A PROJECTION APPROXIMATION MINOR SUBSPACE TRACKING ALGORITHM

Scott C. Douglas and Xiaoan Sun

Department of Electrical Engineering
Southern Methodist University
Dallas, Texas 75275 USA

ABSTRACT

In this paper, we present an extension of the least-squares-based projection approximation subspace tracking (PAST) algorithm to minor subspace analysis (MSA) tasks. Our novel algorithm uses m identical Householder transformations to update the rows of the subspace matrix estimate at each time instant. Unlike other proposed methods for PAST-based MSA, ours has a complexity that scales linearly with the number of adaptive coefficients. Analyses and simulations verify the excellent numerical behavior of the proposed method in MSA tasks.

1. INTRODUCTION

Principal subspace analysis (PSA) is a classic array processing problem in which a set of n-dimensional signal vectors $\mathbf{x}(k) = [x_1(k) \;\; x_2(k) \;\; \cdots \;\; x_n(k)]^T$, $k \geq 1$ are used to calculate a series of $(m \times n)$ orthonormal matrices $\mathbf{W}(k)$, $m < n$, such that $\mathbf{u}(k)$, the reduced-dimensional reconstruction of $\mathbf{x}(k)$ given by

$$\mathbf{u}(k) = \mathbf{W}^T(k-1)\mathbf{y}(k) \tag{1}$$

$$\mathbf{y}(k) = \mathbf{W}(k-1)\mathbf{x}(k) \tag{2}$$

closely approximates $\mathbf{x}(k)$ in a least-squares or mean-square sense. PSA is a useful preprocessing step in numerous tasks, among them statistical pattern recognition, data visualization, and signal separation [1].

Numerous gradient and approximate least-squares algorithms have been developed for PSA [1]–[18]. Two of the more-popular methods are the projection approximation subspace tracking (PAST) and PAST with deflation (PASTd) algorithms proposed by Yang [11]. These two $\mathcal{O}(mn)$ algorithms are approximate least-squares methods whose asymptotic performances are independent of the eigenvalues of the input signal autocorrelation matrix, unlike other gradient-based approaches. The PAST and PASTd algorithms have proven to be useful for several tasks in communications and signal processing [12, 19, 20].

Minor subspace analysis (MSA) is a related array processing problem in which an $(m \times n)$ orthonormal matrix $\mathbf{W}(k)$ that minimizes the least-squares or mean-square energy of the elements of $\mathbf{y}(k)$ is desired [1]. MSA is also useful for numerous tasks, among them frequency estimation and bias removal [21]–[27]. In addition, MSA is practically useful for PSA in situations where the principal subspace dimension m is nearly as large as the signal dimension n, as the principal subspace is the null space of the $(n-m) \times n$ MSA matrix estimate.

To date, almost all proposed MSA algorithms are gradient approaches that have poor convergence performances when the input signal autocorrelation matrix is ill-conditioned [29]–[36]. A notable extension is the algorithm in [37] which is an extension of the PASTd algorithm to the minor component analysis (MCA) task. This algorithm employs a periodic re-orthonormalization of the subspace matrix estimate, and its overall complexity is $\mathcal{O}(m^2 n)$ for a m-component MCA task. No simple $\mathcal{O}(mn)$ method for approximate least-squares tracking of the minor subspace has been proposed.

In this paper, we present a novel PAST-based method for the estimation of minor subspaces. The algorithm employs m identical Householder transformations at each time instant to adjust the rows of $\mathbf{W}(k)$. Our algorithm maintains the orthogonality of the subspace matrix estimation up to the numerical precision of the computing environment. The algorithm requires $9mn + \mathcal{O}(m^2) + \mathcal{O}(n)$ multiply/adds at each time instant, making it of similar complexity as competing gradient-based MSA schemes. Following the analytical methods outlined in [33, 34], we prove that our proposed MSA algorithm performs minor subspace estimation, and simulations verify the superior performance of the method as compared to that of gradient-based schemes in an MSA task.

2. PRINCIPAL SUBSPACE ANALYSIS

Our algorithm is a non-trivial modification of recently-developed methods for principal subspace analysis. In this section, we review the derivation and implementation of the PAST algorithm for principal subspace analysis [11] as well as a numerically-robust implementation of the PAST algorithm [18]. In order to differentiate these existing methods with the proposed PAST-based MSA algorithm, we shall hereafter refer to the original PAST algorithm for PSA as the P-PAST algorithm.

Yang's P-PAST algorithm exactly minimizes the crite-

This material is based in part upon work supported by the Texas Advanced Technology Program under Grant No. 003613-0031-1999.

rion

$$\mathcal{J}(\mathbf{W}(k)) = \sum_{n=1}^{k} \lambda^{k-n} ||\mathbf{x}(n) - \mathbf{W}^T(k)\mathbf{y}(n)||^2 + \lambda^k \delta ||\mathbf{W}(k) - \mathbf{W}(0)||_F^2 \quad (3)$$

at each time instant, where $\mathbf{W}(0)$ is the initial subspace estimate, $||\cdot||_F$ denotes the Frobenius norm, and λ, $0 \ll \lambda < 1$ is the forgetting factor. Using standard techniques, it is straightforward to show that the optimum value for $\mathbf{W}(k)$ is

$$\mathbf{W}(k) = \mathbf{R}_{\mathbf{yy}}^{-1}(k)\mathbf{R}_{\mathbf{yx}}(k) \quad (4)$$

$$\mathbf{R}_{\mathbf{yy}}(k) = \lambda^k \delta \mathbf{I} + \sum_{n=1}^{k} \lambda^{k-n} \mathbf{y}(n)\mathbf{y}^T(n) \quad (5)$$

$$\mathbf{R}_{\mathbf{yx}}(k) = \lambda^k \delta \mathbf{W}(0) + \sum_{n=1}^{k} \lambda^{k-n} \mathbf{y}(n)\mathbf{x}^T(n). \quad (6)$$

One can develop a recursive method for updating $\mathbf{W}(k)$ directly over time using well-known procedures in recursive least-squares (RLS) adaptive filtering [38]. The resulting updates are

$$\mathbf{W}(k) = \mathbf{W}(k-1) + \mathbf{k}(k)\mathbf{e}^T(k) \quad (7)$$

$$\mathbf{e}(k) = \mathbf{x}(k) - \mathbf{u}(k) \quad (8)$$

$$\mathbf{k}(k) = \frac{\mathbf{R}_{\mathbf{yy}}^{-1}(k-1)\mathbf{y}(k)}{\lambda + \mathbf{y}^T(k)\mathbf{R}_{\mathbf{yy}}^{-1}(k-1)\mathbf{y}(k)} \quad (9)$$

$$\mathbf{R}_{\mathbf{yy}}^{-1}(k) = \frac{1}{\lambda}\left(\mathbf{R}_{\mathbf{yy}}^{-1}(k-1) - \mathbf{k}(k)\mathbf{y}^T(k)\mathbf{R}_{\mathbf{yy}}^{-1}(k-1)\right) \quad (10)$$

with $\mathbf{R}_{\mathbf{yy}}^{-1}(0) = \delta^{-1}\mathbf{I}$. Eqns. (7)–(10) constitute the PAST algorithm. This algorithm can be shown to be locally-convergent to the principal subspace of the vector signal sequence; this property and other details of the algorithm's behavior can be found in [11].

The P-PAST algorithm does not exactly maintain the orthogonality constraint given by

$$\mathbf{W}(k)\mathbf{W}^T(k) = \mathbf{I} \quad (11)$$

at each time instant, although it does converge to matrix values whose rows are approximately orthogonal to each other. Recently, a new implementation of the P-PAST algorithm has been develop that exactly maintains the constraint in (11) up to the limits of the numerical accuracy of the computing environment [18]. The coefficient updates for this algorithm are

$$\mathbf{W}(k) = \mathbf{W}(k-1) + \frac{\mathbf{k}(k)\mathbf{v}^T(k)}{1 + \frac{1}{4}||\mathbf{e}(k)||^2||\mathbf{k}(k)||^2}, \quad (12)$$

where the n-dimensional reflection vector $\mathbf{v}(k)$ is

$$\mathbf{v}(k) = \mathbf{e}(k) - \frac{||\mathbf{e}(k)||^2}{2}\mathbf{W}^T(k-1)\mathbf{k}(k). \quad (13)$$

Although it is not obvious from the form of (12), this update applies m identical Householder transformations to each of the rows of $\mathbf{W}(k-1)$ as [18, 39]

$$\mathbf{W}(k) = \mathbf{W}(k-1)\left[\mathbf{I} - 2\frac{\mathbf{v}(k)\mathbf{v}^T(k)}{||\mathbf{v}(k)||^2}\right]. \quad (14)$$

By direct substitution, one can show that $\mathbf{W}(k)\mathbf{W}^T(k) = \mathbf{W}(k-1)\mathbf{W}^T(k-1)$ using (14), such that (11) is satisfied for all $k \geq 0$ if $\mathbf{W}(0)\mathbf{W}^T(0) = \mathbf{I}$. Hence, this algorithm exactly maintains the orthogonality of the subspace matrix estimate, unlike the original P-PAST algorithm. In addition, it is easily shown that, by ignoring all terms of $\mathcal{O}([\mathbf{k}(k)]_i^2)$ and higher on the RHS of (12), one obtains the original P-PAST update in (7). Simulations in [18] indicate that both algorithms have nearly-identical performance within the constraint space defined by (11) for forgetting factors near unity. Only the Householder-based update in (12), however, is successful at maintaining the constraint in (11) to the numerical precision limits of the computing environment. This property is crucial to the development of a PAST-based MSA algorithm, as will become apparent.

3. THE M-PAST ALGORITHM

To formulate the MSA equivalent of the P-PAST algorithm, we consider the original criterion upon which the P-PAST algorithm is based, as given by

$$\overline{\mathcal{J}}(\mathbf{W}(k)) = \sum_{n=1}^{k} \lambda^{k-n} ||\mathbf{x}(n) - \mathbf{W}^T(k)\mathbf{W}(k)x(n)||^2 \quad (15)$$

It is easy to see that $\lim_{k\to\infty} \mathcal{J}(\mathbf{W}(k)) = \overline{\mathcal{J}}(\mathbf{W}(k))$ if $\mathbf{W}(0) = \mathbf{W}(1) = \cdots = \mathbf{W}(k)$. Thus, the two cost functions in (3) and (15) have the same asymptotic stationary points for any given set of constraints on $\mathbf{W}(k)$. Now, if $\mathbf{W}(k)$ satisfies the constraint in (11), then

$$\overline{\mathcal{J}}(\mathbf{W}(k)) = C(k) - \sum_{n=1}^{k} \lambda^{k-n} ||\mathbf{W}(k)x(n)||^2 \quad (16)$$

where $C(k)$ does not depend on $\mathbf{W}(k)$. The second term on the RHS of (16) is the negative of the exponentially-weighted average power of $\mathbf{x}(n)$ in the projected signal subspace of $\mathbf{W}(k)$. Therefore, the optimization problem given by

$$\text{maximize} \quad \overline{\mathcal{J}}(\mathbf{W}(k)) \quad (17)$$

$$\text{such that} \quad \mathbf{W}(k)\mathbf{W}^T(k) = \mathbf{I} \quad (18)$$

corresponds to that of the MSA task. In addition, we may employ the projection approximation in this task by substituting $\mathcal{J}(\mathbf{W}(k))$ for $\overline{\mathcal{J}}(\mathbf{W}(k))$ in (17)–(18). Note that the constraint in (18) is necessary in the case of MSA, as unconstrained maximization of $\mathcal{J}(\mathbf{W}(k))$ would lead to unbounded matrix entries in $\mathbf{W}(k)$.

We can apply the same concepts used to develop the Householder-based P-PAST algorithm in (12) to develop an MSA algorithm that approximately solves (17)–(18). In fact, our starting point is the update in (12), as this algorithm approximately minimizes $\mathcal{J}(\mathbf{W}(k))$ under the constraint in (18). We first replace $\mathbf{k}(k)$ by $-\mathbf{k}(k)$ wherever it appears in the RHS of (12), as this changes corresponds to a reversal in the algorithm's search direction and therefore maximizes $\mathcal{J}(\mathbf{W}(k))$ within the constraint space. The resulting coefficient updates are

$$\mathbf{W}(k) = \mathbf{W}(k-1) - \frac{1}{1+\frac{1}{4}||\mathbf{e}(k)||^2||\mathbf{k}(k)||^2}\Big[\mathbf{k}(k)\mathbf{x}^T(k) - \mathbf{k}(k)\left\{\mathbf{u}^T(k) + \frac{||\mathbf{e}(k)||^2}{2}\mathbf{k}^T(k)\mathbf{W}(k-1)\right\}\Big] \quad (19)$$

Simulations of this algorithm indicate that it does not maintain $\mathbf{W}(k)\mathbf{W}^T(k) = \mathbf{I}$ over time due to accumulation of numerical errors. These error accumulations have been shown to cause catastrophic divergence of $\mathbf{W}(k)$ for other MSA methods [35], and they require one to modify the updates in order to achieve numerically-stable behavior [33, 34, 40, 41].

In [33, 34], a modification to a gradient-based PSA algorithm is developed in which the matrix $\mathbf{W}(k-1)\mathbf{W}^T(k-1)$ is inserted at judicious points within the updates. The resulting algorithm is proven to be self-stabilizing to the space of orthonormal matrices, such that periodic projection of $\mathbf{W}(k)$ back to the space of orthonormal matrices is not required. In this paper, we consider similar modifications to the update in (19) so that it exhibits numerically-stable behavior. The proposed algorithm modification is

$$\mathbf{W}(k) = \mathbf{W}(k-1) - \frac{1}{1+\frac{1}{4}||\mathbf{e}(k)||^2||\mathbf{k}(k)||^2}\Big[\mathbf{z}(k)\mathbf{x}^T(k) - \mathbf{k}(k)\left\{\mathbf{u}^T(k) + \frac{||\mathbf{e}(k)||^2}{2}\mathbf{z}^T(k)\mathbf{W}(k-1)\right\}\Big], \quad (20)$$

where

$$\mathbf{z}(k) = \mathbf{W}(k-1)\mathbf{W}^T(k-1)\mathbf{W}(k-1)\mathbf{W}^T(k-1)\mathbf{k}(k) \quad (21)$$

Eqns. (1), (2), (8)–(10), (20), and (21) describe the proposed M-PAST algorithm for MSA. This algorithm requires $9mn + 2m^2 + 2n + 7m + 4$ multiply/adds and two divides to implement at each iteration. Thus, for $1 \ll m \ll n$, the complexity of this algorithm roughly scales as the number of coefficients within the subspace matrix estimate.

4. ANALYSIS

In this section, we discuss the asymptotic behavior of the proposed M-PAST algorithm. The analysis technique employed here is the ordinary differential equation (ODE) method; for details on this procedure, see [42].

Our analysis makes use of the following assumptions:

1. $\mathbf{x}(k)$ is a zero-mean wide sense stationary random process with covariance $\mathbf{R} = E\{\mathbf{x}(k)\mathbf{x}^T(k)\}$;
2. λ is very close to one such that terms that are of $\mathcal{O}((1-\lambda)^2)$ and higher can be neglected; and
3. the value of $\mathbf{R}_{\mathbf{yy}}^{-1}(k)$ is approximately

$$\mathbf{R}_{\mathbf{yy}}^{-1}(k) \approx (1-\lambda)[\mathbf{W}(k)\mathbf{R}\mathbf{W}^T(k)]^{-1}. \quad (22)$$

The above two assumptions are fairly standard for analyses of this type. The third assumption is reasonable given a value of λ close to one and allows us to decouple the Kalman gain update from the subspace matrix update within the analysis. It is well-known that the $\mathcal{O}(m^2)$ Kalman gain updates in (9)–(10) are robust to perturbations in the $\mathbf{y}(k)$ sequence, so long as the numerical symmetry of $\mathbf{R}_{\mathbf{yy}}^{-1}(k)$ is maintained [43]. For this reason, we will not discuss the numerical properties of this portion of the algorithm in what follows.

Our analysis focuses on the associated ODE of (20) under the above assumptions. This equation is

$$\frac{d\mathbf{W}}{dt} = \mathbf{W} - \mathbf{W}\mathbf{W}^T\mathbf{W}\mathbf{W}^T(\mathbf{W}\mathbf{R}\mathbf{W}^T)^{-1}\mathbf{W}\mathbf{R}. \quad (23)$$

The following two theorems describe the properties of (23), the proofs of which are in the Appendix.

Theorem 1: If $\{\lambda_i : \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-m} > \lambda_{n-m+1} \geq \cdots \geq \lambda_n\}$ are the n eigenvalues of $\mathbf{R}$, then the rows of $\mathbf{W}$ evolve to span the minor subspace of $\mathbf{R}$, such that span$[\mathbf{W}^T] \to$ span$[\mathbf{E}_2]$, where $\mathbf{E}_2$ contains the m eigenvectors corresponding to $\{\lambda_{n-m+1}, \ldots, \lambda_n\}$.

Theorem 2: If rank$[\mathbf{W}] = m$, then the space of locally-stable stationary points satisfy $\mathbf{W}\mathbf{W}^T = \mathbf{I}$, such that the rows of $\mathbf{W}$ are orthogonal and normalized to unit length.

The above theorems imply that the M-PAST algorithm is a valid MSA technique and possesses the "self-stabilizing" property, in which deviations of $\mathbf{W}(k)$ away from orthonormality do not cause the algorithm to fail. The excellent numerical behavior of the proposed algorithm in simulations is shown in the next section.

5. SIMULATIONS

We now explore the behaviors of the M-PAST algorithm via a simulation example. In these simulations, we have compared the performance of the M-PAST algorithm with two other MSA methods:

- The gradient-based minor subspace rule [33, 34]

$$\mathbf{W}(k) = \mathbf{W}(k-1) - \mu[\mathbf{W}(k-1)\mathbf{W}^T(k-1)\mathbf{W}(k-1)\mathbf{u}(k)\mathbf{x}(k) - \mathbf{y}(k)\mathbf{u}^T(k)] \quad (24)$$

- An alternative M-PAST algorithm, in which (20) is replaced by

$$\mathbf{W}(k) = \mathbf{W}(k-1) - \mathbf{k}(k)\mathbf{e}^T(k) \quad (25)$$

  and $\mathbf{W}(k)$ is re-orthonormalized by a singular-value decomposition (SVD) after every coefficient update.

In the example, we have generated sequences of $n = 4$-dimensional independent jointly Gaussian random vectors $\mathbf{x}(k)$ with zero mean and covariance

$$\mathbf{R} = \begin{bmatrix} 0.9 & 0.4 & 0.7 & 0.3 \\ 0.4 & 0.3 & 0.5 & 0.4 \\ 0.7 & 0.5 & 1.0 & 0.6 \\ 0.3 & 0.4 & 0.6 & 0.8 \end{bmatrix}. \quad (26)$$

For each algorithm, we have chosen $m = 2$ and $\delta = 2$, where appropriate. One hundred simulations have been generated

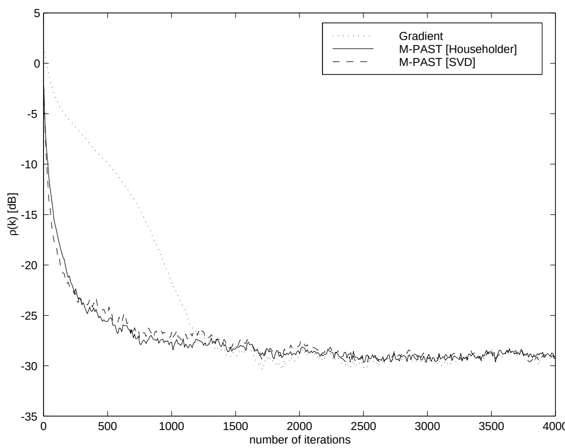

Fig. 1: Evolutions of $E\{\rho(n)\}$ for the three MSA algorithms in the simulation example.

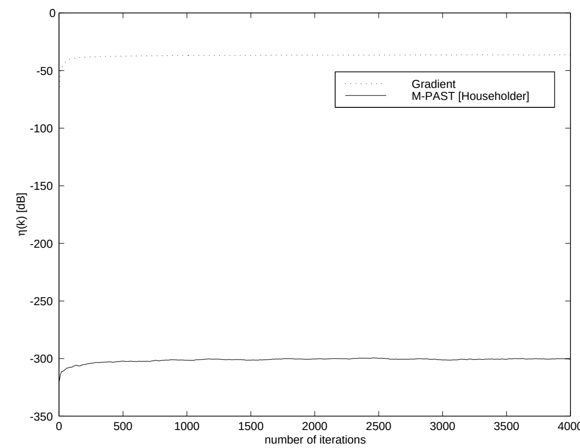

Fig. 2: Evolutions of $E\{\eta(n)\}$ for two of the MSA algorithms in the simulation example.

to evaluate the ensemble averages of the performance factors

$$\rho(k) = ||\mathbf{C}_1(k)||_F^2/||\mathbf{C}_2(k)||_F^2 \quad (27)$$

and $$\eta(k) = ||\mathbf{W}(k)\mathbf{W}^T(k) - \mathbf{I}||_F^2/||\mathbf{W}(k)||_F^2 \quad (28)$$

for each algorithm, where $\mathbf{C}_i(k) = \mathbf{W}(k)\mathbf{E}_i$ for $i \in \{1, 2\}$ and $\mathbf{E}_1$ and $\mathbf{E}_2$ are the two-dimensional principal and minor subspaces, respectively. In each simulation run, we have generated a different random orthogonal coefficient matrix $\mathbf{W}(0)$, and all calculations have been performed using MATLAB signal analysis software.

Fig. 1 shows the average evolutions of $\rho(k)$ for the three algorithms, where the chosen values of $\lambda = 0.99915$, $\mu = 0.011$, and $\lambda = 0.999$ for the algorithms in (20), (24), and (25), respectively, yield the same steady-state performance of approximately -29 dB for these signal statistics. As can be seen, the two M-PAST approaches perform similarly and clearly outperform the gradient-based MSA method in (24). Since the Householder-based M-PAST algorithm avoids using a computationally-intensive SVD procedure, it is to be preferred among the three methods.

Fig. 2 shows the average evolutions of $\eta(k)$ for the Householder-based M-PAST and gradient-based algorithms. As can be seen, the proposed Householder-based update maintains the orthogonality of $\mathbf{W}(k)$ to the precision limits of the computing environment, whereas the minor subspace rule in (24) causes a much-greater deviation of $\mathbf{W}(k)$ from orthogonality.

6. CONCLUSIONS

In this paper, we have proposed a novel efficient method for minor subspace analysis (MSA). The algorithm is a variant of Yang's projection approximation subspace tracking (PAST) method and uses m identical Householder updates at each time instant. The proposed algorithm is simple, requiring about $9mn$ multiply/adds at each time instant to track an m-dimensional minor subspace using n-dimensional signal measurements. Analyses verify that the algorithm behaves in a numerically-robust fashion, and simulations indicate that the method outperforms an existing gradient-based MSA approach.

7. APPENDIX

Proof of Theorem 1:

Let $\mathbf{E}_1 \in \mathcal{R}^{n\times(n-m)}$ and $\mathbf{E}_2 \in \mathcal{R}^{n\times m}$ contain the principal $(n-m)$ and minor m eigenvectors of $\mathbf{R}$, respectively. We can express $\mathbf{W}(t)$ in (23) as

$$\mathbf{W} = \mathbf{A}_1\mathbf{E}_1^T + \mathbf{A}_2\mathbf{E}_2^T, \quad (29)$$

where $\mathbf{A}_1(t) \in \mathcal{R}^{m\times(n-m)}$ and $\mathbf{A}_2(t) \in \mathcal{R}^{m\times m}$. Post-multiplying both sides of (23) by $\mathbf{E}_1$ and $\mathbf{E}_2$ gives

$$\frac{d\mathbf{A}_1}{dt} = \mathbf{A}_1 - \mathbf{B}\mathbf{\Sigma}^{-1}\mathbf{A}_1\mathbf{\Lambda}_1 \quad (30)$$

and $$\frac{d\mathbf{A}_2}{dt} = \mathbf{A}_2 - \mathbf{B}\mathbf{\Sigma}^{-1}\mathbf{A}_2\mathbf{\Lambda}_2, \quad (31)$$

respectively, where $\mathbf{B}(t) = [\mathbf{A}_1(t)\mathbf{A}_1^T(t) + \mathbf{A}_2(t)\mathbf{A}_2^T(t)]\times[\mathbf{A}_1(t)\mathbf{A}_1^T(t) + \mathbf{A}_2(t)\mathbf{A}_2^T(t)]$, $\mathbf{\Lambda}_1 = \mathrm{diag}\{\lambda_1, \ldots, \lambda_{n-m}\}$, $\mathbf{\Lambda}_2 = \mathrm{diag}\{\lambda_{n-m+1}, \ldots, \lambda_n\}$, $\lambda_k > \lambda_l$ for all $1 \le k \le n-m$ and $n-m+1 \le l \le n$, and $\mathbf{\Sigma}(t) = \mathbf{W}(t)\mathbf{R}\mathbf{W}^T(t)$.

Consider the matrix $\mathbf{Q}(t) = \mathbf{A}_2^{-1}(t)\mathbf{A}_1(t)$. If $\mathbf{A}_2(0)$ is nonsingular, then $\mathbf{B}(0)$ is positive-definite, and thus $\mathbf{A}_2(t)$ is nonsingular from (30)–(31). As shown in [1, pp. 88],

$$\frac{d\mathbf{Q}}{dt} = \mathbf{A}_2^{-1}\frac{d\mathbf{A}_1}{dt} - \mathbf{A}_2^{-1}\frac{d\mathbf{A}_2}{dt}\mathbf{A}_2^{-1}\mathbf{A}_1. \quad (32)$$

Substituting (30)–(31) for $d\mathbf{A}_1/dt$ and $d\mathbf{A}_2/dt$ into (32) and simplifying, we obtain after some algebra

$$\frac{d\mathbf{Q}}{dt} = \mathbf{A}_2^{-1}\mathbf{B}\mathbf{\Sigma}^{-1}\mathbf{A}_2\left[\mathbf{\Lambda}_2\mathbf{Q} - \mathbf{Q}\mathbf{\Lambda}_1\right] = \mathbf{M}[\mathbf{\Lambda}\circ\mathbf{Q}], \quad (33)$$

where $\mathbf{M}(t) = \mathbf{A}_2^{-1}(t)\mathbf{B}(t)\mathbf{\Sigma}^{-1}\mathbf{A}_2(t) \in \mathcal{R}^{m\times m}$ and $\mathbf{\Lambda}\circ$ is a linear operator acting on $\mathbf{Q}(t)$. Since $\mathbf{B}(t)$ is the square of a positive-definite symmetric matrix, it is also positive-definite, as is $\mathbf{M}(t)$.

We now show that $\mathbf{\Lambda}\circ$ is a negative operator. We may regard $\mathbf{Q}(t)$ as an extended vector and search for all of the eigenvalues of $\mathbf{\Lambda}\circ$. The eigenvalues of $\mathbf{\Lambda}\circ$ are $\lambda_{kl} = \lambda_k - \lambda_l$, $k \in \{n-m+1, \ldots, n\}$, $l \in \{1, \ldots, n-m\}$, and the corresponding eigenvectors (vectorized matrices) are $\mathbf{Q}_{kl}(t)$ whose components are all zero except for the (k,l)th element that is unity valued, because

$$\mathbf{\Lambda} \circ \mathbf{Q}_{kl} = (\lambda_k - \lambda_l)\mathbf{Q}_{kl}. \quad (34)$$

Since $\lambda_k - \lambda_l < 0$ for any pair $\{k, l\}$, $\mathbf{\Lambda}\circ$ is a negative operator, and all of the eigenvalues of $\mathbf{M}(t)[\mathbf{\Lambda}\circ]$ have negative real parts. Thus, we have

$$\lim_{t\to\infty} \mathbf{Q}(t) = \mathbf{0}, \quad (35)$$

and therefore $\text{span}[\mathbf{W}^T] \to \text{span}[\mathbf{E}_2]$. □

Proof of Theorem 2:

The fixed points of (23) satisfy

$$\mathbf{W}\mathbf{W}^T\mathbf{W}\mathbf{W}^T(\mathbf{W}\mathbf{R}\mathbf{W}^T)^{-1}\mathbf{W}\mathbf{R} = \mathbf{W}. \quad (36)$$

If $\text{rank}[\mathbf{W}] = m$, then from Theorem 1, the only potential stationary points of (23) are of the form $\mathbf{W} = \mathbf{A}_2\mathbf{E}_2^T$, where $\mathbf{A}_2 \in \mathcal{R}^{m\times m}$ is an arbitrary nonsingular matrix and $\mathbf{E}_2$ is defined as previously. Substituting this form into (36) and defining $\mathbf{P} = \mathbf{A}_2^T\mathbf{A}_2$, we obtain after some simplification

$$\mathbf{P}\mathbf{P} = \mathbf{P}. \quad (37)$$

Eqn. (37) implies that $\mathbf{P}$ is a projection operator, and since $\mathbf{P}$ is symmetric, we have immediately that $\mathbf{P} = \mathbf{A}_2^T\mathbf{A}_2 = \mathbf{A}_2\mathbf{A}_2^T = \mathbf{I}$. Therefore, the space of stationary points of (23) satisfy $\text{span}[\mathbf{W}^T] = \text{span}[\mathbf{E}_2]$ and $\mathbf{W}\mathbf{W}^T = \mathbf{I}$.

To determine the local stability characteristics about the space of stationary points, examine $\mathbf{P}(t) = \mathbf{A}_2^T(t)\mathbf{A}_2(t)$ for $t \geq t_0 \geq 0$ when $\mathbf{W}(t) = \mathbf{A}_2(t)\mathbf{E}_2^T$. Using (31), we can obtain

$$\frac{d\mathbf{P}}{dt} = 2\mathbf{P}\left(\mathbf{I} - \mathbf{P}\right). \quad (38)$$

It is straightforward to show that the solution to (38) is

$$\mathbf{P}(t) = \left(\mathbf{I} - e^{-2t}\mathbf{N}_0\right)^{-1}, \quad (39)$$

where $\mathbf{N}_0 = \mathbf{I} - \mathbf{P}^{-1}(t_0)$. Therefore, $\lim_{t\to\infty}\mathbf{P}(t) = \mathbf{I}$, and (23) is locally asymptotically stable about the space of stationary points satisfying $\text{span}[\mathbf{W}^T] = \text{span}[\mathbf{E}_2]$ and $\mathbf{W}\mathbf{W}^T = \mathbf{I}$.

8. REFERENCES

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

[2] T.P. Krasulina, "Method of stochastic approximation in the determination of the largest eigenvalue of the mathematical expectation of random matrices," *Automat. Remote Contr.,* vol. 2, pp. 215-221, 1970.

[3] N.L. Owsley, "Adaptive data orthonormalization," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Tulsa, OK, pp. 109-112, Apr. 1978.

[4] P.A. Thompson, "An adaptive spectral analysis technique for unbiased frequency estimation in the presence of white noise," *Proc. 13th Asilomar Conf. Circ., Syst., Comput.,* Pacific Grove, CA, pp. 529-533, Nov. 1979.

[5] E. Oja and J. Karhunen, "On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix," *J. Math Anal. Appl.,* vol. 106, no. 1, pp. 69-84, 1985.

[6] J. Yang and M. Kaveh, "Adaptive eigensubspace algorithms for direction or frequency estimation and tracking," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 36, pp. 241-251, Feb. 1988.

[7] P.A. Regalia, "An adaptive unit norm filter with applications to signal analysis and Karhunen-Loève transformations," *IEEE Trans. Circuits Syst.,* vol. 37, pp. 646-649, May 1990.

[8] P. Comon and G.H. Golub, "Tracking a few extreme singular values and vectors in signal processing," *Proceedings of the IEEE,* vol. 78, pp. 1327-1343, Aug. 1990.

[9] R.D. DeGroat, "Noniterative subspace tracking," *IEEE Trans. Signal Processing,* vol. 40, pp. 571-577, Mar. 1992.

[10] G.W. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. Signal Processing,* vol. 40, pp. 1535-1541, June 1992.

[11] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing,* vol. 43, pp. 95-107, Jan. 1995.

[12] B. Yang, "Extension of the PASTd algorithm to both rank and subspace tracking," *IEEE Signal Processing Lett.,* vol. 2, pp. 179-182, Sept. 1995.

[13] B. Yang, "Asymptotic convergence analysis of the projection approximation subspace tracking algorithms," *Signal Processing,* vol. 50, pp. 123-136, Apr. 1996.

[14] D.R. Fuhrmann, "A geometric approach to subspace tracking," *Proc. 31st Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, vol. 1, pp. 783-787, Nov. 1997.

[15] V. Solo and X. Kong, "Performance analysis of adaptive eigenanalysis algorithms," *IEEE Trans. Signal Processing,* vol. 46, pp. 636-646, Mar. 1998.

[16] A. Edelman, T. Arias, and S.T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.,* vol. 20, no. 2, pp. 303-353, Apr. 1998.

[17] Y. Hua, Y. Xiang, T. Chen, K. Abed-Meraim, and Y. Miao, "Natural power method for fast subspace tracking," *Proc. IEEE Workshop Neural Networks for Signal Processing,* Madison, WI, pp. 176-185, August 1999.

[18] S.C. Douglas, "Numerically-robust adaptive subspace tracking using Householder transformations," presented at the *1st IEEE Sensor Array Multichannel Signal Processing Workshop,* Boston, MA, Mar. 2000.

[19] X. Wang and H.V. Poor, "Blind adaptive multiuser detection in multipath CDMA channels based on subspace tracking," *IEEE Trans. Signal Processing,* vol. 46, pp. 303-3044, Nov. 1998.

[20] J. Sanchez-Araujo, and S. Marcos, "An efficient PASTd-algorithm implementation for multiple direction of arrival tracking," *IEEE Trans. Signal Processing,* vol. 47, pp. 2321-2324, Aug. 1999.

[21] V.U. Reddy, B. Egardt, and T. Kailath, "Least squares type algorithm for adaptive implementation of Pisarenko's harmonic retrieval method," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-30, pp. 399-405, June 1982.

[22] P.A. Regalia, "An unbiased equation error identifier and reduced-order approximations," *IEEE Trans. Signal Processing,* vol. 42, pp. 1297-1412, June 1994.

[23] K. Gao, M.O. Ahmad, and M.N.S. Swamy, "A constrained anti-Hebbian learning algorithm for total least-squares estimation with applications to adaptive FIR and IIR filtering," *IEEE Trans. Circuits Syst. II: Analog Digital Signal Processing,* vol. 41, pp. 718-729, Nov. 1994.

[24] K.C. Ho and Y.T. Chan, "Bias removal in equation error adaptive IIR filters," *IEEE Trans. Signal Processing,* vol. 43, pp. 51-62, Jan. 1995.

[25] S.C. Douglas, "Analysis of an anti-Hebbian adaptive FIR filtering algorithm," *IEEE Trans. Circuits Syst. II: Analog Digital Signal Processing,* vol. 43, pp. 777-780, Nov. 1996.

[26] S.C. Douglas and M. Rupp, "On bias removal and unit norm constraints in equation-error adaptive IIR filters," *Proc. 30th Asilomar Conf. Sig., Syst., Comput.,* Pacific Grove, CA, vol. 2, pp. 1093-1097, Nov. 1996.

[27] M. Rupp and S.C. Douglas, "Deterministic stability analyses of unit-norm constrained algorithms for unbiased adaptive IIR filtering," *IEEE Int. Conf. Acoust., Speech, Signal Processing,* Munich, Germany, vol. 3, pp. 1937-1940, Apr. 1997.

[28] D.R. Fuhrmann and B. Liu, "An iterative algorithm for locating the minimal eigenvector of a symmetric matrix," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Dallas, TX, pp. 45.8.1-45.8.4, 1984.

[29] L. Wang and J. Karhunen, "A unified natural bigradient algorithm for robust PCA and MCA," *Int. J. Neural Syst.,* vol. 7, no. 1, pp. 53-67, Mar. 1996.

[30] T. Chen, "Modified Oja's algorithms for principal subspace and minor subspace extraction," *Neural Processing Lett.,* vol. 5, no. 2, pp. 105-110, Apr. 1997.

[31] F.-L. Luo and R. Unbehauen, "A minor subspace analysis algorithm," *IEEE Trans. Neural Networks,* vol. 8, pp. 1149-1155, Sept. 1997.

[32] T.-P. Chen, S. Amari, and Q. Lin, "A unified algorithm for principal and minor components extraction," *Neural Networks,* vol. 11, pp 385-390, Apr. 1998.

[33] S.C. Douglas, S.-Y. Kung, and S. Amari, "A self-stabilized minor subspace rule," *IEEE Signal Processing Lett.,* vol. 5, pp. 330-332, Dec. 1998.

[34] S.C. Douglas, S.-Y. Kung, and S. Amari, "On the numerical stabilities of principal, minor, and independent component analysis algorithms," *Proc. IEEE Int. Workshop Indep. Compon. Anal. Signal Separation,* Aussois, France, pp. 419-425, Jan. 1999.

[35] A. Taleb and G. Cirrincione, "Against the convergence of the minor component analysis neurons," *IEEE Trans. Neural Networks,* vol. 10, pp. 207-210, Jan. 1999.

[36] S. Ouyang, Z. Bao, and G. Liao, "Adaptive step-size minor component extraction algorithm," *Electron. Lett.,* vol. 35, no. 6, pp. 443-444, 18 Mar. 1999.

[37] H. Sakai and K. Shimizu, "New adaptive algorithm for minor component analysis," *Signal Processing,* vol. 71, no. 3, pp. 301-308, Dec. 1998.

[38] S. Haykin, *Adaptive Filter Theory,* 3rd ed. (Upper Saddle River, NJ: Prentice-Hall, 1996).

[39] G.H. Golub and C.F. Van Loan, *Matrix Computations* (Baltimore, MD: Johns Hopkins, 1989).

[40] S.C. Douglas, S. Amari, and S.Y. Kung, "Gradient adaptation under unit-norm constraints," *Proc. 9th IEEE Workshop Stat. Signal Array Processing,* Portland, OR, pp. 144-147, Sept. 1998.

[41] S.C. Douglas, S. Amari, and S.-Y. Kung, "On gradient adaptation with unit-norm constraints," accepted for publication in *IEEE Trans. Signal Processing;* to appear.

[42] A. Benveniste, M. Métivier, and P. Piouret, *Adaptive Algorithms and Stochastic Approximations* (Berlin: Springer-Verlag, 1990).

[43] M.H. Verhaegen, "Round-off error propagation in four generally-applicable, recursive, least-squares estimation schemes," *Automatica,* vol. 25, no. 3, pp. 437-444, May 1989.