A Two-Stage MMSE Beamformer for Underdetermined Signal Separation

Zbyněk Koldovský[1,2], Petr Tichavský[2], Anh Huy Phan[3], and Andrzej Cichocki[3]

***Abstract*—Blind separation of underdetermined instantaneous mixtures is a popular solution to inverse problems encountered in audio or biomedical applications where the number of sources exceeds the number of sensors. There are two non-equivalent tasks: to identify the mixing matrix and to separate the original sources. In this paper, we focus on the latter task by proposing a novel beamformer that minimizes the theoretical mean square error distance between the separated and original signals. The beamformer has two stages: one for the estimation of signals and one for their refinement. Within the former stage, the signals are assumed to be random and locally stationary, while the latter stage is based on a semi-deterministic model. The experiments prove superior performance of the proposed method compared to conventional MMSE beamforming.**

***Index Terms*—Blind source separation, underdetermined mixtures, nonstationary processes, beamforming**

I. Introduction

The goal of Underdetermined Blind Source Separation (UBSS) is to retrieve r unknown signals from m mixtures when $m < r$. An instantaneous mixture is described by

$$\mathbf{X} = \mathbf{A}\mathbf{S} \tag{1}$$

where $\mathbf{S}$ is a $r \times N$ matrix whose rows contain N samples of original zero-mean independent signals, $\mathbf{A}$ is an $m \times r$ mixing matrix, and $\mathbf{X}$ is the $m \times N$ matrix of the mixed signals. The tasks of finding $\mathbf{A}$ and $\mathbf{S}$ are not equivalent since $m < r$. Most UBSS methods thus consist of two steps: $\mathbf{A}$ is identified first, and then $\mathbf{S}$ are separated using the estimated $\mathbf{A}$. In this paper, we focus on the latter problem.

Finding $\mathbf{S}$ in (1) when $\mathbf{A}$ is known (or estimated) is a classic inverse problem. Numerous array processing techniques [1] and methods exploiting sparsity of $\mathbf{S}$ [2], [3] have been proposed. In this paper, we revise the popular minimum mean-squared error-based (MMSE) beamforming approaches and apply them for the UBSS. MMSE beamforming techniques were recently studied by Eldar et al.; see e.g. [4].

Nomenclature: Matrices will be denoted by the boldface upper case letter; e.g., $\mathbf{B}$. The ijth element of $\mathbf{B}$ will be denoted by upper case letter B_{ij} or as $(\mathbf{B})_{ij}$, and the ith column of the matrix will be denoted by the boldface lower case letter, that is $\mathbf{b}_i$. The superscripts $(\cdot)^H$, $(\cdot)^T$ and $(\cdot)^*$ denote the Hermitian transpose, the transpose and the conjugate operator, respectively.

The MMSE beamformer estimates S_{jt} as $\widehat{S}_{jt} = \mathbf{w}_{j,t}^H \mathbf{x}_t$ where $\mathbf{w}_{j,t}$ is the minimizer of $\mathrm{E}\left[|S_{jt} - \widehat{S}_{jt}|^2\right]$; $\mathrm{E}[\cdot]$ stands for the expectation operator. The solution is [1]

$$\mathbf{w}_{j,t} = D_{jt}\mathbf{R}_t^{-1}\mathbf{a}_j, \tag{2}$$

where $\mathbf{R}_t$ denotes the covariance matrix of $\mathbf{x}_t$, that is $\mathbf{R}_t = \mathrm{E}[\mathbf{x}_t\mathbf{x}_t^H]$, and $D_{jt} = \mathrm{E}[|S_{jt}|^2]$ is the variance of S_{jt}.

Knowledge of $\mathbf{R}_t$ and D_{jt} is the key problem [5]. In the blind scenario, where only $\mathbf{X}$ are known, a conventional approach assumes local stationarity of signals within blocks of length N_1 and replaces $\mathbf{R}_t$ by the sample covariance estimator $\widehat{\mathbf{R}}_t = \frac{1}{N_1}\sum_{k=t-N_1/2}^{t+N_1/2} \mathbf{x}_k\mathbf{x}_k^H$; see Section 7.3 in [1]. A common estimator of D_{jt} is then $\widehat{D}_{jt} = (\mathbf{A}^+\widehat{\mathbf{R}}_t(\mathbf{A}^+)^H)_{jj}$ where $\mathbf{A}^+$ denotes the Moore-Penrose pseudoinverse of $\mathbf{A}$. We will refer to $\mathbf{w}_{j,t}^{\mathrm{con}} = \widehat{D}_{jt}\widehat{\mathbf{R}}_t^{-1}\mathbf{a}_j$ as to the conventional MMSE beamformer.

In the following section, we propose a novel two-stage MMSE beamformer tailored to the underdetermined mixtures given by (1). In Section III, we derive its spatio-temporal variant exploiting the non-whiteness of $\mathbf{S}$, which achieves improved signal-to-interference ratio. Simulations comparing and verifying the efficiency of the beamformers are conducted in Section IV, and Section V concludes the article.

II. Two-stage MMSE Beamformer

The beamformer retrieves the original signals $\mathbf{S}$ using estimation and refinement stages that are based on different models. Within the former stage, $\mathbf{S}$ are assumed to be random and locally stationary. The refinement is based on a semi-deterministic model where every target signal is assumed to be deterministic while the other signals are stochastic and locally stationary [4].

A. Estimation

The estimation step comes from (2). However, compared to the conventional MMSE beamformer, more accurate estimates of $\mathbf{R}_t$ and D_{jt} are obtained using the structure of $\mathbf{R}_t$. The structure follows from (1), that is,

$$\mathbf{R}_t = \mathbf{A}\,\mathtt{diag}(\mathbf{d}_t)\mathbf{A}^H, \tag{3}$$

This work was supported by the Grant Agency of the Czech Republic through the project P103/11/1947.
[1]Faculty of Mechatronic and Interdisciplinary Studies, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic. E-mail: zbynek.koldovsky@tul.cz.
[2]Institute of Information Theory and Automation, Pod vodárenskou věží 4, P.O.Box 18,182 08 Prague 8, Czech Republic. E-mail: tichavsk@utia.cas.cz.
[3]Brain Science Institute, RIKEN, Wakoshi, Japan. E-mail: phan@brain.riken.jp, cia@brain.riken.jp.

and it holds that $D_{jt} \geq 0$ where D_{jt} is the jth element of $\mathbf{d}_t$. Here, $\mathtt{diag}(\cdot)$ denotes a diagonal matrix with diagonal elements of the argument.

To estimate $\mathbf{d}_t$, we solve the nonnegative quadratic programming optimization problem

$$\widehat{\mathbf{d}}_t^{\mathrm{est}} = \arg\min_{\mathbf{b}} \left\| \mathbf{C}_t(\widehat{\mathbf{R}}_t - \mathbf{A}\,\mathtt{diag}(\mathbf{b})\,\mathbf{A}^H) \right\|_F^2 \text{ w.r.t. } \mathbf{b} \geq 0, \tag{4}$$

where $\|\cdot\|_F$ is the Frobenius norm, and $\mathbf{C}_t$ is a weighting matrix whose choice will be discussed later. An efficient solver of (4) is the interior-point method [6] initialized using the unconstrained solution of (4), which is

$$\mathbf{b}_t^{\mathrm{unc}} = \widetilde{\mathbf{A}}_t^{+}\mathtt{vec}(\mathbf{C}_t\widehat{\mathbf{R}}_t) \tag{5}$$

where $\widetilde{\mathbf{A}}_t = \mathbf{A}^* \odot (\mathbf{C}_t\mathbf{A})$, $\odot$ denotes the Khatri-Rao (a column-wise Kronecker) product, and $\mathtt{vec}(\cdot)$ is the vectorization operator. Provided that all elements of $\mathbf{b}_t^{\mathrm{unc}}$ are nonnegative[1], $\mathbf{b}_t^{\mathrm{unc}}$ is already the solution of (4).

$\mathbf{R}_t$ is estimated as $\widehat{\mathbf{R}}_t^{\mathrm{est}} = \mathbf{A}\,\mathtt{diag}\left(\widehat{\mathbf{d}}_t^{\mathrm{est}}\right)\mathbf{A}^H$, so the estimation beamformer reads

$$\mathbf{w}_{j,t}^{\mathrm{est}} = \widehat{D}_{jt}^{\mathrm{est}}\,(\widehat{\mathbf{R}}_t^{\mathrm{est}})^{-1}\mathbf{a}_j \tag{6}$$

where $\widehat{D}_{jt}^{\mathrm{est}}$ is the jth element of $\widehat{\mathbf{d}}_t^{\mathrm{est}}$. The samples of S_{jt} are thus estimated as $\widehat{S}_{jt}^{\mathrm{est}} = (\mathbf{w}_{j,t}^{\mathrm{est}})^H\mathbf{x}_t$.

In practice, it is reasonable to assume block-stationarity of signals in this stage and to perform the estimation block-by-block with the length of block N_2 where $\widehat{\mathbf{R}}_t$, $\mathbf{C}_t$, $\widehat{\mathbf{R}}_t^{\mathrm{est}}$, $\widehat{\mathbf{d}}_t^{\mathrm{est}}$, and thus $\mathbf{w}_{j,t}^{\mathrm{est}}$ are constant within the block. This leads to computational savings.

B. Refinement

In this stage, the samples S_{jt} are refined assuming, consecutively for each $j = 1,\ldots,d$, that S_{jt} is deterministic while S_{it}, $i \neq j$, are random and locally stationary. The MMSE beamformer (2) for this model reads [4]

$$\mathbf{w}_{j,t} = V_{jt}\left(V_{jt}\mathbf{a}_j\mathbf{a}_j^H + \sum_{i=1,i\neq j}^{r} D_{it}\mathbf{a}_i\mathbf{a}_i^H\right)^{-1}\mathbf{a}_j, \tag{7}$$

where $V_{jt} = |S_{jt}|^2$.

We propose to replace the unknown V_{jt} and D_{it}, $i \neq j$, by the best available estimators obtained due to the estimation stage. Namely, we replace V_{jt} in (7) by $\widehat{V}_{jt} = |\widehat{S}_{jt}^{\mathrm{est}}|^2$ and D_{it}, $i \neq j$, by $\widehat{D}_{it}^{\mathrm{est}}$. Then, the refinement beamformer can be written in the form

$$\mathbf{w}_{j,t}^{\mathrm{ref}} = \widehat{V}_{jt}\left(\widehat{\mathbf{R}}_t^{\mathrm{est}} + (\widehat{V}_{jt} - \widehat{D}_{jt}^{\mathrm{est}})\mathbf{a}_j\mathbf{a}_j^H\right)^{-1}\mathbf{a}_j. \tag{8}$$

Using the matrix inversion lemma, we obtain

$$\mathbf{w}_{j,t}^{\mathrm{ref}} = \beta_{j,t}\mathbf{w}_{j,k}^{\mathrm{est}}, \tag{9}$$

where, after some simplifications,

$$\beta_{j,t} = \frac{\widehat{V}_{jt}}{\widehat{D}_{jt}^{\mathrm{est}} + (\widehat{V}_{jt} - \widehat{D}_{jt}^{\mathrm{est}})\mathbf{a}_j^H\mathbf{w}_{j,t}^{\mathrm{est}}}. \tag{10}$$

Indeed, $\beta_{j,t}$ is just a scaling correction factor due to the refinement stage. If $\widehat{V}_{jt} = \widehat{D}_{jt}^{\mathrm{est}}$, then $\beta_{j,t} = 1$.

[1]It can be shown that $\mathbf{b}_t^{\mathrm{unc}}$ is always a real-valued vector when $\mathbf{C}_t = \mathbf{I}$.

III. Spatio-temporal MMSE Beamforming

This section describes an extension of the proposed beamformer for spatio-temporal (ST) beamforming. An ST beamformer estimates S_{jt} using also $P-1$ previous values of the observation vector $\mathbf{x}_t$ where $P > 1$ is a free integer parameter. The beamformer is represented by a vector $\mathbf{w}_{j,t}$ of length $mP \times 1$, and the output is $\mathbf{w}_{j,t}^H\widetilde{\mathbf{x}}_t$ where $\widetilde{\mathbf{x}}_t = [\mathbf{x}_t^T, \mathbf{x}_{t-1}^T, \ldots \mathbf{x}_{t-P+1}^T]^T$. A straightforward computation yields that the ST MMSE beamformer is given by

$$\mathbf{w}_{j,t} = \mathrm{E}[\widetilde{\mathbf{x}}_t\widetilde{\mathbf{x}}_t^H]^{-1}(\boldsymbol{\delta}_{j,t} \otimes \mathbf{a}_j) \tag{11}$$

where $\otimes$ denotes the Kronecker product, and $\boldsymbol{\delta}_{j,t}$ is a vector of autocovariances of S_{jt} of length P, that is, its pth element is $\mathrm{E}[S_{jt}S_{j(t-p+1)}^*]$, $p = 1,\ldots,P$.

Before deriving the estimation and refinement steps, we explore the structure of the covariance matrix $\mathrm{E}[\widetilde{\mathbf{x}}_t\widetilde{\mathbf{x}}_t^H]$, from here denoted as $\mathbf{G}_t$. It holds that

$$\mathbf{G}_t = \begin{bmatrix} \mathbf{R}_t[0] & \mathbf{R}_t[1] & \ldots & \mathbf{R}_t[P-1] \\ \mathbf{R}_t[1] & \mathbf{R}_t[0] & \ldots & \mathbf{R}_t[P-2] \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{R}_t[P-1] & \mathbf{R}_t[P-2] & \ldots & \mathbf{R}_t[0] \end{bmatrix} \tag{12}$$

where $\mathbf{R}_t[\tau] = \mathrm{E}[\mathbf{x}_t\mathbf{x}_{t-\tau}^H]$, $\tau = 0,\ldots,P-1$ (note that $\mathbf{R}_t = \mathbf{R}_t[0]$). $\mathbf{R}_t[\tau]$ has the structure $\mathbf{R}_t[\tau] = \mathbf{A}\,\mathtt{diag}(\mathbf{d}_{t,\tau})\mathbf{A}^H$ where $\mathbf{d}_{t,\tau}$ denotes the diagonal of $\mathrm{E}[\mathbf{s}_t\mathbf{s}_{t-\tau}^H]$ and it holds that $(\mathbf{d}_{t,\tau})_j = (\boldsymbol{\delta}_{t,j})_{\tau+1}$ and $\mathbf{d}_{t,0} = \mathbf{d}_t$. Hence,

$$\begin{aligned} \mathbf{G}_t &= \mathtt{btoep}\Big[\mathbf{R}_t[0],\ldots,\mathbf{R}_t[P-1]\Big] \\ &= \mathtt{btoep}\left[\mathbf{A}\,\mathtt{diag}(\mathbf{d}_{t,0})\mathbf{A}^H,\ldots,\mathbf{A}\,\mathtt{diag}(\mathbf{d}_{t,P-1})\mathbf{A}^H\right] \\ &= (\mathbf{I}_P \otimes \mathbf{A})\cdot\mathtt{btoep}\Big[\mathtt{diag}(\mathbf{d}_{t,0}),\ldots,\mathtt{diag}(\mathbf{d}_{t,P-1})\Big] \\ &\quad\cdot(\mathbf{I}_P \otimes \mathbf{A}^H) \\ &= (\mathbf{I}_P \otimes \mathbf{A})\,\mathbf{P}\,\mathtt{bdiag}\Big[\mathtt{toep}(\boldsymbol{\delta}_{1,t}),\ldots,\mathtt{toep}(\boldsymbol{\delta}_{r,t})\Big] \\ &\quad\cdot\mathbf{P}^T(\mathbf{I}_P \otimes \mathbf{A}^H) \end{aligned} \tag{13}$$

where $\mathtt{btoep}[\cdot]$, $\mathtt{bdiag}[\cdot]$, and $\mathtt{toep}[\cdot]$, respectively, denote a symmetric block-Toeplitz matrix composed of blocks given by the argument, a block-diagonal matrix, and a symmetric Toeplitz matrix with the first column given by the argument. $\mathbf{I}_P$ is the $P \times P$ identity matrix, and $\mathbf{P}$ is the (p, P) vec-permutation matrix that for every $p \times P$ matrix $\mathbf{B}$ satisfies $\mathtt{vec}(\mathbf{B}) = \mathbf{P}\,\mathtt{vec}(\mathbf{B}^T)$.

A conventional ST MMSE beamformer does not take the structure of $\mathbf{G}_t$ into account and estimates it using the sample estimates of $\mathbf{R}_t[\tau]$, $\tau = 0,\ldots,P-1$, denoted as $\widehat{\mathbf{R}}_t[\tau]$. Namely, $\widehat{\mathbf{G}}_t = \mathtt{btoep}\Big[\widehat{\mathbf{R}}_t[0],\ldots,\widehat{\mathbf{R}}_t[P-1]\Big]$ and the conventional beamformer reads $\mathbf{w}_{j,t}^{\mathrm{con}} = \widehat{\mathbf{G}}_t^{-1}(\widehat{\boldsymbol{\delta}}_{j,t} \otimes \mathbf{a}_j)$ where $\widehat{\boldsymbol{\delta}}_{j,t} = \Big[(\mathbf{A}^+\widehat{\mathbf{R}}_t[0](\mathbf{A}^+)^H)_{jj},\ldots,(\mathbf{A}^+\widehat{\mathbf{R}}_t[P-1](\mathbf{A}^+)^H)_{jj}\Big]^T$.

A. Estimation

As in Section II.A, the estimation stage is based on computing the estimates $\widehat{\mathbf{d}}_{t,0}^{\mathrm{est}},\ldots,\widehat{\mathbf{d}}_{t,P-1}^{\mathrm{est}}$ and building the estimate of $\mathbf{G}_t$ according to (13) as

$$\widehat{\mathbf{G}}_t^{\mathrm{est}} = (\mathbf{I}_P \otimes \mathbf{A})\,\mathtt{btoep}\Big[\mathtt{diag}(\widehat{\mathbf{d}}_{t,0}^{\mathrm{est}}),\ldots,\mathtt{diag}(\widehat{\mathbf{d}}_{t,P-1}^{\mathrm{est}})\Big] \cdot(\mathbf{I}_P \otimes \mathbf{A}^H). \tag{14}$$

While $\widehat{\mathbf{d}}^{\text{est}}_{t,0}$ is estimated as in (4) with the non-negativity constraint, the unconstrained estimation as in (5) is used to estimate the other vectors, that is, $\widehat{\mathbf{d}}^{\text{est}}_{t,\tau} = \widetilde{\mathbf{A}}^{+}_{t}\mathtt{vec}(\mathbf{C}_t\widehat{\mathbf{R}}_t[\tau])$ for $\tau = 1, \ldots, P-1$.

However, the resulting $\widehat{\mathbf{G}}^{\text{est}}_{t}$ need not be positive definite (PD), which is required due to the stability[2] of the beamformer. From the last equation in (13) it follows that $\widehat{\mathbf{G}}^{\text{est}}_{t}$ is PD iff $\mathtt{toep}(\widehat{\boldsymbol{\delta}}^{\text{est}}_{j,t})$, $j = 1, \ldots, r$, are also PD, and this is guaranteed when the autocovariances $\widehat{\boldsymbol{\delta}}^{\text{est}}_{j,t}$ correspond to positive definite functions. We therefore apply a correction procedure to each $\widehat{\boldsymbol{\delta}}^{\text{est}}_{j,t}$ to ensure that all poles of the corresponding AR sequence lie inside the complex unit circle. Poles whose magnitude is greater than one are put equal to their inverse conjugate values.

Once the PD of $\widehat{\mathbf{G}}^{\text{est}}_{t}$ is ensured, (11) is estimated as

$$\mathbf{w}^{\text{est}}_{j,t} = (\widehat{\mathbf{G}}^{\text{est}}_{t})^{-1}(\widehat{\boldsymbol{\delta}}^{\text{est}}_{j,t} \otimes \mathbf{a}_j), \tag{15}$$

and the output is $\widehat{S}^{\text{est}}_{jt} = (\widehat{\boldsymbol{\delta}}^{\text{est}}_{j,t} \otimes \mathbf{a}_j)^H (\widehat{\mathbf{G}}^{\text{est}}_{t})^{-1}\widetilde{\mathbf{x}}_t$.

B. Refinement

The refinement stage of the ST beamformer uses the estimated scale of signals, that is, $\widehat{V}_{jt} = |\widehat{S}^{\text{est}}_{jt}|^2$ in order to make the estimation of (11) more accurate. We define the refined autocovariance of S_{jt} as

$$\widehat{\boldsymbol{\delta}}^{\text{ref}}_{j,t} = \frac{\widehat{V}_{jt}}{(\widehat{\boldsymbol{\delta}}^{\text{est}}_{j,t})_1}\widehat{\boldsymbol{\delta}}^{\text{est}}_{j,t}. \tag{16}$$

This is used to build the refined covariance matrix with the improved estimate of scale of the jth source, as

$$\widehat{\mathbf{G}}^{\text{ref},j}_{t} = (\mathbf{I}_P \otimes \mathbf{A})\,\mathbf{P}\,\mathtt{bdiag}\Big[\mathtt{toep}(\widehat{\boldsymbol{\delta}}^{\text{est}}_{1,t}), \ldots, \mathtt{toep}(\widehat{\boldsymbol{\delta}}^{\text{est}}_{j-1,t}), \mathtt{toep}(\widehat{\boldsymbol{\delta}}^{\text{ref}}_{j,t}), \mathtt{toep}(\widehat{\boldsymbol{\delta}}^{\text{est}}_{j+1,t}), \ldots, \mathtt{toep}(\widehat{\boldsymbol{\delta}}^{\text{est}}_{r,t})\Big]\,\mathbf{P}^T\,(\mathbf{I}_P \otimes \mathbf{A}^H). \tag{17}$$

By comparing (14) and (17), the latter can be written as

$$\widehat{\mathbf{G}}^{\text{ref},j}_{t} = \widehat{\mathbf{G}}^{\text{est}}_{t} - (\mathbf{I}_P \otimes \mathbf{a}_j)\,\mathtt{toep}(\widehat{\boldsymbol{\delta}}^{\text{est}}_{j,t} - \widehat{\boldsymbol{\delta}}^{\text{ref}}_{j,t})(\mathbf{I}_P \otimes \mathbf{a}_j^H).$$

This expression can be used to speed-up the inversion of $\widehat{\mathbf{G}}^{\text{ref},j}_{t}$ using $(\widehat{\mathbf{G}}^{\text{est}}_{t})^{-1}$ known from the estimation stage and the matrix inversion lemma. The refined beamformer is then given by

$$\mathbf{w}^{\text{ref}}_{j,t} = (\widehat{\mathbf{G}}^{\text{ref},j}_{t})^{-1}(\widehat{\boldsymbol{\delta}}^{\text{ref}}_{j,t} \otimes \mathbf{a}_j). \tag{18}$$

It can be verified that both stages described in Section II coincide with their ST variants when $P = 1$.

C. Choice of the weighting matrix

In [7], a method for separation of underdetermined instantaneous real-valued mixtures called UDSEP has been derived. UDSEP assumes block-stationary white Gaussian signals and performs a weighted tensor decomposition to estimate $\mathbf{A}$ and the variances of signals on blocks. The criterion for the tensor decomposition is

$$\sum_t \left\| \mathbf{C}_t(\widehat{\mathbf{R}}_t[0] - \mathbf{A}\,\mathtt{diag}(\mathbf{d}_t)\,\mathbf{A}^T) \right\|_F^2 \tag{19}$$

where the summation goes over the blocks of signals' stationarity, $\mathbf{d}_t$ is the vector of variance of signals on the tth block, and

$$\mathbf{C}_t = (\widehat{\mathbf{R}}_t[0] + \epsilon\mathbf{I})^{-1} \tag{20}$$

are the weighting matrices with a suitable small positive ϵ to restrain the regularity. These weights are optimized for the block-stationary real-valued white Gaussian model of signals.

By comparing (4) with the argument of (19), it follows that (4) is the optimum criterion for the estimation of $\mathbf{d}_t$ ($\mathbf{d}_{t,0}$) when the signals obey the aforementioned model. Therefore, we choose the same weighting matrices as in (20) to estimate the (auto)covariances of signals, although there is no claim of inherited optimality as the beamformer does not rely on a specific model of signals.

[2]The output power of the beamformer, which is $\mathbf{w}^H_{j,t}\mathbf{G}_t\mathbf{w}_{j,t} \geq 0$, must be non-negative. Therefore, $\widehat{\mathbf{G}}^{\text{est}}_{t}$ must be PD.

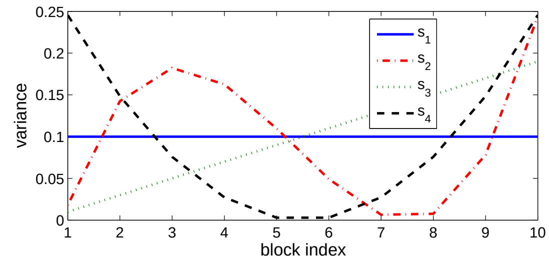

Fig. 1. Variances of the artificial signals on blocks.

IV. Experimental Evaluation

A. Artificial signals

Our first experiment consists of the separation of four artificial signals of the length $N = 10000$ samples that are mixed into three channels. The mixing matrix has columns $\mathbf{a}_j = [1, \cos(\phi_j), \sin(\phi_j)]^T$ where $(\phi_1, \phi_2, \phi_3, \phi_4) = (0, \pi/4, \pi/2, 3\pi/4)$. The signals obey the block-stationary AR Gaussian model, where there are $M = 10$ blocks. Variances of the signals on the blocks are defined as shown in Fig. 1. We define AR coefficients in the following order: $(1, \rho)$, $(1, -\rho)$, $(1, 0, \rho)$, $\ldots$, $(1, 0, 0, 0, 0, -\rho)$. Using this order, the AR model of the jth signal on the kth block has the $[((k + 2j - 3) \mod M) + 1]$th coefficients. Consequently, the signals never have the same variance and spectrum in any block, up to the case $\rho = 0$.

For comparison, we consider two cases, respectively, when $\mathbf{A}$ is known or estimated by an UBSS algorithm. For the latter case, we consider the UDSEP algorithm from [7] with $M = 10$ (the number of blocks in the block Gaussian white stationary model).

The signals $\mathbf{S}$ are separated by the following methods to be compared: the conventional MMSE beamformer (MMSEconv), the estimation stage of the proposed MMSE beamformer (MMSEest), and the output of the novel beamformer after the refinement (MMSEnew). Each beamformer is

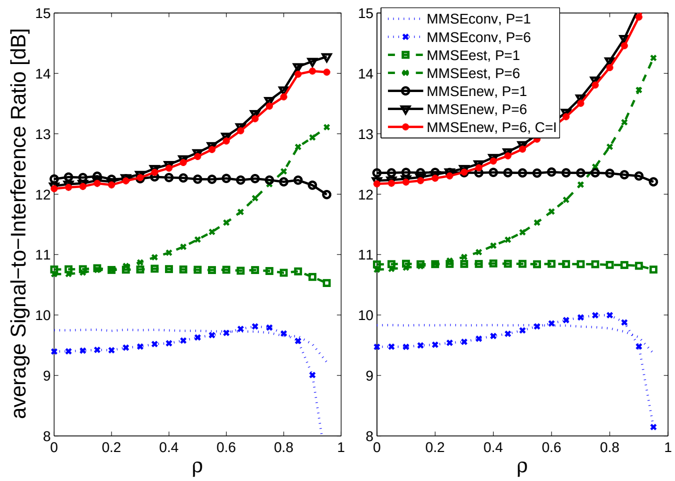

Fig. 2. Average SINR of extracted signals by the compared beamformers.

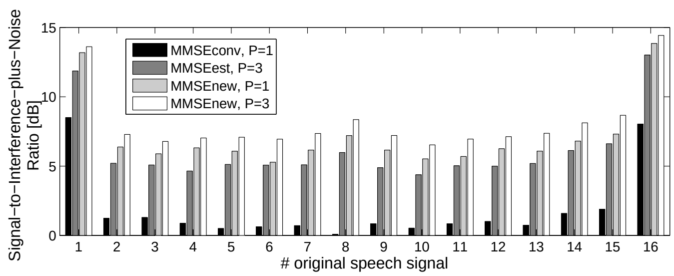

Fig. 3. SINR of 16 separated speech signals from their 9 noisy mixtures.

tested with $P = 1$ and $P = 6$. MMSEnew is also tested when the weighting matrices $\mathbf{C}_t$ are equal to the identity matrix. The sample covariance matrices are estimated with $N_1 = N/M$. The estimation stage is running with $N_2 = N_1$ (described at the end of Section II.A).

100 independent trials were run for each $\rho \in [0, 0.95]$ and the separated signals were evaluated in term of Signal-to-Interference ratio (SIR); averages of SIR over all trials and separated signals are shown in Fig. 2. MMSEconv achieves the lowest SIR of about 10 dB. MMSEest improves the results by about 1 dB, which is improved by one more dB by MMSEnew. For $P = 1$, the beamformers achieve approximately the same SIR for all ρ while their performances grow when $P = 6$. This is due to the fact that the ST beamforming is able to exploit the non-whiteness of signals ($\rho > 0$). Only the ST conventional beamformer fails to improve the SIR due to large errors in the estimation of signals' (auto)covariances.

For ρ close to zero, the beamforming with $P = 6$ is overparameterized and is slightly worse than with $P = 1$. The signals are white for $\rho = 0$ so $P = 1$ is optimal. The last observation is that the weighting matrices $\mathbf{C}_t = (\widehat{\mathbf{R}}_t[0] + \epsilon\mathbf{I})^{-1}$ give a slightly better performance than $\mathbf{C}_t = \mathbf{I}$; here demonstrated for MMSEnew with $P = 6$.

B. Instantaneous noisy mixture of speech signals

16 speech signals were mixed into 9 channels using a mixing matrix whose columns are defined as $\mathbf{a}_j = [1, \cos\phi_j, \cos 2\phi_j, \ldots, \cos 8\phi_j]^T$ where $\phi_j = j\pi/17$, $j = 1, \ldots, 16$, and a white Gaussian noise was added to the mixture at the ratio of 20 dB. The speech signals are utterances 7.5 seconds long sampled at 16 kHz, normalized to zero mean and unit variance[3]. The other parameters of this experiment are as follows: $\mathbf{A}$ is known, $N_1 = 1200$, $N_2 = 150$, the ST beamforming is performed with $P = 3$. For clarity, the results of only selected methods are shown in Figure 3.

MMSEconv achieves significantly lower Signal-to-Interference-plus-Noise Ratio (SINR) than MMSEest and MMSEnew. The performance of its ST variant is even worse, not shown here for clarity. By contrast, the ST variants of MMSEest as well as of MMSEnew achieve improved performance with a higher value of P, here, $P = 3$. Next, MMSEnew with $P = 1$ achieves higher SINR than MMSEest with $P = 3$, which is practical in view of the fact that the former method is computationally simpler than the latter one.

This example demonstrates that the proposed beamformer yields improved SINR on noisy mixtures of real-world signals that do not obey any specific mathematical model such as the block Gaussian AR model in the previous example.

V. Conclusions

The proposed two-stage MMSE beamformer is an effective tool to separate underdetermined mixtures of signals. Its accuracy in separation of nonstationary signals was shown to be better than that of the conventional MMSE beamformer. The beamformer does not rely on sparsity of signals and can be applied to signals of a wide class, such as biological signals EEG, ECG, etc; see also [8]. The Matlab implementation is available at [9].

References

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*, John Wiley & Sons, 2004.

[2] I. F. Gorodnitsky and B D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[3] P. Bofill, M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353-2362, Nov. 2001.

[4] Y. C. Eldar, A. Nehorai, and P. S. La Rosa, "A Competitive Mean-Squared Error Approach to Beamforming," *Signal Processing, IEEE Transactions on*, vol. 55, no. 11, pp. 5143–5154, Nov. 2007.

[5] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, pp. 397–401, July 1988.

[6] J. A. Nocedal and S. J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.

[7] P. Tichavský and Z. Koldovský, "Weight Adjusted Tensor Method for Blind Separation of Underdetermined Mixtures of Nonstationary Sources, " *IEEE Trans. on Signal Processing*, Vol. 59, No. 3, pp. 1037-1047, March 2011.

[8] Z. Koldovský, A. H. Phan, P. Tichavský, and A. Cichocki, "A Treatment of EEG Motor Imagery Data by Underdetermined Blind Source Separation, " *Proc. of EUSIPCO 2012*, pp. 1484–1488, Bucharest, Romania, August 27-31, 2012.

[9] Z. Koldovský, m-code of the two-stage MMSE beamformer [online], `http://itakura.ite.tul.cz/zbynek/downloads.htm`.

[3]The signals are available in the IEEE Xplore database in the multimedia attachment of [7].