

A Deflation Procedure for Subspace Decomposition

Reza Sameni, Christian Jutten, Mohammad B. Shamsollahi

▶ **To cite this version:**

Reza Sameni, Christian Jutten, Mohammad B. Shamsollahi. A Deflation Procedure for Subspace Decomposition. IEEE Transactions on Signal Processing, 2010, 58 (4), pp.2363-2374. 10.1109/TSP.2009.2037353 . hal-00466435

HAL Id: hal-00466435
https://hal.science/hal-00466435

Submitted on 23 Mar 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Deflation Procedure for Subspace Decomposition

Reza Sameni, *Member, IEEE*, Christian Jutten, *Fellow, IEEE*, and Mohammad B. Shamsollahi, *Member, IEEE*

***Abstract*—A general deflation framework is described for the separation of a desired signal subspace of arbitrary dimensions from noisy multichannel observations. The method simultaneously uses single and multichannel priors to split the desired and undesired subspaces, even for coplanar (intersecting) subspaces. By appropriate use of signal priors, it can even extract signals from degenerate mixtures of signals and noise recorded from a few number of channels in low SNR scenarios, without the reduction of the data dimensions. As a case study, the performance of the proposed method is studied for the problem of extracting fetal cardiac signals from maternal abdominal recordings, over simulated and real data. A second case study deals with the degenerate problem of extracting diaphragmatic electromyogram from electrocardiograph artifacts. A provisional patent application based on this method has been filed.**

***Index Terms*—Deflation, nonlinear filtering, nonstationarity, periodicity, prior, rank preservation, Rayleigh quotient, signal denoising, subspace.**

I. Introduction

SIGNAL denoising is among the most developed and competitive areas of stochastic signal processing. To date, numerous approaches have been proposed for this purpose [1]–[3]. Despite their diversity, a common objective of these methods is to find a domain, such as the time, frequency, scale, space, etc., in which the desired and undesired signals are (partially) separable. Each of these algorithms has theoretical upper performance bounds; however, depending on the application, additional processing gain can usually be achieved by using additional priors or by increasing the number of observation channels.

During the past two decades, multichannel analysis based on blind or semi-blind source separation (BSS) has highly developed [4]–[6]. Due to the minimal assumptions that they impose on the data model, these methods have drawn special attention for the separation of desired signals from noisy multichannel recordings [7]. However, due to the so-called "blindness" of these methods, the performance of BSS is limited to moderately low-noise situations with sufficient number of observations, known as the *determined* and *over-determined* cases [4]. Fortunately, real world applications are not totally "blind," and we typically have some prior information about the signals of interest. Currently, a large body of research is focused on the effective use of signal priors, such as signal dynamics [5, Ch. 11], sparseness [8], and time-frequency footprints [9], for improving the signal/interference separability of BSS and solving the *under-determined* case.

In this paper, a framework is presented for the separation of a desired signal subspace from noisy multichannel observations. The proposed method is an iterative procedure that is repeatedly applied to the input signal until all the dimensions of the desired subspace are extracted. This method simultaneously uses single and multichannel priors to split the desired and undesired subspaces, even for coplanar (intersecting) subspaces. It is experimentally shown that by appropriate use of such priors, one can even extract signals from degenerate mixtures in low SNR scenarios. The proposed method is fairly general and may be applied to various applications. As a case study, it is used for the separation of fetal cardiac signals from maternal interference and noise, over simulated and real data. In a second case study, the method is applied to the degenerate problem of extracting diaphragmatic electromyogram from electrocardiograph artifacts. A provisional patent application based on this method has been filed [10].

The rest of the paper is organized as follows. In Sections II and III, the required background and data model are presented. The proposed method is presented in Section IV. Sections V and VI focus on two special case studies, which are followed by a summary and concluding remarks in Section VII.

II. Generalized Eigenvalue Decomposition

For symmetric matrices $A, B \in \mathbb{R}^{N\times N}$, the problem of generalized eigenvalue decomposition (GEVD) [11], of the matrix pair (A, B), consists of finding matrices W and D, such that

$$W^T A W = D, \quad W^T B W = I \tag{1}$$

where D is a diagonal matrix of the generalized eigenvalues corresponding to the eigenmatrix $W = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$, with real eigenvalues sorted in ascending order on its diagonal.[1] In (1), W is a transform that simultaneously diagonalizes A and B. Moreover, the first eigenvector $\mathbf{w}_1$, corresponding to the largest generalized eigenvalue, also maximizes the following ratio, known as the *Rayleigh quotient* [11]

$$J(\mathbf{w}) = \frac{\mathbf{w}^T A \mathbf{w}}{\mathbf{w}^T B \mathbf{w}}. \tag{2}$$

[1]In the problem of interest, A and B are symmetric, and commonly, positive definite matrices; therefore, the eigenvalues are real and positive [11].

Manuscript received February 21, 2009; accepted October 07, 2009. First published November 24, 2009; current version published March 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Z. Jane Wang.

R. Sameni is with the School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran (e-mail: rsameni@shirazu.ac.ir).

C. Jutten is with the Laboratory of Grenoble Image Parole Signal Automatique (GIPSA-lab), Grenoble, France, and also with the Institut Universitaire de France (e-mail: christian.jutten@gipsa-lab.grenoble-inp.fr).

M. B. Shamsollahi is with the Biomedical Signal and Image Processing Laboratory (BiSIPL), School of Electrical Engineering, Sharif University of Technology, Tehran, Iran (e-mail: mbshams@sharif.ir).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2037353

1053-587X/$26.00 © 2010 IEEE

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.

III. Data Model

We consider zero-mean N-dimensional observations $\mathbf{x}(t) \in \mathbb{R}^N$, which follow an additive model

$$\mathbf{x}(t) = \mathbf{x}_s(t) + \mathbf{x}_n(t) \tag{3}$$

where $\mathbf{x}_s(t)$ is the *desired* part of the observations, considered as the signal part, and $\mathbf{x}_n(t)$ is the *undesired* part, considered as noise and/or interference. The covariance matrix of the observations is defined as follows:

$$C_{\mathbf{x}} \doteq E_t\{\mathbf{x}(t)\mathbf{x}(t)^T\} \tag{4}$$

where $E_t\{\cdot\}$ represents averaging over time. We further assume that the desired and undesired parts of the observations be uncorrelated, but the desired parts of the different channels be somehow dependent. For example, they can be linear or nonlinear mixtures of a signal subspace projected onto the observation space

$$\mathbf{x}_s(t) = f[\mathbf{s}(t)] \tag{5}$$

where $\mathbf{s}(t) \in \mathbb{R}^M$ and $f(\cdot) : \mathbb{R}^M \mapsto \mathbb{R}^N$. The undesired part of the signal is also assumed to be a mixture of low-rank [12], and full-rank noise and/or interferences. The objective is to denoise the observations, i.e., to retrieve $\mathbf{x}_s(t)$.

The data model is very general and appears in many applications. For example, the problem of blind source (or subspace) extraction with a latent variable model is a special case for this model

$$\begin{aligned}\mathbf{x}(t) &= H\mathbf{u}(t) + \mathbf{n}(t) \\ &= \underbrace{u_1(t)\mathbf{h}_1 + \cdots + u_M(t)\mathbf{h}_M}_{\mathbf{x}_s(t)} \\ &\quad + \underbrace{\cdots + u_P(t)\mathbf{h}_P + \mathbf{n}(t)}_{\mathbf{x}_n(t)}\end{aligned} \tag{6}$$

where $H \in \mathbb{R}^{N\times P}$, $\mathbf{u}(t) \in \mathbb{R}^P$, $\mathbf{n}(t) \in \mathbb{R}^N$, and $\mathbf{h}_i \in \operatorname{span}\{H\}$. In this case, it is assumed that the signal of interest is a subspace of the column space of H, while the other subspaces and noises form the undesired part.

Under the assumption of independent $u_i(t)$ with $P \leq N$, and a moderately small noise $\mathbf{n}(t)$, the latent variable model can be solved by second and higher-order statistical methods, known as independent component analysis (ICA) [13], [14]. The degenerate (*under-determined*) case, where $P > N$, is yet more challenging and does not generally have a solution, unless when additional assumptions such as sparsity (in the time or transform domain) are valid [8].

With this background, the limitations of the currently existing methods include:

a) the estimation of the demixing matrix is sensitive to the data model and observation noise;
b) the degenerate case can not be solved, except under special assumptions and prior information;
c) dimensions of the desired subspace is commonly required in advance;
d) coplanar (intersecting) subspaces can not be separated;
e) full-rank noise is not separated and may even be amplified in the extracted components [12];
f) the rank of the observations are commonly reduced, if used for denoising.

In what follows, by using *a priori* knowledge of the signal/noise subspaces, we present a method that removes some of the mentioned limitations.

IV. Method

A. Single Channel Denoising

Regardless of the multichannel structure of the data model in (3), in each of the channels the desired and undesired parts of the observation may be separable (although approximately) using some linear or nonlinear denoising scheme applied in the time or transform domain. The optimal denoising scheme is application dependent and its performance depends on the signal-to-noise ratio (SNR) of each channel. Conventional frequency domain filters, wavelet denoisers, optimal Wiener filters, linear or nonlinear Kalman filters are amongst the possible denoising schemes for single channel denoising. However, due to the commonly low SNR of the observations, single channel denoising is not very efficient for many applications and we seek for a method to benefit from the mutual spatial information within the different channels to improve the signal quality before denoising.

Filters used for denoising typically have a monotonic characteristic, i.e., the quality of a signal (in any of the later mentioned senses) should not be degraded by the filtering process. More rigorously, if $\mathbb{Q}(x)$ represents the quality of the signal x and $F(\cdot)$ denotes the filtering operator, the following property holds:

$$\mathbb{Q}(a) \geq \mathbb{Q}(b) \Rightarrow \mathbb{Q}(F(a)) \geq \mathbb{Q}(F(b)). \tag{7}$$

We hereby refer to this feature as *monotonic performance*. The denoisers used for the proposed method are assumed to possess this feature. Moreover, in order to avoid algorithm divergence, it is also assumed that the filters do not increase the signal power.

B. Linear Decomposition Using GEVD

Any linear transform of the observations $\mathbf{x}(t)$, defined in (3), can be represented as follows:

$$\begin{aligned}y(t) &= \mathbf{w}^T\mathbf{x}(t) = \mathbf{w}^T\mathbf{x}_s(t) + \mathbf{w}^T\mathbf{x}_n(t) \\ &\doteq y_s(t) + y_n(t).\end{aligned} \tag{8}$$

The performance of the later proposed method highly relies on the efficient use of prior information about the signal and noise subspaces in such linear transforms. In the following, we review several different cost functions that can appear in different applications, each leading to a different linear transform.

1) SNR Maximization: Suppose that we want to find a linear mixture of the input with a maximal SNR. The SNR of the linear mixture $y(t)$, defined in (8), can be calculated as follows:

$$\begin{aligned}\operatorname{SNR}(\mathbf{w}) &\doteq \frac{E_t\{y_s(t)^2\}}{E_t\{y_n(t)^2\}} \\ &= \frac{\mathbf{w}^T C_{\mathbf{x}_s}\mathbf{w}}{\mathbf{w}^T C_{\mathbf{x}_n}\mathbf{w}} = \frac{\mathbf{w}^T C_{\mathbf{x}}\mathbf{w}}{\mathbf{w}^T C_{\mathbf{x}_n}\mathbf{w}} - 1\end{aligned} \tag{9}$$

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.

where $C_{\mathbf{x}_s} \doteq E_t\{\mathbf{x}_s(t)\mathbf{x}_s(t)^T\}$ and $C_{\mathbf{x}_n} \doteq E_t\{\mathbf{x}_n(t)\mathbf{x}_n(t)^T\}$ are the covariance matrices of the signal and noise parts. The uncorrelatedness of the signal and noise parts has been used for the second part of this equation.

Following the explanations in Section II, the maximum value of the SNR is achieved by GEVD of the matrix pair $(C_{\mathbf{x}_s}, C_{\mathbf{x}_n})$, or equivalently, the GEVD of the matrix pair $(C_{\mathbf{x}}, C_{\mathbf{x}_n})$. This approach is, therefore, applicable for the cases that $C_{\mathbf{x}_n}$ or $C_{\mathbf{x}_s}$ are know or can be estimated.

2) Nonstationarity Maximization: Suppose that the signal (or noise) parts of the observations have a burst-like or sparse behavior, i.e., they only appear over certain periods of time and the active times of the desired signals are known or could be estimated from the dataset. This is the case for many practical applications such as mixtures of electroencephalogram (EEG) signals corrupted by eye movements or blink artifacts [15], EEG evoked responses to ocular or audio stimuli, EEG signals corrupted by switching magnetic resonance artifacts in fMRI experiments [16], or other possible applications in telecommunications. In all of these cases, the observations are nonstationary mixtures of signal and noise. We can, therefore, define a cost function that accounts for this nonstationarity

$$\zeta(\mathbf{w}) \doteq \frac{E_\theta\{y(\theta)^2\}}{E_t\{y(t)^2\}} = \frac{\mathbf{w}^T \tilde{C}_{\mathbf{x}} \mathbf{w}}{\mathbf{w}^T C_{\mathbf{x}} \mathbf{w}} \tag{10}$$

where θ is the active time epochs of the bursty signal (or more generally the time instants with a known nonstationarity), and $\tilde{C}_{\mathbf{x}} \doteq E_\theta\{\mathbf{x}(\theta)\mathbf{x}(\theta)^T\}$ is the covariance matrix of the observations over the periods of nonstationary. This idea is rather similar to the notion of Common Spatial Patterns (CSP) [17]. By maximizing (10), one can find the components that have a maximal contribution in the energy of the burst epochs (the numerator), while having the least contribution in the overall signal energy (the denominator). Following (2), the maximum value of $\zeta(\mathbf{w})$ is achieved by GEVD of the matrix pair $(\tilde{C}_{\mathbf{x}}, C_{\mathbf{x}})$.

3) Spectral Contrast Maximization: The signal/noise separability may be achieved in domains other than the time domain. For instance, suppose that we are interested in extracting band-limited signals; e.g., the extraction of *alpha* or *beta* (or both) rhythms from noisy EEG recordings. In this case, the linear mixture defined in (8), can be transformed into the frequency domain

$$Y(f) \doteq \mathcal{F}\{y(t)\} = \mathbf{w}^T \mathcal{F}\{\mathbf{x}(t)\} = \mathbf{w}^T \mathbf{X}(f)$$

where $\mathcal{F}\{\cdot\}$ represents the Fourier transform. We can now define the contrast function as follows:

$$\sigma(\mathbf{w}) \doteq \frac{E_\nu\{|Y(\nu)|^2\}}{E_f\{|Y(f)|^2\}} = \frac{\mathbf{w}^T S_{\mathbf{x}} \mathbf{w}}{\mathbf{w}^T C_{\mathbf{x}} \mathbf{w}} \tag{11}$$

where ν is the frequency band of interest, f is the whole frequency axis (or the Nyquist band, in the discrete case), and $S_{\mathbf{x}} \doteq E_\nu\{\mathbf{X}(\nu)\mathbf{X}(\nu)^H\}$ is the cross-spectrum of the observation vector averaged over the bandwidth of interest. In the denominator of the last part of (11), we have used the Parseval's relation [18]. In order to have a real symmetric matrix $S_{\mathbf{x}}$, the summation $E_\nu\{\cdot\}$ is taken over both positive and negative values of ν (or symmetric frequencies with respect to the Nyquist frequency, in the discrete case), over the frequency bands of interest. This condition guarantees the realness of $\sigma(\mathbf{w})$. For this cost function, the maximal spectral contrast is achieved by GEVD of the matrix pair $(S_{\mathbf{x}}, C_{\mathbf{x}})$.

Note that here, the objective is to transform the signals into a domain in which the signal and noise are better separated. Therefore, any other transform that preserves the linearity, such as wavelet transforms, may also be used in this procedure.

4) Periodicity Maximization: The desired or undesired parts of the observations can have a periodic or quasi-periodic structure. This is the case in applications in which a periodic or quasi-periodic signal is corrupted by noise, such as mixtures of electrocardiogram (ECG) or magnetocardiogram (MCG) signals and noise. In this case, one may seek for linear transforms that maximize a measure of periodicity, while keeping the signal energy bounded

$$\epsilon(\mathbf{w}) \doteq \frac{E_t\{y(t)y(t+\tau_t)\}}{E_t\{y(t)^2\}} = \frac{\mathbf{w}^T P_{\mathbf{x}} \mathbf{w}}{\mathbf{w}^T C_{\mathbf{x}} \mathbf{w}} \tag{12}$$

where τ_t is a (time-varying) period of the periodic or quasi-periodic signal and $P_{\mathbf{x}} \doteq E_t\{\mathbf{x}(t)\mathbf{x}(t+\tau_t)^T\}$. The AMUSE algorithm with a constant τ_t [19], and periodic component analysis with a time-varying time-lag τ_t [20], are examples of algorithms that maximize such a cost function. In either case, the maximum value of $\epsilon(\mathbf{w})$ is achieved by GEVD of the matrix pair $(P_{\mathbf{x}}, C_{\mathbf{x}})$.

The cost functions defined in (9)–(12), were all defined such that the problem of finding the optimal linear transform would lead to GEVD of the covariance matrix $C_{\mathbf{x}}$ and another positive definite matrix[2] that is designed and calculated from the observations using prior knowledge of the desired signals. For our problem of interest, linear transforms based on GEVD have several advantages over other possible linear decompositions.

- The components are uncorrelated and whitened, which means that there is no redundancy in the extracted components up to second-order statistics, as in principal component analysis (PCA).
- They use all the N^2 degrees of freedom of an $N \times N$ linear transform, unlike PCA or sphering transforms that only use $N(N-1)/2$ and $N(N+1)/2$ degrees of freedom, respectively [11, ch. 5].
- The components are ranked according to the statistical measure used in defining the second diagonalized matrix, unlike conventional ICA that can not rank the components.

C. Iterative Subspace Decomposition

We are now at a point where we can merge linear multichannel decomposition with single channel denoising schemes. The proposed method is presented in Fig. 1. It consists of an iterative procedure that applies a sequence of *linear decomposition* (*projection*), *denoising*, and *recomposition* (*back-projection*) to the input data. Following the explanations in the previous subsection, the linear decomposition unit is a GEVD procedure

[2]The positive definiteness of the matrices guarantees that the cost functions defined in (9)–(12) are always positive, which makes their maximization reasonable. Among the defined matrices, $P_{\mathbf{x}}$ is the only matrix that depending on the choice of τ_t can become nonpositive definite. However, for quasi-periodic signals, such as the ECG, a proper choice of τ_t usually leads to dominant positive eigenvalues of $P_{\mathbf{x}}$.

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.

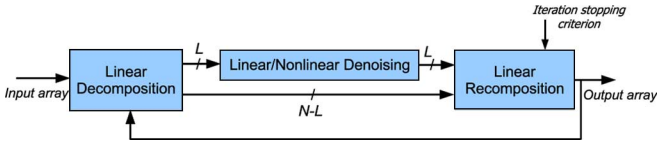


Fig. 1. General iterative subspace decomposition scheme.

based on our prior knowledge of the signal and noise subspaces. The outputs of this unit are ranked in descending (ascending) order of resemblance to the signal (noise) subspace. The objective of this block is to concentrate the components of the desired subspace in the first few components of its output. This unit is followed by a linear or nonlinear monotonic denoising filter that is applied to the first L components ($1 \leq L \ll N$) of the previous block. This filter can be a single channel filter applied to each channel separately, or a multichannel filter applied to the first L components together. As explained before, such denoising could have been directly applied to the original $\mathbf{x}(t)$, but by applying it after the linear decomposition in Fig. 1, we benefit from the improved signal quality of the first (or the first few) components extracted by the linear decomposition block. This improvement is the direct consequence of maximizing any of the cost functions defined in (9)–(12). For example, if the linear decomposition is based on SNR maximization (9), the SNR of the first extracted component is higher than any of the original input channels. Therefore, a monotonic denoiser applied after the linear decomposition is more effective than a denoiser that is applied before the decomposition.

Finally, the residual signals of the L denoised components and the other $N-L$ unchanged components are back-projected to the observation space, using the inverse of the linear decomposition matrix. In each iteration of the algorithm, some portion of the signal and noise subspaces are separated and the procedure is repeated until the output signals satisfy some predefined measure of signal/noise separability.

The overall algorithm can be expressed as follows.

Subspace Decomposition Algorithm:

1: $\mathbf{x}^{(0)}(t) = \mathbf{x}(t)$, $k = 0$
2: **repeat**
3: Calculate $C_{\mathbf{x}}^{(k)}$, the covariance matrix of $\mathbf{x}^{(k)}(t)$
4: Calculate $Q_{\mathbf{x}}^{(k)}$, the matrix containing the desired statistics of $\mathbf{x}^{(k)}(t)$
5: $W^{(k)} \leftarrow \mathrm{GEVD}(Q_{\mathbf{x}}^{(k)}, C_{\mathbf{x}}^{(k)})^T$
6: $A^{(k)} \doteq [\mathbf{a}_1^{(k)}, \ldots, \mathbf{a}_N^{(k)}] \leftarrow {W^{(k)}}^{-1}$
7: $\mathbf{y}^{(k)}(t) \leftarrow W^{(k)}\mathbf{x}^{(k)}(t)$
8: $\mathbf{s}^{(k)}(t) \doteq [s_1^{(k)}(t), \ldots, s_L^{(k)}(t)]^T \leftarrow \mathbf{G}(\mathbf{y}^{(k)}(t))$
9: $\mathbf{x}^{(k+1)}(t) \leftarrow \mathbf{x}^{(k)}(t) - \sum_{j=1}^{L} \mathbf{a}_j^{(k)} s_j^{(k)}(t)$
10: $c \leftarrow \mathcal{P}(\mathbf{x}^{(k+1)}(t))$
11: $k \leftarrow k+1$
12: **until** $c \leq th$

In this algorithm, $Q_{\mathbf{x}}^{(k)}$ is the matrix containing the desired statistics of $\mathbf{x}^{(k)}(t)$, which depending on the separation criterion can represent any of the matrices $C_{\mathbf{x}_n}$, $\tilde{C}_{\mathbf{x}}$, $S_{\mathbf{x}}$, or $P_{\mathbf{x}}$ defined in (9)–(12) (or other possible measures). The index k represents the kth iteration, $W^{(k)}$ is the transpose of the decomposition matrix found by GEVD,[3] $\mathbf{x}^{(k+1)}(t)$ is the output of each iteration, $\mathbf{G}(\cdot)$ is the denoising function for removing (keeping) the undesired (desired) components applied to the first L channels of $\mathbf{y}^{(k)}(t)$, $s_j^{(k)}(t)$ is the output of the denoising block in channel j, $\mathbf{a}_j^{(k)}$ is the jth column vectors of $A^{(k)}$, $\mathcal{P}(\cdot)$ is a measure of the desired subspace removal used as a stopping criterion, and th is a predefined threshold. The threshold and stopping criterion are rather subjective. In Section V, examples of such measures are presented.

The output of each iteration of the algorithm can be represented in the following compact form:

$$\mathbf{x}^{(k)}(t) = \mathbf{x}(t) - \sum_{i=0}^{k-1} \sum_{j=1}^{L} \mathbf{a}_j^{(i)} s_j^{(i)}(t), \quad (k > 0). \tag{13}$$

The total number of iterations depends on the number of dimentions of the target subspace and the denoising strength. On the other hand, the choice of L is rather *ad hoc* and a compromise between the accuracy of denoising and the number of iterations. Due to the ranking property of the decomposition block, for high dimentional data, one can use a larger L to speed-up the convergence of the algorithm, while for low-dimensional and degenerate data, one can use $L = 1$ to avoid losing the delicate components.

It should be noted that since the strongest components of the desired (undesired) subspace are removed from $\mathbf{x}(t)$ in each iteration, the weaker components that were dominated by the stronger ones are found in later iterations. Therefore, the covariance matrices required for the GEVD procedure of the proposed method are recalculated in every iteration from the output signals of the previous iteration. Moreover, the separation criterion used for designing the linear projector and the denoising block of Fig. 1 do not need to be the same in all iterations. For instance, for EEG signals, if we use the spectral contrast function defined in (11), one iteration of the algorithm can be designed to extract the *alpha* rhythms, while another iteration extracts the *beta* rhythms of the EEG. Similar examples are presented in the following for the extraction of maternal and fetal ECG mixtures.

One of the advantages of the proposed method is that we can remove the desired (or undesired) subspaces without losing the dimensionality of the recordings. This is due to the intermediate denoising step that is "breaking" the linearity of the transform. This is an important issue for noisy mixtures recorded from a few number of channels.

As an iterative algorithm, the study of convergence and robustness to changes in parameters is a major concern. Due to the data-driven nature and flexibility of the proposed algorithm, the proof of convergence in its general case is application-dependent. In the Appendix , sufficient conditions are presented that guarantee the local stability and convergence of the signal norms in each iteration. These conditions place constraints on the eigenvalues (and, therefore, structure) of the input data and the output of the denoiser in each iteration.

[3]Using singular value decomposition, it can be shown that if $C_{\mathbf{x}}^{(k)}$ is nonsingular, $W^{(k)}$ is also nonsingular and has an inverse (c.f. [21, sec. 8.7]).

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.

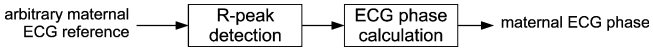


Fig. 2. Maternal phase calculation procedure.

V. Application in Fetal ECG Extraction

Fetal ECG (fECG) recorded from the maternal abdomen are heavily contaminated with maternal ECG (mECG), which depending on the gestational age and electrode locations, can be up to twenty times stronger than the fetal components [22], [23]. Independent subspace analysis (ISA) and other ICA-based methods are the most common approaches for fECG extraction. However, these methods have some limitations: although the maternal and fECGs form two independent subspaces and ISA tends to separate these two subspaces, in presence of noise or with special lead configurations, the subspaces of the fetal and maternal cardiac signals are not fully separated by linear ICA. This usually results in fetal signals that are still contaminated by mECG, or in fetal components removed with the maternal signals. We should add to this the problem of distributedness of the cardiac sources and the ambiguity in the number of dominant dimensions corresponding to the maternal and fetal subspaces [24, Ch. 6].

In the following, the proposed deflation procedure is used for the removal of mECG interference from fetal recordings. Due to the quasi-periodic structure of the ECG and according to criterion (12), the linear step of the algorithm is designed according to the periodic component analysis (πCA) procedure developed in [20], and for the denoising step we use the Kalman filtering framework presented in [25]. The method is applied to both simulated and real ECG recordings. The block diagram of the algorithm is depicted in Figs. 2–4.

A. Simulated Data

1) Dataset: Due to the multidimensional nature of the ECG, a realistic ECG model with multiple dimensions is required. Here, we use a multichannel ECG and noise generator similar to the one proposed in [26]. The model is as follows:

$$\begin{aligned}\mathbf{x}(t) &= \alpha H_m \mathbf{s}_m(t) + H_f \mathbf{s}_f(t) + \beta H_\eta \mathbf{v}(t) + \gamma \mathbf{n}(t) \\ &\doteq \alpha \mathbf{x}_m(t) + \mathbf{x}_f(t) + \beta \boldsymbol{\eta}(t) + \gamma \mathbf{n}(t) \quad (14)\end{aligned}$$

where $\mathbf{s}_m(t)$ and $\mathbf{s}_f(t)$ are 3-D sources representing the maternal and fetal cardiac components[4], $\mathbf{v}(t)$ is structured low-rank noise representing respiration or electrode movement artifacts, $\mathbf{n}(t)$ is full-rank noise representing typical measurement noise, and H_m, H_f, and H_η are the volume conduction transfer matrices for the mother, fetus, and structured noises, respectively. In this model, the maternal signal $\mathbf{x}_m(t)$ is assumed as *interference*, while $\boldsymbol{\eta}(t)$ and $\mathbf{n}(t)$ are assumed as *noises* for the fetal signal $\mathbf{x}_f(t)$. Therefore, the parameters α, β and γ control the signal-to-interference ratio (SIR) and signal-to-noise ratio (SNR) of the fetal ECG.

For this simulation, three independent leads from 24 subjects of the Physikalisch-Technische Bundesanstalt diagnostic ECG database (PTBDB) were used [28]. This database has a sampling rate of 1000 Hz. The signals were further preprocessed for baseline wander removal and bandpass filtered between 0.7 and 150 Hz. For each trial, three channels of the preprocessed data were randomly selected among the 24 subjects to represent $\mathbf{s}_m(t)$ and $\mathbf{s}_f(t)$. The fetal heart beat is typically about twice as fast as the maternal heart beat. Hence, the signals representing $\mathbf{s}_f(t)$ were resampled to 500 Hz to mimic fECGs with higher heart beats. All simulations were performed over 10-s data segments.

The full-rank noise $\mathbf{n}(t)$ was considered as white Gaussian noise, while for the structured noise $\mathbf{v}(t)$, real baseline wander, muscle artifacts, and electrode movement signals from the Noise Stress Test Database (NSTDB) [29], [30] were randomly selected and resampled to $f_s = 1000$ Hz.

In order to have eight channel observations, H_m, H_f, and H_η are chosen as random 8×3 matrices having specific angles between their column subspaces. To control the angle between the maternal and fetal subspaces, we calculate the *principal angles* (PA) between these two subspaces, denoted as follows:

$$\theta_{mf} = \mathrm{PA}(H_m, H_f).$$

The procedure of calculating the PA is explained in [21, ch. 12]. The signal-to-interference ratio (SIR), signal-to-noise ratio ($\mathrm{SNR}_n, \mathrm{SNR}_\eta$), and the signal-to-interference plus noise ratio (SINR), are calculated as measures of fetal signal quality *before* denoising

$$\mathrm{SIR} \doteq \frac{E_{t,i}\{x_{f_i}(t)^2\}}{\alpha^2 E_{t,i}\{x_{m_i}(t)^2\}} \quad (15)$$

$$\mathrm{SNR}_\eta \doteq \frac{E_{t,i}\{x_{f_i}(t)^2\}}{\beta^2 E_{t,i}\{\eta_i(t)^2\}} \quad (16)$$

$$\mathrm{SNR}_n \doteq \frac{E_{t,i}\{x_{f_i}(t)^2\}}{\gamma^2 E_{t,i}\{n_i(t)^2\}} \quad (17)$$

$$\mathrm{SINR} \doteq \frac{E_{t,i}\{x_{f_i}(t)^2\}}{E_{t,i}\{[\alpha x_{m_i}(t) + \beta \eta_i(t) + \gamma n_i(t)]^2\}} \quad (18)$$

where $x_{f_i}(t)$, $x_{m_i}(t)$, $\eta_i(t)$, and $n_i(t)$, are respectively the entries of $\mathbf{x}_f(t)$, $\mathbf{x}_m(t)$, $\boldsymbol{\eta}(t)$, and $\mathbf{n}(t)$ in channel i, and $E_{t,i}\{\cdot\}$ represents averaging over time and channels. For this simulation, the parameters α, β, and γ were selected such that

$$\begin{aligned}\mathrm{SNR}_\eta &= \mathrm{SIR} + 10 \text{ dB} \\ \mathrm{SNR}_n &= \mathrm{SIR} + 20 \text{ dB}\end{aligned}$$

and the overall fetal input SINR was swept in the range of -25 dB to -5 dB, which are in the range of their actual values.[5] With these choices, the mECG is the dominant artifact. Due to the choice of signal and noise dimensions, it is clear that the resulting mixtures in (14) are degenerate.

2) Evaluation: The following three algorithms were used, for evaluation:

i) ISA: As a benchmark method, the Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm was applied

[4]The assumption of three dimensions for the maternal and fetal components is based on the far-field dipole approximation of cardiac potentials [27].

[5]This SINR range corresponds to typical amplitudes of real fetal ECG that range from 5% to 50% of the maternal ECG amplitude and commonly below 20% [22], [23].

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.

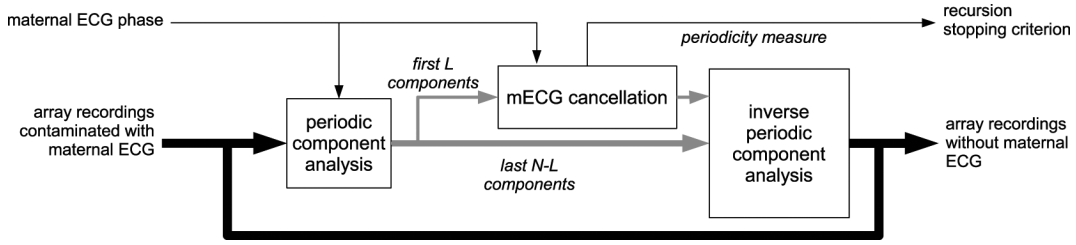


Fig. 3. Overall iterative procedure for maternal ECG cancellation. The details of the mECG cancellation block are depicted in Fig. 4.

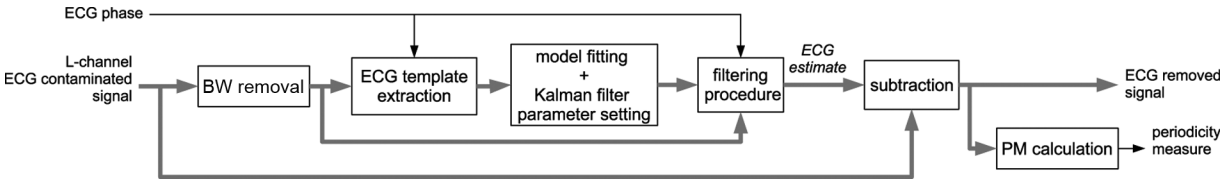


Fig. 4. L-channel Kalman filter for maternal ECG removal.

to the simulated data $\mathbf{x}(t)$ to achieve independent components $\mathbf{y}(t)$ [13]. Next, in order to find and regroup the dominant components that correspond to the fetal subspace, the optimal linear transform was found such that the following error be minimized:

$$e_i = E_t\{\big(y_i(t) - \mathbf{w}_i^T \mathbf{s}_f(t)\big)^2\} \tag{19}$$

where $y_i(t)$ is the ith entry of $\mathbf{y}(t)$ and $\mathbf{w}_i \in \mathbb{R}^3$ is a projection vector to be found. The optimal solution of this problem is

$$\mathbf{w}_i^* = E_t\{\mathbf{s}(t)\mathbf{s}_f(t)^T\}^{-1} E_t\{y_i(t)\mathbf{s}_f(t)\}. \tag{20}$$

Using this transform, the corresponding errors were calculated from (19) and ranked in ascending order. The first three channels of $\mathbf{y}(t)$ with the smallest value of e_i were selected as the fetal subspace.[6] These three components were back-projected to the original signal space using the inverse of the demixing matrix estimated by JADE.

ii) Deflation: The proposed deflation method was also applied to the synthetic signals for removing the mECG. The R-peaks of the mECG required for the πCA method were detected from an arbitrary channel which had dominant mECG artifacts [20]. The algorithm was repeated in four iterations over each dataset, which was empirically found to be sufficient for removing the 3-D maternal subspace used in the simulations.

iii) Deflation and ISA: The described deflation procedure only removes the maternal subspace. The deflation results may be improved by applying a postfiltering, for instance another ISA, to the signals after mECG removal. This ISA step is identical to the one explained in the first algorithm, except that after the deflation step, the mixture is "less-degenerate," in the sense that the mECG has already been removed and there are no more dimensions occupied by the mECG components in the ISA outputs. Note that in this postprocessing step, we again take the first three components corresponding to the fetal subspace.

For each value of the input SINR, the simulations were repeated ten times, each time with different signal and noise samples and for different values of θ_{mf}. Multichannel source separation techniques can be rather sensitive to the angles between the mixing subspaces. Therefore, all experiments were performed for two cases: 1) all principal angles smaller than 10° with an average $\bar{\theta}_{mf} = 5.9^\circ$, which we denote by $\theta_{mf} < 10^\circ$, and 2) all principal values greater than 60° with an average $\bar{\theta}_{mf} = 73.1^\circ$, which we denote by $\theta_{mf} > 60^\circ$. They were achieved by generating random matrices H_m and H_f, and altering them by a Givens rotation to achieve the desired principal angles [21].

These two cases represent very close and very far maternal-fetal subspaces, respectively. After applying the filter, the SINR improvement[7] of the filter outputs were calculated and compared for the different denoising methods. In Fig. 5(a) and (b), the results of this study can be seen for $\theta_{mf} < 10^\circ$ and $\theta_{mf} > 60^\circ$, respectively. We can see from Fig. 5(a) that the SINR improvement is significantly lower as compared with Fig. 5(b), indicating the fact that subspace separation becomes more difficult as the two subspaces become closer to each other. In this case, the deflation method outperforms ISA and the combination of deflation and ISA (denoted by deflation+ISA) outperforms both methods, except in very high SINR where the deflation method alone has outperformed the ISA-based method. For $\theta_{mf} < 10^\circ$, the standard deviation of ISA is higher than the other methods, indicating its lower robustness. On the other hand, in Fig. 5(b) we can see that for $\theta_{mf} > 60^\circ$ ISA and the combination of ISA and deflation have close performances, while the latter has been slightly more effective. Interestingly, in the highest input SINR $(-5$ dB$)$, the deflation method has again outperformed ISA.

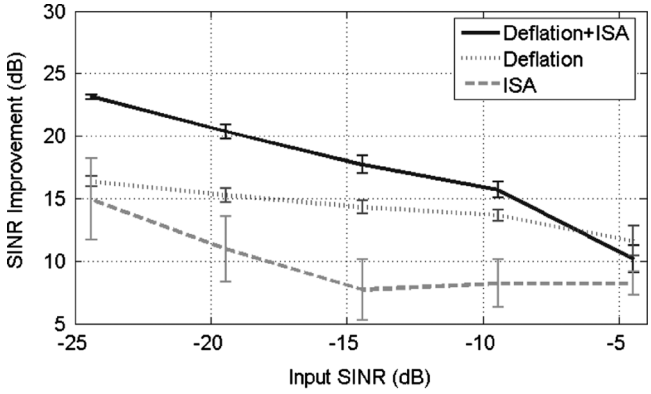
B. Real Data

1) Evaluation Measures: For real data, it is not possible to calculate the SNR as a measure of performance. Therefore, in order to have a quantitative measure, we propose to compare the amount of periodicity of the signal before and after filtering. Here, the idea is that an artifact-free signal should not contain any waveform that is synchronous with the mECG. Based on

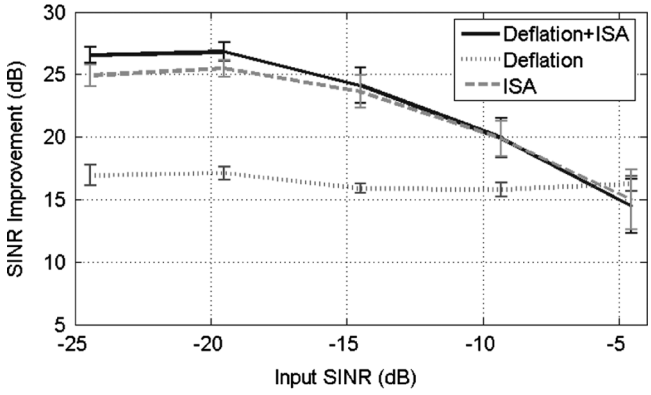
[6]Note that this procedure is somehow unrealistic for real ISA problems; since in reality we do not have access to the actual fetal subspace $\mathbf{s}_f(t)$. In that case, the fetal components are selected and regrouped by visual inspection or other classification criteria.

[7]The SNR improvement is the output SNR, in decibels (dB), minus the input SNR, in dB.

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.



(a)



(b)

Fig. 5. Mean and standard deviation bar of SINR improvements achieved by the three methods in different input SINRs for (a) $\theta_{mf} < 10^\circ$ (b) $\theta_{mf} > 60^\circ$.

this idea, we first find the mECG R-peaks and the *cardiac phase* $\phi(t) \in [-\pi, \pi]$, as explained in [20]. From $\phi(t)$, a time-varying lag τ_t is calculated

$$\tau_t \doteq \min\{\tau | \phi(t+\tau) = \phi(t), \tau > 0\}. \tag{21}$$

Accordingly, τ_t is the time distance between the sample $x(t)$ and its *dual sample* $x(t+\tau_t)$, which is the sample with the same phase value in the successive ECG beat.

From this, the following correlation coefficient is proposed as an overall *periodicity measure* (PM) for a signal $x(t)$

$$\mathrm{PM} = \frac{|E\{x(t)x(t+\tau_t)\}|}{\sqrt{E\{x(t)^2\}E\{x(t+\tau_t)^2\}}} \times 100\% \tag{22}$$

where $E\{\cdot\}$ represents averaging over the time index. From this definition $0 \leq \mathrm{PM} \leq 100\%$, where $\mathrm{PM} = 0$ indicates an aperiodic signal, and $\mathrm{PM} = 100\%$ indicates a fully periodic one (with respect to the target ECG signal). An effective filter should be able to remove any component that is temporally synchronous with the heartbeat and the samples $x(t)$ and $x(t+\tau_t)$ should become uncorrelated; resulting in a PM close to zero. Therefore, this measure is a suitable candidate for the stopping criterion $\mathcal{P}(\cdot)$, in the deflation algorithm.

The definition of PM may also be extended to vectorial data

$$\overline{\mathrm{PM}} = \frac{\mathrm{trace}\big(E\{\mathbf{x}(t)\mathbf{x}(t+\tau_t)^T\}\big)}{\mathrm{trace}\big(E\{\mathbf{x}(t)\mathbf{x}(t)^T\}\big)} \times 100\% \tag{23}$$

which can be used as an overall measure of periodicity for multichannel data. Note that the reduction of PM or $\overline{\mathrm{PM}}$ is a *necessary*, but not a *sufficient* measure of the filtering performance. In fact, the PM might be reduced, e.g., by an increase of the overall noise, without an improvement of the signal quality. Therefore, other evidence such as visual inspection of the resultant waveforms or a comparison of the signal spectra before and after filtering is always required besides this measure.

2) Results: We apply the proposed method to the DaISy fECG database for removing the mECG [31]. As seen in Fig. 6(a), this dataset consists of five abdominal and three thoracic channels, recorded from the abdomen and chest of a pregnant woman, with a sampling rate of 250 Hz. As a preprocessing step, the baseline wander of the data was removed, the maternal R-peaks were detected and the proposed method was applied on it in eight iterations. The results of the first five iterations are seen in Fig. 6(b)–(f).

The PM of all channels in different iterations are listed in Table I. It is seen that the proposed method has effectively reduced the mECG contaminants from the first to the last iteration. These values can be used for defining the appropriate threshold th for stopping the iterative procedure. From Table I, we can also see that PM reduction has been more significant in the abdominal channels which are of greater importance for fECG analysis (channels 1 to 5), as compared to the thoracic channels. This implies that in practice, depending on the application, the stopping threshold can be based upon a number of specific channels.

Note that after removing the maternal interference with the proposed technique, other source separation algorithms, such as ICA, may be applied to the residual signals to separate the fetal components from the background noise.

C. Numerical Convergence and Robustness

We present numerical results on the algorithm performance under some changes of parameters, for the aforementioned application. For this purpose, the real abdominal signals introduced in section V-B are manipulated as follows:

$$\mathbf{x}(t) = G[\mathbf{x}_0(t) + \Lambda \cdot \mathbf{v}(t)] \tag{24}$$

where $\mathbf{x}_0(t)$ is the original multichannel signal in Fig. 6(a), $\mathbf{v}(t)$ is Gaussian white noise, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ is a diagonal matrix for controlling the SNR of each channel, $G \in \mathbb{R}^{N \times N}$ is an arbitrary random matrix with a condition number smaller than 100 (to assure its well-condition), and $\mathbf{x}(t)$ is the modified signal. In this case, to show the generality of the method, a wavelet denoiser is used as an alternative for the Kalman denoising block of Fig. 4 for mECG removal, with $L = 1$, i.e., denoising only the first decomposed signal. Following [2], the parameters of the wavelet denoiser are as follows: Coiflet-5 mother wavelet, soft theresolding of wavelet coefficients based on the heuristic Stein's unbiased risk principle, without threshold rescaling [25], [32]. To show the performance of the method in different denoising schemes, the mentioned parameters are studied in two decomposition levels of one and five. With one level of decomposition, the denoising is very strong, but has the risk of removing useful non-mECG

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.

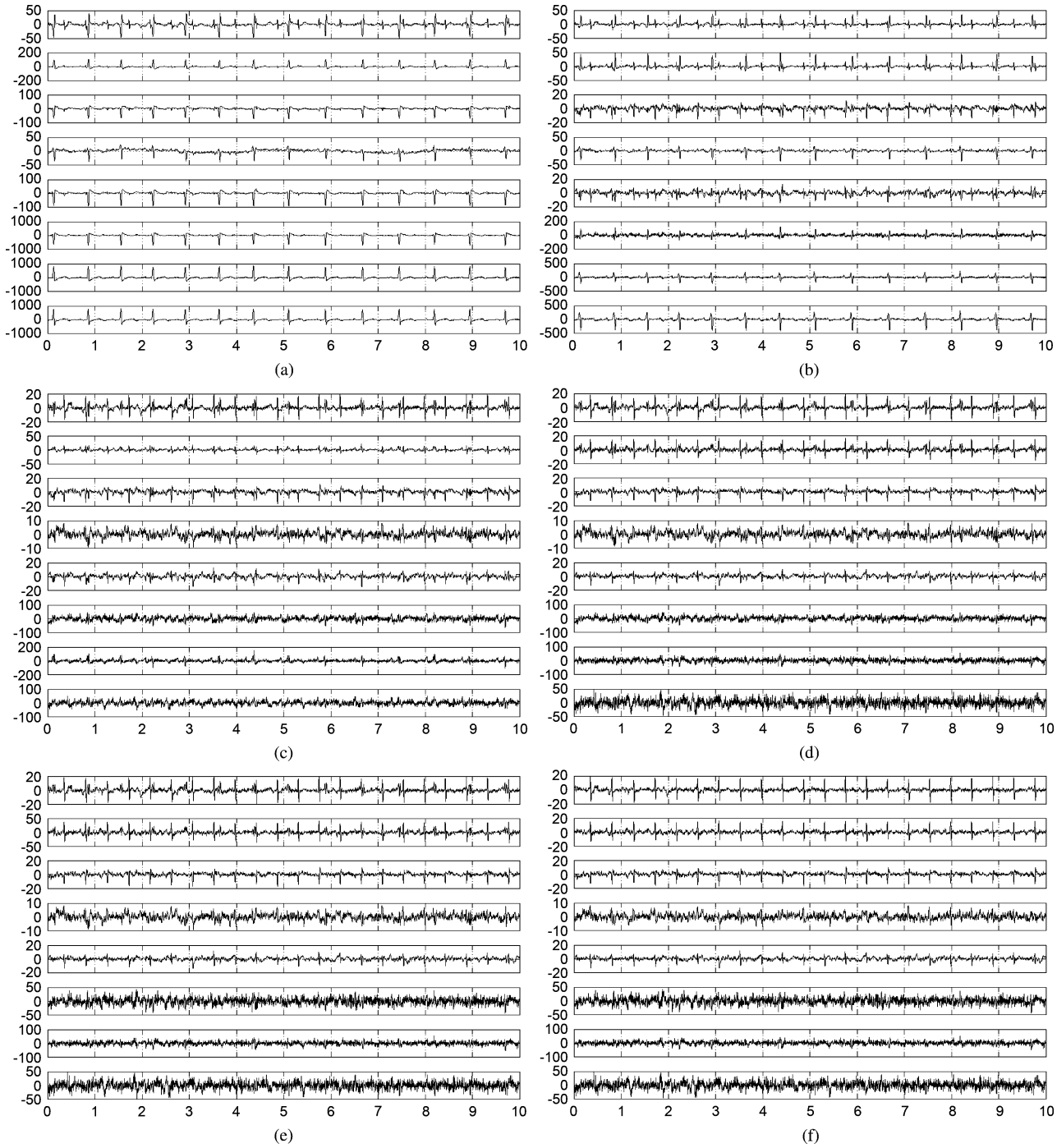


Fig. 6. Results of the proposed method on the DaISy fECG dataset in five iterations. Note that the vertical scales are not the same in all the subfigures. (a) Original data. (b) First iteration. (c) Second iteration. (d) Third iteration. (e) Fourth iteration. (f) Fifth iteration.

signals. With five levels of decomposition, the denoising is weaker, but preserves non-mECG traces that might exist in the first decomposed channels.

The proposed algorithm is run over $\mathbf{x}(t)$ in ten iterations. In each iteration, by changing Λ, the SNR is swept in the range of -10 to 30 dB in 10-dB steps and PM is calculated for each SNR. To study the consistency of the results, the whole procedure is repeated 1000 times, each time with a different noise instance $\mathbf{v}(t)$ and a random mixing matrix G. The average and standard deviation bars of the results over 1000 trials are depicted in Fig. 7 for one and five levels of wavelet decomposition. It is seen that PM monotonically decreases in successive iterations, but depending on the input SNR, it saturates after a number of iterations. Moreover, we can see that a wavelet denoiser based on single level decomposition has been more effective in decreasing the PM. However, as mentioned before, the decrease of PM is only a necessary condition for the evaluation of the method. In order, to evaluate the quality of mECG extraction, we compare the mECG waveforms extracted directly from $\mathbf{x}_0(t)$, with the signals extracted in different SNR from

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.

TABLE I
PERCENTAGE OF MATERNAL ECG PM IN DIFFERENT ITERATIONS FOR THE DAISY DATABASE

channel	1	2	3	4	5	6	7	8	
iteration									$\overline{\text{PM}}$
0 (original)	84	96	92	62	96	98	98	98	98
1	70	80	24	83	41	70	87	95	91
2	31	44	23	16	37	50	59	45	51
3	30	17	08	16	18	45	33	34	36
4	28	17	05	15	14	39	34	34	34
5	26	17	05	15	14	34	28	34	31
6	07	09	05	2	12	34	28	34	30
7	07	06	05	2	11	28	22	23	23
8	06	06	05	2	11	22	23	24	22

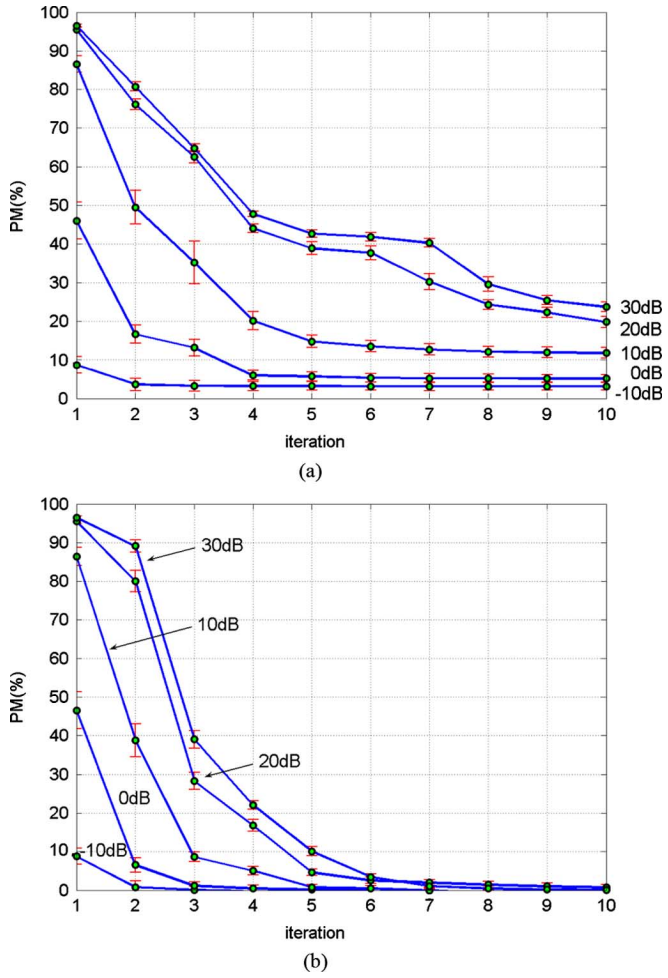


Fig. 7. Average and standard deviation bars of PM on multichannel fetal ECG data in ten iterations with randomized linear mixtures in different SNR; (a) and (b) correspond to five levels and one level of wavelet denoising, respectively. (a) Five-level wavelet denoising. (b) One-level wavelet denoising.

the manipulated signal $\mathbf{x}(t)$. The average and standard deviation bars of the mean-square error (MSE) between the directly extracted mECG and the ones extracted from the manipulated signal can be seen in in Fig. 8, for one and five levels of wavelet decomposition. It can be seen that the MSEs converge to their final values in less than ten iterations. We can also see that the MSE is higher in lower SNRs, which is due to the additive noise in (24) that decreases the input SNR and, therefore, increases the MSE. Moreover, comparing Figs. 7 and 8, one can notice that one-level wavelet decomposition has been faster and more

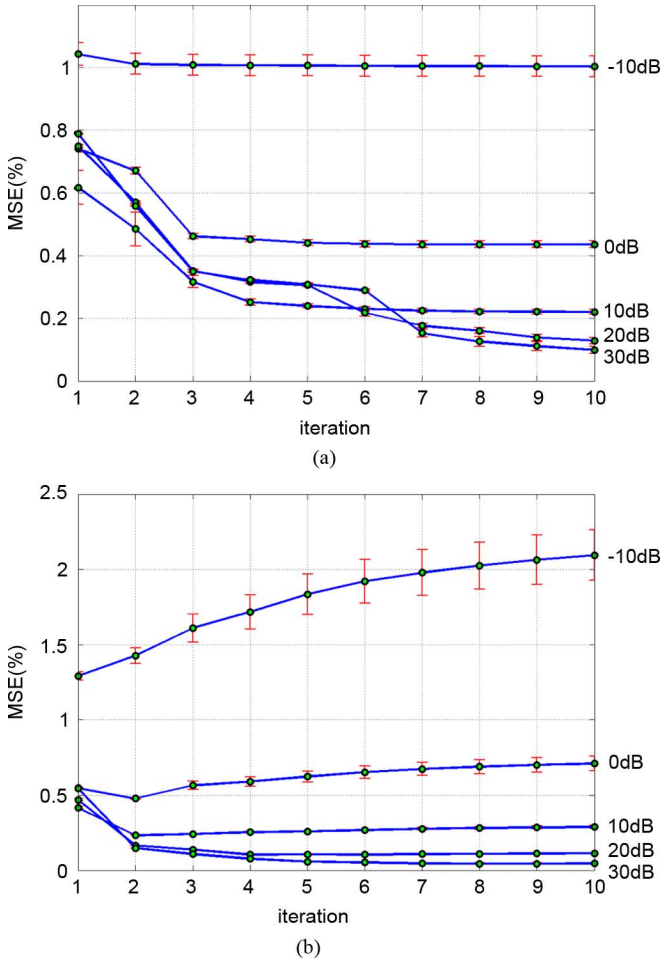


Fig. 8. Average and standard deviation bars of MSE on multichannel fetal ECG data in ten iterations with randomized linear mixtures in different SNR; (a) and (b) correspond to five levels and one level of wavelet denoising, respectively. (a) Five-level wavelet denoising. (b) One-level wavelet denoising.

effective in PM reduction as compared with the five-level decomposition; however, the MSE of one-level decomposition is worse. This shows that in each iteration of the algorithm, there is a compromise between the amount of denoising and the residual components.

The number of iterations required for convergence depends on the dimensions of the estimated subspace. For ECG signals, previous studies have shown that they have between four to eight dominant dimensions [24, Ch. 6]. Therefore, assuming that each iteration of the algorithm almost extracts one of these dimensions, it is reasonable that the algorithm converges within the same number of iterations. This observation is highlighted in the results of Figs. 7 and 8.

VI. APPLICATION IN DIAPHRAGMATIC SIGNAL EXTRACTION

As a second case study, the proposed method is applied to diaphragmatic electromyogram (EMG) signals recorded from a set of *intraesophageal* electrodes. A typical segment of this signal can be seen in Fig. 9(a).[8] The objective of this study

[8]The EMG for this study has been kindly provided by Dr. Vincent Vigneron from the Laboratory of Informatique, Biologie Intégrative et Systèmes Complexes (IBISC), CNRS FRE 2873, Evry, France.

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.

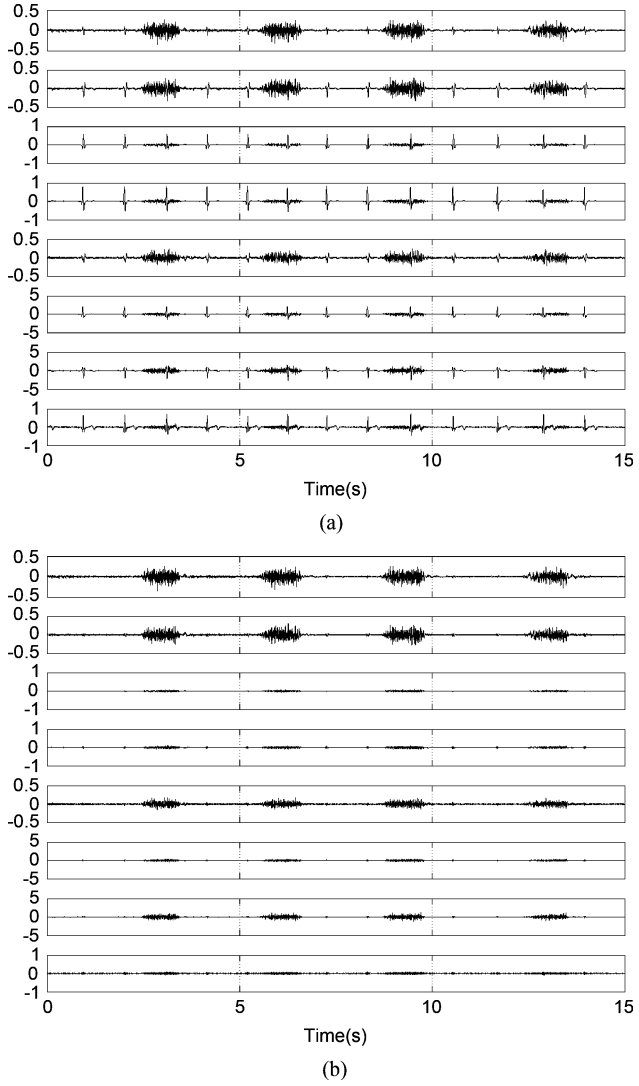


Fig. 9. Typical intraesophageal data (a) before and (b) after ECG artifact removal. (a) Original channels. (b) Eighth iteration.

was to measure diaphragmatic EMG, which are characterized by low amplitude noise-like bursts synchronous with the respiration. The exact detection of the beginning and ending points of the EMG burst are widely used in respiratory studies. However, since the intraesophageal electrode is rather close to the heart, the ECG dominates the desired EMG. Moreover, due to the distributedness of the diaphragm, conventional source separation techniques have limited performance in removing the cardiac interference or require additional ECG references for adaptively removing the interference [33], [34].

The procedure of cardiac interference cancellation is similar to the procedure used in Section V, i.e., πCA is used in the linear decomposition stage of the deflation algorithm and wavelet denoising is used for the denoising block. The parameters of the wavelet denoiser are as follows: Coiflet-5 mother wavelet, four-level wavelet decomposition, soft theresolding of wavelet coefficients based on the heuristic Stein's unbiased risk principal, with multiplicative threshold rescaling. The results of this algorithm can be seen in Fig. 9(b). We can see that the ECG has been effectively removed from the EMG.

VII. Summary and Conclusion

In this paper, a general framework was presented for the separation of a desired signal subspace of arbitrary dimensions from noisy, and possibly degenerate, multichannel mixtures of signal and noise. An important advantage of this method is that it can separate the subspaces without losing the dimensions of the signals, which is an important issue for noisy signals recorded from a few number of channels. Two examples of the applications of this method were presented. Nevertheless, the method is very general and may be adapted to other applications from different contexts.

Many of the existing source separation methods are based on a linear model, which is a simple, but constraining, model. As compared to these methods, the advantage of the proposed method is that although it benefits from a linear model for decomposition, but in the denoising step, it *breaks* this linearity in favor of extracting the desired sources. Nevertheless, it finally preserves the additive property of the subspaces in the back-projection step [c.f. (13)]. Therefore, we are able to extract sources that were otherwise un-extractable using a pure linear projection. However, the cost of this improved performance is the need for *a priori* information about the signal and noise subspaces. The performance of the deflation method highly depends on the performance of its decomposition and denoising steps and as we have seen in the studied applications, these steps should be in accord with one another to improve the signal quality. The assumption of monotonic performance of the denoising procedure described in Section IV-A intuitively guarantees that the overall denoising procedure does not degrade the signal. This point has also been verified in the studied applications, but requires theoretical justification in future works.

A prerequisite for the proposed technique is to formulate the features of the desired subspaces in terms of matrices that are jointly diagonalized by GEVD. In Section IV-B, several measures were proposed for this purpose. In future studies, other measures such as sparsity (in various domains), scale, probability density, etc., can also be considered.

The numerical convergence and robustness of the algorithm were studied for the presented case study. In future works, questions concerning the stability, convergence, robustness, and performance bounds should be studied from a theoretical perspective, for linear and nonlinear mixtures. For these studies, the framework presented in the Appendix and ideas from *operator theory* and *oblique projections* can be used as promising mathematical frameworks [12], [35], [36].

Appendix
Study of Convergence

Following the definitions in Section IV-C, let W, $\mathbf{G}(\cdot)$, and W^{-1}, respectively denote the *projection*, *denoising*, and *back-projection* operators of the kth iteration (the superscript k is dropped for notation simplicity). $\mathbf{y}(t) = W\mathbf{x}^{(k)}(t)$ denotes the signals after GEVD, which are uncorrelated, i.e., $C_y \doteq E_t\{\mathbf{y}(t)\mathbf{y}(t)^T\} = I$. The objective is to find conditions that starting from the original observations $\mathbf{x}(t)$, guarantee the convergence (local stability) of the proposed algorithm as the number of iterations approaches infinity. As a prerequisite,

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.

W, and, therefore, $C_{\mathbf{x}}^{(k)}$, should be nonsingular. A sufficient condition for convergence is to show that the Frobenius norm of the data is not increased in each iteration. Using the aforementioned definitions, we have

$$\begin{aligned}\Delta_k &\doteq \|\mathbf{x}^{(k)}(t)\|_F^2 - \|\mathbf{x}^{(k+1)}(t)\|_F^2 \\ &= \mathrm{tr}[\mathbf{x}^{(k)}(t)\mathbf{x}^{(k)}(t)^T] - \mathrm{tr}[\mathbf{x}^{(k+1)}(t)\mathbf{x}^{(k+1)}(t)^T] \\ &= \mathrm{tr}[W^{-1}\mathbf{y}(t)\mathbf{y}(t)^T W^{-T} - W^{-1}\mathbf{G}(\mathbf{y}(t))\mathbf{G}(\mathbf{y}(t))^T W^{-T}] \\ &= \mathrm{tr}[W^{-1}\big(\mathbf{y}(t)\mathbf{y}(t)^T - \mathbf{G}(\mathbf{y}(t))\mathbf{G}(\mathbf{y}(t))^T\big)W^{-T}] \\ &= \mathrm{tr}[(WW^T)^{-1}\big(\mathbf{y}(t)\mathbf{y}(t)^T - \mathbf{G}(\mathbf{y}(t))\mathbf{G}(\mathbf{y}(t))^T\big)] \\ &= K\,\mathrm{tr}[(WW^T)^{-1}(I - C_G)] \end{aligned} \tag{25}$$

where K is the number of data samples and $C_G \doteq E_t\{\mathbf{G}(\mathbf{y}(t))\mathbf{G}(\mathbf{y}(t))^T\}$. A sufficient condition for convergence is to have $\Delta_k \geq 0$ in all iterations.

The matrices $I - C_G$ and $(WW^T)^{-1}$ are both real symmetric. It is straightforward to show that $\lambda_i(I - C_G) = 1 - \lambda_i(C_G)$ and $\lambda_i((WW^T)^{-1}) = \lambda_i(C_{\mathbf{x}}^{(k)})$, where $\lambda_i(\cdot)$ $(i = 1, \ldots, N)$ denotes the eigenvalues ranked in descending order. Using matrix trace inequalities [37], a lower bound for Δ_k is

$$\Delta_k \geq K \sum_{i=1}^{N} \lambda_i(C_{\mathbf{x}}^{(k)})\big(1 - \lambda_i(C_G)\big). \tag{26}$$

Apparently, $\Delta_k \geq 0$ is achieved when the positive terms in (26) dominate the negative ones. A special case that leads into $\Delta_k \geq 0$ is when the denoising operator $\mathbf{G}(\cdot)$, which is applied to the first L channels, does not increase their norms and preserves the uncorrelatedness of the channels; hence

$$C_G = \mathrm{diag}(\epsilon_1, \ldots, \epsilon_L, 1, \ldots, 1) \tag{27}$$

where $0 \leq \epsilon_i \leq 1$ $(i = 1, \ldots, L)$. This is a practically feasible assumption in many cases. Loosely speaking, due to the GEVD procedure of the algorithm, which decorrelates the statistics in $C_{\mathbf{x}}^{(k)}$ and $Q_{\mathbf{x}}^{(k)}$, the channels of $\mathbf{y}(t)$ are not only uncorrelated, but rather close to independence. Therefore, any functional applied to each channel of $\mathbf{y}(t)$ should preserve their uncorrelatedness (cf. [38, Ch. 6]).

Tighter lower bounds proposed for traces of matrix products [37], can be used to find better lower bounds for Δ_k and necessary conditions for the algorithm convergence.

ACKNOWLEDGMENT

The authors would like to thank Dr. V. Vigneron from the Laboratory of Informatique, Biologie Intégrative et Systèmes Complexes (IBISC), CNRS FRE 2873, Evry, France, for providing the diaphragmatic EMG recordings used in Section VI.

REFERENCES

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. New York: Dover, 1979.

[2] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, pp. 613–627, 1995.

[3] H. Van-Trees, *Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing*. New York: Wiley, 2002.

[4] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[5] A. Cichocki and S.-i. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. New York: Wiley, 2002.

[6] P. D. O'Grady, B. A. Pearlmutter, and S. T. Rickard, "Survey of sparse and non-sparse methods in source separation," *Int. J. Imag. Syst. Technol.*, vol. 15, no. 1, pp. 18–33, 2005.

[7] W. Y. Leonga, W. Liub, and D. P. Mandic, "Blind source extraction: Standard approaches and extensions to noisy and post-nonlinear mixing," *Neurocomputing*, vol. 71, pp. 2344–2355, 2008.

[8] Y. Li, S. Amari, A. Cichocki, D. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 423–437, Feb. 2006.

[9] O. Yılmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, Jul. 2004.

[10] R. Sameni, M. Shamsollahi, C. Jutten, and G. Clifford, "Extraction of Fetal Cardiac Signals," U.S. Patent 61/116,870, Nov. 21, 2008.

[11] G. Strang, *Linear Algebra and Its Applications*, 3rd ed. New York: Brooks/Cole, 1988.

[12] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. Signal Process.*, vol. 42, pp. 1413–1424, 1994.

[13] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proc. F*, vol. 140, pp. 362–370, 1993.

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

[15] C. Gouy-Pailler, R. Sameni, M. Congedo, and C. Jutten, "Iterative subspace decomposition for ocular artifact removal from EEG recordings," in *Proc. 8th Int. Conf. Independent Component*, Paraty, Brazil, 2009, pp. 419–426.

[16] L. Amini, R. Sameni, C. Jutten, G. Hossein-Zadeh, and H. Soltanian-Zadeh, "MR artifact reduction in the simultaneous acquisition of EEG and fMRI of epileptic patients," presented at the 16th European Signal Processing Conf., Lausanne, Switzerland, Aug. 25–29, 2008.

[17] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.

[18] A. V. Oppenheim, A. S. Willsky, and H. Nawab, *Signals and Systems*. Upper Saddle River, NJ: Prentice-Hall, 1997.

[19] L. Tong, R.-W. Liu, V. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, no. 5, pp. 499–509, May 1991.

[20] R. Sameni, C. Jutten, and M. B. Shamsollahi, "Multichannel electrocardiogram decomposition using periodic component analysis," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1935–1940, Aug. 2008.

[21] G. Golub and C. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[22] T. Oostendorp, "Modeling the Fetal ECG," Ph.D. dissertation, K. U. Nijmegen, Nijmegen, The Netherlands, 1989.

[23] R. A. Shepoval'nikov, A. P. Nemirko, A. N. Kalinichenko, and V. V. Abramchenko, "Investigation of time, amplitude, and frequency parameters of a direct fetal ECG signal during labor and delivery," *Pattern Recognit. Image Anal.*, vol. 16, no. 1, pp. 74–76, Jan. 2006.

[24] R. Sameni, "Extraction of Fetal Cardiac Signals from an Array of Maternal Abdominal Recordings," Ph.D. dissertation, Sharif Univ. Technology—Institut National Polytechnique de Grenoble, Grenoble, France, 2008.

[25] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford, "A nonlinear bayesian filtering framework for ECG Denoising," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2172–2185, Dec. 2007.

[26] R. Sameni, G. D. Clifford, C. Jutten, and M. B. Shamsollahi, "Multichannel ECG and noise modeling: Application to maternal and fetal ECG signals," *EURASIP J. Adv. Signal Process.*, vol. 2007, p. 14, 2007.

[27] J. A. Malmivuo and R. Plonsey, *Bioelectromagnetism, Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford, U.K.: Oxford Univ. Press, 1995.

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.

[28] The PTB Diagnostic ECG Database, Physikalisch-Technische Bundesanstalt (PTB) [Online]. Available: http://www.physionet.org/physiobank/database/ptbdb/
[29] G. Moody, W. Muldrow, and R. Mark, The MIT-BIH Noise Stress Test Database [Online]. Available: http://www.physionet.org/physiobank/database/nstdb/.
[30] G. Moody, W. Muldrow, and R. Mark, "A noise stress test for arrhythmia detectors," *Comput. Cardiol.*, pp. 381–384, 1984.
[31] B. De Moor, Database for the Identification of Systems (DaISy) 1997 [Online]. Available: http://homes.esat.kuleuven.be/smc/daisy/
[32] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi, Matlab®, Wavelet Toolbox User's guide version 3 2005 [Online]. Available: http://www.mathworks.com/access/helpdesk/help/toolbox/wavelet/
[33] G. Naik, D. Kumar, S. Arjunan, M. Palaniswami, and R. Begg, "Limitations and applications of ICA for surface electromyogram," in *Proc. IEEE 28th Annu. Int. Conf. Engineering in Medicine and Biology Society*, 3, 2006, pp. 5739–5742.
[34] Y. Deng, W. Wolf, R. Schnell, and U. Appel, "New aspects to event-synchronous cancellation of ECG interference: an application of the method in diaphragmatic EMG signals," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1177–1184, Sep. 2000.
[35] R. Y. Kadison and J. R. Ringrose, *Fundamentals of the Theory of Operator Algebras*. New York: Academic, 1983, vol. I-II.
[36] L. Scharf and M. McCloud, "Blind adaptation of zero forcing projections and oblique pseudo-inverses for subspace detection and estimation when interference dominates noise," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 2938–2946, Dec. 2002.
[37] J. B. Lasserre, "Tight bounds for the trace of a matrix product," *IEEE Trans. Autom. Control*, vol. 42, no. 4, pp. 578–581, Apr. 1997.
[38] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.



Reza Sameni (S'01–M'09) received the B.Sc. degree in electronics engineering from Shiraz University, Iran, the M.Sc. degree in bioelectrical engineering from the Sharif University of Technology, Iran, in 2000 and 2003, respectively, and a joint Ph.D. degree in signal processing and bioelectrical engineering from the Institut National Polytechnique de Grenoble (INPG), France, and the Sharif University of Technology, in 2008. His Ph.D. dissertation focused on the modeling, filtering, and analysis of fetal cardiac signals.

He is now an Assistant Professor and on faculty at the Department of Electrical and Computer Engineering, Shiraz University. His research interests include statistical signal processing and time-frequency analysis of biomedical recordings. He has also worked in industry on the design and implementation of digital electronics and software defined radio systems.



Christian Jutten (A'92–M'03–SM'06–F'08) received the Ph.D. and the Docteurès Sciences degrees from the Institut National Polytechnique of Grenoble, France, in 1981 and 1987, respectively.

He was an Associate Professor in the Electrical Engineering Department, Institut National Polytechnique of Grenoble, from 1982 to 1989. He was Visiting Professor with the Swiss Federal Polytechnic Institute, Lausanne, in 1989, before becoming Full Professor at the University Joseph Fourier of Grenoble, Sciences and Technologies Department: Polytech Grenoble. He is currently Associate Director of the Grenoble Images, Speech, Signal, and Control Laboratory (GIPSA, 300 people) and Head of the Department Images-Signal (DIS) of this laboratory. For 25 years, his research interests have been in blind source separation, independent component analysis, and learning in neural networks, including theoretical aspects (separability, source separation in nonlinear mixtures), applications in signal processing (biomedical, seismic, speech), and data analysis. He is the author or coauthor of more than 50 papers in international journals, four books, 18 invited papers, and 150 communications in international conferences. He was a scientific advisor for signal and images processing at the French Ministry of Research from 1996 to 1998 and for the French National Research Center from 2003 to 2006.

Dr. Jutten was Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (1994–1995) and co-organizer of the First International Conference on Blind Signal Separation and Independent Component Analysis (Aussois, France, January 1999). He was associate editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS from 1992 to 1994. He is a member of the technical committee "Blind signal Processing" of the IEEE Circuits and Systems society and of the technical committee "Machine Learning for signal Processing" of the IEEE Signal Processing society. He is a reviewer of major international journals (IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, IEEE TRANSACTIONS ON NEURAL NETWORKS, *Signal Processing, Neural Computation, Neurocomputing*, etc.) and conferences in signal processing and neural networks (ICASSP, ISCASS, EUSIPCO, IJCNN, ICA, ESANN, IWANN, etc.). He received the EURASIP best paper award in 1992 and Medal Blondel in 1997 from SEE (French Electrical Engineering Society) for his contributions in source separation and independent component analysis, and became an IEEE Fellow and a Senior Member of Institut Universitaire de France in 2008.



Mohammad B. Shamsollahi (M'02) was born in Qom, Iran, in 1965. He received the B.Sc. degree in electrical engineering from Tehran University, Tehran, Iran, in 1988, the M.Sc. degree in electrical engineering and telecommunications from the Sharif University of Technology, Tehran, in 1991, and the Ph.D. degree in electrical engineering and biomedical signal processing from the University of Rennes 1, Rennes, France, in 1997.

Currently, he is an Associate Professor with the Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran. His research interests include biomedical signal processing, brain computer interface, time-scale, and time-frequency signal processing.

Authorized licensed use limited to: MI2S - Universite Joseph Fourier. Downloaded on March 23,2010 at 13:32:05 EDT from IEEE Xplore. Restrictions apply.