ROBUST APPROACH FOR BLIND SOURCE SEPARATION IN NON-GAUSSIAN NOISE ENVIRONMENTS

Mohamed Sahmoudi¹, Hichem Snoussi² and Moeness G. Amin¹

¹Center for Advanced Communications, Villanova University, Villanova, PA 19085, USA
²ISTIT/M2S, Université de Technologie de Troyes, France

E-mail: {mohamed.sahmoudi, moeness.amin}@villanova.edu    hichem.snoussi@utt.fr

ABSTRACT

In this contribution, we address the issue of Blind Source Separation (BSS) in non-Gaussian noise. We propose a two-step approach by combining the fractional lower order statistics (FLOS) for the mixing matrix estimation and minimum entropy criterion for noise-free source components estimation with the gradient-based BSS algorithms in an elegant way. First, we extend the existing gradient algorithm in order to reduce the bias in the denoising matrix caused by the non-Gaussian noise. In the noise cancellation step, we derive a new kind of nonlinear function that depends on the noise distribution and we discuss the optimal choice of this nonlinearity assuming a generalized Gaussian noise model. The optimal choice, in the minimum entropy sense, is robust against the influence of Gaussian and non-Gaussian noise including heavy-tailed model. The effectiveness and the robustness of the proposed separating algorithm are shown on numerical examples.

1. INTRODUCTION

A major problem with the existing BSS approaches is that, a vast majority of them has been conducted with the assumption of noise-free mixtures. While only few approaches have been developed for the additive Gaussian noise case, their performance is degraded if the additive noise has heavy-tailed non-Gaussian distribution. In this paper, we consider the classical noisy linear BSS model with instantaneous mixtures given by

\[ x(t) = As(t) + \varepsilon(t), \quad t = 1 \ldots T \]  

where \( A \) is a \( n \times m \) unknown full column rank mixing matrix. The sources \( s_1(t), \ldots, s_m(t) \) are collected in a \( m \times 1 \) vector denoted \( s(t) \) and are assumed to be mutually statistically independent. The vector \( \varepsilon(t) \) denotes the additive noise assumed to be independent with \( s(t) \) and possibly non-Gaussian impulsive with heavy-tailed probability distribution.

The goal of a BSS method is to find a separating matrix \( W \) i.e. an \( m \times n \) matrix such that the recovered sources \( Wx(t) \) are as independent as possible. In the noiseless case, (1) admits a unique solution up to scaling and permutation indeterminacy \( y(t) = Wx(t) \) such that \( C \triangleq WA = P\Lambda \), where \( \Lambda \) is a diagonal scaling matrix and \( P \) is a permutation matrix (e.g. see [3]).

In recent years several algorithms were proposed for BSS that perform well in many situations [2, 3, 9]. But when the data present an impulsive heavy-tailed distribution behavior, these methods may lead to wrong conclusions [1, 2, 5]. In such scenario, it is important to distinguish between three possible cases: (i) When one or more sources do not have finite second or higher moments (heavy-tailed sources) for which we have introduced several solutions, using FLOS [8], using the concept of normalized statistics [6], and a semi-parametric approach of the maximum likelihood principle [7]; (ii) When the data contain outliers, many existing BSS algorithms are satisfactory robust in this sense like the well known FastICA algorithm [3], and other preprocessing procedures have been proposed to clean properly the data before performing the separation task [1, 2]; and (iii) In the case when the additive noise is non-Gaussian with impulsive nature, existing BSS algorithms can not work because pre-whitening or criteria optimization would cause a breakdown. While many algorithms for BSS have been designed to be robust to outliers and additive Gaussian noise, fewer algorithms for the non-Gaussian noise case have been proposed [5].

In noisy BSS, we also encounter a new problem which is the estimation of the noise-free sources components. The noisy model is not invertible, and therefore estimation of the noise-free sources components requires new methods. In other words, it is not enough to estimate the mixing matrix. Indeed, inverting the mixing matrix in (1), we obtain

\[ Wx(t) = s(t) + W\varepsilon(t) \]

Then, we only get noisy estimates of the independent components. Therefore, we would like to obtain estimates of the
original source components $\hat{s}_i$ that contain minimum noise. Next, we resume the essential ideas of the proposed two-stage BSS procedure.

As it is well known that the fractional lower order statistics (FLOS) are better for signal processing in impulsive noise than second order statistics (SOS) and higher order statistics (HOS) [8], we propose in this paper a two-step procedure based on the FLOS. First, we introduce a new FLOS-based gradient algorithm for the mixing matrix estimation. In the second step, we adopt the minimum entropy criterion to estimate the source components (noise cancelation) using a gradient algorithm. So the global procedure estimation capabilities become greatly improved under both Gaussian and non-Gaussian noise.

2. ROBUST QUASI-WHITENING

The standard decorrelation or pre-whitening algorithm for $x(t)$ in the noiseless case can be presented as

$$W(t+1) = W(t) + \eta(t)[I - E[y(t)y^T(t)]]W(t)$$

where $y(t) = Wx(t)$, $I$ is the identity matrix, and $E[.]$ denotes mathematical expectation. The stochastic gradient version of this algorithm for noisy data with finite second-order statistics is given by (see, e.g. [3])

$$W(t+1) = W(t) + \eta(t)[I - R_y + WR_tW^T]W(t)$$

where $R_y = E[y(t)y^T(t)]$ and $R_t = E[\epsilon(t)\epsilon(t)^T]$. However, in the case of heavy-tailed noise the covariance matrix has infinite values and then can not be exploited. In that case, we have shown in [8] that the covariance matrix normalized by its trace is of finite values. Here, we use this normalized covariance to extend the above pre-whitening procedure to preprocess data in heavy-tailed noise. Indeed, if we replace heuristically $R_y$ by $R_y/\text{Trace}(R_y)$ and $R_t$ by $\gamma I$ in the above algorithm, we obtain a suitable whitening algorithm for heavy-tailed data.

$$W(t+1) = W(t) + \eta(t)[I - \frac{R_y}{\text{Trace}(R_y)} + \gamma WW^T]W(t)$$

where $\gamma$ is the dispersion parameter that measure the impulsive noise level. This procedure is called robust quasi-whitening.

3. FLOS GRADIENT ALGORITHM FOR MIXING MATRIX ESTIMATION

In order to extract one source from the mixture, one computes the inner product of the observations with a vector $w_i$, to obtain the output random variable or estimated source

$$y_i(t) = w_i^T x(t)$$

A well known BSS approach uses a criterion based on HOS like the normalized kurtosis

$$J(w_i) = \frac{1}{4}\kappa_4(y_i) = \frac{E[y_i^4(t)]}{(E[y_i^2(t)])^2} - 3$$

assuming that observed signals are pre-whitened (sphered) and assumes that the sources are non-Gaussian (at most only one can be Gaussian). By analogy to this approach, we will use some FLOS-based measures of non-Gaussianity. Thus as a cost function for minimization, we may employ

$$J(w_i) = \frac{1}{p}E[|y_i(t)|^p] = \frac{1}{p}E[|x^T(t)w_i|^p]$$

where $p > 0$. For the case of impulsive noise $p$ must be in $0 < p < \alpha \leq 2$, where $\alpha$ is the characteristic parameter of the noise heavy-tailed distribution. Thus, (8) become also a sparsity criterion and then is more adequate for heavy-tailed mixtures. To impose the orthogonality constraint $w_i^T w_i^T = I$, we will introduce a new cost function where the separation is achieved by considering a mixed separation criterion combining two terms; the pre-whitening condition and the fractional lower order nonlinearity

$$J(w_i) = \frac{1}{p}E[|x(t)^T w_i|^p]$$

$$+ \frac{1}{2} \beta_{ij} w_i^T w_j - 1 + \frac{1}{2} \sum_{j \neq i} \beta_{ij} w_i^T w_j$$

where $\beta_{ij}$ are the Lagrange multipliers. Using natural gradient descent to increase the non-Gaussianity, we get the so called FLOS gradient algorithm

$$w(l+1) = w(l) - \eta \nabla_{w_i} J(w_i(l))$$

where $\eta$ is a step size. The gradient $\nabla_{w_i} J(w_i(l))$ of the cost function (9) can be expressed as

$$\nabla_{w_i} J(w_i) = E[x(t)x^T(t)w_i] |^{p-2} x^H(t) w_i + \sum_{j=1}^m \beta_{ij} w_j$$

where $w_i$ denotes the complex conjugate of $w_i$, and $x^H(t)$ denotes the conjugate transpose of the complex vector $x(t)$. Then, $w_i$ is a separating vector if

$$\nabla_{w_i} J(w_i) = 0 \quad ; \quad i = 1, \cdots, m$$

$$w_i^T w_j = \delta_{ij}$$

Multiplying from the left side the above gradient (12) by $w_j^T$, we obtain that

$$\beta_{ij} = - w_j^T E[x(t)x^T(t)w_i] |^{p-2} x^H(t) w_i$$

This leads to a simple gradient expression given by

$$\nabla J(w_i) = (I - \sum_{j=1}^m w_j w_j^T) E[x(t)x^T(t)w_i] |^{p-2} x^H(t) w_i$$

(14)
Estimating the fractional moments by their sample averages and using the fact that \( y_i(t) = x^T(t)w_i(t) \), we can simplify the gradient algorithm to

\[
\Delta w = -\eta \langle |y_i(t)|^{p-2}y_i(t) \rangle \left[ w_i(l)y_i(t) - \sum_{j=1}^{m} w_j(l)y_j(t) \right]
\]

where \( \langle \cdot \rangle \) is an averaging operator. If we consider \( \varphi(u) = |u|^{p-2}u = |u|^{p-1} \text{sign}(u) \) as the appropriate nonlinearity chosen to deal with the heavy-tailed noise, then we find again the unified form of a BSS gradient algorithm as

\[
\Delta W = -\eta \varphi(y(t)) \left[ W^T(l)y(t) - W(l)y(t) \right] \varphi(y^T(t))
\]

The term \( \varphi(y) \) is a vector of the so-called \textit{first kind of nonlinearities} whose forms depend on the probability density functions (pdf) of the extracted output signals \([2, 3]\). The used fractional nonlinearity in our FLOS algorithm limits the power of the nonlinearities such that the algorithm will not blow up and then suitable for data with impulsive nature. Our simulations show that multiple choices of weighting functions are applicable and all give generally good performance. In general, we propose to choose

\[
\varphi(u) = \text{sign}(u)\rho(u)|u|^{p-1}\frac{\partial \rho}{\partial u}
\]

where \( \rho(.) \) is a nonlinear function properly chosen with relation to the source distributions, and \( 0 < p < \alpha \) in the case of heavy-tailed distribution.

### 4. SOURCE COMPONENTS ESTIMATION

In this section, we show that there is a \textit{second kind of nonlinearity} that depends on the additive noise distribution. Then, we restrict ourselves to investigate the relation between the so-called second kind nonlinearity and the adopted heavy-tailed noise distribution. Assume that we have successfully obtained an unbiased estimate of the separating matrix \( W \), for example via the previously described approach. Then, we can estimate a mixing matrix \( \hat{A} = W^\top \), where \( W^\top \) is the pseudo-inverse of \( W \). In this second BSS step, we describe a method for cancelling the effects of noise in the estimated source signals. For that, we define the error vector \( e(t) = x(t) - \hat{A}y(t) \), where \( y(t) \) is an estimate of the source \( s(t) \). We consider the minimum entropy criterion to estimate the sources:

\[
\hat{y}(t) = \arg\min E\{\Phi(e(t))\} = \arg\min\left\{-\sum_{k=1}^{m} \log(f_k(e_i(t)))\right\}
\]

where \( f_k(e_i) \) is the p.d.f of the additive noise component \( e_i(t) \). Since the noise components are i.i.d., the gradient descent of the minimum entropy cost yields independence of the error components as well as the minimization of their magnitude. The resulting algorithm is

\[
\hat{y}(t+1) = \hat{y}(t) + \mu \hat{A}^T G(e(t))
\]

where \( G(e(t)) = [G_1(e_1), \ldots, G_m(e_m)] \) is the vector if the second kind of non lineairties \( G_i(e_i) = -\partial \log f_i(e_i)/\partial e_i \). Clearly, in the proposed algorithm, the optimal choice of second kind nonlinearities depends on the noise distributions. Assume that the noise signals have generalized Gaussian distributions of the form

\[
f_i(e_i) = \frac{\alpha_i}{2\gamma_i\Gamma(1/\alpha_i)} \exp\left[-\frac{1}{\alpha_i \gamma_i^{1/\alpha_i}} |e_i|^{\alpha_i} \right]
\]

where \( \alpha_i > 0 \), \( \Gamma \) is the Gamma function and \( \gamma_i \) denotes the dispersion parameter such that \( \gamma_i^{1/\alpha_i} = E(|e_i|^{\alpha_i}) \). We note that the dispersion is a general measure of the noise power and the value \( \alpha_i = 2 \) yields the standard Gaussian distribution. For the heavy-tailed noise case, we have to choose \( \alpha_i \) in \((0, 2]\), in which case the locally optimal second kind nonlinearity functions are of the form

\[
G_i(e_i) = -\frac{\partial \log f_i(e_i)}{\partial e_i} = |e_i|^{\alpha_i - 1} \text{sign}(e_i)
\]

In addition, the optimal parameter \( \alpha_i \) typically takes a value between zero and one. In such case, we can use the modified nonlinearity function

\[
G_i(e_i) = \frac{e_i}{|e_i|^{2-\alpha_i} + \xi}
\]

where \( \xi \) is a small number to avoid the singularity of the absolute value function at zero. In practice, the noise distribution parameters are unknown and then one must choose robustly the parameter \( \alpha_i \). We keep the same minimum noise criterion to update \( \alpha_i \). Thus, a simple gradient-based rule for adjusting each parameter \( \alpha \) is

\[
\alpha_i(t + 1) = \alpha_i(t) + \eta_i \frac{\partial \log f_i(e_i)}{\partial e_i} \\
\approx \eta_i \frac{|e_i|^{\alpha_i(t) - 1}(1 - \log |e_i|^{\alpha_i(t)})}{\alpha_i^2(t)}
\]

The proposed noise cancellation algorithm (20) can be considered as a form of nonlinear postprocessing that efficiently reduces the additive noise component in the estimated source signals.
5. SIMULATIONS

To illustrate the performance of our proposed approach, we have generated measurements according to (1), where \( n = m = 3 \), \( s(t) = [s_1, s_2, s_3]^T \); \( s_1 = \sin(2\pi 90t) \), \( s_2 \) is Gaussian source and \( s_3 \) is a heavy-tailed Cauchy distributed source. The mixing matrix \( \mathbf{A} \) is generated randomly and the FLOS are computed using \( p = 1 \). The total number of data samples was \( T = 5000 \). The sensor signals were contaminated by an additive impulsive noise generated using the generalized Gaussian model (21) with \( \alpha = 1.5 \).

In a first illustrative example, \( \gamma \) has been chosen such that the noise power is \(-15 \) dB. As shown in Fig. 1, the proposed procedure allows accurate estimation of the original sources.

Fig. 1. The first three signals are the original sources, the second three signals are the mixed sensors contaminated by impulsive noise, and the last three signals are the estimated source signals.

\[
GRMSE = \frac{1}{m} \sum_{t=1}^{m} \frac{1}{T} \sum_{t=1}^{T} (s_i(t) - \hat{s}_i(t))^2
\]

versus the noise power of both our FLOS gradient and the FastICA algorithm [3]. In simulation experiments, the results are averaged over 100 iterations. Eventhough the non Gaussian noise power is not very high, the proposed FLOS gradient procedure largely outperforms FastICA.

Fig. 2. GRMSE versus the noise power.

6. REMARKS AND CONCLUSION

In this paper, a robust procedure for performing blind source separation in the presence of non-Gaussian noise is introduced. The derived gradient type algorithms are based on the FLOS statistics and are performed in two stages; the separating matrix estimation and the noise effect reduction on the extracted source components. Simulations show the robustness of the algorithms in the presence of heavy-tailed noise contamination. In addition, we emphasize that FLOS are a relevant information for robust BSS: In a heavy-tailed noise environment, the use of fractional lower order statistics can improve drastically the robustness of the designed BSS procedure.

Generalization and universal nonlinearity: The how a non linearity should related in general to the additive noise is still an open problem. More studies need to be undertaken for selecting optimal FLOS and weighting functions.

7. REFERENCES


