Master 2 Signaux Pour La Santé, les Télécommunications, l'Image et Multimédia

Tensor Models Applied to Atrial Fibrillation Analysis

Supervisors:
Prof. Dr. Vicente Zarzoso
CNRS Research Director Gérard Favier

Author:
Lucas Nogueira Ribeiro

Academic supervisor:
Prof. Dr. Luc Deneire

Sophia Antipolis, France
January 29, 2015

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Vicente Zarzoso, for his guidance, inspiration, patience, constructive comments and good humor throughout this internship. I would also like to thank my co-supervisor, Prof. Dr. Gérard Favier, for his many constructive comments and suggestions about the tensors methods that I worked with. In addition, I would like to thank Prof. Dr. João César for the many fruitful discussions we had about adaptive noise cancellation methods and Prof. Dr. Luc Deneire for welcoming me in France. I thank also Dr. Antonio Hidalgo-Muñoz for his constant support and the many scientific discussions we had during this internship. I would also like to express my gratitude towards CAPES for funding my stay in France through the BRAFITEC project.

Glossary

ABS Average Beat Subtraction.

AF Atrial Fibrillation.

ALS Alternating Least-Squares.

AV Atrioventricular.

BSS Blind Source Separation.

BTD Block Terms Decomposition.

CPD Canonical Polyadic Decomposition.

CWT Continuous Wavelet Transform.

ECG Electrocardiogram.

EEG Electroencephalogram.

FIR Finite Impulse Response.

HOOI High-Order Orthogonal Iteration.

HOSVD High-Order Singular Value Decomposition.

ICA Independent Component Analysis.

IIR Infinite Impulse Response.

MPCA Multilinear Principal Component Analysis.

MRANC Multi Reference Adaptive Noise Cancellation.

MSE Mean Squared Error.

MSL Multilinear Subspace Learning.

PCA Principal Component Analysis.

SA Sinoatrial.

STC Spatiotemporal Cancellation.

STFT Short-Time Fourier Transform.

SVD Singular Value Decomposition.

List of Figures

List of Tables

Contents

Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia detected by clinicians, affecting mostly the elderly population. There is clinical evidence that AF is responsible of about one fourth of the cerebrovascular accidents (brain strokes). Since its mechanisms are not yet well understood, it is considered the "last great frontier" of electrophysiology. The non-invasive analysis of the AF can be performed by analyzing the surface electrocardiogram (ECG). Arranged in matrix form for, the ECG can be factorized by means of certain matrix decomposition techniques like principal component analysis (PCA), independent component analysis (ICA) and their variants, permitting the identification of some latent signals, called "sources". The matrix decomposition approach is very interesting for artifacts removal, AF complexity evaluation and prediction of the success of many therapies.

With the advent of the ECG as an analysis tool to assess the heart's activity, physicians had to learn how to interpret them in order to obtain medical conclusions. However sometimes there are some details in the ECG that can be difficult to perceive or accurately calculate. Fortunately, with the aid of computers and digital signal processing techniques, it is possible to obtain accurate quantitative information from ECG [6]. AF caught the attention of the scientific community because its mechanisms are not fully understood. In this case, ECG can be used to analyze the AF and give some insights about its mechanisms. Specialized signal processing algorithms have been proposed to model and extract the atrial fibrillation activity from ECG recordings [7, 8].

These techniques rely on well established matrix methods, namely *Principal Component Analysis* (PCA) [9], *Independent Component Analysis* (ICA) [10] and their variants. These techniques explore some a priori knowledge of the ECG signals and use them to obtain some "hidden" signal components that describe interesting features, giving insights about the heart. This operation of extracting latent signal components from a set of observation signals is known as blind source separation (BSS).

In recent years, multiway arrays, or tensors, have been applied to signal processing problems such as wireless communications systems [11, 12], audio source separation [13], pattern recognition [14], etc. Tensors can be seen as the generalization of matrices, arranging data in multiple dimensions (or modes). Some advantages of this more flexible structure include the possibility of exploring the inherent multi-dimensionality of some data. Colored images, for example, can be decomposed in three color channels, easily being modeled as a third order tensor. Another advantage concerns the rank. The rank of a matrix is bounded by its smallest dimensions, whereas this bound does not apply to tensors. In fact, there are many definitions of rank for tensors [15].

An ongoing project between the "Biomed" and "Tensors" teams of the I3S laboratory aims at exploring this multidimensional approach focused to the analysis of cardiac signals. This Master 2 internship is engaged on this project. The proposed work consists of modeling ECG signals using the tensor framework with the final objective of analyzing AF ECG signals.

In the following chapters, we will discuss some current methods for extracting AF sources from ECG signals. Next, an introduction of tensor methods and possible ways to apply them to ECG signals will be presented. Subsequently, a tensor-based BSS method will be employed to the extraction of AF sources from ECG observations. Finally, we will discuss Multilinear Subspace Learning, which can be used to describe the most important characteristics of an ECG signal.

Chapter 1

Cardiac Electrophysiology

Cardiac electrophysiology is the study of the electrical activities of the heart. By analyzing it either by invasive methods such as an endocardiac catheter or by noninvasive methods such as ECG, it is possible to have some insight about the heart's condition. Non invasive methods are very important nowadays in clinical practice, since the physician can have a general idea about the heart of a patient in a rather simple way. Initially we will discuss the electrical activity in the cellular level, fundamental to understand the electrical heart's mechanisms. Subsequently we will expand the discussion to the mechanical and electrical cardiac activities in general and we will briefly present the ECG and the 12-lead standard system. Finally, we will introduce the AF, a heart disease which is one keyword in this work.

1.1 Electrical Cellular Activity

The analysis of bioelectrical signals on the body surface aims at observing the general electrical activity of the body cells and interpret them. In order to comprehend how the electrophysiology of the heart and how the electrocardiogram is generated, it is important to understand the ionic exchanges in the cellular level that generate the electrical activities of the human body.

A cell basically consists of a protoplasm enclosed within a membrane, which contains many biomolecules such as proteins and nucleic acids [16]. This membrane is made of a poor conducting material and has a permeability to certain substances, allowing them to enter or quit the cell via specific channels. Even though the intra and extracellular contents are mostly water, an electrically neutral substance, there are some ions that allow the electrical conductance.

Under resting conditions, the interior of the cell is negatively charged with respect to the outside. This difference of voltage is caused by the different concentration of ions, resulting in a difference of electrical charge. When a cell is stimulated by a current, rapid alterations in the membrane permeability take place, resulting in a movement of ions called *action potential*. The action potential consists of two phases. The first is the *depolarization*, in which the inside of the cell becomes less negative and ultimately becomes positive due to the movement of ions. The second is the *repolarization*, during which the potential returns to its resting level, so that the inside becomes more negative once again.

To illustrate the action potential phenomenon, let us consider the situation in which potassium ions are assumed to be inside the cell, sodium ions are outside and the membrane potential is zero. When it is perturbed by an external stimulus such as currents propagating from neighboring cells that are greater than the threshold, the action potential is initiated. The depolarization takes place and the sodium channels at the membrane are opened, so that sodium ions can move into the cell, making the interior plus positive, thus less negative, until the polarity of the cell is reversed. Next, the sodium channels are closed and potassium channels are opened, initiating the repolarization. Here, the potassium ions will quit the cell, making the its interior less positive, thus more negative. Finally the original resting state will be attained. An action potential propagates to neighboring cells, triggering new action potentials until it reaches a point where the voltage is insufficient for further stimulation.

1.2 The Human Heart

The heart is a muscular organ of the circulatory system responsible for pumping oxygenated (arterial) blood throughout the body and desoxygenated (venous) blood to the lungs. The wall of the heart is called

myocardium and is composed of muscle cells that contract in the occurrence of an electrical stimulation. The heart is divided in four cavities: left atria, left ventricle, right atria and right ventricle, as depicted in Figure 1.1. A heart cycle consists of mechanical contractions within the heart cavities that make the blood move. This cycle starts in the right atrium, where the desoxygenated blood arrives from the body via the superior vena cava. Then, the right atrium is triggered to contract and push the blood to the right ventricle. When it has been filled, it will contract and force the blood to the lungs. The venous blood will be oxygenated in the lungs and taken to the left atrium via the pulmonary veins. Next, the left atrium is contracted and forces the blood to the left ventricle, which will pump the oxygenated blood to all the body through the arterial vessels. The metabolism of the cells will transform the oxygen in carbon dioxide, desoxygenating the blood, which will be taken back to the heart via the venous system.

This whole mechanical process is generated and coordinated by electrical impulses. The heart cycle is initiated in a group of cells called sinoatrial (SA) node that has the ability of spontaneously emit electrical impulses. The SA node, situated in the upper part of the right atrium, will fire a depolarization impulse that will make the atria contract. This impulse will propagate through the conduction system until arriving at the atrioventricular (AV) node. This node will cause a delay of propagation which will allow the atrial contraction to further increase the blood volume in the ventricles before they contract. Subsequently, the impulse propagates in the wall between the two ventricles in a path called bundle of His. Eventually this path is divided in a network of conductive fibers called Purkinje fibers, which will trigger a strong and unified contraction in the ventricles. After the depolarization, the heart cells will will undergo a repolarization, returning to the resting state, and then another heart cycle is ready to start.

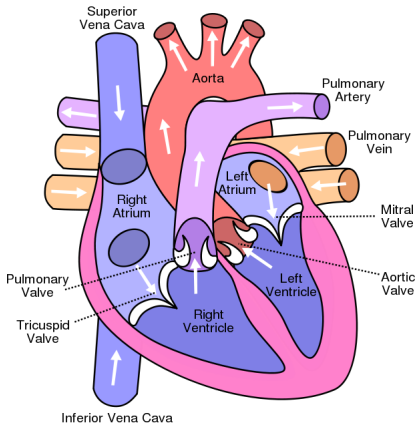

Figure 1.1: The human heart and its cavities. Original image created by Eric Pierce and licensed under the Creative Commons Attribution Share- Alike 3.0 License.

1.3 The Electrocardiogram

The ECG is a tool developed in the 20th century that allows us to assess the electrical activity of the heart. Electrodes are placed over the body surface to measure the voltage variations of some specialized cardiac cell as they make the heart contract. The resulting signal in the ECG is manifested by a series of waves whose morphology and timing can be used for describing the heart's health.

The ECG consists of waves that describe the difference of voltage between a pair of electrodes, called lead. A unipolar lead is measures the voltage variation of a single electrode in relation to a reference electrode defined to have zero potential. A bipolar lead reflects the difference between two electrodes placed over the body. These electrodes are connected to special amplifiers with high gain and large dynamic range.

The most used ECG system is the standard 12-lead [16]. In this system, the 10 electrodes are placed over the body as depicted in Figure 1.2. Let V_{LA} and V_{RA} and V_{LL} denote voltage in left arm, right arm and left leg, respectively. The 12-lead ECG system consists of

- Three bipolar limb leads denoted I, II and III. They measure the voltage difference between the arms and between the arms and the left leg. The position of the electrodes which correspond to

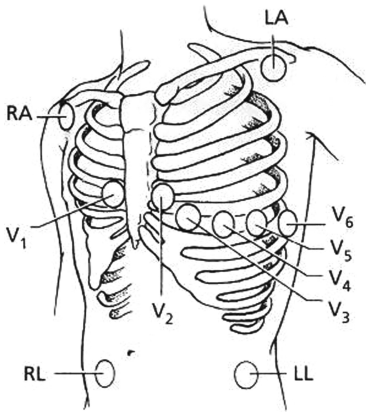

Figure 1.2: The standard placement os electrodes for the 12-lead ECG system [1].

the arms and the left leg can be viewed as the corners of a triangle with the heart at its center. This triangle is known as Einthoven's triangle.

- Three augmented unipolar limb leads denoted aVF, aVL and aVR. These leads use the same electrodes as the bipolar limb leads, but are defined as voltage differences between one corner of Einthoven's triangle and the average of the remaining corners.
- Six unipolar precordial leads $V_1, \ldots, V_6$. They are placed on the left front of the chest, giving a more detailed reflection of the heart's activity. Since these leads are placed close to the heart, they have higher amplitudes than the limb leads.

A typical ECG lead from a healthy subject consists of a P wave, the QRS complex and the T wave (Figure 1.3). The P wave reflects the depolarization of the atria. The QRS complex reflects the depolarization of the ventricles and the T wave reflects the ventricular repolarization. Some ECG intervals describe the heart's state. The RR interval, for example, is measured as the length between two successive R waves and indicates the ventricular rate.

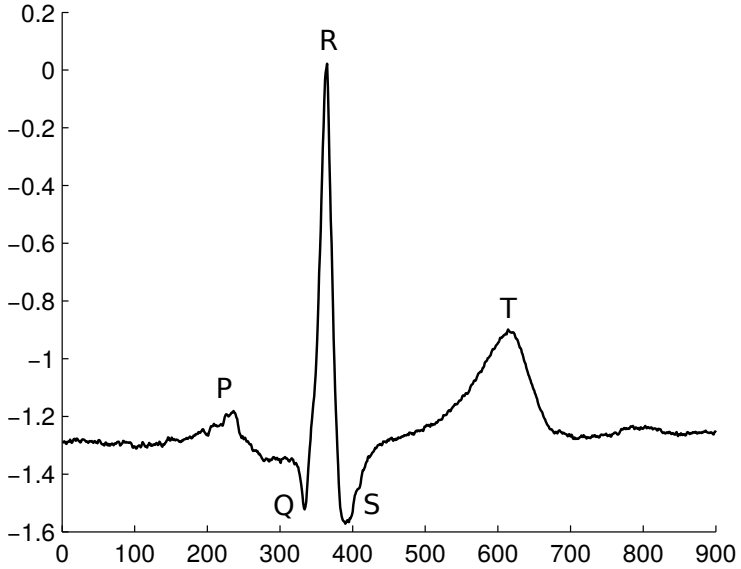

Figure 1.3: Sinus rhythm ECG indicating the P, Q, R, S, T waves [2].

The spectrum of the ECG is mostly concentrated between 0 Hz and 40 Hz as shown in Figure 1.4. The P and T waves are usually below 10-15 Hz, considered to be low-frequency components, while the QRS complex is mostly concentrated between 10 Hz and 50 Hz. The ECG may present some noise

and artifact components that may deteriorate its quality. Baseline wander is a low-frequency component below 1 Hz related to respiration and body movements that makes the isoelectric line fluctuate. Powerline interference is caused by proximity of certain equipment and improper grounding of the ECG equipment. Electromyographic noise, or simply muscle noise, originates from contraction of skeletal muscles during. Moving the electrodes while the ECG is being recorded may generate some artifacts in the ECG. These interferences may be canceled by using signal processing techniques such as notch filters [16] and adaptive filters [17].

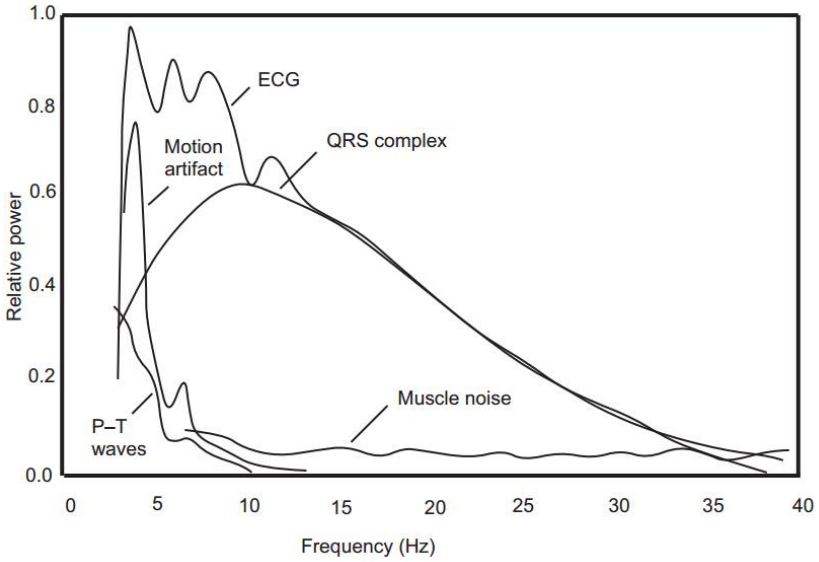

Figure 1.4: Power spectrum of an ECG signal [3].

1.4 Atrial Fibrillation

AF is a cardiac arrhythmia characterized by uncoordinated atrial activation. Its causes are not very well understood. Indeed, it is considered as "last great frontier of cardiac electrophysiology" [18]. The most popular hypothesis that explains its mechanisms is the multiple wavelet reentry. Depolarization wavefronts are fractioned due to alteration in the conduction system of the atria, generating reentry wavelets, which randomly activate the AV node, resulting in a disorganized ventricular contraction.

This heart condition is quite common in the clinical practice, affecting about 10% of the elder population. Due to the irregular atrial and ventricular contractions, there may be a formation of blood clots, which might cause cerebral vascular accidents. AF is often treated with medication to correct the heart rhythm, and catheter ablation methods are used to prevent recurrence of AF.

AF can be classified according to the frequency of episodes in a patient. When it happened after two or more times, AF is considered recurrent. If it terminates spontaneously, it is considered paroxysmal. Persistent AF is considered when the arrhythmia is sustained beyond a week.

In the ECG, the AF is reflected as an irregular oscillation that replaces the P wave and is mostly visible in the TQ interval, specially in leads II, III, aVF and V_1. These random oscillations are known as fibrillatory waves, or simply f-waves (Figure 1.5). They may present several levels of disorganization according to the stage of development of the disease. The atrial flutter is an associated arrhythmia related to AF and it is reflected in the ECG as a sawtooth shaped wave in the TQ segment. If not treated, atrial flutter may degenerate into AF.

In this chapter we have introduced the basic elements of cardiac electrophysiology, AF mechanisms and how they are connected to an ECG. In the following chapters, we will analyze ECG signals in order to obtain information about AF.

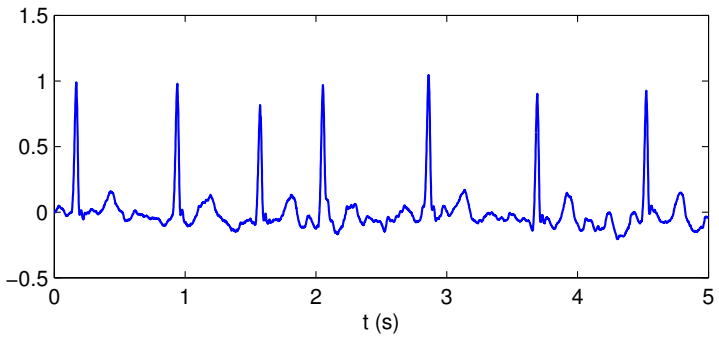

Figure 1.5: An ECG recording of an subject presenting Atrial Fibrillation.

Chapter 2

Atrial activity extraction in ECG recordings

In this work we focus on the problem of atrial activity extraction for ECG. During an AF episode, atrial activity corresponds to the f-waves, which bring information and features very important to understanding AF mechanisms, analyzing the disorganization level of AF, inferring the development of the disease, and even predicting the success of different therapies to treat the arrhythmia, such as catheter ablation. In the following, we will review the existent techniques for atrial activity extraction. We will focus on matrix methods, which can be divided in three different approaches: adaptive noise cancellation, spatiotemporal cancellation and blind source separation.

2.1 Multi-Reference Adaptive Noise Cancellation

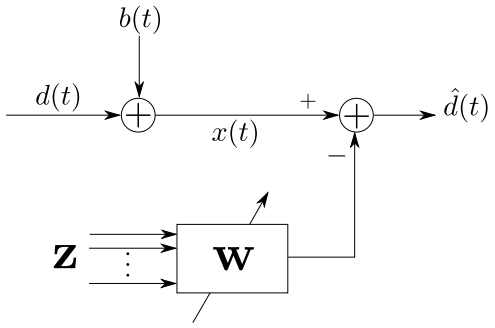

Figure 2.1: Block diagram of the adaptive noise cancellation filter.

Consider an observation signal $x(t)$ composed of a desired signal $d(t)$ (or latent source) and an interference signal $b(t)$. Then,

$$x(t) = d(t) + b(t). \tag{2.1}$$

We seek to suppress the noise signal $b(t)$ to estimate our desired signal $d(t)$. To do so, we suppose that reference signals $\{z_1(t), z_2(t), \ldots, z_R(t)\}$ are available and that they are correlated with the interference signal $b(t)$ but uncorrelated with the desired signal $d(t)$.

This problem formulation can be solved by means of the multi-reference adaptive noise canceling (MRANC) [19, 17], where a filter continuously adapts its coefficients to estimate a desired signal from a set of observation signals. Either Finite Impulse Response (FIR) or Infinite Impulse Response (IIR) filters can be adopted to solve this problem. However, due to its simplicity and assured stability, a FIR filter will be considered.

Consider a set of R FIR filters of order M with impulse response $\{w_r(0), w_r(1), \ldots, w_r(M-1)\}_{r=1}^{R}$. The desired signal will be estimated by subtracting approximated versions of $b(t)$ from the observation signal $x(t)$. These approximations are obtained by filtering the reference signals $z_r(t)$ for $r = 1, \ldots, R$

with the FIR filters. Then, the estimated desired signal is

$$\begin{aligned}\hat{d}(t) &= x(t) - \sum_{r=1}^{R} w_r(t) * z_r(t) && (2.2)\\ &= x(t) - \sum_{r=1}^{R}\left(\sum_{k=0}^{M} w_r(k) z_r(t-k)\right) && (2.3)\\ &= x(t) - \mathbf{w}^T\mathbf{z}(t), && (2.4)\end{aligned}$$

where

$$\mathbf{w} = [w_1(0), \ldots, w_1(M-1), w_2(0), \ldots, w_2(M-1), \ldots, w_R(0), \ldots, w_R(M-1)]^T$$

is the reference filters coefficients vector and

$$\mathbf{z}(t) = [z_1(t), \ldots, z_1(t-M), z_2(t), \ldots, z_2(t-M), \ldots, z_R(t), \ldots, z_R(t-M)]^T$$

is the delayed samples of the reference signals (Figure 2.1).

Consider the output of the filter, denoted $\hat{d}(t)$:

$$\begin{aligned}\hat{d}(t) &= x(t) - \mathbf{w}^T\mathbf{z}(t) && (2.5)\\ &= d(t) + b(t) - \mathbf{w}^T\mathbf{z}(t) && (2.6)\end{aligned}$$

The criterion to be optimized in order to find the filter coefficients $\mathbf{w}$ that will lead to an optimal estimation of $d(t)$ is defined as

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathbb{E}\left\{\left(\hat{d}(t) - d(t)\right)^2\right\}. \quad (2.7)$$

Let $\varepsilon = \mathbb{E}\left\{\left(\hat{d}(t) - d(t)\right)^2\right\}$. Then the problem (2.7) is solved by setting $\nabla\varepsilon = 0$ and solving for $\mathbf{w}$. Then

$$\begin{aligned}\nabla\varepsilon &= \frac{\partial}{\partial\mathbf{w}}\left[\mathbb{E}\left\{\left(\hat{d}(t) - d(t)\right)^2\right\}\right] && (2.8)\\ &= \frac{\partial}{\partial\mathbf{w}}\mathbb{E}\left\{\left(d(t) + b(t) - \mathbf{w}^T\mathbf{z}(t) - d(t)\right)^2\right\} && (2.9)\\ &= \frac{\partial}{\partial\mathbf{w}}\left[\mathbb{E}\left\{\left(b(t) - \mathbf{w}^T\mathbf{z}(t)\right)^2\right\}\right] && (2.10)\\ &= \frac{\partial}{\partial\mathbf{w}}\left[\mathbb{E}\left\{b^2(t) - 2b(t)\mathbf{w}^T\mathbf{z}(t) + \mathbf{w}^T\mathbf{z}(t)\mathbf{z}(t)^T\mathbf{w}\right\}\right] && (2.11)\\ &= -2\mathbb{E}\left\{b(t)\mathbf{z}(t)\right\} + 2\mathbb{E}\left\{\mathbf{z}(t)\mathbf{z}(t)^T\right\}\mathbf{w}. && (2.12)\end{aligned}$$

Let $\mathbf{R}_\mathbf{z} = \mathbb{E}\left\{\mathbf{z}(t)\mathbf{z}(t)^T\right\}$ and $\mathbf{p}(t) = \mathbb{E}\left\{b(t)\mathbf{z}(t)\right\}$, setting the equation (2.12) to $\mathbf{0}$:

$$\mathbf{R}_\mathbf{z}\mathbf{w} = \mathbf{p}(t) \quad (2.13)$$

and the optimal filter is

$$\mathbf{w}^* = \mathbf{R}_\mathbf{z}^{-1}\mathbf{p}(t). \quad (2.14)$$

The equation (2.13) is known as the Wiener-Hopf equation. The optimal solution decorrelates the output of the filter $\hat{d}(t)$ with the reference signals $\mathbf{z}(t)$ which are assumed to model $b(t)$. This decorrelation makes $\hat{d}(t)$ be the best estimation of $d(t)$ in the MSE sense.

In practice, $d(t)$ is not known. So we can reformulate the optimization criterion to

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathbb{E}\left\{\hat{d}(t)^2\right\}. \quad (2.15)$$

Let $\varepsilon' = \mathbb{E}\left\{\hat{d}(t)^2\right\}$. Using the same assumptions as before, let us develop it:

$$\begin{aligned}\varepsilon' &= \mathbb{E}\left\{\hat{d}(t)^2\right\} && (2.16)\\ &= \mathbb{E}\left\{\left(d(t) + b(t) - \mathbf{w}^T\mathbf{z}(t)\right)^2\right\} && (2.17)\\ &= \mathbb{E}\left\{d^2(t) + 2d(t)\left(b(t) - \mathbf{w}^T\mathbf{z}(t)\right) + \left(b(t) - \mathbf{w}^T\mathbf{z}(t)\right)^2\right\} && (2.18)\\ &= \mathbb{E}\left\{d^2(t)\right\} + \mathbb{E}\left\{\left(b(t) - \mathbf{w}^T\mathbf{z}(t)\right)^2\right\}. && (2.19)\end{aligned}$$

Hence

$$\min_{\mathbf{w}} \mathbb{E}\left\{\hat{d}(t)^2\right\} \equiv \min_{\mathbf{w}} \mathbb{E}\left\{\left(b(t) - \mathbf{w}^T\mathbf{z}(t)\right)^2\right\} \equiv \min_{\mathbf{w}} \mathbb{E}\left\{\left(\hat{d}(t) - d(t)\right)^2\right\}.$$

The optimal filter can be obtained either by the equation (2.13) or iteratively by means of a gradient-descent algorithm

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \mu\nabla\varepsilon(\mathbf{w}^{(k)}). \tag{2.20}$$

If we use instantaneous gradient estimates, the expectation will drop and then (2.20) becomes

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - 2\mu\mathbf{z}(t)\left(x(t) - \mathbf{w}^{(k)^T}\mathbf{z}(t)\right). \tag{2.21}$$

The scalar μ is known as adaptation step and must verify $0 < \mu < 1/\lambda_{max}$, where λ_{max} is the dominant eigenvalue of $\mathbf{R_z}$.

In the context of atrial activity extraction, we desire to cancel the QRST complex, which corresponds to the ventricular activity. Then let us model it as the interference $b(t)$ uncorrelated to the atrial activity $d(t)$. The observation input $x(t)$ will correspond to an electrode lead while the reference signals $\mathbf{z}(t)$ are chosen to be correlated with $b(t)$, but uncorrelated to $d(t)$. In practice, we choose them as ECG observation collected next to the ventricular sources. Clearly, the quality of estimation will be subject to a good positioning of the reference electrodes. This technique is also very sensible to ectopic beats, which may disturb the convergence of the transveral filters. Moreover, we can not always assure that the references signals are correlated to the QRST complex $b(t)$ and uncorrelated to the atrial activity $d(t)$. The performance of this technique depends of the well placement of the electrodes, so that the reference signals will be decorrelated, fulfilling the hypothesis cited above.

2.2 Spatiotemporal Cancellation

The QRST complex corresponds to the electrical manifestation of the ventricular activity in the ECG. Thus, suppressing the QRST complex from the ECG may result in an estimate of the f-wave. The Average Beat Subtraction (ABS) [20] is a popular technique that consists of estimating a beat template by synchronously averaging the QRST complex in the entire ECG observation and then subtracting it from the observation in one lead. In a multi-lead scenario, this procedure is repeated for each lead. Before calculating the average beat, the QRST complexes are detected and aligned, then each heart beat is segmented into QRST complex and atrial activity. Subsequently the average beat is calculated by using the ensemble average or the median beat. After generating a template of average beats, it is subtracted from the original ECG signal. Even though the ABS is a simple and intuitive method, it is sensitive to alterations in the electrical axis of the heart, manifested in different leads, resulting in large QRST residuals.

This problem can be mitigated by using the spatiotemporal cancellation (STC) method [4]. Let $\mathbf{X} \in \mathbb{R}^{N\times R}$ be the data vector with R leads and N samples. The average beat $\bar{\mathbf{X}} \in \mathbb{R}^{(N+2\Delta)\times R}$ will contain 2Δ additional samples in order to allow the temporal alignment of $\mathbf{X}$ relative to $\bar{\mathbf{X}}$ performed by the shift matrix

$$\mathbf{J}_\tau = \begin{pmatrix}\mathbf{0}_{N\times(\Delta+\tau)} & \mathbf{I}_{N\times N} & \mathbf{0}_{N\times(\Delta-\tau)}\end{pmatrix}_{N\times(N+2\Delta)}$$

where τ denotes an integer time shift, $\mathbf{0}$ and $\mathbf{I}$ denote the zero and the identity matrix respectively. Assuming that the atrial activity $\mathbf{X}_A$ and the ventricular activity $\mathbf{X}_V$ are statistically decoupled, the data matrix can be decomposed as

$$\mathbf{X} = \mathbf{X}_A + \mathbf{X}_V + \mathbf{N}. \tag{2.22}$$

The ventricular activity is assumed to be an average beat $\bar{\mathbf{X}}$ aligned in time by $\mathbf{J}_\tau$ and in space by $\mathbf{S} \in \mathbb{R}^{R\times R}$ in the following way:

$$\mathbf{X}_V = \mathbf{J}_\tau\bar{\mathbf{X}}\mathbf{S}. \tag{2.23}$$

The matrix $\mathbf{S}$ shifts data between the leads to compensate the variation on the electrical axis and scales the leads to compensate variations in amplitudes in different leads. These compensation can be performed by a rotation matrix $\mathbf{Q} \in \mathbb{R}^{R\times R}$ and a scaling matrix $\mathbf{D} \in \mathbb{R}^{R\times R}$. Then

$$\mathbf{S} = \mathbf{DQ}. \tag{2.24}$$

We desire to calculate $\mathbf{X}_V$ by estimating $\mathbf{D}$, $\mathbf{Q}$ and τ and cancel it from the original data $\mathbf{X}$. We need to devise an optimization criterion to find the optimal parameters of $\mathbf{X}_V$. Combining the equations (2.22) and (2.23), we have

$$\mathbf{X} - \mathbf{J}_\tau \bar{\mathbf{X}} \mathbf{S} = \mathbf{X}_A + \mathbf{N} \tag{2.25}$$

Note that both the noise $\mathbf{N}$ and the atrial activity $\mathbf{X}_A$ limits the model fitting. This problem can be mitigated by introducing an initial estimate $\tilde{\mathbf{X}}_A$ of $\mathbf{X}_A$ and subtracting it from $\mathbf{X}$. This estimate can be obtained from TQ segments in a AF ECG observation. Then (2.25) becomes

$$\mathbf{Y} - \mathbf{J}_\tau \bar{\mathbf{X}} \mathbf{S} = \mathbf{N} + \mathbf{X}_A - \tilde{\mathbf{X}}_A \tag{2.26}$$

where $\mathbf{Y} = \mathbf{X} - \tilde{\mathbf{X}}_A$. The optimal parameters are estimated by minimizing the estimation error

$$\varepsilon^2_{min} = \min_{\mathbf{D},\mathbf{Q},\tau} ||\mathbf{Y} - \mathbf{J}_\tau \bar{\mathbf{X}} \mathbf{D} \mathbf{Q}||^2_F, \tag{2.27}$$

where $\mathrm{tr}(\cdot)$ is the trace operator. This error can be written as

$$\varepsilon^2 = \mathrm{tr}(\mathbf{Y}\mathbf{Y}^T) + \mathrm{tr}(\mathbf{J}_\tau \bar{\mathbf{X}} \mathbf{D}\mathbf{D}^T \bar{\mathbf{X}}^T \mathbf{J}_\tau^T) - 2\mathrm{tr}(\mathbf{D}^T \bar{\mathbf{X}}^T \mathbf{J}_\tau^T \mathbf{Y}\mathbf{Q}^T). \tag{2.28}$$

It is difficult to find a closed-form solution to this problem, thus an alternate iterative minimization algorithm is used. The error is minimized by maximizing the last term of equation (2.28) under the assumption that $\mathbf{D}$ is known. Let $\mathbf{T} = \mathbf{D}^T \bar{\mathbf{X}}^T \mathbf{J}_\tau^T \mathbf{Y}$, then the term to be maximized is now written as $2\mathrm{tr}(\mathbf{T}\mathbf{Q}^T)$. The matrix $\mathbf{T}$ assumes a SVD $\mathbf{T} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Then $2\mathrm{tr}(\mathbf{T}\mathbf{Q}^T)$ is maximized for [21]

$$\hat{\mathbf{Q}} = \mathbf{U}\mathbf{V}^T. \tag{2.29}$$

Next, (2.29) is substituted into (2.28) and the error is minimized with respect to $\mathbf{D}$. Finally the diagonal entries of $\mathbf{D}$ are calculated by

$$\hat{\mathbf{d}}_r = \left(\left[\mathbf{J}_\tau \bar{\mathbf{X}} \right]^T_{:r} \left[\mathbf{J}_\tau \bar{\mathbf{X}} \right]_{:r} \right)^{-1} \left(\left[\mathbf{J}_\tau \bar{\mathbf{X}} \right]^T_{:r} \left[\mathbf{Z}\mathbf{Q}^{-1} \right]_{:r} \right), \quad r = 1, \ldots, R. \tag{2.30}$$

The optimal τ is found by means of a grid search in the possible interval of alignment $[-\Delta, \Delta]$. For each value of τ in the grid search, the equations (2.29) and (2.30) are calculated in an alternate fashion until the difference of the squared error $||\mathbf{Y} - \mathbf{J}_\tau \bar{\mathbf{X}} \mathbf{D}\mathbf{Q}||^2_F$ in two successive iterations is smaller than a certain threshold. Finally, the τ value which attains the smallest error value will be chosen as the optimal. A solution close to $\mathbf{Q} = \mathbf{D} = \mathbf{I}$, indicating that almost no spatial alignment would be required, is desirable. Then we initialize this algorithm with $\mathbf{D}_0 = \mathbf{I}$.

2.3 Blind Source Separation

The MRANC technique is based on the existence of reference signals that respect a couple of constraints, which are not always verified. The ABS and the spatiotemporal cancellation methods suppose that the QRST complex is quasi-periodic, which may not always be true due to the variations that occur in some patients. Furthermore these methods are not well prepared to deal with ectopic beats.

Lately [7, 22, 23, 24], a BSS approach based on linear instantaneous mixtures has been used to extract atrial activity from multilead ECG. Consider the instantaneous linear mixture model:

$$\mathbf{Y} = \mathbf{M}\mathbf{S} \tag{2.31}$$

where $\mathbf{Y} \in \mathbb{R}^{K \times N}$ is the observation matrix of K leads and N time samples, $\mathbf{M} \in \mathbb{R}^{K \times R}$ is an unknown mixing matrix that describes the propagation characteristics of the body and $\mathbf{S} \in \mathbb{R}^{R \times N}$ is a matrix that contains R unknown source signals. In practice, the only available information is the output ECG $\mathbf{Y}$, thus we need to estimate the sources signals and mixing matrix in (2.31), characterizing this as a BSS problem. In order to uniquely determine these matrices, we need to make some assumptions. For instance, suppose that $\mathbf{Y}$ is an ECG data matrix. We can argue that during AF, the atrial and ventricular activities are considered to present a certain statistical independence. Hence, we can impose this independence constraint to uniquely determine $\mathbf{M}$ and $\mathbf{S}$. In the following, we will discuss two techniques used to solve the BSS problem (2.31): PCA, which supposes decorrelation between sources, and ICA, which assumes high-order independence.

The BSS problem is affected by the scale and permutation ambiguities. A scale factor/permutation can be interchanged between a source and its mixing matrix column without modifying the observation. A greedy method, discussed in the appendix A, can be used to estimate the correct scaling and permutation of the estimated vectors and their unmixing columns. Note that although the model (2.31) does not take into account noise, in practice it is present. Nevertheless, the methods discussed in the following subsections are able to tackle this noise, as we will see.

2.3.1 Principal Component Analysis

The Principal Component Analysis (PCA) is a statistical tool that transforms a set of possibly correlated data into a set of uncorrelated data, called principal components [25, 9]. Consider a data vector $\mathbf{y} \in \mathbb{R}^N$ with $\mathbb{E}[\mathbf{y}] = \mathbf{0}$. One of the definitions of the PCA is that it performs the data decorrelation by finding a set of projection vectors, or principal directions. The projection vectors $\mathbf{w}_k \in \mathbb{R}^N$ with $||\mathbf{w}_k|| = 1, k = 1, \ldots, K$, project the data $\mathbf{y}$ to a scalar z_k, called kth principal component, that will have maximum variance. We need to find the set of principal directions $\mathbf{w}$, that will maximize the variance of the corresponding projections $z_k = \mathbf{w}_k^T\mathbf{y}$. The vector $\mathbf{w}_1$ will maximize the projected variance, thus

$$\begin{aligned}
\mathbf{w}_1^* &= \arg\max_{||\mathbf{w}_1||=1} \mathbb{E}[z_1^2] && (2.32)\\
&= \arg\max_{||\mathbf{w}_1||=1} \mathbb{E}[\left(\mathbf{w}_1^T\mathbf{y}\right)^2] && (2.33)\\
&= \arg\max_{||\mathbf{w}_1||=1} \mathbb{E}[\mathbf{w}_1^T\mathbf{y}\mathbf{y}^T\mathbf{w}_1] && (2.34)\\
&= \arg\max_{||\mathbf{w}_1||=1} \mathbf{w}_1\mathbf{R}_\mathbf{y}\mathbf{w}_1. && (2.35)
\end{aligned}$$

Since $\mathbf{w}_1$ has unit-norm, we can write (2.35) as

$$\mathbf{w}_1^* = \arg\max_{\mathbf{w}_1} \frac{\mathbf{w}_1\mathbf{R}_\mathbf{y}\mathbf{w}_1}{\mathbf{w}_1^T\mathbf{w}_1} \quad (2.36)$$

The quantity to be maximized is a Rayleigh quotient. Thus it will be maximized when $\mathbf{w}_1$ is the most significant eigenvector, which corresponds to the largest eigenvalue of $\mathbf{R}_\mathbf{y}$. The kth component can be found by subtracting the first $(k-1)$ first components vectors $z_k\mathbf{w}_k$ from $\mathbf{y}$:

$$\mathbf{y}_{k-1} = \mathbf{y} - \sum_{s=1}^{k-1} z_s\mathbf{w}_s \quad (2.37)$$

and then finding the vector $\mathbf{w}_k$ that will maximize the projected variance of $z_k = \mathbf{w}_k^T\mathbf{y}_{k-1}$:

$$\mathbf{w}_k^* = \arg\max_{\mathbf{w}_k} \frac{\mathbf{w}_k\mathbf{R}_{\mathbf{y}_{k-1}}\mathbf{w}_k}{\mathbf{w}_k^T\mathbf{w}_k}. \quad (2.38)$$

It turns out that this gives the remaining eigenvectors of $\mathbf{R}_\mathbf{y}$. Let $\mathbf{R}_\mathbf{y} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ be the eigendecomposition of the data covariance matrix. Since the principal vectors are the eigenvectors of $\mathbf{R}_\mathbf{y}$, i.e. the columns of $\mathbf{U}$, then

$$\begin{aligned}
\mathbf{z} &= \mathbf{U}^T\mathbf{y} && (2.39)\\
\mathbf{y} &= \mathbf{U}\mathbf{z}, && (2.40)
\end{aligned}$$

where $\mathbf{z} = [z_1, z_2, \ldots, z_N]^T$.

We can extend this definition of PCA to matrices, storing sample realizations of the several random vectors containing N samples of random variable y_k. Let $\mathbf{y}_k \in \mathbb{R}^{N\times 1}$ be a measurement column vector. We arrange K measurement vectors in the rows of the matrix $\mathbf{Y} \in \mathbb{R}^{K\times N}$. The first principal vector $\mathbf{w}_1 \in \mathbb{R}^{K\times 1}$ will maximize the variance of the projected vector $\mathbf{z}_1 = \mathbf{Y}^T\mathbf{w}_1$:

$$\begin{aligned}
\mathbf{w}_1^* &= \arg\max_{||\mathbf{w}_1||=1} \mathbf{z}^T\mathbf{z} && (2.41)\\
&= \arg\max_{||\mathbf{w}_1||=1} \mathbf{w_1}^T\mathbf{Y}\mathbf{Y}^T\mathbf{w_1} && (2.42)\\
&= \arg\max_{\mathbf{w}_1} \frac{\mathbf{w_1}^T\mathbf{R_Y}\mathbf{w_1}}{\mathbf{w_1}^T\mathbf{w_1}}. && (2.43)
\end{aligned}$$

Similarly to the vector case, the quantity to be maximized will be maximum when $\mathbf{w}_1$ is the eigenvector associated with the largest eigenvalue of $\mathbf{R_Y}$. Let

$$\mathbf{R_Y} = \mathbf{U}\mathbf{D}\mathbf{U}^T \tag{2.44}$$

be the eigendecomposition of $\mathbf{Y}\mathbf{Y}^T$. Then

$$\mathbf{Z} = \mathbf{U}^T\mathbf{Y} \tag{2.45}$$
$$\mathbf{Y} = \mathbf{U}\mathbf{Z} \tag{2.46}$$

where $\mathbf{U} \in \mathbb{R}^{K\times K}$ is the principal directions matrix and $\mathbf{Z} \in \mathbb{R}^{K\times N}$ is the principal components matrix. Finally, we can check that the principal components are uncorrelated. Consider the covariance matrix of the principal components $\mathbf{R_Z}$:

$$\mathbf{R_Z} = \mathbb{E}[\mathbf{Z}\mathbf{Z}^T] = \mathbf{U}^T\mathbf{R_Y}\mathbf{U} = \mathbf{D}, \tag{2.47}$$

whose diagonal structure shows indeed that the rows of $\mathbf{Z}$ are uncorrelated.

The PCA will transform original data into a uncorrelated data, which is easier to analyze, due to the second-order independence of its principal components. However, not all principal components need to be kept. If a principal component is not significant, we can ignore it in the principal components matrix $\mathbf{Z}$. Consider the principal components and principal direction matrices with only M rows $\mathbf{Z}_M \in \mathbb{R}^{M\times N}$ and $\mathbf{U}_M \in \mathbb{R}^{K\times M}$, respectively. Then $\mathbf{Z}_M = \mathbf{U}_M^T\mathbf{Y}$. If we we discard some principal components with small correspondent eigenvalue, the total squared reconstruction error $||\mathbf{U}\mathbf{Z} - \mathbf{U}_M\mathbf{Z}_M||^2$ will be small. This property allows us to perform dimensionality reduction on data, by representing it only as its M most significant principal components.

PCA solution to the BSS problem

The data matrix $\mathbf{Y}$ assumes a SVD

$$\mathbf{Y} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^T, \tag{2.48}$$

where $\Sigma = (\sigma_k)$ is a $K \times N$ rectangular diagonal matrix of the singular values σ_k, $\hat{\mathbf{U}} \in \mathbb{R}^{K\times K}$ is the left singular vectors matrix and $\hat{\mathbf{V}} \in \mathbb{R}^{N\times N}$ is the right singular vectors matrix. Both $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ are orthogonal matrices and we assume that the columns of the source matrix $\mathbf{S}$ have unit norm and uncorrelated.

We can link it to the PCA in the calculation of the covariance matrix of $\mathbf{Y}$:

$$\mathbf{R_Y} = \mathbf{Y}\mathbf{Y}^T \tag{2.49}$$
$$= \mathbf{M}\mathbf{S}\mathbf{S}^T\mathbf{M}^T \tag{2.50}$$
$$= \mathbf{M}\mathbf{M}^T. \tag{2.51}$$

On the other hand, the covariance matrix $\mathbf{R_Y}$ can be estimated via the ensemble average:

$$\mathbf{R_Y} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^T. \tag{2.52}$$

Substituting (2.48) into (2.52):

$$\mathbf{R_Y} = \frac{1}{N}\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^T\hat{\mathbf{V}}\hat{\mathbf{\Sigma}}\hat{\mathbf{U}}^T \tag{2.53}$$
$$= \frac{1}{N}\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}^2\hat{\mathbf{U}}^T. \tag{2.54}$$

Comparing (2.51) and (2.54), we see that $\hat{\mathbf{M}}_{PCA} = \frac{1}{\sqrt{N}}\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}$. Hence, the remainder factors of the SVD (2.54) give the source matrix $\hat{\mathbf{S}}_{PCA} = \sqrt{N}\hat{\mathbf{V}}^T$.

Note that the PCA solution to the BSS supposes that the mixing matrix has orthogonal columns due to the left singular matrix. If $\mathbf{M}$ does not conform to this requirement, PCA will fail to perform the BSS. In the problem of atrial activity extraction, imposing an orthogonal structure to the mixing matrix may not correspond to a practical possibility due to the proximity of the atrial and ventricular sources.

2.3.2 Independent Component Analysis

PCA assumes that the mixing matrix $\mathbf{M}$ has orthogonal columns, which is a strong constraint. We can relax these constraints on the mixing matrix and impose a stronger constraint on the source matrix. ICA [26] admits that the sources are independent, typically up to fourth order. Even though it is a strong constraint, it is not an unrealistic assumption. Indeed in AF, the atrial and ventricular components have a certain degree of independence.

Furthermore, to be able to uniquely retrieve the independent sources, we have to assume that the sources are non-Gaussian. To see why, consider the following scenario. Let $\mathbf{s} = [s_1, s_2, \ldots, s_N]^T$ be a vector of independent variables s_n with identical distribution, an arbitrary full-rank mixing matrix $\mathbf{M} \in \mathbb{R}^{N\times N}$ and an observation vector $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$. Considering the instantaneous linear mixture model, we have $\mathbf{y} = \mathbf{Ms}$. To estimate one of the independent components, we consider a linear combination of y_n, expressed as $\hat{s} = \mathbf{w}^T\mathbf{y}$, where $\mathbf{w}$ is a projection vector to be determined. If $\mathbf{w}$ was one of the columns of the inverse of $\mathbf{M}$, it would recover an independent component directly. However, we do not know the $\mathbf{M}$ in practice. Nevertheless, an estimation can be found based on the Central Limit Theorem which affirms that the distribution of a sum of independent random variables tends toward a Gaussian distribution under certain conditions. Let us define $\mathbf{g} = \mathbf{M}^T\mathbf{w}$. Then $\hat{s} = \mathbf{w}^T\mathbf{y} = \mathbf{w}^T\mathbf{Ms} = \mathbf{g}^T\mathbf{s}$, we verify that $\hat{s}$ is a linear combination of s_n weighted by g_n. According to the Central Limit Theorem, the sum $\hat{s} = \mathbf{g}^T\mathbf{s} = g_1s_1 + g_2s_2 + \ldots + g_Ns_N$ is more Gaussian than any of the s_n. This sum $\hat{s}$ becomes less Gaussian when it equals only one of the s_n. In this case, the projection vector would be $\mathbf{g} = [0, 0, \ldots, g_n, \ldots, 0]^T$, leading to $\hat{s} = g_ns_n$. Observe that finding a projection that maximizes the nongaussianity of the projection, would lead to an independent component s_n. Then, to estimate it, we could take a $\mathbf{g}$ with only one nonzero element that maximizes the nongaussianity of $\hat{s} = \mathbf{w}^T\mathbf{y} = \mathbf{g}^T\mathbf{s}$, resulting in an independent component s_n.

In order to measure the nongaussianity of a random variable, we will adopt its fourth order cumulant as measurement defined by

$$\mathrm{Cum}(z_i, z_j, z_k, z_l) = \mu_{ijkl} - R_{ij}R_{kl} - R_{ik}R_{jl} - R_{il}R_{jk}. \tag{2.55}$$

where $\mathbf{z} \in \mathbb{R}^M$ is a random vector, $\mu_{ijkl} = \mathbb{E}\{z_iz_jz_kz_l\}$ is a fourth order moment of $\mathbf{z}$ and R_{mn} represents $[\mathbf{R_z}]_{mn} = \mathbb{E}\{z_mz_n\}$. It can be show that when $\mathbf{z}$ is made of Gaussian random variables, its fourth order cumulant is zero for every index (i, j, k, l). The kurtosis of a random variable z is defined as its fourth order cumulant:

$$\mathrm{kurt}(z) = \mathrm{Cum}(z, z, z, z) = \mathbb{E}\left\{z^4\right\} - 3\left(\mathbb{E}\left\{z^2\right\}\right)^2 \tag{2.56}$$

If the kurtosis of z is negative, it is called a subgaussian variable, exhibiting a "flat" pdf. Moreover, when the kurtosis of z is positive, it is called a supergaussian variable, often taking values around zero.

Before discussing the estimation of the independent sources, let us assume that there are M mutually statistical independent sources in which at least $(M-1)$ are non gaussian. Next, we will discuss an algorithm for performing the estimation of the independent components by maximizing the nongaussianity of the estimated values.

RobustICA

Let $\mathbf{y}$ be an observation vector, $\mathbf{w}$ a vector that will project $\mathbf{y}$ to $\hat{s} = \mathbf{w}^T\mathbf{y}$, the estimation of the independent component s. We seek finding $\mathbf{w}$ that will lead to an optimal estimation of s. Let the Kurtosis Maximization criterion Ψ is defined be defined by:

$$\Psi(\hat{s}) = \frac{\mathrm{kurt}(\hat{s})}{\mathbb{E}\left\{\hat{s}^2\right\}^2} = \frac{\mathbb{E}\left\{\hat{s}^4\right\} - 3\left(\mathbb{E}\left\{\hat{s}^2\right\}\right)^2}{\mathbb{E}\left\{\hat{s}^2\right\}^2} \tag{2.57}$$

The RobustICA algorithm [27] consists of maximizing the criterion by finding a suitable search direction which will lead to the global optimum of the criterion. The search direction can be defined as the gradient of the criterion (2.57), $\nabla\Psi(\hat{s})$. The extraction vector can be iteratively estimated as $\mathbf{w}^{(i+1)} = \mathbf{w}^{(i)} + \eta_i\nabla\Psi(\hat{s})$ until the convergence is attained according to a suitable stopping criterion. The step parameter η_i is automatically found by RobustICA at each iteration. After the convergence, one independent source can be obtained by projecting the observation vector $\mathbf{y}$ into the estimated extraction vector. If more than one source are to be extracted, we employ this method several times. After extracting one source, it has to be removed from the observation vector, so it will not be estimated again.

The ICA framework can be used to extract the atrial activity in initial stages of AF ECG [7]. In this context, the amplitude of the atrial activity signal presents an organized and highly nongaussian behavior, motivating the use of ICA. Indeed, the statistical independence implied by nongaussianity has a physiological justification. During AF, the wavefronts generated by the sino-atrial node affects the atrio-ventricular node in a random way, generating uncoupled atrial and ventricular activities. This rupture of synchronism between these activities leads to their (approximate) statistical independence.

However, in more developed stages of AF, the amplitude of the atrial activity becomes more disorganized, presenting a quasi-Gaussian probability distribution. Thus some additional steps have to be used in order to recover the AF amplitude activity. In [22], the atrial activity has been extracted in a two-step procedure. First of all, classical ICA detects the components that correspond to the ventricular activity, which presents a nongaussian distribution. The remaining components are composed of atrial activity and noise. The second step consists of extracting the atrial activity from these remaining components by employing a technique known as Second-Order Blind Identification (SOBI) [28]. Basically it will separate sources with narrow-band spectra, which is the case of atrial activity. Among the components obtained from SOBI, the atrial activity can be finally retrieved by inspecting the spectral concentration around their fundamental frequency. Narrow peaks in the 3-9 Hz frequency band may indicate the atrial activity.

The narrowband characteristic of the atrial activity during AF is also exploited by a technique called RobustICA-f [27]. It exploits the sparsity of the AF signal in the frequency domain caused by its narrow spectral support. Hence, in the frequency domain, the atrial signal will have a strong kurtosis, justifying the suitability of the ICA approach based on higher-order stationarity.

In this chapter, we have discussed some matrix methods for atrial activity extraction from ECG. Two powerful source extraction techniques, PCA and ICA, were introduced as a BSS approach. In the following chapters, we will introduce a tesor-based BSS technique to atrial activity extraction from ECG signals.

Chapter 3

Tensor models

Lately, multilinear algebra methods, or tensor methods, have been applied to solve problems related to signal processing in fields such as wireless communications systems [11, 12], audio source separation [13], pattern recognition [14], etc. Tensors present some interesting properties. Some tensor decompositions are guaranteed to be unique under mild conditions. This allows blind identification and source separation with fewer sensors or samples in some cases. Aiming at applying tensor methods to biomedical signal processing, now we discuss the fundamentals of tensor models and possible ways of modeling ECG signals as tensors.

3.1 Definitions

Originally the term tensor was introduced in the 19th century to denote a multilinear map from one subspace to another [29]. But several decades ago, it started to be employed to denote a multilinear array of data, as will be the case in this report.

Consider a Nth order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ whose elements are represented by $x_{i_1 i_2 \cdots i_N}$. The order of a tensor is defined as the number of its dimensions. From this definition, it is straightforward to see that scalars can be considered as zeroth-order tensors, vectors as first-order tensors and matrices as second-order tensors. Indeed tensors subsume these data structures.

Subarrays can be obtained from a tensor by fixing one or more indices while varying the others. Fibers are defined as fixing every element index but one. For example, the mode-1 fibers of $\mathcal{X}$ are given by $x_{:i_2 \ldots i_N}$ where ":" denotes the full range of the corresponding index. Matrix slices are defined as fixing all but two indices. For instance, the frontal slices of a third tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ are $\mathbf{A}_{::i_3} \in \mathbb{R}^{I_1 \times I_2}$ for $i_3 = 1, \ldots, I_3$. Fibers are the generalization of row/columns and slices are multidimensional sections of a tensor.

3.2 Operations

Tensors can be transformed into matrices by means of a linear mapping. A tensor element $(i_1, i_2, \ldots, i_N)$ can be mapped into a matrix at (i_n, j) where

$$j = 1 + \sum_{k=1 \neq n}^{N} (i_k - 1) J_k \quad \text{with} \quad J_k = \prod_{m=1 \neq n}^{k-1} I_m$$

However this definition does not express intuitively this operation. Equivalently the n-mode *matricization* (or unfolding) $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times I_1 I_2 \cdots I_{n-1} I_{n+1} \cdots I_N}$ of the Nth order tensor $\mathcal{X}$ is obtained by setting its columns as the mode-n fibers of $\mathcal{X}$.

The outer product $(\mathcal{M} \circ \mathcal{N}) \in \mathbb{R}^{I_1 \times \ldots \times I_N \times J_1 \times \ldots \times J_Q}$ between the tensors $\mathcal{M} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_P}$ and $\mathcal{N} \in \mathbb{R}^{J_1 \times J_2 \ldots \times J_Q}$ is defined as

$$(\mathcal{M} \circ \mathcal{N})_{i_1 i_2 \ldots i_P j_1 j_2 \ldots j_Q} = m_{i_1 i_2 \ldots i_P} n_{j_1 j_2 \ldots j_Q}. \tag{3.1}$$

The Frobenius norm of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is defined as

$$\|\mathcal{X}\|_F = \sqrt{\sum_{i_1}^{I_1} \sum_{i_2}^{I_2} \cdots \sum_{i_N}^{I_N} x_{i_1 i_2 \ldots i_n}^2}. \tag{3.2}$$

The inner product of two identically sized tensors $\mathcal{P}, \mathcal{Q} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$ is defined as

$$\langle \mathcal{P}, \mathcal{Q} \rangle = \sum_{i_1}^{I_1} \sum_{i_2}^{I_2} \ldots \sum_{i_N}^{I_N} p_{i_1 i_2 \ldots i_N} q_{i_1 i_2 \ldots i_N} \tag{3.3}$$

It is easy to see that $\langle \mathcal{X}, \mathcal{X} \rangle = ||\mathcal{X}||_F^2$.

The mode-n product between a Nth order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and a matrix $\mathbf{U}^{(n)} \in \mathbb{R}^{J \times I_n}$ results in a tensor $\mathcal{Y} = \mathcal{X} \times_n \mathbf{U}^{(n)}$ of size $I_1 \times I_2 \ldots \times I_{n-1} \times J \times I_{n+1} \ldots \times I_N$. Its elements are given by

$$y_{i_1 i_2 \ldots i_{n-1} j i_{n+1} \ldots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \ldots i_N} u^{(n)}_{j i_n}$$

The mode-n product can be equivalently defined as the product between the mode-n fibers of $\mathcal{X}$ and the matrix $\mathbf{U}^{(n)}$. This product can be expressed by unfolding the tensor $\mathcal{X}$ in the nth mode and multiplying the unfolded tensor $\mathbf{X}_{(n)}$ by $\mathbf{U}^{(n)}$. The result is the unfolded version of the mode-n product:

$$\mathbf{Y}_{(n)} = \mathbf{U}^{(n)} \mathbf{X}_{(n)}.$$

Let $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, $\mathcal{Y} \in \mathbb{R}^{J_1 \times J_2 \times \ldots \times J_N}$ and $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times J_n}$. The mode-$n$ product

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \ldots \times_N \mathbf{U}^{(N)^T}$$

can be expressed by means of the Kronecker product:

$$\mathbf{Y}_{(n)} = \mathbf{U}^{(n)} \mathbf{X}_{(n)} \left(\mathbf{U}^{(N)} \otimes \ldots \otimes \mathbf{U}^{(n+1)} \otimes \mathbf{U}^{(n-1)} \otimes \ldots \otimes \mathbf{U}^{(1)} \right)^T \tag{3.4}$$

3.3 Rank

Unlike matrices, an interesting aspect of tensors is that they have many definitions of rank. Indeed some definitions of rank are not bounded by the dimensions of the tensor. An Nth order rank-one tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$ is formed by the outer product of N vectors $\mathbf{v}_n \in \mathbb{R}^{I_n \times 1}$ for $n = 1, \ldots, N$:

$$\mathcal{B} = \mathbf{v}_1 \circ \mathbf{v}_2 \ldots \circ \mathbf{v}_N$$

The *rank* of a Nth order tensor $\mathcal{X}$ is defined as the minimum number of rank 1 tensors whose sum generates $\mathcal{X}$. Although its definition is analogue to the matrix rank, they are different in some aspects. First of all, the rank of a real valued tensor may be different over $\mathbb{R}$ and $\mathbb{C}$. Secondly, there are no simple algorithms to calculate it. Let the mode-n rank of $\mathcal{X}$ denote the dimension of the subspace spanned by its mode-n fibers. A tensor $\mathcal{X}$ is said to have *multilinear rank* $(R_1, R_2, \ldots, R_N)$ if the rank of $\mathbf{X}_{(n)}$ equals R_n for $n = 1 \ldots N$. Let the *Kruskal rank* $k_{\mathbf{M}}$ of $\mathbf{M}$ be defined as the maximum value k such that any k columns of $\mathbf{M}$ are linearly independent. This definition will be useful later for assuring the uniqueness of the CPD decomposition.

3.4 Tensor Decompositions

In this section, the Tucker and the Canonical Polyadic decompositions will be discussed for Nth order tensors.

3.4.1 Tucker Decomposition

The Tucker decomposition was first introduced by Tucker in 1963 [30]. It decomposes a tensor in a core tensor multiplied by a matrix along each mode. For a Nth order tensor $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \ldots \times_N I_N}$, its Tucker decomposition is

$$\mathcal{Y} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \ldots \times_N \mathbf{A}^{(N)} \tag{3.5a}$$

$$= [\![\mathcal{G}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!] \tag{3.5b}$$

$$= \sum_{r_1=1}^{R_1} \ldots \sum_{r_N=1}^{R_N} g_{r_1, \ldots, r_N} \mathbf{A}^{(1)}_{:r_1} \circ \mathbf{A}^{(2)}_{:r_2} \circ \ldots \circ \mathbf{A}^{(N)}_{:r_N} \tag{3.5c}$$

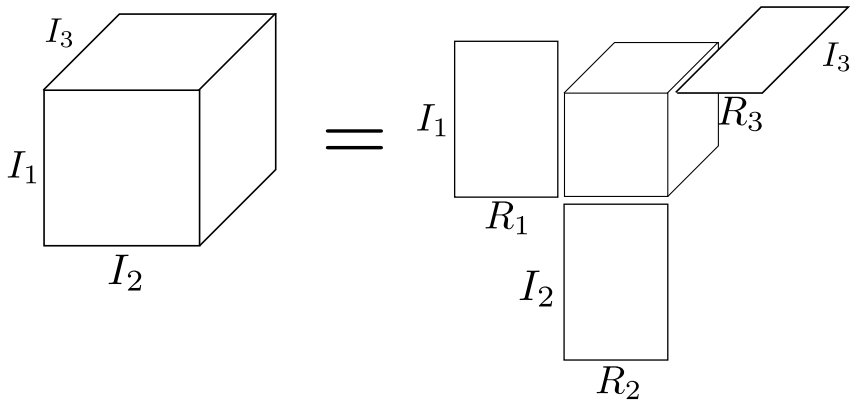

Figure 3.1: The Tucker Decomposition for a third order tensor.

where $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \ldots \times R_N}$ is the core tensor, $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R_n}$ for $n = 1, \ldots, N$ are the factor matrices, $g_{r_1,\ldots,r_N}$ are the elements of $\mathcal{G}$ and $\mathbf{A}^{(n)}_{:r_n}$ is the r_nth column of $\mathbf{A}^{(n)}$. The elements of $\mathcal{Y}$ are

$$y_{i_1 i_2 \ldots i_N} = \sum_{r_1=1}^{R_1} \cdots \sum_{r_N=1}^{R_N} g_{r_1,\ldots,r_N} a^{(1)}_{i_1 r_1} a^{(2)}_{i_2 r_2} \ldots a^{(N)}_{i_N r_N}, \tag{3.6}$$

and its unfolded tensor in the mode-n (3.4) is

$$\mathbf{Y}_{(n)} = \mathbf{A}^{(n)} \mathbf{G}_{(n)} \left(\mathbf{A}^{(N)} \otimes \ldots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \ldots \otimes \mathbf{A}^{(1)} \right)^T. \tag{3.7}$$

Tucker decompositions are not unique. Consider the tensor $\mathcal{G} \in \mathbb{R}^{P \times Q \times R}$, its factor matrices $\mathbf{A}_{P \times P}, \mathbf{B}_{Q \times Q}$, $\mathbf{C}_{R \times R}$ and the non-singular matrices $\mathbf{U}_{P \times P}$, $\mathbf{V}_{Q \times Q}$ and $\mathbf{W}_{R \times R}$. Then the relationship

$$[\![\mathcal{G}, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = [\![\mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}, \mathbf{A}\mathbf{U}^{-1}, \mathbf{B}\mathbf{V}^{-1}, \mathbf{C}\mathbf{W}^{-1}]\!]$$

shows that there are infinite Tucker decomposition for a given tensor. By imposing some constraints, however, it may become unique. For instance, consider a Nth order tensor $\mathcal{T} = [\![\mathcal{S}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!]$. If $\mathbf{A}^{(n)}$ for $n = 1, 2, \ldots, N$ are column-wise orthonormal and $\mathcal{S}$ is all orthogonal, i.e. all tensor slices of order $(N-1)$ are mutually orthogonal in every mode, then its Tucker decomposition will be unique [31]. This constrained decomposition is called Higher-Order SVD (HOSVD). Other constraints such as non-negativity and sparsity can be imposed upon this model, leading to Nonnegative Tucker Decompositions (NTD) [32].

The Tucker decomposition $\mathcal{T} = [\![\mathcal{G}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!]$ can be calculated by means of the HOSVD. It consists of performing the SVD of $\mathbf{T}_{(n)} = \mathbf{U}_n \mathbf{\Sigma}_n \mathbf{V}_n^T$ and making $\mathbf{A}^{(n)} = \mathbf{U}_n$ for $n = 1, \ldots, N$. Finally $\mathcal{G} = \mathcal{X} \times_1 \mathbf{A}^{(1)^T} \times_2 \mathbf{A}^{(2)^T} \ldots \times_N \mathbf{A}^{(N)^T}$. However, the HOSVD is not optimal in giving the best fit considering the norm of the difference. Then, De Lathauwer et al [33], proposed the Higher-Order Orthogonal Iteration (HOOI). Let $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$. Then the HOOI aims at solving the following problem:

$$\min_{\mathcal{G}, \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}} \left\| \mathcal{X} - [\![\mathcal{G}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!] \right\|_F \tag{3.8}$$

subject to

$$\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \ldots \times R_N}$$

and

$$\mathbf{A}^{(n)} \in \mathbf{R}^{I_n \times R_n} \text{ column-wise orthogonal for } n = 1, \ldots, N.$$

Having predefined $R_1, \ldots, R_N$, the solution of the problem (3.8) can be computed by means of a Alternating Least-Squares (ALS) [34]:

Step 1 Initialize $\mathbf{A}^{(n)}$ for $n = 1, \ldots, N$ using HOSVD.

Step 2 While the algorithm has not converged, repeat:

- for $n = 1, \ldots, N$

– Calculate $\hat{\mathcal{G}} = \mathcal{X} \times_1 \mathbf{A}^{(1)^T} \ldots \times_{(n-1)} \mathbf{A}^{(n-1)^T} \times_{(n+1)} \mathbf{A}^{(n+1)^T} \ldots \times_N \mathbf{A}^{(N)^T}$.

– Set $\mathbf{A}^{(n)}$ as the R_n most significant left eigenvectors of $\hat{\mathbf{G}}_{(n)}$.

Step 3 Calculate $\mathbf{G} = \mathcal{X} \times_1 \mathbf{A}^{(1)^T} \times_2 \mathbf{A}^{(2)^T} \ldots \times_N \mathbf{A}^{(N)^T}$.

Timmerman and Kiers [35] devised an automatic procedure to choose the numbers of components to use in a Tucker analysis which they called "DIFFIT" (DIFFerence In fiT).

3.4.2 Canonical Polyadic Decomposition

Another possibility of assuring the uniqueness of the Tucker model is imposing a diagonality constraint upon the core tensor. Consider the elements of the constrained core tensor $\mathcal{D} \in \mathbb{R}^{R \times R \ldots \times R}$:

$$d_{r_1 r_2 \ldots r_n} = \begin{cases} d_r, \text{ if } r_1 = r_2 = \ldots = r_n \\ 0, \quad \text{otherwise.} \end{cases}$$

The equation (3.5) becomes

$$\begin{aligned} \mathcal{Y} &= \mathcal{D} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \ldots \times_N \mathbf{A}^{(N)} && (3.9) \\ &= [\![\mathcal{D}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!] && (3.10) \\ &= \sum_{r=1}^{R} d_r \mathbf{A}^{(1)}_{:r} \circ \mathbf{A}^{(2)}_{:r} \circ \ldots \circ \mathbf{A}^{(N)}_{:r} && (3.11) \end{aligned}$$

where R is a positive constant, $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R_n}$ for $n = 1, \ldots, N$ are the factor matrices and $\mathbf{A}^{(n)}_{:r_n}$ is the r_nth column of $\mathbf{A}^{(n)}$. Often the elements d_r absorb the length of the columns of the factor matrices. The equation (3.9) is called Canonic Polyadic Decomposition (CPD), also known as PARAFAC [36].

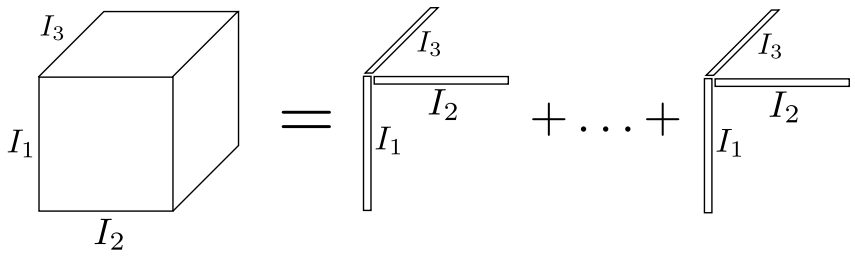

Figure 3.2: The CPD of a third order tensor.

The tensor $\mathcal{D}$ can be considered to be an identity tensor $\mathcal{I}$, with

$$d_{r_1 r_2 \ldots r_n} = \delta_{r_1 r_2 \ldots r_n}.$$

From now on, we will consider the CPD with identity core tensor. The elements of $\mathcal{Y}$ are given by

$$y_{i_1 i_2 \ldots i_N} = \sum_{r=1}^{R} a^{(1)}_{i_1 r} a^{(2)}_{i_2 r} \ldots a^{(N)}_{i_N r}. \tag{3.12}$$

Note that the CPD reveals immediately the rank of a tensor. Let $\mathcal{Y}$ be a Nth order tensor. Its mode-n unfolded tensor is

$$\begin{aligned} \mathbf{Y}_{(n)} &= \mathbf{A}^{(n)} \left(\mathbf{A}^{(N)} \odot \ldots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \ldots \odot \mathbf{A}^{(1)} \right)^T && (3.13a) \\ &= \mathbf{A}^{(n)} \mathbf{V}^{\{n\}^T}, \, 1 \leq n \leq N, && (3.13b) \end{aligned}$$

where "$\odot$" denotes the Khatri-Rao product, defined as the columnwise Kronecker product. A uniqueness condition of CPD was devised by Sidiropoulos and Bro [37], which affirms that CPD is essentially unique if

$$\sum_{n=1}^{N} k_{\mathbf{A}^{(n)}} \geq 2R + (N - 1)$$

where N is the order of the tensor, R is the rank of the tensor and $k_{\mathbf{A}^{(n)}}$ is the Kruskal rank of the factor matrix $\mathbf{A}^{(n)}$. Recently, more relaxed conditions have been obtained when one factor matrix has full column rank [38]. A recent condition devised by Domanov and De Lathauwer requires only that the components are "sufficiently different" and there are not many of them [39].

Several constraints such as sparsity, non-negativity, smoothness, structured matrix factors etc., can be applied to the CPD model. There are many decompositions that result from constraints in this model, like PARATUCK [40], CONFAC [41], etc.

In order to calculate the CPD of $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \ldots I_N}$, we need to solve the following problem

$$\min_{\hat{\mathbf{A}}^{(1)} \ldots \hat{\mathbf{A}}^{(N)}} \left\| \mathcal{Y} - [\![\mathcal{I}, \hat{\mathbf{A}}^{(1)}, \ldots, \hat{\mathbf{A}}^{(N)}]\!] \right\|_F . \tag{3.14}$$

where $\hat{\mathbf{A}}^{(n)} \in \mathbb{R}^{I_n \times R_n}$ for $n = 1, \ldots, N$, are the estimated factor matrices. There are three groups of algorithms to calculate the CPD: alternating, gradient-based and direct (non-iterative) algorithms [42]. We will briefly discuss the calculation of CPD by means of the Alternating Least-Squares (ALS), which belongs to the first group of algorithms. It will solve (3.14) for $\hat{\mathbf{A}}^{(n)}$ while fixing the others factor matrices and will continue the entire procedure until the convergence according to a defined criterion. This method supposes that the number of components R is known. Bro and Kiers [43] proposed a consistency diagnostic called CORCONDIA that indicates if a given R is suitable for a decomposition. The CPD-ALS algorithm can be summarized as follows:

Step 1 Initialize the factor matrices $\hat{\mathbf{A}}^{(n)}$ for $n = 1, \ldots, N$.

Step 2 While the algorithm has not converged, repeat

- for $n = 1, \ldots, N$
  - $\hat{\mathbf{A}}^{(n)} = \mathbf{X}^{(n)} \left(\hat{\mathbf{A}}^{(N)} \odot \ldots \odot \hat{\mathbf{A}}^{(n+1)} \odot \hat{\mathbf{A}}^{(n-1)} \odot \ldots \odot \hat{\mathbf{A}}^{(1)} \right)^T$

3.5 Multilinear ECG signals

Recently, EEG signals have been modeled as Space-Time-Frequency (STF) tensors. In [44, 45], a STF tensor was obtained by calculating Short-Time Fourier Transform (STFT) matrices from observation electrodes and stacking them in a tensor. Subsequently, this tensor was decomposed by means of CPD to perform artifact detection and noise removal. Continuous Wavelet Transform (CWT) could also be used to obtain STF tensors, in which the frequency would correspond to wavelet scale. The EEG signals were later recovered by employing an inverse transformation in each spatial slice of the STF tensor.

Inspired by the recent work on multilinear EEG signals, Huang and Zhang [46] proposed a feature extraction method for ECG signals based on STF ECG tensors. The ECG signals were segmented in heartbeats, and then the data in all leads corresponding to each segmented heartbeat were transformed in a tensor by stacking their STFT matrices. In that paper, there was not a discussion about the trade-off between the STFT's window length and the spectral resolution, which is fundamental for obtaining precise results in ECG analysis. Hunyadi et al [47] built a STF tensor from EEG data using CWT with a Mexican hat wavelet. Analogously, an ECG STF tensor can be built by performing the CWT of each ECG lead using a DB-5 wavelet, which [48] pointed to be well adapted to capture important ECG features.

An interesting fact of STF ECG tensors is that they are considerably sparse. Since the spectrum of ECG is concentrated mostly in the 0 Hz - 50 Hz range (Section 1.3), its frequency-domain representation will be mostly sparse for 1 kHz sampling rates. Sparsity is an interesting property that can be explored, for example, by compressive sensing techniques [49].

In [50], Niknazar et al. proposed an extraction of event-related sources via robust tensor decomposition. They explore the event-related nature of ECG signals to construct a data tensor with space, event-synchronized window and time as modes. Each heartbeat is placed in an observation matrix and synchronized relative to the R peak of different heart beats. These matrices are then stacked in a third-order tensor. They apply this modeling for fetal ECG extraction. After building the event-synchronized tensor, it is decomposed in a modified CPD that permits the extraction of low power sources compared to others sources, such as fetal ECG compared to the maternal ECG.

In this chapter, we have presented some basic definitions of tensors and some decompositions such as CPD and Tucker. In the next chapter, another decomposition will be introduced and applied in the BSS problem. Subsequently, we will use this decomposition for extracting atrial activity from ECG signals.

Chapter 4

Block Term Decomposition

In Chapter 2, we have discussed BSS approaches to the atrial activity extraction based on matrix methods that explore some statistical parameters of the atrial components. Recently, De Lathauwer [51] proposed a new tensorial approach to solve the BSS problem based on a tensor decomposition in block terms. In this chapter, we will discuss the BSS method based on this Block Term Decomposition (BTD) and assess its performance to extract the atrial activity from AF ECG signals.

4.1 Definitions

Consider a third order tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. Its BTD [52] in rank-$(L_r, L_r, 1)$ terms is defined as

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{E}_r \circ \mathbf{c}_r, \tag{4.1}$$

in which $\mathbf{E}_r \in \mathbb{R}^{I_1 \times I_2}$ has rank L_r and $\mathbf{c}_r \in \mathbb{R}^{I_3 \times 1}$. If we decompose

$$\begin{aligned} \mathbf{E}_r &= \mathbf{D}_r^{(1)} \mathbf{D}_r^{(2)^T} & (4.2) \\ &= \sum_{r'=1}^{L_r} \mathbf{d}_{r,r'}^{(1)} \circ \mathbf{d}_{r,r'}^{(2)}, & (4.3) \end{aligned}$$

in which $\mathbf{d}_{r,r'}^{(1)} \in \mathbb{R}^{I_1 \times 1}$, $\mathbf{d}_{r,r'}^{(2)} \in \mathbf{R}^{I_2 \times 1}$, $\mathbf{D}_r^{(1)} = \left[\mathbf{d}_{r,1}^{(1)}, \ldots, \mathbf{d}_{r,L_r}^{(1)}\right]_{I_1 \times L_r}$ and $\mathbf{D}_r^{(2)} = \left[\mathbf{d}_{r,1}^{(2)}, \ldots, \mathbf{d}_{r,L_r}^{(2)}\right]_{I_2 \times L_r}$, then (4.1) can be written as

$$\begin{aligned} \mathcal{T} &= \sum_{r=1}^{R} \left(\mathbf{D}_r^{(1)} \mathbf{D}_r^{(2)^T}\right) \circ \mathbf{c}_r & (4.4) \\ &= \sum_{r=1}^{R} \left(\sum_{r'=1}^{L_r} \mathbf{d}_{r,r'}^{(1)} \circ \mathbf{d}_{r,r'}^{(2)}\right) \circ \mathbf{c}_r & (4.5) \end{aligned}$$

Each rth term of (4.5) is rank-$(L_r, L_r, 1)$. The elements of $\mathcal{T}$ are defined as

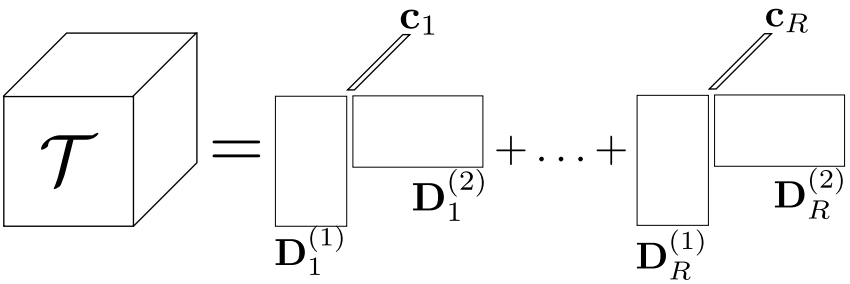

Figure 4.1: Block Term Decomposition in rank-$(L_r, L_r, 1)$ terms.

$$(\mathcal{T})_{i_1 i_2 i_3} = \sum_{r=1}^{R} (\mathbf{E}_r)_{i_1 i_2} (\mathbf{c}_r)_{i_3} . \tag{4.6}$$

Note that

$$(\mathbf{E}_r)_{i_1 i_2} = \sum_{r'=1}^{L_r} \left(\mathbf{D}_r^{(1)}\right)_{i_1 r'} \left(\mathbf{D}_r^{(2)}\right)_{i_2 r'} \tag{4.7}$$

$$= \sum_{r'=1}^{L_r} \left(\mathbf{d}_{r,r'}^{(1)}\right)_{i_1} \left(\mathbf{d}_{r,r'}^{(2)}\right)_{i_2} \tag{4.8}$$

Then substituting (4.8) into equation (4.6):

$$(\mathcal{T})_{i_1 i_2 i_3} = \sum_{r=1}^{R} \left[\sum_{r'=1}^{L_r} \left(\mathbf{d}_{r,r'}^{(1)}\right)_{i_1} \left(\mathbf{d}_{r,r'}^{(2)}\right)_{i_2} \right] (\mathbf{c}_r)_{i_3} . \tag{4.9}$$

The uniqueness of this decomposition is discussed in the following theorem, proved in [52].

Theorem 1. *Let* $\mathbf{D}^{(1)} = \left[\mathbf{D}_1^{(1)}, \mathbf{D}_2^{(1)}, \ldots, \mathbf{D}_R^{(1)}\right]$, $\mathbf{D}^{(2)} = \left[\mathbf{D}_1^{(2)}, \mathbf{D}_2^{(2)}, \ldots, \mathbf{D}_R^{(2)}\right]$ *and* $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_R]$. *The decomposition in block terms of rank-*$(L_r, L_r, 1)$ *of* $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ *with* $I_1, I_2 \geq \sum_{r=1}^{R} L_r$ *will be essentially unique, i.e., unique up to trivial indeterminacies, if*

- $\mathbf{D}^{(1)} \in \mathbb{R}^{I_1 \times \sum_{r=1}^{R} L_r}$ *and* $\mathbf{D}^{(2)} \in \mathbb{R}^{I_2 \times \sum_{r=1}^{R} L_r}$ *have full column rank*
- $\mathbf{C} \in \mathbb{R}^{I_3 \times R}$ *does not have proportional columns,*

This is the power of BTD for underdetermined BSS. While in matrix methods for BSS we need to assume that the unmixing matrix (related to $\mathbf{C}$) is full-column rank, BTD relaxes this assumption. It only needs that the columns of $\mathbf{C}$ are not proportional.

4.2 Blind Source Separation via BTD

This tensorial decomposition can be used to address the Blind Source Separation (BSS) problem assuming a certain source model. Consider the instantaneous linear mixture model (2.31). Let $\mathbf{Y}_{k:} = [y_k(0), y_k(1), \ldots, y_k(N-1)]$ denote the kth row of $\mathbf{Y}$ and $\mathbf{S}_{r:}$ the rth row of $\mathbf{S}$. Let $\mathcal{H}(\cdot)$ be an operator that maps an input vector $\mathbf{v} \in \mathbb{R}^{N \times 1}$ to a Hankel matrix $\mathcal{H}(\mathbf{v}) \in \mathbb{R}^{N \times N}$. Then, consider the Hankel matrices $\mathcal{H}(\mathbf{Y}_{k:})$ obtained from the rows of $\mathbf{Y}$ and stacked in a third-order tensor $\mathcal{Y} \in \mathbb{R}^{N \times N \times K}$ so that the frontal slices are given by:

$$(\mathcal{Y})_{::k} = \mathcal{H}(\mathbf{Y}_{k:}) = \begin{bmatrix} y_k(0) & y_k(1) & \ldots & y_k(N-1) \\ y_k(1) & \ldots & y_k(N-1) & 0 \\ \vdots & \iddots & 0 & \vdots \\ y_k(N-1) & 0 & \ldots & 0 \end{bmatrix}, \tag{4.10}$$

for $k = 1, \ldots, K$. Note that for Hankel matrices, the element (i, j) equals the element $(i-1, j+1)$ for $i, j = 1, \ldots, N$. From equation (2.31), we know that

$$y_k(n) = \sum_{r=1}^{R} m_k(r) s_r(n) \tag{4.11}$$

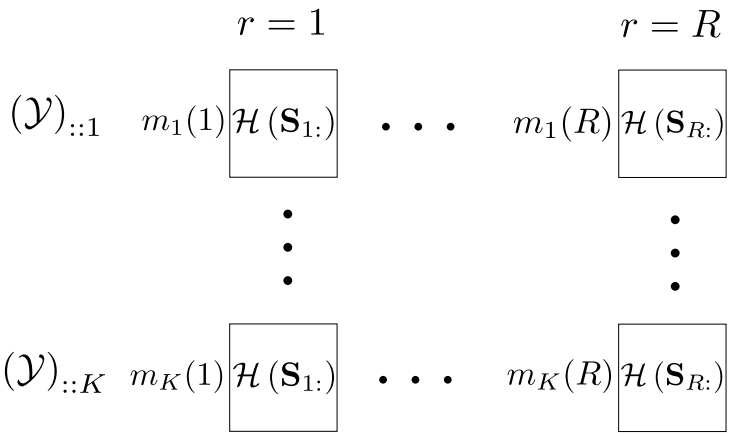

Figure 4.2: Hankel matrices $\mathcal{H}(\mathbf{S}_{r:})$ being scaled by $m_k(r)$, revealing the outer product structure in each r.

in which $m_k(r)$ and $s_r(n)$ are elements of $\mathbf{M}$ and $\mathbf{S}$ respectively. Substituting (4.11) into (4.10):

$$\begin{aligned}
(\mathcal{Y})_{::k} &= \begin{bmatrix} \sum_{r=1}^{R} m_k(r)s_r(0) & \sum_{r=1}^{R} m_k(r)s_r(1) & \dots & \sum_{r=1}^{R} m_k(r)s_r(N-1) \\ \sum_{r=1}^{R} m_k(r)s_r(1) & \dots & \sum_{r=1}^{R} m_k(r)s_r(N-1) & 0 \\ \vdots & \iddots & 0 & \vdots \\ \sum_{r=1}^{R} m_k(r)s_r(N-1) & 0 & \dots & 0 \end{bmatrix} \\
&= \sum_{r=1}^{R} \left\{ m_k(r) \begin{bmatrix} s_r(0) & s_r(1) & \dots & s_r(N-1) \\ s_r(1) & \dots & s_r(N-1) & 0 \\ \vdots & \iddots & 0 & \vdots \\ s_r(N-1) & 0 & \dots & 0 \end{bmatrix} \right\} \\
&= \sum_{r=1}^{R} m_k(r)\mathcal{H}\left(\mathbf{S}_{r:}\right). \qquad (4.12)
\end{aligned}$$

In equation (4.12), for a fixed r and $k = 1, \dots, K$, the matrix $\mathcal{H}\left(\mathbf{S}_{r:}\right)$ is scaled by $m_k(r)$ as depicted in Figure 4.2. This implies that for a fixed r, the outer product between $\mathcal{H}\left(\mathbf{S}_{r:}\right)$ and $\mathbf{M}_{:r}$, the rth column of $\mathbf{M}$, is being performed. Now considering every r, we have:

$$\mathcal{Y} = \sum_{r=1}^{R} \mathcal{H}\left(\mathbf{S}_{r:}\right) \circ \mathbf{M}_{:r}, \qquad (4.13)$$

which is equivalent to the BTD model (4.1). The mixing matrix can be estimated by grouping the vectors $\mathbf{M}_{:r}$ in a matrix. Let $\hat{\mathbf{M}}_{:r}$ be the estimated mixing vectors from (4.13), then the estimated mixing matrix will be $\hat{\mathbf{M}} = \left[\hat{\mathbf{M}}_{:1}, \hat{\mathbf{M}}_{:2}, \dots, \hat{\mathbf{M}}_{:R}\right]$. If $\mathbf{M}$ has full column rank with $K \geq R$, then the estimated source matrix $\hat{\mathbf{S}}$ can be obtained from (2.31) by inverting $\hat{\mathbf{M}}$ as:

$$\hat{\mathbf{S}} = \hat{\mathbf{M}}^{\dagger}\mathbf{Y}, \qquad (4.14)$$

where $(\cdot)^{\dagger}$ is the Moore-Penrose pseudoinverse. However, if $\mathbf{M}$ is rank-deficient or $K \leq R$ (sub-determined problem), then inverting $\hat{\mathbf{M}}$ is not possible. Nevertheless, consider the Hankel matrix associated with the rth row of $\mathbf{S}$ obtained from (4.13):

$$\mathcal{H}\left(\mathbf{S}_{r:}\right) = \begin{bmatrix} s_r(0) & s_r(1) & \dots & s_r(N-1) \\ s_r(1) & \dots & s_r(N-1) & 0 \\ \vdots & \iddots & 0 & \vdots \\ s_r(N-1) & 0 & \dots & 0 \end{bmatrix}.$$

Observe that the element $s_r(n)$, $n = 0, \dots, (N-1)$ can be obtained by averaging the $(n+1)$th anti-diagonal of $\mathcal{H}\left(\mathbf{S}_{r:}\right)$. Let $\mathcal{H}\left(\hat{\mathbf{S}}_{r:}\right)$ be the estimation of $\mathcal{H}\left(\mathbf{S}_{r:}\right)$ obtained from (4.13). Thus, the element

$\hat{s}_r(n)$ of the estimated source matrix $\hat{\mathbf{S}}$ can be obtained by averaging the $(n+1)$th anti-diagonal of the $\mathcal{H}\left(\hat{\mathbf{S}}_{r:}\right)$ for $r = 1, \ldots, R$ and $n = 0, \ldots, (N-1)$.

The BSS will be achieved if the decomposition (4.13) is essentially unique. To assure this uniqueness, a constraint upon the sources obeying the Theorem 1 needs to be imposed. Suppose that they can be expressed as linear combinations of exponentials:

$$s_r(n) = \sum_{l_r=1}^{L_r} c_{l_r,r} z_{l_r,r}^n, \quad 0 \leq n \leq (N-1),\ 1 \leq r \leq R. \tag{4.15}$$

where L_r is the number of exponential terms, $c_{l_r,r}$ is the linear coefficient of the l_rth term in elements of the rth row and $z_{l_r,r}^n$ is the base of the exponential associated to the l_rth term of $s_r(n)$ (also called pole in [53]). Then $\mathbf{S}$ can be written as

$$\mathbf{S} = \begin{bmatrix} \sum_{l_r=1}^{L_r} c_{l_r,1} z_{l_r,1}^0 & \sum_{l_r=1}^{L_r} c_{l_r,1} z_{l_r,1}^1 & \cdots & \sum_{l_r=1}^{L_r} c_{l_r,1} z_{l_r,1}^{N-1} \\ \sum_{l_r=1}^{L_r} c_{l_r,2} z_{l_r,2}^0 & \sum_{l_r=1}^{L_r} c_{l_r,2} z_{l_r,2}^1 & \cdots & \sum_{l_r=1}^{L_r} c_{l_r,2} z_{l_r,2}^{N-1} \\ \vdots & \ddots & & \vdots \\ \sum_{l_r=1}^{L_r} c_{l_r,R} z_{l_r,R}^0 & \sum_{l_r=1}^{L_r} c_{l_r,R} z_{l_r,R}^1 & \cdots & \sum_{l_r=1}^{L_r} c_{l_r,R} z_{l_r,R}^{N-1} \end{bmatrix}$$

and the Hankel matrix $\mathcal{H}\left(\mathbf{S}_{r:}\right) \in \mathbb{R}^{N \times N}$of the rth row will be:

$$\mathcal{H}\left(\mathbf{S}_{r:}\right) = \begin{bmatrix} \sum_{l_r=1}^{L_r} c_{l_r,r} z_{l_r,r}^0 & \sum_{l_r=1}^{L_r} c_{l_r,r} z_{l_r,r}^1 & \cdots & \sum_{l_r=1}^{L_r} c_{l_r,r} z_{l_r,r}^{N-1} \\ \sum_{l_r=1}^{L_r} c_{l_r,r} z_{l_r,r}^1 & \cdots & \sum_{l_r=1}^{L_r} c_{l_r,r} z_{l_r,r}^{N-1} & 0 \\ \vdots & \iddots & 0 & \vdots \\ \sum_{l_r=1}^{L_r} c_{l_r,r} z_{l_r,r}^{N-1} & 0 & \cdots & 0 \end{bmatrix}. \tag{4.16}$$

This Hankel matrix admits the Vandermonde decomposition [54]:

$$\mathcal{H}\left(\mathbf{S}_{r:}\right) = \mathbf{V}_r \mathrm{diag}(c_{1,r}, \ldots, c_{L_r,r}) \mathbf{V}_r^T \tag{4.17}$$

in which the Vandermonde matrix $\mathbf{V}_r \in \mathbb{R}^{N \times L_r}$ is defined by

$$\mathbf{V}_r = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z_{1,r} & z_{2,r} & \cdots & z_{L_r,r} \\ \vdots & \vdots & & \vdots \\ z_{1,r}^{N-1} & z_{2,r}^{N-1} & \ldots & z_{L_r,r}^{N-1} \end{bmatrix}, \tag{4.18}$$

as discussed in [53]. Since a Vandermonde matrix generated by distinct poles has full column rank L_r, then the rth term in (4.13) is rank-$(L_r, L_r, 1)$, $1 \leq r \leq R$. If the matrix $\mathbf{M} \in \mathbb{R}^{K \times R}$ does not have proportional columns and all the poles $z_{l_r,r}$ are distinct for $1 \leq l_r \leq L_r$, $1 \leq r \leq R$, then the matrices

$$\mathbf{D}^{(1)'} = [\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_R]\, \mathrm{diag}\left(c_{1,1}, c_{2,1}, \ldots, c_{L_R,R}\right) \in \mathbb{R}^{N \times \sum_{r=1}^{R} L_r}$$

and

$$\mathbf{D}^{(2)'} = [\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_R] \in \mathbb{R}^{N \times \sum_{r=1}^{R} L_r},$$

have full column rank like matrices $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$ in Theorem 1, fulfilling the conditions of the essential uniqueness of the BTD. Thus the BSS can be performed by either inverting $\hat{\mathbf{M}}$ or averaging the antidiagonals of $\mathcal{H}\left(\mathbf{S}_{r:}\right)$ for $r = 1, \ldots, R$. To tackle the scaling and permutation indeterminacies, refer to the Greedy Method explained in the Appendix A.

The BSS-BTD method can be summarized as the following:

Step 1 Build the observation matrix $\mathbf{Y}$.

Step 2 For each line of $\mathbf{Y}$, create a Hankel matrix.

Step 3 Create a tensor $\mathcal{Y}$ by stacking up the Hankel matrices.

Step 4 Perform the BTD on the tensor $\mathcal{Y}$.

Step 5 Estimate the mixing matrix $\hat{\mathbf{M}}$ from the BTD. If $\mathbf{M}$ is full-rank, estimate the source matrix as in equation (4.14), otherwise estimate the source matrix directly from the Hankel matrices.

4.3 Calculating the BTD

Consider the BTD model (4.1). Let

$$\mathbf{F}_{\mathbf{D}}^{(1)} = \left[\mathbf{d}_{1,1}^{(1)}, \ldots, \mathbf{d}_{1,L_1}^{(1)}, \ldots, \mathbf{d}_{R,1}^{(1)}, \ldots, \mathbf{d}_{R,L_R}^{(1)}\right]_{I_1 \times \sum_{r=1}^{R} L_r} \tag{4.19}$$

$$\mathbf{F}_{\mathbf{D}}^{(2)} = \left[\mathbf{d}_{1,1}^{(2)}, \ldots, \mathbf{d}_{1,L_1}^{(2)}, \ldots, \mathbf{d}_{R,1}^{(2)}, \ldots, \mathbf{d}_{R,L_R}^{(2)}\right]_{I_2 \times \sum_{r=1}^{R} L_r} \tag{4.20}$$

and

$$\mathbf{F}_{\mathbf{C}} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_R]_{I_3 \times R}. \tag{4.21}$$

Note that matrices $\mathbf{F}_{\mathbf{D}}^{(1)}$ and $\mathbf{F}_{\mathbf{D}}^{(2)}$ are built by arranging the $\sum_{r=1}^{R} L_r$ vector components from the rank-L_r matrices $\mathbf{D}_r^{(1)}$ and $\mathbf{D}_r^{(2)}$ for $r = 1, \ldots, R$. This model can be interpreted as a *structured CPD* with factor matrices

$$\mathbf{A}^{(1)} = \mathbf{F}_{\mathbf{D}}^{(1)} \tag{4.22}$$

$$\mathbf{A}^{(2)} = \mathbf{F}_{\mathbf{D}}^{(2)} \tag{4.23}$$

$$\mathbf{A}^{(3)} = \mathbf{F}_{\mathbf{C}}\mathbf{E}, \tag{4.24}$$

in which $\mathbf{E}$ is an $R \times \sum_{r=1}^{R} L_r$ block-diagonal matrix $\text{diag}\left(\mathbf{1}_{1\times L_1}, \ldots, \mathbf{1}_{1\times L_R}\right)$ in which the rth block on the diagonal is the row vector $\mathbf{1}_{1\times L_r}$. It will repeat the rth column of $\mathbf{F}_{\mathbf{C}}$ L_r times along the columns of $\mathbf{A}^{(3)}$ for $r = 1, \ldots, R$.

The unfolded tensor $\mathbf{T}_{(n)}$ can be now defined by means of the CPD factor matrices

$$\mathbf{T}_{(n)} = \mathbf{A}^{(n)}\mathbf{V}^{\{n\}^T},\ 1 \leq n \leq N \tag{4.25}$$

in which $\mathbf{V}^{(n)}$ is the Khatri-Rao product between the factor matrices, not including $\mathbf{A}^{(n)}$, as defined in (3.13). Let $\hat{\mathcal{T}}$ be an estimate of $\mathcal{T}$. Then the residual tensor $\mathcal{F}_{BTD}$ is defined by

$$\mathcal{F}_{BTD} = \hat{\mathcal{T}} - \mathcal{T}. \tag{4.26}$$

The mode-n unfolding of $\mathcal{F}_{BTD}$ is

$$\mathbf{F}_{(n)} = \hat{\mathbf{A}}^{(n)}\hat{\mathbf{V}}^{\{n\}^T} - \mathbf{T}_{(n)}, \tag{4.27}$$

where $\hat{\mathbf{T}}_{(n)} = \hat{\mathbf{A}}^{(n)}\hat{\mathbf{V}}^{\{n\}^T}$ and $\mathbf{T}_{(n)}$ are the mode-n unfolding of $\hat{\mathcal{T}}$ and $\mathcal{T}$ respectively. Based on the residual tensor definition, we formulate a non-linear problem

$$\min_{\mathbf{A}^{(1)},\mathbf{A}^{(2)},\mathbf{A}^{(3)}} f_{BTD}, \tag{4.28}$$

where $f_{BTD} = \frac{1}{2}||\mathcal{F}_{BTD}||_F^2$. In the following subsections, two families of optimization methods will be discussed in order to solve the problem (4.28).

General unconstrained optimization and Alternating Least-Squares

Gradient-based unconstrained optimization methods such as the nonlinear conjugate gradient and quasi-Newton can be used for minimizing the objective function f_{BTD}. The quasi-Newton and Alternating Least-Squares (ALS) approaches will be discussed in this subsection.

Let $\overset{c}{\cdot}$ denote the concatenation of its argument with its complex conjugate, i.e. $\overset{c}{\mathbf{c}} = (\mathbf{c}, \overline{\mathbf{c}})$ and let $\mathbf{z}$ be the vector of unknowns

$$\mathbf{z} = \left[\text{vec}\left(\mathbf{A}^{(1)}\right)^T, \text{vec}\left(\mathbf{A}^{(2)}\right)^T, \text{vec}\left(\mathbf{A}^{(3)}\right)^T\right]^T.$$

Since this vector will be updated over the iterations, it can be indexed as $\mathbf{z}_k$. Consider the second-order model $m_k^{f_{BTD}}(\overset{c}{\mathbf{p}})$ of the objective function $f_{BTD}(\overset{c}{\mathbf{z}}_k)$ with search direction $\overset{c}{\mathbf{p}}$ at the kth iteration

$$m_k^{f_{BTD}}(\overset{c}{\mathbf{p}}) = f_{BTD}(\overset{c}{\mathbf{z}}_k) + \overset{c}{\mathbf{p}}^T \frac{\partial f_{BTD}(\overset{c}{\mathbf{z}}_k)}{\partial \overset{c}{\mathbf{z}}} + \frac{1}{2}\overset{c}{\mathbf{p}}^H B_k \overset{c}{\mathbf{p}}, \tag{4.29}$$

in which B_k is a Hermitian positive definite matrix that is updated at every iteration. The minimizer of the convex quadratic model (4.29) is obtained by setting the model's conjugate complex gradient equal to zero [55], resulting in

$$\overset{c}{\mathbf{p}}_k^* = -B_k^{-1}\overline{\frac{\partial f_{BTD}(\overset{c}{\mathbf{z}}_k)}{\partial \overset{c}{\mathbf{z}}}} \tag{4.30}$$

$$(\mathbf{p}_k^*, \overline{\mathbf{p}}_k^*) = \left(-B_k^{-1}\frac{\partial f_{BTD}(\mathbf{z}_k)}{\partial \mathbf{z}}, -B_k^{-1}\overline{\frac{\partial f_{BTD}(\mathbf{z}_k)}{\partial \mathbf{z}}}\right). \tag{4.31}$$

Deriving an expression for $\frac{\partial f_{BTD}(\mathbf{z}_k)}{\partial \mathbf{z}}$ enables us to compute the the search direction (4.30). Let $\mathbf{W}^\sigma = *_{n=1,n\notin\sigma}^{N} \mathbf{A}^{(n)^H}\mathbf{A}^{(n)}$ where $*$ is the elementwise Hadamard product. Then

$$\frac{\partial f_{BTD}}{\partial \mathbf{z}} = \left(\text{vec}\left(\frac{\partial f_{BTD}}{\partial \mathbf{A}^{(1)}}\right)^T, \text{vec}\left(\frac{\partial f_{BTD}}{\partial \mathbf{A}^{(2)}}\right)^T, \text{vec}\left(\frac{\partial f_{BTD}}{\partial \mathbf{A}^{(3)}}\right)^T\right)^T, \tag{4.32}$$

where

$$2\frac{\partial f_{BTD}}{\partial \mathbf{A}^{(1)}} = \overline{\mathbf{F}_\mathbf{D}^{(1)}}\mathbf{W}^{\{1\}} - \overline{\mathbf{T}}_{(1)}\mathbf{V}^{\{1\}}, \tag{4.33}$$

$$2\frac{\partial f_{BTD}}{\partial \mathbf{A}^{(2)}} = \overline{\mathbf{F}_\mathbf{D}^{(2)}}\mathbf{W}^{\{2\}} - \overline{\mathbf{T}}_{(2)}\mathbf{V}^{\{2\}}, \tag{4.34}$$

$$2\frac{\partial f_{BTD}}{\partial \mathbf{A}^{(3)}} = \overline{\mathbf{F}_\mathbf{C}}\mathbf{E}\mathbf{W}^{\{3\}}\mathbf{E}^T - \overline{\mathbf{T}}_{(3)}\mathbf{V}^{\{3\}}\mathbf{E}^T. \tag{4.35}$$

Refer to [56] for the proof of these relationships. The ALS solution to this problem consists of alternately setting to zero the partial gradients (4.33), (4.34) and (4.35), from which each factor matrix $\{\mathbf{F}_\mathbf{D}^{(1)}, \mathbf{F}_\mathbf{D}^{(2)}, \mathbf{F}_\mathbf{C}\}$ is computed. The factor matrix $\mathbf{F}_\mathbf{D}^{(n)}$ for $n = 1, 2$ is obtained in the following way:

$$\begin{aligned}
\overline{\mathbf{F}_\mathbf{D}^{(n)}}\mathbf{W}^{\{n\}} - \overline{\mathbf{T}}_{(n)}\mathbf{V}^{\{n\}} &= 0 \\
\overline{\mathbf{F}_\mathbf{D}^{(n)}}\mathbf{W}^{\{n\}} &= \overline{\mathbf{T}}_{(n)}\mathbf{V}^{\{n\}} \\
\overline{\mathbf{F}_\mathbf{D}^{(n)}} &= \overline{\mathbf{T}}_{(n)}\mathbf{V}^{\{n\}}\left(\mathbf{W}^{\{n\}}\right)^\dagger
\end{aligned} \tag{4.36}$$

and for $\mathbf{F}_\mathbf{C}$:

$$\begin{aligned}
\overline{\mathbf{F}_\mathbf{C}}\mathbf{E}\mathbf{W}^{\{3\}}\mathbf{E}^T - \overline{\mathbf{T}}_{(3)}\mathbf{V}^{\{3\}}\mathbf{E}^T &= 0 \\
\overline{\mathbf{F}_\mathbf{C}}\mathbf{E}\mathbf{W}^{\{3\}}\mathbf{E}^T &= \overline{\mathbf{T}}_{(3)}\mathbf{V}^{\{3\}}\mathbf{E}^T \\
\overline{\mathbf{F}_\mathbf{C}} &= \overline{\mathbf{T}}_{(3)}\mathbf{V}^{\{3\}}\mathbf{E}^T\left(\mathbf{E}\mathbf{W}^{\{3\}}\mathbf{E}^T\right)^\dagger.
\end{aligned} \tag{4.37}$$

Although the ALS is simple and effective, it may not converge to a stationary point and is sensitive to *swamps*, where the convergence is very slow for many iterations. Severeal approaches such as Tikhonov regularization and exact line search can be used to mitigate the swamp effects.

Nonlinear least-squares methods

Methods like Gauss-Newton or Levenberg-Marquadt can also be used to solve (4.28). In these methods, the residual tensor $\mathcal{F}_{BTD}$ is approximated by the linear model

$$m_k^{\mathcal{F}_{BTD}}(\overset{c}{\mathbf{p}}) = vec\left(\mathcal{F}_{BTD}(\overset{c}{\mathbf{z}}_k)\right) + \frac{\partial vec\left(\mathcal{F}_{BTD}(\overset{c}{\mathbf{z}}_k)\right)}{\partial \overset{c}{\mathbf{z}}^T}\overset{c}{\mathbf{p}} \tag{4.38}$$

and used in the modified quadratic model of the objective function

$$m_k^{f_{BTD}}(\overset{c}{\mathbf{p}}) = \frac{1}{2}||m_k^{\mathcal{F}_{BTD}}(\overset{c}{\mathbf{p}})||^2 + \frac{\lambda_k}{2}||\overset{c}{\mathbf{p}}||^2, \tag{4.39}$$

where λ_k is the Levenberg-Marquadt regularization parameter. In the Gauss-Newton method, $\lambda_k = 0 \quad \forall k$. Subsequently, to minimize the objective function $m_k^{f_{BTD}}(\overset{c}{\mathbf{p}})$, a preconditioned conjugate gradient (PCG) algorithm is used with an ALS-based algorithm to estimate the factor matrices. The heavy calculations are performed in the PCG step, resulting in faster ALS updates. This refined ALS adds robustness to this nonlinear least-squares method while increasing its computational complexity, even though it is still competitive with ALS for simpler problems.

Parameter	case A	case B
Frequency, f_0	6 Hz	8 Hz
Frequency variation, Δf	0.2 Hz	0.3 Hz
Frequency variation frequency, f_f	0.1 Hz	0.23 Hz
Harmonics, M	5	3
Amplitude, a_l	150 μV	60 μV
Amplitude variation, Δa_l	50 μV	18 μV
Amplitude variation frequency, ω_a	0.08 Hz	0.5 Hz

Table 4.1: Parameters for the two AF patterns, as in the paper [4].

4.4 Application to Atrial Signal Extraction in AF ECG

The BSS-BTD method is a new deterministic BSS solution based on a linear combination of exponentials model for the sources, which is not an unrealistic assumption for biomedical signals. Recently, Hunyadi et al [47] successfully employed the BTD-BSS method in the analysis of $channel \times time \times frequency$ EEG tensors during epileptic seizures. Their CPD decomposition is defined by a sum of rank 1 terms with one spatial, temporal and spectral signature. This rank constraint does not allow their approach to model more complex signals such those which present a certain degree of nonstationarity. Thus they have proposed the BTD in rank-$(L_r, L_r, 1)$ terms as a solution to model more involved signals such as EEG signals during an epilepsy episode.

In this section, we seek to extract the atrial activity of ECG during AF episodes by means of BSS-BTD and analyze its performance. To this end, we will adopt the procedure suggested in [6, 8], used to assess the performance of an atrial source extraction algorithm. This procedure consists of generating a synthetic AF signal and superimposing it to a sinus rhythm ECG with the P waves properly removed, and then compare the AF signal estimated by the algorithm and the synthesized AF signal.

The most established AF model in the literature was proposed by Stridh and Sornmo [4], and consists of a non-stationary signal that mimics the sawtooth-similar shape of the atrial flutter, an arrhythmia closely connected to the AF. The model consists of a sinusoid and $(M-1)$ harmonics both amplitude and frequency modulated. Thus the f-wave is modeled by

$$y(n) = -\sum_{i=1}^{M} a_i(n) \sin(i\theta(n)) \tag{4.40}$$

where $a_i(n)$ is the modulated amplitude, given by

$$a_i(n) = \frac{2}{i\pi}\left(a + \Delta a \sin 2\pi \frac{f_a}{F_s} n\right) \tag{4.41}$$

in which a is the sawtooth amplitude, Δa is the modulation peak amplitude and f_a is the amplitude modulation frequency. The phase $\theta(n)$ defined as

$$\theta(n) = 2\pi \frac{f_0}{F_s} n + \left(\frac{\Delta f}{f_f}\right) \sin 2\pi \frac{f_f}{Fs} n \tag{4.42}$$

varying sinusoidally around f_0 with maximum frequency deviation Δf and modulation frequency f_f. The two AF patterns described by Stridh are considered in this paper: large amplitude and long cycle length (case A) and low amplitude with shorter cycle length (case B), as shown in Figure 4.3. The parameters model for each case are summarized in Table 4.1.

A 10 s sinus rhythm 8-leads ECG (I, II, V1-V6) from the PTB Diagnostic ECG Database [2] was acquired at a sampling rate of 1 kHz. A forward-backward bandpass type II Chebyshev IIR filter with cut-off frequencies at 0.5 Hz and 30 Hz was used to remove the baseline wander and powerline interference. Next, the beginning and the end of the P waves associated with the first two heart beats were manually annotated and interpolated by means of a spline interpolation. The resulting signal was stored in a matrix $\mathbf{S} \in \mathbb{R}^{8\times 2000}$.

Subsequently, a window consisting of the first 1200 samples from all leads was extracted from $\mathbf{S}$ and downsampled at a rate of 10, resulting in a data matrix $\mathbf{S}_d \in \mathbb{R}^{8\times 120}$. Afterwards, a synthetic AF signal

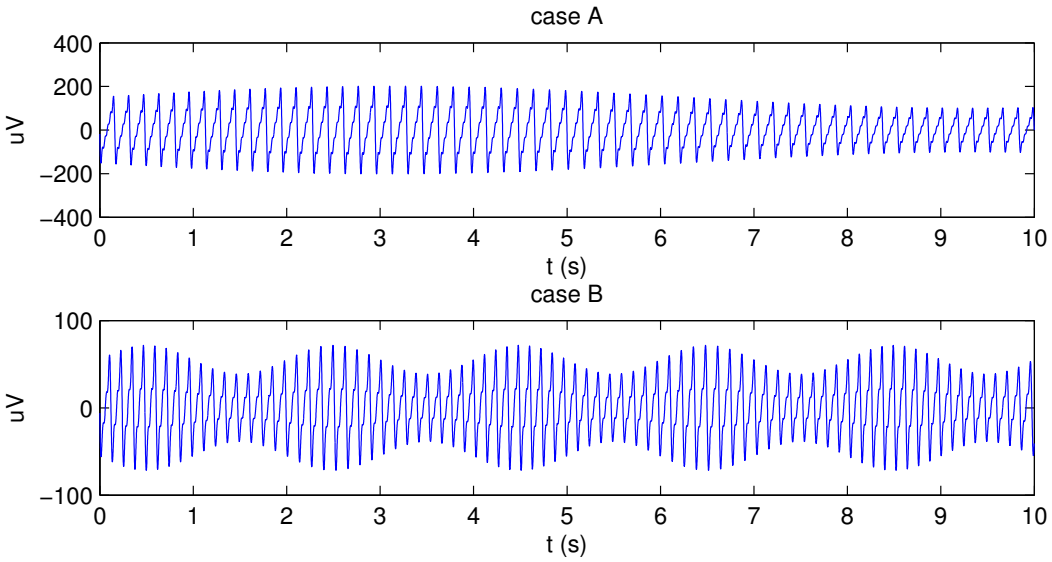

Figure 4.3: Waveforms for the synthetic AF model.

with length of 1.2 s was generated and acquired at a sampling rate of 100 Hz and stored in a vector $\mathbf{y} \in \mathbb{R}^{120\times 1}$. The resulting data matrix is given by

$$\mathbf{Y}(\gamma, \sigma) = \mathbf{S}_d + \gamma \mathbf{a}\mathbf{y}^T + \sigma \mathbf{B}$$

where $\mathbf{Y}(\gamma, \sigma) \in \mathbb{R}^{8\times 120}$ is the observation ECG data matrix, $\mathbf{a} \in \mathbb{R}^{8\times 1}$ is an unit-length random vector of spatial scales of the AF signal, γ is the power factor of the AF components, $\mathbf{B} \in \mathbb{R}^{8\times 120}$ is a normalized white Gaussian noise matrix and σ is the power factor of the noise matrix.

Subsequently, Hankel matrices were obtained from the 8 rows of $\mathbf{Y}(\gamma, \sigma)$ and stacked along the third dimension, as explained in Section 4.2, resulting in a tensor $\mathcal{Y} \in \mathbb{R}^{120\times 120\times 8}$. In order to calculate the BTD decomposition, its parameters need to be selected, i.e. number of block components and their multilinear rank. Since we know the desired AF source signal, we can calculate the rank of its associated Hankel matrix and use it as L_r. Henceforth, let us denote $\mathbf{l} = [L_1, L_2, \ldots, L_R]$ a vector whose components corresponds to the multilinear rank of the rth BTD component. For instance, let $\mathbf{l} = [2, 3, 4, 5]$, then the first BTD component will be rank-$(2, 2, 1)$, the second will be rank-$(3, 3, 1)$, and so on. The L_r value changes with the AF signal parameters as show in Figure 4.4. For each set of parameters proposed by Stridh and Sornmo (Table 5.1), denoted as case A and case B, the rank of their corresponding Hankel matrix is shown in Figure 4.4.

The power of a signal represented in a matrix $\mathbf{M}$ with N elements is defined as $P_{\mathbf{M}} = \frac{1}{N}||\mathbf{M}||_F$. Let $\mathbf{A} = \gamma \mathbf{a}\mathbf{y}^T$ denote the atrial activity matrix, $\mathbf{V} = \mathbf{S}_d$ the ventricular activity matrix and $P_{\mathbf{N}} = \sigma \mathbf{B}$. We define the Atrial-Ventricular Ratio (AVR) as

$$AVR = \frac{P_{\mathbf{A}}}{P_{\mathbf{V}}}, \tag{4.43}$$

and the atrial Signal-to-Noise Ratio (SNR) as

$$SNR = \frac{P_{\mathbf{A}}}{P_{\mathbf{N}}}. \tag{4.44}$$

If we normalize $\mathbf{a}\mathbf{y}^T$ and $\mathbf{B}$ by their power, then $P_{\mathbf{A}} = \gamma$ and $P_{\mathbf{N}} = \sigma$. In the following experiments, we will set some values for AVR and SNR then find the parameters γ and σ for calculating $\mathbf{Y}(\gamma, \sigma)$. Let $\mathbf{M} \in \mathbb{R}^{P\times Q}$. Its root mean squared (RMS) value is $RMS(\mathbf{M}) = \sqrt{\frac{1}{PQ}\sum_{p=1}^{P}\sum_{q=1}^{Q} m_{pq}^2}$.

First experiment

The first experiment consists of extracting the case A synthesized AF source from a noiseless ECG with $AVR = -10\,dB$ (Figure 4.5). Since the AF source is known, its Hankel matrix was calculated and

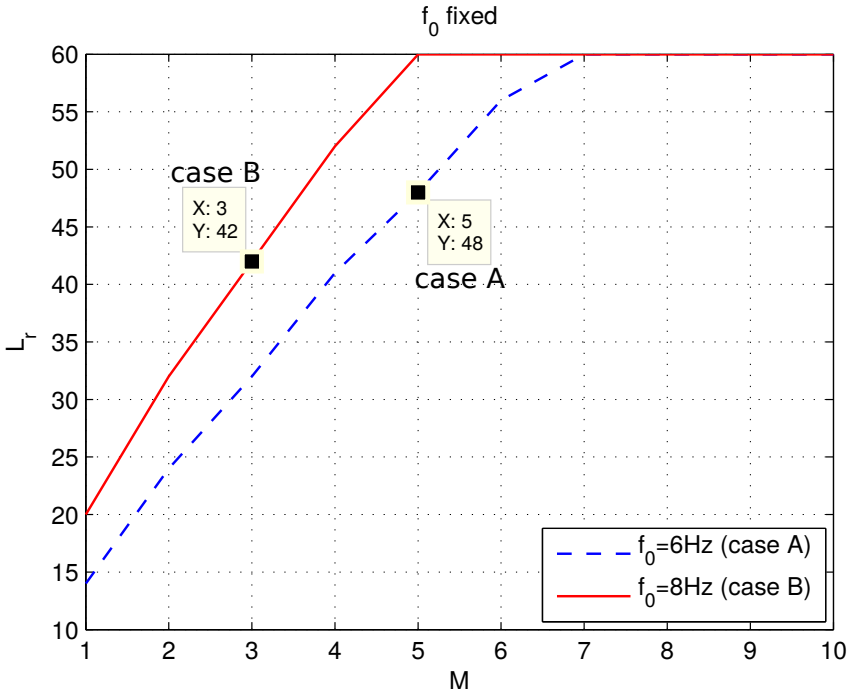

Figure 4.4: Variation of L_r with the number of harmonics M for a certain fundamental frequency f_0 fixed.

the rank of this matrix was used as the multilinear rank of each tensor component. According to the Figure 4.4, $L_r = 48$ for the case A. In this experiment, the AF source was extracted by BSS-BTD with $\mathbf{l} = [48, 48, 48]$ (Figure 4.6) and some benchmark methods, such as PCA (Figure 4.7) and RobustICA-f (Figure 4.8). Next, a Monte Carlo simulation consisting of 20 runs was performed by varying the SNR (Figure 4.9). The noise matrix $\mathbf{N}$ and the vector of spatial scales $\mathbf{a}$ were generated at each experiment. Finally the root mean squared value of the error between the estimated and generated AF signals was calculated.

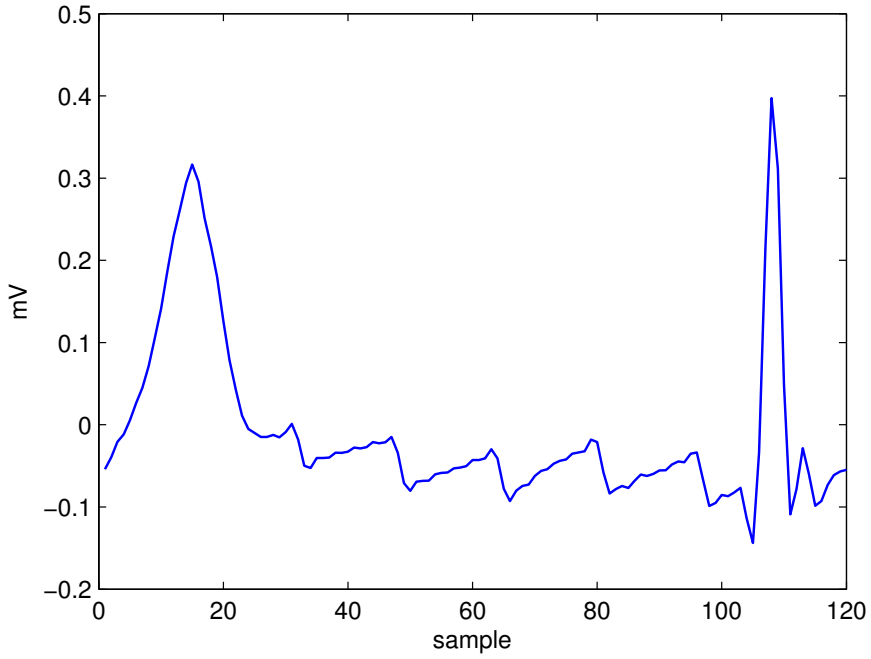

Figure 4.5: Generated ECG signal for the case A AF source with $AVR = -10\,dB$ (lead II).

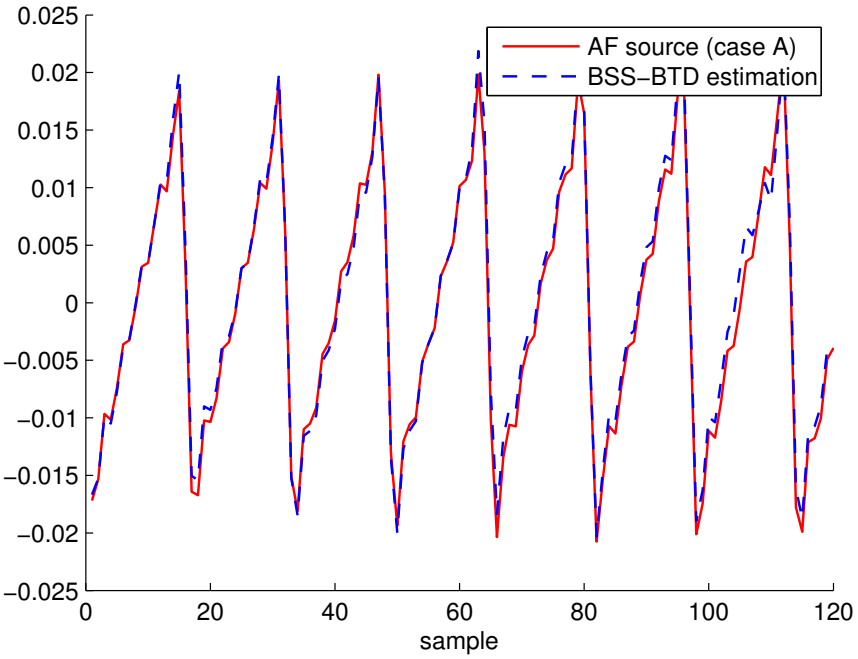

Figure 4.6: Case A AF source extracted by BSS-BTD.

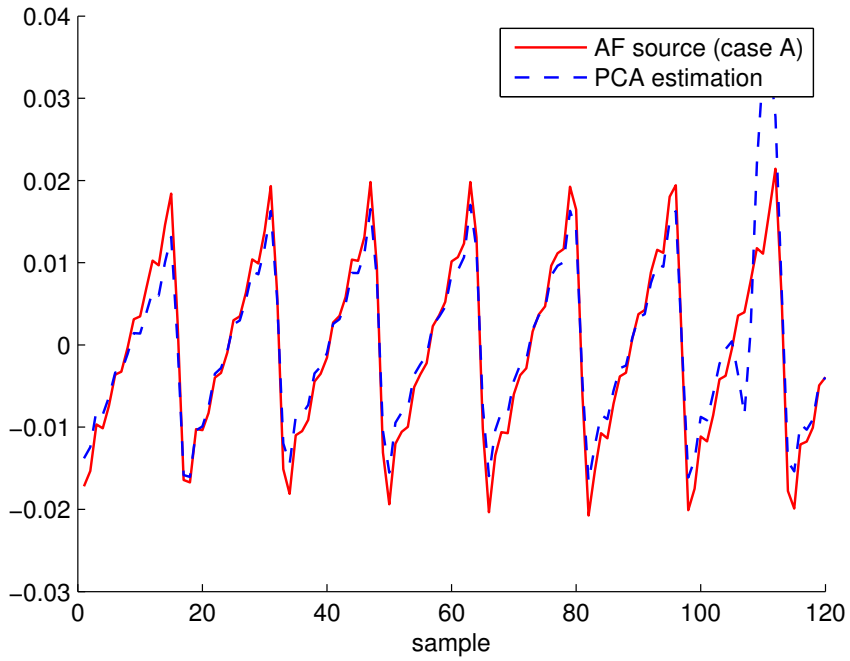

Figure 4.7: Case A AF source extracted by PCA.

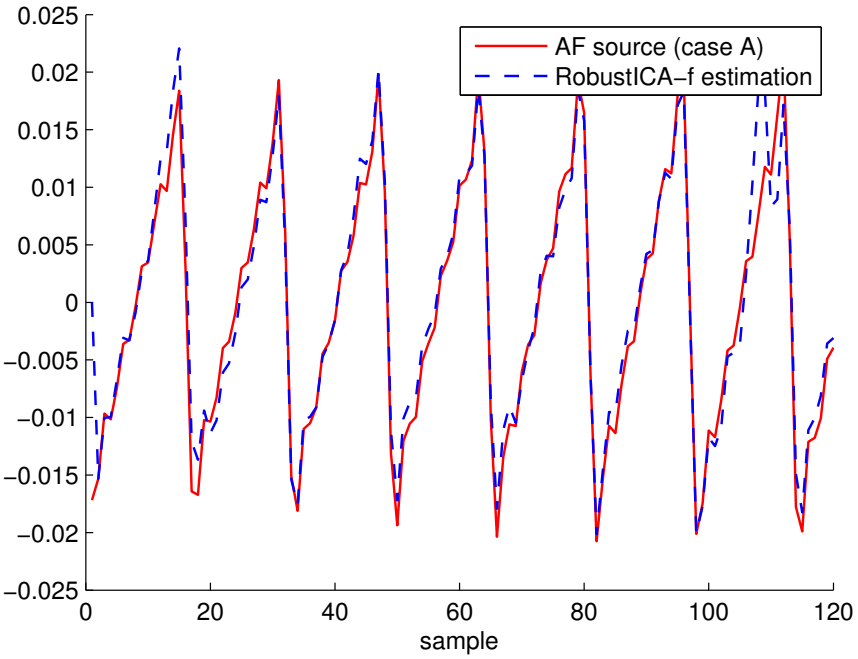

Figure 4.8: Case A AF source extracted by RobustICA-f.

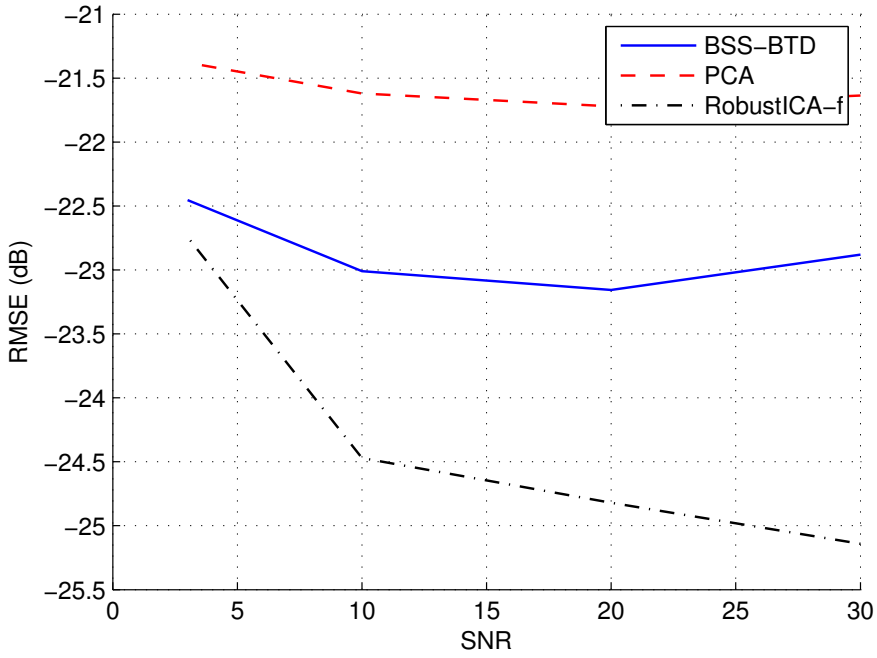

Figure 4.9: Monte Carlo simulation for estimating the performance of the source extraction algorithms under noise in experiment 1 (case A AF source with 8 ECG leads).

Second experiment

Now we consider extracting the case B of AF source from a noiseless ECG with $AVR = -10\,dB$ (Figure 4.10). The AF source was extracted by BSS-BTD with $\mathbf{l} = (42, 42, 42, 42)$ (Figure 4.11) as well as some benchmark methods, such as PCA (Figure 4.12) and RobustICA-f (Figure 4.13). Next, a Monte Carlo simulation consisting of 20 runs was performed in the same fashion as the first experiment (Figure 4.14) and the root mean squared error was calculated.

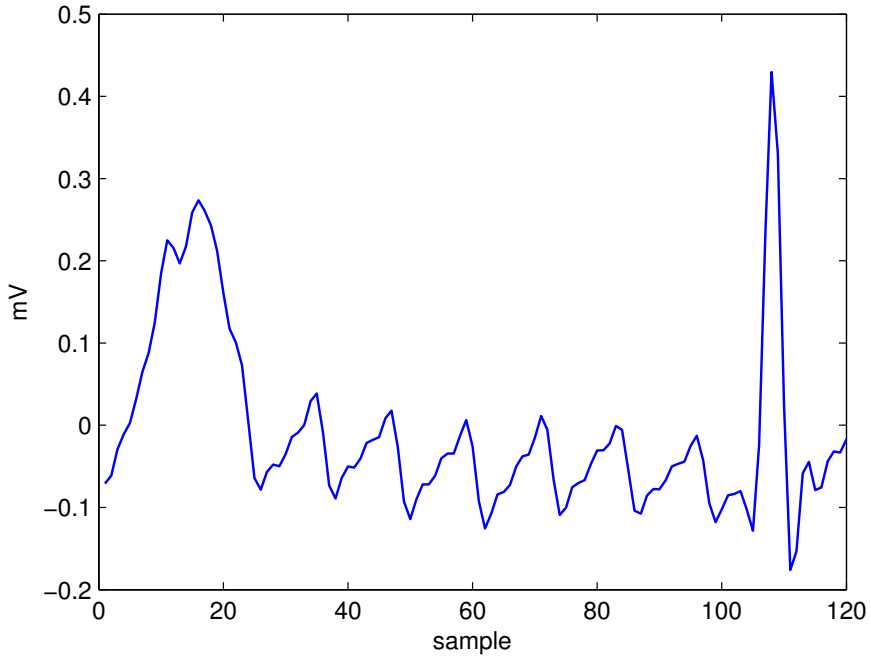

Figure 4.10: Generated ECG signal for the case B AF source with $AVR = -10\,dB$ (lead II).

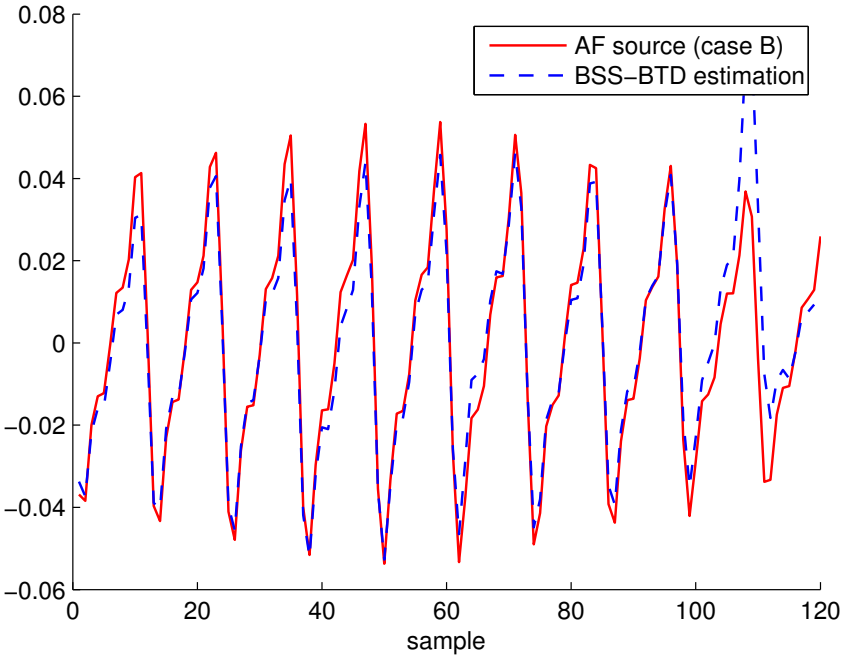

Figure 4.11: Case B AF source extracted by BSS-BTD.

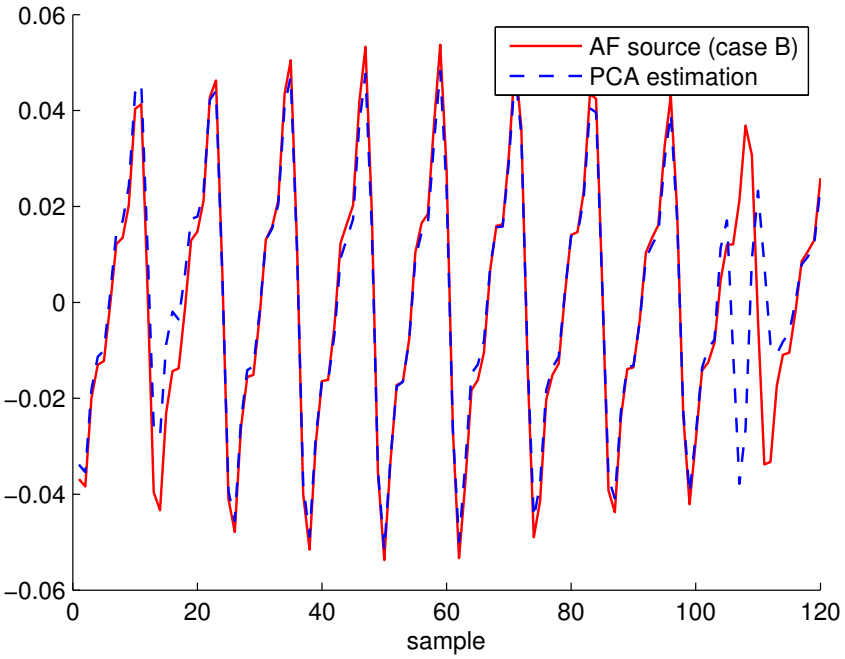

Figure 4.12: Case B AF source extracted by PCA.

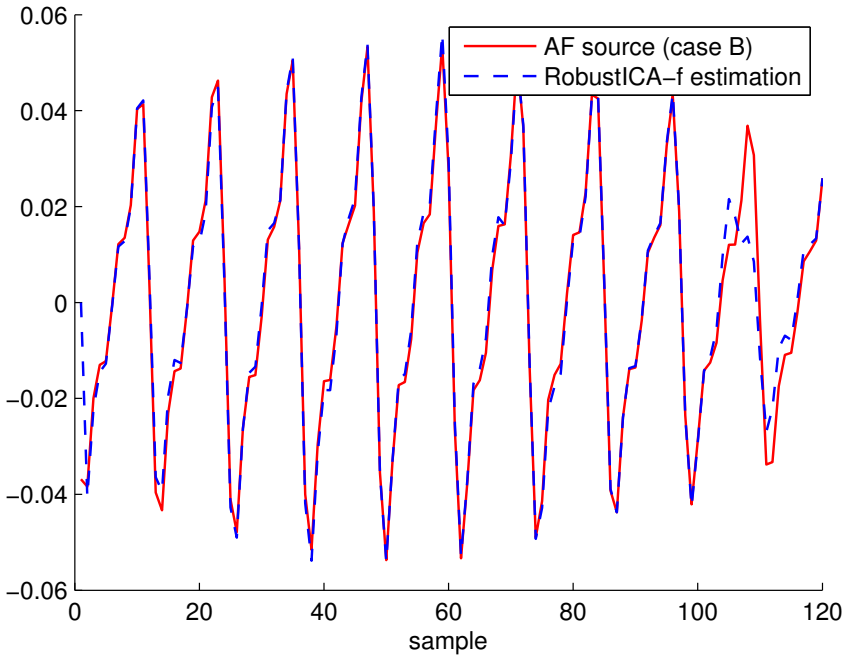

Figure 4.13: Case B AF source extracted by RobustICA-f.

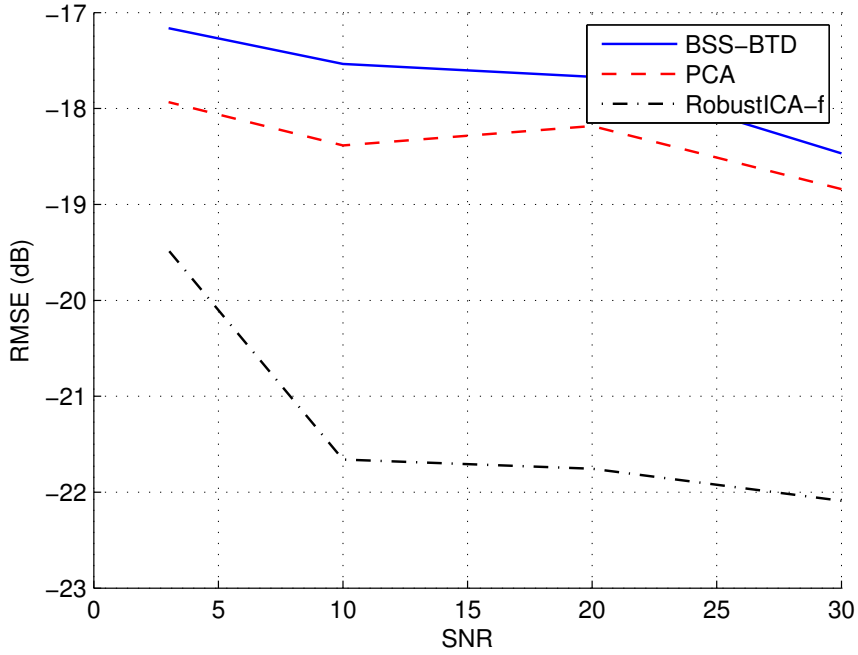

Figure 4.14: Monte Carlo simulation for estimating the performance of the source extraction algorithms under noise in experiment 2 (case B AF source with 8 ECG leads).

Third experiment

Now the BSS-BTD will be compared to PCA and RobustICA-f with minimial spatial diversity, i.e., using only two leads (I and II) for the extraction of case A of AF model. We consider $AVR = -10\,dB$ and an ECG observation of 2.2 s (4.15). The AF source was estimated by BSS-BTD with $\mathbf{l} = (77, 77, 77, 77)$ (Figure 4.16). The sources estimated by PCA and RobustICA-f are depicted in Figures 4.17 and 4.18), respectively. A first Monte Carlo simulation consisting of 20 runs was performed by varying the SNR (Figure 4.19). A second Monte Carlo simulation was performed by varying the number of leads (Figure 4.20) for $SNR = 15$. Like the first simulation, 20 runs was performed.

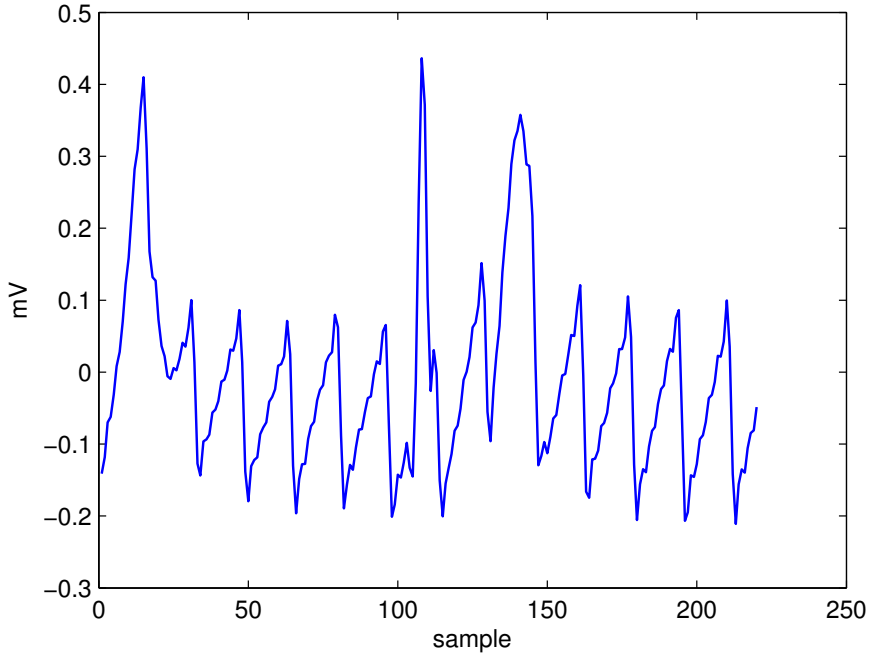

Figure 4.15: Generated ECG signal for the case A of AF source with $AVR = 0.1$ (lead II).

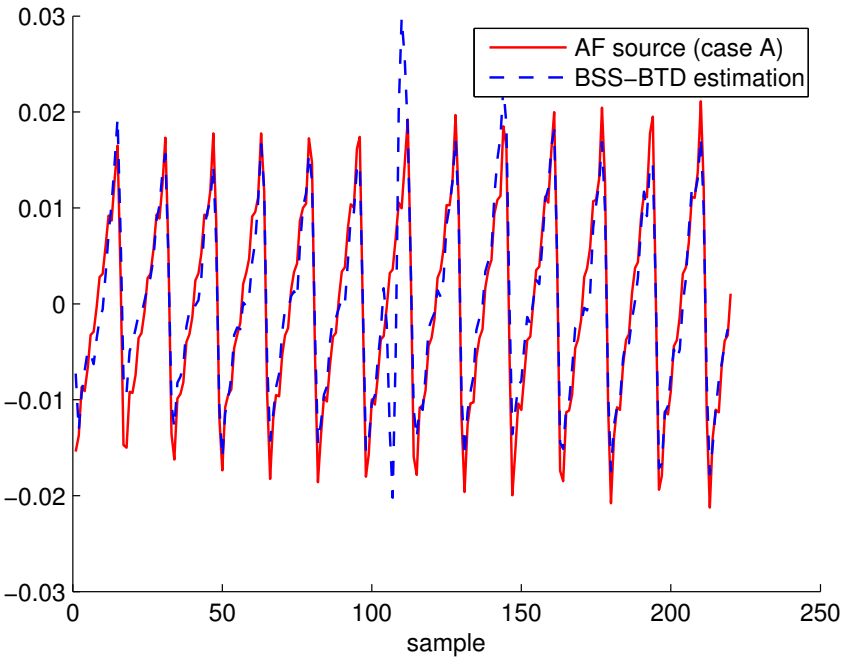

Figure 4.16: Case A AF source extracted by BSS-BTD using two leads.

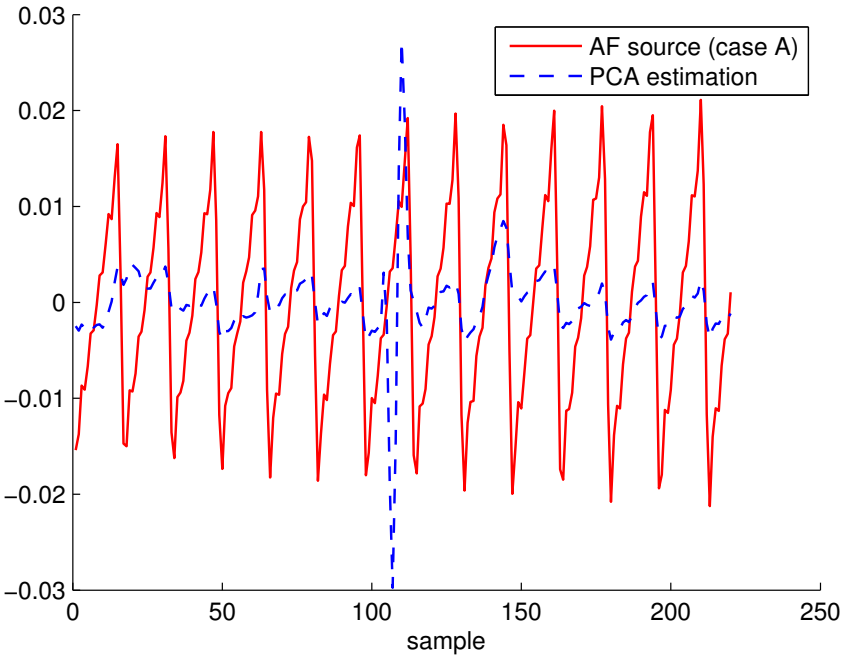

Figure 4.17: Case A AF source extracted by PCA using two leads.

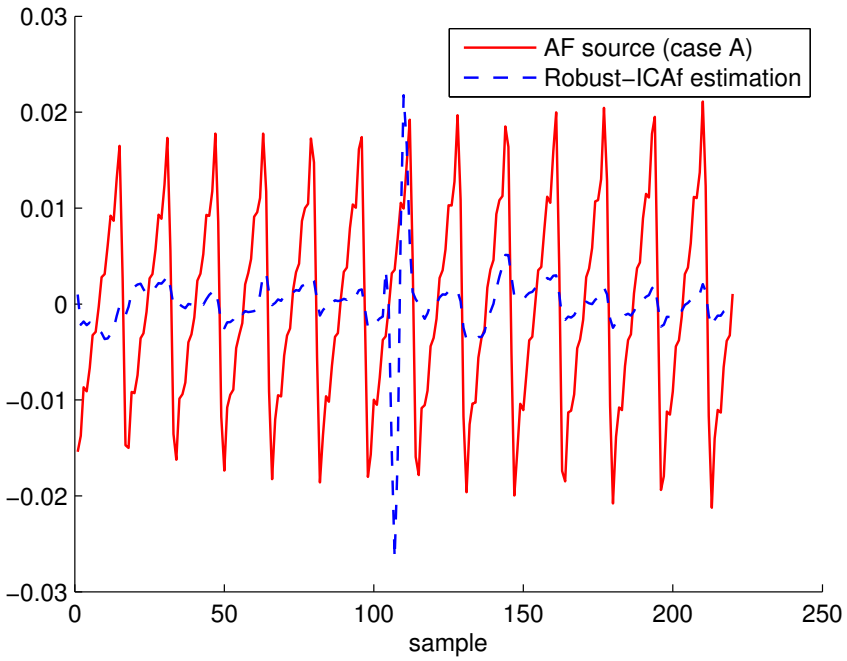

Figure 4.18: Case B AF source extracted by RobustICA-f using two leads.

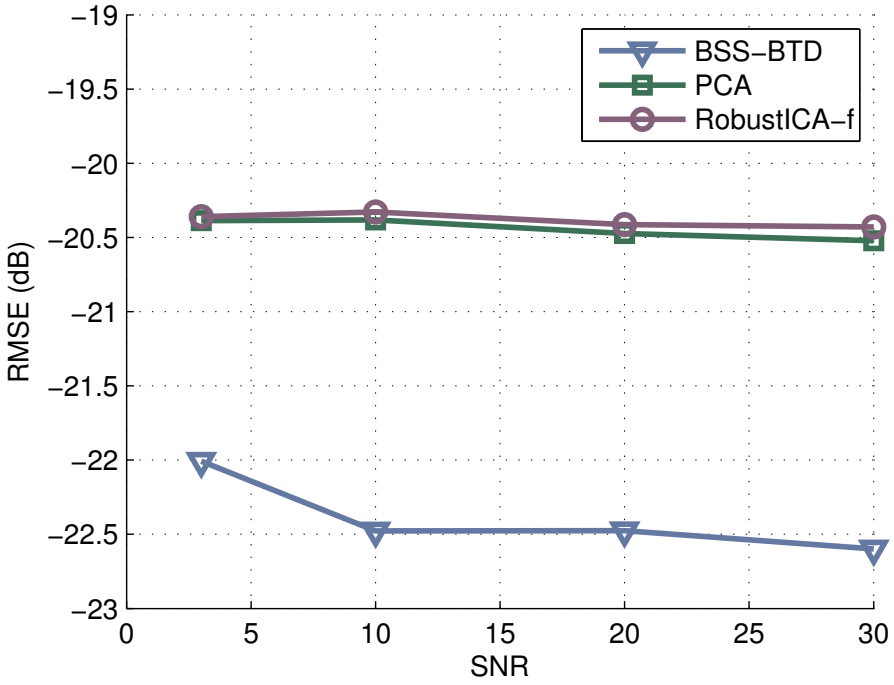

Figure 4.19: Monte Carlo simulation for estimating the performance of the source extraction algorithms under noise in experiment 3 (case A AF source, two leads ECG).

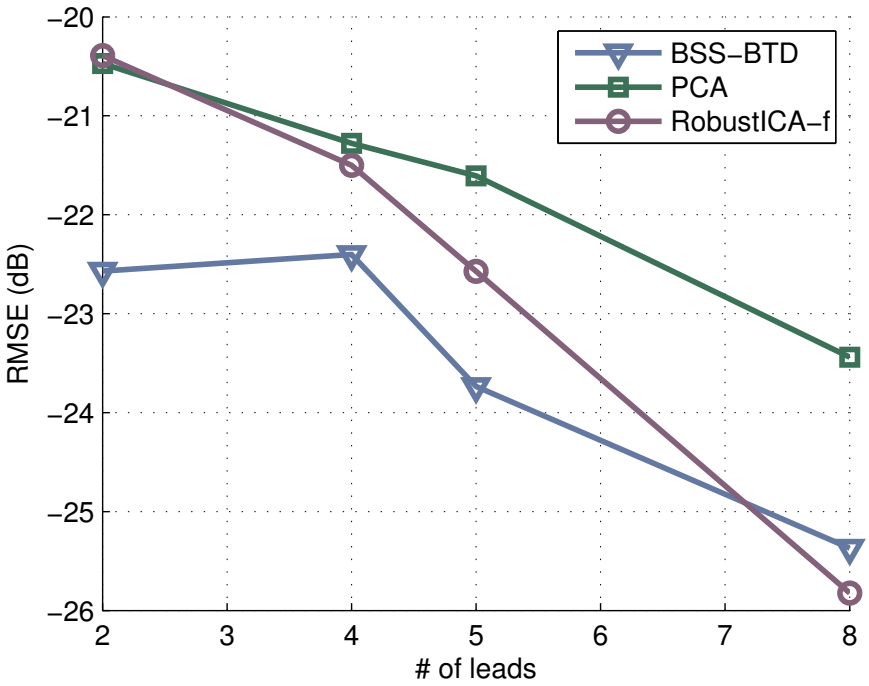

Figure 4.20: Monte Carlo simulation for estimating the performance of the source extraction algorithms with different number of leads (case A AF source, $SNR = 15$, $AVR = -10\,dB$).

Fourth experiment

In the same conditions of the third experiment, but using the case B of AF model, we compare BSS-BTD, PCA and RobustICA-f using only two leads. Lead II is depicted in Figure 4.21. We consider $AVR = -10\,dB$. The AF source was estimated by BSS-BTD with $\mathbf{L} = (66, 66, 66, 66)$ (Figure 4.22). The sources estimated by PCA and RobustICA-f are depicted in Figures 4.23 and 4.24, respectively. A first Monte Carlo simulation consisting of 20 runs was performed by varying the SNR (Figure 4.25). A second Monte Carlo simulation was performed by varying the number of leads (Figure 4.26) for $SNR = 15$. Like the first simulation, 20 runs was performed.

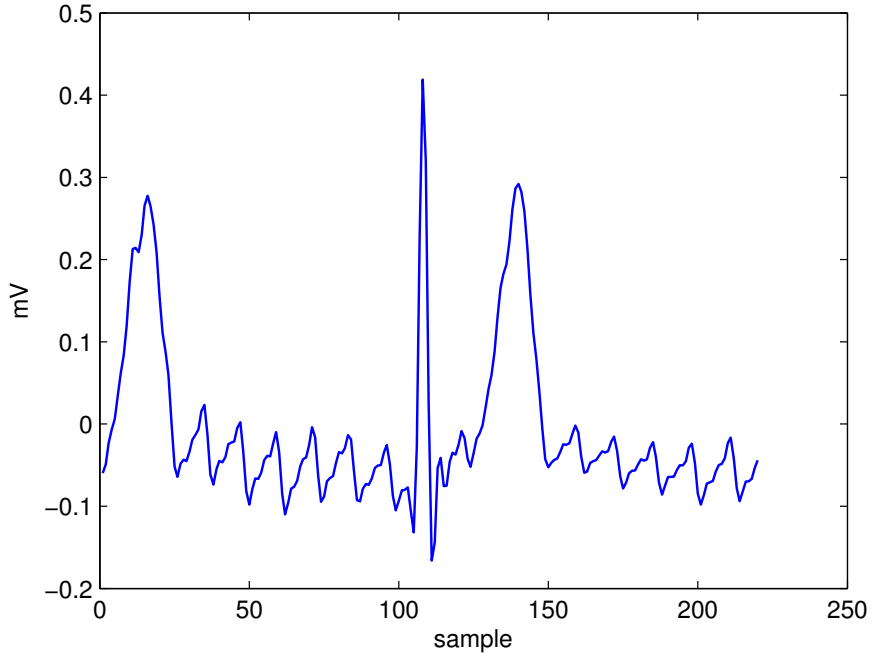

Figure 4.21: Generated ECG signal for the case B of AF source with $AVR = -10\,dB$ (lead II).

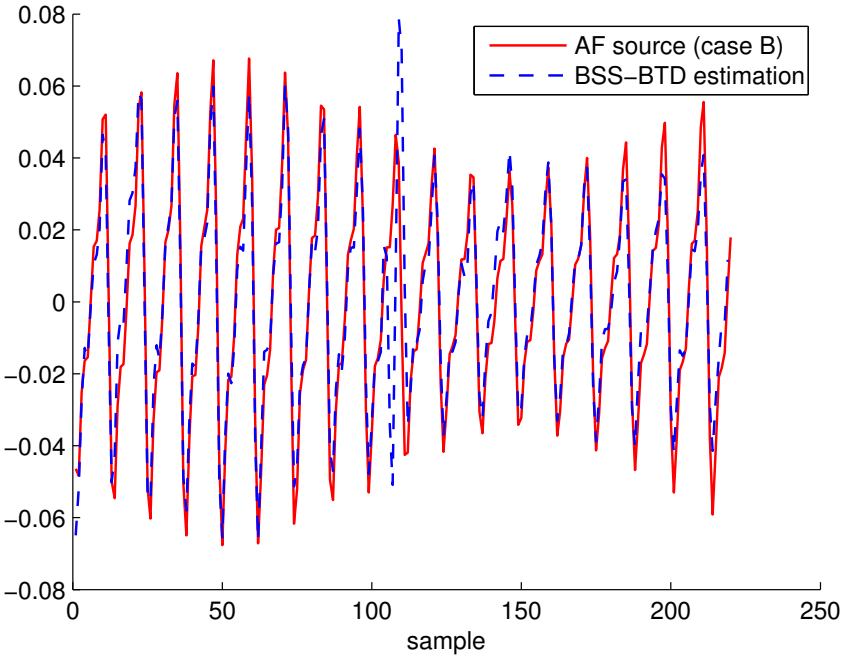

Figure 4.22: Case B AF source extracted by BSS-BTD using two leads.

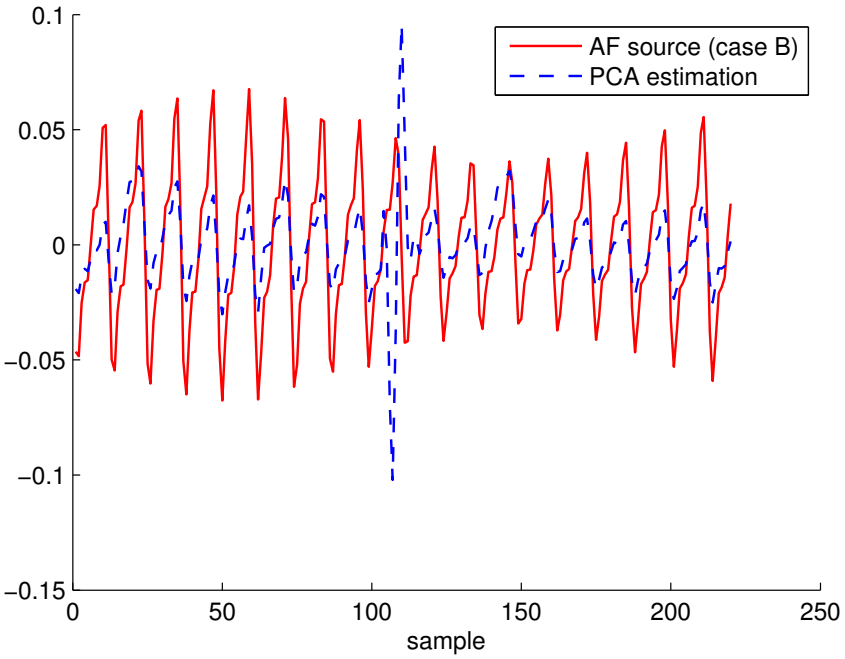

Figure 4.23: Case B AF source extracted by PCA using two leads.

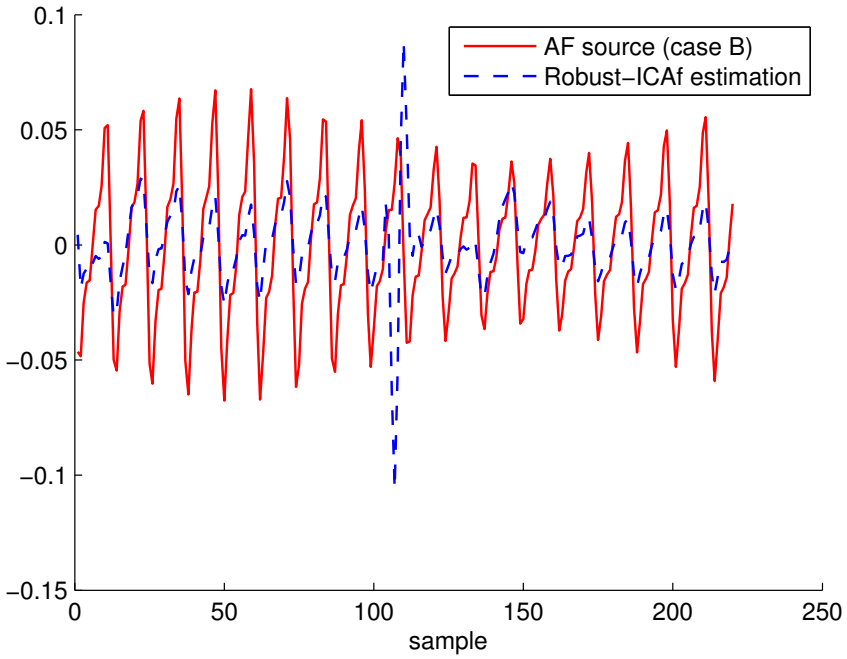

Figure 4.24: Case B AF source extracted by RobustICA-f using two leads.

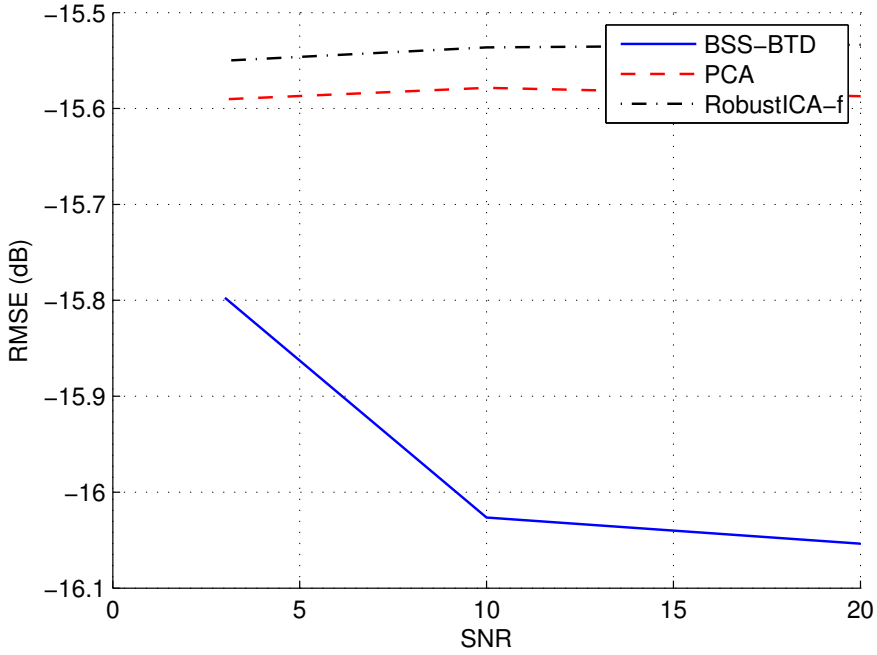

Figure 4.25: Monte Carlo simulation for estimating the performance of the source extraction algorithms under noise in experiment 4 (case B AF source, two leads ECG).

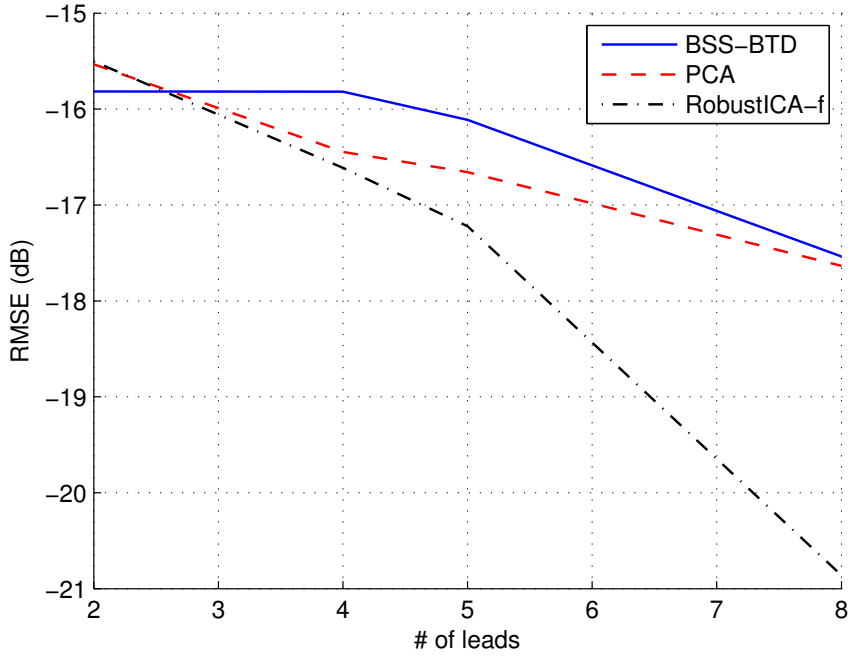

Figure 4.26: Monte Carlo simulation for estimating the performance of the source extraction algorithms with different number of leads (case B AF source, $SNR = 15$, $AVR = -10\,dB$).

Discussion

The BSS-BTD method is comparable to the classic source extraction methods based on matrix decompositions when using 8 leads, as observed in experiments 1 and 2. By contrast, BSS-BTD can extract the AF source using only two leads, while RobustiICA-f and PCA fail, since they do not have enough spatial diversity. Therefore, BSS-BTD presents an advantage over the classical methods using when using only two leads, as observed in experiments 3 and 4.

Even though BSS-BTD can extract Stridh's AF sources, do they fit its deterministic model? In order to use the BSS-BTD, we have to assure that the desired signals can be modeled as (4.15). The model of Stridh and Sornmo consists of:

$$y_l(n) = -\sum_{i=1}^{M} a_{l,i}(n) \sin\left(i\theta(n)\right), \tag{4.45}$$

where

$$a_{l,i}(n) = \frac{2}{\pi i}\left(a_l + \Delta a_l \sin \omega_a n\right) \tag{4.46}$$

and

$$\theta(n) = \omega_0 n + \frac{\Delta f}{f_f} \sin\left(\omega_f n\right). \tag{4.47}$$

Let us extend the model (4.45) in order to allow different fundamental frequencies $\omega_{0,l}$ at each lead l:

$$y_l(n) = -\sum_{i=1}^{M} a_{l,i}(n) \sin\left(i\theta_l(n)\right), \tag{4.48}$$

in which

$$\theta_l(n) = \omega_{0,l} n + \frac{\Delta f}{f_f} \sin\left(\omega_f n\right). \tag{4.49}$$

Note that this modification does not change the structure of the original model. Let us develop the equation (4.48) by substituting (4.49):

$$\begin{aligned} y_l(n) &= -\sum_{i=1}^{M} a_{l,i}(n) \sin\left(i\omega_{0,l} n + i\frac{\Delta f}{f_f} \sin\left(\omega_f n\right)\right) && (4.50)\\ &= -\sum_{i=1}^{M} a_{l,i}(n) \left[\frac{1}{2j} e^{ji\left(\omega_{0,l} n + \frac{\Delta f}{f_f}\sin(\omega_f n)\right)} - \frac{1}{2j} e^{-ji\left(\omega_{0,l} n + \frac{\Delta f}{f_f}\sin(\omega_f n)\right)}\right] && (4.51)\\ &= -\sum_{i=1}^{M} \left[\frac{a_{l,i}(n)}{2j}\left(e^{ji\omega_{0,l}}\right)^n e^{ji\frac{\Delta f}{f_f}\sin(\omega_f n)} - \frac{a_{l,i}(n)}{2j}\left(e^{-ji\omega_{0,l}}\right)^n e^{-ji\frac{\Delta f}{f_f}\sin(\omega_f n)}\right]. && (4.52) \end{aligned}$$

Without loss of generality, the equation (4.52) can be written in a simpler form:

$$y_l(n) = \sum_{r=1}^{R} a_{l,r}(n) p_{l,r}(n) z_{l,r}^n, \tag{4.53}$$

where $p_{l,r}(n)$ is the term responsible for the frequency modulation in (4.52).

We see clearly that (4.53) cannot be modeled as (4.15) due to the time-dependency of the amplitude $a_{l,r}(n)$ and frequency modulation $p_{l,r}(n)$. This observation could help explain why the BSS-BTD does not show a good performance in certain scenarios. We can observe that BSS-BTD performs better extraction for the case A of AF by comparing Figures (4.19) and (4.25). If the BTD-BSS model is modified to deal with these discrepancies of models, its performance might be ameliorated when considering others AF scenarios, such as the case B of AF.

In this chapter, we have introduced a new tensor decomposition called BTD and a source separation method called BSS-BTD. We have used this method for extracting atrial activity from ECG signals and we have compared its performance with benchmark methods such as PCA and RobustICA-f. We have concluded that BSS-BTD performs better than the benchmark methods when the spatial diversity is very constrained. In the next chapter, we will introduce another tensor-based method for extracting characteristics from ECG signals.

Chapter 5

Multilinear Subspace Learning

As discussed in Chapter 2, the PCA and ICA are capable of extracting features (or components) that describe the ECG signal and can be used to interpret and analyze the AF signal and infer some aspects about the heart's condition. Indeed, PCA and ICA can be seen as part of a family of algorithms called Linear Subspace Learning (LSL), in which a high-dimensional data set pertaining to a certain space is projected to a subspace in which the analysis is simplified. Recently, generalizations of LSL methods such as Multilinear PCA (MPCA) [57] and Multilinear ICA (MICA) [58] have been proposed and applied to machine learning [5] and biomedical [46] problems. In this section we discuss the MSL methods and their possible application to AF analysis.

5.1 Multilinear Projections

Data pertaining to a high dimensional space can be transformed to a lower dimensional subspace by means of projections. When dealing with tensor data, multilinear projections are considered. If we are interested in extracting some features from a tensor, or reducing its dimensionality, the best set of projections that will optimize the projected tensor according to a certain criterion need to be found. Before discussing this problem, let us discuss some class of projections.

A *Vector-to-Vector Projection* (VVP) takes a vector $\mathbf{x} \in \mathbb{R}^I$ and projects to another vector $\mathbf{y} \in \mathbb{R}^P$ using a projection matrix $\mathbf{U} \in \mathbb{R}^{I \times P}$:

$$\mathbf{y} = \mathbf{U}^T \mathbf{x}. \tag{5.1}$$

Consider the Nth order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$. Before using a VVP on this tensor, it needs to be vectorized to a vector $\mathbf{x} \in \mathbb{R}^I$ where $I = \prod_{n=1}^{N} I_n$, and then a $(I \times P)$-dimensional matrix will be calculated, resulting in $IP = P \prod_{n=1}^{N} I_n$ parameters to be estimated. Consider a Nth order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$. An *Elementary Multilinear Projection* (EMP) will project $\mathcal{X}$ into a scalar $y \in \mathbb{R}$ through N projection vectors $\mathbf{u}^{(n)} \in \mathbb{R}^{I_n}$:

$$y = \mathcal{X} \times_1 \mathbf{u}^{(1)^T} \times_2 \mathbf{u}^{(2)^T} \ldots \times_N \mathbf{u}^{(N)^T}, \; ||\mathbf{u}^{(n)}|| = 1 \quad \text{for } n = 1, \ldots, N \tag{5.2}$$

Multiplying a tensor by a vector will result in a tensor of order $(N-1)$. Thus, multiplying it by N vectors in all modes will result in a tensor of order 0, i.e., a scalar. Each I_n-dimensional vector needs to be estimated for $n = 1, \ldots, N$, then $\sum_{n=1}^{N} I_n$ parameters needs to be calculated in an EMP. The *Tensor-to-Vector Projection* (TVP) consists of projecting $\mathcal{X}$ to a vector $\mathbf{y} \in \mathbb{R}^P$. The elements of $\mathbf{y}$ are defined as

$$y(p) = \mathcal{X} \times_1 \mathbf{u}_p^{(1)^T} \times_2 \mathbf{u}_p^{(2)^T} \ldots \times_N \mathbf{u}_p^{(N)^T}, \quad \text{for } p = 1, \ldots, P, \tag{5.3}$$

where $\mathbf{u}_p^{(n)} \in \mathbb{R}^{I_n}$. Note that a TVP consists of P EMP, then there are $P \sum_{n=1}^{N} I_n$ parameters to be estimated. In a *Tensor-to-Tensor Projection* (TTP), the tensor $\mathcal{X}$ is projected to a tensor $\mathcal{Y} \in \mathbb{R}^{P_1 \times P_2 \times \ldots \times P_N}$, where $P_n \leq I_n \, \forall n$, by means of projection matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}$:

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \ldots \times_N \mathbf{U}^{(N)^T}, \tag{5.4}$$

where $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times P_n}$. This projection can be performed in N steps. Each step consists of multiplying each I_n-dimensional mode-n fiber of $\mathcal{X}$ by $\mathbf{U}^{(n)}$, resulting in P_n-dimensional fibers. In a TTP, each $(I_n \times P_n)$-dimensional matrix has to be calculated for $n = 1, \ldots, N$, resulting in $\sum_{n=1}^{N} P_n \times I_n$ parameters.

Input	Output	VVP	TVP	TTP
$\prod_{n=1}^{N} I_n$	P	$P\prod_{n=1}^{N} I_n$	$P\sum_{n=1}^{N} I_n$	$\sum_{n=1}^{N} P_n I_n$
10×10	4	400	80	40 ($P_n = 2$)
100×100	4	40000	800	400 ($P_n = 2$)
$100 \times 100 \times 100$	8	8000000	2400	600 ($P_n = 2$)

Table 5.1: Number of parameters of three multilinear projections [5].

Table 5.1 compares the number of parameters of the discussed projections. Note that when the input tensor has a massive quantity of data, VVP needs to estimate an overwhelming number of parameters, while TTP needs to estimate a few hundreds. Therefore, modeling massive data as tensors will allow the utilization of multilinear projections that present the advantage over linear projections relative to the numbers of parameters to be estimated.

5.2 The Multilinear Subspace Learning Framework

Consider a set of M Nth-order tensor samples $\{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M\}$ where $\mathcal{X}_n \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$. The objective of MSL to find a set of projections $\{\mathbf{U}^{(n)}\}$ such that the projected samples will maximize (or minimize) an optimatilty criterion, where the dimensionality of the projected tensors are lower than the original dimensionality. This definition can be mathematically expressed as:

$$\{\mathbf{U}^{(n)}\} = \arg\max_{\{\mathbf{U}^{(n)}\}} \Phi\left(\{\mathbf{U}^{(n)}\}, \{\mathcal{X}_m\}\right) \tag{5.5}$$

for $n = 1, \ldots, N$ and $m = 1, \ldots, M$.

In order to define a MSL algorithm, we need to choose an optimization criterion $\Phi(\cdot)$ and a multilinear projection. The optimization problem (5.5) does not have a closed-form solution, since tensor-based projections have N parameters to be solved in each mode, which depend on projections on others modes. This makes simultaneous estimation very difficult, motivating the use of a suboptimal iterative solutions. Generally, an ALS-based algorithm is used to solve the problem (5.5). This kind of algorithm, requires the selection of some initialization values. A bad initialization may lead the algorithm to converge slowly.

Many MSL methods have been proposed in the literature. They are based on different optimization criteria, such as Least-Squares error minimization, variation maximization, scatter difference maximization and many others [5]. In the next section, we will discuss the MPCA method, a multilinear generalization of the PCA discussed in Section 2.3.1.

5.3 Multilinear Principal Component Analysis

In machine learning applications, data is usually very high-dimensional. Using classical techniques such as PCA to perform dimensionality reduction and feature extraction may be overwhelming when dealing with massive data, as explained in Section 5.1. Lately there have been several attempts to develop multilinear methods of subspace learning.

Lu et al [57] proposed a multilinear algorithm called Multilinear Principal Component Analysis (MPCA) that performs dimensionality reduction in all tensor modes by projecting tensors so that most of the variation in the original tensors is captured.

First of all, let us define a measure of variability and then pose the MPCA problem. Let $\{\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \ldots, \mathcal{X}^{(M)}\}$ be a set of M tensor samples in $\mathbb{R}^{I_1 \times I_2 \ldots \times_N I_N}$. The total scatter of these tensors is defined as

$$\Psi_{\mathcal{X}} = \sum_{m=1}^{M} ||\mathcal{X}^{(m)} - \bar{\mathcal{X}}||_F^2, \tag{5.6}$$

where $\bar{\mathcal{X}}$ is the mean tensor calculated as

$$\bar{\mathcal{X}} = \frac{1}{M} \sum_{m=1}^{M} \mathcal{X}^{(m)}.$$

The mode-n total scatter matrix of these samples is defined as

$$\mathbf{S}_{\mathcal{X}}^{(n)} = \sum_{m=1}^{M} \left(\mathbf{X}_{(n)}^{(m)} - \bar{\mathbf{X}}_{(n)}\right) \left(\mathbf{X}_{(n)}^{(m)} - \bar{\mathbf{X}}_{(n)}\right)^T, \tag{5.7}$$

where $\mathbf{X}_{(n)}^{(m)}$ is the mode-n unfolded matrix of $\mathcal{X}^{(m)}$.

Now consider that a set of M tensor samples $\{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M\}$ is available. Each input tensor $\mathcal{X}_m$ belongs to a tensor space $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \ldots \otimes \mathbb{R}^{I_N}$ where I_n is the mode-n dimension of the tensor. The goal of MPCA is to define a set of multilinear transformations $\{\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times P_n},\ n = 1, \ldots, N\}$ that maps the original tensor space into the subspace $\mathbb{R}^{P_1} \otimes \mathbb{R}^{P_2} \ldots \otimes \mathbb{R}^{P_N}$, $P_n < I_n$, $\forall n$, via a TTP:

$$\mathcal{Y}_m = \mathcal{X}_m \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \ldots \times_N \mathbf{U}^{(N)^T}, \tag{5.8}$$

such that $\mathcal{Y}_m$, for $m = 1, \ldots, M$, captures the most of the variations observed in the original tensor objects. Given the desired projected dimensions P_n for $n = 1, \ldots, N$, these N projection matrices are obtained by maximizing the total scatter of the projected tensors:

$$\left\{\hat{\mathbf{U}}^{(1)}, \hat{\mathbf{U}}^{(2)}, \ldots, \hat{\mathbf{U}}^{(N)}\right\} = \arg \min_{\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}} \Psi_{\mathcal{Y}}. \tag{5.9}$$

Let $\hat{\mathbf{U}}^{(n)}$, $n = 1, \ldots, N$ be the solution of (5.9). It consists of the P_n eigenvectors corresponding to the P_n eigenvalues of the matrix

$$\mathbf{\Phi}^{(n)} = \sum_{m=1}^{M} \left(\mathbf{X}_{(n)}^{(m)} - \bar{\mathbf{X}}_{(n)}^{(m)}\right) \mathbf{U}_{\mathbf{\Phi}^{(n)}} \mathbf{U}_{\mathbf{\Phi}^{(n)}}^T \left(\mathbf{X}_{(n)}^{(m)} - \bar{\mathbf{X}}_{(n)}^{(m)}\right)^T, \tag{5.10}$$

where

$$\mathbf{U}_{\mathbf{\Phi}^{(n)}} = \left(\mathbf{U}^{(N)} \otimes \ldots \otimes \mathbf{U}^{(n+1)} \otimes \mathbf{U}^{(n-1)} \otimes \ldots \otimes \mathbf{U}^{(1)}\right) \tag{5.11}$$

in which $\mathbf{U}^{(n)}$ is the projection matrix for the mode n. The proof of this result can be found in [57]. Note that $P_1, P_2, \ldots, P_N$ are supposed to be known.

A special case of MPCA is called "Full Projection". It consists of a MPCA with $I_n = P_n$ for $n = 1, \ldots, N$. In this case, $\mathbf{\Phi}^{(n)}$ reduces to

$$\mathbf{\Phi}^{(n)^*} = \sum_{m=1}^{M} \left(\mathbf{X}_{(n)}^{(m)} - \bar{\mathbf{X}}_{(n)}^{(m)}\right) \left(\mathbf{X}_{(n)}^{(m)} - \bar{\mathbf{X}}_{(n)}^{(m)}\right)^T. \tag{5.12}$$

The optimal $\mathbf{U}^{(n)} = \mathbf{U}^{(n)^*}$ is obtained as the matrix comprised of the eigenvectors of $\mathbf{\Phi}^{(n)^*}$.

In [57], Lu et al proposed two techniques for automatic dimensionality selection based on including constraints to the objective function in (5.9) to take into account the best dimensionality:

$$\left\{\hat{\mathbf{U}}^{(n)}, \hat{P}_n\right\}_{n=1}^{N} = \arg \min_{\{\mathbf{U}^{(n)}\}_{n=1}^{N}, \{P_n\}_{n=1}^{N}} \Psi_{\mathcal{Y}}. \tag{5.13}$$

subject to

$$\frac{\prod_{n=1}^{N} P_n}{\prod_{n=1}^{N} I_n} < \Omega, \tag{5.14}$$

where Ω is set by the user. The first technique, called Sequential Mode Truncation (SMT), consists of truncating, in a selected mode n, the P_nth n-mode eigenvector of the reconstructed input tensors $\mathcal{X}_m = \mathcal{Y}_m \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_N \mathbf{U}^{(N)}$ until the constraint (5.14) is satisfied. However truncation in one mode affects the eigenvalues in other modes. Another technique, called Q-based method, overcomes this difficulty. For more details, refer to [57].

Since $\mathbf{U}_{\mathbf{\Phi}^{(n)}}$ depends on $\mathbf{U}^{(k)}, k = 1, \ldots, n-1, n+1, \ldots, N$, its optimization depends on the projections in other modes and there is no closed-form solution to this problem. Nevertheless, an iterative approach can be used. The projection matrices are calculated one by one with all the others fixed (local optimization). This optimization is repeated similarly to the ALS method until convergence or the maximum number of iterations is achieved.

An initialization is required for this iterative algorithm. The Full Projection Truncation (FPT) is often used as initialization for the MPCA algorithm. It consists of setting $\mathbf{U}^{(n)}$ as the P_n most significant

eigenvectors of the full projection matrix (5.12) for $n = 1, \ldots, N$. Simulation studies in [57] indicate that a good initialization increases the speed of convergence and, compared to many other initializations, FPT has the faster convergence speed.

The MPCA algorithm will be detailed in the following. Consider the input tensor samples $\{\mathcal{X}_m \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N},\ m = 1, \ldots, M\}$ and their low-dimensional representation $\{\mathcal{Y}_m \in \mathbb{R}^{P_1 \times P_2 \ldots \times P_N},\ m = 1, \ldots, M\}$. First of all, Center the input tensors: $\{\tilde{\mathcal{X}}_m = \mathcal{X}_m - \bar{\mathcal{X}}_m\}$, and then initialize $\mathbf{U}_{\mathbf{\Phi}^{(n)}}$. Subsequently, perform the following local optimization:

- Calculate $\tilde{\mathcal{Y}}_m = \tilde{\mathcal{X}}_m \times_1 \mathbf{U}^{(1)^T} \ldots \times_N \mathbf{U}^{(N)^T}$ for $m = 1, \ldots, M$.
- Calculate $\mathbf{\Psi}_{\mathcal{Y}_0} = \sum_{m=1}^{M} ||\tilde{\mathcal{Y}}||_F^2$.
- For $k = 1, \ldots, K$
  - For $n = 1, \ldots, N$
    - Set $\mathbf{U}^{(n)}$ as the P_n most significant eigenvectors of $\mathbf{\Psi}^{(n)}$.
  - Calculate $\tilde{\mathcal{Y}}_m$ and $\Psi_{\mathcal{Y}_k}$ for $m = 1, \ldots, M$.
  - If $|\Psi_{\mathcal{Y}_k} - \Psi_{\mathcal{Y}_{k-1}}| < \eta$, break the loop.

Finally, project the input tensors into the subspace: $\mathcal{Y}_m = \mathcal{X} \times_1 \mathbf{U}^{(1)^T} \ldots \times_N \mathbf{U}^{(N)^T}$ for $m = 1, \ldots, M$.

It should be noted that while PCA produces uncorrelated principal components, MPCA does not produce uncorrelated tensors. Instead, refer to Uncorrelated MPCA [59]. Moreover, MPCA was not designed to minimize the reconstruction error, but rather maximizing the total variation between the tensor samples. Unlike PCA, whose principal components are ordered according to their respective eigenvalue, MPCA does not produce ordered projected tensors, since the feature tensor $\mathcal{Y}_m$ is obtained directly from the sample tensor $\mathcal{X}_m$ by projection.

MSL methods are interesting since they present some advantages over LSL methods, like reduced quantity of parameters to be estimated. Thus, these methods are well adapted to massive data tensors. The definitions of MPCA were introduced and we have seen that it can be used to extract features from a set of input tensors. It can be used to extract components that strongly contribute to the total variation of the input tensors. In the ECG context, these components correspond to ventricular activity. This correspondence motivates further investigation of applying MPCA to ECG signals.

Chapter 6

Conclusion and further work

This research projected was composed of two parts. The first part was devoted to a bibliographic research aiming at:

- Studying the biophysical mechanisms of the heart.
- Studying the state of the art of AF extraction methods.
- Studying existing tensorial methods proposed tensors techniques for ECG signal processing.

The second part was focused on the implementation and validation of the studied methods:

- Applying the BSS-BTD method for AF extraction.
- Validating the obtained results by comparing them with the results of existent methods.

The tensorial modeling of ECG signals showed to be a very promising research field, because of the advantages that high-order tensor methods present over matrix methods, and due to the many problems that can be more efficiently solved by this modeling.

In particular, the BSS-BTD method is able to better recover AF sources using fewer sensors than matrix methods, offering superior performance in scenarios with limited spatial diversity. This improved performance comes from the fact that tensors are uniquely decomposed in mild conditions, allowing the separation of underdetermined mixtures. It is important to stress that even though we used a synthetic AF model to assess the performance of this method, this model is a established method in the literature [4, 6] for analyzing the performance of AF extraction algorithms.

Since good results were obtained for synthetic signals, applying the BSS-BTD method to real AF signals is motivated. Adapting the BSS-BTD model to Stridh's AF model, extracting features from ECG tensors by means of MPCA, and using tensor decompositions for noise removal and AF extraction are some perspectives for the continuation of the work presented in this report.

Appendix A

Solving permutation and scaling indeterminacies

Consider the instantaneous linear mixture model (2.31). Let the $\hat{\mathbf{M}}$ be the estimate of $\mathbf{M}$. In a perfect separation, we ideally have $\hat{\mathbf{M}} = \mathbf{MDP}$ in which $\mathbf{D} \in \mathbb{R}^{R\times R}$ is a nonsingular scaling diagonal matrix and $\mathbf{P} \in \mathbb{R}^{R\times R}$ a permutation matrix.

To solve these indeterminacies, it is assumed that $\mathbf{M}$ is known. The matrix $\mathbf{C} = |\hat{\mathbf{M}}^T\mathbf{M}|$ is calculated. The element (k, r) of this matrix consists of the correlation between the kth column of $\hat{\mathbf{M}}$ and the rth column of $\mathbf{M}$. The column r of $\mathbf{M}$ that corresponds to the column k of $\hat{\mathbf{M}}$ is equivalent to the column with the largest element of the kth row of $\mathbf{C}$. The permutation of columns can be carried out by multiplying $\hat{\mathbf{M}}$ with a matrix $\mathbf{P}'$ that reflects the column correspondences found in $\mathbf{C}$. Finally the permutation indeterminacy can be undone by calculating $\hat{\mathbf{M}}_p = \hat{\mathbf{M}}\mathbf{P}'$.

Let $\hat{\mathbf{m}}_r$ and $\mathbf{m}_r$ be the rth column of $\hat{\mathbf{M}}_p$ and $\mathbf{M}$, respectively. The scaling indeterminacy between $\hat{\mathbf{m}}_r$ and $\mathbf{m}_r$ can be solved by solving the problem

$$\min_{\alpha_r} ||\mathbf{m}_r - \alpha_r\hat{\mathbf{m}}_r||^2. \tag{A.1}$$

The optimum α_r can be found by setting the gradient of $J(\alpha_r) = ||\mathbf{m}_r - \alpha_r\hat{\mathbf{m}}_r||^2$ to zero:

$$\begin{aligned}
\frac{\partial J(\alpha_r)}{\partial \alpha_r} &= \frac{\partial}{\partial \alpha_r}||\mathbf{m}_r - \alpha_r\hat{\mathbf{m}}_r||^2 \\
&= \frac{\partial}{\partial \alpha_r}\left\{(\mathbf{m}_r - \alpha_r\hat{\mathbf{m}}_r)^T(\mathbf{m}_r - \alpha_r\hat{\mathbf{m}}_r)\right\} \\
&= \frac{\partial}{\partial \alpha_r}\left\{\mathbf{m}_r^T\mathbf{m}_r - 2\alpha_r\mathbf{m}_r^T\hat{\mathbf{m}}_r + \alpha_r^2\hat{\mathbf{m}}_r^T\hat{\mathbf{m}}_r\right\} \\
&= -2\mathbf{m}_r^T\hat{\mathbf{m}}_r + 2\alpha_r\hat{\mathbf{m}}_r^T\hat{\mathbf{m}}_r = 0.
\end{aligned} \tag{A.2}$$

Then

$$\alpha_r^* = \frac{\mathbf{m}_r^T\hat{\mathbf{m}}_r}{\hat{\mathbf{m}}_r^T\hat{\mathbf{m}}_r}. \tag{A.3}$$

The scaling indeterminacy can be resolved by calculating α_r for the $r = 1, \ldots, R$ columns of $\hat{\mathbf{M}}_p$, storing them in a diagonal matrix $\mathbf{D}'$ and calculating $\hat{\mathbf{M}}_p\mathbf{D}'$. Finally, the indeterminacies are undone by $\mathbf{M} = \hat{\mathbf{M}}\mathbf{P}'\mathbf{D}'$.

Bibliography

[1] B. J. Drew, R. M. Califf, M. Funk, E. S. Kaufman, M. W. Krucoff, M. M. Laks, P. W. Macfarlane, C. Sommargren, S. Swiryn, and G. F. Van Hare, "Practice standards for electrocardiographic monitoring in hospital settings an american heart association scientific statement from the councils on cardiovascular nursing, clinical cardiology, and cardiovascular disease in the young: Endorsed by the international society of computerized electrocardiology and the american association of critical-care nurses," *Circulation*, vol. 110, no. 17, pp. 2721–2746, 2004.

[2] M. Oeff, H. Koch, R. Bousseljot, and D. Kreiseler, "The PTB diagnostic ECG database, national metrology institute of germany," 1995.

[3] W. J. Tompkins, "Biomedical digital signal processing," *Editorial Prentice Hall*, 1993.

[4] M. Stridh and L. Sornmo, "Spatiotemporal QRST cancellation techniques for analysis of atrial fibrillation," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 1, pp. 105–111, 2001.

[5] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A survey of multilinear subspace learning for tensor data," *Pattern Recognition*, vol. 44, no. 7, pp. 1540–1551, 2011.

[6] L. Mainardi, L. Sornmo, and S. Cerutti, *Understanding atrial fibrillation: the signal processing contribution*. Morgan & Claypool Publishers, 2008.

[7] J. J. Rieta, F. Castells, C. Sánchez, V. Zarzoso, and J. Millet, "Atrial activity extraction for atrial fibrillation analysis using blind source separation," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1176–1186, 2004.

[8] M. Stridh and L. Sornmo, "Spatiotemporal QRST cancellation techniques for analysis of atrial fibrillation," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 1, pp. 105–111, 2001.

[9] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2005.

[10] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.

[11] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 810–823, 2000.

[12] A. L. De Almeida, G. Favier, and J. C. M. Mota, "PARAFAC-based unified tensor modeling for wireless communication systems with application to blind multiuser equalization," *Signal Processing*, vol. 87, no. 2, pp. 337–351, 2007.

[13] A. Ozerov, C. Févotte, R. Blouet, and J.-L. Durrieu, "Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 257–260, IEEE, 2011.

[14] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A survey of multilinear subspace learning for tensor data," *Pattern Recognition*, vol. 44, no. 7, pp. 1540–1551, 2011.

[15] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[16] L. Sörnmo and P. Laguna, *Bioelectrical signal processing in cardiac and neurological applications*. Academic Press, 2005.

[17] N. V. Thakor and Y.-S. Zhu, "Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection," *IEEE Transactions on Biomedical Engineering*, vol. 38, no. 8, pp. 785–794, 1991.

[18] A. Bollmann, D. Husser, L. Mainardi, F. Lombardi, P. Langley, A. Murray, J. J. Rieta, J. Millet, S. B. Olsson, M. Stridh, *et al.*, "Analysis of surface electrocardiograms in atrial fibrillation: techniques, research, and clinical applications," *Europace*, vol. 8, no. 11, pp. 911–926, 2006.

[19] C. J. James, M. T. Hagan, R. D. Jones, P. J. Bones, and G. J. Carroll, "Multireference adaptive noise canceling applied to the EEG," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 8, pp. 775–779, 1997.

[20] J. Slocum, E. Byrom, L. McCarthy, A. Sahakian, and S. Swiryn, "Computer detection of atrioventricular dissociation from surface electrocardiograms during wide QRS complex tachycardias.," *Circulation*, vol. 72, no. 5, pp. 1028–1036, 1985.

[21] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3. JHU Press, 2012.

[22] F. Castells, J. J. Rieta, J. Millet, and V. Zarzoso, "Spatiotemporal blind source separation approach to atrial activity estimation in atrial tachyarrhythmias," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 2, pp. 258–267, 2005.

[23] R. Phlypo, Y. D'Asseler, I. Lemahieu, and V. Zarzoso, "Extraction of the atrial activity from the ECG based on independent component analysis with prior knowledge of the source kurtosis signs," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 6499–6502, IEEE, 2007.

[24] V. Zarzoso, "Extraction of ecg characteristics using source separation techniques: exploiting statistical independence and beyond," in *Advanced Biosignal Processing*, pp. 15–47, Springer, 2009.

[25] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[26] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4, pp. 411–430, 2000.

[27] V. Zarzoso and P. Comon, "Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size," *IEEE Transactions on Neural Networks*, vol. 21, no. 2, pp. 248–261, 2010.

[28] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[29] P. Comon, "Tensors: a brief introduction," in *IEEE Sig. Proc. Magazine*, vol. 31, 2014.

[30] L. R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change," in *Problems in measuring change.* (C. W. Harris, ed.), pp. 122–137, Madison WI: University of Wisconsin Press, 1963.

[31] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[32] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation.* John Wiley & Sons, 2009.

[33] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-$(R_1, R_2, \ldots, R_n)$ approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

[34] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[35] M. E. Timmerman and H. A. Kiers, "Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima," *British Journal of Mathematical and Statistical Psychology*, vol. 53, no. 1, pp. 1–16, 2000.

[36] R. A. Harshman, "Foundations of the parafac procedure: models and conditions for an" explanatory" multimodal factor analysis," 1970.

[37] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of n-way arrays," *Journal of chemometrics*, vol. 14, no. 3, pp. 229–239, 2000.

[38] A. Stegeman, "On uniqueness conditions for candecomp/parafac and indscal with full column rank in one mode," *Linear Algebra and its Applications*, vol. 431, no. 1, pp. 211–227, 2009.

[39] I. Domanov and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors—part I: Basic results and uniqueness of one factor matrix," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 855–875, 2013.

[40] R. A. Harshman and M. E. Lundy, "Uniqueness proof for a family of models sharing features of tucker's three-mode factor analysis and PARAFAC/CANDECOMP," *Psychometrika*, vol. 61, no. 1, pp. 133–154, 1996.

[41] A. L. De Almeida, G. Favier, and J. C. M. Mota, "A constrained factor decomposition with application to MIMO antenna systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2429–2442, 2008.

[42] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the parafac model," *Computational Statistics & Data Analysis*, vol. 50, no. 7, pp. 1700–1734, 2006.

[43] R. Bro and H. A. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *Journal of chemometrics*, vol. 17, no. 5, pp. 274–286, 2003.

[44] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, and B. Yener, "Multiway analysis of epilepsy tensors," *Bioinformatics*, vol. 23, no. 13, pp. i10–i18, 2007.

[45] F. Miwakeichi, E. Martínez-Montes, P. A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space–time–frequency components using parallel factor analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, 2004.

[46] K. Huang and L. Zhang, "Cardiology knowledge free ECG feature extraction using generalized tensor rank one discriminant analysis," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, pp. 1–15, 2014.

[47] B. Hunyadi, D. C. L. Sorber, W. Van Paesschen, M. De Vos, S. Van Huffel, and L. De Lathauwer, "Block term decomposition for modelling epileptic seizures,"

[48] B. N. Singh and A. K. Tiwari, "Optimal selection of wavelet basis function applied to ECG signal denoising," *Digital Signal Processing*, vol. 16, no. 3, pp. 275–287, 2006.

[49] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

[50] M. Niknazar, H. Becker, B. Rivet, C. Jutten, and P. Comon, "Robust 3-way tensor decomposition and extended state kalman filtering to extract fetal ECG," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO), 2013*, pp. 1–5, IEEE, 2013.

[51] L. De Lathauwer, "Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms," *SIAM Journal on Matrix Analysis and Applications*, vol. 32, no. 4, pp. 1451–1474, 2011.

[52] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms-part ii: Definitions and uniqueness," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1033–1066, 2008.

[53] S.-Y. Kung, K. Arun, and D. Rao, "State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem," *JOSA*, vol. 73, no. 12, pp. 1799–1811, 1983.

[54] D. L. Boley, F. T. Luk, and D. Vandevoorde, "Vandermonde factorization of a hankel matrix," *Scientific computing*, pp. 27–39, 1998.

[55] S. Wright and J. Nocedal, *Numerical optimization*, vol. 2. Springer New York, 1999.

[56] L. Sorber, M. Van Barel, and L. De Lathauwer, "Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-$(L_r, L_r, 1)$ terms, and a new generalization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 695–720, 2013.

[57] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, 2008.

[58] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear independent components analysis," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005.*, vol. 1, pp. 547–553, IEEE, 2005.

[59] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear principal component analysis through successive variance maximization," in *Proceedings of the 25th international conference on Machine learning*, pp. 616–623, ACM, 2008.