Causal Distortionless Response Beamforming by Alternating Direction Method of Multipliers

Yoshiki Masuyama, Kouei Yamaoka, Yuma Kinoshita and Nobutaka Ono
Tokyo Metropolitan University, Tokyo, Japan
E-mail: masuyama-yoshiki@ed.tmu.ac.jp

***Abstract*—We present a low-latency beamforming method by extending the well-known minimum power distortionless response (MPDR) beamformer. Beamforming has been widely used to extract a target signal arriving from a specific direction. It is often conducted in the time-frequency domain, which causes an algorithmic delay for frame analysis. To reduce this delay, it was proposed to truncate the non-causal components of a spatial filter in the time domain and convolve it with audio mixtures. This method can reduce the algorithmic delay, but the truncated filter was not optimal. To address this problem, we optimize the spatial filter so that the power of the extracted signal is minimized under the distortionless and causality constraints. We further propose a relaxation of the distortionless constraint to improve the extraction performance. The alternating direction method of multipliers (ADMM) is used to solve the exact and relaxed optimization problems. Through numerical experiments, we investigate the performance of the causal MPDR beamforming and demonstrate the effectiveness of the relaxation.**

I. Introduction

The aim of beamforming is to extract a target signal arriving from a specific direction from audio mixtures observed by multiple microphones [1], [2]. In the literature, various beamformers have been presented, such as the minimum power distortionless response (MPDR) beamformer and the minimum variance distortionless response (MVDR) beamformer [3]–[7]. These beamformers can suppress interference signals without distorting the sound arriving from the target direction. They have been successfully applied to automatic speech recognition [8]–[12] and mobile communication [13].

Beamforming and other spatial filtering techniques have been applied to real-time audio applications [14]–[19]. In such applications, we should consider two problems: online estimation of spatial information and reduction in system latency. The first problem has been tackled by several methods, including probabilistic-model-based methods [8]–[10] and deep-neural-network-based methods [20]–[22]. It is thus beyond the scope of this paper. The second problem is to reduce system latency. While computational time has been reduced [20], spatial filtering in the time-frequency (T-F) domain causes an algorithmic delay for short-time Fourier transform (STFT). This delay is due to buffering the signal in order to apply the discrete Fourier transform (DFT) and becomes 128 ms when the window length is 2048 samples with 16 kHz sampling. Such a long delay is, however, unacceptable for some applications including hearing aids, because the delayed auditory feedback and the mismatch between lip movements and sounds disturb communication [23]. Specifically, a tolerable delay is 6 ms for open-fitting hearing aids [24]. We should thus avoid the algorithmic delay of spatial filtering in the T-F domain.

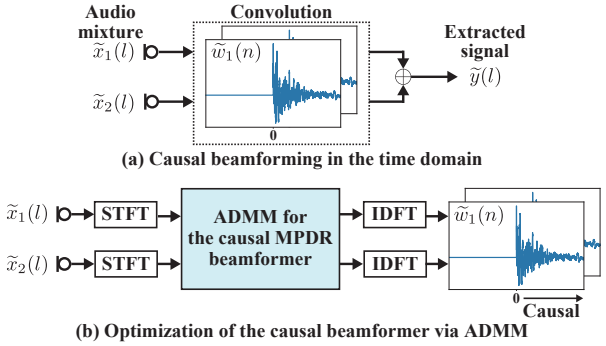

Fig. 1. Block diagram of the proposed causal MPDR beamforming.

To solve this issue, a straightforward approach is to use a shorter window for STFT, but it degrades the extraction performance of spatial filtering [18]. One promising method is to truncate the non-causal components of the spatial filter in the time domain and convolve the truncated filter with audio mixtures [17]. Once the spatial filter is obtained in the T-F domain, this method can reduce the algorithmic delay to an allowable length. Although this method achieved excellent performance with independent vector analysis [17], the directivity and frequency response of the truncated filter are different from those of the original filter designed in the T-F domain. Furthermore, in the case of MPDR beamforming, the truncated filter does not satisfy the distortionless constraint.

In this paper, we propose a causal MPDR beamformer that simultaneously satisfies the distortionless and causality constraints. The spatial filter is optimized to minimize the power of the extracted signal under these constraints. By relaxing the constraints, we further design a quasi-causal beamformer with a slight distortion to improve the extraction performance. These two spatial filters are obtained by solving constrained optimization problems by the alternating direction method of multipliers (ADMM) [25]. They are convolved with the audio mixtures in the time domain, as illustrated in Fig. 1. We investigated the performance of the causal MPDR beamformer and its relaxed version in numerical experiments. The experimental results show the effectiveness of the relaxation.

II. Mathematical Tools

In this section, we briefly explain ADMM and proximity operators since our proposed method is based on them.

A. ADMM

Let us consider an optimization problem with two proper lower semicontinuous convex functions $g(\cdot)$ and $h(\cdot)$:

$$\min_{\boldsymbol{\beta}} \; g(\boldsymbol{\beta}) + h(\boldsymbol{\beta}). \tag{1}$$

The ADMM algorithm [25] for this problem is given by[1]

$$\boldsymbol{\beta}^{[k+1]} = \mathrm{prox}_{g/\rho}(\boldsymbol{\gamma}^{[k]} - \boldsymbol{\zeta}^{[k]}), \tag{2}$$

$$\boldsymbol{\gamma}^{[k+1]} = \mathrm{prox}_{h/\rho}(\boldsymbol{\beta}^{[k+1]} + \boldsymbol{\zeta}^{[k]}), \tag{3}$$

$$\boldsymbol{\zeta}^{[k+1]} = \boldsymbol{\zeta}^{[k]} + \boldsymbol{\beta}^{[k+1]} - \boldsymbol{\gamma}^{[k+1]}, \tag{4}$$

where $\rho > 0$ is a hyperparameter, k is an iteration counter, and $\boldsymbol{\gamma}$ and $\boldsymbol{\zeta}$ are auxiliary variables. One advantage of ADMM is that the two objective functions are handled separately through the proximity operator:

$$\mathrm{prox}_{g}(\boldsymbol{\phi}) = \operatorname*{argmin}_{\boldsymbol{\beta}} \; g(\boldsymbol{\beta}) + \frac{1}{2}\|\boldsymbol{\beta} - \boldsymbol{\phi}\|_2^2, \tag{5}$$

where the existence and uniqueness of the minimizer of the right-hand side are guaranteed [26].

ADMM has been applied to robust beamforming [27]–[30] and beamforming with multiple objective functions [31]. Particularly in [31], ADMM was used to minimize multiple objective functions including the ℓ_2 norm of the frequency-directional second-order derivative of the spatial filter. This objective function reduces the length of the spatial filter in the time domain but does not enforce the causality.

B. Proximity Operator

In this subsection, we show two useful properties of proximity operators. First, if an objective function is separable across two variables, i.e., $g(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = g_1(\boldsymbol{\beta}_1) + g_2(\boldsymbol{\beta}_2)$, its proximity operator can be calculated as follows [26]:

$$\mathrm{prox}_{g}(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2) = (\mathrm{prox}_{g_1}(\boldsymbol{\phi}_1), \mathrm{prox}_{g_2}(\boldsymbol{\phi}_2)). \tag{6}$$

That is, we can separately calculate the proximity operator for each objective function.

Second, if $g(\boldsymbol{\beta})$ can be reformulated as $\widetilde{g}(\mathbf{Q}\boldsymbol{\beta})$ by using a unitary matrix $\mathbf{Q}$, the proximity operator is given by [26]

$$\mathrm{prox}_{g}(\boldsymbol{\phi}) = \mathbf{Q}^{\mathsf{H}} \mathrm{prox}_{\widetilde{g}}(\mathbf{Q}\boldsymbol{\phi}), \tag{7}$$

where $(\cdot)^{\mathsf{H}}$ denotes the Hermitian transpose.

III. Conventional Methods

A. MPDR Beamforming in T-F Domain

Let us denote audio mixtures observed by M microphones as $\mathbf{x}(t,f) = [x_1(t,f), \ldots, x_M(t,f)]^{\mathsf{T}}$ in the T-F domain, where $t = 1, \ldots, T$, $f = 1, \ldots, F$, and $m = 1, \ldots, M$ respectively are the time, frequency, and microphone indices, and $(\cdot)^{\mathsf{T}}$ is the transpose. Beamforming extracts the target signal with a spatial filter $\mathbf{w}(f) = [w_1(f), \ldots, w_M(f)]^{\mathsf{T}}$ as follows:

$$y(t,f) = \mathbf{w}^{\mathsf{H}}(f)\mathbf{x}(t,f), \tag{8}$$

where $y(t,f)$ is the STFT coefficients of the extracted signal.

The MPDR beamformer [3], which minimizes the power of the extracted signal under the distortionless constraint, has been widely used. It can be obtained by solving the following optimization problem:

$$\min_{(\mathbf{w}(1), \ldots, \mathbf{w}(F))} \; \frac{1}{2T} \sum_{t=1}^{T} \sum_{f=1}^{F} |\mathbf{w}^{\mathsf{H}}(f)\mathbf{x}(t,f)|^2 \tag{9a}$$

$$\text{s.t.} \qquad \mathbf{w}^{\mathsf{H}}(f)\mathbf{a}(f) = 1, \; \forall f, \tag{9b}$$

where $\mathbf{a}(f) = [1, a_2(f), \ldots, a_M(f)]^{\mathsf{T}}$ is the relative transfer function (RTF) of the target signal. The optimization problem in (9) can be analytically solved by considering the Karush–Kuhn–Tucker (KKT) condition, and its solution is given by

$$\mathbf{w}(f) = \frac{\mathbf{R}^{-1}(f)\mathbf{a}(f)}{\mathbf{a}^{\mathsf{H}}(f)\mathbf{R}^{-1}(f)\mathbf{a}(f)}, \tag{10}$$

where $\mathbf{R}(f)$ is the spatial covariance matrix (SCM) of the audio mixtures at the fth frequency:

$$\mathbf{R}(f) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}(t,f)\mathbf{x}^{\mathsf{H}}(t,f). \tag{11}$$

B. Low-latency Spatial Filtering via Quasi-causal FIR Filter

When applying beamforming to real-time audio applications, it is crucial to reduce the algorithmic delay. While beamforming is often implemented in the T-F domain because of efficiency, STFT inherently introduces an algorithmic delay of at least the window length.

To reduce this delay, spatial filtering was realized in the time domain by convolving a quasi-causal finite impulse response (FIR) filter with audio mixtures and adding up the results [17]. Assuming that the number of DFT points F is even and the number of non-causal components D is much less than F, a quasi-causal FIR filter can be computed as follows:

$$\widehat{w}_m(n) = \begin{cases} \widetilde{w}_m(n) & (F/2 - D + 1 \le n \le F) \\ 0 & (1 \le n \le F/2 - D) \end{cases}, \tag{12}$$

$$\widetilde{w}_m(n) = (\mathcal{T} \circ \mathcal{F}^{-1})(\overline{\mathbf{w}}_m), \tag{13}$$

$$\mathbf{w}_m = [w_m(1), \ldots, w_m(F)]^{\mathsf{T}}, \tag{14}$$

where $\mathcal{T}(x)(n) = x(n - F/2)$ represents a circular shift of $\mathbf{x}$ with length F, $\mathcal{F}(\cdot)$ is DFT, $\overline{x}$ is the complex conjugate of x, and $n = 1, \ldots, F$ is the sample index.

By using the quasi-causal FIR filter $\widehat{w}_m(n)$, we can extract the target signal in the time domain as follows:

$$\widetilde{y}(l) = \sum_{m=1}^{M} \sum_{n'=-D}^{F/2-1} \widehat{w}_m\left(\frac{F}{2} + 1 + n'\right) \widetilde{x}_m(l - n'), \tag{15}$$

where $\widetilde{y}(l)$ is the time-domain extracted signal, $\widetilde{x}_m(l)$ is the time-domain audio mixture observed at the mth microphone, and $l = 1, \ldots, L$ is the sample index. We stress that the

[1]The ADMM algorithm explained in this paper is simplified to allow a straightforward explanation of the proposed method. It can handle more complicated problems than the problem in (1). We refer the reader to [25] for further details of ADMM.

algorithmic delay for (15) is only D samples, which is much shorter than that of beamforming in the T-F domain.

As one drawback of truncating the non-causal components, the quasi-causal FIR filter does not retain the directivity or the frequency response of the original filter. Furthermore, it does not satisfy the distortionless constraint in (9) even though the original filter is the MPDR beamformer.

IV. Proposed Method

In this section, we present an optimization-based method to realize MPDR beamforming under the quasi-causality constraint. In addition, we propose to relax the distortionless constraint because it is not easy to satisfy the two constraints simultaneously. By using ADMM, we solve the optimization problems to obtain the spatial filters efficiently.

A. Optimization Problem for Causal MPDR Beamforer

To obtain a spatial filter that simultaneously satisfies the distortionless and quasi-causality constraints, we formulate the following optimization problem:

$$\min_{(\mathbf{w}(1),\ldots,\mathbf{w}(F))} \frac{1}{2T}\sum_{t=1}^{T}\sum_{f=1}^{F}|\mathbf{w}^{\mathsf{H}}(f)\mathbf{x}(t,f)|^2 \tag{16a}$$

$$\text{s.t.} \quad \mathbf{w}^{\mathsf{H}}(f)\mathbf{a}(f)=1,\ \forall f, \tag{16b}$$

$$\mathbf{w}_m \in \mathcal{C},\ \forall m, \tag{16c}$$

where $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ are the sets of quasi-causal FIR filters (with D samples of non-causal components) in the frequency and time domains, respectively:

$$\mathcal{C}=\{\mathbf{v}\in\mathbb{C}^F \mid \mathcal{T}\circ\mathcal{F}^{-1}(\overline{\mathbf{v}})\in\widetilde{\mathcal{C}}\}, \tag{17}$$

$$\widetilde{\mathcal{C}}=\{\widetilde{\mathbf{v}}\in\mathbb{C}^F \mid \widetilde{v}(n)=0 \text{ if } 1\le n\le F/2-D\}. \tag{18}$$

Note that $\mathcal{C}$ becomes the set of causal FIR filters when $D=0$, and $\mathcal{C}$ becomes the set of FIR filters when $D=F/2$. In the latter case, the optimization problem in (16) coincides with that for the original MPDR beamformer in (9).

Before solving (16), we show the existence of its solution. Even when $D=0$, the following spatial filter satisfies both the distortionless and quasi-causality constraints simultaneously:

$$w_m(f)=\begin{cases}1 & (m=1)\\ 0 & (m\neq 1)\end{cases}. \tag{19}$$

In addition, the objective function in (16a) is bounded below. Thus, there exists a solution for the optimization problem.

B. Relaxation of Distortionless Constraint

Although a quasi-causal filter satisfying the distortionless constraint can be obtained by solving (16), our preliminary experiment showed that its performance was significantly lower than that of the original MPDR beamformer. To improve the extraction performance, we relax the distortionless constraint:

$$\min_{\substack{(\mathbf{w}(1),\ldots,\mathbf{w}(F),\\ z(1),\ldots,z(F))}} \frac{1}{2T}\sum_{t=1}^{T}\sum_{f=1}^{F}|\mathbf{w}^{\mathsf{H}}(f)\mathbf{x}(t,f)|^2 \tag{20a}$$

$$\text{s.t.} \quad \mathbf{w}^{\mathsf{H}}(f)\mathbf{a}(f)=1+z(f),\ \forall f, \tag{20b}$$

$$z(f)\in\mathbb{R}_+,\ \forall f, \tag{20c}$$

$$\mathbf{w}_m\in\mathcal{C},\ \forall m, \tag{20d}$$

where $z(f)$ is an auxiliary variable related to the gain of the spatial filter. In (20), the gain of the spatial filter for the target direction is allowed to be greater than 1. This may introduce distortion of the extracted signal, but we expect that it is not serious because we minimize the power of the extracted signal.

C. ADMM for Quasi-causal Beamforming

In contrast to the optimization problem for the original MPDR beamformer in (9), the optimization problems in (16) and (20) cannot be solved independently at each frequency owing to the quasi-causality constraint. In addition, the relaxed problem in (20) contains inequality constraints making it difficult to solve analytically. To solve such an optimization problem efficiently, we apply the ADMM algorithm, which allows us to update optimization variables independently for each frequency or microphone.

To solve the optimization problem in (20) by ADMM, we reformulate it with an equivalent problem[2]. Let us concatenate the spatial filter $\mathbf{w}(f)$ and the auxiliary variable $z(f)$ and introduce an augmented RTF as follows:

$$\boldsymbol{\omega}(f)=[w_1(f),\ldots,w_M(f),z(f)]^{\mathsf{T}}, \tag{21}$$

$$\boldsymbol{\alpha}(f)=[1,a_2(f),\ldots,a_M(f),-1]^{\mathsf{T}}. \tag{22}$$

The set of spatial filters $\boldsymbol{\omega}(f)$ that satisfy the constraint in (20b) is represented by

$$\mathcal{A}_f=\{\boldsymbol{\upsilon}\in\mathbb{C}^M \mid \boldsymbol{\upsilon}^{\mathsf{H}}\boldsymbol{\alpha}(f)=1\}. \tag{23}$$

On the basis of these notations, the optimization problems in (20) can be reformulated as

$$\min_{(\boldsymbol{\omega}(1),\ldots,\boldsymbol{\omega}(F))} g(\boldsymbol{\omega}(1),\ldots,\boldsymbol{\omega}(F))+h(\boldsymbol{\omega}(1),\ldots,\boldsymbol{\omega}(F)), \tag{24}$$

where the objective functions are given by

$$\begin{aligned} g(\boldsymbol{\omega}(1),\ldots,\boldsymbol{\omega}(F)) &= \sum_{f=1}^{F} g_f(\boldsymbol{\omega}(f)) \\ &= \sum_{f=1}^{F}\left[\frac{1}{2}\boldsymbol{\omega}^{\mathsf{H}}(f)\boldsymbol{\Sigma}(f)\boldsymbol{\omega}(f)+\iota_{\mathcal{A}_f}(\boldsymbol{\omega}(f))\right], \end{aligned} \tag{25}$$

$$\begin{aligned} &h(\boldsymbol{\omega}(1),\ldots,\boldsymbol{\omega}(F)) \\ &\quad = \sum_{f=1}^{F}\iota_{\mathbb{R}_+}(z(f))+\sum_{m=1}^{M}\iota_{\mathcal{C}}(\mathbf{w}_m), \end{aligned} \tag{26}$$

[2]Through Section IV-C, we only explain the ADMM algorithm for the optimization problems in (20) because that for (16) can be obtained simply by replacing the constraint $z(f)\in\mathbb{R}_+$ with $z(f)=0$.

where $\iota_{\mathcal{Q}}(\cdot)$ is the indicator function with respect to a set $\mathcal{Q}$:

$$\iota_{\mathcal{Q}}(x) = \begin{cases} 0 & (x \in \mathcal{Q}) \\ \infty & (x \notin \mathcal{Q}) \end{cases}. \tag{27}$$

By using the zero vector $\mathbf{0} \in \mathbb{C}^M$, $\mathbf{\Sigma}(f)$ is given by

$$\mathbf{\Sigma}(f) = \begin{pmatrix} (1/T)\sum_{t=1}^{T} \mathbf{x}(t,f)\mathbf{x}^{\mathsf{H}}(t,f) & \mathbf{0} \\ \mathbf{0}^{\mathsf{H}} & 0 \end{pmatrix}. \tag{28}$$

In (24), $g(\cdot)$ consists of the objective function in (20a) and the indicator function corresponding to the constraint in (20b). Meanwhile, $h(\cdot)$ is the sum of the indicator functions corresponding to the constraints in (20c)–(20d). We can solve the optimization problem in (24) by ADMM using the proximity operators for $g(\cdot)$ and $h(\cdot)$.

1) Proximity operator of $g(\cdot)$: The proximity operator of $g(\cdot)$ can be evaluated separately for each frequency because it is separable across all frequencies (see Section II-B). The proximity operator of $g_f(\cdot)$ is defined as

$$\begin{aligned} &\mathrm{prox}_{g_f/\rho}(\boldsymbol{\phi}) \\ &\quad = \underset{\boldsymbol{v}}{\mathrm{argmin}}\ \iota_{\mathcal{A}_f}(\boldsymbol{v}) + \frac{1}{2}\boldsymbol{v}^{\mathsf{H}}\mathbf{\Sigma}(f)\boldsymbol{v} + \frac{\rho}{2}\|\boldsymbol{v}-\boldsymbol{\phi}\|_2^2, \end{aligned} \tag{29}$$

where $\boldsymbol{v}$ is an optimization variable related to $\boldsymbol{\omega}(f)$. The optimization problem in (29) can be reformulated as

$$\min_{\boldsymbol{v}} \frac{1}{2}\boldsymbol{v}^{\mathsf{H}}(\mathbf{\Sigma}(f) + \rho\mathbf{I})\boldsymbol{v} + \rho\boldsymbol{v}^{\mathsf{H}}\boldsymbol{\phi} \tag{30a}$$

$$\text{s.t. } \boldsymbol{v}^{\mathsf{H}}\boldsymbol{\alpha}(f) = 1, \tag{30b}$$

where $\mathbf{I} \in \mathbb{C}^{(M+1)\times(M+1)}$ is the identity matrix. Considering the KKT conditions for the global minimum, we obtain the following linear system:

$$\begin{pmatrix} \mathbf{\Sigma}(f) + \rho\mathbf{I} & \boldsymbol{\alpha}(f) \\ \boldsymbol{\alpha}^{\mathsf{H}}(f) & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{v}^\star \\ \lambda^\star \end{pmatrix} = \begin{pmatrix} \rho\boldsymbol{\phi} \\ 1 \end{pmatrix}, \tag{31}$$

where $\boldsymbol{v}^\star \in \mathbb{C}^{M+1}$ is the solution of the optimization problem in (30), and $\lambda^\star \in \mathbb{C}$ is the KKT multiplier. Since $\mathbf{\Sigma}(f)$ is positive-semidefinite, this KKT system is nonsingular. Hence, the solution to the linear system in (31) can be analytically calculated. Note that KKT matrix does not change at every iteration. That is, we can obtain $\boldsymbol{v}^\star$ by multiplying the inverse of the KKT matrix, which is calculated in advance, by the input of the proximity operator.

2) Proximity operator of $h(\cdot)$: The proximity operator of $h(\cdot)$ can also be evaluated separately for $\iota_{\mathbb{R}_+}(\cdot)$ and $\iota_{\mathcal{C}}(\cdot)$. The proximity operator for $\iota_{\mathbb{R}_+}(\cdot)$ is given by

$$\mathrm{prox}_{\iota_{\mathbb{R}_+}}(\phi) = \max(\mathrm{Re}[\phi], 0), \tag{32}$$

where $\mathrm{Re}[\cdot]$ returns the real part of its input. Meanwhile, since both $\mathcal{T}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ are unitary operators, $\mathrm{prox}_{\iota_{\mathcal{C}}}(\cdot)$ can be reformulated as (see Section II-B)

$$\overline{\mathrm{prox}_{\iota_{\mathcal{C}}}(\boldsymbol{\phi})} = (\mathcal{F}\circ\mathcal{T}^{-1})(\mathrm{prox}_{\iota_{\widetilde{\mathcal{C}}}}((\mathcal{T}\circ\mathcal{F}^{-1})(\overline{\boldsymbol{\phi}}))), \tag{33}$$

where $\mathrm{prox}_{\iota_{\widetilde{\mathcal{C}}}}(\cdot)$ becomes the projection onto the set of quasi-causal FIR filters $\widetilde{\mathcal{C}}$:

$$\mathrm{prox}_{\iota_{\widetilde{\mathcal{C}}}}(\boldsymbol{\phi})(n) = \begin{cases} \phi(n) & (F/2 - D + 1 \le n \le F) \\ 0 & (1 \le n \le F/2 - D) \end{cases}. \tag{34}$$

Algorithm 1 ADMM for (20)

Input: $\boldsymbol{\psi}^{[1]}(f)$, $\boldsymbol{\xi}^{[1]}(f)$, $\mathbf{R}(f)$, $\forall f$, $\rho > 0$
Output: $\boldsymbol{\varphi}_m$, $\forall f$

```
for k = 1, 2, ... do
  for f = 1, ..., F do
    ω^[k+1](f) = prox_{g_f/ρ}(ψ^[k](f) − ξ^[k](f))
  end for
  Split ω^[k+1](f) + ξ^[k](f) into φ_m^[k] and ϑ^[k](f)
  for m = 1, ..., M do
    φ_m^[k+1] = prox_{ι_C}(φ_m^[k])
  end for
  for f = 1, ..., F do
    ϑ^[k+1](f) = prox_{ι_{R_+}}(ϑ^[k](f))
  end for
  ψ^[k+1](f) = [φ_1^[k+1](f), ..., φ_M^[k+1](f), ϑ^[k+1](f)]^T
  for f = 1, ..., F do
    ξ^[k+1](f) = ξ^[k](f) + ω^[k+1](f) − ψ^[k+1](f)
  end for
end for
```

3) Summary of Proposed Algorithm: The proposed algorithm is summarized in Algorithm 1, where $\boldsymbol{\psi}(f)$ and $\boldsymbol{\xi}(f)$ are auxiliary variables for ADMM. Here, $\boldsymbol{\varphi}_m$ is a spatial filter at the mth microphone, and $\vartheta(f)$ is an auxiliary variable related to $z(f)$. In the proposed algorithm, we can update $\boldsymbol{\omega}(f)$ and $\vartheta(f)$ in parallel for each frequency. On the other hand, $\boldsymbol{\varphi}_m$ is updated independently for each microphone. Hence, we do not handle the spatial filter for all frequencies and microphones at the same time. This is desirable for computationally limited applications, including hearing aids.

The procedure for updating $\boldsymbol{\varphi}_m$ in Algorithm 1 is exactly the truncation of the unallowable non-causal components in (12). Although the truncation has been introduced heuristically [17], we show its interpretation as the projection onto the set of quasi-causal filters $\widetilde{\mathcal{C}}$ in a rigorous optimization framework. When truncating the non-causal components of the MPDR beamformer only once, the truncated filters do not satisfy the distortionless constraint. The proposed algorithm, however, can find a quasi-causal FIR filter that satisfies the distortionless constraint or its relaxed version because the convergence of ADMM to the global optimum is guaranteed [25].

Once the spatial filter is obtained using Algorithm 1, we can extract the target signal by (15). Hence, we can extract the target signal with the algorithmic delay of only D samples. In this paper, we focus on reducing the algorithmic delay of MPDR beamforming. Hence, we do not explore the online estimation of RTF or the online optimization of the spatial filter. Implementation of the proposed methods to real-time audio applications will be included in our future work.

V. Experimental Evaluation

To confirm the effectiveness of the proposed methods, we conducted target speech extraction by beamforming with synthesized multichannel audio mixtures. We investigated the effects of the number of iterations, the direction of an interference source, and the number of non-causal components.

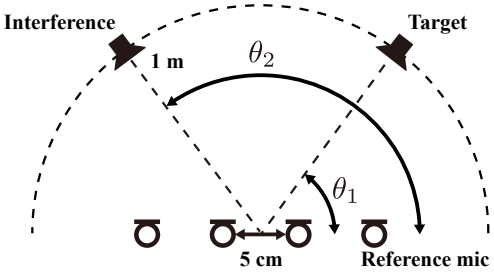

Fig. 2. Spatial arrangements of sound sources and a linear microphone array.

A. Relation between Performance and Number of Iterations

Audio mixtures were generated by convolving room impulse responses (RIRs) to source signals in the evaluation set of the Voice Conversion Challenge (VCC) 2018 dataset [32]. The source signals were resampled at 16 kHz. The half-overlapping Hann window with 2048 samples was used for STFT, and the number of DFT points was 4096. The RIRs were numerically generated by the image method [33] implemented in the `pyroomacoustics` toolbox [34]. The room had dimensions of 5.0 m $\times$ 3.5 m $\times$ 2.5 m, and the reverberation time was randomly sampled from $[0.2, 0.3]$ s. The spatial arrangements of the sound sources and a microphone array are illustrated in Fig. 2, where the four-channel linear microphone array was located at the center of the room. In this experiment, we set $\theta_1 = 45°$ and $\theta_2 = 135°$. The number of allowable non-causal components D was set to 7. This corresponds to the time taken for the sound to propagate from the reference microphone (rightmost one in Fig. 2) to the microphone on the other side.

The proposed methods were compared with the MPDR beamforming in the T-F domain [`MPDR (freq)`], the MPDR beamforming in the time domain by $\widetilde{\mathbf{w}}_m$ [`MPDR (time)`] and its truncated version $\widehat{\mathbf{w}}_m$ (`Truncation`). The proposed methods in (16) and in (20) are abbreviated as `Prop-exact` and `Prop-relax`, respectively. For faster convergence, we set ρ to 0.2 and 0.005 for `Prop-exact` and `Prop-relax`, respectively. We initialized $\boldsymbol{\psi}^{[1]}(f)$ by converting the truncated filter $\widehat{\mathbf{w}}_m$ to the frequency domain and concatenating it with 0.1 as the auxiliary variable related to the gain. We used eigenvalue decomposition to estimate the RTF of the target signal from the SCM of the target signal itself.

The signal-to-noise ratio (SNR) of the extracted signal per iteration is illustrated in Fig. 3. As a result of truncating the non-causal components, `Truncation` resulted in a much lower performance than `MPDR (time)`. The SNR for `Prop-exact` decreased with increasing number of iterations. In ADMM, the quasi-causal filters obtained at the beginning of iterations do not satisfy the distortionless constraint, and they exactly satisfy the constraint only at the convergence point. The decrease in SNR should be due to insufficient degrees of freedom in the spatial filter that simultaneously satisfies both the distortionless and quasi-causality constraints. On the other hand, `Prop-relax` achieved a comparable performance to that of `MPDR (time)` owing to the relaxation of the distortionless constraint.

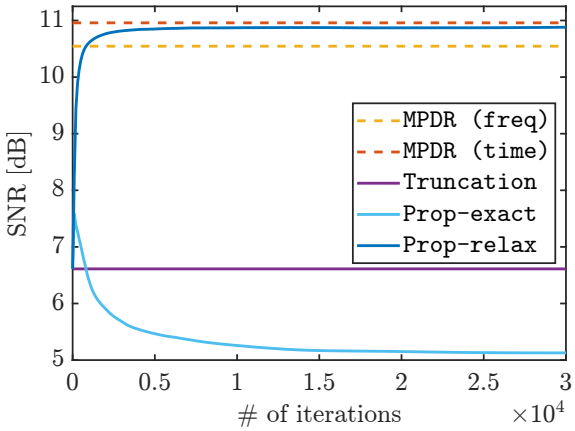

Fig. 3. SNR of the extracted signals averaged over 10 audio mixtures. Solid and dashed lines are for the quasi-causal and non-causal methods, respectively.

B. Effect of Direction of Interference

The position of the interference source affects whether the interference signal observed by the reference microphone reaches the other microphones within D samples. It is important that the interference signal reaches the other microphones in order to suppress the interference signal in the audio mixture observed at the reference microphone. We thus investigated the extraction performance with different directions of $\theta_2 \in \{0°, 30°, 60°, 90°, 120°, 150°, 180°\}$ with various numbers of non-causal components D. For each condition, 10 audio mixtures were evaluated. The number of iterations was set to 20000 in accordance with Fig. 3, and other experimental conditions were the same as described in Section V-A.

Table I shows the SNR of the extracted signals under each condition. When the interference source was close to the reference microphone, i.e., $\theta_2 < 90°$, `Truncation` resulted in a significantly lower performance for $D \leq 10$. This result indicates that the truncated non-causal components were essential for beamforming, especially when the interference signal reached the reference microphone first. On the other hand, `Prop-exact` and `Prop-relax` essentially maintained their extraction performance. We expect that the optimization under the quasi-causality constraint helps low-latency beamforming to be robust to the direction of the interference source. In particular, `Prop-relax` outperformed the other quasi-causal filtering methods under all conditions and numbers of non-causal components.

VI. Conclusions

In this paper, we presented the causal MPDR beamformer that minimizes the power of the extracted signal under both the distortionless and causality constraints. We further proposed to relax these constraints to improve the extraction performance. Experimental results confirmed the robustness of the proposed method to the interference direction and the effectiveness of the relaxation. Our future work includes the application of the proposed methods to hearing aids by integrating them with online RTF estimation methods.

TABLE I
SNR OF THE EXTRACTED SIGNALS WITH DIFFERENT DIRECTIONS OF INTERFERENCE θ_2 AND NUMBERS OF NON-CAUSAL COMPONENTS D.

		SNR [dB]						
D	θ_2	0°	30°	60°	90°	120°	150°	180°
-	MPDR (freq)	10.0	8.5	8.8	11.9	13.1	11.3	12.1
-	MPDR (time)	10.2	8.3	8.7	12.3	13.2	12.0	12.8
0	Truncation	0.6	-1.0	0.3	4.7	4.4	3.4	4.2
	Prop-exact	3.4	2.0	3.1	4.8	6.0	6.3	6.6
	Prop-relax	**7.5**	**5.5**	**7.2**	**11.7**	**12.3**	**11.7**	**12.4**
10	Truncation	0.3	-1.9	1.9	6.0	7.1	4.7	6.1
	Prop-exact	4.3	2.5	3.3	5.1	6.2	6.5	6.9
	Prop-relax	**9.7**	**7.3**	**7.6**	**11.9**	**12.5**	**11.9**	**12.6**
100	Truncation	4.6	3.2	3.0	7.5	8.4	7.0	6.9
	Prop-exact	5.1	3.4	4.9	6.6	7.9	7.7	8.4
	Prop-relax	**10.6**	**8.5**	**9.4**	**12.9**	**13.7**	**12.2**	**13.1**

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Numbers JP20H00613, JP21J21371, and JST CREST Grant Number JPMJCR19A3, Japan.

REFERENCES

[1] M. Brandstein and D. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer, 2001.

[2] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, Springer, 2008.

[3] H. L. V. Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, Wiley, 2004.

[4] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 260–276, Feb. 2010.

[5] 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.

[6] E. Warsitz and R. Haeb-Umbach, "Blind acoustic beamforming based on generalized eigenvalue decomposition," *IEEE Trans. Audio, Speech, Language Process*, vol. 15, no. 5, pp. 1529–1539, 2007.

[7] T. Nakatani and K. Kinoshita, "Simultaneous denoising and dereverberation for low-latency applications using frame-by-frame online unified convolutional beamformer," in *Proc. INTERSPEECH*, Sept. 2019, pp. 111–115.

[8] T. Higuchi, N. Ito, T. Yoshioka, and T. Nakatani, "Robust MVDR beamforming using time-frequency masks for online/offline ASR in noise," in *Proc. IEEE Int. Conf. on Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 5210–5214.

[9] T. Higuchi, N. Ito, S. Araki, T. Yoshioka, M. Delcroix, and T. Nakatani, "Online MVDR beamformer based on complex Gaussian mixture model with spatial prior for noise robust ASR," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 4, pp. 780–793, Apr. 2017.

[10] K. Shimada, Y. Bando, M. Mimura, K. Itoyama, K. Yoshii, and T. Kawahara, "Unsupervised speech enhancement based on multichannel NMF-informed beamforming for noise-robust automatic speech recognition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 5, pp. 960–971, May 2019.

[11] H. Erdogan, J. R. Hershey, S Watanabe, M. I. Mandel, and J. Le Roux, "Improved MVDR beamforming using single-channel mask prediction networks," in *Proc. INTERSPEECH*, Sept. 2016, pp. 1981–1985.

[12] T. Ochiai, S. Watanabe, T. Hori, J. R. Hershey, and X. Xiao, "Unified architecture for multichannel end-to-end speech recognition with neural beamforming," *IEEE J. Selected Topics Signal Process.*, vol. 11, no. 8, pp. 1274–1288, Dec. 2017.

[13] J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*, Springer, 2006.

[14] T. Taniguchi, N. Ono, A. Kawamura, and S. Sagayama, "An auxiliary-function approach to online independent vector analysis for real-time blind source separation," in *Proc. Workshop Hands-free Speech Commun. Microphone Arrays (HSCMA)*, May 2014, pp. 107–111.

[15] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, *Handbook on Array Processing and Sensor Networks*, chapter Acoustic beamforming for hearing aid applications, pp. 269–302, Wiley, 2010.

[16] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[17] M. Sunohara, C. Haruta, and N. Ono, "Low-latency real-time blind source separation for hearing aids based on time-domain implementation of online independent vector analysis with truncation of non-causal components," in *Proc. IEEE Int. Conf. on Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 216–220.

[18] T. Ueda, T. Nakatani, R. Ikeshita, K. Kinoshita, S. Araki, and S. Makino, "Low latency online blind source separation based on joint optimization with blind dereverberation," in *Proc. IEEE Int. Conf. on Acoust., Speech Signal Process. (ICASSP)*, June 2021, pp. 506–510.

[19] Y. Luo, C. Han, N. Mesgarani, E. Ceolini, and S. Liu, "FaSNet: Low-latency adaptive beamforming for multi-microphone audio processing," in *Proc. IEEE Autom. Speech Recognit. Underst. Workshop (ASRU)*, 2019.

[20] T. Higuchi, K. Kinoshita, N. Ito, S. Karita, and T. Nakatani, "Frame-by-frame closed-form update for mask-based adaptive MVDR beamforming," in *Proc. IEEE Int. Conf. on Acoust., Speech Signal Process. (ICASSP)*, 2018, pp. 531–535.

[21] S. Horiguchi, Y. Fujita, and K. Nagamatsu, "Block-online guided source separation," in *Proc. IEEE Spoken Lang. Tech. Workshop (SLT)*, Jan. 2021, pp. 236–242.

[22] C. Boeddeker, H. Erdogan, T. Yoshioka, and R. Haeb-Umbach, "Exploring practical aspects of neural mask-based beamforming for far-field speech recognition," in *Proc. IEEE Int. Conf. on Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 6697–6701.

[23] J. Agnew and J. M. Thornton, "Just noticeable and objectionable group delays in digital hearing aids," *J. Am. Acad. Audiol.*, vol. 11, no. 6, pp. 330–336, June 2000.

[24] M. A. Stone, B. C. J. Moore, K. Meisenbacher, and R. P. Derleth, "Tolerable hearing aid delays. V. Estimation of limits for open canal fittings," *Ear Hear.*, vol. 29, no. 4, pp. 601–617, Aug. 2008.

[25] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Now Publishers Inc., 2010.

[26] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.

[27] X. Jiang, J. Chen, H. C. So, and X. Liu, "Large-scale robust beamforming via ℓ_∞-minimization," *IEEE Trans. Signal Process.*, vol. 66, no. 14, pp. 3824–3837, July 2018.

[28] W. Fan, J. Liang, G. Yu, H. C. So, and J. Li, "Robust capon beamforming via ADMM," in *Proc. IEEE Int. Conf. on Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4345–4349.

[29] W. Liao, Z. Luo, I. Merks, and T. Zhang, "An effective low complexity binaural beamforming algorithm for hearing aids," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2015, pp. 1–5.

[30] W. Pu, J. Xiao, T. Zhang, and Z. Luo, "A penalized inequality-constrained minimum variance beamformer with applications in hearing aids," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2017, pp. 175–179.

[31] R. Sato, K. Niwa, and N. Harada, "Function designable beamformer based on probabilistic assumptions on filter and its auxiliary variables," in *Proc. IEEE Int. Conf. on Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 780–784.

[32] J. Lorenzo-Trueba, J. Yamagishi, T. Toda, D. Saito, F. Villavicencio, T. Kinnunen, and Z. Ling, "The voice conversion challenge 2018: Promoting development of parallel and nonparallel methods," in *Odyssey*, June 2018, pp. 195–202.

[33] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

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