Digital Signal Processing 68 (2017) 154–166

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Blind separation of partially overlapping data packets ☆

Mu Zhou, Alle-Jan van der Veen *

TU Delft, Fac. EEMCS, Mekelweg 4, 2628 CD Delft, The Netherlands

ARTICLE INFO

Article history:
Available online 16 June 2017

Keywords:
Generalized singular value decomposition
Blind source separation
Subspace intersection
Signed URV decomposition
Automatic Identification of Ships (AIS)

ABSTRACT

The paper discusses the separation of partially overlapping data packets by an antenna array in narrowband communication systems. This problem occurs in asynchronous communication systems and several transponder systems such as Radio Frequency Identification (RFID) for wireless tags, Automatic Identification System (AIS) for ships, and Secondary Surveillance Radar (SSR) and Automatic Dependent Surveillance—Broadcast (ADS—B) for aircraft. Partially overlapping data packages also occur as inter-cell interference in mutually unsynchronized communication systems. Arbitrary arrival times of the overlapping packets cause nonstationary scenarios and makes it difficult to identify the signals using standard blind beamforming techniques. After selecting an observation interval, we propose subspace-based algorithms to suppress partially present (interfering) packets, as a preprocessing step for existing blind beamforming algorithms that assume stationary (fully overlapping) sources. The proposed algorithms are based on a subspace intersection, computed using a generalized singular value decomposition (GSVD) or a generalized eigenvalue decomposition (GEVD). In the second part of the paper, the algorithm is refined using a recently developed subspace estimation tool, the Signed URV algorithm, which is closely related to the GSVD but can be computed non-iteratively and allows for efficient subspace tracking. Simulation results show that the proposed algorithms significantly improve the performance of classical algorithms designed for block stationary scenarios in cases where asynchronous co-channel interference is present. An example on experimental data from the AIS ship transponder system confirms the effectiveness of the proposed algorithms in a real application.

1. Introduction

Co-channel interference is a growing concern in wireless communication applications. One approach for interference mitigation is to use an antenna array. Beamforming techniques allow to receive the target signals and suppress the interference signals, assuming the array response vector of each of the signals is known. Blind beamforming techniques aim to estimate these array response vectors.

In many cases, the interference is intermittent and unsynchronized. For example, inter-cell interference reduces channel capacity in Multiple Input Multiple Output (MIMO) cellular networks [6,7]. Also, ad-hoc communication systems or wireless sensor networks where devices transmit whenever data is available result in multiple partially overlapping data packets at the receiver. Other examples are Radio Frequency Identification (RFID) systems with multiple tags, the Automatic Identification System (AIS) for ships, wherein transponders periodically report their locations [8,9], the secondary surveillance radar (SSR) [10,11,2] and similar Automatic Dependent Surveillance—Broadcast (ADS—B) transponder systems for aircraft. Another example is multiple unsynchronized Wireless Local Area Network (WLAN) systems in the same service area.

In this paper we consider the separation of partially overlapping data packets using blind beamforming techniques under narrowband assumptions. We consider an observation interval (typically a sliding window) matched to the length of the data packets, and consider packets that are fully inside this window as target signals, and packets that are partially in the window as interfering signals. It is important to realize that, in this scenario, there is no inherent property that defines a “target” or “interference” signal, the classification is based on the position of packets in the observation interval.

The approach is to collect a block of data from an analysis window. The data block is split into two sub-blocks, and we compare

☆ This work was supported in part by the China Scholarship Council from P.R. China. Some parts of this paper were presented at the Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California, USA, Nov. 2011 [1], IEEE CAMSAP'11 [2], IEEE Sensors Array Multichannel Workshop (SAM) 2012 [3], and IEEE ICASSP'14 [4]. These conference papers showed early versions of the proposed algorithms and did not contain explanations, proofs or simulations of the final algorithm. More extensive results have been published in the PhD thesis of Mu Zhou [5].

* Corresponding author.
E-mail address: a.j.vanderveen@tudelft.nl (A.-J. van der Veen).

http://dx.doi.org/10.1016/j.dsp.2017.06.009
1051-2004/

the subspaces present in each block. Specifically, a generalized singular value decomposition (GSVD) allows to match basis vectors within the subspaces to each other, and target/interference signal classification is based on detecting differences in signal power between the two blocks. The subspace information from the GSVD directly leads to a beamformer to suppress the interfering signals while keeping the target signals. The analysis window can then shift a number of samples and the process is repeated, allowing previously classified "interference signals" to become properly aligned and be detected as target signals.

It could happen that the resulting subspace contains multiple target signals. In that case, the proposed algorithm return a mixture of the nearly fully overlapping target signals, and other properties should provide further separation, such as constant modulus properties (the Algebraic Constant Modulus (ACMA) algorithm [12]) or related algorithms based on fourth-order cumulants (the Joint Approximation Diagonalization of Eigen-matrices (JADE) [13] and the Multi-User Kurtosis (MUK) algorithms [14]). Such algorithms explicitly assume stationary signals and therefore typically cannot handle intermittent interference or signals with non-stationary properties, and the algorithms in this paper can serve as a preprocessing step both to filter out the intermittent signals and to arrive at a nearly synchronous scenario.

To understand why traditional blind source separation algorithms based on cumulants such as ACMA and JADE fail on intermittent sources, consider first a data matrix $\mathbf{X}$ consisting of N samples of a mixture of *stationary* sources. Based on $\mathbf{X}$, these algorithms estimate a cumulant matrix $\mathbf{Q}$ and derive separating beamformers from it. Each entry of $\mathbf{Q}$ can be written as $S_4/N - S_2/N^2$, where S_4 is a sum of fourth-order products of entries of $\mathbf{X}$, and S_2 a combination of sums of second-order products. If we now augment $\mathbf{X}$ with N "zero" columns to $[\mathbf{X}, \mathbf{0}]$, then S_4 and S_2 do not change, while the weights $(1/N)$ and $(1/N^2)$ scale with factors $(1/2)$ and $(1/4)$, respectively. Very quickly, $\mathbf{Q}$ loses the structure on which the computation of the beamformers rely. This simple example shows that ACMA and JADE are not reliable for separating intermittent sources, and this is confirmed in the simulations in Sec. 8.

The paper has two parts. We first propose a generic algorithm based on the GSVD [15] or the related generalized eigenvalue decomposition (GEVD). We then work out an implementation based on a new tool—the Signed URV (SURV) algorithm [16,17]. This leads to a computationally efficient technique that allows for tracking and improved noise processing. Simulations and an experiment using acquired AIS data are provided to confirm the results.

Interference cancellation using oblique projections has been studied in [18,19], assuming the "target" and "'interference" subspaces are known. Here, we focus on the estimation of the required subspace information so that these tools can be applied. Not many papers consider intermittent interference cancellation based on subspace techniques. For the blind separation of partially overlapping SSR signals, Petrochilos et al. proposed a block-based tracking algorithm [20,21] based on detecting and projecting out rank-1 components representing time segments where only a single source is present. The existence of such segments can be considered as a simplified special case of our scenarios.

Notation

Matrices and vectors are denoted by uppercase and lowercase boldface symbols, respectively. $(\mathbf{A})_{ij}$ denotes the i,jth entry of a matrix $\mathbf{A}$. For a matrix $\mathbf{A}$, $\mathbf{A}^H$ denotes the complex conjugate transpose, and $\mathbf{A}^\dagger$ denotes the Moore–Penrose matrix pseudo-inverse. If $\mathbf{A}$ has full column rank, then $\mathbf{A}^\dagger = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$.

$\mathrm{E}\{\cdot\}$ is the expectation operator.

$\|\cdot\|$ denotes the matrix 2-norm, which is equal to the largest singular value of the matrix.

Subspaces are denoted by calligraphic symbols. The column span (range) of a matrix $\mathbf{A}$ is $\mathcal{A} = \mathrm{ran}(\mathbf{A})$.

2. Data model

2.1. Signals

We consider unknown discrete-time intermittent signals (data packets) $s_i[k]$ where i is the signal index and k is the time index. Each signal consists of a stretch of N_p nonzero values, preceded and followed by zeros. For simplicity of notation, all intermittent signals will have the same packet length N_p (this is generalized at a later stage). There are d signals, and they are stacked in a vector $\mathbf{s}[k] = [s_1[k], \cdots, s_d[k]]^T$.

We assume that the receiver has an antenna array with M antennas, and we stack the (complex-valued) antenna signals into a vector $\mathbf{x}[k] \in \mathcal{C}^M$. In a narrowband scenario, the received signal vector is an instantaneous mixture

$$\mathbf{x}[k] = \mathbf{h}_1 s_1[k] + \cdots + \mathbf{h}_d s_d[k] + \mathbf{n}[k] = \mathbf{H}\mathbf{s}[k] + \mathbf{n}[k] \tag{1}$$

where the vectors $\mathbf{h}_i$, $i = 1, \cdots, d$ are the channel vectors (array response vectors) corresponding to each signal, $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_d] \in \mathcal{C}^{M\times d}$ is the channel matrix, $\mathbf{s} = [s_1, \cdots, s_d]^T$ is the source vector, and $\mathbf{n} \in \mathcal{C}^M$ is the noise vector.

We assume that the unknown channel matrix $\mathbf{H}$ has full column rank. This also implies that $d \le M$. In our applications, we have an inherent scaling indeterminacy between signals and channel vectors; without loss of generality we will assume that the channel vectors are all scaled to $\|\mathbf{h}_i\| = 1$ (this can be achieved by exchanging a scaling factor with $s_i[n]$). No further parametric structure is assumed on $\mathbf{H}$, e.g., we do not consider a calibrated array, with channel vectors functions of source directions and antenna locations. Also, multipath and antenna coupling may be present as this leads to the same instantaneous mixture model (1).

The noise is modeled by i.i.d. zero mean Gaussian vectors, with covariance matrix $\mathbf{R}_n = \mathrm{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma^2\mathbf{I}$. We assume that the noise power σ^2 is known.

If we have collected N_s observations $\mathbf{x}[k]$, then we can collect these into a matrix $\mathbf{X} = [\mathbf{x}[1], \cdots, \mathbf{x}[N_s]]$, and similarly for the source signals and the noise. The corresponding data model is

$$\mathbf{X} = \mathbf{H}\mathbf{S} + \mathbf{N}. \tag{2}$$

The sample covariance matrix is $\hat{\mathbf{R}}_x = \frac{1}{N_s}\mathbf{X}\mathbf{X}^H$.

2.2. Separation scenario

We assume that we have obtained N_s samples of data corresponding to an "analysis window". Thus, in the data model (2), $\mathbf{X}$ is known and $\mathbf{H}, \mathbf{S}$ are unknown. Our algorithms are based on splitting this window into two parts and comparing the subspaces determined by each part. The splitting can be done in several ways, each corresponding to different definitions of target signals and interference signals. Here, we limit the presentation to one scenario, explained below. A second scenario applicable to continuously present target signals is described in Appendix A.

We split the analysis window into three blocks (see Fig. 1). A target signal is defined by being centered in the middle block, and the corresponding data matrix is denoted by $\mathbf{X}_1$. Interfering signals are defined by being more present in the first or third block. Samples from these two blocks are combined into a single data matrix $\mathbf{X}_2$ as shown in Fig. 1.

As a refinement of (2), assume that there are d_s target signals and d_f interference signals. The channel vectors of the target signals are collected in a matrix $\mathbf{H}_s$, and those of the interference signals in $\mathbf{H}_f$. We also define $\mathbf{H} = [\mathbf{H}_s, \mathbf{H}_f]$.

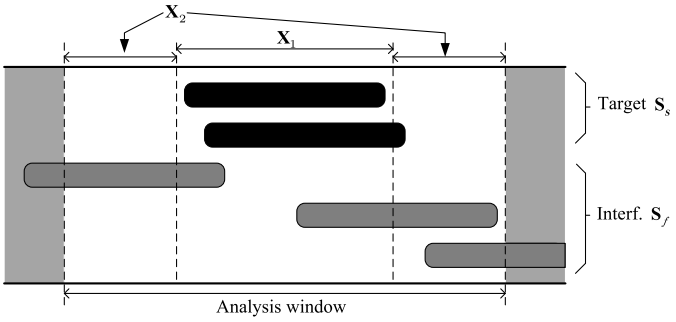

Fig. 1. Separation scenario: target signals are present mostly in $\mathbf{X}_1$. The algorithms are based on comparing the power (or: number of samples) of each signal in $\mathbf{X}_1$ and $\mathbf{X}_2$.

The corresponding data model is then

$$\begin{aligned} \mathbf{x}_1[n] &= \mathbf{H}_s\mathbf{s}_{s1}[n] + \mathbf{H}_f\mathbf{s}_{f1}[n] + \mathbf{n}_1[n], \quad n = 1, \cdots, N_1 \\ \mathbf{x}_2[n] &= \mathbf{H}_s\mathbf{s}_{s2}[n] + \mathbf{H}_f\mathbf{s}_{f2}[n] + \mathbf{n}_2[n], \quad n = 1, \cdots, N_2 . \end{aligned} \tag{3}$$

The available observations $\mathbf{x}_1[n]$ are collected in a data matrix $\mathbf{X}_1$, and likewise for $\mathbf{X}_2$. The target signal samples are collected in $\mathbf{S}_{s1}$ and $\mathbf{S}_{s2}$, respectively, and likewise the interference samples are collected in $\mathbf{S}_{f1}$ and $\mathbf{S}_{f2}$.

In general, $\mathbf{X}_1$ contains N_1 samples, $\mathbf{X}_2$ has N_2 samples, and $N_s = N_1 + N_2$ is the number of samples in the analysis window. For simplicity of notation, we will initially assume that both data matrices have the same number of samples $N = N_s/2$, and also that $N = N_p$, the number of samples in a data packet. The generalization follows in Sec. 7.

The corresponding sample covariance matrices of the observations are $\hat{\mathbf{R}}_1 = \frac{1}{N}\mathbf{X}_1\mathbf{X}_1^H$ and $\hat{\mathbf{R}}_2 = \frac{1}{N}\mathbf{X}_2\mathbf{X}_2^H$.

2.3. Covariance model

The target and interference signals are nonstationary, and this is reflected in the observations $\mathbf{x}_1[n]$ and $\mathbf{x}_2[n]$. For algorithm development, however, we will model the signals within a block as wide sense stationary (WSS) random vectors with the same second-order statistics. I.e., in the first data block we model $\mathbf{s}_{s1}[n]$ as a WSS complex random process with zero mean and a covariance matrix $\mathbf{R}_{s1}$ of size $d_s \times d_s$. This signal is uncorrelated to $\mathbf{s}_{f1}[n]$, which is zero mean with covariance matrix $\mathbf{R}_{f1}$ of size $d_f \times d_f$. Likewise, for the second data block we have uncorrelated WSS signals with covariance matrices $\mathbf{R}_{s2}$ and $\mathbf{R}_{f2}$.

The covariance matrices of the observations are then modeled by

$$\begin{aligned} \mathbf{R}_1 &= \mathbf{H}_s\mathbf{R}_{s1}\mathbf{H}_s^H + \mathbf{H}_f\mathbf{R}_{f1}\mathbf{H}_f^H + \mathbf{R}_{n1} \\ \mathbf{R}_2 &= \mathbf{H}_s\mathbf{R}_{s2}\mathbf{H}_s^H + \mathbf{H}_f\mathbf{R}_{f2}\mathbf{H}_f^H + \mathbf{R}_{n2} \end{aligned} \tag{4}$$

where $\mathbf{R}_{n1} = \mathbf{R}_{n2} = \sigma^2\mathbf{I}$.

The distinction between target signals and interference signals is defined by[1]

$$\mathbf{R}_{s1} > \mathbf{R}_{s2}, \qquad \mathbf{R}_{f1} < \mathbf{R}_{f2}. \tag{5}$$

(We will later refine this by introducing a scaling factor α.) This defines target signals as those signals that are stronger (larger in power) in the first data block than in the second data block. Similarly, in the second data block the interference signals are stronger than in the first data block.

[1] For matrices $\mathbf{A}, \mathbf{B}$ the notation "$\mathbf{A} > \mathbf{B}$" means that $\mathbf{A} - \mathbf{B}$ is positive definite.

2.4. Objective

The objective of the paper is, given the data matrix $\mathbf{X}$ and with $\mathbf{H}, \mathbf{S}$ unknown, to compute a single separating beamforming matrix $\mathbf{W}$ of size $M \times d_s$, such that

$$\hat{\mathbf{S}}_{s1} = \mathbf{W}^H\mathbf{X}_1, \qquad \hat{\mathbf{S}}_{s2} = \mathbf{W}^H\mathbf{X}_2$$

are estimates of the target signals in each block, with the interference signals suppressed. At this stage, we will not aim to separate the individual target signals but allow for an arbitrary linear combination $\mathbf{M}_s$, an unknown full rank matrix of size $d_s \times d_s$. Thus, we will aim for

$$\mathbf{W}^H\mathbf{H}_s = \mathbf{M}_s, \qquad \mathbf{W}^H\mathbf{H}_f = \mathbf{0}. \tag{6}$$

Further separation of the, almost fully overlapping, target signals can be achieved by exploiting other properties, such as constant modulus properties (the ACMA algorithm [12]) or similar algorithms based on higher-order statistics (JADE [13] and MUK [14]). Note that these algorithms include assumptions on stationarity, i.e., they do not work well on non-stationary (intermittent) signals. The algorithms proposed in this paper reduce the data to fit this assumption.

3. Tools from linear algebra

3.1. Singular value decomposition (SVD)

The "skinny" SVD [22] of a matrix $\mathbf{X} : M \times N$ with $M \le N$ is given by $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where $\mathbf{U} : M \times M$ is unitary, $\mathbf{V} : N \times M$ is semi-unitary ($\mathbf{V}^H\mathbf{V} = \mathbf{I}$), and $\mathbf{\Sigma} : M \times M$ is square and diagonal with nonnegative entries. For a given threshold $\epsilon \ge 0$, we can sort the singular values and partition the matrices as

$$\mathbf{X} = [\mathbf{U}_1 \quad \mathbf{U}_2] \begin{bmatrix} \mathbf{\Sigma}_1 & \\ & \mathbf{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix},$$

where $\mathbf{\Sigma}_1 > \epsilon\mathbf{I}$ and $\mathbf{\Sigma}_2 < \epsilon\mathbf{I}$. A low-rank approximation of $\mathbf{X}$ is then given by the truncated SVD (TSVD),

$$\hat{\mathbf{X}} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^H, \tag{7}$$

which is such that $\|\hat{\mathbf{X}} - \mathbf{X}\| < \epsilon$. The columns of $\mathbf{U}_1$ form a basis for the "signal subspace" $\mathrm{ran}(\hat{\mathbf{X}})$ of $\mathbf{X}$, with threshold ϵ. An orthogonal projection onto this subspace is $\mathbf{P} = \mathbf{U}_1\mathbf{U}_1^H$. The columns of $\mathbf{U}_2$ span what is commonly called the "noise subspace".

3.2. Generalized singular value decomposition (GSVD)

The GSVD [15] of two matrices $\mathbf{X}_1, \mathbf{X}_2$ (each of size $M \times N$ with $M \le N$) is denoted by

$$\mathrm{GSVD}(\mathbf{X}_1, \mathbf{X}_2) \quad \Leftrightarrow \quad \begin{cases} \mathbf{X}_1 = \mathbf{F}\mathbf{C}\mathbf{U}^H \\ \mathbf{X}_2 = \mathbf{F}\mathbf{S}\mathbf{V}^H \end{cases}$$

where $\mathbf{F} : M \times M$ is an invertible matrix, $\mathbf{C}$ and $\mathbf{S}$ are square positive diagonal matrices, and $\mathbf{U}$, $\mathbf{V}$ are semi-unitary matrices of size $N \times M$. Without loss of generality (and not following the usual convention), we scale each column of $\mathbf{F}$ to norm 1 by exchanging factors with $\mathbf{C}$ and $\mathbf{S}$.[2]

For a given threshold $\epsilon \ge 0$, we can further sort the entries of $\mathbf{C}$ and $\mathbf{S}$ and partition the matrices as

[2] In comparison to the usual definition of the GSVD and its Matlab implementation, a transpose has been introduced here. The norm scaling of $\mathbf{F}$ is also not standard but necessary to introduce a noise threshold ϵ.

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & & & \\ & \mathbf{C}_2 & & \\ & & \mathbf{C}_3 & \\ & & & \mathbf{C}_4 \end{bmatrix}, \ \mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & & & \\ & \mathbf{S}_2 & & \\ & & \mathbf{S}_3 & \\ & & & \mathbf{S}_4 \end{bmatrix}$$

where the partitioning is defined by

$$\begin{array}{llll} \mathbf{C}_1 > \epsilon\mathbf{I}, & \mathbf{S}_1 > \epsilon\mathbf{I}, & \mathbf{C}_3 < \epsilon\mathbf{I}, & \mathbf{S}_3 > \epsilon\mathbf{I} \\ \mathbf{C}_2 > \epsilon\mathbf{I}, & \mathbf{S}_2 < \epsilon\mathbf{I}, & \mathbf{C}_4 < \epsilon\mathbf{I}, & \mathbf{S}_4 < \epsilon\mathbf{I}. \end{array}$$

This enumerate all possibilities for entries of $\mathbf{C}$ and corresponding entries of $\mathbf{S}$ to be larger or smaller than ϵ. Depending on this, corresponding columns of $\mathbf{F}$ are dominantly present in $\mathbf{X}_1$ and/or $\mathbf{X}_2$. This provides information on the intersection of the column span of $\mathbf{X}_1$ with that of $\mathbf{X}_2$. Indeed, if we define a corresponding partitioning of $\mathbf{F}$ as

$$\mathbf{F} = [\mathbf{F}_1 \quad \mathbf{F}_2 \quad \mathbf{F}_3 \quad \mathbf{F}_4],$$

then the partitioning generates the following subspace information:

- ran($\mathbf{F}_1$) contains the common column span (with tolerance ϵ), i.e., ran($\mathbf{X}_1$) ∩ ran($\mathbf{X}_2$)
- ran($\mathbf{F}_2$) is the subspace of columns that are in ran($\mathbf{X}_1$) but not in ran($\mathbf{X}_2$),
- Similarly, ran($\mathbf{F}_3$) is the subspace of columns that are in ran($\mathbf{X}_2$) but not in ran($\mathbf{X}_1$),
- ran($\mathbf{F}_4$) is a common left null space, the "noise subspace".

Thus, the GSVD provides a general technique for subspace intersection, and this has been exploited in source separation applications, see e.g., [23,24]. Note that $\mathbf{F}$ is invertible but generally not unitary, so that these subspaces are not necessarily orthogonal to each other.

3.3. Generalized eigenvalue decomposition (GEVD)

If we "square" the data matrices in the previous subsection, we obtain a related decomposition for positive definite matrices $\mathbf{R}_1$, $\mathbf{R}_2$:

$$\text{GEVD}(\mathbf{R}_1, \mathbf{R}_2) \quad \Leftrightarrow \quad \begin{cases} \mathbf{R}_1 = \frac{1}{N}\mathbf{F}\mathbf{D}\mathbf{F}^H \\ \mathbf{R}_2 = \frac{1}{N}\mathbf{F}\mathbf{K}\mathbf{F}^H \end{cases}$$

where $\mathbf{F}$ is invertible and $\mathbf{D}$, $\mathbf{K}$ are diagonal and positive. If $\mathbf{R}_1 = \frac{1}{N}\mathbf{X}_1\mathbf{X}_1^H$ and $\mathbf{R}_2 = \frac{1}{N}\mathbf{X}_2\mathbf{X}_2^H$, then the GEVD is related to the GSVD via $\mathbf{D} = \mathbf{C}^2$ and $\mathbf{K} = \mathbf{S}^2$. We can define the same partitioning into four subspaces as for the GSVD. For symmetric matrices, the GEVD is sure to exist if $\mathbf{R}_1 > 0$ or $\mathbf{R}_2 > 0$, but otherwise the existence of the decomposition is unclear ($\mathbf{D}$ or $\mathbf{K}$ may become complex).

4. Separation by comparing two data blocks

Interference suppression by beamforming is related to an oblique projection, wherein the interference subspace is projected out while the target signal subspace is kept [19]. To compute the beamformer, we need to know both subspaces. In this section, we develop algorithms to estimate these subspaces.

Recall that the covariance model for $\mathbf{X}_1, \mathbf{X}_2$ is

$$\begin{aligned} \mathbf{R}_1 &= \mathbf{H}_s\mathbf{R}_{s1}\mathbf{H}_s^H + \mathbf{H}_f\mathbf{R}_{f1}\mathbf{H}_f^H + \mathbf{R}_{n1} \\ \mathbf{R}_2 &= \mathbf{H}_s\mathbf{R}_{s2}\mathbf{H}_s^H + \mathbf{H}_f\mathbf{R}_{f2}\mathbf{H}_f^H + \mathbf{R}_{n2}, \end{aligned} \tag{8}$$

where $[\mathbf{H}_s, \mathbf{H}_f]$ is of full column rank, with columns normalized to norm 1.

The distinction between target signals and interference signals is based on

$$\mathbf{R}_{s1} > \alpha^2\mathbf{R}_{s2}, \qquad \mathbf{R}_{f1} < \alpha^2\mathbf{R}_{f2}, \tag{9}$$

where in comparison to (5) we have introduced a parameter $\alpha > 1$ as a way to control the threshold on detecting a target. The choice of α is discussed later in Sec. 7. Consider first the noise-free case. In this case, we can write

$$\begin{aligned} \mathbf{R}_1 &= [\mathbf{H}_s \quad \mathbf{H}_f] \begin{bmatrix} \mathbf{R}_{s1} & \\ & \mathbf{R}_{f1} \end{bmatrix} \begin{bmatrix} \mathbf{H}_s^H \\ \mathbf{H}_f^H \end{bmatrix} \\ \mathbf{R}_2 &= [\mathbf{H}_s \quad \mathbf{H}_f] \begin{bmatrix} \mathbf{R}_{s2} & \\ & \mathbf{R}_{f2} \end{bmatrix} \begin{bmatrix} \mathbf{H}_s^H \\ \mathbf{H}_f^H \end{bmatrix}. \end{aligned} \tag{10}$$

We can compute a GSVD of $(\mathbf{X}_1, \mathbf{X}_2)$, or equivalently a GEVD of $(\mathbf{R}_1, \mathbf{R}_2)$. This leads to

$$\begin{cases} \mathbf{R}_1 = \frac{1}{N}\mathbf{F}\mathbf{D}\mathbf{F}^H \\ \mathbf{R}_2 = \frac{1}{N}\mathbf{F}\mathbf{K}\mathbf{F}^H. \end{cases} \tag{11}$$

For a given threshold $\epsilon \geq 0$, let us sort the generalized eigenvalues and partition $\mathbf{F}$, $\mathbf{D}$, $\mathbf{K}$ as $\mathbf{F} = [\mathbf{F}_1\ \mathbf{F}_2\ \mathbf{F}_3]$,

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & & \\ & \mathbf{D}_2 & \\ & & \mathbf{D}_3 \end{bmatrix}, \ \mathbf{K} = \begin{bmatrix} \mathbf{K}_1 & & \\ & \mathbf{K}_2 & \\ & & \mathbf{K}_3 \end{bmatrix} \tag{12}$$

where the partitioning is defined such that

$$\begin{array}{ll} \mathbf{D}_1 > \epsilon^2\mathbf{I} & \mathbf{K}_2 > \epsilon^2\mathbf{I} \\ \mathbf{D}_3 < \epsilon^2\mathbf{I} & \mathbf{K}_3 < \epsilon^2\mathbf{I} \end{array}$$

and moreover

$$\mathbf{D}_1 > \alpha^2\mathbf{K}_1, \qquad \mathbf{D}_2 < \alpha^2\mathbf{K}_2. \tag{13}$$

Comparing this decomposition to (9)–(10) and using the uniqueness of the GEVD, we immediately find

$$\text{ran}(\mathbf{F}_1) = \text{ran}(\mathbf{H}_s), \qquad \text{ran}(\mathbf{F}_2) = \text{ran}(\mathbf{H}_f),$$

while $\mathbf{F}_3$ spans the noise subspace. Thus, a GEVD of $(\mathbf{R}_1, \mathbf{R}_2)$, or equivalently the GSVD of $(\mathbf{X}_1, \mathbf{X}_2)$, gives directly the required subspace information. By using $\mathbf{F}$, we can construct a separating beamformer as

$$\mathbf{W}^H = [\mathbf{I} \quad \mathbf{0}][\mathbf{F}_1 \quad \mathbf{F}_2]^\dagger. \tag{14}$$

This beamformer is such that

$$\mathbf{W}^H[\mathbf{H}_s \ \mathbf{H}_f] = [\mathbf{M}_s \ \mathbf{0}]$$

where $\mathbf{M}_s$ is a square invertible matrix which depends on the selected basis for the subspaces ($\mathbf{F}$). Thus, we achieved our source separation objective stated in (6).

So far, we did not assume that the $\mathbf{R}_{si}$ and $\mathbf{R}_{fi}$ are diagonal. However, if they are diagonal (the signals are independent), and if a generalized eigenvalue $(\mathbf{D})_{ii}/(\mathbf{K})_{ii}$ is unique, then the corresponding vector $\mathbf{f}_i$ is a column of $\mathbf{H}_s$ or $\mathbf{H}_f$. If all ratios are unique, then $\mathbf{H}_s = \mathbf{F}_1$ and $\mathbf{H}_f = \mathbf{F}_2$ (up to permutations of the columns). In that case, the beamformer directly provides the individual signals. The technique to use the GEVD of two covariance matrices is reminiscent of the Second Order Blind Identification (SOBI) algorithm [25], which considers a rather different class of applications, namely to separate sources with differing temporal correlation structures, but arrives at similar properties of two (time-domain) covariance matrices.

Now we consider the case where both $\mathbf{X}_1$ and $\mathbf{X}_2$ are contaminated by white noise, and $\mathbf{R}_{n1} = \mathbf{R}_{n2} = \sigma^2\mathbf{I}$. If we compute the GSVD of $(\mathbf{X}_1, \mathbf{X}_2)$, or the GEVD of $(\mathbf{R}_1, \mathbf{R}_2)$, then $\mathbf{F}$ will change compared with the noiseless case. This is unlike the case of the SVD or EVD for a single matrix, where the addition of white noise will shift the singular values (eigenvalues) but leave the singular vectors (eigenvectors) intact. For significant noise powers, the matrix

Table 1
Source separation algorithm using GSVD.

Input: Data matrices $\mathbf{X}_1, \mathbf{X}_2$ each of size $M \times N$; the noise power σ^2; a scaling parameter $\alpha > 1$ that controls the signal classification (choice for α is discussed in Sec. 7)
Output: a separating beamformer $\mathbf{W}$.

1. *Rank reduction:* compute the SVD:

$$[\mathbf{X}_1 \;\; \mathbf{X}_2] = [\mathbf{U}_1 \;\; \mathbf{U}_2] \begin{bmatrix} \mathbf{\Sigma}_1 & \\ & \mathbf{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix},$$

where $\mathbf{\Sigma}_1 > \epsilon\mathbf{I}$ and $\mathbf{\Sigma}_2 < \epsilon\mathbf{I}$, and ϵ is a noise threshold, e.g., $\epsilon = \sigma(\sqrt{2N} + \sqrt{M})$ or slightly larger by a factor β as discussed in (B.2). Then apply a rank and dimension reduction:

$$\tilde{\mathbf{X}}_1 = \mathbf{U}_1^H \mathbf{X}_1, \qquad \tilde{\mathbf{X}}_2 = \mathbf{U}_1^H \mathbf{X}_2$$

2. Estimation of signal subspaces $\mathbf{F}$: Compute

$$\mathrm{GSVD}(\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) \quad \Rightarrow \quad \begin{cases} \tilde{\mathbf{X}}_1 = \mathbf{F}\mathbf{C}\mathbf{U}^H \\ \tilde{\mathbf{X}}_2 = \mathbf{F}\mathbf{S}\mathbf{V}^H \end{cases}$$

3. Using α, sort the entries of $\mathbf{D} = \mathbf{C}^2, \mathbf{K} = \mathbf{S}^2$ and correspondingly partition $\mathbf{F}$ according to (12)–(13).
4. The separating beamformer is

$$\mathbf{W}^H = [\mathbf{I} \;\; \mathbf{0}][\mathbf{F}_1 \;\; \mathbf{F}_2]^{-1} \mathbf{U}_1^H$$

$\mathbf{F}$ as computed in this way will tend to lose its information. Fortunately, for reasonable noise powers (SNR larger than 0 dB, say), the effect is not noticeable in simulations, and the algorithm performs well.

Nonetheless, to limit the effect of the noise, we propose to perform a preprocessing step wherein the noise subspace of $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$ is estimated via an SVD, and projected away. This removes the subspace corresponding to $\mathbf{F}_3$ (the noise subspace).

The proposed algorithm is specified in Table 1. In the algorithm, step 1) is the preprocessing (rank reduction, or projecting out the noise subspace) using a threshold ϵ on the singular values of $\mathbf{X}$ that separates the signal subspace from the noise subspace. This matrix contains $2N$ samples. Asymptotically, we could take $\epsilon = \sigma\sqrt{2N}$, but with finite samples, random matrix theory (see Appendix B) shows that the expected value of the largest singular value due to the noise is $\sigma(\sqrt{2N} + \sqrt{M})$. We should also allow for a variance on top of this. Thus, we suggest to set

$$\epsilon = \beta_1 \sigma(\sqrt{2N} + \sqrt{M}) \tag{15}$$

where the "Tracy–Widom" factor $\beta_1 > 1$ depends on M and N, and is derived in detail in Appendix B, equation (B.2). Typically, this is a small correction factor, e.g., $\beta_1 < 2$ for reasonable values of N.

After this preprocessing step, the term $\mathbf{F}_3$ in the subsequent GSVD step is absent as the noise subspace has been removed. This simplifies the algorithm at that stage (note that there is no literature on threshold selection for the noise subspace in GSVD and by preprocessing this complication is avoided).

In the algorithm, $\alpha > 1$ is an input parameter that controls the threshold on detecting a target (see Sec. 7).

5. Signed URV decomposition

We now enter into the second part of the paper, wherein we discuss a computationally efficient implementation of the algorithm, suitable for online applications.

The proposed algorithm in Table 1 is block-based and requires two steps. It detects the signals that are fully present in segment 1 of the analysis window (the "target signals"). However, in general we would like to detect all signals, over durations much longer than the analysis window. To do so, the analysis window is shifted over one or a few samples, and the detection process is repeated. It is obviously inefficient if we compute the decompositions from scratch. Instead, we would like to update the SVD, and subsequently the GSVD, to take advantage that most of the data in the matrices is the same. Unfortunately, the SVD is not amenable to sliding window updates. Also, the two-step process makes this even harder for the subsequent GSVD.

In the next two sections, we propose an algorithm that replaces the GSVD by a different subspace estimator. This estimator works directly on the data and is suitable to efficient subspace tracking for sliding-window updates. It is based on the Schur Subspace Estimator (SSE) introduced in [16,26] and its algorithm in [17]. After a brief introduction of this estimator, we derive its connection to the GSVD.

5.1. Definition of the SSE

For two given matrices $\mathbf{N} : M \times N_1$ and $\mathbf{X} : M \times N_2$, with $N_1 + N_2 \geq M$, the SSE is obtained from the decomposition

$$[\mathbf{N} \mid \mathbf{X}]\mathbf{\Theta} = [\mathbf{A} \;\; \mathbf{0} \mid \mathbf{B} \;\; \mathbf{0}] \tag{16}$$

where "|" denotes a matrix partitioning, $\mathbf{A} : M \times d_{\mathbf{A}}$, $\mathbf{B} : M \times d_{\mathbf{B}}$, and $\mathbf{\Theta}$ is a $\mathbf{J}$-unitary matrix:

$$\mathbf{\Theta}\mathbf{J}\mathbf{\Theta}^H = \mathbf{J}, \quad \mathbf{\Theta}^H\mathbf{J}\mathbf{\Theta} = \mathbf{J}, \quad \mathbf{J} = \left[\begin{array}{c|c} \mathbf{I}_{N_1} & \mathbf{0} \\ \hline \mathbf{0} & -\mathbf{I}_{N_2} \end{array}\right].$$

$\mathbf{J}$ is called a signature matrix. The partitioning of $\mathbf{J}$ (and hence of $\mathbf{\Theta}$) follows the partitioning indicated by "|" in (16), and we say that $\mathbf{N}$ and $\mathbf{A}$ have a positive signature, whereas $\mathbf{X}$ and $\mathbf{B}$ have a negative signature. The matrix $[\mathbf{A}, \mathbf{B}]$ is square or tall and has full column rank, and the $\mathbf{0}$-blocks augment $\mathbf{A}$ and $\mathbf{B}$ to N_1 and N_2 columns, respectively. After "$\mathbf{J}$-squaring" the data (i.e., computing $[\mathbf{A}, \mathbf{B}]\mathbf{J}[\mathbf{A}, \mathbf{B}]^H$), we obtain from (16)

$$\mathbf{N}\mathbf{N}^H - \mathbf{X}\mathbf{X}^H = \mathbf{A}\mathbf{A}^H - \mathbf{B}\mathbf{B}^H$$

and it is seen that $\mathbf{A}$ and $\mathbf{B}$ capture the positive and negative parts of $\mathbf{N}\mathbf{N}^H - \mathbf{X}\mathbf{X}^H$, respectively, using factors of minimal dimensions. We will assume that there is no "neutral subspace", i.e., $\mathbf{N}\mathbf{N}^H - \mathbf{X}\mathbf{X}^H$ is of full rank; in that case the dimensions of $\mathbf{A}, \mathbf{B}$ satisfy $d_{\mathbf{A}} + d_{\mathbf{B}} = M$, i.e., $[\mathbf{A}, \mathbf{B}]$ is a square matrix. Although the dimensions are well defined, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{\Theta}$ are not unique.

The computation of the decomposition is facilitated by introducing in (16) a QR-factorization of $[\mathbf{A}, \mathbf{B}]$:

$$[\mathbf{A} \quad \mathbf{B}] = \mathbf{Q}[\mathbf{R_A} \quad \mathbf{R_B}]$$

where $\mathbf{Q}$ is unitary and $[\mathbf{R_A}, \mathbf{R_B}]$ is lower triangular, giving rise to a two-sided decomposition of $[\mathbf{N}, \mathbf{X}]$ as

$$\mathbf{Q}^H[\mathbf{N} \,|\, \mathbf{X}]\boldsymbol{\Theta} = [\mathbf{R_A} \quad \mathbf{0} \,|\, \mathbf{R_B} \quad \mathbf{0}] \tag{17}$$

that always exists but is not unique. This "hyperbolic URV" decomposition is reminiscent of the URV [27], except that it involves a $\mathbf{J}$-unitary matrix $\boldsymbol{\Theta}$.

The "signed URV" (SURV) algorithm [17] is a numerically stable algorithm to compute (17), meanwhile posing some implicit structural constraints on $\boldsymbol{\Theta}$ that ensures its stability. The constraints ensure that

$$\mathbf{A}\mathbf{A}^H \le \mathbf{N}\mathbf{N}^H\,, \qquad \mathbf{B}\mathbf{B}^H \le \mathbf{X}\mathbf{X}^H\,. \tag{18}$$

This prevents the introduction of large common components in $\mathbf{A}$ and $\mathbf{B}$ that cancel each other in $\mathbf{A}\mathbf{A}^H - \mathbf{B}\mathbf{B}^H$. The SURV algorithm is non-iterative and easily updated for new columns of $\mathbf{N}$ and $\mathbf{X}$, with a complexity similar to that of a QR-update. Downdating (removing a column of $\mathbf{N}$ or $\mathbf{X}$) is elegantly achieved by updating $\mathbf{X}$ or $\mathbf{N}$, respectively, and is also numerically stable. This makes SURV very suitable for sliding window tracking of subspaces.

5.2. Properties

We list some properties of the SURV to reflect its potential for subspace tracking and its connection to the GSVD. First consider $\mathbf{N} = \epsilon\mathbf{I}$, where ϵ is a threshold, and introduce the SVD of $\mathbf{X}$ as

$$\mathbf{X} = \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{V}_1^H + \mathbf{U}_2\boldsymbol{\Sigma}_2\mathbf{V}_2^H$$

where $\boldsymbol{\Sigma}_1 > \epsilon\mathbf{I}$ and $\boldsymbol{\Sigma}_2 < \epsilon\mathbf{I}$. Assume that $\boldsymbol{\Sigma}_1$ has size $d \times d$ (i.e., $\mathbf{X}$ has d singular values larger than ϵ; we assume none are equal to ϵ). We compute the SURV

$$[\epsilon\mathbf{I} \,|\, \mathbf{X}]\boldsymbol{\Theta} = [\mathbf{A} \;\; \mathbf{0} \,|\, \mathbf{B} \;\; \mathbf{0}] \tag{19}$$

which corresponds to

$$\mathbf{X}\mathbf{X}^H - \epsilon^2\mathbf{I} = \mathbf{B}\mathbf{B}^H - \mathbf{A}\mathbf{A}^H\,. \tag{20}$$

It is seen that the spectrum of $\mathbf{X}\mathbf{X}^H$ is shifted by ϵ^2, which will introduce $M - d$ negative eigenvalues. Thus, $\mathbf{B}$ has d columns and $\mathbf{A}$ has $M - d$ columns. By means of this decomposition, it was shown in [16] that $(\mathbf{A}, \mathbf{B}, \boldsymbol{\Theta})$ can be used to parametrize all rank-d approximants $\hat{\mathbf{X}}$ such that

$$\|\mathbf{X} - \hat{\mathbf{X}}\| < \epsilon\,. \tag{21}$$

In particular, it was shown that the column span of any such $\hat{\mathbf{X}}$ is parametrized as $\mathrm{ran}(\mathbf{B}')$ with

$$\mathbf{B}' = \mathbf{B} - \mathbf{A}\mathbf{M}\,, \quad \|\mathbf{M}\| < 1 \tag{22}$$

where the matrix $\mathbf{M} : d_{\mathbf{A}} \times d_{\mathbf{B}}$ contains free parameters, reflecting the non-uniqueness in the decomposition (17).

It was shown in [17] that the TSVD (7) is a special case of an approximant satisfying (21), corresponding to an SURV decomposition with

$$\mathbf{B} = \mathbf{U}_1(\boldsymbol{\Sigma}_1^2 - \epsilon^2\mathbf{I})^{1/2}\,, \qquad \mathbf{A} = \mathbf{U}_2(\epsilon^2\mathbf{I} - \boldsymbol{\Sigma}_2^2)^{1/2} \tag{23}$$

and a specific $\boldsymbol{\Theta}$. In this case the column span of $\mathbf{B}$ is equal to the principal subspace obtained from the SVD of $\mathbf{X}$.

More in general, consider matrices $\mathbf{N}$, $\mathbf{X}$, each with N columns, and the SURV $[\mathbf{N} \,|\, \mathbf{X}]\boldsymbol{\Theta} = [\mathbf{A} \;\; \mathbf{0} \,|\, \mathbf{B} \;\; \mathbf{0}]$ with $\mathbf{N}$ such that $\mathbf{N}\mathbf{N}^H = \epsilon^2\mathbf{R}_n$. Then a direct generalization of (21) shows that all low-rank approximants $\hat{\mathbf{X}}$ such that

$$\|\mathbf{R}_n^{-1/2}(\mathbf{X} - \hat{\mathbf{X}})\| < \epsilon \tag{24}$$

have a column span parametrized by $\mathbf{B} - \mathbf{A}\mathbf{M}$, $\|\mathbf{M}\| < 1$. The factor $\mathbf{R}_n^{-1/2}$ can be interpreted as a prewhitening of the noise. Thus, properly weighted approximants of the form (24) are obtained without additional effort from the SURV decomposition.

Furthermore, we can relate the SURV in (16) to the GSVD as follows. Introduce the GSVD

$$\begin{cases} \mathbf{N} = \mathbf{F}\mathbf{C}\mathbf{U}^H = [\mathbf{F}_1 \quad \mathbf{F}_2]\begin{bmatrix} \mathbf{C}_1 & \\ & \mathbf{C}_2 \end{bmatrix}\mathbf{U}^H \\ \mathbf{X} = \mathbf{F}\mathbf{S}\mathbf{V}^H = [\mathbf{F}_1 \quad \mathbf{F}_2]\begin{bmatrix} \mathbf{S}_1 & \\ & \mathbf{S}_2 \end{bmatrix}\mathbf{V}^H \end{cases} \tag{25}$$

where the sorting and partitioning is such that $\mathbf{C}_1 > \mathbf{S}_1$ and $\mathbf{C}_2 < \mathbf{S}_2$ (for simplicity of notation, we assume there is no common null space: $\mathbf{F}_3$ is missing). Equivalently, we have the GEVD

$$\begin{cases} \mathbf{N}\mathbf{N}^H = \mathbf{F}\mathbf{D}\mathbf{F}^H = [\mathbf{F}_1 \quad \mathbf{F}_2]\begin{bmatrix} \mathbf{D}_1 & \\ & \mathbf{D}_2 \end{bmatrix}[\mathbf{F}_1 \quad \mathbf{F}_2]^H \\ \mathbf{X}\mathbf{X}^H = \mathbf{F}\mathbf{K}\mathbf{F}^H = [\mathbf{F}_1 \quad \mathbf{F}_2]\begin{bmatrix} \mathbf{K}_1 & \\ & \mathbf{K}_2 \end{bmatrix}[\mathbf{F}_1 \quad \mathbf{F}_2]^H \end{cases}$$

with a partitioning such that $\mathbf{D}_1 > \mathbf{K}_1$, $\mathbf{D}_2 < \mathbf{K}_2$. Then

$$\begin{aligned} \mathbf{N}\mathbf{N}^H - \mathbf{X}\mathbf{X}^H &= \mathbf{F}(\mathbf{D} - \mathbf{K})\mathbf{F}^H \\ &= \mathbf{F}_1(\mathbf{D}_1 - \mathbf{K}_1)\mathbf{F}_1^H - \mathbf{F}_2(\mathbf{K}_2 - \mathbf{D}_2)\mathbf{F}_2^H\,. \end{aligned} \tag{26}$$

But squaring (16) provides

$$\mathbf{N}\mathbf{N}^H - \mathbf{X}\mathbf{X}^H = \mathbf{A}\mathbf{A}^H - \mathbf{B}\mathbf{B}^H$$

so that we obtain

$$\begin{aligned} \mathbf{A} &= \mathbf{F}_1(\mathbf{D}_1 - \mathbf{K}_1)^{1/2} = \mathbf{F}_1(\mathbf{C}_1^2 - \mathbf{S}_1^2)^{1/2}\,, \\ \mathbf{B} &= \mathbf{F}_2(\mathbf{K}_2 - \mathbf{D}_2)^{1/2} = \mathbf{F}_2(\mathbf{S}_2^2 - \mathbf{C}_2^2)^{1/2}\,. \end{aligned} \tag{27}$$

as one possible solution to (16)—recall that this decomposition is not unique. This solution provides, in particular, $\mathrm{ran}(\mathbf{A}) = \mathrm{ran}(\mathbf{F}_1)$, $\mathrm{ran}(\mathbf{B}) = \mathrm{ran}(\mathbf{F}_2)$.

We have thus shown that the GSVD provides a special case of an SURV, where the pair $(\mathbf{F}_1, \mathbf{F}_2)$ in the GSVD is directly related to the pair $(\mathbf{A}, \mathbf{B})$ in the SURV—with coinciding subspaces. Conversely, we will propose in the next section to replace the GSVD in the source separation algorithm by the SURV algorithm, with the expectation that the resulting subspace estimates are close to that of the GSVD. Simulations that confirm this are in Sec. 8. The advantage of using SURV is that it is a simple non-iterative algorithm which is easily updated and thus enables sliding window tracking.

6. Source separation using SURV

We have a pair of data matrices $(\mathbf{X}_1, \mathbf{X}_2)$ each with N samples, with sample covariance matrices $(\hat{\mathbf{R}}_1, \hat{\mathbf{R}}_2)$ converging to covariance matrices $(\mathbf{R}_1, \mathbf{R}_2)$ that satisfy the data model (8). Instead of the GSVD or GEVD, we compute the SURV

$$[\mathbf{X}_1 \,|\, \alpha\mathbf{X}_2]\boldsymbol{\Theta} = [\mathbf{A} \quad \mathbf{0} \,|\, \mathbf{B} \quad \mathbf{0}] \tag{28}$$

where the factor $\alpha > 1$ controls the signal classification (see Sec. 7). Consider first the noiseless case, and assume that $[\mathbf{H}_s, \mathbf{H}_f]$ is square (M columns) and of full rank (this can be ensured by an initial rank reduction). Squaring (28) gives

$$\hat{\mathbf{R}}_1 - \alpha^2\hat{\mathbf{R}}_2 = \frac{1}{N}\left(\mathbf{A}\mathbf{A}^H - \mathbf{B}\mathbf{B}^H\right). \tag{29}$$

As discussed in the previous section, the GSVD or GEVD provides a special case of this decomposition (cf. (26)–(27)), i.e.,

$$\hat{\mathbf{R}}_1 - \alpha^2\hat{\mathbf{R}}_2 = \frac{1}{N}\left(\mathbf{F}_1(\mathbf{D}_1 - \alpha^2\mathbf{K}_1)\mathbf{F}_1^H - \mathbf{F}_2(\alpha^2\mathbf{K}_2 - \mathbf{D}_2)\mathbf{F}_2^H\right)$$

Table 2
Source separation algorithm using SURV.

Input: Data matrices $\mathbf{X}_1, \mathbf{X}_2$; the noise power σ^2; a scaling parameter $\alpha > 1$ that controls the signal classification (choice for α is discussed in Sec. 7)
Output: a separating beamformer $\mathbf{W}$.

1. Compute τ as in (33). Set $\gamma = \sqrt{N|\tau|}$. Compute one SURV:

   (if $\tau > 0$) $\mathbf{Q}^H[\gamma\mathbf{I} \quad \mathbf{X}_1 \mid \alpha\mathbf{X}_2]\mathbf{\Theta} = [\mathbf{R_A} \ \mathbf{0} | \mathbf{R_B} \ \mathbf{0}]$ (35)

   (if $\tau < 0$) $\mathbf{Q}^H[\mathbf{X}_1 \mid \gamma\mathbf{I} \quad \alpha\mathbf{X}_2]\mathbf{\Theta} = [\mathbf{R_A} \ \mathbf{0} | \mathbf{R_B} \ \mathbf{0}]$ (36)

2. Split $\mathbf{Q} = [\mathbf{Q_A} \ \mathbf{Q_B}]$. The separating beamformer is given by $\mathbf{W} = \mathbf{Q_A}$.

(the partitioning of $\mathbf{F}$ is such that $\mathbf{D}_1 - \alpha^2\mathbf{K}_1 > 0$ and $\alpha^2\mathbf{K}_2 - \mathbf{D}_2 > 0$) such that $\mathrm{ran}(\mathbf{A}) = \mathrm{ran}(\mathbf{F}_1)$ and $\mathrm{ran}(\mathbf{B}) = \mathrm{ran}(\mathbf{F}_2)$. On the other hand, the covariance model provides

$$\mathbf{R}_1 - \alpha^2\mathbf{R}_2 = \mathbf{H}_s(\mathbf{R}_{s1} - \alpha^2\mathbf{R}_{s2})\mathbf{H}_s^H - \mathbf{H}_f(\alpha^2\mathbf{R}_{f2} - \mathbf{R}_{f1})\mathbf{H}_f^H .$$

Asymptotically (for sufficiently large N such that $\hat{\mathbf{R}}_1 \to \mathbf{R}_1$ and $\hat{\mathbf{R}}_2 \to \mathbf{R}_2$), we have that $\mathrm{ran}(\mathbf{F}_1) = \mathrm{ran}(\mathbf{H}_s)$ and $\mathrm{ran}(\mathbf{F}_2) = \mathrm{ran}(\mathbf{H}_f)$. Generally, for finite N, we therefore have

$$\begin{cases} \mathrm{ran}(\mathbf{A}) \approx \mathrm{ran}(\mathbf{H}_s) \\ \mathrm{ran}(\mathbf{B}) \approx \mathrm{ran}(\mathbf{H}_f) \end{cases}$$

Thus, the decomposition SURV$(\mathbf{X}_1, \alpha\mathbf{X}_2)$ directly gives estimates of the subspaces required for constructing a separating beamformer $\mathbf{W}$.

Consider now the case with noise, where we assume $\mathbf{R}_{n1} = \mathbf{R}_{n2} = \sigma^2\mathbf{I}$. The GSVD of $(\mathbf{X}_1, \mathbf{X}_2)$, or equivalently the GEVD of $(\hat{\mathbf{R}}_1, \hat{\mathbf{R}}_2)$, and its data model are

$$\hat{\mathbf{R}}_1 = \tfrac{1}{N}\mathbf{F}\mathbf{D}\mathbf{F}^H , \quad \mathbf{R}_1 = \mathbf{H}_s\mathbf{R}_{s1}\mathbf{H}_s^H + \mathbf{H}_f\mathbf{R}_{f1}\mathbf{H}_f^H + \sigma^2\mathbf{I}$$
$$\hat{\mathbf{R}}_2 = \tfrac{1}{N}\mathbf{F}\mathbf{K}\mathbf{F}^H , \quad \mathbf{R}_2 = \mathbf{H}_s\mathbf{R}_{s2}\mathbf{H}_s^H + \mathbf{H}_f\mathbf{R}_{f2}\mathbf{H}_f^H + \sigma^2\mathbf{I}$$

(the parameter α defines the partitioning of $\mathbf{F}$), and

$$\begin{aligned}\mathbf{R}_1 - \alpha^2\mathbf{R}_2 \ &= \mathbf{H}_s(\mathbf{R}_{s1} - \alpha^2\mathbf{R}_{s2})\mathbf{H}_s^H - \mathbf{H}_f(\alpha^2\mathbf{R}_{f2} - \mathbf{R}_{f1})\mathbf{H}_f^H \\ &\quad + (1-\alpha^2)\sigma^2\mathbf{I} \end{aligned} \tag{30}$$

The presence of $(1-\alpha^2)\sigma^2\mathbf{I}$ will tend to disturb the decomposition into "positive" and "negative" subspaces. However, if we consider $\alpha = 1$ and look at $\mathbf{R}_1 - \mathbf{R}_2$, then the noise covariance matrices will cancel each other in (30). In this case, SURV$(\mathbf{X}_1, \alpha\mathbf{X}_2)$ produces (asymptotically) the same result as in the noiseless case.

For $\alpha > 1$, the noise term $(1-\alpha^2)\sigma^2\mathbf{I}$ in (30) is not zero. Although the SURV decomposition will still give reasonable results, we expect improved results if we compensate the noise power by adding a term $\tau\mathbf{I}$ to $\hat{\mathbf{R}}_1 - \alpha^2\hat{\mathbf{R}}_2$. The resulting data model for $\mathbf{R}_1 - \alpha^2\mathbf{R}_2 + \tau\mathbf{I}$ has a noise power term $(1-\alpha^2)\sigma^2\mathbf{I} + \tau\mathbf{I}$. We obtain noise power cancellation if we choose

$$\tau = \sigma^2(\alpha^2 - 1) . \tag{31}$$

Note that $\tau > 0$. Adding $\tau\mathbf{I}$ can be achieved quite simply: instead of (28), compute the SURV

$$[\gamma\mathbf{I} \quad \mathbf{X}_1 \mid \alpha\mathbf{X}_2]\mathbf{\Theta} = [\mathbf{A} \ \mathbf{0} \mid \mathbf{B} \ \mathbf{0}] . \tag{32}$$

If we set $\gamma = \sqrt{N\tau}$, then it follows that

$$\hat{\mathbf{R}}_1 - \alpha^2\hat{\mathbf{R}}_2 + \frac{\gamma^2}{N}\mathbf{I} = \hat{\mathbf{R}}_1 - \alpha^2\hat{\mathbf{R}}_2 + \tau\mathbf{I} = \frac{1}{N}(\mathbf{A}\mathbf{A}^H - \mathbf{B}\mathbf{B}^H)$$

and we obtain noise power compensation, asymptotically.

Several refinements are in order. For finite samples, the resulting noise power $\hat{\mathbf{R}}_{n1} - \alpha^2\hat{\mathbf{R}}_{n2} + \tau\mathbf{I}$ will not be exactly zero, but some indefinite matrix. For optimal accuracy of $\mathbf{A}$, it is better to ensure that this matrix is small but negative definite. As shown in Appendix C, this is achieved by setting

$$\tau = \sigma^2\Big(\alpha^2(1 - \frac{\sqrt{M}}{\sqrt{N}})^2 - (1 + \frac{\sqrt{M}}{\sqrt{N}})^2\Big) \tag{33}$$

which replaces (31) by a slightly smaller τ. If $\tau > 0$, then we set $\gamma = \sqrt{N\tau}$ and compute the SURV in (32). If α or N is small, then it may happen that $\tau < 0$. In that case we set $\gamma = \sqrt{N|\tau|}$ and instead of (32) compute the slightly different SURV

$$[\mathbf{X}_1 \mid \gamma\mathbf{I} \quad \alpha\mathbf{X}_2]\mathbf{\Theta} = [\mathbf{A} \ \mathbf{0} \mid \mathbf{B} \ \mathbf{0}] \tag{34}$$

By squaring this equation, it is seen that we obtain

$$\hat{\mathbf{R}}_1 - \alpha^2\hat{\mathbf{R}}_2 - \frac{\gamma^2}{N}\mathbf{I} = \hat{\mathbf{R}}_1 - \alpha^2\hat{\mathbf{R}}_2 + \tau\mathbf{I} = \frac{1}{N}(\mathbf{A}\mathbf{A}^H - \mathbf{B}\mathbf{B}^H)$$

which is the required expression for noise compensation. Appendix C also lists refinements for matrices $\mathbf{X}_1, \mathbf{X}_2$ of unequal dimensions, and a more accurate result using Tracy–Widom factors.

In the original GSVD algorithm in Table 1, we introduced an initial rank-reduction step as the first step. This required an initial SVD of $[\mathbf{X}_1, \mathbf{X}_2]$. Although this SVD can be replaced by an SURV, this initial rank reduction can also be omitted, since we ensured that the noise components cancel each other implicitly while computing the SURV (32) or (34).

Further, if we consider that $[\mathbf{A}, \mathbf{B}] = \mathbf{Q}[\mathbf{R_A}, \mathbf{R_B}]$ and use the fact that $[\mathbf{R_A}, \mathbf{R_B}]$ is lower triangular, we see that $\mathrm{ran}(\mathbf{B}) = \mathrm{ran}(\mathbf{Q_B})$. This is an estimate of the interference subspace $\mathcal{H}_f$ and the noise subspace. The column span of $\mathbf{Q_A}$ is orthogonal to these subspaces. A beamformer which projects out the interference and noise subspace is simply $\mathbf{W}^H = \mathbf{Q}_\mathbf{A}^H$.

The resulting algorithm is shown in Table 2. Compared to the GSVD-based algorithm, this algorithm is particularly simple, and is suitable for tracking as the single SURV is easily used in a sliding window update setting [17].

7. Considerations for a nonstationary data model

We derived algorithms for a block-stationary data model, but apply the algorithms to nonstationary data consisting of intermittent signals. Moreover, in general $\mathbf{X}_1$ and $\mathbf{X}_2$ have an unequal number of samples N_1 and N_2. In this section, we verify the validity of the algorithms for the more general nonstationary case and discuss the choice of the detection parameter α.

As shown in Appendix D, for nonstationary signals the sample correlation matrix converges faster to its model than it would for corresponding stationary signals. This was shown for the correlation matrix of two signals that are partly zero, but the same holds for the more elaborate model $(\mathbf{X}_1, \mathbf{X}_2)$ of mixed intermittent signals in noise. Thus, the derived algorithms are valid and will perhaps even have better performance for finite samples because of the faster convergence.

For the proper selection of α, define $n_1^{(k)}$ and $n_2^{(k)}$ to be the number of nonzero samples of the kth source (packet) within $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. We detect a target signal if $\mathbf{R}_{s1} > \alpha^2 \mathbf{R}_{s2}$. Assuming independent sources, $\mathbf{R}_{s1}$ and $\mathbf{R}_{s2}$ are diagonal, and it is seen that the power (amplitude) of a source does not play a role, but only the ratio of its number of nonzero samples in block $\mathbf{X}_1$ and $\mathbf{X}_2$. In particular, the kth source is considered a target signal if $n_1^{(k)} > \alpha^2 n_2^{(k)}$. To design α, first choose a parameter n_1 that defines a target packet to have more than n_1 nonzero samples in $\mathbf{X}_1$ (and fewer than $n_2 = N_p - n_1$ samples in $\mathbf{X}_2$). Then set

$$\alpha^2 = \frac{n_1}{n_2}\,. \tag{37}$$

This choice for α is suitable for a large number of samples such that the cross-correlations among the independent sources have diminished.

With a finite number of samples, α plays the role of a detection threshold parameter with regard to the definition of a target packet (defined by n_1, n_2), as its choice determines a certain probability of detection and false alarm. We will not attempt to compute these probabilities as function of α and the number of samples. Instead, we will propose a suitable α such that the probability of false alarm is zero. Assume, in a worst-case scenario, that there are M packets at a marginal position, i.e., with $n_1^{(k)} = n_1$ and $n_2^{(k)} = n_2$. Further assume without loss of generality that the sources have equal unit power, and assume for convenience that all sources have a Gaussian distribution. If for this scenario the packets have to be classified as interferers (i.e., no false alarm), then we require

$$n_1 \hat{\lambda}_1 < \alpha^2 n_2 \hat{\lambda}_M\,,$$

where $\hat{\lambda}_1$ is the largest eigenvalue of a sample covariance matrix generated by n_1 i.i.d. complex white Gaussian noise samples, and $\hat{\lambda}_M$ is the smallest eigenvalue of such a matrix, now for n_2 samples. Results from random matrix theory listed in Appendix B allow to approximate these eigenvalues, and it follows that we have to set

$$\alpha^2 > \alpha_{\min}, \qquad \alpha_{\min} := \frac{\beta_1^2(n_1)}{\beta_M^2(n_2)} \left(\frac{\sqrt{n_1} + \sqrt{M}}{\sqrt{n_2} - \sqrt{M}} \right)^2 \tag{38}$$

where $\beta_1(n)$ and $\beta_M(n)$ are Tracy–Widom factors defined in (B.2), (B.3). This generalizes (37) for a finite number of samples.

Finally, if the two matrices $\mathbf{X}_1, \mathbf{X}_2$ of the two blocks have different lengths N_1, N_2, the noise power threshold ϵ in (15) and similar equations should use $N_1 + N_2 = N_s$ in place of $2N$. The noise power shift parameter γ in (33) also needs to be adjusted. Appendix C lists the appropriate generalizations.

8. Simulation results

8.1. Data model with stationary sub-blocks

To show the algorithms on the data model for which they have been derived, we first show their performance on stationary data matrices, each of size $M \times N$. Specifically, we generate data according to

$$\begin{aligned} \mathbf{X}_1 &= \mathbf{H}_s \mathbf{S}_{s1} + \sqrt{r}\mathbf{H}_f \mathbf{S}_{f1} + \mathbf{N}_1 \\ \mathbf{X}_2 &= \sqrt{r}\mathbf{H}_s \mathbf{S}_{s2} + \mathbf{H}_f \mathbf{S}_{f2} + \mathbf{N}_2 \end{aligned} \tag{39}$$

where the signal and interference data matrices $\mathbf{S}_{s1}$, $\mathbf{S}_{s2}$, $\mathbf{S}_{f1}$, and $\mathbf{S}_{f1}$ correspond to unit-power QPSK sources with symbols taken from the alphabet $\{+1, +j, -1, -j\}$. The "mixing" parameter r ($0 \le r < 1$) controls the signal to interference ratio (SIR), defined by $-10\log_{10}(r)$ dB. The algorithms are tested using $\alpha = 1$, which determines the classification threshold: a signal is classified as "target" if it is more present in segment 1 than in segment 2.

To simplify the comparison, we consider only one signal and one interference in (39). We generate the columns of $\mathbf{H}$ as the array response vectors of a uniform linear antenna array consisting of $M = 4$ elements spaced at half wavelength. The target signal direction of arrival (DOA) is fixed at 0°. The interference DOA is set at 20° unless specified otherwise. The noise matrices are generated from i.i.d. complex Gaussian sources with variance σ^2. The SNR is defined as $\text{SNR} = 10\log_{10}(1/\sigma^2)$. The SNR is set at 15 dB unless specified otherwise. We test the algorithms based on *GSVD* as shown in Table 1 and on *SURV* as in Table 2.

Instead of looking at beamformers, we consider the underlying estimates for $\text{ran}(\mathbf{H}_s)$ and $\text{ran}(\mathbf{H}_f)$ for each algorithm. As performance measure we consider the subspace error $\mathcal{E}(\mathbf{U}, \mathbf{U}_e)$, defined for an orthonormal basis $\mathbf{U}$ for the "true" subspace and an orthonormal basis $\mathbf{U}_e$ for its estimate, as

$$\mathcal{E}(\mathbf{U}, \mathbf{U}_e) = \|(\mathbf{I} - \mathbf{U}_e \mathbf{U}_e^H)\mathbf{U}\|\,.$$

In Fig. 2(a), the subspace error of the algorithms is shown as function of the mixing factor r at SNR = 15 dB. There is not much difference between the algorithms. It is seen that the performance of both algorithms drops when the mixing factor r goes towards 1 (SIR = 0 dB). At this marginal case, we cannot expect to distinguish the subspaces.

Fig. 2(b) shows the performance for varying separation in DOA between both sources, for $r = 0.5$ (SIR = 3 dB). The target DOA is fixed at 0° and the interference DOA goes from 0° to 90°. Both algorithms give the same performance.

Fig. 2(c) shows the performance as function of N, for $r = 0.5$.

Fig. 2(d) shows the performance as function of SNR, for $r = 0.5$ (SIR = 3 dB). Unless the SNR is small, the performance of the algorithms is dominated by the non-zero finite sample cross-correlations of the sources, and essentially independent of the SNR. (We did not consider negative SNRs as the resulting signal estimates would not be meaningful in applications.)

Overall, we can conclude that essentially there is no performance difference between the two algorithms in the block-stationary data model (39).

8.2. Data model with partially overlapping data packets

We now compare the proposed algorithms for separating partially overlapping data packets, as shown in Fig. 1. All packets have the same length $N_p = 256$ symbols. The length of the analysis window is set to $N_s = 3N_p$, where the length of $\mathbf{X}_1$ is $N_1 = N_p$ and the length of $\mathbf{X}_2$ is $N_2 = 2N_p$.

The source data are QPSK symbols. A uniform linear antenna array of $M = 5$ elements spaced at half wavelength is used. We consider up to 2 target packets with DOAs $[-10°, 40°]$, and 3 partially overlapping interference packets with DOAs $[5°, -30°, -60°]$.

We generate packets with arbitrary arrival times. To describe the extent of overlap, let $n_1^{(k)}$ be the number of nonzero samples of the kth packet in block $\mathbf{X}_1$, and define the "overlapping ratio" r_k as $r_k = \frac{n_1^{(k)}}{N_p}$. We generate target packets with r_k randomly selected in the interval [0.85, 1], and interference packets with $r_k \in [0, 0.5]$.

We compare the interference suppression algorithms based on *GSVD* (Table 1) and on *SURV* (Table 2). We consider cases with $d_s = 1$ and $d_s = 2$ target sources. In the latter case, a postprocessing follows to separate the remaining mixture of the two target signals into two individual target signals using ACMA [12]. To ensure that ACMA acts only on the stationary part of these signals, we process only the central 100 samples of the mixture.

For reference, we show as an upper bound the performance of the standard (non-blind) linear minimum mean squared error (LMMSE) receiver obtained when all source data is known. As a lower bound, we show the performance of directly applying ACMA

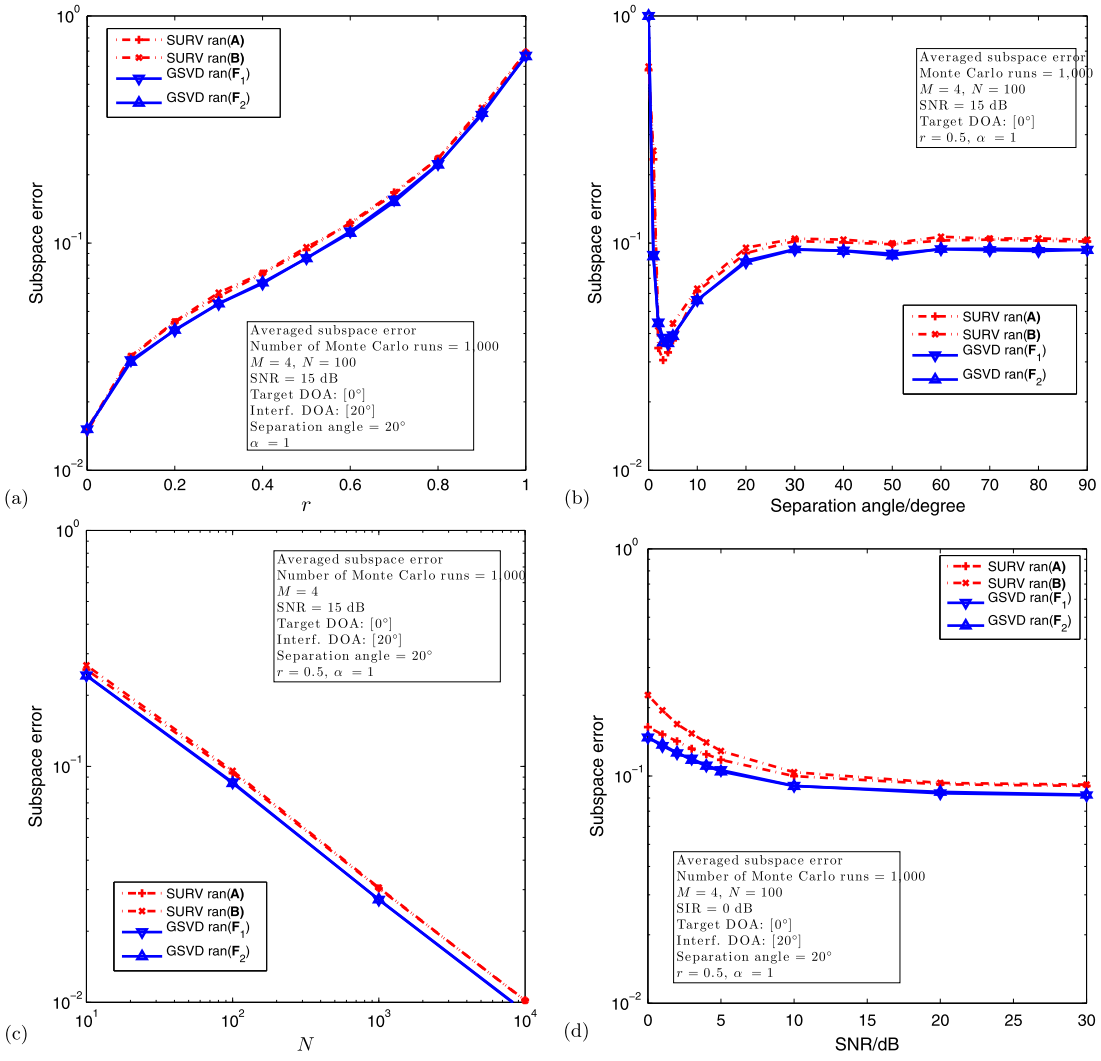

Fig. 2. Separation performance for block-stationary data model in terms of the subspace error $\mathcal{E}$. (a) Varying SIR (parameter r) for SNR = 15 dB; (b) varying separation angle for $r = 0.5$; (c) varying N for $r = 0.5$; (d) varying SNR for $r = 0.5$.

on the central 100 samples of $\mathbf{X}_1$, without interference suppression (denoted by ACMA*). We also show the performance of JADE, acting on the central 100 samples (denoted by JADE*).

For further comparison, we show the performance of the Algebraic Zero/Constant Modulus Algorithm (AZCMA [28,11]), which separates signals with samples $s[n]$ that are either zero or constant modulus, i.e., $s(|s|^2 - 1) = 0$. By inserting $s[n] = \mathbf{w}^H\mathbf{x}[n]$ into this equation, it effectively uses 6-th order statistics of the data. It is known that this algorithm breaks down if two sources are fully non-overlapping or, on the other hand, do not contain zero entries. The output of the algorithm can be post-processed by ZCMA [28], which is a block-iterative algorithm similar to LS-CMA.[3]

The performance measure is the residual signal-to-interference-plus-noise ratio (SINR) at the output of the beamformers. Here, the output SINR is defined as the worst SINR among the d_s output signals, after we make the best assignment of the estimated beamformers to target source indices.

The values of the various parameters ϵ, γ and α in the algorithms are set according to Sec. 7 and Appendix C.

Fig. 3 shows the output SINR of the beamformers for SNR = 15 dB, for $d_s = 1$ and $d_s = 2$ target sources, respectively. Both GSVD and SURV were able to correctly detect the presence of the target signals in all cases. Their performance is very similar, nearly independent on the number of target signals, and very close to that of the non-blind LMMSE receiver.

It is also seen that ACMA and JADE are not performing well, as they are not designed for non-stationary sources. AZCMA performs reasonable for $d_s = 1$ target and high SIR, but fails for low SIR due to the presence of fully non-overlapping interfering sources. It also fails for multiple (fully overlapping) target sources. The results regarding JADE and AZCMA are consistent with what was already reported in [11].

In summary, it is seen that the two proposed algorithms have a rather similar performance, close to that of the (informed) LMMSE. Note that SURV has a lower complexity and admits an attractive sliding window implementation.

9. Example experiment

We have set up a demonstration system for testing the proposed algorithms in a real application: the automatic identification system for ships. AIS data is used for exchanging navigational in-

[3] No other algorithms seem suitable for comparison in the presented scenario. The papers [20,21] consider a similar scenario but the proposed algorithms are insufficiently general as they assume that for each signal there is an interval where that signal is the only source present. The papers [29,30] consider semi-blind algorithms for a single target signal containing known training symbols.

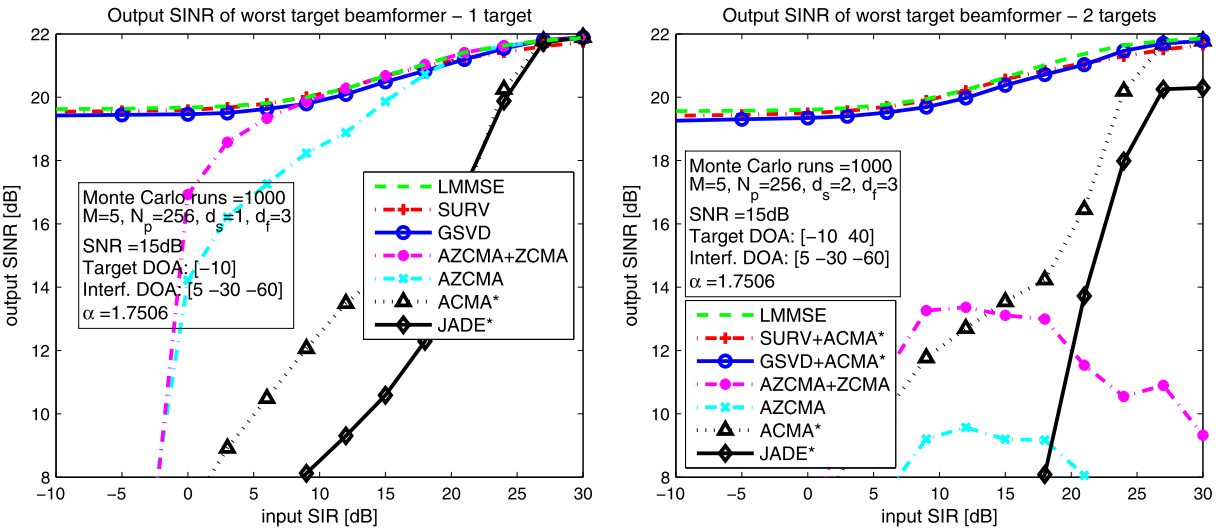

Fig. 3. Separation performance of the proposed algorithms. SINR as function of SIR for SNR = 15 dB, 3 interfering sources, and (*a*) 1 target source, (*b*) 2 target sources. * signifies algorithms that act on a subset of the data (central 100 samples).

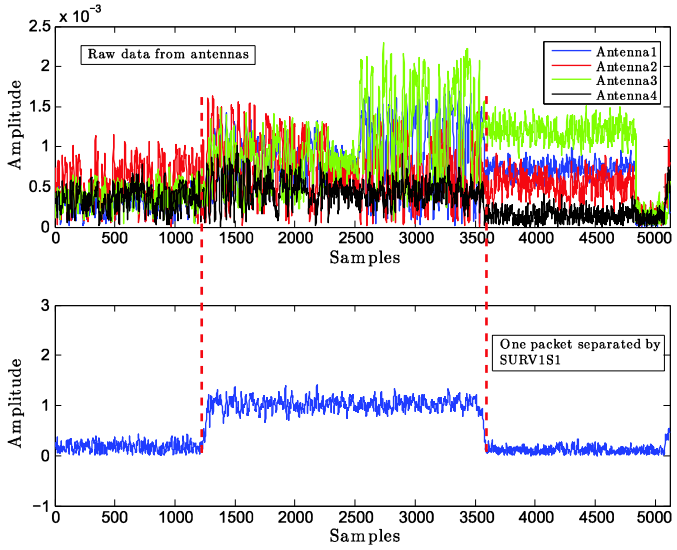

Fig. 4. Example of the proposed source separation algorithm on experimental AIS data. The upper panel shows the amplitude of the raw data from the 4 antennas in one analysis window. The lower panel shows the amplitude of one packet separated by SURV.

formation among ships and between ships and base stations in the maritime frequency band at 162 MHz. Most AIS data packets are 256 symbols long, Gaussian minimum shift keying (GMSK) modulated, and transmitted at a rate of 9.6 kbps. We use an uncalibrated receiver array consisting of $M = 4$ antennas to collect the baseband data samples. The antennas are roughly positioned at half wavelength spacing. The RF front-ends for the 4 antennas are not identical and uncalibrated. The receiver array is placed at an altitude of 68 meters in the window of a building in Delft, facing the extended harbor and coastline of Rotterdam, about 10 nautical miles away. The receiver is behind two layers of window glass, and next to the metal frame of the building, i.e., multipath is expected to be present. Although AIS works in a synchronized network and no packet collision is tolerated in its protocol, the actual AIS packets from different communication cells are frequently seen overlapping by our receiver.

Fig. 4 shows in the top panel as example the amplitude of raw samples from the 4 antennas. The shown analysis window contains at least three overlapping data packets. The bottom panel shows one packet after separation from the other two partially overlapping packets. The length of the analysis window is $N_s = 2N_p = 512$ symbols. Data was sampled using an oversampling ratio of 10, so that the number of samples in one analysis window is 5120. The beamformers are computed after downsampling by a factor of 10. The upper panel in Fig. 4 shows that the antennas have different responses to the same packet, which is mainly caused by multipath fading. In the shown example, one interference packet overlaps the head of the middle packet while another packet overlaps the tail. The middle packet is the target packet in this analysis window. The input SIR is below 0 dB and the output SNR is around 10 dB. The lower panel in Fig. 4 shows the amplitude of one packet separated by SURV. Without the proposed algorithms, it is even difficult to visually identify the start and end of the separated packet.

Experience with this demonstrator shows that the separation performance is good; the algorithms are reliable and robust, and messages with good SNR are easily decoded after separation. Some related results including tracking are presented in [4]; more results are reported in [5].

10. Conclusions

In this paper, we proposed two blind beamforming algorithms for suppressing asynchronous co-channel interference. The first algorithm was based on subspace estimations using GSVD. We subsequently introduced the SURV as an elegant and computationally efficient replacement of the GSVD. Simulations showed that these algorithms have essentially equal performance, close to that of a reference MMSE receiver with completely known target signals. For high SNR, the performance of the algorithms is limited by the assumption that sources are uncorrelated, while for a small number of samples the empirical cross-correlation is not yet zero.

The algorithm was also demonstrated on experimental data from a real application (AIS), which confirmed the effectiveness and robustness of the proposed algorithms.

Future research may include the design of tracking algorithms dedicated to specific applications, and extensions to wideband sources.

Appendix A. Scenario with continuously present target signals

In this Appendix, we consider a second scenario (see Fig. A.5), where we assume that a target signal fills the complete analysis window. This can be used to model target signals that transmit

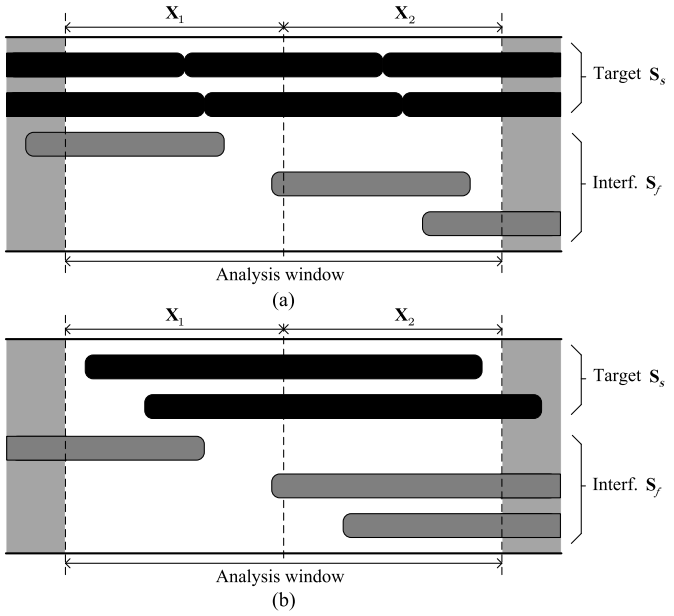

Fig. A.5. Scenario 2: target signals are equally present in both data blocks, whereas each interference signal is concentrated in only one of these blocks. (a) Target sources transmit continuously; (b) the analysis window equals the length of a packet.

continuously, or short analysis windows that may be even shorter than the duration of a data packet. As shown in Fig. A.5, we can split the window into two consecutive blocks, where the assumption is that each target signal is equally present in each block, while each interference signal is concentrated mostly in one of the blocks, but not equally strong in both. The data samples of each block are collected into two corresponding matrices $\mathbf{X}_1$ and $\mathbf{X}_2$, that have the same data model (3) as before. Each data block consists of N samples. The covariance model is the same as in (4).

However, we now define that for target signals the signal power is equal in each block ($\mathbf{R}_{s1} \approx \mathbf{R}_{s2}$), whereas for interference signals the power is unequal ($\mathbf{R}_{f1} \neq \mathbf{R}_{f2}$). To formulate this more precisely, we introduce again a parameter α that controls how "equal" two values of signal power should be to classify as a target signal. Specifically, we define that a target signal satisfies

$$\frac{1}{\alpha^2} < \frac{(\mathbf{R}_{s1})_{ii}}{(\mathbf{R}_{s2})_{ii}} < \alpha^2 \tag{A.1}$$

where $\alpha > 1$.

The GEVD of $(\mathbf{R}_1, \mathbf{R}_2)$ is defined in the same way as before in (11), but the partitioning of the generalized eigenvalues is now based on

$$\frac{1}{\alpha^2} < \frac{(\mathbf{D}_1)_{ii}}{(\mathbf{K}_1)_{ii}} < \alpha^2 \tag{A.2}$$

whereas $\mathbf{D}_2, \mathbf{K}_2$ collect the diagonal entries that do not satisfy any of these conditions. This replaces (13).

Comparing (A.2) to (A.1), we see that $\mathbf{F}_1$ corresponds to the target signal subspace and $\mathbf{F}_2$ to the interference signal subspace. The resulting algorithm is the same as for Scenario 1 (see Table 1), except that now the partitioning rule (A.2) is used. This algorithm could also be implemented using SURV decompositions.

The parameter α is used to tune the detection of a target signal, i.e., how "equal" the power in both blocks has to be to detect a target signal. If we take a finite number of samples into account, then following similar considerations as in Sec. 7, we obtain the two conditions (replacing (38))

$$\alpha^2 > \frac{\beta_1^2(n_2)}{\beta_M^2(n_1)} \left(\frac{\sqrt{n_2}+\sqrt{M}}{\sqrt{n_1}-\sqrt{M}}\right)^2$$
$$\alpha^2 > \frac{\beta_1^2(n_1)}{\beta_M^2(n_2)} \left(\frac{\sqrt{n_1}+\sqrt{M}}{\sqrt{n_2}-\sqrt{M}}\right)^2.$$

The larger one of these lower bounds on α should be used.

Appendix B. Bounds on eigenvalues of random matrices

The following results from random matrix theory are needed in Appendix C. We consider a random complex white Gaussian noise variable $\mathbf{n}[k]$ with covariance matrix $\mathbf{R}_n = \sigma^2\mathbf{I}$. If we have N samples, the sample covariance matrix is $\hat{\mathbf{R}}_n = \frac{1}{N}\sum_{k=1}^{N} \mathbf{n}[k]\mathbf{n}[k]^H$. The largest eigenvalue of $\hat{\mathbf{R}}_n$ is denoted by $\hat{\lambda}_1$ and the smallest by $\hat{\lambda}_M$.

For $N \to \infty$, it follows from Bai and Yin [31] that (almost sure convergence)

$$\begin{aligned} \hat{\lambda}_1 &\to \sigma^2\rho_1^2, \qquad \rho_1 := 1 + \tfrac{\sqrt{M}}{\sqrt{N}}, \\ \hat{\lambda}_M &\to \sigma^2\rho_M^2, \qquad \rho_M := 1 - \tfrac{\sqrt{M}}{\sqrt{N}}. \end{aligned} \tag{B.1}$$

These are equal to the expected values of $\hat{\lambda}_1$ and $\hat{\lambda}_M$. If more accurate upper bounds are needed, we can employ the known distributions of the eigenvalues, as follows. Define centering and scaling constants as

$$\begin{aligned} \mu_1 &= N\rho_1^2, \qquad \nu_1 = N^{1/2}M^{-1/6}\rho_1^{4/3} \\ \mu_M &= N\rho_M^2, \qquad \nu_M = N^{1/2}M^{-1/6}\rho_M^{4/3} \end{aligned}$$

Under some conditions, it was shown in [32] that $w_1 := \frac{N\hat{\lambda}_1/\sigma^2 - \mu_1}{\nu_1}$ converges to the Tracy–Widom distribution of order 2, $w_1 \sim F_2(s)$ [33]. Similarly, it was shown in [34] that $w_M := \frac{N\hat{\lambda}_M/\sigma^2 - \mu_M}{\nu_M}$ converges to the reflected Tracy–Widom distribution of order 2, $w_M \sim 1 - F_2(-s)$.

For the largest eigenvalue, we are interested in a threshold u_1 for which the probability that $\hat{\lambda}_1 > u_1$ is zero. Using the distribution of $\hat{\lambda}_1$, we find

$$\begin{aligned} P(\hat{\lambda}_1 > u_1) = 0 \quad &\Leftrightarrow \quad P\Big(\frac{N\frac{\hat{\lambda}_1}{\sigma^2} - \mu_1}{\nu_1} > \frac{N\frac{u_1}{\sigma^2} - \mu_1}{\nu_1}\Big) = 0 \\ &\Leftrightarrow \quad \frac{N\frac{u_1}{\sigma^2} - \mu_1}{\nu_1} = F_2^{-1}(1) \end{aligned}$$

where $F_2^{-1}(s)$ is the inverse CDF, and $f_1 := F_2^{-1}(1) \approx 2.24$. It follows that

$$\begin{aligned} u_1 &= \tfrac{\sigma^2}{N}(\mu_1 + \nu_1 f_1) \\ &= \tfrac{\sigma^2}{N}\left[N\rho_1^2 + N^{1/2}M^{-1/6}\rho_1^{4/3} f_1\right] \\ &= \sigma^2\rho_1^2\beta_1^2, \qquad \beta_1^2(N) := 1 + N^{-\frac{1}{2}}M^{-\frac{1}{6}}\rho_1^{-\frac{2}{3}} f_1, \end{aligned} \tag{B.2}$$

which shows a refinement of (B.1) by a factor β_1^2. For the smallest eigenvalue, we require a threshold u_M for which the probability that $\hat{\lambda}_M < u_M$ is zero, and we find similarly

$$u_M = \sigma^2\rho_M^2\beta_M^2, \qquad \beta_M^2(N) := 1 - N^{-\frac{1}{2}}M^{-\frac{1}{6}}\rho_M^{-\frac{2}{3}} f_1. \tag{B.3}$$

Appendix C. Noise power shifting using γ

In Eq. (33) ff. we introduced a parameter γ to shift the noise subspace such that it becomes a small but negative component in a covariance difference equation. This appendix shows a more detailed derivation for a slightly more general case.

As the noise power shift is sometimes positive, sometimes negative, we introduce a more general notation that captures this in a single expression. Thus, let

$$\overset{+}{[\mathbf{X}_1}\ \overset{-}{\alpha\mathbf{X}_2}\ \overset{j}{\gamma\mathbf{I}]}\,\mathbf{\Theta} = \overset{+}{[\mathbf{A}}\ \overset{-}{\mathbf{0}}\mid \mathbf{B}\ \mathbf{0}] \tag{C.1}$$

Here, $j = \pm 1$, and the superscripts $+$, $-$ and j indicate the "signature" of each component, i.e., the entries of the corresponding signature matrix $\mathbf{J}$. Details on this notation can be found in [17]. Upon squaring this equation, we obtain

$$\mathbf{X}_1\mathbf{X}_1^H - \alpha^2\mathbf{X}_2\mathbf{X}_2^H + j\gamma^2\mathbf{I} = \mathbf{A}\mathbf{A}^H - \mathbf{B}\mathbf{B}^H\,. \tag{C.2}$$

Thus, for $j = +1$, we add $\gamma^2\mathbf{I}$, and for $j = -1$, we subtract $\gamma^2\mathbf{I}$ in the covariance difference.

Assume $\mathbf{X}_1$ has N_1 samples, and $\mathbf{X}_2$ has N_2 samples. Then the data model for (C.2) is

$$\begin{aligned} N_1\hat{\mathbf{R}}_1 - \alpha^2 N_2\hat{\mathbf{R}}_2 + j\gamma^2\mathbf{I} \approx\ & \mathbf{H}_s(N_1\hat{\mathbf{R}}_{s1} - \alpha^2 N_2\hat{\mathbf{R}}_{s2})\mathbf{H}_s^H \\ & - \mathbf{H}_f(\alpha^2 N_2\hat{\mathbf{R}}_{f2} - N_1\hat{\mathbf{R}}_{f1})\mathbf{H}_f^H + N_1\hat{\mathbf{R}}_{n1} - \alpha^2 N_2\hat{\mathbf{R}}_{n2} + j\gamma^2\mathbf{I} \end{aligned}$$

where we ignored crossterms in the correlations (they diminish by order $1/N$ whereas the term we are interested in is of order $1/\sqrt{N}$). The objective is to choose j and γ such that $N_1\hat{\mathbf{R}}_{n1} - \alpha^2 N_2\hat{\mathbf{R}}_{n2} + j\gamma^2\mathbf{I}$ is guaranteed to be (slightly) negative definite, and thus becomes part of $\mathbf{B}\mathbf{B}^H$ that also contains the interference subspace.

Let $\hat{\lambda}_1$ be the largest eigenvalue of $\hat{\mathbf{R}}_{n1}$ and $\hat{\lambda}_M$ be the smallest eigenvalue of $\hat{\mathbf{R}}_{n2}$, and define

$$t = \alpha^2 N_2\hat{\lambda}_M - N_1\hat{\lambda}_1\,.$$

It is sufficient that $j\gamma^2 < t$. Note that we can have $t > 0$ or $t < 0$, depending on α and the number of samples.

As shown in Appendix B, a good estimate for t is

$$t = \sigma^2[\alpha^2\beta_M^2(N_2)\,(N_2^{\frac{1}{2}} - M^{\frac{1}{2}})^2 - \beta_1^2(N_1)\,(N_1^{\frac{1}{2}} + M^{\frac{1}{2}})^2] \tag{C.3}$$

where $\beta_1(N)$ was defined in (B.2) and $\beta_M(N)$ in (B.3). We set $j = \text{sign}(t)$ and $\gamma = \sqrt{|t|}$. This result is used in (33).

Appendix D. Convergence of correlations

Consider two signals $s_1[k]$ and $s_2[k]$, $k = 0, \cdots, N-1$. Each signal has $N_p < N$ samples unequal to zero and the other samples are zero. The nonzero parts are random signals with zero mean and unit variance, and overlap with $N_x < N_p$ samples. Stack the two signals into a data matrix $\mathbf{S}_x$ of size $2 \times N$ and compute the sample correlation matrix $\hat{\mathbf{R}}_x = \frac{1}{N}\mathbf{S}_x\mathbf{S}_x^H$. Similarly, let $\mathbf{S}_y$ be a $2 \times N$ data matrix with two equal-variance zero-mean stationary random sources, and compute $\hat{\mathbf{R}}_y = \frac{1}{N}\mathbf{S}_y\mathbf{S}_y^H$.

To have equal auto-correlations in $\hat{\mathbf{R}}_x$ and $\hat{\mathbf{R}}_y$, the variance of the sources in $\mathbf{S}_y$ should be set lower by a factor $\ell = N_p/N < 1$. The cross-correlations in $\hat{\mathbf{R}}_x$ go to zero as $O(\frac{N_x}{N^2})$. For $\hat{\mathbf{R}}_y$, taking into account the factor ℓ, they go to zero as $O(\frac{N_p}{N^2})$. As $N_x < N_p$, the cross-correlations go faster to zero for the nonstationary (intermittent) signals. Essentially this is because cross-products of samples in the non-overlapping parts are all zero. In conclusion, the sample correlation matrix for the nonstationary signals converges faster to its model than for stationary signals.

References

[1] M. Zhou, A.-J. van der Veen, Improved subspace intersection based on signed URV decomposition, in: Proc. of the Asilomar Conf. on Signals, Syst., and Comput, Pacific Grove (CA), USA, 2011.

[2] M. Zhou, A.J. van der Veen, Improved blind separation algorithm for overlapping secondary surveillance radar replies, in: Proc. IEEE CAMSAP, San Juan, Puerto Rico, 2011, pp. 181–184.

[3] M. Zhou, A.-J. van der Veen, A subspace tracking algorithm for separating partially overlapping data packets, in: Proc. of the Seventh IEEE Sensor Array and Multichannel Signal Process, Workshop (SAM), Hoboken, NJ, USA, 2012.

[4] M. Zhou, A.-J. van der Veen, Blind beamforming techniques for automatic identification system using GSVD and tracking, in: Proc. Int. Conf. Acoustics, Speech, Signal Proc., ICASSP 2014, IEEE, Florence (Italy), 2014.

[5] M. Zhou, Blind beamforming techniques for global tracking systems, Ph.D. thesis, ISBN 978-94-6259-934-5, TU Delft, Fac. EEMCS, Dec. 2015, http://repository.tudelft.nl/view/ir/uuid78-4ea0-be4a-4794be886bf4/.

[6] J. Andrews, W. Choi, R. Heath, Overcoming interference in spatial multiplexing MIMO cellular networks, IEEE Trans. Wirel. Commun. 14 (6) (2007) 95–104.

[7] S. Karimifar, J. Cavers, Achieving high-capacity narrowband cellular systems by means of multicell multiuser detection, IEEE Trans. Inf. Theory 57 (2) (2008) 945–953.

[8] M.A. Cervera, A. Ginesi, K. Eckstein, Satellite-based vessel automatic identification system: a feasibility and performance analysis, Int. J. Satell. Commun. Netw. 29 (2) (2011) 117–142.

[9] M. Zhou, A.-J. van der Veen, R. van Leuken, Multi-user LEO-satellite receiver for robust space detection of AIS messages, in: Proc. IEEE ICASSP, Kyoto, Japan, 2012.

[10] E. Chaumette, P. Comon, D. Muller, ICA-based technique for radiating sources estimation: application to airport surveillance, IEE Proc., F, Radar Signal Process. 140 (6) (1993) 395–401.

[11] N. Petrochilos, A.-J. van der Veen, Algebraic algorithms to separate overlapping secondary surveillance radar replies, IEEE Trans. Signal Process. 55 (7) (2007) 3746–3759.

[12] A.-J. van der Veen, A. Paulraj, An analytical constant modulus algorithm, IEEE Trans. Signal Process. 44 (5) (1996) 1136–1155.

[13] J. Cardoso, A. Souloumiac, Blind beamforming for non-Gaussian signals, IEE Proc., F, Radar Signal Process. 140 (6) (1993) 362–370.

[14] C. Papadias, Globally convergent blind source separation based on a multiuser kurtosis maximization criterion, IEEE Trans. Signal Process. 48 (12) (2000) 3508–3519.

[15] C.C. Paige, M.A. Saunders, Towards a generalized singular value decomposition, SIAM J. Numer. Anal. 18 (3) (1981) 398–405.

[16] A.-J. van der Veen, A Schur method for low-rank matrix approximation, SIAM J. Matrix Anal. Appl. 17 (1) (1996) 139–160.

[17] M. Zhou, A.-J. van der Veen, Stable subspace tracking algorithm based on a signed URV decomposition, IEEE Trans. Signal Process. 60 (6) (2012) 3036–3051.

[18] R. Behrens, L. Scharf, Signal processing applications of oblique projection operators, IEEE Trans. Signal Process. 42 (6) (1994) 1413–1424.

[19] L. Scharf, M. McCloud, Blind adaptation of zero forcing projections and oblique pseudo-inverses for subspace detection and estimation when interference dominates noise, IEEE Trans. Signal Process. 50 (12) (2002).

[20] N. Petrochilos, G. Galati, L. Mene, E. Piracci, Separation of multiple secondary surveillance radar sources in a real environment by a novel projection algorithm, in: Proc. 5th IEEE Int. Symp. Signal Proc. Information Techn., 2005, pp. 125–130.

[21] N. Petrochilos, G. Galati, E. Piracci, Separation of SSR signals by array processing in multilateration systems, IEEE Trans. Aerosp. Electron. Syst. 45 (3) (2009) 965–982.

[22] G.H. Golub, C.F. Van Loan, Matrix Computations, 3rd ed., Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[23] D. Callaerts, B. De Moor, J. Vandewalle, W. Sansen, G. Vantrappen, J. Janssens, Comparison of SVD methods to extract the foetal electrocardiogram from cutaneous electrode signals, Med. Biol. Eng. Comput. 28 (3) (1990) 217–224.

[24] P. Hansen, S. Jensen, Prewhitening for rank-deficient noise in subspace methods for noise reduction, IEEE Trans. Signal Process. 53 (10) (2005) 3718–3726.

[25] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, E. Moulines, A blind source separation technique based on second order statistics, IEEE Trans. Signal Process. 45 (2) (1997) 434–444.

[26] J. Götze, A.-J. van der Veen, On-line subspace estimation using a Schur-type method, IEEE Trans. Signal Process. 44 (6) (1996) 1585–1589.

[27] G. Stewart, An updating algorithm for subspace tracking, IEEE Trans. Signal Process. 40 (6) (1992) 1535–1541.

[28] A.J. van der Veen, J. Tol, Separation of zero/constant modulus signals, in: Proc. IEEE ICASSP, IEEE, Munich (Germany), 1997, pp. 3445–3448.

[29] A.M. Kuzminskiy, Y.I. Abramovich, Second-order asynchronous interference cancellation: regularized semi-blind technique and non-asymptotic maximum likelihood benchmark, Signal Process. 86 (12) (2006) 3849–3863.

[30] A.M. Kuzminskiy, Y.I. Abramovich, Non-stationary multiple-antenna interference cancellation for unsynchronized OFDM systems with distributed training, Signal Process. 89 (5) (2009) 753–764.

[31] Z.D. Bai, Y.Q. Yin, Limit of the smallest eigenvalue of a large dimensional sample covariance matrix, Ann. Probab. 21 (3) (1993) 1275–1294.

[32] K. Johansson, Shape fluctuations and random matrices, Commun. Math. Phys. 209 (2000) 437–476.
[33] C.A. Tracy, H. Widom, Correlation functions, cluster functions, and spacing distributions for random matrices, J. Stat. Phys. 92 (1998) 809–835.
[34] D. Paul, Asymptotic Distribution of the Smallest Eigenvalue of Wishart(N, n) when $N, n \to \infty$ such that $N/n \to 0$, World Scientific, 2011.

Mu Zhou received the B.Sc. and M.Sc. degrees from the School of Information Science and Engineering, Southeast University, Nanjing, P. R. China, in 2005 and 2008, respectively, all in electrical engineering. He obtained the PhD degree in 2015 from TU Delft, The Netherlands. He is currently with Huawei, China. His research interests include matrix decomposition for array signal processing, receiver design for nano-satellite communication and mixed signal system modeling.

Alle-Jan van der Veen was born in The Netherlands in 1966. He received the Ph.D. degree (cum laude) from TU Delft in 1993. At present, he is a Full Professor in Signal Processing at TU Delft.

Prof. Van der Veen is an IEEE Fellow. He is the recipient of a 1994 and a 1997 IEEE Signal Processing Society (SPS) Young Author paper award. He was Editor-in-Chief of IEEE Signal Processing Letters (2002–2005) and Editor-in-Chief of IEEE Transactions on Signal Processing (2006–2008), Technical Cochair of ICASSP 2011, and chair of the IEEE SPS Fellow Reference Committee. He also is an EURASIP Fellow, and currently serves as the EURASIP Director of Publications.

His research interests relate to algebraic methods for array signal processing, with applications to wireless communications and radio astronomy.