

An Online Subspace Denoising Algorithm for Maternal ECG Removal from Fetal ECG Signals

Marzieh Fatemi¹ · Reza Sameni¹

Received: 30 April 2015 / Accepted: 19 February 2017 / Published online: 27 April 2017
© Shiraz University 2017

Abstract Noninvasive extraction of the fetal electrocardiogram (fECG) from multichannel maternal abdomen recordings is an emerging technology used for fetal cardiac monitoring and diagnosis. The strongest interference for the fECG is the maternal ECG (mECG), which is not always removed through conventional methods, including blind source separation, especially for low-rank abdominal recordings. In this work, we address the problem of maternal cardiac signal removal and introduce an online subspace denoising procedure customized for mECG cancellation. The proposed method is a general online denoising framework, which can be used for the extraction of a signal subspace from noisy multichannel observations in low signal-to-noise ratios, using suitable prior information of the signal and/or noise. The method is fairly generic and may also be useful for the separation of other signals and noises. The performance of the proposed technique is evaluated on both real and synthetic data and benchmarked versus state-of-the-art methods.

Keywords Online subspace denoising · Semi-blind source separation · Maternal ECG cancellation · Noninvasive fetal ECG extraction · Online generalized eigenvalue decomposition

1 Introduction

The fetal electrocardiogram (fECG) provides vital information about the fetal cardiac status. Recent measurement and processing technologies have enabled the noninvasive extraction of the fECG, from an array of sensors placed on the maternal abdomen (Sameni and Clifford 2010). One of the most challenging issues in this context is to remove maternal cardiac (mECG) interferences, without affecting the fECG. The mECG can be up to two orders of magnitude stronger than the fECG (Sameni and Clifford 2010).

To date, various methods have been developed for mECG removal, including spatial filtering (Bergveld and Meijer 1981), adaptive filtering (Widrow et al. 1975; Strobach et al. 1994; Swarnalath and Prasad 2010), template subtraction techniques (Ungureanu and Wolf 2006; Martens et al. 2007) and Kalman filtering (Sameni 2008; Sameni et al. 2007b, 2008b).

Although adaptive and Kalman filters have been very effective for single channel ECG denoising, they have two major limitations for fECG extraction: (1) the inter-channel correlation of the ECG is not used, (2) the fECG is removed with the mECG during periods of mECG and fECG temporal overlap (Sameni 2008). Both issues can be avoided by using multiple channels.

A well-known multichannel technique for extraction of fECG is *blind source separation* (BSS) using *independent component analysis* (ICA), which has been shown to be more accurate and robust as compared to similar approaches (Zarzoso and Nandi 2001). However, a basic limitation in conventional ICA is that the performance highly degrades in presence of full-rank Gaussian noise (Graupe et al. 2007), resulting in residual mECG within the fECG. It is therefore more effective to remove the mECG before applying ICA techniques (Sameni et al. 2010a).

✉ Reza Sameni
rsameni@shirazu.ac.ir

Marzieh Fatemi
marziehfatemi@cse.shirazu.ac.ir

¹ School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

More recently, a deflation subspace decomposition procedure, which we call *denoising by deflation* (DEFL), was proposed for signal subspace separation from full-rank noisy multichannel observations (Sameni et al. 2010a, b; Sameni 2008; Fatemi et al. 2013). An interesting application of this framework is for mECG removal from maternal abdominal recordings (Sameni et al. 2010a). The method has resulted in very good fECG separation, especially in low signal-to-noise ratios (SNR). Yet, a limiting factor of DEFL is the offline block-wise procedure required for generalized eigenvalue decomposition (GEVD), as the core of this algorithm. This issue has been the major obstacle in using DEFL for real-time online fECG extraction.

In this work, using recent developments in online GEVD (Zhao et al. 2008), an online extension of DEFL—called *online denoising by deflation* (ODEFL)—is introduced for eliminating the mECG from noninvasive maternal abdominal recordings. As with the offline version, the proposed method is fairly general and applicable to various scenarios depending on the prior knowledge regarding the signal and noise subspaces.

2 Problem Definition

Electrical signals recorded from the abdomen of a pregnant woman consist of mixtures of various signals including the mECG, fECG, baseline wanders and muscle contractions considered as noise. Bio-potentials recorded at the body surface are low frequency signals compared with the high propagation velocity of the electrical signals and the sensor distances (Geselowitz 1989). Therefore, the following linear instantaneous data model has been shown to be rather realistic for modeling multichannel maternal abdominal signals (Sameni et al. 2010a):

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{H}_m(t)\mathbf{s}_m(t) + \mathbf{H}_f(t)\mathbf{s}_f(t) + \mathbf{H}_\eta(t)\mathbf{v}(t) + \mathbf{n}(t) \\ &\triangleq \mathbf{x}_m(t) + \mathbf{x}_f(t) + \boldsymbol{\eta}(t) + \mathbf{n}(t) \end{aligned} \quad (1)$$

where $\mathbf{s}_m(t)$ is the maternal ECG source, $\mathbf{s}_f(t)$ is the fetal ECG source and $\mathbf{v}(t)$ represents structured noises, such as electrode movements and muscle contractions. $\mathbf{n}(t)$ is full-rank measurement noise and $\mathbf{H}_m(t)$, $\mathbf{H}_f(t)$ and $\mathbf{H}_\eta(t)$ are the transfer functions that model the propagation media from the corresponding source signals onto the body surface (Sameni et al. 2007a). In a realistic model, the cardium (of the mother and fetus) should be considered as a distributed source. Therefore, $\mathbf{s}_m(t)$ and $\mathbf{s}_f(t)$ are generally full-rank (Sameni 2008); but the effective number of dimensions can be relatively small (typically below six Sameni et al. 2006), depending on the sensor positioning and SNR.

The overall objective of noninvasive fECG extraction is to extract $\mathbf{x}_f(t)$ from this mixture. Among the different

interferences and noises, the mECG is the dominant interference, which cannot be fully separated from the fECG through conventional ICA, due to its full-rank nature, high amplitude, and background noise. This results in residual components within the extracted fECG. The DEFL algorithm was proposed to overcome this issue (Sameni et al. 2010a). Before introducing its online version, DEFL is reviewed in the following section.

3 Background

3.1 Denoising by Deflation

The DEFL algorithm is a subspace denoising method, which removes the undesired parts of a multichannel noisy data using a sequence of *linear decomposition*, *denoising* and *linear re-composition*, in a block-wise manner. As shown in Fig. 1, a block of multichannel noisy data $\mathbf{X} = [\mathbf{x}(1), \dots, \mathbf{x}(T)] \in \mathbb{R}^{N \times T}$ is given as input to the DEFL algorithm and a denoised block of the same dimension, namely $\mathbf{Y} = [\mathbf{y}(1), \dots, \mathbf{y}(T)] \in \mathbb{R}^{N \times T}$ is obtained.

The first stage of DEFL consists of finding a suitable invertible spatial filter $\mathbf{W} \in \mathbb{R}^{N \times N}$, which works as a *feature enhancer* for transforming \mathbf{X} to a space in which the data is ranked from the most to least resemblance to the “desired property”. In other words, in the transformed space, the SNR is improved within the first few channels, allowing better signal/noise separability for the first few channels. At the second stage, the signal and noise contents of the first L channels are separated using a suitable denoising method, which is customized per-application, according to the nature of the signals and noises. In the last stage, the residual signals and the $N - L$ unchanged channels are back-projected to the original space. These three stages make the first iteration of the DEFL algorithm. This procedure is repeated in multiple iterations, each time over the output of the previous iteration, until all the undesired components within the data are eliminated. The number of iterations can be selected using a termination criterion that is application-dependent and measures the quality of the signal according to a desired characteristics. For instance, the *periodicity measure* (PM) defined in Sect.

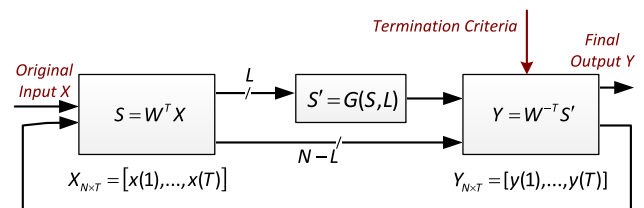


Fig. 1 Block-wise deflation scheme. Adapted from Sameni et al. (2010a)

6.2 can be used to indicate the portion of the maternal ECG that is removed (or retained) after each iteration, in each channel.

Each iteration of DEFL can be summarized as follows:

$$\mathbf{Y} = \mathbf{W}^{-T} \mathbf{G}(\mathbf{W}^T \mathbf{X}, L) \tag{2}$$

where \mathbf{X} is the input data block, \mathbf{Y} is output data block, $\mathbf{G}(\cdot, \cdot)$ is the denoising operator applied to the first L channels of the input, and \mathbf{W} is the spatial filter, as defined above.

The matrix \mathbf{W} is application-dependent. As proposed in Sameni et al. (2010a), it can be obtained by maximizing a *Rayleigh quotient* in a GEVD procedure. For the application of interest, *periodic component analysis* (π CA) (Sameni et al. 2008a) is used for estimating \mathbf{W} .

For multichannel ECG observations $\mathbf{x}(t) \in \mathbb{R}^N$, π CA consists of finding $\mathbf{w} \in \mathbb{R}^N$ in $s(t) = \mathbf{w}^T \mathbf{x}(t)$, such that the following objective function is maximized.

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{E_t\{s(t)s(t + \tau_t)\}}{E_t\{s(t)^2\}} = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{\mathbf{w}^T \mathbf{C}_\tau \mathbf{w}}{\mathbf{w}^T \mathbf{C} \mathbf{w}} \tag{3}$$

$E_t\{\cdot\}$ denotes averaging over time; $\mathbf{C} \triangleq E_t\{\mathbf{x}(t)\mathbf{x}^T(t)\}$ and $\mathbf{C}_\tau \triangleq E_t\{\mathbf{x}(t)\mathbf{x}^T(t + \tau_t)\}$ are the covariance and lagged covariance matrices, respectively; τ_t is a variable period calculated using the reference (here the maternal) ECG R-wave peaks, as defined in Sameni et al. (2008a). Estimating the matrix \mathbf{W} in Eq. (3) is equivalent to solving the following GEVD problem for $\mathbf{W} \in \mathbb{R}^{N \times N}$:

$$\mathbf{W}^H \mathbf{C}_\tau \mathbf{W} = \mathbf{\Lambda}, \quad \mathbf{W}^H \mathbf{C} \mathbf{W} = \mathbf{I}_N \tag{4}$$

where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N]$ is a matrix of generalized eigenvectors, \mathbf{I}_N is an $N \times N$ identity matrix and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_N)$ is a diagonal matrix containing the generalized eigenvalues on its diagonal. It can be shown that $\mathbf{w}^* = \mathbf{w}_1$, i.e., the eigenvector corresponding to the largest generalized eigenvalue λ_1 maximizes (3). Moreover, if \mathbf{C} and \mathbf{C}_τ are symmetric matrices, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ are real and the components of $\mathbf{s}(t) = \mathbf{W}^T \mathbf{x}(t)$ are ranked according to their resemblance with the desired (the maternal) ECG (Sameni et al. 2008a).

An interesting property of the DEFL algorithm is that unlike most PCA and ICA denoising schemes, the data dimensionality is preserved. Moreover, due to the denoising block between the linear projection stages, it overall performs as a nonlinear filtering scheme, which can deal with full-rank and even non-additive mixtures. Apparently, the method is only applicable when prior information about the signal/noise subspaces is available and the maternal ECG is normal (pseudo-periodic). In previous studies, this algorithm has been used for various applications (Amini et al. 2008; Gouy-Pailler 2009; Sameni et al. 2010a;

Sameni and Gouy-Pailler 2014). Despite its vast range of applications, the block-wise nature of the algorithm has limited its application to batch processing. In this work, an online extension of DEFL is presented.

3.2 Incremental Common Spatial Pattern

Common spatial pattern (CSP) has found vast applications in machine learning and signal processing in the recent decade. It has been widely used in biomedical applications such as brain computer interface (Ramoser et al. 2000). From an algebraic viewpoint, CSP consists of finding a matrix \mathbf{W} , which jointly diagonalizes two matrices (\mathbf{R}_l and \mathbf{R}_c) using GEVD.

An online extension of CSP, known as *incremental common spatial pattern* (ICSP), has also been developed for time-varying matrices $\mathbf{R}_l(t)$ and $\mathbf{R}_c(t)$ (Zhao et al. 2008). In ICSP, the sample-wise update of the first spatial pattern is as follows:

$$\mathbf{w}_1(t) = \frac{\mathbf{w}_1^T(t-1)\mathbf{R}_c(t)\mathbf{w}_1(t-1)}{\mathbf{w}_1^T(t-1)\mathbf{R}_l(t)\mathbf{w}_1(t-1)} \mathbf{R}_c^{-1}(t)\mathbf{R}_l(t)\mathbf{w}_1(t-1) \tag{5}$$

The minor patterns are found by repeating (5), after applying a deflation procedure on \mathbf{R}_l :

$$\mathbf{R}_l \leftarrow \left[\mathbf{I}_N - \frac{\mathbf{R}_l \mathbf{w}_1 \mathbf{w}_1^T}{\mathbf{w}_1^T \mathbf{R}_l \mathbf{w}_1} \right] \mathbf{R}_l \tag{6}$$

In Sect. 4, this recursive update algorithm is integrated in the π CA algorithm to develop an online extension of DEFL.

4 Method

Herein, an online extension of DEFL, which we coin as *online denoising by deflation* (ODEFL) is proposed for mECG cancellation. The overall block-diagram of ODEFL is summarized in Algorithm 1. In this algorithm, $\mathbf{x}(t)$ is the input multi-channel data, $\mathbf{y}_i(t)$ ($1 \leq i \leq K$) is the output of each iteration, K is the number of iterations, T is the number of samples, and $\mathbf{G}_i(\cdot, L)$ is the denoising function for removing the undesired parts,¹ applied to the first L channels in iteration i .

In Algorithm 1, unlike DEFL, which works on a block of data, ODEFL proceeds sample-by-sample in parallel units corresponding to the successive iterations of the deflation algorithm. In ODEFL, the matrix \mathbf{W} is recursively

¹ Note that for mixtures of signals with different origins and temporal characteristics, the projection (and back-projection) algorithms and the denoising scheme can generally be customized for each iterations, which is beyond the scope of the current study.

Algorithm 1 Online denoising by deflation (ODEFL)

```

1:  $\mathbf{x}_1(t) \leftarrow \mathbf{x}(t)$  ▷ Initialize with the input data
2: for  $i = 1, \dots, K$  do ▷ In each of the parallel stages of ODEFL
3:    $\mathbf{C}_i(0) \leftarrow \mathbf{I}_N$  ▷ Initialize with identity (or random unitary) matrices
4:    $\mathbf{C}_{\tau,i}(0) \leftarrow \mathbf{I}_N$ 
5:    $\mathbf{W}_i(0) = [\mathbf{w}_{i1}(0), \dots, \mathbf{w}_{iN}(0)] \leftarrow \mathbf{I}_N$ 
6:   for  $t = 1, \dots, T$  do ▷ Repeat for all samples of the data
7:      $\mathbf{C}_i(t) \leftarrow \beta \mathbf{C}_i(t-1) + \mathbf{x}_i(t) \mathbf{x}_i(t)^T$  ▷ Covariance matrix update
8:      $\mathbf{C}_{\tau,i}(t) \leftarrow \gamma \mathbf{C}_{\tau,i}(t-1) + \mathbf{x}_i(t) \mathbf{x}_i(t + \tau_t)^T$  ▷ Lagged covariance matrix update
9:      $\mathbf{C}_i(t) \leftarrow [\mathbf{C}_i(t) + \mathbf{C}_i(t)^T]/2$  ▷ Covariance matrix symmetrization
10:     $\mathbf{C}_{\tau,i}(t) \leftarrow [\mathbf{C}_{\tau,i}(t) + \mathbf{C}_{\tau,i}(t)^T]/2$  ▷ Lagged covariance matrix symmetrization
11:     $\mathbf{C}_{\tau,i}^{-1}(t) \leftarrow \gamma^{-1} \mathbf{C}_{\tau,i}^{-1}(t-1) - \frac{\gamma^{-1} \mathbf{C}_{\tau,i}^{-1}(t-1) \mathbf{x}_i(t) \mathbf{x}_i(t + \tau_t)^T \mathbf{C}_{\tau,i}^{-1}(t-1)}{\gamma + \mathbf{x}_i(t)^T \mathbf{C}_{\tau,i}^{-1}(t-1) \mathbf{x}_i(t + \tau_t)}$  ▷ Matrix inversion lemma
12:     $\tilde{\mathbf{C}} \leftarrow \mathbf{C}_i(t)$ 
13:    for  $j = 1, \dots, N$  do ▷ Perform Online GEVD of  $(\tilde{\mathbf{C}}, \mathbf{C}_{\tau}(t))$  over all channels
14:       $\mathbf{w}_{ij}(t) \leftarrow \frac{\mathbf{w}_{ij}^T(t-1) \mathbf{C}_{\tau,i}(t) \mathbf{w}_{ij}(t-1)}{\mathbf{w}_{ij}^T(t-1) \tilde{\mathbf{C}} \mathbf{w}_{ij}(t-1)} \mathbf{C}_{\tau,i}^{-1}(t) \tilde{\mathbf{C}} \mathbf{w}_{ij}(t-1)$  ▷ Incremental CSP update
15:       $\mathbf{w}_{ij}(t) \leftarrow \mathbf{w}_{ij}(t) / \|\mathbf{w}_{ij}(t)\|$  ▷ Normalization
16:       $\tilde{\mathbf{C}} \leftarrow \left[ \mathbf{I}_N - \frac{\tilde{\mathbf{C}} \mathbf{w}_{ij} \mathbf{w}_{ij}^T}{\mathbf{w}_{ij}^T \tilde{\mathbf{C}} \mathbf{w}_{ij}} \right] \tilde{\mathbf{C}}$  ▷ Deflation procedure
17:    end for
18:     $\mathbf{W}_i(t) = [\mathbf{w}_{i1}(t), \dots, \mathbf{w}_{iN}(t)]$ 
19:     $\mathbf{s}(t) \leftarrow \mathbf{W}_i^T(t) \mathbf{x}_i(t)$  ▷ Go to transform space
20:    Estimate  $L$  ▷ The number of effective mECG dimensions
21:     $\tilde{\mathbf{s}}(t) \leftarrow \mathbf{G}_i(\mathbf{s}(t), L)$  ▷ Apply denoising operator over the first  $L$  channels
22:     $\mathbf{y}_i(t) \leftarrow \mathbf{W}_i^{-T}(t) \tilde{\mathbf{s}}(t)$  ▷ Return to original space
23:     $\mathbf{x}_{i+1}(t) \leftarrow \mathbf{y}_i(t)$  ▷ Use output as input for next stage
24:  end for
25: end for

```

updated from one sample to another and all stages of DEFL are repeated on a sample-wise basis in each iteration. The major stages of Algorithm 1 are detailed below.

4.1 Online Estimation of Covariance Matrices for π CA

For an online formulation, the signal statistics contained in \mathbf{C} and \mathbf{C}_{τ} , should be tracked in time. In order to re-estimate them as the signal evolves, the temporal averaging in the definitions of \mathbf{C} and \mathbf{C}_{τ} can be replaced with a weighted sum as follows (Yang 1995):

$$\mathbf{C}(t) = \sum_{i=0}^{t-1} \beta^i \mathbf{x}(t-i) \mathbf{x}^T(t-i) \quad (7)$$

$$\mathbf{C}_{\tau}(t) = \sum_{i=0}^{t-1} \gamma^i \mathbf{x}(t-i) \mathbf{x}^T(t-i + \tau_{t-i})$$

where $\beta \in [0, 1]$ and $\gamma \in [0, 1]$ are forgetting factors. This is an *infinite impulse response* (IIR) formulation, in which all samples in the range $1 \leq i \leq t$ contribute in estimating the covariance matrices; but with smaller weights to the older samples.²

The weighted sum in (7) can be replaced with the following recursion formulas, in favor of computational and memory efficiency:

² For other applications, one might prefer a *finite impulse response* (FIR) form, in which the samples do not have any effect beyond a finite window length.

$$\mathbf{C}(t) = \beta \mathbf{C}(t-1) + \mathbf{x}(t) \mathbf{x}^T(t) \quad (8)$$

$$\mathbf{C}_{\tau}(t) = \gamma \mathbf{C}_{\tau}(t-1) + \mathbf{x}(t) \mathbf{x}^T(t + \tau_t)$$

The forgetting factors enable the adaptation of the algorithm in stationary and non-stationary environments. For stationary data, selecting $\beta = \gamma = 1$ incorporates all the samples with identical weights. For non-stationary data, the value is chosen less than 1, which for $t \gg 1$ is similar to using a sliding window with the effective window length of $1/(1-\beta)$ (Yang 1995).

In order to guarantee the symmetry of \mathbf{C} and \mathbf{C}_{τ} (to have real generalized eigenvalues extracted by GEVD), the following update is applied after re-estimation of the second order statistics.

$$\mathbf{C}(t) \leftarrow \frac{\mathbf{C}(t) + \mathbf{C}^T(t)}{2}, \quad \mathbf{C}_{\tau}(t) \leftarrow \frac{\mathbf{C}_{\tau}(t) + \mathbf{C}_{\tau}^T(t)}{2} \quad (9)$$

4.2 Online Demixing Matrix Update

In order to obtain an online solution for the GEVD problem in (3) and (4), the time-varying covariance matrix updates are integrated into the online update formula (5) as follows.

$$\mathbf{w}_1(t) = \frac{\mathbf{w}_1^T(t-1) \mathbf{C}_{\tau}(t) \mathbf{w}_1(t-1)}{\mathbf{w}_1^T(t-1) \mathbf{C}(t) \mathbf{w}_1(t-1)} \mathbf{C}_{\tau}^{-1}(t) \mathbf{C}(t) \mathbf{w}_1(t-1) \quad (10)$$

where $\mathbf{w}_1(t)$ is the the first generalized eigenvector corresponding to the largest generalized eigenvalue at time

index t . As noted in Sect. 3.2, the other minor eigenvectors are computed in a sequential (deflation) manner (Zhao et al. 2008).

As shown in Step 11 of Algorithm 1, in order to reduce the computational complexity of the matrix inversion required in (10), $C_\tau^{-1}(t)$ is recursively calculated by applying the *matrix inversion lemma* to the sample-wise covariance matrix update in (8).

It should be noted that since the online π CA algorithm requires the R-peak locations for calculating $C_\tau(t)$, the update of this matrix has a minimum delay of one ECG beat, which can be fixed to the longest expected mECG beat gap (e.g., 1.2 s). Therefore, the ODEFL output has a fixed delay with its input (of the order of a second), which is acceptable for fECG extraction.

4.3 Real-Time Implementation

The parallel structure of Algorithm 1 is specifically appealing for real-time applications. The algorithm can be efficiently implemented using reconfigurable hardware architectures, such as *field-programmable gate arrays* (FPGA), or using real-time processors, embedded systems or *graphics processing units* (GPU). For FPGA implementations, the iteration over K is converted into K parallel units (known as *modules*). For software implementations (e.g. using GPU), parallelization techniques such as *loop unrolling* can be used to implement the algorithm concurrently on K parallel processors. In either case, the iteration over time (t) is performed sample-by-sample as the data flows into the processor in real-time, with a single sample dependency to sample $t - 1$.

As later noted in Sect. 7.4, for real-time implementations (either on FPGA, embedded systems or GPU), the number of iterations K and the number of denoised channels L can be fixed to predefined values to obtain clock-wise accuracy and a constant processing load over time and processing units.

5 Benchmark Algorithms

The proposed algorithm has been evaluated on both real and synthetic data and compared with the block-wise DEFL (Sameni et al. 2010a), single-channel ECG Kalman Filter (Sameni 2008; Sameni et al. 2007b, 2008b), standard ANC (Widrow et al. 1975), a modified multistage ANC (Swarnalath and Prasad 2010), template subtraction (Martens et al. 2007), ICA denoising (Zarzoso and Nandi 2001) and a single-channel wavelet denoiser. In this section, the details of the benchmark methods used for evaluation are reviewed.

5.1 Kalman Filter

The *Kalman filter* (KF) and its nonlinear version, the *extended Kalman filter* (EKF), are methods for estimating hidden states of a system, having its dynamics and a set of observations. In the past decade, this filter has been adapted for estimating ECG signals from noisy measurements and other applications (Sameni et al. 2007b, 2008b; Sameni 2008). In summary, using a polar extension of the morphological ECG model proposed by McSharry et al. (2003), the following *state space* and *observation* models have been used as the ECG dynamic model (Sameni et al. 2007b, 2008b; Sameni 2008):

$$\begin{cases} \theta_{k+1} = (\theta_k + \omega\delta) \bmod 2\pi \\ z_{k+1} = z_k - \sum_i \delta \frac{\alpha_i \omega}{b_i^2} \Delta\theta_i \exp\left[-\frac{\Delta\theta_i^2}{2b_i^2}\right] + \eta \end{cases} \quad (11)$$

$$\begin{cases} s_k = z_k + v_k \\ \phi_k = \theta_k + u_k \end{cases} \quad (12)$$

where $\Delta\theta_i = (\theta_k - \theta_i) \bmod 2\pi$, δ is the sampling period, η is an additive noise, and the summation is taken over finite number of Gaussian waveforms used for modeling P, Q, R, S and T waves with amplitude, center and width parameters α_i , θ_i and b_i , respectively. The variable z_k , the amplitude of the noiseless ECG at time instant k , and θ (known as the *cardiac phase*), are assumed as state variables for this model. The parameters θ_i , ω , α_i , b_i and η are i.i.d Gaussian random variables considered as process noise vectors. In the observation equations, s_k and ϕ_k are amplitude and phase of the noisy observation ECG and v_k and u_k are observation noise vectors of the ECG and its phase.

Using an EKF, the ECG signal z_k can be estimated from the background noise v_k (Sameni et al. 2007b, 2008b). For our application of interest, z_k is the maternal ECG, which should be estimated and removed from the maternal abdominal sensors. Further details can be found at Sameni et al. (2007b, 2008b). The required source codes are online available at Sameni (2010).

5.2 Adaptive Noise Cancellation

Adaptive noise cancellation (ANC) is a well-known method for online signal denoising developed by Widrow et al. (1975). Standard ANC consists of a *primary input* that is the corrupted signal and a *reference input* containing the noise that is correlated with the primary noise. The weights of the filter adaptively change over time to retrieve an estimate of the noise and the weight update algorithm depends on the defined cost function. By subtracting the filter output (noise estimate) from the primary input, the

primary signal is estimated and the corrupted signal is denoised.

For mECG cancellation, the reference input is obtained by a mECG channel recorded directly from the maternal chest. The primary input is obtained by maternal abdomen recordings containing both maternal and fetal ECG.

For multichannel recordings, the ANC is applied to each channel separately. As discussed in Sameni et al. (2007b), the drawback of conventional ANC for ECG denoising is that the reference ECG should be morphologically similar to the contaminating ECG. However, since the ECG morphology highly depends on the lead position, the mECG contaminating the maternal abdominal leads do not necessarily resemble the chest lead ECG morphology. As a result, the performance of this method widely differs from one channel to another, which leads to a weak overall performance over all channels, as compared to other methods. Nonetheless, the method remains as a well-know benchmark for mECG cancellation.

More rigorously, considering $n(t)$ as the mECG, $s(t)$ as the non-mECG (fECG plus background signals), $d(t) = s(t) + n(t)$ as the noisy observations, $x(t)$ as the reference mECG, and $\mathbf{w} = [w_0, \dots, w_{p-1}]^T$ as the weight coefficient of length p , using a *least mean squares* (LMS) algorithm, the output of an ANC is obtained from Algorithm 2.

Algorithm 2 Adaptive noise cancelation (ANC) algorithm

```

1: for  $t = 1 \rightarrow T$  do
2:    $\mathbf{x}(t) = [x(t), x(t-1), \dots, x(t-p+1)]^T$ 
3:    $\hat{n}(t) = \mathbf{w}^T \mathbf{x}(t)$ 
4:    $\hat{s}(t) = d(t) - \hat{n}(t)$ 
5:    $\mathbf{w}(t+1) = \mathbf{w}(t) + 2\mu \hat{s}(t) \mathbf{x}(t)$ 
6: end for

```

In Algorithm 2, T is the number of data samples, $\hat{n}(t)$ and $\hat{s}(t)$ are estimates of primary noise and primary signal, respectively. The parameter μ is a step size that controls the filter stability and convergence rate and should be in the range $[0, \lambda_{\max}]$, where λ_{\max} is the greatest eigenvalue of the covariance matrix $\mathbf{R} = \mathbf{E}\{\mathbf{x}(t)\mathbf{x}(t)^T\}$ (Haykin 1996, Ch. 9).

More recently, other extensions of the ANC have also been introduced for fECG extraction. One of the extensions that is used in this study for comparison is a multistage ANC (Swarnalath and Prasad 2010). The modified ANC consists of two sequential adaptive filters, which enables the application of different adaptive algorithms such as LMS, *recursive least squares* (RLS) and *normalized least mean square* (NLMS) in a single filter. Another aspect of this method is that the primary and reference inputs are applied to the algorithm after a sequence of operations such as squaring and/or rescaling to increase reliability of the algorithm to situations in which the maternal ECG in the primary input is not quite similar to the reference input.

Further details regarding this filtering scheme can be followed from Swarnalath and Prasad (2010).

5.3 ICA-Based BSS and Denoising

ICA-based BSS was first used in Zarzoso and Nandi (2001) for fECG extraction from maternal abdominal sensors. This method exploits the statistical independence and spatial diversity of the sources (here the maternal and fetal heart signals plus noises), for separating fECG from other signals. In classical ICA, it is assumed that the observed signals $\mathbf{x}(t) \in \mathbb{R}^N$ are linear mixtures of N independent sources $\mathbf{s}(t) \in \mathbb{R}^N$:

$$\mathbf{x}(t) = \mathbf{A}(t)\mathbf{s}(t) \quad (13)$$

in which the mixing matrix $\mathbf{A}(t) \in \mathbb{R}^{N \times N}$ models the propagation media and $\mathbf{s}(t)$ contains the source signals. ICA methods are used to find the separating matrix $\mathbf{B}(t)$ such that $\hat{\mathbf{s}}(t) = \mathbf{B}(t)\mathbf{x}(t)$ is an estimate of the sources and $\hat{\mathbf{A}}(t) = \mathbf{B}^{-1}(t)$ is an estimate of the mixing matrix. Among the different ICA algorithms, the *joint diagonalization of eigenmatrices* (JADE) (Cardoso 1998), is used in this work as a benchmark.

In fECG applications, due to the multidimensional nature of the sources, source signals are categorized into sets of multichannel components including mECG, fECG and noise subspaces as described in *multidimensional ICA* (MICA) (Cardoso 1998) and *blind source subspace separation* (BSSS) (Lathauwer et al. 2000) schemes. Suppose that $\hat{\mathbf{s}}_f(t) = [\hat{s}_{f_1}(t), \dots, \hat{s}_{f_M}(t)]$ represents M -dimensional fetal components and the remaining components of $\hat{\mathbf{s}}(t)$ include mECG and noises. Accordingly, the corresponding columns of the mixing matrix are stored in $\hat{\mathbf{A}}_f(t) = [\hat{\mathbf{a}}_{f_1}, \dots, \hat{\mathbf{a}}_{f_M}] \in \mathbb{R}^{N \times M}$. As a result, the contribution of the fetal signals in the observation signals is obtained as follows:

$$\hat{\mathbf{x}}_f(t) = \hat{\mathbf{A}}_f(t)\hat{\mathbf{s}}_f(t) \quad (14)$$

in which $\hat{\mathbf{x}}_f(t)$ is the extracted fECG signal in the original domain. A known drawback of conventional ICA is that they cannot preserve the order, sign and amplitude of the sources (Hyvärinen et al. 2001). Therefore, for automatic applications, reliable source type detection and block-wise sign/amplitude correction is required to identify and correct the fECG sources among the other extracted components. In practice, due to the rather structured morphology of the ECG, the significant amplitude of the mECG compared to the fECG and accessible of prior information about the mECG (from maternal chest leads), the mECG signals can be systematically identified in the transformed space. In this work, we detect mECG signals using a channel assessment criteria based on maternal R-peaks.

6 Evaluation

Both synthetic and real data are used for qualitative and quantitative evaluation of the proposed method. The details of both datasets are presented in this section.

6.1 Real Data

The widely used DaISy fECG dataset, shown in Fig. 6a is used for evaluation (Moor 1997). 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.

6.2 Synthetic ECG Generation

Synthetic maternal and fetal ECG mixtures are generated using a realistic model adopted from the *open-source electrophysiological toolbox* (OSET) (Sameni 2010; Sameni et al. 2007a):

$$\mathbf{x}(t) = \mathbf{H}_m(t)\mathbf{s}_m(t) + \mathbf{H}_f(t)\mathbf{s}_f(t) + \mathbf{H}_\eta(t)\mathbf{v}(t) + \mathbf{n}(t) \quad (15)$$

$$\triangleq \mathbf{x}_m(t) + \mathbf{x}_f(t) + \boldsymbol{\eta}(t) + \mathbf{n}(t)$$

This model is based on the single dipole model of the heart, which assumes three geometrically orthogonal lead pairs, known as the Frank lead electrodes, or the *vectorcardiogram* (VCG), and a linear propagation media for the body volume conductor to map the three dimensions to body-surface potentials, using a Dower-like transformation (Edenbrandt and Pahlm 1988). Although the single dipole model is only an approximation of the true cardiac activity (Sameni et al. 2006), the model was found to be accurate enough for the hereby presented study, as it has all the required spatio-temporal features of the ECG.

Based on this model, we generate three-dimensional $\mathbf{s}_m(t)$ and $\mathbf{s}_f(t)$, representing the ECG signal of maternal and fetus hearts respectively, using a three-dimensional VCG. The ECG sources are then mapped to twelve body surface channels using the $\mathbf{H}_m(t)$ and $\mathbf{H}_f(t)$ matrices, which model the propagation media. As a result, both maternal and fetal ECG are distributed in all body surface ECG channels; but with only three underlying dimensions. A realistic full-rank noise with a desired SNR is also added to the signal using the idea proposed in (Sameni et al. 2007a). Using this model, 10,000 samples (20 s) of twelve lead synthetic maternal/fetal ECG mixtures were generated at a sampling rate of 500 Hz, for evaluation.

6.3 Quantitative Measures

After applying the denoising procedure, various measures can be used to evaluate the effectiveness of mECG cancellation algorithm, which we detail below.

6.3.1 Signal-to-Noise and Signal-to-Interference Ratios

Following (1), consider $\mathbf{x}(t)$ as the noisy input observations, $\mathbf{x}_f(t)$ as the fECG signal, $\mathbf{x}_m(t)$ as maternal interference and $\boldsymbol{\eta}(t) + \mathbf{n}(t)$ as noise for the fECG. The total interference plus noise for the fECG is

$$\mathbf{I}(t) = \mathbf{x}_m(t) + \boldsymbol{\eta}(t) + \mathbf{n}(t) \quad (16)$$

and the overall fetal *signal-to-interference-plus-noise ratio* (SINR) is defined (Sameni et al. 2010a):

$$\text{SINR} \triangleq 10 \log \left(\frac{\text{tr}(\mathbf{E}\{\mathbf{x}_f(t)\mathbf{x}_f^T(t)\})}{\text{tr}(\mathbf{E}\{\mathbf{I}(t)\mathbf{I}^T(t)\})} \right) \quad (17)$$

SINR can be used to quantify the data quality *before denoising*. For synthetic data, the SINR can be set to arbitrary ratios by scaling the mixing matrices $\mathbf{H}_m(t)$, $\mathbf{H}_f(t)$, $\mathbf{H}_\eta(t)$ and the noise variances in (1) by appropriate factors (cf. Sameni et al. 2010a for further details).

In order to assess the mECG cancellation quality, we additionally define the *signal-plus-noise-to-interference ratio* (SIR)

$$\text{SIR} \triangleq 10 \log \left(\frac{\text{tr}(\mathbf{E}\{\mathbf{x}_s(t)\mathbf{x}_s^T(t)\})}{\text{tr}(\mathbf{E}\{\hat{\mathbf{x}}_m(t)\hat{\mathbf{x}}_m^T(t)\})} \right) \quad (18)$$

where $\mathbf{x}_s(t) \triangleq \mathbf{x}_f(t) + \boldsymbol{\eta}(t) + \mathbf{n}(t)$ is the summation of all non-mECG components, which we call the *mECG complement*. $\hat{\mathbf{x}}_m(t)$ is the mECG (noise) residue in the mECG canceler's output:

$$\hat{\mathbf{x}}_m(t) = \mathbf{y}(t) - \mathbf{x}_s(t) \quad (19)$$

and $\mathbf{y}(t)$ denotes the denoised signal. Since the objective of the proposed method is to remove mECG, in an ideal mECG canceler, $\mathbf{y}(t)$ should be equal to $\mathbf{x}_s(t)$. In the later presented results, *SIR improvement* is defined as the output SIR minus the input SIR in dB. Therefore, SIR improvement is a measure of mECG cancellation in dB.

6.3.2 Periodicity Measure

The most dominant characteristic of the ECG is its pseudo-periodicity. We define the ECG *periodicity measure* (PM) as follows

$$\text{PM} \triangleq \left| \frac{\text{tr}(\mathbf{E}\{\mathbf{y}(t)\mathbf{y}^T(t + \tau_t)\})}{\text{tr}(\mathbf{E}\{\mathbf{y}(t)\mathbf{y}^T(t)\})} \right| \times 100 \quad (20)$$

The PM, measures the amount of periodicity of denoised data according to the period of a reference ECG. By definition, $0 \leq \text{PM} \leq 100$ ($\text{PM} = 0$ for fully aperiodic signals and $\text{PM} = 100$ for a fully periodic signal). By computing the PM for mECG, it indicates the amount of mECG components that still exists in the output of the denoiser. It should be noted that the reduction of PM is only a necessary—but not sufficient—measure for the algorithm success; since the PM might decrease due to an increase of noise or at a cost of losing the fECG. Therefore, a complement measure is required, which assures the fidelity of the remaining components. This measure is proposed in what follows.

6.3.3 Similarity Measure

The *similarity measure* (SM) is defined as a complement for the PM:

$$\text{SM} \triangleq \frac{|\text{tr}(\mathbb{E}\{\mathbf{y}(t)\mathbf{x}_s^T(t)\})|}{\sqrt{\text{tr}(\mathbb{E}\{\mathbf{y}(t)\mathbf{y}^T(t)\})\text{tr}(\mathbb{E}\{\mathbf{x}_s(t)\mathbf{x}_s^T(t)\})}} \quad (21)$$

SM is the correlation coefficient between the denoised data and the original signal components, $\mathbf{x}_s(t)$. By definition $0 \leq \text{SM} \leq 1$. A SM value close to 1 indicates that the algorithm has preserved the non-mECG components (including the fECG) in its output.

7 Parameter Selection

All the algorithms used for comparison have parameters that require optimization. The details of the parameter selection is studied in this section.

7.1 Extended Kalman Filter Parameters

For estimating the parameters of the Gaussian kernels used in the extended Kalman filter, the ensemble average of the mECG are extracted as a single beat average template. Next, the parameters are estimated by applying a nonlinear least squares error algorithm to fit the ECG template, using open-source packages available in OSET (Sameni 2010). The other parameters and covariance matrices are initialized following the methods developed in Sameni et al. (2007b).

7.2 ANC Parameters

The standard ANC and the modified multistage ANC are implemented using a 5-tap FIR filter (20 ms window length at a 250 Hz sampling frequency) with a step size equal to $\mu = 10^{-6}$. Both parameters were found as the optimal

values, by searching over a grid of possible values in varying SINR. The maternal ECG reference, required for the ANC is selected directly from $\mathbf{x}_m(t)$ in Eq. (15) during the generation of synthetic data. Since $\mathbf{x}_m(t)$ is a pure mECG without other noise and interferences, each of its channels can play the role of the chest lead ECG required as reference.

7.3 Wavelet Parameters

In Sameni et al. (2007b), Sameni (2017), a comprehensive study has been reported on more than 7000 combinations of wavelet parameters for ECG denoising. Herein, based on these studies, the *Coiflets3* mother wavelet with six levels of signal decomposition, using the *Stein's unbiased risk estimate* (SURE) shrinkage rule, single level rescaling and a soft thresholding strategy is used as the optimal denoising setup for the wavelet-based ECG denoiser (cf. Sameni et al. 2007b; Sameni 2017 for a detailed discussion).

7.4 DEFL and ODEFL Parameters

The optimum number of iterations, K , the number of channels to be denoised in each iteration, L , and the strategy used for denoising are critical (and application-dependent) issues that highly influence the performance of DEFL and ODEFL. The parameter K , provides the capability of eliminating full-rank and possibly nonlinearly superposed noise, which is beyond the capabilities of conventional ICA techniques. The parameter L , may be considered as the *effective number of dimensions* of the signal and noise subspaces.

For typical software-based implementations, the parameters K and L can be dynamically optimized using signal-dependent measures calculated online. This results in variable values for these parameters depending on the signal quality and the ECG channels used during data collection. On the other hand, for clock-wise accurate software implementations (e.g. real-time embedded systems) or parallel hardware implementations (e.g. using FPGA), fixed values of K and L are preferred.

The denoising function $\mathbf{G}(\cdot, \cdot)$, used for signal and noise subspace separation, also influences the overall performance of both DEFL and ODEFL. In practice, all of these parameters should be tuned according to the application.

Herein, a Monte Carlo simulation was carried out to investigate the sensitivity of DEFL and ODEFL algorithms, with respect to the denoising function and the values of L and K . The performance was investigated using 700 simulated data, generated according to the scheme in Sect. 6.2, in different input SINRs, in the range of -35 to -5 dB in 5 dB steps. Figure 2 shows the average SIR improvements versus K and L using four denoising

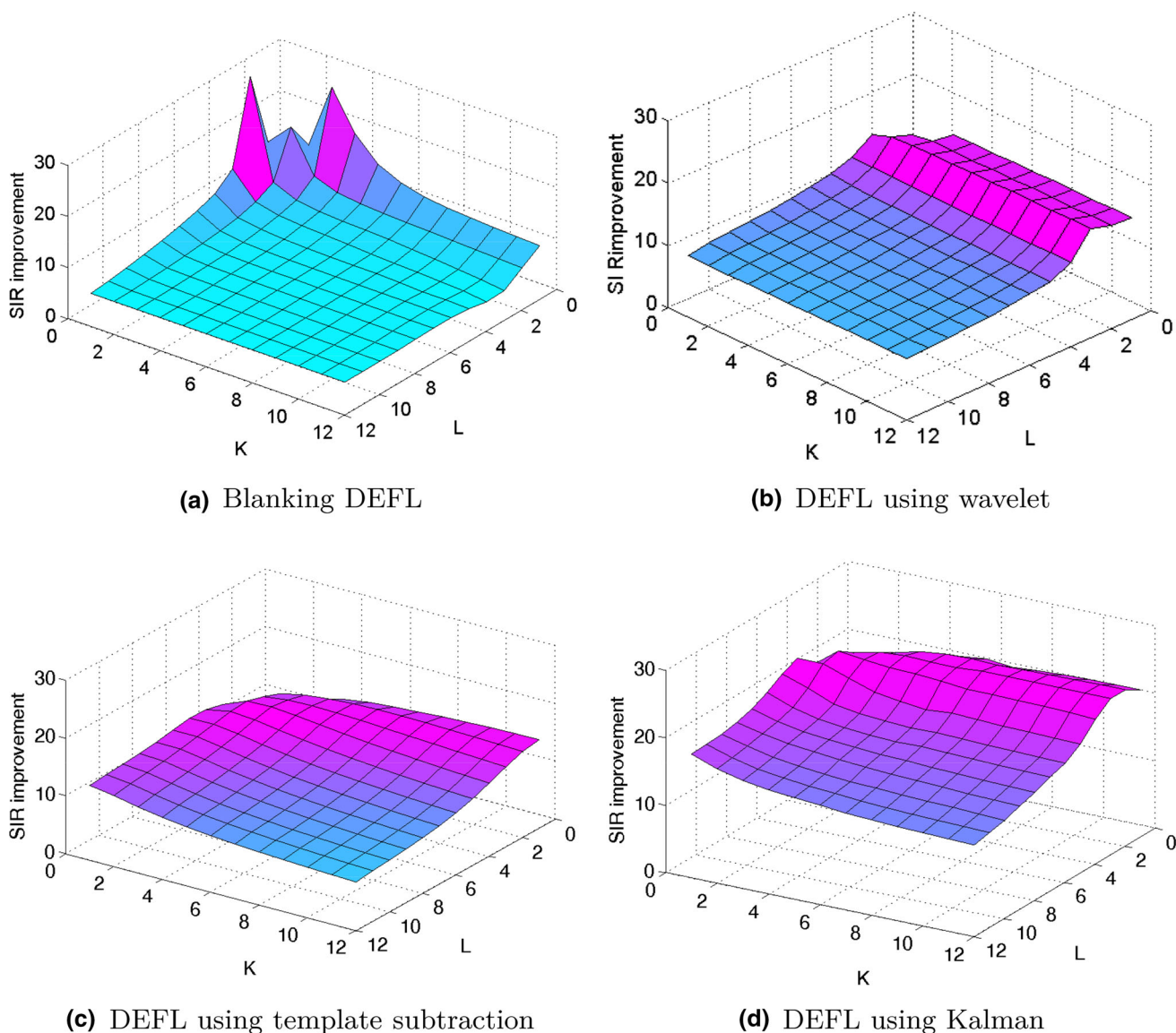


Fig. 2 Sensitivity of the SIR improvement versus K and L parameters in four denoising schemes

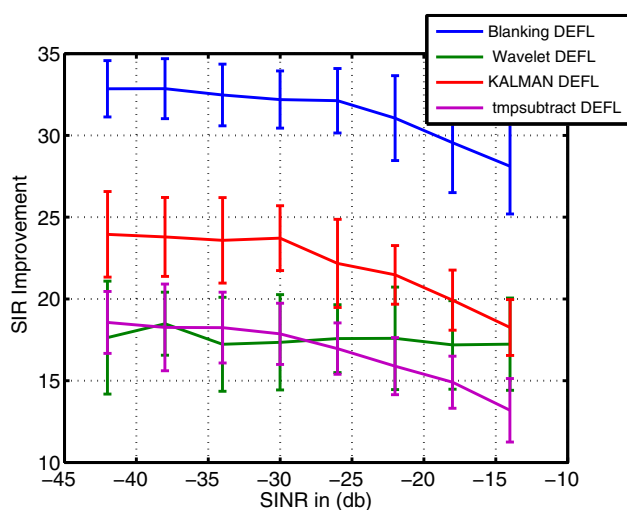
strategies $\mathbf{G}(\cdot, \cdot)$. In the first strategy, which we call *blinking DEFL*, the first L channels of $\mathbf{s}(t)$ are simply set to zero (similar to hard-thresholding in wavelet denoising). In the second strategy, wavelet denoising was used as the denoiser using the optimal parameters explained in Sect. 7.3. In the third strategy, the single-channel extended Kalman filtering scheme proposed in Sameni et al. (2007b, 2008b) is used as the denoiser. In the fourth strategy, the single-channel template subtraction technique proposed in Martens et al. (2007) is used as the denoiser.

The results of optimizing the parameters of all methods are shown in Fig. 3. In Fig. 3a, the SIR improvement versus different SINRs is calculated for the best values of K and L parameters. In Fig. 3b, the average SIR improvement over the average of the whole values of K and

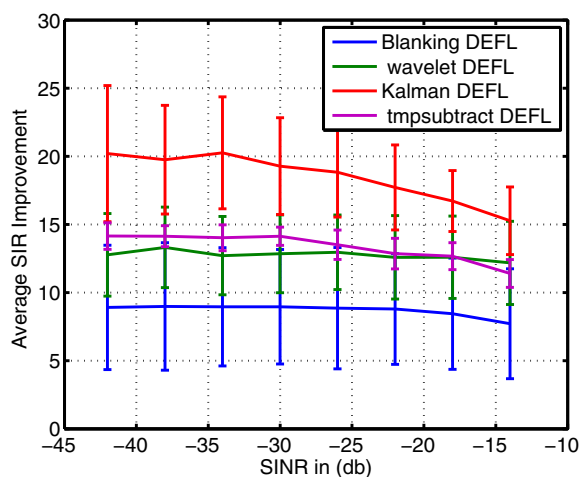
L in the range of studied parameters is calculated versus different SINRs.

According to Figs. 2 and 3a, by setting appropriate values for L and K , blinking DEFL has better performance as compared to wavelet, template subtraction and Kalman denoising strategies, which is due to the fact that when the signal space dimensions are obtained, the algorithm completely removes all the noise space dimensions while it leaves the signal unchanged.

In practice, the appropriate value of K can be estimated using some termination criterion such as the PM criterion. The optimal value of L can also be calculated using related methods for estimating the signal/noise dimensionality (Nadakuditi and Edelman 2008; Lee and Verleysen 2007). For non-stationary data, K and L can also be updated in



(a) For best parameter set



(b) Average over all parameters in range

Fig. 3 SIR improvement versus SINR using four denoising strategies

time.³ According to Fig. 3, although blanking DEFL performs best for the suitable parameters, it is sensitive to the proper choice of K and L and its performance highly degrades in case of inappropriate parameters. On the other hand, wavelet denoiser, template subtraction and Kalman denoising strategies are more robust to the choice of parameters; since increasing K and L beyond their optimal values does not significantly degrade the SIR improvements. As a result, using denoising methods such as wavelets, template subtraction or Kalman filter in DEFL, instead of banking DEFL are more appropriate in practice.

³ According to our empirical results, for ECG signals, the update should be done over long temporal windows (tens of seconds and above) rather than short windows; otherwise the performance degrades.

From Fig. 3 it is also seen that the Kalman filter outperforms template subtraction and wavelet denoiser in terms of SIR improvement and robustness to its parameters. This result was anticipated, as the Kalman filter is a model-based approach, which benefits from prior knowledge of the signal. Besides, as compared to template subtraction, the Kalman filter performance relies on both the model and the observation, which makes it effectively adaptive to different SNR scenarios. Nevertheless, the necessity of a signal model is a limitation of this method in practice as compared to the non model-based methods. In what follows, for simplicity, the first denoising strategy (blanking the first L components) with $K = 1$ and $L = 3$ are used for evaluation of both DEFL and ODEFL algorithms.

The other parameters of ODEFL are the forgetting factors β and γ . These factors should be chosen according to the degree of data (non-)stationarity within the range $[0, 1]$. In the studied database, the ECG signal and noise were both stationary. Hence, we chose $\beta = \gamma = 1$, i.e., the algorithm does not forget the old samples.

7.5 ICA Denoising Parameters

The free parameter in ICA denoising is the number of mECG components (effective number of mECG dimensions) that should be removed after the source separation stage. For synthetic data, according to our prior knowledge, $s_m(t)$ is three-dimensional. Therefore, we set $L = 3$. For real data, this choice was also empirically found to be the optimal value for the studied dataset, in order to eliminate the most dominant components of the mECG. In general, the number of mECG channels can be adaptively obtained during the denoising process by morphological similarity (the PM measure defined in 20), or by using the notion of *effective number of dimensions* (Sameni and Gouy-Pailler 2014). In this work, the mECG identification for both real and synthetic data is accomplished by computing the similarity measure defined in (21) between the maternal reference signal (chest lead ECG) and the different source channels extracted by ICA. The top L channels having the highest correlations are selected as the mECG components. These channels are set to zero and the remaining channels are back-projected to the original subspace. This strategy is rather similar to a single stage of the DEFL algorithm.

8 Results

8.1 Simulated Data

The simulated data generation procedure was discussed in Sect. 6.2. For visual inspection, a typical 20 s length synthetic ECG with SINR of -20 dB, along with the

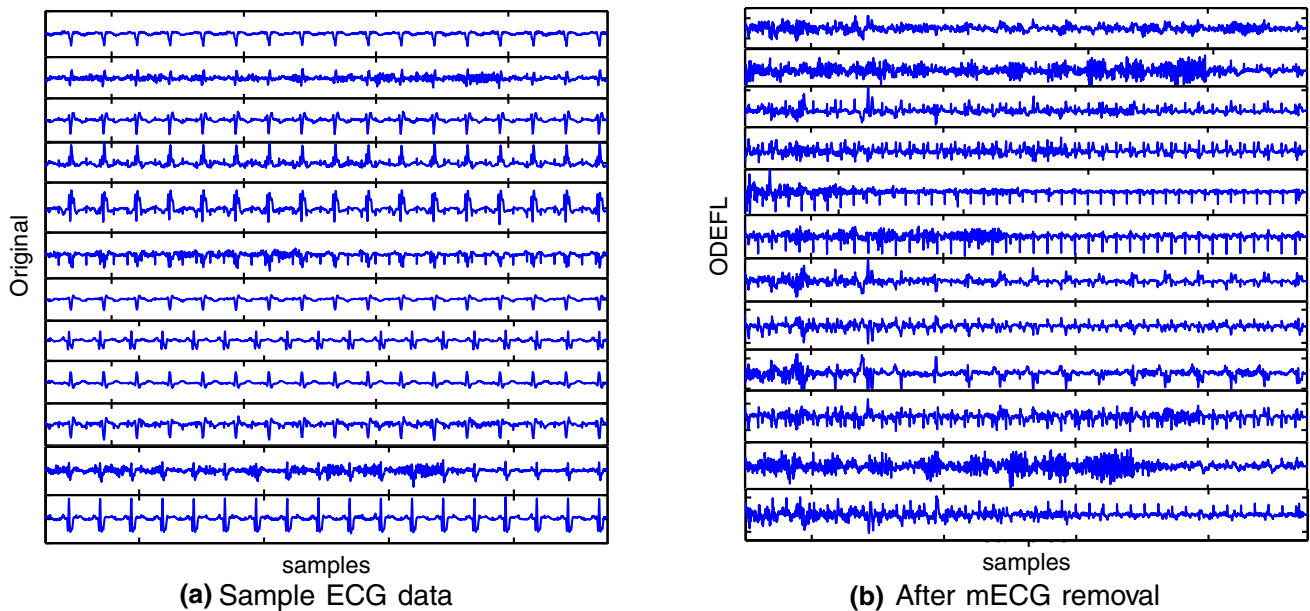


Fig. 4 A sample of synthetic ECG with -20 dB SINR and 20 s length **a** before and **b** after mECG removal. The peaks remaining after mECG removal are the fECG plus background noise

corresponding denoised output (mECG removal) is shown in Fig. 4. It can be seen that the mECG is distributed in all the simulated channels. The denoised output indicates that the maternal ECG is removed in almost all channels without affecting the fetal ECG. The first 500 samples (1 s) of the denoised data, show the transient effect of the filter. The filter has reached steady state after this period.

For a quantitative evaluation, the proposed algorithm was compared with the benchmark methods using 1000 different ensembles of simulated data and noise, in different input SINRs. The average and standard deviation of SIR improvements, PM, and SM are shown in Fig. 5. Accordingly, DEFL outperforms all methods and is only slightly better than the ODEFL. The outperformance of DEFL as compared to ODEFL is reasonable, due to the offline (non-causal) and exact calculation of the covariance matrices used in DEFL. However, the difference is negligible as compared to the advantages of ODEFL for online and nonstationary applications. As shown, DEFL and ODEFL, which are based on prior knowledge of the ECG periodicity have outperformed ICA. This is due to the fact that DEFL and ODEFL can deal with situations that ICA assumptions are not satisfied. In fact, ICA algorithms despite their vast and effective applications have some intrinsic ambiguities due to their simplified assumptions. Typically, it is assumed that the number of independent sources is fixed and equal to the number of sensors. The signal mixture is considered instantaneous and time-invariant. However, these assumptions are not necessarily satisfied in practice. As a result, the performance of ICA degrades in presence of full-rank Gaussian noise and

correlated/distributed sources (Fatemi et al. 2013), resulting in residual mECG within the fECG. Moreover, the ranking property of DEFL and ODEFL (contrary to the permutation ambiguity of ICA) helps the reliable and automatic detection of fECG/mECG signals in long recordings (Fatemi et al. 2013); while for ICA it is necessary to have robust source identification methods, which identify the mECG among others components.

Overall, DEFL, ODEFL and ICA denoising outperform the other benchmarks, in both low and high SNR scenarios. This can be due to the fact that the ANC, wavelet, template subtraction and Kalman filtering schemes are all single-channel, while DEFL, ODEFL and ICA benefit from the spatial information within multiple channels to obtain higher SNR.

Among the single-channel methods, the performance of Kalman filter and template subtraction is similar in low SNR and outperforms other single-channel methods; while in high SNR the Kalman filter has superior performance. The reason is that depending on the signal quality, the Kalman filter dynamically tends towards the observations or the system's prior dynamics; i.e., when the data is too noisy, the Kalman filter tracks the prior dynamic model rather than relying on the observation. Therefore, in low SNR, the Kalman filter performance is identical to template subtraction. On the other hand, in high SNR the Kalman filter benefits from the information within the observations, making it better than template subtraction.

The low performance of ANC, as mentioned before, can be related to the fact that the reference signal used in ANC (here the chest lead mECG) does not necessarily resemble

the morphology of the mECG superposed over the abdominal leads, which significantly downgrades its performance.

8.2 Real Data

The results of ODEFL on the DaSy dataset are shown in Fig. 6. It is seen that after about 4 samples (160 ms), the algorithm has converged and the mECG is almost completely removed in the first channel; but it takes up to 500 samples (2 s) for all channels to converge. This is due to the sequential nature of the proposed ODEFL algorithm. Figure 7 shows a closer view of the results over two successive ECG beats. It is seen that DEFL and ODEFL outperform the ANC, template subtraction, Kalman filter and ICA denoising. While DEFL and ODEFL have effectively removed the mECG, other methods have left some residual mECG or removed parts of the fECG.

For numerical evaluation of the proposed method on real data, we synthetically manipulate the real DaSy abdominal signals as follows (Sameni et al. 2010a):

$$\mathbf{x}(t) = \mathbf{G}[\mathbf{x}_0(t) + \Lambda \mathbf{v}(t)] \quad (22)$$

where $\mathbf{x}_0(t)$ is the original real data in Fig. 6, $\mathbf{v}(t)$ is Gaussian white noise, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$ is a diagonal matrix, which controls the per channel SNR, $\mathbf{G} \in \mathbb{R}^{N \times N}$ is an arbitrary non-singular random matrix and $\mathbf{x}(t)$ is the new noisy signal. The signal $\mathbf{x}(t)$ is generated in three different input SNR: 30, 20 and 10 dB, by changing the entries of Λ . The proposed method is then applied to $\mathbf{x}(t)$ by selecting $L = 3$ and $K = 2$. The algorithm is repeated over 1000 trials using different instances of $\mathbf{v}(t)$ and \mathbf{G} in each trial. The PM was defined in (20) as a measure of algorithm performance in mECG cancellation. But as noted before, the PM should be studied together with the fECG preserving indexes, to assert the overall algorithm performance. For this we define the *overall periodicity measure* (OPM):

$$\text{OPM} \triangleq \text{fPM} - \text{mPM} \quad (23)$$

where mPM and fPM are maternal and fetal PM, respectively. Accordingly, $-100 \leq \text{OPM} \leq 100$, where higher values of OPM are an indication of algorithm success in simultaneously removing the mECG and preserving the fECG. The average and standard deviation of the mPM and OPM are shown in Fig. 8 for the proposed and benchmark methods. We can see that the results on real data follow the same trend and order as the synthetic data results. The only exception is the ICA denoiser, which has inferior results for real data. This might be due to the fact that for real noisy data, mECG identification and estimation of L is difficult, resulting in a degraded performance.

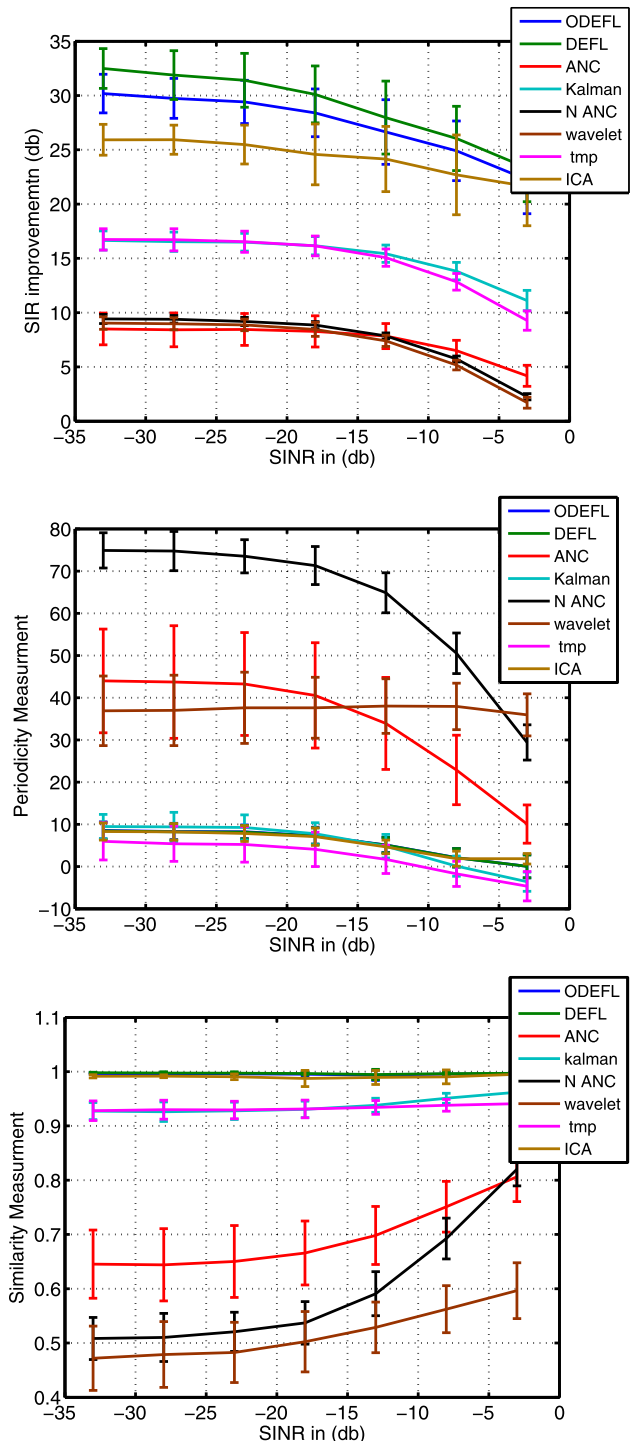


Fig. 5 (Top) SIR improvement, (middle) PM and (bottom) SM versus input SINRs. The PM and SM of DEFL and ODEFL have overlapped

9 Conclusion

In this paper, an online version of an iterative subspace denoising procedure proposed in Sameni et al. (2010a) was presented for removing maternal ECG from noninvasive

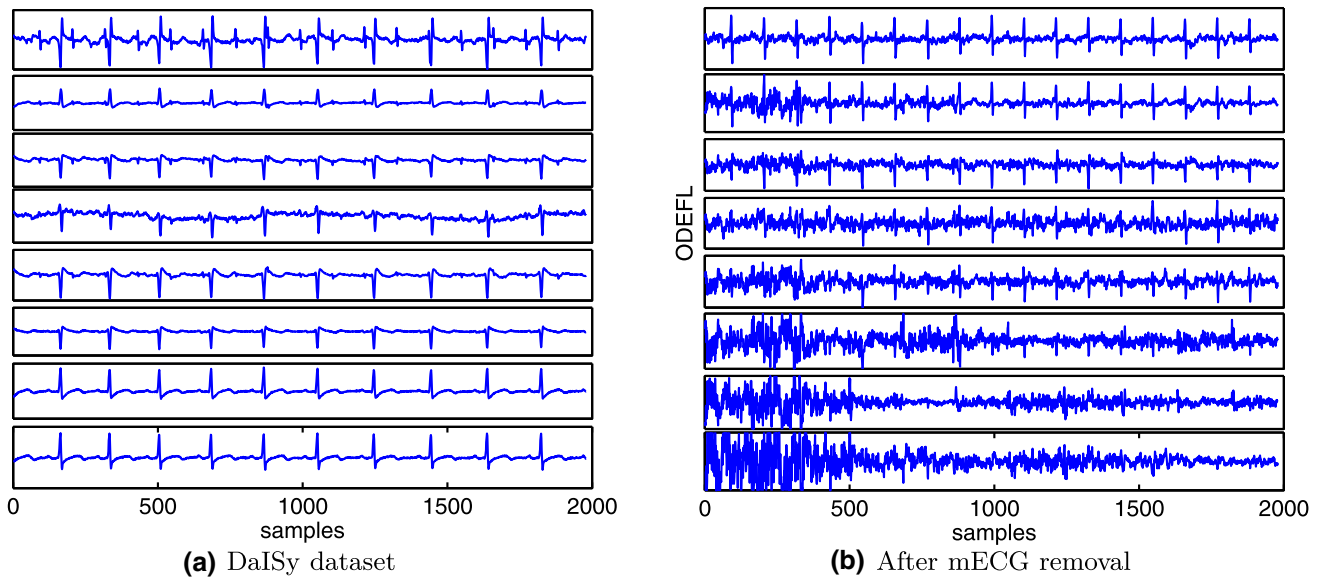


Fig. 6 DaISy dataset **a** before and **b** after mECG removal. The peaks remaining after mECG removal are the fECG plus background noise. Due to the sequential structure of ODEFL, the algorithm converges slower in the last channels

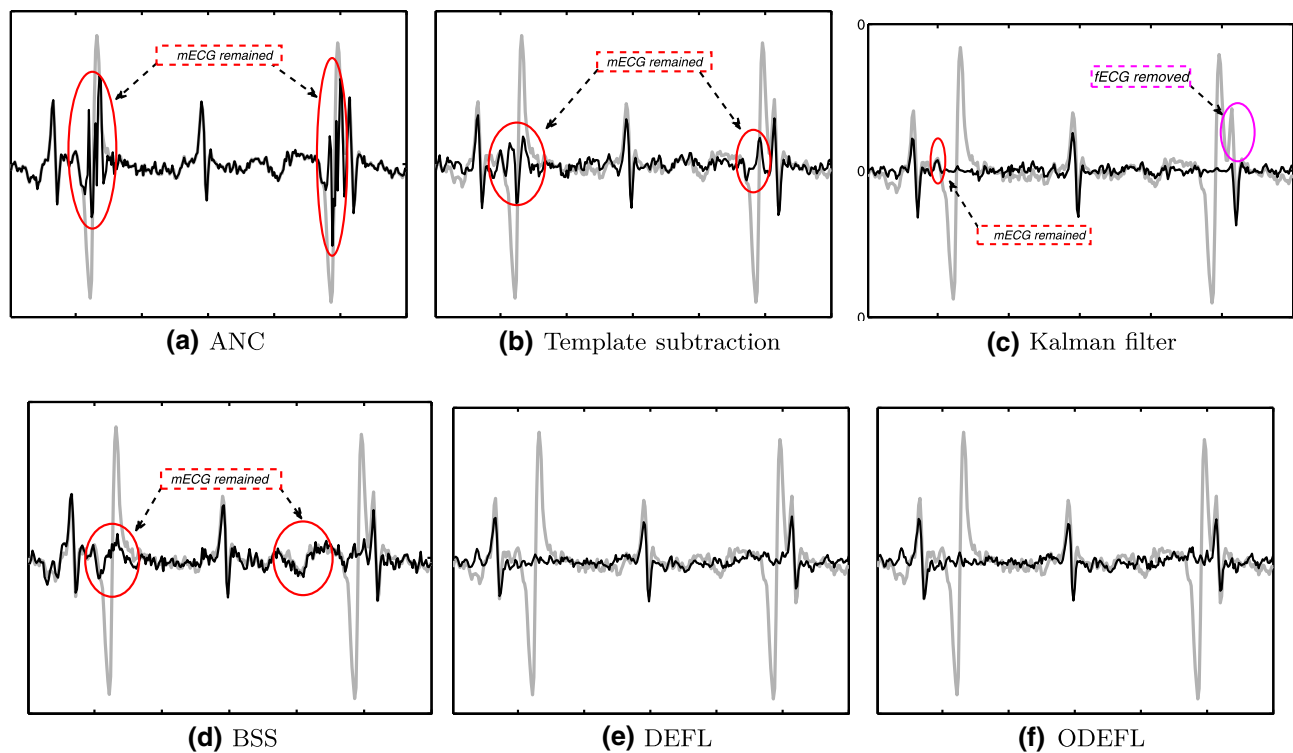


Fig. 7 A typical data segment before (*gray plots*) and after (*black plots*) mECG cancellation. It is observed that the mECG is completely removed in DEFL and ODEFL methods with minimal effect on the fECG

signals recorded from the abdomen of a pregnant woman. The proposed method is rather generic and may be applied to other blind and semi-blind source separation applications, in which the signal and noise mixtures are not separable using conventional source separation and denoising

techniques. It was shown that the proposed method outperforms the state of the art single channel denoising techniques, while it marginally performs as good as its offline version. It was further shown that DEFL and ODEFL algorithms which are based on the GEVD of only

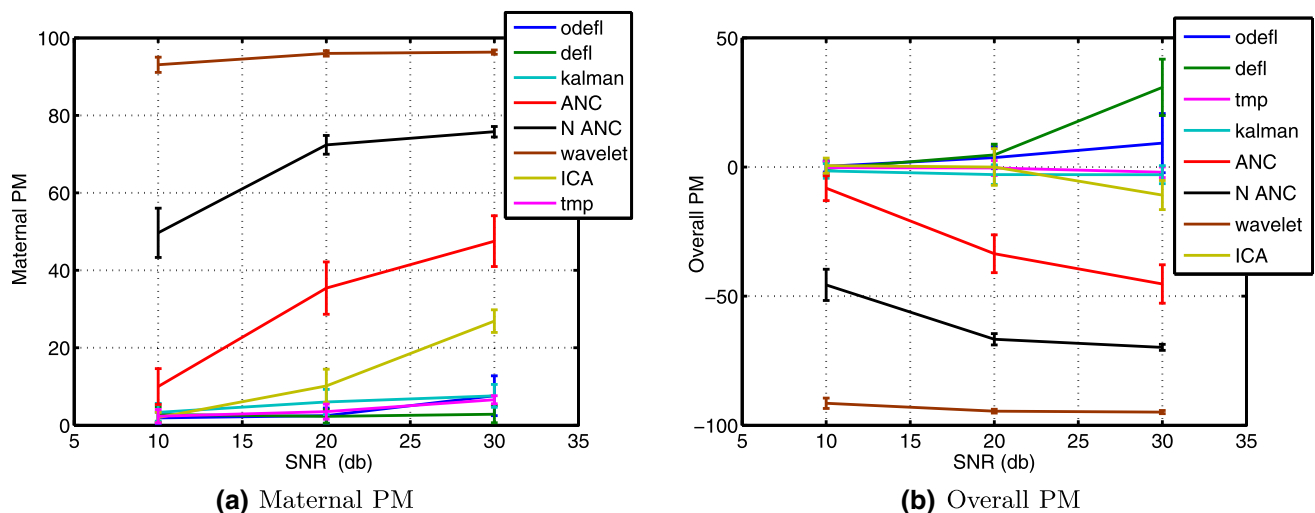


Fig. 8 Output maternal PM and overall PM versus SNR in presence of additive noise

two second-order matrices, outperform classical ICA, which use more than two matrices containing the higher-order statistics of the observations. The outperformance can be related to the fact that DEFL and ODEFL can deal with situations in which some of the underlying assumptions of ICA are not satisfied. Moreover, DEFL and ODEFL benefit from the ranking property of GEVD for mECG detection; while ICA suffers from permutation and sign ambiguities, which require the utilization of a robust mECG identifier. As a result, the proposed method is less complicated and more reliable for long datasets, as compared with batch ICA techniques.

The performance of ODEFL was investigated with different sets of parameters using different denoising strategies including simple blanking, wavelet denoising, template subtraction and Kalman denoising.

According to the hereby presented results and the former experiments reported in Sameni et al. (2007b), we conclude that for single channel data, the Kalman filter outperforms other ECG denoising schemes in different SINR scenarios, while the DEFL and ODEFL algorithms are better for multichannel data as they use inter-channel correlations, without having the mixing matrix of the data. Therefore, in future studies, the combination of the Kalman denoiser and ODEFL may result in superior results. Introducing an online method for automatic calculation of the algorithm parameters L , K , β and γ is also an interesting extension to the current work, which was partially studied in Sameni and Gouy-Pailler (2014); but requires further investigation in future studies.

The performance of ODEFL is influenced by several parameters including the method used for online GEVD. In future studies, other online GEVD algorithms can be compared with incremental common spatial pattern, used in this work. Moreover, theoretical aspects of online GEVD

and the convergence of DEFL and ODEFL should also be considered. A symmetric extension of the method for avoiding the problems of sequential source separation and error propagation is also interesting for practical applications.

In recent studies, the problem of fetal motion tracking using noninvasive ECG recordings has found significant interests (Biglari and Sameni 2016). In future studies, the hereby proposed techniques can be combined with these developments to obtain a unified fetal ECG extraction and motion tracking algorithm.

References

- Amini L, Sameni R, Jutten C, Hossein-Zadeh G, Soltanian-Zadeh H (2008) MR artifact reduction in the simultaneous acquisition of EEG and fMRI of epileptic patients. In: EUSIPCO2008—16th European signal processing conference, Lausanne, Switzerland, August 25–29, 2008
- Bergveld P, Meijer WJH (1981) A new technique for the suppression of the MEG. *IEEE Trans Biomed Eng BME-28*(4):348–354
- Biglari H, Sameni R (2016) Fetal motion estimation from noninvasive cardiac signal recordings. *Physiol Meas* 37(11):2003–2023
- Cardoso J-F (1998) Multidimensional independent component analysis. In: Proceedings of the IEEE international conference on acoustics, speech, and signal processing (ICASSP '98), vol 4, pp 1941–1944, May 1998
- de Lathauwer L, de Moor B, Vandewalle J (2000) Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Trans Biomed Eng* 47(5):567–572
- De Moor B (1997) Database for the identification of systems (DaISy) [Online]. <http://homes.esat.kuleuven.be/smc/daisy/>
- Edenbrandt L, Pahlm O (1988) Vectorcardiogram synthesized from a 12-lead ECG: superiority of the inverse Dower matrix. *J Electrocardiol* 21:361
- Fatemi M, Niknazar M, Sameni R (2013) A robust framework for noninvasive extraction of fetal electrocardiogram signals. In: Computing in cardiology conference (CinC), 2013, pp 201–204

- Geselowitz DB (1989) On the theory of the electrocardiogram. *Proc IEEE* 77:857–876
- Gouy-Pailler C, Sameni R, Congedo M, Jutten C (2009) Iterative subspace decomposition for ocular artifact removal from EEG recordings. In: Proceedings of the 8th international conference on independent component (ICA 2009), Paraty, Brazil, pp 419–426
- Graupe D, Zhong Y, Graupe MH (2007) Extracting fetal from maternal eeg for early diagnosis: theoretical problems and solutions—baf and ica. In: Proceedings of the fifth IASTED international conference: biomedical engineering, ser. BIEN '07. Anaheim, CA, USA: ACTA Press, pp 352–356
- Haykin S (1996) Adaptive filter theory. Prentice-Hall information and system sciences series. Prentice Hall, Upper Saddle River
- Hyvärinen A, Karhunen J, Oja E (2001) Independent component analysis. Wiley-Interscience, New York
- Lee JA, Verleysen M (2007) Nonlinear dimensionality reduction. Springer Science, Berlin
- Martens SMM, Rabotti C, Mischi M, Sluijter RJ (2007) A robust fetal ECG detection method for abdominal recordings. *Physiol Meas* 28(4):373–388
- McSharry PE, Clifford GD, Tarassenko L, Smith LA (2003) A dynamic model for generating synthetic electrocardiogram signals. *IEEE Trans Biomed Eng* 50:289–294
- Nadakuditi R, Edelman A (2008) Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Trans Signal Process* 56(7):2625–2638
- Ramoser H, Muller-Gerking J, Pfurtscheller G (2000) Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans Rehabil Eng* 8(4):441–446
- Sameni R (2008) Extraction of fetal cardiac signals from an array of maternal abdominal recordings. Ph.D. dissertation, Sharif University of Technology—Institut National Polytechnique de Grenoble, July 2008
- Sameni R (2010) The open-source electrophysiological toolbox (OSET), version 2.1 [Online]. <http://www.oset.ir>
- Sameni R (2017) Online filtering using piecewise smoothness priors: application to normal and abnormal electrocardiogram denoising. *Signal Process* 133:52–63
- Sameni R, Clifford GD (2010) A review of fetal ECG signal processing; issues and promising directions. *Open Pacing Electrophysiol Ther J (TOPETJ)* 3:4–20
- Sameni R, Gouy-Pailler C (2014) An iterative subspace denoising algorithm for removing electroencephalogram ocular artifacts. *J Neurosci Methods* 225(3):97–105
- Sameni R, Jutten C, Shamsollahi MB (2006) What ICA provides for ECG processing: application to noninvasive fetal ECG extraction. In: Proceedings of the international symposium on signal processing and information technology (ISSPIT'06), Vancouver, Canada, August 2006, pp 656–661
- Sameni R, Clifford GD, Jutten C, Shamsollahi MB (2007a) Multichannel ECG and noise modeling: application to maternal and fetal ECG signals. *EURASIP J Adv Signal Process* 2007. Article ID 43 407. doi:10.1155/2007/43407. ISSN 1687-6172
- Sameni R, Shamsollahi MB, Jutten C, Clifford GD (2007b) A nonlinear Bayesian filtering framework for ECG denoising. *IEEE Trans Biomed Eng* 54(12):2172–2185
- Sameni R, Jutten C, Shamsollahi MB (2008a) Multichannel electrocardiogram decomposition using periodic component analysis. *IEEE Trans Biomed Eng* 55(8):1935–1940
- Sameni R, Shamsollahi MB, Jutten C (2008b) Model-based Bayesian filtering of cardiac contaminants from biomedical recordings. *Physiol Meas* 29(5):595–613
- Sameni R, Jutten C, Shamsollahi MB (2010a) A deflation procedure for subspace decomposition. *IEEE Trans Signal Process* 58(4):2363–2374
- Sameni R, Jutten C, Shamsollahi M, Clifford G (2010b) Extraction of fetal cardiac signals. U.S. Patent US 2010/0 137 727 A1, June 3, 2010
- Strobach P, Abraham-Fuchs K, Harer W (1994) Event-synchronous cancellation of the heart interference in biomedical signals. *IEEE Trans Biomed Eng* 41(4):343–350
- Swarnalath R, Prasad DV (2010) A novel technique for extraction of FECCG using multi stage adaptive filtering. *J Appl Sci* 10(4):319–324
- Ungureanu M, Wolf W (2006) Basic aspects concerning the event-synchronous interference canceller. *IEEE Trans Biomed Eng* 53(11):2240–2247
- Widrow B, Glover J, McCool J, Kaunitz J, Williams C, Hearn H, Zeidler J, Dong E, Goodlin R (1975) Adaptive noise cancelling: principles and applications. *Proc IEEE* 63(12):1692–1716
- Yang B (1995) Projection approximation subspace tracking. *IEEE Trans Signal Process* 43(1):95–107
- Zarzoso V, Nandi A (2001) Noninvasive fetal electrocardiogram extraction: blind separation versus adaptive noise cancellation. *IEEE Trans Biomed Eng* 48(1):12–18
- Zhao Q, Zhang L, Cichocki A, Li J (2008) Incremental common spatial pattern algorithm for BCI. In: (IEEE World Congress on Computational Intelligence). IEEE international joint conference on neural networks, 2008. IJCNN 2008, June 2008, pp 2656–2659