A Curtosis Based Criterion for Solving the Permutation Ambiguity in Convolutive Blind Source Separation

Radoslaw Mazur, Jan Ole Jungmann, and Alfred Mertins
Universität zu Lübeck, Institute for Signal Processing, Ratzeburger Allee 160, D-23562 Lübeck
Email: mazur@isip.uni-luebeck.de

Abstract

In this work, we present a modification of an algorithm for solving the permutation ambiguity in convolutive blind source separation. A well known approach for separation of convolutive mixtures is the transformation to the time-frequency domain, where the convolution becomes a multiplication. With this approach it is possible to use well known instantaneous ICA algorithms independently in each frequency bin. This simplification leads to reduced computational costs and better separation in each frequency bin. However, this simplification has the major drawback of arbitrary permutation in each frequency bin. Without a correction of this permutation the restored time domain signals still remain mixed. An often used approach for solving this permutation problem is the dyadic sorting, where groups of bins are consecutively depermuted. By recursively joining growing groups all bins gets sorted. In recent works we presented a criterion for the depermutation, which was based on sparsity in the time domain of the restored subband signals. In this work we modify this approach to use a curtosis based criterion which is an alternative measurement for the non-gaussianity of speech signals.

Introduction

Blind Source Separation (BSS) of linear and instantaneous mixtures can be performed using the Independent Component Analysis (ICA). For this case, numerous algorithms have been proposed [1, 2].

When dealing with real-world recordings of speech, this simple approach is not effective anymore. As the signals arrive multiple times with different delays, the mixing procedure becomes convolutive. These characteristics can be modeled using FIR filters. In this case, the separation is only possible when the unmixing system is again a set of FIR filters.

As the direct calculation of the unmixing fitlers in time domain is very demanding, time-frequency approaches are often used. Here, the convolution becomes a multiplication and each frequency bin can be separatated using an instantaneous method. However, this simplification has a major disadvantage. The separated signals usually have arbitrary scaling and are randomly permuted across the frequency bins. Without the correction of the scaling, only filtered versions of the signals are restored. This ambiguity is often solved using the minimal distortion principle [4]. This method accepts the filtering done by the mixing system without adding new distortions.

The random permutation of the single frequency bins has an even bigger impact. Without a correct alignment, different signals appear in the single outputs and the whole process fails.

Many different approaches for solving this problem have been proposed. Often, the time structure of the separated bins is used and the assumption of high correlation between neighboring bins is utilized. This has been used for example in [3] and [8]. Other approaches include a statistical modeling of the single bins using the generalized Gaussian distribution. Small differences of the parameters lead to a depermutation criterion in [5] and [6].

The method from [8] introduced the so called dyadic sorting, where in every stage growing sets of bins are depermuted using the correlation method. Using multiple bins for comparison resulted in a more robust criterion. This approach has been extended in [7], where time domain signals for the sets of bins have been used. This modification allows for a even more robust criterion, as only one coefficient has to be considered for the depermutation. Additionally, in [7] a sparsity based criterion has been introduced.

In this work we modify this approach. The calculation of the sparsity of the time domain signals can be interpreded as an measurement of non-gaussianity. Here, we propose to use a kurtosis based criterion which is an alternative measurement for the non-gaussianity.

BSS for instantaneous mixtures

The instantaneous mixing process of N sources into N observations is modeled by an $N \times N$ matrix $\mathbf{A}$. With the source vector $\boldsymbol{s}(n) = [s_1(n), \ldots, s_N(n)]^T$ and negligible measurement noise, the observation signals $\boldsymbol{x}(n) = [x_1(n), \ldots, x_N(n)]^T$ are given by

$$\boldsymbol{x}(n) = \mathbf{A} \cdot \boldsymbol{s}(n). \tag{1}$$

The separation is again a multiplication with a matrix $\mathbf{B}$:

$$\boldsymbol{y}(n) = \mathbf{B} \cdot \boldsymbol{x}(n) \tag{2}$$

with $\boldsymbol{y}(n) = [y_1(n), \ldots, y_N(n)]^T$. The single source of information for the estimation of $\mathbf{B}$ is the observed process $\boldsymbol{x}(n)$. The separation is successful when $\mathbf{B}$ can be estimated so that $\mathbf{BA} = \mathbf{D\Pi}$ with $\mathbf{\Pi}$ being a permutation matrix and $\mathbf{D}$ being an arbitrary diagonal matrix. These two matrices stand for the two ambiguities of BSS. The signals may appear in any order and can be arbitrarily scaled.

For the separation we use the well known gradient-based update rule [1]

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \Delta\mathbf{B}_k \tag{3}$$

with

$$\Delta\mathbf{B}_k = \mu_k(\boldsymbol{I} - E\left\{\boldsymbol{g}(\boldsymbol{y})\boldsymbol{y}^T\right\})\mathbf{B}_k. \tag{4}$$

The term $\boldsymbol{g}(\boldsymbol{y}) = (g_1(y_1), \ldots g_n(y_n))$ is a component-wise vector function of nonlinear score functions $g_i(s_i) = -p_i'(s_i)/p_i(s_i)$ where $p_i(s_i)$ are the assumed source probability densities.

Convolutive mixtures

When dealing with real-world acoustic scenarios it is necessary to consider reverberation. The mixing system can be modeled by FIR filters of length L:

$$\boldsymbol{x}(n) = \mathbf{H}(n) * \mathbf{s}(n) = \sum_{l=0}^{L-1} \mathbf{H}(l)\boldsymbol{s}(n-l) \quad (5)$$

where $\mathbf{H}(n)$ is a sequence of $N \times N$ matrices containing the impulse responses of the mixing channels. For the separation we use FIR filters of length M and obtain

$$\boldsymbol{y}(n) = \mathbf{W}(n) * \boldsymbol{x}(n) = \sum_{l=0}^{M-1} \mathbf{W}(l)\boldsymbol{x}(n-l) \quad (6)$$

with $\mathbf{W}(n)$ containing the unmixing coefficients.

Using the short-time Fourier transform (STFT), the signals can be transformed to the time-frequency domain, where the convolution approximately becomes a multiplication:

$$\boldsymbol{Y}(\omega_k, \tau) = \boldsymbol{W}(\omega_k)\boldsymbol{X}(\omega_k, \tau), \quad k = 0, 1, \ldots, K-1, \quad (7)$$

where K is the FFT length. The major benefit of this approach is the possibility to estimate the unmixing matrices for each frequency independently, however, at the price of possible permutation and scaling in each frequency bin:

$$\boldsymbol{Y}(\omega_k, \tau) = \boldsymbol{W}(\omega_k)\boldsymbol{X}(\omega_k, \tau) = \boldsymbol{D}(\omega_k)\boldsymbol{\Pi}(\omega_k)\boldsymbol{S}(\omega_k, \tau) \quad (8)$$

where $\boldsymbol{\Pi}(\omega)$ is a frequency-dependent permutation matrix and $\boldsymbol{D}(\omega)$ an arbitrary diagonal scaling matrix.

The scaling can be solved the minimal distortion principle [4]. A modified approach for the permutation problem will be shown in the next section.

Depermutation Algorithm

The method from [7] extends the ideas of [8]. The basis is the dyadic sorting, where at the first step, only pairs of bins are depermuted. In the second step, these pairs are aligned, and then the resulting quadruples are depermuted. This scheme is continued until all bins are processed. Within this procedure, single wrong permuted bins at the early stages do not outbalance the majority.

In the original work, the comparison was based on correlation of the single bins. In [7] the sparsity of time domain representation of the sets of bins has been used. With $z(\omega_{ab}, n)$ being the time-domain representation of the bins in the frequency range $[a, b]$ of $Y(\omega, \tau)$ and the sparsity measurement of the sum of two sets of bins

$$\varrho_{qp}(\omega_{ks}, \omega_{lt}) = \|z_q(\omega_{ks}, n) + z_p(\omega_{lt}, n)\|_{\ell_p} \quad (9)$$

using the ℓ_p pseudo norm

$$\|\boldsymbol{x}\|_{\ell_p} = \left(\sum_{i=0}^{N-1} |x(i)|^p\right)^{\frac{1}{p}} \quad (10)$$

with $0 \leq p \leq 1$, the depermutation is computed an the basis of the ratio

$$r_{kl,st} = \frac{\varrho_{pp}(\omega_{ks}, \omega_{lt}) + \varrho_{qq}(\omega_{ks}, \omega_{lt})}{\varrho_{pq}(\omega_k s, \omega_{lt}) + \varrho_{qp}(\omega_{ks}, \omega_{lt})}. \quad (11)$$

With $r_{kl,st} > 1$ ranges $[k, s]$ and $[l, t]$ of the separated signals q and p are correctly aligned among each other. Otherwise they are permuted.

Here, we propose to use kurtosis for the measurement of the non-gaussianity instead of the sparsity. With

$$\varrho_{qp}(\omega_{ks}, \omega_{lt}) = \text{kurt}\big(z_q(\omega_{ks}, n) + z_p(\omega_{lt}, n)\big) \quad (12)$$

and

$$\text{kurt}(x) = \frac{E\{|x|^4\}}{E\{|x|^2\}^2} - 3 \quad (13)$$

the permutation of the ranges $[k, s]$ and $[l, t]$ can be again determined using equation (11). With this modification the meaning of $r_{kl,st}$ is inverted. Now, with $r_{kl,st} > 1$ the sets are permuted and otherwise correctly aligned. The remaining algorithm, as proposed in [7], can be used without modification.

Preliminary tests show simmilar performance on real-world data sets.

Conclusions

In this work, we presented a modification of an algorithm for solving the permutation ambiguity in convolutive blind source separation. It is based on the kurtosis as a measurement of non-gaussianity of time domain representation of groups of bins.

References

[1] S.-I. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems*, volume 8, MIT Press, Cambridge, MA, 1996.

[2] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.

[3] S. Ikeda and N. Murata. A method of blind separation based on temporal structure of signals. In *Proc. Int. Conf. on Neural Information Processing*, pages 737–742, 1998.

[4] K. Matsuoka. Minimal distortion principle for blind source separation. In *Proceedings of the 41st SICE Annual Conference*, volume 4, pages 2138–2143, 5-7 Aug. 2002.

[5] R. Mazur and A. Mertins. An approach for solving the permutation problem of convolutive blind source separation based on statistical signal models. *IEEE Trans. Audio, Speech, and Language Processing*, 17(1):117–126, Jan. 2009.

[6] R. Mazur and A. Mertins. Simplified formulation of a depermutation criterion in convolutive blind source separation. In *Proc. European Signal Processing Conference*, pages 1467–1470, Glasgow, Scotland, Aug 2009.

[7] R. Mazur and A. Mertins. A sparsity based criterion for solving the permutation ambiguity in convolutive blind source separation. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, pages 1996–1999, Prague, Czech Republic, May 2011.

[8] K. Rahbar and J. P. Reilly. A frequency domain method for blind source separation of convolutive audio mixtures. *IEEE Trans. Speech and Audio Processing*, 13(5):832–844, Sept. 2005.