KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Robbe Van Rompaey, Marc Moonen, (2018), **Using Prior Knowledge on the Local Steering Matrix Subspace for Distributed Adaptive Node-specific Signal Estimation in a Wireless Sensor Network**
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	Klik hier als u tekst wilt invoeren.
Journal homepage	Klik hier als u tekst wilt invoeren.
Author contact	robbe.vanrompaey@esat.kuleuven.be *+ 32 (0)16 37 37 40*
Abstract	This paper first introduces the centralized generalized eigenvalue decomposition (GEVD) based Multichannel Wiener filter (MWF) with prior knowledge for node-specific signal estimation in a wireless sensor network (WSN), where nodes have prior knowledge on the local steering matrix subspace. A distributed adaptive estimation algorithm for a fully-connected WSN is then proposed demonstrating that this MWF can be obtained by letting the nodes work on compressed (i.e. reduced-dimensional) sensor signals compared to the centralized approach. The algorithm can be used in many applications where prior knowledge is available, such as speech enhancement in acoustic sensor networks, where nodes have prior knowledge on the locations of their desired speech sources and on their local microphone array geometry
IR	Klik hier als u tekst wilt invoeren.

(article begins on next page)

USING PRIOR KNOWLEGDE ON THE LOCAL STEERING MATRIX SUBSPACE FOR DISTRIBUTED ADAPTIVE NODE-SPECIFIC SIGNAL ESTIMATION IN A WIRELESS SENSOR NETWORK

Robbe Van Rompaey[*]*, Marc Moonen*

KU Leuven - Dept. ESAT, STADIUS
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
E-mail: robbe.vanrompaey@esat.kuleuven.be; marc.moonen@esat.kuleuven.be

ABSTRACT

This paper first introduces the centralized generalized eigenvalue decomposition (GEVD) based Multichannel Wiener filter (MWF) with prior knowledge for node-specific signal estimation in a wireless sensor network (WSN), where nodes have prior knowledge on the local steering matrix subspace. A distributed adaptive estimation algorithm for a fully-connected WSN is then proposed demonstrating that this MWF can be obtained by letting the nodes work on compressed (i.e. reduced-dimensional) sensor signals compared to the centralized approach. The algorithm can be used in many applications where prior knowledge is available, such as speech enhancement in acoustic sensor networks, where nodes have prior knowledge on the locations of their desired speech sources and on their local microphone array geometry.

Index Terms— Wireless Sensor Networks (WSN), distributed estimation, Multichannel Wiener Filter (MWF), generalized eigenvalue decomposition (GEVD).

1. INTRODUCTION

In a wireless sensor network (WSN) [1], nodes aim to combine their sensor signals in an optimal way to perform a task at hand, such as the estimation of a certain signal. For the estimation of a common signal, the sensor signals are mostly locally combined or fused, possibly in a fusion center (FC) [2, 3], generally leading to superior estimation performance compared to that of the stand-alone estimation, where each node uses only local sensor signals.

When node-specific signal estimation is considered, each node aims to estimate a different node-specific signal, e.g. a desired source signal component in a certain reference sensor [4–6]. The different node-specific signals to be estimated, are then assumed to be dependent on a common low-dimensional signal, such as a mixture of a number of desired speech signals. The algorithms in [4–6] exploit this low-dimensional signal subspace, to significantly compress the sensor signals that are communicated between the nodes, without compromising performance.

To construct the signal correlation matrix of the low-dimensional signal, the algorithms assume to have access to either training frames or to have access to the activity (on-off) pattern of the random signal. Since training frames are difficult to use in an acoustic scenario, the second option is mostly used. However in low SNR scenarios, this might result in a poor estimation of the signal correlation matrix, deteriorating the node-specific signal estimation performance. Inspired by Ali et al. [7], a scenario is considered in this paper where nodes have prior knowledge on the local steering matrix subspace (i.e., the subspace for the signal correlation matrix), which can, for instance, in an acoustic scenario be obtained if nodes have prior knowledge on the locations of their desired speech sources and on their local microphone array geometry [8, 9].

This paper first introduces the centralized generalized eigenvalue decomposition (GEVD) based Multichannel Wiener filter (MWF) with prior knowledge for node-specific signal estimation in a wireless sensor network (WSN), where nodes have prior knowledge on the local steering matrix subspace. A distributed adaptive estimation algorithm for a fully-connected WSN is then proposed demonstrating that this MWF can be obtained by letting the nodes work on compressed (i.e. reduced-dimensional) sensor signals compared to the centralized approach. It turns out that the number of compressed sensor signals broadcast by each node, will be double the number needed in previous algorithms [4–6], since extra compressed sensors signals are needed to propagate the prior knowledge on the local steering matrix subspace to all the other nodes. Still the signal estimation task is enhanced with this prior knowledge, justifying this extra communication requirements.

The paper is organized as follows. The problem formulation and the centralized approach to the node specific signal estimation problem with prior knowledge on the local steering matrix subspace are presented in Section 2. In Section 3 the distributed algorithm is presented. Section 4 provides batch-mode simulations to show convergence of the proposed distributed algorithm. Conclusions are given in Section 5.

2. PROBLEM FORMULATION AND PK-GEVD-MWF

2.1. Node-specific signal estimation

Consider a fully-connected WSN with K nodes, where each node $k \in \mathcal{K} = \{1, ..., K\}$ has access to observations of a different M_k-dimensional complex-valued sensor signal $\mathbf{y}_k$. $\mathbf{y}_k$ is modeled as

$$\mathbf{y}_k = \mathbf{s}_k + \mathbf{n}_k = \mathbf{A}_k \check{\mathbf{s}} + \mathbf{n}_k \tag{1}$$

where $\check{\mathbf{s}}$ is a latent S-dimensional complex-valued signal representing S mutually uncorrelated source signals, $\mathbf{A}_k$ is an unknown $M_k \times$

[*]The work of R. Van Rompaey was supported by a doctoral Fellowship of the Research Foundation Flanders (FWO-Vlaanderen). This work was carried out at the ESAT Laboratory of KU Leuven in the frame of KU Leuven internal funding C2-16-00449 'Distributed Digital Signal Processing for Ad-hoc Wireless Local Area Audio Networking' and FWO/FNRS EOS Poject nr. 30452698 'MUSE-WINET - Multi-Service Wireless Network'. The scientific responsibility is assumed by its authors.

S complex-valued steering matrix and $\mathbf{n}_k$ is an additive noise component that can be correlated over all the sensor signals in the WSN. Define also the centralized M-channel signals $\mathbf{y}, \mathbf{s}, \mathbf{n}$ and the centralized $M \times S$ steering matrix $\mathbf{A}$ as the stacked version of $\mathbf{y}_k, \mathbf{s}_k, \mathbf{n}_k$ and $\mathbf{A}_k$ respectively, where $M = \sum_{k=1}^{M} M_k$. Then (1) can be extended to

$$\mathbf{y} = \mathbf{s} + \mathbf{n} = \mathbf{A}\check{\mathbf{s}} + \mathbf{n}. \tag{2}$$

The node-specific task of each node $k \in \mathcal{K}$ is to find an estimate of the L-channel desired signal $\mathbf{d}_k$, defined w.l.o.g. as the first L channels of $\mathbf{s}_k$:

$$\mathbf{d}_k = [\mathbf{I}_L \; \mathbf{0}]^H \mathbf{s}_k = [\mathbf{0} \; \mathbf{I}_L \; \mathbf{0}]^H \mathbf{s} = \check{\mathbf{E}}_{\mathbf{d}_k}^H \mathbf{s} \quad \forall k \in \mathcal{K} \tag{3}$$

where ^{H} denotes the conjugate transpose operator, $\mathbf{I}_L$ is the $L \times L$ identity matrix and $\mathbf{0}$ is an all-zero matrix with matching dimensions. Each node estimates its desired signal $\mathbf{d}_k$ as a linear combination of all the sensor signals $\mathbf{y}$ by minimizing the following minimum mean square error (MSE) criterion:

$$\boxed{\check{\mathbf{W}}_k = \arg\min_{W_k} E\{\|\mathbf{d}_k - \mathbf{W}_k^H \mathbf{y}\|^2\}} \tag{4}$$

where $E\{.\}$ is the expected value operator. The resulting filter is called the Multichannel Wiener Filter (MWF)[1]. If $\mathbf{R}_{\mathbf{yy}} = E\{\mathbf{y}\mathbf{y}^H\}$ has full rank, the unique solution of (4) is [10]:

$$\boxed{\check{\mathbf{W}}_k = \mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{R}_{\mathbf{yd}_k} = \mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{R}_{\mathbf{ys}}\hat{\mathbf{E}}_{\mathbf{d}_k} = \mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{R}_{\mathbf{ss}}\check{\mathbf{E}}_{\mathbf{d}_k}} \tag{5}$$

with $\mathbf{R}_{\mathbf{yd}_k} = E\{\mathbf{y}\mathbf{d}_k^H\}$, $\mathbf{R}_{\mathbf{ys}} = E\{\mathbf{y}\mathbf{s}^H\}$ and $\mathbf{R}_{\mathbf{ss}} = E\{\mathbf{s}\mathbf{s}^H\}$. The last step in (5) is allowed due to the (often valid) assumption that the additive noise component $\mathbf{n}$ and the source signal $\check{\mathbf{s}}$ are uncorrelated. $\mathbf{R}_{\mathbf{ss}}$ is then given by $\mathbf{A}E\{\check{\mathbf{s}}\check{\mathbf{s}}^H\}\mathbf{A}^H$, where $E\{\check{\mathbf{s}}\check{\mathbf{s}}^H\}$ is a diagonal matrix containing the source signal powers. Notice that $\mathbf{R}_{\mathbf{ss}}$ is not directly observable, since nodes do not have access to the clean desired signal $\mathbf{d}_k$. A robust way to estimate $\mathbf{R}_{\mathbf{ss}}$ is given in the next subsection, based on the exploitation of the on-off behavior of the source signal and on prior knowledge on the steering matrix subspace of $\mathbf{A}$.

2.2. GEVD based estimation of $\mathbf{R}_{\mathbf{ss}}$ with prior knowledge

If the source signal has an on-off behavior and the on-off detection of the signal is available, e.g. via a voice activity detector in speech applications [11, 12], a distinction can be made between the 'signal+noise' correlation matrix $\mathbf{R}_{\mathbf{yy}}$ and 'noise-only' correlation matrix $\mathbf{R}_{\mathbf{nn}}$. These correlation matrices can be estimated by (recursive) time-averaging during 'signal+noise' and 'noise-only' periods if $\mathbf{y}$ is assumed to satisfy (short-term) stationarity and ergodicity conditions.

$\mathbf{R}_{\mathbf{ss}}$ can then be estimated as $\mathbf{R}_{\mathbf{ss}} = \mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}}$. However such an estimate has mostly a rank larger than S, especially in low SNR scenarios, so that better correlation matrix estimation methods are necessary.

There exist different signal correlation matrix estimation methods [13], but recently Ali et al. [7] have introduced a very general signal correlation matrix estimation method, where the on-off behavior of the source signal is exploited and prior knowledge on the local steering matrix subspace is taken into account.

[1]Notice that all above signals and filters are defined as complex-valued signals, permitting the model to include, e.g., convolutive time-domain mixtures, described as instantaneous per-frequency mixtures in the (short-term) Fourier transform domain making it applicable for speech enhancement.

Applying this in the WSN context, one can imagine a scenario where all the nodes $k \in \mathcal{K}$ have prior knowledge on the local steering matrix subspace of $\mathbf{A}_k$. For instance, in an acoustic scenario a node can have an estimate for the relative transfer function between the desired source(s) and all its microphones [8, 9]. This prior knowledge can then be represented by a unitary $M_k \times S$ subspace matrix $\mathbf{H}_k$ ($\mathbf{H}_k^H\mathbf{H}_k = \mathbf{I}_S$) with $col(\mathbf{H}_k) = col(\mathbf{A}_k)$. Denote also the orthogonal complement of $\mathbf{H}_k$ as the unitary $M_k \times (M_k - S)$ blocking matrix $\mathbf{B}_k$, such that $\mathbf{H}_k^H\mathbf{B}_k = \mathbf{0}$ and $\mathbf{B}_k^H\mathbf{B}_k = \mathbf{I}_{M_k - S}$. Stacking these subspace matrices and blocking matrices in one centralized subspace matrix and blocking matrix respectively results in

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{H}_K \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{B}_K \end{bmatrix} \tag{6}$$

where $\mathbf{H}$ is a block-diagonal $M \times KS$ dimensional matrix and $\mathbf{B}$ a block-diagonal $M \times (M - KS)$ dimensional matrix.

The following centralized optimization criterion is then defined to provide an estimate for $\mathbf{R}_{\mathbf{ss}}$:

$$\boxed{\underset{\substack{\mathbf{R}_{\mathbf{ss}} = \text{rank-}S; \\ \mathbf{B}^H\mathbf{R}_{\mathbf{ss}}\mathbf{B} = \mathbf{0}}}{\arg\min} \; \| \, \mathbf{R}_{\mathbf{nn}}^{-1/2} \left(\mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}} - \mathbf{R}_{\mathbf{ss}}\right) \mathbf{R}_{\mathbf{nn}}^{-H/2} \, \|_F^2} \tag{7}$$

where $\| \, . \, \|_F$ denotes the Frobenius norm. Here $\mathbf{R}_{\mathbf{ss}}$ is constrained to be rank-S, the column and row space of $\mathbf{R}_{\mathbf{ss}}$ are constrained to lie in the space formed by the columns of $\mathbf{H}$ and approximation errors are considered relative to the noise correlation matrix $\mathbf{R}_{\mathbf{nn}}$ (cfr. the pre- and post-multiplication with the cholesky factorization [14] of $\mathbf{R}_{\mathbf{nn}}^{-1}$).

The solution (proof omitted) to (7) is based on the GEVD [10, 15] of a reduced $KS \times KS$ dimensional matrix pencil $\{\hat{\mathbf{R}}_{\mathbf{yy}}, \hat{\mathbf{R}}_{\mathbf{nn}}\}$:

$$\begin{aligned} \hat{\mathbf{R}}_{\mathbf{yy}} &= \hat{\mathbf{Q}}\hat{\mathbf{\Sigma}}_{\mathbf{yy}}\hat{\mathbf{Q}}^H \\ \hat{\mathbf{R}}_{\mathbf{nn}} &= \hat{\mathbf{Q}}\hat{\mathbf{\Sigma}}_{\mathbf{nn}}\hat{\mathbf{Q}}^H \end{aligned} \tag{8}$$

where $\hat{\mathbf{Q}} = \hat{\mathbf{X}}^{-H}$ is an invertible matrix. The columns of $\hat{\mathbf{X}}$ are unique up to a scalar and define the generalized eigenvectors. $\hat{\mathbf{\Sigma}}_{\mathbf{yy}}$ and $\hat{\mathbf{\Sigma}}_{\mathbf{nn}}$ are real-valued diagonal matrices where $\hat{\mathbf{\Sigma}}_{\mathbf{yy}} = \text{diag}\{\hat{\sigma}_{y_1}, .., \hat{\sigma}_{y_K}\}, \hat{\mathbf{\Sigma}}_{\mathbf{nn}} = \text{diag}\{\hat{\sigma}_{n_1}, .., \hat{\sigma}_{n_K}\}$ define the generalized eigenvalues sorted from high to low $\{\hat{\sigma}_{y_i}/\hat{\sigma}_{n_i}\}$ ratio.

The reduced $KS \times KS$ dimensional correlation matrices $\{\hat{\mathbf{R}}_{\mathbf{yy}}, \hat{\mathbf{R}}_{\mathbf{nn}}\}$ can be determined by first performing an LCMV-beamformer on the sensor signals $\mathbf{y}$, defined by the following LCMV-criterion:

$$\begin{aligned} \min_{\mathbf{C}} \quad & \text{trace}\{\mathbf{C}^H\mathbf{R}_{\mathbf{nn}}\mathbf{C}\} \\ s.t. \quad & \mathbf{H}^H\mathbf{C} = \mathbf{I}_{KS} \end{aligned} \tag{9}$$

where C is an $M \times KS$ matrix, of which every column represents a specific LCMV-beamformer. The solution, here based on a GSC-implementation [16], is given by

$$\mathbf{C} = \mathbf{H} - \mathbf{B}\mathbf{F} \tag{10}$$

$$\mathbf{F} = (\mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbf{H} \tag{11}$$

The reduced dimension correlation matrices $\{\hat{\mathbf{R}}_{\mathbf{yy}}, \hat{\mathbf{R}}_{\mathbf{nn}}\}$ are then determined as the correlation matrices corresponding to the compressed signal $\hat{\mathbf{y}} = \mathbf{C}^H\mathbf{y}$.

The optimal solution for $\mathbf{R}_{\mathbf{ss}}$ of (7) is finally given by

$$\check{\mathbf{R}}_{\mathbf{ss}} = \mathbf{H}\hat{\mathbf{Q}}\text{diag}\{\hat{\sigma}_{y_1} - \hat{\sigma}_{n_1}, ..., \hat{\sigma}_{y_S} - \hat{\sigma}_{n_S}, 0, ..., 0\}\hat{\mathbf{Q}}^H\mathbf{H}^H. \tag{12}$$

2.3. Centralized GEVD based MWF with prior-knowledge

Substituting estimate (12) in (5) and using $\mathbf{R_{yy}} = \mathbf{R_{nn}} + \check{\mathbf{R}}_{\mathbf{ss}}$, after some manipulations, results in

$$\boxed{\check{\mathbf{W}}_k = \mathbf{C}^H \hat{\mathbf{W}}_{GEVD} \mathbf{H}^H \check{\mathbf{E}}_{\mathbf{d}_k}} \tag{13}$$

where

$$\hat{\mathbf{W}}_{GEVD} = \hat{\mathbf{X}} \text{diag}\{\frac{\hat{\sigma}_{y_1} - \hat{\sigma}_{n_1}}{\hat{\sigma}_{y_1}}, ..., \frac{\hat{\sigma}_{y_S} - \hat{\sigma}_{n_S}}{\hat{\sigma}_{y_S}}, 0, ..., 0\} \hat{\mathbf{Q}}^H \tag{14}$$

is the GEVD-based MWF-filter [6,13] that estimates the compressed source components $\hat{\mathbf{s}} = \mathbf{C}^H \mathbf{s}$ in the compressed signal $\hat{\mathbf{y}}$.

The filter obtained in (13) is referred to as the prior-knowledge GEVD based MWF (PK-GEVD-MWF) and the formula shows that the resulting filter is a concatenation of three different blocks. The first block corresponds to the LCMV-beamformer (9), the second block is a full GEVD-based MWF and the last block is a selection and scaling part to obtain the node-specific desired signal $\mathbf{d}_k$.

To determine the centralized PK-GEVD-MWF, the centralized correlation matrices $\mathbf{R_{yy}}$ and $\mathbf{R_{nn}}$ need to be constructed. This would require the nodes to send all their M_k sensor signals $\mathbf{y}_k$ to FC. The transmission of the sensor signals will require a large communication bandwidth, and the obtained correlation matrices will become large so that the inversion of $\mathbf{B}^H \mathbf{R_{nn}} \mathbf{B}$ in (11) will require significant computational power at the FC.

To overcome this strict requirement, a distributed adaptive estimation algorithm is presented in the next section where nodes only broadcast $2S$ compressed sensor signals and the local computations are performed on a smaller number of signals[2], i.e. only the local sensor signal and the received compressed sensor signals from the other nodes. It will turn out that each node will (upon convergence) still be able to obtain the same filter output as if the node had access to all the sensor signals in the WSN and so could directly compute the centralized PK-GEVD-MWF. The distributed algorithm is here referred to as the Prior Knowledge Distributed Adaptive Node Specific (PK-GEVD-DANSE) algorithm. A drawback of the PK-GEVD-DANSE algorithm is the slower adaptation and tracking speed compared to the centralized approach, due to the block-iterative nature of the algorithm.

3. PK-GEVD-DANSE ALGORITHM

3.1. Algorithm description

In the PK-GEVD-DANSE algorithm, each node k broadcasts $2S$ compressed sensor signals instead of the full M_k-dimensional sensor signal $\mathbf{y}_k$. The $2S$-dimensional compressed sensor signals consists of two parts (supercript i denotes the iteration index):

- the S-dimensional signal $\mathbf{p}_k^i = \mathbf{\Pi}_k^{i^H} \mathbf{y}_k$ where the $S \times M_k$ dimensional compression matrix $\mathbf{\Pi}_k^i$ corresponds to the contribution of all the local sensor signals to the output of the local filter and will be defined later;
- the S-dimensional signal $\boldsymbol{\lambda}_k^i = \mathbf{\Lambda}_k^{i^H} \mathbf{B}_k^H \mathbf{y}_k$ corresponding to a compressed version of the local noise references $\mathbf{B}_k^H \mathbf{y}_k$, where the $S \times (M_k - S)$ dimensional compression matrix $\mathbf{\Lambda}_k^i$ will be defined later.

[2]In an iteration of Algorithm 1 in Section 3, the inversion of a typically much smaller matrix $\tilde{\mathbf{B}}_q^H \mathbf{R}_{\tilde{\mathbf{n}}_q \tilde{\mathbf{n}}_q} \tilde{\mathbf{B}}_q$ is needed, which makes the algorithm also less susceptible to numerical errors.

By consequence, each node k has access to a reduced-dimensional sensor signal $\tilde{\mathbf{y}}_k = [\mathbf{y}_k^H \mathbf{p}_{-k}^{i^H} \boldsymbol{\lambda}_{-k}^{i,H}]^H$ where the subscript $-k$ refers to the concatenation of the broadcast compressed sensor signals of the other nodes: $\mathbf{p}_{-k}^{i^H} = [\mathbf{p}_1^{i^H} ... \mathbf{p}_{k-1}^{i^H} \mathbf{p}_{k+1}^{i^H} ... \mathbf{p}_K^{i^H}]^H$ and $\boldsymbol{\lambda}_{-k}^{i^H} = [\boldsymbol{\lambda}_1^{i^H} ... \boldsymbol{\lambda}_{k-1}^{i^H} \boldsymbol{\lambda}_{k+1}^{i^H} ... \boldsymbol{\lambda}_K^{i^H}]^H$.

To be able to perform the same operations as in the centralized PK-GEVD-MWF, one needs to have a unitary estimate for the steering matrix subspace of $\tilde{\mathbf{y}}_k$. Here the following subspace matrix $\tilde{\mathbf{H}}_k$ and corresponding blocking matrix $\tilde{\mathbf{B}}_k$ are proposed:

$$\tilde{\mathbf{H}}_k = \begin{bmatrix} \mathbf{H}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{S(K-1)} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}; \quad \tilde{\mathbf{B}}_k = \begin{bmatrix} \mathbf{B}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{S(K-1)} \end{bmatrix}. \tag{15}$$

The signal subspace for $\mathbf{y}_k$ is evidently defined by $\mathbf{H}_k$ and since the signal subspace for every entry $\mathbf{p}_q^i$ in $\mathbf{p}_{-k}^i$ is given by $\mathbf{\Pi}_q^{i^H} \mathbf{H}_q$, in general a full rank square matrix, the signal subspace can be represented by the identity matrix. The signal subspace for $\boldsymbol{\lambda}_{-k}^{i,H}$ is empty because these signals are compressed versions of the local noise references $\mathbf{B}_k^H \mathbf{y}_k$, where the signal component is already locally canceled ($\mathbf{B}_k^H \mathbf{y}_k = \mathbf{B}_k^H \mathbf{A}_k \check{\mathbf{s}} + \mathbf{B}_k^H \mathbf{n}_k = \mathbf{B}_k^H \mathbf{n}_k$).

The PK-GEVD-DANSE algorithm is presented in Algorithm 1. This is a block-iterative round-robin algorithm, where the updating node performs the same operations as the operations in the centralized algorithm, but here with locally defined reduced-dimensional variables $\mathbf{F}_k^i, \mathbf{C}_k^i, \hat{\mathbf{W}}_k^i$ and $\tilde{\mathbf{W}}_k^i$. The definition of the local compression matrix $\mathbf{\Pi}_k^i$ is given in (21) and indeed corresponds to the part of the filter $\tilde{\mathbf{W}}_k^i$ working on the local sensor signals $\mathbf{y}_k$ as explained before. The next subsection will provide an intuitive explanation for the definition of the local compression matrix $\mathbf{\Lambda}_k^i$ of the local noise references $\mathbf{B}_k^H \mathbf{y}_k$ in (22). A proof of convergence showing that Algorithm 1 converges to the PK-GEVD-MWF (13) for any random initialization of the compression matrices, will be provided elsewhere.

3.2. Comparison with GEVD-DANSE

A related algorithm to PK-GEVD-DANSE is GEVD-DANSE [6], which aims to estimate the centralized GEVD based MWF in a distributed way, but without the ability to introduce prior knowledge on the local steering matrix subspace. The algorithm only requires S per node broadcast signals, compared to $2S$ signals for PK-GEVD-DANSE. The extra S messages of PK-GEVD-DANSE are the compressed versions of the local noise references $\mathbf{B}_k^H \mathbf{y}_k$. From simulations it is shown that, upon convergence of PK-GEVD-DANSE, $\boldsymbol{\lambda}_k$ is equal to its centralized variant $[\mathbf{0}\ \mathbf{I}_{M_k - S}\ \mathbf{0}] \left(\mathbf{B}^H \mathbf{R_{nn}} \mathbf{B}\right)^{-1} \mathbf{B}^H \mathbf{R_{yy}} \check{\mathbf{W}}_k$. This can be shown to be the optimal compression of the noise references of node k to still be able to let a similar procedure as in Section 2, attain the same PK-GEVD-MWF (13) as when all the noise references are used in (10). One can also show that in the case where $\mathbf{B}^H \mathbf{R_{yy}} \mathbf{B}$ is exactly equal to $\mathbf{B}^H \mathbf{R_{nn}} \mathbf{B}$, $\boldsymbol{\lambda}_k$ becomes equal to $\mathbf{0}, \forall k \in \mathcal{K}$ (so in fact unnecessary for PK-GEVD-DANSE) and the obtained node-specific MWF's in PK-GEVD-DANSE and GEVD-DANSE will be the same. The compressed versions of the local noise references thus accounts for estimation mismatch in $\mathbf{R_{yy}}$ and $\mathbf{R_{nn}}$, since in the ideal case $\mathbf{B}^H \mathbf{R_{yy}} \mathbf{B}$ should be equal to $\mathbf{B}^H \mathbf{R_{nn}} \mathbf{B}$. From the simulations in the next section, it will be clear that the PK-GEVD-MWF is still performing better in terms of minimizing the objective in (4) than the GEVD based MWF, justifying the communication cost for the extra S compressed signals per node.

Algorithm 1: PK-GEVD-DANSE algorithm

1 - Construct $\tilde{\mathbf{H}}_k$ and $\tilde{\mathbf{B}}_k$ using node k's prior knowledge, initialize $\mathbf{\Pi}_k^0$ and $\mathbf{\Lambda}_k^0$ as random matrices, $\forall k \in \mathcal{K}$
- $i \leftarrow 0$ and $q \leftarrow 1$

2 - All nodes $k \in \mathcal{K}$ broadcast N new compressed observations of $\mathbf{p}_k = \mathbf{\Pi}_k^{iH}\mathbf{y}_k$ and $\boldsymbol{\lambda}_k = \mathbf{\Lambda}_k^{iH}\mathbf{B}_k^H\mathbf{y}_k$ to the other nodes and construct locally

$$\tilde{\mathbf{y}}_k = [\mathbf{y}_q^H \mathbf{p}_{-k}^H \boldsymbol{\lambda}_{-k}^H]^H \tag{16}$$

3 - (Updating) node q estimates $\mathbf{R}_{\tilde{\mathbf{y}}_q\tilde{\mathbf{y}}_q}$ and $\mathbf{R}_{\tilde{\mathbf{n}}_q\tilde{\mathbf{n}}_q}$ based on the N new observations and updates its local parameters:

$$\mathbf{F}_q^{i+1} = \left(\tilde{\mathbf{B}}_q^H \mathbf{R}_{\tilde{\mathbf{n}}_q\tilde{\mathbf{n}}_q}\tilde{\mathbf{B}}_q\right)^{-1}\tilde{\mathbf{B}}_q^H\mathbf{R}_{\tilde{\mathbf{n}}_q\tilde{\mathbf{n}}_q}\tilde{\mathbf{H}}_q \tag{17}$$

$$\mathbf{C}_q^{i+1} = \tilde{\mathbf{H}}_q - \tilde{\mathbf{B}}_q\mathbf{F}_q^{i+1}, \quad \hat{\mathbf{y}}_q = {\mathbf{C}_q^{i+1}}^H\tilde{\mathbf{y}}_q \tag{18}$$

$$\hat{\mathbf{W}}_q^{i+1} = \arg\min_{\mathbf{W}} E\{\| \hat{\mathbf{s}}_q - \mathbf{W}^H\hat{\mathbf{y}}_q \|^2\} \tag{19}$$

$$\tilde{\mathbf{W}}_q^{i+1} = {\mathbf{C}_q^{i+1}}^H\hat{\mathbf{W}}_q^{i+1}\tilde{\mathbf{H}}^H[\mathbf{I}_S \;\; \mathbf{0} \;\; \mathbf{0}]^H \tag{20}$$

$$\mathbf{\Pi}_q^{i+1} = \begin{bmatrix}\mathbf{I}_{M_q} & \mathbf{0}\end{bmatrix}\tilde{\mathbf{W}}_q^{i+1} \tag{21}$$

$$\mathbf{\Lambda}_q^{i+1} = \tag{22}$$

$$\begin{bmatrix}\mathbf{I}_{M_q-S} & \mathbf{0}\end{bmatrix}\left(\tilde{\mathbf{B}}_q^H\mathbf{R}_{\tilde{\mathbf{n}}_q\tilde{\mathbf{n}}_q}\tilde{\mathbf{B}}_q\right)^{-1}\tilde{\mathbf{B}}_q^H\mathbf{R}_{\tilde{\mathbf{y}}_q\tilde{\mathbf{y}}_q}\tilde{\mathbf{W}}_q^{i+1}$$

- All other nodes do not change their variables:

$$\mathbf{\Pi}_k^{i+1} = \mathbf{\Pi}_k^i, \mathbf{\Lambda}_k^{i+1} = \mathbf{\Lambda}_k^i, \tilde{\mathbf{W}}_q^{i+1} = \tilde{\mathbf{W}}_q^i \tag{23}$$

4 For the N new observations, each node $k \in \mathcal{K}$ generates an estimate of its desired signal:

$$\mathbf{d}_k{}^i \approx {\tilde{\mathbf{W}}_q^{i+1}}^H\tilde{\mathbf{y}}_q \tag{24}$$

5 $i \leftarrow i+1, q \leftarrow (q \mod K)+1$ and return to step 2

4. SIMULATIONS

To demonstrate the convergence and optimality of the PK-GEVD-DANSE algorithm, batch-mode[3] Monte-Carlo (MC) simulations are conducted and compared with the GEVD-DANSE algorithm and the centralized PK-GEVD-MWF. For every MC run, a new scenario with $K = 8$ nodes is randomly generated, where every node has access to 10 different sensor signals $\mathbf{y}_k$. At every time instance t, the sensor signals consist of a mixture of S spatially located desired sources $\mathbf{A}_k\check{\mathbf{s}}[t]$ and a mixture of 3 (continuously active) spatially located interfering sources $\mathbf{B}_k\check{\mathbf{n}}[t]$. The desired sources $\check{\mathbf{s}}[t]$ have a 50% chance of being all active at a time instant t. A random noise component $\mathbf{n}_k[t]$ is also added to model independent sensor noise. The desired component $\mathbf{d}_k[t]$ is chosen as the first S sensor signals of $\mathbf{A}_k\check{\mathbf{s}}[t]$, and the prior knowledge $\mathbf{H}_k$ at each node is obtained by keeping the Q-matrix in the QR-factorization [14] of $\mathbf{A}_k$.

The observations of $\mathbf{A}_k, \check{\mathbf{s}}[t], \mathbf{B}_k, \check{\mathbf{n}}[t]$ and $\mathbf{n}_k[t]$ are in each MC run, drawn from a 0-mean-Gaussian distribution and the variances are chosen such that the average SNR over all the microphone signals in the WSN is equal to 0 dB. In each MC run, $N = 10000$ time instances are simulated to estimate (local versions of) $\mathbf{R}_{\mathbf{yy}}$ and $\mathbf{R}_{\mathbf{nn}}$.

[3]Note that in reality the algorithm will be executed in an adaptive, time-recursive manner, where each iteration is performed over a different signal segment and the same block of samples will never be broadcast again.

The upper part of Fig. 1 shows the result of the decrease in the $L2$-objective function:

$$\sum_{t,k}\frac{1}{NSK}\|\mathbf{d}_k[t] - \mathbf{d}_k^i[t]\|_2^2 \tag{25}$$

as a function of the nmber of iterations of the PK-GEVD-DANSE and GEVD-DANSE algorithm, for different sizes of S (as an average over 200 MC runs). The PK-GEVD-DANSE algorithm is able to reduce the objective function compared to the GEVD-DANSE algorithm, and the improvement is larger for larger values of S. The bottom part of Fig. 1 shows the median (over 200 MC runs) of the squared error between the centralized filter $\check{\mathbf{W}}_k$ and the local filter $\tilde{\mathbf{W}}_k^i$ (converted to a centralized filter via the compression matrices $\mathbf{\Pi}$ and $\mathbf{\Lambda}$) averaged over all the nodes and the S entries. This same is done for the GEVD-DANSE algorithm. It is observed that the convergence speed is higher for higher values of S and higher then the convergence of the GEVD-DANSE algorithm with the same value of S, due to the fact that nodes in PK-GEVD-DANSE receive more compressed signals ($2S$) from the other nodes and have by consequence more degrees of freedom to solve their local optimization problem better.

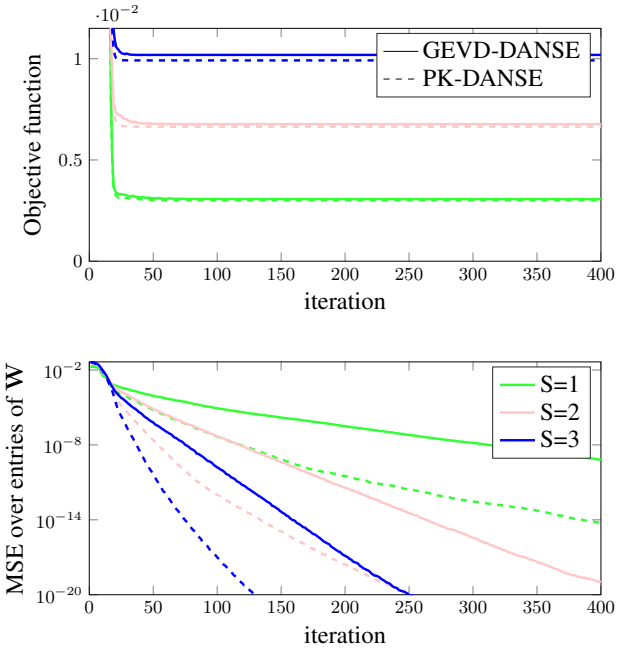

Fig. 1. Convergence properties of the PK-GEVD-DANSE compared with GEVD-DANSE and the centralized PK-GEVD-MWF.

5. CONCLUSIONS

In this paper, the centralized PK-GEVD-MWF has been obtained as an extension to the centralized GEVD-based MWF by introducing prior knowledge on the local steering matrix subspace as an extra constraint. Also a distributed round-robin algorithm has been presented to show that the output of this filter can be computed in a fully-connected WSN in a distributed way. Instead of transmitting all the sensor signals, each node broadcasts a compressed version of its sensor signals, reducing the communication and computational cost, compared to the centralized approach. The algorithm has been validated by means of numerical simulations.

6. REFERENCES

[1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, May 2001, vol. 4, pp. 2033–2036 vol.4.

[2] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, Aug 2007.

[3] I. D. Schizas, G. B. Giannakis, and Z. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4284–4299, Aug 2007.

[4] A. Bertrand and M. Moonen, "Distributed adaptive estimation of correlated node-specific signals in a fully connected sensor network," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 2053–2056.

[5] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks—part i: Sequential node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5277–5291, Oct 2010.

[6] A. Hassani, A. Bertrand, and M. Moonen, "GEVD-based low-rank approximation for distributed adaptive node-specific signal estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 64, no. 10, pp. 2557–2572, May 2016.

[7] R. Ali, G. Bernardi, T. van Waterschoot, and M. Moonen, "Methods of extending a gsc with exteral microphones," Tech. Rep., KU Leuven, Belgium, 2018, ftp://ftp.esat.kuleuven.be/pub/SISTA/rali/Reports/18-125.pdf.

[8] M. W. Hansen, J. R. Jensen, and M. G. Christensen, "Localizing near and far field acoustic sources with distributed microhone arrays.," in *ACSSC*, 2014, pp. 491–495.

[9] H. Teutsch, *Modal array signal processing: principles and applications of acoustic wavefield decomposition*, vol. 348, Springer, 2007.

[10] S. Doclo and M. Moonen, "Gsvd-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sep 2002.

[11] C. Dai, L. Luo, H. Peng, and Q. Sun, "A method based on support vector machine for voice activity detection on isolated words," in *2018 13th International Conference on Computer Science Education (ICCSE)*, Aug 2018, pp. 1–4.

[12] S. M. R. Nahar and A. Kai, "Robust voice activity detector by combining sequentially trained deep neural networks," in *2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA)*, Aug 2016, pp. 1–5.

[13] R. Serizel, M. Moonen, B. Van Dijk, and J. Wouters, "Low-rank approximation based multichannel wiener filter algorithms for noise reduction with application in cochlear implants," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 785–799, April 2014.

[14] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3, JHU Press, 2012.

[15] M. Dendrinos, S. Bakamidis, and G. Carayannis, "Speech enhancement from noise: A regenerative approach," *Speech Communication*, vol. 10, no. 1, pp. 45 – 57, 1991.

[16] B. R. Breed and J. Strauss, "A short proof of the equivalence of lcmv and gsc beamforming," *IEEE Signal Processing Letters*, vol. 9, no. 6, pp. 168–169, June 2002.