ANALYTICAL SOLUTION OF NONLINEAR MICROPHONE ARRAY BASED ON COMPLEMENTARY BEAMFORMING

Shigeki Miyabe,[†*] *Biing-Hwang (Fred) Juang,*[‡] *Hiroshi Saruwatari,*[†] *Kiyohiro Shikano*[†]

† Graduate School of Information Science, Nara Institute of Science and Technology
{shige-m,sawatari,shikano}@is.naist.jp
‡ Center of Signal and Image Processing, Georgia Institute of Technology
juang@ece.gatech.edu

ABSTRACT

An analytical optimization method for nonlinear beamforming is proposed. It has been shown that twice the number of noise sources, compared to conventional linear beamformers, can be cancelled by nonlinear array signal processing with complementary beamforming and spectral subtraction. However, optimization of the beamformer coefficients via conventional adaptation techniques often suffers from a severe convergence problem. In this paper, we propose a new efficient optimization algorithm by re-writing the optimization objective of the two complementary beamformer vectors in a convex minimization problem, which can be solved with an analytical solution.

Index Terms— Adaptive beamformer, microphone array, speech enhancement, spectral subtraction.

1. INTRODUCTION

In many speech applications, from teleconferencing to human-machine interface via voice, signal acquisition without the intervention of the talkers or users is essential. Common microphones when used in such situations at a distance from the signal sources tend to capture an excessive amount of interference (acoustic ambient noise, reverberation, and others) and the quality of the captured signal is often considered undesirable. As a solution to this problem, microphone arrays that provide directivity gains by forming beam patterns towards the desired signal sources have been proposed for some time [1]. The advantage of beamforming is in its flexible control of the beam pattern to enhance or suppress sound sources in specific directions. The suppression ability of an array is particularly strong if the noise is generated from a point source in a free field.

However, conventional linear beamformers are often inadequate due to a number of reasons. First, the number of point sources to be suppressed via placement of directional nulls is bounded by the number of the microphones in the array. Second, in an ordinary situation, acoustic interferences in an enclosure (e.g., a room, a hall or an auditorium) tend to be diffuse rather than from point sources. Interference suppression in many real applications involves more than the placement of directional nulls. Thus, in order to deliver a reasonable performance on the captured signal quality, a beamformer must be integrated with other signal processing techniques and the idea of relying on a directivity pattern to mitigate the interference problem must be broadened.

There are a number of approaches to extending the basic idea of beamforming, such as incorporating single-channel nonlinear post-processing like spectral subtraction (SS) [2, 3, 4, 5, 6], Wiener filtering [7], or other amplitude estimation signal enhancement techniques [8]. In this paper, we focus on the integration of beamforming and SS, which, as shown in [3], can nullify twice the number of directional noise sources compared to conventional linear beamformers. The improved suppression capability is afforded by integrating the SS algorithm, which operates on the primary path (to enhance the target signal) and the reference path (to estimate the interfering noise), and the corresponding directional null placement along both the primary and reference paths in a jointly optimized fashion. Note that spectral subtraction is no longer employed as a post-processing technique to further reduce the residual interference in the output of a beamformer; rather, it is integrated into the array processing and optimization, thus making the resultant beamforming system a nonlinear one.

*Supported by the JSPS Fellowship for the Young Scientists.

The beamformer coefficients of the system need to be computed according to a certain criterion. Conventionally, adaptive algorithms such as the technique proposed in [3] are employed to obtain the optimization solution. However, these techniques usually suffer severe convergence problems. Other integrated approaches of beamforming and SS [2, 5] only adjust the noise estimator adaptively, compromising the overall system performance. Suppression of multiple sources is essential in source localization, and the property of SS-integrated beamformer is effective, as utilized in [4]. Nevertheless, real-time processing is just as critical important in many applications of source localization and poor convergence in the adaptive solution is detrimental. In this paper, we propose an efficient algorithm for the optimization of the integrated nonlinear beamforming system. The algorithm is derived by reformulating the nonlinear beamforming adaptation as a convex optimization problem, to which an analytic solution can be obtained, thereby circumventing the difficulty in algorithmic convergence.

2. OBSERVATION MODEL

Consider D interfering noise sources $n_d(\omega)$, $d = 1, 2, \dots, D$, and a target source $s_0(\omega)$ that are distributed in a room are observed at a K-element microphone array as

$$\boldsymbol{x}(\omega) = [x_1(\omega)\ \ x_2(\omega)\ \ \cdots\ \ x_K(\omega)]^{\mathrm{T}}, \tag{1}$$

where ω is the angular frequency and $\{\cdot\}^{\mathrm{T}}$ denotes matrix transposition. Denoting the transfer function from the d-th noise source to the k-th microphone element as $a_{kd}(\omega)$ and from the target source to the k-th microphone element as $a_{k0}(\omega)$, we have

$$\boldsymbol{x}(\omega) = \boldsymbol{a}_0(\omega)s_0(\omega) + \sum_{d=1}^{D}\boldsymbol{a}_d(\omega)n_d(\omega), \tag{2}$$

$$\boldsymbol{a}_d(\omega) = [a_{1d}(\omega)\ \ a_{2d}(\omega)\ \ \cdots\ \ a_{Kd}(\omega)]^{\mathrm{T}}. \tag{3}$$

In the frequency-domain adaptation of spatial filters, we assume the transfer functions change slowly and may be regarded time-invariant for suitable length of time duration. Ignoring bias caused by the window of the frame analysis, we adopt the usual short-time analysis practice and write each frame of analysis of the observed signals as

$$\boldsymbol{x}(\omega, t) = \boldsymbol{a}_0(\omega)s_0(\omega, t) + \sum_{d=1}^{D}\boldsymbol{a}_d(\omega)n_d(\omega, t), \tag{4}$$

where $x(\omega, t)$ denotes signal $x(\omega)$ in the t-th analysis frame.

3. CONVENTIONAL LINEAR ADAPTIVE BEAMFORMER

The goal of adaptive beamforming is to suppress or cancel the noise components without degrading the target signal. A linear beamformer is an ensemble of FIR filters

$$\boldsymbol{w}(\omega) = [w_1(\omega)\ \ w_2(\omega)\ \ \cdots\ \ w_K(\omega)], \tag{5}$$

applied to the observed signals captured at the microphones, respectively, so as to obtain an enhanced signal $y(\omega)$, formed by summing those filtered signals as

$$y(\omega) = \boldsymbol{w}(\omega)\boldsymbol{x}(\omega). \tag{6}$$

Note that we define the filters in a row vector $\boldsymbol{w}(\omega)$ in Eq. (5). The beamformer will successfully suppress the noise component if

$$\boldsymbol{w}(\omega)\boldsymbol{a}_d(\omega) = 0 \quad \text{for} \quad d = 1, 2 \dots, D. \tag{7}$$

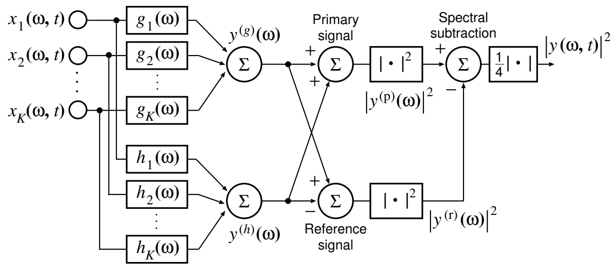

Fig. 1. Block diagram of the complementary beamformer.

In addition, the target source is extracted without degradation as $y(\omega) = s_0(\omega)$ if

$$\boldsymbol{w}(\omega)\boldsymbol{a}_0(\omega) = 1. \tag{8}$$

It is clear that the number of the noise sources to be cancelled in Eq. (7) is bounded by the number of the sensors as $K - 1$.

A typical technique to adaptively obtain the beamformer coefficients is the linearly constrained minimum variance (LCMV) method which uses minimum output power in the training data as the optimization objective [1]:

$$\min_{\boldsymbol{w}(\omega)} \langle |y(\omega, t)|^2 \rangle_{t=l_1,l_2\ldots,l_L} \quad \text{subject to} \quad \boldsymbol{w}(\omega)\boldsymbol{r}(\omega) = 1, \tag{9}$$

where $\langle\cdot\rangle_{t=l_1,l_2,\ldots,l_L}$ denotes averaging operator over the range of frame indices t, $t = l_1, l_2, \ldots, l_L$, and $\boldsymbol{r}(\omega)$ is an estimate of the transfer function $\boldsymbol{a}_0(\omega)$ w.r.t. the target source, refered to as the *steering vector*. The steering vector is synthesized from assumed direction of the target source. Note that "training" data $\boldsymbol{x}(\omega, t), t = l_1, l_2, \ldots, l_L$ must exclude those frames when target source is active; otherwise, the objective of Eq. (9) becomes meaningless. The Lagrange multiplier method gives the solution of LCMV beamformer in an analytical form as

$$\boldsymbol{w}(\omega) = \frac{\boldsymbol{r}(\omega)^{\mathrm{H}} \left(\langle \boldsymbol{x}(\omega, t)\boldsymbol{x}(\omega, t)^{\mathrm{H}} \rangle_{t=l_1,l_2,\ldots,l_L}\right)^{-1}}{\boldsymbol{r}(\omega)^{\mathrm{H}} \left(\langle \boldsymbol{x}(\omega, t)\boldsymbol{x}(\omega, t)^{\mathrm{H}} \rangle_{t=l_1,l_2,\ldots,l_L}\right)^{-1} \boldsymbol{r}(\omega)}, \tag{10}$$

where $\{\cdot\}^{\mathrm{H}}$ and $\{\cdot\}^{-1}$ denotes conjugate transposition and inversion of matrix, respectively.

4. NONLINEAR MICROPHONE ARRAY

4.1. Nonlinear Beamforming with Spectral Subtraction

In the current work, a nonlinear beamformer is formed by integrating complementary beamforming and the technique of spectral subtraction. Complementary beamforming involves two adaptive beamformers, the outputs of which are used as inputs to the SS algorithm, resulting in a strong interference cancellation performance at a compromise of the phase information [3]. While a conventional linear beamformer can deal with as many sources as the sensors, a nonlinear microphone array has the ability to deal with twice the number of the sources.

Figure 1 shows a block diagram of the nonlinear array signal processing. First, two processed signals, $y^{(g)}(\omega) = \boldsymbol{g}(\omega)\boldsymbol{x}(\omega)$ and $y^{(h)}(\omega) = \boldsymbol{h}(\omega)\boldsymbol{x}(\omega)$, are obtained as the output of two beamformers $\boldsymbol{g}(\omega)$ and $\boldsymbol{h}(\omega)$ with different beam patters:

$$y^{(g)}(\omega) = \boldsymbol{g}(\omega)\boldsymbol{a}_0(\omega)s_0(\omega) + \sum_{d=1}^{D} \boldsymbol{g}(\omega)\boldsymbol{a}_d(\omega)n_d(\omega), \tag{11}$$

$$y^{(h)}(\omega) = \boldsymbol{h}(\omega)\boldsymbol{a}_0(\omega)s_0(\omega) + \sum_{d=1}^{D} \boldsymbol{h}(\omega)\boldsymbol{a}_d(\omega)n_d(\omega). \tag{12}$$

The directivity patterns of the two beamformers are designed to be complementary, i.e., they satisfy the condition

$$|\boldsymbol{g}(\omega)\boldsymbol{a}_d(\omega)| \gg |\boldsymbol{h}(\omega)\boldsymbol{a}_d(\omega)|, \tag{13}$$

or

$$|\boldsymbol{h}(\omega)\boldsymbol{a}_d(\omega)| \gg |\boldsymbol{g}(\omega)\boldsymbol{a}_d(\omega)|, \tag{14}$$

for each of $d = 1, 2, \ldots, D$, and unity for the target source as

$$\boldsymbol{g}(\omega)\boldsymbol{a}_0(\omega) = \boldsymbol{h}(\omega)\boldsymbol{a}_0(\omega) = 1. \tag{15}$$

Under the complementary condition, the sum $y^{(\mathrm{p})}(\omega) = y^{(g)}(\omega) + y^{(h)}(\omega)$, refered to as primary signal, and difference $y^{(\mathrm{r})}(\omega) = y^{(g)}(\omega) - y^{(h)}(\omega)$, refered to as reference signal, of the beamformer outputs, can be given as

$$y^{(\mathrm{p})}(\omega) = 2s_0(\omega) + \sum_{d=1}^{D} [\boldsymbol{g}(\omega) + \boldsymbol{h}(\omega)]\, \boldsymbol{a}_d(\omega)n_d(\omega), \tag{16}$$

$$y^{(\mathrm{r})}(\omega) = \sum_{d=1}^{D} [\boldsymbol{g}(\omega) - \boldsymbol{h}(\omega)]\, \boldsymbol{a}_d(\omega)n_d(\omega). \tag{17}$$

If the directivity patterns $\boldsymbol{g}(\omega)\boldsymbol{a}_d(\omega)$ and $\boldsymbol{g}(\omega)\boldsymbol{a}_d(\omega)$ are designed complementary, and if there is no correlation among the sources, the following approximation holds:

$$E\left[\left|\sum_{d=1}^{D} [\boldsymbol{g}(\omega) + \boldsymbol{h}(\omega)]\, \boldsymbol{a}_d(\omega)n_d(\omega)\right|^2\right] \approx E\left[\left|\sum_{d=1}^{D} [\boldsymbol{g}(\omega) - \boldsymbol{h}(\omega)]\, \boldsymbol{a}_d(\omega)n_d(\omega)\right|^2\right]. \tag{18}$$

Thus, the noise component in the primary signal can be approximated by the reference signal, and the estimated power spectrum of the recovered signal is given as

$$\begin{aligned} |y(\omega)|^2 =& \frac{1}{4}\left\{|y^{(\mathrm{p})}(\omega)|^2 - E[|y^{(\mathrm{r})}(\omega)|^2]\right\} \\ =& \frac{1}{4}\Big[\{\boldsymbol{g}(\omega) + \boldsymbol{h}(\omega)\}\, \boldsymbol{x}(\omega)\boldsymbol{x}(\omega)^{\mathrm{H}} \{\boldsymbol{g}(\omega) + \boldsymbol{h}(\omega)\}^{\mathrm{H}} \\ & - \{\boldsymbol{g}(\omega) - \boldsymbol{h}(\omega)\}\, \boldsymbol{x}(\omega)\boldsymbol{x}(\omega)^{\mathrm{H}} \{\boldsymbol{g}(\omega) - \boldsymbol{h}(\omega)\}^{\mathrm{H}}\Big] \\ =& \frac{1}{4}\left\{\boldsymbol{g}(\omega)\boldsymbol{x}(\omega)\boldsymbol{x}(\omega)^{\mathrm{H}}\boldsymbol{h}(\omega)^{\mathrm{H}} + \boldsymbol{h}(\omega)\boldsymbol{x}(\omega)\boldsymbol{x}(\omega)^{\mathrm{H}}\boldsymbol{g}(\omega)^{\mathrm{H}}\right\}. \end{aligned} \tag{19}$$

4.2. Adaptation

A criterion for coefficient optimization is the minimization of a squared error under the constraint to maintain a unity gain with respect to the steering vector $\boldsymbol{r}(\omega)$. However, the solution to the minimization of the squared error is not unique and does not always give the optimal directivity pattern. To avoid the trivial solution, block-averaging technique is applied here. In the technique, the error is averaged in each block, and their squared sum is minimized. Under the unity gain constraint against the steering vector $\boldsymbol{r}(\omega)$, as

$$\boldsymbol{g}(\omega)\boldsymbol{r}(\omega) = \boldsymbol{h}(\omega)\boldsymbol{r}(\omega) = 1, \tag{20}$$

the objective of the optimization is then

$$\begin{aligned} J(\omega) &= \sum_{b=1}^{B} \langle |y(\omega, t)|^2 \rangle_{t\in\Omega(b)}^2 \\ &\equiv \sum_{b=1}^{B} \left\{ \frac{1}{|\Omega(b)|} \sum_{t\in\Omega(b)} |y(\omega, t)|^2 \right\}^2, \end{aligned} \tag{21}$$

where $b = 1, 2, \ldots, B$ denotes index of the blocks, $\Omega(b)$ denotes a group of frame indices in the b-th block, $|\Omega(b)|$ denotes the number of the frames included in $\Omega(b)$, and the target source must be inactive in all the frames in the blocks. In [3], an iterative update formula of the complementary beamformers $\boldsymbol{g}(\omega)$ and $\boldsymbol{h}(\omega)$ based on quasi-Newton method is proposed. However, this method has many local minima and its convergence is quite unstable. This is because of the indefiniteness of the solution. The optimum condition is that all the transfer function vectors $\boldsymbol{a}_d(\omega)$ are orthogonal to at least one of $\boldsymbol{g}(\omega)$ and $\boldsymbol{h}(\omega)$. A pair of the beamformers $\boldsymbol{g}(\omega)$ and $\boldsymbol{h}(\omega)$ to satisfy the complementary condition is not unique.

4.3. Signal Reconstruction with Over-Subtraction

To take advantage of the uncorrelated sources, time-averaging among neighboring frames is taken for the reference signal. In addition, for stronger noise-suppression ability with SS, over-subtraction is conducted, i.e., magnifying the estimate of the noise component by multiplying it with $\beta > 1$, and flooring it to 0 if the estimated signal power becomes negative:

$$|y(\omega)|^2 \equiv \begin{cases} \frac{1}{2}\cdot\left||y^{(\mathrm{p})}(\omega)|^2 - \beta\cdot\left\langle|y^{(\mathrm{r})}(\omega)|^2\right\rangle\right|^{1/2} \\ \quad \left(\text{if } |y^{(\mathrm{p})}(\omega)|^2 > \beta\cdot\left\langle|y^{(\mathrm{r})}(\omega)|^2\right\rangle\right), \\ 0 \ \ (\text{otherwise}). \end{cases} \tag{22}$$

5. PROPOSED ADAPTATION ALGORITHM WITH ANALYTICAL SOLUTION

5.1. Adaptation Problem as Matrix Optimization

As discussed in Sect. 4.2, the optimization procedure in the conventional method is unstable because of the indefiniteness of the solution. In this section, we propose a new closed-form solution to the adaptation problem by obtaining a matrix formed by the two complementary beamformers, without obtaining the beamformers themselves. It will be shown that such a matrix optimization is sufficient and the optimized matrix can give the same nonlinear array signal processing as the complementary beamformers.

The recovered signal $|y(\omega)|^2$ in Eq. (19) can be rewritten as

$$\begin{aligned} |y(\omega)|^2 &= (1/4)\,\boldsymbol{x}(\omega)^{\mathrm{H}}\left\{\boldsymbol{g}(\omega)^{\mathrm{H}}\boldsymbol{h}(\omega) + \boldsymbol{h}(\omega)^{\mathrm{H}}\boldsymbol{g}(\omega)\right\}\boldsymbol{x}(\omega) \\ &= (1/4)\,\boldsymbol{x}(\omega)^{\mathrm{H}}\boldsymbol{G}(\omega)\boldsymbol{x}(\omega), \end{aligned} \tag{23}$$

and optimizing $\boldsymbol{g}(\omega)$ and $\boldsymbol{h}(\omega)$ is equivalent to optimizing a Hermitian matrix

$$\boldsymbol{G}(\omega) = \boldsymbol{g}(\omega)^{\mathrm{H}}\boldsymbol{h}(\omega) + \boldsymbol{h}(\omega)^{\mathrm{H}}\boldsymbol{g}(\omega), \tag{24}$$

whose rank is two with a positive eigenvalue and a negative one. Such a matrix $\boldsymbol{G}(\omega)$ can be expressed as the combination of the autocorrelation of the training data $\boldsymbol{x}(\omega, l)$, $l = l_1, l_2, \ldots, l_L$ as

$$\boldsymbol{G}(\omega) = \sum_{l=l_1}^{l_L} c(\omega, l)\boldsymbol{x}(\omega, l)\boldsymbol{x}(\omega, l)^{\mathrm{H}}, \tag{25}$$

where $c(\omega, l)$ is the weighting coefficient to be optimized, which is real valued and can be positive or negative. With this expression, the optimization of $\boldsymbol{G}(\omega)$ is converted into the optimization of $c(\omega, l)$.

We utilize the same criterion in Eq. (21) as that in the conventional method. Equation (21) can be rewritten as

$$\begin{aligned} &J(\omega) \\ &= \sum_{b=1}^{B}\left\{\frac{1}{|\Omega(b)|}\sum_{t\in|\Omega(b)|}\frac{1}{4}\boldsymbol{x}(\omega,t)^{\mathrm{H}}\boldsymbol{G}(\omega)\boldsymbol{x}(\omega,t)\right\}^2 \\ &= \sum_{b=1}^{B}\left\{\sum_{t\in|\Omega(b)|}\sum_{l=l_1}^{l=l_L}\frac{1}{4|\Omega(b)|}c(\omega,l)\boldsymbol{x}(\omega,t)^{\mathrm{H}}\boldsymbol{x}(\omega,l)\boldsymbol{x}(\omega,l)^{\mathrm{H}}\boldsymbol{x}(\omega,t)\right\}^2 \\ &= \sum_{b=1}^{B}\left\{\sum_{t=l_1}^{l_L}\sum_{l=l_1}^{l_L}M(b,t)c(\omega,l)\boldsymbol{x}(\omega,l)^{\mathrm{H}}\boldsymbol{x}(\omega,t)\boldsymbol{x}(\omega,t)^{\mathrm{H}}\boldsymbol{x}(\omega,l)\right\}^2, \end{aligned} \tag{26}$$

where $M(b, t)$ is an averaging constant determined by the corresponding membership of the data $\boldsymbol{x}(\omega, t)$ in the b-th block and is given by

$$M(b,t) = \begin{cases} \frac{1}{4|\Omega(b)|} & \text{if } t \in \Omega(b), \\ 0 & \text{otherwise.} \end{cases} \tag{27}$$

In addition, $\boldsymbol{G}(\omega)$ must be constrained by the steering vector $\boldsymbol{r}(\omega)$ as in Eq. (20), lead to as

$$\boldsymbol{r}(\omega)^{\mathrm{H}}\boldsymbol{G}(\omega)\boldsymbol{r}(\omega) = 2. \tag{28}$$

Here $J(\omega)$ in Eq. (26) is rewritten in the form of matrix multiplication as

$$J(\omega) = \boldsymbol{c}(\omega)^{\mathrm{H}}\boldsymbol{K}(\omega)^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{K}(\omega)\boldsymbol{c}(\omega), \tag{29}$$

$$\boldsymbol{c}(\omega) = [c(\omega,l_1)\ \ c(\omega,l_2)\ \ \cdots\ \ c(\omega,l_L)]^{\mathrm{T}}, \tag{30}$$

$$\boldsymbol{K}(\omega) = \left[\boldsymbol{x}(\omega,l)^{\mathrm{H}}\boldsymbol{x}(\omega,t)\boldsymbol{x}(\omega,t)^{\mathrm{H}}\boldsymbol{x}(\omega,l)\right]_{tl}, \tag{31}$$

$$\boldsymbol{M} = [M(b,t)]_{bt}. \tag{32}$$

In the above expressions, $[x]_{ij}$ denotes a matrix, the element of which in the i-th column and the j-th row is specified by the substituting i and j in x. Also, the constraint of Eq. (28) is rewritten as

$$\boldsymbol{c}(\omega)^{\mathrm{H}}\boldsymbol{q}(\omega) = 2, \tag{33}$$

$$\begin{aligned} \boldsymbol{q}(\omega) = &[\boldsymbol{r}(\omega)^{\mathrm{H}}\boldsymbol{x}(\omega,l_1)\boldsymbol{x}(\omega,l_1)^{\mathrm{H}}\boldsymbol{r}(\omega), \ldots, \\ &\boldsymbol{r}(\omega)^{\mathrm{H}}\boldsymbol{x}(\omega,l_L)\boldsymbol{x}(\omega,l_L)^{\mathrm{H}}\boldsymbol{r}(\omega)]^{\mathrm{T}}. \end{aligned} \tag{34}$$

Minimization of $J(\omega)$ in Eq. (29) under the constraint of Eq. (33) can be accomplished by introducing a Lagrange multiplier $\gamma(\omega)$, and Lagrangian $\mathcal{L}(\omega)$ is then given by

$$\mathcal{L}(\omega) = \boldsymbol{c}(\omega)^{\mathrm{H}}\boldsymbol{K}(\omega)^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{K}(\omega)\boldsymbol{c}(\omega) + \gamma(\omega)\left\{\boldsymbol{c}(\omega)^{\mathrm{H}}\boldsymbol{q}(\omega) - 2\right\}. \tag{35}$$

The optimality condition satisfies

$$\begin{aligned} \frac{\partial\mathcal{L}(\omega)}{\partial\boldsymbol{c}(\omega)^{\mathrm{H}}} &= \boldsymbol{K}(\omega)^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{K}(\omega)\boldsymbol{c}(\omega) + \gamma(\omega)\boldsymbol{q}(\omega) = 0, \\ \frac{\partial\mathcal{L}(\omega)}{\partial\gamma(\omega)} &= \boldsymbol{c}(\omega)^{\mathrm{H}}\boldsymbol{q}(\omega) - 2 = 0, \end{aligned} \tag{36}$$

which leads to the following optimal solution:

$$\boldsymbol{c}(\omega) = \frac{2\left\{\boldsymbol{K}(\omega)^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{K}(\omega)\right\}^{+}\boldsymbol{q}(\omega)}{\boldsymbol{q}(\omega)^{\mathrm{H}}\left\{\boldsymbol{K}(\omega)^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{K}(\omega)\right\}^{+}\boldsymbol{q}(\omega)}, \tag{37}$$

where $\{\cdot\}^{+}$ denotes the Moore-Penrose pseudo inverse matrix of the argument. By substituting the obtained weighting coefficients $c(\omega, l)$ for $l = l_1, l_2, \ldots, l_L$ in Eq. (25), the optimized filter $\boldsymbol{G}(\omega)$ is obtained.

This formulation is inspired by an extension of principal component analysis to kernel method, i.e., the so-called kernel PCA [9]. In fact, by regarding each element $\boldsymbol{x}(\omega,l)^{\mathrm{H}}\boldsymbol{x}(\omega,t)\boldsymbol{x}(\omega,t)^{\mathrm{H}}\boldsymbol{x}(\omega,l)$ of $\boldsymbol{K}(\omega)$ as kernel function and substituting an identity matrix in $\boldsymbol{M}$, the proposed optimization can be interpreted as a kernel LCMV.

5.2. Factorization of Primary and Reference Path

Although the Hermitian matrix $\boldsymbol{G}(\omega)$ obtained in the previous section leads to minimization of the squared error, we have to separate the primary and reference paths from $\boldsymbol{G}(\omega)$ for the effective processing in Eq. (22). Because of the indefiniteness, $\boldsymbol{g}(\omega)$ and $\boldsymbol{h}(\omega)$ themselves cannot be obtained from $\boldsymbol{G}(\omega)$. However, the equivalent separate processing of the primary and reference paths can be accomplished by proper factorization of the Hermitian matrix.

Before the factorization, let us discuss about the primary and reference paths. Using another matrix filter $\boldsymbol{H}(\omega)$ defined as

$$\boldsymbol{H}(\omega) = \boldsymbol{g}(\omega)^{\mathrm{H}}\boldsymbol{g}(\omega) + \boldsymbol{h}(\omega)^{\mathrm{H}}\boldsymbol{h}(\omega), \tag{38}$$

we express the power of the primary and reference signals as

$$\begin{aligned} |y^{(\mathrm{p})}(\omega)|^2 &= \boldsymbol{x}(\omega)^{\mathrm{H}}\left\{\boldsymbol{g}(\omega)+\boldsymbol{h}(\omega)\right\}^{\mathrm{H}}\left\{\boldsymbol{g}(\omega)+\boldsymbol{h}(\omega)\right\}\boldsymbol{x}(\omega) \\ &= \boldsymbol{x}(\omega)^{\mathrm{H}}\left\{\boldsymbol{H}(\omega)+\boldsymbol{G}(\omega)\right\}\boldsymbol{x}(\omega), \end{aligned} \tag{39}$$

$$\begin{aligned} |y^{(\mathrm{r})}(\omega)|^2 &= \boldsymbol{x}(\omega^{\mathrm{H}}\left\{\boldsymbol{g}(\omega)-\boldsymbol{h}(\omega)\right\}^{\mathrm{H}}\left\{\boldsymbol{g}(\omega)-\boldsymbol{h}(\omega)\right\}\boldsymbol{x}(\omega) \\ &= \boldsymbol{x}(\omega)^{\mathrm{H}}\left\{\boldsymbol{H}(\omega)-\boldsymbol{G}(\omega)\right\}\boldsymbol{x}(\omega). \end{aligned} \tag{40}$$

Since $\boldsymbol{G}(\omega)$ is already computed, all we have to obtain is $\boldsymbol{H}(\omega)$ for separate filtering of the primary and reference signals.

Define the eigenvalue decomposition of the Hermitian matrix $\boldsymbol{G}(\omega)$ as

$$\boldsymbol{G}(\omega) = \boldsymbol{U}(\omega)\boldsymbol{E}(\omega)\boldsymbol{U}(\omega)^{\mathrm{H}}, \tag{41}$$

$$\begin{aligned} &\boldsymbol{E}(\omega) = \mathrm{Diag}\left[e_1(\omega), \cdots, e_K(\omega)\right] \\ &\left(|e_1(\omega)| \geq |e_2(\omega)| \geq \cdots \geq |e_K(\omega)|\right), \end{aligned} \tag{42}$$

$$\boldsymbol{U}(\omega) = \left[\boldsymbol{u}_1(\omega)\ \ \boldsymbol{u}_2(\omega)\ \ \cdots\ \ \boldsymbol{u}_K(\omega)\right]. \tag{43}$$

Since $\boldsymbol{G}(\omega)$ should be written as Eq. (24) and its rank should be two, we regard the later eigenvalues than the third to be 0 as

$$\boldsymbol{G}(\omega) = e_1(\omega)\boldsymbol{u}_1(\omega)\boldsymbol{u}_1(\omega)^{\mathrm{H}} + e_2(\omega)\boldsymbol{u}_2(\omega)\boldsymbol{u}_2(\omega)^{\mathrm{H}}. \tag{44}$$

Moreover, $\boldsymbol{g}(\omega)$ and $\boldsymbol{h}(\omega)$ should be written as

$$\boldsymbol{g}(\omega) = \kappa_1(\omega)\boldsymbol{u}_1(\omega)^{\mathrm{H}} + \kappa_2(\omega)\boldsymbol{u}_2(\omega)^{\mathrm{H}}, \tag{45}$$

$$\boldsymbol{h}(\omega) = \lambda_1(\omega)\boldsymbol{u}_1(\omega)^{\mathrm{H}} + \lambda_2(\omega)\boldsymbol{u}_2(\omega)^{\mathrm{H}}, \tag{46}$$

where $\kappa_i(\omega)$ and $\lambda_i(\omega)$ for $i = 1, 2$ are unknown constant values. Substitution of Eqs. (45) and (46) in Eqs. (24) and (38) gives

$$\begin{aligned}\boldsymbol{G}(\omega) =& (\kappa_1(\omega)^*\lambda_1(\omega) + \lambda_1(\omega)^*\kappa_1(\omega))\boldsymbol{u}_1(\omega)\boldsymbol{u}_1(\omega)^{\mathrm{H}} + \\ & (\kappa_1(\omega)^*\lambda_2(\omega) + \lambda_1(\omega)^*\kappa_2(\omega))\boldsymbol{u}_1(\omega)\boldsymbol{u}_2(\omega)^{\mathrm{H}} + \\ & (\kappa_2(\omega)^*\lambda_1(\omega) + \lambda_2(\omega)^*\kappa_1(\omega))\boldsymbol{u}_2(\omega)\boldsymbol{u}_1(\omega)^{\mathrm{H}} + \\ & (\kappa_2(\omega)^*\lambda_2(\omega) + \lambda_2(\omega)^*\kappa_2(\omega))\boldsymbol{u}_2(\omega)\boldsymbol{u}_2(\omega)^{\mathrm{H}},\end{aligned} \tag{47}$$

$$\begin{aligned}\boldsymbol{H}(\omega) =& (\kappa_1(\omega)^*\kappa_1(\omega) + \lambda_1(\omega)^*\lambda_1(\omega))\boldsymbol{u}_1(\omega)\boldsymbol{u}_1(\omega)^{\mathrm{H}} + \\ & (\kappa_1(\omega)^*\kappa_2(\omega) + \lambda_1(\omega)^*\lambda_2(\omega))\boldsymbol{u}_1(\omega)\boldsymbol{u}_2(\omega)^{\mathrm{H}} + \\ & (\kappa_2(\omega)^*\kappa_1(\omega) + \lambda_2(\omega)^*\lambda_1(\omega))\boldsymbol{u}_2(\omega)\boldsymbol{u}_1(\omega)^{\mathrm{H}} + \\ & (\kappa_2(\omega)^*\kappa_2(\omega) + \lambda_2(\omega)^*\lambda_2(\omega))\boldsymbol{u}_2(\omega)\boldsymbol{u}_2(\omega)^{\mathrm{H}}.\end{aligned} \tag{48}$$

From Eqs. (44) and (47), the following conditions are obtained:

$$\lambda_1(\omega)^*\kappa_1(\omega) + \kappa_1(\omega)^*\lambda_1(\omega) = e_1(\omega), \tag{49}$$

$$\lambda_1(\omega)^*\kappa_2(\omega) + \kappa_1(\omega)^*\lambda_2(\omega) = 0, \tag{50}$$

$$\kappa_2(\omega)^*\lambda_2(\omega) + \lambda_2(\omega)^*\kappa_2(\omega) = e_2(\omega). \tag{51}$$

In addition, the condition Eq. (20) gives other conditions;

$$\kappa_1\boldsymbol{u}_1(\omega)^{\mathrm{H}}\boldsymbol{r}(\omega) + \kappa_2\boldsymbol{u}_2(\omega)^{\mathrm{H}}\boldsymbol{r}(\omega) = 1, \tag{52}$$

$$\lambda_1\boldsymbol{u}_1(\omega)^{\mathrm{H}}\boldsymbol{r}(\omega) + \lambda_2\boldsymbol{u}_2(\omega)^{\mathrm{H}}\boldsymbol{r}(\omega) = 1. \tag{53}$$

The conditions in Eqs. (49)–(53) cannot define unique solution of $\kappa_1(\omega)$, $\kappa_2(\omega)$, $\lambda_1(\omega)$ and $\lambda_2(\omega)$ but uniquely gives

$$\begin{aligned}&\kappa_1(\omega)^*\kappa_1(\omega) + \lambda_1(\omega)^*\lambda_1(\omega) \\ &= |\boldsymbol{u}_1(\omega)^{\mathrm{H}}\boldsymbol{r}(\omega)|^2 e_1(\omega)^2 - e_1(\omega),\end{aligned} \tag{54}$$

$$\begin{aligned}&\kappa_1(\omega)^*\kappa_2(\omega) + \lambda_1(\omega)^*\lambda_2(\omega) \\ &= \frac{2 - |\boldsymbol{u}_1(\omega)^{\mathrm{H}}\boldsymbol{r}(\omega)|^2 e_1(\omega)}{\boldsymbol{u}_2(\omega)^{\mathrm{H}}\boldsymbol{r}(\omega)}\boldsymbol{u}_1(\omega)^{\mathrm{H}}\boldsymbol{r}(\omega)e_1(\omega),\end{aligned} \tag{55}$$

$$\begin{aligned}&\kappa_2(\omega)^*\kappa_2(\omega) + \lambda_2(\omega)^*\lambda_2(\omega) \\ &= \left|\frac{2 - \left|\boldsymbol{u}_1(\omega)^{\mathrm{H}}\boldsymbol{r}(\omega)\right|^2 e_1(\omega)}{\boldsymbol{u}_2(\omega)^{\mathrm{H}}\boldsymbol{r}(\omega)}\right| - e_2(\omega).\end{aligned} \tag{56}$$

By substituting Eqs. (54)–(56) in Eq. (48), $\boldsymbol{H}(\omega)$ is obtained. The speech enhancement is conducted with substitution of Eqs. (39) and (40) in the spectral subtraction of Eq. (22).

6. EXPERIMENT

We compared the performance of the proposed adaptation algorithm with that of two previous methods, namely, a traditional LCMV beamformer, to gain insights on the effectiveness of non-linear beamforming, and the conventional iterative adaptation of nonlinear beamforming of [3], to investigate the quality of the solutions. In addition, we compared the performance of the conventional adaptation of nonlinear microphone arrays with different initial values to evaluate its convergence properties.

The observed signals are made in simulation using measured impulse responses at a room with a reverberation time (T60) of 200 ms. The number of microphone elements is set at two, and the inter-element spacing is 2 cm. The sampling frequency is 16 kHz, the frame length 32 ms, the FFT length 2048, and the block size for averaging is 50 frames with a block shift of one frame. The observation is a convolutive mixture of three speech utterances; one of them is chosen as the target signal. The target speech signal is assumed to arrive from the look direction of 0° and the interfering noise is assumed to arrive from two directions at -40° and 30°, respectively. β is set to be 3.

For the conventional iterative adaptation, four conditions of initialization were evaluated. The condition 1 is the initialization recommended in the paper [3]: all the 153 pairs of null beamformers out of those who have equally-sampled 18 different directional nulls are used as initial beamformer pairs, and they are optimized in parallel; then the best one in each frequency bin is adopted as the optimized complementary beamformers. In the condition 2, two null beamformers given the true look directions of the interfering speech sources are used as the initial values. In the condition 3, null beamformers with directions at -60° and 60° are used as the initial values. In the condition 4, random values are used and we evaluated the average of the 20 trials. Evaluation scores are based on the noise reduction rate (NRR) [3, 5], to evaluate the noise reduction performance, and the cepstral distance (CD) [10], to evaluate the speech quality objectively.

Table 1. Experimental result

Method	NRR [dB]	CD [dB]
LCMV	0.60	1.09
Condition 1	10.42	6.05
Condition 2	10.33	6.97
Condition 3	5.03	5.19
Condition 4	1.00	4.60
Proposed	9.15	6.98

Table 1 shows the evaluation result. LCMV failed to cancel the interfering signals because there were too many sources. The poor performances related to the conditions 3 and 4 show the poor convergence property of the conventional method. The conditions 1 and 2 with the good initial values perform well. However, the former incurs a large computational cost and the latter requires very accurate estimation of the directions of arrival, which is hard to accomplish with the large number of sources compared with the number of the sensors. The performance of the proposed method is as good as the best of the conventional method, yet with a better computational efficiency due to the analytical solution. Furthermore, the proposed method does not require any estimation of the initial values. Thus the proposed algorithm is considered the more reasonable than the conventional iterative algorithm.

7. CONCLUSIONS

We have proposed a new algorithm to solve optimization problem in adaptive nonlinear microphone array based on complementary beamforming. The iterative procedure of the conventional method based on the quasi-Newton algorithm has poor convergence behavior because of the indefiniteness of the intermediate solutions. We reformulate the problem as matrix optimization, a solution to which can be obtained in analytical form. Experimental result shows effectiveness and efficiency of the proposed method. Our future project is extension of the proposed method to unsupervised optimization based on independent component analysis.

8. REFERENCES

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

[2] M. Mizumachi and M. Akagi, "Noise reduction by paired-microphones using spectral subtraction," *Proc. ICASSP,* vol.2, pp.1001–1004, 1998.

[3] H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Speech enhancement using nonlinear microphone array based on noise adaptive complementary beamforming," *IEICE Trans. Fundam.,* vol.E83-A, no.5, pp.866–876, 2000.

[4] H. Kamiyanagida *et al.*, "Direction of arrival estimation using nonlinear microphone array," *IEICE Trans. Fundam.,* vol.E84-A, no.4, pp.999-1010, 2001.

[5] Y. Takahashi, T. Takatani, H. Saruwatari, and K. Shikano, "Blind spatial subtraction array with independent component analysis for hands-free speech recognition," *Proc. IWAENC,* 2006.

[6] H. Shimizu, N. Ono, K. Matsumoto, and S. Sagayama, "Isotropic Noise Suppression on Power Spectrum Domain by Symmetric Microphone Array," *Proc. WASPAA,* pp.54–57, 2007

[7] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Communication,* vol.49, no.7–8, pp.636–656, 2007.

[8] A. Hussain *et al.,* "Nonlinear Speech Enhancement: An Overview," *Proc. WNSP*, pp.2170-248, 2005.

[9] B. Scholkopf, A. Smola, and K.-R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation,* vol.10, pp.1299-1319, 1998.

[10] L. Rabiner and B.H. Juang, *Fundamentals of speech recognition*, Prentice Hall, Englewood Cliffs, NJ, 1993.