Signal Processing 100 (2014) 160–168

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast communication

Underdetermined direction of arrival estimation using acoustic vector sensor

Shengkui Zhao [a,*], Tigran Saluev [a], Douglas L. Jones [b]

[a] Advanced Digital Sciences Center, 1 Fusionopolis Way, #08-10 Connexis North Tower, Singapore 138632, Singapore
[b] Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 1308 West Main Street Urbana, IL 61801, USA

A R T I C L E I N F O

Article history:
Received 30 July 2013
Received in revised form
29 November 2013
Accepted 23 January 2014
Available online 31 January 2014

Keywords:
Microphone arrays
Vector sensors
Direction of arrival estimation
Covariance matrix

A B S T R A C T

This paper presents a new approach for the estimation of two-dimensional (2D) direction-of-arrival (DOA) of more sources than sensors using an Acoustic Vector Sensor (AVS). The approach is developed based on Khatri–Rao (KR) product by exploiting the subspace characteristics of the time variant covariance matrices of the uncorrelated quasi-stationary source signals. An AVS is used to measure both the acoustic pressure and pressure gradients in a complete sound field and the DOAs are determined in both horizontal and vertical planes. The identifiability of the presented KR-AVS approach is studied in both theoretic analysis and computer simulations. Computer simulations demonstrated that 2D DOAs of six speech sources are successfully estimated. Superior root mean square error (RMSE) is obtained using the new KR-AVS array approach compared to the other geometries of the non-uniform linear array, the 2D L-shape array, and the 2D triangular array.

1. Introduction

Two-dimensional (2D) direction of arrival (DOA) estimation with array sensors is essential for source localization in audio surveillance, auditory scene analysis, hearing aids, etc. In these applications, the sources come from not only the horizontal plane but also the vertical plane. In addition, the number of sources can exceed the number of sensors. Using small aperture arrays provides a great convenience in configuration and portability too. Therefore, the DOA estimation in 2D space using small aperture arrays is highly desirable. The conventional linear array approaches [1–4] are only able to estimate the DOAs in the horizontal plane. In addition, they have the front and back ambiguous problem. Therefore, they are less efficient for the situations with sources located at different heights. The four-element Acoustic Vector Sensor (AVS) first presented in [5] for DOA estimation is an acoustic sensor that is capable of measuring acoustic pressure gradient as well as pressure as in a standard microphone. This combination makes it possible for the sensor to measure the complete sound field.

The AVS has been studied for overdetermined DOA estimation where the number of sources is less than the number of sensors [5–8]. In this case, the 2D DOA can be estimated by employing the subspace approach such as the MUSIC (MUltiple SIgnal Classification) [10,11]. By applying eigenvector decomposition to the local covariance matrix, the source subspace and noise subspace are identified based on their eigenvalues, and then the DOAs are estimated based on searching the steering vectors orthogonal to the noise subspace. However, when the number of sources is equal to or more than the number of senors, the noise subspace cannot be identified using the MUSIC approach on the local covariance matrix. To localize more speech sources, the time–frequency spareness was exploited to find low-rank covariance matrices in [3]. However, it proves difficult to

* Corresponding author.
E-mail addresses: shengkui.zhao@adsc.com.sg (S. Zhao), tigran.saluev@gmail.com (T. Saluev), jones@ifp.illinois.edu (D.L. Jones).

0165-1684/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2014.01.021

identify low rank covariance matrices when ambient noise is high.

In this work, we are interested to study the underdetermined 2D DOA estimation problem where the number of sources is more than the number of sensors using an AVS. We would like to deal with the sources of audio or speech signals which are so-called quasi-stationary signals. It is known that for such quasi-stationary signals the covariance matrix of the array output is locally static over a short period of time and exhibits difference over a long time period. Using the subspace characteristics of the time-variant covariance matrices of the quasi-stationary source signals, the underdetermined DOA estimation problem can be transformed to an overdetermined DOA estimation problem. Therefore, we develop a subspace approach based on Khatri–Rao product [4]. By applying the vectorization to local covariance matrices and stacking the resulting vectors into a virtual matrix, the degree of freedom is increased by square of its original value. Then we apply a detection criterion similar to the MUSIC on the virtual matrix and estimate the azimuth and elevation angles. Our identifiable analysis and simulations show that the 2D DOAs of six speech sources are successfully estimated using an AVS.

2. Problem formulation

We consider a four-element AVS [5] and K wideband sources impinging on the array from far field with azimuth angles of $\theta_k \in (-180°, 180°]$ and elevation angles of $\phi_k \in [-90°, 90°]$, $k = 1, \ldots, K(K \geq 4)$. The output signals of the sensors are modeled in time–frequency domain as

$$\mathbf{x}(t,f) = A\mathbf{s}(t,f) + \mathbf{v}(t,f), \quad t = 0, 1, 2, \ldots \tag{1}$$

where t is the time index and f is the frequency index. Here, $\mathbf{x}(t,f) = [x_1(t,f), \ldots, x_4(t,f)]^T$ is the received signal vector, $\mathbf{s}(t,f) = [s_1(t,f), \ldots, s_K(t,f)]^T$ is the source vector, $\mathbf{v}(t,f) \in \mathbb{C}^4$ represents the spatial noise vector. For convenience we omit the frequency index f during the derivation in the rest of the paper. The matrix $A = [\mathbf{a}(\vec{r}_1), \ldots, \mathbf{a}(\vec{r}_K)] \in \mathbb{R}^{4\times K}$ is the array response matrix, and $\mathbf{a}(\vec{r}_k)$ is the 4×1 AVS array manifold for source k [5]:

$$\mathbf{a}(\vec{r_k}) \triangleq [1, \cos\theta_k \cos\phi_k, \sin\theta_k \cos\phi_k, \sin\phi_k]^T, \tag{2}$$

where the vector $\vec{r} = [\cos\theta_k \cos\phi_k, \sin\theta_k \cos\phi_k, \sin\phi_k]^T$ is the unit source bearing vector where the azimuth and elevation angles are defined as $\theta_k \in [-180°, 180°]$ and $\phi_k \in [-90°, 90°]$. Note that only the bearing angle θ_k is considered for a linear array studied in [4,12]. Next, we will derive the approach for the 2D DOA estimation with both θ_k and ϕ_k.

When the source signals $s_k(t)$ and noise signals $\mathbf{v}(t)$ are assumed mutually uncorrelated, a local covariance matrix can be defined as

$$\mathbf{R}_m = E\{\mathbf{x}(t)\mathbf{x}^H(t)\} = \mathbf{A}\mathbf{D}_m\mathbf{A}^T + \mathbf{C}, \tag{3}$$

for $\forall t \in [(m-1)L, mL-1]$, where $m = 1, 2, \ldots$ denotes the frame index, L is the frame size, and $\mathbf{D}_m = \mathrm{Diag}(d_{m1}, d_{m2}, \ldots, d_{mK}) \in \mathbb{R}^{K\times K}$ is the source covariance matrix at frame m. These local covariance matrices may be estimated by local averaging. To estimate the DOAs $\theta_1, \ldots, \theta_K$, the conventional MUSIC criterion is based on a single instance of the local covariance matrix $\mathbf{R}_m \in \mathbb{C}^4$. Since the degree of freedom for $\mathbf{R}_m$ is equal to 4, it is insufficient to identify $K \geq 4$ sources.

3. The proposed KR-AVS approach

In this section, we present the approach using the Khatri–Rao product and AVS (KR-AVS) for 2D DOA estimation of $K \geq 4$ sources. The approach will transform the above underdetermined DOA estimation problem into an overdetermined problem.

3.1. The 2D KR-AVS criterion

Let us apply the vectorization computation to the covariance matrix $\mathbf{R}_m$ to obtain

$$\begin{aligned}\mathbf{y}_m &\triangleq \mathrm{vec}(\mathbf{R}_m) = \mathrm{vec}(\mathbf{A}\mathbf{D}_m\mathbf{A}^T) + \mathrm{vec}(\mathbf{C}) \\ &= (\mathbf{A} \odot \mathbf{A})\mathbf{d}_m + \mathrm{vec}(\mathbf{C})\end{aligned} \tag{4}$$

where $\mathrm{vec}(\cdot)$ stands for vectorization computation; i.e., $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$ then $\mathrm{vec}(\mathbf{V}) = [\mathbf{v}_1^T, \mathbf{v}_2^T, \ldots, \mathbf{v}_n^T]^T$, and the symbol $\odot$ stands for the Khatri–Rao (KR) product: $\mathbf{A} \odot \mathbf{A} = [\mathbf{a}(\vec{r}_1) \otimes \mathbf{a}(\vec{r}_1), \ldots, \mathbf{a}(\vec{r}_K) \otimes \mathbf{a}(\vec{r}_K)] \in \mathbb{R}^{16\times K}$, and $\otimes$ denotes the Kronecker product, the vector is $\mathbf{d}_m = [d_{m1}, \ldots, d_{mK}]^T$.

Now we can see that the expression of (4) has a similar structure as the signal model in (1). The KR product $(\mathbf{A} \odot \mathbf{A})$ can be considered as the transformed array response matrix, which has virtual array dimension 16 much greater than the physical dimension of 4.

Now consider we have the local covariance matrices $\mathbf{R}_1, \ldots, \mathbf{R}_M$, we stack their vectorization vectors to obtain

$$\mathbf{Y} \triangleq [\mathbf{y}_1, \ldots, \mathbf{y}_M] = (\mathbf{A} \odot \mathbf{A})\mathbf{\Psi}^T + \mathrm{vec}(\mathbf{C})\mathbf{1}_M^T, \tag{5}$$

where the matrix is defined as $\mathbf{\Psi} = [\mathbf{d}_1, \ldots, \mathbf{d}_M]^T \in \mathbb{R}^{M\times K}$ and $\mathbf{1}_M = [1, \ldots, 1]^T \in \mathbb{R}^M$. For the quasi-stationary source signals and a large number of frames $M \gg K$, the matrix $[\mathbf{\Psi}\ \mathbf{1}_M] \in \mathbb{R}^{M\times(K+1)}$ can be safely assumed as a full column rank. That is, there exists a collection of K linearly independent columns in $\mathbf{\Psi}$. In a real world, most of the audio and speech signals whose power spectrums are not flat can satisfy this assumption.

Therefore, the noise covariance term $\mathrm{vec}(\mathbf{C})\mathbf{1}_M^T$ in (5) has identical columns and can be eliminated by applying the orthogonal component projection matrix $\mathbf{P}_{\mathbf{1}_M}^{\perp} = \mathbf{I}_M - (1/M)\mathbf{1}_M\mathbf{1}_M^T$ to (5). Then we have the following decomposition form:

$$\mathbf{Y}\mathbf{P}_{\mathbf{1}_M}^{\perp} = (\mathbf{A} \odot \mathbf{A})(\mathbf{P}_{\mathbf{1}_M}^{\perp}\mathbf{\Psi})^T \tag{6}$$

Noted that when a subset $J(J < K)$ of the K sources is stationary, the matrix $[\mathbf{\Psi}\ \mathbf{1}_M] \in \mathbb{R}^{M\times(K+1)}$ cannot be assumed as a full column rank. There exists a collection of $K-J$ linearly independent columns in $\mathbf{\Psi}$. Together with the stationary noise, the J stationary sources are eliminated by the orthogonal component project matrix $\mathbf{P}_{\mathbf{1}_M}^{\perp}$. Therefore, only the $K-J$ sources are to be considered in the DOA estimation process.

Now consider the subspace of $\mathbf{Y}\mathbf{P}_{\mathbf{1}_M}^{\perp}$. For ease of exposition of idea, we assume that the decomposition (6) is unique. We will soon provide the conditions under which