HAL
open science

Spatio-Temporal Source Separation using Temporal Priors with Parameterized Uncertainties

Reza Sameni

To cite this version:

Reza Sameni. Spatio-Temporal Source Separation using Temporal Priors with Parameterized Uncertainties. 2016. hal-01382035

HAL Id: hal-01382035

https://hal.science/hal-01382035v1

Preprint submitted on 15 Oct 2016

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.

Spatio-Temporal Source Separation using Temporal Priors with Parameterized Uncertainties

Reza Sameni*
October 2016

***Abstract*—The objective of this progressive research is to develop a spatio-temporal source separation scheme using temporal priors with parametric uncertainties. The method has potential applications in single and multi-modal data analysis. It is shown that the proposed scheme can be jointly used with classical source separation techniques based on generalized eigenvalue decomposition (GEVD) and joint approximate diagonalization of eigenmatrices (JADE). As a case study, the proposed method is evaluated over noninvasive fetal electrocardiogram signals.**

I. Introduction

When dealing with multi-modal signals— such as the simultaneous recording of the phonocardiogram and electrocardiogram— a recurrent issue is how to deal with simultaneous recordings with a common (yet hidden) origin; but with totally different temporal representations. In a progressive research, we have been studying various source separation techniques, with potential application in multi-modal data analysis. Herein, we present preliminary results of this progressive research, which has led to the development of spatio-temporal filters, which are applicable to signals belonging to some general class of temporal structure; but with some parametric uncertainties in their temporal priors.

II. Problem definition

Consider N-channel discrete-time noisy observations $x_i(t)$ $(t = 1, \cdots, T)$ and a set of linear (and generally time-variant) filters $h_i(t,\tau)$ $(i = 1, \cdots N)$. Defining the filtered observed channels

$$y_i(t) = \langle h_i(t,\tau), x_i(\tau)\rangle_\tau = \sum_{\tau=1}^{T} h_i(t,\tau)x_i(\tau), \tag{1}$$

the objective is to find linear mixtures

$$z_m(t) = \sum_{i=1}^{N} w_{mi} y_i(t), \quad m = 1, \cdots M \tag{2}$$

with maximal total channel energy under some bounding conditions on the weight vector w_{ki} (as detailed below).

To start with, if we simplify the problem to fixed linear filter coefficient sets $h_i(t,\tau) = h(t,\tau)$ over all channels, the problem can be stated in matrix form as follows

$$\mathbf{Z} = \mathbf{WXH} \tag{3}$$

where

$$\mathbf{X} = \mathbf{S} + \boldsymbol{\eta} \tag{4}$$

is the matrix form of the noisy observations $\mathbf{X} \in \mathbb{R}^{N\times T}$, $\mathbf{S} \in \mathbb{R}^{N\times T}$ is the desired multi-channel signal, $\boldsymbol{\eta} \in \mathbb{R}^{N\times T}$ is additive noise, $\mathbf{Z} \in \mathbb{R}^{M\times T}$ is the filtered data, $\mathbf{H} \in \mathbb{R}^{T\times T}$ is the matrix form of $h(t,\tau)$ considered as a finite impulse response (FIR) *temporal filter*, and $\mathbf{W} \in \mathbb{R}^{M\times N}$ is the *spatial filter* (in fact each row of $\mathbf{W}$ is a spatial filter). The desired source signals $\mathbf{S}$ may generally be mixtures of some latent variables $\boldsymbol{\theta}$ (e.g., $\mathbf{S} = \mathbf{A}\boldsymbol{\theta}$ as in classical ICA); but we do not explicitly use this fact in the current study.

With these definitions, we seek the spatial filter, such that

$$\mathbf{W}^* = \arg\max_{\mathbf{W}} \frac{\mathrm{tr}(\mathbf{WXHH}^T\mathbf{X}^T\mathbf{W}^T)}{\mathrm{tr}(\mathbf{WXX}^T\mathbf{W}^T)} \tag{5}$$

which is equivalent to the joint diagonalization of the following two matrices using generalized eigenvalue decomposition (GEVD):

$$\mathbf{A} = \mathbf{XHH}^T\mathbf{X}^T, \quad \mathbf{B} = \mathbf{XX}^T \tag{6}$$

III. Known (Presumed) Temporal Structure

As a linear filter, $\mathbf{H}$ can generally have an arbitrary structure. For time-invariant filtering, the filter takes a column-wise Toeplitz form, obtained from a prototype impulse response vector $\mathbf{h} = [h_1, h_2, \cdots, h_L]^T$ $(L \leq T)$. The filter can be causal, anti-causal, or non-causal, depending on whether $\mathbf{H}$ is upper-triangular, lower-triangular, or non-triangular. The impulse response of the filter may have been obtained from a classical filter design procedure, or an optimal design scheme, such as a matched filter or Wiener filter (smoother) design. It can also be a linear map to a transform domain, such as the wavelet or Fourier domain. In the latter case, $\mathbf{H}$ takes the form of a DFT or DCT matrix.

In the more general case of nonlinear filters with different per-channel filter responses, (2) becomes

$$z_m(t) = \sum_{i=1}^{N} w_{mi} h_i(x_i(t)), \quad (m = 1, \cdots M) \tag{7}$$

(notice the difference from the classical *post-nonlinear* data model [1]). In this case, the diagonalized matrices defined in (6) can be replaced by a symmetric kernel[1]

$$\mathbf{K_X} = [\langle h_i(x_i(t)), h_j(x_j(t))\rangle_t]_{ij} \tag{8}$$

*R. Sameni is with the School of Electrical & Computer Engineering, Shiraz University, Shiraz, Iran. Email: reza.sameni@gmail.com, Web: www.sameni.info.

[1]The kernel symmetry assumption, implies translational symmetry between the correlation functions of the two filtered channels, which is generally achieved in LTI systems. The asymmetric part of the cross-correlation can be of interest for time-variant causal analysis, which is beyond the scope of the current study as it takes the eigenvalue decomposition to the complex domain (due to the asymmetry of the diagonalized matrices).

Depending on the choice of the temporal filter $h(\cdot)$, the problem reduces to classical time-domain separation methods such as AMUSE [2], TDSEP [3], [4], πCA [5], [6], and a few other time and transform domain source separation methods described in [7].

The overall problem is trivial with a known (or set of known) prior temporal filter(s)— either in the linear or kernelized form— using GEVD. But how can one solve the problem if the temporal structure is unknown? In the most general case of unknown temporal structure, the problem is ill-posed as we have more parameters than observations ($\mathbf{H}$ is a huge $T \times T$ unknown matrix). In the kernelized form, the problem reduces to *kernel learning* problem, which can be solved by prior assumptions, such as stochastic independence, on the desired sources. This approach is the basis of the well known kernel ICA [8] and similar techniques.

Herein, we study the special case of temporal priors, with parametric uncertainties in their temporal structure. Smoothness priors are studied as a special case for this framework. It is shown that the problem has a low computational complexity and outperforms the fully blind scenario (especially in presence of noise), as it makes the best of the temporal priors despite their uncertainties.

IV. Parametric Spatio-Temporal Filter Design

Let us suppose that we know some general properties regarding the temporal structure of the desired sources. For instance, one might know (or assume) that the signal follows an auto-regressive model of a certain order and spectra, or that the signal is smooth up to a certain order (for example by knowledge of its spectral decay rate [9, Ch. 2]), etc. At the same time, there can be some uncertainties in the signal model, such as the exact smoothness order or level of background noise (SNR). Based on these assumptions, we propose two spatio-temporal source separation schemes to overcome the uncertainties in the temporal structure.

A. Multiple smoothness parameters using JADE

In the first approach, we assume that we have multiple temporal priors regarding the sources, or a single prior with unknown set of parameters. In case of multiple temporal priors, one can have different temporal filters represented in matrix form as $\mathbf{H}_k$ $(k = 1, \cdots, K)$. The filters can be totally different (each using a different temporal prior) or may belong to the same class of filters; but with different set of parameters. In this case, the problem is to find the spatial filter that jointly (approximately) diagonalizes [10], [11], the matrices

$$\mathbf{A}_k = \mathbf{X}\mathbf{H}_k\mathbf{H}_k^T\mathbf{X}^T (k = 1, \cdots, K), \quad \mathbf{B} = \mathbf{X}\mathbf{X}^T \tag{9}$$

(the matrix $\mathbf{B}$ is only required in the diagonalization set, if we seek uncorrelatedness).

B. Iterative regularization factor estimation using GEVD)

In the second approach, we use the fact that the linear form assumed in (3), implies that the spatial and temporal filters are fully decoupled. In other words, the temporal filtering does not mix the channels in space and the spatial mixture does not mix the data in time and the order of applying the filters in time and space is not important. However, due to the nonlinear (quadratic) form of the cost function in (5), the two filters can enhance the performance of one other, in the sense that one can seek "a spatial filter that enhances some temporal structure and vice versa a temporal filter that improves the spatial structure". Based on this idea, we propose to use successive iterations of temporal and spatial filtering, each time improving the estimate of one another. The general idea is summarized in Algorithm 1. We can notice that the idea is generic and may be extended to other parametric spatio-temporal source separation schemes and for kernelized formulations. In the next section, we will study this iterative approach for the special case of signals with parameterized smoothness priors.

Algorithm 1 Iterative spatio-temporal source separation

Require: Multichannel signals $\mathbf{X} \in \mathbb{R}^{N\times T}$, with zero-mean (per-channel)
Require: Parameterized temporal filter matrix $\mathbf{H}(\theta) \in \mathbb{R}^{T\times T}$, with parameter set θ
Require: Initial temporal filter parameter set θ_0
Require: Number of iterations I

```
Calculate C_X = X X^T
for i = 1, ..., I do
    Calculate Y = X H(θ_{i-1})
    Calculate K_X = Y Y^T
    Find the spatial filter W, using GEVD on the matrix pair (K_X, C_X)
    Find Z = W X = [z_1, ..., z_N]^T
    Optimize the temporal filter parameter set θ_i over z_1, (or over all channels of Z_i in the kernelized form)
return Z and W
```

V. Case Study: Parametric Temporal Smoothness

As proof of concept, we study the problem of extracting smooth signals from multichannel mixtures of signal and noise. The order of signal smoothness and the background noise level are assumed to be unknown[2]. Some related studies in this field, which have used the slow temporal variations of a signal for source separation, include [12], [13].

Numerous methods have been developed for data smoothing in the literature. In this context, smoothing regularization (also known as Tikhonov regularization) is a popular technique based on constrained least squares estimation [14, p. 307]. Consider the second-order difference operator $\mathbf{d}_2 \triangleq [1, -2, 1]$

[2]The extension of this case study to signals with arbitrary spectral models is rather straightforward using, e.g., auto-regressive modeling; but beyond the scope of this work.

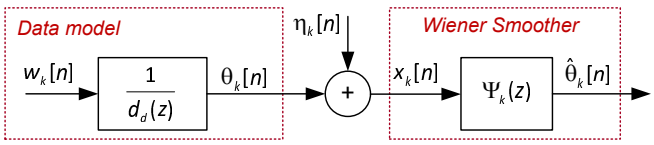

Fig. 1. An equivalent Wiener smoother scheme for Tikhonov regularization, adopted from [15].

and its Toeplitz matrix form

$$\mathbf{D}_2 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -2 & 1 & 0 & \ddots & \vdots \\ 1 & -2 & 1 & \ddots & \vdots \\ 0 & 1 & -2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ \vdots & \ddots & \ddots & \ddots & -2 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \tag{10}$$

The right multiplication of $\mathbf{D}_2$ in any matrix $\mathbf{X}$ containing length T signals as its rows is equivalent to convolving each row with $\mathbf{d}_2$, i.e., the second order difference of the signal. This definition can be extended to order-d difference operators $\mathbf{d}_d$ and $\mathbf{D}_d$ [15]. Using a constrained least squares optimization scheme, it has been shown that the following matrix acts as a smoothing operator

$$\mathbf{H}(d, \lambda) = (\mathbf{I} + \lambda \mathbf{D}_d \mathbf{D}_d^T)^{-1} \tag{11}$$

where d is the difference order and $\lambda \geq 0$ is the regularization parameter ($\lambda = 0$ for no smoothing) considered as an unknown parameter in this work.

In a recent study [15], it was shown that $\mathbf{H}(\cdot)$ is equivalent to a linear zero-phase forward backward Wiener smoother, with the following per-channel transfer function

$$\psi_d(z) = \frac{1}{1 + \lambda \mathbf{d}_d(z)\mathbf{d}_d(z^{-1})} = \frac{S_s(z)}{S_s(z) + S_n(z)} \tag{12}$$

which is in the form of an *optimal Wiener smoother* for a signal $s(t)$ having an auto-regressive all-pole model with a spectra denoted by $S_s(z)$, driven by input white (process) noise with variance ϵ^2, in presence of white observation noise $n(t)$ with a white spectra denoted by $S_n(z)$ and variance δ^2. Fig. 1, illustrates the analogy between Tikhonov regularization and optimal Wiener smoothing in the sample-wise formulation. With this interpretation, the regularization factor λ is the ratio between the observation and process noise variances, i.e.,

$$\lambda = \frac{\delta^2}{\epsilon^2} \tag{13}$$

In [15], it was shown that this smoothing scheme can be implemented as a forward-backward smoother for instance by using the *filtfilt* functions in Matlab, Octave or R.

In a blind or semi-blind source separation scheme, the smoothness order d and the noise variances δ^2 and ϵ^2 (and their ratio λ) can be unknown. Therefore, one should optimize the solution over the unknown parameter pair (d, λ).

Herein, we propose two approaches for optimized source separation over the unknown set (d, λ).

A. A JADE solution

In our first approach, we choose (by intuition) different combinations of integer values for d_k and positive real values λ_k in different orders of magnitude, to cover the possible smoothness orders and signal to noise ratios. Next, we form the following matrices:

$$\begin{aligned} &\mathbf{A}_k = \mathbf{X}\mathbf{H}(d_k, \lambda_k)\mathbf{H}(d_k, \lambda_k)^T \mathbf{X}^T \quad (k = 1, \cdots, K) \\ &\mathbf{B} = \mathbf{X}\mathbf{X}^T \end{aligned} \tag{14}$$

and use the joint approximate diagonalization of eigenmatrices (JADE) to simultaneously diagonalize these matrices (approximately). The joint diagonalization of these matrix set, accounts for various orders of signal smoothness and unknown noise levels.

B. An iterative GEVD solution

In our second approach, following Algorithm 1 detailed in Section IV-B, successive stages of temporal and spatial filtering are used. The only part of Algorithm 1, which requires customization for different applications is to optimize the parametric temporal filter in each iteration. For illustration we fix the temporal filter smoothness order to $d = 2$, which only leaves the regularization factor λ to be optimized in each iteration.

In other contexts, methods such as the *L-curve* have been proposed for finding the optimal regularization factor [14], [16]–[19], which also has interpretations in terms of maximum a posteriori (MAP) estimation [20].

In our case, using the notations of Algorithm 1, the L-curve is the plot of $\|\mathbf{X} - \mathbf{Y}\|$ versus $\|\mathbf{Y}\mathbf{D}_d\|$, as λ is swept from zero to infinity. The λ corresponding to the corner of the L-curve (in our case, the point at which $\|\mathbf{X} - \mathbf{Y}\| + \|\mathbf{Y}\mathbf{D}_d\|$ is minimized) is empirically chosen as the optimal regularization factor, which provides the best balance between the signal smoothness and minimum mean square error. This optimal λ is considered as the parameter for the temporal filter $\mathbf{H}(d, \lambda)$ used in the next iteration of the algorithm.

VI. Results

A potential domain of interest for the proposed methods is for extracting multichannel electrocardiogram (ECG) from background noise. In [6], a semi-blind source separation scheme was developed for extracting maternal and fetal ECG using pseudo-periodicity priors. This method has shown to be very effective in practice; however, it highly relies on the pseudo-periodicity of the ECG and its performance degrades for arrhythmic and highly irregular ECG beats. On the other hand, as noted in [15], due to the relatively high sampling frequency of contemporary ECG recording systems, the ECG (regular or irregular) is rather smooth as compared with the background noise. Therefore, "smoothness priors" can be used for designing ECG source separation schemes.

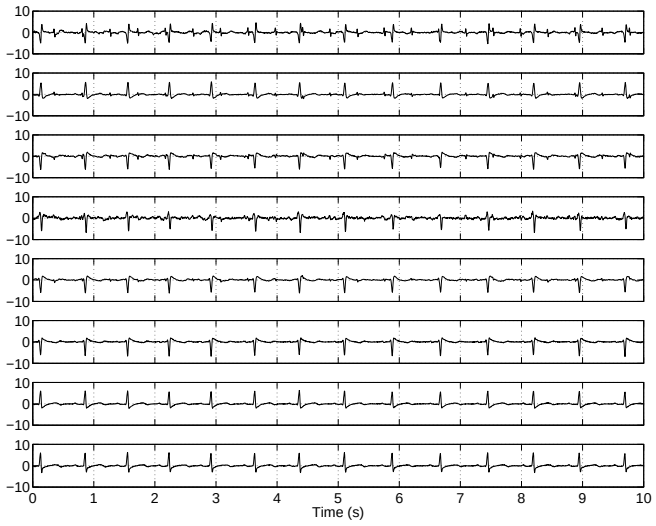

Fig. 2. The DaISy dataset consisting of five maternal abdominal and three thoracic channels [21]; adopted from [6].

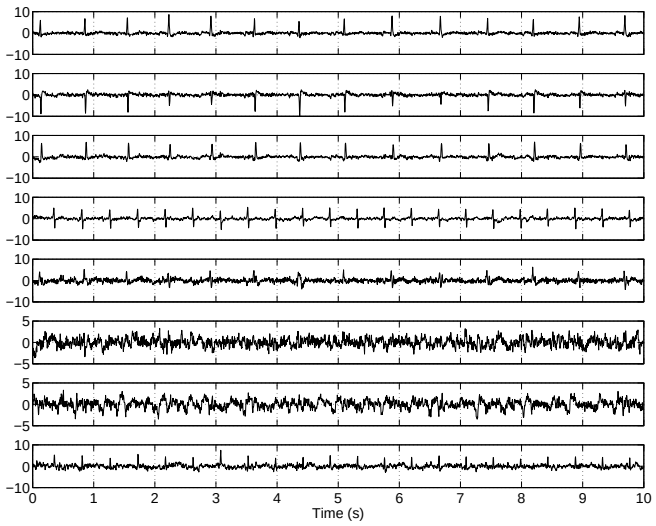

Fig. 3. Independent components extracted from the dataset of Fig. 2, using JADE. Notice that components 1, 2, 3, and 5 correspond to the maternal subspace and components 4 and 8 to the fetal subspace; adopted from [6].

A. Dataset

As proof of concept, we use the well-known DaISy fetal ECG dataset for illustration [21]. The dataset consists of five abdominal and three thoracic channels recorded from the abdomen and chest of a pregnant woman at a sampling rate of 250Hz. The eight channels of the dataset may be seen in Fig. 2.

We use independent subspace decomposition using the JADE algorithm as benchmark [22], [23]. The eight independent components extracted by this algorithm are depicted in Fig. 3.

To evaluate the hereby proposed method, we first use a single temporal smoothing matrix with $d = 2$ and $\lambda = 1$. The covariance of the original and smoothed data are jointly

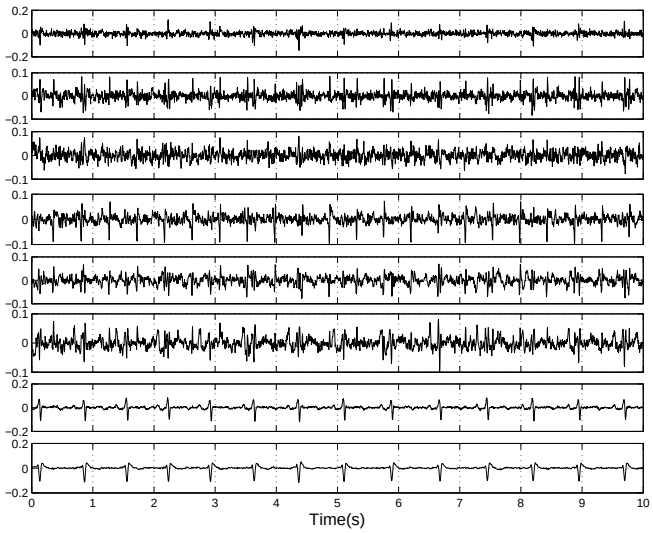

Fig. 4. Smooth components extracted from the dataset of Fig. 2, using GEVD on the covariance matrix and a smoothed version of the data using $d = 2$ and $\lambda = 1$. Notice that algorithm has extracted smooth components; but has not separated the fetal and maternal ECG.

diagonalized using GEVD. The results obtained from this method are shown in Fig. 4, accordingly although the extracted components are smooth, only one fetal ECG channel has been extracted (in channel 4), and the second fetal component (channel 5) is diluted by the maternal peaks. This was already anticipated as the method has only targetted smooth components and not the separation of the fetal and maternal ECGs.

In the next stage we consider a more blind scenario. We first pre-whiten the data and form the smoothing matrix using multiple regularization factors $\{0.008, 0.04, 0.2, 1, 5, 25, 125\}$. In addition, we use the pseudo-periodicity matrices corresponding to the mother and the fetus, as described in the πCA algorithm [6]. We therefore obtain nine covariance matrices, corresponding to pseudo-periodicity and temporal smoothness priors. The joint (approximate) diagonalization of these matrices [10], [11], accounts for pseudo-periodicity and an unknown noise level. The resulting signals are shown in Fig. 5. It is seen that this time the fetal and maternal components have been very well extracted.

VII. Conclusion and Future Work

In this work, a spatio-temporal source separation scheme was proposed for the extraction of sources with prior temporal structure having parametric uncertainties in their temporal model.

The proposed method can be extended from various aspects in future studies, including:

- The hereby developed spatio-temporal filter is based on second order temporal statistics, which was shown to be very effective for signal extraction from noise. The diagonalized matrices used in this study can be used together with higher order statistics used in the classical ICA algorithms such as JADE [24], to obtain simultaneous

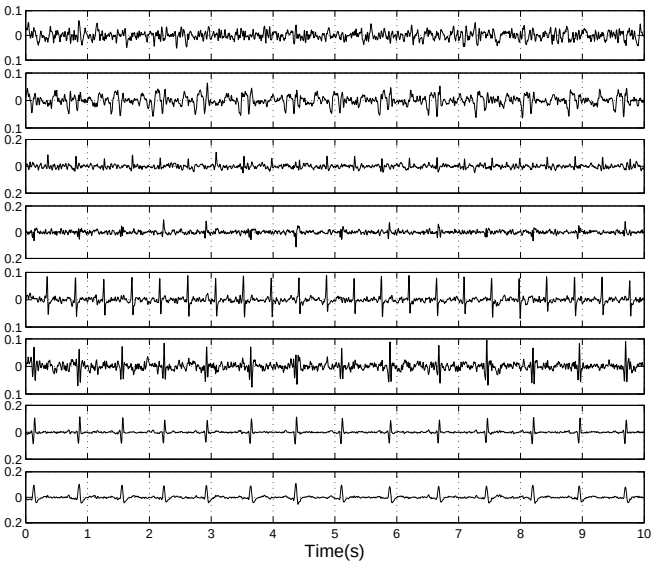

Fig. 5. Smooth components extracted from the dataset of Fig. 2, using JADE on the covariance matrix and several smoothed versions of the data using $d = 2$ and $\lambda = \{0.008, 0.04, 0.2, 1, 5, 25, 125\}$, plus two periodicity matrices obtained using πCA [6]. Notice that algorithm has extracted smooth components and separated the fetal and maternal ECG.

signal denoising (using spectral or temporal priors) and independent components.

- Combining the proposed method with the deflation algorithm proposed in [7] (for rank deficient data).
- Combining the method with generalized multi-view analysis (GMA) [25], for multiple modalities such as simultaneous ECG and PCG databases.
- One might be able to formulate the parameterized temporal filter using kernels with the same eigenstructures (sharing the same eigen-vectors) to guarantee that all matrices are exactly diagonalized using JADE.
- The problem of finding a closed form solution for the optimal temporal filter parameter should also be studied in future studies.

References

[1] A. Taleb and C. Jutten, "Source separation in post-nonlinear mixtures," *IEEE Transactions on signal Processing*, vol. 47, no. 10, pp. 2807–2820, 1999.

[2] L. Tong, V. Soon, Y. Huang, and R. Liu, "Amuse: a new blind identification algorithm," in *Circuits and Systems, 1990., IEEE International Symposium on*, May 1990, pp. 1784–1787 vol.3.

[3] A. Ziehe and K.-R. Müller, "Tdsepan efficient algorithm for blind separation using time structure," in *ICANN 98*. Springer, 1998, pp. 675–680.

[4] L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Physical review letters*, vol. 72, no. 23, p. 3634, 1994.

[5] L. K. Saul and J. B. Allen, "Periodic Component Analysis: An Eigenvalue Method for Representing Periodic Structure in Speech," in *NIPS*, 2000, pp. 807–813. [Online]. Available: http://www.cs.cmu.edu/Groups/NIPS/00papers-pub-on-web/SaulAllen.pdf

[6] R. Sameni, C. Jutten, and M. B. Shamsollahi, "Multichannel Electrocardiogram Decomposition using Periodic Component Analysis," *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 8, pp. 1935–1940, Aug 2008.

[7] ——, "A Deflation Procedure for Subspace Decomposition," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2363–2374, April 2010.

[8] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *Journal of machine learning research*, vol. 3, no. Jul, pp. 1–48, 2002.

[9] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. Elsevier Science, 2008.

[10] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Mat. Anal. Appl.*, vol. 17, no. 1, pp. 161–164, Jan. 1996.

[11] J.-F. Cardoso, "Perturbation of joint diagonalizers. ref# 94d027," Télécom Paris, Tech. Rep., 1994.

[12] T. Blaschke and L. Wiskott, "Independent slow feature analysis and nonlinear blind source separation," in *International Conference on Independent Component Analysis and Signal Separation*. Springer, 2004, pp. 742–749.

[13] T. Blaschke, T. Zito, and L. Wiskott, "Independent slow feature analysis and nonlinear blind source separation," *Neural computation*, vol. 19, no. 4, pp. 994–1021, 2007.

[14] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[15] R. Sameni, "Online Filtering Using Piecewise Smoothness Priors: Application to Normal and Abnormal Electrocardiogram Denoising," Mar. 2016, working paper or preprint. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01352281

[16] W. Gander, "Least squares with a quadratic constraint," *Numerische Mathematik*, vol. 36, no. 3, pp. 291–307, 1980.

[17] G. Golub and C. van Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1996.

[18] P. C. Hansen, *The L-curve and its use in the numerical treatment of inverse problems*. IMM, Department of Mathematical Modelling, Technical Universityof Denmark, 1999.

[19] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari, "Tikhonov regularization and the l-curve for large discrete ill-posed problems," *Journal of Computational and Applied Mathematics*, vol. 123, no. 12, pp. 423 – 446, 2000, numerical Analysis 2000. Vol. III: Linear Algebra. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377042700004143

[20] S. Oraintara, W. Karl, D. Castanon, and T. Nguyen, "A method for choosing the regularization parameter in generalized tikhonov regularized linear inverse problems," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 1, 2000, pp. 93–96 vol.1.

[21] B. De Moor, *Database for the Identification of Systems (DaISy)*, 1997. [Online]. Available: http://homes.esat.kuleuven.be/~smc/daisy/

[22] J.-F. Cardoso, "Multidimensional independent component analysis," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, vol. 4, May 1998, pp. 1941–1944.

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

[24] J.-F. Cardoso, "Source separation using higher order moments," in *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on*. IEEE, 1989, pp. 2109–2112.

[25] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2160–2167.