Unsupervised Feature Extraction, Signal Labeling, and Blind Signal Separation in a State Space World

Nour Zalmai, Raphael Keusch, Hampus Malmberg, and Hans-Andrea Loeliger

ETH Zurich, Dept. of Information Technology & Electrical Engineering
{zalmai, malmberg, loeliger}@isi.ee.ethz.ch, raphael.urs.keusch@alumni.ethz.ch

***Abstract*—The paper addresses the problem of joint signal separation and estimation in a single-channel discrete-time signal composed of a wandering baseline and overlapping repetitions of unknown (or known) signal shapes. All signals are represented by a linear state space model (LSSM). The baseline model is driven by white Gaussian noise, but the other signal models are triggered by sparse inputs. Sparsity is achieved by normal priors with unknown variance (NUV) from sparse Bayesian learning. All signals and system parameters are jointly estimated with an efficient expectation maximization (EM) algorithm based on Gaussian message passing, which works both for known and unknown signal shapes. The proposed method outputs a sparse multi-channel representation of the given signal, which can be interpreted as a signal labeling.**

I. Introduction

Many signals (including, in particular, raw physiological signals) are superpositions of a wandering baseline and several (possibly overlapping) repetitive signal shapes from different sources. For instance, an abdominal electrocardiogram (ECG) of a pregnant woman, as in the upper plot of Fig. 3, is composed of the maternal ECG, the fetal ECG, and a wandering baseline mainly caused by the mother's breathing. The goal of the paper is to decompose a single-channel signal into a baseline and unknown but repetitive signal shapes. This includes learning the signal shapes and finding their occurrences.

An obvious approach to such problems is successive cancellation, where one signal component after the other is identified and subtracted (e.g., see [1], [2]). The baseline is often removed with a frequency-selective filter [1] while other signal shapes are identified based on some criterion such as periodicity [2]. However, errors in the cancellation will accumulate, which limits the robustness of such methods. Moreover, using a frequency-selective filter for the baseline extraction deforms the other signal shapes.

The problem we want to solve can be expressed as a blind separation of sparse convolutive sources from single-channel measurements. Blind source separation (BSS) has been a research topic for decades and many algorithms, such as the popular ICA [3], [4], have been proposed [5]–[7]. Most of these algorithms require more than one measurement channel; therefore, several attempts have been made to first transform single-channel measurements into a multi-channel signal (e.g, with an overcomplete wavelet transform [8]–[10]) before using a standard BSS algorithm. The performance of such an approach is tightly bound to the choice of a meaningful transform, which may be hard to find. For instance, single-channel ICA fails to separate the fetal ECG from the maternal ECG, as reported in [11].

Along with the emergence of compressed sensing [12], the BSS problem has been considered with the assumption that source signals can be sparsely represented in some transform domain [13]–[16], which allows the separation of more sources than measurement channels. Learning the signal dictionary where sources are sparsely represented is an essential part of the problem. Several dictionary learning algorithms such as K-SVD [17] have been proposed (e.g., [14], [16]) and successfully applied for various BSS problems as in [18], [19]. Most of these methods iterate between estimating a sparse source vector given a dictionary and updating a dictionary given a sparse source vector. These two steps do not normally share a common objective function [16], and both steps are usually computationally demanding. Especially, the dictionary update often requires some approximation to maintain affordable computations [16]. Furthermore, with a single channel, most of these algorithms require to strongly regularize the dictionary in order to output good results as in [19], which is not always easy to incorporate in the learning rules. More importantly, such methods can fail since real-world signals such as raw physiological signals and in particular wandering baselines are not necessarily sparse in some domain [20].

In this paper, we advocate the use of linear state space models (LSSM) for solving the joint problem of signal separation and estimation. Indeed, LSSMs are versatile enough to generate, within a common representation, smoothly-varying signals, such as a wandering baseline, with white Gaussian noise inputs [21], [22] and repetitive signal shapes with sparse inputs, modeled as zero-mean normal variables with unknown variances (NUV) [23]–[25] (cf. Sec. II). We provide an efficient expectation maximization (EM) algorithm based on Gaussian message passing for jointly estimating the input and system parameters by maximum likelihood. We demonstrate that all parameters can be updated jointly with closed-form expressions (cf. Sec. III). In addition, the use of the modified Bryson-Frazier smoother [25], [26], which is numerically stable and avoids matrix inversions, allows to compute all expectations required in EM with low complexity (cf. Sec. IV). Finally, in Sec. V, we present results of our method for the single-channel fetal ECG signal separation problem.

ISBN 978-0-9928626-7-1 © EURASIP 2017

II. Problem Setup and Signal Model

Let $y = (y_1, \dots, y_K) \in \mathbb{R}^K$ be a given single-channel discrete-time signal of length $K \in \mathbb{N}$. We wish to explain the signal y as a superposition of a smooth baseline $s^{[0]} \in \mathbb{R}^K$ and $L \in \mathbb{N}$ signal components $s^{[\ell]} \in \mathbb{R}^K$, $\ell \in \{1, \dots, L\}$, with each component consisting of repetitions (across time) of an unknown or known signal shape. Occurrences of a signal shape can vary in amplitude and overlap with other signal shapes, including its own. Specifically, we have

$$y_k = s_k^{[0]} + \sum_{\ell=1}^{L} s_k^{[\ell]} + Z_k, \tag{1}$$

for $k \in \{1, \dots, K\}$, with observation noise $Z_k \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_Z^2)$.

Each signal component $s^{[\ell]}$ is modeled with its own unknown LSSM of (given) order $n_\ell \in \mathbb{N}$ such that

$$\begin{cases} X_k^{[\ell]} &= A_\ell X_{k-1}^{[\ell]} + b_\ell U_k^{[\ell]} + E_k^{[\ell]} \\ s_k^{[\ell]} &= c_\ell X_k^{[\ell]} \ , \end{cases} \tag{2}$$

with states $X_k^{[\ell]} \in \mathbb{R}^{n_\ell}$ ($X_0^{[\ell]} = 0$), $A_\ell \in \mathbb{R}^{n_\ell \times n_\ell}$, $c_\ell \in \mathbb{R}^{1 \times n_\ell}$, $b_\ell \in \mathbb{R}^{n_\ell}$, $U_k^{[\ell]} \in \mathbb{R}$, and $E_k^{[\ell]} \overset{\text{iid}}{\sim} \mathcal{N}(0, V_{E^{[\ell]}})$ with $V_{E^{[\ell]}} \in \mathbb{R}^{n_\ell \times n_\ell}$ a symmetric positive definite matrix.

For the signal components corresponding to repetitive signal shapes (i.e., $\ell > 0$), we assume that the inputs $U_1^{[\ell]}, \dots, U_K^{[\ell]}$ are sparse and that $V_{E^{[\ell]}} = \sigma_{E_\ell}^2 I_{n_\ell}$. These assumptions simply mean that $s^{[\ell]}$ is composed of a few scaled and time-shifted versions of the nominal shape

$$\phi_\ell(i) = \begin{cases} 0, & \text{for } i < 0 \\ c_\ell A_\ell^i b_\ell, & \text{for } i \geq 0 \ . \end{cases} \tag{3}$$

The isotropic Gaussian noise term, introduced via $\sigma_{E_\ell}^2$, accounts for variability in the signal shapes and reveals to be quite useful, as discussed in Sec. V. In case of successful estimation, the variance $\sigma_{E_\ell}^2$ should be small.

We model sparse inputs with independent NUV terms [23]:

$$U_k^{[\ell]} \sim \mathcal{N}\left(0, \sigma_{U_k^{[\ell]}}^2\right). \tag{4}$$

While n_ℓ is assumed fixed, the parameters c_ℓ, A_ℓ, b_ℓ, $\sigma_{E_\ell}^2$, and $\sigma_{U_1^{[\ell]}}^2, \dots, \sigma_{U_K^{[\ell]}}^2$ are unknown and need to be estimated. If we wish to restrict the allowed signal shapes, we can simply constrain c_ℓ, A_ℓ, and b_ℓ. In this way, we can handle a known signal shape. Many real-world signals are well-approximated with such LSSM representation, which comprises signal shapes of both finite (when A_ℓ is nilpotent) and infinite support. The former is popular in the BSS literature [5], [7] but the latter is not.

For the smooth baseline model (i.e., $\ell = 0$), we assume that $U_k^{[0]} = 0$, for all k, and

$$V_{E^{[0]}} = \sigma_{E_0}^2 V_0, \tag{5}$$

with $V_0 \in \mathbb{R}^{n_0 \times n_0}$ a symmetric positive definite matrix. Furthermore, c_0, A_0, and V_0 are fixed, and only $\sigma_{E_0}^2$ is unknown and to be estimated. Such model generates filtered white Gaussian noise, which includes, under a suitable choice of parameters, spline smoothing [21], [22]. Note that the baseline model is the sparsest signal component in the representation (2). In the BSS literature, such a baseline is normally filtered out in a preprocessing step, but we will treat it as just another signal component.

Finally, combining the LSSMs of the signal components, the signal y is the output of the joint LSSM of order $n = \sum_{\ell=0}^{L} n_\ell$

$$\begin{cases} X_k &= AX_{k-1} + BU_k + E_k \\ y_k &= CX_k + Z_k \ , \end{cases} \tag{6}$$

for $k \in \{1, \dots, K\}$, with states $X_k \in \mathbb{R}^n$ (obtained by stacking $X_k^{[0]}, \dots, X_k^{[L]}$), NUV-inputs $U_k = (U_k^{[1]}, \dots, U_k^{[L]})^\mathsf{T} \in \mathbb{R}^L$, white Gaussian noise $E_k \overset{\text{iid}}{\sim} \mathcal{N}(0, V_E)$,

$$V_E = \operatorname{diag}(\sigma_{E_0}^2 V_0, \sigma_{E_1}^2 I_{n_1}, \dots, \sigma_{E_L}^2 I_{n_L}) \in \mathbb{R}^{n \times n} \tag{7}$$

$$A = \operatorname{diag}(A_0, \dots, A_L) \in \mathbb{R}^{n \times n} \tag{8}$$

$$C = \begin{bmatrix} c_0 & \dots & c_L \end{bmatrix} \in \mathbb{R}^{1 \times n} \tag{9}$$

$$B = \begin{bmatrix} 0_{n_0 \times L} \\ \operatorname{diag}(b_1, \dots, b_L) \end{bmatrix} \in \mathbb{R}^{n \times L}. \tag{10}$$

Thus, the joint problem of signal separation and estimation reduces to the estimation of the input and system parameters

$$\theta = \{C, A, B, \sigma_{U_1^{[1]}}^2, \dots, \sigma_{U_K^{[L]}}^2, \sigma_{E_0}^2, \dots, \sigma_{E_L}^2, \sigma_Z^2\}. \tag{11}$$

In the following, we provide an efficient EM algorithm to estimate θ by maximum likelihood. Several authors [23]–[25], [27] pointed out that the local maxima of the likelihood function are such that the number of non-zero elements among $\sigma_{U_k^{[\ell]}}$, $\ell \in \{1, \dots, L\}$, $k \in \{1, \dots, K\}$, is substantially smaller than $L \cdot K$. Essentially, the maximization of the marginal likelihood with respect to a single parameter $\sigma_{U_k^{[\ell]}}$ leads to a thresholding effect that can exactly set $\sigma_{U_k^{[\ell]}}$ to zero.

III. Joint EM Algorithm

Considering $U = (U_1, \dots, U_K)$ and $X = (X_1, \dots X_K)$ as hidden variables, the EM algorithm iteratively updates the estimate $\hat{\theta}$ as

$$\hat{\theta} = \operatorname*{argmax}_{\theta} \mathbb{E}\left[\ln p(y, U, X | \theta)\right], \tag{12}$$

where the expectation is computed with respect to the posterior density $p(u, x | y, \hat{\theta}_{\text{old}})$. The joint density obtained from our model described in (6) factorizes into

$$p(y, u, x | \theta) = \prod_{k=1}^{K} p(y_k | x_k, \theta) p(x_k | x_{k-1}, u_k, \theta) p(u_k | \theta), \tag{13}$$

with

$$-2 \ln p(y_k | x_k, \theta) = (y_k - Cx_k)^2 / \sigma_Z^2 + \ln(2\pi\sigma_Z^2) \tag{14}$$

$$-2 \ln p(u_k | \theta) = \sum_{\ell=1}^{L} \left(u_k^{[\ell]}\right)^2 / \sigma_{U_k^{[\ell]}}^2 + \ln\left(2\pi\sigma_{U_k^{[\ell]}}^2\right), \tag{15}$$

ISBN 978-0-9928626-7-1 © EURASIP 2017

and the decomposition

$$
\begin{aligned}
&-2\ln p(x_k|x_{k-1},u_k,\theta)\\
&\quad\propto (x_k^{[0]}-A_0x_{k-1}^{[0]})^\mathsf{T}\frac{V_0^{-1}}{\sigma_{E_0}^2}(x_k^{[0]}-A_0x_{k-1}^{[0]})+n_0\ln\sigma_{E_0}^2\\
&\qquad+\sum_{\ell=1}^{L}\frac{\left\|x_k^{[\ell]}-A_\ell x_{k-1}^{[\ell]}-b_\ell u_k^{[\ell]}\right\|^2}{\sigma_{E_\ell}^2}+n_\ell\ln\sigma_{E_\ell}^2,
\end{aligned}
\tag{16}
$$

where $\propto$ means equality up to an additive constant independent of x_k, x_{k-1}, u_k, and θ.

Interestingly, the maximization in (12) can be done in closed form since it splits for each individual set of model parameters. Putting everything together, the EM update is

$$\hat{\sigma}^2_{U_k^{[\ell]}}=\mathbb{E}[(U_k^{[\ell]})^2]\ ,\ \text{for } k\in\{1,\dots,K\} \tag{17}$$

$$\hat{b}_\ell=\frac{\mathbb{E}[X_\bullet^{[\ell]}U_\bullet^{[\ell]}]-\hat{A}_\ell\,\mathbb{E}[X_{\bullet-1}^{[\ell]}U_\bullet^{[\ell]}]}{\mathbb{E}[(U_\bullet^{[\ell]})^2]} \tag{18}$$

$$\hat{\sigma}^2_{E_\ell}=\frac{1}{n_\ell K}\,\mathbb{E}\Big[\big\|X_\bullet^{[\ell]}-\hat{A}_\ell X_{\bullet-1}^{[\ell]}-\hat{b}_\ell U_\bullet^{[\ell]}\big\|^2\Big] \tag{19}$$

$$\hat{A}_\ell=\operatorname*{argmin}_{A_\ell}\operatorname{tr}\left(A_\ell V_{A_\ell}A_\ell^\mathsf{T}-2A_\ell\xi_{A_\ell}\right), \tag{20}$$

for $\ell\in\{1,\dots,L\}$, with

$$V_{A_\ell}=\mathbb{E}[X_{\bullet-1}^{[\ell]}(X_{\bullet-1}^{[\ell]})^\mathsf{T}]-\frac{\mathbb{E}[X_{\bullet-1}^{[\ell]}U_\bullet^{[\ell]}]\,\mathbb{E}[X_{\bullet-1}^{[\ell]}U_\bullet^{[\ell]}]^\mathsf{T}}{\mathbb{E}[(U_\bullet^{[\ell]})^2]} \tag{21}$$

$$\xi_{A_\ell}=\mathbb{E}[X_{\bullet-1}^{[\ell]}(X_\bullet^{[\ell]})^\mathsf{T}]-\frac{\mathbb{E}[X_{\bullet-1}^{[\ell]}U_\bullet^{[\ell]}]\,\mathbb{E}[X_\bullet^{[\ell]}U_\bullet^{[\ell]}]^\mathsf{T}}{\mathbb{E}[(U_\bullet^{[\ell]})^2]}. \tag{22}$$

The notation '$\bullet$' means that a single sum over k from 1 to K is placed in front of the direct outer expectation sign while replacing all signs '$\bullet$' by k. The cost function in (20) is a quadratic form in A_ℓ and closed-form expressions can be derived in particular when A_ℓ is in controllable, observable or Jordan canonical form. If some signal shapes are known, their respective parameters c_ℓ, A_ℓ, and b_ℓ need not be updated.

For the baseline model ($\ell=0$), we have

$$
\begin{aligned}
\hat{\sigma}^2_{E_0}=\frac{1}{n_0K}\operatorname{tr}\Big(V_0^{-1}\big(&A_0\,\mathbb{E}[X_{\bullet-1}^{[0]}(X_{\bullet-1}^{[0]})^\mathsf{T}]A_0^\mathsf{T}\\
&-2A_0\,\mathbb{E}[X_{\bullet-1}^{[0]}(X_\bullet^{[0]})^\mathsf{T}]+\mathbb{E}[X_\bullet^{[0]}(X_\bullet^{[0]})^\mathsf{T}]\big)\Big).
\end{aligned}
\tag{23}
$$

Closed-form EM updates for C and σ_Z^2 are also available (see [27]). Under suitable state space parameterization, C can be fixed since $b_1,\dots,b_L$ are estimated anyway. Unfortunately, updating σ_Z^2 often leads to undesirable results where σ_Z^2 goes to zero (i.e., perfect fit), which in turn makes the likelihood go to infinity with non-sparse input variances and trivial signal shapes, such as spikes. Hence, σ_Z^2 is usually fixed to obtain a desired trade-off between sparsity and data fit.

Thus, all unknown parameters in θ are jointly updated with closed-form expressions. Furthermore, all quantities required for EM can be efficiently computed using Gaussian message passing as described in Section IV. While most parameters can be randomly initialized, the way the input variances are initialized may favor a signal shape over the others. It is especially useful to favor unknown models over known ones.

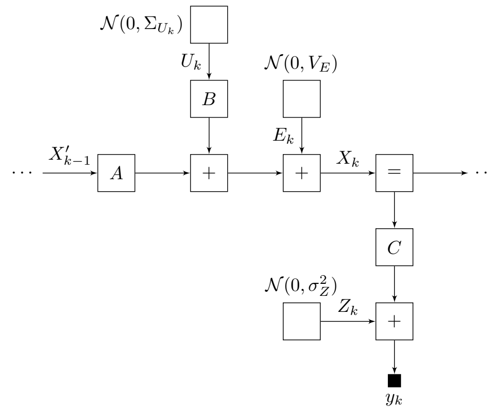

Fig. 1. Factor graph representation of the LSSM.

IV. Stable and Efficient Message Passing

Using Gaussian message passing in the factor graph [28] of Fig. 1, all quantities required for EM can be efficiently computed. The modified Bryson-Frazier smoother [25] is a suitable choice of message passing updates for our purposes. In particular, it is stable while any input (co-)variance goes to zero, which is actually expected in our algorithm. Furthermore, no matrix inversions are needed for both the message passing updates and the expectation computations. This allows the use of relatively high state space orders to include many signal shapes and/or fairly complex signal shapes.

We denote $\Sigma_{U_k}=\operatorname{diag}(\sigma^2_{U_k^{[1]}},\dots,\sigma^2_{U_k^{[L]}})$. We here recall the modified Bryson-Frazier smoother [25] and provide the link to the required expectation quantities for EM. The smoother begins with a forward pass for $k=1,\dots,K$

$$\overrightarrow{m}_{X_k}=A(\overrightarrow{m}_{X_{k-1}}+(y_{k-1}-C\overrightarrow{m}_{X_{k-1}})g_{k-1}\overrightarrow{V}_{X_{k-1}}C^\mathsf{T}) \tag{24}$$

$$\overrightarrow{V}_{X_k}=B\Sigma_{U_k}B^\mathsf{T}+V_E+AF_{k-1}\overrightarrow{V}_{X_{k-1}}A^\mathsf{T} \tag{25}$$

$$g_k=1/(\sigma_Z^2+C\overrightarrow{V}_{X_k}C^\mathsf{T}) \tag{26}$$

$$F_k=I-g_k\overrightarrow{V}_{X_k}C^\mathsf{T}C, \tag{27}$$

with $\overrightarrow{m}_{X_0}$, $\overrightarrow{V}_{X_0}$, and y_0 initialized all zero, followed by a backward pass for $k=K,\dots,1$

$$\tilde{\xi}_{X_k}=F_k^\mathsf{T}A^\mathsf{T}\tilde{\xi}_{X_{k+1}}-g_k(y_k-C\overrightarrow{m}_{X_k})C^\mathsf{T} \tag{28}$$

$$\tilde{W}_{X_k}=F_k^\mathsf{T}A^\mathsf{T}\tilde{W}_{X_{k+1}}AF_k+g_kC^\mathsf{T}C, \tag{29}$$

initialized with $\tilde{W}_{X_{K+1}}=0$ and $\tilde{\xi}_{X_{K+1}}=0$. The joint posterior of X_k, U_k, and X_{k-1} is Gaussian and thus characterized by its mean and covariance matrix, which are obtained from

$$m_{X_k}=\overrightarrow{m}_{X_k}-\overrightarrow{V}_{X_k}\tilde{\xi}_{X_k} \tag{30}$$

$$m_{U_k}=-\Sigma_{U_k}B^\mathsf{T}\tilde{\xi}_{X_k} \tag{31}$$

$$V_{X_k}=\overrightarrow{V}_{X_k}(I-\tilde{W}_{X_k}\overrightarrow{V}_{X_k}) \tag{32}$$

$$V_{U_k}=\Sigma_{U_k}-\Sigma_{U_k}B^\mathsf{T}\tilde{W}_{X_k}B\Sigma_{U_k} \tag{33}$$

$$V_{X_{k-1},X_k^\mathsf{T}}=F_{k-1}\overrightarrow{V}_{X_{k-1}}A^\mathsf{T}(I-\tilde{W}_{X_k}\overrightarrow{V}_{X_k}) \tag{34}$$

$$V_{X_{k-1},U_k^\mathsf{T}}=-F_{k-1}\overrightarrow{V}_{X_{k-1}}A^\mathsf{T}\tilde{W}_{X_k}B\Sigma_{U_k} \tag{35}$$

$$V_{X_k,U_k^\mathsf{T}}=(I-\overrightarrow{V}_{X_k}\tilde{W}_{X_k})B\Sigma_{U_k}, \tag{36}$$

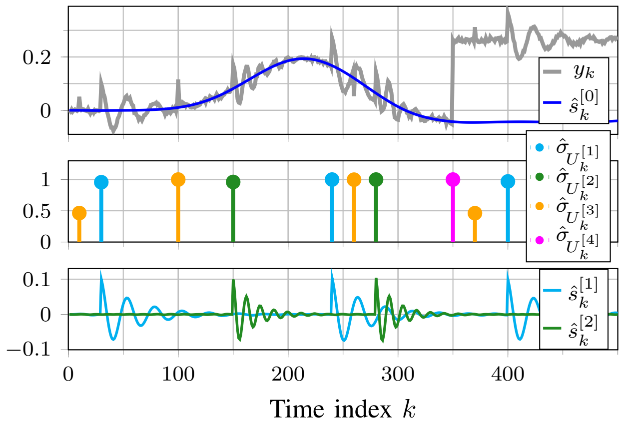

Fig. 2. Synthetic example of signal separation. For each signal component $\hat{s}^{[\ell]}$, $\ell \in \{0, \ldots, 4\}$, the average squared error is below 10^{-5}.

from which we deduce $\mathbb{E}[X_k X_k^\mathsf{T}]$, $\mathbb{E}[U_k U_k^\mathsf{T}]$, $\mathbb{E}[X_{k-1} U_k^\mathsf{T}]$, and $\mathbb{E}[X_k U_k^\mathsf{T}]$. Note that at each EM iteration, the log-likelihood is non-decreasing and can be computed iteratively as

$$\mathrm{L}_k = \mathrm{L}_{k-1} - \frac{(y_k - C\vec{m}_{X_k})^2}{\sigma_Z^2 + C\vec{V}_{X_k} C^\mathsf{T}} - \ln\big(2\pi(\sigma_Z^2 + C\vec{V}_{X_k} C^\mathsf{T})\big), \quad (37)$$

for $k = 1, \ldots, K$, with $\mathrm{L}_0 = 0$ and $\mathrm{L}_k = 2\ln p(y_1, \ldots, y_k|\theta)$. Finally, the ℓ^{th} signal component estimate is given by

$$\hat{s}_k^{[\ell]} = C m_{X_k}^{[\ell]}, \ \ k \in \{1, \ldots, K\}. \quad (38)$$

V. Experimental Results

We first illustrate our algorithm with a synthetic example. We generate a signal as in the upper plot of Fig. 2 which superimposes a baseline generated with filtered white Gaussian noise, two different decaying sinusoids, spikes, an offset, and white Gaussian noise of variance 10^{-2}. For our algorithm, we use a LSSM with $L = 4$ where

- $n_0 = 2$ with (c_0, A_0, V_0) that models a cubic spline smoothing [22]
- $n_1 = 2$ with $c_1 = [1, 0]$ and unknown $A_1 = \rho_1 R(\omega_1)$ and $b_1 \in \mathbb{R}^2$ to model a decaying sinusoid
- $n_2 = 2$ with $c_2 = [1, 0]$ and unknown $A_2 = \rho_2 R(\omega_2)$ and $b_2 \in \mathbb{R}^2$ to model another decaying sinusoid
- $n_3 = 1$ with $c_3 = b_3 = 1$ and $A_3 = 0$ to model spikes
- $n_4 = 1$ with $c_4 = A_4 = b_4 = 1$ to model offsets

(where $R(\omega)$ denotes a 2×2 rotation matrix of angle ω).

While A_1, b_1, A_2 and b_2 are randomly initialized, the input variances are initialized as $\sigma_{E_\ell}^2 = 10^{-5}$, $\sigma_{U_k^{[1]}}^2 = \sigma_{U_k^{[2]}}^2 = 10^{-6}$, and $\sigma_{U_k^{[3]}}^2 = \sigma_{U_k^{[4]}}^2 = 10^{-7}$ in order to favor the unknown models over the known (spike and offset) ones. The noise variance σ_Z^2 is fixed to 10^{-1}. As shown in Fig. 2, our algorithm recovers the individual input positions of each model (indicated in the second plot by the non-zero values of $\hat{\sigma}_{U_k^{[\ell]}}$) and outputs a good estimation for both the individual signal shape components $\hat{s}^{[1]}$ and $\hat{s}^{[2]}$ and the baseline $\hat{s}^{[0]}$. If more unknown shapes (i.e., $L > 4$) than actually present are specified, unnecessary signal shapes are automatically disregarded by the algorithm under adequate initialization of parameters.

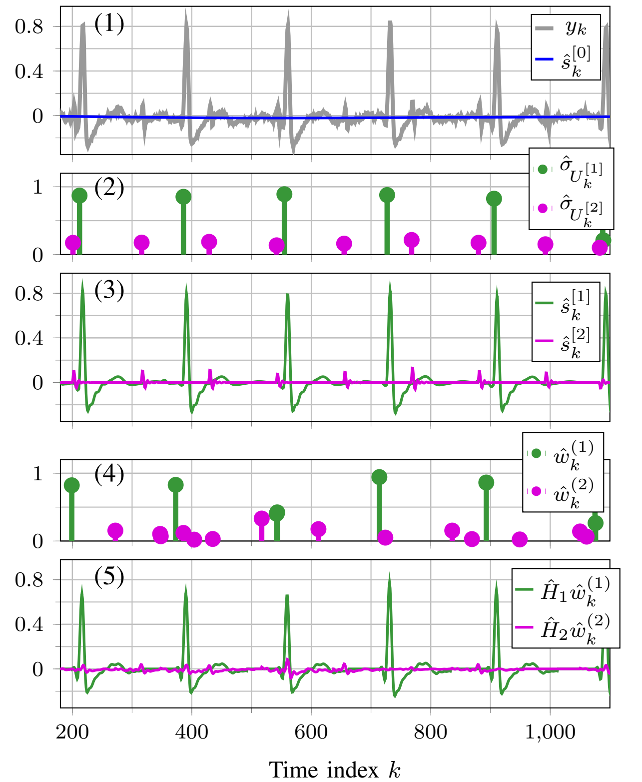

Fig. 3. Fetal ECG separation. 1) raw ECG. 1,2,3) estimated signal components from our algorithm. 4,5) estimated signal components from K-SVD.

In Fig. 3, we display the result of our algorithm on a raw abdominal ECG recording of a pregnant woman from the DaISy dataset [29]. In this recording, the fetal ECG signal is about eight times weaker than the maternal ECG, which makes its detection and estimation challenging. For our algorithm, we use $L = 2$ models consisting of linear combinations of 8 and 3 damped sinusoids. The baseline model still emulates cubic spline smoothing. The weak fetal ECG signal (plotted in purple) is remarkably well estimated and separated from the strong maternal ECG signal (plotted in green), as indicated in the subplots 2 & 3 of Fig. 3.

We also compare our algorithm with an adaptation of the K-SVD algorithm [17] that finds L sparse vectors $w^{(\ell)} \in \mathbb{R}^K$ and dictionaries $H_\ell \in \mathbb{R}^{K \times K}$ consisting of time-shifted versions of a single vector, while minimizing $\|y - \sum_{\ell=1}^{L} H_\ell w^{(\ell)}\|$ subject to, $w^{(\ell)} \geq 0$, $\|w^{(\ell)}\|_1 \leq T_\ell$, for some fixed $T_\ell > 0$. The results, for $L = 2$, are shown in subplots 4 & 5 of Fig. 3. While the maternal ECG signal is well estimated, the fetal ECG signal is poorly separated and estimated. Indeed, K-SVD algorithm does not allow much variations between occurrences of a signal shape. Thus, despite a good estimation of the maternal ECG signal, the remaining errors are still large enough to spoil the estimation and separation of the weak fetal ECG (plotted in purple in subplots 4 & 5 of Fig. 3). Our method is more robust with this respect since small variations of a pulse shape are compensated by state noise terms (via the variances $\sigma_{E_\ell}^2$).

ISBN 978-0-9928626-7-1

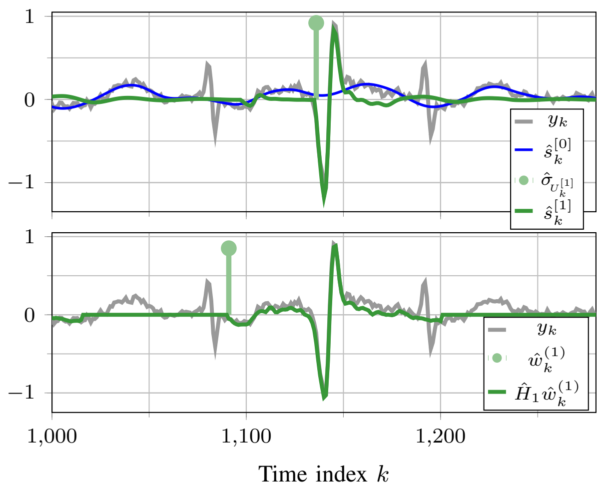

Fig. 4. Estimated maternal ECG signal (green line) using the proposed algorithm (upper plot) and K-SVD (lower plot).

We also apply both methods on an ECG recording where the fetal ECG signal is about two times weaker than the maternal ECG signal. In this scenario, both algorithms successfully estimate and separate the two signal components (not reported here). However, as shown in Fig. 4, the maternal ECG signal is estimated in different ways for each method. Namely, our algorithm includes the P and T waves of the maternal ECG in the baseline component, while K-SVD algorithm neglects the T waves and merges the P waves with the QRS complex. Since the time delays between the P wave, QRS complex, and T wave vary at each heart beat, it is actually worthwhile to include the T and P waves in the baseline component, which in turn allows a more robust estimation of the other signals.

VI. Conclusion

We have advocated a state space approach to decompose a single-channel discrete-time signal into a wandering baseline and signal components consisting of repetitive (unknown or known) signal shapes. The originality of our method resides in the combination of exploiting the state space model capabilities, sparsifying via NUV priors, and learning with an efficient EM algorithm based on a Gaussian message passing algorithm without any matrix inversion. Experimental results have emphasized on robustness and differences compared to a standard algorithm. Finally, the obtained sparse multi-channel representation suggests the use of further hierarchical processing to extract additional structural information such as synchrony of signal shapes.

References

[1] B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, J. E. Dong, and R. C. Goodlin, “Adaptive noise cancelling: Principles and applications,” *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec 1975.

[2] P. P. Kanjilal, S. Palit, and G. Saha, “Fetal ECG extraction from single-channel maternal ECG using singular value decomposition,” *IEEE Trans. on Biomedical Engineering*, vol. 44, no. 1, pp. 51–59, 1997.

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

[4] L. De Lathauwer, B. De Moor, and J. Vandewalle, “Fetal electrocardiogram extraction by blind source subspace separation,” *IEEE Trans. on Biomedical Engineering*, vol. 47, no. 5, pp. 567–572, May 2000.

[5] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[6] J.-F. Cardoso, “Blind signal separation: statistical principles,” *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, 1998.

[7] H.-L. N. Thi and C. Jutten, “Blind source separation for convolutive mixtures,” *Signal processing*, vol. 45, no. 2, pp. 209–229, 1995.

[8] J. Lin and A. Zhang, “Fault feature separation using wavelet-ICA filter,” *NDT & E International*, vol. 38, no. 6, pp. 421–427, 2005.

[9] B. Mijovic, M. De Vos, I. Gligorijevic, J. Taelman, and S. Van Huffel, “Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis,” *IEEE Trans. on Biomedical Eng.*, vol. 57, no. 9, pp. 2188–2196, 2010.

[10] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic press, 1999.

[11] M. E. Davies and C. J. James, “Source separation using single channel ICA,” *Signal Processing*, vol. 87, no. 8, pp. 1819–1832, 2007.

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

[13] M. Zibulevsky and B. Pearlmutter, “Blind source separation by sparse decomposition in a signal dictionary,” *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.

[14] B. A. Olshausen and D. J. Field, “Sparse coding with an overcomplete basis set: A strategy employed by V1?” *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[15] T.-W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, “Blind source separation of more sources than mixtures using overcomplete representations,” *IEEE Signal Processing Letters*, vol. 6, no. 4, pp. 87–90, 1999.

[16] M. S. Lewicki and T. J. Sejnowski, “Learning overcomplete representations,” *Neural Computation*, vol. 12, no. 2, pp. 337–365, 2000.

[17] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[18] B. Mailhé, R. Gribonval, F. Bimbot, M. Lemay, P. Vandergheynst, and J.-M. Vesin, “Dictionary learning for the sparse modelling of atrial fibrillation in ECG signals,” in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2009, pp. 465–468.

[19] T. Blumensath and M. Davies, “Shift-invariant sparse coding for single channel blind source separation,” *SPARS*, vol. 5, no. 75-78, pp. 8–1, 2005.

[20] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, “Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse Bayesian learning,” *IEEE Trans. on Biomedical Engineering*, vol. 60, no. 2, pp. 300–309, 2013.

[21] G. Wahba, *Spline Models for Observational Data*. Siam, 1990, vol. 59.

[22] C. Reller, “State-space methods in statistical signal processing: New ideas and applications,” Ph.D. dissertation, No. 20584, ETH Zurich, 2012.

[23] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[24] D. P. Wipf and B. D. Rao, “Sparse Bayesian learning for basis selection,” *IEEE Trans. on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

[25] H.-A. Loeliger, L. Bruderer, H. Malmberg, F. Wadehn, and N. Zalmai, “On sparsity by NUV-EM, Gaussian message passing, and Kalman smoothing,” in *Information Theory and Applications Workshop (ITA), La Jolla, CA*, Feb 2016.

[26] L. Bruderer, H. Malmberg, and H.-A. Loeliger, “Deconvolution of weakly-sparse signals and dynamical-system identification by Gaussian message passing,” in *IEEE Int. Symp. on Information Theory (ISIT)*, 2015.

[27] N. Zalmai, H. Malmberg, and H.-A. Loeliger, “Blind deconvolution of sparse but filtered pulses with linear state space models,” in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 4194–4198.

[28] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, “The factor graph approach to model-based signal processing,” *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, 2007.

[29] B. De Moor, “DaISy: Database for the identification of systems,” http://homes.esat.kuleuven.be/ smc/daisy, eSAT/STADIUS, KU Leuven, Belgium, Used dataset: [96-012], Accessed: 2017-01-20.