Prewhitening Algorithms of Signals in the Presence of White Noise

Andrzej Cichocki, Stanisław Osowski, Krzysztof Siwek

***Abstract* — The paper presents and compares the performance of four different prewhitening algorithms of the signals in the presence of white noise. They have been applied to the decorrelation of the mixed signals and may find application in the solutions of the blind source separation systems.**

I. Introduction

The temporal, spatial or temporal-spatial decorrelation (prewhitening) plays an important role in signal processing. Prewhitening is often necessary condition for the stronger stochastic independence criteria. It is for example the basic step in blind separation (BSS) or independent component analysis (ICA) tasks. After prewhitening the BSS or ICA become somewhat easier and less ill conditioned, because the subsequent separating system is described by an orthonormal matrix for real valued signals. Furthermore the decorrelation technique can be used to identify the mixing matrix and perform blind signal separation for colored signals.

In this paper we present and compare four methods of prewhitening of the signals in the presence of noise. Three of them are based on the modified eigenvalue analysis and one applies the Gram-Schmidt orthogonalization.

II. Basic prewhitening algorithm

A random, zero mean vector $\mathbf{y} \subset R^n$ of dimension n is said to be white if its covariance (correlation) matrix is an identity matrix, i.e., $\mathbf{R}_{yy}=E\{\mathbf{yy}^T\}=\mathbf{1}_n$. The white signals form the correlation matrix which is diagonal. It means that such signals are not correlated with each other. Any set of vectors $\mathbf{x} \subset R^m$ can be decorrelated (whitened) by applying some preprocessing stage. The whitening procedure is equivalent to the linear transformation of the vector $\mathbf{x}$. The whitened vector $\mathbf{y}$ is described by the relation

$$\mathbf{y} = \mathbf{Wx} \tag{1}$$

where $\mathbf{W}$ is an $n \times m$ whitening matrix. If $n<m$ the matrix $\mathbf{W}$ simultaneously reduces the dimension of the data vectors from m to n. If $n=m$ the size of the whitened vector is the same as original one. The vectors $\mathbf{y}$ are mutually uncorrelated and have unit variance. It means that

$$\mathbf{R}_{yy} = E\{\mathbf{yy}^T\} = E\{\mathbf{Wxx}^T\mathbf{W}^T\} = \mathbf{WR}_{xx}\mathbf{W}^T = \mathbf{1}_n \tag{2}$$

In general situation the measured sensor signals are mutually correlated, i.e., the covariance matrix $\mathbf{R}_{xx}$ is not diagonal one. However it is always symmetric and usually positive definite. It means that it can be decomposed using the eigenvalue decomposition as follows

$$\mathbf{R}_{xx} = \mathbf{V}_x\mathbf{L}_x\mathbf{V}_x^T = \mathbf{V}_x\mathbf{L}_x^{1/2}\mathbf{L}_x^{1/2}\mathbf{V}_x^T \tag{3}$$

where $\mathbf{V}_x$ is an orthogonal matrix and $\mathbf{L}_x$ is a diagonal matrix with all nonnegative eigenvalues λ_i, that is $\mathbf{L}_x=\text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$. The columns of the matrix $\mathbf{V}_x$ are the eigenvectors corresponding to the appropriate eigenvalues. Thus, assuming that the covariance matrix is positive definite, the required decorrelation matrix $\mathbf{W}$ can be computed as follows

$$\mathbf{W} = \mathbf{L}_x^{-1/2}\mathbf{V}_x^T \tag{4}$$

If some eigenvalues of $\mathbf{R}_{xx}$ are zero we can take only positive eigenvalues and the eigenvectors associated with them.

a) b)

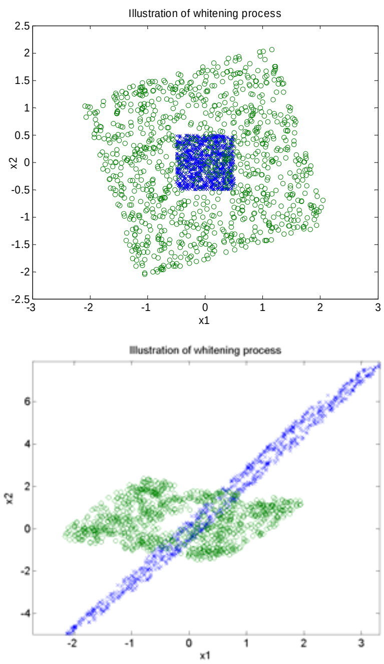

Fig.1 Scatter plots illustrating the whitening transformation for two sensor signals

Fig. 1 presents the results of whitening of two input signals generated according to different distribution. Fig. 1a corresponds to two randomly generated signals

Authors are with the Institute of the Theory of Electrical Engineering, Measurement and Information Systems, Warsaw University of Technology, Warsaw, Poland. A. Cichocki is also with Brain Science Institute RIKEN, Japan, S. Osowski is also with Military University of Technology, Warsaw, e-mail: sto@iem.pw.edu.pl
The work has been supported by Polish Ministry of Science (KBN) as the research project.

and Fig. 1b corresponds to 2 deterministic sinusoidal signals disturbed by white noise of normal distribution. The points denoted by **x** are original points and the symbols o are associated with whitened signals. As it is seen the whitened signals are distributed in a wider space (the typical situation for the uncorrelated signals).

III. THE MODIFIED PREWHITENING ALGORITHMS

The main problem of whitening is due to the noise that is usually contained in the measured values. Let us assume the noisy signal **x**=**s**+**n**, where **s** is the signal and **n** – the random white noise of the standard deviation σ_n. The eigenvalues corresponding to the noise are usually very small, so their inverse very high. It means that the whitening algorithm described by the relations (3)-(4) amplifies the noise in the transformed signals. To process the noisy data we need some modifications of the whitening procedure. In this paper we present three different modified versions of the prewhitening algorithms and compare their performance at the presence of the noise.

A. Algorithm with bias removal

The most important point in this approach is removing the estimated noise components of the signal. Let us denote the noise variance in the system by σ_n^2. It is easy to show that at random white noise, uncorrelated with the signal **s**, the autocorrelation matrix $\mathbf{R}_{xx}$, calculated in a standard way (without delays) may be presented in the form

$$\mathbf{R}_{xx} = \mathbf{R}_{vv} + \mathbf{R}_{nn} \tag{5}$$

where $\mathbf{R}_{vv}$ is the autocorrelation matrix of signal **v=As** (**A** – the mixing matrix) and $\mathbf{R}_{nn}$ – the autocorrelation matrix corresponding to the noise **n**, i.e., $\mathbf{R}_{nn} = E\{\mathbf{nn}^T\}$. Taking into account the uncorrelated character of the noise, we can estimate the autocorrelation matrix of the signal as

$$\mathbf{R}_{vv} = \mathbf{R}_{xx} - \sigma_n^2 \mathbf{1} \tag{6}$$

where in this equation σ_n^2 represents the estimation of the noise variance. It is straightforward to note that the matrix $\mathbf{L}_v$ is equal

$$\mathbf{L}_v = diag\{\lambda_1 - \sigma_n^2, ..., \lambda_m - \sigma_n^2\} \tag{7}$$

where λ_j (j=1, 2, ..., m) are the eigenvalues of the correlation matrix $\mathbf{R}_{xx}$ of measured signals **x**. In such case we can apply the standard eigenvalue decomposition to the matrix $\mathbf{R}_{vv}$,

$$\mathbf{R}_{vv} = \mathbf{V}_v \mathbf{L}_v \mathbf{V}_v^T \tag{8}$$

and define the whitening matrix **W** on the basis of this decomposition

$$\mathbf{W} = \mathbf{L}_v^{-1/2} \mathbf{V}_v^T \tag{9}$$

The key point in this approach is the accurate estimation of the noise variance. Generally this value is not known a’priori and should be estimated on the basis of measurement of the noisy signal. The most straightforward way to perform such estimation is to calculate the autocorrelation matrix $\mathbf{R}_{xx}$ of the observations and to analyze the distributions of the eigenvalues of this matrix. Irrespective of the noise level there is a visible knee point in this distribution. The eigenvalues corresponding to the signals are relatively high. The other small values (usually of negligible magnitudes) represent the noise. The variance of the noise may be estimated as the mean of all these insignificant eigenvalues []. If only K eigenvalues of the autocorrelation matrix are considered, the remaining last (m-K) eigenvalues represent the noise. The variance of the noise may be then estimated as

$$\sigma_n^2 = \frac{\sum_{j=K+1}^{m} \lambda_j}{m-K} \tag{10}$$

This estimated value is substituted in the equations (7).

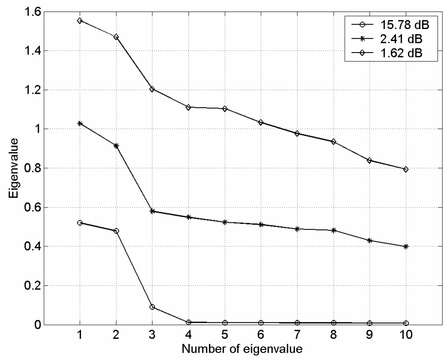

Fig. 2 The distribution of the eigenvalues of the correlation matrix at three different SNR values for the set of signals containing 3 noisy signals and 7 random noise components

The typical situation corresponding to the noisy case is presented in Fig. 2. It depicts the distribution of the eigenvalues of the system created by three signals corrupted by the noise and mixed with additional 7 white random noise signals of different variances (3-dimensional “signal space” and 7-dimensional ”noise space”). The knee point indicating the number of signals (3) is easily visible, irrespective of the noise level.

B. The regularization method

The main drawback of the previous method is its relative sensitivity to the estimation accuracy of the noise, which can be a source of significant error. There is another way to compensate for this poor performance of the main whitening algorithm at high noise. High noise, as was noted before, is reflected by small eigenvalues, resulting in magnifying the noise in the whitened signals due to its inverse relation to the eigenvalues of the considered system, $\mathbf{L}_x^{-0.5} = diag\left\{\frac{1}{\sqrt{\lambda_1 - \sigma_n^2}}, \frac{1}{\sqrt{\lambda_2 - \sigma_n^2}}, ..., \frac{1}{\sqrt{\lambda_m - \sigma_n^2}}\right\}$. In defining the matrix $\mathbf{L}_x$ we take into account the noise variance σ_n^2 and assume its modified form as given by (7). For the eigenvalues λ_i larger than the noise variance we can apply the following approximation

$\frac{1}{\sqrt{\lambda_i - \sigma_n^2}} \cong \frac{\lambda_i}{\sqrt{\lambda_i + \sigma_n^2}}$. As a result we get the whitening matrix **W** in the regularized form, described by

$$\mathbf{W} = diag\left\{\frac{\lambda_1}{\sqrt{\lambda_1 + \sigma_n^2}}, ..., \frac{\lambda_m}{\sqrt{\lambda_m + \sigma_n^2}}\right\}\mathbf{V}_x^T \qquad (11)$$

which is less sensitive to the noise variance.

C. Gram –Schmidt orthogonalization

The Gram-Schmidt orthogonalization performs the whitening of the signals sequentially, vector after vector using simple transformation scheme and as a result decomposes the given matrix **X** into product of the orthogonal matrix **Q** and upper diagonal matrix **U**

$$\mathbf{X} = \mathbf{QU} \qquad (12)$$

In Gram-Schmidt orthogonalization the transformation of **X** into **Q** and **U** is done sequentially step after step and mathematically it can be presented in the form

$$\mathbf{q}_1 = \mathbf{x}_1 \qquad (13)$$

$$u_{ik} = \frac{\mathbf{q}_i^T \mathbf{x}_k}{\mathbf{q}_i^T \mathbf{q}_i} \qquad (14)$$

$$\mathbf{q}_k = \mathbf{x}_k - \sum_{i=1}^{k-1} u_{ik}\mathbf{q}_i \qquad (15)$$

for k=2, 3, ..., K and i = 1, 2, ..., k-1. In these expressions $\mathbf{q}_i$ is the i^{th} column of matrix **Q** and u_{ik} is the ik^{th} element of matrix **U**.

Applying this algorithm to our case let us decompose the matrix $\mathbf{X}^T$ into Gram –Schmidt form as $\mathbf{X}^T$=**QU**. It is easy to prove that the whitening matrix **W** is equal

$$\mathbf{W} = \left[\mathbf{U}^T\right]^{-1} \qquad (16)$$

IV. THE MEASURE OF WHITENING QUALITY

The numerical experiments comparing different whitening algorithms have been performed on the signals generated in the typical structure presented in Fig. 3. The signals are mixed by using the mixing matrix **A**. The noise of the variance σ_n^2 is added and the measured signals **x** are then decorrelated by using the whitening matrix **W**.

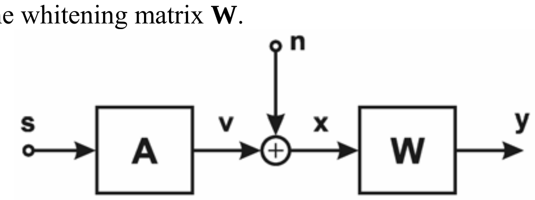

Fig. 3 The structure of signal processing used in numerical experiments of whitening

To compare different prewhitening algorithms we have to define the quality measure, determining the accuracy of prewhitening. The most important requirement is the orthogonality of the output signal, irrespective of the input noise. Let us assume that the source signals are statistically independent, that is the autocorrelation matrix $\mathbf{R}_{ss}$ is diagonal, $\mathbf{R}_{ss}$=**D**. Taking into account that the observed signals forming vector **x** are described by the matrix relation **x**=**As**+**n**, the correlation matrix $\mathbf{R}_{xx}$ fulfills the relation

$$\mathbf{R}_{xx} = \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^T + \mathbf{R}_{nn} \qquad (17)$$

The whitened signals forming vector **y**=**Wx** are characterized by the correlation matrix $\mathbf{R}_{yy}$ satisfying the following equation

$$\mathbf{R}_{yy} = \mathbf{W}\mathbf{R}_{xx}\mathbf{W}^T = \mathbf{W}\mathbf{A}\mathbf{R}_{ss}(\mathbf{WA})^T + \mathbf{W}\mathbf{R}_{nn}\mathbf{W}^T \qquad (18)$$

Taking into account that $\mathbf{R}_{ss}$=**D** the last relation can be presented in the form

$$\mathbf{R}_{yy} = \mathbf{WAD}^{0.5}(\mathbf{WAD}^{0.5})^T + \sigma_n \mathbf{WW}^T \qquad (19)$$

The whitening of the output signals corresponding to the source signals **s** means that the following condition must be satisfied

$$\mathbf{WAD}^{0.5} = \mathbf{1} \qquad (20)$$

On the basis of this we can define the error indicator of the whitening process as the Frobenius norm of the residue matrix

$$e_w = \left\|\mathbf{WAD}^{0.5} - \mathbf{1}\right\|_F \qquad (21)$$

The smaller the value of the error e_w, the better the quality of whitening process.

V. THE RESULTS OF NUMERICAL EXPERIMENTS

The numerical experiments have been performed for different arrangements of signals, corrupted by the white noise of uniform and normal distributions.

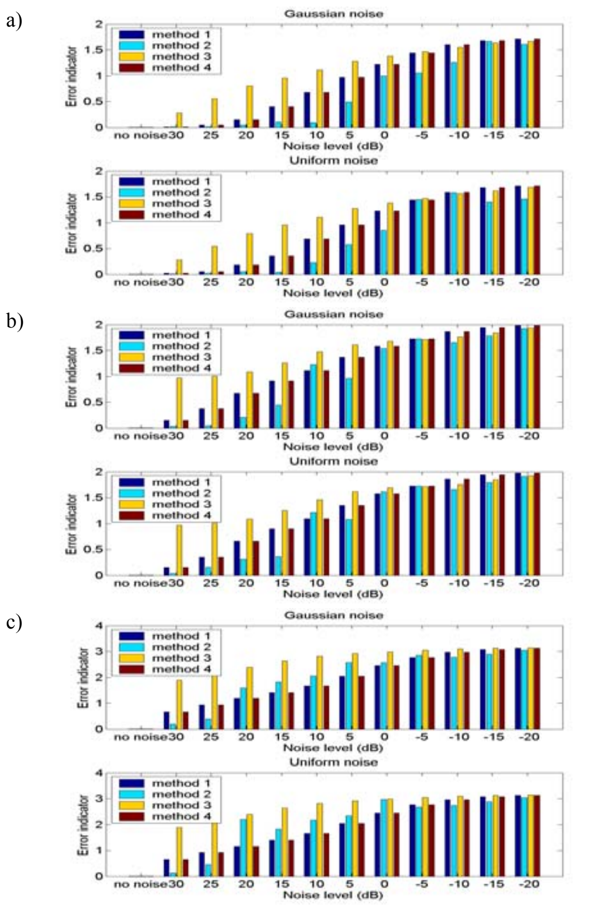

Fig. 4 The dependence of whitening error indicator on SNR for different noise distributions and different signals: a) 3 sinusoidal signals, b) 4 independent speech signals, c) 10 dependent speech signals all corrupted by noise (in all cases dimension of signal plus noise was equal 10)

All methods of whitening have been investigated at different signal-to-noise (SNR) ratios, changing from -20 to 30dB. The quality of whitening procedures has been measured by an error indicator (21). Different sets of signals and noise have been tried. The chosen results are presented in Fig. 4. In the first case (Fig. 4a) the "signal space" has been formed by 3 synthetic sinusoidal signals and the "noise space" by 7 noise components. In the next experiments we have changed the types of signals, as well as their number. We have applied: 4 independent speech signals combined with 6 noise components (Fig. 4b), and 10 dependent speech signals (the same sentence told by 10 different people) presented in Fig. 4c. As it is seen in almost all cases the best performance at different SNR values has been observed for bias removal method. In our opinion this method should be suggested for practical applications. However it is also seen that in the case of dependent signals (case c) the basic methods (the first and Gram-Schmidt) seem to be competitive.

The important observation drawn from these experiments is that the performance of the algorithm is practically independent on the type of distribution of the white noise. On the other side the dependency of the signals is an important factor. When assumption of independency is not valid the performance of all algorithms deteriorates.

a)

b)

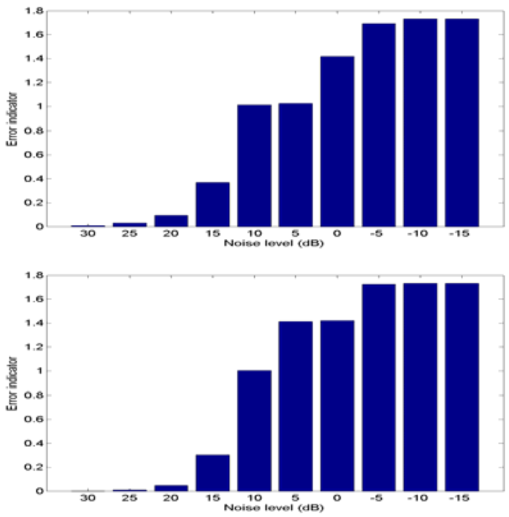

Fig. 5 The dependence of error of whitening on the SNR value for different signal compositions: a) 3 sinusoidal and 7 noise signals, b) 3 sinusoidal and 97 noise signals,

Our next experiments were aimed at the discovering how sensitive the whitening method is at different number of noise components **n** (the dimension of the noise space). Fig. 5 presents the results for the bias removal method applied at 2 different noise space dimensions at normal distribution of noise. Fig. 5a corresponds to 3-dimensional "signal space" and 7-dimensional "noise space", while in Fig. 5b the noise dimension has been changed to 97. In both cases the results are similar, which means that the method is rather insensitive to the number of noise components corrupting the measurements. The whitening quality depends only on the SNR ratio and its value did not exceed 1.8.

a)

b)

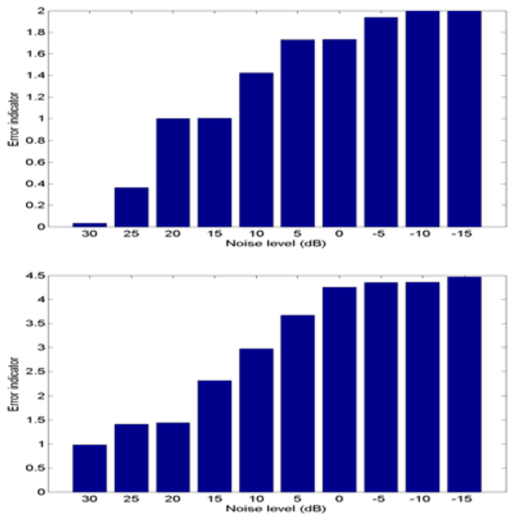

Fig. 6 The dependence of error of whitening on the SNR value for different type of signals: a) 4 independent speech and 96 noise signals, b) 20 dependent speech and 80 noise signals

The important variation is observed when we change the signals into dependent. Fig. 6a and b present the results for two kinds of signals. Fig. 6a corresponds to 4 independent and Fig. 6b – to 20 dependent speech signals, all corrupted with white noise. The dimension of signal and noise space was in both cases the same and equal 100. It is evident that lack of independence of signals significantly deteriorates the performance of the algorithm. The error has been doubled.

VI. Conclusions

The paper has presented and compared different methods of prewhitening of the measured signals. They have been based either on eigenvalue decomposition or on Gram-Schmidt orthogonalization procedure. The numerical experiments performed on different sets of either synthetic or real life signals have confirmed that the most robust is the method with bias removal. Its superiority is evident in almost whole region of applied noise variance. The important factor implicating the performance of the algorithm is the independency of the signals. Any dependency among signals deteriorates the results of whitening.

References

[1] A. Cichocki, S. I. Amari, Adaptive blind signal and image processing, Wiley, 2000

[2] S. Osowski, Sieci neuronowe w ujęciu algorytmicznym, WNT, Warszawa, 1996

[3] S. Haykin, Neural networks, a comprehensive foundation, Macmillan, 1994, N. Y.

[4] K. Diamantras, S. Y. Kung, Principal component neural networks, Wiley, 1996

[5] S. Chen, C.F. Cowan, P.M. Grant, Orthogonal least squares learning algorithm for RBF networks, IEEE Trans. Neural Networks, 1991, vol. 2, pp. 302-309

[6] T. Minka, Automatic choice of dimensionality for PCA, NIPS13, 2000, pp. 234-240