A BAYESIAN PCA APPROACH FOR FETAL ECG EXTRACTION

V.P. Oikonomou* and D.I. Fotiadis*

* Unit of Medical Technology and Intelligent Information Systems, Dept. of Computer Science, University of Ioannina, GR 45110 Ioannina, Greece
Biomedical Research Institute - FORTH, GR 45110 Ioannina, Greece

voikonom@cs.uoi.gr, fotiadis@cs.uoi.gr

Abstract: This paper addresses the problem of fetal electrocardiogram (FECG) extraction. A novel algorithm is proposed, which takes advantage of the nonstationarity nature of ECG signal. The FECG problem is formulated in a state - space form. This leads to an iterative algorithm. The algorithm was tested in a real life example. Visual test results are obtained and show that the proposed algorithm is capable of reliably extracting the FECG beats.

Introduction

Physiological signals may not be directly measurable and we might have to determine the signal from measurable complex signals. Fetal Electrocardiogram (FECG) belongs to this category of signals. The FECG must be extracted from the abdominal ECG which is dominated by the maternal ECG. FECG contains information on the condition of the fetus. Routine examinations such as evaluation of fetal heart rate might prove a significant tool for the doctor for pregrancy monitoring.

Several approaches have been proposed in the literature addressing FECG extraction. Windrow et al. [1] propose an adaptive filtering and adaptive noise cancellation method to extract FECG from the composite maternal ECG (MECG). Multiple MECG signals obtained from the chest leads were used to cancel the MECG component identified as noise in the composite MECG signal. Blind Source Seperation (BSS) methods have also been used for FECG extraction [2, 3]. This is illustrated in [2] where a two stage BSS algorithm is shown to be superior to Widrow's method. In [3] the performance of high order Independent Component Analysis (ICA) methods and Principal Component Analysis (PCA) is reported. The wavelet transform is used in [4] to separate sources in the time-scale domain. The FECG extraction is modeled as a BSS problem assuming that the mixed sources are statistically time invariant. However, source nonstationarity is a characteristic of biomedical signals. In our case, heart rate varies with time.

ECG recordings from a pregnant woman are affected by various bioelectric sources and noise. As indicated in [3] the measurements can be considered as instantaneous linear mixtures of signals generated by underlying bioelectric phenomena and the noise can be taken into account as an additive pertubation.

In this work we present a method for the extraction of FECG from the abdominal ECG, introducing nonstationarity for the sources and using a Bayesian PCA approach. Our work is based on the approach proposed in [5]. The authors demonstrate how the principal axes of a set of observed vectors of data can be determined through maximum likelihood estimation of a latent variable model, which is called propabilisitc PCA (pPCA). However, their approach is limited to stationary sources, which is not realistic and it is addressed in our work. The pPCA model is formulated in a state - space form and the Expectation - Maximization algorithm is used to derive the maximum likelihood (ML) values of the parameters. The above procedure is extended also to a bayesian approach.

Methods and Materials

A latent variable model seeks to relate a d-dimensional observation vector $\mathbf{y}$ to a corresponding q-dimensional vector of latent variables $\mathbf{x}$:

$$\mathbf{y} = W\mathbf{x} + \mathbf{n} \tag{1}$$

The dxq matrix W relates the observation and latent variables vectors. The latent variables are defined to be independent and Gaussian with unit variance, i.e. $\mathbf{x} \sim N(0, I_q)$, where I_q is the q-dimensional unit matrix. The noise model, $\mathbf{n}$, is Gaussian, i.e. $\mathbf{n} \sim N(0, \Psi)$. The observations follows a Gaussian, i.e. $\mathbf{y} \sim N(0, WW^T + \Psi)$. The model parameters, W and Ψ, may thus be determined by maximum likelihood. Since there isn't closed form analytic solution, their values may be obtained via an iterative procedure. It can be proved [5] that in the case of isotropic noise model, i.e. $\mathbf{n} \sim N(0, \sigma^2 I_d)$, the stationary points of the log likelihood with respect to W satisfy:

$$W_{ML} = U_q(\Lambda_q - \sigma^2 I_q)^{1/2} \tag{2}$$

where the columns of U_q are the eigenvectors of sample covariance S, with corresponding eigenvalues in the diagonal matrix Λ_q. It has also been proved that the maximum of likelihood is achieved when the q largest eigenvalues of S are chosen so that the columns of U_q correspond to the principal eigenvectors. Also, it can be shown that in the limit $\sigma^2 \to 0$ the conventional PCA is recovered.

The pPCA is formulated by Eq. (1). To derive a dynamic version of the above model the latent variables $\mathbf{x}$

ISSN: 1727-1983

must evolve with time. If the evolution in time is linear then we have the following dynamical linear model:

$$\mathbf{x}[t] = \mathbf{x}[t-1] + \mathbf{v} \quad (3)$$

$$\mathbf{y}[t] = W\mathbf{x}[t] + \mathbf{n} \quad (4)$$

where $\mathbf{v} \sim N(0, \sigma_v^2 I)$ and $\mathbf{n} \sim N(0, \sigma^2 I)$. The log likelihood of the data is:

$$\log p(X,Y) = -\sum_{t=1}^{N} \frac{1}{2\sigma^2}(\mathbf{y}[t] - W\mathbf{x}[t])^T(\mathbf{y}[t] - W\mathbf{x}[t]) - \frac{Nq}{2}\log\sigma^2 - \frac{(N-1)q}{2}\log\sigma_v^2 - \sum_{t=2}^{N} \frac{1}{2\sigma_v^2}(\mathbf{x}[t] - \mathbf{x}[t-1])^T(\mathbf{x}[t] - \mathbf{x}[t-1]) \quad (5)$$

where N is the number of samples, $Y = \{\mathbf{y}[1], \mathbf{y}[2], \cdots, \mathbf{y}[N]\}$ and $X = \{\mathbf{x}[1], \mathbf{x}[2], \cdots, \mathbf{x}[N]\}$. The log likelihood $\log p(X,Y)$ is a function of the parameters W and σ^2, The variable σ_v is supposed to be a known quantity. If the value of parameters are known the latent variables $\mathbf{x}$ can be calculated using the Kalman Smoother. However, the values of W and σ^2 is not known in advance. Thus, we employ maximum likelihood (ML) to estimate W and σ^2. A common technique for ML estimation of the parameters is the Expectation - Maximization (EM) algorithm [7]. The EM algorithm alternates between estimating the latent variables given the current model and refitting the model given the estimated, complete data. More specifically, the EM algorithm constitutes from two steps: the E - step and the M-step. In the E-step the expectation of the log likelihood of complete data is calculated given the observations and the current values of parameters. In the M-step the maximization of the expected log likelihood with respect to parameters is performed. These two steps are iterated until convergence.

To calculate the expected log likelihood, $E\{\log p(X,Y)|Y\}$, we define the following conditional mean:

$$\hat{\mathbf{x}}[t] = E\{\mathbf{x}[t]|Y\}, \quad (6)$$

and the conditional covariance:

$$P[t] = E\{(\mathbf{x}[t] - \hat{\mathbf{x}}[t])(\mathbf{x}[t] - \hat{\mathbf{x}}[t])^T|Y\} \quad (7)$$

Calculating the expectations in the log likelihood and discarding terms irrelevant of the parameters W and σ^2 we have:

$$L(W,\sigma^2) \propto -\sum_{t=1}^{N} \frac{1}{2\sigma^2}\Big[\mathbf{y}[t]\mathbf{y}[t]^T - 2\mathbf{y}[t]^T W\hat{\mathbf{x}}[t] + tr(W^T W < \mathbf{x}[t]\mathbf{x}[t]^T >)\Big] - \frac{Nd}{2}\log\sigma^2 \quad (8)$$

Maximizing the log likelihood with respect to W and σ^2 we obtain:

$$\hat{W} = \left[\sum_{t=1}^{N} \mathbf{y}[t]^T\hat{\mathbf{x}}[t]\right]\left[\sum_{t=1}^{N}(P[t] + \hat{\mathbf{x}}[t]\hat{\mathbf{x}}[t]^T)\right]^{-1}, \quad (9)$$

$$\hat{\sigma}^2 = \frac{1}{Nd}\sum_{t=1}^{N}\Big(\mathbf{y}[t]^T\mathbf{y}[t] - 2\hat{\mathbf{x}}[t]^T W^T\mathbf{y}[t] + tr(W^T W(P[t] + \hat{\mathbf{x}}[t]\hat{\mathbf{x}}[t]^T))\Big). \quad (10)$$

The Eqs. (9, 10), consists the M - step. The E - step related to the calculation of conditional mean $\hat{\mathbf{x}}[t]$ and conditional covariance $P[t]$. Those are calculated using the Kalman Smoother recursions (see Appendix A). Alternativelly an approximation to the above expectations may be obtained with Kalman Filter [8].

A bayesian treatment of a model is obtained by first introducing a prior ditribution over the parameters of the model. Then the posterior distribution is obtained by multiplying the prior by the likelihood function. Finally the predictive density is obtained by marginalizing over the parameters. To implement this framework we must address two issues: (i) the choise of a prior and (ii) the use of a tractable algortihm. In this work we focus to control the effective dimensionality of the latent space. This may be achieved by the use of a hierarchical prior $p(W|\alpha)$ over the matrix W, governed by the vector of hyperparameters $\alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_{d-1}\}$. Each hyperparameter controls one of the columns of W through a conditional distribution:

$$p(W|\alpha) = \prod_{i=1}^{d-1}\left(\frac{\alpha_i}{2\pi}\right)^{d/2}\exp\left(-\frac{\alpha_i}{2} \parallel \mathbf{w}_i \parallel^2\right) \quad (11)$$

Each α_i control the inverse of the variance of corresponding $\mathbf{w}_i$. If an α_i is concentrated on large values, the corresponding $\mathbf{w}_i$ tends to be small, this means that the corresponding direction will be "switched off". This formulation of prior is reported as Automatic Relevance Determination (ARD)[9]. We treat the parameter σ^2 as a deterministic variable and not as a random variable, so it can be determined by maximum likelihood. To use this model we must marginalize over the posterior distribution of W. In this case the solution is analytically intractable and we employ approximation methods. One such method is the local Gaussian approximation to a local mode (type - II maximum likelihood,[9]). The mode W_{MP} can be found by maximizing:

$$\log p(W|D) = L(W,\sigma^2) - \frac{1}{2}\sum_{i=1}^{d-1}\alpha_i \parallel \mathbf{w}_i \parallel^2 \quad (12)$$

This leads to:

$$W_{MP} = \left[\sum_{t=1}^{N}\mathbf{y}[t]^T\hat{\mathbf{x}}[t]\right]\left[\sum_{t=1}^{N}(P[t] + \hat{\mathbf{x}}[t]\hat{\mathbf{x}}[t]^T) + \sigma^2 A\right]^{-1} \quad (13)$$

where A is diagonal matrix with elements $\{\alpha_1, \alpha_2, \cdots, \alpha_{d-1}\}$. To estimate α we maximize the marginal likelihood in which we have integrated over W using a quadratic approximation (type - II maximum likelihood)[9], [10]. This leads to the estimation formula of the form:

$$\alpha_i = \frac{d}{\parallel \mathbf{w}_i \parallel^2} \quad (14)$$

ISSN: 1727-1983

where we have supposed that all model parameters are "well - determined". In the bayesian approach the learning algorithm proceeds with repeated application of Eqs. (6, 7,10,13,14). Optimization of W and σ^2 is alternated with the estimation of α_i using Eq. (14) until all parameters satisfy a covergence criterion. The above iterative algorithm does not gaurantee local maximization of likelihood, although convergence is achieved.

In our case the vector **y** represents the abdominal ECG measurements and the latent variables **x** are the underlying biolectric phenomena, i.e. MECG and FECG. The estimation of bioelectric phenomena, and hence the estimation of FECG, is a by-product of the proposed algorithm.

Results

To investigate the effectiveness of the proposed algorithm experiments were conducted on real ECG signals [11]. These signals represent 5 sec. recordings from eight different skin electrodes located on different points on a pregrant woman's body with sampling frequency 500 Hz. Five of these signals were obtained from the mother's abdominal region while the other three were obtained from the mother's thoracic region. In all experiments we use $\sigma_v = 1$. For the initialization of W we use the maximum likelihood solution, while the initialization of σ is random. Fig. 1 depicts 5 sec. of these signals while Fig. 2 depicts the extracted sources using the proposed algorithm. As we can see the seventh and sixth source contains the MECG component, the fifth source contains the FECG component, the fourth source contains a signal that is composed by a low frequency component and contibutions from the MECG signal, the third and second sources contain mainly noise contributions. Finally, the first source is a low periodic signal and deserves further medical interpretation, it might e.g. be due to the respiration.

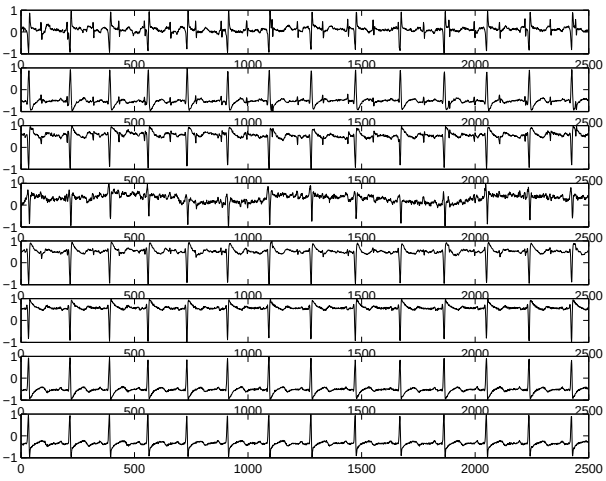

Figure 1: ECG recording of a pregnant woman

To show how the proposed approach extracts the

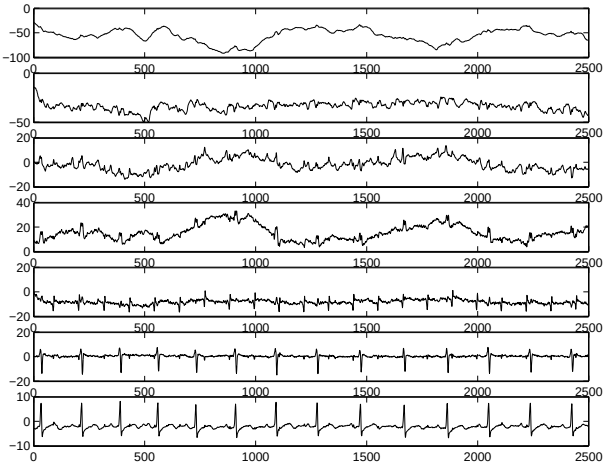

Figure 2: Extracted Sources using the proposed algorithm

FECG component from the composite ECG we provide with three representative examples. The first example contains non overlapping FECG beats, the second partially overlapping FECG beats and the third example fully overlapping FECG beats. In Fig. 3(a) we can see three nonoverlapping FECG beats and two MECG beats. Fig. 3(b) depicts the extracted FECG signal, which clearly shows the FECG beats by supressing the MECG beats. In Fig. 4(a) we can see three nonoverlapping FECG

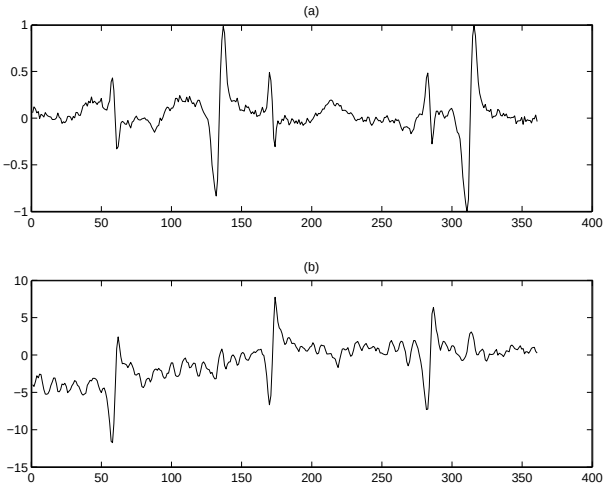

Figure 3: (a) Composite ECG and (b) Extracted FECG .

beats, one partially overlapping FECG and two MECG beats. Fig. 4(b) depicts the extracted FECG signal, which clearly show the FECG beats by supressing the MECG beats. A more difficult situation is presented in Fig. 5, where one FECG beat is fully overlapping with one MECG beat. Fig. 5(a) depicts the composite signal which contains one FECG beat fully overlapping with one MECG beat (first MECG beat).

ISSN: 1727-1983

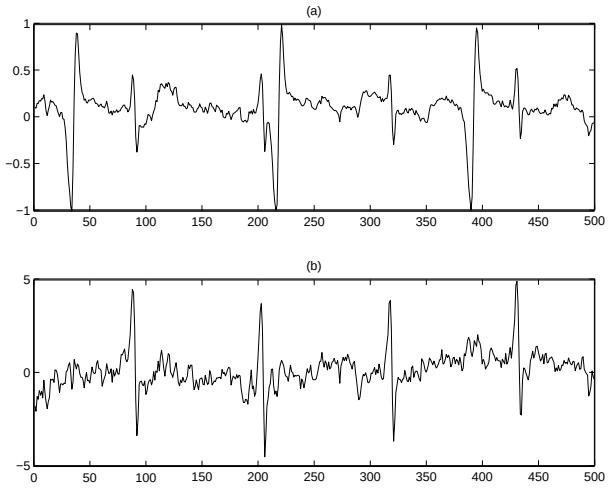

Figure 4: (a) Composite ECG and (b) Extracted FECG.

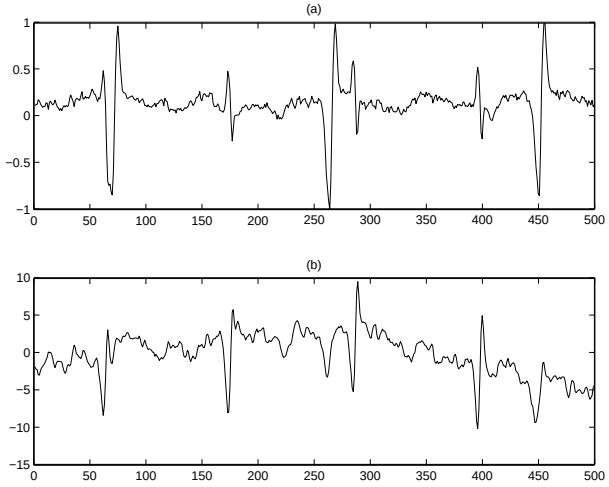

Figure 5: (a) Composite ECG and (b) Extracted FECG.

Discussion

Fig. 6(a) depicts the first recording which contains FECG beats as well as MECG beats. Fig. 6(b)-(c) depict the extracted FECG using the proposed method, Principal Component Analysis and Independent Component Analysis, respectively. As we can observe the proposed method presented comparative results with that of ICA, while both methods presented superior results compared to PCA. The proposed method and ICA supressed the MECG beats very well while the PCA source for FECG contains contributions from the MECG beats.

As reported in [3] only the different sources subspaces have to be separated, instead of all source components. This leads to a reduction of computional cost without loss of medical information. Fig. 7 depicts this situation where only four sources have been estimated instead of the seven sources of previous experiments. As we observe the signal depicted in Fig.7(b) contains the FECG contributions, MECG is presented in Fig.7(c)-(d), while the noise components are concentrated in the signal depicted in Fig.7(a).

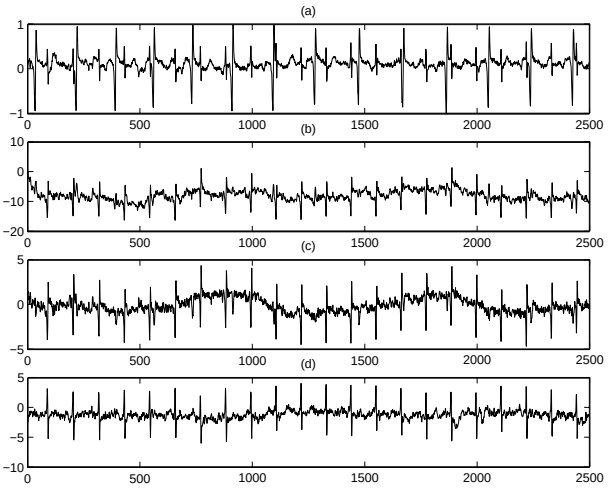

Figure 6: (a) Composite ECG, (b) Extracted FECG using the proposed algorithm, (c) Extracted FECG using PCA and (d) Extracted FECG using ICA

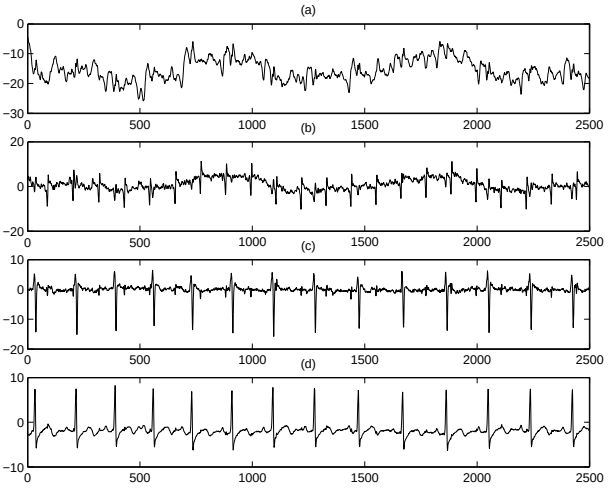

Figure 7: (a)-(d) Extracted sources using the proposed algorithm

In many practical situations the use of a sequential algorithm is essential, i.e. in real time applications. In our case the algorithm iterates between state estimation using Kalman Smoother and parameter identification. To obtain a sequential version of the above algorithm we can replace the Kalman Smoother equations with the Kalman Filter equations. However, this is beyond the scope of present study and is the subject of our future work.

Conclusions

In this paper a method for the extraction of the FECG is presented. The FECG problem formulated as a BSS problem with nonstationary sources. The BSS problem was defined as a special case of a linear dynamical model. As a result, and based on the assumptions we have made,

ISSN: 1727-1983

we finally have an iterative algorithm. In the limiting case the algorithm coincides with conventional PCA. The algorithm was tested on eight leads ECG data and resulted in efficient FECG extraction. The algorithm has shown excellent separation of nonoverlapping FECG and MECG beats, and it was also capable of separating completely overlapping FECG and MECG beats. In the future we intend to introduce a more general formulation of the particular state-space model and the use of variational bayesian method for the estimation of the parameters and the sources, and a sequential version of it.

Acknowledgment

This research was funded by the program "Pythagoras" of the Operational Program for Education and Initial Vocational Training of the Hellenic Ministry of Education under the 3rd Community Support Framework and the European Social Fund.

Appendix A

The Kalman Smoother estimator

$$\mathbf{x}_t^n = E\{\mathbf{x}_t|\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_n\} \tag{15}$$

for the model defined by equations (3, 4) is obtained by minimizing the mean square error

$$P_t^n = E\{(\mathbf{x}_t-\mathbf{x}_t^n)(\mathbf{x}_t-\mathbf{x}_t^n)^T|\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_n\} \tag{16}$$

The estimator can be obtained recursively using the following equations [12]. For $t=1,\cdots,n$ we have:

$$\mathbf{x}_t^{t-1} = \mathbf{x}_{t-1}^{t-1}, \tag{17}$$

$$P_t^{t-1} = P_t^{t-1} + Q, \tag{18}$$

$$K_t = P_t^{t-1}W^T(WP_t^{t-1}W^T+\sigma^2)^{-1}, \tag{19}$$

$$\mathbf{x}_t^t = \mathbf{x}_t^{t-1} + K_t(\mathbf{y}_t - W^T\mathbf{x}_t^{t-1}), \tag{20}$$

$$P_t^t = P_t^{t-1} - K_tWP_t^{t-1}, \tag{21}$$

where we take $\mathbf{x}_0^0=\mu$ and $P_0^0=\Sigma$ and μ,Σ are the initial values of the estimated quantities, arbitralily chosen. To calculate $\mathbf{x}_t^n$ and P_t^n, we use the following set of recursive equations for $t=n,n-1,\cdots,1$

$$J_{t-1} = P_{t-1}^{t-1}(P_t^{t-1})^{-1}, \tag{22}$$

$$\mathbf{x}_{t-1}^n = \mathbf{x}_{t-1}^{t-1} + J_{t-1}(\mathbf{x}_t^n - \mathbf{x}_{t-1}^{t-1}), \tag{23}$$

$$P_{t-1}^n = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}^T. \tag{24}$$

References

[1] WIDROW, B., MCCOOL, J. M., KANMITZ, J., WILLIAMS, C, HEAM, R., ZEIDLER, J., DONG, E., and GOODLIN, R. Adaptive noise cancelling: Principles and applications. *Proc. IEEE*, 63:1692 – 1716, 1975.

[2] ZARDOSO, V. and NANDI, A. K. Noninvasive fetal electrocardiogram extraction: Blind separation versus adaptive noise cancellation. *IEEE Transactions on Biomedical Engineering*, 48:12–18, 2001.

[3] DE LATHAUWER, L., DE MOOR, B., and VANDEWALLE, J. Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Transactions on Biomedical Engineering*, 47:567–572, 2000.

[4] JAFARI, M.G and J.A. CHAMBERS, J.A. Fetal Electrocardiogram Extraction by Sequential Source Separation in the Wavelet Domain. *IEEE Transactions on Biomedical Engineering*, 52:390–400, 2005.

[5] TIPPING, M. E. and BISHOP, C. M. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 21:611–622, 1999.

[6] BISHOP, C. M. Bayesian PCA. *Advances in Neural Processing Systems*, 11:382 – 388, 1999.

[7] DEMPSTER, A., LAIRD, A., and RUBIN, D. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society, Series B*, 39:1–38, 1977.

[8] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. PTR Prentice Hall, Inc., Englewood Cliffs, NJ, 1993.

[9] MACKAY, D. Probable networks and plausible predictions - a review of practical bayesians methods for supervised neural networks. *Network: Computation in Neural Systems*, 6:469–505, 1995.

[10] BISHOP, C. M. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[11] ICA Central. Internet site address: http://sig.enst.fr/ cardoso/icacentral/index.html.

[12] SHUMWAY, R. H. and STOFFER, D. S. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3:253–264, 1982.

ISSN: 1727-1983