Title

On Convergence of the NIC Algorithm for Subspace Computation

Journal
IEEE Transactions on Signal Processing, 52(4)

ISSN
1053-587X

Authors
Hua, Yingbo
Chen, Tianping

Publication Date
2004-04-01

Peer reviewed

Correspondence

On Convergence of the NIC Algorithm for Subspace Computation

Yingbo Hua and Tianping Chen

***Abstract*—The "novel information criterion" (NIC) algorithm was developed by Miao and Hua in 1998 for fast adaptive computation of the principal subspace of a vector sequence. The NIC algorithm is as efficient computationally as the PAST method, which was devised by Yang in 1995, and also has an attractive orthonormal property. Although all available evidence suggests that the NIC algorithm converges to the desired solution for any fixed leakage factor between zero and one, a complete proof (or disproof) has not been found, except for an arbitrarily small leakage factor. This paper presents this long-standing open problem with a discussion of what is known so far. The results shown in this paper provide a new insight into the orthonormal property of the NIC algorithm at convergence.**

I. Introduction

Subspace computation is a fundamental tool for data compression, feature extraction, parameter estimation, model detection, and multiuser communications (e.g., see [4], [5], [9]–[11]). A key objective of subspace computation is to compute the principal subspace spanned by a sequence of vectors. Namely, given a $n \times m$ complex matrix $\mathbf{Y} = [\mathbf{y}(1)\ \mathbf{y}(2)\ \cdots\ \mathbf{y}(m)]$, one needs to compute a p-column basis matrix $\mathbf{W}$ such that the range of $\mathbf{W}$ is the rank-p principal subspace of $\mathbf{Y}$. A systematic treatment of a class of power-based algorithms is available in [2]. These algorithms are recursive and globally convergent (under a weak condition) to a desired principal subspace. They are well suited for adaptive subspace computations (or subspace tracking). As a very brief description, the power-based algorithms update the estimate of a principal subspace by multiplying the previous estimate by some forms of the original data. Besides the power-based algorithms, there are nonpower-based algorithms as well. Examples of nonpower-based algorithms are available in [8] and the references therein. Among the power-based algorithms is the "novel information criterion" (NIC) algorithm [1]. The NIC algorithm is a generalization of an earlier algorithm called "projection approximation for subspace tracking" (PAST) [3]. Although fast in convergence to a desired principal subspace, the PAST algorithm does not generally yield an orthonormal basis matrix. Simulations have shown that the NIC algorithm not only converges as fast as the PAST algorithm in subspace tracking but also always yields an orthonormal basis matrix at convergence [2].

In this correspondence, we first review the NIC algorithm and then present a long-standing conjecture that the NIC algorithm (with a fixed leakage factor between zero and one) converges to an orthonormal basis matrix of the desired principal subspace. This conjecture has been observed in all examples known today. We will also present a new understanding of this conjecture, which explains the orthonormal property of the NIC method at convergence.

Manuscript received June 28, 2002; revised June 17, 2003. This work was supported in part by the Australian Research Council and the U.S. National Science Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Paul D. Fiore.

Y. Hua is with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA (e-mail: yhua@ee.ucr.edu).

T. Chen is with the Department of Mathematics, Fudan University, Shanghai, China (e-mail: tchen@fudan.edu.cn).

Digital Object Identifier 10.1109/TSP.2004.823488

II. NIC Algorithm and an Open Problem

The NIC algorithm was based on the maximization of the following NIC criterion [1]:

$$\begin{aligned} J_{\text{NIC}}(\mathbf{W}) &= \frac{1}{2} tr\left(\log(\mathbf{W}^H\mathbf{C}\mathbf{W})\right) - \frac{1}{2} tr(\mathbf{W}^H\mathbf{W}) \\ &= \frac{1}{2}\log\left(\det(\mathbf{W}^H\mathbf{C}\mathbf{W})\right) - \frac{1}{2} tr(\mathbf{W}^H\mathbf{W}) \end{aligned} \tag{1a}$$

where $\mathbf{C} = \mathbf{Y}\mathbf{Y}^H \in C^{n\times n}$, and the superscript ^{H} denotes the complex conjugate transpose. The identity $tr(\log(\mathbf{A})) = \log(\det(\mathbf{A}))$ holds for any positive definite matrix $\mathbf{A}$. This follows from the definition $\log(\mathbf{A}) \triangleq \mathbf{E}\text{diag}(\log\lambda_1,\ \log\lambda_2,\ \cdots,\ \log\lambda_n)\mathbf{E}^H$, where $\mathbf{A} = \mathbf{E}\text{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)\mathbf{E}^H$ is the eigenvalue decomposition of $\mathbf{A}$. In [1], only real matrices are considered. For complex matrices, we need a slightly different treatment, as described below. Following the gradient ascent principle to maximize the NIC criterion, we have the following algorithm:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \alpha\frac{\partial}{\partial\mathbf{W}(k)} J_{\text{NIC}}(\mathbf{W}(k)) \tag{1b}$$

where α is a step size, and $\mathbf{W}(k) \in C^{n\times p}$ $(n > p)$ is the estimate of the basis matrix of the principal subspace spanned by the columns of $\mathbf{Y}$ (or equivalently, by the columns of $\mathbf{C}$) after the kth iteration. The complex matrix gradient in (1b) is defined as[1]

$$\frac{\partial}{\partial\mathbf{W}} J_{\text{NIC}}(\mathbf{W}) = \frac{\partial}{\partial\text{Re}(\mathbf{W})} J_{\text{NIC}}(\mathbf{W}) + j\frac{\partial}{\partial\text{Im}(\mathbf{W})} J_{\text{NIC}}(\mathbf{W}) \tag{1c}$$

where $j = \sqrt{-1}$. It can be shown (after a straightforward but slightly tedious procedure) that

$$\frac{\partial}{\partial\text{Re}(\mathbf{W})} J_{\text{NIC}}(\mathbf{W}) = \frac{1}{2}\left\{\left(\mathbf{C}\mathbf{W}(\mathbf{W}^H\mathbf{C}\mathbf{W})^{-1}\right)^* + \mathbf{C}\mathbf{W}(\mathbf{W}^H\mathbf{C}\mathbf{W})^{-1} - \mathbf{W}^* - \mathbf{W}\right\}$$

and

$$\frac{\partial}{\partial\text{Im}(\mathbf{W})} J_{\text{NIC}}(\mathbf{W}) = \frac{1}{2}\left\{j\left(\mathbf{C}\mathbf{W}(\mathbf{W}^H\mathbf{C}\mathbf{W})^{-1}\right)^* - j\mathbf{C}\mathbf{W}(\mathbf{W}^H\mathbf{C}\mathbf{W})^{-1} - j\mathbf{W}^* + j\mathbf{W}\right\}$$

where * denotes complex conjugation. Note that both of the above expressions are real valued (as they should be). Using these two expressions in (1c) yields

$$\frac{\partial}{\partial\mathbf{W}} J_{\text{NIC}}(\mathbf{W}) = \mathbf{C}\mathbf{W}(\mathbf{W}^H\mathbf{C}\mathbf{W})^{-1} - \mathbf{W}. \tag{1d}$$

Applying (1d) to (1b), we have the NIC algorithm in its batch form:

$$\mathbf{W}(k+1) = (1-\alpha)\mathbf{W}(k) + \alpha\mathbf{C}\mathbf{W}(k)\left(\mathbf{W}(k)^H\mathbf{C}\mathbf{W}(k)\right)^{-1}. \tag{2}$$

The range of $\mathbf{W}(k)$ is used as an estimate of the principal subspace of the range of $\mathbf{Y}$ at iteration k. $\mathbf{W}(k)$ is also referred to as the weight matrix in a context of linear neural networks [10]. When $\alpha = 1$, (1) becomes the batch form of the PAST algorithm [3]. When $0 < \alpha < 1$, the old estimate $\mathbf{W}(k)$ "leaks" through the term $(1-\alpha)\mathbf{W}(k)$ to yield the new estimate $\mathbf{W}(k+1)$. Hence, α is also referred to as a leakage

[1]This is a natural definition of complex matrix gradient although complex (vector or matrix) gradients are not commonly addressed in text books. For real matrix gradients, see [4, p. 275].

factor. An adaptive version of (2) follows if $\mathbf{C}$ is replaced by the recursive equation $\mathbf{C}(k+1) = \mathbf{C}(k) + \mathbf{y}(k+1)\mathbf{y}^H(k+1)$, where $\mathbf{y}(k+1)$ is the new data vector at time $k+1$. One can also add some standard forgetting factor (e.g., see [2] and [7, p. 354]) to the above recursion of $\mathbf{C}(k)$ to achieve faster adaptation. Following the principle of "projection approximation" [3] (i.e., replacing $\mathbf{C}(k)\mathbf{W}(k)$ by $\mathbf{C}(k)\mathbf{W}(k-1)$ in a propagation term), an adaptive version of (2) can be implemented such that its computational order at each iteration is linearly proportional to n. This linear complexity is the key computational advantage of both the NIC and PAST algorithms in comparison to many other methods [2]. To conduct a convergence analysis of (2), however, one needs to assume a constant $\mathbf{C}$.[2]

The open problem here is either a proof or a disproof of the following conjecture.

Conjecture: If a) the pth and $(p+1)$th largest eigenvalues of $\mathbf{C}$ are distinct, and b) $\mathbf{W}(0)$ has a nonsingular (full rank[3]) projection onto the span S of the first p eigenvectors (principal subspace) of $\mathbf{C}$, then the limit $\mathbf{W}(\infty)$ of the recursive (1) with $0 < \alpha < 1$ is an orthonormal matrix (i.e., $\mathbf{W}(\infty)^H\mathbf{W}(\infty) = \mathbf{I}$) with its column span equal to S. Note that the conditions a) and b) are weak and almost always satisfied in practice.

This conjecture is proven to be true when α is arbitrarily small [1], where the convergence is established via a Lyapunov function. When α is arbitrarily small, there is a corresponding differential NIC flow equation, whose convergence is further established in a more recent paper [6]. The conjecture, however, does not cover the case when $\alpha = 1$. For this case, $\mathbf{W}(\infty)$ spans S but is not necessarily an orthonormal matrix [2]. A simple examination of (2) confirms the intuition that the dynamics of (2) is only slowed down when α is decreased from 1. However, the conjecture suggests that by choosing a fixed α satisfying $0 < \alpha < 1$, $\mathbf{W}(\infty)$ not only spans S but is also an orthogonal matrix. The property $range(\mathbf{W}(\infty)) = S$ seems obvious since it is proven to be true for both arbitrarily small α and for $\alpha = 1$. However, this property is not yet proven today. On the other hand, the property $\mathbf{W}(\infty) = orthonormal$ for any α satisfying $0 < \alpha < 1$ is not at all obvious, especially considering the fact that $\mathbf{W}(\infty)$ is not necessarily orthonormal if $\alpha = 1$. We show next a new understanding of the conjecture with a focus on the property $\mathbf{W}(\infty) = orthonormal$.

III. New Understanding

Equation (2) can be transformed into the following:

$$\mathbf{X}(k+1) = (1-\alpha)\mathbf{X}(k) + \alpha\Lambda\mathbf{X}(k)\left(\mathbf{X}(k)^H\Lambda\mathbf{X}(k)\right)^{-1} \tag{2a}$$

where $\mathbf{X}(k) = \mathbf{U}^H\mathbf{W}(k)$, $\mathbf{U} \in C^{n\times n}$ is the eigenvector matrix of $\mathbf{C}$, and $\Lambda \in R^{n\times n}$ is the (corresponding) diagonal matrix of the eigenvalues that are in descending order. Let $\mathbf{X}(k) = \begin{bmatrix}\mathbf{X}_1(k)\\ \mathbf{X}_2(k)\end{bmatrix}$, where $\mathbf{X}_1(k) \in C^{p\times p}$ and $\mathbf{X}_2(k) \in C^{(n-p)\times p}$. It follows from $\mathbf{X}(k) = \mathbf{U}^H\mathbf{W}(k)$ that $\mathbf{W}(\infty)$ is orthonormal and spans S if and only if $\mathbf{X}_2(\infty) = 0$ and $\mathbf{X}_1(\infty)$ is unitary. Condition b) in the conjecture simply means that $\mathbf{X}_1(0)$ is nonsingular.

We now show two lemmas. The first lemma shows that the weight matrix in the NIC algorithm does not diverge to infinity nor degenerate into a reduced-rank matrix. The second lemma shows the orthonormal property of the NIC method at convergence.

Lemma 1: If $\mathbf{X}_1(0)$ is nonsingular and $0 < \alpha < 1$, then for all $k > 0$, $\mathbf{X}(k)$ has a full column rank and a finite norm.

[2]It is also a convention of convergence analyzes in adaptive filter theory [7].

[3]This will be explained later.

Proof: Let the ith singular value of a matrix $\mathbf{A}$ be denoted by $\sigma_i\{\mathbf{A}\}$. Assuming that $\mathbf{X}(k)$ has a full column rank for a fixed k, it follows from (2a) that

$$\begin{aligned}
&\mathbf{X}(k+1)^H\mathbf{X}(k+1)\\
&\quad = (1-\alpha)^2\mathbf{X}(k)^H\mathbf{X}(k) + 2\alpha(1-\alpha)\mathbf{I}\\
&\qquad + \alpha^2\left(\mathbf{X}(k)^H\Lambda\mathbf{X}(k)\right)^{-1}\\
&\qquad \times \mathbf{X}(k)^H\Lambda^2\mathbf{X}(k)\left(\mathbf{X}(k)^H\Lambda\mathbf{X}(k)\right)^{-1}
\end{aligned}$$

which implies that $\sigma_p\{\mathbf{X}(k+1)^H\mathbf{X}(k+1)\} > 2\alpha(1-\alpha)$. Therefore, under condition b), $\sigma_p\{\mathbf{X}(k)\} > c_2 > 0$ for all $k > 0$, which means that $\mathbf{X}(k)$ has a full column rank. Now, from (2a), we have

$$\begin{aligned}
&\sigma_1\{\mathbf{X}(k+1)\}\\
&\quad \le (1-\alpha)\sigma_1\{\mathbf{X}(k)\}\\
&\qquad + \sigma_1\left\{\alpha\Lambda^{\frac{1}{2}}\right\}\sigma_1\left\{\Lambda^{\frac{1}{2}}\mathbf{X}(k)\left(\mathbf{X}(k)^H\Lambda\mathbf{X}(k)\right)^{-1}\right\}\\
&\quad = (1-\alpha)\sigma_1\{\mathbf{X}(k)\} + \sigma_1\left\{\alpha\Lambda^{\frac{1}{2}}\right\}\frac{1}{\sigma_p\left\{\Lambda^{\frac{1}{2}}\mathbf{X}(k)\right\}}.
\end{aligned}$$

Note that Λ is nonsingular, and hence, there is a constant $c_3 > 0$ such that $\sigma_p\{\Lambda^{1/2}\mathbf{X}(k)\} > c_3$ for all $k > 0$. Therefore, there is a constant $c_4 < \infty$ such that for all $k > 0$

$$\sigma_1\{\mathbf{X}(k+1)\} \le (1-\alpha)\sigma_1\{\mathbf{X}(k)\} + c_4.$$

This expression implies that for all $k > 0$

$$\sigma_1\{\mathbf{X}(k)\} \le (1-\alpha)^k\sigma_1\{\mathbf{X}(0)\} + c_4\sum_{l=0}^{k}(1-\alpha)^l < c_1 < \infty$$

i.e., $\mathbf{X}(k)$ has a finite norm. ∎

The next lemma provides an important statement, and the proof that follows is even more insightful.

Lemma 2: If $\mathbf{X}_1(0)$ is nonsingular, $0 < \alpha < 1$, and $\mathbf{X}_2(\infty) = 0$, then $\mathbf{X}_1(\infty)$ is unitary.

Proof: Let Σ be the $p\times p$ top left submatrix of Λ. Then, taking the top p rows of (2a), we have

$$\begin{aligned}
&\mathbf{X}_1(k+1)\\
&\quad = (1-\alpha)\mathbf{X}_1(k) + \alpha\Sigma\mathbf{X}_1(k)\left(\mathbf{X}_1(k)^H\Sigma\mathbf{X}_1(k)\right)^{-1} + \mathbf{E}(k)\\
&\quad = (1-\alpha)\mathbf{X}_1(k) + \alpha\mathbf{X}_1(k)^{-H} + \mathbf{E}(k)
\end{aligned}$$

where $\mathbf{E}(k)$ is a matrix of small norm for large k, and $\mathbf{E}(\infty) = 0$. This property of $\mathbf{E}(k)$ follows from $\mathbf{X}_2(\infty) = 0$. Note that the above equation requires $\mathbf{X}_1(k)$ to be nonsingular. This is guaranteed by Lemma 1 and the assumption that $\mathbf{X}_2(k)$ is arbitrarily small for large k. Substituting the singular value decomposition of $\mathbf{X}_1(k)$ into the above matrix equation leads to the following singular value equation:

$$d_i(k+1) = (1-\alpha)d_i(k) + \alpha\frac{1}{d_i(k)} + e_i(k) \tag{3}$$

where $1 \le i \le p$, $d_i(k)$ is the ith singular value of $\mathbf{X}_1(k)$, i.e., $d_i(k) = \sigma_i\{\mathbf{X}_1(k)\}$, $e_i(k)$ is a small perturbation for large k, and $e_i(\infty) = 0$. Note that $e_i(k)$ is not necessarily a singular value of $\mathbf{E}(k)$. The validity of the singular value (3) holds despite the fact that the left and/or right subspaces of $\mathbf{X}_1(k)$ are generally different (unless $\mathbf{E}(k) = 0$) from those of $\mathbf{X}_1(k+1)$. It follows from Lemma 1 and the assumption $\mathbf{X}_2(\infty) = 0$ that for large enough k and all i, $0 < c_2 < d_i(k) < c_1 < \infty$.

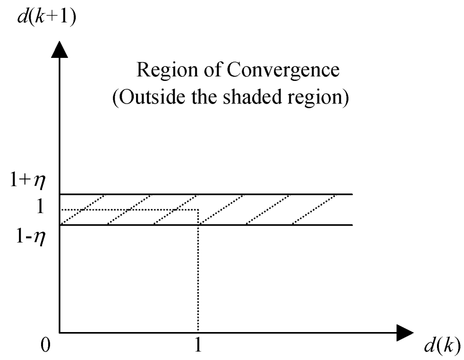

Fig. 1. Region of convergence where $[d(k),\ d(k+1)]$ is attracted toward [1, 1]. As $\max\{|e(k)|, |d(k)e(k)|, |(e(k)/d(k))|\} < \varepsilon$ becomes arbitrarily small, so does η. The index k denotes the iteration index of the NIC algorithm.

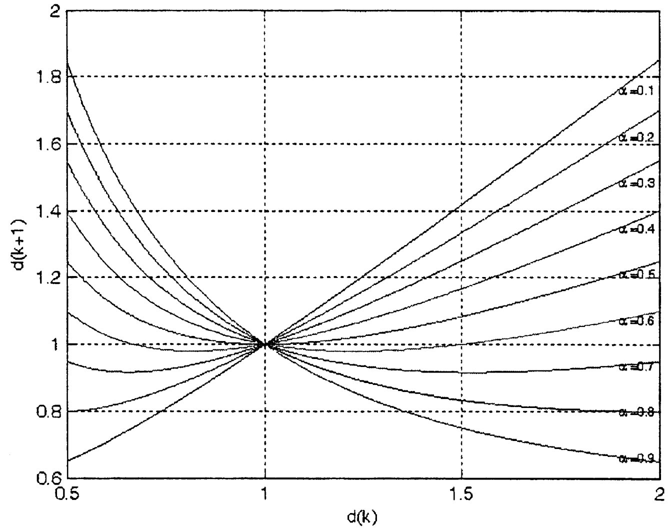

Fig. 2. Curves of $d(k+1)$ versus $d(k)$ for varied α according to (3) with $e_i(k) = 0$. Each point on these curves satisfies (9). The statement in footnote 6 is verified here.

All we need to show now is that $d_i(k)$ converges to one for all i. For convenience, we will drop the subscript $_i$ from (3) and write an alternative form of (3), which is easy to verify, as follows:

$$d(k+1) - 1 = ((1-\alpha)d(k) - \alpha)\,\frac{d(k)-1}{d(k)} + e(k). \tag{4}$$

We define a positive number ε and an integer k_0 such that for all $k > k_0, \max\{|e(k)|, |d(k)e(k)|, |(e(k)/d(k))|\} < \varepsilon$. It follows from $\mathbf{X}_2(\infty) = 0$ that for any arbitrarily small ε, there exists such k_0. Since (4) is a nonlinear (difference) equation, it is no surprise that we need to examine its convergence property in separate regions. In our treatment of (4), we use $d(k)$ and $d(k+1)$ as two coordinates, as shown in Fig. 1. We now analyze the convergence property of (4) in four separate cases of $d(k)$ and $d(k+1)$, which constitute all possible cases of interest.[4] We will show that for each case, the distance between $d(k+1)$ and 1 is upper bounded by an arbitrarily small number $O(\varepsilon)$ plus a down-scaled distance between $d(k)$ and 1. The distance between $d(k)$ and 1 is measured by $d(k) - 1$ if $d(k) > 1$ and by $(1/d(k)) - 1$ if $d(k) < 1$. A similar rule applies to $d(k+1)$.

Case 1) $d(k) > 1$, and $d(k+1) > 1 + \eta$. In this case, (4) implies that for any given η, there is a small enough (but finite) ε such that $(1-\alpha)d(k) - \alpha > 0$. We also know that $(1-\alpha)d(k) - \alpha < (1-\alpha)d(k)$. Then, (4) implies

$$d(k+1) - 1 < (1-\alpha)\,(d(k) - 1) + \varepsilon. \tag{5}$$

Case 2) $d(k) < 1$, and $d(k+1) > 1 + \eta$. In this case, (4) implies that for any given η, there is a small enough (but finite) ε such that $\alpha - (1-\alpha)d(k) > 0$. Then, (4) also implies

$$\begin{aligned} d(k+1) - 1 = (\alpha - (1-\alpha)d(k))&\left(\frac{1}{d(k)} - 1\right)\\ +e(k) < \alpha&\left(\frac{1}{d(k)} - 1\right) + \varepsilon. \end{aligned} \tag{6}$$

Case 3) $d(k) > 1$, and $d(k+1) < 1 - \eta$. In this case, (4) implies that for any given η, there is a small enough (but finite) ε such that $\alpha - (1-\alpha)d(k) > 0$. We can use (3) and (4) to write

$$\begin{aligned} \frac{1}{d(k+1)} - 1 &= \frac{1 - d(k+1)}{d(k+1)}\\ &= \frac{(\alpha - (1-\alpha)d(k))\,(d(k)-1) - d(k)e(k)}{(1-\alpha)d(k)^2 + \alpha + d(k)e(k)}\\ &= \frac{\alpha - (1-\alpha)d(k)}{(1-\alpha)d(k)^2 + \alpha}\,(d(k)-1) + O(\varepsilon) \end{aligned} \tag{7a}$$

where $O(\varepsilon)$ denotes a term of the order of ε. Furthermore, since $\alpha - (1-\alpha)d(k) > 0$, $0 < \alpha < 1$, and $d(k) > 1$, we have[5]

$$0 < \frac{\alpha - (1-\alpha)d(k)}{(1-\alpha)d(k)^2 + \alpha} < 2\alpha - 1 < \alpha.$$

Hence, (7a) implies

$$\frac{1}{d(k+1)} - 1 < \alpha\,(d(k) - 1) + O(\varepsilon). \tag{7b}$$

Case 4) $d(k) < 1$, and $d(k+1) < 1 - \eta$. In this case, (4) implies that for any given η, there is a small enough (but finite) ε such that[6] $(1-\alpha)d(k) - \alpha > 0$. We now use (3) and (4) to write

$$\begin{aligned} \frac{1}{d(k+1)} - 1 &= \frac{1 - d(k+1)}{d(k+1)}\\ &= \frac{((1-\alpha)d(k) - \alpha)\left(\frac{1}{d(k)} - 1\right) - e(k)}{(1-\alpha)d(k) + \alpha\frac{1}{d(k)} + e(k)}\\ &= \frac{\left((1-\alpha) - \frac{\alpha}{d(k)}\right)\left(\frac{1}{d(k)} - 1\right) - \frac{e(k)}{d(k)}}{(1-\alpha) + \alpha\frac{1}{d^2(k)} + \frac{e(k)}{d(k)}}\\ &= \frac{(1-\alpha) - \frac{\alpha}{d(k)}}{(1-\alpha) + \alpha\frac{1}{d^2(k)}}\left(\frac{1}{d(k)} - 1\right) + O(\varepsilon). \end{aligned} \tag{8a}$$

[4]Other cases such as $d(k) = 1$ or $1 + \eta \geq d(k+1) \geq 1 - \eta$ are of no value in establishing the convergence property as they are already confined in an arbitrarily small region around 1.

[5]The following inequality implies that $\alpha > 1/2$. This simply means that Case 3 is possible only if $\alpha > 1/2$.

[6]This inequality implies that $\alpha < 1/2$ and, hence, that Case 4 is possible only if $\alpha < 1/2$ (see Fig. 2).

Since $(1-\alpha)d(k)-\alpha > 0$ and $0 < c_2 < d(k) < 1$, we know

$$0 < \frac{(1-\alpha)-\frac{\alpha}{d(k)}}{(1-\alpha)+\alpha\frac{1}{d^2(k)}} < \frac{(1-\alpha)}{(1-\alpha)+\alpha\frac{1}{d^2(k)}} < 1-\alpha.$$

Then, (8a) implies

$$\frac{1}{d(k+1)} - 1 < (1-\alpha)\left(\frac{1}{d(k)}-1\right) + O(\varepsilon). \tag{8b}$$

We now combine the above results (5)–(8) for all four cases. We obtain that for $|d(k+1)-1| > \eta$ and some small enough (but finite) ε

$$\max\left(d(k+1)-1, \quad \frac{1}{d(k+1)}-1\right) < \max(\alpha, \quad 1-\alpha) \times \max\left(d(k)-1, \quad \frac{1}{d(k)}-1\right) + O(\varepsilon). \tag{9}$$

Note that as ε becomes arbitrarily small, so does η. Since $\max(\alpha, \ 1-\alpha) < 1$, (9), together with $e(\infty)=0$, implies that

$$\max\left(d(\infty)-1, \quad \frac{1}{d(\infty)}-1\right) = 0 \tag{10}$$

i.e., $d(\infty)=1$. ■

IV. Conclusions

We have reviewed the NIC algorithm for subspace computation and provided a complex data version of the algorithm. Despite its excellent performance as demonstrated by all existing theories and numerical experiments, the NIC algorithm still poses a long standing question about its convergence property. This question demands a complete proof (or disproof) of the conjecture that the NIC algorithm (under a weak condition) converges to an orthonormal basis matrix of the desired principal subspace. In this correspondence, we have provided a new understanding of this conjecture with regard to its orthonormal property at convergence. A full proof (or disproof) of the conjecture remains open.

Acknowledgment

The authors would like to thank the following people for their time and efforts in discussing the conjecture shown in this letter on separate occasions: Prof. X. Cao (Hong Kong University of Science and Technology), Prof. T. Luo (McMaster University), Prof. L. Qiu (Hong Kong University of Science and Technology), Prof. L. Xie (Nanyang Technological University), and Prof. I. Mareels (University of Melbourne). The reviewers and the Associate Editor P. Fiore have paid meticulous attention to every detail of this correspondence. Fig. 2 was provided by Fiore.

References

[1] Y. Miao and Y. Hua, "Fast subspace tracking and neural network learning by a novel information criterion," *IEEE Trans. Signal Processing*, vol. 46, pp. 1967–1979, July 1998.
[2] Y. Hua, Y. Xiang, T. Chen, K. Abed-Meraim, and Y. Miao, "A new look at the power method for fast subspace tracking," in *Digital Signal Processing*. New York: Academic, Oct. 1999, vol. 9, pp. 297–314.
[3] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, pp. 95–107, Jan. 1995.
[4] L. L. Scharf, *Statistical Signal Processing—Detection, Estimation, and Time Series Analysis*, First ed. Reading, MA: Addison-Wesley, 1991.
[5] W. Wang and H. V. Poor, "Blind multiuser detection: a subspace approach," *IEEE Trans. Inform. Theory*, vol. 44, pp. 677–690, Mar. 1998.
[6] W. Liu, W.-Y. Yan, V. Sreeram, and K. L. Tao, "Global convergence analysis for the NIC flow," *IEEE Trans. Signal Processing*, vol. 49, pp. 2422–2430, Oct. 2001.
[7] S. Haykin, *Adaptive Filter Theory*, Fourth ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
[8] E. C. Real, D. W. Tufts, and J. W. Cooley, "Two algorithms for fast approximate subspace tracking," *IEEE Trans. Signal Processing*, vol. 47, pp. 1936–1945, July 1999.
[9] Y. Yamashita and H. Ogawa, "Relative Karhunen-Loeve transform," *IEEE Trans. Signal Processing*, vol. 44, pp. 371–378, Feb. 1996.
[10] K. I. Diamantara and S. Y. Kung, "Multilayer neural networks for reduced-rank approximation," *IEEE Trans. Neural Networks*, vol. 5, pp. 684–697, Sept. 1994.
[11] D. Tufts and R. Kumaresan, "Frequency estimation of multiple sinusoids: making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, pp. 975–990, Mar. 1983.

Analysis of the Power Spectral Deviation of the General Transfer Function GSC

Sharon Gannot, David Burshtein, and Ehud Weinstein

***Abstract*—In recent work, we considered a microphone array located in a reverberated room, where general transfer functions (TFs) relate the source signal and the microphones, for enhancing a speech signal contaminated by interference. It was shown that it is sufficient to use the ratio between the different TFs rather than the TFs themselves in order to implement the suggested algorithm. An unbiased estimate of the TFs ratios was obtained by exploiting the nonstationarity of the speech signal.**

In this correspondence, we present an analysis of a distortion indicator, namely *power spectral density (PSD) deviation*, imposed on the desired signal by our newly suggested *transfer function generalized sidelobe canceller* (TF-GSC) algorithm. It is well known that for speech signals, PSD deviation between the reconstructed signal and the original one is the main contribution for speech quality degradation. As we are mainly dealing with speech signals, we analyze the *PSD deviation* rather than the regular waveform distortion. The resulting expression depends on the TFs involved, the noise field, and the quality of estimation of the TF's ratios. For the latter dependency, we provide an approximated analysis of estimation procedure that is based on the signal's nanstationarity and explore its dependency on the actual speech signal and on the signal-to-noise ratio (SNR) level. The theoretical expression is then used to establish empirical evaluation of the *PSD deviation* for several TFs of interest, various noise fields, and a wide range of SNR levels. It is shown that only a minor amount of *PSD deviation* is imposed on the beamformer output. The analysis presented in this correspondence is in good agreement with the actual performance presented in the former TF-GSC paper.

***Index Terms*—Beamforming, nonstationarity, speech enhancement.**

I. Introduction

Adaptive microphone arrays are widely used for speech enhancement. Most of the methods are based on the *generalized sidelobe canceller* (GSC) proposed by Griffiths and Jim [1]. Since this structure is usually based on the assumption that the different sensors receive a delayed version of the desired signal, we refer to it as the *delayed-GSC* (D-GSC). In more complex environments such as the reverberating

Manuscript received November 27, 2002; revised May 21, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Olivier Besson.
S. Gannot is with the School of Engineering, Bar-Ilan University, Ramat-Gan, Israel (e-mail: gannot@eng.biu.ac.il).
D. Burshtein and E. Weinstein are with the Department of Electrical Engineering—Systems, Tel-Aviv University, Tel-Aviv, Israel.
Digital Object Identifier 10.1109/TSP.2004.823487