Blind Separation of Noisy Multivariate Data Using Second-Order Statistics: Remote-Sensing Applications

Blind Separation of Noisy Multivariate Data Using Second-Order Statistics: Remote-Sensing Applications

Keith T. Herring, Amy V. Mueller, and David H. Staelin, *Life Fellow, IEEE*

***Abstract*—In this paper a second-order method for blind source separation of noisy instantaneous linear mixtures is presented for the case where the signal order k is unknown. Its performance advantages are illustrated by simulations and by application to Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) multichannel visible/infrared data. The model assumes that m mixtures $\mathbf{x}$ of dimension n are observed, where $\mathbf{x} = \mathbf{A}\mathbf{p} + \mathbf{G}\mathbf{w}$, and the underlying signal vector $\mathbf{p}$ has $k \lesssim n/3$ independent unit-variance elements. $\mathbf{A}$ is the mixing matrix, $\mathbf{G}$ is diagonal, and $\mathbf{w}$ is a normalized white-noise vector. The algorithm estimates the Second-Order separation matrix $\mathbf{A}$, signal Order k, and Noise and is therefore designated as SOON. SOON first iteratively estimates k and $\mathbf{G}$ using a scree metric, singular-value decomposition, and the expectation-maximization algorithm, and then determines the values of $\mathbf{AP}$ and $\mathbf{W}$. The final step estimates $\mathbf{A}$ and the set of m signal vectors $\mathbf{p}$ using a variant of the joint-diagonalization method used in the Second-Order Blind Identification (SOBI) and Second-Order NonStationary (SONS) source-separation algorithms. The SOON extension of SOBI and SONS significantly improves their separation of simulated sources hidden in noise. SOON also reveals interesting thermal dynamics within AVIRIS multichannel visible/infrared imaging data not found by noise-adjusted principal-component analysis.**

***Index Terms*—Blind signal separation (BSS), estimation, image representation, remote sensing, separation.**

I. Introduction

BLIND signal separation (BSS) has recently been applied to diverse research areas such as remote sensing, telecommunications, neurobiology, and speech processing. For example, when classifying regions and phenomena in multispectral images, it is often useful to first transform the original spectral data to a more compact and physically significant representation using BSS [1]–[3]. The BSS problem generally involves estimating a set of source signals when only their mixtures are observed. For the case of interest here, the observed vectors $\mathbf{x}$ of dimension n are noisy instantaneous linear mixtures of $k \lesssim n/3$ sources. The problem can be characterized by the following system:

$$\mathbf{X} = \mathbf{AP} + \mathbf{GW} \tag{1}$$

where $\mathbf{X}$ is an $n \times m$ matrix of n parameters observed m times, $\mathbf{A}$ is an $n \times k$ mixing matrix, and $\mathbf{P}$ is a $k \times m$ matrix of k underlying signals sampled m times. $\mathbf{GW}$ is a noise matrix where it is assumed that $\mathbf{W}$ contains n white-noise vectors and that the noise covariance matrix $\mathbf{GG}^{\mathrm{H}}$ is diagonal with generally unequal diagonal elements. Any noise correlated across multiple instantaneous observations is considered to be an additional source signal. We also adopt the standard assumptions that the mixing matrix $\mathbf{A}$ is of full column rank and that $\mathbf{P}$ and $\mathbf{W}$ have zero mean and unit variance so as to avoid any ambiguity in the separation. The problem explored in this paper involves estimating $\mathbf{A}$, k, $\mathbf{P}$, $\mathbf{G}$, and $\mathbf{W}$ using only knowledge of $\mathbf{X}$.

The remainder of this introduction summarizes the contributions of this paper relative to some popular BSS methods. Section II describes the Second-Order estimates of signal Order k and Noise $\mathbf{GW}$ (SOON) algorithm, Section III compares SOON's performance to that of other separation algorithms operating on simulated Gaussian data sets, and Section IV illustrates how SOON reveals interesting physical phenomena in an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) multispectral image.

Although many BSS methods take advantage of higher order statistics that distinguish the underlying sources [4]–[6], multispectral environments often exhibit sufficient second-order diversity that second-order separation methods exploiting it are desired. One such method for separating Gaussian sources is factor analysis [7], which assumes that the order k is known in order to calculate prior probabilities for the unobserved signals. Bayesian BSS [8] assumes that k is unknown, although its probability distribution is known. The Second-Order Blind Identification (SOBI) algorithm separates sources by taking advantage of any spectral diversity among the source signals [9], and the Second-Order NonStationary (SONS) algorithm improves SOBI by also utilizing any second-order nonstationarity present within the sources [10]; both assume that k is known. The algorithm presented in this paper uses both nonstationarity and spectral diversity to separate signals, but first produces second-order estimates of signal order k and noise

Manuscript received June 17, 2008; revised January 13, 2009 and March 23, 2009. First published June 26, 2009; current version published September 29, 2009. This work was supported in part by the MIT Lincoln Laboratory under Contract BX-8463 and in part by the Department of the Air Force under Contract F19628-00-C-0002.

K. T. Herring and D. H. Staelin are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139-4307 USA (e-mail: kherring@alum.mit.edu; staelin@mit.edu).

A. V. Mueller is with the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139-4307 USA (e-mail: amym@mit.edu).

Digital Object Identifier 10.1109/TGRS.2009.2022325

$\mathbf{GW}$ and is designated as "SOON" [11]. $\mathbf{GW}$ and k are jointly estimated iteratively using both a scree-plot method [12], [13] and the estimation-maximization (EM) algorithm. Most standard alternative methods for estimating $\mathbf{GW}$, such as Weiner and Kalman filtering and maximum-likelihood (ML) covariance-matrix estimation, are discussed or referenced within the BSS references cited earlier.

SOON then computes a noise-reduced estimate of the mixing matrix $\mathbf{A}$ and signal set $\mathbf{P}$ by employing an improved form of the joint-diagonalization method used in the SOBI and SONS algorithms. SOON's sensitivity to diversity and nonstationary spatial correlations of spectral features helped it separate, for example, the varied thermal effects in the sample AVIRIS multispectral image presented later. The contribution of this paper thus lies in demonstrating and quantifying SOON's signal-separation capabilities for real and simulated multispectral data and in definition of SOON, which combines iterative estimation of the unknown signal order k and noise matrix $\mathbf{G}$ with a modification of SONS that better captures information in nonstationary data sets by first SNR weighting the data correlation matrices before jointly diagonalizing them for chosen time lags and image regimes. The blind separation of solar influences in the AVIRIS spectral data is also believed to be novel.

II. Second-Order Separation and Order Noise (SOON) Estimation

SOON extends SONS [10], which is an extension of SOBI [9]. SOBI is a second-order source-separation method that capitalizes on source spectral diversity. SOBI first prewhitens the observation matrix $\mathbf{X}$ using a whitening transform $\mathbf{H}$ found through an eigenvalue decomposition of the sample covariance matrix of $\mathbf{X}$. The final step of SOBI finds a unitary transform $\mathbf{U}$ that jointly diagonalizes a set of sample covariance matrices $\mathbf{R_{HX}}$ time lagged by τ and calculated from the whitened data $\mathbf{HX}$, where[1]

$$\mathbf{R_{HX}}(\tau) = \mathbf{U}\mathbf{R_P}(\tau)\mathbf{U}^H \quad \forall \tau \neq 0. \tag{2}$$

As pointed out in [9], it follows from the spectral theorem for normal matrices that the existence of a unitary matrix $\mathbf{V}$ is guaranteed such that, for a nonzero time lag τ

$$\mathbf{V}^{\mathrm{H}}\mathbf{R_{HX}}(\tau)\mathbf{V} = \mathrm{diag}\{d_1, \ldots, d_k\}. \tag{3}$$

However, in order for $\mathbf{U}$ to be obtained directly from $\mathbf{V}$, the condition

$$\rho_i(\tau) \neq \rho_j(\tau) \quad \forall i \neq j \tag{4}$$

must hold, where

$$\rho_i(\tau) = E\left[\mathbf{P}_i(t+\tau)\mathbf{P}_i^*(t)\right] \tag{5}$$

and row vector $\mathbf{P}_i$ denotes the ith source; this is not always true for an arbitrary nonzero time lag τ. The joint-diagonalization step therefore seeks a unitary matrix $\mathbf{V}$ such that (2) holds for a set of nonzero time lags, $\{\tau_i | \mathrm{i} = 1, \ldots, L\}$. This increases the probability that $\mathbf{U}$ and $\mathbf{A}$ can be reasonably identified. An overview of SOBI appears in Table I.

[1]The superscript H on a matrix signifies that it is Hermetian, and $*$ signifies the transpose.

TABLE I
SOBI Algorithm

Step	
1.	Calculate the sample covariance matrix, $\hat{\mathbf{R}}_{\mathrm{X}}(0)$, of $\mathbf{X} = \mathbf{AP} + \mathbf{GW}$ $\Rightarrow \hat{\mathbf{R}}_{\mathrm{X}}(0) \approx \mathbf{A}\mathbf{R}_p(0)\mathbf{A}^H + \mathbf{G}\mathbf{R}_W(0)\mathbf{G}^H$ $\approx \mathbf{A}\mathbf{A}^H + \mathbf{G}\mathbf{G}^H$. Let $\lambda_1, \ldots, \lambda_k$ denote the k largest eigenvalues and $\mathbf{h}_1, \ldots, \mathbf{h}_k$ the corresponding eigenvectors of $\hat{\mathbf{R}}_{\mathrm{X}}(0)$.
2.	Assuming white noise, an estimate $\hat{\sigma}^2$ of the noise variance is the average of the n - k smallest eigenvalues of $\hat{\mathbf{R}}_{\mathrm{X}}(0)$. Calculate the whitening matrix, $\mathbf{H}$, as $\mathbf{H} = \left[\sqrt{\lambda_1 - \hat{\sigma}^2}\,\mathbf{h}_1, \ldots, \sqrt{\lambda_k - \hat{\sigma}^2}\,\mathbf{h}_k\right]^{\mathrm{H}}$.
3.	Let the whitened data be denoted as $\mathbf{Z} = \mathbf{HX}$. Calculate the sample covariance matrices of $\mathbf{Z}$, $\hat{\mathbf{R}}_z(\tau)$, for a fixed set of time lags $\tau \in \{\tau_i \mid i = 1, \ldots, L\}$.
4.	Find $\mathbf{U}$ as the joint diagonalizer of the set $\{\hat{\mathbf{R}}_z(\tau_i) \mid i = 1, \ldots, L\}$, recalling $\hat{\mathbf{R}}_z(\tau_j) \approx \mathbf{U}\mathbf{R}_P(\tau_i)\mathbf{U}^T$.
5.	Estimate $\mathbf{A}$ and $\mathbf{P}$: $\hat{\mathbf{A}} = \mathbf{H}^{\#}\mathbf{U}$ (# indicates the Moore-Penrose pseudo-inverse) $\hat{\mathbf{P}} = (\mathbf{U}^T\mathbf{H})\mathbf{X}$.

It is often computationally desirable in the BSS prewhitening step to find a whitening transform $\mathbf{H}$ such that $\mathbf{HA}$ is a unitary matrix, i.e.,

$$(\mathbf{HA})(\mathbf{HA})^{\mathrm{H}} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}. \tag{6}$$

The search space for the mixing transform can then be reduced to that of unitary matrices. SOBI assumes that the noise energy is either known or is approximately equal for each parameter so that it can be estimated by averaging the smallest eigenvalues of the data covariance matrix. If this equal-energy assumption does not hold, which is commonly the case, the whitening step will be biased by noise. More specifically, $\mathbf{HA}$ will generally not be orthogonal for the calculated whitening transform $\mathbf{H}$; this can degrade separation performance for low SNR data as the exact new mixing transform $\mathbf{HA}$ may not be contained in the search space. In addition, if any sources happen to be nonstationary, which is common for real data sets, the performance of SOBI can suffer.

The second-order separation method, SONS, extends the capabilities of SOBI. First, any noise bias is reduced through a robust whitening procedure in which the zero-time-lag covariance matrix is replaced by a linear combination of time-delayed data covariance matrices when whitening the data. This involves starting first with an arbitrary weight vector α for the

TABLE II
SONS ALGORITHM

Step	
1.	Calculate the set of sample covariance matrices, $\{\hat{\mathbf{R}}_{\mathrm{X}}(\tau_i) \mid i = 1, ..., J\}$ where $\mathbf{X} = \mathbf{AP} + \mathbf{GW}$. Choose any non-zero initial vector of weights $\alpha = [\alpha_1, ...\alpha_j]$. Use the global convergence algorithm to iteratively update α until $\mathbf{R} = \alpha_1 \hat{\mathbf{R}}_{\mathrm{X}}(\tau_1) + ... + \alpha_j \hat{\mathbf{R}}_{\mathrm{X}}(\tau_j)$ is positive definite. Let $\lambda_1, ..., \lambda_k$ denote the **k** largest eigenvalues, and $\mathbf{h}_1, ..., \mathbf{h}_k$ denote the corresponding eigenvectors of $\mathbf{R}$.
2.	Calculate the whitening matrix, $\mathbf{H}$, as $\mathbf{H} = \left[\sqrt{\lambda_1}\,\mathbf{h}_1, ..., \sqrt{\lambda_k}\,\mathbf{h}_k\right]^{\mathrm{H}}$.
3.	Let the whitened data be denoted as $\mathbf{Z} = \mathbf{HX}$. Divide $\mathbf{Z}$ into L non-overlapping blocks (time windows T_i) and calculate the sample covariance matrices $\hat{\mathbf{R}}_Z(T_i, \tau_j)$ for $i = 1, \ldots, L$ and $j = 1, ..., M$.
4.	Find the unitary matrix $\mathbf{U}$ that jointly diagonalizes the set $\left\{\hat{\mathbf{R}}_Z(T_i, \tau_j)\right\}$
5.	Estimate $\mathbf{A}$ as: $\hat{\mathbf{A}} = \mathbf{H}^{\#}\mathbf{U}$ (# indicates the Moore-Penrose pseudo-inverse).

TABLE III
EM ALGORITHM

Step	
1.	Initialization: Iteration index $j = 1$; number of iterations j_{max} $\mathbf{GG}_1^H = 0.5\mathbf{I}$ $\mathbf{A}_1$ = {unit-variance n x $\hat{k}_i$ array of Gaussian random noise}.
2.	Expectation Step: $\mathbf{S}_j = \mathbf{A}_j^H \left(\mathbf{GG}_j^H\right)^{-1} \mathbf{A}_j + \mathbf{I}$ $\mathbf{C}_j = E\left[\mathbf{P}^{\mathrm{H}} \mid \mathbf{X}, \mathbf{A}_j, \mathbf{G}_j\right] = \mathbf{X}^H \left(\mathbf{GG}_j^H\right)^{-1} \mathbf{A}_j \mathbf{S}_j^{-1}$ $\mathbf{D}_j = E\left[\mathbf{PP}^{\mathrm{H}} \mid \mathbf{X}, \mathbf{A}_j, \mathbf{G}_j\right]$ $= m\mathbf{S}_j^{-1} + \mathbf{S}_j^{-1}\mathbf{A}_j^H \left(\mathbf{GG}_j^H\right)^{-1} \mathbf{XX}^H \left(\mathbf{GG}_j^H\right)^{-1} \mathbf{A}_j \mathbf{S}_j^{-1}$.
3.	Maximization Step: $\mathbf{A}_{j+1} = \mathbf{XC}_j \mathbf{D}_j^{-1}$ $G_{q,j+1} = \left[\left(\mathbf{X}^H\right)_q^H \left(\mathbf{X}^H\right)_q - \mathbf{A}_{q,j+1}\mathbf{C}_j^H \left(\mathbf{X}^H\right)_q\right] \frac{1}{m}$; where $q = 1, \ldots, n$ indexes the columns of $\mathbf{X}^T$ and $\mathbf{A}$; G_q is the q^{th} diagonal element of $\mathbf{GG}^H$.
4.	If $j < j_{max}$, increment j and go to Step 2 for further iteration.
5.	After last iteration, output $\mathbf{G}_{i+1}$ and $\hat{\mathbf{X}}_{i+1} = \left[\mathbf{X}^H \left(\mathbf{GG}_{j+1}^H\right)^{-1} \mathbf{A}_{j+1}\mathbf{S}_{j+1}^{-1}\right]\mathbf{A}_{j+1}^H$.

coefficients of the sample covariance matrices used in the linear combination, i.e.,

$$\mathbf{R} = \alpha_1 \hat{\mathbf{R}}_{\mathrm{X}}(\tau_1) + \cdots + \alpha_J \hat{\mathbf{R}}_{\mathrm{X}}(\tau_J), \quad \tau_i \neq 0 \,\forall\, i \qquad (7)$$

where

$$\hat{\mathbf{R}}_{\mathrm{X}}(\tau_i) \approx \mathbf{A}\mathbf{R}_{\mathbf{P}}(\tau_i)\mathbf{A}^{\mathrm{H}} \qquad (8)$$

follows from the white-noise assumption. However, $\mathbf{R}$ is not necessarily positive definite, and the whitening transform $\mathbf{R}^{-1/2}$ may not be valid. Therefore, the next step of the procedure is to use the finite-step global-convergence algorithm [14] to adapt the initial weight vector in such a way that $\mathbf{R}$ becomes positive definite. This improves performance by increasing the likelihood that $\mathbf{HA}$ will be approximately calculated as a unitary matrix. Additionally, SONS enables separation of second-order non-stationary sources; the data are partitioned in time, and sample covariance matrices are calculated separately for each partition. This new set of covariance matrices is then used in the same joint-diagonalization step implemented in SOBI. The SONS algorithm appears in Table II.

One potential problem associated with SONS is that if the data are partitioned arbitrarily in time or space rather than at boundaries where the statistics change, then the poorer sample statistics will degrade the source-separation performance. Methods to detect statistical boundaries can be used [15]. In addition, the whitening procedure can be quite sensitive to the chosen initial weight vector α. For these data sets, the separation performance may be degraded by uncharacterized noise, as is the case with SOBI.

The SOON algorithm iteratively estimates both the number k of sources and the noise covariance matrix, which is used to improve the SNR of the data matrix. In particular, SOON uses a scree metric and singular-value decomposition (SVD) to estimate k, and then the EM algorithm to estimate the noise covariance matrix $\mathbf{GG}^{\mathrm{H}}$. The consecutively improved estimates of k and $\mathbf{G}$ improve the estimate of each parameter. The EM algorithm is summarized in Table III.

For the trivial case in which the noise covariance matrix $\mathbf{GG}^{\mathrm{H}}$ is an identity matrix multiplied by a known scalar, the minimum eigenvalue of $\mathbf{R}_{\mathrm{X}}$ is replicated $n - k$ times, where k equals the number of sources. However, the noise covariance is generally not a scaled identity matrix, and k cannot be determined by counting, since the order estimate is calculated using a finite sample covariance matrix $\hat{\mathbf{R}}_{\mathrm{X}}$ with typically distinct eigenvalues. Information-theoretic approaches to the estimation of k have been studied in [16] and [17], and other methods in [18] and [19]. In each, however, it is assumed that either $\mathbf{GG}^{\mathrm{H}}$ is a scaled identity matrix or that $\mathbf{GG}^{\mathrm{H}}$ is known *a priori*.

Although SVD is commonly used to estimate the signal order k [9], [10] along with other approaches [1]–[5], simulations favor a more robust iterative method that combines SVD with a scree plot to estimate the order k and, then, uses the EM algorithm (described in Table III) together with the newly estimated signal order to estimate the noise $\mathbf{G}$, which iteratively improves the subsequent signal-order estimate until the loop terminates [12], [20]. The scree plot displays the ordered log-magnitude eigenvalues of the sample correlation matrix $\hat{\mathbf{R}}_{\mathrm{X}}$ for the data matrix $\mathbf{X}$ as a function of eigenvector number [12]. Fig. 1 shows a typical scree plot for simulated data having the

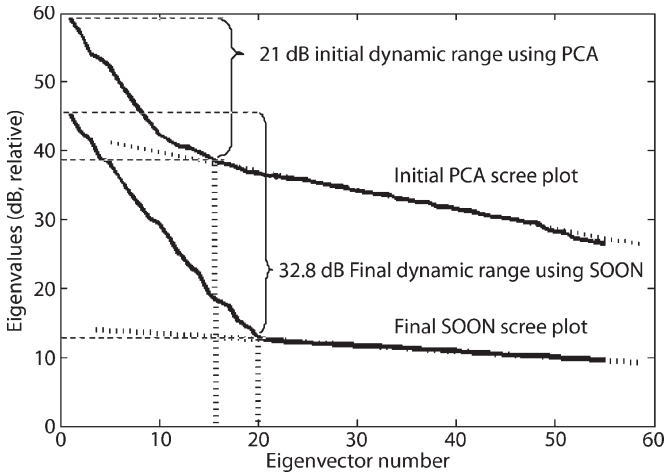

Fig. 1. Typical scree plots displaying the ordered eigenvalues of the data matrix $\mathbf{X}$ at the start (PCA only) and conclusion of SOON, based on Table VI. Straight dotted lines fit the noise plateaus of the scree plots and vertical dotted lines mark the approximate order estimates. Horizontal lines indicate the resulting signal dynamic range above the noise floor before and after using SOON.

parameters listed in Table VI, where the plateau to the right of the break generally represents pure-noise components. The order k is estimated using the intersection between the vertical axis and the extrapolated straight line that best fits the central 20% of the plot. Thus, k should be less than approximately one-third of the number of eigenvalues, since the choice of segment that is fit can depend on the character of the data. If consecutive estimates of k have a derivative of opposite sign to that determined using SVD, then the more robust SVD value is substituted (which seldom happens), where the SVD estimate for k was defined in this paper as the number of eigenvalues that individually exceed 0.001% of the total variance of $\mathbf{X}$. The EM algorithm then uses this estimate of k to estimate $\mathbf{G}$, which can be used to yield an improved scree plot and an improved estimate of k. A few iterations of order and noise estimation generally suffice.

Eigenvalues corresponding to pure normalized noise have approximately equal amplitudes and form a plateau of length $n-k$, where n shown in Fig. 1 is 55. The estimated signal order $\hat{k}$ is the number of eigenvalues that lie above the extrapolated plateau, where $\hat{k}=16$ after principal-component analysis (PCA), and $\hat{k}=20$, the correct value, after use of SOON. The noise being estimated is the additive component on each channel that is uncorrelated with variations on any other channel, and it can include contributions from both sensor noise and the image; conversely, additive contributions correlated across two or more channels are considered by SOON to be signal, not noise.

The EM method estimates the noise variance $\mathbf{G}$ as follows. In [8], a variant of the EM method was derived to estimate the noise variances when the observation matrix $\mathbf{X}$ is temporally spatially white. As follows, we derive a variant of the EM algorithm for estimating these variances for data sets that may contain time and space correlations. In (1), $\mathbf{A}$ and $\mathbf{G}$ are unknown but fixed, whereas $\mathbf{P}$ and $\mathbf{X}$ are stochastic matrices. The goal of the EM method then is to find the ML estimates of both $\mathbf{A}$ and $\mathbf{G}$, given the stochastic parameters $\mathbf{X}$ and an assumed $\mathbf{P}$, i.e., to find $\mathbf{A}_{\mathrm{ML}}$ and $\mathbf{G}_{\mathrm{ML}}$ such that

$$\{\mathbf{A}_{\mathrm{ML}},\mathbf{G}_{\mathrm{ML}}\}=\max_{\mathbf{A},\mathbf{G}} f(x_1,\ldots,x_m,p_1,\ldots,p_m|\mathbf{A},\mathbf{G}) \quad (9)$$

where $f(x)$ is the probability distribution of x and

$$\mathbf{X}=[x_1 \quad x_2 \quad \cdots \quad x_m], \qquad x_i\,\varepsilon\, C^{nx1} \quad (10)$$

$$\mathbf{P}=[p_1 \quad p_2 \quad \cdots \quad p_m], \qquad p_i\,\varepsilon\, C^{kx1}. \quad (11)$$

Using Bayes' rule, (9) can be rewritten as

$$\{\mathbf{A}_{\mathrm{ML}},\mathbf{G}_{\mathrm{ML}}\}=\max_{\mathbf{A},\mathbf{G}} f(p_1,\ldots,p_m|\mathbf{A},\mathbf{G}) \times f(x_1,\ldots,x_m|p_1,\ldots,p_m,\mathbf{A},\mathbf{G}) \quad (12)$$

where $\mathbf{P}$ does not depend on $\{\mathbf{A},\mathbf{G}\}$, so $f(p_1,\ldots,p_m|\mathbf{A},\mathbf{G})$ can be omitted. Additionally

$$x_i=\mathbf{A}p_i+\mathbf{G}w_i \qquad \forall\, i \quad (13)$$

where w_i is the ith time-space sample of the stochastic matrix $\mathbf{W}$. Therefore, given $\mathbf{P}$, $\mathbf{A}$, and $\mathbf{G}$, the x_i's are independent, since they depend stochastically only on the w_i's, which are assumed independent. Thus, (9) can be rewritten as

$$\{\mathbf{A}_{\mathrm{ML}},\mathbf{G}_{\mathrm{ML}}\}=\max_{\mathbf{A},\mathbf{G}} f(x_1|p_1,\mathbf{A},\mathbf{G}),\ldots,f(x_m|p_m,\mathbf{A},\mathbf{G}) \quad (14)$$

$$=\max_{\mathbf{A},\mathbf{G}}\prod_{i=1}^{m}\frac{1}{\sqrt{(2\pi)^n\det(\mathbf{G}\mathbf{G})^{\mathrm{H}}}}\exp^{-B_1} \quad (15)$$

where

$$B_1=(x_i-\mathbf{A}p_i)^{\mathrm{H}}(\mathbf{G}\mathbf{G}^{\mathrm{H}})^{-1}(x_i-\mathbf{A}p_i)/2. \quad (16)$$

After taking the log of (15), we arrive at the final expression for the ML estimate of $\mathbf{A}$ and $\mathbf{G}$

$$\{\mathbf{A}_{\mathrm{ML}},\mathbf{G}_{\mathrm{ML}}\}=\min_{\mathbf{A},\mathbf{G}}\left(\sum_{i=1}^{n} m\log(G_i)+\frac{1}{G_i}B_2\right) \quad (17)$$

where

$$B_2=X_iX_i^{\mathrm{H}}-2X_i\mathbf{P}^HA_i^H+A_i\mathbf{P}\mathbf{P}^HA_i^H \quad (18)$$

A_i is the ith row of $\mathbf{A}$, X_i is the ith row of $\mathbf{X}$, and G_i is the ith diagonal element of $\mathbf{G}\mathbf{G}^{\mathrm{H}}$. To find the values of $\mathbf{A}$ and $\mathbf{G}$ that minimize (17), we take the partial derivative with respect to each, which results in the solutions shown in step 3 of Table III. The maximization equations include the terms $\mathbf{P}^{\mathrm{H}}$ and $\mathbf{P}\mathbf{P}^{\mathrm{H}}$, which are unknown, so that (17) cannot be directly minimized. The EM algorithm finds an approximation to the minimum of (17) by replacing these terms with their expectations given $\mathbf{X}$ and current estimates of $\mathbf{A}$ and $\mathbf{G}$. To obtain these expectations, we can first solve for the distribution $f(p_i|x_i,\mathbf{A},\mathbf{G})$

$$f(p_i|x_i,\mathbf{A},\mathbf{G})=\frac{f(x_i|p_i,\mathbf{A},\mathbf{G})f(p_i|\mathbf{A},\mathbf{G})}{2f(x_i|\mathbf{A},\mathbf{G})}. \quad (19)$$

Since x_i, p_i, and w_i are all jointly Gaussian

$$f(p_i|x_i,\mathbf{A},\mathbf{G})=N\left(\mathbf{S}^{-1}\mathbf{A}^{\mathrm{H}}(\mathbf{G}\mathbf{G}^{\mathrm{H}})^{-1}x_i\mathbf{S}^{-1}\right) \quad (20)$$

where

$$\mathbf{S} = \mathbf{A}^{\mathrm{H}}(\mathbf{G}\mathbf{G}^{\mathrm{H}})^{-1}\mathbf{A} + \mathbf{I}. \tag{21}$$

The expressions for $E[\mathbf{P}^{\mathrm{H}}|\mathbf{X}, \mathbf{A}, \mathbf{G}]$ and $E[\mathbf{P}\mathbf{P}^{\mathrm{H}}|\mathbf{X}, \mathbf{A}, \mathbf{G}]$ follow from (19) and complete the expectation step of the algorithm. The entire iterative EM algorithm can be seen in Table III. The scree-plot/SVD estimate of k and the EM algorithm are iterated together in steps 2–5 of the algorithm in Table III until k reaches its asymptote.

SOON further diverges from SONS in the way the initial weight vector is chosen in the robust whitening step. When calculating time-delayed sample covariance matrices for noisy data using finite sample windows, it is clearly best to make greater use of those matrices that have high SNR. SOON first calculates the set of sample covariance matrices

$$\left\{\hat{\mathbf{R}}_{\mathrm{X}}(\tau_{\mathrm{i}})|\mathrm{i} = 1, \ldots, \mathrm{J}\right\} \tag{22}$$

and then the energy of $\hat{R}_{\mathrm{X}}(\tau_{\mathrm{i}})$

$$E_i \equiv \sum_{j,k}\left[\hat{\mathbf{R}}_{\mathrm{X}}(\tau_i)_{jk}\right]\left[\hat{\mathbf{R}}_{\mathrm{X}}(\tau_i)_{jk}\right]^{*} \qquad \forall\, i. \tag{23}$$

The initial weight vector α in step 6 of SOON is then

$$\alpha = \{E_1, E_2, \ldots, E_J\}. \tag{24}$$

This weights the delayed covariance matrices proportional to their energy in the initial linear combination rather than arbitrarily choosing these weights. Since the measurement noise is white, it can be assumed that any sample noise in these estimated covariance matrices will have approximately equal energy over each matrix, i.e.,

$$\sum_{j,k}\left\|\mathbf{R}_{\mathrm{X}}(\tau_i)_{jk} - \hat{\mathbf{R}}_{\mathrm{X}}(\tau_i)_{jk}\right\| \approx C \qquad \forall\, i \tag{25}$$

for some constant C, where $\mathbf{R}_X(\tau_i)$ and $\hat{\mathbf{R}}_X(\tau_i)$ are the time-delayed covariance and sample covariance matrices of the data, respectively.

Like SONS, SOON yields improved covariance matrices by apt partitioning of the data $\mathbf{X}$ in step 8 prior to joint diagonalization. Proper partitioning of remote-sensing data can simply involve identification of visibly distinct regions in the images, as demonstrated in Section IV. Table IV provides a complete algorithm description of SOON.

The global-convergence and joint-diagonalization algorithms [9] in steps 6 and 10 can be briefly summarized. For global convergence, the vector α is updated by adding δ if $\mathbf{R}$ is not positive definite, where

$$\delta = \frac{[\mathbf{U}^{\mathrm{H}}\mathbf{R}_1\mathbf{U}, \ldots, \mathbf{U}^{\mathrm{H}}\mathbf{R}_M\mathbf{U}]^{\mathrm{H}}}{\|\mathbf{U}^{\mathrm{H}}\mathbf{R}_1\mathbf{U}, \ldots, \mathbf{U}^{\mathrm{H}}\mathbf{R}_M\mathbf{U}\|}. \tag{26}$$

Joint diagonalization of $\mathbf{U}$ and the set $\mathbf{R}_z(T_i, t_j)$ in step 10 minimizes $\sum_{i=1}^{\mathbf{M}}$ off $(\mathbf{U}^{\mathrm{H}}\mathbf{R}_i\mathbf{U})$ for the $\mathbf{M}$ sample covariance matrices, where the function off $(\mathbf{Q}) \equiv \sum_{i\neq j}|\mathbf{Q}_{ij}|^2$. Step 11 simply estimates the mixing matrix $\mathbf{A}$ and the k signal vectors using the equations given in Table IV.

TABLE IV
SOON ALGORITHM

Step	
1.	Optionally normalize rows of $\mathbf{X}$ to zero mean and unit variance. Initialize $\hat{\mathbf{G}}_0 = \mathbf{I}$.
2.	Noise-normalize $\mathbf{X}$: $\mathbf{X}_i = \hat{\mathbf{G}}_{i-1}^{-1}\mathbf{X}$. (i = iteration index)
3.	Estimate signal order k_i using a scree plot of $\mathbf{X}_i$ and SVD.
4.	Estimate $\mathbf{G}_i\mathbf{G}_i^H$ and noise-reduced data $\hat{\mathbf{X}}_i$ using the EM algorithm and k_i.
5.	Check for iteration termination conditions; if none, increment the index i and return to Step 2.
6.	Calculate the set $\{\hat{\mathbf{R}}_{\hat{\mathrm{X}}}(\tau_i)\}$ for i =1,...,L. Initialize whitening vector α according to energy of this set. Use the global convergence algorithm to iteratively update α until $\mathbf{R} = \alpha_1\hat{\mathbf{R}}_{\mathrm{X}}(\tau_1) + \ldots + \alpha_J\hat{\mathbf{R}}_{\mathrm{X}}(\tau_J)$ is positive definite. Let $\lambda_1, \ldots, \lambda_k$ denote the k largest eigenvalues, and $\mathbf{h}_1, \ldots, \mathbf{h}_k$ the corresponding eigenvectors of $\mathbf{R}$.
7.	Calculate the whitening matrix, $\mathbf{H}$, as $\mathbf{H} = \left[\sqrt{\lambda_1}\ \mathbf{h}_1, \ldots, \sqrt{\lambda_k}\ \mathbf{h}_k\right]^H$.
8.	Characterize the non-stationarity of $\hat{\mathrm{X}}$ and construct partitioning set $\{T_i\}$ for i =1,...,J accordingly.
9.	Calculate the sample covariance matrices $\hat{\mathbf{R}}_Z(T_i, \tau_j)$ for i = 1,...,L and j = 1,...,M.
10.	Find the unitary matrix $\mathbf{U}$ as the joint diagonalizer of the set $\{\hat{\mathbf{R}}_Z(T_i, \tau_j)\}$) using the joint diagonalization method.
11.	Estimate **A, P** as: $\hat{\mathbf{A}} = \mathbf{H}^{\#}\mathbf{U}$ (# indicates the Moore-Penrose pseudo-inverse) $\hat{\mathbf{P}} = \mathbf{U}^{\mathrm{H}}\mathbf{H}\hat{\mathbf{X}}$.

III. EVALUATION OF SOON

To compare the performance of SOON to that of basic SOBI and SONS, we constructed two test data sets having 55 observations of 20 sources. The test mixing matrix $\mathbf{A}$ has normalized columns with 45° of separation between them, i.e.,

$$\mathbf{A}_i^{\mathrm{H}}\mathbf{A}_i = \mathbf{A}_j^{\mathrm{H}}\mathbf{A}_j \qquad \forall\, i, j \tag{27}$$

$$\frac{\mathbf{A}_i^{\mathrm{H}}\mathbf{A}_j}{\|\mathbf{A}_i\|\|\mathbf{A}_j\|} = \cos(45^\circ) \qquad \forall\, i \neq j. \tag{28}$$

Additionally, the noise variances were independently and randomly chosen from a normalized exponential distribution, i.e.,

$$f_{\mathbf{G}_{ii}}(x) = e^{-x}, \qquad x \geq 0 \tag{29}$$

where $f_{\mathbf{G}_{ii}}$ represents the distribution function of the noise variance for the ith observation.

In order to measure separation performance, we use a normalized inverse residual metric of the estimated mixing matrix

TABLE V
SONS VARIANTS

Variant Number	Initial Weight Vector	Partitioning Scheme
1	proportional to energy	1 segment
2	proportional to energy	10 equal size segments
3	equally weighted	1 segment
4	equally weighted	10 equal size segments

$\hat{\mathbf{A}}$ of $\mathbf{A}$, which we define as

$$\hat{r}_{\mathbf{A}} \equiv \frac{Tr\{\mathbf{A}_S\}}{Tr\{\mathbf{A}_N\}} \tag{30}$$

where

$$\mathbf{A}_S = E[\mathbf{A}\mathbf{A}^{\mathrm{H}}] \tag{31}$$

$$\mathbf{A}_N = E\left[(\mathbf{A}-\hat{\mathbf{A}})(\mathbf{A}-\hat{\mathbf{A}})^{\mathrm{H}}\right]. \tag{32}$$

Thus, $\hat{r}_{\mathbf{A}}$ is roughly a signal-to-noise ratio for $\hat{\mathbf{A}}$. The performance for each test was calculated for multiple values of SNR of the observation matrix $\mathbf{X}$. In each test, both SOBI and SONS were given the correct number of sources k, whereas SOON was not.

The first data set illustrates the performance of SOON on stationary time-correlated Gaussian sources. In this paper, the underlying signals are modeled using a standard autoregressive model

$$\mathbf{P}_{i,j} = \rho_i \mathbf{P}_{i,j-1} + \varepsilon_{i,j} \tag{33}$$

where $\mathbf{P}_{ij}$ is the jth time sample of the ith source signal, ρ_i is the correlation coefficient of the ith source, and $\varepsilon_{i,j}$ is a Gaussian random number. The ρ_i's were chosen independently and uniformly in (0,1). Using this test set, we compare the performance of SOON to that of SOBI and four variants of the SONS algorithm. This comparison suggests the utility of the noise-reduction aspects of SOON. As detailed in Table V, the initial weight vector was either proportional to energy, as described in Section II, or equally weighted; and the data were either not partitioned, which is optimal for this stationary case, or divided into ten equal segments. The results for the first test is shown in Fig. 2.

The second data set tested the performance of SOON on nonstationary white Gaussian sources with time-varying second moments that shift independently from one another at discrete Poisson-distributed points in time; the statistics are stable over the time intervals between these points. The time samples are

$$\mathbf{P}_{ijk} = \sigma_{ij}\varepsilon_k \quad f_{\sigma_{ij}} = e^{-x}, \qquad x > 0 \tag{34}$$

where the index i indicates the source, j indicates the time interval, k indicates the time sample within that interval, and f_σ indicates the unnormalized probability distribution of σ. In this paper, we adjusted the density of the transition points so that, on average, there would be ten segments with independent statistics.

Robust whitening is not applicable in this paper, since the underlying sources $\mathbf{P}_i$ are uncorrelated. We therefore compared

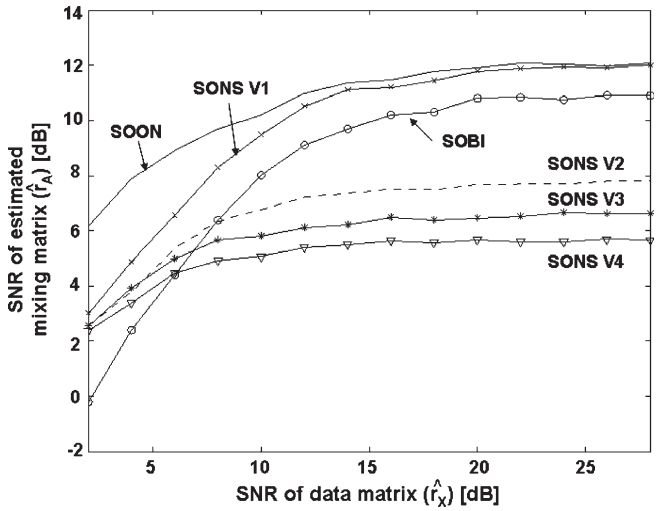

Fig. 2. Test 1 results. SNR ($\hat{r}_{\mathbf{A}}$, in decibels) for SOON, SOBI, and four variants of SONS as a function of $\hat{r}_{\mathrm{X}}$ for stationary sources with diverse spatial- and spectral-correlation functions.

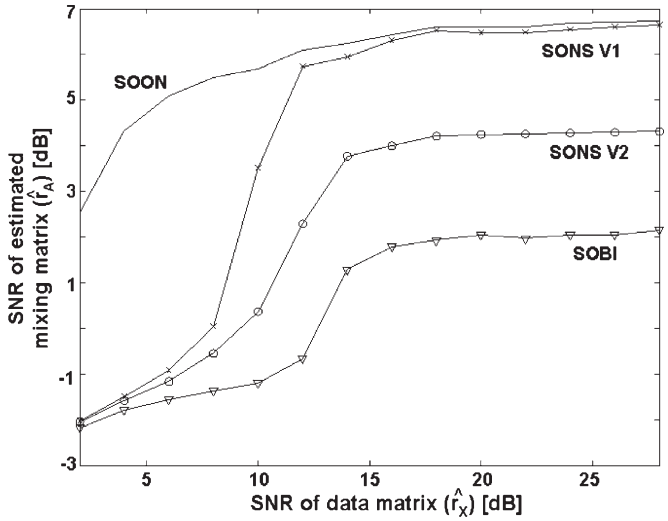

Fig. 3. Test 2 results. SNR ($\hat{r}_{\mathbf{A}}$, in decibels) for SOON, SOBI, and two variants of SONS for data sets comprised of white SONS sources.

SOON with SOBI and two variants of the SONS algorithm that each use traditional data whitening. For SONS variant 1 and SOON, the partitioning again matches exactly the nonstationarity of the data set, whereas the data in variant 2 are now partitioned into five equal-length segments. Fig. 3 shows the resulting SNR, $\hat{r}_{\mathbf{A}}$; SOON estimates the mixing transform better than both SOBI and SONS over the entire range of data SNR, particularly at low SNR values, as expected. This can be attributed to the inability of SONS to use robust whitening for nonstationary white data. This figure also illustrates how sensitive the SNR is to proper characterization of data nonstationarity; each successive improvement in data segmenting, from SOBI to SONS, significantly improves the separation. The sensitivity of the mixing matrix SNR to the data matrix SNR is greater in Fig. 3 than in Fig. 2 because the inaccuracies in the delayed sample covariance matrices used for Fig. 2 arise predominately from finite sample sizes, not SNR.

TABLE VI
NOMINAL TEST PARAMETERS

Parameter	Value
Number of parameters n	55
Signal order k	$0.36n$ ($k = 20$)
Number of observations m	10,000
SNR of $\mathbf{X}$	16 dB
Mean angle between columns of $\mathbf{A}$	$\angle_A = 45^\circ$

TABLE VII
SEPARATION AND ORDER ESTIMATION PERFORMANCE OF SOON ON COLORED STATIONARY GAUSSIAN TEST DATA (IN DECIBELS)

Exp.	Nom.	$k = 0.5n$	$m = 1000$	SNR_X = 3 dB	$\angle_A = 20^\circ$
$\hat{\mathbf{r}}_\text{G}$	37.3	19.8	32.4	32.7	33.7
$\hat{\mathbf{r}}_\text{W}$	4.4	3.1	4.4	4.5	4.2
$\hat{\mathbf{r}}_\text{A}$	11.9	11.8	5.4	6.5	10.1
$\hat{\mathbf{r}}_\text{P}$	10.9	10.1	3.7	3.3	5.3
μ_k	-0.002	-0.006	-0.002	-0.004	-0.002
σ_k	0.01	0.03	0.01	0.02	0.01

TABLE VIII
SEPARATION AND ORDER ESTIMATION PERFORMANCE OF SOON ON WHITE NONSTATIONARY GAUSSIAN TEST DATA (IN DECIBELS)

Exp.	Nom.	$k = 0.5n$	$m = 1000$	SNR_X = 3 dB	$\angle_A = 20^\circ$
$\hat{\mathbf{r}}_\text{G}$	36.8	19.8	31.9	32.2	33.9
$\hat{\mathbf{r}}_\text{W}$	4.4	3.1	4.5	4.3	4.6
$\hat{\mathbf{r}}_\text{A}$	6.2	6.6	2.2	3.3	6.3
$\hat{\mathbf{r}}_\text{P}$	3.7	4.5	0.3	0.8	0.5
μ_k	0	-0.008	-0.002	-0.002	-0.004
σ_k	0	0.04	0.01	0.01	0.02

The applicability of SOON to blind separation is illustrated further by the data SNR metric $\hat{r}_\mathbf{A}$ for estimates of $\mathbf{A}$, $\mathbf{P}$, $\mathbf{G}$, and $\mathbf{W}$ for both stationary and nonstationary sources. For both source types, we begin by assuming the nominal parameters given in Table VI. Each additional test case differs from Table VI in one parameter: the source-to-observation ratio (k/n), the number of data samples m, the data SNR $\hat{r}_\mathbf{A}$, or the angle between columns of the mixing transform $\angle_A$. The SNR $\hat{r}_\mathbf{A}$ for the estimates of $\mathbf{A}$, $\mathbf{P}$, $\mathbf{G}$, and $\mathbf{W}$ are given in Tables VII and VIII for colored stationary and white nonstationary Gaussian data as characterized by (33) and (34), respectively. Tables VII and VIII also present the mean errors μ_k and standard deviations σ_k for the estimated signal order k

$$\mu_k = E\left[\frac{k-k'}{k'}\right] \tag{35}$$

$$\sigma_k = \sqrt{E\left[\left(\frac{k-k'}{k'} - \mu_k\right)^2\right]} \tag{36}$$

where k' is the actual number of sources.

Tables VII and VIII show that, for the nominal test case, SOON estimates all parameters reasonably well. They also show that increasing the ratio k/n from 0.36 to 0.5 significantly degrades only the noise estimates $\mathbf{G}$, and that, it slightly improves the separation metric $\hat{r}_\text{P}$. Decreasing the SNR for the observations $\mathbf{X}$ significantly degrades only the separation performance, i.e., the estimates of $\mathbf{A}$ and $\mathbf{P}$. This is consistent with the degradation in separation found when the sample size m was decreased from 10 000 to 1000, thus decreasing the accuracy of the estimated covariance matrices used in the whitening and joint-diagonalization steps.

Decreasing the linear independence of the columns of the mixing transform $\mathbf{A}$ by reducing $\angle_\mathbf{A}$ from 45° to 20° primarily degraded the estimated source signals $\mathbf{P}$. These results are similar for both the stationary colored sources and the white nonstationary sources except that the separation SNR was better for the stationary case, as expected.

IV. APPLICATION TO REMOTE SENSING

Basis conversion of multispectral remote-sensing data is often used to help compress data or gain insight. Classic second-order methods include, for example, PCA, noise-adjusted PCA (NAPCA) [21], [22], SOBI [9], and SONS [10]. The simulations in Section III demonstrated that SOON can enhance these techniques. The potential value of this enhancement for remote sensing is illustrated in this section by using SOON for signal separation of an AVIRIS multispectral image without exploiting any of its higher order moments. Methods exploiting higher order moments [4], [5], [7] might perform better or worse, depending on the amplitudes of those moments relative to the statistical richness of the second-order moments.

PCA is a method for changing the linear coordinate system of a data set so that, for any j, the first j principal components (basis functions) span that j-dimensional subspace of the data having the greatest total variance. The problem is that variance alone does not distinguish between physical signals and noise, so that high-variance PCA components may have lower SNR than low-variance components. This problem motivated definition of NAPCA, which normalizes the noise variance for each input variable prior to PCA. The drawback with NAPCA is the same for SOBI and SONS: The noise covariance matrix must either be known or accurately estimated in order to normalize the noise. The principal advantage of SOON over PCA for characterizing physically significant multispectral features is its ability to estimate, normalize, and thereby mitigate both instrument and image noise, thus revealing more potentially useful underlying variables p_i.

This advantage is clearly revealed by 224-channel data collected over Moffet Field, CA, by the AVIRIS instrument. AVIRIS was developed and operated by the California Institute of Technology Jet Propulsion Laboratory [23], [24] and observes from ultraviolet to near infrared, i.e., wavelengths from 370 to 2500 nm. The chosen image in Fig. 4 shows natural terrain in its upper half and urban terrain in the lower half for seven representative bands uniformly distributed over the 224 channels. The image was manually partitioned into natural and urban zones along this boundary in step 8 of SOON.

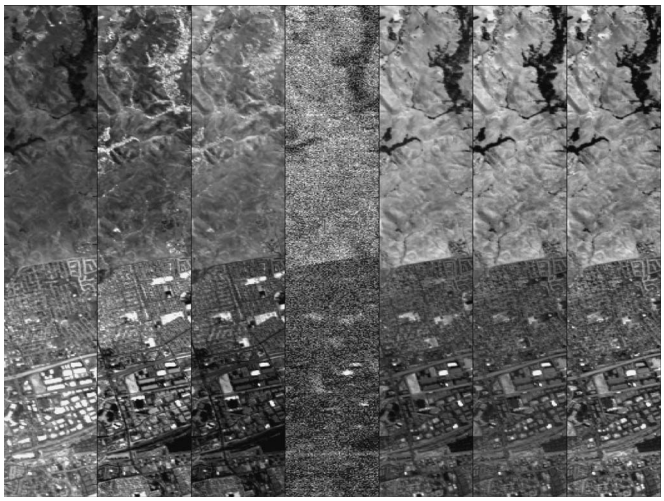

Fig. 4. Data collected over Moffet Field, CA, by AVIRIS. Channels 20, 50, 80, 110, 140, 170, and 200 are shown from left to right, in the order of decreasing wavelength.

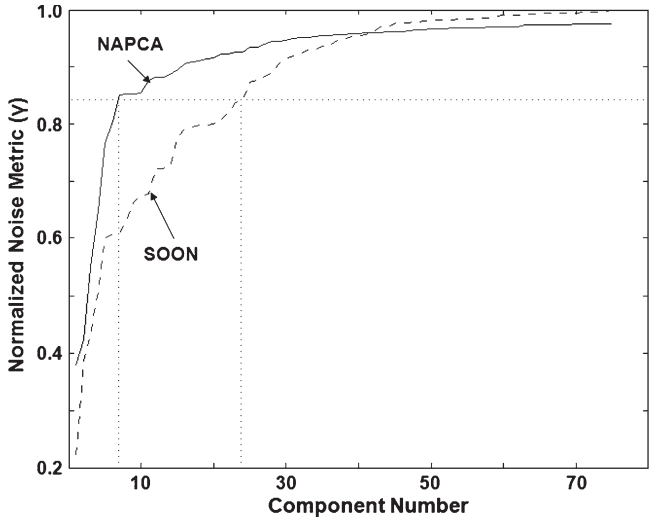

Fig. 5. Noise metric γ for each principal component found by SOON and NAPCA, sorted in ascending order.

Since only the observations and instrument-noise specifications are known for the AVIRIS image, a noise-estimation metric was used to compare SOON with the well-known benchmark NAPCA, where NAPCA was noise normalized using the noise covariance-matrix estimate provided by SOON. To compute the noise-estimation metric, each extracted principal-component image or $\mathbf{P}_i$ was rated by the degree to which it resembled white noise, where less noise suggests better separation. The chosen noise metric γ_i is the total energy in the highest spatial frequency components j, k of an energy-normalized $\mathbf{P}_i$ image and was defined using the 2-D spatial Fourier transform of $\mathbf{P}_i$

$$F_i \equiv \mathrm{FFT}(\mathbf{P}_i) \tag{37}$$

$$\gamma_i = \sum_{j,k:j+k>\tau} |\mathrm{F}_{ijk}| \tag{38}$$

where τ is the threshold for being considered high frequency. For images with transforms of dimensions x_1 and x_2, we define τ as

$$\tau = \min\left\{\frac{x_1}{2}, \frac{x_2}{2}\right\}. \tag{39}$$

This metric γ_i was evaluated for the full Moffet Field image, which is five times larger than the segment shown in Fig. 4; the results for SOON and NAPCA are shown in Fig. 5. The vertical scale of Fig. 5 is normalized so that $\gamma_i = 1$ for pure-white-noise images $\mathbf{P}_i$. The components on the horizontal axis have been rank ordered by the metric γ_i. Since the thermal noise relative to dynamic range is roughly uniform across the AVIRIS channels, the performance differences between PCA and NAPCA are sufficiently small that only the NAPCA results are plotted. Below the arbitrary noise threshold of $\gamma_i = 0.85$ shown in Fig. 5, there are 24 SOON components and only 7 for NAPCA; although those NAPCA components above rank 42 are slightly less noisy than the corresponding SOON components, they are too noisy to be useful. Thus, this noise metric γ_i, based on the fractional energy in $\mathbf{P}_\mathrm{i}$ at its highest spatial frequencies, suggests the improved signal-separation abilities of SOON.

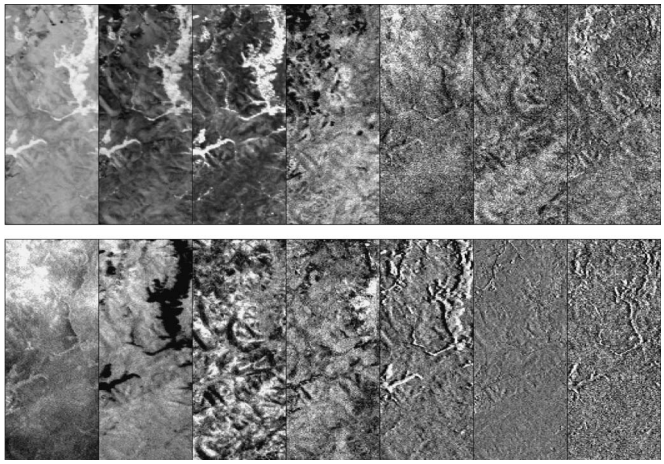

Fig. 6. Top halves of the visually most interesting components P_i found by (top row) NAPCA and (bottom row) SOON.

Visual inspection reveals the differences between the separated image components. Fig. 6 shows the top halves (natural terrain) of the visually most interesting and distinct components found by NAPCA and SOON, and Fig. 7 shows the urban bottom halves of these same images. The sign of each component was reversed as necessary to maximize the visual similarity of the images in each set. Component orthogonality was imposed only for the original larger image, not within the subimages. In both figures, NAPCA components reveal redundant information in each subimage more often than does SOON. For example, the first three NAPCA subimages (ordered by eigenvalue) shown in Fig. 6 capture the same terrain features, albeit somewhat differently, while the seven SOON terrain images are all sufficiently different that it is not even obvious that the terrain is the same. In Fig. 7, NAPCA tends to highlight the same buildings repeatedly, although with different combinations of values, while the SOON image set is again more diverse. Consistent with Fig. 5 but not shown, the NAPCA

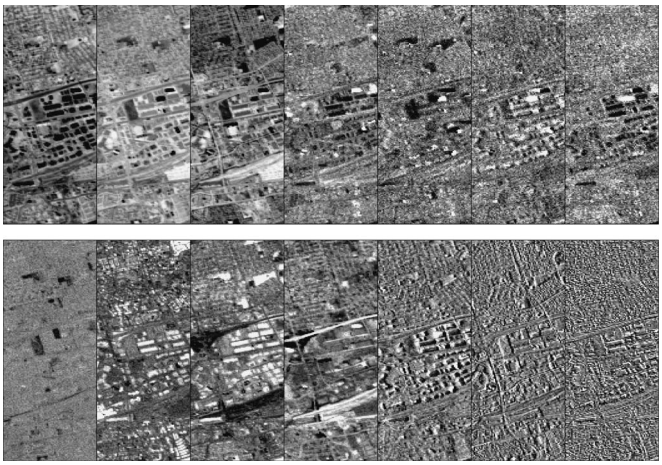

Fig. 7. Bottom halves of the visually most interesting components P_i found by (top row) NAPCA and (bottom row) SOON.

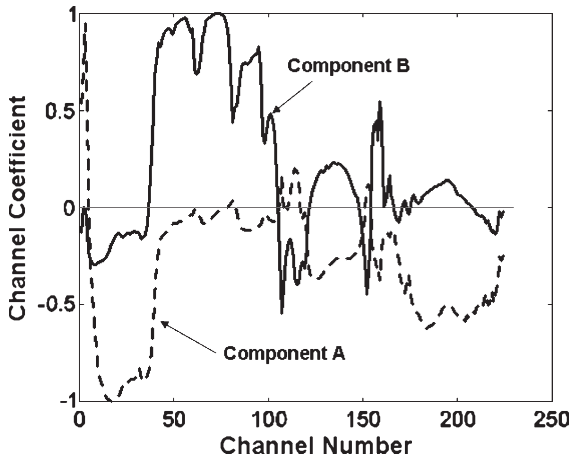

Fig. 8. Dashed line (component A): Normalized channel coefficients for the fifth SOON component P_5 of Fig. 7. Its greatest sensitivity is to the short visible wavelength channels 1–40 and the infrared channels above 170. Solid line (component B): Normalized channel coefficients for the sixth SOON component P_6 of Fig. 7. Its greatest sensitivity is to the near-infrared channels 40–100.

images also appear to drop into the noise sooner than those of SOON, although a few of the higher order NAPCA images still have utility. One reason SOON outperforms NAPCA in this way is that SOON treats different spatial frequencies differently, thereby recognizing more potential degrees of freedom. For example, if red/blue features have different spatial spectra than do red/green features, they will be more easily separated than if they had the same spatial spectra.

Another illustration of SOON's ability to better separate multispectral features having diverse spatial correlations is its ability to capture a phenomenon not well revealed by NAPCA principal components—the shadowing signatures on different spatial scales shown in the last three subimages of Figs. 6 and 7. The physical origin of these signatures is suggested partly by the shadowing, which is centered near 4 o'clock, and by the corresponding principal components A and B, as shown in Fig. 8. Component A forms the low-spatial-frequency image (the fifth of seven SOON subimages in Fig. 7) and is dominated by the UV/visible band plus longer wavelengths, while component B forms the medium spatial frequency image (the sixth subimage in Fig. 7) and is dominated by the red and near-infrared channels. Based on this and the apparent illumination effect, these three very distinct subimages are plausibly responding differently to solar heating, radiation cooling, and the magnitude of specular and diffuse solar reflections. In contrast, NAPCA principal components are insensitive to localized spatial differences in spectral correlations and use instead those spectral features to distinguish the various $\mathbf{P}_i$.

It would also be interesting to compare SOON to other source-separation algorithms such as SOBI and SONS that utilize only second-order statistical information and also ICA algorithms such as JADE, Fast ICA, and others that primarily utilize higher order information. In order to be useful, however, multiple images with distinctive characteristics should be compared. For example, SOON more effectively reduces noise than do algorithms such as SOBI and SONS, so its relative performance would be sensitive to the image's noise level and statistical stationarity, as demonstrated in Section III. Comparisons with ICA algorithms [4]–[6] are further challenged by the diversity of both the algorithms and the statistical character of the images, for SOON would perform well with images having rich second-order structures but no higher order correlations, while methods that primarily use higher order correlations would be disadvantaged. Conversely, images with significant high-order correlations and little second-order complexity would presumably be separated more effectively by ICA methods that exploit them. Still, better performance should be obtainable by coupling SOON to ICA algorithms to exploit all orders, most simply by using SOON only for noise reduction. Such comparisons should be quite interesting but are beyond the scope of this paper.

V. Conclusion

This paper has introduced SOON, a BSS algorithm for additive jointly Gaussian signal mixtures $\mathbf{X}$ that consist of m samples of n variables. SOON estimates the k underlying independent additive variables $\mathbf{P}_i$ for each sample, the mixing transform $\mathbf{A}$, the noise variances $\mathbf{G}$, and thereby even the additive noise $\mathbf{W}$ on each sample. SOON combines and extends the best features of prior algorithms such as the following: 1) noise adjustment prior to computing principal components (NAPCA); 2) Iterative estimation of signal Order and Noise (ION) using the EM algorithm and scree plots; and 3) joint consideration and utilization of spatial and spectral correlations (SONS). For SOON to work well, however, the order k should be less than approximately one-third of the number n of independent spectral channels because the scree-plot method for order estimation degrades otherwise. Similarly, the number of variables n should exceed 15 or 20, and the number of observations m should substantially exceed the number of variables n.

SOON's estimates of $\mathbf{P}_i$ can provide clues to underlying independent physical characteristics and can help improve blind categorization of multispectral images and automated feature recognition. The estimates $\mathbf{P}_i$ can also facilitate data compression and data recovery from isolated errors. For

example, comparison of an observed multispectral image $\mathbf{X}$ with the corresponding estimated $\hat{\mathbf{X}}$ reconstructed from $\hat{\mathbf{P}}$ and $\hat{\mathbf{A}}$ under the assumption of zero noise $\mathbf{W}$ yields the estimated noise $\mathbf{GW}$ for each channel and pixel, permitting excessive noise to be detected and optionally replaced by $\mathbf{AP}$, as shown using ION instead of SOON [25].

Hopefully, these encouraging results will stimulate future development of SOON by the community and experiments to determine the dependence of SOON's performance on choices of image, segmentation strategies in step 8, correlation offsets τ_j in step 9, metric definitions, and use of SONS versus JADE versus Fast-ICA, etc.

ACKNOWLEDGMENT

The authors would like to thank J. Lee and C. Cabrera-Mercader for useful conversations.

REFERENCES

[1] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "A survey of convolutive blind source separation methods," *Springer Handbook on Speech Processing and Speech Communication*, Springer-Verlag, Berlin, Germany, 2007, pp. 1–34.

[2] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[3] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[4] J.-F. Cardoso, "Dependence, correlation and Gaussianity in independent component analysis," *J. Mach. Learn. Res.*, vol. 4, no. 7/8, pp. 1177–1203, Dec. 2003.

[5] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Chichester, U.K.: Wiley, 2001.

[6] D.-T. Pham, "Blind separation of instantaneous mixture of sources based on order statistics," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 363–375, Feb. 2000.

[7] H. Attias, "Independent factor analysis," *Neural Comput.*, vol. 11, no. 4, pp. 803–851, May 1999.

[8] D. B. Rowe, "A Bayesian approach to blind source separation," *J. Interdiscip. Math.*, vol. 5, no. 1, pp. 49–76, Feb. 2002.

[9] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[10] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing*. Chichester, U.K.: Wiley, 2002.

[11] K. T. Herring, "Blind separation of noisy multivariate data using second-order statistics," M.S. thesis, Dept. Elect. Eng., MIT, Cambridge, MA, 2005.

[12] J. Lee and D. H. Staelin, "Iterative signal-order and noise estimation for multivariate data," *Electron. Lett.*, vol. 37, no. 2, pp. 134–135, Jan. 2001.

[13] J. Lee, "Blind noise estimation and compensation for improved characterization of multivariate processes," Ph.D. dissertation, Dept. Elect. Eng., Massachusetts Inst. Technol., Cambridge, MA, 2000.

[14] L. Tong, Y. Inouye, and R.-W. Liu, "A finite-step global convergence algorithm for the parameter estimation of multichannel MA processes," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2547–2558, Oct. 1992.

[15] K. Fukuda, H. E. Stanley, and L. A. N. Amaral, "Heuristic segmentation of a nonstationary time series," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 2, pp. 021 108-1–021 108-12, Feb. 2004.

[16] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.

[17] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Multivar. Anal.*, vol. 20, no. 1, pp. 1–25, Oct. 1986.

[18] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[19] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.

[20] A. V. Mueller, "Iterative blind separation of Gaussian data of unknown order," M.S. thesis, Dept. Elect. Eng., Massachusetts Inst. Technol., Cambridge, MA, 2003.

[21] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.

[22] J. B. Lee, S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 3, pp. 295–304, May 1990.

[23] *Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Homepage*, Feb.–May 2003, Jet Propulsion Lab., California Inst. Technol. [Online]. Available: http://aviris.jpl.nasa.gov/

[24] G. Vane, Ed., *Airborne Visible/Infrared Imaging Spectrometer (AVIRIS): A Description of the Sensor, Ground Data Processing Facility, Laboratory Calibration, and First Results*. Pasadena, CA: Jet Propulsion Lab., 1987. Publ. 87-38. [Online]. Available: ftp://popo.jpl.nasa.gov/pub/docs/workshops/87_docs/JPLPUB87-38.pdf

[25] C. R. Cabrera-Mercader, "Robust compression of multispectral remote sensing data," Ph.D. dissertation, Dept. Elect. Eng., Massachusetts Inst. Technol., Cambridge, MA, 1999.

Keith T. Herring received the B.S. degree in computer science from the University of Illinois at Urbana–Champaign, Urbana, in 2003 and the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2005 and 2008, respectively.

He currently is a Post-Doctoral Associate with the Research Laboratory of Electronics, MIT, where his research was focused on blind signal identification and wireless software receiver development and propagation research from 2003 to 2008.

Amy V. Mueller received the B.S. and M.Eng. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002 and 2003, respectively, where she is currently working toward the Ph.D. degree in the Department of Civil and Environmental Engineering.

From 2002 to 2003, she was with the Remote Sensing and Estimation Group, Research Laboratory of Electronics, MIT, where she was focusing on environmental applications of signal-processing algorithms. She is currently working to improve sensor technologies through the application of nonlinear signal-processing models at MIT.

David H. Staelin (S'59–M'65–SM'75–F'79–LF'04) received the B.S., M.S., and D.Sc. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1960, 1961, and 1965, respectively.

In 1965, he joined the faculty of MIT, where he is currently a Professor of electrical engineering with the Research Laboratory of Electronics. He teaches electromagnetics and signal processing. From 1990 to 2001, he was an Assistant Director of the Lincoln Laboratory, MIT. He was a Principal Investigator for the NASA Nimbus E Microwave Spectrometer and Scanning Microwave Spectrometer experiments on the Nimbus 5 and Nimbus 6 satellites and a Co-investigator for the NASA AIRS/AMSU/HSB experiment on Aqua, Scanning Multichannel Microwave Radiometer experiment on Nimbus 7, and planetary radio-astronomy experiment on Voyagers 1 and 2. Other research has involved estimation, radio astronomy, video coding, milliarc-second optical astrometry, and wireless communications. He is a member of the science teams for the NASA Precipitation Measurement Mission and the NPOESS Preparatory Program.