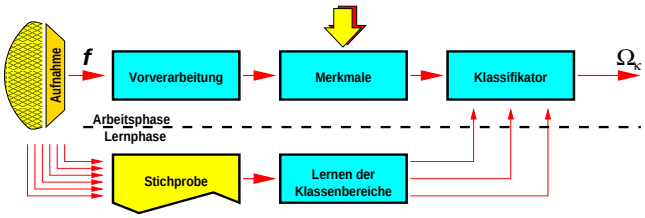
Vorlesung im Sommersemester

MUSTERERKENNUNG

E.G. Schukat-Talamazzini

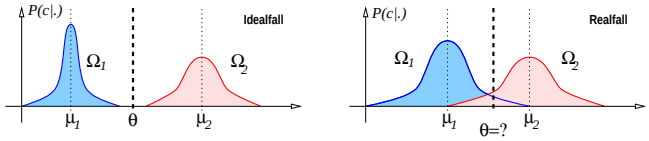
3. Merkmalgewinnung

Reihenentwicklungen • Fraktale Merkmale •
Spektrale Merkmale • Eigentransformationen



Merkmalgewinnung $\widehat{=}$ Transformation

- eines vorverarbeiteten Musters $\boldsymbol{f} \in \mathbb{R}^N$ in ...
- ➪ Merkmalvektor $\boldsymbol{c} \in \mathbb{R}^{N'}$
- ➪ Zeichenfolge $\boldsymbol{v} = v_1 \ldots v_L \in \mathcal{A}^\star$



ZIELSETZUNG:

- *Komprimierung* der Darstellung ($N' \leq N$, besser: $N' \ll N$)
- Gewinnung *trennscharfer* Merkmale
- Unterdrückung/Hervorhebung von *Un/wesentlichem*

Definition 3.1 *Eine Folge $\phi_1, \ldots, \phi_N$ linear unabhängiger Vektoren des $\mathbb{R}^N$ mit der Eigenschaft*

$$\phi_\nu^\top \phi_\mu \;=\; \begin{cases} 1 & \nu = \mu \\ 0 & \nu \neq \mu \end{cases}$$

heißt **Orthonormalsystem**.

Φ bezeichne die **Orthogonalmatrix** mit Basisvektor ϕ_ν in der ν-ten Zeile, d.h. es gilt $\Phi\Phi^\top = \boldsymbol{E}$ (Einheitsmatrix).

Satz 3.1 (Vollständige Reihenentwicklung)

Es sei Φ ein Orthonormalsystem und $\boldsymbol{f} \in \mathbb{R}^N$. Für die Summe

$$\widehat{\boldsymbol{f}} \;=\; \Phi^\top \boldsymbol{c} \;=\; \sum_{\nu=1}^{N} c_\nu \cdot \phi_\nu$$

mit den Entwicklungskoeffizienten

$$c_\nu \;=\; \phi_\nu^\top \boldsymbol{f} \quad \text{bzw.} \quad \boldsymbol{c} \;=\; \Phi \boldsymbol{f}$$

gilt die Aussage $\widehat{\boldsymbol{f}} = \boldsymbol{f}$.

Satz 3.2 (Unvollständige Reihenentwicklung)

Für die unvollständige Darstellung

$$\widehat{\boldsymbol{f}} \;=\; \sum_{\nu=1}^{N'} c_\nu \cdot \phi_\nu \;, \qquad N' < N$$

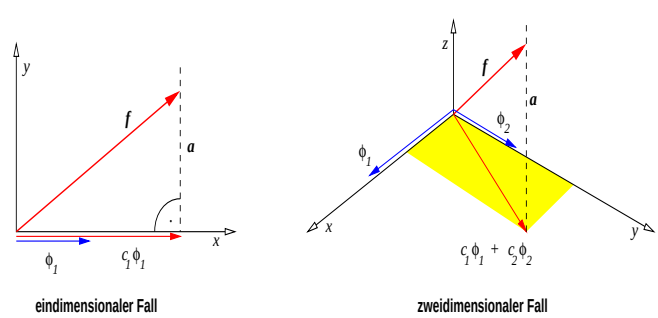
von $\boldsymbol{f}$ nach dem Orthonormalsystem Φ gilt: der mittlere quadratische **Rekonstruktionsfehler**

$$\varepsilon \;=\; \left\| \boldsymbol{f} - \widehat{\boldsymbol{f}} \right\|^2 \;=\; (\boldsymbol{f} - \widehat{\boldsymbol{f}})^\top (\boldsymbol{f} - \widehat{\boldsymbol{f}})$$

wird minimal, wenn die Reihenkoeffizienten gemäß

$$c_\nu \;=\; \phi_\nu^\top \boldsymbol{f} \;, \qquad \nu = 1, \ldots, N'$$

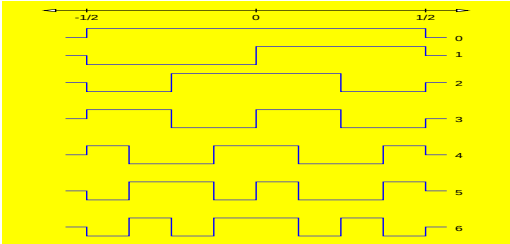
gewählt werden.



Walsh-Hadamard-Transformation

Definition 3.2 *Die kontinuierlichen Walshfunktionen* $wal_n(x)$ *sind definiert durch:*

$$
\begin{aligned}
wal_0(x) &= \begin{cases} 1 & -\frac{1}{2} \le x \le +\frac{1}{2} \\ 0 & \textit{sonst} \end{cases} \\
wal_{2n}(x) &= (-1)^{\lfloor \frac{n}{2} \rfloor} \cdot \left(wal_n(\frac{4x+1}{2}) + (-1)^n wal_n(\frac{4x-1}{2}) \right) \\
wal_{2n+1}(x) &= (-1)^{\lfloor \frac{n}{2} \rfloor + 1} \cdot \left(wal_n(\frac{4x+1}{2}) + (-1)^{n+1} wal_n(\frac{4x-1}{2}) \right)
\end{aligned}
$$



Satz 3.3 *Die Walshfunktionen sind orthogonal. d.h. es gilt für alle* n, m*:*

$$\int_{-1/2}^{+1/2} wal_n(x) \cdot wal_m(x)\, dx = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}$$

Definition 3.3 *Die* **Hadamardmatrizen** *sind definiert durch:*

$$\boldsymbol{H}_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \boldsymbol{H}_N = \boldsymbol{H}_2 \otimes \boldsymbol{H}_{N/2} = \begin{pmatrix} \boldsymbol{H}_{N/2} & \boldsymbol{H}_{N/2} \\ \boldsymbol{H}_{N/2} & -\boldsymbol{H}_{N/2} \end{pmatrix}$$

Für $N = 2^p$ *heißt die Abbildung*

$$\boldsymbol{c} = \mathrm{WHT}(\boldsymbol{f}) = \boldsymbol{H}_N \boldsymbol{f} \quad \textit{bzw.} \quad \boldsymbol{f} = \mathrm{WHT}^{-1}(\boldsymbol{c}) = \frac{1}{N} \boldsymbol{H}_N \boldsymbol{c}$$

(inverse) **Walsh-Hadamard**-*Transformation d.Ordnung* N.

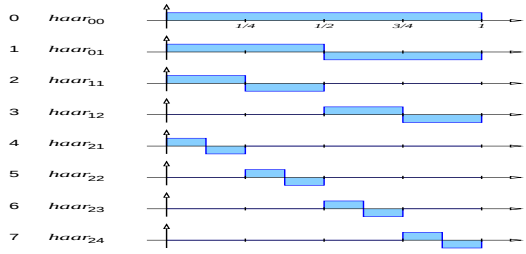
Haar-Transformation

Definition 3.4
Die **Haarfunktionen** $haar_{pq} : [0,1] \to \mathbb{R}$ *sind definiert durch:*

$$
\begin{aligned}
haar_{00}(x) &= 1 \Big/ \sqrt{N} \\
haar_{pq}(x) &= \frac{1}{\sqrt{N}} \cdot \begin{cases} 2^{p/2} & q-1 \le 2^p x < q - 1/2 \\ -2^{p/2} & q - 1/2 \le 2^p x < q \\ 0 & \textit{sonst} \end{cases}
\end{aligned}
$$

Für eine gegebene Ordnung $N = 2^n \in \mathbb{N}$ *läuft*

$$k = 0, 1, \ldots, N-1 \quad \textit{mit} \quad k = 2^p + q - 1$$



Diskrete Haartransformation, für $N = 8$:

$$\mathrm{DHT}(\boldsymbol{f}) = \frac{1}{\sqrt{8}} \cdot \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
\sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\
2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 2 & -2
\end{pmatrix} \cdot \boldsymbol{f}$$

Die DHT ist Grundlage für eine spezielle WAVELET-Transformation.

Definition 3.5
Die **Legendre-Polynome** *sind rekursiv definiert durch:*

$$
\begin{aligned}
\mathrm{Q}_0(x) &= 1 \\
\mathrm{Q}_1(x) &= x \\
\mathrm{Q}_p(x) &= \frac{2p-1}{p} x \cdot \mathrm{Q}_{p-1}(x) - \frac{p-1}{p} \cdot \mathrm{Q}_{p-2}(x)
\end{aligned}
$$

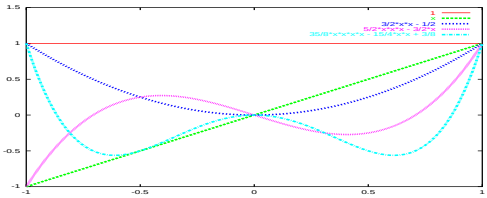
Für die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ *bezeichnet*

$$\lambda_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-1}^{+1} \int_{-1}^{+1} \mathrm{Q}_p(x) \mathrm{Q}_q(y) f(x,y)\, dx\, dy$$

das (p,q)*-te* **Legendre-Moment**. *Für* $[f_{nm}]$ *bezeichnet*

$$\lambda_{pq} = \frac{(2p+1)(2q+1)}{N \cdot M} \sum_{n=1}^{N} \sum_{m=1}^{M} \mathrm{Q}_p(n\Delta x) \mathrm{Q}_q(m\Delta y) f_{nm}$$

das (p,q)*-te* **diskrete Legendre-Moment**.



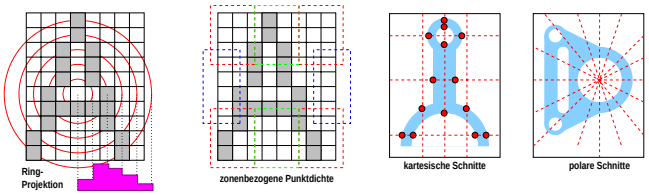
Satz 3.4 (Legendre-Interpolation)
Für ein Muster $[f_{nm}]$ *und seine Legendre-Momente* $\{\lambda_{pq}\}$, $p \le n$, $q \le m$ *gilt die Näherungsformel*

$$f_{nm} \approx \hat{f}(n\Delta x, m\Delta y)$$

mit

$$\hat{f}(x,y) \stackrel{\text{def}}{=} \sum_{p=0}^{n} \sum_{q=0}^{m} \lambda_{pq} \mathrm{Q}_p(x) \mathrm{Q}_q(y)$$

Heuristische Kennzahlen



Projektionen

- x/y-Achse
- Winkelprojektion
- Ringprojektion

Schnittpunkte mit Testlinien

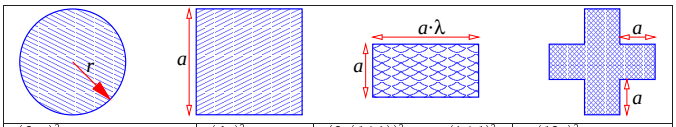
- kartesische Koordinaten (x, y)
- polare Koordinaten (r, ϕ)

Zonenpunktdichten

$$c_{\mathcal{D}}(\boldsymbol{f}) = \sum_{(n,m) \in \mathcal{D}} f_{nm}$$

Formfaktoren

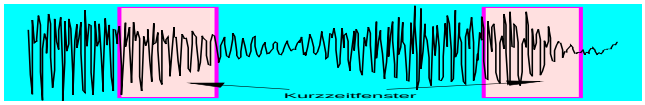
$$c = \frac{(\text{Umfang})^2}{\text{Flächeninhalt}}$$



$\frac{(2\pi r)^2}{\pi r^2} = 4\pi = 12.6$	$\frac{(4a)^2}{a^2} = 16$	$\frac{(2a(1+\lambda))^2}{\lambda^2} = 4\frac{(\lambda+1)^2}{\lambda}$	$\frac{(12a)^2}{5a^2} = 28.8$

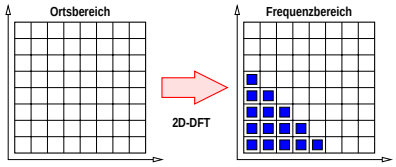
SPEKTRALE MERKMALE

Schallwellen



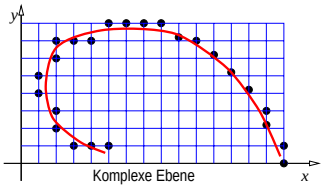
Grauwertbilder

Ausgewählte Merkmale (Betrag/Phase) des 2D-Spektrums



Linien und Konturen

$[f_{nm}] \;\mapsto\; [z_k]$ mit $z_k = x_k + iy_k \;(k = 0, \ldots, K-1)$



FOURIERDESKRIPTOREN (verschiebungsinvariant)

$$c_\kappa \;=\; |Z_\kappa|^2 \quad \text{mit} \quad [Z_\kappa] \;=\; \mathrm{DFT}\{[z_k]\}$$

Lineares Vorhersagemodell

$$f_n \;\approx\; -\sum_{\nu=1}^{p} a_\nu \cdot f_{n-\nu}$$

Vorhersage aus den vergangenen p Werten im Bereich $N_0 \le n \le N_1$

Vorhersageformel

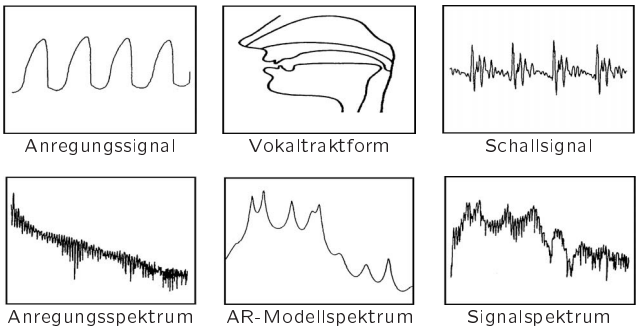
$$\hat{f}_n \;\stackrel{\text{def}}{=}\; -\sum_{\nu=1}^{p} a_\nu \cdot f_{n-\nu}$$

Fehlersignal

$$e_n \;\stackrel{\text{def}}{=}\; f_n - \hat{f}_n \;=\; \sum_{\nu=0}^{p} a_\nu \cdot f_{n-\nu} \qquad (a_0 = 1)$$

Übergang zur z-Transformierten

$$\begin{aligned} E(z) &= F(z) \cdot A(z) \\ F(z) &= E(z) \cdot 1/A(z) \end{aligned}$$



Satz 3.5 (Markel 1972)
Der mittlere quadratische Vorhersagefehler

$$\varepsilon \;=\; \sum_{n=N_0}^{N_1} \left(f_n - \hat{f}_n\right)^2$$

der Prädiktionsformel p-ter Ordnung

$$\hat{f}_n \;=\; -\sum_{\nu=1}^{p} a_\nu \cdot f_{n-\nu}$$

wird gerade durch diejenigen Prädiktionskoeffizienten $a_0, \ldots, a_p$ minimiert, welche Lösung des linearen Gleichungssystems

$$\sum_{\nu=1}^{p} a_\nu \cdot \psi_{\nu\mu} \;=\; -\,\psi_{0\mu}\;, \qquad \mu = 1, \ldots, p$$

sind, wobei die Abkürzung

$$\psi_{\nu\mu} \;\stackrel{\text{def}}{=}\; \sum_{n=N_0}^{N_1} f_{n-\nu} \cdot f_{n-\mu}\;, \qquad \nu, \mu = 0, \ldots, p$$

vereinbart sei.

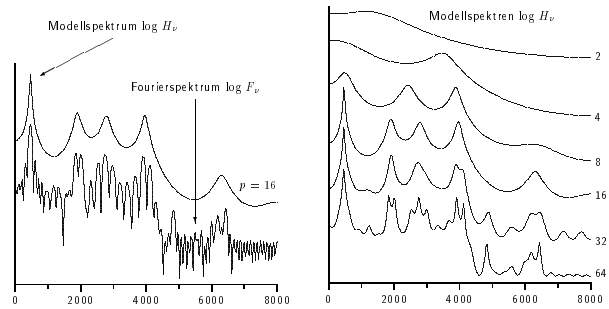
Definition 3.6 *Ist $[f_n]$ ein Sprachsignal mit den Vorhersagekoeffizienten $[a_k]$ der Ordnung p, so heißt*

$$H(e^{i\omega}) \;=\; \frac{1}{A(e^{i\omega})} \;=\; \frac{1}{\sum_{k=0}^{p} a_k \cdot e^{-i\omega k}}$$

das **Modellspektrum** *von $[f_n]$.*

Das logarithmierte Betragsmodellspektrum wird praktisch wie folgt berechnet:

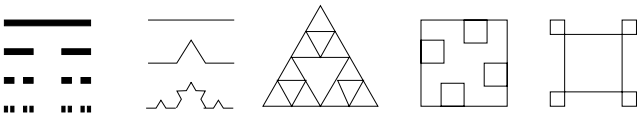
$$\log|H_\nu|^2 \;=\; -\log|A_\nu|^2 \qquad [A_\nu] \;=\; \mathrm{DFT}\left\{(1, a_1, \ldots, a_p, 0, \ldots, 0)^\top\right\}$$



Fraktale Dimension

Maß für den Ausfüllungsgrad einer Punktmenge

Selbstähnliche fraktale Punktmengen



Cantormenge | von Koch Kurve | Sierpinskidreieck | Cantorstaub | Falconermäander

Definition 3.7 *Besteht die selbstähnliche Menge A aus genau m vielen, mit Faktor $0 < r < 1$ skalierten Kopien ihrer selbst, so beträgt die* **Ähnlichkeitsdimension**

$$\dim_S A \;=\; \log m / \log(1/r)$$

Für die Ähnlichkeitsdimension gilt folglich

$$m \;=\; \left(\frac{1}{r}\right)^{\dim_S A} \;=\; r^{-\dim_S A}$$

Linie	$\log m / \log m = 1$
Quadrat	$\log m^2 / \log m = 2$
Kubus	$\log m^3 / \log m = 3$
Cantormenge	$\log 2 / \log 3 = 0.631$
von Kochkurve	$\log 4 / \log 3 = 1.262$
Sierpinskidreieck	$\log 3 / \log 2 = 1.585$
Kantorstaub	$\log 4 / \log 4 = 1$
Falconermenge	?

Hausdorff-Dimension

Für $U \subseteq \mathbb{R}^n$ definiere den **Durchmesser** als

$$|U| \stackrel{\text{def}}{=} \sup\{|x-y| \mid x, y \in U\}$$

Das Mengensystem $\{U_i\}$ ist **δ-Überdeckung** von A gdw.

$$A \;\subseteq\; \bigcup_{i=1}^{\infty} U_i \quad \text{und} \quad \forall i : 0 < |U_i| \le \delta$$

Für $\delta > 0$, $s \ge 0$ definiere die Zahlen

$$\mathfrak{h}^s_\delta(A) \stackrel{\text{def}}{=} \inf\left\{ \sum_{i=1}^{\infty} |U_i|^s \mid \{U_i\} \text{ ist } \delta\text{-Überdeckung von } A \right\}$$

Das **s-dimensionale Hausdorffmaß** von A beträgt dann

$$\mathfrak{h}^s(A) \stackrel{\text{def}}{=} \lim_{\delta \to 0} \mathfrak{h}^s_\delta(A)$$

SKALIERUNGSEIGENSCHAFT:

$$\mathfrak{h}^s(\lambda A) \;=\; \lambda^s \cdot \mathfrak{h}^s(A)$$

mit der Skalierung $\lambda A = \{\lambda x \mid x \in A\}$

Es gibt für jedes A genau ein $s_0 \in \mathbb{R}$ mit der Eigenschaft

$$\mathfrak{h}^s(A) \;=\; \begin{cases} \infty & s < s_0 \\ c & s = s_0 \\ 0 & s > s_0 \end{cases}$$

Dieses s_0 ist die **Hausdorff-Dimension** ($\dim_H A = s_0$).

Kolmogorov-Entropie

$N^{(a)}_\delta(A)$	minimale Anzahl von Mengen mit Durchmesser $\le \delta$, die A überdecken
$N^{(b)}_\delta(A)$	Anzahl der δ-Netz Hyperwürfel, die A schneiden
$N^{(c)}_\delta(A)$	minimale Anzahl von δ-Hyperwürfeln, die A überdecken
$N^{(d)}_\delta(A)$	minimale Anzahl von δ-Hyperkugeln, die A überdecken
$N^{(e)}_\delta(A)$	maximale Anzahl disjunkter δ-Hyperkugeln mit Mittelpunkt in A

Definition 3.8 (Boxdimensionen)
Für eine Punktmenge $A \subseteq \mathbb{R}^n$ und $x \in \{a, b, c, d, e\}$ definiere

$$\uparrow\dim^{(x)}_B A \;=\; \liminf_{\delta \to 0} \frac{\log N^{(x)}_\delta(A)}{-\log \delta}$$

$$\downarrow\dim^{(x)}_B A \;=\; \limsup_{\delta \to 0} \frac{\log N^{(x)}_\delta(A)}{-\log \delta}$$

sowie

$$\dim^{(x)}_B A \stackrel{\text{def}}{=} \uparrow\dim^{(x)}_B A\,, \quad \textit{falls} \quad \uparrow\dim^{(x)}_B A \;=\; \downarrow\dim^{(x)}_B A$$

Satz 3.6 *Die Unter- und Ober-Boxdimensionen $\uparrow\dim^{(x)}_B$ bzw. $\downarrow\dim^{(x)}_B$ stimmen für alle $x \in \{a, b, c, d, e\}$ überein.*

Satz 3.7 *Für jede Punktmenge $A \subseteq \mathbb{R}^n$ gilt der folgende Zusammenhang zwischen Hausdorff- und Boxdimensionen:*

$$\dim_H A \;\le\; \uparrow\dim_B A \;\le\; \downarrow\dim_B A$$

Definition 3.9 (Minkowski-Dimension)
Für eine Punktmenge $A \subseteq \mathbb{R}^n$ und $\delta > 0$ heißt

$$A_\delta \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \; |x-y| \le \delta \text{ für ein } y \in A \;\}$$

die **δ-parallele Umgebung** *von A.*
Verhält sich das n-dimensionale Volumen $vol_n(A_\delta)$ asymptotisch gemäß

$$vol_n(A_\delta) \;\sim\; c \cdot \delta^{n-s}\,,$$

so heißt s die **Minkowski-Dimension** *und c der* **Minkowski-Gehalt** *von A.*

Satz 3.8
Für eine Punktmenge $A \subseteq \mathbb{R}^n$ gelten die Aussagen

$$\uparrow\dim_B A \;=\; n - \liminf_{\delta \to 0} \frac{\log vol_n(A_\delta)}{\log \delta}$$

$$\downarrow\dim_B A \;=\; n - \limsup_{\delta \to 0} \frac{\log vol_n(A_\delta)}{\log \delta}$$

Definition 3.10 (Dimension einer Kurve)
Sei A das Bild der stetigen Bijektion $f : [a, b] \to \mathbb{R}^n$ und $M_\delta(A)$ die maximale Länge $M + 1$ von Punktfolgen $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_M \in A$ mit Abstand $|\boldsymbol{x}_m - \boldsymbol{x}_{m-1}| = \delta$. Dann heißt

$$\dim_E A \stackrel{\text{def}}{=} \lim_{\delta \to 0} \frac{\log M_\delta(A)}{-\log \delta}$$

die **Etappendimension** *der Kurve A bzw. der Abbildung f.*

BEISPIEL: Für die britische Küstenlinie gilt $\dim_E = 1.2$

Potenzgesetz für die fraktale Dimension

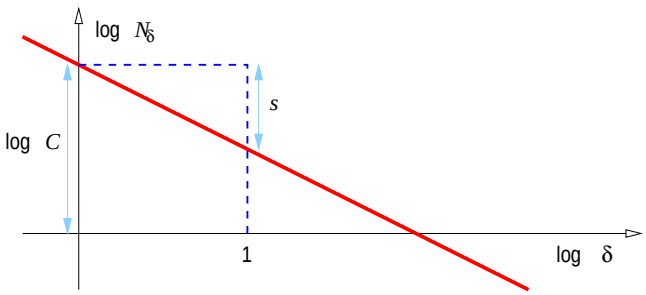
$$N_\delta \;\approx\; C \cdot \delta^{-s}$$

für die Würfel-, Kugel- oder Etappenzahl $N_\delta(A)$

Die logarithmierte Form

$$\log N_\delta \;\approx\; \log C - s \cdot \log \delta$$

suggeriert einen **linearen** Zusammenhang zwischen $\log \delta$ und $\log N_\delta$



BERECHNUNG:

- Berechne N_δ für ausgewählte Werte $\delta_i > 0$
- Plaziere die Punkte $(\log \delta_i, \log N_{\delta_i})$
- Bestimme die Steigung der Ausgleichsgeraden

Analytische Methoden zur Merkmalgewinnung

Systematische Berechnung der $n < N$ besten Merkmale $c_1, \ldots, c_n$

Lineare Transformation

$$\boldsymbol{c} \;=\; \boldsymbol{\Phi}^\top \boldsymbol{f}\,, \qquad \boldsymbol{f} \in \mathbb{R}^N$$

Matrix mit orthonormalen Spaltenvektoren

$$\boldsymbol{\Phi} \;=\; (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n)\,, \qquad \boldsymbol{\phi}_\nu \in \mathbb{R}^N$$

Optimierung eines Gütekriteriums

- Konzentration der Merkmalsgebiete aller Muster
- Kompaktheit der Merkmalsgebiete von Mustern *einer* Klasse
- Trennung der Merkmalsgebiete von Mustern *verschiedener* Klassen
- Übereinstimmung mit einem statistischen Verteilungsmodell
- ⇩ Approximation einer *idealen* Klassenzugehörigkeitsfunktion
- ⇩ Minimaler Klassifikationsfehler

bezogen auf eine **Lernstichprobe** von Mustern

Stichproben von Mustern und Merkmalen

Definition 3.11
Es sei Ω ein Problemkreis mit Mustern aus K Klassen. Eine Teilmenge $\omega \subseteq \Omega$ heißt **Stichprobe** *aus Ω. Ein System*

$$\{\omega_\kappa \mid \kappa = 1, \ldots, K\}$$

von Mengen $\omega_\kappa \subseteq \Omega$ von Mustern der Klasse κ heißt **klassifizierte** *oder* **etikettierte Stichprobe** *aus Ω.*

Wir verwenden dieselben Bezeichnungen auch für die Ergebnisse

$$\omega^{\mathfrak{T}} \qquad \text{bzw.} \qquad \{\omega_\kappa^{\mathfrak{T}} \mid \kappa = 1, \ldots, K\}$$

der Anwendung einer Transformation $\mathfrak{T} : \Omega \to \mathbb{R}^n$ von Mustern in den Merkmalraum $\mathbb{R}^n$.

Definition 3.12
Für eine (endliche) Stichprobe $\omega = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(T)}\}$ bezeichne $|\omega| = T$ die **Mächtigkeit** *(den Umfang) der Stichprobe,*

$$\boldsymbol{\mu} \;=\; \boldsymbol{\mu}(\omega) \;=\; \frac{1}{T} \cdot \sum_{\boldsymbol{x} \in \omega} \boldsymbol{x}$$

den (empirischen) **Mittelwertvektor** *und*

$$\boldsymbol{S} \;=\; \boldsymbol{S}(\omega) \;=\; \frac{1}{T} \cdot \sum_{\boldsymbol{x} \in \omega} (\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top \;=\; \underbrace{\frac{1}{T} \cdot \sum_{\boldsymbol{x} \in \omega} \boldsymbol{x}\boldsymbol{x}^\top}_{\boldsymbol{R}(\omega)} - \boldsymbol{\mu}\boldsymbol{\mu}^\top$$

die (empirische) **Kovarianzmatrix** *(bzw.* **Momentenmatrix** *$\boldsymbol{R}(\omega)$).*

Für die Mittelwertvektoren, Momenten- und Kovarianzmatrizen der klassenbezogenen Komponenten ω_κ einer etikettierten Stichprobe verwenden wir auch die Kurzbezeichnungen $\boldsymbol{\mu}_\kappa$, $\boldsymbol{R}_\kappa$ und $\boldsymbol{S}_\kappa$.

Kompaktheitskriterien

K1 Mittlerer quadratischer Abstand *aller* Merkmalvektoren einer Stichprobe $\omega \subseteq \mathbb{R}^n$ von *allen* anderen

$$s_1 \;=\; \underbrace{\frac{1}{|\omega|^2} \sum_{\boldsymbol{x} \in \omega} \sum_{\boldsymbol{y} \in \omega} (\boldsymbol{x} - \boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y})}_{d(\omega,\omega)}$$

K2 Mittlerer quadratischer Abstand aller Merkmalvektoren einer Klasse Ω_κ von denen einer *anderen* Klasse Ω_λ

$$s_2 \;=\; \frac{2}{K(K-1)} \sum_{\kappa=1}^{K} \sum_{\lambda<\kappa} \underbrace{\frac{1}{|\omega_\kappa| \cdot |\omega_\lambda|} \sum_{\boldsymbol{x} \in \omega_\kappa} \sum_{\boldsymbol{y} \in \omega_\lambda} (\boldsymbol{x} - \boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y})}_{d(\omega_\kappa,\omega_\lambda)}$$

K3 Mittlerer quadratischer Abstand aller Merkmalvektoren einer Stichprobe $\{\omega_\kappa\}$ von Merkmalvektoren der *eigenen* Klasse

$$s_3 \;=\; \frac{1}{K} \sum_{\kappa=1}^{K} \underbrace{\frac{1}{|\omega_\kappa|^2} \sum_{\boldsymbol{x} \in \omega_\kappa} \sum_{\boldsymbol{y} \in \omega_\kappa} (\boldsymbol{x} - \boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y})}_{d(\omega_\kappa,\omega_\kappa)}$$

K4 Kombination der Kriterien K2 und K3

$$s_4 \;=\; s_2 + \theta \cdot s_3 \qquad (\theta = \text{Lagrange-Multiplikator})$$

Satz 3.9
Ist $\Phi = (\phi_1, \ldots, \phi_n) \in \mathbb{R}^{N\times n}$ eine orthonormale Transformationsmatrix ($\Phi^\top\Phi = E$) und sind ω' und $\{\omega'_\kappa\}$ die (un)klassifizierten Merkmalstichproben zu ω und $\{\omega_\kappa\}$ hinsichtlich der Transformation $c = \Phi^\top f$, so besitzen die Kompaktheitskriterien (1)–(3) — bezogen auf die transformierten Muster — die Gestalt

$$s_i \;=\; 2\cdot\sum_{\nu=1}^{n}\phi_\nu^\top Q^{(i)}\phi_\nu$$

mit den symmetrischen, positiv-definiten **Kernmatrizen**

$$Q^{(1)} \;\stackrel{\text{def}}{=}\; S(\omega)$$

$$Q^{(2)} \;\stackrel{\text{def}}{=}\; \frac{1}{K}\sum_{\kappa=1}^{K} R(\omega_\kappa) \;-\; \frac{1}{K(K-1)}\sum_{\kappa\neq\lambda}\mu(\omega_\kappa)\mu(\omega_\lambda)^\top$$

$$Q^{(3)} \;\stackrel{\text{def}}{=}\; \frac{1}{K}\sum_{\kappa=1}^{K} S(\omega_\kappa)$$

Satz 3.10 *Ist die Matrix $A \in \mathbb{R}^{N\times N}$ symmetrisch, so gilt für alle Vektoren $u, v \in \mathbb{R}^N$:*

$$u^\top A v \;=\; \text{spur}\left(A u v^\top\right)$$

Satz 3.11 *Ist die Matrix $A \in \mathbb{R}^{N\times N}$ symmetrisch, und ist $\Phi = (\phi_1, \ldots, \phi_n) \in \mathbb{R}^{N\times n}$, so gilt:*

$$\text{spur}\left(\Phi\Phi^\top\cdot A\right) \;=\; \sum_{\nu=1}^{n}\phi_\nu^\top A\phi_\nu$$

EIGENWERTPROBLEME

Definition 3.13 *Sind A, $B \in \mathbb{R}^{n\times n}$ zwei quadratische Matrizen, so heißt*

$$Ax \;=\; \lambda x \qquad \text{bzw.} \qquad Ax \;=\; \lambda B x \qquad (\lambda\in\mathbb{R}, x\in\mathbb{R}^n)$$

spezielles (allgemeines) Eigenwertproblem. *Ist obiges EWP erfüllt, heißt x* **Eigenvektor** *zum* **Eigenwert** *λ.*

Es gelten die folgenden Aussagen:

- Skalierungseigenschaft
$$Ax \;=\; \lambda x \quad\Rightarrow\quad A(\alpha x) \;=\; \lambda(\alpha x)$$
- Das spezielle EWP ist äquivalent zum homogenen LGS
$$(A - \lambda E)\cdot x \;=\; 0$$
- Das EWP besitzt nichttriviale Lösungen $x \neq 0$ gdw.
$$P(\lambda) \stackrel{\text{def}}{=} \det(A - \lambda E) \;=\; 0 \qquad \text{(charakt. Polynom)}$$
- A besitzt genau n nicht notwendig verschiedene Eigenwerte
$$\lambda_1, \ldots, \lambda_n$$
- A symmetrisch ➪ $\lambda_i \in \mathbb{R}$ für alle i
- alle λ_i verschieden ➪ $\{x_1, \ldots, x_n\}$ linear unabhängig

Für **reelle symmetrische** Matrizen ($A^\top = A$) gilt:

- Eigenvektoren x_i, x_j zu verschiedenen Eigenwerten λ_i, λ_j sind orthogonal:
$$x_i^\top x_j \;=\; 0$$
- Die Eigenvektoren mehrfacher Eigenwerte sind orthogonalisierbar (Gram-Schmidt)
- Die normierten Eigenvektoren $U = (u_1, \ldots, u_n)$ von A bilden ein Orthonormalsystem:
$$U^\top U \;=\; E$$
- Für Orthonormalsysteme $U \in \mathbb{R}^{n\times n}$ gilt auch
$$UU^\top \;=\; E$$
- Eigenwertproblem in Matrixschreibweise:
$$AU \;=\; UD\,, \qquad D = \text{diag}(\lambda_1, \ldots, \lambda_n)$$

Satz 3.12
Sind $U = (u_1, \ldots, u_n)$ bzw. $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$ die (normierten) Eigenvektoren und Eigenwerte der reellen symmetrischen Matrix $A \in \mathbb{R}^{n\times n}$, so gilt:

$$D \;=\; U^\top A U$$

$$A \;=\; U D U^\top$$

$$A \;=\; \sum_{\nu=1}^{n}\lambda_\nu u u^\top$$

$$A^{-1} \;=\; U D^{-1} U^\top$$

Quadratische Formen

Definition 3.14 *Ist A reell und symmetrisch, so heißt die* **quadratische Form**

$$Q : x \;\mapsto\; x^\top A x$$

positiv/negativ-definit, *wenn für alle x gilt:*

$$x \neq 0 \quad\Rightarrow\quad Q(x) > 0 \qquad (\text{bzw. } Q(x) < 0)$$

Die quadratische Form $Q(x) = x^\top A x$ ist positiv-definit (negativ-definit) genau dann, wenn A ausschließlich positive (negative) Eigenwerte besitzt.

Satz 3.13 *Die quadratische Form $Q(x) = x^\top A x$ nimmt — unter der Nebenbedigung $\|x\| = 1$ — ihren maximalen (minimalen) Wert an, wenn x der Einheitsvektor u_k zum größten (kleinsten) Eigenwert λ_k ist. Es gilt dann*

$$Q(u_k) \;=\; u_k^\top A u_k \;=\; \lambda_k$$

Satz 3.14 *Die Summe*

$$s \;=\; \sum_{\mu=1}^{m} Q(\phi_\mu) \;=\; \sum_{\mu=1}^{m}\phi_\mu^\top A\phi_\mu$$

quadratischer Formen nimmt — unter der Voraussetzung normierter, paarweise orthogonaler Vektoren ϕ_μ — ihr Maximum (Minimum) an, wenn für die ϕ_μ die zu den m größten (kleinsten) Eigenwerten $\lambda_1, \ldots, \lambda_m$ von A gehörenden Eigenvektoren $u_1, \ldots, u_m$ eingesetzt werden. Es gilt dann

$$s \;=\; \sum_{\mu=1}^{m} Q(u_\mu) \;=\; \sum_{\mu=1}^{m}\lambda_\mu$$

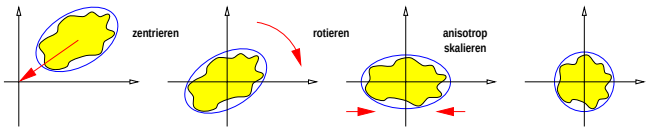
Definition 3.15 (KL-Transformation)
Sind $\boldsymbol{\mu}$ und $\boldsymbol{S}$ Mittelwertvektor und Kovarianzmatrix der Muster einer Stichprobe $\omega \subseteq \mathbb{R}^N$, und ist $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N)$ die Matrix der zu den nach absteigender Größe sortierten Eigenwerten $\boldsymbol{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ gehörenden Eigenvektoren von $\boldsymbol{S}$, so heißt die Abbildung

$$\mathfrak{T}_{\mathrm{KL}} : \boldsymbol{f} \;\mapsto\; \boldsymbol{U}^\top \boldsymbol{f}$$

die **Karhunen-Loéve-Transformation** *von ω. Wir bezeichnen ferner*

$$\boldsymbol{f} \;\mapsto\; \boldsymbol{U}^\top(\boldsymbol{f} - \boldsymbol{\mu}) \quad \text{und} \quad \boldsymbol{f} \;\mapsto\; (\boldsymbol{U}\boldsymbol{D}^{-1/2})^\top(\boldsymbol{f} - \boldsymbol{\mu})$$

als **zentrierte** *bzw.* **normierte** *Karhunen-Loéve-Transformation*



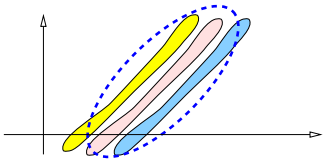
Satz 3.15 *Die oben definierte KL-Transformation bildet Mustervektoren auf dekorrelierte Merkmalvektoren ab, d.h. die Kovarianzmatrix wird zur Diagonalmatrix (oder zur Einheitsmatrix bei der normierten KLT).*

Die unvollständige Reihenentwicklung

$$\mathfrak{T}_{\mathrm{KL}}^{(n)} : \boldsymbol{f} \;\mapsto\; (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)^\top \boldsymbol{f}$$

erzielt unter allen Reihenentwicklungen den niedrigsten mittleren quadratischen Rekonstruktionsfehler für die Muster der Stichprobe.

Lineare Diskriminanyanalyse



„Adidas-Problem“

- Ballungsgebiete kompakt halten
- Zentren voneinander entfernen
- Muster in eine der Klassentrennung förderliche Lage drehen

Definition 3.16 *Ist $\{\omega_\kappa\}$ eine klassifizierte Stichprobe, so heißt*

$$\boldsymbol{S}_W \;=\; \sum_{\kappa=1}^{K} \frac{|\omega_\kappa|}{|\omega|} \boldsymbol{S}_\kappa \;=\; \frac{1}{|\omega|} \sum_{\kappa=1}^{K} \sum_{\boldsymbol{x} \in \omega_\kappa} (\boldsymbol{x} - \boldsymbol{\mu}_\kappa)(\boldsymbol{x} - \boldsymbol{\mu}_\kappa)^\top$$

die (empirische) **Intraklassenstreuungsmatrix** *('within-class scatter') und*

$$\boldsymbol{S}_B \;=\; \sum_{\kappa=1}^{K} \frac{|\omega_\kappa|}{|\omega|} \cdot (\boldsymbol{\mu}_\kappa - \boldsymbol{\mu})(\boldsymbol{\mu}_\kappa - \boldsymbol{\mu})^\top$$

die (empirische) **Interklassenstreuungsmatrix** *('between-class scatter'); dabei bezeichne $\boldsymbol{\mu}_\kappa$ den κ-ten klassenbezogenen und $\boldsymbol{\mu}$ den globalen Mittelwertvektor der Stichprobe.*

Satz 3.16 *Für die Gesamtstreuungsmatrix*

$$\boldsymbol{S} \;=\; \frac{1}{|\omega|} \sum_{\boldsymbol{x} \in \omega} (\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top \;=\; \frac{1}{|\bigcup \omega_\kappa|} \sum_{\kappa=1}^{K} \sum_{\boldsymbol{x} \in \omega_\kappa} (\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top$$

der Stichprobe $\{\omega_\kappa\}$ gilt die Aussage

$$\boxed{\boldsymbol{S} \;=\; \boldsymbol{S}_W + \boldsymbol{S}_B}$$

Transformation von Kitano (1973)

1 Transformation $\tilde{\boldsymbol{x}} = \boldsymbol{\Phi}_1^\top \boldsymbol{x}$ mit

$$\boldsymbol{\Phi}_1 \stackrel{\text{def}}{=} \boldsymbol{U}\boldsymbol{D}^{-1/2} \quad \text{und} \quad \boldsymbol{S}_W \boldsymbol{U} \stackrel{!}{=} \boldsymbol{U}\boldsymbol{D}$$

Für die Intraklassenstreuung der transformierten Stichprobe gilt dann:

$$\begin{aligned} \tilde{\boldsymbol{S}}_W &= \boldsymbol{\Phi}_1^\top \boldsymbol{S}_W \boldsymbol{\Phi}_1 \\ &= (\boldsymbol{U}\boldsymbol{D}^{-1/2})^\top (\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top)(\boldsymbol{U}\boldsymbol{D}^{-1/2}) \\ &= \boldsymbol{D}^{-1/2}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{D}^{-1/2} \;=\; \boldsymbol{D}^{-1/2}\boldsymbol{D}\boldsymbol{D}^{-1/2} \;=\; \boldsymbol{E} \end{aligned}$$

Die $\tilde{\boldsymbol{x}}$ sind also *invariant* gegenüber orthonormalen Transformationen.

2 Transformation $\hat{\boldsymbol{x}} = \boldsymbol{\Phi}_2^\top \tilde{\boldsymbol{x}}$ mit

$$\boldsymbol{\Phi}_2 \stackrel{\text{def}}{=} \boldsymbol{V} \quad \text{und} \quad \tilde{\boldsymbol{S}}_B \boldsymbol{V} \stackrel{!}{=} \boldsymbol{V}\boldsymbol{C}$$

Klassenzentren auseinanderziehen, ohne Kompaktheit zu zerstören!

Satz 3.17 *Die Gesamttransformationsmatrix $\boldsymbol{\Phi} \stackrel{\text{def}}{=} \boldsymbol{\Phi}_1\boldsymbol{\Phi}_2$ von Kitano ist gerade die Lösung des Eigenwertproblems*

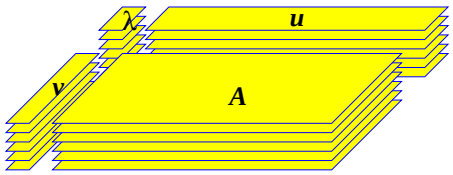
$$(\boldsymbol{S}_W^{-1}\boldsymbol{S}_B) \cdot \boldsymbol{U} \stackrel{!}{=} \boldsymbol{U}\boldsymbol{D}$$

SINGULÄRWERTZERLEGUNG

Nach Satz 3.12 bietet die Summe

$$\tilde{\boldsymbol{A}} \;=\; \sum_{n=1}^{N'} \lambda_n \boldsymbol{u}_n \boldsymbol{u}_n^\top$$

mit den betragsmäßig geordneten Eigenwerten $\{\lambda_n\}$ eine interessante Approximationsformel für *symmetrische quadratische* Grauwertbilder. *Und allgemein?*



Satz 3.18 *Sei $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ eine reellwertige Rechteckmatrix. Die (symmetrischen quadratischen) Matrizen $\boldsymbol{A}^\top\boldsymbol{A}$ und $\boldsymbol{A}\boldsymbol{A}^\top$ besitzen genau dieselben Eigenwerte $\{\lambda_i\}$.*

Definition 3.17 *Sei $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ eine reellwertige Rechteckmatrix. Die Vektoren $\boldsymbol{u}_i \in \mathbb{R}^N$ mit*

$$\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{u}_i \;=\; \lambda_i \boldsymbol{u}_i$$

heißen **Rechtssingulärvektoren** *von $\boldsymbol{A}$, die Vektoren $\boldsymbol{v}_i \in \mathbb{R}^M$ mit*

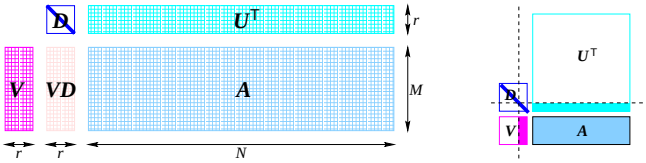
$$\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{v}_i \;=\; \lambda_i \boldsymbol{v}_i$$

heißen **Linkssingulärvektoren** *von $\boldsymbol{A}$. Die Wurzeln $d_i = \sqrt{\lambda_i}$ der gemeinsamen Eigenwerte von $\boldsymbol{A}^\top\boldsymbol{A}$ und $\boldsymbol{A}\boldsymbol{A}^\top$ heißen* **Singulärwerte** *von $\boldsymbol{A}$.*

Satz 3.19 *Für zwei zusammengehörende normierte Rechts- und Linkssingulärvektoren $\boldsymbol{u}$, $\boldsymbol{v}$ mit Singulärwert d gelten die Aussagen*

$$\boldsymbol{A}\boldsymbol{u} \;=\; d\boldsymbol{v} \qquad \text{und} \qquad \boldsymbol{A}^\top\boldsymbol{v} \;=\; d\boldsymbol{u}$$

Satz 3.20 *Besitzt die Matrix $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ den Rang r, so existieren genau r positive sowie $N-r$ bzw. $M-r$ verschwindende Singulärwerte, zusammen mit ihren orthonormierten Rechts- und Linkssingulärvektoren.*



Satz 3.21 (Singulärwertzerlegung)
Besitzt die Matrix $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ den Rang r, und sind $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r)$ und $\boldsymbol{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r)$ die Matrizen, welche die zu den nichtverschwindenden Singulärwerten $d_1, \ldots, d_r$ gehörenden orthonormierten Rechts- und Linkssingulärvektoren als Spalten enthalten, so gilt für $\boldsymbol{A}$ die Darstellungsformel

$$\boldsymbol{A} \;=\; \boldsymbol{V}\boldsymbol{D}\boldsymbol{U}^\top \;=\; \sum_{\rho=1}^{r} d_\rho \boldsymbol{v}_\rho \boldsymbol{u}_\rho^\top$$

Dabei bezeichne $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_r)$.