RLS-Based Adaptive Algorithms for Generalized Eigen-Decomposition

Jian Yang, Hongsheng Xi, Feng Yang, and Yu Zhao

***Abstract*—The aim of this paper is to develop efficient online adaptive algorithms for the generalized eigen-decomposition problem which arises in a variety of modern signal processing applications. First, we reinterpret the generalized eigen-decomposition problem as an unconstrained minimization problem by constructing a novel cost function. Second, by applying projection approximation method and recursive least-square (RLS) technique to the cost function, a parallel adaptive algorithm for a basis for the r-dimensional $(r > 0)$ dominant generalized eigen-subspace and a sequential algorithm based on deflation technique for the first r-dominant generalized eigenvectors are derived. These algorithms can be viewed as counterparts of the extended projection approximation subspace tracking (PAST) and PASTd algorithms, respectively. Furthermore, we modify the parallel algorithm to explicitly estimate the first r-generalized eigenvectors in parallel, not the generalized eigen-subspace. More important, the modified parallel algorithm can be used to extract multiple generalized eigenvectors of two nonstationary sequences, while the proposed sequential algorithm lacks this ability because of slow convergence of minor generalized eigenvectors due to error propagation of the deflation technique. Third, following convergence analysis methods for PAST and PASTd, we prove the asymptotic convergence properties of the proposed algorithms. Finally, computer simulations are performed to investigate the accuracy and the speed advantages of the proposed algorithms.**

***Index Terms*—Adaptive algorithms, generalized eigen-decomposition, generalized eigenvector, matrix pencil, recursive least-square (RLS).**

I. Introduction

PRINCIPAL component analysis (PCA) and generalized eigen-decomposition (GED) play very important roles in various signal processing applications, e.g., in data compression, feature extraction, noise filtering, antenna array processing, classification [1]–[4]. PCA, which is the special case of GED problem, has been widely studied. Many efficient and powerful adaptive algorithms have been developed for PCA [5], [8], [9]. However, there are very few efficient online adaptive algorithms for the GED problem that can be applied in real-time applications. The purpose of this paper is to propose novel efficient online adaptive algorithms for the generalized eigen-decomposition problem.

Manuscript received September 21, 2004; revised April 20, 2005. This work was partially supported by the National "863" Project of China (Grant no. 2005AA103320) and the Innovation foundation (Projection no. KD2004046). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andrew C. Singer.

The authors are with the Laboratory of Network Communication System and Control (218), Department of Automation, University of Science and Technology of China, Hefei, Anhui, China 230027 (e-mail: jian_y@ustc.edu; xihs@ustc.edu.cn; youngf@ustc.edu; universe@mail.ustc.edu.cn).

Digital Object Identifier 10.1109/TSP.2005.863040

Some concepts used in the generalized eigen-decomposition problem are given as follows [10]:

$$\mathbf{R}_y\mathbf{w} = \lambda\mathbf{R}_x\mathbf{w} \tag{1}$$

where $\mathbf{R}_x$ and $\mathbf{R}_y$ are $N \times N$ symmetric definite matrices. The positive scalar λ and the corresponding vector $\mathbf{w}$ are called the generalized eigenvalue and eigenvector, respectively, of the matrix pencil $(\mathbf{R}_y, \mathbf{R}_x)$. According to the matrix theory, the matrix pencil has N positive generalized eigenvalues, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$, and corresponding $\mathbf{R}_x$-orthonormal generalized eigenvectors $\mathbf{v}_i$, $i = 1, \cdots, N$:

$$\mathbf{R}_y\mathbf{v}_i = \lambda_i\mathbf{R}_x\mathbf{v}_i \tag{2}$$

$$\mathbf{v}_i^H\mathbf{R}_x\mathbf{v}_j = \delta_{ij}, \quad i, j \in \{1, \cdots, N\} \tag{3}$$

where δ_{ij} is the Kronecker delta function.

The task of the generalized eigen-decomposition is to find the first r principal generalized eigenvectors of the pencil $(\mathbf{R}_y, \mathbf{R}_x)$. Many researchers have proposed many methods for solving it assuming given $\mathbf{R}_x$ and $\mathbf{R}_y$ [6], [7], [11]. But these analytical techniques are computationally intensive. In fact, in adaptive signal processing applications, the matrices $\mathbf{R}_x$ and $\mathbf{R}_y$ are time-variant and need to be estimated from the samples. Hence, these algorithms based on given $\mathbf{R}_x$ and $\mathbf{R}_y$ are not feasible for real time signal processing. In [10], Mathew *et al.* propose a quasi-Newton adaptive algorithm derived from a cost function based on the penalty function method. This algorithm sequentially estimates the generalized eigenvector corresponding to the minimum generalized eigenvalue up to the generalized eigenvector corresponding to the maximum generalized eigenvalue. However, in many applications, we are only interested in estimating the principal generalized eigenvectors. Moreover, as shown in [10], the choice of the penalty factor needs *a priori* information of the data covariance matrices. In [17], by assuming $\mathbf{R}_x$ given, adaptive online algorithms for the generalized eigen-decomposition problem are proposed. The applicability of this type of algorithms is limited in the applications where $\mathbf{R}_x$ is not directly available. In [12], Chaterjee *et al.* present new adaptive algorithms to extract the generalized eigenvectors from two sequences of random vectors or matrices. Most algorithms in literatures including [12] are gradient-based algorithms [4], [13], [15]. The main problem of this type of algorithms is slow convergence and difficulty of selecting an appropriate step size which is essential: too small a value will lead to slow convergence and too large a value will lead to overshooting and instability [16]. Rao *et al.* have developed a fast recursive least squares (RLS)-like, not true RLS, sequential algorithm for GED [14]. Whether its performance is better than

the RLS-type algorithm or not should be further investigated. Unfortunately, there are few true RLS-type algorithms.

In this paper, in order to derive efficient online adaptive algorithms, we formulate a novel unconstrained quartic cost function for the GED problem. By applying appropriate projection approximation [8], the cost function is modified to be fit for the RLS learning rule. We first derive a parallel iterative algorithm for estimating the basis for r-dimensional dominant generalized eigen-subspace. Then, starting from the parallel algorithm in one vector case ($r = 1$), we develop a sequential algorithm for explicitly estimating the first r dominant generalized eigenvectors by using a deflation method. Furthermore, we extend the algorithm for the generalized eigen-subspace to estimate the first r principal generalized eigenvectors in parallel. Following [21], we analyze the convergence properties of the proposed algorithms.

The rest of this paper is organized as follows. In Section II, a novel unconstrained criterion is presented. In Section III, adaptive algorithms are introduced to find the basis for dominant generalized eigen-subspace and the dominant generalized eigenvectors by minimizing the unconstrained criterion. In Section IV, the convergence of the proposed algorithms is discussed. Section V presents some simulation results to illustrate the applicability and the performance of these algorithms. Finally, the paper is concluded in Section VI.

II. Novel Unconstrained Criterion for GED

Let $\mathbf{x}$ and $\mathbf{y}$ be two complex valued random $N \times 1$ vector processes with the covariance matrices $\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^H]$ and $\mathbf{R}_y = E[\mathbf{y}\mathbf{y}^H]$. We assume that $\lambda_1 > \lambda_2 > \cdots > \lambda_r > \lambda_{r+1} \geq \cdots \geq \lambda_N$ are the generalized eigenvalues of $(\mathbf{R}_y, \mathbf{R}_x)$ and the corresponding $\mathbf{R}_x$-orthonormal generalized eigenvectors are $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_N$. In this paper, we focus on finding the first r principal generalized eigenvectors or the basis for r-dimensional dominant generalized eigen-subspace.

Consider the following scalar function:

$$\begin{aligned} J(\mathbf{W}) &= E\left\|\mathbf{R}_x^{-1}\mathbf{y} - \mathbf{W}\mathbf{W}^H\mathbf{y}\right\|_{\mathbf{R}_x}^2 \\ &= E\left[\mathbf{y}^H\mathbf{R}_x^{-1}\mathbf{y}\right] - 2E\left[\mathbf{y}^H\mathbf{W}\mathbf{W}^H\mathbf{y}\right] \\ &\quad + E\left[\mathbf{y}^H\mathbf{W}\mathbf{W}^H\mathbf{R}_x\mathbf{W}\mathbf{W}^H\mathbf{y}\right] \\ &= \mathrm{tr}(\mathbf{C}^H\mathbf{R}_y\mathbf{C}) - 2\mathrm{tr}(\mathbf{W}^H\mathbf{R}_y\mathbf{W}) \\ &\quad + \mathrm{tr}\left((\mathbf{W}^H\mathbf{R}_y\mathbf{W})(\mathbf{W}^H\mathbf{R}_x\mathbf{W})\right) \end{aligned} \tag{4}$$

where $\mathbf{R}_x = (\mathbf{C}^{-1})^H\mathbf{C}^{-1}$, $\mathbf{C}$ is an invertible $N \times N$ matrix and $\mathbf{W}$ is a $N \times r (r < N)$ matrix. In this paper, we assume $\mathbf{W}$ to have full rank r. In fact, if the rank of $\mathbf{W}$ is $\tilde{r} < r$, $\mathbf{W}$ in (4) can be replaced by a full rank $N \times \tilde{r}$ matrix $\tilde{\mathbf{W}}$ satisfying $\tilde{\mathbf{W}}\tilde{\mathbf{W}}^H = \mathbf{W}\mathbf{W}^H$.

As will be shown next, it is a novel criterion for GED problem. Note that when $\mathbf{R}_x = \mathbf{I}$, the cost function $J(\mathbf{W})$ is reduced to the criterion in [8] which is used for PCA. Therefore, the proposed cost function (4) is a universal criterion for PCA and GED problem.

Theorem 1: A $N \times r$ matrix $\mathbf{W}$ is a stationary point of $J(\mathbf{W})$ if and only if $\mathbf{W} = \mathbf{V}_r\mathbf{Q}$ where $\mathbf{V}_r$ contains any r distinct generalized eigenvectors of $(\mathbf{R}_y, \mathbf{R}_x)$ and Q is an arbitrary $r \times r$ unitary matrix. At each stationary point, $J(\mathbf{W})$ equals the sum of generalized eigenvalues whose generalized eigenvectors are not involved in $\mathbf{V}_r$.

Proof: See Appendix I. ∎

Theorem 2: All stationary points of $J(\mathbf{W}_{N\times r})$ are saddle points except when $\mathbf{V}_r$ contains the first r dominant generalized eigenvectors of $(\mathbf{R}_y, \mathbf{R}_x)$. In this case, $J(\mathbf{W})$ attains the global minimum.

Proof: See Appendix I. ∎

The above two theorems show that the proposed cost function $J(\mathbf{W})$ has a global minimum at which the columns of $\mathbf{W}$ span r-dimensional dominant generalized eigen-subspace of $(\mathbf{R}_y, \mathbf{R}_x)$ and no other local minima. This implies that one can search the global minimum point of $J(\mathbf{W})$ by iterative methods.

Therefore, the GED problem can be reformulated as the following unconstrained optimization problem:

$$\min\left\{J(\mathbf{W}) = E\left\|\mathbf{R}_x^{-1}\mathbf{y} - \mathbf{W}\mathbf{W}^H\mathbf{y}\right\|_{\mathbf{R}_x}^2\right\}. \tag{5}$$

According to the theorems, minimizing $J(\mathbf{W})$ will automatically result in a solution $\mathbf{W}$ with $\mathbf{R}_x$-orthonormal columns. This means that when we estimate the dominant generalized eigenvectors or the basis for eigen-subspace by minimizing $J(\mathbf{W})$ with an iterative algorithm, any $\mathbf{R}_x$-orthonormalization operation during the iteration is not required. Therefore, the proposed criterion for the generalized eigen-decomposition problem is different from those in the literature where the problem is often formulated as a constrained optimization problem [12].

III. Adaptive Algorithms for GED

Assume that at the instant n, the N-dimensional input vectors $\mathbf{x}[i]$ and $\mathbf{y}[i]$ are available and that $\mathbf{W}[i-1]$ are known, for $1 \leq i \leq n$. In practical applications, we are interested in updating $\mathbf{W}[n-1]$ to $\mathbf{W}[n]$ recursively. In this paper, we apply the RLS learning rule to estimate the generalized eigenvectors or the basis for eigen-subspace, since RLS-type algorithms offer fast convergence and tracking capability. We use similar steps in [8] to develop algorithms. As a result, two versions of algorithms can be derived: the parallel algorithm can be used to estimate the basis for dominant generalized eigen-subspace recursively, and the sequential algorithm can be applied to explicitly compute the dominant generalized eigenvectors. In addition, we modify the algorithm for the dominant generalized eigen-subspace to extract the first r principal generalized eigenvectors.

A. Parallel RLS-Based Adaptive Algorithm for Principal Generalized Eigen-Subspace

Consider the following exponentially weighted sum instead of the expectation (4):

$$J(\mathbf{W}[n]) = \sum_{j=1}^{n} \beta^{n-j}\left\|\mathbf{R}_x^{-1}\mathbf{y}[j] - \mathbf{W}[n]\mathbf{W}^H[n]\mathbf{y}[j]\right\|_{\mathbf{R}_x}^2 \tag{6}$$

where the forgetting factor β is between 0 and 1. If $\beta = 1$, all the samples are given the same weight, and no forgetting of old

data takes place. Choosing $\beta < 1$ is especially useful in tracking nonstationary changes.

Note that the cost function (6) is fourth order in the elements of $\mathbf{W}[n]$. Considering the projection approximation technique [8], it can be simplified by approximating the vector $\mathbf{W}^H[n]\mathbf{y}[j]$ in the sum (6) by the vector $\mathbf{z}[j] = \mathbf{W}^H[j-1]\mathbf{y}[j]$. These vectors can be easily computed because the estimated weight matrices $\mathbf{W}[j-1]$ for the previous iteration steps $j = 1, \cdots, n$ are already known at step n. This approximation yields the modified least-square-type criterion

$$J'(\mathbf{W}[n]) = \sum_{j=1}^{n} \beta^{n-j} \left\| \mathbf{R}_x^{-1}\mathbf{y}[j] - \mathbf{W}[n]\mathbf{z}[j] \right\|_{\mathbf{R}_x}^2. \quad (7)$$

The rationality of using the projection approximation in the above derivation is same as presented in [8]. The advantage of the modified cost function $J'(\mathbf{W}[n])$ is that it is a quadratic function in the elements of $\mathbf{W}[n]$ and the standard formula for the RLS algorithm.

Applying the RLS technique to minimize the modified criterion $J'(\mathbf{W}[n])$, we can derive the following recursive algorithm to solve the basis for r-dimensional dominant generalized eigen-subspace:

$$\mathbf{z}[n] = \mathbf{W}^H[n-1]\mathbf{y}[n] \quad (8)$$

$$\mathbf{h}[n] = \mathbf{P}[n-1]\mathbf{z}[n] \quad (9)$$

$$\mathbf{g}[n] = \frac{\mathbf{h}[n]}{(\beta + \mathbf{z}^H[n]\mathbf{h}[n])} \quad (10)$$

$$\mathbf{P}[n] = \mathrm{Tri}\left(\frac{1}{\beta}\left(\mathbf{P}[n-1] - \mathbf{g}[n]\mathbf{h}^H[n]\right)\right) \quad (11)$$

$$\mathbf{e}[n] = \mathbf{R}_x^{-1}\mathbf{y}[n] - \mathbf{W}[n-1]\mathbf{z}[n] \quad (12)$$

$$\mathbf{W}[n] = \mathbf{W}[n-1] + \mathbf{e}[n]\mathbf{g}^H[n] \quad (13)$$

where the notation Tri means that only the upper triangular part of the argument is computed and its transpose is copied to the lower triangular part, making thus the matrix $\mathbf{P}[n]$ symmetric.

Note that the step (12) contains the determined matrix $\mathbf{R}_x^{-1}$. Therefore, the algorithm (8)–(13) is fit for the signal processing applications with a known matrix $\mathbf{R}_x$. However, in many applications it is unknown, and should be estimated from sequence vectors $\mathbf{x}[n]$. We use the exponentially weighted sample correlation matrix $\mathbf{R}_x[n]$ instead of $\mathbf{R}_x$. The recursive update equation can be written as

$$\mathbf{R}_x[n] = \mu\mathbf{R}_x[n-1] + \mathbf{x}[n]\mathbf{x}^H[n] \quad (14)$$

where μ is the forgetting factor between 0 and 1. Using the matrix inversion lemma [16], we can write the time update for the inverse matrix $\mathbf{Q}[n] = \mathbf{R}_x^{-1}[n]$ as

$$\mathbf{Q}[n] = \frac{1}{\mu}\left(\mathbf{I} - \frac{\mathbf{Q}[n-1]\mathbf{x}[n]\mathbf{x}^H[n]}{\mu + \mathbf{x}^H[n]\mathbf{Q}[n-1]\mathbf{x}[n]}\right)\mathbf{Q}[n-1]. \quad (15)$$

By replacing $\mathbf{R}_x^{-1}$ in step (12) with $\mathbf{Q}[n]$, a totally online adaptive parallel algorithm is derived. The simplest way to choose the initial values is to set $\mathbf{P}[0] = \alpha_1\mathbf{I}$, $\mathbf{Q}[0] = \alpha_2\mathbf{I}$, where α_1 and α_2 are appropriate positive scalar. Let $\mathbf{W}[0] = \alpha_3[\mathbf{m}_1, \cdots, \mathbf{m}_r]$, where $\mathbf{m}_i$ is the ith column of the $N \times N$ identity matrix and $\alpha_3 > 0$.

In the above algorithm, a recursive computation of the $r \times r$ matrix $\mathbf{P}[n]$, the $N \times r$ matrix $\mathbf{W}[n]$ and the $N \times N$ matrix $\mathbf{Q}[n]$ requires $O(r^2)$, $O(Nr)$ and $O(N^2)$. The total computation complexity of the algorithm is $O(N^2) + 3Nr + O(r^2)$ per update. Note that the update steps for $\mathbf{P}[n]$ and $\mathbf{Q}[n]$ involve subtraction. Therefore, accumulation of numerical and quantization errors with time update may lead to the loss of Hermitian positive definiteness, while $\mathbf{P}[n]$ and $\mathbf{Q}[n]$ are theoretically Hermitian positive definite (being the inverse of a Hermitian positive definite matrix). An efficient and numerical robust way is to apply the QR-update technique to update the square root matrices $\mathbf{P}^{1/2}[n]$ and $\mathbf{Q}^{1/2}[n]$ [16]. Since $\mathbf{P}[n] = \mathbf{P}^{1/2}[n]\mathbf{P}^{H/2}[n]$ and $\mathbf{Q}[n] = \mathbf{Q}^{1/2}[n]\mathbf{Q}^{H/2}[n]$, they remain Hermitian positive definite regardless of any numerical or quantization errors in the square roots.

B. Sequential RLS-Based Adaptive Algorithm for r Dominant Generalized Eigenvectors

For the case $r = 1$, $\mathbf{w} = \pm\mathbf{v}_1$ is the global minimum of $J(\mathbf{w})$ according to Theorem 2. Based on this fact, we present the following theorem to develop the sequential algorithm.

Theorem 3: Let $\lambda_1 > \lambda_2 > \cdots \lambda_r > \lambda_{r+1} \geq \cdots \geq \lambda_N > 0$. Suppose that $\mathbf{w}_i = \pm\mathbf{v}_i$ $(i = 1, \cdots, j-1)$ are the first $j-1$ dominant generalized eigenvectors of $(\mathbf{R}_y, \mathbf{R}_x)$. Define the following deflation:

$$\mathbf{y}_j = \mathbf{y}_{j-1} - \mathbf{R}_x\mathbf{w}_{j-1}\mathbf{w}_{j-1}^H\mathbf{y}_{j-1} \quad (j = 2, \cdots, r) \quad (16)$$

where $\mathbf{y}_1 = \mathbf{y}$. Then $\mathbf{w}_j = \pm\mathbf{v}_j$ is the unique global minimal point of $J_j(\mathbf{w}_j)$ and the others are saddle points of $J_j(\mathbf{w}_j)$, where $J_j(\mathbf{w}_j)$ is defined as

$$J_j(\mathbf{w}_j) = E\left\| \mathbf{R}_x^{-1}\mathbf{y}_j - \mathbf{w}_j\mathbf{w}_j^H\mathbf{y}_j \right\|_{\mathbf{R}_x}^2 \quad (17)$$

Proof: See Appendix II. ∎

Thus, we apply the parallel algorithm with $r = 1$ to minimize $J_j(\mathbf{w}_j)$ for estimating the jth dominant generalized eigenvector, i.e., $\mathbf{v}_j$, of $(\mathbf{R}_y, \mathbf{R}_x)$. Furthermore, the sequential adaptive algorithm for the first r principal generalized eigenvectors can be summarized as

$$\mathbf{y}_1[n] = \mathbf{y}[n] \quad (18)$$

$$\textbf{for } i = 1, 2, \cdots, r \textbf{ do} \quad (19)$$

$$z_i[n] = \mathbf{w}_i^H[n-1]\mathbf{y}_i[n] \quad (20)$$

$$d_i[n] = \beta d_i[n-1] + |z_i[n]|^2 \quad (21)$$

$$\mathbf{e}_i[n] = \mathbf{R}_x^{-1}\mathbf{y}_i[n] - \mathbf{w}_i[n-1]z_i[n] \quad (22)$$

$$\mathbf{w}_i[n] = \mathbf{w}_i[n-1] + \mathbf{e}[n]\left(\frac{z_i^H[n]}{d_i[n]}\right) \quad (23)$$

$$\mathbf{y}_{i+1}[n] = \mathbf{y}_i[n] - \mathbf{R}_x\mathbf{w}_i[n]z_i[n]. \quad (24)$$

In the above algorithm, $d_i[n]$ is equal to $1/\mathbf{P}_i[n]$ $(r = 1)$ in the algorithm (8)–(13). Then $g_i[n]$ can be expressed as $z_i[n]/d_i[n]$. Note that step (24) contains $\mathbf{R}_x$ which remains to be estimated. Actually, the update rule (15) has been used to estimate the matrix $\mathbf{R}_x^{-1}$. It is not appropriate to use another iterative equation to estimate $\mathbf{R}_x$. Because using two independent iterative equations for $\mathbf{Q}[n]$ and $\mathbf{R}_x[n]$, the numerical errors will be produced, and

it cannot be guaranteed that $\mathbf{R}_x[n]\mathbf{Q}[n] = \mathbf{I}$. Hence, we will reformulate the algorithm to avoid estimating $\mathbf{R}_x$.

Let $\mathbf{c}_i[n] = \mathbf{R}_x[n]\mathbf{w}_i[n]$ and $\mathbf{s}_i[n] = \mathbf{R}_x[n]\mathbf{e}_i[n]$. We first estimate $\mathbf{c}_i[n]$ recursively instead of directly estimating $\mathbf{w}_i[n]$. Then, use the equation $\mathbf{w}_i[n] = \mathbf{Q}[n]\mathbf{c}_i[n]$ to calculate $\mathbf{w}_i[n]$. The modified algorithm can be given as

$$\mathbf{y}_1[n] = \mathbf{y}[n] \tag{25}$$

$$\mathbf{Q}[n] = \mathrm{Tri}\left(\frac{1}{\mu}\left(\mathbf{I} - \frac{\mathbf{Q}[n-1]\mathbf{x}[n]\mathbf{x}^H(n)}{\mu + \mathbf{x}^H[n]\mathbf{Q}[n-1]\mathbf{x}[n]}\right) \times \mathbf{Q}[n-1]\right) \tag{26}$$

$$\textbf{for } i = 1, 2, \cdots, r \textbf{ do} \tag{27}$$

$$z_i[n] = \mathbf{c}_i^H[n-1]\mathbf{Q}[n]\mathbf{y}_i[n] \tag{28}$$

$$d_i[n] = \beta d_i[n-1] + |z_i[n]|^2 \tag{29}$$

$$\mathbf{s}_i[n] = \mathbf{y}_i[n] - \mathbf{c}_i[n-1]z_i[n] \tag{30}$$

$$\mathbf{c}_i[n] = \mathbf{c}_i[n-1] + \mathbf{s}_i[n]\left(\frac{z_i^H[n]}{d_i[n]}\right) \tag{31}$$

$$\mathbf{w}_i[n] = \mathbf{Q}[n]\mathbf{c}_i[n] \tag{32}$$

$$\mathbf{y}_{i+1}[n] = \mathbf{y}_i[n] - \mathbf{c}_i[n]z_i[n]. \tag{33}$$

The initial values of the above algorithm can be set as $\mathbf{Q}[0] = \alpha_1\mathbf{I}$ ($\alpha_1 > 0$), $d_i[0] = \alpha_2$ ($\alpha_2 > 0$), $\mathbf{c}_i[0] = \alpha_3\mathbf{m}_i$, and $\mathbf{w}_i[0] = \mathbf{Q}[0]\mathbf{c}_i[0]$.

The above algorithm requires $4Nr + N^2r + O(N^2) + O(r)$ operations per update. In contrast to the parallel algorithm, this method enables the explicit computation of the generalized eigenvectors. On the other hand, the deflation technique causes a slightly increased computational complexity.

C. Modified Parallel Algorithm for r-Dominant Generalized Eigenvectors

It is important to note that the columns of $\mathbf{W}[n]$ estimated according to the proposed parallel algorithm is the basis spanning the principal generalized eigen-subspace, not the r principal generalized eigenvectors. Although the sequential version of the algorithm can result r principal generalized eigenvectors, the minor generalized eigenvectors may converge slowly because of the estimation error propagation (see simulation results in Section V). In order to overcome these difficulties, we will extend the parallel algorithm to find the r principal generalized eigenvectors in parallel.

As Theorem 1 and 2 show, $\mathbf{W} = \mathbf{V}_r\mathbf{Q}$ is the global minimum where the columns of $\mathbf{V}_r$ are r principal $\mathbf{R}_x$-orthonormal generalized eigenvectors and $\mathbf{Q}$ is any $r \times r$ unitary matrix. Let us consider the following constraint:

$$\mathbf{W}^H\mathbf{R}_y\mathbf{W} = \mathbf{A} \tag{34}$$

where $\mathbf{A} = \mathrm{diag}(a_1, a_2, \cdots, a_r)$ and $a_1, a_2, \cdots, a_r > 0$. Then, we can get that

$$\mathbf{Q}^H\Sigma_r = \mathbf{A}\mathbf{Q}^H \tag{35}$$

where the diagonal matrix Σ_r is a generalized eigenvalue matrix. Since the r principal generalized eigenvalues are different, it is easy to show that

$$\begin{aligned}\mathbf{A} &= \Sigma_r \\ \mathbf{Q} &= \mathbf{I} \\ \mathbf{W} &= \mathbf{V}_r.\end{aligned} \tag{36}$$

This means that if we combine the cost function (4) with the constraint (34), we can find the r principal generalized eigenvectors in parallel. It is a very important propriety for development of the modified algorithm. In fact, the constraint (34) implies that $\mathbf{W}$ should be $\mathbf{R}_y$-orthogonal. This can be achieved by $\mathbf{R}_y$-orthogonalizing the matrix $\mathbf{W}$ during every update of the parallel algorithm for dominant eigen-subspace. The steps of $\mathbf{R}_y$-orthogonalization are given as

$$\mathbf{a}_1 = \mathbf{w}_1 \tag{37}$$

$$\mathbf{a}_j = \mathbf{w}_j - \sum_{i=1}^{j-1}\frac{\mathbf{a}_i^H\mathbf{R}_y\mathbf{w}_j}{\mathbf{a}_i^H\mathbf{R}_y\mathbf{a}_i}\mathbf{a}_i \tag{38}$$

where $\mathbf{w}_i$ is the ith column of $\mathbf{W}$. When $\mathbf{R}_y = \mathbf{I}$, the $\mathbf{R}_y$-orthogonalization is same as the Gram–Schmidt orthogonalization (GSO) method. Therefore, it is a generalized form of the GSO method. If in the $\mathbf{R}_y$-orthogonalization method each $\mathbf{a}_j$ is further divided by its $\mathbf{R}_y$-norm, the vectors $\mathbf{a}_j$ will be $\mathbf{R}_y$-orthonormal. Note that the matrix $\mathbf{R}_y$ is unknown, it should be estimated from the input vector sequence. We use exponential average of the correlation matrix to approximate $\mathbf{R}_y$

$$\mathbf{R}_y[n] = \eta\mathbf{R}_y[n] + \mathbf{y}[n]\mathbf{y}^H[n] \tag{39}$$

where $\eta < 1$ is the forgetting factor.

Therefore, the above idea results the following modified parallel algorithm for r principal generalized eigenvectors.

$$\mathbf{z}[n] = \mathbf{W}^H[n-1]\mathbf{y}[n] \tag{40}$$

$$\mathbf{h}[n] = \mathbf{P}[n-1]\mathbf{z}[n] \tag{41}$$

$$\mathbf{g}[n] = \frac{\mathbf{h}[n]}{(\beta + \mathbf{z}^H[n]\mathbf{h}[n])} \tag{42}$$

$$\mathbf{P}[n] = \mathrm{Tri}\left(\frac{1}{\beta}\left(\mathbf{P}[n-1] - \mathbf{g}[n]\mathbf{h}^H[n]\right)\right) \tag{43}$$

$$\mathbf{Q}[n] = \mathrm{Tri}\left(\frac{1}{\mu}\left(\mathbf{I} - \frac{\mathbf{Q}[n-1]\mathbf{x}[n]\mathbf{x}^H[n]}{\mu + \mathbf{x}^H[n]\mathbf{Q}[n-1]\mathbf{x}[n]}\right) \times \mathbf{Q}[n-1]\right) \tag{44}$$

$$\mathbf{e}[n] = \mathbf{Q}[n]\mathbf{y}[n] - \mathbf{W}[n-1]\mathbf{z}[n] \tag{45}$$

$$\mathbf{W}[n] = \mathbf{W}[n-1] + \mathbf{e}[n]\mathbf{g}^H[n] \tag{46}$$

$$\mathbf{R}_y[n] = \eta\mathbf{R}_y[n] + \mathbf{y}[n]\mathbf{y}^H[n] \tag{47}$$

$$\mathbf{W}[n] = \mathbf{R}_y - \textbf{orthogonalize}\left(\mathbf{W}[n]\right). \tag{48}$$

Compared with the parallel algorithm for principal generalized eigen-subspace, it can explicitly track first r principal generalized eigenvectors, although the computation complexity is increased about $O(rN^2)$ during every update. In addition, we can determine the generalized eigenvectors' order according to diagonal elements of $\mathbf{W}^H[n]\mathbf{R}_y[n]\mathbf{W}[n]$.

D. Discussion

Since we use the modified criterion (7) instead of (6) to derive three proposed algorithms, the columns of $\mathbf{W}[n]$ may not be exactly $\mathbf{R}_x$-orthonormalized. This is not to say that every update of these algorithms needs $\mathbf{R}_x$-orthonormalization. Via computer simulations (see Section V), we have been able to confirm that $\mathbf{R}_x$-orthonormalization does not imply performance improvement of these algorithms. If some applications require explicit $\mathbf{R}_x$-orthonormalization, $\mathbf{R}_x$-orthonormalization steps described in Section III-C can be incorporated into the corresponding algorithms.

It should be also noted that the proposed algorithms have different properties. Although complexity of the parallel algorithm (8)–(13) is lowest, it converges to the principal generalized eigen-subspace. The sequential algorithm can be used to explicitly compute the first r principal generalized eigenvectors, but the convergence of the minors may be slow because of error propagation. The modified parallel algorithm makes one seek the r principal generalized eigenvectors fast; however it needs $\mathbf{R}_y$-orthogonalization during every update.

IV. Convergence Analysis

Since the proposed parallel algorithm for principal generalized eigen-subspace and the proposed sequential algorithm for dominant generalized eigenvectors can be viewed as the generalized forms of PAST and PASTd algorithms, following [21], which applies Ljung's stochastic approximation theory [20] to prove the convergence of PAST and PASTd, the convergence properties of the proposed algorithms can be proved similarly. It is unnecessary to analyze concretely as [21], since the analysis methods are almost same and some results can be directly applied. In this section, we only list the theorems for the proposed algorithms and point out the differences when proving these theorems.

In order to apply Ljung's theorem, we assume the stochastic processes $\mathbf{x}$ and $\mathbf{y}$ are stationary and choose the forgetting factor $\beta = 1$. Using similar derivation of the ordinary differential equation (ODE) for PAST [21], we can derive the ODE for the proposed parallel algorithm for principal generalized eigen-subspace

$$\frac{d\mathbf{R}(t)}{dt} = \mathbf{W}^H(t)\mathbf{R}_y\mathbf{W}(t) - \mathbf{R}(t) \tag{49}$$

$$\frac{d\mathbf{W}(t)}{dt} = \left(\mathbf{R}_x^{-1} - \mathbf{W}(t)\mathbf{W}^H(t)\right)\mathbf{R}_y\mathbf{W}(t)\mathbf{R}^{-1}(t) \tag{50}$$

where $\mathbf{W}(t)$ is a $N \times r$ matrix. Note that when $\mathbf{R}_x = \mathbf{I}$, the above ODE is same as ODE ((18a) and (18b)) for PAST in [21]. In the following, we will start from the ODE (49) and (50) to show the convergence of the proposed algorithm for principal generalized eigen-subspace and the sequential algorithm for dominant generalized eigenvectors. For notational convenience, we use $\mathbf{W}$ and $\mathbf{R}$ to denote $\mathbf{W}(t)$ and $\mathbf{R}(t)$, respectively.

Theorem 4: Given a symmetric and positive definite matrix pencil $(\mathbf{R}_y, \mathbf{R}_x)$ with generalized eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_r > \lambda_{r+1} \geq \cdots \geq \lambda_N$. If $\mathbf{R}(0)$ is symmetric and positive definite, the asymptotically stable equilibrium state of autonomous system (49), (50) is $\tilde{\mathbf{W}} = \mathbf{V}_r\mathbf{Q}$, where $\mathbf{Q}$ is an arbitrary $r \times r$ unitary matrix.

Proof: See Appendix III. ∎

According to Ljung's theorem, Theorem 4 shows that the proposed algorithm for the principal generalized eigen-subspace converges to $\mathbf{V}_r\mathbf{Q}$.

Note that if $r = 1$, then the $\mathbf{w}(t)$ will converge to $-\mathbf{v}_1$ or $\mathbf{v}_1$ according to Theorem 4. By using this fact, we now prove the convergence of the sequential algorithm for dominant generalized eigenvectors.

Theorem 5: Assume $\lambda_1 > \cdots > \lambda_r > \lambda_{r+1} \geq \cdots \geq \lambda_N$. Then for $d_i(0) > 0$, $\mathbf{w}_i(k)$ tend to $\pm\mathbf{v}_i (i = 1, \ldots, r)$.

Proof: See Appendix III. ∎

While rigorous analysis of the modified parallel algorithm for principal generalized eigenvectors may be greatly difficult, we expect it to retain numerical stability. In fact, every update of the algorithm involves two steps. The first step is to find the principal generalized eigen-subspace. Therefore, the convergence property should be same as the parallel algorithm for dominant generalized eigen-subspace. The second step is only to make the columns of $\mathbf{W}[n]$ computed according to the first step $\mathbf{R}_y$-orthogonal. Therefore, by applying these two steps alternately, after enough iterations, $\mathbf{W}[n]$ should satisfy that the columns of $\mathbf{W}[n]$ span the first r principal generalized eigen-subspace and are $\mathbf{R}_y$-orthogonal, simultaneously. This means that the columns of $\mathbf{W}[n]$ converge to the first r dominant generalized eigenvectors. In addition, simulations in Section V are performed to demonstrate the behavior of the modified algorithm for principal generalized eigenvectors.

V. Simulations

In this section, we present some simulation results to show the applicability and the performance of the proposed RLS-based adaptive GED algorithms. Two experiments are performed: one is to extract several dominant generalized eigenvectors from two random vector processes and investigate the effect of the projection approximation on the proposed algorithms; the other one is to apply the discussed algorithm in the one-vector case $r = 1$ to cancel interference in code-division multiple-access (CDMA) adaptive antenna arrays systems. The latter example demonstrates the proposed algorithm has fast convergence and excellent tracking capability. By adopting the proposed algorithm for the beamforming problem, the interference and noise to the desired user can be suppressed largely and the capacity of the system can be improved.

In the simulations, we compare our proposed algorithms with the following:

- FRLSL: fast RLS-like algorithm proposed by Rao [14];
- RDEVD: direct EVD method with exponentially weighted sample correlation matrices $\mathbf{R}_x[n]$ and $\mathbf{R}_y[n]$.

For notational convenience, we call the three proposed algorithms as follows:

- Algorithm 1: parallel RLS algorithm for principal generalized eigen-subspace;
- Algorithm 2: sequential RLS algorithm for the first r principal generalized eigenvectors;

- Algorithm 3: parallel RLS algorithm for the first r principal generalized eigenvectors.

A. Experiment 1

In order to evaluate the convergence speed and the estimated accuracy of the proposed algorithms, the direction cosine and subspace distance are defined, respectively, as

$$\text{direction cosine}[k] = \frac{\left|\mathbf{w}_i^H[k]\mathbf{v}_i\right|}{||\mathbf{w}_i[k]||\,||\mathbf{v}_i||} \tag{51}$$

and

$$\text{dist}[k] = \left\| \frac{\mathbf{W}[k]\mathbf{W}^H[k]}{\text{trace}\,(\mathbf{W}^H[k]\mathbf{W}[k])} - \frac{\mathbf{V}_r\mathbf{V}_r^H}{\text{trace}\,(\mathbf{V}_r^H\mathbf{V}_r)} \right\|_F \tag{52}$$

where $\mathbf{W}[k] = [\mathbf{w}_1[k], \cdots, \mathbf{w}_r[k]]$, $\mathbf{V}_r = [\mathbf{v}_1, \cdots, \mathbf{v}_r]$ and the subscript F denotes the Frobenius norm of a matrix. Note that $\mathbf{v}_i$ is the ith actual dominant generalized eigenvector of $(\mathbf{R}_y, \mathbf{R}_x)$ and can be computed via the direct EVD.

The input samples are generated by

$$\begin{aligned} y[k] &= 10\sqrt{2}\sin(0.46\pi k + \theta_1) \\ &\quad + 10\sqrt{2}\sin(0.74\pi k + \theta_2) + n_1[k] \end{aligned} \tag{53}$$

$$x[k] = \sqrt{2}\sin(0.62\pi k + \theta_3) + n_2[k] \tag{54}$$

where $\theta_i (i = 1, \cdots, 3)$ is the initial phase which has a uniform distribution in $[0, 2\pi]$, $n_1[k]$ and $n_2[k]$ are zero-mean white noises with variances $\sigma_1^2 = 1$ and $\sigma_2^2 = 0.01$. The input vectors $\mathbf{y}[k]$ and $\mathbf{x}[k]$ are arranged in blocks of size eight $(N = 8)$. The first four $(r = 4)$ dominant generalized eigenvectors of $(\mathbf{R}_y, \mathbf{R}_x)$ are considered to be extracted via FRLSL, RDEVD, Algorithm 2, and Algorithm 3.

Algorithm 2 starts with $\mathbf{c}_i(0) = 0.1\mathbf{m}_i (i = 1, \cdots, 4)$. Let $d_i(0) = 1$, $\beta = 0.92$, $\mu = 0.998$ and $\mathbf{Q}(0) = 0.01\mathbf{I}$. For algorithm 3, $\mathbf{W}[0] = 0.1[\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4]$, $\mathbf{Q}[0] = 0.1\mathbf{I}_N$, $\mathbf{P}[0] = \mathbf{I}_r$ and forgetting factors $\beta = 0.9$, $\mu = \eta = 0.998$. The initial parameters of FRLSL algorithm are taken according to [14]. For RDEVD method, the initial values are given as $\mathbf{R}_x[0] = 10\mathbf{I}_N$, $\mathbf{R}_y[0] = 10\mathbf{I}_N$, and the forgetting factor 0.998. 500 Monte Carlo simulations are performed for these algorithms. Figs. 1–5 show the simulation results.

As the simulation results show, the RDEVD outperforms other algorithms, but its computation complexity is too expensive for practical applications. The proposed sequential algorithm (Algorithm 2) has better performance than FRLSL. However, because of using deflation technique, it suffers the error propagation which causes the slow convergence of the minor generalized eigenvectors. In fact, this problem exists in all sequential algorithms including FRLSL algorithm. This disadvantage may cause sequential-type algorithms unfit for extracting the multiple principal generalized eigenvectors of nonstationary sequences. It can also be seen that the four different generalized eigenvectors corresponding to the modified algorithm (Algorithm 3) have almost same convergence speed. This is not surprising because Algorithm 3 seeks four principal generalized eigenvectors in parallel, and is not influenced by the error propagation. Therefore, it is more suitable for nonstationary sequences than the sequential-type algorithms. From

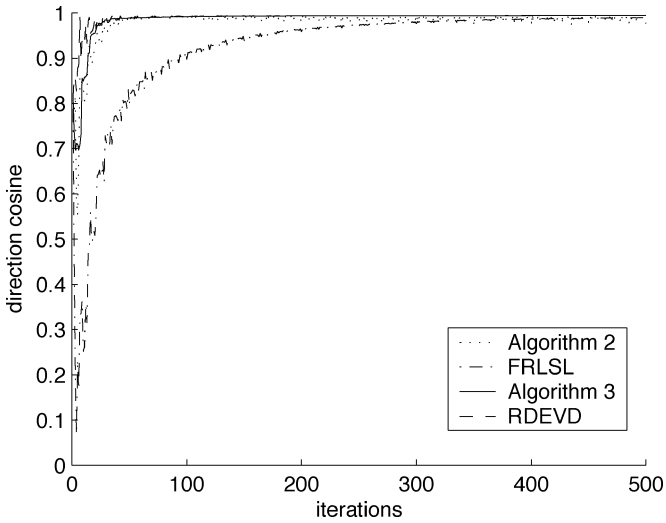


Fig. 1. Direction cosines of the first dominant generalized eigenvector estimated by using four different algorithms.

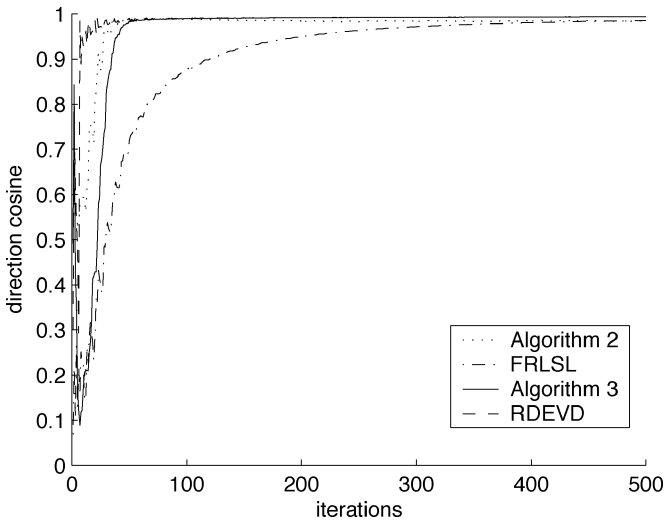


Fig. 2. Direction cosines of the second dominant generalized eigenvector estimated by using four different algorithms.

Figs. 1–5, we can observe that the steady distance between Algorithm 3 and RDEVD are almost same after about 100 iterations. This means that the performance of Algorithm 3 approaches to RDEVD algorithm. In addition, it can be also noted that the steady estimation accuracy of the sequential algorithm is almost same as that of FRLS algorithm which does not use projection approximation method, and similarly the steady performance of the modified parallel algorithm is nearly equal to that of RDEVD method which also does not adopt the projection approximation technique. This shows that at the beginning, the projection approximation can influence the performance of the proposed algorithms, but after enough iterations, its effect will be weakened largely, and even can be neglected. Hence, simulation results also confirm the reasonability of using the projection approximation in the derivation of the proposed algorithms.

In the following simulations, we investigate the influence of the projection approximation on the $\mathbf{R}_x$-orthonormality of

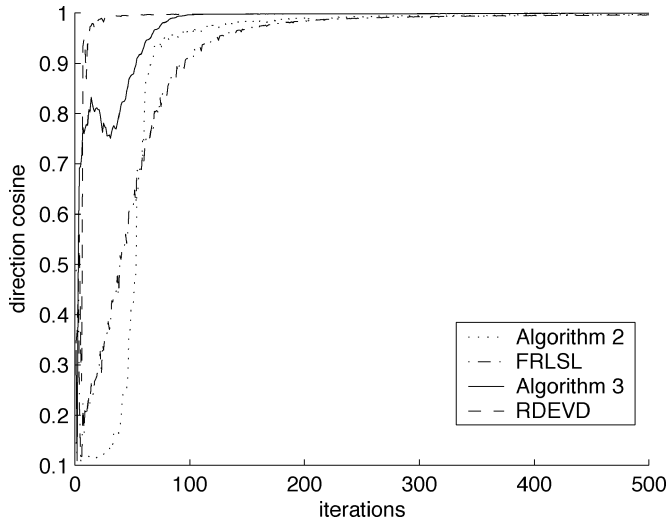


Fig. 3. Direction cosines of the third dominant generalized eigenvector estimated by using four different algorithms.

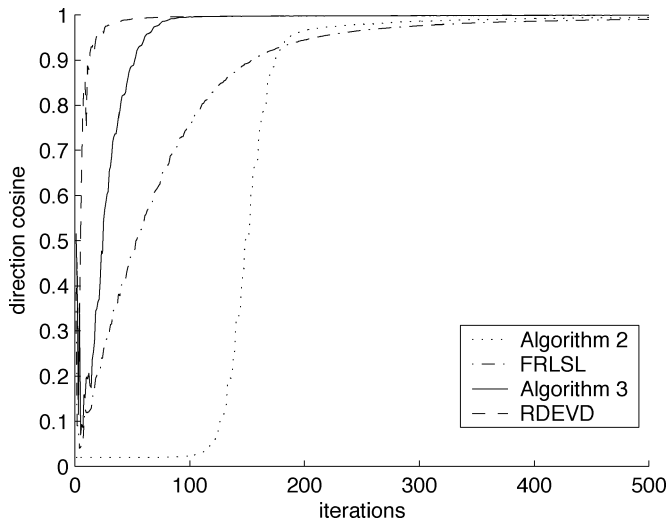


Fig. 4. Direction cosines of the fourth dominant generalized eigenvector estimated by using four different algorithms.

$\mathbf{W}[k]$ estimated by using the proposed algorithms. For comparison, the deviation from $\mathbf{R}_x$-orthonormality is defined, as

$$\text{deviation}[k] = \left\|\mathbf{W}^H[k]\mathbf{R}_x\mathbf{W}[k] - \mathbf{I}\right\|_F. \qquad (55)$$

It should be noted that because different algorithms have used different methods to estimate the correlation matrix $\mathbf{R}_x$, the resulted correlation matrix $\mathbf{R}_x[n]$ may be not unbiased estimation of $\mathbf{R}_x$, such as the exponentially averaging estimation method. Hence, before using the formula to calculate the deviation, we $\mathbf{R}_x$-normalize the columns of $\mathbf{W}[n]$. Fig. 6 and Fig. 7 show the simulation results. From Fig. 6, it can be seen that $\mathbf{R}_x$-orthonormalization during every update of the proposed sequential algorithm does not help to improve the performance after about 200 iterations, although it may work at the beginning. Moreover, the deviations of Algorithm 2, Algorithm 2 with $\mathbf{R}_x$-orthonormalization and FRLSL algorithm which does not use the projection approximation are almost uniform after about 200 iterations.

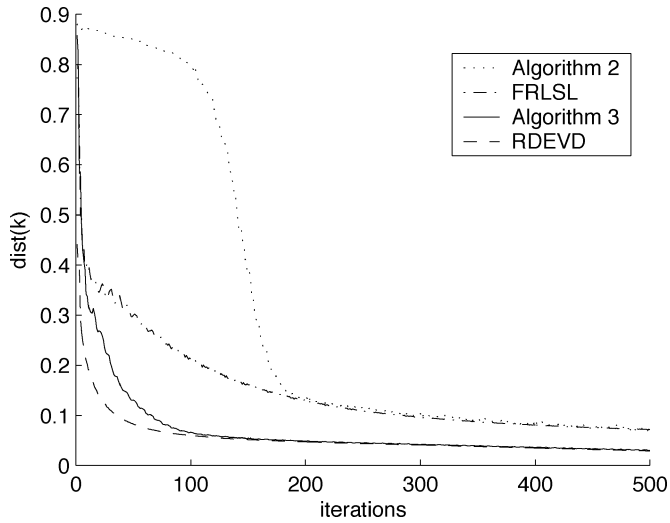


Fig. 5. Subspace distance for four different algorithms.

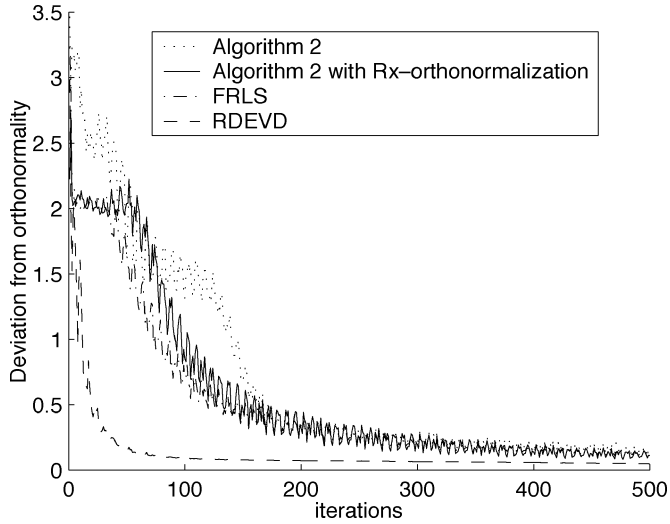


Fig. 6. Deviation from $\mathbf{R}_x$-orthonormality for sequential algorithms.

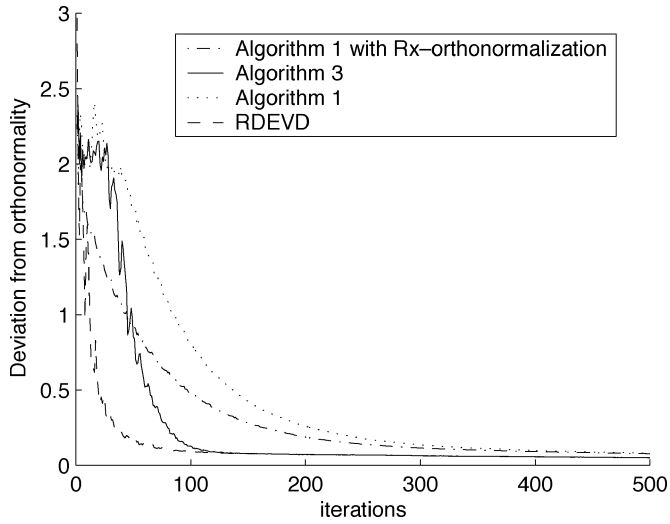


Fig. 7. Deviation from $\mathbf{R}_x$-orthonormality for parallel algorithms.

This illustrates that the influence of the projection approximation on the proposed sequential algorithm are reduced largely as

iteration increases. From Fig. 7, we also observe that the $\mathbf{R}_x$-orthonormalization does not improve the performance of the parallel algorithm after about 400 iterations. Furthermore, it can be observed that the steady deviations of the modified parallel algorithm and RDEVD which does not employ the projection approximation technique are almost same. This implies that after enough iterations, the influence of the projection approximation is neglectable. This also confirms rationality of using the projection approximation method in proposed algorithms. In addition, we note that the curves corresponding to the sequential-type algorithms are somewhat fluctuant in comparison with the parallel-type algorithms. This may imply that the sequentially processing affects the deviation because of error propagation.

B. Experiment 2

In CDMA systems, the signal model at the base station equipped with a linear uniform array with N antenna elements can be described as [19]

$$\mathbf{x}(t)=\sum_{i=1}^{K}\alpha_i A_i b_i(t-\tau_i)c_i(t-\tau_i)e^{-j\phi_i}\mathbf{a}(\theta_i)+\mathbf{n}(t) \quad (56)$$

where $x_i(t)(1\le i\le N)$, the ith element of the vector $\mathbf{x}(t)$, is the output of the ith receiving antenna, K is the number of the users, A_i is the transmitting amplitude of the ith user, α_i accounts for the overall effects of path loss and/or fading for the propagation multipath, ϕ_i and τ_i are the phase shift and the propagation delay of the ith user. $b_i(t)$ and $c_i(t)$ are the symbol waveform with interval T and the spreading waveform with chip interval T_c. $G=T/T_c$ is the processing gain. $\mathbf{a}(\theta_i)$ is the array response vector in the direction θ_i. $\mathbf{n}(t)$ is the additive white Gaussian noise vector.

Assuming perfect synchronization, after matched filter with interval T_c, the received signal vector of the ith user at the chip time $g(0\le g\le G-1)$ during the nth bit can be written as

$$\mathbf{x}_i[g;n]=[x_{i,1}[g;n],\cdots,x_{i,N}[g;n]]^T. \quad (57)$$

Through the PN-correlators [4], the received signal vector of the ith user during the nth bit can be given as

$$\mathbf{y}_i[n]=[y_{i,1}[n],\cdots,y_{i,N}[n]]^T. \quad (58)$$

The received signal vectors $\mathbf{x}_i[g;n]$ and $\mathbf{y}_i[n]$ are referred to as un-despreaded and despreaded received signal vectors, respectively. Assuming the first user is the desired user. For notational convenience, we will omit the user index 1.

The beamforming problem is to find the weight vector to generate an optimal beam for the desired user. In [4], a modified criterion that is equivalent to the maximum signal-to-interference-plus-noise ratio (MSINR) performance criterion [18] is provided

$$\mathbf{w}_{\mathrm{MSINR}}=\arg\max_{\mathbf{w}}\frac{\mathbf{w}^H\mathbf{R}_y\mathbf{w}}{\mathbf{w}^H\mathbf{R}_x\mathbf{w}} \quad (59)$$

where $\mathbf{R}_y=E[\mathbf{y}[n]\mathbf{y}^H[n]]$ and $\mathbf{R}_x=E[\mathbf{x}[g;n]\mathbf{x}^H[g;n]]$. According to the matrix theory, the optimal vector $\mathbf{w}_{\mathrm{MSINR}}$ is the most dominant generalized eigenvector of $(\mathbf{R}_y,\mathbf{R}_x)$.

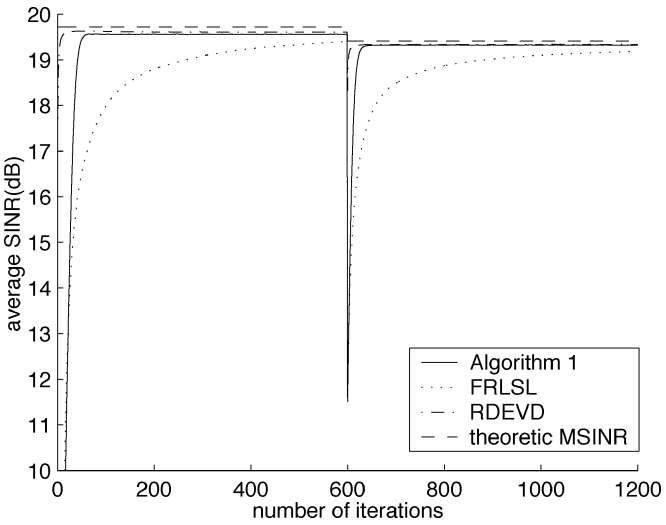


Fig. 8. SINR performance for three different algorithms.

In the simulation environment, we consider the system with an uniform linear array of $N=12$ antenna elements. Each user is assigned a golden sequence with length of $G=31$ as his signature sequence. It is assumed that the direction of arrival (DOA) of the desired user is zero and those of other users are distributed uniformly in $[-\pi/2,\pi/2]$. The desired user has the unit energy received power and SNR of 20 dB. At the beginning of the simulation, powers of three, two, and two interferers are 10-dB, 20-dB, and 30-dB stronger than that of the desired user, respectively. At the time 600, two 10-dB, three 20-dB and one 30-dB stronger interfering users are added to the system.

For comparison, in the simulation we will implement Algorithm 1 with $r=1$, fast RLS-like algorithm and RDEVD method simultaneously. For Algorithm 1, the initial values $\mathbf{w}[0]=[1\ 0\cdots 0]^T$, $P(0)=1$, $\mathbf{Q}[0]=\mathbf{I}$ and the forgetting factors $\beta=0.9$ and $\mu=0.998$ are taken. The fast RLS-like algorithm presented in [14] is fit for real signals. We reformulate it to be able to process complex signals. The initial value $\mathbf{w}[0]=[1\ 0\cdots 0]^T$ is taken, and other parameters are same as those presented in [14]. For RDEVD, the initial values are same as those in Experiment 1. 500 Monte Carlo simulations are performed for each algorithm. Figs. 8–10 show the simulation results.

Fig. 8 shows that the proposed algorithm has faster convergence than FRLSL. At time 600, although the environment suddenly changes, Algorithm 1 can adapt fast to track the optimal beamforming vector. FRLSL algorithm is also able to track the optimal beamforming vector in dynamical environments, but its convergence is somewhat slower. It can be also seen that the steady SINR performance of the proposed algorithm approaches to that of RDEVD algorithm. Fig. 9 and Fig. 10 show the optimum beam pattern and the one corresponding to the proposed algorithm at iteration 80 and 680, respectively. The point marked with the black point represents the desired user and the others marked with the asterisks represent the interferers. As shown in figures, the main beam is accurately along the DOA of the desired user, and the interference from the undesired users is suppressed largely. Moreover, the beam pattern corresponding to the proposed algorithm almost coincides with the optimum

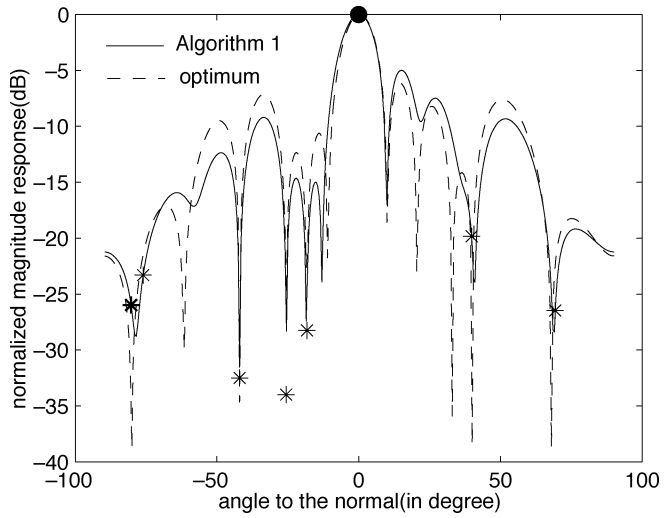


Fig. 9. Beam pattern for the proposed algorithm at time 80.

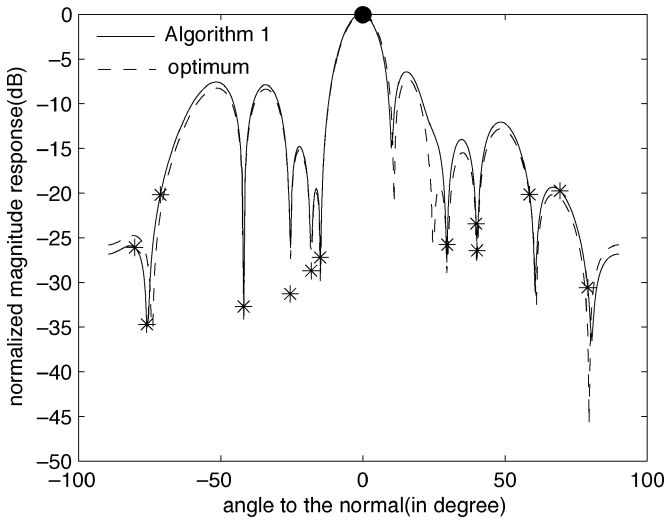


Fig. 10. Beam pattern for the proposed algorithm at time 680.

one. The fast convergence and excellent tracking ability of the proposed algorithm make it very useful in practical signal environments, because the interfering power increases due to many practical reasons, e.g., incorrect power control, various transmitting power, and too many interference users, etc.

VI. Conclusion

In this paper, we study adaptive algorithms for the generalized eigen-decomposition from two sequences of sample vectors. A novel unconstrained cost function for the generalized eigen-decomposition has been presented, and an adaptive recursive parallel algorithm has been developed to track the dominant generalized eigen-subspace with RLS learning rule. By the deflation technique, we have extended the parallel algorithm in one vector case to explicitly compute the dominant generalized eigenvectors sequentially. Furthermore, a modified algorithm is also presented to extract multiple generalized eigenvectors in parallel. We also have provided convergence analysis for the proposed algorithms. Several computer simulations are performed to investigate the performance of the proposed algorithms, and the simulation results show the applicability and efficiency of the algorithms.

Appendix I
Proof of the Two Theorems in Section II

Proof: (for Theorem I): Let $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_r]$ and $\nabla = [\nabla_1, \cdots, \nabla_r]$. ∇_i is the gradient operator with respect to $\mathbf{w}_i$. According to [8], [16], the complex gradient operator is defined as $\nabla_i = (1/2)[\nabla_{R,i} + j\nabla_{I,i}]$, where $\nabla_{R,i}$ and $\nabla_{I,i}$ are the gradient operators with respect to the real and imaginary part of $\mathbf{w}_i$. Then, we have

$$\nabla_i J = \left[-2\mathbf{R}_y + \mathbf{R}_x\mathbf{W}\mathbf{W}^H\mathbf{R}_y + \mathbf{R}_y\mathbf{W}\mathbf{W}^H\mathbf{R}_x\right]\mathbf{w}_i \quad (60)$$

$$\nabla J = \left[-2\mathbf{R}_y + \mathbf{R}_x\mathbf{W}\mathbf{W}^H\mathbf{R}_y + \mathbf{R}_y\mathbf{W}\mathbf{W}^H\mathbf{R}_x\right]\mathbf{W}. \quad (61)$$

If $\mathbf{W} = \mathbf{V}_r\mathbf{Q}$, where $\mathbf{Q}$ is unitary and $\mathbf{V}_r$ contains any r distinct generalized eigenvectors of $(\mathbf{R}_y, \mathbf{R}_x)$. Then, we can get following results:

$$\mathbf{R}_y\mathbf{V}_r = \mathbf{R}_x\mathbf{V}_r\mathrm{diag}\{\lambda_1, \cdots, \lambda_r\} \quad (62)$$

$$\mathbf{R}_y\mathbf{W} = \mathbf{R}_x\mathbf{W}\mathbf{Q}^H\mathrm{diag}\{\lambda_1, \cdots, \lambda_r\}\mathbf{Q} \quad (63)$$

$$\mathbf{W}^H\mathbf{R}_x\mathbf{W} = \mathbf{I} \quad (64)$$

$$\mathbf{W}^H\mathbf{R}_y\mathbf{W} = \mathbf{Q}^H\mathrm{diag}\{\lambda_1, \cdots, \lambda_r\}\mathbf{Q}. \quad (65)$$

Hence

$$\begin{aligned}\nabla J|_{W=V_rQ} =& -2\mathbf{R}_y\mathbf{W} + \mathbf{R}_x\mathbf{W}\mathbf{W}^H\mathbf{R}_y\mathbf{W} \\ &+ \mathbf{R}_y\mathbf{W}\mathbf{W}^H\mathbf{R}_x\mathbf{W} \\ =& -2\mathbf{R}_x\mathbf{W}\mathbf{Q}^H\mathrm{diag}\{\lambda_1, \cdots, \lambda_r\}\mathbf{Q} \\ &+ \mathbf{R}_x\mathbf{W}\mathbf{Q}^H\mathrm{diag}\{\lambda_1, \cdots, \lambda_r\}\mathbf{Q} \\ &+ \mathbf{R}_x\mathbf{W}\mathbf{Q}^H\mathrm{diag}\{\lambda_1, \cdots, \lambda_r\}\mathbf{Q} \\ =& 0. \end{aligned} \quad (66)$$

Therefore, $\mathbf{W} = \mathbf{V}_r\mathbf{Q}$ is the stationary point of $J(\mathbf{W})$.

Contrarily, from the fact $\nabla J = 0$, we have

$$\begin{aligned}\mathbf{W}^H\nabla J =& -2\mathbf{W}^H\mathbf{R}_y\mathbf{W} + \mathbf{W}^H\mathbf{R}_x\mathbf{W}\mathbf{W}^H\mathbf{R}_y\mathbf{W} \\ &+ \mathbf{W}^H\mathbf{R}_y\mathbf{W}\mathbf{W}^H\mathbf{R}_x\mathbf{W} \\ =& 0. \end{aligned} \quad (67)$$

Then

$$\mathbf{W}^H\mathbf{R}_y\mathbf{W}(\mathbf{W}^H\mathbf{R}_x\mathbf{W} - \mathbf{I}_r) + (\mathbf{W}^H\mathbf{R}_x\mathbf{W} - \mathbf{I}_r)\mathbf{W}^H\mathbf{R}_y\mathbf{W} = 0. \quad (68)$$

Since both $\mathbf{W}^H\mathbf{R}_y\mathbf{W}$ and $\mathbf{W}^H\mathbf{R}_x\mathbf{W}$ are symmetric and $\mathbf{W}^H\mathbf{R}_y\mathbf{W}$ is positive definite, according to Lemma in [8] we conclude $\mathbf{W}^H\mathbf{R}_x\mathbf{W} = \mathbf{I}_r$, and the rank of $\mathbf{W}$ is r. Therefore, combining $\nabla J = 0$, we have

$$\mathbf{R}_y\mathbf{W} = \mathbf{R}_x\mathbf{W}\mathbf{W}^H\mathbf{R}_y\mathbf{W}. \quad (69)$$

Let $\mathbf{W}^H\mathbf{R}_y\mathbf{W} = \mathbf{Q}^H\Sigma_r\mathbf{Q}$, where the matrix $\mathbf{Q}$ is orthogonal and Σ_r is a diagonal matrix. Then, we can attain $\mathbf{R}_y\mathbf{V}_r = \mathbf{R}_x\mathbf{V}_r\Sigma_r$, where $\mathbf{V}_r = \mathbf{W}\mathbf{Q}^H$. Furthermore, from the fact that $(\mathbf{C}^H\mathbf{R}_y\mathbf{C})(\mathbf{C}^{-1}\mathbf{V}_r) = (\mathbf{C}^{-1}\mathbf{V}_r)\Sigma_r$ and Σ_r is a diagonal matrix, the columns of the full rank matrix $\mathbf{C}^{-1}\mathbf{V}_r$ must be r distinct eigenvectors of $\mathbf{C}^H\mathbf{R}_y\mathbf{C}$, which also means that

$\mathbf{V}_r$ contains r distinct generalized eigenvectors of $(\mathbf{R}_y, \mathbf{R}_x)$. Therefore, we have $\mathbf{W} = \mathbf{V}_r\mathbf{Q}$. Without loss of generality, we assume the generalized eigenvalues are arbitrarily ordered and $\mathbf{V}_r$ contains the first r generalized eigenvectors. Then, we get following result:

$$J(\mathbf{W})|_{W=V_rQ} = \sum_{i=1}^{N}\lambda_i - 2\sum_{i=1}^{r}\lambda_i + \sum_{i=1}^{r}\lambda_i = \sum_{i=r+1}^{N}\lambda_i. \quad (70)$$

■

Proof (for Theorem 2): Let λ_i and $\mathbf{v}_i$, $1 \le i \le N$, are, respectively, the generalized eigenvalues and the corresponding $\mathbf{R}_x$-orthonormal generalized eigenvectors of $(\mathbf{R}_y, \mathbf{R}_x)$. Without loss of generality, the nonincreasingly ordered generalized eigenvalues are given by

$$\lambda_1 > \cdots > \lambda_r > \lambda_{r+1} \ge \cdots \ge \lambda_N > 0. \quad (71)$$

According to Theorem 1, any $\mathbf{W} = \mathbf{V}_r\mathbf{Q}$ is the stationary point of $J(\mathbf{W})$. Let the set $\mathbf{S}_I = \{i_1, \cdots, i_r\}$, where the elements of $\mathbf{S}_I$ are the indexes of the generalized eigenvectors, which make up the matrix $\mathbf{V}_r$. Let $\tilde{\mathbf{S}}_I = \{1, \cdots, r\}$, the corresponding matrix $\tilde{\mathbf{V}}_r$ and $\tilde{\mathbf{W}} = \tilde{\mathbf{V}}_r\mathbf{Q}$.

For any $\mathbf{S}_I (\mathbf{S}_I \ne \tilde{\mathbf{S}}_I)$, there exists j, $j \in \mathbf{S}_I$ and $j \notin \tilde{\mathbf{S}}_I = \{1, \cdots, r\}$. Then the component $\mathbf{v}_j$ of the matrix $\mathbf{V}_r$ is substituted by component $\mathbf{v}_j + \varepsilon\mathbf{v}_k$, where $k \in \tilde{\mathbf{S}}_I$, $k \notin \mathbf{S}_I$ and $\forall\varepsilon > 0$. Moreover, we have $\lambda_j < \lambda_k$. Let $\mathbf{V}'_r$ be the new matrix, and $\mathbf{W}' = \mathbf{V}'_r\mathbf{Q}$. Then, after some calculations, we get

$$J(\mathbf{W}') - J(\mathbf{W}) = (\lambda_j - \lambda_k)\varepsilon^2 + O(\varepsilon^4). \quad (72)$$

It is shown that along the direction of the component $\mathbf{v}_k$, $J(\mathbf{W})$ will decrease. In addition, if the component $\mathbf{v}_j$ is substituted by component $\mathbf{v}_j + \varepsilon\mathbf{v}_j$. Let $\mathbf{V}''_r$ be the new matrix, and $\mathbf{W}'' = \mathbf{V}''_r\mathbf{Q}$. Then, it is easy to show that

$$J(\mathbf{W}'') - J(\mathbf{W}) = 4\lambda_j\varepsilon^2 + O(\varepsilon^3). \quad (73)$$

This means along the direction of the component $\mathbf{v}_j$, $J(\mathbf{W})$ will increase. Therefore, $\mathbf{W}$ is a saddle point.

Conversely, it can be shown that if any component of $\tilde{\mathbf{V}}_r$ is perturbed by the component $\mathbf{v}_k$, where $1 \le k \le N$, $J(\tilde{\mathbf{W}})$ will increase.

Therefore, $\tilde{\mathbf{W}}$ is the unique local minimum. This means $J(\mathbf{W})$ has a global minimum without any other local minimum. It can be easily computed

$$J(\tilde{\mathbf{W}}) = \sum_{i=r+1}^{N}\lambda_i \quad (74)$$

■

APPENDIX II
PROOF OF THE THEOREM IN SECTION III

Proof (for Theorem 3): Let $\mathbf{R}_{\mathbf{y}_i} = E[\mathbf{y}_i\mathbf{y}_i^H]$. We then have

$$\mathbf{R}_{\mathbf{y}_i} = \left(\mathbf{I} - \mathbf{R}_x\mathbf{w}_{i-1}\mathbf{w}_{i-1}^H\right)\mathbf{R}_{\mathbf{y}_{i-1}}\left(\mathbf{I} - \mathbf{w}_{i-1}\mathbf{w}_{i-1}^H\mathbf{R}_x\right). \quad (75)$$

Since $\mathbf{w}_i = \pm\mathbf{v}_i$, for $1 \le i \le j-1$, it is easy to verify that

$$\mathbf{R}_{\mathbf{y}_j}\mathbf{v}_i = \mathbf{0}, \quad \text{for } 1 \le i \le j-1 \quad (76)$$

$$\mathbf{R}_{\mathbf{y}_j}\mathbf{v}_i = \lambda_i\mathbf{R}_x\mathbf{v}_i, \quad \text{for } j \le i \le N. \quad (77)$$

Hence, $\pm\mathbf{v}_i (1 \le i \le N)$, the generalized eigenvectors of $(\mathbf{R}_y, \mathbf{R}_x)$, are also the generalized eigenvectors of $(\mathbf{R}_{\mathbf{y}_j}, \mathbf{R}_x)$, and the corresponding generalized eigenvalues are $(0, \cdots, 0, \lambda_j, \cdots, \lambda_N)$. Since $\lambda_j > \cdots > \lambda_r > \lambda_{r+1} \ge \cdots \ge \lambda_N > 0$, the generalized eigenvector $\pm\mathbf{v}_j$ is the most dominant one of $(\mathbf{R}_{\mathbf{y}_j}, \mathbf{R}_x)$.

On the other hand, from Theorem 1 and Theorem 2, the most dominant generalized eigenvector of $(\mathbf{R}_{\mathbf{y}_j}, \mathbf{R}_x)$ is the unique global minimal point of $J_j(\mathbf{w}_j)$, and the others are saddle points of $J_j(\mathbf{w}_j)$. Since the generalized eigenvector $\pm\mathbf{v}_j$ is the most dominant one of $(\mathbf{R}_{\mathbf{y}_j}, \mathbf{R}_x)$, $\mathbf{w}_j = \pm\mathbf{v}_j$ is the unique global minimal point of $J_j(\mathbf{w}_j)$ and $\pm\mathbf{v}_i (i \ne j)$ are saddle points of $J_j(\mathbf{w}_j)$. ■

APPENDIX III
PROOF OF THE THEOREMS IN SECTION IV

Proof (for Theorem 4): Let $\mathbf{Z} = \mathbf{W}^H\mathbf{R}_x\mathbf{W}$. We use the same Liapunov function as that in [21]

$$L_1(\mathbf{Z}) = \left\|\mathbf{R}^{\frac{1}{2}}(\mathbf{Z} - \mathbf{I})\mathbf{R}^{\frac{1}{2}}\right\|_F^2 = \mathrm{tr}\left[(\mathbf{Z} - \mathbf{I})\mathbf{R}(\mathbf{Z} - \mathbf{I})\mathbf{R}\right] \quad (78)$$

Note that the (49) is same as (18a) in [21]. We can conclude that the equilibrium state $\mathbf{Z} = \mathbf{I}$ is the asymptotical state by using same derivation.

In order to prove that $\mathbf{W}(t)$ converges to $\mathbf{V}_r\mathbf{Q}$, we use the cost function (4) as Liapunov function

$$L_2(\mathbf{W}) = \mathrm{tr}(\mathbf{C}^H\mathbf{R}_y\mathbf{C}) - 2\mathrm{tr}(\mathbf{W}^H\mathbf{R}_y\mathbf{W}) + \mathrm{tr}\left((\mathbf{W}^H\mathbf{R}_y\mathbf{W})(\mathbf{W}^H\mathbf{R}_x\mathbf{W})\right). \quad (79)$$

Since $\tilde{\mathbf{W}} = \mathbf{V}_r\mathbf{Q}$ is the global minimum according to Theorem 2, we have following result:

$$L_2(\mathbf{W}) > L_2(\tilde{\mathbf{W}}) \quad (80)$$

where $\mathbf{W} \neq \mathbf{V}_r\mathbf{Q}$. Since $\mathbf{W}^H\mathbf{R}_x\mathbf{W} \to \mathbf{I}$, dL_2/dW for $t \to \infty$ can be simplified as $(\mathbf{I}-\mathbf{R}_x\mathbf{W}\mathbf{W}^H)\mathbf{R}_y\mathbf{W}$. Then combining the (50), the time derivative of L_2 can be expressed as

$$\begin{aligned}\dot{L}_2 &= \operatorname{tr}\left(\dot{\mathbf{W}}\left(\frac{dL_2}{dW}\right)^H\right)\\ &= -\operatorname{tr}\Big(\mathbf{R}_x^{-1}(\mathbf{I}-\mathbf{R}_x\mathbf{W}\mathbf{W}^H)\\ &\qquad \times\mathbf{R}_y\mathbf{W}\mathbf{R}^{-1}\mathbf{W}^H\mathbf{R}_y(\mathbf{I}-\mathbf{W}\mathbf{W}^H\mathbf{R}_x)\Big)\\ &= -\operatorname{tr}\Big(\mathbf{R}_x^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}_x\mathbf{W}\mathbf{W}^H)\mathbf{R}_y\mathbf{W}\mathbf{R}^{-1}\mathbf{W}^H\mathbf{R}_y\\ &\qquad \times(\mathbf{I}-\mathbf{W}\mathbf{W}^H\mathbf{R}_x)\mathbf{R}_x^{-\frac{1}{2}}\Big) \qquad (81)\end{aligned}$$

According to Lemma 1 in [21], there exists $\xi > 0$ such that $\mathbf{R} \leq \xi\mathbf{I}$. Hence, we have

$$\begin{aligned}\dot{L}_2 &\leq -\frac{1}{\xi}\operatorname{tr}\Big[\mathbf{R}_x^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}_x\mathbf{W}\mathbf{W}^H)\\ &\qquad \times\mathbf{R}_y\mathbf{W}\mathbf{W}^H\mathbf{R}_y(\mathbf{I}-\mathbf{W}\mathbf{W}^H\mathbf{R}_x)\mathbf{R}_x^{-\frac{1}{2}}\Big]\\ &= -\frac{1}{\xi}\left\|\mathbf{R}_x^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}_x\mathbf{W}\mathbf{W}^H)\mathbf{R}_y\mathbf{W}\right\|_F^2. \qquad (82)\end{aligned}$$

According to Theorem 1, $\dot{L}_2(\tilde{\mathbf{W}}) = 0$ and $\dot{L}_2(\mathbf{W}) < 0$ for $\mathbf{W} \in \{\mathbf{W}|\mathbf{W}^H\mathbf{v}_i \neq 0, 1 \leq i \leq r, \mathbf{W} \text{ full rank}\}$ and $\mathbf{W} \neq \tilde{\mathbf{W}}$. This means that $\tilde{\mathbf{W}}$ is the asymptotically stable equilibrium. Its domain of attraction is $\{\mathbf{W}|\mathbf{W}^H\mathbf{v}_i \neq 0, 1 \leq i \leq r, \mathbf{W} \text{ full rank}\}$. This completes the proof. ■

Proof (for Theorem 5): In the sequential algorithm, the procedure for the second dominant generalized eigenvector $\mathbf{w}_2(k)$ is same as for the first one except for the use of the modified input data $\mathbf{y}_2(k) = [\mathbf{I}-\mathbf{R}_x\mathbf{w}_1(k)\mathbf{w}_1^H(k)]\mathbf{y}(k)$ due to the deflation step. Because $\mathbf{w}_1(k)$ converges to $\pm\mathbf{v}_1$, $\mathbf{y}_2(k)$ tends to $[\mathbf{I}-\mathbf{R}_x\mathbf{v}_1\mathbf{v}_1^H]\mathbf{y}(k)$ as well. Hence, the correlation matrix of $\mathbf{y}_2$ is $\mathbf{R}_{y_2} = \mathbf{R}_y - \lambda_1\mathbf{R}_x\mathbf{v}_1\mathbf{v}_1^H\mathbf{R}_x$. It is easy to show that the generalized eigenvalues of $(\mathbf{R}_{y_2}, \mathbf{R}_x)$ are $\lambda_2, \cdots, \lambda_n, 0$ and corresponding generalized eigenvectors are $\mathbf{v}_2, \cdots, \mathbf{v}_N, \mathbf{v}_1$. Therefore, according to Theorem 4 in one vector case $r = 1$, $\mathbf{w}_2(k)$ converges to $\pm\mathbf{v}_2$. Using the same method, the convergence of the other minor dominant generalized eigenvectors can be also shown sequentially. ■

REFERENCES

[1] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*. New York: Wiley, 1996.

[2] H. Tao, H. H. Chen, W. Wu, and T. S. Huang, "Compression of MPEG-4 facial animation parameters for transmission of talking heads," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 2, pp. 264–276, Mar. 1999.

[3] L. Matthews, T. F. Cootes, J. A. Bangham, S. Cox, and R. Harvey, "Extraction of visual features for lipreading," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 198–213, Feb. 2002.

[4] S. Choi, J. Choi, H. J. Im, and B. Choi, "A novel adaptive beamforming algorithm for antenna array CDMA systems with strong interferers," *IEEE Trans. Veh. Technol.*, vol. 51, no. 5, pp. 808–816, Sep. 2002.

[5] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," *Proc. IEEE*, vol. 78, pp. 1327–1343, Aug. 1990.

[6] A. Bunse-Gerstner, "An algorithm for the symmetric generalized eigenvalue problem," *Linear Algebra Appl.*, vol. 58, pp. 43–68, 1984.

[7] W. Shougen and Z. Shuqin, "An algorithm for $\mathbf{Ax} = \lambda\mathbf{Bx}$ with symmetric and positive-definite $\mathbf{A}$ and $\mathbf{B}$," *SIAM J. Matrix Anal. Appl.*, vol. 12, pp. 654–660, Oct. 1991.

[8] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[9] Y. F. Miao and Y. Hua, "Fast subspace tracking and neural network learning by a novel information criterion," *IEEE Trans. Signal Process.*, vol. 46, no. 7, pp. 1967–1979, Jul. 1998.

[10] G. Mathew and V. U. Reddy, "A Quasi-Newton adaptive algorithm for generalized symmetric eigenvalue problem," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2413–2422, Oct. 1996.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computation*. Baltimore, MD: The John Hopkins Univ. Press, 1991.

[12] C. Chatterjee, V. P. Roychowdhury, J. Ramos, and M. D. Zoltowski, "Self-organizing algorithms for generalized eigen-decomposition," *IEEE Trans. Neural Netw.*, vol. 8, no. 6, pp. 1518–1530, Nov. 1997.

[13] D. X. Xu, J. C. Principe, and H. C. Wu, "Generalized eigendecomposition with an on-line local algorithm," *IEEE Signal Process. Lett.*, vol. 5, no. 11, pp. 298–301, Nov. 1998.

[14] Y. N. Rao, J. C. Principe, and T. F. Wong, "Fast RLS-like algorithm for generalized eigendecomposition and its applications," *J. VLSI Signal Process.*, vol. 37, pp. 333–344, Jun. 2004.

[15] T. F. Wong, T. M. Lok, J. S. Lehnert, and M. D. Zoltowski, "A linear receiver for direct-sequence spread-spectrum multiple-access systems with antenna arrays and blind adaptation," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 659–676, Mar. 1998.

[16] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[17] P. Strobach, "Fast orthogonal iteration adaptive algorithms for the generalized symmetric eigenproblem," *IEEE Trans. Signal Process.*, vol. 46, no. 12, pp. 3345–3359, Dec. 1998.

[18] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*. New York: Wiley, 1980.

[19] A. F. Naguib, "Adaptive Antennas for CDMA Wireless Networks," Ph.D. dissertation, Elect. Eng. Dept., Stanford Univ., Stanford, CA, 1996.

[20] L. Lung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Autom. Control*, vol. AC-22, no. 4, pp. 551–575, Aug. 1977.

[21] B. Yang, "Asymptotic convergence analysis of the projection approximation subspace tracking algorithms," *Signal Process.*, vol. 50, pp. 123–136, 1996.

Jian Yang received the B.S. and M.S. degrees from the University of Science and Technology of China (USTC), Hefei, China, in 2001 and 2003, respectively. He is currently working toward the Ph.D. degree in the School of Information Science and Technology of USTC.

His research interests include signal processing for communications with focus on multiuser detection, adaptive antenna array processing, and timing estimation in code-division multiple access.

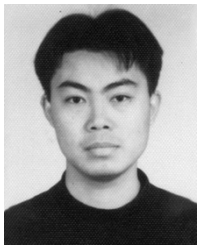
Mr. Yang received the Huawei Award in 2004 and the President Scholarship of Chinese Academy of Sciences in 2005.

Hongsheng Xi received the B.S. and M.S. degrees in applied mathematics from the University of Science and Technology of China (USTC), Hefei, China, in 1980 and 1985, respectively.

He is currently the Dean of the Department of Automation at USTC. He also directs the Laboratory of Network Communication System and Control. His research interests include stochastic control systems, discrete-event dynamic systems, network performance analysis and optimization, and wireless communications.

Feng Yang received the B.S. degree in electrical engineering from TongJi University, Shanghai, China, in 2001 and the M.S. degree from the University of Science and Technology of China (USTC), Hefei, China, in 2003. He is currently working toward the Ph.D. degree at the same university.

His current research interests include adaptive filtering theory, multicarrier code-division multiple-access systems, and multiple-input multiple-output systems.



Yu Zhao received the B.S. and M.S. degrees from the University of Science and Technology of China (USTC), Hefei, China, in 2001 and 2003, respectively. He is currently working toward the Ph.D. degree in the School of Information Science and Technology of USTC.

His current research interests include support vector machine (SVM) and its application for wireless communications.