# Sparse Distortionless Beamformer Based on Nonconvex Optimization

Taiga Kawamura[1], Kohei Yatabe[2], Ryoichi Miyazaki[1]

[1] *National Institute of Technology, Tokuyama College, Yamaguchi, Japan*
{i15kawamura,miyazaki}@tokuyama.kosen-ac.jp
[2] *Waseda University, Tokyo, Japan*
k.yatabe@asagi.waseda.jp

***Abstract*—Minimum power distortionless response (MPDR) beamformer is a popular beamformer that minimizes its output power under the distortionless constraint. To improve the performance of the MPDR beamformer, a sparse distortionless beamformer has been proposed. It minimizes the $\ell_1$ norm of the output signal under the same constraint so that the output has sparse time-frequency representation. While the $\ell_1$-norm-based formulation is theoretically advantageous owing to its convexity, its practical performance might not be excellent because of the bias of the $\ell_1$ norm. To reduce the bias and improve the performance, we propose a sparse distortionless beamformer based on a nonconvex sparsity-inducing objective function. The proposed beamformer is performed via a heuristic application of a primal-dual splitting algorithm. The experiments showed that the proposed beamformer can achieve higher performance and is more robust against mismatch of the target direction.**

***Index Terms*—Microphone array, speech enhancement, sparse time-frequency representation, proximal algorithm.**

## I. Introduction

Beamformers are powerful tools for speech enhancement when a sufficient number of microphones are available. The minimum power distortionless response (MPDR) beamformer [1], [2] is one of the beamformers that has been widely studied [3]–[9]. It minimizes the power of output signals under the constraint that allows no distortion for the target signal. Since the distortionless constraint maintains the target signal, minimization of the output power only eliminates the interference signals in the ideal situation. However, MPDR beamformers may not perform well in practice because of the mismatch of the distortionless constraint.

To improve the practical performance of MPDR beamformers, a sparse distortionless beamformer has been proposed [10]. It minimizes the $\ell_1$ norm under the same distortionless constraint. Since minimization of the $\ell_1$ norm induces sparsity, the sparse beamformer can enhance the target signal that is approximately sparse in the time-frequency domain such as a speech signal. That is, sparsity serves as a prior knowledge of speech signals, and therefore the sparse distortionless beamformer can achieve higher performance even when the distortionless constraint contains error.

However, the $\ell_1$ norm is not ideal for promoting sparsity due to its bias. The $\ell_1$-norm minimization attenuates not only small components but also larger components. This effect tries to suppress the target signal, which results in mixing of the target and interference signals. By reducing the bias of $\ell_1$-norm minimization, it should be able to further improve the performance of the sparse distortionless beamformer.

In this paper, we propose a sparse distortionless beamformer based on a nonconvex sparsity-inducing objective function. The proposed method reduces the bias by using a nonconvex function [11], [12] in place of the $\ell_1$ norm. A primal-dual splitting algorithm [13] is heuristically applied to the resultant optimization problem for obtaining the enhanced signal. The experiments were performed using three types of relative transfer functions (RTFs) [14], [15] for the distortionless constraint. The experimental results showed that the proposed method outperformed the conventional beamformers. It was also shown that the proposed method is robust against the mismatch of the constraint.

## II. Preliminaries

### A. Problem Formulation

Let $\boldsymbol{x}_t(f) = [X_1(f,t), X_2(f,t), \cdots, X_M(f,t)]^\mathsf{T}$ be an $M$-dimensional vector of signals observed by $M$ microphones, where $f$ and $t$ are indices of frequency bin and time frame, respectively, $X_m(f,t)$ represents the $m$th signal in the time-frequency domain, and $\cdot^\mathsf{T}$ denotes the transpose of a vector or matrix. Its observation model is given by

$$\boldsymbol{x}_t(f) = \boldsymbol{h}(f)S(f,t) + \boldsymbol{n}_t(f), \tag{1}$$

where $\boldsymbol{h}(f) = [H_1(f), H_2(f), \cdots, H_M(f)]^\mathsf{T}$ is an $M$-dimensional vector of the transfer functions, $H_m(f)$ is the transfer function from the target signal to the $m$th microphone, $S(f,t)$ is the time-frequency representation of the target signal, and $\boldsymbol{n}_t(f) = [N_1(f,t), N_2(f,t), \cdots, N_M(f,t)]^\mathsf{T}$ represents noise.

The output signal of a beamformer $Z(f,t)$ is given by

$$Z(f,t) = \boldsymbol{w}(f)^\mathsf{H}\boldsymbol{x}_t(f), \tag{2}$$

where $\boldsymbol{w}(f) = [W_1(f), W_2(f), \cdots, W_M(f)]^\mathsf{T}$ is a vector of beamformer coefficients, and $\cdot^\mathsf{H}$ is the Hermitian transpose.

The beamformer coefficients $\boldsymbol{w}(f)$ are designed by minimizing some criteria related the output signal $\boldsymbol{w}(f)^{\mathsf{H}}\boldsymbol{x}_t(f)$. Hereafter, the frequency index $f$ is dropped from the beamformer output as $\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}_t$ for simpler notation.

### B. MPDR Beamformer

The MPDR beamformer is obtained by minimizing the output energy under the distortionless constraint. Let the RTF be $\boldsymbol{a}(f) = [a_1(f), a_2(f), \cdots, a_M(f)]^{\mathsf{T}}$. Then, the coefficients of the MPDR beamformer is given as a solution to the following minimization problem:

$$\min_{\boldsymbol{w}} \quad \|\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}_t\|_2^2 \quad \text{s.t.} \quad \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a} = 1, \tag{3}$$

where $\|\cdot\|_2$ is the $\ell_2$ norm. Its closed-form solution is

$$\boldsymbol{w} = (\boldsymbol{\Phi}_x^{-1}\boldsymbol{a})/(\boldsymbol{a}^{\mathsf{H}}\boldsymbol{\Phi}_x^{-1}\boldsymbol{a}), \tag{4}$$

where $\boldsymbol{\Phi}_x = E[\boldsymbol{x}_t\boldsymbol{x}_t^{\mathsf{H}}]$ is the $M \times M$ spatial correlation matrix, and $E[\,\cdot\,]$ represents the expectation.

### C. $\ell_1$-norm-based Sparse Distortionless Beamformer [10]

By assuming that the target signal is sparse in the time-frequency domain, the sparse distortionless beamformer is obtained by solving the following minimization problem [10]:

$$\min_{\boldsymbol{w}} \quad \|\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}_t\|_1 \quad \text{s.t.} \quad \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a} = 1, \tag{5}$$

where $\|\cdot\|_1$ is the $\ell_1$ norm. This problem is a modified version of the MPDR beamformer given by Eq. (3). Since the $\ell_1$ norm induces sparsity, the sparse distortionless beamformer aims to make the output signal more sparse than the MPDR beamformer while maintining the distortionless condition of the target signal. To solve the optimization problem in Eq. (5), any appropriate convex optimization algorithm can be used.

## III. Proposed Method

The proposed method is given by an improved formulation and application of the primal-dual splitting algorithm. For convenience of explanation, we first explain the primal-dual splitting algorithm and then explain the proposed method.

### A. Primal-dual Splitting Algorithm [13]

The primal-dual splitting algorithm given in Algorithm 1 solves the following convex optimization problem:

$$\min_{\boldsymbol{w}} \quad g(\boldsymbol{w}) + h(\boldsymbol{L}\boldsymbol{w}), \tag{6}$$

where $g(\cdot)$ and $h(\cdot)$ are proper lower-semicontinuous convex functions, and $\boldsymbol{L}$ is a bounded linear operator. The proximity operator in the 2nd and 4th lines is defined as

$$\mathrm{prox}_{\mu g}(\boldsymbol{z}) = \arg\min_{\boldsymbol{x}} \quad g(\boldsymbol{x}) + \frac{1}{2\mu}\|\boldsymbol{z} - \boldsymbol{x}\|_2^2. \tag{7}$$

In Algorithm 1, the step sizes $\mu_1, \mu_2 > 0$ are chosen so that

$$\mu_1\mu_2\|\boldsymbol{L}\|_{\mathrm{op}}^2 \leq 1, \tag{8}$$

where $\|\cdot\|_{\mathrm{op}}^2$ is the operator norm. The relaxation parameter $\alpha$ can be arbitrarily set as $0 < \alpha < 2$.

**Algorithm 1** Primal-dual Splitting Algorithm [13]

**Input:** $\boldsymbol{L}, \boldsymbol{w}^{[1]}, \boldsymbol{y}^{[1]}, \mu_1, \mu_2, \alpha$
**Output:** $\boldsymbol{w}^{[J+1]}$

1: **for** $j = 1, 2, \cdots, J$ **do**
2: $\quad \bar{\boldsymbol{w}} = \mathrm{prox}_{\mu_1 g}(\boldsymbol{w}^{[j]} - \mu_1\mu_2\boldsymbol{L}^{\mathsf{H}}\boldsymbol{y}^{[j]})$
3: $\quad \boldsymbol{z} = \boldsymbol{y}^{[j]} + \boldsymbol{L}(2\bar{\boldsymbol{w}} - \boldsymbol{w}^{[j]})$
4: $\quad \bar{\boldsymbol{y}} = \boldsymbol{z} - \mathrm{prox}_{\frac{1}{\mu_2}h}(\boldsymbol{z})$
5: $\quad \boldsymbol{w}^{[j+1]} = \alpha\bar{\boldsymbol{w}} + (1-\alpha)\boldsymbol{w}^{[j]}$
6: $\quad \boldsymbol{y}^{[j+1]} = \alpha\bar{\boldsymbol{y}} + (1-\alpha)\boldsymbol{y}^{[j]}$
7: **end for**

By applying Algorithm 1 to Eq. (3) or (5), one can obtain an MPDR or sparse distortionless beamformer based on the primal-dual splitting algorithm. To do so, the minimization problems are adapted to Eq. (6). The distortionless constraint is handled by the following indicator function:

$$g(\boldsymbol{w}) = \begin{cases} 0 & (\boldsymbol{a}^{\mathsf{H}}\boldsymbol{w} = 1) \\ \infty & (\boldsymbol{a}^{\mathsf{H}}\boldsymbol{w} \neq 1) \end{cases}, \tag{9}$$

where the complex conjugate is taken for both side of the constraint for convenience. The complex conjugate of the output signal is written as $\boldsymbol{L}\boldsymbol{w}$ by the following matrix:

$$\boldsymbol{L} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T]^{\mathsf{H}}. \tag{10}$$

By setting $h(\,\cdot\,) = \|\cdot\|_2^2$ or $h(\,\cdot\,) = \|\cdot\|_1$, the primal-dual splitting algorithm for solving Eq. (3) or (5) is given by inserting the following proximity operators into Algorithm 1:

$$\mathrm{prox}_{\mu g}(\boldsymbol{w}) = \boldsymbol{w} + \frac{1 - \boldsymbol{a}^{\mathsf{H}}\boldsymbol{w}}{\|\boldsymbol{a}\|_2^2}\boldsymbol{a}, \tag{11}$$

$$\mathrm{prox}_{\frac{\mu}{2}\|\cdot\|_2^2}(\boldsymbol{z}) = \frac{1}{1+\mu}\,\boldsymbol{z}, \tag{12}$$

$$\left(\mathrm{prox}_{\mu\|\cdot\|_1}(\boldsymbol{z})\right)_i = \max\left(0,\, 1 - \frac{\mu}{|z_i|}\right) z_i, \tag{13}$$

where $\max(\cdot,\cdot)$ returns the larger value of the input scalars.

### B. Bias of Soft-thresholding Operator

The proximity operator of the $\ell_1$ norm given in Eq. (13) is called soft-thresholding operator [16]. Since it attenuates both small and large components, soft-thresholding cannot remain the important components intact. This effect can be seen in Fig. 1. The soft-thresholding operator changes every input value to the smaller one. Therefore, the conventional sparse distortionless beamformer tries to attenuate the target signal, which results in mixing of the signals. One solution to this problem is to consider an objective function whose proximity operator does not attenuate large values. An extreme example of such a function is the so-called $\ell_0$ norm whose proximity operator is the hard-thresholding operator and is also shown in Fig. 1. Although $\ell_0$ norm has no bias, minimizing it is difficult owing to its discontinuous nature. To alleviate both problems, a sparsity-inducing function that is continuous and has less bias should be used instead of the $\ell_1$ norm.

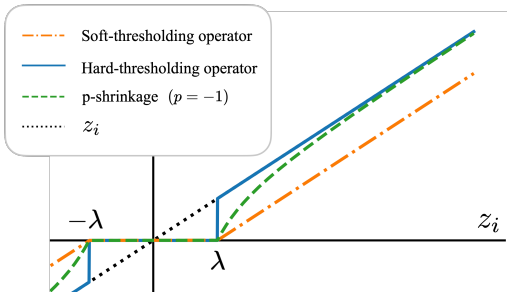


Fig. 1. Soft-thresholding operator (yellow dot-dashed), hard-thresholding operator (blue solid), and $p$-shrinkage operator ($p = -1$) (green dashed).

### C. Proposed Nonconvex Sparse Distortionless Beamformer

The proposed method uses a continuous sparsity-inducing function $\psi_p$ whose bias is less than the $\ell_1$ norm. That is, the proposed beamformer is given via the following problem:

$$\min_{\boldsymbol{w}} \quad \psi_p(\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}_t) \quad \text{s.t.} \quad \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a} = 1, \tag{14}$$

where $\psi_p$ is a real-valued function with a parameter $p$. By heuristically applying the primal-dual algorithm to this problem, an algorithm for the proposed method can be obtained by inserting Eq. (11) and the proximity operator of $\psi_p$ into Algorithm 1.

For the sparsity-inducing function $\psi_p$, this paper uses the nonconvex function given in [11], [12]. It is implicitly defined through its proximity operator called $p$-shrinkage:

$$\left(\mathrm{prox}_{\mu\psi_p}(\boldsymbol{z})\right)_i = \max\left(0,\, 1 - \left(\frac{\mu}{|z_i|}\right)^{2-p}\right) z_i, \tag{15}$$

where $p$ controls the degree of approximation of the $\ell_0$ norm, i.e., $p = 1$ corresponds to the soft-thresholding operator in Eq. (13) whereas $p \to -\infty$ gives the hard-thresholding operator. An example of the $p$-shrinkage operator ($p = -1$) is also given in Fig. 1. It can be seen that the two shortcomings, the bias and discontinuous nature, can be balanced by using the $p$-shrinkage operator.

## IV. Experiments

In this section, four types of experiments were performed:

- **Qualitative assessment of the effect of $p$:** Spectrograms of the output signals obtained by the proposed method with $p \in \{-10, -1, 0, 1\}$ were qualitatively compared.
- **Comparison of the performance at each iteration:** The performance for each iteration was evaluated with four conditions of the reverberation time $\mathrm{RT}_{60}$.
- **Comparison with error in the constraint:** The performance was compared by varying the target direction to evaluate the robustness against the error in the constraint.
- **Comparison using different types of constraint:** The performance was compared by using three types of RTFs for the distortionless constraint.

TABLE I
Comparison Methods

| Method | Objective function | Eq. | $\mathrm{prox}_h$ |
|---|---|---|---|
| Proposed ($p = -1$) | $\min_{\boldsymbol{w}}\ \psi_p(\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}_t)$ | (14) | (15) |
| PDS sparse ($\ell_1$) | $\min_{\boldsymbol{w}}\ \lVert\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}_t\rVert_1$ | (5) | (13) |
| PDS MPDR ($\ell_2$) | $\min_{\boldsymbol{w}}\ \lVert\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}_t\rVert_2^2$ | (3) | (12) |
| Conventional MPDR | $\min_{\boldsymbol{w}}\ \lVert\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}_t\rVert_2^2$ | (4) | — |

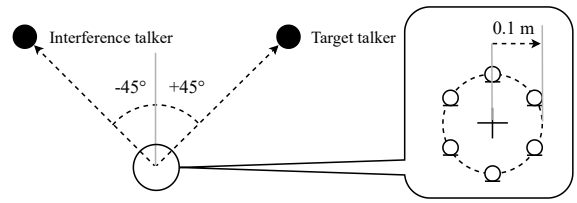


Fig. 2. Arrangement of microphone array, target talker, and interference talker.

### A. Experimental Conditions

The simulations were performed by using the audio signal processing software `Pyroomactoustics` [17]. The size of the room was set to $2.8 \times 4.2 \times 2.5$ m$^3$, and a 6-microphone uniform circular array of radius 0.1 m was placed at the center of the room. The target and interference talkers were placed 1.0 m away from the center of the microphone array, as shown in Fig. 2. The target and interference signals were male/female speech signals contained in the JNAS corpus [18] with no sensor noise. For STFT, the window size was set to 512 samples, and frame shift was set to 64 samples, with the Hamming window. The sampling frequency was $16,000$ Hz. For evaluation, the signal-to-distortion ratio (SDR) and PESQ were used.

To compare the difference due to formulations, the primal-dual splitting algorithm was applied to all minimization problems in Eqs. (3), (5), and (14). The methods are summarized in Table I, where PDS stands for primal-dual splitting.

### B. Effect of the Parameter $p$

To assess the effect of bias, qualitative comparison of the output signals was performed by varying the parameter $p$ of the $p$-shrinkage operator in Eq. (15). Since $p = 1$ and $p \to -\infty$ correspond to the soft- and hard-thresholding operators, respectively, decreasing $p$ from 1 reduces the bias of the proposed method compared to the $\ell_1$-norm-based sparse distortionless beamformer in Eq. (5). Here, $p$ was set to $1, 0, -1$, and $-10$. The reverberation time was 100 ms, and the number of iterations $J$ of Algorithm 1 was set to 2000.

Figure 3 shows the spectrograms of (a) the observed signal, (b) the target signal observed by the reference microphone, (c)–(f) and the output signals of the proposed beamformer for each $p$. The area dominated by the interference signal is indicated by the white dotted lines (see Fig. 3 (a) and (b)). Since a smaller value of $p$ leads to a smaller bias, $p = -1$ in (e) removed the interference signal better than $p = 1$ and $p = 0$ in (c) and (d), respectively. However, $p = -10$ in

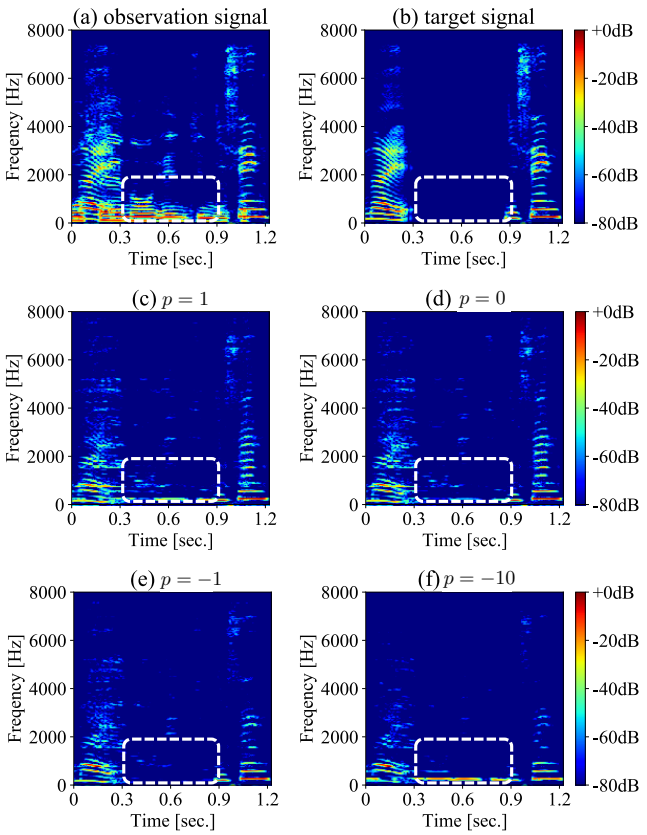


Fig. 3. Spectrograms of (a) observed signal, (b) target signal at reference microphone, and output signals of proposed beamformer whose parameter was set to (c) $p = 1$, (d) $p = 0$, (e) $p = -1$, and (f) $p = -10$.

(f) obtained a result worse than $p = -1$ in (e). Since the smallness of $p$ corresponds to the degree of approximation of the $\ell_0$ norm, smaller $p$ increases the risk of getting stuck into a local minimum. According to this observation, we set $p = -1$ for the proposed method in the rest of the experiments.

### C. Performance for Each Iteration

Figures 4 and 5 show SDR and PESQ, respectively, for each iteration. The reverberation times $\mathrm{RT}_{60}$ were set to $0, 100, 200,$ and 300 ms. As in the figures, while it requires more iterations, the proposed method $(p = -1)$ performed better than the other methods for all situations. However, the number of iterations to reach the peak has increased, indicating that there is a tradeoff between the performance and number of iterations. Interestingly, even though the conventional MPDR and PDS MPDR $(\ell_2)$ solve the same optimization problem in Eq. (3), PDS MPDR tended to result in the better performance. This should be because the conventional MPDR has instability due to the inversion of the spatial correlation matrix as in Eq. (4).

### D. Robustness Against Error on Target Direction

To evaluate the sensitivity to the error on target direction, the direction of the RTF was varied $\pm 10°$ from the true direction. Reverberation time $\mathrm{RT}_{60}$ was set to 100 ms, and the number of

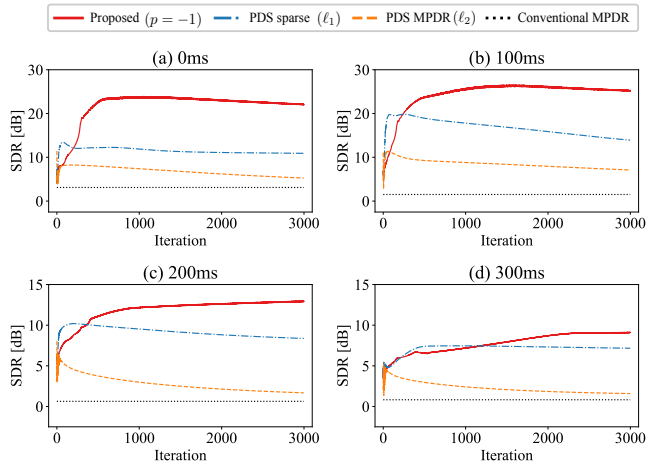


Fig. 4. SDR for each iteration. The reverberation time $\mathrm{RT}_{60}$ was set to (a) 0 ms, (b) 100 ms, (c) 200 ms, and (d) 300 ms.

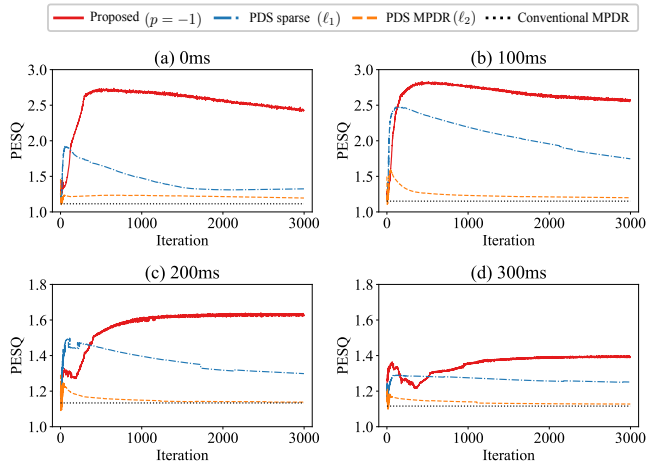


Fig. 5. PESQ for each iteration. The reverberation time $\mathrm{RT}_{60}$ was set to (a) 0 ms, (b) 100 ms, (c) 200 ms, and (d) 300 ms.

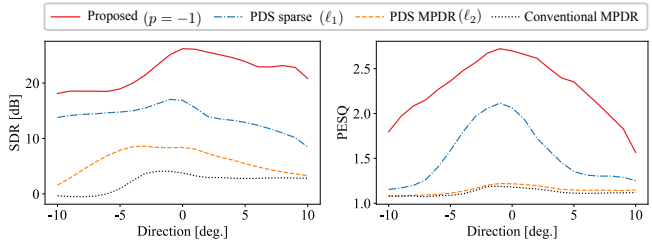


Fig. 6. SDR (left) and PESQ (right) for each angle of target direction.

iterations $J$ of Algorithm 1 was 2000. Figure 6 shows SDR and PESQ for each direction of the RTF. The direction of the target signal is denoted $0°$, where the interference signal is $-90°$ in this scale. For all situations, the proposed method $(p = -1)$ performed better than the other methods. Note that the proposed method with $-10°$ error obtained better SDR than the $\ell_1$-norm-based sparse distortionless beamformer with the true RTF. This result indicates the robustness of the proposed

TABLE II
SDR AVERAGED OVER 15 RESULTS

| | Point source | Ideal RTFs | CS |
|---|---|---|---|
| Proposed ($p = -1$) | **13.76** | **16.84** | 14.76 |
| PDS sparse ($\ell_1$) | 8.67 | 15.25 | **14.77** |
| PDS MPDR ($\ell_2$) | 3.76 | 10.77 | 11.71 |
| Conventional MPDR | 2.27 | 8.51 | 10.83 |

TABLE III
PESQ AVERAGED OVER 15 RESULTS

| | Point source | Ideal RTFs | CS |
|---|---|---|---|
| Proposed ($p = -1$) | **1.56** | **2.40** | **1.77** |
| PDS sparse ($\ell_1$) | 1.29 | 2.25 | 1.76 |
| PDS MPDR ($\ell_2$) | 1.15 | 1.62 | 1.55 |
| Conventional MPDR | 1.15 | 1.52 | 1.54 |

method against the error of the distortionless constraint.

### *E. Performance for Different Constraint using RTFs [14]*

For the distortionless constraint, the above experiments used the RTF calculated based on the point source model. Since the presence of reverberation changes the transfer function, the point source RTF has model mismatch unless the reverberation time is zero. To alleviate the model mismatch contained in the distortionless constraint, we consider two RTFs in addition to the point source model RTF.

In this experiment, the ideal RTF $\boldsymbol{a}_{\text{Ideal}}$ and an RTF estimated by using the covariance-subtraction (CS) method $\boldsymbol{a}_{\text{CS}}$ were considered [15]. Based on the observation model in Eq. (1), the ideal RTF is given by the following normalization:

$$\boldsymbol{a}_{\text{Ideal}} = \boldsymbol{h}/(\boldsymbol{e}_1^{\mathsf{T}}\boldsymbol{h}), \tag{16}$$

where $\boldsymbol{e}_1 = [1, 0, \cdots, 0]^{\mathsf{T}}$ is the unit vector selecting the reference microphone. The CS-based RTF is given by

$$\boldsymbol{a}_{\text{CS}} = (\boldsymbol{\Phi}_{\hat{S}}\boldsymbol{e}_1)/(\boldsymbol{e}_1^{\mathsf{T}}\boldsymbol{\Phi}_{\hat{S}}\boldsymbol{e}_1), \tag{17}$$

where $\boldsymbol{\Phi}_{\hat{S}}$ is the spatial correlation matrix of the target signal estimated by subtracting that of the interference signal $\boldsymbol{\Phi}_n = E[\boldsymbol{n}_k\boldsymbol{n}_k^{\mathsf{H}}]$, which was calculated ideally, as $\boldsymbol{\Phi}_{\hat{S}} = \boldsymbol{\Phi}_x - \boldsymbol{\Phi}_n$.

Here, the directions of the target and interference signals were given by 3 situations: $(+15°, -15°)$, $(+30°, -30°)$, and $(+45°, -45°)$. For each situation, 5 combinations of signals were used, which resulted in 15 types of observations. Reverberation time $\mathrm{RT}_{60}$ was 100 ms, and the number of iterations $J$ of Algorithm 1 was set to 5000.

The results are summarized in Tables II and III. Except CS in Table II, the proposed method ($p = -1$) performed better than the other methods. When the better RTFs ($\boldsymbol{a}_{\text{Ideal}}$ and $\boldsymbol{a}_{\text{CS}}$) are given for the distortionless constraint, the proposed method and the $\ell_1$-norm-based method performed similarly. In contrast, the proposed method outperformed the other methods when the given RTF was not accurate. The proposed method performed robustly regardless of the quality of the given RTF, which should be advantageous in a practical situation because estimating a high-quality RTF is not an easy task.

## V. CONCLUSION

In this paper, the sparse distortionless beamformer was proposed by improving the existing $\ell_1$-norm-based beamformer. The proposed method is based on the nonconvex sparse-inducing function whose proximity operator is given by the $p$-shrinkage operator. It is iteratively applied in the primal-dual splitting algorithm utilized to minimize the objective function. The simulation experiments confirmed that the proposed method performed better than the comparison methods even when the given target direction was erroneous. Since the proposed method required a lot of iterations when the reverberation time was long, investigation of a faster algorithm should be considered in the future works.

## REFERENCES


[1] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," IEEE ASSP Mag., vol. 5, no. 2, pp. 4–24, 1988.

[2] H. L. V. Trees, "Detection, estimation, and modulation theory, Part IV: Optimum array processing," N. Y., NY, USA: Wiley, 2002.

[3] L. Ehrenberg, S. Gannot, A. Leshem, E. Zehavi, "Sensitivity analysis of MVDR and MPDR beamformers," presented at the IEEE 26th Convention of Electr. Electron. Eng. in Israel, Eliat, Israel, 2018.

[4] 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 TASLP, vol. 25, no. 4, pp. 780–793, 2017.

[5] H. Erdogan, J. Hershey, S. Watanabe, M. Mandel, and J. L. Roux, "Improved MVDR beamforming using single-channel mask prediction networks," INTERSPEECH, pp. 1981–1985, 2016.

[6] T. Nakatani, N. Ito, T. Higuchi, S. Araki, and K. Kinoshita, "Integrating DNN-based and spatial clustering-based mask estimation for robust MVDR beamforming," IEEE ICASSP, pp. 286–290, 2017.

[7] Y. P. Liu, Q. Wan, and X. L. Chu, "A robust beamformer based on weighted sparse constraint," Prog. Electromagn. Res. Lett., Vol. 16, 53–60, 2010.

[8] J. Benesty and Y. Huang, "A single-channel noise reduction MVDR filter," Proc. Int. Conf. Acoust. Speech Signal Process., pp. 273–276, 2011.

[9] A. Schasse and R. Martin, "Estimation of subband speech correlations for noise reduction via MVDR processing," IEEE/ACM TASLP, vol. 22, no. 9, pp. 1355–1365, 2014.

[10] S. Emura, S. Araki, T. Nakatani, and N. Harada, "Distortionless beamforming optimized with $\ell_1$-norm minimization," IEEE Signal Process. Lett., vol. 25, no. 7, 2018.

[11] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in IEEE Int. Symp. Biomed. Imaging, 2009.

[12] R. Chartrand, "Shrinkage mappings and their induced penalty functions," in IEEE Int. Conf. Acoust., Speech Signal Process., pp. 1026–1029, 2014.

[13] N. Komodakis and J. C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," IEEE Signal Process. Mag., vol. 32, no. 6, pp. 31–54, 2015.

[14] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," IEEE Trans. Signal Processing, vol. 49, pp. 1614–1626, 2001.

[15] S. Markovich-Golan, S. Gannot, and W. Kellermann, "Performance analysis of the covariance-whitening and the covariance-subtraction methods for estimating the relative transfer function," Proc. IEEE EUSIPCO, 2018, pp. 2513–2517.

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

[17] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulations and array processing algorithms," Proc. IEEE ICASSP, pp. 351–355, 2018.

[18] K. Ito, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, and S. Itahashi, "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research," J. Acoust. Soc. Jpn, vol. 20, pp. 196–206, 1999.