Distributed Adaptive Node-Specific Signal Estimation in a Wireless Sensor Network with Partial Prior Knowledge of the Desired Source Steering Vector

Robbe Van Rompaey
Dept. of Electrical Engineering-ESAT, STADIUS
KU leuven
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
robbe.vanrompaey@esat.kuleuven.be

Marc Moonen
Dept. of Electrical Engineering-ESAT, STADIUS
KU leuven
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
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 (some of) the nodes have partial prior knowledge of the desired source steering vector. 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 applications such as speech enhancement in an acoustic sensor network, where (some of) the nodes nodes have prior knowledge on the location of the desired speech source and on their local microphone array geometry or have access to clean noise reference signals.**

***Index Terms*—Wireless Sensor Networks (WSN), distributed estimation, multichannel Wiener filter (MWF), generalized eigenvalue decomposition (GEVD).**

I. Introduction

In a wireless sensor network (WSN) [1], nodes aim to combine their sensor signals with (possibly compressed) sensor signals of other nodes in an optimal way to perform a task at hand, such as the estimation of a node-specific desired signal. This generally leads to superior estimation performance compared to that of the stand-alone estimation, where each node uses only local sensor signals. The goal for every node is to obtain the same performance as if all the sensor signals were collected in a fusion center (FC) [2], [3], but in a distributed fashion while minimizing the local computations and communication with the other nodes [4], [5].

Node-specific signal estimation is considered here, where the different node-specific desired signals are assumed to be dependent on a common desired source signal. The algorithms in [4], [5] exploit this common signal subspace, to significantly compress the sensor signals that are communicated between the nodes, without compromising performance. To construct the corresponding signal correlation matrix, the algorithms assume to have access to the activity (on-off) pattern of the desired source signal. However in low SNR scenarios, this might result in a poor estimation of the signal correlation matrix, deteriorating the node-specific signal estimation performance [6]. Inspired by Ali et al. [7], a scenario is considered in this paper where (some of) the nodes have partial prior knowledge of the desired source steering vector, which can, for instance, in an acoustic scenario be obtained if nodes have prior knowledge on the location of the desired speech source and on their local microphone array geometry [8] or have access to clean noise reference signals [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 WSN, where (some of) the nodes have partial prior knowledge of the desired source steering vector. 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 amount of compressed sensor signals communicated by a node that has prior knowledge, will be twice the amount needed in previous algorithms [4], [5], since extra compressed sensor signals are needed to propagate the prior knowledge to all the other nodes. Still the signal estimation task is enhanced with this prior knowledge, justifying this extra communication.

The paper is organized as follows. The problem formulation and the centralized approach to the node specific signal estimation problem with prior knowledge are presented in Section II. In Section III the distributed algorithm is presented. In Section IV batch-mode simulations are provided to show convergence

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.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

of the proposed distributed algorithm. Conclusions are given in Section V.

II. Problem formulation and PK-GEVD-MWF

A. Node-specific signal estimation

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

$$\mathbf{y}_k = \mathbf{s}_k + \mathbf{n}_k = \mathbf{a}_k \breve{s} + \mathbf{n}_k \quad (1)$$

where $\breve{s}$ is a latent complex-valued signal representing the desired source signal, $\mathbf{a}_k$ is an (for the time being) unknown M_k-dimensional complex-valued steering vector and $\mathbf{n}_k$ is an additive noise signal that can be correlated with other noise signals in the WSN. Define also the centralized M-dimensional signals $\mathbf{y}, \mathbf{s}, \mathbf{n}$ and the centralized M-dimensional steering vector $\mathbf{a}$ as the stacked version of $\mathbf{y}_k, \mathbf{s}_k, \mathbf{n}_k$ and $\mathbf{a}_k$ respectively, where $M = \sum_{k=1}^{K} M_k$. Then (1) can be extended to

$$\mathbf{y} = \mathbf{s} + \mathbf{n} = \mathbf{a}\breve{s} + \mathbf{n}. \quad (2)$$

The node-specific task of each node $k \in \mathcal{K}$ is to find an estimate of the desired signal d_k, defined w.l.o.g. as the desired source signal component in the node's first channel:

$$d_k = [1 \ \mathbf{0}] \ \mathbf{s}_k = \mathbf{e}_{d_k}^H \mathbf{s} \quad (3)$$

where ^{H} denotes the conjugate transpose operator, $\mathbf{0}$ is an all-zero matrix with matching dimensions and $\mathbf{e}_{d_k} = [\mathbf{0} \ 1 \ \mathbf{0}]^H$ selects the correct desired source signal component in $\mathbf{s}$. Each node estimates its desired signal d_k as a linear combination of all the sensor signals $\mathbf{y}$ by minimizing the following mean squared error (MSE) criterion:

$$\boxed{\breve{\mathbf{w}}_k = \arg\min_{\mathbf{w}_k} E\{\|d_k - \mathbf{w}_k^H \mathbf{y}\|^2\}} \quad (4)$$

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

$$\boxed{\breve{\mathbf{w}}_k = \boldsymbol{R}_{\mathbf{yy}}^{-1} \boldsymbol{R}_{\mathbf{y}d_k} = \boldsymbol{R}_{\mathbf{yy}}^{-1} \boldsymbol{R}_{\mathbf{ys}} \mathbf{e}_{d_k} = \boldsymbol{R}_{\mathbf{yy}}^{-1} \boldsymbol{R}_{\mathbf{ss}} \mathbf{e}_{d_k}} \quad (5)$$

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

[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 also applicable for speech enhancement.

B. Centralized prior knowledge GEVD-based $\boldsymbol{R}_{\mathbf{ss}}$ estimation

If the desired 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], a distinction can be made between the *signal+noise* correlation matrix $\boldsymbol{R}_{\mathbf{yy}}$ and *noise-only* correlation matrix $\boldsymbol{R}_{\mathbf{nn}} = E\{\mathbf{n}\mathbf{n}^H\}$. 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 and will be denoted by $\mathbf{R}_{\mathbf{yy}}$ and $\mathbf{R}_{\mathbf{nn}}$ respectively.

$\boldsymbol{R}_{\mathbf{ss}}$ can then be estimated from $\mathbf{R}_{\mathbf{ss}} = \mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}}$. However such an estimate has mostly a rank larger than 1, especially in low SNR scenarios [6], so that a better correlation matrix estimation method is needed.

There exist different signal correlation matrix estimation methods [6], but recently Ali et al. [7] have introduced a signal correlation matrix estimation method, where the on-off behavior of the desired source signal is exploited and partial prior knowledge of the desired source steering vector is taken into account.

Extending this method in the WSN context, one can consider a scenario where a node $k \in \mathcal{K}$ has prior knowledge on the subspace of the steering vector subspace $\mathbf{a}_k$, represented by a unitary $M_k \times L_k$ subspace matrix $\mathbf{H}_k$. An example scenario is presented in the simulations in Section IV. Denote the orthogonal complement to the column space of $\mathbf{H}_k$ as the column space of the unitary $M_k \times (M_k - L_k)$ blocking matrix $\mathbf{B}_k$, such that $\mathbf{H}_k^H \mathbf{B}_k = \mathbf{0}$. Stacking these subspace matrices and blocking matrices in one centralized subspace matrix and blocking matrix respectively results in

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

where $\mathbf{H}$ is a block-diagonal $M \times L$ dimensional matrix and $\mathbf{B}$ a block-diagonal $M \times (M - L)$ dimensional matrix with $L = \sum_{k=1}^{K} L_k$. Here $M - L$ is representative of how much prior knowledge is available, summed over all the nodes. One extreme case is the case where node k does not have any prior knowledge, then $\mathbf{H}_k = \mathbf{I}_{M_k}$ (the $M_k \times M_k$ identity matrix) and $\mathbf{B}_k = [\ ]$ and so $M_k - L_k = 0$ ('zero prior knowledge'). The other extreme case is where node k knows its steering vector $\mathbf{a}_k$ (up to a scalar α) , then $\mathbf{H}_k = \alpha \mathbf{a}_k$ and so $M_k - L_k = M_k - 1$. An in-between case is where some of the sensor signals in node k are known to be clean noise reference signals, for instance when the first signal is a clean noise reference signal, then

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{M_k - 1} \end{bmatrix}; \mathbf{B}_k = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix} \quad (7)$$

and so $M_k - L_k = 1$.

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

$$\boxed{\underset{\substack{\text{rank}(\mathbf{R}_{\mathbf{ss}})=1 \\ \mathbf{B}^H \mathbf{R}_{\mathbf{ss}} \mathbf{B} = \mathbf{0}}}{\arg\min} \ \| \mathbf{R}_{\mathbf{nn}}^{-1/2} \left(\mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}} - \mathbf{R}_{\mathbf{ss}} \right) \mathbf{R}_{\mathbf{nn}}^{-H/2} \|_F^2} \quad (8)$$

where $||\ .\ ||_F$ denotes the Frobenius norm. Here $\mathbf{R_{ss}}$ is constrained to be rank 1, the column and row space of $\mathbf{R_{ss}}$ are constrained to lie in the column space of $\mathbf{H}$ and approximation errors are considered relative to the estimated noise correlation matrix $\mathbf{R_{nn}}$ (cfr. the pre- and post-multiplication with the Cholesky factor of $\mathbf{R_{nn}^{-1}}$).

The solution (proof omitted) to (8) is based on the GEVD [10], [12] of a reduced $L \times L$ dimensional matrix pencil $\{\mathbf{R_{\hat{y}\hat{y}}}, \mathbf{R_{\hat{n}\hat{n}}}\}$:

$$\begin{aligned}\mathbf{R_{\hat{y}\hat{y}}} &= \hat{\mathbf{Q}}\mathbf{\Sigma_{\hat{y}\hat{y}}}\hat{\mathbf{Q}}^H \\ \mathbf{R_{\hat{n}\hat{n}}} &= \hat{\mathbf{Q}}\mathbf{\Sigma_{\hat{n}\hat{n}}}\hat{\mathbf{Q}}^H\end{aligned} \tag{9}$$

where $\mathbf{R_{\hat{y}\hat{y}}}$ and $\mathbf{R_{\hat{n}\hat{n}}}$ are defined in the next paragraph. Here, $\hat{\mathbf{Q}} = \hat{\mathbf{X}}^{-H}$ is an invertible matrix, the columns of $\hat{\mathbf{X}}$ are unique up to a scalar multiplication and define the generalized eigenvectors. $\mathbf{\Sigma_{\hat{y}\hat{y}}}$ and $\mathbf{\Sigma_{\hat{n}\hat{n}}}$ are real-valued diagonal matrices where $\mathbf{\Sigma_{\hat{y}\hat{y}}} = \text{diag}\{\hat{\sigma}_{y_1}, .., \hat{\sigma}_{y_L}\}, \mathbf{\Sigma_{\hat{n}\hat{n}}} = \text{diag}\{\hat{\sigma}_{n_1}, .., \hat{\sigma}_{n_L}\}$ define the generalized eigenvalues sorted from high to low $\{\hat{\sigma}_{y_i}/\hat{\sigma}_{n_i}\}$ ratio.

The reduced $L \times L$ dimensional correlation matrices $\{\mathbf{R_{\hat{y}\hat{y}}}, \mathbf{R_{\hat{n}\hat{n}}}\}$ can be determined by first performing an LCMV-beamforming on the sensor signals $\mathbf{y}$, defined by the following LCMV-criterion:

$$\begin{aligned}\mathbf{C} = \arg\min_{\mathbf{C}} \quad & \text{trace}\{\mathbf{C}^H\mathbf{R_{nn}}\mathbf{C}\} \\ s.t. \quad & \mathbf{H}^H\mathbf{C} = \mathbf{I}_L\end{aligned} \tag{10}$$

where C is an $M \times L$ matrix, of which every column represents a specific LCMV-beamformer. The solution, here based on a GSC-implementation [13], is given by

$$\mathbf{C} = \mathbf{H} - \mathbf{BF} \tag{11}$$

$$\mathbf{F} = (\mathbf{B}^H\mathbf{R_{nn}}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R_{nn}}\mathbf{H}. \tag{12}$$

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

The optimal solution for $\mathbf{R_{ss}}$ of (8) is finally given by

$$\check{\mathbf{R}}_{\mathbf{ss}} = \mathbf{H}\hat{\mathbf{Q}}\text{diag}\{\hat{\sigma}_{y_1} - \hat{\sigma}_{n_1}, 0, ..., 0\}\hat{\mathbf{Q}}^H\mathbf{H}^H. \tag{13}$$

C. Centralized prior knowledge GEVD-based MWF

Substituting estimate (13) in (5) and using $\mathbf{R_{yy}} = \mathbf{R_{nn}} + \check{\mathbf{R}}_{\mathbf{ss}}$, after some manipulations, results in

$$\boxed{\check{\mathbf{w}}_k = \mathbf{C}\hat{\mathbf{W}}_{GEVD}\mathbf{H}^H\mathbf{e}_{d_k}} \tag{14}$$

where

$$\hat{\mathbf{W}}_{GEVD} = \hat{\mathbf{X}}\text{diag}\{\frac{\hat{\sigma}_{y_1} - \hat{\sigma}_{n_1}}{\hat{\sigma}_{y_1}}, 0, ..., 0\}\hat{\mathbf{Q}}^H \tag{15}$$

is the GEVD-based MWF [5], [6] that estimates $\hat{\mathbf{s}} = \mathbf{C}^H\mathbf{s}$ from $\hat{\mathbf{y}}$.

The filter obtained in (14) 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-beamformers (10), the second block is a full GEVD-based MWF and the last block is a selection and scaling part, specific to node k, to estimate the desired signal d_k.

To determine the centralized PK-GEVD-MWF, the 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 a FC. This will require a large communication bandwidth, and furthermore, as these correlation matrices are large, the inversion of $\mathbf{B}^H\mathbf{R_{nn}}\mathbf{B}$ in (12) and the GEVD in (9) will require significant computational power at the FC.

To overcome this complexity problem, a distributed adaptive estimation algorithm is presented in the next section where nodes only broadcast 2 compressed sensor signals and the computations in each node are performed on a smaller number of signals[2], i.e. only the local sensor signals 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 referred to as the Prior Knowledge GEVD-based Distributed Adaptive Node Specific Signal Estimation (PK-GEVD-DANSE) algorithm. A drawback of the PK-GEVD-DANSE algorithm is the slower adaptation and tracking speed compared to the centralized algorithm, due to the block-iterative nature of the algorithm.

III. PK-GEVD-DANSE ALGORITHM

A. Algorithm description

In the PK-GEVD-DANSE algorithm, each node k communicates 2 compressed sensor signals instead of the full M_k-dimensional sensor signal $\mathbf{y}_k$, namely:

- the signal $z_k = \mathbf{p}_k^H\mathbf{y}_k$ where the M_k-dimensional compression vector $\mathbf{p}_k$ corresponds to the current estimate of the MWF coefficients corresponding to the local sensor signals;
- the signal $\underline{z}_k = \boldsymbol{\lambda}_k^H\mathbf{B}_k^H\mathbf{y}_k$ corresponding to a compressed version of the local noise references $\mathbf{B}_k^H\mathbf{y}_k$, where the $(M_k - L_k)$-dimensional compression vector $\boldsymbol{\lambda}_k$ will be defined later.

Consequently, each node k has access to reduced-dimensional sensor signals $\tilde{\mathbf{y}}_k = [\mathbf{y}_k^H\mathbf{z}_{-k}^H\underline{\mathbf{z}}_{-k}^H]^H$ where the subscript $-k$ refers to the concatenation of the compressed sensor signals of the other nodes: $\mathbf{z}_{-k}^H = [z_1^H...z_{k-1}^H z_{k+1}^H...z_K^H]^H$ and $\underline{\mathbf{z}}_{-k}^H = [\underline{z}_1^H...\underline{z}_{k-1}^H\underline{z}_{k+1}^H...\underline{z}_K^H]^H$.

To be able to perform the same operations as in the centralized PK-GEVD-MWF on these reduced-dimensional sensor signals, the following subspace matrix $\tilde{\mathbf{H}}_k$ and corresponding blocking matrix $\tilde{\mathbf{B}}_k$ are defined:

$$\tilde{\mathbf{H}}_k = \begin{bmatrix}\mathbf{H}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{K-1} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}; \quad \tilde{\mathbf{B}}_k = \begin{bmatrix}\mathbf{B}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{K-1}\end{bmatrix}. \tag{16}$$

[2] In an iteration of Algorithm 1 in Section III, the inversion of a reduced-dimensional matrix $\tilde{\mathbf{B}}_q^H\mathbf{R}_{\tilde{\mathbf{n}}_q\tilde{\mathbf{n}}_q}\tilde{\mathbf{B}}_q$ and the GEVD of a reduced dimensional matrix pencil $\{\mathbf{R}_{\hat{\mathbf{y}}_q\hat{\mathbf{y}}_q}, \mathbf{R}_{\hat{\mathbf{n}}_q\hat{\mathbf{n}}_q}\}$ are needed.

The steering vector subspace for $\mathbf{y}_k$ is defined by $\mathbf{H}_k$ and so represented by $\mathbf{H}_k$ and $\mathbf{B}_k$ in $\tilde{\mathbf{H}}_k$ and $\tilde{\mathbf{B}}_k$ respectively. The steering vector subspace for $\mathbf{z}_{-k}$ is unknown and thus represented by $\mathbf{I}_{K-1}$ in $\tilde{\mathbf{H}}_k$ and by $\left[\ \right]$ in $\tilde{\mathbf{B}}_k$. The steering vector subspace for $\underline{\mathbf{z}}_{-k}$ is empty because these signals are compressed versions of local noise references ($\mathbf{B}_k^H \mathbf{y}_k = \mathbf{B}_k^H \mathbf{a}_k \breve{s} + \mathbf{B}_k^H \mathbf{n}_k = \mathbf{B}_k^H \mathbf{n}_k$), where the signal component is already locally canceled and is represented by $\left[\ \right]$ in $\tilde{\mathbf{H}}_k$ and by $\mathbf{I}_{K-1}$ in $\tilde{\mathbf{B}}_k$.

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 in the centralized algorithm, but here with locally defined reduced-dimensional variables $\mathbf{F}_k, \mathbf{C}_k, \hat{\mathbf{W}}_{\text{GEVD},k}$ and $\tilde{\mathbf{w}}_k$. The definition of the local compression matrix $\mathbf{p}_k^i$ is given in (22) and indeed corresponds to the current estimate of the MWF coefficients $\tilde{\mathbf{w}}_k^i$ corresponding to 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 vector $\boldsymbol{\lambda}_k^i$ of the local noise references $\mathbf{B}_k^H \mathbf{y}_k$ in (21). A proof of convergence showing that Algorithm 1 converges to the PK-GEVD-MWF (14) for any random initialization of the compression matrices, will be provided elsewhere.

Algorithm 1: PK-GEVD-DANSE algorithm

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

2 - All nodes $k \in \mathcal{K}$ broadcast N compressed observations of $z_k = \mathbf{p}_k^{iH} \mathbf{y}_k$ and $\underline{z}_k = \boldsymbol{\lambda}_k^{iH} \mathbf{B}_k^H \mathbf{y}_k$ and construct locally

$$\tilde{\mathbf{y}}_k = [\mathbf{y}_k^H \mathbf{z}_{-k}^H \underline{\mathbf{z}}_{-k}^H]^H. \tag{17}$$

3 - Node q estimates $\mathbf{R}_{\tilde{\mathbf{n}}_q \tilde{\mathbf{n}}_q}$ based on the observations and updates its local LCMV-beamformer:

$$\mathbf{F}_q^{i+1} = \left(\tilde{\mathbf{B}}_q^H \mathbf{R}_{\tilde{\mathbf{n}}_q \tilde{\mathbf{n}}_q} \tilde{\mathbf{B}}_q\right)^{-1} \tilde{\mathbf{B}}_q^H \mathbf{R}_{\tilde{\mathbf{n}}_q \tilde{\mathbf{n}}_q} \tilde{\mathbf{H}}_q \tag{18}$$

$$\mathbf{C}_q^{i+1} = \tilde{\mathbf{H}}_q - \tilde{\mathbf{B}}_q \mathbf{F}_q^{i+1}, \quad \hat{\mathbf{y}}_q = \mathbf{C}_q^{i+1^H} \tilde{\mathbf{y}}_q. \tag{19}$$

- Node q estimates $\mathbf{R}_{\hat{\mathbf{y}}_q \hat{\mathbf{y}}_q}$ and $\mathbf{R}_{\hat{\mathbf{n}}_q \hat{\mathbf{n}}_q}$ based on the observations and constructs $\hat{\mathbf{W}}_{\text{GEVD},q}^{i+1}$ as in (15) using the GEVD of $\{\mathbf{R}_{\hat{\mathbf{y}}_q \hat{\mathbf{y}}_q}, \mathbf{R}_{\hat{\mathbf{n}}_q \hat{\mathbf{n}}_q}\}$ and updates its local variables:

$$\tilde{\mathbf{w}}_q^{i+1} = \mathbf{C}_q^{i+1} \hat{\mathbf{W}}_{\text{GEVD},q}^{i+1} \tilde{\mathbf{H}}^H [1 \quad \mathbf{0}]^H \tag{20}$$

$$\boldsymbol{\lambda}_q^{i+1} = \left[\mathbf{I}_{M_q - L_q} \quad \mathbf{0}\right] \left(\tilde{\mathbf{B}}_q^H \mathbf{R}_{\tilde{\mathbf{n}}_q \tilde{\mathbf{n}}_q} \tilde{\mathbf{B}}_q\right)^{-1} \tilde{\mathbf{B}}_q^H \mathbf{R}_{\tilde{\mathbf{y}}_q \tilde{\mathbf{y}}_q} \tilde{\mathbf{w}}_q^{i+1} \tag{21}$$

$$\mathbf{p}_q^{i+1} = \left[\mathbf{I}_{M_q} \quad \mathbf{0}\right] \tilde{\mathbf{w}}_q^{i+1}. \tag{22}$$

- All other nodes do not change their variables:

$$\tilde{\mathbf{w}}_q^{i+1} = \tilde{\mathbf{w}}_q^i, \boldsymbol{\lambda}_k^{i+1} = \boldsymbol{\lambda}_k^i, \mathbf{p}_k^{i+1} = \mathbf{p}_k^i. \tag{23}$$

4 - For the N new observations, each node $k \in \mathcal{K}$ generates an estimate of its desired signal ${d_k}^i \approx \tilde{\mathbf{w}}_q^{i+1^H} \tilde{\mathbf{y}}_q$.

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

B. Comparison with GEVD-DANSE

A related algorithm to PK-GEVD-DANSE is GEVD-DANSE [5], which also aims to estimate the centralized GEVD-based MWF in a distributed way, but without the ability to introduce prior knowledge. GEVD-DANSE only requires 1 signal to be communicated per node, compared to 2 signals for PK-GEVD-DANSE. The extra communicated signal of PK-GEVD-DANSE is a compressed version of the local noise references $\mathbf{B}_k^H \mathbf{y}_k$. From simulations it is observed that, upon convergence of PK-GEVD-DANSE, $\boldsymbol{\lambda}_k$ is equal to the corresponding part in its centralized variant $\boldsymbol{\lambda}^k = \left(\mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbf{B}\right)^{-1} \mathbf{B}^H \mathbf{R}_{\mathbf{yy}} \breve{\mathbf{w}}_k$. This can be shown to be the optimal compression of the noise references $\boldsymbol{\lambda}^H \mathbf{B}^H \mathbf{y}$ to still be able to let a similar procedure as in Section II, attain the same PK-GEVD-MWF (14) as when all the noise references $\mathbf{B}^H \mathbf{y}$ are used. One can also show that in the case where $\mathbf{B}^H \mathbf{R}_{\mathbf{yy}} \mathbf{B}$ is exactly equal to $\mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbf{B}$, $\boldsymbol{\lambda}^k$ and so $\boldsymbol{\lambda}_k$ become 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 version of the local noise references thus accounts for estimation mismatch in $\mathbf{R}_{\mathbf{yy}}$ and $\mathbf{R}_{\mathbf{nn}}$, since in the ideal case $\mathbf{B}^H \mathbf{R}_{\mathbf{yy}} \mathbf{B}$ should be equal to $\mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbf{B}$. 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 extra communication.

IV. Simulations

To demonstrate the convergence and optimality of the PK-GEVD-DANSE algorithm, the following scenario is used. The scenario consists of 4 nodes with each $M_k = 10$ sensor signals, one desired source $\breve{s}[t]$ and 5 undesired noise sources $\breve{\mathbf{n}}[t]$. Node 1 and node 2 have access to an exact[3] (normalized) estimate of their steering vector $\mathbf{a}_1$ and $\mathbf{a}_2$, denoted by $\mathbf{H}_1$ and $\mathbf{H}_2$ respectively. These are for instance linear microphone arrays or binaural hearing aids with the desired source in their broadside direction. The first 3 sensor signals of node 3 correspond directly to 3 of the 5 undesired noise sources. This is for example the case if node 3 has access to a signal that is played by a loudspeaker present in the scene. Finally, node 4 has the prior knowledge that its first 3 sensors do not observe the desired source, which is for example located at the other side of a signal dampening wall, but they do observe the 5 noise sources. The prior knowledge of node 3 and 4 can thus be captured by the following unitary matrix:

$$\mathbf{H_3} = \mathbf{H_4} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{7\times 7} \end{bmatrix}. \tag{24}$$

For simplicity the PK-GEVD-MWF algorithm is run in batch-mode[4] for a single frequency bin (in the case of speech

[3]The influences of estimation errors in the prior knowledge is part of future work.

[4]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.

signals). Monte-Carlo (MC) simulations are conducted and compared with the convergence of GEVD-DANSE and the output of the centralized eigenvalue decomposition based-MWF (EVD-MWF), where the best rank 1 approximation of $\mathbf{R_{yy}} - \mathbf{R_{nn}}$ is used to approximate $\mathbf{R_{ss}}$ [6], the centralized GEVD-MWF and the centralized PK-GEVD-MWF. In every MC run and $\forall k \in \{1,2,3,4\}$, a new random desired speech source steering vector $\mathbf{a}_k$ and new random undesired noise steering vectors $\mathbf{D}_k$ are generated from a 0-mean complex Gaussian distribution with variance 1, with the following constraints to satisfy the scenario. The first 3 components of $\mathbf{a}_3$ and $\mathbf{a}_4$ are always equal to zero and the first 3 rows of $\mathbf{D}_3$ are $\begin{bmatrix}\mathbf{I}_{3\times3} & \mathbf{0}\end{bmatrix}$. Also $N = 1000$ samples of $\breve{s}[t]$ (being active for 50% of the time), $\breve{\mathbf{n}}[t]$ and a random noise component $\mathbf{n}_k[t]$ (to model sensor noise) are generated to create the sensor signals

$$\mathbf{y}_k[t] = \mathbf{a}_k\breve{s}[t] + \mathbf{D}_k\breve{\mathbf{n}}[t] + \mathbf{n}_k[t] \quad \forall k \in \{1,2,3,4\} \tag{25}$$

by drawing them from a 0-mean complex Gaussian distribution. The variances are chosen such that the average SNR over all the sensors observing the desired source, is equal to 0 dB.

The upper part of Fig. 1 shows the median (over 200 MC runs) of the decrease in the $L2$-objective function:

$$\sum_{t,k} \frac{1}{NK} \|d_k[t] - d_k^i[t]\|_2^2 \tag{26}$$

as a function of the number of iterations of the PK-GEVD-DANSE and shows the result when the centralized EVD-MWF, GEVD-MWF and PK-GEVD-MWF are used to estimate $d_k[t]$. The GEVD-MWF is able to reduce the objective function compared to the EVD-MWF, but the addition of the prior knowledge to obtain the centralized PK-GEVD-MWF reduces the objective even further. The bottom part of Fig. 1 shows the median (over 200 MC runs) of the squared error between the centralized filter $\breve{\mathbf{w}}_k$ and the local filter $\tilde{\mathbf{w}}_k^i$ (converted to a centralized filter via the compression vectors $\mathbf{p}_k^i$ and $\boldsymbol{\lambda}_k^i$) averaged over all the nodes. This same is done for the GEVD-DANSE algorithm. Convergence to the machine machine precision is observed. The convergence speed of PK-GEVD-DANSE is higher then the convergence of the GEVD-DANSE algorithm, due to the fact that nodes in PK-GEVD-DANSE receive more compressed signals from the other nodes and have by consequence more degrees of freedom to solve their local optimization problem better.

V. Conclusions

In this paper, the centralized PK-GEVD-MWF has been derived as an extension to the centralized GEVD-based MWF by introducing partial prior knowledge of the desired source steering vector. 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 communicating all the sensor signals, each node communicates 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.

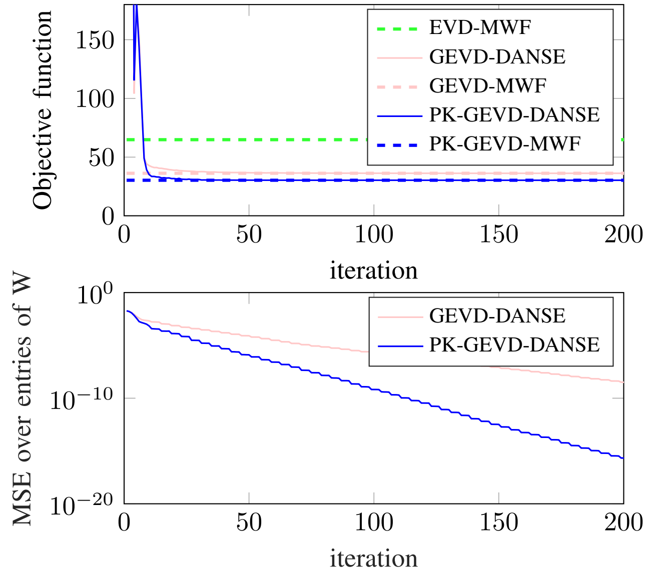

Fig. 1. Convergence properties of PK-GEVD-DANSE compared with GEVD-DANSE and the centralized EVD-MWF, GEVD-MWF and PK-GEVD-MWF.

References

[1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *ICASSP*, May 2001, vol. 4, pp. 2033–2036.

[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 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.

[5] 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.

[6] 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.

[7] R. Ali, G. Bernardi, T. van Waterschoot, and M. Moonen, "Methods of extending a generalised sidelobe canceller with exteral microphones," *IEEE/ACM Trans Audio, Speech, Lang. Process*, to appear, 2019.

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

[9] A. Geusens, A. Bertrand, B. Cornelis, and M. Moonen, "Multi-channel noise reduction in hearing aids with wireless access to an external reference signal," in *IWAENC*, Aachen, Germany, Sep. 2012.

[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] S. M. R. Nahar and A. Kai, "Robust voice activity detector by combining sequentially trained deep neural networks," in *CAICTA*, Aug 2016, pp. 1–5.

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

[13] 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.