.4.31), respectively.

Analogously, for Problem (A2.1.6) one obtains
\[
\bar{\phi}_h(\alpha) := \frac{1}{8} \sum_{T \in \mathcal{T}_h} \frac{1}{m(T)} \left[ \left( \sum_{k=1}^{3} \beta_{T_k} \alpha_{T_k} \right)^2 + \left( \sum_{k=1}^{3} \gamma_{T_k} \alpha_{T_k} \right)^2 \right]
- \sum_{i \in \mathcal{I}_h} \alpha_i \int_{\text{supp} \varphi_i} f(x, y) \varphi_i(x, y) dx dy,
\]
(A2.4.38)

and \(K_h\) is defined by (A2.4.33).

Note that
\[
\int_{\text{supp} \varphi} f(x, y) \varphi_i(x, y) dx dy = \sum_{T \in \mathcal{T}_i} \int_T f(x, y) (b_{iT}x + c_{iT}y + d_{iT}) dx dy,
\]
with \(\mathcal{T}_i\) the set of triangles of \(\mathcal{T}_h\) having a common vertex \(P_i\). Here the coefficients \(b_{iT}, c_{iT}, d_{iT}\) are defined such that the function
\[
b_{iT}x + c_{iT}y + d_{iT}
\]
is equal to 1 in \(P_i\), and 0 in all the other vertices of \(T\).

Now, using the uniform triangulation, we give an illustrative description of the approximation of Problem (A2.1.3). The rectangular domain is drawn in Figure A2.4.3

where the mesh-sizes of this rectangular domain are given by \(h_x = \frac{a}{N_x}\) and \(h_y = \frac{b}{N_y}\), \((N_x, N_y\) are integers). The vertex \((i, j)\) has the coordinates \((ih_x, jh_y)\).

In this case the solution is sought in the form
\[
u_h(x, y) := \sum_{j=0}^{N_y} \sum_{i=0}^{N_x} u_{ij}(x, y),
\]
(A2.4.39)
with $u_{ij} = u_{i0} = u_{N_x,j} = u_{iN_y} = 0$ for all $i, j$; $\varphi_{ij}$ is a usual piece-wise affine basis function with

$$\varphi_{ij}(ih_x, jh_y) = 1, \quad \varphi_{ij}(kh_x, lh_y) = 0 \quad \text{for} \ (k, l) \neq (i, j).$$

With a fixed vertex $(i, j), (0 \leq i \leq N_x - 1, \ 0 \leq j \leq N_y - 1)$, two triangles $T'$ and $T''$ are connected as shown in Figure A2.4.3. Using the formulas (A2.4.35) for the description of the function $u_h$ on the triangles $T'$ and $T''$, we obtain the following representation of the derivatives of $u_h$ on $T' \cup T''$:

$$\frac{\partial u_h}{\partial x}(x, y) = \frac{1}{\sigma_{T'}} (h_y u_{i,j} - h_y u_{i+1,j}) \theta_{T'}(x, y)$$

$$+ \frac{1}{\sigma_{T''}} (h_y u_{i+1,j+1} - h_y u_{i,j+1}) \theta_{T''}(x, y),$$

$$\frac{\partial u_h}{\partial y}(x, y) = \frac{1}{\sigma_{T'}} (-h_x u_{i+1,j+1} + h_x u_{i+1,j}) \theta_{T'}(x, y)$$

$$+ \frac{1}{\sigma_{T''}} (-h_x u_{i,j} + h_x u_{i,j+1}) \theta_{T''}(x, y),$$

with $|\sigma_{T'}| = |\sigma_{T''}| = h_x h_y$.

Hence, in order to describe the quadratic term of the objective function, we have to calculate

$$\frac{1}{2} \int_{T' \cup T''} \left[ \left( \frac{\partial u_h}{\partial x} \right)^2 + \left( \frac{\partial u_h}{\partial y} \right)^2 \right] dx dy =$$

$$= \frac{1}{4} h_x h_y \left( \frac{(u_{i+1,j} - u_{i,j})^2}{h_x^2} + \frac{(u_{i+1,j+1} - u_{i+1,j})^2}{h_y^2} \right)$$

$$+ \left( \frac{(u_{i+1,j+1} - u_{i,j+1})^2}{h_x^2} + \frac{(u_{i,j+1} - u_{i,j})^2}{h_y^2} \right)$$
A2. APPROXIMATION TECHNIQUES AND ESTIMATES OF SOLUTIONS

\[ \frac{1}{2} \int_{\Omega} \left[ \left( \frac{\partial u_h}{\partial x} \right)^2 + \left( \frac{\partial u_h}{\partial y} \right)^2 \right] \, dx \, dy = \]
\[= \frac{1}{4} h_x h_y \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} \left( \frac{(u_{i+1,j} - u_{i,j})^2}{h_x^2} + \frac{(u_{i,j+1} - u_{i,j})^2}{h_y^2} \right) \]
\[+ \frac{1}{4} h_x h_y \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} \left( \frac{(u_{i+1,j+1} - u_{i,j+1})^2}{h_x^2} + \frac{(u_{i+1,j+1} - u_{i,j+1})^2}{h_y^2} \right) \]
\[+ \frac{1}{4} h_x h_y \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} \left( \frac{(u_{i+1,j} - u_{i,j})^2}{h_x^2} + \frac{(u_{i,j+1} - u_{i,j})^2}{h_y^2} \right). \]

Taking into account the values \( u_{i,j} \) on the boundary, we obtain

\[ \frac{1}{2} \int_{\Omega} \left[ \left( \frac{\partial u_h}{\partial x} \right)^2 + \left( \frac{\partial u_h}{\partial y} \right)^2 \right] \, dx \, dy = \]
\[= \frac{1}{4} h_x h_y \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} \left( \frac{(u_{i+1,j} - u_{i,j})^2}{h_x^2} + \frac{(u_{i,j+1} - u_{i,j})^2}{h_y^2} \right). \]

The linear part of the objective functional has the form

\[ - \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y-1} u_{ij} \int_{\text{supp} \varphi} f(x,y) \varphi_{ij}(x,y) \, dx \, dy, \]

in which, according to Figure A2.4.4, we have

Figure A2.4.4:

\[ \varphi_{ij}(x,y) := \begin{cases} - \frac{y}{h_y} + \frac{x}{h_x} + (j + 1) & \text{on triangle 1,} \\ \frac{x}{h_x} - \frac{y}{h_y} + 1 + (j - 1) & \text{on triangle 2,} \\ \frac{x}{h_x} + (1 - i) & \text{on triangle 3,} \\ \frac{y}{h_y} + (1 - j) & \text{on triangle 4,} \\ - \frac{x}{h_x} + \frac{y}{h_y} + 1 + (i - j) & \text{on triangle 5,} \\ - \frac{x}{h_x} + (i + 1) & \text{on triangle 6,} \end{cases} \]

and

\[ K_h := \{ u = (u_{ij})_{i=0,j=0}^{N_x,N_y} : u_{0j} = u_{i0} = u_{N_x,j} = u_{iN_y} = 0, u_{i,j} \geq \varphi(ih_x,jh_y) \forall i,j \}. \]
Note that in the case of a uniform triangulation the matrices $A_h$ and $\bar{A}_h$ have a regular five-diagonal structure, and for $h_x = h_y$ their elements do not depend on $h$. Concerning such matrices, see Young [417].

If, for example, $a = b = 1$ and $h_x = h_y = \tilde{h} = \frac{1}{N}$, then the matrix $A_h$ has the eigenvalues

$$\lambda_{k,l} = 8 \left( \sin^2 \left( \frac{k}{2} \tilde{h} \pi \right) + \sin^2 \left( \frac{1}{2} \tilde{h} \pi \right) \right), \quad k, l = 1, ..., N - 1.$$ 

Hence, its condition number is

$$\text{cond}(A_h) = \frac{\max_{k,l} \lambda_{k,l}}{\min_{k,l} \lambda_{k,l}} = \frac{\sin^2 \frac{N-1}{2N} \pi}{\sin^2 \frac{1}{2N} \pi} \approx 4 \tilde{h}^{-2} \pi^2 = 8 \tilde{h}^{-2} \pi^2$$

(recall that $h = \sqrt{h_x^2 + h_y^2}$, i.e. $h = \sqrt{2} \tilde{h}$).

In a more general situation, using quasi-uniform triangulations, we obtain for the Dirichlet Problem (A2.1.1) and the Obstacle Problem (A2.1.3) that

$$h^2 m \|v\|_{\mathbb{R}^n} \leq v^T A_h v \leq M \|v\|_{\mathbb{R}^n} \quad \forall v \in \mathbb{R}^n,$$

with $m > 0$ and $M$ independent of $h$, consequently,

$$\text{cond}(A_h) \leq \frac{M}{m}$.

Numerical experiments and the theoretical analysis enable us to consider these matrices as well-conditioned.

In the Boundary-obstacle Problem (A2.1.6) the set $K$ contains constant functions and for $u = c = \text{const}$, using the formulas (A2.4.35), (A2.4.36), we immediately obtain

$$\alpha^T \bar{A}_h \alpha = 0 \quad \text{with } \alpha = (c, ..., c) \in \mathbb{R}^n.$$

Hence, due to the Rayleigh principle,

$$\text{cond}^{-1}(A_h) = \frac{\min_{\|\alpha\| \neq 0} \alpha^T \bar{A}_h \alpha}{\max_{\|\alpha\| \neq 0} \alpha^T A \alpha} = 0.$$

Of course, degeneration of the approximate problems is a consequence of the degeneration of the quadratic form $a(u, u)$ on $V$.

We also need the following result about the distance between the solution $u$ of the Obstacle Problem (A2.1.3) and the finite-element solutions $u_h$ of the approximate problems (A2.4.29), (A2.4.31).

**A2.4.12 Proposition.** (Ciarlet [75])

Assume that the solution $u$ of Problem (A2.1.3) belongs to $H^2(\Omega)$. Then there exists a constant $c(u, f, \varphi)$ independent of $h$ such that for every quasi-uniform family of triangulations

$$\|u - u_h\|_{1,\Omega} \leq c(u, f, \varphi) h.$$  \hfill (A2.4.40)

Assuming regularity of the Obstacle Problem, this estimate is also true if $\Omega$ is a convex set with a boundary of the class $C^2$ (cf. Brezis and Stampacchia [53]).

Obviously, as a conclusion of the proposition above, a similar estimate

$$\|u - u_h\|_{1,\Omega} \leq c(u, f) h$$

can be obtained for the Dirichlet Problem (A2.1.1).
A3  Selected Methods of Convex Minimization

In this section standard methods of convex unconstrained and constrained minimization are briefly described in form of conceptual algorithms. Their selection depends on specific properties of the investigated variational problems. In particular, the variational inequalities occurring in applications have usually simple constraints and their discretization leads to finite-dimensional convex (often quadratic) problems of minimization with linear constraints.

Moreover, some of the methods considered here are studied in the main part of the book in connection with ill-posed finite-dimensional problems.

The majority of the convergence results is described without proofs. Only arguments, which help to understand the behavior of the procedures and methods applied, will be given. Concerning proofs, the reader is referred to the literature.

Our aim is to consider numerical methods as a whole, starting with discretization and including elementary procedures within the optimization methods. We omit linear programming methods, nevertheless, they are used in particular for solving the auxiliary problems in algorithms of feasible directions. Quadratic programming methods will be sketched only in a special setting.

Taking into account the object of our efforts, among numerous publications related to convex optimization methods, the books of Polyak [330], Fletcher [117], Bazaraa, Sheraly and Shetty [36], Bertsekas [41], Dennis and Schnabel [90], Gill, Murray and Wright [128], Mangasarian [285] and Geiger and Kanzow [126] are quite suitable. The monograph of Neittaanmaki, Sprekels and Tiba [304] is devoted, in particular, to optimization methods of elliptic systems.

A3.1  Notions, definitions and convergence statements

I order to investigate numerical methods we need the following notions which characterize the behavior of iterative methods.

A3.1.1  Definition. 

(i) A sequence \( \{v^k\} \) is convergent to a set \( Q \) if

\[
\lim_{k \to \infty} \rho(v^k, Q) = 0.
\]

(ii) A sequence \( \{v^k\} \) is convergent to the optimal value (especially of Problem (A1.5.11)) if

\[
\lim_{k \to \infty} J(v^k) = \min_{u \in K} J(u) \quad \text{and} \quad \lim_{k \to \infty} \rho(v^k, K) = 0.
\]

♦

A3.1.2  Remark. In this case \( \{v^k\} \) is called a generalized minimizing sequence. The notion ”generalized” will be omitted if, starting with some iteration number, \( v^k \in K \) holds true.

♦

A3.1.3  Definition. A sequence \( \{v^k\} \) converges to \( v \)
(i) *linearly* (with geometric progression), if for some \( c \) and \( q \in (0,1) \)
\[
\|v^k - v\| \leq cq^k, \quad \forall \ k \in \{1, 2, \ldots \} = \mathbb{N};
\]

(ii) *super-linearly* if
\[
\|v^k - v\| \leq cq_1q_2\ldots q_k, \quad \forall \ k \in \mathbb{N}, \ q_k \in (0,1), \ \lim_{k \to \infty} q_k = 0;
\]

(iii) *quadratically* if \( q \in (0,1) \) and
\[
\|v^k - v\| \leq cq^{2k}, \quad \forall \ k \in \mathbb{N}.
\]

Usually, the real possibility to estimate the values \( \|v^k - v\| \) is more important, i.e., knowledge of the particular values of \( c \) and \( q \) is required. In that case it makes sense to speak about different types of rates of convergence for different large intervals of the iterations index \( k \). Sometimes we make use only of the knowledge of the existence of such constants without knowing their real values.

In the following we remind on some results about the convergence of sequences of numbers, which are useful for proving of convergence results.

**A3.1.4 Lemma.** Let \( \{\nu_k\} \) be a sequence in \( \mathbb{R} \), bounded from below such that
\[
\nu_{k+1} \leq \nu_k + \beta_k, \quad \beta_k \geq 0, \quad \sum_{k=0}^{\infty} \beta_k < \infty. \tag{A3.1.1}
\]

Then \( \{\nu_k\} \) is convergent.

**Proof:** From the first inequality in (A3.1.1) we get
\[
\nu_{i+1} \leq \nu_0 + \sum_{k=0}^{\infty} \beta_k, \quad \forall \ i.
\]

Hence, \( \{\nu_k\} \) is bounded. Let \( \{\nu_{k_j}\} \) be a convergent subsequence with limit \( \bar{\nu} \). For each \( \gamma > 0 \) there exists a number \( j_0 \) such that for \( j \geq j_0 \)
\[
\nu_{k_j} < \bar{\nu} + \frac{\gamma}{2} \quad \text{and} \quad \sum_{k=k_j}^{\infty} \beta_k < \frac{\gamma}{2}.
\]

Due to (A3.1.1), for \( i > k_{j_0} \), the inequality
\[
\nu_i \leq \nu_{k_{j_0}} + \sum_{k=k_{j_0}}^{\infty} \beta_k
\]
holds, hence, \( \nu_i < \bar{\nu} + \gamma \).

Since \( \gamma \) is an arbitrary chosen positive number, the sequence \( \{\nu_k\} \) cannot have a limit greater than \( \bar{\nu} \), and because \( \{\nu_{k_j}\} \) is an arbitrary convergent subset, \( \bar{\nu} \) must be the unique limit. \( \square \)

The following results can be found, for instance by Polyak [330].
A3. SELECTED METHODS OF CONVEX MINIMIZATION

A3.1.5 Lemma. Let
\[ \nu_{k+1} \leq q \nu_k + \mu, \quad 0 \leq q < 1, \mu > 0. \] (A3.1.2)

Then it holds
\[ \nu_k \leq \frac{\mu}{1-q} + \left( \nu_0 - \frac{\mu}{1-q} \right) q^k. \] (A3.1.3)

A3.1.6 Lemma. Let
\[ \nu_{k+1} \leq q_k \nu_k + \mu_k \] (A3.1.4)

with
\[ 0 \leq q_k < 1, \mu_k \geq 0, \sum_{k=1}^{\infty} (1-q_k) = \infty, \frac{\mu_k}{1-q_k} \to 0. \]

Then
\[ \lim_{k \to \infty} v_k \leq 0. \]

In particular, if \( v_k \geq 0 \), then \( v_k \to 0 \).

But, if in (A3.1.4)
\[ q_k \equiv q < 1, \mu_k \to 0, v_k \geq 0, \]

then \( v_k \to 0 \).

A3.1.7 Lemma. Let \( \{v_k\} \) be a sequence of positive numbers satisfying
\[ \nu_{k+1} \leq (1 + \mu_k) v_k + \beta_k, \] (A3.1.5)

with
\[ \mu_k \geq 0, \beta_k \geq 0, \sum_{k=0}^{\infty} \mu_k < \infty, \sum_{k=0}^{\infty} \beta_k < \infty. \]

Then \( v_k \to v \geq 0 \).

A3.1.8 Lemma. Let \( \{v_k\} \) be a sequence of positive numbers satisfying
\[ \nu_{k+1} \leq (1 - \mu_k) v_k + \beta_k, \] (A3.1.6)

with
\[ 0 \leq \mu_k < 1, \beta_k \geq 0, \sum_{k=1}^{\infty} \mu_k = \infty, \frac{\beta_k}{\mu_k} \to 0. \]

Then \( v_k \to 0 \).

A3.2 Methods of unconstrained minimization

A3.2.1 Minimization of convex univariate functions

We consider two variants of bisection methods which computes an approximate minimizer \( u^* \) of a convex function \( J \) on an interval \([a, b] \). The first method makes use of function values only, whereas the second one is applicable if \( J \) is differentiable.

A3.2.9 Algorithm. (Bisection algorithm)
CHAPTER 11. APPENDIX

S0: Choose \( \delta > 0 \) (desirable accuracy of the minimizer \( u^* \));

S1: calculate \( J \left( \frac{1}{2}(a+b-\delta) \right) \) and \( J \left( \frac{1}{2}(a+b+\delta) \right) \);

S2: a) if \( J \left( \frac{1}{2}(a+b-\delta) \right) < J \left( \frac{1}{2}(a+b+\delta) \right) \), continue on the interval \( [a, \frac{1}{2}(a+b+\delta)] \);

   b) if \( J \left( \frac{1}{2}(a+b-\delta) \right) > J \left( \frac{1}{2}(a+b+\delta) \right) \), continue on the interval \( \left[ \frac{1}{2}(a+b-\delta), b \right] \);

S3: stop if the obtained interval is less than \( \delta \).

S4: if \( J \left( \frac{1}{2}(a+b-\delta) \right) = J \left( \frac{1}{2}(a+b+\delta) \right) \), then \( u^* := \frac{1}{2}(a+b) \).

\( \Diamond \)

A3.2.10 Algorithm. (Bisection algorithm using derivatives)

S0: Choose \( \delta > 0 \) (desirable accuracy of the minimizer);

S1: calculate \( J'(a) \) and \( J'(b) \);

S2: if \( J'(a) \geq 0 \), set \( u^* := a \); if \( J'(b) \leq 0 \) set \( u^* := b \);

   otherwise, compute \( J' \left( \frac{1}{2}(a+b) \right) \);

   if \( J' \left( \frac{1}{2}(a+b) \right) < 0 \), continue on the interval \( \left[ \frac{1}{2}(a+b), b \right] \);

   if \( J' \left( \frac{1}{2}(a+b) \right) > 0 \), continue on \( \left[ a, \frac{1}{2}(a+b) \right] \);

S3: stop if the obtained interval is less than \( \delta \);

S4: if \( J' \left( \frac{1}{2}(a+b) \right) = 0 \), then \( u^* := \frac{1}{2}(a+b) \).

\( \Diamond \)

A3.2.2 Line search methods

Our investigations are mainly concerned with smooth variational problems or with problems permitting to apply efficient smoothing procedures. Therefore, we will focus mainly on techniques of the differentiable optimization. Only in connection with variational inequalities with multi-valued operators we will meet non-smooth techniques.

Now, we consider unconstrained minimization methods for convex, differentiable functions, i.e. we deal with problems

\[
\min \{ J(x) : x \in \mathbb{R}^n \},
\]

assuming, if necessary, the existence of derivatives of higher order.

The common structure of these methods is as follows. Let the iterate \( u^k \) be calculated at the \( k \)-th step, then \( u^{k+1} \) is determined by moving from \( u^k \) in a suitable direction \( p^k \) such that

\[
u^{k+1} := u^k + \alpha_k p^k \quad k = 1, 2, \ldots .
\]

The value \( \alpha_k \) is called step-length parameter.

Particular methods differ in the choice of the direction vector \( p^k \) and the step-length \( \alpha_k \). We consider only methods in which

\[
\langle p^k , \nabla J(u^k) \rangle < 0 , \quad \text{for } \nabla J(u^k) \neq 0.
\]
If
\[ p_k := -\nabla J(u^k) \quad \text{and} \quad \alpha_k := \arg \min_{\alpha \geq 0} J(u^k + \alpha p_k), \]
(A3.2.7)
then we deal with a method of steepest descent, and the choice of the Newton direction
\[ p_k := -[\nabla^2 J(u^k)]^{-1} \nabla J(u^k), \quad \alpha_k := 1, \]
(A3.2.8)
leads to the classical Newton method. Since, in general, the calculation of \( u^{k+1} \) is difficult (especially with the Newton direction), there exists a number of modifications of these methods. Most of them can be embedded in the following type of algorithms:

A3.2.11 Method. (Gradient-type method)

Data: \( \mu \in [-1, 0) \), \( u^0 \in \mathbb{R} \);

S0: Set \( k := 0 \);

S1: if \( \nabla J(u^k) = 0 \) stop, \( u^k \) is a minimizer;
otherwise, determine \( p_k \neq 0 \) such that
\[ \langle p_k, \nabla J(u^k) \rangle \leq \mu \| p_k \| \| \nabla J(u^k) \| ; \]
(A3.2.9)

S2: find \( \alpha_k > 0 \) such that
\[ J(u^k + \alpha_k p_k) < J(u^k), \]
(A3.2.10)
set
\[ u^{k+1} := u^k + \alpha_k p_k; \]

S3: set \( k := k + 1 \) and go to Step 1.

\[ \diamond \]

A3.2.12 Proposition. Let \( U^* = \arg \min_{u \in \mathbb{R}^n} J(u) \) be non-empty and bounded and the sequence \( \{u^k\} \) is generated by Method A3.2.11. Assume that a decreasing function \( \delta : \mathbb{R}_+ \to \mathbb{R}_- \) exists with \( \delta(t) < 0 \) for \( t > 0 \) such that
\[ J(u^{k+1}) \leq J(u^k) + \delta(\| \nabla J(u^k) \|), \quad \forall \ k \in \mathbb{N}. \]
(A3.2.11)

Then \( \{u^k\} \) is bounded and each cluster point of \( \{u^k\} \) belongs to \( U^* \).

Now, if we suppose that for some \( \mu \in [-1, 0) \) in S1 of Algorithm (A3.2.9) an efficient method is available, then, in order to prove convergence of the algorithm, it suffices to construct a suitable function \( \delta \) or at least to establish its existence.

In the sequel we investigate different line search methods. In order to do so the following statement is useful (see Demyanov and Malozemov [88]).

A3.2.13 Proposition. Let \( J \) be continuously differentiable on an open set \( \Omega' \) and \( \Omega \subset \Omega' \) be a compact set. Then there exists a function \( o : \mathbb{R}_+ \to \mathbb{R}_+ \) such that \( \lim_{\gamma \downarrow 0} \frac{o(\gamma)}{\gamma} = 0 \), and for all \( u \in \Omega \), \( p \in \mathbb{R}^n \) and \( \alpha \geq 0 \) the inequality
\[ J(u + \alpha p) \leq J(u) + \alpha \langle p, \nabla J(u) \rangle + o(\alpha \| p \|) \]
(A3.2.12)
is true.
Goldstein Search:
With fixed $\sigma \in (0, \frac{1}{2})$ and a given direction $p^k$ the step-length parameters $\alpha_k > 0$ in (A3.2.10) are chosen such that
\[
\alpha_k (1 - \sigma) \langle p^k, \nabla J(u^k) \rangle \leq J(u^k + \alpha_k p^k) - J(u^k) \leq \alpha_k \sigma \langle p^k, \nabla J(u^k) \rangle, \quad \forall k.
\]
This inequality means that the ratio $q$ between both, the ascent of the secant going through the points $(u^k, J(u^k))$, $(u^k + \alpha_k p^k, J(u^k + \alpha_k p^k))$ and the ascent of the tangent at $u^k$, is bounded by $1 - \sigma \geq q \geq \sigma$.

If inequality (A3.2.9) holds, then the function $\delta$ in Proposition A3.2.12 exists. Indeed, since $U^*$ is non-empty and bounded, due to Proposition A1.7.55, the set $\{ u : J(u) \leq J(u^0) \}$ is compact. In view of the differentiability of the objective function $J$ Proposition A3.2.13 implies the existence of a function $o : \mathbb{R}_+ \to \mathbb{R}_+$ such that
\[
\lim_{\gamma \downarrow 0} \frac{o(\gamma)}{\gamma} = 0
\]
and the relation
\[
J(u^k + \alpha_k p^k) \leq J(u^k) + \alpha_k \langle p^k, \nabla J(u^k) \rangle + o(\|p^k\|), \quad \forall k
\]
is satisfied. Consequently, a non-increasing function $\bar{\alpha} : \mathbb{R}_- \to \mathbb{R}_+$ exists with $\bar{\alpha}(\gamma) > 0$ for $\gamma < 0$ such that
\[
J(u^k + \alpha_k p^k) < J(u^k) + \alpha (1 - \sigma) \langle p^k, \nabla J(u^k) \rangle
\]
holds with $\alpha \|p^k\| \in \left(0, \bar{\alpha}(\mu\|\nabla J(u^k)\|)\right)$.

Due to the left inequality in (A3.2.13), as well as the relations (A3.2.9), (A3.2.14) and the monotonicity of $\bar{\alpha}$, we obtain
\[
\alpha_k \|p^k\| \geq \bar{\alpha}(\mu\|\nabla J(u^k)\|).
\]
Moreover, on account of (A3.2.9), (A3.2.15) and the right-hand side in (A3.2.13)
\[
J(u^{k+1}) \leq J(u^k) + \frac{\sigma}{\|p^k\|} \langle p^k, \nabla J(u^k) \rangle \bar{\alpha}(\mu\|\nabla J(u^k)\|) \\
\leq J(u^k) + \sigma \mu \|\nabla J(u^k)\| \bar{\alpha}(\mu\|\nabla J(u^k)\|), \quad \forall k.
\]
Thus, we can choose $\delta(t) = \sigma \mu \bar{\alpha}(\mu t)$.

Another often used procedure to calculate the step-length parameters $\alpha_k > 0$ in (A3.2.10) is the so-called

Armijo Search:
Suppose that there exists a constant $c > 0$ such that
\[
\|p^k\| \geq c\|\nabla J(u^k)\|, \quad \forall k \in \mathbb{N}.
\]
We note that this assumption holds for almost all methods used in practice, in particular, if $p^k$ is determined as a Quasi-Newton direction, i.e.,
\[
p^k := -H_k \nabla J(u^k)
\]
with
\[ \langle y, H_k y \rangle \geq m \|y\|^2, \quad \forall \ y \in \mathbb{R}^n \]
is fulfilled with a constant \( m > 0 \) for all \( k \).

Now the Armijo Search reads as follows:
Choose some \( \tau > 0 \), \( \beta \in (0,1) \) and \( \sigma \in (0,1) \). Define on the \( k \)-th step the smallest integer \( s_k \geq 0 \) such that \( \alpha_k := \tau \beta^{s_k} \) satisfies the relation
\[ J(u^k + \alpha_k p^k) - J(u^k) \leq \alpha_k \sigma \langle p^k, \nabla J(u^k) \rangle. \tag{A3.2.17} \]
As in the Goldstein Search, for \( \alpha \|p^k\| \in (0, \bar{\alpha} \|p^k\|, \nabla J(u^k)) \), and consequently for \( \alpha \|p^k\| \in (0, \bar{\alpha}(\mu \|\nabla J(u^k)\|)) \), the inequality
\[ J(u^k + \alpha p^k) \leq J(u^k) + \alpha \sigma \langle p^k, \nabla J(u^k) \rangle \]
holds. If \( s_k > 0 \) then, instead of (A3.2.13), the inequalities
\[ \tau \beta^{s_k - 1} \langle p^k, \nabla J(u^k) \rangle < J(u^k + \tau \beta^{s_k - 1} p^k) - J(u^k), \]
\[ J(u^k + \tau \beta^{s_k} p^k) - J(u^k) \leq \sigma \tau \beta^{s_k} \langle p^k, \nabla J(u^k) \rangle \]
can be used and we obtain analogously
\[ J(u^k + \alpha_k p^k) \leq J(u^k) + \beta \sigma \mu \|\nabla J(u^k)\| \|\alpha \|\nabla J(u^k)\|. \]
If \( s_k = 0 \) then, due to (A3.2.16) and (A3.2.17)
\[ J(u^k + \alpha_k p^k) \leq J(u^k) + \tau \sigma \mu \|\nabla J(u^k)\|^2. \]
Consequently, (A3.2.11) is satisfied with
\[ \delta(t) = \mu \sigma t \min \{\varepsilon t, \beta \bar{\alpha}(\mu t)\}, \quad \forall \ t. \tag{A3.2.18} \]

Minimization along a direction:
In this method \( \alpha_k \) is defined by an one-dimensional minimization problem along the given direction \( p^k \)
\[ \alpha_k := \arg \min_{\alpha \geq 0} J(u^k + \alpha p^k). \tag{A3.2.19} \]

Obviously, (A3.2.11) holds with any function \( \delta \) constructed above.

It is easy to verify that the existence of a suitable \( \alpha_k \) is guaranteed in each of the methods described above if the optimal set \( U^* \) is non-empty and bounded.

In the Goldstein Search the value \( \alpha_k \) is any positive root of the equation \( \zeta(\alpha) = 0 \), with \( \zeta : \mathbb{R}_+ \rightarrow \mathbb{R} \) defined by
\[ \zeta(\alpha) = \begin{cases} 
\zeta(\alpha) & \text{if } \zeta(\alpha) > 0, \\
\zeta(\alpha) & \text{if } \zeta(\alpha) < 0, \\
0 & \text{otherwise} 
\end{cases} \]
and

$$
\zeta(\alpha) = J(u^k + \alpha p^k) - J(u^k) - \alpha \sigma(p^k, \nabla J(u^k)),
$$

$$
\bar{\zeta}(\alpha) = J(u^k + \alpha p^k) - J(u^k) - \alpha (1 - \sigma)(p^k, \nabla J(u^k)).
$$

In Polak [322] a bisection procedure is described which solves the equation \( \zeta(\alpha) = 0 \) in a finite number of iterations. The Armijo Search of \( \alpha_k \) has an algorithmic structure, and minimizing the univariate function \( J(u^k + \alpha p^k) \) according to (A3.2.19), the methods considered above can be used.

### A3.2.3 Gradient methods

Using formula (A1.5.15) in order to compute the directional derivative, one can define a direction of the local steepest descent at a point \( u \) as

$$
p := \arg \min \{ \langle \nabla J(u), y \rangle \} = -\frac{\nabla J(u)}{\|\nabla J(u)\|}
$$

Thus, setting \( p^k := -\nabla J(u^k) \), we obtain a gradient method. The conditions (A3.2.9) and (A3.2.16) are obviously satisfied. Using the step-size procedures considered above, convergence statements for gradient methods are corollaries of Proposition A3.2.12.

If \( \nabla J \) satisfies a Lipschitz condition with a known constant \( L \), then there exists a simpler rule to determine \( \{\alpha_k\} \). In that case one can choose \( \{\alpha_k\} \) such that

$$
0 < \epsilon_1 \leq \alpha_k \leq \frac{2}{L} - \epsilon_2 \quad (A3.2.20)
$$

with arbitrary \( \epsilon_1 > 0, \epsilon_2 > 0 \). In particular, one can take \( \alpha_k = \alpha \forall k \in \mathbb{N} \), with \( \alpha \in (0, \frac{2}{L}) \). Indeed, in view of (A1.4.10) for \( u := u^k \) and \( w := -\alpha_k \nabla J(u^k) \) it follows

$$
J(u^{k+1}) = J(u^k) - \alpha_k \|\nabla J(u^k)\|^2
- \alpha_k \int_0^1 \langle \nabla J(u^k - \tau \alpha_k \nabla J(u^k)) - \nabla J(u^k), \nabla J(u^k) \rangle d\tau
\leq J(u^k) - \alpha_k \|\nabla J(u^k)\|^2 + L \alpha_k^2 \|\nabla J(u^k)\|^2 \int_0^1 \tau d\tau
= J(u^k) - \alpha_k \left(1 - \frac{1}{2}L \alpha_k \right) \|\nabla J(u^k)\|^2 \leq J(u^k) - \frac{1}{2}L \epsilon_1 \epsilon_2 \|\nabla J(u^k)\|^2
$$

and inequality (A3.2.11) is satisfied with \( \delta(t) = -\frac{1}{2}L \epsilon_1 \epsilon_2 \).}

### A3.2.14 Proposition. Let \( J \) be a convex, differentiable function and \( \nabla J \) be Lipschitz-continuous (with constant \( L \)). Then, for \( \alpha_k := \alpha \forall k, \alpha \in (0, \frac{2}{L}) \), the gradient method converges to a minimizer \( u^* \) of the function \( J \) on \( \mathbb{R}^n \).

If, in addition, \( J \) is strongly convex (with constant \( \kappa \)), then the convergence is linear, i.e.,

$$
\|u^k - u^*\| \leq cq^k \quad k = 1, 2, \ldots ,
$$

with \( c = \sqrt{\frac{1}{\kappa}(J(u^0) - J(u^*))} \) and \( q = \sqrt{1 - 4\kappa \sigma + 2L \kappa \alpha^2} \).

Obviously, \( q^* = \sqrt{1 - \frac{2}{L}} \) is minimal and is attained for \( \alpha^* = \frac{1}{L} \).
A3. Selected Methods of Convex Minimization

A3.2.15 Proposition. Let \( J \) be twice-differentiable and
\[
m \langle v, v \rangle \leq \langle v, \nabla^2 J(u)v \rangle \leq M \langle v, v \rangle
\]
for some \( m > 0, M \) and any \( v \in \mathbb{R}^n, u \in \mathbb{R}^n \).
Then, for \( \alpha_k := \gamma k, \alpha \in (0, \frac{2}{M}) \), the estimates
\[
\|u^k - u^*\| \leq q^k \|u^0 - u^*\|, \quad k = 1, 2, \ldots
\]
hold with
\[
q(\gamma) := \max \{|1 - \gamma m|, |1 - \gamma M|\}.
\]
The value \( q^* := \frac{M - m}{M + m} \) becomes minimal if \( \gamma^* := \frac{2}{M+m} \).
Estimate (A3.2.22) is sharp in the following sense: It cannot be improved for any quadratic function
\[
J(u) := \frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle,
\]
where \( A \) is a positive definite matrix with eigenvalues \( \lambda_{\min}(A) = m \) and \( \lambda_{\max}(A) = M \).

Estimates of the rate of convergence give no clue for a preference of any variant of the gradient methods. These methods converge slowly in the case of ill-conditioning of the problems. If we know a suitable value of the Lipschitz constant \( L \), then the choice of \( \alpha_k \) according to (A3.2.20) is more convenient. However, the determination of \( L \) is usually a challenge.

Nevertheless, gradient methods are simple and converge globally under relatively weak assumptions.

A3.2.4 Conjugate gradient methods

Methods of conjugate directions make use of successive quadratic approximations of convex functions, because convex quadratic functions can be minimized by means of these methods in a finite number of iterations.

First we consider a quadratic function
\[
J(u) := \frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle,
\]
with a positive definite \([n \times n]\)-matrix \( A \) and recall that the minimization of \( J \) is equivalent to the solution of the equation \( Au = f \).

A3.2.16 Definition. The pair of vectors \( p^i, p^j \) is called conjugate with respect to \( A \) or \( A \)-orthogonal if \( \langle Ap^i, p^j \rangle = 0 \).

If a system of pair-wise conjugate vectors \( p^i \ (i = 0, \ldots, n - 1) \) is known, then the minimizer of \( J \) can be calculated easily. Indeed, using for the sought solution the representation
\[
\sum_{i=0}^{n-1} \alpha_i p^i,
\]
the equation \( Au = f \) provides
\[
\sum_{i=0}^{n-1} \alpha_i Ap^i = f.
\]
CHAPTER 11. APPENDIX

Multiplying this equation successively by $p_0, \ldots, p^{n-1}$, it follows that

$$\alpha_i = \frac{\langle f, p_i \rangle}{\langle Ap_i, p_i \rangle}.$$ 

Now, this determination of $u$ can be described recurrently:

$$u^{k+1} := u^k + \alpha_k p^k, \quad k = 0, \ldots, m \ (m \leq n - 1),$$

with

$$\alpha_k := \arg \min_{\alpha \in \mathbb{R}} J(u^k + \alpha p^k).$$

Moreover, $u^{k+1}$ minimizes $J$ on the affine set $u^0 + \text{span}(p_0, \ldots, p^k)$. This enables us to construct conjugate directions also by means of the recurrent formula described above.

Now we describe one of the variants of a conjugate gradient method (cg-method).

**A3.2.17 Method.** *(Fletcher-Reeves cg-method)*

S0: Choose an arbitrary $u^0 \in \mathbb{R}^n$ and set

$$r^0 := \nabla J(u^0), \quad p^0 := -r^0, \quad k := 0.$$ (A3.2.24)

S1: Determine

$$\alpha_k := \arg \min_{\alpha \in \mathbb{R}_+} J(u^k + \alpha p^k),$$ (A3.2.25)

S2: Set

$$u^{k+1} := u^k + \alpha_k p^k,$$ (A3.2.26)

$$r^{k+1} := \nabla J(u^{k+1}),$$ (A3.2.27)

$$\beta_{k+1} := \frac{\|r^{k+1}\|^2}{\|p^k\|^2},$$ (A3.2.28)

$$p^{k+1} := -r^{k+1} + \beta_{k+1} p^k.$$ (A3.2.29)

S3: Set $k := k + 1$ go to S1.

♦

**A3.2.18 Proposition.** The gradients $r^0, r^1, \ldots$ in Method **A3.2.17** are pair-wise orthogonal, and the directions $p^0, p^1, \ldots$ are pair-wise $A$-orthogonal.

**Proof:** Assume that for fixed $k \geq 2$

$$\langle r^i, r^j \rangle = 0, \quad \langle Ap^i, p^j \rangle = 0 \quad \text{for } i < j \leq k \text{ and } r^i \neq 0, \ i = 0, \ldots, k.$$

Obviously, $r^i = 0$ implies the end of the iteration process. For $k = 2$ the first two relations can be verified immediately.

Then, using (A3.2.26) and (A3.2.29), we get

$$u^{k+1} := u^k - \alpha_k r^k + \frac{\alpha_k \beta_k}{\alpha_{k-1}} (u^k - u^{k-1})$$
and, consequently,
\[ r^{k+1} := r^k - 2\alpha_k A r^k + \frac{\alpha_k \beta_k}{\alpha_{k-1}} (r^k - r^{k-1}). \]

Due to the choice of \(\alpha_i\), it follows for all \(i\) that
\[ 0 = J'_\alpha(u^i + \alpha p^i)|_{\alpha=\alpha_i} = \langle r^{i+1}, p^i \rangle, \]
and since \(r^k \neq 0\),
\[ \langle r^k, p^k \rangle = -\|r^k\|^2 + \beta_k \langle r^k, p^{k-1} \rangle < 0. \]
Thus, \(\alpha_k \neq 0\) and \(Ar^k\) is a linear combination of the vectors \(r^{k+1}, r^k\) and \(r^{k-1}\).

Henceforward we write
\[ Ar^k = \text{span}(r^{k+1}, r^k, r^{k-1}). \]

Analogously,
\[ Ar^i = \text{span}(r^{i+1}, r^i, r^{i-1}), \quad 0 < i < k \]
can be established. In view of (A3.2.31) and the assumption of induction we obtain
\[ \langle Ar^i, r^k \rangle = 0, \quad \text{for } i < k - 1, \]
hence,
\[ \langle r^{k+1}, r^i \rangle = \langle r^k - 2\alpha_k A r^k + \frac{\alpha_k \beta_k}{\alpha_{k-1}} (r^k - r^{k-1}), r^i \rangle = 0 \quad \text{for } i = 0, \ldots, k - 2. \]

But (A3.2.24), (A3.2.29) imply
\[ p^k = \text{span}(r^k, r^{k-1}, \ldots, r^0) \]
and
\[ \langle r^{k+1}, p^i \rangle = 0 \quad \text{for } i = 0, \ldots, k - 2. \]

Furthermore,
\[ r^{k+1} := r^k + 2\alpha_k Ap^k \]
and
\[ \langle r^{k+1}, p^{k-1} \rangle = \langle r^k, p^{k-1} \rangle + 2\alpha_k \langle Ap^k, p^{k-1} \rangle = 2\alpha_k \langle p^{k-1}, Ap^k \rangle = 0. \]

Now, due to (A3.2.29) and (A3.2.30), we obtain
\[ \langle r^{k+1}, r^k \rangle = -\langle r^{k+1}, p^k \rangle + \beta_k \langle r^{k+1}, p^{k-1} \rangle = 0 \]
(A3.2.33)
and, analogously,
\[ \langle r^{k+1}, r^{k-1} \rangle = \langle r^{k+1}, -p^{k-1} + \beta_{k-1} p^{k-2} \rangle = 0. \]

According to (A3.2.31) and (A3.2.32) for \(i \leq k - 1\) it follows that
\[ Ap^i = \text{span}(r^{i+1}, \ldots, r^0) \]
and using \(p^{k+1} := -r^{k+1} + \beta_{k+1} p^k\) together with the assumption of induction, one can conclude that
\[ \langle Ap^i, p^{k+1} \rangle = \beta_{k+1} \langle Ap^i, p^k \rangle - \langle \text{span}(r^{i+1}, \ldots, r^0), r^{k+1} \rangle = 0 \quad \text{for } i \leq k - 1. \]
Finally, from (A3.2.26) we get
\[ Ap_k = \frac{1}{2\alpha_k} (r^{k+1} - r^k), \]
and the relations (A3.2.29), (A3.2.33), (A3.2.30) and (A3.2.28) lead to
\[
\langle Ap_k, p^{k+1} \rangle = \frac{1}{2\alpha_k} \langle r^{k+1} - r^k, -r^{k+1} + \beta_{k+1}p^k \rangle \\
= \frac{1}{2\alpha_k} (-\|r^{k+1}\|^2 - \beta_{k+1} \langle r^k, p^k \rangle) \\
= \frac{1}{2\alpha_k} (-\|r^{k+1}\|^2 - \beta_{k+1} \langle r^k, -r^k + \beta_k p^{k-1} \rangle) \\
= \frac{1}{2\alpha_k} (-\|r^{k+1}\|^2 + \beta_{k+1} \|r^k\|^2) = 0.
\]

\[ \square \]

This statement implies immediately the following result.

**A3.2.19 Proposition.** Method A3.2.17 defines a minimizer of the quadratic function (A3.2.23) in at most \( n \) iterations.

**A3.2.20 Remark.** Proposition A3.2.19 remains true if the matrix \( A \) is positive semi-definite and \( \min_{u \in \mathbb{R}^n} J(u) > -\infty \). The number of iterations required does not exceed \( q = \text{rank}(A) \).

Method A3.2.17 can be applied without any change also to the minimization of convex, non-quadratic and differentiable functions \( J : \mathbb{R}^n \to \mathbb{R} \). Of course, in that case the determination of the solution requires infinitely many iterations. For corresponding convergence statements see Daniel [86].

In order to avoid an accumulation of the errors, for non-quadratic functions Method A3.2.17 has to be slightly changed. A restart procedure has to be implemented with \( \beta_k = 0 \) after some steps. Usually, instead of (A3.2.28), \( \beta_k \) is determined by
\[
\beta_k := \begin{cases} 
0 & \text{if } k = n, 2n, 
\frac{\|r^k\|^2}{\|r^k - 1\|^2} & \text{if } k \neq n, 2n, \ldots \end{cases} \tag{A3.2.34}
\]

**A3.2.21 Proposition.**

(i) Let \( J \) be a convex, differentiable function and the optimal set \( U^* \) be non-empty and bounded. Then, the sequence \( \{u^k\} \), generated by Algorithm A3.2.17 with an infinite number of restarts via (A3.2.34), is bounded and each limit point belongs to \( U^* \).

(ii) If, moreover, \( J \) is strongly convex and twice-continuously differentiable, then \( \{u^k\} \) converges super-linearly.

If the Hessian \( \nabla^2 J(u^*) \) fulfills a Lipschitz condition in a neighborhood \( O(u^*) \), then the estimate
\[
\|u^{(m+1)n} - u^*\| \leq c\|u^{mn} - u^*\|^2 \quad \forall \ m \in \mathbb{N}
\]
holds in \( O(u^*) \).
A3. SELECTED METHODS OF CONVEX MINIMIZATION

Solving large-scale problems, an estimate of the distance between the iterates $\mathbf{u}^k$ and the minimizer $\mathbf{u}^*$ is of interest even in the case of quadratic functions for $k < n$.

For the quadratic function (A3.2.23) with eigenvalues $\lambda(A)_{\text{min}} > m > 0$, $\lambda(A)_{\text{max}} < M$, the estimate

$$\|\mathbf{u}^k - \mathbf{u}^*\| \leq 2 \sqrt{\frac{M}{m}} q^k \|\mathbf{u}^0 - \mathbf{u}^*\|$$

with

$$q = \frac{\sqrt{M} - \sqrt{m}}{\sqrt{M} + \sqrt{m}}$$

holds, and for $m = \lambda(A)_{\text{min}}$, $M = \lambda(A)_{\text{max}}$ this estimate cannot be improved.

There are a lot of cg-methods that differ mainly in the choice of $\beta_k$ (see, in particular, Fletcher [116], Gill, Murray and Wright [128], Al-Baali [4]). A sufficient general approach for cg-methods is described by Pšeničnyj and Danilin [341].

For large-scale problems, in practice, the solution procedure becomes essentially faster in a neighborhood of the solution if the eigenvalues of the Hessian $\nabla^2 J(\mathbf{u}^*)$ are clustered into sets containing the eigenvalues of $\nabla J(\mathbf{u}^k)$ with similar magnitude. This happens often when penalty methods are used, especially, when the problem under consideration results from a discretization.

For the minimization of a quadratic function with a large number of variables special preconditioning procedures have been developed in order to accelerate the cg-method (see Concus and Golub [80], Glowinski et al. [136] and Blaheta [45]). The large size and the condition of many technical or physical applications result in the need for efficient parallelization and preconditioning techniques of the cg-methods, see for instance Basermann, Reichel and Schelthoff [33].

A3.2.5 Quasi-Newton methods

Suppose that the objective function $J : \mathbb{R}^n \to \mathbb{R}$ is strongly convex (with constant $\kappa$) and twice-differentiable.

First we consider Newton’s method for solving unconstrained minimization problems. Starting with $\mathbf{u}^0 \in \mathbb{R}^n$ the method determines recursively

$$\mathbf{u}^{k+1} := \mathbf{u}^k - [\nabla^2 J(\mathbf{u}^k)]^{-1} \nabla J(\mathbf{u}^k), \quad \forall k \in \mathbb{N},$$

and can be considered as the classical (undamped) Newton method for solving the equation $\nabla J(u) = 0$.

A3.2.22 Proposition. Let the Hessian $\nabla^2 J$ satisfy a Lipschitz condition (with constant $L$). If the starting point $\mathbf{u}^0$ satisfies the condition

$$q := \frac{L}{2\kappa^2} \|\nabla J(\mathbf{u}^0)\| < 1,$$

then Method (A3.2.35) converges to the minimizer $\mathbf{u}^*$ with a quadratic rate, i.e.,

$$\|\mathbf{u}^k - \mathbf{u}^*\| \leq \frac{2\kappa}{L} q^k.$$
A3.2.23 Remark. For the damped Newton method

\[ u^{k+1} := u^k - \alpha_k [\nabla^2 J(u^k)]^{-1} \nabla J(u^k), \quad \forall k \in \mathbb{N}, \]  

(A3.2.36)

with \( \alpha_k \) chosen according to the Armijo Search (with \( \tau = 1 \) and \( \sigma \in (0, \frac{1}{2}) \)), in Pšeničnyj and Danilin [341] it is shown that under the assumptions of Proposition A3.2.22 the damped Newton method converges super-linearly for arbitrary starting point \( u^0 \), and after some iterations it turns into the undamped Newton method, i.e., \( \alpha_k = 1 \) for \( k \geq k' \).

The computation of the Hessians \( \nabla^2 J(u^k) \) and especially the solution of the system

\[ \nabla^2 J(u^k)(u - u^k) = \nabla J(u^k) \]

are expensive procedures. Therefore, numerous modifications of Newton’s method have been developed, in which, instead of \( \{\nabla^2 J(u^k)\} \), a sequence of symmetric matrices \( H_k \in \mathbb{R}^{n \times n} \) with

\[ \|H_k - [\nabla^2 J(u^k)]^{-1}\| \to 0 \quad \text{for} \quad k \to \infty \]  

(A3.2.37)

is constructed in a simpler way. In the majority of such methods the current matrix \( H_k \) is efficiently used to calculate \( H_k+1 \).

These so-called Quasi-Newton methods have the form

\[ u^{k+1} := u^k - \alpha_k H_k \nabla J(u^k) \quad k = 1, 2, \ldots . \]  

(A3.2.38)

In order to give a reason for formula (A3.2.37) we consider the Taylor expansion of \( \nabla J \):

\[ \nabla J(u^{k+1}) - \nabla J(u^k) = \nabla^2 J(u^k)(u^{k+1} - u^k) + o(\|u^{k+1} - u^k\|). \]

If \( u^k \to u^* \), then due to the continuity of \( \nabla^2 J \)

\[ \|\nabla^2 J(u^k) - \nabla^2 J(u^*)\| \to 0. \]

It is obvious that for arbitrary matrices \( B_k := H_k^{-1} \) the relation

\[ \nabla J(u^{k+1}) - \nabla J(u^k) = B_k (u^{k+1} - u^k) + o(\|u^{k+1} - u^k\|) \]  

(A3.2.39)

cannot be guaranteed, because \( H_k \) is calculated before \( u^{k+1} \) is.

However, together with \( H_k \to [\nabla^2 J(u^*)]^{-1} \) the characteristic property of quasi-Newton methods is their finiteness for strongly convex, quadratic functions.

Since for a quadratic \( J \)

\[ \nabla J(u^{k+1}) - \nabla J(u^k) = \nabla^2 J(u^k)(u^{k+1} - u^k) = 2A(u^{k+1} - u^k), \]

it is natural to preserve the validity of the so-called quasi-Newton condition

\[ H_{k+1} y^k = \alpha_k p^k, \]  

(A3.2.40)

with \( y^k := \nabla J(u^{k+1}) - J(u^k) \) and \( p^k := -H_k \nabla J(u^k) \). In other words, we "fulfil" condition (A3.2.39), in which the terms of higher order are deleted, with one modification: \( B_{k+1} \) has to be used instead of \( B_k \).

Moreover, it is important that these matrices \( H_k \) are symmetric and the
transition from $H_k$ to $H_{k+1}$ is as simple as possible. In particular, if $H_{k+1}$ is defined such that
\[ H_{k+1} := H_k + a_k u v^T, \]
i.e., with a rank-1-update, then symmetry of $H_{k+1}$ requires that $v = u$. Due to the quasi-Newton condition \((A3.2.40)\) we get
\[ H_k y^k + a_k u v^T y^k = \alpha_k p^k. \]
Consequently, the vector $u$ is proportional to $\alpha_k p^k - H_k y^k$, and inserting the coefficient of proportionality into $a_k$, we take
\[ u = \alpha_k p^k - H_k y^k, \]
with $a_k := \langle u, y^k \rangle^{-1}$.

The corresponding formula
\[ H^B_{k+1} := H_k + \frac{(\alpha_k p^k - H_k y^k)(\alpha_k p^k - H_k y^k)^T}{(\alpha_k p^k - H_k y^k)^T y^k} \] (A3.2.41)
leads to the Broyden method and exploits the possibilities of rank-1-updates.

Using a symmetric rank-2-update
\[ H_{k+1} := H_k + a_k uu^T + b_k vv^T, \]
the quasi-Newton condition
\[ H_k y^k + (a_k uu^T + b_k vv^T) y^k = \alpha_k p^k \]
defines a one-parametric family of formulas (Broyden class). In particular, for
\[ u := \alpha_k p^k, \quad v := H_k y^k, \quad a_k u^T y^k = 1, \quad b_k v^T y^k = -1, \]
the Davidon-Fletcher-Powell method
\[ H^DFP_{k+1} := H_k + \frac{p^k (p^k)^T}{y^k} - \frac{H_k y^k (y^k)^T H_k}{(y^k)^T H_k y^k} \] (A3.2.42)
can be obtained.

We refer also to an often used other formula of this family: the Broyden-Fletcher-Goldfarb-Shanno class
\[ H^BFGS_{k+1} := H_k + \left( 1 + \frac{(y^k)^T H_k y^k}{\alpha_k (p^k)^T y^k} \right) \frac{\alpha_k (p^k)(p_k)^T}{(p^k)^T y^k} - \frac{p^k (y^k)^T H_k + H_k y^k (p^k)^T}{(p^k)^T y^k} \] (A3.2.43)

Any formula of the Broyden family can be written as
\[ H^\xi_{k+1} := (1 - \xi) H^DFP_{k+1} + \xi H^BFGS_{k+1} \]
and for an arbitrary $\xi \geq 0$ the following properties can be guaranteed for arbitrary starting point $u^0$ and arbitrary positive definite matrix $H_0$:

(i) positive definites of the matrices $H_k \forall k$;
(ii) super-linear convergence with respect to $m$ for the sequence $\{\|u^{mn} - u^*\|\}$ in a neighborhood of the minimizer $u^*$, in particular, for some methods quadratic convergence can be established in $O(u^*)$ if $\nabla^2 J$ is Lipschitz-continuous.

Moreover, for quadratic functions (A3.2.23) it holds also:

(iii) termination of the iteration after $m \leq n$ steps, and $H_m = A^{-1}$ if $m = n$;

(iv) independence of the sequence $\{u^k\}$ from parameter $\xi$;

(v) coincidence of the iterates of these methods and the iterates of conjugate gradient methods, if $H_0 := I$ is chosen.

It is useful to consider quasi-Newton methods also from another point of view. If, together with the scalar product $\langle \cdot, \cdot \rangle$ in $\mathbb{R}^n$ a new scalar product $\langle \cdot, \cdot \rangle_B$ is introduced by means of an arbitrary positive definite matrix $B$, i.e.

$$\langle u, v \rangle_B := \langle Bu, v \rangle,$$

then the gradient $\nabla_B J$ of function $J$ (on the space with this new scalar product) is related with $\nabla J$ by

$$\nabla_B J(u) := B^{-1}\nabla J(u).$$

In this new metric the gradient method reads as

$$u^{k+1} := u^k - \alpha_k B_k^{-1}\nabla J(u^k) \quad k = 1, 2, \cdots.$$  \hspace{1cm} (A3.2.44)

It is easily seen that gradient methods are not invariant with respect to the choice of the metric, and in the case $B_k := \nabla^2 J(u^k)$ an essentially faster gradient method is created. Hence, one can attempt to accelerate a gradient method by means of a suitable metric.

In order to obtain a faster decreasing of the function $J$, the choice of $B_k := \nabla^2 J(u^k)$ can be considered as the best one. Therefore, quasi-Newton methods can be interpreted as gradient methods from the point of view that at each step a new metric is chosen as close as possible to the best one. Note that the Newton methods and quasi-Newton methods considered above, if $H_0 := [\nabla^2 J(u^0)]^{-1}$ is taken for the latter ones, are invariant with respect to the choice of any metric.

Hence, it makes sense that often quasi-Newton methods are called variable metric methods, but the latter class is substantially broader. For a review of such methods see Lukšan and Spedicato [280]. Investigations about global and super-linear convergence of certain classes of variable metric methods can be found by Ritter [346]. In Vlček and Lukšan [410] variable metric methods with a limited-memory effect are studied for large-scale unconstrained optimization.

The number of publications concerning quasi-Newton methods and cg-methods is growing very quickly. Here we refer only to some monographs and special collection papers (cf. Fletcher [116], Gill, Murray and Wright [128], Hestens [172], Dixon and Szego [91], Lootsma [278]). In Broyden and Vespucci [58] a brief review of cg-methods is given.

There are also important investigations taking into account inexact line search methods (cf. Powell [333], Al-Baali [4]), functions with sparse Hessians (cf. Toint [397], Griewank and Toint [149], Powell and Toint [280]).
and not strongly convex functions (cf. Powell [333], Pšeničnyj and Danilin [341], Dennis and Schnabel [90]).

Concerning a comparison of different methods from the numerical point of view we refer to the analysis of Fletcher [117], Schittkowski [363] and Brezinski [49]. For ill-posed problems the cg-method is studied by Frommer and Maass [122].

For non-quadratic functions, numerical experiments show that quasi-Newton methods are more efficient. However, cg-methods require to store only some current \( n \)-dimensional vectors, whereas in quasi-Newton methods it is necessary to store and to update \( n \times n \)-matrices. Thus, the latter methods may be inapplicable for large-scale problems which still can be solved by cg-methods quite successful. Although, there are recently some developments concerning inexact quasi-Newton and cg-methods, which allow to work in lower dimensional spaces. These algorithms forces the iterates to stay on an appropriate manifold, so called Krylow subspace, of a dimension much smaller than that of the Hessian itself and reinitializes approximate curvature along directions off the manifold. In this way, typical difficulties associated with ill-conditioned can be overcome, see for instance Gill and Leonhard [129], Fasano [109].

**A3.2.24 Remark.** Quasi-Newton- and variable metric methods can also applied to the minimization of non-differentiable functions. Here the transition to a new metric is performed by stretching the space in the direction of some subgradients at the previous point or in the direction of the difference between some subgradients at the last two points, see for instance, Uryas’ev [404], Kiwiel [236] and Lemaréchal and Sagastizábal [262]. ♦

### A3.3 Minimization methods subject to simple constraints

In this subsection we describe methods which are applicable for solving minimization problems arising from the discretization of elliptic variational inequalities.

They are destined to solve Problem (A1.5.11), with \( V = \mathbb{R}^n \) and box-constraints like

\[
K = \otimes_{i=1}^n S_i, \quad S_i = [a_i, b_i].
\]

The cases \( a_i = -\infty \) and \( b_i = +\infty \) are not excluded.

Discretization of variational inequalities by means of finite element methods or finite difference methods leads to problems in which the Hessian of the objective functions have a special structure, i.e., non-zero elements are located only along some diagonals. Usually, the approximate problems of real-life applications have a large dimension, which confines essentially the use of universal optimization methods.

Relaxation methods, based on the coordinate descent, and gradient type methods make use in an effective manner of the structure and sparseness of the Hessian of the objective function in the approximate problems.

#### A3.3.1 Method of coordinate descent

Now, we consider the following method to minimize a strictly convex and differentiable function \( J : \mathbb{R}^n \to \mathbb{R} \) on box-constraints (A3.3.45). Assume that for
fixed \( u^0 \in K \) the set 
\[
\bar{K} = K \cap \{ u : J(u) \leq J(u^0) \}
\]
is compact.

**A3.3.25 Method. (Coordinate descent method)**

S0: Set \( k := 0, i := 0, u^{0,1} := u^0 \).

S1: Denote \( u^{i,k+1} := (u_1^{k+1}, \ldots, u_i^{k+1}, u_{i+1}, \ldots, u_n) \), \( 0 \leq i \leq n - 1 \).

Compute 
\[
u_{i+1}^{k+1} := \arg \min_{v \in S_{i+1}} J(u_1^{k+1}, \ldots, u_i^{k+1}, v, u_{i+2}^k, \ldots, u_n^k)
\]
and set 
\[
u^{i+1,k+1} := (u_1^{k+1}, \ldots, u_i^{k+1}, u_{i+1}^{k+1}, u_{i+2}^k, \ldots, u_n^k).
\]

S2: If \( i < n - 1 \), set \( i := i + 1 \) and go to S1; if \( i = n - 1 \), set \( k := k + 1, i := 0 \) and go to S1.

Convergence of this method can be proved by means of the following property of strictly convex functions.

**A3.3.26 Proposition.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a strictly convex function and \( C \) be a convex, compact set. Then there exists a strictly increasing function \( \tau : \mathbb{R}_+ \to \mathbb{R}_+ \), with \( \tau(0) = 0 \), such that for all \( u, v \in C \) and all \( q \in \partial f(v) \)
\[
f(u) - f(v) \geq \langle q, u - v \rangle + \tau(\|u - v\|).
\]

**A3.3.27 Proposition.** If the objective function \( J \) in Problem (A1.5.11) is strictly convex and continuously differentiable and \( \bar{K} \) is a compact set (with \( K \) given by (A3.3.45)), then the iterates, generated by Method A3.3.25, converge to
\[
u^* = \arg \min_{u \in \bar{K}} J(u).
\]

The idea of the proof consists in the following: Since \( J(u^{k+1}) \leq J(u^k) \) holds for arbitrary \( k \) and \( \bar{K} \) is a compact set, there exists \( \lim_{k \to \infty} J(u^k) = \bar{J} \). From Proposition A3.3.26 we obtain
\[
\|u^{k+1} - u^k\| \to 0, \quad \text{for } k \to \infty.
\]
If \( \bar{J} = J(u^*) + d \), with \( d > 0 \), then due to (A3.3.47), the relation
\[
-d > J(u^*) - J(u^k) \geq \langle \nabla J(u^k), u^* - u^k \rangle,
\]
and the continuity of \( \nabla J \), it holds
\[
\liminf_{k \to \infty} |u^{k+1}_{i_0(k)} - u^k_{i_0(k)}| > 0
\]
for indices $i_0(k)$ for which
\[
\frac{\partial J(u^k)}{\partial u_{i_0(k)}} (u^*_{i_0(k)} - u^k_{i_0(k)}) = \min_{1 \leq i \leq n} \frac{\partial J(u^k)}{\partial u_i} (u^* - u^k).
\]
But this contradicts (A3.3.47).

If
\[
J(u) = \frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle
\]
with a positive definite matrix $A$, then the method can be generalized in the following way. In order to determine $u^k_{i+1}$, instead of (A3.3.46), the procedure
\[
u^k_{i+1} = \arg\min_{v \in \mathbb{R}} \{ J(u^k_1, ..., u^k_i, v, u^k_{i+2}, ..., u^k_n) \},
\]
with $\omega \in (0, 2)$ is implemented. The fixed value $\omega$ is called relaxation parameter. Algorithm (A3.3.25) is included in the generalized method (A3.3.48) if one takes $\omega = 1$. In the case $K = \mathbb{R}^n$ and $\omega = 1$ we obtain the Gauss-Seidel method. Proposition A3.3.27 remains true also for the generalized method (A3.3.48).

In order to compute the component $u^k_{i+1}$ according to (A3.3.46), the methods of univariate minimization described in Subsection A3.2.1 can be used. If
\[
J
\]
is a quadratic functional, then $u^k_{i+1}$ is calculated as solution of the linear equation
\[
\frac{\partial}{\partial v} J(u^k_1, ..., u^k_i, v, u^k_{i+2}, ..., u^k_n) = 0.
\]

Note that a generalization of the method of coordinate descent towards a block-coordinate-wise descent is possible in which, instead of the scalar component $u_i$, a vector component $u_i \in \mathbb{R}^{m_i}$ is considered with $\mathbb{R}^n = \otimes_{i=1}^N \mathbb{R}^{m_i}$, $(n = \sum_{i=1}^N m_i)$, $K = \otimes_{i=1}^N S_i$ and $S_i \subset \mathbb{R}^{m_i}$ are assumed to be convex and closed. In this case we deal with the method of block relaxation, see Glowinski, Lions and Trémolières [135].

### A3.3.2 Gradient projection methods

Assume that $K \subset \mathbb{R}^n$ is an arbitrary convex, closed set. Let
\[
\Pi_K(u) = \arg\min\{ \| v - u \| : v \in K \},
\]
be the projection of the point $u$ onto the set $K$. Choosing $u^0 \in K$, according to the gradient method we determine an intermediate iterate
\[
u^k_{i+1} := \Pi_K(u^k_{i+1/2}) := u^k_{i+1/2} - \alpha_k \nabla J(u^k)
\]
and then
\[
u^{k+1} := \Pi_K(u^{k+1/2}).
\]
In case $K = \otimes_{i=1}^N S_i$, $S_i := [a_i, b_i]$, the projection (A3.3.49) can be performed very easily:
\[
u^{k+1}_i := \min \{ \max \{ u^{k+1/2}_i, a_i \}, b_i \}.
\]

We formulate a convergence result for Method (A3.3.49) in case $\alpha_k = \alpha$, $k = 1, 2, \cdots$. 

A3.3.28 Proposition. Let $K$ be a convex, closed subset of $\mathbb{R}^n$. Assume that $J: \mathbb{R}^n \to \mathbb{R}$ is a convex, differentiable function and $\nabla J$ fulfills a Lipschitz condition on $K$ with constant $L$. If the optimal set $U^* \neq \emptyset$, then the gradient projection method with $\alpha \in (0, \frac{2}{L})$ guarantees that $u_k \to u^* \in U^*$.

If, moreover, $J$ is twice-differentiable and

$$m(v, v) \leq \langle \nabla^2 J(u)v, v \rangle$$

for some $m > 0$ and any $u \in K, v \in \mathbb{R}^n$, then Method (A3.3.49) converges linearly and for $\alpha = \frac{1}{L}$ the factor of the geometric progression is $q^* = \max\{|1 - \alpha m|, |1 - \alpha L|\}$.

A3.3.3 Conjugate gradient method

As a last method in this subsection we consider a specification of the Fletcher-Reeves Method A3.2.17, which was developed by Polyak [329] in order to minimize a convex, differentiable function $J: \mathbb{R}^n \to \mathbb{R}$ on the set $K$ given in (A3.3.45).

A3.3.29 Method. (Polyak’s conjugate gradient method)

S0: Choose an arbitrary $u^0 \in K = \otimes_{i=1}^n S_i$.

Define

$$I_0 := \left\{ i : u_i^0 := a_i, \frac{\partial J(u^0)}{\partial u_i} > 0 \right\} \cup \left\{ i : u_i^0 := b_i, \frac{\partial J(u^0)}{\partial u_i} < 0 \right\},$$

$$p_i^0 := \begin{cases} -\frac{\partial J(u^0)}{\partial u_i} & \text{if } i \notin I_0, \\ 0 & \text{if } i \in I_0, \end{cases}$$

set $p^0 := (p_1^0, ..., p_n^0)$ and $k := 0$.

S1: Calculate

$$\alpha_k = \arg \min \{ J(u^k + \alpha p^k) : \alpha \geq 0, a \leq u^k + \alpha p^k \leq b \},$$

set $u^{k+1} := u^k + \alpha_k p^k$.

S2: If $\frac{\partial J(u^{k+1})}{\partial u_i} = 0 \forall i \notin I_k$, set

$$I_{k+1} := \left\{ i : u_i^{k+1} := a_i, \frac{\partial J(u^{k+1})}{\partial u_i} > 0 \right\} \cup \left\{ i : u_i^{k+1} := b_i, \frac{\partial J(u^{k+1})}{\partial u_i} < 0 \right\}$$

and go to S3; otherwise, set

$$I_{k+1} := I_k \cup \{ i : u_i^{k+1} := a_i \} \cup \{ i : u_i^{k+1} := b_i \}.$$ 

S3: Compute

$$\beta_{k+1} := \begin{cases} \sum_{i \notin I_{k+1}} \left( \frac{\partial J(u^{k+1})}{\partial u_i} \right)^2 \left( \sum_{i \notin I_{k+1}} \left( \frac{\partial J(u^k)}{\partial u_i} \right)^2 \right)^{-1} & \text{if } I_{k+1} = I_k, \\ 0 & \text{if } I_{k+1} \neq I_k, \end{cases}$$

$$p_i^{k+1} := \begin{cases} -\frac{\partial J(u^{k+1})}{\partial u_i} + \beta_{k+1} p_i^k & \text{if } i \notin I_{k+1}, \\ 0 & \text{if } i \in I_{k+1}, \end{cases}$$

(A3.3.51)

and set $p^{k+1} := (p_1^{k+1}, ..., p_n^{k+1})$. 
A3. SELECTED METHODS OF CONVEX MINIMIZATION

S4: Set $k := k + 1$ and go to S1.

\[\leftarrow\]

**A3.3.30 Proposition.** In order to minimize the quadratic function $J$ in (A3.2.23) (with positive semi-definite matrix $A$) on the feasible set $K$ given by (A3.3.45), Method A3.3.29 takes a finite number of iterations.

Note, if $I_k = I_{k+1} = \ldots = I_{k+r}$, the method performs a conjugate direction search on a corresponding subspace. In the non-quadratic case one can not exclude that $r \to \infty$. The possible infiniteness of Method A3.3.29 requires to control the accuracy of the search on the hyperplane and to introduce restart procedures analogous to those in the Fletcher-Reeves Method A3.2.17.

Now, we briefly consider methods for solving quadratic programming problems

\[(QP) \quad \begin{aligned} \text{minimize} & \quad J(u) = \frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle \\ \text{subject to} & \quad u \in K := \{v \in \mathbb{R}^n : Bv \leq d\}, \end{aligned} \tag{A3.3.52} \]

with $B \in \mathbb{R}^{m \times n}$, $d \in \mathbb{R}^m$, and $A \in \mathbb{R}^{n \times n}$ is supposed to be positive definite.

Note that for problems with linear constraints the Kuhn-Tucker-Theorem A1.7.51 remains true without Slater condition. Using this statement, it is possible to solve the dual problem of (A3.3.52), which can be written in the form

\[
\begin{aligned}
\min_{\lambda \geq 0} \{ \psi(\lambda) = & \frac{1}{2} \langle A^{-1}(f - B^T \lambda), f - B^T \lambda \rangle + \langle d, \lambda \rangle \}, \\
\end{aligned} \tag{A3.3.53}
\]

(cf. the transition from Problem (A3.4.56) to Problem (A3.4.67) below).

Thus, methods described in Subsection A3.3 can be applied for solving Problem (A3.3.53). But, in general, this approach is not quite efficient, because we need to know the inverse matrix $A^{-1}$ or have to solve an equation system with matrix $A$ in each step of the method considered. Indeed,

\[
\nabla \psi(\lambda) = BA^{-1}B^T \lambda + d - BA^{-1}f,
\]

with $BA^{-1}B^T$ a positive semi-definite matrix, and in order to compute $\nabla \psi(\lambda)$ we have to solve the systems

\[
\begin{aligned}
Ax &= f \\
Ay &= B^T \lambda.
\end{aligned} \tag{A3.3.54}
\]

The second system must be solved for fixed $\lambda$ in each step.

However, quadratic problems resulting from finite element discretization of elliptic variational inequalities have often matrices with diagonal and sparse structure, which enables us to solve the equation systems (A3.3.54) sufficiently fast. Here sparse $LU$-factorization is used for the matrix $A$ and row and column updates are performed for the constraint matrix $B$, see for instance, Gould [144]. Another approach to solve large-scale quadratic programming problems is due to the application of interior point methods, which are very efficient (cf. Cafieri et al. [67]).
A3.4 Methods for convex minimization problems

We start with the description of the cone of feasible directions for Problem (A1.7.35).

A3.4.31 Definition. Let $K$ be a convex set in the Hilbert space $V$. The set

$$\mathcal{C}_u(K) := \{ p \in V : p \neq 0, u + \gamma p \in K \text{ for some } \gamma > 0 \}$$

is said to be a cone of feasible directions at point $u \in K$, and its elements $p$ are called feasible directions at the point $u$.

In (A1.5.23) this set is used in order to characterize the solutions of Problem (A1.5.11). Obviously, $\mathcal{C}_u(K)$ is a convex cone. The closure of this cone can be described as follows.

A3.4.32 Proposition. Assume that the constraints of Problem (A1.7.35) fulfill the Slater condition and that the constraint functions $g_i$ are Gâteaux-differentiable on $V$. Then for $u \in K$ the relation

$$\text{cl}(\mathcal{C}_u(K)) = \{ p \in V : \langle \nabla g_j(u), p \rangle \leq 0, \ j \in I_0(u) \},$$

(A3.4.55)

holds with $I_0(u) = \{ j : g_j(u) = 0 \}$.

Now we consider the finite-dimensional variant of Problem (A1.7.35):

\[
\begin{align*}
\text{minimize} & \quad J(u), \\
\text{subject to} & \quad u \in K, \\
K & := \{ v \in \mathbb{R}^n : g_j(v) \leq 0, \ j = 1, \ldots, m \},
\end{align*}
\]

(A3.4.56)

with $J, g_j : \mathbb{R}^n \to \mathbb{R}$ are convex, differentiable functions, the Slater condition is fulfilled and $U^* \neq \emptyset$.

Obviously, Proposition A3.4.32 is applicable also to this problem.

A3.4.33 Remark. The Kuhn-Tucker Theorem A1.7.51 and Proposition A3.4.32 remain also true for Problem (A3.4.56) if the Slater condition is weakened in the following way:

there exists a point $\tilde{u} \in K$ such that $g_j(\tilde{u}) < 0$ for those functions which are not affine.

Due to the Propositions A1.5.34 (with $J_2 \equiv 0$) and A3.4.32, for $u \in K$ with $J(u) > J^*$ there exists a vector $p'$ such that

$$\langle \nabla J(u), p' \rangle < 0,$$
$$\langle \nabla g_j(u), p' \rangle \leq 0, \ \forall \ j \in I_0(u).$$

Taking $p := p' + \alpha (\tilde{u} - u)$, then for sufficiently small $\alpha > 0$ we obtain

$$\langle \nabla J(u), p \rangle < 0$$
and
\[
\langle \nabla g_j(u), p \rangle = \langle \nabla g_j(u), p' \rangle + \alpha (\nabla g_j(u), \tilde{u} - u) \\
\leq \alpha (g_j(\tilde{u}) - g_j(u)) < 0 \quad \forall \ j \in I_0.
\]

Conversely, if for \( u \in K \) the inequalities
\[
\langle \nabla J(u), p \rangle < 0, \\
\langle \nabla g_j(u), p \rangle \leq 0, \quad j \in I_0(u)
\]
are satisfied, then there exists a constant \( \alpha > 0 \) such that
\[
u + \alpha p \in K \quad \text{and} \quad J(u + \alpha p) < J(u).
\]

Indeed, the inequalities
\[
J(u + \alpha p) < J(u), \\
g_j(u + \alpha p) < 0, \quad j \notin I_0(u)
\]
are obvious for sufficiently small \( \alpha > 0 \), and for \( j \in I_0(u) \) we get
\[
g_j(u + \alpha p) = g_j(u + \alpha p) - g_j(u) \leq \alpha \langle \nabla g_j(u + \alpha p), p \rangle
\]
and there has nothing else to be done than to use the continuity of \( \nabla g_j \).

This explains the choice of the directions \( p \) in the following methods.

A3.4.1 Feasible direction methods

A3.4.34 Method. (Method of feasible directions)

Data: \( u^0 \in K \), \( \{\delta_k\} \downarrow 0 \).

S0: Set \( k := 0 \).

S1: Define \( I_\delta(u^k) := \{j : g_j(u^k) \geq -\delta_k\} \) and determine the vector \( (p^k, \sigma_k) \in \mathbb{R}^{n+1} \) as solution of the direction search problem
\[
\min \sigma \quad \text{subject to} \\
\langle \nabla J(u^k), p \rangle \leq \sigma \\
\langle \nabla g_j(u^k), p \rangle \leq \sigma, \quad \forall \ j \in I_\delta(u^k) \\
\langle p, p \rangle \leq 1 \quad (\text{or} \ \max_{1 \leq i \leq n} |p_i| \leq 1).
\]

S2: If \( -\sigma_k > \delta_k \), compute
\[
\alpha_k \approx \arg \min \{J(u^k + \alpha p^k) : u^k + \alpha p^k \in K\}, \quad (A3.4.59)
\]
\[
u^{k+1} := u^k + \alpha_k p^k \\
\delta_{k+1} := \delta_k \quad \text{and go to S3.}
\]

Otherwise set \( u^{k+1} := u^k, \ \delta_{k+1} := \frac{1}{2} \delta_k \).
S3: Set $k := k + 1$ and go to S1.

In (A3.4.59) the step-length $\alpha_k$ can be chosen according to

$$
\alpha_k := \arg \min \{J(u^k + \alpha p^k) : u^k + \alpha p^k \in K\} 
$$  \hspace{1cm} (A3.4.60)

or

$$
\alpha_k : \quad u^{k+1} := u^k + \alpha_k p^k \in K, \quad J(u^{k+1}) \leq \inf_{u^* + \alpha p^* \in K} J(u^k + \alpha p^k) + \epsilon_k 
$$  \hspace{1cm} (A3.4.61)

with

$$
\sum_{k=0}^{\infty} \epsilon_k < \infty. 
$$  \hspace{1cm} (A3.4.62)

Also Armijo search, described in Subsection A3.2.1, can be implemented with simple modifications guaranteeing that $u^k + \alpha_k p^k \in K$.

A3.4.35 Remark. The calculation of the direction $p^k$ can be improved if for affine functions $g_j$ the constraints $\langle \nabla g_j(u^k), p \rangle \leq \sigma$ are replaced by homogenous ones $\langle \nabla g_j(u^k), p \rangle \leq 0$.

A3.4.36 Proposition. Suppose the assumptions concerning Problem (A3.4.56) are fulfilled. Moreover, assume that the gradients $\nabla J$ and $\nabla g_j (j = 1, \ldots, m)$ satisfy a Lipschitz condition on $K$ and the set $U^*$ is non-empty and bounded. Then the sequence $\{u^k\}$, generated by Method A3.4.34, converges to $U^*$ if

(i) $\alpha_k$ is chosen according (A3.4.60) or

(ii) $\alpha_k$ is calculated by means of the conditions (A3.4.61) (A3.4.62).

The parameter $\delta_k$ in S1 has to be chosen such that too short steps are avoided. Using $I_0(u^k)$ instead of $I_3(u^k)$, examples are known which show that $J(u^k) \to J > J^*$. Constraint (A3.4.58) is introduced to guarantee the solvability of the direction search problems in S1. Usually, the constraint max $1 \leq i \leq n |p_i| \leq 1$ is chosen in order to deal in the directions search with a linear programming problem.

A3.4.37 Remark. Any method ensuring convergence of the sequence $\{u^k\}$ to $U^*$ for Problem (A3.4.56) determines a feasible starting point $u^0$ after a finite number of steps if it is applied to the auxiliary problem

$$
\min \{u_{n+1} : g_j(u) \leq u_{n+1}, j = 1, \ldots, m, J(u) \leq c\}
$$

with $c$ sufficient large.

For this problem a feasible starting point can be found trivially. The constraint $J(u) \leq c$ becomes superfluous if the convergence conditions of the method used do not require boundedness of the optimal set, or if $K$ itself is bounded.

It is easily seen in S1 of Method A3.4.34, if there are no constraints of the original problem and the constraint $\langle p, p \rangle \leq 1$ is used, then the method of feasible directions turns into a gradient method.
A3.4.2 Linearization methods

The idea of the linearization of the original problem is basic for many algorithms in convex programming. Now a method, suggested by Pšeníčnyj [337], [340], will be described using a different line search procedure as given in the papers mentioned.

A3.4.38 Method. (Pšeníčnyj’s linearization method)

Data: \( u^0 \in \mathbb{R}^n, \alpha > 0, \{\epsilon_k\} \downarrow 0; \)

S0: Set \( k := 0. \)

S1: Determine
\[
I_k(u^k) = \{j : g_j(u^k) \geq \max[0, \max_{1 \leq j \leq n} g_j(u^k)] - \epsilon\}. \tag{A3.4.63}
\]

S1: Compute an approximate \( \hat{u}^k \) of the quadratic programming problem
\[
\begin{align*}
\min \psi_k(u) &= \langle \nabla J(u^k), u - u^k \rangle + \frac{1}{2} \|u - u^k\|^2 \\
g_j(u^k) + \langle \nabla g_j(u^k), u - u^k \rangle &\leq 0, \quad \forall j \in I_k(u^k),
\end{align*} \tag{A3.4.64}
\]

such that
\[
|\nabla \psi_k(\hat{u}^k) - \nabla \psi_k(u^k)| \leq \epsilon_k, \tag{A3.4.65}
\]
with \( \hat{u}^k \) the exact solution of Problem (A3.4.64).

S2: Set
\[
\begin{align*}
u^{k+1} &:= u^k + \alpha(\hat{u}^k - u^k), \\
k &:= k + 1 \text{ and go to S1.}
\end{align*}
\]

♦

A3.4.39 Proposition. Suppose that

(i) the assumptions for Problem (A3.4.56) are satisfied;

(ii) the gradients of the functions \( J, g_j \) \( (j = 1, ..., m) \) obey a Lipschitz condition with constants \( L, L_j \) \( (j = 1, ..., m) \), respectively;

(iii) the step-length \( \alpha \) is chosen such that
\[
0 < \alpha < 2 \left( L + 2 \sum_{j=1}^{m} L_j \lambda_j^* \right)^{-1}
\]
with \( \lambda^* \) a Lagrange multiplier of Problem (A3.4.56);

(iv) the series \( \sum_{k=1}^{\infty} \epsilon_k \) converges and \( \epsilon = +\infty. \)

Then the sequence \( \{u^k\} \), generated by Method A3.4.38, converges to some element \( u^* \in U^* \). If, moreover, \( J \) is strongly convex and \( \hat{u}^k \) is the exact solution of Problem (A3.4.64), then
\[
\|u^k - u^*\| \leq c q^k, \quad q \in (0, 1).
\]
In POLYAK [330] it is noted that an analogous proposition is true for any $\epsilon > 0$.

The idea to decrease the number of constraints of Problem (A3.4.56) by means of a chosen small $\epsilon$ in (A3.4.63) seems to be promising. However, in [340] PˇšENIČNÝJ notes that the analysis of numerical experiments leads to a converse conclusion: $\epsilon$ should be chosen large enough in order to regard as many constraints of the initial problem as possible.

Now, we deal briefly with the numerical solution of the quadratic programming problem (A3.4.64). The feasible set of this problem contains the original feasible set $K$ and, because $\psi_k$ is strongly convex, the problem is uniquely solvable.

Due to the condition
$$\nabla_u \mathcal{L}_k(u, \lambda) = 0,$$
(A3.4.66)
with
$$\mathcal{L}_k(u, \lambda) := \psi_k(u) + \sum_{j \in I_k(u^k)} \lambda_j [g_j(u^k) + \langle \nabla g_j(u^k), u - u^k \rangle]$$
the Lagrangian function of Problem (A3.4.64), we obtain
$$u = u(\lambda) := u^k - \sum_{j \in I_k(u^k)} \lambda_j \nabla g_j(u^k).$$
Hence, the determination of the Lagrange multiplier vector $\bar{\lambda}^k$ leads to the problem
$$\max \{ \mathcal{L}_k(u(\lambda), \lambda) : \lambda_j \geq 0 \ \forall \ j \in I_k(u^k) \},$$
which can be reformulated as
$$\min \left\{ \frac{1}{2} \| \nabla J(u^k) + \sum_{j \in I_k(u^k)} \lambda_j \nabla g_j(u^k) \|^2 - \sum_{j \in I_k(u^k)} \lambda_j g_j(u^k) \right\}$$
(A3.4.67)
s.t. $\lambda_j \geq 0$, $\ \forall \ j \in I_k(u^k)$.

In this way, in order to calculate $\bar{\lambda}^k$, we deal with a quadratic programming problem with simple constraints. This problem can be solved in a finite number of iterations, for instance, by means of the conjugate gradient Method A3.3.29. Thus,
$$\nabla_u \mathcal{L}_k(\bar{u}^k, \bar{\lambda}^k) = 0,$$
implies
$$\bar{u}^k := u^k - \left[ \nabla J(u^k) + \sum_{j \in I_k(u^k)} \bar{\lambda}^k_j \nabla g_j(u^k) \right],$$
i.e., the iterate $\bar{u}^k$ can be simply calculated after the determination of the multiplier $\bar{\lambda}^k$.

For unconstrained problems this linearization method (with $\epsilon_k = 0$ $\ \forall \ k$ turns into a gradient method with constant step-length.

**A3.4.40 Remark.** In order to simplify the linearization method it seems to be natural to omit in the quadratic program (A3.4.64) the quadratic term of the objective. But usually the numerical algorithm breaks off, because the linear...
objective function can be unbounded on the polyhedral set. Some linearization
methods are known, in which solvability of the linearized auxiliary problems
and convergence of the iterates are obtained by looking for other ways and
means (see Mangasarian [284], Pironneau and Polak [321], which include
methods with linear auxiliary problems, too.

A3.4.3 Penalty methods
We pay special attention to this class of methods, because they are often used
for the construction of stable algorithms for ill-posed problems. Here they are
studied in a more general setting than given in Problem (A3.4.56). We consider
the problem

$$ \min \{ J(u) : u \in K \}, \quad (A3.4.68) $$

with $J : \mathbb{R}^n \to \mathbb{R}$ a convex function and $K = G \cap H$, where $G \subset \mathbb{R}^n$ is a convex,
closed set and $H \subset \mathbb{R}^n$ is an affine set such that $H \cap \text{int} G \neq \emptyset$.
We introduce families $\{ \phi^1_k \}$ and $\{ \phi^2_k \}$ of convex functions mapping from
$\mathbb{R}^n$ into $\mathbb{R}$. Their properties will be specialized in Theorem A3.4.41 below.

The principal idea of penalty methods consists in the following:
Choose a sequence $\{ \epsilon_k \} \downarrow 0$, set

$$ F_k = J + \phi^1_k + \phi^2_k, $$

and compute $u^k$ according to

$$ F_k(u^k) \leq \inf_{u \in \mathbb{R}^n} F_k(u) + \epsilon_k. \quad (A3.4.69) $$

A3.4.41 Theorem. Assume that the solution set $U^*$ of Problem (A3.4.68) is
not-empty and bounded and that the systems of convex functions $\{ \phi^1_k \}, \{ \phi^1_k \}$ have
the following properties

$$ \lim_{k \to \infty} \phi^1_k = \begin{cases} 0 & \text{if } u \in \text{int} G, \\ +\infty & \text{if } u \notin G, \end{cases} \quad (A3.4.70) $$

and

$$ \begin{align*}
\phi^2_k(u) &\geq 0, \quad u \in \mathbb{R}^n, \\
\lim_{k \to \infty} \phi^2_k(u) &\to 0, \quad u \in H, \\
\lim_{k \to \infty} \phi^2_k(v^k) &\to +\infty \quad \text{as } v^k \to u \notin H.
\end{align*} \quad (A3.4.71) \quad (A3.4.72) $$

Then, starting with some iteration, there exist points $u^k$ satisfying (A3.4.69),
moreover, the sequence $\{ u^k \}$ is bounded and each cluster point belongs to the
optimal set $U^*$.

Proof: For fixed $\theta > 0$ and $\delta > 0$ we define the sets

$$ \bar{K} := \{ u \in K : J(u) \leq J^* + \theta \} \quad (A3.4.73) $$

and

$$ K_\delta := \{ u \in \mathbb{R}^n : \rho(u, \bar{K}) \leq \delta \}, $$
which are bounded due to Proposition A1.7.55. Now, we choose a convex, compact set $S \subset \text{int}G \cap \text{int}K_{\delta}$ such that $H \cap \text{int}S \neq \emptyset$ and for every $u \in S$ the inequality $J(u) \leq J^* + \tau$ is true with some fixed $\tau \in (0, \frac{\alpha}{2})$. Let

$$\bar{z} \in H \cap \text{int}S, \quad w \in \partial K_{\delta}, \quad \bar{z} \in [\bar{z}, w] \cap \text{int}(K_{\delta} \setminus \hat{K}), \quad z \notin \partial G \cup S.$$ 

Then the following tree cases are possible:

(a) $z \notin G$: In this case, on the basis of (A3.4.70) and Proposition A1.5.19, starting with some iteration $k_1$, we get

$$J(z) + \phi_k(z) > J^* + \theta - \tau,$$

and

$$J(u) + \phi_k(u) \leq J^* + 2\tau, \quad \forall u \in S.$$ 

Consequently, for $k \geq k_1$ and $u \in S$ the inequality

$$J(z) + \phi_k(z) > J(u) + \phi_k(u) + \theta - 3\tau$$  \hspace{1cm} (A3.4.74)

is satisfies. Since the set $S_\lambda(z) := \{z + \lambda(z - u) : u \in S, \lambda \geq 0\}$ contains the point $w$ together with some neighborhood $O(w)$, we obtain, with regard to the convexity of $J + \phi_k$ and (A3.4.74),

$$J(v) + \phi_k(v) > J(z) + \phi_k(z) + \theta - 3\tau, \quad \forall k \geq k_1, \quad \forall v \in O(w).$$ 

Hence, in view of (A3.4.71), (A3.4.72), there exists a number $k_2 \geq k_1$ such that

$$F_k(v) > F_k(z) + \theta - 4\tau, \quad \forall k \geq k_2, \quad \forall v \in O(w).$$  \hspace{1cm} (A3.4.75)

(b) $z \notin H$: Due to (A3.4.70) and Proposition A1.5.19 we can conclude that for sufficiently large $k$ ($k \geq k_3$) and all $u \in S$ the estimate

$$|\phi_k(u)| < \frac{\rho_1 \tau}{2\rho_2}$$

is true with

$$\rho_1 := \rho(\bar{z}, \partial S), \quad \rho_2 := \max\{\|\bar{z} - \xi\| : \xi \in \partial K_{28}\}.$$

Hence, in view of the convexity of $\phi_k$, for arbitrary $q \in K_{28} \setminus S$ and $\eta \in [\bar{z}, q] \cap \partial S$, we obtain for $k \geq k_3$

$$\phi_k(q) \geq \frac{\|\bar{z} - q\|}{\|\eta - \bar{z}\|} \phi_k(\bar{z}) - \frac{\|q - \eta\|}{\|\eta - \bar{z}\|} \phi_k(z)$$

$$\geq \frac{\|\bar{z} - q\|}{\|\eta - \bar{z}\|} \left( -\frac{\rho_1 \tau}{2\rho_2} \right) - \frac{\|q - \eta\|}{\|\eta - \bar{z}\|} \left( \frac{\rho_1 \tau}{2\rho_2} \right)$$

$$\geq 2 \frac{\|\bar{z} - q\|}{\|\eta - \bar{z}\|} \left( -\frac{\rho_1 \tau}{2\rho_2} \right) \geq -\tau.$$
Since \( z \notin H \), it is obvious that \( w \notin H \). Hence, for a sufficiently small neighborhood \( \mathcal{O}(w) \subset S_\lambda(z) \), the relation
\[
\text{cl}(\mathcal{O}(w)) \cap H = \emptyset
\]  
(A3.4.76)
holds and
\[
\phi_k^1(v) \geq -\tau, \quad \forall \ v \in \mathcal{O}(w), \ \forall \ k \geq k_3.
\]
Increasing \( k_3 \) (if necessary), due to (A3.4.70), we convince ourselves that
\[
\phi_k^1(\bar{z}) \leq \tau \quad \forall \ k \geq k_3.
\]
The relations (A3.4.76), (A3.4.71), (A3.4.72) lead to
\[
J(v) + \phi_k^2(v) > J(\bar{z}) + \phi_k^2(\bar{z}) + \theta - 2\tau, \quad \forall \ v \in \mathcal{O}(w), \ \forall \ k \geq k_4 \geq k_3.
\]
Indeed, for \( \gamma_k := \inf_{u \in \mathcal{O}(w)} \phi_k^2(u) \), we obtain from (A3.4.76) and (A3.4.72)
\[
\lim_{k \to \infty} \gamma_k = +\infty,
\]
and (A3.4.71) implies that \( \phi_k^2(\bar{z}) = 0 \).
Consequently, inequality (A3.4.75) is satisfied for all \( v \in \mathcal{O}(w) \) if \( k \) is chosen large enough.

(c) \( z \in G \cap H \): Due to the choice of \( z \) and (A3.4.73) the inequality
\[
J(z) \geq J^* + \theta
\]
holds and there exists a number \( k_5 \) such that for \( k \geq k_5 \)
\[
\phi_k^1(z) \geq -\tau, \quad \phi_k^1(u) \leq \tau, \quad \forall \ u \in S.
\]
Thus, we get
\[
J(z) + \phi_k^1(z) > J(u) + \phi_k^1(u) + \theta - 3\tau, \quad \forall \ k \geq k_5,
\]
and the same inequality (A3.4.75) as in case (a) can be obtained.
Therefore, inequality (A3.4.75) is true in any case.

Furthermore, varying the point \( z \), we obtain a cover of the compact set \( \partial K_\delta \) by means of the neighborhoods \( \mathcal{O}(w) \) determined above. Choosing a finite subcover, we can state that, starting with some iteration \( k' \), inequality (A3.4.75) holds for all \( v \in \partial K_\delta \). Since \( \theta - 4\tau \geq 0 \ \forall \ k \geq k' \), the functions \( F_k \) attain their minima on \( \mathbb{R}^n \) only in points belonging to \( K_\delta \).
Moreover, if \( \theta \) and \( \tau \) satisfy the relation \( \theta - 4\tau > \sum_{k \geq k'} \epsilon_k \), then
\[
u^k \in K_\delta \ \forall \ k \geq k'.
\]
But with regard to the arbitrarily chosen \( \delta > 0 \), the cluster points of the sequence \( \{u^k\} \) must belong to \( K \subset K \).
Now, varying \( \theta \) and \( \tau \) such that \( \theta - 4\tau > 0, \ \theta \downarrow 0 \), (we emphasize that \( k' \) depends on \( \tau \) and \( \theta \) and \( k' \) may tend to \( +\infty \) if \( \tau \downarrow 0 \)), relation (A3.4.73) provides that
\[
\min_{k \to \infty} \rho(u^k, U^*) = 0.
\]
\( \Box \)
Assuming $F_k \in C^1(\mathbb{R}^n)$, an analogous result is true if, instead of (A3.4.69), the stopping criterion

$$\| \nabla F_k(u^k) \| \leq \epsilon_k, \quad \epsilon_k \downarrow 0$$

is used. Following the proof of Lemma 3.1 in Grossmann and Kaplan [151] it is easily seen that under the assumptions of Theorem A3.4.41 the condition

$$\lim_{k \to \infty} \| \nabla F_k(u^k) \| = 0$$

leads to

$$\lim_{k \to \infty} [F_k(u^k) - \inf_{u \in \mathbb{R}^n} F_k(u)] = 0.$$ 

Now, let us specify the penalty functions $\phi^1_k$ and $\phi^2_k$. Assume the feasible set $K = G \cap H$ is described by

\begin{align*}
G & := \{ u \in \mathbb{R}^n : g_j(u) \leq 0, \quad j = 1, \ldots, m \}, \\
H & := \{ u \in \mathbb{R}^n : h_i(u) = 0, \quad i = 1, \ldots, m_1 \},
\end{align*}

with $g_j$ convex and $h_i$ affine functions, and if in addition there exists a Slater point $\tilde{u} \in H$ with $g_j(\tilde{u}) < 0$, $(j = 1, \ldots, m)$, then

$$\phi^2_k(u) := r_k \sum_{k=1}^{m_1} |h_i(u)|^2$$

is chosen usually.

As for the penalty function $\phi^1_k$, satisfying condition (A3.4.70), the following ones are often applied in practice:

$$\phi^1_k(u) := r_k \sum_{k=1}^{m} \max\{0, g_j(u)\}^s, \quad s \geq 1, \quad (A3.4.77)$$

$$\phi^1_k := \left\{ \begin{array}{ll}
- \frac{1}{r_k} \sum_{j=1}^{m} \frac{1}{g_j(u)} & \text{if } u \in \text{int}G, \\
+ \infty & \text{otherwise.} \end{array} \right. \quad (A3.4.78)$$

Sometimes, the latter one is called barrier function.

In order to satisfy the hypothesis of Theorem A3.4.41 one has to guarantee that $r_k \downarrow 0$.

In the sequel we suppose that there are no equality constraints and describe some estimates of the rate of convergence for the penalty method with penalty functions of the type (A3.4.77) and (A3.4.78).

**A3.4.42 Proposition.** For the penalty method the following estimates are true:

(i) Using penalty function (A3.4.77) with $s = 2$:

$$-\|\tilde{\lambda}\| \sqrt{\frac{1}{r_k} \|\tilde{\lambda}\|^2 + \frac{1}{r_k} 2\epsilon_k} \leq J(u^k) - J^* \leq \epsilon_k, \quad (A3.4.79)$$

where $\tilde{\lambda}$ is a Lagrange multiplier vector.
(ii) Using function (A3.4.78):

\[
J(u^k) - J^* < \frac{2}{\sqrt{r_k}} \sqrt{m\sigma_0[J(\tilde{u}) - J^*]} + \epsilon_k,
\]

with

\[
r_k > \frac{m\sigma_0}{J(\tilde{u}) - J^*}, \quad \sigma_0 := \frac{1}{\min_{1 \leq j \leq m} |g_j(\tilde{u})|}
\]

and \(\tilde{u}\) a Slater point.

On the class of problems considered the both estimates cannot be improved with respect to the order. For general approaches to a priori estimates of the rate of convergence for penalty methods we refer to Lootsma [277], Skarin [373] and Grossmann and Kaplan [151]. It should be noted that the conditioning of unconstrained minimization problems generated by penalty methods, becomes worse within the increase of the iterations. Often these methods are used as an effective tool to compute a comparatively hoarse solution of an optimization problem. However, if a more precise solution is required, then special modifications are needed.

The following a posteriori estimates for the rate of convergence provide a convenient stopping criterion for the penalty methods considered above, see [151].

**A3.4.43 Proposition.** Suppose that \(J \in C^1(\mathbb{R}^n)\), \(g_j \in C^1(\mathbb{R}^n), (j = 1, ..., m)\), \(\phi_k := \varphi_k(g(\cdot))\), with \(\varphi_k \in C^1(\mathbb{R}^m)\). Furthermore, assume that the hypotheses of Theorem A3.4.41 and the Slater condition are fulfilled. The functions

\[\eta^j_k(\cdot) := \frac{\partial \varphi_k(g(\cdot))}{\partial g_j}\]

are assumed to be non-negative on \(\mathbb{R}^n\) and for each \(j\) and \(\{z^k\} \subset \mathbb{R}^n\) the sequence \(\{\eta^j_k(z^k)\}\) converges to zero if

\[
\lim_{k \to \infty} \inf g_j(z^k) > -\infty, \quad \lim_{k \to \infty} \sup g_j(z^k) < 0.
\]

Then, for the iterates \(\{u^k\}\), generated according to

\[\|\nabla F_k(u^k)\| \leq \epsilon_k, \quad \epsilon_k \downarrow 0,
\]

the following holds true:

(i) setting \(\lambda^k := \eta^k(u^k)\) and \(\lambda^k = (\lambda^k_1, ..., \lambda^k_m)^T\), the sequence \(\{(u^k, \lambda^k)\}\) is bounded and each cluster point is a saddle point of the corresponding Lagrangian function;

(ii) the estimate

\[J^* \geq J(u^k) + \langle \lambda^k, g(u^k) \rangle - \epsilon_k \rho(u^k, U^*)\]

is true.
If $u^k \in K$, then in view of (A3.4.81), we get

$$J(u^k) \geq J^* \geq J(u^k) + \langle \lambda^k, g(u^k) \rangle - \epsilon_k \rho(u^k, U^*),$$

(A3.4.82)

and

$$\langle \lambda^k, g(u^k) \rangle - \epsilon_k \rho(u^k, U^*) \to 0, \quad k \to \infty.$$

Instead of the unknown value $\rho(u^k, U^*)$ an upper estimate can be applied.

If $u^k \notin K$, then by means of a linear interpolation of the function

$$\bar{g}(\cdot) := \max_{1 \leq j \leq m} g_j(\cdot)$$

on the interval $[\bar{u}, u^k]$, a point $z^k \in K$ can be easily defined such that

$$J(z^k) - J(u^k) \to 0,$$

and estimate (A3.4.82) can be replaced by

$$J(z^k) \geq J^* \geq J(u^k) + \langle \lambda^k, g(u^k) \rangle - \epsilon_k \rho(u^k, U^*).$$

(A3.4.83)

**A3.4.44 Remark.** Problem (A3.4.56) is equivalent to the unconstrained minimization of the non-differentiable function

$$F(u) := J(u) + r \sum_{i=1}^{m} \max[0, g_i(u)]$$

if $r > \max_{1 \leq j \leq m} |\bar{\lambda}|$, with $\bar{\lambda}$ any Lagrange multiplier vector of the problem. Hence, using penalty function (A3.4.77) with $s = 1$, it is not necessary to increase the penalty parameter $r$. However, in that case special methods of non-differentiable minimization must be applied, see Remark A3.2.24.

A3.4.4 Dual methods

The statement of the Kuhn-Tucker Theorem A1.7.51 about the equivalence of convex optimization problems and corresponding saddle point problems is the foundation for a number of methods in which a saddle point of a convex-concave function is sought under simple constraints, only non-negativity of the dual variables is required.

Let us consider Problem (A1.7.35), given for $V \equiv \mathbb{R}^n$, and the corresponding saddle point problem (A1.7.36), (A1.7.37).

One of the first *multiplier method* for solving this kind of problems is the Arrow-Hurwicz-Uzawa method:

$$u^{k+1} := u^k - \alpha \nabla_u \mathcal{L}(u^k, \lambda^k),$$

$$\lambda^{k+1} := [\lambda^k + \alpha \nabla_\lambda \mathcal{L}(u^k, \lambda^k)]_+, \quad \text{(A3.4.84)}$$

with $\alpha > 0$ suitably chosen, $([y]_+ := \max[0, y])$.

Hence, the pair of iterates $(u^{k+1}, \lambda^{k+1})$ can be calculated by means of a gradient step with respect to the $u$ component and a gradient projection step with respect to the component $\lambda$. 
In the same paper another multiplier method is described which requires the fully minimization of the Lagrangian function \( L(u, \lambda_k) \) instead of one gradient step in each iteration. In other words, a gradient projection method for solving the dual problem (cf. (A1.7.38))

\[
L(\lambda) := \inf_{u \in \mathbb{R}^n} L(u, \lambda) \to \max_{\lambda \geq 0},
\]

has to be considered.

The analysis of methods based on the implementation of Lagrangian functions shows slow convergence and a comparatively limited field of applications. Their convergence is ensured under sufficiently strict requirements for the initial problem, mainly caused by the instability of the set of saddle points of the Lagrangian function with respect to block relaxation as well as by the insufficient smoothness of the corresponding function \( L \).

Indeed, without additional assumptions, for instance strict convexity of \( J \), we have in general

\[
\text{Arg min}_{u \in \mathbb{R}^n} L(u, \lambda^*) \neq U^*, \quad (\lambda^* \text{ Langrange multiplier vector})
\]

and \( L \) is not continuous on \( \mathbb{R}^n \) and not differentiable on \( \text{dom}(-L) \), even if the functions involved are analytic.

More efficient multiplier methods can be constructed by means of modifications of the Lagrangian function that overcome the defects mentioned above. Often the \textit{augmented Lagrangian function} \( L_A : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \)

\[
L_A(u, \lambda) := J(u) + \frac{1}{2r} \left\{ ||\lambda + rg(u)||^2 - ||\lambda||^2 \right\}
\]

(A3.4.85)
is suggested, with \( r > 0 \) arbitrarily, but fixed chosen. The set of saddle points of \( L_A \) on \( \mathbb{R}^n \times \mathbb{R}^m \) coincides with the set \( U^* \times \Lambda^* \) of saddle points of the original Lagrangian function \( L \) on \( \mathbb{R}^n \times \mathbb{R}_{+}^m \).

Moreover, function \( L_A \) has the following desirable properties:

(i) \( L_A \) is convex with respect to \( u \) and concave with respect to \( \lambda \);

(ii) \( \text{Arg min}_{u \in \mathbb{R}^n} L_A(u, \lambda^*) = U^* \);

(iii) \( L_M(\lambda) = \inf_{u \in \mathbb{R}^n} L_A(u, \lambda) \) is concave and differentiable on \( \mathbb{R}^m \) and its gradients fulfill a Lipschitz condition with constant \( L = \frac{1}{r} \).

If we replace the original Lagrangian function by the modified one, then a similar multiplier method can be described:

\[
u^{k+1} := u^k - \alpha \nabla_u L_A(u^k, \lambda^k),
\]

\[
\lambda^{k+1} := \lambda^k + \alpha \nabla_\lambda L_A(u^k, \lambda^k),
\]

(A3.4.86)

If \( \alpha := r \), then \( \lambda^{k+1} := [\lambda^k + rg(u^k)]_+ \) can be chosen.

A3.4.45 Proposition. Let the assumptions for Problem (A3.4.56) be fulfilled. Assume further that the gradients of the functions \( J, g_j \) satisfy a Lipschitz condition with a common Lipschitz constant \( L(\delta) \) on the ball \( B_\delta := \{ u \in \mathbb{R}^n : ||u|| \leq \delta \} \) for any \( \delta > 0 \). Then, for

\[
\delta^2 \geq ||u^0||^2_{\mathbb{R}^n} + ||\lambda^0||^2_{\mathbb{R}^m},
\]
there exists a constant \( \alpha(\delta) \) such that the sequence \( \{(u^k, \lambda^k)\} \), generated by Method (A3.4.86) with \( \alpha \in (0, \alpha(\delta)] \), converges to some saddle point \((u^*, \lambda^*) \in U^* \times \Lambda^*\).

The constant \( \alpha(\delta) \) can be efficiently calculated if \( L(\alpha) \) is known (cf. Gol’stein and Tretyakov [141]).

If, for some solution \( u^* \) of Problem (A3.4.56), the vectors \( \{\nabla g_j(u^*)\}_{j \in I_0(u^*)} \),

\[ I_0(u^*):=\{i: g_i(u^*) = 0\}, \]

are linearly independent, then the Lagrange multiplier vector \( \lambda^* \) is unique. This is a conclusion of the Kuhn-Tucker theorem.

**A3.4.46 Proposition.** Suppose in addition to the assumptions of Proposition A3.4.45 that

(i) \( J \) and \( g_j \) are twice-differentiable on \( \mathbb{R}^n \) and their second derivatives satisfy a Lipschitz condition;

(ii) \( J \) is strongly convex;

(iii) the vectors \( \{\nabla g_j(u^*)\}_{j \in I_0(u^*)} \) are linearly independent and

\[ \lambda^*_j > 0 \ \forall \ j \in I_0(u^*). \]

Then, starting with an arbitrary point \((u^0, \lambda^0)\), a constant \( \bar{\alpha} > 0 \) exists such that the iterates \( \{(u^k, \lambda^k)\} \), generated by Method (A3.4.86) with \( \alpha \in (0, \bar{\alpha}) \), converge linearly to \((u^*, \lambda^*) \in U^* \times \Lambda^*\).

Due to property (iii) of the augmented Lagrangian function \( L_A \), it is possible to use a gradient method to maximize \( L_M \). In this situation we have to calculate \( \arg \min_{u \in \mathbb{R}^n} L_A(u, \lambda^k) \) for each \( \lambda^k \).

Let us briefly describe this method under the assumption that

\[ \lambda^* \]

is calculated approximately:

\[ u^k: \quad L_A(u^k, \lambda^k) \leq \min_{u \in \mathbb{R}^n} L_A(u, \lambda^k) + \min\left[ \delta_k, \beta \|\nabla_A L_A(u^k, \lambda^k)\|^2 \right], \]

\[ \lambda^{k+1} := \lambda^k + \alpha \nabla_A L_A(u^k, \lambda^k), \]

(A3.4.87)

with \( \alpha \in (0, 2r) \), \( \beta \geq 0 \), \( \delta_k \geq 0 \) fixed and \( \sum_{k=1}^{\infty} \delta_k < \infty \).

**A3.4.47 Proposition.** Let the hypothesizes of Proposition A3.4.46 be fulfilled. Then for arbitrary \( \lambda^0 \geq 0 \) and sufficiently small \( \beta \) the sequence \( \{(u^k, \lambda^k)\} \), generated by Method (A3.4.87), converges linearly to \((u^*, \lambda^*) \in U^* \times \Lambda^*\).

Choosing \( \alpha := r \) the factor of the geometric progression is \( q = \alpha(\frac{1}{r}) \).

Hence, factor \( q \) can be made as small as desired by means of the chosen parameter \( r \) in function (A3.4.85). But one should be aware that an increase of \( r \) makes the computation of \( u^k \) more difficult, because the condition number of the Hessian of \( L_A(\cdot, \lambda) \) becomes worse.
Bibliography


[198] Jarre, F. Comparing two interior-point approaches for semi-infinite programs.


Index

algorithm
  splitting, 420, 440
approximation
  inner, 265
  Mosco, 47
Armijo search, 490
auxiliary problem principle, 419

barrier algorithm, 204
BFGS class, 499
bilinear form, 445
  coercive, 454
  symmetric, 445
bisection method, 487
block relaxation, 503
boundary
  Lipschitz, 446
boundary coercivity, 389
boundary-obstacle problem, 471
Bregman function, 388
Bregman-like function, 395
Broyden class, 499
bundle method, 93, 372

Cauchy Problem, 2
closure, 444
cocercivity, 454
  boundary, 389
  zone, 389
cone
  normal, 462
  of feasible directions, 506
conjugate
  gradient method, 493, 504
  operator, 456
  space, 443
  vectors, 493
constraint qualification, 466
contact zone, 289
controlling parameter
  choice of, 131
convection-diffusion problem, 357
convergence, 444
  linear, 486
  quadratic, 486
  superlinear, 486
  to optimal value, 485
  to set, 485
  weak, 444
convergence sensing condition, 389
convexification, 82
coordinate descent method, 502
Davidon-Fletcher-Powell method, 499
derivative
  directional, 455
  Fréchet, 449
  Gâteaux, 449
  generalized, 447
direction search, 507
Dirichlet Problem, 351, 470
displacements
  admissible, 291
  virtual, 291
distance function, 402
domain
  effective, 445
dual method, 516
dual problem, 285, 377
duality pairing, 443
Dunn property, 421, 424, 464
Einstein's summation convention, 290
elasticity coefficient, 290
elliptic regularization method, 339
ellipticity property, 290
embedding
  compact embedding, 448
  continuous embedding, 448
epigraph, 451
exchange method, 168
feasible direction, 506
field
  of displacements, 290
  of strains, 290
finite element
  discretization, 471
  space, 472
function
  barrier, 514
  Bregman, 388
  Bregman-type, 388
  DC, 94
  distance, 389, 402
  entropy-like, 402
  generalized distance, 406
  logarithmic-quadratic, 402
  Lyapunov, 404
  penalty, 514
functional
  concave, 451
  continuous, 445
    weakly, 445
  convex, 451
  Hölder continuous, 445
  indicator, 16
  Lipschitz, 445
  M-convergence, 47
  Moreau-Yosida, 8
  proper convex, 451
  semicontinuous
    lower, 445, 451
    upper, 445
    weakly lower, 445
    weakly upper, 445
  strictly convex, 454
  strongly convex, 454
  trace, 448
  uniformly convex, 454
gap function, 370
Gauss-Seidel method, 503
Gelfand triple, 446
generalized distance, 406
generalized proximal point method, 340, 408
Goldstein search, 490
gradient method, 492
gradient projection method, 30, 503
gradient-type methods, 489
growth condition, 347
Hausdorff distance, 468
Hooke’s law, 289
inequality
  Cauchy-Schwarz, 443
  Friedrichs, 447
  Jensen, 452
  Korn, 303
  Poincaré, 335
inner approximation
  set, 477
  space, 471
interior proximal method, 387
interpolant, 473
Korn’s inequality, 303
Krylov subspace, 501
Lagrangian multiplier, 466
Lagrangian function, 466
  augmented, 89, 517
lemma
  Opial, 444
  Polyak, 486
linearization method, 509
Lyapunov function, 404
mapping
  convex, 452
  proximal point, 22
  proximal-point, 55
method
  Armijo search, 490
  augmented Lagrangian, 73
  bisection, 487
  block relaxation, 503
  Bregman-function-based, 390
  Broyden, 499
  bundle, 372
  conjugate gradient, 493, 504
  coordinate descent, 502
  Davidon-Fletcher-Powell, 499
  dual, 516
  elliptic regularization, 339
  exact PPR, 387, 404
  feasible directions, 507
  Fletcher-Reeves, 494
  Gauss-Seidel, 503
  generalized proximal point, 340, 408
  Goldstein search, 490
INDEX

gradient, 492
gradient projection, 30, 503
gradient-type, 489
implicit Euler, 77
interior proximal, 387
iterative regularization, 38
Levenberg-Marquard, 78
linearization, 79, 509
multiplier, 516
Newton, 489, 497
penalty, 32, 511
proximal auxiliary problem, 421
proximal gradient, 60
proximal point, 57
proximal point regularization, 23
Quasi-Newton, 498
quasi-solution, 15
regularized barrier, 68
regularized multiplier, 89
regularized penalty, 42, 60, 247
regularized subgradient projection, 39
residual, 15
steepest descent, 489, 491
Tikhonov regularization, 15
trust-region, 90
variable metric, 500
minimal surface problem, 351
with obstacle, 351
minimizing sequence, 485
model, 90
Moreau-Yosida functional, 8
Moreau-Yosida-Regularization, 56
Moroeau-Yosida, 439
Mosco convergence, 47
Mosco scheme, 47
MSR-method, 112
multi-step regularization, 38, 112, 248
Newton method, 489, 497
norm, 443
operator, 445
seminorm, 443
normal solution, 20
obstacle problem, 470
one-step regularization, 38, 101, 247
operator
\( \epsilon \)-enlargement, 367, 390
co-coercive, 464
coercive, 464
divergence, 285
effective domain, 461
firmly non-expansive, 55
graph, 461
inverse, 461
locally bounded, 461
locally hemi-bounded, 411, 461
locally hemi-continuous, 461
maximal monotone, 462
monotone, 461
multi-valued, 461
non-expansive, 56
normality, 462
paramonotone, 388, 414, 463
pseudo-monotone, 414, 463
range, 461
resolvent, 365
single-valued, 461
strongly monotone, 461
weakly coercive, 464
operator equation of first kind, 15
OSR-method, 101
modified, 109
outward normal, 449
penalty method, 32, 60, 247, 511
Poincaré inequality, 335
precondition, 497
problem
boundary control, 244
Cauchy, 2
chattering, 266
convection-diffusion, 357
Dirichlet, 351, 470
distributed control, 243
dual, 377, 466
Fuller, 279
ill-posed, 2, 6
inclusion, 419
inverse, 4
minimal surface, 351
obstacle on the boundary, 283, 471
Poisson, 270
quadratic programming, 376, 505
saddle point, 287, 466
semi-infinite programming, 167
Signorini, 288
Signorini with friction, 328
singularly perturbed, 357
INDEX

trust-region, 91

two-body contact, 288

variational, 453

weakly well-posed, 6

well-posed, 1, 6

proximal gradient methods, 60

proximal point method

exact, 57

inexact, 59

Quasi-Newton method, 498

quasi-solution, 15

quasi-solution method, 15

rank-1-update, 499

rank-2-update, 499

reduction set, 468

regularization

multi-step, 38

on subspace, 247, 297, 410

one-step, 38

weak, 410

with weaker norm, 298, 307, 322, 334

relaxation, 366

relaxation parameter, 503

residual method, 15

saddle point, 466

scalar product, 443

semi-infinite programming, 167

set

compact, 445

dense, 445

feasible, 453

inner approximation, 477

interior, 451

M-convergence, 47

of feasible directions, 458

of well-posedness, 2, 15

optimal, 453

relative interior, 451

solution set, 453

weakly compact, 445

SIP

$T$-well-posed, 7

convex, 7, 167

discretization approach, 168

exchange approach, 168

linear, 4

reduction approach, 170

Slater condition, 466, 467

space

Hilbert, 447

inner approximation, 265, 471

Sobolev, 447

stabilizer, 14

weak, 14

steepest descent method, 491

step-length, 488

stress tensor, 290

subdifferential, 455, 462

subgradient, 455

subgradient method, 420

support, 446

Taylor formula, 450

theorem

Kuhn-Tucker, 467

generalized Weierstrass, 445

mean value, 450

Moreau-Rockafellar, 411, 456

Riesz, 444

Tikhonov regularization, 15

trace, 283

triangulation

quasi-uniform, 472

uniform, 474

unilateral constraint, 287

variable metric method, 500

variational inequality, 459

hemi-, 421, 437, 459

weak closure, 444

weak limit, 444

weak solution, 292

zone coercivity, 389