Datenkompression: Verlustbehaftete Komprimierungsverfahren

Einführung

H. Fernau

email: fernau@uni-trier.de

WiSe 2008/09

Universität Trier

Verlustbehaftete Komprimierung: Grundlagen

$$\underset{\text{Eingabe}}{\mathcal{X}} \xrightarrow{C} \underset{\text{Repräsentation}}{\mathcal{X}_c} \xrightarrow{D} \underset{\text{Rekonstruktion}}{\mathcal{Y}}$$

Ein Verfahren (C, D) heißt *verlustbehaftete Komprimierung*, wenn $\mathcal{X} \neq \mathcal{Y}$.

Maße für die Qualität der DK-Verfahren

- Kompressionsquotient,
- Geschwindigkeit der Komprimierung,
- Geschwindigkeit der Dekomprimierung,
- *Entstellung* oder *Verzerrung* (engl. distortion): $\mathcal{Y} - \mathcal{X}$.

Typische *Verzerrungsmaße*

quadratisches Fehlermaß	$d(x_n, y_n) = (x_n - y_n)^2$
Betragsfehlermaß	$d(x_n, y_n) = \lvert x_n - y_n \rvert$
mittlerer quadratischer Fehler (MSE)	$\sigma^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n - y_n)^2$
mittlerer Betragsfehler (AD)	$d_1 = \frac{1}{N} \sum_{n=1}^{N} \lvert x_n - y_n \rvert$
Verhältnis von Signal zu Verzerrung	$SNR = \frac{\sigma_X^2}{\sigma^2}$
SNR: signal-to-noise	σ_X^2 : (MSE des Signals)
SNR logarithmisch skaliert [Dezibel]	$SNR(dB) = 10 \log_{10} \frac{\sigma_X^2}{\sigma^2}$
PSNR (Peak SNR) log. skal. [Dezibel]	$SNR(dB) = 10 \log_{10} \frac{\text{max-pixel}^2}{\sigma^2}$

Zwei Beispiele

Verzerrung am Beispiel von JPEG

Bild	Bit Map	GIF	JPEG
Brücke	65 536	76 511	17 272
Kamera	65 536	55 484	10 889

$\mathcal{X}$	$\mathcal{Y}$	σ^2	d_1
Brücke	JPEG-Repräsent. für Brücke	44,20	5,07
Kamera	JPEG-Repräsent. für Kamera	22,34	3,07
Brücke	Kamera	6 674,66	66,93

Problem: Wie korrelieren diese Maße mit der wahrgenommenen Störung?

Original Bild Lena

Komprimiertes Bild mit jpeg 25%

Exkurs: StirMark

Fabien A. P. Petitcolas, Ross J. Anderson, Markus G. Kuhn. Attacks on copyright marking systems, in David Aucsmith (Ed), Information Hiding, Second International Workshop, IH'98, Portland, Oregon, U.S.A., April 15-17, 1998, Proceedings, LNCS 1525, Springer-Verlag, ISBN 3-540-65386-4, pp. 219-239.

Digitale Wasserzeichen dienen u.a. zum Kopierschutz:
Geistiges Eigentum wird so geschützt, indem z.B. Bilder mit einer für das menschliche Auge unsichtbaren Information durchzogen werden.

Hinweis: Vorlesungen zur Kryptographie (speziell Steganographie)

Exkurs: Einfachster StirMark-Trick:

Entferne bzw. Dopple einige Spalten im Original

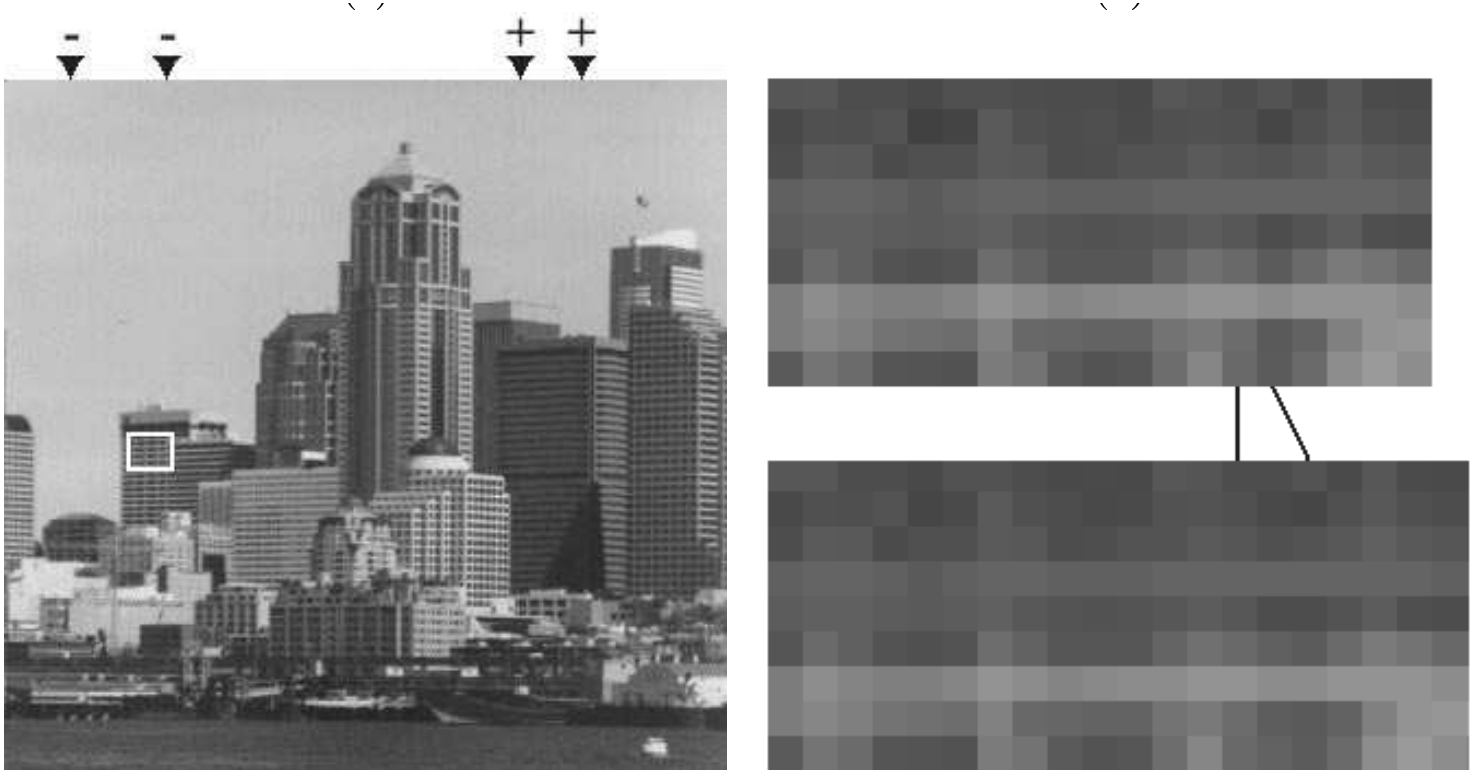


Bild nach $N(0, 0.05)$ Rauschen

Bild nach StirMark-Angriff (mit $q = 2.0$)

Bild	Lena vs. Bild AD	MSE	SNR	PSNR
Lena	0.0000	0.0000	$+\infty$	$+\infty$
`jpeg 25%`	0.0143	0.0004	28.4136	33.4729
$N(0, 0.05)$ Rauschen	0.0398	0.0025	20.5478	25.6071
`StirMark 2.0`	0.0580	0.0113	13.9883	19.0476

Ein praktischer Ansatz bei der Bildkompression

— "Verluste" kann man sich am ehesten bei Graustufenbildern und bei Farbbildern "leisten".
— Einfachste Idee: die fortschreitende Bildübertragung legt nahe, zunächst nicht alle Bits für jede Graustufe / Farbgebung zu übertragen, sondern nur "die wichtigsten".
— Alternative Interpretation: Betrachte n-te Stufe der fortschreitenden Bildübertragung als "verlustbehaftete Näherung"
$\rightsquigarrow$ Aus dem Übertragungsproblem für Farbbilder (mit n Bits Farbinformation pro Pixel) werden n Übertragungsprobleme für "Schwarz-Weiß-Bilder". (*Bitebenen*)
Problem/Frage: Sind "ähnliche Farben" auch "bitweise ähnlich"? Selbst zwei "benachbarte Bytes" können sich um 7 Bits unterscheide. . .

Gray-Codes zur Auflistung von allen n-Bit-Zahlen, sodass aufeinanderfolgende Zahlen sich um genau ein Bit unterscheiden.

Für 1-Bit-Zahlen: $(0, 1)$.
Ist Gray-Code für k-Bit-Zahlen bekannt, $(x_1, \ldots, x_{2^k})$, so entsteht die Folge $(y_1, \ldots, y_{2^k}, z_1, \ldots, z_{2^k})$ der Gray-Codes für $(k+1)$-Bit-Zahlen wie folgt:
(1) y_i entsteht durch Voranstellen einer Null aus x_i.
(2) z_{2^k+1-i} entsteht durch Voranstellen einer Eins aus x_i.

Da y_{2^k} und z_1 beide aus x_{2^k} durch Konkatenation eines Bits entstehen, folgt die gewünschte Eigenschaft leicht durch Induktion.

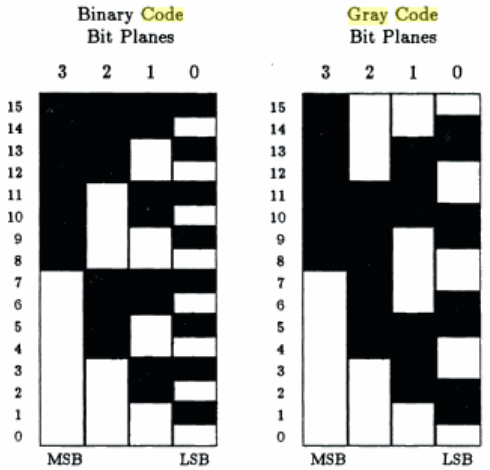
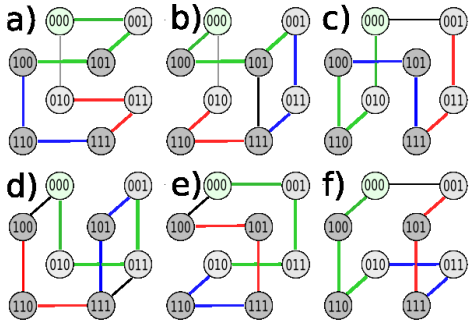


Figure 6.1: 4-bit binary and Gray codes.

Wir nutzen Graycodes zur Verschlüsselung von z.B. Farbinformationen vor dem Aufspalten in *Bitebenen*.
Technische Anwendung bei Zahlenübertragung, damit das "Kippen" einzelner Bits unerheblich ist.

Mehr Gray-Codes. . .

Hamiltonsche Pfade in Hyperwürfeln



binär	a)	b)	c)	d)	e)	f)
000	000	000	000	000	000	000
001	001	100	010	010	001	100
010	101	101	110	011	011	110
011	100	001	100	001	010	010
100	110	011	101	101	110	011
101	111	111	111	111	111	111
110	011	110	011	110	101	101
111	010	010	001	100	100	001

Variante e) entspräche dem bisher vorgestellten Gray-Code, der aufgrund seiner rekursiven Erzeugung bevorzugt wird und daher auch *reflektierter Gray-Code* heißt.

Verschiedene Bitebenen

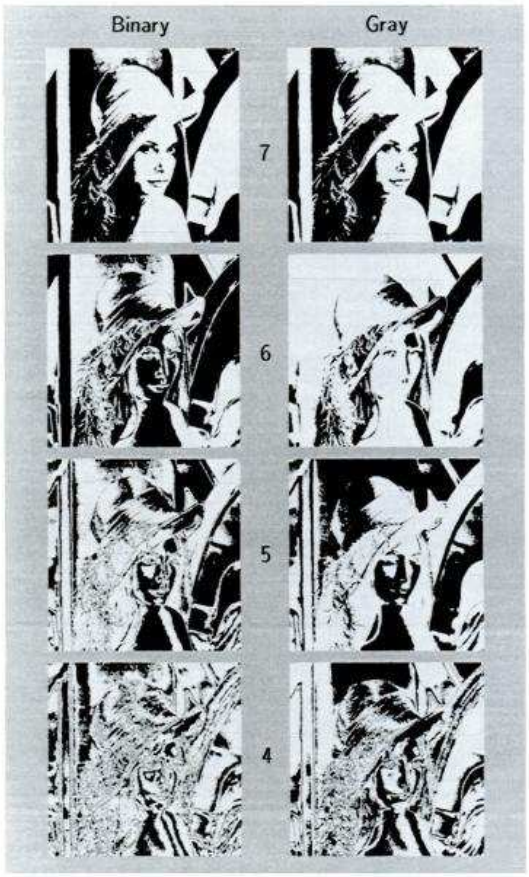


Figure 6.2: Binary and Gray code bit planes for LENA, (BP 7 through 4).

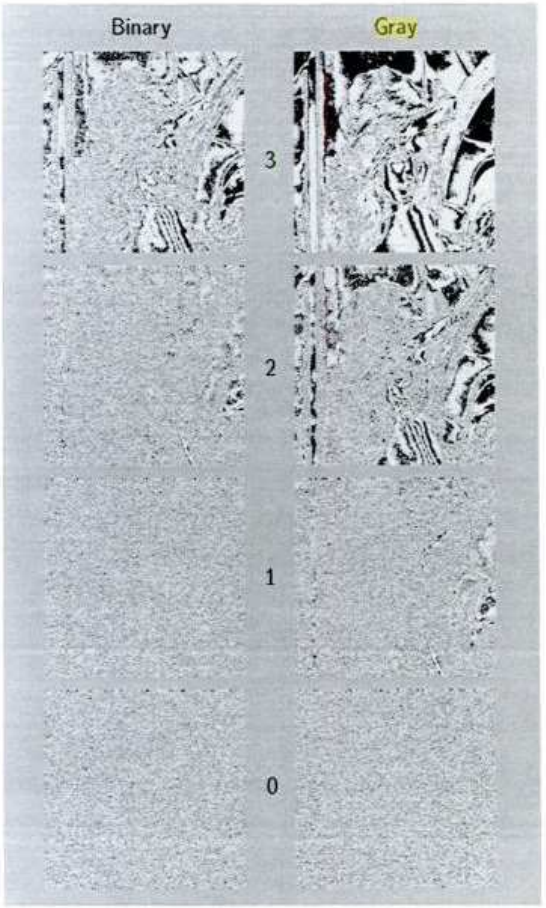


Figure 6.3: Binary and Gray code bit planes for LENA, (BP 3 through 0).

Vorteile in Zahlen

Bit Plane	1-D Runlength Entropy		7-pixel 2-D Q-Coder Bit Rate		7-pixel 3-D Q-Coder Bit Rate	
	BC	GC	BC	GC	BC	GC
LENA						
BP 7	0.274	0.274	0.144	0.144	0.144	0.144
BP 6	0.452	0.285	0.279	0.150	0.162	0.153
BP 5	0.614	0.391	0.450	0.235	0.263	0.236
BP 4	0.838	0.645	0.736	0.490	0.541	0.493
BP 3	0.965	0.802	0.955	0.698	0.753	0.692
BP 2	0.999	0.966	1.039	0.957	0.976	0.931
BP 1	1.000	0.999	1.043	1.039	1.040	1.028
BP 0	1.000	1.000	1.043	1.042	1.042	1.042
Total	6.142	5.362	5.689	4.755	4.921	4.719
BOOTS						
BP 7	0.292	0.292	0.187	0.187	0.187	0.187
BP 6	0.487	0.344	0.370	0.228	0.234	0.228
BP 5	0.690	0.566	0.607	0.454	0.479	0.445
BP 4	0.829	0.715	0.795	0.631	0.663	0.623
BP 3	0.915	0.839	0.933	0.808	0.836	0.797
BP 2	0.942	0.918	0.982	0.938	0.948	0.924
BP 1	0.945	0.942	0.986	0.981	0.981	0.975
BP 0	0.946	0.945	0.988	0.987	0.986	0.985
Total	6.046	5.561	5.848	5.214	5.316	5.164

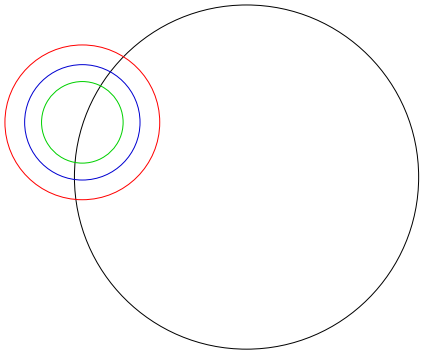
Table 6.1: Runlength entropies and Q-coder bit rates for bit plane encoding.

Der Q-Codierer ist eine Variante der arithmetischen Codierung.
Als Kontextinformationen kann entweder die Bitebene alleine (2-dimensionale Sicht) oder auch “benachbarte Bitebenen” herangezogen werden (3D).

Türme von Hanoi und reflektierte Gray-Codes:

z.B. mit drei Scheiben

binär	e)	Vorschrift
000	000	
001	001	Bewege kleinste Scheibe (*)
010	011	Bewege mittlere Scheibe
011	010	Bewege kleinste Scheibe (*)
100	110	Bewege größte Scheibe
101	111	Bewege kleinste Scheibe (*)
110	101	Bewege mittlere Scheibe
111	100	Bewege kleinste Scheibe (*)

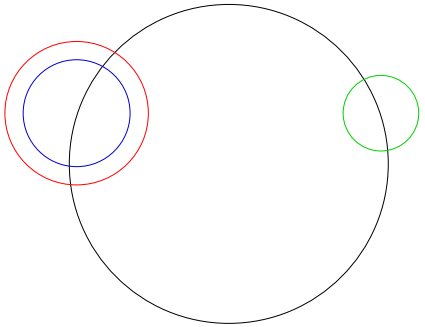


Nur Fall (*) ist nicht eindeutig, es gibt aber einfache Auflösung.
Seien die drei Stäbe auf einem Kreis angeordnet: links der Ausgangsstab, rechts der Zielstab und unten der Hilfsstab.
Ist # Scheiben gerade, so bewegt sich die kleine Scheibe entgegen dem Uhrzeigersinn; andernfalls bewegt sie sich im Uhrzeigersinn.

Türme von Hanoi und reflektierte Gray-Codes:

z.B. mit drei Scheiben

binär	e)	Vorschrift
000	000	
001	001	Bewege kleinste Scheibe (*)
010	011	Bewege mittlere Scheibe
011	010	Bewege kleinste Scheibe (*)
100	110	Bewege größte Scheibe
101	111	Bewege kleinste Scheibe (*)
110	101	Bewege mittlere Scheibe
111	100	Bewege kleinste Scheibe (*)

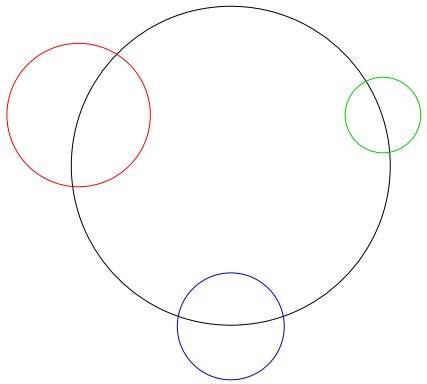


Nur Fall (*) ist nicht eindeutig, es gibt aber einfache Auflösung.
Seien die drei Stäbe auf einem Kreis angeordnet: links der Ausgangsstab, rechts der Zielstab und unten der Hilfsstab.
Ist # Scheiben gerade, so bewegt sich die kleine Scheibe entgegen dem Uhrzeigersinn; andernfalls bewegt sie sich im Uhrzeigersinn.

Türme von Hanoi und reflektierte Gray-Codes:

z.B. mit drei Scheiben

binär	e)	Vorschrift
000	000	
001	001	Bewege kleinste Scheibe (*)
010	011	Bewege mittlere Scheibe
011	010	Bewege kleinste Scheibe (*)
100	110	Bewege größte Scheibe
101	111	Bewege kleinste Scheibe (*)
110	101	Bewege mittlere Scheibe
111	100	Bewege kleinste Scheibe (*)

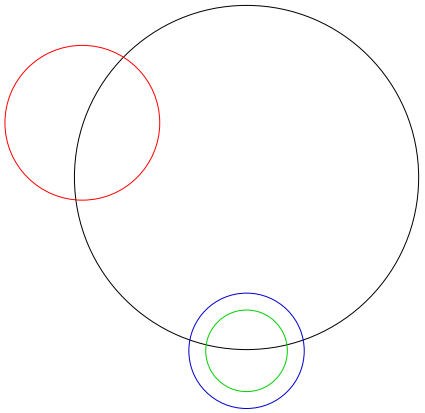


Nur Fall (*) ist nicht eindeutig, es gibt aber einfache Auflösung.
Seien die drei Stäbe auf einem Kreis angeordnet: links der Ausgangsstab, rechts der Zielstab und unten der Hilfsstab.
Ist # Scheiben gerade, so bewegt sich die kleine Scheibe entgegen dem Uhrzeigersinn; andernfalls bewegt sie sich im Uhrzeigersinn.

Türme von Hanoi und reflektierte Gray-Codes:

z.B. mit drei Scheiben

binär	e)	Vorschrift
000	000	
001	001	Bewege kleinste Scheibe (*)
010	011	Bewege mittlere Scheibe
011	010	Bewege kleinste Scheibe (*)
100	110	Bewege größte Scheibe
101	111	Bewege kleinste Scheibe (*)
110	101	Bewege mittlere Scheibe
111	100	Bewege kleinste Scheibe (*)

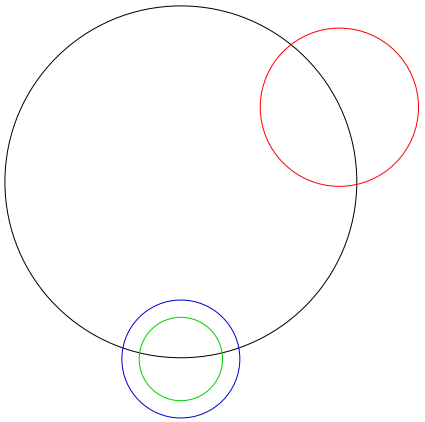


Nur Fall (*) ist nicht eindeutig, es gibt aber einfache Auflösung.
Seien die drei Stäbe auf einem Kreis angeordnet: links der Ausgangsstab, rechts der Zielstab und unten der Hilfsstab.
Ist # Scheiben gerade, so bewegt sich die kleine Scheibe entgegen dem Uhrzeigersinn; andernfalls bewegt sie sich im Uhrzeigersinn.

Türme von Hanoi und reflektierte Gray-Codes:

z.B. mit drei Scheiben

binär	e)	Vorschrift
000	000	
001	001	Bewege kleinste Scheibe (*)
010	011	Bewege mittlere Scheibe
011	010	Bewege kleinste Scheibe (*)
100	110	Bewege größte Scheibe
101	111	Bewege kleinste Scheibe (*)
110	101	Bewege mittlere Scheibe
111	100	Bewege kleinste Scheibe (*)

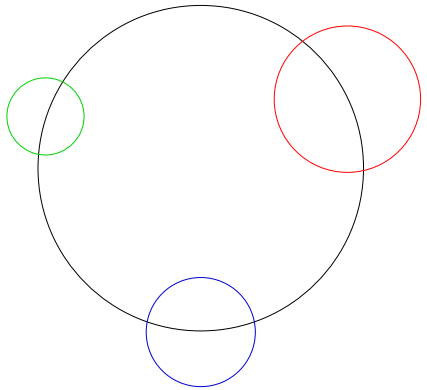


Nur Fall (*) ist nicht eindeutig, es gibt aber einfache Auflösung.
Seien die drei Stäbe auf einem Kreis angeordnet: links der Ausgangsstab, rechts der Zielstab und unten der Hilfsstab.
Ist # Scheiben gerade, so bewegt sich die kleine Scheibe entgegen dem Uhrzeigersinn; andernfalls bewegt sie sich im Uhrzeigersinn.

Türme von Hanoi und reflektierte Gray-Codes:

z.B. mit drei Scheiben

binär	e)	Vorschrift
000	000	
001	001	Bewege kleinste Scheibe (*)
010	011	Bewege mittlere Scheibe
011	010	Bewege kleinste Scheibe (*)
100	110	Bewege größte Scheibe
101	111	Bewege kleinste Scheibe (*)
110	101	Bewege mittlere Scheibe
111	100	Bewege kleinste Scheibe (*)

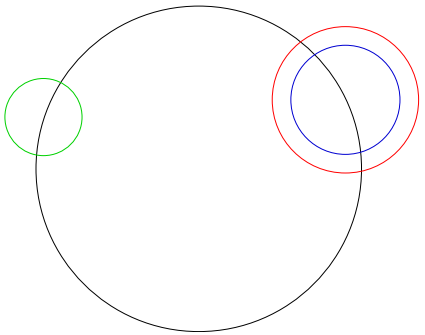


Nur Fall (*) ist nicht eindeutig, es gibt aber einfache Auflösung.
Seien die drei Stäbe auf einem Kreis angeordnet: links der Ausgangsstab, rechts der Zielstab und unten der Hilfsstab.
Ist # Scheiben gerade, so bewegt sich die kleine Scheibe entgegen dem Uhrzeigersinn; andernfalls bewegt sie sich im Uhrzeigersinn.

Türme von Hanoi und reflektierte Gray-Codes:

z.B. mit drei Scheiben

```
binär e)    Vorschrift
000   000
001   001   Bewege kleinste Scheibe (*)
010   011   Bewege mittlere Scheibe
011   010   Bewege kleinste Scheibe (*)
100   110   Bewege größte Scheibe
101   111   Bewege kleinste Scheibe (*)
110   101   Bewege mittlere Scheibe
111   100   Bewege kleinste Scheibe (*)
```

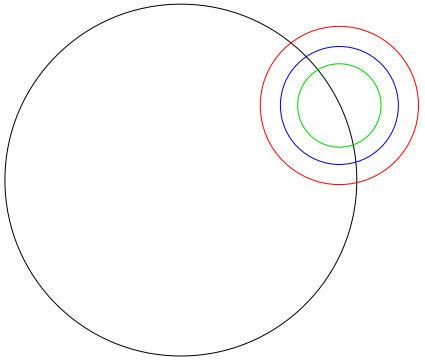


Nur Fall (*) ist nicht eindeutig, es gibt aber einfache Auflösung.
Seien die drei Stäbe auf einem Kreis angeordnet: links der Ausgangsstab, rechts der Zielstab und unten der Hilfsstab.
Ist # Scheiben gerade, so bewegt sich die kleine Scheibe entgegen dem Uhrzeigersinn; andernfalls bewegt sie sich im Uhrzeigersinn.

Türme von Hanoi und reflektierte Gray-Codes:

z.B. mit drei Scheiben

```
binär e)    Vorschrift
000   000
001   001   Bewege kleinste Scheibe (*)
010   011   Bewege mittlere Scheibe
011   010   Bewege kleinste Scheibe (*)
100   110   Bewege größte Scheibe
101   111   Bewege kleinste Scheibe (*)
110   101   Bewege mittlere Scheibe
111   100   Bewege kleinste Scheibe (*)
```



Nur Fall (*) ist nicht eindeutig, es gibt aber einfache Auflösung.
Seien die drei Stäbe auf einem Kreis angeordnet: links der Ausgangsstab, rechts der Zielstab und unten der Hilfsstab.
Ist # Scheiben gerade, so bewegt sich die kleine Scheibe entgegen dem Uhrzeigersinn; andernfalls bewegt sie sich im Uhrzeigersinn.

Mathematische Grundlagen

Es sei X eine Zufallsvariable für das Eingabesymbol aus dem Quellenalphabet Σ. Wir schreiben $P(X)$ für eine Zufallsfunktion s.d. $\forall a \in \Sigma\ P(a) = \Pr[X = a]$.

Entropie (zurückgehend auf Shannon)

$$H(X) = -\mathbb{E}[\log P(X)] = -\sum_{a \in \Sigma} P(a) \log P(a).$$

Mehrere Zufallsvariablen: es sei Y eine Zufallsvariable für Σ'.

Die *gemeinsame Entropie*

$$H(X,Y) = -\mathbb{E}[\log P(X,Y)] = -\sum_{a \in \Sigma} \sum_{a' \in \Sigma'} P(a,a') \log P(a,a').$$

Die *bedingte Entropie* von X unter (der Bedingung) Y:

$$\begin{aligned} H(X|Y) &= -\mathbb{E}_{P(X,Y)}[\log P(X|Y)] = -\textstyle\sum_{a \in \Sigma} \sum_{a' \in \Sigma'} P(a,a') \log P(a|a') \\ &= \textstyle\sum_{a \in \Sigma} P(a) H(Y|X=a). \end{aligned}$$

$H(X|Y)$ gleich Null genau dann, wenn X durch Kenntnis von Y vollständig bestimmt werden kann.

Die *relative Entropie* oder **Kullback-Leibler-Abstand** zwischen zwei Verteilungen P und Q über gleichem Raum:

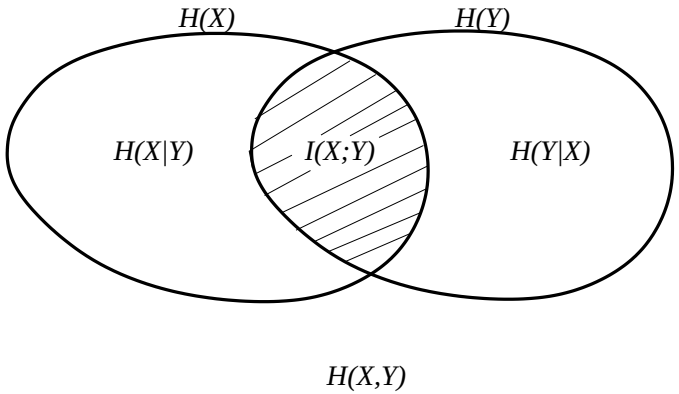
$$D(P||Q) = \mathbb{E}_P \left[\log \frac{P(X)}{Q(X)} \right] = \sum_{a \in \Sigma} P(a) \log \frac{P(a)}{Q(a)},$$

wobei $0 \log 0/Q := 0$ und $P \log P/0 = \infty$.

Die *wechselseitige Information* (engl.: **mutual information**) ist:

$$\begin{aligned} I(X;Y) &= \mathbb{E}_{P(X,Y)} \left[\log \tfrac{P(X,Y)}{P(X)P(Y)} \right] \\ &= \textstyle\sum_{a \in \Sigma} \sum_{a' \in \Sigma'} P(a,a') \log \left[\tfrac{P(a,a')}{P(a)P(a')} \right]. \end{aligned}$$

Eine graphische Illustration der obigen Definition:



Satz 1 *Es gilt:*

$$\begin{aligned} H(X,Y) &= H(X) + H(Y|X), \\ I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X,Y) = I(Y;X), \\ I(X;X) &= H(X). \end{aligned}$$

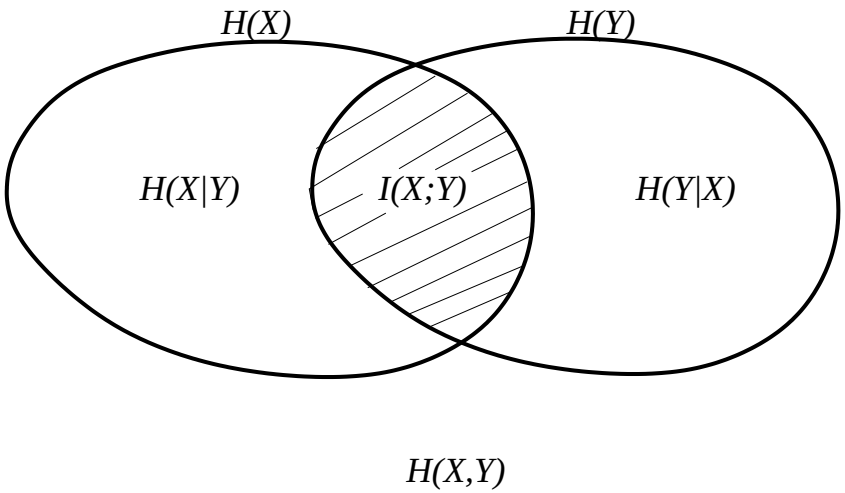
Satz 2 *Die folgenden Ungleichungen gelten:*

$$\begin{aligned} D(P||Q) &\geq 0, && (1) \\ I(X;Y) &\geq 0, && (2) \\ H(X) &\geq H(X|Y), && (3) \\ H(X,Y) &\leq H(X) + H(Y). && (4) \end{aligned}$$

(1) gleich genau dann, wenn $P(a) = Q(a)$ für jedes $a \in \Sigma$, und
(2), (3) und (4) gleich genau dann, wenn X und Y unabhängig sind.

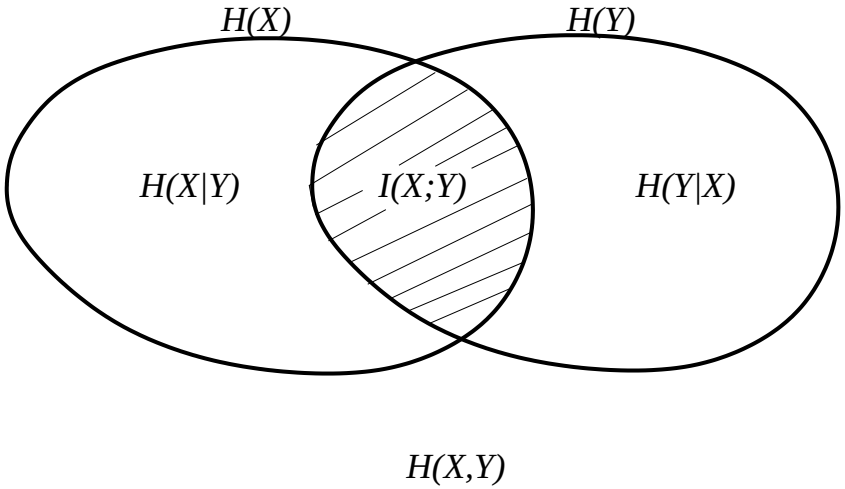
Bilder sind oft einprägsamer als Formeln. . .

$$
\begin{aligned}
H(X,Y) &= H(X) + H(Y|X), \\
I(X;Y) &= H(X) - H(X|Y) \\
&= H(X) + H(Y) - H(X,Y) = I(Y;X), \\
I(X;X) &= H(X).
\end{aligned}
$$



$$\begin{aligned} I(X;Y) &\geq 0, \\ H(X) &\geq H(X|Y), \\ H(X,Y) &\leq H(X) + H(Y). \end{aligned}$$

Gleichheit gilt genau dann, wenn X und Y unabhängig sind.



Beispiel 3 *Betrachte*
$\Sigma = \{0,1,2,\ldots,15\}$, $\Gamma = \{0,1,2,\ldots,7\}$, $\Sigma' = \{0,2,4,\ldots,14\}$.

$$\textit{Schema:} \quad \underset{\text{Eingabe}}{x \in \Sigma} \xrightarrow{C} \underset{\text{Repräsentation}}{x_c \in \Gamma} \xrightarrow{D} \underset{\text{Rekonstruktion}}{y \in \Sigma'}$$

Es sei: $\forall 0 \leq i \leq 15 \; P(i) = 1/16$.
Betrachten wir Codierung: $C(i) = \lfloor i/2 \rfloor$ *und Dekodierung:* $D(j) = 2j$.

$$\begin{aligned} H(X) &= \textstyle\sum_{i=0}^{15} \frac{1}{16} \log 16 \;=\; 4. \\ H(Y) &= -\textstyle\sum_{j=0}^{7} Pr[Y = 2j] \times \log Pr[Y = 2j] \\ &= -\textstyle\sum_{i=0}^{7} Pr[X = 2j \vee X = 2j+1] \times \log Pr[X = 2j \vee X = 2j+1] \\ &= 3. \end{aligned}$$

$$\begin{aligned} H(X|Y) &= -\textstyle\sum_{i=0}^{15}\sum_{j=0}^{7} Pr[X = i|Y = 2j] \times Pr[Y = 2j] \times \log Pr[X = i|Y = 2j] \\ &= -\textstyle\sum_{j=0}^{7} [Pr[X = 2j|Y = 2j] \times Pr[Y = 2j] \times \log Pr[X = 2j|Y = 2j] + \\ &\qquad Pr[X = 2j+1|Y = 2j] \times Pr[Y = 2j] \times \log Pr[X = 2j+1|Y = 2j]] \\ &= -8[-\tfrac{1}{2} \cdot \tfrac{1}{8} \cdot 1 - \tfrac{1}{2} \cdot \tfrac{1}{8} \cdot 1] = 1. \end{aligned}$$

$\rightsquigarrow$ *Ungewiss bleibt 1 Bit, wenn man* Y *kennt, aber nicht* X.

Rate-Distortion-Theorem (Kompression versus Verzerrung)

Wir definieren:

$$D = \sum_{i=0}^{N-1} P(x_i) \sum_{j=0}^{M-1} P(y_j|x_i) \times d(x_i, y_j) \;=\; \sum_{i,j} P(y_j, x_i) \times d(x_i, y_j).$$

Wir nennen D die *Verzerrung*. Beachte: D hängt ab von

1. der Art der Quelle (modelliert durch $P(x_i)$),
2. dem gewählten Kompressionsverfahren (die die Verteilungsfunktion $P(y_j|x_i)$ beeinflusst) und
3. dem Verzerrungsmaß d.

Sind Quelle und Verzerrungsmaß fixiert, können wir D als Funktion auffassen, die lediglich von der Verteilungsfunktion (also dem Kompressionsverfahren) $P(y_j|x_i)$ abhängt.

Beispiel 4 *Für die Quelle und die Codierung wie im vorigen Beispiel und für* $d(i,j) = |i-j|$ *ist die Verzerrung* $D = 1/2$.

Frage: Wie ist der Trade-off zwischen der Verzerrung und dem Kompressionsquotienten?

Oder anders: *Wie ist der bestmögliche Kompressionsquotient R für die Verzerrung, die kleiner oder gleich D ist?*

$R(D)$ bezeichne diese Abhängigkeit.

Es sei $\Gamma(D)$ die Menge von Verteilungen für Y, sodass gilt:

$$\sum P(x_i) \sum P(y_j|x_i) \times d(x_i, y_j) \ \leq \ D.$$

Das heißt: für eine bestimmte Eingabe $P(x_0), P(x_1), \ldots, P(x_{N-1})$ sowie für ein gewisses D ist jede Ausgabe des Kompressionsverfahrens, das die Verteilung für Y aus $\Gamma(D)$ hat, mit der Verzerrung höchstens D behaftet.

Satz 5 *(Shannon 1959)*

$$R(D) = \min_{P(\cdot|\cdot)\in\Gamma(D)} I(X;Y).$$

Der bestmögliche Kompressionsquotient wird also nach oben durch die wechselseitige Information der Zufallsvariablen für Ein- und Ausgabe beschränkt, sofern die Verteilungsfunktion dieser Zufallsvariablen in $\Gamma(D)$ liegt.

Kompression versus Verzerrung für binäre Quelle

Es sei $\Sigma, \Sigma' = \{0, 1\}$ und $d(x, y) = x \oplus y$.

Nehmen wir an: $\Pr[X = 0] = p, \Pr[X = 1] = 1 - p$. Zusätzlich sei $p \leq 1/2$.

Problem: Für gegebenes D berechne $R(D)$.

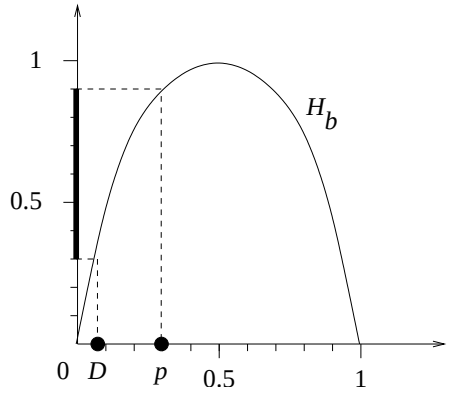
Wir betrachten zwei Fälle:

Fall 1 $D \geq p$. Dann $R(D) = 0$
(Kompressionsverfahren: $C(0) = C(1) = 1;\ D(1) = 1$).

Fall 2 $0 \leq D < p$. Dann $R(D) = H_b(p) - H_b(D)$.

Insgesamt:

$$R(D) = \begin{cases} H_b(p) - H_b(D) & \text{wenn } D < \min\{p, 1-p\} \\ 0 & \text{sonst.} \end{cases}$$



Beweis (Fall 2)
Wir haben: $0 \leq D < p \leq 1/2$. Um $R(D)$ zu berechnen benutzen wir Shannons Satz. Das Ziel ist $I(X;Y)$ zu minimieren über alle Y, sodass

$$\mathbb{E}[d(X,Y)] \leq D. \tag{1}$$

Wir haben:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) = H(X) - H(X \oplus Y|Y) \\ &\geq H(X) - H(X \oplus Y). \end{aligned}$$

Um $I(X;Y)$ zu minimieren ($H(X)$ ist fixiert!), werden wir

$$H(X \oplus Y) = H_b(P(X \oplus Y = 1)) \tag{2}$$

maximieren. Aus (1) und aus der Definition von $d(x,y)$ folgt:

$$\begin{aligned} \mathbb{E}[d(X,Y)] &= P(X=0,Y=1) + P(X=1,Y=0) \\ &= P(X \oplus Y = 1) \ \leq \ D \ < \ p. \end{aligned}$$

Der Wert (2) ist maximal für $P(X \oplus Y = 1) = D$. Das heißt:

$$I(X;Y) \geq H_b(p) - H_b(D)$$

und wir erhalten die Gleichheit für

$$P(X=0|Y=1) \ = \ P(X=1|Y=0) \ = \ D.$$

Was heißt das praktisch ?

Shannons Interpretation ist folgende:

X ist die Eingabe und Y die Ausgabe eines Übertragungskanals.
Wir suchen nach einem Kanal, der $I(X;Y)$ minimiert.
Wir ermitteln die Ausgabewahrscheinlichkeiten solch eines optimalen Kanals.
Dann suche (willkürlich) Codewörter, deren aufmultiplizierte Symbolwahrscheinlichkeiten den errechneten Ausgabewahrscheinlichkeiten (möglichst) gleich sind.

Codierung: Ermittle Codewort, welches (gemessen bezüglich Verzerrung d) der tatsächlichen Nachricht möglichst nahe kommt.
Versende einen Index dieses Codewortes (also: bei 2^K Codewörtern braucht man K Bits).
Decodierung: Liefere das Codewort, welches dem übermittelten Index entspricht.

Hinweis: Die Prozedur ähnelt der Vektorquantisierung (s.u.).