This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

Distributed Adaptive Signal Estimation in Wireless Sensor Networks with Partial Prior Knowledge of the Desired Sources Steering Matrix

Robbe Van Rompaey, Marc Moonen

***Abstract*— In the past two decades, wireless sensor networks (WSNs) and their applications have been the topic of many studies. Different multi-sensor nodes are used to collect, process and distribute data over wireless links to perform different tasks such as smart detection, target tracking, node localization, etc. In this paper, the problem of distributed adaptive estimation of node-specific signals for signal enhancement or noise reduction is addressed. First the centralized rank R generalized eigenvalue decomposition (GEVD) based multichannel Wiener filter (MWF) with prior knowledge for node-specific signal estimation in a WSN is introduced, where (some of) the nodes have partial prior knowledge of the desired sources steering matrix. 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 algorithm. The distributed algorithm can be used in applications such as speech enhancement in a wireless acoustic sensor network (WASN), where (some of) the nodes have prior knowledge on the location of the desired speech sources and on their local microphone array geometry or have access to clean noise reference signals. Foundations for a proof of convergence using a Lagrangian framework, are given, since convergence is observed in batch-mode simulations. Finally, numerical simulation results are provided for a speech enhancement scenario.**

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

I. Introduction

In a wireless sensor network (WSN) [1]–[3], nodes aim to combine their sensor signals with (possibly compressed) sensor signals of other nodes in some optimal way to perform a task at hand, such as the estimation of a node-specific desired signal [4], smart detection [5], target tracking [6], node localization [7], etc... This generally leads to superior estimation performance compared to the stand-alone performance, where each node uses only local sensor signals. The goal for every node could be to obtain the same estimation performance as if all the sensor signals were collected in a fusion center (FC) [8], [9], but then in a distributed fashion while minimizing the local computations and communications with the other nodes [10]–[12].

Node-specific signal estimation is considered here, where the different node-specific desired signals are assumed to be dependent on the same desired sources. The so-called distributed adaptive node-specific signal estimation algorithm (DANSE) and GEVD-DANSE in [10], [11] 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 sources. However in low SNR scenarios, this might result in a poor estimation of the signal correlation matrix, deteriorating the node-specific signal estimation performance [13]. Inspired by Ali et al. [14], [15], a scenario is considered in this paper where (some of) the nodes have partial prior knowledge of the desired sources steering matrix, which can, for instance, in an acoustic scenario be obtained if nodes have prior knowledge on the location of the desired speech sources and on their local microphone array geometry [16] or have access to clean noise reference signals [17]. In previous work [18], the case where the nodes' desired signals are dependent on only one desired source is investigated and the so-called prior-knowledge GEVD-DANSE (PK-GEVD-DANSE) is proposed. Here this algorithm is extended to multiple desired sources and further theoretical analysis and numerical simulations are provided.

Algorithms in [10], [11] are using either an oracle to find the signal correlation matrix or a fully data-driven approach based on a GEVD and the on-off behavior of the sources to estimate the signal correlation matrix. The introduction of local partial prior knowledge on the signal correlation matrix, which has beneficial effects on the estimation performance as illustrated by [14], [15], [19], has only been done in a centralized fashion (e.g., using a fusion center), since the partial prior knowledge is measured with respect to the noise correlation matrix and the noise is typically correlated between the nodes. However, the PK-GEVD-DANSE shows that the local partial prior knowledge can still be included in the algorithm by allowing the nodes to broadcast twice the number of compressed signals compared to [10], [11]. Still the performance is improved by exploiting this prior knowledge, justifying this extra communication.

The introduction of prior information to a WSN has already been done for different applications, like inverse source problems in acoustic wave propagation [20], [21], distributed

The work of R. Van Rompaey was supported by a doctoral Fellowship of the Research Foundation Flanders (FWO-Vlaanderen). This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of Research Council KU Leuven: C3-19-00221 "Cooperative Signal Processing Solutions for IoT-based Multi-User Speech Communication Systems", Research Council KU Leuven: C24/16/019 "Distributed Digital Signal Processing for Ad-hoc Wireless Local Area Audio Networking" and Fonds de la Recherche Scientifique - FNRS and the Fonds Wetenschappelijk Onderzoek - Vlaanderen under EOS Project no 30452698 '(MUSE-WINET) MUlti-SErvice WIreless NETwork'. The scientific responsibility is assumed by its authors.

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

optimization [19] and failure detection [22]. Distributed node-specific signal enhancement algorithms with constraints on the node-specific filters, which is mostly posed as a linearly constrained minimum variance (LCMV) beamforming problem [23], has already been investigated [12], [24]–[26]. However to construct the constraints in these approaches, it is assumed to either have access to the network-wide steering matrix or an estimate thereof. Here a different approach is considered where it is assumed that (some of) the nodes have partial prior knowledge on the local steering matrix and the relation to the network-wide steering matrix is estimated by the algorithm.

In the rest of the paper, bold lower case letters are used to denote complex vectors, bold upper case letters are used to represent complex matrices and italic upper/lower case letters are used to denote optimization variables. $\mathbf{I}_R$ is the $R \times R$ identity matrix and $\mathbf{0}$ denotes the all zeros matrix whose dimensions will be provided using subscript or will be clear from the context. $||.||$, $||.||_F$ and $E\{.\}$ denote the Euclidean norm, Frobenius norm and expected value operator respectively. Lastly tr$\{.\}$,$.^*$ and $.^H$ are used to denote the trace, conjugate and conjugate transpose operation. The paper will introduce many different symbols, so a non-exhaustive overview of the most relevant notation is given in Table I. Each symbol in the table will be further explained in the text.

The paper is organized as follows. Section II introduces the problem formulation of node-specific signal estimation and the centralized rank R 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 sources steering matrix are presented . In Section III an alternative derivation of the centralized algorithm is provided using a Lagrangian framework, where the Lagrange parameters will turn out to play an important role in the distributed algorithm. A distributed adaptive estimation algorithm for a fully-connected WSN is then proposed in Section IV demonstrating that this MWF can be obtained by letting the nodes work on compressed (i.e. reduced-dimensional) sensor signals compared to the centralized algorithm. It turns out that the number of compressed sensor signals communicated by a node that has prior knowledge, will be twice the number of communicated signals needed in previous algorithms [10], [11], since extra compressed sensor signals are needed to propagate the prior knowledge to all the other nodes. An argument for convergence is also provided in Section IV as well as the extension to a topology-independent algorithm. Some special cases for which changes to the distributed algorithm are required, are listed in Section V. Section VI provides numerical simulations to show convergence of the proposed distributed algorithm and the performance gains obtained from the availability of prior knowledge. Also the effect of estimation errors in the correlation matrices is experimentally verified. An application of the distributed algorithm to speech enhancement in a wireless acoustic sensor network (WASN) is provided.

II. Problem Formulation and rank R PK-GEVD-MWF

A. Node-specific signal estimation

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

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

where $\breve{\mathbf{s}}$ is a vector containing the R complex-valued desired source signals, $\mathbf{A}_k$ is a (for the time being) unknown $M_k \times R$-dimensional complex-valued steering matrix 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 $M \times R$-dimensional steering matrix $\mathbf{A}$ as the stacked version of $\mathbf{y}_k, \mathbf{s}_k, \mathbf{n}_k$ and $\mathbf{A}_k$ respectively, where $M = \sum_{k=1}^{K} M_k$. Then (1) can be extended to

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

The node-specific task of node k is to find an estimate of the R-dimensional desired signal $\breve{\mathbf{d}}_k$, defined w.l.o.g. as the mixture of desired source signals in the node's first R channels[1]:

$$\bar{\mathbf{d}}_k = [\mathbf{I}_R \ \mathbf{0}_{R\times(M_k-R)}] \ \mathbf{s}_k = \mathbf{E}_k^T \mathbf{s}_k = \mathbf{E}_{\mathbf{d}_k}^T \mathbf{s} \tag{3}$$

where $\mathbf{E}_{\mathbf{d}_k} = [\mathbf{0}_{R\times(M_1+...+M_{k-1})} \ \mathbf{E}_k \ \mathbf{0}_{R\times(M_{k+1}+...+M_K)}]^T$ selects the correct mixture of desired source signals in $\mathbf{s}$. Each node estimates its desired signal $\bar{\mathbf{d}}_k$ as a linear combination of all the sensor signals in $\mathbf{y}$ by minimizing the following mean squared error (MSE) criterion:

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

The resulting filter is referred to as the multichannel Wiener filter (MWF)[2]. If $\bar{\mathbf{R}}_{\mathbf{yy}} = E\{\mathbf{y}\mathbf{y}^H\}$ has full rank, the unique solution of (4) is [27]:

$$\boxed{\begin{aligned}\bar{\mathbf{W}}_k &= \bar{\mathbf{R}}_{\mathbf{yy}}^{-1}\bar{\mathbf{R}}_{\mathbf{y}\mathbf{d}_k} = \bar{\mathbf{R}}_{\mathbf{yy}}^{-1}\bar{\mathbf{R}}_{\mathbf{ys}}\mathbf{E}_{\mathbf{d}_k} \\ &= \left(\bar{\mathbf{R}}_{\mathbf{nn}} + \bar{\mathbf{R}}_{\mathbf{ss}}\right)^{-1} \bar{\mathbf{R}}_{\mathbf{ss}}\mathbf{E}_{\mathbf{d}_k}\end{aligned}} \tag{5}$$

with $\bar{\mathbf{R}}_{\mathbf{y}\mathbf{d}_k} = E\{\mathbf{y}\bar{\mathbf{d}}_k^H\}$, $\bar{\mathbf{R}}_{\mathbf{ys}} = E\{\mathbf{y}\mathbf{s}^H\}$, $\bar{\mathbf{R}}_{\mathbf{nn}} = E\{\mathbf{n}\mathbf{n}^H\}$ and $\bar{\mathbf{R}}_{\mathbf{ss}} = E\{\mathbf{s}\mathbf{s}^H\}$. The last step in (5) is allowed due to the (often valid) assumption that the additive noise signal $\mathbf{n}$ and the desired source signals in $\breve{\mathbf{s}}$ are uncorrelated. The signal correlation matrix $\bar{\mathbf{R}}_{\mathbf{ss}}$ is then given by $\mathbf{A}E\{\breve{\mathbf{s}}\breve{\mathbf{s}}^H\}\mathbf{A}^H$, where $E\{\breve{\mathbf{s}}\breve{\mathbf{s}}^H\}$ is a diagonal matrix containing the desired source signal powers. Notice that $\bar{\mathbf{R}}_{\mathbf{ss}}$ 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 $\bar{\mathbf{R}}_{\mathbf{ss}}$ is given in the next subsection, based on

[1]The reason why at least R desired signals are estimated will become clear later on. Special cases such as when $M_k < R$ or rank$([\mathbf{I}_R \ \mathbf{0}]\mathbf{A}_k) < R$ are discussed in Section V.

[2]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.

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

TABLE I: Overview of used notation.

Notation	Examples of usage	Description
$\bar{\cdot}$	$\bar{\mathbf{R}}_{\mathbf{yy}}, \bar{\mathbf{W}}_k, \bar{\mathbf{d}}_k$	Quantities related to the true (unknown) signals.
$\hat{\cdot}$	$\hat{\mathbf{y}}, \hat{\mathbf{R}}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}, \hat{\mathbf{Q}}, \hat{\mathbf{W}}, \hat{\mathbb{X}}$	Quantities related to the reduced dimensional signals after centralized LCMV-filtering.
$\cdot_k$	$\mathbf{y}_k, \mathbf{H}_k, \mathbf{W}_{kk}, \mathbf{\Lambda}_{kk}, \mathbb{X}_{kk}, \mathbb{\Lambda}_{kk}, \mathbf{z}_k$	Local quantities related to the sensor signals of node k.
$\tilde{\cdot}_k$	$\tilde{\mathbf{y}}_k, \mathbf{R}_{\tilde{\mathbf{n}}_k\tilde{\mathbf{n}}_k}, \tilde{\mathbf{H}}_k, \tilde{\mathbf{C}}_k, \tilde{\mathbf{F}}_k, \tilde{\mathbf{W}}_k, \tilde{\mathbf{\Lambda}}_{kk}, \tilde{\mathbb{X}}_k, \tilde{\mathbb{\Lambda}}_k$	Local quantities related to all the available signals at node k including compressed signals form other nodes.
$\hat{\cdot}_k$	$\hat{\mathbf{y}}_k, \mathbf{R}_{\hat{\mathbf{n}}_k\hat{\mathbf{n}}_k}, \hat{\mathbf{W}}_k, \hat{\mathbb{X}}$	Local quantities related to the reduced dimensional signals after local LCMV-filtering at node k.
$\underline{\cdot}$	$\underline{\mathbf{z}}_k$	Compressed signals related to the local noise references at node k.

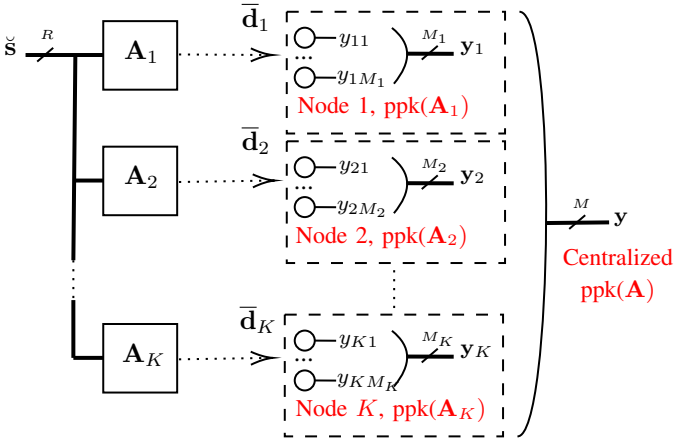

Fig. 1: Bird's eye view on system model. The network contains K nodes, where node k collects M_k sensor signals and estimates a node-specific R-dimensional desired signal $\bar{\mathbf{d}}_k$, which is a mixture of the R source signals and where ppk($\mathbf{A}_k$) denotes possible partial prior knowledge on the steering matrix $\mathbf{A}_k$ available at node k.

the exploitation of the on-off behavior of the desired source signals to first estimate $\bar{\mathbf{R}}_{\mathbf{yy}}$ and $\bar{\mathbf{R}}_{\mathbf{nn}}$ and on partial prior knowledge of the desired sources steering matrix $\mathbf{A}$. A bird's eye view on the system model is given in Figure 1.

B. Centralized rank R $\bar{\mathbf{R}}_{\mathbf{ss}}$ estimation

If the desired source signals have a simultaneous on-off behavior and the on-off detection is available, e.g. via a centralized [28], [29] or distributed [30] multi-speaker voice activity detector in speech applicationsft, a distinction can be made between the *signal+noise* and *noise-only* periods to estimate the correlation matrix $\bar{\mathbf{R}}_{\mathbf{yy}}$ and $\bar{\mathbf{R}}_{\mathbf{nn}}$ respectively. The estimation can be done 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 the estimates will be denoted by $\mathbf{R}_{\mathbf{yy}}$ and $\mathbf{R}_{\mathbf{nn}}$ respectively.

$\bar{\mathbf{R}}_{\mathbf{ss}}$ can then be estimated[3] as $\mathbf{R}_{\mathbf{ss}} = \mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}}$. However such an estimate has mostly a rank larger than R and might not even be positive definite, especially in low SNR scenarios [13], so that a better correlation matrix estimation method is needed.

[3]Note that the true signal correlation matrices are denoted by $\bar{\mathbf{R}}_{\mathbf{yy}}, \bar{\mathbf{R}}_{\mathbf{nn}}$ and $\bar{\mathbf{R}}_{\mathbf{ss}}$, the measured correlation matrices are denoted by $\mathbf{R}_{\mathbf{yy}}$ and $\mathbf{R}_{\mathbf{nn}}$ and the estimated signal correlation matrices using the measured correlation matrices is denoted by $\mathbf{R}_{\mathbf{ss}}$.

A better method to estimate the signal correlation matrix, used in the so-called GEVD-MWF, is given by solving the following optimization problem [13], [31]:

$$\mathbf{R}_{\mathbf{ss}} = \underset{\text{rank}(\boldsymbol{R}_{\mathbf{ss}})=R}{\arg\min} \parallel \mathbf{R}_{\mathbf{nn}}^{-1/2} \left(\mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}} - \boldsymbol{R}_{\mathbf{ss}}\right) \mathbf{R}_{\mathbf{nn}}^{-H/2} \parallel_F^2 \tag{6}$$

of which the solution is based on the GEVD of the matrix pencil $\{\mathbf{R}_{\mathbf{yy}}, \mathbf{R}_{\mathbf{nn}}\}$ and will be defined later on.

Recently [14], [15], [18] extended optimization problem (6) by introducing partial prior knowledge of the desired source steering vector for a single desired source signal scenario. Partial prior knowledge is here imposed by constraining the row- and column space of $\mathbf{R}_{\mathbf{ss}}$ to a given subspace.

Extending this method in the WSN context and to multiple desired source signals, one can consider a scenario where a node k has prior knowledge on the columnspace of the local steering matrix $\mathbf{A}_k$, represented by a unitary $M_k \times L_k$ subspace matrix $\mathbf{H}_k$. An example scenario is presented in Section VI. The orthogonal complement to the column space of $\mathbf{H}_k$ is then represented by a 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 and blocking matrices in one centralized subspace and blocking matrix respectively results in

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{H}_K \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{B}_K \end{bmatrix} \tag{7}$$

where $\mathbf{H}$ is a block-diagonal $M \times L$ matrix and $\mathbf{B}$ a block-diagonal $M \times (M - L)$ matrix with $L = \sum_{k=1}^{K} L_k$. Here $M - L$ is representing the amount of prior knowledge available, summed over all the nodes, with $L \geq R$ since $\mathbf{A}$ is assumed to be rank R. An extreme case is when node k does not have any prior knowledge, then $\mathbf{H}_k = \mathbf{I}_{M_k}$ and $\mathbf{B}_k = \left[\ \right]$ and so $M_k - L_k = 0$ ('zero prior knowledge'). Another extreme case is when node k knows its local steering matrix $\mathbf{A}_k$ (up to a random $R \times R$ invertible matrix $\mathbf{G}$) , then $\mathbf{H}_k = \mathbf{A}_k\mathbf{G}$ and so $M_k - L_k = M_k - R$. 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 sensor signal in node k is a clean noise reference signal, then

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{M_k - 1} \end{bmatrix}; \mathbf{B}_k = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix} \tag{8}$$

and so $M_k - L_k = 1$.

Note that this representation of the prior knowledge is very relevant for an ad-hoc WSN, since a node only has to be aware of the locations of its local sensors, but not of the location of the sensors of other nodes. So if one node changes the position

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

of its sensors, the prior knowledge of the other nodes remains unchanged.

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

$$\mathbf{R}_{\mathbf{ss}} = \underset{\substack{\text{rank}(\boldsymbol{R}_{\mathbf{ss}})=R \\ \mathbf{B}^H \boldsymbol{R}_{\mathbf{ss}} \mathbf{B}=\mathbf{0}}}{\arg\min} \| \mathbf{R}_{\mathbf{nn}}^{-1/2} \left(\mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}} - \boldsymbol{R}_{\mathbf{ss}}\right) \mathbf{R}_{\mathbf{nn}}^{-H/2} \|_F^2 . \tag{9}$$

Here $\mathbf{R}_{\mathbf{ss}}$ is constrained to be rank R, the column and row space of $\mathbf{R}_{\mathbf{ss}}$ are constrained to lie in the column space of $\mathbf{H}$ and approximation errors are considered relative to the estimated noise correlation matrix $\mathbf{R}_{\mathbf{nn}}$ (cfr. the pre- and post-multiplication with the Cholesky factor of $\mathbf{R}_{\mathbf{nn}}^{-1} = \mathbf{R}_{\mathbf{nn}}^{-H/2}\mathbf{R}_{\mathbf{nn}}^{-1/2}$).

As shown in Appendix B, the solution of (9) can equivalently be obtained by solving

$$\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}} = \underset{\text{rank}(\boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}})=R}{\arg\min} \| \mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}^{-1/2} \left(\mathbf{R}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} - \mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}} - \boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\right) \mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}^{-H/2} \|_F^2 . \tag{10}$$

The reduced $L \times L$-dimensional correlation matrices $\{\mathbf{R}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}, \mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}\}$ defined as

$$\begin{aligned} \mathbf{R}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} &= \mathbf{C}^H \mathbf{R}_{\mathbf{yy}} \mathbf{C} \\ \mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}} &= \mathbf{C}^H \mathbf{R}_{\mathbf{nn}} \mathbf{C} \end{aligned} \tag{11}$$

are determined as the correlation matrices corresponding to the compressed signal $\hat{\mathbf{y}} = \mathbf{C}^H \mathbf{y}$, where $\mathbf{C}$ can be defined using the following LCMV-criterion:

$$\mathbf{C} = \underset{\mathbf{H}^H \boldsymbol{C} = \mathbf{I}_L}{\arg\min} \text{tr}\{\boldsymbol{C}^H \mathbf{R}_{\mathbf{nn}} \boldsymbol{C}\} \tag{12}$$

where every column of $\mathbf{C}$ represents a specific LCMV-beamformer. The solution, here based on a generalized sidelobe canceller (GSC)-implementation [32], is given by

$$\mathbf{C} = \mathbf{H} - \mathbf{B}\mathbf{F} \tag{13}$$

$$\mathbf{F} = (\mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbf{B})^{-1} \mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbf{H}. \tag{14}$$

where $\mathbf{F}$ is the filter estimating the noise component in the fixed beamformer branch $\mathbf{H}^H\mathbf{n}$ from the noise references $\mathbf{B}^H\mathbf{n}$:

$$\mathbf{F} = \underset{\boldsymbol{F}}{\arg\min}\, E\{\|\mathbf{H}^H\mathbf{n} - \boldsymbol{F}^H \mathbf{B}^H \mathbf{n}\|^2\}. \tag{15}$$

$\mathbf{C}$ can be verified to satisfy (63) since $\mathbf{C}^H\mathbf{H} = \mathbf{I}_L$ and $\mathbf{C}^H\mathbf{R}_{\mathbf{nn}}\mathbf{B} = \mathbf{0}$.

The solution to (10) is based on the GEVD [27], [33] of the reduced $L \times L$-dimensional matrix pencil $\{\mathbf{R}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}, \mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}\}$:

$$\begin{aligned} \mathbf{R}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} &= \hat{\mathbf{Q}} \hat{\boldsymbol{\Sigma}} \hat{\mathbf{Q}}^H \\ \mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}} &= \hat{\mathbf{Q}} \hat{\mathbf{Q}}^H \end{aligned} \tag{16}$$

Here, $\hat{\mathbf{Q}} = \hat{\mathbf{X}}^{-H}$ is an invertible matrix, the columns of $\hat{\mathbf{X}}$ are unique (up to a sign-ambiguity) and define the generalized eigenvectors. $\hat{\boldsymbol{\Sigma}}$ is a real-valued diagonal matrix where $\hat{\boldsymbol{\Sigma}} = \text{diag}\{\hat{\sigma}_1, ..., \hat{\sigma}_L\}$ defines the generalized eigenvalues sorted from large to small.

Finally the estimate $\mathbf{R}_{\mathbf{ss}}$ of (9) is given by [34]

$$\begin{aligned} \mathbf{R}_{\mathbf{ss}} &= \mathbf{H}\hat{\mathbf{Q}}\text{diag}\{\hat{\sigma}_1 - 1, ..., \hat{\sigma}_R - 1, 0, ..., 0\}\hat{\mathbf{Q}}^H\mathbf{H}^H \\ &= \mathbf{H}\hat{\mathbb{Q}}\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\hat{\mathbb{Q}}^H\mathbf{H}^H \end{aligned} \tag{17}$$

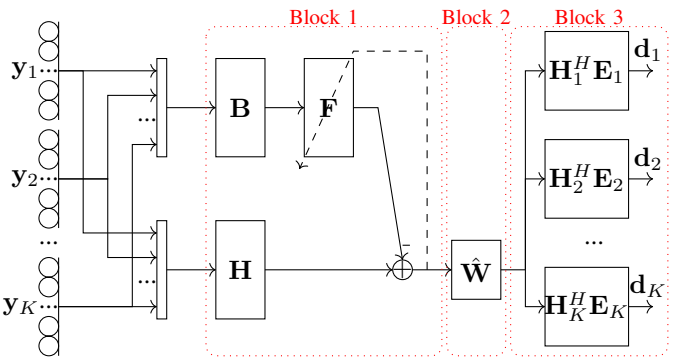

Fig. 2: Schematic overview of the rank R PK-GEVD-MWF

where $\hat{\mathbb{Q}}$ contains the first R columns of $\hat{\mathbf{Q}}$ and $\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}} = \text{diag}\{\hat{\sigma}_1 - 1, ..., \hat{\sigma}_R - 1\}$.

C. Centralized rank R PK-GEVD-MWF

Substituting estimate $\mathbf{R}_{\mathbf{ss}}$ of (17) and $\mathbf{R}_{\mathbf{nn}}$ as $\bar{\mathbf{R}}_{\mathbf{ss}}$ and $\bar{\mathbf{R}}_{\mathbf{nn}}$ respectively in (5), after some manipulations provided in Appendix C, results in the following estimate of the MWF in (4):

$$\boxed{\mathbf{W}_k = \mathbf{C}\hat{\mathbf{W}}\mathbf{H}^H\mathbf{E}_{\mathbf{d}_k}} \tag{18}$$

where

$$\begin{aligned} \hat{\mathbf{W}} &= \hat{\mathbf{X}}\text{diag}\{\frac{\hat{\sigma}_1 - 1}{\hat{\sigma}_1}, ..., \frac{\hat{\sigma}_R - 1}{\hat{\sigma}_R}, 0, ..., 0\}\hat{\mathbf{Q}}^H \\ &= \hat{\mathbb{X}}\text{diag}\{\frac{\hat{\sigma}_1 - 1}{\hat{\sigma}_1}, ..., \frac{\hat{\sigma}_R - 1}{\hat{\sigma}_R}\}\hat{\mathbb{Q}}^H \end{aligned} \tag{19}$$

is the GEVD-based MWF [11], [13] that estimates $\hat{\mathbf{s}} = \mathbf{C}^H\mathbf{s}$ from $\hat{\mathbf{y}} = \mathbf{C}^H\mathbf{y}$. Here $\hat{\mathbb{X}}$ contains the first R columns of $\hat{\mathbf{X}}$.

The filter obtained in (18) is referred to as the rank R prior-knowledge GEVD-based MWF (rank R 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 (12), the second block is a rank R GEVD-based MWF and the last block is a selection and scaling part, specific to node k, to produce an estimate $\mathbf{d}_k$ of the desired signal $\bar{\mathbf{d}}_k$. A schematic overview of the rank R PK-GEVD-MWF is given in Figure 2.

To determine the centralized rank R PK-GEVD-MWF, the correlation matrices $\mathbf{R}_{\mathbf{yy}}$ and $\mathbf{R}_{\mathbf{nn}}$ have to be constructed first. This would require each node k to send all of its M_k sensor signals in $\mathbf{y}_k$ to a FC, which will require a large communication bandwidth. Furthermore, as these correlation matrices are large, the inversion of $\mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbf{B}$ in (14) and the GEVD in (19) will require significant computational power at the FC.

To overcome this complexity problem, a distributed adaptive estimation algorithm is presented in this paper. However in the next section the centralized rank R PK-GEVD-MWF is first derived alternatively using a Lagrangian framework, since the expression for the Lagrange parameters will turn out to play an important in the distributed filter.

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

III. Lagrangian perspective on rank R PK-GEVD-MWF

In this section expression (18) for the rank R PK-GEVD-MWF is derived via a different optimization problem. First notice that all the node-specific rank R PK-GEVD-MWFs have a common column space, which can be observed from

$$\mathbf{W}_k = \underbrace{\mathbf{C}\hat{\mathbb{X}}}_{\mathbb{X}} \underbrace{\text{diag}\{\frac{\hat{\sigma}_1 - 1}{\hat{\sigma}_1}, ..., \frac{\hat{\sigma}_R - 1}{\hat{\sigma}_R}\}\hat{\mathbb{Q}}^H \mathbf{H}^H \mathbf{E}_{\mathbf{d}_k}}_{\mathbf{G}_k} \tag{20}$$

where $\mathbf{G}_k$ is a node-specific non-singular $R \times R$ matrix and $\mathbb{X}$ is common for all filters $\mathbf{W}_k$. In the following, the optimization problem is proposed for the node-independent variable $\mathbb{X} \in \mathbb{C}^{M \times R}$ and by later using the substitution

$$\mathbb{X} = \mathbf{W}_k \mathbf{G}_k^{-1} \tag{21}$$

expression (18) and the corresponding Lagrange parameters will be obtained.

A different optimization problem that, as shown in the following derivation, will result in the same expression for $\mathbb{X}$, is given as follows:

$$\mathbb{X} = \underset{\substack{\mathbb{X}^H \mathbf{R}_{\mathbf{nn}} \mathbb{X} = \mathbf{I}_R \\ \mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbb{X} = \mathbf{0}}}{\arg\max} \text{tr}\{\mathbb{X}^H \mathbf{R}_{\mathbf{yy}} \mathbb{X}\}. \tag{22}$$

Without the last equality constraint ($\mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbb{X} = \mathbf{0}$), $\mathbb{X}$ would contain the R principal generalized eigenvectors of the matrix pencil $\{\mathbf{R}_{\mathbf{yy}}, \mathbf{R}_{\mathbf{nn}}\}$ corresponding to the largest generalized eigenvalues. Note that the optimal solution of (22) will never be unique, since $\mathbb{X}\mathbf{U}$ with $\mathbf{U}\mathbf{U}^H = \mathbf{I}_R$ gives the same objective function while satisfying all the constraint. This non-uniqueness problem is averted by constraining $\mathbb{X}^H \mathbf{R}_{\mathbf{yy}} \mathbb{X}$ to be a diagonal matrix with its (real) diagonal elements sorted from large to small.[4]

Since problem (22) is a constrained optimization problem, it can be solved using a Langrangian framework implying that the solution of (22) is given by saddle points (or in this case fixed points) of the Lagrangian function [35]:

$$\begin{aligned} L(\mathbb{X}, \boldsymbol{M}, \mathbb{\Lambda}) = \text{tr}\{\mathbb{X}^H \mathbf{R}_{\mathbf{yy}} \mathbb{X}\} + \text{tr}\{\boldsymbol{M}^H (\mathbb{X}^H \mathbf{R}_{\mathbf{nn}} \mathbb{X} - \mathbf{I}_R) \\ + (\mathbb{X}^H \mathbf{R}_{\mathbf{nn}} \mathbb{X} - \mathbf{I}_R) \boldsymbol{M}\} \\ - \text{tr}\{\mathbb{\Lambda}^H \mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbb{X} + \mathbb{X}^H \mathbf{R}_{\mathbf{nn}} \mathbf{B} \mathbb{\Lambda}\}. \end{aligned} \tag{23}$$

The fixed points are given by the following set of equations

$$\begin{aligned} \nabla_{\mathbb{X}^*} L(.) &= \mathbf{R}_{\mathbf{yy}} \mathbb{X} + \mathbf{R}_{\mathbf{nn}} \mathbb{X} (\boldsymbol{M} + \boldsymbol{M}^H) - \mathbf{R}_{\mathbf{nn}} \mathbf{B} \mathbb{\Lambda} = \mathbf{0} \\ \nabla_{\boldsymbol{M}^*} L(.) &= \mathbb{X}^H \mathbf{R}_{\mathbf{nn}} \mathbb{X} - \mathbf{I}_R = \mathbf{0} \\ \nabla_{\mathbb{\Lambda}^*} L(.) &= \mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbb{X} = \mathbf{0}. \end{aligned} \tag{24}$$

The uniqueness constraint of $\mathbb{X}$ can then be imposed by forcing $(\boldsymbol{M} + \boldsymbol{M}^H)$ to be a (real) diagonal matrix with its (real) diagonal elements sorted from large to small. Substituting $\mathbb{X} = \mathbf{C}\hat{\mathbb{X}}$ in (24), where $\mathbf{C}$ is defined in (63), results in

$$\begin{aligned} &\mathbf{R}_{\mathbf{yy}} \mathbf{C}\hat{\mathbb{X}} + \mathbf{R}_{\mathbf{nn}} \mathbf{C}\hat{\mathbb{X}} (\boldsymbol{M} + \boldsymbol{M}^H) - \mathbf{R}_{\mathbf{nn}} \mathbf{B} \mathbb{\Lambda} = \mathbf{0} \\ &\hat{\mathbb{X}}^H \mathbf{C}^H \mathbf{R}_{\mathbf{nn}} \mathbf{C}\hat{\mathbb{X}} - \mathbf{I}_R = \mathbf{0} \\ &\mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbf{C}\hat{\mathbb{X}} = \mathbf{0} \hat{\mathbb{X}} = \mathbf{0}. \end{aligned} \tag{25}$$

[4]There is still a sign ambiguity for each column of $\hat{\mathbb{X}}$, but this sign ambiguity is compensated by $\mathbf{G}_k$ to obtain a unique PK-GEVD-MWF $\mathbf{W}_k$.

The last equation is then automatically fulfilled for every $\hat{\mathbb{X}}$. Multiplying the first equation first with $\mathbf{C}^H$ and then with $\mathbf{B}^H$, results in (using (16))

$$\begin{aligned} &\mathbf{R}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} \hat{\mathbb{X}} + \mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}} \hat{\mathbb{X}} (\boldsymbol{M} + \boldsymbol{M}^H) = \mathbf{0} \\ &\mathbf{B}^H \mathbf{R}_{\mathbf{yy}} \mathbf{C}\hat{\mathbb{X}} - \mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbf{B} \mathbb{\Lambda} = \mathbf{0} \\ &\hat{\mathbb{X}}^H \mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}} \hat{\mathbb{X}} - \mathbf{I}_R = \mathbf{0}. \end{aligned} \tag{26}$$

Looking at the first and last equation and taking the objective ($\text{tr}\{\mathbb{X}^H \mathbf{R}_{\mathbf{yy}} \mathbb{X}\} = \text{tr}\{\hat{\mathbb{X}}^H \mathbf{R}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} \hat{\mathbb{X}}\}$) into account, one can see that the optimal $\hat{\mathbb{X}}$ contains the R generalized eigenvectors of the matrix pair $\{\mathbf{R}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}, \mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}\}$ corresponding to the largest generalized eigenvalues. The optimal $\mathbb{X}$ can then be found using $\mathbb{X} = \mathbf{C}\hat{\mathbb{X}}$ which is exactly the same expression for $\mathbb{X}$ as in (20). The optimal value of $\mathbb{\Lambda}$ is given by

$$\mathbb{\Lambda} = (\mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbf{B})^{-1} \mathbf{B}^H \mathbf{R}_{\mathbf{yy}} \mathbb{X}. \tag{27}$$

Both $\mathbb{X}$ and $\mathbb{\Lambda}$ will play an important role as compression matrices for the distributed algorithm presented in the next section.

Using the substitution defined in (21), the node-specific Lagrange parameter is given by

$$\begin{aligned} \mathbf{\Lambda}_k &= (\mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbf{B})^{-1} \mathbf{B}^H \mathbf{R}_{\mathbf{yy}} \mathbb{X} \mathbf{G}_k \\ &= (\mathbf{B}^H \mathbf{R}_{\mathbf{nn}} \mathbf{B})^{-1} \mathbf{B}^H \mathbf{R}_{\mathbf{yy}} \mathbf{W}_k. \end{aligned} \tag{28}$$

Before presenting the distributed PK-GEVD-DANSE, some extra definitions are introduced. A partitioning of the filter $\mathbf{W}_k$ is defined as $\mathbf{W}_k = [\mathbf{W}_{k1}^T \ ... \ \mathbf{W}_{kN}^T]^T$ where $\mathbf{W}_{kq}$ denotes the $M_k \times R$ submatrix of $\mathbf{W}_k$ that is applied to $\mathbf{y}_q$ in (4). Similarly a partitioning of $\mathbf{\Lambda}_k$ is defined as $\mathbf{\Lambda}_k = [\mathbf{\Lambda}_{k1}^T \ ... \ \mathbf{\Lambda}_{kN}^T]^T$ where $\mathbf{\Lambda}_{kq}$ denotes the $(M_k - L_k) \times R$ submatrix of $\mathbf{\Lambda}_k$ that is multiplied with $\mathbf{B}_q$ in (23). Note that for the optimal filters $\mathbf{W}_k$ and Lagrange parameters $\mathbf{\Lambda}_k$, the following relations hold:

$$\begin{aligned} \mathbf{W}_{kq} &= \mathbb{X}_{kk} \mathbf{G}_q \\ \mathbf{\Lambda}_{kq} &= \mathbb{\Lambda}_{kk} \mathbf{G}_q \end{aligned} \tag{29}$$

with $\mathbb{X} = [\mathbb{X}_{11}^T \ ... \ \mathbb{X}_{NN}^T]^T$ and $\mathbb{\Lambda} = [\mathbb{\Lambda}_{11}^T \ ... \ \mathbb{\Lambda}_{NN}^T]^T$ defined similarly.

IV. rank R PK-GEVD-DANSE

To overcome the complexity problem that emerges when all M sensor signals in $\mathbf{y}$ are sent to a FC, a distributed adaptive estimation algorithm is presented in this section. In a fully-connected WSN each node then broadcasts only $2R$ compressed sensor signals and the computations in each node are performed on a smaller number of signals[5], 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 obtain the same desired signal estimates as if the node had access to all the sensor signals in the WSN and so could directly compute the centralized rank R PK-GEVD-MWF. The distributed algorithm is referred to as the rank

[5]In an iteration of Algorithm 1, 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.

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

R Prior Knowledge GEVD-based Distributed Adaptive Node-Specific Signal Estimation (rank R PK-GEVD-DANSE) after DANSE and GEVD-DANSE of [10] and [11]. A drawback of the rank R PK-GEVD-DANSE is the slower adaptation and tracking speed compared to the centralized rank R PK-GEVD-MWF, since the rank R PK-GEVD-DANSE needs multiple iterations on the same data to converge to the optimal centralized rank R PK-GEVD-MWF.

A. Algorithm description

In the rank R PK-GEVD-DANSE, in one iteration i, each node k broadcasts, instead of observations of the full M_k sensor signal $\mathbf{y}_k$, observations of $2R$ compressed sensor signals:

- the signal $\mathbf{z}_k^i = \mathbf{W}_{kk}^{i,H}\mathbf{y}_k$ where the $M_k \times R$-dimensional compression matrix $\mathbf{W}_{kk}^i$ is obtained by partitioning $\mathbf{W}_k^i$, which is an estimate of $\mathbf{W}_k$ in iteration i;
- the signal $\underline{\mathbf{z}}_k^i = \mathbf{\Lambda}_{kk}^{i,H}\mathbf{B}_k^H\mathbf{y}_k$ corresponding to compressed versions of the local noise references $\mathbf{B}_k^H\mathbf{y}_k$, where the $(M_k - L_k) \times R$ compression matrix $\mathbf{\Lambda}_{kk}^i$ is obtained by partitioning $\mathbf{\Lambda}_k^i$, which is an estimate of $\mathbf{\Lambda}_k$ in iteration i.

Consequently, each node k has access to observations of the reduced-dimensional sensor signal $\tilde{\mathbf{y}}_k^i = [\mathbf{y}_k^T \ \mathbf{z}_{-k}^{i,T} \ \underline{\mathbf{z}}_{-k}^{i,T}]^T$ where the subscript $_{-k}$ refers to the concatenation of the compressed sensor signals of the other nodes, i.e.

$$\mathbf{z}_{-k}^i = [\mathbf{z}_1^{i,T} \dots \mathbf{z}_{k-1}^{i,T} \ \mathbf{z}_{k+1}^{i,T} \dots \mathbf{z}_K^{i,T}]^T \tag{30}$$

$$\underline{\mathbf{z}}_{-k}^i = [\underline{\mathbf{z}}_1^{i,T} \dots \underline{\mathbf{z}}_{k-1}^{i,T} \ \underline{\mathbf{z}}_{k+1}^{i,T} \dots \underline{\mathbf{z}}_K^{i,T}]^T. \tag{31}$$

To be able to perform the same operations as in the centralized rank R 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 for node k:

$$\tilde{\mathbf{H}}_k = \begin{bmatrix} \mathbf{H}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{R(K-1)} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}; \quad \tilde{\mathbf{B}}_k = \begin{bmatrix} \mathbf{B}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{R(K-1)} \end{bmatrix}. \tag{32}$$

The steering matrix 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 matrix subspace for $\mathbf{z}_{-k}^i$ is unknown and thus represented by $\mathbf{I}_{R(K-1)}$ in $\tilde{\mathbf{H}}_k$ and by $\begin{bmatrix} \ \end{bmatrix}$ in $\tilde{\mathbf{B}}_k$. The steering matrix subspace for $\underline{\mathbf{z}}_{-k}^i$ is empty because this signal is a compressed version of the 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 $\begin{bmatrix} \ \end{bmatrix}$ in $\tilde{\mathbf{H}}_k$ and hence by $\mathbf{I}_{R(K-1)}$ in $\tilde{\mathbf{B}}_k$.

The rank R PK-GEVD-DANSE is presented in Algorithm 1. This is a block-iterative round-robin algorithm[6], where the updating node performs the same operations as in the centralized rank R PK-GEVD-MWF, but here with locally defined reduced-dimensional variables $\tilde{\mathbf{F}}_k, \tilde{\mathbf{C}}_k, \hat{\mathbf{W}}_k$ and $\tilde{\mathbf{W}}_k$. A schematic overview of the local rank R PK-GEVD-MWF is given in Figure 3.

[6]Instead of round-robin updates, simultaneous and asynchronous updating as discussed in [36], requiring relaxed updates in (38) and (39), can also be performed, but this is not discussed here for ease of exposition.

Algorithm 1: rank R PK-GEVD-DANSE

1 - All nodes $k \in \mathcal{K}$ construct $\tilde{\mathbf{H}}_k$ and $\tilde{\mathbf{B}}_k$ using node k's prior knowledge, initialize $\mathbf{W}_{kk}^0$ and $\mathbf{\Lambda}_{kk}^0$ as random matrices
- $i \leftarrow 0$ and $q \leftarrow 1$.

2 - All nodes $k \in \mathcal{K}$ broadcast B observations of $\mathbf{z}_k = \mathbf{W}_{kk}^{i,H}\mathbf{y}_k$ and $\underline{\mathbf{z}}_k = \mathbf{\Lambda}_{kk}^{i,H}\mathbf{B}_k^H\mathbf{y}_k$ and construct locally

$$\tilde{\mathbf{y}}_k^i = [\mathbf{y}_k^T \ \mathbf{z}_{-k}^{i,T} \ \underline{\mathbf{z}}_{-k}^{i,T}]^T. \tag{33}$$

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

$$\tilde{\mathbf{F}}_q^{i+1} = \left(\tilde{\mathbf{B}}_q^H\mathbf{R}_{\tilde{\mathbf{n}}_q\tilde{\mathbf{n}}_q}\tilde{\mathbf{B}}_q\right)^{-1}\tilde{\mathbf{B}}_q^H\mathbf{R}_{\tilde{\mathbf{n}}_q\tilde{\mathbf{n}}_q}\tilde{\mathbf{H}}_q \tag{34}$$

$$\tilde{\mathbf{C}}_q^{i+1} = \tilde{\mathbf{H}}_q - \tilde{\mathbf{B}}_q\tilde{\mathbf{F}}_q^{i+1}, \quad \hat{\mathbf{y}}_q = \tilde{\mathbf{C}}_q^{i+1^H}\tilde{\mathbf{y}}_q. \tag{35}$$

- Updating 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 B observations and constructs $\tilde{\mathbf{W}}_q^{i+1}$ as in (19) 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:

$$\tilde{\mathbf{W}}_q^{i+1} = \tilde{\mathbf{C}}_q^{i+1}\hat{\mathbf{W}}_q^{i+1}\tilde{\mathbf{H}}_q^H[\mathbf{I}_R \ \ \mathbf{0}]^H \tag{36}$$

$$\tilde{\mathbf{\Lambda}}_q^{i+1} = \left(\tilde{\mathbf{B}}_q^H\mathbf{R}_{\tilde{\mathbf{n}}_q\tilde{\mathbf{n}}_q}\tilde{\mathbf{B}}_q\right)^{-1}\tilde{\mathbf{B}}_q^H\mathbf{R}_{\tilde{\mathbf{y}}_q\tilde{\mathbf{y}}_q}\tilde{\mathbf{W}}_q^{i+1} \tag{37}$$

$$\mathbf{W}_{qq}^{i+1} = \begin{bmatrix}\mathbf{I}_{M_q} & \mathbf{0}\end{bmatrix}\tilde{\mathbf{W}}_q^{i+1} \tag{38}$$

$$\mathbf{\Lambda}_{qq}^{i+1} = \begin{bmatrix}\mathbf{I}_{M_q-L_q} & \mathbf{0}\end{bmatrix}\tilde{\mathbf{\Lambda}}_q^{i+1}. \tag{39}$$

- All other nodes do not change their variables:

$$\tilde{\mathbf{W}}_k^{i+1} = \tilde{\mathbf{W}}_k^i, \mathbf{\Lambda}_{kk}^{i+1} = \mathbf{\Lambda}_{kk}^i, \mathbf{W}_{kk}^{i+1} = \mathbf{W}_{kk}^i. \tag{40}$$

4 - All nodes $k \in \mathcal{K}$ generate B observations of the estimated desired signal $\mathbf{d}_k^{i+1} \approx \tilde{\mathbf{W}}_k^{i+1,H}\tilde{\mathbf{y}}_k^i$.

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

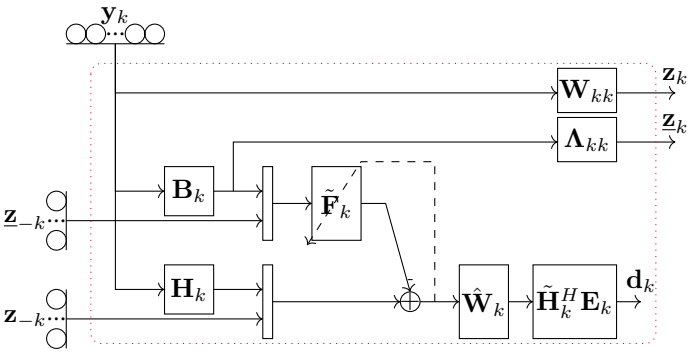

Fig. 3: Schematic overview of the per-node operations in the rank R PK-GEVD-DANSE. The signals outside the dotted box are the signals that the node receives and sends to the other nodes.

B. Sufficient statistic

A formal convergence proof for PK-GEVD-DANSE is not available, but convergence is conjectured as follows:

Conjecture IV.1. *Assume that $\mathbf{R}_{\mathbf{yy}}$ and $\mathbf{R}_{\mathbf{nn}}$ are positive definite matrices. The rank R PK-GEVD-DANSE converges for any initialization of its compression matrices $\mathbf{W}_{kk}^0$ and $\mathbf{\Lambda}_{kk}^0$ to the network-wide PK-GEVD-MWF* (18) *for any k, i.e., when $i \to \infty$ then $\mathbf{W}_{kk}^i = \mathbf{W}_{kk}, \mathbf{\Lambda}_{kk}^i = \mathbf{\Lambda}_{kk}$ and $\mathbf{d}_k^i = \mathbf{W}_k^H\mathbf{y}$ $\forall k$.*

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

In Section VI extensive simulations are provided in which convergence of the rank R PK-GEVD-DANSE is consistently observed, supporting this conjecture. In addition, in this section a formal explanation for the seemingly arbitrary choice of the expressions for the compression matrices $\mathbf{W}_{kk}^i$ and $\mathbf{\Lambda}_{kk}^i$ in Algorithm 1 is provided. It is shown that when the compression matrices reach a certain state where $\mathbf{W}_{kk}^i = \mathbf{W}_{kk}$ and $\Lambda_{kk}^i = \breve{\mathbf{\Lambda}}_{kk}$ for all k, this state is an equilibrium state of Algorithm 1 and the desired signal estimates are the same as the the desired signal estimates of the rank R PK-GEVD-MWF for all nodes. This property of Algorithm 1 is referred to here as the sufficient statistic property of the compressions $\mathbf{W}_{kk}$ and $\mathbf{\Lambda}_{kk}$ for all nodes. To next main theorem provides a sufficient statistic property of the compression $\mathbf{W}_{kk}, \mathbf{\Lambda}_{kk}$, for each node k. It is noted that, similar to [10], [11], the proof for the sufficient statistic property assumes that the local correlation matrices are perfectly estimated using infinitely many observations, to make the theoretical analysis mathematically tractable. In practice, a finite observation window is used to estimate local correlation matrices, and therefore this analysis should be viewed as an asymptotic analysis for which the approximation accuracy improves when a larger observation window is used.

Theorem IV.2. *If the rank R PK-GEVD-DANSE has reached a state in iteration $i^\star$ where $\mathbf{W}_{kk}^{i^\star}$ and $\mathbf{\Lambda}_{kk}^{i^\star}$ have the same column space as $\mathbb{X}_{kk}$ and $\mathbb{\Lambda}_{kk}$ respectively, i.e. there exist transformation matrices $\mathbf{G}_k^{i^\star,1}$ and $\mathbf{G}_k^{i^\star,2}$ such that*

$$\mathbf{W}_{kk}^{i^\star} = \mathbb{X}_{kk}\mathbf{G}_k^{i^\star,1}, \;\; \mathbf{\Lambda}_{kk}^{i^\star} = \mathbb{\Lambda}_{kk}\mathbf{G}_k^{i^\star,2} \tag{41}$$

for all k, then this state corresponds to an equilibrium point, i.e. $\tilde{\mathbf{W}}_k^{i+1} = \tilde{\mathbf{W}}_k^i$ and $\tilde{\mathbf{\Lambda}}_k^{i+1} = \tilde{\mathbf{\Lambda}}_k^i, \forall k$ where $i \geq i^\star$, and this equilibrium point is given by

$$\tilde{\mathbf{W}}_k^{i^\star} = \begin{bmatrix} \mathbb{X}_{kk}\mathbf{G}_k \\ \mathbf{G}_1^{-1}\mathbf{G}_k \\ \vdots \\ \mathbf{G}_K^{-1}\mathbf{G}_k \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \tilde{\mathbf{\Lambda}}_k^{i^\star} = \begin{bmatrix} \mathbb{\Lambda}_{kk}\mathbf{G}_k \\ \mathbf{G}_1^{-1}\mathbf{G}_k \\ \vdots \\ \mathbf{G}_K^{-1}\mathbf{G}_k \end{bmatrix} \tag{42}$$

such that $\mathbf{G}_k^{i^\star,1}$ and $\mathbf{G}_k^{i^\star,2}$ are equal to $\mathbf{G}_k$ and the final filter for every node k is given by the node-specific rank R PK-GEVD-MWF $\mathbf{W}_k = \mathbb{X}\mathbf{G}_k$.

Proof. See Appendix D. □

A similar sufficient statistic property is also valid for the DANSE and GEVD-DANSE of [10], [11], showing that the optimal node-specific filters are stable solutions under the update rules of the considered round-robin updates. To prove convergence of DANSE and GEVD-DANSE, the updates of the algorithm (using transformation matrices) are linked to a centralized objective function that follows a monotonically nondecreasing path over the different updates. For GEVD-DANSE this is optimization problem (22) without the constraint $\mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbb{X} = \mathbf{0}$. Since these centralized objective functions are bounded below/above, this shows that the updates result in filters converging to a fixed point. The sufficient statistic property is then used to show that the optimal centralized filter is the only stable solution.

TABLE II: Comparison of communication cost and computational complexity at updating node q.

	Transmitted signals	Received signals	Computational Complexity
Centralized PK-GEVD-MWF	$M_q B$	$(M-M_q)B$	$O(L^3 + (M-L)^3)$
PK-GEVD-DANSE	$2RB$	$2R(K-1)B$	$O\big((L_q + R(K-1))^3 + (M_q - L_q + R(K-1))^3\big)$

Conjecture IV.1 is based on the convergence theorems of DANSE and GEVD-DANSE. However, linking the updates of PK-GEVD-DANSE to a centralized objective function, for example based on (22), that experience a nonincreasing/nondecreasing path over the different updates, is seen to be particularly hard. The constraint $\mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbb{X} = \mathbf{0}$ is only satisfied upon convergence of PK-GEVD-DANSE, allowing the objective function $\text{tr}\{\mathbb{X}^H\mathbf{R}_{\mathbf{yy}}\mathbb{X}\}$ to reach values higher than the optimal one during intermediate updates. Therefor, Conjecture IV.1 can only be given as a statement, a complete proof is still missing.

C. Communication cost and computational complexity

In each iteration of the rank R PK-GEVD-DANSE, each node transmits $2RB$ and receives $2(K-1)RB$ complex scalars, where B is the number of observations in one iteration. Assuming the same communication cost for the transmitter and the receiver modules, the total communication cost for node k in each iteration will be $2KRB$, leading to a network communication cost of $2K^2RB$. However, in a setup where each node functions as a FC, the communication cost of each individual node and the communication cost of the whole network would be KM_kB and KMB, respectively. If $2R < M_k$, then obviously the rank R PK-GEVD-DANSE reduces the communication cost for this node with a factor $M_k/2R$. This dimensionality reduction then also results in a reduced per-node computational complexity. The network-wide PK-GEVD-MWF has a complexity of $O(L^3)$ for the GEVD-computation and $O((M-L)^3)$ for the matrix inversion in (15), whereas the rank R PK-GEVD-DANSE has a complexity of $O((L_q + R(K-1))^3)$ and $O((M_q - L_q + R(K-1))^3)$ respectively at the updating node q. When compared to the the setup where each node functions as a FC, this reduction comes at the cost of having a slower adaptation speed or tracking performance due to the iterative nature of the rank R PK-GEVD-DANSE. A overview of the discussed communication costs and computational complexities is given in Table II.

D. Topology independent rank R PK-GEVD-DANSE

Algorithm 1 operates in a fully-connected network, i.e. each node communicates with all of the other nodes in the network via a wireless link. As the number of wireless links scales with $O(K^2)$, which becomes intractable for large K, other topologies are investigated in which the number of links is smaller. In [37], [38] simple changes are made to let the

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

original DANSE [10] operate in different topologies, such as tree topologies.

To show that similar changes will also work for the proposed rank R PK-GEVD-DANSE, a topology-independent variant is considered. For this topology independent rank R PK-GEVD-DANSE, it is assumed that a node q has access to the sum of the $\mathbf{z}_k^i$ and $\underline{\mathbf{z}}_k^i$ of all other nodes, denoted by $\mathbf{z}_{\circ q}^i$ and $\underline{\mathbf{z}}_{\circ q}^i$ respectively given as

$$\mathbf{z}_{\circ q}^i = \sum_{k\neq q} \mathbf{z}_k^i = \sum_{k\neq q} \mathbb{X}_{kk}^{i,H} \mathbf{y}_k \tag{43}$$

$$\underline{\mathbf{z}}_{\circ q}^i = \sum_{k\neq q} \underline{\mathbf{z}}_k^i = \sum_{k\neq q} \lambda_{kk}^{i,H} \mathbf{B}_k^H \mathbf{y}_k. \tag{44}$$

Efficient methods for constructing these network-wide sums in (43) and (44) are presented in [38]. This is compared to the case where updating node q has access to the $(K-1)R$-dimensional signals $\mathbf{z}_{-q}^i$ and $\underline{\mathbf{z}}_{-q}^i$. The local operations of this topology independent rank R PK-GEVD-DANSE are similar to those of Algorithm 1, except for the fact that the subspace matrix and corresponding blocking matrix have to be redefined as

$$\tilde{\mathbf{H}}_k = \begin{bmatrix} \mathbf{H}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_R \\ \mathbf{0} & \mathbf{0} \end{bmatrix}; \quad \tilde{\mathbf{B}}_k = \begin{bmatrix} \mathbf{B}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_R \end{bmatrix} \tag{45}$$

and that the compression matrices are redefined as $\mathbb{X}_{qq}^i$ and $\mathbb{\Lambda}_{qq}^i$ compared to $\mathbf{W}_{qq}^i$ and $\mathbf{\Lambda}_{qq}^i$ in Algorithm 1. Since each node has to compute a (reduced dimensional) GEVD as presented in (19) from which it obtains $\hat{\mathbb{X}}_q^i$ as a byproduct, $\mathbb{X}_{qq}^i$ and $\mathbb{\Lambda}_{qq}^i$ can be computed as

$$\mathbb{X}_{qq}^i = \begin{bmatrix} \mathbf{I}_{M_q} & \mathbf{0} \end{bmatrix} \tilde{\mathbf{C}}_q^i \hat{\mathbb{X}}_q^i \tag{46}$$

$$\mathbb{\Lambda}_{qq}^i = \begin{bmatrix} \mathbf{I}_{M_q - L_q} & \mathbf{0} \end{bmatrix} \left(\tilde{\mathbf{B}}_q^H \mathbf{R}_{\tilde{\mathbf{n}}_q \tilde{\mathbf{n}}_q} \tilde{\mathbf{B}}_q \right)^{-1} \tilde{\mathbf{B}}_q^H \mathbf{R}_{\tilde{\mathbf{y}}_q \tilde{\mathbf{y}}_q} \tilde{\mathbf{C}}_q^i \hat{\mathbb{X}}_q^i \tag{47}$$

It is important that these compression matrices are redefined, such that the different compressed signals add coherently. A similar sufficient statistic property as in Section IV-B can than be derived since it is clear that at convergence

$$\mathbb{X}^H \mathbf{y} = \mathbb{X}_{qq}^H \mathbf{y}_q + \sum_{k\neq q} \mathbb{X}_{kk}^H \mathbf{y}_k = [\mathbb{X}_{qq}^H \mathbf{I}_R] \begin{bmatrix} \mathbf{y}_q \\ \mathbf{z}_{\circ q} \end{bmatrix} \tag{48}$$

and

$$\mathbb{\Lambda}^H \mathbf{B}^H \mathbf{y} = \mathbb{\Lambda}_{qq}^H \mathbf{B}_q^H \mathbf{y}_q + \sum_{k\neq q} \mathbb{\Lambda}_{kk}^H \mathbf{B}_k^H \mathbf{y}_k = [\mathbb{\Lambda}_{qq}^H \mathbf{B}_q^H \ \mathbf{I}_R] \begin{bmatrix} \mathbf{y}_q \\ \underline{\mathbf{z}}_{\circ q}^i \end{bmatrix}. \tag{49}$$

The local desired signal estimate can then be computed as $\mathbf{W}_q^H \mathbf{y} = \mathbf{G}_q^H \mathbb{X}^H \mathbf{y}$.

V. Rank R PK-GEVD-DANSE with rank deficiencies

Algorithm 1 assumes that $L_k > R$ and $M_k - L_k > R \ \forall k$ in order to obtain the communication cost and computational complexity reductions as discussed in Subsection IV-C. However if this does not hold for certain nodes, Algorithm 1 needs to be adapted as follows:

1) $L_k < R$, $M_k - L_k < R$: In this case, it is more efficient that node k defines its compressed sensor signals as $\mathbf{z}_k = \mathbf{H}_k^H \mathbf{y}_k$ and $\underline{\mathbf{z}}_k = \mathbf{B}_k^H \mathbf{y}_k$, such that the total number of broadcast signals is equal to M_k instead of $2R$ when the (rank-deficient) compressions matrices defined in Algorithm 1 are used. There is no communication cost reduction compared to the setup where each node functions as a FC in this case.

2) $L_k > R$, $M_k - L_k < R$: In this case, it is more efficient that node k redefines $\underline{\mathbf{z}}_k$ in as $\underline{\mathbf{z}}_k = \mathbf{B}_k^H \mathbf{y}_k$. The definition for $\mathbf{z}_k^i$ remains the same as in Algorithm 1. A specific case is when $M_k = L_k$. This corresponds to the case where $B_k = [\,]$. This means that node k has no partial prior knowledge on the local steering matrix. Since this node does not have any noise-references, this leads to $\underline{\mathbf{z}}_k = [\,]$, thus this node only transmits $\mathbf{z}_k$ and has a further communication cost reduction (R broadcast signals i.s.o $2R$). When this is the case for all nodes in the WSN, Algorithm 1 reduces to the GEVD-DANSE of [11].

3) $L_k < R$, $M_k - L_k > R$: This case can only happen when $\text{rank}(A_k) < R$, i.e. when node k is not observing all the R desired sources. The rank of $\tilde{\mathbf{W}}_k^{i+1}$ in (36) is then equal to $L_k < R$, and the resulting compression matrices $\mathbf{W}_{kk}^{i+1}$ and $\mathbf{\Lambda}_{kk}^{i+1}$ have also the same rank. However for the sufficient statistic property, it is required that the column space of the compression matrices should contain the R-dimensional column space of $\mathbb{X}_{kk}$ and $\mathbb{\Lambda}_{kk}$. This can be fixed by redefining the compression matrices as in the topology-independent PK-GEVD-DANSE, i.e. by using (46) and (47).

VI. Numerical Simulations

A. Distributed node-specific signal estimation in a WSN

To demonstrate the convergence and optimality of the rank R PK-GEVD-DANSE, the following scenario is used. The scenario consists of 4 nodes with each $M_k = 15$ sensor signals, 3 desired source $\check{\mathbf{s}}[t]$ and 10 undesired noise sources $\check{\mathbf{n}}[t]$. Node 1 and node 2 have access to an exact (normalized) estimate of their steering matrix $\mathbf{A}_1$ and $\mathbf{A}_2$, which are then equal to $\mathbf{H}_1$ and $\mathbf{H}_2$ respectively. The first 5 sensor signals of node 3 correspond directly to 5 of the 10 undesired noise sources. Finally, node 4 has the prior knowledge that its first 5 sensors do not observe the desired sources. The prior knowledge of node 3 and 4 can thus be captured by the following matrices:

$$\mathbf{H_3} = \mathbf{H_4} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{10} \end{bmatrix}. \tag{50}$$

The rank 3 PK-GEVD-DANSE is run in batch-mode[7]. Monte-Carlo (MC) simulations are conducted and compared with the convergence of rank 3 GEVD-DANSE and rank 6 GEVD-DANSE (same number of broadcast signals as rank 3 PK-GEVD-DANSE). Also the results of the centralized GEVD-MWF and the centralized PK-GEVD-MWF are shown. In every MC run and $\forall k \in \{1,2,3,4\}$, a new random desired

[7]In batch-mode, iterations are performed over the same block of observations. Note that in reality the algorithm is 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 multiple times.

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

sources steering matrix $\mathbf{A}_k$ and new random undesired noise steering matrix $\mathbf{D}_k$ are generated from a 0-mean complex Gaussian distribution with variance 1, with the following constraints to satisfy the scenario. The first 5 rows of $\mathbf{A}_3$ and $\mathbf{A}_4$ are always equal to zero and the first 3 rows of $\mathbf{D}_3$ are $\begin{bmatrix}\mathbf{I}_5 & \mathbf{0}\end{bmatrix}$. Also $B = 1000$ samples of $\check{\mathbf{s}}[t]$ (being active for 50% of the time), $\check{\mathbf{n}}[t]$ and a random noise component $\mathbf{n}_k[t]$ (to model sensor noise) are generated to create the sensor signals

$$\mathbf{y}_k[t] = \mathbf{A}_k\check{\mathbf{s}}[t] + \mathbf{D}_k\check{\mathbf{n}}[t] + \mathbf{n}_k[t] \quad \forall k \in \{1,2,3,4\} \tag{51}$$

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 sources, is equal to 0 dB. To algorithms are run with a perfect activity detection (AD-error = 0%), and when 5% or 10% of the samples are wrongly classified, introducing estimation error in $\mathbf{R}_{\mathbf{yy}}$ and $\mathbf{R}_{\mathbf{nn}}$.

The upper part of Fig. 4 shows the average (over 200 MC runs) of the decrease in the $L2$-objective function:

$$\sum_{t,k} \frac{1}{BK} \|\bar{\mathbf{d}}_k[t] - \mathbf{d}_k^i[t]\|_2^2 \tag{52}$$

as a function of the number of iterations of the rank 3 PK-GEVD-DANSE and shows the result when the centralized rank 3 GEVD-MWF and PK-GEVD-MWF are used to estimate $\bar{\mathbf{d}}_k[t]$. The rank 3 GEVD-DANSE is able to reduce the objective function compared to the rank 6 GEVD-DANSE showing that increasing the rank in the GEVD is only beneficial when it matches the considered setup, but the addition of the prior knowledge as done in rank 3 PK-GEVD-DANSE, reduces the objective even further. It is also clear that prior knowledge makes the algorithm more robust to AD-errors compared to GEVD-DANSE with the same estimation errors. This happens because PK-GEVD-DANSE is only influenced by mistakes in the lower dimensional subspace (determined by the prior knowledge), while the performance of GEVD-DANSE is influenced by all mistakes in the measured correlation matrices. The bottom part of Fig. 4 shows the average (over 200 MC runs) of the squared error between the centralized filter $\mathbf{W}_k$ and the local filter $\tilde{\mathbf{W}}_k^i$ (converted to a centralized filter via the compression matrices $\mathbf{W}_{kk}^i$ and $\mathbf{\Lambda}_{kk}^i$) averaged over all the nodes. This same is done for the rank 3 GEVD-DANSE. The PK-GEVD-DANSE is also run in a rank 1 and rank 2 version on the same data (with the same partial prior knowledge). Convergence to machine precision is observed for each algorithm, independent of the rank. This supports the statement that Algorithm 1 converges for every positive definite matrix pencil $\{\mathbf{R}_{\mathbf{yy}}, \mathbf{R}_{\mathbf{nn}}\}$. The convergence speed of rank 3 PK-GEVD-DANSE is higher then the convergence of the rank 3 GEVD-DANSE, due to the fact that nodes in rank 3 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.

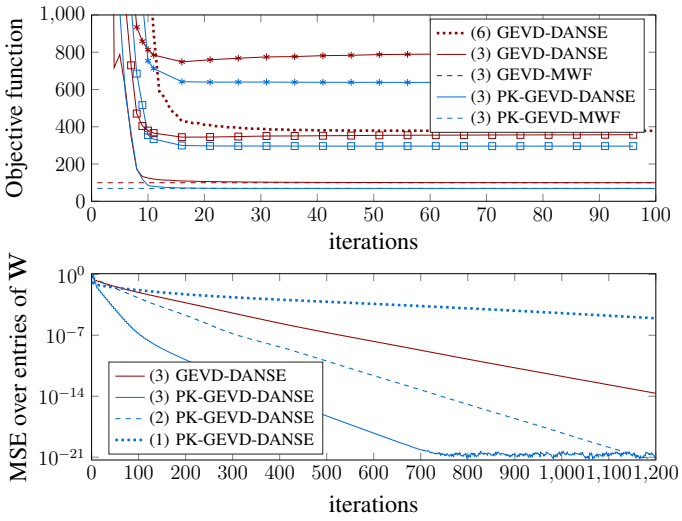

Fig. 4: Convergence properties of PK-GEVD-DANSE. (Upper figure) Comparison between rank 3 PK-GEVD-DANSE, rank 3 GEVD-DANSE, rank 6 GEVD-DANSE and the centralized rank 3 GEVD-MWF and PK-GEVD-MWF by considering the objective function (52). Lines without marks have no AD-errors, lines with square marks (□) have 5% AD-errors and lines with asterisk marks (∗) have 10% AD-errors. (Lower figure) Comparison between rank 3 PK-GEVD-DANSE, rank 2 PK-GEVD-DANSE, rank 1 PK-GEVD-DANSE and rank 3 GEVD-DANSE in terms of convergence speed to the corresponding centralized filter. The number between brackets denotes the rank used for the considered algorithm or considered rank approximation problem.

B. Distributed speech enhancement in a WASN

To evaluate the performance of the proposed distributed PK-GEVD-DANSE and the effect of adding prior knowledge in a realistic environment, an acoustic scenario is simulated using the image method [39]. Note that the signal model (1) is approximately valid for convolutive mixtures that are evaluated in the short-time Fourier transform domain. A WASN with 3 nodes is located in a square room with reflective walls. Figure 5 shows a schematic illustration of the scenario and the positioning of the nodes. Table III describes the parameters used to generate the acoustic scenario. There is one desired speaker in the center of the room surrounded by microphone arrays with omnidirectional microphones. Two noise sources (one white noise source and one babble noise source) are also present. Four loudspeakers are playing four different undesired signals (2 speech signals and 2 white noise signals) and are each located in a different corner of the room. All speech signals are sentences from the HINT database [40]. In addition to the aforementioned noise sources, the microphone signals have an uncorrelated noise component consisting of white noise with power that is 10%-of the power of the desired signal in the first microphone of node 1. The power of the other noise sources is equal and normalized to obtain an input broadband SNR of approximately -10 dB in the reference microphone of each node. All nodes and all sound sources are in the same horizontal plane, 1m above ground level. A sound absorbing wall prevents that 2 microphones of node 3 observe the desired speaker, but the microphones are still observing all the other sources in the scenario. A fourth node

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

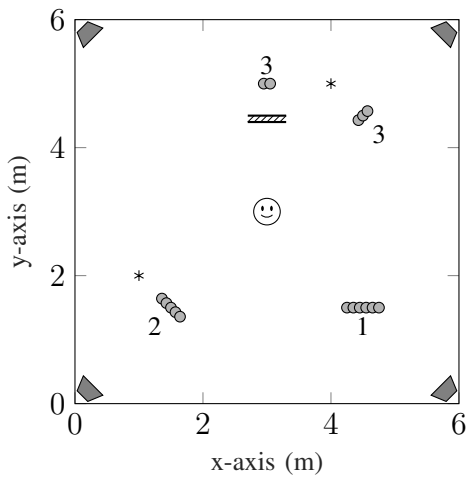

Fig. 5: The acoustic scenario considered in the simulations. The desired speech source is located in the center of the room, surrounded by microphone arrays belonging to specific nodes. The two other noise sources are depicted by $(*)$ and the four loudspeakers are in the corners of the room. A sound absorbing wall prevents that 2 microphones of node 3 observe the desired speaker, while still observing all the other sources in the scenario.

TABLE III: Parameters used to generate the acoustic scenario.

Room dimension	6m × 6m × 3m
T_{60}	0.24s
Microphone spacing	10 cm
Sampling frequency	16kHz
DFT size Hann-window (50% overlap)	256

is considered that consists of all the signals played by the four loudspeaker.

The prior knowledge for each considered frequency bin in the short-time Fourier transform domain, is determined as follows. Node 1 is assumed to have an estimate of the impulse response between the desired speaker and each of its microphones. As a consequence the prior knowledge matrix $\mathbf{H}_1$ is computed by computing the Fourier transform of this impulse responses and combining the corresponding frequency bins to obtain the steering vector $\mathbf{h}_1[n]$ (where n is the bin index). $\mathbf{H}_1[n]$ is then given by normalizing $\mathbf{h}_1[n]$ for each bin. Node 2 know that the desired speaker is in its broadside direction and approximate its prior knowledge matrix $\mathbf{H}_2[n]$ (using the planar wave assumption [16]) as $[\frac{1}{\sqrt{M_2}}...\frac{1}{\sqrt{M_2}}]^T$ for each frequency bin n. No prior knowledge is assumed for the microphones of node 3, but since the 2 (last) microphones are not observing the desired source, the subspace matrix $\mathbf{H}_3$ and blocking matrix $\mathbf{B}_3$ can be represented by

$$\mathbf{H}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ \mathbf{B}_3 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{53}$$

for each frequency bin n. Since node 4 is not observing the desired signal, the prior knowledge in each frequency bin is represented by $\mathbf{H}_4 = [\,]$ and $\mathbf{B}_4 = \mathbf{I}_4$.

Three performance measures are used to assess the quality of the noise reduction and the convergence of the algorithm, namely the broadband signal-to-noise ratio (SNR), the signal-to-distortion ratio (SDR) and the signal to error ratio (SER). These performance measures are defined at node k as

$$\text{SNR}_k[i] = \frac{E\{|s_k^i|^2\}}{E\{|n_k^i|^2\}} \tag{54}$$

$$\text{SDR}_k[i] = \frac{E\{|s_k^i|^2\}}{E\{|s_k^i - s_{k1}|^2\}} \tag{55}$$

$$\text{SER}_k[i] = \frac{E\{|s_{k1}|^2\}}{E\{|s_{k1} - d_k^i|^2\}} \tag{56}$$

with d_k^i the output at node k in iteration i and s_k^i, n_k^i the corresponding time domain desired speech component and noise component respectively. Finally s_{k1} is the desired time domain speech in the first microphone signal of node k. Since the performance of the noise reduction in all the microphone signals of the WASN is similar, the performance measures are only provided for the first microphone signal of node 1. The desired signal component of this reference microphone serves also as reference for the perfect VAD.

1) Batch-Mode Simulations: At first, the algorithm is run in batch-mode on the first 5s of the signals. In each iteration the local correlation matrices are re-estimated over the full 5s using a simple time-averaging procedure. The performance measures for both the rank 1 GEVD-DANSE iterations and the rank 1 PK-GEVD-DANSE iterations are provided in Figure 6, where the performance of centralized (PK-)GEVD-MWF are also provided to observe the convergence. It is clear that the performance of both algorithms converges to the performance of the centralized (PK-)GEVD-MWF and that the performance is almost the same after 10 iterations. The 3 performance measures for the PK-GEVD-MWF are higher than the performance measures for the GEVD-MWF (without prior knowledge), but the strongest benefit is obtained when looking at the speech distortion and the estimation error.

2) Block-Mode Simulations: The algorithm is also run as presentend in Algorithm 1, where blocks of observations equivalent to 5s are used to compute the signal correlation matrices and update the compression matrices. The performance measures for the both the rank 1 GEVD-DANSE iterations and the rank 1 PK-GEVD-DANSE iterations are provided in Figure 7, where the performance of the centralized (PK-)GEVD-MWF are also provided as a reference. It is clear that the the perfomance of PK-GEVD-DANSE is mostly superior to the GEVD-DANSE performance except in terms of speech distortion where the performance is similar.

VII. Conclusion

In this paper, the centralized rank R PK-GEVD-MWF has been derived as an extension to the centralized rank R GEVD-based MWF by introducing partial prior knowledge on the desired sources steering matrix. The centralized rank R PK-GEVD-MWF is also considered in a Lagrangian framework

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

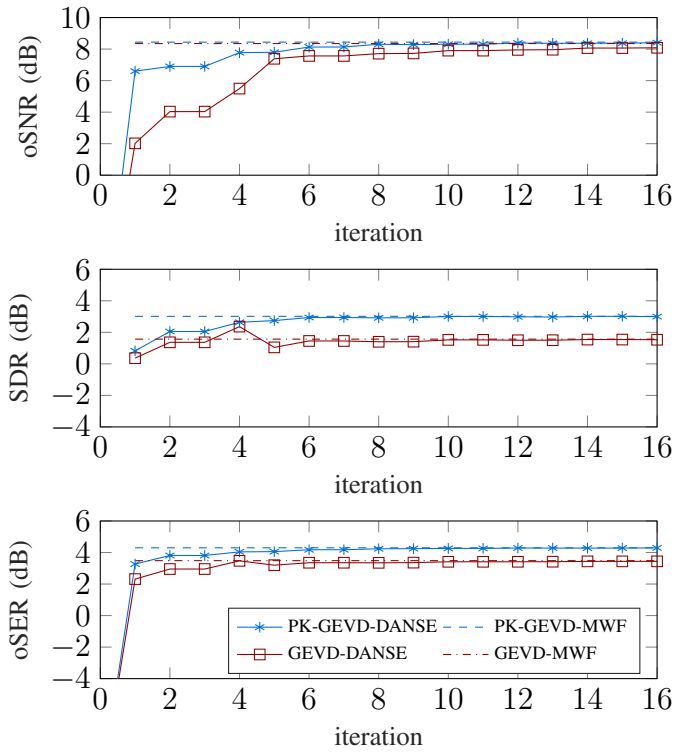

Fig. 6: Performance measures of the GEVD-DANSE and PK-GEVD-DANSE run in batch-mode compared to the centralized GEVD-MWF en PK-GEVD-MWF.

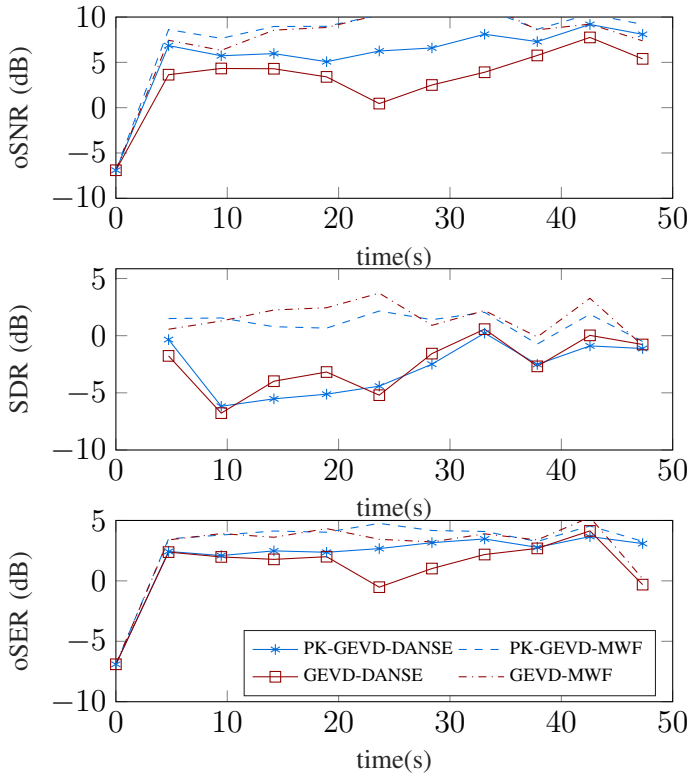

Fig. 7: Performance measures of the GEVD-DANSE and PK-GEVD-DANSE run in block-mode compared to the centralized GEVD-MWF and PK-GEVD-MWF.

to facilitate the step towards the distributed algorithm. The so-called PK-GEVD-DANSE, has then been presented to show that the output of this filter can be computed in a fully-connected WSN in a distributed way. Instead of broadcasting all its sensor signals, each node broadcasts only a compressed versions of its sensor signals, reducing the communication cost and computational complexity, compared to a setup where each node functions as a fusion center. The algorithm has been validated by means of numerical simulations, showing the benefits of including prior knowledge, even in situations with estimation error in the correlation matrices. The algorithm is also applied to a speech enhancement scenario in a WASN, where it is observed that the addition of partial prior knowledge indeed has beneficial effect on the noise reduction performance.

Future research directions include: finding efficient centralized and distributed algorithms for scenarios where the prior knowledge is not accurate but can be introduced as a soft constraint [15] (e.g. $\| \mathbf{B}^H \boldsymbol{R}_{\mathbf{ss}} \mathbf{B} \|_F^2 \leq \epsilon$ with ϵ small) and finding ways to make the distributed algorithm robust to communication channel effects (noise and quantization errors) [4], [41].

APPENDIX A
USEFUL IDENTITIES

For the derivations in Appendix B and C, the following theorem is needed:

Theorem A.1. *Given an $M \times L$-dimensional matrix $\mathbf{A}$ with $M \geq L$ and linearly independent columns. Let $\mathbf{A}_\perp$ denote the $M \times M - L$-dimensional matrix whose columns span the nullspace of $\mathbf{A}^H$ such that $\mathbf{A}^H \mathbf{A}_\perp = \mathbf{0}$ and $[\mathbf{A}\mathbf{A}_\perp]$ is full rank. The identity matrix $\mathbf{I}_M$ can then uniquely be decomposed as follows:*

$$\mathbf{I}_M = \mathbf{A} \left(\mathbf{A}^H \mathbf{A}\right)^{-1} \mathbf{A}^H + \mathbf{A}_\perp \left(\mathbf{A}_\perp^H \mathbf{A}_\perp\right)^{-1} \mathbf{A}_\perp^H \tag{57}$$

Proof. This can be seen from

$$\begin{aligned} \mathbf{I}_M &= [\mathbf{A}\ \mathbf{A}_\perp][\mathbf{A}\ \mathbf{A}_\perp]^{-1}[\mathbf{A}\ \mathbf{A}_\perp]^{-H}[\mathbf{A}\ \mathbf{A}_\perp]^H \\ &= [\mathbf{A}\ \mathbf{A}_\perp] \left([\mathbf{A}\ \mathbf{A}_\perp]^H [\mathbf{A}\ \mathbf{A}_\perp]\right)^{-1} [\mathbf{A}\ \mathbf{A}_\perp]^H \\ &= [\mathbf{A}\ \mathbf{A}_\perp] \begin{bmatrix} \mathbf{A}^H\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_\perp^H\mathbf{A}_\perp \end{bmatrix}^{-1} [\mathbf{A}\ \mathbf{A}_\perp]^H \end{aligned} \tag{58}$$

□

Applying Theorem A.1 using the matrices $\mathbf{A} = \mathbf{R}_{\mathbf{nn}}^{-1/2}\mathbf{H}$ and $\mathbf{A}_\perp = \mathbf{R}_{\mathbf{nn}}^{H/2}\mathbf{B}$ results in

$$\begin{aligned} \mathbf{I}_M = \mathbf{R}_{\mathbf{nn}}^{-1/2}\mathbf{H} \left(\mathbf{H}^H\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}\right)^{-1} \mathbf{H}^H\mathbf{R}_{\mathbf{nn}}^{-H/2} \\ +\mathbf{R}_{\mathbf{nn}}^{H/2}\mathbf{B} \left(\mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbf{B}\right)^{-1} \mathbf{B}^H\mathbf{R}_{\mathbf{nn}}^{1/2} \end{aligned} \tag{59}$$

or equivalent

$$\begin{aligned} \mathbf{R}_{\mathbf{nn}} = \mathbf{H} \left(\mathbf{H}^H\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}\right)^{-1} \mathbf{H}^H \\ +\mathbf{R}_{\mathbf{nn}}\mathbf{B} \left(\mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbf{B}\right)^{-1} \mathbf{B}^H\mathbf{R}_{\mathbf{nn}} \end{aligned} \tag{60}$$

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

APPENDIX B
EQUIVALENCE BETWEEN (9) AND (10)

To remove the constraint $\mathbf{B}^H\mathbf{R_{ss}}\mathbf{B}=\mathbf{0}$, $\mathbf{R_{ss}}$ is parameterized as

$$\mathbf{R_{ss}}=\mathbf{H}\boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\mathbf{H}^H \tag{61}$$

and the constraint on the rank of $\mathbf{R_{ss}}$ can be directly transfered to the same constraint on the $L\times L$-dimensional matrix $\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}$.

The optimization problem is then reduced to

$$\underset{\text{rank}(\boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}})=R}{\arg\min} \parallel \mathbf{R}_{\mathbf{nn}}^{-1/2}\left(\mathbf{R_{yy}}-\mathbf{R_{nn}}-\mathbf{H}\boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\mathbf{H}^H\right)\mathbf{R}_{\mathbf{nn}}^{-H/2}\parallel_F^2. \tag{62}$$

Define the $M\times L$-dimensional matrix $\mathbf{C}$ and $M\times(M-L)$-dimensional matrix $\mathbf{C}^c$ as follows (assuming $\begin{bmatrix}\mathbf{H} & \mathbf{R_{nn}}\mathbf{B}\end{bmatrix}$ is full rank):

$$\begin{bmatrix}\mathbf{C}^H\\ \mathbf{C}^{c,H}\end{bmatrix}\begin{bmatrix}\mathbf{H} & \mathbf{R_{nn}}\mathbf{B}\end{bmatrix}=\begin{bmatrix}\mathbf{I}_L & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{M-L}\end{bmatrix} \tag{63}$$

Inserting an identity matrix on certain positions in (62) does not change the optimal value for $\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}$, so (62) is equivalent to:

$$\underset{\text{rank}(\boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}})=R}{\arg\min} \parallel \begin{bmatrix}\mathbf{R}_{\mathbf{nn}}^{-1/2}\mathbf{H} & \mathbf{R}_{\mathbf{nn}}^{1/2}\mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{C}^H\\ \mathbf{C}^{c,H}\end{bmatrix}(\mathbf{R_{yy}}-\mathbf{R_{nn}} -\mathbf{H}\boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\mathbf{H}^H)\begin{bmatrix}\mathbf{C} & \mathbf{C}^c\end{bmatrix}\begin{bmatrix}\mathbf{H}^H\mathbf{R}_{\mathbf{nn}}^{-H/2}\\ \mathbf{B}^H\mathbf{R}_{\mathbf{nn}}^{H/2}\end{bmatrix}\parallel_F^2. \tag{64}$$

(64) can then be further expanded to (65).

One can notice that only the left upper block part of the resulting matrix inside the Frobenius norm of (65) is influenced by $\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}$, so the optimization problem (65) can be simplified to:

$$\underset{\text{rank}(\boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}})=R}{\arg\min} \parallel \mathbf{R}_{\mathbf{nn}}^{-1/2}\mathbf{H}(\mathbf{C}^H\mathbf{R_{yy}}\mathbf{C}-\mathbf{C}^H\mathbf{R_{nn}}\mathbf{C}-\boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}})\mathbf{H}^H\mathbf{R}_{\mathbf{nn}}^{-H/2}\parallel_F^2 \tag{66}$$

One can also show that $\mathbf{C}^H\mathbf{R_{nn}}\mathbf{C}=\left(\mathbf{H}^H\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}\right)^{-1}$ by combining (60) and (63). So now by plugging in the definitions of (16) in (66), the reduced equivalent optimization problem to (9) is finally given by

$$\underset{\text{rank}(\boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}})=R}{\arg\min} \parallel \mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}^{-1/2}\left(\mathbf{R}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}-\mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}-\boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\right)\mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}^{-H/2}\parallel^2. \tag{67}$$

APPENDIX C
DERIVATION OF RANK R PK-GEVD-MWF

(5) Substituting estimate $\mathbf{R_{ss}}$ of (17) and $\mathbf{R_{nn}}$ as $\bar{\mathbf{R}}_{\mathbf{ss}}$ and $\bar{\mathbf{R}}_{\mathbf{nn}}$ respectively in (5) reveals to the following identities:

$$\begin{aligned}
\mathbf{W}_k&=\left(\mathbf{R_{nn}}+\mathbf{R_{ss}}\right)^{-1}\mathbf{R_{ss}}\mathbf{E}_{\mathbf{d}_k}\\
&=\left(\mathbf{R_{nn}}+\mathbf{H}\hat{\mathbb{Q}}\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\hat{\mathbb{Q}}^H\mathbf{H}^H\right)^{-1}\mathbf{H}\hat{\mathbb{Q}}\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\hat{\mathbb{Q}}^H\mathbf{H}^H\mathbf{E}_{\mathbf{d}_k}\\
&=\left(\mathbf{I}_M-\left(\mathbf{R_{nn}}+\mathbf{H}\hat{\mathbb{Q}}\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\hat{\mathbb{Q}}^H\mathbf{H}^H\right)^{-1}\mathbf{R_{nn}}\right)\mathbf{E}_{\mathbf{d}_k}\\
&=\Big(\mathbf{I}_M-\Big(\mathbf{R}_{\mathbf{nn}}^{-1}-\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}\hat{\mathbb{Q}}\Big((\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}})^{-1}+\hat{\mathbb{Q}}^H\mathbf{H}^H\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}\hat{\mathbb{Q}}\Big)^{-1}\hat{\mathbb{Q}}^H\mathbf{H}^H\mathbf{R}_{\mathbf{nn}}^{-1}\Big)\mathbf{R_{nn}}\Big)\mathbf{E}_{\mathbf{d}_k}\\
&=\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}\hat{\mathbb{Q}}\left((\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}})^{-1}+\hat{\mathbb{Q}}^H\mathbf{H}^H\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}\hat{\mathbb{Q}}\right)^{-1}\hat{\mathbb{Q}}^H\mathbf{H}^H\mathbf{E}_{\mathbf{d}_k}\\
&=\mathbf{C}\mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}^{-1}\hat{\mathbb{Q}}\left((\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}})^{-1}+\hat{\mathbb{Q}}^H\mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}^{-1}\hat{\mathbb{Q}}\right)^{-1}\hat{\mathbb{Q}}^H\mathbf{H}^H\mathbf{E}_{\mathbf{d}_k}\\
&=\mathbf{C}(\mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}+\hat{\mathbb{Q}}\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\hat{\mathbb{Q}}^H)^{-1}\hat{\mathbb{Q}}\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\hat{\mathbb{Q}}^H\mathbf{H}^H\mathbf{E}_{\mathbf{d}_k}\\
&=\mathbf{C}\hat{\mathbf{W}}\mathbf{H}^H\mathbf{E}_{\mathbf{d}_k}.
\end{aligned} \tag{68}$$

The matrix inversion lemma is used in the fourth step. In the sixth step, the identity $\mathbf{C}^H\mathbf{R_{nn}}\mathbf{C}=\mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}=\left(\mathbf{H}^H\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}\right)^{-1}$ derived from (60) and (64) is used, which implies also that $\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}=\mathbf{C}\mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}^{-1}$. In the last step, the GEVD-MWF expression for the reduce matrix pencil $\{\mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}},\mathbf{R}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}\}$ is identified:

$$\hat{\mathbf{W}}=(\mathbf{R}_{\hat{\mathbf{n}}\hat{\mathbf{n}}}+\hat{\mathbb{Q}}\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\hat{\mathbb{Q}}^H)^{-1}\hat{\mathbb{Q}}\Sigma_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\hat{\mathbb{Q}}^H \tag{69}$$

which is equivalent to the expression in (19).

APPENDIX D
PROOF OF THEOREM IV.2

To prove the sufficient statistic property of the compression $\mathbf{W}_{kk},\boldsymbol{\Lambda}_{kk}$, for each node k, an extra lemma is needed. For this theorem, the node-specific compression matrices $\mathbb{X}_B^k$, $\mathbb{\Lambda}_B^k$ and $\mathbf{J}_B^k$ are defined as

$$\mathbb{X}_B^k=\begin{bmatrix}\mathbf{0} & \mathbb{X}_{<k} & \mathbf{0}\\ \mathbf{I}_{M_k} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbb{X}_{>k}\end{bmatrix} \tag{70}$$

$$\mathbb{\Lambda}_B^k=\begin{bmatrix}\mathbf{0} & \mathbb{\Lambda}_{<k} & \mathbf{0}\\ \mathbf{I}_{L_k} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbb{\Lambda}_{>k}\end{bmatrix} \tag{71}$$

$$\mathbf{J}_B^k=\begin{bmatrix}\mathbf{0} & \mathbb{X}_{<k} & \mathbf{0} & \mathbf{B}_{<k}\mathbb{\Lambda}_{<k} & \mathbf{0}\\ \mathbf{I}_{M_k} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbb{X}_{>k} & \mathbf{0} & \mathbf{B}_{>k}\mathbb{\Lambda}_{>k}\end{bmatrix} \tag{72}$$

where[8] $\{.\}_{<k}=\text{Blkdiag}\{\{.\}_1,...,\{.\}_{k-1}\}$ and $\{.\}_{>k}=\text{Blkdiag}\{\{.\}_{k+1},...,\{.\}_K\}$.

Lemma D.1. *Consider the (compressed) maximization problem (related to* (22)*)*

$$\begin{aligned}\underset{\tilde{\mathbb{X}}_k\in\mathbb{C}^{M_k+2R(K-1)\times R}}{\arg\max}\quad & tr\{\tilde{\mathbb{X}}_k\mathbf{J}_B^{k,H}\mathbf{R_{yy}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k\}\\ s.t.\quad & \tilde{\mathbb{X}}_k^H\mathbf{J}_B^{k,H}\mathbf{R_{nn}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k=\mathbf{I}_R\\ & \mathbb{\Lambda}_B^{k,H}\mathbf{B}^H\mathbf{R_{nn}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k=\mathbf{0}.\end{aligned} \tag{73}$$

[8]It is noted that the diagonal blocks are not square here, i.e. in this case Blkdiag(.) is not a block-diagonal matrix in the strict sense.

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

$$\underset{\text{rank}(\boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}})=R}{\arg\min} \| \begin{bmatrix} \mathbf{R}_{\mathbf{nn}}^{-1/2}\mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{nn}}^{H/2}\mathbf{B} \end{bmatrix} \left(\begin{bmatrix} \mathbf{C}^H\mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}}\mathbf{C} & \mathbf{C}^H\mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}}\mathbf{C}_\perp \\ \mathbf{C}_\perp^H\mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}}\mathbf{C} & \mathbf{C}_\perp^H\mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}}\mathbf{C}_\perp \end{bmatrix} - \begin{bmatrix} \boldsymbol{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right) \begin{bmatrix} \mathbf{H}^H\mathbf{R}_{\mathbf{nn}}^{-H/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^H\mathbf{R}_{\mathbf{nn}}^{1/2} \end{bmatrix} \|_F^2 \tag{65}$$

The solution

$$\tilde{\mathbb{X}}_k = \begin{bmatrix} \mathbb{X}_{kk} \\ \mathbf{I}_{-k} \\ \mathbf{0}_{-k} \end{bmatrix} \tag{74}$$

fulfills the KKT-condition of (73) *with the same objective* $tr\{\tilde{\mathbb{X}}_k\mathbf{J}_B^{k,H}\mathbf{R}_{\mathbf{yy}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k\}$ *as the objective of the (non-compressed) maximization problem in* (22) *corresponding to its optimal solution* $\mathbb{X}$. *Furthermore the Lagrange parameter corresponding to the last constraint of* (73) *and solution* (74) *is given by*

$$\tilde{\mathbb{\Lambda}}_k = \begin{bmatrix} \mathbb{\Lambda}_{kk} \\ \mathbf{I}_{-k} \end{bmatrix}. \tag{75}$$

Proof. See Appendix E. □

Lemma D.1 states that problem (22) can be solved equivalently by solving (73) instead and this is the case for every node k. The proof of Theorem IV.2 can now be stated.

Proof. The proof starts by noticing that the obtained local sensor signals $\tilde{\mathbf{y}}_k^{i^\star}$ at node k can be related to the sensor signal $\mathbf{y}$ using the node-specific compression matrix $\mathbf{V}_B^{i^\star,k}$:

$$\mathbf{V}_B^{i^\star,k} = \mathbf{J}_B^k\mathbf{M}_B^{i^\star,k} \tag{76}$$

via $\tilde{\mathbf{y}}_k^{i^\star} = \mathbf{V}_B^{i^\star,k,H}\mathbf{y}$ where

$$\begin{aligned} \mathbf{M}_B^{i^\star k} = \text{Blkdiag}\{&\mathbf{I}_{M_k}, \mathbf{G}_1^{i^\star 1}, ..., \mathbf{G}_{k-1}^{i^\star 1}, \mathbf{G}_{k+1}^{i^\star 1}, ..., \mathbf{G}_K^{i^\star 1}, \\ &\mathbf{G}_1^{i^\star 2}, ..., \mathbf{G}_{k-1}^{i^\star 2}, \mathbf{G}_{k+1}^{i^\star 2}, ..., \mathbf{G}_K^{i^\star 2}\} \end{aligned} \tag{77}$$

is a fixed block-diagonal matrix, which is actually unknown at node k (except for the first $M_k \times M_k$ identity matrix in $\mathbf{M}_B^{i^\star,k}$). As shown in Section III, $\tilde{\mathbf{W}}_k^{i^\star+1}$ of node k can be computed via a transformation of the optimal solution $\tilde{\mathbb{Y}}_k$ of the following optimization problem:

$$\begin{aligned} \underset{\tilde{\mathbb{Y}}_k \in \mathbb{C}^{M_k+2R(K-1)\times R}}{\arg\max} \quad & \text{tr}\{\tilde{\mathbb{Y}}_k\mathbf{M}_B^{i^\star,k,H}\mathbf{J}_B^{k,H}\mathbf{R}_{\mathbf{yy}}\mathbf{J}_B^k\mathbf{M}_B^{i^\star,k}\tilde{\mathbb{Y}}_k\} \\ s.t. \quad & \tilde{\mathbb{Y}}_k^H\mathbf{M}_B^{i^\star,k,H}\mathbf{J}_B^{k,H}\mathbf{R}_{\mathbf{nn}}\mathbf{J}_B^k\mathbf{M}_B^{i^\star,k}\tilde{\mathbb{Y}}_k = \mathbf{I}_R \\ & \tilde{\mathbf{B}}_k^H\mathbf{M}_B^{i^\star,k,H}\mathbf{R}_{\mathbf{nn}}\mathbf{J}_B^k\mathbf{M}_B^{i^\star,k}\tilde{\mathbb{Y}}_k = \mathbf{0}. \end{aligned} \tag{78}$$

The last constraint can be written as

$$\begin{aligned} \text{Blkdiag}\{\mathbf{I}_{M_k}, \mathbf{G}_1^{i^\star 2}, ..., \mathbf{G}_{k-1}^{i^\star 2}, \mathbf{G}_{k+1}^{i^\star 2}, ..., \mathbf{G}_K^{i^\star 2}\}\mathbb{\Lambda}_B^{k,H} \\ \mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbf{J}_B^k\mathbf{M}_B^{i^\star,k}\tilde{\mathbb{Y}}_k = \mathbf{0} \end{aligned} \tag{79}$$

or equivalently (since $\mathbf{G}_k^{i^\star 2}$ are invertible matrices) as

$$\mathbb{\Lambda}_B^{k,H}\mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbf{J}_B^k\mathbf{M}_B^{i^\star,k}\tilde{\mathbb{Y}}_k = \mathbf{0}. \tag{80}$$

Since $\mathbf{M}_B^{i^\star,k}$ is non-singular, problem (78) is equivalent to problem (73) and as such Lemma D.1 gives the optimal solution of (78) via the relation

$$\mathbf{M}_B^{i^\star,k}\tilde{\mathbb{Y}}_k = \tilde{\mathbb{X}}_k = \begin{bmatrix} \mathbb{X}_{kk} \\ \mathbf{I}_{-k} \\ \mathbf{0}_{-k} \end{bmatrix}. \tag{81}$$

Since $\tilde{\mathbf{W}}_k^{i^\star+1}$ is given by a transformation of $\tilde{\mathbb{Y}}_k$, this implies

$$\tilde{\mathbf{W}}_k^{i^\star+1} = (\mathbf{M}_B^{i^\star,k})^{-1}\tilde{\mathbb{X}}_k\mathbf{G}_k^{i^\star+1} \tag{82}$$

with $\mathbf{G}_k^{i^\star+1}$ a transformation matrix. The first M_k rows of $\tilde{\mathbf{W}}_k^{i+1}$ can then be expressed as $\tilde{\mathbb{X}}_{kk}\mathbf{G}_k^{i^\star+1}$. Comparing this to (29), shows that $\mathbf{G}_k^{i^\star+1}$ can be chosen equal to $\mathbf{G}_k$ such that the optimal solution (18) is obtained. Equation (82) is then equivalent to (42).

As similar procedure can be used to show the same result for $\tilde{\mathbf{\Lambda}}_k^{i^\star}$, which concludes the proof. □

APPENDIX E
PROOF OF LEMMA D.1

Proof. As before, the Langrangian function [35] corresponding to the problem given in (73) can be written as:

$$\begin{aligned} L(\tilde{\mathbb{X}}_k, \boldsymbol{M}, \tilde{\mathbb{\Lambda}}_k) =& \text{tr}\{\tilde{\mathbb{X}}_k\mathbf{J}_B^{k,H}\mathbf{R}_{\mathbf{yy}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k\} \\ &+ \text{tr}\{\boldsymbol{M}^H(\tilde{\mathbb{X}}_k^H\mathbf{J}_B^{k,H}\mathbf{R}_{\mathbf{nn}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k - \mathbf{I}_R) \\ &+ (\tilde{\mathbb{X}}_k^H\mathbf{J}_B^{k,H}\mathbf{R}_{\mathbf{nn}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k - \mathbf{I}_R)\boldsymbol{M}\} \\ &- \text{tr}\{\tilde{\mathbb{\Lambda}}_k^H\mathbb{\Lambda}_B^{k,H}\mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k \\ &+ \tilde{\mathbb{X}}_k^H\mathbf{J}_B^{k,H}\mathbf{R}_{\mathbf{nn}}\mathbf{B}\mathbb{\Lambda}_B^k\tilde{\mathbb{\Lambda}}_k\}. \end{aligned} \tag{83}$$

The KKT condition can then be defined as the following set of equations

$$\begin{aligned} \nabla_{\tilde{\mathbb{X}}_k^*}L =& \mathbf{J}_B^{k,H}\mathbf{R}_{\mathbf{yy}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k + \mathbf{J}_B^{k,H}\mathbf{R}_{\mathbf{nn}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k(\boldsymbol{M} + \boldsymbol{M}^H) \\ &- \mathbf{J}_B^{k,H}\mathbf{R}_{\mathbf{nn}}\mathbf{B}\mathbb{\Lambda}_B^k\tilde{\mathbb{\Lambda}}_k = \mathbf{0} \\ \nabla_{\boldsymbol{M}}L =& \tilde{\mathbb{X}}_k^H\mathbf{J}_B^{k,H}\mathbf{R}_{\mathbf{nn}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k - \mathbf{I}_R = \mathbf{0} \\ \nabla_{\tilde{\mathbb{\Lambda}}_k^*}L =& \mathbb{\Lambda}_B^{k,H}\mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbf{J}_B^k\tilde{\mathbb{X}}_k = \mathbf{0}. \end{aligned} \tag{84}$$

Using (74) for $\tilde{\mathbb{X}}_k$ and (75) for $\tilde{\mathbb{\Lambda}}_k$, the equations can be transformed to

$$\begin{aligned} &\mathbf{J}_B^{k,H}\left(\mathbf{R}_{\mathbf{yy}}\mathbb{X} + \mathbf{R}_{\mathbf{nn}}\mathbb{X}(\boldsymbol{M} + \boldsymbol{M}^H) - \mathbf{R}_{\mathbf{nn}}\mathbf{B}\mathbb{\Lambda}\right) = \mathbf{0} \\ &\mathbb{X}^H\mathbf{R}_{\mathbf{nn}}\mathbb{X} - \mathbf{I}_R = \mathbf{0} \\ &\mathbb{\Lambda}_B^{k,H}\mathbf{B}^H\mathbf{R}_{\mathbf{nn}}\mathbb{X} = \mathbf{0}. \end{aligned} \tag{85}$$

The correctness of these equations can easily be verified since (24) holds for $\mathbb{X}$ and $\mathbb{\Lambda}$. Consequently, the same value for the (real) diagonal matrix $\boldsymbol{M}$ can be picked. Since for both problems (22) and (73), the objective is given by $\text{tr}\{\boldsymbol{M}\}$, this shows that these problems are equivalent. □

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks

REFERENCES

[1] D. Estrin, L. Girod, G. J. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001, pp. 2033–2036.

[2] C. Buratti, A. Conti, D. Dardari, and R. Verdone, "An overview on wireless sensor networks technology and evolution," *Sensors*, vol. 9, no. 9, pp. 6869–6896, 2009.

[3] T. Hee, S. Krishnamurthy, L. Luo, T. Yan, R. Stoleru, G. Zhou, Q. Cao, P. Vicaire, J. A. Stankovic, T. F. Abdelzaher, J. Hui, B. Krogh, S. Krishnamurthy, T. Yan, L. Gu, R. Stoleru, G. Zhou, P. Vicaire, J. A. Stancovic, J. Hui, and B. Krogh, "VigilNet: An Integrated Sensor Network System for Energy-Efficient Surveillance," *ACM Transactions on Sensor Networks*, vol. 2, no. 1, pp. 1–38, 2006.

[4] A. Shirazinia, S. Dey, D. Ciuonzo, and P. S. Rossi, "Massive MIMO for Decentralized Estimation of a Correlated Source," *IEEE Transactions on Signal Processing*, vol. 64, no. 10, pp. 2499–2512, 2016.

[5] D. Ciuonzo, P. S. Rossi, and P. K. Varshney, "Distributed Detection in Wireless Sensor Networks under Multiplicative Fading via Generalized Score-tests," *IEEE Internet of Things Journal*, pp. 1–13, 2021.

[6] L. Hu, X. Wang, and S. Wang, "Decentralized Underwater Target Detection and Localization," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 2385–2399, 2021.

[7] J. Kuriakose, S. Joshi, R. Vikram Raju, and A. Kilaru, "A Review on Localization in Wireless Sensor Networks," in *Advances in Signal Processing and Intelligent Recognition Systems*, S. M. Thampi, A. Gelbukh, and J. Mukhopadhyay, Eds. Cham: Springer International Publishing, 2014, pp. 599–610.

[8] 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, 2007.

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

[10] A. Bertrand and M. Moonen, "Distributed Adaptive Node-Specific Signal Estimation in Fully Connected Sensor NetworksPart I: Sequential Node Updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5277–5291, 2010.

[11] 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, 2016.

[12] S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 2, pp. 343–356, 2013.

[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, apr 2014.

[14] R. Ali, G. Bernardi, T. Van Waterschoot, and M. Moonen, "Methods of Extending a Generalized Sidelobe Canceller with External Microphones," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 27, no. 9, pp. 1349–1364, 2019.

[15] R. Ali, T. V. Waterschoot, and M. Moonen, "An integrated MVDR beamformer for speech enhancement using a local microphone array and external microphones," *EURASIP Journal on Audio, Speech, and Music Processing*, no. 10, pp. 1–20, 2021.

[16] H. Teutsch, "Modal array signal processing: Principles and applications of acoustic wavefield decomposition," *Lecture Notes in Control and Information Sciences*, vol. 348, pp. 1–258, 2007.

[17] 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 *International Workshop on Acoustic Signal Enhancement, IWAENC 2012*, no. September, 2012, pp. 4–6.

[18] R. Van Rompaey and M. Moonen, "Distributed adaptive node-specific signal estimation in a wireless sensor network with partial prior knowledge of the desired source steering vector," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5.

[19] R. Nassif, S. Vlaski, and A. H. Sayed, "Adaptation and Learning Over Networks Under Subspace Constraints Part I : Stability Analysis," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1346–1360, 2020.

[20] S. Kitić, L. Albera, N. Bertin, and R. Gribonval, "Physics-Driven Inverse Problems Made Tractable With Cosparse Regularization," *IEEE Transactions on Signal Processing*, vol. 64, no. 2, pp. 335–348, 2016.

[21] J. Murray-bruce and P. L. Dragotti, "A Sampling Framework for Solving Physics-Driven Inverse Source Problems," *IEEE Transactions on Signal Processing*, vol. 65, no. 24, pp. 6365–6380, 2017.

[22] M. Bellalouna and A. Ghabri, "A priori methods for fault tolerance in wireless sensor networks," *2013 World Congress on Computer and Information Technology, WCCIT 2013*, pp. 1–6, 2013.

[23] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A Consolidated Perspective on Multimicrophone Speech Enhancement and Source Separation," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, no. 4, pp. 692–730, 2017.

[24] A. Bertrand and M. Moonen, "Distributed node-specific LCMV beamforming in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 233–246, 2012.

[25] ——, "Distributed LCMV beamforming in a wireless sensor network with single-channel per-node signal transmission," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3447–3459, 2013.

[26] Z. Li, K. Fai, C. Yiu, Y.-h. Dai, and S. Nordholm, "Distributed LCMV beamformer design by randomly permuted ADMM," *Digital Signal Processing*, vol. 1, p. 102820, 2020.

[27] 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.

[28] 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)*. IEEE, aug 2016, pp. 1–5.

[29] X. L. Zhang and J. Wu, "Deep belief networks based voice activity detection," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 4, pp. 697–710, 2013.

[30] L. K. Hamaidi, M. Muma, and A. M. Zoubir, "Robust distributed multi-speaker voice activity detection using stability selection for sparse non-negative feature extraction," in *25th European Signal Processing Conference, EUSIPCO 2017*, vol. 2017-Janua, 2017, pp. 161–165.

[31] R. Van Rompaey and M. Moonen, "GEVD based speech and noise correlation matrix estimation for multichannel Wiener filter based noise reduction," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 2544–2548.

[32] B. 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, jun 2002.

[33] M. Dendrinos, S. Bakamidis, and G. Carayannis, "Speech enhancement from noise: A regenerative approach," *Speech Communication*, vol. 10, no. 1, pp. 45–57, feb 1991. [Online]. Available: https://www.sciencedirect.com/science/article/pii/016763939190027Q

[34] R. Serizel, M. Moonen, B. Van Dijk, and J. Wouters, "Rank-1 approximation based multichannel wiener filtering algorithms for noise reduction in cochlear implants," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, may 2013, pp. 8634–8638.

[35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[36] A. Bertrand and M. Moonen, "Distributed Adaptive Node-Specific Signal Estimation in Fully Connected Sensor NetworksPart II: Simultaneous and Asynchronous Node Updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5292–5306, 2010.

[37] J. Szurley, A. Bertrand, and M. Moonen, "Distributed adaptive node-specific signal estimation in heterogeneous and mixed-topology wireless sensor networks," *Signal Processing*, vol. 117, pp. 44–60, dec 2015.

[38] ——, "Topology-Independent Distributed Adaptive Node-Specific Signal Estimation in Wireless Sensor Networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 1, pp. 130–144, 2017.

[39] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[40] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," *Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1085–1099, 1994.

[41] A. S. Behbahani, A. M. Eltawil, and H. Jafarkhani, "Linear Decentralized Estimation of Correlated Data for Power-Constrained Wireless Sensor Networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 6003–6016, 2012.

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2021.3098941, IEEE Transactions on Signal and Information Processing over Networks



Robbe Van Rompaey Robbe Van Rompaey received the M.Sc. degree in electrical engineering from KU Leuven, Leuven, Belgium, in 2017 (magna cum laude), where he is currently pursuing the Ph.D. degree with the Electrical Engineering Department, under the supervision of Prof. M. Moonen. He is currently a Fellow of the Research Foundation Flanders (FWO). His research interests include multi-channel signal processing, wireless sensor networks, distributed optimization and distributed signal processing.



Marc Moonen Marc Moonen (M'94, SM'06, F'07) is a Full Professor at the Electrical Engineering Department of KU Leuven, where he is heading a research team working in the area of numerical algorithms and signal processing for digital communications, wireless communications, DSL and audio signal processing. He is a Fellow of the IEEE (2007) and a Fellow of EURASIP (2018). He received the 1994 KU Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon), and was a 1997 Laureate of the Belgium Royal Academy of Science. He received journal best paper awards from the IEEE Transactions on Signal Processing (with Geert Leus and with Daniele Giacobello) and from Elsevier Signal Processing (with Simon Doclo). He was chairman of the IEEE Benelux Signal Processing Chapter (1998-2002), a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications, and President of EURASIP (European Association for Signal Processing, 2007-2008 and 2011-2012). He is currently Vice-President for Publications of the IEEE Signal Processing Society (2021-2023). He has served as Editor-in-Chief for the EURASIP Journal on Applied Signal Processing (2003-2005), Area Editor for Feature Articles in IEEE Signal Processing Magazine (2012-2014), and has been a member of the editorial board of Signal Processing, IEEE Transactions on Circuits and Systems II, IEEE Signal Processing Magazine, Integration-the VLSI Journal, EURASIP Journal on Wireless Communications and Networking and EURASIP Journal on Advances in Signal Processing.

2373-776X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: KU Leuven Libraries. Downloaded on July 29,2021 at 12:01:26 UTC from IEEE Xplore. Restrictions apply.