



Herta Reich Gymnasium und Realgymnasium  
Mürzzuschlag; 613016  
8680 Mürzzuschlag, Roseggergasse 10

# Principal Component Analysis und Independent Component Analysis

Eine Einführung in die Standardalgorithmen  
zweier Verfahren zur Signalanalyse und  
mögliche Verbesserungen des ICA-Algorithmus

Fachbereichsarbeit aus Mathematik

Vorgelegt bei OStR. Prof. Mag. Ursula Brünner

von Georg Andreas Maierhofer, 8.B

Mürzzuschlag, abgegeben am 01.03.2013

# Inhaltsverzeichnis

<b>Abstract</b>	<b>4</b>
<b>1 Vorwort</b>	<b>5</b>
<b>2 Einleitung</b>	<b>7</b>
2.1 Notation und wichtige Begriffe . . . . .	10
2.2 Das Modell . . . . .	11
<b>3 Einführung in das Verfahren Principal Component Analysis (PCA)</b>	<b>13</b>
3.1 Motivation . . . . .	13
3.2 Das Verfahren in der Praxis . . . . .	14
3.2.1 Das Prinzip der Korrelation . . . . .	14
3.2.2 Zentrieren . . . . .	16
3.2.3 Bedingungen zur Bestimmung der Hauptkomponenten . . . . .	17
3.2.4 Die Kovarianzmatrix . . . . .	19
3.2.5 Eigenräume . . . . .	20
3.2.6 Bestimmung der Hauptkomponenten . . . . .	21
3.2.7 Reduktion der Datendimension - Darstellung durch wenige Hauptkomponenten . . . . .	23
<b>4 Einführung in das Verfahren Independent Component Analysis (ICA)</b>	<b>25</b>
4.1 Motivation . . . . .	25
4.2 Das Verfahren in der Praxis . . . . .	26
4.2.1 Vergleich Unkorreliertheit und Unabhängigkeit . . . . .	26
4.2.2 Zentrieren . . . . .	29
4.2.3 Bedingungen zur Bestimmung der unabhängigen Komponenten . . . . .	29
4.2.4 Die Bedingung der Unabhängigkeit und der Zentrale Grenzwertsatz . . . . .	31
4.2.5 Messung der Gaußianität . . . . .	33

4.2.6	Minimierung der Zielfunktion $J_1$ durch Gradienten-Abstiegsverfahren	35
4.2.7	Algorithmus zur Bestimmung der unabhängigen Komponenten . . . . .	36
4.2.8	Hervorheben von Komponenten - Darstellung durch weniger unabhängige Komponenten . . . . .	38
<b>5</b>	<b>Verbesserung des Standard-Algorithmus von Independent Component Analysis (ICA)</b>	<b>40</b>
5.1	Motivation . . . . .	40
5.2	Die Messung der Gaußianität . . . . .	41
5.2.1	Schwachpunkt der Wölbung . . . . .	41
5.2.2	Verbesserte Zielfunktion . . . . .	45
5.3	Minimierung der Zielfunktion . . . . .	48
5.3.1	Schwachpunkte des Gradienten-Abstiegsverfahrens . . . . .	48
5.3.2	Verbesserte Minimierung . . . . .	48
5.4	Verbesserter Algorithmus zur Bestimmung der unabhängigen Komponenten .	51
<b>6</b>	<b>Visualisierung der Verfahren</b>	<b>55</b>
6.1	Motivation . . . . .	55
6.2	Das Modell in der Anwendung . . . . .	56
6.3	Anwendung von PCA . . . . .	56
6.4	Anwendung des Standard-Algorithmus von ICA . . . . .	57
6.5	Anwendung des verbesserten Algorithmus von ICA . . . . .	58
<b>7</b>	<b>Zusammenfassung</b>	<b>60</b>
	<b>Abbildungsverzeichnis</b>	<b>61</b>
	<b>Literaturverzeichnis</b>	<b>64</b>
	<b>Beilagen</b>	<b>65</b>
	<b>Eidesstattliche Erklärung</b>	<b>66</b>

## **Abstract**

Diese Fachbereichsarbeit setzt sich zum Ziel, die beiden Verfahren Principal Component Analysis und Independent Component Analysis sowohl aus theoretischer, als auch aus praktischer Sicht vorzustellen. Ihr grundlegendes Prinzip wird im Kapitel 2 besprochen und anhand von zwei Beispielen gezeigt. In Kapitel 3 und 4 werden die mathematischen Grundlagen und die Vorgehensweise in den Verfahren Principal Component Analysis und Independent Component Analysis gezeigt. Dabei wird vor allem auf die dahinterliegenden Prinzipien - Korrelation und statistische Unabhängigkeit - eingegangen. In Kapitel 5 werden Verbesserungsmöglichkeiten des Standard-Algorithmus von Independent Component Analysis dargelegt, welche zu einem der momentan aktuellsten Algorithmen aus der Forschungsarbeit von a.o. Univ.-Prof. Mag. Dr. Stephen Keeling führen. Abschließend wird in Kapitel 6 ein Beispiel für die praktische Anwendbarkeit der Verfahren dargestellt. Dieses Kapitel wird durch eine DVD ergänzt, auf welcher dieses Beispiel, sowie zwei weitere, detaillierter gezeigt werden.

## **Abstract**

The aim of this subject-specific thesis is to give a theoretical as well as practical insight into the methods of Principal Component Analysis and Independent Component Analysis. Their fundamental principle is explained in chapter 2 - underlined with two examples. In chapter 3 and 4 the mathematical basics and the approach of Principal Component Analysis and Independent Component Analysis are presented, with particular focus on the statistical theory behind the methods - correlation and statistical independence. In chapter 5 improvements of the standard algorithm of Independent Component Analysis are shown, which lead to one of the most current algorithms from the research of a.o. Univ.-Prof. Mag. Dr. Stephen Keeling. In the concluding chapter 6 an example for a realistic application of the methods is displayed. In addition to this chapter a DVD is added, on which one can find this example and two more explained in more detail.

# 1 Vorwort

Als ich zum ersten Mal von dem Themenbereich hörte, welcher in dieser Fachbereichsarbeit behandelt wird, wusste ich sehr wenig über die Vielseitigkeit von praktischen Anwendungen der Mathematik. Ich hatte durch die Teilnahme an der Österreichischen Mathematikolympiade schon etwas Erfahrung mit weiterführender Mathematik, welche über den normalen Schulunterrichtsstoff hinausreicht, gemacht, aber keine wirkliche Vorstellung davon, welchen praktischen Nutzen man aus komplexeren mathematischen Zusammenhängen ziehen kann. Als ich dann in der von Herrn a.o. Univ.-Prof. Mag. Dr. Keeling organisierten Modellierungswoche von gleich fünf komplett unterschiedlichen Anwendungen höherer Mathematik hörte - angefangen bei Spielstärkeeinschätzung und Ergebnisvorhersage bei Paarvergleichen bis hin zu mathematischer Signalverarbeitung - war meine Begeisterung und mein Interesse geweckt. Aufgeteilt in Kleingruppen wurden diese Themenbereiche bearbeitet. Dabei konnte ich in der Gruppe von Herrn a.o. Univ.-Prof. Mag. Dr. Keeling, unter seiner Anleitung, die beiden Verfahren Principal Component Analysis und Independent Component Analysis näher kennenlernen.

Als nach der Woche für mich noch einige Fragen offen waren, blieb er mit mir in Kontakt und half mir dabei die Problemstellung selbstständig zu erarbeiten. Er nahm sich für mich Zeit, um mir noch offene Fragen zu beantworten, und hielt mich gleichzeitig über die aktuelle Forschung am Laufenden, welche er an der Karl-Franzens-Universität Graz zur Verbesserung des ICA Algorithmus betreibt.

Nachdem ich mich eine Zeitlang mit dem Thema auseinandergesetzt hatte, wollte ich noch mehr darüber wissen und noch mehr davon verstehen. Da für mich schon lange feststand, dass ich eine Fachbereichsarbeit schreiben wollte, bot sich mir dadurch die einmalige Gelegenheit, mich noch intensiver mit dem Thema zu beschäftigen.

In der Zeit nach meinem Entschluss erhielt ich sowohl von Frau OStR. Prof. Mag. Brünner, als auch von Herrn a.o. Univ.-Prof. Mag. Dr. Keeling tolle Unterstützung, was die Planung der Möglichkeiten, sowie des Inhaltes der FBA betrifft. Herr a.o. Univ.-Prof. Mag. Dr. Kee-

ling stand mir weiterhin für alle Fragen geduldig zur Verfügung, hat die Arbeit für mich korrigiert und hält mich bis heute mit seiner Forschungsarbeit auf dem aktuellsten Stand des mathematischen Wissens in diesem Bereich.

Aus diesem Grund möchte ich mich ganz besonders bei ihm für seine Unterstützung, Geduld und Mühe bedanken - er hat mir eine Möglichkeit geboten, welche für einen Schüler etwas wirklich Besonderes ist. Auch möchte ich mich bei Frau OStR. Prof. Mag. Brünner für ihre Offenheit meine Fachbereichsarbeit zu betreuen, ihre beharrliche Anstrengung, die Informationen für die Planung der Fachbereichsarbeit einzuholen und die Zeit, welche sie ohne zu zögern in den Kontakt und die Koordination mit Herrn a.o. Univ.-Prof. Mag. Dr. Keeling investiert hat, bedanken.

## 2 Einleitung

**Principal Component Analysis (PCA) - Hauptkomponentenanalyse** und **Independent Component Analysis (ICA) - Unabhängigkeitsanalyse** sind zwei statistische Techniken, um Signalen unterliegende versteckte Faktoren ausfindig zu machen. Principal Component Analysis wurde Anfang des 20. Jahrhunderts entwickelt<sup>1</sup>, während Independent Component Analysis erst im Jahr 1982 eingeführt wurde<sup>2</sup>. Der ursprüngliche Auslöser für das Entwickeln von ICA war eine Problemstellung der Neurophysiologie<sup>3</sup>. Um dem Leser ein leichteres Verständnis der grundlegenden Problemstellung zu ermöglichen, folgen zwei einleitende Beispiele.

### Das Cocktail-Party-Problem

Man stelle sich folgende Situation vor: Während einer Cocktail-Party befinden sich  $m$  Menschen in einem Gespräch untereinander. Aus bestimmten Gründen möchte eine weitere Person  $P$  das Gesprochene jedes Einzelnen der  $m$  Personen aufzeichnen. Dabei stellt sich allerdings das Problem, dass die Personen alle zur selben Zeit sprechen.  $P$  hat nun genau  $m$  Mikrophone zur Verfügung und platziert diese im Raum. Später betrachtet er die Aufzeichnungen, aber muss dabei feststellen, dass auf allen Aufnahmen alle  $m$  Personen zu hören sind, somit wurden lediglich Mischungen der ursprünglichen Stimmen aufgezeichnet.  $P$  lässt sich dadurch allerdings nicht entmutigen und sucht nach einer Möglichkeit, aus diesen akustischen Mischungen die ursprünglichen Stimmen - die akustischen Quellen - zu rekonstruieren.

---

<sup>1</sup>vgl. Jolliffe, Ian T.: Principal Component Analysis. 2. Aufl., New York, 2002 , S. ix

<sup>2</sup>vgl. Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis. 1. Aufl., New York u.a., 2001 , S. 11

<sup>3</sup>vgl. ebda, S.11

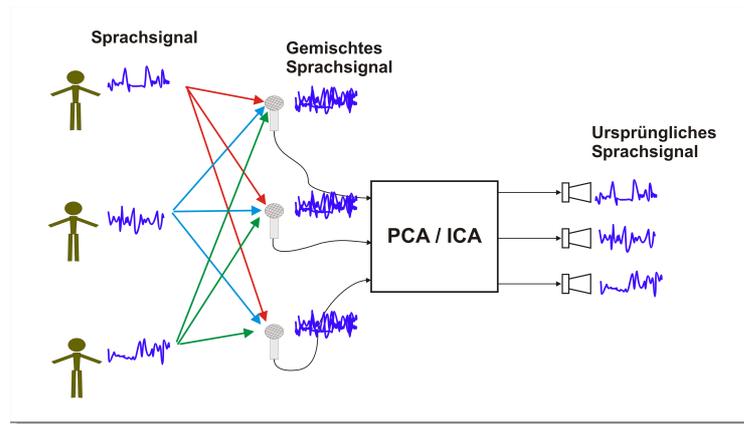


Abb. 2.1: Cocktail-Party-Problem

Seine Suche ist nicht erfolglos: Die beiden Verfahren PCA und ICA bieten die Möglichkeit, die Stimmen nahezu fehlerlos wieder zu rekonstruieren, dies wird in der Abbildung 2.1 schematisch veranschaulicht.

Im Gegensatz zum Cocktail-Party-Problem gibt es viele reale Einsatzgebiete der beiden Verfahren in der Signalanalyse, dazu gehört auch folgendes Beispiel:

### **Inverse Probleme in der Bildverarbeitung**

Dieses Problem bezieht sich auf die Verarbeitung von Bild-Signalen.

Als Beispiel dazu kann ein Kontrastvideo (hier ein Video aus Magnetresonanztomographie-Aufnahmen) aus der Medizin dienen. Dabei stellt sich in diesem Beispiel für den Arzt das Problem, dass er einen Tumor im Kopf des Patienten näher diagnostizieren muss. Bei dieser Untersuchung wird dem Patienten ein Kontrastmittel verabreicht, welches sich durch die Zellen des Gehirnes bewegt und diese somit erkennbar werden lässt. Das Kontrastmittel kann allerdings in den größeren Blutgefäßen stärker als erwünscht erscheinen. Dadurch wird die Beurteilung des Tumors stark eingeschränkt. Die Unterscheidung zwischen den Blutgefäßen mit Kontrastmittel und dem kranken Gewebe kann nicht exakt getroffen werden. Mithilfe von PCA und ICA kann man die versteckten Komponenten des Videos separieren und gezielt die Gefäße ausblenden, um ein besseres Sichtbarwerden des Tumors zu ermöglichen.

Während in der Abbildung 2.2 der Tumor im linken oberen Bereich des Gehirnes nur schwach bis gar nicht sichtbar ist, kann dieser im bearbeiteten Video in der Abbildung 2.3 deutlicher erkannt werden.

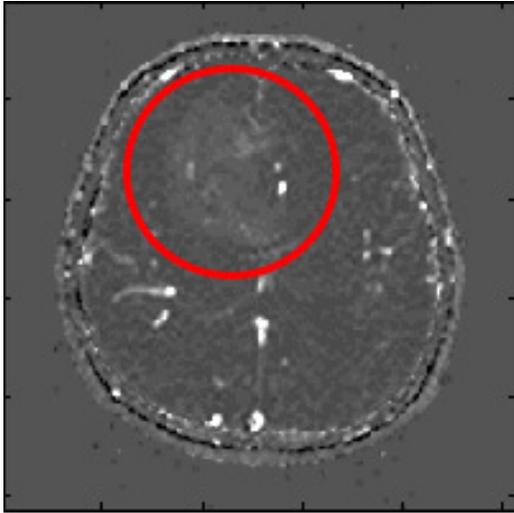


Abb. 2.2: MRT-Aufnahme von Tumor unbearbeitet<sup>4</sup>

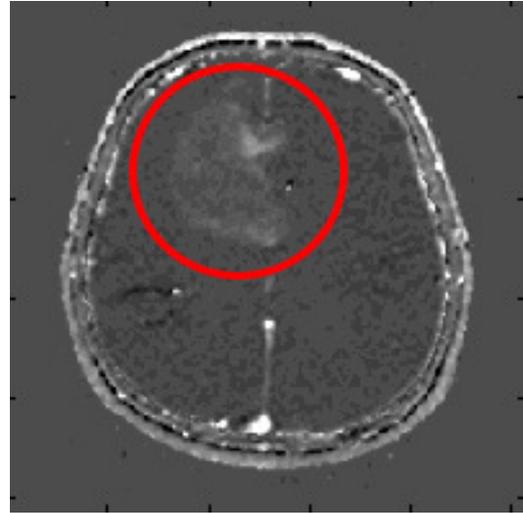


Abb. 2.3: MRT-Aufnahme von Tumor mit PCA und ICA bearbeitet<sup>4</sup>

---

<sup>4</sup>Medizinische Aufnahme von Herrn a.o. Univ.-Prof. Mag. Dr. Keeling erhalten, Bild daraus selbst erstellt

## 2.1 Notation und wichtige Begriffe

### Notation

Prinzipiell wird in der Arbeit die allgemein übliche Notation der Mathematik verwendet. Lediglich einzelne Ausnahmen wurden in der Schreibweise verändert oder neu hinzugefügt, um die Übersichtlichkeit der mathematischen Ausdrücke zu gewährleisten.

- Fettgedruckte Kleinbuchstaben (wie zum Beispiel  $\mathbf{x}$ ) bezeichnen Vektoren.
- Großbuchstaben (wie zum Beispiel  $X$ ) bezeichnen Matrizen.
- In Klammern geschriebene Zahlen direkt nach Vektoren oder Matrizen (wie zum Beispiel  $\mathbf{x}(i)$  oder  $X(i)$ ) bezeichnen die mit der Zahl angegebene Zeile des Vektors oder der Matrix. So wäre zum Beispiel  $\mathbf{x}(i)$  ein Skalar und  $X(i)$  ein Zeilenvektor.
- Die Einheitsmatrix wird mit dem Buchstaben  $I$  bezeichnet. Deren Dimensionen sind aus dem jeweiligen Zusammenhang ablesbar.
- Der Ausdruck  $\text{diag}\{a_1, a_2, \dots, a_n\}$  bezeichnet eine diagonale Matrix, an dessen Diagonale die Werte  $a_1, a_2, \dots, a_n$  stehen.
- Der Ausdruck  $e_{a,i}$  bezeichnet einen Vektor in  $a$  Dimensionen, dessen Einträge alle 0 sind, ausgenommen dem  $i$ -ten, dieser Eintrag ist 1.
- Ausdrücke mit hochgestelltem "T" (wie zum Beispiel  $X^T$ ) bezeichnen die transponierte Matrix des entsprechenden Ausdruckes.
- Fettgedruckte Kleinbuchstaben mit Querbalken (wie zum Beispiel  $\bar{\mathbf{x}}$ ) bezeichnen den Mittelwert der Einträge des entsprechenden Vektors  $\mathbf{x}$ .
- Großbuchstaben mit Querbalken (wie zum Beispiel  $\bar{X}$ ) bezeichnen einen Spaltenvektor, dessen Einträge die Mittelwerte der Einträge in den entsprechenden Zeilenvektoren der Matrix  $X$  sind.
- $P(A)$  bezeichnet die relative Häufigkeit mit der ein Ereignis  $A$  eintritt.
- Wird geschriebenermaßen einem Spalten- oder Zeilenvektor ein Skalar addiert (wie zum Beispiel  $\mathbf{x}+a$ ) so wird das Skalar tatsächlich zu jedem Eintrag des Vektors addiert.

## Wichtige Begriffe

- **Datenvektor:** Ein Datenvektor beinhaltet die Ergebniswerte einer bestimmten Messung. Ein Beispiel dafür ist eine akustische Tonspur, welche die Ergebniswerte der in Abhängigkeit der Zeit gemessenen Tonhöhe beinhaltet.
- **Datensatz:** Ein Datensatz ist eine aus mehreren Datenvektoren bestehende Menge, welche mehrere verschiedene Messungen des ähnlichen Experimentes beinhaltet. Alle Datenvektoren in einem Datensatz besitzen die selbe Dimension. Ein Beispiel dazu ist der Datensatz einer Tonaufzeichnung: Es besteht aus mehreren Tonspuren (= Datenvektoren), welche alle die Messung eines ähnliches Ereignisses - die Tonhöhe in Abhängigkeit der Zeit - sind.
- **Ausreißer:** Ein Ausreißer ist ein Wert eines Datenvektors, welcher ungewöhnlich stark von dessen Mittelwert abweicht. Dies kann in der Realität zum Beispiel eine Störung bei einer Tonaufzeichnung sein.
- **Streudiagramm:** Ein Streudiagramm bietet eine sehr gute grafische Veranschauligungsmöglichkeit von mehrdimensionalen Datensätzen. Besteht der Datensatz aus den beiden Datenvektoren  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ , so werden  $n$  Punkte in ein ebenes Koordinatensystem so aufgetragen:

Die Koordinaten des  $i$ -ten Punktes  $P_i$  mit  $1 \leq i \leq n$  sind die Werte der beiden Vektoren in der  $i$ -ten Zeile, also hat  $P_i$  die Form:

$$P_i = \begin{pmatrix} \mathbf{x}_1(i) \\ \mathbf{x}_2(i) \end{pmatrix}$$

Das Prinzip lässt sich an sich auch für mehr als zwei Datenvektoren verallgemeinern, wird aber hier nur in 2-Dimensionen als Diagramm verwendet.

In der Arbeit sind alle abgebildeten Diagramme, falls nicht anderes bemerkt wird, Streudiagramme.

## 2.2 Das Modell

Um eine Problemstellung mit mathematischen Mitteln bearbeiten zu können, wird zunächst ein Modell des Ausgangspunktes für das Bearbeiten benötigt. Im Falle der Problemstellung in dieser Arbeit - die Verarbeitung von Signalen mittels der beiden Verfahren PCA und

ICA - ist somit ein geeignetes Modell des Ausgangs-Datensatzes beziehungsweise der zu verarbeitenden Signale von Nöten. Es ist sowohl aus mathematischer, als auch aus anwendungsbezogener Sicht sinnvoll, für die ursprünglichen Daten das folgende Modell anzunehmen:

**Modell 1.** Sei  $X \in \mathbb{R}^{m \times n}$  der ursprüngliche Datensatz mit  $m, n \in \mathbb{N}$  und  $n > m \geq 2$ , welcher die  $m$  Datenvektoren  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \in \mathbb{R}^n$  in der Form

$$X = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_m^T \end{pmatrix}$$

beinhaltet. Die Datenvektoren  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  sind linear unabhängig und Abtastungen von  $m$  Zufallsvariablen, welche Messungen eines ähnlichen Experimentes sind.

Dabei ist lineare Unabhängigkeit wie folgt definiert:

**Definition 2.2.1.** Für Vektoren  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \in \mathbb{R}^n$  mit  $n > m \geq 2$  gilt: Besitzt die Gleichung

$$\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \dots + \alpha_m \mathbf{x}_m = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

nur die Lösung  $\alpha_1 = \alpha_2 = \dots = \alpha_m = 0$ , so heißen die Vektoren  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  linear unabhängig.

(Definitions-Quelle: <sup>5</sup>)

---

<sup>5</sup>vgl. Leydold, Josef: Lineare Unabhängigkeit. Online in WWW unter URL:

<http://statistik.wu-wien.ac.at/~leydold/MOK/HTML/node25.html>. [Stand 28.02.2013]

# 3 Einführung in das Verfahren Principal Component Analysis (PCA)

## 3.1 Motivation

Die grundlegende Idee der **Hauptkomponentenanalyse (Principal Component Analysis)** ist es die Datendimension eines Datensatzes zu reduzieren, in welchen sich eine große Anzahl von voneinander abhängigen Variablen befindet, während zur gleichen Zeit so viel Information wie möglich über die Struktur erhalten bleiben soll<sup>1</sup>.

Um diese Reduktion zu erreichen sucht man nach den sogenannten Hauptkomponenten (Principal Components) des Datensatzes. Es ist sinnvoll nach Komponenten mit möglichst wenig Zusammenhang zu suchen, da solche die wichtigsten Eigenschaften des Datensatzes einzeln beinhalten. In der Hauptkomponentenanalyse wird als Maß der Zusammenhänge die Korrelation verwendet. Die Hauptkomponenten sind praktisch die dahinterliegenden Faktoren, welche unkorreliert als Grundlage für die Zusammensetzung der Daten dienen. Diese Hauptkomponenten sind nach ihren jeweiligen Varianzen sortiert. Deswegen enthalten üblicherweise die ersten paar Komponenten bereits den Großteil der Information aller ursprünglichen Variablen<sup>2</sup>.

Durch eine Darstellung der Daten durch die lediglich ersten paar Hauptkomponenten lässt sich die Dimension des Datensatzes stark vermindern, ohne dabei viel der ursprünglichen Information zu verlieren.

In den folgenden Darstellungen entspricht Abbildung 3.1 einem Datensatz und Abbildung 3.2 eine Darstellung desselben durch lediglich die stärkste Hauptkomponente. Man erkennt, dass obwohl lediglich ca. die Hälfte der Informationen für die Darstellung benutzt wurde, diese dennoch den wichtigsten Aspekt des Datensatzes hinreichend gut wiedergibt.

---

<sup>1</sup>vgl. Jolliffe, Ian T.: Principal Component Analysis , S. ix

<sup>2</sup>vgl. ebda, S.ix

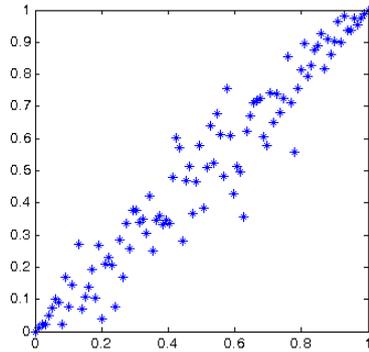


Abb. 3.1: Beispiel eines zweidimensionalen Datensatzes

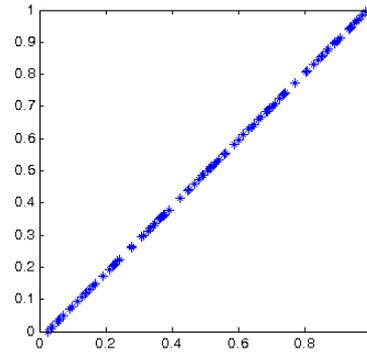


Abb. 3.2: Darstellung durch die stärkste Hauptkomponente

## 3.2 Das Verfahren in der Praxis

### 3.2.1 Das Prinzip der Korrelation

**Definition 3.2.1.** Seien die beiden Datenvektoren  $\mathbf{x}_1 \in \mathbb{R}^n$  und  $\mathbf{x}_2 \in \mathbb{R}^n$  gegeben, so ist die Korrelation der beiden definiert als

$$\chi(\mathbf{x}_1, \mathbf{x}_2) = \frac{\kappa(\mathbf{x}_1, \mathbf{x}_2)}{\sigma(\mathbf{x}_1)\sigma(\mathbf{x}_2)}.$$

Wobei  $\kappa(\mathbf{x}_1, \mathbf{x}_2)$  als Kovarianz der beiden gegeben ist durch

$$\kappa(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{n} (\mathbf{x}_1 - \bar{\mathbf{x}}_1)^T (\mathbf{x}_2 - \bar{\mathbf{x}}_2)$$

und die Standardabweichung  $\sigma(\mathbf{x})$  gegeben ist durch

$$\sigma(\mathbf{x}) = \sqrt{\kappa(\mathbf{x}, \mathbf{x})}.$$

(Definitions-Quelle: <sup>3</sup>)

Nach dieser Definition gilt offensichtlich

$$\chi(\mathbf{x}_1, \mathbf{x}_2) = 0 \iff \kappa(\mathbf{x}_1, \mathbf{x}_2) = 0. \quad (3.1)$$

Für die Korrelation gilt weiters immer:

$$|\chi(\mathbf{x}_1, \mathbf{x}_2)| \leq 1.$$

<sup>3</sup>vgl. Arens, Tilo/ Hettlich, Frank/ Karpfinger, Christian/ et al: Mathematik. 2. Aufl., Heidelberg, 2012 , S. 1255

*Beweis.* Setzt man die Definition der Kovarianz und der Standardabweichung in jene der Korrelation in der Definition 3.2.1 ein, so erhält man:

$$\chi(\mathbf{x}_1, \mathbf{x}_2) = \frac{\frac{1}{n}(\mathbf{x}_1 - \bar{\mathbf{x}}_1)^T(\mathbf{x}_2 - \bar{\mathbf{x}}_2)}{\sqrt{\frac{1}{n}(\mathbf{x}_1 - \bar{\mathbf{x}}_1)^T(\mathbf{x}_1 - \bar{\mathbf{x}}_1)}\sqrt{\frac{1}{n}(\mathbf{x}_2 - \bar{\mathbf{x}}_2)^T(\mathbf{x}_2 - \bar{\mathbf{x}}_2)}}.$$

Substituiert man

$$\mathbf{a} = \mathbf{x}_1 - \bar{\mathbf{x}}_1$$

$$\mathbf{b} = \mathbf{x}_2 - \bar{\mathbf{x}}_2$$

so ist die Korrelation nach Vereinfachung

$$\chi(\mathbf{x}_1, \mathbf{x}_2) = \frac{\mathbf{a}^T \mathbf{b}}{\sqrt{\mathbf{a}^T \mathbf{a}} \sqrt{\mathbf{b}^T \mathbf{b}}}.$$

Es ist nun  $\mathbf{a}^T \mathbf{b}$  das skalare Produkt  $\langle \mathbf{a}, \mathbf{b} \rangle$  der beiden Vektoren und  $\sqrt{\mathbf{a}^T \mathbf{a}}$  bzw.  $\sqrt{\mathbf{b}^T \mathbf{b}}$  ist der Betrag  $\|\mathbf{a}\|$  bzw.  $\|\mathbf{b}\|$  des entsprechenden Vektors. Damit lässt sich der absolute Betrag der Korrelation anders schreiben als:

$$\begin{aligned} |\chi(\mathbf{x}_1, \mathbf{x}_2)| &= \left| \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\| \|\mathbf{b}\|} \right| \\ &= \frac{|\langle \mathbf{a}, \mathbf{b} \rangle|}{\|\mathbf{a}\| \|\mathbf{b}\|}. \end{aligned} \tag{3.2}$$

Nach der Cauchy-Schwarzschen Ungleichung gilt nun:

**Satz 3.2.1.** Für beliebige zwei Vektoren  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  gilt:

$$|\langle \mathbf{a}, \mathbf{b} \rangle| \leq \|\mathbf{a}\| \|\mathbf{b}\|.$$

(Satz-Quelle: <sup>4</sup>)

Satz 3.2.1 ist äquivalent zu

$$\begin{aligned} \frac{|\langle \mathbf{a}, \mathbf{b} \rangle|}{\|\mathbf{a}\| \|\mathbf{b}\|} &\leq 1 \\ \stackrel{(3.2)}{\iff} |\chi(\mathbf{x}_1, \mathbf{x}_2)| &\leq 1. \end{aligned}$$

□

---

<sup>4</sup>vgl. Maierhofer, Georg/ Tichy, Robert: Skriptum zur Vorlesung Ungleichungen. Online in WWW unter URL: [http://www.math.tugraz.at/OeMO/F-Kurs/Unterlagen/Ungleichungen\\_Maierhofer.pdf](http://www.math.tugraz.at/OeMO/F-Kurs/Unterlagen/Ungleichungen_Maierhofer.pdf). [Stand 22.03.2012], S. 14

Die Korrelation quantifiziert wie linear der Zusammenhang zweier Datenvektoren ist<sup>5</sup>. Sie ist wie oben bewiesen betragsmäßig immer kleiner 1, allerdings gilt je näher der absolute Betrag der Korrelation bei 1 liegt, desto stärker ist der lineare Zusammenhang zweier Datenvektoren<sup>6</sup>.

Zwei Datenvektoren besitzen im Grunde einen linearen Zusammenhang, wenn prinzipiell für große Werte des einen die entsprechenden Werte des anderen ebenfalls groß sind, sowie umgekehrt für kleine Werte des einen die entsprechenden Werte des anderen klein sind. Die nächsten beiden Abbildungen zeigen dieses Prinzip grafisch veranschaulicht. Dabei hat der Datensatz aus Abbildung 3.3 die gemessene Korrelation  $\chi(\mathbf{x}_1, \mathbf{x}_2) \approx 0,4$  und der Datensatz aus Abbildung 3.4 die gemessene Korrelation  $\chi(\mathbf{x}_1, \mathbf{x}_2) \approx 0$ .

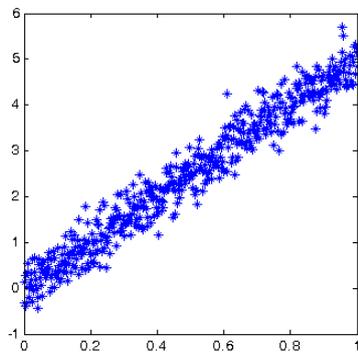


Abb. 3.3: Daten mit linearem Zusammenhang

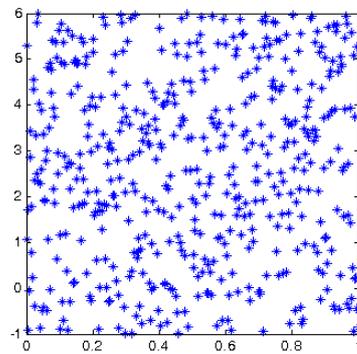


Abb. 3.4: Daten ohne linearem Zusammenhang

### 3.2.2 Zentrieren

Wenn man das Prinzip der Korrelation ausnutzen möchte, erkennt man, dass aufgrund der Definition 3.2.1 Daten mit Mittelwert 0 deutlich leichter zu verarbeiten sind. Auch erkennt man, dass die Kovarianz invariant gegenüber einer Verschiebung des Datenursprungs ist:

*Beweis.* Seien die Daten wie oben als  $(\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n)$  und die Skalare  $a, b \in \mathbb{R}$  gegeben. Dann sind  $\mathbf{y}_1$  und  $\mathbf{y}_2$  mit  $\mathbf{y}_1 = \mathbf{x}_1 + a$  und  $\mathbf{y}_2 = \mathbf{x}_2 + b$  lineare Transformationen der Ausgangsdaten. Dabei gilt

$$\begin{aligned} \overline{\mathbf{y}_1} &= \overline{\mathbf{x}_1} + a \\ \overline{\mathbf{y}_2} &= \overline{\mathbf{x}_2} + b. \end{aligned} \tag{3.3}$$

<sup>5</sup>vgl. Arens, Tilo/ Hettlich, Frank/ Karpfinger, Christian/ et al: Mathematik , S. 1257

<sup>6</sup>vgl. ebda, S. 1259

Es ist nun nach der Definition 3.2.1:

$$\begin{aligned}
 \kappa(\mathbf{x}_1, \mathbf{x}_2) &= \frac{1}{n} (\mathbf{x}_1 - \bar{\mathbf{x}}_1)^T (\mathbf{x}_2 - \bar{\mathbf{x}}_2) \\
 &= \frac{1}{n} ((\mathbf{x}_1 + a) - (\bar{\mathbf{x}}_1 + a))^T ((\mathbf{x}_2 + b) - (\bar{\mathbf{x}}_2 + b)) \\
 &\stackrel{(3.3)}{=} \frac{1}{n} (\mathbf{y}_1 - \bar{\mathbf{y}}_1)^T (\mathbf{y}_2 - \bar{\mathbf{y}}_2) \\
 &= \kappa(\mathbf{y}_1, \mathbf{y}_2)
 \end{aligned}$$

□

Ähnlich lässt sich beweisen, dass die Korrelation invariant gegenüber einer Verschiebung des Datenursprungs ist.

Man kann also die Daten ohne den Verlust von Information bezüglich deren linearer Zusammenhänge zentrieren<sup>7</sup>.

Es sind die zentrierten Daten gegeben als

$$\mathbf{x}_{zi} = \mathbf{x}_i - \bar{\mathbf{x}}_i$$

für  $i = 1, 2, \dots, m$ . Das Zentrieren der Daten hat grafisch veranschaulicht den Effekt einer Verschiebung. Dies erkennt man in den nächsten beiden Grafiken.

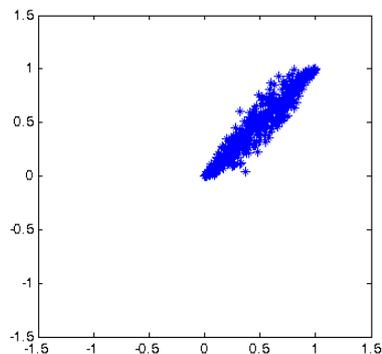


Abb. 3.5: Zweidimensionaler Datensatz

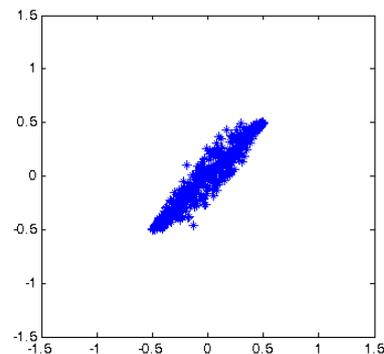


Abb. 3.6: Zentrierte Daten

### 3.2.3 Bedingungen zur Bestimmung der Hauptkomponenten

**Definition 3.2.2.** Sind die Datenvektoren  $\mathbf{x}_{z1}, \mathbf{x}_{z2}, \dots, \mathbf{x}_{zm} \in \mathbb{R}^n$  mit  $n \geq m \geq 2$  gegeben, so sind die Hauptkomponenten  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$  des Datensatzes die Menge jener Datenvektoren, welche die folgenden Eigenschaften besitzen:

<sup>7</sup>Zentrierte Daten haben Mittelwert 0

1. Jeder Vektor  $\mathbf{x}_{z_i}$  lässt sich als eine Linearkombination der Hauptkomponenten darstellen.
2. Die Hauptkomponenten sind paarweise unkorreliert mit jeweiliger Varianz 1.

(Definitions-Quelle: <sup>8</sup>)

Sei die in Definition 3.2.2 besprochene Darstellung der Daten gegeben als

$$\mathbf{x}_{z_i} = \sum_{j=1}^m a_{i,j} \mathbf{y}_j$$

für  $i = 1, 2, \dots, m$ . Sei ab nun

$$X_z = \begin{pmatrix} \mathbf{x}_{z_1}^T \\ \mathbf{x}_{z_2}^T \\ \vdots \\ \mathbf{x}_{z_m}^T \end{pmatrix}, A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,m} \end{pmatrix}, Y = \begin{pmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \vdots \\ \mathbf{y}_m^T \end{pmatrix}.$$

Dann lässt sich die Darstellung schreiben als:

$$X_z = AY.$$

Um die Hauptkomponenten zu bestimmen wird somit nach einer inversen Matrix  $W \approx A^{-1}$  gesucht, sodass die durch die Transformation

$$Y = WX_z$$

entstehenden Komponenten  $Y$  unkorreliert sind.

Die zweite Bedingung der Definition lässt sich anders formulieren als:

Für alle paarweise verschiedenen  $i, j \in \mathbb{Z}^+$  mit  $1 \leq i, j \leq m$  gilt

$$\chi(\mathbf{y}_i, \mathbf{y}_j) = 0.$$

Dies ist nach (3.1) äquivalent zu

$$\kappa(\mathbf{y}_i, \mathbf{y}_j) = 0.$$

---

<sup>8</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten. Online in WWW unter URL: <http://math.uni-graz.at/keeling/skripten/Tutorium.pdf>. [Stand 19.02.2013], S. 31

### 3.2.4 Die Kovarianzmatrix

**Definition 3.2.3.** Für einen Datensatz aus  $m \geq 2$  Datenvektoren  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  ist die Kovarianzmatrix  $K$  definiert als

$$K = \begin{pmatrix} \kappa(\mathbf{x}_1, \mathbf{x}_1) & \kappa(\mathbf{x}_1, \mathbf{x}_2) & \cdots & \kappa(\mathbf{x}_1, \mathbf{x}_m) \\ \kappa(\mathbf{x}_2, \mathbf{x}_1) & \kappa(\mathbf{x}_2, \mathbf{x}_2) & \cdots & \kappa(\mathbf{x}_2, \mathbf{x}_m) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(\mathbf{x}_m, \mathbf{x}_1) & \kappa(\mathbf{x}_m, \mathbf{x}_2) & \cdots & \kappa(\mathbf{x}_m, \mathbf{x}_m) \end{pmatrix}.$$

(Definitions-Quelle: <sup>9</sup>)

Für den noch nicht bestimmten Datensatz der Hauptkomponenten  $Y$  wäre die Kovarianzmatrix gegeben als

$$K = \frac{1}{n} Y Y^T.$$

Damit hat  $K \in \mathbb{R}^{m \times m}$  im Fall der Hauptkomponenten folgende Form:

$$K = \begin{pmatrix} \sigma^2(\mathbf{y}_1) & \kappa(\mathbf{y}_1, \mathbf{y}_2) & \cdots & \kappa(\mathbf{y}_1, \mathbf{y}_m) \\ \kappa(\mathbf{y}_2, \mathbf{y}_1) & \sigma^2(\mathbf{y}_2) & \cdots & \kappa(\mathbf{y}_2, \mathbf{y}_m) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(\mathbf{y}_m, \mathbf{y}_1) & \kappa(\mathbf{y}_m, \mathbf{y}_2) & \cdots & \sigma^2(\mathbf{y}_m) \end{pmatrix}.$$

Da die Datenvektoren  $\mathbf{y}_i$  als Hauptkomponenten paarweise unkorreliert sind und Varianz 1 haben, muss die Kovarianzmatrix die Form

$$K = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = I$$

haben. Dies bedeutet nun, dass zur Bestimmung der Hauptkomponenten  $Y$  eine Transformation  $W$  der Matrix  $X_z$  gesucht wird, so dass die Kovarianzmatrix von  $Y$  die Einheitsmatrix ist, also

$$K = \frac{1}{n} (W X_z)(W X_z)^T = I.$$

---

<sup>9</sup>vgl. Arens, Tilo/ Hettlich, Frank/ Karpfinger, Christian/ et al: Mathematik , S. 1262

### 3.2.5 Eigenräume

**Definition 3.2.4.** Man nennt einen Vektor  $\mathbf{v} \in \mathbb{R}^m$  einen Eigenvektor einer quadratischen Matrix  $A \in \mathbb{R}^{m \times m}$ , wenn es einen Skalar  $\lambda \in \mathbb{R}$  gibt, sodass gilt

$$A\mathbf{v} = \lambda\mathbf{v}.$$

In diesem Fall heißt  $\lambda$  Eigenwert von  $A$  zum Eigenvektor  $\mathbf{v}$ .

(Definitions-Quelle: <sup>10</sup>)

Weitere Informationen dazu findet man in <sup>11</sup> und <sup>12</sup>.

Eine quadratische Matrix  $A$  der Form  $A \in \mathbb{R}^{m \times m}$  hat genau  $m$  normalisierte<sup>13</sup> Eigenvektoren und jeweils einen zu jedem Vektor gehörigen Eigenwert - also insgesamt auch  $m$  Eigenwerte. Diese sind geordnet durch  $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_m|$ <sup>14</sup>, wobei der Eigenwert  $\lambda_i$  jener des Eigenvektors  $\mathbf{v}_i$  ist.

**Definition 3.2.5.** Zu der Matrix  $A \in \mathbb{R}^{m \times m}$  seien  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  Eigenvektoren mit dazugehörigen Eigenwerten  $\lambda_1, \lambda_2, \dots, \lambda_m$ . Fasst man nun diese in Matrizen  $V \in \mathbb{R}^{m \times m}$  und  $\Lambda \in \mathbb{R}^{m \times m}$  wie folgt zusammen:

$$V = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{pmatrix}$$
$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{pmatrix}$$

so ist eine Eigenraumzerlegung der Matrix  $A$  gegeben durch:

$$AV = V\Lambda.$$

(Definitions-Quelle: <sup>15</sup>)

---

<sup>10</sup>vgl. Arens, Tilo/ Hettlich, Frank/ Karpfinger, Christian/ et al: Mathematik , S. 594

<sup>11</sup>Arens, Tilo/ Hettlich, Frank/ Karpfinger, Christian/ et al: Mathematik

<sup>12</sup>Keeling, Stephen: Grundlagen der Numerischen Mathematik. Online in WWW unter URL: <http://math.uni-graz.at/keeling/skripten/numerik.pdf>. [Stand 15.12.2012]

<sup>13</sup>Normalisierte Vektoren haben den Betrag 1

<sup>14</sup>vgl. Keeling, Stephen: Grundlagen der Numerischen Mathematik , S. 99

<sup>15</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten , S. 12

### 3.2.6 Bestimmung der Hauptkomponenten

Die Lösung zur Bestimmung der Hauptkomponenten lässt sich sehr elegant durch Umschreiben mittels der Eigenraumzerlegung erkennen. Dazu wird die Eigenraumzerlegung der Kovarianzmatrix  $K = \frac{1}{n}X_z X_z^T$  betrachtet. Die Matrix ist offensichtlich symmetrisch, und auch positiv semidefinit wegen:

*Beweis.* Für jeden Vektor  $\mathbf{v} \in \mathbb{R}^m$  mit  $\|\mathbf{v}\| > 0$  gilt:

$$\begin{aligned} \mathbf{v}^T K K^T \mathbf{v} &= (K^T \mathbf{v})^T (K^T \mathbf{v}) \\ &= \|K^T \mathbf{v}\|^2 \\ &\geq 0. \end{aligned}$$

□

Solche Matrizen haben orthonormale<sup>16</sup> Eigenvektoren und nicht negative Eigenwerte<sup>17</sup>. Aus diesem Grund ist  $\Lambda^{-\frac{1}{2}}$  wohl definiert und

$$\begin{aligned} V^T V &= I \\ \iff V^T &= V^{-1}. \end{aligned} \tag{3.4}$$

Es ist nun nach der Eigenraumzerlegung der Matrix  $K$ :

$$\begin{aligned} KV &= V\Lambda \\ \iff K &= V\Lambda V^T \\ \stackrel{K = \frac{1}{n}X_z X_z^T}{\iff} V\Lambda V^T &= \frac{1}{n}X_z X_z^T \\ \stackrel{(3.4)}{\iff} \Lambda &= \frac{1}{n}V^T X_z X_z^T V. \end{aligned} \tag{3.5}$$

Es gilt außerdem, da nach Definition 3.2.5  $\Lambda$  diagonal ist

$$I = \Lambda^{-\frac{1}{2}} \Lambda \Lambda^{-\frac{1}{2}}.$$

Einsetzen von (3.5) ergibt nun:

$$\begin{aligned} I &= \frac{1}{n} \Lambda^{-\frac{1}{2}} V^T X_z X_z^T V \Lambda^{-\frac{1}{2}} \\ \iff I &= \frac{1}{n} \left( \Lambda^{-\frac{1}{2}} V^T X_z \right) \left( \Lambda^{-\frac{1}{2}} V^T X_z \right)^T \end{aligned} \tag{3.6}$$

<sup>16</sup>Orthonormale Vektoren sind paarweise orthogonal und haben jeweils den Betrag 1

<sup>17</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten, S. 22

da für drei Matrizen  $A, B, C$  mit entsprechenden Dimensionen immer gilt:

$$(ABC)^T = C^T B^T A^T.$$

Wählt man nun die Hauptkomponenten  $Y$  als

$$Y = \Lambda^{-\frac{1}{2}} V^T X_z$$

so ist deren Kovarianzmatrix nach (3.6) tatsächlich die Einheitsmatrix.

Also ist die gesuchte Transformation gegeben durch

$$W = \Lambda^{-\frac{1}{2}} V^T \approx A^{-1}.$$

Und damit sind die Hauptkomponenten gegeben durch

$$Y = \Lambda^{-\frac{1}{2}} V^T X_z.$$

Der grafische Effekt der Transformation entspricht einer Drehstreckung der Daten wie man in den folgenden Abbildungen erkennen kann. Die Transformation legt also die Achsen eines gedrehten orthogonalen Achsensystemes fest (siehe Abbildung 3.7), welches so gedreht und gestreckt wird, dass es nach der Transformation dem natürlichen Achsensystem entspricht (siehe Abbildung 3.8).

Diese Transformation wird auch Sphären genannt - dementsprechend heißen die Daten  $Y$  auch die gesphärten Daten<sup>18</sup>.

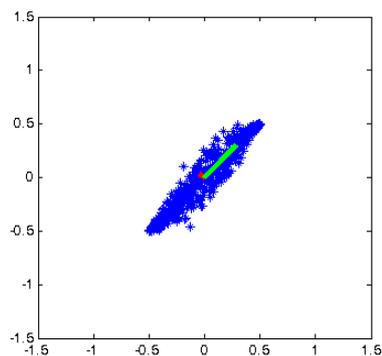


Abb. 3.7: Zentrierte Daten

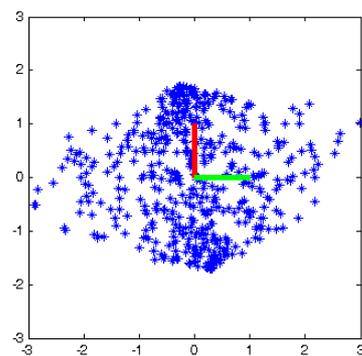


Abb. 3.8: Gesphärte Daten

<sup>18</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten, S. 22

### 3.2.7 Reduktion der Datendimension - Darstellung durch wenige Hauptkomponenten

Sind die Hauptkomponenten durch die obige Transformation bereits bestimmt, so kann man umgekehrt daraus den ursprünglichen Datensatz bestimmen mit:

$$X_z = W^{-1}Y \approx AY.$$

Nach (3.4) ist

$$W^{-1} = \Lambda^{\frac{1}{2}}V$$

und deshalb sind die ursprünglichen zentrierten Daten darstellbar als:

$$X_z = \Lambda^{\frac{1}{2}}VY. \quad (3.7)$$

Möchte man nun die ursprünglichen Daten nur durch die stärksten  $r$  Hauptkomponenten darstellen so genügt die folgende Transformation:

Sei die Matrix  $R \in \mathbb{R}^{m \times m}$  gegeben als

$$R = \text{diag}\{\underbrace{1, \dots, 1}_r, 0, \dots, 0\}.$$

Hierbei sind die ersten  $r$  Einträge 1 und die restlichen 0.

Die Darstellung der zentrierten Daten durch die stärksten  $r$  Hauptkomponenten sei  $X_{z,r}$  und ist gegeben durch:

$$X_{z,r} = \Lambda^{\frac{1}{2}}VRY.$$

Sei die Darstellung der ursprünglichen Daten durch die stärksten  $r$  Hauptkomponenten  $X_r$ , so ist diese gegeben als

$$X_r = X_{z,r} + \bar{X}$$

für

$$\bar{X} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_m \end{pmatrix}.$$

Durch diese Darstellung der Daten lässt sich die Zielstellung von PCA - die Verringerung der Datendimension bei gleichzeitigem Beibehalt möglichst vieler Informationen erfüllen.

In den folgenden Abbildungen wird eine Darstellung durch weniger Hauptkomponenten im Vergleich zum ursprünglichen Datensatz gezeigt.

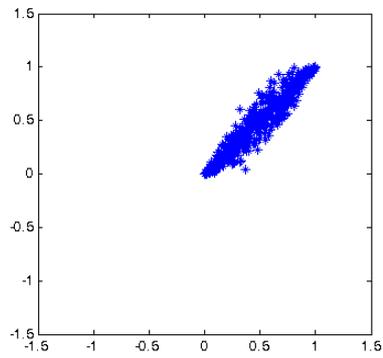


Abb. 3.9: Zweidimensionaler Datensatz

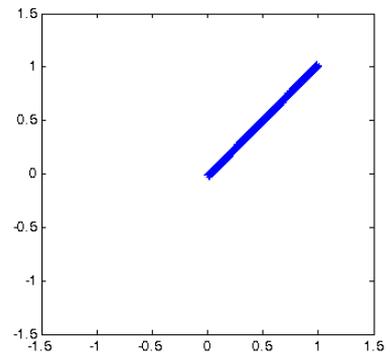


Abb. 3.10: Darstellung der Daten durch weniger Hauptkomponenten

# 4 Einführung in das Verfahren

## Independent Component Analysis (ICA)

### 4.1 Motivation

Die **Unabhängigkeitsanalyse (Independent Component Analysis)** ist eine Methode, um einem Datensatz zugrundeliegende Komponenten zu finden<sup>1</sup>, die sich in einer Weise unterscheiden, die fundamentaler ist, als jene welche im vorherigen Kapitel im Zuge der Hauptkomponentenanalyse vorgestellt wurde. Damit können in Datensätzen bestimmte Komponenten in der Darstellung hervorgehoben oder entfernt werden und somit diese Daten gezielter betrachtet und verarbeitet werden.

Diese Komponenten heißen unabhängige Komponenten (Independent Components). Die Zielstellung und Modellierung des Problems ist ähnlich zu jener der Hauptkomponentenanalyse, allerdings wird im Gegensatz zu der in ihr verwendeten Korrelation das weitaus mächtigere Prinzip der statistischen Unabhängigkeit genutzt, um die unabhängigen Komponenten zu definieren und zu bestimmen.

In den folgenden Grafiken erkennt man, dass bei der Bestimmung der unabhängigen Komponenten der Zusammenhang jener deutlich besser entfernt wird, als bei der Bestimmung der Hauptkomponenten. Im ursprünglichen Datensatz gibt es durch die diamantenartige Form einen deutlichen Zusammenhang (vgl. Abbildung 4.1). In den durch PCA bestimmten Hauptkomponenten bleibt dieser erhalten (die Daten werden lediglich skaliert, haben allerdings noch immer die diamantenartige Form) (vgl. Abbildung 4.2). Erst durch die Verwendung des mächtigen Prinzipes der statistischen Unabhängigkeit in ICA gelingt es die Daten so zu drehen, dass sie tatsächlich ohne Zusammenhang sind (vgl. Abbildung 4.3).

---

<sup>1</sup>vgl. Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis , S. xvii

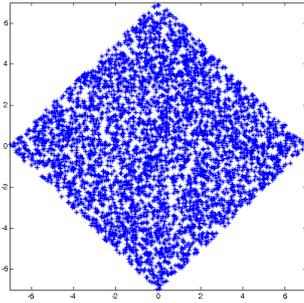


Abb. 4.1: Diamantförmiger Datensatz

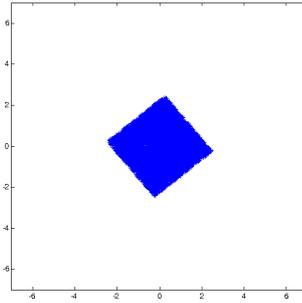


Abb. 4.2: Durch PCA transformierte Daten

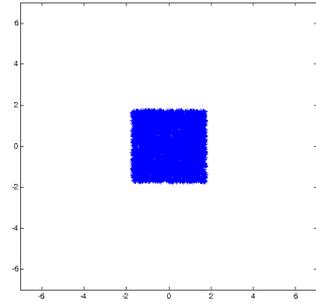


Abb. 4.3: Durch ICA transformierte Daten

## 4.2 Das Verfahren in der Praxis

### 4.2.1 Vergleich Unkorreliertheit und Unabhängigkeit

**Definition 4.2.1.** Zwei Ereignisse  $A$  und  $B$  heißen statistisch unabhängig, wenn gilt

$$P(A \cap B) = P(A)P(B).$$

(Definitions-Quelle: <sup>2)</sup>)

Für Zufallsvariablen wird dies zu:

**Definition 4.2.2.** Zwei Zufallsvariablen  $x, y$  mit den Ergebnisräumen (Ein Ereignisraum einer Zufallsvariable ist die Menge aller möglichen Ergebnisse des dazugehörigen Zufallsexperimentes<sup>3</sup>.)  $\Omega_1, \Omega_2$  heißen statistisch unabhängig, wenn für jedes Ereignis  $\omega_1 \in \Omega_1$  und jedes Ereignis  $\omega_2 \in \Omega_2$  gilt:

$$P(x = \omega_1 \cap y = \omega_2) = P(x = \omega_1)P(y = \omega_2).$$

Da in dem Verfahren der Unabhängigkeitsanalyse lediglich mit Datenvektoren gearbeitet wird, ist es sinnvoll die statistische Unabhängigkeit auch für Datenvektoren zu definieren:

**Definition 4.2.3.** Man betrachte die zwei Datenvektoren  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  als Abtastungen zweier Zufallsvariablen  $x, y \in \mathbb{R}$ . Die Datenvektoren  $\mathbf{x}, \mathbf{y}$  heißen statistisch unabhängig, wenn die entsprechenden Zufallsvariablen  $x, y$  statistisch unabhängig nach der Definition 4.2.2 sind.

<sup>2</sup>vgl. Arens, Tilo/ Hettlich, Frank/ Karpfinger, Christian/ et al: Mathematik , S. 1282

<sup>3</sup>vgl. mathe online: Zufall und Wahrscheinlichkeit. Online in WWW unter URL:

<http://www.mathe-online.at/mathint/wstat1/i.html#Beispiele123>. [Stand 19.02.2013]

Wie bereits in Kapitel 3.2.1 beschrieben gibt die Korrelation Aufschluss darüber, ob zwei Datenvektoren linear zusammenhängen. Der Nachteil der Korrelation ist allerdings, dass diese nur über den linearen Zusammenhang und über keinen weiteren (z.B.: quadratischen oder kubischen Zusammenhang) Informationen beinhaltet. Sind zwei Datenvektoren aber unabhängig, so bedeutet dies, dass tatsächlich kein Zusammenhang zwischen den beiden besteht.

Der Unterschied, beziehungsweise, die weitaus umfassendere Wirkung der Unabhängigkeit wird am folgenden Beispiel deutlich:

**Beispiel 4.2.1.** *Man betrachte die beiden Datenvektoren  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ . Seien  $a_i \in [0, 1]$  mit  $1 \leq i \leq n$  zufällig gewählte Zahlen, und die beiden Datenvektoren für alle  $1 \leq i \leq n$  gegeben als*

$$\mathbf{x}(i) = -5 + \frac{1}{n-1}10(i-1)$$

$$\mathbf{y}(i) = (\mathbf{x}(i))^2 + a_i.$$

Dies bedeutet, dass in diesem Beispiel die Werte des Vektors  $\mathbf{x}$  regelmäßig im Intervall  $[-5, 5]$  verteilt sind, während die Werte des zweiten Vektors  $\mathbf{y}$  zufällig aber annähernd in der Nähe der Funktion  $f(\mathbf{x}(i)) = (\mathbf{x}(i))^2$  verteilt sind. Es besteht, wie man auch in der Grafik erkennen kann, ein eindeutiger quadratischer Zusammenhang zwischen den beiden Datenvektoren.

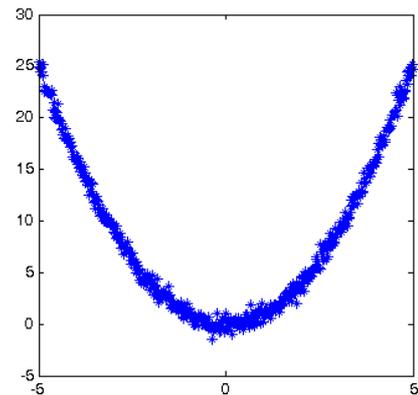


Abb. 4.4: Datenvektoren mit quadratischem Zusammenhang

Es ist nun allerdings nach der Definition 3.2.1.:

$$\kappa(\mathbf{x}, \mathbf{y}) = \frac{1}{n} (\mathbf{x} - \bar{\mathbf{x}})^T (\mathbf{y} - \bar{\mathbf{y}}).$$

Da nach der Wahl in diesem Beispiel die Schätzung  $\mathbf{y}(i) \approx (\mathbf{x}(i))^2$  sehr gut ist, ist also:

$$\begin{aligned} \kappa(\mathbf{x}, \mathbf{y}) &\approx \frac{1}{n} \sum_{i=1}^n \left( \mathbf{x}(i) - \frac{1}{n} \sum_{j=1}^n \mathbf{x}(j) \right) \left( \mathbf{x}(i)^2 - \frac{1}{n} \sum_{j=1}^n \mathbf{x}(j)^2 \right) \\ \overset{\sum_{j=1}^n \mathbf{x}(j)=0}{\iff} \kappa(\mathbf{x}, \mathbf{y}) &\approx \frac{1}{n} \sum_{i=1}^n \left( \mathbf{x}(i)^3 - \mathbf{x}(i) \frac{1}{n} \sum_{j=1}^n \mathbf{x}(j)^2 \right) \\ \overset{\sum_{i=1}^n \mathbf{x}(i)^3=0}{\iff} \kappa(\mathbf{x}, \mathbf{y}) &\approx -\frac{1}{n} \sum_{i=1}^n \mathbf{x}(i) * \frac{1}{n} \sum_{j=1}^n \mathbf{x}(j)^2 \\ \overset{\sum_{i=1}^n \mathbf{x}(i)=0}{\iff} \kappa(\mathbf{x}, \mathbf{y}) &\approx 0. \end{aligned}$$

Somit ist für diese Wahl die Korrelation ungefähr 0. Im nächsten Schritt wird betrachtet, ob die Variablen auch unabhängig sind: Da in diesem Beispiel die Werte des Vektors  $\mathbf{x}$  regelmäßig im Intervall  $[-5, 5]$  verteilt sind, ist nun

$$P(\mathbf{x}(i) > 3) = \frac{2}{10}.$$

Desweiteren ist für nahezu jeden Wert in  $\mathbf{x}$  mit  $|\mathbf{x}(i)| < 2$  der entsprechende Wert in  $\mathbf{y}$  kleiner als 4, also ist

$$P(\mathbf{y}(i) < 4) \approx P(|\mathbf{x}(i)| < 2) = \frac{4}{10}.$$

Wären die Datenvektoren unabhängig, so müsste nach der Definition 4.2.3 gelten:

$$\begin{aligned} P(\mathbf{x}(i) > 3 \cap \mathbf{y}(i) < 4) &= P(\mathbf{x}(i) > 3)P(\mathbf{y}(i) < 4) \\ &\approx \frac{2}{10} * \frac{4}{10} > 0 \end{aligned}$$

Da offensichtlich für kein  $\mathbf{x}(i) > 3$  das entsprechende  $\mathbf{y}(i)$  kleiner sein kann als  $3^2 - 1 = 8$ , ist aber tatsächlich

$$P(\mathbf{x}(i) > 3 \cap \mathbf{y}(i) < 4) = 0.$$

Darum sind die beiden Zufallsvariablen nicht unabhängig. Mithilfe der Korrelation alleine kann dies aber nicht festgestellt werden. Zu bemerken ist noch, dass unabhängige Variablen auch immer zwangsläufig unkorreliert sind<sup>4</sup>.

---

<sup>4</sup>vgl. Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis , S. 159

## 4.2.2 Zentrieren

Auch im Verfahren der Unabhängigkeitsanalyse wird die Theorie für Daten mit Mittelwert 0 um einiges einfacher und wie vorhin kann man, ohne jegliche Information über den Zusammenhang in den ursprünglichen Daten zu verlieren, diese zentrieren<sup>5</sup>. Es sind die zentrierten Daten somit erneut gegeben als

$$\mathbf{x}_{z_i} = \mathbf{x}_i - \bar{\mathbf{x}}_i$$

für  $i = 1, 2, \dots, m$ .

## 4.2.3 Bedingungen zur Bestimmung der unabhängigen Komponenten

**Definition 4.2.4.** Sind die Datenvektoren  $\mathbf{x}_{z_1}, \mathbf{x}_{z_2}, \dots, \mathbf{x}_{z_m} \in \mathbb{R}^n$  gegeben, so sind unabhängige Komponenten  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m$  des Datensatzes definiert durch:

1. Die unabhängigen Komponenten sind Hauptkomponenten des Datensatzes wie in der Definition 3.2.2 beschrieben.
2. Die unabhängigen Komponenten sind statistisch unabhängig, wie in Definition 4.2.3 beschrieben.

(Definitions-Quelle: <sup>6</sup>)

Nach 1. der obigen Definition gibt es gleich wie in Kapitel 3.2.3 eine Darstellung der ursprünglichen Daten als Linearkombination der unabhängigen Komponenten. Sei diese gegeben als

$$\mathbf{x}_{z_i} = \sum_{j=1}^m b_{i,j} \mathbf{z}_j$$

für  $i = 1, 2, \dots, m$ . Sei ab nun ähnlich wie im Kapitel 3.2.3

$$X_z = \begin{pmatrix} \mathbf{x}_{z_1}^T \\ \mathbf{x}_{z_2}^T \\ \vdots \\ \mathbf{x}_{z_m}^T \end{pmatrix}, B = \begin{pmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,m} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,m} \end{pmatrix}, Z = \begin{pmatrix} \mathbf{z}_1^T \\ \mathbf{z}_2^T \\ \vdots \\ \mathbf{z}_m^T \end{pmatrix}.$$

<sup>5</sup>vgl. Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis , S. 154

<sup>6</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten , S. 31

Dann lässt sich die Darstellung schreiben als:

$$X_z = BZ.$$

Um die unabhängigen Komponenten zu bestimmen, wird somit nach einer inversen Matrix  $V \approx B^{-1}$  gesucht, sodass die durch die Transformation

$$Z = VX_z$$

entstehenden Komponenten in den Reihen von  $Z$  möglichst unabhängig sind.

**Definition 4.2.5.** *Nachdem eine, wie oben beschriebene, inverse Matrix  $V \approx B^{-1}$  nicht immer existiert - beziehungsweise nicht immer numerisch eindeutig bestimmbar ist - werden die, mit der bestmöglichen (nach der Optimierung einer ausgewählten Zielfunktion<sup>7</sup>) Approximation für  $V \approx B^{-1}$  bestimmten, unabhängigen Komponenten als "möglichst unabhängig" bezeichnet.*

#### 4.2.3.1 Sphären als Vorschritt

Nimmt man an, dass die Daten bereits wie im Kapitel 3.2.6 durch die Matrix  $W$  gesphärt sind, so bedeutet dies, dass nicht mehr die Matrix  $V$  (vgl. Kapitel 4.2.3) zur Bestimmung der unabhängigen Komponenten gesucht ist, sondern nun eine Matrix  $U$  bestimmt werden muss, sodass

$$Z = UY = U\Lambda^{-\frac{1}{2}}V^T X_z \quad (4.1)$$

gilt. Bereits im Kapitel 4.2.1 war erkennbar, dass diese gesphärten Daten nicht zwangsläufig die unabhängigen Komponenten sind. Aus diesem Grund ist nun eine Matrix  $U$  gesucht, sodass die durch

$$Z = UY$$

dargestellten Daten (möglichst) unabhängig sind.

Der Nutzen, den man nun daraus zieht, die gesphärten Daten zu betrachten, ist, dass hier die Transformationsmatrix  $U$  orthonormal sein muss.

---

<sup>7</sup>Der verwendete ICA-Algorithmus bestimmt welches Optimierungsverfahren und welche Zielfunktion verwendet werden

*Beweis.* Man kann dies erkennen, wenn man die Kovarianzmatrix  $K$  der noch nicht bestimmten unabhängigen Komponenten betrachtet. Nach der Definition 4.2.4 ist diese die Einheitsmatrix, deshalb ist

$$\begin{aligned}
 I &= \frac{1}{n}ZZ^T \\
 &\stackrel{(4.1)}{=} \frac{1}{n}UYY^TU^T \\
 &= U\frac{1}{n}YY^TU^T \\
 &\stackrel{(3.6)}{=} UIU^T \\
 &= UU^T \\
 \iff I &= UU^T. \tag{4.2}
 \end{aligned}$$

Somit muss  $U$  tatsächlich orthonormal sein. □

Dies reduziert nun die Suche nach der Matrix  $U$  auf den Bereich von orthonormalen Matrizen und ermöglicht somit eine leichtere Bestimmung dieser Matrix.

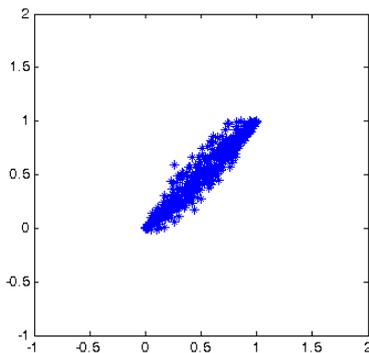


Abb. 4.5: Zweidimensionaler Datensatz

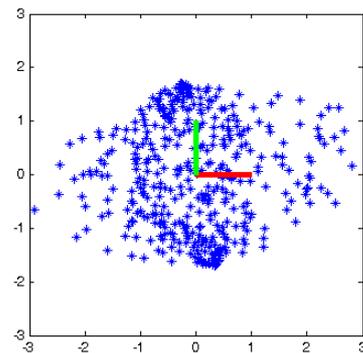


Abb. 4.6: Gesphärte Daten

Zur Bestimmung der unabhängigen Komponenten werden deshalb zunächst die Daten  $X_z$  gesphärnt und danach die Matrix  $U$  so bestimmt, dass die unabhängigen Komponenten in der Darstellung  $Z = UY$  (möglichst) unabhängig sind und dass die Matrix  $U$  orthonormal ist.

## 4.2.4 Die Bedingung der Unabhängigkeit und der Zentrale Grenzwertsatz

Um die zweite Bedingung der Definition 4.2.4 für die Bestimmung der unabhängigen Komponenten auszunützen, kann der zentrale Grenzwertsatz verwendet werden.

**Satz 4.2.1.** Seien  $k$  unabhängige identisch verteilte<sup>8</sup> Zufallsvariablen  $z_1, z_2, \dots, z_k \in \mathbb{R}$  mit  $k \in \mathbb{N}$  gegeben und sei die Zufallsvariable  $x_k \in \mathbb{R}$  gegeben als:

$$x_k = \sum_{i=1}^k z_i.$$

Da nun der Mittelwert und die Varianz von  $x_k$  ohne Grenze wachsen können, wenn  $k \rightarrow \infty$  betrachtet man die standardisierte Zufallsvariable  $y_k$  (wobei hier  $\sigma(x_k)$  die Standardabweichung von  $x_k$  ist, welche anhand der Definition 3.2.1 bestimmt werden kann) mit

$$y_k = \frac{x_k - \bar{x}_k}{\sigma(x_k)},$$

so besagt der Zentrale Grenzwertsatz, dass die Zufallsvariable  $y_k$  zu einer gauß-verteiltern Zufallsvariable mit Mittelwert 0 und Varianz 1 konvergiert<sup>9</sup>, wenn  $k \rightarrow \infty$ .

(Satz-Quelle: <sup>10</sup>)

Der Zentrale Grenzwertsatz weist somit darauf hin, dass die Mischung von Datenvektoren gaußscher verteilt ist, als die Datenvektoren selbst. In der Realität ist tatsächlich bereits für eine relativ kleine Anzahl an Datenvektoren die Mischung aus jenen nahe einer gaußschen Verteilung. Dies trifft meist auch dann zu, wenn die gegebenen Datenvektoren nicht identisch verteilt sind.

In den folgenden Balkendiagrammen wird die Verteilung der jeweiligen Daten gezeigt. Man kann erkennen, dass die Mischung in der Abbildung 4.9 von zwei - obwohl nicht gleich verteilten - Datenvektoren aus den Abbildungen 4.7 und 4.8 bereits einen deutlichen gaußschen Charakter besitzt.

---

<sup>8</sup>Zwei Zufallsvariablen heißen identisch verteilt, wenn deren Wahrscheinlichkeitsverteilungen gleich sind - im Fall der Unabhängigkeitsanalyse werden nur diskrete Wahrscheinlichkeitsverteilungen betrachtet

<sup>9</sup>Eine solche Zufallsvariable hat die Wahrscheinlichkeitsdichtefunktion

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

<sup>10</sup>vgl. Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis , S. 34

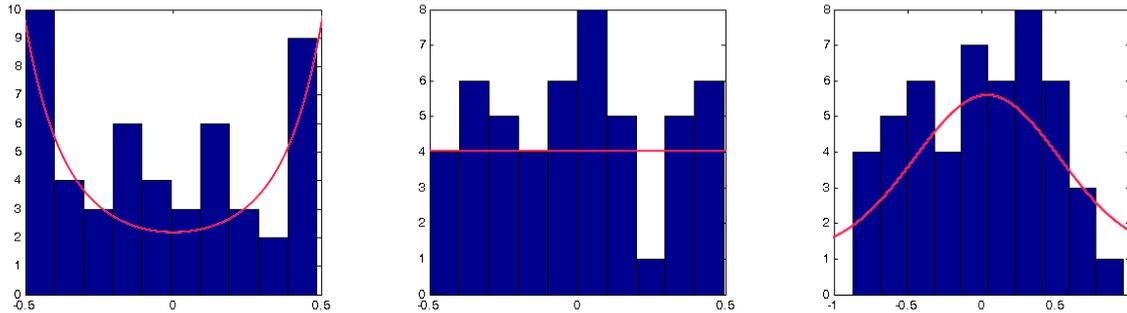


Abb. 4.7: Datenvektor mit an den Enden gewichteter Verteilung  
 Abb. 4.8: Datenvektor mit gleichmäßiger Verteilung  
 Abb. 4.9: Mischung der beiden nicht-gauß-verteilten Datenvektoren

Nach den obigen Überlegungen ist es naheliegend, das Prinzip<sup>11</sup> des Zentralen Grenzwertsatzes bei der Bestimmung der unabhängigen Komponenten eines Datensatzes auszunützen. Da nach der Definition 4.2.4 die ursprünglichen Datenvektoren eine Mischung aus den unabhängigen Komponenten darstellen, kann man annehmen, dass die ursprünglichen Datenvektoren gaußscher verteilt sind, als die entsprechenden unabhängigen Komponenten. Aus diesem Grund ist es sinnvoll, zur Bestimmung jener Komponenten nach einer Transformation  $U$  mit

$$Z = UY,$$

so zu suchen, dass die Verteilungen der entstehenden Komponenten in  $Z$  so wenig gaußschen Charakter wie möglich besitzen.

## 4.2.5 Messung der Gaußianität

**Definition 4.2.6.** Sei der Datenvektor  $\mathbf{x} \in \mathbb{R}^n$  gegeben, so ist das sogenannte  $k$ -te Moment dieses Datenvektors definiert als:

$$M_k(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n |\mathbf{x}(i) - \bar{\mathbf{x}}|^k.$$

(Definitions-Quelle: <sup>12</sup>)

<sup>11</sup>Gemeint ist hier die naheliegende Vermutung, dass Mischungen aus Datenvektoren meist gaußscher verteilt sind, als die Datenvektoren selbst.

<sup>12</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten, S. 30

Die Momente haben die folgende Eigenschaft, welche der Grund für deren Nützlichkeit zur Bestimmung der unabhängigen Komponenten ist:

**Satz 4.2.2.** Sei  $\mathbf{v} \in \mathbb{R}^n$  eine Abtastung einer gauß-verteiltern Zufallsvariable mit Mittelwert  $\mu$  und Varianz  $\sigma^2$ , dann gilt

$$\lim_{n \rightarrow \infty} M_k(\mathbf{v}) = \begin{cases} (k-1)!!\sigma^k, & k \text{ gerade} \\ \sqrt{\frac{2}{\pi}}(k-1)!!\sigma^k, & k \text{ ungerade} \end{cases}$$

(Satz-Quelle: <sup>13</sup>)

Aufgrund dieser Grenzwerteigenschaft eignen sich die Momente sehr gut für die Messung der Gaußianität einer Datenstruktur. Man kann deshalb die sogenannte Wölbung für die Messung der Gaußianität verwenden:

**Definition 4.2.7.** Sei der Datenvektor  $\mathbf{x} \in \mathbb{R}^n$  gegeben, dann ist die Wölbung dieses Vektors definiert als:

$$W(\mathbf{x}) = M_4(\mathbf{x}) - 3M_2(\mathbf{x})^2.$$

(Definitions-Quelle: <sup>14</sup>)

Die Wölbung hat nach den obigen Grenzwerteigenschaften der Momente  $M_k$  die Eigenschaft, bei einem Vektor mit normalverteilten Werten annähernd 0 zu sein. Zusätzlich weiß man, dass für eher gleichmäßig verteilte Werte  $W < 0$  und für eher spitzig verteilte Werte  $W > 0$  gilt<sup>15</sup>.

Seien die Zeilen der gesuchten Matrix  $U$  die Vektoren  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$  mit

$$U = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_m^T \end{pmatrix}.$$

Dann kann man, um die Matrix  $U$  - und damit die möglichst unabhängigen Komponenten zu bestimmen - nach dem Prinzip des Satzes 4.2.1 die Zielfunktion  $J_1(\mathbf{u}_i)$  für alle  $1 \leq i \leq m$  mit

$$J_1(\mathbf{u}_i) = -W(Y^T \mathbf{u}_i)^2 = -(M_4(Y^T \mathbf{u}_i) - 3M_2(Y^T \mathbf{u}_i)^2)^2$$

minimieren.

<sup>13</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten, S. 30

<sup>14</sup>vgl. ebda, S. 30

<sup>15</sup>vgl. ebda, S. 30

## 4.2.6 Minimierung der Zielfunktion $J_1$ durch Gradienten-Abstiegsverfahren

**Definition 4.2.8.** Der Gradient einer partiell differenzierbaren Funktion  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  im Punkt  $\mathbf{p}$  ist der Vektor der partiellen Ableitungen in diesem Punkt:

$$D_{\mathbf{x}}f(\mathbf{p}) = \begin{pmatrix} \frac{\partial f}{\partial x(1)}(\mathbf{p}) \\ \frac{\partial f}{\partial x(2)}(\mathbf{p}) \\ \vdots \\ \frac{\partial f}{\partial x(m)}(\mathbf{p}) \end{pmatrix}.$$

(Definitions-Quelle: <sup>16</sup>)

Um eine bessere Vorstellung des Gradienten zu ermöglichen, wird dies nun anhand eines kurzen Beispiels veranschaulicht:

**Beispiel 4.2.2.** Gegeben ist die Funktion  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  mit  $f(\mathbf{x}) = \mathbf{x}(1)^2 + \mathbf{x}(2)^2$  und der Punkt  $\mathbf{p} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ . Dann ist der Gradient der Funktion  $f(\mathbf{x})$  im Punkt  $\mathbf{p}$  gegeben als:

$$D_{\mathbf{x}}f(\mathbf{p}) = \begin{pmatrix} \frac{\partial f}{\partial x(1)}(\mathbf{p}) \\ \frac{\partial f}{\partial x(2)}(\mathbf{p}) \end{pmatrix}.$$

Es ist nun:

$$\begin{aligned} \frac{\partial f}{\partial x(1)}(\mathbf{p}) &= 2\mathbf{p}(1) \\ \frac{\partial f}{\partial x(2)}(\mathbf{p}) &= 2\mathbf{p}(2). \end{aligned}$$

Und deshalb ist:

$$D_{\mathbf{x}}f(\mathbf{p}) = \begin{pmatrix} 2\mathbf{p}(1) \\ 2\mathbf{p}(2) \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.$$

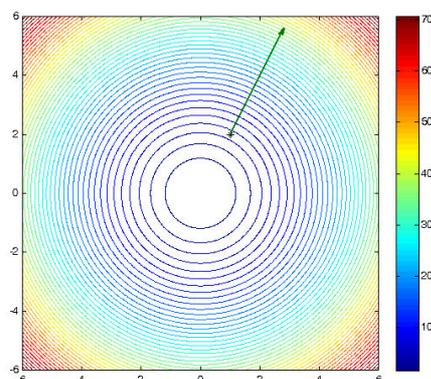


Abb. 4.10: Niveaulinien der Funktion  $f(\mathbf{x})$  und Gradient an der Stelle  $\mathbf{p}$

“Der Gradient [einer Funktion, d. Verf.] zeigt in [die, d. Verf.] Richtung des steilsten Anstiegs.”<sup>17</sup> Dies kann man in Abbildung 4.10 gut erkennen. Dementsprechend ist die Richtung

<sup>16</sup>vgl. Arens, Tilo/ Hettlich, Frank/ Karpfinger, Christian/ et al: Mathematik , S. 802

<sup>17</sup>ebda, S. 803

$\mathbf{d}^*(\mathbf{p})$  des steilsten Anstiegs der Funktion  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  im Punkt  $\mathbf{p}$  gegeben durch

$$\mathbf{d}^*(\mathbf{p}) = \frac{D_{\mathbf{x}}f(\mathbf{p})}{\|D_{\mathbf{x}}f(\mathbf{p})\|}. \quad (4.3)$$

Aus diesem Grund ist ein sinnvoller Algorithmus für die Minimierung einer Funktion  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  gegeben durch die Aktualisierungsregel für den minimalen Punkt  $\mathbf{m} \in \mathbb{R}^m$ :

$$\mathbf{m} \leftarrow \mathbf{m} - \alpha \mathbf{d}^*(\mathbf{m}).$$

Wobei vernünftige Werte für  $\alpha$  durch  $\alpha \in \{0, \frac{1}{2}, 1, 2\}$  gegeben sind<sup>18</sup>.

### 4.2.7 Algorithmus zur Bestimmung der unabhängigen Komponenten

Durch dieses Wissen ergibt sich der folgende Algorithmus - der Standard-Algorithmus von Independent Component Analysis - zur tatsächlichen Bestimmung der unabhängigen Komponenten. Dazu wird die Matrix  $U$  zeilenweise bestimmt, also werden die Vektoren  $\mathbf{u}_i$  für  $1 \leq i \leq m - 1$  im Algorithmus nacheinander wie folgt bestimmt:

1. Man beginnt mit einer zufälligen Wahl von  $\mathbf{u}_i$ , welche orthogonal zu allen bereits bestimmten Vektoren  $\mathbf{u}_j$  mit  $j < i$  ist und den Betrag 1 besitzt - dies sind notwendige Bedingungen aufgrund von (4.2).
2. Die Richtung  $\mathbf{d}^*(\mathbf{u}_i)$  des steilsten Anstiegs der Funktion  $J_1(\mathbf{u}_i)$  wird nach (4.3) berechnet durch

$$\mathbf{d}^*(\mathbf{u}_{i,alt}) = \frac{D_{\mathbf{u}_i} J_1(\mathbf{u}_{i,alt})}{\|D_{\mathbf{u}_i} J_1(\mathbf{u}_{i,alt})\|}.$$

3. Der Vektor  $\mathbf{u}_{i,alt}$  wird aktualisiert durch

$$\mathbf{u}_{i,neu} \leftarrow \mathbf{u}_{i,alt} - \alpha \mathbf{d}^*(\mathbf{u}_{i,alt})$$

wobei von den möglichen Werten für  $\alpha \in \{0, \frac{1}{2}, 1, 2\}$  jener gewählt wird, an dem  $J_1(\mathbf{u}_{i,neu})$  am kleinsten ist.

4. Ist  $i > 1$ , so wird der neue Vektor  $\mathbf{u}_{i,neu}$  in den zu den bereits berechneten Vektoren  $\mathbf{u}_j$  mit  $j < i$  normalen Raum aufgrund von (4.2) projiziert durch

$$\mathbf{u}_{i,neu} \leftarrow \mathbf{u}_{i,neu} - \frac{\mathbf{u}_j^T \mathbf{u}_{i,neu}}{\|\mathbf{u}_j\|^2} \mathbf{u}_j$$

für  $j \in \{1, 2, \dots, i - 1\}$ .

---

<sup>18</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten, S. 42

5. Der Vektor wird aufgrund von (4.2) normalisiert durch

$$\mathbf{u}_{ineu} \leftarrow \frac{\mathbf{u}_{ineu}}{\|\mathbf{u}_{ineu}\|}.$$

6. Ist  $\frac{\mathbf{u}_{alt}^T \mathbf{u}_{ineu}}{\|\mathbf{u}_{alt}\| \|\mathbf{u}_{ineu}\|} < 0$ , so wird, damit der Betrag im Schritt 7. tatsächlich gegen 0 konvergieren kann,  $\mathbf{u}_{ineu}$  ersetzt durch

$$\mathbf{u}_{ineu} \leftarrow -\mathbf{u}_{ineu}.$$

7. Die obigen Schritte 2. - 6. werden so lange wiederholt bis der Betrag  $\|\mathbf{u}_{alt} - \mathbf{u}_{ineu}\|$  annähernd 0 ist, das heißt, der Vektor  $\mathbf{u}_i$  konvergiert.

Der letzte Vektor  $\mathbf{u}_m$  ist aufgrund von (4.2) normal zu allen bereits bestimmten und hat zusätzlich den Betrag 1. Die Orientierung kann beliebig gewählt werden und damit kann dieser sehr leicht bestimmt werden.

Dadurch sind nun die Matrix  $U$  und damit die unabhängigen Komponenten in  $Z$  des Datensatzes  $X_z$  berechnet mit

$$Z = UY = UW X_z.$$

Die Multiplikation der gesphärten Daten mit der Matrix  $U$  hat den Effekt einer Drehung, wie man auch in den folgenden Abbildungen erkennen kann. Jede Drehung der gesphärten Daten ist gleichwertig bezüglich der Korrelation<sup>19</sup>, durch Independent Component Analysis wird allerdings eine Drehung gefunden, so dass die entstehenden Daten möglichst unabhängig sind.

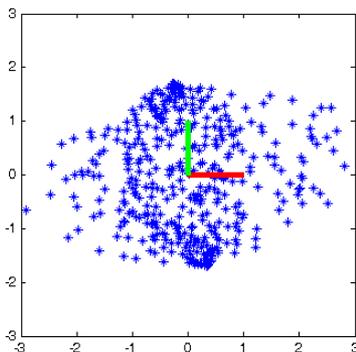


Abb. 4.11: Gesphärte Daten

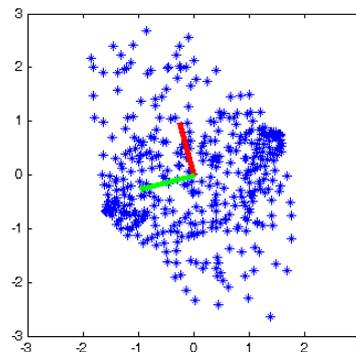


Abb. 4.12: Unabhängige Komponenten

<sup>19</sup>vgl. Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis , S. 160

## 4.2.8 Hervorheben von Komponenten - Darstellung durch weniger unabhängige Komponenten

Es gilt

$$\begin{aligned} Z &= UY \\ \Leftrightarrow U^{-1}Z &= Y. \end{aligned}$$

Nach (4.2) gilt

$$\begin{aligned} U^T U &= I \\ \Leftrightarrow U^T &= U^{-1} \end{aligned}$$

und dementsprechend lassen sich die Hauptkomponenten des Datensatzes wieder rekonstruieren durch

$$Y = U^T Z.$$

Der zentrierte Datensatz lässt sich dann aufgrund von (3.7) rekonstruieren durch

$$X_z = V \Lambda^{-\frac{1}{2}} U^T Z.$$

Möchte man nun gezielt einzelne unabhängige Komponenten ausblenden, so kann man dies durch die Multiplikation mit folgender Matrix machen:

Zum Ausblenden der  $i$ -ten Komponente muß man im  $i$ -ten Eintrag in der Matrix  $P = \text{diag}\{1, 0, 1, 1, \dots, 0\} \in \mathbb{R}^{m \times m}$  eine 0 setzen, soll sie eingeblendet bleiben, muss eine 1 gesetzt werden. Sei  $X_{z,p}$  die Darstellung der zentrierten Daten durch lediglich die gewünschten unabhängigen Komponenten, so gilt:

$$X_{z,p} = V \Lambda^{-\frac{1}{2}} U^T P Z.$$

Durch

$$X_p = X_{z,p} + \bar{X}$$

findet man somit eine Darstellung  $X_p$  der ursprünglichen Daten durch lediglich die gewünschten unabhängigen Komponenten.

Anhand der beiden folgenden Grafiken ist erkenntlich, dass die neue Darstellung des Datensatzes aus Abbildung 4.13 durch weniger unabhängige Komponenten in Abbildung 4.14

den Charakter der länglichen Verteilung des ursprünglichen Datensatzes besser erkennbar werden lässt. Die Darstellung durch weniger unabhängige Komponenten entspricht in diesem Fall Punkten auf einer Geraden. Dies erscheint nachvollziehbar, da im Grunde zwei Richtungen für das Aussehen des ursprünglichen Datensatzes verantwortlich sind: die Längsrichtung, welche in der Abbildung 4.14 zu sehen ist, und die Querrichtung, welche annähernd normal dazu steht. Die unabhängigen Komponenten entsprechen prinzipiell diesen Richtungen. Dementsprechend liegen die Punkte der Darstellung durch lediglich eine unabhängige Komponente auf einer dieser Richtungen - also einer Geraden.

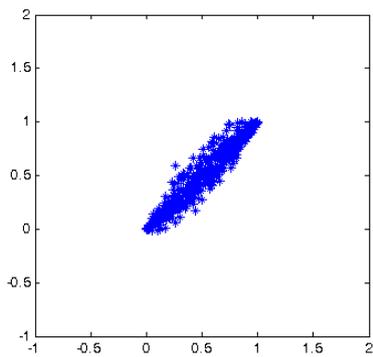


Abb. 4.13: Zweidimensionaler Datensatz

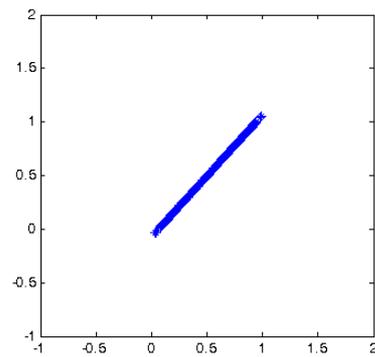


Abb. 4.14: Darstellung der Daten durch weniger unabhängige Komponenten

# 5 Verbesserung des Standard-Algorithmus von Independent Component Analysis (ICA)

## 5.1 Motivation

Die beiden wichtigsten Schritte des im Kapitel 4.2.7 besprochenen Standard-Algorithmus zur Bestimmung der unabhängigen Komponenten sind die Messung der Gaußianität und die Minimierung der Zielfunktion. Genau in diesen beiden Schritten kann man Schwachpunkte im Standard-Algorithmus finden, ebenso wie hier Verbesserungsmöglichkeiten ansetzen können.

Während das Verfahren bereits Anfang der 1980er Jahre entwickelt wurde<sup>1</sup>, sind die Verbesserungsmöglichkeiten desselben aktuelle Forschungsthemen. Aufgrund der intensiven Forschung wurde 2001 ein Buch zum Thema Independent Component Analysis von finnischen Forschern veröffentlicht (siehe <sup>2</sup>).

Noch aktueller ist die zur Zeit von Herrn a.o. Univ.-Prof. Mag. Dr. Keeling an der Karl Franzens Universität Graz durchgeführte Forschung. Durch seine Forschungsarbeit ist es nun möglich, in diesem Kapitel die neueste, von ihm entwickelte Verbesserung vorzustellen.

---

<sup>1</sup>vgl. Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis , S. 11

<sup>2</sup>Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis

## 5.2 Die Messung der Gaußianität

### 5.2.1 Schwachpunkt der Wölbung

Im vorherigen Kapitel wurde die Gaußianität mittels der Zielfunktion

$J_1(\mathbf{u}_i) = -W(\mathbf{Y}^T \mathbf{u}_i)^2$  gemessen. Diese Funktion hat einen entscheidenden Schwachpunkt:

Die Wölbung ist nicht robust gegenüber Ausreißern.

Was genau das bedeutet, wird am folgenden Beispiel verdeutlicht:

**Beispiel 5.2.1.** Die beiden Datenvektoren  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$  seien gewählt als die Koordinaten der gewählten Punkte aus der Menge  $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}$  mit der selben Häufigkeit - ausgenommen von 2 Ausreißern, nämlich:

$$\mathbf{x}_1(1) = \alpha, \quad \mathbf{x}_2(1) = 0$$

$$\mathbf{x}_1(2) = -\alpha, \quad \mathbf{x}_2(2) = 0.$$

Der Rest der Werte in  $\mathbf{x}_1$  und  $\mathbf{x}_2$  erfüllt:

$$\mathbf{x}_1(3) = 0, \quad \mathbf{x}_2(3) = 1$$

$$\mathbf{x}_1(4) = 0, \quad \mathbf{x}_2(4) = -1$$

und

$$\overline{\mathbf{x}_1} = \overline{\mathbf{x}_2} = 0. \tag{5.1}$$

Damit ist nach Definition 3.2.1.:

$$\begin{aligned} \sigma^2(\mathbf{x}_1) &= \frac{1}{n} \left( \underbrace{2\alpha^2}_{i=1,2} + \underbrace{0}_{i=3,4} + \sum_{i=5}^n \mathbf{x}_1(i)^2 \right) = \frac{1}{n} \left( 2\alpha^2 + \frac{1}{2}(n-4) \right) = \frac{1}{2} \left( 1 + \frac{4}{n}(\alpha^2 - 1) \right) \\ \sigma^2(\mathbf{x}_2) &= \frac{1}{n} \left( \underbrace{0}_{i=1,2} + \underbrace{2}_{i=3,4} + \sum_{i=5}^n \mathbf{x}_2(i)^2 \right) = \frac{1}{n} \left( 2 + \frac{1}{2}(n-4) \right) = \frac{1}{2} \end{aligned} \tag{5.2}$$

Da wegen (5.1)  $X_z = X$  gilt, ist die Kovarianzmatrix von  $X_z$  nach der Definition 3.2.3. gegeben als

$$K = \frac{1}{n} X X^T.$$

Aufgrund der Wahl der Daten in diesem Beispiel ist immer eine Koordinate 0, deshalb muss nach der Definition 3.2.1 die Korrelation  $\chi(\mathbf{x}_1, \mathbf{x}_2)$  von  $\mathbf{x}_1$  und  $\mathbf{x}_2$  0 sein. Aus diesem Grund ist die Kovarianzmatrix bereits diagonal und hat deshalb die Eigenraumzerlegung:

$$\begin{aligned} KI &= I\Lambda \\ \iff K &= \Lambda. \end{aligned}$$

Wobei  $V$  und  $\Lambda$  somit gegeben sind als:

$$\begin{aligned} V &= I \\ \Lambda &= \begin{pmatrix} \sigma^2(\mathbf{x}_1) & 0 \\ 0 & \sigma^2(\mathbf{x}_2) \end{pmatrix}. \end{aligned} \tag{5.3}$$

Nach Kapitel 3.2.6 sind die gesphärten Daten gegeben als:

$$Y = \Lambda^{-\frac{1}{2}} V X_z.$$

Dies vereinfacht sich mittels (5.3) zu:

$$Y = \begin{pmatrix} \left( \begin{pmatrix} \mathbf{x}_1 \\ \sigma(\mathbf{x}_1) \end{pmatrix} \right)^T \\ \left( \begin{pmatrix} \mathbf{x}_2 \\ \sigma(\mathbf{x}_2) \end{pmatrix} \right)^T \end{pmatrix}.$$

(Beispiel-Quelle: <sup>3</sup>)

Zur Bestimmung der unabhängigen Komponenten in  $Z$  des Datensatzes sucht man wie im Kapitel 4.2.7 eine Drehungsmatrix  $U$  mit

$$Z = UY \quad \wedge \quad UU^T = I.$$

Wie im Kapitel 4.2.7 wird diese Matrix  $U$  wieder zeilenweise bestimmt, wobei wie im Kapitel 4.2.7 der Vektor  $\mathbf{u}_i$  die  $i$ -te Zeile von  $U$  beinhaltet. Zur Berechnung dieser  $\mathbf{u}_i$  mit  $1 \leq i \leq 2$  werden diese Vektoren als Minimum der Zielfunktion bestimmt da, wie in Kapitel 4.2.4 erklärt, die gemessene Gaußianität am wenigsten sein sollte. Als Zielfunktion dafür wurde bisher verwendet:

$$J_1(\mathbf{u}_i) = -W(Y^T \mathbf{u}_i)^2 = - (M_4(Y^T \mathbf{u}_i) - 3M_2(Y^T \mathbf{u}_i)^2)$$

---

<sup>3</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten, S. 54 ff

In dem gewählten Beispiel ist nun:

$$\begin{aligned}
M_2(Y^T \mathbf{u}_i) &= \frac{1}{n} \sum_{j=1}^n (e_{n,j}^T Y^T \mathbf{u}_i)^2 = \\
&= \frac{1}{n} (Y^T \mathbf{u}_i)^T Y^T \mathbf{u}_i = \\
&= \frac{1}{n} \mathbf{u}_i^T Y Y^T \mathbf{u}_i = \\
&\stackrel{(3.6)}{=} \mathbf{u}_i^T \mathbf{u}_i \\
&\stackrel{(4.2)}{\iff} M_2(\mathbf{u}_i) = 1
\end{aligned}$$

Die Zeilen von  $Y$  sind wie im Kapitel 3.2.3 die Hauptkomponenten  $\mathbf{y}_i$  mit  $1 \leq i \leq 2$ .

Deshalb ist in diesem Beispiel:

$$\begin{aligned}
M_4(Y^T \mathbf{u}_i) &= \frac{1}{n} \sum_{j=1}^n (e_{n,j}^T Y^T \mathbf{u}_i)^4 = \\
&= \frac{1}{n} \sum_{j=1}^n \left( \frac{\mathbf{y}_1(j)}{\sigma(\mathbf{y}_1)} \mathbf{u}_i(1) + \frac{\mathbf{y}_2(j)}{\sigma(\mathbf{y}_2)} \mathbf{u}_i(2) \right)^4 = \\
&= \frac{1}{n} \left( \sum_{j=1}^2 \left( \frac{\mathbf{y}_1(j)}{\sigma(\mathbf{y}_1)} \mathbf{u}_i(1) \right)^4 + \sum_{j=3}^4 \left( \frac{\mathbf{y}_2(j)}{\sigma(\mathbf{y}_2)} \mathbf{u}_i(2) \right)^4 \right) + \\
&+ \frac{1}{n} \left( \sum_{j=5}^n \left( \frac{\mathbf{y}_1(j)}{\sigma(\mathbf{y}_1)} \mathbf{u}_i(1) + \frac{\mathbf{y}_2(j)}{\sigma(\mathbf{y}_2)} \mathbf{u}_i(2) \right)^4 \right) = \\
&\stackrel{(5.2)}{=} \frac{1}{n} \left( \sum_{j=1}^2 \frac{\alpha^4}{\left( \frac{1}{\sqrt{2}} \sqrt{1 + \frac{4}{n}(\alpha^2 - 1)} \right)^4} (\mathbf{u}_i(1))^4 + \sum_{j=3}^4 \frac{1}{\left( \frac{1}{\sqrt{2}} \right)^4} (\mathbf{u}_i(2))^4 \right) + \\
&+ \frac{1}{n} \left( \sum_{j \geq 5, \mathbf{y}_2(j)=0}^n \frac{1}{\left( \frac{1}{\sqrt{2}} \sqrt{1 + \frac{4}{n}(\alpha^2 - 1)} \right)^4} (\mathbf{u}_i(1))^4 + \sum_{j \geq 5, \mathbf{y}_1(j)=0}^n \frac{1}{\left( \frac{1}{\sqrt{2}} \right)^4} (\mathbf{u}_i(2))^4 \right) \\
\iff M_4(Y^T \mathbf{u}_i) &= \frac{4}{n} \left( \frac{2\alpha^4 + \frac{1}{2}(n-4)}{\left(1 + \frac{4}{n}(\alpha^2 - 1)\right)^2} \mathbf{u}_i(1)^4 + \left(2 + \frac{1}{2}(n-4)\right) \mathbf{u}_i(2)^4 \right).
\end{aligned}$$

Setzt man

$$v_n(\alpha) = \frac{1 + \frac{4}{n}(\alpha^4 - 1)}{\left(1 + \frac{4}{n}(\alpha^2 - 1)\right)^2}$$

so erhält man<sup>4</sup>:

$$M_4(Y^T \mathbf{u}_i) = 2 \left( v_n(\alpha) (\mathbf{u}_i(1))^4 + (\mathbf{u}_i(2))^4 \right).$$

<sup>4</sup>vgl. Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis , S. 56

Damit wird die Zielfunktion für dieses Beispiel zu:

$$J_1(\mathbf{u}_i) = -4 \left( v_n(\alpha)(\mathbf{u}_i(1))^4 + (\mathbf{u}_i(2))^4 - \frac{3}{2} \right)^2.$$

Nach einer Drehung um  $\frac{\pi}{4}$  wären die Daten unabhängig, dies entspräche dem Vektor  $\mathbf{u}_i = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$  mit  $\theta = \frac{\pi}{4}$ .

*Beweis.* Damit die ursprünglichen Datenvektoren unabhängig wären, müsste nach Definition 4.2.3

$$P(\mathbf{x}_1(i) = a \cap \mathbf{x}_2(i) = b) = P(\mathbf{x}_1(i) = a)P(\mathbf{x}_2(i) = b) \quad (5.4)$$

für alle möglichen Werte von  $a, b$  gelten. Für  $a = b = 1$  ist dies in den ursprünglichen Datenvektoren nicht der Fall: In den gegebenen Daten ist für jeden Eintrag 1 im Vektor  $\mathbf{x}_1$  der entsprechende Eintrag im Vektor  $\mathbf{x}_2$  0. Deshalb ist die Häufigkeit auf der linken Seite  $P(\mathbf{x}_1(i) = 1 \cap \mathbf{x}_2(i) = 1) = 0$ . Im Gegensatz dazu sind ungefähr ein Viertel der Einträge in  $\mathbf{x}_1$  und  $\mathbf{x}_2$  1, darum ist  $P(\mathbf{x}_1(i) = 1) \approx \frac{1}{4}$  und  $P(\mathbf{x}_2(i) = 1) \approx \frac{1}{4}$  also ist

$$0 = P(\mathbf{x}_1(i) = 1 \cap \mathbf{x}_2(i) = 1) \neq P(\mathbf{x}_1(i) = 1)P(\mathbf{x}_2(i) = 1) \approx \frac{1}{16}.$$

Darum sind die ursprünglichen Daten nicht unabhängig. Werden diese allerdings um  $\frac{\pi}{4}$  gedreht, so entsprechen sie (ausgenommen den beiden Ausreißern) den Koordinaten der gewählten Punkte aus der Menge  $\left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right\}$  mit der selben Auswahlhäufigkeit. Nun ist leicht überprüfbar, dass (5.4) für alle möglichen Werte von  $a, b$  erfüllt ist und damit die, durch die Transformation entstehenden, Datenvektoren nach der Definition 4.2.3 tatsächlich unabhängig sind. Aus diesem Grund ist die natürliche Lösung für  $\mathbf{u}_i$  mit  $\theta = \frac{\pi}{4}$  gegeben.  $\square$

In der folgenden Abbildung 5.1 wird die erste Zielfunktion mit den Werten des obigen Beispiels gezeigt. Dabei ist die Zielfunktion in Abhängigkeit der möglichen Werte des Vektors  $\mathbf{u}_i = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$  mit  $\theta \in [0, 2\pi]$  und der möglichen Werte für  $\alpha \in [0, 6]$  für  $n = 2000$  dargestellt. Die natürliche Lösung für die den Drehungswinkel  $\theta$  und damit den Vektor  $\mathbf{u}_i$  wäre wie oben gezeigt  $\theta = \frac{\pi}{4}$ .

Abbildung 5.2 stellt die Zielfunktion mit  $\alpha = 6$  und  $n = 2000$  dar. Man erkennt deutlich, dass dort das Minimum bei  $\theta = 0$  liegt - mit dieser würden die Daten somit nicht gedreht.

Deshalb ist die Zielfunktion nicht robust gegenüber Ausreißern.

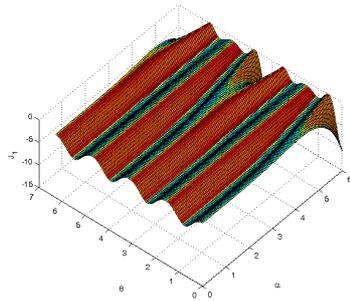


Abb. 5.1: Zielfunktion<sup>5</sup>  $J_1$  in Abhängigkeit von  $\theta$  und  $\alpha$

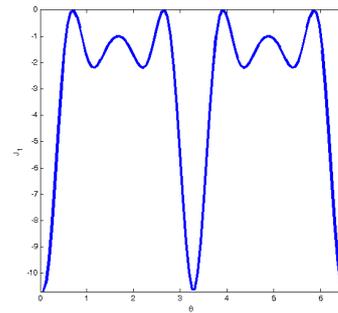


Abb. 5.2: Zielfunktion<sup>5</sup>  $J_1$  bei  $\alpha = 6$

## 5.2.2 Verbesserte Zielfunktion

Anhand des Beispiels im vorherigen Kapitel erkennt man, dass die erste Zielfunktion  $J_1(\mathbf{u}_i)$  nicht ideal zur Messung der Gaußianität geeignet ist, da sie nicht robust gegen Ausreißer ist. Aus diesem Grund sucht man nach einer alternativen Zielfunktion, welche robuster gegen solche Ausreißer ist. Eine solche alternative Zielfunktion ist durch die folgende gegeben:

$$J_2(\mathbf{u}_i) = - \left( M_1(Y^T \mathbf{u}_i) - \sqrt{\frac{2}{\pi}} M_2(Y^T \mathbf{u}_i) \right)^2 .$$

(Zielfunktions-Quelle: <sup>6</sup>)

Diese ist wie  $J_1$  aufgrund der in Satz 4.2.2. beschriebenen Grenzwerteigenschaft der Momente genau bei einem Datenvektor mit gauß-verteiltern Werten annähernd 0, also würde man auch hier durch eine Minimierung der Funktion eine Minimierung der Gaußianität erreichen. Am folgenden Beispiel wird verdeutlicht, dass die Funktion robuster ist:

<sup>5</sup>Vorbild für selbst erstellte Grafik: Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten , S.57

<sup>6</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten , S. 59

Erneut seien die Datenvektoren  $\mathbf{x}_1$  und  $\mathbf{x}_2$  wie im Beispiel 5.2.1 gegeben. Es ist hier nun:

$$\begin{aligned}
M_1(Y^T \mathbf{u}_i) &= \frac{1}{n} \sum_{j=1}^n |e_{n,j}^T Y^T \mathbf{u}_i| = \\
&= \frac{1}{n} \left( \sum_{j=1}^2 \left| \frac{\mathbf{y}_1(j)}{\sigma(\mathbf{y}_1)} \mathbf{u}_i(1) \right| + \sum_{j=3}^4 \left| \frac{\mathbf{y}_2(j)}{\sigma(\mathbf{y}_2)} \mathbf{u}_i(2) \right| \right) \\
&+ \frac{1}{n} \left( \sum_{j=5}^n \left| \frac{\mathbf{y}_1(j)}{\sigma(\mathbf{y}_1)} \mathbf{u}_i(1) + \frac{\mathbf{y}_2(j)}{\sigma(\mathbf{y}_2)} \mathbf{u}_i(2) \right| \right) = \\
&= \frac{1}{n} \left( \sum_{j=1}^2 \frac{|\alpha|}{\left| \frac{1}{\sqrt{2}} \sqrt{1 + \frac{4}{n}(\alpha^2 - 1)} \right|} |\mathbf{u}_i(1)| + \sum_{j=3}^4 \frac{1}{\frac{1}{\sqrt{2}}} |\mathbf{u}_i(2)| \right) + \\
&+ \frac{1}{n} \left( \sum_{j \geq 5, \mathbf{y}_2(j)=0} \frac{1}{\left| \frac{1}{\sqrt{2}} \sqrt{1 + \frac{4}{n}(\alpha^2 - 1)} \right|} |\mathbf{u}_i(1)| + \sum_{j \geq 5, \mathbf{y}_1(j)=0} \frac{1}{\frac{1}{\sqrt{2}}} |\mathbf{u}_i(2)| \right) = \\
&= \frac{\sqrt{2}}{n} \left( \frac{2|\alpha| + \frac{1}{2}(n-4)}{\sqrt{1 + \frac{4}{n}(\alpha^2 - 1)}} |\mathbf{u}_i(1)| + \left( 2 + \frac{1}{2}(n-4) \right) |\mathbf{u}_i(2)| \right) \\
\iff M_1(Y^T \mathbf{u}_i) &= \frac{1}{\sqrt{2}} \left( \frac{1 + \frac{4}{n}(|\alpha| - 1)}{\sqrt{1 + \frac{4}{n}(\alpha^2 - 1)}} |\mathbf{u}_i(1)| + |\mathbf{u}_i(2)| \right).
\end{aligned}$$

Setzt man nun

$$\nu_n(\alpha) = \frac{1 + \frac{4}{n}(|\alpha| - 1)}{\sqrt{1 + \frac{4}{n}(\alpha^2 - 1)}}$$

so erhält man<sup>7</sup>:

$$M_1(Y^T \mathbf{u}_i) = \frac{1}{\sqrt{2}} (\nu_n(\alpha) |\mathbf{u}_i(1)| + |\mathbf{u}_i(2)|).$$

Dies bedeutet für die neue Zielfunktion  $J_2$  in diesem Beispiel:

$$J_2(\mathbf{u}_i) = -\frac{1}{2} \left( \nu_n(\alpha) |\mathbf{u}_i(1)| + |\mathbf{u}_i(2)| - \frac{2}{\sqrt{\pi}} \right)^2.$$

In der folgenden Abbildung 5.3 wird die neue Zielfunktion mit den Werten des obigen Beispiels gezeigt. Dabei ist die Zielfunktion in Abhängigkeit der möglichen Werte des Vektors  $\mathbf{u}_i = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$  mit  $\theta \in [0, 2\pi]$  und der möglichen Werte für  $\alpha \in [0, 6]$  für  $n = 2000$  dargestellt. Die natürliche Lösung für den Drehungswinkel  $\theta$  und damit den Vektor  $\mathbf{u}_i$  wäre, wie bereits vorhin bewiesen,  $\theta = \frac{\pi}{4}$ .

<sup>7</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten, S. 58

Abbildung 5.4 stellt die Zielfunktion mit  $\alpha = 6$  und  $n = 2000$  dar. Man erkennt deutlich, dass dort das Minimum noch immer bei  $\theta = \frac{\pi}{4}$  liegt - mit dieser werden die Daten somit richtig gedreht. Deshalb scheint die Zielfunktion robust gegenüber Ausreißern zu sein.

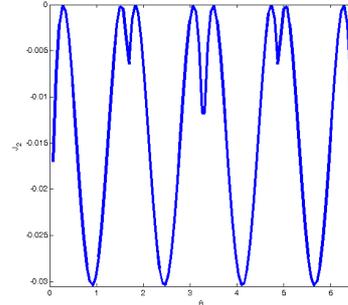
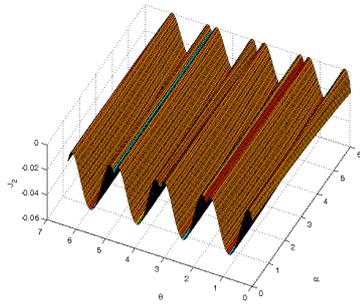


Abb. 5.3: Zielfunktion<sup>8</sup>  $J_2$  in Abhängigkeit von  $\theta$  und  $\alpha$

Abb. 5.4: Zielfunktion<sup>8</sup>  $J_2$  bei  $\alpha = 6$

Deshalb ist es sinnvoll anstatt der Zielfunktion  $J_1(\mathbf{u}_i)$  die Zielfunktion

$$J_2(\mathbf{u}_i) = - \left( M_1(Y^T \mathbf{u}_i) - \sqrt{\frac{2}{\pi}} M_2(Y^T \mathbf{u}_i) \right)^2$$

für die Messung und Minimierung der Gaußianität zu verwenden.

Es ist allerdings

$$\begin{aligned} M_2(Y^T \mathbf{u}_i) &= \frac{1}{m} \sum_{j=1}^m (e_{m,j}^T Y^T \mathbf{u}_i)^2 = \\ &= \frac{1}{n} (Y^T \mathbf{u}_i)^T Y^T \mathbf{u}_i = \\ &= \frac{1}{n} \mathbf{u}_i^T Y Y^T \mathbf{u}_i = \\ &\stackrel{(3.6)}{=} \mathbf{u}_i^T \mathbf{u}_i \end{aligned}$$

$$\stackrel{(4.2)}{\iff} M_2(\mathbf{u}_i) = 1.$$

Somit vereinfacht sich die neue Zielfunktion  $J_2((\mathbf{u})_i)$  zu

$$J_2(\mathbf{u}_i) = - \left( M_1(Y^T \mathbf{u}_i) - \sqrt{\frac{2}{\pi}} \right)^2 .$$

<sup>8</sup>Vorbild für selbst erstellte Grafik: Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten , S.59

## 5.3 Minimierung der Zielfunktion

### 5.3.1 Schwachpunkte des Gradienten-Abstiegsverfahrens

Bearbeitet man bereits vollständig vorhandene<sup>9</sup> Daten, so ist ein entscheidender Nachteil des Gradienten-Abstiegsverfahrens, dass dessen Konvergenz zu einem Minimum langsam ist<sup>10</sup>. Aus diesem Grund sucht man nach einer alternativen Optimierungsmethode mit schnellerer Konvergenz.

### 5.3.2 Verbesserte Minimierung

Eine solche alternative Optimierungsmethode ist ein sogenannter Fixed-Point-Algorithmus<sup>11</sup>. Man bedient sich dabei der Lagrange'schen Multiplikatorenregel. Diese besagt:

**Satz 5.3.1.** *Man bezeichnet mit  $L(\mathbf{x}, \lambda) = f + \sum_{j=1}^k \lambda(j)g_j$  die Lagrange-Funktion zu einem Optimierungsproblem*

$$\min_{\mathbf{x} \in D} f(\mathbf{x})$$

mit einer Zielfunktion  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  und  $k$  Nebenbedingungen

$$D = \{\mathbf{x} \in \mathbb{R}^n \mid g_j(\mathbf{x}) = 0, j = 1, \dots, k\}.$$

Ist  $\hat{\mathbf{x}} \in D$  ein lokales Minimum der Funktion  $f$  auf der Menge  $D$  und sind die  $k$  Gradienten

$$D_{\mathbf{x}}g_1(\hat{\mathbf{x}}), D_{\mathbf{x}}g_2(\hat{\mathbf{x}}), \dots, D_{\mathbf{x}}g_n(\hat{\mathbf{x}})$$

an der Stelle  $\hat{\mathbf{x}}$  linear unabhängig, dann existieren Lagrange'sche Multiplikatoren

$\hat{\lambda}(1), \hat{\lambda}(2), \dots, \hat{\lambda}(k) \in \mathbb{R}$ , sodass die  $n + k$  Gleichungen mit

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{x}^{(i)}}(\hat{\mathbf{x}}, \hat{\lambda}) &= 0, \quad i = 1, \dots, n \\ \frac{\partial L}{\partial \lambda^{(j)}}(\hat{\mathbf{x}}, \hat{\lambda}) &= 0, \quad j = 1, \dots, k \end{aligned}$$

erfüllt sind. In diesem Fall heißt  $\hat{\mathbf{x}}$  stationärer Punkt der Lagrange-Funktion  $L(\mathbf{x}, \lambda)$ .

(Satz-Quelle: <sup>12</sup>)

<sup>9</sup>Im Gegensatz zu ständig neu ankommenden Daten

<sup>10</sup>vgl. Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis, S. 178

<sup>11</sup>Fixed-Point ist die englische Bezeichnung für einen stationären Punkt

<sup>12</sup>vgl. Arens, Tilo/ Hettlich, Frank/ Karpfinger, Christian/ et al: Mathematik, S. 1205

Dieser Satz bietet also die Möglichkeit, eine Funktion unter Nebenbedingungen zu minimieren. In der Unabhängigkeitsanalyse möchte man nun die neue Zielfunktion

$$J_2(\mathbf{u}_i) = - \left( M_1(Y^T \mathbf{u}_i) - \sqrt{\frac{2}{\pi}} \right)^2$$

unter der Nebenbedingung  $\mathbf{u}_i^T \mathbf{u}_i - 1 = 0$  (aufgrund von (4.2)) minimieren.

Nach dem Satz 5.3.1 entspricht ein Minimum von  $J_2$  unter dieser Nebenbedingung einem stationären Punkt der folgenden Lagrange-Funktion:

$$L(\mathbf{u}_i, \lambda) = J_2(\mathbf{u}_i) + \lambda(1)(\mathbf{u}_i^T \mathbf{u}_i - 1). \quad (5.5)$$

Für einen stationären Punkt  $\hat{\mathbf{u}}_i$  dieser Lagrange-Funktion existiert nach Satz 5.3.1 ein  $\hat{\lambda}$ , sodass

$$D_{\mathbf{u}_i} L(\hat{\mathbf{u}}_i) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (5.6)$$

Es ist nun

$$D_{\mathbf{u}_i} L(\hat{\mathbf{u}}_i) = D_{\mathbf{u}_i} J_2(\hat{\mathbf{u}}_i) + \hat{\lambda}(1)2\hat{\mathbf{u}}_i \quad (5.7)$$

wobei die Zielfunktion  $J_2(\hat{\mathbf{u}}_i)$  den folgenden Gradienten hat:

$$D_{\mathbf{u}_i} J_2(\hat{\mathbf{u}}_i) = -2 \left( M_1(Y^T \hat{\mathbf{u}}_i) - \sqrt{\frac{2}{\pi}} \right) D_{\mathbf{u}_i} M_1(Y^T \hat{\mathbf{u}}_i). \quad (5.8)$$

Schätzt man  $M_1(Y^T \hat{\mathbf{u}}_i)$  mit  $\epsilon \ll 1$  als:

$$M_1(Y^T \hat{\mathbf{u}}_i) \approx \frac{1}{m} \sum_{j=1}^m ((e_{m,j}^T Y^T \hat{\mathbf{u}}_i)^2 + \epsilon^2)^{\frac{1}{2}}$$

so erhält man<sup>13</sup> für den Gradienten von  $M_1(Y^T \mathbf{u}_i)$ :

$$D_{\mathbf{u}_i} M_1(Y^T \hat{\mathbf{u}}_i) \approx \frac{1}{m} \sum_{j=1}^m ((e_{m,j}^T Y^T \hat{\mathbf{u}}_i)^2 + \epsilon^2)^{-\frac{1}{2}} (e_{m,j}^T Y^T \hat{\mathbf{u}}_i) (Y e_{m,j}).$$

---

<sup>13</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten, S. 60

Sei ab nun:

$$\phi(\hat{\mathbf{u}}_i) = 2 \left( M_1(Y^T \hat{\mathbf{u}}_i) - \sqrt{\frac{2}{\pi}} \right)$$

und

$$G(\hat{\mathbf{u}}_i) = \frac{1}{m} \sum_{j=1}^m ((e_{m,j} Y^T \hat{\mathbf{u}}_i)^2 + \epsilon^2)^{-\frac{1}{2}} (Y e_{m,j})(e_{m,j}^T Y^T). \quad (5.9)$$

Dann ist  $\phi(\hat{\mathbf{u}}_i) \in \mathbb{R}$  ein Skalar und  $G(\hat{\mathbf{u}}_i) \in \mathbb{R}^{n \times n}$  eine Matrix der Größe  $n \times n$ . Wie leicht anhand Verdeutlichung der Inhalte in den folgenden Vektoren und Matrizen erkenntlich ist gilt

$$(e_{m,j}^T Y^T \hat{\mathbf{u}}_i)(Y e_{m,j}) = (Y e_{m,j})(e_{m,j}^T Y^T) \hat{\mathbf{u}}_i.$$

Aus diesem Grund und wegen (5.8) lässt sich  $D_{\mathbf{u}_i} J_2(\hat{\mathbf{u}}_i)$  anders darstellen als<sup>14</sup>

$$\phi(\hat{\mathbf{u}}_i) G(\hat{\mathbf{u}}_i) \hat{\mathbf{u}}_i = -D_{\mathbf{u}_i} J_2(\hat{\mathbf{u}}_i).$$

Nach (5.6) und (5.7) ist nun:

$$-D_{\mathbf{u}_i} J_2(\hat{\mathbf{u}}_i) = \hat{\lambda} \hat{\mathbf{u}}_i.$$

Wählt man nun eine Zahl  $\hat{\mu}(\hat{\mathbf{u}}_i)$  sodass  $\hat{\lambda}(1) = \hat{\mu}(\hat{\mathbf{u}}_i) \phi(\hat{\mathbf{u}}_i)$ , dann ist dies äquivalent zu

$$G(\hat{\mathbf{u}}_i) \hat{\mathbf{u}}_i = \hat{\mu}(\hat{\mathbf{u}}_i) \hat{\mathbf{u}}_i$$

wobei dies erneut unter der Nebenbedingung  $\hat{\mathbf{u}}_i^T \hat{\mathbf{u}}_i - 1 = 0$  gelten soll.

Dies entspricht einem (nicht linearen) Eigenraumproblem, ähnlich zu dem in Kapitel 3.2.5 besprochenen<sup>15</sup>. Die Lösung eines solchen Problem es ist nach der Definition 3.2.5 bekannt. Zur Bestimmung von  $\mathbf{u}_i$  sucht man also nach einem Eigenvektor der Matrix  $G(\hat{\mathbf{u}})$ . Ein solcher Eigenvektor wird durch eine Vektoriteration bestimmt<sup>16</sup>. Diese Vektoriteration ist durch Computer deutlich schneller berechenbar als das im Standard-Algorithmus der Unabhängigkeitsanalyse verwendete Gradienten-Abstiegsverfahren. Aus diesem Grund ist es sinnvoll, die neue Zielfunktion durch diese Methode zu minimieren.

<sup>14</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten , S. 61

<sup>15</sup>vgl. ebda, S. 61

<sup>16</sup>vgl. Keeling, Stephen: Grundlagen der Numerischen Mathematik , S. 100

## 5.4 Verbesserter Algorithmus zur Bestimmung der unabhängigen Komponenten

Mithilfe dieses Wissens kann man die Werte in  $U$  und damit die möglichst unabhängigen Komponenten durch den folgenden Algorithmus bestimmen<sup>17</sup>:

Zunächst wird die erste Zeile  $\mathbf{u}_1$  in  $U$  bestimmt durch:

1. Es wird mit einer ersten Schätzung des Vektors  $\mathbf{u}_1$  begonnen, welche den Betrag 1 besitzt (aufgrund von (4.2)).
2. Der Vektor wird aktualisiert durch die Multiplikation der durch (5.9) gegebenen Matrix:

$$\mathbf{u}_{1neu} \leftarrow G(\mathbf{u}_{1alt})\mathbf{u}_{1alt}.$$

3. Der Vektor wird erneut aufgrund von (4.2) normalisiert durch:

$$\mathbf{u}_{1neu} \leftarrow \frac{\mathbf{u}_{1neu}}{\|\mathbf{u}_{1neu}\|}.$$

4. Ist  $\frac{\mathbf{u}_{1alt}^T \mathbf{u}_{1neu}}{\|\mathbf{u}_{1alt}\| \|\mathbf{u}_{1neu}\|} < 0$ , so wird, damit der Betrag im Schritt 5. tatsächlich gegen 0 konvergieren kann,  $\mathbf{u}_{1neu}$  ersetzt durch

$$\mathbf{u}_{1neu} \leftarrow -\mathbf{u}_{1neu}.$$

5. Die Schritte 2. - 4. werden wiederholt, bis der Betrag  $\|\mathbf{u}_{1alt} - \mathbf{u}_{1neu}\|$  annähernd 0 ist, also der Vektor konvergiert. Die Konvergenz tritt tatsächlich ein<sup>18</sup>.

Die nächsten Zeilen  $\mathbf{u}_i$  mit  $2 \leq i \leq n - 1$  werden durch einen etwas abgeänderten Algorithmus bestimmt<sup>19</sup>:

1. Erneut wird mit einer ersten Schätzung des Vektors  $\mathbf{u}_i$  begonnen, welche den Betrag 1 besitzt und orthogonal zu allen bereits bestimmten Vektoren  $\mathbf{u}_j$  mit  $1 \leq j \leq i - 1$  ist (aufgrund von (4.2)).

---

<sup>17</sup>vgl. Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten, S. 61 f

<sup>18</sup>vgl. ebda, S. 62

<sup>19</sup>vgl. ebda, S. 62

2. Sei die Matrix  $U_{i-1}$  jene, welche die bereits bestimmten Vektoren  $\mathbf{u}_j$  mit  $1 \leq j \leq i-1$  beinhaltet, also

$$U_{i-1} = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_{i-1}^T \end{pmatrix}.$$

Seien dann die projizierten Daten  $Y_p$  bestimmt als

$$Y_p = (I - U_{i-1}^T U_{i-1})Y$$

und damit die Matrix  $\tilde{G}$  berechnet als:

$$\tilde{G}(\mathbf{u}_{i_{alt}}) = \frac{1}{m} \sum_{j=1}^m ((e_{m,j} Y_p^T \mathbf{u}_{i_{alt}})^2 + \epsilon^2)^{-\frac{1}{2}} (Y_p e_{m,j}) (e_{m,j}^T Y_p^T).$$

Dann wird der Vektor  $\mathbf{u}_i$  aktualisiert durch:

$$\mathbf{u}_{i_{neu}} \leftarrow \left( U_{i-1}^T U_{i-1} + \tilde{G}(\mathbf{u}_{i_{alt}}) \right) \mathbf{u}_{i_{alt}}.$$

3. Der neue Vektor  $\mathbf{u}_{i_{neu}}$  wird in den zu den bereits berechneten Vektoren  $\mathbf{u}_j$  mit  $1 \leq j \leq i-1$  normalen Raum aufgrund von (4.2) projiziert durch

$$\mathbf{u}_{i_{neu}} \leftarrow \mathbf{u}_{i_{neu}} - \frac{\mathbf{u}_j^T \mathbf{u}_{i_{neu}}}{\|\mathbf{u}_j\|^2} \mathbf{u}_j$$

für  $j \in \{1, 2, \dots, i-1\}$ .

4. Der Vektor wird aufgrund von (4.2) normalisiert durch

$$\mathbf{u}_{i_{neu}} \leftarrow \frac{\mathbf{u}_{i_{neu}}}{\|\mathbf{u}_{i_{neu}}\|}.$$

5. Ist  $\frac{\mathbf{u}_{i_{alt}}^T \mathbf{u}_{i_{neu}}}{\|\mathbf{u}_{i_{alt}}\| \|\mathbf{u}_{i_{neu}}\|} < 0$ , so wird, damit der Betrag im Schritt 6. tatsächlich gegen 0 konvergieren kann,  $\mathbf{u}_{i_{neu}}$  ersetzt durch

$$\mathbf{u}_{i_{neu}} \leftarrow -\mathbf{u}_{i_{neu}}.$$

6. Die obigen Schritte 2. - 5. werden so lange wiederholt bis der Betrag  $\|\mathbf{u}_{i_{alt}} - \mathbf{u}_{i_{neu}}\|$  annähernd 0 ist, das heißt, der Vektor  $\mathbf{u}_i$  konvergiert.

Der letzte Vektor  $\mathbf{u}_n$  ist aufgrund von (4.2) normal zu allen bereits bestimmten und hat zusätzlich den Betrag 1. Die Orientierung kann beliebig gewählt werden und damit kann dieser sehr leicht bestimmt werden.

Dadurch sind nun die Matrix  $U$  und damit die möglichst unabhängigen Komponenten in  $Z$  des Datensatzes  $X_z$  berechnet mit

$$Z = UY = UWX_z.$$

Damit wurde ein schnellerer und robusterer Algorithmus als der Standard-Algorithmus der Unabhängigkeitsanalyse gefunden. Die ursprünglichen Daten können ab hier, gleich wie im Kapitel 4.2.8, rekonstruiert und durch weniger unabhängige Komponenten dargestellt werden.

In den folgenden Grafiken erkennt man den sehr deutlichen Unterschied zwischen dem Standard- und dem verbesserten Algorithmus der Unabhängigkeitsanalyse. In Abbildung 5.5 werden die beiden gedrehten Koordinatensysteme gezeigt, anhand derer die Hauptkomponenten in der Unabhängigkeitsanalyse gedreht werden, sodass die neuen Koordinatensysteme nach der Drehung dem natürlichen entsprechen. Die strichlierten Achsen wurden durch den Standard-, die durchgehenden durch den verbesserten Algorithmus bestimmt. In den Abbildungen 5.6 und 5.7 werden die gedrehten Daten - die unabhängigen Komponenten - nach der Berechnung durch die beiden unterschiedlichen Algorithmen gezeigt.

Die ursprünglichen Daten waren diamantförmig im Streudiagramm mit 2 Ausreißern angeordnet. Die Struktur und die Ausreißer sind auch nach dem Sphären erhalten geblieben (vgl. Abbildung 5.5). Solange die Daten allerdings diese diamantförmige Struktur besitzen, sind sie nicht unabhängig. Die durch den Standard-Algorithmus von ICA bestimmten unabhängigen Komponenten sind deshalb nicht unabhängig - sie werden aufgrund der Ausreißer nicht korrekt bestimmt. Jene unabhängigen Komponenten, welche durch den verbesserten Algorithmus bestimmt werden, bilden allerdings annähernd ein Quadrat im Streudiagramm, was bedeutet, dass sie unabhängig sind - somit entsteht eine deutliche Verbesserung durch die Verwendung des neuen Algorithmus von ICA.

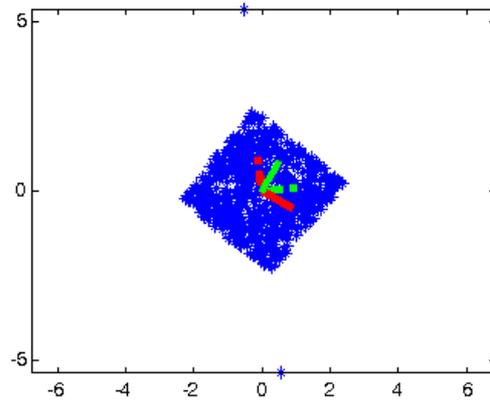


Abb. 5.5: Gesphärte Daten mit gedrehten Koordinatensystemen

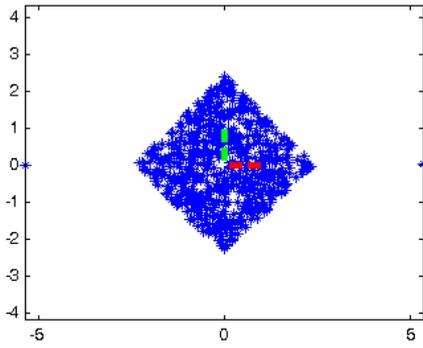


Abb. 5.6: Durch Standard-Algorithmus bestimmte unabhängige Komponenten

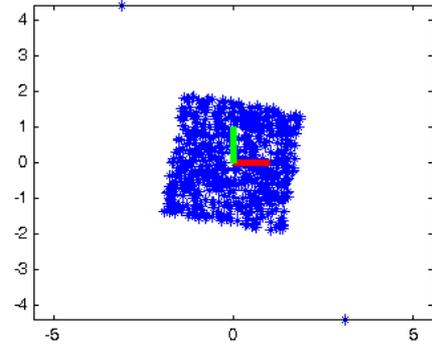


Abb. 5.7: Durch verbesserten Algorithmus bestimmte unabhängige Komponenten

# 6 Visualisierung der Verfahren

## 6.1 Motivation

In den vorherigen Abschnitten der Arbeit wurden die Verfahren PCA und ICA aus theoretischer Sicht vorgestellt und behandelt. In diesem letzten Teil der Arbeit soll nun anhand eines Beispiels gezeigt werden, wie in der Realität mit den beiden Verfahren gearbeitet werden kann.

**Beispiel 6.1.1.** *Mithilfe einer Magnetresonanztomographie wurden die Organe im Bauchraum (Abdomen) aufgezeichnet und in einem Video dargestellt. Deutlich sind die Organe, die Wirbelsäule und die äußeren Körpergrenzen zu erkennen. Während der Aufnahme kommt es durch unwillkürliche Bewegungen des Patienten (Atmung, Herzschlag, etc.) zu Bewegungsartefakten.*

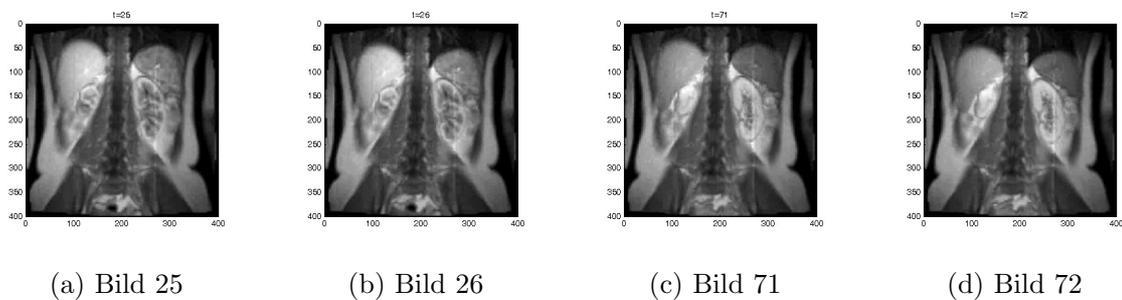


Abb. 6.1: Ausschnitte aus dem originalen Kontrastvideo<sup>1</sup>

*Das Ziel bei der Anwendung der Verfahren PCA und ICA soll es sein die durch Bewegung verursachten Intensitätsänderungen zu entfernen, um die Beurteilung des Videos zu verbessern.*

---

<sup>1</sup>Medizinische Aufnahme von Herrn a.o. Univ.-Prof. Mag. Dr. Keeling erhalten, Bilder daraus selbst erstellt

## 6.2 Das Modell in der Anwendung

Das Video im Beispiel besteht aus 134 Bildern mit einer jeweiligen Auflösung von  $400 \times 400$  Bildpunkten. Jeder dieser Bildpunkte besitzt einen skalaren Wert, der die Graustufe angibt, in welcher der Punkt bei der Anzeige erscheint. Es ist nun aufgrund der Dimensionen des Videos sinnvoll den ursprünglichen Datensatz für die weitere Verarbeitung des Videos als den folgenden zu wählen:

Es seien die ursprünglichen Datenvektoren  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{134} \in \mathbb{R}^{160000}$  die 134 Bilder des Videos. Dabei werden die Graustufenwerte der einzelnen Bildpunkte wie folgt in die Datenvektoren verpackt: Beginnend von links oben werden die Werte nach der Regel "links vor rechts und oben vor unten" (ähnlich wie das Lesen eines deutschsprachigen Buches) in den entsprechenden Datenvektor eingeschrieben. Somit entspricht das Bild in der Abbildung 6.1a dem Vektor  $\mathbf{x}_1$ , das Bild in der Abbildung 6.1b dem Vektor  $\mathbf{x}_{40}$  und so weiter.

## 6.3 Anwendung von PCA

Nachdem die ursprünglichen Daten in  $X$  gegeben sind, können die Hauptkomponenten in  $Y$  des Datensatzes, wie in Kapitel 3.2.6 besprochen, bestimmt werden durch:

$$Y = \Lambda^{-\frac{1}{2}} V^T X_Z.$$

Die Eigenwerte in  $\Lambda$  geben Aufschluss darüber, wie stark die entsprechende Hauptkomponente im ursprünglichen Video vertreten ist.

In der Abbildung 6.2 wird der Wert des natürlichen Logarithmus der Eigenwerte dargestellt. Die horizontale Linie entspricht einer Grenze von 1‰ des größten Eigenwertes. Lediglich die größten 7 Eigenwerte liegen über dieser Grenze. Deshalb genügt es die 7 am stärksten vertretenen Hauptkomponenten für weitere Berechnungen zu verwenden.

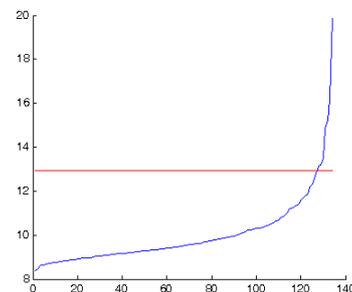


Abb. 6.2: Logarithmus der Eigenwerte in

$\Lambda$

Sei deshalb ab nun:

$$Y = \begin{pmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \mathbf{y}_3^T \\ \mathbf{y}_4^T \\ \mathbf{y}_5^T \\ \mathbf{y}_6^T \\ \mathbf{y}_7^T \end{pmatrix} \in \mathbb{R}^{7 \times 160000} \quad (6.1)$$

## 6.4 Anwendung des Standard-Algorithmus von ICA

Sind die 7 relevanten Hauptkomponenten in  $Y$  gegeben, wie in (6.1), dann lässt sich mit dem in Kapitel 4.2.7 besprochenen Algorithmus eine Matrix  $U$  bestimmen, sodass die Komponenten in  $Z$  mit

$$Z = UY$$

möglichst unabhängig sind. Durch den Standard-Algorithmus werden folgende Komponenten bestimmt:

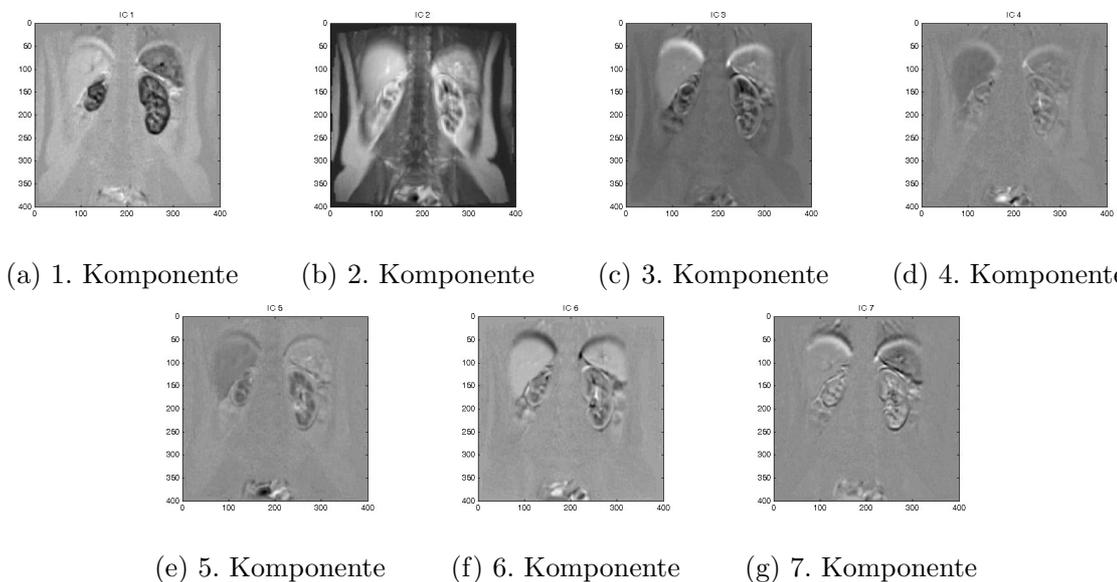


Abb. 6.3: Durch Standard-Algorithmus bestimmte unabhängige Komponenten<sup>2</sup>

Nun können, wie in Kapitel 4.2.8 besprochen, gezielt Komponenten in der Rekonstruktion des ursprünglichen Videos ausgeblendet und dieses somit gezielt verändert werden.

<sup>2</sup>Medizinische Aufnahme von Herrn a.o. Univ.-Prof. Mag. Dr. Keeling erhalten, Bilder daraus selbst erstellt

Hier wird allerdings zunächst der Unterschied zu den durch den verbesserten Algorithmus bestimmten möglichst unabhängigen Komponenten gezeigt.

## 6.5 Anwendung des verbesserten Algorithmus von ICA

Wie mittels des Standard-Algorithmus kann durch den verbesserten Algorithmus eine Matrix  $U$  aus den 7 relevanten Hauptkomponenten bestimmt werden, sodass die Komponenten in  $Z$  mit

$$Z = UY$$

möglichst unabhängig sind. Dabei entstehen die folgenden 7 Komponenten:

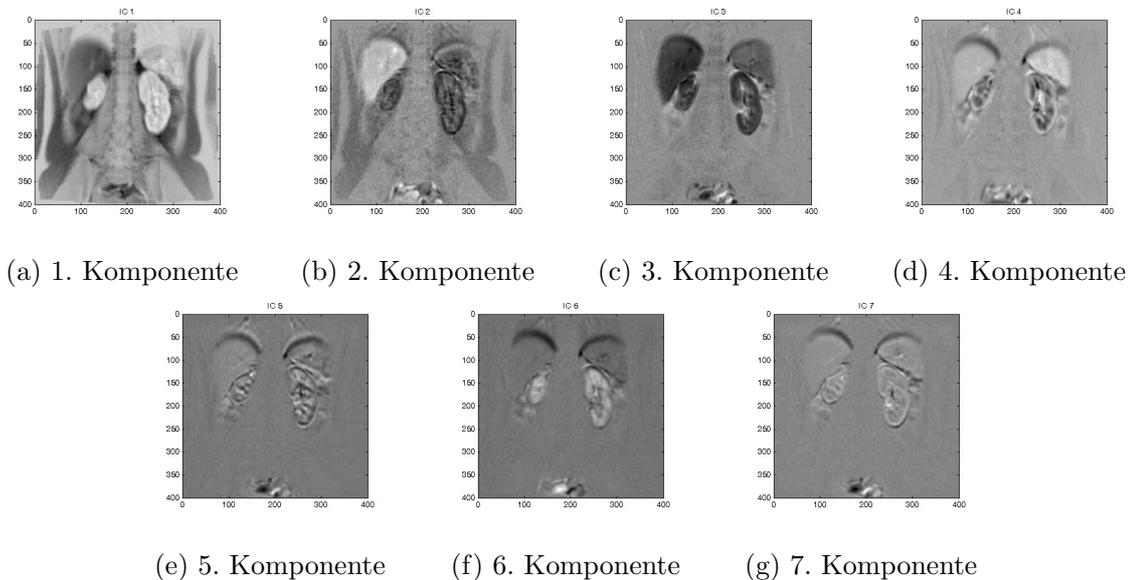


Abb. 6.4: Durch verbesserten Algorithmus bestimmte unabhängige Komponenten<sup>3</sup>

Im Vergleich zu den durch den Standard-Algorithmus bestimmten möglichst unabhängigen Komponenten scheinen jene, welche durch den verbesserten Algorithmus bestimmt sind, etwas weniger Gemeinsamkeiten zu besitzen (dies ist vor allem am Vergleich der Abbildungen 6.4e - 6.4g mit den Abbildungen 6.3e - 6.3g erkennbar).

Nun können gezielt einzelne Komponenten in der Rekonstruktion des ursprünglichen Videos ausgeblendet werden. Nachdem es das Ziel ist, die Bewegung im Video zu entfernen, können die 4., 5., 6. und 7. Komponente ausgeblendet werden.

<sup>3</sup>Medizinische Aufnahme von Herrn a.o. Univ.-Prof. Mag. Dr. Keeling erhalten, Bilder daraus selbst erstellt

Das Video sieht nach der Bearbeitung an den selben Stellen, wie im ursprünglichen Video, wie folgt aus:

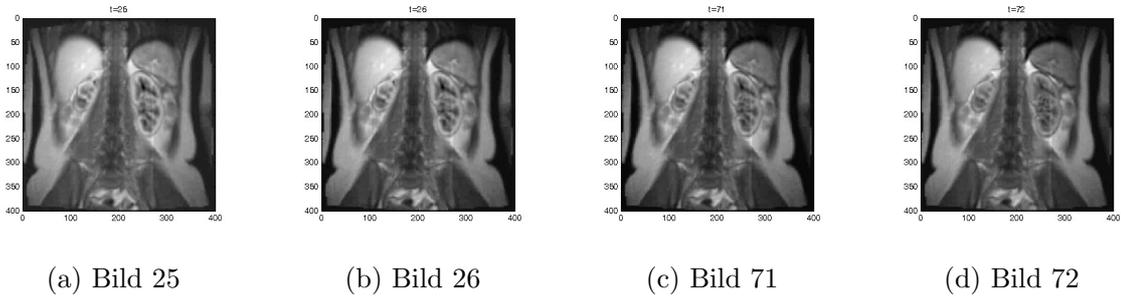


Abb. 6.5: Ausschnitte aus dem bearbeiteten Kontrastvideo<sup>4</sup>

Im Vergleich dazu waren die Bilder des ursprünglichen Videos:

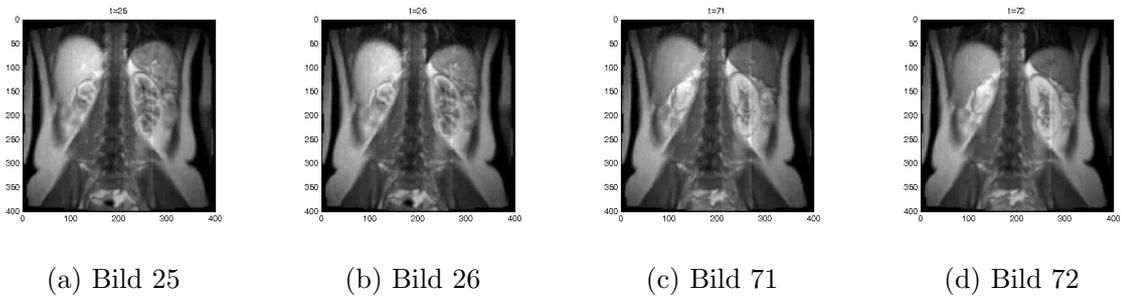


Abb. 6.6: Ausschnitte aus dem originalen Kontrastvideo<sup>4</sup>

Die Bewegung in der Aufnahme konnte deutlich reduziert bzw. teilweise sogar vollständig entfernt werden. Damit war die Anwendung der Verfahren erfolgreich.

Eine vollständige Aufnahme des Videos beziehungsweise des Verbesserungsprozesses befindet sich auf der beiliegenden DVD.

---

<sup>4</sup>Medizinische Aufnahme von Herrn a.o. Univ.-Prof. Mag. Dr. Keeling erhalten, Bilder daraus selbst erstellt

# 7 Zusammenfassung

Die Fachbereichsarbeit *Principal Component Analysis und Independent Component Analysis* beschäftigt sich mit zwei Verfahren der mathematischen Signalanalyse. Dabei werden die wichtigsten mathematischen Aspekte - auf Basis von anfänglichen Vorkenntnissen über lineare Algebra und Statistik - vorgestellt und auch auf etwas weiterführende Themen hinsichtlich der Verbesserung des ICA Algorithmus eingegangen.

Während zu Beginn eine kurze, deskriptive Einleitung in die grundlegende Idee hinter den beiden Methoden gegeben wird, gliedert sich der Hauptteil der Arbeit in vier wesentliche Teile.

Beginnend mit dem Kapitel 3 wird der Leser mit dem Verfahren Principal Component Analysis, beziehungsweise dessen mathematischer Vorgehensweise, vertraut gemacht. Dabei liegt ein besonderer Schwerpunkt auf dem Prinzip der Korrelation, welche die Vorgänge in Principal Component Analysis erst ermöglicht.

Im Kapitel 4 folgt eine Erläuterung des ersten entwickelten Algorithmus für Independent Component Analysis. Dabei wird die mathematische Vorgehensweise und das grundlegende Prinzip der statistischen Unabhängigkeit erklärt, welches im Standardalgorithmus mit der Wölbung und einem Gradienten-Abstiegsverfahren ausgenützt wird.

Darauf folgend werden im Kapitel 5 Schwachstellen in zwei Punkten des Standard Algorithmus aufgezeigt und weiters ein verbesserter Algorithmus, welcher aus der aktuellen Forschungsarbeit von Herrn a.o. Univ.-Prof. Mag. Dr. Keeling (Professor an der Karl-Franzens-Universität Graz) stammt, vorgestellt und mathematisch erläutert. In diesen Kapiteln wird ganz besonders die tatsächliche Vorgehensweise in den Berechnungen dargelegt, welche die Implementierung am Computer nachvollziehbar werden lassen soll.

Dies führt zum Kapitel 6, in welchem die Ergebnisse gezeigt werden, welche mittels der Anwendung der Verfahren in der Videoaufbereitung erzielt werden können. Für ergänzende Demonstrationen dient die beigelegte DVD.

Das Ziel dieser Fachbereichsarbeit war es, einerseits anhand des Themas zu zeigen, welche

vielversprechende praktische Anwendungen von Mathematik es gibt, und andererseits eine mathematische Einführung in die beiden Methoden zu bieten, beziehungsweise ein Gefühl für die realistische Arbeit mit diesen zu vermitteln.

Natürlich wird in dieser Arbeit lediglich ein Bruchteil des gesamten Wissens behandelt, welches im Bereich dieser und ähnlicher Methoden bereits erlangt wurde und auch aktuell erlangt wird. Detailliertere Informationen zu Principal Component Analysis bzw. Independent Component Analysis - vor allem bezüglich Anwendungen und alternativen mathematischen Herangehensweisen - findet man in den Büchern <sup>1</sup> und <sup>2</sup>. Die aktuelle Forschung zu diesem Thema an der Karl-Franzens-Universität Graz kann auf <sup>3</sup> und <sup>4</sup> mitverfolgt werden.

---

<sup>1</sup>Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis

<sup>2</sup>Jolliffe, Ian T.: Principal Component Analysis

<sup>3</sup>Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten

<sup>4</sup>Keeling, Stephen/ Kunisch, Karl: Robust  $l_1$  Approaches to Principal and Independent Component Analysis. Online in WWW unter URL: <http://math.uni-graz.at/keeling/manuskripten/pcaica.pdf>. [Stand 21.02.2013]

# Abbildungsverzeichnis

2.1	Cocktail-Party-Problem . . . . .	8
2.2	MRT-Aufnahme von Tumor unbearbeitet . . . . .	9
2.3	MRT-Aufnahme von Tumor mit PCA und ICA bearbeitet . . . . .	9
3.1	Beispiel eines zweidimensionalen Datensatzes . . . . .	14
3.2	Darstellung durch die stärkste Hauptkomponente . . . . .	14
3.3	Daten mit linearem Zusammenhang . . . . .	16
3.4	Daten ohne linearem Zusammenhang . . . . .	16
3.5	Zweidimensionaler Datensatz . . . . .	17
3.6	Zentrierte Daten . . . . .	17
3.7	Zentrierte Daten (ähnlich zu Abb. 3.6) . . . . .	22
3.8	Gesphärte Daten . . . . .	22
3.9	Zweidimensionaler Datensatz (entspricht Abb. 3.5) . . . . .	24
3.10	Darstellung der Daten durch weniger Hauptkomponenten . . . . .	24
4.1	Diamantförmiger Datensatz . . . . .	26
4.2	Durch PCA transformierte Daten . . . . .	26
4.3	Durch ICA transformierte Daten . . . . .	26
4.4	Datenvektoren mit quadratischem Zusammenhang . . . . .	27
4.5	Zweidimensionaler Datensatz (ähnlich zu Abb. 3.5) . . . . .	31
4.6	Gesphärte Daten (ähnlich zu Abb. 3.8) . . . . .	31
4.7	Datenvektor mit an den Enden gewichteter Verteilung . . . . .	33
4.8	Datenvektor mit gleichmäßiger Verteilung . . . . .	33
4.9	Mischung der beiden nicht gauß-verteilter Datenvektoren . . . . .	33
4.10	Niveaulinien der Funktion $f(\mathbf{x})$ und Gradient an der Stelle $\mathbf{p}$ . . . . .	35
4.11	Gesphärte Daten (entspricht Abb. 4.6) . . . . .	37
4.12	Unabhängige Komponenten . . . . .	37

4.13	Zweidimensionaler Datensatz (entspricht Abb. 4.5)	39
4.14	Darstellung der Daten durch weniger unabhängige Komponenten	39
5.1	Zielfunktion $J_1$ in Abhängigkeit von $\theta$ und $\alpha$	45
5.2	Zielfunktion $J_1$ bei $\alpha = 6$	45
5.3	Zielfunktion $J_2$ in Abhängigkeit von $\theta$ und $\alpha$	47
5.4	Zielfunktion $J_2$ bei $\alpha = 6$	47
5.5	Gespährte Daten mit gedrehten Koordinatensystemen	54
5.6	Durch Standard-Algorithmus bestimmte unabhängige Komponenten	54
5.7	Durch verbesserten Algorithmus bestimmte unabhängige Komponenten	54
6.1	Ausschnitte aus dem originalen Kontrastvideo	55
6.2	Logarithmus der Eigenwerte in $\Lambda$	56
6.3	Durch Standard-Algorithmus bestimmte unabhängige Komponenten	57
6.4	Durch verbesserten Algorithmus bestimmte unabhängige Komponenten	58
6.5	Ausschnitte aus dem bearbeiteten Kontrastvideo	59
6.6	Ausschnitte aus dem originalen Kontrastvideo (entspricht Abb. 6.1)	59

# Literaturverzeichnis

- Arens, Tilo/ Hettlich, Frank/ Karpfinger, Christian/ et al: Mathematik. 2. Aufl., Heidelberg, 2012.
- Hyvärinen, Aapo/ Karhunen, Juha/ Oja, Erkki: Independent Component Analysis. 1. Aufl., New York u.a., 2001.
- Jolliffe, Ian T.: Principal Component Analysis. 2. Aufl., New York, 2002.
- Keeling, Stephen: Grundlagen der Numerischen Mathematik. Online in WWW unter URL:  
<http://math.uni-graz.at/keeling/skripten/numerik.pdf>. [Stand 15.12.2012]
- Keeling, Stephen: Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten. Online in WWW unter URL:  
<http://math.uni-graz.at/keeling/skripten/Tutorium.pdf>. [Stand 19.02.2013]
- Keeling, Stephen/ Kunisch, Karl: Robust  $l_1$  Approaches to Principal and Independent Component Analysis. Online in WWW unter URL:  
<http://math.uni-graz.at/keeling/manuskripten/pcaica.pdf>. [Stand 21.02.2013]
- Leydold, Josef: Lineare Unabhängigkeit. Online in WWW unter URL:  
<http://statistik.wu-wien.ac.at/~leydold/MOK/HTML/node25.html>.  
[Stand 28.02.2013]
- Maierhofer, Georg/ Tichy, Robert: Skriptum zur Vorlesung Ungleichungen. Online in WWW unter URL:  
[http://www.math.tugraz.at/OeMO/F-Kurs/Unterlagen/Ungleichungen\\_Maierhofer.pdf](http://www.math.tugraz.at/OeMO/F-Kurs/Unterlagen/Ungleichungen_Maierhofer.pdf).  
[Stand 22.03.2012]
- mathe online: Zufall und Wahrscheinlichkeit. Online in WWW unter URL:  
<http://www.mathe-online.at/mathint/wstat1/i.html#Beispiele123>. [Stand 19.02.2013]

# Beilagen

DVD Auf der DVD befinden sich drei Anwendungsbeispiele<sup>5</sup> von Principal Component Analysis und Independent Component Analysis. Dabei werden in einer grafischen Oberfläche drei Kontrastvideos mit den beiden Verfahren bearbeitet. Die detailliertere Darstellung von Videos in bewegten Bildern soll ein leichteres Erkennen der erzielten Verbesserung der Videos ermöglichen.

---

<sup>5</sup>Medizinische Aufnahmen von Herrn a.o. Univ.-Prof. Mag. Dr. Keeling erhalten

# Eidesstattliche Erklärung

Ich, Georg Andreas Maierhofer, erkläre hiermit, dass ich diese Arbeit selbst verfasst und dabei keine anderen als die von mir angeführten Quellen und Hilfsmittel benutzt habe.

Langenwang, 1. März 2013

.....

Georg Andreas Maierhofer