Subspace Analysis of Spatial Time–Frequency Distribution Matrices

Yimin Zhang, *Member, IEEE*, Weifeng Mu, and Moeness G. Amin, *Fellow, IEEE*

***Abstract*—Spatial time–frequency distributions (STFDs) have been recently introduced as the natural means to deal with source signals that are localizable in the time–frequency domain. Previous work in the area has not provided the eigenanalysis of STFD matrices, which is key to understanding their role in solving direction finding and blind source separation problems in multisensor array receivers. The aim of this paper is to examine the eigenstructure of the STFDs matrices. We develop the analysis and statistical properties of the subspace estimates based on STFDs for frequency modulated (FM) sources. It is shown that improved estimates are achieved by constructing the subspaces from the time–frequency signatures of the signal arrivals rather than from the data covariance matrices, which are commonly used in conventional subspace estimation methods. This improvement is evident in a low signal-to-noise ratio (SNR) environment and in the cases of closely spaced sources. The paper considers the MUSIC technique to demonstrate the advantages of STFDs and uses it as grounds for comparison between time–frequency and conventional subspace estimates.**

***Index Terms*—Array signal processing, spatial time–frequency distribution, subspace analysis, time–frequency distribution, time–frequency MUSIC.**

I. Introduction

While time–frequency distributions (TFDs) [1]–[4] have been sought out and successfully used in the areas of speech, biomedicine, the automotive industry, and machine monitoring, their use in sensor and spatial signal processing has not been properly investigated. The evaluation of quadratic TFDs of the data snapshots across the array yields the "spatial time-frequency distributions" (STFDs) [5], [6]. These spatial distributions permit the application of eigenstructure subspace techniques to the solution of a large class of channel estimation and equalization, blind source separation, and high-resolution direction-of-arrival (DOA) estimation problems. In the area of blind source separation, the STFDs allow the separation of Gaussian sources with identical spectral shape but with different time–frequency localization properties, i.e., different signatures in the time–frequency domain. For DOA estimation problems, the construction of the signal and noise subspaces using the source time–frequency signatures improves angular resolution performance.

Although the applications of the spatial time–frequency distributions to blind source separation and DOA estimation problems using multiple antenna arrays in nonstationary environments have been introduced in [5], [7], and [8], thus far, there has been no sufficient analysis that explains their offerings and justifies their performance. The aim of this paper is to examine the eigenstructure of the STFD matrices and provide statistical analysis of their respective signal and noise subspaces. The paper focuses on the class of frequency modulated (FM) signals as they represent a clear case of nonstationary signals that are localizable in the time–frequency domain. It shows that the subspaces obtained from the STFDs are robust to both noise and angular separation of the FM waveforms incident on the array. This robustness is primarily due to spreading the noise power while localizing the source energy in the time–frequency domain. By forming the STFD matrices from the points residing on the source time–frequency signatures, we increase, in essence, the input signal-to-noise ratio (SNR) and, hence, improve the accuracy of the subspace estimates.

This paper is organized as follows. Section II presents the signal model and considers nonstationary environments defined by FM source signals. The statistical properties of signal and noise subspace estimates for uncorrelated FM signals over the observation period are delineated. In Section III, we give a brief review of the definition and basic properties of the STFDs and derive the signal and noise subspaces using STFD matrices for the general class of FM signals. We demonstrate the robustness of the STFD-based subspace estimates to both noise and angular source separation, as compared with those obtained in Section II, using covariance matrices. The analytical results of Sections II and III are used in Section IV to examine the performance of the direction-finding MUSIC technique based on the covariance matrix and STFD noise subspace estimates. Numerical simulations are given in Section V.

II. Subspace Analysis for FM Signals

A. Signal Model

In narrowband array processing, when n signals arrive at an m-element array, the linear data model

$$\mathbf{x}(t) = \mathbf{y}(t) + \mathbf{n}(t) = \mathbf{A}\mathbf{d}(t) + \mathbf{n}(t) \tag{1}$$

is commonly assumed, where the $m \times n$ spatial matrix $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$ represents the mixing matrix or the steering matrix. In direction-finding problems, we require $\mathbf{A}$ to have a known structure, and each column of $\mathbf{A}$ corresponds to a single arrival and carries a clear bearing. On the other hand, when we consider blind source separation problems, $\mathbf{A}$ is a mixture of several steering vectors, due to multipaths, and its columns may

Manuscript received August 6, 1999; revised December 28, 2000. This work was supported by the Office of Naval Research under Grant N00014-98-1-0176. The associate editor coordinating the review of this paper and approving it for publication was Dr. Paulo J. S. G. Ferreira.

The authors are with the Department of Electrical and Computer Engineering, Villanova University, Villanova, PA 19085 USA (e-mail: zhang@ece.vill.edu; weifang@ece.vill.edu; moeness@ece.vill.edu).

Publisher Item Identifier S 1053-587X(01)02247-4.

assume any structure. The analytical treatment in this paper does not depend on any special structure of matrix $\mathbf{A}$.

Due to the mixture of the signals at each sensor, the elements of the $m \times 1$ data vector $\mathbf{x}(t)$ are multicomponent signals, whereas each source signal $d_i(t)$ of the $n \times 1$ signal vector $\mathbf{d}(t)$ is often a monocomponent signal. $\mathbf{n}(t)$ is an additive noise vector whose elements are modeled as stationary, spatially and temporally white, zero-mean complex random processes that are independent of the source signals. That is

$$E[\mathbf{n}(t+\tau)\mathbf{n}^H(t)] = \sigma\delta(\tau)\mathbf{I}$$

and

$$E[\mathbf{n}(t+\tau)\mathbf{n}^T(t)] = \mathbf{0}, \qquad \text{for any } \tau \tag{2}$$

where $\delta(\tau)$ is the delta function, $\mathbf{I}$ denotes the identity matrix, σ is the noise power at each sensor, superscripts H and T, respectively, denote conjugate transpose and transpose, and $E(\cdot)$ is the statistical expectation operator.

In (1), it is assumed that the number of sensors is larger than the number of sources, i.e., $m > n$. Further, matrix $\mathbf{A}$ is full column rank, which implies that the steering vectors corresponding to n different angles of arrival are linearly independent. We further assume that the correlation matrix

$$\mathbf{R_{xx}} = E[\mathbf{x}(t)\mathbf{x}^H(t)] \tag{3}$$

is nonsingular and that the observation period consists of N snapshots with $N > m$.

Under the above assumptions, the correlation matrix is given by

$$\mathbf{R_{xx}} = E[\mathbf{x}(t)\mathbf{x}^H(t)] = \mathbf{A}\mathbf{R_{dd}}\mathbf{A}^H + \sigma\mathbf{I} \tag{4}$$

where $\mathbf{R_{dd}} = E[\mathbf{d}(t)\mathbf{d}^H(t)]$ is the source correlation matrix.

Let $\lambda_1 > \lambda_2 > \cdots > \lambda_n > \lambda_{n+1} = \lambda_{n+2} = \cdots = \lambda_m = \sigma$ denote the eigenvalues of $\mathbf{R_{xx}}$. It is assumed that λ_i, $i = 1, \cdots, n$ are distinct. The unit-norm eigenvectors associated with $\lambda_1, \cdots, \lambda_n$ constitute the columns of matrix $\mathbf{S} = [\mathbf{s}_1, \cdots, \mathbf{s}_n]$, and those corresponding to $\lambda_{n+1}, \cdots, \lambda_m$ make up matrix $\mathbf{G} = [\mathbf{g}_1, \cdots, \mathbf{g}_{m-n}]$. Since the columns of $\mathbf{A}$ and $\mathbf{S}$ span the same subspace, then $\mathbf{A}^H\mathbf{G} = \mathbf{0}$.

In practice, $\mathbf{R_{xx}}$ is unknown and, therefore, should be estimated from the available data samples (snapshots) $\mathbf{x}(i)$, $i = 1, 2, \cdots, N$. The estimated correlation matrix is given by

$$\hat{\mathbf{R}}_{\mathbf{xx}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}(i)\mathbf{x}^H(i). \tag{5}$$

Let $\{\hat{\mathbf{s}}_1, \cdots, \hat{\mathbf{s}}_n, \hat{\mathbf{g}}_1, \cdots, \hat{\mathbf{g}}_{m-n}\}$ denote the unit-norm eigenvectors of $\hat{\mathbf{R}}_{\mathbf{xx}}$ that are arranged in the descending order of the associated eigenvalues, and let $\hat{\mathbf{S}}$ and $\hat{\mathbf{G}}$ denote the matrices defined by the set of vectors $\{\hat{\mathbf{s}}_i\}$ and $\{\hat{\mathbf{g}}_i\}$, respectively. The statistical properties of the eigenvectors of the sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{xx}}$ for signals modeled as independent processes with additive white noise are given in [9].

B. *Subspace Analysis for FM Signals*

In this paper, we focus on analytic frequency modulation (FM) signals, which are modeled as

$$\mathbf{d}(t) = [d_1(t), \cdots, d_n(t)]^T = [D_1 e^{j\psi_1(t)}, \cdots, D_n e^{j\psi_n(t)}]^T \tag{6}$$

where D_i and $\psi_i(t)$ are the fixed amplitude and time-varying phase of the ith source signal. For each sampling time t, $d_i(t)$ has an instantaneous frequency $f_i(t) = (1/2\pi)(d\psi_i(t)/dt)$.

FM signals are often encountered in applications such as radar and sonar [2]. The consideration of FM signals in this paper is motivated by the fact that these signals are uniquely characterized by their instantaneous frequencies, and therefore, they have clear time–frequency signatures that can be utilized by the STFD approach. In addition, FM signals have constant amplitudes. To simplify the analysis, we assume that the transmitted signals propagate in a stationary environment and are mutually uncorrelated over the observation period $[1 : N]$. Subsequently, the corresponding covariance matrices are time independent. Under these assumptions, we have

$$\frac{1}{N}\sum_{k=1}^{N} d_i(k)d_j^*(k) = 0 \qquad \text{for } i \neq j, \quad i, j = 1, \cdots, n. \tag{7}$$

In this case, the signal correlation matrix in (4) is

$$\mathbf{R_{dd}} = \text{diag}[D_i^2,\, i = 1, 2, \cdots, n]$$

where diag[$\cdot$] is the diagonal matrix formed with the elements of its vector valued arguments. From the above assumptions, we have the following Lemma.

Lemma 1: For uncorrelated FM signals with additive white Gaussian noise, we have the following.

a) The estimation errors $(\hat{\mathbf{s}}_i - \mathbf{s}_i)$ are asymptotically (for large N) jointly Gaussian distributed with zero means and covariance matrices given by

$$\begin{aligned} E[(\hat{\mathbf{s}}_i - \mathbf{s}_i)(\hat{\mathbf{s}}_j - \mathbf{s}_j)^H] &= \frac{\sigma}{N}\left[\sum_{\substack{k=1\\k\neq i}}^{n}\frac{\lambda_i+\lambda_k-\sigma}{(\lambda_k-\lambda_i)^2}\mathbf{s}_k\mathbf{s}_k^H + \sum_{k=1}^{m-n}\frac{\lambda_i}{(\sigma-\lambda_i)^2}\mathbf{g}_k\mathbf{g}_k^H\right]\delta_{i,j} \\ &\triangleq \tilde{\mathbf{W}}_i\delta_{i,j}, \end{aligned} \tag{8}$$

$$\begin{aligned} E[(\hat{\mathbf{s}}_i - \mathbf{s}_i)(\hat{\mathbf{s}}_j - \mathbf{s}_j)^T] &= -\frac{\sigma}{N}\frac{(\lambda_i+\lambda_j-\sigma)}{(\lambda_j-\lambda_i)^2}\mathbf{s}_j\mathbf{s}_i^T(1-\delta_{i,j}) \\ &\triangleq \tilde{\mathbf{V}}_{i,j}. \end{aligned} \tag{9}$$

b) The orthogonal projections of $\{\hat{\mathbf{g}}_i\}$ onto the column space of $\mathbf{S}$ are asymptotically (for large N) jointly Gaussian distributed with zero means and covariance matrices given by

$$\begin{aligned} E[(\mathbf{S}\mathbf{S}^H\hat{\mathbf{g}}_i)(\mathbf{S}\mathbf{S}^H\hat{\mathbf{g}}_j)^H] &= \frac{\sigma}{N}\left[\sum_{k=1}^{n}\frac{\lambda_k}{(\sigma-\lambda_k)^2}\mathbf{s}_k\mathbf{s}_k^H\right]\delta_{i,j} \\ &\triangleq \frac{1}{N}\tilde{\mathbf{U}}\delta_{i,j} \end{aligned} \tag{10}$$

$$E[(\mathbf{S}\mathbf{S}^H\hat{\mathbf{g}}_i)(\mathbf{S}\mathbf{S}^H\hat{\mathbf{g}}_j)^T] = 0 \qquad \text{for all } i, j. \tag{11}$$

The proof of part a) is given in Appendix A. The proof of part b) follows the same exact steps of the respective results derived in [9] and is not given here. Equations (8) and (9) hold strong similarity to those of [9]. The only difference is that the term $(\lambda_i\lambda_k)$ in [9, (3.8a) and (3.8b)] is replaced by $\sigma(\lambda_i+\lambda_k-\sigma)$ in (8) and (9), due to the uncorrelation property (7). Accordingly, for high input SNR $(\lambda_k \gg \sigma,\, k=1,\,2,\,\cdots,\,n)$, the estimation error of $(\hat{\mathbf{s}}_i - \mathbf{s}_i)$ can be greatly reduced. From (8) and (9), each column of the signal subspace will be perfectly estimated when $\sigma = 0$. This is in contrast with the estimation error that would result under the same noise-free condition if we use the temporally independent signal characteristics considered in [9].

Equations (10) and (11) are identical to [9, (3.9) and (3.10)]. The reason of such identity is that despite the difference in the signal eigenvectors in the two different scenarios, which are discussed in this paper and in [9], the signal subspaces in both cases are identical and are spanned by the columns of matrix $\mathbf{A}$. Accordingly, the projection of the estimated noise eigenvectors on the true signal subspace for either FM signals or white random processes yield equal results.

III. Subspace Analysis for STFD Matrices

The purpose of this section is to show that the signal and noise subspaces based on TFDs for nonstationary signals are more robust to noise than those obtained from conventional array processing.

A. Spatial Time–Frequency Distributions

We first review the definition and basic properties of the STFDs. STFDs based on Cohen's class of TFD were introduced in [5], and its applications to direction finding and blind source separation have been discussed in [5], [7], and [8]. In this paper, we consider the simplest member of Cohen's class, namely, the pseudo Wigner–Ville distribution (PWVD) [1] and its respective spatial distribution. Only the time-frequency (t-f) points in the autoterm regions of PWVD are considered for STFD matrix construction. The autoterm region refers to the t-f points along the true instantaneous frequency (IF) of each signal. The crossterms may intrude on the autoterms through the power in their mainlobes or/and sidelobes. This intrusion depends on the signal temporal structures and the window size. In this paper, however, we assume that the crossterms are negligible over the autoterm regions.

The discrete form of pseudo Wigner–Ville distribution of a signal $x(t)$, using a rectangular window of odd length L, is given by

$$D_{xx}(t,\,f) = \sum_{\tau=-((L-1)/2)}^{(L-1)/2} x(t+\tau)x^*(t-\tau)e^{-j4\pi f\tau} \quad (12)$$

where $*$ denotes complex conjugate. It should be noted that incorporating multiple t-f points, via t-f averaging, over the autoterm region causes the crossterm components present at the signal IF to cancel each other, rendering their overall effect negligible.

The spatial pseudo Wigner–Ville distribution (SPWVD) matrix is obtained by replacing $x(t)$ by the data snapshot vector $\mathbf{x}(t)$

$$\mathbf{D_{xx}}(t,\,f) = \sum_{\tau=-((L-1)/2)}^{(L-1)/2} \mathbf{x}(t+\tau)\mathbf{x}^H(t-\tau)e^{-j4\pi f\tau}. \quad (13)$$

Substituting (1) into (13), we obtain

$$\mathbf{D_{xx}}(t,\,f)=\mathbf{D_{yy}}(t,\,f)+\mathbf{D_{yn}}(t,\,f)+\mathbf{D_{ny}}(t,\,f)+\mathbf{D_{nn}}(t,\,f). \quad (14)$$

We note that $\mathbf{D_{xx}}(t,\,f), \mathbf{D_{yy}}(t,\,f), \mathbf{D_{yn}}(t,\,f), \mathbf{D_{ny}}(t,\,f)$, and $\mathbf{D_{nn}}(t,\,f)$ are matrices of dimension $m \times m$. Under the uncorrelated signal and noise assumption and the zero-mean noise property, the expectation of the crossterm STFD matrices between the signal and noise vectors is zero, i.e., $E[\mathbf{D_{yn}}(t,\,f)] = E[\mathbf{D_{ny}}(t,\,f)] = \mathbf{0}$, and it follows that

$$\begin{aligned} E[\mathbf{D_{xx}}(t,\,f)] &= \mathbf{D_{yy}}(t,\,f) + E[\mathbf{D_{nn}}(t,\,f)] \\ &= \mathbf{A}\mathbf{D_{dd}}(t,\,f)\mathbf{A}^H + E[\mathbf{D_{nn}}(t,\,f)] \end{aligned} \quad (15)$$

where the source TFD matrix

$$\mathbf{D_{dd}}(t,\,f) = \sum_{\tau=-((L-1)/2)}^{(L-1)/2} \mathbf{d}(t+\tau)\mathbf{d}^H(t-\tau)e^{-j4\pi f\tau} \quad (16)$$

is of dimension $n \times n$. For narrowband array signal processing applications, the mixing matrix $\mathbf{A}$ holds the spatial information and maps the auto- and cross-TFDs of the source signals into auto- and cross-TFDs of the data.

Equation (15) is similar to the formula that has been commonly used in DOA estimation and blind source separation problems, relating the signal correlation matrix to the data spatial correlation matrix. In the above formulation, however, the correlation matrices are replaced by the STFD matrices. The well-established results in conventional array signal processing could, therefore, be utilized, and key problems in various applications of array processing, specifically those dealing with nonstationary signal environments, can be approached using bilinear transformations.

It is noted that (15) holds true for every $(t,\,f)$ point. In order to reduce the effect of noise and ensure the full column rank property of the STFD matrix, we consider multiple t-f points, instead of a single one. That is, the signal and noise subspaces are constructed using as many $(t,\,f)$ points in the source autoterm regions as possible. This allows more information of the source signal t–f signatures to be included into their respective subspace formulation and, as such, enhances direction finding and source separation performance. Joint-diagonalization [10] and t-f averaging are the two main approaches that have been used for this purpose [5], [7], [11]. In this paper, however, we only consider averaging over multiple t–f points.

B. SNR Enhancement

The TFD maps one-dimensional (1-D) signals in the time domain into two-dimensional (2-D) signals in the t–f domain. The

TFD property of concentrating the input signal around its instantaneous frequency while spreading the noise over the entire t–f domain increases the effective SNR and proves valuable in the underlying problem.

The ith diagonal element of the TFD matrix $\mathbf{D_{dd}}(t, f)$ is given by

$$D_{d_i d_i}(t, f) = \sum_{\tau=-((L-1)/2)}^{(L-1)/2} D_i^2 e^{j[\psi_i(t+\tau)-\psi_i(t-\tau)]-j4\pi f\tau}. \tag{17}$$

Assume that the third-order derivative of the phase is negligible over the window length L, then along the true t–f points of the ith signal, $f_i(t) = (1/2\pi)(d\psi_i(t)/dt)$, and $\psi_i(t+\tau) - \psi_i(t-\tau) - 4\pi f_i(t)\tau = 0$. Accordingly, for $(L-1)/2 \le t \le N-(L-1)/2$

$$D_{d_i d_i}(t, f_i(t)) = \sum_{\tau=-((L-1)/2)}^{(L-1)/2} D_i^2 = LD_i^2. \tag{18}$$

Similarly, the noise STFD matrix $\mathbf{D_{nn}}(t, f)$ is

$$\mathbf{D_{nn}}(t, f) = \sum_{\tau=-((L-1)/2)}^{(L-1)/2} \mathbf{n}(t+\tau)\mathbf{n}^H(t-\tau)e^{-j4\pi f\tau}. \tag{19}$$

Under the spatially and temporally white assumptions, the statistical expectation of $\mathbf{D_{nn}}(t, f)$ is given by

$$\begin{aligned} E[\mathbf{D_{nn}}(t, f)] &= \sum_{\tau=-((L-1)/2)}^{(L-1)/2} E[\mathbf{n}(t+\tau)\mathbf{n}^H(t-\tau)]e^{-j4\pi f\tau} \\ &= \sigma\mathbf{I}. \end{aligned} \tag{20}$$

Therefore, when we select the t–f points along the t–f signature or the IF of the ith FM signal, the SNR in model (15) is LD_i^2/σ, which has an improved factor L over the one associated with model (4). The IF of the FM signals can be estimated from the employed TFD, which in this case is the PWVD. It may also be given separately using any appropriate IF estimator. It is noted that the STFD equation (15) provides a natural platform for the direct incorporation of any *a priori* information or estimates of the IF into DOA estimation.

The pseudo Wigner–Ville distribution of each FM source has a constant value over the observation period, providing that we leave out the rising and falling power distributions at both ends of the data record. For convenience of analysis, we select those $N-L+1$ t–f points of constant distribution value for each source signal. In the case where the STFD matrices are averaged over the t–f signatures of n_o sources, i.e., a total of $n_o(N-L+1)$ t–f points, the result is given by

$$\hat{\mathbf{D}} = \frac{1}{n_o(N-L+1)} \sum_{q=1}^{n_o} \sum_{i=1}^{N-L+1} \mathbf{D_{xx}}(t_i, f_{q,i}(t_i)) \tag{21}$$

where $f_{q,i}(t_i)$ is the instantaneous frequency of the qth signal at the ith time sample. $\mathbf{x}(t)$ is an instantaneous mixture of the FM signals $d_i(t)$, $i = 1, \cdots, n$ and, hence, features the same IFs. The expectation of the averaged STFD matrix is

$$\begin{aligned} \mathbf{D} = E[\hat{\mathbf{D}}] &= \frac{1}{n_o(N-L+1)} \sum_{q=1}^{n_o} \sum_{i=1}^{N-L+1} E[\mathbf{D_{xx}}(t_i, f_{q,i}(t_i))] \\ &= \frac{1}{n_o} \sum_{q=1}^{n_o} [LD_q^2 \mathbf{a}_q \mathbf{a}_q^H + \sigma\mathbf{I}] = \frac{L}{n_o} \mathbf{A}^o \mathbf{R}_{\mathbf{dd}}^o (\mathbf{A}^o)^H + \sigma\mathbf{I} \end{aligned} \tag{22}$$

where $\mathbf{R}_{\mathbf{dd}}^o = \text{diag}[D_i^2, i = 1, 2, \cdots, n_o]$, and $\mathbf{A}^o = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{n_o}]$ represent the signal correlation matrix and the mixing matrix formulated by considering n_o signals out of the total number of n signal arrivals, respectively.

It is clear from (22) that the SNR improvement $G = L/n_o$ (we assume $L > n_o$ throughout this paper) is inversely proportional to the number of sources contributing matrix $\mathbf{D}$. Therefore, from the SNR perspective, it is best to set $n_o = 1$, i.e., to select the sets of $N-L+1$ t–f points that belong to individual signals one set at a time and then separately evaluate the respective STFD matrices.

This procedure is made possible by the fact that STFD-based direction finding is, in essence, a discriminatory technique in the sense that it does not require simultaneous localization and extraction of all unknown signals received by the array. With STFDs, direction finding can be performed using STFDs of a subclass of the impinging signals with specific t–f signatures. In this respect, the proposed direction finding technique acts as a spatial filter, removing all other signals from consideration and, subsequently, saving any downstream processing that is required to separate interference and signals of interest. It is also important to note that with the ability to construct the STFD matrix from one or few signal arrivals, the well-known $m > n$ condition on source localization using arrays can be relaxed, i.e., we can perform direction finding or source separation with the number of array sensors smaller than the number of impinging signals [6]. From the angular resolution perspective, closely spaced sources with different t–f signatures can be resolved by constructing two separate STFDs, each corresponding to one source, and then proceeding with subspace decomposition for each STFD matrix, followed by an appropriate source localization method (MUSIC, for example). The drawback of performing direction finding several times using different STFD matrices is, of course, the need for repeated computations of eigen-decompositions and source localizations.

C. Signal and Noise Subspaces Using STFDs

The following Lemma provides the relationship between the eigen-decompositions of the STFD matrices and the data covariance matrices used in conventional array processing.

Lemma 2: Let $\lambda_1^o > \lambda_2^o > \cdots > \lambda_{n_o}^o > \lambda_{n_o+1}^o = \lambda_{n_o+2}^o = \cdots = \lambda_m^o = \sigma$ denote the eigenvalues of $\mathbf{R}_{\mathbf{xx}}^o = \mathbf{A}^o \mathbf{R}_{\mathbf{dd}}^o (\mathbf{A}^o)^H + \sigma\mathbf{I}$, which is defined from a data record of a mixture of the n_o selected FM signals. Denote the unit-norm eigenvectors associated with $\lambda_1^o, \cdots, \lambda_{n_o}^o$ by the columns of $\mathbf{S}^o = [\mathbf{s}_1^o, \cdots, \mathbf{s}_{n_o}^o]$ and those corresponding to $\lambda_{n_o+1}^o, \cdots, \lambda_m^o$ by the columns of $\mathbf{G}^o = [\mathbf{g}_1^o, \cdots, \mathbf{g}_{m-n_o}^o]$. We also denote $\lambda_1^{tf} > \lambda_2^{tf} > \cdots > \lambda_{n_o}^{tf} > \lambda_{n_o+1}^{tf} =$

$\lambda_{n_o+2}^{tf} = \cdots = \lambda_m^{tf} = \sigma^{tf}$ as the eigenvalues of $\mathbf{D}$ defined in (22). The superscript tf denotes that the associated term is derived from the STFD matrix $\mathbf{D}$. The unit-norm eigenvectors associated with $\lambda_1^{tf}, \cdots, \lambda_{n_o}^{tf}$ are represented by the columns of $\mathbf{S}^{tf} = [\mathbf{s}_1^{tf}, \cdots, \mathbf{s}_{n_o}^{tf}]$, and those corresponding to $\lambda_{n_o+1}^{tf}, \cdots, \lambda_m^{tf}$ are represented by the columns of $\mathbf{G}^{tf} = [\mathbf{g}_1^{tf}, \cdots, \mathbf{g}_{m-n_o}^{tf}]$. Then, we have the following.

a) The signal and noise subspaces of $\mathbf{S}^{tf}$ and $\mathbf{G}^{tf}$ are the same as $\mathbf{S}^o$ and $\mathbf{G}^o$, respectively.
b) The eigenvalues have the following relationship:

$$\lambda_i^{tf} = \begin{cases} \dfrac{L}{n_o}(\lambda_i^o - \sigma) + \sigma = \dfrac{L}{n_o}\lambda_i^o + \left(1 - \dfrac{L}{n_o}\right)\sigma \\ \qquad\qquad i \le n_o \\ \sigma^{tf} = \sigma, \quad n_o < i \le m. \end{cases} \tag{23}$$

The proof of Lemma 2 is shown in Appendix B.

An important conclusion from Lemma 2 is that the largest n_o eigenvalues are amplified using STFD analysis. Fig. 1 shows the two principal (largest) eigenvalues λ_i^o (i.e., $L = 1$) and λ_i^{tf} (for $L = 33$ and $L = 129$), where a uniform linear array of eight sensors ($m = 8$) separated by half a wavelength and receiving signal from two sources ($n_o = n = 2$) is used. The two signals are of equal power ($D_1 = D_2 = D$), and their angular separation $\Delta\theta$ is defined as $\theta_2 - \theta_1$. We choose $\theta_1 + \theta_2 = 0$, that is, the two signals are symmetric with respect to the broadside direction. Denote

$$\beta = \frac{\mathbf{a}_1^H \mathbf{a}_2}{||\mathbf{a}_1||_2 ||\mathbf{a}_2||_2}$$

as the spatial correlation coefficient between the two directional vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, corresponding to the angles θ_1 and θ_2. $||\mathbf{a}||_2$ is the 2-norm of a vector $\mathbf{a}$. The two largest eigenvalues for the two uncorrelated signals are given by [12]

$$\lambda_{1,2}^o = mD^2[1 \pm |\beta|] + \sigma. \tag{24}$$

Hence, combining (23) and (24), we obtain

$$\lambda_{1,2}^{tf} = \frac{mL}{n_o} D^2[1 \pm |\beta|] + \sigma. \tag{25}$$

The amplification of the largest n_o eigenvalues improves detection of the number of the impinging signals on the array as it widens the separation between dominant and noise-level eigenvalues. Determination of the number of signals is key to establishing the proper signal and noise subspaces and subsequently plays a fundamental role in subspace-based applications [13]. When the input SNR is low or the signals are closely spaced, the number of signals may often be underdetermined. Fig. 2 shows, for the same signal scenario of Fig. 1, the threshold level of the input SNR required to determine the correct number of signals $\hat{n} = 2$ according to the Akaike information criterion (AIC) [14]

$$\min_{\hat{n}} N(m - \hat{n}) \log \left[\frac{f_1(\hat{n})}{f_2(\hat{n})}\right] + f_3(\hat{n}) \tag{26}$$

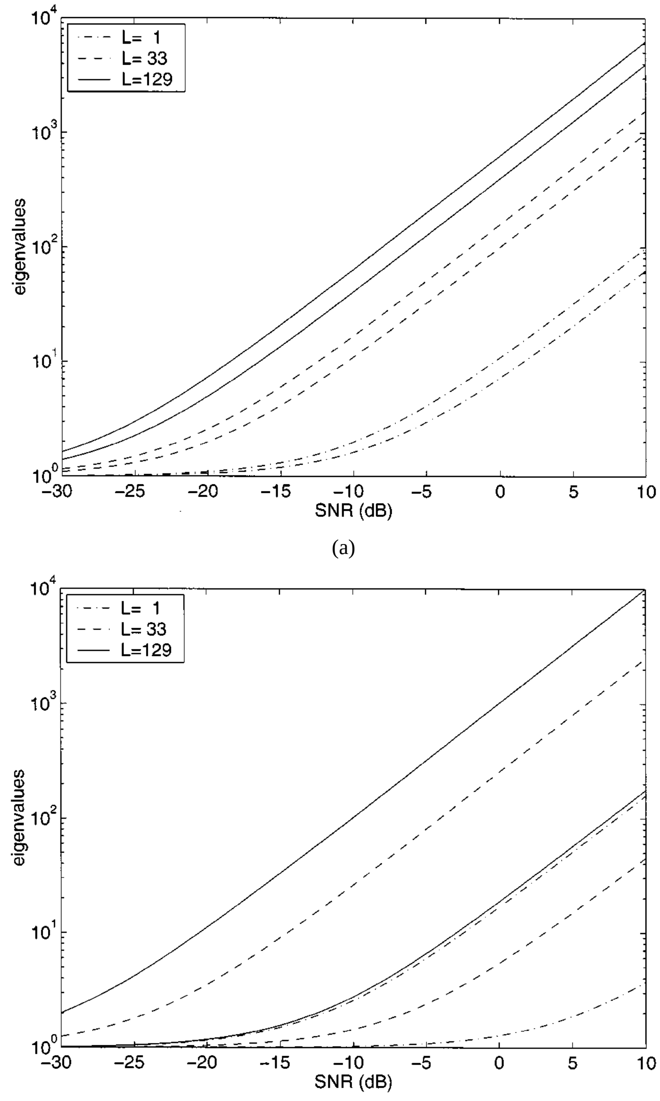

Fig. 1. Principal eigenvalues of the correlation and the STFD matrices. (a) $\theta_1 = -10°, \theta_2 = 10°$. (b) $\theta_1 = -1°, \theta_2 = 1°$.

where

$$f_1(\hat{n}) \triangleq \frac{1}{m - \hat{n}} \sum_{i=\hat{n}+1}^{m} \lambda_i,$$
$$f_2(\hat{n}) \triangleq \left(\prod_{i=\hat{n}+1}^{m} \lambda_i\right)^{1/(m-\hat{n})}$$
$$f_3(\hat{n}) \triangleq \hat{n}(2m - \hat{n}). \tag{27}$$

It is clear from Fig. 2 that when the STFD is applied, the SNR threshold level that is necessary for the correct determination of the number of signals is greatly reduced.

Next, we consider the signal and noise subspace estimates from a finite number of data samples. We form the STFD matrix based on the true (t, f) points along the IF of the n_o FM signals.

Lemma 3: If the third-order derivative of the phase of the FM signals is negligible over the time period $[t - L + 1, t + L - 1]$, then we have the following.

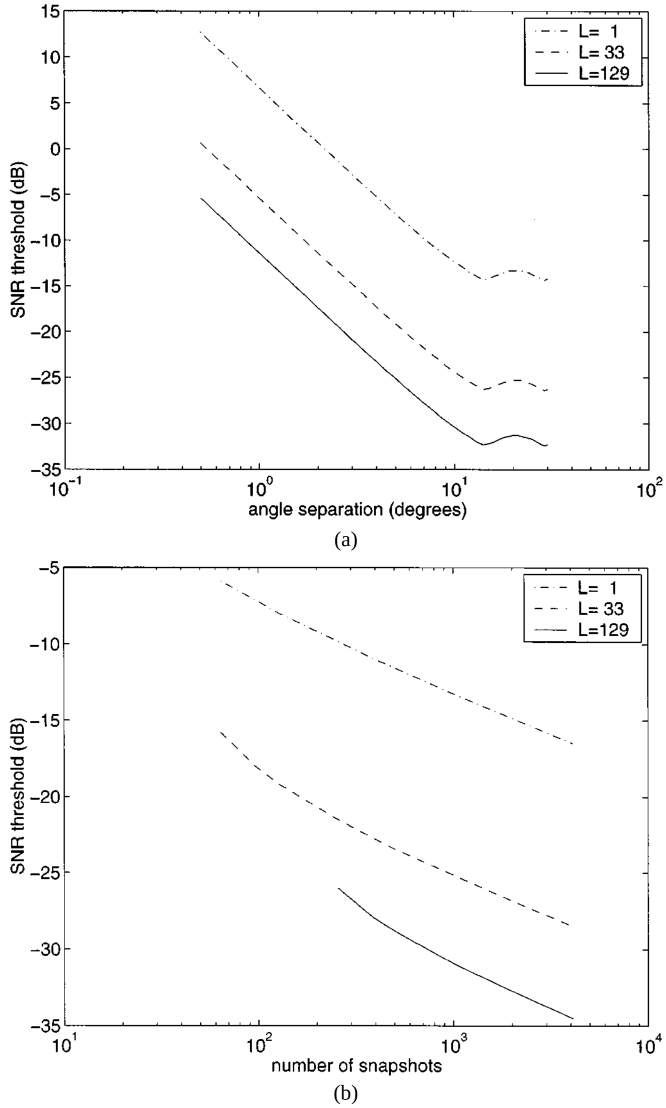

Fig. 2. SNR thresholds to identify two signals ($m = 8$). (a) SNR threshold versus angle separation ($N = 1024$). (b) SNR threshold versus number of snapshots ($\Delta\theta = 20°$).

a) The estimation errors in the signal vectors are asymptotically (for $N \gg L$) jointly Gaussian distributed with zero means and covariance matrices given by

$$
\begin{aligned}
&E(\hat{\mathbf{s}}_i^{tf} - \mathbf{s}_i^{tf})(\hat{\mathbf{s}}_j^{tf} - \mathbf{s}_j^{tf})^H \\
&= \frac{\sigma L}{n_o(N-L+1)} \left[\sum_{\substack{k=1\\k\neq i}}^{n_o} \frac{\lambda_i^{tf} + \lambda_k^{tf} - \sigma}{(\lambda_k^{tf} - \lambda_i^{tf})^2} \mathbf{s}_k^{tf}(\mathbf{s}_k^{tf})^H \right. \\
&\qquad \left. + \sum_{k=1}^{m-n_o} \frac{\lambda_i^{tf}}{(\sigma - \lambda_i^{tf})^2} \mathbf{g}_k^{tf}(\mathbf{g}_k^{tf})^H \right] \delta_{i,j} \\
&\triangleq \mathbf{W}_i^{tf} \delta_{i,j}
\end{aligned} \tag{28}
$$

$$
\begin{aligned}
&E(\hat{\mathbf{s}}_i^{tf} - \mathbf{s}_i^{tf})(\hat{\mathbf{s}}_j^{tf} - \mathbf{s}_j^{tf})^T \\
&= -\frac{\sigma L}{n_o(N-L+1)} \frac{(\lambda_i^{tf} + \lambda_j^{tf} - \sigma)}{(\lambda_j^{tf} - \lambda_i^{tf})^2} \mathbf{s}_j^{tf}(\mathbf{s}_i^{tf})^T (1 - \delta_{i,j}) \\
&\triangleq \mathbf{V}_{i,j}^{tf}.
\end{aligned} \tag{29}
$$

b) The orthogonal projections of $\{\hat{\mathbf{g}}_i^{tf}\}$ onto the column space of $\mathbf{S}^{tf}$ are asymptotically (for $N \gg L$) jointly Gaussian distributed with zero means and covariance matrices given by

$$
\begin{aligned}
&E(\mathbf{S}^{tf}(\mathbf{S}^{tf})^H \hat{\mathbf{g}}_i^{tf})(\mathbf{S}^{tf}(\mathbf{S}^{tf})^H \hat{\mathbf{g}}_j^{tf})^H \\
&= \frac{\sigma L}{n_o(N-L+1)} \left[\sum_{k=1}^{n_o} \frac{\lambda_k^{tf}}{(\sigma - \lambda_k^{tf})^2} \mathbf{s}_k^{tf}(\mathbf{s}_k^{tf})^H \right] \delta_{i,j} \\
&\triangleq \frac{1}{(N-L+1)} \mathbf{U}^{tf} \delta_{i,j}
\end{aligned} \tag{30}
$$

$$
\begin{aligned}
&E(\mathbf{S}^{tf}(\mathbf{S}^{tf})^H \hat{\mathbf{g}}_i^{tf})(\mathbf{S}^{tf}(\mathbf{S}^{tf})^H \hat{\mathbf{g}}_j^{tf})^T \\
&= \mathbf{0} \qquad \text{for all } i, j.
\end{aligned} \tag{31}
$$

The proof of (28) and (29) is given in Appendix C, and the proof of (30) and (31) is given in Appendix D.

To demonstrate the performance advantage of using STFDs, we substitute (23) into (28)–(30)

$$
\begin{aligned}
&E(\hat{\mathbf{s}}_i^{tf} - \mathbf{s}_i^{tf})(\hat{\mathbf{s}}_j^{tf} - \mathbf{s}_j^{tf})^H \\
&= \frac{\sigma}{N-L+1} \\
&\quad \times \left[\sum_{\substack{k=1\\k\neq i}}^{n_o} \frac{(\lambda_i^o - \sigma) + (\lambda_k^o - \sigma) + \frac{n_o}{L}\sigma}{(\lambda_k^o - \lambda_i^o)^2} \mathbf{s}_k^o(\mathbf{s}_k^o)^H \right. \\
&\qquad \left. + \sum_{k=1}^{m-n_o} \frac{(\lambda_i^o - \sigma) + \frac{n_o}{L}\sigma}{(\sigma - \lambda_i^o)^2} \mathbf{g}_k^o(\mathbf{g}_k^o)^H \right] \delta_{i,j}
\end{aligned} \tag{32}
$$

$$
\begin{aligned}
&E(\hat{\mathbf{s}}_i^{tf} - \mathbf{s}_i^{tf})(\hat{\mathbf{s}}_j^{tf} - \mathbf{s}_j^{tf})^T \\
&= -\frac{\sigma}{N-L+1} \cdot \frac{(\lambda_k^o - \sigma) + (\lambda_i^o - \sigma) + \frac{n_o}{L}\sigma}{(\lambda_k^o - \lambda_i^o)^2} \\
&\quad \times \mathbf{s}_j^o(\mathbf{s}_i^o)^T (1 - \delta_{i,j})
\end{aligned} \tag{33}
$$

and

$$
\begin{aligned}
&E(\mathbf{S}^{tf}(\mathbf{S}^{tf})^H \hat{\mathbf{g}}_i^{tf})(\mathbf{S}^{tf}(\mathbf{S}^{tf})^H \hat{\mathbf{g}}_j^{tf})^H \\
&= \frac{\sigma}{N-L+1} \left[\sum_{k=1}^{n_o} \frac{(\lambda_k^o - \sigma) + \frac{n_o}{L}\sigma}{(\sigma - \lambda_k^o)^2} \mathbf{s}_k^o(\mathbf{s}_k^o)^H \right] \delta_{i,j}.
\end{aligned} \tag{34}
$$

From (32)–(34), two important observations are in order. First, if the signals are both localizable and separable in the t–f domain, then the reduction of the number of signals from n to n_o greatly reduces the estimation error, specifically when the signals are closely spaced. The examples, which are given in the following section, show the advantages of using t-f MUSIC with partially selected signals. The second observation relates to SNR enhancements. The above equations show that error reductions using STFDs are more pronounced for the cases of low SNR and/or closely spaced signals. It is clear from (32)–(34) that when $\lambda_k^o \gg \sigma$ for all $k = 1, 2, \cdots, n_o$, the results are almost independent of L (suppose $N \gg L$ so that $N - L + 1 \simeq N$), and therefore, there would be no obvious improvement in using the STFD over conventional array processing. On the other hand, when some of the eigenvalues are close to σ ($\lambda_k^o \simeq \sigma$ for some $k = 1, 2, \cdots, n_o$), which is the case of weak or closely spaced signals, all the results of above three equations are reduced by a factor of up to $G = L/n_o$, respectively. This factor represents, in essence, the gain achieved from using STFD processing.

IV. Time–Frequency MUSIC

To demonstrate the robustness of the eigen-decomposition of the STFDs when used in practical applications, we consider in this section the recently proposed time-frequency MUSIC (t-f MUSIC) algorithm [7]. The DOA estimation based on time-frequency maximum likelihood (t-f ML) is investigated in [8].

We first recall that the DOAs are estimated in the MUSIC technique by determining the n values of θ for which the following spatial spectrum is maximized [15]:

$$f_{MU}(\theta) = [\mathbf{a}^H(\theta)\hat{\mathbf{G}}\hat{\mathbf{G}}^H\mathbf{a}(\theta)]^{-1} = [\mathbf{a}^H(\theta)(\mathbf{I}-\hat{\mathbf{S}}\hat{\mathbf{S}}^H)\mathbf{a}(\theta)]^{-1} \tag{35}$$

where $\mathbf{a}(\theta)$ is the steering vector corresponding to θ. The variance of those estimates in the conventional MUSIC technique, assuming white noise processes, is given by [9]

$$E(\hat{\omega}_i - \omega_i)^2 = \frac{1}{2N}\frac{\mathbf{a}^H(\theta_i)\mathbf{U}\mathbf{a}(\theta_i)}{h(\theta_i)} \tag{36}$$

where ω_i is the spatial frequency associated with DOA θ_i, and $\hat{\omega}_i$ is its estimate obtained by the conventional MUSIC. In the above equation

$$\begin{aligned}\mathbf{U} &= \sigma\left[\sum_{k=1}^{n}\frac{\lambda_k}{(\sigma-\lambda_k)^2}\mathbf{s}_k\mathbf{s}_k^H\right]\\ \mathbf{d}(\theta_i) &= d\mathbf{a}(\theta_i)/d\omega\\ h(\theta_i) &= \mathbf{d}^H(\theta_i)\mathbf{G}\mathbf{G}^H\mathbf{d}(\theta_i).\end{aligned} \tag{37}$$

From the results of Lemma 1, part b), $\tilde{\mathbf{U}} = \mathbf{U}$, which implies that (36) also holds true when the conventional MUSIC algorithm is applied to FM signals in white noise.

Similarly, for t-f MUSIC with n_o signals selected, the DOAs are determined by locating the n_o peaks of the spatial spectrum defined from the n_o signals' t-f regions.

$$\begin{aligned}f_{MU}^{tf}(\theta) &= [\mathbf{a}^H(\theta)\hat{\mathbf{G}}^{tf}(\hat{\mathbf{G}}^{tf})^H\mathbf{a}(\theta)]^{-1}\\ &= [\mathbf{a}^H(\theta)(\mathbf{I}-\hat{\mathbf{S}}^{tf}(\hat{\mathbf{S}}^{tf})^H)\mathbf{a}(\theta)]^{-1}.\end{aligned} \tag{38}$$

Following the same procedure in [9] and using the results of Lemmas 2 and 3, we obtain the variance of the DOA estimates based on t-f MUSIC

$$E(\hat{\omega}_i^{tf} - \omega_i)^2 = \frac{1}{2(N-L+1)}\frac{\mathbf{a}^H(\theta_i)\mathbf{U}^{tf}\mathbf{a}(\theta_i)}{h^{tf}(\theta_i)} \tag{39}$$

where $\hat{\omega}_i^{tf}$ is the estimate of ω_i obtained by the t-f MUSIC, $\mathbf{U}^{tf}$ is defined in (30), and

$$h^{tf}(\theta_i) = \mathbf{d}^H(\theta_i)\mathbf{G}^{tf}(\mathbf{G}^{tf})^H\mathbf{d}(\theta_i) \tag{40}$$

which is equal to $h(\theta_i)$, if $n_o = n$.

V. Simulation Results

Consider a uniform linear array of eight sensors spaced by half a wavelength and an observation period of 1024 samples. Two chirp signals emitted from two sources positioned at angle θ_1 and θ_2. The start and end frequencies of the signal source at θ_1 are $\omega_{s1} = 0$ and $\omega_{e1} = \pi$, whereas the corresponding two frequencies for the other source at θ_2 are $\omega_{s2} = \pi$ and $\omega_{e2} = 0$, respectively. The noise used in this simulation is zero-mean, Gaussian distributed, and temporally white. The SNR of the ith FM signal is defined as $\mathrm{SNR}_i = 10\log(D_i^2/\sigma)$. Fig. 3 shows the PWVD of the mixed noise-free signals for $L = 129$.

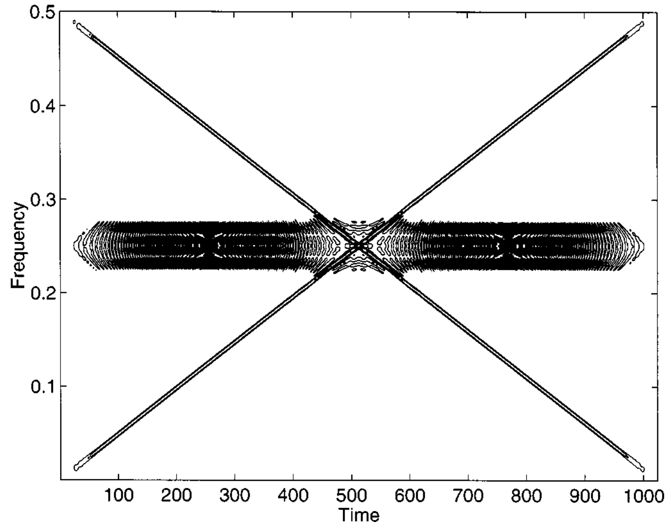

Fig. 3. Pseudo Wigner–Ville distribution of the mixture of the two signals.

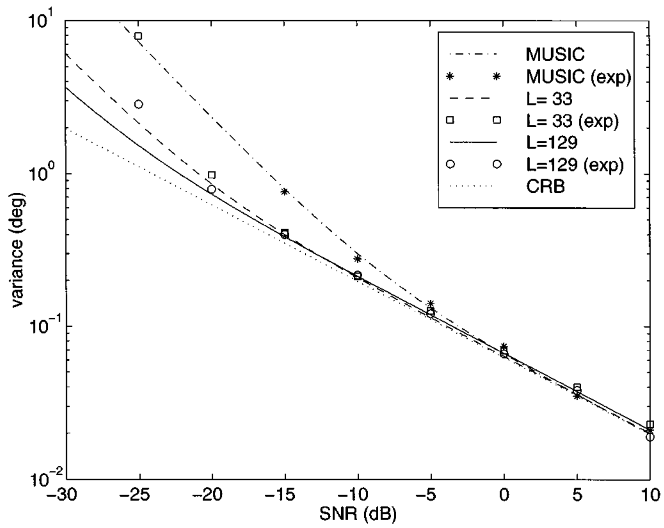

Fig. 4. Variance of DOA estimation versus SNR.

Fig. 4 displays the variance of the estimated DOA $\hat{\theta}_1$ versus SNR for the case $(\theta_1, \theta_2) = (-10°, 10°)$. The curves in this figure show the theoretical and experimental results of the conventional MUSIC and t-f MUSIC (for $L = 33$ and 129). The Cramér–Rao bound (CRB) is also shown in Fig. 4 for comparison. Both signals were selected when performing t-f MUSIC ($n_o = n = 2$). We assume that the number of signals is correctly estimated for each case. Simulation results were averaged over 100 independent trials of Monte Carlo experiments. The advantages of t-f MUSIC in low SNR cases are evident from this figure. The experiment results deviate from the theoretical results for low SNR since we only considered the lowest order of the coefficients of the perturbation expansion of $\check{\mathbf{v}}_i$ in deriving

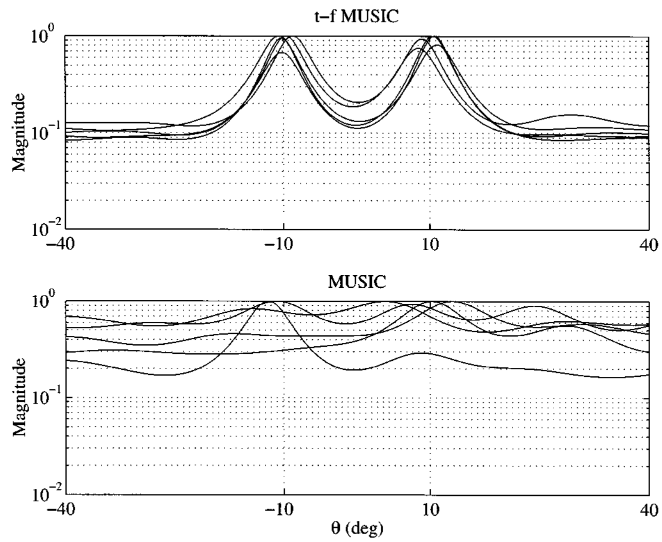

Fig. 5. Estimated spatial spectra ($m = 8$, $N = 1024$, SNR $= -20$ dB, $L = 129$ for t-f MUSIC).

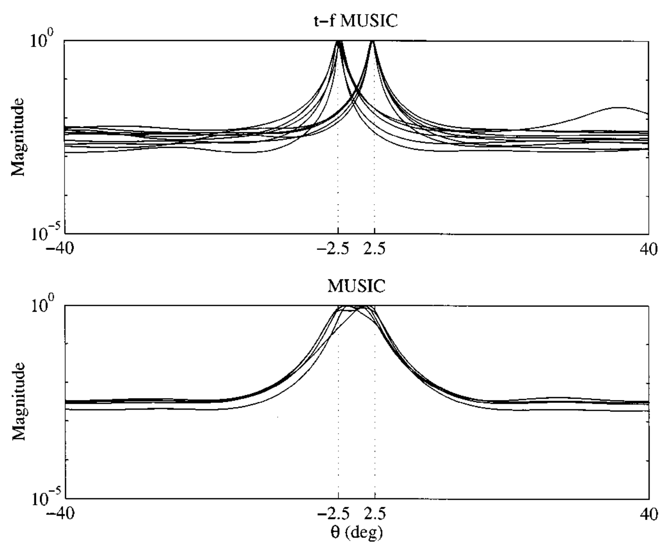

Fig. 6. Estimated spatial spectra for closely spaced signals ($m = 8$, N=1024, SNR $= -5$ dB, $L = 129$ for t-f MUSIC).

the theoretical results (see Appendix A). Fig. 5 shows estimated spatial spectra at SNR $= -20$ dB based on t-f MUSIC ($L = 129$) and the conventional MUSIC. The t-f MUSIC spectral peaks are clearly resolved.

Fig. 6 shows examples of the estimated spatial spectrum based on t-f MUSIC ($L = 129$) and the conventional MUSIC where the angle separation is small ($\theta_1 = -2.5°$, $\theta_2 = 2.5°$). The input SNR is -5 dB. Two t-f MUSIC algorithms are performed using two sets of t-f points, where each set belongs to the t-f signature of one source ($n_o = 1$). It is evident that the two signals cannot be resolved when MUSIC is applied, whereas by applying t-f MUSIC separately for each signal, the two signals become clearly separated, and reasonable DOA estimation is achieved. This is attributed to the signal's distinct t-f signatures. It is noted that there is a small bias in the estimates of t-f MUSIC due to the imperfect separation of the two signals in the t-f domain.

It should be noted that the computation cost used to implement the t-f MUSIC is higher than the conventional MUSIC because it involves the additional processing based on bilinear t-f distributions. Nevertheless, the pseudo Wigner–Ville distribution considered in this paper is relatively simple and only requires a bilinear product and one FFT operation. Moreover, there now exist several computationally efficient t-f kernels that allow TFDs to be provided via spectrogram-based implementations, recursive, and multiplication-free processing. On the issue of practical implementation, many procedures have been devised so that any distribution can be calculated quickly with minimum computer resources [16]. Recently, kernels have been devised, specifically tailored to very fast computation, with the binomial kernel devised by Jeong and Williams being the prime example [17]. Methods for decomposition of kernels, leading to fast computation and increased understanding, have also been carried out by White [18], Amin [19], [20], Venkatesan and Amin [21], and Cunningham and Williams [22].

VI. Conclusions

Subspace analyzes of spatial time-frequency distribution (STFD) matrices have been presented. It has been shown that for signals with clearly defined t-f signatures, such as FM signals, smaller estimation errors in the signal and noise subspaces can be achieved using STFD matrices over the subspace estimates obtained from the data covariance matrix approach. This performance improvement is the result of incorporating the t-f points along the instantaneous frequencies of the impinging signals on the array into the subspace estimation procedure. Under the assumption that the instantaneous frequencies are ideally located, these points belong to autoterm regions of high power concentrations, and as such, when used in constructing STFDs, they provide high SNR matrices with improved eigendecompositions.

The advantages of STFD-based direction finding over traditional direction finding methods using data covariance matrices were demonstrated using the MUSIC algorithm. It was shown that the t-f MUSIC outperforms conventional MUSIC in the two situations of low SNR and closely spaced sources.

Unlike conventional array processing techniques, which are nondiscriminatory, and must therefore spatially localize all signals incident on the array, the STFD-based array processing provides the flexibility of dealing with all signal arrivals or only a subset of them. In this respect, it does not suffer from the drawback of requiring a higher number of sensors than sources. The ability to select fewer sources depends on the differences of their t-f signatures from those of other source signals. The eigenstructure of the STFD matrix constructed from the t-f points that belong to the autoterm regions of a number of sources will only yield the signal subspace of these sources. It was shown that the maximum improvement offered using STFD over data covariance matrices is obtained when constructing the STFD from only one source signal.

Appendix A

For notational simplicity, we denote $\mathbf{v}_i$, $i = 1, 2, \cdots, m$ as the eigenvectors of the correlation matrix $\mathbf{R_{xx}}$, where the first n vectors form the signal subspace ($\mathbf{s}_i$, $i = 1, 2, \cdots, n$), and the last $m - n$ vectors form the noise subspace ($\mathbf{g}_i$, $i = 1, 2, \cdots, m-n$).

To derive the covariance matrices, we follow the same procedure in [9] and [23] but note the fact that the underlying signals are deterministic rather than white random processes, which are considered in [9] and [23]. We define $\hat{\mathbf{R}}_{\mathbf{xx}}$ in terms of a random perturbation to $\mathbf{R}_{\mathbf{xx}}$ with a perturbation factor p, $0 < p \ll 1$. Thus

$$\hat{\mathbf{R}}_{\mathbf{xx}} = \mathbf{R}_{\mathbf{xx}} + (\hat{\mathbf{R}}_{\mathbf{xx}} - \mathbf{R}_{\mathbf{xx}}) = \mathbf{R}_{\mathbf{xx}} + p\mathbf{B}. \tag{A.1}$$

When the source signals are FM and the noise vector forms a multivariate white Gaussian process, then $\mathbf{B}$ is a Hermitian, zero-mean random matrix whose elements are asymptotically jointly Gaussian. Let $\check{\mathbf{v}}_i$ denote the unnormalized perturbed version of the eigenvector $\mathbf{v}_i$. According to [24]

$$\check{\mathbf{v}}_i = \mathbf{v}_i + \sum_{\substack{k=1\\k\neq i}}^{m} \left(\sum_{l=1}^{\infty} t_{lk}^{(i)} p^l\right) \mathbf{v}_k \tag{A.2}$$

where $t_{lk}^{(i)}$, $l = 1, 2, \cdots$ are the coefficients of the perturbation expansion of $\check{\mathbf{v}}_i$ along $\mathbf{v}_k$. By keeping the term with the lowest order of p, then [23]

$$t_{1k}^{(i)} = \frac{\mathbf{v}_k^H \mathbf{B} \mathbf{v}_i}{\lambda_k - \lambda_i}, \qquad k \neq i. \tag{A.3}$$

The mean square value of $t_{1k}^{(i)}$ is given by

$$E\left[\left|t_{1k}^{(i)}\right|^2\right] = E\left[\frac{\mathbf{v}_k^H \mathbf{B} \mathbf{v}_i \mathbf{v}_i^H \mathbf{B} \mathbf{v}_k}{(\lambda_k - \lambda_i)^2}\right]. \tag{A.4}$$

To evaluate the numerator in the above equation, we consider the following general case:

$$\begin{aligned} E[\mathbf{v}_i^H \mathbf{B} \mathbf{v}_j \mathbf{v}_k^H \mathbf{B} \mathbf{v}_l] &= \frac{1}{p^2} E[\mathbf{v}_i^H (\hat{\mathbf{R}}_{\mathbf{xx}} - \mathbf{R}_{\mathbf{xx}}) \mathbf{v}_j \mathbf{v}_k^H (\hat{\mathbf{R}}_{\mathbf{xx}} - \mathbf{R}_{\mathbf{xx}}) \mathbf{v}_l] \\ &= \frac{1}{(Np)^2} E\left[\left(\sum_{r=1}^{N} \mathbf{v}_i^H \mathbf{x}(t_r) \mathbf{x}^H(t_r) \mathbf{v}_j\right) \times \left(\sum_{q=1}^{N} \mathbf{v}_k^H \mathbf{x}(t_q) \mathbf{x}^H(t_q) \mathbf{v}_l\right)\right] \\ &\quad - \frac{1}{p^2} \mathbf{v}_i^H \mathbf{R}_{\mathbf{xx}} \mathbf{v}_j \mathbf{v}_k^H \mathbf{R}_{\mathbf{xx}} \mathbf{v}_l. \end{aligned} \tag{A.5}$$

It can be easily realized that the expected value in (A.5) is taken from a product of four nonzero mean Gaussian random variables. It is well known that for Gaussian random variable x_1, x_2, x_3, x_4 with nonzero means

$$\begin{aligned} E[x_1x_2x_3x_4] &= E[x_1x_2]E[x_3x_4] + E[x_1x_3]E[x_2x_4] \\ &\quad + E[x_1x_4]E[x_2x_3] - 2E[x_1]E[x_2]E[x_3]E[x_4]. \end{aligned} \tag{A.6}$$

Using the properties of the zero-mean circular complex Gaussian noise vector and the deterministic source signal vector

$$\begin{aligned} E[\mathbf{x}(t_r)] &= \mathbf{y}(t_r) \\ E[\mathbf{x}(t_r)\mathbf{x}^H(t_q)] &= \mathbf{y}(t_r)\mathbf{y}^H(t_q) + \sigma \mathbf{I} \delta_{r,q} \\ E[\mathbf{x}(t_r)\mathbf{x}^T(t_q)] &= \mathbf{y}(t_r)\mathbf{y}^T(t_q). \end{aligned}$$

Accordingly, (A.5) can be written as

$$\begin{aligned} &E[\mathbf{v}_i^H \mathbf{B} \mathbf{v}_j \mathbf{v}_k^H \mathbf{B} \mathbf{v}_l] \\ &= \frac{1}{(Np)^2} \sum_{r=1}^{N}\sum_{q=1}^{N} E[\mathbf{v}_i^H \mathbf{x}(t_r) \mathbf{x}^H(t_r) \mathbf{v}_j] \times E[\mathbf{v}_k^H \mathbf{x}(t_q) \mathbf{x}^H(t_q) \mathbf{v}_l] \\ &\quad + \frac{1}{(Np)^2} \sum_{r=1}^{N}\sum_{q=1}^{N} E[\mathbf{v}_i^H \mathbf{x}(t_r) \mathbf{v}_k^H \mathbf{x}(t_q)] \times E[\mathbf{x}^H(t_r) \mathbf{v}_j \mathbf{x}^H(t_q) \mathbf{v}_l] \\ &\quad + \frac{1}{(Np)^2} \sum_{r=1}^{N}\sum_{q=1}^{N} E[\mathbf{v}_i^H \mathbf{x}(t_r) \mathbf{x}^H(t_q) \mathbf{v}_l] \times E[\mathbf{v}_k^H \mathbf{x}(t_q) \mathbf{x}^H(t_r) \mathbf{v}_j] \\ &\quad - 2\frac{1}{(Np)^2} \sum_{r=1}^{N}\sum_{q=1}^{N} E[\mathbf{v}_i^H \mathbf{x}(t_r)] E[\mathbf{x}^H(t_r) \mathbf{v}_j] \times E[\mathbf{v}_k^H \mathbf{x}(t_q)] E[\mathbf{x}^H(t_q) \mathbf{v}_l] \\ &\quad - \frac{1}{p^2} \mathbf{v}_i^H \mathbf{R}_{\mathbf{xx}} \mathbf{v}_j \mathbf{v}_k^H \mathbf{R}_{\mathbf{xx}} \mathbf{v}_l \\ &= \frac{1}{(Np)^2} \sum_{r=1}^{N}\sum_{q=1}^{N} [\mathbf{v}_i^H \mathbf{v}_l \sigma \delta_{r,q} \mathbf{v}_k^H \mathbf{y}(t_q) \mathbf{y}^H(t_r) \mathbf{v}_j] \\ &\quad + \frac{1}{(Np)^2} \sum_{r=1}^{N}\sum_{q=1}^{N} [\mathbf{v}_i^H \mathbf{y}(t_r) \mathbf{y}^H(t_q) \mathbf{v}_l \mathbf{v}_k^H \mathbf{v}_j \sigma \delta_{r,q}] \\ &\quad + \frac{1}{(Np)^2} \sum_{r=1}^{N}\sum_{q=1}^{N} [\mathbf{v}_i^H \mathbf{v}_l \mathbf{v}_k^H \mathbf{v}_j \sigma^2 \delta_{r,q}] \\ &= \frac{1}{(Np)^2} \sum_{r=1}^{N} [\delta_{i,l} \sigma \mathbf{v}_k^H \mathbf{y}(t_r) \mathbf{y}^H(t_r) \mathbf{v}_j] \\ &\quad + \frac{1}{(Np)^2} \sum_{r=1}^{N} [\mathbf{v}_i^H \mathbf{y}(t_r) \mathbf{y}^H(t_r) \mathbf{v}_l \delta_{j,k} \sigma] \\ &\quad + \frac{1}{(Np)^2} \sum_{r=1}^{N} [\delta_{i,l} \delta_{j,k} \sigma^2]. \end{aligned} \tag{A.7}$$

By using the uncorrelation assumption (7)

$$\begin{aligned} \frac{1}{N} \sum_{r=1}^{N} \mathbf{y}(t_r) \mathbf{y}^H(t_r) &= \mathbf{A} \left[\frac{1}{N} \sum_{r=1}^{N} \mathbf{d}(t_r) \mathbf{d}^H(t_r)\right] \mathbf{A}^H \\ &= \mathbf{A} \mathbf{R}_{\mathbf{dd}} \mathbf{A}^H = \mathbf{R}_{\mathbf{yy}} = \mathbf{R}_{\mathbf{xx}} - \sigma \mathbf{I} \end{aligned} \tag{A.8}$$

(A.7) simplifies to

$$\begin{aligned} &E[\mathbf{v}_i^H \mathbf{B} \mathbf{v}_j \mathbf{v}_k^H \mathbf{B} \mathbf{v}_l] \\ &= \frac{\sigma}{Np^2} [\delta_{i,l} \mathbf{v}_k^H \mathbf{R}_{\mathbf{yy}} \mathbf{v}_j + \mathbf{v}_i^H \mathbf{R}_{\mathbf{yy}} \mathbf{v}_l \delta_{j,k}] + \frac{\sigma^2}{Np^2} \delta_{i,l} \delta_{j,k} \\ &= \frac{\sigma}{Np^2} [\delta_{i,l} \mathbf{v}_k^H (\mathbf{R}_{\mathbf{xx}} - \sigma \mathbf{I}) \mathbf{v}_j + \mathbf{v}_i^H (\mathbf{R}_{\mathbf{xx}} - \sigma \mathbf{I}) \mathbf{v}_l \delta_{j,k}] \\ &\quad + \frac{\sigma^2}{Np^2} \delta_{i,l} \delta_{j,k} \\ &= \frac{\sigma}{Np^2} [\delta_{i,l} \mathbf{v}_k^H \mathbf{R}_{\mathbf{xx}} \mathbf{v}_j + \mathbf{v}_i^H \mathbf{R}_{\mathbf{xx}} \mathbf{v}_l \delta_{j,k} - \sigma \delta_{i,l} \delta_{j,k}] \\ &= \frac{\sigma}{Np^2} [\lambda_i + \lambda_j - \sigma] \delta_{i,l} \delta_{j,k}. \end{aligned} \tag{A.9}$$

Therefore

$$E\left[\left|t_{1k}^{(i)}\right|^2\right] = E\left[\frac{\mathbf{v}_k^H \mathbf{B}\mathbf{v}_i \mathbf{v}_i^H \mathbf{B}\mathbf{v}_k}{(\lambda_k - \lambda_i)^2}\right] = \frac{\sigma}{Np^2}\,\frac{(\lambda_i + \lambda_k - \sigma)}{(\lambda_i - \lambda_k)^2}, \qquad k \neq i \quad \text{(A.10)}$$

and

$$E\left[t_{1k}^{(i)}\left(t_{1k}^{(j)}\right)^*\right] = E\left[\frac{\mathbf{v}_k^H \mathbf{B}\mathbf{v}_i \mathbf{v}_j^H \mathbf{B}\mathbf{v}_k}{(\lambda_k - \lambda_i)(\lambda_k - \lambda_j)}\right] = 0, \qquad k \neq i, \quad k \neq j. \quad \text{(A.11)}$$

It is shown in [13] that

$$cov(\hat{\mathbf{v}}_i, \hat{\mathbf{v}}_j) = cov(\check{\mathbf{v}}_i, \check{\mathbf{v}}_j) + o(N^{-2}).$$

By ignoring the terms of N^{-2}, then

$$\begin{aligned} cov(\hat{\mathbf{v}}_i, \hat{\mathbf{v}}_j) &\simeq cov(\check{\mathbf{v}}_i, \check{\mathbf{v}}_j) \\ &\simeq E\left[\left(\sum_{\substack{k=1\\k\neq i}}^{m} t_{1k}^{(i)} p\mathbf{v}_k\right)\left(\sum_{\substack{k=1\\k\neq j}}^{m} t_{1k}^{(j)} p\mathbf{v}_k\right)^H\right] \\ &= E\left[\sum_{\substack{k=1\\k\neq i}}^{m} \left|t_{1k}^{(i)}\right|^2 p^2 \mathbf{v}_k \mathbf{v}_k^H \delta_{i,j}\right] \\ &= \frac{\sigma}{N}\sum_{\substack{k=1\\k\neq i}}^{m} \frac{\lambda_i + \lambda_k - \sigma}{(\lambda_i - \lambda_k)^2}\,\mathbf{v}_k \mathbf{v}_k^H \delta_{i,j}. \end{aligned} \quad \text{(A.12)}$$

Replacing $\mathbf{v}_k$ by $\mathbf{s}_k$ or $\mathbf{g}_k$ leads to (8). Similarly

$$\begin{aligned} cov(\hat{\mathbf{v}}_i, \hat{\mathbf{v}}_j^*) &\simeq cov(\check{\mathbf{v}}_i, \check{\mathbf{v}}_j^*) \\ &\simeq E\left[\left(\sum_{\substack{k=1\\k\neq i}}^{m} t_{1k}^{(i)} p\mathbf{v}_k\right)\left(\sum_{\substack{k=1\\k\neq j}}^{m} t_{1k}^{(j)} p\mathbf{v}_k\right)^T\right] \\ &= p^2 E\left[\sum_{\substack{k_1=1\\k_1\neq i}}^{m}\sum_{\substack{k_2=1\\k_2\neq j}}^{m} t_{1k_1}^{(i)} t_{1k_2}^{(j)} \mathbf{v}_{k_1}\mathbf{v}_{k_2}^T\right] \\ &= p^2 \sum_{\substack{k_1=1\\k_1\neq i}}^{m}\sum_{\substack{k_2=1\\k_2\neq j}}^{m} E\left[\frac{\mathbf{v}_{k_1}^H \mathbf{B}\mathbf{v}_i}{(\lambda_{k_1}-\lambda_i)}\,\frac{\mathbf{v}_{k_2}^H \mathbf{B}\mathbf{v}_j}{(\lambda_{k_2}-\lambda_j)}\,\mathbf{v}_{k_1}\mathbf{v}_{k_2}^T\right]. \end{aligned} \quad \text{(A.13)}$$

From (A.9), it is clear that the above equation has nonzero value only when $k_1 = j$ and $k_2 = i$. Noting the fact that $k_1 \neq i$ and $k_2 \neq j$, (A.13) becomes

$$\begin{aligned} cov(\hat{\mathbf{v}}_i, \hat{\mathbf{v}}_j^*) &\simeq -\frac{p^2}{(\lambda_j-\lambda_i)^2}\,E[\mathbf{v}_j^H \mathbf{B}\mathbf{v}_i \mathbf{v}_i^H \mathbf{B}\mathbf{v}_j]\mathbf{v}_j \mathbf{v}_i^T (1-\delta_{i,j}) \\ &= -\frac{\sigma}{N}\,\frac{\lambda_i + \lambda_j - \sigma}{(\lambda_j - \lambda_i)^2}\,\mathbf{v}_j \mathbf{v}_i^T (1-\delta_{i,j}). \end{aligned} \quad \text{(A.14)}$$

For the signal subspace, $\mathbf{v}_i$ is $\mathbf{s}_i$, $i = 1, \cdots, n$, and (A.14) yields (9), and this concludes the proof for part a) of Lemma 1.

APPENDIX B

Using eigendecomposition theory, we have

$$\mathbf{A}^o \mathbf{R}_{\mathbf{dd}}^o (\mathbf{A}^o)^H = \mathbf{Q}\mathbf{\Xi}\mathbf{Q}^H \quad \text{(B.1)}$$

where $\mathbf{\Xi} = \text{diag}[\xi_1, \cdots, \xi_{n_o}, 0, \cdots, 0]$ is a diagonal matrix whose elements are the eigenvalues of $\mathbf{A}^o \mathbf{R}_{\mathbf{dd}}^o (\mathbf{A}^o)^H$, and $\mathbf{Q}$ is the corresponding eigenvector matrix. Clearly, $\xi_i = \lambda_i^o - \sigma$, $i = 1, \cdots, n_o$, and $\mathbf{Q} = [\mathbf{S}^o | \mathbf{G}^o]$.

From the definition of λ_i^o and λ_i^{tf}, it is evident that

$$\mathbf{D} = \frac{L}{n_o}\mathbf{Q}\mathbf{\Xi}\mathbf{Q}^H + \sigma\mathbf{I} = \mathbf{Q}\left[\frac{L}{n_o}\mathbf{\Xi} + \sigma\mathbf{I}\right]\mathbf{Q}^H. \quad \text{(B.2)}$$

Therefore, $\mathbf{R}_{\mathbf{xx}}^o$ and $\mathbf{D}$ share the same set of eigenvectors, which proves part a) of Lemma 2. The ith eigenvalue of $\mathbf{D}$ is $(L/n_o)\xi_i + \sigma = (L/n_o)(\lambda_i^o - \sigma) + \sigma$ for $i \leq n_o$ and is σ for $n_o < i \leq m$, subsequently leading to part b).

APPENDIX C

Similar to Appendix A, we let $\mathbf{v}_i, i = 1, 2, \cdots, m$ represent the whole eigenvectors of the STFD matrix $\mathbf{D}$, where the first n_o vectors form the signal subspace ($\mathbf{s}_i^{tf}$, $i = 1, 2, \cdots, n_o$), whereas the last $m - n_o$ vectors form the noise subspace ($\mathbf{g}_i^{tf}$, $i = 1, 2, \cdots, m - n_o$). As discussed in Section III, we assume that the selected t-f points belong to regions where no crossterm components are present.

For an array mixture of FM signals, we select points from n_o signals at the t-f domain, where the pseudo Wigner–Ville distribution matrix is defined in (13). We define $\hat{\mathbf{D}}$ in terms of a random perturbation to $\mathbf{D}$ with a perturbation factor p, $0 < p \ll 1$. Thus

$$\hat{\mathbf{D}} = \mathbf{D} + (\hat{\mathbf{D}} - \mathbf{D}) = \mathbf{D} + p\mathbf{B}. \quad \text{(C.1)}$$

Matrix $\mathbf{B}$ is a Hermitian, zero-mean random matrix whose elements are asymptotically jointly Gaussian [8]. Similar to Appendix A, we derive

$$\begin{aligned} &E[\mathbf{v}_i^H \mathbf{B}\mathbf{v}_j \mathbf{v}_k^H \mathbf{B}\mathbf{v}_l] \\ &\quad= \frac{1}{p^2}E[\mathbf{v}_i^H(\hat{\mathbf{D}}-\mathbf{D})\mathbf{v}_j \mathbf{v}_k^H(\hat{\mathbf{D}}-\mathbf{D})\mathbf{v}_l] \\ &\quad= \frac{1}{(n_o p(N-L+1))^2} \\ &\qquad\times E\left[\left(\sum_{q=1}^{n_o}\sum_{i=1}^{N-L+1}\mathbf{v}_i^H \mathbf{D}_{\mathbf{xx}}(t_i, f_{q,i})\mathbf{v}_j\right)\right. \\ &\qquad\qquad\left.\times\left(\sum_{q=1}^{n_o}\sum_{i=1}^{N-L+1}\mathbf{v}_k^H \mathbf{D}_{\mathbf{xx}}(t_i, f_{q,i})\mathbf{v}_l\right)\right] \\ &\qquad- \frac{1}{p^2}\mathbf{v}_i^H \mathbf{D}\mathbf{v}_j \mathbf{v}_k^H \mathbf{D}\mathbf{v}_l. \end{aligned} \quad \text{(C.2)}$$

Substituting (21) and (A.7) into (C.2), we obtain

$$\begin{aligned}
&E[\mathbf{v}_i^H \mathbf{B} \mathbf{v}_j \mathbf{v}_k^H \mathbf{B} \mathbf{v}_l] \\
&= \frac{1}{(n_o p(N-L+1))^2} \sum_{q_1=1}^{n_o} \sum_{q_2=1}^{n_o} \sum_{i_1=1}^{N-L+1} \sum_{i_2=1}^{N-L+1} \\
&\quad \times \sum_{\tau_1=-((L-1)/2)}^{(L-1)/2} \sum_{\tau_2=-((L-1)/2)}^{(L-1)/2} e^{-j4\pi[f_{q_1,i_1}\tau_1 + f_{q_2,i_2}\tau_2]} \\
&\quad \times \{E[\mathbf{v}_i^H \mathbf{x}(t_{i_1}+\tau_1)\mathbf{x}^H(t_{i_1}-\tau_1)\mathbf{v}_j] \\
&\qquad \times E[\mathbf{v}_k^H \mathbf{x}(t_{i_2}+\tau_2)\mathbf{x}^H(t_{i_2}-\tau_2)\mathbf{v}_l] \\
&\qquad + E[\mathbf{v}_i^H \mathbf{x}(t_{i_1}+\tau_1)\mathbf{v}_k^H \mathbf{x}(t_{i_2}+\tau_2)] \\
&\qquad \times E[\mathbf{x}^H(t_{i_1}-\tau_1)\mathbf{v}_j \mathbf{x}^H(t_{i_2}-\tau_2)\mathbf{v}_l] \\
&\qquad + E[\mathbf{v}_i^H \mathbf{x}(t_{i_1}+\tau_1)\mathbf{x}^H(t_{i_2}-\tau_2)\mathbf{v}_l] \\
&\qquad \times E[\mathbf{v}_k^H \mathbf{x}(t_{i_2}+\tau_2)\mathbf{x}^H(t_{i_1}-\tau_1)\mathbf{v}_j] \\
&\qquad - 2E[\mathbf{v}_i^H \mathbf{x}(t_{i_1}+\tau_1)]E[\mathbf{x}^H(t_{i_1}-\tau_1)\mathbf{v}_j] \\
&\qquad \times E[\mathbf{v}_k^H \mathbf{x}(t_{i_2}+\tau_2)]E[\mathbf{x}^H(t_{i_2}-\tau_2)\mathbf{v}_l]\} \\
&\quad - \frac{1}{p^2}\mathbf{v}_i^H \mathbf{D} \mathbf{v}_j \mathbf{v}_k^H \mathbf{v}_l \\
&= \frac{1}{(n_o p(N-L+1))^2} \sum_{q_1=1}^{n_o} \sum_{q_2=1}^{n_o} \sum_{i_1=1}^{N-L+1} \sum_{i_2=1}^{N-L+1} \\
&\quad \times \sum_{\tau_1=-((L-1)/2)}^{(L-1)/2} \sum_{\tau_2=-((L-1)/2)}^{(L-1)/2} e^{-j4\pi[f_{q_1,i_1}\tau_1 + f_{q_2,i_2}\tau_2]} \\
&\quad \times [\mathbf{v}_i^H \mathbf{y}(t_{i_1}+\tau_1)\mathbf{y}^H(t_{i_2}-\tau_2)\mathbf{v}_l \sigma \delta_{j,k}\delta_{t_{i_1}-\tau_1, t_{i_2}+\tau_2} \\
&\qquad + \sigma\delta_{i,l}\delta_{t_{i_1}+\tau_1, t_{i_2}-\tau_2}\mathbf{v}_k^H \mathbf{y}(t_{i_2}+\tau_2)\mathbf{y}^H(t_{i_1}-\tau_1)\mathbf{v}_j \\
&\qquad + \sigma^2 \delta_{i,l}\delta_{j,k}\delta_{t_{i_1},t_{i_2}}\delta_{\tau_1,\tau_2}].
\end{aligned} \tag{C.3}$$

Under the assumption of no crossterms, q_1 should be equivalent to q_2 to have nonzero values, and in this case, $q_1 = q_2 = q$. Note that within the t-f region of the qth signal, $\mathbf{y}(t) = \mathbf{y}_q(t) \triangleq \mathbf{A}\mathbf{d}_q(t)$. When the third-order derivative of the phase is negligible over $[t-L+1, t+L-1]$ for any signal and any t, we have

$$\begin{aligned}
&E[\mathbf{v}_i^H \mathbf{B} \mathbf{v}_j \mathbf{v}_k^H \mathbf{B} \mathbf{v}_l] \\
&\simeq \frac{1}{(Np(N-L+1))^2} \sum_{q=1}^{n_o} \sum_{i_1=1}^{N-L+1} \sum_{i_2=1}^{N-L+1} \\
&\quad \times \sum_{\tau_1=-((L-1)/2)}^{(L-1)/2} \sum_{\tau_2=-((L-1)/2)}^{(L-1)/2} \\
&\quad \times [\mathbf{v}_i^H \mathbf{y}(t_{i_1}+\tau_1)\mathbf{y}^H(t_{i_2}-\tau_2)\mathbf{v}_l \sigma\delta_{j,k}\delta_{t_{i_1}-\tau_1, t_{i_2}+\tau_2} \\
&\qquad + \sigma\delta_{i,l}\delta_{t_{i_1}+\tau_1, t_{i_1}-\tau_1}\mathbf{v}_k^H \mathbf{y}(t_{i_2}+\tau_2)\mathbf{y}^H(t_{i_1}-\tau_1)\mathbf{v}_j] \\
&\quad + \frac{\sigma^2 L}{n_o(N-L+1)p^2}\delta_{i,l}\delta_{j,k} \\
&\simeq \frac{\sigma L}{n_o(N-L+1)p^2}\left[\frac{L}{n_o}(\lambda_i - \sigma) + (\lambda_j - \sigma) + \sigma\right]\delta_{i,l}\delta_{j,k} \\
&= \frac{\sigma L}{n_o(N-L+1)p^2}[(\lambda_i^{tf} + \lambda_j^{tf} - \sigma)]\delta_{i,l}\delta_{j,k}.
\end{aligned} \tag{C.4}$$

Let $\check{\mathbf{v}}_i$ denote the unnormalized eigenvector given in a perturbation expansions by

$$\check{\mathbf{v}}_i = \mathbf{v}_i + \sum_{\substack{k=1\\k\neq i}}^{m}\left(\sum_{l=1}^{\infty} t_{lk}^{(i)} p^l\right)\mathbf{v}_k \tag{C.5}$$

where $t_{lk}^{(i)}$, $l = 1, 2, \cdots$ are the coefficients of the perturbation expansion of $\check{\mathbf{v}}_i$ along $\mathbf{v}_k$, and keeping the term with the lowest order of p, then

$$t_{1k}^{(i)} = \frac{\mathbf{v}_k^H \mathbf{B} \mathbf{v}_i}{\lambda_k^{tf} - \lambda_i^{tf}}, \qquad k \neq i. \tag{C.6}$$

Therefore

$$\begin{aligned}
E\left[\left|t_{1k}^{(i)}\right|^2\right] &= E\left[\frac{\mathbf{v}_k^H \mathbf{B} \mathbf{v}_i \mathbf{v}_i^H \mathbf{B} \mathbf{v}_k}{(\lambda_k^{tf} - \lambda_i^{tf})^2}\right] \\
&= \frac{\sigma L}{n_o(N-L+1p^2}\,\frac{(\lambda_i^{tf} + \lambda_k^{tf} - \sigma)}{(\lambda_i^{tf} - \lambda_k^{tf})^2}, \qquad k \neq i
\end{aligned} \tag{C.7}$$

and

$$\begin{aligned}
E\left[t_{1k}^{(i)}\left(t_{1k}^{(j)}\right)^*\right] &= E\left[\frac{\mathbf{v}_k^H \mathbf{B} \mathbf{v}_i \mathbf{v}_j^H \mathbf{B} \mathbf{v}_k}{(\lambda_k^{tf} - \lambda_i^{tf})(\lambda_k^{tf} - \lambda_j^{tf})}\right] \\
&= 0, \qquad k \neq i, \quad k \neq j.
\end{aligned} \tag{C.8}$$

Similar to Appendix A, we follow

$$\begin{aligned}
cov(\hat{\mathbf{v}}_i, \hat{\mathbf{v}}_j) &\simeq cov(\check{\mathbf{v}}_i, \check{\mathbf{v}}_j) \\
&\simeq E\left[\left(\sum_{\substack{k=1\\k\neq i}}^{m} t_{1k}^{(i)} p \mathbf{v}_k\right)\left(\sum_{\substack{k=1\\k\neq j}}^{m} t_{1k}^{(j)} p \mathbf{v}_k\right)^H\right] \\
&= E\left[\sum_{\substack{k=1\\k\neq i}}^{m}\left|t_{1k}^{(i)}\right|^2 p^2 \mathbf{v}_k \mathbf{v}_k^H \delta_{i,j}\right] \\
&= \frac{\sigma L}{n_o(N-L+1)}\sum_{\substack{k=1\\k\neq i}}^{m}\frac{\lambda_i^{tf} + \lambda_k^{tf} - \sigma}{(\lambda_i^{tf} - \lambda_k^{tf})^2}\mathbf{v}_k \mathbf{v}_k^H \delta_{i,j}.
\end{aligned} \tag{C.9}$$

Equation (28) follows by properly replacing $\mathbf{v}_k$ by $\mathbf{s}_k^{tf}$ or $\mathbf{g}_k^{tf}$. Similarly

$$\begin{aligned}
&cov(\hat{\mathbf{v}}_i, \hat{\mathbf{v}}_j^*) \\
&\quad \simeq cov(\check{\mathbf{v}}_i, \check{\mathbf{v}}_j^*) \\
&\quad \simeq E\left[\left(\sum_{\substack{k=1\\k\neq i}}^{m} t_{1k}^{(i)} p \mathbf{v}_k\right)\left(\sum_{\substack{k=1\\k\neq j}}^{m} t_{1k}^{(j)} p \mathbf{v}_k\right)^T\right] \\
&\quad = p^2 \sum_{\substack{k_1=1\\k_1\neq i}}^{m}\sum_{\substack{k_2=1\\k_2\neq j}}^{m} E\left[\frac{\mathbf{v}_{k_1}^H \mathbf{B} \mathbf{v}_i}{(\lambda_{k_1}^{tf} - \lambda_i^{tf})}\,\frac{\mathbf{v}_{k_2}^H \mathbf{B} \mathbf{v}_j}{(\lambda_{k_2}^{tf} - \lambda_j^{tf})}\mathbf{v}_{k_1}\mathbf{v}_{k_2}^T\right] \\
&\quad = -\frac{p^2}{(\lambda_j^{tf} - \lambda_i^{tf})^2} E[\mathbf{v}_j^H \mathbf{B} \mathbf{v}_i \mathbf{v}_i^H \mathbf{B} \mathbf{v}_j]\mathbf{v}_j \mathbf{v}_i^T (1 - \delta_{i,j}) \\
&\quad = -\frac{\sigma L}{n_o(N-L+1)}\,\frac{\lambda_i^{tf} + \lambda_j^{tf} - \sigma}{(\lambda_j^{tf} - \lambda_i^{tf})^2}\mathbf{v}_j \mathbf{v}_i^T (1 - \delta_{i,j}).
\end{aligned} \tag{C.10}$$

For the columns of signal subspace, $\mathbf{v}_i$ becomes $\mathbf{s}_i^{tf}$, and (C.10) becomes (29).

APPENDIX D

This Appendix follows the procedure of [9]. Denote

$$\mathbf{\Gamma} = (\mathbf{S}^{tf})^H \hat{\mathbf{D}} \mathbf{G}^{tf}$$

and γ_i the ith column of $\mathbf{\Gamma}$. Using the results of (C.2)–(C.4) and the fact $(\mathbf{S}^{tf})^H \mathbf{D} \mathbf{G}^{tf} = \mathbf{0}$, we have

$$\begin{aligned} E[\gamma_i \gamma_j^H]_{t,q} &= E[((\mathbf{s}_t^{tf})^H \hat{\mathbf{D}} \mathbf{g}_i^{tf})((\mathbf{g}_j^{tf})^H \hat{\mathbf{D}} \mathbf{s}_q^{tf})] \\ &= \frac{\sigma L}{n_o(N-L+1)} \lambda_t^{tf} \delta_{t,q} \delta_{i,j}. \end{aligned} \tag{D.1}$$

Subsequently

$$\begin{aligned} E[\gamma_i \gamma_j^H] &= E[((\mathbf{S}^{tf})^H \hat{\mathbf{D}} \mathbf{g}_i^{tf})((\mathbf{g}_j^{tf})^H \hat{\mathbf{D}} \mathbf{S}^{tf})] \\ &= \frac{\sigma L}{n_o(N-L+1)} \mathbf{\Lambda}^{tf} \delta_{i,j} \end{aligned} \tag{D.2}$$

where $\mathbf{\Lambda}^{tf} = \text{diag}[\lambda_1^{tf}, \cdots, \lambda_{n_o}^{tf}]$. Similarly,

$$E[\gamma_i \gamma_j^T]_{t,q} = E[((\mathbf{s}_t^{tf})^H \hat{\mathbf{D}} \mathbf{g}_i^{tf})((\mathbf{s}_q^{tf})^H \hat{\mathbf{D}} \mathbf{g}_j^{tf})] = 0 \tag{D.3}$$

and subsequently

$$E[\gamma_i \gamma_j^T] = \mathbf{0}. \tag{D.4}$$

Since $\mathbf{S}^{tf}(\mathbf{S}^{tf})^H \hat{\mathbf{g}}_i^{tf}$ has the same limiting distribution as that of $-\mathbf{S}^{tf}(\mathbf{\Gamma} - \sigma \mathbf{I})^{-1} \gamma_i$ [8], then it follows that

$$\begin{aligned} &E(\mathbf{S}^{tf}(\mathbf{S}^{tf})^H \hat{\mathbf{g}}_i^{tf})(\mathbf{S}^{tf}(\mathbf{S}^{tf})^H \hat{\mathbf{g}}_j^{tf})^H \\ &\quad = \frac{\sigma L}{n_o(N-L+1)} [\mathbf{S}(\mathbf{\Lambda}^{tf} - \sigma \mathbf{I})^{-1} \mathbf{\Lambda}^{tf} (\mathbf{\Lambda}^{tf} - \sigma \mathbf{I})^{-1} \mathbf{S}^H] \delta_{i,j} \\ &\quad = \frac{\sigma L}{n_o(N-L+1)} \left[\sum_{k=1}^{n_o} \frac{\lambda_k^{tf}}{(\sigma - \lambda_k^{tf})^2} \mathbf{s}_k^{tf} (\mathbf{s}_k^{tf})^H \right] \delta_{i,j} \end{aligned} \tag{D.5}$$

and

$$\begin{aligned} &E(\mathbf{S}^{tf}(\mathbf{S}^{tf})^H \hat{\mathbf{g}}_i^{tf})(\mathbf{S}^{tf}(\mathbf{S}^{tf})^H \hat{\mathbf{g}}_j^{tf})^T \\ &\quad = \mathbf{0} \qquad \text{for all } i, j. \end{aligned} \tag{D.6}$$

ACKNOWLEDGMENT

The authors would like to thank Dr. K. Yang of ATR Adaptive Communications Research Laboratories, Japan, for his helpful discussions on this subject during his visit to Villanova University.

REFERENCES

[1] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[2] S. Qian and D. Chen, *Joint Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[3] B. Boashash, "Time-frequency signal analysis," in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1990, vol. I.
[4] L. Cohen, "Time-frequency signal analysis—A review," *Proc. IEEE*, vol. 77, pp. 941–981, July 1989.
[5] A. Belouchrani and M. Amin, "Blind source separation based on time-frequency signal representation," *IEEE Trans. Signal Processing*, vol. 46, pp. 2888–2898, Nov. 1998.
[6] M. G. Amin, "Spatial time-frequency distributions for direction finding and blind source separation," *Proc. SPIE: Wavelet Applications IV*, vol. 3723, pp. 62–70, Apr. 1999.
[7] A. Belouchrani and M. Amin, "Time-frequency MUSIC," *IEEE Signal Processing Lett.*, vol. 6, pp. 109–110, May 1999.
[8] Y. Zhang, W. Mu, and M. G. Amin, "Time-frequency maximum likelihood methods for direction finding," *J. Franklin Inst.*, vol. 337, no. 4, pp. 483–497, July 2000.
[9] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer–Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.
[10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[11] K. Sekihara, S. Nagarajan, D. Poeppel, and Y. Miyashita, "Time-frequency MEG-MUSIC algorithm," *IEEE Trans. Med. Imag.*, vol. 18, pp. 92–97, Jan. 1999.
[12] J. E. Hudson, *Adaptive Array Principles*. London, U.K.: Peter Peregrinus, 1981.
[13] L. C. Godara, "Application of antenna arrays to mobile communications—Part II: Beam-forming and direction-of-arrival considerations," *Proc. IEEE*, vol. 85, pp. 1195–1245, Aug. 1997.
[14] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.
[15] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276–280, Mar. 1986.
[16] L. Cohen, "Recent developments in the core of digital signal processing: Time-frequency analysis," *IEEE Signal Processing Mag.*, vol. 16, pp. 22–28, Jan. 1999.
[17] J. Jeong and W. Williams, "Kernel design for reduced interference distributions," *IEEE Trans. Signal Processing*, vol. 40, pp. 402–412, Feb. 1992.
[18] L. White, "Transition kernels for bilinear time-frequency signal representations," *IEEE Trans. Signal Processing*, vol. 39, pp. 542–544, Feb. 1991.
[19] M. Amin, "Spectral decomposition of the time-frequency distribution kernels," *IEEE Trans. Signal Processing*, vol. 42, pp. 1156–1166, May 1994.
[20] ——, "Recursive kernels for time-frequency signal representations," *IEEE Signal Processing Lett.*, vol. 3, pp. 16–18, Jan. 1996.
[21] G. Venkatesan and M. Amin, "Time-frequency distribution kernel design over a discrete powers-of-two space," *IEEE Signal Processing Lett.*, vol. 3, pp. 305–306, Dec. 1996.
[22] G. Cunningham and W. Williams, "Fast implementation of generalized discrete time-frequency distributions," *IEEE Trans. Signal Processing*, vol. 42, pp. 1496–1508, June 1994.
[23] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 331–340, Apr. 1986.
[24] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. New York: Oxford Univ. Press, 1965.

Yimin Zhang (M'88) received the M.S. and Ph.D. degrees from the University of Tsukuba, Tsukuba, Japan, in 1985 and 1988, respectively.

He joined the faculty of the Department of Radio Engineering, Southeast University, Nanjing, China, in 1988. He served as a Technical Manager at Communication Laboratory Japan, Kawasaki, from 1995 to 1997, and as a Visiting Researcher at ATR Adaptive Communications Research Laboratories, Kyoto, Japan, from 1997 to 1998. Currently, he is a Research Fellow with the Department of Electrical and Computer Engineering, Villanova University, Villanova, PA. His current research interests are in the areas of array signal processing, space-time adaptive processing, multiuser detection, blind signal processing, digital mobile communications, and time–frequency analysis.

Weifeng Mu received the B.S. degree from Tsinghua University, Beijing, China, in 1993 and the M.S. degree from Villanova University, Villanova, PA, in 2000, both in electrical engineering.

His research interests include statistical and array signal processing, smart antennas, adaptive filtering, digital signal processing, nonstationary signal, and time–frequency distributions. Currently, he is an Electrical Engineer with Metron Communication, Inc., Denville, NJ.

Moeness Amin (F'00) received the Ph.D. degree in electrical engineering in 1984 from University of Colorado, Boulder.

He has been with the Faculty of the Department of Electrical and Computer Engineering, Villanova University, Villanova, PA, since 1985, where he is now a Professor. His current research interests are in the areas of time–frequency analysis, spread spectrum communications, smart antennas, and blind signal processing.

Dr. Amin was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and a member of the Technical Committee of the IEEE Signal Processing Society on Statistical Signal and Array Processing from 1995 to 1997. He is currently a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications. He was the General Chair of the 1994 IEEE International Symposium on Time–Frequency and Time-Scale Analysis and is the General Chair of the 2000 IEEE Workshop on Statistical Signal and Array Processing. He received the 1997 IEEE Philadelphia Section Service Award and the IEEE Third Millennium Medal. He is also the recipient of the 1997 Villanova University Outstanding Faculty Research Award. He serves on the Committee of Science and Arts of the Franklin Institute.