PERGAMON Journal of the Franklin Institute 337 (2000) 483–497

Journal of The Franklin Institute

www.elsevier.nl/locate/jfranklin

Time–frequency maximum likelihood methods for direction finding

Yimin Zhang, Weifeng Mu, Moeness G. Amin*

Department of Electrical and Computer Engineering, Villanova University, 800 Lancaster Avenue, Villanova, PA 19085, USA

Abstract

This paper proposes a novel time–frequency maximum likelihood (t–f ML) method for direction-of-arrival (DOA) estimation for nonstationary signals impinging on a multi-sensor array receiver, and compares this method with conventional maximum likelihood DOA estimation techniques. Time–frequency distributions localize the signal power in the time–frequency domain, and as such enhance the effective SNR, leading to improved DOA estimation. The localization of signals with different time–frequency signatures permits the division of the time–frequency domain into smaller regions, each containing fewer signals than those incident on the array. The reduction of the number of signals within different time–frequency regions not only reduces the required number of sensors, but also decreases the computational load in multidimensional optimizations. Compared to the recently proposed time–frequency MUSIC (t–f MUSIC), the proposed t–f ML method can be applied to coherent environments, without the need to perform any type of preprocessing that is subject to both array geometry and array aperture.

Keywords: Time–frequency distribution; Direction finding; Maximum likelihood; Spatial time–frequency distribution; Array processing

1. Introduction

The localization of spatial sources by passive sensor array is one of the important problems in radar, sonar, radio-astronomy, and seismology. So far, numerous methods have been proposed for direction finding, most of which are based on the

* Corresponding author. Tel.: +1-610-519-7305; fax: +1-610-519-4436.

E-mail address: moeness@ece.vill.edu (M.G. Amin).

PII: S0016-0032(00)00031-4

estimates of the data covariance matrix. Among these methods, the maximum likelihood (ML) technique was one of the first to be devised and investigated [1]. It has a superior performance compared to other methods, particularly when the input signal-to-noise ratio (SNR) is low, the number of data samples is small, or the sources are highly correlated [2]. Therefore, despite its complexity, the ML technique remains of practical interest.

The evaluation of quadratic time–frequency distributions of the data snapshots across the array yields what is known as spatial time–frequency distributions (STFDs) [3,4]. STFD techniques are most appropriate to handle sources of nonstationary waveforms. STFDs spread the noise power while localizing the energy of the impinging signals in the time–frequency domain. This property leads to increasing the robustness of eigenstructure signal and noise subspace estimates with respect to the channel and receiver noise, and hence improves spatial resolution performance.

In this paper, we propose the time–frequency maximum likelihood (*t–f* ML) method for direction finding and provide the analysis that explains its performance. It is shown that the superior performance of the *t–f* ML method relative to other methods is attributed to the following three fundamental reasons: (1) Time–frequency distributions localize the signal power in the time–frequency domain, and as such enhance the effective SNR and improve the direction-of-arrival (DOA) estimation. (2) The localization of signals with different time–frequency signatures permits the division of the time–frequency domain into smaller regions, each containing fewer signals than those incident on the array. The reduction of the number of signals within different time–frequency regions relaxes the condition on the size of the array aperture as well as simplifies the multidimensional optimization estimation procedure. (3) Compared with the previously proposed time–frequency MUSIC (*t–f* MUSIC), the *t–f* ML method can be applied when the signal arrivals are highly correlated, whereas the *t–f* MUSIC algorithm cannot do so without some sort of preprocessing, such as spatial smoothing.

This paper is organized as follows. In Section 2, the signal model is established, and a brief review of the spatial time–frequency distributions is given. In Section 3, we discuss the SNR enhancement based on time–frequency distributions and its effect on the signal and noise subspace estimates using STFD matrices. The subspace estimates obtained from the STFD matrices are more robust to SNR and angular separation compared to those obtained from data covariance matrices. Section 4 presents the *t–f* ML and shows its performance in time-varying environments.

2. Background

2.1. Signal model

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

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

is commonly assumed, where the $m \times n$ spatial matrix $\mathbf{A}(\boldsymbol{\Theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_n)]$ represents the mixing matrix or the steering matrix, and $\mathbf{a}(\theta_i)$ are the steering vectors corresponding to angle of arrival θ_i. 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, independent of the source signals. That is,

$$E[\mathbf{n}(t+\tau)\mathbf{n}^{\mathrm{H}}(t)] = \sigma\delta(\tau)\mathbf{I} \quad \text{and} \quad E[\mathbf{n}(t+\tau)\mathbf{n}^{\mathrm{T}}(t)] = \mathbf{0} \quad \text{for any } \tau, \tag{2}$$

where $\delta(\tau)$ is the Kronecker 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 Eq. (1), it is assumed that the number of sensors is greater than the number of sources, i.e., $m > n$, and the number of snapshots is greater than the number of array sensors, i.e., $N > m$. We also assume that matrix $\mathbf{A}$ is full column rank, which implies that the steering vectors corresponding to n different angles of arrival are linearly independent.

Under the above assumptions, the correlation matrix is given by

$$\mathbf{R}_{\mathbf{xx}} = E[\mathbf{x}(t)\mathbf{x}^{\mathrm{H}}(t)] = \mathbf{A}(\boldsymbol{\Theta})\mathbf{R}_{\mathbf{dd}}\mathbf{A}^{\mathrm{H}}(\boldsymbol{\Theta}) + \sigma\mathbf{I}, \tag{3}$$

where $\mathbf{R}_{\mathbf{dd}} = E[\mathbf{d}(t)\mathbf{d}^{\mathrm{H}}(t)]$ is the signal correlation matrix. For notational convenience, we drop the argument $\boldsymbol{\Theta}$ and simply use $\mathbf{A}$ instead of $\mathbf{A}(\boldsymbol{\Theta})$. If $\hat{\boldsymbol{\Theta}}$ is an estimate of $\boldsymbol{\Theta}$, then we also use $\hat{\mathbf{A}}$ instead of $\mathbf{A}(\hat{\boldsymbol{\Theta}})$.

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

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

$$\hat{\mathbf{R}}_{\mathbf{xx}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}(i)\mathbf{x}^{\mathrm{H}}(i). \tag{4}$$

Let $\{\hat{\mathbf{s}}_1, \ldots, \hat{\mathbf{s}}_n, \hat{\mathbf{g}}_1, \ldots, \hat{\mathbf{g}}_{m-n}\}$ denote the unit-norm eigenvectors of $\hat{\mathbf{R}}_{\mathbf{xx}}$, arranged in the descending order of the associated eigenvalues, and let $\hat{\mathbf{S}}$ and $\hat{\mathbf{G}}$ denote the matrices made of 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 is given in [5].

In this paper, we focus on frequency-modulated (FM) signals, modeled as

$$\mathbf{d}(t) = [d_1(t), \ldots, d_n(t)]^{\mathrm{T}} = [D_1 \mathrm{e}^{\mathrm{j}\psi_1(t)}, \ldots, D_n \mathrm{e}^{\mathrm{j}\psi_n(t)}]^{\mathrm{T}}, \tag{5}$$

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

$$\text{(IF) } f_i(t) = \frac{1}{2\pi}\frac{\mathrm{d}\psi_i(t)}{\mathrm{d}t}.$$

FM signals are often encountered in applications such as radar and sonar. The consideration of FM signals in this paper is motivated by the fact that these signals are uniquely characterized by their IFs and, therefore, they have clear time–frequency signatures that are utilized by the STFD approach. Also, FM signals have constant amplitudes and, subsequently, yield time-independent covariance matrices. This property makes them amenable to conventional array processing based on second-order statistics.

2.2. Spatial time–frequency distributions

The STFDs based on Cohen's class of time–frequency distribution were introduced in [3] and its applications to direction finding has been discussed in [4]. However, the performance of direction finding based on STFD has not been made clear yet. In this paper, we focus on one key member of Cohen's class, namely the pseudo-Wigner–Ville distribution (PWVD) and its respective spatial distribution. Only the time–frequency points in the autoterm regions of PWVD are considered for STFD matrix construction. In these regions, it is assumed that the crossterms are negligible. This assumption serves to simplify the analysis and does not present any condition on performance. It is noted that the crossterms in STFD matrices play a similar role to the cross-correlation between source signals [6], and therefore they do not always impede the direction finding process.

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

$$D_{xx}(t,f) = \sum_{\tau=-(L-1)/2}^{(L-1)/2} x(t+\tau)x^{*}(t-\tau)\mathrm{e}^{-\mathrm{j}4\pi f\tau}, \tag{6}$$

where $*$ denotes complex conjugation. The spatial pseudo-Wigner–Ville distribution (SPWVD) matrix is obtained by replacing $x(t)$ by the data snapshot vector $\mathbf{x}(t)$,

$$\mathbf{D}_{\mathbf{xx}}(t,f) = \sum_{\tau=-(L-1)/2}^{(L-1)/2} \mathbf{x}(t+\tau)\mathbf{x}^{\mathrm{H}}(t-\tau)\mathrm{e}^{-\mathrm{j}4\pi f\tau}. \tag{7}$$

Substituting (1) into (7), we obtain

$$\mathbf{D}_{\mathbf{xx}}(t,f) = \mathbf{D}_{\mathbf{yy}}(t,f) + \mathbf{D}_{\mathbf{yn}}(t,f) + \mathbf{D}_{\mathbf{ny}}(t,f) + \mathbf{D}_{\mathbf{nn}}(t,f). \tag{8}$$

Under the assumption of uncorrelated signal and noise components and the zero-mean noise property, the expectation of the crossterm TFD matrices between the signal and noise vectors is zero, i.e., $E[\mathbf{D}_{\mathbf{yn}}(t,f)] = E[\mathbf{D}_{\mathbf{ny}}(t,f)] = \mathbf{0}$, and it follows that

$$\begin{aligned} E[\mathbf{D}_{\mathbf{xx}}(t,f)] &= \mathbf{D}_{\mathbf{yy}}(t,f) + E[\mathbf{D}_{\mathbf{nn}}(t,f)] \\ &= \mathbf{A}\mathbf{D}_{\mathbf{dd}}(t,f)\mathbf{A}^{\mathrm{H}} + E[\mathbf{D}_{\mathbf{nn}}(t,f)]. \end{aligned} \tag{9}$$

It is noted that relationship (9) holds true for every (t,f) point. Therefore, multiple time–frequency points can be used to reduce the effect of noise and ensure the full column rank property of the STFD matrix. In this paper, the STFD matrices over multiple time–frequency points are averaged, as is discussed in the next section.

3. Subspace analysis for STFD matrices

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

3.1. SNR enhancement

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

$$D_{d_i d_i}(t,f) = \sum_{\tau=-(L-1)/2}^{(L-1)/2} D_i^2 \mathrm{e}^{\mathrm{j}[\psi_i(t+\tau)-\psi_i(t-\tau)]-\mathrm{j}4\pi f\tau}. \tag{10}$$

Assume that the third-order derivative of the phase is negligible over the window length L, then along the true time–frequency points of ith signal, $f_i = (1/2\pi)\mathrm{d}\psi_i(t)/\mathrm{d}t$, and $\psi_i(t+\tau) - \psi_i(t-\tau) - 4\pi f_i \tau \simeq 0$. Accordingly,

$$D_{d_i d_i}(t,f_i) = \sum_{\tau=-(L-1)/2}^{(L-1)/2} D_i^2 = LD_i^2. \tag{11}$$

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

$$\mathbf{D}_{\mathbf{nn}}(t,f) = \sum_{\tau=-(L-1)/2}^{(L-1)/2} \mathbf{n}(t+\tau)\mathbf{n}^{\mathrm{H}}(t-\tau)\mathrm{e}^{-\mathrm{j}4\pi f\tau}. \tag{12}$$

Under the spatial white and temporal white assumptions, the statistical expectation of $\mathbf{D}_{\mathbf{nn}}(t,f)$ is given by

$$E[\mathbf{D}_{\mathbf{nn}}(t,f)] = \sum_{\tau=-(L-1)/2}^{(L-1)/2} E[\mathbf{n}(t+\tau)\mathbf{n}^{\mathrm{H}}(t-\tau)]\mathrm{e}^{-\mathrm{j}4\pi f\tau} = \sigma\mathbf{I}. \tag{13}$$

Therefore, when we select the time–frequency points along the time–frequency signature or the IF of the ith FM signal, the SNR in model (9) is LD_i^2/σ, which has an improved factor L over the one associated with model (3).

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 $N-L+1$ time–frequency points of constant distribution value for each source signal. Therefore, the averaged STFD over the time–frequency signatures of n_0 signals, i.e., a total of $n_0(N-L+1)$ time–frequency points, is given by

$$\hat{\mathbf{D}} = \frac{1}{n_0(N-L+1)} \sum_{q=1}^{n_0} \sum_{i=1}^{N-L+1} \mathbf{D}_{\mathbf{xx}}(t_i, f_{q,i}), \tag{14}$$

where $f_{q,i}$ is the instantaneous frequency of the qth signal at the ith time sample. The expectation of the averaged STFD matrix is

$$\begin{aligned} \mathbf{D} = E[\hat{\mathbf{D}}] &= \frac{1}{n_0(N-L+1)} \sum_{q=1}^{n_0} \sum_{i=1}^{N-L+1} E[\mathbf{D}_{\mathbf{xx}}(t_i, f_{q,i})] \\ &= \frac{1}{n_0} \sum_{q=1}^{n_0} [LD_q^2 \mathbf{a}(\theta_q)\mathbf{a}^{\mathrm{H}}(\theta_q) + \sigma \mathbf{I}] = \frac{L}{n_0} \mathbf{A}^0 \mathbf{R}_{\mathbf{dd}}^0 (\mathbf{A}^0)^{\mathrm{H}} + \sigma \mathbf{I}, \end{aligned} \tag{15}$$

where $\mathbf{R}_{\mathbf{dd}}^0$ and $\mathbf{A}^0$, respectively, represent the signal correlation matrix and the mixing matrix constructed by only considering n_0 signals out of the total number of signal arrivals n.

3.2. Signal and noise subspaces based on STFDs

The statistical properties of the eigenstructures using the STFD matrix are provided in this subsection.

Lemma 1. *Let* $\lambda_1^0 > \lambda_2^0 > \cdots > \lambda_{n_0}^0 > \lambda_{n_0+1}^0 = \lambda_{n_0+2}^0 = \cdots = \lambda_m^0 = \sigma$ *denote the eigenvalues of* $\mathbf{R}_{\mathbf{xx}}^0$, *which is defined from a data record of a mixture of the* n_0 *selected FM signals. Denote the unit-norm eigenvectors associated with* $\lambda_1^0, \ldots, \lambda_{n_0}^0$ *by the columns of* $\mathbf{S}^0 = [\mathbf{s}_1^0, \ldots, \mathbf{s}_{n_0}^0]$, *and those corresponding to* $\lambda_{n_0+1}^0, \ldots, \lambda_m^0$ *by the columns of* $\mathbf{G}^0 = [\mathbf{g}_1^0, \ldots, \mathbf{g}_{m-n_0}^0]$. *We also denote* $\lambda_1^{tf} > \lambda_2^{tf} > \cdots > \lambda_{n_0}^{tf} > \lambda_{n_0+1}^{tf} = \lambda_{n_0+2}^{tf} = \cdots = \lambda_m^{tf} = \sigma^{tf}$ *as the eigenvalues of* $\mathbf{D}$ *defined in* (15). *The unit-norm eigenvectors associated with* $\lambda_1^{tf}, \ldots, \lambda_{n_0}^{tf}$ *are represented by the columns of* $\mathbf{S}^{tf} = [\mathbf{s}_1^{tf}, \ldots, \mathbf{s}_{n_0}^{tf}]$, *and those corresponding to* $\lambda_{n_0+1}^{tf}, \ldots, \lambda_m^{tf}$ *are represented by the columns of* $\mathbf{G}^{tf} = [\mathbf{g}_1^{tf}, \ldots, \mathbf{g}_{m-n_0}^{tf}]$. *Accordingly,*

(a) *The signal and noise subspaces of* $\mathbf{S}^{tf}$ *and* $\mathbf{G}^{tf}$ *are the same as* $\mathbf{S}^0$ *and* $\mathbf{G}^0$, *respectively.*
(b) *The eigenvalues have the following relationship*:

$$\lambda_i^{tf} = \begin{cases} \dfrac{L}{n_0}(\lambda_i^0 - \sigma) + \sigma = \dfrac{L}{n_0}\lambda_i^0 + \left(1 - \dfrac{L}{n_0}\right)\sigma & i \leqslant n_0, \\ \sigma^{tf} = \sigma & n_0 < i \leqslant m. \end{cases} \tag{16}$$

Both parts of the above equations are direct results of (15). From Lemma 1 it is clear that the largest n_0 eigenvalues are amplified using STFD analysis.

Lemma 2. *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* $\hat{\mathbf{D}} - \mathbf{D}$ *is a zero-mean, random matrix whose elements are asymptotically jointly Gaussian. The proof is given in Appendix A.*

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 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*

$$E(\mathbf{S}^{tf}(\mathbf{S}^{tf})^{\mathrm{H}}\hat{\mathbf{g}}_i^{tf})(\mathbf{S}^{tf}(\mathbf{S}^{tf})^{\mathrm{H}}\hat{\mathbf{g}}_j^{tf})^{\mathrm{H}} = \frac{1}{(N-L+1)}\mathbf{U}^{tf}\delta_{i,j}, \tag{17}$$

$$E(\mathbf{S}^{tf}(\mathbf{S}^{tf})^{\mathrm{H}}\hat{\mathbf{g}}_i^{tf})(\mathbf{S}^{tf}(\mathbf{S}^{tf})^{\mathrm{H}}\hat{\mathbf{g}}_j^{tf})^{\mathrm{T}} = \mathbf{0} \quad \textit{for all } i, j, \tag{18}$$

where

$$\begin{aligned}\mathbf{U}^{tf} &= \frac{\sigma L}{n_0}\left[\sum_{k=1}^{n_0} \frac{\lambda_k^{tf}}{(\sigma-\lambda_k^{tf})^2}\mathbf{s}_k^{tf}(\mathbf{s}_k^{tf})^{\mathrm{H}}\right] \\ &= \sigma\left[\sum_{k=1}^{n_0} \frac{(\lambda_k^0-\sigma)+(n_0/L)\sigma}{(\sigma-\lambda_k^0)^2}\mathbf{s}_k^0(\mathbf{s}_k^0)^{\mathrm{H}}\right].\end{aligned} \tag{19}$$

The proof is given in [7]. For comparison, we quote the results from reference [5], which were provided using the data covariance matrix

$$E(\mathbf{S}\mathbf{S}^{\mathrm{H}}\hat{\mathbf{g}}_i)(\mathbf{S}\mathbf{S}^{\mathrm{H}}\hat{\mathbf{g}}_j)^{\mathrm{H}} = \frac{\sigma}{N}\left[\sum_{k=1}^{n} \frac{\lambda_k}{(\sigma-\lambda_k)^2}\mathbf{s}_k\mathbf{s}_k^{\mathrm{H}}\right]\delta_{i,j}, \tag{20}$$

$$E(\mathbf{S}\mathbf{S}^{\mathrm{H}}\hat{\mathbf{g}}_i)(\mathbf{S}\mathbf{S}^{\mathrm{H}}\hat{\mathbf{g}}_j)^{\mathrm{T}} = \mathbf{0} \quad \text{for all } i, j, \tag{21}$$

where $\mathbf{S}, \mathbf{s}_k, \hat{\mathbf{g}}_k, \lambda_k$ are analogous to $\mathbf{S}^0, \mathbf{s}_k^0, \hat{\mathbf{g}}_k^0, \lambda_k^0$, respectively, except that they are defined for all n signals instead of only n_0 signals.

Comparing (17) and (19) with (20), two important observations are in order. First, if the signals are both localizable and separable in the time–frequency domain, then the reduction of the number of signals from n to n_0 reduces the estimation error, specifically when the signals are closely spaced. 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 (17) and (19) that, when $\lambda_k^0 \gg \sigma$ for all $k = 1, 2, \ldots, n_0$, 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 eigenvalues are close to σ ($\lambda_k^0 \simeq \sigma$, for some $k = 1, 2, \ldots, n_0$), which is the case of weak or closely spaced signals, the result of (17) is reduced by a factor of up to $G = L/n_0$. This factor represents the gain achieved using STFD processing.

4. The time–frequency maximum likelihood methods

In this section, we analyze the performance of the maximum likelihood methods based on time–frequency distributions (t–f ML). For conventional ML methods, the joint density function of the sampled data vectors $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N)$, is given by [2]

$$f(\mathbf{x}(1), \ldots, \mathbf{x}(N)) = \prod_{i=1}^{N} \frac{1}{\pi^m \det[\sigma \mathbf{I}]} \exp\left(-\frac{1}{\sigma}[\mathbf{x}(i) - \mathbf{A}\mathbf{d}(i)]^{\mathrm{H}}[\mathbf{x}(i) - \mathbf{A}\mathbf{d}(i)] \right), \tag{22}$$

where $\det[\cdot]$ denotes the determinant. It follows from (22) that the log-likelihood function of the observations $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N)$, is given by

$$L = -mN \ln \sigma - \frac{1}{\sigma} \sum_{i=1}^{N} [\mathbf{x}(i) - \mathbf{A}\mathbf{d}(i)]^{\mathrm{H}}[\mathbf{x}(i) - \mathbf{A}\mathbf{d}(i)]. \tag{23}$$

To carry out this minimization, we fix $\mathbf{A}$ and minimize (23) with respect to $\mathbf{d}$. This yields the well-known solution

$$\hat{\mathbf{d}}(i) = [\mathbf{A}^{\mathrm{H}}\mathbf{A}]^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{x}(i). \tag{24}$$

We can obtain the concentrated likelihood function as [2,8]

$$F_{\mathrm{ML}}(\mathbf{\Theta}) = tr\{[\mathbf{I} - \hat{\mathbf{A}}(\hat{\mathbf{A}}^{\mathrm{H}}\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^{\mathrm{H}}]\hat{\mathbf{R}}_{xx}\}, \tag{25}$$

where $tr(\mathbf{A})$ denotes the trace of $\mathbf{A}$. The ML estimate of $\mathbf{\Theta}$ is obtained as the minimizer of (25). Let ω_i and $\hat{\omega}_i$, respectively, denote the spatial frequency and its ML estimate associated with θ_i, then the estimation error $(\hat{\omega}_i - \omega_i)$ are asymptotically (for large N) jointly Gaussian distributed with zero means and the covariance [9]

$$E[(\hat{\omega}_i - \omega_i)^2] = \frac{1}{2N}\{[\mathrm{Re}(\mathbf{H} \odot \mathbf{R}_{\mathbf{dd}}^{\mathrm{T}})]^{-1} \times \mathrm{Re}[\mathbf{H} \odot (\mathbf{R}_{\mathbf{dd}}\mathbf{A}^{\mathrm{H}}\mathbf{U}\mathbf{A}\mathbf{R}_{\mathbf{dd}})^{\mathrm{T}}][\mathrm{Re}(\mathbf{H} \odot \mathbf{R}_{\mathbf{dd}}^{\mathrm{T}})]^{-1}\}_{ii}, \tag{26}$$

where $\odot$ denotes Hadamard product. Moreover,

$$\mathbf{U} = \sum_{k=1}^{n} \frac{\lambda_k \sigma}{(\sigma - \lambda_k)^2} \mathbf{s}_k \mathbf{s}_k^{\mathrm{H}},$$

$$\mathbf{H} = \mathbf{C}^{\mathrm{H}}[\mathbf{I} - \mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}]\mathbf{C},$$

$$\mathbf{C} = \frac{\mathrm{d}\mathbf{A}}{\mathrm{d}\omega}. \tag{27}$$

Next we consider the t–f ML method. As we discussed in the previous section, we select $n_0 \leqslant n$ signals in the time–frequency domain. The concentrated likelihood function defined from the STFD matrix is similar to (25) and is obtained by replacing $\hat{\mathbf{R}}_{\mathbf{xx}}$ by $\hat{\mathbf{D}}$ (Appendix B),

$$F_{\mathrm{ML}}^{tf}(\mathbf{\Theta}) = tr[\mathbf{I} - \hat{\mathbf{A}}^0((\hat{\mathbf{A}}^0)^{\mathrm{H}}\hat{\mathbf{A}}^0)^{-1}(\hat{\mathbf{A}}^0)^{\mathrm{H}}]\hat{\mathbf{D}}. \tag{28}$$

Therefore, the estimation error $(\hat{\omega}_i^{tf} - \omega_i)$ associated with the t–f ML method are asymptotically (for $N \gg L$) jointly Gaussian distributed with zero means and the covariance

$$\begin{aligned}
&E[(\hat{\omega}_i^{tf} - \omega_i)^2] \\
&= \frac{\sigma}{2(N-L+1)}\{[\mathrm{Re}(\mathbf{H}^0 \odot \mathbf{D}_{\mathbf{dd}}^{\mathrm{T}})]^{-1} \\
&\quad \times \mathrm{Re}[\mathbf{H}^0 \odot (\mathbf{D}_{\mathbf{dd}}(\mathbf{A}^0)^{\mathrm{H}}\mathbf{U}^{tf}\mathbf{A}^0\mathbf{D}_{\mathbf{dd}})^{\mathrm{T}}][\mathrm{Re}(\mathbf{H}^0 \odot \mathbf{D}_{\mathbf{dd}}^{\mathrm{T}})]^{-1}\}_{ii} \\
&= \frac{\sigma}{2(N-L+1)}\{[\mathrm{Re}(\mathbf{H}^0 \odot (\mathbf{R}_{\mathbf{dd}}^0)^{\mathrm{T}})]^{-1} \\
&\quad \times \mathrm{Re}[\mathbf{H}^0 \odot (\mathbf{R}_{\mathbf{dd}}^0(\mathbf{A}^0)^{\mathrm{H}}\mathbf{U}^{tf}\mathbf{A}^0\mathbf{R}_{\mathbf{dd}}^0)^{\mathrm{T}}][\mathrm{Re}((\mathbf{H}^0 \odot \mathbf{R}_{\mathbf{dd}}^0)^{\mathrm{T}})]^{-1}\}_{ii},
\end{aligned} \tag{29}$$

where $\mathbf{U}^{tf}$ is defined in (19), and

$$\mathbf{H}^0 = (\mathbf{C}^0)^{\mathrm{H}}[\mathbf{I} - \mathbf{A}^0((\mathbf{A}^0)^{\mathrm{H}}\mathbf{A}^0)^{-1}(\mathbf{A}^0)^{\mathrm{H}}]\mathbf{C}^0,$$

$$\mathbf{C}^0 = \frac{\mathrm{d}\mathbf{A}^0}{\mathrm{d}\omega}. \tag{30}$$

In the case of $n_0 = n$, then $\mathbf{H}^0 = \mathbf{H}$, and $\mathbf{C}^0 = \mathbf{C}$.

The signal localization in the time–frequency domain enables us to select fewer signal arrivals. This fact is not only important in improving the estimation performance, particularly when the signals are closely spaced, but also reduces the dimension of optimization problem solved by the maximum likelihood algorithm, and subsequently reduces the computational requirement.

To demonstrate the advantages of t–f ML over the conventional ML and the time–frequency MUSIC (t–f MUSIC), consider a uniform linear array of eight sensors separated by half a wavelength. Two FM signals arrive from $(\theta_1, \theta_2) = (-10°, 10°)$ with the instantaneous frequencies $f_1(t) = 0.2 + 0.1t/N + 0.2 \times \sin(2\pi t/N)$ and $f_2(t) = 0.2 + 0.1t/N + 0.2\sin(2\pi t/N + \pi/2)$, $t = 1, \ldots, N$. The SNR of both signals is -20 dB, and the number of snapshots used in the simulation is $N = 1024$. We used $L = 129$ for t–f ML. Fig. 1 shows the PWVD of the mixed noise-free signals at the reference sensor. Fig. 2 shows (θ_1, θ_2) that yield the minimum values of the likelihood function of the t–f ML and the ML methods for 20 independent trials. It is evident that the t–f ML provides much improved DOA estimation over the conventional ML.

In the next example, we compare the t–f ML and the t–f MUSIC for coherent sources. The two coherent FM signals have common instantaneous frequencies $f_{1,2}(t) = 0.2 + 0.1t/N + 0.2\sin(2\pi t/N), t = 1, \ldots, N$, with $\pi/2$ phase difference. The signals arrive at $(\theta_1, \theta_2) = (-2°, 2°)$. The SNR of both signals is 5 dB and the number of snapshots is 1024. Again, we used $L = 129$ for both t–f ML and t–f MUSIC. Fig. 3 shows the PWVD of the mixed noise-free signals, and Fig. 4 shows the contour plots of the likelihood function of the t–f ML and the estimated spectra of t–f MUSIC for five independent trials. It is clear that the t–f ML can separate the two signals whereas the t–f MUSIC cannot.

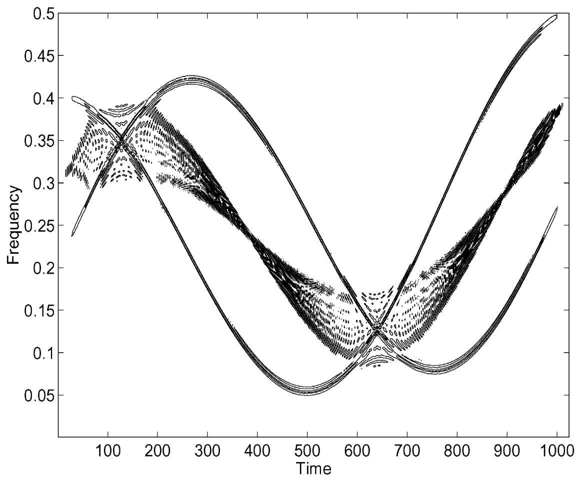
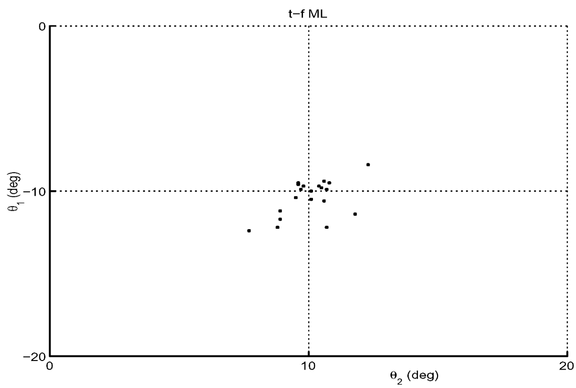
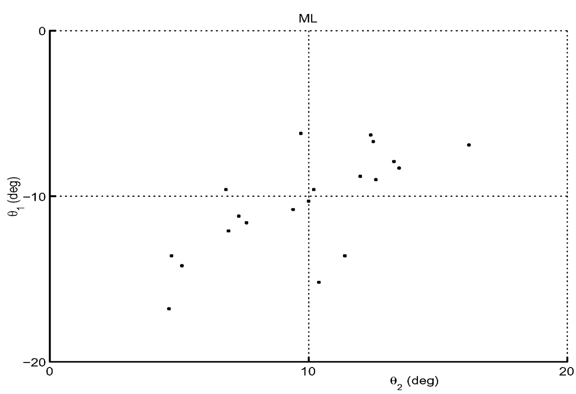


Fig. 1. Pseudo-Wigner–Ville distribution of the mixture of the two FM signals.



(a) t-f ML



(b) ML

Fig. 2. (θ_1, θ_2) which minimize the *t–f* ML and ML likelihood functions.

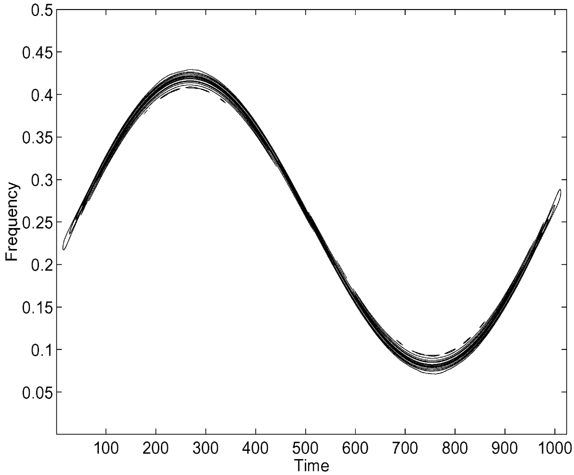
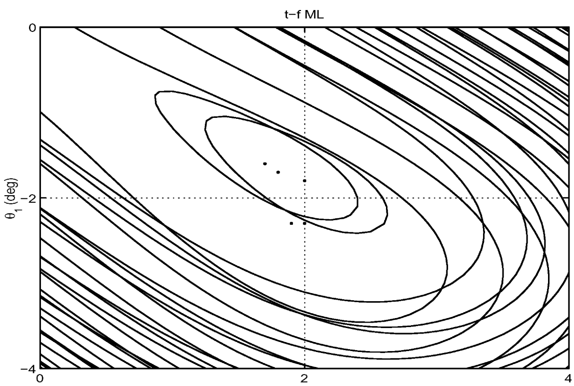
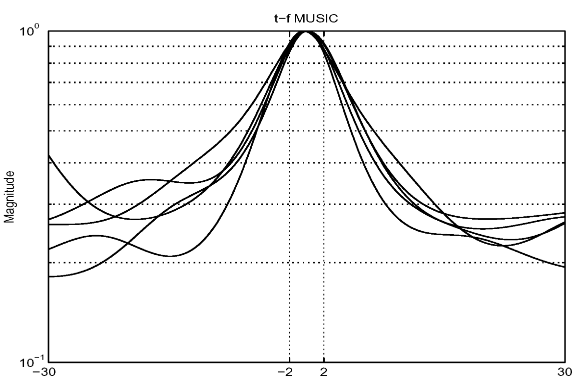


Fig. 3. Pseudo-Wigner–Ville distribution of the mixture of the two coherent FM signals.



(a) t-f ML



(b) t-f MUSIC

Fig. 4. Contour plots of *t–f* ML likelihood function and spatial spectra of *t–f* MUSIC.

5. Conclusions

The time–frequency maximum likelihood (t–f ML) method has been proposed for direction finding, which is based on the spatial time–frequency distribution (STFD) matrices. By taking frequency-modulated (FM) signals as example, we show that the STFD matrices provide more robust eigen-decomposition than covariance matrices. The analysis and simulation results showed that the t–f ML improves over the conventional maximum likelihood technique for low SNR, and outperforms the t–f MUSIC in coherent signal environments.

Acknowledgements

This work was supported by the Office of Naval Research under Grant N00014-98-1-0176.

Appendix A

Proof of Lemma 2

From (1), (14), and (15),

$$
\begin{aligned}
\hat{\mathbf{D}} - \mathbf{D}
&= \frac{1}{n_0(N-L+1)} \sum_{q=1}^{n_0} \sum_{i=1}^{N-L+1} \sum_{\tau=-(L-1)/2}^{(L-1)/2} \mathbf{y}(t_i+\tau)\mathbf{n}^{\mathrm{H}}(t_i-\tau)\mathrm{e}^{-\mathrm{j}4\pi f_{q,i}\tau} \\
&+ \frac{1}{n_0(N-L+1)} \sum_{q=1}^{n_0} \sum_{i=1}^{N-L+1} \sum_{\tau=-(L-1)/2}^{(L-1)/2} \mathbf{n}(t_i+\tau)\mathbf{y}^{\mathrm{H}}(t_i-\tau)\mathrm{e}^{-\mathrm{j}4\pi f_{q,i}\tau} \\
&+ \frac{1}{n_0(N-L+1)} \sum_{q=1}^{n_0} \sum_{i=1}^{N-L+1} \sum_{\tau=-(L-1)/2}^{(L-1)/2} \mathbf{n}(t_i+\tau)\mathbf{n}^{\mathrm{H}}(t_i-\tau)\mathrm{e}^{-\mathrm{j}4\pi f_{q,i}\tau} - \sigma\mathbf{I}.
\end{aligned}
\tag{A.1}
$$

We first consider the first term in (A.1). Denoting $t_i' = t_i - \tau$, and noting the fact that, when the third-order derivative of the phase is negligible over $[t-L+1, t+L-1]$ for any signal and any t, $d_q(t_i'+2\tau)\mathrm{e}^{-\mathrm{j}4\pi f_{q,i}\tau} \simeq d_q(t_i')$ at the time–frequency point $(t_i, f_{q,i})$, then

$$
\begin{aligned}
&\sum_{q=1}^{n_0} \sum_{i=1}^{N-L+1} \sum_{\tau=-(L-1)/2}^{(L-1)/2} \mathbf{y}(t_i+\tau)\mathbf{n}^{\mathrm{H}}(t_i-\tau)\mathrm{e}^{-\mathrm{j}4\pi f_{q,i}\tau} \\
&= \sum_{q=1}^{n_0} \sum_{t_i'=1}^{N-L+1} \sum_{\tau=-(L-1)/2}^{(L-1)/2} \mathbf{y}(t_i'+2\tau)\mathbf{n}^{\mathrm{H}}(t_i')\mathrm{e}^{-\mathrm{j}4\pi f_{q,i}\tau} \\
&\simeq \sum_{q=1}^{n_0} \sum_{t_i'=1}^{N-L+1} L d_q(t_i')\mathbf{a}(\theta_q)\mathbf{n}^{\mathrm{H}}(t_i') = \sum_{t_i'=1}^{N-L+1} L\mathbf{y}(t_i')\mathbf{n}^{\mathrm{H}}(t_i').
\end{aligned}
\tag{A.2}
$$

Therefore, the elements of the first term in Eq. (A.1) are clearly asymptotically jointly Gaussian from the multivariate Central Limit Theorem [10]. Similar result can be obtained for the second term of (A.1). The elements of the third term in (A.1) are also jointly Gaussian from the fact that the covariance between the (p, r)th element of $\mathbf{n}(t+\tau)\mathbf{n}^{\mathrm{H}}(t-\tau)$ at time t_i and t_k is given by

$$
\begin{aligned}
&E\Bigg\{\Bigg[\sum_{\tau_1=-(L-1)/2}^{(L-1)/2} n_p(t_i+\tau_1)n_r^*(t_i-\tau_1) \\
&\quad - E\Bigg(\sum_{\tau_1=-(L-1)/2}^{(L-1)/2} n_p(t_i+\tau_1)n_r^*(t_i-\tau_1)\Bigg)\Bigg]\mathrm{e}^{-\mathrm{j}4\pi f_{q,i}\tau_1} \\
&\quad \times\Bigg[\sum_{\tau_2=-(L-1)/2}^{(L-1)/2} n_p^*(t_k+\tau_2)n_r(t_k-\tau_2) \\
&\quad - E\Bigg(\sum_{\tau_2=-(L-1)/2}^{(L-1)/2} n_p^*(t_k+\tau_2)n_r(t_k-\tau_2)\Bigg)\Bigg]\mathrm{e}^{-\mathrm{j}4\pi f_{q,k}\tau_2}\Bigg\} \\
&= \sum_{\tau_1=-(L-1)/2}^{(L-1)/2}\ \sum_{\tau_2=-(L-1)/2}^{(L-1)/2} E[n_p(t_i+\tau_1)n_r^*(t_i-\tau_1)] \\
&\quad \times E[n_p^*(t_k+\tau_2)n_r(t_k-\tau_2)]\mathrm{e}^{-\mathrm{j}4\pi(f_{q,i}\tau_1-f_{q,k}\tau_2)} \\
&\quad + \sum_{\tau_1=-(L-1)/2}^{(L-1)/2}\ \sum_{\tau_2=-(L-1)/2}^{(L-1)/2} E[n_p(t_i+\tau_1)n_p^*(t_k+\tau_2)] \\
&\quad \times E[n_r^*(t_i-\tau_1)n_r(t_k-\tau_2)]\mathrm{e}^{-\mathrm{j}4\pi(f_{q,i}\tau_1-f_{q,k}\tau_2)} \\
&\quad + \sum_{\tau_1=-(L-1)/2}^{(L-1)/2}\ \sum_{\tau_2=-(L-1)/2}^{(L-1)/2} E[n_p(t_i+\tau_1)n_r(t_k-\tau_2)] \\
&\quad \times E[n_p^*(t_k+\tau_2)n_r^*(t_i-\tau_1)]\mathrm{e}^{-\mathrm{j}4\pi(f_{q,i}\tau_1-f_{q,k}\tau_2)} \\
&\quad - \sum_{\tau_1=-(L-1)/2}^{(L-1)/2}\ \sum_{\tau_2=-(L-1)/2}^{(L-1)/2} \sigma^2\delta_{p,r}\mathrm{e}^{-\mathrm{j}4\pi(f_{q,i}\tau_1-f_{q,k}\tau_2)} \\
&= L\sigma^2\delta_{i,k}.
\end{aligned}
\tag{A.3}
$$

Since the linear combination of joint-Gaussian processes is jointly Gaussian, then $\hat{\mathbf{D}}-\mathbf{D}$ is a random matrix whose elements are asymptotically jointly Gaussian. Also $\hat{\mathbf{D}}-\mathbf{D}\to 0$ as $N\to\infty$.

Appendix B

Derivation of (28)

The number of data samples available for the construction of the STFD matrix is $N-L+1$, where the selected n_0 signals are included. Denote $\hat{\mathbf{u}}_k$ as the kth column of $\hat{\mathbf{D}}$, and $\mathbf{u}_k$ the kth column of $\mathbf{D}$. From Lemma 2, we know that $\hat{\mathbf{u}}_k$ is asymptotically jointly Gaussian, and its density function is

$$f_{tf}(\hat{\mathbf{u}}_k) = \frac{1}{\pi}\det\left[\frac{1}{N-L+1}\mathbf{\Delta}_k\right]^{-1/2} \times \exp\left[-\frac{1}{2}(\hat{\mathbf{u}}_k - \mathbf{u}_k)^{\mathrm{H}}\left(\frac{1}{N-L+1}\mathbf{\Delta}_k\right)^{-1}(\hat{\mathbf{u}}_k - \mathbf{u}_k)\right], \tag{B.1}$$

where $\mathbf{\Delta}_k$ stands for the asymptotic covariance matrix of $\mathbf{u}_k$,

$$\mathbf{\Delta}_k \triangleq \lim_{N\to\infty} (N-L+1)E[(\hat{\mathbf{u}}_k - \mathbf{u}_k)(\hat{\mathbf{u}}_k - \mathbf{u}_k)^{\mathrm{H}}]. \tag{B.2}$$

From the results of Lemma 2, it is clear that $\mathbf{\Delta}_k$ is a diagonal matrix with equal diagonal elements. Denoting $\mathbf{\Delta}_k = \beta\mathbf{I}$, the log-likelihood function is given by

$$L_{tf} = -\frac{1}{2m}\frac{1}{N-L+1}\log\beta - \frac{1}{2\beta}(\hat{\mathbf{u}}_k - \mathbf{u}_k)^{\mathrm{H}}(\hat{\mathbf{u}}_k - \mathbf{u}_k). \tag{B.3}$$

Maximizing L_{tf} is equivalent to minimizing

$$h_k \triangleq [\hat{\mathbf{u}}_k - \mathbf{u}_k]^{\mathrm{H}}[\hat{\mathbf{u}}_k - \mathbf{u}_k]. \tag{B.4}$$

For different k, we may construct the following cost function.

$$\begin{aligned} h &\triangleq \sum_{k=1}^{m} h_k \\ &= \sum_{k=1}^{m} [\hat{\mathbf{u}}_k - \mathbf{u}_k]^{\mathrm{H}}[\hat{\mathbf{u}}_k - \mathbf{u}_k] \\ &= tr\{[\hat{\mathbf{D}} - \mathbf{D}]^{\mathrm{H}}[\hat{\mathbf{D}} - \mathbf{D}]\}. \end{aligned} \tag{B.5}$$

Similar to (24), and by taking into account that we used n_0 signals instead of n signals, the estimation of $\mathbf{D}$ is obtained as $\hat{\mathbf{A}}^0((\hat{\mathbf{A}}^0)^{\mathrm{H}}\hat{\mathbf{A}}^0)^{-1}(\hat{\mathbf{A}}^0)^{\mathrm{H}}\hat{\mathbf{D}}\hat{\mathbf{A}}^0 \times ((\hat{\mathbf{A}}^0)^{\mathrm{H}}\hat{\mathbf{A}}^0)^{-1}(\hat{\mathbf{A}}^0)^{\mathrm{H}}$, and the minimization of Eq. (B.5) leads to (28).

References

[1] F.C. Schweppe, Sensor array data processing for multiple signal estimation, IEEE Trans. Inform. Theory IT-14 (1968) 294–305.

[2] I. Ziskind, M. Wax, Maximum likelihood localization of multiple sources by alternating projection, IEEE Trans. Acoust. Speech Signal Process. ASSP-36 (10) (1988) 1553–1560.

[3] A. Belouchrani, M. Amin, Blind source separation based on time–frequency signal representation, IEEE Trans. Signal Process. 46 (11) (1998) 2888–2898.

[4] A. Belouchrani, M. Amin, Time–frequency MUSIC, IEEE Signal Process. Lett. 6 (5) (1999) 109–110.

[5] P. Stoica, A. Nehorai, MUSIC, maximum likelihood, and Cramer–Rao bound, IEEE Trans. Acoust. Speech Signal Process. 37 (5) (1989) 720–741.

[6] Y. Zhang, M. Amin, Spatial averaging of time–frequency distributions, Proceedings of ICASSP, Vol. III, Pheonix, AZ, March 1999, pp. 1337–1340.

[7] Y. Zhang, W. Mu, M.G. Amin, Maximum likelihood methods for array processing based on time–frequency analysis, Proceedings of SPIE: Advanced Signal Processing Algorithms, Architectures, and Implementations IX, Vol. 3807, Denver, CO, July 1999.
[8] P. Stoica, K.C. Sharman, Maximum likelihood methods for direction-of-arrival estimation, IEEE Trans. Acoust. Speech Signal Process. ASSP-38 (7) (1990) 1132–1142.
[9] P. Stoica, A. Nehorai, MUSIC, maximum likelihood, and Cramer–Rao bound: further results and comparisons, IEEE Trans. Acoust. Speech Signal Process. 38 (12) (1990) 2140–2150.
[10] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, New York, 1971.

Dr. Yimin Zhang received his M.S. and Ph.D. degrees from the University of Tsukuba, Japan, in 1985 and 1988, respectively. He joined the faculty of the Department of Radio Engineering, Southeast University, China, in 1988. He served as a Technical Manager at Communication Laboratory Japan from 1995 to 1997, and a Visiting Researcher at ATR Adaptive Communications Research Laboratories, Japan, from 1997 to 1998. Currently, he is a Research Associate at the Department of Electrical and Computer Engineering, Villanova University. His current research interests are in the areas of array signal processing, space-time adaptive processing, blind signal processing, digital mobile communications, and time–frequency analysis.

Mr. Weifeng Mu received the B.S. degree from Tsinghua University, Beijing, China in 1993 and the M.S. degree from Villanova University, Villanova in 2000, both in Electrical Engineering. His current research interests include statistical and array signal processing, smart antenna, adaptive filtering, digital signal processing, nonstationary signal and time–frequency distributions.

Dr. Moeness Amin received his Ph.D. degree in Electrical Engineering in 1984 from University of Colorado, Boulder. He has been on the Faculty of the Department of Electrical and Computer Engineering at Villanova University since 1985, where he is now a Professor. From 1995 to 1997, 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. 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. Dr. Amin is the General Chair of the 2000-IEEE Workshop on Statistical Signal and Array Processing. Dr. Amin is the recipient of 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. His current research interests are in the areas of time–frequency analysis, spread spectrum communications, smart antennas, and blind signal processing.