Statistical analysis of biomedical data

Dissertation

zur Erlangung des Doktorgrades
der Naturwissenschaften (Dr. rer. nat.)
der Naturwissenschaftlichen Fakultät II – Physik
der Universität Regensburg

vorgelegt von

Andreas Jung

aus München

Dezember 2003

Das Promotionsgesuch wurde am 17. Dezember 2003 eingereicht.
Das Promotionskolloquium fand am 29. Januar 2004 statt.

Prüfungsausschuss:

Vorsitzender:	Prof. Dr. Werner Wegscheider
1. Gutachter:	Prof. Dr. Klaus Richter
2. Gutachter:	Prof. Dr. Gustav Obermair
Weiterer Prüfer:	Prof. Dr. Matthias Brack

Meinen Eltern

Contents

Glossary

ABP	Arterial Blood Pressure
BSS	Blind Source Separation
CbO_2	Oxygen content in the blood
$C(a\|v)O_2$	Oxygen content in the (arterial|venous) blood
cdf	cumulative distribution function
CSF	Cerebrospinal Fluid
CT	Computer Tomography
DFT	Discrete Fourier Transform
ECG	Electro-Cardiography
EEG	Electro-Encephalography
Hb	Deoxyhaemoglobin
HbO_2	Oxyhaemoglobin
IC	Independent Component
ICA	Independent Component Analysis
ICP	Intracranial Pressure
Invos	In-Vivo Optical Spectroscopy
Licox	Liquor Oxygenation
MABP	Mean Arterial Blood Pressure
MTM	Multi Taper Method
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
$\bar{p}_{ti}$	mean partial oxygen pressure in the tissue
pbO_2	partial oxygen pressure in the blood
$p(a\|v)O_2$	partial oxygen pressure in the (arterial|venous) blood
pdf	probability density function
RMT	Random Matrix Theory
SbO_2	Saturation of the blood with oxygen
$S(a\|v)O_2$	Saturation of the (arterial|venous) blood with oxygen
SNR	Signal to Noise Ratio
STFT	Short Time Fourier Transform
TSA	Time Series Analysis

Introduction

Recently, the development of computer applications in the field of life sciences, in particular for the clinical and biomedical environments, has gained increasing attention due to the promising results in the treatment of patients.

Today, the recording of the patient's status in multivariate data sets is a standard procedure in everydays clinical life. Apart from the immediate evaluation of the data by the physicians, the (off-line) analysis by dedicated computer processing tools can be a valuable information source.

In the context of data analysis, the methods derived in different fields of *physics* raise the question of applying these well known methods to life sciences, in particular to *biomedical data analysis*: Is it possible to obtain a deeper understanding of the data with new (non)linear methods and physical models describing the biological system? Can we further improve the analysis to reveal the hidden information in biomedical data by developing new algorithms overcoming limitations of the existing methods?

At the university hospital in Regensburg, the department of neurosurgery records different biomedical data sets in the clinical environment. These recordings have motivated this work and we will focus on two of these data sets. On one hand, *neuromonitoring* data is recorded on the intensive care unit from patients with severe head injuries. These data sets reflect mainly the following brain status parameters: oxygen content in the blood and tissue of the brain, the arterial blood pressure and the internal brain pressure. In addition, the patient's brain temperature is measured. On the other hand, the neural brain activity of patients is recorded in the *electro-encephalography* (EEG), in particular in the post-operative treatment, to monitor neurological diseases. These recordings represent – in contrast to the neuromonitoring data – highly multivariate data sets with 21 or more signals.

The questions arising in the analysis of the data are very diverse. They all focus on a deeper understanding of the mechanisms of the investigated system, namely the human brain. In neuromonitoring data we may ask: Do the signals influence each other, are they correlated in some sense? Which processes trigger the system? What is the underlying biological system generating the signals? In the analysis of the EEG data we are mainly interested in whether new methods can reveal more information or enhance the highly multivariate data.

The techniques typically used in data analysis can be divided into three different categories. This depends on the amount of knowledge available about the investigated system:

- Time Series Analysis:
  No knowledge is available about the system, only the time series originating from the outputs of the system can be analysed. Two main techniques are used: either unimodal (fourier transforms, wavelets, (non)linear analysis using time embedding) or bimodal analysis methods (correlations and couplings using nonlinear statistics). Symptoms or phenomena detectable in the time series can be quantified in such a way that statistical tests can be applied.

- Independent Component Analysis (ICA):
  Hidden sources are underlying the measured signals. ICA is in particular useful for the analysis of highly multivariate data sets ($\gg 2$ signals). Using the concept of ICA, the recorded signals can be described by a (non)linear mixing of unknown statistically independent sources.

- Model:
  The basic mechanisms describing the system are known. Based on this knowledge, the behaviour of the system can be modeled, however some free parameters have to be fitted by comparing the simulations to the measured signals. Properties of the system can be analysed by a sensibility and stability analysis including their parameter dependence. The reaction of the system with respect to parameter drifts, external influences etc. can be extrapolated to make possible "predictions".

With the methods listed above, using complementary approaches with very different assumptions on the knowledge about the system, we will try to reveal the mechanisms generating the recorded signals and show how, for the various data sets, different methods have to be applied.

First, the reader will be made familiar with the basic anatomy of the human head including a description of the fluid dynamical and metabolic (oxygen supply) processes. Furthermore, a detailed description of the data sets and the basic physical principles of the measurement methods will be presented. Finally, the origin of the neural activity, their recording on the scalp with electro-encephalography (EEG) and their typical waveforms will be described in the first chapter.

The second chapter focuses on the analysis of the neuromonitoring data using time series analysis. To gain basic information about possible interconnections between the time series, a correlation analysis in the *frequency domain* will be used, to ensure that all processes are treated equally, independent of their origin and their frequency band. Beside the well established correlations between

the oxygen supply processes, more surprising couplings are found. In particular connections between the arterial blood pressure, as a hydrodynamical process, and the oxygen content in the blood and brain tissue, as a metabolic process, are seen. First explanations of these observed correlations can be given with some intuitive descriptions using basic knowledge about the biological mechanisms in the human brain.

In a more global context in chapter three, the results of the time series analysis can be understood in more detail by using biomedical knowledge about the human brain. Based on this knowledge, we develop a model combining the fluid dynamical processes with a description of the oxygen supply in the human brain tissue. Thereby, the amount of blood flowing through the brain, the so-called cerebral blood flow, represents the main connection between the hydrodynamical and oxygen supply model. One advantage of the model is that this parameter, which is in practice hardly measureable, can be numerically estimated and therefore provides the physicians with important information about the patient's state. Furthermore, this *combined model* allows to calculate all other parameters measured in neuromonitoring, in particular the values describing the oxygen supply processes. Based on a pure physical description, such a combined model enables us to describe and understand in a more detailed way the couplings identified in the time series analysis.

Furthermore, the influence of variations of the model parameters, i.e. the state of health of the patient, on the evolution of the recorded brain status parameters can now be described and can provide suggestions for the treatment of patients in various situations. Interestingly, the relationship between some of the parameters obtained from the model show a surprisingly well established *linear* correlation, although the physical processes are highly complex and *nonlinear*. This observation explains, why the linear correlation analysis of the time series produced such good results.

Before the investigation of the applicability of the independent component analysis (ICA) to biomedical data, an elaborate theoretical presentation of the method will be given in chapter four. Two new algorithms will be presented. First an intuitive approach where geometrical consideration of the transformation of scatter plots will be used to develop a theoretical framework. From these considerations, a fast histogram based algorithm (FastGeo) is derived. The second algorithm (FastTeICA) shows how time structures in the data can be included in the framework of ICA by using methods known in time series analysis – in particular by using independent time embedding vectors.

Finally, we present the results of the application of ICA to two biomedical data sets, namely the neuromonitoring data and the data recorded by electroencephalography (EEG). The examples demonstrate both the power and the weakpoints of the concept based on ICA. In particular the interpretation of the extracted independent components have to be considered in real world applications from a biomedical point of view.

In the last chapter, the conclusions of the preceding chapters will be summarised and an outlook to further developments in the field of biomedical data analysis will be given.

This work was conducted in close cooperation with the department of neurosurgery of the university hospital in Regensburg.

Chapter 1

Survey of the biomedical data sets

At the neurosurgery department of the university hospital in Regensburg different data sets are recorded as a standard procedure for immediate evaluation and later off-line analysis. In this chapter a survey of these recorded data sets and the measurement methods is presented.

In the following work two kinds of data sets are of primary interest: first the neuromonitoring data recorded at the intensive care unit from patients with severe head injuries and second the electro-encephalography (EEG) data used to monitor neurological diseases. Both data sets consist of multi-channel recordings often referred to as multivariate measurements. Before going into a detailed description of the data sets and the measurement methods, an introductory overview of the anatomy and physiology of the human brain is given.

1.1 Anatomy and physiology of the human brain

As an introduction to this chapter, a survey of the anatomy and physiology of the human head will be given. Two schematic illustrations in figure 1.1 visualise the two main fluid circulations in the head, the cerebral blood flow (CBF) and the cerebrospinal fluid (CSF) flow. A further illustration in figure 1.4 shows a cross-section of the human head obtained by a nuclear magnetic resonance recording of the authors brain.

The cranial bone which acts as a closed compartment accommodates the human brain. Between the cranial bone and the brain tissue, a fluid, the so-called cerebrospinal fluid, acts as a protection against external shocks.

The brain itself can be divided into two horizontal layers. The surface of the brain is composed of nerve cell rich tissue which is called the gray brain matter (due to its colour) where the main neural activity of the brain takes place. To connect the nerve cells with each other throughout the brain, the subjacent tissue

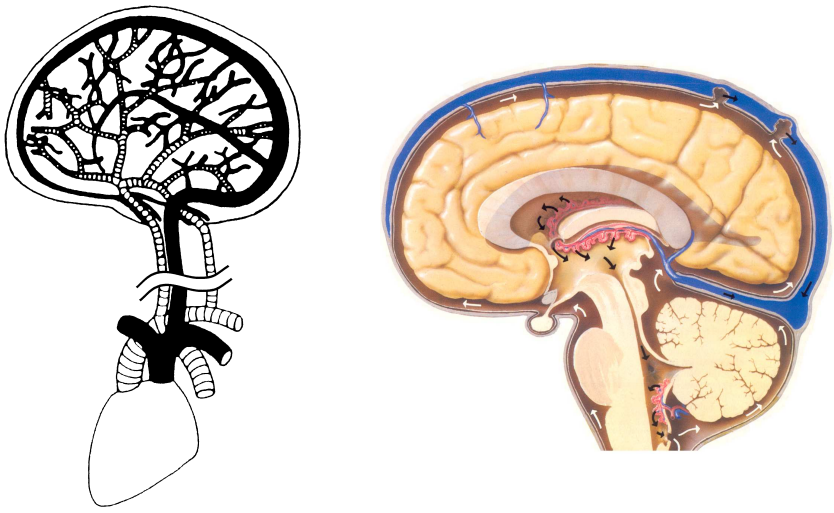

Figure 1.1: Simplified figure of the blood supply and cerebrospinal fluid flow in the brain. *Left:* Two main arterial blood vessels (shown as white striped vessels) supply the brain with oxygenated blood. The vessels branch into smaller ones reaching the capillary level. Deoxygenated blood flows from the capillaries through the veins (black vessels) to the central vein (sagittal sinus) leading back to the heart. *Right:* The cerebrospinal fluid (CSF) is surrounding the brain tissue. The production of CSF takes place at the ventricles (centre of the brain). The CSF flows around the brain and gets absorbed at the sagittal sinus, the big venous blood vessel at the top of the brain.

(the white matter) consists primarily of connecting paths or the so-called axons of the nerve cells. In figure 1.4 the different layers of the brain tissue can be associated by the different gray values. Note that due to the representation of the recorded image, the gray brain matter is lighter than the white matter.

Monro and Kellie already formulated their doctrine in the 18th century stating the volume of the head is fixed by the cranial bone and the only exchange of fluids is due to the blood flow from and to the heart.

The cerebral blood flow is illustrated in the left drawing of figure 1.1. All oxygen necessary for the brain as well as for other tissues in the body is transported through the blood. In the lungs the blood exchanges gases with the air, CO_2 from the cells is diffusing into the air and O_2 is absorbed from it into the blood. Since the amount of dissolved oxygen in the blood would never satisfy the demands from the cells, the blood has a further oxygen buffer, the haemoglobin, which is responsable for the red colour of blood. One O_2 molecule can bound to each of the four iron (Fe) sites on the haemoglobin molecule. 98% of the oxygen in human blood is bound to the haemoglobin whereas only 2% are dissolved. Therefore the haemoglobin can be seen as an oxygen buffer.

The nearly 100% saturated (haemoglobin) blood flows from the lungs through

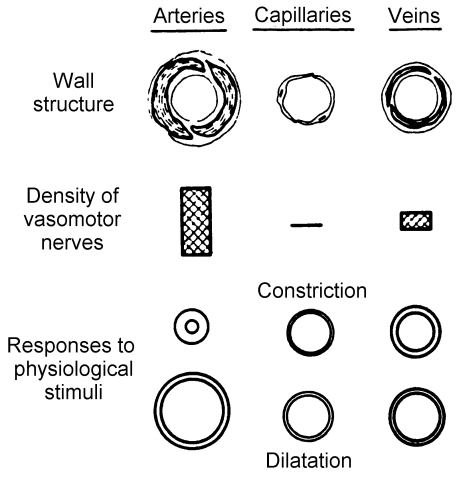

Figure 1.2: Scheme of the behaviour of different blood vessels (arteries, capillaries and veins) related to their wall structure. The comparatively thick muscular layer and high density of vasomotor nerves account for considerable constriction and dilatation ability of cerebral arteries. These responses are less pronounced in the veins (especially in the cerebral veins, which are devoid of a continuous muscles layer), and absent in the capillaries (adapted from Mchedlishvili [1986]).

the arteries to the brain. The blood vessels branch into smaller and smaller vessels until they reach the capillary level with diameters of only a few μm. Arteries are active elements distinguishing themselves from the other blood vessels by having a muscular layer around the vessel (for an illustration of the different blood vessels see figure 1.2). This muscular layer comprises the brains possibility to regulate the cerebral blood flow by dilating or constricting the arteries. The mechanism of keeping the CBF constant, the cerebral autoregulation, guarantees a steady supply of oxygen over a wide range of the arterial blood pressure.

In contrast to the arteries, the capillaries are thin-walled vessels with typically diameters of 7 μm. The ratio of the surface of a capillary to the volume of the contained blood is much higher at this level than at any other point in the blood vessel arrangement. This enables the dissolved oxygen as well as other metabolic products like glucose to diffuse easily into the surrounding tissue. At the capillary level the main exchange of the metabolic products happens. But waste products like CO_2 can also diffuse from brain tissue back into the capillaries. The CO_2 molecules can then be bind to the free Fe sites on the haemoglobin so that it can transport the CO_2 back to the lungs where it is again exchanged with a new O_2 molecule. All diffusion processes, as for example between the blood vessels (capillaries) and the brain tissue or the air and the capillaries in the lungs, are based on the pressure difference of the dissolved oxygen. Therefore the oxygen diffusion processes in the human body are a pure pressure depending processes.

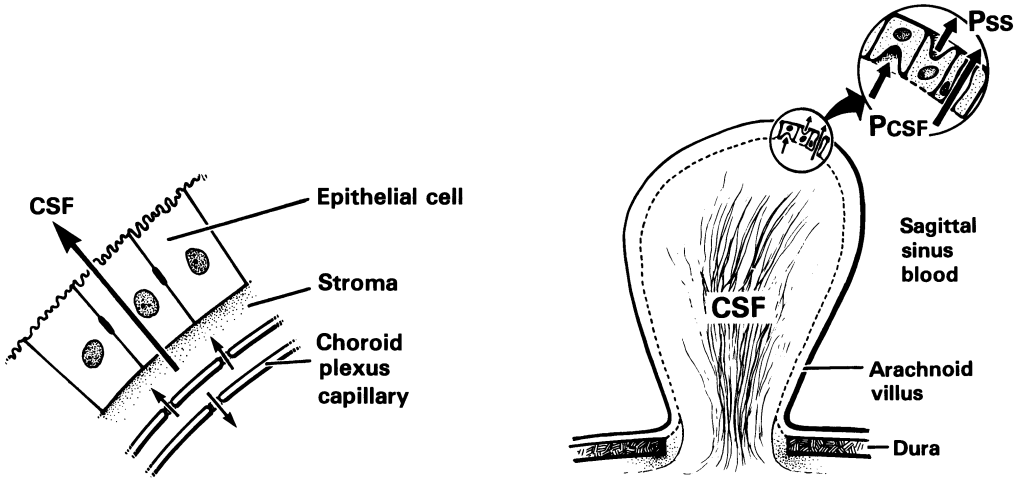

Figure 1.3: Production and absorption of cerebrospinal fluid (CSF). *Left:* The CSF is produced at the capillary level in the ventricles. The driving force is the pressure difference between the blood in the capillaries and the CSF compartment. *Right:* The absorption of the CSF occurs at the sagittal sinus by the difference between the pressure in the CSF and the sagittal sinus (figures from Sullivan and Allison [1985]).

Finally, the veins are the blood vessels transporting the deoxygenated blood back to the heart, taking mainly the CO_2 but also other waste products from the cells. The veins finally reach the sagittal sinus, a big hard-walled venous blood vessel in the top of the head, which is the main path leading back to the heart.

The fact, that 600 ml blood/min (i.e. 30% of the total amount of blood flowing through the body) is transported through the human brain, gives a rough impression of the importance of the cerebral blood flow in the brain.

As mentioned before, the brain is not directly touching the cranial bone, but is surrounded by the so-called cerebrospinal fluid (CSF), which is produced at the capillary level, the ventricles, in the middle of the brain (see right illustration of figure 1.1). The fluid, which consists of around 170 ml mainly pure water, protects the brain tissue from external shocks. The absorption of the CSF takes place at the sagittal sinus. To get an impression of the microbiological arrangement at the two locations, a schematic illustration is shown in figure 1.3.

Around 600 ml of CSF are produced and absorbed per day which corresponds to an exchange of the whole fluid more than three times a day. While the dynamics of the CSF circulation reaches just a 1/1500 of the blood circulation, it plays a crucial role in volume compensation during a brain swelling as will be seen later in description of the hydrodynamical model in chapter 3.

1.2 Neuromonitoring

Throughout the treatment of a patient with a severe head injury at an intensive care unit, different important physiological parameters are continuously monitored. A distinction is drawn between two different mechanisms which can be monitored by sensing appropriate brain status parameters: first the cerebral metabolism, i.e. oxygen supply and consumption in the brain tissue, and second the cerebral haemodynamics, i.e. the flow and the pressures of the cerebral fluids.

Primarily the oxygen supply of the brain is the most important factor for medical observation due to the fact that many of the patients die because of oxygen undersupply. Therefore a constant monitoring of the patient is necessary. This monitoring or recording of the data is called "neuromonitoring".

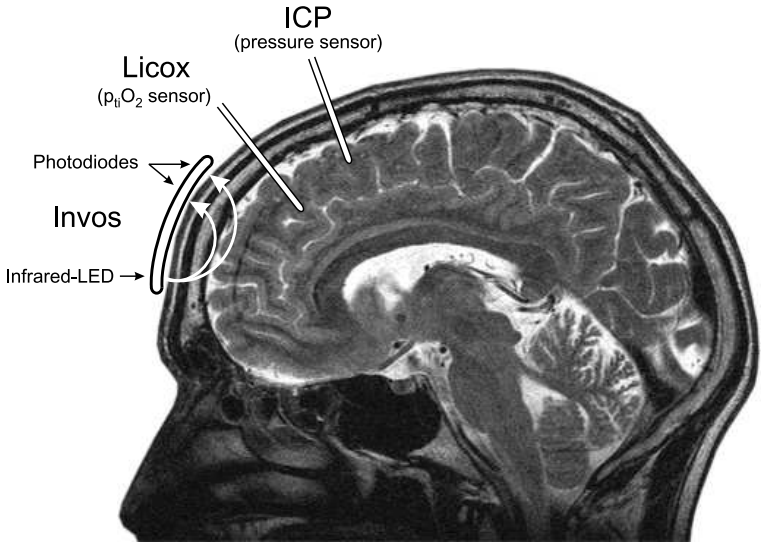

Figure 1.4: Location of the neuromonitoring sensors – all sensors are placed at the front part of the brain. i) Licox: the sensor, measuring the partial oxygen pressure in the brain tissue ($p_{ti}O_2$) as well as the temperature, is placed 2.7 cm deep into the brain matter. ii) Invos sensor: placed on the surface of the scalp measuring the oxygen saturation of blood. The distance of the photodiodes to the infrared-LEDs is 3 and 4 cm. iii) ICP-sensor: localised close to the surface (1–2 cm depth) and measuring the intracranial pressure (internal brain pressure). Note: The image is an NMR measurement of the authors brain :).

In figure 1.4 the usual positioning of the relevant neuromonitoring sensors is shown and will be described in the following:

- Cerebral metabolism – Two sensors are used for the measurement of the oxygen supply and consumption:

Licox (Liquor Oxygenation) is a brain tissue oxygenation monitoring system. This measurement method is a standard method in clinical applications to monitor the supply of the brain tissue with oxygen. Due to the temperature dependence of the measurement method the result need to be corrected by an integrated temperature sensor for a more accurate reading. Therefore the temperature values of the patients brain are also available.

Invos (In-Vivo Optical Spectroscopy) is a relatively new device to monitor the oxygen supply in-vivo (i.e. in the living organism) without the need of placing a sensor into the patients brain tissue. The sensor is placed on the front part of the scalp using infrared light to measure the oxygen content of the blood (more precisely the degree of saturation of haemoglobin with oxygen).

- Cerebral haemodynamics – Mainly two parameters are monitored to get more information about the flow and pressure of the cerebral fluids:

  ICP (Intracranial Pressure) depicts the pressure inside the cranial bone. It is one of the most basic and important parameters recorded in standard neuromonitoring, since values above 40 mmHg are perilous. Long time measurements are done by a small piezo-electric sensor placed into the brain tissue.

  ABP (Arterial Blood Pressure) depicts the pressure of the artery blood, measured in the central arterial of the patient. Typically the *mean* arterial blood pressure (MABP) is given in the recordings with typical values in the range of 80 to 120 mmHg.

In the following sections, the basic physical principles of the measurement methods will be described in more detail.

Licox including temperature measurement

The Licox sensor is placed into the brain tissue to measure the amount of available oxygen for the nerve cells. The cells receive the oxygen by diffusion, this means, that the partial oxygen pressure in the tissue is proportional to the available oxygen for the nerve cells. Measuring this value gives an indication of the oxygen supply of the brain tissue.

In figure 1.5 the basic configuration of a Licox sensor is shown. The sensor is based on a polarographic Clark-type probe which is injected into the frontal part of the brain. Oxygen diffuses from the tissue through the polyethylene tube diffusion membrane into the inner electrolyte chamber filled with KCl. Between the polarographic silver anode and gold cathode a current flows due to the applied voltage of a 800 mV. The measured current is proportional to the oxygen concentration in the electrolytic chamber. Furthermore it is also proportional to the oxygen concentration in the surrounding tissue due to the diffusion membrane.

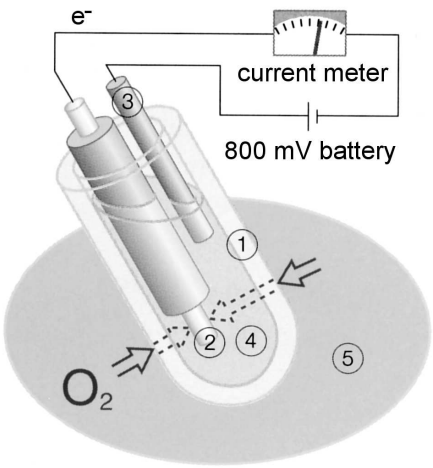

Figure 1.5: Licox sensor: The basic design of the Revoxode $p_{ti}O_2$ probe is a polarographic Clark-type cell. (1)=polyethylene diffusion membrane (diameter of sensor is less than 1 mm), (2)=gold cathode, (3)=silver anode, (4)=electrolytic chamber (KCl) and (5)=brain tissue (figure adapted from the Gesellschaft für Medizinische Sondentechnik mbH).

The chemical reactions at the silver and gold cathode – placed into a KCl electrolyte – are

- an oxidation at the gold cathode: $O_2 + 2\,H_2O + 4e^- \rightarrow 4\,OH^-$
- a reduction at the silver anode: $4\,Ag + 4\,Cl^- - 4e^- \rightarrow 4\,AgCl.$

As seen above, the chemical reaction with one O_2 molecule results in a flow of four electrons through the circuit, i.e. the flowing current is proportional to the oxygen concentration.

Furthermore a temperature compensation is necessary for the Licox sensor, since the permeability of the membrane increases with temperature, permitting more oxygen to enter the probe. The temperature measurement is done by an additional sensor in the probe and digitally processed in the Licox device. Therefore the temperature values of the patients brain are also available.

In a typical Licox sensor, the surface of the diffusive membrane is about 14 mm^2 in size. Assuming a capillary density of around 300–1000 capillaries/mm^2, the surface of the sensor touches hundreds of capillaries measuring therefore an average partial oxygen pressure. Still, compared to the size of the human brain, the sensor is measuring only a local oxygen concentration.

Typical measurement values of the Licox sensor are in the range of 10 mmHg to 40 mmHg. Under special conditions (supply of the patient with 100% oxygen) the sensor measurements can reach values above 100 mmHg. Values below 10 mmHg are in general described as perilous.

Invos

The Invos sensor measures the percentage of oxygen saturated haemoglobin in the blood. This type of measurement is called "oximetry".

It was discovered in the 1860's that the coloured substance in blood, the haemoglobin, is also its carrier of oxygen. Haemoglobin is a protein which is bound to the red blood cells and can bind oxygen to its 4 Fe (iron) sites. At the same time, it was noticed that two common forms of the molecule, oxidised haemoglobin (HbO_2) and reduced haemoglobin (Hb), have significantly different optical spectra in the wavelength range from 600 nm to 1000 nm (see figure 1.6). The oxygen chemically combined with haemoglobin inside the red blood cells makes up nearly all of the oxygen present in the blood, only a small amount of around 2% is dissolved in the plasma.

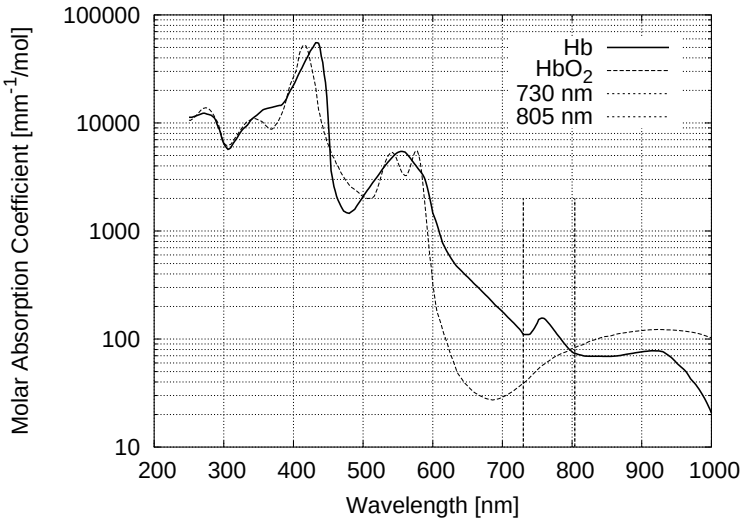

Figure 1.6: Absorption spectra of haemoglobin in deoxygenated (Hb) and oxygenated (HbO_2) state. The Invos sensor is using two different wavelengths (730 and 805 nm) to distinguish between these two states and to calculate their ratio. At 805 nm the absorption of Hb and HbO_2 is nearly equal in contrast to the absorption at 730 nm.

The functional principle of the Invos sensor is based on the difference in the optical spectra of Hb (deoxyhaemoglobin) and HbO_2 (oxyhaemoglobin), by sending light of two different wavelengths through the tissue and measuring their absorption. The tissue itself is relatively transparent in the range between 650 and 1100 nm. Therefore the wavelengths $\lambda_1 = 730\,\text{nm}$ and $\lambda_2 = 805\,\text{nm}$ are chosen for the oxymetric measurements by the Invos sensor. With the Beer-Lambert law ($I = I_0\, e^{-\alpha c l}$) we obtain two equations for the intensities at the two wavelength

λ_1 and λ_2:

$$\ln(\frac{I(\lambda)}{I_0(\lambda)}) = -\alpha_{\mathrm{Hb}}(\lambda) \cdot c_{\mathrm{Hb}} \cdot l - \alpha_{\mathrm{HbO_2}}(\lambda) \cdot c_{\mathrm{HbO_2}} \cdot l - t, \quad \text{with } \lambda = \lambda_1, \lambda_2 \tag{1.1}$$

where α is the absorption coefficient of Hb and HbO_2, c their unknown concentrations, l the length of the light path through the tissue and $t = \alpha_t \cdot c_t \cdot l$ the absorption of the tissue which does not depend on the wavelength and oxygen content, but is also unknown. To actually determine the concentration ratio between Hb and HbO_2, either a further measurement at a wavelength λ_3 is needed or the mean absorption of the tissue t is directly determined in a separate experiment. Ignoring absorption and scattering in the tissue would lead to wrong values, since 83% of the light is attenuated by tissue as shown by Kauper [1999].

The saturation of HbO_2 is then given by the ratio of the concentration of HbO_2 to the sum of the concentrations of haemoglobin

$$\mathrm{HbO_2\ saturation} = \frac{c_{\mathrm{HbO_2}}}{c_{\mathrm{Hb}} + c_{\mathrm{HbO_2}}} \tag{1.2}$$

which takes typically values in the range between 60% and 80%. However, Metz [2001] emphasised with his knowledge from clinical studies that rather the variation of the values of the Invos sensor can indicate an oxygen undersupply of the brain, a so-called ischemia, than the absolute value. Therefore in the analysis of the Invos data we will focus mainly on its dynamical aspects.

Furthermore, the use of two photodiodes with different distances from the light emitting diodes (LED) can help to remove the surface effects and to calculate only the HbO_2 saturation in the deeper areas of the brain. Calculations on the light paths of the infrared light through the scalp, the cerebrospinal fluid and the brain tissue were done by Okada et al. [1997]. It was shown, that for smaller distances between infrared LED and photodiode the light travels more through the scalp and the cerebrospinal fluid, while for distances above 4 cm, the photons travel more through the brain tissue. When "subtracting" the two measurements from each other, primarily the contribution from the deeper brain tissue is obtained. Using these results, a good balance between the removal of surface effects and a too strong absorption is achieved for a separation of 3 and 4 cm between the infrared LED and photodiode, as used by the Invos sensor.

The actual value given by this sensor will reflect therefore the local averaged saturation of the cerebral blood with oxygen. Generally it is assumed, that the cerebral tissue contains approximately 75% of venous and only 25% of arterial blood, as was also stated in a publication by Kim et al. [2000]. Therefore the HbO_2 saturation measured – also abbreviated by SbO_2 – is given by

$$\mathrm{SbO_2} = 0.25 \cdot \mathrm{SaO_2} + 0.75 \cdot \mathrm{SvO_2} \tag{1.3}$$

where SaO_2 respectively SvO_2 is the HbO_2 saturation in the arterial and venous blood.

Finally, a discussion about the relation between the Licox and the Invos sensors and their measured values will be given in great detail in chapter 3.3.

Intracranial pressure and arterial blood pressure

The intracranial pressure (ICP) is measured in the brain tissue using the piezoresistive effect. A membrane attached to the piezo element is exposed on one side to the brain tissue, on the other side to the (external) air pressure. The pressure difference measured at the element gives the intracranial pressure (ICP) value. Typical values range between 5 mmHg and 20 mmHg. Every value above 40 mmHg is perilous.

Equivalent methods are used for the measurement of the arterial blood pressure. The sensor is normally placed in the central artery of the patient. For the analysis, only the mean long time behaviour of the arterial blood pressure is of interest. Therefore only the time average is given in the recordings, which is abbreviated by MABP (*mean* arterial blood pressure) or often just by ABP. Typical MABP values are in the range between 80 to 120 mmHg, the true high resolution curve of the blood pressure pulse can of course take values between 60 and 200 mmHg.

An example of neuromonitoring data

A typical example of neuromonitoring data recorded at the intensive care unit of the department of neurosurgery (university hospital Regensburg) is shown in figure 1.7. The data presented are the unfiltered (raw) time series from a 30 hours recording. In principle, the recordings of the patients parameters can be taken over any length in time with a sampling interval of up to one second. Our longest continuous measurement is a 6 days recording.

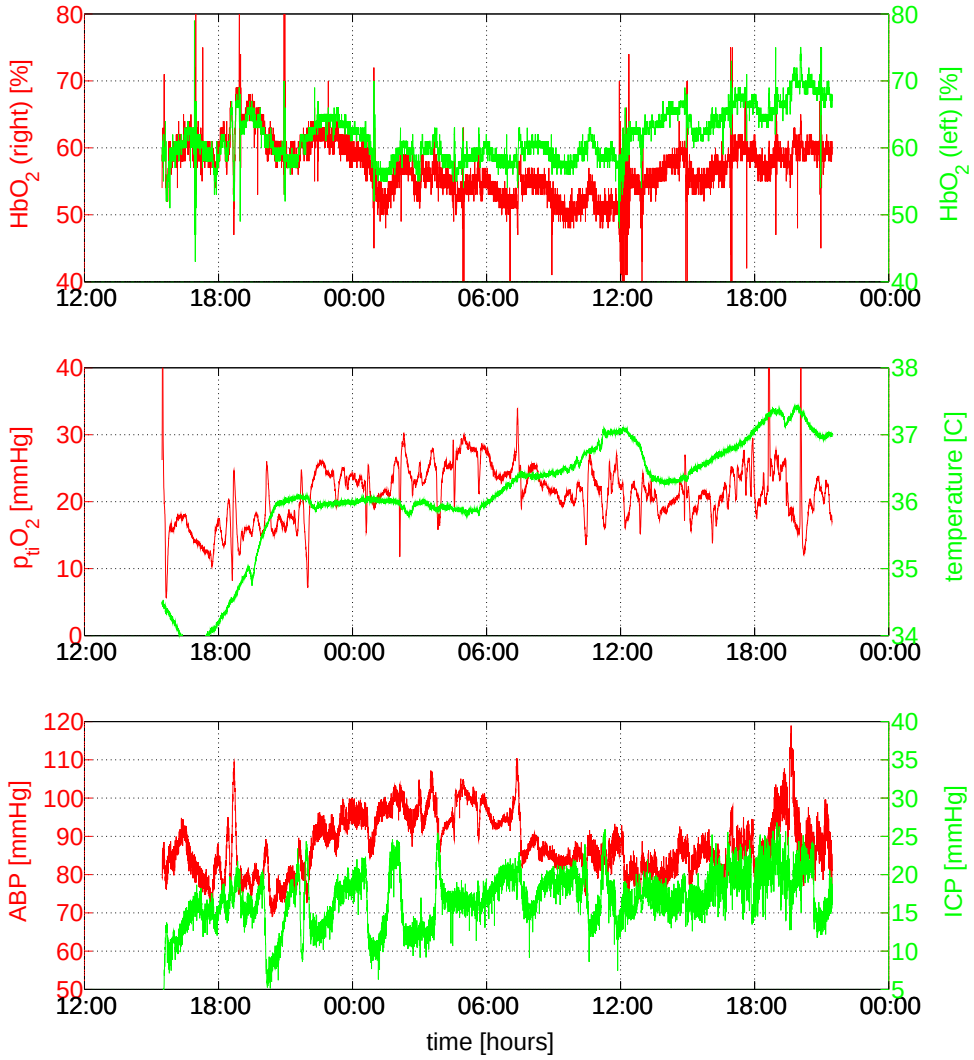

Figure 1.7: Recording of neuromonitoring data from the intensive care unit at the department of neurosurgery (University hospital Regensburg). The two upper plots show the parameters of the cerebral metabolism: the saturation of the haemoglobin with oxygen (HbO_2) in the left and right hemisphere of the brain, the partial oxygen pressure in the brain tissue ($p_{ti}O_2$) and the body temperature of the patient. The lower plot shows the haemodynamic parameters: the mean arterial blood pressure (ABP) and the intracranial pressure (ICP). A section of only 30 hours of unfiltered data (sampling every 15 seconds) is shown but recordings of up to 144 hours = 6 days are also available.

1.3 Electro-Encephalography (EEG)

The recording of electrical patterns at the surface of the scalp which primarily reflect cortical electrical activity or "brainwaves" is called electro-encephalography and abbreviated by EEG[1]. In 1929, Hans Berger was the first who discovered that tiny rhythms of electrical wave activity could be detected at the human scalp.

In the following years typical potentials and rhythms of patients suffering from epilepsy were categorised. In 1945 the first multi-channel recording with four channels has been achieved. Quickly the electronic equipment developed to a standard 21 channel recording, called the Ten-Twenty-System. It was introduced as a standardised layout for EEG measurements in 1958 (see figure 1.8). Today up to 256 channels are used for scientific research.

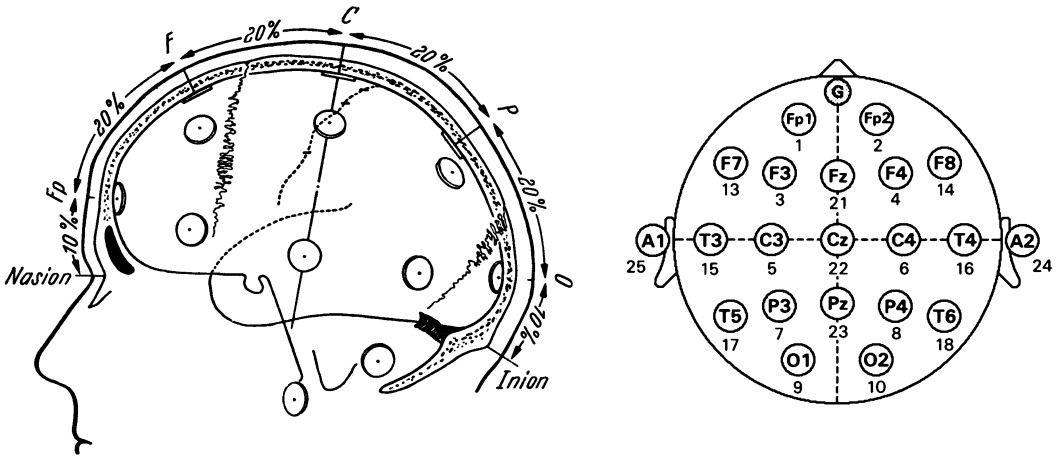

Figure 1.8: Location of the electrodes in the ten-twenty-system (21 channels) which is the international standard for EEG measurements in clinical application since 1958 (*left*=side-view and *right*=top-view of the patients head). The positions are chosen to cover the whole scalp and make a location of a potential anomaly as good as possible (adapted from Kugler [1981] and Zschocke [1995]).

The EEG equipment consists of small, non-invasive electrodes which are placed carefully with paste or a glue-like substance on a patient's scalp. Low voltage signals (5–500 microvolts) are amplified and recorded with sampling rates usually around 166 Hz by the EEG equipment. The measured potentials originate from the nerve cells which generate electric potentials by chemical processes. Due to the huge number of nerve cells in the cortex of the brain, the measured signal is the sum of a huge collection of nerve cells.

The recordings are mostly taken with the eyes closed to get an undisturbed EEG, although the patient is sometimes asked to open them for short periods.

[1]Electro-encephalograph: electro=electrical; encephalon=head; graph=drawing/picture

A typical recording time with eyes closed would be about 5 minutes. For an illustration, a sample of 6 seconds is shown in figure 1.9. The evolution of the electric potentials at the electrodes over time is shown for all 21 channels. By visual inspection the neurosurgeon can identify four clinically relevant spectral bands: the delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), and beta (above 12 Hz) waves. A very dominant signal in closed eyes condition is the alpha wave activity, which is easily identifyable also in the example shown here.

Today the EEG measurements are used in clinical applications to monitor neurological diseases. The method turns out to be often more sensitive for epilepsy and tumour evolution studies than computer tomography (CT) and nuclear magnetic resonance (NMR) imaging. At the university hospital in Regensburg in particular the tumour evolution after a neurosurgical operation is of interest. Pathological regions show typically an increase of slow activity (delta, theta waves) and diminishing fast activity (alpha, beta waves).

In general, EEG interpretation requires considerable skill and often years of clinical experience due to the complex structure of the signals. Therefore an automatic detection and removal of artifacts could enhance the interpretation of an EEG and the identification of potential neurological diseases. An example of a typical artifact can be seen in channel "Fp2" in figure 1.9 which corresponds to an eye movement or an eye blink.

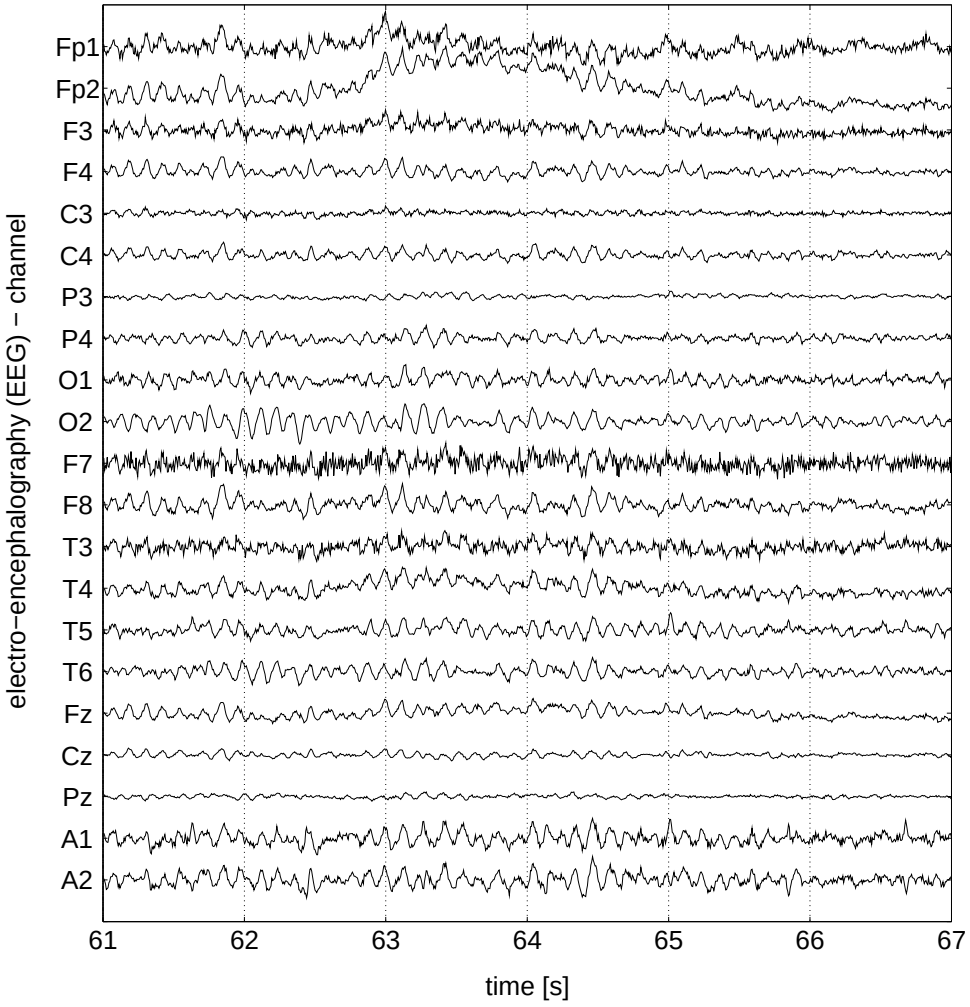

Figure 1.9: An electro-encephalography (EEG) measurement from a patient recorded at the department of neurosurgery (University hospital Regensburg). The EEG-channels are labelled by a abbreviation which corresponds to a given electrode on the head of the patient (see figure 1.8). The plotted lines show the evolution of the electric potentials at each electrode over time. One clear artifact (an eye-blink) is visible in channel "Fp2", furthermore the alpha-wave activity of the brain ($\sim 8\,\mathrm{Hz}$) is visible in nearly all channels.

Chapter 2

Time series analysis

Typically, when applying time series analysis, only little knowledge is available about the system to be investigated. The only information source are the "outputs", i.e. the time series recorded from the system. All information is gained by the analysis of these outputs.

In this chapter we will use time series analysis to gain more information about the interrelation between the different time series recorded in neuromonitoring. It will be shown that in particular the correlation in the frequency domain contains valuable information which leads to a better insight into intracranial dynamics.

2.1 Introduction

Time series analysis is typically used for uni- or bimodal signals, that means analysing a system just by using one respectively two time series of the system. Neuromonitoring data is in contrast to EEG data a good candidate for the time series analysis. In particular the coupling between the time series of the neuromonitoring data is of interest and can be treated by the so-called correlation analysis. In the following a short overview over the two main techniques in time series analysis is given.

Nonlinear unimodal time series analysis is mainly based on a method called time-embedding as described by Kantz and Schreiber [1997]. It can be shown that by analysing time delayed samples of one time series and plotting them in $(2n+1)$ dimensions, it is possible to unfold the attractor of a n-dimensional system. Such an analysis was also performed on the data of the neuromonitoring recordings, but it yielded no results. This was mainly due to two reasons. Firstly the high noise level and the short length of the recorded data, which makes it hard to unfold the attractor of the system – as was also shown in the diploma thesis by Meier [2000] – and secondly the nonstationarity of the underlying system. The system parameter in the background can change on the same time scale as the variations of the system itself. Such conditions make the application of unimodal

time series analysis nearly impossible.

Bimodal time series analysis concentrates more on the correlation or coupling between two time series. The question to be answered is: Which signals have common information contents and do they influence each other? Correlation analysis and higher order statistics like mutual information and conditional entropy, e.g. transfer entropy, can help to reveal such interconnections.

The problem in neuromonitoring focuses in particular on the signals triggering the system. To solve this question, the transfer entropy could be applied, since it can find the direction and coupling strength between two processes as described by Kaiser and Schreiber [2002]. But the high noise level and the nonstationarity of the data makes it nearly impossible to apply such methods. Further the time synchronisation between the time series of the neuromonitoring data is not guaranteed. The measurement devices for the metabolic brain status parameters use different non-synchronised clocks. Due to this fact, an absolute time stamp was not available and an estimation of coupling directions of the biological or physical processes is therefore difficult. This problem will be solved in the future by using radio controlled synchronised clocks.

In the following analysis we will therefore only use the correlation analysis, since it works more stable with noisy and nonstationary data. To be able to interpret the flow of information and the coupling directions, we will use our biological and medical knowledge about the system.

2.2 Theory

In the following sections the method for the calculation of correlations between two time series will be presented. The straight calculation of the correlation in time domain can have some drawbacks. For example when large oscillations and drifts are superimposed to the data or if the correlation of the high frequency content between the processes is of interest. Therefore the change from time domain to frequency domain is an appropriate solution, since all frequencies are treated equally.

Correlation in the frequency domain

The calculation of the correlation in the frequency domain is equivalent to the correlation in time domain after filtering the time series with the corresponding frequency filter. In the appendix the formula for the correlation in the frequency domain between two time series $x(t)$ and $y(t)$ in a rectangular frequency window ω_1 to ω_2 is derived as

$$c_{x,y} = \frac{\frac{1}{2}\int_{\omega_1}^{\omega_2} \tilde{x}^*\tilde{y} + \tilde{x}\tilde{y}^* \,\mathrm{d}\omega}{\sqrt{\int_{\omega_1}^{\omega_2} \tilde{x}\tilde{x}^* \,\mathrm{d}\omega \int_{\omega_1}^{\omega_2} \tilde{y}\tilde{y}^* \,\mathrm{d}\omega}} \tag{2.1}$$

where $\tilde{x}(\omega)$ depicts the fourier transform of $x(t)$. Note: The equation is only valid for long time series with zero mean ($\langle x\rangle, \langle y\rangle = 0$).

It turns out, that the definition of the correlation of frequency filtered time series is equivalent to calculating the correlation between the (complex) fourier coefficients in the corresponding frequency window. The fourier coefficients can either be calculated by using a discrete fourier transform (DFT) or more advanced spectra estimation methods as for example the multi taper method (MTM) used in the analysis of Brawanski et al. [2002]. A comparison of the two spectra estimation methods applied to the neuromonitoring data showed the same basic behaviour. Therefore we will only present results obtained with the discrete fourier transform.

To take the nonstationarity of long data sets into account, usually a windowed discrete fourier transform or often called short time fourier transform (STFT) is applied. Such methods can give more insight into the variability of the correlation over time. The window length is chosen to be much smaller than the length of the full time series. The data in the chosen window is then assumed to be stationary.

Smoothing data and removing drifts in time domain

For the removal of unwanted artifacts in time domain, the time series can be convoluted with a kernel. This procedure is often called "smoothing". In particular the removal of long term drifts and the smoothing of the discrete data as well as removing the noise from the analog measurements is of interest for the neuromonitoring data analysis. We can write the convoluted/smoothed time series $x(t)$ as

$$x_{\text{smooth}}(t) = \frac{1}{\tau}\int_{t-\tau}^{t+\tau} x(t')\cdot k(\frac{t'-t}{\tau})\,\mathrm{d}t' \tag{2.2}$$

with τ depicting the width of the kernel. Different kernels for the convolution can be used, as for examples a rectangular ($k(x) = \frac{1}{2}$) or parabolic ($k(x) = \frac{15}{16}(1-x^2)^2$) kernel, where $x \in [-1, 1]$. For the following analysis we will always choose the parabolic kernel, since its behaviour is smoother than the rectangular one.

Cross-correlation of two time series

To find the most probable time delay between two time series, the method of cross-correlations is used. The time series are shifted in time relatively to each other and the correlation is calculated for every time shift τ.

$$\langle x(t)y(t+\tau)\rangle = \int x(t)\,y(t+\tau)\,\mathrm{d}t \tag{2.3}$$

The τ at which $\langle x(t)y(t+\tau)\rangle$ has a maximum, denotes in practical applications the most probable time shift between the time series. This method will be applied in the further analysis for two different purposes. If the time series are recorded by

the same measurement device, than the detected time delay τ between the time series can give an indication to the physical processes producing such a time shift. An interpretation using some basic biological and medical knowledge will then be possible. But if the time series were recorded by two different measurement devices and no absolute time stamp was available, the detected time delay will give no valuable information.

In general, any time shift between the signals should be removed before applying the frequency correlation analysis. Ignoring this time delay would lead to virtual phase shifts in the fourier coefficients and distortions in the correlations. A simple example for this phenomenon is the shifting of two sinusoidal signals by 1/4 of the period (corresponding to a phase shift of $\pi/2$). This shifting leads to a correlation coefficient between the two signals of zero even if the signals are meant to be highly "correlated".

2.3 Application: Correlation between ...

To analyse the recorded neuromonitoring data, i.e. Licox, Invos, arterial blood pressure, intracranial pressure and the temperature, the method of correlation in frequency domain will be applied.

First analyses between the Invos and Licox data were performed by Brawanski et al. [2002] showing the possible application of the method to neuromonitoring data. In the following sections we will discuss all interesting correlations between the recorded brain status parameters and give some indications to the origin of the observed phenomena. Explanations of the phenomena will be discussed in more detail in chapter three when a model for the haemodynamic and metabolic processes in the brain is presented.

2.3.1 Invos on left and right hemisphere

First of all we will investigate the data from the Invos sensor on the left and right hemisphere to test on one hand the reliability of this measurement method and on the other hand to show in detail the analysis techniques used in this chapter. All further analysis of the neuromonitoring data will be done in the same way, but only the most important results are then presented.

The monitoring of patients at the intensive care unit with two Invos sensors to get an indication of the blood supply of *both* hemispheres is a standard procedure. Naturally we would expect an equal behaviour of the sensor data, but under some circumstances as for example after a severe injury on one hemisphere or an injury close to the Invos sensor, differences can appear.

As an example, the raw data from an Invos measurement on the left and right hemisphere over a time period of 34 hours with a sample rate of 5 seconds is shown in figure 2.1.

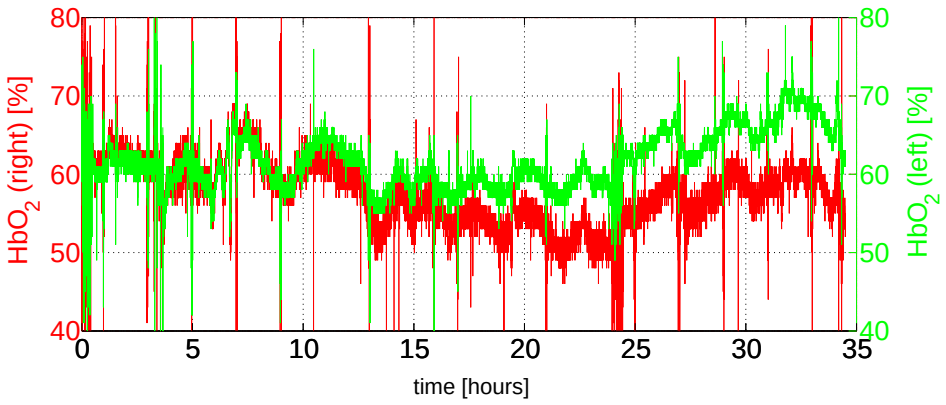

Figure 2.1: Raw data from the Invos sensor on the left and right hemisphere of a patient. The relative saturation of the haemoglobin (Hb) with oxygen (O_2) is plotted for both sensors over a period of 34 hours (samples every 5 s).

Time shift analysis

To determine the time shift between the time series of the Invos sensors, we use the method of cross-correlation. The maximum of the correlation gives the most probable time delay. For these two data sets, which were recorded by the same measurement device, no time shift between the time series was detected. This is expected since the Invos sensor measures the saturation of the haemoglobin with oxygen in the blood and as long as the blood supply of one of the hemispheres is not impaired no time delay should be seen.

Power spectrum analysis

Further the power spectrum or more precisely the absolute value of the coefficients of the fourier transformed time series can give an insight into the underlying processes. Due to the spectral separation of different processes, we find two fundamental contributions in the Invos sensor data as shown in figure 2.2.

The typical one-over-f behaviour seen in many natural systems can also be identified in the data from the Invos sensor. Further a white noise contribution as it is often seen as noise from measurement devices can be detected. One-over-f time series, i.e. the power behaves as 1/f, are often interpreted as self-similar or fractal time series, because they show interesting fluctuations on many different time scales. A very intuitive introduction with many examples to this topic is given by Gardner [1978]. Compared to one-over-f noise, the white noise has a power spectrum of the time series which is independent of the frequency, i.e. equal power in every frequency. A white noise contribution typically shows a complete random behaviour in the time series.

The crossing of the two contributions occurs in the power spectrum at around 2.5 mHz or a cycle length of 6 minutes. Above this frequency the white noise con-

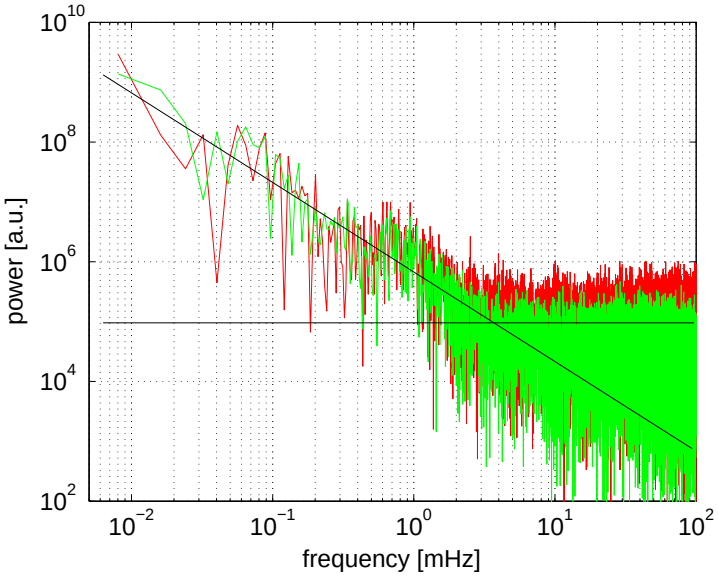

Figure 2.2: Power spectrum of the Invos data shown in figure 2.1. The power is given in arbitrary units (a.u.) over the frequency using a double logarithmic plot. Two contributions can be identified: firstly the one-over-f type signal (more accurate a $\frac{1}{f^\alpha}$ signal with $\alpha = \frac{3}{4}$) represented here as a line with a negative slope and secondly a contribution of white noise (horizontal line). The intersection of both lines occurs at a frequency of $\sim 2.5\,\text{mHz}$ or a cycle length of $\sim 6\,\text{min}$.

tribution dominates the one-over-f behaviour. Interestingly the power spectrum suggests that even for high frequencies the one-over-f noise contribution never vanishes.

In the whole analysis of the neuromonitoring data, this phenomenon of two contributions to the power spectrum is seen in all time series analysed.

Correlation in frequency domain

After having performed the analysis on the common behaviour of the data, the correlation between the time series in the frequency domain is investigated. Small frequency windows are taken for the analysis, usually 10 to 30 fourier coefficients per window, to calculate the correlation between the time series. For a more intuitive representation of the results a plot of the correlation over the cycle length instead of the frequency is chosen in figure 2.3.

A clear difference can be seen between the correlation of the high and the low frequency content. The Invos data is well correlated for cycle lengths above 5 minutes (correlation coefficient is above 0.5), while for smaller cycle lengths the correlation breaks down and is negligible. Note, with statistical tests using white and coloured (time correlated) noise the level of the correlation coefficient for

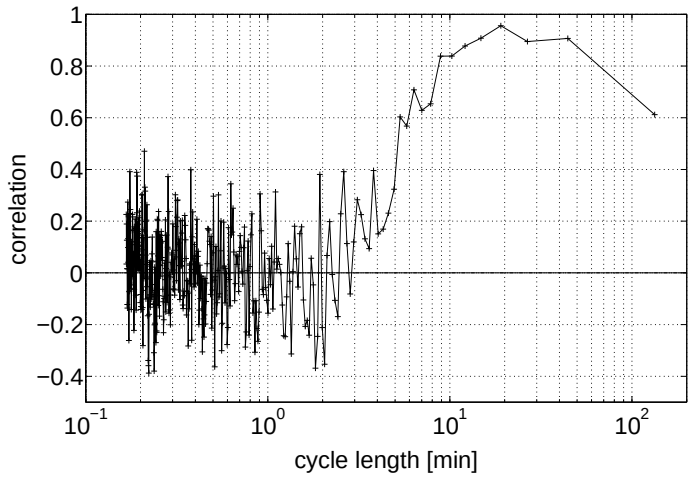

Figure 2.3: Correlation in the frequency domain between the Invos signals shown in figure 2.1. Every point corresponds to the correlation between ~ 20 fourier coefficients plotted at the mean cycle length of the chosen frequency window. Both Invos signals are significantly correlated (correlation coefficient above the noise level of ~ 0.5) for cycle lengths above 5 minutes. For long cycle lengths (> 2 h) as well as for high frequencies/short cycle lengths the correlation breaks down.

significantly correlated signals can be determined as shown in Brawanski et al. [2002].

Furthermore, a small drop of the correlation for very long cycle length is seen in the plot. This can be explained by the drifts of the sensors coming either from the sensor equipment itself or by possible changed environment in the tissue beneath the Invos sensor.

Interestingly, the signal in the correlated frequency range (above 5 minutes cycle length) corresponds to the one-over-f behaviour in the power spectrum plot (figure 2.2) which typically corresponds to the behaviour of natural systems. Further the crossing of the lines of the one-over-f noise and the white noise occurs at around 5 minutes cycle length, where the power of the one-over-f signal is about 10 times less then the one from the white noise. Tests with synthetic data have shown a robustness of the correlation method down to a signal to noise ratio (SNR) of 1:10. This SNR corresponds exactly to the crossing of the two lines in the power spectrum. Therefore, correlations between the two Invos signals in the higher frequencies can not be excluded.

Reconstruction and filtering of the data

To get an impression of the waveform of the significantly correlated part of the signals, the data can be reconstructed by using only the significantly correlated frequencies in the inverse fourier transform. The resulting signals of the Invos

sensor for the left and right hemisphere are shown in figure 2.4.

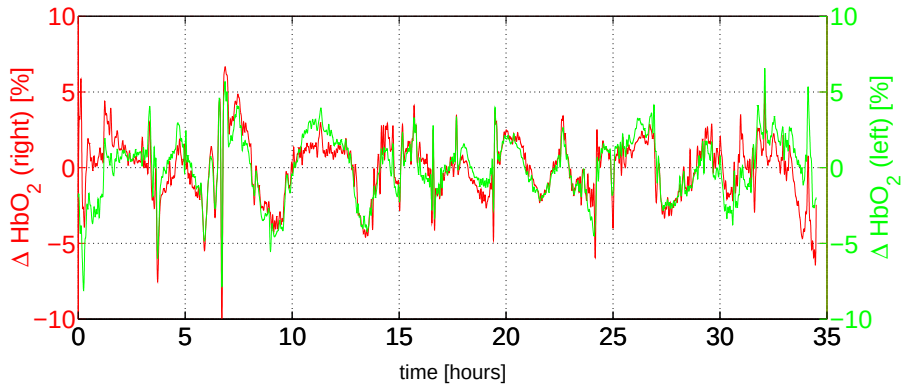

Figure 2.4: Reconstructed Invos signals using only the significantly correlated frequencies shown in figure 2.3. This corresponds to a frequency filtering using a rectangular window between 5 minutes and 2 hours.

Comparing these signals to the raw data (shown in figure 2.1), the removal of the noise and the drift by the reconstruction can clearly be seen. With this knowledge, we could try to obtain similar data just by filtering or smoothing the data in time domain using the kernel convolution method. Using a drift removal method of ±2 hours (i.e. smoothing the data with a ±2 hours kernel and subtracting this from the original data) gives the drift removed signal. Further smoothing the data with a ±2 minutes parabolic kernel, the signals shown in figure 2.5 are obtained.

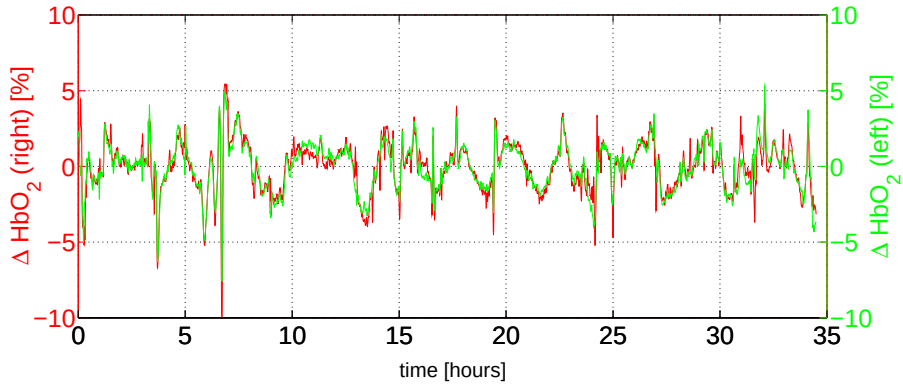

Figure 2.5: Raw data from the Invos sensors filtered in time domain using a smoothing of ±2 min and a drift removal of ±2 hours. Filtering in time domain shows the same result as the frequency filtered time series in figure 2.4.

It turns out that the signals from both Invos sensors (after applying the kernel convolution/smoothing method) match even better than the ones obtained by the

frequency reconstruction method. The deviations at the boundaries seen in the frequency reconstruction are due to the sharp filtering in the frequency domain and disappear when using the time domain filtering.

Conclusions

From the theoretical point of view, we can assume to have a one-over-f process from the natural system (the patient) which is covered by white noise originating from the measurement device (the Invos sensor). Filtering the data in time domain using a ± 2 hours drift removal and a ± 2 minutes smoothing, we obtain reasonable data showing the correlated content of both Invos channels.

From the medical point of view we can conclude, that the Invos sensor is a reliable measurement method, since the correlation between the time series is stable over long time periods. For the further analysis, we are only interested in the variations (above 5 minutes) of the signals, since Metz [2001] already indicated that only the dynamics of the signals is of interest for the clinical analysis. Further, the long time drifts seem not to contain any valuable information and can therefore be neglected.

2.3.2 Licox and Invos

After having presented the detailed analysis of the Invos data, the data from the two sensors measuring the oxygen supply of the brain is now analysed. The partial oxygen pressure in the brain tissue (measured by the *Licox* sensor) is compared to the oxygen content in the blood or more precisely to the saturation of haemoglobin with oxygen in the blood (measured by the *Invos* sensor).

A correlation between the two parameters is expected since both sensors measure the available oxygen for the nerve cells. First analyses have been done for previously recorded (Invos and Licox) data by Brawanski et al. [2002] where the authors presented how advanced spectral estimation methods can be used to calculate correlations between biomedical time series. The question on the common information content between Invos and Licox could be positively answered, showing the stability of the correlation on a large group of patients.

Problems in the measurements of Licox and Invos can arise if they are placed too close to or into an injured tissue. In such a case no correlations between the sensors can be expected.

First the time shift between the signals is investigated. As described in the introduction, the problem of time synchronisation can arise if two different measurement devices are used. For the pair of sensor signals analysed in this section, the devices were started at different points in time. Therefore any coupling directions or time delays due to processes in the brain, can not be investigated by using the cross-correlation analysis. But for the further analysis the data was aligned in time.

Performing the correlation analysis in the frequency domain on this time aligned data sets, the plot in figure 2.6 is obtained. As previously seen, the plot can be divided into two main parts, the high frequency content (below 6 minutes cycle length) which is not significantly correlated and a low frequency content (between 6 minutes and 3 hours) which is highly correlated. Therefore both signals seem to have only common information down to a cycle length of 6 minutes. Possible correlations in the higher frequency range are either not resolvable by the method or not existing. The breakdown of the correlation above 3 hours is again due to the drifts respectively trends in the data.

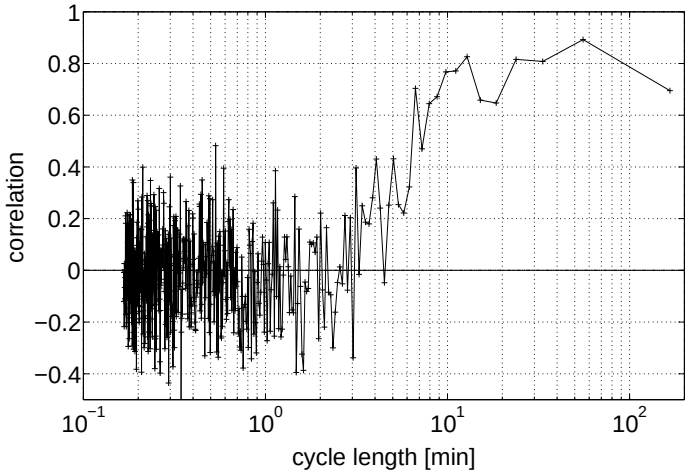

Figure 2.6: Correlation in the frequency domain between Licox and Invos signals based on continuous data of 27 hours. Both signals are significantly correlated (correlation coefficient above the noise level of ~ 0.5) for cycle lengths above 6 minutes. For very long cycle lengths (> 3 h) as well as for high frequencies/short cycle lengths the correlation breaks down.

Overall the signals are well correlated over long time periods as was expected and seen in previously recorded data. Still, the result of a *linear* correlation is an interesting fact for the description of the oxygen supply in the brain. Both sensors are measuring in some sense the oxygen content in the brain, but their measured values are connected by a *nonlinear* dissociation curve, the diffusion process and the cerebral blood flow as will be seen in chapter 3.3.

The actual variations of the signals are thought to be changes in the cerebral blood flow (CBF). The CBF depends on one hand on the arterial blood pressure (as will be shown in the next section) and on the other hand on additional parameters influencing the arteries regulating the CBF.

Sometimes huge variations are seen in the data sets (mainly in the Licox values) originating from a medical manoeuvre such as the cleaning of the air tube of the patient. The patient is then supplied with pure oxygen prior to the manoeuvre and therefore high oxygen values are measured.

A detailed description of the relation between Licox and Invos, taking all physical processes into account, is given in chapter 3.3.

2.3.3 Arterial blood pressure and oxygen supply

In this section we will focus on the connection between the haemodynamics, i.e. primarily the arterial blood pressure, and the oxygen supply, i.e. the measured signals from Invos and Licox. As shortly mentioned in the previous section, the variations of the signals from the Licox and Invos sensors are assumed to be due to variations of the cerebral blood flow. If more blood flows through the brain, more oxygen per time is available for diffusion into the tissue and therefore a higher oxygen content should be measured.

The variations in the cerebral blood flow itself are expected to be due to the variations in the arterial blood pressure (ABP) as a higher ABP leads to a higher blood flow and in the end to a higher oxygen content in the tissue. This connection between ABP and the values measured with the Licox respectively Invos sensor is investigated in the following.

Performing the same analysis as done in the previous sections, we obtain a clear correlation between the signals for selected parts of the data. Again a look at the correlations in the frequency domain (see figure 2.7) gives an insight into the possible processes.

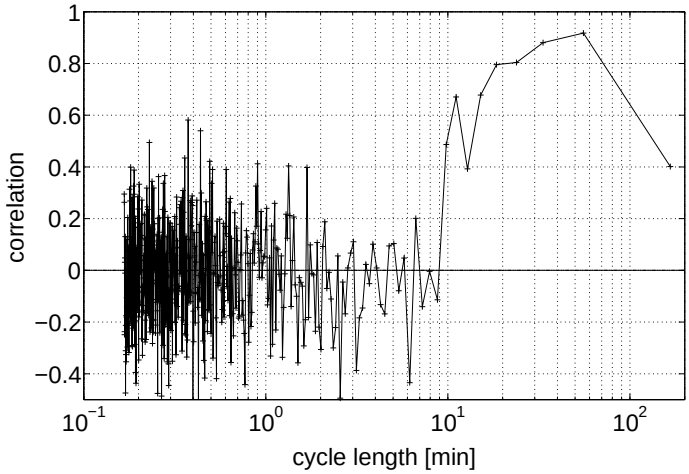

Figure 2.7: Correlation in the frequency domain between the arterial blood pressure and the oxygen content in the brain tissue (Licox sensor). The analysis is based on continuous data of 21 hours. Both signals are significantly correlated (correlation coefficient above the noise level of ~ 0.5) for cycle lengths above 10 minutes. For longer cycle lengths (> 1.5 h) as well as for high frequencies/short cycle lengths the correlation breaks down.

Noticeable is the small but clearly correlated region in the frequency domain ranging from 10 minutes to just over one hour. Outside this range the correlation breaks down quickly. On the low frequency end this can be explained by clearly visible trends in the data and on the high frequency end by either a high noise level and/or a damping mechanism from the brain.

The time shift for this pair of signals could be measured, because both measurements were simultaneously recorded by the same device. A cross-correlation analysis shows that the variations of the arterial blood pressure are seen $\sim$ 135 seconds earlier than the ones from the Licox sensor. This can be explained by two processes. Firstly by the reaction time (the so-called $t_{90\%}$ time) of the Licox sensor of around 90 seconds and secondly by the time the blood needs to reach the capillaries and oxygen to diffuse into the tissue. From this knowledge we can conclude that the arterial blood pressure must be the driving force of the seen variation in the data measured by the Licox sensor.

But as mentioned before, we can only see this correlation in parts of the data sets (less than 10% of the data). Later, during the discussion of the model, we will learn that the autoregulation plays a major role in the correlation of the ABP with the oxygen supply. Only if the autoregulation, i.e. the regulation of the blood flow by the brain's arteries, is impaired, a correlation between ABP and Licox/Invos should be seen.

2.3.4 Arterial blood pressure and intracranial pressure

Focusing on the haemodynamics, we will now investigate the connection between the driving force of the system, the arterial blood pressure, and the intracranial pressure, which is one of the most important brain status parameters monitored on the intensive care unit. The relation between these two parameters is an important factor to understand the hydrodynamic system in the brain.

The human head can be seen as a closed compartment in which the brain and the intracranial fluid is contained. A local swelling in the brain can lead to an increase of the intracranial pressure by a diminishment of the available volume and therefore a compression of the brain tissue. The question to be asked, when analysing the sensor data, is, whether an increase of the arterial blood pressure can lead to an increased intracranial pressure or not.

This connection could be expected, since the increase of the arterial blood pressure leads to an increase of the (arterial) blood vessels and a reduction of the available volume in the brain. But this increase of volume can be compensated by an increased absorption of the cerebrospinal fluid as described in the first chapter.

In the analysis of more than 670 hours of data we could identify a section of 9 hours of continuous data showing a significant correlation. As usual, the correlation between the two time series is plotted over the cycle length of the corresponding frequency window (figure 2.8).

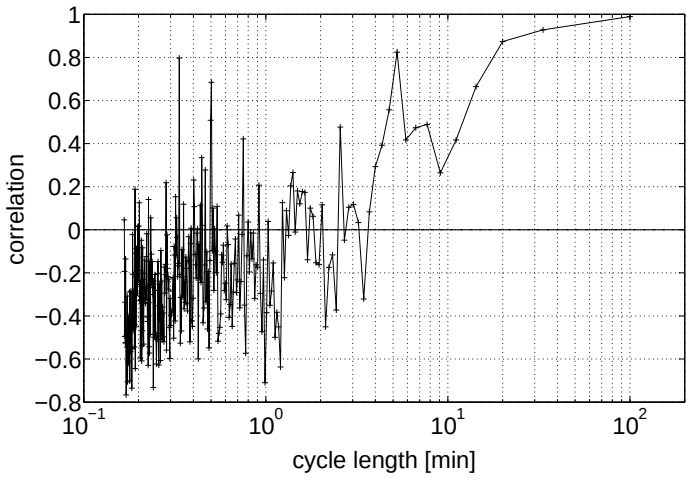

Figure 2.8: Correlation in the frequency domain between the haemodynamical signals: the arterial blood pressure and the intracranial pressure. The analysis is based on continuous data of 9 hours. A significant correlation (correlation coefficient above the noise level of ~ 0.5) is only given for cycle lengths above 15 minutes. For shorter cycle lengths one can see a good correlation around 5 minutes, before it breaks down for high frequencies/short cycle lengths.

Again the plot showing the correlation can be divided into two main parts. The high frequency content is not well correlated, except a few high peaks below 1 minute cycle length which are probably artifacts of the measurement device. Whereas the low frequency content has a high correlation, in particular for cycle lengths above 20 minutes. To get a better impression of the actual waveform, the arterial blood pressure and intracranial pressure are plotted (after a filtering in the time domain) in figure 2.9. From the plotted time series, a very good correlation of the larger and longer oscillations can be seen in contrast to the ones on the smaller time scale.

Since both measurements are recorded with the same analog/digital converter, the time shift between the signals can be calculated. The variations of the arterial blood pressure are seen 30 seconds earlier than those from the intracranial pressure. This supports the hypothesis that the arterial blood pressure is the driving force of the system.

As mentioned before, the correlation between the haemodynamic parameters (ABP and ICP) could only be seen very rarely (in 1.3% of the data). This will get more clear in chapter 3, where the model of the haemodynamic and metabolic processes in the brain is presented. We will see, that the ICP is stable, as long as cerebrospinal fluid (CSF) can be absorbed into the venous blood to compensate the increased volume of the arteries and other blood vessels. Without CSF no compensation is possible and a rise in the ABP is followed by an increasing ICP.

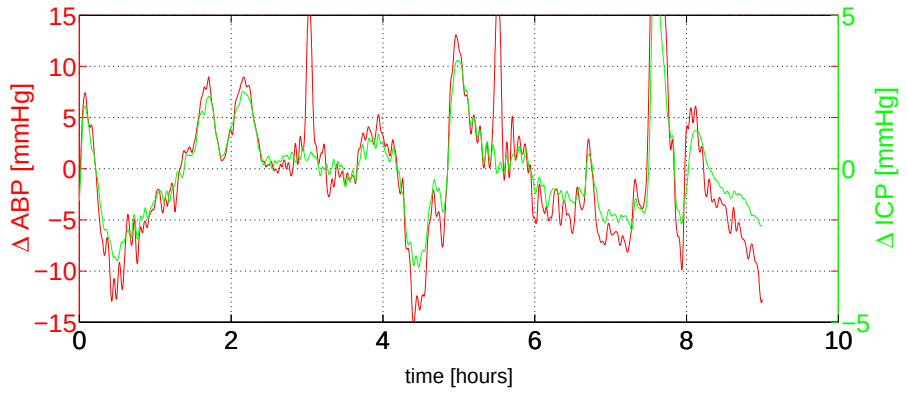

Figure 2.9: Filtering the raw data from the arterial blood pressure and the intracranial pressure in time domain with a smoothing of ±2 min and drift removal of ±4 h. The correlation of the long and big oscillations are clearly visible, in contrast to the short time oscillations which are not well correlated. This observation is in good agreement with the plot in figure 2.8.

Then, the arteries transmit the pressure directly to the brain tissue.

From the medical documents and nuclear magnetic resonance images we can support this theory. The cerebrospinal fluid was (up to a small amount) fully absorbed. The patient had a severe swelling of the brain, the scalp had to be opened to overcome a further increase of the brain pressure.

2.4 Conclusions

Concluding the results of this chapter, we distinguish between a theoretical and a medical part.

Focusing on the theory: With the correlation analysis in the frequency domain we can extract a lot of valuable information. It seems that the signals are often linearly correlated. Higher order statistics like mutual information or transfer entropy seems not to be necessary to reveal higher order correlations. In preceding tests, these (higher order statistics) methods were also applied to the data, but showed equivalent results. However, difficulties with the statistics arise, since these methods need longer time series with a better signal to noise ratio.

The analysis of the correlation in the low frequency range (variations above 5 to 10 minutes) shows a high correlation between the signals. In the power spectrum this frequency range corresponds to a one-over-f type noise behaviour which is typically seen in natural systems. The correlation in this frequency range seems to be a good measure on which statistical tests can be applied. This measure could therefore be used in future devices monitoring the patient. Changes in the correlation of the signals could indicate a variation in the patients condition and raise an alert.

The drifts or trends in the time series (cycle length above 2 hours) seem not to contain any valuable information, no correlation is seen between these signals. Therefore we suggest to remove the drift in time domain by kernel-smoothing. But one exception should be mentioned here. The drift of the Licox sensor is correlated with the temperature seen in at least two data sets of 30 and 50 hours. An explanation for this correlation can not yet be given, since the Licox sensor is temperature compensated and calibrated. The phenomenon is still under investigation by the company providing the Licox sensor, but maybe a physiological explanation is more reasonable than a technical one.

Furthermore the signals below 2 minutes cycle length have either no common information content or the method used in this analysis cannot resolve these correlations. Still the typical one-over-f type behaviour is seen in this frequency range, even if it is covered by a dominating white noise process. Therefore we suggest to smooth the data in time domain with a kernel of ± 2 minutes.

From the medical point of view: The Invos sensor gives reproduceable and stable measurements over longer periods of time (34 hours – as shown in the example above), but the placement of the sensor is important. When correctly placed (not too close to the injured tissue) the variations/dynamics of the data can be trusted. The absolute values of the sensor are not always reliable due to unknown trends.

Under normal conditions we can also see stable correlations between Licox and Invos – in the example shown above on a period of 27 hours. The origin of the actual variations of the data values can be explained on one hand by variations in the cerebral blood flow, which produces only small variations and on the other hand by clinical manoeuvres (supply of the patient with air consisting of 100% oxygen) resulting in huge oscillations in the data. These phenomena are explained in great detail in chapter 3.3.

The arterial blood pressure seems to be the driving force of the hydrodynamic respectively the haemodynamic system if a disease is present. In the case of an impaired autoregulation (was seen only in less than 10% of the time) it can lead to the coupling between the ABP and the Licox values due to the linear dependence of the cerebral blood flow on the ABP. If a swelling of the brain is present, then the ABP couples also with the ICP due to the missing CSF to compensate the volume changes (this was only seen very rarely in 1.3% of the time).

Chapter 3

Model of the haemodynamic and metabolic processes in the brain

With the aid of biological and medical knowledge about the dynamical properties of the brain – mainly the haemodynamic and metabolic (oxygen supply) processes – a global model for the dynamical behaviour of the system "brain" can be designed.

Previous analysis methods only focused on parts of the system, using for example correlation analysis of the time series. With the capability of investigating the system using a model it should be possible to explain the observed phenomena within a global context.

The final goal of such an analysis is to find a model which can reproduce and explain all the measured data, i.e. the combination of haemodynamics and metabolics. In this case it would then be possible to compare simulations of the model with the measured data – after fitting free parameters of the model. Deviations between the prediction of the model and the measured data could indicate a possible change of the parameters i.e. the patients state of health.

In this chapter we propose a basic hydrodynamic model with realistic extensions to existing models which fits the needs of neurosurgical applications. In a second step we will couple this to a newly derived model describing the metabolic processes in the brain. This will result in a new so-called *combined model*. The main connection between the two models is the cerebral blood flow which is hardly measureable in clinical applications over longer periods. Therefore this parameter determined by the model can give valuable information to the physicians for the treatment of the patient.

In the following the fluid dynamical and the oxygen supply model will be described in detail and validated on typical curves and measured data. It will get clear, how the observed phenomena seen in the previous section can be explained within the context of this model.

3.1 Introduction

In this chapter, first a hydrodynamical model based on interacting subunits, the compartments, will be derived. Compartment models of the present type have a long history, dating back to the earliest of such models formulated by Monro and Kellie in the end of the 18th and begin of the 19th century using a three compartment model (arterial, venous and brain tissue). More recent works on this type of models are using an increased number of compartments and relaxations of some assumptions. But many of these models lag of the correct description of one or the other processes in the brain. Often, instead of a full description of the autoregulation processes, a constant blood flow is assumed or the compressibility of the brain tissue is neglected (Kadas et al. [1997], Lakin et al. [1996]). For the modelling of the autoregulations, Czosnyka et al. [1992] and later Bekker et al. [1996] proposed the use of a pressure depending nonlinear resistance, regulating the blood flow. Extensions to the description of the capacitance of the blood vessels and the brain tissue were then given in more detail by Czosnyka et al. [1997] and comparisons with measurements by Stevens [2000].

The model presented in this work will more closely follow the one proposed by Ursino and Lodi [1997], since it uses a more realistic description of the autoregulation process and takes into account the possible collapse of the veins in a preliminary step. Still, some more consistent extensions will be presented here, in particular regarding the compression of the veins and using a simpler equation for the autoregulation. Many basic concepts of the later works of Ursino were already presented in the excellent paper from 1988 (Ursino [1988]).

After presenting a hydrodynamical model, which is adapted to the requirements of neurosurgical applications, the details of the oxygen supply will be discussed. Krogh [1918-1919] proposed as the first a model for the description of the oxygen supply in tissue. An elaborated discussion of the literature and the assumption of the Krogh cylinder is given in Reneau et al. [1967] as well as results of numerical calculations. In the following years Reneau and others published further works considering countercurrent flows and the undersupply of tissue (Reneau and Knisely [1971], Reneau et al. [1970], Hudetz et al. [1982]). Still, the derivation of the analytical expression for the mean oxygen pressure in tissue, as measured by a device like the Licox sensor, and the oxygen content in the blood, as measured by the Invos sensor, and in particular their relationship, were not presented in any works published. The reason for this is that until a few years ago the data of both sensors couldn't be recorded over long periods. Therefore, the possibility to compare experimental measured data with theoretical predictions was not existing.

In this work we will also show, which further difficulties have to be overcome to get a meaningful comparison of the absolute values of the predictions with measured data.

3.2 Fluid dynamics

In this model approach, fluid and matter constituents within the human head are subdivided into a number of interacting subunits, namely the compartments. The interactions are mainly determined by fluid dynamical processes. Basically the blood and the cerebrospinal fluid (CSF) will interact as separate flows between the compartments, whereas the brain tissue will be depicted by one compartment consisting of soft matter.

The compartment model used in this work is presented in figure 3.1 showing a closed overall compartment with only one inflow and one outflow of blood. All other fluid dynamical processes take place in this overall compartment. In this model we will distinguish between the following 7 compartments:

A - arteries, C - capillaries, V - veins, S - sagittal sinus, B - brain tissue,
F - cerebrospinal fluid and E - extra volume describing swelling of brain tissue.

The details of each compartment will be discussed in the following sections.

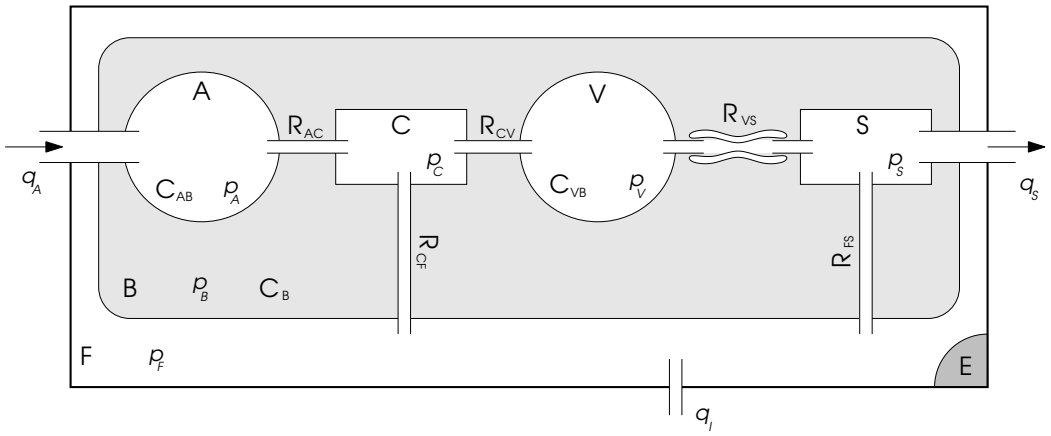

Figure 3.1: A hydrodynamical model describing the processes in the brain using 7 distinguishable compartments: A - arteries, C - capillaries, V - veins, S - sagittal sinus, B - brain tissue, F - cerebrospinal fluid and E - extra volume describing swelling of brain tissue. For every compartment a pressure p_X can be defined as well as flows (q_{XY}) respectively resistances (R_{XY}) between the compartments. Furthermore for volumes with a membrane a capacitance C_{XY} is defined.

To make it easier to specify for example the pressures in the compartments or the flows between compartments, the following conventions will be applied. For the pressure in compartment "A" the variable p_A will be used. When two compartments are involved as for example the flow between compartment "A" and "C", the variable q_{AC} is used. To depict the standard value of a variable an index n will be attached as for example q_{ACn}.

The physical equation describing the hydrodynamical processes is based upon the conservation of mass and reads as

$$\sum q_i = \frac{dm}{dt} = \underbrace{\frac{\partial m}{\partial \rho}}_{=V} \cdot \frac{\partial \rho}{\partial t} + \underbrace{\frac{\partial m}{\partial V}}_{=\rho} \cdot \frac{\partial V}{\partial t} \quad \text{with} \quad m(t) = \rho(t) \cdot V(t). \tag{3.1}$$

The change of mass in the compartments depends on the fluxes in and out of the compartment (q_i) as well as on the changes of the density ($\frac{\partial \rho}{\partial t}$) and on the change of the volume of the compartment ($\frac{\partial V}{\partial t}$). For incompressible media, like the fluids, $\frac{\partial \rho}{\partial t}$ is of course zero. Therefore the first term in (3.1) can be neglected.

3.2.1 Assumption for the hydrodynamical model

In the following, the assumptions for the compartment models used in this work are stated from which the equations are derived later:

i) The blood and the CSF (mainly consisting of water) are *fluids* and can be assumed as *incompressible* media. Therefore a constant density in the compartments ($\rho = \text{const} \Rightarrow \frac{d\rho}{dt} = 0$) due to the *isotropy* is assumed.

ii) The brain tissue is, in contrast to the fluids, a *compressible* medium, i.e. $\rho \neq \text{const} \Rightarrow \rho(t) = f(p(t))$ for a given volume and temperature ($V, T = \text{const}$). In our model we always assume the temperature to be constant.

iii) The flow in the blood vessels is *laminar*, since the Reynolds number of blood in the arteries and the capillaries is well below the transition Reynolds number of 2000, which defines the change from a laminar to a turbulent flow. Therefore we can use the *Hagen-Poiseuille law* to describe the flow between the compartments. According to this law, the flux depends only on the pressure difference between compartments A and C and the resistance of the vessels

$$q_{AC} = \frac{\Delta p_{AC}}{R_{AC}} = \frac{p_A - p_C}{R_{AC}} \tag{3.2}$$

iv) The membranes between the compartments, i.e. the blood vessels, are assumed to be *elastic*. Accordingly, we define the *capacitance* of the vessels as $dV = C(p)dp$. For the volume of a compartment with an elastic membrane it follows:

$$V - V_0 = \int_0^p C(\varphi)\mathrm{d}\varphi \tag{3.3}$$

where $p = p_{external} - p_{internal}$ is the membrane pressure. For the volume change $\frac{dV_A}{dt}$ of volume A and an external pressure p_B we derive from

$$V_A - V_{A_0} = \int_0^{p_A - p_B} C(\varphi)\mathrm{d}\varphi \quad \text{the equation} \quad \frac{dV_A}{dt} = C(p_A - p_B) \cdot \frac{d(p_A - p_B)}{dt}. \tag{3.4}$$

If the capacitance is independent of the pressure we obtain from

$$V_A - V_{A_0} = C \cdot (p_A - p_B) \quad \text{the equation} \quad \frac{dV_A}{dt} = C \cdot \frac{d(p_A - p_B)}{dt}. \tag{3.5}$$

v) The cranial bone acts as a closed compartment, i.e. the total volume is *constant*, and therefore the Monro-Kellie doctrine can be formulated as

$$\sum_i V_i = V_{\text{total}} \quad \Rightarrow \quad \sum_i \frac{dV_i}{dt} = 0. \tag{3.6}$$

3.2.2 The compartments

Before going into the details of the fluid compartments, we will shortly address the role of the brain tissue (compartment "B") in the model. The brain tissue consists of soft material which can be compressed. An increase of fluids (CSF and blood) in the brain can therefore be compensated by the brain tissue. But this has of course as consequence an increase of the intracranial pressure.

The compliance of compartment B is assumed to be depending on the intracranial pressure p_B. In Ursino and Lodi [1997]) an equation for the compliance of the brain tissue is given, which will be slightly modified by adding an extra pressure p_{B0} to overcome an unrealistic divergence at $p_B = 0$,

$$C_B(p_B) = \frac{1}{k_B(|p_B| + p_{B0})} \tag{3.7}$$

where k_B defines the stiffness of the brain tissue. Furthermore, the membrane between the brain tissue and the cerebrospinal fluid compartment is very thin and soft, i.e. equal pressures in both compartments ($p_B = p_F$) can be assumed. First measurements with dual ICP sensors in the brain tissue and in the CSF compartment support this assumption (Woertgen et al. [2003]).

In the following subsections, we will go through the compartments of the model and shortly mention their important properties.

Autoregulation of the arteries

The arteries, described by compartment "A", have due to their muscular layer the capability of controlling the cerebral blood flow. The blood flow through the brain is kept constant over a wide range of the blood pressure to guarantee a continuous supply of the tissue with oxygen. This mechanism is called *autoregulation*. A measurement on baboons describes the dependence between blood flow and arterial blood pressure (figure 3.2).

Physiologically, the arteries have via their muscles the ability to dilate or constrict which results in a changed vessel resistance. In the model, this regulation is therefore described by adjusting the resistance R_{AC} of the arteries. In the case

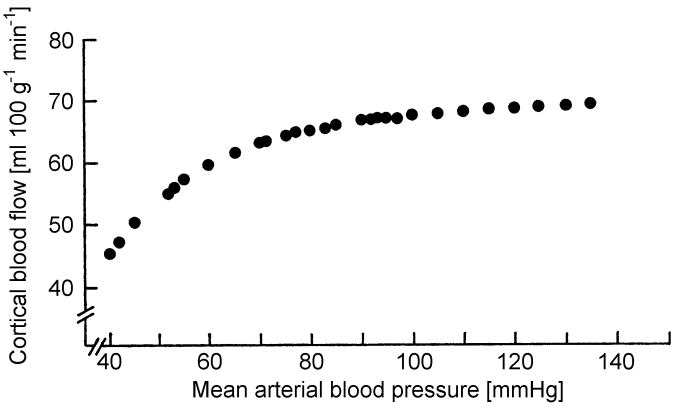

Figure 3.2: The cerebral blood flow in relation to the mean arterial blood pressure shows the autoregulation of the brain. In a wide pressure range the blood flow is constant. In the case of an impaired autoregulation one obtains a linear relationship between pressure and flow (not shown here). Values are taken from measurements on baboons (adapted from Purves [1972]).

of severe head injuries or deep anaesthesia, the autoregulation can be impaired resulting in a nearly constant resistance of the blood vessels and finally in a linear dependence between flow and pressure.

Assuming a microvascular bed consisting of a parallel arrangement of several microvessels with equal inner radius r, the blood volume is then directly proportional to r^2. Whereas the resistance can be assumed to be inversely proportional to r^4 (Hagen-Poiseuille law). We can then write for the resistance

$$R_{AC} = \frac{k'_{R_{AC}}}{r^4} = \frac{k_{R_{AC}}}{{V_A}^2} = \frac{k_{R_{AC}}}{{C_{AB}}^2 (p_A - p_B)^2}. \tag{3.8}$$

where $V_A = C_{AB} \cdot (p_A - p_B)$ and $k_{R_{AC}}$ is a constant determined by the standard values of the volume and the resistance of the arteries. Since the autoregulation is done by constriction and dilatation of the arteries (volume changes), depending on the blood flow, the compliance varies accordingly. The autoregulation is therefore a regulation mechanism of the compliance which can be described by the following differential equation

$$\frac{dC_{AB}}{dt} = -\frac{1}{\tau} \left(C_{AB} - C_{ABreg}(\alpha \cdot x) \right) \quad \text{with} \quad x = \frac{q_{AC} - q_{ACn}}{q_{ACn}} \tag{3.9}$$

where C_{ABreg} is an "ideal" adapted capacitance to guarantee the constant blood flow through the brain. The constant α specifies the degree of the autoregulation and τ the time constant of the mechanism. The function C_{ABreg} has the following form (proposed by Ursino and Lodi [1997]) with an upper and lower limit of the

compliance corresponding to maximal constriction and dilatation

$$C_{ABreg}(x) = C_{ABn} - \Delta C_{AB} \cdot \tanh\left(\frac{x}{\Delta C_{AB}}\right). \tag{3.10}$$

However the autoregulation curve is not symmetrical since the increase in blood volume (and compliance) induced by dilatation is higher than the decrease in blood volume induced by constriction. Hence, two different values must be chosen for the parameter ΔC_{AB}, depending on whether dilatation or constriction is considered. Therefore we have

$$\Delta C_{AB} = \begin{cases} \Delta C_{AB1} & \text{if } x < 0 \text{ i.e. dilatation,} \\ \Delta C_{AB2} & \text{if } x > 0 \text{ i.e. constriction.} \end{cases} \tag{3.11}$$

Cerebrospinal fluid (CSF)

A good general review about the main aspects of the production and absorption of the cerebrospinal fluid is given in Sullivan and Allison [1985]. The cerebrospinal fluid surrounding the brain tissue is interacting with three compartments: C - capillary blood vessels, S - sagittal sinus & F - the fluid compartment itself.

The production of the CSF is taking place at the capillary level, the ventricles, and the absorption in the sagittal sinus (see chapter 1.1). The driving force of both processes is the pressure difference between the capillaries and the CSF compartment respectively between the CSF compartment and the pressure in the sagittal sinus.

From a medical point of view it is clear, that a reversal of both fluxes is not possible. Therefore the resistances must be modelled as diodes. For the absorption of the CSF the amount of CSF volume must additionally be taken into account. When no fluid is available for absorption, the resistance must be infinite, since otherwise negative volumes would appear. It follows for the flows q_{CF} (= production) and q_{FS} (= absorption)

$$q_{CF} = \frac{p_C - p_B}{R_{CF}} \quad \text{and} \quad q_{FS} = \frac{p_B - p_S}{R_{FS}} \tag{3.12}$$

where

$$R_{CF}(p_C - p_B) = \begin{cases} R_{CFn} & \text{if } p_C - p_B > 0 \\ \infty & \text{otherwise} \end{cases} \quad \text{and} \tag{3.13}$$

$$R_{FS}(p_B - p_S, V_F) = \begin{cases} R_{FSn} & \text{if } (p_B - p_S > 0) \text{ and } (V_F > 0) \\ \infty & \text{otherwise} \end{cases} \tag{3.14}$$

Furthermore, an injection of fluid from outside of the brain into the CSF compartment can be modelled by a flow q_I. An inflation of a balloon or a swelling is modelled by the artificially appended compartment "E", where q_E defines the in- and outflow to this compartment.

The venous system

The venous system can be separated into two main parts: firstly the part leading from the small capillaries (compartment "C") to the bigger veins (compartment "V") and secondly from the bigger veins to the sagittal sinus (compartment "S"), which has a stiff wall and is therefore incompressible.

For the first part, the resistance (R_{CV}) is assumed to be constant, since the blood vessels are small and therefore relatively stiff. But when the veins are getting bigger, they behave like a rubber hose, because they have no active elements like muscles. The diameter – respectively the volume – depends only on the pressure difference between the venous blood pressure and the pressure surrounding the veins, the intracranial pressure. The resistance can be modelled in the same way as for the arteries (equation (3.8)) by using the Hagen-Poiseuille law

$$R_{VS} = \frac{k_{R_{VS}}}{{V_V}^2}. \tag{3.15}$$

The volume of the veins is determined by their compliance which is depending on the pressure difference. Using the formula proposed by Ursino [1988], but taking the absolute value of the difference $p_V - p_B$ to overcome a divergence, the compliance for the veins can be written as

$$C_{VB} = \frac{1}{k_V(|p_V - p_B| + p_{V_0})} \tag{3.16}$$

where k_V defines the stiffness of the veins. As we will see later, it is possible to derive the basic structure for the compliances of the veins and the brain tissue from the global pressure-volume curve shown in figure 3.5. This pressure-volume curve shows an exponential increase of the pressure, if the volume of a balloon is increased: $p = e^{(V-V_0)} + p_0$. With simple algebraic manipulations, (first taking the logarithm, then solving for V and finally taking the derivative with respect to p) one obtains for the compliance $\frac{dV}{dp} = C \sim \frac{1}{p-p_0}$.

3.2.3 Final set of equations

To obtain the final set of differential equations, we write down the equation of mass conservation for every compartment. This is the basic equation which must be fulfilled in every compartment. For the compartment "B" we use equation (3.6) to integrate the assumption of the head to be a closed compartment (Monro-Kellie doctrine).

Now putting together all equations including the ones presented in the previ-

ous section, we obtain for each compartment

$$A: \quad q_A - \frac{p_A - p_C}{R_{AC}} = \dot{C}_{AB} \cdot (p_A - p_B) + C_{AB} \cdot (\dot{p}_A - \dot{p}_B) \tag{3.17}$$

$$C: \quad \frac{p_A - p_C}{R_{AC}} - \frac{p_C - p_V}{R_{CV}} - \frac{p_C - p_B}{R_{CF}} = 0 \tag{3.18}$$

$$V: \quad \frac{p_C - p_V}{R_{CV}} - \frac{p_V - p_S}{R_{VS}} = C_{VB} \cdot (\dot{p}_V - \dot{p}_B) \tag{3.19}$$

$$B: \quad q_I + q_E + \frac{p_C - p_B}{R_{CF}} - \frac{p_B - p_S}{R_{FS}} = C_B \dot{p}_B - \dot{C}_{AB} \cdot (p_A - p_B) + C_{AB} \cdot (\dot{p}_A - \dot{p}_B) - C_{VB} \cdot (\dot{p}_V - \dot{p}_B) \tag{3.20}$$

$$S: \quad \frac{p_V - p_S}{R_{VS}} - q_S + \frac{p_B - p_S}{R_{FS}} = 0 \tag{3.21}$$

and for the autoregulation respectively the capacitance of the arteries

$$\dot{C}_{AB} = -\frac{1}{\tau} \left(C_{AB} - C_{ABreg}(\alpha \cdot x) \right) \quad \text{with} \quad x = \frac{q_{AC} - q_{ACn}}{q_{ACn}}. \tag{3.22}$$

Furthermore we have to take the temporal evolution of the volumes of compartment "V" and "F" into account

$$\dot{V}_V = \frac{p_C - p_V}{R_{CV}} - \frac{p_V - p_S}{R_{VS}} \tag{3.23}$$

$$\dot{V}_F = q_I + \frac{p_C - p_B}{R_{CF}} - \frac{p_B - p_S}{R_{FS}} \tag{3.24}$$

since the resistances $R_{FS} = R_{FS}(p_B - p_S, V_F)$ and $R_{VS} = R_{VS}(V_V)$ are depending on these volumes.

A closer look at the equations shows, that the mass conservation equations of compartment "A" and "S" (grey shaded) are redundant. Therefore, we finally obtain

5 coupled non-linear differential equations (3.19), (3.20), (3.22)-(3.24)
plus one extra constraint from compartment "C" (equation (3.18))
and as variables: p_B, p_V, C_{AB}, V_V and V_F.

This set of differential equations has to be solved numerically, where the constraint of compartment "C" causes difficulties, since in general it can't be explicitly solved for p_C. The external driving force of the system is described by $p_A(t)$ and $\dot{p}_A(t)$. At the outflow of the system (sagittal sinus), a constant pressure p_S is assumed.

3.2.4 Standard values

Before verifying the model on data known from medical literature, the standard values for the parameters of a healthy patient have to be determined. First the pressure distribution in the cerebral blood vessels will be investigated. This is shown in figure 3.3 for two different medical conditions. The pressure drops from the arteries over the capillaries to the veins, where at the end of the veins – in the sagittal sinus – the pressure is just above 0 mmHg.

As standard values for the pressures in the different compartments we therefore use

$$p_{An} = 100\,\text{mmHg} \qquad p_{Vn} = 15\,\text{mmHg} \qquad p_{Bn} = p_{Fn} = 10\,\text{mmHg} \tag{3.25}$$

$$p_{Cn} = 25\,\text{mmHg} \qquad p_{Sn} = 5\,\text{mmHg}. \tag{3.26}$$

In the medical literature, the standard unit of pressure is mmHg. To make it easier to compare the results with other works, we also use in this work mmHg as the unit for the pressure.

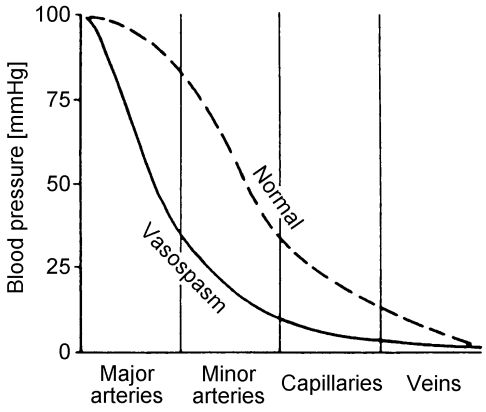

Figure 3.3: Distribution of the pressure in the blood vessels of the brain (more exactly in the cerebral vascular bed) under normal conditions and in a situation with constricted blood vessels (vasospasm) (adapted from Mchedlishvili [1986]).

For the corresponding volumes, we use the standard values given in the medical literature as for example in Purves [1972]

$$V_{An} = 15\,\text{ml} \qquad V_{Vn} = 40\,\text{ml} \qquad V_{Bn} = 1000\,\text{ml} \tag{3.27}$$

$$V_{Cn} = 10\,\text{ml} \qquad V_{Sn} = 80\,\text{ml} \qquad V_{Fn} = 30\,\text{ml} \tag{3.28}$$

where for volume V_F it has to be noticed, that from medical experiments it is known that only 30 ml of the total 170 ml cerebrospinal fluid can be absorbed. Most of the CSF is located in the spinal chanal, so that it can not be absorbed in the sagittal sinus.

The main flows of interest are the blood flow from the heart into compartment "A" and the production of cerebrospinal fluid from the capillaries in the ventricles into the compartment "F", where in medical literature the values

$$q_{An} = 600\,\tfrac{\text{ml}}{\text{min}} \qquad\qquad q_{CFn} = 0.4\,\tfrac{\text{ml}}{\text{min}} \tag{3.29}$$

are well established. With these basic parameters the standard values for the resistances can be calculated. The resistance between two compartments is defined as the ratio of the pressure difference and the flow between the compartments. For R_{AC} this would be $\frac{p_A - p_C}{q_{An}}$. Taking this into account, we obtain for the blood and the cerebrospinal fluid circulation (units of resistance is $\frac{\text{mmHg}}{\text{ml/min}}$):

$$R_{ACn} = 0.125 \qquad R_{VSn} = 0.0166 \qquad R_{CFn} = 37.5 \tag{3.30}$$

$$R_{CVn} = 0.0166 \qquad\qquad R_{FSn} = 12.5. \tag{3.31}$$

Since the resistance R_{VS} is not constant but depending on the volume of compartment "V" – see equation (3.15) – we calculate the parameter $k_{R_{VS}}$ using the volume V_{Vn} to obtain $k_{R_{VS}} = 26.66$.

To determine the parameter $k_{R_{AC}}$ in equation (3.8) we first need the capacitance of the arteries since the capacitance determines the volume of the compartment. Following Ursino and Lodi [1997] the standard value for the capacitance is

$$C_{ABn} = 0.15\,\tfrac{\text{ml}}{\text{mmHg}} \qquad \begin{aligned} \Delta C_{AB1} &= 0.165\,\tfrac{\text{ml}}{\text{mmHg}} \\ \Delta C_{AB2} &= 0.065\,\tfrac{\text{ml}}{\text{mmHg}} \end{aligned} \tag{3.32}$$

while $\Delta C_{AB1/2}$ is determined by the upper (140 mmHg) and lower (70 mmHg) limit of the autoregulation curve. A good review on values of the pressure dependence of the autoregulation curve is given in Gao et al. [1998]. The dimensionless parameter of the strength of the autoregulation can vary from $\alpha = 1.5$ to $\alpha = 0.2$ for a full functional respectively an impaired autoregulation. The time constant of this process is assumed to be in the range of $\tau = 10\,\text{s}$ (in Giulioni and Ursino [1996] a value of 5 s and in Ursino and Lodi [1997] a value of 20 s is given).

We are using for the capacitances C_{VB} and C_B the following parameters as first proposed by Ursino [1988] and slightly modified by us for a more realistic behaviour:

$$k_V = 0.3\,\tfrac{1}{\text{ml}} \qquad\qquad k_B = 0.26\,\tfrac{1}{\text{ml}} \tag{3.33}$$

$$p_{V0} = 2.5\,\text{mmHg} \qquad\qquad p_{B0} = 2.5\,\text{mmHg}. \tag{3.34}$$

With this set of parameters (valid for a healthy patient) the behaviour of our model can now be validated. After having done this, the same model can then be used to reproduce measured data of patients from the intensive care unit.

3.2.5 Validation of the model

In the following, typical curves known in the medical literature are reproduced to show the applicability of the model and to explain the phenomena seen in the time series analysis in the previous chapter in more detail. Furthermore a discussion on the mechanism and measurements of the CSF production can now be supported by calculations of the model.

Autoregulation curve

One of the basic and most important mechanisms determining cerebral perfusion is the autoregulation. To reproduce this process, is a basic necessity of a hydrodynamical model of the brain. In figure 3.2 the measurements on a baboon are shown and can now be compared to the behaviour of our model (figure 3.4).

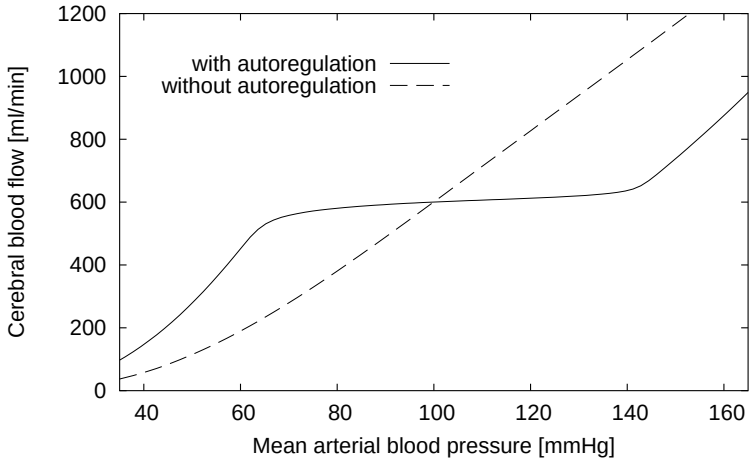

Figure 3.4: The autoregulation curve from the model shows the same behaviour as the measurements in figure 3.2. The increase of the cerebral blood flow above the critical blood pressure was not possible to measure in the experiment. Chosen parameters for the autoregulation (see equation (3.9)) are $\alpha = 1.5$ (with) and $\alpha = 0.2$ (without autoregulation).

A good agreement between the two curves can be noticed, the plateau is well established between 70 and 140 mmHg. Below an arterial blood pressure of 70 mmHg the blood flow shows a quasi-linear dependence. Above a critical blood pressure of 140 mmHg the arteries can not compensate the higher pressure, since they can't constrict further and therefore the blood flow above this pressure shows a linear increase due to the nearly constant resistance of the arteries.

In our model, we can also show, how the blood flow relates to the arterial blood pressure if the autoregulation is impaired. In this case the arteries are no longer active elements, their resistance is nearly constant, and therefore the blood flow increases quasi linearly with the pressure.

Global pressure-volume curve

An important curve showing the behaviour of the system with increase of extra volumes in the brain – for example due to a swelling – is the global pressure-volume curve. In particular the behaviour of the intracranial pressure is of interest, since it provides an opportunity to determine the overall compliance of the brain.

A good review on determining the pressure-volume curve experimentally is given in Sklar and Elashvili [1977], where constant flow injection and balloon inflation measurements are shown. The difference between these two measurement methods shows up at that point in the curve where the exponential relationship starts growing.

In figure 3.5 a balloon inflation measurement is shown. The volume can be increased by some amount, without any increase of the pressure. This originates from the possibility that some of the CSF is used to compensate this increase of volume by absorbing CSF into the blood at the sagittal sinus. Whereas in a flow injection measurement, the pressure would start growing immediately, since the "effective" volume increase is calculated, where the absorption of the CSF has already been taken into account.

In the literature mainly experiments with animals are presented. Data from human patients using the constant flow injection method can be found in Fridén and Ekstedt [1983] as well as similar measurements by Gaab et al. [1983].

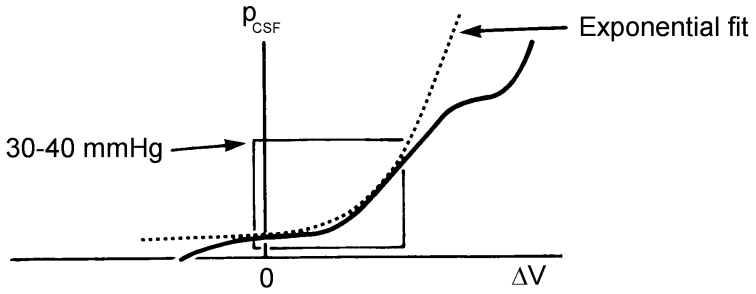

Figure 3.5: Global pressure-volume curve/compliance of the brain. The measurement of the global compliance is realized by increasing the volume (ΔV) of a balloon in the brain or by injecting and extracting fluid from the CSF compartment. An exponential function can be fitted for the increase of the intracranial pressure. The resting pressure is then given at $\Delta V = 0$ (adapted from Sullivan and Allison [1985]).

As mentioned before, the pressure-volume curve determines the compliances of the veins and the brain tissue. To check our model, a balloon inflation experiment was simulated. The volume of compartment "E" was steadily increased and the volume of this compartment was plotted against the intracranial pressure, showing the results in figure 3.6. Due to the increase of the compartment

"E", in particular the veins and the brain tissue are compressed. The arterial blood vessels have in contrast to that a higher internal blood pressure and active elements regulating their volume. Therefore their compliance doesn't play any role in the resulting curve.

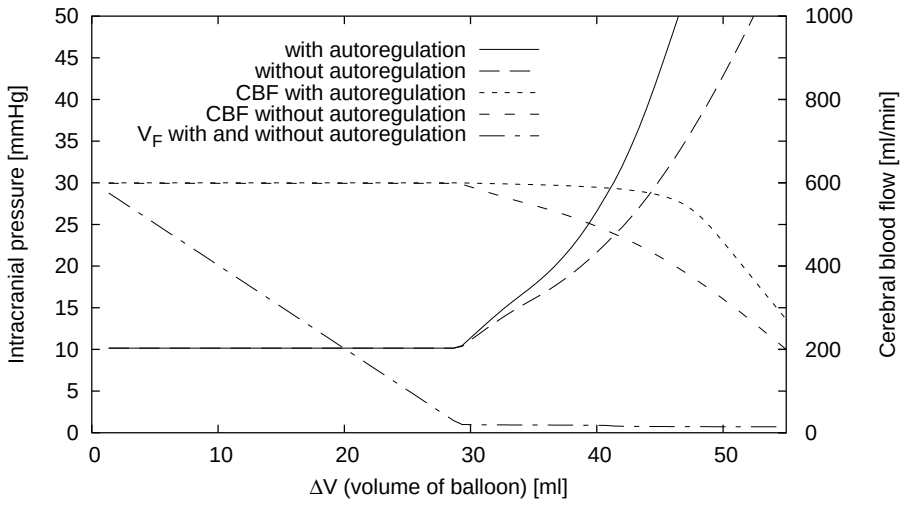

Figure 3.6: The global pressure-volume curve produced by the model. The exponential-type increase of the pressure (when the absorbable CSF volume V_F reaches zero) is consistent with figure 3.5. Note, a reduction of the increase of the intracranial pressure for *impaired* autoregulation can be observed accompanied by a decrease in CBF.

The curve clearly shows that the intracranial pressure only rises when no further cerebrospinal fluid can be absorbed ($V_F \sim 0$). This means that the CSF is used for the compensation of the expanding volume ΔV by absorbing it into the sagittal sinus.

Furthermore a difference in the increase of the ICP for different states of the autoregulation can be seen. The increase of the ICP is smaller for impaired autoregulation than for a healthy patient. In this case, the arteries are no longer acting against the external pressure and therefore they will also be compressed. As a result of this, the cerebral blood flow in the impaired autoregulation situation decreases (due to the smaller volume of the arteries and therefore higher resistance) in contrast to a healthy patient.

Cerebrospinal fluid production

From various clinical measurements the production of the cerebrospinal fluid is either assumed to be depending on the cerebral blood flow or to be stable over a wide range of the arterial blood pressure. In Sullivan and Allison [1985] and Hoffmann [1987] these two possibilities are stated. In this section, a short

explanation of the phenomena and the underlying process will be given by using results from simulations conducted with the model. The knowledge about the production is also important to explain the coupling seen between ABP and ICP.

For our model we used a pure pressure dependent production of the CSF, i.e. the pressure difference between the blood in the capillaries and the pressure of the CSF in compartment "F" is the driving force. Simulations with and without working autoregulation results in the following behaviour for the CBF and CSF-production (figure 3.7).

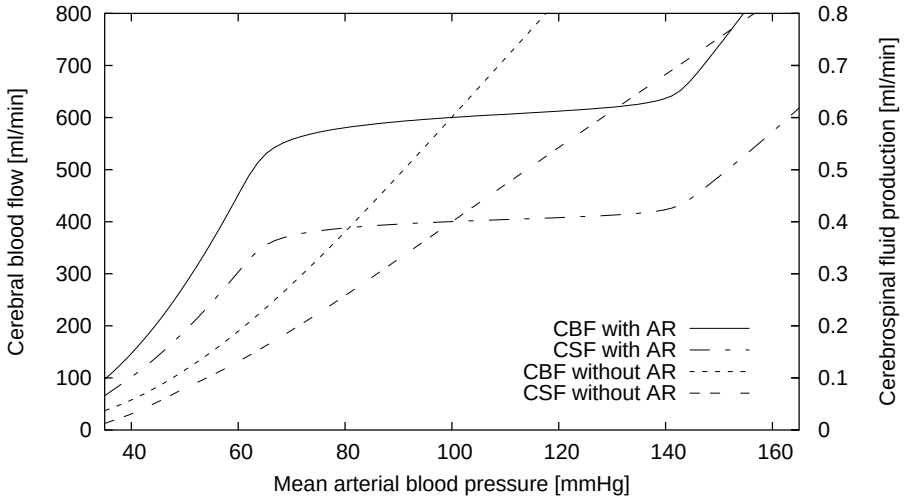

Figure 3.7: Dependence of the production of cerebrospinal fluid on the arterial blood pressure. For both states (with and without autoregulation) a good correlation between CBF and CSF-production can be observed, as documented by experiments. But the origin of this correlation is the linear dependence of the CSF production on the pressure difference between the capillary and the CSF compartment!

The plot suggests that the cerebral blood flow is the origin of the CSF production, since both curves show the same behaviour over the full range of the arterial blood pressure. But effectively a secondary process is seen, the primary process is the pressure depending CSF production. The autoregulation keeps the CBF as well as the pressure in the capillaries constant, therefore the CSF production is independent of the ABP. But if the autoregulation is impaired the pressure in the capillaries varies with the ABP, due to the constant resistance of the arteries, and therefore the CSF production grows nearly linearly with the ABP.

Arterial blood pressure - Intracranial pressure

In time series analysis, we have seen a coupling between the arterial blood pressure and the intracranial pressure. The question to be asked is, does our model

reproduce this coupling and if so, under which circumstances? Can we learn anything from the results of such simulations for the treatment of the patients?

We simulated three different situations describing different states of physical health. A patient with...

- active autoregulation and CSF to compensate volume changes ($V_F > 0$)
- impaired autoregulation and CSF to compensate volume changes ($V_F > 0$)
- impaired autoregulation and *no* CSF to compensate volume changes ($V_F = 0$)

To visualise a possible coupling, the intracranial pressure is plotted over the arterial blood pressure. The results of the simulations are shown in figure 3.8.

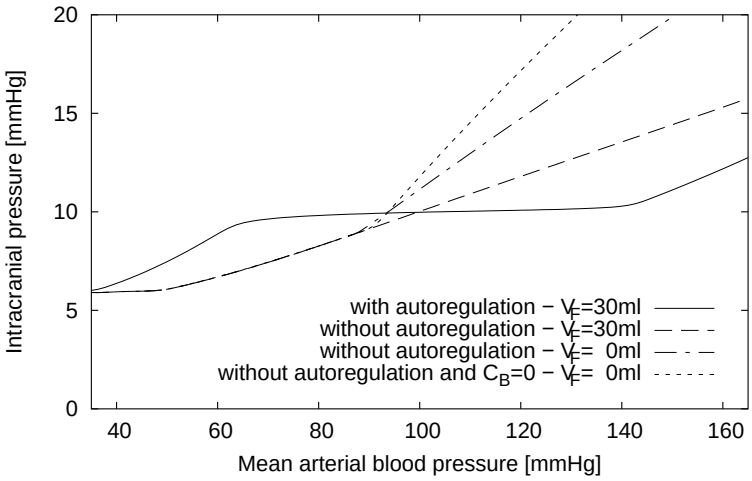

Figure 3.8: Relation between the intracranial pressure (ICP) and the arterial blood pressure (ABP) for different states of the autoregulation, the volume of the CSF fluid (V_F) and the capacitances of the brain tissue (C_B). No or only moderate increase of the ICP is seen for situations with fluid in the CSF compartment ($V_F = 30\,\text{ml}$). For impaired autoregulation and no compensation fluid in the CSF compartment, the increase of the ICP gets stronger especially when the brain tissue is not compressible anymore ($C_B = 0$).

As expected for a healthy patient, the intracranial pressure is independent of the ABP. But if the autoregulation is impaired we obtain a linear increase of the ICP. Both phenomena can be explained by the CSF production. As we have seen in the previous section, the CSF production is depending on the ABP, i.e. for impaired autoregulation a higher CSF production is expected for higher ABP. Due to the constant resistance of the absorption of the CSF, the pressure rises if more CSF is produced.

The situation changes clearly if the volume of the cerebrospinal fluid (V_F) is fully absorbed, i.e. no compensation fluid is available. This situation was

simulated for blood pressures above 100 mmHg. A sharp bend is seen in the curves showing a stronger increase of the ICP which can be increased even further by assuming the brain tissue to be stiff ($C_B = 0$). The increase is then as strong as the one seen in the time series analysis where the relation between ABP and ICP was 3:1 (see figure 2.8).

The main reason for the ICP rising is the increase of the volume of the arteries. Since the autoregulation is impaired, the arteries are passive elements and act like a rubber hose. The volume is then proportional to the ABP. This is equivalent to the phenomenon seen in the global pressure-volume curve. Instead of an expansion of compartment "E" – by a swelling or a balloon – the compartment "A" increases its volume due to the impaired autoregulation and results in a rise of the ICP.

Stability analysis of the system of differential equations

To analyse the stability of our system of 5 coupled non-linear differential equations, we calculate the eigenvalues of the Jacobi matrix at their fixed points. This gives of course only information about the linear stability as shown in Wiggins [1990]. Interestingly, all fixed points – for physiological reasonable parameters – are stable, i.e. all eigenvalues are negative and real. Furthermore the system has not only one fixed point, but a fixed point line. The location of the fixpoints on this line depends on the initial condition from which the simulation is started.

Oscillating, diverging or chaotic trajectories are not found which seems to be typical for biomedical systems. Still there could be a parameter range – probably outside physiologically reasonable parameters – which shows asymptotically no stable solutions.

3.3 Oxygen transport

As described briefly in the first chapter, the blood circulation is responsible for the transport of oxygen from the lungs to the cells. As will be seen later, the Krogh cylinder is a good description for the oxygen diffusion processes between the blood in the capillaries and the tissue as well as for the tissue itself. A good overview of the oxygen transport in the human body is given in the textbooks by Klinke and Silbernagl [2003] and Greger and Windhorst [1996], as well as in Schmidt et al. [2000].

The intention of this section is the calculation of the mean oxygen pressure in the tissue, which can be measured by a device like the Licox sensor. To be able to do so, all diffusion processes including the cerebral blood flow have to be taken into account. With such a model the relationship between the values of the Licox and Invos sensor will be shown. The cerebral blood flow is thereby the origin of the variations as was seen in time series analysis.

3.3.1 The blood

To supply the cells in the human body with oxygen, the blood first flows through the lungs, where due to diffusion processes – triggered by the pressure gradient between the oxygen in the air and in the blood – the blood is saturated with oxygen. The oxygen in the blood is mainly bound to the haemoglobin, which acts as an oxygen buffer. 98% of the oxygen molecules are buffered on the haemoglobin, only 2% of them are dissolved in the blood. The relation between the oxygen saturated haemoglobin (HbO_2) and the dissolved oxygen is described by the oxygen dissociation curve (ODC), which will be discussed in the following section.

Oxygen dissociation curve

The haemoglobin releases the bound oxygen into dissolved oxygen depending on the oxygen dissociation curve (ODC) as shown in figure 3.9.

From the steep slope between 20 and 50 mmHg of the S-shape curve it can be seen, how the haemoglobin acts as a buffer. Large variations of the saturation of the haemoglobin will be reflected only in small variations of the partial oxygen pressure (the dissolved oxygen). Therefore, the dissolved oxygen which is the only oxygen able for the diffusion into the tissue, is kept stable over a wide range.

First measurements on human blood including a formula for the ODC were presented by Hill [1910]. Later Adair [1925] proposed a formula with four constants reflecting the 4 Fe sites on the haemoglobin to which the O_2 molecules can bind. A good explanation for the four constants is given in the book of Crystal et al. [1991], where transition probabilities between the states of bound and unbound Fe sites are used.

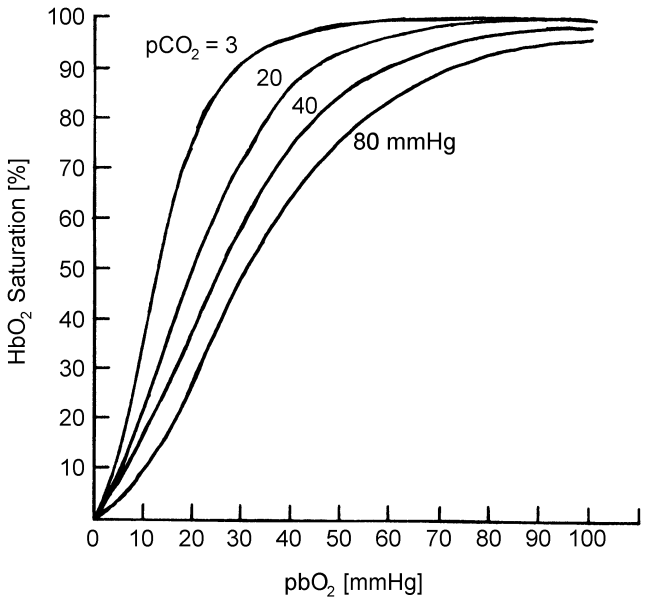

Figure 3.9: Oxygen dissociation curve of human blood. The dissolved oxygen respectively the partial pressure of oxygen in the blood (pbO_2) is associated with the saturation of haemoglobin with oxygen (HbO_2) in an S-shape curve. Furthermore, the dissociation curve depends on the carbon dioxide content (pCO_2) in the blood, where $pCO_2 = 40\,\text{mmHg}$ represents normal conditions (adapted from Reneau et al. [1969]).

For our purposes the equation describing the ODC proposed by Hill is sufficient, reading

$$\text{HbO}_2 \text{ saturation} = \text{SbO}_2 = \frac{\left(\frac{p}{p_{50}}\right)^n}{1 + \left(\frac{p}{p_{50}}\right)^n} \tag{3.35}$$

where SbO_2 is an abbreviation for the saturation of haemoglobin in the blood with O_2. p_{50} defines the oxygen pressure in the blood for a SbO_2 of 50%. Typical values for the equation of Hill, using the two constants n and p_{50}, were measured for normal conditions by Zwart et al. [1984] ($n = 2.5$, $p_{50} = 26\,\text{mmHg}$) and Severingshaus [1979] ($n = 2.6$, $p_{50} = 27.6\,\text{mmHg}$) as well as derived from the Adair formula with the coefficients given by Roughton and Severinghaus [1973] ($n = 2.7$, $p_{50} = 26.6\,\text{mmHg}$). For our model we will use the following parameters

$$n = 2.6 \quad \text{and} \quad p_{50} = 26\,\text{mmHg}. \tag{3.36}$$

The saturation of haemoglobin with oxygen in the arterial blood is up to 99–100%, i.e. the blood is fully saturated and could only transport more oxygen, if the partial pressure of (dissolved) oxygen in the blood is raised. In contrast to that, the venous blood shows a HbO_2 saturation of "only" 65%, which indicates still a good reserve of oxygen.

Total oxygen content in the blood

First we present a convention for the expressions of the amount of oxygen in the blood. An example would be pbO_2, which defines the partial oxygen pressure in the blood. Now, the first letter can be either

$$\begin{aligned} &\text{p} = \text{pressure [mmHg]},\\ &\text{S} = \text{saturation of haemoglobin with oxygen [\%] or}\\ &\text{C} = \text{content/amount of oxygen } [\tfrac{\text{cm}^3 \text{ of } O_2}{\text{cm}^3 \text{ of blood}}] \end{aligned}$$

and the second letter can be either

$$\text{b} = \text{blood}, \quad \text{a} = \text{arterial} \quad \text{or} \quad \text{v} = \text{venous}.$$

The rest defines the molecule which can be either O_2 or CO_2. The total amount of oxygen in the blood can be written as a sum of the two terms, the buffered and the dissolved oxygen

$$CbO_2 = \underbrace{\beta \cdot SbO_2}_{\sim 98\%} + \underbrace{\gamma \cdot pbO_2}_{\sim 2\%} \tag{3.37}$$

where $\beta = 0.201 \, \frac{\text{cm}^3 O_2}{\text{cm}^3}$ and $\gamma = 3 \cdot 10^{-5} \, \frac{\text{cm}^3 O_2}{\text{cm}^3 \cdot \text{mmHg}}$, as given in Schmidt et al. [2000] and other books. CbO_2 has therefore the unit $\frac{\text{cm}^3 \text{ of } O_2}{\text{cm}^3 \text{ of blood}}$. Using normal values for the saturation and the oxygen pressure, a ratio of 98% of bound oxygen to 2% of dissolved oxygen is obtained.

3.3.2 The Krogh cylinder

The oxygen supply of tissue was first described by Krogh [1918-1919] using a tissue cylinder around a capillary. In general it is assumed that the oxygen in the blood can reach the tissue only by diffusion. Since the surface to volume ratio in capillaries is the highest, due to their small diameter, here the largest amount of exchange of oxygen takes place. Therefore arteries and veins can be neglected for modelling the oxygen transport into the tissue. To get an impression of the Krogh cylinder an illustration is given in figure 3.10.

For simplicity, cylinder coordinates are used in the description of the Krogh cylinder. The z axis lies along the capillary and r defines the radius. At $r = r_c$ the capillary wall touches the tissue cylinder, whereas $r = r_t$ defines its overall radius. The length of the tissue cylinder is given by L.

For the description of the Krogh cylinder and the diffusion processes, the following assumptions are made:

i) General: The oxygen pressure gradient as the driving force of the diffusion process. In this work, only radial diffusion is taken into account. Calculations of the oxygen pressure in the brain tissue using radial or radial *and*

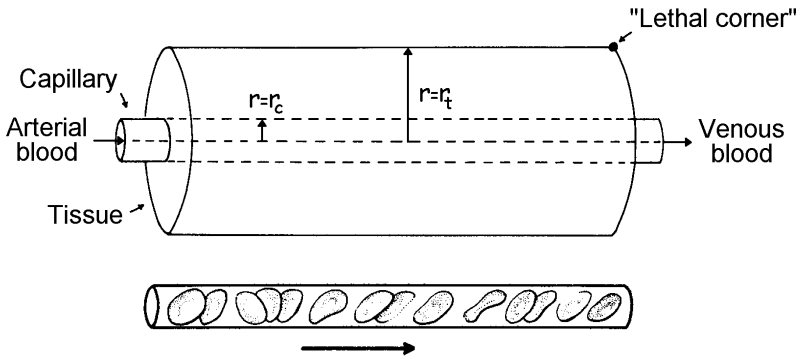

Figure 3.10: The Krogh cylinder representing a tissue cylinder around a capillary. Red blood cells flow through the capillary carrying the bound oxygen, which, after dissociation into the blood, diffuses into the tissue. r_c defines the capillary radius, r_t the radius of the tissue cylinder and "lethal corner" the area in the cylinder with the lowest oxygen content (adapted from Reneau et al. [1967]).

axial gradients were performed by Reneau et al. [1967]) and the results are shown in figure 3.11. Only small deviations are seen in the first part of the tissue cylinder. Since we are only interested in the mean oxygen pressure in the tissue cylinder, this deviation cancels out and the axial gradients can therefore be neglected.

ii) Capillary: The capillaries are straight pipes, running in parallel and having a unidirectional blood flow. They are arranged in a hexagonal lattice as shown in figure 3.12. The diameter is assumed to be constant and the velocity profile is uniform, since the red blood cells can only pass the capillary one by one as shown in the lower part of figure 3.10. Furthermore the blood flow is assumed to be constant over the whole capillary length. The capillary wall represents no resistance to the oxygen diffusion. The oxygen pressure in the capillary is assumed to be constant in r and φ.

iii) Tissue: The oxygen consumption in the tissue is assumed to be independent of the local oxygen pressure (zero-order reaction). The cells may be represented as a homogeneous volume distribution of sinks of oxygen, independent of time and position. Further the diffusion coefficient can be assumed to be homogeneous.

A way to relax this assumption (which is only necessary for pathological cases) is given at the end of this chapter.

Typical values for the radius of the capillary and the tissue cylinder as well as the length of the cylinder can be found in different books as shown in the

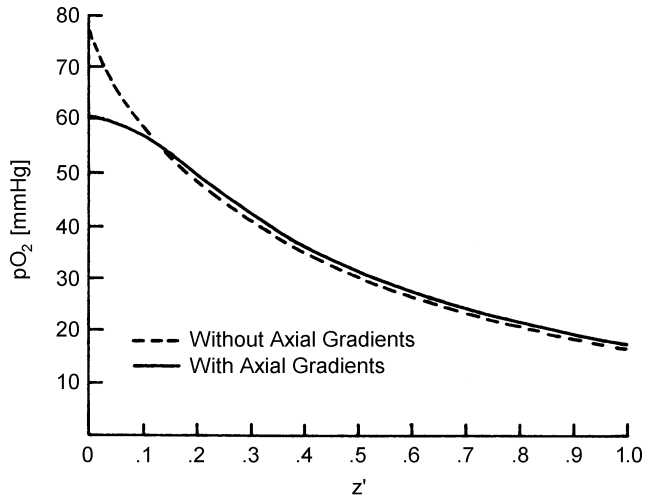

Figure 3.11: Oxygen pressure distribution in the tissue cylinder at maximum radius (r_t) along the capillary. Calculations were done with and without axial gradients in the oxygen diffusion equations. Only small variations at the arterial end of the capillary ($z' < 0.1$) are seen between the two calculations (adapted from Reneau et al. [1967]).

following. It is necessary to distinguish between grey and white matter, since they have different capillary densities.

The values for humans given in Bargmann [1977] for the radius of the capillaries are around $3.5\,\mu$m. The red blood cells, which are squeezed in the capillaries, have a diameter of $7.7\,\mu$m. For the capillary density of grey matter 1000–1100 capillaries/mm^2 and white matter 300 capillaries/mm^2 are typical. In Purves [1972] values of 800–870 capillaries/mm^2 are given for cats in grey matter and in whiter matter of 370 capillaries/mm^2.

The length of the capillaries respectively the tissue cylinder is hard to determine. Sharan and Popel [2002] used for their multi-compartment model values (observed in sheeps brain tissue) in the range of 600–1050 μm.

From medical observations we know that the Licox sensor is placed into the white matter of the brain tissue. Therefore we use the following parameters in our model

$$r_c = 3.5\,\mu\text{m} \quad \text{and} \quad r_t = 31\,\mu\text{m} \mathrel{\widehat{=}} 330\,\tfrac{\text{capillaries}}{\text{mm}^2} \quad \text{and} \quad L = 770\,\mu\text{m} \tag{3.38}$$

where r_c is the radius of the capillary, r_t the radius respectively L the length of the tissue cylinder. One should mention that the ratio of grey and white matter in the human brain was determined by Raff et al. [1994] as 57% to 43%. These values were obtained by using magnetic resonance imaging techniques. Nevertheless we can be sure, that the Licox sensor is placed in the white brain tissue.

Under normal conditions of the cerebral blood flow, we can assume for the

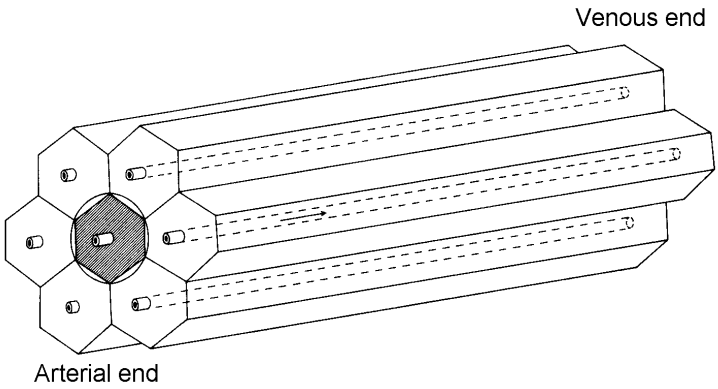

Figure 3.12: Theoretical arrangement of capillaries in the brain using a hexagonal lattice to guarantee an optimal supply of the tissue with oxygen (adapted from Reneau et al. [1967]).

velocity of the blood respectively the red blood cells typical values of

$$v = 400 \, \tfrac{\mu\mathrm{m}}{\mathrm{s}} \tag{3.39}$$

as mentioned in Mintun et al. [2001]. Further, we need for the tissue cylinder the parameters of the consumption, diffusivity and solubility of O_2. A good review of typical values in oxygen transport in blood and tissue is also given in Mintun et al. [2001], from where we take the following values

$$A = 4.5 \cdot 10^{-4} \, \tfrac{\mu\mathrm{m}^3\mathrm{O}_2}{\mu\mathrm{m}^3 \cdot \mathrm{s}} \quad \text{and} \quad D = 1.8 \cdot 10^{3} \, \tfrac{\mu\mathrm{m}^2}{s} \tag{3.40a}$$

$$\text{and} \quad c = 2.6 \cdot 10^{-5} \, \tfrac{\mu\mathrm{m}^3\mathrm{O}_2}{\mu\mathrm{m}^3 \cdot \mathrm{mmHg}} \tag{3.40b}$$

where A is the O_2 consumption, D the O_2 diffusivity and c the O_2 solubility. In Mintun et al. [2001] only a consumption value for grey matter ($8.2{\cdot}10^{-4} \, \tfrac{\mu\mathrm{m}^3\mathrm{O}_2}{\mu\mathrm{m}^3 \cdot \mathrm{s}}$) is given, but from nuclear magnetic resonance imaging measurements in An et al. [2001] the values for grey and white matter were determined as $9.2{\cdot}10^{-4}$ and $4.5{\cdot}10^{-4} \, \tfrac{\mu\mathrm{m}^3\mathrm{O}_2}{\mu\mathrm{m}^3 \cdot \mathrm{s}}$.

3.3.3 Theory

Our intention is to calculate the mean oxygen pressure in brain tissue ($\langle p_{ti}O_2 \rangle$). First of all, we will give a rough outline of the calculation before going into the details. The calculations can be divided into two parts, firstly the calculation of the oxygen pressure in the blood and secondly the diffusion of oxygen into and within the tissue.

The content of oxygen in the *blood* along the capillary ($CbO_2(z)$) can be calculated by assuming a constant metabolism in the tissue and using the velocity from the cerebral blood flow. As boundary condition at the arterial end of the capillary, the content $CbO_2(z{=}0)$ will be set to the oxygen content of the arterial blood CaO_2. For the diffusion processes, the partial pressure of free dissolved oxygen in the blood ($pbO_2(z)$) is needed. This can be obtained by solving the equation of $CbO_2(z)$ for $pbO_2(z)$. Unfortunately the pressure can't be derived analytically, therefore either an approximation has to be made (contribution of $pbO_2 \ll CbO_2$) or a nonlinear equation has to be solved numerically.

The oxygen pressure in the *tissue* will be calculated by using only the radial gradients as mentioned in the assumptions. For the boundary conditions we assume that the pressure is continuous at the interface between capillary and tissue and that the pressure gradient vanishes at the tissue surface ($r = r_t$). This can be satisfied by reasons of symmetry, since the flux through the surface of two neighbouring tissue cylinders should be zero (flux $\sim$ pressure gradient – Fick's first law). Therefore we obtain the pressure in the tissue as function of radius and position along the capillary ($p_{ti}O_2(r, z)$). Finally we have to integrate over the tissue volume to obtain a value for the mean oxygen pressure in the tissue (as measured by a device like the Licox sensor).

In the following a detailed derivation of the equations will be given.

Oxygen pressure in the blood

From equation (3.37) we obtain the amount of oxygen per blood volume (CbO_2). Using the equation of continuity, the CbO_2 along the capillary can be calculated. For simplicity we use $\rho{=}CbO_2$, i.e.

$$\frac{\partial \rho}{\partial t} + \mathrm{div} j = \alpha, \quad \text{where} \quad j = \rho \cdot v, \tag{3.41}$$

v is the velocity of the blood in the capillaries (depending on the CBF!) and α depicts the oxygen consumption from the surrounding tissue. The consumption can be calculated from the product of the volume of the tissue cylinder times the oxygen consumption of the tissue

$$\alpha = -\frac{\pi({r_t}^2 - {r_c}^2)}{\pi {r_c}^2} \cdot A \tag{3.42}$$

where A is the constant metabolism of the tissue in units of $\frac{\mu m^3 O_2}{\mu m^3 \cdot s}$. As a stationary solution ($\frac{\partial \rho}{\partial t} = 0$ and $\mathrm{div} j = \frac{dj}{dz}$) and the boundary condition of $CaO_2 = CbO_2(z{=}0)$, we obtain

$$CbO_2(z) = CaO_2 - \frac{{r_t}^2 - {r_c}^2}{{r_c}^2} \frac{A}{v} \cdot z \tag{3.43}$$

where z is the coordinate along the capillary in units of μm with the arterial end at $z = 0$. For the diffusion of oxygen into the tissue we need the dissolved oxygen (pbO_2) in the blood, which is given implicitly by equation (3.37). The equation consists of the hill equation plus the linear contribution of the dissolved oxygen

$$CbO_2(z) = \beta \cdot \frac{\left(\frac{p_c(z)}{p_{50}}\right)^n}{1 + \left(\frac{p_c(z)}{p_{50}}\right)^n} + \gamma \cdot p_c(z) \tag{3.44}$$

and has to solved for $p_c(z)$ – equivalent to pbO_2(z) – with the coefficients $\beta = 0.201 \frac{cm^3 O_2}{cm^3}$ and $\gamma = 3 \cdot 10^{-5} \frac{cm^3 O_2}{cm^3 \cdot mmHg}$. We obtain as solution an equation of the form

$$\frac{\gamma}{{p_{50}}^n} \cdot p_c^{n+1}(z) - \frac{CbO_2(z) - \beta}{{p_{50}}^n} \cdot p_c^n(z) + \gamma \cdot p_c(z) - CbO_2(z) = 0 \tag{3.45}$$

where $CbO_2(z)$ has to be replaced by equation (3.43). The solution for $p_c(z)$ of this equation will give the correct *oxygen pressure in the blood* (pbO_2(z)), but unfortunately there is no general analytic solution for arbitrary n – not even for $n = 3$.

Two possibilities exist to obtain a value for the pressure of the dissolved oxygen in the blood (pbO_2(z)): *Either* we solve the equation numerically, but this will lead to a stepwise numerical integration of the final integral for the mean oxygen pressure in the tissue *or* we use an approximation ignoring the last part of equation (3.44), since for normal physiological conditions the contribution is only 2%, leading to an analytical solution. Using the approximation, we then only have to deal with the saturation of the haemoglobin (SbO_2) and not with the full oxygen content of the blood (CbO_2). This simplifies the calculation to

$$SbO_2(z) = \frac{\left(\frac{p_c(z)}{p_{50}}\right)^n}{1 + \left(\frac{p_c(z)}{p_{50}}\right)^n} \quad \text{solved for } p_c\text{:} \quad p_c(z) = p_{50} \cdot \left(\frac{SbO_2(z)}{1 - SbO_2(z)}\right)^{\frac{1}{n}} \tag{3.46}$$

with SbO_2 depending on the position z along the capillary

$$SbO_2(z) = SaO_2 - \frac{{r_t}^2 - {r_c}^2}{{r_c}^2} \frac{A}{v} \cdot z \tag{3.47}$$

which is similar to equation (3.43).

Oxygen diffusion in the tissue

Calculating the diffusion from the capillary wall into and within the tissue we use Fick's second law. The derivation is based on definition of the flux as $j = -D\nabla p$ from the oxygen pressure field and applying the mass conservation, i.e. the mass balance for diffusive O_2 transport and O_2 consumption: $\nabla \cdot j + \alpha = 0$. For the time depending equation, we write

$$\frac{\partial p}{\partial t} = D \cdot \Delta p - \frac{A}{c} \tag{3.48}$$

with A the metabolism of the tissue, c the oxygen solubility coefficient and D the oxygen diffusivity in the tissue. Due to the geometry of the model, we use cylindrical coordinates. Since we know that we can ignore axial diffusion, i.e. along the z coordinate, and can assume the tissue to be homogeneous, i.e. no gradients depending on φ, we write Fick's second law as

$$\frac{\partial p}{\partial t} = D \cdot \left(\frac{\partial^2 p}{\partial r^2} + \frac{1}{r}\frac{\partial p}{\partial r} \right) - \frac{A}{c}. \tag{3.49}$$

To solve this equation, we use the following boundary conditions as they were mentioned in the introduction

$$p = p_c \;\text{ at }\; r = r_c \quad \text{and} \quad \frac{\partial p}{\partial r} = 0 \;\text{ at }\; r = r_t \tag{3.50}$$

where p_c is the oxygen pressure in the capillary obtained either from equation (3.45) or (3.46). Assuming further a steady state, we obtain the solution for the oxygen pressure in the tissue

$$p_{ti}(r, z) = p_c(z) - \frac{A}{2Dc}\left({r_t}^2 \ln \frac{r}{r_c} + \frac{1}{2}({r_c}^2 - r^2) \right) \tag{3.51}$$

as was already shown in Krogh [1918-1919], where a Danish friend of him, the mathematician Mr-. K. Erlang, derived this solution. Finally, to calculate the mean partial pressure of oxygen in the tissue we have to integrate over the tissue cylinder

$$\bar{p}_{ti} = \langle p_{ti}(r, z) \rangle_{V_{\text{tissue}}} = \frac{1}{V_{\text{tissue}}} \int_0^L \int_0^{2\pi} \int_{r_c}^{r_t} p_{ti}(r, z) \cdot r \, \mathrm{d}r \, \mathrm{d}\varphi \, \mathrm{d}z \tag{3.52}$$

with $V_{\text{tissue}} = \pi({r_t}^2 - {r_c}^2)L$. This is the general form of the mean oxygen pressure in the tissue. Note, our expression of $p_{ti}(r, z)$ is only valid if the pressure is positive at any point in the tissue cylinder. For normal physiological conditions, as we will assume for the following calculations, a positive pressure in the tissue can be assumed. Therefore the constant extraction of O_2 (independent of z), as

given in equation (3.42), is still valid. The integral can then be solved partially and the mean oxygen pressure in the tissue can be written as

$$\bar{p}_{ti} = \frac{1}{L}\int_0^L p_c(z)\,\mathrm{d}z - \frac{A}{8Dc}\cdot\frac{-{r_c}^4 + 4{r_c}^2{r_t}^2 - 3{r_t}^4 + 4{r_t}^4\ln\frac{r_t}{r_c}}{{r_t}^2 - {r_c}^2}. \tag{3.53}$$

But the integral over $p_c(z)$ has still to be solved separately. Either we use equation (3.45), but then the nonlinear equation has to be solved in every integration step, or we use the approximation (ignoring the dissolved oxygen in the calculation of the total amount of oxygen in the blood) given in equation (3.46) and (3.47). Using this approximation, we then obtain

$$\bar{p}_{ti} = \frac{p_{50}}{L}\int_0^L \left(\frac{a-bz}{1-(a-bz)}\right)^{\frac{1}{n}}\mathrm{d}z - \frac{A}{8Dc}\cdot\frac{-{r_c}^4 + 4{r_c}^2{r_t}^2 - 3{r_t}^4 + 4{r_t}^4\ln\frac{r_t}{r_c}}{{r_t}^2 - {r_c}^2} \tag{3.54a}$$

$$\text{where}\quad a = \mathrm{SaO_2}\quad\text{and}\quad b = \frac{{r_t}^2 - {r_c}^2}{{r_c}^2}\frac{A}{v}. \tag{3.54b}$$

The integral has unfortunately no closed solution for arbitrary n, but a solution involving hypergeometric functions can be given and solved numerically.

3.4 Validation: Theory ↔ Experimental data

In the following a validation of the theoretical description of the oxygen diffusion processes in the human brain as well as a validation of the combined hydrodynamical and oxygen supply model will be given. We will compare the theoretical prediction from the model with the experimentally measured data from patients at the intensive care unit. Finally we will address the problem of the to high absolute values of the theoretical predictions and give a suggestion for a solution.

Oxygen supply model: Licox ↔ Invos

In the time series analysis chapter, the question of the functional relationship between the measured values of the Licox and Invos sensors was raised. With the here presented theoretical framework we are able to give an answer to this question and to proof the validity of it by a comparison of the predictions with the measured data.

A plot showing the relationship between the pressure of oxygen in the tissue (Licox) and the saturation of haemoglobin with oxygen (Invos) is given in figure 3.13. Due to the long term trends in the data sets, only the variations are shown to enable a comparison with the theoretical predictions. From a 48 hours recording, only those time sections where no external influence on the patient took place, were used. This resulted in 30 hours of undisturbed data.

From the medical point of view, the variations in the measured values of both sensors are mainly due to the changes of the cerebral blood flow. To be able to compare the theoretical predictions with the measured data, we calculate the theoretical values for Licox and Invos for varying cerebral blood flow using equation (3.53). Thereby the velocity in the capillaries is assumed to be proportional to the cerebral blood flow. The variable used for the theoretical Licox value is $\bar{p}_{ti}$, whereas for the Invos sensor a ratio of 1:3 of the arterial and venous saturated blood (SaO_2 and SvO_2) is used as was stated in equation (1.3). Overlaying the measured data with the theoretical curve results in the plot shown in figure 3.13.

At first sight, the relation between Licox and Invos obtained from the theoretical calculation shows a nearly linear correlation for the variation of the blood velocity. The measured signals are not exactly following the theoretical description – which was expected for the description of such a complex system – but the main behaviour is well described. One should propably notice, that the resolution of the Invos sensors is given as $\pm 1\%$. Therefore only for the larger variations (where the influence of the measurement inaccuracy is small) it can be expected that the measured data follows the theoretical predictions. The linear behaviour of the prediction explains now, why a good correlation was seen between the two sensor signals in the (linear) time series analysis.

Before taking the rest of the 48 hours time series into account, we should mention the air tube cleaning manoeuvre performed by nurses on the patient.

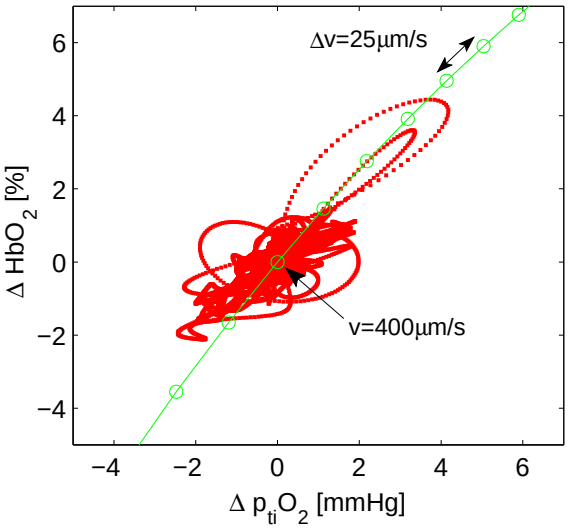

Figure 3.13: Relation between Licox ($p_{ti}O_2$) and Invos (HbO_2). Shown is the data from a patients recording (red dots) compared to the theoretical prediction (green line). The variations are assumed to depend on the variability of the cerebral blood flow respectively the velocity of the blood in the capillaries. Note: The plotted data originates from times with no external influence on the patient (30 h of 48 h recording).

Due to the artificial ventilation of the patients, the air tube fills itself with a mucus (a slime) leading to a smaller tube radius. The tube has therefore to be cleaned regularly. During the cleaning, the patient has no artificial ventilation, which could lead to an oxygen undersupply. To overcome this problem, the patient is supplied (a few minutes before the manoeuvre) with air containing 100% oxygen.

In figure 3.14 the remaining 18 hours of the data are plotted as grey dots, while still showing the 30 hours data (red dots) from the previous figure 3.13. The blue line is a theoretically calculated curve based on the variation of the arterial oxygen pressure paO_2. The velocity of the blood was kept constant at 400 μm/s. The paO_2 values – corresponding to the circles on the line – are increased quadratically from 50 mmHg to 700 mmHg.

We can clearly see a strong increase of $p_{ti}O_2$ when paO_2 is increased above 100 mmHg. In contrast to the $p_{ti}O_2$ value, the HbO_2 value saturates slowly. This can be described by the higher dissolved oxygen content in the blood. The haemoglobin saturation is not strongly increased by the higher paO_2 (due to the nonlinearity of the ODC), but more dissolved oxygen is available for diffusion into the tissue. This results automatically in a higher oxygen content in the tissue and therefore higher values from the Licox sensor. The behaviour of the curves shown by the grey dots – mostly originating from the air tube cleaning manoeuvre – are thus described and understood using the variation of the arterial oxygen pressure.

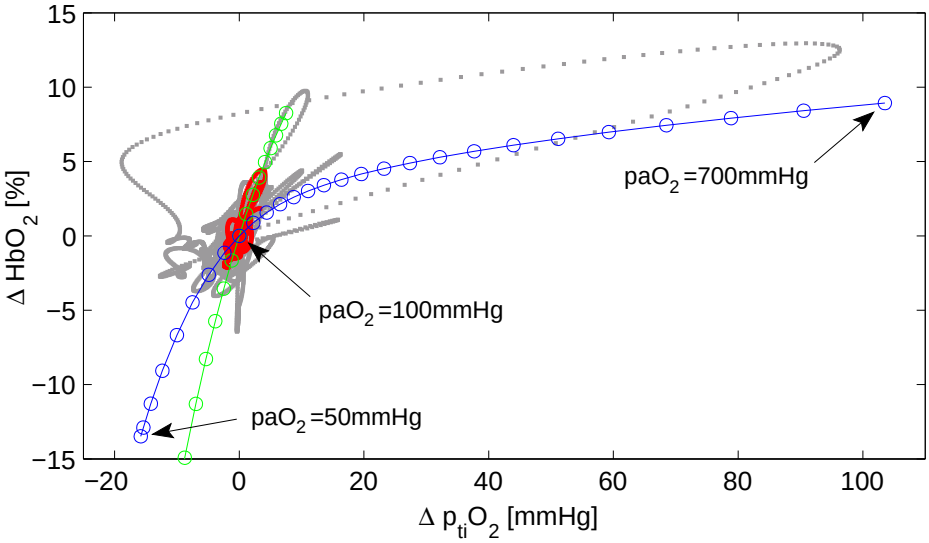

Figure 3.14: Relation between Licox ($p_{ti}O_2$) and Invos (HbO_2). The plotted points shown here originate from a 48 h recording, where the red points are the data from figure 3.13 and the gray ones are the remaining 18 h. They are compared to two theoretical curves: green the variation of the cerebral blood flow (see figure 3.13) and blue the variation of the oxygen pressure in the arterial blood (paO_2). The high Licox values in the recording can be explained by an increase of the arterial oxygen pressure up to 700 mmHg.

Combined model: ABP ↔ Oxygen supply

For a comparison of the predictions of the combined hydrodynamical and metabolic model with the measured data, we refer once again to the observed correlations seen in the time series analysis. In some cases, a coupling between the mean arterial blood pressure and the oxygen content in the blood was detected.

As already discussed in section 2.3.3, we assume an impaired autoregulation to be a reason for this kind of coupling. In such a case, the cerebral blood flow (CBF) is nearly linear dependend on the arterial blood pressure (ABP). This would have as consequence that for higher ABP more oxygen per time is transported to the brain, due to the higher CBF, and finally a higher oxygen content in the blood and in the brain tissue is measured.

The question is now, how does the relation between the ABP and the oxygen supply look like? A nonlinear relationship could be expected since the two parameters are coupled with each other via four processes: ABP ↔ CBF ↔ velocity of the blood ↔ diffusion of the oxygen from the blood into the tissue ↔ diffusion of the oxygen in the tissue.

In figure 3.15 a prediction by the model for the relationship between the mean arterial blood pressure and the values for the Licox and Invos sensor in the case of an impaired autoregulation is presented.

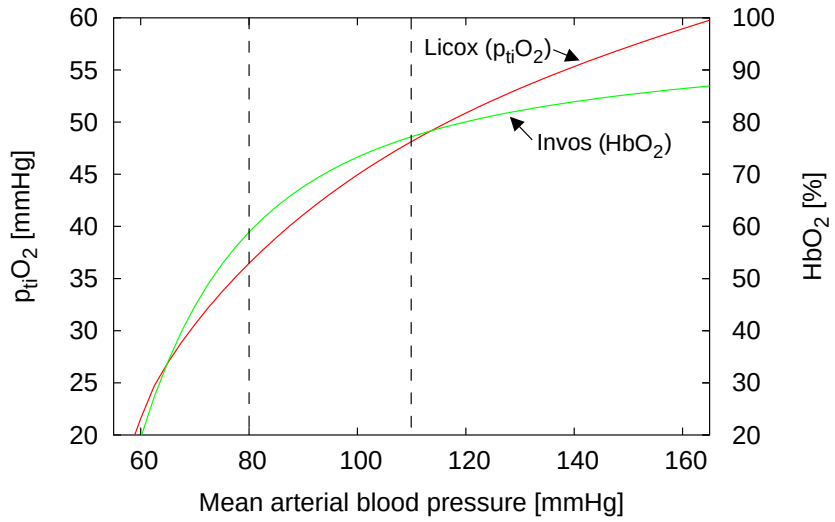

Figure 3.15: Predictions from the combined model for the relation between the mean ABP and Licox ($p_{ti}O_2$) respectively Invos (HbO_2) in the case of an impaired autoregulation. A nonlinear relationship is observed for both sensors. But in the typical ABP range of 80 to 110 mmHg, in particular the Licox values correlates nearly linear with the ABP.

In this figure, two interesting observations can be made. First the global behaviour of both values ($p_{ti}O_2$ and HbO_2) over the full range of the mean arterial blood pressure shows a nonlinear relationship. In particular the predicted values for the Invos sensor show a steep slope for low ABP values and a strong saturation for high ABP values. But in the typically observed ABP range of 80 to 110 mmHg, both predicted curves can be well approximated by a linear relationship. In particular the Licox values follow nicely a straight line.

This linear relationship explains now the well established linear correlation found in time series analysis, making us confident that the combined model describes both metabolic and hydrodynamical processes correctly. To support this claim, we will compare the 21 hours recording used for the correlation analysis in figure 2.7 with the theoretical predictions. In figure 3.16 the variations of the recorded Licox values are plotted versus the mean arterial blood pressure (red dots), after removing trends and noise from the data as described in chapter 2.2. Further the theoretical prediction from the model for the case of an impaired autoregulation is shown (green line).

During this 21 hours recording, the patient is assumed to be in the same state of health, in particular having an impaired autoregulation. Of course external influences to the patient will be reflected in the data. Nevertheless the measured data follows mainly the prediction, even when sometimes larger deviations are seen. Furthermore inaccuracies from the Licox sensor or a wrong drift removal induce large deviations.

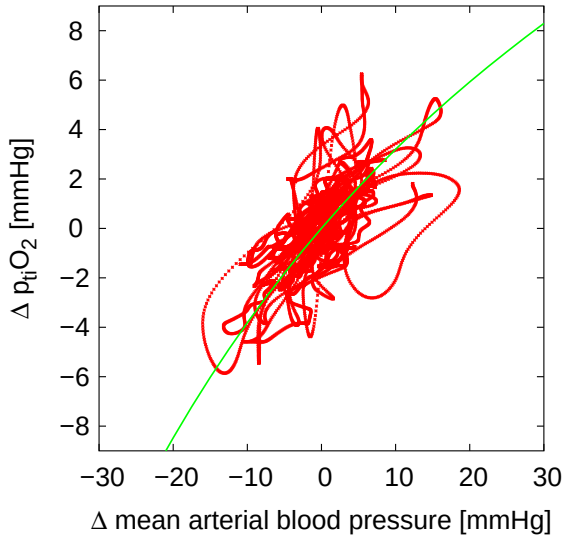

Figure 3.16: Comparison of the theoretical predictions (green line) for the relation between the mean arterial blood pressure (ABP) and Licox ($p_{ti}O_2$) in the case of an impaired autoregulation with the data from a patient's recording (red dots). The plotted values correspond to the 21 hours recording which was already used in figure 2.7. The measured values follow mainly the theoretical curve. Deviations are probably due to the inaccuracy of the Licox sensor as well as external influences on the patient.

The presented comparison is a first "proof" of the validity of the *combined* model which use just physically based descriptions of the processes in the brain. This model is just a first step of the description of the complex phenomena in the human brain. More extensive investigations are necessary, but a promissing foundation is laid.

Taking the true capillary distribution into account

In the plots, only the *variations* of the values from Licox and Invos were shown. A comparison of the absolute values of the measurements with the theoretical predictions shows that the theoretical values are slightly to high, in particular the mean Licox value (measured 25, theoretical 45 mmHg). The values for the Invos sensor lie closer to each other, instead of 63 in the measurements, a value of 72 is gained from the theoretical predictions. These deviations are due to two reasons:

On the one hand, the measurements are from patients on the intensive care unit, while the parameters for the theoretical predictions are only valid for a healthy patient. Furthermore, it has to be made sure that the sensors are placed not to close to an injured tissue area.

On the other hand, a perfect arrangement of the capillaries in a hexagonal

lattice is assumed. This is of course not true for natural systems. The distance distribution of the nearest neighbours will be widely spread and we don't have (as assumed here) a delta peak at fixed r_t. In the book of Hoofd [1992] it was shown – using a large lattice on which the diffusion equation was solved – that a randomly distributed arrangement of capillaries shows a decrease of the oxygen pressures in the tissue. Since in the equations given in this work, the radius of the tissue cylinder r_t can be varied, the overall mean oxygen pressure for *true* natural systems can be calculated by integrating over all possible tissue radii

$$\langle \bar{p}_{ti}(r_t) \rangle_{\text{all possible tissue radii}} = \int_0^\infty \varphi(r) \cdot \bar{p}_{ti}(r_t = r) \cdot r^2 \cdot L \, \mathrm{d}r \tag{3.55}$$

where $\varphi(r)$ is the probability to find the nearest neighbour capillary at distance r.

Note: Attention should be paid to the fact that the oxygen pressure in the tissue ($p_{ti}(r, z)$) can get negative for large tissue cylinders r_t. A negative partial oxygen pressure is of course non-physical and therefore the formulas/boundary conditions have to be adapted. This can be done by choosing the tissue radius r_t in equation (3.50) so that it fulfils the conditions

$$p_{ti} = 0 \quad \text{and} \quad \frac{\partial p}{\partial r} = 0 \;\text{ at }\; r = r_t^* \tag{3.56}$$

as it was mentioned briefly in Reneau et al. [1970]. If the radius of the tissue cylinder r_t is larger than r_t^*, the oxygen pressure in the tissue (p_{ti}) is zero, i.e. $p_{ti}(r > r_t^*) = 0$.

The two boundary conditions are getting clear when we think about, which conditions must be fulfilled at the point, where the oxygen pressure reaches zero. Since the oxygen pressure reaches zero ($p_{ti} = 0$ at $r = r_t^*$), the flux throughout this point is zero (no oxygen molecules will diffuse through out this point). But this is equivalent to a vanishing pressure gradient ($\frac{\partial p}{\partial r} = 0$ at $r = r_t^*$).

First preliminary tests showed, that the calculation with this modification to the formulas results in a much better agreement with the measured values. The overall mean oxygen pressure in the tissue reaches now values of 30 mmHg instead of 45 mmHg in the previous calculations.

3.5 Conclusions

In this chapter we have shown that mathematical models of the haemodynamic and metabolic processes are useful for a deeper understanding of complex processes occurring in the human brain in normal and pathophysiological states.

A hydrodynamical compartment model was established by a pure physical description and adapted to the special requirements of the neurosurgical intensive medicine. The special properties of the individual compartments, as for example the autoregulation, the cerebrospinal fluid (CSF) production and the compression of the veins, were described. From the equation of continuity, a system of 5 coupled nonlinear differential equations plus one constrain was derived. We were able to reproduce the important interdependencies known from medical literature as for example the autoregulation or the pressure-volume curve. Furthermore, the observation of the (indirect) cerebral blood flow dependent production of the CSF could be explained, by showing that the observations are only secondary processes. The primary (underlying) processes is the pressure depending production of the CSF. Finally, the main advantage of this model is the possibility to estimate the cerebral blood flow continuously in contrast to already existing imaging techniques.

Further, an oxygen supply model was derived using the theory of the Krogh cylinder. With this relatively simple model consisting of a capillary surrounded by a tissue cylinder, we are able to describe the oxygen supply processes in the brain. As the basic mechanism, a pressure gradient driven diffusion was used to determine the oxygen content in brain matter. Interestingly, the theoretical predictions for the values of Licox and Invos show a nearly linear correlation for varying blood velocity. The nonlinear oxygen diffusion curve as well as the diffusion processes only play a role, when huge changes to the physiological conditions takes place, as for example in the air tube cleaning manoeuvre. We could show how these large oscillations in the Licox measurements could be described by a variation of the partial oxygen pressure in the arterial blood.

Finally both models were combined to simulate the measurements obtained by recordings from the intensive care unit. Thereby, the cerebral blood flow estimated by the hydrodynamical model acts as the link to the oxygen supply model. By means of the combined model the results of the correlation analysis could be confirmed and the measurements quantitatively be reproduced. Therefore, a better insight into the underlying processes generating these correlations was achieved by this model.

Chapter 4

Independent component analysis

Independent component analysis (ICA) is a multimodal time series analysis technique to find statistical independent signals within a given data set.

A typical example for the application of ICA is the cocktail party problem, a so-called blind source separation (BSS) problem. Imagine a group of n people standing around chatting with each other. This mixture of signals is recorded by n microphones. The goal is to extract from the mixture of speech signals the voices of the speakers (the sources) without knowing the sources and the mixture process (therefore called blind). Assuming the voices to be independent – which is a valid assumption – the method of ICA can be applied and a corresponding algorithm is then able to recover, just by using the mixed signals, the voice of each speaker.

ICA seems not only to be able to find the correct solution in the cocktail party problem, but also to reveal the essential structures of the data in many (real world) applications. The assumption of the independence of the signals appears to be a very strong property and in particular useful when applying it to the analysis of highly multimodal data sets.

In this chapter the derivation of two new algorithms using a geometric approach and an information theoretical one including time structures is presented. Furthermore the application to biomedical data sets – in particular to EEG data – is shown while the limitations of the method are also discussed.

4.1 Introduction

Independent component analysis is a signal processing tool to decompose *observed signals* $\mathbf{x}(t) \in \mathbb{R}^m$ into a set of statistically independent signals $\mathbf{y}(t) \in \mathbb{R}^m$, which are called the *independent components*. Statistical independence has to be seen in the mathematical sense of stochastically/statistically independent probability

densities, i.e. the joint probability density factorises into the marginal densities

$$p(\mathbf{y}) = \prod_{i=1}^{n} p(y_i) \tag{4.1}$$

where $\mathbf{y}$ is an n-dimensional random variable with y_i its ith component. The idea of decomposing signals into independent signals was first expressed by Jutten and Herault [1991] in the framework of blind separation of sources, while the term "ICA" was later coined by Comon [1994] in his fundamental paper on the theory of linear ICA.

However the field became quickly popular with the seminal paper by Bell and Sejnowski [1995] who elaborated upon the Infomax-principle, first introduced by Linsker [1992]. Later Amari [1998] introduced the concept of a natural gradient simplifying the Infomax learning rule. Among others, the most efficient ICA algorithm developed in the following year was the FastICA-algorithm by Hyvärinen [1999] using negentropy and kurtosis as a contrast function. Recently, geometric ICA algorithms first proposed by Puntonet et al. [1995], which are based on neural learning clustering algorithms, have received further attention due to their relative ease of implementation.

For a good introduction to ICA together with extensions to various cases and applications of ICA we refer to one of the following text-books: Hyvärinen et al. [2001], Cichocki and Amari [2002] and Lee [1998].

The blind source separation problem

The most prominent application of ICA is the blind source separation problem. There an underlying mixture model generates the observed signals $\mathbf{x}(t)$ from the sources $\mathbf{s}(t)$ by $\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t))$, where $\mathbf{f}(\cdot)$ can be a (non-)linear function mapping from $\mathbb{R}^n \rightarrow \mathbb{R}^m$. For the linear case, which is commonly investigated and on which will be focused in this work, the observed signals are generated by

$$\mathbf{x}(t) = \mathbf{A} \cdot \mathbf{s}(t) \tag{4.2}$$

where $\mathbf{s}(t)$ are the n-dimensional independent sources, $\mathbf{x}(t)$ depicts the m dimensional mixed/measured signals and $\mathbf{A}$ is the $m \times n$ mixing matrix. The goal of the BSS problem is to find the unknown sources $\mathbf{s}(t)$ and the unknown transformation matrix $\mathbf{A}$.

An ICA algorithm is trying to find as independent signals as possible from the mixture of signals $\mathbf{x}(t)$ by using – in the linear case – an unmixing matrix $\mathbf{W}$, so that the independent components can be written as $\mathbf{y}(t) = \mathbf{W} \cdot \mathbf{x}(t)$. Often only the quadratic case is investigated, where the number of sources equals the number of observation ($m = n$).

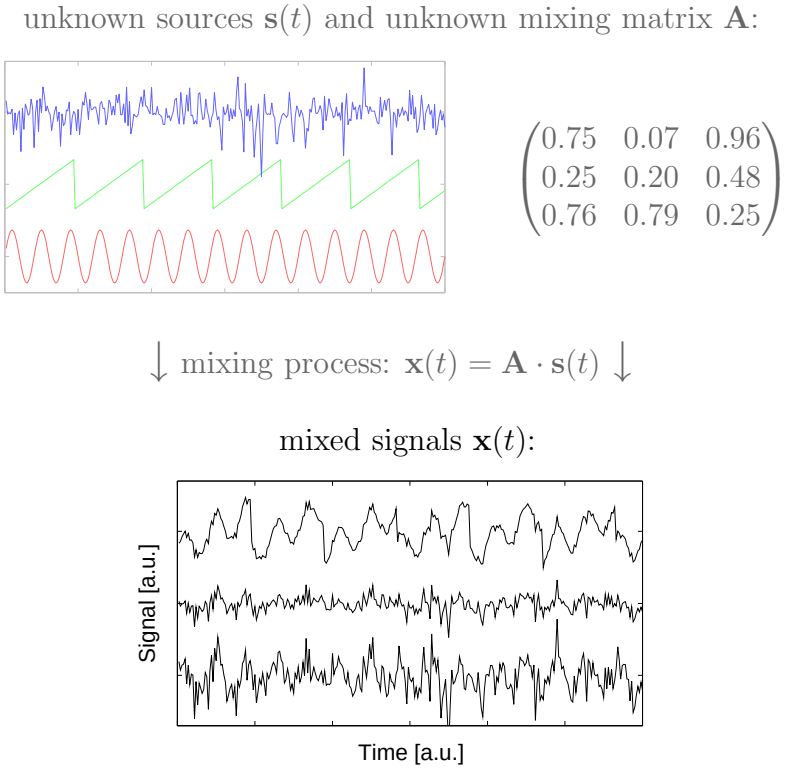

Figure 4.1: Application of the independent component analysis (ICA) to the blind source separation (BSS) problem. The sources $\mathbf{s}(t)$ as well as the mixing process (here the matrix) are unknown – depicted by the light grey colour. Only the mixed signals $\mathbf{x}(t)$ are available, hence the problem is called blind source separation. Goal of ICA is to extract independent components (ICs) from a set of mixed signals. Since the sources $\mathbf{s}(t)$ are independent, the extracted ICs correspond (up to permutation and scaling) to the sources, while the ICA algorithm also recovers the mixing matrix $\mathbf{A} \sim \mathbf{W}^{-1}$.

The reason for the restriction to the quadratic case is that it can be proofen (as done by Comon [1994]) that the assumption of independent sources is strong enough to solve the (quadratic) BSS problem correctly up to an indeterminacy of scaling and permutation of the recovered sources. A further restriction, which will get obvious in the following theory section, is that maximally one source may have a probability function equal to a gaussian density. Typically this proof from Comon is called the proof of the identifiability of quadratic linear ICA.

For BSS problems with more signals than sources ($m > n$), the problem can easily be reduced by a principal component analysis to the quadratic case, since no additional information is gained by further linear combinations of the sources as shown in great detail in the book of Hyvärinen et al. [2001]. In contrast to that, the BSS problem with more sources than measured signals ($n > m$) has – without any additional knowledge – no unique solutions and is referred to as overcomplete ICA, since the basis (the sources) represents an overcomplete basis for the measured signals. First approaches to algorithms solving overcomplete ICA problems were presented by Lewicki and Sejnowski [1998]. Later Theis et al. [2003a] presented a procedure using geometric considerations.

In the quadratic case, the recovered mixing matrix $\mathbf{W}^{-1}$ is said to be similar to $\mathbf{A}$, when the recovered sources differ to the original sources only by scaling and permutation. The unmixed signals respectively the independent components can then be written as

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{W} \cdot \mathbf{x}(t) \\ &= \mathbf{W} \cdot \mathbf{A} \cdot \mathbf{s}(t) \\ &= \mathbf{S} \cdot \mathbf{P} \cdot \mathbf{s}(t) \end{aligned} \tag{4.3}$$

where $\mathbf{S}$ is a scaling matrix with coefficients only on the diagonal and $\mathbf{P}$ is the permutation matrix. Both matrices are for the quadratic case of the size $n \times n$. For an illustration of the application of ICA to a BSS problem see figure 4.1.

ICA can be applied in many other areas. The cocktail party problem is a nice and intuitive example, which shows the basics of speech processing. Beside this more artificial application, the method is also of great interest for remote sensing, image recovery and biomedical applications. In this chapter, we will show how the ICA method can be applied to biomedical data sets and where the limitations of this method are.

Before presenting the applications, a short introductory section will make the reader familiar with the notation and the basic concepts used in probability and information theory. Subsequently, two new algorithms are derived using two very different approaches. Firstly a geometrically supported algorithm using scatter plots and their properties under transformation and secondly an approach using information theory including the knowledge of time structures in the data, as it is used in the framework of time series analysis.

4.2 Basic theory

In this section a short review on the basic probability- and information-theoretical definitions is given to make the reader familiar with the notation used in this work. The introduction follows mainly the theory chapters of Hyvärinen et al. [2001] using the same notation.

4.2.1 Probability theory

For a more detailed introduction to probability theory the book by Papoulis [1991] is recommended, which includes also the proofs of statements given in these sections.

Random variables and probability distributions

In this work we are assuming random variables as continuously-valued unless stated otherwise. The *cumulative distribution function* (cdf) F_x of a random variable at point $x = x_0$ is defined as the probability that $x \leq x_0$

$$F_x(x_0) = P(x \leq x_0) \tag{4.4}$$

where x_0 can take values between $-\infty$ to ∞, defining the whole cumulative distribution function for all values of x. The cdf for continuous random variables is a nonnegative, nondecreasing continuous function whose values lie in the range $0 \leq F_x(x) \leq 1$. From equation (4.4) it follows directly that $F_x(-\infty) = 0$ and $F_x(+\infty) = 1$.

The *probability density function* (pdf) of a random variable x can then be defined as the derivative of its cumulative distribution function

$$p_x(x_0) = \left.\frac{\mathrm{d}F_x(x)}{\mathrm{d}x}\right|_{x=x_0}. \tag{4.5}$$

The normalisation of the probability density function follows directly from the definition of the cdf

$$\int_{-\infty}^{+\infty} p_x(x)\,\mathrm{d}x = 1. \tag{4.6}$$

The concept of probability distributions can easily be generalised to random vectors. Assuming that $\mathbf{x}$ is an n-dimensional random vector $\mathbf{x} = (x_1, x_2, ..., x_n)^T$, where the components $x_1, x_2, ..., x_n$ are continuous random variables. T denotes the transpose of the vector. We will deal in this work only with column vectors unless stated otherwise. The same definition for the cdf and pdf can be formulated for random vectors as shown in Papoulis [1991] and Hyvärinen et al. [2001].

A special set of probability densities are the conditional densities. They are necessary to answer the question: "What is the probability density of a random vector $\mathbf{x}$ given another random vector $\mathbf{y}$?" Assuming that the joint density

$p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})$ of $\mathbf{x}$ and $\mathbf{y}$ and the (marginal) density $p_{\mathbf{y}}(\mathbf{y})$ exist, the *conditional probability density of* $\mathbf{x}$ *given* $\mathbf{y}$ is defined as

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}) = \frac{p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})}{p_{\mathbf{y}}(\mathbf{y})}. \tag{4.7}$$

Expectations and moments

Often the expectations of some functions of a random variable are of interest for the analysis and the processing of data. A great advantage of expectations is the possibility to directly derive them from the data, although they are formally defined in terms of the probability density function.

Let $\mathbf{g}(\mathbf{x})$ be any quantity of the random variable, which can be either a scalar, a vector or a matrix. The *expectation of* $\mathbf{g}(\mathbf{x})$ is denoted by $E\{\mathbf{g}(\mathbf{x})\}$ and is defined as

$$E\{\mathbf{g}(\mathbf{x})\} = \int_{-\infty}^{+\infty} \mathbf{g}(\mathbf{x})\, p_{\mathbf{x}}(\mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{4.8}$$

The integration operation is applied separately to every component of the vector or every element of the matrix resulting in another vector or matrix of the same size.

Usually the probability density of a random vector is unknown, but often a set of N samples $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$ from $\mathbf{x}$ is available, as for example in the case of data measured in real world applications. The expectation can then be estimated by averaging over the samples using

$$E\{\mathbf{g}(\mathbf{x})\} \simeq \frac{1}{N} \sum_{i=1}^{N} \mathbf{g}(\mathbf{x}_i). \tag{4.9}$$

If $g(x)$ is of the form x^n, i.e. products of the random variable, we end up at the nth *moment*, denoted by α_n and defined by

$$\alpha_n = E\{x^n\} = \int_{-\infty}^{+\infty} x^n\, p_x(x)\, \mathrm{d}x. \tag{4.10}$$

Often the *central moments* μ_n are more useful, which are computed around the mean ($m_x = \alpha_1 =$ 1st moment) of x

$$\mu_n = E\{(x - m_x)^n\} = \int_{-\infty}^{+\infty} (x - m_x)^n\, p_x(x)\, \mathrm{d}x. \tag{4.11}$$

From the definition it follows directly that the central moments $\mu_0 = 1$, due to normalisation, and $\mu_1 = 0$, due to the removal of the mean, are insignificant.

But the higher central moments, used in the so-called higher order statistics, can contain valuable information. In the literature the following variables are defined

$$\sigma^2 = \mu_2 \qquad \text{the } \textit{variance} \text{ of } x, \tag{4.12}$$

$$\gamma = \frac{\mu_3}{\sigma^3} \qquad \text{the } \textit{skewness} \text{ of } x, \tag{4.13}$$

$$\kappa = \frac{\mu_4}{\sigma^4} - 3 \qquad \text{the } \textit{kurtosis} \text{ of } x. \tag{4.14}$$

In words, the variance gives an estimate of the width of the distribution, the skewness of the asymmetry of the distribution and the kurtosis an estimate of the deviation from a gaussian distribution. As we will see later, the gaussian distribution has very special properties, its kurtosis is zero and due to its symmetry it has also zero skewness. Therefore it is often used as a reference. In one dimension one can write

$$p_x(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{x^2}{2\sigma^2}}. \tag{4.15}$$

Correlation and independence

The *correlation* between ith and jth component of a random vector $\mathbf{x}$ is denoted by r_{ij} and given by

$$r_{ij} = E\{x_i x_j\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_i\, x_j\, p_{x_i,x_j}(x_i, x_j)\, \mathrm{d}x_j\, \mathrm{d}x_i \tag{4.16}$$

where the correlation can be positive or negative. Is the correlation equal to zero, we speak of *uncorrelated random variables*. For the calculation of all correlations between the components of the random vector $\mathbf{x}$ we can write a $n \times n$ matrix, the *correlation matrix*

$$\mathbf{R}_\mathbf{x} = E\{\mathbf{x}\mathbf{x}^T\}. \tag{4.17}$$

Using the central moments we can also define a sort of correlation matrix, the so-called *covariance matrix*, where the mean of the random vector is removed prior to the calculation

$$\mathbf{C}_\mathbf{x} = E\{(\mathbf{x} - \mathbf{m}_\mathbf{x})(\mathbf{x} - \mathbf{m}_\mathbf{x})^T\}. \tag{4.18}$$

If the mean $\mathbf{m}_\mathbf{x}$ is zero, the correlation and covariance matrices become equal. The expectation operation can be extended to functions $\mathbf{g}(\mathbf{x}, \mathbf{y})$ of two different random vectors $\mathbf{x}$ and $\mathbf{y}$ in terms of their joint probability density

$$E\{\mathbf{g}(\mathbf{x}, \mathbf{y})\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbf{g}(\mathbf{x}, \mathbf{y})\, p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})\, \mathrm{d}\mathbf{y}\, \mathrm{d}\mathbf{x} \tag{4.19}$$

where the integrals are computed over all components of the random vectors. In analogy to the correlation and covariance matrix definitions, we can define the so-called *cross-correlation* and *cross-variance matrix* of $\mathbf{x}$ and $\mathbf{y}$ as

$$\mathbf{R_{xy}} = E\{\mathbf{x}\mathbf{y}^T\} \quad \text{and} \quad \mathbf{C_{xy}} = E\{(\mathbf{x} - \mathbf{m_x})(\mathbf{y} - \mathbf{m_y})^T\}. \tag{4.20}$$

Two random vectors $\mathbf{x}$ and $\mathbf{y}$ are called *uncorrelated* if their cross-covariance matrix $\mathbf{C_{xy}}$ is a zero matrix $\mathbf{C_{xy}} = \mathbf{0}$.

The mathematical definition of independence of two random variable x and y (as it is used in ICA) is defined by the factorisation of the joint probability distribution $p_{x,y}(x, y)$ into the product of their marginal densities $p_x(x)$ and $p_y(y)$, i.e. we write

$$p_{x,y}(x, y) = p_x(x) \cdot p_y(y). \tag{4.21}$$

More general the components of a random vector $\mathbf{x}$ are *independent*, if

$$p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i). \tag{4.22}$$

It follows directly from this definition that independent variables satisfy the basic property

$$E\{g(x)h(y)\} = E\{g(x)\} \cdot E\{h(y)\} \tag{4.23}$$

where $g(x)$ and $h(y)$ are any absolutely integrable functions of x and y, respectively. This reveals that statistical independence is a much stronger property than assuming uncorrelated variables. For uncorrelated variables, $g(x)$ and $h(y)$ are just linear functions, taking only the so-called second-order statistics into account.

However, if the random variables have gaussian distributions and are uncorrelated, then they are also independent. This is a very special property of gaussian distributions. It gets obvious when we write down the n-dimensional probability density of a gaussian random vector $\mathbf{x}$

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbf{C_x}}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m_x})^T \mathbf{C_x}^{-1}(\mathbf{x}-\mathbf{m_x})}. \tag{4.24}$$

Assuming that $\mathbf{x}$ is uncorrelated, it follows directly that $\mathbf{C_x}$ is a diagonal matrix and further the inverse $\mathbf{C_x}^{-1}$ is of a diagonal form. This results in a sum in the exponent with which the probability density $p_{\mathbf{x}}(\mathbf{x})$ factorises. But the factorisation is the criterion of independence. Therefore uncorrelated gaussian random vectors are independent.

Furthermore it becomes obvious, why independent component analysis can not find the correct (de)mixing matrix, if more than one source is gaussian distributed. Assume $\mathbf{x}$ is an independent gaussian distributed random vector with zero mean ($\mathbf{m_x} = 0$) and its probability density equals equations (4.24). Then,

under every transformation of an orthogonal matrix $\mathbf{O}$ so that $\mathbf{y} = \mathbf{O}\mathbf{x}$, the probability density of $\mathbf{y}$ does not change. Using the property $\mathbf{O}^T = \mathbf{O}^{-1}$ of the orthogonal matrix shows directly the equality of the probability densities ($p(\mathbf{y}) = p(\mathbf{x})$). Therefore ICA can only find the (de)mixing matrix of a mixture of gaussian distributed random variables up to an orthogonal matrix.

Markov chains

A class of random processes that is rich enough to capture a large variety of temporal dependencies is the class of finite state Markov processes. In the Markov chain model, the current state of the process depends on a certain number of previous values of the process. The number depicts the order of the process.

Such a process is characterised by its order m, and by a conditional probability matrix $\mathbf{P}_m$ whose rows may be interpreted as probability functions according to which the next random variable X_i is generated when the process is in a state $x_{i-1}, x_{i-2}, \ldots, x_{i-m}$

$$P(X_i = x_i | X_{i-1} = x_{i-1}, \ldots, X_{i-m} = x_{i-m}) = \mathbf{P}_m(x_i | x_{i-1}, \ldots, x_{i-m}). \quad (4.25)$$

A process X_i is a Markov chain of order m, if the conditional probability

$$P(X_i = x_i | X_{i-1} = x_{i-1}, \ldots, X_1 = x_1) = P(X_i = x_i | X_{i-1} = x_{i-1}, \ldots, X_{i-m} = x_{i-m}) \quad (4.26)$$

is independent of x_{i-h} for all $h > m$.

4.2.2 Information theory

An alternative approach to characterise random variables – instead of using moments – is information theory. Here the emphasis lies on coding and code length. Finding a suitable code depends on the statistical properties of the data and therefore it is another measure for the distribution of the random variable.

Entropy and mutual information

The basic concept of information theory is *entropy* which is defined for a discrete-valued random variable X as

$$H(X) = -\sum_i P(X = a_i) \log P(X = a_i) \quad (4.27)$$

where the a_i are the possible values of X. The object entropy can be seen as a measure for the uncertainty or randomness of a random variable. In the context of (information) coding it describes roughly the minimum necessary code length to transmit most efficiently a large number of observations of X.

For a continuous random variable x a similar object can be defined, the so-called *differential entropy*

$$H(x) = -\int p_x(x) \log p_x(x) \, \mathrm{d}x \tag{4.28}$$

which can be seen as a generalisation of the entropy of discrete random variables by a discretisation of the probability distribution. But a straight forward reduction of the discretisation size would lead to a diverging quantity in the formula ($log\Delta x$ diverges for $\Delta x \to 0$), which has therefore to be ignored. This leads to the above definition of the differential entropy.

Still, the differential entropy can be interpreted as a measure of randomness in the same way as entropy. In the following we will leave out the term "differential" since it gets obvious in the calculations which definition is used to calculate the quantity entropy. A generalisation to random vectors is straight forward.

The transformation of the (differential) entropy can be calculated for an invertible transformation $\mathbf{y} = \mathbf{f}(\mathbf{x})$ of the random vector $\mathbf{x}$. We obtain

$$H(\mathbf{y}) = H(\mathbf{x}) + E\{\log |\det J\mathbf{f}(\mathbf{x})|\} \tag{4.29}$$

where $J\mathbf{f}(\mathbf{x})$ denotes the Jacobian matrix of the function $\mathbf{f}$, i.e. the matrix of the partial derivatives of $\mathbf{f}$ at point $\mathbf{x}$. For the special case of linear transformations ($\mathbf{y} = \mathbf{W}\mathbf{x}$) we obtain for the last term of the above equation $\log |\det \mathbf{W}|$. This shows that the (differential) entropy is not scale-invariant.

A quantity measuring the amount of common information between two random variables is the *mutual information*. It is defined by using the sum of the following entropies

$$I(\mathbf{x}) = \sum_i H(x_i) - H(\mathbf{x}). \tag{4.30}$$

In the framework of the entropies, the mutual information can be interpreted as the difference between the sum of the minimum code length for each random variable and the code length for the full random vector. If the random variables are all independent of each other, a coding of the full random vector would not reduce the code length, since there is no common information, which can be used to reduce the size. With these arguments we can see, that the mutual information is nonnegative and only zero for independent variables.

Inserting the definition of the (differential) entropy we obtain a different formulation of the mutual information

$$I(\mathbf{x}) = \int p_{\mathbf{x}}(\mathbf{x}) \log \frac{p_{\mathbf{x}}(\mathbf{x})}{\prod_i p_i(x_i)} \, \mathrm{d}\mathbf{x} \tag{4.31}$$

where $p_i(\cdot)$ are the marginal densities of the x_i. This is exactly the Kullback-Leibler divergence which measures a kind of distances between two probability

densities $p_{\mathbf{x}}(\mathbf{x})$ and $\prod_i p_i(x_i)$. It has the property of being nonnegative and only zero if both probability densities in the logarithm are equivalent, which is equivalent to the independence of the random variables of the vector.

From the definition it follows directly, that the mutual information is invariant under the permutation of the arguments $I(x, y) = I(y, x)$.

Maximum entropy and negentropy

A natural question arises when defining a measure like the entropy: which random variable has the highest entropy or in other words, the highest randomness for a fixed variance? For continuous valued random variables the gaussian distribution fulfils this criterion.

This maximal entropy property for gaussian random variables shows that entropy can be used to define a measure of nongaussianity. Every other probability distribution has a lower entropy. Therefore we can define a measure called negentropy of the following form

$$J(\mathbf{x}) = H(\mathbf{x}^G) - H(\mathbf{x}) \tag{4.32}$$

where $\mathbf{x}^G$ is a gaussian random vector with the same covariance matrix as $\mathbf{x}$. From the definition and the maximum entropy property of gaussian random variables it follows that the negentropy is nonnegative and only zero if the distribution of $\mathbf{x}$ is gaussian.

An interesting and useful property is the invariance of the negentropy under an invertible linear transformation: $J(\mathbf{y}) = J(\mathbf{M}\mathbf{x}) = ... = J(\mathbf{x})$. We refer for the derivations of these and the following statements to Hyvärinen et al. [2001].

Furthermore, we can write the mutual information in terms of the negentropy as it will be derived in the following. Using the definition of the mutual information in terms of the entropies (equation (4.30)) and letting $\mathbf{y} = \mathbf{W}\mathbf{x}$ with $\mathbf{W} \in \mathrm{Mat}(n \times n, \mathbb{R})$ of full rank, $E\{\mathbf{x}\} = 0$ and $\mathbf{W}$ so that the variance of $\mathbf{y}$ is one ($E\{\mathbf{y}\mathbf{y}^T\} = \mathbf{I}$), we can write

$$I(\mathbf{y}) = \sum_i H(y_i) - H(\mathbf{y}) \tag{4.33}$$

$$= \sum_i H(y_i) - H(\mathbf{x}) - \log|\det \mathbf{W}|. \tag{4.34}$$

Adding and subtracting the term $\sum_i H(y_i^G)$ with y_i^G being a gaussian random variable with the same variance as $\mathbf{y}$, we obtain

$$I(\mathbf{y}) = -\sum_i H(y_i^G) + \sum_i H(y_i) + \sum_i H(y_i^G) - H(\mathbf{x}) - \log|\det \mathbf{W}|. \tag{4.35}$$

Since $\sum_i H(y_i^G)$ and $H(\mathbf{x})$ are independent of $\mathbf{W}$, both terms give only a constant contribution. The last term $\log|\det \mathbf{W}|$ is also constant since

$$E\{\mathbf{y}\mathbf{y}^T\} = E\{\mathbf{W}\mathbf{x}\mathbf{x}^T\mathbf{W}^T\} = \mathbf{W}E\{\mathbf{x}\mathbf{x}^T\}\mathbf{W}^T = \mathbf{I} \tag{4.36}$$

where $\mathbf{I}$ is the identity matrix and therefore we can write

$$\det \mathbf{I} = \det(\mathbf{W} E\{\mathbf{x}\mathbf{x}^T\}\mathbf{W}^T) = \det \mathbf{W} \det E\{\mathbf{x}\mathbf{x}^T\} \det \mathbf{W}^T = 1 \tag{4.37}$$

which leads to $\det \mathbf{W} = \text{const}$ since $\det \mathbf{W} = \det \mathbf{W}^T$ and $\det E\{\mathbf{x}\mathbf{x}^T\}$ is independent of $\mathbf{W}$. Thus, the mutual information can be written in terms of the negentropy as

$$I(\mathbf{y}) = -\sum_i J(y_i) + \text{const}\,. \tag{4.38}$$

The calculation of the entropy from a given set of observations of a random variable is non trivial since for the estimation a large number of observations is necessary. An approximation can reduce the necessary number of observations. For the negentropy, we can expand the probability density function of a random variable x around the standardised gaussian density, not unlike a Taylor expansion using moments or cumulants (the so-called Gram-Charlier expansion). Doing so for a one-dimensional random variable, we obtain as an approximation for the negentropy

$$J(x) \approx \frac{1}{12} E\{x^3\}^2 + \frac{1}{48}(E\{x^4\} - 3)^2 \tag{4.39}$$

corresponding to a sum of the skewness and kurtosis of x, assuming the variance of x to be one ($E\{x^2\} = 1$). The disadvantage of such an approximation is the lag of robustness to outliers. A term of x^4 is very sensible to observations far from the mean of the distribution and can therefore destroy a correct estimation.

Using entropy approximations based on an approximative maximum entropy method, we can overcome such problems. Instead of using the functions x^3 and x^4 two new functions G^1 (odd) and G^2 (even) are used, which do not grow too fast and therefore are more robust to outliers. For the new approximation of the negentropy we then obtain

$$J(x) \approx k_1 E\{G^1(x)\}^2 + k_2 \left(E\{G^2(x)\} - E\{G^2(\nu)\}\right)^2 \tag{4.40}$$

where k_1 and k_2 are positive costants and ν is a standardised gaussian random variable with zero mean. This is a generalisation of equation (4.39), since for $G^1(x) = x^3$ and $G^2(x) = x^4$ we end up with the expansion of Gram-Charlier.

4.3 A geometric approach

Geometric algorithms for linear independent component analysis represent a type of ICA algorithms that is easy to comprehend and simple to implement. The geometric approach to ICA has been proposed first by Puntonet et al. [1995] in order to separate linear mixtures using a kind of neural algorithm.

We reconsider geometric ICA in a theoretic framework showing that fixed points of (neural) geometric ICA fulfil a so-called geometric convergence condition (GCC), which the mixed images of the unit vectors satisfy, too. This leads to a conjecture claiming that in the nongaussian unimodal symmetric case there is only one stable fixed point, thus demonstrating uniqueness of geometric ICA after convergence.

Guided by the principles of neural geometric ICA, a new approach to linear geometric ICA based on histograms (FastGeo) is presented observing a considerable improvement in separation quality of different distributions and a sizable reduction in computational cost by a factor of 100 compared to the neural learning approach.

Furthermore we explore the accuracy of the algorithm and compare the performance of the geometric algorithms with classical ICA algorithms based on information theory. Finally we discuss the problem of high dimensional datasets within the realm of geometrical ICA algorithms.

The theoretical results on geometric ICA have been published in Theis et al. [2001] and Theis et al. [2003b], where also the FastGeo algorithm by Jung et al. [2001] has been presented and studied.

The notation

In the following mathematical derivations it will be important to distinguish between the random variables themselves and their realizations. Therefore we denote $S : \Omega \to \mathbb{R}^n$ as a n-dimensional random vector, where Ω is a fixed probability space. The samples of a random vector S are denoted by $\mathbf{s}_i \in R^n$ or if they originate from a time series by $\mathbf{s}(t)$. The mixing process can therefore be written either in terms of the random vectors $X = \mathbf{A} \circ S$ or by using the samples directly $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ where $\mathbf{A} \in \mathrm{Mat}(n \times n, \mathbb{R})$ is a mixing matrix of full rank.

4.3.1 Geometric considerations

We will explain the principles of the geometric algorithm using a simple two-dimensional blind sources separation model $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$.

Assuming S to be an independent two-dimensional random vector with zero mean and a peaklike probability distribution (kurtosis > 0) and let $\mathbf{s}_1, ..., \mathbf{s}_N \in \mathbb{R}^2$ be independent identically distributed (i.i.d.) samples of S. The whole set of samples can be visualised in a scatter plot, where every point in the plot

corresponds to a given sample $\mathbf{s}_i$ – see left illustration in figure 4.2. In this so-called source space, the areas with larger probability density p_S have a higher density of samples than those with a lower p_S value. Therefore the scatter plot can be used to visualise the density of S.

Let

$$\mathbf{A} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \tag{4.41}$$

be the mixture matrix in this example, then the samples $\mathbf{x}(t) = \mathbf{A}s(t)$ in the mixture space can be visualised in the same way – see right illustration in figure 4.2.

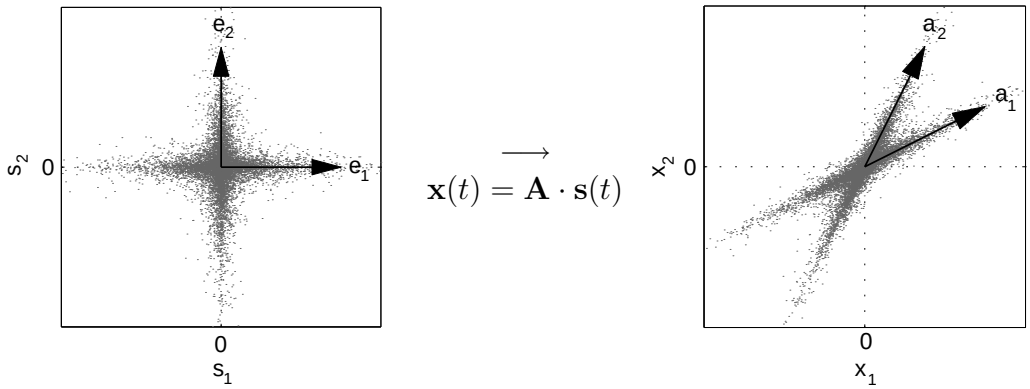

Figure 4.2: Schematic example of a two-dimensional scatter plot for the source $\mathbf{s}(t)$ (left), which have identical peaklike distributions, and their mixtures after the transformation $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ (right). Every point in the scatter plot corresponds to a given sample $\mathbf{s}(t)$, where $s_{1/2}$ is the first/second component of the vector. The unit vectors $\mathbf{e}_{1/2}$ are mapped onto the vectors $\mathbf{a}_{1/2}$ corresponding to the columns of the mixing matrix A.

The unit vectors $\mathbf{e}_{1/2}$ are transformed under the mixture matrix $\mathbf{A}$ to $\mathbf{a}_{1/2} = \mathbf{A} \cdot \mathbf{e}_{1/2}$ which corresponds exactly to the columns of the matrix $\mathbf{A}$. In order to find $\mathbf{A}$, we have to find the "principle" axis defined by $\mathbf{a}_1$ and $\mathbf{a}_2$ in the mixture space – indeed we only need lines and not the points themselves, so we have to recover $\mathbf{a}_1$ and $\mathbf{a}_2$ up to scaling and permutation which corresponds exactly to the indeterminacy of the BSS problem.

The recovery can be achieved by projecting the samples $\mathbf{x}_i$ onto a unit sphere and find the clusters using a Kohonen-like clustering algorithm which moves iteratively so-called neurons (w_1, w_1', w_2, w_2') on the circle into the direction of the chosen sample. Figure 4.3 shows the starting and end configuration of the neurons in the two-dimensional example and demonstrates visually that the fixed points of the algorithm correspond to the lines defined by $\mathbf{a}_1$ and $\mathbf{a}_2$.

Recapitulating we can say that the basic idea of the geometric separation method lies in the fact that in the source space $\{\mathbf{s}_1, \ldots, \mathbf{s}_N\} \subset \mathbb{R}^n$, where $\mathbf{s}_i$ represent a fixed number of samples of the source vector S with zero mean,

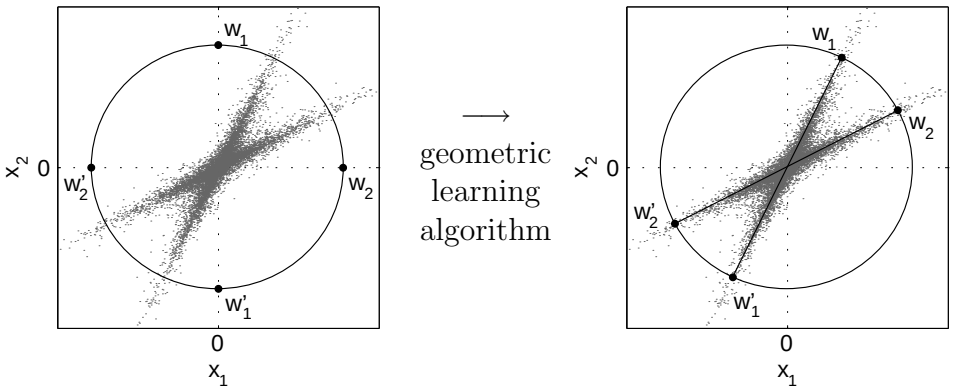

Figure 4.3: The basic idea of the geometric learning algorithm in two dimensions: 4 neurons are initially located at the crossings of the sphere with the axes of the coordinate system (left). Then iteratively, a winner neuron defined by a minimum distance to the chosen sample vector projected onto the sphere, is moved towards the sample by a given amount. This process continues until a convergence is reached and the axes fall together with the main distribution axes.

the data cluster along the axes of the coordinate system. After applying the transformation (matrix $\mathbf{A}$), the data cluster along the transformed coordinate axes. The detection of these n new axes allows the determination of a demixing matrix $\mathbf{B}$ with an inverse that is equivalent to $\mathbf{A}$.

The detection of these lines can either be done with a (neural) geometric learning algorithm – as briefly mentioned here and described in the next section in more detail – or by an algorithm based on histograms using the knowledge from the theoretical fundaments of the (neural) geometric learning algorithm, the so-called FastGeo algorithm.

A precise description how to recover the matrix $\mathbf{A}$ after the axes, which span the observation space, have been successfully extracted from the data and a proof of the uniqueness of the geometric algorithm are given in the appendix A.2.

4.3.2 The (neural) geometric learning algorithm

We will for now restrict ourselves to the two-dimensional case for simplicity – extension to higher dimensions is in principal straight forward, but problems due to sparseness of samples will be discussed in detail in section 4.3.7.

So, let $S : \Omega \to \mathbb{R}^2$ be an independent two-dimensional random vector describing the source pattern distribution; its density function is denoted by $\rho : \mathbb{R}^2 \to \mathbb{R}$. As S is independent, ρ factorises in the following way

$$\rho(s_1, s_2) = \rho_1(s_1) \cdot \rho_2(s_2), \tag{4.42}$$

with $\rho_i : \mathbb{R} \to \mathbb{R}$ denoting the corresponding marginal source density functions.

As a further simplification we will assume the source variables S_i to have zero mean $E(S) = 0$ and to be distributed symmetrically, i.e. $\rho_i(x) = \rho_i(-x)$ for $x \in \mathbb{R}$ and $i = 1, 2$. To assure stability of the geometric algorithm we further have to assume that the source distributions are nongaussian and unimodal – in practice these restrictions are often met at least approximately.

As usual let $X = \mathbf{A} \circ S$ denote the sensor signal vector. The *geometric learning algorithm* or often called *neural geometric algorithm* for symmetric distributions in its simplest form then goes as follows:

Pick four starting neurons w_1, w_1', w_2 and w_2' on S^1 such that w_i and w_i' are opposite each other, i.e. $w_i = -w_i'$ for $i = 1, 2$, and w_1 and w_2 are linearly independent vectors in R^2. Usually, one takes the unit vectors $w_1 = \mathbf{e}_1$ and $w_2 = \mathbf{e}_2$. Furthermore fix a learning rate $\eta : \mathbb{N} \to \mathbb{R}$. The usual hypothesis – as given in Cottrell et al. [1994] – is $\eta(t) > 0$, $\sum_{n \in \mathbb{N}} \eta(n) = \infty$ and $\sum_{n \in \mathbb{N}} \eta(n)^2 < \infty$. Then iterate the following step until an appropriate abort condition has been met:

Choose a sample $\mathbf{x}(t) \in \mathbb{R}^2$ according to the distribution of X. If $\mathbf{x}(t) = 0$ pick a new one – note that this case happens with probability zero since the probability density function (pdf) ρ_X of X is assumed to be continuous. Project $\mathbf{x}(t)$ onto the unit sphere to get $\mathbf{y}(t) := \frac{\mathbf{x}(t)}{|\mathbf{x}(t)|}$. Let i be in $\{1, 2\}$ such that w_i or w_i' is the neuron closest to $\mathbf{y}$ with respect to a Euclidean metric. Then update $w_i(t)$ according to the following update-rule

$$w_i(t+1) := \mathrm{pr}\left(w_i(t) + \eta(t) \frac{y(t) - w_i(t)}{|y(t) - w_i(t)|} \right), \tag{4.43}$$

where $\mathrm{pr} : \mathbb{R}^2 \setminus \{0\} \to S^1$ represents the projection onto the unit sphere, and

$$w_i'(t+1) := -w_i(t+1). \tag{4.44}$$

The other two neurons are not moved in this iteration.

Later the following theorem will be proven (under some more restrictions) for $n = 2$.

Theorem 4.1 (Geometric ICA). *Let S be unimodal, symmetric and nongaussian. Assume the algorithm has converged against $w_1, ..., w_n$. Then $(w_1|...|w_n)^{-1}$ solves the BSS problem.*

In figures 4.4 the learning algorithm has been visualised both on the S^1 sphere and after the projection onto $[0, \pi)$.

This competitive weight update-rule may be called *absolute winner-takes-all learning*. It resembles Kohonen's learning algorithm for self-organising maps with a trivial neighbourhood function, but with the modification that the step size along the direction of a sample does not depend on distance, and that the learning process takes place on S^1 not in $\mathbb{R}$.

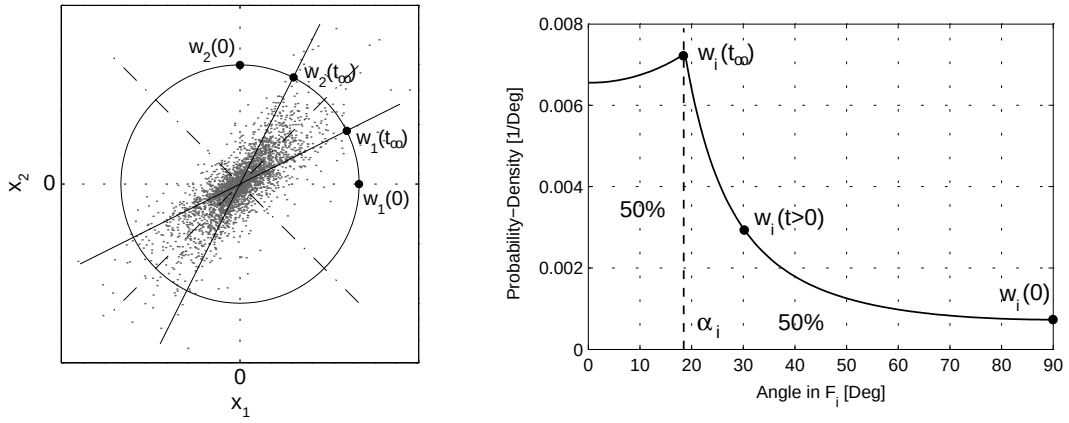

Figure 4.4: *Left:* Visualisation of the geometric algorithm with starting points $w_1(0)$ and $w_2(0)$ and end points $w_1(\infty)$ and $w_2(\infty)$ in a scatterplot of a mixture of two Laplacian signals with identical variance. Dash-dotted lines mark receptive field borders. *Right:* Plot of the density ρ_Y on the sphere S^1 of the mixture of the two Laplacian signals with identical variance.

4.3.3 Theoretical framework for the geometric ICA algorithm

Now, we present a formal theoretical framework for geometric ICA which will be used in the next section to formulate a proper convergence condition. Without loss of generality, assume that $\mathbf{A}$ is of the form

$$\mathbf{A} = \begin{pmatrix} \cos\alpha_1 & \cos\alpha_2 \\ \sin\alpha_1 & \sin\alpha_2 \end{pmatrix}. \tag{4.45}$$

where $\alpha_i \in [0, \pi)$ denote two angles.

First, we show using the symmetry of S that it is in fact not necessary to have two neurons w_i and w_i' moving around on the same axis. Indeed, we should not speak of neurons but of lines in $\mathbb{R}^2$ – so the w_i's would be living in the real projective space $\mathbb{RP}^1 = S^1/\sim$, where $\sim$ identifies antipodal points. This is the manifold of all 1-dimensional subvectorspaces of $\mathbb{R}^2$. A metric is defined by setting

$$d([x],[y]) := \min\{|x-y|, |x+y|\} \tag{4.46}$$

for $[x], [y] \in \mathbb{RP}^1$. Alternatively, one can picture the w's in

$$S^1_+ := S^1 \cap \{(x_1, x_2) \in \mathbb{R}^2 | x_2 \geq 0\}/\sim, \tag{4.47}$$

where $\sim$ identifies the two points $(1,0)$ and $(-1,0)$. Let $\zeta : S^1 \to S^1_+$ represent the canonical projection. Furthermore, it is useful to introduce polar coordinates $\varphi : S^1_+ \to [0, \pi)$ on S^1_+ with the stratification $\varphi' : \mathbb{R} \to S^1_+$ such that $\varphi' \circ \varphi = \mathrm{id}$, where id denotes the identity. Let $\chi := \varphi \circ \varphi' : \mathbb{R} \to [0, \pi)$ be the 'modulo π' map.

We are interested in the projected random sensor signal vector $\mathrm{pr} \circ X : \Omega \to S^1$, so, after cutting open the circle S^1 and identifying opposite points, we want to approximate the transformed random variable $Y := \varphi \circ \zeta \circ \mathrm{pr} \circ X : \Omega \to [0, \pi)$. in a suitable manner. Note that using the symmetry of ρ, the density function ρ_Y of the transformed sensor signal Y can be calculated from the density ρ_X of the original sensor signal X by

$$\begin{aligned} \rho_Y(\varphi) &= \int_{-\infty}^{\infty} \rho_X(r\cos\varphi, r\sin\varphi)\, r\, \mathrm{d}r \\ &= |\det \mathbf{A}|^{-1} \int_{-\infty}^{\infty} \rho(\mathbf{A}^{-1}(r\cos\varphi, r\sin\varphi)^T)\, r\, \mathrm{d}r \\ &= 2|\det \mathbf{A}|^{-1} \int_{0}^{\infty} \rho(\mathbf{A}^{-1}(r\cos\varphi, r\sin\varphi)^T)\, r\, \mathrm{d}r \end{aligned} \tag{4.48}$$

where the last identity follows for any $\varphi \in [0, \pi)$ because of the symmetry of ρ. Then the geometric learning algorithm induces the following discrete Markov process $W(t) : \Omega \to \mathbb{R}^2$ defined recursively by $W(0) = (w_1, w_2)$ and $W(t+1) = \chi^2(W(t) + \eta(t)\theta((Y(t), Y(t)) - W(t)))$, where

$$\theta(x, y) := \begin{cases} (\mathrm{sgn}(x), 0) & |y| \geq |x| \\ (0, \mathrm{sgn}(y)) & |x| > |y| \end{cases} \tag{4.49}$$

and $Y(0), Y(1), \ldots$ is a sequence of independent identically distributed random variables $\Omega \to \mathbb{R}^2$ with the same distribution as Y. These random variables will be needed to represent the independence of the successive sampling experiments. Note that the 'modulo π' map χ guarantees that $W(t+1) \in [0, \pi)$. Indeed, this is just winner-takes-all learning with a signum function in $\mathbb{R}$, but taking into account the fact that we have to stay in $[0, \pi)$. Note that the metric used here is the planar metric, which obviously is equivalent to the metric on S^1_+ induced by the Euclidean metric on $S^1 \subset \mathbb{R}^2$.

We furthermore can assume that after a sufficient number of iterations there is one point $a \in S^1$ that will not be transversed any more, and without loss of generality, we assume a to be $\mathbf{0}$ (otherwise cut S^1 open at a and project along this resulting arc), so that the above algorithm simplifies to the planar case with the recursion rule

$$W(t+1) = W(t) + \eta(t)\theta\left((Y(t), Y(t)) - W(t)\right). \tag{4.50}$$

Without the sign function, and the additional fact that the probability distribution of Y is *log-concave*, it has been shown in Cottrell and Fort [1987], Ritter and Schulten [1988] and Benaim et al. [1998] that the process $W(t)$ converges to a unique constant fixed point process $W \equiv w \in \mathbb{R}^2$ such that

$$\int_{\beta_1(F_i)}^{w_i} \rho_Y(\varphi)\, \mathrm{d}\varphi = \int_{w_i}^{\beta_2(F_i)} \rho_Y(\varphi)\, \mathrm{d}\varphi \tag{4.51}$$

for $i = 1, 2$, where

$$F_i := F(w_i) := \{\varphi \in [0, \pi) \mid \chi(|\varphi - w_i|) \leq \chi(|\varphi - w_j|) \text{ for all } j \neq i\} \tag{4.52}$$

denotes the *receptive field* of w_i and $\beta_j(F_i)$ designate the receptive field borders. However, it is not clear how to generalise the proof to the geometric case, especially because we do not have (and also do not want) log-concavity of Y as this would lead to a unique fixed point. Therefore we will assume convergence in a sense stated in the following section.

4.3.4 Limit points of the geometric algorithm

In this section, we want to study the end points of geometric ICA, so we will assume that the algorithm has already converged. The idea then is to formulate a condition which the end points will have to satisfy and to show that the solutions do so, too.

Definition 4.2 (Geometric Convergence Condition). *We say that two angles $l_1, l_2 \in [0, \pi)$ satisfy the Geometric Convergence Condition (GCC) if they are the medians of Y restricted to their receptive fields respectively i.e. if l_i is the median of $\rho_Y|F(l_i)$ for $i = 1, 2$.*

Definition 4.3. *A constant random vector $\hat{W} \equiv (\hat{w}_1, \hat{w}_2) \in \mathbb{R}^2$ is called fixed point of geometric ICA in the expectation if*

$$E\{\theta(Y - \hat{W}(t))\} = 0.$$

Hence, the expectation of a Markov process $W(t)$ starting at a fixed point of geometric ICA will indeed not be changed by the geometric update rule because

$$E\{W(t+1)\} = E\{W(t)\} + \eta(t)E\{\theta(Y(t) - W(t))\} = E\{W(t)\}. \tag{4.53}$$

Theorem 4.4. *Given that the geometric algorithm converges to a constant random vector $W(\infty) \equiv (w_1(\infty), w_2(\infty))$, then $W(\infty)$ is a fixed point of geometric ICA in the expectation if and only if the $w_i(\infty)$ satisfy the GCC.*

Proof. Assume $W(\infty)$ is a fixed point of geometric ICA in the expectation. Without loss of generality, let $[\beta_1, \beta_2]$ be the receptive field of $w_1(\infty)$ such that $\beta_i \in [0, \pi)$. Since $W(\infty)$ is a fixed point of geometric ICA in the expectation, we have

$$E\left\{\chi_{[\beta_1,\beta_2]}(Y(t))\,\mathrm{sgn}(Y(t) - w_1(\infty))\right\} = 0 \tag{4.54}$$

where $\chi_{[\beta_1,\beta_2]}$ denotes the characteristic function of that interval. But this means

$$\int_{\beta_1}^{w_1(\infty)} (-1)\rho_Y(\varphi)\,\mathrm{d}\varphi + \int_{w_1(\infty)}^{\beta_2} 1\rho_Y(\varphi)\,\mathrm{d}\varphi = 0 \tag{4.55}$$

and therefore

$$\int_{\beta_1}^{w_1(\infty)} \rho_Y(\varphi)\,\mathrm{d}\varphi = \int_{w_1(\infty)}^{\beta_2} \rho_Y(\varphi)\,\mathrm{d}\varphi, \tag{4.56}$$

so $w_1(\infty)$ satisfies GCC. The same calculation for $w_2(\infty)$ shows one direction of the claim. The other direction follows by simply reading the above proof backwards, which completes the proof. □

As before, let $\mathbf{p}_i := \mathbf{A}\mathbf{e}_i$ be the transformed unit vectors, and let $q_i := \varphi \circ \zeta \circ \mathrm{pr}(\mathbf{p}_i) \in [0, \pi)$ be the corresponding angles for $i = 1, 2$.

Theorem 4.5. *The transformed angles q_i satisfy the GCC.*

Proof. Because of the symmetry of the claim it is enough to show that q_1 is at the median of $\rho_Y|F(q_1)$. Without loss of generality let $0 < \alpha_1 < \alpha_2 < \pi$ using the symmetry of ρ. Then, due to construction $q_i = \alpha_i$. Let $\beta_1 := \frac{\alpha_1+\alpha_2}{2} - \frac{\pi}{2}$ and $\beta_2 := \beta_1 + \frac{\pi}{2}$. Then the receptive field of q_1 can be written (modulo π) as $F(q_1) = [\beta_1, \beta_2]$. Therefore, we have to show that $q_1 = \alpha_1$ is the median of ρ_Y restricted to $[\beta_1, \beta_2]$, which means

$$\int_{\beta_1}^{\alpha_1} \rho_Y(\varphi)\,\mathrm{d}\varphi = \int_{\alpha_1}^{\beta_2} \rho_Y(\varphi)\,\mathrm{d}\varphi. \tag{4.57}$$

We will reduce this to the orthogonal standard case $\mathbf{A} = \mathrm{id}$ by transforming the integral as follows:

$$\int_{\beta_1}^{\alpha_1} \rho_Y(\varphi)\,\mathrm{d}\varphi = 2|\det \mathbf{A}|^{-1} \int_{\beta_1}^{\alpha_1} \int_0^{\infty} \rho(\mathbf{A}^{-1}(r\cos\varphi, r\sin\varphi)^T)\, r\,\mathrm{d}r\,\mathrm{d}\varphi \tag{4.58}$$

$$= 2|\det \mathbf{A}|^{-1} \int_K \rho(\mathbf{A}^{-1}(x, y)^T)\,\mathrm{d}y\,\mathrm{d}x \tag{4.59}$$

where

$$K := \{(x, y) \in \mathbb{R}^2 | \beta_1 \le \arctan(y/x) \le \alpha_1\} \tag{4.60}$$

denotes the cone of opening angle $\alpha_1 - \beta_1$ starting from angle β_1. Using the transformation formula, we continue

$$\int_{\beta_1}^{\alpha_1} \rho_Y(\varphi)\,\mathrm{d}\varphi = 2 \int_{\mathbf{A}^{-1}(K)} \rho(x, y)\,\mathrm{d}y\,\mathrm{d}x. \tag{4.61}$$

Now note that the transformed cone $\mathbf{A}^{-1}(K)$ is a cone ending at the x-axis of opening angle $\pi/4$, because $\mathbf{A}$ is linear. Therefore we are left with the following integral

$$\int_{\beta_1}^{\alpha_1} \rho_Y(\varphi)\,\mathrm{d}\varphi = 2 \int_0^{\infty} \int_{-x}^{0} \rho(x, y)\,\mathrm{d}y\,\mathrm{d}x = 2 \int_0^{\infty} \int_0^{x} \rho(x, -y)\,\mathrm{d}y\,\mathrm{d}x \tag{4.62}$$

$$= 2 \int_0^{\infty} \int_0^{x} \rho(x, y)\,\mathrm{d}y\,\mathrm{d}x = \int_{\alpha_1}^{\beta_2} \rho_Y(\varphi)\,\mathrm{d}\varphi \tag{4.63}$$

where we have used the same calculation for $[\alpha_1, \beta_2]$ as for $[\beta_1, \alpha_1]$ at the last step. This completes the proof of the theorem. □

Combining both theorems, we have therefore shown:

Theorem 4.6. *Let Φ be the set of fixed points of geometric ICA in the expectation. Then there exists $(\hat{w}_1, \hat{w}_2) \in \Phi$ such that $(\hat{w}_1|\hat{w}_2)^{-1}$ solves the BSS problem. The stable fixed points in Φ can be found by the geometric ICA algorithm.*

Furthermore, we believe that in the special case of unimodal, symmetric and nongaussian signals, the set Φ consists of only two elements: a stable and an unstable fixed point, where the stable fixed point will be found by the algorithm:

Conjecture 4.7. *Assume that the sources S_i are unimodal, symmetric and nongaussian. Then there are only two fixed points of geometric ICA in the expectation.*

We can prove this conjecture for the special case of two sources with identical distributions which are nicely super- or subgaussian in the sense that ρ_Y has only four extremal points. This together with the preceding theorem is as close as we can get with proofs to theorem 4.1

Theorem 4.8. *Assume that the sources S_i are unimodal, symmetric and that $\rho_1 = \rho_2$. Assume that $\rho_Y|[0, \pi)$ with $\mathbf{A} = \mathrm{id}$ has exactly two local maxima and two local minima. Then there exist only two fixed points of geometric ICA in the expectation.*

Proof. Due to the lengthiness, the proof can be found in appendix A.3. □

The above conjecture states that there are only two fixed points of geometric ICA. In fact, we claim that of those two, only one fixed point is stable in the sense that slight perturbations of the initial conditions preserve the convergence. Then, depending on the kurtosis of the sources, either the stable (supergaussian case: kurtosis > 0) or the instable (subgaussian case: kurtosis < 0) fixed point represents the image of the unit vectors. This is stated in the following conjecture.

Conjecture 4.9. *Assume that the sources S_i are unimodal, symmetric and nongaussian. Then by conjecture 4.7 there are only two fixed points $(\hat{w}_1, \hat{w}_2)$ and $(\tilde{w}_1, \tilde{w}_2)$ of geometric ICA in the expectation. We claim:*

i. There is only one stable fixed point $(\hat{w}_1, \hat{w}_2)$.

ii. If the sources are supergaussian, $(\hat{w}_1|\hat{w}_2)^{-1}$ solves the BSS problem.

iii. If the sources are subgaussian, $(\tilde{w}_1|\tilde{w}_2)^{-1}$ solves the BSS problem.

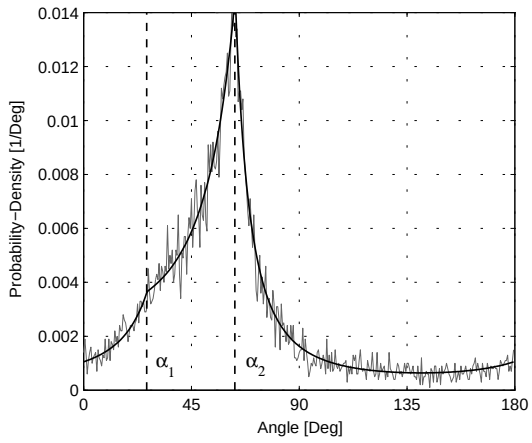

Figure 4.5: Projected density distribution ρ_Y of a mixture of two Laplacian signals with different variances, with the mixture matrix mapping the unit vectors e_i to $(\cos\alpha_i, \sin\alpha_i)$ for $i = 1, 2$ (smooth line = theoretical density function, ragged line = histogram of a mixture of 10.000 samples).

Update rules without sign functions

We have shown that in two dimensions the geometric update step after projection onto $[0, \pi)$ requires the signum function as follows

$$w_i(t+1) = w_i(t) + \eta(t)\,\mathrm{sgn}(y(t) - w_i(t)). \tag{4.64}$$

Then the w_i converge to the medians in their receptive field. Note that the medians don't have to coincide with any maxima of the sensor signal density distribution on the sphere as shown in figure 4.5. Therefore, in general, any algorithm searching for the maxima of the distribution – as done in Prieto et al. [1999] – will not end at the medians, which are the correct images of the unit vectors under the given mixing transformation. Only given special restrictions to the sources (same super-gaussian distribution of each component, as for example speech signals), the medians correspond to the maxima and a maximum searching algorithm will converge to the correct fixed points of geometric ICA.

4.3.5 FastGeo: A histogram based algorithm

So far in geometric ICA – Puntonet and Prieto [1995] and Puntonet and Prieto [1998] – mostly 'neural' algorithms as in section 4.3.2 have been applied. As shown above, these algorithms search for points satisfying the GCC. In the following we will establish a new geometric algorithm called *FastGeo* using the GCC only. The theoretical derivations and the applications were published in Jung et al. [2001] and Theis et al. [2003b].

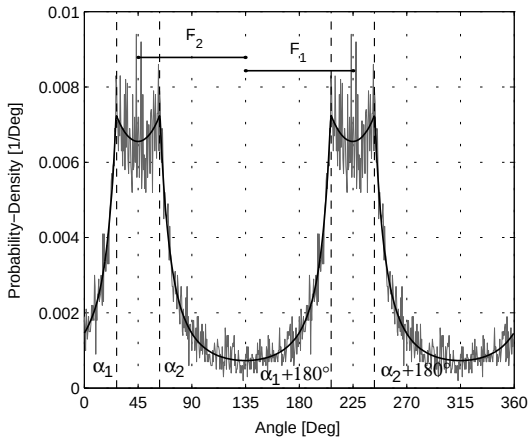

Figure 4.6: Probability density function ρ_Y of Y – from the scatter plot in figure 4.4 – with the mixing angles α_i and their receptive fields F_i for $i = 1, 2$ (smooth line = theoretical density function, ragged line = histogram from the scatter plot).

Let again the dimension $n = 2$ and let

$$\mathbf{A} := \begin{pmatrix} \cos\alpha_1 & \cos\alpha_2 \\ \sin\alpha_1 & \sin\alpha_2 \end{pmatrix}. \tag{4.65}$$

Theorem 4.5 shows that the vectors $(\cos\alpha_i, \sin\alpha_i)^T$ satisfy the GCC. Therefore the vectors w_i will converge to the medians in their receptive fields. This enables us to compute these positions directly using a search on the histogram of Y, which reduces the computation time by a factor of about 100 or more. In the *FastGeo-algorithm* we scan through all different receptive fields and test GCC. In practice this means discretising the distribution f_Y of Y using a given bin-size $\beta > 0$ and then testing the π/β different receptive fields. The algorithm will be formulated more precisely in the following:

For simplicity let us assume that the cumulative distribution F_Y of Y is invertible – this means that F_Y is nowhere constant. Define a function

$$\begin{aligned} \mu : [0, \pi) &\to \mathbb{R} \\ \varphi &\mapsto \frac{l_1(\varphi) + l_2(\varphi)}{2} - (\varphi + \frac{\pi}{2}) \end{aligned} \tag{4.66}$$

where

$$l_i(\varphi) := F_Y^{-1}\left(\frac{F_Y(\varphi + i\frac{\pi}{2}) + F_Y(\varphi + (i-1)\frac{\pi}{2})}{2}\right) \tag{4.67}$$

is the median of $Y|[\varphi + (i-1)\frac{\pi}{2}, \varphi + i\frac{\pi}{2}]$ in $[\varphi + (i-1)\frac{\pi}{2}, \varphi + i\frac{\pi}{2}]$ for $i = 1, 2$.

Lemma 4.10. *Let φ be a zero of μ in $[0, \pi)$. Then the $l_i(\varphi)$ satisfy the GCC.*

Proof. By definition,

$$\left[\frac{l_1(\varphi)+l_2(\varphi)}{2}-\frac{\pi}{2},\frac{l_1(\varphi)+l_2(\varphi)}{2}\right] \tag{4.68}$$

is the receptive field of $l_1(\varphi)$. Since $\mu(\varphi)=0$, the starting point of the above interval is φ, because

$$\varphi=\frac{l_1(\varphi)+l_2(\varphi)}{2}-\frac{\pi}{2}. \tag{4.69}$$

Hence we have shown that the receptive field of $l_1(\varphi)$ is $[\varphi,\varphi+\frac{\pi}{2}]$, and by construction $l_1(\varphi)$ is the median of Y restricted to the above interval. The claim for $l_2(\varphi)$ then follows. □

Algorithm 4.11 (FastGeo). *Find the zeros of μ.*

Hence by lemma 4.10 and conjecture 4.9 FastGeo solves the BSS problem.

μ always has at least two zeros which represent the stable and the unstable fixed point of the neural algorithm. In practice we extract the fixed point which then gives the proper demixing matrix A^{-1} by picking φ_0 such that

$$f_Y(l_1(\varphi_0))+f_Y(l_2(\varphi_0)) \tag{4.70}$$

is maximal. For unimodal and super-gaussian source distributions conjecture 4.9 claims that this results in a stable fixed point. For sub-gaussian sources choosing φ_0 with

$$f_Y(l_1(\varphi_0))+f_Y(l_2(\varphi_0)) \tag{4.71}$$

being minimal induces the corresponding demixing matrix. Hence, one advantage of this histogram-based algorithm is that without any modifications we can solve the ICA problem also for subgaussian signals. Furthermore, the sophisticated parameter choice of the 'neural' algorithm is not necessary any more; only one parameter, the bin size, has to be chosen.

In practice, one sometimes notices that due to the discretization of the distribution, the approximated distribution has a rather noisy shape on small scales. This results from μ having zeros being split up into multiple zeros close together. Therefore, a useful improvement of convergence can be established by smoothing this distribution with a kernel function with sufficiently small halfwidth. This smoothing should be performed preferably during the discretization process of the original distribution.

4.3.6 Accuracy and performance of FastGeo

In this section we want to test the accuracy of the results from the FastGeo algorithm and compare the performance with other established ICA algorithms.

Accuracy

First we want to consider the dependence of the FastGeo ICA algorithm on the number of samples after the bin-size β has been fixed. As seen in the previous section, the accuracy of the histogram based algorithm then *only* depends on the distribution of the samples X respectively Y i.e. we can estimate the error made by approximating the mixing matrix $\mathbf{A}$ by a finite number of samples. In the following we will give some results of test-runs made with this algorithm.

When choosing two arbitrary angles $\alpha_i \in [0, \pi), i = 1, 2$ for the mixing matrix $\mathbf{A}$, we define α as the distance between these two angles modulo $\frac{\pi}{2}$. This will give us an angle in the range between 0 and $\frac{\pi}{2}$ respectively 0° and 90°.

First let us consider the accuracy of the recovered angles $\Delta\alpha = |\alpha_i - \alpha_i^{\text{recovered}}|$, when varying the angle α for a fixed number of samples. Choosing a mixture of two Laplacian source signals ($p(x) \sim e^{-|x|}$) with identical variances, figure 4.7 shows a nearly linear decrease of the error $\Delta\alpha$ with decreasing α. The 95% confidence interval decreases similar to the standard deviation, i.e. both parameters provide a good estimate for the error. Note that the mean of the error $\Delta\alpha$ is very close to zero.

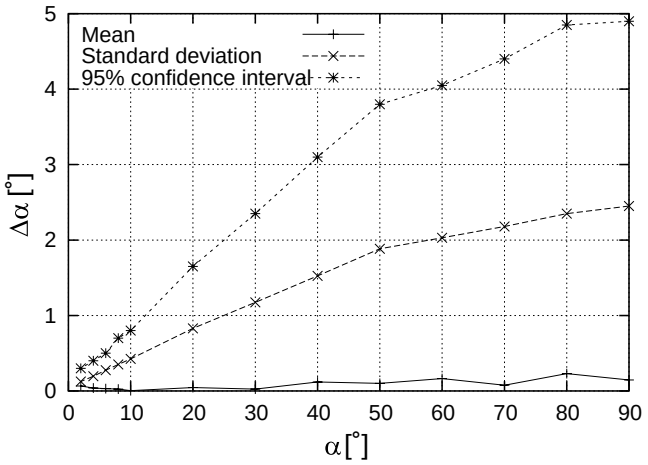

Figure 4.7: Mixture of 1.000 samples of two Laplacian source signals with identical variances. Plotted is the mean, standard deviation and 95% confidence interval of $\Delta\alpha$ calculated from 100 runs for each angle α.

An alternative way to demonstrate this proportionality $\Delta\alpha \propto \alpha$ is to plot $\Delta\alpha/\alpha$ versus α in figure 4.8, which should result in a horizontal line. Obviously the resulting graph is reasonably constant over a wide range of α demonstrating that the estimate of the α_i with respect to α is good for a wide range of α ($\alpha > 10°$) and gets only slightly worse for smaller values of α. Note that the

distortions around the origin are due to the finite bin-size. Increasing the number of bins increases the accuracy for small α's, but also the computational effort.

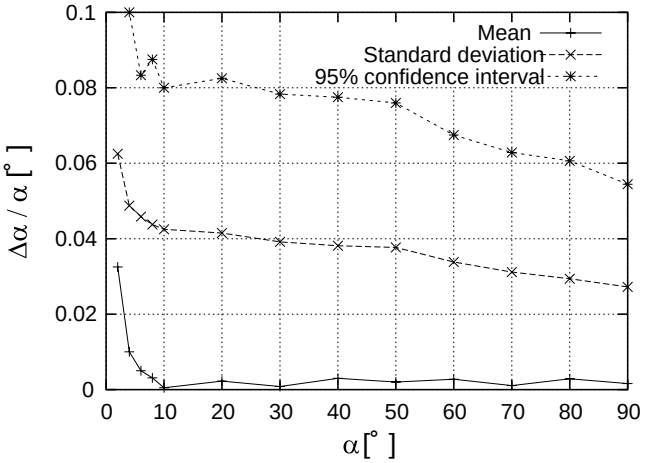

Figure 4.8: Same mixture as in figure 4.7, plotting $\Delta\alpha/\alpha$ versus α.

Varying the number of samples used for estimating the angles α_i shows that with increasing number of samples the quality of the estimate enhances as denoted in figure 4.9. This fact is well know in statistics and signal processing.

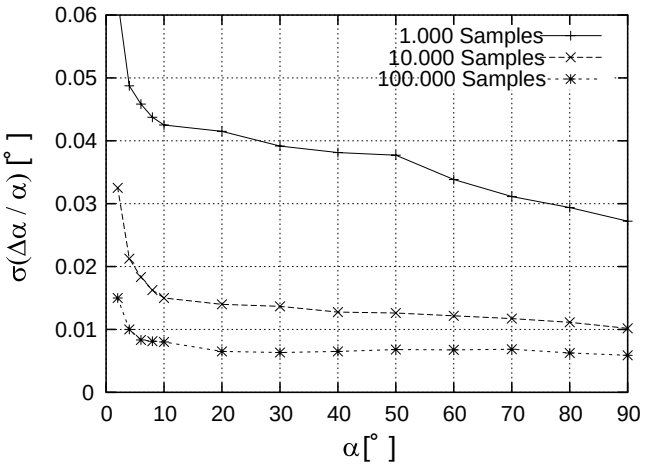

Figure 4.9: Mixture of two Laplacian source signals with identical variances for different number of samples. The standard deviation of $\Delta\alpha/\alpha$ calculated from 100 runs for each angle α is plotted.

To investigate the relation between the error $\Delta\alpha$ and the number of samples, we plot for different α, the standard deviation of $\Delta\alpha$ versus the number of samples, see figure 4.10.

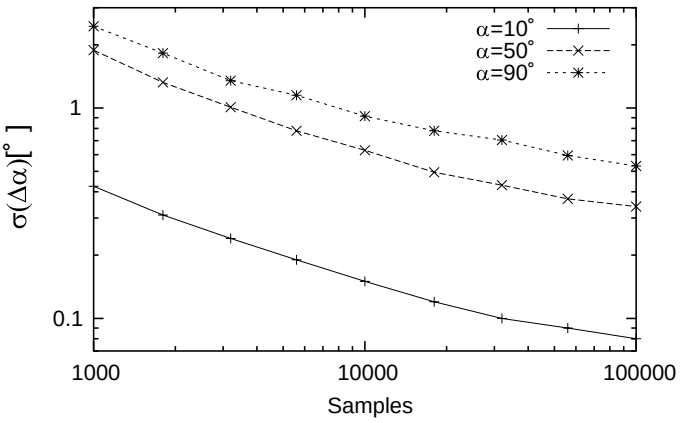

Figure 4.10: Dependence of standard deviation of $\Delta\alpha$ with the number of samples for estimating the α_i for three different angles $\alpha = 10°$, $50°$ and $90°$. Note the logarithmic scale on both axes.

The above results have been collected numerically in table 4.1, where we chose the following mixing matrix $\mathbf{A}$:

$$\mathbf{A} := \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}. \tag{4.72}$$

The table entry gives the standard deviation of the non-diagonal terms after normalising each column of the mixing matrix, so that the diagonal elements are unity. For comparison, we also calculated the performance index E_1 as proposed by Amari Amari et al. [1996].

$$E_1 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{|p_{ij}|}{\max_k |p_{ik}|} - 1 \right) + \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \frac{|p_{ij}|}{\max_k |p_{kj}|} - 1 \right) \tag{4.73}$$

where $\mathbf{P} = (p_{ij}) = \mathbf{A}_{recon}^{-1} \cdot \mathbf{A}$.

In statistics, an often used technique for estimating the error made by approximating a probability density function by a finite number of samples is the so-called *confidence interval*, which is the interval around the estimate obtained from a given number of samples such that the probability that the real value lies outside this interval is less than a fixed *error probability* ε:

$$P(|X - \hat{x}| < c) = 1 - \varepsilon. \tag{4.74}$$

number of samples	standard deviation	index E_1
1.000	0.033	0.18
10.000	0.013	0.07
100.000	0.007	0.038

Table 4.1: Standard deviations of the non-diagonal terms and the performance index E_1 with different number of samples.

For estimating the median of a probability density function, we refer to Bosch [1993]: Let $x_1, \ldots, x_n$ be independent identically distributed (i.i.d.) samples of the random variable X, such that $x_i < x_{i+1}$, then the *estimated median* $\hat{x}$ of the samples is defined as $x_{\frac{n+1}{2}}$ if n is odd or $\frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$ if n is even. For large n ($n > 36$), let z be the inverse of the cumulative standard distribution of $1 - \frac{\varepsilon}{2}$ and

$$k = -0,5 + \frac{n}{2} + \frac{\sqrt{n}}{2} z. \tag{4.75}$$

Then the confidence interval of $\hat{x}$ is given by

$$[x_{1+k}, x_{n-k}]. \tag{4.76}$$

In our case, an approximate confidence interval can be calculated by first running the algorithm to estimate approximate angles α_i. Then use the equations given above to derive confidence intervals for the samples of Y restricted to the corresponding receptive fields F_i. This is useful for error estimation in real world problems.

Performance

Now, we compare geometric ICA algorithms with other ICA algorithms, namely the *Extended Infomax* algorithm by Lee et al. [1999], which is based on the classical Bell-Sejnowski ICA algorithm (Bell and Sejnowski [1995]) using the maximum entropy criteria, and the *FastICA* algorithm by Hyvärinen and Oja [1997a] and Hyvärinen [1999] in version 2.1, with different non-linearities as described in the articles. As geometric algorithms, we use the classical *Neural Geometric* algorithm presented in Puntonet and Prieto [1995] and Puntonet and Prieto [1998] and the *FastGeo* algorithm from above. Calculations were conducted on a PIII-850 PC using Matlab 6.0.

In our first example, we consider a mixture of two Laplacian signals. The results of the different algorithms are shown in table 4.2: for each algorithm we measure the mean elapsed CPU-time per run and the mean crosstalking error E_1 with its standard deviation.

Both the Extended Infomax and the FastICA algorithm perform best in terms of accuracy. Using more complicated non-linearities where accuracy, and

Algorithm	elapsed time [s]	index E_1
Extended Infomax	11.1	0.072±0.002
FastICA (pow3=default)	0.068	0.076±0.004
FastICA (tanh)	0.11	0.052±0.001
FastICA (gauss)	0.12	0.048±0.001
Neural Geometric	>60	0.18±0.10
FastGeo	0.84	0.110±0.071

Table 4.2: Comparison of time per run and crosstalking error of ICA algorithms for a random mixture of two Laplacian signals. Means and standard deviations were taken over 1000 runs (100 runs for Extended Infomax and Geo) with 10000 samples and uniformly distributed mixing matrix elements.

in terms of computational speed, FastICA lives up to its name being followed by FastGeo, then Infomax and the neural algorithms, always by one order of magnitude slower. The neural algorithm lacks of accuracy, and also shows some convergence problems, whereas FastGeo lies between the geometric algorithm and FastICA/Infomax regarding accuracy.

Algorithm	elapsed time [s]	index E_1
Extended Infomax	41.2	0.058±0.002
FastICA (pow3=default)	0.14	0.050±0.005
FastICA (tanh)	0.24	0.022±0.001
FastICA (tanh)	0.26	0.019±0.001
Neural Geometric	>60	0.49±0.29
FastGeo	0.89	0.136±0.087

Table 4.3: Comparison of time per run and crosstalking error of ICA algorithms for a random mixture of two sound signals with 22000 samples. Means and standard deviations were taken over 1000 runs (100 runs for Extended Infomax and Geo) with uniformly distributed mixing matrix elements.

The second example deals with real-world data: two audio signals (one speech and one music signal) – see table 4.3. The results are similar to the Laplacian toy example. FastICA outperforms the other algorithms in terms of speed. The accuracy of Extended Infomax and FastICA are comparable, and FastGeo is slightly (factor 4) worse, but faster than the extended Infomax. The neural geometric algorithm again is both slower and less accurate, mainly because of

convergence problems.

The conclusion of our test-runs is the confirmation of the exceptional performance of the FastICA algorithm; nonetheless, the FastGeo algorithm is much faster than both Extended Infomax and the Neural Geometric algorithm. Therefore, when geometric algorithms are the choice, we suggest to use FastGeo instead of any neural approach, also because of a much higher stability of FastGeo.

4.3.7 Higher dimensions

So far we have explicitly considered two-dimensional data sets only. In real world problems, however, the sensor signals are usually high-dimensional data sets like, for example, EEG-data with 21 dimensions. Therefore it is necessary to generalise geometric algorithms to higher dimensions. The neural geometric algorithm can be easily translated to higher dimensional cases, but one faces serious problems in the explicit calculation: In order to approximate higher dimensional probability density functions it becomes necessary to have an exponentially growing number of samples available, as will be shown in the following.

The number of samples in a ball B^{d-1} of radius ϑ on the unit sphere $S^{d-1} \subset \mathbb{R}^d$ divided by the number of samples on the whole S^{d-1} can be calculated as follows, if we assume a uniformly distributed random vector.

Let $B^d := \{x \in \mathbb{R}^d || x| \leq 1\}$ and $S^{d-1} := \{x \in \mathbb{R}^d || x| = 1\}$ – referring to Pathria [1998], the volume of B^d can be calculated by

$$\mathrm{vol}(B^d) = \frac{\pi^{\frac{d}{2}}}{\left(\frac{d}{2}\right)!} = c_d. \tag{4.77}$$

It follows for $d > 3$:

$$\frac{\text{Number of Samples in Ball}}{n} = \frac{n\frac{\mathrm{vol}(B^{d-1})\vartheta^{d-1}}{\mathrm{vol}(S^{d-1})}}{n} \tag{4.78}$$

$$= \frac{\vartheta^{d-1}c_{d-1}}{c_d} \tag{4.79}$$

$$\leq \frac{\vartheta^{d-1}c_{d-1}}{c_{d+1}} \tag{4.80}$$

$$= \frac{\vartheta^{d-1}d}{\pi} \tag{4.81}$$

Obviously the number of samples in the Ball decreases by $\vartheta^{d-1}d$ if $\vartheta < 1$, which is the interesting case. To have the same accuracy when estimating the medians, the decrease must be compensated by an exponential growth in the number of samples. For three dimensions using the standard geometric learning algorithm, we have found a good approximation for the demixing matrix by using 10^5 samples, in four dimensions the reconstructed mixing matrix couldn't be reconstructed correctly, even with a larger number of samples.

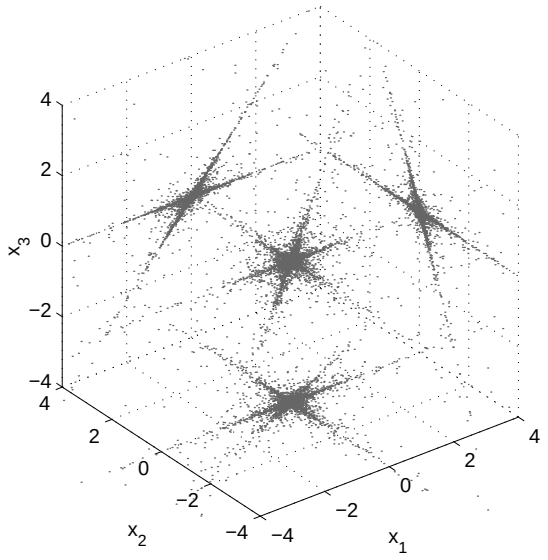

Figure 4.11: Projection of a three dimensional mixture of Laplacian signals onto the three coordinate planes. Note that the projection into the x_1-x_2-plane does not have two distinct lines which are needed for the geometric algorithms.

Bauer et al. [2000] and Puntonet et al. [2000] have taken a different approach for higher dimensions, where A has been calculated by using $\frac{d(d-1)}{2}$ projections of X from $\mathbb{R}^d$ onto $\mathbb{R}^2$ along the different coordinate axes and reconstructed the multidimensional matrix from the two-dimensional solutions. However, this approach works only satisfactorily if the mixing matrix A is close to the unit matrix up to permutation and scaling. Otherwise, even in three dimensions, this projection approach won't give the desired results, as can be seen in figure 4.11, where the mixing matrix has been chosen as

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0.65 \\ 0 & 1 & 0.3 \\ 0 & 0 & 0.7 \end{pmatrix}. \tag{4.82}$$

4.3.8 Conclusions

The geometric ICA algorithm has been studied in a concise theoretical framework resembling the one of Kohonen's learning algorithm. The fixed points of the geometric ICA learning algorithm have been examined in detail. We have introduced a Geometric Convergence Condition, which has to be fulfilled by the fixed points of the learning algorithm. We further showed that it is also fulfilled by the mixed unit vectors spanning the sensor signal space. Hence geometric ICA can solve the BSS problem. Finally, we have finished the theoretic part by giving

two conjectures for the unimodal case where the fixed point property is expected to be very rigid.

We have then presented a new algorithm for linear geometric ICA (FastGeo) based on histograms, which is both robust and computationally much more efficient than the neural geometric ICA algorithm. The accuracy of the algorithm concerning the estimation of the relevant medians of the underlying data distributions, when varying both the mixing matrices and the sample numbers, has been explored quantitatively, showing a rather good performance of the algorithm. Simulations with non-symmetrical and non-unimodal distributions have shown promising results so far, indicating that the new algorithm will perform well with almost any distribution. This is the subject of ongoing research in our group.

When comparing FastGeo with classical ICA algorithms and the neural geometric one, we noticed that FastGeo performs only slightly worse than the classical ones in terms of accuracy, and better than the neural one; in terms of speed, FastGeo places itself between FastICA and the Extended Infomax, and is much faster than the neural approach, which also suffers from severe convergence problems. Furthermore, the fact that geometric algorithms and especially FastGeo are very easy to implement makes FastGeo a good choice even in comparison with the classical ICA algorithms in practical 2-dimensional applications.

We also considered the problem of high dimensional data sets with respect to the geometrical algorithms and discussed how projections to low-dimensional subspaces could solve this problem for a special class of mixing matrices.

In future work, the two conjectures will have to be proven in full, as well as the Kohonen proof of convergence to be translated into the above model. In addition, the histogram based algorithm could be extended to the non-linear case similar to Puntonet et al. [1999], using multiple centred spheres for projection on the surface on which the projected data histograms could then be evaluated.

4.4 An information theoretical approach including time structures

So far we have treated the independent component analysis as a tool in which the structure of a time series was not been taken into account. But in many real world applications the signals are not just independent samples of a random variable but show an ordering, a temporal structure. This structure can be used to extend the classical ICA to cases in which it would fail.

In this section we will show how temporal structures in time series can be used in the framework of independent component analysis assuming the signals arise from Markov chains with finite order. Taking the past of the underlying processes into account by using *time embedding* vectors, not only *instantaneous independent* – as done in the classical ICA – but also *uncoupled* sources can be found. As a result signals which are gaussian distributed at each time can be decomposed as long as the time embedding vectors are nongaussian. Using the model of independent time embedding vectors, we derive an algorithm – we will call it Fast Time Embedding ICA (FastTeICA) – which is similar to the well known FastICA algorithm introduced by Hyvärinen [1999].

A comparison will be conducted between the known ICA algorithm already dealing with time structures and the new derived FastTeICA algorithm, showing the performance and the more flexible application due to less assumptions for the FastTeICA algorithm.

At the end a new concept will be presented which weakens the strict assumption of independent time embedding vectors but still taking into account the dynamics of the processes for signal decomposition. This can be achieved by assuming independent increments, i.e. the change of state respectively the dynamics of the processes is assumed to be independent, which is correct for Markov chains of order one and a good assumption for higher order Markov chains. Both approaches, independent states and independent dynamics/increments, are a special case of the independence of the time embedding vectors.

The results presented in this section are published in Jung and Kaiser [2003].

4.4.1 Introduction

In many real world applications one deals with time signals which are given by stochastic processes. Hence, the random variables of each source are usually correlated in time. The so far proposed algorithms by Hyvärinen [1999] using negentropy and by Bell and Sejnowski [1995] using the maximum information principle, do not consider these additional temporal structures. In particular, scrambling the time series leads to the same result, since these algorithms only use as criteria the independence between the signals at the same time step. Thus, one could try to improve the recovery of the sources, using the extra structure

contained in the signals.

Approaches using time structures

A set of algorithms dealing with time structures have been proposed in the past, using different approaches:

Tong et al. [1991], Molgedey and Schuster [1994], Belouchrani et al. [1997] and Ziehe and Müller [1998] used autocorrelation functions to separate temporal correlated signals. Simple and intuitive algorithms can be derived from this approach, but they all lag of two disadvantages: first the autocorrelation functions of all sources have to be different since otherwise the matrix of the autocorrelation functions is singular and second, nonlinear temporal correlations can not be detected and therefore the sources not be separated.

Pearlmutter and Parra [1997] and Amari [2000] proposed to use autoregressive models based on maximum likelihood approaches. There the coefficients of the autoregressive models are estimated, i.e. one is trying to find the model generating the signals. But in the derivation of the algorithm one has to assume, that the coefficients of the models have to be different for each signals.

In Matsuoka et al. [1995] the idea was proposed to use the non-stationarity of the signals, which works very successful for speech signals. Other sources having no non-stationarity will not be treated with this approach.

Finally, an approach using the the past time steps in a probabilistic theory was proposed by Hosseini and Jutten [2001] using for the estimation of the sources a quasi maximum likelihood approach. There the nonlinear correlations are also taken into account, but the maximum likelihood approach needs for the correct convergence the knowledge of the probability density functions of the sources, which are normally not known.

Our approach: independent time embedding vectors

In our approach we extend the idea of looking "only" for *(instantaneously) independent* signals, i.e. the states of the sources $s_1(t), \ldots, s_n(t)$ at time t are independent, to the idea of independent *and* uncoupled signals.

For this, we assume that the processes can be approximated by Markov chain of order m. We say, the Markov processes x and y, both of order m, are *(stochastically) uncoupled* from each other, if the transition probability of the process (x, y) factorise into the transition probabilities of the processes x and y such that the transition probabilities of x does not dependent on the past of y and vice versa, i.e.

$$\begin{aligned} p(x(t), y(t)|x(t-1), ..., x(t-m), y(t-1), ..., y(t-m)) = \\ p(x(t)|x(t-1), ..., x(t-m)) \cdot p(y(t)|y(t-1), ..., y(t-m)) \end{aligned} \quad (4.83)$$

A schematic illustration of (in)dependent and (un)coupled Markov processes is shown in figure 4.12.

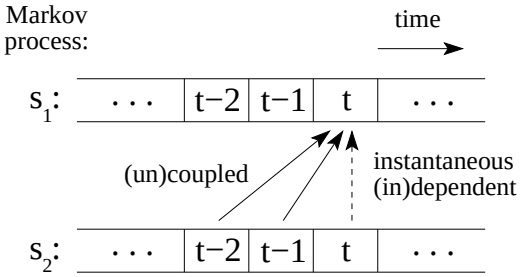

Figure 4.12: Schematic illustration of the possible couplings between two Markov processes s_1 and s_2. In the classical ICA only the coupling between the same time steps of the processes is investigated, hence we call it instantaneous (in)dependence. Taking also the past time steps into account, one speaks of (un)coupled processes.

Based on this model of coupling, it is possible to determine the coupling direction and to quantify the information transfer between two stochastic processes as first presented by Schreiber [2000] and later also in the context of point processes in Kaiser and Schreiber [2002].

In particular, x and y are uncoupled if the $(m+1)$-dimensional *time embedding vectors* $(x(t), ..., x(t-m))^T$ and $(y(t), ..., y(t-m))^T$ are stochastically independent. The generalisation on more than two processes is straight forward.

In the following we show, how temporal structures using the concept of independent time embedding vectors can be included in ICA. In particular, we derive an algorithm analog to the FastICA algorithm, first proposed in Hyvärinen and Oja [1997b] and Hyvärinen [1999], to show in a "proof of concept" the applicability of the idea. Although the assumption of independent time embedding vectors is very powerful, one has to deal with numerical instabilities. In order to solve this difficulty the strict requirement of independent time embedding vectors is weakened. Consider uncoupled Markov chains of order one where the dynamics of each process is independent of the other. In this case, the increments of each process, i.e. the change of state within one time step, are independent. Hence, an ICA algorithm which searches for sources with independent increments represents an alternative approach to the classical ICA. In order to apply this approach, one only has to modify the classical ICA algorithms marginally.

4.4.2 Theory

Consider n time series generated by n sources $\mathbf{s} = (s_1, \ldots, s_n)$ with $\mathbf{s}(t) \in \mathbb{R}^n$. The states of $\mathbf{s}(t)$ are recorded at time $t = 1, 2, \ldots$. The sources should be centred

($E\{\mathbf{s}\} = 0$) and should be Markov chains of order m. Furthermore, the sources should be stochastically uncoupled,

$$p(\mathbf{s}(t)|\mathbf{s}(t-1),\ldots,\mathbf{s}(t-m)) = \prod_{i=1}^{n} p(s_i(t)|s_i(t-1),\ldots,s_i(t-m))\,, \qquad (4.84)$$

which means, that the state $s_i(t)$, $(i = 1,\ldots,n)$ of each process s_i is only affected by the process's own past $(s_i(t-1),\ldots,s_i(t-m))$ and not by the past of the other processes $(s_j(t-1),\ldots,s_j(t-m))$, $(i \neq j)$.

This is fulfilled if the joint probability $p(\mathbf{s}(t),\mathbf{s}(t-1),\ldots,\mathbf{s}(t-m))$ factorises into the probabilities of the $(m+1)$-dimensional embedding vectors $(s_i(t), s_i(t-1),\ldots,s_i(t-m))$, $(i = 1,\ldots,n)$, thus we write

$$p(\mathbf{s}(t),\mathbf{s}(t-1),\ldots,\mathbf{s}(t-m)) = \prod_{i=1}^{n} p(s_i(t), s_i(t-1),\ldots,s_i(t-m))\,. \qquad (4.85)$$

From the independence of these embedding vectors, the instantaneous independence of the processes follows immediately, so that one has both uncoupled and (instantaneously) independent sources. If equation (4.85) is fulfilled, then the mutual information of all time embedding vectors is zero as well,

$$I(\begin{pmatrix} s_1(t) \\ \vdots \\ s_1(t-m) \end{pmatrix},\ldots,\begin{pmatrix} s_n(t) \\ \vdots \\ s_n(t-m) \end{pmatrix}) = 0. \qquad (4.86)$$

Equation (4.86) is obviously an extension of the classical ICA assumption where the sources $\mathbf{s}(t)$ fulfil $I(s_1(t),\ldots,s_n(t)) = 0$.

In the following, we concentrate on linear instantaneous mixtures $\mathbf{x}(t)$ of the sources $\mathbf{s}(t)$ with a square matrix $\mathbf{A} \in \mathrm{Mat}(n \times n; \mathbb{R})$ of full rank

$$\mathbf{x}(t) = \mathbf{A} \cdot \mathbf{s}(t). \qquad (4.87)$$

To simplify the calculations, the signals $\mathbf{x}(t)$ can be decorrelated and normalised resulting in a new variable $\mathbf{z}(t) = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{x}(t)$, where $\mathbf{E}$ are the eigenvectors of $\mathbf{C}_\mathbf{x} = E\{\mathbf{x}\mathbf{x}^T\}$ and $\mathbf{D}$ is a diagonal matrix with the corresponding eigenvalues. Then for $\mathbf{z}(t)$ holds $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. This step is called in the literature whitening or sphering of the data (Hyvärinen and Oja [1997b]). Since we assume a linear mixing, we want to find a demixing matrix $\mathbf{W}$, so that the demixed signals

$$\mathbf{y}(t) = \mathbf{W} \cdot \mathbf{z}(t) \qquad (4.88)$$

are equivalent to the sources or at least as independent as possible. Taking our new definition of independent *and* uncoupled sources (equation (4.86)), i.e. independent time embedding vectors, we can rewrite the mutual information in terms

of the negentropy J as given in equation (4.38), but now taking $m+1$ time steps into account

$$I(\mathbf{y}) = -\sum_{i=1}^{n} J(\begin{pmatrix} y_i(0) \\ \dots \\ y_i(m) \end{pmatrix}) + \text{const}\,. \tag{4.89}$$

Here, we dropped the time index t for easy reading and introduced the notation $y_i(k) := y_i\,(t-k)$ with $k = 0, 1, \dots, m$.

In analogy to the classical ICA, where the mutual information is minimised to obtain as independent signals as possible, we are now able to maximise the negentropy of the time embedding vectors. The maximisation of the negentropy is simpler than the estimation of the mutual information, since we don't have to estimate any probability densities but can use the approximations for the negentropy as given in equations (4.39) and (4.40).

In equation (4.89) we have made so far no approximation for finding independent and uncoupled signals. In contrast, other approaches, e.g. AMUSE Molgedey and Schuster [1994]; Tong et al. [1991], SOBI Belouchrani et al. [1997], and TD-SEP Ziehe and Müller [1998], as well as Amari [2000], have to assume, that the autocorrelation functions of the sources are different or that the sources are modelled by different autoregressive processes. Further, we do not have to assume a non-stationary of the sources as it was done in Matsuoka et al. [1995]. Merely, the time embedding vectors of the sources may *not* be gaussian distributed, since under every orthogonal transformation the independent components of a multi-dimensional Gaussian process remain independent. However, in real world problems the distribution of the time embedding vector is usually nongaussian.

4.4.3 Algorithm

In this section we derive an algorithm – close to the FastICA algorithm – to show in a proof of concept, the advantages of the criteria of independent time embedding vectors. Therefore we use equation (4.89) which uses temporal structures of the signals but not limiting it to a special set of sources – except of sources where the (m+1)-dimensional time embedding vectors are gaussian distributed, since these sources cannot be separated as mentioned above. In analogy to the FastICA algorithm by Hyvärinen [1999] a deflation scheme first introduced by Delfosse and Loubaton [1995] is used including a Gram-Schmidt orthogonalisation method.

Since the estimate for the negentropy is only given for an one dimensional random variable, we generalise the function $J(\mathbf{w})$ – equation (4.40), but dropping the odd function, since we assume symmetric probability density functions as also done in the (classical) FastICA algorithm – to m+1 dimensional random vectors in the following way

$$J(\mathbf{w}) = \Big[E\left\{G\left(\mathbf{w}^T\mathbf{z}(0), \dots, \mathbf{w}^T\mathbf{z}(m)\right)\right\} - E\left\{G\left(\nu\right)\right\}\Big]^2 \tag{4.90}$$

but instead of a one dimensional function $G : \mathbb{R} \to \mathbb{R}$ a different function $G(u_0, \ldots, u_m)$, in particular, $G : \mathbb{R}^{m+1} \to \mathbb{R}$ is used. Thus, by considering the time embedding vectors the function is now mapping from a m+1 dimensional space to $\mathbb{R}$, taking the time structures into account. ν is a $(m+1)$-dimensional Gaussian random variable with the same variance as the unmixed signals $\mathbf{y}$. The shape of the function G will be discussed later, for the following derivation it has only to be twice differentiable.

In order to maximise equation (4.90) respecting to $\mathbf{w}$ one has to find the extrema of the first expectation $E\{G(\mathbf{w}^T\mathbf{z}(0), \ldots, \mathbf{w}^T\mathbf{z}(m))\}$. Furthermore the demixed signals $\mathbf{y}$ shall be normalised, thus we obtain the constraint $E\{(\mathbf{w}^T\mathbf{z})^2\} = ||\mathbf{w}||^2 = 1$, which acts as a regularisation. Using the method of the Lagrangian multipliers, by which constrained extrema can be found, we get the function

$$L(\mathbf{w}) = E\left\{G\left(\mathbf{w}^T\mathbf{z}(0), \ldots, \mathbf{w}^T\mathbf{z}(m)\right)\right\} - \lambda(||\mathbf{w}||^2 - 1)\,. \tag{4.91}$$

λ is the Lagrangian multiplier. The constrained extrema are given by the zeros of the first derivative of $L(\mathbf{w})$. To solve this numerically, Newton's method can be applied. For this, the first $\frac{\partial L}{\partial \mathbf{w}}$ and second derivatives $\frac{\partial^2 L}{\partial \mathbf{w}^2}$ of $L(\mathbf{w})$ are needed

$$\frac{\partial L}{\partial \mathbf{w}} = E\left\{\sum_{k=0}^{m} \mathbf{z}(k)\frac{\partial}{\partial u_k} G\left(\mathbf{w}^T\mathbf{z}(0), \ldots, \mathbf{w}^T\mathbf{z}(m)\right)\right\} - 2\lambda\mathbf{w} \stackrel{!}{=} 0 \tag{4.92}$$

$$\text{with } \lambda = \frac{1}{2}E\left\{\sum_{k=0}^{m} \mathbf{w}^T\mathbf{z}(k)\frac{\partial}{\partial u_k} G\left(\mathbf{w}^T\mathbf{z}(0), \ldots, \mathbf{w}^T\mathbf{z}(m)\right)\right\} \tag{4.93}$$

and

$$\begin{aligned}
\frac{\partial^2 L}{\partial \mathbf{w}^2} &= E\left\{\sum_{k,l=0}^{m} \mathbf{z}(k)\mathbf{z}^T(l)\frac{\partial^2}{\partial u_k \partial u_l} G\left(\mathbf{w}^T\mathbf{z}(0), \ldots, \mathbf{w}^T\mathbf{z}(m)\right)\right\} - 2\lambda\mathbf{I} \\
&\approx \sum_{k,l=0}^{m} E\left\{\mathbf{z}(k)\mathbf{z}^T(l)\right\} E\left\{\frac{\partial^2}{\partial u_k \partial u_l} G\left(\mathbf{w}^T\mathbf{z}(0), \ldots, \mathbf{w}^T\mathbf{z}(m)\right)\right\} - 2\lambda\mathbf{I} \qquad (4.94) \\
&\approx \sum_{k,l=0}^{m} E\left\{\frac{\partial^2}{\partial u_k \partial u_l} G\left(\mathbf{w}^T\mathbf{z}(0), \ldots, \mathbf{w}^T\mathbf{z}(m)\right)\right\}\mathbf{I} - 2\lambda\mathbf{I}\,.
\end{aligned}$$

The stabilised Newton method to find the zeros of a function $f(\mathbf{w})$ reads

$$\mathbf{w} \leftarrow \mathbf{w} - \mu \cdot J_f^{-1} \cdot f(\mathbf{w}) \tag{4.95}$$

where J_f is the Jacobian matrix of f, $\mu \in (0, 1]$ is the stabilisation factor and "$\leftarrow$" denotes the update step for $\mathbf{w}$. Inserting $f = \frac{\partial L}{\partial \mathbf{w}}$ and $J_f = \frac{\partial^2 L}{\partial \mathbf{w}^2}$, we obtain the following iteration scheme for finding the unmixing vectors $\mathbf{w}_i$ using the deflation

described in Delfosse and Loubaton [1995]:

$$\mathbf{w}_i \leftarrow \mathbf{w}_i - \mu \cdot \left(\frac{\partial^2 L}{\partial \mathbf{w}^2}\right)^{-1} \cdot \frac{\partial L}{\partial \mathbf{w}} \tag{4.96}$$

$$\mathbf{w}_i \leftarrow \mathbf{w}_i - \sum_{j=1}^{i-1} (\mathbf{w}_i^T \mathbf{w}_j)\mathbf{w}_j \tag{4.97}$$

$$\mathbf{w}_i \leftarrow \frac{\mathbf{w}_i}{||\mathbf{w}_i||}. \tag{4.98}$$

To ensure the constraint $||\mathbf{w}||^2 = 1$, the unmixing vector $\mathbf{w}_i$ is normalised at the end of the iteration step again.

In analogy to the classical FastICA algorithm, in the second derivative of $L(\mathbf{w})$, equation (4.94), we made the approximation that the $(\mathbf{z}(k)\mathbf{z}^T(l))$ are independent of the other part of the addendum. Furthermore, we assumed that the expected value $E\{\mathbf{z}(k)\,\mathbf{z}^T(l)\}$ reduces to the identity matrix if $k \neq l$ as well. If $k = l$ then $E\{\mathbf{z}(k)\,\mathbf{z}^T(l)\} = \mathbf{I}$ holds due to whitening. Using this approximation for $\frac{\partial^2 L}{\partial \mathbf{w}^2}$ one gets good convergence, while using the exact second derivative, the Newton's method has difficulties to converge. This was also observed for the embedding dimension 1 (m=0) when our algorithm is identical to the classical one. We do not have any satisfying explanation yet.

Following the proof for the FastICA algorithm given in Hyvärinen and Oja [1997b], one can also show, that if $\mathbf{w}_j$ is a sought demixing vector for s_j ($s_j = \mathbf{w}_j \cdot \mathbf{z}$), then $E\left\{G\left(\mathbf{w}_j^T\mathbf{z}(0), \ldots, \mathbf{w}_j^T\mathbf{z}(m)\right)\right\}$ has a maximum/minimum under the constraint $E\{(\mathbf{w}_j^T\mathbf{z})^2\} = ||\mathbf{w}_j||^2 = 1$ if the condition

$$\sum_{k,l=0}^{m} E\left\{s_i(k)s_i(l)\right\} \cdot E\left\{\frac{\partial^2}{\partial u_k \partial u_l} G(s_j(0), \ldots, s_j(m))\right\} - \sum_{k=0}^{m} E\left\{s_j(k)\frac{\partial}{\partial u_k} G(s_j(0), \ldots, s_j(m))\right\} \lesseqgtr 0 \tag{4.99}$$

holds for all $i = 1, \ldots, n$, $i \neq j$. Furthermore, each sought demixing vector $\mathbf{w}_j$ is a zero of $\frac{\partial L(\mathbf{w}_j)}{\partial \mathbf{w}} = 0$. Equation (4.99) represents a guideline for choosing the function G, so that one obtains a converging algorithm.

Since we want to approximate the negentropy of a (m+1)-dimensional random variable, we have to take the higher dimensional structure of the probability distribution into account. Using a function like $G(u_0, \ldots, u_m) = u_0^4 + \ldots + u_m^4$ would not yield any improvement because it does not consider any correlation between the time steps. As a starting point we use

$$G(u_0, \ldots, u_m) = F(u_0) + \sum_{k=1}^{m} F(u_0 - u_k) \tag{4.100}$$

with $F(u) = \frac{1}{4}u^4$.

Even when the existence of an extremum for the original sources for this algorithm can be proven, we like to mention the critical point: Due to approximations and the estimation of high dimensional functions, which is necessary for including the information from the temporal structure, the numerical stability of this algorithm is more critical than for the classical FastICA algorithm. However, the proposed idea works with any mixture of sources, as long as the time embedding vectors are nongaussian and independent.

An alternative approach

Note, one can also derive an algorithm analog to the Bell and Sejnowski [1995] algorithm by optimising the mutual information of the time embedding vectors. However, more knowledge about the $(m+1)$-dimensional source distributions is necessary. Even for simple distributions the implemented algorithm works unstable and convergence is often not achieved. Therefore the derivation of this algorithm is leaved out here.

4.4.4 Applications

In an example, we want to demonstrate the application of our algorithm. We have mixed 4 sources with gaussian distributed states at each time step according to our mixing model (4.87), using a random 4x4 matrix with full rank. In particular, the source data set consists of two synthetic data sets (a sinusoidal and triangle signal), one autoregressive process of order one and a voice signal. All sources were made gaussian and each time series had a length of 10,000 samples. The first few hundred samples of the time series of the four sources are shown in figure 4.13.

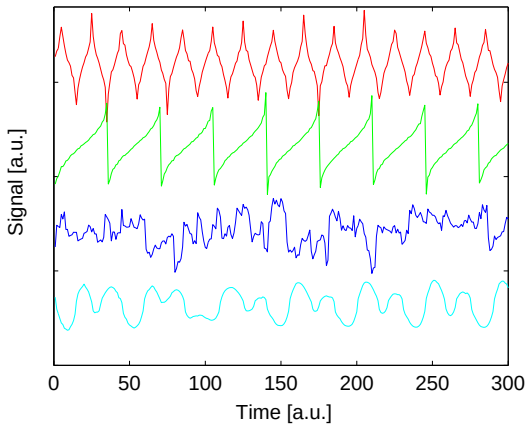

Figure 4.13: Time series used in the example showing the first 300 samples of the four sources having gaussian distributed states at each time step.

To visualise the temporal structure of the signals, in figure 4.14 the scatter-plots of the sources $(s_i(t-1)$ over $s_i(t))$, $i = 1, \ldots, 4$ are given. All four plots show a strong deviation from a 2-dimensional gaussian distribution. In particular, the sinusoidal and the triangle signals, which is in addition asymmetric, can be clearly identified (upper two plots in figure 4.14).

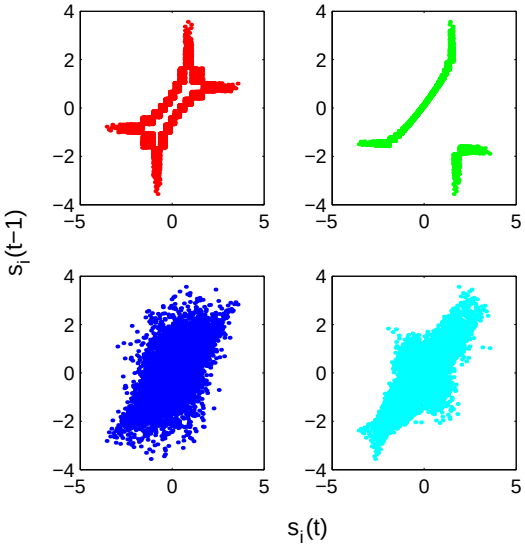

Figure 4.14: Scatter-plot $s_i(t-1)$ over $s_i(t)$ of the sources s_i shown in figure 4.13. The deviation of the scatter-plots from a two dimensional gaussian distribution is clearly visible, showing the temporal structure in the data.

When plotting the histograms of the sources $s_i(t)$ one obtains perfect gaussian distributions (see left column of figure 4.15). In contrast, the histograms of $s_i(t) - s_i(t-1)$ (right column of figure 4.15) deviate strongly from a gaussian distribution. One can interpret the histograms as a projection of the 2-dimensional distribution (figure 4.14) along the diagonal. These deviations are used in the time embedding algorithm to separate the sources.

Using the function given in (4.100) and a time embedding vector of dimension 2 (m=1), we can separate the mixed signals applying the algorithm described in the previous section. The best convergence in our example was achieved for a time delay τ of 10 samples, $(s_i(0), s_i(1)) \rightarrow (s_i(t), s_i(t-\tau))$. To determine the correct time embedding dimension, one calculates the conditional entropy $H(x(t)|x(t-1), ..., x(t-m))$ and varies m until a plateau is reached. A misestimation of m results in ignoring longer time correlation, but still a good separation was achieved in real world applications with small m.

To verify how close the estimated signals are to the original sources, we calculate the matrix $E\{\mathbf{s}\mathbf{y}^T\}$, where $\mathbf{s}$ are the sources and $\mathbf{y}$ the sources estimated by the algorithm. Both signals are normalised to a variance of one so that this matrix has only entries in the range from -1 to 1.

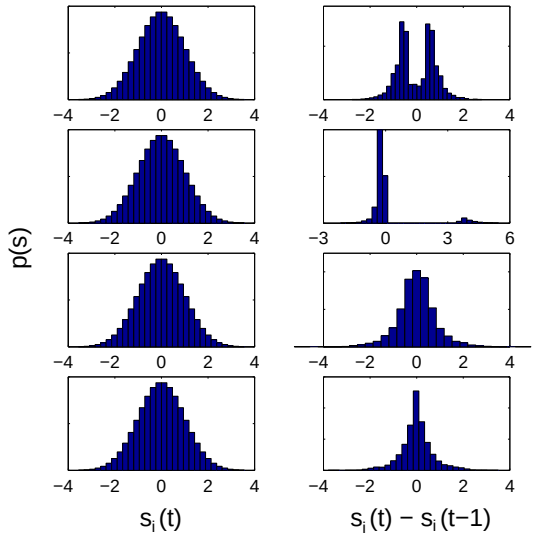

Figure 4.15: Histograms of the sources $s_i(t)$ (left column) and of the increments $s_i(t) - s_i(t-1)$ (right column). The histograms of the increments show the temporal structure of the data whereas the left column shows the gaussian distribution of the states.

A good measure for the accuracy of the estimate is the mean off-diagonal element of this matrix. Due to a finite number of samples, we obtain already a value of 0.036 when calculating $E\{\mathbf{s}\mathbf{s}^T\}$ for the sources. For the estimated sources of our algorithm, we obtain a value of 0.049 (= mean off-diagonal element of $E\{\mathbf{s}\mathbf{y}^T\}$), which can be seen as a good reconstruction of the sources. The signal to noise (S/N) or more accurate the signal to error ratio for our estimated sources is in the range of 200 to 1,000, which indicates also a good reconstruction.

For comparison we made calculations with the TDSEP algorithm where we obtain values of 0.021 for the mean off-diagonal element. This indicates a better performance of the algorithm, but one should realize, that TDSEP, SOBI etc. can only be used in cases with temporal correlations and different autocorrelation functions of the sources. In contrast, our algorithm can in principal deal with any independent sources. Still this algorithm should presently only be seen as a proof of concept to test the presented approach.

Furthermore it should be mentioned, that classical ICA algorithms will give no results for these mixture of sources. For real world applications, e.g EEGs with 21 signals, the proposed algorithm converges and finds sources as expected by physicians.

4.4.5 A new concept for ICA – independent increments

As seen in the previous sections, the independence of the time embedding vectors is very strict but also powerful. Asking for "independent sources" in real world application – as done by an independent component analysis – one does not only want instantaneous independent sources, *but those* for which the current state does not depend on the past of the other sources. In principal, one has to take the complete past of the time series into account in order to guarantee full independence. However, if the sources are Markov chains of order m then it is sufficient to look for sources whose $(m+1)$-dimensional time embedding vectors are independent.

Assuming that the sources are only random variables (Markov chains of order zero), we can reduce the embedding dimension to 1. Hence, we end up at the assumption of the classical ICA (the $s_i(t)$ are independent of each other). Therefore the classical ICA is a special case of the time embedding ICA.

Due to numerical difficulties when trying to find sources with independent time embedding vectors of dimension $m+1$, one would like to weaken the assumption of the time embedding ICA. We want to propose a new assumption for independent sources:

Assume the sources are Markov chains of first order and their time embedding vector with dimension two are independent. Then the increments $(s_i(t)-s_i(t-\tau))$ of the processes s_i, $i = 1, \ldots, n$ are also independent. On the other hand, the increments display the dynamics of the sources. Therefore, why don't we look for independent increments instead of instantaneous independent signals?

This assumption weakens the time embedding ICA, but for Markov chains of first order it is equivalent with the time embedding ICA and even for Markov chains with order $m > 1$ it takes the dynamics of the system in a certain degree into account.

In order to perform an ICA with the aim to seek sources with independent increments the classical FastICA algorithm can be applied on the increments $z_i(t) - z_i(t-\tau)$. In contrast to the time embedding ICA the classical FastICA algorithm is more stable. Using this algorithm to look for independent increments, it gives not only the correct results for Markov chains of order 1 but also for random variables (Markov chains of order 0). For the example above we obtain a mean off-diagonal element of 0.039 and a signal to error ratio of 1,000 to 5,000 using increments, showing the power of the new concept.

Performing ICA using the states of the processes or the increments of the processes depends on the question one wants to be answered. When searching for sources which show little *similarities* (synchronisation), i.e. the sources shall be instantaneous independent, then the classical ICA has to be applied while ICA performed on the increments is more appropriate to seek sources with *independent dynamics*, i.e. which are uncoupled.

4.4.6 Conclusions

We have shown that in the framework of ICA temporal structure in the time series can – in the most general case – be taken into account by using time embedding vectors. This is a generalisation of the classical ICA. Instead of considering only instantaneous independent sources, the sources are assumed to be uncoupled, i.e. the dynamics of the processes are independent of each other. This extension of the ICA model is very powerful and we have shown how to implement it analogously to the FastICA algorithm resulting in an algorithm which we call Fast Time Embedding ICA (FastTeICA). As long as the $(m+1)$-dimensional time embedding vectors are nongaussian, the sources can be separated with the FastTeICA algorithm.

A weakening of the strict requirement of independent time embedding vectors which still takes the dynamics of the processes into account, can be achieved by assuming that the increments of the processes are independent. This is an alternative approach to the decomposition into instantaneous independent sources. For this, standard ICA algorithms can be used just by applying them on the increments of the signals. Thus, the advantage of the better stability in convergence of the standard ICA algorithms compared with the FastTeICA algorithm can be combined with the aim of seaching for dynamically independent sources.

4.5 Application to biomedical data

The concept of describing observed signals by a mixture of underlying hidden sources is an intuitive approach and suggests to use it in biomedical applications. Not only the underlying sources and their waveform can be of interest, but also the mixing process (in the linear case the mixing matrix) can be useful to get a better understanding of the origin of the measured signals

In this section, we want to explore how far independent component analysis can be applied to the biomedical data sets (neuromonitoring and electroencephalography data) presented in the first chapter and how the method can improve the interpretation of medical measurements in the clinical environment.

4.5.1 Neuromonitoring data

In the chapter of time series analysis we have seen a high correlation in the low frequency range between some of the measured neuromonitoring signals. This linear correlation suggests, that an underlying source generates the signals and is mixed linearly with different weight into the measured signals.

But, we have also seen, that this correlation is not stable over time, i.e. the data is non-stationary, which is a main assumption when using random variables with constant probability density functions. By using ICA on smaller windows, we can overcome this problem, but a second more severe problem arises with the neuromonitoring data.

As an example, the plotted time series of the Invos sensors in figure 2.1 and their power spectra in figure 2.2 illustrate the problem. These measured signals can be divided into different contributions. The signals originating from the patients oxygen content in the blood is overlaid by two other contributions, a long term drift – clearly visible in figure 2.1 – and a white noise signal. Therefore not only two sources generate the two signals, but in total five different sources (the oxygen content signal, two noise signals from the sensor device and two long term drifts).

Thus, we have an overcomplete ICA problem with non-stationary signals bringing us to an unsolvable problem. The true signals originating from the oxygen content in the blood are not recoverable by ICA. But as we have seen, time series analysis can be used to uncover the original signal by using other techniques.

4.5.2 Electro-Encephalography (EEG) data

EEG data, as shown in figure 1.9, is a highly multimodal data set and often more than 21 signals are recorded. In the medical research up to 256 electric potentials are recorded from the scalp of patients to be examined. Therefore it seems to be an interesting question, what we can expect when applying ICA to EEG data.

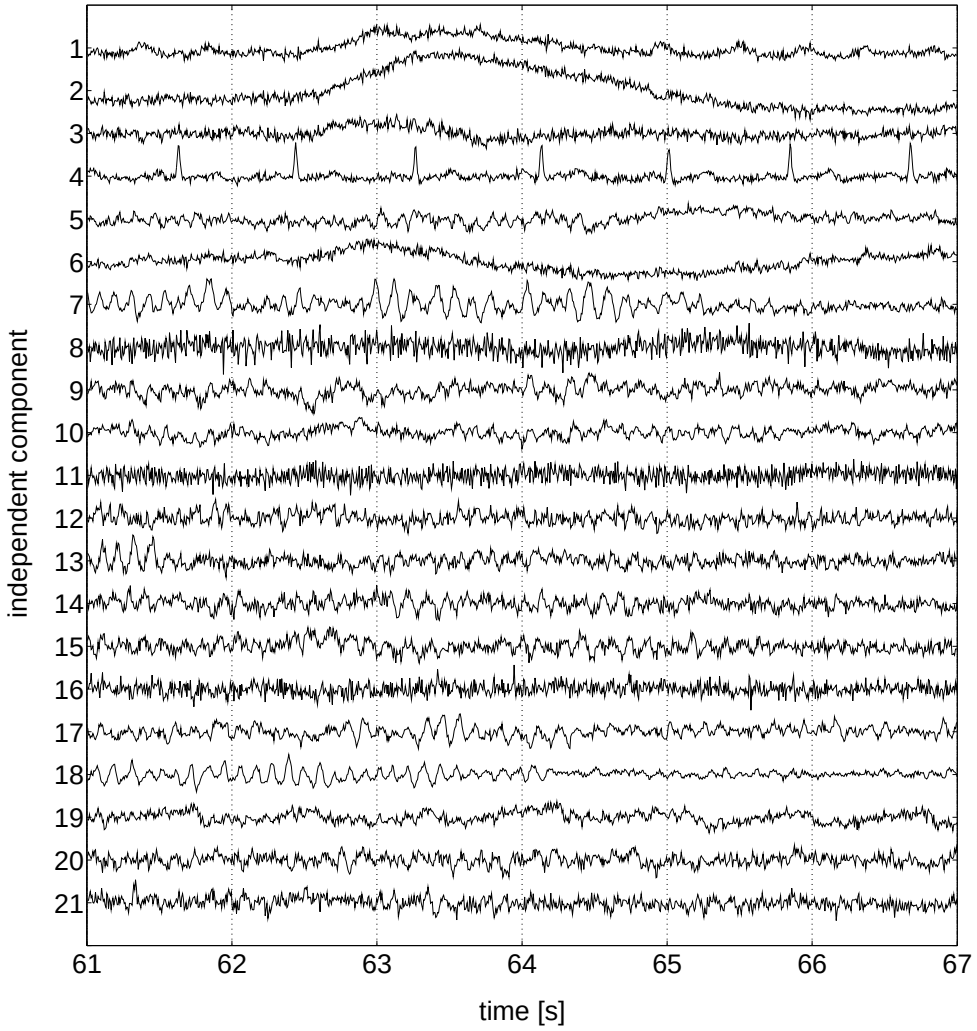

Figure 4.16: Independent components (ICs) found by the FastICA algorithm from the EEG data shown in figure 1.9. An interpretation of the some selected ICs can be given as following: #2 - eye movement, #4 - heart beat, #7 - α-wave activity ($\sim 8\,\mathrm{Hz}$), #11 - noise, #19 - pulse amplitude from the heart.

To give the reader a simple and intuitive example for an EEG measurement, we will use the following analogy. Imagine you are listening to a football game from outside the stadium using many microphones. The goal is to understand, what happens inside the stadium. Surely the main behaviour will be recognisable – the goals of one or the other team can be identified by the different songs they sing and even where the supporters are located. But trying to identify a single person or smaller group of people will not be possible, we only see an average behaviour. More microphones will only increase the spatial resolution. But what will happen, if an airplane flys above the stadium? This additional (noise) signal

will also be recorded by the microphones. Surely one would like to remove such signals from the recordings, since they will give no information about the events inside the stadium.

Different ICA algorithms were applied to EEG signals in the past, as for example done in Puntonet et al. [2000] and Bauer et al. [2000] by using geometric ICA, to reveal deeper information and identify possible tumour signals. To get an impression of the resulting independent components found by an ICA algorithm (FastICA), the signals are shown in figure 4.16.

Very different signals arise from such an analysis, but a trained eye can sort them into different categories. The first three independent components in figure 4.16 can be identified as artifacts from the eye movement due to their slow changes, but also noise signals (as for example independent component #8 and #11) are identifiable. The typical alpha-wave activity ($\sim 8\,\mathrm{Hz}$) of the brain is also separated into an independent component (#7). The other signals originating from the brain or from other external sources are difficult to identify.

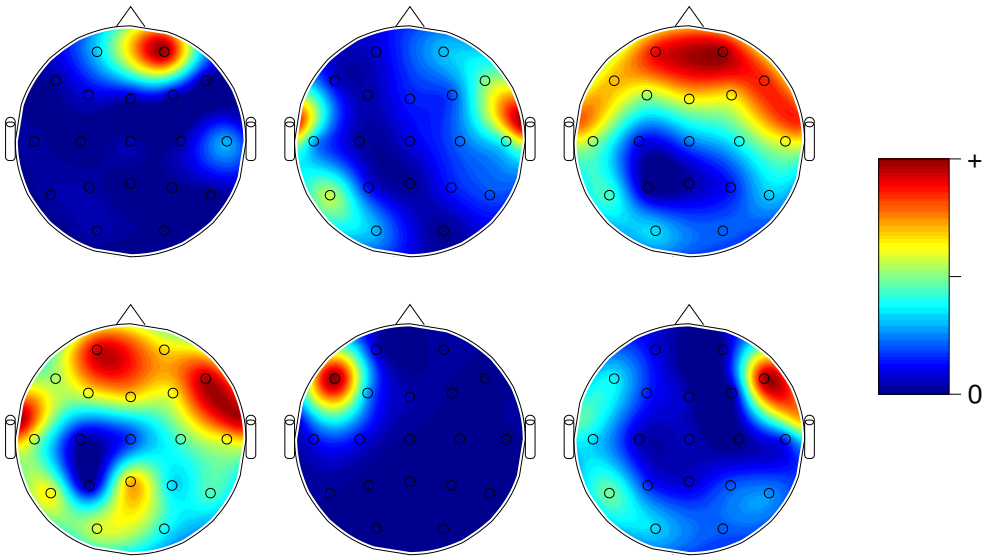

Figure 4.17: Origin of the independent components (ICs) – shown in figure 4.16 – projected on to the scalp using the absolute values of the entries in the recovered mixing matrix. The positions of the electrodes are represented by the circles. Some selected ICs (from top left to bottom right): #2 - eye movement, #4 - heart beat, #7 - α-wave activity ($\sim 8\,\mathrm{Hz}$), #10 - unknown distortion (random origin of IC), #11 - noise and #19 - pulse amplitude from the heart.

Additionally to the waveform, the algorithm allows us to determine the origin of the signals by investigating the entries of the recovered mixing matrix. Each entry corresponds to the amount of signal of the independent component which is mixed into the recorded channel (the recording of each electrode). To visualise

these entries and the origin of the signals, the absolute value of the entries is used to generate a density plot mapped on to the scalp, as is shown in figure 4.17.

With these two different informations – the waveform and the origin of the sources – we can try to determine the true nature of the signal. Either the signals originate from internal events (like the alpha wave activity corresponding for example to the songs in the stadium) or external events. The signals originating from external events are often called artifacts, since they give no information about the investigated system (as for example the electric activity of an eye movement which in the analogy corresponds to the noise of the airplane flying above the stadium).

Independent components = independent biomedical signals?

An open question which needs to be answered when applying ICA to biomedical signals is: how much can we trust in the independent components found by an ICA algorithm? Are these independent components true biomedical signals or artifacts from the algorithm?

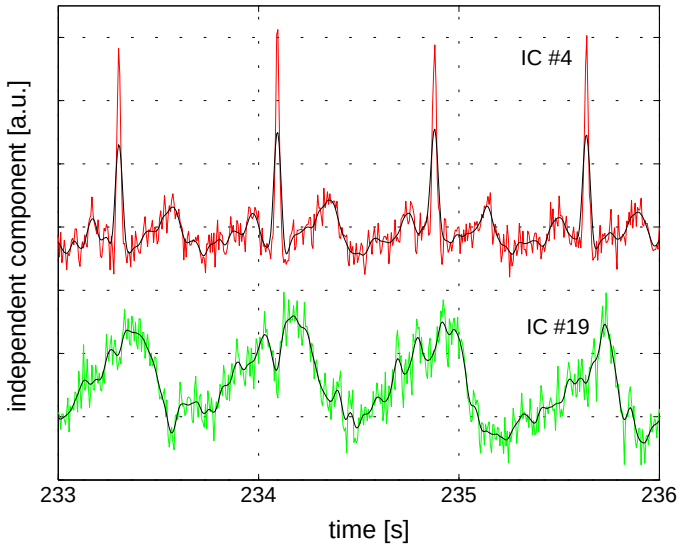

Figure 4.18: Independent components #4 (heart beat – red) and #19 (pulse amplitude from the heart – green) from figure 4.16. Plotted is the raw data and the data smoothed with a parabolic kernel of ± 5 samples $= \pm 30$ ms (black line). By visual inspection both signals show a high correlation suggesting that the heart beat occurs 0.1 s earlier than the maximum of the pulse amplitude.

The following example will illustrate that algorithms can generate signals, which are *statistically* independent, but are in the medical sense not independent

biomedical signals. The independent components (ICs) #4 and #19 from figure 4.16 are such an example. For a better comparison both signals are plotted enlarged in figure 4.18. The IC #4 corresponds clearly to the (electrically measurable) heart beat signal recorded close to the ears, as it is well known. But IC #19 is a similar looking signal, which turns out to be the pulse of the heart beat from an arterial close by the right ear – see also the projection in figure 4.17.

To understand the origin of these biomedically correlated signals, we will now analyse both signals in more detail. The signals generated by the algorithm are all statistically independent, this can be shown by calculating the cross-correlation and mutual information between the signals. Figure 4.19 illustrates this for both signals while also investigating the behaviour of both measures for time shifts between the signals.

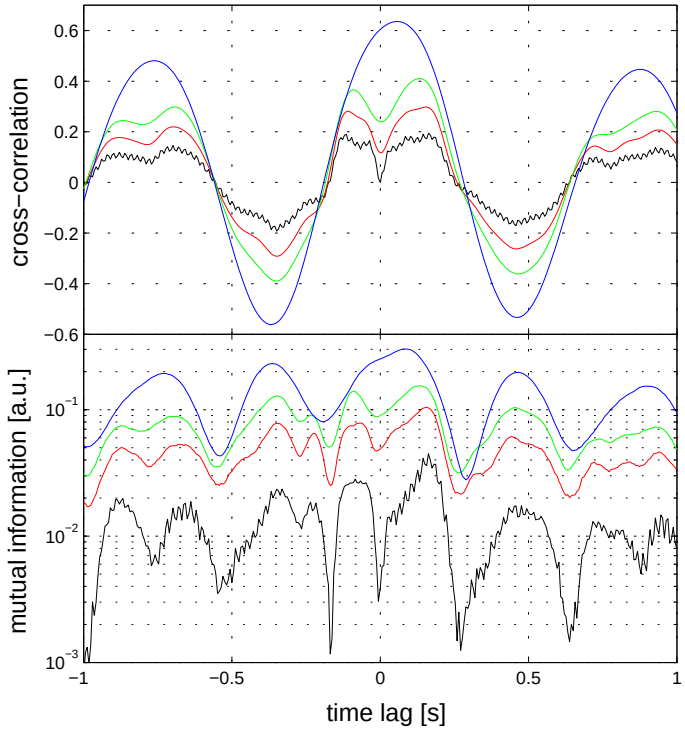

Figure 4.19: Time lagged cross-correlation and mutual information between the independent components #4 (heart beat) and #19 (pulse amplitude from the heart). Calculation performed with the raw data (black line) and with data smoothed with a parabolic kernel of ±5 samples = ±30 ms (red), ±60 ms (green) and ±180 ms (blue). At time lag = 0 s the raw data of the ICs have no cross-correlation and very low mutual information. But with an increase of the smoothing, both values rise drastically. With a smoothing of ±180 ms a maximum of the correlation between the signals is obtained at a time shift of 0.1 s, as was also visible in figure 4.18. Conclusion: "numerical noise" made the signals independent.

First of all, the raw data of the signals (black line) shows at time lag = 0 s no correlation and a very low mutual information. The small positive value of the mutual information originates from the estimation procedure. For more details on estimating entropies and mutual information see the PhD thesis of Kaiser [2002]. But already for small time lags the correlation and the mutual information rise quickly. In particular the cross-correlation over the time lag of ±1 second shows a sinusoidal behaviour, except for time lags close to zero where it drops quickly to zero. This suggests, that the algorithm forced the resulting independent components to be uncorrelated and instantaneous independent.

But this was only achieved by adding artificial "numerical noise" to the signals. When the signals are smoothed with increasing width of the kernel – as explained in the chapter of time series analysis using equation (2.2) – the signals show also an increasing correlation and mutual information. Especially at time lag zero the cross-correlation and the mutual information rise quickly with the kernel width.

For a smoothing window of ±180 ms – removing nearly all noise – the cross-correlation as well as the mutual information reach a maximum at a time lag of around +0.1 s. That means that the electrical heart beat pattern occurs 0.1 s earlier than the pulse amplitude measured near the right ear. This value fits well with a visual inspection of the time series in figure 4.18 and can also be supported by medical observations.

Thus, we have shown, that not all independent components found by an ICA algorithm are always independent in the biomedical sense. The ICA algorithm can – under some circumstances – generate "quasi" independent signals which are just artifacts from the algorithm. The requirement of producing as many independent components as measured signals can lead to such consequences. Especially for a larger number of measured signals one has to be careful with the resulting independent signals.

Removal of artifacts from Electro-Encephalography (EEG) data

As we have seen in the previous section, some of the independent components correspond to external influences disturbing the EEG recordings. Muscle activity such as the eye movements strongly disturbs the EEG recordings. But also noise from some of the electrodes can make the EEG recordings difficult to interpret.

From the medical point of view, these external influences can easily and accurately be identified by their waveform and their origin on the scalp. As in our analogy, we would like to remove these external signals (the sound from an airplane flying above the stadium) to make it easier for the physician to examine the true electric potential originating from the brain.

Therefore we suggest to use ICA *not* to try to identify electric patterns originating from the brain, *but* to identify external influences and remove them from the data. Since we can fully reconstruct the data from the independent compo-

nents (using the (de)mixing matrix) we can also reconstruct an artifact free EEG recording by using only those independent components, which are not identified as an external influence.

For our EEG example from figure 1.9, we identified in figure 4.16 the following independent components as artifacts: #1,2,3,6 (originating from eye movements), #4 (heart beat signal respectively the ECG) and #8,11,16 (noise signals). Reconstructing the EEG recording without those independent components, we obtain the EEG shown in figure 4.20.

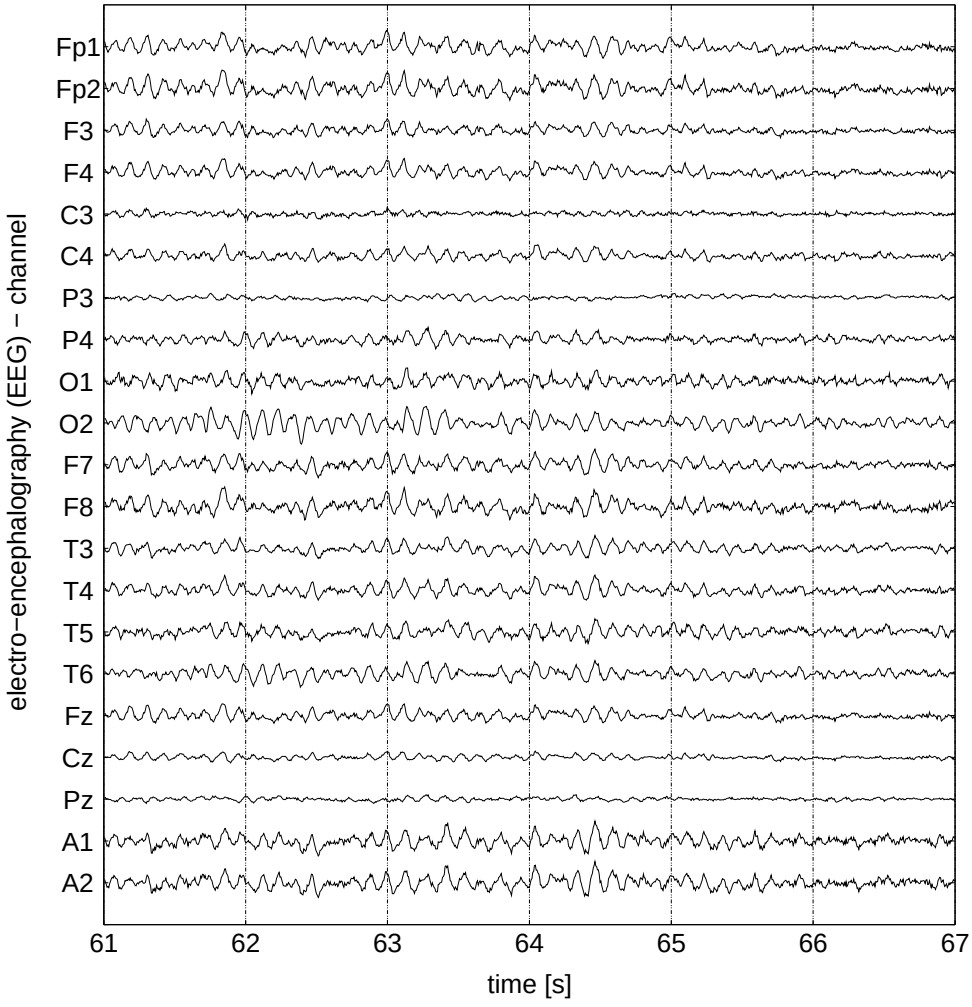

Figure 4.20: EEG data cleaned by associating independent components (ICs) with artifacts and using for the reconstruction of the data only non-artifact ICs. The ICs identified as artifacts were #1,2,3,6 (origin from eye movement), #4 (heart beat) and #8,11,16 (noise signals).

Compared to the original recording, which was already a quite good (artifact-free) EEG recording, nearly all artifacts were removed leaving only those signals,

which from our and the physicians perspective originate from the electric activity in the brain. The author is not aware of any other method which can remove artifacts so clearly without disturbing the rest of the signals. Smoothing procedures to remove noise or nonlinear time series analysis techniques don't work that well on high dimensional data.

4.6 Conclusions

In this chapter we have presented two different approaches to find independent components in a mixture of signals.

The geometrical approach uses scatter plots and their behaviour under transformation, to recover the unknown sources and the mixture matrix in the case of the blind source separation problem. With the presented derivation of the theoretical fundaments of the neural geometric learning algorithm (first presented by Puntonet et al. [1995]), a new algorithm based on histograms (FastGeo) could be derived. The computational effort was reduced by two orders of magnitude and also the accuracy was enhanced, making it comparable in performance to other ICA algorithms. In particular the easy implementation of the algorithm makes it a good choice for two dimensional problems.

The information theoretical approach using the temporal structures in time series was motivated by the time series analysis technique, which uses time embedding vectors to reconstruct the underlying system. With the concept of independent time embedding vectors, restrictions of other algorithms based on time structures could be weakened. Any linear quadratic mixture of sources can be recovered as long as the complete time embedding vector is nongaussian distributed. The derived algorithm (FastTeICA), which was derived in analogy to the classical FastICA algorithm, shows in a proof of concept the applicability to temporal correlated signals. Furthermore, the theoretical description lead to a new concept of ICA which uses independent increments (respectively independent dynamics) instead of searching for instantaneous independence (removing similarities between the signals). Depending on the question which shall be answered by ICA, one of the two concepts should be used. Though, in real world applications the question for independent dynamical processes often arises, wherefore the concept of independent increments should be favoured in such cases.

The application of ICA in biomedical data analysis demonstrates both the power and weakness of the concept. In the case of the neuromonitoring data, the mixture of signals represents a so-called overcomplete problem (more unknown sources than signals, which can not be solved uniquely), while the electroencephalography (EEG) data seem to be an ideal application. The artifacts in the data can easily be identified and by reconstructing the data without these signals, a cleaned EEG recording can be produced. Still we have to keep in mind, that independent components do not always have to be independent biomedical signals. Furthermore not every independent component must correspond to a biomedical process.

Conclusions and Outlook

In the following, a summary of the conclusions of the preceding chapters will be given. For more details we refer to the conclusions in the corresponding chapters. The second part of this chapter gives an outlook on further developments in the field of biomedical data analysis, in particular how enhanced biological/physical models could be developed.

Conclusions

In the presentation of the time series analysis applied to neuromonitoring data we have shown that well known methods with nearly no assumptions on the data can reveal interesting relations.

The correlation analysis in the frequency domain and the power spectrum of the data uncovered three fundamental characteristics of the data. First the signals originating from the patient behave as a one-over-f-type process and represent the only correlated part in the time series of the neuromonitoring data. These so-called selfsimilar time series are typical processes observed in many natural systems. In contrast to that, the two other contributions, a long term drift of the signal and a white noise contribution in the higher frequency band, seem not to contain any valuable information from the medical point of view.

The correlation between the time series representing the oxygen supply was well established in a large group of patients. But also other more surprising couplings were seen between the time series describing the hydrodynamical processes and the oxygen supply. With some basic medical knowledge, we gave an intuitive description of the origin of these couplings. In general, these interconnections between the measured brain status parameters can be used for statistical tests to indicate a change in the state of health of the patient.

In a next step we adapted a hydrodynamical model to the special requirements necessary in neurosurgical applications. This model reproduces all patterns known from medical experiments such as the autoregulation and the pressure-volume curve. To be able to reproduce the brain status parameters measured at the neurosurgical intensive care unit, we added a mathematical description of the oxygen supply of brain matter based on the theory of the Krogh cylinder. This *combined model* enables us to describe and understand in a far more detailed way the couplings between the measured signals, as was investigated by the correlation

analysis, as well as their generating processes. Furthermore, other parameters as for example the cerebral blood flow or the autoregulation capability of the brain, which are only hardly measureable in the clinical environment, can now be estimated by the model. This gives the physicians an important indication of the state of health of the patients and the possibility of a more accurate treatment.

All fundamental biological processes used in this model are based on pure physical descriptions, making us confident to reproduce the mechanisms in the brain correctly. The comparison of the simulations of the model with the measured data, in particular the relation between the time series describing the oxygen supply, supports the model.

An alternative approach to analyse biomedical data is the independent component analysis (ICA). First we presented in an elaborate theoretical discussion the derivation of two new algorithms based on very different assumptions. The geometrical approach uses the properties of the scatter plots under transformation, making it a simple and intuitive algorithm for low dimensional problems. From the presented theoretical framework, where a proof for the correct solution of the method is given, a new fast histogram based algorithm is derived. The second approach is based on information theoretical considerations including the knowledge about the time structures in the data. A method known in the nonlinear time series analysis is used to overcome limitations of other ICA algorithms using time structures. The approach of independent time embedding vectors lead further to a new concept of ICA. Instead of looking for instantaneous independence, i.e. removing similarities between the signals, the concept of independent dynamics respectively independent increments is proposed. In many real world applications these independent dynamical processes are of interest.

The results of the application of ICA to biomedical data demonstrate both the power and the weakpoints of the concept based on ICA. While ICA is not well suited for the analysis of neuromonitoring, due to the high number of unknown sources generating the measured signals, the application to the EEG data showed the excellent data enhancement possibilities of ICA. Artifacts respectively external influences in the recordings can be removed in a very accurate way without disturbing the data. This makes it a superior data processing method compared to others known for this purpose.

However, not all independent components extracted by the ICA algorithms have to be independent in the biomedical sense. In an example we have shown that the extracted electrical heart beat pattern was correlated to a signal similar to the pulse amplitude of a heart beat after marginal smoothing. Therefore the independent components have to be verified to their biomedical significance.

Outlook

From the medical point view, the presented combined model can support the detection of changes in the patient's physiological state. Possible clinical applications could be...

- ...the determination of the *state of the autoregulation* by the analysis of the coupling between the arterial blood pressure and the parameters of the oxygen supply ($p_{ti}O_2$ and HbO_2) as it was shown in time series analysis and described by the model. For the medical treatment, the knowledge of the state of the autoregulation is vital as already emphasised by Golding et al. [1999]. Questions on the behaviour of the system on possible changes of the ABP to higher or lower values can be answered by the model.

- ...the analysis of the current status of the patient by fitting the parameters of the model to the measured data. With these parameters, a prediction of the brain status for different conditions is then possible. In particular, the model can describe under which conditions and in which form (linear or exponential) the *intracranial pressure* (ICP) rises. Correspondingly the treatment of the patient can be adapted.

- ...the calculation of the *metabolic brain status parameters* – in particular the oxygen supply including the oxygen content in blood and tissue. This is only possible since we obtain from the hydrodynamical model the cerebral blood flow (CBF), which is vital for the calculation of the oxygen transport. A monitoring of the oxygen measurements and a comparison with the model predictions can give an indication on the change of the patients physiological state.

In the context of the physical description of the model, further extensions are imaginable. In particular the increase of the spatial resolution in the model could be of interest. Two approaches could be followed:

- A "top/down" approach by subdividing the model into smaller subunits taking the local anatomic structure of the brain into account. A swelling or injury of the brain tissue could be described by dedicated subunits. Local measurements obtained by a set of devices like the Invos sensor could then reveal a better insight into the regional processes occurring in the brain, as for example the local velocity in the blood vessels and the diffusion of the oxygen into the tissue. Such – in some sense – phenomenological description would of course need a more detailed incorporation of the knowledge about the anatomy of the human brain.

- Or a "bottom/up" approach by using computer tomography or nuclear magnetic resonance imaging to reconstruct the true anatomy of the blood

vessels in the human head. A full hydrodynamical calculation, i.e. solving the Navier-Stokes equation, coupled to a oxygen diffusion model as described in this work could then be performed. But the questions of how to implement the autoregulation mechanisms in the brain and the calculation of the long term predictions have to be answered. If this could be achieved, we would have the ideal model.

In addition, other chemical substances beside the oxygen could be taken into account. In particular the autoregulation mechanism, which regulates the blood flow in the brain is influenced by a variety of chemical substances and neural processes. For example the CO_2, which originates from the nerve cells, determines the pH-value of the blood and finally influence the autoregulation process. A description which takes all important chemical substances into account, would be favourable but also highly complex. A good balance between the phenomenological and the pure chemical/physical description has to be found.

Nevertheless, the presented model are a first step in a long iterative process to develop sophisticated diagnostic tools for neurosurgical intensive care units. In particular the fitting of the free parameters to the current patients state defines the next step in this challenge.

Apart from the "classic" analysis techniques, a method first used in nuclear physics to describe the energy spectra of the nucleus, has been used for the analysis of the EEG data, namely the random matrix theory (RMT). Without the knowledge about the interactions in the core of the atom, the resonances are measured and statistically analysed. In electro-encephalography (EEG) a similar problem arises, only little is known about the interactions of billions of neurons in the brain. However, one can measure the electric potential on the scalp of the head. Correlations between the measured signals can be written in a matrix and the fluctuations of the eigenvalues (corresponding to the resonances of the nucleus) can be compared to the predictions of RMT. Interestingly, a behaviour similar to the gaussian orthogonal ensemble (GOE) is found in the EEG data, i.e. the entries in the correlation matrix seem to be random and gaussian distributed, but a perfect agreement to RMT is not achieved.

The reason for this is not yet clear and will require further investigation. Probably a preprocessing step to remove the artifacts with the ICA method as well as a higher temporal resolution of the EEG data is necessary, to be able to understand in a more detailed way the origin of this behaviour. Still, this analysis method is an interesting approach since no prior knowledge about the system is necessary.

In general, the removal of artifacts by the ICA method is a promising application for devices recording electro-encephalography (EEG) data. In clinical environments such a method could enhance the quality of the EEGs and simplify the inspection of the EEG recordings for the physicians. A possible implementation of this method into EEG recording devices is under consideration.

Appendix A

Mathematical tools and proofs

A.1 Correlation in the frequency domain

In the following the correlation coefficient between two frequency filtered time series $x_f(t)$ and $y_f(t)$ will be derived. To filter the time series $x(t)$ and $y(t)$ in frequency domain, one has to fourier transform them by

$$\tilde{x}(\omega) = \frac{1}{\sqrt{2\pi}} \int x(\tau) e^{i\omega\tau} \mathrm{d}\tau \tag{A.1}$$

and multiply them with a (here rectangular) windowing function

$$S(\omega) = \begin{cases} 1 & \omega_1 \leqslant |\omega| \leqslant \omega_2 \\ 0 & \text{else} \end{cases} . \tag{A.2}$$

The filtered fourier coefficients read as $\tilde{x}_f(\omega) = \tilde{x}(\omega) S(\omega)$ and the filtered time series can be written as

$$x_f(t) = \frac{1}{\sqrt{2\pi}} \int \tilde{x}(\omega) S(\omega) e^{-i\omega\tau} \mathrm{d}\tau. \tag{A.3}$$

For the correlation coefficient between the frequency filtered time series (normalised by the standard deviations) we write

$$c_{x_f, y_f} = \frac{\langle x_f y_f \rangle - \langle x_f \rangle \langle y_f \rangle}{\sqrt{\langle x_f^2 \rangle - \langle x_f \rangle^2} \sqrt{\langle y_f^2 \rangle - \langle y_f \rangle^2}} \tag{A.4}$$

where $\langle . \rangle$ is defined as

$$\langle f \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t) \, \mathrm{d}t \tag{A.5}$$

For simplicity and clarity of the equations, we will leave out the limits $T \to \infty$ in the equations, however still taking it in the calculations into account. Furthermore we can remove the mean from the time series before calculating the correlation, i.e. $\langle x_f \rangle, \langle y_f \rangle = 0$, since this simplifies the correlation coefficient to

$$c_{x_f,y_f} = \frac{\int x_f y_f \,\mathrm{d}t}{\sqrt{\int x_f^2 \,\mathrm{d}t \int y_f^2 \,\mathrm{d}t}}. \tag{A.6}$$

In the following, the numerator will be written in terms of the fourier coefficients instead of the time series itself, which leads to a simpler equation

$$\int x_f(t) y_f(t) \,\mathrm{d}t = \frac{1}{2\pi} \iiint \tilde{x}(\omega) S(\omega) e^{-i\omega t} \, \tilde{y}(\omega') S(\omega') e^{-i\omega' t} \,\mathrm{d}\omega \,\mathrm{d}\omega' \,\mathrm{d}t \tag{A.7}$$

$$. \tag{A.8}$$

The integration dt over $e^{-i(\omega+\omega')t}$ simplifies to $\delta(\omega + \omega')$

$$= \frac{1}{2\pi} \iint \tilde{x}(\omega) \tilde{y}(\omega') \, S(\omega) S(\omega') \, \delta(\omega + \omega') \,\mathrm{d}\omega \,\mathrm{d}\omega', \tag{A.9}$$

after the integration $\mathrm{d}\omega'$ over $\delta(\omega + \omega')$ one substitutes ω' by $-\omega$

$$= \frac{1}{2\pi} \int \tilde{x}(\omega) \tilde{y}(-\omega) \, S(\omega) S(-\omega) \,\mathrm{d}\omega. \tag{A.10}$$

Since $S(\omega)$ is symmetric, i.e. $S(\omega) = S(-\omega)$, and x(t) and y(t) are real valued functions, i.e. $\tilde{y}(-\omega) = \tilde{y}^*(\omega)$ where $\tilde{y}^*$ is the conjugate complex of $\tilde{y}$, and the frequency window is chosen to be rectangular (ω_1 and ω_2 are the limits as defined in A.2), the equation simplifies to

$$= \frac{1}{2\pi} \int_{-\omega_2}^{-\omega_1} \tilde{x}(\omega) \tilde{y}^*(\omega) \,\mathrm{d}\omega + \frac{1}{2\pi} \int_{\omega_1}^{\omega_2} \tilde{x}(\omega) \tilde{y}^*(\omega) \,\mathrm{d}\omega. \tag{A.11}$$

Exchanging the limits of the first integral gives the final result for the numerator

$$= \frac{1}{2\pi} \int_{\omega_1}^{\omega_2} \tilde{x}^*(\omega) \tilde{y}(\omega) + \tilde{x}(\omega) \tilde{y}^*(\omega) \,\mathrm{d}\omega. \tag{A.12}$$

Therefore one obtains for the correlation coefficient

$$c_{x_f,y_f} = \frac{\frac{1}{2} \int_{\omega_1}^{\omega_2} \tilde{x}^* \tilde{y} + \tilde{x} \tilde{y}^* \,\mathrm{d}\omega}{\sqrt{\int_{\omega_1}^{\omega_2} \tilde{x} \tilde{x}^* \,\mathrm{d}\omega \int_{\omega_1}^{\omega_2} \tilde{y} \tilde{y}^* \,\mathrm{d}\omega}}. \tag{A.13}$$

Note: This formula is only strictly valid for long time series and zero mean of the time series $x(t)$ and $y(t)$!

A.2 Proof: Uniqueness of geometric ICA

We now describe precisely how to recover the matrix A after the axes, which span the observation space, have been extracted successfully from the data using for example a (neural) geometric learning algorithm. This proof is given in Theis et al. [2003b] and in a shorter version in Theis [2002].

So we assume that each $\mathbf{w}_i$ is linear-dependent on one $\mathbf{A}\mathbf{e}_j$. Let

$$\Lambda := \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid \exists i\, x_i > 0, x_j = 0 \text{ for all } j \neq i\} \tag{A.14}$$

be the set of positive coordinate axes and denote with $\Lambda' := \mathbf{A}\Lambda$ the image of this set under the transformation $\mathbf{A}$. Note, due to $\mathbf{A}$ being bijective Λ' intersects the unit $(n-1)$-sphere

$$S^{n-1} := \{x \in \mathbb{R}^n \mid |x| = 1\} \tag{A.15}$$

in exactly n distinct points $\{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ and that those $\mathbf{p}_i$'s form a basis of $\mathbb{R}^n$

Define the matrix $\mathbf{M}_{\mathbf{p}_1,\dots,\mathbf{p}_n} \in \mathrm{Gl}(n)$, where $\mathrm{Gl}(n) := \{\mathbf{W} \in \mathrm{Mat}(n \times n, \mathbb{R}) \mid \det(\mathbf{W}) \neq 0\}$, to be the linear mapping of $\mathbf{e}_i$ onto $\mathbf{p}_i$ for $i = 1, \dots, n$, i.e.

$$\mathbf{M}_{\mathbf{p}_1,\dots,\mathbf{p}_n} = (\mathbf{p}_1 | \dots | \mathbf{p}_n)\,. \tag{A.16}$$

This matrix thus effects the linear coordinate change from the standard coordinates $(\mathbf{e}_i)$ to the new basis $(\mathbf{p}_i)$. Note that for this coordinate transformation the following lemma holds:

Lemma A.1. *For any permutation $\sigma \in \mathcal{S}_n$, the two matrices $\mathbf{M}_{\mathbf{p}_1,\dots,\mathbf{p}_n}$ and $\mathbf{M}_{\mathbf{p}_{\sigma(1)},\dots,\mathbf{p}_{\sigma(n)}}$ are equivalent.*

Proof. The relation

$$\mathbf{M}_{\mathbf{p}_1,\dots,\mathbf{p}_n} = \mathbf{P}\mathbf{M}_{\mathbf{p}_{\sigma(1)},\dots,\mathbf{p}_{\sigma(n)}} \tag{A.17}$$

holds with $\mathbf{P}$ representing a permutation matrix and $\mathbf{M}_{\mathbf{p}_{\sigma(1)},\dots,\mathbf{p}_{\sigma(n)}}$ designating the matrix $\mathbf{M}_{\mathbf{p}_1,\dots,\mathbf{p}_n}$ with permuted columns according to $\sigma \in \mathcal{S}_n$. Note that any additional permutation does not interfere with the equivalence condition given above. □

Therefore $\mathbf{M}_{\mathbf{p}_1,\dots,\mathbf{p}_n}$ is equivalent to $\mathbf{M}_{\mathbf{w}_1,\dots,\mathbf{w}_n} = (\mathbf{w}_1|\dots|\mathbf{w}_n)$. Now we can state the following theorem:

Theorem A.2 (Uniqueness of the geometric method). *The matrix $\mathbf{M}_{\mathbf{p}_1,\dots,\mathbf{p}_n}$ is equivalent to $\mathbf{A}$.*

Proof. By construction of $\mathbf{M}_{\mathbf{p}_1,\dots,\mathbf{p}_n}$, we have

$$\mathbf{M}_{\mathbf{p}_1,\dots,\mathbf{p}_n}(\mathbf{e}_i) = \mathbf{p}_i = f(\mathbf{e}_i) = \frac{\mathbf{A}\mathbf{e}_i}{|\mathbf{A}\mathbf{e}_i|}, \tag{A.18}$$

so there exists a $\lambda_i \in \mathbb{R} \setminus \{0\}$ such that

$$\mathbf{M}_{\mathbf{p}_1,\ldots,\mathbf{p}_n}(\mathbf{e}_i) = \lambda_i \mathbf{A}\mathbf{e}_i. \tag{A.19}$$

Setting $L(e_i) := \lambda_i e_i$ yields an invertible diagonal matrix $\mathbf{L} \in \mathrm{Gl}(n)$, such that $\mathbf{M}_{\mathbf{p}_1,\ldots,\mathbf{p}_n} = \mathbf{L}\mathbf{A}$. This shows the claim. □

Having proven this theorem, the following corollary follows immediately:

Corollary A.3. *Under the assumption of successful convergence of the geometric learning algorithm, the matrix* $\mathbf{M}^{-1}_{\mathbf{w}_1,\ldots,\mathbf{w}_n}$ *solves the BSS problem.*

A.3 Proof: Existence of only two fixed points in geometric ICA

The proof of the following theorem is given in Theis et al. [2003b] and the PhD thesis of Theis [2002].

Theorem A.4. *Assume that the sources* S_i *are unimodal, symmetric and that* $\rho_1 = \rho_2$*. Assume that* $\rho_Y|[0,\pi)$ *with* $\mathbf{A} = \mathrm{id}$ *has exactly two local maxima and two local minima. Then there exist only two fixed points of geometric ICA in the expectation.*

Proof. The same calculation as in the proof of theorem 4.5 shows that we can assume $\mathbf{A} = \mathrm{id}$ without loss of generality. Then by theorem 4.4 and by $\rho_1 = \rho_2$, the two pairs $\{0, \frac{\pi}{2}\}$ and $\{\frac{\pi}{4}, \frac{3\pi}{4}\}$ satisfy GCC. We have to show that no other pair than those two fulfils this condition.

First note that the symmetry of ρ_1 and ρ_2 shows that

$$\rho_Y(\frac{n\pi}{2} - \varphi) = \rho_Y(\frac{n\pi}{2} + \varphi) \tag{A.20}$$

for $n \in \mathbb{Z}$ and $\varphi \in \mathbb{R}$, and $\rho_1 = \rho_2$ induces even

$$\rho_Y(\frac{n\pi}{4} - \varphi) = \rho_Y(\frac{n\pi}{4} + \varphi). \tag{A.21}$$

Due to assumption ρ_Y has only two maxima and two minima in $[0,\pi)$, the above equations then show that those have to be at $\{0, \frac{\pi}{4}, \frac{\pi}{2}\, \frac{3\pi}{4}\}$.

Now we claim that for $\beta_1 \neq \frac{n\pi}{4}$, $n \in \mathbb{Z}$, the median of $\rho_Y|[\beta_1, \beta_1 + \frac{\pi}{2}]$ does not equal $\beta_1 + \frac{\pi}{4}$. Note that this shows the theorem.

For the proof of the claim, consider the smallest integer $p \in \mathbb{Z}$ such that $\gamma_1 := \frac{p\pi}{2} > \beta_1$, and set $\gamma_2 := \gamma_1 + \frac{\pi}{4}$. Let $\beta_2 := \beta_1 + \frac{\pi}{2}$ and $\alpha := \beta_1 + \frac{\pi}{4}$. We have to show that the median of $\rho_Y|[\beta_1, \beta_2]$ does not equal α. A figure visualising these relations and the following definitions is given in figure A.1.

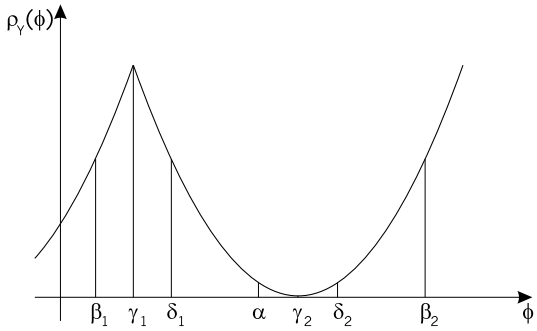

Figure A.1: Visualisation of the proof of theorem 4.8. $\alpha, \beta_i, \gamma_i$ and δ_i are defined in the text.

Using the symmetry noted above, we have for $\delta_1 := \gamma_1 + (\gamma_1 - \beta_1) = 2\gamma_1 - \beta_1$:

$$C_1 := \int_{\beta_1}^{\gamma_1} \rho_Y = \int_{\gamma_1}^{\delta_1} \rho_Y. \tag{A.22}$$

Note that $\delta_1 \neq \alpha$, or else $p\pi = 2\gamma_1 = \alpha + \beta_1 = 2\beta_1 + \frac{\pi}{2}$, which contradicts the assumption to β_1. Without loss of generality, let $\alpha > \delta_1$; the other case can be proven similarly.

Setting $\delta_2 := \gamma_2 + (\gamma_2 - \alpha) = 2\gamma_1 - \alpha$, again symmetry shows that

$$C_2 := \int_{\alpha}^{\gamma_2} \rho_Y = \int_{\gamma_2}^{\delta_2} \rho_Y \tag{A.23}$$

and

$$C_3 := \int_{\delta_1}^{\alpha} \rho_Y = \int_{\delta_2}^{\beta_2} \rho_Y; \tag{A.24}$$

the second equation follows because $\gamma_2 - \delta_1 = \beta_2 - \gamma_2$ and $\gamma_2 - \alpha = \delta_2 - \gamma_2$.

Now assume in contradiction to our claim that α is the median of $\rho_Y|[\beta_1, \beta_2]$. Then we have

$$2C_1 + C_3 = \int_{\beta_1}^{\alpha} \rho_Y = \int_{\alpha}^{\beta_2} \rho_Y = 2C_2 + C_3 \tag{A.25}$$

and therefore $C_1 = C_3$. As shown above $\rho_Y(\gamma_1)$ and $\rho_Y(\gamma_2)$ are the only extremal points of $\rho_Y|[\beta_1, \beta_2]$; without loss of generality, let $\rho_Y(\gamma_1)$ be a maximum, then $\rho_Y(\gamma_2)$ has to be a minimum. But this means that

$$C_1 = \int_{\beta_1}^{\gamma_1} \rho_Y > (\gamma_1 - \beta_1)\rho_Y(\alpha) \geq \int_{\alpha}^{\gamma_2} \rho_Y = C_2 \tag{A.26}$$

which contradicts the above. This completes the proof of the theorem. □

Bibliography

G. S. Adair. The oxygen dissociation curve of hemoglobin. *Journal of Biological Chemistry*, LXIII:529–545, 1925.

S. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10 (2):251–276, 1998.

S. Amari, A. Cichocki, and H. Yang. A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems*, 8:757–763, 1996.

S.-I. Amari. Estimating functions of independent component analysis for temporally correlated signals. *Neural Computation*, 12(9):2083–2107, 2000.

H. An, W. Lin, A. Celik, and Y. Z. Lee. Quantitative measurements of cerebral metabolic rate of oxygen utilization using MRI: a volunteer study. *NMR in Biomedicine*, 14:441–447, 2001.

W. Bargmann. *Histologie und mikroskopische Anatomie des Menschen*. Georg Thieme Verlag, 7th edition, 1977.

C. Bauer, C. Puntonet, M. Rodriguez-Alvarez, and E. Lang. Separation of EEG signals with geometric procedures. *Engineering of Intelligent Systems (Proc. EIS'2000)*, pages 104–108, 2000.

A. Bekker, S. Wolk, H. Turndorf, D. Kristol, and A. Ritter. Computer simulation of cerebralvascular circulation: Assessment of intracranial hemodynamics during induction of anesthesia. *Journal of Clinical Monitoring*, 12:433–444, 1996.

A. J. Bell and T. J. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Signal Processing*, 45(2):434–444, 1997.

M. Benaim, J.-C. Fort, and G. Pagés. Convergence of the one-dimensional Kohonen algorithm. *Adv. Appl. Prob.*, 30:850–869, 1998.

Bosch. *Statistik-Taschenbuch*, pages 697–701. Oldenbuch Verlag, 2nd edition, 1993.

A. Brawanski, R. Faltermeier, R. D. Rothoerl, and C. Woertgen. Comparison of near-infrared spectroscopy and tissue pO_2 time series in patients after severe head injury and aneurysmal subarachnoid hemorrhage. *Journal of Cerebral Blood Flow & Metabolism*, 22:605–611, 2002.

A. Cichocki and S. Amari. *Adaptive blind signal and image processing.* John Wiley & Sons, 2002.

P. Comon. Independent component analysis: A new concept? *Signal Processing*, 36:287–314, 1994.

M. Cottrell and J.-C. Fort. Étude d'un processus d'auto-organisation. *Annales de l'Institut Henri Poincaré*, 23(1):1–20, 1987. (in French).

M. Cottrell, J. C. Fort, and G. Pagès. Two or three things that we know about the Kohonen algorithm. In M. Verleysen, editor, *European Symp. on Artificial Neural Networks (Proc. ESANN'94)*, pages 235–244, Brussels, Belgium, 1994.

R. G. Crystal, J. B. West, P. J. Barnes, and N. S. Cherniack, editors. *The Lung: Scientific Foundations*, volume 2. Raven Press Ltd., New York, 1991.

M. Czosnyka, J. Pickard, H. Whitehouse, and S. Piechnik. The hyperaemic response to a transient reduction in cerebral perfusion pressure. *Acta Neurochirurgica*, pages 90–97, 1992.

M. Czosnyka, S. Piechnik, H. K. Richards, P. Kirkpatrick, P. Smielewski, and J. D. Pickard. Contribution of mathematical modelling to the interpretation of bedside tests of cerebrovascular autoregulation. *Journal of Neurology, Neurosurgery and Psychiatry*, 63:721–731, 1997.

N. Delfosse and P. Loubaton. Adaptive blind separation of independent sources: a deflation approach. *Signal Processing*, 45:59–83, 1995.

H. G. Fridén and J. Ekstedt. Volume/pressure relationship of the cerebrospinal space in humans. *Neurosurgery*, 13(4):351–366, 1983.

M. R. Gaab, I. Haubitz, A. Brawanski, J. Faulstich, and H. E. Heißler. *Intracranial Pressure V*, chapter "Pressure-Volume diagram, pulse amplitude and intracranial pulse volume - analysis and significance". Springer Verlag, 1983.

E. Gao, W. L. Young, J. Pile-Spellman, E. Ornstein, and Q. Ma. Mathematical considerations for modeling cerebral blood flow autoregulation to systemic arterial pressure. *American Journal of Physiology 274 (Heart Circ. Physiol. 43)*, pages 1023–1031, 1998.

M. Gardner. Mathematical games: White and brown music, fractal curves and one-over-f fluctuations. *Scientific American*, 238(4):16–32, 1978.

M. Giulioni and M. Ursino. Impact of cerebral perfusion and autoregulation on intracranial dynamics: A modeling study. *Neurosurgery*, 39(5):1005–1015, November 1996.

E. M. Golding, C. S. Robertson, and R. M. Bryan. The consequences of traumatic brain injury on cerebral blood flow and autoregulation: A review. *Clinical and Experimental Hypertension*, 21(4):299–332, 1999.

R. Greger and U. Windhorst. *Comprehensive Human Physiology*, volume 2. Springer Verlag, 1996.

A. V. Hill. The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *Journal of Physiology*, XL:iv–vii, 1910.

O. Hoffmann. Biomathematics of intracranial CSF and haemodynamics. Simulation and analysis with the aid of a mathematical model. *Acta Neurochirurgica*, 40:117–130, 1987.

L. Hoofd. *Oxygen transport in biological systems: modeling of pathways from enviroment to cell*, chapter "Updating the Krogh model – assumptions and extensions", pages 197–229. Press Syndicate of the University of Cambridge, 1992.

S. Hosseini and C. Jutten. Blind seperation of temporally correlated sources using a quasi maximum likelihood approach. *Independent Component Analysis and Blind Signal Separation (Proc. ICA2001) - Proceedings, San Diego, USA*, 2001.

A. G. Hudetz, J. H. Hasley, C. R. Horton, K. A. Conger, and D. D. Reneau. Mathematical simulation of cerebral blood flow in focal ischemia. *Stroke*, 13 (5), 1982.

A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 10(3):626–634, 1999.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997a.

A. Hyvärinen and E. Oja. Independent component analysis by general nonlinear hebbian-like learning rules. *Signal Processing*, 64(3):301–313, 1997b.

A. Jung and A. Kaiser. Considering temporal structures in independent component analysis. *Independent Component Analysis and Blind Signal Separation (Proc. ICA 2003) - Nara, Japan*, 2003.

A. Jung, F. J. Theis, C. G. Puntmnet, and E. W. Lang. FastGeo – A histogram based approach to linear geometric ICA. *Independent Component Analysis and Blind Signal Separation (Proc. ICA2001), San Diego, USA*, 2001.

C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

Z. M. Kadas, W. D. Lakin, J. Yu, and P. L. Penar. A mathematical mdel of the intracranial system including autoregulation. *Neurological Research*, 19: 441–450, 1997.

A. Kaiser. *Nichtlineare Methoden zur Quantifizierung von Abhängigkeiten und Kopplungen zwischen stochastischen Prozessen basierend auf Informationstheorie*. PhD thesis, Universität Wuppertal, 2002.

A. Kaiser and T. Schreiber. Information transfer in continuous processes. *Physica D*, 166:43–62, 2002.

H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 1997.

F. U. Kauper. *Transkranielle Oximetrie mittels Nahinfrarotspektroskopie*. PhD thesis, Universität Erlangen, 1999.

M. B. Kim, D. S. Ward, C. R. Cartwright, J. Kolano, S. Chlebowski, and L. C. Henson. Estimation of jugular venous O_2 saturation from cerebral oximetry or arterial O_2 saturation during isocapnic hypoxia. *Journal of Clinical Monitoring and Computing*, 16:191–199, 2000.

R. Klinke and S. Silbernagl. *Lehrbuch der Physiologie*. Georg Thieme Verlag, 4th edition, 2003.

A. Krogh. The number and distribution of capillaries in muscles with calculations of the oxygen pressure head necessary for supplying the tissue. *Journal of Physiology*, LII:409–415, 1918-1919.

J. Kugler. *Elektroenzephalographie in Klinik und Praxis – Eine Einführung*. Georg Thieme Verlag, 1981.

W. D. Lakin, J. Yu, and P. Penar. Mathematical modeling of pressure dynamics in the intracranial system. *Nova Journal of Mathematics, Game Theory and Algebra*, 5(2):103–130, 1996.

T. Lee. *Independent Component Analysis – Theory and Applications*. Kluwer Academic Publishers, 1998.

T.-W. Lee, M. Girolami, and T. Sejnowski. Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. *Neural Computation*, 11:417–441, 1999.

M. Lewicki and T. Sejnowski. Learning nonlinear overcomplete representations for efficient coding. In *Advances in Neural Information Processing System*, volume 10. The MIT Press, 1998.

R. Linsker. Local synaptic learning rules suffice to maximize mutual information in a linear network. *Neural Computation*, 4:691–702, 1992.

K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):411–419, 1995.

G. I. Mchedlishvili. *Arterial behaviour and blood circulation in the brain*. Plenum Press, New York, 1986.

G. Meier. Untersuchung nichtlinearer Zeitreihenanalysemethoden. Diplomarbeit - Universität Regensburg, March 2000.

C. Metz. Vergleichende untersuchungen zur morbidität und mortalität bei arteria carotis endarteriektomien in allgemeinanästhesie versus regionalanästhesie unter besonderer berücksichtigung der zerebralen ischaemie. Habilitation - Universität Regensburg, 2001.

M. A. Mintun, B. N. Lundstrom, A. Z. Snyder, A. G. Vlassenko, G. L. Shulman, and M. E. Raichle. Blood flow and oxygen delivery to human brain during functional activity: Theoretical modeling and experimental data. *Proceedings of the National Academy of Sciences*, 98(12):6859–6864, 2001.

L. Molgedey and H. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72:3634–3636, 1994.

E. Okada, M. Firbank, M. Schweiger, S. R. Arridge, M. Cope, and D. T. Delpy. Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head. *Applied Optics*, 36(1):21–31, 1997.

A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 3rd edition, 1991.

R. K. Pathria. *Statistical Mechanics*, pages 504–505. Butterworth, Heinemann, 2nd edition, 1998.

B. Pearlmutter and L. Parra. Maximum likelihood blind source separation: A context-sensitive generalization of ICA. *Advances in Neural Information Processing Systems*, 1997.

A. Prieto, B. Prieto, C. Puntonet, A. Canas, and P. Martin-Smith. Geometric separation of linear mixtures of sources: Application to speech signals. *Independent Component Analysis and Blind Signal Separation (Proc. ICA'99)*, pages 295–300, 1999.

C. Puntonet, M. Alvarez, A. Prieto, and B. Prieto. Separation of speech signals for nonlinear mixtures. *Lecture Notes in Computer Science*, 1607:665–673, 1999.

C. Puntonet, C. Bauer, E. Lang, M. Alvarez, and B. Prieto. Adaptive-geometric methods: Application to the separation of EEG signals. *Independent Component Analysis and Blind Signal Separation (Proc. ICA'2000)*, pages 273–278, 2000.

C. Puntonet and A. Prieto. An adaptive geometrical procedure for blind separation of sources. *Neural Processing Letters*, 2(5), 1995.

C. G. Puntonet and A. Prieto. Neural net approach for blind separation of sources based on geometric properties. *Neurocomputing*, 18:141–164, 1998.

C. G. Puntonet, A. Prieto, C. Jutten, M. R. Alvarez, and J. Ortega. Separation of sources: a geometry-based procedure for reconstruction of n-valued signals. *Signal Processing*, 46:267–284, 1995.

M. J. Purves. *The Physiology of the Cerebral Circulation*. Cambirdge University Press, 1972.

U. Raff, A. L. Scherzinger, P. F. Vargas, and J. H. Simon. Quantitation of grey matter, white matter, and cerebrospinal fluid from spin-echo magnetic resonance images using artificial neural network techniques. *Medical Physics*, 21(12):1933–1942, 1994.

D. D. Reneau, D. F. Bruley, and M. H. Knisely. *Chemical Engineering in Medincine and Biology*, chapter "A mathematical simulation of oxygen release, diffusion, and consumption in the capillaries and tissue of the human brain", pages 135–241. Plenum Press, New York, 1967.

D. D. Reneau, D. F. Bruley, and M. H. Knisely. A digital simulation of transient oxygen transport in capillary-tissue systems (cerebral grey matter). *AIChE Journal*, 15(6):916–925, November 1969.

D. D. Reneau, D. F. Bruley, and M. H. Knisely. A computer simulation for prediction of oxygen limitations in cerebral gray matter. *Journal of the Association for the Advancement of Medical Instrumentation*, 4(6):211–223, 1970.

D. D. Reneau and M. H. Knisely. A mathematical simulation of oxygen transport in the human brain under conditions of countercurrent capillary blood flow. *Advances in Bioengineering – Chemical Engineering Progress Symposium Series*, 67(114):18–27, 1971.

H. Ritter and K. Schulten. Convergence properties of Kohonen's topology conserving maps: Fluctuations, stability and dimension selection. *Biological Cybernetics*, 60:59–71, 1988.

F. J. W. Roughton and J. W. Severinghaus. Accurate determination of O_2 dissociation curve of human blood above 98.7% saturation with data on O_2 solubility in unmodified human blood from 0° to 37° C. *Journal of Applied Physiology*, 35(6):861–869, December 1973.

R. F. Schmidt, G. Thews, and F. Lang. *Physiologie des Menschen*. Springer Verlag, 28th edition, 2000.

T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85:461, 2000.

J. W. Severingshaus. Simple, accurate equations for human blood O_2 dissociation computations. *Journal of Applied Physiology*, 46:599–602, January-June 1979.

M. Sharan and A. S. Popel. A compartment model for oxygen transport in brain microcirculation in the presence of blood substitutes. *Journal of theoretical Biology*, 216:479–500, 2002.

F. H. Sklar and I. Elashvili. The pressure-volume function of brain elasticity – physiological considerations and clinical applications. *Journal of Neurosurgery*, 47:670–679, 1977.

S. A. Stevens. Mean pressures and flows in the human intracranial system, determined by mathematical simulations of a steady-state infusion test. *Neurological Research*, 22:809–814, 2000.

H. G. Sullivan and J. D. Allison. Physiology of cerebrospinal fluid. In R. Wilkins and S. Rengachary, editors, *Neurosurgery*, volume 3, pages 2125–2135. McGraw Hillbook Co., New York, 1985.

F. Theis, E. Lang, and C. Puntonet. A geometric algorithm for overcomplete linear ICA. *Neurocomputing*, 2003a. accepted.

F. J. Theis. *Mathematics in Independent Component Analysis*. PhD thesis, University Regensburg, 2002.

F. J. Theis, A. Jung, and E. W. Lang. A theoretic model for linear geometric ICA. *Independent Component Analysis and Blind Signal Separation (Proc. ICA2001), San Diego, USA*, 2001.

F. J. Theis, A. Jung, C. G. Puntonet, and E. W. Lang. Linear geometric ICA: Fundamentals and algorithms. *Neural Computation*, 15:419–439, 2003b.

L. Tong, R.-W. Liu, V. Soon, and Y. Huang. Indeterminacy and identifiability of blind identification. *IEEE Trans. on Circuits and Systems*, 38:499–509, 1991.

M. Ursino. A mathematical study of human intracranial hydrodynamics: Part 1 - The cerebrospinal fluid pulse pressure. *Annals of Biomedical Engineering*, 16:379–401, 1988.

M. Ursino and C. Lodi. A simple mathematical model of the interaction between intracranial pressure and cerebral hemodynamics. *Journal of Applied Physiology*, 82(4):1256–1269, 1997.

S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer Verlag, 1990.

C. Woertgen, R. Faltermeier, R. Rothörl, and A. Brawanski. Vergleich von venrikelnahen und kortikalen intracraniellen drücken. In *Jahrestagung der Sektion Intracranieller Druck, Hirndurchblutung, Hirnödem der Deutschen Gesellschaft für Neurochirurgie*, Berlin, Unfallkrankenhaus, 10.-11.Oktober 2003.

A. Ziehe and K.-R. Müller. TDSEP – an efficient algorithm for blind separation using time structure. *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, pages 675–680, 1998. Skövde, Sweden.

S. Zschocke. *Klinische Elektroenzephalographie*. Springer Verlag, 1995.

A. Zwart, G. Kwant, B. Oeseburg, and W. G. Zijlstra. Human whole-body oxygen affinity: effect of temperature. *Journal of Applied Physiology*, 57:429–434, July-December 1984.

Dank

Ich möchte mich besonders bei den folgenden Personen bedanken, die diese Arbeit in dieser Form möglich gemacht haben. Speziell geht mein Dank an meinen Doktorvater und Betreuer Klaus Richter, da ohne ihn diese Arbeit nicht hier – lieber Leser – in ihren Händen liegen würde.

Für die Unterstützung und Hilfe als wichtiges Glied zwischen Physik und Medizin bedanke ich mich bei Rupert Faltermeier der auch innerhalb kürzester Zeit die verschiedensten Versionen meiner Dissertation korregiert hat. Bei Ralf Rothörl möchte ich mich für die manchmal für ihn etwas anstrengenden medizinischen Diskussionen bedanken, wie auch bei Professor Brawanski für die immer währende Unterstützung zum Fortführen der Arbeit im medizinischen Gebiet. Aber ebenso bei Professor Gustav Obermair, der die Kooperation zwischen Medizin und Physik erst initiiert hat und damit an der Entstehung dieses Teils der Arbeit wesentlich beigetragen hat.

Dank auch an Andreas Kaiser, ohne ihn der ICA Teil nicht seine jetzige Form hätte und mir immer geduldig meine Fragen bzgl. Wahrscheinlichkeits- und Informationstheorie beantwortet hat. Ebenso aber auch Fabian Theis für die Zusammenarbeit und Diskussionen auf dem Gebiet der (geometrischen) ICA und für die für uns beide interessanten Konferenzreisen.

Jörg Kaidel danke ich für die vielen Ratschläge und Diskussionen, speziell während der Entstehung dieser Arbeit – operation "falling snow" ist also doch noch pünktlich fertig geworden. Meinem Zimmerkollegen Juan-Diego Urbina dank ich ebenfalls, speziell für die Hilfe bei den vielen "illposed questions" – wie er sie immer nannte – und Peter Schlagheck für das geduldige Zuhören bei meinen Problemen mit der Dissertation. Ausserdem danke ich Gustav Mecke für die Arbeit als gründlicher Lektor.

Für die professionelle EDV-Unterstützung möchte ich mich besonders bei Fritz Wünsch und Rudi Holzer bedanken, die für unsere Probleme immer ein offenes Ohr hatten. Aber auch Christian Guggenberger, der uns unermüdlich bei unserem Physik-Cluster und allen anderen Probleme von Linux geholfen hat. Dank auch an Christoph Bauer, der einerseits mich an Anfang an der Uni und in Regensburg willkommen geheissen hat, aber auch jetzt sich sehr um die Belange der Physik im Rechenzentrum kümmert. Ausserdem Dank an meine EDV-Kollegen Josef Kainz und Christian Lechner für die herzliche Atmosphäre.

Der Deutschen Forschungsgemeinschaft danke ich für die finanzielle Unterstützung im Rahmen eines Stipendiums im Graduiertenkolleg *"Nichtlinearität und Nichtgleichgewicht in kondensierter Materie"*.

Natürlich hat die angenehme und warme Atmosphäre am Institut ebenfalls zu der glücklichen Entstehung der Arbeit beigetragen, daher Dank an alle Mitarbeiter am Institut. Besonders danke ich aber Angela Reißer, die ich immer als hilfsbereite und liebe Sekretärin erleben durfte.

Doch ohne die Liebe und Unterstützung meiner Eltern wäre diese Arbeit nie entstanden. Danke!

Allen nochmal herzlichen Dank, auch jenen,
die aus reiner Vergesslichkeit hier nicht aufgeführt sind.