

Blind Source Separation: Eine Einführung

A Tutorial for the Course *Computational Intelligence*

<http://www.igi.tugraz.at/lehre/CI>

Michael Wohlmayr

Signal Processing and Speech Communication Laboratory
Inffeldgasse 16c

Abstract

This tutorial introduces Principal Component Analysis (PCA) and Independent Component Analysis (ICA). The differences are described and the application to Blind Source Separation is mentioned.

Usage

To make full use of this tutorial you have to

1. Download the file `BSS.zip` which contains this tutorial and the accompanying Matlab program(s).
2. Unzip `BSS.zip` which will generate a subdirectory `BSS/matlab` where you can find the Matlab program(s).
3. Add the path `BSS/matlab` to the matlab search path with a command like `addpath('C:\Work\BSS\matlab')` if you are using a Windows machine or `addpath('/home/jack/BSS/matlab')` if you are using a Unix/Linux machine.

1 Einführung

Blind Source Separation beschäftigt sich mit der Problemstellung, aus einer Menge von Beobachtungen die darin enthaltenen unabhängigen Komponenten zu extrahieren. Eine unabhängige Komponente ist dabei ein Ereignis, das ursächlich nicht mit anderen beobachteten Ereignissen in Verbindung steht. Ein zur Veranschaulichung dieser Problemstellung oft formuliertes Beispiel ist das sog. *Cocktail Party Problem* [1], siehe auch Abbildung 1:

In einem Veranstaltungssaal befinden sich mehrere Personen, die sich in Gruppen miteinander unterhalten. Im Raum sind an verschiedenen Positionen Mikrophone angebracht, die die gesamte Geräuschkulisse im Raum aufzeichnen. Ziel der Blind Source Separation ist es hier, unter Zuhilfenahme der Mikrophonsignale jeweils eine Stimme von allen anderen Stimmen zu trennen.

Zur Lösung dieses Problems existieren mehrere Ansätze. Einer davon ist die sogenannte *Independent Component Analysis* (ICA) [5] [4], [3] auf die in diesem Artikel besonderes Augenmerk gelegt wird.

In Sektion 2 wird zunächst auf die mathematische Formulierung der Problemstellung eingegangen. Sektion 3 beschäftigt sich mit der sog. *Principal Component Analysis* (PCA), die in vielen Fällen einen wichtigen Vorverarbeitungsschritt für ICA darstellt. In Sektion 4 werden grundlegende Methoden der Independent Component Analysis vorgestellt.

2 Modellierung der Problemstellung

Independent Component Analysis ist eine statistische Methode, die auf vereinfachenden Annahmen über den zugrundeliegenden Prozess und Messvorgang basiert:

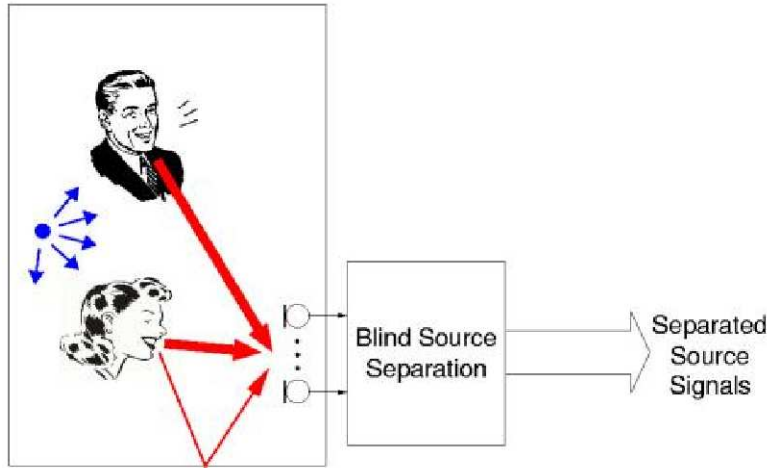


Figure 1: Das Cocktail Party Problem.

- Typischerweise werden die vermischten Signale, aus der die unabhängigen Komponenten extrahiert werden sollen, durch ein Array von M Mikrophonen/Sensoren aufgezeichnet. Die Anzahl M der Mikrophone spielt hier eine große Rolle, denn ICA ist nur durchführbar, wenn die Anzahl der unabhängigen Komponenten N kleiner oder höchstens gleich der Anzahl der Sensoren M ist.
- Der zugrundeliegende Mischprozess wird in der Modellierung stark vereinfacht. Während in der Realität ein unabhängiges Quellsignal $s[n]$ zwischen Quelle i und Sensor j durch eine beliebige Übertragungsfunktion $H_{ji}(\omega)$ gefiltert wird (z.B.: Raumimpulsantwort in einem Büro), wird bei ICA der Übertragungsweg lediglich durch ein reines Skalar a_{ji} modelliert.
- Der Effekt von additivem Rauschen, wie er in jedem physikalischen System auftritt, wird im Modell ignoriert.

2.1 Modellierung des Signals

Ein zeitdiskretes Signal $x[n]$ wird im Folgenden als Zufallsvariable (ZV) x mit einer bestimmten Wahrscheinlichkeitsdichtefunktion (pdf) modelliert. Jeder zeitdiskrete Abtastwert $x[n]$ stellt dabei eine *Realisierung* des durch die zugehörige pdf vollständig charakterisierten Zufallsexperiments dar. Dementsprechend wird ein Array $\mathbf{x}[n]$ von Signalen

$$\mathbf{x}[n] = (x_1[n], x_2[n], \dots, x_N[n])^T$$

als Zufallsvektor

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

modelliert.

2.2 Modellierung des Mischvorgangs

Wie bereits erwähnt, wird die Übertragungsfunktion des Signalpfads zwischen Quelle i und Sensor j durch ein Skalar a_{ji} modelliert. Da ein Sensor x_j die Summe aller unabhängigen Signale s_i aufzeichnet, ergibt sich:

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jN}s_N$$

Für die Gesamtheit der Sensorsignale ergibt sich daher (siehe auch Abbildung 2):

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

Dabei ist

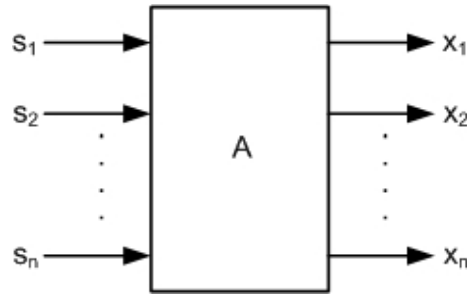


Figure 2: Modellierung des Mischvorgangs durch eine Mischmatrix \mathbf{A} .

- der Sensorvektor \mathbf{x} ein Zufallsvektor der Dimension $M \times 1$,
- der Quellvektor \mathbf{s} ein Zufallsvektor der Dimension $N \times 1$,
- die Mischmatrix \mathbf{A} eine Matrix der Dimension $M \times N$:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \dots & a_{MN} \end{pmatrix}.$$

2.3 Prinzip des Lösungsverfahrens

2.3.1 Statistische Unabhängigkeit

Ziel der Independent Component Analysis ist die Extraktion der unabhängigen Komponenten in \mathbf{s} . Dazu benötigt man eine *Entmischungsmatrix* \mathbf{W} :

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}. \quad (2)$$

Dabei soll $\hat{\mathbf{s}}$ eine möglichst genaue Annäherung bzw. Schätzung von \mathbf{s} sein.

Die Problematik besteht offensichtlich darin, dass keine Informationen über \mathbf{A} und \mathbf{s} vorliegen. (Wäre die Mischmatrix \mathbf{A} bekannt, könnte die Entmischungsmatrix \mathbf{W} sofort durch Invertierung von \mathbf{A} gefunden werden). Die einzigen Annahmen, die zur Vereinfachung über \mathbf{s} getroffen werden, sind:

- Die unabhängigen Komponenten in \mathbf{s} (und somit auch in \mathbf{x}) sind mittelwertfrei. Sollte dies nicht zutreffen, kann \mathbf{x} stets durch Subtraktion des Mittelwerts zentriert werden.
- Die Varianz der einzelnen Komponenten in \mathbf{s} ist 1. Selbst wenn dies nicht der Wahrheit entsprechen sollte - das Gegenteil kann nur schwer bewiesen werden. Das willkürliche Normieren der Varianz auf 1 bewirkt dann lediglich eine Änderung der entsprechenden Koeffizienten in der Mischmatrix \mathbf{A} - die Modellierung bleibt gültig.

Der eigentliche Schlüssel zur Lösung des Problems liegt jedoch in der Annahme, dass die einzelnen Komponenten in \mathbf{s} *statistisch unabhängig* voneinander sind:

Zwei Zufallsvariablen X, Y sind statistisch voneinander unabhängig, wenn die Information über den Wert der einen Variable keine Schlüsse auf den Wert der anderen Variable zulässt.

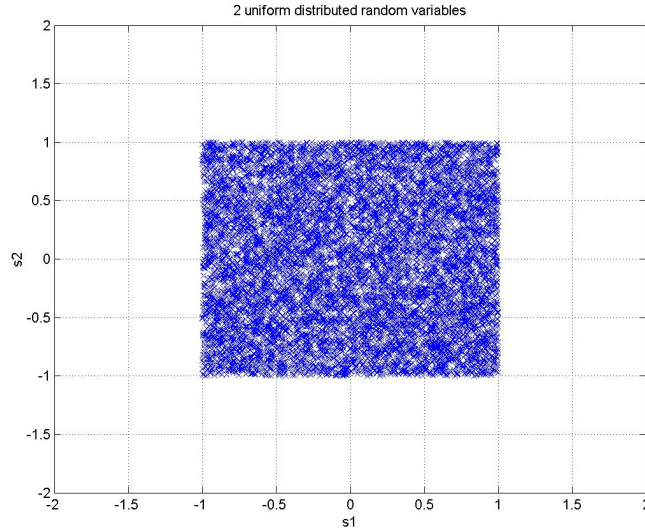


Figure 3: Verteilung von zwei statistisch unabhängigen ZV s_1 und s_2 .

Mathematisch formuliert:

$$f_{XY}(x, y) = f_X(x)f_Y(y). \quad (3)$$

Dabei ist f_{XY} die gemeinsame Wahrscheinlichkeitsdichte zweier Zufallsvariablen X und Y und f_X ist die Wahrscheinlichkeitsdichte einer ZV X.

2.3.2 Beispiel

Bevor näher auf die Lösungsidee eingegangen wird, soll das Prinzip der statistischen Unabhängigkeit zweier unabhängiger Komponenten und die Auswirkung des Mischvorgangs anhand eines Beispiels erläutert werden. Abbildung 3 zeigt Realisierungen eines Zufallsvektors $\mathbf{s} = (s_1, s_2)^T$, wobei s_1 und s_2 statistisch voneinander unabhängig sind und jeweils gleichverteilt sind gemäß:

$$f_{s_i}(s_i) = \begin{cases} 0.5 & |s_i| \leq 1, \\ 0 & \text{sonst.} \end{cases}$$

Da beide Signale statistisch unabhängig voneinander sind, kann allein durch Kenntnis des Wertes einer Realisierung von s_1 in keiner Weise auf den zugehörigen Wert der Realisierung von s_2 geschlossen werden (und umgekehrt). Die zugehörige Kovarianzmatrix \mathbf{C}_s lautet:

$$\mathbf{C}_s = E\{\mathbf{s}\mathbf{s}^T\} = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{pmatrix},$$

wobei $E\{\cdot\}$ der Erwartungswert ist und wir annehmen, dass \mathbf{s} mittelwertfrei ist. Aufgrund der statistischen Unabhängigkeit der beiden Signale s_1 und s_2 ist die Kovarianzmatrix \mathbf{C}_s eine Diagonalmatrix, wobei der Wert in der i-ten Zeile und i-ten Spalte der Varianz des Signals s_i entspricht. Die Kovarianz *zwischen* den einzelnen Signalen ist Null. Die Umkehrung gilt jedoch im Allgemeinen nicht. Zwei Signale sind nicht automatisch statistisch unabhängig, wenn die Kovarianz zwischen diesen Signalen gleich Null ist. Statistische Unabhängigkeit erfordert Trennung der Momente jeder Ordnung, nicht nur der zweiten Ordnung (Varianz: Zentraler Moment zweiter Ordnung).

Nun sei angenommen, dass die Signale s_i gemäß Gleichung (1) gemischt werden und in zwei Sensorsignalen x_1 und x_2 resultieren. Die resultierenden Realisierungen werden in Abbildung 4 dargestellt, die zugehörige Mischmatrix \mathbf{A} lautet:

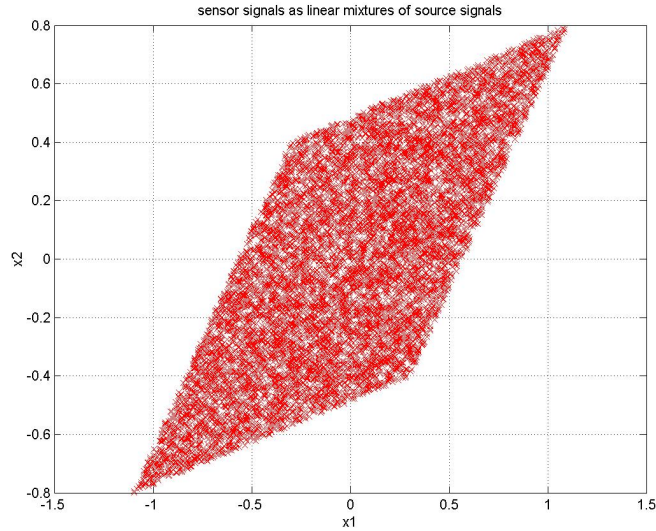


Figure 4: Verteilung der ZV nach dem Mischprozess.

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.7 \\ 0.6 & 0.2 \end{pmatrix}.$$

Wie hier gut beobachtet werden kann, sind die resultierenden Sensorsignale x_i nicht mehr statistisch unabhängig. Beispielsweise kann durch Kenntnis von $x_1 = 1.1$ darauf geschlossen werden, dass der Wert des zugehörigen Samples von x_2 im Bereich $0.62 - 0.8$ liegen *muss*. Die zugehörige Kovarianzmatrix \mathbf{C}_x lautet nun:

$$\mathbf{C}_x = E\{\mathbf{x}\mathbf{x}^T\} = \begin{pmatrix} 0.215 & 0.125 \\ 0.125 & 0.131 \end{pmatrix}.$$

Die Kovarianz zwischen x_1 und x_2 ist somit 0.125, die Signale sind nicht statistisch unabhängig.

2.3.3 Erreichen von statistischer Unabhängigkeit

Das grundlegende Prinzip von ICA ist es nun, die Werte von \mathbf{x} auf eine andere Basis zu projizieren, sodass die aus dieser Projektion resultierenden Signale statistisch unabhängig sind:

$$\mathbf{u} = \mathbf{W}\mathbf{x} \tag{4}$$

Wähle \mathbf{W} so, dass die einzelnen Komponenten in \mathbf{u} paarweise statistisch unabhängig sind.

Dabei sind die Basisvektoren \mathbf{w}_i der Projektion (Basiswechsel) die Reihen der Entmischungsmatrix \mathbf{W} :

$$\mathbf{W} = \begin{pmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_n^T \end{pmatrix}$$

Wie kann nun eine solche Basis gefunden werden? Mit diesem Thema beschäftigen sich die nächsten beiden Sektionen. Sektion 3 beschreibt das Verfahren der Principal Component Analysis, die oft einen wertvollen Vorverarbeitungsschritt für ICA darstellt. In Sektion 4 wird schließlich auf das Verfahren der Independent Component Analysis eingegangen.

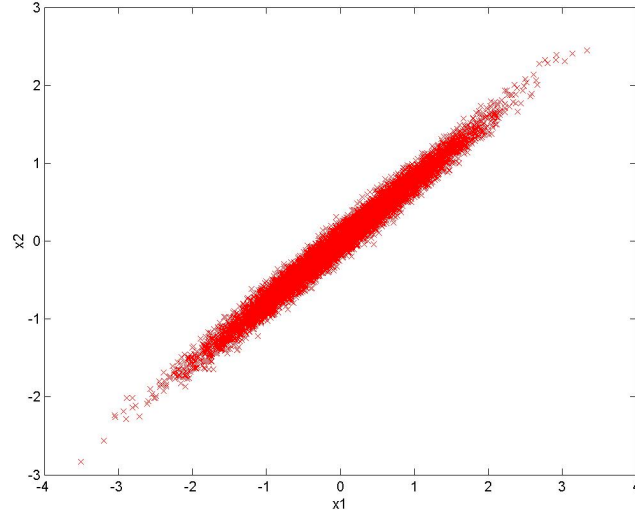


Figure 5: Annahme von Principal Component Analysis: In Richtung der größten Energieanteile steckt am meisten Information.

3 Principal Component Analysis

3.1 Einleitung

Principal Component Analysis ist eine klassische Methode der Statistik, aus der später auch andere Methoden wie Independent Component Analysis hervorgingen. Ziel der PCA ist es, die Daten \mathbf{x} auf eine orthonormale Basis \mathbf{M} zu projizieren, wobei ein Basisvektor in die Richtung der grössten Varianz von \mathbf{x} zeigt. Wie wir sehen werden, hat dies gleichzeitig den Effekt dass die einzelnen Signale eines Signalvektors \mathbf{x} paarweise dekorreliert werden.

In Abbildung 5 sind Realisierungen eines 2-dimensionalen Zufallsvektors zu sehen. Intuitiv könnte angenommen werden, dass in der 45° Richtung am meisten Information vorhanden ist, da in dieser Richtung die Varianz am größten ist. PCA bietet eine mathematische Methode, diese Richtung zu finden die Daten auf diese neue Basis zu projizieren.

3.2 Herleitung

Zunächst suchen wir die Richtung \mathbf{p}_1 , in der die Daten \mathbf{x} die grösste Varianz besitzen. Die Varianz der projizierten Daten ist:

$$E \{ (\mathbf{p}_1^T \mathbf{x})^2 \} = E \{ \mathbf{p}_1^T \mathbf{x} \mathbf{x}^T \mathbf{p}_1 \} = \mathbf{p}_1^T E \{ \mathbf{x} \mathbf{x}^T \} \mathbf{p}_1 = \mathbf{p}_1^T \mathbf{C}_x \mathbf{p}_1$$

Wir wollen also die Varianz bezüglich \mathbf{p}_1 maximieren. Um $\mathbf{p}_1 \rightarrow \infty$ als mögliche Lösung auszuschliessen, führen wir die Nebenbedingung $\mathbf{p}_1^T \mathbf{p}_1 = 1$ ein. Mit Hilfe der Methode der Lagrange-Multiplikatoren erhalten wir das folgende Optimierungsproblem ohne Nebenbedingung:

$$\max_{\mathbf{p}_1} \mathbf{p}_1^T \mathbf{C}_x \mathbf{p}_1 + \lambda(1 - \mathbf{p}_1^T \mathbf{p}_1).$$

Um dieses Optimierungsproblem zu lösen, leiten wir diesen Ausdruck nach \mathbf{p}_1 ab und setzen auf Null:

$$\begin{aligned} \mathbf{C}_x \mathbf{p}_1 - \lambda \mathbf{p}_1 &= 0 \\ \mathbf{C}_x \mathbf{p}_1 &= \lambda \mathbf{p}_1 \end{aligned} \tag{5}$$

Multiplikation von Gleichung (5) mit \mathbf{p}_1^T ergibt $\lambda = \mathbf{p}_1^T \mathbf{C}_x \mathbf{p}_1$, λ ist also die Varianz der Daten in Richtung \mathbf{p}_1 . Die gesuchte Richtung der maximalen Varianz, \mathbf{p}_1 , ist also der zum grössten Eigenwert gehörende Eigenvektor von \mathbf{C}_x . Wir erhalten folglich die gesuchte Basis aus der Eigenwertzerlegung von \mathbf{C}_x :

$$\mathbf{C}_x = E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$$

$\mathbf{\Lambda}$...Diagonalmatrix mit den Eigenwerten von \mathbf{C}_x

\mathbf{P} ...Matrix mit orthonormierten Eigenvektoren von \mathbf{C}_x als Spalten

Es gilt: $\mathbf{P}^{-1} = \mathbf{P}^T$

Durch Normalisierung der Eigenvektoren durch die entsprechenden Eigenwerte erhalten wir schliesslich die gesuchte Basis:

$$\mathbf{M} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{P}^T. \quad (6)$$

Die projizierten Daten $\mathbf{x}_s = \mathbf{M}\mathbf{x}$ sind nun dekorreliert und besitzen Varianz 1:

$$\begin{aligned} E\{\mathbf{x}_s\mathbf{x}_s^T\} &= \mathbf{M}E\{\mathbf{x}\mathbf{x}^T\}\mathbf{M}^T \\ &= \mathbf{M}\mathbf{C}_x\mathbf{M}^T \\ &= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{P}^T\mathbf{C}_x\mathbf{P}\mathbf{\Lambda}^{-\frac{1}{2}} \\ &= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{P}^T\mathbf{P}\mathbf{\Lambda}\mathbf{P}^T\mathbf{P}\mathbf{\Lambda}^{-\frac{1}{2}} \\ &= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Lambda}\mathbf{\Lambda}^{-\frac{1}{2}} \\ &= \mathbf{I}. \end{aligned}$$

3.3 Bedeutung für ICA

Durch Anwendung dieser Transformation auf einen Sensorvektor \mathbf{x} können die einzelnen Komponenten des Vektors paarweise dekorreliert werden. Allerdings bedeutet Dekorrelation nicht automatisch statistische Unabhängigkeit, weshalb PCA das Problem der Quelltrennung nicht lösen kann.

Trotzdem erweist sich PCA in vielen Fällen als nützlicher *Vorverarbeitungsschritt*, denn die Anwendung dieser Transformation auf einen Sensorvektor \mathbf{x} bedeutet eine Modifikation der Mischmatrix des zugrundeliegenden Mischmodells (1):

$$\mathbf{x}_s = \mathbf{M}\mathbf{x} = \mathbf{M}\mathbf{A}\mathbf{s} = \mathbf{B}\mathbf{s} \quad (7)$$

Die *neue* Mischmatrix \mathbf{B} hat die besonders nützliche Eigenschaft der Orthogonalität, das bedeutet die einzelnen Spalten von \mathbf{B} sind linear unabhängig:

$$\begin{aligned} E\{\mathbf{x}_s\mathbf{x}_s^T\} &= E\{\mathbf{B}\mathbf{s}\mathbf{s}^T\mathbf{B}^T\} = \mathbf{B}E\{\mathbf{s}\mathbf{s}^T\}\mathbf{B}^T = \mathbf{B}\mathbf{B}^T \\ &\Rightarrow \mathbf{B}\mathbf{B}^T = \mathbf{I} \end{aligned} \quad (8)$$

Während es sich bei der ursprünglichen Mischmatrix \mathbf{A} um eine *beliebige* Matrix mit insgesamt $n \times m$ unbekannten Parametern handeln konnte, ist die neue Mischmatrix \mathbf{B} *orthogonal*, was geometrisch einer reinen Rotation entspricht.

Dieser Sachverhalt ist in Abbildung 6 und Abbildung 7 dargestellt. Durch Vergleich von Abbildung 7 und Abbildung 3 fällt auf, dass die Mischung geometrisch nur mehr einer Rotation entspricht (abgesehen von der Skalierung der Amplituden, die durch die Normierung der Vektorkomponenten auf Varianz 1 eingeführt wird).

Somit wird durch die Vorverarbeitung des Sensorvektors durch PCA das Problem des Entmischens bis auf eine zusätzliche Rotation gelöst.

4 Independent Component Analysis

4.1 Prinzip der Gaußähnlichkeit

Wie im vorherigen Abschnitt erläutert wurde, kann mit Hilfe von Principal Component Analysis das Problem des Findens einer Entmischungsmatrix \mathbf{W} bis auf eine Rotation gelöst werden. Um auch diese

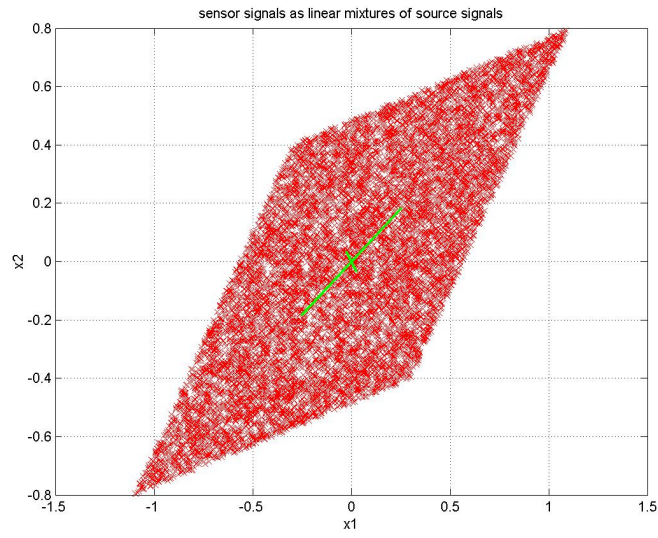


Figure 6: Sensordaten nach Mischung durch Mischmatrix \mathbf{A} ; In grün: Richtungen der größten Energieanteile.

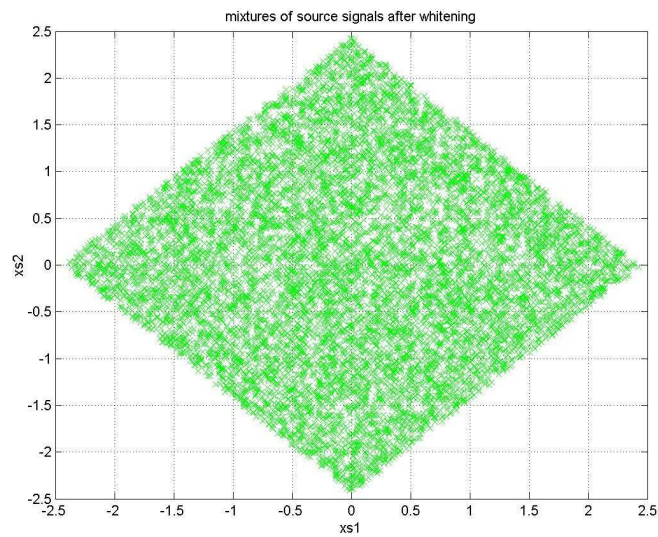


Figure 7: Verteilung der dekorrelierten Sensordaten.

Rotation rückgängig zu machen und somit die unabhängigen Signale s_i zu erhalten, ist es notwendig die inverse Rotationsmatrix zu finden.

Wie schon in Abschnitt 2.3.3 erläutert, basiert das Prinzip von ICA auf dem Auffinden einer neuen Basis (notiert durch eine Matrix \mathbf{W}), sodass die Signalvektorkomponenten nach Projektion auf diese Basis paarweise statistisch unabhängig sind. Die *Qualität* der Projektion wird dabei von einer sog. *Kontrastfunktion* J bewertet. Dabei wird unterschieden zwischen einer

- one-unit contrast function $J(\mathbf{w}_i)$: Diese Funktion bewertet die Qualität eines einzelnen Basisvektors
- multi-unit contrast function $J(\mathbf{W})$: Diese Funktion bewertet die Qualität der gesamten Basis

Mit Hilfe einer geeigneten Kontrastfunktion kann dann mit Hilfe von numerischen Optimierungsverfahren (z.B. Gradient Descent) eine optimale Basis \mathbf{W} gefunden werden.

Nun stellt sich die grundlegende Frage, *wie* die Qualität einer Projektion quantitativ bewertet werden kann. Diese Frage ist gleichbedeutend mit der Frage, wie die statistische Unabhängigkeit der auf die neue Basis projizierten Variablen \mathbf{u} bewertet werden kann. Einen möglichen Ansatz zur Beantwortung dieser Frage liefert der *zentrale Grenzwertsatz* aus der Statistik:

Die Summe von vielen beliebig verteilten Zufallsgrößen ist näherungsweise normalverteilt.

Angenommen, die unabhängigen Quellsignale s_i sind *nicht* gaußverteilt. Dann gilt für die Sensorsignale x_i , dass diese gaußähnlicher sein müssen als die unabhängigen Signale s_i , da jedes x_i eine Summe der einzelnen s_i darstellt. Unter der Voraussetzung, dass die unabhängigen Signale s_i nicht gaußverteilt sind, kann somit statistische Unabhängigkeit durch *Nicht-Gaußähnlichkeit* bewertet werden. Ziel einer Kontrastfunktion J ist es deshalb, die Gaußähnlichkeit eines projizierten Sensorvektors \mathbf{u} und dadurch die Qualität der Projektion \mathbf{W} zu bewerten.

Nun stellt sich im weiteren die Frage, wie die (Nicht-)Gaußähnlichkeit einer Zufallsvariable bestimmt werden kann. Zwei dafür häufig herangezogene Konzepte sind die *Kurtosis* und die *Negentropie*, die in den folgenden Abschnitten beschrieben werden.

4.2 Kurtosis

Die Kurtosis γ_2 einer Zufallsvariable X ist ein statistisches Moment 4. Ordnung und ist definiert als:

$$\gamma_2 = \frac{E\{(X - E\{X\})^4\}}{(E\{X^2\})^2} - 3 \quad (9)$$

Für die Berechnung der Kurtosis eines Sensorsignals reduziert sich dieser Ausdruck auf

$$\gamma_2 = E\{X^4\} - 3 \quad (10)$$

da das Signal mittelwertfrei und auf Varianz 1 normiert ist.

Die Kurtosis einer Zufallsvariable X bewertet die Konzentration der zugehörigen Wahrscheinlichkeitsdichte (Peak der Wahrscheinlichkeitsdichte) um den Mittelwert von X . Dieser Zusammenhang ist in Abbildung 8 dargestellt. Die Gaußverteilung (rot) hat per Definition die Kurtosis Null. Eine pdf mit einer höheren Konzentration der Dichte um den Mittelwert - wie beispielsweise die Laplaceverteilung (grün) - besitzt eine positive Kurtosis. Eine pdf mit einer gleichmäßiger verteilten Dichte - wie beispielsweise die Gleichverteilung (blau) - besitzt eine negative Kurtosis. In diesem Zusammenhang wird eine pdf mit positiver Kurtosis auch als *super-gauß* und eine pdf mit negativer Kurtosis auch als *sub-gauß* bezeichnet. In den Anwendungen der ICA wird oft die quadratische Kurtosis als Kontrastfunktion definiert:

$$J(\mathbf{w}_i) = (E\{(\mathbf{w}_i^T \mathbf{x}_i)^4\} - 3)^2 \quad (11)$$

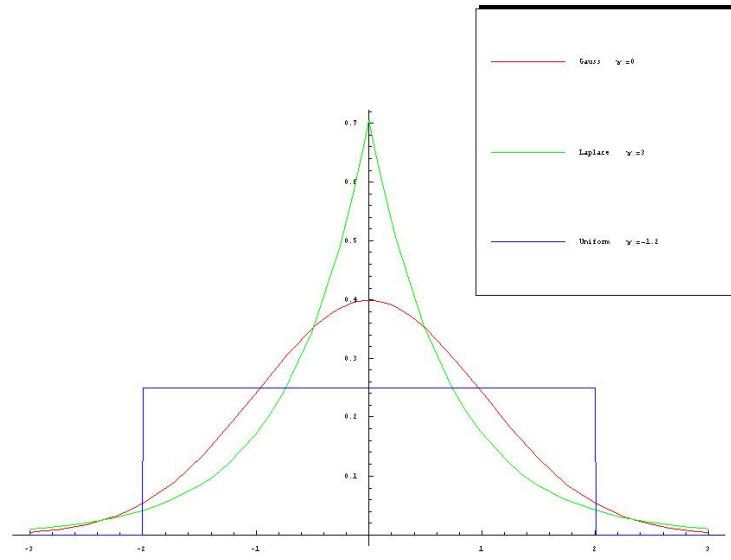


Figure 8: Kurtosis von Gleichverteilung, Normalverteilung und Laplaceverteilung.

Diese ist leicht berechenbar und stets positiv. Je gaußähnlicher ein Zufallsvektor ist, desto kleiner wird die zugehörige quadratische Kurtosis.

Ein Nachteil der Kurtosis ist allerdings deren Anfälligkeit gegenüber *Ausreißern*: Bei der Schätzung der Kurtosis anhand nur weniger Realisierungen von X kann es passieren, dass die Häufigkeitsverteilung dieser Realisierungen an manchen Stellen nicht der tatsächlichen pdf entsprechen. Gerade für solche Realisierungen mit hohem Betrag wird dann die Schätzung der Kurtosis völlig falsch.

Aus diesem Grund ist es wünschenswert, *robustere* Verfahren zur Schätzung der Gaußähnlichkeit herzuleiten. Eines dieser robusten Verfahren ist die Negentropie.

4.3 Negentropie

Die Negentropie leitet sich aus dem Konzept der Entropie ab. Die Entropie H eines Signals ist ein Maß für dessen Informationsgehalt. Beispielsweise trägt ein über die Zeit stets konstantes Signal offensichtlich überhaupt keine Information, die Entropie dieses Signals ist dementsprechend Null. Je weniger vorhersagbar bzw. je überraschender ein zukünftiges Sample eines Signals andererseits ist, desto mehr Information *kann* in diesem Signal stecken. Daraus wird klar, dass die Entropie eines Signals eng mit dessen zugrundeliegenden Wahrscheinlichkeitsverteilung verknüpft ist. Die Entropie H einer Zufallsvariable X ist definiert als:

$$H(X) = - \int f_X(\eta) \log(f_X(\eta)) d\eta \quad (12)$$

Dabei ist f_X die X zugrundeliegende Wahrscheinlichkeitsdichte.

Auf der Suche nach einer geeigneten Bewertung der Gaußähnlichkeit kommt nun ein Theorem aus der Informationstheorie zu Hilfe:

In der Menge aller möglichen Zufallsvariablen mit gleicher Varianz hat die gaußverteilte Zufallsvariable die größte Entropie.

Aus diesem Theorem kann nun das Konzept der Negentropie abgeleitet werden. Die Negentropie J einer Zufallsvariable X ist definiert als:

$$J(X) = H(X_{gauss}) - H(X) \quad (13)$$

Dabei ist X_{gauss} eine gaußverteilte Zufallsvariable mit gleicher Varianz wie X . Auf diese Weise wird $J(X)$ dann und nur dann gleich Null, wenn X gaußverteilt ist. Je weniger gauß-ähnlich X ist, desto größer wird $J(X)$.

Aus Sicht der Statistik ist die Negentropie ein optimaler Schätzoperator für die Gaußähnlichkeit einer Zufallsvariable. Ein großer Nachteil der Negentropie ist allerdings, dass zur Berechnung die pdf von X bekannt oder geschätzt werden muss, was sich als aufwendig und rechenlastig herausstellt.

Aus diesem Grund wird versucht, die Negentropie mit weniger rechenlastigen Funktionen zu approximieren. Für solche Funktionen hat sich folgende Struktur als nützlich herausgestellt:

$$J(X) = k_1 (E\{G_1(X)\})^2 + k_2 (E\{G_2(X)\} - E\{G_2(\nu)\})^2 \quad (14)$$

Dabei ist ν die standardisierte Gaußvariable und G_1 bzw. G_2 sind nichtlineare Funktionen, die so gewählt werden müssen dass die Schätzung robust ist. Eine beliebte Wahl für G_1 und G_2 ist beispielsweise:

$$G_1(x) = \frac{1}{a_1} \log(\cosh(a_1 x)) \quad (15)$$

$$G_2(x) = -e^{-\frac{x^2}{2}} \quad (16)$$

Insgesamt stellt die Approximation der Negentropie einen guten Kompromiss zwischen den beiden Varianten Kurtosis und Negentropie dar, da sie einerseits effizient berechenbar und andererseits robust ist.

4.4 Zusammenfassung

Independent Component Analysis basiert auf der Annahme, dass

- die einzelnen Quellen s_i *statistisch unabhängig* sind
- nur maximal eine Quellen s_i gaußverteilt ist (alle anderen Quellen müssen *nicht gaußverteilt* sein).
- die Mischmatrix A muss vollen Rang besitzen.

Sind diese Annahmen erfüllt, kann der Mischprozess gemäß Gleichung (1) durch den Entmischprozess gemäß Gleichung (2) rückgängig gemacht werden.

Ein großer Nachteil von ICA besteht in der Tatsache, dass die Anzahl der Sensoren M größer oder mindestens gleich der Anzahl der unabhängigen Quellen N sein muss. Dies ist in der Praxis besonders dann ein Problem, wenn im Vorhinein nicht abgeschätzt werden kann, wie viele unabhängige Quellen auftreten werden bzw. wenn aufgrund der zu hohen Anzahl an Quellen das Setup nicht mehr durchführbar ist.

References

- [1] Albert S. Bregman. *Auditory Scene Analysis*. MIT Press, 1990.
- [2] Carsten Henning. ICA-Kurs. Technical report, Technische Universität Berlin, June 1999.
- [3] Aapo Hyvärinen. Survey on Independent Component Analysis. *Neural Computing Surveys*, 1999.
- [4] Aapo Hyvärinen and Erkki Oja. Independent Component Analysis: A Tutorial. 1999.
- [5] Aapo Hyvärinen and Erkki Oja. Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 2000.