GENERALIZED ANTI-HEBBIAN LEARNING FOR SOURCE SEPARATION

Hsiao-Chun Wu and Jose C. Principe

Computational Neuro-Engineering Laboratory
Department of Electrical and Computer Engineering
University of Florida, Gainesville, FL 32611
E-mail: wu@cnel.ufl.edu, principe@cnel.ufl.edu

ABSTRACT

The information-theoretic framework for source separation is highly suitable. However the choice of the nonlinearity or the estimation of the multidimensional joint probability density function are nontrivial. We propose here a generalized Gaussian model to construct a generalized blind source separation network based on the minimum entropy principle. This new separation network can suppress the interference to a significant amount compared to the traditional LMS-echo-canceler. The simulation is given to show the disparity of the performance as α varies. Finally how to choose the appropriate α in our generalized anti-Hebbian rule is discussed.

1. INTRODUCTION

Recently, source separation or independent component analysis has been a new direction in signal processing research because of its relevance for medical instrumentation, interference removal in communication networks and speech enhancement in noisy environments. Previous research [1, 2] has addressed an unrealistic simplified model: the instantaneous mixture model. In this case the maximum likelihood estimation of the demixing matrix can be related to the Kullback-Leibler divergence between the source probability density function and the underlying model density function associated with the nonlinearity used in the processing elements (PEs).

However, in the more realistic case of convolutive mixtures, the characterization of the joint probability density function is not available. The simple relationship between the output and input through a Jacobian determinant does not exist any more [2]. In addition, the nonlinearity restricts the separation performance due to the mismatch problem [2]. Hence, in this paper, we utilize the minimum entropy principle applied to the marginal probability density function to compensate for the lack of the Jacobian in convolutive mixtures. Moreover, we provide a family of symmetric generalized Gaussian distributions to match the source statistics and derive the generalized anti-Hebbian rule for source separation for sources with different kurtoses.

2. PROBLEM STATEMENT

The problem of blind separation of independent sources can be depicted in Figure 1. A vector of n source signals $\boldsymbol{s}(t) \in \mathbb{R}^n$ are transmitted through an unknown linear channel with time delays. The mathematical expression is:

$$\boldsymbol{x}(t) = \sum_{k=1}^{L} \boldsymbol{H}(k)\boldsymbol{s}(t-k) \quad (1)$$

where $\boldsymbol{H}(k)$ parameterizes a series of mixing matrices composed by the transforms of the echo and interference paths, and t denotes time. The problem is to reconstruct $\boldsymbol{s}(t)$ in the form of

$$\boldsymbol{y}(t) = \sum_{k=1}^{M} \boldsymbol{W}(k)\boldsymbol{x}(t-k) = \boldsymbol{s}(t), \quad (2)$$

or its scaled version

$$\boldsymbol{s}'(t) = \begin{bmatrix} D_1(z) & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & D_n(z) \end{bmatrix} \boldsymbol{s}(t) . \quad (3)$$

from the given input $\boldsymbol{x}(t)$.

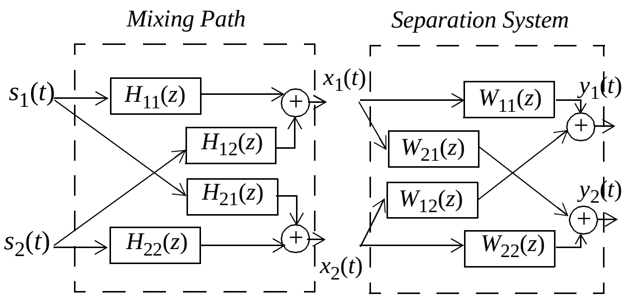


Figure 1. The BSS problem.

Generally speaking, the recovered sources at the outputs of Figure 1 may not always be the true original sources as described in Equation (3), since the statistical independence condition is still verified among $\boldsymbol{s}'(t)$. However, it is satisfactory to obtain any version of $\boldsymbol{s}'(t)$ under the independence assumption.

3. MINIMUM ENTROPY PRINCIPLE

The minimum entropy deconvolution (MED) was utilized in early seismography processing [3]. The MED design of an inverse filter was based on an estimator such that its outputs produced the best possible approximation to a delta function. The inverse filtered outputs are thus less Gaussian, have higher kurtosis and less entropy than the inputs. Hence the single channel MED blind deconvolution algorithms [3] or the multichannel blind separation algorithms [5] maximize the Gray norms.

On the other hand, we may minimize the output entropy directly by estimating the separating matrix by

$$\boldsymbol{W}(z) = \arg min\left\{-E\left\{\sum_{i=1}^{n} \log(f((y_i|\boldsymbol{W}(z))))\right\}\right\}. \quad (4)$$

where f(.) denotes the probability density function.

4. GENERALIZED GAUSSIAN DISTRIBUTION

In Equation (4), it is obvious that we need the *a priori* knowledge of the probability density function, i.e., the statistical model of the estimator. The generalized Gaussian distribution model [3, 4] can be a good candidate. That specific family of symmetric distribution can be characterized by a two parameters set (α, β) as

$$f(y_i) = \frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)} \exp\left(-\frac{|y_i|}{\beta}\right)^{\alpha} \quad (5)$$

where Γ is the gamma function. From Equation (5), we can easily derive

$$\beta^{\alpha} = \alpha E\{|y_i{}^{\alpha}|\}. \quad (6)$$

Hence the parameter β is associated with parameter α, i.e. α is the only parameter which needs to be determined. Figure 2 shows some members from the generalized Gaussian distribution family with different α's all with unity variance.

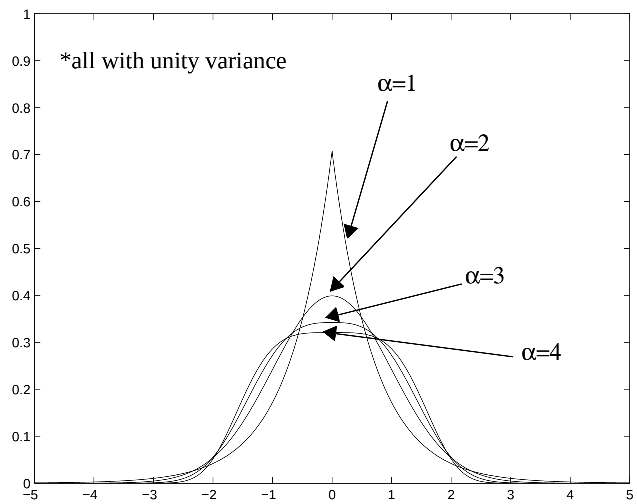


Figure 2. The generalized Gaussian distribution.

5. GOODNESS-OF-FIT TEST

From Figure 2, we can clearly see the relationship between the parameter α and the tailweight of the distribution. For the generalized Gaussian distribution we may obtain the measure of the tailweight, generalized kurtosis, "$K(\rho, y_i)$" as

$$K(\rho, y_i) = \frac{E\{|y_i|^{2\rho}\}}{E\{|y_i|^{\rho}\}^2} = \frac{\Gamma\left(\frac{2\rho+1}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{\rho+1}{\alpha}\right)^2} \quad (7)$$

Equation (7) can be a test measure for goodness-of-fit between the distribution model and the true source PDF. We can utilize Equation (7) to choose the optimal α.

If we choose $\rho = 2$, $K(2, y_i)$ is the kurtosis. Figure 3 will show the kurtoses for different α variables.

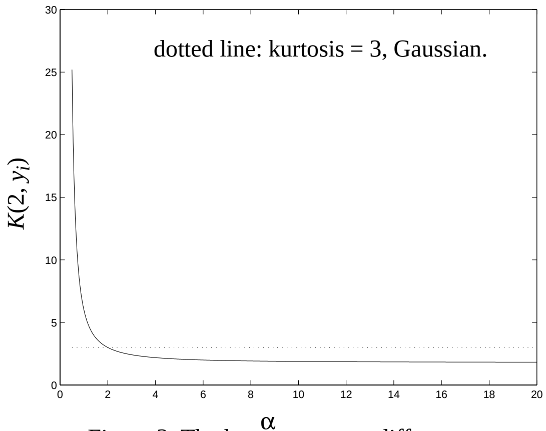


Figure 3. The kurtoses versus different α.

In Figure 3, we may see that we need to model the super-Gaussian sources with α>2, and sub-Gaussian sources with α<2. α can be fractional.

6. GENERALIZED ANTI-HEBBIAN LEARNING

If we replace the marginal distribution $f(y_i|\boldsymbol{W}(z))$ in Equation (4) by the generalized Gaussian model in Equation (5), we can derive the learning rule for source separation. For a feedforward separating system, we can formulate the i^{th}outputs as

$$y_i(t) = x_i + \sum_{j=1, j \neq i} \sum_{k=0}^{M} w_{ij}(k) x_j(t-k) . \quad (8)$$

On the other hand, for a recurrent separating system, we can formulate the i^{th}outputs as

$$y_i(t) = x_i + \sum_{j=1, j \neq i} \sum_{k=1}^{M} w_{ij}(k) y_j(t-k) . \quad (9)$$

Hence we have formulated the generalized anti-Hebbian learning rule for source separation:

$$\Delta w_{ij}(k) = -\eta \frac{x_j(t-k)|y_i(t)|^{\alpha-1} sign(y_i(t))}{E_T\{|y_i(t)|^{\alpha}\}} \quad (10)$$

for the feedforward system, and

$$\Delta w_{ij}(k) = -\eta \frac{y_j(t-k)|y_i(t)|^{\alpha-1} sign(y_i(t))}{E_T\{|y_i(t)|^{\alpha}\}} \quad (11)$$

for the recurrent system, where η is the step size and the E_T operator is the time averaging α-moment estimator with window-length *T*. Two special cases of temporal anti-Hebbian rules have been shown before for Laplacian distribution (kurtosis = 6, α = 1) [6] and Gaussian distribution (kurtosis = 3, α = 2) [7].

7. SIMULATION

7.1 The tailweight of the original sources: kurtoses of audio signals

In this section we will conduct two sets of experiments. For quantification convenience, we use CD audio sources and mix them in a computer with $H_{12}(z) = 0.8z^{-2} + 0.1z^{-3}$ and $H_{21}(z) = 0.7z^{-1} + 0.4z^{-2} + 0.25$, with $H_{11}(z) = H_{22}(z) = 1$. The first pair of sources are two male speakers from the TIMIT database, and the pair of sources in the second experiment are two rock music. The waveforms of the sources are depicted in Figure 4. Before we choose the appropriate anti-Hebbian rule to separate the mixing signals, we utilize the kurtoses as the tailweight measure in order to quantify the goodness-of-fit of the distribution model. Again the kurtosis is defined as

$$k(2, s(t)) = \frac{E_T\{s^4(t)\}}{E_T^{\ 2}\{s^2(t)\}}, \quad (12)$$

where E_T denotes the window averaging operator. This window averaging can be long-term (averaging over the whole data set) or short-time (averaging over the local data only). Table 1 gives the kurtosis estimate computed with the whole data sets.

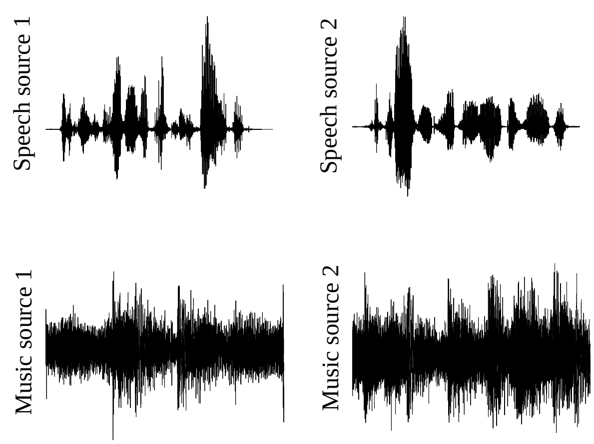
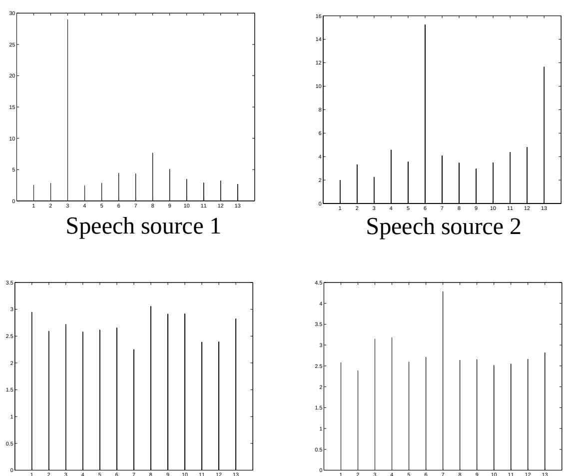


Figure 4. Two pairs of sources.

Figure 5 depicts the short-time kurtoses (20 msec. widow).

TABLE 1. Sources and mixed signals kurtoses

	Speech 1	Speech 2	Music1	Music2
Original	28.65	26.55	4.20	3.82
Mixing	9.62	10.41	3.07	3.07



Speech source 1 Speech source 2

Music source 1 Music source 2

Figure 5. Short-time kurtosis estimates.

7.2 Anti-Hebbian learning with different α values

First we use the recurrent separating system in Equation (9) with the learning rule described in Equation (11) to separate the speech mixture and to compare the signal-to-interference values in dB among different α values. The learning curve is plotted in Figure 6.

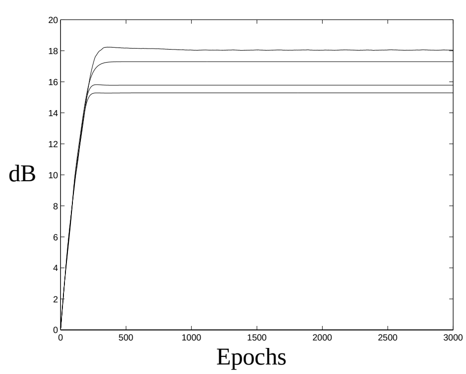


Figure 6. The learning curves by a recurrent system. (From top to bottom: α = 1, 2, 4, 6).

Then we use the feedforward system in Equation (8) with the adaptation rule in Equation (10) to separate the speech mixture as well as the music mixture. Both the results are listed in Table 2.

TABLE 2. The performance of feedforward separator

speech α=0.6	speech α=1	speech α=2	music α=1	music α=2
17.72	18.44	16.20	8.09	8.47

Although speech has large global kurtosis (in Table 1), the local kurtoses will be much smaller (in Figure 5). Hence α = 1 for speech and α = 2 for music gave us the best results as shown in Table 2.

8. CONCLUSION

We derive the generalized anti-Hebbian learning rule for source separation and propose to utilize the kurtoses as a measure to choose the appropriate α. Although the original sources are not observable, we may use the kurtosis estimated from the mixing signals instead to roughly determine the appropriate α. Simulation shows that the appropriate α will lead to better separation performance in generalized anti-Hebbian learning.

REFERENCE

[1] A. J. Bell and T. J. Sejnowski, “An information maximization approach to blind separation and blind deconvolution,” *Neural Computation*, 7, pp. 1129-1159, 1995.

[2] J.-F. Cardoso, “Infomax and maximum likelihood for blind source separation,” *IEEE Signal Processing Letters*, vol. 4, no. 4, April 1997.

[3] W. Gray, *Variable Norm Deconvolution*, Ph. D. dissertation, Stanford University, 1979.

[4] J. Miller and J. Thomas, “Detectors for discrete-time signals in non-Gaussian noise,” *IEEE Transactions on Information Theory*, vol. IT-18, no. 2, March 1972, pp. 241-250.

[5] R. H. Lambert, “A new method for source separation,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 2116-2119, 1995

[6] M. Girolami, “Noise reduction and speech enhancement via temporal anti-Hebbian learning” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 1233-1236, 1998.

[7] C. Wang, H.-C. Wu and J. Principe, “Correlation estimation using teacher forcing Hebbian learning and its application,” in *Proc. 1996 IEEE Int. Conf. on Neural Networks*, Washington D. C., U. S. A., June 1996, pp. 282-287.