

複素数に対する独立成分分析のための極座標表示に基づく活性化関数

A Polar-Coordinate based Activation Function for Complex-valued Independent Component Analysis

澤田 宏 (PY), 向井 良, 荒木 章子, 牧野 昭二 H. Sawada, R. Mukai, S. Araki and S. Makino
NTT コミュニケーション科学基礎研究所 NTT Communication Science Laboratories
{sawada,ryo,shoko,maki}@cslab.kecl.ntt.co.jp

Abstract

This paper presents a new activation function for an ICA algorithm to process complex-valued signals. It is based on the polar coordinates of a complex number, whereas the conventional one is based on the Cartesian coordinates of a complex number. The new activation function eliminates an undesirable constraint occurred by the conventional one.

1 はじめに

近年、独立成分分析 (ICA: Independent Component Analysis) により、線形混合された複素の信号を分離する技術が進展してきている [1, 2, 3]。しかしながら、実際に室内で混合された音の分離に関しては、残響の影響により十分な分離性能が得られていないのが現状である。残響に対処する一つの手法は、周波数領域での解法である [4, 5, 6, 7]。その大きな利点は、時間領域における畳み込み (残響のモデル化) が、周波数領域では単なる係数の掛け算として扱えるため、瞬時混合の ICA アルゴリズムが適用できるということにある。

周波数領域での解法では、時間領域での信号を短時間フーリエ変換を用いて各周波数ビン毎の系列に変換する。そのため、ICA を適用する際には複素数を扱う必要がある。一般に ICA アルゴリズムでは非線形の活性化関数を用いて学習を進めていくが、これも複素数に拡張する必要がある。これまでには、複素数を実部と虚部に分解して別々に非線形関数を適用する方法が提案されている。しかし、この方法では余分な制約が発生して収束を阻むことがある。本稿では、複素数の極座標表示に基づく活性化関数を提案し、上記の制約が発生しないこと、およびその他の有効性について議論する。

2 複素数を扱う独立成分分析

まず、瞬時混合に対する独立成分分析 (ICA) の定式化を行い、そのアルゴリズムを説明する。互いに独立な N 個の信号 $s(t) = [s_1(t), \dots, s_N(t)]^T$ が行列 A により線形混合 $x(t) = As(t)$ されて、 M 個のセンサ $x(t) = [x_1(t), \dots, x_M(t)]^T$ で観測されたとする。ICA の目的は、混合系 A や源信号 $s(t)$ を知らずに、 $x(t)$ を互いに独立な N 個の信号 $y(t) = [y_1(t), \dots, y_N(t)]^T = Wx(t)$ に分離する行列 W を求めることである。

分離行列 W は、 y の各要素間の相互情報量の最小化を目指して、学習則 $W_{i+1} = W_i + \Delta W_i$ により徐々に改良される [1, 2]。 ΔW の計算には自然勾配法 (natural gradient) [3]、 $\Delta W = \mu [I - \langle \varphi(y)y^T \rangle] W$ が広く用いられている。ここで、 μ はステップサイズ、 $\langle \cdot \rangle$ は時間平均を表す。 $\varphi(\cdot)$ は活性化関数であり、 $\varphi(y) = \tanh(\eta \cdot y)$

が非線形の活性化関数として広く用いられている [1, 2]。 η は φ の非線形性の強さを制御するパラメータである。

残響下での音源分離に対する周波数領域での解法では、ICA が各周波数毎に、しかも複素数に対して行われる。そのため、 ΔW と活性化関数 $\varphi(\cdot)$ の複素数への拡張 $\Phi(\cdot)$ が提案された [4]。

$$\Delta W = \mu [I - \langle \Phi(y)y^H \rangle] W \quad (1)$$

$$\Phi(y) = \varphi[re(y)] + j \cdot \varphi[im(y)] \quad (2)$$

I は単位行列、 y^H は y の共役転置、 $re(y)$ と $im(y)$ はそれぞれ y の実部と虚部である。式 (1) に従うと W は

$$\langle \Phi(y_p)y_q^* \rangle = 0 \quad (p \neq q) \quad (3)$$

$$\langle \Phi(y_p)y_p^* \rangle = 1 \quad (p = q) \quad (4)$$

を満たす点に収束する。ここで、 y_q^* は y_q の複素共役である。式 (3) は y_p と y_q が互いに独立になるように働く。式 (4) は y_p の振幅の平均値をある値に近づける。式 (4) を実部と虚部に分解すると、

$$\langle \varphi[re(y_p)]re(y_p) + \varphi[im(y_p)]im(y_p) \rangle = 1 \quad (5)$$

$$\langle \varphi[im(y_p)]re(y_p) - \varphi[re(y_p)]im(y_p) \rangle = 0 \quad (6)$$

となる。ここで式 (6) が余分な制約を課していることがわかる [6]。例えば $re(y_p)$ と $im(y_p)$ が互いに独立であればこの制約を満たすが、このような制約は望ましくない。そこで、 ΔW を計算する別の式が提案された [5]。

$$\Delta W = \mu [\text{diag}(\langle \Phi(y)y^H \rangle) - \langle \Phi(y)y^H \rangle] W \quad (7)$$

この式によると、 W は式 (3) のみを満たす点に収束し、 y の振幅の平均はそれほど変化しない。

3 極座標表示に基づく活性化関数

余分な制約 (6) の問題を別の方法で解決するため、複素数の極座標表示に基づく新しい活性化関数 $\Phi(\cdot)$ を提案する。

$$\Phi(y) = \varphi[abs(y)] \cdot e^{j \cdot angle(y)}, \quad (8)$$

ここで、 $abs(y)$ と $angle(y)$ はそれぞれ y の各要素の絶対値と偏角を表す。この活性化関数は、元の y の絶対値のみを変更し、その偏角は変化させない。これは、実数に対する活性化関数 $\varphi(\cdot)$ の自然な拡張であり、実数に対しては双方とも同じ値を出力する。また、この関数は [8] に示されている活性化関数としての要件を満たしている。

この活性化関数を使うことで、式 (6) のような余分な制約は発生しない。 $\theta = angle(y_p)$ とすると、 y_p^* が y_p の複素共役であることから、

$$\begin{aligned} \Phi(y_p)y_p^* &= \varphi[abs(y_p)] \cdot e^{j\theta} \cdot abs(y_p) \cdot e^{-j\theta} \\ &= \varphi[abs(y_p)] \cdot abs(y_p) \end{aligned}$$

従って、式 (4) の $\langle \Phi(y_p)y_p^* \rangle$ における虚部は 0 となる。

表 1: 異なる $\Phi(\cdot)$ と ΔW に対する SNR (dB)

	$T_R = 150$ ms		$T_R = 300$ ms	
	Ref	Imp	Ref	Imp
Polar-I	18.3	19.7	12.7	16.3
Cartesian-I	17.9	19.4	12.3	15.6
Cartesian-diag	17.8	18.0	11.9	14.6

$\mu = 0.1, \eta = 100, W$ の更新回数 = 100
 Ref: 音声信号自身で計測した SNR
 Imp: インパルスで計測した SNR [7]

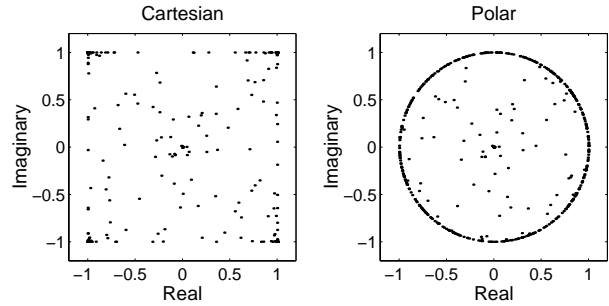


図 1: $\Phi(y)$ の出力値

4 実験結果および考察

提案した活性化関数の有効性を示すため、以下に示す 3 種類の組合せに関して、残響下で混合された音声进行分離する実験を行った。実際には、ASJ 研究用音声コーパス中から選んだ 8 秒の音声データと部屋のインパルス応答を計算機上で畳み込んで混合した。音源数、マイク数は共に 2 である。サンプリング周波数は 8kHz である。2 種類の残響時間 $T_R = 150$ ms (1200 サンプル), 300ms (2400 サンプル) に対して行い、短時間フーリエ変換のフレーム長は、それぞれ 1024, 2048 サンプルとした。

Polar-I $\Phi(\cdot)$ に式 (8), ΔW に式 (1) を用いる
Cartesian-I $\Phi(\cdot)$ に式 (2), ΔW に式 (1) を用いる
Cartesian-diag $\Phi(\cdot)$ に式 (2), ΔW に式 (7) を用いる

表 1 に、2 つの出力 y_1, y_2 の SNR (signal-to-noise ratio) の平均を示す。総じて“Polar-I”の結果が他を上回った。“Cartesian-diag”の場合には、周波数ビン毎の収束速度に大きなばらつきが見られた。これは、式 (4) が働かないことから、周波数ビン毎で y の大きさが揃わないためであると認識している。以下では、“Polar-I”が“Cartesian-I”より結果が良かった理由を考察する。

2 つの $\Phi(\cdot)$ がどのように振る舞うかを見るため、 $\Phi(y)$ の出力値を図 1 にプロットした。これらのデータは、 $T_R = 150$ ms の場合の第 200 番目 (1554.7 Hz) の周波数ビンにおける最終ステップのものである。“Cartesian”では多くの点が 4 つの角 $1+j, 1-j, -1+j, -1-j$ に集中しているのに対し、“Polar”では多くの点が単位円 $e^{j\theta}$ 上に分布している。これにより、式 (8) は、式 (2) よりも、 y に関する多くの情報 (特に位相情報) を表現していることがわかる。さらに、 $\Phi(y)$ のエントロピーも、式 (8) によるものの方が大きいことが分かる。

次に、“Cartesian-I”の場合に課せられている余分な制約 (6) に関して考察する。図 2 は、活性化関数 (2) および (8) を用いた場合の $[I - \langle \Phi(y)y^H \rangle]$ の各成分の絶対値を示している。これらのデータは、 $T_R = 150$ ms の場合の第 100 番目 (773.4 Hz) の周波数ビンのものである。“Cartesian”には収束を妨げている振動が見られるが、“Polar”では滑らかに収束しており、“Cartesian”の場合と比べて y の相互情報量は明らかに小さくなっている。なお、“Cartesian”の場合の振動は、2 章での議論の通り、 $[I - \langle \Phi(y)y^H \rangle]$ の対角成分の虚部で発生していることを確認した。

5 おわりに

複素数 ICA のための新たな活性化関数を提案した。これは極座標表示に基づくものであり、実数の場合の自然な拡張である。これを用いると、制約 (6) が影響しないため、式 (1) を使っても収束を妨げる振動は発生しない。従って、制約 (6) を避けるために、周波数毎で収束速度

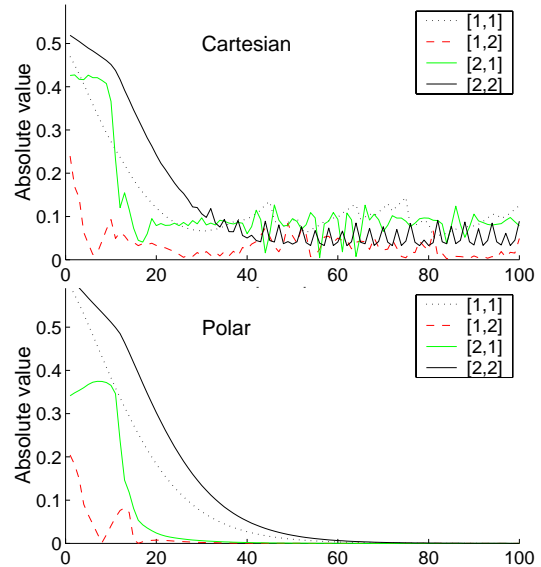


図 2: $[I - \langle \Phi(y)y^H \rangle]$ の各成分の絶対値

がばらつく式 (7) を使わなくても良くなる。残響下で混合された音声进行分離する実験により、新たな活性化関数の有効性を示した。

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, A. Cichocki, and H. H. Yang, “A new learning algorithm for blind signal separation,” in *Advances in Neural Information Processing Systems*. 1996, vol. 8, pp. 757–763, The MIT Press.
- [4] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, vol. 22, pp. 21–34, 1998.
- [5] N. Murata and S. Ikeda, “An on-line algorithm for blind source separation on speech signals,” in *Proc. Int. Symposium on Nonlinear Theory and Its Application (NOLTA '98)*, 1998, pp. 923–926.
- [6] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, “Blind signal separation using directivity pattern,” in *Technical Report of Japanese Society for Artificial Intelligence*, Nov. 1999, pp. 21–26.
- [7] R. Mukai, S. Araki, and S. Makino, “Separation and dereverberation performance of frequency domain blind source separation for speech in a reverberant environment,” in *Proc. Eurospeech2001*, Sept. 2001.
- [8] G. M. Georgiou and C. Koutsougeras, “Complex domain backpropagation,” *IEEE Trans. Circuits and Systems II*, vol. 39, no. 5, pp. 330–334, May 1992.