Relaxed Binaural LCMV Beamforming

Andreas I. Koutrouvelis, Richard C. Hendriks, Richard Heusdens and Jesper Jensen

Abstract—In this paper we propose a new binaural beamforming technique which can be seen as a relaxation of the linearly constrained minimum variance (LCMV) framework. The proposed method can achieve simultaneous noise reduction and exact binaural cue preservation of the target source, similar to the binaural minimum variance distortionless response (BMVDR) method. However, unlike BMVDR, the proposed method is also able to preserve the binaural cues of multiple interferers to a certain predefined accuracy. Specifically, it is able to control the trade-off between noise reduction and binaural cue preservation of the interferers by using a separate trade-off parameter per-interferer. Moreover, we provide a robust way of selecting these trade-off parameters in such a way that the preservation accuracy for the binaural cues of the interferers is always better than the corresponding ones of the BMVDR. The relaxation of the constraints in the proposed method achieves approximate binaural cue preservation of more interferers than other previously presented LCMV-based binaural beamforming methods that use strict equality constraints.

Index Terms—Beamforming, binaural cue preservation, hearing aids, LCMV, multi-microphone noise reduction, MVDR.

I. INTRODUCTION

COMPARED to normal-hearing people, hearing-impaired people generally have more difficulties in understanding a target talker in complex acoustic environments with multiple interfering sources. To reduce noise and improve speech comfort, single-microphone (see e.g. [1] for an overview) or multi-microphone noise reduction methods (see e.g., [2] for an overview) can be used. While the former are mostly effective in reducing listening effort, the latter are also effective in improving speech intelligibility [3]. Examples of multi-microphone noise reduction methods include the multi-channel Wiener filter (MWF) [4], [5], the minimum variance distortionless response (MVDR) beamformer [6], [7], or, its generalization, the linearly constrained minimum variance (LCMV) beamformer [7], [8].

Traditionally, hearing aids (HAs) have been fitted *bilaterally*, i.e., the user wears a HA on each ear, and the HAs are operating essentially independently of each other. As such, the noise reduction algorithm in each HA estimates the signal of interest using only the recordings of the microphones from that specific HA [9]. Such a setup with an independent multi-microphone algorithm per ear may severely distort the binaural cues since phase and magnitude relations of the sources reaching the two ears are modified [10]. This is harmful for the naturalness of the total sound field as received by the hearing-aid user. Ideally, all sound sources (including the undesired ones) that are present after processing should still sound as if originating from the original direction. This does not only lead to a more natural perception of the acoustic environment,

but can also lead to an improved intelligibility of a target talker in certain cases; more specifically, in spatial unmasking experiments [11] it has been shown that a target talker in a noisy background is significantly easier to understand when the noise sources are separated in space from the talker, as compared to the situation where talker and noise sources are co-located.

Binaural HAs are able to wirelessly exchange microphone signals between HAs. This facilitates the use of multi-microphone noise reduction methods which combine all microphone recordings from both HAs, hence allowing the usage of more microphone recordings than with the bilateral noise reduction. As such, the increased number of microphone recordings can potentially lead to better noise suppression and, thus, to a higher speech intelligibility. Moreover, by introducing proper constraints on the beamformer coefficients, binaural cue preservation of the sources can be achieved.

The LCMV method [7], [8] minimizes the output noise power under multiple linear equality constraints. One of these equality constraints is typically used to guarantee that the target source remains undistorted with respect to a certain reference location or microphone. The remaining constraints can be used for additional control on the final filter response. For example, they can be used to steer nulls in the directions of the interferers [7], [12], or to broaden the beam towards the target source in order to avoid steering vector mismatch problems [13], [14]. A special case of the LCMV method is the minimum variance distortionless response (MVDR) beamformer, which only uses the distortionless constraint of the target source [6], [7].

An alternative multi-microphone noise reduction method is the MWF [4], [5] which leads to the minimum mean square error (MMSE) estimate of the target source if the estimator is constrained to be linear, or, the target source and the noise are assumed to be jointly Gaussian distributed [15]. However, in [16]–[18], it was demonstrated that speech signals in time and frequency domains tend to be super-Gaussian distributed rather than Gaussian distributed. Thus, the MWF is generally not MMSE optimal. The MWF does not include a distortionless constraint for the target source and, thus, it generally introduces speech distortion in the output [4]. Several generalizations of the MWF have been proposed, among which the speech distortion weighted MWF (SDW-MWF) [5], which introduces a parameter in the minimization procedure to control the trade-off between speech distortion and noise reduction. A well-known property of the MWF is the fact that it can be decomposed into an MVDR beamformer and a single-channel Wiener filter as a post-processor [19].

There are several binaural multi-microphone noise reduction methods known from the literature. These can be devided into two main categories [20]: a) methods based on the linearly

constrained minimum variance (LCMV) framework and b) methods based on the multi-channel Wiener filter (MWF).

The binaural version of the SDW-MWF (BSDW-MWF) [21], [22] preserves the binaural cues of the target. However, it was theoretically proven that the binaural cues of the interferers collapse on the binaural cues of the target source [23] (i.e., after processing the binaural cues of the interferers become identical to the binaural cues of the target source). In [22], a variation of the BSDW-MWF (called BSDW-MWF-N) was proposed which tries to partially preserve the binaural cues of the interferers. This method inserts a portion of the unprocessed noisy signal at the reference microphones to the coresponding BSDW-MWF enhanced signals. The larger the portion of the unprocessed noisy signals, the lower the noise reduction, but the better the preservation of binaural cues of the interferers and vice versa. As such, this solution exhibits a trade-off between the preservation of binaural cues and the amount of noise reduction. In [24], a subjective evaluation of BSDW-MWF and BSDW-MWF-N shows that for a moderate input SNR indeed the subjects localized the processed interferer correctly with BSDW-MWF-N and incorrectly with BSDW-MWF. However, for a small input SNR the processed interferer was also localized correctly for BSDW-MWF. This is mainly due to the inaccurate estimates of the cross power spectral density (CPSD) matrix of the target, and due to masking effects when the processed target and processed interferer are represented to the subjects simultaneously [24]. In [25], two other variations of the BSDW-MWF were proposed. The first one is capable of preserving the binaural cues of the target and completely cancel one interferer. The second one is capable of accurately preserving the binaural cues of only one interferer, while distorting the binaural cues of the target.

Similarly to SDW-MWF, the BSDW-MWF can be decomposed into the binaural MVDR (BMVDR) beamformer and a single-channel Wiener filter [25]. The BMVDR can preserve the binaural cues of the target source, but the binaural cues of the interferers collapse to the binaural cues of the target source. In [26], [27], the binaural linearly constrained minimum variance (BLCMV) method was proposed, which achieves simultaneous noise reduction and binaural cue preservation of the target source and multiple interferers. Unlike the BMVDR, the BLCMV uses two additional linear constraints per interferer to preserve its binaural cues. A fixed interference rejection parameter is used in combination with these constraints to control the amount of noise reduction. The BLCMV is thus capable of controlling the amount of noise reduction using two constraints per interferer. However, in hearing-aid systems with a rather limited number of microphones, the degrees of freedom (DOF) for noise reduction are exhausted quickly when increasing the number of interferers. This makes the BLCMV less suitable for this application.

In [28], a similar method to BLCMV, called optimal BLCMV (OBLCMV), was proposed which is able to achieve simultaneous noise reduction and binaural cue preservation of the target source and only one interferer. Unlike the BLCMV, the OBLCMV uses an optimal interference rejection parameter with respect to the binaural output SNR. In [29], [30] two independent works proposed the same LCMV-based method (we call it joint BLCMV (JBLCMV)) as an alternative to the BLCMV, which preserves the binaural cues of the target source and *more* than twice the number of interferers compared to the BLCMV [29]. Unlike the BLCMV, the JBLCMV requires only one linear constraint per interferer and, as a result, it has more DOF left for noise reduction. The linear constraints for the preservation of the binaural cues of the interferers have the same form as the linear constraint used in [25]. However, unlike the method in [25], the JBLCMV can preserve the binaural cues of a limited number of interferers and does not distort the binaural cues of the target source.

In this paper, we present an iterative, relaxed binaural LCMV beamforming method. Similar to the other binaural LCMV-based approaches, the proposed method strictly preserves the binaural cues of the target source. However, the proposed method is flexible to control the accuracy of binaural cue preservation of the interferers and, therefore, trade-off against additional noise reduction. This is achieved by using inequality constraints instead of the commonly used equality constraints. The task of each inequality constraint is the (approximate) preservation of the binaural cues of a single interferer in a controlled way. The proposed method is flexible to select a different value for the trade-off parameter of each interferer according to importance. The BMVDR and the JBLCMV can be seen as two extreme cases of the proposed method. On one hand, the BMVDR can achieve the best possible overall noise suppression compared to all the other aforementioned binaural LCMV-based methods, but causes full collapse of the binaural cues of the interferers towards the binaural cues of the target source. On the other hand, the JBLCMV can achieve the preservation of the maximum possible number of interferers compared to the other aforementioned binaural LCMV-based methods, but at the expense of less noise suppression. Unlike the JBLCMV and the BMVDR, the proposed method, is flexible to control the amount of noise suppression and binaural cue preservation according to the needs of the user. The relaxations used in the proposed method allow the usage of a substantially larger number of constraints for the approximate preservation of more interferers compared to all the other binaural LCMV-based methods including JBLCMV.

The remainder of this paper is organized as follows. In Section II, the signal model and the notation are presented. In Section III the key idea of the binaural beamforming is explained and several existing binaural LCMV-based methods are summarized. In Sections IV and V, a novel non-convex binaural beamforming problem and its iterative convex approximation are presented, respectively. In Section VI, the evaluation of the proposed method is provided. Finally, in Section VII, we draw some conclusions.

II. SIGNAL MODEL AND NOTATION

Assume for convenience that each of the two HAs consists of $M/2$ microphones, where M is an even number. Thus, the microphone array consists of M microphones in total. The multi-microphone noise reduction methods considered in this paper operate in the frequency domain on a frame-by-frame

basis. Let l denote the frame index and k the frequency-bin index. Assume that there is only one target source and there are r interferers. The k-th frequency coefficient of the l-th frame of the j-th microphone noisy signal, $y_j(k,l)$, $j = 1, \cdots, M$, is given by

$$y_j(k,l) = \underbrace{a_j(k,l)s(k,l)}_{x_j(k,l)} + \sum_{i=1}^{r} \underbrace{b_{ij}(k,l)u_i(k,l)}_{n_{ij}(k,l)} + v_j(k,l), \quad (1)$$

where

- $s(k,l)$ denotes the target signal at the source location.
- $u_i(k,l)$, is the i-th interfering signal at the source location.
- $a_j(k,l)$ is the acoustic transfer function (ATF) of the target signal with respect to the j-th microphone.
- $b_{ij}(k,l)$ is the ATF of the i-th interfering signal with respect to the j-th microphone.
- $x_j(k,l)$ is the received target signal at the j-th microphone.
- $n_{ij}(k,l)$ is the i-th received interfering signal at the j-th microphone.
- $v_j(k,l)$ is additive noise at the j-th microphone.

Here we use in the signal model the ATFs for notational convinience. However, note that the ATFs can be replaced with relative acoustic transfer functions (RATF)s which can often be identified easier than the ATFs [12], [20].

In the remainder of the paper, the frequency and frame indices are neglected to simplify the notation. Using vector notation, Eq. (1) can be written as

$$\mathbf{y} = \mathbf{x} + \sum_{i=1}^{r} \mathbf{n}_i + \mathbf{v}, \quad (2)$$

where $\mathbf{y} \in \mathbb{C}^{M \times 1}$, $\mathbf{x} \in \mathbb{C}^{M \times 1}$, $\mathbf{n}_i \in \mathbb{C}^{M \times 1}$ and $\mathbf{v} \in \mathbb{C}^{M \times 1}$ are the stacked vectors of the y_j, x_j, n_{ij}, v_j (for $j = 1, \cdots, M$) components, respectively. Moreover, $\mathbf{x} = \mathbf{a}s$ and $\mathbf{n}_i = \mathbf{b}_i u_i$, where $\mathbf{a} \in \mathbb{C}^{M \times 1}$ and $\mathbf{b}_i \in \mathbb{C}^{M \times 1}$ are the stacked vectors of the a_j and b_{ij} (for $j = 1, \cdots, M$) components, respectively.

Assuming that all sources and the additive noise are mutually uncorrelated, the CPSD matrix of $\mathbf{y}$ is given by

$$\mathbf{P_y} = E[\mathbf{yy}^H] = \mathbf{P_x} + \underbrace{\sum_{i=1}^{r} \mathbf{P_{n_i}} + \mathbf{P_v}}_{\mathbf{P}}, \quad (3)$$

where

- $\mathbf{P_x} = E[\mathbf{xx}^H] = p_s \mathbf{aa}^H \in \mathbb{C}^{M \times M}$ is the CPSD matrix of $\mathbf{x}$, with $p_s = E[|s|^2]$ the power spectral density (PSD) of s.
- $\mathbf{P_{n_i}} = E[\mathbf{n}_i \mathbf{n}_i^H] = p_{u_i} \mathbf{b}_i \mathbf{b}_i^H \in \mathbb{C}^{M \times M}$ is the CPSD matrix of $\mathbf{n}_i$, with $p_{u_i} = E[|u_i|^2]$ the PSD of u_i.
- $\mathbf{P_v} = E[\mathbf{vv}^H] \in \mathbb{C}^{M \times M}$ is the CPSD matrix of $\mathbf{v}$.
- $\mathbf{P}$ is the total CPSD matrix of all disturbances.

III. BINAURAL BEAMFORMING

Binaural multi-microphone noise reduction methods aim at the simultaneous noise reduction and binaural cue preservation of the sources. In order to preserve the binaural cues, two different spatial filters $\hat{\mathbf{w}}_L \in \mathbb{C}^{M \times 1}$ and $\hat{\mathbf{w}}_R \in \mathbb{C}^{M \times 1}$, are applied to the left and right HA, respectively, where constraints can be used to guarantee that certain phase and magnitude relations between the left and right HA outputs are preserved. Note that both spatial filters use all microphone recordings from both HAs.

Without loss of generality, assume that the reference microphone for the left and right HA is indexed as $j = 1$ and $j = M$, respectively. In the sequel, for ease of notation, the reference terms of Eq. (1) use the subscripts L and R instead of $j = 1$ and $j = M$, respectively. The two enhanced output signals at the left and right HAs are then given by

$$\hat{x}_L = \hat{\mathbf{w}}_L^H \mathbf{y} \quad \text{and} \quad \hat{x}_R = \hat{\mathbf{w}}_R^H \mathbf{y}. \quad (4)$$

In Section III-A, objective measures for the preservation of binaural cues are presented. In Sections III-C—III-F, the BMVDR, the BLCMV, the OBLCMV, and the JBLCMV are reviewed, respectively. All reviewed methods are special cases of the general binaural LCMV (GBLCMV) framework, presented in Section III-B. Finally, the basic properties of all reviewed methods are summarized in Section III-G.

A. Binaural Cues

The extent to which the binaural cues of a specific source are preserved can be expressed using the input and output interaural tranfer function (ITF) [31], [32]. Often the ITF is decomposed into its magnitude, describing the interaural level differences (ILDs) and its phase, describing the interaural phase differences (IPDs). The input and output ITFs of the i-th interferer are defined as [32]

$$\mathrm{ITF}_{\mathbf{n}_i}^{\mathrm{in}} = \frac{n_{iL}}{n_{iR}} = \frac{b_{iL}}{b_{iR}}, \quad \mathrm{ITF}_{\mathbf{n}_i}^{\mathrm{out}} = \frac{\hat{\mathbf{w}}_L^H \mathbf{n}_i}{\hat{\mathbf{w}}_R^H \mathbf{n}_i} = \frac{\hat{\mathbf{w}}_L^H \mathbf{b}_i}{\hat{\mathbf{w}}_R^H \mathbf{b}_i}. \quad (5)$$

The input and output ILDs are defined as [32]

$$\mathrm{ILD}_{\mathbf{n}_i}^{\mathrm{in}} = |\mathrm{ITF}_{\mathbf{n}_i}^{\mathrm{in}}|^2, \quad \mathrm{ILD}_{\mathbf{n}_i}^{\mathrm{out}} = |\mathrm{ITF}_{\mathbf{n}_i}^{\mathrm{out}}|^2. \quad (6)$$

The input and output IPDs are given by [32]

$$\mathrm{IPD}_{\mathbf{n}_i}^{\mathrm{in}} = \angle \mathrm{ITF}_{\mathbf{n}_i}^{\mathrm{in}}, \quad \mathrm{IPD}_{\mathbf{n}_i}^{\mathrm{out}} = \angle \mathrm{ITF}_{\mathbf{n}_i}^{\mathrm{out}}. \quad (7)$$

Note that frequently, the IPDs are converted and measured as time delays [33], i.e., interaural time differences (ITDs). The IPDs and ILDs are the dominant cues for binaural localization for low and high frequencies, respectively [34]. Typically, the IPDs become more important for frequencies below 1 kHz, while ILDs become more important for frequencies above 3 kHz [34]. In [35] it was experimentally shown that for broadband signals, the IPDs are perceptually much more important than the ILDs for localizing a source. More specifically, it was shown that the low frequency IPDs play the most important role perceptually for correct localization. Based on this observation several proposed multi-microphone noise reduction techniques [33], [36] leave the low frequency content of the noisy measurements unprocessed, and process only the higher frequency content. Unfortunately, if a large portion of the power of the noise is concentrated at low frequencies, the noise reduction capabilities are reduced significantly. Therefore, in

this paper we aim at the simultaneous preservation of binaural cues of all sources and noise reduction at all frequencies.

A binaural spatial filter, $\hat{\mathbf{w}} = [\hat{\mathbf{w}}_L^T \quad \hat{\mathbf{w}}_R^T]^T$, exactly preserves the binaural cues of the i-th interferer if $\text{ITF}_{\mathbf{n}_i}^{\text{in}} = \text{ITF}_{\mathbf{n}_i}^{\text{out}}$ [32]. Exact preservation of ITFs also implies preservation of ILDs and IPDs [32], i.e., $\text{ILD}_{\mathbf{n}_i}^{\text{in}} = \text{ILD}_{\mathbf{n}_i}^{\text{out}}$ and $\text{IPD}_{\mathbf{n}_i}^{\text{in}} = \text{IPD}_{\mathbf{n}_i}^{\text{out}}$. Non-exact preservation of binaural cues implies that there is some positive ITF error given by

$$\mathcal{E}_{\mathbf{n}_i} = |\text{ITF}_{\mathbf{n}_i}^{\text{out}} - \text{ITF}_{\mathbf{n}_i}^{\text{in}}|. \quad (8)$$

Moreover, non-exact presevation of binaural cues implies that there is some ILD and/or IPD errors, given by

$$\mathcal{L}_{\mathbf{n}_i} = |\text{ILD}_{\mathbf{n}_i}^{\text{out}} - \text{ILD}_{\mathbf{n}_i}^{\text{in}}|, \quad \mathcal{T}_{\mathbf{n}_i} = \frac{|\text{IPD}_{\mathbf{n}_i}^{\text{out}} - \text{IPD}_{\mathbf{n}_i}^{\text{in}}|}{\pi}, \quad (9)$$

where $0 \leq \mathcal{T}_{\mathbf{n}_i} \leq 1$ [32]. Eqs. (5), (6), (7), (8) and (9) apply also for the target source $\mathbf{x}$. As it will become obvious in the sequel, for all methods that will be discussed in this paper, the errors in Eqs. (8), (9) with respect to the target source are always zero.

As explained before, the IPD error is perceptually more important measure for binaural localization than the ILD error for broadband signals (such as speech signals contaminated by broadband noise signals), because the IPDs are perceptually more important than the ILDs for this category of signals. Moreover, the IPD error is perceptually more informative at low frequencies, while the ILD error is perceptually more informative at high frequencies.

B. General Binaural LCMV Framework

All binaural LCMV-based methods discussed in this section are based on a general binaural LCMV (GBLCMV)[1] framework which is the binaural version of the classical LCMV framework [7], [8]. The GBLCMV minimizes the sum of the left and right output noise powers under multiple linear equality constraints. That is,

$$\hat{\mathbf{w}}_{\text{GBLCMV}} = \arg\min_{\mathbf{w} \in \mathbb{C}^{2M \times 1}} \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{\Lambda} = \mathbf{f}^H, \quad (10)$$

where $\hat{\mathbf{w}}_{\text{GBLCMV}} = [\hat{\mathbf{w}}_{\text{GBLCMV},L}^T \quad \hat{\mathbf{w}}_{\text{GBLCMV},R}^T]^T \in \mathbb{C}^{2M \times 1}$, $\mathbf{\Lambda} \in \mathbb{C}^{2M \times d}$ is assumed to be a full column rank matrix (i.e., $\text{rank}(\mathbf{\Lambda}) = d$), $\mathbf{f} \in \mathbb{C}^{d \times 1}$, d is the number of linear equality constraints, and

$$\tilde{\mathbf{P}} = \begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} \end{bmatrix} \in \mathbb{C}^{2M \times 2M}. \quad (11)$$

Similarly to the classical LCMV framework [7], [8], if $d \leq 2M$, and $\mathbf{\Lambda}$ is full column rank, the GBLCMV has a closed-form solution given by

$$\hat{\mathbf{w}}_{\text{GBLCMV}} = \begin{cases} \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda} \left(\mathbf{\Lambda}^H \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda} \right)^{-1} \mathbf{f} & \text{if } d < 2M \\ (\mathbf{\Lambda}^H)^{-1} \mathbf{f} & \text{if } d = 2M. \end{cases} \quad (12)$$

In GBLCMV, the total number of DOF devoted to noise reduction is $\text{DOF}_{\text{GBLCMV}} = 2M - d$. Note that in the special

[1]We used the word *general* in order to distinguish it from the BLCMV method [26], [27].

case where $d = 2M$, there are no DOF left for *controlled* noise reduction, i.e., $\hat{\mathbf{w}}_{\text{GBLCMV}}$ cannot reduce the objective function of the GBLCMV problem in a controlled way. Finally, if $d > 2M$, the feasible set is $\{\mathbf{w} : \mathbf{w}^H \mathbf{\Lambda} = \mathbf{f}^H\} = \emptyset$ and the GBLCMV problem has no solution. In conclusion, the matrix $\mathbf{\Lambda}$ has to be "tall" (i.e., $d < 2M$), to be able to simultaneously achieve controlled noise reduction and satisfy the constraints of the GBLCMV problem. The maximum number of constraints that the GBLCMV framework can handle, while achieving controlled noise reduction, is $d_{\text{max}} = 2M - 1$, i.e., there should be always left at least one DOF for noise reduction. Generally, the more DOF (i.e., the larger $\text{DOF}_{\text{GBLCMV}}$), the more controlled noise reduction can be achieved.

The set of linear constraints of the GBLCMV framework in Eq. (10) can be devided into two parts,

$$\mathbf{w}^H [\mathbf{\Lambda}_1 \mid \mathbf{\Lambda}_2] = [\mathbf{f}_1^H \mid \mathbf{f}_2^H]. \quad (13)$$

The first part consists of two distortionless constraints $\mathbf{w}_L^H \mathbf{a} = a_L$ and $\mathbf{w}_R^H \mathbf{a} = a_R$ which preserve the target source at the two reference microphones. This can be written compactly as

$$\mathbf{w}^H \mathbf{\Lambda}_1 = \mathbf{f}_1^H, \quad (14)$$

where

$$\mathbf{\Lambda}_1 = \begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{bmatrix} \in \mathbb{C}^{2M \times 2}, \quad \mathbf{f}_1 = \begin{bmatrix} a_L^* \\ a_R^* \end{bmatrix} \in \mathbb{C}^{2 \times 1}.$$

All binaural methods discussed in this section are special cases of the GBLCMV framework and they share the constraints in Eq. (14), while the constraints $\mathbf{w}^H \mathbf{\Lambda}_2 = \mathbf{f}_2^H$ are different.

In the sequel of the paper we use the term m (m_{max}) to indicate the number (maximum number) of interferers that a special case of the GBLCMV framework can preserve, while at the same time achieving controlled noise reduction. Recall that controlled noise reduction means that there is at least one DOF left for noise reduction. Moreover, $m_{\text{max}} \leq r$ which means that some methods may be unable to preserve all simultaneously present interferers of the acoustic scene, because there are not enough available DOF.

C. BMVDR

The BMVDR beamformer [30] can be formulated using the combination of the following two beamformers

$$\hat{\mathbf{w}}_{\text{BMVDR},L} = \arg\min_{\mathbf{w}_L \in \mathbb{C}^{M \times 1}} \mathbf{w}_L^H \mathbf{P} \mathbf{w}_L \text{ s.t. } \mathbf{w}_L^H \mathbf{a} = a_L, \quad (15)$$

$$\hat{\mathbf{w}}_{\text{BMVDR},R} = \arg\min_{\mathbf{w}_R \in \mathbb{C}^{M \times 1}} \mathbf{w}_R^H \mathbf{P} \mathbf{w}_R \text{ s.t. } \mathbf{w}_R^H \mathbf{a} = a_R, \quad (16)$$

with closed-form solutions

$$\hat{\mathbf{w}}_{\text{BMVDR},L} = \frac{\mathbf{P}^{-1} \mathbf{a} a_L^*}{\mathbf{a}^H \mathbf{P}^{-1} \mathbf{a}}, \quad \hat{\mathbf{w}}_{\text{BMVDR},R} = \frac{\mathbf{P}^{-1} \mathbf{a} a_R^*}{\mathbf{a}^H \mathbf{P}^{-1} \mathbf{a}}. \quad (17)$$

The BMVDR is the simplest special case of the GBLCMV framework in the sense that it has the minimum number of constraints ($d = 2$) given by Eq. (14). Specifically, the two optimization problems in Eqs. (15) and (16) can be reformulated as the following joint optimization problem,

$$\hat{\mathbf{w}}_{\text{BMVDR}} = \arg\min_{\mathbf{w} \in \mathbb{C}^{2M \times 1}} \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{\Lambda}_1 = \mathbf{f}_1^H, \quad (18)$$

where $\hat{\mathbf{w}}_{\text{BMVDR}} = [\hat{\mathbf{w}}_{\text{BMVDR},L}^T \quad \hat{\mathbf{w}}_{\text{BMVDR},R}^T]^T \in \mathbb{C}^{2M \times 1}$. Since, the BMVDR has the minimum possible number of constraints, the total number of DOF which can be devoted to noise reduction is $\text{DOF}_{\text{BMVDR}} = 2M - 2$.

The BMVDR preserves the binaural cues of the target source, but distorts the binaural cues of all the interferers [30], i.e., $m_{\max} = 0$. More specifically, after processing, the binaural cues of the interferers collapse on the binaural cues of the target source. It can be shown [30] that the binaural cues of the target source are preserved due to the satisfaction of the two distortionless constraints of the problems in Eqs. (15) and (16). That is,

$$\text{ITF}_{\mathbf{x}}^{\text{in}} = \text{ITF}_{\mathbf{x}}^{\text{out}} = \frac{a_L}{a_R}. \tag{19}$$

Therefore, the ITF error is $\mathcal{E}_{\mathbf{x},\text{BMVDR}} = 0$. Furthermore, it can be shown that the binaural cues of the interferers collapse to the binaural cues of the target source [30]. More specifically, the $\text{ITF}_{\mathbf{n}_i}^{\text{in}}$ is given by

$$\text{ITF}_{\mathbf{n}_i}^{\text{in}} = \frac{b_{iL}}{b_{iR}}, \tag{20}$$

while $\text{ITF}_{\mathbf{n}_i}^{\text{out}}$ is given by

$$\text{ITF}_{\mathbf{n}_i}^{\text{out}} = \frac{\hat{\mathbf{w}}_{\text{BMVDR},L}^H \mathbf{b}_i}{\hat{\mathbf{w}}_{\text{BMVDR},R}^H \mathbf{b}_i} = \frac{\frac{\mathbf{a}^H \mathbf{P}^{-1} \mathbf{b}_i a_L}{\mathbf{a}^H \mathbf{P}^{-1} \mathbf{a}}}{\frac{\mathbf{a}^H \mathbf{P}^{-1} \mathbf{b}_i a_R}{\mathbf{a}^H \mathbf{P}^{-1} \mathbf{a}}} = \frac{a_L}{a_R} = \text{ITF}_{\mathbf{x}}^{\text{in}}. \tag{21}$$

Thus, after processing, the interferers will have the same ITF as the target source and their ITF error is given by

$$\mathcal{E}_{\mathbf{n}_i,\text{BMVDR}} = \left| \text{ITF}_{\mathbf{n}_i}^{\text{out}} - \text{ITF}_{\mathbf{n}_i}^{\text{in}} \right| = \left| \frac{a_L}{a_R} - \frac{b_{iL}}{b_{iR}} \right|. \tag{22}$$

D. BLCMV

Another special case of the GBLCMV framework is the binaural linearly constrained minimum variance (BLCMV) beamformer [26], [27] which, unlike the BMVDR, uses additional constraints for the preservation of the binaural cues of m interferers. The left and right spatial filters of the BLCMV are given by [26], [27]

$$\hat{\mathbf{w}}_{\text{BLCMV},L} = \underset{\mathbf{w}_L \in \mathbb{C}^{M \times 1}}{\arg \min} \ \mathbf{w}_L^H \mathbf{P} \mathbf{w}_L$$
$$\text{s.t.} \quad \mathbf{w}_L^H \mathbf{a} = a_L$$
$$\mathbf{w}_L^H \mathbf{b}_1 = \eta_L b_{1L}, \ \dots, \ \mathbf{w}_L^H \mathbf{b}_m = \eta_L b_{mL}, \tag{23}$$

and

$$\hat{\mathbf{w}}_{\text{BLCMV},R} = \underset{\mathbf{w}_R \in \mathbb{C}^{M \times 1}}{\arg \min} \ \mathbf{w}_R^H \mathbf{P} \mathbf{w}_R$$
$$\text{s.t.} \quad \mathbf{w}_R^H \mathbf{a} = a_R$$
$$\mathbf{w}_R^H \mathbf{b}_1 = \eta_R b_{1R}, \ \dots, \ \mathbf{w}_R^H \mathbf{b}_m = \eta_R b_{mR}, \tag{24}$$

where the constraints $\mathbf{w}_L^H \mathbf{a} = a_L$ and $\mathbf{w}_R^H \mathbf{a} = a_R$ are the two common distortionless constraints used in all special cases in the GBLCMV framework, while the constraints $\mathbf{w}_L^H \mathbf{b}_i = \eta_L b_{iL}$ and $\mathbf{w}_R^H \mathbf{b}_i = \eta_R b_{iR}$, for $i = 1, \dots, m$, aim at a) preserving the binaural cues and b) supressing the m interferers. The amount of supression is controlled via the interference rejection parameters η_L and η_R which are pre-defined $(0 \leq \eta_L, \eta_R < 1)$ real-valued scalars. Binaural cue

preservation is achieved only if $\eta = \eta_L = \eta_R$ [26], [28]. The two problems in Eqs. (23) and (24) can be compactly formulated as a joint optimization problem. That is,

$$\hat{\mathbf{w}}_{\text{BLCMV}} = \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}}{\arg \min} \ \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \ \text{s.t.} \ \mathbf{w}^H \mathbf{\Lambda} = \mathbf{f}^H, \tag{25}$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_1 & \mathbf{\Lambda}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{a} & \mathbf{0} & \mathbf{b}_1 & \mathbf{0} & \cdots & \mathbf{b}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{a} & \mathbf{0} & \mathbf{b}_1 & \cdots & \mathbf{0} & \mathbf{b}_m \end{bmatrix}}_{\mathbb{C}^{2M \times (d=2+2m)}},$$

and

$$\mathbf{f}^T = \begin{bmatrix} \mathbf{f}_1^T & \mathbf{f}_2^T \end{bmatrix}$$
$$= \underbrace{\begin{bmatrix} a_L^* & a_R^* & \eta_L b_{1L}^* & \eta_R b_{1R}^* & \cdots & \eta_L b_{mL}^* & \eta_R b_{mR}^* \end{bmatrix}}_{\mathbb{C}^{1 \times (d=2+2m)}}.$$

The available DOF for noise reduction are $\text{DOF}_{\text{BLCMV}} = 2M - d = 2M - 2m - 2$. Since $d_{\max} = 2M - 1$ (see Section III-B), BLCMV can simultaneously achieve controlled noise suppression and binaural cue preservation of at most $m_{\max} = M - 2$ interferers.

The ITF errors of the target source and of the m interferers that are included in the constraints are zero, i.e., $\mathcal{E}_{\mathbf{x},\text{BLCMV}} = 0$ and $\mathcal{E}_{\mathbf{n}_i,\text{BLCMV}} = 0$, for $i = 1, \cdots, m \leq r$. However, if some interferers are not included in the constraints, their ITF error will be non-zero, i.e., $\mathcal{E}_{\mathbf{n}_i,\text{BLCMV}} > 0$, for $i = m + 1, \cdots, r$.

E. OBLCMV

The OBLCMV [28] can be seen as a special case of the BLCMV (and, hence, the GBLCMV) since it solves the same optimization problem. However, it preserves the binaural cues of only one interferer (e.g., the k-th interferer) using an optimal complex-valued interference rejection parameter $\hat{\eta} = \hat{\eta}_L = \hat{\eta}_R$ with respect to the binaural output SNR. More specifically, OBLCMV solves the problem in Eq. (25) where $\mathbf{\Lambda}$ and $\mathbf{f}^T$, are given by [28]

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_1 & \mathbf{\Lambda}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{a} & \mathbf{0} & \mathbf{b}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{a} & \mathbf{0} & \mathbf{b}_k \end{bmatrix} \in \mathbb{C}^{2M \times 4},$$

$$\mathbf{f}^T = \begin{bmatrix} \mathbf{f}_1^T & \mathbf{f}_2^T \end{bmatrix} = \begin{bmatrix} a_L^* & a_R^* & \hat{\eta} b_{kL}^* & \hat{\eta} b_{kR}^* \end{bmatrix} \in \mathbb{C}^{1 \times 4} \tag{26}$$

where $1 \leq k \leq r$. The available DOF for noise reduction are $\text{DOF}_{\text{OBLCMV}} = 2M - 4$. The ITF errors of the target source and of the k-th interferer that are included in the constraints are zero, i.e., $\mathcal{E}_{\mathbf{x},\text{OBLCMV}} = 0$ and $\mathcal{E}_{\mathbf{n}_k,\text{OBLCMV}} = 0$. However, the binaural cues of all the other $r - 1$ interferers will be distorted, i.e., $\mathcal{E}_{\mathbf{n}_i,\text{BLCMV}} > 0$, for $i \in \{1, \cdots, r\} - \{k\}$.

F. JBLCMV

Recall from Section III-A that preserving binaural cues of the i-th interferer implies that the following constraint has to be satisfied

$$\text{ITF}_{\mathbf{n}_i}^{\text{in}} = \text{ITF}_{\mathbf{n}_i}^{\text{out}} \implies \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} = \frac{b_{iL}}{b_{iR}}, \tag{27}$$

which can be reformulated as:

$$\mathbf{w}_L^H \mathbf{b}_i b_{iR} - \mathbf{w}_R^H \mathbf{b}_i b_{iL} = 0. \tag{28}$$

Compared to (O)BLCMV this unified constraint reduces the number of constraints, used for binaural cue preservation, by a factor 2. As a result, for a given number of interferers, more DOF can be devoted to noise reduction. The JBLCMV [29], [30] uses this type of equality constraints for the preservation of the binaural cues of m interferers. More specifically, the JBLCMV problem is given by

$$\hat{\mathbf{w}}_{\text{JBLCMV}} = \arg\min_{\mathbf{w} \in \mathbb{C}^{2M \times 1}} \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{\Lambda} = \mathbf{f}^H, \qquad (29)$$

where

$$\begin{aligned}
\mathbf{\Lambda} &= \begin{bmatrix} \mathbf{\Lambda}_1 \mid \mathbf{\Lambda}_2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{a} & \mathbf{0} & \mathbf{b}_1 b_{1R} & \cdots & \mathbf{b}_m b_{mR} \\ \mathbf{0} & \mathbf{a} & -\mathbf{b}_1 b_{1L} & \cdots & -\mathbf{b}_m b_{mL} \end{bmatrix} \in \mathbb{C}^{2M \times (2+m)}
\end{aligned} \tag{30}$$

and $\mathbf{w}_{\text{JBLCMV}} = [\mathbf{w}_{\text{JBLCMV},L}^T \quad \mathbf{w}_{\text{JBLCMV},R}^T]^T$. Moreover,

$$\begin{aligned}
\mathbf{f}^T &= \begin{bmatrix} \mathbf{f}_1^T \mid \mathbf{f}_2^T \end{bmatrix} \\
&= \begin{bmatrix} a_L^* & a_R^* \mid 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{C}^{1 \times (2+m)}.
\end{aligned} \tag{31}$$

Similarly to all other special cases of the GBLCMV framework, $\mathbf{w}^H \mathbf{\Lambda_1} = \mathbf{f}_1^H$ is used for the exact binaural cue preservation of the target source, while $\mathbf{w}^H \mathbf{\Lambda_2} = \mathbf{f}_2^H$ is used for the preservation of the binaural cues of m interferers.

The JBLCMV can simultaneously achieve controlled noise reduction and binaural cue preservation of up to $m_{\max} = 2M - 3$ interferers [29]. Moreover, the DOF devoted to noise reduction is $\text{DOF}_{\text{JBLCMV}} = 2M - m - 2$.

G. Summary of GBLCMV methods

We summarize some of the properties of the methods discussed in Section III. Table I gives an overview of two important factors: a) the maximum number of interferers' binaural cues that can be preserved while achieving controlled noise reduction $m_{\max}$, and b) the degrees of freedom (DOF) available for noise reduction. The following conclusions can be drawn from this table:

- The BMVDR has the maximum DOF, which means that it can achieve the best possible noise reduction. It preserves the binaural cues of the target source, but not the binaural cues of the interferers.
- Unlike (O)BLCMV which uses two constraints per interferer, JBLCMV uses only one constraint per interferer. Therefore, the JBLCMV can preserve the binaural cues of more interferers, or equivalently, given the same number of interferers it has more available DOF devoted to noise reduction.

In this paper, if the number of simultaneously present interferers is $r > m_{\max}$, the extra interferers $r - m_{\max}$ are *not* included in the constraints in the GBLCMV methods, in order to always have one DOF left for controlled noise reduction.

IV. PROPOSED NON-CONVEX PROBLEM

In this section, we present a general optimization problem of which BMVDR and JBLCMV are special cases. More specifically, we relax the constraints on the binaural cues of the interferers, while keeping the strict equality constraints on

TABLE I
SUMMARY OF A) MAXIMUM NUMBER OF INTERFERERS' BINAURAL CUES THAT CAN BE PRESERVED WHILE ACHIEVING CONTROLLED NOISE REDUCTION (m_{MAX}), AND B) NUMBER OF AVAILABLE DEGREES OF FREEDOM FOR NOISE REDUCTION (DOF). ALL METHODS ARE SPECIAL CASES OF THE GBLCMV FRAMEWORK. M IS THE TOTAL NUMBER OF MICROPHONES, AND m IS THE NUMBER OF THE CONSTRAINED INTERFERERS.

Method	$m_{\max}$	DOF
BMVDR [30]	0	$2M - 2$
BLCMV [27]	$M - 2$	$2M - 2m - 2$
OBLCMV [28]	1	$2M - 4$
JBLCMV [29], [30]	$2M - 3$	$2M - m - 2$

the target source (i.e., $\mathbf{w}^H \mathbf{\Lambda}_1 = \mathbf{f}_1^H$). The relaxation allows to trade-off the amount of noise reduction and binaural cue preservation per interferer in a controlled way. The proposed optimization problem is defined as

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w} \in \mathbb{C}^{2M \times 1}} \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{\Lambda}_1 = \mathbf{f}_1^H,$$

$$\underbrace{\left| \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} - \frac{b_{iL}}{b_{iR}} \right|}_{\mathcal{E}_{\mathbf{n}_i}} \le e_i, \quad i = 1, \cdots, m. \tag{32}$$

The inequality constraints bound the ITF error (see Eq. (8)), for the interferers $i = 1, \cdots, m$ to be less than a positive trade-off parameter $e_i, i = 1, \cdots, m$. These inequality constraints will be transformed, in the sequel of this section (see Eqs. (34), (35)), in such a way that they can be viewed as relaxations of the strict equality constraints in Eq. (28) used in the JBLCMV method. Note that the proposed method is flexible to choose a different e_i for every interferer according to its importance. For instance, maybe certain locations are more important to be preserved than others and, therefore, a smaller e_i must be used. The trade-off parameter, e_i, is selected as

$$e_i(c_i) = c_i \mathcal{E}_{\mathbf{n}_i,\text{BMVDR}}, \tag{33}$$

where $0 \le c_i \le 1$ controls the amount of binaural cue collapse towards the target source, and the amount of noise reduction of the i-th interferer. If $c_i = 1, \forall i$ is used in the optimization problem in Eq. (32), then $\hat{\mathbf{w}} = \hat{\mathbf{w}}_{\text{BMVDR}}$ which is seen as a worst case, with respect to binaural cue preservation, because there is total collapse of binaural cues of the interferers towards the binaural cues of the target source. If $c_i = 0, \forall i$ we have perfect preservation of binaural cues of the m interferers, and $\hat{\mathbf{w}} = \hat{\mathbf{w}}_{\text{JBLCMV}}$. Without any loss of generality, for notational convenience, we assume that the binaural cues of all interferers are of equal importance and, therefore, $c_i = c, \forall i$. Moreover, we keep c fixed over all frequency bins. It is worth noting that other strategies for choosing c may exist, which might lead to a better trade-off between maximum possible noise reduction and perceptual binaural cue preservation. As explained in Section III-A, low frequency content is perceptually more important for binaural cue preservation than high frequency content. Thus, smaller c values for low frequencies and larger c values for higher frequencies may give a better perceptual trade-off.

The problem in Eq. (32) is not a convex problem and it is hard to solve. In Section V we propose a method that approximately solves the non-convex problem in an iterative way by solving at each iteration a convex problem.

V. PROPOSED ITERATIVE CONVEX PROBLEM

By doing some simple algebraic manipulations, the optimization problem in Eq. (32) can equivalently be written as

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{C}^{2M \times 1}} \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{\Lambda}_1 = \mathbf{f}_1^H,$$

$$\frac{|\mathbf{w}_L^H \mathbf{b}_i b_{iR} - \mathbf{w}_R^H \mathbf{b}_i b_{iL}|}{|\mathbf{w}_R^H \mathbf{b}_i b_{iR}|} \le e_i(c), \text{ for } i = 1, \cdots, m. \quad (34)$$

Furthermore, the problem in Eq. (34) can be re-written as

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{C}^{2M \times 1}} \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{\Lambda}_1 = \mathbf{f}_1^H,$$

$$|\mathbf{w}^H \mathbf{\Lambda}_{2,i}| \le \underbrace{|e_i(c) \mathbf{w}_R^H \mathbf{b}_i b_{iR}|}_{f_{2,i}}, \text{ for } i = 1, \cdots, m, \quad (35)$$

where $\mathbf{\Lambda}_{2,i}$ is the i-th column of $\mathbf{\Lambda}_2$ in Eq. (30).

We approximately solve the non-convex problem in Eq. (35) in an iterative way using $\mathbf{w}_R^H$ of the previous iteration in $f_{2,i}, i = 1, \cdots, m$. The new iterative problem is convex at each iteration and is given by

$$\hat{\mathbf{w}}_{(k)} = \arg \min_{\mathbf{w} \in \mathbb{C}^{2M \times 1}} \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{\Lambda}_1 = \mathbf{f}_1^H,$$

$$|\mathbf{w}^H \mathbf{\Lambda}_{2,i}| \le \underbrace{|e_i(c) \hat{\mathbf{w}}_{R,(k-1)}^H \mathbf{b}_i b_{iR}|}_{f_{2,i,(k)}}, \text{ for } i = 1, \cdots, m, \quad (36)$$

where $\hat{\mathbf{w}}_{(k)} = [\hat{\mathbf{w}}_{L,(k)}^T \quad \hat{\mathbf{w}}_{R,(k)}^T]^T$ is the estimated binaural spatial filter of the k-th iteration, which is initialized as $\hat{\mathbf{w}}_{(0)} = \hat{\mathbf{w}}_{\text{BMVDR}}$. Similarly to other existing minimum variance beamformers with inequality constraints [37], [38], the convex optimization problem in Eq. (36) can be equivalently written as a second order cone programming (SOCP) problem with equality and inequality constraints (see Appendix) and it can be solved efficiently with interior point methods [39].

The ITF error of the i-th interferer at the k-th iteration is given by

$$\mathcal{E}_{\mathbf{n}_i,(k)} = \left| \frac{\hat{\mathbf{w}}_{L,(k)}^H \mathbf{b}_i}{\hat{\mathbf{w}}_{R,(k)}^H \mathbf{b}_i} - \frac{b_{iL}}{b_{iR}} \right|. \quad (37)$$

This iterative method is stopped when all the constraints of the original problem in Eq. (32) are satisfied. Therefore, the stopping criterion that we use is given by

$$\mathcal{E}_{\mathbf{n}_i,(k)} \le e_i(c), \text{ for } i = 1, \cdots, m, \quad (38)$$

where $e_i(c)$ is given in Eq. (33). Recall that $\mathbf{f}_2 = \mathbf{0}$ (i.e., $f_{2,i} = 0, \forall i$) is used in JBLCMV. Unlike JBLCMV, the proposed method uses $f_{2,i,(k)} \ge 0, \forall i$ and, therefore, the constraints dedicated for the preservation of binaural cues of the interferers are seen as relaxations of the strict equality constraints of the JBLCMV method. These relaxations enlarge the feasible set of the problem, allowing more constraints to be used compared to JBLCMV. The JBLCMV can be seen as a special case of the proposed method for $c = 0$, $f_{2,i,(1)} = 0, i = 1, \cdots m$. In this case, the relaxed constraints in the proposed method

become identical to the strict constraints of the JBLCMV. Hence, the JBLCMV needs to run only one iteration of the problem in Eq. (36). If $c = 0$, the proposed method follows the same strategy for handling $r > m_{\max}$ simultaneously present interferers as in Section III-G. However, if $c > 0$, then there is a typically large, difficult to predict $m_{\max}$[2], due to the inequality constraints and, therefore, the proposed method uses $m = r, \forall r$ constraints for the preservation of the binaural cues of all simultaneously present interferers. Finally, if $c = 1$, the proposed method does not iterate and stops immediately giving as output the initialization $\hat{\mathbf{w}}_{(0)} = \hat{\mathbf{w}}_{\text{BMVDR}}$.

The termination of the proposed iterative method *may* need a large amount of iterations because of the fixed c in Eq. (36). The reason for this is explained in detail in Section V-A. To control the speed of termination we replace in Section V-B the fixed c in Eq. (36) with a decreasing parameter $\tau_{(k)}$ (initialized with $\tau_{(0)} = c$) which controls the speed of termination. In Section V-C we show under which conditions the proposed method: a) guarantees that it will find a feasible solution satisfying the stopping criterion in Eq. (38) in a finite number of iterations, and b) guarantees a bounded amount of binaural cue preservation and a bounded amount of noise reduction. An overview of the proposed method using the adaptive $\tau_{(k)}$ is given in Algorithm 1.

A. Speed of Termination

The proposed iterative method may have slow termination due to the fixed choice of c. In this section we explain the reason and in Section V-B we explain how to control the speed of termination.

Let $\Phi_{(k)}$ denote the convex feasible set in the k-th iteration of the iterative optimization problem in Eq. (36) given by

$$\Phi_{(k)} = \bigcap_{i=1}^{m} \left\{ \mathbf{w}_{(k)} : \mathbf{\Lambda}_1^H \mathbf{w}_{(k)} = \mathbf{f}_1, |\mathbf{w}_{(k)}^H \mathbf{\Lambda}_{2,i}| \le f_{2,i,(k)} \right\}, \quad (39)$$

and $\Psi(c)$ the non-convex feasible set of the original non-convex problem of Eqs. (32), (33) given by

$$\Psi(c) = \bigcap_{i=1}^{m} \left\{ \mathbf{w} : \mathbf{\Lambda}_1^H \mathbf{w} = \mathbf{f}_1, \left| \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} - \frac{b_{iL}}{b_{iR}} \right| \le e_i(c) \right\}, \quad (40)$$

where $\hat{\mathbf{w}}_{\text{JBLCMV}} \in \Psi(0)$, and $\Psi(0) \subseteq \Psi(c), 0 \le c \le 1$ and, therefore, $\hat{\mathbf{w}}_{\text{JBLCMV}} \in \Psi(c), 0 \le c \le 1$. In words, $\hat{\mathbf{w}}_{\text{JBLCMV}}$ is an element of the set $\Psi(0)$, which gives the minimum output noise power compare to the other elements of $\Psi(0)$. Note that the $\Phi_{(k)}$ changes for every next iteration, while $\Psi(c)$ is constant over time. We can think of $\Phi_{(k)}$ as a convex approximation set of $\Psi(c)$ at iteration k (see a simplistic example of the two sets in Fig. 1(a)).

Note that the proposed iterative method will typically try to find a solution on the boundary of $\Phi_{(k)}$. Some parts of the boundary of $\Phi_{(k)}$ will be inside or on the boundary of $\Psi(c)$, while other parts can be outside the set $\Psi(c)$. Therefore, it is possible that the estimated $\hat{\mathbf{w}}_{(k)}$ will be outside of $\Psi(c)$ (see

---

[2]The feasible set of the proposed method typically reduces by adding more inequality constraints. However it is difficult to predict after how many constraints, m, it becomes empty, i.e., what is the value of $m_{\max}$.

Algorithm 1 Proposed Iterative Method

Input: $c, k_{\max}, \mathbf{a}, \mathbf{b}_i, i = 1, \cdots, m$

Output: $\hat{\mathbf{w}}_{(k)}$

    Initialisation : $\hat{\mathbf{w}}_{(0)} \leftarrow \hat{\mathbf{w}}_{\text{BMVDR}}, k \leftarrow 1, \tau_{(0)} \leftarrow c$

    General comments :

        {SC stands for stopping criterion in Eq. (38)}.

        {SP stands for solving problem in Eq. (36)}.

1: **if** $\text{SC}(\hat{\mathbf{w}}_{(0)}, c) =$ **true then**

2:    go to 17

3: **end if**

    start iterations

4: **while** $k \leq k_{\max}$ **do**

5:    **if** $k = k_{\max}$ **then**

6:        $\hat{\mathbf{w}}_{(k)} \leftarrow \text{SP}\left(\hat{\mathbf{w}}_{(k-1)}, \tau_{(k)}, \mathbf{a}, \mathbf{b}_i, i = 1, \cdots, 2M - 3\right)$

7:        go to 17

8:    **else**

9:        $\hat{\mathbf{w}}_{(k)} \leftarrow \text{SP}\left(\hat{\mathbf{w}}_{(k-1)}, \tau_{(k)}, \mathbf{a}, \mathbf{b}_i, i = 1, \cdots, m\right)$

10:    **end if**

11:    **if** $\text{SC}(\hat{\mathbf{w}}_{(k)}, c) =$ **true then**

12:        go to 17

13:    **end if**

14:    $k \leftarrow k + 1$

15:    $\tau_{(k)} = \tau_{(k-1)} - c/k_{\max}$

16: **end while**

17: **return** $\hat{\mathbf{w}}_{(k)}$

Fig. 1. Simplistic visualization of two successive iterations (k and $k + 1$) of the proposed method with (a) a fixed c, (b) a reducing $\tau(k)$. In $k+1$ iteration the stopping criterion is satisfied in (b). On the contrary, in (a) the stopping criterion is not satisfied, because $\hat{\mathbf{w}}_{(k+1)} \notin \Psi(c)$.

Fig. 1(a) for instance). In this case, obviously, the stopping criterion is not satisfied and, therefore, the problem goes to the next iteration. In the next iteration, $\Phi_{(k+1)}$ changes and a new $\hat{\mathbf{w}}_{(k+1)}$ is estimated which can be again outside of $\Psi(c)$ (see Fig. 1(a) for instance). This repetition can happen many times leading to a very slow termination because the new estimate $\hat{\mathbf{w}}_{(k+1)}$ is not selected according to a binaural-cue error descent direction. To avoid this undesirable situation, we propose in Section V-B to replace the fixed c in Eq. (36) with an adaptive reduction parameter $\tau_{(k)}$, in order to make sure that solutions that are on the boundary of $\Phi_{(k)}$ and that are outside $\Psi(c)$ will progressively provide a reduced binaural-cue error, i.e., to move towards the direction of the interior of $\Psi(c)$ (see Fig. 1(b) for instance).

B. Avoiding Slow Termination

The termination of the proposed iterative method *may* need a large amount of iterations because of the fixed c in Eq. (36), as explained in Section V-A. Therefore, the replacement of c with an adaptive reduction parameter $\tau_{(k)}$ *only* in Eq. (36) is useful for guaranteed termination within a *pre-selected* finite maximum number of iterations, $k_{\max}$. More specifically, the new adaptive reduction parameter that we use in Eq. (36) instead of c is given by

$$\tau_{(k)} = \tau_{(k-1)} - \alpha_{(k_{\max})}, \qquad (41)$$

where $\tau_{(0)} = c$ is selected according to the initial desired amount of collapse of binaural cues in the original non-convex problem in Eqs. (32), (33). The step $\alpha_{(k_{\max})}$ controls the speed

of termination, and is a function of the maximum allowed number of iterations for termination given from the user, i.e.,

$$\alpha_{(k_{\max})} = \frac{c}{k_{\max}}. \qquad (42)$$

Note that we replace c with $\tau_{(k)}$ only in Eq. (36) and not in the stopping criterion in Eq. (38). This is because, the stopping criterion is based on the fixed feasible set $\Psi(c)$ of the non-convex problem in Eq. (32) which should remain constant over iterations (see an example of two consecutive iterations in Fig. 1). Moreover, the $\tau_{(k)}$ is always non-negative, because $\tau_{(k_{\max})} = 0$. Small $k_{\max}$, speeds up the reduction of $\tau_{(k)}$ and, thus, it also speeds up the termination of the proposed method. Of course a very small $k_{\max}$ can lead to a feasible solution, $\hat{\mathbf{w}}_{(k)}$, for which $\sum_i \mathcal{E}_{\mathbf{n}_i,(k)} \ll \sum_i e_i(c)$, i.e., to be far away from the boundary of $\Psi(c)$. This means that $\hat{\mathbf{w}}_{(k)}$ provides better binaural cue preservation than the desired amount of binaural cue preservation, $e_i(c)$. As a result, there will be less noise suppression. Ideally, we would like to arrive as close as possible to the controlled trade-off between noise reduction and binaural cue preservation given by our initial specifications (i.e., amount of collapse). Therefore, a careful choice of $k_{\max}$ is needed in order to find a feasible solution $\hat{\mathbf{w}}_{(k)}$ that:

- achieves a total ITF error $\sum_i \mathcal{E}_{\mathbf{n}_i,(k)} \approx \sum_i e_i(c)$, i.e., to be as close as possible to the boundary of $\Psi(c)$[3].
- to terminate as fast as possible.

Of course there is a trade-off between the two goals.

C. Guarantees

In this section, we prove that the proposed iterative method using the adaptive reduction parameter in Eq. (41) *guarantees* termination, a bounded binaural cue preservation accuracy,

---

[3]Note that there may not be any element on the boundary (or in the interior) of $\Psi(c)$, which provides a total ITF error of $\sum_i e_i(c)$. The max possible total ITF error of $\Psi(c)$ may be less than $\sum_i e_i(c)$. This depends mainly on the number of constraints. Nevertheless, in general, the smaller the difference $\sum_i \mathcal{E}_{\mathbf{n}_i,(k)} - \sum_i e_i(c)$ is, the closer to the boundary of $\Psi(c)$ is the solution.

9

and a bounded amount of noise reduction, in *at most* k_{max} iterations, for a limited number of interferers $m \leq 2M - 3$. Nevertheless, our simulation experiments (see Section VI-C) show that our algorithm a) is capable of simultaneously achieving the same bounds for binaural cue preservation accuracy and for noise reduction of *more* interferers than $2M - 3$ for $c > 0$, and b) finds a feasible solution in much fewer iterations, on average, than k_{max}, for $k_{max} = 10, 50$.

The adaptive decreasing of $\tau_{(k)}$ (see Eq. (41)) results in an adaptive shrinking of $\Phi_{(k)}$. Therefore, in the case where the estimated $\hat{\mathbf{w}}_{(k)}$ will be outside of $\Psi(c)$, the stopping criterion is not satisfied and, therefore, the algorithm continues with the next iteration. In the next iteration, $\Phi_{(k)}$ typically shrinks due to the decreased value of $\tau_{(k)}$ according to Eq. (41). The algorithm continues until there is a solution $\hat{\mathbf{w}}_{(k)} \in \Psi(c)$. Note that this does not necessarily mean that the algorithm will stop if and only if $\Phi_{(k)} \subseteq \Psi(c)$ (see e.g., Fig. 1(b) where the algorithm stops before $\Phi_{(k)} \subseteq \Psi(c)$). Only in the worst case scenario a solution is found when $\Phi_{(k)} \subseteq \Psi(c)$.

We show below that, for $m \leq 2M-3$, the proposed method guarantees termination within a pre-defined finite maximum number of iterations, k_{max}, while achieving a bounded binaural cue preservation accuracy and a bounded amount of noise reduction. This is written more formally in Theorem 1.

Theorem 1. *If $m \leq 2M - 3$, the proposed method a) will always find a solution in a finite number of iterations $k \leq k_{max}$ satisfying the stopping criterion of Eq. (38), and b) will always have a bounded ITF error, i.e.,*

$$0 \leq \mathcal{E}_{\mathbf{n}_i,(k)} \leq e_i(c), \text{ for } i = 1, \cdots, m, \quad (43)$$

and a bounded noise output power

$$\hat{\mathbf{w}}_{BMVDR}^H \tilde{\mathbf{P}} \hat{\mathbf{w}}_{BMVDR} \leq \hat{\mathbf{w}}_{(k)}^H \tilde{\mathbf{P}} \hat{\mathbf{w}}_{(k)} \leq \hat{\mathbf{w}}_{JBLCMV}^H \tilde{\mathbf{P}} \hat{\mathbf{w}}_{JBLCMV}. \quad (44)$$

Proof. Note that for $m \leq 2M-3$, after k_{max} iterations $\tau_{(k_{max})} = 0$ (see Eqs. (41) and (42)) and, therefore, $\hat{\mathbf{w}}_{(k_{max})} = \hat{\mathbf{w}}_{JBLCMV}$ because the relaxations of the proposed method in Eq. (36) become $\hat{\mathbf{w}}_{(k_{max})}^H \mathbf{\Lambda}_2 = 0$, which is the same as in JBLCMV as explained in Section V. Note also that $\hat{\mathbf{w}}_{JBLCMV}$ always satisfies the stopping criterion, i.e., $\hat{\mathbf{w}}_{JBLCMV} \in \Psi(c)$, for $0 \leq c \leq 1$ (see Section V-A). Therefore, for $m \leq 2M - 3$, the algorithm, in the worst case scenario, will terminate after k_{max} iterations. Consequently, the first part of the theorem has been proved. Thus, in the worst case scenario, the algorithm gives the solution $\hat{\mathbf{w}}_{JBLCMV}$ which results in $\mathcal{E}_{\mathbf{n}_i,(k)} = 0$ for $i = 1, \cdots, m$. Since the algorithm always terminates (i.e., satisfies the stopping criterion), the ITF error will always be $\mathcal{E}_{\mathbf{n}_i,(k)} \leq e_i(c)$, for $i = 1, \cdots, m$. Thus, Eq. (43) has been proved. Moreover, the algorithm in the worst case scenario (after k_{max}) will have the noise output power $\hat{\mathbf{w}}_{JBLCMV}^H \tilde{\mathbf{P}} \hat{\mathbf{w}}_{JBLCMV}$. Finally, the noise output power cannot be less than $\hat{\mathbf{w}}_{BMVDR}^H \tilde{\mathbf{P}} \hat{\mathbf{w}}_{BMVDR}$ (because $\hat{\mathbf{w}}_{BMVDR}$ achieves the best noise reduction over all the aforementioned methods, because it has the largest feasible set). Thus, Eq. (44) has been proved. □

Note that, for $k = k_{max}$ and $m > 2M - 2$, $\Phi_{(k_{max})} = \emptyset^4$. However, for $k < k_{max}$ and $m > 2M - 2$, $\Phi_{(k)}$ may not be

---

^{4}Recall that for $m = 2M - 2$ (i.e., $d = 2M$), there is a feasible solution which does not provide controlled noise reduction (see Section III-B).
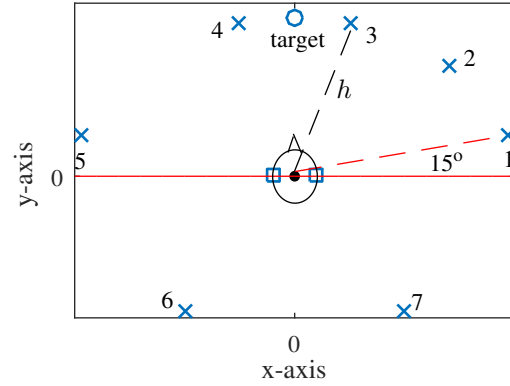
Fig. 2. Experimental setup: □ HAs, 'o' target source, 'x' speech shaped interferers. Each source has the same distance, h, from the center of the head.

empty. As we will show in our experiments, indeed, usually it is not empty and, thus, we may achieve simultaneous bounded approximate binaural cue preservation and bounded noise reduction of $m > 2M - 2$ interferers. This can be observed experimentally in Sections VI-C2 and VI-C3.

VI. EXPERIMENTAL RESULTS

In this section, the proposed method, summarized in Algorithm 1, is experimentally evaluated. In Section VI-A, the setup of our experiments is demonstrated. In Section VI-B, the performance measures are presented. In Section VI-C, the proposed method is compared to other LCMV-based methods with regard to binaural cue preservation and noise reduction. Moreover, we provide results with regard to the speed of the proposed method in terms of number of iterations.

A. Experiment Setup

Fig. 2 shows the experimental setup that we used. Two behind-the-ear (BTE) HAs, with two microphones each, are simulated and, therefore, the total number of microphones is $M = 4$. The publicly available database with the BTE impulse responses (IRs) in [40] is used to simulate the head IRs (we used the front and middle microphone for each HA). The front microphones are selected as reference microphones.

We placed all sources on a $h = 80$ cm radius circle centered at the origin $(0,0)$ (center of head) with an elevation of $0°$ degrees. The index of each interferer (denoted by 'x' marker) is indicated in Fig. 2. The interferers $1, 2, 3, 4, 5, 6$ and 7 are speech shaped noise realizations with the same power and are placed at $15°$, $45°$, $75°$, $105°$, $165°$, $240°$ and $300°$ degrees, respectively. The target source (denoted by 'o' marker) is a speech signal in the look direction, i.e., $90°$ degrees.

The duration of all sources is 60 sec. The microphone self noise at each microphone is simulated as white Gaussian noise (WGN) with $\mathbf{P_V} = \sigma^2 \mathbf{I}$, where $\sigma = 3.8 * 10^{-5}$ which corresponds to an SNR of 50 dB with respect to the target signal at the left reference microphone. The noise CPSD matrices, $\mathbf{P}$, are calculated (as in Eq. (3)) using the ATFs of the truncated true BTE IRs, from the database, and the estimated PSDs of the sources using all available data without

voice activity detection (VAD) errors. Also, the constraints of all the aforementioned methods use the ATFs of the truncated true BTE IRs. The truncated BTE IRs length is 50 ms. The sampling frequency is $f_s = 16$ kHz. We use a simple overlap-and-add analysis/synthesis method [41] with frame length 10 ms, overlap 50% and an FFT size of 1024. The analysis/synthesis window is a square-root-Hann window. The ATFs are also computed with an FFT size of 1024. The microphone signals are computed by convolving the truncated BTE IRs with the source signals at the original locations.

B. Performance Evaluation

In this section we define the performance evaluation measures that we use to evaluate the results.

1) ITFs, IPDs & ILDs: Here we define four average performance measures for binaural cue preservation: the total ILD error, the total IPD error, the total ITF error, and the average ITF error ratio. As explained in Section III-A, the IPD errors are perceptually more important for frequencies below 1 kHz, and the ILD errors are perceptually more important for frequencies above 3 kHz. Thus, the evaluation of IPDs and ILDs will be done only for these frequency regions. We evaluate the total ILD and IPD errors as follows. Let $\mathcal{L}_{\mathbf{n}_i}(k,l)$ and $\mathcal{T}_{\mathbf{n}_i}(k,l)$ denote the ILD and IPD errors (for the k-th frequency bin and l-th frame), respectively, defined in Eq. (9). Then the total ILD and IPD errors are defined as

$$\text{TotER}^{\text{ILD}} = \sum_{i=1}^{r} \left(\frac{1}{N - k_{\text{ILD}}} \sum_{k=k_{\text{ILD}}}^{N} \left(\frac{1}{T} \sum_{l=1}^{T} \mathcal{L}_{\mathbf{n}_i}(k,l) \right) \right), \quad (45)$$

and

$$\text{TotER}^{\text{IPD}} = \sum_{i=1}^{r} \left(\frac{1}{k_{\text{IPD}}} \sum_{k=1}^{k_{\text{IPD}}} \left(\frac{1}{T} \sum_{l=1}^{T} \mathcal{T}_{\mathbf{n}_i}(k,l) \right) \right), \quad (46)$$

where N and T are the number of frequency bins and the number of frames, respectively, k_{ILD} and k_{IPD} are the first and last frequency-bin indices in the frequency regions $3 - 8$ kHz and $0 - 1$ kHz, respectively. Note that since the maximum possible value of $\mathcal{T}_{\mathbf{n}_i}(k,l)$ is 1, the maximum value of $\text{TotER}^{\text{IPD}}$ is r. Moreover, we evaluate the total ITF error given by

$$\text{TotER}^{\text{ITF}} = \sum_{i=1}^{r} \left(\frac{1}{N} \sum_{k=1}^{N} \left(\frac{1}{T} \sum_{l=1}^{T} \mathcal{E}_{\mathbf{n}_i}(k,l) \right) \right), \quad (47)$$

where $\mathcal{E}_{\mathbf{n}_i}$ is the ITF error defined in Eq. (8). Finally, we evaluate the average ITF error ratio given by

$$\text{AvER}^{\text{ITF}}(c) = \frac{1}{r} \sum_{i=1}^{r} \frac{1}{N} \sum_{k=1}^{N} \frac{1}{T} \sum_{l=1}^{T} \frac{\mathcal{E}_{\mathbf{n}_i}(k,l)}{\mathcal{E}_{\mathbf{n}_i,\text{BMVDR}}(k,l)}, \quad (48)$$

which measures the average amount of binaural cue collapse by comparing the ITF error of the proposed method with the ITF error of the BMVDR. Since the proposed method will always satisfy the condition $\mathcal{E}_{\mathbf{n}_i}(k,l) \leq c\mathcal{E}_{\mathbf{n}_i,\text{BMVDR}}(k,l)$ for $r \leq 2M - 3$ (see Theorem 1), obviously $\text{AvER}^{\text{ITF}}(c) \leq c$ for $r \leq 2M - 3$. Note that ideally the proposed method will provide a solution as close as possible to the boundary of

$\Psi(c)$, i.e., $\text{AvER}^{\text{ITF}}(c) - c$ to be as small as possible (see Section V-B). Moreover, for the proposed method $\text{AvER}^{\text{ITF}}(0) = 0$ and $\text{AvER}^{\text{ITF}}(1) = 1$ because for $c = 0$, $\mathcal{E}_{\mathbf{n}_i}(k,l) = 0$ (for $r \leq 2M - 3$), and for $c = 1$, $\mathcal{E}_{\mathbf{n}_i}(k,l) = \mathcal{E}_{\mathbf{n}_i,\text{BMVDR}}(k,l)$.

It is worth mentioning that there are other more perceptually relevant methods (see e.g., [42], [43]) determining the ability of a user to correctly localize (before and after applying the binaural spatial filter) concurrent multiple sound sources in reverberant environments than the simple objective performance measures given in Eqs. (45)-(48). In this paper, we focus on the aforementioned simplified instrumental measures.

Note that we use the true ATFs in the constraints of the optimization problems of all competing methods. Therefore, we do not measure the corresponding error measures for the binaural cues of target source since they are always zero, because in all compared methods the distortionless constraints perfectly preserve the binaural cues of the target source.

2) SNR measures: We define the binaural global segmental signal-to-noise-ratio (gsSNR) gain as

$$\text{gsSNR}^{\text{gain}} = \text{gsSNR}^{\text{out}} - \text{gsSNR}^{\text{in}} \text{ dB}, \quad (49)$$

where the gsSNR input and output are defined as

$$\text{gsSNR}^{\text{in}} = \frac{1}{T} \sum_{l=1}^{T} \min\left(\max\left(\text{SNR}^{\text{in}}(l), -20\right), 50\right) \text{ dB}, \quad (50)$$

$$\text{gsSNR}^{\text{out}} = \frac{1}{T} \sum_{l=1}^{T} \min\left(\max\left(\text{SNR}^{\text{out}}(l), -20\right), 50\right) \text{ dB}, \quad (51)$$

respectively, where for the l-th frame, the binaural input signal-to-noise-ratio (SNR) is defined as

$$\text{SNR}^{\text{in}}(l) = 10\log_{10}\left(\frac{\sum_{k=1}^{N} \mathbf{e}^T \tilde{\mathbf{P}}_{\mathbf{x}}(k,l)\mathbf{e}}{\sum_{k=1}^{N} \mathbf{e}^T \tilde{\mathbf{P}}(k,l)\mathbf{e}} \right) \text{ dB}, \quad (52)$$

where $\mathbf{e}^T = [\mathbf{e}_L^T \quad \mathbf{e}_R^T]$, $\mathbf{e}_L^T = [1, 0, \cdots, 0]$ and $\mathbf{e}_R^T = [0, \cdots, 0, 1]$, $\tilde{\mathbf{P}}$ is defined in Eq. (11) and $\tilde{\mathbf{P}}_{\mathbf{x}}$ is similarly defined but it uses as diagonal block matrices the $\mathbf{P}_{\mathbf{x}}$ matrix. The binaural output SNR for the l-th frame, is defined as

$$\text{SNR}^{\text{out}}(l) = 10\log_{10}\left(\frac{\sum_{k=1}^{N} \mathbf{w}^H(k,l)\tilde{\mathbf{P}}_{\mathbf{x}}(k,l)\mathbf{w}(k,l)}{\sum_{k=1}^{N} \mathbf{w}^H(k,l)\tilde{\mathbf{P}}(k,l)\mathbf{w}(k,l)} \right) \text{ dB}, \quad (53)$$

where $\mathbf{w} = [\mathbf{w}_L^T(k,l) \quad \mathbf{w}_R^T(k,l)]^T$. Note that gsSNR$^{\text{out}}$ and gsSNR$^{\text{in}}$ can be seen as average measures of the binaural SNR measures defined in [30]. We also use the frequency-weighted segmental SNR (fwsSNR) [44], [45] to measure the amount of noise suppression at the left and right HA. The fwsSNR gain at the left reference microphone is given by

$$\text{fwsSNR}_L^{\text{gain}} = \text{fwsSNR}_L^{\text{out}} - \text{fwsSNR}_L^{\text{in}} \text{ dB}, \quad (54)$$

where the input and output fwsSNR at the left reference microphone are given by [45]

$$\text{fwsSNR}_L^{\text{in}} = \frac{1}{T} \sum_{l=1}^{T} \min\left(\max\left(\sum_{j=1}^{N_{fb}} g_j \text{SNR}_{j,L}^{\text{in}}, -20\right), 50\right) \text{ dB}, \quad (55)$$
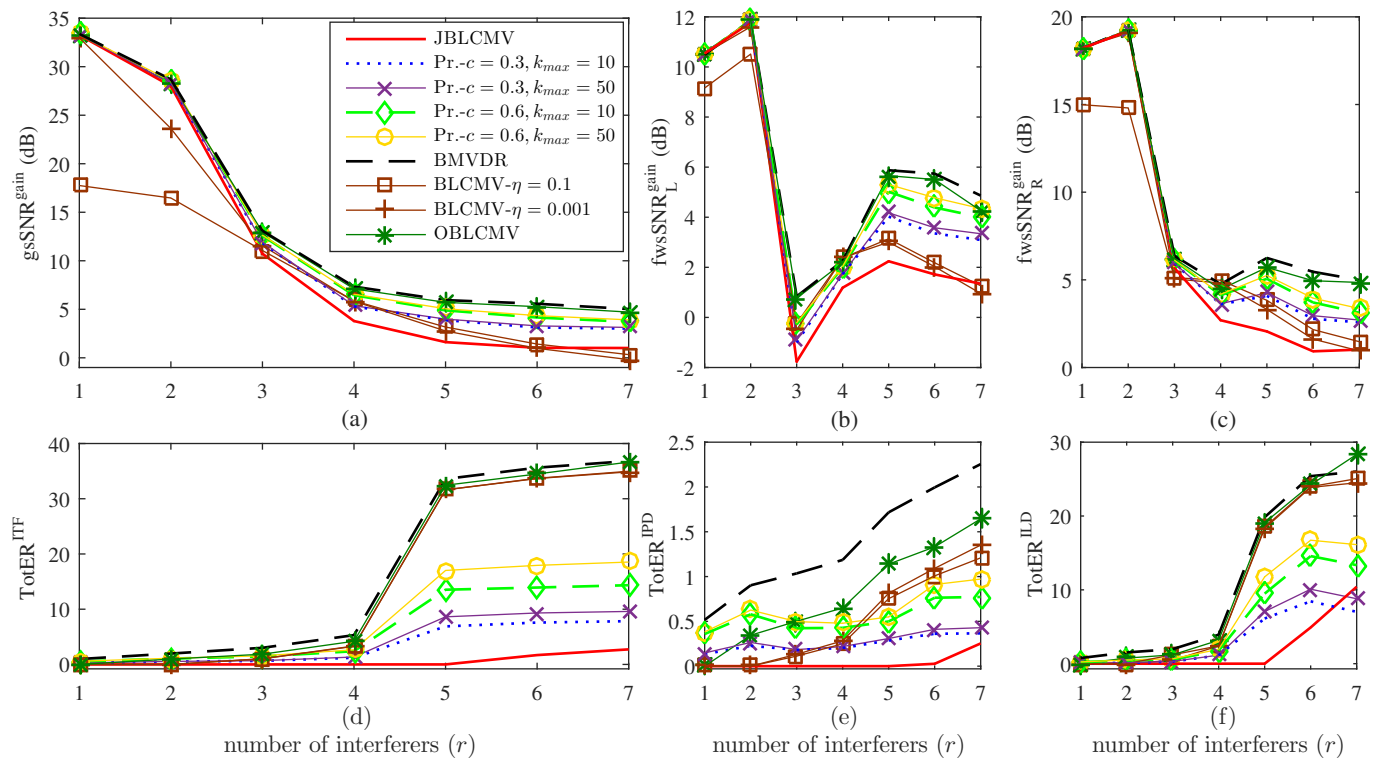
Fig. 3. Anechoic environment: Performance of the competing methods in terms of (a,b,c) noise reduction, (d) ITF error, (e) IPD error, (f) ILD error.

$$\text{fwsSNR}_L^{\text{out}} = \frac{1}{T} \sum_{l=1}^{T} \min\left(\max\left(\sum_{j=1}^{N_{fb}} g_j \text{SNR}_{j,L}^{\text{out}}, -20\right), 50\right) \text{ dB},$$
(56)

where $\text{SNR}_{j,L}^{\text{in}}$ and $\text{SNR}_{j,L}^{\text{out}}$ are the input and output SNRs, respectively, of the j-th frequency band at the left reference microphone. The SNR values of the N_{fb} frequency bands are weighted differently with weights g_j. The ranges and central frequencies of the frequency bands, and the values of $g_j, i = 1, \cdots, N_{fb}$ are selected as described in [46]. The input and output fwsSNR for the right reference microphone are defined similarly to Eqs. (55) and (56), respectively. Note that the noise-only frames are excluded from the evaluation.

C. Results

In the following experiments we evaluate the performance of the proposed and reference methods (i.e., BLCMV [27] with two different values of η, OBLCMV [28], BMVDR [30] and JBLCMV [29], [30]) as a function of the number of simultaneously present interferers, $1 \leq r \leq 7$. For instance, for $r = 1$, only the interferer with index 1 is enabled while all the others are silent. For $r = 2$, only the interferers with indices $1, 2$ are enabled, while the others are silent, and so on. Recall that each method has a different $m_{\max}$, except for the proposed method for $c > 0$ where $m_{\max}$ is difficult to be estimated, as explained in Section V, and, therefore, m is always set to $m = r$. For each of the reference methods and the proposed method in the case of $c = 0$ and if $r > m_{\max}$, we will use in the constraints only the first $m_{\max}$ interferers and the

TABLE II
ANECHOIC ENVIRONMENT: INPUT NOISE LEVELS FOR $r = 1, 2, 3, 4, 5, 6, 7$.

Measure \ r	1	2	3	4	5	6	7
gsSNR$^{\text{in}}$	0.67	−1.26	−2.11	−2.76	−3.58	−4.00	−4.36
fwsSNR$_L^{\text{in}}$	6.88	4.34	0.47	−2.76	−4.84	−5.68	−5.84
fwsSNR$_R^{\text{in}}$	−1.50	−3.93	−4.87	−5.14	−5.26	−5.43	−5.98

last $r - m_{\max}$ will not be preserved. For simplicity, we used the same $c = c_j$, for $j = 1, \cdots, m$, for all interferers in the proposed method. In other words, we assumed that the binaural cues of all interferers are equally important. Moreover, we selected for the adaptive change of $\tau_{(k)}$ the step parameter $\alpha(k_{\max})$ with $k_{\max} \in \{10, 50\}$. In Sections VI-C1, VI-C2 the simulations are carried out without taking into account room acoustics. In Section VI-C3 the simulations are carried out by taking into account room acoustics.

1) SNR & Binaural Cue Preservation: In this section and in Section VI-C2 the evaluation is undertaken in an anechoic environment. The binaural gsSNR$^{\text{in}}$, fwsSNR$_L^{\text{in}}$ and fwsSNR$_R^{\text{in}}$ values for $r = 1, 2, 3, 4, 5, 6$ and 7 are given in Table II.

Figs. 3 and 4 show the comparison of the proposed method (denoted by Pr.$-c = $ value, $k_{\max} = $ value) with the aforementioned reference methods in terms of binaural cue preservation and noise reduction. Note that BMVDR and the JBLCMV are the two extreme special cases of our method which can be denoted as Pr.$-c = 1$ and Pr.$-c = 0$, respectively. However,

Fig. 7. Anechoic environment: Top view of 3D histogram of number of frequency bins that have pairs (k, c) for the proposed method for (a) $k_{\max} = 10$ and (b) $k_{\max} = 50$.

TABLE III
REVERBERANT ENVIRONMENT (OFFICE): INPUT NOISE LEVELS FOR
$r = 1, 2, 3, 4, 5, 6, 7$.

Measure \ r	1	2	3	4	5	6	7
gsSNR$^{\text{in}}$	0.36	-1.68	-2.61	-3.38	-4.07	-4.63	-5.03
fwsSNR$_L^{\text{in}}$	5.99	3.54	0.20	-3.10	-4.98	-5.95	-6.66
fwsSNR$_R^{\text{in}}$	-2.25	-4.59	-5.69	-5.96	-6.12	-6.26	-6.33

stopping criterion) as a function of the simultaneously present interferers, r, of the four configurations of the proposed method that are tested in Figs. 3 and 4. It is clear that the proposed method terminates after 3-4 iterations on average, even for $r = 6, 7 > 2M - 3$. Note that for both tested values of $k_{\max}$, for all frames and frequency bins the proposed method terminated before reaching $k_{\max}$.

Fig. 7 shows a 3D histogram which depicts the statistical termination behaviour of the proposed method. Specifically, the proposed method is evaluated with different c values from 0.1 to 0.9 with a step-size 0.1. For each c value it is evaluated for all numbers of simultaneously present interferers, i.e., for $r = 1, \cdots, 7$ as in Fig 6. Hence, this histogram represents all gathered pair-values (c, k) of all frequency bins for all $r = 1, \cdots, 7$. The pairs (c, k) express the number of iterations (per frequency bin), k, that the proposed method need in order to terminate for a certain initial c. The z-axis, which is depicted with different colors, is the number of frequency bins that are associated with a certain pair (c, k) in the x-y axes. Again we see that, on average, after 3-4 iterations the algorithm terminates for $c = 0.1 : 0.1 : 0.9$.

3) Reverberation: Figs. 8, 9, 10 and 11 show the same experiments as in Figs. 3, 4, 5, and 6, respectively, but this time in a reverberant office environment. The same signals for the interferers and the target are used here. The reverberant BTE IRs are also taken from the database in [40]. Note that, the aforementioned database does not have the reverberant (for the office environment) BTE IRs corresponding to 240º and 300º degrees [40]. Therefore, we used the avalaible angles, 125º, 145º for the 6-th and 7-th interferer, respectively. Moreover, the sources are now placed on a $h = 100$ cm radius circle centered at the origin $(0, 0)$ (center of head) with an elevation of 0º degrees (because only this distance is available for the office environment in [40]). Similarly to the anechoic experiment, the microphone self noise at each microphone is simulated as WGN with $\mathbf{P_V} = \sigma^2 \mathbf{I}$, where $\sigma = 6.1 * 10^{-5}$ which corresponds to an SNR of 50 dB with respect to the target signal at the left reference microphone. The binaural gsSNR$^{\text{in}}$, fwsSNR$_L^{\text{in}}$ and fwsSNR$_R^{\text{in}}$ values for $r = 1, 2, 3, 4, 5, 6$ and 7 are given in Table III.

As it is shown in Figs. 8(a,d) and 9, again the performance of the proposed method is bounded (see Theorem 1) even for $m > 2M - 3$. In Fig. 10 it is clear that the proposed method has

very similar behavior as in Fig. 5, i.e., by increasing $k_{\max}$, the proposed method approaches closer to the boundary. Finally, in Fig. 11 it is shown that the speed of termination is not effected significantly due to reverberation.

VII. CONCLUSION

In this paper we proposed a new multi-microphone iterative binaural noise reduction method. The proposed method is capable of controlling the amount of noise reduction and the accuracy of binaural cue preservation per interferer using a robust methodology. Specifically, the inequality constraints introduced for the binaural cue preservation of the interferers, are selected in such a way that a) the total ITF error is always less or equal than a fraction of the corresponding total ITF error of the BMVDR method, and b) the achieved amount of noise reduction is larger or equal to the one achieved via JBLCMV. Therefore, the proposed method provides the flexibility to the users to parametrize the proposed method according to their needs. Moreover, the proposed method always preserves strictly the binaural cues of the target source. Although the proposed method guarantees a bounded binaural cue preservation accuracy and a bounded amount of noise reduction only for $m \leq 2M - 3$ interferers, it is experimentally demonstrated that is also capable of doing the same for more interferers and terminate in just a few iterations.

APPENDIX

In this section, we show how the optimization problem in Eq. (36) can be equivalently written as a second order cone programming (SOCP) problem. For convenience, we reformulate the optimization problem in Eq. (36) using RATFs instead of ATFs. The left and right RATFs of the i-th interferer are $\bar{\mathbf{b}}_{i,L} = (1/b_{iL})\mathbf{b}_i$ and $\bar{\mathbf{b}}_{i,R} = (1/b_{iR})\mathbf{b}_i$, respectively, while the left and right RATFs of the target are $\bar{\mathbf{a}}_L = (1/a_L)\mathbf{a}$ and $\bar{\mathbf{a}}_R = (1/a_R)\mathbf{a}$, respectively. It is easy to show that the constraints of the optimization problem in Eq. (36) can be equivalently written as

$$\underbrace{\begin{bmatrix} \bar{\mathbf{a}}_L^H & \mathbf{0}^H \\ \mathbf{0}^H & \bar{\mathbf{a}}_R^H \end{bmatrix}}_{\boldsymbol{\Phi}_1^H} \mathbf{w} = \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\mathbf{q}_1}, \tag{57}$$

$$\left| \boldsymbol{\Phi}_{2,i}^H \mathbf{w} \right| \leq \underbrace{\left| \tau_{(k)} \zeta \bar{\mathbf{b}}_{i,R}^H \hat{\mathbf{w}}_{R,(k-1)} \right|}_{\mathbf{q}_{2,i}}, i = 1, \cdots, m, \tag{58}$$
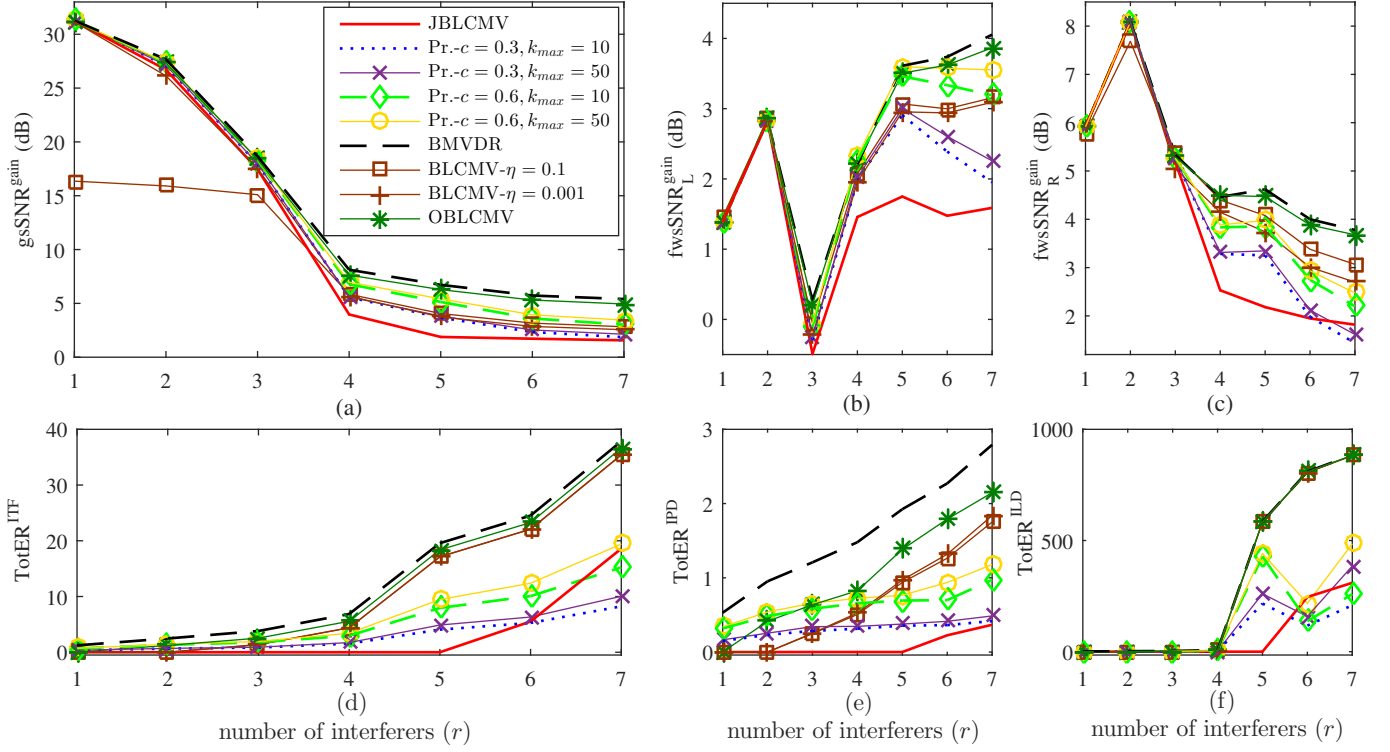
Fig. 8. Reverberant environment (office): Performance of the competing methods in terms of (a,b,c) noise reduction, (d) ITF error, (e) IPD error, (f) ILD error.
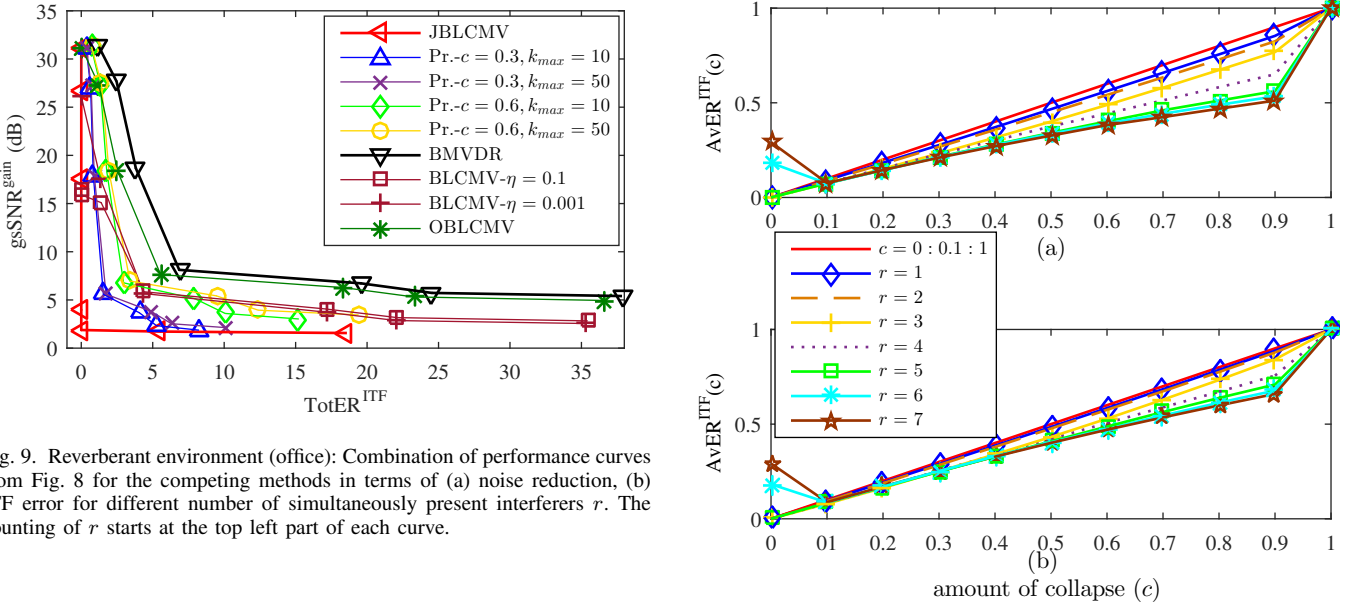

Fig. 9. Reverberant environment (office): Combination of performance curves from Fig. 8 for the competing methods in terms of (a) noise reduction, (b) ITF error for different number of simultaneously present interferers r. The counting of r starts at the top left part of each curve.

Fig. 10. Reverberant environment (office): Average ITF error ratio as a function of c for $1 \leq r \leq 7$ for (a) $k_{\max} = 10$ and (b) $k_{\max} = 50$. The solid line is the c values.

where $\zeta = |\bar{a}_{R,1}^* \bar{b}_{i,L,M}^* - 1|$ (with $\bar{a}_{R,1}^*$ the first element of $\bar{\mathbf{a}}_R^H$ and $\bar{b}_{i,L,M}^*$ is the last element of $\bar{\mathbf{b}}_{i,L}$) and $\boldsymbol{\Phi}_{2,i}$ is the i-th column of the matrix $\boldsymbol{\Phi}_2$ given by

$$\boldsymbol{\Phi}_2 = \begin{bmatrix} \bar{\mathbf{b}}_{1L}, \cdots, \bar{\mathbf{b}}_{mL} \\ -\bar{\mathbf{b}}_{1R}, \cdots, -\bar{\mathbf{b}}_{mR} \end{bmatrix}. \quad (59)$$

Similar to [37], [38], we convert the complex vectors and matrices to real-valued ones, i.e.,

$$\check{\mathbf{w}} = \begin{bmatrix} \check{\mathbf{w}}_L \\ \check{\mathbf{w}}_R \end{bmatrix}, \check{\mathbf{w}}_L = \begin{bmatrix} \text{Re}\{\mathbf{w}_L\} \\ \text{Im}\{\mathbf{w}_L\} \end{bmatrix}, \check{\mathbf{w}}_R = \begin{bmatrix} \text{Re}\{\mathbf{w}_R\} \\ \text{Im}\{\mathbf{w}_R\} \end{bmatrix}, \quad (60)$$

$$\check{\mathbf{a}}_L = \begin{bmatrix} \text{Re}\{\bar{\mathbf{a}}_L\} \\ \text{Im}\{\bar{\mathbf{a}}_L\} \end{bmatrix}, \check{\mathbf{a}}_R = \begin{bmatrix} \text{Re}\{\bar{\mathbf{a}}_R\} \\ \text{Im}\{\bar{\mathbf{a}}_R\} \end{bmatrix} \quad (61)$$

$$\check{\mathbf{a}}_L = \begin{bmatrix} -\text{Im}\{\bar{\mathbf{a}}_L\} \\ \text{Re}\{\bar{\mathbf{a}}_L\} \end{bmatrix}, \check{\mathbf{a}}_R = \begin{bmatrix} -\text{Im}\{\bar{\mathbf{a}}_R\} \\ \text{Re}\{\bar{\mathbf{a}}_R\} \end{bmatrix} \quad (62)$$
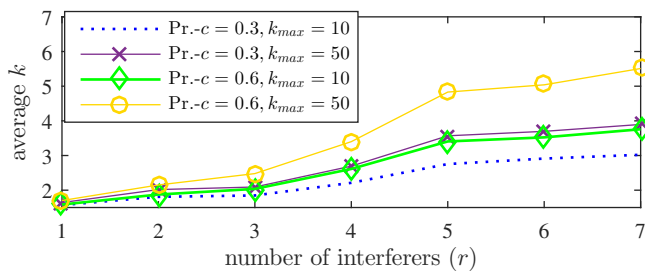
Fig. 11. Reverberant environment (office): Average number of iterations as a function of simultaneously present interferers, r.

$$\breve{\mathbf{b}}_{iL} = \begin{bmatrix} \text{Re}\{\bar{\mathbf{b}}_{iL}\} \\ \text{Im}\{\bar{\mathbf{b}}_{iL}\} \end{bmatrix}, \breve{\mathbf{b}}_{iR} = \begin{bmatrix} \text{Re}\{\bar{\mathbf{b}}_{iR}\} \\ \text{Im}\{\bar{\mathbf{b}}_{iR}\} \end{bmatrix}, \quad (63)$$

$$\breve{\mathbf{b}}_{iL} = \begin{bmatrix} -\text{Im}\{\bar{\mathbf{b}}_{iL}\} \\ \text{Re}\{\bar{\mathbf{b}}_{iL}\} \end{bmatrix}, \breve{\mathbf{b}}_{iR} = \begin{bmatrix} -\text{Im}\{\bar{\mathbf{b}}_{iR}\} \\ \text{Re}\{\bar{\mathbf{b}}_{iR}\} \end{bmatrix}, \quad (64)$$

$$\breve{\mathbf{P}} = \begin{bmatrix} \text{Re}\{\mathbf{P}\} & -\text{Im}\{\mathbf{P}\} \\ \text{Im}\{\mathbf{P}\} & \text{Re}\{\mathbf{P}\} \end{bmatrix}, \tilde{\breve{\mathbf{P}}} = \begin{bmatrix} \breve{\mathbf{P}} & \mathbf{0} \\ \mathbf{0} & \breve{\mathbf{P}} \end{bmatrix}, \quad (65)$$

$$\breve{\boldsymbol{\Phi}}_1 = \begin{bmatrix} \breve{\mathbf{a}}_L & \mathbf{0} & \breve{\mathbf{a}}_L & \mathbf{0} \\ \mathbf{0} & \breve{\mathbf{a}}_R & \mathbf{0} & \breve{\mathbf{a}}_R \end{bmatrix}, \quad (66)$$

$$\breve{\boldsymbol{\Phi}}_2 = \begin{bmatrix} \breve{\mathbf{b}}_{1L}, \cdots, \breve{\mathbf{b}}_{mL} \\ -\breve{\mathbf{b}}_{1R}, \cdots, -\breve{\mathbf{b}}_{mR} \end{bmatrix}, \breve{\boldsymbol{\Phi}}_2 = \begin{bmatrix} \breve{\mathbf{b}}_{1L}, \cdots, \breve{\mathbf{b}}_{mL} \\ -\breve{\mathbf{b}}_{1R}, \cdots, -\breve{\mathbf{b}}_{mR} \end{bmatrix}. \quad (67)$$

Note that $\mathbf{w}^T \tilde{\breve{\mathbf{P}}} \mathbf{w} = ||\tilde{\breve{\mathbf{P}}}^{1/2} \mathbf{w}||_2^2$, where $\tilde{\breve{\mathbf{P}}}^{1/2}$ is the principal square root of $\tilde{\breve{\mathbf{P}}}$. The convex optimization problem in Eq. (36) can be equivalently written as

$$\hat{\tilde{\mathbf{w}}}_{(k)} = \arg \min_{t, \breve{\mathbf{w}}} t \text{ s.t. } \breve{\mathbf{w}}^T \breve{\boldsymbol{\Phi}}_1 = \breve{\mathbf{q}}_1^T,$$

$$||\tilde{\breve{\mathbf{P}}}^{1/2} \breve{\mathbf{w}}||_2 \leq t,$$

$$\left\| \begin{bmatrix} \breve{\boldsymbol{\Phi}}_{2,i}^T \\ \breve{\boldsymbol{\Phi}}_{2,i}^T \end{bmatrix} \breve{\mathbf{w}} \right\|_2 \leq q_{2,i,(k)}, \text{ for } i = 1, \cdots, m, \quad (68)$$

where $\breve{\mathbf{q}}_1^T = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}$, $\breve{\boldsymbol{\Phi}}_{2,i}$ is the i-th column of $\breve{\boldsymbol{\Phi}}_2$, and $\breve{\boldsymbol{\Phi}}_{2,i}$ is the i-th column of $\breve{\boldsymbol{\Phi}}_2$. Note that the problem in Eq. (68) is a standard-form SOCP problem [39].

ACKNOWLEDGMENT

REFERENCES

[1] R. C. Hendriks, T. Gerkmann, and J. Jensen, *DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement: A Survey of the State of the Art*. Morgan & Claypool, 2013.

[2] M. Brandstein and D. Ward (Eds.), *Microphone arrays: signal processing techniques and applications*. Springer, 2001.

[3] K. Eneman et al., "Evaluation of signal enhancement algorithms for hearing instruments," in *EURASIP Europ. Signal Process. Conf. (EU-SIPCO)*, Aug. 2008.

[4] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2230–2244, Sept. 2002.

[5] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Process.*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.

[6] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[7] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 5, pp. 4–24, Apr. 1988.

[8] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[9] J. M. Kates, *Digital hearing aids*. Plural publishing, 2008.

[10] T. Van den Bogaert, T. J. Klasen, L. Van Deun, J. Wouters, and M. Moonen, "Horizontal localization with bilateral hearing aids: without is better than with," *J. Acoust. Soc. Amer.*, vol. 119, no. 1, pp. 515–526, Jan. 2006.

[11] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acoustica*, vol. 86, no. 1, pp. 117–128, 2000.

[12] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beam-forming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech, Language Process.*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.

[13] H. Schmidt, A. B. Baggeroer, W. A. Kuperman, and E. K. Scheer, "Environmentally tolerant beamforming for high-resolution matched field processing: deterministic mismatch," *J. Acoust. Soc. Amer.*, vol. 88, no. 4, Oct. 1990.

[14] S. A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," *ELSEVIER Signal Process.*, vol. 93, no. 12, pp. 3264–3277, Dec. 2013.

[15] R. C. Hendriks, R. Heusdens, U. Kjems, and J. Jensen, "On optimal multichannel mean-squared error estimators for speech enhancement," *IEEE Signal Process. Lett.*, vol. 16, no. 10, pp. 885–888, Oct. 2009.

[16] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Process. Lett.*, vol. 10, no. 7, pp. 204–207, Jul. 2003.

[17] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 845–856, Sep. 2005.

[18] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete Fourier coefficients with generalized gamma priors," *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, no. 6, pp. 1741–1752, Aug. 2007.

[19] P. Vary and R. Martin, *Digital speech transmission: Enhancement, coding and error concealment*. John Wiley & Sons, 2006.

[20] S. Doclo, W. Kellermann, S. Makino, and S. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[21] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Preservation of interaural time delay for binaural hearing aids through multi-channel Wiener filtering based noise reduction," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2005, pp. 29–32.

[22] T. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.

[23] S. Doclo, T. J. Klasen, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions," in *Int. Workshop Acoustic Echo, Noise Control (IWAENC)*, Sep. 2006.

[24] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids," *J. Acoust. Soc. Amer.*, vol. 124, no. 1, pp. 484–497, July 2008.

[25] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Theoretical analysis of linearly constrained multi-channel Wiener filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids," *IEEE Trans. Audio, Speech, Language Process.*, vol. 23, no. 12, Sept. 2015.

[26] E. Hadad, S. Gannot, and S. Doclo, "Binaural linearly constrained minimum variance beamformer for hearing aid applications," in *Int. Workshop Acoustic Signal Enhancement (IWAENC)*, Sep. 2012, pp. 1–4.

[27] E. Hadad, S. Doclo, and S. Gannot, "The binaural LCMV beamformer and its performance analysis," *IEEE Trans. Audio, Speech, Language Process.*, vol. 24, no. 3, pp. 543–558, Mar. 2016.

[28] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Optimal binaural lcmv beamformers for combined noise reduction and binaural cue preservation," in *Int. Workshop Acoustic Signal Enhancement (IWAENC)*, Sep. 2014, pp. 288–292.

[29] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, "Improved multi-microphone noise reduction preserving binaural cues," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2016.