ECG signal,
weighted averaging,
Bayesian inference

Alina MOMOT*, Michał MOMOT**, Jacek ŁĘSKI**,***

EMPIRICAL BAYESIAN AVERAGING OF BIOMEDICAL SIGNALS

The paper presents empirical Bayesian approach to problem of signal averaging which is commonly used to extract a useful signal distorted by a noise. The averaging is especially useful for biomedical signal such as ECG signal, where the spectra of the signal and noise significantly overlap. In reality can be observed variability of noise power from cycle to cycle which is motivation for using methods of weighted averaging. Performance of the new method is experimentally compared with the traditional averaging by using arithmetic mean and weighted averaging method based on criterion function minimization.

1. INTRODUCTION

In the most of biomedical signal processing systems noise reduction plays very important role. Accuracy of all later operations performed on signal, such as detections or classifications, depends on the quality of noise-reduction algorithms. Using the fact that certain biological systems produce repetitive patterns, an averaging in the time domain may be used for noise attenuation. Traditional averaging technique assumes the constancy of the noise power cycle-wise, however the most types of noise are not stationary. In these cases a need for using weighted averaging occurs, which reduces influence of hardly distorted cycles on resulting averaged signal (or even eliminates them).

The paper presents new method for resolving of signal averaging problem which incorporates empirical Bayesian inference. By exploiting a probabilistic Bayesian framework [1], [4] and an expectation-maximization technique [2] it can be derived an algorithm of weighted averaging which application to electrocardiographic (ECG) signal averaging is competitive with alternative methods as will be shown in the later part of the paper.

Let us assume that in each signal cycle $y_i(j)$ is the sum of a deterministic (useful) signal $x(j)$, which is the same in all cycles, and a random noise $n_i(j)$ with zero mean and variance for the *i*th cycle equal to σ_i^2. Thus, $y_i(j) = x(j) + n_i(j)$, where *i* is the cycle index $i = 1,2,\ldots,M$, and the *j* is the sample index in the single cycle $j = 1,2,\ldots,N$ (all cycles have the same length *N*). The weighted average is given by

---

* Silesian University of Technology, Institute of Computer Science, 16 Akademicka St., 44-101 Gliwice, Poland

** Institute of Medical Technology and Equipment, 118 Roosevelt St., 41-800 Zabrze, Poland

*** Silesian University of Technology, Institute of Electronics, 16 Akademicka St., 44-101 Gliwice, Poland

$$v(j)=\sum_{i=1}^{M} w_i y_i(j)\,, \tag{1}$$

where w_i is a weight for ith signal cycle and $v(j)$ is the averaged signal.

2. SIGNAL AVERAGING METHODS

2.1. ARITHMETIC AVERAGING

The traditional ensemble averaging with arithmetic mean as the aggregation operation gives all the weights w_i equal to M^{-1}. If the noise variance is constant for all cycles, then these weights are optimal in the sense of minimizing the mean square error between v and x, assuming Gaussian distribution of noise. When the noise has a non-Gaussian distribution, the estimate (1) is not optimal, but it is still the best of all linear estimators of x [5].

2.2. WEIGHTED AVERAGING METHOD BASED ON CRITERION FUNCTION MINIMIZATION

As it is shown in [6], for $\boldsymbol{y}_i = [y_i(1), y_i(2),\ldots, y_i(N)]^T$, $\boldsymbol{w} = [w_1, w_2,\ldots, w_M]^T$ and $\boldsymbol{v} = [v(1), v(2),\ldots, v(N)]^T$ minimization the following scalar criterion function

$$I_m(\boldsymbol{w},\boldsymbol{v})=\sum_{i=1}^{M}(w_i)^m \rho(\boldsymbol{y}_i-\boldsymbol{v})\,, \tag{2}$$

where $\rho(\cdot)$ is a measure of dissimilarity for vector argument and $m\in(1,\infty)$ is a weighting exponent parameter, with respect to the weights vector yields

$$w_i=\frac{\left[\rho(\boldsymbol{y}_i-\boldsymbol{v})\right]^{(1-m)^{-1}}}{\sum_{j=1}^{M}\left[\rho(\boldsymbol{y}_i-\boldsymbol{v})\right]^{(1-m)^{-1}}}\,. \tag{3}$$

for $i=1,2,\ldots,M$. When the most frequently used quadratic function $\rho(\cdot)=\|\cdot\|_2^2$ is used, the averaged signal can be obtained as

$$\boldsymbol{v}=\frac{\sum_{i=1}^{M}(w_i)^m \boldsymbol{y}_i}{\sum_{i=1}^{M}(w_i)^m}\,, \tag{4}$$

for the weights vector given by (2) with the quadratic function. The optimal solution for minimization (2) with respect to $\boldsymbol{w}$ and $\boldsymbol{v}$ is a fixed point of (3) and (4) and it is obtained from the Picard iteration.

If m tends to one then the trivial solution is obtained where only one weight, corresponding to the signal cycle with the smallest dissimilarity to averaged signal, is equal to one. If m tend to infinity then weights tend to M^{-1} for all i. Generally, a larger m results in a smaller influence of dissimilarity measures. The most common value of m is 2 which results in greater decrease of medium weights [6].

2.3. EMPIRICAL BAYESIAN WEIGHTED AVERAGING METHOD

Given a data set $y = \{y_i(j)\}$, where i is the cycle index $i = 1,2,\ldots,M$ and the j is the sample index in the single cycle $j = 1,2,\ldots,N$, there are made assumptions that $y_i(j) = x(j) + n_i(j)$, where a random noise $n_i(j)$ is zero-mean Gaussian with variance for the ith cycle equal to σ_i^2, and signal x has also Gaussian distribution with zero mean and covariance matrix $B = \mathrm{diag}(\eta_1^2, \eta_2^2, \ldots, \eta_N^2)$. Thus, from the Bayes rule, the posterior distribution over x and the noise variance is proportional to

$$p(x,\alpha \mid y,\beta) = \frac{p(y \mid x,\alpha)\, p(x \mid \beta)\, p(\alpha)}{p(y)} \propto$$
$$\propto \left(\prod_{i=1}^{M} \alpha_i\right)^{\frac{N}{2}} \prod_{j=1}^{N} \beta_j^{\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N}\big(y_i(j) - x(j)\big)^2 \alpha_i - \frac{1}{2}\sum_{j=1}^{N}\big(x(j)\big)^2 \beta_j\right), \tag{5}$$

where $\alpha_i = \sigma_i^{-2}$ and $\beta_j = \eta_j^{-2}$, because of assumption that the prior $p(\alpha)$ is approximately constant (for large M the influence of this prior is very small). The values x and α which maximize (5) are given by

$$\alpha_i = N\left(\sum_{j=1}^{N}\big(y_i(j) - x(j)\big)^2\right)^{-1} \qquad x(j) = \frac{\sum_{i=1}^{M} \alpha_i y_i(j)}{\beta_j + \sum_{i=1}^{M} \alpha_i}, \tag{6}$$

for $i = 1,2,\ldots,M$ and $j = 1,2,\ldots,N$. Since β_j could not be observed, the iterative EM algorithms is used like in [3]. As values of β_j it is taken

$$E(\beta_j) = \int_0^{\infty} \beta_j\, p(\beta_j \mid x)\, d\beta_j = \frac{\int_0^{\infty} \beta_j\, p(x \mid \beta_j)\, p(\beta_j)\, d\beta_j}{\int_0^{\infty} p(x \mid \beta_j)\, p(\beta_j)\, d\beta_j} = \frac{3}{\big(x(j)\big)^2 + 2\lambda}, \tag{7}$$

assuming an exponential prior $p(\beta_j) = \lambda \exp(-\lambda \beta_j)$ for all *j*. The estimate $\hat{\lambda}$ of hyperparameter λ can be calculated by applying empirical method [7], because

$$E(|x|) = 2\int_0^\infty x p(x|\lambda) dx = 2\int_0^\infty x\lambda\left(x^2 + 2\lambda\right)^{-\frac{3}{2}} dx = \sqrt{2\lambda}, \tag{8}$$

thus

$$\hat{\lambda} = \frac{1}{2}\left(\frac{1}{N}\sum_{j=1}^{N} |x(j)|\right)^2. \tag{9}$$

Therefore the proposed Bayesian weighted averaging algorithm can be described as follows, where ε is a preset parameter:

1. Initialize $v^{(0)} \in R^N$. Set the iteration index $k = 1$.
2. Calculate the hyperparameter $\lambda^{(k)}$ using (9), next $\beta_j^{(k)}$ using (7) for $j = 1,2,\ldots,N$ and $\alpha_i^{(k)}$ using (6) for $i = 1,2,\ldots,M$, assuming $x = v^{(k-1)}$.
3. Update the averaged signal for *k*th iteration $v^{(k)}$ using (6) and $\beta_j^{(k)}$ and $\alpha_i^{(k)}$, assuming $v^{(k)} = x$.
4. If $\left\|v^{(k)} - v^{(k-1)}\right\| > \varepsilon$ then $k \leftarrow k+1$ and go to 2, else stop.

3. NUMERICAL EXPERIMENTS

In all experiments using Weighted Averaging method based on Criterion Minimization Function (WACMF) and Empirical Bayesian Weighted Averaging method (EBWA) calculations were initialised as the means of disturbed signal cycles. Iteration were stopped as soon as the L_2 norm for a successive pair of vectors was less than 10^{-6}, respectively ***w*** vectors for the WACMF and ***v*** vectors for the EBWA. For a computed averaged signal the performance of tested methods was evaluated by the maximal absolute difference between the deterministic component and the averaged signal. The root mean-square error (RMSE) between the deterministic component and the averaged signal was also computed. All experiments were run in the MATLAB environment.

The simulated ECG signal cycles were obtained as the same deterministic component with added realizations of random noise. The deterministic component presented in Fig. 1 was obtained by averaging 500 real ECG signal cycles (2000-Hz and 16-bit resolution) with high signal to noise ratio. Before averaging these cycles were time-aligned using the cross correlation method. A series of 100 ECG cycles was generated with the same deterministic component and zero-mean white Gaussian noise with four different standard deviations. For the first, second, third and fourth 25 cycles, the noise standard deviations were 10, 50, 100, 200 μV, respectively. These signal cycles were averaged using the following methods: Arithmetic Averaging (AA), WACFM with *m* = 2 and EBWA. Subtraction of deterministic component from these averaged signal gives a residual noise.

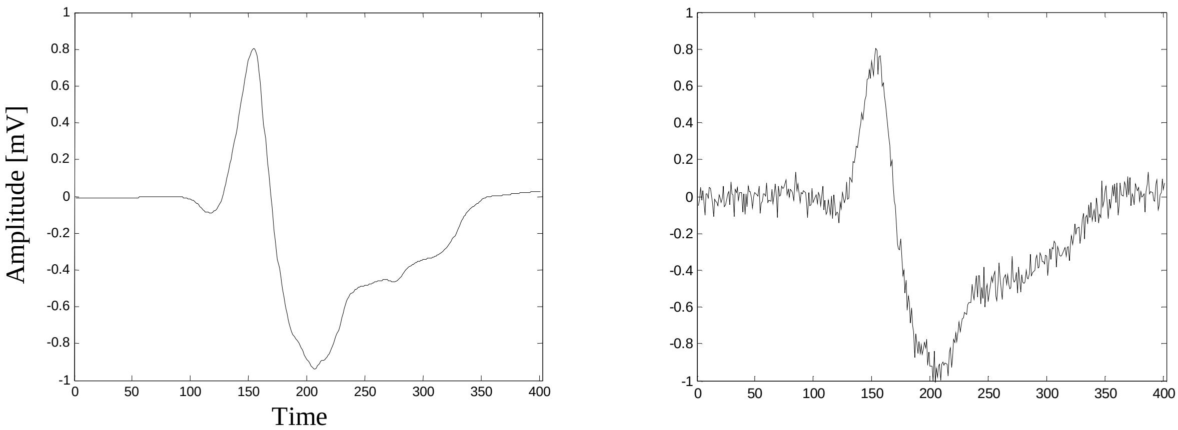

Fig. 1 The simulated ECG signal and this signal with 50μV standard deviation noise.

The RMSE and the maximal value (MAX) of residual noise for all tested method are presented in Table 1. In this table there are also presented result when power of noise was multiplied by 2 and 10 respectively. The best results for each power of noise are bolded. It shows that in all experiments the smallest RMSE were obtained by EBWA method and a little bit worse results were obtained by WACFM, but the smallest MAX error for noise power increased by 10 was obtained by WACFM.

Noise power	Type of error	AA	WACFM	BWA
1×	RMSE [μV]	12.0964	1.9381	**1.9131**
	MAX [μV]	39.1728	5.7307	**5.1671**
2×	RMSE [μV]	24.1927	3.8762	**3.7788**
	MAX [μV]	78.3457	11.4615	**10.3521**
10×	RMSE [μV]	120.9635	19.3810	**17.2270**
	MAX [μV]	391.7285	**57.3073**	58.9807

Tab. 1 RMSE and maximum error for averaged ECG signals with Gaussian noise.

4. CONCLUSION

In this work the new approach to weighted averaging of biomedical signal was presented along with the application to averaging ECG signals. Presented method uses the results of empirical Bayesian methodology which leads to improved reduction of noise comparing with alternative methods. The new method is introduced as Bayesian inference together with expectation-maximization procedure. It is worth noting that the new algorithm does not require setting of additional parameters in contrast to for example WACFM which needs value of an exponential parameter *m*. The only parameter which influences performance of the procedure λ is estimated during iterations from input values by empirical method. The results of numerical experiments show usefulness of the presented method in the noise reduction in ECG signal competitively to existing algorithms.

BIBLIOGRAPHY

[1] CARLIN B., LOUIS T., Bayes and Empirical Bayes Methods for Data Analysis, Chapman & Hall, New York, 1996.
[2] DUDA R., HART P., STORK D., Pattern Classification, John Wiley & Sons, Inc, New York, 2001.
[3] FIGUEIREDO M., Adaptive Sparseness for Supervised Learning, IEEE Transaction on Pattern Analysis and Machine Learning, 25(9), pp. 1150-1159, 2003.
[4] GELMAN A., CARLIN J., STERN H., RUBIN D., Bayesian Data Analysis, Chapman & Hall, New York, 2004.
[5] ŁĘSKI J., Application of time domain signal averaging and Kalman filtering for ECG noise reduction, Ph.D. dissertation, Silesian University of Technology, Gliwice, 1989.
[6] ŁĘSKI J., Robust Weighted Averaging, IEEE Transactions on Biomedical Engineering 49(8), pp. 796-804, 2002.
[7] WHITE H., Estimation, Inference and Specification Analysis, Cambridge University Press, Cambridge, 1996.