





John von Neumann-Institut für Computing (NIC)

Harald Stögbauer

**Zerlegung von Signalen in  
unabhängige Komponenten:  
Ein informationstheoretischer Zugang**

Dissertation

NIC-Serie Band 26

ISBN 3-00-013620-7

---

Forschungsgruppe Komplexe Systeme

Die Deutsche Bibliothek - CIP-Einheitsaufnahme  
Ein Titeldatensatz für diese Publikation ist bei  
Der Deutschen Bibliothek erhältlich.

Herausgeber: NIC-Direktorium  
Vertrieb: NIC-Sekretariat  
Forschungszentrum Jülich  
52425 Jülich  
Deutschland  
Internet: [www.fz-juelich.de/nic](http://www.fz-juelich.de/nic)  
Druck: Graphische Betriebe, Forschungszentrum Jülich

© 2005 John von Neumann-Institut für Computing

Es ist erlaubt, dieses Werk oder Teile davon digital oder auf Papier zum persönlichen Gebrauch oder zu Lehrzwecken zu vervielfältigen, vorausgesetzt die Kopien werden nicht kommerziell genutzt. Kopien müssen diese Copyright-Notiz und das volle Zitat auf ihrer Titelseite enthalten. Andere Vervielfältigung bedarf der vorherigen schriftlichen Genehmigung des oben genannten Herausgebers.

NIC-Serie Band 26

ISBN 3-00-013620-7



Bergische Universität  
Wuppertal

**Zerlegung von Signalen in unabhängige Komponenten:  
Ein informationstheoretischer Zugang**

**Dissertation**

von

**Harald Stögbauer**

**Januar 2005**

(WUB-DIS 2005-02)



# Kurzfassung

Das Hauptziel der vorliegenden Dissertation ist die Zerlegung von multivariaten Signalen in möglichst unabhängige Komponenten. Zu diesem Zweck werden neue Methoden entwickelt, die durch die Informationstheorie motiviert sind. Diese erlauben nicht nur die Lösung des Minimierungsproblems, sondern auch weitere Analysen der Komponenten.

Im ersten Teil dieser Dissertation wird ein exakter Schätzer für die Abhängigkeit zwischen zwei reellwertigen Zufallsvariablen, gemessen durch die wechselseitige Information (Mutual Information (MI)), eingeführt, der als Ausgangspunkt für weitere Entwicklungen dient. Mit Hilfe numerischer Tests wird für eine Vielzahl von Verteilungen bewiesen, dass der neue Schätzer im Vergleich zu anderen bekannten Schätzern minimalen systematischen und statistischen Fehler besitzt. Darüber hinaus wird eine neuartige Cluster-Methode entwickelt, die eine sehr natürliche Definition für die Abstände zwischen beliebigen Clustern benutzt, wobei auf eine Mittelung über Eigenschaften der einzelnen Objekte verzichtet werden kann.

Im zweiten Teil wird der MI-Schätzer verwendet, um einen neuen Algorithmus für die Independent Component Analysis (ICA) zu formulieren. Dieser zeigt in numerischen Tests, dass er hinsichtlich Genauigkeit bisherigen ICA-Algorithmen überlegen ist. Des Weiteren wird mit Hilfe einer genauen Abschätzung der absoluten Abhängigkeit ein Zuverlässigkeits-Test für beliebige ICA-Ausgaben entwickelt. Außerdem wird im Anschluss ein erweiterter ICA-Algorithmus formuliert, der es erlaubt, für verrauschte Daten bessere Ergebnisse zu erzielen.

Der dritte Teil der Dissertation beschäftigt sich mit der Anwendung dieser neu entwickelter Algorithmen auf experimentelle Messdaten. Anhand der Untersuchung des EKGs einer schwangeren Frau wird gezeigt, dass mit kleinen Erweiterungen des Standard-ICA-Modells sehr gute Ergebnisse in der Trennung des Mutter- und Fötus-Signals erzielt werden. In einer weiteren Anwendung werden die Infrarotspektren von organischen Mischungen, trotz gewisser Abhängigkeiten zwischen den Spektren der Reinsubstanzen, durch eine passende Vorverarbeitung hervorragend separiert. Außerdem werden die neuen Algorithmen zur Reduzierung des systematischen Fehlers in der Synchronisationsanalyse, der durch eine Überlagerung der beiden Signale entsteht, eingesetzt. Als letzte Anwendung multivariater Methoden wird eine Schlafphasenklassifikation in intrakraniellen EEG-Daten durchgeführt.



# Abstract

The main object of this dissertation is to decompose multivariate signals into components as independent as possible. For this purpose a new approach based on information theory is developed that allows not only to solve the minimization problem but also to further analyze the components obtained.

In the first part of this dissertation a precise estimator of mutual information (MI) is introduced which is used as starting point for further developments. In numerical tests it is proven that for a variety of distributions this new estimator shows minimal systematical and statistical errors in comparison to other estimators. Furthermore, by means of MI a novel cluster method is formulated which defines distances of higher dimensional objects without having to average over properties of the single objects.

In the second part the MI estimator is used to formulate a new independent component analysis (ICA) algorithm. Numerically, it is found that the algorithm outperforms previous algorithms in regard to accuracy. In addition, the exact estimation of the absolute dependency allows to express a reliability test for any ICA component. Subsequently, an extended ICA algorithm is formulated which achieves improved results with noisy data.

In the third part of this dissertation these new methods are applied to experimental data. First, the ECG of a pregnant woman is separated into signals of the mother and fetus by an extended version of the standard ICA model. In the next application the infrared spectra of the pure substances are identified from the spectra of organic mixtures, despite the fact that the first ones are not really independent. Subsequently, the new algorithms are used to eliminate the bias in a synchronization analysis due to an overlap of two signals. As a last application it is shown that multivariate methods are also able to perform a sleep stage classification on intracranial EEG data.

Erster Gutachter: Prof. Dr. Peter Grassberger, Universität Wuppertal  
Zweiter Gutachter: Prof. Dr. Andreas Klümper, Universität Wuppertal  
Dritter Gutachter: Prof. Dr. Klaus-Robert Müller, Universität Potsdam  
Eingereicht am: 24. Januar 2005

# Inhaltsverzeichnis

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Einführung</b>  | <b>1</b>  |
| <b>2</b> | <b>Messung von Abhängigkeiten</b>                              | <b>5</b>  |
| 2.1      | Informationstheorie . . . . .                                  | 5         |
| 2.2      | Wechselseitige Information - Mutual Information (MI) . . . . . | 6         |
| 2.3      | Schätzung der MI . . . . .                                     | 8         |
| 2.4      | Praktische Überlegungen . . . . .                              | 13        |
| 2.5      | Numerische Simulationen . . . . .                              | 14        |
| 2.6      | Eigenschaften der MI . . . . .                                 | 19        |
| 2.7      | Anwendungen der MI . . . . .                                   | 20        |
| 2.7.1    | Anhängigkeitsmaß für die Zeitreihenanalyse . . . . .           | 20        |
| 2.7.2    | Clusteranalyse . . . . .                                       | 21        |
| <b>3</b> | <b>Behandlung multivariater Daten</b>                          | <b>25</b> |
| 3.1      | Blind Source Separation (BSS) . . . . .                        | 26        |
| 3.2      | Independent Component Analysis (ICA) . . . . .                 | 27        |
| 3.2.1    | Einführung . . . . .   | 27        |
| 3.2.2    | Unabhängig kontra unkorreliert . . . . .                       | 28        |
| 3.2.3    | ICA-Algorithmen . . . . .                                      | 31        |
| 3.2.4    | Leistungsvergleich verschiedener ICA-Algorithmen . . . . .     | 44        |
| 3.2.5    | Von der ICA zur Blind Source Separation . . . . .              | 47        |
| 3.2.6    | Von der Theorie zur Praxis . . . . .                           | 50        |

|          |   |           |
|----------|---|-----------|
| <b>4</b> | <b>Anwendungen der ICA</b>                                | <b>57</b> |
| 4.1      | Elektrokardiogramm (EKG) einer schwangeren Frau . . . . . | 57        |
| 4.2      | Infrarotspektren organischer Moleküle . . . . .           | 65        |
| 4.3      | Synchronisation oder Superposition? . . . . .             | 73        |
| 4.4      | Schlafphasenklassifikation anhand des EEG . . . . .       | 76        |
| <b>5</b> | <b>Zusammenfassung und Ausblick</b>                       | <b>81</b> |

# Kapitel 1

## Einführung

Bei der experimentellen Untersuchung komplexer Systeme werden oft eine Vielzahl von Signalen gemessen. Aus dem Wissen über ihre wechselseitigen Abhängigkeiten kann eine Fülle an Information über das untersuchte System gewonnen werden. So werden Schätzungen dieser Abhängigkeiten z.B. in der Synchronisations- [51] oder Clusteranalyse [36] verwendet. Das Wissen über die Abhängigkeiten kann aber auch für andere Zwecke verwendet werden. So kann oft angenommen werden, dass Signale, die von unterschiedlichen Quellen ausgesandt werden, unabhängig sind. Bei einer Messung werden dann aber meistens nur Mischungen von diesen Quellsignalen aufgenommen. Durch Minimieren der so auftretenden Abhängigkeiten können dann wieder die ursprünglichen Signale zurückgewonnen werden. Diese Rückgewinnung wird "Blind Source Separation (BSS)" genannt, und wie schon angedeutet, sind dabei nicht nur die Quellsignale unbekannt, sondern auch die Parameter der Mischung. Die Methode, die mit Hilfe von Schätzungen der Abhängigkeiten versucht, diese Aufgabe zu lösen, wird "Independent Component Analysis (ICA)" [27] genannt.

Es gibt eine Vielzahl von Abhängigkeitsmaßen, wobei die Mutual Information (MI) [60] durch ihren informationstheoretischen Hintergrund ausgezeichnet ist. Dieses Maß ist dann und nur dann Null, wenn die untersuchten Signale völlig unabhängig sind. Hierin unterscheidet sie sich z.B. vom Korrelationskoeffizienten, der nur lineare Abhängigkeiten misst. Ein weiterer Vorteil der MI ist, dass sie auch für höherdimensionale Signale definiert ist. Diese Eigenschaften zeichnen die MI gegenüber anderen Abhängigkeitsmaßen aus und ermöglichen eine Reihe von zusätzlichen Anwendungen. Es kann zum Beispiel eine Clustermethode formuliert werden, die bei der Berechnung mehrdimensionaler Cluster auf die sonst übliche Mittelung über Eigenschaften der einzelnen Objekte verzichtet. Des weiteren stellt die MI ein Maß der absoluten Abhängigkeit dar, was z.B. in der ICA von Nutzen ist, um zu entscheiden, ob tatsächlich unabhängige Komponenten erhalten wurden. Dies ist mit anderen Maßen nicht möglich, da diese nur relative Abhängigkeiten abschätzen bzw. nicht auf alle Arten von Abhängigkeiten sensitiv sind.

Die ICA ist eine relativ junge Technik und wurde das erste Mal vor etwa 20 Jahren in der Literatur erwähnt [31]. Bis jetzt lag das Hauptaugenmerk auf der Entwicklung von schnellen Algorithmen für die Trennung von instantanen linearen Mischungen [43, 28, 10, 79]. Es sind bereits Algorithmen erstellt worden, die z.B. zeitversetzte [39, 73] bzw. nichtlineare Mischungen [69, 24] separieren können, diese sind aber bis jetzt viel weniger getestet worden und somit sind ihre Fehlerquellen noch recht unbekannt.

Ein Beweis dafür, dass die ICA noch immer in ihren Anfängen steckt, ist, dass erst seit kurzem die Zuverlässigkeit der geschätzten ICA-Komponenten diskutiert wird [45, 23, 26, 64]. Die Überprüfung der tatsächlichen Abhängigkeit der Komponenten und ihre Eindeutigkeit ist unerlässlich, um diesen eine physikalische Bedeutung zu geben. Nur wenn diese Prüfungen positiv ausfallen, entspricht die ICA-Ausgabe den tatsächlichen Quellsignalen, und die BSS ist erfolgreich durchgeführt worden.

Ein weiteres fundamentales Problem, welches noch weitgehend ungelöst ist, ist die Trennung von verrauschten Mischungen. Bis jetzt existieren nur Algorithmen [27], die eine Vielzahl von Anforderungen an das Rauschen stellen. Diese sind aber für experimentelle Daten meistens nicht erfüllt. Um die Trennung von Messsignalen zu ermöglichen, die mit stärkerem Messrauschen überlagert sind, müssen daher erst passende ICA-Algorithmen entwickelt werden.

Trotz der intensiven und relativ langen Beschäftigung mit dem instantanen linearen ICA-Modell sind erst kürzlich neue Algorithmen entwickelt worden [6, 42, 65], die eine höhere Trenngenauigkeit erreichen als die bisherigen. Dies wurde zwar mit einem numerischen Mehraufwand erkauft, dieser ist aber für die meisten Anwendungen akzeptabel.

Dennoch sind schon eine Reihe von Erfolgen bei der Anwendung der ICA auf experimentelle Daten erzielt worden, insbesondere in den Gebieten der Akustik [72], Medizin [76, 65] und Spektroskopie [5, 49]. Dabei handelt es sich z.B. um die Trennung von Gemischen verschiedener Geräusche, die Separation von Überlagerungen physiologischer Signale und die Gewinnung der Spektren von Reinsubstanzen aus chemischen Mischungen.

Aufgrund der Vielzahl von Situationen, in denen durch simultane bzw. sequentielle Aufzeichnungen multidimensionale Datensätze erzeugt werden, ist mit einer Anzahl weiterer Anwendungen zu rechnen. Außerdem kann in manchen Gebieten, wo bis jetzt erfolglos bestimmte ICA-Algorithmen angewandt wurden, durch die Verwendung neuer Techniken mit einem Durchbruch gerechnet werden.

Die vorliegende Doktorarbeit ist wie folgt gegliedert. Der erste Teil beinhaltet eine Einführung in die Messung von Abhängigkeiten (Kap.(2.2)). Nach einer allgemeinen Diskussion

---

wird dann ein neuer Schätzer der MI [37] vorgestellt (Kap.(2.3)). Nach der numerischen Validierung des Schätzers in Kap.(2.4) werden dann praktische Überlegungen zur Implementierung des Schätzers angestellt. Im letzten Teil dieses Kapitels (Kap.(2.7)) werden die Anwendungsmöglichkeiten der MI als Abhängigkeitsmaß in der Synchronisations- und der Clusteranalyse gezeigt.

Im zweiten Teil der Arbeit wird zuerst eine Einführung in die ICA gegeben (Kap.(3.2)), wobei die bekanntesten Algorithmen vorgestellt werden. Anschließend wird in Kap.(3.2.3) ein neuer ICA-Algorithmus vorgestellt, der im Anschluss an einer Vielzahl von Signalen mit anderen Algorithmen verglichen wird. Danach wird mit Hilfe des neuen Schätzers für die MI ein Zuverlässigkeitstest (Kap.(3.2.5)) für die ICA-Ausgabe formuliert. Vor dem Abschluss dieses Kapitels werden praktische Überlegungen (Kap.(3.2.6)) angestellt, z.B. wie zu entscheiden ist, welcher ICA-Algorithmus am besten für eine Anwendung geeignet ist, oder was getan werden kann, wenn die Quellsignale abhängig sind, oder wie Rauschen in der Mischung behandelt werden kann.

Im letzten Teil dieser Arbeit werden Anwendungen der ICA auf experimentelle Daten besprochen. Im Kap.(4.1) wird erfolgreich das EKG einer schwangeren Frau in einen Mutter- und Fötus-Anteil separiert. Die nächste Anwendung (Kap.(4.2)) beschäftigt sich mit der Rückgewinnung von Infrarotspektren der Reinsubstanzen aus Spektren von Mischungen organischer Verbindungen. In Kap.(4.3) wird gezeigt, dass die ICA in der Lage ist, den systematischen Fehler in der Synchronisationsanalyse zu reduzieren, der durch eine Überlagerung der Signale entsteht. Die letzte Anwendung (Kap.(4.4)) multivariater Methoden ist eine Schlafphasenklassifikation anhand des intrakraniellen EEGs von Epilepsiepatienten. Die Arbeit wird mit einer Zusammenfassung und einem Ausblick abgeschlossen.



# Kapitel 2

## Messung von Abhängigkeiten

Bei der Untersuchung komplexer Systeme stellt sich oft die Frage "Besteht zwischen dem System A und System B ein Zusammenhang, gab es einen Informationsaustausch zwischen ihnen?". Antworten auf diese Frage ermöglichen, die Beziehungen in einem komplexen Netzwerk und damit auch die zugrundeliegenden Mechanismen zu verstehen. Im Allgemeinen ist es aber nicht so einfach, diese Abhängigkeiten zu bestimmen.

Eine Möglichkeit, den gemeinsamen Informationsinhalt von Systemen zu bestimmen, bietet die Informationstheorie. Sie ermöglicht es, ausgehend von Messungen der Systeme unter Zuhilfenahme von statistischen Methoden, z.B. mit der wechselseitigen Information ("mutual information", MI), eine Schätzung ihrer Abhängigkeit zu geben. Die Bestimmung der Abhängigkeiten zweier oder mehrerer Messreihen ermöglicht auch eine Reihe weiterer Anwendungen. In dem Fall, dass diese Abhängigkeiten nur dadurch zustande kamen, dass die verschiedenen Messreihen Überlagerung der ursprünglichen Signale sind, kann versucht werden durch Minimierung eben dieser Abhängigkeiten die "originalen" (unge-mischten) Signale wieder zurückzugewinnen. Dieser Ansatz wird oft als ICA (Independent Component Analysis) bezeichnet und wird im nächsten Kapitel besprochen. Es gibt also eine Vielzahl von Anwendungsmöglichkeiten, wenn man in der Lage ist, die Abhängigkeiten zwischen Signalen zu bestimmen. Im Folgenden werden wir einen neu entwickelten Schätzer für die MI vorstellen, der sich durch besonders geringe systematische und statistische Fehler auszeichnet. Außerdem werden wir die so genannte Gruppierungseigenschaft der MI ausnützen, um eine neue Methode der Clusteranalyse vorzustellen. Alle diese Ansätze werden durch umfassende numerische Simulationen überprüft.

### 2.1 Informationstheorie

In der Informationstheorie stellt man sich zunächst die Frage, wie Messsignale kodiert werden können. Eine beliebige Messung kann nach erfolgreicher Kodierung in einem Speicher

abgelegt, übertragen und verarbeitet werden. Die Auffindung optimaler Lösungen für die Datenverarbeitung ist eine grundlegende Fragestellung für viele modernen Wissenschaften wie Mathematik, Physik, Nachrichtentechnik, Informatik, Elektrotechnik und Medizin. Grundlegende Konzepte der Informationstheorie wurden bereits in den vierziger Jahren des zwanzigsten Jahrhunderts von Claude Shannon veröffentlicht [59]. Ein Beispiel ist der aus der Thermodynamik übernommene Begriff der Entropie. Eine Basisgröße der Informationstheorie ist die Shannon-Entropie, die für eine diskret verteilte Zufallsvariable  $X$  wie folgt definiert ist:

$$H(X) = - \sum_i p(X = a_i) \log p(X = a_i). \quad (2.1)$$

Hier stehen die  $a_i$  für mögliche Werte von  $X$ , und  $p(\dots)$  ist die Wahrscheinlichkeit dafür, dass eine Messung von  $X$  den Wert  $a_i$  ergibt. Die Einheit der Entropie ist dabei abhängig von der Basis des Logarithmus. Häufig wird als Basis 2 gewählt, da dies der Kodierung in binären Zahlen entspricht (Einheit: *bit*). Die Entropie einer Zufallsvariable ist ein Maß für den mittleren Gewinn an Information durch eine Messung. Je unstrukturierter und zufälliger eine Variable ist, um so größer ist ihre Entropie. Im anderen Fall, wenn die Zufallsvariable zum Beispiel die meiste Zeit einen konstanten Wert annimmt und nur manchmal davon abweicht, kann im Mittel nicht sehr viel Information bei der Beobachtung einer Messung gewonnen werden, da diese sehr wahrscheinlich dem konstanten Wert entspricht. Diese Situation zeichnet sich durch eine kleine Entropie aus.

Ein anschauliches Beispiel kann in Ref. [27] gefunden werden: Betrachten wir eine Zufallsvariable  $X$ , welche acht verschiedene Werte mit den Wahrscheinlichkeiten  $(1/2, 1/4, 1/8, 1/16, 1/64, 1/64, 1/64, 1/64)$  annehmen kann. Die Entropie von  $X$  ist dann genau 2 bit. Wenn wir in herkömmlicher Weise jeder Variablen eine gleich lange binäre Zahl zuordnen, benötigen wird aber 3 bit. Ein bessere Methode wäre aber, häufige Variablen durch kurze binäre Folgen zu kodieren und weniger häufige durch längere. Wir könnten zum Beispiel 0, 10, 110, 1110, 111100, 111101, 111110, 111111 verwenden (Diese Folgen können direkt in einer Bitsequenz verbunden werden, da sie aufgrund ihrer Konstruktion trotzdem eindeutig zugeordnet werden können). Mit dieser Kodierung sind im Mittel nun nurmehr  $\frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \frac{1}{16} \times 4 + \frac{4}{64} \times 5 = 2$  bit erforderlich, was der berechneten Entropie (= minimale Kodierungslänge) entspricht.

## 2.2 Wechselseitige Information - Mutual Information (MI)

Wie bereits eingangs erwähnt, ist die Mutual Information ein Maß für die Abhängigkeit von Zufallsvariablen. Wir definieren sie zunächst für zwei Zufallsvariablen  $X$  und  $Y$ . Eine Verallgemeinerung auf beliebig viele Zufallsvariablen erfolgt später.

In Gl.(2.1) ist davon ausgegangen worden, dass die Zufallsvariable diskret verteilt ist. In den meisten Anwendungen sind aber die Werte der Zufallsvariable kontinuierlich verteilt, was in der Abschätzung der Entropie berücksichtigt werden muss. Dazu wird der Wertebereich der Zufallsvariable in Intervalle aufgeteilt ("binning") und die Wahrscheinlichkeitsverteilung durch eine im Intervall konstante Wahrscheinlichkeitsdichte  $p_X(x) \approx \mu_X(\tilde{x})\Delta^m$  ersetzt, wobei  $m$  der Dimension der Variable entspricht. Eine Verkürzung der Intervalllänge  $\Delta$  gegen Null führt zur Zerlegung in die "Differential-Entropie"  $\tilde{H}(X)$  und einen konstanten Beitrag

$$H_{\text{bin}}(X) \approx \tilde{H}(X) - m \log \Delta \quad , \quad \tilde{H}(X) = - \int dx \mu_x(x) \log \mu_x(x) \quad . \quad (2.2)$$

Die Mutual Information ist allgemein definiert als

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \quad . \quad (2.3)$$

Im Fall kontinuierlicher Variablen hebt sich der zweite Beitrag der Entropie  $\log \Delta$  durch die Differenzenbildung der Entropien auf, und es ergibt sich

$$I(X, Y) = \lim_{\Delta \rightarrow 0} I_{\text{bin}} = \iint dx dy \mu(x, y) \log \frac{\mu(x, y)}{\mu_x(x)\mu_y(y)} \quad . \quad (2.4)$$

Hierbei ist angenommen, dass die Verbundvariable  $Z = (X, Y)$  die Wahrscheinlichkeitsdichte  $\mu(x, y)$  hat. Die Wahrscheinlichkeitsdichten von  $X$  und  $Y$  sind durch  $\mu_x(x) = \int dy \mu(x, y)$  und  $\mu_y(y) = \int dx \mu(x, y)$  bestimmt.

Die Gl.(2.4) ist genau dann Null, wenn  $X$  und  $Y$  unabhängig sind, d.h. dass die Wahrscheinlichkeitsverteilung faktorisiert,  $\mu(x, y) = \mu_x(x)\mu_y(y)$ . In allen anderen Fällen nimmt sie größere Werte als Null an. Das Problem ist nun,  $I(X, Y)$  anhand von Daten zu schätzen. Die zu untersuchenden Daten bestehen aus einem Satz aus  $N$  bivarianten Messungen  $z_i = (x_i, y_i)$ ,  $i = 1, \dots, N$ , von denen angenommen wird, dass sie iid (independent identically distributed) Realisationen einer Zufallsvariable  $Z = (X, Y)$  sind. Die Aufgabe ist es,  $I(X, Y)$  von dem Satz  $z_i$  abzuschätzen, ohne dabei explizites Wissen über  $\mu$ ,  $\mu_x$ , und  $\mu_y$  zu benutzen.

Im Gegensatz zu anderen Schätzern, die auf Kumulantenentwicklungen [27], Parametrisierung von Dichten [60], Kern-Schätzern [60] oder Binning [16] basieren, wollen wir hier einen Algorithmus behandeln, der die  $k$ -nächsten Nachbarn jedes Punktes  $z_i$  zur Schätzung benutzt [37]. Dieser ist dateneffizient und adaptiv, das heißt, mit  $k=1$  wird bis zum kleinsten möglichen Maßstab aufgelöst, und die Auflösung ist dort höher, wo mehr Daten vorhanden sind. Weiterhin besitzt er einen minimalen systematischen Fehler, für unabhängige Verteilungen wird dieser, bis auf einen statistischen Fehler, sogar exakt Null. Diese Eigenschaften

zeichnen den neuen Schätzer gegenüber anderen in der Literatur (ein Überblick ist in Ref. [37, 74] zu finden) diskutierten Algorithmen aus und machen ihn besonders nützlich, wenn eine möglichst genaue Bestimmung der Abhängigkeiten erforderlich ist.

## 2.3 Schätzung der MI

Die naheliegendste und am meisten verbreitete Methode die Mutual Information abzuschätzen besteht darin, die Träger von  $X$  und  $Y$  in endliche Intervalle zu zerlegen ("binning") und Gl.(2.4) durch eine Summe abzuschätzen

$$I(X, Y) \approx I_{\text{binned}}(X, Y) \equiv \sum_{ij} p(i, j) \log \frac{p(i, j)}{p_x(i)p_y(j)}, \quad (2.5)$$

wobei  $p_x(i) = \int_i dx \mu_x(x)$ ,  $p_y(j) = \int_j dy \mu_y(y)$  und  $p(i, j) = \int_i \int_j dx dy \mu(x, y)$  ist. ( $\int_i$  entspricht dem Integral über das Intervall  $i$ ). Um die MI jetzt abzuschätzen, müssen die Punkte abgezählt werden, die in die einzelnen Intervalle fallen. Wenn  $n_x(i)$  ( $n_y(j)$ ) die Anzahl der Punkte darstellt, die in das  $i$ -te Intervall von  $X$  ( $j$ -te Intervall von  $Y$ ) fallen, und  $n(i, j)$  die Anzahl, die in die Schnittfläche der Intervalle fallen, dann ist  $p_x(i) \approx n_x(i)/N$ ,  $p_y(j) \approx n_y(j)/N$ , und  $p(i, j) \approx n(i, j)/N$ , wenn alle Wahrscheinlichkeitsdichten durch (nicht notwendigerweise glatte) Funktionen beschreibbar sind. Die rechte Seite von Gl.(2.5) konvergiert dann zu  $I(X, Y)$ , wenn die Datenpunktzahl gegen unendlich geht und die Intervalllänge hinreichend langsam gegen Null tendiert. Im Fall, dass aber die Dichten (multi-)fraktal sind, würde Gl.(2.5) auflösungsabhängige Entropien definieren, die im Grenzfalle von unendlicher Auflösung divergieren würden [19, 20, 61]. Obgleich die im Folgenden vorgestellte Methode auch für diesen Fall adaptiert werden kann, gehen wir dieser Problemstellung hier nicht weiter nach.

Die in Gl.(2.5) verwendete Intervalllänge muss nicht für alle Intervalle gleich gewählt werden. Optimierte Schätzer [18, 16] verwenden in der Tat adaptive Intervalllängen, die so gewählt sind, dass die  $n(i, j)$  möglichst gleich sind. Obwohl solche Schätzer viel besser sind als solche, die eine feste Länge verwenden, haben sie trotzdem noch systematische Fehler, die einerseits von der Abschätzung von  $I(X, Y)$  durch  $I_{\text{binned}}(X, Y)$  stammen und andererseits durch die Abschätzung der Wahrscheinlichkeitsdichten durch Häufigkeiten zustande kommen. Der Fehler der Letzteren kann im Prinzip durch Korrekturterme für endliche  $n_x(i)$  bzw.  $n_y(j)$  minimiert werden [20], was aber bei den Algorithmen in Ref. [18, 16] nicht gemacht wurde.

Wir wollen hier einen anderen Weg gehen, und die Mutual Information mittels einer Statistik über die  $k$ -nächsten Nachbarn abschätzen. Es wird zunächst der Kozachenko-Leonenko Schätzer für die Shannon Entropie [33, 19, 61, 75] besprochen, der als Ausgangspunkt für die Schätzung der MI dient.

### 2.3. SCHÄTZUNG DER MI

---

Sei  $X$  eine kontinuierliche Zufallsvariable mit Werten in einem metrischen Raum, d.h. es existiert eine Abstandsfunktion  $\|x - x'\|$  zwischen zwei beliebigen Realisationen von  $X$ , und sei die Wahrscheinlichkeitsdichte  $\mu(x)$  eine Funktion im eigentlichen Sinn. Dann ist die differentielle Shannon-Entropie durch Gl.(2.2) definiert. Das Ziel ist jetzt,  $\tilde{H}(X)$  aus  $N$  Stichproben  $(x_1 \dots x_N)$  abzuschätzen.

Im ersten Schritt der Herleitung wird verwendet, dass Gl.(2.2) auch als Mittelwert von  $\log \widehat{\mu(x)}$  verstanden werden kann. Wenn es uns möglich ist, einen fehlerfreien<sup>1</sup> Schätzer für  $\log \widehat{\mu(x)}$  zu finden, hätten wir auch sofort einen fehlerfreien Schätzer für die Entropie

$$\hat{H}(X) = -N^{-1} \sum_{i=1}^N \log \widehat{\mu(x_i)}. \quad (2.6)$$

Um  $\log \widehat{\mu(x_i)}$  abzuschätzen, benutzen wir die Wahrscheinlichkeitsdichte  $P_k(\epsilon)$  für die Abstände zwischen  $x_i$  und den  $k$ -ten nächsten Nachbarn. Die Größe  $P_k(\epsilon)d\epsilon$  ist gleich der Wahrscheinlichkeit, einen Punkt innerhalb des Radius  $r \in [\epsilon/2, \epsilon/2 + d\epsilon/2]$  um  $x_i$  zu finden,  $k - 1$  Punkte innerhalb kleinerer Abstände und  $N - k - 1$  Punkte außerhalb des Radius. Ferner sei  $p_i$  die Masse des um  $x_i$  zentrierten  $\epsilon$ -Balls,  $p_i(\epsilon) = \int_{\|\xi - x_i\| < \epsilon/2} d\xi \mu(\xi)$ . Unter Verwendung der Trinomialgleichung erhalten wir

$$\begin{aligned} P_k(\epsilon)d\epsilon &= \frac{(N-1)!}{1!(k-1)!(N-k-1)!} \times \\ &\times \frac{dp_i(\epsilon)}{d\epsilon} d\epsilon \times p_i^{k-1} \times (1-p_i)^{N-k-1} \end{aligned} \quad (2.7)$$

bzw.

$$P_k(\epsilon) = k \binom{N-1}{k} \frac{dp_i(\epsilon)}{d\epsilon} p_i^{k-1} (1-p_i)^{N-k-1}. \quad (2.8)$$

Es kann leicht überprüft werden, dass die Normierung  $\int d\epsilon P_k(\epsilon) = 1$  der Wahrscheinlichkeitsdichte erfüllt ist. Mit Hilfe der Gl.(2.8) kann auch sofort der Erwartungswert von  $\log p_i(\epsilon)$  berechnet werden

$$\begin{aligned} E(\log p_i) &= \int_0^\infty d\epsilon P_k(\epsilon) \log p_i(\epsilon) \\ &= k \binom{N-1}{k} \int_0^1 dp p^{k-1} (1-p)^{N-k-1} \log p \\ &= \psi(k) - \psi(N), \end{aligned} \quad (2.9)$$

---

<sup>1</sup>Mit "fehlerfrei" meinen wir hier, dass das Ensemblemittel von  $\hat{H}(X)$  mit  $H(X)$  übereinstimmt.

wobei  $\psi(x)$  die Digamma-Funktion ist,  $\psi(x) = \Gamma(x)^{-1}d\Gamma(x)/dx$ . Sie wird durch die Rekursionformel  $\psi(x+1) = \psi(x) + 1/x$  und  $\psi(1) = -C$  erzeugt, wobei  $C = 0.5772156\dots$  die Euler-Mascheroni Konstante ist. Für große  $x$  gilt  $\psi(x) \approx \log x - 1/2x$ . Der Erwartungswert ist über die Position aller anderen  $N - 1$  Punkte berechnet worden, wobei  $x_i$  festgehalten wurde. Ein Schätzer für  $\log \mu(x)$  wird dadurch erhalten, dass  $\mu(x)$  über den gesamten  $\epsilon$ -Ball als konstant angenommen wird. Dies ergibt

$$p_i(\epsilon) \approx c_d \epsilon^d \mu(x_i) , \quad (2.10)$$

wobei  $d$  die Dimension von  $x$  ist, und  $c_d$  das Volumen des  $d$ -dimensionalen Einheitsballs darstellt. Für die Maximumsnorm ergibt dies  $c_d = 1$  und für die euklidische Norm  $c_d = \pi^{d/2} / \Gamma(1 + d/2) / 2^d$ . Im späteren Verlauf werden wir stets die Maximumsnorm verwenden. Aus der Kombination von Gl.(2.9) und (2.10) erhält man

$$\log \mu(x_i) \approx \psi(k) - \psi(N) - d E(\log \epsilon) - \log c_d , \quad (2.11)$$

und daraus

$$\hat{H}(X) = -\psi(k) + \psi(N) + \log c_d + \frac{d}{N} \sum_{i=1}^N \log \epsilon(i) , \quad (2.12)$$

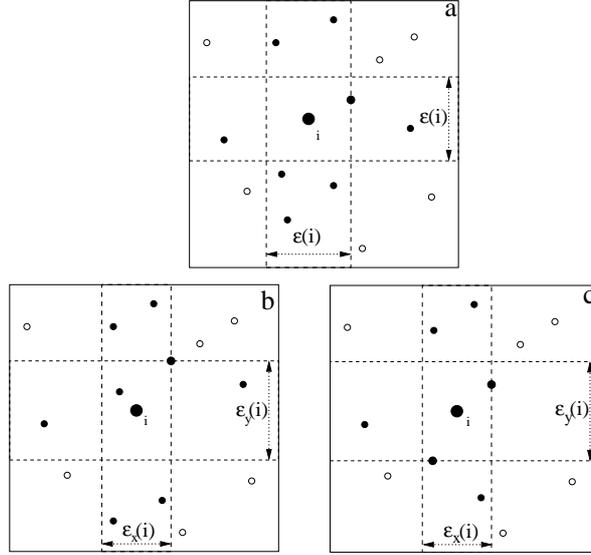
wobei  $\epsilon(i)$  der doppelte Abstand zwischen  $x_i$  und seinem  $k$ -ten Nachbarn ist.

Die einzige Näherung, die wir verwenden, ist Gl.(2.10). Wenn die Dichte  $\mu(x)$  konstant wäre, hätten wir also einen fehlerfreien Schätzer.

Mit Hilfe des Kozachenko-Leonenko Schätzers für die Entropie und Gl.(2.3) kann jetzt die MI abgeschätzt werden. Dafür betrachten wir eine zweidimensionale Zufallsvariable  $Z(X, Y)$ , die wiederum durch  $N$  Realisierungen vertreten wird, und  $\epsilon/2$  sei der Abstand zwischen einem Punkt  $z_i$  und dem  $k$ -ten Nachbarn. Die Gleichungen (2.8) und (2.9) gelten unverändert. Die erste Veränderung tritt in Gl.(2.10) auf, wo  $d$  durch  $d_Z = d_X + d_Y$ ,  $c_d$  durch  $c_{d_X} c_{d_Y}$  und natürlich  $x_i$  durch  $z_i = (x_i, y_i)$  ersetzt werden muss. Mit diesen Modifikationen erhalten wir

$$\begin{aligned} \hat{H}(X, Y) &= \psi(k) - \psi(N) - \log(c_{d_X} c_{d_Y}) \\ &\quad - \frac{d_X + d_Y}{N} \sum_{i=1}^N \log \epsilon(i) . \end{aligned} \quad (2.13)$$

Um  $I(X, Y)$  zu erhalten, muss nur noch Gl.(2.13) von den Schätzungen für  $H(X)$  und  $H(Y)$  subtrahiert werden. Für Letztere kann direkt Gl.(2.12) mit dem selben  $k$ -Wert verwendet werden. Dies würde aber bedeuten, dass wir die Entropien auf verschiedenen Skalen schätzen. Für jedes  $k$  würden dann der Abstand zum nächsten Nachbarn im zweidimensionalen Raum größer sein, als die Abstände zu den Nachbarn in den eindimensionalen Räumen. Da aber der Fehler in Gl.(2.12), durch die Ungleichverteilung der Dichten,



**Abbildung 2.1:** (a): Bestimmung von  $\epsilon(i)$ ,  $n_x(i)$  und  $n_y(i)$  für den ersten Algorithmus mit  $k = 1$  und einem fixen  $i$ . In diesem Beispiel ist  $n_x(i) = 5$  und  $n_y(i) = 3$ .

(b),(c): Bestimmung von  $\epsilon_x(i)$ ,  $\epsilon_y(i)$ ,  $n_x(i)$  und  $n_y(i)$  für den zweiten Algorithmus. (b) zeigt den Fall, bei dem  $\epsilon_x(i)$  und  $\epsilon_y(i)$  durch den selben Punkt bestimmt werden, während in (c) der Fall gezeigt wird, bei dem beide Größen durch zwei verschiedene Punkte bestimmt werden.

natürlich vom Abstand abhängt, würden sich die Fehler von  $\hat{H}(X)$ ,  $\hat{H}(Y)$  und  $\hat{H}(X, Y)$  nicht gegenseitig aufheben.

Um dies zu vermeiden, stellen wir fest, dass Gl.(2.12) für beliebige Wert von  $k$  gültig ist, und wir auch unterschiedliche  $k$ -Werte für die Schätzung von  $\hat{H}(X, Y)$  und  $\hat{H}(X)$  bzw.  $\hat{H}(Y)$  verwenden können. Im Folgenden werden zwei Varianten vorgestellt, die  $k$  für die verschiedenen Räume so wählen, dass der Fehler des Schätzers möglichst klein wird.

Im ersten Algorithmus, der in Abb.(2.1a) dargestellt ist, liegt der  $k$ -te Nachbar von  $x_i$  auf einer der Seiten des Quadrates mit der Kantenlänge  $\epsilon(i)$ , ohne Beschränkung der Allgemeinheit können wir annehmen, dass dies eine vertikale Seite ist. Wenn jetzt  $n_x(i)$  Punkte innerhalb der vertikalen Linien  $x = x_i \pm \epsilon(i)/2$  liegen, dann ist  $\epsilon(i)/2$  der Abstand zu den  $(n_x(i) + 1)$ -ten Nachbarn von  $x_i$  und

$$\begin{aligned} \hat{H}(X) &= \frac{1}{N} \sum_{i=1}^N \psi(n_x(i) + 1) - \psi(N) \\ &- \log c_{d_X} - \frac{d_X}{N} \sum_{i=1}^N \log \epsilon(i) . \end{aligned} \quad (2.14)$$

Für die andere Richtung (die  $y$  Richtung in Abb.(2.1a)) ist dies nicht exakt gültig, da  $\epsilon(i)/2$  nicht exakt der Abstand zum  $(n_y(i) + 1)$ -ten Nachbarn ist, wenn analog  $n_y(i)$  als die Anzahl

der Punkte innerhalb  $\|y_j - y_i\| < \epsilon(i)/2$  definiert wird. Trotzdem können wir Gl.(2.14) als gute Näherung für  $H(Y)$  betrachten, wenn alle  $X$  durch  $Y$  ersetzt werden. Einsetzen von Gl.(2.13) und Gl.(2.14) in Gl.(2.3) führt dann zu

$$I^{(1)}(X, Y) = \psi(k) - \langle \psi(n_x + 1) + \psi(n_y + 1) \rangle + \psi(N). \quad (2.15)$$

In gleicher Weise kann auch die MI für  $m$  Zufallsvariable abgeschätzt werden. Die höher dimensionale MI, auch allgemeine Redundanz genannt, kann wieder als eine Summe von Entropien geschrieben werden

$$\begin{aligned} I(X_1, X_2, \dots, X_m) &= H(X_1) + H(X_2) + \dots + H(X_m) \\ &- H(X_1, X_2, \dots, X_m) \end{aligned} \quad (2.16)$$

und somit ergibt sich

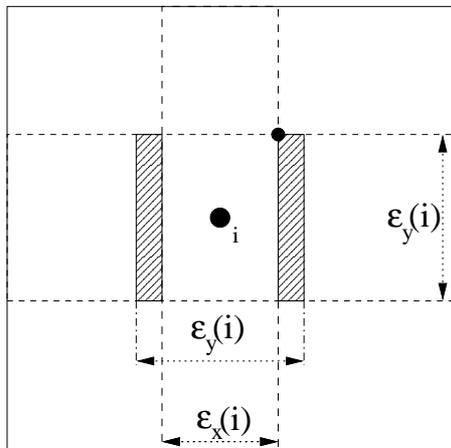
$$\begin{aligned} I^{(1)}(X_1, X_2, \dots, X_m) &= \psi(k) + (m - 1)\psi(N) \\ &- \langle \psi(n_{x_1}) + \psi(n_{x_2}) + \dots + \psi(n_{x_m}) \rangle. \end{aligned} \quad (2.17)$$

Wir werden im nächsten Abschnitt sehen, dass dieser Schätzer allen bis jetzt in der Literatur verfügbaren Schätzern für die MI überlegen ist und nur einen geringen systematischen Fehler besitzt. Trotzdem kann er noch verbessert werden, da die Verwendung von Quadraten bzw. (Hyper-)Kuben die Nachbarschaften unnötig groß macht. Vor allem für hochdimensionale Zufallsvariablen zahlt es sich daher aus, von (Hyper-)Kuben auf (Hyper-)Rechtecke überzugehen.

Wie in Abb.(2.1b) und (c) ersichtlich, müssen wir für die zweite Variante des Schätzers zwischen zwei Fällen unterscheiden. Entweder sind  $\epsilon_x(i)$  und  $\epsilon_y(i)$  durch einen Punkt (Abb.2.1b) oder durch zwei verschiedene Punkte (Abb.2.1c) bestimmt. Nachdem die Wahrscheinlichkeitsdichte aus Gl.(2.9) für diesen Fall adaptiert worden ist [34], erhält man einen neuen Schätzer für die allgemeine Redundanz

$$\begin{aligned} I^{(2)}(X_1, X_2, \dots, X_m) &= \psi(k) - (m - 1)/k \\ &+ (m - 1)\psi(N) \\ &- \langle \psi(n_{x_1}) + \psi(n_{x_2}) + \dots + \psi(n_{x_m}) \rangle. \end{aligned} \quad (2.18)$$

Bei der Herleitung ist leicht geschummelt worden, da wir wissen, dass im Fall  $\epsilon_x(i) < \epsilon_y(i)$  (wie in Abb.(2.1b),(c)) in den zwei Rechtecken  $[x_i - \epsilon_y(i)/2, x_i + \epsilon_x(i)/2] \times [y_i - \epsilon_y(i)/2, y_i + \epsilon_y(i)/2]$  und  $[x_i + \epsilon_x(i)/2, x_i + \epsilon_y(i)/2] \times [y_i - \epsilon_y(i)/2, y_i + \epsilon_y(i)/2]$  keine



**Abbildung 2.2:** Innerhalb der schattierten Rechtecke können keine Punkte liegen. Dies heißt für die Methode (2), dass die Schätzung für die Entropien  $H(X)$  ( $H(Y)$ ) eigentlich modifiziert werden müsste, da ein Teil der Fläche außerhalb (innerhalb) des Streifens  $\epsilon_x$  ( $\epsilon_y$ ) für Punkte verboten ist. Diese Abweichung wird in Gl.2.18 nicht berücksichtigt.

Punkte enthalten sein können (Abb.(2.2)). In der Schätzung für  $H(X, Y)$  wurde dies korrekt berücksichtigt, nicht aber in denen für  $H(X)$  und  $H(Y)$ . Die Korrekturterme sind von der Ordnung  $O(1/n_x)$  bzw.  $O(1/n_y)$ , und es ist wahrscheinlich, dass sich beide Terme zum Großteil aufheben, da sie mit unterschiedlichen Vorzeichen zu  $H(X)$  bzw.  $H(Y)$  beitragen. Wir haben aber keinen Beweis dafür. Unabhängig davon kann durch die Annahme der Konstanz der Wahrscheinlichkeitsdichte innerhalb des Rechtecks nicht erwartet werden, dass die Schätzer für endliche  $N$  exakt sind. Die Bestimmung des eigentlichen Fehlers von  $I^{(1)}$  und  $I^{(2)}$  beruht letztendlich auf numerischen Tests, die in Kap.(2.5) durchgeführt werden.

## 2.4 Praktische Überlegungen

Die MI ist invariant unter Reparametrisierung der einzelnen Zufallsvariablen. Wenn  $X' = F(X)$  und  $Y' = G(Y)$  Homomorphismen sind, dann gilt  $I(X, Y) = I(X', Y')$ . Im Gegensatz dazu ist aber die Entropie  $H(X)$  nicht invariant. Diese Eigenschaft der MI erlaubt es nicht nur, die beiden Variablen zum Beispiel auf gleiche Varianz zu bringen, sondern auch ihre Verteilungen so gleichförmig wie möglich zu gestalten. Dies ist zwar streng genommen für die nächste Nachbarsuche nicht nötig, reduziert aber trotzdem den numerischen Fehler des Schätzers, wenn die Dichten der Zufallsvariablen hohe Spitzen aufweisen, wie es zum Beispiel in der weiter unten diskutierten Gamma-Exponentialverteilung,  $\mu(x, y) = x^\theta \exp(-x - xy)/\Gamma(\theta)$  für  $x, y > 0$ , der Fall ist.

Den größten Teil der CPU-Zeit benötigt der Algorithmus für die Suche nach Nachbarn. In der einfachsten Implementierung werden zwei verschachtelte Schleifen über alle Punkte verwendet und somit eine CPU-Zeit der Ordnung  $O(N^2)$  benötigt. Diese Version ist aber nur für sehr kurze Messreihen ( $N$  kleiner als ungefähr 300) vertretbar. Einen Algorithmus der Komplexität  $O(N\sqrt{kN})$  erhält man, wenn zuerst die  $x_i$  nach ihrer Größe geordnet werden, was automatisch die Ordnung für die  $y_i$  vorgibt [52]. Die nächsten Nachbarn zu einem Punkt  $(x_i, y_i)$  werden dadurch gefunden, dass die an beiden Seiten von  $x_i$  gelegenen Punkte auf ihre Abstände in  $y$ -Richtung überprüft werden. Den schnellsten Algorithmus für die Nachbarsuche erhält man aber durch die Verwendung von Gittern [21, 25]. Der zweidimensionale Raum wird in Boxen der Größe  $O(\sqrt{k/N})$  unterteilt und die beiden eindimensionalen Räume in Boxen der Größe  $O(1/N)$ . Zuerst wird  $\epsilon$  bestimmt, indem die  $k$ -nächsten Nachbarn mit Hilfe des 2D-Gitters gefunden werden. Danach wird  $n_x$  und  $n_y$  durch die beiden anderen Gitter bestimmt. Falls die Verteilungen von  $X$  und  $Y$  glatt sind, führt dies zu einer Komplexität der Ordnung  $O(\sqrt{kN})$ .

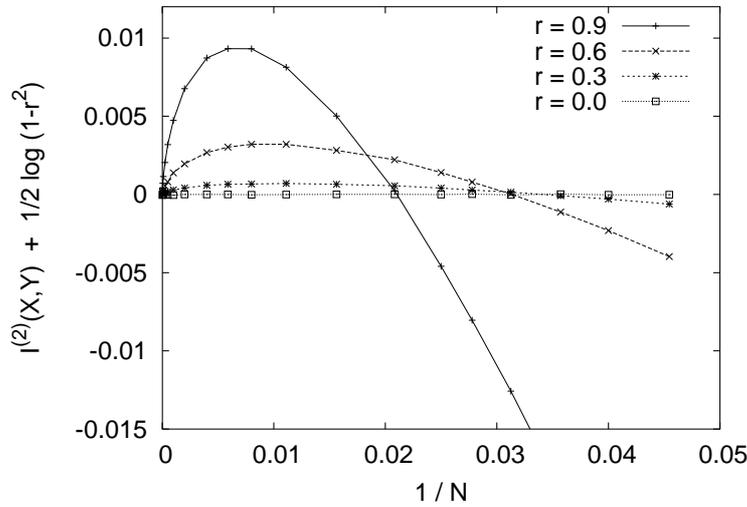
In der Datenaufzeichnung werden üblicherweise A/D-Wandler mit einer Auflösung von 12 bzw. 16 Bit verwendet. Dies führt für längere Datensätze zu vielen Punkten mit identischen Koordinaten. Die Bestimmung der  $k$ -nächsten Nachbarn ist dadurch nicht mehr eindeutig, weil manchmal Nachbarn exakt den gleichen Abstand vom untersuchten Punkt aufweisen. Durch Addieren von sehr geringem Rauschen (Amplitude viel niedriger als der kleinste Abstand in den Messsignalen) zu den Daten, kann diese Entartung aufgehoben werden. In allen untersuchten Fällen führte dies zu zufriedenstellenden Ergebnissen.

## 2.5 Numerische Simulationen

In diesem Kapitel wird gezeigt, wie sich der statistische und systematische Fehler der beiden Varianten des MI-Schätzers verhalten. Wir zeigen hier nur Ergebnisse für zweidimensionale Verteilungen, da Untersuchungen an höherdimensionalen Verteilungen sehr ähnliches Verhalten [37] für die Fehler zeigten. Im ersten Test wenden wir den Schätzer auf korrelierte Gaußverteilungen an, da für diesen Fall das analytische Ergebnis bekannt ist. Seien  $X$  und  $Y$  gaußverteilt mit verschwindendem Mittelwert, einer Varianz von eins und einer Kovarianz von  $r$ . Die MI ist dann [16]:

$$I_{\text{Gauss}}(X, Y) = -\frac{1}{2} \log(1 - r^2). \quad (2.19)$$

Abb.(2.3) zeigt den Mittelwert des Fehlers  $I^{(2)}(X, Y) - I_{\text{Gauss}}(X, Y)$  für verschiedene Werte von  $r$ , gemittelt über eine große Anzahl von Realisationen ( $10^5 - 10^7$ ). Es werden nur die Ergebnisse von  $I^{(2)}(X, Y)$  für  $k = 1$  gezeigt, da sie sich kaum von denen mit  $k > 1$  unterscheiden. Auch der Unterschied zu  $I^{(1)}(X, Y)$  ist so klein, dass er nicht in der



**Abbildung 2.3:** Schätzungen von  $I^{(2)}(X, Y) - I_{\text{exact}}(X, Y)$  für zwei gaußverteilte Zufallsvariable mit Varianz eins und Kovarianz  $r = 0.9, 0.6, 0.3$ , und  $0.0$ , aufgetragen gegen  $1/N$  ( $k = 1$ ). Für  $N \leq 1000$  betrug die Anzahl der Realisationen  $> 2 \times 10^6$ , die dann für  $N = 40000$  bis auf  $\approx 10^5$  verkleinert wurde. Die Fehlerbalken sind kleiner als die Symbole.

Abbildung erkennbar ist. Auffällig in Abb.(2.3) ist nicht nur die Konvergenz des systematischen Fehlers  $I^{(2)}(X, Y) - I_{\text{Gauss}}(X, Y) \rightarrow 0$  für  $N \rightarrow \infty$ , sondern auch, dass dieser für unkorrelierte Gaußverteilungen ( $r = 0$ ), unabhängig von  $N$ , verträglich mit Null ist.

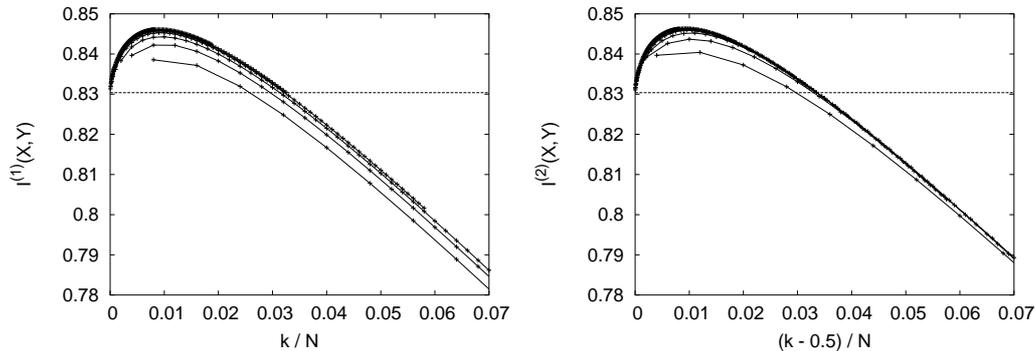
Um Unterschiede zwischen den beiden Variablen des Schätzers zu zeigen, werden in Abb.(2.4) die Werte von  $I^{(1)}(X, Y)$  mit denen von  $I^{(2)}(X, Y)$  für verschiedene Werte von  $N$  gegenübergestellt ( $r = 0.9$ ). An der Abszisse wird  $k/N$  bzw.  $(k - 1/2)/N$  aufgetragen. Abgesehen von sehr kleinen Werten für  $k$  und  $N$ , kann beobachtet werden, dass die Werte der Schätzer ein Skalierungsverhalten der Form

$$I^{(1)}(X, Y) \approx \Phi\left(\frac{k}{N}\right), \quad I^{(2)}(X, Y) \approx \Phi\left(\frac{k - 1/2}{N}\right) \quad (2.20)$$

zeigen.

Dies ist ein für verschiedene Verteilungen gültiges Ergebnis. Das Verhalten von  $I^{(1)}(X, Y)$  resultiert aus dem Umstand, dass die Anzahl von Nachbarn  $k$  innerhalb eines fixen Abstandes mit  $N$  skaliert. Für  $I^{(2)}(X, Y)$  muss dieses Argument leicht modifiziert werden, da das kleinere Epsilon,  $\epsilon_x$  bzw.  $\epsilon_y$ , meistens durch den  $(k - 1)$ -ten anstatt durch den  $k$ -ten Nachbarn festgelegt wird.

Der statistische Fehler der beiden Schätzer zeigt ebenfalls, dass  $I^{(2)}(X, Y)$  für ein gegebenes  $k$  zwischen den Werten von  $I^{(1)}(X, Y)$  mit  $k - 1$  und  $k$  liegt. Die Abb.(2.5) zeigt die Varianz der Schätzungen für zwei Gaußverteilungen mit der Kovarianz  $r = 0.9$ . Der statistische Fehler hängt nur sehr schwach von  $r$  ab und ist für  $r = 0$  etwa 10% kleiner.



**Abbildung 2.4:** Mutual Information Schätzungen  $I^{(1)}(X, Y)$  (linkes Bild) und  $I^{(2)}(X, Y)$  (rechtes Bild) für zwei gaußverteilte Zufallsvariable mit Kovarianz  $r = 0.9$ , aufgetragen gegen  $k/N$  (linkes Bild) bzw.  $(k - 1/2)/N$  (rechtes Bild). Jede Kurve entspricht einem fixen Wert für  $N$ , mit  $N = 125, 250, 500, 1000, 2000, 4000, 10000$  und  $20000$ , von unten nach oben. Die Fehlerbalken sind kleiner als die Symbole. Die gestrichelte Linie stellte den exakten Wert  $I(X, Y) = 0.830366$  dar.

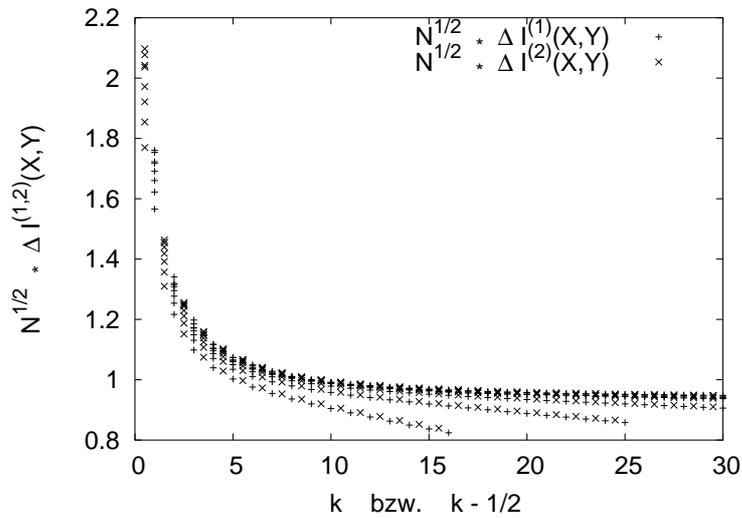
Aus Abb.(2.5) ist ersichtlich, dass der Fehler, außer für sehr große Werte von  $k/N$ , etwa mit  $\sim \sqrt{N}$  skaliert. Die Tatsache, dass der statistische Fehler für abnehmende  $k$  zunimmt, ist nicht verwunderlich, da auch die Verteilung der Epsilons breiter wird. Für praktische Anwendungen sollten daher  $k$ -Werte größer als 1 verwendet werden, um den statistischen Fehler zu reduzieren, aber auch nicht zu große Werte, weil dadurch auf der anderen Seite der systematische Fehler erhöht wird. Für die meisten Anwendungen wird es sinnvoll sein,  $k$ -Werte zwischen 3 und 5 zu wählen. Falls speziell auf Unabhängigkeit getestet werden soll, spielt der systematische Fehler, wie gezeigt, keine Rolle und es können Werte bis  $k \approx N/2$  verwendet werden.

Auch für andere Verteilungen ist es möglich, Werte der MI analytisch zu berechnen. In Ref. [37] werden Ergebnisse der beiden Schätzer für die Gamma-Exponentialverteilung [15], die sortierte Weinman-Exponentialverteilung [15] und die "Kreis-Verteilung" [14] diskutiert. Im Folgenden sind nur noch Resultate für die erste der drei Verteilungen gezeigt, da sich keine unterschiedlichen Schlussfolgerungen aus den letzten beiden ergeben. Die Wahrscheinlichkeitsdichte der Gamma-Exponentialverteilung ist für  $x > 0$  und  $y > 0$

$$\mu(x, y; \theta) = \frac{1}{\Gamma(\theta)} x^\theta e^{-x-xy}, \quad (2.21)$$

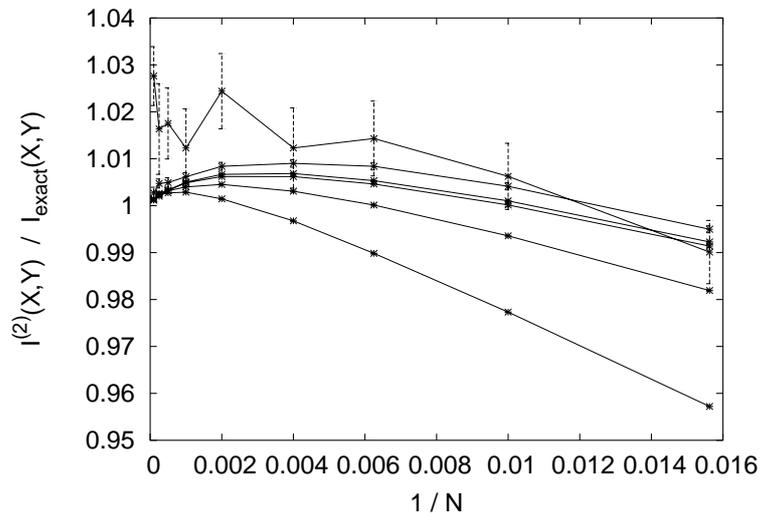
wobei  $\theta$  ein skalarer Parameter ist [15]. Für alle anderen Werte von  $x$  und  $y$  gilt  $\mu(x, y; \theta) = 0$ . Die MI ist durch  $I(X, Y)_{\text{exact}} = \psi(\theta + 1) - \log \theta$  bestimmt [15]. Da die Verteilung für  $\theta > 1$  eng um  $x = 0$  und  $y = 0$  lokalisiert ist, werden die Eingangsdaten für die Schätzer durch  $x'_i = \log x_i$  bzw.  $y'_i = \log y_i$  umskaliert, mit dem Ziel den numerischen Fehler zu verkleinern (siehe Kap.(2.4)). In Abb.(2.6) wird für fünf verschiedene Werte von

## 2.5. NUMERISCHE SIMULATIONEN

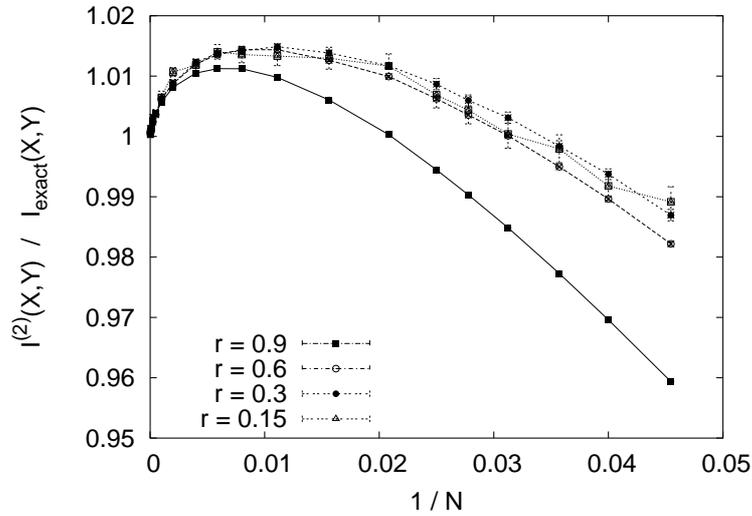


**Abbildung 2.5:** Standardabweichungen des Schätzers  $I^{(1)}(X, Y)$  (+) und  $I^{(2)}(X, Y)$  (×) für zwei gaußverteilte Zufallsvariable mit Kovarianz  $r = 0.9$ , multipliziert mit  $\sqrt{N}$  und dargestellt gegen  $k$  ( $I^{(1)}(X, Y)$ ) bzw.  $k - 1/2$  ( $I^{(2)}(X, Y)$ ). Jede Kurve entspricht einem festen Wert für  $N$ , mit  $N = 125, 250, 500, 1000, 2000, 4000, 10000$  und  $20000$ , von unten nach oben.

$\theta$  der relative Fehler  $I^{(2)}(X', Y') / I(X, Y)_{\text{exact}}$  für  $k = 1$  gegen  $1/N$  dargestellt. Es muss beachtet werden, dass der Abschnitt der  $y$ -Achse sehr klein ist, d.h. dass auch für diese Verteilung hervorragende Ergebnisse erhalten wurden. Zum Vergleich ist in Abb.(2.7) der relative Fehler für die Gaußverteilung mit den selben  $r$ -Werten gezeigt.



**Abbildung 2.6:** Verhältnis  $I(X, Y)_{\text{estim}} / I_{\text{exact}}(X, Y)$  für die Gamma-Exponentialverteilung, dargestellt gegen  $1/N$ . Diese Werte wurden, nach einer logarithmischen Transformation von  $x_i$  und  $y_i$ , mit  $I^{(2)}$  und  $k = 1$  erhalten. Die fünf Kurven entsprechen den Werten  $\theta = 0.1, 0.3, 1.0, 2.0, 10.0$  und  $100.0$  (von unten nach oben).



**Abbildung 2.7:** Verhältnis  $I^{(2)}(X, Y)/I_{\text{exact}}(X, Y)$  für  $k = 1$  gezeichnet gegen  $1/N$ , für 4 verschiedene Werte von  $r$ .

Das Verhalten des Schätzers in Abb.(2.6) als auch in Abb.(2.7) gibt Anlass, eine Annahme zu formulieren:

Sei  $\mu(x, y; r)$  eine zweidimensionale Verteilung mit  $r$  als skalaren Parameter. Weiters wird angenommen, dass  $\mu$  für  $r = r_0$  faktorisiere und in der Umgebung von  $r_0$  glatt ist, so dass  $\partial\mu(x, y; r)/\partial r$  endlich bleibt. Dann schlagen wir für viele Verteilungen, nicht notwendigerweise für alle, vor, dass

$$I^{(1,2)}(X, Y)/I_{\text{exact}}(X, Y) \rightarrow F(k, N) \quad (2.22)$$

gültig ist, wenn  $r \rightarrow r_0$ . Hierbei ist  $F(k, N)$  eine Funktion, die für alle  $k$  und  $N \gg 1$  in der Nähe von 1 verläuft und für  $N \rightarrow \infty$  gegen 1 konvergiert.

Wir wissen nicht, für welche Gruppen von Verteilungen diese Annahme zutrifft, aber wir fanden, dass  $I^{(1)}$  und  $I^{(2)}$  für unabhängige Zufallsvariablen  $X$  und  $Y$  für alle untersuchten Verteilungen exakt sind. Wir testeten auch gleichförmige und exponentielle Verteilungen und Daten mit unterschiedlichen Verteilungen für die beiden Zufallsvariablen und fanden immer, dass  $I^{(1,2)}(X, Y) = 0$  ist. Der statistische Fehler lag für diese numerischen Experimente in der Größenordnung von  $\approx 10^{-3}$  bis  $10^{-4}$ . Abschließend sei festgestellt, dass in Ref. [37] die systematischen und statistischen Fehler von  $I^{(1)}(X, Y)$  und  $I^{(2)}(X, Y)$  mit denen des MI-Schätzers von Darbelly und Vajda [16] verglichen werden, der in der Literatur als der genaueste und effizienteste Algorithmus gilt. Für alle untersuchten Verteilungen zeigt sich, dass der systematische Fehler von  $I^{(1,2)}$  gegenüber diesem Schätzer um etwa eine Zehnerpotenz verbessert wird. Der statistische Fehler verhält sich vergleichbar und zeigt für große  $N$  eine Verbesserung.

## 2.6 Eigenschaften der MI

Die Mutual Information besitzt eine Reihe von interessanten Eigenschaften. Die erste wichtige Eigenschaft, die bereits in Kap.(2.4) erwähnt wurde, ist die Unabhängigkeit von  $I(X, Y)$  gegenüber einer Umskalierung der Variablen. Jede glatte und eindeutig invertierbare Funktion ist für die Umformung  $X' = F(X)$  und  $Y' = G(Y)$  zulässig. Diese Eigenschaft kann leicht bewiesen werden, da die neue Wahrscheinlichkeitsdichte durch

$$\mu'(x', y') = J_X(x')J_Y(y')\mu(x, y) \quad (2.23)$$

gegeben ist, wobei  $J_X = \|\partial X/\partial X'\|$  und  $J_Y = \|\partial Y/\partial Y'\|$  die Jacobi-Determinanten sind. Weil für die eindimensionalen Dichten das Gleiche gilt, ergibt sich für die Mutual Information

$$\begin{aligned} I(X', Y') &= \iint dx' dy' \mu'(x', y') \log \frac{\mu'(x', y')}{\mu'_x(x')\mu'_y(y')} \\ &= \iint dx dy \mu(x, y) \log \frac{\mu(x, y)}{\mu_x(x)\mu_y(y)} \\ &= I(X, Y) . \end{aligned} \quad (2.24)$$

Diese Eigenschaft kann einerseits dazu verwendet werden, die Variablen so gleichförmig wie möglich zu verteilen, um numerische Fehler zu minimieren (Kap.2.4). Andererseits stellt sie eine Robustheit des Schätzers gegenüber gewissen Störungen des Signals im eigentlichen Messvorgang dar.

Die nächste Eigenschaft ("Gruppierungseigenschaft")

$$I(X, Y, Z) = I((X, Y), Z) + I(X, Y) . \quad (2.25)$$

kann ebenfalls aus der Definition der MI (Gl.(2.4)) hergeleitet werden. Sie ist analog zu dem Additionsgrundsatz der Shannon-Entropien [13] und besagt, dass die MI in hierarchische Ebenen zerlegt werden kann. Für  $n > 2$  kann  $I(X_1 \dots X_n)$ , mit einer beliebigen Aufteilung von  $(X_1 \dots X_n)$ , in die MI zwischen den Elementen innerhalb eines Clusters und die MI zwischen den Clustern zerlegt werden.

Eine weitere nützliche Eigenschaft ergibt sich, wenn Gl.(2.24) und (2.25) kombiniert werden

$$\begin{aligned}
 I(X', Y', Z) &= I((X', Y'), Z) + I(X', Y') \\
 &= I((X, Y), Z) + I(X', Y') \\
 &= I(X, Y, Z) + [I(X', Y') - I(X, Y)] .
 \end{aligned}
 \tag{2.26}$$

Hierbei wurde ausgenutzt, dass die MI auch invariant gegenüber Homomorphismen von höherdimensionalen Variablen  $(X', Y') = F(X, Y)$  ist. Diese Eigenschaft besagt, dass es für die Bestimmung der Änderung der Redundanz einer hochdimensionalen Variable, welche eine Transformation in einem Unterraum erfährt, ausreicht, nur die Änderung der MI in diesem Unterraum zu bestimmen. Dies reduziert den numerischen Aufwand für viele Anwendungen erheblich, siehe Kap.(3.2.3.4). Alle hier besprochenen Eigenschaften für die MI können auf jede beliebige Dimension der Variablen verallgemeinert werden.

## 2.7 Anwendungen der MI

### 2.7.1 Anhängigkeitsmaß für die Zeitreihenanalyse

Eine der zentralen Aufgaben in der Datenanalyse ist es, Abhängigkeit von Signalen zu testen. Sobald eine Messung aus mehreren Komponenten besteht, stellt sich die Frage, ob diese unabhängig sind oder gemeinsame Informationen tragen. Zum Beispiel ist für die Neurophysiologie interessant, ob zwei Hirnareale Information austauschen. Durch die Aufzeichnung der elektrischen Potentiale in den relevanten Bereichen des Gehirns erhält man Zeitreihen, die dann mit verschiedenen Anhängigkeitsmaßen untersucht werden können. In Ref. [53] werden die Kovarianz, die Kohärenz, die Phasensynchronisation, die nichtlineare Interdependenz sowie die Mutual Information für diesen Fall miteinander verglichen. Diese Verfahren unterscheiden sich in der ihnen zugrundeliegenden Definition von Abhängigkeit. Die ersten beiden, die Kovarianz und Kohärenz, sind nur auf lineare Abhängigkeiten sensitiv.

Nichtlineare Abhängigkeiten können mit den verbleibenden drei Verfahren gemessen werden. Für die Messung der Phasensynchronisation wird meist die Phase mit Hilfe einer Hilbert- oder Wavelettransformation aus den beiden Zeitreihen  $X(t)$  und  $Y(t)$  extrahiert [51]. Aus der Phasendifferenz  $\phi_{XY}(t) = \varphi_X(t) - \varphi_Y(t)$  kann dann mit unterschiedlichen statistischen Methoden ein Maß für die Phasensynchronisation gewonnen werden. Im Späteren wird die "mean phase coherence" aus Ref. [47] verwendet, die wie folgt definiert ist:

$$\phi_{xy}(t) = \varphi_x(t) - \varphi_y(t) .
 \tag{2.27}$$

Es kommen für diese Methode nur Anwendungen in Betracht, bei denen die Phase vernünftig extrahiert werden kann, dies ist z.B. für gekoppelte Oszillatoren erfüllt. Die nichtlineare Interdependenz hingegen wird über die Ähnlichkeit in der zeitlichen Entwicklung bestimmt. Dazu wird untersucht, ob ähnliche Zustände in der einen Zeitreihe zu den selben Zeiten auch ähnlichen Zuständen in der anderen entsprechen. Mit dem Begriff Zustand sind hier die Größen eines dynamischen Systems zu einem bestimmten Zeitpunkt gemeint, die durch eine Einbettung, aus der Zeitreihe approximiert werden [32]. Dieser Ansatz wird als Test auf generalisierte Synchronisation [51] verwendet.

Die MI kommt hingegen aus der Informationstheorie (Kap.(2.1)) und ist sensitiv auf jede Art von Abhängigkeiten. Wenn die Abhängigkeit von zwei Zufallsvariablen  $X$  und  $Y$  bestimmt werden soll, kann einfach  $I^{(1)}(X, Y)$  bzw.  $I^{(2)}(X, Y)$  berechnet werden. Falls wir es aber stattdessen mit zwei Zeitreihen  $X(t)$  bzw.  $Y(t)$  zu tun haben, sind im Allgemeinen die Messwerte zur Zeit  $t$  von vergangenen Messwerten abhängig und es können somit auch zeitliche Abhängigkeiten auftreten. Um jetzt sensitiv auf jede Art von Abhängigkeit zu sein, muss  $I^{(1,2)}((X(t), X(t-1), \dots, X(t-n)), ((Y(t), Y(t-1), \dots, Y(t-n))))$  berechnet werden, wobei der Parameter  $n$  durch die Ausdehnung der zeitlichen Korrelationen in den einzelnen Zeitreihen bestimmt ist.

Alle diese Maße versuchen Abhängigkeiten in Zeitreihen zu identifizieren. Durch ihre unterschiedlichen Konstruktionen hängt es aber immer von der Fragestellung ab, ob ein Maß geeignet ist. Hier zeichnet sich die Eigenschaft der MI aus, die besagt, dass sie dann und nur dann Null ist, wenn die untersuchten Zeitreihen absolut unabhängig sind, d.h. ihre Wahrscheinlichkeitsdichten faktorisieren und somit kein Informationsaustausch stattfand. Für alle anderen Maße ist dies nicht zutreffend, da sehr wohl zwei Zeitreihen abhängig sein können, ohne dass sich dies in der linearen Abhängigkeit, der Phasensynchronisation oder der nichtlinearen Interdependenz manifestiert. Somit ist die MI zu bevorzugen, wenn es um die Bestimmung der Unabhängigkeit geht, und genau in diesem Bereich liegt auch die größte Stärke der Schätzer  $I^{(1)}(X, Y)$  und  $I^{(2)}(X, Y)$ . Dieser Umstand macht diese Schätzer auch besonders attraktiv für die Independent Component Analyse (ICA) (Kap.3.2). Ein guter Überblick über die verschiedenen Abhängigkeitsmaße findet sich in Ref. [51].

### 2.7.2 Clusteranalyse

Es wird hier eine vom Konzept her sehr einfache Methode zur hierarchischen Clusteranalyse vorgestellt, die wir Mutual Information Clustering (MIC) [36] nennen wollen. Die Einfachheit des Algorithmus wird durch die Gruppierungseigenschaft der MI (Gl.(2.25)) ermöglicht. Sie erlaubt in jeder Ebene der Hierarchie, die Verzweigung der Cluster zu berechnen, so dass ein Zurückgreifen auf Ergebnissen aus tieferen Ebenen entfallen kann. Im

Folgenden sei eine kurze Einführung in die Clusteranalyse gegeben.

Die Klassifikation von Daten ist fundamental für das Verständnis in allen Disziplinen der Wissenschaften [29]. Eine Klassifikation kann überlappend oder nicht überlappend sein, beaufsichtigt oder unbeaufsichtigt erfolgen. Wir sind hier speziell an der Clusteranalyse interessiert, bei der die Klassifikation nicht überlappend und unbeaufsichtigt erfolgt.

Den Ausgangspunkt der Clusteranalyse stellt ein Satz von Objekten dar, deren Eigenschaften durch einen charakteristischen Vektor beschrieben werden. Die Aufgabe besteht jetzt darin, diese Objekte in Gruppen einzuteilen, wobei dafür ausschließlich die charakteristischen Vektoren verwendet werden. Aufgrund der Vielfältigkeit der Anwendungsmöglichkeiten, existieren auch eine große Anzahl von Methoden um Daten zu clustern (für einen Überblick siehe [29]). Hier wollen wir uns mit der hierarchischen Clustermethode (HC) beschäftigen, die die Objekte in verschachtelte Strukturen verknüpft.

Das Herz eines jeden Clusteralgorithmus bildet das benutzte Abstandsmaß. Dieses kann die Ähnlichkeit der charakteristischen Vektoren bestimmen, respektive ihre Unähnlichkeit. Es wird aber nicht strikt gefordert, dass das Abstandsmaß den Grundsätzen einer Metrik genügt (Positivität, Symmetrie, Dreiecksungleichung). In der Mehrzahl von HC-Methoden wird ein Abstandsmaß verwendet, welches nur für zwei Objekte definiert ist. Für diese Fälle können nur die Abstände zwischen den Objekten in der ersten Ebenen der Hierarchie berechnet werden und müssen für alle anderen Ebenen gemittelt werden. Im Gegensatz dazu, erlaubt die MI auch Abstände zwischen beliebig dimensional Objekten zu berechnen und somit auch auf jeder Ebene der Hierarchie. Da in der Gruppierungseigenschaft der MI (Gl.2.25)  $X, Y$  und  $Z$  selbst wieder Cluster höherer Dimension darstellen können, kann dies rekursiv angewandt werden und erlaubt somit die Berechnung einer vollständigen Hierarchie.

Die MI direkt als Abstandsmaß zu verwenden, bringt allerdings einige Probleme mit sich. Der Abstand wäre dann von der Menge an Information in den einzelnen Objekten abhängig. Es muss daher versucht werden, eine Normierung für die MI zu finden. Eine Möglichkeit hierfür stellt das Abstandsmaß

$$D(X, Y) = 1 - \frac{I(X, Y)}{H(X, Y)}, \quad (2.28)$$

in Ref. [44] dar, welches die Voraussetzung einer Metrik erfüllt und normiert ist:  $D(X, Y) \leq 1$  für alle Paare  $(X, Y)$ .

Für kontinuierliche Zufallsvariablen tritt das Problem auf, dass  $\tilde{H}(X, Y)$  (Gl.(2.2)) nicht invariant unter Reskalierung und nicht positiv definit ist, und andererseits  $H_{bin}$  für  $\Delta \rightarrow 0$

divergiert. Wir ersetzen daher in Gl.(2.28)  $H(X, Y)$  durch  $H_{bin}(X, Y)$  und  $D(X, Y)$  durch ein Ähnlichkeitsmaß

$$S(X, Y) = \lim_{\Delta \rightarrow 0} (D(X, Y) - 1) \log \Delta = \frac{I(X, Y)}{m_x + m_y} . \quad (2.29)$$

Dieses Maß berücksichtigt nun, dass unterschiedlich dimensionale Cluster auch unterschiedliche Mengen an Information tragen, aber nicht, dass Cluster mit gleicher Dimension sich auch in ihrem Informationsgehalt unterscheiden können. In vielen Anwendungen wird dies näherungsweise erfüllt sein, für sehr heterogene Daten muss versucht werden eine passende Transformation zu finden, die es erlaubt den Informationsgehalt der einzelnen Objekte zu homogenisieren.

Im Einzelnen ist der MIC-Algorithmus wie folgt aufgebaut:

- (1) Berechnung der Nachbarschaftsmatrix - paarweiser Abstand  $S(X_i, X_j)$  zwischen allen  $n$  charakteristischen Vektoren ( $n$  Objekte  $\equiv n$  Cluster)
- (2) Auffinden der zwei Cluster,  $k$  und  $l$ , mit dem geringsten Abstand
- (3) Bilden eines neuen Clusters ( $kl$ ) durch Vereinigen von  $k$  und  $l$
- (4) Entfernen der Zeilen/Spalten mit den Indices  $k$  und  $l$  aus der Nachbarschaftsmatrix und Einfügen einer neuen Zeile/Spalte, die die Abstände zwischen ( $kl$ ) und allen anderen Clustern enthält
- (5) Falls die Anzahl der Cluster  $n > 2$  ist, zurück zu (2); anderenfalls Verbinden der beiden Cluster; Abbruch

Eine Anwendung des MIC-Algorithmus wird in Kap.(4.1) gezeigt.



# Kapitel 3

## Behandlung multivariater Daten

In der experimentellen Untersuchung komplexer Systeme ist es häufig möglich, eine Vielzahl an Signalen simultan aufzunehmen. Die einzelnen Signale, im Späteren auch Kanäle oder Komponenten genannt, sind aber in den meisten Fällen nicht separat interpretierbar, d.h. die interessante Information ist über mehrere Kanäle verteilt. Eine anspruchsvolle Aufgabe in der Statistik und verwandter Gebiete ist es, eine passende Darstellung dieser multivariaten Daten zu finden, d.h. die Daten so zu transformieren, dass ihre essentielle Struktur sichtbar beziehungsweise zugänglich wird. Es wird versucht, Projektionen der Messdaten zu finden, die jeweils nur eine interessante Größe des Systems enthalten.

Die Mischungen können unterschiedlicher Natur sein, im Moment wollen wir uns aber auf lineare instantane Zusammenhänge beschränken, um das Konzept so klar wie möglich darzustellen.

Nehmen wir an, uns liege ein multivariates Messsignal  $x_j(t)$ ,  $j = 1, \dots, m$  vor, welches aus  $m$  Kanälen mit je  $T$  Beobachtungen besteht. Es können jetzt  $n$  lineare Mischung erzeugt werden

$$y_i(t) = \sum_{j=1}^m W_{ij} x_j(t) \quad i = 1, \dots, n . \quad (3.1)$$

In Matrixschreibweise ergibt sich eine kompaktere Darstellung

$$\mathbf{y} = \mathbf{W}\mathbf{x} . \quad (3.2)$$

Die  $n \times m$ -dimensionale Matrix  $\mathbf{W}$  kann jetzt unter verschiedenen Gesichtspunkten bestimmt werden, je nach der Natur der Messdaten und der Fragestellung.

Für manche Anwendungen ist es zum Beispiel sinnvoll, die Dimension von  $\mathbf{y}$  zu verkleinern ( $n < m$ ) und so die verfügbare Information auf eine geringere Anzahl von Komponenten zu komprimieren. Dies ist insbesondere dann von Interesse, wenn die Daten stark

verrauscht sind und so der Rauschpegel unterdrückt wird. Die hierfür verwendete Technik wird als Principle Component Analysis (PCA) oder Faktor Analyse bezeichnet. Als anschauliches Beispiel aus der Psychologie sei hier eine Arbeit von Spearman [62] angeführt. Er führte verschiedene Leistungstests an Schülern durch, wobei die Ergebnisse eines jeden Tests eine Komponente von  $\mathbf{x}$  darstellten. Es wurde dann mit Hilfe der PCA eine lineare Mischung gefunden, so dass die erste Komponente von  $\mathbf{y}$  so viel Information wie möglich von allen Tests enthält. Diese Komponente wurde dann von Spearman als genereller Faktor der Intelligenz bezeichnet. Diese Argumentation ist natürlich nicht ganz unumstritten.

Eine andere Technik behält alle Komponenten, versucht aber die transformierten Komponenten so statistisch unabhängig wie möglich zu machen. Das heißt, die Werte einer Komponente sollen so wenig Information wie möglich über die anderen Komponenten enthalten. Dieser Ansatz ermöglicht es, Mischungen von Quellsignalen wieder zu entmischen, und ist unter dem Namen "Blind Source Separation" (BSS) bekannt.

Wiederum eine andere Technik kombiniert eine starke Dimensionsreduktion mit der Suche nach möglichst interessanten Wahrscheinlichkeitsverteilungen der einzelnen Komponenten (ein Spezialfall sind wieder unabhängige Komponenten). Diese Methode ist unter dem Namen "Projection Pursuit" bekannt.

### 3.1 Blind Source Separation (BSS)

In vielen experimentellen Messsituationen besteht keine Möglichkeit, direkt die eigentlichen interessanten Systemgrößen zu messen, zum Beispiel wenn man keinen direkten Zugang zum Ursprung ihres Entstehens hat. Vielmehr werden mehrere Sensoren in dem zu untersuchenden System platziert, und es wird versucht, das so erhaltene multivariate Signal zu interpretieren. Beispiele für solche Situationen sind mehrere sprechende Personen in einem Raum, mehrere aktive Gehirnregionen zur selben Zeit, Mobiltelefone mit der selben Trägerfrequenz, die gleichzeitig emittieren, Infrarotspektren von chemischen Mischungen.

Mit der Nomenklatur des letzten Kapitels, wobei  $x_i(t)$  jetzt für die Messung am Sensor  $i$  zur Zeit  $t$  steht, können wir schreiben

$$\mathbf{x} = \mathbf{A}\mathbf{s} . \quad (3.3)$$

Hierbei steht  $s_i(t)$  für das Signal der Quellen, z.B. das Gesprochene nur einer Person, und  $\mathbf{A}$  ist eine Matrix, die angibt, wie eine Quelle an den Sensoren gemischt wird. Letztere hängt z.B. von den Positionen der Mikrophone und Personen ab. Sowohl  $\mathbf{A}$  als auch  $\mathbf{s}$  ist unbekannt, und die Aufgabe besteht darin, allein aus der Messung der Überlagerung  $\mathbf{x}$  beide Größen zu ermitteln. Unter der Annahme, dass  $\mathbf{A}$  nicht singulär (d.h. invertierbar) ist, suchen wir eine Matrix  $\mathbf{W}$ , die wieder die ursprünglichen Quellen erzeugt

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} . \quad (3.4)$$

Wenn die unabhängigen Quellsignale  $s_i(t)$  irgendeine Struktur enthalten, räumlicher oder zeitlicher Natur, d.h. keine gaußsche Verteilung mit deltaförmiger Autokorrelationsfunktion aufweisen, liefert die Transformation in statistisch unabhängige Komponenten eine eindeutige Lösung für  $\mathbf{W}$ . Genau betrachtet ist die Lösung nur eindeutig, bis auf eine Skalierungs- $\mathbf{S}$  und Permutationsmatrix  $\mathbf{P}$ , da die Aufgabe unterbestimmt ist:

$$\mathbf{x} = \mathbf{A}'\mathbf{s}' = (\mathbf{A}\mathbf{S}^{-1}\mathbf{P}^{-1})(\mathbf{P}\mathbf{S}\mathbf{s}) . \quad (3.5)$$

Wenn aber manche Quellsignale  $s_i(t)$  abhängig von anderen sind, muss dies in der Interpretation der geschätzten Quellsignale berücksichtigt werden. Da im vorhinein in keiner Anwendung sichergestellt werden kann, dass alle Quellen wirklich unabhängig sind und eine raum/zeitliche Struktur aufweisen, muss immer die Eindeutigkeit der gefundenen Lösung überprüft werden. Die im Folgenden vorgestellte Methode, ICA, erzeugt nur so unabhängige Komponenten wie möglich. Es muss daher nachträglich überprüft werden, ob tatsächlich eine BSS erzielt worden ist.

## 3.2 Independent Component Analysis (ICA)

### 3.2.1 Einführung

Independent Component Analysis ist eine statistische Methode, welche ein multivariates Signal in möglichst unabhängige Komponenten zerlegt. Dazu benötigt wird ein Schätzer für die Abhängigkeit von Signalen (Kontrastfunktion) und ein Algorithmus, der effizient die Komponenten mit der geringsten Abhängigkeit findet.

Die Verwandtschaft von ICA mit der PCA wird in Kapitel (3.2.2) diskutiert. Im Kapitel (3.2.3) werden die am häufigsten verwendeten Algorithmen sowie eigene Entwicklungen vorgestellt. Wie wir im letzten Kapitel gesehen haben, ist die Tatsache, dass der Minimierungsalgorithmus konvergiert, nicht ausreichend für die Interpretation der Ergebnisse, da die Quellen vielleicht nicht unabhängig sind oder das lineare Modell nicht passend ist. In Kapitel (3.2.4) werden die verbleibenden Abhängigkeiten der Komponenten und die Güte ihrer Separation untersucht. Praktische Überlegungen wie Filterung des Messsignals, Wahl des Algorithmus und Verhalten bei Messrauschen werden in Kapitel (3.2.6) besprochen. Im Folgenden sei hier noch ein geschichtlicher Überblick der ICA skizziert.

Die erste Anwendung fand dieser Ansatz 1982 in einer neurophysiologischen Fragestellung [31]. In einem einfachen Modell für die Bewegungskoordination wurde versucht, den Winkel  $s_1$  und die Geschwindigkeit  $s_2$  des bewegten Gelenks aus zwei EMG-Sensoren  $x_1$  und  $x_2$  zurück zu gewinnen. In den achtziger Jahren war ICA hauptsächlich unter französischen Wissenschaftlern bekannt. Es wurden vorwiegend Algorithmen entwickelt,

die Kumulanten höherer Ordnung verwendeten. Erst Mitte der 90er erhielt ICA größere Beachtung. Es wurden schnellere und robustere Algorithmen entwickelt, die es auch dank der fortgeschrittenen Computertechnik ermöglichten, ICA auf höherdimensionale Probleme anzuwenden [27]. Der erste internationale ICA-Workshop wurde 1999 in Frankreich abgehalten.

### 3.2.2 Unabhängig kontra unkorreliert

In der Alltagssprache besteht kein Unterschied in der Bedeutung der Begriffe unabhängig und unkorreliert, mathematisch aber sehr wohl. Im späteren werden wir unter Unkorreliertheit verstehen, dass zwei Zufallsvariablen  $y_1$  und  $y_2$  Kovarianz Null haben

$$\text{cov}(x_1, x_2) = \sum_j^N (x_{1j} - \sum_i^N x_{1i})(x_{2j} - \sum_i^N x_{2i}) = E\{x_1 x_2\} - E\{x_1\}E\{x_2\} = 0. \quad (3.6)$$

Im Sinne einer vereinfachten Darstellung, aber ohne Beschränkung der Allgemeinheit, setzen wir den Mittelwert der Zufallsvariablen  $E\{y_1\} = E\{y_2\} = 0$ . Somit heißt für uns unkorreliert einfach

$$E\{x_1 x_2\} = \sum_j^N x_{1j} x_{2j} = 0. \quad (3.7)$$

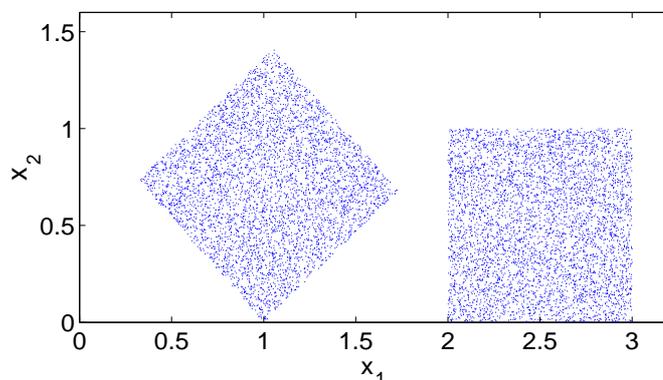
Im Gegensatz dazu sind zwei Zufallsvariablen unabhängig, wenn ihre Wahrscheinlichkeitsdichte faktorisieren.

$$p_{x_1 x_2}(x_1, x_2) = p_{x_1}(x_1) p_{x_2}(x_2). \quad (3.8)$$

Wie wir bereits in Gl.(2.4) gesehen haben, ist nur für diesen Fall  $I(x_1, x_2) = 0$ . Ist das gegeben, ist auch  $E\{x_1 x_2\} = 0$ . Zwei unabhängige Zufallsvariablen sind also auch unkorreliert, während der Umkehrschluss nicht zwingend gilt. Durch Integration von Gl.(3.8) erhält man außerdem

$$E\{g(x_1)h(x_2)\} = E\{g(x_1)\}E\{h(x_2)\} \quad (3.9)$$

wobei  $g$  und  $h$  beliebige integrable Funktionen sind. Der Vergleich der Definitionen von Unkorreliertheit (Gl.(3.7)) und Unabhängigkeit (Gl.(3.9)) zeigt, dass erste nur ein Spezialfall von der letzteren ist, nämlich wenn  $g$  und  $h$  linear sind. Die Unabhängigkeit zweier Zufallsvariablen ist eine stärkere Forderung, weil die Gleichung für alle integrablen



**Abbildung 3.1:** Zwei gleichförmig verteilte Zufallsvariable, wobei sich die linke von der rechten Abbildung nur durch eine Rotation unterscheidet. Beide Bilder stellen unkorrelierte Variable dar, aber nur rechts sind sie auch unabhängig.

Funktionen erfüllt sein muss. Ein anschauliches Beispiel ist in Abb.(3.1) gegeben. Beide Darstellungen zeigen unkorrelierte Variablenpaare. Im linken Bild erkennt man aber, dass Wissen über  $y_1$ , bereits Auskunft über den Wertebereich von  $y_2$  gibt und somit eine Abhängigkeit zwischen den beiden Variablen besteht. Das rechte Bild zeigt hingegen völlig unabhängige Zufallsvariable.

Oft ist es hilfreich, das Messsignal vor der Anwendung des ICA-Algorithmus durch eine Hauptachsentransformation zu dekorrelieren. Zusammen mit einer darauf folgenden Normierung wird dieser Schritt in der Literatur als *Whitening* bezeichnet.

Die Kovarianzmatrix von  $x$  kann mit Hilfe der Eigenvektormatrix  $\mathbf{E}$  und einer Diagonalmatrix der Eigenwerte  $\mathbf{D}$  umgeschrieben werden

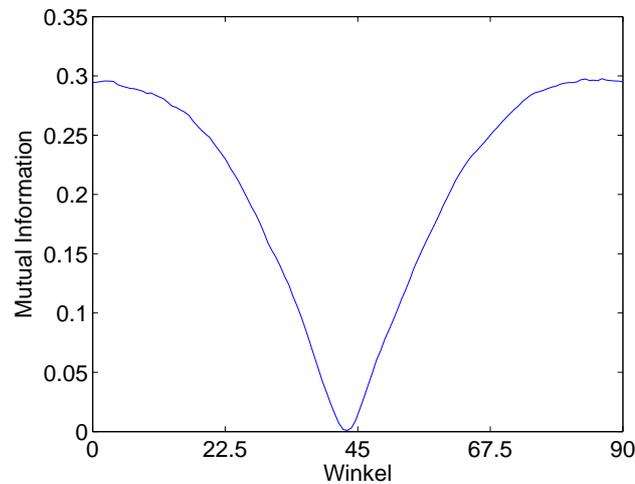
$$E\{\mathbf{xx}^T\} = \mathbf{E}\mathbf{D}\mathbf{E}^T . \tag{3.10}$$

Die Transformation ist dann durch

$$\mathbf{z} = \mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{x} = \mathbf{V}\mathbf{x} \tag{3.11}$$

bestimmt und liefert

$$\mathbf{C}_z = E\{\mathbf{zz}^T\} = \mathbf{1} . \tag{3.12}$$



**Abbildung 3.2:** Abhängigkeit der MI vom Rotationswinkel für die Datenpunkte aus Abb.(3.1 links). Für einen Winkel von etwa 43 wird ein Minimum der MI erreicht, welches in Abb.(3.1 rechts) dargestellt ist.

Die Matrix  $\mathbf{V}$  kann noch mit jeder beliebigen orthogonalen Matrix von links multipliziert werden, ohne die Kovarianzmatrix von  $\mathbf{z}$  zu verändern, d.h. eine beliebige Rotation hat keinen Einfluss auf die Korrelation der Komponenten: mit

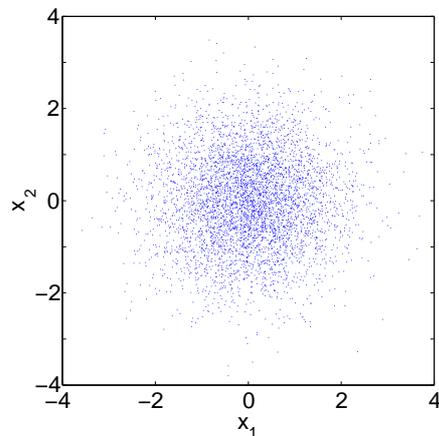
$$\mathbf{z}' = \mathbf{U}\mathbf{z} \quad (3.13)$$

gilt

$$E\{\mathbf{z}'\mathbf{z}'^T\} = \mathbf{U}E\{\mathbf{z}\mathbf{z}^T\}\mathbf{U}^T = \mathbf{U}\mathbf{1}\mathbf{U}^T = \mathbf{1} . \quad (3.14)$$

Diese Freiheit kann dazu benutzt werden, bereits dekorrelierte Quellen noch unabhängig zu machen. Jede Wahl von  $\mathbf{U}$  erzeugt ein unkorreliertes  $\mathbf{z}'$ -Signal, aber nur ein bestimmtes  $\mathbf{U}$  erzeugt im Allgemeinen unabhängige Signale. Die Aufgabe der ICA kann also auf das Auffinden einer orthogonalen Matrix  $\mathbf{R} \subset \mathbf{U}$  beschränkt werden.

Die Anzahl der zu bestimmenden Parameter wird durch das Whitening um etwa einen Faktor zwei reduziert, da eine orthogonale Matrix durch  $n(n - 1)/2$  Werte vollkommen bestimmt ist. Dies bedeutet für den zweidimensionalen Fall, dass nur noch ein Parameter, der Winkel der Drehmatrix, bestimmt werden muss. Verständlich wird dies auch durch Abb.(3.1), da sich die linke von der rechten Seite nur durch eine Rotation unterscheidet. Die Eindeutigkeit des passenden Rotationswinkels wird in Abb.(3.2) veranschaulicht. Dort sind die Werte des in Kap.(2.3) besprochenen Schätzers der MI gegen den Winkel aufgetragen. Ein ausgeprägtes Minimum ist für den Winkel erkenntlich, bei dem die Kante des Quadrates parallel zu den Achsen verläuft (Abb.(3.1) rechts), bei dem also die Zufallsvariablen minimal abhängig sind.



**Abbildung 3.3:** Zwei gaußverteilte Zufallsvariable

In der Einleitung ist erwähnt worden, dass ICA nur für Signale sinnvoll ist, deren Struktur nicht bereits durch die Korrelationsmatrix festgelegt ist. Letzteres ist für gauß'sche Variable der Fall, wenn sie zeitlich  $\delta$ -korreliert sind. Abb.(3.3) zeigt zwei Zufallsvariable mit gaußverteilter Wahrscheinlichkeitsdichte. Es ist ersichtlich, dass in diesem Fall jede Rotation unabhängige Komponenten liefert (wenn die MI nur zwischen gleichzeitig aufgezeichneten Punkten berechnet wird) und somit  $\mathbf{R}$  in diesem Fall nicht eindeutig bestimmt werden kann. Nur eine zeitliche Struktur im Signal, zum Beispiel eine Mischung aus weißem und rotem gauß'schen Rauschen, würde jetzt eine Entmischung ermöglichen. Hierfür müsste dann aber die MI auch zwischen zeitversetzten Punkten berücksichtigt werden.

Um die ursprünglichen Quellsignale  $\mathbf{s}$  und die Mischmatrix  $\mathbf{A}$  erfolgreich im linearen Fall abzuschätzen, können wir nun die nötigen Voraussetzungen zusammenfassen:  $\mathbf{A}$  muss vollen Rang haben und ist somit invertierbar,  $\mathbf{s}$  darf nicht mehr als einen Kanal weißes Rauschen enthalten und alle Quellsignale müssen unabhängig sein. Die letzte Forderung kann ein wenig abgeschwächt werden, mögliche Anwendungen werden in Kap. (3.2.6.2) diskutiert.

### 3.2.3 ICA-Algorithmen

Wie bereits eingangs erwähnt, wird für jeden ICA-Algorithmus ein Schätzer der Abhängigkeiten ("Kontrastfunktion") und ein effizienter Minimierungsalgorithmus benötigt. Im Allgemeinen kann man die verschiedenen Kontrastfunktionen in zwei Familien einteilen. Die erste beruht auf der Schätzung der Wahrscheinlichkeitsdichten Gl. (3.8) [6, 42, 65, 50]. Die andere Möglichkeit besteht darin, für  $g$  und  $h$  aus Gl.(3.9) geeignete Funktionen zu finden mit dem Ziel, ein möglichst großes Anwendungsspektrum abzudecken. Diese Möglichkeit, Abhängigkeiten abzuschätzen, wird auch durch den zentralen Grenzwertsatz motiviert, welcher besagt, dass eine Summe von Zufallsvariablen näher an einer Gaußverteilung ist

als die einzelnen Variablen. Es gibt daher zahlreiche Algorithmen, die Statistiken höherer Ordnung verwenden, um die Nähe zur Gaußverteilung abzuschätzen. Die günstigen mathematischen Eigenschaften dieser Statistik erlauben eine sehr effiziente Minimierung.

Im Laufe der Zeit ist eine große Anzahl von Algorithmen entwickelt worden (für einen Überblick siehe [27]). Wir stellen hier nur die häufigst verwendeten Algorithmen vor. Da in diesen mehr Wert auf Schnelligkeit als auf Genauigkeit gelegt worden ist, werden wir auch eigene Entwicklungen vorstellen, die diesem Umstand Rechnung tragen.

### 3.2.3.1 Bell-Sejnowski-Algorithmus und Abkömmlinge

Dieser Algorithmus verwendet den in der Statistik weitverbreiteten Maximum-Likelihood (ML) Schätzer. Die Annahme ist, dass die ML gerade die höchste Wahrscheinlichkeit für die Beobachtung der Quellsignale  $s_i$  ergibt. Die Likelihood für  $\mathbf{x}(t)$  wird nach Gl. (3.3) aus den Wahrscheinlichkeitsdichten der Einzelmessungen hergeleitet

$$p_x(\mathbf{x}) = |\det \mathbf{W}| p_s(\mathbf{s}) = |\det \mathbf{W}| \prod_i p_i(s_i) = |\det \mathbf{W}| \prod_i p_i(w_i^T \mathbf{x}). \quad (3.15)$$

Im vorletzten Schritt wurde die Unabhängigkeit der Quellsignale  $\mathbf{s}$  verwendet. Wenn wir nun  $T$  Messungen von  $\mathbf{x}$  vorliegen haben, ergibt sich die Likelihood aus dem Produkt der Dichten

$$L(\mathbf{W}) = \prod_{t=1}^T \prod_{i=1}^n p_i(\mathbf{w}_i^T \mathbf{x}(t)) |\det \mathbf{W}|. \quad (3.16)$$

Oft wird der Logarithmus von Gl. (3.16) verwendet, da Summen in der Maximierung leichter zu handhaben sind

$$\log L(\mathbf{W}) = \sum_{t=1}^T \sum_{i=1}^n p_i(\mathbf{w}_i^T \mathbf{x}(t)) + T |\det \mathbf{W}|. \quad (3.17)$$

Wir haben nun die Likelihood als Funktion der Matrix  $W$  ausgedrückt. Tatsächlich hängt sie aber auch von den Funktionen  $p_i$  ab. Zum Glück hat sich gezeigt, dass eine genaue Kenntnis der  $p_i$  nicht nötig ist. Dies folgt aus dem folgenden Theorem:

*Bezeichnen wir mit  $\tilde{p}_i$  vermutliche Wahrscheinlichkeitsdichten der unabhängigen Komponenten und sei*

$$g_i(s_i) = \frac{\partial}{\partial s_i} \log \tilde{p}_i(s_i) = \frac{\tilde{p}'_i(s_i)}{\tilde{p}_i(s_i)}. \quad (3.18)$$

### 3.2. INDEPENDENT COMPONENT ANALYSIS (ICA)

---

Fordern wir von den Schätzungen der unabhängigen Komponenten  $y_i = \mathbf{b}_i^T \mathbf{x}$ , dass sie unkorreliert sind und Varianz eins haben. Dann ist der ML-Schätzer mit  $p_i$  ersetzt durch  $\tilde{p}_i$  lokal konsistent, wenn jedes  $\tilde{p}_i$  die Gleichung

$$\Phi\{\tilde{p}_i\} \equiv E\{s_i g_i(s_i) - g'(s_i)\} > 0 \quad (3.19)$$

erfüllt.

Um das Theorem anzuwenden, brauchen wir nur zwei Funktionen  $\tilde{p}_i^+(s)$  und  $\tilde{p}_i^-(s)$ , von denen wir wissen, dass  $\Phi\{\tilde{p}_i^+\}$  und  $\Phi\{\tilde{p}_i^-\}$  unterschiedliche Vorzeichen haben. Denn dann können wir in jedem Maximierungsschritt diejenige der beiden Funktionen wählen, für die  $\Phi\{\tilde{p}_i^\pm\}$  positiv ist (der Fall  $\Phi = 0$  hat Wahrscheinlichkeit Null).

Zwei solcher Funktionen sind z.B.

$$\begin{aligned} \log \tilde{p}_i^+(s) &= \alpha - 2 \log \cosh(s) \\ \log \tilde{p}_i^-(s) &= \beta - [s^2/2 - \log \cosh(s)] \end{aligned} \quad (3.20)$$

gegeben.

Für die Herleitung des Bell-Sejnowski-Algorithmus muss zunächst der Gradient von Gl. (3.17) berechnet werden

$$\frac{1}{T} \frac{\partial \log L}{\partial \mathbf{W}} = [\mathbf{W}^T]^{-1} + E\{\mathbf{g}(\mathbf{W}\mathbf{x})\mathbf{x}^T\}. \quad (3.21)$$

Hier ist  $g$  eine vektorwertige Funktion der  $\tilde{p}_i$ , mit den Komponenten

$$g_i(\mathbf{W}\mathbf{x}) = (\log \tilde{p}_i(\mathbf{W}\mathbf{x}))' = \frac{\tilde{p}_i'(\mathbf{W}\mathbf{x})}{\tilde{p}_i(\mathbf{W}\mathbf{x})} = \begin{cases} -2 \tanh(\mathbf{W}\mathbf{x}) & f(\mathbf{W}\mathbf{x}) > 0 \\ \tanh(\mathbf{W}\mathbf{x}) - \mathbf{W}\mathbf{x} & f(\mathbf{W}\mathbf{x}) < 0 \end{cases} \quad (3.22)$$

wobei  $f(\mathbf{W}\mathbf{x}) = E\{-\tanh(\mathbf{W}\mathbf{x})\mathbf{W}\mathbf{x} + (1 - \tanh(\mathbf{W}\mathbf{x})^2)\}$  ist. Die Matrix  $W$  wird somit in jedem Maximierungsschritt um

$$\Delta \mathbf{W} \propto [\mathbf{W}^T]^{-1} + E\{\mathbf{g}(\mathbf{W}\mathbf{x})\mathbf{x}^T\}. \quad (3.23)$$

aktualisiert. Dieser Algorithmus ist nicht sehr effizient, da in jedem Schritt  $W$  invertiert werden muss. Die Konvergenz kann verbessert werden, indem das Eingangssignal  $x$  vorher *gewhited* wird. Eine weitere Verbesserung kann durch Verwendung des Natürlichen-Gradienten-Verfahrens

$$\Delta \mathbf{W} \propto (\mathbf{1} + E\{\mathbf{g}(\mathbf{y})\mathbf{y}^T\})\mathbf{W} \quad (3.24)$$

erreicht werden. Da die Skalierung von  $W$  nicht von Bedeutung ist, kann in dieser Schreibweise gut erkannt werden, dass der Algorithmus konvergiert, wenn  $E\{\mathbf{g}(\mathbf{y})\mathbf{y}^T\} = \mathbf{1}$  erfüllt ist. Das heißt,  $y_i$  und  $g_j(y_j)$  sind unkorreliert für  $i \neq j$ . Dies ist die nichtlineare Erweiterung der Kovarianz Gl. (3.9) und unterstreicht, dass Unabhängigkeit mehr als nur lineare Unkorreliertheit bedeutet.

Es wird in Ref. [27] gezeigt, dass dieser Algorithmus ähnlich dem Informationsmaximierungs-Prinzip (infomax) in der Neuronalen-Netzwerk Literatur ist. Darüber hinaus wird dargestellt, dass der ML auch mit Hilfe eines Fixpunkt-Algorithmus (siehe Kap. (3.2.3.2)) gefunden werden kann, der sehr gute Konvergenzeigenschaften besitzt.

### 3.2.3.2 FastICA

Anstelle die Abhängigkeiten zwischen den Komponenten abzuschätzen, kann alternativ auch die Abweichung der einzelnen Komponenten von einer Gaußverteilung berechnet werden und als Kontrastfunktion in einem ICA-Algorithmus verwendet werden. Dieser Ansatz ermöglicht es, die Komponenten nacheinander abzuschätzen, so dass bei Bedarf nicht alle Komponenten berechnet werden müssen. Die Abweichung von einer Gaußverteilung kann beispielsweise durch Momente höherer Ordnung abgeschätzt werden, welche einfach numerisch zu berechnen sind und außerdem günstige mathematische Eigenschaften besitzen. Aus diesem Grund findet man die zeitlich effizientesten ICA-Algorithmen, darunter auch FastICA, in dieser Kategorie.

Die einfachste Variante von FastICA verwendet die Kurtosis, um Abweichungen vom gauß'schen Verhalten abzuschätzen:

$$kurt(y) = Ey^4 - 3(Ey^2)^2 \begin{cases} > 0 & \text{subgaussisch, z.B. gleichförmige Verteilung} \\ < 0 & \text{supergaussisch, z.B. Laplaceverteilung} \end{cases} \quad (3.25)$$

Unter der Annahme, dass unser Eingangssignal unkorreliert ist und Varianz eins hat, ergibt sich für die Ableitung der Kurtosis

$$\frac{\partial |kurt(\mathbf{w}^T \mathbf{z})|}{\partial \mathbf{w}} = 4 \text{sign}(kurt(\mathbf{w}^T \mathbf{x})) [E\{(\mathbf{w}^T \mathbf{z})^3\} - 3\mathbf{w}\|\mathbf{w}\|^2] . \quad (3.26)$$

Da wir nur an Richtungsänderungen des Vektors  $\mathbf{w}$  interessiert sind, können wir den zweiten Term der Ableitung verwerfen. Außerdem spielen multiplikative Faktoren keine Rolle, da wir ein  $\mathbf{w}$  auf dem Einheitskreis suchen und nach jeder Iteration den Vektor durch seine

Länge dividieren. Somit folgt die Aktualisierungsregel des Gradientenalgorithmus, der den Betrag der Kurtosis jeder Komponente maximiert

$$\begin{aligned}\Delta \mathbf{w} &\propto \text{sign}(\text{kurt}(\mathbf{w}^T \mathbf{z})) E\{\mathbf{z}(\mathbf{w}^T \mathbf{z})^3\} \\ \mathbf{w} &\leftarrow (\mathbf{w} + \Delta \mathbf{w}) / \|\mathbf{w} + \Delta \mathbf{w}\| .\end{aligned}\tag{3.27}$$

Dieser Algorithmus kann auch leicht in eine online<sup>1</sup>-Version umgeschrieben werden, welche den Vorteil hat, auch für nichtstationäre Umgebungen geeignet zu sein. Diese besitzt aber eine langsamere Konvergenzgeschwindigkeit, die stark von der Wahl der Lernrate abhängt und somit einen kritischen Parameter darstellt. Eine Alternative ist die Verwendung eines Fixpunkt-Algorithmus [28], der parameterfrei ist und sich durch ein gutes Konvergenzverhalten auszeichnet. Hierfür wird der Gradient der Kurtosis Gl. (3.26) direkt als neue Richtung für  $\mathbf{w}$  verwendet

$$\begin{aligned}\Delta \mathbf{w} &\propto E\{\mathbf{z}(\mathbf{w}^T \mathbf{z})^3\} - 3\mathbf{w} \\ \mathbf{w} &\leftarrow \mathbf{w} / \|\mathbf{w}\| .\end{aligned}\tag{3.28}$$

Dies ist die einfachste FastICA-Variante und findet vor allem für störungsfreie Signale Verwendung. Sobald Störungen wie Ausreißer in den Messdaten vorhanden sind, ist es oft notwendig, andere Schätzer für die Abweichung von der Gaußverteilung zu verwenden, da die Kurtosis durch die Verwendung der vierten Potenz fehleranfällig ist. Es wird in Ref. [28] gezeigt, dass hierfür eine bestimmte Klasse von Funktionen besonders gut geeignet ist (Herleitung siehe [27]). Im Folgenden ist das Gerüst des FastICA-Algorithmus dargestellt,

1. Subtrahieren des Mittelwerts von den Daten
2. Whiten der Daten  $\mathbf{z} = \mathbf{V}\mathbf{x}$
3. Initialisieren von  $\mathbf{w}$
4. Aktualisieren von  $\mathbf{w} \leftarrow E\{\mathbf{z}g(\mathbf{w}^T \mathbf{z})\} - E\{\mathbf{z}g'(\mathbf{w}^T \mathbf{z})\}\mathbf{w}$ , wobei ein  $g$  aus Gl.(3.29) gewählt wird
5. Projektion auf den Einheitskreis  $\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\|$
6. Zurück zu Punkt 4, bis Konvergenz erreicht ist

---

<sup>1</sup>Jeder neue Messpunkt kann sofort im Algorithmus verwendet werden. Im Gegensatz dazu müssen bei einem Algorithmus in Blockverarbeitung (Gl.(3.27)) alle Messpunkte vorher bekannt sein.

wobei für die Funktion  $g$  eine gewünschte Schätzfunktion aus Gl.(3.29) für die Abweichung von der Gaußverteilung eingesetzt werden kann.

$$\begin{aligned} g_1(y) &= \tanh(y) \\ g_2(y) &= y \exp(-y^2/2) \\ g_3(y) &= y^3 \end{aligned} \tag{3.29}$$

Der Algorithmus findet nur eine unabhängige Komponente. Für die Suche der weiteren Komponenten wird die Eigenschaft ausgenutzt, dass die gesuchte Matrix  $W$  eine Rotationsmatrix ist. Somit müssen einfach nur die erhaltenen Komponenten nach jedem Iterationsschritt orthogonalisiert werden. Hier kann man zwei Vorgehensweisen unterscheiden. Die erste Methode sucht nach den unabhängigen Komponenten hintereinander, das heißt jede neue Komponenten wird zum Beispiel mit der Gram-Schmidt-Methode orthonormal auf alle bisherig gefunden erzeugt. Die zweite Methode sucht alle Komponenten gleichzeitig, so dass nach jedem Iterationsschritt immer alle Komponenten aufeinander orthogonal erzwungen werden. Diese Variante zeigt zwar meistens ein schlechteres Konvergenzverhalten, hat aber den Vorteil, dass keine Fehler aufsummiert werden.

FastICA stellt zur Zeit den am meisten verwendeten Algorithmus dar, weil er, wie der Name schon suggerieren will, schnell konvergiert und somit auch für höherdimensionale Probleme geeignet ist. Wie wir im nächsten Kapitel sehen werden, gibt er aber im Vergleich zu anderen ICA-Algorithmen nicht die präzisesten Abschätzungen für die unabhängigen Komponenten. Algorithmen, die mehr auf die Genauigkeit der Schätzung der Abhängigkeiten bedacht sind, erlauben auch eine exaktere Ermittlung der Quellsignale [6, 42, 65]. Durch diesen Mehraufwand verlängert sich aber auch die Rechenzeit, so dass besonders für höherdimensionale Probleme oft FastICA bevorzugt wird.

### 3.2.3.3 TDSEP

Bis jetzt haben wir ICA-Algorithmen besprochen, denen die Annahme zugrunde liegt, dass die Quellsignale unabhängige Realisationen eines stationären Zufallsprozesses sind, das heißt, es ist keinerlei Information zwischen Messungen enthalten, sie sind zufällig verteilt. In vielen Situationen kommt es aber vor, dass aufeinanderfolgenden Messungen korreliert sind. Beispiele hierfür sind etwa Audiosignale, biomedizinische Daten, Bilddaten und Spektren. In diesem Fall lässt sich diese zeitliche Information zur Formulierung neuer ICA-Algorithmen verwenden. Im Kapitel (3.2.2) haben wir gesehen, dass eine Unkorreliertheit der Signale noch nicht ausreichend ist, um unabhängige Komponenten zu erhalten. Diese Unkorreliertheit bezog sich aber immer nur auf simultan aufgezeichnete Messpunkte. Wenn man nun diese Unkorreliertheit auch für zu unterschiedlichen Zeitpunkten aufgezeichnete Signale fordert, kann durch alleinige Verwendung von Statistiken zweiter Ordnung eine Zerlegung in unabhängige Komponenten erreicht werden. Dabei muss man

allerdings fordern, dass die Mischmatrix  $\mathbf{A}$  instantan ist und somit eine Verallgemeinerung auf zeitverzögerte Mischungen nicht möglich ist.

Im theoretischen Fall würden bereits zwei Zeitpunkte ausreichen, die simultan unkorreliert gemacht werden, um einen ICA-Algorithmus zu definieren [46]. Dieser Ansatz erlaubt zwar eine sehr einfache und schnelle Berechnung der Komponenten, da nur zwei Hauptachsentransformationen durchgeführt werden müssen, ist aber aufgrund der unterschiedlichen Zeitstrukturen in Messsignalen und des zu wählenden fixen Zeitintervalls für viele Anwendungen ungeeignet. Vielmehr hat sich gezeigt, dass eine simultane Diagonalisierung der Korrelationsmatrizen mehrerer Zeitintervalle, typischerweise 20, zu einem robusten und effizienten Algorithmus führt. Implementierungen dieses Ansatzes sind in der Literatur unter zwei Namen bekannt SOBI [7] und TDSEP [79], die unterschiedliche Minimierungsverfahren verwenden. Hier sei in aller Kürze TDSEP (Temporal Decorrelation source SEparation) vorgestellt. Die Kontrastfunktion setzt sich aus dem Beitrag zur instantanen Dekorrelation und der zeitlich Dekorrelation zusammen

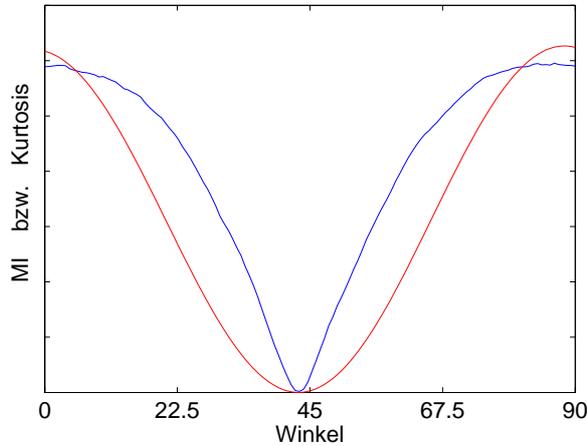
$$J(\mathbf{y} = \mathbf{B}\mathbf{x}) = \sum_{i \neq j} \langle y_i(t) y_j(t) \rangle^2 + \sum_{k=1}^N \sum_{i \neq j} \langle y_i(t) y_j(t + \tau_k) \rangle^2. \quad (3.30)$$

Um diese Funktion zu minimieren, wird ein Trick verwendet, der von Cardoso vorgeschlagen wurde [11]. Dieser basiert auf der von Jacobi verwendeten Zerlegung einer Rotationsmatrix in elementare zweidimensionale Rotationen zur Erzeugung von Diagonalmatrizen. Auf diese Weise kann das Minimum von Gl.(3.30) effizient gefunden werden.

Der TDSEP-Algorithmus eignet sich also nur zur Trennung von Quellsignalen, die nicht unabhängig verteilt sind. Diese müssen nicht unbedingt nur Zeitreihen sein. So können auch zum Beispiel Überlagerungen von Bildern getrennt werden, weil ein Bildpunkt meistens in Helligkeit und Farbe seinen Nachbarn ähnlich ist. Es ist auch nahe liegend, Algorithmen wie den Bell-Sejnowski oder FastICA, die ausschließlich die instantane Information im Signal verwenden, mit Algorithmen wie TDSEP zu verbinden, um bestmögliche Separationen zu erzielen. Ein solcher Ansatz wird im Kapitel (4) verfolgt und ist besonders für Messsignale geeignet, die starkes Rauschen der Quellen und/oder eine steil abfallende Autokorrelation besitzen.

#### 3.2.3.4 MI basierte Least Dependent Component Analysis (MILCA)

Im Prinzip können die Kontrastfunktionen der FastICA und des Bell-Sejnowski-Algorithmus als sehr grobe Approximationen der Mutual Information gesehen werden (Kumulantenexpansion siehe [27]). Da der absolute Wert der Abhängigkeiten für ein Minimierungsproblem irrelevant ist und es nur auf die richtige Ordnung der Abhängigkeiten ankommt, ist eine solche Approximation zulässig. Dennoch kann erwartet werden, dass die Verwendung eines exakten Schätzers für die MI die Genauigkeit eines ICA-Algorithmus erhöht.



**Abbildung 3.4:** Verhalten der Kurtosis und MI unter Rotation einer zweidimensionalen gleichförmigen Verteilung, siehe Abb.(3.1).

In Abb.(3.4) wird für zwei gleichförmig verteilte Zufallsvariablen das Verhalten bei einer Rotation der Datenpunkte für die Kurtosis mit der des MI-Schätzers aus Kap.(2.3) verglichen. Wie zu erwarten, zeigen beide Abhängigkeitsmaße ein ähnliches Verhalten, es ist aber ersichtlich, dass die MI ein schärfer ausgeprägtes Minimum zeigt und so im Mittel eine genauere Bestimmung des tatsächlichen Winkels erlaubt. Darüber hinaus erlaubt die Verwendung eines absoluten Abhängigkeitsmaßes, wie die MI, auch direkt eine Abschätzung der Zuverlässigkeit der ICA-Ausgabe (Kap.(3.2.4)). Durch diese Vorzüge der MI motiviert formulieren wir jetzt einen neuen ICA-Algorithmus, der den MI-Schätzer der vorherigen Kapitel verwendet.

Nehmen wir wieder an,  $\mathbf{x}$  besteht aus  $n$  Messungen mit jeweils  $T$  Messpunkten, die wiederum Gl.(3.3) genügen. Des Weiteren wurde  $\mathbf{x}$  durch eine Hauptachsentransformation und eine Skalierung in  $\mathbf{z}$  transformiert, so dass  $C(\mathbf{z}) = \mathbf{1}$  ist. Das Ziel ist jetzt  $I(z_1 \dots z_n)$  unter einer Rotationsmatrix  $R$  zu minimieren. Jede höherdimensionale Rotation kann in ein Produkt aus Rotationen zerlegt werden, welche nur in einem  $2 \times 2$  Unterraum agieren

$$\mathbf{R} = \prod_{i,j} \mathbf{R}_{ij}(\phi) , \quad (3.31)$$

wobei

$$\mathbf{R}_{ij}(\phi)(x_1 \dots x_i \dots x_j \dots x_n) = (x_1 \dots x'_i \dots x'_j \dots x_n) \quad (3.32)$$

mit

$$x'_i = \cos \phi x_i + \sin \phi x_j, \quad x'_j = -\sin \phi x_i + \cos \phi x_j . \quad (3.33)$$

ist.

Die minimal abhängigen Komponenten können jetzt dadurch erhalten werden, dass iterativ über alle Kanäle jeweils in zweidimensionalen Unterräumen das Minimum der MI gefunden wird, solange bis das globale Minimum erreicht wird. Im Prinzip müßten auch für Rotationen in den Unterräumen, nach jedem Iterationsschritt die allgemeinen Redundanzen  $I(z_1 \dots z_n)$  berechnet werden. Glücklicherweise kann aber der numerische Aufwand durch Verwendung der Eigenschaft (Gl.(2.26))

$$I(\mathbf{R}_{ij}(\phi)\mathbf{X}) - I(\mathbf{X}) = I(X'_i, X'_j) - I(X_i, X_j) \quad (3.34)$$

erheblich verringert werden, da sie es erlaubt, die MI nur in den Unterräumen zu berechnen in denen gedreht worden ist.

Zur Bestimmung des optimalen Rotationswinkels in einer Ebene  $(i, j)$  wird  $\hat{I}_{ij}(\phi) = \hat{I}(z'_i, z'_j)$  an üblicherweise 128 verschiedenen Winkeln im Intervall  $[0, \pi/2]$  berechnet. Die Verwendung der  $k$ -nächsten Nachbar-Statistik im Schätzer bewirkt aber, dass die Werte der MI über die Winkeln keine glatte Funktion ergibt, da bei sehr kleinen Winkeländerungen sich die Nachbarschaftsverhältnisse sprunghaft ändern können. In numerischen Simulationen hat sich gezeigt, dass das Minimum am zuverlässigsten gefunden werden kann, wenn  $\hat{I}(z'_i, z'_j)$  durch 3 bis 15 Fourierkomponenten gefittet wird und dann das Minimum des Fits berechnet wird. Ein weiterer Trick zur Glättung wird in Ref. [42] vorgestellt, wobei einzelne Punkte durch Punktwolken ersetzt werden. Diese Vermehrung der Punkte hilft den relativen numerischen Fehler der  $\hat{I}(z'_i, z'_j)$  weiter zu verkleinern. Die erzielte Verbesserung ist recht bescheiden, wie wir in Kap. (3.2.4) sehen werden, und verursacht einen beträchtlichen numerischen Mehraufwand (Faktor 30). Somit muss im Einzelfall entschieden werden, ob dieser Aufwand gerechtfertigt ist.

Der MILCA-Algorithmus kann jetzt zusammengefasst werden:

1. Vorverarbeitung (Mittelwertentfernung, Filterung, Trendentfernung, ...) und *whiten* der Daten;
2. Für jedes Paar  $(i, j)$  mit  $i, j = 1 \dots n$  Auffinden des Winkels  $\phi$ , der den geglätteten Fit von  $\hat{I}_{ij}(\phi) = \hat{I}(Z'_i, Z'_j)$  minimiert;
3. Wenn  $\hat{I}(Z'_1 \dots Z'_n)$  noch nicht konvergiert ist, zurück zu (2). Andernfalls ist  $\hat{s}_i = Z'_i$ .

Die Reihenfolge der  $(i, j)$  in Punkt (2) ist im Allgemeinen nicht ausschlaggebend. Sowohl eine systematische wie auch eine zufällige Wahl der Paare zeigten in numerischen Simulationen ähnliche Konvergenzverhalten. Um die Effizienz des Algorithmus zu steigern, wäre eine selektive Wahl der Paare nützlich, die z.B. durch die Änderungen der Winkel im letzten Iterationsschritt bestimmt werden kann. Weitere Möglichkeiten, den Algorithmus zu verbessern, wären, die Anzahl der Winkel schrittweise zu vergrößern, die Dichte

der Schätzungen im Minimum zu erhöhen oder mit einer Ausgabe eines schnelleren ICA-Algorithmus zu starten. Darüber hinaus könnte in der nächsten Nachbarsuche berücksichtigt werden, dass sich die Nachbarschaften nur minimal bei der Rotation ändern, anstatt wie bisher diese für jeden Winkel neu zu berechnen.

In der vorgestellten Variante des MILCA-Algorithmus wird davon ausgegangen, dass keine zeitliche Abhängigkeiten in den Signalen vorhanden sind. Im Folgenden wollen wir aber annehmen, dass uns ein Signal  $x$  vorliegt, welches eine endliche Autokorrelationszeit besitzt. Algorithmen wie JADE [10], FastICA [28] und Infomax [43] können weiter verwendet werden. Dies heißt aber, dass die zeitliche Information des Signals verschenkt wird und somit auch eine Möglichkeit, die Separation zu verbessern. Auf der anderen Seite minimieren Algorithmen wie AMUSE [46] und TDSEP [79] zeitliche Korrelationen, sie verwenden dabei aber ausschließlich Statistiken zweiter Ordnung. Somit ist die erste Gruppe von Algorithmen nicht in der Lage, zwei gaußverteilte Signale mit unterschiedlichen Frequenzspektren zu trennen, während die zweite Gruppe nicht in der Lage ist, zeitlich weiße Signale mit verschiedenen Amplitudenverteilungen zu trennen. Offensichtlich ist es erforderlich, sowohl Statistiken höherer Ordnung, als auch die zeitlichen Strukturen in den Signalen zu nützen, um optimale Ergebnisse zu erhalten [48]. Der MILCA-Algorithmus kann sehr einfach und unkompliziert für diesen Fall formuliert werden.

Der erste Schritt in der nichtlinearen Zeitreihenanalyse von eindimensionalen Daten ist eine Verzögerungseinbettung [32]. Dabei wird ein  $m$ -dimensionales Signal ( $m > 1$ ) dadurch erzeugt, dass  $m$  Messwerte mit passend gewählten Verzögerungen  $\tau$  zu sogenannten Verzögerungsvektoren zusammengefasst werden

$$\mathbf{x}(t) = [x(t - \tau), x(t - 2\tau), \dots, x(t - m\tau)]^T. \quad (3.35)$$

Demnach wird der "Zustand" eines Signals zur Zeit  $t$  durch  $m$  vorherige Zeitpunkte beschrieben. Dies macht natürlich nur Sinn, wenn eine Zeitstruktur in dem Signal enthalten ist. Gleichermäßen können wir auch multivariate Signale einbetten. Für  $n$  Messkanäle ergibt sich somit eine  $n \times m$  Verzögerungsmatrix

$$\mathbf{X}(t) = [\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)]. \quad (3.36)$$

Um jetzt eine instantane lineare Mischung aus  $n$  Signalen, die entweder nicht gaußverteilt sind oder keine nicht-triviale Zeitstruktur besitzen, zu entmischen, schlagen wir vor, die MI der Verzögerungsmatrix zu minimieren

$$\hat{I}(\mathbf{s}_1(t), \dots, \mathbf{s}_n(t)) \stackrel{!}{=} \min. \quad (3.37)$$

Es muss dabei beachtet werden, dass die Verzögerungsvektoren als eine Einheit betrachtet werden, d.h. in Gl.(3.37) werden nicht die MI zwischen den verschiedenen Verzögerungen des selben  $x_i$  berechnet. Ausführlicher heißt das

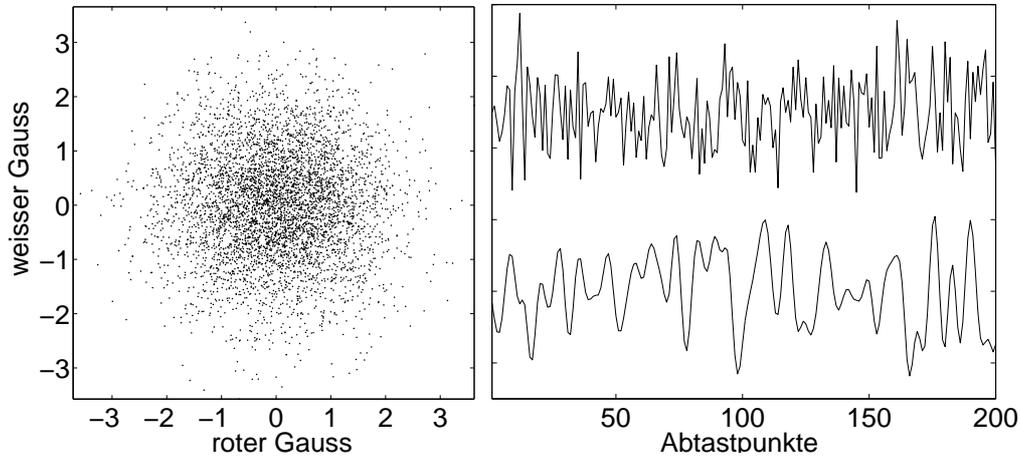
$$\begin{aligned}
 I(\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)) &= I(x_1(t - \tau), \dots, x_1(t - m\tau), \\
 &\quad x_2(t - \tau), \dots, x_2(t - m\tau), \dots \\
 &\quad x_n(t - \tau), \dots, x_n(t - m\tau)) \\
 &= \sum_{i=1}^n I(x_i(t - \tau), \dots, x_i(t - m\tau)) \\
 &= \sum_{i=1}^n H(\mathbf{x}_i(t)) - H(\mathbf{x}_1(t), \dots, \mathbf{x}_n(t))
 \end{aligned} \tag{3.38}$$

In der Minimierung kann wie in der ersten Variante des Algorithmus vorgegangen werden, d.h. wir zerlegen die Rotationen, die  $\hat{I}(\mathbf{x}_1(t), \dots, \mathbf{x}_n(t))$  minimieren, in  $n(n - 1)/2$  Rotationen. Jede dieser Rotationen besteht immer noch aus Rotationen von  $m$  Verzögerungs-koordinaten-Paaren, aber diese können wiederum in  $m$  Rotationen zerlegt werden, wobei jeweils nur ein Paar rotiert wird. Somit erhalten wir

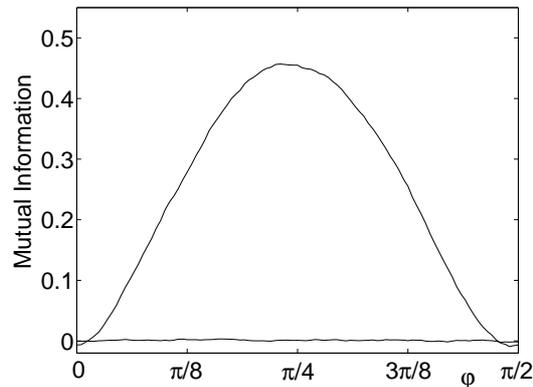
$$\begin{aligned}
 &I(\dots \mathbf{x}'_i(t), \dots \mathbf{x}'_j(t) \dots) - I(\dots \mathbf{x}_i(t), \dots \mathbf{x}_j(t) \dots) \\
 &= I(\mathbf{x}'_i(t), \mathbf{x}'_j(t)) - I(\mathbf{x}_i(t), \mathbf{x}_j(t)) \\
 &= I(x_i(t - \tau), \dots, x_i(t - m\tau)) \\
 &\quad + I(x_j(t - \tau), \dots, x_j(t - m\tau)) \\
 &\quad - I(x'_i(t - \tau), \dots, x'_i(t - m\tau)) \\
 &\quad - I(x'_j(t - \tau), \dots, x'_j(t - m\tau)) \\
 &\quad + m [I(x'_i(t), x'_j(t)) - I(x_i(t), x_j(t))] ,
 \end{aligned} \tag{3.39}$$

wobei wir für den letzten Term angenommen haben, dass die  $x'_i(t)$  stationär sind und somit  $I(x'_i(t), x'_j(t))$  unabhängig von  $t$  ist. Im Fall  $m = 2$  werden somit im Algorithmus nur die paarweise MI benötigt. Erst für  $m > 2$  ist es nötig, die  $m$ -dimensionale MI abzuschätzen.

Die Erweiterung des MILCA-Algorithmus kann an einem einfachen Beispiel veranschaulicht werden. Seien  $x_1(t)$  und  $x_2(t)$  zwei Kanäle mit einer gauß'schen Amplitudenverteilung, aber mit unterschiedlichen Spektren:  $x_1(t)$  ist weiß (iid), während  $x_2(t)$  rot ist. Der Einfachheit halber nehmen wir eine Rotationsmatrix als Mischung an. Die Punktwolke der beiden Signale in Abb.(3.5 links) bieten keinerlei Ansatzpunkt für eine Separation der Signale. Wie wir aber in Abb.(3.6) sehen können, ist es vollkommen ausreichend,  $m = 2$  zu



**Abbildung 3.5:** Linkes Bild: Scatterplot von zwei Gaußverteilungen mit unterschiedlichen Spektren. Rechtes Bild: Ausgabe des modifizierten MILCA-Algorithmus ( $\tau = 1$  und  $m = 2$ ), wobei die weiße gaußverteilte Zufallsvariable oben und die rote unten dargestellt ist (die Signale stimmen bis auf einen Skalierungsfaktor sehr genau mit den ursprünglichen Quellsignalen überein.)



**Abbildung 3.6:** Änderung von  $\hat{I}$  unter Rotation für das Signal aus Abb.(3.5). Die annähernd horizontale Kurve zeigt das Verhalten, wenn die MI nur zu instantanen Zeitpunkten berechnet wird, die sinusförmige Kurve zeigt die Werte der MI unter Berücksichtigung von Zeitverzögerungen (Der tatsächliche Winkel der Mischung ist 0).

wählen, um die originalen Quellen (Abb.(3.5 rechts)) wieder zurückzugewinnen.

Der MILCA-Algorithmus kann aber auch für andere Anwendungen erweitert werden. Der allgemeinste Ansatz für eine lineare Entmischung ist [27]

$$\begin{aligned}\hat{s}_i(t) &= \sum_{j=1}^N \sum_{k=1}^m w_{ij}^k x_j(t - k\tau) \\ &= \sum_{j=1}^N \mathbf{w}_{ij} \mathbf{x}_j(t),\end{aligned}\quad (3.40)$$

wobei die  $\mathbf{x}_j(t)$  wieder Verzögerungsvektoren wie in Gl.(3.35) sind, und

$$\mathbf{w}_{ij} = [w_{ij}^1 \dots w_{ij}^k]. \quad (3.41)$$

Somit erhöht sich gewissermaßen die Anzahl der Messkanäle  $x_j(t - k\tau)$  von  $n$  auf  $n \times m$  und auch die Zahl der Komponenten der  $\hat{s}_i(t)$  auf  $n \times m$ , da im Unterschied zur Gl.(3.37) jetzt jeder Kanal als eigenständig betrachtet wird. Somit können wir die instantane Variante des MILCA-Algorithmus unverändert verwenden, und es erhöht sich lediglich die Anzahl der Komponenten.

$$\hat{I}(s_1(t), \dots, s_{nm}(t)) \stackrel{!}{=} \min. \quad (3.42)$$

Dieser Ansatz ist offensichtlich passender für Signale  $x_i(t)$ , die selbst Mischungen zeitverzögerter Quellsignale sind. Für eine endliche Anzahl an Verzögerungszeiten

$$x_i(t) = \sum_j \sum_{k=1}^{m'} a_{ij}^k s_j(t - k\tau), \quad (3.43)$$

ergibt Gl.(3.40) mit endlichen  $m$  allerdings keine exakte Entmischung, da wir für die Invertierung von Gl.(3.43) eine unbegrenzte Zahl an Verzögerungstermen benötigen würden. Außerdem sind die Lösung von Gl.(3.40) im Allgemeinen keine Verzögerungsvektoren. Trotzdem sollten sie die Quellsignale in Gl.(3.43) besser als die eines instantanen Ansatzes abschätzen können (siehe Anwendungen in Kap.(4.1)).

Für die letzte Variante des MILCA-Algorithmus wird einerseits berücksichtigt, dass die Mischung nicht nur instantan erfolgen muss (Gl.(3.42)) und außerdem, dass es nützlich sein kann, auch die zeitlichen Abhängigkeiten (Gl.(3.37)) zu minimieren. Es werden somit die letzten beiden Varianten des Algorithmus kombiniert.

Wir verwenden  $m$  Verzögerungen in der Entmischung und  $m'$  für die Minimierung der MI. Mit dem selben Ansatz wie in Gl.(3.40) ist nun die Aufgabe

$$\hat{I}(\mathbf{s}_1(t), \dots, \mathbf{s}_{nm}(t)) \stackrel{!}{=} \min. \quad (3.44)$$

zu finden, wobei wir die Definition von  $I(\mathbf{s}_1(t), \dots)$  aus Gl.(3.39) mit  $\hat{\mathbf{s}}_i(t) = [\hat{s}_i(t - \tau), \hat{s}_i(t - 2\tau), \dots, \hat{s}_i(t - m'\tau)]^T$  verwendet haben.

In Kap.(4) werden dann verschiedene Anwendungen vorgestellt, in denen die unterschiedlichen Varianten des MILCA-Algorithmus Verwendung finden.

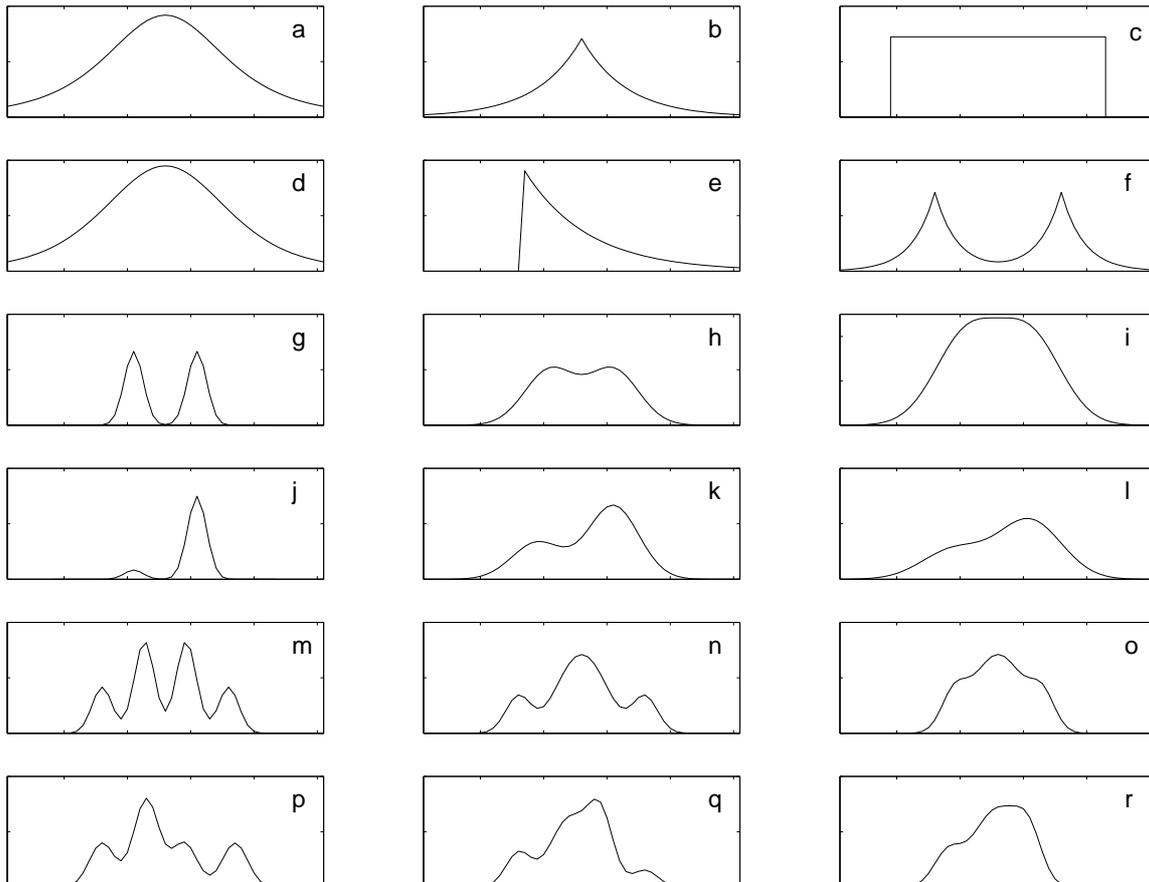
### 3.2.4 Leistungvergleich verschiedener ICA-Algorithmen

In den letzten zehn Jahren sind eine Vielzahl von ICA-Algorithmen entwickelt worden. Im Kap.(3.2.3) sind dabei nur die am häufigst verwendeten Algorithmen behandelt worden. Die Motivationen für neue Entwicklungen sind ebenso vielfältig. Sie sind z.B. darauf zurückzuführen, dass die Datensätze in den verschiedenen Anwendungen spezifische Eigenschaften besitzen, so dass ein Algorithmus auf diese optimiert werden kann. Aber da wir es meistens mit Situationen zu tun haben, in denen kein Wissen über die Quellsignale vorhanden ist (daher auch der Name "Blind Source Separation"), muss im allgemeinen von einem ICA-Algorithmus verlangt werden, dass er für beliebige Wahrscheinlichkeitsdichten der Signale gute Ergebnisse liefert. Als Test, wie gut die verschiedenen Algorithmen in der Lage sind, blind Mischungen zu trennen, messen wir die Separationsgenauigkeit an 18 unterschiedlich verteilten Quellsignalen. Die verschiedenen Verteilungen sind in Abb.(3.7) dargestellt. Um jetzt die Güte einer Separation zu bestimmen, wird der Amari-Leistungsindex [1]

$$P_{err} = \frac{1}{2N} \sum_{i,j=1}^N \left( \frac{|p_{ij}|}{\max_k |p_{ik}|} + \frac{|p_{ij}|}{\max_k |p_{kj}|} \right) - 1 \quad (3.45)$$

verwendet. Hierbei ist  $p_{ij} = (\hat{\mathbf{A}}^{-1}\mathbf{A})_{ij}$  das Matrixprodukt der invertierten geschätzten Mischmatrix mit der tatsächlichen Mischmatrix und sollte daher bei optimaler Rückgewinnung der Quellsignale die Einheitsmatrix (bis auf Zeilenpermutationen und -skalierung) darstellen.

Für den Test folgen wir exakt Ref. [6, 42] um eine Vergleichbarkeit zu gewährleisten. Jedes Problem besteht aus einer der Verteilungen  $p(x)$  aus Abb.(3.7). Es werden eintausend Paare von Zufallszahlen  $x$  und  $y$  unabhängig aus  $p(x)p(y)$  gezogen. Diese werden dann mit einem zufällig gewählten Winkel  $\phi$  durch  $x' = x \cos \phi + y \sin \phi, y' = -x \sin \phi + y \cos \phi$



**Abbildung 3.7:** Die 18 Wahrscheinlichkeitsdichtefunktionen der Quellsignale [6, 42]: (a) Studentverteilung mit drei Freiheitsgraden; (b) doppel-exponentiell; (c) gleichförmig; (d) Student mit fünf Freiheitsgraden; (e) exponentiell; (f) Mischung zweier Doppel-exponentieller; (g-h-i) symmetrische Mischung zweier Gauß: multimodal, transitional und unimodal; (j-k-l) nicht symmetrische Mischung zweier Gauß: multimodal, transitional und unimodal; (m-n-o) symmetrische Mischung von vier Gauß: multimodal, transitional und unimodal; (p-q-r) nicht symmetrische Mischung von vier Gauß: multimodal, transitional und unimodal.

| pdfs   | FastICA | Jade | Imax       | KCCA | KGV        | RADICAL    | MILCA      | MILCA<br>(verfeinert) |
|--------|---------|------|------------|------|------------|------------|------------|-----------------------|
| a      | 4.4     | 3.7  | <b>1.8</b> | 3.7  | 3.0        | 2.1        | 2.7        | 2.4                   |
| b      | 5.8     | 4.1  | 3.4        | 3.7  | 2.9        | 2.7        | 2.9        | <b>2.5</b>            |
| c      | 2.3     | 1.9  | 2.0        | 2.7  | 2.4        | 1.2        | 1.5        | <b>1.0</b>            |
| d      | 6.4     | 6.1  | 6.9        | 7.1  | 5.7        | 5.3        | 7.0        | <b>4.3</b>            |
| e      | 4.9     | 3.9  | 3.2        | 1.7  | 1.5        | <b>0.9</b> | <b>0.9</b> | 1.0                   |
| f      | 3.6     | 2.7  | 1.0        | 1.7  | 1.5        | 1.0        | <b>0.9</b> | <b>0.9</b>            |
| g      | 1.8     | 1.4  | <b>0.6</b> | 1.5  | 1.4        | <b>0.6</b> | <b>0.6</b> | <b>0.6</b>            |
| h      | 5.1     | 4.1  | <b>3.1</b> | 4.6  | 3.6        | 3.7        | 3.4        | 3.3                   |
| i      | 10.0    | 6.8  | 7.8        | 8.3  | <b>6.4</b> | 8.3        | 7.9        | 8.0                   |
| j      | 6.0     | 4.5  | 50.6       | 1.4  | 1.3        | 0.8        | <b>0.7</b> | 0.8                   |
| k      | 5.8     | 4.4  | 4.2        | 3.2  | 2.8        | 2.7        | 2.4        | <b>2.3</b>            |
| l      | 11.0    | 8.3  | 9.4        | 4.9  | 3.8        | 4.2        | 4.1        | <b>3.3</b>            |
| m      | 3.9     | 2.8  | 3.9        | 6.2  | 4.7        | 1.0        | 1.0        | <b>0.8</b>            |
| n      | 5.3     | 3.9  | 32.1       | 7.1  | 3.0        | 1.8        | 2.0        | <b>1.6</b>            |
| o      | 4.4     | 3.3  | 4.1        | 6.3  | 4.5        | 3.4        | 3.4        | <b>2.9</b>            |
| p      | 3.7     | 2.9  | 8.2        | 3.6  | 2.8        | <b>1.1</b> | 1.6        | 1.2                   |
| q      | 19.0    | 15.3 | 43.3       | 5.2  | 3.6        | 2.3        | 2.9        | <b>1.9</b>            |
| r      | 5.8     | 4.3  | 5.9        | 4.1  | 3.7        | 3.2        | 3.5        | <b>2.7</b>            |
| Mittel | 6.1     | 4.7  | 10.6       | 4.3  | 3.3        | 2.6        | 2.7        | <b>2.3</b>            |

**Tabelle 3.1:** Leistungsindex (multipliziert mit 100) für die BSS von 2-Komponenten-Mischungen. Die Ergebnisse der ersten 6 Spalten (FastICA [28], Jade [10], Imax [43], KCCA [6], KGV [6], and RADICAL [42]) sind von [42] kopiert. Die letzten beiden Spalten zeigen die Ergebnisse für MILCA, zuerst für die Standardversion (Spalte 7) und dann mit Datenvermehrung (Spalte 8). Der Leistungsindex ergibt sich aus einer Mittelung über 100 Datensätzen, jeder Satz besteht aus 1000 Wertepaaren, welche zufällig aus den pdfs gezogen werden. Für MILCA verwenden wir  $k = 10$  und fitten  $\hat{I}(\phi)$  mit drei (MILCA) bzw. fünf Fouriertermen (MILCA verfeinert).

gemischt (d.h.  $\mathbf{A}$  ist eine reine Rotation). Dies wird 100 mal mit unterschiedlichen Winkeln und Sätzen  $(x, y)$  wiederholt, über welche anschließend ein mittlerer Leistungsindex (Gl.(3.45)) gebildet wird. Die Ergebnisse sind zusammen mit Resultaten verschiedener anderer ICA-Algorithmen [42] in Tabelle (3.1) (Spalte MILCA) gezeigt. Sie sind im Mittel hervorragend und werden nur von denen des RADICAL-Algorithmus [42] übertroffen. Dieser ist MILCA ähnlich, da er ebenfalls die nächste Nachbar-Statistik benutzt, aber dabei nur die Schätzung der differentiellen Shannon-Entropien  $H(x')$  und  $H(y')$  verwendet. Eine weitere Eigenschaft des Algorithmus in Ref. [42] ist die Verwendung der in Kap.(3.2.3.4) besprochenen Datenvermehrung, die zu einer glatteren Abhängigkeit vom Winkel  $\phi$  führt. Der selbe Trick kann auch im MILCA-Algorithmus verwendet werden und verbessert aus ähnlichen Gründen auch dort das Ergebnis. Die Resultate werden in der

Spalte MILCA(verfeinert) gezeigt und sind nun im Mittel besser als die von RADICAL.

In Anwendungen mit einer kleinen Zahl von Komponenten ist somit MILCA zu bevorzugen, nur für höherdimensionale Probleme kann es aufgrund des numerischen Mehraufwandes effizienter sein, schnellere ICA-Algorithmen zu verwenden.

### 3.2.5 Von der ICA zur Blind Source Separation

Das Ziel der ICA ist es, Komponenten zu finden, die voneinander so unabhängig wie möglich sind. Das Ziel der BSS ist es, die originalen Quellsignale  $s$  aus einer Mischung  $x$  wieder zurückzugewinnen. Die lineare instantane ICA liefert genau dann eine Lösung für die BSS, wenn das Eingangssignal eine lineare instantane Mischung aus unabhängigen Quellsignalen<sup>2</sup> darstellt. Diese Voraussetzungen müssen immer für jeden Datensatz überprüft werden. Ob die Unabhängigkeit der Quellsignale gegeben ist, kann wiederum mit Hilfe der Mutual Information bestimmt werden, da sie ein Maß für die absolute Abhängigkeit ist. Gefundene Abhängigkeiten zwischen den vermeintlichen Quellsignalen können darauf hinweisen, dass die ursprünglichen Quellsignale tatsächlich abhängig sind, aber auch, dass die Mischung nicht linearer instantaner Natur ist.

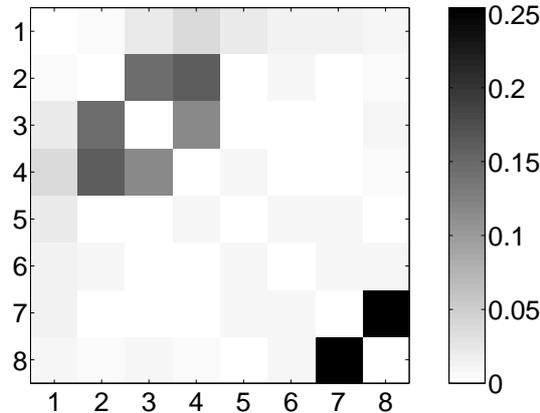
Auf der anderen Seite ist die Struktur der Quellsignale entscheidend für die Eindeutigkeit der Lösung. So ist es möglich, dass andere Mischungen der Daten ähnliche Abhängigkeiten zeigen und somit, die gefundene Entmischung nicht eindeutig ist. Dies kann mit Hilfe der Variabilität der MI unter kontrollierter Mischung verifiziert werden. Eine rasche Änderung der MI durch eine erneute Mischung identifiziert stabile Komponenten. Andererseits zeichnen sich unzuverlässige Komponenten durch eine geringere Änderung der MI aus. Diese Tests werden jetzt ausführlich erklärt und dann an einem künstlich erzeugten Datensatz illustriert. In Kap.(4.1) wird dann die Anwendung auf Messdaten gezeigt. Außerdem sei noch erwähnt, dass in Ref. [45, 23, 26] Zuverlässigkeitstests vorgestellt werden, die auf der Bootstrapmethode bzw. der Rauschinjektion basieren.

#### **Abhängigkeitsmatrix / Modelltest:**

Für die meisten Anwendungen kann angenommen werden, dass die Quellsignale unabhängig sind. Somit weisen Abhängigkeiten in dem ICA-Ausgabevektoren meistens auf Fehlannahmen im gewählten Modell für die Mischung hin. Es muss sich immer im Voraus für ein Modell entschieden werden, ob es linear oder nichtlinear sei, ist es instantan oder enthält es zeitliche Mischungen. Für die Bewertung der Stimmigkeit des gewählten Modells mit der

---

<sup>2</sup>Die Quellsignale müssen noch zusätzlich die Bedingung erfüllen, dass maximal ein Kanal eine Gaußverteilung ohne zeitliche Korrelationen aufweist.



**Abbildung 3.8:** Exemplarische Abhängigkeitsmatrix  $\hat{I}(X_i, X_j)$  ( $k=6$ ) zwischen allen Kanalpaaren (die Diagonale ist auf Null gesetzt). Die Kanäle 1, 5 und 6 sind eindimensionale Komponenten, die Kanäle 2, 3 und 4 bilden eine dreidimensionale Komponente und die Kanäle 7 und 8 eine zweidimensionale Komponente.

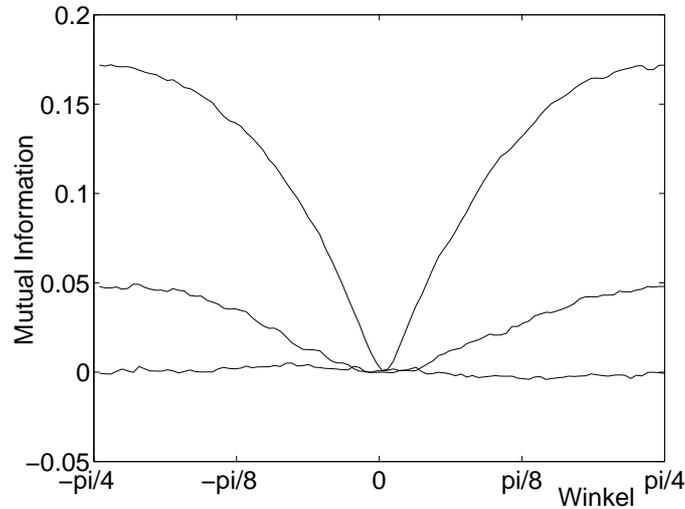
Realität berechnen wir die paarweise MI zwischen allen ICA-Komponenten,  $\hat{I}(y_i, y_j)$  mit  $i, j = 1, \dots, N$  und  $i \neq j$ . Eine exemplarische Abhängigkeitsmatrix ist in Abb.(3.8) gezeigt. Ein Paar von Komponenten, deren MI unter einem definierten Grenzwert  $\hat{I}(y_i, y_j) < D_{MAX}$  fällt, wird als wirklich unabhängig angesehen.

Eine Komponente, die wirklich unabhängig von allen anderen Komponenten ist, wird als eindimensionale Komponente bezeichnet<sup>3</sup>. Eine Gruppe von Komponenten, die untereinander abhängig sind<sup>4</sup> ( $\hat{I}(y_{i_1}, \dots, y_{i_n}) > D_{MAX}$ ), aber die wirklich unabhängig von den restlichen Komponenten sind, werden als multidimensionale Komponenten bezeichnet. Auf diese Weise erlaubt die Abhängigkeitsmatrix, unabhängig vom verwendeten ICA-Algorithmus, ein- und multidimensionale Komponenten zu identifizieren.

Eine eindimensionale Komponente, deren Eindeutigkeit noch mit Hilfe eines zweiten Tests überprüft werden muss, entspricht dann, bis auf einen Skalierungsfaktor, einem ursprünglichen Quellsignal. Eine multidimensionale Komponente ist schwieriger zu interpretieren, da es mehrere mögliche Gründe für die Abhängigkeiten zwischen ihren Kanälen gibt. So könnten diese Kanäle noch zusätzlich durch eine weitere Überlagerung (unterschiedlicher Natur) untereinander gemischt worden sein, wobei auch die einzelnen Quellsignale durch Filterung unterschiedliches Aussehen an den Sensoren zeigen können und somit unterschiedlich zu den Messsignalen  $x_i$  beitragen (Kap.(4.1)). Eine andere Möglichkeit für das Auffinden von multidimensionalen Komponenten können Abhängigkeiten zwischen einzelnen Quellsignalen sein (Kap.(4.2)).

<sup>3</sup>Eine theoretisch saubere Definition wäre  $\hat{I}(y_k, (y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_N)) < D$  zu fordern, aber das würde zu einem numerischen Mehraufwand führen. Diese strengere Definition wird jedoch für die meisten Anwendungen nicht benötigt.

<sup>4</sup>Es sei noch angemerkt, dass nicht alle möglichen Paare von Komponenten innerhalb dieser Gruppe paarweise abhängig sein müssen.



**Abbildung 3.9:**  $\hat{I}(\mathbf{R}(\phi)(X, Y))$  ( $k=6$ ) gegen den Rotationswinkel. Von oben nach unten:  $X$ = gleichförmig, gleichförmig+Gauß, Gauß;  $Y$ = immer Gauß.

Für den Fall, dass nur eine einzige multidimensionale Komponente gefunden wird, d.h. alle Komponenten sind voneinander abhängig, konnte kein einziges Quellsignal zurückgewonnen werden und es wurde somit eine BSS erzielt.

**Eindeutigkeitstest:**

Nehmen wir an, unter der Annahme einer linearen instantanen Mischung habe sich im Modelltest gezeigt, dass  $X$  und  $Y$  zwei eindimensionale Komponenten sind. Aufgrund der Unabhängigkeit von allen anderen Komponenten ist somit für diese Komponenten das gewählte Modell passend. Um jetzt noch die Eindeutigkeit von  $X$  und  $Y$  zu prüfen, wird eine Rotation im zweidimensionalen Unterraum durchgeführt, wobei  $\hat{I}(\mathbf{R}(\phi)(X, Y))$  für unterschiedliche Rotationswinkel  $\phi$  berechnet wird. In Abb.(3.9) zeigen wir dafür drei Fälle.

In diesen ist  $X$  gleichförmig verteiltes Rauschen, gleichförmig verteiltes Rauschen überlagert mit gauß'schem Rauschen, bzw. reines gauß'sches Rauschen, wobei  $y$  für alle Fälle gauß'sches Rauschen darstellt. Die Tiefe des Minimums von  $\hat{I}(\mathbf{R}(\phi))$  spiegelt die Eindeutigkeit von  $X$  und  $Y$  wieder. Eine flache Kurve deckt eine Mehrdeutigkeit der Lösung für  $X$  und  $Y$  auf, während ein ausgeprägtes Minimum ihre Eindeutigkeit zeigt. Es wird nun die Variabilität von  $\hat{I}(\mathbf{R}(\phi))$  zwischen allen Kanalpaaren der ICA-Ausgabe als ein Maß für die Eindeutigkeit verwendet:

$$\sigma_{ij} = \overline{I(X_i, X_j)} - I(\mathbf{R}(0)(X_i, X_j)) \tag{3.46}$$

wobei das globale Minimum von  $\hat{I}$  bei  $\phi = 0$  ist und

$$\overline{I(X_i, X_j)} = \frac{1}{2\pi} \int_0^{2\pi} d\phi I(\mathbf{R}(\phi)(X_i, X_j)) . \quad (3.47)$$

Aus Abb.(3.9) könnte man schließen, dass es ausreichend ist,  $\overline{I(X_i, X_j)}$  aus MI-Werten an nur zwei Winkeln  $\phi = (0, \pi/4)$  abzuschätzen, aber dies könnte aufgrund numerischer Fluktuationen gefährlich sein. Außerdem kann die Form der  $\hat{I}(\mathbf{R}(\phi))$ -Kurve auch komplexere Erscheinungen als in Abb.(3.9) annehmen, z.B. für Signale mit multimodaler Verteilung.

Der Modelltest ist für verschiedene ICA-Modelle direkt verwendbar, und im Eindeutigkeitsstest muss nur noch die Mischoperation entsprechend angepasst werden.

Um zu überprüfen, ob die gefundene ICA-Komponenten eine physikalische Bedeutung tragen, d.h. ob die BSS erfolgreich war, ist es unerlässlich, die Komponenten auf ihre Zuverlässigkeit (Modell- und Eindeutigkeits-Test) zu testen. Entsprechende Anwendungen werden in Kap.(4) für das lineare instantane Modell bzw. lineare Modell mit zeitversetzter Mischung gezeigt.

## 3.2.6 Von der Theorie zur Praxis

### 3.2.6.1 Rauschen am Sensor und der Quelle

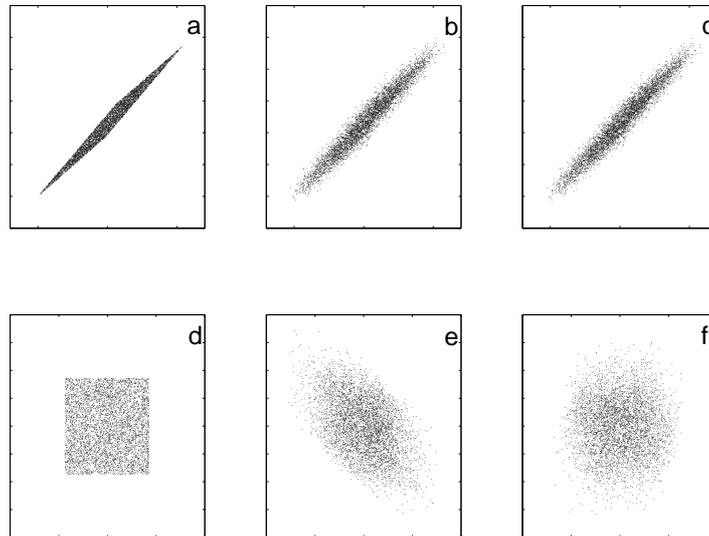
Da es unser Ziel ist, eine BSS an realen Messdaten durchzuführen, müssen wir den Einfluss von Rauschen in der ICA diskutieren. Hierbei können prinzipiell zwei Fälle unterschieden werden<sup>5</sup>. In ersten Fall sind die Quellsignale selbst mit Rauschen überlagert, im letzteren wird das Rauschen durch die Messung selbst induziert.

Für den ersten der beiden Fälle kann im Allgemeinen weiterhin ein Standard-ICA-Algorithmus verwendet werden, weil die Mischung der Signale erst nach der Überlagerung mit Rauschen erfolgt. Somit kann die Entmischung wie bisher durchgeführt werden, wobei nur mit geringen Leistungseinbußen zu rechnen ist, die dadurch entstehen, dass die Quellsignale jetzt näher einer Gaußverteilung sind. Dies heißt, dass im Idealfall die Mischmatrix exakt bestimmt werden kann und die primäre Zusatzaufgabe darin besteht, aus den zurückgewonnenen verrauschten Quellsignalen die tatsächlichen Quellsignale abzuschätzen. Hierfür können verschiedene Rauschunterdrückungsmethoden [32] verwendet werden, wobei die einfachste wahrscheinlich ein Tiefpassfilter wäre.

Im Fall von Messrauschen gestaltet sich die BSS als viel schwieriger, da hier das Signal noch nach der eigentlichen Mischung verändert wird. Infolgedessen steht man vor

---

<sup>5</sup>Tatsächlich gibt es noch den weiteren Fall "dynamisches Rauschen".



**Abbildung 3.10:** Erste Zeile: gemischte gleichförmig verteilte Zufallsvariablen: (a) ohne gauß'sches Messrauschen; (b,c) mit Messrauschen; Zweite Zeile: (d) MILCA-Ausgabe der Daten von (a); (e) Anwendung von  $A^{-1}$  auf die Daten aus (b); (f) MILCA-Ausgabe der Daten von (c).

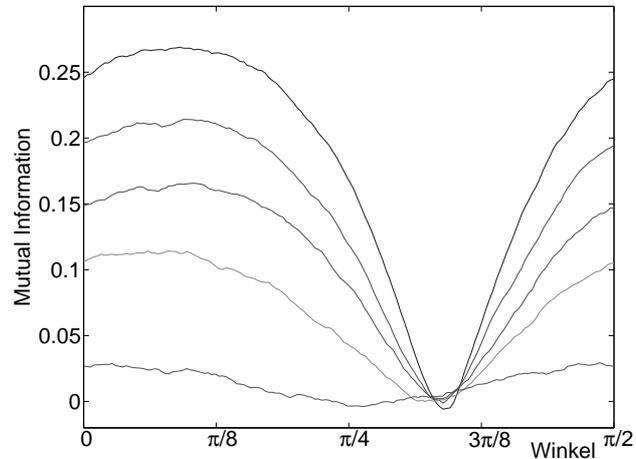
dem Problem, dass bei einer Invertierung der Mischung das Rauschen plötzlich abhängig gemacht wird. Veranschaulicht wird dies in Abb.(3.10), wo zwei gleichförmig verteilte Quellsignale  $s_i$  nach der Mischung mit unabhängig verteiltem gauß'schem Rauschen  $\eta$  überlagert werden

$$x_i = \sum_{j=1}^2 A_{ij}s_j + \eta_i , \quad (3.48)$$

wobei

$$\langle \eta_i(t)\eta_j(t') \rangle = r\delta_{ij}\delta_{tt'} \quad (3.49)$$

gilt. In der mittleren Spalte der Abbildung wird die Inverse der tatsächlichen Mischmatrix angewandt und man sieht, dass damit korreliertes Rauschen erzeugt wird. In der rechten Spalte wird MILCA direkt verwendet und in diesem Fall ist zu erkennen, dass hier ein Kompromiss zwischen der Unabhängigkeit der Quellsignale und der Unabhängigkeit des Rauschens gemacht werden muss. Um einen tieferen Einblick zu gewinnen, wird diese Mischung für ein fixes  $A$  und verschiedenen Signal-Rausch-Verhältnissen ( $SNR = \text{var}(s_i(t))/r$ ) wiederholt durchgeführt. Das Verhalten von  $\hat{I}(\phi)$  wird dann in Abb.(3.11)

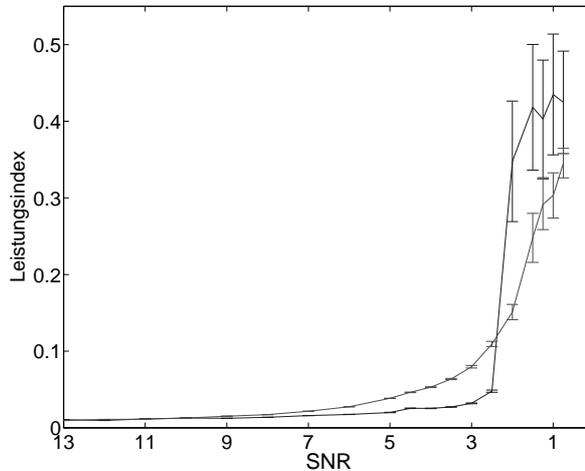


**Abbildung 3.11:** Ungeglättete Schätzungen von  $\hat{I}(\phi)$  für zwei rotierte gleichförmig verteilte Zufallsvariablen, die mit isotropen gauß'schem Messrauschen überlagert sind, abgetragen gegen  $\phi$ . Das Signal-Rausch-Verhältnis beträgt von oben nach unten  $SNR = \infty, 13, 7, 4, 1$ .

gezeigt. Wir können sehen, dass  $\hat{I}$  durch die Annäherung des Signals an die Gaußverteilung, mit sinkendem SNR flacher wird, d.h. die Variabilität bezüglich des Mischwinkels verringert sich. Der Einfluss des Rauschens führt durch die vorher besprochene Problematik darüberhinaus zu einer Verschiebung des Minimum von  $\hat{I}$ . Diese beiden Effekte führen daher bei einer direkten Verwendung eines Standard-ICA-Algorithmus zu Fehlern in der Abschätzung der originalen Mischmatrix.

Es kann versucht werden, spezielle Algorithmen zu entwickeln, die weniger vom Messrauschen beeinflusst werden. Nahezu alle in der Literatur vorgestellten Algorithmen (für einen Überblick, siehe [27, 12]) erzielen aber nur gute Ergebnisse, wenn das Rauschen sehr spezielle Voraussetzungen erfüllt. So wird oft verlangt, dass es gaußverteilt, additiv, unkorreliert und isotrop ist. Für diesen einfachen Fall kann auch der MILCA-Algorithmus zu n-MILCA (noisy-MILCA) modifiziert werden. Im ersten Schritt wird ein "Quasiwhitening" mit der geschätzten Kovarianzmatrix  $\mathbf{V} = (\mathbf{C}_x - r\mathbf{1})^{-1/2}$  der reinen Signale<sup>6</sup> durchgeführt, um die Quellsignale unkorreliert zu machen [28]. Im Vergleich zum herkömmlichen Whitening heißt das aber, dass unsere rekonstruierten Signale  $\hat{\mathbf{z}}_j(t)$  (siehe Gl.(3.11)) wegen des Messrauschens korreliert sind. Aus diesem Grund darf jetzt aber nicht mehr  $\hat{I}(\phi)$  minimiert werden, da es durch diese Abhängigkeiten irreführend werden würde. Da wir aber davon ausgehen, dass das Messrauschen die oben erwähnten Bedingungen erfüllt, d.h. die Abhängigkeiten vollkommen durch die lineare Korrelation bestimmt sind, können wir den linearen Beitrag zur MI subtrahieren. Für den neuen MILCA-Algorithmus wird daher  $\hat{I}(\phi) + \frac{1}{2} \ln(1 - C_{ij}(\phi)^2)$  minimiert.

<sup>6</sup>Unter reinen Signalen ist hier das Messsignal ohne Messrauschen gemeint.



**Abbildung 3.12:** Gemittelter Amari-Leistungindex gegen das Signal-Rausch-Verhältnis. Die Konditionszahl der Mischmatrix ist 6. Die obere Kurve (im SNR-Bereich von 7 bis 3) ist für den Standard-MILCA-Algorithmus und die untere für n-MILCA

Um jetzt MILCA mit n-MILCA zu vergleichen, mischen wir wiederum zwei gleichförmig verteilte Zufallsvariable mit Matrizen mit fixierter Konditionszahl<sup>7</sup>. In Abb.(3.12) zeigen wir den durchschnittlichen Amari-Leistungindex über 100 verschiedene Mischungen mit verschiedenem Rauschen. Es ist ersichtlich, dass die Berücksichtigung des Rauschens im Algorithmus das Ergebnis deutlich verbessert (abgesehen für sehr kleine SNR). Das Ausmaß der Verbesserung hängt aber von der Konditionszahl der Mischmatrix ab. Für Matrizen, die weit von einer Singularität entfernt sind (kleine Konditionszahl) hat das Quasiwhitening nur einen sehr geringen Einfluss und der Unterschied zwischen MILCA und n-MILCA ist minimal. Im anderen Extremfall einer sehr großen Konditionszahl sind die Mischungen annähernd gleich,  $x_i(t) \approx x_j(t)$ , und keiner der Algorithmen ist in der Lage, ein gutes Ergebnis zu liefern.

Bevor wir diesen Abschnitt verlassen, sollen noch ein paar Worte über Ausreißer in Messdaten gesagt werden. Im Prinzip sind Ausreißer nur eine spezielle Form von Rauschen. Da aber unser MI-Schätzer auf der  $k$  nächsten Nachbar-Statistik beruht, stellen Ausreißer ein viel geringeres Problem dar als für Algorithmen [6, 42], die auf der Kurtosis basieren.

#### 3.2.6.2 Abhängige Quellen

Eine der Voraussetzungen an die Quellsignale ist ihre gegenseitige Unabhängigkeit. Dies ist in der Tat die Grundidee der ICA und ist auch intuitiv verständlich. Wenn sich zum

<sup>7</sup>Die Konditionszahl ist durch das Verhältnis des größten Eigenwerts zum kleinsten Eigenwert gegeben, d.h. annähernd singuläre Matrizen führen zu großen Werten.

Beispiel mehrere Personen in einem Raum unterhalten, so ist es offensichtlich, dass die einzelnen Personen voneinander unabhängig unterschiedliche akustische Signale ausstoßen. Es ist somit die Voraussetzung erfüllt, dass die an mehreren Orten im Raum verteilten Mikrophone Mischungen von unabhängigen Quellsignalen aufzeichnen. Die Situation sieht aber vollkommen anders aus, wenn z.B. die Quellsignale durch einen äußeren Taktgeber synchronisiert werden. Dies ist in einem Musikkonzert der Fall, wo nicht mehr die Voraussetzung erfüllt ist, dass die Quellsignale unabhängig sind. Für diesen Fall würde die direkte Verwendung einer ICA auf die Mikrophonsignale nicht die akustischen Signale der einzelnen Musiker liefern, auch nicht wenn die Art der Mischung bekannt wäre. Der ICA-Algorithmus würde nämlich Mischungen der Messsignale finden, die unabhängiger voneinander sind als die tatsächlichen Quellsignale.

Eine mögliche Lösung dieses Problem stellt die Filterung der Messdaten dar. Dies kann damit erklärt werden, dass die Quellsignale nur in einem bestimmten Frequenzbereich abhängig sind. Am Beispiel des Musikkonzerts kann davon ausgegangen werden, dass vor allem die niedrigen Frequenzen Abhängigkeiten zeigen und die höheren Frequenzen durch die Charakteristik der einzelnen Musikinstrumente bestimmt werden und somit unabhängig für verschiedene Instrumente sind. Die Anwendung eines ICA-Algorithmus auf hochpassgefilterte Messsignale erhöht somit die Chance, eine gute Schätzung für die Mischmatrix zu erhalten und damit doch noch eine BSS zu erzielen. Eine Anwendung für einen ähnlichen Fall wird in Kap.(4.2) gezeigt, wo Infrarotspektren organischer Mischungen getrennt werden.

### 3.2.6.3 Anzahl der Quellen und Anzahl der Komponenten

In den meisten Veröffentlichungen findet man Sätze wie "Der Einfachheit halber nehmen wir an, dass die Anzahl der Quellsignale gleich der Anzahl der Messkanäle ist". Natürlich macht dies vieles einfacher, aber diese Annahme ist nur für die wenigsten Anwendungen erfüllt. Im Allgemeinen ist es immer von Vorteil, wenn die Anzahl der Messkanäle grösser als die Anzahl der Quellen ist. Im Fall einer linearen Mischung können dann einfach die Komponenten gewählt werden, die nach der Hauptachsentransformation eine Varianz über einem bestimmten Grenzwert aufweisen, d.h. die Anzahl der von Null verschiedenen Eigenwerte entspricht im Idealfall der Anzahl der Quellen. Diese Reduzierung der Anzahl der Messkanäle ist nicht in allen Fällen einfach und gestaltet sich speziell bei verrauschten Messdaten als schwierig.

Ein anderes Problem kann auftreten, wenn die Anzahl der Messkanäle etwa von der gleichen Größenordnung wie die Anzahl der Messpunkte ist. Dann kann es zu einem "Überlernen" (overlearning) des Algorithmus kommen, das sich dadurch auszeichnet, dass die geschätzten ICA-Komponenten hohe Spitzen aufweisen und an den restlichen Stellen annähernd Null sind. Im Extremfall, falls die Anzahl der Messkanäle identisch mit der der

Messpunkte ist, kann jedes Messsignal in Komponenten zerlegt werden, die nur einen einzigen Punkt ungleich Null enthalten, d.h. jeder einzelne Messpunkt wird durch eine einzelne Komponente beschrieben. Da aber in den meisten Fällen die Anzahl der tatsächlichen Quellen viel kleiner ist als die der Messkanäle, kann dieses Problem meist mit Hilfe einer Dimensionsreduktion gelöst werden [27].

Noch problematischer ist es, wenn die Anzahl der Quellen größer ist als die Anzahl der Messkanäle. Das bedeutet nämlich, dass die Mischung nicht invertierbar ist. Selbst wenn die Mischmatrix exakt bekannt wäre, könnten wir nicht die exakten Werte der unabhängigen Signale zurückgewinnen. Der Grund dafür ist die durch die Mischung verloren gegangene Information. Diese Situation wird oft als ICA mit überkompletter Basis bezeichnet, da wir es mit einem ICA-Modell

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \sum_i \mathbf{a}_i s_i \quad (3.50)$$

zu tun haben, bei dem die Anzahl der Basisvektoren  $\mathbf{a}_i$  größer ist als die Raumdimension von  $\mathbf{x}$ , folglich die Basis zu groß ist. Solche eine Situation ist z.B. manchmal bei der Extraktion von Merkmalen aus Bildern zu finden, wo die Anzahl der Quellen gegen unendlich tendiert.

Um dieses Problem zu lösen, muss versucht werden, zunächst die Mischmatrix abzuschätzen und dann im nächsten Schritt die unabhängigen Quellsignale. In der Praxis zeigt sich aber, dass es oft sehr schwierig ist, eine effiziente BSS zu erzielen [27]. Der Erfolg ist sehr stark von den jeweiligen Messdaten abhängig, und es empfiehlt sich oft, einen einfachen linearen ICA-Algorithmus auf die Daten anzuwenden, um etwa einen Teil der Quellsignale zu gewinnen. Diese können dann vom originalen Messsignal entfernt werden, um so eventuell weitere Quellen in einer zweiten ICA zu identifizieren. Dieser Ansatz liefert aber im Allgemeinen nur grobe Abschätzungen der ursprünglichen Quellsignale.

### 3.2.6.4 Welcher ICA-Algorithmus ist der Beste?

Die Wahl des ICA-Algorithmus wird entscheidend von der Fragestellung der Anwendungen und vom vorhandenen Wissen über die Messdaten bestimmt. Wenn die Anwendung sehr zeitkritisch ist, müssen Abstriche bei der Genauigkeit akzeptiert werden und FastICA wäre meist eine gute Wahl. Bei bekannter Verteilung der Quellsignale kann z.B. aus Tab. (3.1) ein ICA-Algorithmus gewählt werden. In den meisten Fällen ist aber die Verteilung der Quellsignale unbekannt und so ist es ratsam, einen Algorithmus zu wählen, der im Durchschnitt über viele verschiedene Verteilungen die beste Leistung zeigt, wie MILCA oder RADICAL.

Für jede Anwendung muss entschieden werden, wie die Messdaten gemischt worden sind. Wenn kein Wissen darüber vorhanden ist, ist es zweckmäßig im ersten Entmischungsversuch ein lineares instantanes Modell zu wählen. Mit Hilfe eines Zuverlässigkeitstest (siehe Kap.(3.2.5)) muss dann entschieden werden, ob das gewählte Modell für die untersuchten Messdaten passend ist. Wenn dies nicht der Fall ist, müssen Änderungen des Modells vorgenommen werden. Aufgrund der Vielzahl von unterschiedlichen Arten einer Mischung ist es meist unerlässlich, Vermutungen über ihre Natur aufzustellen. Diese können z.B. darin bestehen, dass Punktwolken wie in Abb.(3.10) geplottet werden und nach Strukturen darin gesucht wird [27]. Es ist mit Sicherheit keine leichte Aufgabe, die Art der Mischung zu bestimmen, nicht nur aufgrund der oft sehr geringen Anzahl von vorhandenen Messpunkten. Falls es aber gelingt, eine vernünftige Schätzung des Modells der Mischung zu erhalten, können die Messdaten mit seiner Hilfe so transformiert werden [12, 27], dass wieder die linearen instantanen ICA-Algorithmen aus Kap.(3.2.3) verwendet werden können. Somit muss nur noch, wie am Anfang des Kapitels besprochen, ein passender Algorithmus, abhängig von der Fragestellung der Anwendungen und dem vorhandenen Wissen über die Messdaten, gewählt werden.

# Kapitel 4

## Anwendungen der ICA

In den seltensten Fällen erfüllen Messdaten die Voraussetzungen, dass sie lineare instantane Mischungen von unabhängigen Quellen darstellen. Erschwert wird die Rückgewinnung der Quellsignale außerdem dadurch, dass die Anzahl der Messkanäle manchmal kleiner ist als die Anzahl der Quellen. Rauschen sowohl an der Quelle als auch am Sensor komplizieren das Problem noch zusätzlich. Trotz alledem ist ein effizienter Algorithmus zur Dekomposition linearer instantaner Mischungen (siehe Kap.(3.2.3)) ein hilfreiches Werkzeug, welches für viele Anwendungen adaptiert werden kann. Im Folgenden werden verschiedene Anwendungen dargestellt, in denen jeweils eine unterschiedliche Erweiterung des linearen instantanen ICA-Modells verwendet wird. Es werden Beispiele gezeigt, bei denen entweder die Mischung zeitversetzt erfolgt, die Quellen nicht unabhängig sind oder starkes Rauschen die Signale überlagert.

### 4.1 Elektrokardiogramm (EKG) einer schwangeren Frau

Die Trennung eines Elektrokardiogramms einer schwangeren Frau in die Herzschlagkomponente der Mutter und des Fötus stellt eine sehr interessante Anwendung für die ICA in der Medizin dar. Dafür wurde zuerst der lineare instantane MILCA-Algorithmus auf einen in der Literatur viel diskutierten EKG-Datensatz [9, 45, 37, 40] angewendet, der im Netz frei verfügbar ist [17]. Dieser besteht aus der Ableitung von 8 Elektroden, die über den Thorax und das Abdomen der Frau verteilt waren. Der in Abb.(4.1) dargestellte Datensatz enthält insgesamt 8 mal 2500 Messpunkte, die mit einer Abtastrate von 500 Hz aufgezeichnet wurden<sup>1</sup>.

---

<sup>1</sup>Wir sind uns der Tatsache bewusst, dass der Herzschlag sehr hoch ist und die tatsächliche Abtastrate womöglich geringer ist als die in Ref. [17] angegebene. Aber die Richtigkeit der Daten wurde von den Autoren bestätigt. Darüber hinaus sind unsere Schlussfolgerungen unabhängig von der tatsächlichen Abtastrate.



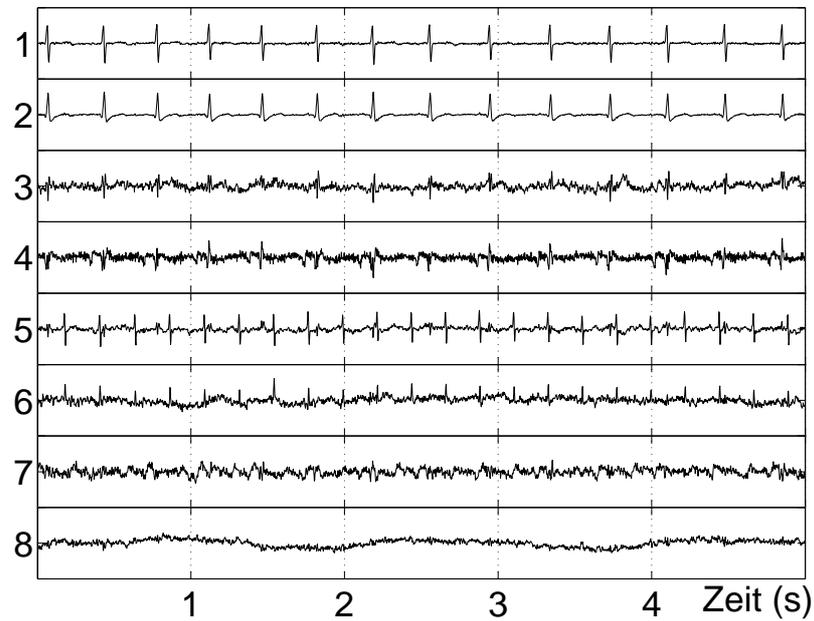
**Abbildung 4.1:** 8-Kanal EKG einer schwangeren Frau.

Die Ausgabe des MILCA-Algorithmus ist in Abb.(4.2) gezeigt, wobei  $k = 30$  nächste Nachbarn für die Abschätzung der Mutual Information verwendet wurden. Für die Minimumsuche im Raum der Drehungen wurde  $\hat{I}_{ij}(\phi)$  mit 3 Fourierkomponenten angepasst. Der Erfolg der Separation kann für diese Anwendung bereits visuell erkannt werden. Ganz offensichtlich sind die ersten beiden Kanäle durch den Herzschlag der Mutter bestimmt und der Kanal 5 durch die des Fötus. Die Kanäle 3, 4 und 6 enthalten aber noch immer Anteile von beiden und sind stärker verrauscht. Die Kanäle 7 und 8 sind demgegenüber überwiegend von Rauschen mit unterschiedlichen Frequenzanteilen dominiert.

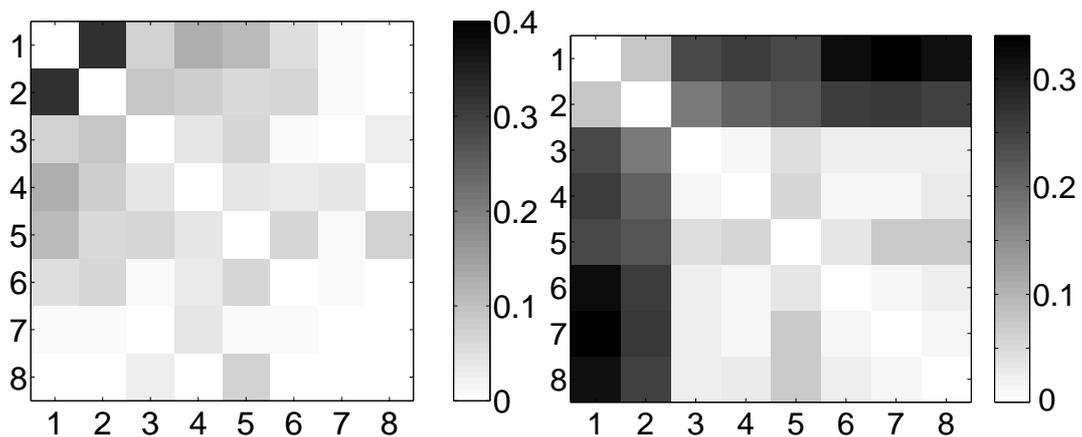
Bei einer genaueren Untersuchung der ICA-Komponenten erkennt man aber, dass die Komponente des Fötus noch einen Anteil vom EKG der Mutter enthält. Dieser Verdacht wird auch von der Zuverlässigkeitsüberprüfung der Komponenten bestätigt. Die Ergebnisse des Modelltests, die paarweisen Abhängigkeiten, sind in Abb.(4.3 links) gezeigt. Wir sehen, dass die meisten MI in der Tat kleine Werte annehmen, ausgenommen der ersten beiden Komponenten. Dies ist ein weiterer Hinweis, dass die Kanäle 1 und 2 zu der selben Quelle gehören, nämlich dem Herzschlag der Mutter. Aber es ist auch zu erkennen, dass die übrigen Komponenten nicht wirklich unabhängig voneinander sind.

Um in einer automatisierten Weise zu entscheiden, welche Komponenten abhängig sind und welche nicht, d.h. ein- bzw. multidimensionale Komponenten in der ICA-Ausgabe zu identifizieren, wenden wir den Clusteralgorithmus aus Kap.(2.7.2) auf die ICA-Komponen-

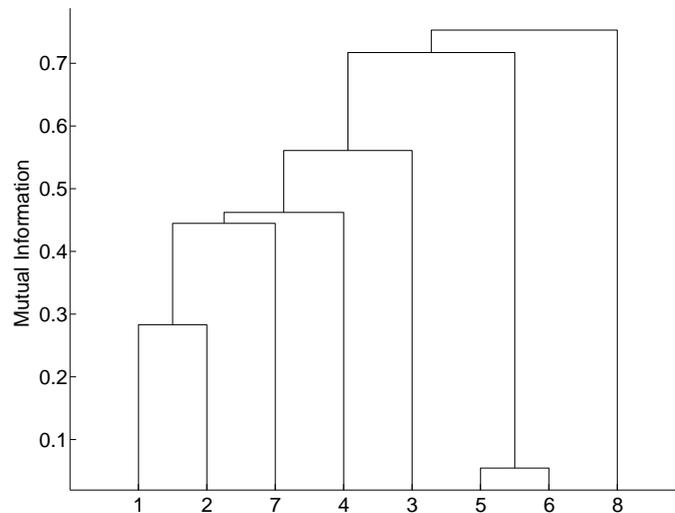
#### 4.1. ELEKTROKARDIOGRAMM (EKG) EINER SCHWANGEREN FRAU



**Abbildung 4.2:** MILCA-Ausgabe: Komponenten nach der Minimierung von  $I(X_1 \dots X_8)$  für die EKG-Signale aus Abb.(4.1).



**Abbildung 4.3:** Linkes Bild:  $\hat{I}$  zwischen allen paarweisen Kombinationen des Signals aus Abb.(4.2). Rechtes Bild: Quadratwurzel der Variabilität  $\sigma_{ij}$  von  $\hat{I}_{ij}(\phi)$ . In beiden Bildern ist die Diagonale auf Null gesetzt.



**Abbildung 4.4:** Dendrogramm für Abb.(4.2). Die Höhe der Cluster  $(ij)$  entspricht  $\hat{I}(X_i, X_j)$  ( $k=6$ ).

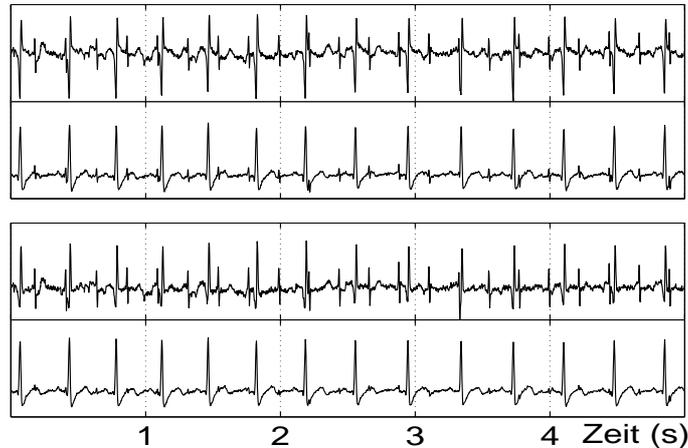
ten an. Das so erhaltene Dendrogramm wird in Abb.(4.4) gezeigt. In der Abbildung sind zwei Cluster zu erkennen: der Muttercluster, der die Kanäle 1, 2, 3, 4 und 7 enthält, und der Fötuscluster mit den Kanälen 5 und 6. Dies entspricht dem Ergebnis aus dem Modelltest, zeigt aber, dass auch hier noch keine ausgezeichnete Trennung der beiden Cluster erreicht wurde, da das Dendrogramm keine ausgeprägte Abtrennung aufweist.

Der Vollständigkeit halber ist in Abb.(4.3 rechts) das Ergebnis des Eindeigkeitstest gezeigt. Es ist ersichtlich, dass die Kanäle 1, 2 und 5, die die deutlichste Präsenz eines Herzschlags zeigen, auch am zuverlässigsten abgeschätzt werden können. Die restlichen Kanäle sind durch ihren starken Rauschanteil nicht mehr eindeutig voneinander trennbar.

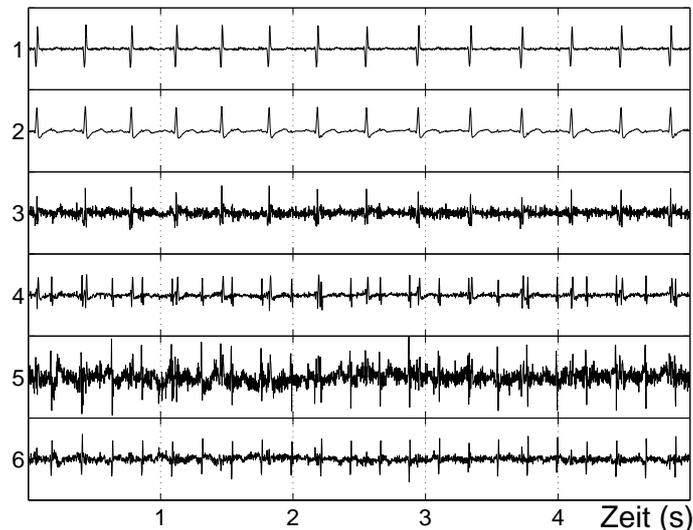
Daher kann man schließen, dass die Separation noch nicht optimal ist und Veränderungen am verwendeten ICA-Modell durchgeführt werden sollten. Zuerst müssen wir uns aber klar werden, in welche Richtung das Modell verbessert werden muss. Dazu hilft es sich vorzustellen, dass nur eine Quelle aktiv wäre. Im Fall des EKGs, sähen dann die verschiedenen Elektroden typischerweise verschiedene Projektionen dieser Quelle und die Messsignale  $x_i(t)$  können deshalb als unterschiedliche dynamische Koordinaten aufgefasst werden. Solche Koordinaten sind dann nicht unabhängig voneinander, können aber auch nicht linear abhängig sein. Wir erwarten in diesem Fall also eine mehrdimensionale Komponente, wie in Kap.(3.2.5) diskutiert. Wie schon von Takens [68] festgestellt, können die zeitverzögerten Varianten eines Kanals ebenfalls als unterschiedliche Koordinaten aufgefasst werden. Somit stellte der ICA-Ansatz aus Gl.(3.40), der mit Verzögerungsvektoren arbeitet, eine Möglichkeit dar, das EKG-Signal besser zu trennen, wenn gleichzeitig auch mehrdimensionale Komponenten zugelassen werden.

#### 4.1. ELEKTROKARDIOGRAMM (EKG) EINER SCHWANGEREN FRAU

---

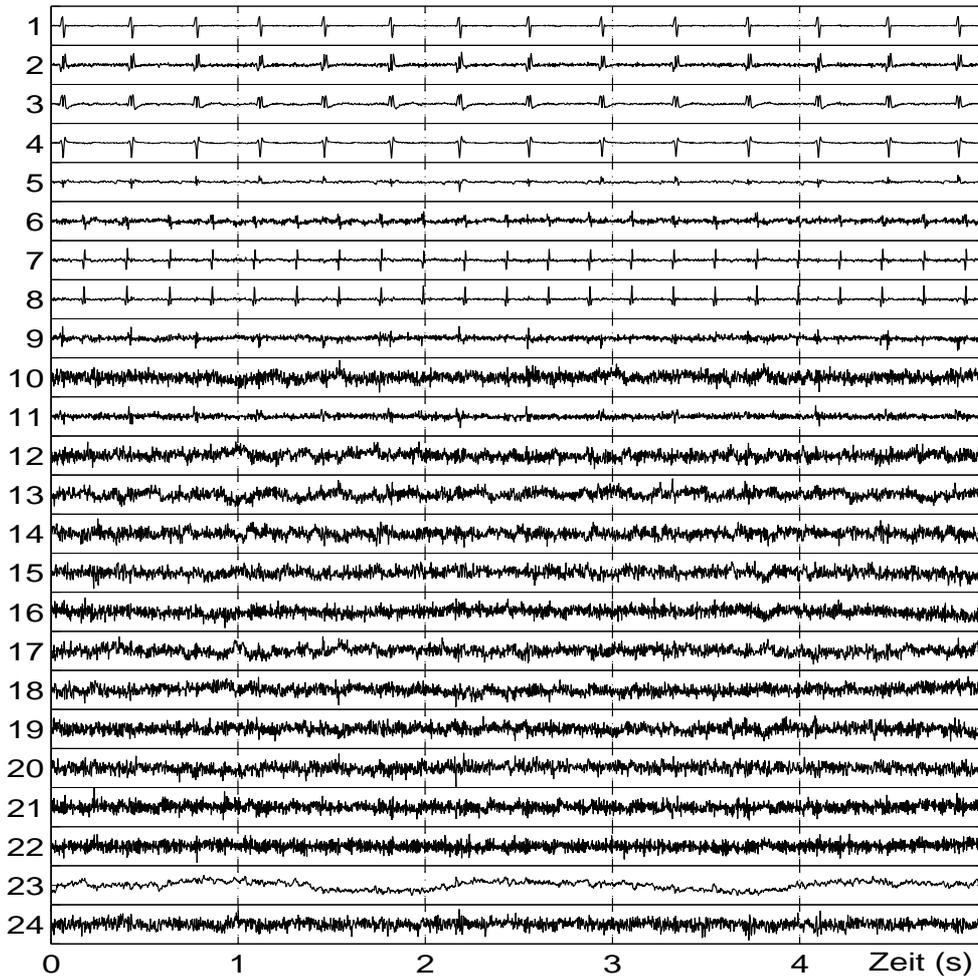


**Abbildung 4.5:** Oberes Bild: Kanal 1 und 2 des EKGs aus Abb.(4.1). Unteres Bild: MILCA-Ausgabe für diese beiden Kanäle.



**Abbildung 4.6:** MILCA-Ausgabe für das eingebettete 2-Kanal-EKG-Signal aus Abb.(4.5 oben) mit Einbettungsdimension  $m = 3$ .

Wir nehmen zuerst einmal an, dass wir nur Zugang zu den ersten beiden Kanäle aus Abb.(4.1) hätten. In beiden Kanälen, die in Abb.(4.5 oben) gezeigt sind, ist der Herzschlag des Fötus gut zu erkennen und eine Separation in Mutter- und Fötus-EKG sollte möglich sein. Bei genauem Hinsehen sieht man, dass beide Kanäle von dem Herzschlag der Mutter dominiert werden, aber dass die R-Zacken in beiden Kanälen eine unterschiedliche Form haben: In der oberen Spur sind sie hauptsächlich positiv, wohingegen sie in der unteren sowohl positiv wie auch negativ sind. Es ist daher klar, dass es keine instantane Mischung der beiden Kanäle geben kann, zu welcher der Herzschlag der Mutter nicht beiträgt. Eine

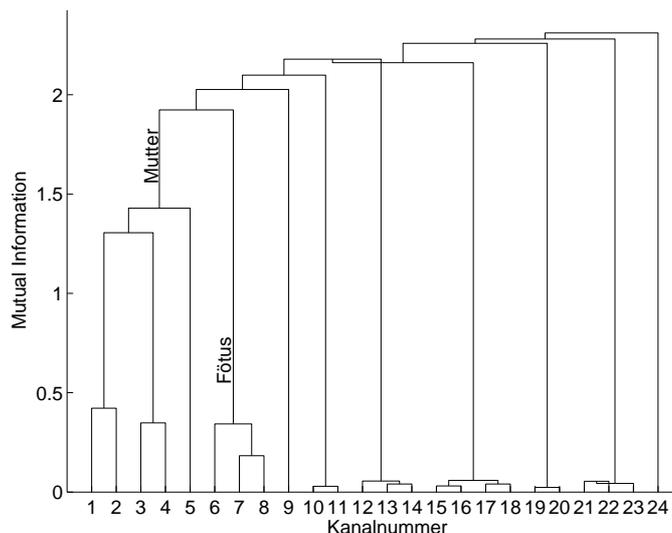


**Abbildung 4.7:** MILCA-Ausgabe für das eingebettete 8-Kanal-EKG ( $k = 100, m = 3$ ).

instantane lineare ICA muss daher für solch einen Fall versagen, wie es in Abb.(4.5 unten) gezeigt ist.

Für den nächsten Versuch, das EKG zu trennen, verwenden wir jetzt den Ansatz aus Gl.(3.40), wobei wir die Einbettungsdimension  $m = 3$  und die kleinstmögliche Verzögerungszeit  $\tau = 1/500 \text{ s}^{-1}$  wählen. Der einzige Unterschied in der Minimierung ist, dass wir jetzt sechs Komponenten haben, die wie bisher sukzessive in den zweidimensionalen Unterräumen minimal abhängig gemacht werden. Die Ergebnisse für die beiden EKG-Kanäle sind in Abb.(4.6) gezeigt. Die Separation ist deutlich verbessert worden. Obwohl wir noch immer eine ICA-Komponente haben, die sowohl Beiträge von der Mutter als auch vom Fötus enthält (Komponente 4), ist der Komponente 6 praktisch ein reines Fötus-EKG.

Schließlich wenden wir noch diese Methode auf alle verfügbaren 8 EKG-Kanäle an. Als Einbettungsdimension wird wieder  $m = 3$  gewählt. Die ICA-Ausgabe enthält somit 24

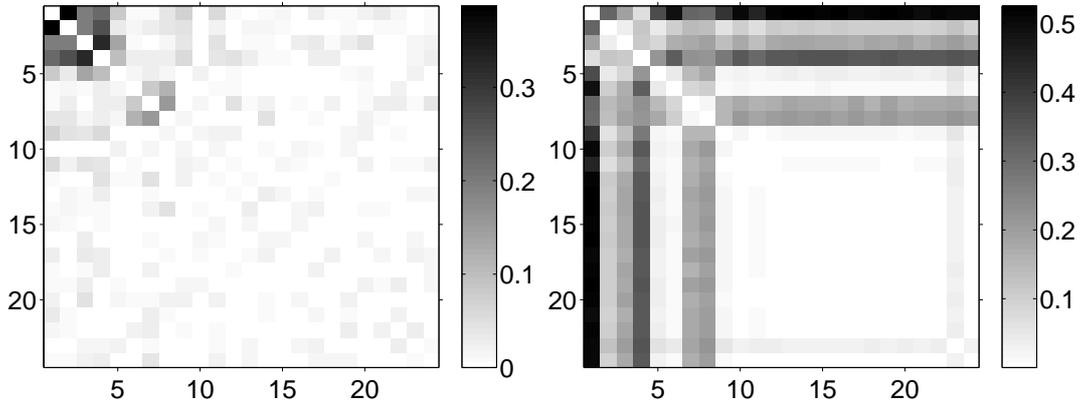


**Abbildung 4.8:** Dendrogramm für Abb.(4.7). Die Höhe jedes Clusters entspricht der MI  $I(X_i, X_j)$  der Cluster  $ij$  ( $k=3$ ).

Kanäle, die in Abb.(4.7) gezeigt sind. Es können klar die Komponenten erkannt werden, die durch den Herzschlag der Mutter, durch den des Fötus und durch Rauschen dominiert werden. Um eine objektivere Aussage zu erhalten, wenden wir wieder unsere Clustermethode auf die ICA-Ausgabe an. Das in Abb.(4.8) dargestellte Dendrogramm zeigt zwei ausgeprägte Cluster, die die Mutter- bzw. die Fötuskanäle enthalten. Es gibt auch etliche kleine Cluster, die als Rauschen betrachtet werden können.

Für beliebige zwei Cluster  $X = X_1 \dots X_p$  und  $Y = Y_1 \dots Y_q$  gilt immer  $I(X, Y) \geq I(X) + I(Y)$ . Dies garantiert, wenn die MI korrekt abgeschätzt wird, dass der Baum korrekt wächst, d.h. jeder Elternknoten über seinen zwei Tochterknoten angeordnet ist. Die zwei kleinen Störungen (wenn die Cluster (1-14) und (15-18) sich verbinden und wenn (21-22) sich mit 23 verbindet) resultieren aus einem kleinen Fehler in der Abschätzung der Abhängigkeiten. Dies beeinflusst aber nicht unsere Schlussfolgerungen.

In Abb.(4.9) zeigen wir die Matrizen der paarweisen MI (links) und der paarweisen Variabilität unter Mischung (rechts). Wie zu erwarten zeigen sie eine viel ausgeprägtere Struktur als die Matrizen in Abb.(4.3). In der Abhängigkeitsmatrix kann nun eine klare Blockstruktur erkannt werden, d.h. die Mutter- und die Fötus-Komponenten sind jetzt in der Tat unabhängig. Aus dem rechten Bild, dem Eindeutigkeitsstest, kann ersehen werden, dass die Mutterkanäle (1-4) und die Fötuskanäle (7-8) stabile Schätzungen darstellen. Die restlichen Kanäle sind hauptsächlich gaußverteilt Rauschen, bis auf Kanal 23, der dem überlagerten Trend des originalen vierten EKG-Kanal entspricht.



**Abbildung 4.9:** Linkes Bild:  $\hat{I}$  zwischen allen paarweisen Kombinationen des Signals aus Abb.(4.7). Rechtes Bild: Quadratwurzel der Variabilität  $\sigma_{ij}$  von  $\hat{I}_{ij}(\phi)$ . In beiden Bildern ist die Diagonale auf Null gesetzt.

Das Endergebnis der BSS erhält man dadurch, dass alle Komponenten, die nicht zum ausgewählten Cluster gehören, gelöscht werden

$$\hat{s}_i(t) \rightarrow \mathcal{P}_C \hat{s}_i(t) \equiv \begin{cases} \hat{s}_i(t) & i \in \text{Cluster } C \\ 0 & \text{sonst} \end{cases} \quad (4.1)$$

und anschließend eine Rücktransformation durchgeführt wird. An dieser Stelle steht man vor dem Problem, dass das rekonstruierte Signal

$$\hat{x}_{j,k}(t; C) = \mathbf{W}_{(j,k),i}^{-1} \mathcal{P}_C \hat{s}_i(t), \quad \mathbf{W}_{i,(j,k)} = w_{ij}^k \quad (4.2)$$

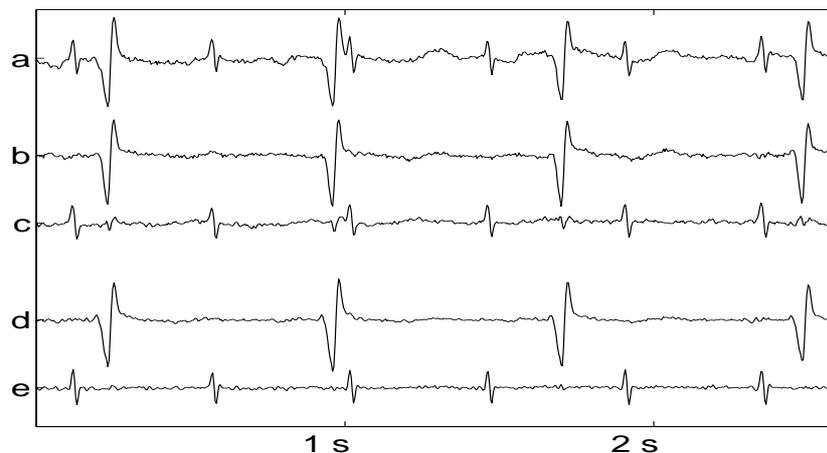
im Allgemeinen kein Verzögerungsvektor ist, d.h.

$$\hat{x}_{j,k+1}(t; C) \neq \hat{x}_{j,k}(t - \tau; C). \quad (4.3)$$

In Anbetracht dessen muss eine heuristische Entscheidung gefällt werden, was als gereinigtes Signal verwendet werden kann. Wir verwenden einfach den Mittelwert

$$\hat{x}_j(t; C) = \frac{1}{m} \sum_{k=1}^m \hat{x}_{j,k}(t + k\tau; C). \quad (4.4)$$

Wir zeigen hier nicht alle acht bereinigten Kanäle des Mutter- bzw. Fötus-Signals, da dies nicht sehr informativ wäre. Die Ergebnisse sind nämlich zu sauber, um eine Beurteilung auf dieser Skala durchzuführen. Anstatt dessen zeigen wir in Abb.(4.10) einen Ausschnitt von einem der originalen EKG-Kanäle und die Beiträge des Herzschlags der Mutter bzw.



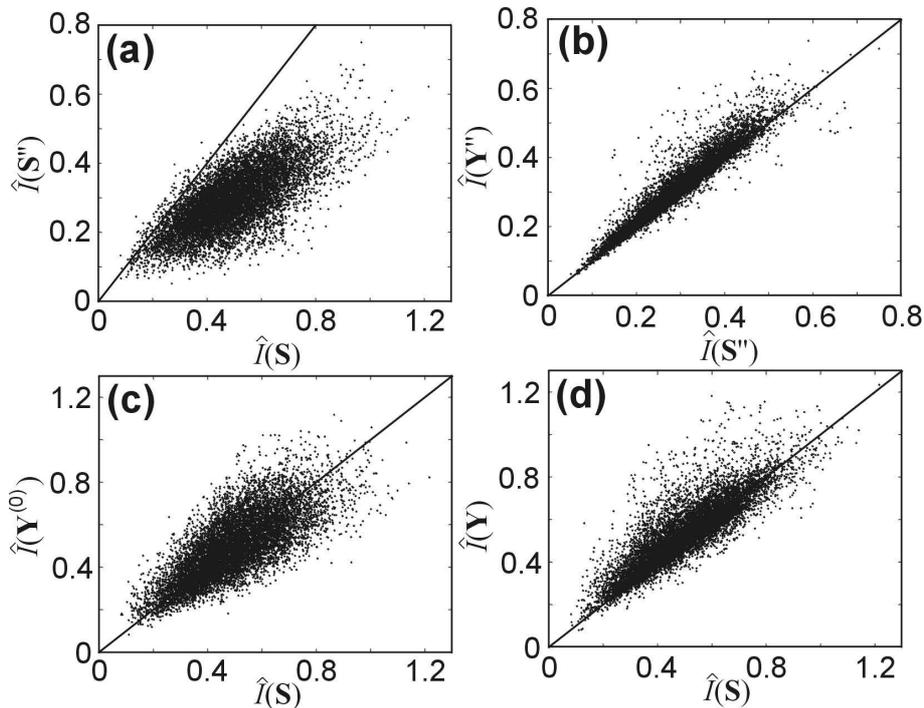
**Abbildung 4.10:** (a) Ausschnitt des originalen EKGs; (b,c) der Mutter- und Fötus-Beitrag, erhalten ohne Verzögerungseinbettung; (d,e) beide Beiträge bei Verwendung der Verzögerungseinbettung.

des Fötus aus der Rücktransformation des jeweiligen Clusters. Die Trennung ist nahezu perfekt.

Das sehr gute Ergebnis für den untersuchten EKG-Datensatz [17] zeigt die Leistungsfähigkeit der ICA in realen Messdaten, obgleich wir eingestehen müssen, dass das Signal-Rausch-Verhältnis in diesem Fall sehr vorteilhaft ist. Die Gewinnungen eines reinen Fötus-EKGs ist von klinischer Relevanz, wird aber aufgrund der erheblichen Schwierigkeiten bei der Extraktion bis jetzt noch nicht praktiziert. Daher wäre es interessant zu sehen, wie MILCA bei ungünstigeren Signal-Rausch-Verhältnissen arbeitet. In diesem Zusammenhang seien noch die Arbeiten [58, 55] erwähnt, in denen versucht wurde, das Fötus-EKG sogar aus univariaten Signalen mit Hilfe lokaler nichtlinearer Methoden zu gewinnen. Es wäre interessant, unsere Methode mit einer Erweiterung der Methode aus [58, 55] auf multivariaten EKG-Signalen zu vergleichen.

## 4.2 Infrarotspektren organischer Moleküle

Eine der Anwendungen, bei denen man einen starken Hinweis auf die Art der Mischung besitzt, ist sicherlich die Spektroskopie. Wir betrachten hier Absorptionsspektren im Infrarot. Das Lambert-Beersche Gesetz besagt, dass sich die gemessene Absorption linear zu den Konzentrationen der untersuchten Stoffe verhält. Somit kann davon ausgegangen werden, dass z.B. ein aufgezeichnetes Infrarotspektrum  $x_i(f)$  einer chemischen Mischung, sich aus einer linearen Mischung der Spektren  $s_i(f)$  der reinen Substanzen zusammensetzt:



**Abbildung 4.11:** Abhängigkeiten zwischen 3-Komponenten-Spektren (10000 Realisationen): (a) MI der zweiten Ableitung der originalen Quellsignale  $S''$  gegen die MI der originalen Quellsignale  $S$ ; (c) MILCA wurde auf die ungefilterten Signale angewandt: MI der originalen Quellsignale  $S$  gegen die MI der geschätzten Quellsignale  $Y^{(0)}$ ; (b,d) MILCA wurde auf die gefilterten Signale angewandt: (b) MI der zweiten Ableitung der originalen Quellsignale  $S''$  gegen die MI der gefilterten geschätzten Quellsignale  $Y''$ , (d) MI der originalen Quellsignale  $S$  gegen die MI der geschätzten Quellsignale  $Y$ ;

$$x_i(f) = \sum_j \mathbf{a}_j s_j(f) . \quad (4.5)$$

Dieser Umstand vereinfacht die Anwendung und im weiteren beschränken wir uns auf ein instantanes lineares ICA-Modell, um einerseits die unbekannte Konzentrationsmatrix und andererseits die unbekanntes Spektren der Reinsubstanzen aus den Spektren einer Mischung organischer Moleküle zu bestimmen.

Um eine BSS zu erzielen, müssen aber auch andere Voraussetzungen in den Messdaten erfüllt werden. So darf für maximal eines der Quellsignale, in diesem Fall sind das die Spektren der Reinsubstanzen, die Werte einer Gaußverteilung<sup>2</sup> folgen. Für den Fall der IR-Spektroskopie ist diese Voraussetzung gewissermaßen immer erfüllt, da die Spektren aufgrund ihrer ausgeprägten Supergaußianität meist sehr große Kurtosiswerte aufweisen.

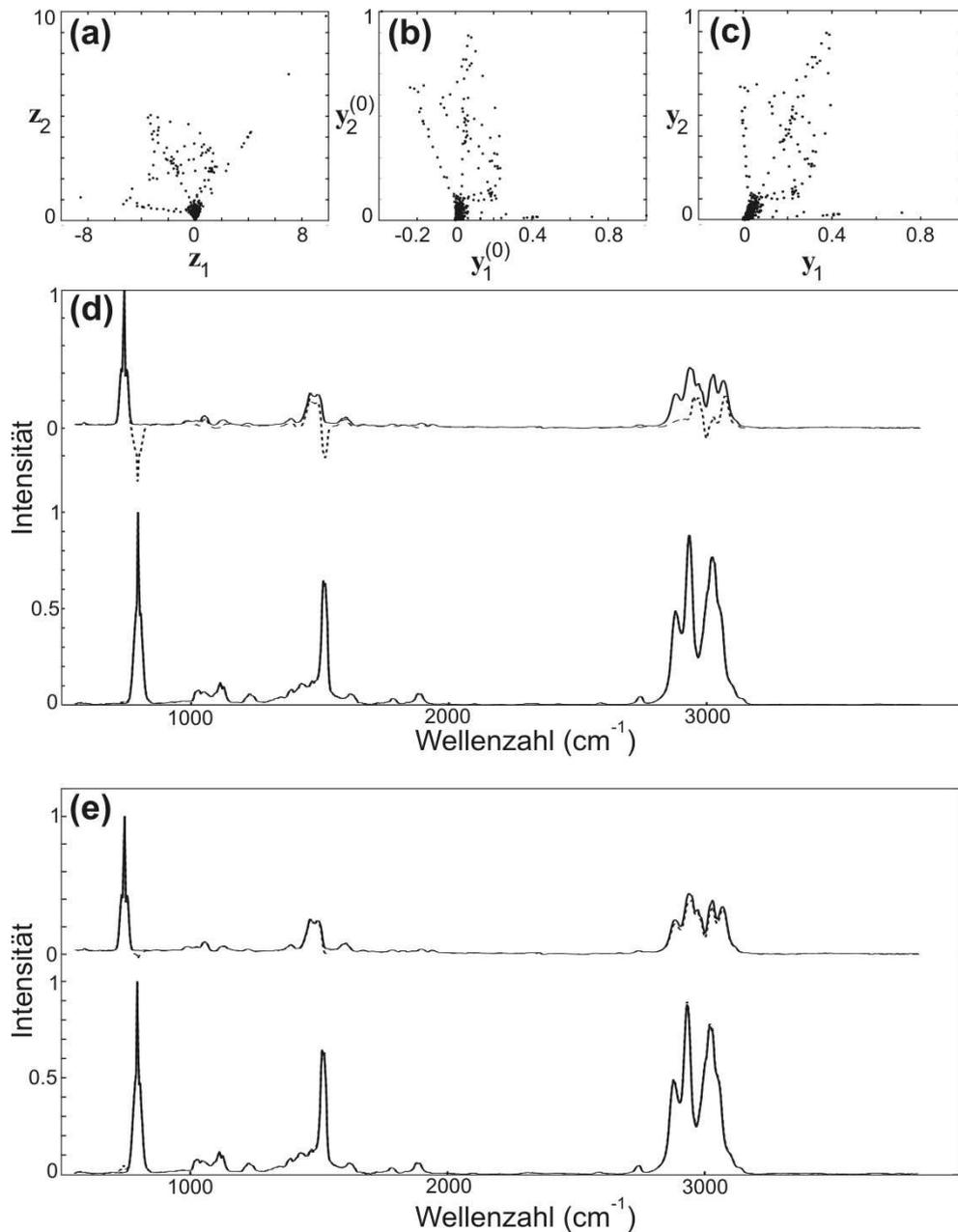
<sup>2</sup>ohne Korrelationen zwischen den Frequenzen.

Die letzte noch zu erfüllende Anforderung an die Messdaten ist die Unabhängigkeit der Quellsignale. Leider treten für IR-Spektren hier aber oft grobe Verletzungen auf. Die Erklärung liegt darin, dass funktionelle Gruppen in einer Substanz, z.B. Hydroxidgruppe, Aminogruppe, Aldehydgruppe usw., immer einen Beitrag zu einem bestimmten Frequenzband liefern. Somit sind die Spektren zweier unterschiedlicher Moleküle, die dieselbe (oder auch nur eine ähnliche) funktionelle Gruppe enthalten, abhängig. Eine Möglichkeit, die Quellsignale unabhängiger zu machen, besteht darin, die Signale Hochpass zu filtern (siehe Kap.(3.2.6.2)), da somit ein stärkeres Gewicht auf die Feinheiten im Spektrum gelegt werden kann. Diese Feinheiten sind charakteristisch für die betreffende Substanz, weil diese die Information über die chemischen Verbindungen zu den nächsten Nachbarn der funktionellen Gruppe enthalten.

Um die generelle Gültigkeit der Verbesserung der Unabhängigkeit durch Hochpassfilterung zu zeigen, wird die MI zwischen einer großen Zahl an IR-Spektren berechnet. Hierfür werden 99 IR-Spektren organischer Moleküle aus der NIST-Datenbank [63] ausgewählt, die im Frequenzbereich  $550\text{-}3830\text{ cm}^{-1}$  (je 822 Datenpunkte) aufgenommen wurden. Daraus wurden 10000mal jeweils drei Spektren ausgewählt und dann ihre Abhängigkeit vor und nach der Filterung bestimmt. Als Hochpassfilter wird hier die zweite Ableitung verwendet, da sie sich bereits für andere BSS-Algorithmen [41, 77, 78, 22] bewährt hat. In Abb.(4.11a) ist zu sehen, dass die Abhängigkeiten zwischen den gefilterten Quellsignalen  $s''$  deutlich erniedrigt wurden. Zwar ist die MI noch immer für die meisten Tripel nicht vertretbar mit Null, aber es ist mit einer signifikanten Leistungssteigerung für die Separation zu rechnen.

Zuerst wollen wir uns aber ansehen, was mit den ICA-Komponenten passiert, wenn die Quellen abhängig sind. Dafür mischen wir die zwei Spektren von o-Xylol und p-Xylol, die eine sehr ähnliche molekulare Struktur und somit stark überlappende Spektren aufweisen. Die Punktvolke in Abb.(4.12a) zeigt, dass der Prewhitening-Schritt aus den anfänglich streng positiven Signalen zu unkorrelierten Vektoren führt, die durch eine reine Rotation nicht mehr vollkommen positiv gemacht werden können. Somit ist es unmöglich, dass die fehlende Rotation, die durch einen beliebigen ICA-Algorithmus bestimmt wird, die exakten Spektren der Reinsubstanzen liefert. Das Ergebnis von MILCA ist in Abb.(4.12b,d) zu sehen. Eine der beiden Komponenten ist zwar sehr gut rekonstruiert worden, die andere aber zeigt wie erwartet grobe Abweichungen vom originalen Quellsignal. Wenn nun aber die Mischung der Spektren vor der Anwendung der ICA gefiltert werden und dann die Inverse der gefundenen Konzentrationsmatrix auf das Mischsignal angewandt wird, erhält man die in Abb.(4.12c,e) gezeigten Schätzungen für die Reinsubstanzen. Diese sind nahezu perfekt und beweisen, dass in manchen Fällen eine BSS auch möglich ist, wenn die originalen Quellen abhängig sind.

Um die Verbesserung in der Trennung durch die Filterung zu untermauern, verwenden wir wieder 99 Spektren aus der NIST-Datenbank und mischen 10000mal jeweils drei Spektren



**Abbildung 4.12:** Mischung von o-Xylol und p-Xylol: (a) gewhitetes Signal; (b) MILCA-Ausgabe; (c) MILCA-Ausgabe, wobei die ICA im gefilterten Raum durchgeführt wurde; (d) die gestrichelte Linie entspricht den Punkten aus (b), die durchgezogene Linie den tatsächlichen Quellsignalen; (e) wie (d), nur jetzt sind die Ergebnisse aus Bild (c) gezeigt.

mit einer zufällig verteilten Matrix. In Abb.(4.13) ist der Amari-Leistungsindex sowohl für MILCA im originalen Raum (diese ICA-Komponenten werden wir mit  $\mathbf{Y}$  bezeichnen), als auch für den gefilterten Raum (diese ICA-Komponenten werden wir mit  $\mathbf{Y}''$  bezeichnen) gegen die MI der Quellsignale aufgetragen. Wie zu erwarten wird die Trennung mit steigender Abhängigkeit der Quellsignale erschwert. Dieser Effekt ist aber für die unbehandelten Signale (Abb.(4.13a)) viel stärker ausgeprägt, als für die gefilterten (Abb.(4.13b)). In den Abb.(4.13c,d) ist ersichtlich, um wie viel sich die Separation verbessert hat, der Mittelwert des Leistungsindex konnte von 0.3 auf weniger als 0.1 verringert werden.

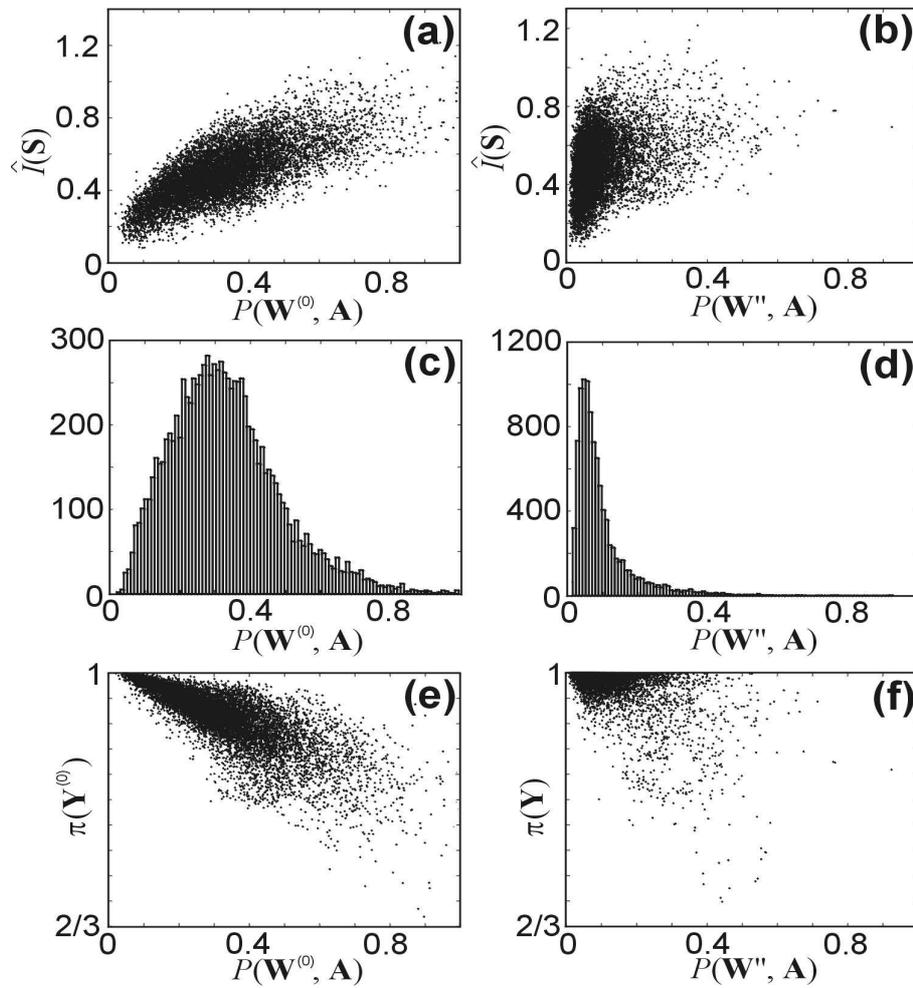
Auch die Positivität der ICA-Komponenten kann als Leistungsmerkmal herangezogen werden, da ein IR-Spektrum im Allgemeinen immer positiv ist<sup>3</sup>. Als ein Maß für die Positivität von  $K$  Spektren  $x_i$  verwenden wir

$$\pi(x_1, \dots, x_K) = \frac{1}{K} \sum_{i=1}^K \frac{\sum_{j: x_{ij} > 0} x_{ij}}{\sum_{j=1}^N |x_{ij}|}. \quad (4.6)$$

Wenn alle Spektren streng positiv sind, ergibt der Index  $\pi(x_1, \dots, x_K) = 1$ , in allen anderen Fällen ist er kleiner als 1. In den Abb.(4.13e,f) ist der Positivitätsindex gegen den Amari-Leistungsindex aufgetragen. Das linke Bild, bei dem MILCA auf die unveränderten Spektren angewandt wurde, enthält eine Vielzahl von ICA-Komponenten, die markante negative Bereiche aufweisen. Außerdem zeigt die Punktwolke eine starke Korrelation zwischen beiden Indices, was ein Hinweis darauf ist, dass die Einbeziehung der Positivität in die Kontrastfunktion der ICA eine Leistungssteigerung verspricht. Anders sieht es im rechten Bild aus, bei dem MILCA auf die gefilterten Spektren angewandt wurde. Hier ist für die Mehrheit der Punkte nur eine minimale Abweichung von einer strikten Positivität der ICA-Komponenten zu sehen und es ist keinerlei Korrelation zwischen den beiden Indices zu erkennen. Dies bedeutet, dass nur geringe Aussichten auf eine Steigerung der Trennleistung durch eine Einbindung der Positivität in die Kontrastfunktion der ICA bestehen.

Um die Unterschiede in den Abhängigkeiten der ICA-Komponenten zu zeigen, werden für die 10000 Realisationen die MI der tatsächlichen Quellsignale  $S$  bzw.  $S''$  gegen die MI der geschätzten Quellsignale  $Y$  bzw.  $Y''$  in Abb.(4.11b,c,d) aufgetragen. Wie zu erwarten war, ist die Punktwolke für die im gefilterten Raum durchgeführte ICA näher um die Diagonale verteilt als für die ungefilterten Spektren. Zusammenfassend lässt sich somit sagen, dass eine Hochpassfilterung, im untersuchten Fall die zweite Ableitung, die Abhängigkeiten von IR-Spektren erheblich reduzieren kann, was zu einer signifikanten Leistungssteigerung der BSS von chemischen Mischungen führt.

<sup>3</sup>In den Messdaten kann es aufgrund von messtechnischen Grundlinienkorrekturen manchmal vorkommen, dass sehr kleine negative Werte im Spektrum enthalten sind.



**Abbildung 4.13:** Leistungsvergleich der MILCA durchgeführt im originalen Raum (linke Spalte) mit der MILCA im gefilterten Raum (rechte Spalte): (a,b) MI der Quellsignale gegen den Amari-Leistungsindex; (c,d) Histogramm des Leistungsindex; (e,f) Positivitätsindex gegen Leistungsindex.

## 4.2. INFRAROTSPEKTREN ORGANISCHER MOLEKÜLE

|          | OPA-ALS | IPCA  | SIMPLISMA | SIMPLISMA-ALS | BTEM  | MILCA |
|----------|---------|-------|-----------|---------------|-------|-------|
| toluene  | 0.996   | 0.968 | 0.971     | 0.973         | 0.954 | 0.987 |
| n-hexane | 0.523   | 0.994 | 0.994     | 0.995         | 0.992 | 0.990 |
| acetone  | 0.639   | 0.873 | 0.866     | 0.899         | 0.886 | 0.933 |
| aldehyde | 0.579   | 0.937 | 0.943     | 0.953         | 0.899 | 0.901 |
| 33DMB    | 0.555   | 0.584 | 0.576     | 0.963         | 0.983 | 0.964 |
| DCM      | 0.502   | 0.971 | 0.969     | 0.967         | 0.904 | 0.909 |

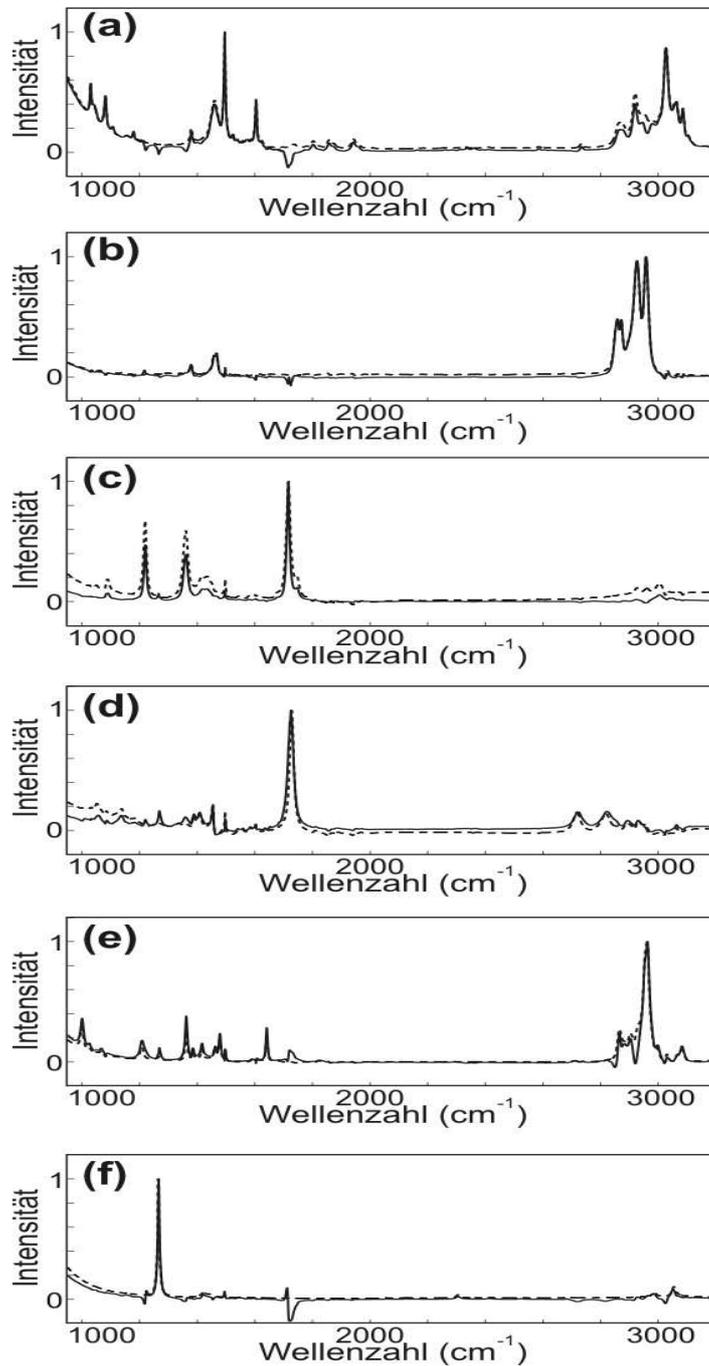
**Tabelle 4.1:** Vergleich der Separationsgüte von verschiedenen BSS-Algorithmen für experimentell gemessene IR-Spektren einer 6-Komponenten-Mischung. Als Leistungsindex ist das normierte Skalarprodukt  $i$  dargestellt. Die Werte der ersten 5 Spalten sind aus Ref. [77] genommen.

Die Analyse von IR-Spektren chemischer Mischungen hat bereits eine lange Tradition und es sind eine Vielzahl an Algorithmen über die Jahre entwickelt worden, die versuchen, die Spektren der Reinsubstanzen zurückzugewinnen. Diese Algorithmen sind mit der Zeit sehr ausgeklügelt worden und so wird meistens über verschiedene Verfahren eine Frequenzbandselektion vorgenommen, um die Überlappungen in den Spektren zu minimieren. Außerdem enthält die Kontrastfunktion meistens viele verschiedene Terme, wie die Entropie, Positivität und andere Größen, die optimiert werden.

Wir wollen hier die Trennleistung von bekannten effizienten Algorithmen mit der von MILCA vergleichen [5]. Dazu verwenden wir vierzehn experimentelle gemessene IR-Spektren [77], die aus sechs Reinsubstanzen (Toluol, n-Hexan, Aceton, 3-Phenylpropanolaldehyd (Aldehyd), 3,3-Dimethylbut-1-en (33DMB) und Dichlormethan (DCM)) zusammen gemischt wurden. Der Frequenzbereich erstreckt sich dabei von 950 bis 3200  $cm^{-1}$ , wobei jedes Spektrum aus 5626 Datenpunkten besteht. Um das Hintergrundrauschen zu unterdrücken, wendeten wir einen Savitzky-Golay-Filter [57] mit 81 Punkten Fensterlänge und einem Polynom siebenter Ordnung auf die Messdaten an. Das Ergebnis des MILCA-Algorithmus ist zusammen mit den Spektren der Reinsubstanzen in Abb.(4.14) zu sehen. Alle Spektren wurden relativ rein rekonstruiert und es sind nur geringe Unterschiede zwischen den originalen Quellsignalen und ihren Schätzungen zu erkennen. Das Ergebnis von MILCA wird in Tab.(4.1) mit denen von SIMPLISMA [78], SIMPLISMA-ALS [71], BTEM [77, 22], IPCA [8] und OPA [56] verglichen. Um konsistent zu sein, verwenden wir ebenfalls den in Ref. [77] benutzten Leistungsindex, das normierte Skalarprodukt

$$i(\mathbf{y}, \mathbf{s}) = \frac{(\mathbf{y} \cdot \mathbf{s})}{|\mathbf{y}| |\mathbf{s}|} \quad (4.7)$$

zwischen dem Spektrum der Reinsubstanz  $\mathbf{s}$  und dem geschätzten Spektrum  $\mathbf{y}$ . Aus der Tabelle ist ersichtlich, dass SIMPLISMA-ALS am effizientesten für diesen Datensatz arbeiten, dicht gefolgt von MILCA und BTEM. Dies ist ein durchaus motivierendes Ergebnis, da bis jetzt nur die einfachste Version des MILCA-Algorithmus verwendet wurde. Dies



**Abbildung 4.14:** Geschätzte Spektren von MILCA (durchgezogene Linie) und Spektren der Reinsubstanzen (gestrichelte Linie) [77]; (a) Toluol; (b) n-Hexan; (c) Aceton; (d) Aldehyd; (e) 33DMB; (f) DCM.

steht im Gegensatz zu anderen Algorithmen, bei denen eine Vielzahl weiterer Tricks und eine Reihe von weiteren Beiträgen zur Kostenfunktion benutzt wurden. Möglichkeiten die Leistung unseres Algorithmus weiter zu steigern, sind z.B. die Einbeziehung der Positivität oder die Nutzung der Korreliertheit der Nachbarfrequenzen (siehe Kap.(3.2.3.4) in die Kontrastfunktion. Außerdem sollte überprüft werden, ob eine Frequenzbandselektion, wie sie in den meisten Algorithmen erfolgt, ebenfalls zu einer Steigerung der Güte der Separation führt.

### 4.3 Synchronisation oder Superposition?

In vielen Bereichen der Wissenschaft wird versucht, immer komplexere Systeme zu untersuchen. An der Front der Forschung steht hier sicherlich das menschliche Gehirn. Es gibt eine Vielzahl von Problemen, die das Verständnis eines solch komplexen Netzwerks erschweren. Um diese Probleme zu umgehen, ist es oft notwendig, das gesamte System konzeptionell in Teilsysteme zu zerlegen, die dann aber sicherlich nicht unabhängig voneinander sind. Eine Methode, um Zusammenhänge zwischen diesen Teilen darzustellen, ist die Synchronisationsanalyse [51]. Dabei werden Maße wie z.B. die Phasenkohärenz [51, 47, 35, 38] dazu benutzt, Verbindungen in diesen Netzwerken zu erkennen. Ein großes Problem ist dabei, dass die Messdaten meist Überlagerungen verschiedener Quellsignale enthalten und somit eine Synchronisation vorgetäuscht wird, obwohl zwischen den Quellen gar keine Verbindung bestehen muss. In dem Fall kann die ICA helfen, um diesen systematischen Fehler der Synchronisationsmaße zu beheben.

Die erzielte Verbesserung der Synchronisationsanalyse wird exemplarisch an einem System von zwei Rössler-Oszillatoren<sup>4</sup> gezeigt, da diese als ein Modell für viele biomedizinische Anwendungen dienen [70, 51]. Im Folgenden werden wir untersuchen, wie sich die Synchronisationsmaße in den beiden folgenden Fällen verhalten:

- Beide Oszillatoren sind unabhängig voneinander und ihre Signale werden mit variierender Stärke gemischt. (Dynamik A)
- Die Signale der Oszillatoren sind dynamisch gekoppelt. Der Kontrollparameter ist die Kopplungsstärke. (Dynamik B)

---

<sup>4</sup>Es wird hier auf eine genaue Beschreibung des Rössler-Modells verzichtet. Interessierte finden eine ausführliche Beschreibung z.B. in Ref. [32, 51]

Im Detail sieht das so aus, dass zuerst aus den Differentialgleichungen der Rössler-Systeme

$$\begin{aligned}
 \dot{v}_1 &= -0.985 v_2 - v_3 + \varepsilon (w_1 - v_1) \\
 \dot{v}_2 &= 0.985 v_1 + 0.15 v_2 \\
 \dot{v}_3 &= 0.2 + v_3 (v_1 - 10) \\
 \dot{w}_1 &= -1.015 w_2 - w_3 + \varepsilon (v_1 - w_1) \\
 \dot{w}_2 &= 1.015 w_1 + 0.15 w_2 \\
 \dot{w}_3 &= 0.2 + w_3 (w_1 - 10).
 \end{aligned} \tag{4.8}$$

die Zeitreihen erzeugt werden ( $dt = 0.1$ ). Für die Dynamik  $B$  wird der Kopplungsparameter  $\varepsilon$  von  $\varepsilon = 0$  (keine Kopplung) bis  $\varepsilon = 0.36$  mit einer Schrittgröße von 0.01 verändert. In der späteren Analyse wird dann die Synchronisation zwischen  $x_1 = v_1$  und  $x_2 = w_1$  berechnet. Für Dynamik  $A$  ist  $\varepsilon = 0$  und die beiden Quellsignale  $v_1$  und  $w_1$  werden durch

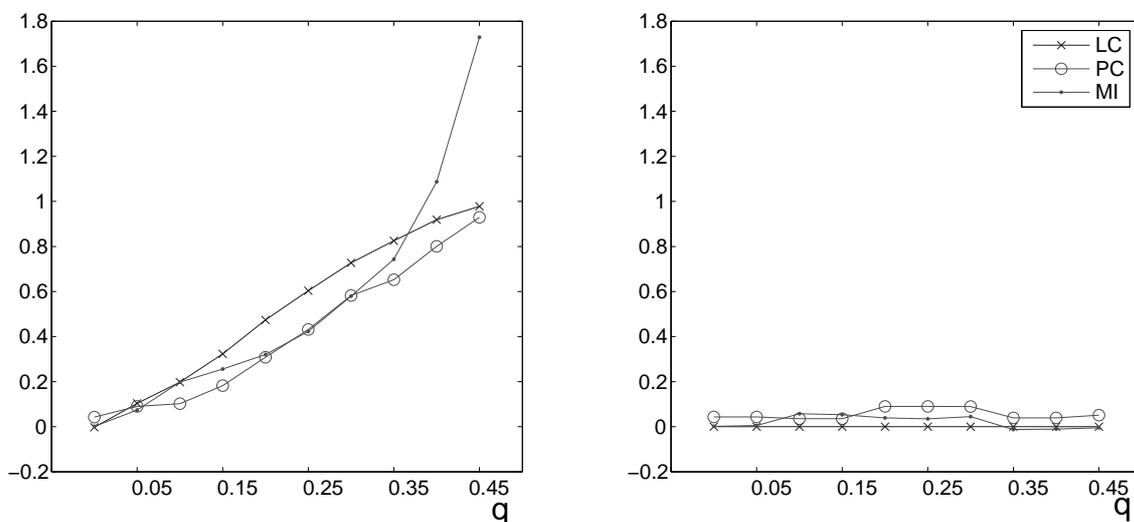
$$\begin{aligned}
 x_1 &= (1 - q)v_1 + qw_1 \\
 x_2 &= qv_1 + (1 - q)w_1
 \end{aligned} \tag{4.9}$$

gemischt, wobei der Mischparameter  $q$  von  $q = 0$  (keine Mischung) bis  $q = 0.45$  mit einer Schrittgröße von 0.05 geändert wird. Somit erhält man für die Dynamik  $A$  10 und für die Dynamik  $B$  insgesamt 37 Zeitreihenpaare  $(x_1, x_2)$ .

Im nächsten Schritt wird die Synchronisation zwischen den Zeitreihenpaaren sowohl vor als auch nach der Anwendung von MILCA berechnet. Als Synchronisationsmaße werden dabei die lineare Korrelation (Gl.(3.7)), die Phasenkohärenz (Gl.(2.27)) und die Mutual Information (Gl.(2.19)) verwendet. Die Ergebnisse für die Dynamik  $A$  sind in Abb.(4.15) zu sehen. Die Werte aller Maße erhöhen sich mit steigender Stärke der Mischung, obwohl die beiden Rössler-Systeme unabhängig sind (linkes Bild). Es ist somit mit diesen Maßen unmöglich zwischen einer echten Synchronisation und einer Überlagerung zu unterscheiden. Mit Verwendung der ICA als Vorverarbeitungsschritt (rechtes Bild) gelingt es aber, diesen systematischen Fehler praktisch zu eliminieren, da wie gewünscht alle Maße über den gesamten Wertebereich von  $q$  annähernd null sind.

Es bleibt noch zu beweisen, dass bei der Analyse von gekoppelten Systemen die ICA nicht die tatsächliche vorhandene Synchronisation zerstört. In Abb.(4.16) ist zu sehen, dass diese Bedingung erfüllt ist. Im linken Bild, in dem die Werte der Maße gegen die Kopplungsstärke  $\varepsilon$  aufgetragen sind, dient gewissermaßen als Sollvorlage. Wenn jetzt die ICA als Vorverarbeitungsschritt verwendet wird, bleiben die Werte der Phasenkohärenz nahezu unverändert. Die Werte der MI verringern sich gerade um den linearen Anteil der Abhängigkeiten, da dieser durch die ICA auf null gesetzt wird (siehe lineare Korrelation in Abb.(4.16 rechts)). Daher empfiehlt es sich, als neues Maß nur den nichtlinearen Anteil der MI,  $\hat{I}(\phi) + \frac{1}{2} \ln(1 - C_{ij}(\phi)^2)$ , zu verwenden. Dieser bleibt dann wie die Phasenkohärenz

### 4.3. SYNCHRONISATION ODER SUPERPOSITION?

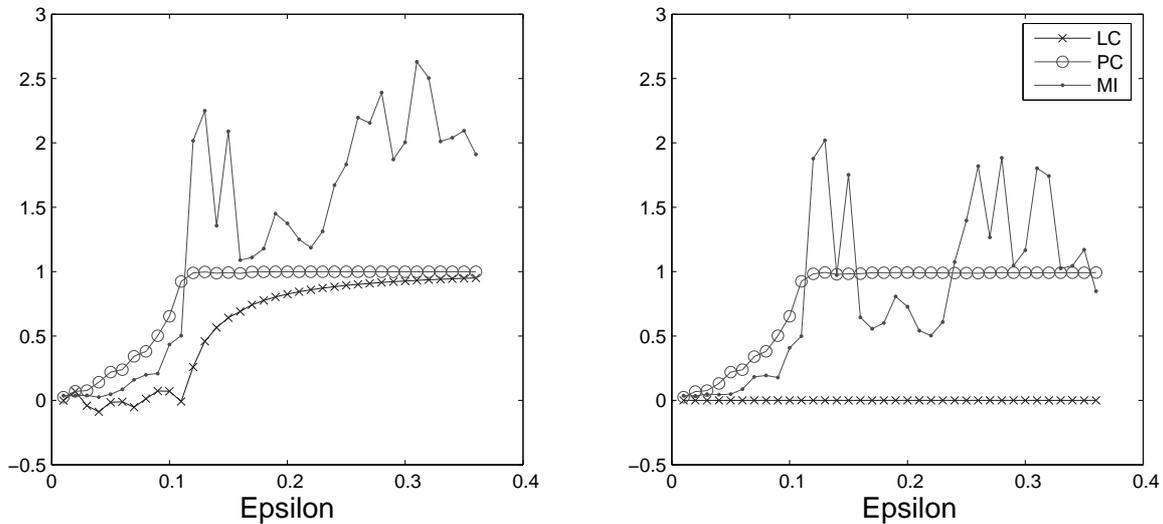


**Abbildung 4.15:** Werte der linearen Korrelation, Phasenkohärenz und Mutual Information gegen den Parameter  $q$  der Mischung (Dynamik  $A$ ). Linkes Bild: ohne ICA-Vorverarbeitung; Rechtes Bild: mit ICA-Vorverarbeitung.

unverändert. Es zeigt sich somit, dass die ICA als Vorverarbeitungsschritt in der Synchronisationsanalyse ein sehr hilfreiches Werkzeug, ist und manche Anwendungen überhaupt erst ermöglicht.

Die gezeigten Ergebnisse gelten für Modelldaten, d.h. genaues Wissen über die Art der Mischung, kein Messrauschen und eine kontrollierte Anzahl von Quellen. Es muss noch gezeigt werden, wie sich die Synchronisationsmaße unter realistischeren Bedingungen verhalten, um die tatsächliche Leistungsgewinnung zu beurteilen. Vor allem ist zu erwarten, dass Probleme bei stark gekoppelten Systemen auftreten werden, weil es Mischungen der Messsignale geben wird, die unabhängiger sind als die ursprünglichen Quellsignale. Dies würde dann zu erniedrigten Wert für die Synchronisation führen.

Bis jetzt werden in der Synchronisationsanalyse so genannte "Surrogate" verwendet, die es erlauben, auf gewisse Null-Hypothesen zu testen. Es ist aber bekannt, dass diese nicht zwischen Daten von gemischten unabhängigen Signalen und nichtlinear gekoppelten Systemen unterscheiden können (siehe z.B. Ref. [3]). Somit ist die ICA eine der vielversprechendsten Methoden, um solche Probleme der Synchronisationsanalyse in realen Messdaten zu lösen.



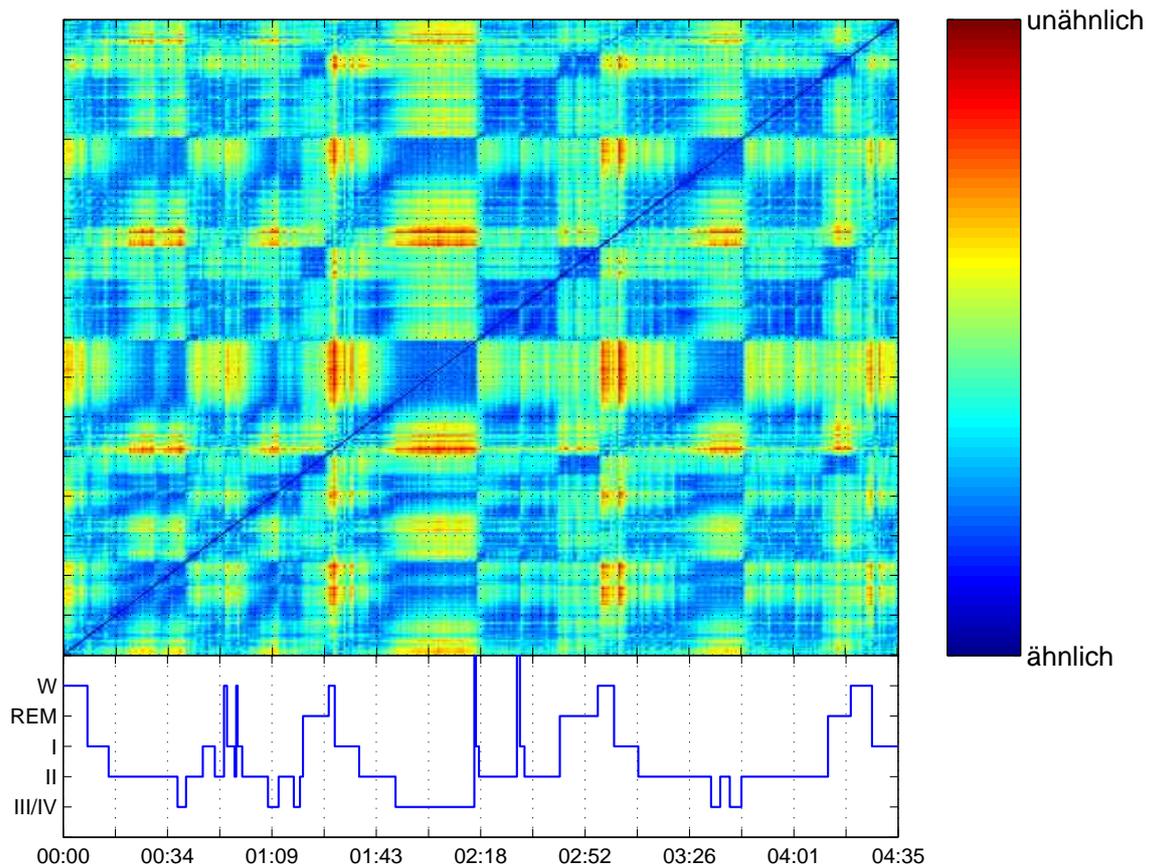
**Abbildung 4.16:** Werte der linearen Korrelation, Phasenkohärenz und Mutual Information gegen den Kopplungsparameter  $q$  (Dynamik  $B$ ). Linkes Bild: ohne ICA-Vorverarbeitung; Rechtes Bild: mit ICA-Vorverarbeitung.

## 4.4 Schlafphasenklassifikation anhand des EEG

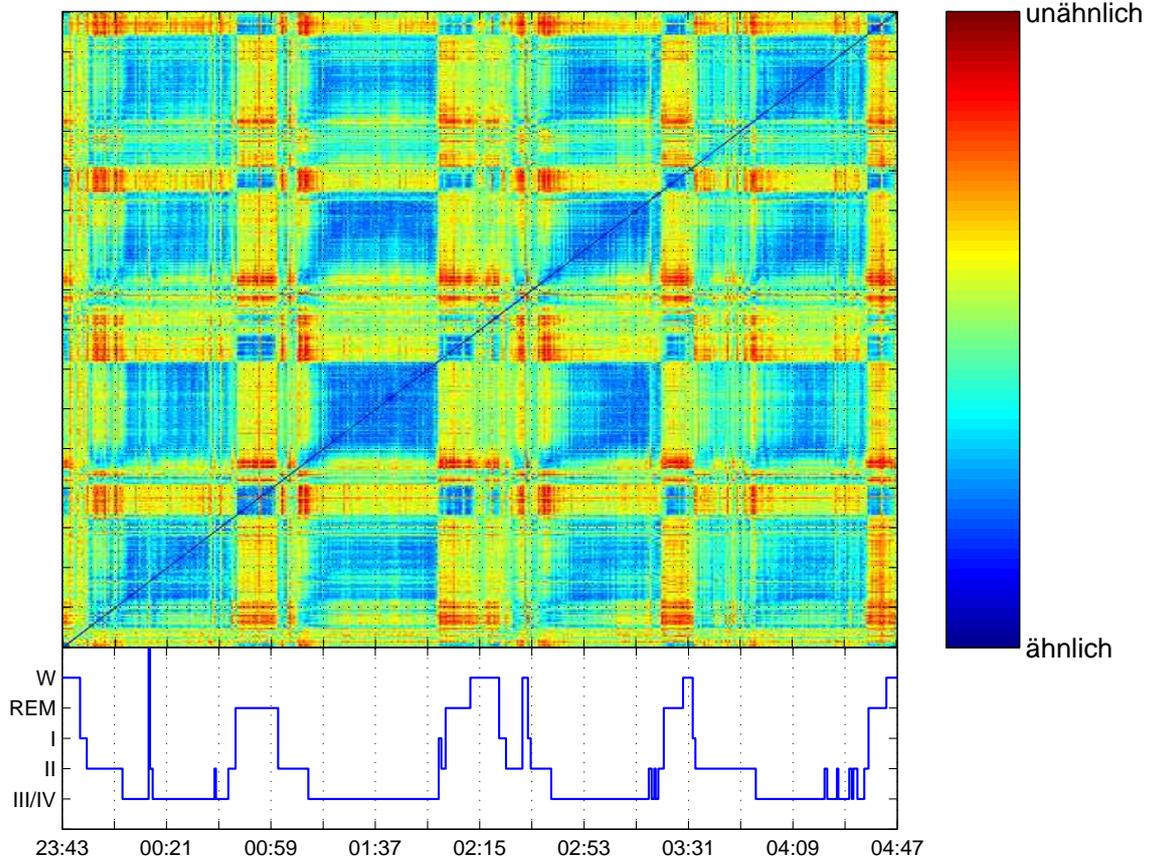
In der klassischen Schlafphasenklassifikation [54] werden die einzelnen Messkanäle auf bestimmte Frequenzanteile und Muster hin untersucht. Dabei wird jeder Kanal als eigenständig betrachtet und für die Zuordnung eines Zeitintervalls (meist erfolgt eine Unterteilung der Signale in 30 Sekundenintervalle) zu einer Schlafphase wird dann die Information aus allen Kanälen zusammengeführt. Wir wollen hier einen anderen Weg gehen. Die Idee ist dabei, nicht nach charakteristischen Merkmalen in den einzelnen Kanälen zu suchen, sondern vielmehr die Verteilung der Quellsignale über alle Kanäle für die Klassifikation zu verwenden [67, 66]. Dazu müsste nur die ICA auf die einzelnen Zeitfenster angewandt werden und die erhaltenen Mischmatrizen miteinander verglichen werden. Wenn jetzt der Patient zu unterschiedlichen Zeiten in die selbe Schlafphase eintritt, müsste sich dies in einer Ähnlichkeit der Matrizen zeigen.

Die im Rahmen dieser Arbeit untersuchten EEG-Daten<sup>5</sup> sind von der Klinik für Epileptologie der Universität Bonn zur Verfügung gestellt worden. Sie wurden mit Hilfe zweier intrahippokampalen Tiefenelektroden [30, 2, 38, 35, 4] (die in der rechten bzw. linken Gehirnhemisphäre platziert waren) mit je 10 Kontakten und einer Abtastrate von 390 Hz aufgezeichnet. Die Auflösung der Analog-Digital-Wandler betrug dabei 12 Bit und zusätzlich erfolgte noch eine Bandpassfilterung (0,05-85 Hz). Zusätzlich wurden simultan noch

<sup>5</sup>Die Hauptziele der EEG-Analyse an der Universität Bonn ist die Verbesserung der Lokalisation des epileptischen Fokus und eine mögliche Anfallsvorhersage.



**Abbildung 4.17:** Zwei unterschiedliche Schlafphasenklassifikationen; Oberes Bild: Ähnlichkeitsindex der Korrelationsmatrizen des Tiefen-EEGs (Segmentlänge: 41 Sekunden); Unteres Bild: klassische Klassifikation aus dem Oberflächen-EEG nach Ref. [54] (Segmentlänge: 30 Sekunden)



**Abbildung 4.18:** Wie Abb.(4.17) jedoch für einen anderen Datensatz.

Signale an 5 Oberflächenelektroden [80], 2 Elektroden an den Augen (Elektrooculographie (EOG)) und 3 Elektroden am Kinn (Elektromyographie (EMG)) aufgezeichnet, mit deren Hilfe eine klassische Schlafphasenklassifikation erfolgte. Diese dient aber ausschließlich Vergleichszwecken, und für die spätere Analyse wurden nur die Signale der Tiefenelektroden verwendet.

Mit dem Ziel, eine ICA-basierte Schlafphasenklassifikation durchzuführen, werden von einer gewählten Hirnseite die 10 Messsignale  $x_i$  in Segmente mit jeweils 16000 Datenpunkten zerlegt, was einer zeitlichen Länge von etwa 41 Sekunden entspricht. Im nächsten Schritt wird eine instantane lineare ICA, in unserem Fall der MILCA-Algorithmus, auf die Segmente angewandt und die Zuverlässigkeit der einzelnen ICA-Komponenten überprüft. Die Komponenten zeigen in dem Modell- und Eindeutigkeits-Test ein sehr ähnliches Verhalten wie die des EKGs der schwangeren Frau (Kap.(4.1)). Wiederum sehen die verschiedenen Elektroden verschiedene Projektionen der Quellen und es werden somit abhängige Komponenten gefunden. Darüber hinaus können viele der Komponenten nicht eindeutig abgeschätzt werden, da sie annähernd gaußverteilt sind. Somit macht es keinen Sinn, die in der ICA gewonnenen Matrizen  $W$  miteinander zu vergleichen, da ein Maß für ihre Ähnlichkeit nur einen aussagekräftigen Wert liefert, wenn alle Komponenten zuverlässig geschätzt werden können.

Als Alternative zur ICA bietet sich aber an, nur eine Hauptachsentransformation mit einer anschließenden Skalierung (Whitening) der Signale durchzuführen, weil durch sie ein multivariates Signal in unkorrelierte, nach ihrer Varianz geordnet, Komponenten zerlegt werden kann. Diese Ordnung wird in allen Segmenten eingeführt und somit ist ein Vergleich ihrer Korrelationsmatrizen sinnvoll. Als Maß für die Ähnlichkeit der Matrizen eignet sich wiederum der Amari-Leistungsindex. Obwohl wir jetzt keine ICA durchführen und keine Chance haben, die eigentlichen Quellsignale zu erhalten, entspricht eine Ähnlichkeit in den Matrizen mit einer hohen Wahrscheinlichkeit<sup>6</sup> auch ähnlichen Quellverteilungen.

In Abb.(4.17) ist ein beispielhaftes Ergebnis für einen Datensatz gezeigt. Die Ähnlichkeit zwischen den Segmenten unterschiedlicher Zeiten ist farblich kodiert, wobei dunkelblau für identische Matrizen steht und dunkelrot den größten gefunden Wert darstellt. Im unteren Teil des Bildes ist zum Vergleich das aus den Oberflächenelektroden gewonnene Somnogramm gezeigt (beide Klassifikationen sind vollkommenen unabhängig voneinander durchgeführt worden). Um das komplexe Bild zu verstehen, ist es am einfachsten, sich zuerst auf einen schmalen horizontalen Streifen zu konzentrieren. Das Auftreten vereinzelter blauer Blöcke ist ein Zeichen dafür, dass die Quellenverteilung über einen längeren

---

<sup>6</sup>Es gibt die Möglichkeit, dass zwei ähnliche Korrelationsmatrizen vollkommen unterschiedliche Quellverteilungen aufweisen. Aber in Anbetracht der Konstanz der Matrix über einen längeren Zeitraum ist es sehr unwahrscheinlich, dass sich die Quellverteilung gerade so verändert hat, dass die Korrelationsmatrix konstant bleibt.

Zeitraum konstant bleibt. Das wiederholte Auftreten eines blauen Blocks in einem Streifen ist ein Hinweis darauf, dass die gleiche Quellenverteilung mehrere Male während der Nacht auftritt. Wenn wir jetzt die Anfangs- und Endzeiten dieser Blöcke mit dem Somnogramm vergleichen, stellen wir fest, dass diese in etwa mit den Anfangs- und Endzeiten der einzelnen Schlafphasen übereinstimmen. Somit ist es möglich, den einzelnen blauen Blöcken entsprechende Schlafstadien zu zuordnen. In Abb.(4.18) ist noch eine Klassifikation für eine andere Nacht mit stärker ausgeprägten Tiefschlafphasen gezeigt. Die sehr gute Übereinstimmung der zwei Klassifikationen zeigt, dass allein aus der Analyse der Korrelationen der Messsignale eine Unterscheidung der Schlafstadien ermöglicht wird. Um eine Aussage über die Trenngüte der Klassifikation zu machen, muss jedoch erst die Zuverlässigkeit an einer größeren Anzahl von Datensätzen überprüft werden.

Die vorgestellte Methode verzichtet vollkommen auf die Analyse der Kurvenform der einzelnen Kanäle und ist somit unabhängig von der Referenzschaltung (z.B. bipolar, Laplace oder gemeinsame Ableitung) der Kanäle. Außerdem werden gegenüber der klassischen Schlafphasenklassifikation weder EOG noch EMG-Signale benötigt. Es konnte somit gezeigt werden, dass multivariate Techniken ebenso eine Klassifikation der Schlafphasen zulassen wie die bisher verwendeten univariaten Techniken. Als ein Vorteil gegenüber letzterer ist die konzeptionelle Einfachheit der neuen Methode zu sehen, die es erleichtern sollte, Unterschiede zwischen verschiedenen Datensätzen besser zu erkennen. Außerdem erhält man mehr bzw. zusätzliche Information über die Schlafphasen, da die Beziehung zwischen ihnen ebenfalls im Ähnlichkeitsbild zu erkennen sind. Es bieten sich somit eine Reihe von Möglichkeiten, weiterführende klinische Untersuchungen mit dieser Methode durchzuführen.

# Kapitel 5

## Zusammenfassung und Ausblick

Das Hauptaugenmerk dieser Dissertation lag auf der Zerlegung von multivariaten Signalen in möglichst unabhängige Komponenten. Um das zu erreichen, wurde ein Maß benötigt, welches so exakt wie möglich die Abhängigkeiten zwischen den Signalen bestimmt. Da die Genauigkeit bisheriger Algorithmen als nicht ausreichend betrachtet wurde, stellte die Entwicklung eines exakten Schätzers [37] für die wechselseitige Abhängigkeit (Mutual Information) den Ausgangspunkt der Arbeit dar. Für eine Vielzahl von Verteilungen, wurde dann in numerischen Tests bewiesen, dass der neue Schätzer im Vergleich zu anderen Schätzern minimalen systematischen und statistischen Fehler besitzt. Darüber hinaus wurde eine auf der MI basierende neuartige Clustermethode [36] entwickelt, die es erlaubt auch die Abstände zwischen höherdimensionalen Objekten zu messen und somit auf eine Mittelung über die Eigenschaften der einzelnen Objekte verzichtet.

Für die Zerlegung von Signalen in möglichst unabhängige Komponenten (Independent Component Analyse (ICA)) zeigte sich, dass der neue MI-Schätzer es ermöglicht, einen ICA-Algorithmus (MILCA) [65] zu formulieren, der hinsichtlich Genauigkeit bisherige Algorithmen übertrifft. Dieser Zugewinn an Präzision muss jedoch mit einem numerischen Mehraufwand erkaufte werden, welcher aber für die meisten Anwendungen vertretbar ist. Ein Vorteil der Verwendung eines exakten Abhängigkeitsmaßes ist, dass damit die ICA zur LCA (Least dependent Component Analyse) erweitert werden kann, da nämlich nicht in allen Anwendungen die Unabhängigkeit der Quellsignale erfüllt ist (Kap.(4.2)). Desweiteren wurde eine MILCA-Variante vorgestellt, die mit Hilfe von Verzögerungsvektoren nicht nur die Abhängigkeiten zwischen instantanen Messpunkten minimiert, sondern auch von zeitlich versetzten, was von Vorteil ist, wenn die Signale hauptsächlich zeitliche Information tragen. Eine weitere Variante des MILCA-Algorithmus, die in Kap.(4.1) erfolgreich angewandt wurde, ermöglicht auch eine Separation der Signale, wenn an den Sensoren verschiedene Projektionen der Quellsignale gemessen werden.

Die Abschätzung der absoluten Abhängigkeit bietet auch die Möglichkeit einen Zuverlässigkeitstest [64] für beliebige ICA-Ausgabe zu formulieren. Ein solcher Test stellt eine notwendige Voraussetzung für die Interpretation jeglicher ICA-Komponenten dar, dies wurde aber erst kürzlich wahrgenommen [45, 23, 26]. Erst nachdem ermittelt worden ist, welche der Ausgabesignale ein- bzw. multidimensionale Komponenten darstellen, kann die Richtigkeit des verwendeten ICA-Modells überprüft werden. Die letzte Neuentwicklung stellt eine spezielle Erweiterung des MILCA-Algorithmus (Kap.(3.2.6.1) dar, die es erlaubt, bei Anwesenheit von Messrauschen in den Daten, bessere Ergebnisse zu erzielen.

Im dritten Teil der Doktorarbeit wurden die vorgestellten Algorithmen auf experimentelle Daten angewandt. Im ersten Beispiel wurde gezeigt, dass sich das EKG einer schwangeren Frau sehr gut in das Mutter- und Fötus-Signal trennen lässt. Eine der Neuheiten dabei ist, dass nach der Anwendung einer erweiterten Variante des MILCA-Algorithmus und des Zuverlässigkeitstests eine Clustermethode dazu benutzt wurde, um multidimensionale Komponenten in der ICA-Ausgabe zu identifizieren.

In einer weiteren Anwendung wurde gezeigt, dass Infrarotspektren von organischen Mischungen, trotz gewisser Abhängigkeiten zwischen den Spektren der Reinsubstanzen, durch eine Vorverarbeitung mit einem Hochpassfilter hervorragend separiert werden können. In einem Vergleich mit verschiedenen modernen Algorithmen, die speziell für die Trennung von Infrarotspektren konstruiert wurden, konnte gezeigt werden, dass MILCA in der Spitzengruppe lag. In weiteren Analysen könnte daher untersucht werden, ob eine Einbeziehung der Positivität die Trennleistung noch weiter steigert. Außerdem könnte in einem erweiterten Algorithmus die Korrelation in den Nachbarfrequenzen oder die häufig angewandte Frequenzbandselektion genutzt werden.

In der dritten Anwendung wird gezeigt, dass die ICA in der Synchronisationsanalyse eingesetzt werden kann, um den systematischen Fehler zu reduzieren, der durch die Überlagerung der Signale entsteht. Dies ist besonderes für viele biomedizinische Anwendungen wichtig, da man dort sehr häufig mit solchen Situationen konfrontiert ist. In zukünftigen Arbeiten sollte der Einfluss von Messrauschen auf diese Korrektur untersucht werden.

In der letzten Anwendung wird gezeigt, dass die alleinige Verwendung der Korrelationen zwischen den Messkanälen in der Lage ist, eine Schlafphasenklassifikation in intrakraniellen EEG-Daten durchzuführen. Die Einfachheit der vorgestellten Methode steht im starken Kontrast zu bisherigen Klassifikationsverfahren und könnte in klinischen Anwendungen erlauben, Unterschiede zwischen z.B. physiologischem und pathologischem Verhalten leichter zu erkennen.

---

In dieser Arbeit wurden neue Algorithmen entwickelt, die in einer Vielzahl von Anwendungen ihre Nützlichkeit bewiesen haben. Die Verbesserung in der Genauigkeit der Schätzung der Abhängigkeit, die Einführung eines Zuverlässigkeitstests, die Formulierung einer auf der MI basierten Clustermethode und die Erweiterung des instantanen linearen Modells stellen für viele Fragestellungen in der BSS hilfreiche Werkzeuge dar. Es sind daher in Zukunft noch eine Reihe weitere Anwendungen zu erwarten.



# Literaturverzeichnis

- [1] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind source separation. In D.S. Touretzky et al., editor, *Advances in Neural Information Processing 8 (Proc. NIPS'95)*, page 757. MIT Press, Cambridge, 1996.
- [2] R. G. Andrzejak. *Epilepsie als eine nichtlinear deterministische Dynamik: Eine Untersuchung hirnelektrischer Aktivität mit Methoden der linearen und nichtlinearen Zeitreihenanalyse*. PhD thesis, Dissertation, Universität Bonn, Deutschland, 2001.
- [3] R. G. Andrzejak, A. Kraskov, H. Stögbauer, F. Mormann, and T. Kreuz. Bivariate surrogate techniques: Necessity, strengths, and caveats. *Phys. Rev. E*, 68:066202, 2003.
- [4] R.G. Andrzejak, K. Lehnertz, R. Rocamora, F. Mormann, P. David, and C. E. Elger. Focus localization using the fraction of nonlinear determinism: influence of sleep depth. *Epilepsia*, 42((Suppl. 2)):176, 2001.
- [5] S. A. Astakhov, H. Stögbauer, A. Kraskov, and P. Grassberger. Spectral mixture decomposition by least dependent component analysis. *Anal. Chim. Acta (submitted)*, 2005. E-print, [ul.arxiv.org/abs/physics/0412029](http://ul.arxiv.org/abs/physics/0412029).
- [6] F. R. Bach and M. I. Jordan. Kernel independent component analysis. *J. Machine Learning Res.*, 3:1, 2002.
- [7] A. Belouchrani, K. Abed Meraim, J. F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Signal Processing*, 45(2):434, 1997.
- [8] D. Bu and C. W. Brown. Self-modeling mixture analysis by interactive principal component analysis. *Appl. Spectrosc.*, 54:1214, 2000.
- [9] J. F. Cardoso. Multidimensional independent component analysis. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 98)*, page 1941, 1998.
- [10] J. F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 1:157, 1999.
- [11] J. F. Cardoso and A. Souloumiac. Joint angles for simultaneous diagonalization. *SIAM Journal of Matrix Analysis and Applications*, 17(1):161, 1996.

- 
- [12] A. Cichocki and S. Amari. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. Wiley, 2002.
- [13] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.
- [14] G. A. Darbellay. An Estimator of the Mutual Information Based on a Criterion for Independence. *Computational Statistics and Data Analysis*, 32:1, 1999.
- [15] G. A. Darbellay and I. Vajda. Technical report no. 1921. Inst. of Information Theory and Automation, 1998. <http://siprint.utia.cas.cz/darbellay/>.
- [16] G. A. Darbellay and I. Vajda. Estimation of the information by an adaptive partitioning of the observation space. *IEEE Trans. Inform. Th.*, 45:1315, 1999.
- [17] B. L. R. De Moor. Daisy: Database for the identification of systems, 1997. [www.esat.kuleuven.ac.be/sista/daisy](http://www.esat.kuleuven.ac.be/sista/daisy).
- [18] A. Fraser and H. Swinney. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, 33:1134, 1986.
- [19] P. Grassberger. Generalizations of the Hausdorff dimension of fractal measures. *Phys. Lett. A*, 107(3):101–105, 1985.
- [20] P. Grassberger. Finite sample corrections to entropy and dimension estimates. *Phys. Lett. A*, 128:369, 1988.
- [21] P. Grassberger. An optimized box-assisted algorithm for fractal dimensions. *Phys. Lett. A*, 148(1–2):63–68, 1990.
- [22] L. Guo, F. Kooli, and M. Garland. A General Method for the recovery of Pure Powder XRD Patterns from Complex Mixtures Using no a priori Information. Application of Band-Target Entropy Minimization (BTEM) to Materials Characterization of Inorganic Mixtures. *Anal. Chim. Acta*, 517:229, 2004.
- [23] S. Harmeling, F. Meinecke, and K. R. Müller. Analysing ICA component by injection noise. In *Proc. Int. Workshop on Independent Component Analysis*, Nara, Japan, 2003.
- [24] S. Harmeling, A. Ziehe, M. Kawanabe, and K. R. Müller. Kernel-based Nonlinear Blind Source Separation. *Neural Computation*, 15:1089, 2003.
- [25] R. Hegger, H. Kantz, and T. Schreiber. Practical implementation of nonlinear time series methods: The TISEAN package. *CHAOS*, 9:413, 1999.
- [26] J. Himberg and A. Hyvärinen. Icasso: software for investigating the reliability of ICA estimates by clustering and visualization. In *Proceedings of the Workshop on Neural Networks and Signal Processing*, 2003.
- [27] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis*. John Wiley and Sons, New York, 2001.

- [28] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483, 1997.
- [29] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ, 1988.
- [30] J. Engel (Jr.), editor. *Surgical Treatment of the Epilepsies*. Raven Press, New York, 1993.
- [31] C. Jutten. Source separation: from dusk till dawn. In *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA '2000)*, page 15, Helsinki, Finland, 2000.
- [32] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge Univ. Press, Cambridge, UK, 1997.
- [33] L. F. Kozachenko and N. N. Leonenko. Sample estimate of the entropy of a random vector. *Problem. Inf. Transm.*, 23:95, 1987.
- [34] K. Kraisek. Charakterisierung der Dynamik räumlich ausgedehnter Systeme mit der multivariaten spatio-temporalen Zustandsraumkonstruktion. Master's thesis, Diplomarbeit, Universität Bonn, Deutschland, 2003.
- [35] A. Kraskov. *Synchronization and Interdependence Measures and their Applications to the Electroencephalogram of Epilepsy Patients and Clustering of Data*. PhD thesis, Dissertation, Universität Wuppertal, Deutschland, 2004.
- [36] A. Kraskov, H. Stögbauer, R. G. Andrzejak, and P. Grassberger. Hierarchical clustering based on mutual information. *Europhysics Letters (submitted)*, 2004. E-print, [arXiv.org/q-bio/0311039](http://arXiv.org/q-bio/0311039).
- [37] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Phys. Rev. E*, 69:66138, 2003.
- [38] T. Kreuz. *Measuring Synchronization in Model Systems and Electroencephalographic Time Series from Epilepsy Patients*. PhD thesis, Dissertation, Universität Wuppertal, Deutschland, 2003.
- [39] R. H. Lambert and C. L. Nikias. *Unsupervised adaptive filtering*, volume I, chapter Blind deconvolution of multipath mixtures, pages 377–436. Wiley, 2000.
- [40] L. D. Lathauwer, B. D. Moor, and J. Vandewalle. Fetal electrocardiogram extraction by source subspace separation. In *Processing of HOS*, Aiguabla, Spain, 1995.
- [41] W. H. Lawton and E. A. Sylvestre. Self modeling curve resolution. *Technometrics*, 13:617, 1971.
- [42] E. G. Learned-Miller and J. W. Fisher III. ICA using spacings estimates of entropy. *J. Maschine Learning Res.*, 4:1271, 2002.
- [43] T. W. Lee, M. Girolami, and T. J. Sejnowski. Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. *Neural Computation*, 11(2):417, 1999.

- 
- [44] M. Li, J. H. Badger, X. Chen, S. Kwong, P. Kearney, and H. Zhang. An information-based sequence distance and its application to whole mitochondrial genome phylogeny. *Bioinformatics*, 17(2):149–154, 2001.
- [45] F. Meinecke, A. Ziehe, M. Kawanabe, and K. R. Müller. A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE Trans. Biomed. Eng.*, 49:1514, 2002.
- [46] L. Molgedey and H.G. Schuster. Separation of a Mixture of Independent Signals Using Time Delayed Correlations. *Phys. Rev. Letter*, 72:3634, 1994.
- [47] F. Mormann, K. Lehnertz, P. David, and C. E. Elger. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D*, 144:358, 2000.
- [48] K. R. Müller, P. Philips, and A. Ziehe. JADE-TD: Combining higher-order statistics and temporal information for blind source separation (with noise). In J. F. Cardoso, C. Jutten, and P. Loubaton, editors, *Proceedings of the International Workshop on Independent Component Analysis*, page 87, Aussois, France, 1999.
- [49] D. Nuzillard, S. Bourg, and J. M. Nuzillard. Model-Free Analysis of Mixtures by NMR Using Blind Source Separation. *J. Magn. Reson.*, 133:358, 1998.
- [50] D. T. Pham. Blind Separation of Instantaneous Mixtures of Sources Based on Order Statistics. *IEEE Trans. Signal Process.*, 48:363, 2000.
- [51] A. S. Pikovsky, M. G. Rosenblum, and J. Kurths. *Synchronization. A universal concept in nonlinear sciences*. Cambridge Univ. Press, Cambridge, UK, 2001.
- [52] W. H. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical recipes in C: The art of scientific computing*. Cambridge Univ. Press, Cambridge, UK, second edition, 1992.
- [53] R. Quiñero, A. Kraskov, T. Kreuz, and P. Grassberger. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Phys. Rev. E*, 65:041903, 2002.
- [54] A. Rechtschaffen and A. Kales. A manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects. NIH Publication 204, Public Health Service, US Government Printing Office, Washington DC, 1968.
- [55] M. Richter, T. Schreiber, and D. T. Kaplan. Fetal ECG extraction with nonlinear state-space projections. *Bio-Med. Eng.*, 45:133, 1998.
- [56] C. Sánchez, J. Toft, B. van den Bogaert, and D. L. Massart. Orthogonal projection approach applied to peak purity assessment. *Anal. Chem.*, 68:79, 1996.
- [57] A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 36:1627, 1964.
- [58] T. Schreiber and D. T. Kaplan. Signal separation by nonlinear projections: The fetal electrocardiogram. *Phys. Rev. E*, 53:4326, 1996.

- [59] C. E. Shannon. A mathematical theory of communication. *Bell System Technol. J.*, 27:379, 1948.
- [60] B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26 of *Monographs on Statistical and Applied Probability*. Chapman & Hall, London, 1986.
- [61] R. L. Somorjai. *Methods for Estimating the Intrinsic Dimensionality of High-Dimensional Point Sets*. Springer, 1986.
- [62] C. Spearman. General intelligence, objectively determined and measured. *American J. of Psychology*, 15:201, 1904.
- [63] NIST Mass Spec Data Center. S. E. Stein. “infrared spectra” in nist chemistry web-book, March 2003. <http://webbook.nist.gov>.
- [64] H. Stögbauer, R. G. Andrzejak, A. Kraskov, and P. Grassberger. Reliability of ICA estimates with mutual information. In *Lecture Notes in Computer Science*, page 3195, Granada, Spain, 2004. Springer.
- [65] H. Stögbauer, A. Kraskov, and P. Grassberger. Least dependent component analysis based on mutual information. *Phys. Rev. E*, 70:66123, 2004.
- [66] H. Stögbauer, A. Kraskov, T. Kreuz, R. G. Andrzejak, and P. Grassberger. Multivariate analysis: Characterization of physiological state from intracranial eeg recordings. international symposium on topical problems of nonlinear wave physics. In *Proceeding of NWP*, page 119, Nizhny Novgorod, Russia, 2003.
- [67] H. Stögbauer, L. Yang, P. Grassberger, R. G. Andrzejak, T. Kreuz, A. Kraskov, C. E. Elger, and K. Lehnertz. Lateralization of the focal hemisphere in mesial temporal lobe epilepsy using independent component analysis. *Epilepsia*, 43(7):51, 2002.
- [68] F. Takens. Detecting strange attractors in turbulence. In D. A. Rand and L. S. Young, editors, *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics*, page 366. Springer-Verlag, Berlin, 1980.
- [69] A. Taleb and C. Jutten. Nonlinear source separation: The postlinear mixtures. In *Proc. European Symposium on Artificial Neural Networks*, page 279, Bruges, Belgium, 1997.
- [70] P. A. Tass, M. G. Rosenblum, J. Weule, J. Kurths, A. Pikovsky, J. Volkmann, A. Schnitzler, and H. J. Freund. Detection of n:m phase locking from noisy data: Application to magnetoencephalography. *Phys. Rev. Lett.*, 81(15):3291, 1998.
- [71] R. Tauler, B. Kowalski, and S. Fleming. Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Anal. Chem.*, 65:2040, 1993.
- [72] K. Torkkola. Blind separation for audio signals - are we there yet? In *Proc. int. workshop on independent component analysis*, Aussois, France, 1999.
- [73] K. Torkkola. *Unsupervised adaptive filtering*, volume I, chapter Blind separation of delayed and convolved sources, pages 321–375. Wiley, 2000.

- [74] V. N. Vapnik. *Estimation of dependences based on empirical data*. Springer, New York, 1982.
- [75] J. D. Victor. Binless strategies for estimation of information from neural data. *Phys. Rev. E*, 66:051903–1, 2002.
- [76] R. Vigario, J. Särelä, V. Jousmäki, and E. Oja. Independent component approach to the analysis of eeg and meg recordings. *IEEE Trans. Biomedical Engineering*, 47(5):589, 2000.
- [77] E. Widjaja, C. Li, W. Chew, and M. Garland. Band-target entropy minimization. a robust algorithm for pure component spectral recovery. application to complex randomized mixtures of six components. *Anal. Chem.*, 75:4499, 2003.
- [78] W. Windig and J. Guilment. Interactive self-modeling mixture analysis. *Anal. Chem.*, 63:1425, 1991.
- [79] A. Ziehe and K. R. Müller. TDSEP-an efficient algorithm for blind separation using time structure. In *Proc. Int. Conf. on Artificial Neural Networks*, page 675, Skövde, Schweden, 1998.
- [80] S. Zschocke. *Klinische Elektroenzephalographie*. Springer, Berlin, 2002.

# Danksagung

Bei meinem Doktorvater Prof. Dr. Peter Grassberger möchte ich mich bedanken, dass er es mir ermöglicht hat, diese Doktorarbeit am John von Neumann-Institut für Computing im Forschungszentrum Jülich durchzuführen. Besonders dankbar bin ich ihm für sein ständiges Interesse, die vielen lehrreichen Gespräche und die spannenden Diskussionen, die zum Gelingen dieser Doktorarbeit maßgeblich beigetragen haben. Außerdem bin ich ihm dankbar dafür, dass er es mir ermöglicht hat, meine Forschungsergebnisse auf zahlreichen nationalen und internationalen Konferenzen zu präsentieren und dort Kontakte zu vielen interessanten Wissenschaftlern zu knüpfen.

Den Kollegen am John von Neumann-Institut für Computing, Dr. Ralph G. Andrzejak, Dr. Hsiao-Ping Hsu, Dr. Thomas Kreuz, Dr. Walter Nadler, Dr. Alexander Kraskov, Dr. Sergey Astakhov, Dr. Lei Yang, Dr. Vishal Mehra und Dr. Rodrigo Quian Quiroga, danke ich für das ausgezeichnete Arbeitsklima sowie für viele interessante und weiterführende Diskussionen. Bei Dr. Ralph G. Andrzejak, Dr. Sergey Astakhov, Dr. Lei Yang, Dr. Thomas Kreuz und Dr. Alexander Kraskov möchte ich mich sehr für ihre kollegiale Zusammenarbeit und ständige Hilfsbereitschaft bedanken. Außerdem sei Dr. Ralph G. Andrzejak und Dr. Thomas Kreuz für das aufmerksame Korrekturlesen der vorliegenden Arbeit gedankt. Darüber hinaus möchte ich mich bei Dr. Lei Yang bedanken, dass er es mir ermöglicht hat, eine Vielzahl an Universitäten in China zu besuchen, um dort meine Arbeit vorzustellen.

Herrn Prof. Dr. Christian E. Elger und Herrn Priv. Doz. Dr. Klaus Lehnertz von der Klinik für Epileptologie in Bonn gilt mein Dank für die Bereitstellung zahlreicher EEG-Datensätze. Desweiteren möchte ich mich bei Dr. Dr. Florian Mormann für die vielen anregenden Diskussionen bedanken.

Außerdem gilt mein Dank Frau Helga Frank, die eine unersetzliche Hilfe für mich war.

Besonders möchte ich meinen Eltern für den Rückhalt und die Unterstützung während meiner wissenschaftlichen Ausbildung danken.

Bei meiner Frau Lucija Stögbauer bedanke ich mich sehr für ihre Geduld und ständige Unterstützung in allen Lebenslagen. Ohne sie wäre dies alles gar nicht möglich gewesen.

Jülich, im Januar 2005

Harald Stögbauer



Bisher sind erschienen:

**Modern Methods and Algorithms of Quantum Chemistry -  
Proceedings**

Johannes Grotendorst (Hrsg.)

Winterschule, 21. - 25. Februar 2000, Forschungszentrum Jülich

NIC-Serie Band 1

ISBN 3-00-005618-1, Februar 2000, 562 Seiten

*nicht mehr lieferbar*

**Modern Methods and Algorithms of Quantum Chemistry -  
Poster Presentations**

Johannes Grotendorst (Hrsg.)

Winterschule, 21. - 25. Februar 2000, Forschungszentrum Jülich

NIC-Serie Band 2

ISBN 3-00-005746-3, Februar 2000, 77 Seiten

*nicht mehr lieferbar*

**Modern Methods and Algorithms of Quantum Chemistry -  
Proceedings, Second Edition**

Johannes Grotendorst (Hrsg.)

Winterschule, 21. - 25. Februar 2000, Forschungszentrum Jülich

NIC-Serie Band 3

ISBN 3-00-005834-6, Dezember 2000, 638 Seiten

**Nichtlineare Analyse raum-zeitlicher Aspekte der  
hirnelektrischen Aktivität von Epilepsiepatienten**

Jochen Arnold

NIC-Serie Band 4

ISBN 3-00-006221-1, September 2000, 120 Seiten

**Elektron-Elektron-Wechselwirkung in Halbleitern:  
Von hochkorrelierten kohärenten Anfangszuständen  
zu inkohärentem Transport**

Reinhold Löwenich

NIC-Serie Band 5

ISBN 3-00-006329-3, August 2000, 146 Seiten

**Erkennung von Nichtlinearitäten und  
wechselseitigen Abhängigkeiten in Zeitreihen**

Andreas Schmitz

NIC-Serie Band 6

ISBN 3-00-007871-1, Mai 2001, 142 Seiten

**Multiparadigm Programming with Object-Oriented Languages -  
Proceedings**

Kei Davis, Yannis Smaragdakis, Jörg Striegnitz (Hrsg.)

Workshop MPOOL, 18. Mai 2001, Budapest

NIC-Serie Band 7

ISBN 3-00-007968-8, Juni 2001, 160 Seiten

**Europhysics Conference on Computational Physics -  
Book of Abstracts**

Friedel Hossfeld, Kurt Binder (Hrsg.)

Konferenz, 5. - 8. September 2001, Aachen

NIC-Serie Band 8

ISBN 3-00-008236-0, September 2001, 500 Seiten

**NIC Symposium 2001 - Proceedings**

Horst Rollnik, Dietrich Wolf (Hrsg.)

Symposium, 5. - 6. Dezember 2001, Forschungszentrum Jülich

NIC-Serie Band 9

ISBN 3-00-009055-X, Mai 2002, 514 Seiten

**Quantum Simulations of Complex Many-Body Systems:  
From Theory to Algorithms - Lecture Notes**

Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Hrsg.)

Winterschule, 25. Februar - 1. März 2002, Rolduc Conference Centre,

Kerkrade, Niederlande

NIC-Serie Band 10

ISBN 3-00-009057-6, Februar 2002, 548 Seiten

**Quantum Simulations of Complex Many-Body Systems:  
From Theory to Algorithms - Poster Presentations**

Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Hrsg.)

Winterschule, 25. Februar - 1. März 2002, Rolduc Conference Centre,

Kerkrade, Niederlande

NIC-Serie Band 11

ISBN 3-00-009058-4, Februar 2002, 194 Seiten

**Strongly Disordered Quantum Spin Systems in Low Dimensions:  
Numerical Study of Spin Chains, Spin Ladders and  
Two-Dimensional Systems**

Yu-cheng Lin

NIC-Serie Band 12

ISBN 3-00-009056-8, Mai 2002, 146 Seiten

**Multiparadigm Programming with Object-Oriented Languages -  
Proceedings**

Jörg Striegnitz, Kei Davis, Yannis Smaragdakis (Hrsg.)

Workshop MPOOL 2002, 11. Juni 2002, Malaga

NIC-Serie Band 13

ISBN 3-00-009099-1, Juni 2002, 132 Seiten

**Quantum Simulations of Complex Many-Body Systems:  
From Theory to Algorithms - Audio-Visual Lecture Notes**

Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Hrsg.)

Winterschule, 25. Februar - 1. März 2002, Rolduc Conference Centre,

Kerkrade, Niederlande

NIC-Serie Band 14

ISBN 3-00-010000-8, November 2002, DVD

**Numerical Methods for Limit and Shakedown Analysis**

Manfred Staat, Michael Heitzer (Hrsg.)

NIC-Serie Band 15

ISBN 3-00-010001-6, Februar 2003, 306 Seiten

**Design and Evaluation of a Bandwidth Broker that Provides  
Network Quality of Service for Grid Applications**

Volker Sander

NIC-Serie Band 16

ISBN 3-00-010002-4, Februar 2003, 208 Seiten

**Automatic Performance Analysis on Parallel Computers with  
SMP Nodes**

Felix Wolf

NIC-Serie Band 17

ISBN 3-00-010003-2, Februar 2003, 168 Seiten

**Haptisches Rendern zum Einpassen von hochaufgelösten  
Molekülstrukturdaten in niedrigaufgelöste  
Elektronenmikroskopie-Dichteverteilungen**

Stefan Birmanns

NIC-Serie Band 18

ISBN 3-00-010004-0, September 2003, 178 Seiten

**Auswirkungen der Virtualisierung auf den IT-Betrieb**

Wolfgang Gürich (Hrsg.)

GI Conference, 4. - 5. November 2003, Forschungszentrum Jülich

NIC-Serie Band 19

ISBN 3-00-009100-9, Oktober 2003, 126 Seiten

**NIC Symposium 2004)**

Dietrich Wolf, Gernot Münster, Manfred Kremer (Hrsg.)

Symposium, 17. - 18. Februar 2004, Forschungszentrum Jülich

NIC-Serie Band 20

ISBN 3-00-012372-5, Februar 2004, 482 Seiten

**Measuring Synchronization in Model Systems and  
Electroencephalographic Time Series from Epilepsy Patients**

Thomas Kreuz

NIC-Serie Band 21

ISBN 3-00-012373-3, Februar 2004, 138 Seiten

**Computational Soft Matter: From Synthetic Polymers to Proteins -  
Poster Abstracts**

Norbert Attig, Kurt Binder, Helmut Grubmüller, Kurt Kremer (Hrsg.)

Winterschule, 29. Februar - 6. März 2004, Gustav-Stresemann-Institut Bonn

NIC-Serie Band 22

ISBN 3-00-012374-1, Februar 2004, 120 Seiten

**Computational Soft Matter: From Synthetic Polymers to Proteins -  
Lecture Notes**

Norbert Attig, Kurt Binder, Helmut Grubmüller, Kurt Kremer (Hrsg.)

Winterschule, 29. Februar - 6. März 2004, Gustav-Stresemann-Institut Bonn

NIC-Serie Band 23

ISBN 3-00-012641-4, Februar 2004, 440 Seiten

**Synchronization and Interdependence Measures and their Applications  
to the Electroencephalogram of Epilepsy Patients and Clustering of Data**

Alexander Kraskov

NIC-Serie Band 24

ISBN 3-00-013619-3, Mai 2004, 106 Seiten

**High Performance Computing in Chemistry**

Johannes Grotendorst (Hrsg.)

Bericht des Verbundprojekts:

High Performance Computing in Chemistry - HPC-Chem

NIC-Serie Band 25

ISBN 3-00-013618-5, Dezember 2004, 160 Seiten

Alle Bände stehen online zur Verfügung unter <http://www.fz-juelich.de/nic-series/>.