AN APPLICATION OF THE $L_p$-NORM IN ROBUST WEIGHTED AVERAGING OF BIOMEDICAL SIGNALS

Averaging is one of the basic methods of statistical analysis of experimental data where the response of the system is periodic or quasi-periodic. As long as the noise are Gaussian, the standard averaging leads to good results and effective noise reduction. However, when the distortions have impulsive nature, then such an approach leads to a deterioration of the system. In this case the robust methods should be applied which are characterized by resistance to a statistical sample spoken. In this work a robust averaging method based on the minimization of a scalar criterion function using a $L_p$-norm functions are presented. The effectiveness of the proposed method was tested in an averaging periods aligned ECG signal cycles in the presence of impulse noise.

1. INTRODUCTION

Signal averaging is a classic way to improve the signal to noise ratio (SNR) in many biomedical applications [2], [6], [9]. Traditional linear filtering fails when the frequency spectrum of the useful signal and the noise superimposed on a broad range of frequencies. One common example is high resolution analysis of ECG (HRECG) where a surface ECG involves analysis of small segments of a standard ECG in order to detect abnormal cardiac micropotentials, called ventricular late potentials [6].

The classical method of averaging is optimal if the following conditions are satisfied: the signal periods should be centered (aligned in the time domain for the selected centering point), noise should be characterized by a lack of correlation with the averaging signal, has a zero mean value and satisfies the stationarity condition [2]. Let assume that each $i$-th period of the signal is the sum of a deterministic component and a noise $x_i(n) = s(n) + n_i(n)$, where $i$ is the number of cycle ($i=1,\ldots,N$), $n$ is a time index ($n=1,\ldots,M$), $M$ is the signal cycle length, $s(n)$ is the deterministic component, invariant from cycle to cycle, $n(n)$ is the noise uncorrelated with the useful part of processed signal. The averaged signal has the following form:

$$y(n) = \sum_{i=1}^{N} w_i x_i(n),$$

where $w_i$ is the weight of $i$-th cycle and satisfies the following conditions:

$$\forall_{1 \leq i \leq N} \quad w_i \in [0, 1], \quad \sum_{i=1}^{N} w_i = 1.$$
At this assumptions, the expected value of $y(n)$ is the deterministic component $s(n)$. In an ideal situation where the noise has a Gaussian distribution, the signal-to-noise ratio is improved by a factor of $\sqrt{N}$ [1]. Unfortunately, the hypothesis of stationary noise does not always hold, especially in HRECG where the noise variance may change according to muscle activity or acquisition variation [2], [9].

The disadvantage of the ensemble averaging is sensitive to outliers caused by spikes artifacts and bursts of noise that may be related to the activity of skeletal muscles or changing modes of recording devices. Then the noise components have impulsive nature. Non-gaussianity results in significant performance degradation for systems optimized under the Gaussian assumption [15]. For these reasons the robust, weighted averaging method should be applied. The robust averaging method based on minimization a criterion function is presented in [10]. Similar approach to robust averaging based on Bayessian theory is presented in [11].

The objective of this work is to establish the robust method of weighted averaging based on minimization of a certain criterion function which uses the $L_p$-norm and $L_1$-$L_2$-norm as the cost functions. The paper is divided into four sections. Section 2 presents the idea of the weighted averaging method based on the minimization of the scalar criterion function and introduces the proposed methods. Section 3 describes the numerical experiment and contains some results. Finally, the conclusions are given in Section 4.

2. THE IDEA OF THE WEIGHTED AVERAGING

The general idea of weighted averaging rests on the assumption that each cycle of the signal $x_i(n)$ affects the resulting averaged signal $y(n)$ in the manner specified by the value of weight $w_i$. Estimation of the weights values is a crucial for the process of averaging. The weighted averaging method based on criterion function minimization (WACFM) is based on minimization the following scalar function [10]:

$$I_m(w, y) = \sum_{i=1}^{N} w_i^m \rho(z_i),$$

where $z_i = x_i - y$, $m \in (1, \infty)$ is a weighting exponent parameter and the $\rho(\cdot)$ function is a measure of dissimilarity between the averaged signal $y$ and signal cycles $x_i$. The $\rho(\cdot)$ function can be called the cost function. The task of searching for an optimal averaged signal $y^*$ and an optimal weight vector $w^*$, can be formulated as follows $I_m(w^*, v^*) = \min_{w,v} I_m(w, y)$. Minimization of $I_m(w, y)$ can be regarded as the optimization problem with constraints and can be solved using the method of Lagrangian multipliers [10]. If vector of weights $w$ is fixed and satisfy the condition (2), then Lagrangian of expression (3) is written as:

$$L(w, \lambda) = \sum_{i=1}^{N} (w_i)^m \rho(z_i) - \lambda \left[ \sum_{i=1}^{N} w_i - 1 \right],$$

where $\lambda$ is Lagrange multiplier. Setting the Lagrangian’s gradient to zero, we obtain:

$$\forall \ 1 \leq i \leq N \ \frac{\partial}{\partial w_i} L(w, \lambda) = m (w_i)^{m-1} \rho(z_i) - \lambda = 0,$$

and

$$\forall \ 1 \leq i \leq N \ \frac{\partial}{\partial \lambda} L(w, \lambda) = \sum_{i=1}^{N} w_i - 1 = 0.$$

Transforming (5) and (6) yields value of $i$-th weight in the form:

$$w_i = \frac{\rho(z_i)^{1/(1-m)}}{\sum_{j=1}^{N} \rho(z_j)^{1/(1-m)}}.$$
In order to obtain the y averaged signal the gradient of (3) with respect to y is set to zero:

$$\frac{\partial}{\partial y} I_m(w, y) = \frac{\partial}{\partial y} \left( \sum_{i=1}^{N} (w_i)^m \rho(z_i) \right) = 0. \quad (8)$$

In order to solve (8) the influence function $\psi(z) = \frac{\partial \rho(z)}{\partial z}$ and the weight function $\varpi(z) = \frac{\psi(z)}{z}$ known from the theory of maximum likelihood estimator (M-estimator) are introduced [5]. Combining it with (8) yields the iterative form of the averaged vector y:

$$y_{(k+1)} = \frac{\sum_{i=1}^{N} (w_i)^m \cdot \varpi(z_{(k)}) \cdot x_i}{\sum_{i=1}^{N} (w_i)^m \cdot \varpi(z_{(k)})}, \quad (9)$$

where $z_{(k)} = x_i - y_{(k)}$ and superscript $k$ denotes the iteration number. The algorithm is taken as convergent when the condition is satisfied

$$\|w_{(k+1)} - w_{(k)}\| < \epsilon \quad (10)$$

and $\epsilon$ is a small positive value (e.g. $\epsilon = 10^{-6}$). The robust properties of (9) strictly depend on a selection of the $\rho(\cdot)$ function. This function should be symmetric, positive-definite function with a unique minimum at zero and its influence function needs to be bounded for a robust estimate [5]. The $\rho(\cdot)$ dissimilarity function behaves like a distance measure between $x_i$ and $y_i$. In literature there are number of distance measure which can be used like $L_2$-norm when the data are distributed according to a Gaussian distribution. However, it is prone to outliers which dominate the objective function. To overcome this problem the robust metrics should be applied. There are norms that allow to reduce effect of outliers like $L_1$-norm, $L_1$-$L_2$-norm or generally $L_p$-norm [12].

3. $L_p$-NORM AND WEIGHTED AVERAGING METHODS

Let the $L_p$ norm be defined as follows:

$$\|z\|_p = \left( \sum_{i=1}^{s} |z_i|^p \right)^{\frac{1}{p}}, \quad (11)$$

where $z$ is an s–dimensional real vector (i.e. $z \in \mathbb{R}^s$). As the $\rho(\cdot)$ dissimilarity function the $L_p$-norm to the $p$ power is used and then we get $\rho(z) = ||z||_p^p$, where $||\cdot||_p^p$ is the $L_p$-norm to the $p$ power. Finally, we obtain:

$$I_m(w, y) = \sum_{i=1}^{N} w_i^m ||z_i||_p^p. \quad (12)$$

For simplicity the following form of the $\rho(\cdot)$ function and corresponding the influence and the weight functions are used in this paper:

$$\rho(z) = \frac{|z|^p}{p}, \quad \psi(z) = \text{sgn}(z)|z|^{p-1} \quad \text{and} \quad \varpi(z) = |z|^{p-2}. \quad (13)$$

Using (7) and (13) the w weights vector is calculated as:

$$u_i = \frac{\left( |z_i|^p \right)^{2/(1-m)}}{\sum_{j=1}^{N} \left( |z_j|^p \right)^{2/(1-m)}}, \quad (14)$$

and combining (9) and (13) the y averaged signal is given as:

$$y_{(k+1)} = \frac{\sum_{i=1}^{N} (w_i)^m \cdot |z_{(k)}|^p \cdot x_i}{\sum_{i=1}^{N} (w_i)^m \cdot |z_{(k)}|^p}. \quad (15)$$

The algorithm is taken as convergent when the condition (10) is satisfied. In this paper $1 \leq p \leq 2$ and the method is denoted as WACFM$L_p$ (the weighted averaging based on criterion function minimization with using $L_p$-norm). There are two special cases for $p = 1$ and $p = 2$ described in next two subsections.
3.1. THE L₁-NORM

The L₁-norm \((p = 1)\), sometimes called the Manhattan norm, uses the absolute value function \(\rho(z) = |z|\). This norm is well known as the median function which is optimized under Laplace distribution of noise [8], [16]. The influence function and the weighted functions have the forms:

\[
\psi(z) = \text{sgn}(z) \quad \text{and} \quad \varpi = \frac{1}{|z|}.
\]  
\((16)\)

According to (7) and the given absolute value cost function, the \(w\) weights vector is calculated as:

\[
\forall 1 \leq i \leq N \quad w_i = \frac{|z_i|^{1/(1-m)}}{\sum_{j=1}^{N} |z_j|^{1/(1-m)}}.
\]  
\((17)\)

The \(y\) averaged signal is obtained according to (9) and (16) in the following form:

\[
y_{(k+1)} = \frac{\sum_{i=1}^{N} (w_i)^m \cdot x_i}{\sum_{i=1}^{N} (w_i)^m \cdot \frac{1}{|z_i|(k)}}.
\]  
\((18)\)

The algorithm is taken as convergent when the condition (10) is satisfied. This method is called the weighted averaging based on the criterion function minimization with the absolute value cost function (WACFMed).

3.2. THE L₂-NORM

The L₂-norm \((p = 2)\) corresponds to Euclidean measure and then the square function is used as the dissimilarity measure \(\rho(z) = z^2\) (or the cost function if we use term from the robust techniques) [10]. The \(y\) averaged signal can be obtained using the weighted function that for this case has the simplest form \(\varpi(z) = 1\) and then the \(y\) averaged signal is given as:

\[
y = \frac{\sum_{i=1}^{N} (w_i)^m \cdot x_i}{\sum_{i=1}^{N} (w_i)^m},
\]  
\((19)\)

and the \(w\) weights vector is calculated as:

\[
\forall 1 \leq i \leq N \quad w_i = \frac{(z_i)^{2/(1-m)}}{\sum_{j=1}^{N} (z_j)^{2/(1-m)}}.
\]  
\((20)\)

The algorithm is taken as convergent when the condition (10) is satisfied. This method is called the weighted averaging based on the criterion function minimization (WACFM). Unfortunately, the influence function of this cost function is not bounded. It causes a lack of resistance to outliers.

3.3. THE L₁-L₂-NORM

There exists combinations of metrics L₁ and L₂ which is known as L₁-L₂-norm [3]. This measure takes both the advantage of the L₁-norm to reduce the influence of outliers and that of L₂-norm to be convex. The following form of the \(\rho(\cdot)\) function and corresponding the influence and the weight functions are used in this paper:

\[
\rho(z) = 2 \left( \sqrt{1 + z^2/2} - 1 \right), \quad \psi(z) = \frac{z}{\sqrt{1 + z^2/2}} \quad \text{and} \quad \varpi(z) = \frac{1}{\sqrt{1 + z^2/2}}.
\]  
\((21)\)
According to (7) and the cost function (21), the \( w \) weights vector is calculated as:

\[
\forall 1 \leq i \leq N \quad w_i = \frac{2 \left( \sqrt{1 + z_i^2/2} - 1 \right)^{1/(1-m)}}{\sum_{j=1}^{N} \left( 2 \left( \sqrt{1 + z_j^2/2} - 1 \right)^{1/(1-m)} \right)}.
\] (22)

The \( y \) averaged signal is obtained according to (9) and the weight function (21) in the following form:

\[
y_{(k+1)} = \frac{\sum_{i=1}^{N} (w_i)^m \cdot x_i \cdot \frac{1}{\sqrt{1+z_i^2(k)/2}}}{\sum_{i=1}^{N} (w_i)^m \cdot \frac{1}{\sqrt{1+z_i^2(k)/2}}}. \] (23)

The algorithm is taken as convergent when the condition (10) is satisfied. This method is called the weighted averaging based on the criterion function minimization with the absolute value cost function (WACFM\( L_{12} \)).

4. RESULTS AND DISCUSSION

In this section there is presented the performance of the described above methods. All the methods based on minimization of scalar criterion function are initialized with the vector of all ones and \( m = 2 \) which results in a greater decrease of medium weights [10]. For testing requirements the ECG signal from [10] is chosen. This signal is obtained by averaging 500 real ECG cycles (sampled at 2 kHz with 16-bit resolution) with a high signal-to-noise ratio (Fig. 1(a)). Before averaging these cycles are time-aligned.

The examined methods are evaluated using the root mean squared error (RMSE) between the averaged signal and the deterministic cycle and the maximum (MAX) absolute difference between the averaged signal and the deterministic signal. These indicators allow the evaluation of the effectiveness of the noise reduction. All experiments were calculated in an Mathworks Matlab 2010b. The proper selection of the \( p \) parameter for the WACFM\( L_p \) and a particular type of noise is a very difficult task, as appropriate optimization criteria are not known [12]. In this work this parameter is selected in empirical way and \( p = 1.5 \) in all cases the calculations. The reference methods are the WACFM method and the trimmed mean (TM, with trimming parameter 25%).

The purpose of this experiment is to investigate the proposed in this paper methods in the presence of impulsive noise. This kind of noise is modelled by the symmetrical \( \alpha \)-stable (S\( \alpha \)S) distribution [7], [14]. In order to simulate the real conditions of acquisition a series of 100 ECG cycles are generated with the same deterministic component and an impulsive noise with known five values of the \( \gamma \) parameter

![Fig. 1. Clean ECG cycle (a) and ECG cycle with impulsive type of noise (\( \alpha = 1.6, \gamma = 50\mu V \)) (b).](image-url)
(known as the scale parameter or dispersion) which behaves like variance for Gaussian distribution [4]. To maintain compliance with the $\sigma^2$ variance of Gaussian distribution and the case where $\alpha = 2$ for $\alpha$-stable distributions, it is assumed that $\sigma^2 = \gamma^2 / 2$. For the first, second, third, fourth and fifth 20 cycles, the $\gamma$ dispersion values are 10, 50, 100, 200 and 500 $\mu V$. The level of impulsiveness in SαS process is controlled with the $\alpha$ characteristic exponent and in this paper $\alpha$ changes from 1 to 2 with step 0.1. An example of ECG cycle with the impulsive noise is presented in the Fig. 1(b). Such disturbed cycles of ECG signal have been averaged. The results are presented in Table 1.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>TM</th>
<th>WACFM</th>
<th>WACFM$_{ed}$</th>
<th>WACFM$<em>{L</em>{12}}$</th>
<th>WACFM$_{L_p}$</th>
<th>TM</th>
<th>WACFM</th>
<th>WACFM$_{ed}$</th>
<th>WACFM$<em>{L</em>{12}}$</th>
<th>WACFM$_{L_p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td><strong>21.55</strong></td>
<td>115.05</td>
<td>73.19</td>
<td>82.88</td>
<td>71.21</td>
<td>68.78</td>
<td>3931.71</td>
<td>695.59</td>
<td>2633.96</td>
<td>1983.80</td>
</tr>
<tr>
<td>1.1</td>
<td><strong>18.83</strong></td>
<td>43.72</td>
<td>31.99</td>
<td>45.29</td>
<td>25.15</td>
<td>63.67</td>
<td>584.60</td>
<td>332.50</td>
<td>250.74</td>
<td>150.55</td>
</tr>
<tr>
<td>1.2</td>
<td>17.83</td>
<td>29.62</td>
<td>20.67</td>
<td>17.01</td>
<td>18.47</td>
<td>81.72</td>
<td>403.94</td>
<td>230.74</td>
<td>150.55</td>
<td>213.52</td>
</tr>
<tr>
<td>1.3</td>
<td>16.83</td>
<td>24.29</td>
<td>13.71</td>
<td>18.90</td>
<td><strong>12.26</strong></td>
<td>58.85</td>
<td>410.16</td>
<td>139.70</td>
<td>299.62</td>
<td>161.10</td>
</tr>
<tr>
<td>1.4</td>
<td>15.32</td>
<td>10.00</td>
<td>8.97</td>
<td>9.30</td>
<td><strong>7.43</strong></td>
<td>50.39</td>
<td>135.47</td>
<td>43.58</td>
<td>116.54</td>
<td>46.10</td>
</tr>
<tr>
<td>1.5</td>
<td>15.75</td>
<td>7.86</td>
<td>6.96</td>
<td>7.60</td>
<td><strong>6.33</strong></td>
<td>47.54</td>
<td>89.97</td>
<td>84.74</td>
<td>84.05</td>
<td>49.96</td>
</tr>
<tr>
<td>1.6</td>
<td>14.65</td>
<td>4.95</td>
<td>4.40</td>
<td>4.97</td>
<td><strong>4.24</strong></td>
<td>45.54</td>
<td>29.69</td>
<td>31.67</td>
<td>30.37</td>
<td>23.90</td>
</tr>
<tr>
<td>1.7</td>
<td>14.12</td>
<td>4.62</td>
<td>4.31</td>
<td>4.64</td>
<td><strong>3.97</strong></td>
<td>58.31</td>
<td>29.53</td>
<td>40.99</td>
<td>30.02</td>
<td>23.84</td>
</tr>
<tr>
<td>1.8</td>
<td>14.33</td>
<td>3.24</td>
<td>3.10</td>
<td>3.24</td>
<td><strong>3.06</strong></td>
<td>47.69</td>
<td>18.28</td>
<td>15.41</td>
<td>18.29</td>
<td>16.81</td>
</tr>
<tr>
<td>1.9</td>
<td>14.20</td>
<td>2.72</td>
<td>2.93</td>
<td>2.71</td>
<td><strong>2.62</strong></td>
<td>47.94</td>
<td>11.38</td>
<td>34.69</td>
<td>11.36</td>
<td>14.81</td>
</tr>
<tr>
<td>2.0</td>
<td>13.86</td>
<td>2.30</td>
<td>2.28</td>
<td>2.30</td>
<td><strong>2.29</strong></td>
<td>48.29</td>
<td>8.75</td>
<td>8.92</td>
<td>8.75</td>
<td>8.80</td>
</tr>
</tbody>
</table>

The obtained results show that for $\alpha = 1.0$ and 1.1 the best results are obtained with the trimmed mean. This can be explained by the fact that at this level of impulsiveness any changes of the $\gamma$ dispersion are meaningless. But for $\alpha \geq 1.3$ the best noise reduction in process of the weighted averaging is obtained by the WACFM$_{L_p}$ method. This method is "localized" between the WACFM and WACFM$_{ed}$ methods but requires the knowledge of $p$ value. In the case of Gaussian noise ($\alpha = 2.0$), the differences between all investigated methods are very small. The results obtained with the WACFM$_{L_{12}}$ method are better than for the WACFM method and worse than for the WACFM$_{ed}$ for $1.3 \leq \alpha \leq 1.5$. But when $\alpha \geq 1.6$ the obtained results with the WACFM$_{L_{12}}$ method are similar to that obtained with the WACFM method. The advantage of the WACFM$_{L_{12}}$ method is that it requires no additional parameters.

In the second experiment, the set of ECG cycles has been re-created, where the first, second, third and fourth of 25 cycles of the ECG signal are disturbed with noise of a given level of impulsiveness $\alpha$.
= 1.7, 1.8, 1.9 and 2.0 for one value of $\gamma$. Such set of signals were generated for the following values of $\gamma = 10, 50, 100, 200$ and 500 $\mu$V. Then a noisy cycles of ECG have been averaged. The results are presented in Table 2.

Table 2. The values of the performance indices obtained with the proposed methods using the artificial impulsive noise for the second experiments (the best results for RMSE are bolded).

<table>
<thead>
<tr>
<th>$\gamma [\mu V]$</th>
<th>TM WACFM $p = 2$</th>
<th>WACFM$_{ed}$ $p = 1$</th>
<th>WACFM$_{L12}$ $p = 1.5$</th>
<th>TM WACFM $p = 2$</th>
<th>WACFM$_{ed}$ $p = 1$</th>
<th>WACFM$_{L12}$ $p = 1.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.04</td>
<td>1.34</td>
<td>1.27</td>
<td>1.54</td>
<td>1.27</td>
<td>1.18</td>
</tr>
<tr>
<td>50</td>
<td>5.22</td>
<td>6.43</td>
<td>6.36</td>
<td>6.15</td>
<td>20.49</td>
<td>24.45</td>
</tr>
<tr>
<td>100</td>
<td>10.71</td>
<td>13.18</td>
<td>12.85</td>
<td>12.61</td>
<td>36.67</td>
<td>57.83</td>
</tr>
<tr>
<td>200</td>
<td>20.89</td>
<td>26.43</td>
<td>25.68</td>
<td>25.37</td>
<td>66.31</td>
<td>78.21</td>
</tr>
<tr>
<td>500</td>
<td>50.73</td>
<td>63.14</td>
<td>61.28</td>
<td>60.07</td>
<td>183.09</td>
<td>230.22</td>
</tr>
</tbody>
</table>

The obtained results in the second experiments are quite different than for the first experiments. The highest level of noise reduction is obtained for the trimmed mean. But differences are relatively small. Such results can be explained by the fact that the $\gamma$ dispersion of averaged cycles is constant and impulsiveness variation does not affected the result of averaging.

The last numerical experiment concerns the real muscle noise. This type of noise frequently has the impulsive nature. According to the first experiment, for the first, second, third, fourth and fifth 20 cycles, the Geometric-SNR (GSNR) values are 0, 5, 10, 15, 20 dB [14]. The results are presented in Table 3.

Table 3. The values of the performance indices obtained with the proposed methods using the real muscle noise for the third experiments (the best results for RMSE are bolded).

| RMSE [$\mu V$] | WACFM $p = 2$ | TM WACFM $p = 2$ | WACFM$_{ed}$ WACFM$_{L12}$ WACFM$_{Lp}$ |
|----------------|----------------|------------------|------------------|------------------|------------------|
| 23.67          | 14.26          | 23.67            | WACFM$_{ed}$     | WACFM$_{L12}$   | WACFM$_{Lp}$    |
| 72.35          | 43.00          | 14.24            | 13.25            | 14.24            | 13.25            |

The results from the last experiments show that the trimmed mean is insufficient in averaging of ECG cycles. The best results of averaging for this type of noise is obtained for the WACFM$_{ed}$. But other WACFM$_*$ methods provide very similar results. The weighted averaging methods offers high performance of averaging. The obtained results show that the family of WACFM methods are useful when the set of average cycles characterized by the dispersion variation and the level of noise impulsiveness is constant.

5. CONCLUSIONS

This paper presents a new approach to the weighted averaging method based on the minimization of a criterion function introducing different types of norms as a dissimilarity function. The principle of this method is similar to the clustering method especially the fuzzy clustering method. The presented measures ($L_p$-norm, $L_1$-$L_2$-norm) allow a significant reduction of impulsive noise compared with the reference methods, like $L_2$-norm or the trimmed mean, especially in the case of significant dispersion variation. All methods were evaluated in the presence of artificial impulsive noise and the real muscle noise as well. Numerical experiments have shown that the proposed weighted averaging methods lead to better results than the reference methods.
BIBLIOGRAPHY


