

POLAR COORDINATE BASED NONLINEAR FUNCTION FOR FREQUENCY-DOMAIN BLIND SOURCE SEPARATION

Hiroshi Sawada Ryo Mukai Shoko Araki Shoji Makino

NTT Communication Science Laboratories
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
{sawada, ryo, shoko, maki}@cslab.kecl.ntt.co.jp

ABSTRACT

This paper presents a new type of nonlinear function for independent component analysis to process complex-valued signals, which is used in frequency-domain blind source separation. The new function is based on the polar coordinates of a complex number, whereas the conventional one is based on the Cartesian coordinates. The new function is derived from the probability density function of frequency-domain signals that are assumed to be independent of the phase. We show that the difference between the two types of functions is in the assumed densities of independent components. Experimental results for separating speech signals show that the new nonlinear function behaves better than the conventional one.

1. INTRODUCTION

Blind source separation (BSS) is a technique to estimate original source signals using only sensor observations that are mixtures of the original signals. Independent component analysis (ICA) [1–3] works well for BSS, if the mixture is instantaneous (non-convolutive). In a real room environment, however, sounds are mixed in a convolutive manner with reverberations, and long reverberations make the problem difficult. One of the major methods to cope with reverberations is frequency-domain BSS [4–7]. In this approach, a convolutive mixture in the time domain is converted into multiple instantaneous mixtures in the frequency domain, and ICA is applied to the instantaneous mixture in every frequency bin.

In frequency-domain BSS based on ICA, we have to deal with complex-valued signals. An extension of an ICA algorithm to process complex numbers was proposed [4], where the nonlinear function was based on the Cartesian coordinates of a complex number: nonlinearities are applied to the real and imaginary parts separately. This nonlinear function actually works and widely used by other researchers [5–7]. However, there has been presented no appropriate interpretation of this function. Moreover, it imposes an additional constraint that prevents a learning algorithm from converging unless a non-holonomic algorithm [8] is employed.

In this paper, we propose a new type of nonlinear function for an ICA algorithm to process complex numbers.¹ It is derived from the probability density function of frequency-domain signals that are assumed to be independent of the phase. As a result, the new function turns out to be based on the polar coordinates of a complex number. We also give an interpretation of the Cartesian coordinate based function. With experimental results for separating speech signals in a reverberant environment, we compare the behaviors of these two types of nonlinear functions, and discuss the differences between them.

2. FREQUENCY-DOMAIN BSS BASED ON ICA

Suppose that there are N source signals $s_i(t)$ that are mutually independent, and these signals are observed at M microphones $x_k(t) = \sum_{i=1}^N h_{ki}(t) * s_i(t)$, where $h_{ki}(t)$ represents the impulse response from source i to microphone k , and $*$ denotes the convolution operator. The goal of BSS is to separate observed signals $x_k(t)$ into N unmixed signals $y_i(t)$ that are mutually independent. The separation has to be done without knowing impulse responses $h_{ki}(t)$ nor original source signals $s_i(t)$.

To cope with convolutive mixtures, time-domain signals $x_k(t)$ are converted into frequency-domain time-series signals $X_k(\omega, m)$ by a T -point windowed DFT (discrete Fourier transform):

$$X_k(\omega, m) = \sum_{\tau=0}^{T-1} x_k(\tau + mS) w(\tau) e^{-j\omega\tau} \quad (1)$$

where $w(\tau)$ denotes a window function, S is a shifting interval of the window, and $\omega = 0, \frac{1}{T}2\pi, \dots, \frac{T-1}{T}2\pi$. Now, we have $\mathbf{X}(\omega, m) = [X_1(\omega, m), \dots, X_M(\omega, m)]^T$ for each frequency ω . Then, an unmixing $N \times M$ matrix $\mathbf{W}(\omega)$ and unmixed signals $\mathbf{Y}(\omega, m) = [Y_1(\omega, m), \dots, Y_N(\omega, m)]^T$ are obtained by solving an ICA problem:

$$\mathbf{Y}(\omega, m) = \mathbf{W}(\omega)\mathbf{X}(\omega, m)$$

in each frequency bin.

¹Our preliminary work on this topic is presented in [9].

Before explaining a complex-valued ICA, let us review an ordinary real-valued ICA algorithm. Based on the information maximization approach [1, 2] combined with the natural gradient [3], an unmixing matrix \mathbf{W} is gradually improved by the learning rule:

$$\Delta \mathbf{W} = \mu [\mathbf{I} - \langle \varphi(\mathbf{Y}) \mathbf{Y}^T \rangle] \mathbf{W}.$$

In this formula, μ is a step size parameter that has an effect on the speed of convergence, $\langle \cdot \rangle$ denotes the averaging operator, and $\varphi(\cdot)$ is a nonlinear function defined as:

$$\begin{aligned} \varphi(\mathbf{Y}) &= [\varphi(Y_1), \dots, \varphi(Y_N)]^T \\ \varphi(Y_i) &= -\frac{\partial}{\partial Y_i} \log p(Y_i) \end{aligned} \quad (2)$$

where $p(Y_i)$ is the probability density function (pdf) of Y_i . If we assume $p(Y_i) = \alpha / \cosh^2(Y_i)$, then the function is hyperbolic tangent $\varphi(Y_i) = 2 \tanh(Y_i)$, which is widely used for super-gaussian distributions [1, 2].

In frequency-domain BSS, signals obtained by DFT are complex. To deal with complex signals in ICA at each frequency, the calculation of $\Delta \mathbf{W}$ and the nonlinear function were extended [4]:

$$\begin{aligned} \Delta \mathbf{W} &= \mu [\mathbf{I} - \langle \Phi(\mathbf{Y}) \mathbf{Y}^H \rangle] \mathbf{W} \\ \Phi(Y_i) &= \tanh[\operatorname{re}(Y_i)] + j \cdot \tanh[\operatorname{im}(Y_i)] \end{aligned} \quad (3)$$

where \mathbf{Y}^H represents the conjugate transpose of \mathbf{Y} , and $\operatorname{re}(Y_i)$ and $\operatorname{im}(Y_i)$ are the real and imaginary parts of Y_i , respectively. In nonlinear function $\Phi(Y_i)$, $\tanh(\cdot)$ is applied separately in the real and imaginary parts. We call this type of function a Cartesian coordinate based function.

Although function (3) actually works, no appropriate interpretation of this function has been presented yet. Moreover, it has a convergence problem. Looking into the diagonal elements of $[\mathbf{I} - \langle \Phi(\mathbf{Y}) \mathbf{Y}^H \rangle]$, we see that $\Delta \mathbf{W}$ converges to a point that satisfies

$$\langle \Phi(Y_i) Y_i^* \rangle = 1 \quad (4)$$

where Y_i^* is the complex conjugate of Y_i . This makes the average amplitude of Y_i converge to some value. Extracting the imaginary part of this equation, we have

$$\langle \tanh[\operatorname{im}(Y_i)] \operatorname{re}(Y_i) - \tanh[\operatorname{re}(Y_i)] \operatorname{im}(Y_i) \rangle = 0. \quad (5)$$

Equation (5) imposes the additional constraint that $\operatorname{re}(Y_i)$ and $\operatorname{im}(Y_i)$ should be mutually independent. This constraint is too strong, and there are cases where $\Delta \mathbf{W}$ does not converge well because of this. We show such a case in Sec. 5.

If we use non-holonomic algorithm [8]:

$$\Delta \mathbf{W} = \mu [\operatorname{diag}(\langle \Phi(\mathbf{Y}) \mathbf{Y}^H \rangle) - \langle \Phi(\mathbf{Y}) \mathbf{Y}^H \rangle] \mathbf{W},$$

we can avoid constraint (5). Researchers [5, 6] used this algorithm combined with a Cartesian coordinate based function. However, there is still another convergence problem, which is also shown in Sec. 5.

3. NEW NONLINEAR FUNCTION

In this section, we propose a new type of nonlinear function derived from the complex counterpart of equation (2):

$$\Phi(Y_i) = -\frac{\partial}{\partial Y_i} \log p(Y_i). \quad (6)$$

At first, we make an assumption on the density $p(Y_i)$ of a complex-valued signal Y_i in the frequency domain.

Assumption 1 Let $Y_i = |Y_i| e^{j\theta(Y_i)}$ be a complex-valued signal. The pdf $p(Y_i)$ of Y_i is independent of the phase: $p(Y_i) = \alpha \cdot p(|Y_i|)$, where $p(|Y_i|)$ is the pdf of $|Y_i|$ and α is a constant.

This assumption is natural for a frequency-domain signal, since the phase of Y_i depends on the position of windows $w(\tau)$ of a windowed DFT (1) and the windows can be shifted arbitrarily.

Then, let us consider the derivative of a real-valued function $\log p(Y_i)$. Generally speaking, a real-valued function whose argument is a complex is not analytic: the derivative is not well-defined. Throughout this paper, we use the following definition of the derivative of a real-valued function.

Definition 1 Let $Y = Y_R + jY_I$ be a complex and $f(Y)$ be a real-valued function: $\mathcal{C} \rightarrow \mathcal{R}$. We define the derivative as:

$$\frac{\partial f(Y)}{\partial Y} \stackrel{\text{def}}{=} \frac{\partial f(Y)}{\partial Y_R} + j \frac{\partial f(Y)}{\partial Y_I}.$$

The relevance of this definition is in the fact that the result points to a direction in which $f(Y)$ increases. Using this definition, we derive

$$\frac{\partial |Y|}{\partial Y} = \left(\frac{\partial}{\partial Y_R} + j \frac{\partial}{\partial Y_I} \right) \sqrt{Y_R^2 + Y_I^2} = e^{j\theta(Y)} \quad (7)$$

which is used in the following theorem.

Theorem 1 Taking Assumption 1, $\Phi(Y_i)$ in (6) is

$$\Phi(Y_i) = \varphi(|Y_i|) e^{j\theta(Y_i)}$$

$$\text{where } \varphi(|Y_i|) = -\frac{\partial}{\partial |Y_i|} \log p(|Y_i|)$$

Proof: By transforming equation (6) into

$$\begin{aligned} \Phi(Y_i) &= -\frac{\partial}{\partial Y_i} \log \alpha \cdot p(|Y_i|) = -\frac{1}{p(|Y_i|)} \frac{\partial p(|Y_i|)}{\partial Y_i} \\ &= -\frac{1}{p(|Y_i|)} \frac{\partial p(|Y_i|)}{\partial |Y_i|} \frac{\partial |Y_i|}{\partial Y_i} \end{aligned}$$

and taking equation (7) into account, we prove the theorem.

Here, we have a nonlinear function based on the polar coordinates of a complex number. By using this type of function, constraint (5) does not appear. Since Y_i^* is a complex conjugate of Y_i ,

$$\Phi(Y_i) Y_i^* = \varphi(|Y_i|) e^{j\theta(Y_i)} |Y_i| e^{-j\theta(Y_i)} = \varphi(|Y_i|) |Y_i|.$$

Hence, the imaginary part of (4) becomes 0.

If we assume a super-gaussian distribution $p(|Y_i|) = \alpha / \cosh(|Y_i|)$, $\Phi(Y_i) = \tanh(|Y_i|) e^{j\theta(Y_i)}$.

Table 1. Experimental conditions

direction of sources	-30° and 40° (two sources)
distance of two microphones	4 cm
length of source signal	3 seconds
reverberation time	$T_R = 300$ ms
sampling rate	8kHz
window function	$w(\tau)$: Hanning
window length	$T = 2048$ points (256 ms)
shifting interval	$S = 512$ points (64 ms)
step size	$\mu = 0.2$
gain parameter	$\eta = 100$
number of iterations	100

4. INTERPRETATION OF THE CARTESIAN COORDINATE BASED FUNCTION

In this section, we give an interpretation of the conventional Cartesian coordinate based function (3). We assume the definitions in Sec. 3 in the following theorem.

Theorem 2 If $p(Y) = p(Y_R) \cdot p(Y_I)$, then

$$\Phi(Y) = \varphi(Y_R) + j \cdot \varphi(Y_I)$$

where $\varphi(Y_R) = -\frac{\partial}{\partial Y_R} \log p(Y_R)$, $\varphi(Y_I) = -\frac{\partial}{\partial Y_I} \log p(Y_I)$.

Proof:
$$\begin{aligned} \Phi(Y) &= -\frac{\partial}{\partial Y_R} \log p(Y) - j \cdot \frac{\partial}{\partial Y_I} \log p(Y) \\ &= -\frac{\partial}{\partial Y_R} \log p(Y_R) - j \cdot \frac{\partial}{\partial Y_I} \log p(Y_I). \end{aligned}$$

This shows that a Cartesian coordinate based function assumes that the pdf $p(Y)$ can be factorized into the product of $p(Y_R)$ and $p(Y_I)$, and therefore Y_R and Y_I are mutually independent. We can notice that the additional constraint (5) is satisfied if this assumption is met. By this theorem, the assumed density in the nonlinearity (3) turns out to be $p(Y) = \alpha_R / \cosh(Y_R) \cdot \alpha_I / \cosh(Y_I)$.

5. EXPERIMENTS AND DISCUSSIONS

To compare the two types of nonlinear functions, we conducted experiments to separate speech signals. Actually, we used the following two nonlinear functions:

$$\begin{aligned} \text{Polar} \quad & \Phi(Y) = \tanh(\eta|Y|) e^{j \cdot \theta(Y)} \\ \text{Cartesian} \quad & \Phi(Y) = \tanh(\eta Y_R) + j \cdot \tanh(\eta Y_I) \end{aligned}$$

and the following two gradients of \mathbf{W}

$$\begin{aligned} | \quad & \Delta \mathbf{W} = \mu [\mathbf{I} - \langle \Phi(\mathbf{Y}) \mathbf{Y}^H \rangle] \mathbf{W} \\ \text{Diag} \quad & \Delta \mathbf{W} = \mu [\text{diag}(\langle \Phi(\mathbf{Y}) \mathbf{Y}^H \rangle) - \langle \Phi(\mathbf{Y}) \mathbf{Y}^H \rangle] \mathbf{W}. \end{aligned}$$

The other conditions are summarized in Table 1.

The experiments were performed for 48 combinations (2 nonlinear functions, 2 gradients, and 12 pairs of speech signals). Figure 1 shows the overall results. The measurement is the average of two output SNRs (signal-to-noise ratios) in

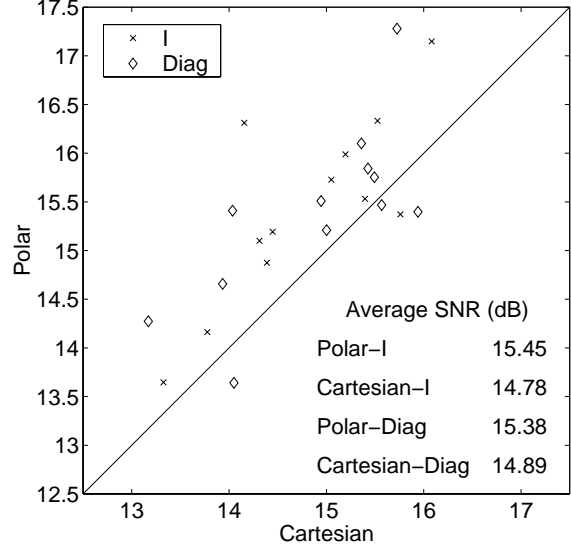


Fig. 1. SNRs compared between Polar and Cartesian

dB. Each point represents the results of Polar and Cartesian with the same gradient and with the same combination of speech signals.

In order to compare the results as correctly as possible, we avoided the influence of the permutation problem [5–7] of frequency-domain BSS. We selected the best permutation by actually calculating the SNR in each frequency bin. Therefore, the results in Fig. 1 are ideal ones under the condition that the permutation problem is perfectly solved. Before applying ICA, we whitened observed signals \mathbf{X} to be uncorrelated and to have unit variances. This pre-process was very important to make the ICA algorithm stable especially for the Diag cases, where equation (4) is not concerned. Without this process, the Diag cases could have exhibited irregular convergence speeds among the different frequency bins.

We can see that the result of Polar is better than that of Cartesian in most cases. At first, we discuss the additional constraint (5) imposed in the Cartesian-I case. Figure 2 shows the values of $[\mathbf{I} - \langle \Phi(\mathbf{Y}) \mathbf{Y}^H \rangle]$ at some frequency bin. The horizontal axis corresponds to the number of iterations. The first graph shows the absolute values of each element for the Cartesian-I case. We see oscillations that hinder convergence. They come from the imaginary parts of the diagonals as shown in the second graph. If we use a polar coordinate based function, we can eliminate such oscillations as discussed in Sec. 3. The third graph shows the Polar-I case. We can see a smooth convergence. Clearly, the mutual information among \mathbf{Y} is well minimized in this case unlike in the Cartesian-I case.

If we use Diag as the calculation of $\Delta \mathbf{W}$, we can eliminate the additional constraint (5) even in the Cartesian case. Accordingly, by investigating the results of Diag, we can see the differences that purely come from the difference

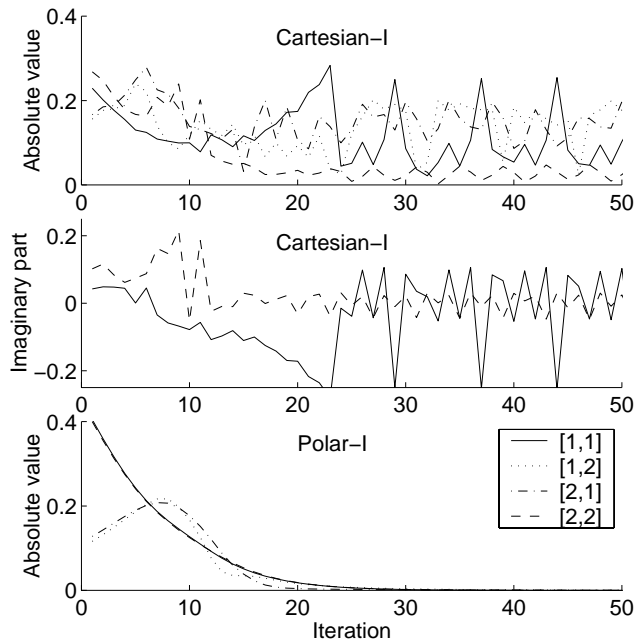


Fig. 2. Values of $[I - \langle \Phi(Y) Y^H \rangle]$

of the nonlinearities between Polar and Cartesian. In fact, we found another convergence problem in a Cartesian–Diag case. Figure 3 shows the trajectory of element W_{11} of \mathbf{W} at some frequency bin. We see that the direction of the movement changes gradually in Polar–Diag, whereas it changes sharply and frequently in Cartesian–Diag. The difference comes from the assumed densities, as discussed in Sec. 3 and Sec. 4. Figure 4 shows the contour and gradient of $-\log p(Y)$, being $p(Y) = \alpha / \cosh(\eta |Y|)$ in Polar, and $p(Y) = \alpha / [\cosh(\eta Y_R) \cdot \cosh(\eta Y_I)]$ in Cartesian. The gradient corresponds to $\Phi(Y)$. We see that the direction of the gradient smoothly changes around the neighborhood in the Polar case, whereas it changes steeply near the vertical and horizontal axes in the Cartesian case. By increasing the number of samples, this steepness may be smoothed out by the averaging operator $\langle \Phi(Y) Y^* \rangle$. However, we conjecture that the jag in Fig. 3 comes from this steepness.

6. CONCLUSIONS

We proposed a polar coordinate based nonlinear function to process complex-valued signals in ICA. Compared to the Cartesian coordinate based function, the main difference is in the assumed densities of independent signals. In frequency-domain BSS, the assumption that the density is phase independent is more natural than the assumption that the real and imaginary parts are mutually independent, since signals are produced by a windowed DFT. This consideration is supported by the experimental results: the Polar results were better than the Cartesian ones in most cases, and there were some convergence problems in the Cartesian case.

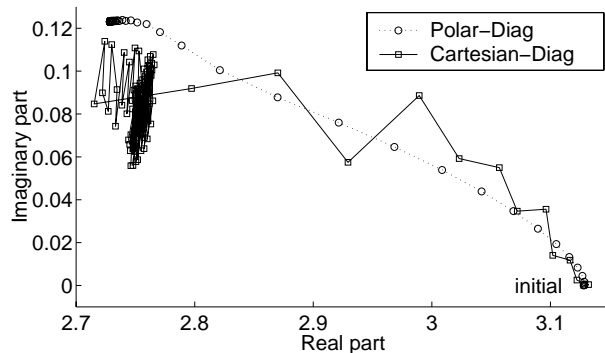


Fig. 3. Trajectory of W_{11} on the complex plane

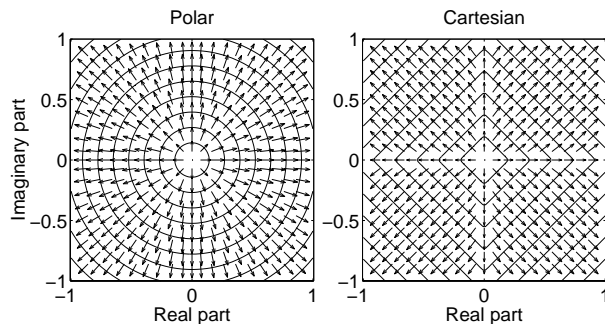


Fig. 4. Contour and gradient of $-\log p(Y)$

REFERENCES

- [1] A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [2] T. W. Lee, *Independent component analysis - Theory and applications*, Kluwer academic publishers, 1998.
- [3] S. Amari, “Natural gradient works efficiently in learning,” *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.
- [4] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, vol. 22, pp. 21–34, 1998.
- [5] S. Ikeda and N. Murata, “A method of ICA in time–frequency domain,” in *Proc. ICA '99*, Jan., pp. 365–370.
- [6] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, “Evaluation of blind signal separation method using directivity pattern under reverberant conditions,” in *Proc. ICASSP2000*, pp. 3140–3143.
- [7] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, “A combined approach of array processing and independent component analysis for blind separation of acoustic signals,” in *Proc. ICASSP2001*, pp. 2729–2732.
- [8] S. Amari, T. P. Chen, and A. Cichocki, “Nonholonomic orthogonal learning algorithm for blind source separation,” *Neural Computation*, vol. 12, no. 6, pp. 1463–1484, 2000.
- [9] H. Sawada, R. Mukai, S. Araki, and S. Makino, “A polar-coordinate based activation function for frequency domain blind source separation,” in *Proc. ICA 2001*, (to appear).