A noisy data model for *MaxNG*

E. Salerno
Istituto di Scienza e Tecnologie dell'Informazione, CNR,
Area della Ricerca di Pisa, via G. Moruzzi 1, 56124 Pisa, Italy.

7/09/2006

Abstract

A noise-insensitive Euclidean distance function is derived for the *MaxNG* algorithm. This is a dependent-component-analysis source-separation algorithm based on the maximization of a nongaussianity measure, and has recently been developed for a noiseless mixture model. It is shown that, in the case of observations corrupted by signal-independent stationary Gaussian noise, the probability density function of the output process can be easily made independent of noise if it is approximated via the Parzen-windows method with Gaussian kernels. The role assumed by the aperture parameter is shown to be similar to the one of the regularization parameter in any inverse problem.

1 Introduction

Blind separation of dependent sources is becoming an actively researched area for the interesting applications that can be envisaged for this relatively new signal processing approach.

An algorithm based on the maximization of nongaussianity (*MaxNG*) proposed by Caiafa and Proto [1] has given very promising results in the field of remote-sensed image analysis [2]. Based on some considerations in information geometry, these authors conjecture that an efficient blind separation of dependent sources can be obtained by locally maximizing the Euclidean distance between a standard Gaussian and the pdf of the transformed data vector. No theoretical proof is provided yet, but the effectiveness of the algorithm has been demonstrated experimentally.

Despite its ability to efficiently separate dependent sources, *MaxNG* shows a poor robustness against noise. This can be due to the fact that the algorithm has been derived from a noise-free mixture model, thus assuming a biased estimate of the nongaussianity measure.

In this note, I introduce uniform, signal-independent Gaussian noise in the data model and show that, by following the *MaxNG* strategy from this standpoint, it is very easy to alter the distance function in order to make it less sensitive to noise.

The derivation of the new nonaussianity measure is provided in Section 2, and an intuitive interpretation is attempted in Section 3 where the aperture parameter of the Parzen-windows approximation is shown to act as a regularization parameter.

2 A noise-insensitive nongaussinity measure

Let us assume the usual linear, instantaneous noisy mixture data model with M observed channels, P source processes and N data samples:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t). \tag{1}$$

with the usual meaning for all the symbols. Let the noise process $\mathbf{n}(t)$ be zero-mean, i.i.d., Gaussian, signal-independent. Let us assume that the data $\mathbf{x}(t)$ have already been sphered and that a unit vector $\mathbf{d}$ is available, through which the noiseless data would give a copy of one of the source processes, say, s_k. When applied to the noisy data, this would give

$$y(t) = \mathbf{d}^T\mathbf{x}(t) = \mathbf{d}^T\mathbf{A}\mathbf{s}(t) + \mathbf{d}^T\mathbf{n}(t) = s_k(t) + n'(t). \tag{2}$$

Let us now consider the measure of nongaussianity assumed by the *MaxNG* blind separation algorithm presented in [1].

$$\Gamma(\mathbf{d}) = \int [\Phi(y) - p_y(y)]^2 dy \tag{3}$$

where Φ is the standard Gaussian pdf and y is the random variable obtained as in (2). *MaxNG* tries to estimate a vector $\mathbf{d}$ by maximizing function Γ. This means that the nongaussianity of variable y is maximized. Actually, in the noisy situation described by (1), the presence of the noise process biases the estimation. Indeed, the pdf $p_y(y)$ in (3), for the hypothesis of signal independent noise, will be:

$$p_y(y) = p_{s_k+n'}(y) = [p_{s_k} * p_{n'}](y), \tag{4}$$

where the asterisk denotes convolution. It is now clear how the presence of $p_{n'}$ in (3) produces a biased estimate, as the statistics of n' depend on $\mathbf{d}$. The problem is now to modify function Γ in order to make it insensitive to noise. Since, for fixed $\mathbf{d}$, p_y can be estimated from the output samples $\mathbf{d}^T\mathbf{x}$ and $p_{n'}$ can be derived from the statistics of $\mathbf{n}$, we can try to obtain p_{s_k} by deconvolution. Taking the Fourier transform of (4), we have

$$\tilde{p}_y(\omega) = \tilde{p}_{s_k}(\omega) \cdot \tilde{p}_{n'}(\omega), \tag{5}$$

then

$$\tilde{p}_{s_k}(\omega) = \frac{\tilde{p}_y(\omega)}{\tilde{p}_{n'}(\omega)}, \tag{6}$$

from which the desired pdf can be obtained by inverse transformation.

Let us now see the forms assumed by p_y and $p_{n'}$. For p_y, we follow the same strategy adopted in [1] to estimate the output pdf, namely the nonparametric approximation based on Parzen windows:

$$p_y(y) \approx \frac{1}{Nh} \sum_{t=0}^{N-1} \Phi\left[\frac{y - y(t)}{h}\right], \tag{7}$$

where h is the standard deviation of the Gaussian kernels used as window functions. For $p_{n'}$, it is easy to see that it is zero-mean Gaussian with variance $\mathbf{d}^T\Sigma\mathbf{d}$, where Σ is the covariance matrix of the M-dimensional process $\mathbf{n}$:

$$p_{n'}(y) = \frac{1}{\sqrt{\mathbf{d}^t \Sigma \mathbf{d}}} \Phi \left[\frac{y}{\sqrt{\mathbf{d}^T \Sigma \mathbf{d}}} \right] \tag{8}$$

To estimate function p_{s_k} on the basis of Equation (6), we need to evaluate the characteristic functions $\tilde{p}_y(\omega)$ and $\tilde{p}_{n'}(\omega)$ from their density counterparts in (7) and (8), respectively. This is an easy task, since the Fourier transform of a Gaussian is still a Gaussian. For the characteristic function of the output variable, we have

$$\tilde{p}_y(\omega) = \frac{1}{Nh} \sum_{t=0}^{N-1} \tilde{\Phi}_{y(t)}(\omega), \tag{9}$$

with

$$\tilde{\Phi}_{y(t)}(\omega) = h e^{-\frac{1}{2} h^2 \omega^2} e^{-j\omega y(t)}. \tag{10}$$

Similarly, for the output noise, we have the following characteristic function:

$$\tilde{p}_{n'}(\omega) = e^{-\frac{1}{2} \mathbf{d}^T \Sigma \mathbf{d} \omega^2}. \tag{11}$$

From Equations (6) and (9)-(11), we easily obtain

$$\tilde{p}_{s_k}(\omega) = \frac{1}{N} e^{-\frac{1}{2}(h^2 - \mathbf{d}^T \Sigma \mathbf{d})\omega^2} \sum_{t=0}^{N-1} e^{-j\omega y(t)} \tag{12}$$

whose inverse transform is

$$\begin{aligned} p_{s_k}(y) &= \tfrac{1}{2\pi} \tfrac{1}{N} \sqrt{\tfrac{2\pi}{h^2 - \mathbf{d}^T \Sigma \mathbf{d}}} \cdot \textstyle\sum_{t=0}^{N-1} e^{-\frac{(y-y(t))^2}{2(h^2 - \mathbf{d}^T \Sigma \mathbf{d})}} \\ &= \tfrac{1}{N\sqrt{h^2 - \mathbf{d}^T \Sigma \mathbf{d}}} \textstyle\sum_{t=0}^{N-1} \Phi \left[\tfrac{y - y(t)}{\sqrt{h^2 - \mathbf{d}^T \Sigma \mathbf{d}}} \right]. \end{aligned} \tag{13}$$

Taking signal-independent Gaussian noise into account thus simply means to modify the aperture of the Parzen windows. The distance to be maximized, instead of (3), now becomes

$$\Gamma(\mathbf{d}) = \int [\Phi(y) - p_{s_k}(y)]^2 dy \tag{14}$$

where p_{s_k} is evaluated from (13). This does not imply any significant increase in computational needs.

3 Discussion

Let us consider the intuitive meaning of the result in (13). In the noiseless case, we estimate the unit vectors $\mathbf{d}$ as the local maximizers of function $\Gamma(\mathbf{d})$ in (3). The presence of noise makes the estimate biased because the transformation of the sphered data does not yield the source pdf, but a blurred version of it. The entity of the blur is proportional to the standard deviation of the noise process. A deconvolution is particularly simple to perform when the noise is Gaussian and the output pdf is estimated via the Parzen method with Gaussian kernels. In this case, an important role is played by the aperture parameter h. Reducing h^2 by a quantity $\mathbf{d}^T\Sigma\mathbf{d}$ means to try to recover the smoothing caused by the noise on the output pdf. If the noise is very strong, however, the estimation of p_{s_k} as in Equation (13) becomes too "spiky", due to the excessively narrow windows used. The reasons of this phenemenon are also clear by observing Equation (6). Having a strong noise component means to let $\tilde{p}_{n'}$ go to zero faster than $\tilde{p}_y$, thus making the result unstable. In the limit, when noise becomes so strong that $\mathbf{d}^T\Sigma\mathbf{d} \geq h^2$, Equation (13) becomes meaningless. Parameter h thus plays the role of a *regularization parameter*. It establishes a compromise between an assumed smoothness of the source pdf and the oversmoothing caused by the presence of noise. Normally, h is established on the basis of the number N of samples available (see [1]). If the output sequence, however, is made by a source sequence plus Gaussian noise, a reduction of h can recover the excessive smoothness of its pdf when compared to the pdf of the noiseless variable. For increasing noise, parameter h must be increased too, to avoid an incorrectly rough estimation of the source pdf. Increasing h, however, means to gain in stability but losing in resolution. As always, a compromise between the reliability of the data and the reliability of the prior information is needed. If prior information is generically "smoothness", its prevalence over the data means oversmoothing, that is loss of resolution. The situation is illustrated visually in Figures 1 and 2, for two different levels of noise. In these figures, the Parzen windows approximated pdfs are shown, as obtained from two noiseless sequences of 10,000 samples uniformly distributed between 0 and 1. Along with these plots, there are also the plots related to the same sequences corrupted by Gaussian noise, with and without the correction in Equation (13). The aperture h used for the original sequences is 0.084. For the case of Figure 1, the noise standard deviation was 0.05, corresponding to an SNR of about $15dB$. The pdf estimated from the noisy sequence with the same aperture is apparently smoother than the original. Conversely, the estimate corrected via (13) ($h = 0.067$) coincides almost prefectly with the estimate in the noiseless case. In Figure 2, the differences are more evident since the noise standard deviation is now 0.08 ($SNR \approx 11dB$). In this case, the original slope in the flanks of the estimated pdf is well recovered by the proposed correction, but the resulting aperture ($h = 0.026$) is too small to get a sufficiently smooth estimate in the interval $0 - 1$.

References

[1] C. F. Caiafa, A. N. Proto, "Separation of statistically dependent sources using an L^2-distance non-Gaussianity measure", *Signal Processing*, doi:10.1016/j.sigpro.2006.02.032

[2] C. F. Caiafa, E. Salerno, A. N. Proto, L. Fiumi, "Dependent component analysis as a tool for blind spectral unmixing of remote sensed images", *Proc. EUSIPCO 2006*, Florence, Italy, 4-8 September 2006.

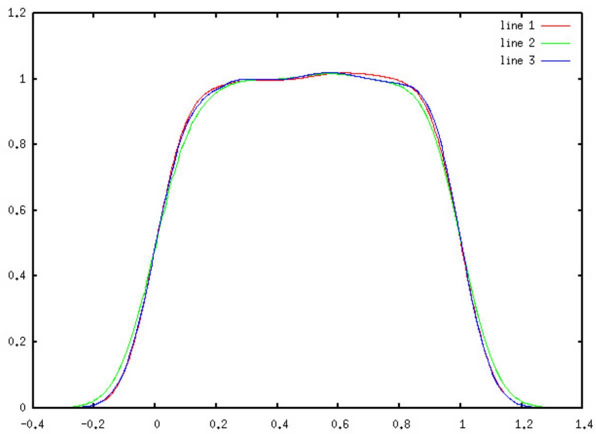

Figure 1: Parzen windows pdf approximation, 10,000 samples uniformly distributed in the range 0-1. Line 1: noiseless sequence, $h = 0.084$. Line 2: noisy sequence, $SNR = 15dB$, $h = 0.084$. Line 3: noisy sequence, $SNR = 15dB$, $h = 0.067$.

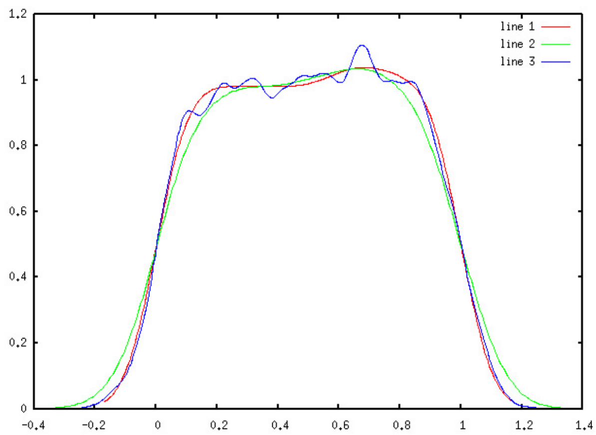

Figure 2: Parzen windows pdf approximation, 10,000 samples uniformly distributed in the range 0-1. Line 1: noiseless sequence, $h = 0.084$. Line 2: noisy sequence, $SNR = 11dB$, $h = 0.084$. Line 3: noisy sequence, $SNR = 11dB$, $h = 0.026$.