ECG signal,
weighted averaging,
Bayesian inference

Alina MOMOT[*], Michał MOMOT[**], Jacek ŁĘSKI[**, ***]

EMPIRICAL BAYESIAN AVERAGING METHOD AND ITS APPLICATION TO NOISE REDUCTION IN ECG SIGNAL

An electrocardiogram (ECG) is the prime tool in non-invasive cardiac electrophysiology and has a prime function in the screening and diagnosis of cardiovascular diseases. However one of the greatest problems is that usually recording an electrical activity of the heart is performed in the presence of noise. 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 the variability of noise can be observed, with power from cycle to cycle, which is motivation for weighted averaging methods usage. It is demonstrated that by exploiting a probabilistic Bayesian learning framework, it can be derived accurate prediction models offering significant additional advantage, namely automatic estimation of 'nuisance' parameters. Performance of the new method is experimentally compared to the traditional averaging by using arithmetic mean and weighted averaging method based on criterion function minimization.

1. INTRODUCTION

In majority of biomedical signal processing systems (as electro-cardiographer signal that is the main case of interest in this contribution) 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. Usually in case of electro-cardiographic (ECG) signal, two principal sources of noise can be distinguished: the 'technical' caused by the physical parameters of the recording equipment and the 'physiological' representing the bioelectrical activity of living cells not belonging to the area of diagnostic interest (also called background activity). Both sources produce noise of random occurrence, overlapping the ECG signal in both time and frequency domains [2].

Using this 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). Therefore, many recently developed noise removal techniques involve weighted signal averaging [8].

The paper presents new method for resolving of signal averaging problem, which incorporates empirical Bayesian inference. By exploiting a probabilistic Bayesian framework [3], [6] and an

[*] 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

expectation-maximization technique [4] it can be derived an algorithm of weighted averaging which application to electro-cardiographic (ECG) signal averaging is competitive with alternative methods as will be shown in the later part of the paper.

2. SIGNAL AVERAGING METHODS

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

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

,where w_i is a weight for *i*th signal cycle and $v(j)$ is the averaged signal.

2.1. TRADITIONAL 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* [7].

2.2. WEIGHTED AVERAGING METHOD BASED ON CRITERION FUNCTION MINIMIZATION

As it is shown in [8], 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

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

with respect to the weights vector $\boldsymbol{w}$ yields

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

for $i = 1,2,\ldots,M$, where $\rho(\cdot)$ is a measure of dissimilarity for vector argument and $m \in (1,\infty)$ is a weighting exponent parameter. 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 [8].

2.3. EMPIRICAL BAYESIAN WEIGHTED AVERAGING METHOD

Given a data set $\boldsymbol{y} = \{y_i(j)\}$, 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 $\boldsymbol{x}$ 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

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

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

$$\alpha_i = N\left(\sum_{j=1}^{N} \big(y_i(j) - x(j)\big)^2\right)^{-1}, \tag{6}$$

$$x(j)=\frac{\sum_{i=1}^{M}\alpha_i y_i(j)}{\beta_j+\sum_{i=1}^{M}\alpha_i} \tag{7}$$

for $i=1,2,\ldots,M$ and $j=1,2,\ldots,N$. The conditions in equations (6) and (7) are obtained by differentiating logarithm of (5) with respect to $\boldsymbol{x}$ and $\boldsymbol{\alpha}$ respectively and setting the results equal to zero.

Since β_j could not be observed, the iterative EM algorithms is used like in [5]. Assuming an exponential prior $p(\beta_j)=\lambda\exp(-\lambda\beta_j)$ for all *j*, as values of β_j it is taken

$$\begin{aligned} E(\beta_j) &= \int_0^\infty \beta_j p(\beta_j \mid x)d\beta_j = \\ &= \frac{\int_0^\infty \beta_j p(x\mid\beta_j)p(\beta_j)d\beta_j}{\int_0^\infty p(x\mid\beta_j)p(\beta_j)d\beta_j} = \\ &= \frac{3}{\left(x(j)\right)^2+2\lambda}. \end{aligned} \tag{8}$$

The estimate $\hat{\lambda}$ of hyperparameter λ can be calculated by applying empirical method [9]. The probability distribution function $p(x\mid\lambda)$ can be written in the form

$$\begin{aligned} p(x\mid\lambda) &= \int_0^\infty \frac{1}{\sqrt{2\pi}}\sqrt{\beta}\exp\left(-\frac{1}{2}x^2\beta\right)\lambda\exp\left(-\lambda\beta\right)d\beta = \\ &= \frac{\lambda}{\sqrt{2\pi}}\int_0^\infty \sqrt{\beta}\exp\left(-\left(\frac{1}{2}x^2+\lambda\right)\beta\right)d\beta = \\ &= \lambda\left(x^2+2\lambda\right)^{-\frac{3}{2}}, \end{aligned} \tag{9}$$

where the index *j* in $x(j)$ and β_j is omitted for clarity.

Since

$$\begin{aligned} E(|x|) &= 2\int_0^\infty xp(x\mid\lambda)dx = \\ &= \sqrt{2\lambda}, \end{aligned} \tag{10}$$

the estimate $\hat{\lambda}$ of hyperparameter λ can be calculated based on first absolute sample moment

$$\hat{\lambda} = \frac{1}{2}\left(\frac{1}{N}\sum_{j=1}^{N}|x(j)|\right)^2 . \tag{11}$$

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 (11), next $\beta_j^{(k)}$ using (8) for $j = 1,2,\dots,N$ and $\alpha_i^{(k)}$ using (6) for $i = 1,2,\dots,M$, assuming $x = v^{(k-1)}$.
3. Update the averaged signal for *k*th iteration $v^{(k)}$ using (7) 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 Minimisation 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 realisations of random noise. The deterministic component presented in Figure 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 was 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: Traditional Arithmetic Averaging (TAA), WACFM with $m = 2$ and EBWA. Subtraction of deterministic component from these averaged signals gives a residual noise.

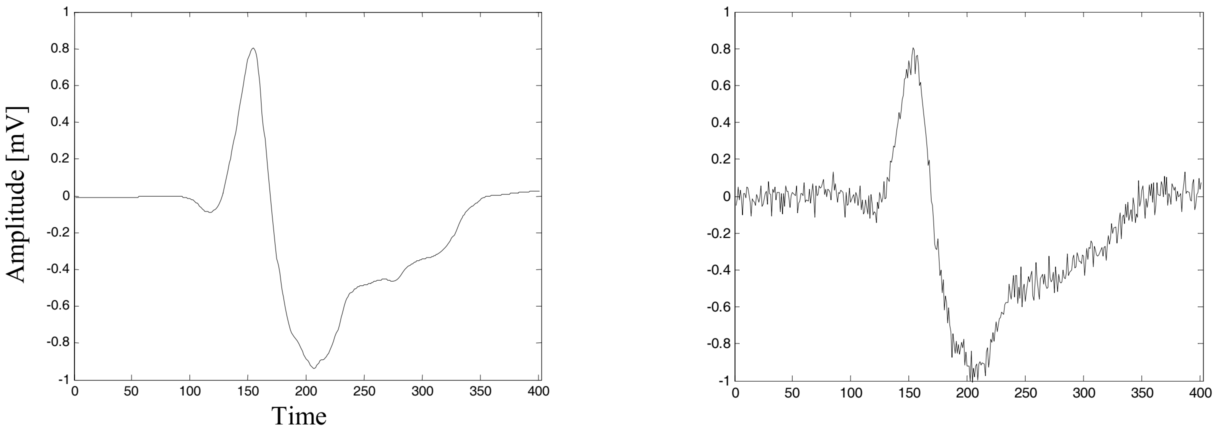

Fig. 1 The example of ECG signal and this signal with 50μV standard deviation noise.

The RMSE and the maximal absolute value (MAX) of residual noise for all tested methods are presented in Table 1. In this table there are also presented results when noise was multiplied by scale factor equal to 2 and 10 respectively. The best results for each case 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 multiplied by 10 was obtained by WACFM.

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

Scale factor	Type of error	TAA	WACFM	EBWA
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

Additionally, the RMSE and the absolute maximum value of residual noise for all tested methods are shown in Figure 2 and Figure 3 respectively. For better comparison the vertical axes present the values of errors in logarithmic scale.

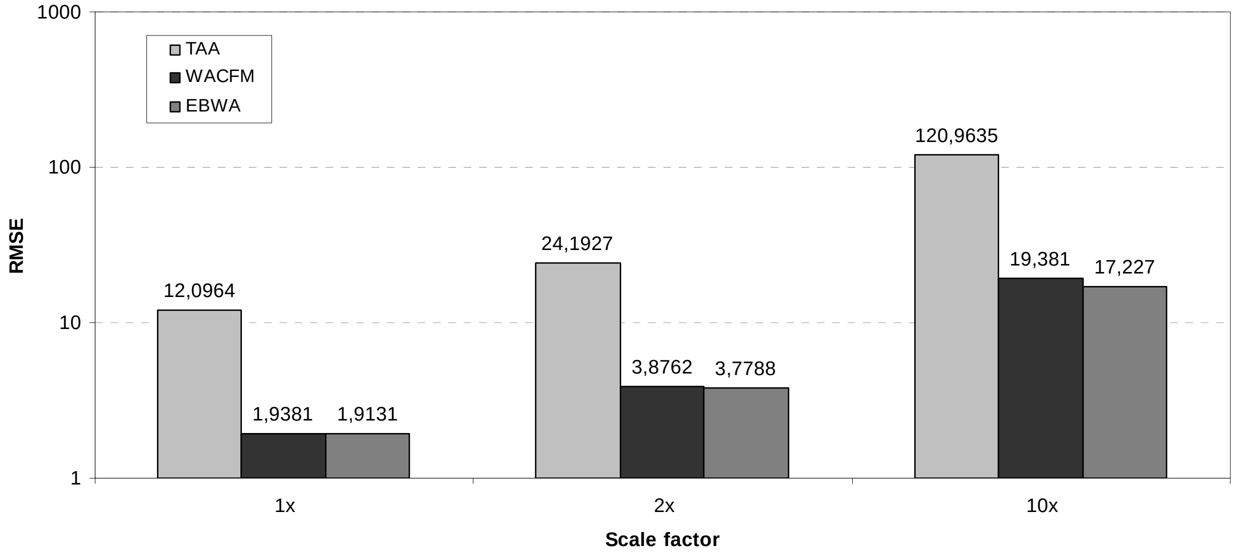

Fig. 2 The root mean-square error between the deterministic component and the averaged signal.

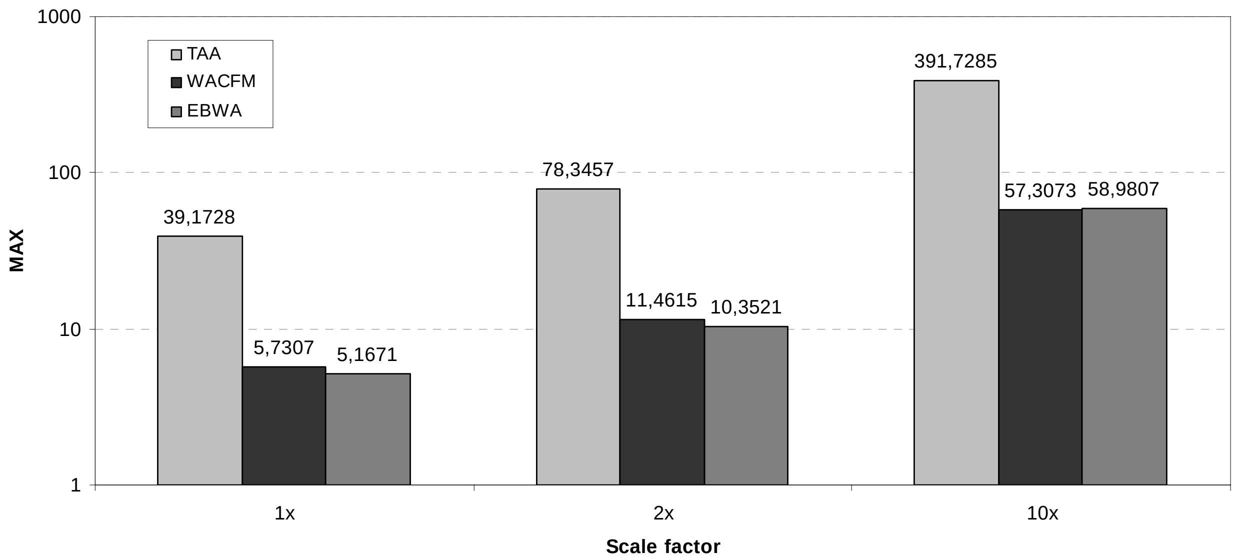

Fig. 3 The absolute maximum error between the deterministic component and the averaged signal

Another experiment was performed with uniformly distributed random noise with zero-mean and 10, 50, 100, 200 μV standard deviations for the first, second, third and fourth 25 cycles, respectively. The RMSE and the maximal absolute value (MAX) of residual noise for all tested methods are presented in Table 2. In this table there are also presented results when noise was multiplied by scale factor equal to 2 and 10 respectively. The best results for each case are bolded.

Table 2. RMSE and maximum error for averaged ECG signals with uniform noise

Scale factor	Type of error	TAA	WACFM	EBWA
1×	RMSE [μV]	11.6650	2.1155	**2.0503**
	MAX [μV]	30.6427	7.1478	**6.1428**
2×	RMSE [μV]	23.3300	4.2310	**4.0775**
	MAX [μV]	61.2854	14.2956	**12.3065**
10×	RMSE [μV]	116.6498	21.1551	**18.6963**
	MAX [μV]	306.4268	71.4781	**62.6855**

As can be seen in the Table 2, for the uniformly distributed random noise, which obviously does not satisfy the assumption for the algorithm, the results of the experiment are similar to those obtained in the case of Gaussian noise. However, in practice besides of Gaussian or uniform types of noise, it can be observed random noise with heavy-tailed distribution [1]. The example of such distribution is Cauchy distribution with probability density function given by

$$f(x) = \frac{1}{\pi s}\left(1+\left(\frac{x-l}{s}\right)^2\right)^{-1}, \tag{12}$$

where l is the location parameter and s is the scale parameter. The parameters are commonly used instead of expected value and standard deviation because of the absence of first two moments. In next experiment to the ECG signal was added Cauchy distributed random noise with l = 0 and s = 10 μV. In Figure 4 the deterministic component was presented along with example of Cauchy noise.

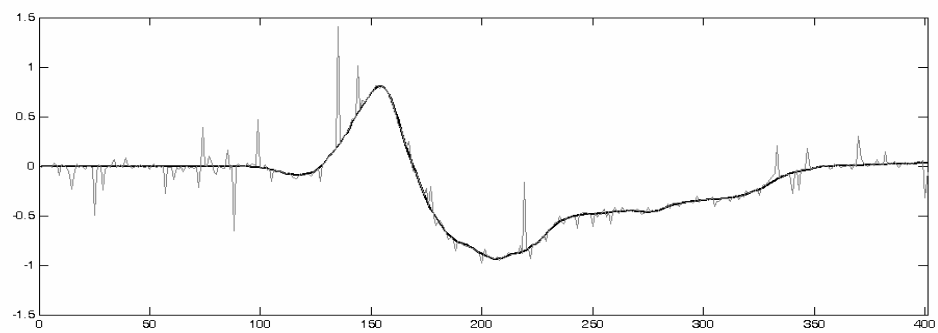

Fig. 4 The simulated ECG signal and this signal with Cauchy noise

The RMSE and the maximal value (MAX) of residual noise for all tested method are presented in Table 3 and the best results are bolded. In this experiment the signal cycles were averaged using WACFM with m = 3, because for most common value of m = 2, the method does not reach stop condition even after 1000 iterations (although in previous cases it did not require more than 20 iterations to stop). It shows that again the smallest RMSE were obtained by EBWA method and a little bit worse results were obtained by WACFM and again despite the fact that the assumption of Gaussian distributed random noise is not satisfied, the results of the experiment are much better compared to the ones obtained by traditional arithmetic averaging.

Table3. RMSE and maximum error for averaged ECG signals with Cauchy noise

Type of error	TAA	WACFM	EBWA
RMSE [μV]	424.4	17.4098	**14.3475**
MAX [μV]	2242.3	59.7734	**42.5693**

4. CONCLUSION

In this work 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 λ which influences performance of the procedure is estimated during iterations from input values by empirical method. Another advantage of presented method is fast convergence to the optimal result. In all performed experiments it did not require more than 12 iterations to stop.

The results of numerical experiments show usefulness of the presented method in the noise reduction in ECG signal competitively to existing algorithms and the empirical Bayesian methodology should be evaluated for other prior distributions in the future.

BIBLIOGRAPHY

[1] ADLER R., FELDMAN R., TAQQU M., A Practical Guide to Heavy Tails. Birkhauser, Boston, 1998.
[2] AUGUSTYNIAK P., Time–frequency modelling and discrimination of noise in the electrocardiogram, Physiological Measurement, 24, pp. 1–15, 2003.
[3] CARLIN B., LOUIS T., Bayes and Empirical Bayes Methods for Data Analysis, Chapman & Hall, New York, 1996.
[4] DUDA R., HART P., STORK D., Pattern Classification, John Wiley & Sons, Inc, New York, 2001.
[5] FIGUEIREDO M., Adaptive Sparseness for Supervised Learning, IEEE Transaction on Pattern Analysis and Machine Learning, 25(9), pp. 1150-1159, 2003.
[6] GELMAN A., CARLIN J., STERN H., RUBIN D., Bayesian Data Analysis, Chapman & Hall, New York, 2004.
[7] ŁĘSKI J., Application of time domain signal averaging and Kalman filtering for ECG noise reduction, Ph.D. dissertation, Silesian University of Technology, Gliwice, 1989.
[8] ŁĘSKI J., Robust Weighted Averaging, IEEE Transactions on Biomedical Engineering 49(8), pp. 796-804, 2002.
[9] WHITE H., Estimation, Inference and Specification Analysis, Cambridge University Press, Cambridge, 1996.