# DOA Estimation of Quasi-Stationary Signals With Less Sensors Than Sources and Unknown Spatial Noise Covariance: A Khatri–Rao Subspace Approach

Wing-Kin Ma, Member, IEEE, Tsung-Han Hsieh, and Chong-Yung Chi, Senior Member, IEEE

Abstract—In real-world applications such as those for speech and audio, there are signals that are nonstationary but can be modeled as being stationary within local time frames. Such signals are generally called quasi-stationary or locally stationary signals. This paper considers the problem of direction-of-arrival (DOA) estimation of quasi-stationary signals. Specifically, in our problem formulation we assume: i) sensor array of uniform linear structure; ii) mutually uncorrelated wide-sense quasi-stationary source signals; and iii) wide-sense stationary noise process with unknown, possibly nonwhite, spatial covariance. Under the assumptions above and by judiciously examining the structures of local second-order statistics (SOSs), we develop a Khatri-Rao (KR) subspace approach that has two notable advantages. First, through an identifiability analysis, it is proven that this KR subspace approach can operate even when the number of sensors is about half of the number of sources. The idea behind is to make use of a "virtual" array structure provided inherently in the local SOS model, of which the degree of freedom is about twice of that of the physical array. Second, the KR formulation naturally provides a simple yet effective way of eliminating the unknown spatial noise covariance from the signal SOSs. Extensive simulation results are provided to demonstrate the effectiveness of the KR subspace approach under various situations.

*Index Terms*—Khatri–Rao subspace, Kruskal-rank, quasi-stationary signals (QSS), second-order statistics, underdetermined direction-of-arrival (DOA) estimation, unknown noise covariance.

## I. INTRODUCTION

**D** IRECTION-OF-ARRIVAL (DOA) estimation using an array of sensors, or direction finding (DF) is an important topic since there are many real-world problems where ac-

Manuscript received September 24, 2008; accepted August 27, 2009. First published October 20, 2009; current version published March 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Olivier Besson. This work was supported in part by General Research Funds awarded by the Research Grant Council of Hong Kong (Project Nos. CUHK415509, CUHK415908) and by the National Science Council, R.O.C., by Grant NSC 97-2221-E-007-073-MY3. Part of this work was presented at the Thirty-Fourth IEEE International Conference on Acoustics, Speech, and Signal Processing, Taipei, Taiwan, April 2009.

W.-K. Ma is with the Department of Electrical Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong (e-mail: wkma@ieee.org). T.-H. Hsieh is with Realtek Semiconductor Corporation, Hsinchu, Taiwan

(e-mail: tsunghan.hsieh1984@gmail.com).

C.-Y. Chi is with the Institute of Communications Engineering and Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, R.O.C. (e-mail: cychi@ee.nthu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2034935

curate acquisition of source directions is essential; for example, in radar, sonar, and microphone array systems, to name a few. The development in this discipline over the decades has resulted in a number of elegant signal processing concepts and techniques, for example, the high-resolution subspace approach to narrowband DOA estimation [1], [2], and its wideband extension using the frequency-domain signal subspace processing methods [3]–[5].

The main interest here is in DOA estimation of quasi-stationary signals, or direction finding of quasi-stationary signals (DF-QSS) for short. Quasi-stationary signals represent a class of nonstationary signals in which the statistics are locally static over a short period of time, but exhibit differences from one local time frame to another. Speech and audio signals, for instance, are often recognized as quasi-stationary signals. In fact, DOA estimation of audio signals has a practically very relevant application where the objective is to monitor birds in an airport for avoiding collisions of birds and aircrafts [6]. It also finds applications in microphone array processing of speech signals [7]. These real-world applications provide strong motivations for studying DF-QSS.

In the context of blind source separation (BSS), utilizing quasi-stationarity has received certain attention [8]-[14]. In BSS of quasi-stationary signals (BSS-QSS), we may classify the presently available methods into two main streams. The first stream is based on the least squares fitting (LSF) criterion [10]–[14]. Using a beautiful linear algebra tool called parallel factor analysis (PARAFAC) [15]-[17], an appealing identifiability result can be proven for LSF: Quasi-stationarity actually gives us the opportunity to identify sources in underdetermined mixing systems [12], [14]; i.e., when the number of sensors is less than the number of sources. However, this advantage comes with a challenge, namely that LSF is a multidimensional nonlinear optimization problem. Currently, LSF is handled by gradient descent [8], or various forms of alternating minimization such as the trilinear alternating least squares [17] and alternating-columns diagonal-centers (AC-DC) [18]. The second stream is based on the joint diagonalization (JD) criterion [9]. Under some mild conditions such as Gaussian (and quasi-stationary) sources, JD was shown to be equivalent to blind maximum-likelihood estimation [9]. JD is a very interesting matrix problem in essence and has attracted much attention; see the literature such as [9], [19], and [20], and the references therein. In JD, we are also faced with an optimization problem that is nonlinear and multidimensional.

In this paper, we propose a subspace approach to DF-QSS. Our work is based on the assumptions that:

- i) the sources are mutually uncorrelated and wide-sense quasi-stationary;
- ii) the noise is wide-sense stationary, but with an unknown, possibly nonwhite, spatial covariance; and
- iii) the sensor array exhibits a uniform linear structure.

Our idea is to exploit the subspace characteristics of the time-variant second-order statistics (SOSs) of the quasi-stationary source signals. Since this development involves subspace formed by the self Khatri-Rao (KR) product of the array response, we call the proposed approach the KR subspace approach. We should point out that the established KR subspace approach has its criterion different from the LSF and JD criteria used in BSS-QSS. A meaningful result with the KR subspace formulation is that for a physically underdetermined problem, the DOA estimation problem under the KR subspace can be 'virtually overdetermined' under some conditions. In essence, our identifiability analysis proves that for a sensor array of N elements, the degree of freedom under the KR subspace formulation is 2N - 2. This translates into an advantage that a KR subspace method can unambiguously identify up to 2N-2 sources. This is a significant improvement, compared to the conventional subspace SOS-based approach where the limit is N-1 sources. The KR subspace approach is also convenient to implement, since available subspace algorithms or concepts (e.g., MUSIC1) can be carried forward at a certain point of the KR formulation. In addition, the KR subspace formulation provides a convenient way of removing the spatial noise covariance matrix from the received signal SOSs, without knowing the noise covariance. Thus, the KR subspace approach is effective against spatially colored noise with unknown spatial covariance. Moreover, we will propose a novel idea where the KR problem dimension is reduced prior to the subspace processing. This dimension reduction idea helps save complexity in implementations, and it does so without losing the effective degree of freedom of the virtual array.

Since the proposed KR subspace approach enables underdetermined DOA estimation of quasi-stationary signals, it would be appropriate to mention other existing underdetermined DOA estimation approaches. Underdetermined DOA estimation is possible when the source signals are non-Gaussian stationary, and that necessitates the use of higher order statistics (HOSs). In the HOS-DF approach [21]–[24], it has been discovered that the higher order cumulants of the observed signals provide a virtual array structure that coincides with the one found in this paper when such cumulants are of fourth order. Nevertheless, the DF-QSS and HOS-DF approaches follow different formulations and they target different signals and applications. For speech signals, our simulation results will show that the proposed DF-QSS approach yields better DOA estimation accuracies than the HOS-DF approach (though it is fair to say that we would expect to see the opposite for strongly non-Gaussian, weakly quasi-stationary signals).

The development in this paper concentrates on the narrowband DOA estimation scenario. Using the frequency-domain signal subspace processing methods [3]–[5], one can readily extend the proposed KR subspace approach to the wideband DOA estimation scenario. The wideband extension is also considered and tested, where the simulation results will illustrate that the proposed wideband KR subspace algorithm is successful in identifying DOAs of real speech signals under underdetermined environments.

*Notations:* We denote matrices and vectors by boldfaced capital letters and lower-case letters, respectively. The space of complex (real) N-dimensional vectors is denoted by  $\mathbb{C}^N (\mathbb{R}^N)$ . Likewise, the space of complex (real)  $M \times N$  matrices is denoted by  $\mathbb{C}^{M \times N} (\mathbb{R}^{M \times N})$ . The *i*th element of a vector  $\mathbf{a} \in \mathbb{C}^N$  is denoted by  $a_i$ . The *i*th column of a matrix  $\mathbf{A} \in \mathbb{C}^{M \times N}$  is denoted by either  $\mathbf{a}_i \in \mathbb{C}^M$  or  $[\mathbf{A}]_i$ . The superscripts "T" and "H" stand for the transpose and conjugate transpose, respectively. For a given vector  $\mathbf{x} \in \mathbb{C}^N$ ,  $||\mathbf{x}||$  denotes its Euclidean norm. Its matrix counterpart, namely the Frobenius norm, is also denoted by  $|| \cdot ||$ . The expectation operator is denoted by  $\mathbf{E}\{\cdot\}$ . The  $M \times M$  identity matrix is denoted by  $\mathbf{I}_M$ . The notation  $\mathbf{1}_M$  stands for the  $M \times 1$  all-one vector.

For a given matrix  $\mathbf{A} \in \mathbb{C}^{M \times N}$ , the range space and the orthogonal complement subspace are denoted by  $\mathcal{R}(\mathbf{A})$  and  $\mathcal{R}^{\perp}(\mathbf{A})$ , respectively. The notation  $\operatorname{vec}(\cdot)$  stands for vectorization; i.e., if  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_N]$  then  $\operatorname{vec}(\mathbf{A}) = [\mathbf{a}_1^T, \dots, \mathbf{a}_N^T]^T$ . For a given vector  $\mathbf{a} \in \mathbb{C}^M$ ,  $\operatorname{Diag}(\mathbf{a})$  means a diagonal matrix with the diagonals given by  $a_1, \dots, a_M$ .

#### **II. PROBLEM STATEMENT**

We will first describe the signal model and assumptions of the DOA estimation problem considered in this paper. Then, we will study the second-order statistics model of the problem.

## A. Signal Model and Assumptions

Consider a scenario in which a number of K narrowband far-field sources are observed by a sensor array of N elements. The array is assumed to have a uniform linear array (ULA) structure. We denote by  $x_n(t)$  the observed signal of the *n*th sensor, and  $s_k(t)$  the signal emitted by the *k*th source. By letting  $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^T$  and  $\mathbf{s}(t) = [s_1(t), \dots, s_K(t)]^T$ , the received signal is modeled as

$$\mathbf{x}(t) = \mathbf{As}(t) + \mathbf{v}(t), \quad t = 0, 1, 2, \dots$$
 (1)

Here,  $\mathbf{v}(t) \in \mathbb{C}^N$  represents the spatial noise,  $\mathbf{A} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{N \times K}$  is the array response matrix where  $\theta_k \in -(\pi/2), (\pi/2)$  is the direction of arrival (DOA) of source k, and

$$\mathbf{a}(\theta) = \left[1, e^{-\frac{j2\pi d}{\lambda}\sin(\theta)}, \dots, e^{-\frac{j2\pi d}{\lambda}(N-1)\sin(\theta)}\right]^T \quad (2)$$

denotes the steering vector function with d and  $\lambda$  being the intersensor spacing and the signal wavelength, respectively. Some common assumptions are made as follows:

- A1) The source signals  $s_k(t), k = 1, ..., K$ , are mutually uncorrelated and have zero-mean.
- A2) The source DOAs  $\theta_k, k = 1, ..., K$ , are distinct to one another, i.e.,  $\theta_k \neq \theta_\ell$  for all  $k \neq \ell$ .

<sup>1</sup>MUSIC stands for multiple signal classification.

A3) The noise  $\mathbf{v}(t)$  is zero-mean wide-sense stationary (WSS) with covariance matrix  $\mathbf{C} \triangleq \mathrm{E}\{\mathbf{v}(t)\mathbf{v}^{H}(t)\}$ , and it is statistically independent of the source signals.

In this paper, the source signals are modeled as quasi-stationary processes. Specifically, our assumption is in a wide sense or up to second-order statistics [8]–[11].

A4) Each source signal  $s_k(t)$  is wide-sense quasi-stationary with frame length L; i.e.,

$$E\{|s_k(t)|^2\} = d_{mk}, \quad \forall t \in [(m-1)L, mL-1]$$
 for  $m = 1, 2, \dots$  (3)

Assumption A4) means that the second-order statistics of the source signals are time-varying, but that they remain static over a short period of time. Quasi-stationarity is a reasonable assumption for certain real-world signals, such as speech and audio.

## B. Local Covariance Model for Quasi-Stationary Signals

Under the quasi-stationarity assumption A4), we can define a local covariance matrix

$$\mathbf{R}_m = \mathbf{E}\{\mathbf{x}(t)\mathbf{x}^H(t)\} \in \mathbb{C}^{N \times N}$$
$$\forall t \in [(m-1)L, mL - 1] \quad (4)$$

where m = 1, 2, ... denotes the frame index. These local covariance matrices may be estimated by local averaging; that is,  $\hat{\mathbf{R}}_m = (1/L) \sum_{t=(m-1)L}^{mL-1} \mathbf{x}(t) \mathbf{x}^H(t)$ . With A1), A3), and A4), we can express  $\mathbf{R}_m$  as

$$\mathbf{R}_m = \mathbf{A}\mathbf{D}_m\mathbf{A}^H + \mathbf{C} \tag{5}$$

where  $\mathbf{D}_m = \text{Diag}(d_{m1}, d_{m2}, \dots, d_{mK}) \in \mathbb{R}^{K \times K}$  is the source covariance matrix at frame m. Now, suppose that we have a number of local covariance matrices  $\mathbf{R}_1, \dots, \mathbf{R}_M$  available, where M is the total number of frames. Our goal is to estimate the DOAs  $\theta_1, \dots, \theta_K$  from  $\mathbf{R}_1, \dots, \mathbf{R}_M$ , without information of both the local source covariances  $\mathbf{D}_1, \dots, \mathbf{D}_M$  and the spatial noise covariance  $\mathbf{C}$ .

## III. KHATRI-RAO SUBSPACE APPROACH

This section presents the proposed approach, the Khatri–Rao subspace approach to DOA estimation of quasi-stationary signals. In Section III-A, we first give a brief review on some basic concepts regarding Khatri–Rao product. Then, we describe the KR subspace formulation for the DOA estimation problem in Section III-B. In particular, the KR subspace DOA estimation criterion, the noise covariance elimination method, and the DOA identifiability conditions of the established criterion are considered. These are followed by Section III-C, where we introduce a KR dimension reduction method for computational efficiency and use that to develop KR subspace algorithms.

## A. Basic Concepts of Khatri-Rao Product

We use the symbol " $\odot$ " to denote the Khatri–Rao (KR) product. For two matrices  $\mathbf{A} \in \mathbb{C}^{n \times k}$  and  $\mathbf{B} \in \mathbb{C}^{m \times k}$  of identical number of columns, the KR product of them is given by

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1, \dots, \mathbf{a}_k \otimes \mathbf{b}_k] \in \mathbb{C}^{nm \times k}$$
(6)

where  $\otimes$  denotes the Kronecker product. In this work it is sufficient to know the definition of the Kronecker product of two vectors  $\mathbf{a} \in \mathbb{C}^n$  and  $\mathbf{b} \in \mathbb{C}^m$ , which is given by

$$\mathbf{a} \otimes \mathbf{b} = \begin{bmatrix} a_1 \mathbf{b} \\ a_2 \mathbf{b} \\ \vdots \\ a_n \mathbf{b} \end{bmatrix} = \operatorname{vec}(\mathbf{b} \mathbf{a}^T).$$
(7)

The KR product plays an important role in PARAFAC [17]. One useful property is given as follows.

*Property 1:* Let  $\mathbf{A} \in \mathbb{C}^{n \times k}$ ,  $\mathbf{B} \in \mathbb{C}^{m \times k}$ , and  $\mathbf{d} \in \mathbb{C}^k$ . Also denote  $\mathbf{D} = \text{Diag}(\mathbf{d})$ . Then

$$\operatorname{vec}(\mathbf{A}\mathbf{D}\mathbf{B}^{H}) = (\mathbf{B}^{*} \odot \mathbf{A})\mathbf{d}.$$
 (8)

Property 1 has been used in the literature such as [17] and [18], often in a very concise manner. To make this paper self-contained, we provide the proof of Property 1 in Appendix A.

The rank properties of KR product have a strong connection to a concept called *Kruskal rank*, or *k*-rank for short. The *k*-rank of a matrix **A**, denoted by krank(**A**), is said to be equal to *r* when every collection of *r* columns of **A** is linearly independent but there exists a collection of r+1 linearly dependent columns. By contrast, the rank of **A** is the maximal number of linearly independent columns, the definition of which is more relaxed than that of the *k*-rank. Thus we have rank(**A**)  $\geq$  krank(**A**). *k*-rank has an interesting property for KR product [16], [25], [26]:

*Property 2:* For two matrices  $\mathbf{A} \in \mathbb{C}^{n \times k}$  and  $\mathbf{B} \in \mathbb{C}^{m \times k}$ , with krank $(\mathbf{A}) \geq 1$  and krank $(\mathbf{B}) \geq 1$ , it holds true that

$$\operatorname{krank}(\mathbf{A} \odot \mathbf{B}) \ge \min\{k, \operatorname{krank}(\mathbf{A}) + \operatorname{krank}(\mathbf{B}) - 1\}.$$
(9)

#### B. Khatri–Rao Subspace Criterion

For the DOA estimation problem formulated in Section II, let us apply Property 1 to the local covariance model in (5) to obtain

$$\mathbf{y}_m \triangleq \operatorname{vec}(\mathbf{R}_m) = \operatorname{vec}(\mathbf{A}\mathbf{D}_m\mathbf{A}^H) + \operatorname{vec}(\mathbf{C}) = (\mathbf{A}^* \odot \mathbf{A})\mathbf{d}_m + \operatorname{vec}(\mathbf{C})$$
(10)

By stacking  $[\mathbf{y}_1, \ldots, \mathbf{y}_M] \triangleq \mathbf{Y}$ , we can write

$$\mathbf{Y} = (\mathbf{A}^* \odot \mathbf{A}) \boldsymbol{\Psi}^T + \operatorname{vec}(\mathbf{C}) \mathbf{1}_M^T$$
(11)

where we recall  $\mathbf{1}_M = [1, \dots, 1]^T \in \mathbb{R}^M$ , and

$$\Psi = [\boldsymbol{d}_{1}, \dots, \boldsymbol{d}_{M}]^{T}$$

$$= \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1K} \\ d_{21} & d_{22} & \cdots & d_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ d_{M1} & d_{M2} & \cdots & d_{MK} \end{bmatrix}.$$
(12)

It is interesting to note from (10) or (11) that  $\mathbf{y}_m$  is reminiscent of an array signal model where  $(\mathbf{A}^* \odot \mathbf{A}) \in \mathbb{C}^{N^2 \times K}$  is virtually the array response matrix and  $d_m$  the source signal vector. The virtual array dimension, given by  $N^2$ , is greater than the physical array dimension N for N > 1, and that essentially provides us with the capability of processing cases where there are less sensors than sources. This will be shown in the ensuing development.

We should note that each column of  $\Psi$  in (12) describes the powers of the respective source signal with respect to frames, or simply speaking, the long-term power profile over time. Let us assume the following:

A5) The matrix  $[\Psi \ \mathbf{1}_M] \in \mathbb{R}^{M \times (K+1)}$  is of full column rank.

Assumption A5) physically implies the followings: First, the source power distributions over the time frames (or the columns of  $\Psi$ ) are different so that  $\Psi$  can maintain a full column rank condition. Second, any linear combination of the sources cannot result in a WSS source (i.e., for any coefficients  $c_1, \ldots, c_K \in \mathbb{C}$ , the signal  $\sum_{k=1}^{K} c_k s_k(t)$  cannot be WSS), otherwise  $\mathbf{1}_M$  can be a linear combination of the columns of  $\Psi$  which violates A5). While the necessary condition for fulfilling A5) is to have M > K + 1, in practice it would be desirable to use a much larger M so as to obtain sufficient long-term source power variations to well satisfy A5). Certainly, availability of large Mdepends on applications. For example, for speech applications where L is generally proportional to a physical time duration of 20 to 25 ms, we can obtain more than 40 frames (or M > 40) in 1 s. Moreover, increasing the value of M would generally be useful in improving the conditioning of  $\Psi$ . Hence, in the presence of estimation errors in the local covariances  $\mathbf{R}_m$ , employing a larger M would be helpful in suppressing the subsequent error effects.

Under A5), we can eliminate the unknown noise covariance effectively and easily. The noise covariance elimination is done by denoting an orthogonal complement projector  $\mathbf{P}_{\mathbf{I}_M}^{\perp} = \mathbf{I}_M - (1/M)\mathbf{1}_M\mathbf{1}_M^T$ , and then by performing a projection

$$\mathbf{Y}\mathbf{P}_{\mathbf{1}_{M}}^{\perp} = \left[ (\mathbf{A}^{*} \odot \mathbf{A}) \boldsymbol{\Psi}^{\mathbf{T}} + \operatorname{vec}(\mathbf{C}) \mathbf{1}_{\mathrm{M}}^{\mathrm{T}} \right] \mathbf{P}_{\mathbf{1}_{M}}^{\perp} = (\mathbf{A}^{*} \odot \mathbf{A}) \left( \mathbf{P}_{\mathbf{1}_{M}}^{\perp} \boldsymbol{\Psi} \right)^{T}.$$
(13)

Under A5), we have rank( $\mathbf{P}_{\mathbf{1}_M}^{\perp} \Psi$ ) = rank( $\Psi$ ) = *K*. In other words, the noise covariance elimination operation does not damage the rank condition of the covariance model.

Now consider the subspaces of  $\mathbf{YP}_{\mathbf{1}_M}^{\perp}$ . For ease of exposition of idea, we assume for the time being that  $(\mathbf{A}^* \odot \mathbf{A})$  is of full column rank. We will soon provide conditions under which this assumption is valid. When both  $(\mathbf{A}^* \odot \mathbf{A})$  and  $\mathbf{P}_{\mathbf{1}_M}^{\perp} \Psi$  in (13) have full column rank, we can have [27]

$$\mathcal{R}(\mathbf{A}^* \odot \mathbf{A}) = \mathcal{R}\left(\mathbf{Y}\mathbf{P}_{\mathbf{1}_M}^{\perp}\right). \tag{14}$$

Denote the singular value decomposition (SVD) of  $\mathbf{YP}_{\mathbf{1}_M}^{\perp}$  by

$$\mathbf{Y}\mathbf{P}_{\mathbf{1}_{M}}^{\perp} = \begin{bmatrix} \mathbf{U}_{s} \, \mathbf{U}_{n} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{s} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{s}^{H} \\ \mathbf{V}_{n}^{H} \end{bmatrix}$$
(15)

where  $\mathbf{U}_s \in \mathbb{C}^{N^2 \times K}$  and  $\mathbf{V}_s \in \mathbb{C}^{M \times K}$  are the left and right singular matrices associated with the nonzero singular values, respectively,  $\mathbf{U}_n \in \mathbb{C}^{N^2 \times (N^2 - K)}$  and  $\mathbf{V}_n \in \mathbb{C}^{M \times (N^2 - K)}$  are the left and right singular matrices associated with the zero singular values, respectively, and  $\mathbf{\Sigma}_s \in \mathbb{R}^{K \times K}$  is a diagonal matrix whose diagonals contain the nonzero singular values. Based on the standard SVD result that

$$\mathcal{R}^{\perp}(\mathbf{A}^* \odot \mathbf{A}) = \mathcal{R}^{\perp}(\mathbf{U}_s) = \mathcal{R}(\mathbf{U}_n)$$
(16)

we know that the source DOAs satisfy

$$\mathbf{U}_{n}^{H}[\mathbf{A}^{*}\odot\mathbf{A}]_{k} = \mathbf{U}_{n}^{H}(\mathbf{a}^{*}(\theta_{k})\otimes\mathbf{a}(\theta_{k})) = \mathbf{0}$$
(17)

for k = 1, ..., K. We hence propose the following KR subspace criterion for DOA estimation of quasi-stationary sources

find 
$$\theta$$
  
such that  $\mathbf{U}_{n}^{H}(\mathbf{a}^{*}(\theta) \otimes \mathbf{a}(\theta)) = \mathbf{0}, \quad \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$  (18)

To satisfy the KR subspace criterion in (18), we can use the idea of MUSIC in the conventional subspace-based DOA estimation approach. The further details will be considered in the next subsection.

Like the development in the subspace discipline, it is crucial to determine conditions under which (18) is satisfied *only if*  $\theta$  is a true DOA, viz., the identifiability conditions. The following two propositions provide the key results of the theoretical identifiability of the KR subspace approach.

*Proposition 1:* Under A2), the sufficient and necessary condition for the virtual array response matrix  $(\mathbf{A}^* \odot \mathbf{A})$  to yield full column rank (or simply rank = K) is when

$$K \le 2N - 1.$$

Proposition 2: Assume that A1)–A5) hold. The KR subspace criterion in (18) is achieved by any one of the true angles  $\theta_k$ ,  $k = 1, \dots, K$ , if and only if

$$K \le 2N - 2. \tag{19}$$

The proofs of Propositions 1 and 2 are given in Appendix B. Simply speaking, the idea behind the proofs is to use the KR product krank property in Property 2, and to have explicit exposition of the structures of  $\mathbf{A}^* \odot \mathbf{A}$ . Proposition 1 justifies our assumption in the KR subspace development above, and Proposition 2 provides the identifiability condition of the KR subspace criterion. In particular, Proposition 2 gives the appealing implication that DOA estimation can be done for underdetermined cases under the KR subspace framework.

## C. KR-Based DOA Estimators With Dimension Reduction

We could develop a KR-based DOA estimator by directly applying a subspace method (say, MUSIC) to the KR subspace criterion (18). But, prior to applying a subspace method, we can reduce the problem dimension as hinted in the proof of Propositions 1 and 2 in Appendix B.

Authorized licensed use limited to: National Tsing Hua University. Downloaded on March 22,2010 at 08:29:46 EDT from IEEE Xplore. Restrictions apply.

The dimension reduction idea is proposed as follows. In Appendix B, it was indicated that the virtual array response matrix  $\mathbf{A}^* \odot \mathbf{A}$  can actually be characterized as

$$\mathbf{A}^* \odot \mathbf{A} = \mathbf{G}\mathbf{B} \tag{20}$$

where  $\mathbf{G} \in \mathbb{C}^{N^2 \times (2N-1)}$  is given by

$$\mathbf{G} = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{C}^{N^2 \times (2N-1)}$$

$$(21)$$

 $\mathbf{B} = [\mathbf{b}(\theta_1), \dots, \mathbf{b}(\theta_K)] \in \mathbb{C}^{(2N-1) \times K}$  is a dimension-reduced virtual array response matrix with

$$\mathbf{b}(\theta) = \left[e^{(N-1)\frac{j2\pi d}{\lambda}\sin(\theta)}, \dots, e^{\frac{j2\pi d}{\lambda}\sin(\theta)}, 1, \dots, e^{-(N-1)\frac{j2\pi d}{\lambda}\sin(\theta)}\right]^T.$$
 (22)

It can be verified from (21) that **G** is column orthogonal. Let  $\mathbf{W} = \mathbf{G}^T \mathbf{G}$ . One can show that

$$\mathbf{W} = \text{Diag}(1, 2, \dots, N - 1, N, N - 1, \dots, 2, 1).$$
(23)

From the noise-covariance-eliminated data matrix in (13), we can reduce the problem size by performing a linear processing

$$\tilde{\mathbf{Y}} = \mathbf{W}^{-1/2} \mathbf{G}^T \left[ \mathbf{Y} \mathbf{P}_{\mathbf{1}_M}^{\perp} \right] = \mathbf{W}^{-1/2} \mathbf{G}^T \left[ (\mathbf{A}^* \odot \mathbf{A}) (\mathbf{P}_{\mathbf{1}_M}^{\perp} \Psi)^T \right] = \mathbf{W}^{1/2} \mathbf{B} \left( \mathbf{P}_{\mathbf{1}_M}^{\perp} \Psi \right)^T.$$
(24)

Note that the dimension reducing transformation  $\mathbf{W}^{-1/2}\mathbf{G}^T$  has orthonormal rows, and hence in the presence of covariance estimation errors the transformation would not cause additional coloring of the errors. A subspace method is then applied to  $\tilde{\mathbf{Y}} \in \mathbb{C}^{(2N-1)\times M}$ , instead of  $\mathbf{YP}_{\mathbf{I}_M}^{\perp} \in \mathbb{C}^{N^2 \times M}$ . Apparently, incorporating this dimension reduction has the advantage of complexity reduction, especially with the SVD subspace extraction.

In order to benefit from the dimension-reduced KR formulation described earlier, we propose to apply MUSIC to the KR subspace criterion in (18) with  $\mathbf{YP}_{\mathbf{I}_M}^{\perp}$  being replaced by  $\tilde{\mathbf{Y}}$  in (24). We call the resultant method *the KR-MUSIC method*. The pseudocode of KR-MUSIC is provided in Table I. We can also consider variations where the MUSIC steps are replaced by some other conceptually related procedure, such as the Capon

TABLE I SUMMARY OF THE KR-MUSIC ALGORITHM

**Given** a received signal sequence  $\{\mathbf{x}(t)\}_{t=0}^{T-1}$ , a source number K, and a frame length L where L divides T.

Step 1. Compute the local covariance estimates

$$\hat{\mathbf{R}}_m = \frac{1}{L} \sum_{t=(m-1)L}^{mL-1} \mathbf{x}(t) \mathbf{x}^H(t)$$

for m = 1, ..., M, where M = T/L. Then, form a data matrix  $\hat{\mathbf{Y}} = [\operatorname{vec}(\hat{\mathbf{R}}_1), ..., \operatorname{vec}(\hat{\mathbf{R}}_M)].$ 

Step 2. (Noise covariance elimination)  $\dot{\mathbf{Y}} = \hat{\mathbf{Y}} \mathbf{P}_{\mathbf{1}_M}^{\perp}$ , where  $\mathbf{P}_{\mathbf{1}_M}^{\perp} = \mathbf{I}_M - \frac{1}{M} \mathbf{1}_M \mathbf{1}_M^T$ .

Step 3. (Dimension reduction)  $\tilde{\mathbf{Y}} = \mathbf{W}^{-1/2} \mathbf{G}^T \bar{\mathbf{Y}}$ , where  $\mathbf{G}$  and  $\mathbf{W}$  are defined in (21) and (23), respectively.

Step 4. (Subspace extraction) Perform SVD on  $\tilde{\mathbf{Y}}$ 

$$\tilde{\mathbf{Y}} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$$

where  $\mathbf{U} \in \mathbb{C}^{(2N-1) \times K}$  and  $\mathbf{V} \in \mathbb{C}^{M \times K}$  are the left and right singular matrices, respectively, and  $\boldsymbol{\Sigma} \in \mathbb{R}^{K \times K}$  is the (diagonal) singular value matrix with the singular values being arranged ir descending order. Then, obtain the noise subspace matrix

 $\mathbf{U}_n = [\mathbf{u}_{K+1}, \dots, \mathbf{u}_{2N-1}] \in \mathbb{C}^{(2N-1) \times (2N-1-K)}.$ 

Step 5. (MUSIC operation) Compute the DOA spectrum

$$\mathbf{P}_{\text{KR-MUSIC}}(\theta) = \frac{1}{\|\mathbf{U}_n^H \mathbf{W}^{1/2} \mathbf{b}(\theta)\|^2}$$

over  $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ , and pick the K largest peaks of P<sub>KR-MUSIC</sub>( $\theta$ ) as the DOA estimates.

and ESPRIT<sup>2</sup> methods [2]. In the application of the Capon method to the dimension-reduced KR formulation, or simply the *KR-Capon algorithm*, the procedures are basically the same as those in Table I except that the spatial spectrum in Step 5 should be replaced by

$$P_{\text{KR-Capon}}(\theta) = \frac{1}{\mathbf{b}^{H}(\theta)\mathbf{W}^{1/2}(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{H})^{-1}\mathbf{W}^{1/2}\mathbf{b}(\theta)}$$
(25)

and that the SVD in Step 4 is not needed.

## **IV. SIMULATION RESULTS**

We provide several sets of simulation results to demonstrate the performance of the proposed KR subspace approach. In all the simulation examples below, the signal-to-noise ratio (SNR) is defined as

SNR = 
$$\frac{\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{\|\mathbf{As}(t)\|^2\}}{\mathbb{E}\{\|\mathbf{v}(t)\|^2\}}$$

where T = LM is the total number of samples. The root mean square (RMS) angle error is used as our performance measure and it is defined as

$$\sqrt{\mathbf{E}\left\{\frac{1}{K}\sum_{k=1}^{K}|\theta_{k}-\hat{\theta}_{k}|^{2}\right\}}$$

<sup>2</sup>ESPRIT stands for estimation of signal parameters via rotational invariance techniques.

TABLE II GENERATION OF SYNTHETIC QUASI-STATIONARY SIGNALS

Given	bounds $L_{low}$ , $L_{upp}$ , and a sequence length T.
Step 1.	$T_{cur} := 0.$
Step 2.	Randomly generate L following a uniform distribution on
	$[L_{low}, L_{upp}].$
Step 3.	Randomly generate $\sigma_s$ following a uniform distribution on [0, 1].
Step 4.	For $t = T_{cur}, T_{cur} + 1, \dots, T_{cur} + L - 1$ , randomly generate
	$s(t) = s_R(t) + js_I(t)$ where $s_R(t)$ and $s_I(t)$ are i.i.d. Laplacian
	distributed with zero mean and variance $\sigma_s^2/2$ .
Step 5.	$T_{cur} := T_{cur} + L$ . If $T_{cur} < T$ then goto Step 2.
Output	$\{s(t)\}_{t=0}^{T-1}$ .
-	

where  $\theta_k$  and  $\hat{\theta}_k$  denote the true and estimated DOAs, respectively. The first three simulation examples are based on the standard narrowband scenario, where the performance of the KR subspace methods is examined under various conditions. We also compared the KR subspace methods with some other existing methods, such as the standard MUSIC algorithm, and a fourth-order cumulant based HOS-MUSIC algorithm called *4-MUSIC* [21]–[23].

In the first three narrowband simulation examples, the quasistationary source signals  $s_k(t)$  were synthetically generated by a random generation procedure given in Table II. This procedure generates a locally stationary zero-mean complex Laplacian process, with the variances randomly varying from one frame to another. Moreover, the duration of each local time frame (L) is randomly drawn following a uniform distribution on  $[L_{low}, L_{upp}]$ . The purpose of doing so is to simulate a more realistic situation where the local stationary periods are uncertain and varying; e.g., in speech. Strictly speaking, such signals violate A4) which assumes all local stationary intervals to be the same and known. Nevertheless, the simulation results to be presented soon will show that the KR subspace approach is not too sensitive to the effects of uncertain local stationary intervals.

In the last simulation example, we simulate a realistic wideband microphone array processing system with real speech signals employed as the sources.

# A. Simulation Example 1: Underdetermined Narrowband DOA Estimation

We consider a narrowband, underdetermined case where (N,K) = (4,6). The true DOAs are  $\{\theta_1,\ldots,\theta_K\} =$  $\{-65^{\circ}, -40^{\circ}, -20^{\circ}, 10^{\circ}, 25^{\circ}, 42^{\circ}\}$ . The sensor array is uniform linear with  $d/\lambda = 1/2$ . The SNR is 10 dB. The spatial noise  $\mathbf{v}(t)$  is zero-mean, uniformly white complex Gaussian. The synthetic signal generation procedure in Table II is used to generate the source signals  $s_k(t)$ , with allowable range of the frame periods  $[L_{\text{low}}, L_{\text{upp}}] = [300, 700]$ . In Fig. 1 we illustrate a segment of the synthetic quasi-stationary signals. Notice that the dashed lines in the figure mark the actual local stationary intervals of the signals. As we can see, the frame intervals of the 6 source signals are not uniform and not synchronized. We apply the KR subspace methods by choosing a fixed frame period of L = 512, which is approximately the mean of the actual frame periods. The number of frames is set to M = 50. Fig. 2(a) shows one realization of the KR-MUSIC DOA spectrum. As seen, the peaks of the spectra are in good agreement with the true DOAs. This also demonstrates that the



Fig. 1. Illustration of a segment of synthetic quasi-stationary source signals  $s_k(t)$ . The dashed lines mark the local stationary intervals. Due to space limit, only the real parts of the signals are plotted. The appearance of the imaginary signal components is similar to that of the real.



Fig. 2. DOA spectra of the various algorithms in the narrowband underdetermined case. The dashed lines mark the true DOA positions. (a) KR-MUSIC. (b) 4-MUSIC.

KR subspace approach exhibits robustness against uncertainties and asynchronism of the actual locally stationary intervals. Fig. 2(b) shows the 4-MUSIC DOA spectrum corresponding to the same realization. We see that 4-MUSIC also exhibits consistent spectral peaks with respect to the true DOAs, but KR-MUSIC shows sharper spectral peaks in comparison.

A Monte Carlo simulation was carried out to evaluate the RMS angle error performance of KR-MUSIC and 4-MUSIC. The number of Monte Carlo trials is 1000. The source signals were randomly generated by Table II on a per-trial basis. The other simulation settings are the same as the previous. In this simulation we included KR-Capon, the variation of KR-MUSIC in (25). The results are shown in Fig. 3. In Fig. 3(a) the RMS angle errors are plotted against the SNRs. We can see that KR-MUSIC gives the best RMS angle error performance. 4-MUSIC is competitive in performance at high SNRs, but its performance gap relative to KR-MUSIC becomes significant for SNRs lower than 0 dB.

In Fig. 3(b) we show the RMS angle error performance of the KR methods when the number of frames M increases. The frame length is fixed at L = 512, and the SNR is 14 dB. 4-MUSIC was also tested with the same amount of total samples; i.e., T = LM. We see that KR-MUSIC generally offers the best RMS angle error performance, except at M = 10 where 4-MUSIC prevails. This shows that M should not be too small in the KR subspace approach.

An interesting question is how the chosen frame length L affects the performance of the KR subspace methods, fixing the data size T = LM. Fig. 3(c) shows a simulation result where we fix the total signal length to T = 24000 and then vary L to evaluate the respective RMS angle errors. The SNR is fixed at 14 dB. Interestingly, it is observed that the best frame length for KR-MUSIC is about L = 200, which is even smaller than the minimum locally stationary interval of the generated signals  $L_{\text{low}} = 300$ . However, Fig. 3(c) also illustrates that L should not be overly small. In the figure we can see that KR-MUSIC shows mild performance degradation when L is decreased down to 50, and that KR-Capon shows considerable performance degradation when  $L \leq 100$ . This is caused by estimation errors of the local covariances, which, for insufficient number of local samples L, become significant. Nevertheless, this set of empirical results suggests that for a fixed data size, trying to obtain more frames by decreasing the frame length L tends to a better option for performance improvement than the vice versa.

## B. Simulation Example 2: Narrowband DOA Estimation of Locally Gaussian Sources

In the KR subspace approach, what is utilized is time-variant SOSs. Hence, the KR subspace methods should be insensitive to the distributions of the sources. This simulation example aims to verify this expectation. We employ locally complex Gaussian sources in place of the locally complex Laplacian sources used in Simulation Example 1. To do so, we simply modify Step 4 of the source generation procedure in Table II, from Laplacian to Gaussian. All other simulation settings are exactly the same as Simulation Example 1. Fig. 4 shows the resultant RMS angle errors with respect to SNRs. The figure indicates that 4-MUSIC is able to provide reasonable DOA estimation performance, but yields quite a significant performance gap compared to KR-MUSIC or KR-Capon. Such a performance gap

- KR-Capon -KR-MUSIC 1.2 Root mean square angle error (Degree) 0

Fig. 3. RMS angle error performance of the various algorithms in the narrowband underdetermined case. (a) (M, L) fixed at (50,512). (b) SNR fixed at 14 dB, L fixed at 512. (c) T = ML fixed at 24 000, SNR fixed at 14 dB.

(c)

is somewhat expected: The fourth-order cumulants, on which 4-MUSIC crucially depends, are supposed to vanish for locally Gaussian processes. Nevertheless, the estimated fourth-order cumulants, due to sampling of finite lengths of data, are not totally zero; and thus 4-MUSIC can still give reasonable DOA estimates. Comparing Fig. 4 and its locally Laplacian counterpart





Fig. 4. RMS angle error performance of the various algorithms when the sources are locally complex Gaussian.

in Fig. 3(a), we found that the RMS angle errors of KR-MUSIC and KR-Capon do not change with the distributions. This verifies the distribution insensitivity of the KR-based methods.

# C. Simulation Example 3: Narrowband DOA Estimation in the Presence of Spatially Nonuniform and Nonwhite Noise

This simulation example considers a narrowband overdetermined case where (N, K) = (4, 3), (M, L) = (50, 512), and where the true DOAs are given by  $\{\theta_1, \theta_2, \theta_3\} = \{-18^\circ, 5^\circ, 25^\circ\}$ . We tried two kinds of noise models. The first is a spatially nonuniform white complex Gaussian noise with covariance

$$\mathbf{C} = \sigma_v^2 \text{Diag}(0.8, 0.9, 1.1, 1.2).$$
(26)

With the other simulation settings being the same as Simulation Example 1, we performed a Monte Carlo simulation for KR-MUSIC, 4-MUSIC, and the conventional MUSIC algorithm which can be applied in this overdetermined example. Fig. 5(a) shows the performance of the various algorithms. We observe that at high SNRs where the noise covariance becomes negligible, it is the conventional MUSIC algorithm that prevails. However, at low SNRs, KR-MUSIC gives lower RMS estimation errors than the conventional MUSIC. This verifies that the proposed KR subspace approach is immune to the unknown noise covariance effects. In fact, the conventional MUSIC is known to be sensitive to colored or nonuniform spatial noise covariance effects [28]. Fig. 5(a) also shows that 4-MUSIC is robust against Gaussian noise with unknown noise covariance.

The second noise model we tested is a uniform, spatially colored complex Gaussian noise with the (i, k)th entry of the covariance given by

$$C_{i,k} = \sigma_v^2 0.85^{|i-k|}.$$
 (27)

The results are shown in Fig. 5(b). We see that the performance behaviors are the same as those of the previous noise model.

# D. Simulation Example 4: Narrowband DOA Estimation of Two Closely Spaced Sources

This example examines the performance of KR-MUSIC when dealing with two closely spaced sources. We set



Fig. 5. RMS angle error performance of the various algorithms in the presence of nonuniform or colored noise. (a) Spatially nonuniform white noise. (b) Spatially colored noise.



Fig. 6. RMS angle error performance for two closely spaced sources.

(N, K) = (4, 2). We tried two different settings of the DOAs, namely  $\{\theta_1, \theta_2\} = \{0^\circ, 2^\circ\}$  and  $\{\theta_1, \theta_2\} = \{0^\circ, 5^\circ\}$ . The rest of the simulation settings are the same as Simulation Example 1. The results are shown in Fig. 6. For the very closely spaced case  $\{\theta_1, \theta_2\} = \{0^\circ, 2^\circ\}$ , we observe that KR-MUSIC

yields better RMS error performance than the conventional MUSIC algorithm. The performance gain of KR-MUSIC is likely due to the higher degree of freedom (which is 2N - 2).

For the case  $\{\theta_1, \theta_2\} = \{0^\circ, 5^\circ\}$ , we can see that KR-MUSIC provides better performance than MUSIC for SNRs lower than 21 dB; and vice versa for higher SNRs. Our speculation with this performance behavior is the following: While the degree of freedom available in MUSIC (i.e., N - 1) is lower than that of KR-MUSIC, MUSIC may be comparatively less susceptible to SOS estimation errors. The latter may be the reason for MUSIC to surpass KR-MUSIC at high SNRs.

# *E. Simulation Example 5: Underdetermined Wideband DOA Estimation of Speech Signals*

This example considers a wideband microphone array processing system, with real speech signals employed as the sources. While our development in the last section is primarily based on the standard narrowband array processing model, the proposed KR subspace approach can be readily applied to the wideband scenario by incorporating the frequency-domain approach [3]–[5].

To keep this work self-contained, we give a concise description of the ideas and the problem settings. Let  $x_n(t)$  be the signal recorded directly by the *n*th sensor. In the wideband case, each  $x_n(t)$  is modeled as a superposition of time-shifted source signals [29]

$$x_n(t) = \sum_{k=1}^{K} s_k(t - \Delta(n-1)\sin(\theta_k)) + v_n(t),$$
  
$$t = 0, 1, 2, \dots \quad (28)$$

where uniform linear array structures are assumed once again. Here,  $\Delta = d/(cT_s)$  is a constant where d is the inter-sensor spacing, c is the signal propagation speed, and  $T_s = 1/F_s$  is the sampling interval (in seconds). The idea of frequency-domain processing is to decouple the wideband model (28) into a multitude of narrowband models. Let us define the short-time Fourier transform (STFT) of a signal y(t) to be

$$\check{y}(f,t) = \sum_{u=0}^{N_{\rm STFT}-1} y(t+u) e^{-j2\pi f u}$$
(29)

where  $N_{\text{STFT}}$  is the STFT window length, and  $f \in [-1/2, 1/2]$ is the frequency (normalized). By applying STFT to  $x_n(t)$  and by denoting  $\check{\mathbf{x}}(f,t) = [\check{x}_1(f,t), \dots, \check{x}_N(f,t)]^T$ , we obtain the following approximate signal formulation [3]

$$\check{\mathbf{x}}(f,t) \approx \mathbf{A}(f)\check{\mathbf{s}}(f,t) + \check{\mathbf{v}}(f,t), \tag{30}$$

where  $\check{\mathbf{s}}(f,t)$  and  $\check{\mathbf{v}}(f,t)$  are defined in the same way as  $\check{\mathbf{x}}(f,t)$ , and  $\mathbf{A}(f) = [\mathbf{a}(f,\theta_1), \dots, \mathbf{a}(f,\theta_K)]$  is a frequency dependent array response matrix with

$$\mathbf{a}(f,\theta) = \left[1, e^{-j2\pi\Delta f \sin(\theta)}, \dots, e^{-j2\pi(N-1)\Delta f \sin(\theta)}\right]^T.$$
(31)

Equation (30) is known to be a good approximation for sufficiently large  $N_{\text{STFT}}$ . We see that fixing f, (30) is essentially equivalent to the narrowband signal model. Hence, for each fixed f we can apply KR-MUSIC to (30).



Fig. 7. The speech waveforms for the wideband microphone array simulation.

In the frequency-domain approach, a necessary step is to combine the subspaces at various frequencies to obtain a DOA spectrum fusion. Let  $\mathcal{B} \subset [0, 1/2]$  be the (normalized) frequency band of the source signals, and denote  $\mathcal{B}_d = \{f = n/N_{\text{STFT}} | n = 0, 1, \dots, N_{\text{STFT}}/2, f \in \mathcal{B}\}$  to be a discretized version of  $\mathcal{B}$ . The following formula is used to compute a DOA spectrum:

$$P_{\text{KR-MUSIC}}(\theta) = \frac{1}{\sum_{f \in \mathcal{B}_d} \|\mathbf{U}_n^H(f)\mathbf{W}^{1/2}\mathbf{b}(f,\theta)\|^2} \quad (32)$$

where  $\mathbf{U}_n(f)$  denotes the KR noise subspace obtained from (30) with f being fixed,  $\mathbf{b}(f,\theta)$  is an extended steering vector of (31) [cf., (22)], and  $\mathbf{W}^{1/2}$  is given in (23). The combining in (32) is known as the *incoherent signal subspace method* (ISSM) in the literature [5].

The settings for simulating the wideband microphone array system are as follows: The sampling frequency is  $F_s = 8$  kHz. The inter-sensor spacing is d = 4.25 cm, which, under a sound propagation speed of c = 340 m/sec., leads to  $\Delta = 1$ . The STFT window length is  $N_{\rm STFT} = 64$ . The frame length is L = 200, which corresponds to a time period of 25 ms. In speech processing, the stationary time of speech signals is generally assumed to be around 20 to 25 ms. The frequency band processed is  $\mathcal{B} = [500/8000, 3500/8000]$  (in practice, frequency points close to DC or Nyquist may have very weak signal components and hence should be discarded). The source signals are real speech, and they are plotted in Fig. 7.

We set up a underdetermined case where (N, K) = (4,5). The true DOAs are  $\{\theta_1, \ldots, \theta_K\} = \{-65^\circ, -40^\circ, -20^\circ, 10^\circ, 25^\circ\}$ . Using 100 frames (or M = 100) which corresponds to a total signal duration of 2.51 s, we ran (32) to obtain a wideband KR-MUSIC spectrum, plotted in Fig. 8(a). We see that the spectral peaks are in proximity to the true DOAs. Fig. 8(b) shows a result where the KR-MUSIC core is replaced by the 4-MUSIC in the frequency domain processing in (32). As seen, the resultant wideband 4-MUSIC spectrum manages to distinguish two sources only. We repeated the simulation by increasing the number of frames to M = 450 (total signal duration = 11.26 s), and the results are shown in Fig. 9. We see that the true DOA positions are roughly distinguishable in 4-MUSIC, but still not as good as what KR-MUSIC offers.



Fig. 8. DOA spectra of KR-MUSIC and 4-MUSIC in the wideband real-speech case. L = 200, M = 100. The gray dashed lines mark the true DOA positions. (a) KR-MUSIC, with ISSM. (b) 4-MUSIC, with ISSM.

## V. CONCLUSION AND DISCUSSION

This paper has addressed the DOA estimation problem of mutually uncorrelated wide-sense quasi-stationary signals in the presence of wide-sense stationary noise and with uniform linear arrays. We have established a KR subspace approach that judiciously utilizes the long-term time-variant characteristics of quasi-stationary SOSs to achieve two advantages. First, the KR subspace approach can be applied to cases where there are less sensors than sources. Our theoretical identifiability analysis has revealed that the proposed KR subspace approach can uniquely identify the source DOAs when

$$K \le 2(N-1)$$

where K and N denote the numbers of sources and sensors, respectively. In addition, the approach can effectively cope with the effects of wide-sense stationary noise with unknown covariance. We have used synthetic signals and real speech signals to perform a number of simulations, where the effectiveness of the KR subspace approach has been successfully demonstrated under various problem settings.

In essence, this paper has revealed how quasi-stationarity can be utilized to provide advantages in DOA estimation. As a future



Fig. 9. DOA spectra of KR-MUSIC and 4-MUSIC in the wideband real-speech case. L = 200, M = 450. The gray dashed lines mark the true DOA positions. (a) KR-MUSIC, with ISSM. (b) 4-MUSIC, with ISSM.

direction, it would be interesting to study how quasi-stationarity may bring a difference to DOA estimation performance limits; e.g., by exploring the Cramér–Rao lower bound in the quasi-stationary case. In addition, it would be interesting to perform a theoretical analysis on the error performance of the KR subspace methods.

#### APPENDIX

A. Proof of Property 1

The matrix product  $ADB^H$  can be written as

$$\mathbf{ADB}^{H} = \sum_{i=1}^{k} d_{i} \mathbf{a}_{i} \mathbf{b}_{i}^{H}.$$
 (33)

Applying vectorization to (33) and using (7), we get

$$\operatorname{ec}(\mathbf{A}\mathbf{D}\mathbf{B}^{H}) = \sum_{i=1}^{k} d_{i}\operatorname{vec}(\mathbf{a}_{i}(\mathbf{b}_{i}^{*})^{T})$$
$$= \sum_{i=1}^{k} d_{i}(\mathbf{b}_{i}^{*} \otimes \mathbf{a}_{i})$$
$$= (\mathbf{B}^{*} \odot \mathbf{A})\mathbf{d}.$$
(34)

Authorized licensed use limited to: National Tsing Hua University. Downloaded on March 22,2010 at 08:29:46 EDT from IEEE Xplore. Restrictions apply.

## B. Proof of Propositions 1 and 2

We describe the analysis that leads to Propositions 1 and 2. In essence, the whole idea lies in studying the following matrix analysis problem: Let

$$\mathbf{V} = \begin{bmatrix} 1 & \cdots & 1\\ z_1^{-1} & \cdots & z_k^{-1}\\ \vdots & \ddots & \vdots\\ z_1^{-(n-1)} & \cdots & z_k^{-(n-1)} \end{bmatrix} \in \mathbb{C}^{n \times k}$$
(35)

be a Vandermonde matrix with roots  $z_1, \ldots, z_k \in \mathbb{C}$ . What is the rank of its self KR product  $\mathbf{V}^* \odot \mathbf{V}$ ? We answer this question by resorting to two distinctively different concepts, namely the KR product properties (summarized in Section III-A) and the virtual array observation [21]–[24]. Interestingly, in the literature we often saw the use of either one of the ideas, but probably not both. First, consider the following lemma.

*Lemma 1:* Consider the matrix V in (35) with  $z_1, \ldots, z_k \in \mathbb{C}, z_k \neq z_\ell$  for all  $k \neq \ell$ . Then we have the following results:

- 1) The KR product  $\mathbf{V}^* \odot \mathbf{V}$  is of full column rank if  $k \leq 2n-1$ .
- 2) For the case where  $|z_1| = \cdots = |z_k| = 1$ , the KR product  $\mathbf{V}^* \odot \mathbf{V}$  is of full column rank only if  $k \leq 2n 1$ .

*Proof of Lemma 1:* We first consider 1), namely the sufficient condition for  $\mathbf{V}^* \odot \mathbf{V}$  to have full column rank. It has been proven in [30] that for a Vandermonde matrix  $\mathbf{V}$ , krank( $\mathbf{V}$ ) = rank( $\mathbf{V}$ ). And for distinct  $z_1, \ldots, z_k$ , the rank of  $\mathbf{V}$  is full [31, pp. 349–350]; i.e., rank( $\mathbf{V}$ ) = min{n, k}. As a consequence, we have

$$\min\{n^2, k\} \ge \operatorname{rank}(\mathbf{V}^* \odot \mathbf{V}) \ge \operatorname{krank}(\mathbf{V}^* \odot \mathbf{V})$$
$$\ge \min\{k, 2\min\{n, k\} - 1\}$$
(36)

where the last inequality is due to Property 2 and krank( $\mathbf{V}^*$ ) = krank( $\mathbf{V}$ ) = min{n, k}. For  $1 \le k \le 2n - 1$ , (36) reduces to  $k \ge \operatorname{rank}(\mathbf{V}^* \odot \mathbf{V}) \ge k$ . This implies that  $\mathbf{V}^* \odot \mathbf{V}$  has full column rank.

Next, we consider 2), the necessary full column rank condition when the roots  $z_k$  have unit magnitude. We notice that for a vector  $\mathbf{v} = [1, z^{-1}, \dots, z^{-(n-1)}]^T \in \mathbb{C}^n, |z| = 1$ ,

$$\mathbf{v}^* \otimes \mathbf{v} = \begin{bmatrix} \mathbf{v} \\ z\mathbf{v} \\ \vdots \\ z^{n-1}\mathbf{v} \end{bmatrix} = \begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ \frac{z^{-(n-1)}}{z} \\ 1 \\ \vdots \\ \frac{z^{-(n-2)}}{\vdots} \\ \vdots \\ 1 \end{bmatrix}.$$
 (37)

This means that  $\mathbf{v}^* \otimes \mathbf{v}$  can be alternatively represented by

$$\mathbf{v}^* \otimes \mathbf{v} = \mathbf{Gb}(z)$$

where

$$\mathbf{G} = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{C}^{n^2 \times (2n-1)}$$

$$(39)$$

and

$$\mathbf{b}(z) = \begin{bmatrix} z^{n-1} & z^{n-2} & \cdots & z & 1 & z^{-1} & \cdots & z^{-(n-1)} \end{bmatrix}^T$$
(40)

Using this representation, we re-express  $\mathbf{V}^* \odot \mathbf{V}$  as

$$\mathbf{V}^* \odot \mathbf{V} = \mathbf{G}\mathbf{B} \tag{41}$$

where  $\mathbf{B} = [\mathbf{b}(z_1), \dots, \mathbf{b}(z_k)] \in \mathbb{C}^{(2n-1)\times k}$ . The matrix  $\mathbf{G}$  can be shown to have orthogonal columns, and thus rank( $\mathbf{G}$ ) = 2n-1. Moreover,  $\mathbf{B}$  has a Vandermonde structure with distinct roots and, therefore, rank( $\mathbf{B}$ ) = min{k, 2n-1}. Applying standard rank results on (41), we obtain

$$\operatorname{rank}(\mathbf{V}^* \odot \mathbf{V}) \le \min\{\operatorname{rank}(\mathbf{G}), \operatorname{rank}(\mathbf{B})\} = \min\{k, 2n-1\}.$$
(42)

Hence, the full column rank condition holds only if  $k \leq 2n-1$ .

With Lemma 1, we are ready to explain how Propositions 1 and 2 are concluded.

*Proof of Propositions 1 and 2:* Proposition 1 is a direct consequence of Lemma 1, considering the unit-magnitude-root Vandermonde structure of the array response matrix **A**. For Proposition 2, we use contradiction. Suppose that there exists an angle  $\varphi \notin \{\theta_1, \ldots, \theta_K\}$  such that the KR subspace criterion (18) is satisfied; i.e.

$$\mathbf{U}_n^H(\mathbf{a}^*(\varphi) \otimes \mathbf{a}(\varphi)) = \mathbf{0}.$$
 (43)

That implies that  $\mathbf{a}^*(\varphi) \otimes \mathbf{a}(\varphi)$  has to be a linear combination of  $\mathbf{a}^*(\theta_1) \otimes \mathbf{a}(\theta_1), \dots, \mathbf{a}^*(\theta_K) \otimes \mathbf{a}(\theta_K)$ ; *viz.* 

$$\mathbf{a}^{*}(\varphi) \otimes \mathbf{a}(\varphi) = \sum_{k=1}^{K} \alpha_{k} \mathbf{a}^{*}(\theta_{k}) \otimes \mathbf{a}(\theta_{k})$$
(44)

for some coefficients  $\alpha_1, \ldots, \alpha_K$ . Equation (44) is equivalent to saying that the matrix

$$\begin{bmatrix} \mathbf{a}^{*}(\theta_{1}) \otimes \mathbf{a}(\theta_{1}), \dots, \mathbf{a}^{*}(\theta_{K}) \otimes \mathbf{a}(\theta_{K}), \mathbf{a}^{*}(\varphi) \otimes \mathbf{a}(\varphi) \end{bmatrix} \\ = \begin{bmatrix} \mathbf{A} \ \mathbf{a}(\varphi) \end{bmatrix}^{*} \odot \begin{bmatrix} \mathbf{A} \ \mathbf{a}(\varphi) \end{bmatrix} \in \mathbb{C}^{N^{2} \times (K+1)}$$
(45)

(38)

has linearly dependent columns. But, by Lemma 1, (45) is linearly independent if and only if  $K+1 \le 2N-1$ . This completes the proof of Proposition 2.

#### ACKNOWLEDGMENT

The authors would like to thank Prof. Kon Max Wong at Mc-Master University, Canada, for his valuable advice that have helped us improve this manuscript significantly. Many thanks also go to the anonymous reviewers, who gave useful and insightful advice that have enabled us to further enhance the paper in many different ways.

#### REFERENCES

- R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [2] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.
- [3] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 823–831, Aug. 1985.
- [4] E. D. Di Claudio and R. Parisi, "WAVES: Weighted average of signal subspaces for robust wideband direction finding," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2179–2191, Oct. 2001.
- [5] Y.-S. Yoon, L. M. Kaplan, and J. H. McClellan, "TOPS: New DOA estimator for wideband signals," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pt. 1, pp. 1977–1989, Jun. 2006.
- [6] C. Kwan, K. C. Ho, G. Mei, Y. Li, Z. Ren, R. Xu, Y. Zhang, D. Lao, M. Stevenson, V. Stanford, and C. Rochet, "An automated acoustic system to monitor and classify birds," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–19, 2006, Article ID 96706.
- [7] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, "Speech enhancement based on the subspace method," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 497–507, Sep. 2000.
- [8] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 320–327, May 2000.
- [9] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sep. 2001.
- [10] K. Rahbar, J. P. Reilly, and J. H. Manton, "Blind identification of MIMO FIR systems driven by quasistationary sources using second-order statistics: A frequency domain approach," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 406–417, Feb. 2004.
- [11] K. Rahbar, J. P. Reilly, and J. H. Manton, "A frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Trans. Signal Process.*, vol. 13, no. 5, pp. 832–844, Sep. 2005.
- [12] R. K. Olsson and L. K. Hansen, "Blind separation of more sources than sensors in convolutive mixtures," in *Proc. ICASSP 2006*, May 2006, pp. V657–V660.
- [13] K. N. Mokios, N. D. Sidiropoulos, and A. Potamianos, "Blind speech separation using parafac analysis and integer least squares," in *Proc. ICASSP 2006*, May 2006, pp. V73–V76.
- [14] K. N. Mokios, A. Potamianos, and N. D. Sidiropoulos, "On the effectiveness of PARAFAC-based estimation for blind speech separation," in *Proc. ICASSP 2008*, Mar. 2008, pp. 153–156.
- [15] R. A. Harshman and M. E. Lundy, "The PARAFAC model for threeway factor analysis and multidimensional scaling," in *Research Methods for Multimode Data Analysis*, H. G. Law, C. W. Snyder, Jr., J. Hattie, and R. P. McDonald, Eds. New York: Praeger, 1984, pp. 122–215.
- [16] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," J. Chemomet., vol. 14, no. 3, pp. 229–239, May 2000.
- [17] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.
- [18] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545–1553, Jul. 2002.

- [19] A. Ziehe, P. Laskov, and G. Nolte, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, vol. 5, pp. 801–818, 2004.
- [20] R. Vollgraf and K. Obermayer, "Quadratic optimization for simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3270–3278, Sep. 2006.
- [21] B. Porat and B. Friedlander, "Direction finding algorithms based on high-order statistics," *IEEE Trans. Signal Process.*, vol. 39, no. 9, pp. 2016–2023, Sep. 1991.
- [22] M. Dogan and J. Mendel, "Applications of cumulants to array processing. I: Aperture extension and array calibration," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1200–1216, May 1995.
- [23] P. Chevalier and A. Ferreol, "On the virtual array concept for the fourth-order direction finding problem," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2592–2595, Sep. 1999.
- [24] P. Chevalier, L. Albera, A. Ferreol, and P. Comon, "On the virtual array concept for higher order array processing," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1254–1271, Apr. 2005.
- [25] C. G. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," *The Indian J. Stat. Series A*, vol. 30, no. 2, pp. 167–180, 1968.
- [26] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics," *Linear Algebra and Its Appl.*, vol. 18, pp. 95–138, 1977.
- [27] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.
- [28] D. Torrieri and K. Bakhru, "The effects of nonuniform and correlated noise on superresolution algorithms," *IEEE Trans. Antennas Propag.*, vol. 45, no. 8, pp. 1214–1218, Aug. 1997.
- [29] H. L. V. Trees, Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing. New York: Wiley, 2002.
- [30] N. D. Sidiropoulos and X. Liu, "Identifiability results for blind beamforming in incoherent multipath with small delay spread," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 228–236, Jan. 2001.
- [31] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Englewood CLiffs, NJ: Pearson Prentice-Hall, 2005.



Wing-Kin Ma (S'96–M'01) received the B.Eng. (first-class honors) degree in electrical and electronic engineering from the University of Portsmouth, Portsmouth, U.K., in 1995 and the M.Phil. and Ph.D. degrees, both in electronic engineering, from the Chinese University of Hong Kong (CUHK), Hong Kong, in 1997 and 2001, respectively. His Ph.D. dissertation was commended to be "of very high quality and well deserved honorary mentioning" by the Faculty of Engineering, CUHK, in 2001.

He is currently an Assistant Professor with the De-

partment of Electronic Engineering, CUHK. From 2005 to 2007, he was also an Assistant Professor with the Institute of Communications Engineering, National Tsing Hua University, Taiwan, R.O.C., where he is still holding an adjunct position. Prior to becoming a faculty member, he held various research positions with McMaster University, Canada; CUHK; and the University of Melbourne, Australia. His research interests are in signal processing and communications, with a recent emphasis on MIMO techniques and convex optimization.

Dr. Ma currently serves as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, and a Guest Editor of IEEE SIGNAL PROCESSING MAGAZINE on the Special Issue titled Convex Optimization for Signal Processing.



**Tsung-Han Hsieh** received the B.S. degree from the Department of Electrical Engineering, National Chi Nan University, Nantou, Taiwan, in 2006, and the M.S. degree from the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan, in 2008.

He is currently working as a Digital IC Design Engineer at Realtek Semiconductor Corp., Hsinchu. His research interests are in signal processing and highspeed interface implementation.



**Chong-Yung Chi** (S'83–M'83–SM'89) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1983.

From 1983 to 1988, he was with the Jet Propulsion Laboratory, Pasadena, CA. He has been a Professor with the Department of Electrical Engineering since 1989 and the Institute of Communications Engineering (ICE) since 1999 (also the Chairman of ICE for 2002–2005), National Tsing Hua University, Hsinchu, Taiwan. He coauthored a technical book Blind Equalization and System Identification (New

York: Springer 2006), and published more than 150 technical (journal and conference) papers. His current research interests include signal processing for wireless communications, convex analysis and optimization for blind source separation, biomedical imaging, and hyperspectral imaging.

Dr. Chi has been a Technical Program Committee member for many IEEE sponsored and cosponsored workshops, symposiums, and conferences on signal processing and wireless communications, including Coorganizer and General Co-Chairman of IEEE SPAWC 2001, and Co-Chair of Signal Processing for Communications (SPC) Symposium, ChinaCOM 2008, and Lead Co-Chair of SPC Symposium, ChinaCOM 2009. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (May 2001-April 2006), the IEEE TRANSACTIONS ON SIGNAL PROCESSING (May 2001-April 2006), the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II (January 2006–December 2007), and a member of Editorial Board of EURASIP Signal Processing Journal (June 2005-May 2008), and an editor (July 2003-December 2005) as well as a Guest Editor (2006) of EURASIP Journal on Applied Signal Processing. Currently, he is an Associate Editor of the IEEE SIGNAL PROCESSING LETERS and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I. He is also a member of the IEEE Signal Processing Committee on Signal Processing Theory and Methods.