On Convergence Properties of Subspace Trackers Based on Orthogonal Iteration

Chao-Cheng Tu and Benoît Champagne
Department of Electrical and Computer Engineering
McGill University, Montréal, Québec, Canada
chao-cheng.tu@mail.mcgill.ca, benoit.champagne@mcgill.ca

Abstract

This paper studies convergence properties of subspace trackers using orthogonal iteration. In the context of blind estimation of a time-varying channel, orthogonal iteration and its variants have been widely considered for tracking the channel parameters by updating the eigen-decomposition of an exponentially weighted correlation matrix. While it is well known that orthogonal iteration converges exponentially with arbitrary initial conditions, orthogonal-iteration-based subspace trackers can only inherit these merits when the channels considered undergo extremely slow time-variations. In this paper, we generalize the traditional (i.e. fixed subspace) convergence analysis of the orthogonal iteration to include non-stationary situations as well. We use the results to investigate the convergence behavior of subspace trackers based on orthogonal iteration under slow, moderate and rapid time-variations of the underlying subspace. In the latter case, we expose a fundamental limitation of the orthogonal iteration, i.e. practical limit on subspace variations to ensure effective tracking.

1 Introduction

Subspace decomposition has proved to be an important tool in various signal processing applications. To this end, a straightforward approach is simply to employ either an eigenvalue decomposition (EVD) or singular value decomposition (SVD). These approaches, which belong to the family of direct or block processing techniques, are characterized as computationally demanding procedures and unsuitable for online processing due to their lack of repetitive structure [1]. Furthermore, they are often implemented in a batch mode, using an estimated correlation matrix obtained by collecting time samples over a sufficiently long observation interval. Therefore, these approaches, which rely on the assumption of statistical stationarity of the data, cannot be used in situations where the characteristics of the received signals change with time [2]. Computationally efficient and sequential algorithms that produce an exact or approximate EVD or SVD at each time step are generally favoured in signal processing applications.

Thus, a considerable effort has gone into the development of sequential adaptive algorithms, also known as *subspace trackers*. To date, several signal-subspace trackers have been proposed for non-stationary environments. Instead of recomputing the EVD or SVD from scratch with every update, these algorithms attempt to recursively update the EVD or SVD so as to minimize the amount of computations involved [2]. While there are many more signal-subspace trackers than noise-subspace trackers in the literature [3], nonetheless, in the applications of blind channel estimation, we can transform the noise-subspace problems into signal-subspace ones without much effort [4].

Orthogonal iteration is a simple technique that can be used to compute higher-dimensional invariant subspaces [5]. It is shown to have a global and exponential convergence property under a mild assumption on the distribution of eigenvalues, with arbitrary initial conditions [6]. In addition, it is suitable for real-time processing because it is well structured [7]. Therefore, orthogonal iteration and its variants have been considered for blind adaptive channel estimation to a great extent. Existing subspace tracking algorithms can be broadly categorized as whether or not they are based on orthogonal iteration. For the orthogonal-iteration-based subspace trackers, their variants include the low rank adaptive filter (LORAF) [7], the orthogonal projection approximation and subspace tracking (OPAST) [8], the Oja's method, and the novel information criterion (NIC) [9]. Recently, improvements on these existing approaches can also be found [3][10][11].

In this paper, given that numerous subspace trackers in the literature are fundamentally derived from the concept of orthogonal iteration, we first investigate the convergence properties when orthogonal iteration is applied in non-stationary scenarios. Specifically, we are interested in the distance between the *true* and the orthogonal-iterated subspaces. Then we study a fundamental limitation on the use of orthogonal-iterated subspace trackers when they are

978-1-4244-4561-5/09/$25.00 ©2009 IEEE PACRIM'09

applied in time-varying scenarios. Our results will be useful for better understanding the behaviors of subspace trackers based on orthogonal iteration when applied to estimate time-varying subspaces.

The notation used in this paper is as follows: A vector is denoted by a bold lower-case letter and a matrix is denoted by a bold upper-case letter. The range of $\mathbf{A} \in \mathbb{C}^{m\times n}$ is defined by $\mathfrak{R}(\mathbf{A}) := \{\mathbf{Ax} : \mathbf{x} \in \mathbb{C}^{n\times 1}\}$. $\|\mathbf{x}\|_p$ and $\|\mathbf{X}\|_p$ represent the p-norms of a vector $\mathbf{x}$ and a matrix $\mathbf{X}$, respectively. $\mathfrak{D}_r(\mathbf{X})$ denotes the subspace spanned by the eigenvectors corresponding to the r largest eigenvalues of $\mathbf{X}$. $\lambda_r(\mathbf{X})$ and $\sigma_r(\mathbf{X})$ represent the rth largest eigenvalue and singular value of the matrix $\mathbf{X}$, respectively.

2 Orthogonal iteration and its applications

Given a tall, column orthonormal matrix $\mathbf{Q}_0 \in \mathbb{C}^{N\times r}$, the so-called method of *orthogonal iteration* generates a sequence of matrices $\mathbf{Q}_m$, whose column span is assumed to approximate the span of the $r-$ dimensional dominant subspace of the matrix $\mathbf{W} \in \mathbb{C}^{N\times N}$, according to the following recurrence:

$$\begin{aligned}\mathbf{A}_m &= \mathbf{W}\mathbf{Q}_{m-1}, \quad m = 1, 2, \cdots \\ \mathbf{Q}_m\mathbf{R}_m &= \mathbf{A}_m, \end{aligned} \tag{1}$$

where $\mathbf{Q}_m$ and $\mathbf{R}_m$ denote the QR decomposition of the matrix $\mathbf{A}_m$. If $\mathbf{W}$ does not change over time, one can show that the subspace $\mathfrak{R}(\mathbf{Q}_m)$ converges to $\mathfrak{D}_r(\mathbf{W})$ at a rate proportional to $|\lambda_{r+1}(\mathbf{W})/\lambda_r(\mathbf{W})|^m$ [5]. Therefore, the usefulness of the method depends on this ratio, since it determines the rate of convergence. Note that when $r = 1$, (1) is just the well-known *power* method [12].

In several applications of interest in signal processing and communications, however, the assumption on the stationarity of $\mathbf{W}$ is usually not valid. Instead, a time-varying sequence $\{\mathbf{W}_m\}_{m=1}^{\infty}$ is often used, which is updated recursively as in e.g.:

$$\mathbf{W}_m = \alpha\mathbf{W}_{m-1} + (1-\alpha)\mathbf{z}_m\mathbf{z}_m^H, \tag{2}$$

where m now represents the discrete-time index, $\alpha \in [0, 1]$ represents the forgetting factor (typically close to 1), and $\mathbf{z}_m \in \mathbb{C}^{N\times 1}$ denotes an observation vector at time m, often modeled as an i.i.d. sequence of random vectors. In this case, we may sequentially track the $r-$ dimensional dominant subspace of the time-varying sequence $\{\mathbf{W}_m\}_{m=1}^{\infty}$ simply by replacing the stationary matrix $\mathbf{W}$ in (1) with $\mathbf{W}_m$, given the forgetting factor α is fairly close to one [13].

3 Convergence analysis

In order to motivate the method and to derive its convergence properties in non-stationary scenarios, we follow the analysis as well as the notation for the stationary case given in [5], and generalize the orthogonal iteration as follows.

To begin, let us consider k iterations of the recurrence in (1) and use induction to express it by

$$\underbrace{\mathbf{W}_k\mathbf{W}_{k-1}\cdots\mathbf{W}_1}_{:=\bar{\mathbf{W}}_k}\mathbf{Q}_0 = \mathbf{Q}_k\mathbf{R}_k\mathbf{R}_{k-1}\cdots\mathbf{R}_1, \tag{3}$$

where $\mathbf{W}_1, \cdots, \mathbf{W}_k$ represent matrices of interest over the first k time iterations, respectively. Assume that

$$\bar{\mathbf{U}}_k^H\bar{\mathbf{W}}_k\bar{\mathbf{U}}_k = \bar{\mathbf{\Lambda}}_k = \mathrm{diag}(\bar{\lambda}_{i,k}) \tag{4}$$

is an EVD of $\bar{\mathbf{W}}_k$ with $\bar{\lambda}_{1,k} \geq \bar{\lambda}_{2,k} \geq \cdots \geq \bar{\lambda}_{N,k} \geq 0$ and $\bar{\mathbf{U}}_k^H\bar{\mathbf{U}}_k = \mathbf{I}$. Partition $\bar{\mathbf{U}}_k$ and $\bar{\mathbf{\Lambda}}_k$ as follows:

$$\bar{\mathbf{U}}_k = \left[\bar{\mathbf{U}}_{1,k}\ \bar{\mathbf{U}}_{2,k}\right], \quad \bar{\mathbf{\Lambda}}_k = \begin{bmatrix} \bar{\mathbf{\Lambda}}_{1,k} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{\Lambda}}_{2,k} \end{bmatrix}, \tag{5}$$

where $\bar{\mathbf{U}}_{1,k} \in \mathbb{C}^{N\times r}$, $\bar{\mathbf{U}}_{2,k} \in \mathbb{C}^{N\times(N-r)}$, $\bar{\mathbf{\Lambda}}_{1,k} \in \mathbb{C}^{r\times r}$, and $\bar{\mathbf{\Lambda}}_{2,k} \in \mathbb{C}^{(N-r)\times(N-r)}$. Then we can arrive at

$$\begin{bmatrix} \bar{\mathbf{\Lambda}}_{1,k} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{\Lambda}}_{2,k} \end{bmatrix}\begin{bmatrix} \bar{\mathbf{U}}_{1,k}^H\mathbf{Q}_0 \\ \bar{\mathbf{U}}_{2,k}^H\mathbf{Q}_0 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{U}}_{1,k}^H\mathbf{Q}_k \\ \bar{\mathbf{U}}_{2,k}^H\mathbf{Q}_k \end{bmatrix}(\mathbf{R}_k\mathbf{R}_{k-1}\cdots\mathbf{R}_1). \tag{6}$$

If we let

$$\bar{\mathbf{U}}_k^H\mathbf{Q}_l = \begin{bmatrix} \bar{\mathbf{U}}_{1,k}^H\mathbf{Q}_l \\ \bar{\mathbf{U}}_{2,k}^H\mathbf{Q}_l \end{bmatrix} := \begin{bmatrix} \mathbf{V}_l \\ \mathbf{Y}_l \end{bmatrix}, \quad l = 0, 1, \cdots, k, \tag{7}$$

then

$$\mathbf{Y}_k = \bar{\mathbf{\Lambda}}_{2,k}\mathbf{Y}_0\mathbf{V}_0^{-1}\bar{\mathbf{\Lambda}}_{1,k}^{-1}\mathbf{V}_k \tag{8}$$

can be obtained by using (6) and (7). We can define the distance between the two subspaces $\mathfrak{D}_r(\bar{\mathbf{W}}_k)$ and $\mathfrak{R}(\mathbf{Q}_k)$ according to [5]

$$\mathrm{dist}\left(\mathfrak{D}_r(\bar{\mathbf{W}}_k), \mathfrak{R}(\mathbf{Q}_k)\right) = \|\bar{\mathbf{U}}_{2,k}^H\mathbf{Q}_k\|_2 = \|\mathbf{Y}_k\|_2. \tag{9}$$

By invoking (8), we can obtain

$$\|\mathbf{Y}_k\|_2 \leq \|\bar{\mathbf{\Lambda}}_{2,k}\|_2\|\mathbf{Y}_0\|_2\|\mathbf{V}_0^{-1}\|_2\|\bar{\mathbf{\Lambda}}_{1,k}^{-1}\|_2\|\mathbf{V}_k\|_2. \tag{10}$$

Let $\bar{\theta}_k \in [0, \pi/2]$ be defined to provide another measure of the closeness of the two subspaces $\mathfrak{D}_r(\bar{\mathbf{W}}_k)$ and $\mathfrak{R}(\mathbf{Q}_0)$, according to

$$\cos(\bar{\theta}_k) = \min_{\mathbf{u}\in\mathfrak{D}_r(\bar{\mathbf{W}}_k),\ \mathbf{v}\in\mathfrak{R}(\mathbf{Q}_0)} \frac{|\mathbf{u}^H\mathbf{v}|}{\|\mathbf{u}\|_2\|\mathbf{v}\|_2}. \tag{11}$$

Then $\cos(\bar{\theta}_k) = \sigma_r(\bar{\mathbf{U}}_{1,k}^H\mathbf{Q}_0) = \sigma_r(\mathbf{V}_0)$ and $\|\mathbf{Y}_0\|_2 = \sin(\bar{\theta}_k)$ [5]. Combining with (9)-(11), we can finally arrive at

$$\mathrm{dist}(\mathfrak{D}_r(\bar{\mathbf{W}}_k), \mathfrak{R}(\mathbf{Q}_k)) \leq \tan(\bar{\theta}_k)\left(\frac{\bar{\lambda}_{r+1,k}}{\bar{\lambda}_{r,k}}\right). \tag{12}$$

In the following, we categorize the non-stationary scenarios into three main cases and show how the result in (12) can be used to study the convergence properties in each case.

Case 1: Very small time variations

In general, we can express $\mathbf{W}_{k-i} = \mathbf{W}_k + \Delta\mathbf{W}_{k-i}$ for $i = 1, 2, \cdots, k-1$. Therefore, $\bar{\mathbf{W}}_k$ in (3) can be re-written as

$$\begin{aligned} \bar{\mathbf{W}}_k &= \mathbf{W}_k \mathbf{W}_{k-1} \cdots \mathbf{W}_1 \\ &= \mathbf{W}_k(\mathbf{W}_k + \Delta\mathbf{W}_{k-1}) \cdots (\mathbf{W}_k + \Delta\mathbf{W}_1) \\ &= (\mathbf{W}_k)^k + \Delta\bar{\mathbf{W}}_k, \end{aligned} \tag{13}$$

where $\Delta\bar{\mathbf{W}}_k = \bar{\mathbf{W}}_k - (\mathbf{W}_k)^k$ comprises products of $\mathbf{W}_k$ and $\Delta\mathbf{W}_{k-i}$ of different powers, $i = 1, 2, \cdots, k-1$. Let us further assume that

$$\mathbf{U}_k^H \mathbf{W}_k \mathbf{U}_k = \mathbf{\Lambda}_k = \mathrm{diag}(\lambda_{i,k}) \tag{14}$$

with $\lambda_{1,k} \geq \lambda_{2,k} \geq \cdots \geq \lambda_{N,k} \geq 0$. If $\Delta\bar{\mathbf{W}}_k \to \mathbf{0}$, then $\bar{\mathbf{W}}_k$ can be approximated by $(\mathbf{W}_k)^k$ alone. Hence, we can rewrite (12) as

$$\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k), \mathfrak{R}(\mathbf{Q}_k)) \leq \tan(\bar{\theta}_k) \left(\frac{\lambda_{r+1,k}}{\lambda_{r,k}} \right)^k. \tag{15}$$

Note that $\tan(\bar{\theta}_k)$ can be seen as a constant in this case due to the small time variations in $\mathbf{W}_i$'s. We may interpret the above result as follows: Given *very* small variations of $\mathbf{W}_i$ for $i = 1, 2, \cdots, k$, the distance between $\mathfrak{D}_r(\mathbf{W}_k)$ and $\mathfrak{R}(\mathbf{Q}_k)$ converges to zero with a rate equal to $(\frac{\lambda_{r+1,k}}{\lambda_{r,k}})^k$ (assuming $\lambda_{r,k} > \lambda_{r+1,k}$), which is the well-known property of the subspace-tracking algorithms using orthogonal iteration.

Case 2: Moderate time variations

For moderate variations of $\mathbf{W}_i$ over $i = 1, 2, \cdots, k$, however, the above property generally does not hold anymore. We first notice from (12) that the orthogonal iteration actually converges towards to $\mathfrak{D}_r(\bar{\mathbf{W}}_k)$, which may be largely different from $\mathfrak{D}_r(\mathbf{W}_k)$ now. Apart from this, we also wonder how the convergence rate is affected. To answer this question, it is of interest to view the effect $\Delta\bar{\mathbf{W}}_k$ in (13) as a perturbation to the matrix $(\mathbf{W}_k)^k$. Therefore, the corresponding perturbation in the eigenvalues of $(\mathbf{W}_k)^k$ can be described according to

$$\left| \lambda_i(\bar{\mathbf{W}}_k) - \lambda_i((\mathbf{W}_k)^k) \right| \leq \|\Delta\bar{\mathbf{W}}_k\|_2, \tag{16}$$

assuming $(\mathbf{W}_k)^k$ is normal [14], which is the case in applications of general interests. Fig. 1(a) illustrates the boundaries of the perturbed eigenvalues $\lambda_{r+1}(\bar{\mathbf{W}}_k)$ and $\lambda_r(\bar{\mathbf{W}}_k)$, which are bounded by a circle of radius $\|\Delta\bar{\mathbf{W}}_k\|_2$, with centers located at $\lambda_{r+1}((\mathbf{W}_k)^k)$ and $\lambda_r((\mathbf{W}_k)^k)$, respectively. Therefore, the ratio $(\frac{\bar{\lambda}_{r+1,k}}{\bar{\lambda}_{r,k}})$ in (12) governing the convergence rate is now bounded by

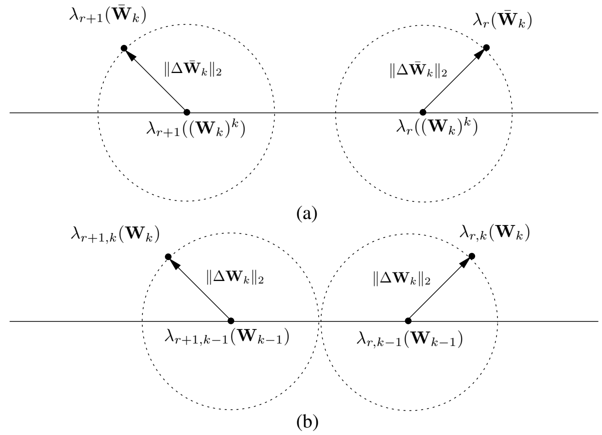

Figure 1. (a)Perturbation of the eigenvalues $\lambda_{r\bullet\,\bullet}((\mathbf{W}_k)^k)$ and $\lambda_r((\mathbf{W}_k)^k)$ due to $\|\Delta\bar{\mathbf{W}}_k\|_{\bullet}$. (b)Perturbation of the eigenvalues $\lambda_{r\bullet\,\bullet,k-\bullet}((\mathbf{W}_{k-\bullet}))$ and $\lambda_{r,k-\bullet}((\mathbf{W}_k))$ due to $\|\Delta\bar{\mathbf{W}}_k\|_{\bullet}$.

$$\frac{\lambda_{r+1}((\mathbf{W}_k)^k) - \delta}{\lambda_r((\mathbf{W}_k)^k) + \delta} \leq \left(\frac{\bar{\lambda}_{r+1,k}}{\bar{\lambda}_{r,k}} \right) \leq \frac{\lambda_{r+1}((\mathbf{W}_k)^k) + \delta}{\lambda_r((\mathbf{W}_k)^k) - \delta}, \tag{17}$$

where $\delta := \|\Delta\bar{\mathbf{W}}_k\|_2 \geq 0$. This implies that the convergence rate may be slightly increased or decreased, depending on the specific nature of the perturbation source $\Delta\bar{\mathbf{W}}_k$.

Case 3: Large time variations

For large variations of $\mathbf{W}_i$ over $i = 1, 2, \cdots, k$, we can generally assume that $\mathfrak{D}_r(\bar{\mathbf{W}}_k)$ can be significantly different from $\mathfrak{D}_r(\mathbf{W}_k)$, potentially making the subspace tracking ineffectual. It is therefore natural to ask, to what extent can we still *track* the subspace by orthogonal iteration, given the matrices $\mathbf{W}_i$ are rapidly changing. In other words, we are curious to know what is the maximum allowable time-variation of $\mathbf{W}_i$.

One possible way is to restrain the variation from $\mathbf{W}_{k-1}$ to $\mathbf{W}_k$ to at most half the distance between the rth and $(r+1)$th eigenvalues of $\mathbf{W}_{k-1}$, i.e.,

$$\|\Delta\mathbf{W}_k\|_2 < \frac{1}{2} \left| \lambda_{r,k-1}(\mathbf{W}_{k-1}) - \lambda_{r+1,k-1}(\mathbf{W}_{k-1}) \right|. \tag{18}$$

Focusing on the kth iteration alone, i.e. $\mathbf{W}_k \mathbf{Q}_{k-1} = \mathbf{Q}_k \mathbf{R}_k$, we can re-state the problem from the viewpoint of initial condition $\mathbf{Q}_{k-1}$ with one-step iteration. On the basis of earlier discussions, we know that

$$\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k), \mathfrak{R}(\mathbf{Q}_k)) \leq \tan(\theta_{k-1}) \left(\frac{\lambda_{r+1,k}(\mathbf{W}_k)}{\lambda_{r,k}(\mathbf{W}_k)} \right),$$

where $\theta_{k-1} \in [0, \pi/2]$ is defined according to

$$\cos(\theta_{k-1}) = \min_{\mathbf{u} \in \mathfrak{D}_r(\mathbf{W}_k),\ \mathbf{v} \in \Re(\mathbf{Q}_{k-1})} \frac{|\mathbf{u}^H \mathbf{v}|}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2}. \quad (19)$$

We can then clearly see that the distance between the $(r+1)$th and the rth eigenvalue of the matrix $\mathbf{W}_k$ should be maximized in order to minimizing the ratio $(\frac{\lambda_{r+1,k}(\mathbf{W}_k)}{\lambda_{r,k}(\mathbf{W}_k)})$, implying that the boundaries of the perturbed eigenvalue as illustrated in Fig. 1(b) should not be touching each other.

Summary

1. *Very small time variations*: orthogonal iteration converges toward to $\mathfrak{D}_r(\bar{\mathbf{W}}_k)$ at the rate according to (15).
2. *Moderate time variations*: for moderate time variations, the convergence rate may be increased or decreased according to (17).
3. *Large time variations*: to ensure effective tracking, the rate of change in the underlying subspace cannot exceeds the fundamental limit provided by (18).

4 Numerical experiments

In order to support the above claims, we provide numerical results as follows. We start by constructing a fixed Hermitian matrix $\mathbf{W} \in \mathbb{C}^{16\times16}$. We first show that when a noisy sample $\mathbf{W}'$ is used instead of $\mathbf{W}$ in the orthogonal iteration, the algorithm may converge to a subspace that is very different from $\mathfrak{D}_r(\mathbf{W})$. In this experiment, $\mathbf{W}'$ is modeled as $\mathbf{W} + \Delta\mathbf{W}$ and 200 runs in total are considered. At each run, each entry of $\Delta\mathbf{W}$ is a realization of an i.i.d. Gaussian r.v. with zero mean and variance σ^2. The experimental results for $r = 2$ and $\sigma^2 = 0, 10^{-3}, 10^{-2}, 10^{-1}$ are shown in Fig. 2, where the distance between $\mathfrak{D}_r(\mathbf{W})$ and $\Re(\mathbf{Q}_k)$, i.e.,

$$d := \mathrm{dist}(\mathfrak{D}_r(\mathbf{W}), \Re(\mathbf{Q}_k)), \quad (20)$$

is plotted versus the iteration index k. As we can observe in the steady-state, the distance between $\mathfrak{D}_r(\mathbf{W})$ and $\Re(\mathbf{Q}_k)$ increases as σ^2 is increased, simply because the orthogonal iteration converges to $\mathfrak{D}_r(\mathbf{W}')$ instead of $\mathfrak{D}_r(\mathbf{W})$. In the context of subspace tracking a wireless channel, this situation occurs when an *estimated* correlation matrix is actually employed for the algorithm. The estimation errors can be due mainly to: insufficient time samples for the correlation matrix averaging, fast time-varying nature of the wireless channel, improper coefficients from the exponential or rectangular windowing, or even a combination of above. In such cases, we are inevitably falling into the above situation.

Next, we want to show that the convergence rate in Case 2 might be slightly increasing or decreasing as described in (17). Considering several realizations of $\Delta\mathbf{W}$ with $\sigma^2 = 10^{-2}$, we compare the convergence curves with the ideal case (i.e., $\sigma^2 = 0$). With such a perturbation on the matrix $\mathbf{W}$, we can observe the convergence rates ranging from below to above that of the ideal case, as clearly shown in Fig. 3. Note that we also show the curve $(\lambda_{r+1}(\mathbf{W})/\lambda_r(\mathbf{W}))^k$ in the logarithmic scale for reference.

Finally, let us verify (18) by introducing another fixed Hermitian matrix $\Delta\mathbf{W}$. In this experiment, $\mathbf{W}' = \mathbf{W} + \beta\,\Delta\mathbf{W}$, where $\beta \in \mathbb{R}^+$ and $\|\Delta\mathbf{W}\|_2 = \|\mathbf{W}\|_2$. Given $\lambda_r(\mathbf{W}) = 7.16$, $\lambda_{r+1}(\mathbf{W}) = 6.61$ and $\|\Delta\mathbf{W}\|_2 = 9.67$, we need to have

$$\beta \le \frac{|\lambda_r(\mathbf{W}) - \lambda_{r+1}(\mathbf{W})|}{2\|\Delta\mathbf{W}\|_2} \approx 0.028. \quad (21)$$

to restrain the variations. We consider a sudden change of $\mathbf{W}$ to $\mathbf{W}'$ at the 50th iteration for various β's. As can be observed in Fig. 4, we can see that for cases with $\beta > 0.028$, the sudden change of $\mathbf{W}$ to $\mathbf{W}'$ at the 50th iteration substantially enlarges the distance between $\mathfrak{D}_r(\mathbf{W}')$ and $\Re(\mathbf{Q}_k)$.

For practical concerns, we also consider the popular time-varying model as mentioned in (2), with $\mathbf{z}_m \in \mathbb{C}^{16\times1}$ modeled by

$$\mathbf{z}_m = \mathbf{H}\mathbf{x}_m + \mathbf{n}_m, \quad (22)$$

where $\mathbf{H} \in \mathbb{C}^{16\times16}$ is a fixed channel matrix with $\mathrm{rank}(\mathbf{H}) = 2$, $\mathbf{x}_m$ is an i.i.d. random vector from a QAM constellation, i.e., with entries randomly selected from $(1/\sqrt{2})(\pm1 \pm j)$ with equal probability, and $\mathbf{n}_m$ is an i.i.d. Gaussian random vector with zero mean and variance σ_n^2. We choose $\mathbf{W}_0 = (1/500)\sum_{j=1}^{500} \mathbf{z}_j\mathbf{z}_j^H$ as our initial condition. Fig. 5 shows the probability $p := \mathrm{prob}\,(|\lambda_{r,k-1}(\mathbf{W}_{k-1}) - \lambda_{r+1,k-1}(\mathbf{W}_{k-1})| \le 2\|\Delta\mathbf{W}_k\|_2)$ versus the forgetting factor α for some SNR's, where SNR $:= 10\log_{10}(1/2\sigma_n^2)$. From the curve of SNR = 0dB in the figure, we can observe that the probability $p = 0.5$, 0.15, and 0 when $\alpha = 0.93$, 0.95, and 0.98, respectively. Indeed, when α is close to one, the rate of change in $\mathbf{W}_m$ in (2) is very small, and therefore the probability of $\|\Delta\mathbf{W}_k\|_2$ exceeding the limit (18) is close to 0. To see how the forgetting α affects the tracking process, Fig. 6-8 give some realizations of d versus the number of iterations, when the above mentioned α's are considered. From these figures, we can conclude that orthogonal iteration can only achieve satisfactory performance when the probability p is fairly small. Hence, (18) can serve as a fundamental limitation to determine whether or not orthogonal iteration can be applied in a rapidly time-varying scenario.

5 Conclusion

In this paper, we extended the convergence analysis of orthogonal iteration from stationary cases to non-stationary ones. In particular, we investigated certain properties of orthogonal iteration when it is applied to subspace tracking of practical wireless time-varying channels. In the context of blind subspace tracking problems, we can conclude that the performance of blind channel estimation using orthogonal iteration is mainly determined by whether we can obtain a good estimate of the *time-varying* correlation matrix. In the case of moderate time variations, we showed that the rate of convergence may be increased or decreased, depending on the nature of the perturbation source. We also discussed a fundamental limitation on the use of orthogonal iteration over rapidly time-varying wireless channels.

References

[1] X. G. Doukopoulos and G. V. Moustakides, "Blind adaptive channel estimation in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1716–1725, July 2006.

[2] B. Champagne, "Adaptive eigendecomposition of data covariance matrices based on first-order perturbations," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2758–2770, Oct. 1994.

[3] X. G. Doukopoulos and G. V. Moustakides, "Fast and stable subspace tracking," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1452–1465, Apr. 2008.

[4] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

[5] G. H. Golub and C. F. V. Loan, *Matrix Computation (third edition)*. The Johns Hopkins University Press, 1996.

[6] Y. Hua, Y. Xiang, T. Chen, K. Abed-Meraim, and Y. Miao, "A new look at the power method for fast subspace tracking," *Digit. Signal Process.*, vol. 9, no. 4, pp. 297–314, 1999.

[7] P. Strobach, "Low-rank adaptive filters," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 2932–2947, Dec. 1996.

[8] K. Abed-Meraim, A. Chkeif, and Y. Hua, "Fast orthonormal PAST algorithm," *IEEE Signal Process. Lett.*, vol. 7, no. 3, pp. 60–62, Mar. 2000.

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

[10] R. Badeau, B. David, and G. Richard, "Fast approximated power iteration subspace tracking," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2931–2941, Aug. 2005.

[11] D.-Z. Feng and W. X. Zheng, "Fast approximate inverse power iteration algorithm for adaptive total least-squares FIR filtering," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 4032–4039, Oct. 2006.

[12] D. S. Watkins, *Fundamentals of Matrix Computations*. New York: John Wiley & Sons, Inc., 2002.

[13] N. L. Owsley, "Adaptive data orthogonalization," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process.*, vol. 3, Apr. 1978, pp. 109–112.

[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

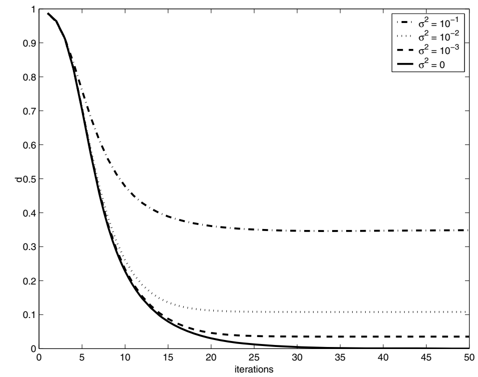

Figure 2. $\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}), \mathfrak{R}(\mathbf{Q}_k))$ **versus the number of iterations for various** σ^2**'s.**

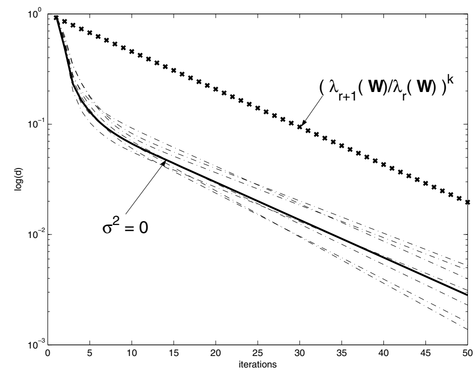

Figure 3. $\log(\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k'), \mathfrak{R}(\mathbf{Q}_k)))$ **versus the number of iterations for** σ^2 **=** 0 **(solid line) and** 10^{-2} **(dash-dot lines).**

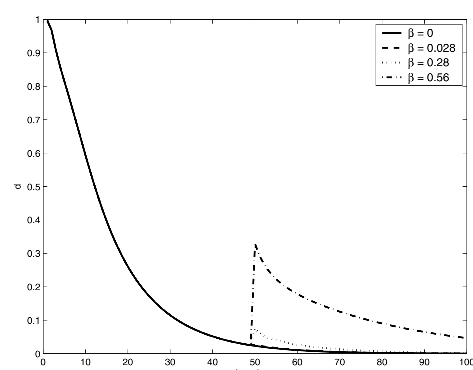

Figure 4. A sudden change of $\mathbf{W}$ **to** $\mathbf{W}'$ **at the** 50**th iteration.**

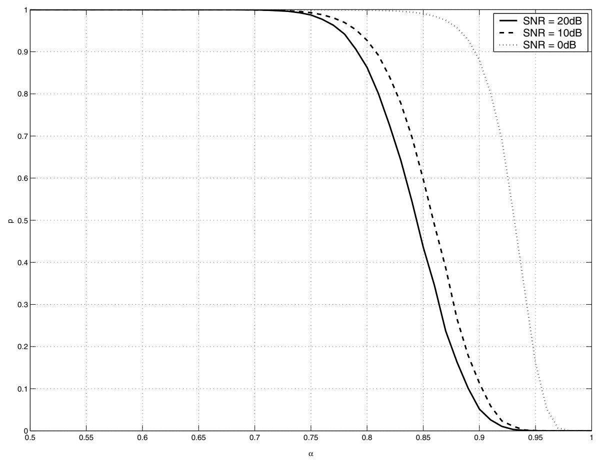

Figure 5. $\mathrm{prob}\,(|\lambda_{r,k-1}(\mathbf{W}_{k-1}) - \lambda_{r+1,k-1}(\mathbf{W}_{k-1})| \leq 2\|\Delta \mathbf{W}_k\|_2)$ **versus the forgetting factor α in the time-varying model.**

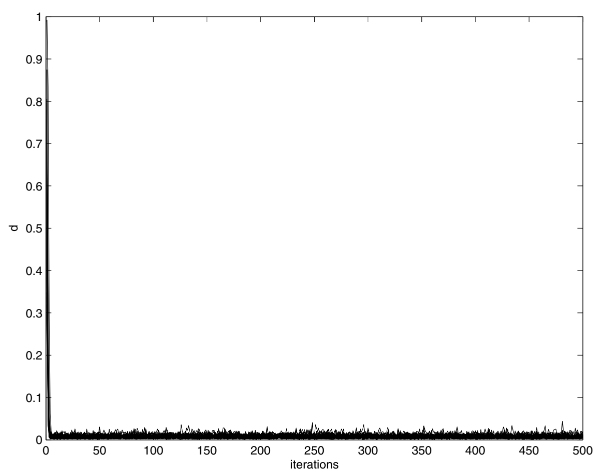

Figure 6. $\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k), \Re(\mathbf{Q}_k))$ **versus the number of iterations in the time-varying model with $\alpha = 0.98$ when SNR = 0dB.**

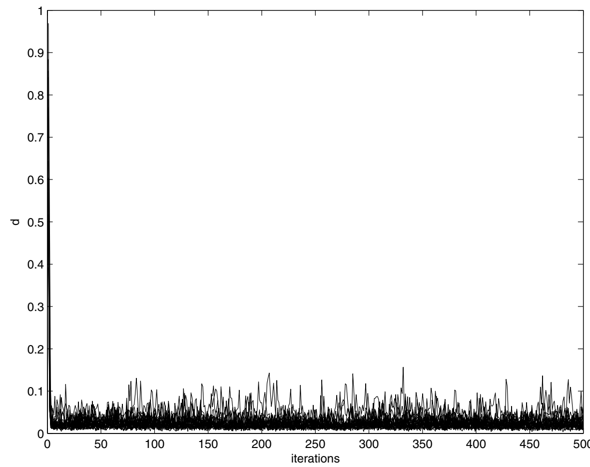

Figure 7. $\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k), \Re(\mathbf{Q}_k))$ **versus the number of iterations in the time-varying model with $\alpha = 0.95$ when SNR = 0dB.**

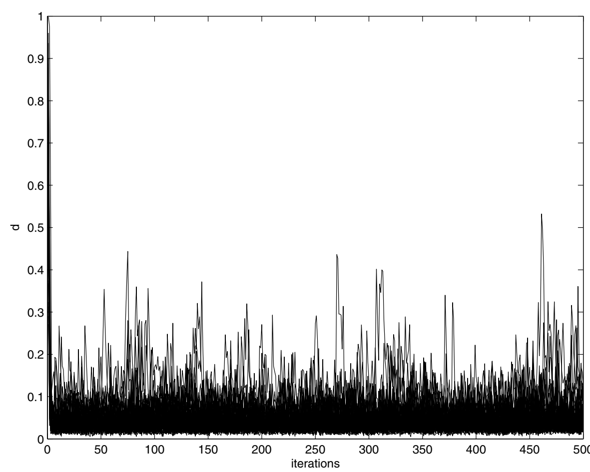

Figure 8. $\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k), \Re(\mathbf{Q}_k))$ **versus the number of iterations in the time-varying model with $\alpha = 0.93$ when SNR = 0dB.**