KERNEL-BASED NONLINEAR INDEPENDENT COMPONENT ANALYSIS FOR UNDERDETERMINED BLIND SOURCE SEPARATION

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

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

ABSTRACT

In this paper we propose a new unsupervised training method for nonlinear spatial filter using a new independent component analysis based on kernel infomax. The nonlinearity of the spatial filter used in this paper is equivalent to the integration of beamforming and spectral subtraction, and the whole structure is optimized by independent component analysis in the reproducing kernel Hilbert space. The optimized filter is shown to be capable of achieving better quality output than the conventional method based on time-frequency binary masking.

Index Terms— Blind source separation, independent component analysis, reproducing kernel Hilbert space, underdetermined problem, beamforming.

1. INTRODUCTION

Blind source separation (BSS) has been widely studied in over the past decade, and is expected to be an important tool in many speech applications [1]. Recently several research groups succeeded in implementations of real-time BSS with high performance. However, there still remain many issues about robustness and limitations for practical use of the system. This paper focuses on the problem of the number of the sources to be separated.

There are mainly two approaches in BSS, namely, independent component analysis (ICA) [1] and time-frequency binary masking (TFBM) [2]. ICA is an unsupervised training framework of beamformers and has wide variations in the domain where ICA is trained, i.e., time domain or frequency domain [3], and in the criterion, i.e., mutual information, temporal correlations and non-stationarity [4]. The performance of ICA is bounded by the limitation of beamforming, which is the underlying physical mechanism of the source separation accomplished by ICA. In ICA, the number of the sources has to be smaller than or equal to that of the sensors so that all sources except the target can be excluded from the output signals by constructing directional nulls. Also, the sources have to be of a point source type which can be cancelled by a directional null. Under the situation where these conditions are satisfied, ICA has a potential to perform extremely well. TFBM method implements discrete suppression of little overlap of signal activities from different sources in the time-frequency domain assuming very small overlap of the amplitude among sources. This mechanism does not assume knowledge of the number of sources, and can solve the problem of the so-called underdetermined BSS, where more sources than the sensors exist. However, the quality is not very high if the assumption does not hold. Also, filtering of TFBM does not take advantage the number of sensors like ICA does.

As a candidate for the solution of underdetermined problem, one of the authors has proposed a nonlinear beamformer, i.e., integration of linear beamforming and nonlinear processing of spectral subtraction [5]. Note that in this paper, we use the terms 'linear' or 'nonlinear' from the viewpoint of signal processing aspect, not array configuration. This method can cancel twice as many sources as a linear beamformer with the same number of sensors can. Processing in the method *does not assume sparseness* in the signal activity from different sources; rather, it is a continuous processing and capable of separating signals even with overlapping time-frequency activities without creating serious artifacts. We have have also proposed a sufficiently analytical adaaptation algorithm of the nonlinear beamformer using kernel method [6]. Kernel method is an optimization technique of nonlinear function, and is popular in the research field of machine learning [7]. By utilizing nonlinear mapping to the higher dimensional Hilbert space so-called reproducing kernel Hilbert space (RKHS) [8], a linear optimization technique is extended to a nonlinear optimization in a straightforward manner. In this paper we propose a new underdetermined BSS solution by adaptation of nonlinear beamformer with ICA in the RKHS. We derive ICA of the class of infomax with natural gradient [9] in the RKHS, and optimize the nonlinear beamformer in an unsupervised way.

Here we describe the difference of the proposed method from the other kernel-based ICA algorithms. Several research groups have studied independence measure with kernel method [10]. These researches aim to linear separation problem unlike our purpose. Another group has already formulated nonlinear ICA with kernel method with non-Gaussianity criterion [11]. However, the output here ends up with feature extraction, and the reconstruction of separated signal in the observation space is not obtained because of an inappropriate choice of kernel and normalization in the RKHS. Also, although the difference may not be significant, the Kullback-Leibler divergence with suitable assumption in the marginal distribution we use is more suitable criterion than non-Gaussianity, because Gaussian signal in the observation space is not Gaussian any more in the RKHS because of the nonlinear scaling in the higher-dimensional mapping.

2. BLIND SOURCE SEPARATION

2.1. Problem description

Observations of multiple sound mixture captured by distant-talking microphones are modeled as convolutive mixtures. In this paper, we discuss the separation of such mixture in the frequency domain, where the convolution is modeled as a simple memoryless multiplication. The observed signal $x_m(\omega)$ of the angular frequency ω at the m-th microphone generated by N sources $s_n(\omega)$ for $n = 1, \dots, N$ is written as

$$x_m(\omega) = \sum_{n}^{N} h_{mn}(\omega) s_n(\omega), \tag{1}$$

where $h_{mn}(\omega)$ is a transfer function from the n-th source to the m-th microphone. Using matrix expression, the observation with M microphones is given by

$$\boldsymbol{x}(\omega) = [x_1(\omega) \ \cdots \ x_M(\omega)]^{\mathrm{T}} = \boldsymbol{H}(\omega)\boldsymbol{s}(\omega) = \sum_{n=1}^{N} \boldsymbol{h}_n(\omega) s_n(\omega), \tag{2}$$

$$\boldsymbol{H}(\omega) = [h_{mn}(\omega)]_{mn} = [\boldsymbol{h}_1(\omega) \ \cdots \ \boldsymbol{h}_N(\omega)], \tag{3}$$

$$\boldsymbol{s}(\omega) = [s_1(\omega) \ \cdots \ s_N(\omega)]^{\mathrm{T}}, \tag{4}$$

where $\{\cdot\}^{\mathrm{T}}$ denotes matrix transposition and $[x]_{ij}$ denotes a matrix which has an entry x in the i-th row and the j-th column. The goal of the BSS is to obtain an estimate $y_n(\omega)$ of $s_n(\omega)$ only from the observed sequences such that

$$y_n(\omega) \approx s_n(\omega). \tag{5}$$

2.2. ICA-based approach

BSS of convolutive mixture based on ICA can be interpreted as simultaneous unsupervised adaptation of multiple beamformers. Since the separation is obtained by linear FIR filtering, this method does not suffer from unpleasant artifact effects caused by nonlinear filtering. Its disadvantage, though, is that the separation is limited

* Research Fellow of the Japan Society for the Promotion of Science.

by directional nulls that the beamformers can construct. ICA with M microphones can separate less than or equal to M sources, and the sources must be strictly point sources.

Here we describe BSS of convolutive mixture using infomax algorithm in the frequency domain [3], which is shown to be effective in the application of speech separation [12]. The ICA separates the output signals $\boldsymbol{y}(\omega)$ with an $N \times M$ filter matrix $\boldsymbol{W}(\omega)$ as

$$\boldsymbol{y}(\omega) = [y_1(\omega) \ \cdots \ y_N(\omega)]^{\mathrm{T}} = \boldsymbol{W}(\omega)\boldsymbol{x}(\omega). \tag{6}$$

Each row of $\boldsymbol{W}(\omega)$ constructs a beamformer to enhance a different source. Assuming statistical independence among the original sources $s_n(\omega)$, achieving independence among output signals through $\boldsymbol{W}(\omega)$ leads to successful separation [1]. To accomplish maximization of independence, we first obtain the a sequence $\boldsymbol{x}(\omega, t)$ of short-time Fourier analysis of the observed signal $\boldsymbol{x}(\omega)$, and assume the following approximation:

$$\boldsymbol{y}(\omega, t) \approx \boldsymbol{W}(\omega)\boldsymbol{x}(\omega, t). \tag{7}$$

We then update the demixing matrix $\boldsymbol{W}(\omega)$ to enhance a prescribed independence criterion. A typical solution to this problem is infomax with natural gradient [3]:

$$\boldsymbol{W}(\omega) \leftarrow \boldsymbol{W}(\omega) + \mu\left(\boldsymbol{I}_L - E\left[\boldsymbol{f}\left(\boldsymbol{y}(\omega,t)\right)\boldsymbol{y}(\omega,t)^{\mathrm{H}}\right]_t\right)\boldsymbol{W}(\omega), \tag{8}$$

where $E[\cdot]_t$ is expectation over t, $\{\cdot\}^{\mathrm{H}}$ is conjugate transposition, and $\boldsymbol{f}(\cdot)$ is the J-dimensional vector of nonlinear function given by

$$\boldsymbol{f}(\boldsymbol{y}(\omega,t)) = \left[f\left(y_1(\omega,t)\right) \ \cdots \ f\left(y_J(\omega,t)\right)\right]^{\mathrm{T}}. \tag{9}$$

The nonlinear function $f(\cdot)$ must beproperly chosen and consistent with the model on which prescribed independence criterion can be sensibly evaluated. Here the independence criterion does not specify the output order in ω. This is the so-called permutation problem [13]. Hence, to extract the separated sources in the time domain, after the iterative optimization, the rows of $\boldsymbol{W}(\omega)$ have to be reordered to produce the same source across the frequency.

2.3. TFBM-based approach

TFBM [2] is a nonlinear filtering with discrete classification for underdetermined BSS problem assuming sparseness among sources. The signal components in the time-frequency domain are masked to be zero when they are not classified to the target singal sources. This method can realize underdetermined source separation; the number of the sources N can be larger than M of the microphones. However, the performance degrades when sparseness is not clear because of reverberation or existence of non-harmonic signal sources such as stationary noise. The residual component of the nonlinear signal processing appears as annoying artifacts such as musical noise. The masking itself is a single-channel processing and the number of the microphnes does not benefit the performance or the quality.

3. NONLINEAR ICA ALGORITHM

3.1. Reproducing kernel Hilbert spaces

In this section, we introduce the kernel method to give the geometric background of underdetermined source separation with nonlinear beamforming. The kernel method employs nonlinear mapping from the observation space to another Hilbert space of a higher, possibly infinite dimensionality. As a result, the dimension of the null realizable space increases with the higher-dimensional mapping. By constructing beamformers in the higher-dimensional RKHS, the beamformers can deal with a larger number of sources.

Figure 1 shows an illustration of mapping with the kernel method. By using a positive-definite kernel function $K(\boldsymbol{x}_1, \boldsymbol{x}_2)$ defined over two vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{C}^M$, we can construct a Hilbert space $\mathcal{H}$. In particular, a vector $\boldsymbol{x} \in \mathbb{C}^M$ is mapped with nonlinear function $\boldsymbol{\phi} : \mathbb{C}^M \to \mathcal{H}$, and the inner product of the two mapped vectors $\boldsymbol{\phi}(\boldsymbol{x}_1)$ and $\boldsymbol{\phi}(\boldsymbol{x}_2)$ is prescribed as the kernel function:

$$\boldsymbol{\phi}(\boldsymbol{x}_1)^{\mathrm{H}}\boldsymbol{\phi}(\boldsymbol{x}_2) = K(\boldsymbol{x}_1, \boldsymbol{x}_2). \tag{10}$$

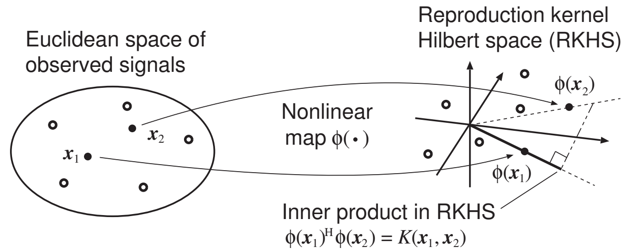

Fig. 1. Relation among the RKHS $\mathcal{H}$, the kernel function $K(\cdot,\cdot)$, and nonlinear mapping $\boldsymbol{\phi}(\cdot)$. Definition of a positive-definite kernel function $K(\boldsymbol{x}_1, \boldsymbol{x}_2)$ ($\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{C}^M$) corresponds to definition of an RKHS $\mathcal{H}$, and the kernel function works as an inner product between $\boldsymbol{\phi}(\boldsymbol{x}_1)$ and $\boldsymbol{\phi}(\boldsymbol{x}_2)$ with the mapping $\boldsymbol{\phi} : \mathbb{C}^M \to \mathcal{H}$. The vector $\boldsymbol{\phi}(\boldsymbol{x})$ in the RKHS is expressed as a functional $K(\boldsymbol{x}, \cdot)$, and we do not need to obtain the mapping function $\boldsymbol{\phi}$ itself.

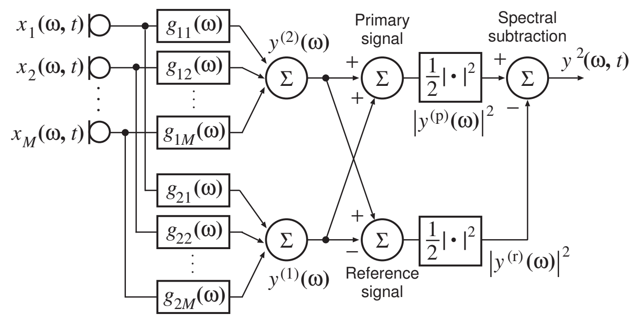

Fig. 2. Configuration of Nonlinear beamformer integrated with spectral subtraction.

One of the most popular positive-definite kernel functions in machine learning is the d-th order polynomial kernel which is also used in [11], given by

$$K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \left(\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{x}_2 + 1\right)^d. \tag{11}$$

Since linearity is satisfied in the RKHS, we can formulate the optimization problem in $\mathcal{H}$ with $K(\cdot,\cdot)$ but mapping $\boldsymbol{\phi}$ itself. Generally the dimensionality of $\mathcal{H}$ is higher than M of the vectors that are mapped from. This is a useful feature for pattern classification problems because linear separability of the classes is improved by the increase of dimensionality [7].

3.2. Nonlinear beamforming with a quadratic kernel

Among many kernel functions to realize higher-dimensional mapping, we have to choose the kernel carefully because the mapping should maintain both the high-dimensional and the relation of the signal bases in the observation space as much as possible. In fact, the polynomial kernel in Eq. (11) is not suitable. If the directional null in the RKHS has no relationship to the directivity in the observation space, the optimization is trivial. In this paper we use the following kernel function:

$$K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_2^{\mathrm{H}}\boldsymbol{x}_1\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{x}_2. \tag{12}$$

We will call this kernel function *quadratic kernel.* As noted earlier, one of the authors proposed a nonlinear beamformer integrated with spectral subtraction [5], which can deal with $2(M-1)$ sources. Recently we have discovered an equivalence between this beamforming and the quadratic kernel in [6], where we have formulated a sufficient supervised adaptation of the nonlinear beamformer using this kernel function.

Figure 2 shows the configuration of the nonlinear beamformer. This method constructs two different linear adaptive beamformers

$\boldsymbol{g}_1(\omega)$ and $\boldsymbol{g}_2(\omega)$, and the output $y^2(\omega)$ of the nonlinear beamforming is a result of spectral subtraction between sum and difference of the two beamformers, written as

$$\begin{aligned} y^2(\omega,t) &= \frac{1}{2}\left|\left(\boldsymbol{g}_1(\omega)+\boldsymbol{g}_2(\omega)\right)\boldsymbol{x}(\omega,t)\right|^2 \\ &\quad - \frac{1}{2}\left|\left(\boldsymbol{g}_1(\omega)-\boldsymbol{g}_2(\omega)\right)\boldsymbol{x}(\omega,t)\right|^2 \\ &= \boldsymbol{x}(\omega,t)^{\mathrm{H}}\boldsymbol{G}(\omega)\boldsymbol{x}(\omega,t), \end{aligned} \tag{13}$$

$$\boldsymbol{g}_l(\omega) = [g_{l1}(\omega)\ \cdots\ g_{lN}(\omega)] \quad \text{for} \quad l=1,2, \tag{14}$$

$$\boldsymbol{G}(\omega) = \boldsymbol{g}_1(\omega)^{\mathrm{H}}\boldsymbol{g}_2(\omega)+\boldsymbol{g}_2(\omega)^{\mathrm{H}}\boldsymbol{g}_1(\omega). \tag{15}$$

Thus the adaptation of $\boldsymbol{g}_1(\omega)$ and $\boldsymbol{g}_2(\omega)$ is equivalent to that of the symmetric matrix $\boldsymbol{G}(\omega)$. Moreover, $\boldsymbol{G}(\omega)$ can be expressed by linear combination of training data $\boldsymbol{x}(\omega,l)$ with the weighting factor $\alpha(\omega,l)$ for the frame index $l=l_1,\ldots,l_L$ of the training data, as

$$\boldsymbol{G}(\omega) = \sum_{l=l_1}^{l_L}\alpha(\omega,l)\boldsymbol{x}(\omega,l)\boldsymbol{x}(\omega,l)^{\mathrm{H}}, \tag{16}$$

and optimization of $\boldsymbol{G}(\omega)$ leads to that of $\omega(\omega,l)$. Using the kernel expression of Eq. (12), Eq. (13) is rewritten as the output of linear beamformer $\boldsymbol{\gamma}(\omega)^{\mathrm{H}}$ in the RKHS as

$$\begin{aligned} y^2(\omega,t) &= \boldsymbol{\gamma}(\omega)^{\mathrm{H}}\boldsymbol{\phi}\left(\boldsymbol{x}(\omega,t)\right) \\ &= \sum_{l=l_1}^{l_L}\alpha(\omega,l)K\left(\boldsymbol{x}(\omega,l),\boldsymbol{x}(\omega,t)\right) \end{aligned} \tag{17}$$

with the beamformer

$$\boldsymbol{\gamma}(\omega)^{\mathrm{H}} = \sum_{l=l_1}^{l_L}\alpha(\omega,l)\boldsymbol{\phi}\left(\boldsymbol{x}(\omega,l)\right)^{\mathrm{H}}. \tag{18}$$

Also, by conducting oversubtraction [5], the nonlinear beamformer obtains a stronger noise reduction ability and robustness against non-point source. To conduct oversubtraction, we need to filter of primary and secondary paths, which is written in [6] and omitted in this paper.

3.3. ICA in RKHS: Kernel infomax

The proposed algorithm in this paper assumes that the quadratic kernel in Eq.(12) is used. However, since the infomax-based formulation of ICA in the RKHS has not been introduced to the best of our knowledge, we generalize the problem and derive a kernel infomax without specifying the type of kernel. The objective is to separate the mixture $\boldsymbol{\phi}(\boldsymbol{x}(\omega))$ in $\mathcal{H}$ into N statistically independent signals $\boldsymbol{\upsilon}(\omega)$ by projecting it onto N vectors $\boldsymbol{\gamma}_n(\omega)$, $n=1,\ldots,N$:

$$\upsilon_n(\omega,t) = \boldsymbol{\gamma}_n(\omega)^{\mathrm{H}}\boldsymbol{\phi}\left(\boldsymbol{x}(\omega,t)\right), \tag{19}$$

or in the matrix form,

$$\begin{aligned} \boldsymbol{\upsilon}(\omega,t) &= [\upsilon_1(\omega,t)\ \cdots\ \upsilon_N(\omega,t)]^{\mathrm{T}} \\ &= \boldsymbol{\Gamma}(\omega)\boldsymbol{\phi}(\boldsymbol{x}(\omega,t)), \end{aligned} \tag{20}$$

where $\boldsymbol{\Gamma}(\omega)$ is a linear operator to map from $\mathcal{H}$ to $\mathbb{C}^N$, written as

$$\boldsymbol{\Gamma}(\omega) = [\boldsymbol{\gamma}_1(\omega)\ \cdots\ \boldsymbol{\gamma}_N(\omega)]^{\mathrm{H}}. \tag{21}$$

Here the domain of $\boldsymbol{\Gamma}(\omega)$ can be limited to the subspace spanned by the training data $\boldsymbol{\phi}\left(\boldsymbol{x}(\omega,l)\right)$. Since the RKHS $\mathcal{H}$ satisfies linearity, we express $\boldsymbol{\gamma}_n(\omega)$ by a linear combination of the mapped training data, similarly to almost all of the kernel methods, as

$$\boldsymbol{\gamma}_n(\omega)^{\mathrm{H}} = \sum_{l=l_1}^{l_L}\alpha_{nl}(\omega)\boldsymbol{\phi}\left(\boldsymbol{x}(\omega,l)\right)^{\mathrm{H}}, \tag{22}$$

$$\boldsymbol{\Gamma}(\omega) = [\boldsymbol{\gamma}_1(\omega)\ \cdots\ \boldsymbol{\gamma}_N(\omega)]^{\mathrm{H}} = \boldsymbol{A}(\omega)\boldsymbol{\kappa}(\omega), \tag{23}$$

$$\boldsymbol{A}(\omega) = [\alpha_{nl}(\omega)]_{nl}, \tag{24}$$

$$\boldsymbol{\kappa}(\omega) = [\boldsymbol{\phi}\left(\boldsymbol{x}(\omega,l_1)\right)\ \cdots\ \boldsymbol{\phi}\left(\boldsymbol{x}(\omega,l_L)\right)]^{\mathrm{H}}, \tag{25}$$

where $\alpha_{nl}(\omega)$ is a real weighting factor. Thus the optimization of $\boldsymbol{\Gamma}(\omega)$ results in optimization of $\alpha_{nl}(\omega)$.

Since the joint probability and the product of the marginal probabilities are equal when the random variables are independent, infomax utilizes the Kullback-Leibler divergence (KLD) between joint and marginal distributions of the output signals as measure of independence. Such a KLD about the product of the $\boldsymbol{\upsilon}(\omega)$ is written as

$$I(\boldsymbol{\upsilon}(\omega)) = \int p(\boldsymbol{\upsilon}(\omega,t))\log\frac{p\left(\boldsymbol{\upsilon}(\omega,t)\right)}{\prod_{n=1}^{N}p\left(\upsilon_n(\omega,t)\right)}\mathrm{d}\boldsymbol{\upsilon}(\omega,t). \tag{26}$$

With the step-size parameter $\mu(\omega)$, the update formula of $\boldsymbol{\Gamma}(\omega)$ is given by the gradient of the KLD as

$$\boldsymbol{\Gamma}(\omega) \leftarrow \boldsymbol{\Gamma}(\omega) - \mu\frac{\partial I\left(\boldsymbol{\upsilon}(\omega)\right)}{\partial\boldsymbol{\Gamma}(\omega)}. \tag{27}$$

Since $\boldsymbol{\phi}\left(\boldsymbol{x}(\omega)\right)$ is mapped to q J-dimensional space linearly from $\mathcal{H}$ with more than J dimensions, the problem corresponds to an overdetermined source separation problem. Thus we can substitute the parameters in the update formula for overdetermined infomax with natural gradient [9] as

$$\boldsymbol{\Gamma}(\omega) \leftarrow \boldsymbol{\Gamma}(\omega) + \mu\left(\boldsymbol{I}_J - E\left[\boldsymbol{f}\left(\boldsymbol{\upsilon}(\omega,t)\right)\boldsymbol{\upsilon}(\omega,t)^{\mathrm{H}}\right]_t\right)\boldsymbol{\Gamma}(\omega), \tag{28}$$

where $\boldsymbol{I}_J$ is J-dimensional identity matrix. Utilizing the relation

$$[\upsilon_n(\omega,l)]_{nl} = \boldsymbol{A}(\omega)\boldsymbol{K}(\omega), \tag{29}$$

with $n=1,\ldots,N$, $l=l_1,\ldots,l_L$, where

$$\boldsymbol{K}(\omega) = \boldsymbol{\kappa}(\omega)\boldsymbol{\kappa}(\omega)^{\mathrm{H}}, \tag{30}$$

the expectation is substituted by the sample average as

$$\begin{aligned} E\left[\boldsymbol{f}\left(\boldsymbol{y}(\omega,t)\right)\boldsymbol{y}(\omega,t)^{\mathrm{H}}\right]_t &\approx \frac{1}{L}\sum_{l=l_1}^{l_L}\boldsymbol{f}\left(\boldsymbol{y}(\omega,l)\right)\boldsymbol{y}(\omega,l)^{\mathrm{H}} \\ &= \frac{1}{L}\boldsymbol{F}\left(\boldsymbol{A}(\omega)\boldsymbol{K}(\omega)\right)\boldsymbol{K}(\omega)^{\mathrm{H}}\boldsymbol{A}(\omega)^{\mathrm{H}}, \end{aligned} \tag{31}$$

where

$$\boldsymbol{F}\left([x_{ij}]_{ij}\right) = [f\left(x_{ij}\right)]_{ij}. \tag{32}$$

Substituting Eq. (23) and multiplying $\boldsymbol{\kappa}(\omega)^{\mathrm{H}}\boldsymbol{K}(\omega)^{+}$, where $\{\cdot\}^{+}$ is the Moore-Penrose Inverse matrix, from the right side of Eq. (28), equivalent update formula of $\boldsymbol{A}(\omega)$ is given by

$$\boldsymbol{A}(\omega) \leftarrow \boldsymbol{A}(\omega) - \mu\left(\boldsymbol{I}_J - \frac{1}{L}\boldsymbol{F}\left(\boldsymbol{A}(\omega)\boldsymbol{K}(\omega)\right)\boldsymbol{K}(\omega)\boldsymbol{A}(\omega)^{\mathrm{H}}\right)\boldsymbol{A}(\omega). \tag{33}$$

Note that we do not normalize signals in the RKHS, which is recommended in the paper [11], and in fact popular in kernel methods to analyze subspaces, e.g., [10]. The normalization is applicable even in this algorithm by modifying the Gramian matrix $\boldsymbol{K}(\omega)$. However, we did not apply normalization to retain the geometric relationship among the observed vectors as much as possible, which is important to recover the separated signal in the time domain. If the observed signal has a bias, normalization should be done in the observation space.

3.4. System identification to reformat the separation filters

By using the weighting factor obtained by kernel infomax with $K(\boldsymbol{x}_1,\boldsymbol{x}_2) = \boldsymbol{x}_2^{\mathrm{H}}\boldsymbol{x}_1\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{x}_2$ in Eq. (12), the estimate $y_n^2(\omega,t)$ of the n-th source is obtained as

$$y_n^2(\omega,t) = \boldsymbol{x}(\omega,t)^{\mathrm{H}}\boldsymbol{G}_n(\omega)\boldsymbol{x}(\omega,t), \tag{34}$$

$$\boldsymbol{G}_n(\omega) = \sum_{l=l_1}^{l_L}\alpha_n(\omega,l)\boldsymbol{x}(\omega,l)\boldsymbol{x}(\omega,l)^{\mathrm{H}}. \tag{35}$$

However, $\boldsymbol{G}_n(\omega)$ still has several problems to be used as a beamformer. First, similar to the conventional frequency-domain ICA, permutation of the output has to be re-aligned. Second, also similar to the conventional ICA, amplitude of the output has to be formatted. Third, the phase cannot be estimated from $\boldsymbol{G}_n(\omega)$. Fourth, estimation of the transfer system, the so-called steering vector, is required to conduct oversubtraction in [6]. To solve these problems, estimation of the mixing system is indispensable.

By calculating the inverse matrix of the demixing matrix, difference of phase and amplitude among channels can be estimated. Similarly, inversion of the linear separation operator $\boldsymbol{\Gamma}(\omega)$ gives a similar estimation. The weighting factor to construct the inverse linear operator is obtained by inversion of the Gramian matrix $\boldsymbol{K}(\omega)$. However, here for the efficiency of calculation, we use a more compact Gramian matrix $\boldsymbol{K}_{\boldsymbol{G}}(\omega)$ of $\boldsymbol{G}_n(\omega)$ given by

$$\boldsymbol{K}_{\boldsymbol{G}}(\omega) = [K\left(\boldsymbol{G}_i(\omega), \boldsymbol{G}_j(\omega)\right)]_{ij}. \tag{36}$$

Here, because $\boldsymbol{G}_n(\omega)$ is a matrix, a calculation of the kernel function $K\left(\boldsymbol{G}_i(\omega), \boldsymbol{G}_j(\omega)\right)$ is slightly different from $K(\boldsymbol{x}_1, \boldsymbol{x}_2)$ in Eq. (12), and is given by

$$K(\boldsymbol{G}_i(\omega), \boldsymbol{G}_j(\omega)) = \sum_k d_k^{(i)}(\omega)\boldsymbol{v}_k^{(i)}(\omega)^{\mathrm{H}}\boldsymbol{G}_j(\omega)\boldsymbol{v}_k^{(i)}(\omega), \tag{37}$$

where $\boldsymbol{v}_k^{(i)}(\omega)$ and $d_k^{(i)}(\omega)$ are the k-th eigenvector and the k-th eigenvalue of $\boldsymbol{G}_i(\omega)$, respectively. Then, the transfer function $\boldsymbol{M}_n(\omega)$ in $\mathcal{H}$, which is emphasized by $\boldsymbol{G}_n(\omega)$, is given by

$$\boldsymbol{M}_n(\omega) = \sum_i \beta_{in}(\omega)\boldsymbol{G}_i(\omega), \tag{38}$$

$$\boldsymbol{B}(\omega) = [\beta_{in}(\omega)]_{in} = \boldsymbol{K}_{\boldsymbol{G}}(\omega)^{+}. \tag{39}$$

Then the vector $\boldsymbol{v}_n(\omega)$ which is emphasized the most by $\boldsymbol{W}_n(\omega)$ is given by the most significant eigenvector $\hat{\boldsymbol{h}}_n$ of the transfer function $\boldsymbol{M}_n(\omega)$ in $\mathcal{H}$ as

$$\hat{\boldsymbol{h}}_n(\omega) = \underset{\boldsymbol{v} \,|\, \|\boldsymbol{v}\|^2=1}{\arg\max}\, \boldsymbol{v}^{\mathrm{H}}\boldsymbol{M}_n(\omega)\boldsymbol{v}. \tag{40}$$

By formatting phase with a suitable criterion, $\hat{\boldsymbol{h}}_n(\omega)$ is used as an estimate of the mixing system for the above objectives.

4. EXPERIMENTS

Tables 1 and 2 show the evaluation results of the separation performance and quality of the extracted speech. NRR stands for noise reduction ratio [5], which denotes the improvement of SNR with the processing. PESQ is an evaluation score of speech coding quality and predicts subjective evaluation result [14]. The evaluation is conducted with measured impulse responses in a room with T60 of approximately 400 ms, and the distance from the sources to the microphones is 1.5 m. The two-element microphone array with interelement spacing of 2 cm is used. Sources are distributed in the directions of $[-40^\circ, -10^\circ, 30^\circ]$ and $[-50^\circ, 10^\circ, 60^\circ]$ from the front. To exclude effects of permutation solution algorithm from the score of the proposed method, we assumed the best permutation alignment. MENUET is a reasonable realization of TFBM proposed in [15]. We evaluated three oversubtraction parameters $\beta = 3, 5, 7$ [5, 6].

Table 1 shows the separation performance of the mixture of three sources chosen from two female and two male speech with the same power (thus three sources are all speech). Since the sparseness assumption is satisfied, the performance of the TFBM method is very good. Table 2 shows the separation performance of speech from the mixture of three sources, which are chosen from female and male speech, music and stationary noise, with the same power. Since signal sparseness is not as strong as the previous one, the TFBM method performs worse, and especially the PESQ score degrades substantially. On the other hand, the proposed method is not affected by the sparseness assumption and performs more robustly than TFBM.

Table 1. Separation performance with three speech mixture

Method	NRR [dB]	PESQ
MENUET	8.87	2.04
Proposed ($\beta = 3$)	8.78	2.04
Proposed ($\beta = 5$)	9.56	2.04
Proposed ($\beta = 7$)	9.52	2.02

Table 2. Speech extraction performance from music and noise

Method	NRR [dB]	PESQ
MENUET	7.36	1.30
Proposed ($\beta = 3$)	7.14	1.92
Proposed ($\beta = 5$)	8.22	1.93
Proposed ($\beta = 7$)	8.87	1.92

5. CONCLUSIONS

We proposed a new nonlinear ICA algorithm based on the kernel method. Also, we used a kernel to works as nonlinear beamforming integrated with spectral subtraction, which can separate twice as many sources as does the linear ICA. The proposed method separated three sources with two microphone with higher performance than the conventional method when sparseness assumption is not satisfied well.

Our future work is to find other effective kernel mapping to higher dimensional RKHS than quadratic kernel does while retaining well the relationship among the signal bases well.

6. REFERENCES

[1] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis,* John and Wiley, New York, 2001.

[2] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Process.,* vol.52, no.7, pp. 1830–1847, 2004.

[3] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[4] J.F. Cardoso, "The three easy routes to independent component analysis: Contrasts and geometry," *Proc. ICA,* pp. 1–6, 2001.

[5] 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.

[6] S. Miyabe, B.H. Juang, H. Saruwatari, and K. Shikano, "Analytical solution of nonlinear microphone array based on complementary beamforming," *Proc. IWAENC,* #9084, 2008.

[7] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning,* vol.20, 273–297, 1995.

[8] S. Saitoh, *Integral Transforms, Reproducing Kernels and Their Applications,* Addison Wesley, 1997.

[9] L.Q. Zhang, A. Cichocki, and S. Amari, "Natural gradient algorithm for blind separation of overdetermined mixture with additive noise," *IEEE Signal Process. Lett.,* Vol.6, no.11, pp. 293–295, 1999.

[10] F.R. Bach and M.I. Jordan, "Kernel independent component analysis," *J. Machine Learning Research,* vol.3, pp. 1–48, 2003.

[11] J Yang, X. Gao, D. Zhang, and J.-Y Yang, "Kernel ICA: An alternative formulation and its application to face recognition," *Pattern Recognition*, vol.38, no.10, pp. 1784–1787, 2005.

[12] H. Sawada, S. Araki, and S. Makino, "MLSP 2007 Data Analysis Competition: Frequency-Domain Blind Source Separation for Convolutive Mixtures of Speech/Audio Signals," *Proc. MLSP,* pp. 45–50, 2007.

[13] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. Speech & Audio Process,*, vol. 12, no. 5, pp. 530–538, 2004.

[14] ITU-T Recommendation P.862, Perceptual evaluation of speech quality (PESQ), an objective method for endto-end speech quality assessment of narrow-band telephone networks and speech codecs, 2001.

[15] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing,* doi:10.1016/j.sigpro.2007.02.003, 2007.