Kronecker Product Adaptive Beamforming for Microphone Arrays

Xuehan Wang*†, Gongping Huang†, Israel Cohen†, Jacob Benesty‡, and Jingdong Chen*
* CIAIC and Shaanxi Provincial Key Laboratory of Artificial Intelligence,
Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China
wangxuehan.123@mail.nwpu.edu.cn, jingdongchen@ieee.org
† Faculty of Electrical and Computer Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel
gongpinghuang@gmail.com, icohen@ee.technion.ac.il
‡ INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Montreal, QC H5A 1K6, Canada,
benesty@emt.inrs.ca

***Abstract*—Microphone array adaptive beamforming is needed in a wide range of audio and speech applications for acquiring the acoustic signal of interest while suppressing noise and interference. In the design and application of adaptive beamformers, particular attention has to be paid to the issues of robustness and computational efficiency. One way to deal with these issues is through the use of the recently developed Kronecker product beamforming framework. However, the existing Kronecker product beamformers were formulated based on special array geometries that can be straightforwardly decomposed into subarrays and, as a result, the application of such formulation is limited to a small range of arrays. To generalize the formulation, we introduce in this paper a framework that can be applied to arbitrary array geometries, where the beamformer is represented as a sum of Kronecker products of several subfilters. Based on this new framework, an iterative optimization algorithm is derived for designing the subfilters of the minimum variance distortionless response (MVDR) beamformer. Simulation results demonstrate the advantages of the proposed method in different conditions.**

***Index Terms*—Microphone arrays, adaptive beamforming, minimum variance distortionless response (MVDR), Kronecker product beamforming.**

I. Introduction

Microphone arrays equipped with beamforming techniques play an important role in a wide range of applications such as teleconferencing and smart home systems [1]–[4]. Generally, beamforming methods can be divided into two categories, i.e., fixed [5]–[8] and adaptive beamformers [9], depending on how the beamforming filters are determined. In comparison, adaptive beamformers are more effective than the fixed ones in dynamic and time-varying acoustic environments as their beamforming filters are updated according to the statistics of the incoming data; but they are also less robust and can introduce signal distortion and signal self cancelation as has been widely discussed with the minimum variance distortionless response (MVDR) beamformer [10]–[13].

Recently, the principle of linear filtering with Kronecker product [14], [15] has been applied in multiple fields, such as system identification [16], noise reduction [17], dereverberation [18], and time-difference-of-arrival (TDOA) estimation [19]. It was also investigated in the context of beamforming [20]–[25]. The basic idea underlying Kronecker product beamforming is to partition the given microphone array into two or multiple smaller virtual subarrays so the (global) beamforming filter associated with the entire array can be expressed as the Kronecker product of subfilters, each corresponding to one subarray. By properly designing those subfilters and then combining them in the framework of Kronecker product, the global beamformer can be flexibly optimized to achieve different properties, e.g., high spatial gain, high output signal-to-noise ratio (SNR), and/or good robustness. This leads to a new way to deal with some issues with the traditional adaptive beamformers.

However, the existing Kronecker product beamformers were formulated based on special array geometries, e.g., linear, rectangle, and cubic ones, which can be straightforwardly decomposed into subarrays. A legitimate question is then how to generalize the existing framework so that it can work with any (arbitrary) array geometry, which is the objective of this work. We propose a framework, which reformulates the beamforming filter as a sum of Kronecker products of several shorter subfilters as developed in [16] in the context of system identification. In comparison with the existing one, the proposed framework explicitly eliminates the restriction on the array geometry and, as a result, it can be applied more generally to any geometry in the three-dimensional space as long as the positions of the microphone sensors are given. Based on this framework, we derive an iterative optimization algorithm for designing the subfilters associated with the MVDR beamformer, resulting a Kronecker MVDR (KMVDR) adaptive beamformer. Simulations are presented to illustrate some properties of this new beamformer.

II. Signal Model and Problem Formulation

We consider a microphone array with an arbitrary geometry in three-dimensional space that includes $M = M_1 M_2$ acoustic sensors as illustrated in Fig. 1, where M_1, M_2 are arbitrary positive integer numbers and the position of the mth microphone in the Cartesian coordinate system is denoted by $\mathbf{p}_m = [x_m \; y_m \; z_m]^T$, $m = 1, 2, \ldots, M$. In an anechoic environment, we assume that a far-field source (plane wave)

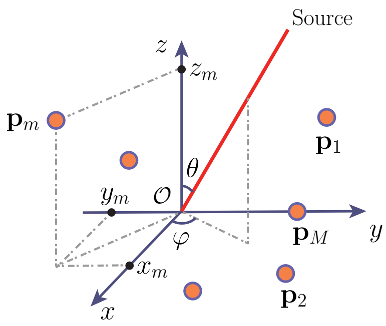

Fig. 1. Illustration of a microphone array with an arbitrary geometry in three-dimensional space.

is incident to the array from the direction $\boldsymbol{\rho}_{\rm s} = (\theta_{\rm s}, \varphi_{\rm s})$ at the speed of sound, i.e., $c = 340$ m/s, with $\theta_{\rm s}$ and $\varphi_{\rm s}$ specifying the elevation and azimuth angles, respectively. In case that an interference (coming from a point source) and ambient noise co-exist in the environment, the array observation can be represented in the frequency domain as

$$\begin{aligned} \mathbf{y}(\omega) &= \begin{bmatrix} Y_1(\omega) & Y_2(\omega) & \cdots & Y_M(\omega) \end{bmatrix}^T \\ &= \mathbf{x}(\omega) + \mathbf{u}(\omega) + \mathbf{v}(\omega) \\ &= \mathbf{d}(\omega, \boldsymbol{\rho}_{\rm s}) X(\omega) + \mathbf{u}(\omega) + \mathbf{v}(\omega), \end{aligned} \tag{1}$$

where $\omega = 2\pi f$ is the angular frequency with f being the temporary frequency, $Y_m(\omega)$ is the signal received at the mth microphone, the superscript ^{T} is the transpose operator, $\mathbf{x}(\omega)$, $\mathbf{u}(\omega)$, and $\mathbf{v}(\omega)$ are the desired, interference, and noise vectors, respectively, having a similar form as $\mathbf{y}(\omega)$, $X(\omega)$ is the desired signal, and

$$\mathbf{d}(\omega, \boldsymbol{\rho}_{\rm s}) = \begin{bmatrix} e^{-\jmath \mathbf{k}_{\rm s}^T \mathbf{p}_1} & e^{-\jmath \mathbf{k}_{\rm s}^T \mathbf{p}_2} & \ldots & e^{-\jmath \mathbf{k}_{\rm s}^T \mathbf{p}_M} \end{bmatrix}^T \tag{2}$$

is the phase vector corresponding to the desired direction, with $\mathbf{k}_{\rm s} = -\frac{\omega}{c} \begin{bmatrix} \sin\theta_{\rm s}\cos\varphi_{\rm s} & \sin\theta_{\rm s}\sin\varphi_{\rm s} & \cos\theta_{\rm s} \end{bmatrix}^T$ being the wavenumber and $\jmath$ being the imaginary unit. In the rest of the paper, we drop the variable ω from the notation for the sake of simplification. The covariance matrix of $\mathbf{y}$ can be written as

$$\boldsymbol{\Phi}_{\mathbf{y}} = E\left(\mathbf{y}\mathbf{y}^H\right), \tag{3}$$

where $E(\cdot)$ denotes the mathematical expectation and the superscript ^{H} is the conjugate-transpose operator. The objective of beamforming is to recover the desired signal, which is achieved by applying a spatial filter to the noisy array observations, i.e.,

$$Z = \mathbf{h}^H \mathbf{y}, \tag{4}$$

where $\mathbf{h}$ is a spatial (or beamforming) filter of length M. Then, the variance of the beamformer's output can be deduced as

$$\phi_Z = E\left(|Z|^2\right) = \mathbf{h}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{h}. \tag{5}$$

In order to preserve the desired signal, the distortionless constraint is always desired, i.e.,

$$\mathbf{h}^H \mathbf{d}(\boldsymbol{\rho}_{\rm s}) = 1. \tag{6}$$

Minimizing the variance of the beamformer's output under the distortionless constraint gives the well-known MVDR beamformer:

$$\mathbf{h}_{\rm MVDR} = \frac{\boldsymbol{\Phi}_{\mathbf{y}}^{-1}\mathbf{d}(\boldsymbol{\rho}_{\rm s})}{\mathbf{d}^H(\boldsymbol{\rho}_{\rm s})\boldsymbol{\Phi}_{\mathbf{y}}^{-1}\mathbf{d}(\boldsymbol{\rho}_{\rm s})}. \tag{7}$$

III. Kronecker MVDR Beamformer

In our study, we propose to rewrite the beamforming filter as the sum of Kronecker products of subfilters [16], i.e.,

$$\mathbf{h} = \sum_{p=1}^{P} \mathbf{h}_{1,p} \otimes \mathbf{h}_{2,p}, \tag{8}$$

where $\mathbf{h}_{1,p}$ and $\mathbf{h}_{2,p}, p = 1, 2, \ldots, P$ are subfilters of length M_1 and M_2, respectively, and P is an integer ranging from 1 to $\min\{M_1, M_2\}$.

We have the following obvious relationships:

$$\begin{aligned} \mathbf{h}_{1,p} \otimes \mathbf{h}_{2,p} &= \left(\mathbf{h}_{1,p} \otimes \mathbf{I}_{M_2}\right)\mathbf{h}_{2,p} \\ &= \left(\mathbf{I}_{M_1} \otimes \mathbf{h}_{2,p}\right)\mathbf{h}_{1,p}, \end{aligned} \tag{9}$$

where $\mathbf{I}_{M_1}$ and $\mathbf{I}_{M_2}$ are the identity matrices of sizes $M_1 \times M_1$ and $M_2 \times M_2$, respectively. Using (9) and (8), we obtain

$$\mathbf{h} = \sum_{p=1}^{P} \mathbf{H}_{1,p}\mathbf{h}_{2,p} = \sum_{p=1}^{P} \mathbf{H}_{2,p}\mathbf{h}_{1,p}, \tag{10}$$

where

$$\mathbf{H}_{1,p} = \mathbf{h}_{1,p} \otimes \mathbf{I}_{M_2}, \tag{11}$$

$$\mathbf{H}_{2,p} = \mathbf{I}_{M_1} \otimes \mathbf{h}_{2,p} \tag{12}$$

are matrices of sizes $M \times M_2$ and $M \times M_1$, respectively.

When $\mathbf{h}_{1,p}, p = 1, 2, \ldots, P$ are fixed, by substituting (10) into (4) we can rewrite the estimated signal as

$$\begin{aligned} Z &= \sum_{p=1}^{P} \mathbf{h}_{2,p}^H \mathbf{H}_{1,p}^H \mathbf{y} \\ &= \sum_{p=1}^{P} \mathbf{h}_{2,p}^H \mathbf{y}_{1,p} \\ &= \underline{\mathbf{h}}_2^H \underline{\mathbf{y}}_1, \end{aligned} \tag{13}$$

where

$$\begin{aligned} \mathbf{y}_{1,p} &= \mathbf{H}_{1,p}^H \mathbf{y}, \; p = 1, 2, \ldots, P, \\ \underline{\mathbf{h}}_2 &= \begin{bmatrix} \mathbf{h}_{2,1}^T & \mathbf{h}_{2,2}^T & \cdots & \mathbf{h}_{2,P}^T \end{bmatrix}^T, \\ \underline{\mathbf{y}}_1 &= \begin{bmatrix} \mathbf{y}_{1,1}^T & \mathbf{y}_{1,2}^T & \cdots & \mathbf{y}_{1,P}^T \end{bmatrix}^T. \end{aligned}$$

Then, the variance of the beamformer's output can be written as

$$\phi_Z\left(\underline{\mathbf{h}}_2|\underline{\mathbf{h}}_1\right) = \underline{\mathbf{h}}_2^H \boldsymbol{\Phi}_{\underline{\mathbf{y}}_1} \underline{\mathbf{h}}_2, \tag{14}$$

where

$$\boldsymbol{\Phi}_{\underline{\mathbf{y}}_1} = \begin{bmatrix} \mathbf{H}_{1,1}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{1,1} & \mathbf{H}_{1,1}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{1,2} & \cdots & \mathbf{H}_{1,1}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{1,P} \\ \mathbf{H}_{1,2}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{1,1} & \mathbf{H}_{1,2}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{1,2} & \cdots & \mathbf{H}_{1,2}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{1,P} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{1,P}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{1,1} & \mathbf{H}_{1,P}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{1,2} & \cdots & \mathbf{H}_{1,P}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{1,P} \end{bmatrix} \tag{15}$$

is a matrix of size $PM_2 \times PM_2$. Using (10) and (6) lead to the distortionless constraint for $\underline{\mathbf{h}}_2$, i.e.,

$$\begin{aligned} \mathbf{h}^H \mathbf{d}(\boldsymbol{\rho}_\mathrm{s}) &= \sum_{p=1}^{P} \mathbf{h}_{2,p}^H \mathbf{H}_{1,p}^H \mathbf{d}(\boldsymbol{\rho}_\mathrm{s}) \\ &= \sum_{p=1}^{P} \mathbf{h}_{2,p}^H \mathbf{d}_{1,p}(\boldsymbol{\rho}_\mathrm{s}) \\ &= \underline{\mathbf{h}}_2^H \underline{\mathbf{d}}_1(\boldsymbol{\rho}_\mathrm{s}) = 1, \end{aligned} \tag{16}$$

where

$$\begin{aligned} \mathbf{d}_{1,p}(\boldsymbol{\rho}_\mathrm{s}) &= \mathbf{H}_{1,p}^H \mathbf{d}(\boldsymbol{\rho}_\mathrm{s}), \ p = 1, 2, \ldots, P, \\ \underline{\mathbf{d}}_1(\boldsymbol{\rho}_\mathrm{s}) &= \begin{bmatrix} \mathbf{d}_{1,1}^T(\boldsymbol{\rho}_\mathrm{s}) & \mathbf{d}_{1,2}^T(\boldsymbol{\rho}_\mathrm{s}) & \cdots & \mathbf{d}_{1,P}^T(\boldsymbol{\rho}_\mathrm{s}) \end{bmatrix}^T. \end{aligned} \tag{17}$$

By minimizing $\phi_Z(\underline{\mathbf{h}}_2 | \underline{\mathbf{h}}_1)$ under the distortionless constraint for $\underline{\mathbf{h}}_2$, we can derive

$$\underline{\mathbf{h}}_2 = \frac{\boldsymbol{\Phi}_{\underline{\mathbf{y}}_1}^{-1} \underline{\mathbf{d}}_1(\boldsymbol{\rho}_\mathrm{s})}{\underline{\mathbf{d}}_1^H(\boldsymbol{\rho}_\mathrm{s}) \boldsymbol{\Phi}_{\underline{\mathbf{y}}_1}^{-1} \underline{\mathbf{d}}_1(\boldsymbol{\rho}_\mathrm{s})}. \tag{18}$$

Alternatively, when $\mathbf{h}_{2,p}, p = 1, 2, \ldots, P$ are fixed, the beamformer's output can be written as

$$\begin{aligned} Z &= \sum_{p=1}^{P} \mathbf{h}_{1,p}^H \mathbf{H}_{2,p}^H \mathbf{y} \\ &= \sum_{p=1}^{P} \mathbf{h}_{1,p}^H \mathbf{y}_{2,p} \\ &= \underline{\mathbf{h}}_1^H \underline{\mathbf{y}}_2, \end{aligned} \tag{19}$$

where

$$\begin{aligned} \mathbf{y}_{2,p} &= \mathbf{H}_{2,p}^H \mathbf{y}, \ p = 1, 2, \ldots, P, \\ \underline{\mathbf{h}}_1 &= \begin{bmatrix} \mathbf{h}_{1,1}^T & \mathbf{h}_{1,2}^T & \cdots & \mathbf{h}_{1,P}^T \end{bmatrix}^T, \\ \underline{\mathbf{y}}_2 &= \begin{bmatrix} \mathbf{y}_{2,1}^T & \mathbf{y}_{2,2}^T & \cdots & \mathbf{y}_{2,P}^T \end{bmatrix}^T. \end{aligned}$$

In this case, the variance of the beamformer's output is

$$\phi_Z(\underline{\mathbf{h}}_1 | \underline{\mathbf{h}}_2) = \underline{\mathbf{h}}_1^H \boldsymbol{\Phi}_{\underline{\mathbf{y}}_2} \underline{\mathbf{h}}_1, \tag{20}$$

where

$$\boldsymbol{\Phi}_{\underline{\mathbf{y}}_2} = \begin{bmatrix} \mathbf{H}_{2,1}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{2,1} & \mathbf{H}_{2,1}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{2,2} & \cdots & \mathbf{H}_{2,1}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{2,P} \\ \mathbf{H}_{2,2}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{2,1} & \mathbf{H}_{2,2}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{2,2} & \cdots & \mathbf{H}_{2,2}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{2,P} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{2,P}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{2,1} & \mathbf{H}_{2,P}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{2,2} & \cdots & \mathbf{H}_{2,P}^H \boldsymbol{\Phi}_{\mathbf{y}} \mathbf{H}_{2,P} \end{bmatrix} \tag{21}$$

Algorithm 1 KMVDR beamformer.

```
Estimate Φ_y
Initialization: h_{2,p}, p = 1, 2, ..., P
repeat
    Compute H_{2,p} using (12)
    Compute Φ_{y_2} and d_2(ρ_s) using (21) and (23)
    Update h_1 using (24)
    Compute H_{1,p} using (11)
    Compute Φ_{y_1} and d_1(ρ_s) using (15) and (17)
    Update h_2 using (18)
until iteration stops
return h_KMVDR = Σ_{p=1}^{P} h_{1,p} ⊗ h_{2,p}
```

is a matrix of size $PM_1 \times PM_1$. The distortionless constraint for $\underline{\mathbf{h}}_1$ is derived as

$$\begin{aligned} \mathbf{h}^H \mathbf{d}(\boldsymbol{\rho}_\mathrm{s}) &= \sum_{p=1}^{P} \mathbf{h}_{1,p}^H \mathbf{H}_{2,p}^H \mathbf{d}(\boldsymbol{\rho}_\mathrm{s}) \\ &= \sum_{p=1}^{P} \mathbf{h}_{1,p}^H \mathbf{d}_{2,p}(\boldsymbol{\rho}_\mathrm{s}) \\ &= \underline{\mathbf{h}}_1^H \underline{\mathbf{d}}_2(\boldsymbol{\rho}_\mathrm{s}) = 1, \end{aligned} \tag{22}$$

where

$$\begin{aligned} \mathbf{d}_{2,p}(\boldsymbol{\rho}_\mathrm{s}) &= \mathbf{H}_{2,p}^H \mathbf{d}(\boldsymbol{\rho}_\mathrm{s}), \ p = 1, 2, \ldots, P, \\ \underline{\mathbf{d}}_2(\boldsymbol{\rho}_\mathrm{s}) &= \begin{bmatrix} \mathbf{d}_{2,1}^T(\boldsymbol{\rho}_\mathrm{s}) & \mathbf{d}_{2,2}^T(\boldsymbol{\rho}_\mathrm{s}) & \cdots & \mathbf{d}_{2,P}^T(\boldsymbol{\rho}_\mathrm{s}) \end{bmatrix}^T. \end{aligned} \tag{23}$$

Similarly, minimizing $\phi_Z(\underline{\mathbf{h}}_1 | \underline{\mathbf{h}}_2)$ under the distortionless constraint for $\underline{\mathbf{h}}_1$ yields

$$\underline{\mathbf{h}}_1 = \frac{\boldsymbol{\Phi}_{\underline{\mathbf{y}}_2}^{-1} \underline{\mathbf{d}}_2(\boldsymbol{\rho}_\mathrm{s})}{\underline{\mathbf{d}}_2^H(\boldsymbol{\rho}_\mathrm{s}) \boldsymbol{\Phi}_{\underline{\mathbf{y}}_2}^{-1} \underline{\mathbf{d}}_2(\boldsymbol{\rho}_\mathrm{s})}. \tag{24}$$

The Kronecker MVDR (KMVDR) beamformer is obtained in an iterative manner, as summarized in Algorithm 1, where we first fix (initialize) $\underline{\mathbf{h}}_1$ to compute $\underline{\mathbf{h}}_2$, and then fix $\underline{\mathbf{h}}_2$ to compute $\underline{\mathbf{h}}_1$. The iteration stops if the performance of the proposed algorithm is no longer improved or the beamformer does no longer change as the iteration continues. And the final KMVDR beamformer is given by

$$\mathbf{h}_{\mathrm{KMVDR}} = \sum_{p=1}^{P} \mathbf{h}_{1,p} \otimes \mathbf{h}_{2,p}. \tag{25}$$

IV. Simulation results

In this section, we study the performance of the proposed beamforming method in simulated reverberant environments. We consider a uniform linear microphone array consisting of 16 microphones with an interelement spacing of $\delta = 3$ cm in a room of size $10 \times 8 \times 4$ m, as shown in Fig. 2, where the microphones are located, respectively, at $(x, 4, 2)$ with $x = 5 : 0.03 : 5.45$. The room impulse responses

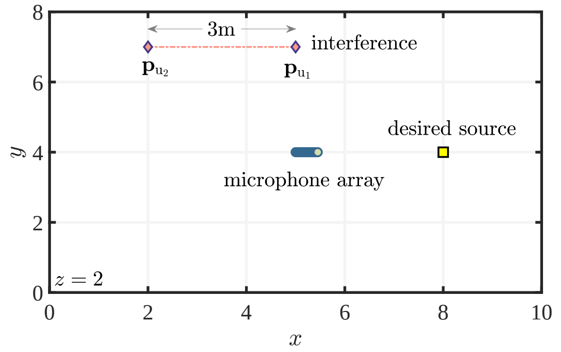

Fig. 2. Layout of the simulation setup (coordinate values are in meters). The 16 microphones of the linear array are located, respectively, at $(x, 4, 2)$, where $x = 5 : 0.03 : 5.45$. The desired source is located at $(8, 4, 2)$. The interference source is located at $(5, 7, 2)$ (static source) or moving from $(5, 7, 2)$ to $(2, 7, 2)$ along a linear path (dynamic source).

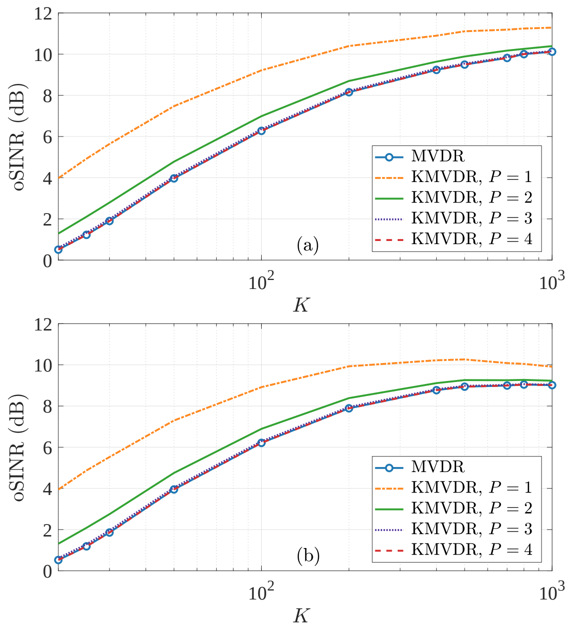

Fig. 3. Output SINR of the MVDR and KMVDR beamformers as a function of the number of snapshots, K: (a) static interference source and (b) dynamic interference source. Conditions of simulation: $P \in \{1, 2, 3, 4\}$, $n = 20$, and with a stationary desired source signal.

(RIRs) from the source to each microphone are generated using the image model method (the reverberation time, T_{60}, is approximately 90 ms) and the microphone observations are generated by convolving the RIRs with the clean source signal, and then adding interference and white Gaussian noise. So, the received signal at each microphone consists of three parts: desired signal, interference signal, and Gaussian white noise. In our simulation, the input signal-to-noise ratio (iSNR) is controlled to be 20 dB. For the KMVDR beamformer, we set $M_1 = M_2 = 4$, the number of iteration, $n = 20$, and initialize $\mathbf{h}_{2,p} = \mathbf{e}_p$, $p = 1, 2, \ldots, P$, where $\mathbf{e}_p$ is a unit vector corresponding to the pth column of the identity matrix $\mathbf{I}_{M_2}$.

We consider two kinds of interference signals. 1) Static interference, where the source is located at the position $(5, 7, 2)$. 2) Dynamic interference, where the source moves from $(5, 7, 2)$ to $(2, 7, 2)$ along a linear path (length of 3 m) with a constant speed of 20 cm/s. Along the path we choose 3000 positions uniformly distributed with an interval of 1 mm and for each position an interference signal segment of 5 ms is generated by convolving the source signal with the corresponding RIR. This is equivalent to producing the interference signal by convolving the source signal with the RIR changing every 5 ms. The interference source signal is a white Gaussian process. The desired source is located at $(8, 4, 2)$. We also consider two kinds of source signals. 1) Stationary signal, where the source signal is an AR process produced by filtering a white Gaussian process through a first order system $1/\left(1 - 0.9z^{-1}\right)$ and the covariance matrix is estimated by averaging the covariance matrix of K available snapshots. In this case, the input signal-to-interference ratio (iSIR) is fixed to 0 dB. 2) Clean speech with a sampling rate of 48 kHz, where an entropy based voice activity detection technique [28]–[30] is used to identify the unvoiced snapshots and then the covariance matrix is estimated with a recursive method as in [31]. In this case, we set iSIR $\in \{-10 : 5 : 15\}$ dB. The array observations are truncated into overlapping frames with a frame length of 256 points and an overlapping ratio of 75%. Each frame is then transformed into the short-time-Fourier-transform (STFT) domain with a Kaiser window. The beamformer is designed and applied in each subband and the output signal in the time domain is obtained with the inverse STFT.

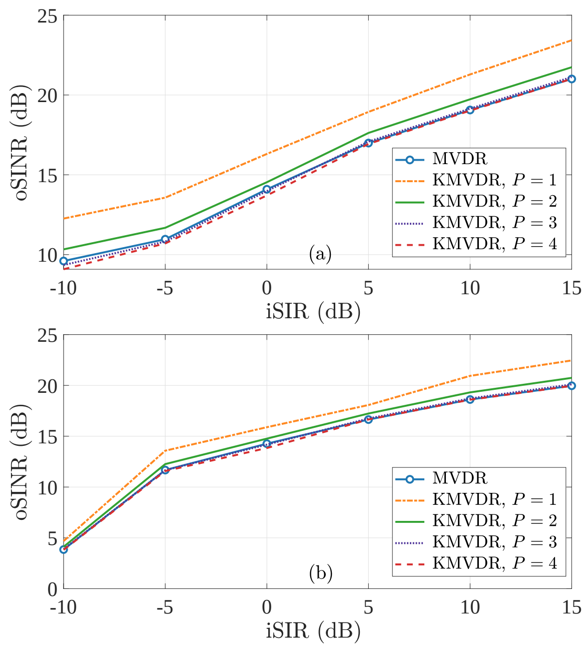

Fig. 4. Output SINR of the MVDR and KMVDR beamformers as a function of the input SIR: (a) static interference source and (b) dynamic interference source. Conditions of simulation: $P \in \{1, 2, 3, 4\}$, $n = 20$, and with a speech source signal.

We first study the performance of the MVDR and KMVDR beamformers, with $P \in \{1, 2, 3, 4\}$ and different numbers of available snapshots, K, for the stationary source case. The results are plotted in Fig. 3, which shows the output signal-to-interference-plus-noise ratio (oSINR) for both the static and dynamic interference cases. It is clearly observed that the oSINR for all the studied beamformers increase as the value of K increases. This is reasonable since the covariance matrix can be estimated more reliably with a larger number of snapshots. For the dynamic interference source, the oSINR is always lower than that obtained with the static interference source, and when the oSINR reaches up to a certain value, a further increase of the value of K does no longer improve the performance. In comparison, the KMVDR beamformers produce a higher level of oSINR for both cases. Figure 4 shows plots of the oSINR of the MVDR and KMVDR beamformers (the source is speech) with different iSIR in the static and dynamic interference cases. Again, the KMVDR beamformers achieve a better performance than MVDR in both cases. The best performance is obtained with the KMVDR beamformer when $P = 1$.

V. Conclusions

In this paper, we introduced a framework, which reformulates the beamformer as a sum of Kronecker products of shorter subfilters. We showed that the Kronecker product can be applied to derive the MVDR beamformer for arbitrary three-dimensional microphone arrays. Alternately, fixing each one of the subfilters, we obtain explicit expressions for the output variance and the distortionless constraint of the other subfilter. By iteratively minimizing the deduced output variance under the distortionless constraint, we derived the KMVDR beamformer. We studied the proposed KMVDR beamformer in different conditions using simulations, where limited observation data and dynamic interferences were considered. Experimental results showed that the proposed KMVDR beamformer can obtain better performance than the conventional one in terms of oSINR.

Acknowledgment

This work was supported by the National Key Research and Development Program of China under Grant No. 2018AAA0102200 and the Key Program of NSFC Grant 61831019.

[1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

[2] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing*. Hoboken, NJ, USA: Wiley, 2018.

[3] G. W. Elko and J. Meyer, "Microphone arrays," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2008, ch. 48, pp. 1021–1041.

[4] M. Brandstein and D. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001.

[5] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2004.

[6] G. W. Elko, "Superdirectional microphone arrays," in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, Eds. Berlin, Germany: Springer, 2000, ch. 10, pp. 181–237.

[7] G. Huang, J. Benesty, and J. Chen, "On the design of frequency-invariant beampatterns with uniform circular microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 1140–1153, May 2017.

[8] G. Huang, J. Chen, and J. Benesty, "Insights into frequency-invariant beamforming with concentric circular microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 12, pp. 2305–2318, Dec. 2018.

[9] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2008, ch. 47, pp. 945–978.

[10] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[11] M. Souden, J. Benesty, and S. Affes, "A study of the LCMV and MVDR noise reduction filters," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4925–4935, Sep. 2010.

[12] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.

[13] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[14] C. F. Van Loan, "The ubiquitous Kronecker product," *J. Comput. Appl.Math.*, vol. 123, no. 1–2, pp. 85–100, Nov. 2000.

[15] C. F. Van Loan and N. Pitsianis, "Approximation with Kronecker products," in *Linear Algebra for Large Scale and Real-Time Applications*, M. S. Moonen, G. H. Golub, and B. L. R. De Moor, Eds. Dordrecht: Springer Netherlands, 1993, pp. 293–314.

[16] C. Paleologu, J. Benesty, and S. Ciochina, "Linear system identification based on a Kronecker product decomposition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 10, pp. 1793–1809, Oct. 2018.

[17] G. Itzhak, J. Benesty, and I. Cohen, "Nonlinear Kronecker product filtering for multichannel noise reduction," *Speech Commun.*, vol. 114, pp. 49–59, Oct. 2019.

[18] W. Yang, G. Huang, J. Chen, J. Benesty, I. Cohen, and W. Kellermann, "Robust dereverberation with Kronecker product based multichannel linear prediction," *IEEE Signal Process. Lett.*, vol. 28, pp. 101–105, Dec. 2020.

[19] X. Wang, G. Huang, J. Benesty, J. Chen, and I. Cohen, "Time difference of arrival estimation based on a Kronecker product decomposition," *IEEE Signal Process. Lett.*, vol. 28, pp. 51–55, Dec. 2020.

[20] I. Cohen, J. Benesty, and J. Chen, "Differential Kronecker product beamforming," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 5, pp. 892–902, Jan. 2019.

[21] J. Benesty, I. Cohen, and J. Chen, *Array Processing: Kronecker Product Beamforming*. Berlin, Germany: Springer-Verlag, 2019.

[22] L. N. Ribeiro, A. L. de Almeida, and J. C. M. Mota, "Tensor beamforming for multilinear translation invariant arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Shanghai, May 2016, pp. 2966–2970.

[23] G. Huang, J. Benesty, J. Chen, and I. Cohen, "Robust and steerable Kronecker product differential beamforming with rectangular microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Barcelona, May 2020, pp. 211–215.

[24] L. N. Ribeiro, A. L. de Almeida, and J. C. M. Mota, "Separable linearly constrained minimum variance beamformers," *Signal Process.*, vol. 158, pp. 15–25, May 2019.

[25] R. Sharma, I. Cohen, and J. Benesty, "Adaptive and hybrid Kronecker product beamforming for far-field speech signals," *Speech Commun.*, vol. 120, pp. 42–52, Jun. 2020.

[26] G. Huang, I. Cohen, J. Benesty, and J. Chen, "Kronecker product beamforming with multiple differential microphone arrays," in *Proc. IEEE Sensor Array Multichannel Signal Process.,*, Hangzhou, Jun. 2020, pp. 1–5.

[27] W. Yang, G. Huang, J. Benesty, I. Cohen, and J. Chen, "On the design of flexible Kronecker product beamformers with linear microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brighton, Apr. 2019, pp. 441–445.

[28] X. Huang, A. Acero, H.-W. Hon, and R. Reddy, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[29] P. Renevey and A. Drygajlo, "Entropy based voice activity detection in very noisy conditions," in *Proc. Eur. Conf. on Speech Commun. and Technol.*, Aalborg, Sep. 2001, pp. 1887–1890.

[30] Z. Tuske, P. Mihajlik, Z. Tobler, and T. Fegyó, “Robust voice activity detection based on the entropy of noise-suppressed spectrum,” in *Proc. Eur. Conf. on Speech Commun. and Technol.*, Lisbon, Sep. 2005, pp. 245–248.

[31] G. Huang, J. Benesty, T. Long, and J. Chen, “A family of maximum SNR filters for noise reduction,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 2034–2047, Sep. 2014.