STABILIZATION OF ADAPTIVE EIGENVECTOR EXTRACTION BY CONTINUATION IN NESTED ORTHOGONAL COMPLEMENT STRUCTURE

Kenji Kakimoto *Daichi Kitahara* *Masao Yamagishi* *Isao Yamada*

Department of Communications and Computer Engineering, Tokyo Institute of Technology, Japan

ABSTRACT

Nguyen and Yamada [NY'13] proposed an adaptive algorithm for fast and stable extraction of the first generalized Hermitian eigenvector and mentioned the extension to the first r generalized eigenvector extraction based on the nested orthogonal complement structure [NTY'12]. However, we recently found that the estimates of the eigenvectors are not expressed ideally in the time-varying coordinate system and can change drastically in a certain situation, which may cause numerical instability. In this paper, we propose a new expression of the estimates along with time-varying coordinate system. This modification can be done efficiently with additional multiplications of orthogonal complement matrices. Numerical experiments show that the modified scheme has better stability compared with the original scheme [NTY'12].

Index Terms— Generalized Hermitian eigenvalue problem (GHEP), Adaptive algorithm, Stabilization, Orthogonal complement matrix, Nested orthogonal complement structure

1. INTRODUCTION

Generalized Hermitian eigenvalue problem (GHEP) is an estimation problem, for a pair of Hermitian positive definite matrices $(R_y, R_x) \in \mathbb{C}^{N\times N} \times \mathbb{C}^{N\times N}$, of vectors $\boldsymbol{v}_i \in \mathbb{C}^N \setminus \{\boldsymbol{0}\}$ satisfying

$$R_y \boldsymbol{v}_i = \lambda_i R_x \boldsymbol{v}_i \quad \text{s.t. } \boldsymbol{v}_i^H R_x \boldsymbol{v}_j = \delta_{i,j}\ (i,j = 1,2,\ldots,N), \quad (1)$$

where λ_i $(0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_N)$ is called the ith minor generalized eigenvalue, $(\cdot)^H$ stands for the conjugate transpose of a vector or a matrix, and $\delta_{i,j}$ is the Kronecker delta function. In this paper, (R_y, R_x) and $\boldsymbol{v}_i$ are respectively called a matrix pencil and the ith minor generalized eigenvector of (R_y, R_x). The GHEP has been attracting great attention in many branches of signal processing, e.g., subspace tracking [1], [2], blind source separation [3], fault detection [4], pattern recognition [5], and array signal processing [6]–[10].

Adaptive estimation of $\boldsymbol{v}_i$ is also required and some adaptive estimators are proposed [11]–[13]. In our previous work [14], we proposed an adaptive estimator of the first minor (or principal) generalized eigenvector for fast and stable estimation. In [15], we also proposed the scheme for tracking the first r minor (or principal) generalized eigenvectors, which achieved the lower estimation error and the robustness against additive noise compared with [11]–[13] (see [15, Figs. 1 and 3]) while keeping the orthogonality. This scheme reduced estimation problem of $\boldsymbol{v}_i$ $(i = 1,\ldots,r)$ of (R_y, R_x) to that of the first minor generalized eigenvector $\boldsymbol{v}_1^{(i)} \in \mathbb{C}^{N-i+1}$ of a certain smaller matrix pencil $(R_y^{(i)}, R_x^{(i)})$ $(i = 1,\ldots,r)$ by using the nested orthogonal complement structure (Section 2.1). The ith minor generalized eigenvector $\boldsymbol{v}_i$ can be calculated from $\boldsymbol{v}_1^{(i)}$ because $\boldsymbol{v}_1^{(i)}$ is an expression of $\boldsymbol{v}_i$ in a certain coordinate system $\perp_i \in \mathbb{C}^{N\times(N-i+1)}$, i.e., $\boldsymbol{v}_i = \perp_i \boldsymbol{v}_1^{(i)}$. In application to adaptive estimation (Section 2.2), this relation is extended with the use of the time-varying coordinate system $\perp_i(k)$. By combining the algorithm [14] with the scheme [15], a fast adaptive estimator of the first r minor generalized eigenvectors is obtained. However, we recently found that the scheme [15] sometimes causes numerical instability in a certain situation.

In this paper, we analyse the reason of such instability and propose an effective way to resolve this issue. Since the scheme [15] assumed implicitly that the time-varying coordinate system changes smoothly, the expression in the previous coordinate system is reused in the current coordinate system. However, if the time-varying coordinate system changes relatively fast, then the consistency between the previous expression and the current coordinate system is lost, which leads to the numerical instability (Section 3.1). To cope with this significant change of the coordinate system, we propose to use the expression of the nearest vector from the previous estimate in the current coordinate system (Section 3.2). This modification can be done with low computational complexity just by multiplying orthogonal complement matrices (Section 3.3). Numerical experiments in a scenario of adaptive subspace tracking show the excellent performance of the proposed stabilization.

2. PRELIMINARIES

Let $\mathbb{R}$ and $\mathbb{C}$ be respectively the set of all real numbers and complex numbers. Capital and bold face small letters respectively express a matrix and a vector. The Euclidean norm and the B-norm of $\boldsymbol{x} := (x_1, x_2, \ldots, x_N)^T \in \mathbb{C}^N$ are respectively defined as $\|\boldsymbol{x}\| := \sqrt{\sum_{i=1}^N |x_i|^2}$ and $\|\boldsymbol{x}\|_B := \sqrt{\boldsymbol{x}^H B \boldsymbol{x}}$, where $B \in \mathbb{C}^{N\times N}$ is a Hermitian positive definite matrix and $(\cdot)^T$ stands for the transpose.

For self-containedness, we present a summary of the adaptive eigenvector extraction in [15].

2.1. Nested Orthogonal Complement Structure

Definition 1 (B-orthogonal complement matrix) *Let $B \in \mathbb{C}^{N\times N}$ be a Hermitian positive definite matrix. For $\boldsymbol{u} \in \mathbb{C}^N \setminus \{\boldsymbol{0}\}$, $U_\perp \in \mathbb{C}^{N\times(N-1)}$ is called a B-orthogonal complement matrix of $\boldsymbol{u}$ if*

$$U_\perp^H B \boldsymbol{u} = \boldsymbol{0} \ \text{ and } \ U_\perp^H U_\perp = I_{N-1}, \quad (2)$$

where $I_{N-1} \in \mathbb{R}^{(N-1)\times(N-1)}$ is the identity matrix.

Fact 1 (Example of a B-orthogonal complement matrix) *One of B-orthogonal complement matrices of $\boldsymbol{u}$ can be calculated as*

$$U_\perp = \begin{bmatrix} I_{N-1} - \frac{1}{1+|\bar{u}_{\rm low}|}\bar{\boldsymbol{u}}_{\rm up}\bar{\boldsymbol{u}}_{\rm up}^H \\ -\theta(\bar{u}_{\rm low})\bar{\boldsymbol{u}}_{\rm up}^H \end{bmatrix}, \quad (3)$$

where $\bar{\boldsymbol{u}}_{\rm up} \in \mathbb{C}^{N-1}$ and $\bar{u}_{\rm low} \in \mathbb{C}$ are respectively the first $(N-1)$ components and the last component of a normalized vector $\bar{\boldsymbol{u}} := B\boldsymbol{u}/\|B\boldsymbol{u}\|$, i.e., $\bar{\boldsymbol{u}} = [\bar{\boldsymbol{u}}_{\rm up}^T, \bar{u}_{\rm low}]^T$, and $\theta : \mathbb{C} \to \mathbb{C}$ is defined as

$$\theta(\bar{u}_{\rm low}) := \begin{cases} 1, & \text{if } \bar{u}_{\rm low} = 0; \\ \bar{u}_{\rm low}/|\bar{u}_{\rm low}|, & \text{otherwise.} \end{cases} \quad (4)$$

This work was supported in part by JSPS Grants-in-Aid (15K13986) (e-mail: {kakimoto, kitahara, myamagi, isao}@sp.ce.titech.ac.jp).

Algorithm 1 Normalized Quasi-Newton Algorithm [14]

With any B-normalized vector $\boldsymbol{u}(0) \in \mathbb{C}^N$ and $\lambda(0) \geq 0$, generate the sequence $(\boldsymbol{u}(k), \lambda(k)) \in (\mathbb{C}^N \setminus \{\boldsymbol{0}\}) \times \mathbb{R}$ $(k = 0, 1, \dots)$ by

$$\begin{aligned}\widehat{\boldsymbol{u}}(k+1) &:= \boldsymbol{u}(k) + \eta\Big[A^{-1}B\boldsymbol{u}(k)\lambda(k) \\ &\qquad + \boldsymbol{u}^H(k)A\boldsymbol{u}(k)\boldsymbol{u}(k)\lambda^{-1}(k) - 2\boldsymbol{u}(k)\Big] \\ \boldsymbol{u}(k+1) &:= \widehat{\boldsymbol{u}}(k+1)/\|\widehat{\boldsymbol{u}}(k+1)\|_B \\ \lambda(k+1) &:= (1-\gamma)\lambda(k) + \gamma\boldsymbol{u}^H(k+1)A\boldsymbol{u}(k+1)\end{aligned}$$

with step sizes $\eta > 0$ and $\gamma \in (0, 1]$, where $(\boldsymbol{u}(k), \lambda(k))$ are the estimates of the first minor generalized eigenvector and eigenvalue.

Remark 1 *The B-orthogonal complement matrix in (2) is not a trivial extension of the standard orthogonal complement matrix [16] satisfying $U_\perp^H \boldsymbol{u} = \boldsymbol{0}$ and $U_\perp^H U_\perp = I_{N-1}$. In Definition 1, we use the condition $U_\perp^H U_\perp = I_{N-1}$ in place of a trivial one $U_\perp^H B U_\perp = I_{N-1}$, which is a key for efficient computation of (3).*

We reduce the estimation problem of the ith minor generalized eigenvector $\boldsymbol{u}_i^{(1)} \in \mathbb{C}^N$ of a matrix pencil $(A, B) \in \mathbb{C}^{N\times N} \times \mathbb{C}^{N\times N}$ to that of the first minor generalized eigenvector $\boldsymbol{u}_1^{(i)} \in \mathbb{C}^{N-i+1}$ of a certain smaller matrix pencil $(A^{(i)}, B^{(i)})$ as follows.

Fact 2 (Nested orthogonal complement structure) *Define N matrix pencils $(A^{(i)}, B^{(i)})$ recursively as $(A^{(1)}, B^{(1)}) := (A, B)$ and*

$$\begin{cases} A^{(i+1)} := \left(U_\perp^{(i)}\right)^H A^{(i)} U_\perp^{(i)} \in \mathbb{C}^{(N-i)\times(N-i)} \\ B^{(i+1)} := \left(U_\perp^{(i)}\right)^H B^{(i)} U_\perp^{(i)} \in \mathbb{C}^{(N-i)\times(N-i)} \end{cases} \quad (i = 1, 2, \dots, N-1),$$

where $U_\perp^{(i)}$ is a $B^{(i)}$-orthogonal complement matrix of the first minor generalized eigenvector $\boldsymbol{u}_1^{(i)}$ of $(A^{(i)}, B^{(i)})$. Then the ith minor generalized eigenvector $\boldsymbol{u}_i^{(1)}$ of (A, B) is expressed as

$$\boldsymbol{u}_i^{(1)} = U_\perp^{(1)} U_\perp^{(2)} \cdots U_\perp^{(i-1)} \boldsymbol{u}_1^{(i)} =: \perp_i \boldsymbol{u}_1^{(i)}. \tag{5}$$

In (5), we can regard $\boldsymbol{u}_1^{(i)} \in \mathbb{C}^{N-i+1}$ as an expression of $\boldsymbol{u}_i^{(1)} \in \mathbb{C}^N$ in the coordinate system $\perp_i \in \mathbb{C}^{N\times(N-i+1)}$. By combining Fact 2 with Algorithm 1 (Normalized Quasi-Newton Algorithm [14]), which iteratively estimates the first minor generalized eigenvector, we established the following scheme for estimation of the first r minor generalized eigenvectors.

Scheme 1 (Extraction of the first r minor generalized eigenvectors $\boldsymbol{u}_i^{(1)} (i = 1, \dots, r)$ of a matrix pencil (A, B) by using Fact 2 [15])

1. *Set $A^{(1)} = A$ and $B^{(1)} = B$.*
2. *For $i = 1, \dots, r$*
   a. *Extract the first minor generalized eigenvector $\boldsymbol{u}_1^{(i)}$ of the matrix pencil $(A^{(i)}, B^{(i)})$.*
   b. *Compute a $B^{(i)}$-orthogonal complement matrix $U_\perp^{(i)}$ of $\boldsymbol{u}_1^{(i)}$.*
   c. *Set $A^{(i+1)} = \left(U_\perp^{(i)}\right)^H A^{(i)} U_\perp^{(i)}$, $B^{(i+1)} = \left(U_\perp^{(i)}\right)^H B^{(i)} U_\perp^{(i)}$.*
3. *For $i = 2, \dots, r$, compute $\boldsymbol{u}_i^{(1)}$ by (5).*

2.2. Adaptive Estimation of Generalized Eigenvectors

In many signal processing applications, the matrix pencil (R_y, R_x) in (1) are defined as a pair of covariance matrices of input sequences $(\boldsymbol{y}(k))_{k\geq 0}$ and $(\boldsymbol{x}(k))_{k\geq 0}$, where k denotes discrete time index. In adaptive case, we have to estimate the matrix pencil (R_y, R_x) and its generalized eigenvectors simultaneously. As an adaptive version of Scheme 1, we proposed the following scheme [15] where $\boldsymbol{w}_i^{(1)}(k)$ is the estimate of the ith minor generalized eigenvector $\boldsymbol{v}_i$ at time k.

Scheme 2 (Adaptive version of Scheme 1)

1. *Update the estimate $(R_y^{(1)}(k), R_x^{(1)}(k))$ of (R_y, R_x).*
2. *For $i = 1, \dots, r$*
   a. *Update the estimate $\boldsymbol{w}_1^{(i)}(k)$, from $\boldsymbol{w}_1^{(i)}(k-1)$, of the first minor generalized eigenvector of the matrix pencil $(R_y^{(i)}(k), R_x^{(i)}(k))$.*
   b. *Compute an $R_x^{(i)}(k)$-orthogonal complement matrix $W_\perp^{(i)}(k)$ of $\boldsymbol{w}_1^{(i)}(k)$.*
   c. *Set $R_y^{(i+1)}(k) = \left(W_\perp^{(i)}(k)\right)^H R_y^{(i)}(k) W_\perp^{(i)}(k)$, $R_x^{(i+1)}(k) = \left(W_\perp^{(i)}(k)\right)^H R_x^{(i)}(k) W_\perp^{(i)}(k)$.*
3. *For $i = 2, \dots, r$, compute $\boldsymbol{w}_i^{(1)}(k) = \left(\prod_{s=1}^{i-1} W_\perp^{(s)}(k)\right)\boldsymbol{w}_1^{(i)}(k)$.*
4. *$k \leftarrow k+1$. Repeat 1–4 until $\boldsymbol{w}_i^{(1)}(k)$ $(i = 1, \dots, r)$ converge.*

In [15], $W_\perp^{(i)}(k)$ in Scheme 2-2b is calculated as an adaptive version of (3), i.e.,

$$W_\perp^{(i)}(k) = \begin{bmatrix} I_{N-i} - \frac{1}{1+|\bar{w}_{\mathrm{low}}^{(i)}(k)|}\, \bar{\boldsymbol{w}}_{\mathrm{up}}^{(i)}(k)\left(\bar{\boldsymbol{w}}_{\mathrm{up}}^{(i)}(k)\right)^H \\ -\theta\left(\bar{w}_{\mathrm{low}}^{(i)}(k)\right)\left(\bar{\boldsymbol{w}}_{\mathrm{up}}^{(i)}(k)\right)^H \end{bmatrix}, \tag{6}$$

where $\bar{\boldsymbol{w}}_{\mathrm{up}}^{(i)}(k) \in \mathbb{C}^{N-i}$ and $\bar{w}_{\mathrm{low}}^{(i)}(k) \in \mathbb{C}$ satisfy $\bar{\boldsymbol{w}}_1^{(i)}(k) := R_x^{(i)}(k)\boldsymbol{w}_1^{(i)}(k)/\|R_x^{(i)}(k)\boldsymbol{w}_1^{(i)}(k)\| = [(\bar{\boldsymbol{w}}_{\mathrm{up}}^{(i)}(k))^T, \bar{w}_{\mathrm{low}}^{(i)}(k)]^T$. Moreover, an example of $(R_y^{(1)}(k), R_x^{(1)}(k))$ is shown in Section 4.

3. PROPOSED STABILIZATION OF ADAPTIVE EIGENVECTOR EXTRACTION

3.1. Instability in Time-Varying Coordinate System

We find that Scheme 2 has certain instability in the estimation of the non-first minor generalized eigenvectors $\boldsymbol{v}_i$ $(i = 2, \dots, r)$. In Scheme 2, since the non-first minor generalized eigenvectors are estimated by using the time-varying coordinate systems

$$\perp_i(k) := \prod_{s=1}^{i-1} W_\perp^{(s)}(k) \tag{7}$$

as $\boldsymbol{w}_i^{(1)}(k) = \perp_i(k)\boldsymbol{w}_1^{(i)}(k)$, the significant change of $\perp_i(k)$ directly influences the estimation accuracy of $\boldsymbol{v}_i$. To see this, we focus on Scheme 2-2a. In this step, if the time-varying coordinate systems change smoothly, i.e., $\perp_i(k-1) \approx \perp_i(k)$, then the estimation of $\boldsymbol{v}_i$ based on the update of $\boldsymbol{w}_1^{(i)}(k)$ from $\boldsymbol{w}_1^{(i)}(k-1)$ is stable because $\perp_i(k-1)\boldsymbol{w}_1^{(i)}(k-1) \approx \perp_i(k)\boldsymbol{w}_1^{(i)}(k)$ can be expected. However, if the coordinate systems change significantly, i.e., $\perp_i(k-1) \not\approx \perp_i(k)$, then the estimation of $\boldsymbol{v}_i$ based on the update of $\boldsymbol{w}_1^{(i)}(k)$ from $\boldsymbol{w}_1^{(i)}(k-1)$ is unstable because $\perp_i(k-1)\boldsymbol{w}_1^{(i)}(k-1) \not\approx \perp_i(k)\boldsymbol{w}_1^{(i)}(k)$.

Such a significant change of $\perp_i(k)$ is based on those of $W_\perp^{(s)}(k)$ $(s = 1, \dots, i-1)$ due to (7), and a significant change of $W_\perp^{(s)}$ is caused by the discontinuity of θ in (4) (see Example 1 below).

Example 1 (Simple example of the significant change of $W_\perp^{(s)}(k)$) *For simplicity, set $N = 2$, $R_x^{(1)}(k) = R_x^{(1)}(k+1) = I_2$, and $\boldsymbol{w}_1^{(1)}(k) = \bar{\boldsymbol{w}}_1^{(1)}(k) \in \mathbb{R}^2$ (in this case, $\theta(\bar{w}_{\mathrm{low}}^{(1)}(k))$ returns the sign of $w_{\mathrm{low}}^{(1)}(k)$). Figure 1 shows an example of the significant change of the orthogonal complement matrix $W_\perp^{(1)}(k)$ (in this case, $W_\perp^{(1)}(k) \in \mathbb{R}^{2\times 1}$). In Fig. 1, green dotted line stands for $\boldsymbol{w}_1^{(1)}(k) = (w_{\mathrm{up}}(k), \delta)^T$ with a small positive value $\delta \in (0, \sqrt{2}/2]$, and red dotted line stands for $W_\perp^{(1)}(k) = (\delta, -w_{\mathrm{up}}(k))^T$ computed by (6). Suppose that $\boldsymbol{w}_1^{(1)}(k)$ is updated*

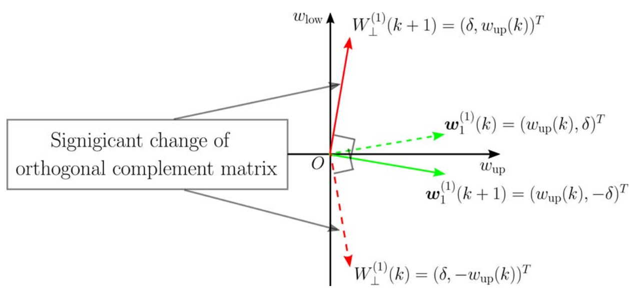


Fig. 1. Example of significant change of $W_\perp^{(1)}(k) \in \mathbb{R}^{2\times 1}$ in (6).

to $\boldsymbol{w}_1^{(1)}(k+1) = (w_{\rm up}(k), -\delta)^T$ *(green solid line), and* $W_\perp^{(1)}(k)$ *is updated to* $W_\perp^{(1)}(k+1) = (\delta, w_{\rm up}(k))^T$ *(red solid line). Obviously, the orthogonal complement matrix changes significantly. In fact, the distance between* $W_\perp^{(1)}(k+1)$ *and* $W_\perp^{(1)}(k)$

$$\left\|W_\perp^{(1)}(k+1) - W_\perp^{(1)}(k)\right\| = 2\sqrt{w_{\rm up}^2(k)} = 2\sqrt{1-\delta^2}$$

implies that $W_\perp^{(1)}(k)$ *changes significantly as* $\delta \to +0$ *while* $\|\boldsymbol{w}_1^{(1)}(k+1) - \boldsymbol{w}_1^{(1)}(k)\| = 2\delta \to +0$ *as* $\delta \to +0$.

On the other hand, the absolute value of the inner product between $W_\perp^{(1)}(k+1)$ *and* $W_\perp^{(1)}(k)$

$$\left|\left(W_\perp^{(1)}(k+1)\right)^T W_\perp^{(1)}(k)\right| = \left|\delta^2 - w_{\rm up}^2(k)\right| = 1 - 2\delta^2.$$

approaches 1 *as* $\delta \to +0$*, which means that the range spaces of* $W_\perp^{(1)}(k+1)$ *and* $W_\perp^{(1)}(k)$ *approach the same subspace as* $\delta \to +0$.

3.2. Continuation in Nested Orthogonal Complement Structure

We stabilize Scheme 2 by maximally utilizing the information of the previous estimate $\boldsymbol{w}_i^{(1)}(k-1) = \perp_i(k-1)\boldsymbol{w}_1^{(i)}(k-1)$ for the update of $\boldsymbol{w}_1^{(i)}(k)$. From Example 1, even if the orthogonal complement matrix $W_\perp^{(s)}(k)$ ($s = 1, \ldots, i-1$) changes significantly, its range space is expected to change smoothly. Therefore, the range space of the time-varying coordinate system $\perp_i(k)$ is also expected to change smoothly from (7). This motivates the use of the expression which stands for the best approximation of the previous estimate $\boldsymbol{w}_i^{(1)}(k-1)$ in the current coordinate system $\perp_i(k)$, i.e.,

$$\begin{aligned}
&\mathop{\arg\min}_{\boldsymbol{w}\in\mathcal{R}(\perp_i(k))} \left\|\boldsymbol{w}_i^{(1)}(k-1) - \boldsymbol{w}\right\|^2 \\
&= \perp_i(k)\left((\perp_i(k))^H \perp_i(k)\right)^{-1}(\perp_i(k))^H \boldsymbol{w}_i^{(1)}(k-1) \quad (8)\\
&= \perp_i(k)\left(\perp_i(k)\right)^H \boldsymbol{w}_i^{(1)}(k-1), \quad (9)
\end{aligned}$$

where $\mathcal{R}(\perp_i(k))$ is the range space of $\perp_i(k)$, (8) is derived as the projection of $\boldsymbol{w}_i^{(1)}(k-1)$ onto $\mathcal{R}(\perp_i(k))$, and (9) is derived from (2) and (7). Since $\boldsymbol{w}_i^{(1)}(k-1) = \perp_i(k-1)\boldsymbol{w}_1^{(i)}(k-1) \approx \perp_i(k)\left(\perp_i(k)\right)^H \boldsymbol{w}_i^{(1)}(k-1)$, it is stable to update $\boldsymbol{w}_1^{(i)}(k)$ from $(\perp_i(k))^H \boldsymbol{w}_i^{(1)}(k-1)$ instead of $\boldsymbol{w}_1^{(i)}(k-1)$. More precisely, we propose to update $\boldsymbol{w}_1^{(i)}(k)$ from an $R_x^{(i)}(k)$-normalized vector

$$\widetilde{\boldsymbol{w}}_1^{(i)}(k-1) := \frac{\left(\perp_i(k)\right)^H \boldsymbol{w}_i^{(1)}(k-1)}{\left\|\left(\perp_i(k)\right)^H \boldsymbol{w}_i^{(1)}(k-1)\right\|_{R_x^{(i)}(k)}} \quad (10)$$

instead of $\boldsymbol{w}_1^{(i)}(k-1)$, and we propose the following scheme.

Scheme 3 (Proposed scheme for stable adaptive estimation)

1. *Update the estimate* $(R_y^{(1)}(k), R_x^{(1)}(k))$ *of* (R_y, R_x).
2. *For* $i = 1, \ldots, r$
   a. *If* $i \geq 2$*, compute* $\widetilde{\boldsymbol{w}}_1^{(i)}(k-1)$ *by (10).*
   b. *Update the estimate* $\boldsymbol{w}_1^{(i)}(k)$*, from* $\widetilde{\boldsymbol{w}}_1^{(i)}(k-1)$*, of the first minor generalized eigenvector of the matrix pencil* $(R_y^{(i)}(k), R_x^{(i)}(k))$.
   c. *Compute an* $R_x^{(i)}(k)$*-orthogonal complement matrix* $W_\perp^{(i)}(k)$ *of* $\boldsymbol{w}_1^{(i)}(k)$.
   d. *Set* $R_y^{(i+1)}(k) = \left(W_\perp^{(i)}(k)\right)^H R_y^{(i)}(k) W_\perp^{(i)}(k)$, $R_x^{(i+1)}(k) = \left(W_\perp^{(i)}(k)\right)^H R_x^{(i)}(k) W_\perp^{(i)}(k)$.
3. *For* $i = 2, \ldots, r$*, compute* $\boldsymbol{w}_i^{(1)}(k) = \left(\prod_{s=1}^{i-1} W_\perp^{(s)}(k)\right)\boldsymbol{w}_1^{(i)}(k)$.
4. $k \leftarrow k+1$. *Repeat 1–4 until* $\boldsymbol{w}_i^{(1)}(k)$ *(*$i = 1, \ldots, r$*) converge.*

3.3. Additional Complexity of Proposed Stabilization

The complexity of Scheme 3 does not increase so much compared with Scheme 2. To see this, we shall observe the complexity of Scheme 3-2a. The calculation of (10) can be separated into two steps, $(\perp_i(k))^H \boldsymbol{w}_i^{(1)}(k-1)$ and $R_x^{(i)}(k)$-normalization. From (7),

$$\begin{aligned}
&(\perp_i(k))^H \boldsymbol{w}_i^{(1)}(k-1)\\
&\quad = \left(W_\perp^{(i-1)}(k)\right)^H \cdots \left(W_\perp^{(1)}(k)\right)^H \boldsymbol{w}_i^{(1)}(k-1), \quad (11)
\end{aligned}$$

which requires $(i-1)$ times multiplications of vectors and orthogonal complement matrices. Fortunately, we can calculate each multiplication with low complexity by using (6). In the following discussion, time index k is omitted for simplicity. For $W_\perp^{(s)} \in \mathbb{C}^{(N-s+1)\times(N-s)}$ and any $\boldsymbol{t} = [\boldsymbol{t}_{\rm up}^T, t_{\rm low}]^T \in \mathbb{C}^{N-s+1}$, we have

$$\begin{aligned}
\left(W_\perp^{(s)}\right)^H \boldsymbol{t} &= \left[I_{N-s} - \frac{\bar{\boldsymbol{w}}_{\rm up}^{(s)}\left(\bar{\boldsymbol{w}}_{\rm up}^{(s)}\right)^H}{1+|\bar{w}_{\rm low}^{(s)}|},\ -\theta\left(\bar{w}_{\rm low}^{(s)}\right)\bar{\boldsymbol{w}}_{\rm up}^{(s)}\right]\begin{bmatrix}\boldsymbol{t}_{\rm up}\\ t_{\rm low}\end{bmatrix}\\
&= \boldsymbol{t}_{\rm up} - \left(\frac{\left(\bar{\boldsymbol{w}}_{\rm up}^{(s)}\right)^H \boldsymbol{t}_{\rm up}}{1+|\bar{w}_{\rm low}^{(s)}|} + \theta\left(\bar{w}_{\rm low}^{(s)}\right)t_{\rm low}\right)\bar{\boldsymbol{w}}_{\rm up}^{(s)}. \quad (12)
\end{aligned}$$

Since (12) can be computed with $\mathcal{O}(2(N-s))$ multiplications, by counting all multiplications over $s = 1, \ldots, i-1$, the calculation of (11) requires $\mathcal{O}(2Ni - i^2)$ multiplications.

The $R_x^{(i)}$-normalization needs $\mathcal{O}((N-i+1)^2)$ multiplications for calculation of $(\boldsymbol{w}_1^{(i)})^H R_x^{(i)} \boldsymbol{w}_1^{(i)}$. Since (10) is calculated for $i = 2, \ldots, r$, the total computational complexity of Scheme 3-2a is $\mathcal{O}(Nr^2 - r^3/6) + \mathcal{O}(N^2 r - Nr^2 + r^3/6) = \mathcal{O}(N^2 r)$. This computational complexity is much smaller than the total computational complexity $\mathcal{O}(12N^2 r - 11Nr^2 + 11r^3/6)$ of Scheme 2 [15].

4. APPLICATION TO SUBSPACE TRACKING

In Scheme 2 and Scheme 3, we estimate the matrix pencil (R_y, R_x) as a pair of exponential weighted sample covariance matrices

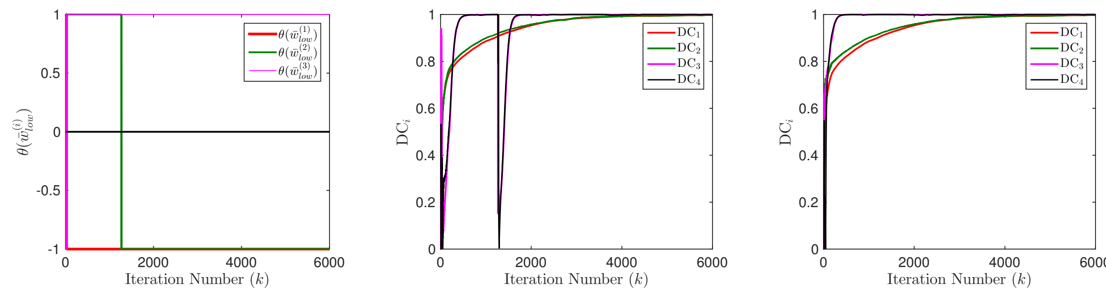
$$\begin{cases} R_y^{(1)}(k) = \beta R_y^{(1)}(k-1) + \boldsymbol{y}(k)\boldsymbol{y}(k)^H \\ R_x^{(1)}(k) = \alpha R_x^{(1)}(k-1) + \boldsymbol{x}(k)\boldsymbol{x}(k)^H \end{cases} \quad (13)$$

with forgetting factors $\alpha, \beta \in (0, 1)$.

4.1. Performance Criteria

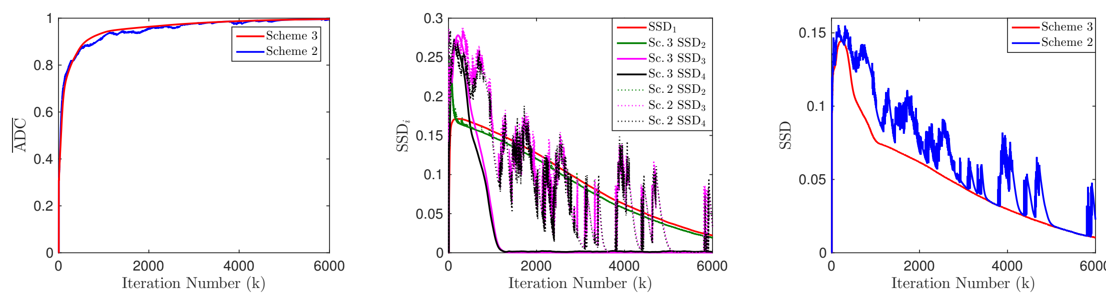
In a scenario of application to subspace tracking, we evaluate the performance of Scheme 3 (proposed scheme) compared with Scheme 2 through L $(= 100)$ independent runs. For comparison, we observe the similarity between $\boldsymbol{v}_i$ and $\boldsymbol{w}_{i,j}^{(1)}(k)$ (the ith minor generalized eigenvector of (R_y, R_x) and the estimate of $\boldsymbol{v}_i$ at time k in the jth independent run) in terms of Direction Cosine and its average

$$\mathrm{DC}_{i,j}(k) := \frac{|(\boldsymbol{w}_{i,j}^{(1)}(k))^H \boldsymbol{v}_i|}{\|\boldsymbol{w}_{i,j}^{(1)}(k)\|\|\boldsymbol{v}_i\|} \quad \text{and} \quad \mathrm{ADC}_j(k) := \frac{1}{r}\sum_{i=1}^{r}\mathrm{DC}_{i,j}(k).$$



(a) $\theta(\bar{w}^{(i)}_{\text{low}})$ of Scheme 2 (b) $\mathrm{DC}_{i,j}$ of Scheme 2 (c) $\mathrm{DC}_{i,j}$ of Scheme 3

Fig. 2. One of the outcomes in 100 independent runs.



(a) $\overline{\mathrm{ADC}}$ (b) SSD_i (c) SSD

Fig. 3. Results in 100 independent runs (Sc. 3 and Sc. 2 respectively stand for Scheme 3 and Scheme 2 in (b)).

Define the averages of $\mathrm{DC}_{i,j}(k)$ and $\mathrm{ADC}_j(k)$ in L independent runs as $\overline{\mathrm{DC}}_i(k) := \frac{1}{L}\sum_{j=1}^{L}\mathrm{DC}_{i,j}(k)$ and $\overline{\mathrm{ADC}}(k) := \frac{1}{L}\sum_{j=1}^{L}\mathrm{ADC}_j(k)$. We also measure numerical stabilities by two kinds of Sample Standard Deviations

$$\mathrm{SSD}_i(k) := \sqrt{\frac{1}{L-1}\sum_{j=1}^{L}\left(\mathrm{DC}_{i,j}(k) - \overline{\mathrm{DC}}_i(k)\right)^2}$$

and

$$\mathrm{SSD}(k) := \sqrt{\frac{1}{L-1}\sum_{j=1}^{L}\left(\mathrm{ADC}_j(k) - \overline{\mathrm{ADC}}(k)\right)^2}.$$

4.2. Numerical Experiments

The input samples are generated by

$$y(k) = \sqrt{2}\sin(0.37\pi k + \theta_1) + n_1(k)$$

and

$$x(k) = \sqrt{2}\sin(0.42\pi k + \theta_2) + \sqrt{2}\sin(0.65\pi k + \theta_3) + n_2(k),$$

where the initial phase θ_i $(i = 1, 2, 3)$ has the uniform distribution in $[0, 2\pi]$, $n_1(k)$ and $n_2(k)$ are white Gaussian noise with variance $\sigma^2 = 0.1$. The input vectors $\boldsymbol{y}(k) \in \mathbb{R}^N$ and $\boldsymbol{x}(k) \in \mathbb{R}^N$ $(N = 8)$ are defined as $\boldsymbol{y}(k) := (y(k), y(k-1), \ldots, y(k-N+1))^T$ and $\boldsymbol{x}(k) := (x(k), x(k-1), \ldots, x(k-N+1))^T$ $(k \geq N)$. [1]

We adaptively estimate the first four $(r = 4)$ minor generalized eigenvectors $\boldsymbol{v}_i$ $(i = 1, 2, 3, 4)$ of the matrix pencil (R_y, R_x) with the parameters $\alpha = \beta = 0.998$, $\eta = 1/(\lambda_N/\lambda_1 - 1)$, $\gamma = 0.998$ and the initial estimates $R_y^{(1)}(0) = R_x^{(1)}(0) = I_N$.

Figure 2 shows one of the outcomes in L $(= 100)$ independent runs. Figure 2(a) depicts $\theta(\bar{w}^{(i)}_{\text{low}})$ $(i = 1, 2, 3)$. In this case, $\theta(\bar{w}^{(i)}_{\text{low}})$ returns the sign of $\bar{w}^{(i)}_{\text{low}}$. Figures 2(b) and 2(c) respectively depict $\mathrm{DC}_{i,j}$ $(i = 1, 2, 3, 4)$ of Scheme 2 and Scheme 3. From Fig. 2, we observe that when $\theta(\bar{w}^{(i)}_{\text{low}})$ changes significantly (i.e., when the sign of $\bar{w}^{(i)}_{\text{low}}$ changes from positive/negative to negative/positive), the estimation by Scheme 2 is unstable, as mentioned in Section 3.1, while the estimation by Scheme 3 is stable. Figure 3 shows the result in 100 independent runs. Figures 3(a), 3(b) and 3(c) respectively depict $\overline{\mathrm{ADC}}$, SSD_i and SSD of Scheme 2 and Scheme 3 (note that in Fig. 3(b), there is no difference between Scheme 2 and Scheme 3 for the estimation of the first minor generalized eigenvector $\boldsymbol{v}_1$). From these figures, we observe that Scheme 3 has better numerical stability than Scheme 2. Especially, SSD_i $(i = 3, 4)$ in Fig. 3(b) are significantly improved.

5. CONCLUSIONS

We found that the numerical instability of the original adaptive eigenvector extraction (Scheme 2) is caused by the significant change of the time-varying coordinate systems. To cope with the significant change, we proposed to use the expression of the nearest vector from the previous estimate in the current coordinate system. This proposed stabilization is realized by the projection of the previous estimate onto the range space of the current coordinate system, and its complexity is low compared with the total complexity of the original scheme. Numerical experiments showed the excellent performance of the proposed stabilization.

Acknowledgement We would like to thank Tuan Duong Nguyen (NTT Communications) for fruitful discussions.

[1] The covariance matrices $R_y, R_x \in \mathbb{R}^{N\times N}$ are given as

$$\begin{cases}(R_y)_{i,j} := \cos(0.37\pi(j-i)) + \delta_{i,j}\sigma^2,\\ (R_x)_{i,j} := \cos(0.42\pi(j-i)) + \cos(0.65\pi(j-i)) + \delta_{i,j}\sigma^2,\end{cases}$$

where $(\cdot)_{i,j}$ stands for the (i,j)-component of the matrix. These matrices are used for computing true generalized eigenvectors $\boldsymbol{v}_i$.

6. REFERENCES

[1] B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 95–107, 1995.

[2] Y. Hua, T. Chen, and W. Y. Yan, "Global convergence of Oja's subspace algorithm for principal component extraction," *IEEE Transactions on Neural Networks*, vol. 9, no. 1, 1998.

[3] C. Chang, Z. Ding, S. F. Yau, and F. H. Chan, "A matrix-pencil approach to blind separation of colored nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 900–907, 2000.

[4] H. Chen, G. Jiang, and K. Yoshihira, "Failure detection in large-scale internet services by principal subspace mapping," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 10, pp. 1308–1320, 2007.

[5] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic press, 2013.

[6] 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 Transactions on Information Theory*, vol. 44, no. 2, pp. 659–676, 1998.

[7] S. Choi, J. Choi, H. J. Im, and B. Choi, "A novel adaptive beamforming algorithm for antenna array CDMA systems with strong interferers," *IEEE Transactions on Vehicular Technology*, vol. 51, no. 5, pp. 808–816, 2002.

[8] S. Shahbazpanahi, A. B. Gershman, Z. Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2257–2269, 2003.

[9] D. R. Morgan, "Downlink adaptive array algorithms for cellular mobile communications," *IEEE Transactions on Communications*, vol. 51, no. 3, pp. 476–488, 2003.

[10] ——, "Adaptive algorithms for solving generalized eigenvalue signal enhancement problems," *Signal Processing*, vol. 84, no. 6, pp. 957–968, 2004.

[11] J. Yang, H. Xi, F. Yang, and Y. Zhao, "RLS-based adaptive algorithms for generalized eigen-decomposition," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1177–1188, 2006.

[12] S. Attallah and K. Abed-Meraim, "A fast adaptive algorithm for the generalized symmetric eigenvalue problem," *IEEE Signal Processing Letters*, vol. 15, pp. 797–800, 2008.

[13] T. Tanaka, "Fast generalized eigenvector tracking based on the power method," *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 969–972, 2009.

[14] T. D. Nguyen and I. Yamada, "Adaptive normalized quasi-newton algorithms for extraction of generalized eigen-pairs and their convergence analysis," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1404–1418, 2013.

[15] T. D. Nguyen, N. Takahashi, and I. Yamada, "An adaptive extraction of generalized eigensubspace by using exact nested orthogonal complement structure," *Multidimensional Systems and Signal Processing*, vol. 24, no. 3, pp. 457–483, 2013.

[16] M. Misono and I. Yamada, "An efficient adaptive minor subspace extraction using exact nested orthogonal complement structure," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 91, no. 8, pp. 1867–1874, 2008.