Microphone Array Wiener Post Filtering Using Monotone Operator Splitting

Kenta Niwa[1] *Member, IEEE*, Hironobu Chiba[1] *Non-Member, IEEE*, Noboru Harada[1] *Senior Member, IEEE*, Guoqiang Zhang[2] *Member, IEEE*, and W. Bastiaan Kleijn[3] *Fellow, IEEE*

***Abstract*—For array-based acoustic source enhancement, variants of multi-channel Wiener filters are commonly used. The approach includes a Wiener post-filter that requires the simultaneous estimation of the power spectral density (PSD) of the target source and of noise sources for each time-frame. Conventional methods generally do not exploit *prior knowledge*, such as sparsity of the source, in solving this simultaneous estimation problem. We show that, for common scenarios, the simultaneous PSD estimation with consideration of prior knowledge can be formulated as a convex optimization problem with linear constraints. We use monotone operator splitting (MOS) to solve the constrained optimization problem. Our experiments confirm that the proposed method improves the accuracy of the noise PSD estimation, and that the resulting enhanced target signal is of higher quality.**

***Index Terms*—Microphone array, Wiener post-filter, power spectral density (PSD) estimation, convex optimization, monotone operator splitting**

I. Introduction

SPEECH interfaces are commonly used for applications such as teleconferencing, manipulating a navigation system while driving a car, and communicating with a clerk robot. These environments are often noisy, which may severely degrade system performance. Hence, technology to improve the fidelity of the target sound source, which may be located in any direction, must be used. This has led to extensive research into microphone array based enhancement methods over the last few decades.

A commonly used and practical microphone array based enhancement approach is multi-channel Wiener filtering, or, more generally, beamforming with Wiener post-filtering (e.g., [1], [2]). The Wiener post-filter requires the simultaneous estimation of the power spectral density (PSD) of a target source and of that of surrounding noises. In classic papers [3], [4], toy problems were studied, where a coherent target source and incoherent/diffuse background noise were mixed in the observed signals. More recently, effective methods of low computational complexity have been developed for practical situations [5], [6], [7]. Representative examples are (i) speech distortion weighted multichannel Wiener filter (SDW-MWF) [8], [9] (ii) the PSD-estimation-in-beamspace method [10] and its extensions [11], [12], which assumes that coherent interference noises and background noises are mixed in the observation and (iii) the method of [13], where both the PSD of late reverberation and background noise are estimated.

Although the recent methods are effective even in practical situations, further improvement of the PSD estimation remains possible. Existing methods do not exploit *prior knowledge* about the PSD. For example, the sparsity of the target source PSD in the short time Fourier transform (STFT)-domain and the stationarity of the background noise over short time intervals were not considered before.

In this paper, we propose a method for accurately estimating PSDs by exploiting prior knowledge derived from reasonable assumptions, such as sparsity of the target source PSD. The estimation procedure can often be based on a convex cost function (e.g., an L_1 norm). By taking several constraints into account, such as the non-negativity of the PSDs, the overall cost function can be cast into the form of a primal optimization problem consisting of a sum of convex functions with several linear equality and/or inequally constraints. Even if the primal problem is complicated, it can generally be reformulated as a minimization problem in the form of a sum of convex functions. It can then be solved with monotone operator splitting (MOS) techniques (e.g., [14], [15]), where the optimal values of the variables are found by iterating over a set of simple problems. To confirm the feasibility of the proposed procedure, the algorithm will be investigated in the context of real-world signals captured in reverberant rooms.

This paper is organized as follows. In Sec. II, the problem formulation and conventional solutions are introduced. Our proposed method for improving PSD estimation accuracy is then presented in Sec. III. The effectiveness of the proposed method is further investigated in Sec. IV. Finally, we conclude the paper in Sec. V.

II. Conventional Method

We first formulate the problem in Sec. II-A. After that, a state-of-the-art conventional solution is described in Sec. II-B.

A. Problem Formulation

We consider $M\ (\geq 2)$ microphone signals in the STFT domain $\mathbf{x} : \mathbb{J}\times\mathbb{Z} \rightarrow \mathbb{C}^M$, where $\mathbb{J}$ is a set of discrete frequencies. We denote the frequency index by ω and the frame-time index by τ. The signals x are the sum of contributions by a target source $s : \mathbb{J} \times \mathbb{Z} \rightarrow \mathbb{C}$, K noise point sources (coherent

Manuscript received 23, December, 2019; revised 26, April, 2020; accepted ?? June, 2020.

[1]: NTT Media Intelligence Laboratories, NTT Corporation, Japan
[2]: University of Technology Sydney, Australia
[3]: Victoria University of Wellington, New Zealand

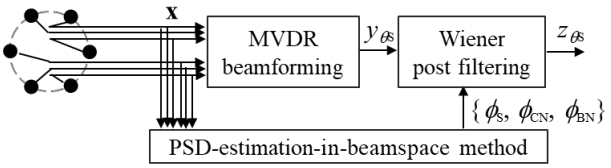

Fig. 1. Signal flow in conventional methods

noise) $\nu_k : \mathbb{J} \times \mathbb{Z} \to \mathbb{C}$, and an incoherent background noise $\boldsymbol{\epsilon} : \mathbb{J} \times \mathbb{Z} \to \mathbb{C}^M$:

$$\mathbf{x}(\omega,\tau) = \mathbf{h}_{\mathrm{S}}(\omega)s(\omega,\tau) + \sum_{k=1}^{K} \mathbf{h}_{\mathrm{CN},k}(\omega)\nu_k(\omega,\tau) + \boldsymbol{\epsilon}(\omega,\tau), \tag{1}$$

where $\mathbf{h}_{\mathrm{S}} : \mathbb{J} \to \mathbb{C}^M$ and $\mathbf{h}_{\mathrm{CN},k} : \mathbb{J} \to \mathbb{C}^M$ are the transfer functions between the target and the k-th noise point source and the microphones, respectively. The direction of arrival (DOA) of the target source θ_{S} is assumed to be known. However, the noise information, such as the noise source DOA, the number of noise point sources, and background noise level are assumed unknown.

When the DOA of the target source is known, a beamforming filter $\mathbf{w}_{\theta_{\mathrm{S}}} : \mathbb{J} \to \mathbb{C}^M$ is applied to emphasize the target source as:

$$y_{\theta_{\mathrm{S}}}(\omega,\tau) = \mathbf{w}_{\theta_{\mathrm{S}}}^{\mathrm{T}}(\omega)\mathbf{x}(\omega,\tau), \tag{2}$$

where the superscript $^{\mathrm{T}}$ denotes transposition. The minimum variance distortionless response (MVDR) method [16] is a method to design $\mathbf{w}_{\theta_{\mathrm{S}}}$ and it is optimal in the L_2 sense. A Wiener filter is then applied to the beamforming output as

$$z_{\theta_{\mathrm{S}}}(\omega,\tau) = \frac{\phi_{\mathrm{S}}(\omega,\tau)}{\phi_{\mathrm{S}}(\omega,\tau) + \phi_{\mathrm{N}}(\omega,\tau)} y_{\theta_{\mathrm{S}}}(\omega,\tau), \tag{3}$$

where $\phi_{\mathrm{S}} : \mathbb{J} \times \mathbb{Z} \to \mathbb{R}$ and $\phi_{\mathrm{N}} : \mathbb{J} \times \mathbb{Z} \to \mathbb{R}$ denote the nonnegative PSD of the target and that of other noises included in the beamforming output $y_{\theta_{\mathrm{S}}}$. In this paper, any other spatial pre-processing was not used. By applying the inverse STFT to $z_{\theta_{\mathrm{S}}}$, we obtain an output signal in the time domain.

Our goal is to estimate the nonnegative PSDs $\{\phi_{\mathrm{S}}, \phi_{\mathrm{N}}\}$ for each time-frame by analyzing the observed signals $\mathbf{x}$.

B. Conventional Method

We now describe a conventional method for estimating the PSDs. As modeled in (1), various noises are randomly mixed in the observed microphone signals. In such situations, it is particularly difficult to estimate the noise PSD ϕ_{N} because coherent noise and incoherent (diffuse) background noise are mixed. Only prior knowledge on their spatial and temporal characteristics is available. In this case, it is more straightforward to estimate the PSD of coherent noises ϕ_{CN} and that of background noise ϕ_{BN} separately.

An approach for separate estimation of the PSDs $\{\phi_{\mathrm{S}}, \phi_{\mathrm{CN}}, \phi_{\mathrm{BN}}\}$ is introduced in [11]. The basic signal flow is shown in Fig. 1. In this approach, the Wiener filtering is modified to

$$z_{\theta_{\mathrm{S}}}(\omega,\tau) = \frac{\phi_{\mathrm{S}}(\omega,\tau)}{\phi_{\mathrm{S}}(\omega,\tau) + \underbrace{\phi_{\mathrm{CN}}(\omega,\tau) + \phi_{\mathrm{BN}}(\omega,\tau)}_{\approx\phi_{\mathrm{N}}(\omega,\tau)}} y_{\theta_{\mathrm{S}}}(\omega,\tau), \tag{4}$$

where the noise PSD is approximated as a sum of the PSDs of the point noise sources and that of background noise.

To estimate ϕ_{CN}, the PSD-estimation-in-beamspace method [10], [12] can be used. In this method, multiple fixed beamforming outputs, whose focusing/null directions are different, are used to analyze observed signals. One of the beamformers $\mathbf{w}_{\theta_{\mathrm{S}}}$ is focused on the target. Assuming that $K+1$ sound sources are independent in the STFT domain, the PSDs of $L\,(\geq 2)$ beamforming outputs $\boldsymbol{\phi}_{\mathrm{BF}} : \mathbb{J} \times \mathbb{Z} \to \mathbb{R}^L$ and the PSDs of the sound sources grouped into $N\,(\geq 2)$ directions $\boldsymbol{\phi}_{\mathrm{GS}} : \mathbb{J} \times \mathbb{Z} \to \mathbb{R}^N$ are linearly related as

$$\underbrace{\begin{bmatrix} \phi_{\mathrm{BF},1}(\omega,\tau) \\ \vdots \\ \phi_{\mathrm{BF},L}(\omega,\tau) \end{bmatrix}}_{\boldsymbol{\phi}_{\mathrm{BF}}(\omega,\tau)} = \underbrace{\begin{bmatrix} D_{1,1}(\omega) & \dots & D_{1,N}(\omega) \\ \vdots & \ddots & \vdots \\ D_{L,1}(\omega) & \dots & D_{L,N}(\omega) \end{bmatrix}}_{\mathbf{D}(\omega)} \underbrace{\begin{bmatrix} \phi_{\mathrm{GS},1}(\omega,\tau) \\ \vdots \\ \phi_{\mathrm{GS},N}(\omega,\tau) \end{bmatrix}}_{\boldsymbol{\phi}_{\mathrm{GS}}(\omega,\tau)} \tag{5}$$

where $\mathbf{D} : \mathbb{J} \to \mathbb{R}^{L \times N}$ is the power sensitivity of the l-th beamformer to the n-th direction, $D_{l,n}(\omega)$. The elements in $\mathbf{D}$ can be calculated by averaging sensitivities, which are obtained by convolving beamformers with array manifold vectors, for each angular spaces/frequency band. The inverse problem of (5) can be solved by

$$\hat{\boldsymbol{\phi}}_{\mathrm{GS}}(\omega,\tau) = \left[\mathbf{D}^{\dagger}(\omega)\boldsymbol{\phi}_{\mathrm{BF}}(\omega,\tau)\right]_{+}, \tag{6}$$

where † and $[\cdot]_{+}$ are the (pseudo) inverse of a matrix and an operator that sets non-negative elements of a vector to zero, respectively. For stably estimating the PSDs, $N \leq L$ is usually assumed. To reduce computational requirements, the frequency bins can be grouped into frequency bands. When the first source $(n=1)$ is selected to be the target direction θ_{S} and point noise sources are in other directions $(n=2,\dots,N)$, the PSDs of the target source and that of incoherent and coherent noises can be estimated by

$$\hat{\phi}_{\mathrm{S}}(\omega,\tau) = \hat{\phi}_{\mathrm{GS},1}(\omega,\tau), \;\; \hat{\phi}_{\mathrm{CN}}(\omega,\tau) = \sum_{n=2}^{N} \hat{\phi}_{\mathrm{GS},n}(\omega,\tau). \tag{7}$$

Next, a method to estimate ϕ_{BN} is explained. When the background noise can be assumed to be stationary, then minimum statistics tracking [17] in the estimated target PSD is an effective approach towards this goal:

$$\hat{\phi}_{\mathrm{BN}}(\omega,\tau) = \min\{\hat{\phi}_{\mathrm{S}}(\omega,\tau-\Gamma),\dots,\hat{\phi}_{\mathrm{S}}(\omega,\tau)\}, \tag{8}$$

where Γ is the time interval to calculate minimum statistics.

Although good experimental results were obtained for the described conventional approaches [10], [11], further performance improvement is possible. The conventional methods do not exploit *prior knowledge* on $\{\phi_{\mathrm{S}}, \phi_{\mathrm{CN}}, \phi_{\mathrm{BN}}\}$. For example, the following attributes can additionally be exploited: (i) the PSD of the target source and those of the noise

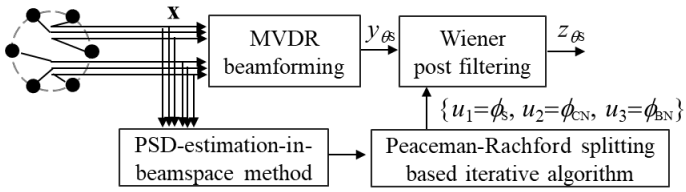

Fig. 2. Signal flow in our proposed method

point sources are sparse and (ii) the sum of the nonnegative PSDs $\{\phi_{\rm S}, \phi_{\rm CN}, \phi_{\rm BN}\}$ equals the PSD of the target focusing beamforming output $\phi_{y_{\theta_{\rm S}}}$. As will be shown in the following sections, by exploiting this prior knowledge, we can obtain improved estimates of the PSDs $\{\phi_{\rm S}^{\rm new}, \phi_{\rm CN}^{\rm new}, \phi_{\rm BN}^{\rm new}\}$ using the values $\{\hat{\phi}_{\rm S}, \hat{\phi}_{\rm CN}, \hat{\phi}_{\rm BN}\}$ already obtained in (7) and (8).

III. Proposed Method

A cost function for accurate PSD estimation and its reformulation are introduced in Sec. III-A and III-B, respectively. In Sec. III-C, a fast PSD estimation algorithm based on monotone operator splitting (MOS) is derived.

A. New Cost Formulation for Accurate PSD Estimation

Our basic strategy for estimating the PSDs is to represent prior knowledge by convex, closed and proper (CCP) functions and linear equality or inequality constraints. Although imposing prior knowledge makes the formulation of the overall cost function complex, the resulting constrained convex minimization problem can be solved by using MOS techniques.

The proposed method assumes that the PSDs estimates obtained in Sec. II-B provide a rough first estimate of the PSDs. We then obtain the signal flow shown in Fig. 2. The estimated PSDs will be denoted as $\mathbf{u} = [u_1, u_2, u_3]^{\rm T}$, where $(u_1 = \phi_{\rm S}^{\rm new}(\omega,\tau))$ is the PSD of the target source, $(u_2 = \phi_{\rm CN}^{\rm new}(\omega,\tau))$ describes the interfering noises, and $(u_3 = \phi_{\rm BN}^{\rm new}(\omega,\tau))$ is the background noise. In the following, we will estimate the PSD values separately for each time-frame and each frequency band. The arguments $\{\omega,\tau\}$ will be omitted to simplify notation.

Our method is based on the notion that it is possible to find cost functions that represent prior knowledge. We provide three examples, which we will use in our PSD estimation method:

Prior knowledge 1: Our assumption is that the PSDs $\{u_1, u_2, u_3\}$ have values that are near a set of known values $\{\hat{\phi}_{\rm S}, \hat{\phi}_{\rm CN}, \hat{\phi}_{\rm BN}\}$. For this purpose we will use the values derived in Sec. II-B. A natural cost function is:

$$F_1(\mathbf{u}) = \frac{\psi_1}{2}(u_1 - \hat{\phi}_{\rm S})^2 + \frac{\psi_2}{2}(u_2 - \hat{\phi}_{\rm CN})^2 + \frac{\psi_3}{2}(u_3 - \hat{\phi}_{\rm BN})^2, \tag{9}$$

where the ψ_i are positive coefficients that facilitate adjustment of the cost balance. For later notation, variables/parameters are summarized by

$$\hat{\boldsymbol{\phi}} = \begin{bmatrix} \hat{\phi}_{\rm S} \\ \hat{\phi}_{\rm CN} \\ \hat{\phi}_{\rm BN} \end{bmatrix}, \quad \boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & \psi_3 \end{bmatrix}. \tag{10}$$

Prior knowledge 2: The PSD of the target source (speech) and that of interfering noises (speech) are sparse in the STFT domain. This can be achieved by using an L_1 norm:

$$F_2(\mathbf{u}) = \mu_1|u_1| + \mu_2|u_2|, \tag{11}$$

where the μ_i are positive weight coefficients. In our experiments in Sec. IV, either the target or the interfering noise is assumed to be sparse, i.e., either μ_1 or μ_2 has non zero value.

Prior knowledge 3: The sum of estimated PSDs equals the PSD of beamforming output $\phi_{y_{\theta_{\rm S}}}$. This assumption is used in the modified Wiener filter design (4), but it is not considered in the PSD estimation procedure (5). Thus, we will enforce the linear constraint

$$u_1 + u_2 + u_3 = \phi_{y_{\theta_{\rm S}}}. \tag{12}$$

The optimization problem is now defined: our objective is to find the primal variable $\mathbf{u}$ that minimizes a sum of F_1 of (9) and F_2 of (11) subject to the constraint (12). This is a linearly constrained convex minimization problem.

Anticipating the optimization approach employed in section III-C, we use lifting: we introduce an auxiliary variable $\mathbf{v}$ in combination with a constraint that the pairs (u_1, u_2) and (v_1, v_2) are equal:

$$\inf_{\mathbf{u},\mathbf{v}} F_1(\mathbf{u}) + F_2(\mathbf{v}) \quad \text{s.t. } \mathbf{A}\mathbf{u} = \mathbf{v}, \mathbf{B}\mathbf{u} = \mathbf{c}, \mathbf{u} \succeq \mathbf{0}, \tag{13}$$

where the curled inequality symbol $\succeq$ is used to denote generalized inequality, i.e., it represents componentwise inequality between vectors [18] and the parameters are given by

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{B} = [1,1,1], \quad \mathbf{c} = \phi_{y_{\theta_{\rm S}}}.$$

We note that (13) is an example cost function. In general our approach can be used to solve the linearly constrained minimization of a sum of convex cost functions.

B. Dual problem formulation

It is convenient to solve the dual problem of (13) because it is an *unconstrained* convex minimization problem. We first formulate the Lagrangian of the constrained optimization problem. The objective of the dual problem is to find the dual variables $\{\mathbf{p}, \mathbf{q}, \mathbf{r}\}$ that correspond to the supremum of

$$\sup_{\substack{\mathbf{p},\mathbf{q},\\ \mathbf{r}\succeq\mathbf{0}}} \inf_{\mathbf{u},\mathbf{v}} \Big[F_1(\mathbf{u}) + F_2(\mathbf{v}) + \langle \mathbf{p}, \mathbf{v} - \mathbf{A}\mathbf{u} \rangle + \langle \mathbf{q}, \mathbf{c} - \mathbf{B}\mathbf{u} \rangle - \langle \mathbf{r}, \mathbf{u} \rangle \Big], \tag{14}$$

where $\langle , \rangle$ denotes the inner product of vectors and the dual variable $\mathbf{r}$ is constrained to be nonnegative because the last term $\langle \mathbf{r}, \mathbf{u} \rangle$ corresponds to an inequality constraint. The dual problem (14) can be reformulated as:

$$\inf_{\substack{\mathbf{p},\mathbf{q},\\ \mathbf{r}\succeq\mathbf{0}}} F_1^\star(\mathbf{A}^{\rm T}\mathbf{p} + \mathbf{B}^{\rm T}\mathbf{q} + \mathbf{r}) - \mathbf{q}^{\rm T}\mathbf{c} + F_2^\star(-\mathbf{p}), \tag{15}$$

where $F_i^\star$ denotes the convex conjugate of F_i:

$$F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p}+\mathbf{B}^\mathrm{T}\mathbf{q}+\mathbf{r}) = \sup_{\mathbf{u}}\Big(\big\langle \mathbf{u}, \mathbf{A}^\mathrm{T}\mathbf{p}+\mathbf{B}^\mathrm{T}\mathbf{q}+\mathbf{r}\big\rangle - F_1(\mathbf{u})\Big), \tag{16}$$

$$F_2^\star(-\mathbf{p}) = \sup_{\mathbf{v}}\Big(\langle \mathbf{v}, -\mathbf{p}\rangle - F_2(\mathbf{v})\Big). \tag{17}$$

Since convex conjugate functions $F_i^\star$ are guaranteed to be CCP, the dual problem is an unconstrained minimization problem of a sum of CCP functions as in (15).

The nonnegative constraint on the dual variable $\mathbf{r}$ can be represented by an indicator function. Hence, the problem (15) can be reformulated as:

$$\inf_{\boldsymbol{\xi}}\ F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p}+\mathbf{B}^\mathrm{T}\mathbf{q}+\mathbf{r}) - \mathbf{q}^\mathrm{T}\mathbf{c}+F_2^\star(-\mathbf{p})+\iota_{(\mathbf{r}\succeq\mathbf{0})}(\mathbf{r}), \tag{18}$$

where the dual variables are summarized by $\boldsymbol{\xi}=[\mathbf{p}^\mathrm{T}, \mathbf{q}^\mathrm{T}, \mathbf{r}^\mathrm{T}]^\mathrm{T}$ and $\iota_{(\mathbf{r}\succeq\mathbf{0})}(\mathbf{r})$ is an indicator function to guarantee nonnegativity of $\mathbf{r}$:

$$\iota_{(\mathbf{r}\succeq\mathbf{0})} = \begin{cases} 0 & (\mathbf{r}\succeq\mathbf{0}) \\ +\infty & (\text{otherwise}) \end{cases}. \tag{19}$$

Since (18) is composed of CCP functions, it is a convex optimization problem. However, it is difficult to find update rules for the variables to reduce the overall cost in (18) because it includes two different convex conjugate functions and an indicator function.

C. Monotone Operator Splitting-based Variable Update Rule

Monotone operator splitting (MOS) is a method for finding update rules when a cost function can be written as the sum of two relatively simple cost functions. It leads to a set of interlaced update rules that correspond to the simple component cost functions. We will use Peaceman-Rachford (P-R) splitting [19] with a Newton method, which is a particular form of MOS.

MOS is a natural approach to finding update rules for the problem (18). The overall cost function (18), is first decomposed into a sum of two CCP functions G_1 and G_2:

$$\inf_{\boldsymbol{\xi}}\ G_1(\boldsymbol{\xi}) + G_2(\boldsymbol{\xi}), \tag{20}$$

where

$$G_1(\boldsymbol{\xi}) = F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p}+\mathbf{B}^\mathrm{T}\mathbf{q}+\mathbf{r}) - \mathbf{q}^\mathrm{T}\mathbf{c}, \tag{21}$$

$$G_2(\boldsymbol{\xi}) = F_2^\star(-\mathbf{p}) + \iota_{(\mathbf{r}\succeq\mathbf{0})}(\mathbf{r}). \tag{22}$$

We denote the subdifferential operators for G_1 and G_2 as $T_1(\boldsymbol{\xi}) = \partial G_1(\boldsymbol{\xi})$ and $T_2(\boldsymbol{\xi}) = \partial G_2(\boldsymbol{\xi})$, respectively. Note that these operators are monotone as G_1 and G_2 are CCP functions. Maximizing (20) corresponds to finding a fixed point of (20):

$$\mathbf{0} \in T_1(\boldsymbol{\xi}) + T_2(\boldsymbol{\xi}), \tag{23}$$

where $\in$ reflects that its output can be multi-valued when F_i includes non-differentiable points. The monotone operators T_i are

$$T_1(\boldsymbol{\xi}) = \partial G_1(\boldsymbol{\xi}) = \begin{bmatrix} \mathbf{A}\partial_p F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p}+\mathbf{B}^\mathrm{T}\mathbf{q}+\mathbf{r}) \\ \mathbf{B}\partial_q F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p}+\mathbf{B}^\mathrm{T}\mathbf{q}+\mathbf{r}) - \mathbf{c} \\ \partial_r F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p}+\mathbf{B}^\mathrm{T}\mathbf{q}+\mathbf{r}) \end{bmatrix}, \tag{24}$$

$$T_2(\boldsymbol{\xi}) = \partial G_2(\boldsymbol{\xi}) = \begin{bmatrix} -\partial_p F_2^\star(-\mathbf{p}) \\ \mathbf{0} \\ \partial_r \iota_{(\mathbf{r}\succeq\mathbf{0})}(\mathbf{r}) \end{bmatrix}. \tag{25}$$

Before introducing P-R splitting, we define the resolvent operator R_i and the Cayley operator C_i of T_i, as summarized in e.g. [15]. For real-time PSD estimation, it is desirable to accelerate the optimization process of (20). Towards this aim, the metric in the operators R_i, C_i is adjusted each iteration. The basic metric in a conventional R_i, C_i is Euclidean, with a convergence rate that resembles first-order gradient descent. However, if the metric is linearly generalized, the convergence rate can benefit from the attributes of a second-order Newton method. The generalized resolvent and Cayley operators are defined by:

$$R_i = (\mathrm{Id} + \mathbf{M}^{-1}T_i)^{-1}, \tag{26}$$

$$\begin{aligned} C_i &= (\mathrm{Id} - \mathbf{M}^{-1}T_i)(\mathrm{Id} + \mathbf{M}^{-1}T_i)^{-1} \\ &= 2(\mathrm{Id} + \mathbf{M}^{-1}T_i)^{-1} - (\mathrm{Id} + \mathbf{M}^{-1}T_i)(\mathrm{Id} + \mathbf{M}^{-1}T_i)^{-1} \\ &= 2R_i - \mathrm{Id}, \end{aligned} \tag{27}$$

where Id denotes the identity operator, $(\cdot)^{-1}$ denotes the inverse operator, and $\mathbf{M}$ is a positive definite matrix to accelerate convergence speed (see next subsection). We use a block-diagonal matrix because we aim to separate the update procedure for each variable:

$$\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{M}_p^{-1} & & \boldsymbol{O} \\ & \mathbf{M}_q^{-1} & \\ \boldsymbol{O} & & \mathbf{M}_r^{-1} \end{bmatrix}, \tag{28}$$

where $\{\mathbf{M}_p^{-1}, \mathbf{M}_q^{-1}, \mathbf{M}_r^{-1}\}$ are positive definite matrices. Their design for a fast convergence rate will be discussed later.

We can now derive the P-R splitting method. We first reformulate the fixed point condition (23) as

$$\begin{aligned} &\mathbf{0} \in \mathbf{M}^{-1}T_1(\boldsymbol{\xi}) + \mathbf{M}^{-1}T_2(\boldsymbol{\xi}), \\ &\mathbf{0} \in (\mathrm{Id} + \mathbf{M}^{-1}T_2)(\boldsymbol{\xi}) - (\mathrm{Id} - \mathbf{M}^{-1}T_1)(\boldsymbol{\xi}), \end{aligned} \tag{29}$$

Let us define $\boldsymbol{\xi}'$ by $\boldsymbol{\xi} \in R_1(\boldsymbol{\xi}')$. Then, (29) can be written as

$$\begin{aligned} &\mathbf{0} \in (\mathrm{Id} + \mathbf{M}^{-1}T_2)R_1(\boldsymbol{\xi}') - (\mathrm{Id} - \mathbf{M}^{-1}T_1)R_1(\boldsymbol{\xi}') \\ &\mathbf{0} \in (\mathrm{Id} + \mathbf{M}^{-1}T_2)R_1(\boldsymbol{\xi}') - C_1(\boldsymbol{\xi}'), \\ &\mathbf{0} \in R_1(\boldsymbol{\xi}') - R_2C_1(\boldsymbol{\xi}'), \\ &\mathbf{0} \in \frac{1}{2}(C_1 + \mathrm{Id})(\boldsymbol{\xi}') - \frac{1}{2}(C_2 + \mathrm{Id})C_1(\boldsymbol{\xi}'), \end{aligned} \tag{30}$$

which implies that the fixed point condition can be written as

$$\boldsymbol{\xi}' \in C_2C_1(\boldsymbol{\xi}') \tag{31}$$

The Cayley operator is a non-expansive operator [14]. If C_1 and C_2 are contractive, then the Banach fixed-point theorem shows that (31) specifies a Picard sequence that converges to the fixed point. We then have a recursive update rule that cycles through the two Cayley operators.

It is convenient to rewrite the iteration C_2C_1 in more elementary terms. We can then write the iterations in terms of primal variables $\{\mathbf{u},\mathbf{v}\}$ and dual variables $\{\mathbf{p},\mathbf{q},\mathbf{r}\}$. Using the result (27), we can write (31) as four iterative steps:

$$\boldsymbol{\xi}^{t+1/4} = R_1(\boldsymbol{\xi}^t), \tag{32}$$

$$\boldsymbol{\xi}^{t+1/2} = C_1(\boldsymbol{\xi}^t) = 2\boldsymbol{\xi}^{t+1/4} - \boldsymbol{\xi}^t, \tag{33}$$

$$\boldsymbol{\xi}^{t+3/4} = R_2(\boldsymbol{\xi}^{t+1/2}), \tag{34}$$

$$\boldsymbol{\xi}^{t+1} = C_2(\boldsymbol{\xi}^{t+1/2}) = 2\boldsymbol{\xi}^{t+3/4} - \boldsymbol{\xi}^{t+1/2}, \tag{35}$$

where t is the iteration step index. The resolvent mappings (32) and (34) are nontrivial because R_i includes the subdifferential of a convex conjugate function, cf. (24) and (25). Appendix A shows that (32) can be implemented using an alternating primal-dual variable update rule as

$$\begin{aligned}\mathbf{u}^{t+1} = \arg\min_{\mathbf{u}}\Big(&F_1(\mathbf{u}) + \frac{1}{2}\left\langle \mathbf{M}_p^{-1}(\mathbf{A}\mathbf{u}-\widetilde{\mathbf{p}}^t), \mathbf{A}\mathbf{u}-\widetilde{\mathbf{p}}^t\right\rangle \\ &+ \frac{1}{2}\left\langle \mathbf{M}_q^{-1}(\mathbf{B}\mathbf{u}-\mathbf{c}-\widetilde{\mathbf{q}}^t), \mathbf{B}\mathbf{u}-\mathbf{c}-\widetilde{\mathbf{q}}^t\right\rangle \\ &+ \frac{1}{2}\left\langle \mathbf{M}_r^{-1}(\mathbf{u}-\widetilde{\mathbf{r}}^t), \mathbf{u}-\widetilde{\mathbf{r}}^t\right\rangle\Big),\end{aligned} \tag{36}$$

$$\widetilde{\mathbf{p}}^{t+1/4} = \widetilde{\mathbf{p}}^t - \mathbf{A}\mathbf{u}^{t+1}, \tag{37}$$

$$\widetilde{\mathbf{q}}^{t+1/4} = \widetilde{\mathbf{q}}^t - (\mathbf{B}\mathbf{u}^{t+1}-\mathbf{c}), \tag{38}$$

$$\widetilde{\mathbf{r}}^{t+1/4} = \widetilde{\mathbf{r}}^t - \mathbf{u}^{t+1}, \tag{39}$$

where $\widetilde{\ }$ indicates linear mappings $\widetilde{\mathbf{p}} = \mathbf{M}_p\mathbf{p}$, $\widetilde{\mathbf{q}} = \mathbf{M}_q\mathbf{q}$, and $\widetilde{\mathbf{r}} = \mathbf{M}_r\mathbf{r}$. By combining the dual variable update (37)-(39) with (33), the dual variable update for each dual variable is given by

$$\widetilde{\mathbf{p}}^{t+1/2} = 2\widetilde{\mathbf{p}}^{t+1/4} - \widetilde{\mathbf{p}}^t = \widetilde{\mathbf{p}}^t - 2\mathbf{A}\mathbf{u}^{t+1}, \tag{40}$$

$$\widetilde{\mathbf{q}}^{t+1/2} = 2\widetilde{\mathbf{q}}^{t+1/4} - \widetilde{\mathbf{q}}^t = \widetilde{\mathbf{q}}^t - 2(\mathbf{B}\mathbf{u}^{t+1}-\mathbf{c}), \tag{41}$$

$$\widetilde{\mathbf{r}}^{t+1/2} = 2\widetilde{\mathbf{r}}^{t+1/4} - \widetilde{\mathbf{r}}^t = \widetilde{\mathbf{r}}^t - 2\mathbf{u}^{t+1}. \tag{42}$$

Using the second resolvent operator (34) is given by

$$\mathbf{v}^{t+1} = \arg\min_{\mathbf{v}}\left(F_2(\mathbf{v}) + \frac{1}{2}\left\langle \mathbf{M}_p^{-1}(\mathbf{v}+\widetilde{\mathbf{p}}^{t+1/2}), \mathbf{v}+\widetilde{\mathbf{p}}^{t+1/2}\right\rangle\right), \tag{43}$$

$$\widetilde{\mathbf{p}}^{t+3/4} = \widetilde{\mathbf{p}}^{t+1/2} + \mathbf{v}^{t+1}, \tag{44}$$

$$\widetilde{\mathbf{q}}^{t+1} = \widetilde{\mathbf{q}}^{t+1/2}, \tag{45}$$

$$\widetilde{r}_i^{t+3/4} = \begin{cases} \widetilde{r}_i^{t+1/2} & (\text{if } \widetilde{r}_i^{t+1/2} \geq 0, \{1,2,3\} \in i) \\ 0 & (\text{otherwise}) \end{cases}, \tag{46}$$

where the derivation is explained in Appendix A. By combining the dual variable update (44)-(46) with (35), the dual variable updates for the dual variables are given by

$$\widetilde{\mathbf{p}}^{t+1} = 2\widetilde{\mathbf{p}}^{t+3/4} - \widetilde{\mathbf{p}}^{t+1/2} = \widetilde{\mathbf{p}}^{t+1/2} + 2\mathbf{v}^{t+1}, \tag{47}$$

$$\widetilde{r}_i^{t+1} = \begin{cases} 2\widetilde{r}_i^{t+1/2} - \widetilde{r}_i^{t+1/2} = \widetilde{r}_i^{t+1/2} & (\text{if } \widetilde{r}_i^{t+1/2} \geq 0, \{1,2,3\} \in i) \\ 2\cdot 0 - \widetilde{r}_i^{t+1/2} = -\widetilde{r}_i^{t+1/2} & (\text{otherwise}) \end{cases}. \tag{48}$$

Algorithm 1 summarises the optimization procedure (36), (40)-(42), (43), (47), (48). The procedure is repeated for each time-frame and frequency band. A small fixed number of iterations T can be used.

Algorithm 1 P-R splitting based PSD estimation algorithm

Initialization of $\widetilde{\mathbf{p}}^0, \widetilde{\mathbf{q}}^0, \widetilde{\mathbf{r}}^0$ for each frequency band
for $t = 0, \ldots, T-1$ **do**
$\mathbf{u}^{t+1} = \arg\min_{\mathbf{u}}\Big(F_1(\mathbf{u}) + \frac{1}{2}\left\langle \mathbf{M}_p^{-1}(\mathbf{A}\mathbf{u}-\widetilde{\mathbf{p}}^t), \mathbf{A}\mathbf{u}-\widetilde{\mathbf{p}}^t\right\rangle + \frac{1}{2}\left\langle \mathbf{M}_q^{-1}(\mathbf{B}\mathbf{u}-\mathbf{c}-\widetilde{\mathbf{q}}^t), \mathbf{B}\mathbf{u}-\mathbf{c}-\widetilde{\mathbf{q}}^t\right\rangle + \frac{1}{2}\left\langle \mathbf{M}_r^{-1}(\mathbf{u}-\widetilde{\mathbf{r}}^t), \mathbf{u}-\widetilde{\mathbf{r}}^t\right\rangle\Big)$
$\widetilde{\mathbf{p}}^{t+1/2} = \widetilde{\mathbf{p}}^t - 2\mathbf{A}\mathbf{u}^{t+1}$
$\widetilde{\mathbf{q}}^{t+1} = \widetilde{\mathbf{q}}^t - 2\left(\mathbf{B}\mathbf{u}^{t+1}-\mathbf{c}\right)$
$\widetilde{\mathbf{r}}^{t+1/2} = \widetilde{\mathbf{r}}^t - 2\mathbf{u}^{t+1}$
$\mathbf{v}^{t+1} = \arg\min_{\mathbf{v}}\left(F_2(\mathbf{v}) + \frac{1}{2}\left\langle \mathbf{M}_p^{-1}(\mathbf{v}+\widetilde{\mathbf{p}}^{t+1/2}), \mathbf{v}+\widetilde{\mathbf{p}}^{t+1/2}\right\rangle\right)$
$\widetilde{\mathbf{p}}^{t+1} = \widetilde{\mathbf{p}}^{t+1/2} + 2\mathbf{v}^{t+1}$
for $i = 1, \ldots, 3$ **do**
$\widetilde{r}_i^{t+1} = \begin{cases} \widetilde{r}_i^{t+1/2} & (\text{if } \widetilde{r}_i^{t+1/2} \geq 0) \\ -\widetilde{r}_i^{t+1/2} & (\text{otherwise}) \end{cases}$
end for
end for

D. Design of $\mathbf{M}$ *Matrix for Fast Convergence*

We discuss a method to design the positive definite matrices in (28). To facilitate real-time implementation, a suitable selection of $\{\mathbf{M}_p^{-1}, \mathbf{M}_q^{-1}, \mathbf{M}_r^{-1}\}$ is important. In the following, we summarize the derivations provided in Appendix B and C. Let the operator $\mathbf{M}^{-1}T_i$ satisfy

$$\sigma_{\mathrm{LB},i}\|\boldsymbol{\xi}-\boldsymbol{\xi}'\|_2 \leq \|\mathbf{M}^{-1}T_i(\boldsymbol{\xi}) - \mathbf{M}^{-1}T_i(\boldsymbol{\xi}')\|_2 \leq \sigma_{\mathrm{UB},i}\|\boldsymbol{\xi}-\boldsymbol{\xi}'\|_2, \tag{49}$$

where $\{\sigma_{\mathrm{LB},i}, \sigma_{\mathrm{UB},i}\} \in [0,\infty]$ and their value range is dependent on $\mathbf{M}^{-1}$. Then, the convergence rate is given by

$$\|\,\boldsymbol{\xi}^t - \boldsymbol{\xi}^*\,\|_2 \leq (\eta_1\eta_2)^t\,\|\,\boldsymbol{\xi}^0 - \boldsymbol{\xi}^*\,\|_2, \tag{50}$$

where $\boldsymbol{\xi}^*$ is the fixed point of $\boldsymbol{\xi}$ and $\eta_i \in [0,1]$ is defined by

$$\eta_i = \sqrt{1 - \frac{4\sigma_{\mathrm{LB},i}}{(1+\sigma_{\mathrm{UB},i})^2}}. \tag{51}$$

(50) indicates that fast convergence will be achieved by modifying $\{\sigma_{\mathrm{LB},i}, \sigma_{\mathrm{UB},i}\}$ such that $\eta_i\,(i=1,2)$ approaches zero. It is clear from (51) that the optimal value for η_i is obtained when $\sigma_{\mathrm{LB},i} = \min(\sigma_{\mathrm{UB},i}, \frac{1}{4}(1+\sigma_{\mathrm{UB},i})^2)$. This means that $\sigma_{\mathrm{LB},i} = \sigma_{\mathrm{UB},i} = \frac{1}{4}(1+\sigma_{\mathrm{UB},i})^2$ only if $\sigma_{\mathrm{UB},i} = 1$ and the contraction factor η_i is then equal to 0. For $0 \leq \sigma_{\mathrm{UB},i} < 1$ or $\sigma_{\mathrm{UB},i} > 1$, the optimal contraction ratio results when $\sigma_{\mathrm{LB},i} = \sigma_{\mathrm{UB},i}$. Thus, the contraction ratio η_i satisfies

$$0 \leq \sqrt{1 - \frac{4\sigma_{\mathrm{UB},i}}{(1+\sigma_{\mathrm{UB},i})^2}} \leq \eta_i \leq 1. \tag{52}$$

We conclude that optimal contraction ratio, i.e., $\eta_i = 0$ is obtained when

$$\sigma_{\mathrm{LB},i} = 1, \quad \sigma_{\mathrm{UB},i} = 1. \tag{53}$$

Substituting (53) into (49) illustrates the meaning of (53). We then have

$$\|\mathbf{M}^{-1}T_i(\boldsymbol{\xi}) - \mathbf{M}^{-1}T_i(\boldsymbol{\xi}')\|_2 = \|\boldsymbol{\xi}-\boldsymbol{\xi}'\|_2. \tag{54}$$

TABLE I
EXPERIMENTAL PARAMETERS

Sampling frequency	16 kHz
FFT analyzing window length	21.3 ms
# of microphones, M	4
# of beamformers, L	3 ($L{=}N$)
# of noise point sources, K	2, 4, 6
Length of observed signal	14.0 sec
# of iteration times, T	5
# of frames in minimum statistics, Γ	62
Weight coefficients, $\mathbf{\Psi}$	$(\psi_1, \psi_2, \psi_3){=}(1.0, 1.0, 2.0)$
Weight coefficients, μ_i	$\mu_1{=}[\hat{\phi}_{\mathrm{CN}}(\omega,\tau) - \hat{\phi}_{\mathrm{S}}(\omega,\tau)]_+$, $\mu_2 = [\hat{\phi}_{\mathrm{S}}(\omega,\tau) - \hat{\phi}_{\mathrm{CN}}(\omega,\tau)]_+$

(54) indicates that a semi-optimal design is to select $\mathbf{M}^{-1}$ such that it locally inverts the linearized T_i at the current variable value $\boldsymbol{\xi}^t$.

Although the overall problem (23) is composed of two monotone operators T_1 and T_2, we will try to select $\mathbf{M}$ such that $\mathbf{M}^{-1}T_1 \approx \mathrm{Id}$. This is motivated by T_2 in (25) being composed of an L_1 norm and an indicator function and its linearity in a subdifferential domain is unaffected by multiplication with $\mathbf{M}^{-1}$. From the (9), subdifferential of F_1 can be written as

$$\partial F_1(\mathbf{u}) = \mathbf{\Psi}(\mathbf{u} - \hat{\boldsymbol{\phi}}) \tag{55}$$

and its inverse operator is given by

$$\partial F_1^{-1}(\mathbf{u}) = \mathbf{\Psi}^{-1}(\mathbf{u}) + \hat{\boldsymbol{\phi}}. \tag{56}$$

A basic property of the subdifferential of a convex conjugate function is $\partial F_1^\star = \partial F_1^{-1}$ [20]. Hence, T_1 in (24) is equal to

$$\begin{aligned} T_1(\boldsymbol{\xi}) &= \begin{bmatrix} \mathbf{A}\partial_p F_1^{-1}(\mathbf{A}^{\mathrm{T}}\mathbf{p}{+}\mathbf{B}^{\mathrm{T}}\mathbf{q}{+}\mathbf{r}) \\ \mathbf{B}\partial_q F_1^{-1}(\mathbf{A}^{\mathrm{T}}\mathbf{p}{+}\mathbf{B}^{\mathrm{T}}\mathbf{q}{+}\mathbf{r}) - \mathbf{c} \\ \partial_r F_1^{-1}(\mathbf{A}^{\mathrm{T}}\mathbf{p}{+}\mathbf{B}^{\mathrm{T}}\mathbf{q}{+}\mathbf{r}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}\mathbf{\Psi}^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{p} + \mathbf{A}(\mathbf{\Psi}^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{q}+\mathbf{r}) + \hat{\boldsymbol{\phi}}) \\ \mathbf{B}\mathbf{\Psi}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{q} + \mathbf{B}(\mathbf{\Psi}^{-1}(\mathbf{A}^{\mathrm{T}}\mathbf{p}{+}\mathbf{r}) + \hat{\boldsymbol{\phi}}) - \mathbf{c} \\ \mathbf{\Psi}^{-1}\mathbf{r} + \mathbf{\Psi}^{-1}(\mathbf{A}^{\mathrm{T}}\mathbf{p}{+}\mathbf{B}^{\mathrm{T}}\mathbf{q}) + \hat{\boldsymbol{\phi}} \end{bmatrix}. \end{aligned} \tag{57}$$

To satisfy (54) by choosing a good $\mathbf{M}^{-1}$ and recalling that $\boldsymbol{\xi}{=}[\mathbf{p}^{\mathrm{T}}, \mathbf{q}^{\mathrm{T}}, \mathbf{r}^{\mathrm{T}}]^{\mathrm{T}}$, we select following positive definite matrices:

$$\mathbf{M}_p^{-1} = (\mathbf{A}\mathbf{\Psi}^{-1}\mathbf{A}^{\mathrm{T}})^{-1} \approx (\mathbf{A}^{\mathrm{T}})^\dagger \mathbf{\Psi}\mathbf{A}^\dagger, \tag{58}$$

$$\mathbf{M}_q^{-1} = (\mathbf{B}\mathbf{\Psi}^{-1}\mathbf{B}^{\mathrm{T}})^{-1} \approx (\mathbf{B}^{\mathrm{T}})^\dagger \mathbf{\Psi}\mathbf{B}^\dagger, \tag{59}$$

$$\mathbf{M}_r^{-1} = (\mathbf{\Psi}^{-1})^{-1} = \mathbf{\Psi}. \tag{60}$$

IV. EXPERIMENTS

To verify the practical feasibility of the proposed method, experiments were conducted using signals recorded in a real environment with reverberation and background noise. We first describe the experimental setup and then the results.

A. Experimental Setup

The experimental setup, including the array structure and impulse response measurement points, is depicted in Fig. 3. The radius of the microphone array was 0.03 m. The array observation was replicated by convolving dry speech

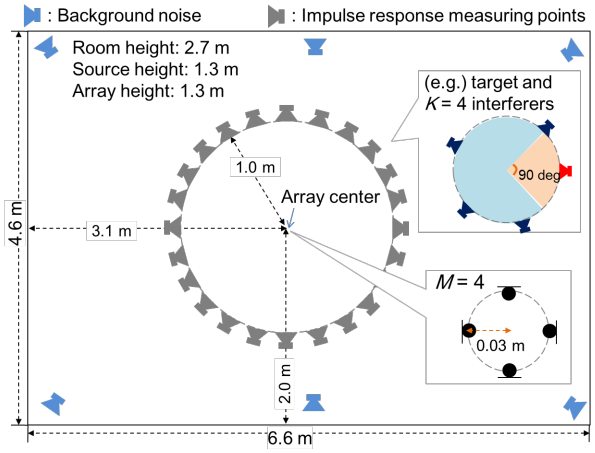

Fig. 3. Recording situations

with the impulse responses measured in two reverberant rooms (RT_{60} at 1.0 kHz: 230, 330 ms). As shown in Fig. 3, the interfering noise sources ($K = 2, 4, 6$ in number) were placed at randomly chosen positions located outside an angular region (of 90 degrees width) that included the target source. Twenty source arrangements were prepared for each condition. We convolved speech signals at a loudspeaker with impulse responses. In addition, we prepared two kinds of background noise (stationary pink noise and non-stationary cafeteria noise [21]). They were convolved with impulse responses from loudspeakers at the edge of the room floor and superimposed over the microphone observation at various noise levels, which was adjusted to $\{-20, -10, 0\}$ dB relative to the averaged target source level. In total, the size of the data sets was 1.4 hours ($= 2$ [reverberant rooms] $\times$ 3 [# of interference noises] $\times$ 3 [background noise level] $\times$ 20 [source arrangements] $\times$ 14 [sec]). The parameters are summarized in Table I.

$L = 3$ beamforming filters were designed with the minimum variance distortionless response (MVDR) method [16], with one focused on the target direction θ_{S} and the remaining two with nulls in directions centered on θ_{S}, with a directional spacing of 120 degrees. PSD estimation was conducted for frequency bands spaced corresponding to the equivalent rectangular bandwidth (ERB) [22] scale (44 bands with no overlap corresponding from 125 to 8000 Hz). The Wiener filter gain was lower bound at 0.05 to suppress musical noise.

The proposed method was compared with four conventional methods. These were: fixed MVDR beamformer (Conv. #1) where the noise spatial correlation matrix is calculated by using array manifold vectors assuming plane wave propagation in free field, speech distortion weighted multichannel Wiener filter (SDW-MWF) (Conv. #2), PSD estimation-in-beamspace method [10] (Conv. #3), and its extension [11] (Conv. #4). Note that the target focusing beamformer used in Conv. #2-4 and the proposed method is the same as the beamformer used in Conv. #1. In [8], the overall procedure is shown in its Fig. 1.1 and sidelobe-canceling filter is specified in its (1.15). For sidelobe-canceling filter design, it is necessary to estimate the voice activity of the observed signals. However, the output sound quality varies with VAD accuracy. To avoid this effect, we have provided accurate VAD information and

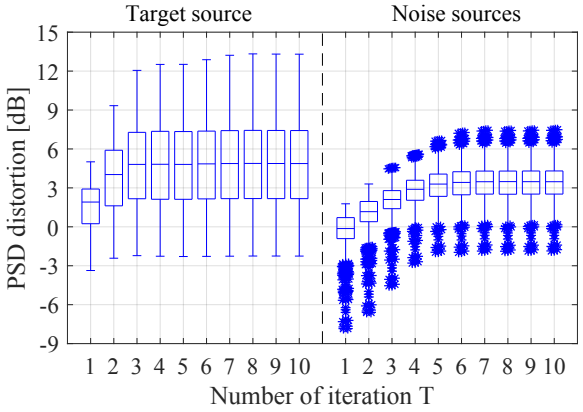

Fig. 4. Relationship between PSD distortion and iteration number T in proposed method. (left) PSD distortion in target source, (right) PSD distortion in noise sources. To show behavior of the data distributions, the box plot was used. After eliminating outliersamples drawn by asterisk $*$, the maximum, upper quartile (75th percentile), median, lower quartile (25th percentile), and minimum are marked.

the target/noise spatial correlation matrix can then be estimated. For calculating the sidelobe-canceling filters according to (1.15) in [8], the target spatial correlation matrix was mixed in the noise spatial correlation matrix with strength 1/10.

The PSD distortions of target source and noise sources were computed using the estimated PSDs $\{u_1, u_2, u_3\}$ as,

$$\mathrm{PSDdist}_{\mathrm{target}} = 10\log_{10}\frac{\mathbb{E}_{\omega,\tau}[\bar{\phi}_{\mathrm{S}}^2]}{\mathbb{E}_{\omega,\tau}[(\bar{\phi}_{\mathrm{S}}-u_1)^2]} \quad [\mathrm{dB}],$$

$$\mathrm{PSDdist}_{\mathrm{noise}} = 10\log_{10}\frac{\mathbb{E}_{\omega,\tau}[\bar{\phi}_{\mathrm{N}}^2]}{\mathbb{E}_{\omega,\tau}[(\bar{\phi}_{\mathrm{N}}-(u_2+u_3))^2]} \quad [\mathrm{dB}],$$

where $\{\bar{\phi}_{\mathrm{S}}, \bar{\phi}_{\mathrm{N}}\}$ is the ground truth of the PSD of target/noise sources. Since the proposed method (Algorithm 1) is constructed with an iterative update form and its performance varies with the iteration number T, the relationship between the PSD distortion and T was investigated first. As shown in Fig. 4, these measures were saturated for around $T = 5$. Thus, we selected $T = 5$ times iteration for the proposed method. The resulting average calculation times for signals of 14 seconds duration are shown in Table II. All methods facilitate real-time computation.

B. Experimental Results

As evaluation measures, we used (i) PSD distortion, (ii) signal-to-interference and background noise ratio improvement (SINR-I) and (iii) signal-to-distortion ratio (SDR) [23], and PESQ [24] as a perceptually relevant metric.

The PSD distortion was calculated for Conv. #3, #4 and the proposed method because these procedures operate in the power spectral domain. The results were shown in Fig. 5 that both target and noise PSDs were estimated most accurately with the proposed method. The proposed method was particularly superior over conventional methods w.r.t noise PSD estimation accuracy. This can be explained by the newly added constraint (12), i.e., PSD estimation accuracy of noise sources can be improved when that of target source is accurately estimated.

TABLE II
AVERAGED COMPUTATION TIME FOR 14 SECONDS SIGNALS. (2.60 GHZ INTEL XEON CPU E5-2650, UNIT: SECONDS)

Conv. #1	Conv. #2	Conv. #3	Conv. #4	Prop
0.27	1.71	0.56	0.86	1.05

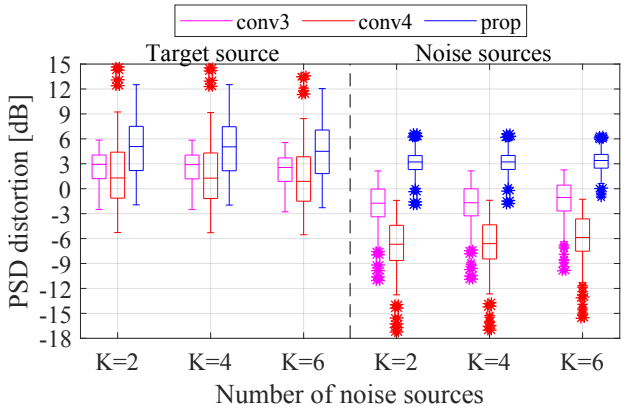

Fig. 5. PSD distortion for each number of noise sources K. (left) PSD distortion in target source, (right) PSD distortion in noise sources.

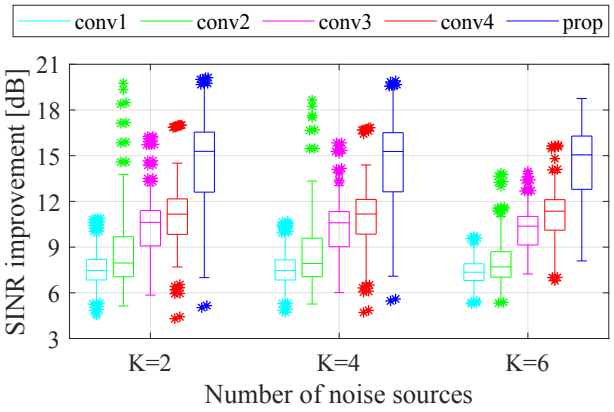

Fig. 6. SIR improvement for each number of noise sources K.

In Figs. 6 and 7, the SINR-I and SDR for each method are shown. For both measures, the best performance was obtained with the proposed method. The SINR-I of the proposed method was improved around 3.4 dB compared with that of Conv. #4. This is consistent with the results for PSD distortion, i.e., the proposed method effectively reduces the noise with only a minimal distortion of the target source. Whereas the SINR-I scores with Conv. #4 were better than those with Conv. #3, the SDR scores gave the opposite result. The evaluation scores were not significantly affected by the reverberation time and background noise pattern (stationary pink noise/non-stationary cafeteria noise). Instead of subjective sound quality tests, we measured the PESQ score for each sample. The results are shown in Fig. 8. The average PESQ score with the proposed method was 2.5 while it was 2.1 for Conv. #1 and 2.3 for Conv. #4.

Audio samples used in experiments are available on our web-cite [1].

[1]http://www.kecl.ntt.co.jp/icl/ls/members/niwa/AudioSamples_IEEETASLP_micmos.zip

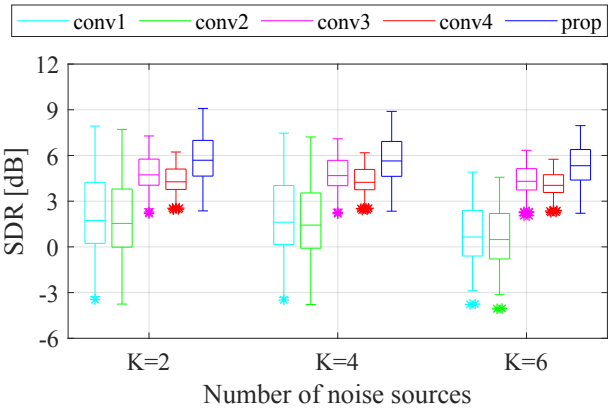

Fig. 7. SDR for each number of noise sources K.

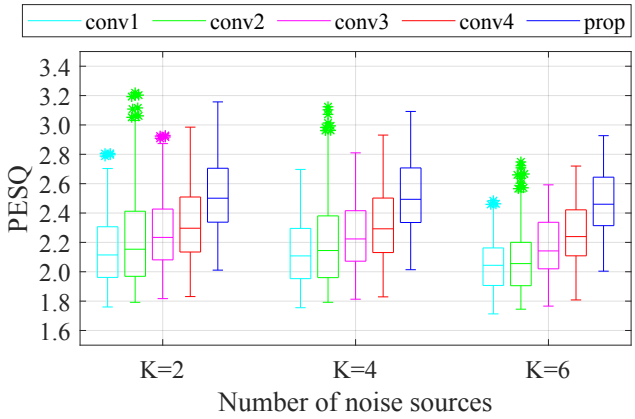

Fig. 8. PESQ score for each number of noise sources K

V. Conclusions

A PSD estimation method based on MOS was proposed. We showed that commonly available prior knowledge on the PSD, such as sparseness, can be represented by convex cost functions and linear constraints. The corresponding dual problem can be written as the sum of two convex functions. We used Peaceman-Rachford splitting with a Newton method as a solver for the optimization problem, because it facilitates the estimation of the PSDs in real-time. From the experimental results we can conclude that the proposed method is an effective noise reduction method that results in minimal distortion of the target source.

Appendix

A. Derivation of Algorithm 1

In this Appendix, the variable update rule in **Algorithm 1** is derived. First, the update rule associated with the resolvent operator R_1, which corresponds to (32), is derived. Assuming that two points $\{\mathbf{p}, \mathbf{p}'\}$ are related through the resolvent operator R_1, their relationship is

$$\begin{aligned}
&\mathbf{p} \in R_{1,p}(\mathbf{p}') = (\mathrm{Id} + \mathbf{M}_p^{-1}\partial_p G_1)^{-1}(\mathbf{p}'),\\
&\mathbf{0} \in \mathbf{M}_p^{-1}\partial_p G_1(\mathbf{p}) + \mathbf{p} - \mathbf{p}',\\
&\widetilde{\mathbf{p}} \in \widetilde{\mathbf{p}}' - \partial_p G_1(\mathbf{p}),
\end{aligned} \tag{61}$$

where $\widetilde{\ }$ denotes a transformation of the form $\widetilde{\mathbf{p}}' = \mathbf{M}_p\mathbf{p}'$. Since the subdifferential operator of G_1 w.r.t $\mathbf{p}$ is $\partial_p G_1(\mathbf{p}) = \mathbf{A}\partial_p F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p}+\mathbf{B}^\mathrm{T}\mathbf{q}+\mathbf{r})$ and since $\mathbf{u} \in \partial_p F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p}+\mathbf{B}^\mathrm{T}\mathbf{q}+\mathbf{r})$, we can rewrite (61) as

$$\begin{aligned}
&\widetilde{\mathbf{p}} \in \widetilde{\mathbf{p}}' - \mathbf{A}\partial_p F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p} + \mathbf{B}^\mathrm{T}\mathbf{q} + \mathbf{r}),\\
&\widetilde{\mathbf{p}} \in \widetilde{\mathbf{p}}' - \mathbf{A}\mathbf{u}, \quad \mathbf{u} \in \partial_p F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p} + \mathbf{B}^\mathrm{T}\mathbf{q} + \mathbf{r}).
\end{aligned} \tag{62}$$

(62) indicates that $\widetilde{\mathbf{p}}$ can be updated by using the updated $\mathbf{u}$ and the previous point $\widetilde{\mathbf{p}}'$. (The $\mathbf{u}$-update rule is derived later.)

For the remaining dual variables $\mathbf{q} \in R_{1,q}(\mathbf{q}')$, the relationship among $\{\mathbf{q}, \mathbf{q}', \mathbf{u}\}$ is obtained in a similar way, as

$$\begin{aligned}
&\mathbf{q} \in R_{1,q}(\mathbf{q}') = (\mathrm{Id} + \mathbf{M}_q^{-1}\partial_q G_1)^{-1}(\mathbf{q}'),\\
&\mathbf{0} \in \mathbf{M}_q^{-1}\partial_q G_1(\mathbf{q}) + \mathbf{q} - \mathbf{q}',\\
&\widetilde{\mathbf{q}} \in \widetilde{\mathbf{q}}' - \partial_q G_1(\mathbf{q}), \qquad (\widetilde{\mathbf{q}}' = \mathbf{M}_q\mathbf{q}')\\
&\widetilde{\mathbf{q}} \in \widetilde{\mathbf{q}}' - (\mathbf{B}\partial_q F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p} + \mathbf{B}^\mathrm{T}\mathbf{q} + \mathbf{r}) - \mathbf{c}),\\
&\widetilde{\mathbf{q}} \in \widetilde{\mathbf{q}}' - (\mathbf{B}\mathbf{u} - \mathbf{c}), \quad \mathbf{u} \in \partial_q F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p} + \mathbf{B}^\mathrm{T}\mathbf{q} + \mathbf{r}).
\end{aligned} \tag{63}$$

For $\mathbf{r} \in R_{1,r}(\mathbf{r}')$, we obtain

$$\begin{aligned}
&\mathbf{r} \in R_{1,r}(\mathbf{r}') = (\mathrm{Id} + \mathbf{M}_r^{-1}\partial_r G_1)^{-1}(\mathbf{r}'),\\
&\mathbf{0} \in \mathbf{M}_r^{-1}\partial_r G_1(\mathbf{r}) + \mathbf{r} - \mathbf{r}',\\
&\widetilde{\mathbf{r}} \in \widetilde{\mathbf{r}}' - \partial_r G_1(\mathbf{r}), \qquad (\widetilde{\mathbf{r}}' = \mathbf{M}_r\mathbf{r}')\\
&\widetilde{\mathbf{r}} \in \widetilde{\mathbf{r}}' - \partial_r F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p} + \mathbf{B}^\mathrm{T}\mathbf{q} + \mathbf{r}),\\
&\widetilde{\mathbf{r}} \in \widetilde{\mathbf{r}}' - \mathbf{u}, \quad \mathbf{u} \in \partial_r F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p} + \mathbf{B}^\mathrm{T}\mathbf{q} + \mathbf{r}).
\end{aligned} \tag{64}$$

Applying a basic property of the subdifferential of a convex conjugate function, $\partial F_1 = (\partial F_1^\star)^{-1}$ (e.g. [20]) to $\mathbf{u} \in \partial F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p} + \mathbf{B}^\mathrm{T}\mathbf{q} + \mathbf{r})$, we obtain

$$\begin{aligned}
&\mathbf{u} \in \partial F_1^\star(\mathbf{A}^\mathrm{T}\mathbf{p} + \mathbf{B}^\mathrm{T}\mathbf{q} + \mathbf{r}),\\
&\partial F_1(\mathbf{u}) \in \mathbf{A}^\mathrm{T}\mathbf{p} + \mathbf{B}^\mathrm{T}\mathbf{q} + \mathbf{r},\\
&\partial F_1(\mathbf{u}) \in \mathbf{A}^\mathrm{T}\mathbf{M}_p^{-1}\widetilde{\mathbf{p}} + \mathbf{B}^\mathrm{T}\mathbf{M}_q^{-1}\widetilde{\mathbf{q}} + \mathbf{M}_r^{-1}\widetilde{\mathbf{r}}.
\end{aligned} \tag{65}$$

By using (62)-(64), the $\mathbf{u}$-update procedure using dual variables $\{\widetilde{\mathbf{p}}', \widetilde{\mathbf{q}}', \widetilde{\mathbf{r}}'\}$ is formulated by

$$\begin{aligned}
\mathbf{0} \in \partial F_1(\mathbf{u}) &+ \mathbf{A}^\mathrm{T}\mathbf{M}_p^{-1}(\mathbf{A}\mathbf{u} - \widetilde{\mathbf{p}}')\\
&+ \mathbf{B}^\mathrm{T}\mathbf{M}_q^{-1}(\mathbf{B}\mathbf{u} - \mathbf{c} - \widetilde{\mathbf{q}}')\\
&+ \mathbf{M}_r^{-1}(\mathbf{u} - \widetilde{\mathbf{r}}').
\end{aligned} \tag{66}$$

The integral of (66) leads the $\mathbf{u}$-update procedure:

$$\begin{aligned}
\mathbf{u}^{t+1} = \arg\min_{\mathbf{u}} \Big(&F_1(\mathbf{u}) + \frac{1}{2}\left\langle \mathbf{M}_p^{-1}(\mathbf{A}\mathbf{u} - \widetilde{\mathbf{p}}^t), \mathbf{A}\mathbf{u} - \widetilde{\mathbf{p}}^t \right\rangle\\
&+ \frac{1}{2}\left\langle \mathbf{M}_q^{-1}(\mathbf{B}\mathbf{u} - \mathbf{c} - \widetilde{\mathbf{q}}^t), \mathbf{B}\mathbf{u} - \mathbf{c} - \widetilde{\mathbf{q}}^t \right\rangle\\
&+ \frac{1}{2}\left\langle \mathbf{M}_r^{-1}(\mathbf{u} - \widetilde{\mathbf{r}}^t), \mathbf{u} - \widetilde{\mathbf{r}}^t \right\rangle \Big).
\end{aligned} \tag{67}$$

By using the updated $\mathbf{u}$, the dual variables are updated in accord with (62)-(64) as

$$\widetilde{\mathbf{p}}^{t+1/4} = \widetilde{\mathbf{p}}^t - \mathbf{A}\mathbf{u}^{t+1}, \tag{68}$$

$$\widetilde{\mathbf{q}}^{t+1/4} = \widetilde{\mathbf{q}}^t - (\mathbf{B}\mathbf{u}^{t+1} - \mathbf{c}), \tag{69}$$

$$\widetilde{\mathbf{r}}^{t+1/4} = \widetilde{\mathbf{r}}^t - \mathbf{u}^{t+1}. \tag{70}$$

Thus, the primal-dual update w.r.t. (32) is composed of (67) and (68)-(70). By using the updated dual variables, (33) results

in:

$$\widetilde{\mathbf{p}}^{t+1/2} = 2\widetilde{\mathbf{p}}^{t+1/4} - \widetilde{\mathbf{p}}^{t} = \widetilde{\mathbf{p}}^{t} - 2\mathbf{A}\mathbf{u}^{t+1}, \tag{71}$$

$$\widetilde{\mathbf{q}}^{t+1/2} = 2\widetilde{\mathbf{q}}^{t+1/4} - \widetilde{\mathbf{q}}^{t} = \widetilde{\mathbf{q}}^{t} - 2(\mathbf{B}\mathbf{u}^{t+1} - \mathbf{c}), \tag{72}$$

$$\widetilde{\mathbf{r}}^{t+1/2} = 2\widetilde{\mathbf{r}}^{t+1/4} - \widetilde{\mathbf{r}}^{t} = \widetilde{\mathbf{r}}^{t} - 2\mathbf{u}^{t+1}. \tag{73}$$

Second, we derive variable update rules associated with the resolvent operator R_2

$$\begin{aligned} &\mathbf{p} \in R_{2,p}(\mathbf{p}') = (\mathrm{Id} + \mathbf{M}_p^{-1}\partial_p G_2)^{-1}(\mathbf{p}'), \\ &\mathbf{0} \in \mathbf{M}_p^{-1}\partial_p G_2(\mathbf{p}) + \mathbf{p} - \mathbf{p}', \\ &\widetilde{\mathbf{p}} \in \widetilde{\mathbf{p}}' - \partial_p G_2(\mathbf{p}), \ \ (\widetilde{\mathbf{p}}' = \mathbf{M}_p(\mathbf{p}')). \end{aligned} \tag{74}$$

Since the subdifferential operator of G_2 w.r.t. $\mathbf{p}$ is $\partial_p G_2(\mathbf{p}) = -\partial_p F_2^\star(-\mathbf{p})$ and since $\mathbf{v} \in \partial_p F_2^\star(-\mathbf{p})$, we can rewrite (74) as

$$\begin{aligned} &\widetilde{\mathbf{p}} \in \widetilde{\mathbf{p}}' + \partial_p F_2^\star(-\mathbf{p}), \\ &\widetilde{\mathbf{p}} \in \widetilde{\mathbf{p}}' + \mathbf{v}. \end{aligned} \tag{75}$$

(75) indicates that $\widetilde{\mathbf{p}}$ can be updated by using updated $\mathbf{v}$ and the previous point $\widetilde{\mathbf{p}}'$.

The $\mathbf{v}$-update rule is derived from $\mathbf{v} \in \partial F_2^\star(-\mathbf{p})$, as

$$\begin{aligned} &\mathbf{v} \in \partial F_2^\star(-\mathbf{p}), \\ &\partial F_2(\mathbf{v}) \in -\mathbf{p}, \\ &\partial F_2(\mathbf{v}) \in -\mathbf{M}_p^{-1}(\widetilde{\mathbf{p}}), \\ &\mathbf{0} \in \partial F_2(\mathbf{v}) + \mathbf{M}_p^{-1}(\mathbf{v} + \widetilde{\mathbf{p}}'). \end{aligned} \tag{76}$$

The integral of (76) leads the $\mathbf{v}$-update procedure:

$$\mathbf{v}^{t+1} = \arg\min_{\mathbf{v}} \Big(F_2(\mathbf{v}) + \frac{1}{2}\Big\langle \mathbf{M}_p^{-1}(\mathbf{v} + \widetilde{\mathbf{p}}^{t+1/2}), \mathbf{v} + \widetilde{\mathbf{p}}^{t+1/2} \Big\rangle \Big). \tag{77}$$

Following (75), dual variable $\mathbf{p}$ is updated by using updated $\mathbf{v}$ and previous point as

$$\widetilde{\mathbf{p}}^{t+3/4} = \widetilde{\mathbf{p}}^{t+1/2} + \mathbf{v}^{t+1}, \tag{78}$$

where the notation follows (34). Thus, the update procedure in (35) results in

$$\widetilde{\mathbf{p}}^{t+1} = 2\widetilde{\mathbf{p}}^{t+3/4} - \widetilde{\mathbf{p}}^{t+1/2} = \widetilde{\mathbf{p}}^{t+1/2} + 2\mathbf{v}^{t+1}. \tag{79}$$

Since R_2 and C_2 do not include a $\mathbf{q}$-update procedure, $\mathbf{q}$ remains unchanged:

$$\mathbf{q}^{t+1} = \mathbf{q}^{t+1/2}. \tag{80}$$

Finally, the $\mathbf{r}$-update procedure using the resolvent operator R_2 is provided. Two points $\{\mathbf{r}, \mathbf{r}'\}$ are associated by

$$\begin{aligned} &\mathbf{r} \in R_{2,r}(\mathbf{r}') = (\mathrm{Id} + \mathbf{M}_r^{-1}\partial_r G_2)^{-1}(\mathbf{r}'), \\ &\mathbf{0} \in \mathbf{M}_r^{-1}\partial_r G_2(\mathbf{r}) + \mathbf{r} - \mathbf{r}', \\ &\mathbf{0} \in \mathbf{M}_r^{-1}\partial_r \iota_{(\mathbf{r}\succeq\mathbf{0})}(\mathbf{r}) + \mathbf{r} - \mathbf{r}', \\ &\mathbf{0} \in \partial_r \iota_{(\widetilde{\mathbf{r}}\succeq\mathbf{0})}(\widetilde{\mathbf{r}}) + \widetilde{\mathbf{r}} - \widetilde{\mathbf{r}}'. \end{aligned} \tag{81}$$

The integration of (81) gives

$$\widetilde{\mathbf{r}}^{t+1} = \arg\min_{\widetilde{\mathbf{r}}} \Big(\iota_{(\widetilde{\mathbf{r}}\succeq\mathbf{0})}(\widetilde{\mathbf{r}}) + \frac{1}{2}\|\widetilde{\mathbf{r}} - \widetilde{\mathbf{r}}^{t+1/2}\|_2^2 \Big), \tag{82}$$

where the notation follows (34). (82) can be calculated bv

$$\widetilde{r}_i^{t+3/4} = \begin{cases} \widetilde{r}_i^{t+1/2} & (\text{if } \widetilde{r}_i^{t+1/2} \geq 0) \\ 0 & (\text{otherwise}) \end{cases}. \tag{83}$$

Then, the update procedure in (35) results in

$$\widetilde{r}_i^{t+1} = \begin{cases} 2\widetilde{r}_i^{t+1/2} - \widetilde{r}_i^{t+1/2} = \widetilde{r}_i^{t+1/2} & (\text{if } \widetilde{r}_i^{t+1/2} \geq 0) \\ 2\cdot 0 - \widetilde{r}_i^{t+1/2} = -\widetilde{r}_i^{t+1/2} & (\text{otherwise}) \end{cases}. \tag{84}$$

B. Attributes of resolvent and Cayley operators

We investigate basic properties of the resolvent operator (26) and the Cayley operator (27). Let us assume that the monotonicity of T_i is determined by the following attribute for any two different two points $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$:

$$\gamma_{\mathrm{LB},i}\|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_2 \leq \|T_i(\boldsymbol{\xi}) - T_i(\boldsymbol{\xi}')\|_2 \leq \gamma_{\mathrm{UB},i}\|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_2, \tag{85}$$

where $\gamma_{\mathrm{LB},i} \in [0, \infty)$ and $\gamma_{\mathrm{UB},i} \in (0, \infty]$. The value of the constant varies with the properties of the cost. For example $\gamma_{\mathrm{LB},i} \in (0, \infty)$ when T_i is strongly monotone and $\gamma_{\mathrm{UB},i} \in (0, \infty)$ when T_i is Lipschitz continuous. Applying $\mathbf{M}^{-1}$ to T_i results in

$$\sigma_{\mathrm{LB},i}\|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_2 \leq \|\mathbf{M}^{-1}T_i(\boldsymbol{\xi}) - \mathbf{M}^{-1}T_i(\boldsymbol{\xi}')\|_2 \leq \sigma_{\mathrm{UB},i}\|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_2, \tag{86}$$

where $\sigma_{\mathrm{LB},i} \in [0, \infty)$ and $\sigma_{\mathrm{UB},i} \in (0, \infty]$. The values of $\sigma_{\mathrm{LB},i}$ and $\sigma_{\mathrm{UB},i}$ change with the design of $\mathbf{M}$.

From (85) and (86) follows the nonexpansive properties of resolvent operator and Cayley operator:

Theorem A.1: Nonexpansive property of resolvent operator

The contractive ratio for the input/output pairs $\boldsymbol{\xi}, \boldsymbol{\xi}'$ on the resolvent operator R_i is given by

$$\frac{1}{1+\sigma_{\mathrm{UB},i}}\|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_2 \leq \|R_i(\boldsymbol{\xi}) - R_i(\boldsymbol{\xi}')\|_2 \leq \frac{1}{1+\sigma_{\mathrm{LB},i}}\|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_2. \tag{87}$$

Let $\sigma_{\mathrm{LB},i} \in [0, \infty)$ and $\sigma_{\mathrm{UB},i} \in (0, \infty]$ hold. Then R_i is a nonexpansive operator.

Proof: The input/output pairs for the resolvent operator are $\boldsymbol{\zeta} = R_i(\boldsymbol{\xi})$, $\boldsymbol{\zeta}' = R_i(\boldsymbol{\xi}')$. They are reformulated by

$$(\mathrm{Id} + \mathbf{M}^{-1}T_i)(\boldsymbol{\zeta}) = \boldsymbol{\xi}, \ \ (\mathrm{Id} + \mathbf{M}^{-1}T_i)(\boldsymbol{\zeta}') = \boldsymbol{\xi}'.$$

By subtracting these, we obtain

$$\big(\mathrm{Id} + \mathbf{M}^{-1}T_i\big)(\boldsymbol{\zeta}) - \big(\mathrm{Id} + \mathbf{M}^{-1}T_i\big)(\boldsymbol{\zeta}') = \boldsymbol{\xi} - \boldsymbol{\xi}'. \tag{88}$$

Since $(\mathrm{Id} + \mathbf{M}^{-1}T_i)$ is strongly monotone with $(1 + \sigma_{\mathrm{LB},i})$, its inverse operator $(\mathrm{Id} + \mathbf{M}^{-1}T_i)^{-1} = R_i$ is Lipschitz continuous with $(1 + \sigma_{\mathrm{LB},i})^{-1}$. By taking the norm of (88), we obtain

$$\|\boldsymbol{\zeta} - \boldsymbol{\zeta}'\|_2 + \|\mathbf{M}^{-1}T_i(\boldsymbol{\zeta}) - \mathbf{M}^{-1}T_i(\boldsymbol{\zeta}')\|_2 \geq \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_2. \tag{89}$$

Since the property of $\mathbf{M}^{-1}T_i$ is given by (86), the lower bound in (87) is obtained. ■

Theorem A.2: Nonexpansive property of Cayley operator

The contractive ratio for the input/output pairs on the Cayley operator C_i satisfies

$$\| C_i(\boldsymbol{\xi}) - C_i(\boldsymbol{\xi}') \|_2 \leq \eta_i \| \boldsymbol{\xi} - \boldsymbol{\xi}' \|_2, \tag{90}$$

where η_i $(0 \leq \eta_i \leq 1)$ is defined by

$$\eta_i = \sqrt{1 - \frac{4\sigma_{\mathrm{LB},i}}{(1+\sigma_{\mathrm{UB},i})^2}}. \tag{91}$$

Let $\{\sigma_{\mathrm{LB},i}, \sigma_{\mathrm{UB},i}\} \in [0, \infty]$ hold. Then C_i is a nonexpansive operator.

Proof: By multiplying $(\boldsymbol{\zeta}-\boldsymbol{\zeta}')^{\mathrm{T}}$ with (88), we obtain

$$\begin{aligned}&\|\boldsymbol{\zeta}-\boldsymbol{\zeta}'\|_2^2 \\ &\quad+\left\langle \boldsymbol{\zeta}-\boldsymbol{\zeta}', \mathbf{M}^{-1}T_i(\boldsymbol{\zeta}) - \mathbf{M}^{-1}T_i(\boldsymbol{\zeta}') \right\rangle = \langle \boldsymbol{\zeta}-\boldsymbol{\zeta}', \boldsymbol{\xi}-\boldsymbol{\xi}' \rangle.\end{aligned}$$

From the lower bound in (86), we obtain

$$(1+\sigma_{\mathrm{LB},i}) \, \|\boldsymbol{\zeta}-\boldsymbol{\zeta}'\|_2^2 \leq \langle \boldsymbol{\zeta}-\boldsymbol{\zeta}', \boldsymbol{\xi}-\boldsymbol{\xi}' \rangle \, . \tag{92}$$

By taking the squared norm for the input/output pairs $\boldsymbol{\chi} = C_i(\boldsymbol{\xi})$, $\boldsymbol{\chi}' = C_i(\boldsymbol{\xi}')$, we obtain

$$\begin{aligned}\| \boldsymbol{\chi} - \boldsymbol{\chi}' \|_2^2 &= \| 2(\boldsymbol{\zeta} - \boldsymbol{\zeta}') - (\boldsymbol{\xi} - \boldsymbol{\xi}') \|_2^2 \\ &= 4 \|\boldsymbol{\zeta}-\boldsymbol{\zeta}'\|_2^2 - 4 \langle \boldsymbol{\zeta}-\boldsymbol{\zeta}', \boldsymbol{\xi}-\boldsymbol{\xi}' \rangle + \|\boldsymbol{\xi}-\boldsymbol{\xi}'\|_2^2 \quad \text{(93a)} \\ &\leq \| \boldsymbol{\xi} - \boldsymbol{\xi}' \|_2^2, \quad \text{(93b)}\end{aligned}$$

where (92) is used for reforming (93a) into (93b), and this proves the nonexpansive property of C_i. Combining (92) and (93a) results in

$$\| \boldsymbol{\chi} - \boldsymbol{\chi}' \|_2^2 \leq \| \boldsymbol{\xi} - \boldsymbol{\xi}' \|_2^2 - 4\sigma_{\mathrm{LB},i} \, \| \boldsymbol{\zeta} - \boldsymbol{\zeta}' \|_2^2 \, .$$

With the lower bound of (87), we obtain

$$\| \boldsymbol{\chi} - \boldsymbol{\chi}' \|_2^2 \leq \left(1 - \frac{4\sigma_{\mathrm{LB},i}}{(1+\sigma_{\mathrm{UB},i})^2} \right) \| \boldsymbol{\xi} - \boldsymbol{\xi}' \|_2^2 \, .$$

Hence, we obtain (90). ■

C. Convergence rate on P-R splitting

In this appendix we investigate the convergence rates of P-R splitting. Since the contractive ratio of the Cayley operator C_i is provided by η_i from Theorem A.2, the contractive ratio for subsequent input/output $\boldsymbol{\xi}^{t+1} = C_2 C_1(\boldsymbol{\xi}^t), \boldsymbol{\xi}^t = C_2 C_1(\boldsymbol{\xi}^{t-1})$ can be bounded by

$$\| \boldsymbol{\xi}^{t+1} - \boldsymbol{\xi}^t \|_2 \leq \eta_1 \eta_2 \, \| \boldsymbol{\xi}^t - \boldsymbol{\xi}^{t-1} \|_2 \, . \tag{94}$$

The difference between variable $\boldsymbol{\xi}$ and its fixed point $\boldsymbol{\xi}^*$ is represented by

$$\begin{aligned}\| \boldsymbol{\xi}^t - \boldsymbol{\xi}^* \|_2 &= \| \boldsymbol{\xi}^t - \boldsymbol{\xi}^{t+1} + \boldsymbol{\xi}^{t+1} - \boldsymbol{\xi}^{t+2} + \cdots - \boldsymbol{\xi}^* \|_2 \\ &\leq \sum_{r=t}^{\infty} \| \boldsymbol{\xi}^r - \boldsymbol{\xi}^{r+1} \|_2 \\ &\leq \left(\sum_{j=1}^{\infty} (\eta_1 \eta_2)^j \right) \| \boldsymbol{\xi}^{t+2} - \boldsymbol{\xi}^{t+1} \|_2 \\ &= \frac{\eta_1 \eta_2}{1 - \eta_1 \eta_2} \, \| \boldsymbol{\xi}^{t+2} - \boldsymbol{\xi}^{t+1} \|_2 \, . \end{aligned} \tag{95}$$

Similarly, we obtain

$$\| \boldsymbol{\xi}^{t+1} - \boldsymbol{\xi}^* \|_2 \leq \frac{1}{1 - \eta_1 \eta_2} \, \| \boldsymbol{\xi}^{t+2} - \boldsymbol{\xi}^{t+1} \|_2 \, . \tag{96}$$

From (95) and (96), the following inequality is obtained:

$$\| \boldsymbol{\xi}^{t+1} - \boldsymbol{\xi}^* \|_2 \leq \eta_1 \eta_2 \, \| \boldsymbol{\xi}^t - \boldsymbol{\xi}^* \|_2 \, . \tag{97}$$

Thus, the convergence rate on the Peaceman-Rachford splitting is

$$\| \boldsymbol{\xi}^t - \boldsymbol{\xi}^* \|_2 \leq (\eta_1 \eta_2)^t \, \| \boldsymbol{\xi}^0 - \boldsymbol{\xi}^* \|_2 \, . \tag{98}$$

REFERENCES

[1] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone arrays*. Springer, 2001, pp. 39–60.

[2] T. Wolff and M. Buck, "A generalized view on microphone array postfilters," in *International Workshop on Acoustic Signal Enhancement*, 2010.

[3] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 1988, pp. 2578–2581.

[4] I. A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 709–716, 2003.

[5] N. Madhu, , and R. Martin, "A versatile framework for speaker separation using a model-based speaker localization approach," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 1900–1912, 2010.

[6] M. Taseska and E. A. Habets, "MMSE-based blind source extraction in diffuse noise fields using a complex coherence-based a priori sap estimator," in *International Workshop on Acoustic Signal Enhancement(IWAENC) 2012*, 2012, pp. 1–4.

[7] R. Serizel, M. Moonen, B. V. Dijk, and J. Wouters, "Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, pp. 785–799, 2014.

[8] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Speech distortion weighted multichannel Wiener filtering techniques for noise reduction," in *Speech enhancement*. Springer, 2005, pp. 199–228.

[9] ——, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Communication*, vol. 49, pp. 636–656, 2007.

[10] Y. Hioka, K. Furuya, K. Kobayashi, K. Niwa, and Y. Haneda, "Underdetermined sound source separation using power spectrum density estimated by combination of directivity gain," vol. 21. IEEE, 2013, pp. 1240–1250.

[11] K. Niwa, Y. Hioka, and K. Kobayashi, "Post-filter design for speech enhancement in various noisy environments," in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2014, pp. 35–39.

[12] ——, "Microphone array source enhancement using subtractive PSD estimation model," *Applied Acoustics*, vol. 143, pp. 239–249, 2019.

[13] A. Kuklasi'nski, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood PSD estimation for speech enhancement in reverberation and noise," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1599–1612, 2016.

[14] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011, vol. 408.

[15] E. K. Ryu and S. Boyd, "Primer on monotone operator methods," *Applied and computational mathematics*, vol. 15, pp. 3–43, 2016.

[16] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[17] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on speech and audio processing*, vol. 9, pp. 504–512, 2001.

[18] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[19] D. W. Peaceman and H. H. Rachford, "The numerical solution of parabolic and elliptic differential equations," *The society for industrial and applied mathematics (SIAM) journal of society for industrial and applied mathematics*, vol. 3, pp. 28–41, 2017.

[20] R. T. Rockafellar, *Convex analysis*. Princeton University Press, 1970.

[21] J. Barker, R. Marxer, E. Vincent, and S. Watanabe, "The third 'CHiME' speech separation and recognition challenge: dataset, task and baselines," in *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, 2015, pp. 504–511.
[22] C. R. Glasberg and B. C. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing research*, vol. 47, no. 1-2, pp. 103–138, 1990.
[23] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE transactions on audio, speech, and language processing*, vol. 14, no. 4, pp. 1462–1469, 2006.
[24] E. Cristobal, C. Flavian, and M. Guinaliu, "Perceived e-service quality (PeSQ)," *Managing service quality: An international journal*, 2007.

Kenta Niwa (M'09) received his B.E., M.E., and Ph.D. in information science from Nagoya University in 2006, 2008, and 2014. Since joining the Nippon Telegraph and Telephone Corporation (NTT) in 2008, he has been engaged in research on microphone array signal processing. In 2017-18, he has joined as a visiting researcher of Victoria university of Wellington and involved with research on distributed machine learning and mathematical optimization. He is now a senior research engineer at NTT Media Intelligence Laboratories. He was awarded the Awaya Prize by the Acoustical Society of Japan (ASJ) in 2010. He is a member of IEEE, ASJ, and the Institute of Electronics, Information and Communication Engineers (IEICE).

Hironobu Chiba received the B.E. degree in information engineering and the M.S. degree in engineering from the University of Tsukuba, Japan, in 2013 and 2015, respectively. He joined Pioneer Corporation in 2015 and started the career as a Researcher in acoustic signal processing and music information retrieval with Pioneer R&D. He is currently a Researcher with Nippon Telegraph and Telephone Corporation (NTT) Media Intelligence Laboratories. He is a Member of the Acoustic Society of Japan.

Noboru Harada (M'99, SM'18) received the B.S. and M.S. degrees in computer science from the Department of Computer Science and Systems Engineering, Kyushu Institute of Technology, Fukuoka, Japan, in 1995 and 1997, respectively, and the Ph.D. degree in computer science from the Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Japan, in 2017. Since joining NTT Corporation, Tokyo, Japan, in 1997, he has been involved with research on speech and audio signal processing, such as high efficiency coding and lossless compression. His current research interests include acoustic signal processing and machine learning for acoustic event detection, including Anomaly Detection in Sound. Dr. Harada is the recipient of the Technical Development Award from the Acoustical Society of Japan (ASJ) in 2016, the Industrial Standardization Encouragement Awards from the Ministry of Economy Trade and Industry of Japan in 2011, and the Telecom System Technology Paper Encouragement Award from the Telecommunications Advancement Foundation of Japan in 2007. He is a member of the ASJ, the Institute of Electronics, Information and Communication Engineers, and the Information Processing Society of Japan.

Guoqiang Zhang received the B. Eng. from University of Science and Technology of China (USTC) in 2003, M.Phil. degree from University of Hong Kong in 2006, and Ph.D. degree from Royal Institute of Technology in 2010. From the spring of 2011, he worked as a Postdoctoral Researcher at Delft University of Technology. From the spring of 2015, he worked as a senior researcher at Ericsson AB, Sweden. Since 2017, he has been a senior lecturer in the School of Electrical and Data Engineering, University of Technology Sydney, Australia. He is an IEEE member. His current research interests include distributed optimization, large scale optimization, deep learning, and multimedia signal processing.

Bastiaan Kleijn (f99) received the M.Sc degree in physics from the University of California, Riverside, Riverside, CA, the M.S.E.E. degree from Stanford University, Stanford, CA, USA, and the Ph.D. degrees in soil science and electrical Engineering from the University of California, Riverside and TU Delft, Delft, Netherlands. He is a Professor with the Victoria University of Wellington, Wellington, New Zealand and a Professor (part-time) with the Delft University of Technology, Delft, Netherlands. He was a Professor and the Head of the Sound and Image Processing Laboratory with KTH, Stockholm, 1996–2010. He was a Founder of Global IP Solutions, a company that provided the enabling audio technology to Skype. It was acquired by Google in 2010. He has served on the editorial Boards of the IEEE Transactions Audio, Speech, Language Processing, Signal Processing, the IEEE Signal Processing Letters, and the IEEE Signal Processing Magazine and is currently on the Board of the IEEE Journal of Selected Topics on Signal Processing. He was the Technical Chair of ICASSP 1999 and EUSIPCO 2010, and two IEEE workshops.