Blind Room Acoustic System Identification in Wireless Acoustic Sensor Networks

Matthias Blochberger, *Member, IEEE,* Jesper Jensen, *Member, IEEE,* Filip Elvander, *Member, IEEE,* Jan Østergaard, *Senior Member, IEEE,* Marc Moonen, *Fellow, IEEE,* Toon van Waterschoot, *Member, IEEE*

***Abstract*—Blind room acoustic system identification is an important task in room-acoustic signal processing applications and as wireless acoustic sensor networks (WASNs) become a ubiquitous reality, it is natural to extend this task to the distributed signal processing domain. In this work, we propose a distributed blind system identification (BSI) algorithm for room acoustic systems based on the cross-relation (CR) method and the alternating direction method of multipliers (ADMM). The algorithm operates over WASNs, enabling the identification of room impulse responses (RIRs) without centralized data processing. We formulate the BSI problem as a modified general-form consensus problem, where sub-problems are solved by individual nodes in the network using a local quasi-Newton method, leading to a scalable solution for large sensor networks. The proposed method extends previous works by incorporating voice activity detection (VAD) for robust performance under non-stationary noisy signals. Additionally, we introduce regularisation to improve convergence and estimation performance whenever identifiability conditions of the BSI problem are not met. Simulation results demonstrate that the algorithm achieves significant improvements in terms of estimation accuracy and adaptability in both simulated and measured acoustic environments.**

***Index Terms*—blind system identification, distributed signal processing, multichannel signal processing, alternating direction method of multipliers, wireless sensor networks, quasi-Newton methods**

I. Introduction

WIRELESS sensor networks are becoming more relevant in various technical fields, driven by the ubiquity of network-connected devices and the growing Internet of Things. They find use in various applications such as environmental monitoring, industrial automation, and smart cities and are also gaining traction in acoustic and audio signal processing applications. Some of the first proposals for networked audio sensors, such as [1], and [2], started the trend, and now, with the advent of smart homes and smart cities, the potential of using network-connected devices with built-in microphones and loudspeakers, such as phones, tablets and more is ever-increasing. Generally, devices described as "smart" in the current marketing language possess the functionality necessary to be used within a wireless acoustic sensor network (WASN); however, state-of-the-art devices often do not use this full potential. Therefore, a valuable research direction is still to enable these devices' distributed signal processing capabilities maximally. In the context of the latter, distributed signal processing tasks such as, e.g., distributed signal estimation [3], [4], noise control and echo cancellation [5], [6], and beamforming [6], [7] have been studied in recent years. Fig. 1 shows an example of a WASN within a room, where each sensor node comprises a microphone, a processing unit, and a transmission unit.

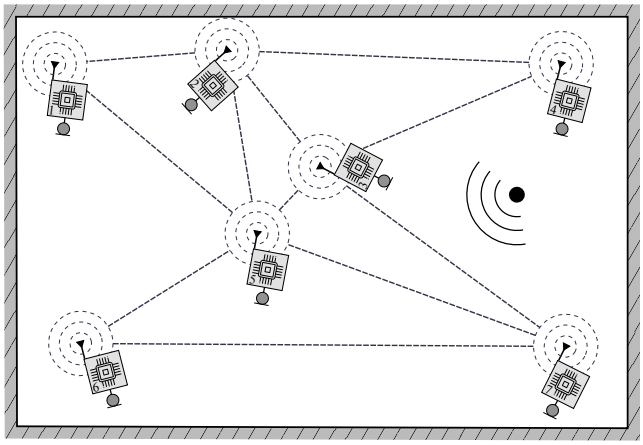

Fig. 1. A wireless acoustic sensor network within a room, where each sensor node comprises a microphone, a processing unit, and a transmission unit. Also shown is a speaker emitting sound into the room, i.e., the system input. The microphones receive the sound and each microphone signal represents a system output.

Blind system identification (BSI) refers to estimating the impulse responses of an unknown system without knowing the input signal, i.e., only the output signals are available. More specifically, we consider the single-input multiple-output (SIMO) room acoustic system. In this case, the term system refers to the combination of the room, the microphones, and the source with their respective positions, which results in room impulse responses (RIRs). A RIR describes the acoustic path the sound waves take from the source to the microphone, which comprises the direct path - if existent - and reflections from the room boundaries and objects within the room. The task of BSI is to estimate these RIRs, which can be used

This research work was carried out at the ESAT Laboratory of KU Leuven, and has received funding from KU Leuven internal funds (C14/21/075), Research Foundation Flanders (FWO projects S005525N and G0A0424N), the Flemish Government (AI Research Program), and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369. The scientific responsibility is assumed by the authors.

Matthias Blochberger, Marc Moonen and Toon van Waterschoot are with the Department of Electrical Engineering (ESAT), STADIUS, KU Leuven, Leuven, Belgium (e-mail: matthias.blochberger@kuleuven.be)

Filip Elvander is with the Dept. of Information and Communications Engineering, Aalto University, Espoo, Finland

Jesper Jensen and Jan Østergaard are with the Department of Electronic Systems, Aalborg University, Aalborg, Denmark

for signal enhancement tasks such as dereverberation, source separation, or acoustic echo cancellation as well as for room acoustics analysis, the computation of room acoustic parameters, e.g., the reverberation time RT60 or speech intelligibility predictors C_{50} [8], or room geometry estimation [9], [10]. The BSI problem was introduced in [11] and has since been studied extensively. Early algorithms used higher-order statistics [12]–[14] for channel estimation; however, the associated high computational complexity motivated research into second-order methods such as the cross-relation (CR) algorithm [15], [16], subspace algorithms [17]–[20], and maximum-likelihood algorithms [21]. Among multichannel BSI algorithms, adaptive CR based least mean squares (LMS) algorithms in the time and frequency domain are widely used, including the normalized multichannel frequency-domain LMS (NMCFLMS) [22], [23], its noise-robust and constrained variants [24]–[26], the multichannel quasi-Newton algorithm (MCQN) [27], and cross-correlation-based methods for early IR estimation [28]. Recently, [29] proposed a blind estimation algorithm for early-echo parameters in directional impulse responses. Significant contributions to BSI have also been made by work analysing identifiability with near-common zeros and noise [30], [31], state-space model-based approaches [32], [33], and Bayesian inference-based algorithms [34]–[40] partly applied in acoustic echo control, speech denoising, and dereverberation.

However, research into BSI in WASNs is limited, even more so in acoustics and audio signal processing applications. The work [41] proposes a distributed algorithm for estimating and equalising of room acoustics based on the common acoustical pole (CAP) room model and the distributed averaging approach. Further, although not for acoustic systems, [42]–[44] introduce time-domain distributed gradient descent and recursive estimation algorithms for distributed BSI. The distributed gradient descent (DGD)-based approach of [42] and the recursive approach in [44] based on the distributed stochastic approximation algorithm with expanding truncations (DSAAWET) [45] implemented as adaptive algorithms are taken into consideration in our study for performance comparison with the proposed algorithm in this paper.

This paper extends our previous work [46], [47], on distributed CR-based adaptive BSI in WASNs. We introduce a distributed algorithm for multichannel BSI in WASNs. We refer to it as a distributed adaptive alternating blind system identification (DABSI) algorithm. The BSI problem is formulated as a modified general-form consensus optimisation problem and solved in a distributed manner by the network nodes. For that, the alternating direction method of multipliers (ADMM) [48] is applied. This method is known to converge under mild conditions and is widely used in distributed optimisation problems. Our problem formulation separates the multichannel BSI problem into multiple lower-dimensional sub-problems with subsets of channels according to the sensor network's topology. The inter-channel cross-relations of the BSI problem are spread over nodes, such that each solves part of the original problem, i.e., each node solves a sub-problem using data from its network neighbours, and for each of the system's channels, multiple nodes form a consensus estimate. The ADMM update steps are applied in a block processing scheme forming an adaptive algorithm called Online-ADMM [49]–[51] with local quasi-Newton updates based on a Broyden class method, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [52], [53]. Compared to our previous work [46], [47], in this paper, we improve the distributed algorithm in several aspects to increase its robustness and performance, especially in speech-processing applications. The non-stationarity of a speech signal is challenging in adverse acoustic environments, with noise spectrally masking parts of the signal. We introduce a voice activity detection (VAD)-based extension, which introduces separate estimates for "signal-plus-noise" and "noise-only" parts of the signal. These are then used to increase the estimation accuracy by solving a generalized eigenvalue decomposition (GEVD) problem to find the parameter subspace. This concept is also used for distributed signal estimation in existing literature, such as [54], [55]. In [44], the concept of known noise characteristics is also used, making the comparison to our work interesting. Furthermore, we introduce a regularisation term to the distributed BSI problem to improve convergence behaviour. This regularisation term is based on prior knowledge of the system's RIRs, i.e., a covariance matrix of RIRs for a specific set of channels, similar to [56]. In this case, the covariance matrix can be of a rather general set of RIRs, e.g., rooms within a specific size range. We conduct simulation studies using simulated rooms and measurement data to evaluate the performance of the proposed algorithm.

The paper is structured as follows: Sec. II introduces the signal model, the cross-relation criterion, and the proposed noise-aware regularized objective. Sec. III derives the distributed formulation and the local BSI subproblems. Sec. IV presents the online-ADMM algorithm and distributed communication protocol. Sec. V discusses convergence behaviour and computational complexity. Sec. VI presents the numerical simulations, and Sec. VII concludes the paper.

Notation: Scalars, vectors, and matrices are denoted by lowercase, bold lowercase, and bold uppercase letters, respectively. Superscripts $(\cdot)^{\top}$, $(\cdot)^{\mathrm{H}}$, and $(\cdot)^{-1}$ denote transpose, Hermitian transpose, and inverse. Sets are denoted by calligraphic letters, with cardinality $|\cdot|$. The Euclidean norm is $\|\cdot\|$, $\mathrm{E}[\cdot]$ denotes expectation, and $\mathrm{Re}\{\cdot\}$ denotes the real part.

II. Problem Formulation

A. Signal and channel representation

We consider a SIMO acoustic system with M microphone signals. Let $\mathcal{M} \triangleq \{1, \dots, M\}$ denote the ordered set of microphone channels. The observed discrete-time microphone signal at channel $i \in \mathcal{M}$ is modeled as

$$x_i[n] = \sum_{m=0}^{L-1} h_i[m]s[n-m] + v_i[n], \qquad i \in \mathcal{M}, \quad (1)$$

where $s[n]$ is the unknown source signal, $h_i[n]$ is the RIR between the source and microphone i, L is the assumed RIR length, and $v_i[n]$ is additive noise. The time-domain RIR vector of channel i is denoted by

$$\bar{\mathbf{h}}_i = \begin{bmatrix} h_i[0] & \cdots & h_i[L-1] \end{bmatrix}^{\top} \in \mathbb{R}^L. \quad (2)$$

Its frequency-domain representation analogous to [22], [57] is

$$\mathbf{h}_i = \mathbf{F}_L \bar{\mathbf{h}}_i \in \mathbb{C}^L, \tag{3}$$

where $\mathbf{F}_L$ is the L-point discrete Fourier transform (DFT) matrix. In the following, $\mathbf{w}_i \in \mathbb{C}^L$ denotes the estimate of $\mathbf{h}_i$, and

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1^\top & \cdots & \mathbf{w}_M^\top \end{bmatrix}^\top \in \mathbb{C}^{ML} \tag{4}$$

denotes the stacked vector of all estimated RIRs.

B. Cross-relation blind system identification

BSI aims to estimate the RIRs $\{\mathbf{h}_i\}_{i\in\mathcal{M}}$ from the observed microphone signals only, without access to the source signal $s[n]$. Since the input signal is unknown, the RIRs are identifiable only up to a common scale factor. Moreover, under the usual causality assumption, the RIRs are identifiable only up to a common delay. The CR method exploits the fact that, in the noiseless case, cross-filtered microphone signals are equal for every pair of channels. In the time domain, the cross-convolved signal is

$$y_{ij}[n] \triangleq \sum_{m=0}^{L-1} w_j[m] x_i[n-m], \qquad i,j \in \mathcal{M}, \tag{5}$$

where $w_j[m]$ is the time-domain representation of the estimate of the j-th channel response. The CR identity is then

$$y_{ij}[n] = y_{ji}[n], \qquad i,j \in \mathcal{M},\ i \neq j. \tag{6}$$

This identity holds under the standard CR identifiability conditions, namely that the channel transfer functions have no common zeros and that the input signal is sufficiently persistently exciting [16], [58]. For the adaptive implementation used in this paper, the CR criterion is formulated in a frame-based overlap-save frequency-domain representation based on [23]. Let $\bar{\mathbf{x}}_i^{(\tau)} \in \mathbb{R}^{2L}$ denote the time-domain signal frame

$$\bar{\mathbf{x}}_i^{(\tau)} = \begin{bmatrix} x_i[\tau L] & x_i[\tau L - 1] & \cdots & x_i[\tau L - (2L-1)] \end{bmatrix}^\top,$$

with hop size L. Its DFT-domain representation is

$$\mathbf{x}_i^{(\tau)} = \mathbf{F}_{2L} \bar{\mathbf{x}}_i^{(\tau)} \in \mathbb{C}^{2L}.$$

The frequency-domain convolution matrix associated with channel i is defined as

$$\mathbf{X}_i^{(\tau)} = \mathbf{W}_{01} \mathbf{D}_i^{(\tau)} \mathbf{W}_{10} \in \mathbb{C}^{L\times L}, \tag{7}$$

where

$$\mathbf{D}_i^{(\tau)} = \operatorname{diag}\left\{ \mathbf{x}_i^{(\tau)} \right\} \tag{8}$$

contains the DFT-domain signal frame on its diagonal. The overlap-save matrices are

$$\mathbf{W}_{01} = \mathbf{F}_L \begin{bmatrix} \mathbf{0}_L & \mathbf{I}_L \end{bmatrix} \mathbf{F}_{2L}^{-1} \in \mathbb{C}^{L\times 2L}, \tag{9a}$$

$$\mathbf{W}_{10} = \mathbf{F}_{2L} \begin{bmatrix} \mathbf{I}_L & \mathbf{0}_L \end{bmatrix}^\top \mathbf{F}_L^{-1} \in \mathbb{C}^{2L\times L}. \tag{9b}$$

The corresponding frame-domain representation of the cross-convolved signal in (5) is

$$\mathbf{y}_{ij}^{(\tau)} = \mathbf{X}_i^{(\tau)} \mathbf{w}_j \in \mathbb{C}^L. \tag{10}$$

Hence, the frame-domain CR error between channels i and j is

$$\mathbf{e}_{ij}^{(\tau)} = \mathbf{y}_{ij}^{(\tau)} - \mathbf{y}_{ji}^{(\tau)} = \mathbf{X}_i^{(\tau)} \mathbf{w}_j - \mathbf{X}_j^{(\tau)} \mathbf{w}_i. \tag{11}$$

Collecting the pairwise CR relations into an equivalent stacked operator gives

$$\mathbf{e}^{(\tau)} = \mathbf{X}^{(\tau)} \mathbf{w} \in \mathbb{C}^{ML}, \tag{12}$$

where the CR operator $\mathbf{X}^{(\tau)} \in \mathbb{C}^{ML\times ML}$ is

$$\mathbf{X}^{(\tau)} \triangleq \begin{bmatrix} \sum_{i\neq 1} \mathbf{X}_i^{(\tau)} & -\mathbf{X}_2^{(\tau)} & \cdots & -\mathbf{X}_M^{(\tau)} \\ -\mathbf{X}_1^{(\tau)} & \sum_{i\neq 2} \mathbf{X}_i^{(\tau)} & \cdots & -\mathbf{X}_M^{(\tau)} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{X}_1^{(\tau)} & -\mathbf{X}_2^{(\tau)} & \cdots & \sum_{i\neq M} \mathbf{X}_i^{(\tau)} \end{bmatrix}. \tag{13}$$

In the noiseless case, $\mathbf{e}^{(\tau)} = \mathbf{0}$, and the BSI problem becomes a null-space problem for $\mathbf{X}^{(\tau)}$. In noisy environments, the exact null-space relation is replaced by the quadratic CR criterion

$$\begin{aligned} \min_{\mathbf{w}} \quad & f_{\mathrm{CR}}(\mathbf{w}) \\ \text{s.t.} \quad & \|\mathbf{w}\| = 1, \end{aligned} \tag{14}$$

with

$$f_{\mathrm{CR}}(\mathbf{w}) = \mathrm{E}\left[\left\| \mathbf{X}^{(\tau)} \mathbf{w} \right\|^2 \right] = \mathbf{w}^{\mathrm{H}} \mathbf{C} \mathbf{w}, \tag{15}$$

where

$$\mathbf{C} = \mathrm{E}\left[\mathbf{X}^{(\tau)\mathrm{H}} \mathbf{X}^{(\tau)} \right] \tag{16}$$

is the CR matrix. The unit-norm constraint avoids the trivial solution $\mathbf{w} = \mathbf{0}$. In practice, $\mathbf{C}$ is estimated by exponential recursive averaging of $\mathbf{X}^{(\tau)\mathrm{H}} \mathbf{X}^{(\tau)}$ with forgetting factor $0 \ll \eta < 1$, which is specialized to separate VAD-gated estimates in the next subsection. The classical CR objective is the starting point of the proposed method. The next subsection introduces a noise-aware modification based on separate signal-plus-noise and noise-only CR matrices.

C. Noise-aware generalized eigenvalue formulation

The classical CR criterion treats all frames equally. For speech signals, however, the observed microphone signals contain both signal-plus-noise periods and noise-only periods [54], [55], [59], [60]. We therefore assume that a VAD is available and provides a binary decision for each frame. This recursive estimate is split into two VAD-gated CR matrices, a signal-plus-noise matrix $\mathbf{C}_{\mathrm{x}}$ and a noise-only matrix $\mathbf{C}_{\mathrm{v}}$, updated as

$$\mathbf{C}_{\mathrm{x}}^{(\tau+1)} = \begin{cases} \eta\, \mathbf{C}_{\mathrm{x}}^{(\tau)} + (1-\eta) \mathbf{X}^{(\tau)\mathrm{H}} \mathbf{X}^{(\tau)}, & \text{if VAD} = 1, \\ \mathbf{C}_{\mathrm{x}}^{(\tau)}, & \text{otherwise}, \end{cases} \tag{17}$$

$$\mathbf{C}_{\mathrm{v}}^{(\tau+1)} = \begin{cases} \eta\, \mathbf{C}_{\mathrm{v}}^{(\tau)} + (1-\eta) \mathbf{X}^{(\tau)\mathrm{H}} \mathbf{X}^{(\tau)}, & \text{if VAD} = 0, \\ \mathbf{C}_{\mathrm{v}}^{(\tau)}, & \text{otherwise}. \end{cases} \tag{18}$$

Assuming that the source component and the noise are independent, the CR matrix associated with the clean source component is

$$\mathbf{C}_{\mathrm{s}} = \mathbf{C}_{\mathrm{x}} - \mathbf{C}_{\mathrm{v}}. \tag{19}$$

In the ideal case, the BSI solution satisfies

$$\mathbf{C}_{\mathrm{s}}\mathbf{w} = (\mathbf{C}_{\mathrm{x}} - \mathbf{C}_{\mathrm{v}})\,\mathbf{w} = \mathbf{0}. \tag{20}$$

Equivalently,

$$\mathbf{C}_{\mathrm{x}}\mathbf{w} = \lambda \mathbf{C}_{\mathrm{v}}\mathbf{w}, \tag{21}$$

which is a generalized eigenvalue problem. The desired estimate corresponds to the smallest generalized eigenvalue. Instead of computing a full generalized eigenvalue decomposition, we minimize the corresponding generalized Rayleigh quotient,

$$\begin{aligned} \min_{\mathbf{w}} \quad & f_{\mathrm{GEV}}(\mathbf{w}) \\ \text{s.t.} \quad & \|\mathbf{w}\| = 1, \end{aligned} \tag{22}$$

where

$$f_{\mathrm{GEV}}(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{H}}\mathbf{C}_{\mathrm{x}}\mathbf{w}}{\mathbf{w}^{\mathrm{H}}\mathbf{C}_{\mathrm{v}}\mathbf{w}}. \tag{23}$$

Since this formulation is based on $\mathbf{C}_{\mathrm{s}} = \mathbf{C}_{\mathrm{x}} - \mathbf{C}_{\mathrm{v}}$, it preserves the classical CR null-space solution in the ideal case. The use of this criterion in the distributed online-ADMM algorithm is described in Sec. III and Sec. IV.

D. Regularized estimation problem

The BSI problem can become ill posed when the identifiability conditions are only approximately satisfied, for example when the input signal is not persistently exciting or when the channel transfer functions contain near-common zeros [30], [31]. To improve robustness in such cases, we incorporate prior information about plausible channel responses through a regularisation term. Although the CR objective is not a likelihood function and therefore does not directly yield a Bayesian estimator, a regularizer can be constructed in the spirit of a zero-mean Gaussian prior on the RIRs. Let $\mathbf{\Sigma}_{h_i} \in \mathbb{C}^{L\times L}$ denote a known covariance matrix describing plausible RIRs in the DFT domain. For a stacked channel vector $\mathbf{w}$, we define the block-diagonal covariance matrix

$$\mathbf{\Sigma}_{\mathbf{w}} = \mathrm{blkdiag}\left(\mathbf{\Sigma}_{h_i}, \ldots, \mathbf{\Sigma}_{h_i}\right), \tag{24}$$

with one block for each channel participating in the corresponding problem. The regularisation term is then

$$g(\mathbf{w}) = \frac{\lambda}{2}\mathbf{w}^{\mathrm{H}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{w}, \tag{25}$$

where $\lambda \in \mathbb{R}_+$ controls the strength of the prior. This prior as a regularisation term is consistent with the observations in [56] that the optimal regularizer comprises the inverse of the covariance matrix of the room impulse responses, here channel responses in the DFT domain. Since the power of the non-stationary speech signal changes over time, the relative scaling between the data-dependent objective and the regularisation term also changes. We therefore use an adaptive regularisation parameter

$$\lambda^{(\tau)} = \lambda_0 \frac{\mathrm{tr}\left(\mathbf{C}_{\mathrm{x}}^{(\tau)}\right)}{(M-1)\,\mathrm{tr}\left(\mathbf{\Sigma}_{\mathbf{w}}^{-1}\right)}, \tag{26}$$

where $0 < \lambda_0 \ll 1$. Combining the noise-aware CR criterion with the prior-based regularisation yields the centralized estimation problem

$$\begin{aligned} \min_{\mathbf{w}} \quad & f_{\mathrm{GEV}}(\mathbf{w}) + g(\mathbf{w}) \\ \text{s.t.} \quad & \|\mathbf{w}\| = 1. \end{aligned} \tag{27}$$

This is the objective addressed in the remainder of the paper. A centralized solution of (27) would require access to all microphone signals and would involve optimization over the full stacked channel vector $\mathbf{w}$. The following section therefore derives a distributed formulation in which each node solves a local subproblem using only its neighbourhood information, while consensus constraints enforce agreement between overlapping channel estimates.

III. Distributed Formulation

A. Wireless acoustic sensor network

We consider a WASN with M sensor nodes. Each node comprises a microphone, a processing unit, and a transmission unit. For notational simplicity, each node is assumed to contain one microphone, such that the node index and microphone-channel index coincide. Generalization to nodes with multiple microphones is straightforward. We use the ordered set $\mathcal{M} = \{1, \ldots, M\}$ to denote both the set of nodes and the set of acoustic channels. A centralized solution of (27) would require collecting all microphone signals at a central processor and solving an optimization problem over the full stacked channel vector $\mathbf{w} \in \mathbb{C}^{ML}$. In contrast, the distributed approach considered here lets each node solve a lower-dimensional BSI subproblem based only on signals from a local neighbourhood. The nodes exchange intermediate variables so that overlapping channel estimates agree through consensus constraints. The communication topology is described by a directed graph

$$\mathcal{G} \triangleq \{\mathcal{M}, \mathcal{E}\}, \tag{28}$$

where $\mathcal{E}$ is the set of directed edges. The edge $(i, j) \in \mathcal{E}$ means that signal information originating from node i is available at node j. We define the transmit-neighbourhood and receive-neighbourhood of node i as

$$\mathcal{T}_i = (o_1, \ldots, o_{T_i}) = \{j \in \mathcal{M} \mid (i, j) \in \mathcal{E}\}, \tag{29a}$$

$$\mathcal{R}_i = (r_1, \ldots, r_{R_i}) = \{j \in \mathcal{M} \mid (j, i) \in \mathcal{E}\}, \tag{29b}$$

where the elements are ordered according to their order in $\mathcal{M}$, and where

$$T_i \triangleq |\mathcal{T}_i|, \qquad R_i \triangleq |\mathcal{R}_i|. \tag{30}$$

If $\mathcal{T}_i = \mathcal{R}_i$ for all $i \in \mathcal{M}$, the directed graph is equivalent to an undirected graph. Fig. 2 (left) and Fig. 2 (right) show examples of such network topologies, one densely, the other sparsely connected. We assume that each node has access to its own signal and estimate, i.e., $i \in \mathcal{R}_i$ and $i \in \mathcal{T}_i$, while the corresponding self-loop is omitted from the topology figures for visual clarity. The direction of an edge refers to the mathematical topology of the distributed BSI problem, i.e., the direction in which signal information is used. Further, we assume $R_i \geq 2$ for all nodes, since at least two channels are required to form a local cross-relation subproblem. The actual

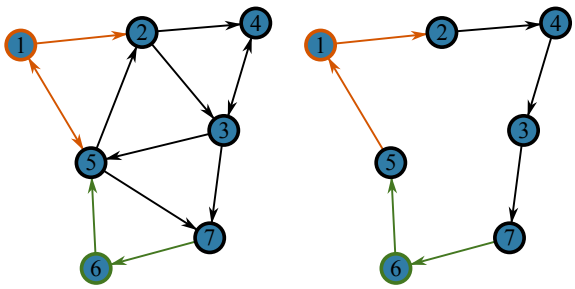

Fig. 2. Two example network topologies: (left) densely connected, i.e., nodes have many neighbours and (right) sparsely connected, i.e., nodes have few neighbours. Omitted are the self-loops (i, i) for all nodes $i \in \mathcal{M}$.

communication during the ADMM iterations is bidirectional, since intermediate variables are exchanged both along and against these edges. We assume throughout that inter-node communication is error-free and instantaneous. Furthermore, the algorithm is described in a synchronous frame-based setting: at each frame (τ), all nodes perform one local update, exchange variables with their neighbours, update consensus variables, and then update their dual variables.

B. Local BSI subproblems

At node i, the local BSI problem is formed from the channels in the receive-neighbourhood $\mathcal{R}_i$. Let

$$\mathcal{R}_i = (r_1, \ldots, r_{R_i}) \tag{31}$$

and define the local primal variable

$$\mathbf{w}_i = \begin{bmatrix} \mathbf{w}_{i|r_1}^\top & \cdots & \mathbf{w}_{i|r_{R_i}}^\top \end{bmatrix}^\top \in \mathbb{C}^{R_i L}. \tag{32}$$

The notation $\mathbf{w}_{i|j}$ denotes the estimate, stored at node i, of the channel response associated with node j. Thus, the subscript before the vertical bar denotes the holder of the variable, whereas the subscript after the vertical bar denotes the node/channel to which the variable corresponds. Node i constructs a local CR operator using only the signal matrices $\{\mathbf{X}_r^{(\tau)}\}_{r\in\mathcal{R}_i}$. We denote this local operator by

$$\mathbf{X}_{i,\mathrm{loc}}^{(\tau)} \in \mathbb{C}^{R_i L \times R_i L}. \tag{33}$$

It is obtained from the same block construction as the centralized CR operator (13), restricted to the channels in $\mathcal{R}_i$. More explicitly, the (a, b)-th $L \times L$ block of $\mathbf{X}_{i,\mathrm{loc}}^{(\tau)}$, with $a, b \in \{1, \ldots, R_i\}$, is

$$\left[\mathbf{X}_{i,\mathrm{loc}}^{(\tau)}\right]_{a,b} = \begin{cases} \sum\limits_{\substack{\ell\in\mathcal{R}_i \\ \ell\neq r_a}} \mathbf{X}_\ell^{(\tau)}, & a = b, \\ -\mathbf{X}_{r_b}^{(\tau)}, & a \neq b. \end{cases} \tag{34}$$

The corresponding local CR error is

$$\mathbf{e}_i^{(\tau)} = \mathbf{X}_{i,\mathrm{loc}}^{(\tau)} \mathbf{w}_i. \tag{35}$$

The local matrices $\mathbf{C}_{\mathrm{x},i}$ and $\mathbf{C}_{\mathrm{v},i}$ are obtained by the same VAD-dependent recursion as in (17) and (18) with $\mathbf{X}^{(\tau)}$ replaced by $\mathbf{X}_{i,\mathrm{loc}}^{(\tau)}$. For simplicity, we assume that all nodes use the same binary VAD decision per frame; in practice, this may be obtained locally and exchanged as a one-bit side-information signal, or computed from a reference node. The local generalized eigenvalue (GEV) objective is

$$f_i(\mathbf{w}_i) = \frac{\mathbf{w}_i^{\mathrm{H}} \mathbf{C}_{\mathrm{x},i} \mathbf{w}_i}{\mathbf{w}_i^{\mathrm{H}} \mathbf{C}_{\mathrm{v},i} \mathbf{w}_i}. \tag{36}$$

The local regularisation term is

$$g_i(\mathbf{w}_i) = \frac{\lambda_i}{2} \mathbf{w}_i^{\mathrm{H}} \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1} \mathbf{w}_i, \tag{37}$$

where

$$\boldsymbol{\Sigma}_{\mathbf{w}_i} = \mathrm{blkdiag}\left(\{\boldsymbol{\Sigma}_{h_r}\}_{r\in\mathcal{R}_i}\right) \in \mathbb{C}^{R_i L \times R_i L}. \tag{38}$$

If the same prior covariance is used for all channels, then $\boldsymbol{\Sigma}_{h_r} = \boldsymbol{\Sigma}_h$ for all $r \in \mathcal{R}_i$. The local adaptive regularisation parameter is chosen as

$$\lambda_i^{(\tau)} = \lambda_0 \frac{\mathrm{tr}\left(\mathbf{C}_{\mathrm{x},i}^{(\tau)}\right)}{(R_i - 1)\,\mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\right)}, \tag{39}$$

where $0 < \lambda_0 \ll 1$. The factor $R_i - 1$ accounts for the number of CR partners per local channel estimate and is the local counterpart of the corresponding normalization in the centralized problem. Thus, the unconstrained local objective at node i is

$$f_i(\mathbf{w}_i) + g_i(\mathbf{w}_i). \tag{40}$$

Since neighbouring nodes estimate overlapping channel responses, the local variables in (32) must be coupled through consensus constraints.

C. Consensus formulation

For every channel $j \in \mathcal{M}$, we introduce a consensus variable

$$\mathbf{z}_{j|j} \in \mathbb{C}^L, \tag{41}$$

which represents the network-wide estimate of the channel response associated with node j. Copies of this consensus variable may be available at neighbouring nodes and are denoted by $\mathbf{z}_{i|j}$, where i is the node holding the copy. Let

$$\mathbf{A}_{i|j} \in \mathbb{R}^{L \times R_i L} \tag{42}$$

be the selection matrix that extracts from $\mathbf{w}_i$ the local copy $\mathbf{w}_{i|j}$, for $j \in \mathcal{R}_i$, i.e.,

$$\mathbf{A}_{i|j} \mathbf{w}_i = \mathbf{w}_{i|j}. \tag{43}$$

The distributed BSI problem is then formulated as the modified general-form consensus problem [46], [48], [61]

$$\begin{aligned} \min_{\{\mathbf{w}_i, \mathbf{z}_{i|i}\}_{i\in\mathcal{M}}} \quad & \sum_{i\in\mathcal{M}} \left(f_i(\mathbf{w}_i) + g_i(\mathbf{w}_i)\right) \\ \text{s.t.} \quad & \mathbf{A}_{i|j}\mathbf{w}_i = \mathbf{z}_{i|j} \quad \forall\, i \in \mathcal{M},\ j \in \mathcal{R}_i, \\ & \mathbf{z}_{i|i} \in \mathcal{C}_i \quad \forall\, i \in \mathcal{M}. \end{aligned} \tag{44}$$

The first constraint enforces agreement between the local estimate $\mathbf{w}_{i|j}$ and the corresponding consensus variable. The second constraint imposes a distributed version of the global non-triviality constraint, preventing the zero solution while avoiding the need for any node to know all channel estimates. The construction of the feasible sets $\mathcal{C}_i$ is described in Sec. IV-B.

IV. Distributed Online-ADMM Algorithm

A. Online-ADMM updates

We solve (44) using an online ADMM procedure, where one ADMM iteration is performed per signal frame. The frame index (τ) is therefore both the time index of the adaptive signal-processing problem and the iteration index of the online optimization algorithm [50], [51], [62]. The real-valued augmented Lagrangian is

$$\mathcal{L}_\rho = \sum_{i\in\mathcal{M}} \left(f_i(\mathbf{w}_i) + g_i(\mathbf{w}_i) + r_i(\mathbf{w}_i, \mathbf{z}_{i|i}, \{\mathbf{u}_{i|j}\}_{j\in\mathcal{R}_i}) \right), \tag{45}$$

where $\rho > 0$ is the ADMM penalty parameter and $\mathbf{u}_{i|j} \in \mathbb{C}^L$ is the dual variable associated with the consensus constraint on edge (j, i). The augmented term at node i is

$$\begin{aligned} &r_i(\mathbf{w}_i, \mathbf{z}_{i|i}, \{\mathbf{u}_{i|j}\}_{j\in\mathcal{R}_i}) \\ &\quad= \sum_{j\in\mathcal{R}_i} 2\,\mathrm{Re}\left\{ \mathbf{u}_{i|j}^{\mathrm{H}} \left(\mathbf{A}_{i|j}\mathbf{w}_i - \mathbf{z}_{i|j} \right) \right\} \\ &\qquad+ \frac{\rho}{2} \sum_{j\in\mathcal{R}_i} \left\| \mathbf{A}_{i|j}\mathbf{w}_i - \mathbf{z}_{i|j} \right\|^2 + \mathcal{I}_{\mathcal{C}_i}(\mathbf{z}_{i|i}), \end{aligned} \tag{46}$$

with

$$\mathcal{I}_{\mathcal{C}_i}(\mathbf{z}_{i|i}) = \begin{cases} 0, & \text{if } \mathbf{z}_{i|i} \in \mathcal{C}_i, \\ \infty, & \text{otherwise.} \end{cases} \tag{47}$$

The non-triviality constraint is imposed on the consensus variables rather than the local variables. This keeps the local objectives smooth while still avoiding the trivial all-zero solution. At every frame (τ), all nodes synchronously perform the following updates:

$$\mathbf{w}_i^{(\tau+1)} = \mathbf{w}_i^{(\tau)} - \mu \mathbf{V}_i^{(\tau)} \mathbf{g}_i^{(\tau)}(\mathbf{w}_i^{(\tau)}), \tag{48a}$$

$$\mathbf{y}_{i|j}^{(\tau+1)} = \mathbf{A}_{i|j}\mathbf{w}_i^{(\tau+1)} + \rho^{-1}\mathbf{u}_{i|j}^{(\tau)}, \tag{48b}$$

$$\mathbf{y}_{j|i}^{(\tau+1)} = \mathbf{y}_{i|j}^{(\tau+1)}, \tag{48c}$$

$$\mathbf{z}_{i|i}^{(\tau+1)} = \Pi_{\mathcal{C}_i}\left(\frac{1}{T_i} \sum_{j\in\mathcal{T}_i} \mathbf{y}_{i|j}^{(\tau+1)} \right), \tag{48d}$$

$$\mathbf{z}_{j|i}^{(\tau+1)} = \mathbf{z}_{i|i}^{(\tau+1)}, \tag{48e}$$

$$\mathbf{u}_{i|j}^{(\tau+1)} = \mathbf{u}_{i|j}^{(\tau)} + \rho\left(\mathbf{A}_{i|j}\mathbf{w}_i^{(\tau+1)} - \mathbf{z}_{i|j}^{(\tau+1)} \right). \tag{48f}$$

The copy equalities (48c) and (48e) for the $\mathbf{y}$- and $\mathbf{z}$-variables denote communication operations. As for $\mathbf{w}_{i|j}$ and $\mathbf{z}_{i|j}$, the subscript before the vertical bar denotes the node holding the variable, and the subscript after the vertical bar denotes the associated node/channel. Thus, node i computes $\mathbf{y}_{i|j}$ for $j \in \mathcal{R}_i$ and transmits it to node j, where it is stored as $\mathbf{y}_{j|i}$. The consensus update at node i then averages the received variables $\{\mathbf{y}_{i|j}\}_{j\in\mathcal{T}_i}$. The local gradient in (48a) is

$$\begin{aligned} \mathbf{g}_i^{(\tau)}(\mathbf{w}_i) &= \nabla_{\mathbf{w}_i^{\mathrm{H}}} \mathcal{L}_\rho^{(\tau)} \\ &= \nabla_{\mathbf{w}_i^{\mathrm{H}}} f_i(\mathbf{w}_i) + \nabla_{\mathbf{w}_i^{\mathrm{H}}} g_i(\mathbf{w}_i) \\ &\quad+ \sum_{j\in\mathcal{R}_i} \mathbf{A}_{i|j}^{\mathrm{H}} \left(\mathbf{u}_{i|j} + \rho\left(\mathbf{A}_{i|j}\mathbf{w}_i - \mathbf{z}_{i|j} \right) \right). \end{aligned} \tag{49}$$

For the GEV objective in (36), define $a_i(\mathbf{w}_i) = \mathbf{w}_i^{\mathrm{H}}\mathbf{C}_{\mathrm{x},i}\mathbf{w}_i$ and $b_i(\mathbf{w}_i) = \mathbf{w}_i^H\mathbf{C}_{\mathrm{v},i}\mathbf{w}_i$. Assuming $\mathbf{C}_{\mathrm{x},i}$ and $\mathbf{C}_{\mathrm{v},i}$ are Hermitian and $b_i(w_i) > 0$, the Wirtinger derivative with respect to $\mathbf{w}_i^{\mathrm{H}}$ is

$$\nabla_{\mathbf{w}_i^{\mathrm{H}}} f_i(\mathbf{w}_i) = \frac{\left(\mathbf{C}_{\mathrm{x},i} - f_i(\mathbf{w}_i)\mathbf{C}_{\mathrm{v},i} \right)\mathbf{w}_i}{\mathbf{w}_i^{\mathrm{H}}\mathbf{C}_{\mathrm{v},i}\mathbf{w}_i}. \tag{50}$$

In the implementation, the positive scalar factor $(\mathbf{w}_i^{\mathrm{H}}\mathbf{C}_{\mathrm{v},i}\mathbf{w}_i)^{-1}$ is absorbed into the step size. Hence, the local quasi-Newton update uses the equivalent descent direction

$$\left(\mathbf{C}_{\mathrm{x},i} - f_i(\mathbf{w}_i)\mathbf{C}_{\mathrm{v},i} \right)\mathbf{w}_i. \tag{51}$$

The regularisation gradient is

$$\nabla_{\mathbf{w}_i^{\mathrm{H}}} g_i(\mathbf{w}_i) = \lambda_i \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\mathbf{w}_i, \tag{52}$$

after complex-valued Wirtinger derivatives. The matrix $\mathbf{V}_i^{(\tau)}$ in (48a) is an estimate of the inverse Hessian used for a quasi-Newton update similar to [27]. We update it using the BFGS [52], [53] recursion

$$\Delta\mathbf{w}_i^{(\tau+1)} = \mathbf{w}_i^{(\tau+1)} - \mathbf{w}_i^{(\tau)}, \tag{53a}$$

$$\Delta\mathbf{g}_i^{(\tau+1)} = \mathbf{g}_i^{(\tau+1)}(\mathbf{w}_i^{(\tau+1)}) - \mathbf{g}_i^{(\tau+1)}(\mathbf{w}_i^{(\tau)}), \tag{53b}$$

$$\gamma_i^{(\tau+1)} = \mathrm{Re}\left\{ \Delta\mathbf{g}_i^{(\tau+1)\mathrm{H}} \Delta\mathbf{w}_i^{(\tau+1)} \right\}, \tag{53c}$$

$$\begin{aligned} \mathbf{V}_i^{(\tau+1)} &= \left(\mathbf{I} - \frac{\Delta\mathbf{w}_i^{(\tau+1)}\Delta\mathbf{g}_i^{(\tau+1)\mathrm{H}}}{\gamma_i^{(\tau+1)}} \right) \mathbf{V}_i^{(\tau)} \\ &\quad\cdot \left(\mathbf{I} - \frac{\Delta\mathbf{g}_i^{(\tau+1)}\Delta\mathbf{w}_i^{(\tau+1)\mathrm{H}}}{\gamma_i^{(\tau+1)}} \right) \\ &\quad+ \frac{\Delta\mathbf{w}_i^{(\tau+1)}\Delta\mathbf{w}_i^{(\tau+1)\mathrm{H}}}{\gamma_i^{(\tau+1)}}. \end{aligned} \tag{53d}$$

The quasi-Newton step improves stability for a potentially poorly conditioned BSI problem while avoiding the explicit matrix inversion required by a full Newton method.

B. Distributed non-triviality constraint

The centralized BSI problems (14) and (27) require a non-triviality constraint to avoid the zero solution. In the distributed formulation, this constraint must be enforced without requiring any node to know all RIR estimates. We therefore impose the constraint through the consensus variables [47]

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_{1|1}^\top & \cdots & \mathbf{z}_{M|M}^\top \end{bmatrix}^\top \in \mathbb{C}^{ML}. \tag{54}$$

The centralized feasible set is $\mathcal{C} = \left\{ \mathbf{z} \mid \|\mathbf{z}\|^2 = 1 \right\}$ and the corresponding projection is

$$\Pi_{\mathcal{C}}(\mathbf{z}) = \frac{\mathbf{z}}{\|\mathbf{z}\|} = \frac{\mathbf{z}}{\sqrt{\sum_{i\in\mathcal{M}} \|\mathbf{z}_{i|i}\|^2}}. \tag{55}$$

This projection is not directly computable at a single node unless the squared norms of all consensus variables are available. To avoid global communication, we replace the denominator by a distributed estimate. For each node i, define the local feasible set

$$\mathcal{C}_i = \left\{ \mathbf{z}_{i|i} \mid \sum\nolimits_{j\in\mathcal{M}} \|\mathbf{z}_{j|j}\|^2 = 1 \right\}. \tag{56}$$

The exact projection onto this set would still require the global denominator in (55). We therefore use the approximate local projection

$$\tilde{\Pi}_{\mathcal{C}_i}^{(\tau)}(\mathbf{z}_{i|i}) = \frac{\mathbf{z}_{i|i}}{\sqrt{M\phi_{i|i}^{(\tau)}}}, \tag{57}$$

where

$$\phi_{i|i}^{(\tau)} \approx \frac{1}{M}\sum_{j\in\mathcal{M}} \|\mathbf{z}_{j|j}^{(\tau)}\|^2 \tag{58}$$

is a local estimate of the average squared norm of the consensus variables. The estimate $\phi_{i|i}^{(\tau)}$ is obtained by adaptive distributed averaging [63]. Define

$$\boldsymbol{\eta}_i^{(\tau)} = \left[\{\|\mathbf{z}_{i|j}^{(\tau)}\|^2\}_{j\in\mathcal{T}_i}\right]^\top, \boldsymbol{\phi}_i^{(\tau)} = \left[\{\phi_{i|j}^{(\tau)}\}_{j\in\mathcal{T}_i}\right]^\top. \tag{59}$$

The copy equality $\phi_{i|j}^{(\tau)} = \phi_{j|j}^{(\tau)}$ denotes transmission of the local squared-norm estimate from node j to node i. The adaptive averaging update is

$$\phi_{i|i}^{(\tau)} = \mathbf{d}^\top\left(\zeta_i^{(\tau)}\boldsymbol{\eta}_i^{(\tau)} + \left(1-\zeta_i^{(\tau)}\right)\boldsymbol{\phi}_i^{(\tau-1)}\right), \tag{60}$$

where $\mathbf{d}$ contains the distributed averaging weights, computed from the network topology, and

$$\zeta_i^{(\tau)} = \frac{\|\boldsymbol{\eta}_i^{(\tau)} - \boldsymbol{\eta}_i^{(\tau-1)}\|}{\|\boldsymbol{\eta}_i^{(\tau)}\| + \|\boldsymbol{\eta}_i^{(\tau-1)}\|}. \tag{61}$$

The value $0 \le \zeta_i^{(\tau)} \le 1$ follows from the triangle inequality. If the channel estimates converge, then the squared norms also converge, so that $\zeta_i^{(\tau)} \to 0$, and the distributed averaging estimate approaches the desired network-wide average. Consequently,

$$\lim_{\tau\to\infty}\sqrt{M\phi_{i|i}^{(\tau)}} = 1, \tag{62}$$

assuming convergence of the response estimates. This yields an adaptive distributed implementation of the global non-triviality constraint using only neighbourhood communication.

C. Distributed protocol

The update equations (48) define the complete per-frame distributed protocol. At frame (τ), each node $i \in \mathcal{M}$ performs the following synchronously:

1) receive the microphone signal-frames for $\mathcal{R}_i$, construct $\mathbf{X}_{i,\mathrm{loc}}^{(\tau)}$, and update either $\mathbf{C}_{\mathrm{x},i}^{(\tau)}$ or $\mathbf{C}_{\mathrm{v},i}^{(\tau)}$;
2) update $\mathbf{w}_i^{(\tau+1)}$ using (48a);
3) form $\mathbf{y}_{i|j}^{(\tau+1)}$ for all $j \in \mathcal{R}_i$ using (48b) and transmit it to node j;
4) receive $\{\mathbf{y}_{i|j}^{(\tau+1)}\}_{j\in\mathcal{T}_i}$, average them, and project onto $\mathcal{C}_i$ to obtain $\mathbf{z}_{i|i}^{(\tau+1)}$;
5) transmit $\mathbf{z}_{i|i}^{(\tau+1)}$ and $\phi_{i|i}$ to nodes holding copies of i;
6) update $\mathbf{u}_{i|j}$ using (48f);

Thus, each node exchanges microphone-signal frames, intermediate ADMM variables $\mathbf{y}_{i|j}$, consensus variables $\mathbf{z}_{i|i}$, and the scalar norm estimates $\phi_{i|i}$ required for the distributed non-triviality constraint.

V. Theoretical Analysis

A. Convergence Behaviour

We briefly discuss the convergence behaviour of the proposed algorithm. A rigorous global convergence proof is outside the scope of this paper, since the proposed method combines non-convex local objectives, a non-convex normalization constraint, online covariance estimates, and inexact quasi-Newton local updates. Nevertheless, the structure of the objective and existing results for non-convex and inexact ADMM provide useful insight.

Consider first a static acoustic system, i.e., fixed channel responses. Each local data term $f_{i,\mathrm{GEV}}(\mathbf{w}_i)$ is a generalized Rayleigh quotient, while $g_i(\mathbf{w}_i)$ is a quadratic regularizer. Generalized Rayleigh quotient minimization is well studied [64]–[69]. Although the quotient is non-convex, its restriction to the hypersphere $\|\mathbf{w}_i\| = c$, $c > 0$, admits exactly two local minima, which are both global [64], [65]. In the BSI setting, these two minimizers are equivalent because the channel responses are identifiable only up to an arbitrary scalar factor.

Adding the regularizer changes this geometry. In general, the sum of a generalized Rayleigh quotient and a quadratic term can introduce local minima that are not global [70]. In the proposed method, the regularizer is chosen to encode prior information on plausible room impulse responses. If this prior is representative of the actual channel responses, the regularisation improves conditioning when the CR identifiability conditions are only approximately satisfied, for example due to insufficient excitation or common or near-common zeros. The weak curvature of the generalized Rayleigh quotient may otherwise lead to slow convergence with first-order methods, which motivates the BFGS quasi-Newton step. While classical convergence results for BFGS mainly concern static convex problems, quasi-Newton updates have been shown to be effective for online acoustic BSI in practice [27].

Convergence of ADMM is well understood for convex objectives [48], [71], but those results do not apply directly here because both the local objective functions and the normalization constraint are non-convex. The exact offline problem is nevertheless structurally related to the setting considered in [61, Theorem 1]: the normalization constraint can be placed in the non-smooth constrained block, while the remaining variables are coupled through linear consensus constraints and smooth quadratic penalties. Moreover, since the normalization excludes the origin, the generalized Rayleigh-type objective is evaluated on a domain where it is continuous. Under the corresponding assumptions and for sufficiently large penalty parameter ρ, the ADMM sequence has limit points, and each limit point is a stationary point of the augmented Lagrangian. Thus, the proposed method should be interpreted as a local optimization method rather than as a method with a general global optimality guarantee. The implemented algorithm differs from exact ADMM because the local primal subproblem is not solved exactly, but approximated by a BFGS-type quasi-Newton step. Therefore, stability depends on choosing the penalty parameter ρ, step size μ, and BFGS metric update such that the local model remains well conditioned and the primal-solve errors remain controlled. In this sense, the proposed

TABLE I
DOMINANT COMPUTATIONAL COMPLEXITY AND TRANSMITTED QUANTITIES AT NODE i AND FRAME (τ). B DENOTES THE NUMBER OF BITS PER COMPLEX NUMBER.

	$\mathbf{x}_i^{(\tau)}$	FFT	$\mathbf{X}_{i,\mathrm{loc}}^{(\tau)}$	$\mathbf{C}_{\mathrm{x},i}^{(\tau)}, \mathbf{C}_{\mathrm{v},i}^{(\tau)}$	$\mathbf{g}_i^{(\tau)}$	$\mathbf{w}_i^{(\tau+1)}$	$\mathbf{V}_i^{(\tau+1)}$	$\mathbf{y}_{i\|j}^{(\tau+1)}$	$\mathbf{y}_{j\|i}^{(\tau+1)}$	$\mathbf{z}_{i\|i}^{(\tau+1)}$	$\mathbf{z}_{j\|i}^{(\tau+1)}$	$\mathbf{u}_{i\|j}^{(\tau+1)}$	$\boldsymbol{\eta}_i^{(\tau)}$	$\gamma_i^{(\tau)}$
Add.	–	$L\log_2 L$	–	–	$4R_i^2L^2+3R_iL$	$R_i^2L^2+R_iL$	$5R_i^2L^2+2R_iL$	R_iL	–	T_iL	–	$2R_iL$	T_iL	L
Mult.	–	$\frac{L}{2}\log_2 L$	$2R_iL^3$	$R_i^3L^3$	$3R_i^2L^2+2R_iL$	$R_i^2L^2$	$6R_i^2L^2+2R_iL$	R_iL	–	T_iL+2L	–	R_iL	–	L
Transm.	$(T_i-1)LB$	–	–	–	–	–	–	–	$(R_i-1)LB$	–	$(T_i-1)LB$	–	T_i-1	–

local update can be viewed as an inexact ADMM step, for which stability is tied to bounded inexactness and a stable curvature approximation rather than exact minimization of each local subproblem [72], [73]. For time-varying systems, the algorithm should be interpreted as an adaptive tracking method.

If the acoustic system changes slowly relative to the adaptation rate, the estimate from the previous frame provides a suitable initialization for the next frame, as in standard adaptive-filtering arguments [74]. The recursive averaging of the CR matrices and the consensus constraints further smooth the updates. In the numerical simulations, a sufficiently large penalty parameter ρ consistently led to stable behaviour. Increasing ρ enforces consensus more strongly, but can reduce convergence speed. We observed that $\rho \geq 1$ leads to stable behaviour in the reported experiments.

B. Computational Complexity

The distributed formulation reduces the dimension of the optimization problem solved at each node. Instead of optimizing over the full channel vector $\mathbf{w} \in \mathbb{C}^{ML}$, node i optimizes over $\mathbf{w}_i \in \mathbb{C}^{R_iL}$, where R_i is the size of its receive-neighbourhood. Hence, for topologies with bounded neighbourhood size, the per-node computational complexity is independent of the total number of nodes M, and all local subproblems can be solved in parallel. The largest per-node reduction relative to a centralized formulation is obtained for sparse topologies, such as the ring topology in Fig. 2, where $R_i = 2$.

Table I lists the dominant arithmetic operations and transmitted quantities per node and frame. In matrix form, the construction and update of the local CR matrices scale with L^3. However, the implementation need not explicitly form these matrices. Instead, the pairwise CR errors and the corresponding gradients can be computed directly using fast Fourier transform (FFT)-based convolution operations, reducing the dominant convolution cost to $R_iL\log L$. The remaining dominant local costs are the matrix-vector operations in the gradient evaluation and the BFGS inverse-Hessian update, both scaling with $R_i^2L^2$.

VI. NUMERICAL SIMULATIONS

Let us define the error metrics, comparing the estimated channel responses $\mathbf{w}_{i|i}^{(\tau)}$ to the ground-truth channel responses $\mathbf{h}_i^{(\tau)}$, which we consider both time-varying. The normalized projection misalignment [75] for two vectors $\mathbf{a}, \mathbf{b}$ of equal length is defined as

$$\mathrm{npm}(\mathbf{a},\mathbf{b}) = 20\,\log_{10}\frac{\|\epsilon(\mathbf{a},\mathbf{b})\|}{\|\mathbf{b}\|}, \tag{63}$$

with

$$\epsilon(\mathbf{a},\mathbf{b}) = \mathbf{b} - \frac{\mathbf{b}^{\mathrm{H}}\mathbf{a}}{\mathbf{a}^{\mathrm{H}}\mathbf{a}}\mathbf{a}. \tag{64}$$

We define the global normalized projection misalignment (NPM) for frame τ as

$$\mathrm{NPM}^{(\tau)} = \mathrm{npm}(\hat{\mathbf{h}}^{(\tau)}, \mathbf{h}^{(\tau)}), \tag{65}$$

where $\hat{\mathbf{h}}^{(\tau)}$ is the stacked vector of the estimated channel responses $\mathbf{w}_{i|i}^{(\tau)}$ for all $i \in \mathcal{M}$ and $\mathbf{h}^{(\tau)}$ is the stacked vector of the ground-truth channel responses $\mathbf{h}_i^{(\tau)}$. Additionally, we define the NPM for each channel i and frame (τ) as

$$\mathrm{NPM}_i^{(\tau)} = \mathrm{npm}(\mathbf{w}_{i|i}^{(\tau)}, \mathbf{h}_i^{(\tau)}). \tag{66}$$

Further, we also use microphone-averaged NPM, defined as

$$\mathrm{NPM}_{\mathrm{avg}}^{(\tau)} = \frac{1}{M}\sum_{i\in\mathcal{M}} \mathrm{NPM}_i^{(\tau)}, \tag{67}$$

which when compared with (65) is invariant to the relative scaling between the channels.

We define the signal-to-noise ratio (SNR) in the acoustic scenarios as the mean SNR at the microphones, given by

$$\mathrm{SNR}_{\mathrm{dB}} = \frac{1}{M}\sum_{i\in\mathcal{M}} 10\log_{10}\frac{\sigma_{\mathrm{y},i}^2 t_{\mathrm{y},i}}{\sigma_{\mathrm{v},i}^2 t_{\mathrm{x}}}, \tag{68}$$

where $\sigma_{\mathrm{y},i}^2$ is the variance of the source signal component in the ith microphone signal during signal-plus-noise periods and $\sigma_{\mathrm{v},i}^2$ is the noise variance at microphone i, respectively. The scalars t_{x} and $t_{\mathrm{y},i}$ are the total and cumulative signal length of signal-plus-noise periods, respectively.

For the following simulation studies, we generate random positive definite noise covariance matrices by multiplying a random matrix where each entry is drawn from a zero-mean Normal distribution with its transpose and scaling it to fulfil (68) for a given target SNR. To generate the noise signal, we generate random sequences of white Gaussian noise and filter it with the Cholesky decomposition of the noise covariance matrix.

We consider two types of input signals: (a) white Gaussian noise (WGN) signal with speech-like non-stationarity in amplitude and signal pauses (b) speech recordings from the LibriSpeech dataset [76]. Since the recordings in the dataset are of varying lengths, we choose a random set of recordings from the same speaker and concatenate them until reaching the desired length. The parameters such as step sizes, regularisation parameters among others are listed in Table II. For each scenario they have been tuned such that convergence speed is approximately the same whenever possible, while the baseline speed is the fastest rate of convergence of the proposed algorithm in the specific scenario.

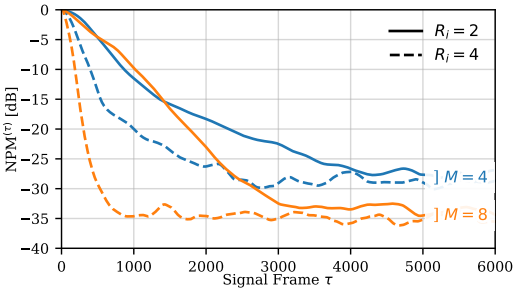

Fig. 3. NPM over time for the proposed algorithm with different WASN topology properties. Convergence speed increases and steady-state error decreases with a larger number of nodes M; convergence speed increases without affecting the steady-state error with a larger number of neighbours R_i. Shown is the median of 20 Monte-Carlo runs. SNR is 20 dB.

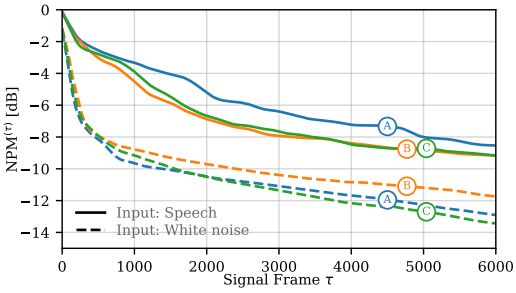

Fig. 4. NPM over time for the proposed algorithm with regularisation parameter $\lambda_0 = 0$ (A), $\lambda_0 = 10^{-4}$ (B) and $\lambda_0 = 10^{-5}$ (C). Show is the median of 20 simulation runs for the SNR in the range [10, 40] dB. Regularization improves the estimation performance when the input signal is non-white.

A. Topology

We evaluate the effect of the WASN's topology density on the algorithm's convergence speed and estimation performance in a simplified scenario. The number of nodes is chosen as $M \in \{4, 8\}$, and we vary the number of neighbours such that $\mathcal{T}_i = R_i \in \{2, 4\}$. We simulate 20 simulation runs of random RIRs drawn from a normal distribution with speech-like WGN input. Fig. 3 shows that the convergence speed increases significantly with a larger network size M and neighbourhood $\mathcal{R}_i$. However, the estimation performance in terms of steady-state NPM is only significantly affected by network size M, while only marginally by the neighbourhood size.

B. Regularization

We show the effect of the regularisation term, which is described in Sec. IV. As a prior, we estimate the covariance matrix $\mathbf{\Sigma_h}$ of the room channel responses in the DFT domain from 20000 randomly generated rooms. For each simulation run, we randomly sample each of the room's dimensions from a uniform distribution on the interval $[2, 6]$ m. The absorption coefficients of the walls are sampled from a uniform distribution on the interval of $[0.05, 0.5]$. The single source and $M = 4$ sensor nodes are placed at random locations within the room with a minimum distance of 0.1 m from the walls, again uniformly sampled. Then, 20 simulation runs are generated with speech-like WGN and speech input. We compare the proposed algorithm with regularisation parameter $\lambda_0 = 0$, $\lambda_0 = 10^{-4}$, and $\lambda_0 = 10^{-5}$ and show the median of all runs for the SNR in the range [10, 40] dB. The simulation results plotted in Fig. 4 show that introducing prior information improves the estimation performance when the input signal is non-white by increasing convergence speed and reducing the steady-state error slightly. Unsurprisingly, the regularisation term does not affect the estimation performance when the input signal is white.

C. Evaluation in randomized simulated environment

As our first simulation study, we generate acoustic systems using the *pyroomacoustics* library [77]. For each simulation run, we generate a random room with each of its dimensions sampled from a uniform distribution on the interval $[2, 6]$ m. The absorption coefficients of the walls are sampled from a uniform distribution on the interval of $[0.05, 0.5]$. The single source and M sensor nodes are placed at random locations within the room with a minimum distance of 0.1 m from the walls, again uniformly sampled. We construct the WASN topology by generating a directed ring topology minimizing the overall distance between connected nodes and subsequently cycling through nodes to add further edges to the nearest node that has the smallest R_i. This ensures that the graph is connected, i.e., there is a path from each node i to each node j. In the case where $R_i = 2$, the smallest sub-problem size possible, the resulting topology is a ring topology; see Fig. 2 (right) for an example. We generate room impulse responses using the randomized image source method [78] contained in the "pyroomacoustics" library and shift the impulse responses such that the first non-zero impulse response sample with the earliest arrival time is at 0. As our ground truth and estimates, we use a 128-bin DFT of the impulse responses, appropriately downsampled and truncated. While this significantly simplifies the real-world scenario, it allows us to evaluate the algorithm's performance in a controlled environment while keeping simulations tractable.

We compare the proposed algorithm to various existing BSI algorithms:

- (MCQN) The centralized MCQN algorithm for the M channels problem as described in [27].
- (S-MCQN) The MCQN algorithm applied to the sub-problems independently.
- (DGD) An adaptive implementation of the distributed (stochastic) distributed gradient descent (DGD) as described in [42].
- (DSAAWET) The distributed stochastic approximation algorithm with expanding truncations (DSAAWET) as described in [44], [45].
- (DABSI) Our proposed algorithm with local BFGS update step on GEVD local cost functions and regularisation.

The MCQN algorithm is chosen as a representative of centralized algorithms, as it has been the best performing in the given scenario. We simulate 20 simulation runs of random

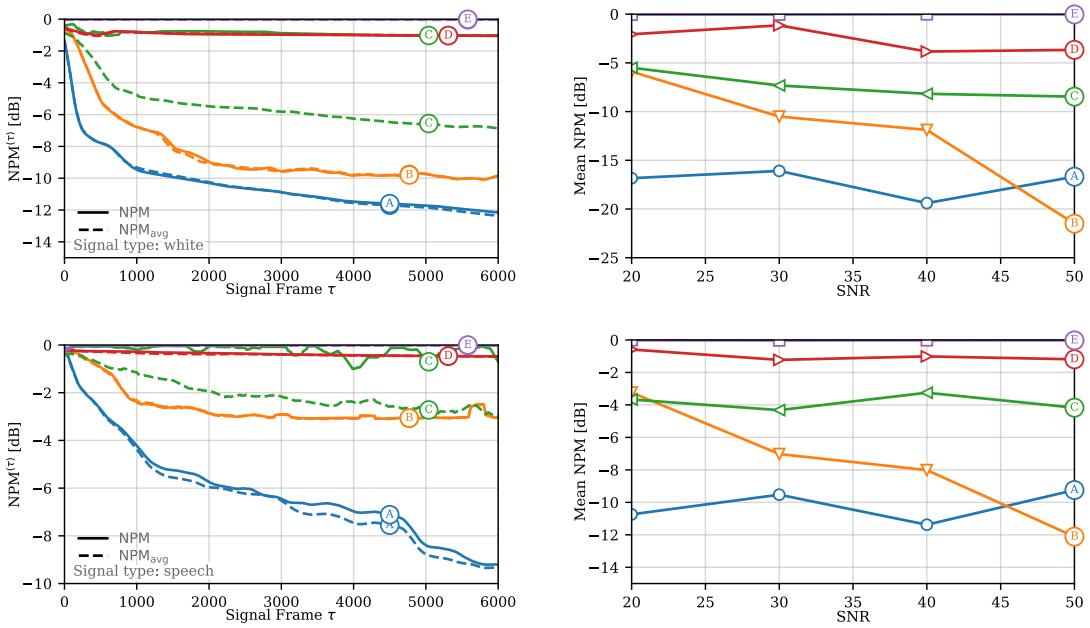

Fig. 5. Proposed DABSI algorithm (A) compared to MCQN (B), S-MCQN (C), DGD (D), and DSAAWET (E). Shown is the median of 20 simulation runs with randomly generated rooms and WGN and input. **(Top row)** WGN input signal. **(Bottom row)** speech input signal. **(Left)** Convergence over time for SNR 20 dB. Full lines indicate global NPM (65), dashed lines indicate microphone-averaged NPM (67). **(Right)** Global NPM after convergence for different SNR levels.

rooms as described above with both speech-like WGN input and speech input. The scenario is simulated at a set of SNR levels $\{10, 20, 30, 40\}$ dB. Fig. 5 shows the error over time for the different algorithms and input types as well as the average NPM after convergence. It can be observed that the proposed algorithm outperforms the compared algorithms in this scenario, especially in the case of speech input. Improved convergence speed and estimation accuracy can be observed. To note is the difference in estimation performance of the S-MCQN algorithm in $\mathrm{NPM}^{(\tau)}$ and $\mathrm{NPM}_{\mathrm{avg}}^{(\tau)}$ coming from the lost relative scaling between channels. The difference in performance between the MCQN and S-MCQN algorithms in both convergence speed and steady-state error is due to the lost cross-channel information by solving the sub-problems with fewer channels independently. This underlines the importance of the consensus step in the distributed algorithm: Even though the sub-problems are solved independently, it ensures that cross-channel information that is lost by sub-problem independence is regained and improved upon. The convergence behaviour of the DGD algorithm is slower than the other algorithms while the DSAAWET algorithm does not converge in the given scenario. Further, it can be observed that the proposed algorithm is less sensitive to the SNR level than the next best performing algorithm, the MCQN algorithm. In turn, however, for high SNR levels, the MCQN algorithm outperforms the proposed.

D. Evaluation in measured environment

In addition to the simulated environments described in Sec. VI-C, we also consider a set of measured impulse responses from the MYRiAD database [79], [80]. We select random combinations of impulse responses from the dataset to construct the scenario. The source position is randomly selected from the given set of loudspeakers. The impulse responses are resampled to a sampling rate of 8 kHz and truncated at 128 samples, with removal of leading zeros, before convolving with the source signal, which is a speech-like WGN on the one hand and speech input on the other.

Again, we simulate 50 Monte Carlo runs of the constructed scenario using the measured impulse responses. The proposed DABSI algorithm converges faster and to a lower NPM than the other algorithms, as shown in Fig. 6. While overall the estimation accuracy is worse, the observed behaviour is consistent with the simulated environment. In the case of the speech input, the proposed DABSI algorithm and the centralized MCQN algorithm perform similarly, while the other algorithms perform worse.

VII. Conclusions

The proposed distributed algorithm for blind acoustic system identification is a step towards a scalable solution for large sensor networks. The algorithm is based on the cross-relation method, which is well-established, and set into a distributed context using the alternating direction method of multipliers. For improved estimation performance, we extend the sub-problems with a VAD-based approach and a regularisation term based on prior knowledge of general room channel responses. Moreover, we use a quasi-Newton method to solve the sub-problems which improves the convergence speed and estimation accuracy. The algorithm is evaluated on both synthetic and real-world data which shows good performance when compared to other BSI algorithms. Further, the effectiveness of the consensus approach is verified by comparing the results to a centralized algorithm applied to

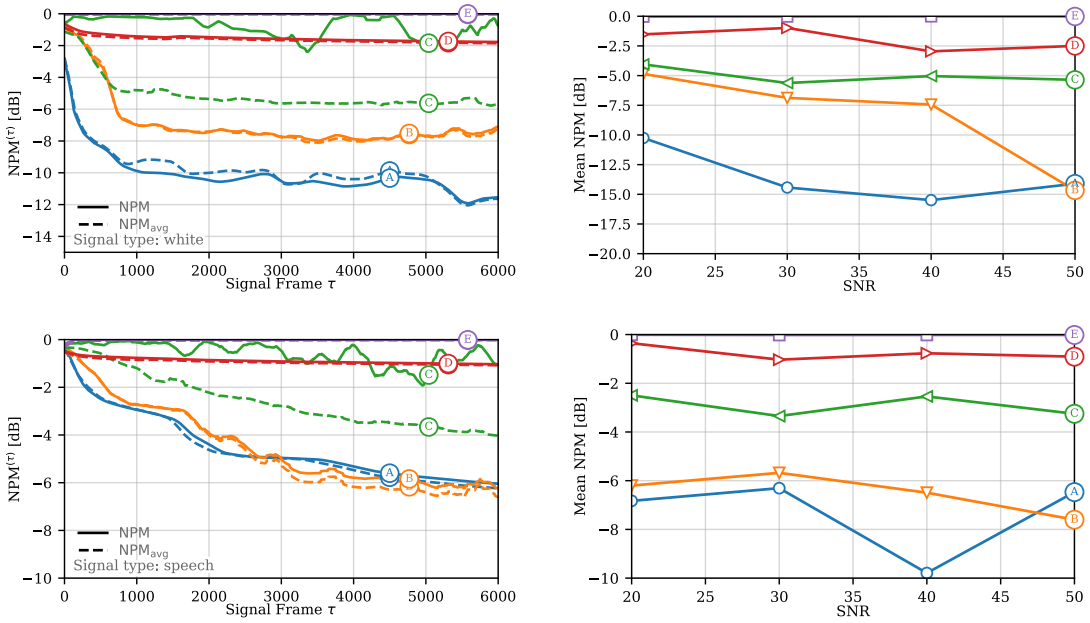

Fig. 6. Proposed DABSI algorithm (A) compared to MCQN (B), S-MCQN (C), DGD (D), and DSAAWET (E). Shown is the median of 20 simulation runs with randomly selected impulse responses from the MYRiAD database and WGN input. **(Top row)** WGN input signal. **(Bottom row)** speech input signal. **(Left)** Convergence over time for SNR 20 dB. Full lines indicate global NPM (65), dashed lines indicate microphone-averaged NPM (67). **(Right)** Global NPM after convergence for different SNR levels.

TABLE II
SIMULATION PARAMETERS FOR DABSI AND COMPARED METHODS. THE FIRST FOUR PARAMETERS ARE USED FOR DABSI; THE REMAINING PARAMETERS ARE STEP SIZES FOR THE COMPARED METHODS, TUNED FOR BEST PERFORMANCE. FOR DSAAWET, NO SUITABLE STEP SIZE FOR CONVERGENCE WAS FOUND.

Parameter	**WGN RIRs**	**Simulated/Measured RIRs**
ρ	1.0	1.0
μ	0.4	0.4
η	0.98	0.98
λ_0	10^{-5}	10^{-5}
μ_{MCQN}	0.2	0.02
$\mu_{\text{S-MCQN}}$	0.1	0.01
α_{DGD}	10^{-5}	10^{-6}
α_{DSAAWET}	n/a	

separate sub-problems. The algorithm is able to effectively estimate the channel responses in an adaptive way while avoiding the need for a centralized processing unit.

There are some limitations to the algorithm, such as the relatively high computational complexity and the assumption of a known length of the impulse responses. In simulated scenarios, the latter is dealt with by truncating impulse responses used to generate the data, however, in real-world applications, the length of the impulse responses is unknown. In sensor networks, the direct path between source and sensors differs for each sensor, which leads to different overall initial delays in impulse responses and therefore subproblems. This requires an alignment operation in the consensus step. Moreover, the computational complexity of the algorithm is relatively high, due to matrix-matrix multiplications. For applicability in real-world scenarios, i.e., a real-time system, this needs to be further investigated and optimized. Further, we made the assumption that there is a VAD at each node, yielding perfect data. In practice, this is not the case and the algorithm should be extended to be robust to inaccurate VAD decisions.

REFERENCES

[1] M. Maroti, G. Simon, A. Ledeczi, and J. Sztipanovits, "Shooter localization in urban terrain," *Computer*, vol. 37, no. 8, pp. 60–61, 2004.
[2] W.-P. Chen, J. Hou, and L. Sha, "Dynamic clustering for acoustic target tracking in wireless sensor networks," *IEEE Trans. Mob. Comput.*, vol. 3, no. 3, pp. 258–271, 2004.
[3] A. Bertrand and M. Moonen, "Distributed Adaptive Node-Specific Signal Estimation in Fully Connected Sensor Networks—Part I: Sequential Node Updating," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5277–5291, Oct. 2010.
[4] ——, "Distributed Adaptive Node-Specific Signal Estimation in Fully Connected Sensor Networks—Part II: Simultaneous and Asynchronous Node Updating," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5292–5306, Oct. 2010.
[5] M. Ferrer, M. de Diego, G. Piñero, and A. Gonzalez, "Affine projection algorithm over acoustic sensor networks for active noise control," *IEEE Trans. Audio Speech Lang. Process.*, vol. 29, pp. 448–461, 2021.
[6] S. Ruiz, T. van Waterschoot, and M. Moonen, "Distributed combined acoustic echo cancellation and noise reduction using GEVD-based distributed adaptive node specific signal estimation with prior knowledge," in *Proc. 28th European Signal Process. Conf. (EUSIPCO '20)*. Amsterdam, Netherlands: IEEE, Jan. 2021, pp. 206–210. [Online]. Available: https://ieeexplore.ieee.org/document/9287337/
[7] Y. Zeng and R. C. Hendriks, "Distributed Delay and Sum Beamformer for Speech Enhancement via Randomized Gossip," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 1, pp. 260–273, Jan. 2014. [Online]. Available: http://ieeexplore.ieee.org/document/6663655/
[8] J. S. Bradley, "Predictors of speech intelligibility in rooms," *J. Acoust. Soc. Am.*, vol. 80, no. 3, pp. 837–845, Sep. 1986. [Online]. Available: https://doi.org/10.1121/1.393907
[9] Y. E. Baba, A. Walther, and E. A. P. Habets, "3D Room Geometry Inference Based on Room Impulse Response Stacks," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 5, pp. 857–872, May 2018.

[10] K. MacWilliam, F. Elvander, and T. van Waterschoot, "Simultaneous Acoustic Echo Sorting and 3-D Room Geometry Inference," in *Proc. 2023 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '23)*. Rhodes Island, Greece: IEEE, 2023, pp. 1–5.

[11] Y. Sato, "A Method of Self-Recovering Equalization for Multilevel Amplitude-Modulation Systems," *IEEE Trans. Commun.*, vol. 23, no. 6, pp. 679–682, Jun. 1975. [Online]. Available: http://ieeexplore.ieee.org/document/1092854/

[12] D. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980. [Online]. Available: http://ieeexplore.ieee.org/document/1094608/

[13] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Conf. Rec. 25th Asilomar Conf. Signals, Syst. Computers*. Pacific Grove, CA, USA: IEEE Comput. Soc. Press, 1991, pp. 856–860. [Online]. Available: http://ieeexplore.ieee.org/document/186568/

[14] J. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991. [Online]. Available: http://ieeexplore.ieee.org/document/75086/

[15] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994. [Online]. Available: http://ieeexplore.ieee.org/document/312157/

[16] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995. [Online]. Available: http://ieeexplore.ieee.org/document/476442/

[17] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995. [Online]. Available: http://ieeexplore.ieee.org/document/348133/

[18] S. Gannot and M. Moonen, "Subspace Methods for Multimicrophone Speech Dereverberation," *EURASIP J. Adv. Signal Process.*, vol. 2003, no. 11, p. 769285, Dec. 2003. [Online]. Available: https://asp-eurasipjournals.springeropen.com/articles/10.1155/S1110865703305049

[19] K. I. Diamantaras and T. Papadimitriou, "An Efficient Subspace Method for the Blind Identification of Multichannel FIR Systems," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5833–5839, Dec. 2008.

[20] Q. Mayyala, K. Abed-Meraim, and A. Zerguine, "Structure-Based Subspace Method for Multichannel Blind System Identification," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1183–1187, Aug. 2017.

[21] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Process.*, vol. 44, no. 3, pp. 661–672, Mar. 1996.

[22] Y. A. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," *Signal Processing*, p. 12, 2002.

[23] ——, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[24] M. Hu, S. Doclo, D. Sharma, M. Brookes, and P. A. Naylor, "Noise robust blind system identification algorithms based on a Rayleigh quotient cost function," in *Proc. 23th European Signal Process. Conf. (EUSIPCO '15)*, Aug. 2015, pp. 2476–2480.

[25] H. He, J. Chen, J. Benesty, and T. Yang, "Noise Robust Frequency-Domain Adaptive Blind Multichannel Identification With ℓ_p-Norm Constraint," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 9, pp. 1608–1619, Sep. 2018.

[26] B. Jo and P. Calamia, "Robust Blind Multichannel Identification based on a Phase Constraint and Different ℓp-norm Constraints," in *Proc. 28th European Signal Process. Conf. (EUSIPCO '20)*, Jan. 2021, pp. 1966–1970.

[27] E. A. Habets and P. A. Naylor, "An online quasi-Newton algorithm for blind SIMO identification," in *Proc. 2010 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '10)*. Dallas, TX: IEEE, Mar. 2010, pp. 2662–2665. [Online]. Available: http://ieeexplore.ieee.org/document/5496248/

[28] M. Hu, "Cross-relation based blind identification of acoustic SIMO systems and applications," Ph.D. dissertation, Imperial College London, 2017.

[29] S. Kitić and J. Daniel, "Blind identification of ambisonic reduced room impulse response," *IEEE Trans. Audio Speech Lang. Process.*, vol. 32, pp. 443–458, 2024.

[30] D. Schmid and G. Enzner, "Cross-Relation-Based Blind SIMO Identifiability in the Presence of Near-Common Zeros and Noise," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 60–72, Jan. 2012.

[31] G. Enzner and P. Thüne, "On the statistics and the detection of multichannel common zeros," in *Proc. 2014 Int. Workshop Acoustic Signal Enhancement (IWAENC '14)*, Sep. 2014, pp. 21–25.

[32] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Processing*, vol. 86, no. 6, pp. 1140–1156, Jun. 2006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0165168405003178

[33] S. Malik, D. Schmid, and G. Enzner, "A State-Space Cross-Relation Approach to Adaptive Blind SIMO System Identification," *IEEE Signal Process. Lett.*, vol. 19, no. 8, pp. 511–514, Aug. 2012.

[34] S. Malik and G. Enzner, "Recursive Bayesian Control of Multichannel Acoustic Echo Cancellation," *IEEE Signal Process. Lett.*, vol. 18, no. 11, pp. 619–622, Nov. 2011.

[35] D. Schmid, S. Malik, and G. Enzner, "A Maximum A Posteriori Approach to Multichannel Speech Dereverberation and Denoising," in *Proc. 2012 Int. Workshop Acoustic Signal Enhancement (IWAENC '12)*, Sep. 2012, pp. 1–4.

[36] S. Malik, "Bayesian learning of linear and nonlinear acoustic system models in hands-free communication," Ph.D. dissertation, Ruhr-Universität Bochum, Bochum, 2012.

[37] D. Schmid, S. Malik, and G. Enzner, "An expectation-maximization algorithm for multichannel adaptive speech dereverberation in the frequency-domain," in *Proc. 2012 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '12)*, Mar. 2012, pp. 17–20.

[38] S. Malik and G. Enzner, "A Variational Bayesian Learning Approach for Nonlinear Acoustic Echo Control," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5853–5867, Dec. 2013.

[39] D. Schmid, G. Enzner, S. Malik, D. Kolossa, and R. Martin, "Variational Bayesian Inference for Multichannel Dereverberation and Noise Reduction," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 8, pp. 1320–1335, Aug. 2014.

[40] P. Thuene and G. Enzner, "Maximum-Likelihood and Maximum-A-Posteriori Perspectives for Blind Channel Identification on Acoustic Sensor Network Data," in *Speech Communication; 13th ITG-Symposium*, Oct. 2018, pp. 1–5.

[41] T. van Waterschoot and M. Moonen, "Distributed estimation and equalization of room acoustics in a wireless acoustic sensor network," in *Proc. 20th European Signal Process. Conf. (EUSIPCO '12)*, Aug. 2012, pp. 2709–2713.

[42] C. Yu, L. Xie, and Y. C. Soh, "Distributed blind system identification in sensor networks," in *Proc. 2014 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '14)*, May 2014, pp. 5065–5069.

[43] Y. Liu, H. Liu, and C. Li, "Distributed blind identification of sparse channels in sensor networks," in *Proc. 35th Chinese Control Conf.(CCC '16)*, Jul. 2016, pp. 5122–5127.

[44] R. Liu and H.-F. Chen, "Distributed and recursive blind channel identification to sensor networks," *Control Theory Technol.*, vol. 15, no. 4, pp. 274–287, Nov. 2017. [Online]. Available: http://link.springer.com/10.1007/s11768-017-7086-x

[45] J. Lei and H.-F. Chen, "Distributed Stochastic Approximation Algorithm With Expanding Truncations," *IEEE Trans. Autom. Control*, vol. 65, no. 2, pp. 664–679, Feb. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8695094/

[46] M. Blochberger, F. Elvander, R. Ali, M. Moonen, T. van Waterschoot, J. Østergaard, and J. Jensen, "Distributed Cross-Relation-Based Frequency-Domain Blind System Identification Using Online-ADMM," in *Proc. 2022 Int. Workshop Acoustic Signal Enhancement (IWAENC '22)*. Bamberg, Germany: IEEE, Sep. 2022, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/9914714/

[47] M. Blochberger, F. Elvander, R. Ali, J. Østergaard, J. Jensen, M. Moonen, and T. van Waterschoot, "Distributed Adaptive Norm Estimation for Blind System Identification in Wireless Sensor Networks," in *Proc. 2023 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '23)*. Rhodes Island, Greece: IEEE, Jun. 2023, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/10096574/

[48] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011. [Online]. Available: http://dx.doi.org/10.1561/2200000016

[49] H. Wang and A. Banerjee, "Online alternating direction method," in *Proc. 29th Int. Conf. Machine Learning (ICML '12)*, 2012, pp. 1119–1126.

[50] S. Hosseini, A. Chapman, and M. Mesbahi, "Online distributed ADMM via dual averaging," in *Proc. 53rd IEEE Conf. Decision Control (CDC '14)*, Dec. 2014, pp. 904–909.

[51] ——, "Online Distributed Convex Optimization on Dynamic Networks," *IEEE Trans. Autom. Control*, vol. 61, no. 11, pp. 3545–3550, Nov. 2016.

[52] K. W. Brodlie, "An assessment of two approaches to variable metric methods," *Mathematical Programming*, vol. 12, no. 1, pp. 344–355, Dec. 1977. [Online]. Available: http://link.springer.com/10.1007/BF01593802

[53] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed., ser. Springer Series in Operations Research. New York: Springer, 2006.

[54] A. Hassani, A. Bertrand, and M. Moonen, "Distributed GEVD-based signal subspace estimation in a fully-connected wireless sensor network," in *Proc. 22nd European Signal Process. Conf. (EUSIPCO '14)*, Lisbon, Portugal, 2014, pp. 1292–1296.

[55] ——, "GEVD-Based Low-Rank Approximation for Distributed Adaptive Node-Specific Signal Estimation in Wireless Sensor Networks," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2557–2572, May 2016.

[56] T. van Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Processing*, vol. 88, no. 3, pp. 594–611, Mar. 2008. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0165168407003179

[57] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*, ser. Signals and Communication Technology. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. [Online]. Available: http://link.springer.com/10.1007/978-3-540-37631-6

[58] Y. Hua, "Blind methods of system identification," *Circuits Systems and Signal Process.*, vol. 21, pp. 91–108, 2002.

[59] 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 Trans. Audio Speech Lang. Process.*, vol. 22, no. 4, pp. 785–799, Apr. 2014. [Online]. Available: http://ieeexplore.ieee.org/document/6730918/

[60] R. Varzandeh, M. Taseska, and E. A. P. Habets, "An iterative multichannel subspace-based covariance subtraction method for relative transfer function estimation," in *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*. San Francisco, CA, USA: IEEE, 2017, pp. 11–15. [Online]. Available: http://ieeexplore.ieee.org/document/7895552/

[61] Y. Wang, W. Yin, and J. Zeng, "Global Convergence of ADMM in Nonconvex Nonsmooth Optimization," *J Sci Comput*, vol. 78, no. 1, pp. 29–63, Jan. 2019. [Online]. Available: http://link.springer.com/10.1007/s10915-018-0757-z

[62] S. Hosseini, A. Chapman, and M. Mesbahi, "Online Distributed ADMM on Networks," Oct. 2015. [Online]. Available: http://arxiv.org/abs/1412.7116

[63] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, Sep. 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167691104000398

[64] X.-B. Gao, G. H. Golub, and L.-Z. Liao, "Continuous methods for symmetric generalized eigenvalue problems," *Linear Algebra Its Appl.*, vol. 428, no. 2-3, pp. 676–696, Jan. 2008. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0024379507004016

[65] G. H. Golub, "Eigenvalue computation in the 20th century," *J. Comput. Appl. Math.*, vol. 123, pp. 35–65, Mar. 2000.

[66] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Natl. Bur. Stand. (U. S.)*, vol. 49, pp. 409–435, 1952. [Online]. Available: https://api.semanticscholar.org/CorpusID:2207234

[67] D. K. Faddeev and V. N. Faddeeva, "Computational methods of linear algebra," *Journal of Soviet Mathematics*, vol. 15, pp. 531–650, 1981.

[68] Y. T. Feng and D. R. J. Owen, "Conjugate gradient methods for solving the smallest eigenpair of large symmetric eigenvalue problems," *Int. J. Numer. Meth. Engng.*, vol. 39, no. 13, pp. 2209–2229, Jul. 1996. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0207(19960715)39:13⟨2209::AID-NME951⟩3.0.CO;2-R

[69] P. Arbenz, U. L. Hetmaniuk, R. B. Lehoucq, and R. S. Tuminaro, "A comparison of eigensolvers for large-scale 3D modal analysis using AMG-preconditioned iterative methods," *Int. J. Numer. Meth. Engng*, vol. 64, no. 2, pp. 204–236, Sep. 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/nme.1365

[70] L.-H. Zhang, "On optimizing the sum of the Rayleigh quotient and the generalized Rayleigh quotient on the unit sphere," *Comput Optim Appl*, vol. 54, no. 1, pp. 111–139, Jan. 2013. [Online]. Available: http://link.springer.com/10.1007/s10589-012-9479-6

[71] J. Eckstein and D. P. Bertsekas, "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, Apr. 1992. [Online]. Available: http://link.springer.com/10.1007/BF01581204

[72] J. Eckstein and W. Yao, "Approximate ADMM algorithms derived from Lagrangian splitting," *Comput Optim Appl*, vol. 68, no. 2, pp. 363–405, Nov. 2017. [Online]. Available: http://link.springer.com/10.1007/s10589-017-9911-z

[73] Y. Gu and N. Yamashita, "An alternating direction method of multipliers with the BFGS update for structured convex quadratic optimization," *Comp. Appl. Math.*, vol. 40, no. 3, p. 81, Apr. 2021. [Online]. Available: http://link.springer.com/10.1007/s40314-021-01467-w

[74] S. Haykin, *Adaptive Filter Theory*, 5th ed., ser. Always Learning. Upper Saddle River, NJ Munich: Pearson, 2014.

[75] D. Morgan, J. Benesty, and M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 174–176, Jul. 1998.

[76] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *Proc. 2015 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '15)*, South Brisbane, Australia, Apr. 2015, pp. 5206–5210. [Online]. Available: http://ieeexplore.ieee.org/document/7178964/

[77] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *Proc. 2018 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '18)*, 2018, pp. 351–355.

[78] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE Trans. Audio Speech Lang. Process.*, vol. 23, no. 4, pp. 774–786, 2015.

[79] T. Dietzen, R. Ali, M. Taseska, and T. van Waterschoot, "MYRiAD: A multi-array room acoustic database," *EURASIP J. Audio, Speech, Music Process.*, vol. 2023, no. 1, p. 17, Apr. 2023. [Online]. Available: https://asmp-eurasipjournals.springeropen.com/articles/10.1186/s13636-023-00284-9

[80] ——, "Data repository for MYRiAD: A multi-array room acoustic database," Dec. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.7322754