ADAPTIVE ESTIMATION OF THE STRONG UNCORRELATING TRANSFORM WITH APPLICATIONS TO SUBSPACE TRACKING

Scott C. Douglas

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

Jan Eriksson and Visa Koivunen

Signal Processing Laboratory, SMARAD CoE
Helsinki University of Technology
Espoo 02015 Finland

ABSTRACT

In some signal processing tasks involving complex-valued multichannel measurements, classical whitening approaches do not completely remove the second-order statistical dependencies of the data. This paper describes adaptive procedures for estimating the strong uncorrelating transform for jointly diagonalizing the covariance and pseudo-covariance matrices of multidimensional signals. Novel algorithms are derived that extend and combine the power method and orthogonal iterations with ordinary fixed and iterative whitening procedures. Finally, we show how to combine our procedures with orthogonal PAST algorithms to perform subspace tracking and source signal clustering based on non-circularity.

1. INTRODUCTION

For zero-mean m-dimensional signals $\mathbf{x}(k)$, the sample covariance matrix and pseudo-covariance matrices

$$\widehat{\mathbf{R}}_{\mathbf{xx},t}=\frac{1}{t}\sum_{k=1}^{t}\mathbf{x}(k)\mathbf{x}^H(k) \quad \text{and} \quad \widehat{\mathbf{P}}_{\mathbf{xx},t}=\frac{1}{t}\sum_{k=1}^{t}\mathbf{x}(k)\mathbf{x}^T(k) \quad (1)$$

where $\cdot^H$ and $\cdot^T$ denote Hermitian (complex) and ordinary transpose operations, respectively, give complete descriptions of the second-order statistical properties of the measurements. The covariance matrix $\widehat{\mathbf{R}}_{\mathbf{xx},t}$ is widely used in parameter estimation, subspace tracking, and data compression. Recently, pseudo (co)variance has also been used in algorithms for carrier offset estimation in wireless communications [1], channel equalization [2], and blind separation of complex-valued signal mixtures [3, 4]. A key property in these applications is that $\widehat{\mathbf{P}}_{\mathbf{xx},t} \neq \mathbf{0}$ due to the underlying problem structure, implying that the measurement statistics are complex *non-circular*. For example, in subspace tracking of multidimensional signals impinging on an antenna array, some of the transmitted signals may have zero energy in their imaginary components, and the matrix $\widehat{\mathbf{P}}_{\mathbf{xx},t}$ along with $\widehat{\mathbf{R}}_{\mathbf{xx},t}$ can be used to characterize the array response for these signals.

In this paper, we develop adaptive algorithms for estimating the strong-uncorrelating transform (SUT) $\mathbf{W}$ for jointly diagonalizing and normalizing $\widehat{\mathbf{R}}_{\mathbf{xx}}$ and $\widehat{\mathbf{P}}_{\mathbf{xx}}$, such that

$$\mathbf{W}\widehat{\mathbf{R}}_{\mathbf{xx}}\mathbf{W}^H = \mathbf{I} \quad \text{and} \quad \mathbf{W}\widehat{\mathbf{P}}_{\mathbf{xx}}\mathbf{W}^T = \mathbf{\Lambda}, \quad (2)$$

where $\widehat{\mathbf{R}}_{\mathbf{xx},t}$ is positive definite and $\mathbf{\Lambda}$ is a diagonal matrix of circularity coefficients $\{\lambda_i\}$, $i \in \{1,\ldots,m\}$ satisfying $0 \leq \lambda_i \leq \lambda_l \leq 1$ for $i > l$. The SUT is defined in [3, 4] in the context of independent component analysis and blind source separation, and very few methods for computing it have been described. In this paper, we extend classic techniques for eigenvector estimation of Hermitian symmetric matrices to computing the Takagi factorization of a complex symmetric matrix needed for the SUT. We then illustrate the importance of the SUT for subspace tracking tasks. In particular, we derive an extension of the projection approximation subspace tracking (PAST) algorithm in which the signal subspace is further decomposed into smaller subspaces due to signal non-circularity.

2. AN ITERATIVE PROCEDURE FOR FINDING THE STRONG-UNCORRELATING TRANSFORM

As defined in [3, 4], the SUT for any $\widehat{\mathbf{R}}_{\mathbf{xx}}$ and $\widehat{\mathbf{P}}_{\mathbf{xx}}$ is

$$\mathbf{W} \;=\; \mathbf{Q}^H\mathbf{G}, \quad (3)$$

where $\mathbf{G}$ is any prewhitening transformation of the measurements $\mathbf{x}(k)$ such that

$$\mathbf{G}\mathbf{R}_{\mathbf{xx}}\mathbf{G}^H \;=\; \mathbf{I}, \quad (4)$$

and the unitary matrix $\mathbf{Q}$ comes from the Takagi factorization or symmetric SVD of the matrix

$$\overline{\mathbf{P}} \;=\; \mathbf{G}\widehat{\mathbf{P}}_{\mathbf{xx}}\mathbf{G}^T \;=\; \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T, \quad (5)$$

in which $\mathbf{\Lambda}$ contains the ordered circularity coefficients λ_i. For $\mathbf{G}$, any prewhitening transform can be used, such as (i) the inverse of the Cholesky factor of $\widehat{\mathbf{R}}_{\mathbf{xx}}$, (ii) the inverse symmetric square root of $\widehat{\mathbf{R}}_{\mathbf{xx}}$ computed from the eigenvalue decomposition $\widehat{\mathbf{R}}_{\mathbf{xx}} = \mathbf{E}\mathbf{\Gamma}\mathbf{E}^H$ as $\mathbf{G} = \mathbf{E}\mathbf{\Gamma}^{-1/2}\mathbf{E}^H$, and (iii) the principal component whitening transformation $\widehat{\mathbf{R}}_{\mathbf{xx}} = \mathbf{\Gamma}^{-1/2}\mathbf{E}^H$. Note that $\overline{\mathbf{P}}$ in (5) is symmetric but not Hermitian symmetric; hence, a normal (complex-valued) eigenvalue decomposition of $\overline{\mathbf{P}}$ will not yield $\mathbf{Q}$ and $\mathbf{\Lambda}$.

In linear algebra, the power method and orthogonal iterations are classical procedures for finding one or more eigenvectors of a Hermitian symmetric complex matrix [5]. In this section, we extend these methods to computing one or more columns of $\mathbf{Q}$ in (5). For finding the first column $\mathbf{q}_1$ of $\mathbf{Q}$, consider a novel modification of the power method given by

$$\widetilde{\mathbf{q}}_{1,t} = \overline{\mathbf{P}}\mathbf{q}^*_{1,t-1} + \beta\mathbf{q}_{1,t-1}, \qquad \mathbf{q}_{1,t} = \frac{\widetilde{\mathbf{q}}_{1,t}}{\sqrt{\widetilde{\mathbf{q}}^H_{1,t}\widetilde{\mathbf{q}}_{1,t}}}, \quad (6)$$

where β is a non-zero real-valued shift parameter designed to speed the convergence of the iteration. Eqn. (6) differs from the classic power method, as (6) uses both $\mathbf{q}_{1,t-1}$ and $\mathbf{q}^*_{1,t-1}$. The following theorem and corollary pertain to (6).

Theorem 1: *If the λ_i values are distinct and $\beta > 0$, (6) causes $\mathbf{q}_{1,t}$ to converge exponentially to $\pm\mathbf{q}_1$. The rate of convergence is upper bounded by*

$$\max\left[\left|\frac{\beta-\lambda_1}{\beta+\lambda_1}\right|, \left|\frac{\beta+\lambda_2}{\beta+\lambda_1}\right|\right]. \quad (7)$$

Corollary 1.1: The optimum value of β is $\beta_{opt} = (\lambda_1-\lambda_2)/2$, in which case the rate is $(\lambda_1+\lambda_2)/(3\lambda_1-\lambda_2)$.

Proof: Consider (6) using the vector $\mathbf{b}_t = \mathbf{Q}^H\mathbf{q}_{1,t}$, for which

$$\widetilde{\mathbf{b}}_t \;=\; \mathbf{\Lambda}\mathbf{b}^*_{t-1} + \beta\mathbf{b}_{t-1}. \quad (8)$$

The real and imaginary parts of the ith element of $\mathbf{b}_t$ are

$$\widetilde{b}_{R,i,t} = (\beta+\lambda_i)b_{R,i,t-1} \text{ and } \widetilde{b}_{I,i,t} = (\beta-\lambda_i)b_{I,i,t-1}. (9)$$

Considering the normalization step, the evolution of the $2m$-element real vector $[b_{R,1,t}\; b_{I,1,t}\; \cdots\; b_{R,m,t}\; b_{I,m,t}]^T$ is identical to the power method applied to the real-valued diagonal matrix $\mathbf{D}$ with diagonal entries $\{\beta+\lambda_1, \beta-\lambda_1, \ldots, \beta+\lambda_m, \beta-\lambda_m\}$. Hence, $\mathbf{b}_t$ converges to $\pm[1\; 0\; \cdots\; 0]^T$, implying that $\mathbf{q}_{1,t} \rightarrow \pm\mathbf{q}_1$. The rate is upper-bounded by the ratio of the two largest magnitude eigenvalues of $\mathbf{D}$, which leads to (7). The optimum choice of β minimizes the maximum of these ratios, which is achieved when $\beta = \beta_{opt}$ $\circ$.

Given (6), it is straightforward to develop a parallel implementation that extracts all of the columns of $\mathbf{Q}$. The method extends the method of orthogonal iterations to the symmetric SVD for complex symmetric matrices with distinct circularity coefficients. Let $\mathbf{Q}_t = [\mathbf{q}_{1,t}\; \cdots\; \mathbf{q}_{m,t}]$ denote the estimate of $\mathbf{Q}$ at iteration t, and consider the update

$$\mathbf{Q}_t\overline{\mathbf{R}}_t \;=\; \mathtt{qr}[\overline{\mathbf{P}}\mathbf{Q}^*_{t-1} + \beta\mathbf{Q}_{t-1}]. \quad (10)$$

where $\mathtt{qr}[\cdot]$ denotes the QR decomposition of a matrix and $\overline{\mathbf{R}}_t$ is upper-triangular. Given Theorem 1, several properties of (10) can be stated:

1. If the $\{\lambda_i\}$ values are distinct, $\mathbf{Q}_t$ converges exponentially to $\mathbf{QJ}$, where $\mathbf{J}$ is a diagonal matrix of ±1 entries.

2. The matrix $\overline{\mathbf{R}}_t$ converges exponentially to $\mathbf{\Lambda} + \beta\mathbf{I}$.

3. Convergence is guaranteed for any $\beta > 0$. Adaptive procedures could be developed for computing individual shifts for each column of $\mathbf{Q}_t$, but for algorithm simplicity, we propose

$$\beta \;=\; \frac{1}{2m}. \quad (11)$$

3. ADAPTIVE ESTIMATION OF THE STRONG UNCORRELATING TRANSFORM

The iterative technique in (10) is now used to define an adaptive procedure for estimating the SUT as

$$\mathbf{W}_t \;=\; \mathbf{Q}^H_t\mathbf{G}_t, \quad (12)$$

where both $\mathbf{G}_t$ and $\mathbf{Q}_t$ are adapted according to time-varying estimates of $\widehat{\mathbf{R}}_{\mathbf{xx},\mathbf{t}}$ and $\overline{\mathbf{P}}_t = \mathbf{G}_t\widehat{\mathbf{P}}_{\mathbf{xx},\mathbf{t}}\mathbf{G}^T_t$. Using two separate matrices to represent $\mathbf{W}_t$ in (12) is both inconvenient and computationally undesirable. For this reason, consider the situation in which either (a) $\mathbf{G}_t$ does not change with time, or (b) $\mathbf{G}_t$ is adapted using an update of the form

$$\mathbf{G}_t \;=\; \mathbf{G}_{t-1}\cdot\mathbf{F}[\mathbf{G}^H_{t-1}\mathbf{G}_{t-1}, \widehat{\mathbf{R}}_{\mathbf{xx},t}, \mathbf{x}(t)], \quad (13)$$

where the matrix function $\mathbf{F}[\mathbf{M},\mathbf{N},\mathbf{x}]$ depends on the matrices $\mathbf{M}$ and $\mathbf{N}$ and/or the vector $\mathbf{x}$. Adaptive prewhitening procedures are given at the end of this section.

Using (12), we obtain from (10) the relation

$$\mathbf{Q}_t\overline{\mathbf{R}}_t \;=\; \mathbf{G}_{t-1}\widehat{\mathbf{P}}_{\mathbf{xx},t}\mathbf{G}^T_{t-1}\mathbf{Q}^*_{t-1} + \beta\mathbf{Q}_{t-1}. \quad (14)$$

Next, we pre-multiply both sides of (14) by $\mathbf{Q}^{-1}_{t-1} = \mathbf{Q}^H_{t-1}$ and use the fact that $\mathbf{Q}_{t-1}$ is unitary to obtain

$$\mathbf{Q}^{-1}_{t-1}\mathbf{Q}_t\overline{\mathbf{R}}_t \;=\; \mathbf{W}_{t-1}\widehat{\mathbf{P}}_{\mathbf{xx},t}\mathbf{W}^T_{t-1} + \beta\mathbf{I} \quad (15)$$

Define the orthogonal matrix $\mathbf{\Theta}_t = \mathbf{Q}^{-1}_{t-1}\mathbf{Q}_t$. Then, by combining (13) and (15), an update expression for $\mathbf{W}_t$ is

$$\mathbf{\Theta}_t\overline{\mathbf{R}}_t \;=\; \mathtt{qr}[\mathbf{W}_{t-1}\widehat{\mathbf{P}}_{\mathbf{xx},t}\mathbf{W}^T_{t-1} + \beta\mathbf{I}] \quad (16)$$

$$\mathbf{W}_t \;=\; \mathbf{\Theta}^H_t\mathbf{W}_{t-1}\cdot\mathbf{F}[\mathbf{W}^H_{t-1}\mathbf{W}_{t-1}, \widehat{\mathbf{R}}_{\mathbf{xx},t}, \mathbf{x}(t)], \quad (17)$$

where we have used $\mathbf{W}^H_t\mathbf{W}_t = \mathbf{G}^H_t\mathbf{Q}_t\mathbf{Q}^H_t\mathbf{G}_t = \mathbf{G}^H_t\mathbf{G}_t$.

The updates in (16) and (17) are our main algorithm iterations. Eqn. (16) computes an orthogonal update matrix $\mathbf{\Theta}_t$ that pre-multiplies $\mathbf{W}_{t-1}$ on the left in (17), the latter of which is also post-multiplied by the update matrix function $\mathbf{F}$. The function $\mathbf{F}$ should be specified such that $\mathbf{W}_t\widehat{\mathbf{R}}_{\mathbf{xx},t}\mathbf{W}^H_t$ is close to or exactly equal to an identity matrix. Note that no approximations were used in deriving (16)–(17), *i.e.*, $\mathbf{W}_t$ exactly equals (12), where $\mathbf{Q}_t$ evolves according to (10) with $\overline{\mathbf{P}} = \overline{\mathbf{P}}_t \equiv \mathbf{G}_{t-1}\widehat{\mathbf{P}}_{\mathbf{xx},t}\mathbf{G}^T_{t-1}$.

Eqns. (16)–(17) require estimates $\widehat{\mathbf{R}}_{\mathbf{xx},t}$ and $\widehat{\mathbf{P}}_{\mathbf{xx},t}$ of the covariance and pseudo-covariance of $\mathbf{x}(t)$. Time averaging of the sequences $\mathbf{x}(t)\mathbf{x}^H(t)$ and $\mathbf{x}(t)\mathbf{x}^T(t)$ can be used to estimate these matrices. Of particular interest are the exponentially-windowed estimates generated by

$$\widehat{\mathbf{R}}_{\mathbf{xx},t} \;=\; \alpha\widehat{\mathbf{R}}_{\mathbf{xx},t-1} + \rho\mathbf{x}(t)\mathbf{x}^H(t) \quad (18)$$

$$\widehat{\mathbf{P}}_{\mathbf{xx},t} \;=\; \alpha\widehat{\mathbf{P}}_{\mathbf{xx},t-1} + \rho\mathbf{x}(t)\mathbf{x}^T(t) \quad (19)$$

for $0 \ll \alpha \leq 1$ and $\rho > 0$, which are computationally-attractive. For $\rho = 1-\alpha$, these estimates are unbiased.

$$
\begin{aligned}
&\mathbf{y}(t) = \mathbf{W}_{t-1}\mathbf{x}(t)\\
&\mathbf{u}(t) = \mathbf{W}_{t-1}^H\mathbf{y}(t)\\
&\zeta(t) = \frac{1}{\mathbf{y}^H(t)\mathbf{y}(t)}\left(1-\sqrt{\frac{\alpha}{\alpha+\rho\mathbf{y}^H(t)\mathbf{y}(t)}}\right)\\
&\widehat{\mathbf{P}}_{\mathbf{xx},t} = \alpha\widehat{\mathbf{P}}_{\mathbf{xx},t-1}+\rho\mathbf{x}(t)\mathbf{x}^T(t)\\
&\overline{\mathbf{P}}_t = \mathbf{W}_{t-1}\widehat{\mathbf{P}}_{\mathbf{xx},t}\mathbf{W}_{t-1}^T\\
&\boldsymbol{\Theta}_t\overline{\mathbf{R}}_t = \mathrm{qr}[\overline{\mathbf{P}}_t+\beta\mathbf{I}]\\
&\mathbf{W}_t = \frac{1}{\sqrt{\alpha}}\boldsymbol{\Theta}_t^H\left(\mathbf{W}_{t-1}-\zeta(t)\mathbf{y}(t)\mathbf{u}^H(t)\right)
\end{aligned}
$$

Table 1: The adaptive SUT algorithm employing orthogonal iterations.

As for the matrix function $\mathbf{F}$, we are motivated by existing prewhitening algorithms to choose specific forms. The choice

$$\mathbf{F}[\mathbf{M},\mathbf{N},\cdot] \;=\; \frac{3}{2}\mathbf{I}-\frac{1}{2}\mathbf{NM} \qquad (20)$$

yields a fast-converging iteration that is related to a classic procedure for vector orthogonalization [6]. Convergence of this iteration to a prewhitening solution is at least quadratic if $\widehat{\mathbf{R}}_{\mathbf{xx},t}$ does not change with time; the value of $\mathbf{W}_t$ must be scaled, however, to maintain stability. The choice

$$\mathbf{F}[\mathbf{M},\cdot,\mathbf{x}] \;=\; (1+\mu)\mathbf{I}-\mu\mathbf{x}\mathbf{x}^H\mathbf{M} \qquad (21)$$

can be viewed as a stochastic implementation of (17) and (20) with an adjustible step size μ [7, 8]. Finally, the choice

$$\mathbf{F}[\mathbf{M},\cdot,\mathbf{x}] \;=\; \frac{1}{\sqrt{\alpha}}\left(\mathbf{I}-\zeta\mathbf{x}\mathbf{x}^H\mathbf{M}\right) \qquad (22)$$

$$\zeta \;=\; \frac{1}{\mathbf{x}^H\mathbf{M}\mathbf{x}}\left(1-\sqrt{\frac{\alpha}{\alpha+\rho\mathbf{x}^H\mathbf{M}\mathbf{x}}}\right) \qquad (23)$$

yields a modified Householder RLS algorithm [9, 10] in which a $\sqrt{\rho}$ scaling is imparted on $\mathbf{W}_t$. This algorithm has particular *deterministic* prewhitening properties.

Theorem 2: *For an arbitrary pseudo-covariance matrix estimate $\widehat{\mathbf{P}}_{\mathbf{xx},t}$, the algorithm in (16)–(17) and (22)–(23) produces a sequence $\mathbf{W}_t$ that exactly satisfies $\mathbf{W}_t\widehat{\mathbf{R}}_{\mathbf{xx},t}\mathbf{W}_t^H = \mathbf{I}$ for the sequence of estimates $\widehat{\mathbf{R}}_{\mathbf{xx},t}$ generated by (18).*

Proof: The proof can be constructed using the Householder RLS relations in [10] with the additional fact that $\boldsymbol{\Theta}_t$ is unitary such that $\boldsymbol{\Theta}_t\mathbf{W}_t\mathbf{R}_{\mathbf{xx},t}\mathbf{W}_t^H\boldsymbol{\Theta}_t^H = \boldsymbol{\Theta}_t\boldsymbol{\Theta}_t^H = \mathbf{I}$ if and only if $\mathbf{W}_t\mathbf{R}_{\mathbf{xx},t}\mathbf{W}_t^H = \mathbf{I}$. $\circ$

Table 1 gives the proposed adaptive strong-uncorrelating transform based on the method of orthogonal iterations. The complexity of this procedure is $2.5m^3+\mathcal{O}(m^2)$ complex operations at each time step. By comparison, the exact SUT in [3] requires the eigenvalue decomposition of $\widehat{\mathbf{R}}_{\mathbf{xx},t}$ which is at least of $\mathcal{O}(9m^3)$ complexity [5], and the symmetric SVD of $\overline{\mathbf{P}}_t$ using an iterative procedure with specialized numeric code. Clearly, the adaptive SUT in Table 1 is much simpler to implement in an on-line setting.

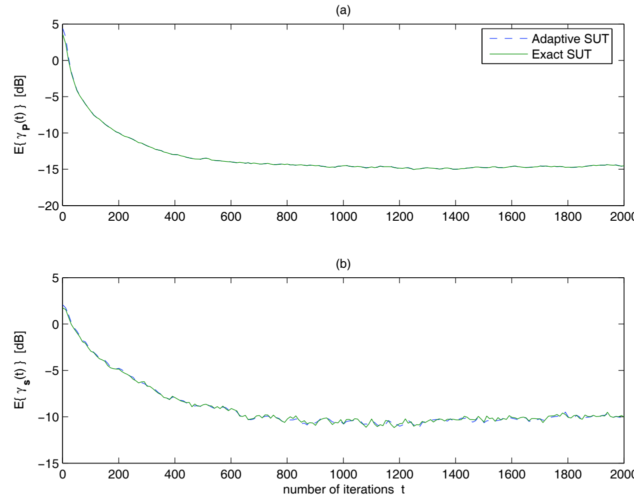

Fig. 1: Evolutions of (a) $E\{\gamma_{\mathbf{P}}(t)\}$ and (b) $E\{\gamma_{\mathbf{s}}(t)\}$ in the first simulation example.

4. SIMULATIONS

We now illustrate the performance of the proposed algorithms via simulations, in which

$$\mathbf{x}(t) \;=\; \mathbf{A}\mathbf{s}(t)+\boldsymbol{\nu}(t), \qquad (24)$$

where the (6×6) matrix $\mathbf{A}$ has complex jointly-Gaussian entries, $\mathbf{s}(t)$ contains realizations of zero-mean uncorrelated complex jointly Gaussian non-circular random variables with $E\{\mathbf{s}(t)\mathbf{s}^H(t)\} = \mathbf{I}$ and $E\{\mathbf{s}(t)\mathbf{s}^T(t)\} = \boldsymbol{\Lambda}$ with $\{\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6\} = \{1, 0.8, 0.6, 0.4, 0.2, 0.1\}$, and $\boldsymbol{\nu}(t) = \mathbf{0}$. Because $E\{|s_i(t)|^2\} = 1$ and $\lambda_i \neq \lambda_j$ for all $i \neq j$, the SUT of $\mathbf{x}(t)$ yields $\mathbf{W} = \mathbf{A}^{-1}$ if $\mathbf{R}_{\mathbf{xx},t}$ and $\mathbf{P}_{\mathbf{xx},t}$ are known exactly. Both the adaptive SUT and an exact SUT as described in [3] have been applied to 100 different realizations of this data, where $\alpha = 1-\rho = 0.995$, $\mathbf{W}_0 = \sqrt{10}\mathbf{I}$, and $\widehat{\mathbf{R}}_{\mathbf{xx},t} = \widehat{\mathbf{P}}_{\mathbf{xx},t} = 0.1\mathbf{I}$. Shown in Fig. 1 are the evolutions of

$$\gamma_{\mathbf{P}}(t) \;=\; \frac{||\mathbf{C}_t\boldsymbol{\Lambda}\mathbf{C}_t^T-\mathrm{diag}[\mathbf{C}_t\boldsymbol{\Lambda}\mathbf{C}_t^T]||_F^2}{||\mathrm{diag}[\mathbf{C}_t\boldsymbol{\Lambda}\mathbf{C}_t^T]||_F^2} \qquad (25)$$

$$\gamma_{\mathbf{s}}(t) \;=\; \frac{1}{2m}\left(\sum_{i=1}^{n}\sum_{l=1}^{n}\frac{|c_{ilt}|^2}{\max\limits_{1\le i\le n}|c_{ilt}|^2}+\frac{|c_{ilt}|^2}{\max\limits_{1\le l\le n}|c_{lit}|^2}\right)-1 \qquad (26)$$

with $\mathbf{C}_t = \mathbf{W}_t\mathbf{A}$, which measure the degree of pseudo-covariance diagonality and the separation quality, respectively, of $\mathbf{W}_t$. The adaptive SUT gives nearly identical performance to an exact SUT of the estimated $\widehat{\mathbf{R}}_{\mathbf{xx},t}$ and $\widehat{\mathbf{P}}_{\mathbf{xx},t}$, respectively.

5. USING THE STRONG-UNCORRELATION TRANSFORM IN SUBSPACE TRACKING

The goal of subspace tracking is to calculate an $(m\times n)$, $m < n$ matrix $\mathbf{C}$ whose rows span the m-dimensional principal or minor subspace of $\mathbf{x}(k)$. For principal subspace tracking, we

$$
\begin{aligned}
&\mathbf{v}(t) = \mathbf{C}_{t-1}\mathbf{x}(t)\\
&\mathbf{e}(t) = \mathbf{x}(t) - \mathbf{C}_{t-1}^H\mathbf{v}(t)\\
&\mathbf{y}(t) = \mathbf{W}_{t-1}\mathbf{v}(t)\\
&\mathbf{u}(t) = \mathbf{W}_{t-1}^H\mathbf{y}(t)\\
&\zeta(t) = \frac{1}{\mathbf{y}^H(t)\mathbf{y}(t)}\left(1 - \sqrt{\frac{\alpha}{\alpha + \mathbf{y}^H(t)\mathbf{y}(t)}}\right)\\
&\mathbf{k}(t) = \frac{\mathbf{u}(t)}{\alpha + \mathbf{y}^H(t)\mathbf{y}(t)}\\
&\mathbf{z}(t) = \mathbf{e}(t) - \mathbf{C}_{t-1}^H\left[0.5||\mathbf{e}(t)||^2\mathbf{k}(t)\right]\\
&\mathbf{C}_t = \mathbf{C}_{t-1} + \frac{\mathbf{k}(t)\mathbf{z}^H(t)}{1 + 0.25||\mathbf{k}(t)||^2||\mathbf{e}(t)||^2}\\
&\widehat{\mathbf{P}}_{\mathbf{xx},t} = \alpha\widehat{\mathbf{P}}_{\mathbf{xx},t-1} + \mathbf{v}(t)\mathbf{v}^T(t)\\
&\overline{\mathbf{P}}_t = \mathbf{W}_{t-1}\widehat{\mathbf{P}}_{\mathbf{xx},t}\mathbf{W}_{t-1}^T\\
&\mathbf{\Theta}_t\overline{\mathbf{R}}_t = \mathtt{qr}[\overline{\mathbf{P}}_t + \beta\mathbf{I}]\\
&\mathbf{W}_t = \frac{1}{\sqrt{\alpha}}\mathbf{\Theta}_t\left(\mathbf{W}_{t-1} - \zeta(t)\mathbf{y}(t)\mathbf{u}^H(t)\right)
\end{aligned}
$$

Table 1. The OPAST-SUT algorithm.

desire $\mathbf{C}\mathbf{e}_i \approx 0$ for $i \in \{m+1, \ldots, n\}$, whereas for minor subspace tracking, we desire $\mathbf{C}\mathbf{e}_i \approx 0$ for $i \in \{1, \ldots, n-m\}$. The rest of this paper focuses on the former problem.

Among the many principal subspace tracking algorithms, the orthogonal PAST algorithms are a family of procedures use an approximate least-squares subspace estimator embedded within a Householder update to maintain exact unitarity of $\mathbf{C}_t$ [11, 12]. All PAST algorithms calculate the Kalman gain vector within the subspace for the update relations.

The strong-uncorrelating transform complements the orthogonal PAST algorithm family. In particular, [(16)–(17), (22)–(23)] exactly represents $\widehat{\mathbf{R}}_{\mathbf{xx},t}^{-1}$ in $\mathbf{W}_t^H\mathbf{W}_t$, so it can replace the traditional Kalman gain update relationships within PAST. A similar design was used in the combined OPAST and kurtosis-based source separation procedure in [13]. Table 1 lists the update equations for the proposed OPAST-SUT adaptive procedure. The overall complexity of this technique is $3mn + 2.5m^3 + \mathcal{O}(m^2)$ operations at each sample time.

Our OPAST-SUT algorithm is mathematically equivalent yet simpler to implement than the disjoint combination of an orthogonal PAST algorithm and an adaptive SUT procedure. It possesses additional signal partitioning capabilities beyond normal subspace tracking. Consider a multiantenna communication example where different modulation schemes may be used adaptively depending on the link quality. The receiver has six antennas, and a mixture of two BPSK signals and one 16-QAM signal is observed. The channel matrix is

$$
\mathbf{A}^T = \left[\begin{array}{cccccc}
0.9+j0.4 & -0.3-j0.7 & 0.8+j0.9 & 0.1+j0.3 & -0.9-j0.2 & 0.2+j0.1\\
0.9+j0.6 & 0.5-j0.6 & -0.6-j0.1 & -0.2-j0.2 & -0.1+j0.4 & -0.7-j0.9\\
-0.3-j0.8 & 0.7-j0.7 & -0.2-j0.1 & -0.9+j0.4 & 0.3+j0.5 & -0.5+j0.4
\end{array}\right] \quad (27)
$$

and $\boldsymbol{\nu}(t)$ contains uncorrelated complex circular jointly-Gaussian noise with power $E\{|\nu_i(t)|^2\} = 0.01$. The circularity coefficients for the three signals are $\{\lambda_1, \lambda_2, \lambda_3\} = \{1, 1, 0\}$. A six-input, three-output combined OPAST-SUT algorithm both estimates the principal subspace *and* partitions this subspace according to the circularity coefficient values, such that the first two outputs in $\mathbf{y}(t)$ contain mixtures of the BPSK sources and the third output in $\mathbf{y}(t)$ contains the one 16-QAM source. Figure 2 shows the output signal constellations of $y_1(t)$, $y_2(t)$, and $y_3(t)$ for $t \in \{1001, \ldots, 2000\}$ with the OPAST-SUT algorithm applied to data from this model with $\mathbf{W}_0 = 10\mathbf{I}$ and $\alpha = 0.99$. The first two outputs show (phase-rotated) mixtures of the two real-valued BPSK sources, and the last output contains the 16-QAM source.

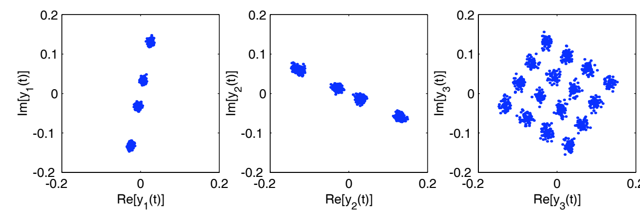

Fig. 2: Signal outputs generated by the OPAST-SUT algorithm indicating source clustering.

6. CONCLUSIONS

In this paper, we have systematically derived adaptive procedures for estimating the strongly-uncorrelating transform using modifications of classic iterations from eigenvalue analysis. The techniques are computationally- and numerically-efficient. Finally, a novel OPAST-SUT algorithm is described that clusters signal subspaces based on source non-circularity.

7. REFERENCES

[1] P. Ciblat and L. Vandenthorpe, "Blind carrier frequency offset estimation for non-circular constellation-based transmissions," *IEEE Trans. Signal Processing,* vol. 51, pp. 1378-1389, May 2003.

[2] D. Darsena, G. Gelli, L. Paura, and F. Verde, "Widely linear equalization and blind channel identification for interference-contaminated multicarrier systems," *IEEE Trans. Signal Processing,* vol. 53, pp. 1163-1177, Mar. 2005.

[3] J. Eriksson and V. Koivunen, "Complex-valued ICA using second order statistics," *Proc. IEEE Workshop on Machine Learning Signal Processing,* Sao Luis, Brazil, pp. 183-191, Oct. 2004.

[4] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Trans. Inform.Theory,* in press.

[5] G.H. Golub and C.F. Van Loan, *Matrix Computations,* 3rd ed. (Baltimore: Johns Hopkins, 1996).

[6] A. Björck and C. Bowie, "An iterative algorithm for computing the best estimate of an orthogonal matrix," *SIAM J. Numer. Anal.,* vol. 8, no. 2, pp. 358-364, June 1971.

[7] S.C. Douglas and A. Cichocki, "Neural networks for blind decorrelation of signals," *IEEE Trans. Signal Processing,* vol. 45, pp. 2829-2842, Nov. 1997.

[8] T. Chen and Q. Lin, "Dynamic behavior of the whitening process," *IEEE Signal Processing Lett.,* vol. 5, pp. 25-26, Jan. 1998.

[9] A.A. Rontogiannis and S. Theodoridis, "Inverse factorization adaptive least-squares algorithms," *Signal Processing,* vol. 52, no. 1, pp. 35-47, July 1996.

[10] S.C. Douglas, "Numerically-robust $\mathcal{O}(N^2)$ RLS algorithms using least-squares prewhitening," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Istanbul, Turkey, vol. I, pp. 412-415, June 2000.

[11] K. Abed-Meraim, A. Chkeif, and Y. Hua, "Fast orthogonal PAST algorithm," *IEEE Signal Processing Lett,* vol. 7, pp. 60-62, Mar. 2000.

[12] S.C. Douglas and X. Sun, "Designing orthonormal subspace tracking algorithms," *Proc. 34th Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, vol. 2, pp. 1441-1445, Nov. 2000.

[13] S.C. Douglas, "Combining subspace tracking, prewhitening, and contrast optimization for noisy blind signal separation," *Proc. 2nd Workshop Independent Compon. Anal. Signal Separation,* Espoo, Finland, pp. 579-584, June 2000.