A Bayesian Approach for Biomedical Signal Denoising

Vangelis P. Oikonomou
Unit of Medical Technology and
Intelligent Information Systems
Dept. of Computer Science, University of Ioannina
GR 45110 Ioannina, Greece
Email: voikonom@cs.uoi.gr

Dimitrios 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
Email: fotiadis@cs.uoi.gr

***Abstract*— Biomedical signal recordings present different kinds of interference and several noise cancelling methods have been proposed. In this paper we present a bayesian approach for signal denoising. The maximum likelihood solution of this problem is the observations, which is not useful at all. To obtain a meaningfull solution we introduce constraints in our problem. We choose the desired signal to belong to the class of smooth signals. The introduction of constraints lead us to a bayesian formalism of the problem. Often the use of priors introduce hyperpriors in the problem. To infer the parameters and the hyperparameters of the problem the variational bayesian methodology is used. Our method is compared with one widely used method for signal denoising, the spectral substraction. The comparison is made in terms of signal enhancement, noise reduction and signal distortion and the results have shown that our method performs better in noisy environment. Another advantage of our method is that no tunable parameters are used, since it is data driven, which means that the method is fully automated.**

I. Introduction

Signal Enhancement (SE) attempts to improve one or more perceptual aspects when the signal is corrupted by noise (e.g. overall quality, intelligibility of human or machine recognizers). The improvement is related to the minimization of the effects of the noise on a processing system performance [1]. The Signal Enhancement problem consists of a family of problems characterized by the type of noise, the way the noise interacts with the signal and the number of channels available for enhancement.

The Signal Enhancement problem has been attracted much attention in the last two decades. The SE algorithms can be classified according to the signal model as parametric or non parametric and according to the number of channels as multichannel or single channel. In the parametric techniques the signal is modeled using a stochastic autoregressive (AR) model embedded in Gaussian noise. Signal Enhancement is related to the estimation of AR parameters applying a Wiener [2] or a Kalman filter [3], [4] to the noisy signal. Non-parametric techniques do not estimate the signal parameters and require a noise fingerprint in a transform domain (DFT or KLT domain), which is used during signal-and-noise periods to obtain an estimate of the clean signal. Well known non-parametric techniques include spectral substraction [5] and signal subspace - based techniques [6]–[8].

Many of the algorithms proposed to estimate the desired signal formulate the problem as maximum likelihood (ML) estimation problem [9]. However, this approach has one serious drawback, it is sensitive to overfitting. This can be avoided using a bayesian approach, where the priors over the signal and/or the coupling systems will act as regularizers. Bayesian estimation is a framework for the formulation of statistical inference problems. The bayesian philosophy is based on the combination of the evidence contained in the data with prior knowledge. To calculate the evidence one just multiplies the model likelihood by the priors and then intergrate the parameters. In some cases this integral can be computed analytically, but when the evidence integral is analytically intractable one has to resort to approximation techniques such as the Variational Bayesian (VB) Methodology, the Markov Chain Monte Carlo (MCMC) methods and the Laplace Approximation [10]. However, in the Laplace approximation the Gaussian assumption is based on the existence of a large number of data, and the posterior will be presented poorly for a small dataset, besides that we need many operations to compute the derivatives of the Hessian. Similarly, in the MCMC methods the number of samples required for accurate estimates must be large. In addition, the lack of the acceptable global measure indicating if the Markov chain has reached equilibrium is a problem. In contrast, the VB methodology is an efficient computational method because gives closed form solutions and a universally accepted criterion to stop the process, which is the convergence of the variational bound.

In this work we present a bayesian formulation of signal denoising in the presence of white gaussian noise. Given the observations we can obtain the desired signal. We introduced priors, which however make the problem intractable. To overcome it and obtain an approximate solution the Variational Bayesian (VB) formalism is used. In this case the priors which depend on parameters such as mean and variance are either assumed known or can be determined as part of the inference problem [11]. This finally leads us to an hierarchical model from which an estimation for the signal and the parameters can be derived. Our approach is evaluated using ECG recordings, where artificial noise is added, and the results are compared with a widely applied methodology for signal denoising, using well accepted measures.

II. Methodology

We consider a signal $s(k)$ corrupted by additive white gaussian noise $n(k)$. The noisy signal can be expressed as follows:

$$y(k) = s(k) + n(k), \tag{1}$$

where k is the index sample and $k = 1, \cdots, N$ with N being the number of samples. The above equation can be written in vector notation as:

$$\mathbf{y} = \mathbf{s} + \mathbf{n}, \tag{2}$$

where $\mathbf{y} = [y(1), \cdots, y(N)]$. The signal $\mathbf{s}$ is independent of the noise and stationary. The problem is to estimate the signal $\mathbf{s}$. The ML approach in this problem is meaningless because the ML estimator is the observations. So the problem is ill posed. To overcome this difficulty one has to regularize the original problem. Typically, regularization means the introduction of a constraint in the original problem. The constraint is chosen ad hoc or it is based on some a priori information of the quantities to be estimated. We choose to introduce smoothness prior over the signal $\mathbf{s}$ [11]. However, the use of smoothness prior introduce a new parameter, α, in our model. To deal with this parameter we introduce a hyperprior over it. The prior over the signal $\mathbf{s}$ is:

$$p(\mathbf{s}|\alpha) \sim \Big(\frac{\alpha}{2\pi}\Big)^{\frac{N}{2}} \exp\Big\{-\frac{\alpha}{2}\mathbf{s^T L^T L s}\Big\}. \tag{3}$$

The matrix $\mathbf{L}$ is a discrete approximation of the d^{th} derivative operator. In this case we use the second difference matrix. In our problem we assume white gaussian noise, i.e.

$$p(\mathbf{n}|\lambda) = \Big(\frac{\lambda}{2\pi}\Big)^{\frac{N}{2}} \exp\Big\{-\frac{\lambda}{2}\mathbf{n^T n}\Big\}. \tag{4}$$

where λ is the precision of the noise (inverse variance). The two parameters, α and λ, introduced into the problem follow gamma distribution:

$$p(\alpha) = \Gamma(\alpha; b_\alpha, c_\alpha), \tag{5}$$

$$p(\lambda) = \Gamma(\lambda; b_\lambda, c_\lambda), \tag{6}$$

where

$$\Gamma(x; b, c) = \frac{1}{\Gamma(c)}\frac{x^{c-1}}{b^c}\exp(-\frac{x}{b}). \tag{7}$$

The choice of this distribution is based on the fact the Normal and Gamma distributions are conjugates [12]. Now we need to estimate the signal $\mathbf{s}$ and the parameters α and λ. The signal $\mathbf{s}$ is uncorrelated to the noise, and hence to the parameter λ, so the overall prior can be expressed as:

$$p(\mathbf{s}, \alpha, \lambda) = p(\mathbf{s} \mid \alpha)p(\alpha)p(\lambda). \tag{8}$$

III. Variational Bayesian Methodology

Consider the problem of evaluating the marginal likelihood:

$$p(y) = \int p(y, \theta)d\theta, \tag{9}$$

where $\theta = \{\theta_i\}$ denotes the set of all the parameters and hidden variables in the model and y are the observations. Sometimes such integrations are analytically intractable. Variational methods involve the introduction of a distribution $q(\theta)$ which provides an approximation to the true likelihood. The true marginal log-likelihood, then, can be bounded as:

$$\begin{aligned} \ln p(y) &= \ln \int p(y,\theta)d\theta & (10)\\ &= \ln \int q(\theta)\frac{p(y,\theta)}{q(\theta)}d\theta & (11)\\ &\geq \int q(\theta)\ln\frac{p(y,\theta)}{q(\theta)}d\theta & (12)\\ &= F(q). & (13) \end{aligned}$$

In Eq. (12) we have applied Jensen's inequality. The function $F(q)$ forms a lower bound on the true marginal likelihood. The quantity $F(q)$ is tractable through a suitable choise to the q-distribution, even though the true log-likelihood is not. The difference between the true log-likelihood $\ln p(y)$ and the bound $F(q)$ is the Kullback - Leibler (KL) divergence between the approximating distribution $q(\theta)$ and the true posterior $p(\theta|y)$

$$KL(q||p) = -\int q(\theta)\ln\frac{p(\theta|y)}{q(\theta)}d\theta. \tag{14}$$

The goal in a variational approach is to choose a suitable form of $q(\theta)$ so the lower bound can be evaluated. In general, we choose a family of q-distributions and we seek the best approximation within this family by maximizing the lower bound. Since the true log-likelihood is independent of q this is equivalent to the minimization of KL divergence. The KL divergence between the two distribution $q(\theta)$ and $p(\theta|y)$ is minimized when $q(\theta) = p(\theta|y)$ and thus the optimal solution for $q(\theta)$ is the true posterior. This solution does not simplify the problem, so to make progress we consider a more restricted range of q-distribution. One approach is to consider a parametric form for $q(\theta)$ such that $q(\theta, \phi)$, governed by a set of parameters ϕ [13]. We then minimize the KL divergence with respect to ϕ, finding the best approximation within this family. An alternative approach is to restrict the functional form of $q(\theta)$ by assuming that it factorizes over the component variables $\{\theta_i\}$ in θ [10]:

$$q(\theta) = \prod_i q_i(\theta_i). \tag{15}$$

Minimizing the KL divergence over all the factorial distributions $q_i(\theta_i)$, we have the following result:

$$q_i(\theta_i) \propto \exp < \ln p(y,\theta) >_{k\neq i}, \tag{16}$$

where $< \cdot >_{k\neq i}$ denotes expectation with respect to the distributions $q_k(\theta_k)$ for all $k \neq i$. Incorporating the prior knowledge for the parameter θ_i, $p(\theta_i)$, we have:

$$q_i(\theta_i) \propto \exp < \ln p(y,\theta) >_{k\neq i} p(\theta_i). \tag{17}$$

As a result the estimated posterior is proportional to the expected likelihood over all parameters except from parameter θ_i multiplied by the prior of parameter θ_i. In our problem the parameters θ is the signal $\mathbf{s}$ and the parameters α and λ.

Now to apply the VB methodology in our problem we approximate the posterior distribution with the factorized density

$$q(\mathbf{s},\alpha,\lambda \mid \mathbf{y}) = q(\mathbf{s} \mid \mathbf{y},\alpha)q(\alpha \mid \mathbf{y})q(\lambda \mid \mathbf{y}). \tag{18}$$

Maximizing $F(q)$ with respect to $q(\mathbf{s} \mid \mathbf{y},\alpha)$, $q(\alpha \mid \mathbf{y})$ and $q(\lambda \mid \mathbf{y})$ the following solutions are obtained. The posterior over the signal $\mathbf{s}$ is a Normal distribution with mean and covariance $N(\hat{\mathbf{s}},\mathbf{C_s})$:

$$\hat{\mathbf{s}} = \hat{\lambda}\mathbf{C_s}\mathbf{y}, \tag{19}$$
$$\mathbf{C_s} = (\hat{\lambda} + \hat{\alpha}\mathbf{L^T L})^{-1}. \tag{20}$$

The posterior over parameter λ is a Gamma distribution with parameters:

$$\frac{1}{b'_\lambda} = \frac{1}{2}(\mathbf{y^T y} - \mathbf{2y^T}\hat{\mathbf{s}} + Tr(\mathbf{C_s} + \hat{\mathbf{s}}\hat{\mathbf{s}}^T)) + \frac{1}{b_\lambda}, \tag{21}$$
$$c'_\lambda = \frac{N}{2} + c_\lambda, \tag{22}$$
$$\hat{\lambda} = b'_\lambda c'_\lambda. \tag{23}$$

Finally the posterior over parameter α is a Gamma distribution with parameters

$$\frac{1}{b'_\alpha} = \frac{1}{2}Tr(\mathbf{L^T L}(\mathbf{C_s} + \hat{\mathbf{s}}\hat{\mathbf{s}}^T)) + \frac{1}{b_\alpha}, \tag{24}$$
$$c'_\alpha = \frac{N}{2} + c_\alpha, \tag{25}$$
$$\hat{\alpha} = b'_\alpha c'_\alpha. \tag{26}$$

To reduce the complexity of the algorithm the above equations can be written in the Discrete Fourier Domain using the fact of the asymptotic equivalence of the eigenvalues between of Toeplitz and circulant matrices [14]. So in the DFT domain we have:

$$P_s(i) = \frac{1}{(\hat{\lambda} + \hat{\alpha} \mid L(i) \mid^2)}, i = 1,\cdots,N, \tag{27}$$
$$S(i) = \frac{\hat{\lambda}Y(i)}{(\hat{\lambda} + \hat{\alpha} \mid L(i) \mid^2)}, i = 1,\cdots,N, \tag{28}$$
$$\frac{1}{b'_\alpha} = \frac{1}{2}\sum_{i=1}^{N} \mid L(i) \mid^2 (P_s(i) + \frac{1}{N} \mid S(i) \mid^2) + \frac{1}{b_\alpha} \tag{29}$$
$$c'_\alpha = \frac{N}{2} + c_\alpha, \tag{30}$$
$$\hat{\alpha} = b'_\alpha c'_\alpha, \tag{31}$$
$$\frac{1}{b'_\lambda} = \frac{1}{2}\sum_{i=1}^{N}(\frac{1}{N}|Y(i)|^2 - \frac{2}{N}Y^*(i)S(i) + P_s(i) + \frac{1}{N} \mid S(i) \mid^2) + \frac{1}{b_\lambda}, \tag{32}$$
$$c'_\lambda = \frac{N}{2} + c_\lambda, \tag{33}$$
$$\hat{\lambda} = b'_\lambda c'_\lambda. \tag{34}$$

The estimated signal can be obtained by inverse transformation of the quantity $S(i), i = 1,\cdots,N$. The learning algorithm consists of the equations (27)-(34). These equations are applied iteratively until the convergence of the lower bound $F(q)$ or the convergence of the parameters. The computational complexity of this algorithm is low. To obtain an estimate of the signal $\mathbf{s}$ we need to estimate only two parameters, α and λ. This is a very strong feature of the proposed algorithm compared to other algorithms proposed in the literature, which need the estimation of a stationary covariance matrix such as the algorithms based on the subspace approach [6]–[8].

IV. RESULTS

To evaluate our approach we have used two recordings (100 and 222) from the BIH/MIT Arrhythmia database [15]. The noisy signal was created using Eq.(2) for SNR values 0, 5, 10, 15, 20 dB. The algorithm was applied to frames of the noisy signal which are overlapped by 25%. The analysis window was a rectangular window and the length of each frame was $N = 1024$. We compare our method with the spectral substraction method [16]. The compromise between signal distortion and the level of residual noise is a well known problem. To quantify this trade-off we introduce two objective measures. The first measure is the signal distortion index which measures the degree of signal deformation. The other measure is the noise reduction factor which quantifies the amount of noise being attenuated. The signal distortion (SD) index is defined (in dB) as:

$$SD = 10\log_{10}\frac{\sum_{n=0}^{N-1}[s(n) - \hat{s}(n)]^2}{\sum_{n=0}^{N-1}[y(n) - s(n)]^2}. \tag{35}$$

The SD index compares the energy of the difference between the true and the estimated signal, $\hat{s}(n)$, with the energy of the noise. When $SD \rightarrow -\infty$ the estimation of the signal is perfect and when $SD = 0$ the estimated signal is the same with the noisy signal.

The noise reduction is defined (in dB) as:

$$NR = 10\log_{10}\frac{\sum_{n=0}^{N-1}[y(n)-s(n)]^2}{\sum_{n=0}^{N-1}[y(n)-\hat{s}(n)]^2}. \tag{36}$$

The NR index compares the energy of true noise to the energy of the estimated noise. When $NR \rightarrow \infty$ there is no noise reduction and when $NR = 0$ the noise has been fully removed.

Finally, we introduce the total output SNR measure, defined as:

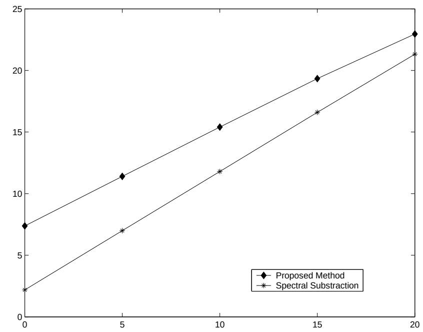
$$SNRtot = 10\log_{10}\frac{\sum_{n=0}^{N-1}s^2(n)}{\sum_{n=0}^{N-1}[s(n)-\hat{s}(n)]^2}. \tag{37}$$

When $SNRtot \rightarrow \infty$ we have complete reconstruction of the signal. In the above equations, $\hat{s}(n)$ is the estimated signal and N the number of samples.

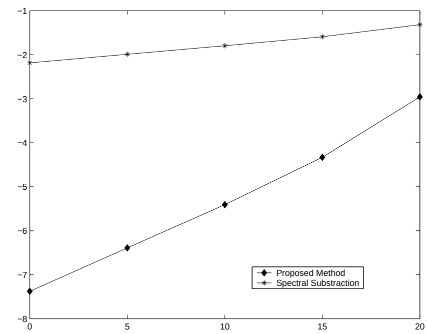
The objective measures for various SNR levels for the two selected recordings are shown in Figs. 1 and 2. We can observe that the results of our approach are better than spectral substraction. In general, our approach produces an estimate of the signal with less signal distortion and higher noise reduction. However, the quality of an estimate of biomedical signal can not be based only in quantitative results. For example, the quality of denoising for ECG recordings depends on the ability to accurately obtain waveforms of the recording such as P wave, QRS complex and T wave, which are usually hidden by noise. We illustrate it in Fig.3, which shows a segment of recording 222 with SNR=0dB, and two estimates obtained by the proposed approach and spectral substraction, respectively. We have shown the results to experienced cardiologists and they agreed that the P wave, QRS complex and T wave are more obvious in the estimate produced by our approach, and the signal itself is less distorted.

V. Conclusions

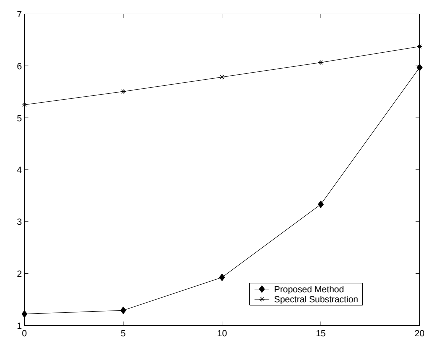
In this work we have presented a method for signal denoising when the noise is white gaussian. The problem is formulated in a bayesian framework and the variational bayesian methodology has been used for its solution, choosing the smoothness prior. Comparing our method with the sell known spectral substraction better results are obtained in terms of signal enhancement, signal distortion and noise reduction, when the two methods are applied in ECG recordings contaminated with various levels of noise. In addition a strong advantage of the proposed approach compared to other methods for signal denoising [6]–[8] or more specifically for ECG denoising [17], is its simplicity. Our method needs to estimate only two parameters, the precision of the noise and an hyperparameter for the signal. Furthermore, it doesn't need a noise fingerprint such as the spectral substraction or an accurate method for the estimation of the noisy signal covariance matrix such as subspace approaches [6]–[8]. In the future we intend to use different priors on the signal, such as the subspace prior [11], and noise conditions, and test the applicability of these models in biomedical signals.



(a) Output SNR Measure



(b) Signal Distortion Measure



(c) Noise Reduction Measure

Fig. 1. Signal Distortion, Noise Reduction and output SNR indexes for record 100. The x-axis represents the level of SNR input noise and the y-axis the values of the measures

Acknowledgments

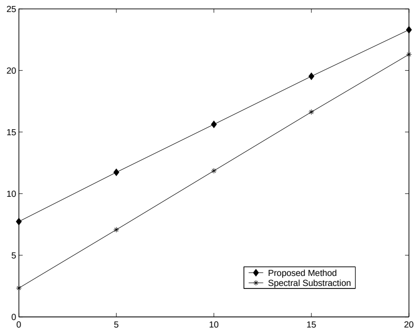
This research was funded by the program "Pythagoras" of the Operational Program for Education and Initial Vocational Training of the Hellenic Ministry of Education under the 3rd Community Support Framework and the European Social Fund.

References

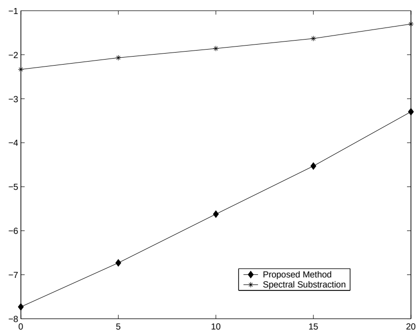
[1] L. Rabiner and R. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice Hall, Inc., 1978.

[2] M. Feder, A. V. Openheim, and E. Weinstein, "Maximum Likelihood Noise Cancellation using the EM Algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, pp. 204 – 216, 1989.

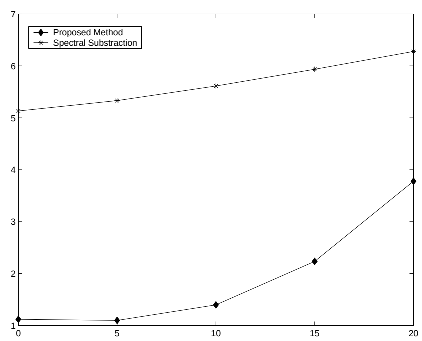
[3] E. Weinstein, A. V. Oppenheim, M. Feder, and J. R. Buck, "Iterative and sequential algorithms for multisensor signal enhancement," *IEEE Transactions on Signal Processing*, vol. 42, pp. 846–859, 1994.



(a) Output SNR Measure



(b) Signal Distortion Measure



(c) Noise Reduction Measure

Fig. 2. Signal Distortion, Noise Reduction and output SNR indexes for record 222. The x-axis represents the level of SNR input noise and the y-axis the values of the measures

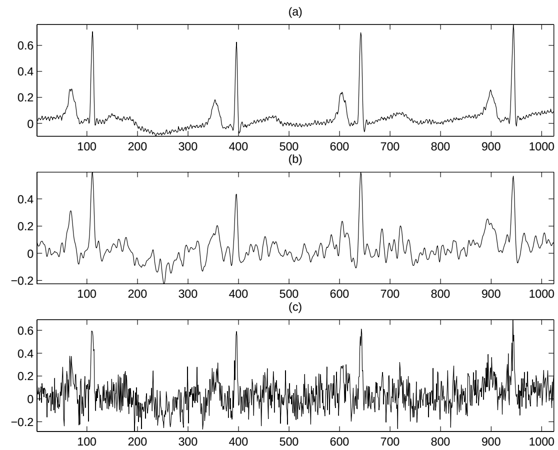


Fig. 3. (a) ECG segment, (b) Estimate using the proposed approach and (c) Estimate using spectral substraction. The x-axis represents time(samples) and the y-axis the amplitude of signal

[4] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and Sequential Kalman Filter - Based Speech Enhancement Algorithms," *IEEE Transactions on Speech and Audio Processing*, vol. 6, pp. 373–385, 1998.

[5] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean - Square Error Short-Time Spectral Amplitude Estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, pp. 1109–1121, 1984.

[6] Y. Ephraim and H. L. V. Trees, "A Signal Subspace Approach for Speech Enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 53–69, 2005.

[7] A. Rezayee and S. Gazor, "An Adaptive KLT Approach for Speech Enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 9, pp. 87–95, 2001.

[8] Y. Hu and P. Loizou, "A Generalized Subspace Approach for Enhancing Speech Corrupted by Colored Noise," *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 334–341, 2003.

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

[10] M. Beal, "Variational algorithms for approximate bayesian inference," Ph.D. dissertation, Gatsby Computational Neuroscience Unit, Univ. College London, London, U.K., 2003.

[11] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*. Springer, 2004.

[12] G. E. P. Box and G. C. Tiao, *Bayesian inference in statistical analysis*. John Wiley and Sons, Inc, 1973.

[13] T. Jaakola, "Variational methods for inference and learning in graphical models," Ph.D. dissertation, Mass.Inst.Technol., Campribge, MA, 1997.

[14] T. Moon and W. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000.

[15] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13), circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215.

[16] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*. John Wiley and Sons Ltd, 2000.

[17] N. Nikolaev, A. Gotchev, K. Egiazarian, and Z. Nikolov, "Supression of electromyogram interference on the electrocardiogram by transform domain denoising," *Medical and Biological Engineering and Computing*, vol. 39, pp. 649–655, 2001.