



ICA-Kurs

Technische Universität Berlin

Institut für Elektronik und Lichttechnik

Quelle: Auszüge aus der Diplomarbeit von Carsten Hennig

Berlin, den 30. Juni 1999

Inhaltsverzeichnis

1	Einleitung	2
2	Verfahren zur ICA	3
2.1	Higher Order Statistics-basierte algebraische Verfahren	3
2.1.1	Statistische Kennwerte höherer Ordnung	3
2.1.2	Dekorrelation zweiter Ordnung als Vorverarbeitung	4
2.1.3	JADE-Algorithmus	6
2.2	Informationstheoretisch basierte Verfahren	7
2.2.1	Informationstheoretische Grundlagen zur ICA	7
2.2.2	Infomax-Algorithmus	10
2.2.3	Neuronale Verfahren zur ICA	11
2.2.4	Fixed-Point-Algorithmus	12
2.3	Second Order Statistics-basierte Verfahren	13
	Literatur	17

1 Einleitung

Das Ziel der ICA bei der blinden Quellentrennung ist die Rekonstruktion von Quellsignalen $\mathbf{s}(k)$ aus einer Mischung $\mathbf{x}(k)$ dieser Signale, wobei sowohl die Quellsignale als auch die Mischung nicht bekannt sind (siehe 1). Allerdings werden Annahmen bezüglich der Signale und der Mischung dieser Signale gemacht. Für die zu rekonstruierenden Signale $\mathbf{s}(k)$ wird angenommen, daß diese statistisch unabhängig voneinander sind.

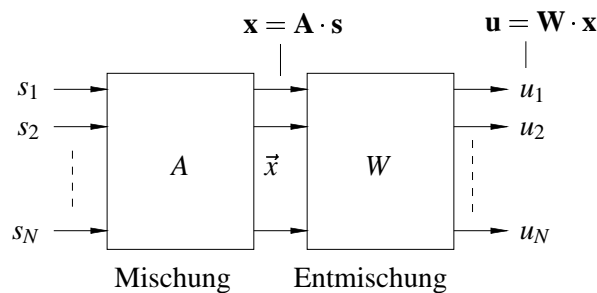


Abbildung 1: Modell der Mischung der Quellsignale \mathbf{s} und der Rekonstruktion von \mathbf{s} durch Entmischung $\mathbf{u} = \hat{\mathbf{s}}$.

Für die Mischungsmatrix \mathbf{A} wird für alle diese Arbeit betreffenden Algorithmen vorausgesetzt, daß \mathbf{A} nur reellwertige Elemente hat; ebenso nehmen alle Signale nur reelle Werte an. Die gemischten Signale entstehen durch Superposition der unterschiedlich skalierten Quellsignale (*instantaneous mixing*). Eine solche Mischung ist bei realen Anwendungen allerdings häufig nicht gegeben. Weitere ICA-Verfahren behandeln die Fälle frequenzabhängiger Mischungen und nichtlinearer Mischungen. Bei der Trennung realer Mehrkanal-Audiosignale oder bei Mobilfunkanwendungen besteht das Problem darin, daß sich neben der einfachen Überlagerung der Quellsignale auch Reflexionen der Quellsignale zu den Mischungen $\mathbf{x}(k)$ addieren (*convolutive mixing*).

Abbildung 2 zeigt den zweidimensionalen Fall einer instantanen Mischung. \mathbf{A} und \mathbf{W} sind in diesem Fall 2×2 -Matrizen, wobei sich die Mischsignale wie folgt berechnen lassen:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}.$$

Mit der Annahme der statistischen Unabhängigkeit der Quellsignale besteht die gemeinsame Zielsetzung der ICA-Algorithmen darin, durch die Wahl von \mathbf{W} statistisch möglichst unabhängige Signale \mathbf{u} zu erhalten. Die Entmischungsmatrix muß dann so gewählt werden, daß

$$\mathbf{W} \cdot \mathbf{A} = \mathbf{I}$$

gilt, wobei für die Spalten der Einheitsmatrix \mathbf{I} Permutationen zugelassen werden. Abhängig vom prinzipiellen Ansatz, wie die statistische Unabhängigkeit der

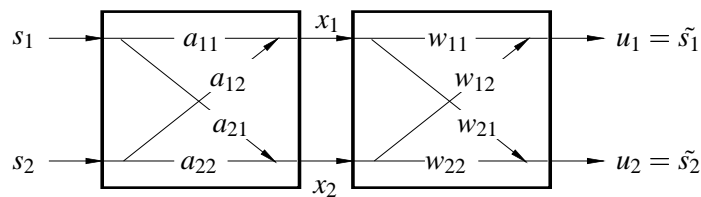


Abbildung 2: Mischung und Entmischung eines Stereosignals mit skalaren Matrizen

Signale \mathbf{s} erreicht wird, unterscheidet man drei Gruppen von Algorithmen zur blinden Quellentrennung:

1. Algebraische Verfahren, die die statistische Unabhängigkeit der Signale \mathbf{s} durch explizite Berechnung statistischer Kennwerte zweiter und höherer Ordnung messen. Die Entmischungsmatrix \mathbf{W} wird durch algebraische Operationen (i.a. Lösung von Eigenwertproblemen) berechnet.
2. Informationstheoretisch basierte Verfahren, die eine Beurteilung der statistischen Unabhängigkeit über den Informationsgehalt vornehmen. Es wird eine Informations-Kostenfunktion definiert, die z.B. über das Gradientenverfahren optimiert werden kann. Für die Messung des Informationsgehalts eines Signals über die in Abschnitt 2.2.1 beschriebenen Maße werden nicht-lineare Transformationen auf das Signal angewendet (siehe 2.2.1). Viele dieser Algorithmen werden auch als neuronale Verfahren bezeichnet, da die Informationsübertragung über einen nichtlinearen Kanal durch ein konnektionistisches Neuron dargestellt werden kann. Da diese Verfahren die Gestalt der Wahrscheinlichkeitsdichteverteilung (PDF) auswerten, verwenden diese Verfahren implizit ebenfalls statistische Kennwerte höherer Ordnung.
3. Verfahren, die auf der Auswertung ausschließlich statistischer Kennwerte zweiter Ordnung basieren. Im Gegensatz zu den algebraischen Verfahren und den informationstheoretisch basierten Verfahren nutzen die Second Order-Verfahren spektrale Unterschiede zur Quellentrennung, die im allgemeinen mit algebraischen Verfahren ausgewertet werden

2 Verfahren zur ICA

2.1 Higher Order Statistics-basierte algebraische Verfahren

2.1.1 Statistische Kennwerte höherer Ordnung

Die algebraischen Verfahren zur ICA nutzen Kumulanten zur Messung der statistischen Unabhängigkeit der Ausgangssignale. Kumulanten sind wie Momente statistische Kennwerte, die eine Aussage über die Gestalt der Wahrscheinlichkeitsdichteverteilung (PDF) machen [Hen98]. Die Momente *Mittelwert* (Moment

und Kumulant erster Ordnung) und *Standardabweichung* (Moment und Kumulant zweiter Ordnung) beschreiben die Lage der PDF und ihre Dispersion. Häufig verwendete Kumulanten sind die *Schiefte* oder *Skewness* (Kumulante dritter Ordnung) und der *Exzess* oder *Kurtosis* als Kumulante vierter Ordnung. Die Skewness beschreibt die Symmetrie der PDF, die Kurtosis beschreibt die Abweichung ihres Verlaufs vom Verlauf einer Gaußverteilung. Flachere Verläufe der PDF werden *subgauß-förmig* genannt, spitzere heißen *supergaußsche* Verteilungen. Da viele Signale symmetrische Verteilungen aufweisen, Momente und Kumulanten dritter Ordnung bei symmetrischen Verteilungen jedoch Null sind, werden in der blinden Signalverarbeitung hauptsächlich statistische Kennwerte vierter Ordnung ausgewertet.

Für die Ermittlung der Entmischungsmatrix werden in einigen Algorithmen Kreuzkumulantenmatrizen berechnet. Ausgehend von diesen Matrizen führen die verschiedenen Verfahren zu Eigenwertproblemen. Das größte Problem besteht in der expliziten Schätzung der Kreuzkumulanten mit ausreichender Genauigkeit. Beispielsweise erfordert die Schätzung der Kurtosis

$$\gamma_{4,x} = E\{x^4(k)\} - 3 \cdot E^2\{x^2(k)\}$$

über die Schätzung des Erwartungswertes

$$E\{x(k)\} = \frac{1}{N} \sum_{k=1}^N x(k) \quad (1)$$

eine Mittelung über eine große Anzahl N von Abtastwerten.

2.1.2 Dekorrelation zweiter Ordnung als Vorverarbeitung

Die Kumulanten von Mittelwert bis Kurtosis erlauben die Bewertung der statistischen Unabhängigkeit bis zur vierten Ordnung. Bei Verwendung von Kennwerten zweiter Ordnung wie der Kovarianz, können nur Transformationen durchgeführt werden, die die Sensorsignale \mathbf{x} in Signale \mathbf{x}' überführen, welche die Kriterien der statistischen Unabhängigkeit nur für die Statistik zweiter Ordnung erfüllen. Bei Verwendung von Kumulanten vierter Ordnung sind Transformationen der Daten möglich, so daß diese auch im Hinblick auf Statistiken bis zur vierten Ordnung statistisch unabhängig sind. Allerdings wird die Dekorrelation bis zur zweiten Ordnung in vielen HOS-basierten Algorithmen (explizit und implizit HOS-basiert) in einer Vorverarbeitungsstufe durchgeführt (siehe Abbildung 3), weil sich der Entmischungsalgorithmus dadurch vereinfacht. Diese Vorverarbeitung ist im Prinzip eine Hauptachsentransformation (principle component analysis, PCA), die einen Spezialfall der ICA darstellt. Bei der Hauptachsentransformation (oder Hauptkomponenten-Analyse) wird ein Signalvektor

$$\mathbf{x}(k) = (x_1(k), x_2(k), \dots, x_M(k))^T$$

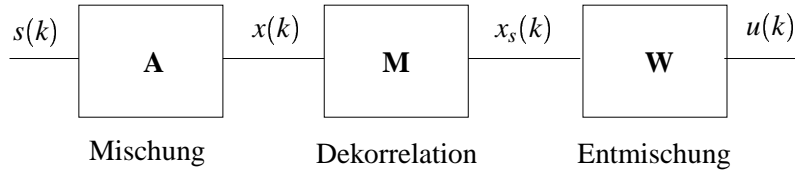


Abbildung 3: ICA mit Vorverarbeitungsstufe

bezüglich einer neuen orthonormalen Basis

$$\mathbf{p}_i = (p_{i_1}, p_{i_2}, \dots, p_{i_M})^T$$

dargestellt durch

$$\mathbf{x}(k) = x'_1(k) \cdot \mathbf{p}_1 + x'_2(k) \cdot \mathbf{p}_2 + \dots + x'_M(k) \cdot \mathbf{p}_M,$$

wobei die Koeffizienten der Transformation gegeben sind durch $x'_i = \mathbf{p}_i^T \cdot \mathbf{x}$ bzw. die *Synthesegleichung der PCA*

$$\mathbf{x}' = \mathbf{P}^T \mathbf{x} \quad \text{mit} \quad \mathbf{P}^T = [\mathbf{p}_1^T \mathbf{p}_2^T \dots \mathbf{p}_M^T]^T. \quad (2)$$

Die Rücktransformation erfolgt mit der *Analysegleichung*

$$\mathbf{x} = \mathbf{P} \mathbf{x}' \quad \text{mit} \quad \mathbf{P} = [\mathbf{p}_1 \mathbf{p}_2 \dots \mathbf{p}_M]. \quad (3)$$

Das Ziel der PCA ist es, eine orthonormale Basis zu finden, deren Basisvektoren in die Richtungen der größten Energieanteile (Hauptkomponenten) des Signals zeigen. Als Bedingung für optimale Basisvektoren folgt unter Ausnutzung der Orthonormalität der Basisvektoren das Eigenwertproblem

$$\mathbf{R} \cdot \mathbf{p}_i = \lambda_i \cdot \mathbf{p}_i, \quad (4)$$

womit sich die Basisvektoren als Eigenvektoren der Autokorrelationsmatrix $\mathbf{R} = E\{\mathbf{x} \cdot \mathbf{x}^T\}$ mit den Eigenwerten λ_i ergeben. Für mittelwertfreie Signale ist die Autokorrelationsmatrix gleich der Kovarianzmatrix \mathbf{C} . Da \mathbf{C} symmetrisch ist, sind alle ihre Eigenwerte reell, und sie kann in ein Produkt dreier Matrizen zerlegt werden (Gleichung 5) [Hof98]. Wegen der Orthogonalität der zu den Eigenwerten gehörenden Eigenräume kann für die symmetrische Matrix \mathbf{C} immer eine orthogonale Matrix \mathbf{P} gefunden werden, so daß $\mathbf{P}^{-1} \mathbf{C} \mathbf{P}$ eine Diagonalmatrix ist, d.h. \mathbf{C} ist darstellbar durch

$$\mathbf{C} = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^T, \quad (5)$$

mit \mathbf{P} als Orthogonalmatrix mit den normierten Eigenvektoren in den Spalten und $\mathbf{\Lambda}$ als Diagonalmatrix mit den Eigenwerten von \mathbf{C} . \mathbf{C} wird durch \mathbf{P} diagonalisiert:

$$\mathbf{\Lambda} = \mathbf{P}^{-1} \mathbf{C} \mathbf{P} \quad (6)$$

$$\mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{\Lambda} \mathbf{\Lambda}^{-\frac{1}{2}} = \mathbf{I} = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{P}^{-1} \cdot \mathbf{C} \cdot \mathbf{P} \mathbf{\Lambda}^{-\frac{1}{2}} \quad (7)$$

Mit der Wahl von $\mathbf{M} = \Lambda^{-\frac{1}{2}} \mathbf{P}^{-1} = (\mathbf{P} \Lambda^{\frac{1}{2}})^{-1}$ ist $\mathbf{C}_s = \mathbf{I}$, d.h. die Kovarianzmatrix $\mathbf{C} = E\{\mathbf{x}_s \cdot \mathbf{x}_s^T\}$ des durch die Dekorrelationsmatrix \mathbf{M} entstandenen Signals $\mathbf{x}_s = \mathbf{M} \cdot \mathbf{x}$ ist eine Einheitsmatrix. Die Signale in \mathbf{x}_s sind damit dekorreliert und auf die Varianzen $\sigma_{x_n}^2 = 1$ normiert. Die PCA diagonalisiert also die Kovarianzmatrix des Signals \mathbf{x} und führt damit eine Dekorrelation bis zur zweiten Ordnung durch. Die Normierung auf die Varianzen $\sigma_{x_n}^2 = 1$ ist eigentlich nicht Teil der PCA.

Mit der Dekorrelationsmatrix \mathbf{M} ergeben sich die dekorrelierten Sensorsignale \mathbf{x}_s aus den Sensorsignalen \mathbf{x} als

$$\mathbf{x}_s = (\mathbf{P} \Lambda^{\frac{1}{2}})^{-1} \cdot \mathbf{x} \quad (8)$$

und aus den zu rekonstruierenden Signalen \mathbf{s} gemäß $\mathbf{x}_s = \mathbf{M} \mathbf{A} \mathbf{s}$. Es gilt

$$E\{\mathbf{x}_s \cdot \mathbf{x}_s^T\} = E\{(\mathbf{M} \mathbf{A} \mathbf{s} \cdot \mathbf{s}^T (\mathbf{M} \mathbf{A})^T)\} = \mathbf{M} \mathbf{A} \cdot E\{\mathbf{s} \cdot \mathbf{s}^T\} \cdot (\mathbf{M} \mathbf{A})^T = \mathbf{I},$$

und mit $E\{\mathbf{s} \cdot \mathbf{s}^T\} = \mathbf{I}$ (es wird statistische Unabhängigkeit der Quellen \mathbf{s} und Varianzen $\sigma_{s_n}^2 = 1$ bei der ICA angenommen) ist

$$\mathbf{M} \mathbf{A} \cdot (\mathbf{M} \mathbf{A})^T = \mathbf{I},$$

d.h. die durch \mathbf{W} zu kompensierende gesamte Mischungsmatrix $\mathbf{M} \mathbf{A}$ ist orthogonal. Diese durch die Vorverarbeitung herbeigeführte Orthogonalität stellt eine wesentliche Vereinfachung für die Suche der Entmischungsmatrix \mathbf{W} bei allen HOS-basierten ICA-Algorithmen dar. Außerdem gingen aus der PCA und der Erweiterung zur nichtlinearen PCA Algorithmen zur ICA hervor (siehe z.B. [OKH96]).

2.1.3 JADE-Algorithmus

Im vorangegangenen Abschnitt wurde dargestellt, daß die Trennung eines N -dimensionalen Zufallsvektors \mathbf{x} in statistisch bis zur zweiten Ordnung unabhängige Anteile durch die Diagonalisierung der Kovarianzmatrix $\mathbf{C}_x = E\{\mathbf{x} \mathbf{x}^T\}$ erfolgt. Die Elemente dieser $N \times N$ -Matrix sind

$$C_{ij} = E\{x_i \cdot x_j\} = \mu_{ij}.$$

Zur Erzeugung einer statistischen Unabhängigkeit höherer Ordnung wertet der JADE-Algorithmus [CS93] Kumulanten κ_x höherer Ordnung von \mathbf{x} aus. Eine wichtige Eigenschaft dieser statistischen Kennwerte ist, daß für statistisch unabhängige Signale die Kreuzkumulanten Null sind, ebenso wie die Kreuzkorrelation bzw. Kovarianz für bis zur zweiten Ordnung unabhängige Signale Null ist. Analog zu \mathbf{C}_x ergibt sich zur Darstellung der Kumulantenstatistik von \mathbf{x} für Kumulanten vierter Ordnung ein Tensor vierter Stufe, der auch als Quadrikovarianz bezeichnet wird:

$$\mathbf{Q}_x = \text{Cum}\{\mathbf{x}, \mathbf{x}^T, \mathbf{x}, \mathbf{x}^T\}$$

Mit $\mu_{ijkl} \stackrel{\text{def}}{=} E\{x_i x_j x_k x_l\}$ sind die Elemente des Tensors \mathbf{Q}_x

$$\text{Cum}(x_i, x_j, x_k, x_l) = \mu_{ijkl} - \mu_{ij} \cdot \mu_{kl} - \mu_{il} \cdot \mu_{kj}.$$

Mit \mathbf{Q}_x als Abbildung $\mathbf{M} \rightarrow \mathbf{N} = \mathbf{Q}_x(\mathbf{M})$ stellt \mathbf{N}_p , $p = 1 \dots N^2$, eine Kumulantenmatrix mit der Dimension $N \times N$ aus einem Satz von N^2 Kumulantenmatrizen dar. Diese Matrizen mit den Elementen $N_p(k, l)$, $1 \leq k, l \leq N$ und $k, l \in \mathbb{N}$ sind die (i, j) -ten Elemente des Tensors \mathbf{Q}_x .

$\mathbf{Q}_x(\mathbf{M})$ als Kumulantenmatrix ist das Gegenstück zur Kovarianzmatrix \mathbf{C}_x für die Darstellung statistischer Kennwerte vierter Ordnung. Wie \mathbf{C}_x ist $\mathbf{Q}_x(\mathbf{M})$ symmetrisch und damit darstellbar durch

$$\mathbf{Q}_x(\mathbf{M}) = \mathbf{V} \Lambda_M \mathbf{V}^T, \quad \Lambda_M = \text{Diag}(\kappa_1 \mathbf{v}_1^T \mathbf{M} \mathbf{v}_1, \dots, \kappa_N \mathbf{v}_N^T \mathbf{M} \mathbf{v}_N). \quad (9)$$

Es gibt N^2 Eigenwerte λ_{M_p} und orthonormale Matrizen \mathbf{M}_p , $p = 1 \dots N^2$, die das Eigenwertproblem $\mathbf{Q}(\mathbf{M}_p) = \lambda_{M_p} \cdot \mathbf{M}_p$ erfüllen. Für eine Dekorrelation bis zur vierten Ordnung ist analog zu Gleichung 4 eine Diagonalisierung des Tensors \mathbf{Q}_x erforderlich, d.h. alle Kumulantenmatrizen $\mathbf{Q}_x(\mathbf{M})$ müssen gleichzeitig diagonalisiert werden. Die Verbunddiagonalisierung (*joint diagonalization*) ermittelt eine orthonormale Basis, mit der alle Matrizen eines Satzes (näherungsweise) diagonalisiert werden können [CC96]. Als Algorithmus zur Verbunddiagonalisierung dient eine Erweiterung des Jacobi-Algorithmus. Der Jacobi-Algorithmus diagonalisiert eine Matrix in einem iterativen Verfahren, bei dem in jedem Iterationsschritt ein Off-Diagonalelement zu Null gesetzt wird.

Da die Kumulanten in der Praxis anhand einer endlichen Anzahl von Werten geschätzt werden und daher ungenau sind, ist die Diagonalisierung nur näherungsweise möglich. Das Ziel des **joint approximate diagonalization of Eigenmatrices** (JADE)-Algorithmus ist daher die Optimierung eines Verbunddiagonalisierungskriteriums. Dabei ist nicht die Diagonalisierung von N^2 Matrizen erforderlich. Es kann gezeigt werden, daß es aufgrund der bei der ICA angenommenen Unabhängigkeit der Quellen nur N Eigenmatrizen mit Eigenwerten $\lambda_{M_p} \neq 0$ gibt; es ist also nur die Diagonalisierung von N Matrizen nötig. Zur Durchführung der ICA mit dem JADE-Algorithmus werden die Sensorsignale \mathbf{x} zunächst durch eine Prewhitening-Matrix vorverarbeitet. Anschließend wird der Satz der Kumulantenmatrizen anhand der dekorrelierten Signale \mathbf{x}_s berechnet. Nach der Berechnung der Eigenpaare $(\lambda_{M_p}, \mathbf{M}_p | 1 \leq p \leq N)$ der größten Eigenwerte wird die Matrix \mathbf{V} bestimmt, die den Satz der Kumulantenmatrizen diagonalisiert.

2.2 Informationstheoretisch basierte Verfahren

2.2.1 Informationstheoretische Grundlagen zur ICA

Die in diesem Abschnitt beschriebenen Ansätze zur blinden Quellentrennung werden oft als Online-Verfahren bezeichnet, weil die Entmischungsmatrix \mathbf{W} bei jedem neu eintreffenden Abtastwert aktualisiert wird. Die Regeln für die Aktualisierung entstehen aus informationstheoretischen Überlegungen. Aufgrund der Annahme statistisch unabhängiger Quellen \mathbf{s} übertragen die einzelnen Kanäle s_i Informationen, die in keinem der anderen Kanäle enthalten sind. Durch die Mischung enthalten die gemischten Signale x_i auch Anteile der Signalinformation aus den

anderen Kanälen, die Signale x_i weisen also Redundanz auf. Die von der Quelle zum Ausgang übertragene Information wird als Transinformation bezeichnet; durch die *gemeinsame Information* (mutual information) der Kanäle ergibt sich die Redundanz der Signale x_i . Eine Trennung kann durch die Reduktion der Redundanz erfolgen.

Eine Aussage über den Informationsgehalt einer Zufallsvariable (bzw. eines Zufallsvektors als mehrdimensionaler Zufallsvariable) macht die Entropie [Gra90]. Die Abtastwerte eines stochastischen Signals werden Realisierungen der Zufallsvariable genannt. Die Information, die mit einer Realisierung verbunden ist, ergibt sich aus der Wahrscheinlichkeit dieser Realisierung, d.h. hier aus der Auftretenswahrscheinlichkeit eines bestimmten Wertes für einen Abtastwert. Die Wahrscheinlichkeitsdichtefunktion (probability density function, PDF) $f_x(x)$ einer Zufallsvariable beschreibt den Verlauf der Wahrscheinlichkeit über der Menge der möglichen Realisierungen. Bei diskreten Zufallsvariablen, die nur diskrete Werte annehmen können, wird anstatt der Funktion $f_x(x)$ die Wahrscheinlichkeit einer Realisierung x_i mit $p_i = P(x = x_i)$ bezeichnet. Mit der PDF $f_x(x)$ der Zufallsvariable x bzw. den Wahrscheinlichkeiten p_i der N möglichen Realisierungen x_i ist die *differentielle Entropie* bzw. im wertdiskreten Fall die *Shannon-Entropie*

$$H(x) = - \int_x f_x(x) \cdot \log f_x(x) dx \quad \text{bzw.} \quad H(x) = - \sum_{k=1}^N p_k \cdot \log p_k. \quad (10)$$

Für einen Vektor \mathbf{x} mit N kontinuierlichen Zufallsvariablen x_1, x_2, \dots, x_N und multivariater PDF ist die differentielle *Verbundentropie*

$$H(\mathbf{x}) = - \int_{\mathbf{x}} f_{\mathbf{x}}(\mathbf{x}) \cdot \log f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}. \quad (11)$$

Eine weitere wichtige Größe ist die *Negentropie*

$$J(\mathbf{x}) = H(\mathbf{x}_{\text{Gauss}}) - H(\mathbf{x}). \quad (12)$$

Sie ist im Gegensatz zur differentiellen Entropie invariant gegenüber Skalierungen. In Gleichung 12 hat die Zufallsvariable $\mathbf{x}_{\text{Gauss}}$ die gleiche Kovarianzmatrix wie \mathbf{x} . Damit ist $J(\mathbf{x})$ ein Maß für die Gaußähnlichkeit von \mathbf{x} . Falls \mathbf{x} eine Gaußverteilung aufweist, nimmt die Negentropie ihren Minimalwert $J(\mathbf{x}_{\text{Gauss}}) = 0$ an, ansonsten ist $J(\mathbf{x})$ größer als Null. Die gemeinsame Information von N Zufallsvariablen x_i als Maß für die Abhängigkeit der Zufallsvariablen voneinander ist gegeben durch

$$I(x_1, \dots, x_N) = \sum_{i=1}^N H(x_i) - H(\mathbf{x}). \quad (13)$$

Bei Zufallsvariablen x_i , die bis zur zweiten Ordnung dekorreliert sind, gilt [Com94]

$$I(x_1, \dots, x_N) = - \sum_{i=1}^N J(x_i) + J(\mathbf{x}). \quad (14)$$

Die Negentropie und die gemeinsame Information werden bei den informationstheoretisch basierten Verfahren verwendet, um Kostenfunktionen zu definieren, die zu Regeln für die korrekte Bildung der Entmischungsmatrix W führen. Prinzipiell erfolgt die Trennung in unabhängige Anteile durch die Maximierung der Information (Redundanzreduktion). Die **Maximierung der Negentropie** entspricht der Suche nach einer PDF, die möglichst weit von der Gaußverteilung entfernt ist. Dieses Ziel deckt sich mit der Aussage des zentralen Grenzwertsatzes.

Die **Minimierung der gemeinsamen Information** entspricht der Maximierung der Transinformation zwischen dem Eingang und dem Ausgang einer Nichtlinearität. Es kann gezeigt werden, daß die Maximierung der Transinformation einer Minimierung der *Kullback-Leibler-Distanz*

$$D(f_{\mathbf{u}}(\mathbf{u})||f_{\mathbf{s}}(\mathbf{u})) = \int_{-\infty}^{\infty} \log \left(\frac{f_{\mathbf{u}}(\mathbf{u})}{f_{\mathbf{s}}(\mathbf{u})} \right) f_{\mathbf{u}}(\mathbf{u}) d\mathbf{u} \quad (15)$$

entspricht. Die Kullback-Leibler-Distanz ist ein Maß für die Unterschiedlichkeit zweier Wahrscheinlichkeitsdichteverteilungen. Da für die Verbundverteilungsdichte statistisch unabhängiger Zufallsvariablen eines Zufallsvektors gilt, daß sie sich aus dem Produkt der Randverteilungsdichten des Zufallsvektors ergibt, ist $D(f_{\mathbf{u}}(\mathbf{u})||f_{\mathbf{s}}(\mathbf{u}))$ ein Maß für die Dekorrelation der Zufallsvariablen. Dabei ist $f_{\mathbf{u}}(\mathbf{u})$ die Verbundverteilung am Ausgang und $f_{\mathbf{s}}(\mathbf{u}) = \prod_k f_{u_k}(u_k)$ das Produkt der Randverteilungen.

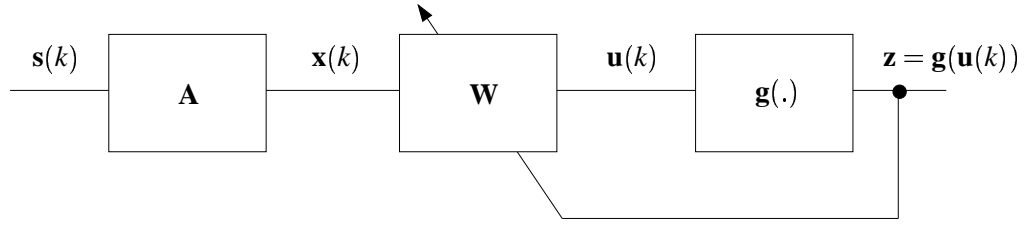
Im allgemeinen wird die Information von Signalen an den Ausgängen von Nichtlinearitäten gemessen (siehe 2.2.2). Durch die Anwendung von nichtlinearen Funktionen auf die Ausgangssignale der Entmischung wird der gesamte Verlauf der PDF ausgewertet, und damit werden implizit statistische Kennwerte höherer Ordnung herangezogen. Die Kostenfunktion wird durch ein Gradientenverfahren oder das Newton-Verfahren optimiert.

Die Alternative zur Anwendung von Nichtlinearitäten ist die explizite Auswertung von Kumulanten. Für eine Minimierung der gemeinsamen Information ist die Kenntnis der Wahrscheinlichkeitsdichten erforderlich, deren direkte Schätzung allerdings schwierig ist. Deshalb werden als Kostfunktionen Approximationen z.B. der Negentropie oder der gemeinsamen Information verwendet. Eine mögliche Schätzung der Negentropie des Signals x ist [Hyv97]

$$\hat{J}(x) = \frac{1}{12} \gamma_{3,x}^2 + \frac{1}{48} \gamma_{4,x}^2. \quad (16)$$

Man approximiert dafür die PDF $f_x(x)$ durch eine Funktion, deren Parameter die Skewness $\gamma_{3,x}$ und die Kurtosis $\gamma_{4,x}$ des Signals x sind. Mit den Gleichungen 10 und 12 erhält man die Schätzung 16.

Die Optimierung der Kostenfunktion mit einer numerischen Methode führt auf eine Vorschrift zur Bildung von \mathbf{W} , die als Lernregel für eine Realisierung der ICA durch ein neuronales Netz angesehen werden kann. Die Spalten \mathbf{w}_i von \mathbf{W} sind dabei die Gewichte der Neuroneneingänge; die nichtlineare Funktion ist die Aktivierungsfunktion des Neurons. Durch die Anpassung der Gewichte lernt das neu-

Abbildung 4: Infomax-Algorithmus mit Nichtlinearität $g(\cdot)$.

ronale Netz eine Entmischungsmatrix, die zu statistisch unabhängigen Ausgangssignalen des Netzes führt. Allgemein gilt für die aktualisierte Entmischungsmatrix

$$\mathbf{W}(k+1) = \mathbf{W}(k) - \mu(k) \cdot f(\mathbf{W}(k), \mathbf{x}(k)).$$

Die speziellen Algorithmen sind durch die Festlegung von $f(\cdot, \cdot)$ gegeben. $\{\mu(k)\}$ ist eine Folge von Werten, die die Konvergenzgeschwindigkeit und die Genauigkeit bestimmt. Entsprechend kann auch eine Vorverarbeitungs-(Dekorrelations-) Matrix durch ein neuronales Netz gelernt werden [Kar96]:

$$\mathbf{M}(k+1) = \mathbf{M}(k) - \mu(k) (\mathbf{x}(k)\mathbf{x}^T(k) - \mathbf{I})\mathbf{M}(k)$$

2.2.2 Infomax-Algorithmus

Das Ziel des Infomax-Algorithmus von Bell und Sejnowski [BS95] [Car97] ist die Maximierung der Entropie am Ausgang der Nichtlinearität. Die zu optimierende Kontrastfunktion ist

$$\phi_I(\mathbf{W}) = H(\mathbf{g}(\mathbf{W}\mathbf{x})).$$

Dabei wird nicht die Entropie von $\mathbf{u}(k)$ maximiert, sondern von $\mathbf{z} = \mathbf{g}(\mathbf{u}(k))$ mit der monoton steigenden nichtlinearen Funktion $\mathbf{g}(\mathbf{u}) = (g_1(u_1), g_2(u_2), \dots, g_N(u_N))$. Sie bildet die Menge der reellen Zahlen als Wertebereich von $\mathbf{u}(k)$ auf das Intervall $[0, 1]$ ab und führt damit eine Amplitudenbegrenzung durch (*squashing function*). Durch die Amplitudenbegrenzung führt die Maximierung der Entropie zu einer Änderung der PDF hin zu einer uniformen Wahrscheinlichkeitsdichteverteilung, da bei amplitudenbegrenzten Signalen uniform verteilte Signale die größte Entropie aufweisen. Bei nicht amplitudenbegrenzten Signalen haben gaußverteilte Signale die größte Entropie.

Wenn $\mathbf{g}(\cdot)$ differenzierbar ist, dann kann $\mathbf{g}(\cdot)$ als Verteilungsfunktion

$$g_i(s) = \int_{-\infty}^s f_u(u) du$$

einer Wahrscheinlichkeitsdichtefunktion $f_u(u)$ angesehen werden. Es kann gezeigt werden, daß die Kontrastfunktion durch die Kullback-Leibler-Distanz ausgedrückt werden kann [Car97]:

$$\phi_I(\mathbf{W}) = -D(\mathbf{g}(\mathbf{W}\mathbf{x}) || \mathbf{x}_1)$$

Dabei ist \mathbf{x}_1 ein Zufallsvektor mit uniformen Verteilungen. Die Minimierung der Kullback-Leibler-Distanz zwischen der Ausgangs-Verbundverteilungsdichte und der Gleichverteilung entspricht damit der Maximierung der Entropie bzw. der Maximierung der Information. Für die Bildung von \mathbf{W} ergibt sich mit einem Gradientenverfahren zur Minimierung der Kullback-Leibler-Distanz die Vorschrift [Kar96] [BS95]

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k) \left(\mathbf{W}^{-T}(k) + \mathbf{v}(k) \mathbf{x}^T(k) \right) \quad (17)$$

mit

$$v_i = \frac{\partial}{\partial z_i} \frac{\partial z_i}{\partial u_i} \quad \text{und} \quad \mathbf{z} = \mathbf{g}(\mathbf{u}) . \quad (18)$$

Die Wahl von $\mathbf{g}(\cdot)$ muß unter Berücksichtigung der für das Eingangssignal \mathbf{s} angenommenen Verteilung erfolgen, weil das Ziel des Algorithmus darin besteht, die Kullback-Leibler-Distanz zwischen der PDF der Ausgangsverteilung und der vermuteten Eingangsverteilung zu minimieren.

Die Realisierung der Vorschrift 17 durch ein neuronales Netz ist wegen der Matrixinvertierung \mathbf{W}^{-T} nicht möglich. Eine Möglichkeit zur Vermeidung der Berechnung von \mathbf{W}^{-T} besteht in der in Abschnitt 2.1.2 beschriebenen Vorverarbeitung. Durch die Dekorrelation zweiter Ordnung ist \mathbf{W} orthogonal, und es gilt $\mathbf{W}^{-1} = \mathbf{W}^T$ bzw. $\mathbf{W}^{-T} = \mathbf{W}$; zusätzlich erhöht sich die Konvergenzgeschwindigkeit. Als weitere Variante ergibt sich bei der Verwendung des natürlichen Gradienten (siehe [ACY96]) zur Optimierung der Kontrastfunktion mit der Nichtlinearität $\mathbf{z} = \tanh(\mathbf{u})$, also $\mathbf{v} = -2 \tanh(\mathbf{u})$, die Lernregel

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k) \left(\mathbf{I} - 2 \tanh(\mathbf{u}(k)) \mathbf{u}^T(k) \right) \mathbf{W}(k) , \quad (19)$$

die auch eine bessere Konvergenz als Gleichung 17 aufweist. Eine Reihe von Algorithmen kommt aufgrund ähnlicher Ansätze zu weiteren Kontrastfunktionen, deren Optimierung auf prinzipiell sehr ähnliche Lernregeln für \mathbf{W} führt. Für einige Algorithmen, die als Feedforward-Netze realisiert werden können, werden hier exemplarisch die Lernregeln aufgeführt.

2.2.3 Neuronale Verfahren zur ICA

Einer der ersten ICA-Algorithmen war der **Herault-Jutten-Algorithmus** (siehe [Kar96]). Er weist eine schlechte Trennungsleistung für mehr als zwei Quellen auf, diente aber als Ausgangspunkt für einige andere Algorithmen. Die Entmischungsmatrix ist

$$\mathbf{W} = (\mathbf{I} + \mathbf{S})^{-1} ,$$

wobei \mathbf{S} eine Matrix ist, deren Diagonalelemente Null sind. Die Bildung dieser Matrix erfolgt nach der Vorschrift [BS95]

$$\mathbf{S}(k+1) = \mathbf{S}(k) + \mu(k) \cdot \mathbf{g}(\mathbf{u}(k)) \cdot \mathbf{h}(\mathbf{u}^T(k)) .$$

Für $g(\cdot)$ und $h(\cdot)$ werden zwei verschiedene ungerade Funktionen gewählt, z.B. $g(x) = x^3$ und $h(x) = x$ für die Trennung subgaußscher Quellen und $g(x) = x$ und $h(x) = x^3$ für die Trennung supergauß-verteilter Quellsignale [Kar96]. Weitere häufig verwendete Nichtlinearitäten, die ebenfalls in vielen anderen Algorithmen verwendet werden, sind

$$g(x) = \operatorname{sgn}(x) \quad \text{und} \quad g(x) = \tanh(x) .$$

Für die Umsetzung des Algorithmus in einem neuronalen Netz wird die Matrixinvertierung durch die Approximation $\mathbf{W} = \mathbf{I} - \mathbf{S}$ vermieden, wobei die Trennungsleistung unter dieser Approximation kaum leidet.

Aufbauend auf dem Herault-Jutten-Algorithmus haben S. Amari et al. weitere neuronale Netze mit Lernregeln für die blinde Quellentrennung und blinde Identifikation vorgestellt [ACY96]. Mit der allgemeinen Lernregel

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k) \cdot \mathbf{G}(\mathbf{u}(k)) \quad (20)$$

und einer Matrixfunktion $\mathbf{G}(\cdot)$ ergibt sich mit

$$\mathbf{G}(\mathbf{u}) = \mathbf{I} - \mathbf{g}(\mathbf{u}) \cdot \mathbf{h}(\mathbf{u}^T)$$

ein modifizierter Herault-Jutten-Algorithmus. Ebenfalls vorgeschlagen für die Wahl der Matrixfunktion wurde

$$\mathbf{G}(\mathbf{u}) = \mathbf{I} - \mathbf{u} \cdot \mathbf{u}^T - \mathbf{g}(\mathbf{u})\mathbf{h}(\mathbf{u}^T) + \mathbf{h}(\mathbf{u})\mathbf{g}(\mathbf{u}^T) .$$

\mathbf{G} entspricht dann mit $h(x) = x$ dem Aktualisierungsterm für \mathbf{W} des EASI-Algorithmus (*Equivalent Adaptive Separation via Independence*) von Cardoso (siehe [CL96]). Dessen Lernregel lautet

$$\mathbf{W}(k+1) = \mathbf{W}(k) - \mu(k) \left(-\mathbf{I} + \mathbf{u} \cdot \mathbf{u}^T + \mathbf{g}(\mathbf{u})\mathbf{u}^T - \mathbf{u}\mathbf{g}(\mathbf{u}^T) \right) \mathbf{W}(k) \quad (21)$$

2.2.4 Fixed-Point-Algorithmus

Die neuronalen Verfahren des vorangegangenen Abschnitts bieten bei nichtstationären Signalen eine schnelle Adaption, ihre Konvergenzgeschwindigkeit ist allerdings gering und die Lernrate muß geeignet gewählt werden. Der Fixed-Point-Algorithmus [Hyv97] weist eine hohe Konvergenzgeschwindigkeit auf, und es muß keine Lernrate gewählt werden. Trotz des informationstheoretischen Ansatzes ist der fixedpoint-Algorithmus kein neuronales Verfahren, weil hier statistische Kennwerte des Signals über den Erwartungswert geschätzt werden. Deshalb werden hier im Gegensatz zu den anderen vorgestellten informationstheoretisch basierten Verfahren pro Iteration mehrere Abtastwerte anstatt nur eines Wertes verwendet (*batch-Algorithmus*).

In Abschnitt 2.2.1 wurde die Darstellung der Negentropie durch statistische Kennwerte gemäß Gleichung 16 beschrieben. Bei symmetrischen Verteilungen mit $\gamma_{3,x} = 0$ gilt für die Negentropie

$$J \propto \gamma_{4,x}^2.$$

Eine allgemeine Darstellung der Negentropie über statistische Kennwerte ist z.B. mit der Approximation über N Erwartungswerte nichtlinearer Transformationen des Ausgangssignals \mathbf{u} gegeben durch

$$J(\mathbf{u}) = \frac{1}{2} \sum_{i=1}^N E^2 \{g_i(\mathbf{w}_i^T \mathbf{x})\}. \quad (22)$$

Mit der Beschränkung auf eine Nichtlinearität $g(\cdot)$ ergibt sich die Kontrastfunktion für die Bestimmung der Spaltenvektoren \mathbf{w}_i von \mathbf{W} :

$$J(\mathbf{w}_i) = \left| E\{g(\mathbf{w}_i^T \mathbf{x})\} - E\{g(\mathbf{v}_{\text{Gauss}})\} \right|$$

Dabei wird die Trennung $\mathbf{u} = \mathbf{w}_i^T \mathbf{x}$ auf ein gaußverteilttes Signal bezogen (siehe Definition der Negentropie in Gleichung 12). Zur Optimierung der Kontrastfunktion müssen die Minima oder Maxima aller $J(\mathbf{w}_i)$, $i = 1 \dots N$ für N Ausgangssignale \mathbf{u} , ermittelt werden. Die Bestimmung einer Extremalstelle einer Kontrastfunktion $J(\mathbf{w}_i)$ erfolgt über das Newton-Verfahren, das in einem iterativen Vorgang eine Nullstelle einer Funktion ermittelt (siehe [GZZ95], S. 465). Wenn die Anfangsschätzung der Nullstelle hinreichend gut ist, weist das Newton-Verfahren quadratische Konvergenz auf. Mit dem Newton-Verfahren wird eine Nullstelle der Ableitung der Kontrastfunktion und damit eine Extremalstelle der Kontrastfunktion bestimmt. Die Schätzung mehrerer \mathbf{w}_i geschieht durch die Anwendung des Gram-Schmidt-Orthogonalisierungsverfahrens (siehe [GZZ95], S. 390) unter Ausnutzung der Orthogonalität von \mathbf{W} , die durch die Vorverarbeitung gemäß Abschnitt 2.1.2 gegeben ist.

2.3 Second Order Statistics-basierte Verfahren

Bei allen Verfahren zur blinden Quellentrennung ist es das gemeinsame Ziel, $N^2 - N$ linear unabhängige Gleichungen zur eindeutigen Bestimmung von insgesamt $N^2 - N$ Unbekannten aufzustellen. N^2 Unbekannte ergeben sich durch die unbekannte Mischungsmatrix. Da die Skalierungen der N Zeilen der Entmischungsmatrix frei wählbar sind, reduziert sich die Anzahl der Unbekannten um N . Die bisher vorgestellten Verfahren nutzen implizit oder explizit Kumulanten höherer Ordnung, um alle für eine ICA notwendigen Informationen zu erhalten, da alleine durch die Verwendung der Kovarianz eine Bestimmung aller $N^2 - N$ Unbekannten nicht möglich ist.

Bei den Verfahren dieses Abschnitts, die ausschließlich statistische Kennwerte zweiter Ordnung nutzen, wird weiterhin angenommen, daß die Quellsignale gegenseitig unkorreliert sind. Zusätzlich wird vorausgesetzt, daß die Quellsignale

autokorreliert sind, d.h. ihre Spektren sollen nicht weiß sein und müssen sich unterscheiden. Die Ausnutzung der Autokorrelation bzw. der spektralen Gestalt liefert bei diesen Verfahren die Gleichungen zur Bestimmung der restlichen Unbekannten.

Bekannte Verfahren, die auf Statistiken zweiter Ordnung basieren, sind u.a. der Algorithmus von Molgedey und Schuster [MS94], der SOBI-Algorithmus von Belouchrani [AMBCM94], die Verfahren von Köhler und Orglmeister [KO98] [KO99] [KOB98] und die hier nicht betrachteten Verfahren von Tong und Féty (Referenzen in [NP97]). Alle Verfahren basieren auf dem gleichen Ansatz: Die zusätzliche spektrale Information wird durch die Berücksichtigung des zeitlichen Verlaufs der Signale bestimmt. Das Ziel ist, die Korrelation nicht nur zwischen gegeneinander unverschobenen Signalen zu minimieren, sondern auch zwischen zueinander zeitlich verzögerten Signalen.

Ausgehend von der Annahme, daß die Quellsignale für beliebige Zeitverzögerungen τ voneinander unabhängig sind, d.h.

$$E\{s_i(k)s_j(k+\tau)\} = 0, \quad 1 \leq i, j \leq N, \quad \tau \in \mathbb{Z},$$

aber farbige Spektren aufweisen, also

$$E\{s_i(k)s_i(k+\tau)\} \neq 0, \quad 1 \leq i \leq N, \quad \tau \in \mathbb{Z}$$

gilt, werden neben der Kovarianzmatrix \mathbf{C}_x weitere Kovarianzmatrizen berechnet:

$$\begin{aligned} \mathbf{C}_x(\tau_i) &= E\{\mathbf{x}(k)\mathbf{x}^T(k+\tau_i)\} = \\ &= E\{\mathbf{A}\mathbf{s}(k)(\mathbf{A}\mathbf{s}(k+\tau_i))^T\} = \mathbf{A} \cdot E\{\mathbf{s}(k)\mathbf{s}^T(k+\tau_i)\} \cdot \mathbf{A}^T = \mathbf{A}\mathbf{\Lambda}_i\mathbf{A}^T. \end{aligned} \quad (23)$$

Die Matrizen $\mathbf{\Lambda}_i$ sind wegen der statistischen Unabhängigkeit der Quellen Diagonalmatrizen. Der Satz $\mathbf{C}_x(\tau_i)$ von Kovarianzmatrizen enthält neben der Information über die Unabhängigkeit der Ausgangssignale \mathbf{x} auch Informationen über die Spektren, da bei Betrachtung des gesamten Satzes $\mathbf{C}_x(\tau_i)$ auch eine Aussage über die Autokorrelationen gemacht wird.

Der **Algorithmus von Molgedey und Schuster** [MS94], der eines der bekanntesten Second Order-Verfahren ist, betrachtet nur die zwei Matrizen \mathbf{C}_x und $\mathbf{C}_x(\tau)$. Die Kombination dieser beiden Matrizen liefert

$$\mathbf{C}_x\mathbf{C}_x^{-1}(\tau)\mathbf{A} = \mathbf{A}(\mathbf{\Lambda}\mathbf{\Lambda}_\tau^{-1}) \quad \text{bzw.} \quad \mathbf{C}_x\mathbf{C}_x^{-1}(\tau) = \mathbf{A}(\mathbf{\Lambda}\mathbf{\Lambda}_\tau^{-1})\mathbf{A}^{-1}. \quad (24)$$

Diese Kombination der Kovarianzmatrizen überführt das Problem der Verbunddiagonalisierung in ein Eigenwertproblem, das mit Standardmethoden gelöst werden kann. Molgedey und Schuster beschreiben außerdem eine Lernregel zur Lösung von Gleichung 24 durch ein neuronales Netz zum Vergleich mit dem Algorithmus von Herault und Jutten. Wegen der Unsymmetrie der Matrix $\mathbf{C}_x\mathbf{C}_x^{-1}(\tau)$ können sich auch bei Mischungsmatrizen mit ausschließlich reellen Elementen komplexe Eigenwerte ergeben. Ebenfalls problematisch ist bei diesem Verfahren die Wahl

von τ .

Der **second order blind identification**-Algorithmus (SOBI) von Belouchrani et al. [AMBCM94] nutzt den bereits beschriebenen Vorteil der Dekorrelation (Prewhitening) in einer Vorverarbeitungsstufe: Durch das Prewhitening mit der Matrix \mathbf{M} beschränkt sich die Suche nach der Mischungsmatrix auf eine orthogonale Matrix $\mathbf{O} = \mathbf{M}\mathbf{A}$, und für den Satz von Kovarianzmatrizen gilt mit $\mathbf{x}_s = \mathbf{M}\mathbf{x}$:

$$\mathbf{C}_x(\tau_i) = E\{\mathbf{x}_s \mathbf{x}_s^T(k + \tau_i)\} = \dots = \mathbf{O} \Lambda_i \mathbf{O}^T.$$

Aus den $\mathbf{C}_x(\tau_i)$ wird im Gegensatz zum Molgedey-Schuster-Algorithmus keine kombinierte Kovarianzmatrix gebildet, sondern eine Verbunddiagonalisierung an diesem Satz vorgenommen. Als Verfahren zur Bestimmung der orthogonalen Matrix \mathbf{O} , die alle Matrizen $\mathbf{C}_x(\tau_i)$ (näherungsweise) diagonalisiert,

$$\mathbf{O}^T \mathbf{C}_x(\tau_i) \mathbf{O} = \Lambda_i,$$

wird die erweiterte Jacobi-Transformation zur Minimierung eines Verbunddiagonalisierungskriteriums vorgeschlagen. Für einen Satz mit mehr als zwei Kovarianzmatrizen führt die Verbunddiagonalisierung zu besseren Ergebnissen als die Bildung einer kombinierten Kovarianzmatrix. Durch die große Anzahl von Matrizen und das Prewhitening wird außerdem das Problem der Wahl von τ entschärft. Die Verfahren von Köhler und Orglmeister verwenden Filter anstatt verschiedener Zeitverzögerungen τ_i zur Bildung eines Satzes von Kovarianzmatrizen. Da Filter als lineare Operatoren keinen Einfluß auf die statistische Unabhängigkeit der Signale haben, aber die Spektren beeinflussen, können sie verwendet werden, um weitere Kovarianzmatrizen aufzustellen.

Die Anwendung linearer Operatoren $\mathbf{T}_i(\mathbf{x}) = (T_i(x_1), T_i(x_2), \dots, T_i(x_n))^T$ durch die Berechnung von Kovarianzmatrizen der dekorrelierten (Prewhitening-Matrix \mathbf{M}) und transformierten (Operator \mathbf{T}) Sensorsignale \mathbf{x}_s führt zu verschiedenen Eigenwertproblemen. Mit $\mathbf{O} = \mathbf{M}\mathbf{A}$ bzw. $\mathbf{A} = \mathbf{M}^{-1}\mathbf{O}$ und $\mathbf{x}_s = \mathbf{M}\mathbf{x} = \mathbf{M}\mathbf{A}\mathbf{s} = \mathbf{O}\mathbf{s}$ ist

$$\begin{aligned} \mathbf{C}_x(\mathbf{T}_i) &= E\{\mathbf{T}_i(\mathbf{x}_s) \mathbf{T}_i(\mathbf{x}_s^T)\} = E\{\mathbf{T}_i(\mathbf{O}\mathbf{s}) \mathbf{T}_i[(\mathbf{O}\mathbf{s})^T]\} = \\ &= \mathbf{O} E\{\mathbf{T}_i(\mathbf{s}) \mathbf{T}_i(\mathbf{s}^T)\} \mathbf{O}^T = \mathbf{O} \cdot \mathbf{D}_{T_i} \cdot \mathbf{O}^T. \end{aligned} \quad (25)$$

Die Matrix \mathbf{D}_{T_i} ist eine Diagonalmatrix, da trotz der linearen Transformation \mathbf{T} die statistische Unabhängigkeit der Quellen \mathbf{s} erhalten bleibt. Die Wahl der Transformationen \mathbf{T}_i beim **AR-Modell**-Ansatz [KO99] soll so erfolgen, daß durch die Transformation Kreuzkorrelationen hervorgehoben werden, um bei der Diagonalisierung zur Minimierung der Kreuzkorrelationen bessere Ergebnisse zu erhalten. Für jedes der N Signale wird dessen Spektrum durch ein AR-Modell approximiert. Bei der Filterung aller Signale mit dem AR-Filter eines Signals werden gemeinsame Anteile hervorgehoben. Insgesamt werden N AR-Filter bestimmt. Nach der

Filterung aller N Signale mit allen N Filtern werden N Kovarianzmatrizen berechnet und gemeinsam diagonalisiert, entweder durch Bildung einer kombinierten Kovarianzmatrix oder durch das JADE-Diagonalisierungsverfahren.

Alternativ wird in [KOB98] die Verwendung von signalangepaßten Filtern für \mathbf{T}_i vorgeschlagen (**matched filter**-Ansatz). Durch die Filterung sollen nicht nur neue Bedingungen für die Bestimmung der Unbekannten aufgestellt werden, sondern gezielt spektrale Komponenten hervorgehoben werden. Dafür muß die Wahl von \mathbf{T}_i speziell für die Bearbeitung einer Gruppe von Signalen (z.B. EEG-Daten) so erfolgen, daß bestimmte Signaleigenschaften hervorgehoben bzw. unterdrückt werden; die Impulsantwort der Filter \mathbf{T}_i muß also an gesuchte Signalmuster angepaßt sein.

Eine weitere Variante des Ansatzes 25 stellt das **Wavelet-basierte ICA**-Verfahren [KO98] dar. Die Operatoren \mathbf{T}_i sind hier die Filter einer Filterbank, die eine Wavelet-Zerlegung durchführt. Dabei wird die wichtige Eigenschaft der Orthogonalität der Wavelet-Transformation ausgenutzt.

Literatur

- [ACY96] S. Amari, A. Chichocki, and H. Yang. A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems 8 (Proc. NIPS'95)*, Cambridge, MA, 1996. MIT Press.
- [AMBCM94] K. Abed-Meraim, A. Belouchrani, J.-F. Cardoso, and E. Moulines. Asymptotic performance of second order blind source separation. *Proceedings of the ICASSP*, 4:277–280, April 1994.
- [BS95] A. Bell and T. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.
- [Car97] J.-F. Cardoso. Infomax and maximum likelihood for source separation. *IEEE Letters on Signal Processing*, 4(4):112–114, April 1997.
- [CC96] J.-F. Cardoso and P. Comon. Independent component analysis, a survey of some algebraic methods. In *Proceedings ISCAS'96*, vol. 2, pages 93–96, 1996.
- [CL96] J.-F. Cardoso and B. Laheld. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 44(12):3017–3030, December 1996.
- [Com94] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36:287–314, 1994.
- [CS93] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non gaussian signals. *IEE Proceedings-F*, 140(6):362–370, 1993.
- [Gra90] Robert M. Gray. *Entropy and Information Theory*. Springer Verlag, New York, 1990. ISBN 0-387-97371-0.
- [GZZ95] Dr. G. Grosche, Prof. Dr. R. Zeidler, and D. Ziegler, editors. *Teubner-Taschenbuch der Mathematik Teil II*. ISBN 3-8154-2100-4. B.G. Teubner Verlagsgesellschaft Leipzig, 1995.
- [Hen98] Carsten Hennig. Untersuchung von verfahren der blinden entfaltung für die segmentierung elektroenzephalographischer daten, 1998. Studienarbeit am Institut für Elektronik und Lichttechnik, TU-Berlin.
- [Hof98] Rüdiger Hoffmann. *Signalanalyse und -erkennung*. ISBN 3-540-63443-6. Springer-Verlag Berlin Heidelberg New York, 1998.
- [Hyv97] A. Hyvärinen. Independent component analysis by minimization of mutual information. Technical Report A46, Helsinki University

- of Technology, Laboratory of Computers and Information Science, August 1997.
- [Kar96] Juha Karhunen. Neural approaches to independent component analysis and source separation. In *Proceedings of the 4th European Symposium on Artificial Neural Networks*, pages 249–266, Bruges, Belgium, April 24–26 1996. ESANN 96.
- [KO98] B.-U. Köhler and R. Orglmeister. Independent component analysis of electroencephalographic data using wavelet decomposition. In *Proceedings of the VIII Mediterranean Conference on Medical and Biological Engineering and Computing*, Lemesos, Cyprus, Juni 1998. Medicon’98.
- [KO99] B.-U. Köhler and R. Orglmeister. Independent component analysis using autoregressive models. In *Proceedings of the First International Workshop on Independent Component Analysis and Blind Separation of Signals*, pages 359–363, Aussois, France, January 1999.
- [KOBF98] B.-U. Köhler, R. Orglmeister, and B. Brehmeier-Flick. Blind source separation of eeg data using matched filters. In *20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Hong Kong, October 1998. EMBS98.
- [MS94] L. Molgedey and H.G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.*, 72(3634), 1994.
- [NP97] J.-P. Nadal and N. Parga. Redundancy reduction and independent component analysis: Conditions on cumulants and adaptive approaches. *Neural Computing*, 9(7):1421–1456, October 1997.
- [OKH96] E. Oja, J. Karhunen, and A. Hyvärinen. From neural pca to neural ica. In *NIPS Post-Conference Workshop on Blind Signal Processing*, Snowmass, Colorado, December 7 1996.