Convex Combination of Two Recursive Least Squares-type Algorithms for Adaptive Whitening

Shifeng Ou, Guoting Song, Ying Gao, Xiaohui Zhao

1 Institute of Opto-electronic Information, Yantai University, Yantai 264005, P. R. China, ousfeng@126.com
2 Institute of Opto-electronic Information, Yantai University, Yantai 264005, P. R. China, cailiaoytu@126.com
*3 Corresponding Author Institute of Opto-electronic Information, Yantai University, Yantai 264005, P. R. China, claragaoying@126.com
4 College of Communication Engineering, Jilin University, Changchun 130012, P. R. China, xhzhao@jlu.edu.cn

Abstract

The recursive least squares (RLS)-type whitening algorithm is shown to be a useful solution for improving the convergence rate of the least mean square (LMS)-type whitening algorithm, but the RLS-type algorithm with a constant forgetting factor requires a tradeoff between the convergence speed and steady-state misadjustment. By using a convex combination of two RLS algorithms with different forgetting factor, we present an effective method to solve the tradeoff problem of the RLS-type whitening algorithm. Firstly, two RLS-type whitening algorithms with different forgetting factor are convex combined in a reasonable manner to put together the best properties of them. Then, a mixing parameter is introduced to adjust the proportion of the two RLS algorithms, and the adaptive updating rule of the mixing parameter is obtained based on the stochastic gradient of the cost function. Finally, the weight transfer procedure and momentum term technique are developed to further improve the convergence performance of the convex combination algorithm. Since two RLS algorithms with different forgetting factor are applied and the best properties of them are acquired, the proposed combination algorithm work more efficiently than the existing adaptive whitening algorithms, as is verified by computer simulations in both stationary and non-stationary environments.

Keywords: Whitening; Recursive least squares; Convex combination; Blind source separation; Blind decorrelation

1. Introduction

Blind source separation (BSS) aims to recover the original statistically independent signals or otherwise unobserved source signals from their mixtures without knowing any prior information about the mixing channels. Since BSS is a fundamental problem encountered in various practical applications, it has drawn lots of attention in speech signal processing, wireless communication, biomedical and image signal processing fields [1-3]. On the other hand, Whitening or blind decorrelation is a very important tool in BSS problem, for some BSS algorithms such as nonlinear principle component analysis and fast independent component analysis, whitening is considered as a necessary condition for blind separation [4]. After whitening, the BSS tasks usually become somewhat easier and well-posed (less ill-conditioned), because the subsequent separating operation is described by an orthogonal matrix for source signals. Furthermore, time-delayed whitening technique can be used to identify the mixing matrix or directly perform the blind separation of colored and non-stationary source signals [5,6].

Over the past two decades, a lot of algorithms have been developed for whitening. Most of the existing whitening approaches can be categorized into either batch or adaptive techniques. Compared with the batch algorithms, the adaptive whitening algorithms have particularly practical advantages due to their computational simplicity and potentially improved performance in tracking non-stationary environment. However, most of these adaptive algorithms belong to the stochastic gradient or LMS-type algorithms, there exits a big problem that such adaptive algorithms often converges slowly. One solution to solve this problem is allowing the step size of the adaptive algorithm can be adjusted in a
proper way (e.g., variable step size algorithm) [7,8], but the step size adaptation technique often results in introducing some hyper parameters which are not easy to establish. The other effective solution to increase the convergence rate of LMS-type whitening algorithm is to adopt RLS formulation. It is well known that the RLS algorithm offers superior convergence rate and tracking capability as compared to the LMS algorithm. Based on this formulation, Zhu proposed a novel RLS algorithm for adaptive whitening [9], and comparison studies showed that this RLS algorithm is faster than other adaptive algorithms proposed for whitening. However, The performance of RLS-type algorithms in terms of convergence rate, tracking, stability and misadjustment depends on the forgetting factor. The RLS-type whitening algorithm using a constant forgetting factor requires a tradeoff between the previous performance criteria. When the forgetting factor is set very close to unity, the algorithm achieves low misadjustment and good stability, but its convergence rate and tracking capability are reduced. A smaller value of the forgetting factor improves the performance in tracking but increases the steady-state misadjustment, and could affect the stability performance of the algorithm.

Recently, a novel approach based on the convex combination of adaptive filters has been proposed [10-12]. The basic idea is that two or more adaptive filters with complementary properties adaptively combine their outputs by using a mixing parameter to obtain the improved performance. The merit of this combination is that it can obtain both fast convergence rate and small steady-state misadjustment without any estimate about the system noise power.

Motivated by these aspects, our proposal in this paper to improve the performance of RLS-type whitening algorithm is to convex combine two RLS algorithms with different forgetting factor adaptively, and the overall adaptive whitening system is carried out in an attempt to put together the best properties of the two RLS algorithms by means of a mixing parameter, which can control the combination proportion adaptively. Since the combination is adaptive itself, the overall system is equivalent to a variable forgetting factor procedure in which the forgetting factor is optimally selected at every time point. In addition, two modified schemes are also presented to eliminate the proposed algorithm’s insufficiency in convergence rate. The remainder of this paper is organized as follows. In section 2, we formulate the whitening problem and RLS-type whitening algorithm. In section 3, the adaptive combination algorithm using two RLS-type whitening systems and its improved version are derived. Simulation results in Section 4 describe the improved performance of the proposed method in both stationary and non-stationary conditions. Finally, in section 5 we give our conclusions.

2. Problem formulation and RLS-type whitening algorithm

In this paper, we consider the standard signal model for whitening, which is

\[ x(t) = As(t) + d(t) \]  \hspace{1cm} (1)

here, \( s(t) \in \mathbb{R}^n \) are \( n \) unknown, statistically independent source signals, with at most one Gaussian distributed at time \( t \), which are mixed by an unknown mixing matrix \( A \in \mathbb{R}^{m \times n} \), and \( x(t) \in \mathbb{R}^m \) are the \( m \) observed mixed signals. \( d(t) \in \mathbb{R}^m \) represents the additive noise which can be ignored. To simplify the problem, we assume that the number of sources matches the number of mixtures, i.e., \( m=n \). The main problem of whitening is to estimate whitening matrix \( W \) so that the output signal vector

\[ y(t) = Wx(t) \]  \hspace{1cm} (2)

are mutually uncorrelated and have the unit variance

\[ R_y = E\left[ y(t)y^T(t) \right] = WE\left[ x(t)x^T(t) \right]W^T = I \]  \hspace{1cm} (3)

where \( I \) is an identity matrix, and \( E(\cdot) \) is the operation of statistical expectation. Suppose \( D \) and \( H \) denote the eigenvectors matrix and eigenvalue matrix of the covariance matrix \( R_y = E[x(t)x^T(t)] \), which can be easily computed by using eigenvalue decomposition technique. Then a linear whitening matrix \( W \) is obtained [1]

\[ W = UH^{-1/2}D^T \]  \hspace{1cm} (4)

where \( U \) is an arbitrary orthogonal matrix. However, the result given in (4) is the batch-based algorithm, which is a very time consuming task and can not track time-varying environments effectively. The other useful tool for eliminating the disadvantage batch-based algorithm is to use an adaptive whitening algorithm, whose whitening matrix can be on-line updated in sympathy with the time-
varying dynamics. There exist many possible criteria to determine the whitening matrix in an adaptive manner. The focus of this paper is the classical adaptive algorithm proposed by Cardoso, which attempts to minimize the cost function $J_M$ as follows [13]

$$J_M = \text{Trace}\{W R W^\top\} - \log \det \{W R W^\top\} - n$$

(5)

where $\text{Trace}(\cdot)$ is the trace operator. From $J_M$, the online whitening procedure can be derived as follows [13]

$$W(t + 1) = W(t) + \mu[I - y(t) y^\top(t)]W(t)$$

(6)

where $\mu$ represents the step size parameter, and $y(t) = W(t)x(t)$ denotes the whitened output vector. Unfortunately, the adaptive whitening in (6) is typically stochastic gradient or LMS-type algorithm, which require careful choice of the step size parameters and often converges slowly. As we know that the RLS algorithm offers superior convergence rate and better tracking capability than the LMS algorithm. By incorporating the exponentially weighted average covariance matrix, Zhu presented a RLS-type algorithm for adaptive whitening, which replacing the cost function in (5) with the following [9]

$$J_L = \text{Trace}\{W \tilde{R} W^\top\} - \log \det \{W \tilde{R} W^\top\} - n$$

(7)

where $\tilde{R}$ is the exponentially weighted covariance matrix

$$\tilde{R} = (1-\lambda)\sum_{i=1}^{\infty} x(i)x^\top(i)$$

(8)

where $\lambda$ represents the forgetting factor which is kept in interval $(0, 1]$. The gradient of $J_L$ with respect to $W$ is

$$\nabla J_L = 2W \tilde{R} - 2\left[W R W^\top\right] W \tilde{R}$$

Postmultiplying both sides of (9) by $W^\top$, and letting $\nabla J_L = 0$, we can get the recursive formula, namely

$$W(t) = \left[(1-\lambda)\sum_{i=1}^{\infty} x(i)x^\top(i)W^\top(t-1)\right]^{-1}\left[\lambda W(t-1) + (1-\lambda)x(t)y^\top(t)\right]$$

(10)

here, the output whitened vector is $y(t) = W(t)x(t)$. By applying the well known matrix inversion lemma, it yields an adaptive RLS-type whitening algorithm as follows [9]

$$W(t) = \frac{1}{\lambda}\left[W(t-1) - \frac{y(t)y^\top(t)}{\lambda/(1-\lambda) + y^\top(t)y(t)}W(t-1)\right]$$

(11)

where the forgetting factor $\lambda$ is close to unity, it not only controls the convergence rate of the RLS algorithm but also determines the final misadjustment. A large value for this factor results in low misadjustment but slow convergence rate, while a small value will lead to fast convergence rate but increased misadjustment or error residual.

3. The proposed algorithm

To solve this tradeoff problem inherent in the RLS-type whitening algorithm presented in [9], the fundamental idea of our proposed algorithm is that two RLS algorithms with different forgetting factors are adopted separately, and the output vectors of each algorithm are convex combined in such a way that the advantages of both RLS algorithms are retained: the reduced misadjustment from the slow algorithm (large $\lambda$ ) and the rapid convergence from the fast algorithm (small $\lambda$ ).

The proposed scheme using such an convex combination of two RLS algorithms is shown in Figure 1. The output vector of the overall whitening system is

$$y(t) = W(t)x(t) = \zeta(t)y_1(t) + [1-\zeta(t)]y_2(t)$$

(12)

where $y_1(t)$ and $y_2(t)$ are the respective outputs of each RLS-type whitening algorithm, i.e., $y_1(t) = W(t-1)x(t)$, $l=1,2$. $\zeta(t)$ is the mixing parameter which is kept in interval $[0, 1]$, and $W(t)$ is the overall whitening matrix that can be computed as
Figure 1. Convex combination of two RLS-type whitening algorithms

\[ W(t) = \zeta(t)W_1(t-1) + \left[1 - \zeta(t)\right]W_2(t-1) \]  

(13)

The idea is that if \( \zeta(t) \) is assigned appropriate values at each iteration, then the above combination would extract the best properties from the individual matrix \( W_1(t-1) \) and \( W_2(t-1) \). Both whitening matrix operate completely decoupled from each other using their own RLS adaptation rule:

\[
W_1(t) = \frac{1}{\lambda_1} \left[ W_1(t-1) - \frac{y_1(t)y_1^T(t)}{\lambda_1/(1-\lambda_1) + y_1^T(t)y_1(t)} W_1(t-1) \right]
\]

(14)

and

\[
W_2(t) = \frac{1}{\lambda_2} \left[ W_2(t-1) - \frac{y_2(t)y_2^T(t)}{\lambda_2/(1-\lambda_2) + y_2^T(t)y_2(t)} W_2(t-1) \right]
\]

(15)

where \( \lambda_1 < \lambda_2 \), which means the first whitening system is faster but results in more misadjustment than the second system.

The key point of the convex combination algorithm is to control the mixing parameter according to the performance of the two components. For the adaptation of \( \zeta(t) \), we propose to use a gradient descent method to minimize the cost function in (5), and so the updating equation for \( \zeta(t) \) is given by

\[
\zeta(t+1) = \zeta(t) - \mu_{\zeta} \nabla_\zeta J_{\mu} \bigg|_{\zeta(t)}
\]

(16)

where \( \mu_{\zeta} \) is a very small constant, and \( J_{\mu} \) is the cost function from which the gradient based adaptive whitening algorithm is derived [13]. To proceed, we use an inner product of matrices defined as [14]

\[
\langle C, D \rangle = \text{trace}(C^T D)
\]

(17)

where \( \langle \rangle \) denotes the inner product, and two matrices \( C, D \in \mathbb{M}^{m\times n} \). Therefore, exploiting (14), the gradient term on the right hand side of (16) can be evaluated as

\[
\nabla_\zeta J_{\mu} \bigg|_{\zeta(t)} = \left[ \frac{\partial J_{\mu}}{\partial W(t)} \right]^T \frac{\partial W(t)}{\partial \zeta(t)}
\]

(18)

From [13], we can get

\[
\frac{\partial J_{\mu}}{\partial W(t)} = -[I - y(t)y^T(t)]W(t)
\]

(19)

according to (13), we have

\[
\frac{\partial W(t)}{\partial \zeta(t)} = W_1(t-1) - W_2(t-1)
\]

(20)

Denoting

\[
\Gamma(t) = [I - y(t)y^T(t)]W(t)
\]

(21)
and
\[
H(t) = W_1(t-1) - W_2(t-1)
\]  
(22)

Substituting (19), (20) into (18), finally, we can obtain
\[
\nabla_{\zeta} J_{sw}[\zeta(t)] = -\text{trace}\left[ \Gamma^T(t)H(t) \right]
\]  
(23)

and thus the adaptive update equation for the mixing parameter \( \zeta(t) \) with the form of (16) can be written as
\[
\zeta(t+1) = \zeta(t) + \mu \zeta \text{trace}\left[ \Gamma^T(t)H(t) \right]
\]  
(24)

Here, the mixing parameter \( \zeta(t) \) should be restricted to the interval \([0, 1]\).


<table>
<thead>
<tr>
<th>Table 1. Summary of the proposed convex combination algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialize: ( W_1(0), W_2(0), \lambda_1, \lambda_2, \mu_\zeta ) and ( T )</td>
</tr>
<tr>
<td>For ( t = 0, 1, \cdots, N ), Do</td>
</tr>
<tr>
<td>( y_i(t) = W_i(t)x(t-1), \quad i = 1, 2 )</td>
</tr>
<tr>
<td>( W(t) = \zeta(t)W_1(t-1) + [1 - \zeta(t)]W_2(t-1) )</td>
</tr>
<tr>
<td>( y(t) = W(t)x(t) )</td>
</tr>
<tr>
<td>( \zeta(t+1) = \zeta(t) + \mu_\zeta \text{trace}\left[ \Gamma^T(t)H(t) \right] + 0.3[\zeta(t) - \zeta(t-1)] )</td>
</tr>
<tr>
<td>( W_1(t) = \frac{1}{\lambda_1} \left[ W_1(t-1) - \frac{y_1(t)y_1^T(t)}{\lambda_1/(1-\lambda_1) + y_1^T(t)y_1(t)} W_1(t-1) \right] )</td>
</tr>
<tr>
<td>if ( \zeta(t) &gt; T )</td>
</tr>
<tr>
<td>( W_2(t) = 0.05 * W_1(t) + 0.95 * \frac{1}{\lambda_2} \left[ W_2(t-1) - \frac{y_2(t)y_2^T(t)}{\lambda_2/(1-\lambda_2) + y_2^T(t)y_2(t)} W_2(t-1) \right] )</td>
</tr>
<tr>
<td>else</td>
</tr>
<tr>
<td>( W_2(t) = \frac{1}{\lambda_2} \left[ W_2(t-1) - \frac{y_2(t)y_2^T(t)}{\lambda_2/(1-\lambda_2) + y_2^T(t)y_2(t)} W_2(t-1) \right] )</td>
</tr>
<tr>
<td>end if</td>
</tr>
<tr>
<td>end;</td>
</tr>
</tbody>
</table>

This proposed scheme has a very simple interpretation: in situations where a fast convergence speed would be desirable, the fast RLS-type whitening algorithm will outperform the slow one, making \( \zeta(t) \) approaches towards unity and \( W(t) \approx W_1(t-1) \). However, in stationary intervals, it is the slow algorithm which operates better, making \( \zeta(t) \) get close to zero and \( W(t) \approx W_2(t-1) \), while low misadjustment would be obtained.

It is also possible to further improve the performance of the convex combination algorithm by using the good convergence properties of the fast RLS algorithm to speed up the convergence rate of the slow one. We can do this by step-by-step transferring a part of weight matrix \( W_1(t) \) to \( W_2(t) \), and this transfer procedure is only applied if the fast algorithm is outperforming the slow one. So, the specific procedure is as follows:

\[
W_2(t) = \rho W_1(t) + (1-\rho) \frac{1}{\lambda_2} \left[ W_2(t-1) - \frac{y_2(t)y_2^T(t)}{\lambda_2/(1-\lambda_2) + y_2^T(t)y_2(t)} W_2(t-1) \right]
\]  
(25)

where \( \rho \) is the parameter to control the proportion in the transfer procedure, and a good choice for it is 0.05 in practice. It need to note that this transfer procedure is applied if the fast algorithm is significantly outperforming the slow one, so we only use this speeding up procedure provided that \( \zeta(t) \) close to unity.

Convex Combination of Two Recursive Least Squares-type Algorithms for Adaptive Whitening
Shifeng Ou, Guoting Song, Ying Gao, Xiaohui Zhao
Another modification of the algorithm is adding a momentum term in (24) for improving the convergence performance of the mixing parameter $\zeta(t)$

$$
\zeta(t+1) = \zeta(t) + \mu \zeta \text{ trace}[I^\top(t)H(t)] + a[\zeta(t) - \zeta(t-1)]
$$

(26)

where $a$ is momentum factor, and in general, 0.3 is a good choice for it. Compared with the basic convex combination algorithm, these two modifications will improve the convergence rate of the overall whitening system. The summary of the proposed algorithm are generalized in table 1.

4. Experiments and results

In this section, the performance of the proposed convex combination algorithm was compared to that of the classical adaptive algorithm and the RLS algorithm in [9] for whitening.

![Figure 2. EF achieved by four algorithms (stationary environment)](image)

![Figure 3. Evolution of $\zeta(t)$ in the proposed algorithm (stationary environment)](image)

Firstly, three zero mean sources $s(t) = [\sin(2\pi 200t), \text{sign}(\sin(2\pi 100t)), 2 \ast (\text{rand}(t) - 0.5)]^\top$ were mixed by a $3 \times 3$ mixing matrix $A$. The value of step size for the classical adaptive algorithm was initialized at 0.002, and the forgetting factor parameters of the proposed algorithm were set to $\zeta_1 = 0.98$, $\zeta_2 = 0.999$, the threshold parameter $T$ was 0.8, and the constant $\mu_c$ was set to 0.06. The value of forgetting factor for the RLS-type whitening algorithm in [9] were set at 0.98 and 0.999, respectively.
The error factor (EF) as a function of the system matrix $\mathbf{G} = \mathbf{WA}$, was used to evaluate the performance of these whitening algorithms [7]

$$\text{EF}(\mathbf{G}) = \|\mathbf{I} - \mathbf{G}\mathbf{G}^\top\|_F = \|\mathbf{I} - \mathbf{W}(t)\mathbf{A}\mathbf{A}^\top\mathbf{W}^\top(t)\|_F$$

(27)

where $g_{ij} = [\mathbf{G}]_{ij}$, which ideally attains its minimum value, zero, when whitening is achieved. Figure 2 plots the EF values obtained from the simulations of each whitening algorithm averaged over 100 Monte Carlo trials. From this figure, we can see that the classical adaptive algorithm has the worst whitening performance, and the RLS-type whitening algorithm with big forgetting factor is able to achieve reduced misalignment in steady-state at the cost of slower convergence, while the algorithm with small forgetting factor has very faster convergence rate but increased misadjustment. The proposed algorithm inherits the best properties of each RLS algorithms, presenting fast convergence together with the low residual error. Furthermore, it is important to remark that the weight transfer procedure and the momentum term technique incorporated in our algorithm allows the convex combined whitening algorithm to achieve the steady-state misalignment of the slow algorithm very soon, and so the more smoother property is obtained in the proposed algorithm. Figure 3 plots the evolution of the mixing parameter $\zeta(t)$ in the our algorithm, it can be seen from the figure that the estimated mixing parameter $\zeta(t)$ can adaptively adjust its value according to the evolution of algorithm’s performance: it can increases toward unity initially, to provide a fast convergence rate and good tracking performance, and then converges to a small value to obtain a small steady-state error from the second RLS component. The evolution of this parameter clearly matches the requirements of the convex combination.

![Figure 4](image_url)

**Figure 4.** EF achieved by four algorithms (non-stationary environment)

![Figure 5](image_url)

**Figure 5.** Evolution of $\zeta(t)$ in the proposed algorithm (non-stationary environment)
In the second experiment, the mixing matrix \( A \) was changed abruptly at sample number 5000, which means that the performance of these adaptive whitening algorithms was compared in a non-stationary environment. The parameter values of all the whitening algorithms were initialized the same as in the stationary environment. The convergence performance of the \( \zeta(t) \) averaged over 100 Monte Carlo trials were showed in Figure 4 and Figure 5, which again confirmed the advantage of the proposed algorithm while following the abrupt change in the mixing channel.

5. Conclusions

The RLS-type whitening algorithm can effectively improve the performance of the LMS-type whitening algorithm, but the RLS algorithm also requires a tradeoff between convergence rate and final misadjustment. To solve this inherent problem, this paper proposed to use an convex combination of two RLS-type algorithms with different forgetting factor. The mixing parameter was designed to adaptively control the proportion of this two RLS algorithms in the overall whitening system, and so the proposed algorithm can effectively extract the best properties of each RLS-type algorithm. In addition, some modified measure was also presented to improve the performance of the convex combination algorithm. Simulation results showed the good performance of the proposed approach.

6. Acknowledgement

The work presented in this paper is supported by the National Natural Science Foundation of China under Grants No. 61005021, and the Research Fund for the Doctoral Program of Higher Education of China under Grants No. 200801830037.

7. References


