Improving Stability in Blind Source Separation with the Stochastic Median Gradient

J.J. Murillo-Fuentes and F.J. González-Serrano

22nd November 2000

Abstract

A new on-line learning algorithm is derived for blind separation of mixed signals with symmetric distributions. Stability of the method is analyzed in comparison to the natural gradient method. It is proved that the set of stability conditions obtained is less stringent. Some experiments are included.

1 Introduction

The problem of blind source separation (BSS) arises in many areas. Accordingly, several on-line algorithms have been proposed. Many of them may be rewritten using the maximum likelihood (ML) approach and the stochastic natural gradient (SNG) [1]. The method was studied in [2] providing the conditions for stability.

We first present a new gradient form: the “stochastic median gradient” (SMG). We relate it to the SNG and show how the concept of median replace the mean, reducing stability conditions. The Hessian is computed for both of the gradients to prove it.

Only the case of sources with symmetric distributions is considered in this paper.

2 Model of mixtures and Natural Gradient

The simplest source separation model is that of an $n \times 1$ vector of observations with structure $\mathbf{x}(t) = A \mathbf{s}(t)$ where $A$ is an invertible $n \times n$ unknown matrix and $\mathbf{s}$ is an unobserved $n \times 1$ vector. The task is to recover the source signals or/and to identify matrix $\mathbf{W} = A^{-1}$ such that $\mathbf{y} = \mathbf{Wx} = \mathbf{WA}s$ using only the assumption of source independence. Matrix $A$ may be computed up to permutations and changes of scales and signs.
Given the source distributions \( q_1(y_1), q_2(y_2), \ldots, q_n(y_n) \), the ML approach proposes to minimize:

\[
L(W) = \log |\text{det}(W)| - E \left[ \sum_{i=1}^{n} \log(q_i(y_i)) \right] = l_W(W) + E[l_y(x, W)]
\]  

(1)

The steepest descent method updates \( W \) according to the direction of the gradient of the contrast function \( \nabla L(W) \). The natural gradient proposes to use \( \nabla L(W) = \nabla L(W)W^T W \). The stochastic version substitutes \( \nabla E[l_y(x, W)] \) by its instantaneous value. The learning law yields

\[
W_{t+1} = W_t + \mu (I - \varphi(y)y^T)W_t
\]

(2)

where in ML approach \( \varphi_i(y_i) = -d_i(y_i)/q_i(y_i) \). However, as source distributions are unknown each author introduces his own activation function \( \varphi_i(y_i) \). These functions should meet stability conditions. The associated stability conditions for (2) were found to be [2]

\[
m_i + 1 = E[y_i^2 \varphi_i'(y_i)] + 1 > 0
\]

(3)

\[
k_i = E[\varphi_i'(y_j)] > 0
\]

(4)

\[
\sigma_i^2 \sigma_j^2 k_i k_j > 1 \text{ for all } i, j \ (i \neq j)
\]

(5)

where \( \sigma_i^2 = E[y_i^2] \). Besides, the following normalization condition holds at separation:

\[
E[\varphi_i(y_i)y_i] = 1
\]

(6)

We now analyze the stochastic median gradient.

3 Stochastic Median Gradient

This paper proposes to introduce the change \( W \leftarrow D^{-1/2}W \) where \( D = \text{diag}([y_1, y_2, \ldots, y_n]) \) (recall that matrix \( W \) may be computed up to scales) so that the SMG yields

\[
\nabla l_y(W) = \nabla l_y(x, W) W^T D^{-1} W
\]

(7)

and (2)

\[
W_{t+1} = W_t + \mu (I - \varphi(y)\text{sign}(y)^T)W_t
\]

(8)

**Proposition 1** the learning law in (8) stops at separation.
At separation, outputs are statistically independent. So that diagonal and crossed elements of

\[ H = E \left[ I - \varphi(y) \text{sign}(y)^T \right] \]  

(9)

cancel. In effect

\[ h_{ii} = E[1 - \varphi_i(y_i) \text{sign}(y_i)] = 0 \]  

(10)

if

\[ E[\varphi_i(y_i) \text{sign}(y_i)] = 1 \]  

(11)

(normalization condition). On the other hand

\[ h_{ij} = E[\varphi_i(y_i) \text{sign}(y_j)] = 0; \]  

(12)

since sources have symmetric distributions. A geometrical perspective of the problem may be found in [3]: if the SNG makes crossed terms cancel by imposing zero mean sources, the SMG requires zero median signals. Next, we compute the Hessian to study the stability of the algorithm at separation. Operating as in [2] we obtain

\[ d^2l = \text{sign}(y)^T dX^T \varphi(y) dX \ y \]  

(13)

where \( dX = dW \ W^{-1} \) and \( \varphi \) is the diagonal matrix with diagonal entries \( \varphi_i(y_i) \). Then

\[
E[d^2l] = \sum_{i \neq j} E[\text{sign}(y_i) \ dx_{ij} \ \varphi_j'(y_j) \ dx_{jk} \ y_k] = \\
= \sum_{i \neq j} E[\varphi_i(y_i)] E[\varphi_j'(y_j)] dx_{ij}^2 + \sum_i E[\varphi_i(y_i)] dx_{ii}^2 = \\
= \sum_{i \neq j} \hat{m}_i dx_{ij}^2 + \sum_i \hat{m}_i dx_{ii}^2 = \sum_{i \neq j} q_{ij} + \sum_i (\hat{m}_i) dx_{ii}^2
\]  

(14)

Conditions for stability yield

\[ \hat{m}_i = E[\varphi_i(y_i)] > 0 \]  

(15)

\[ k_i = E[\varphi_j'(y_j)] > 0 \]  

(16)

In the following, conditions (3)-(5) and (15)-(16) will be compared taking into account their respective normalization conditions (6) and (11).

First of all, expressions (4) and (16) are the same one. On the other hand, inequality (5) does not need to hold using SMG. Finally, we review equations (3) and (15).
Proposition 2 If sources have symmetric distributions, inequalities (3) and (15) hold for any $q_i(y_i)$ if $\varphi_i(y_i)$ has all its odd Taylor’s coefficients positive or null.

Function $\varphi_i(y_i)$ may be decomposed using its Taylor’s coefficients into even and odd parts. As distributions are symmetric, the even part cancels when computing $k_i$, $m_i$, $\tilde{m}_i$ and normalization conditions. On the other hand, if all odd coefficients are negative or null, normalization conditions never hold. Besides, if any odd coefficient is negative, then the odd part of $\varphi_i(y_i)$ is negative at some $y_i$. Then, learning laws in (8) and (2) do not converge to the normalization conditions for any distribution of $y_i$.

Corollary 1 If sources have symmetric distributions, the SMG is stable at separation if the odd Taylor’s coefficients of the activation function $\varphi_i(y_i)$ are greater or equal to zero.

Corollary 1 provides a sufficient condition. If stability were analysed together with convergence, it would turn into a necessary condition.

4 Experimental Results

The performance index

$$Q = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} \frac{|p_{ij}|}{\max_k |p_{ik}|} - 1 \right) + \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \frac{|p_{ij}|}{\max_k |p_{kj}|} - 1 \right)$$

(17)

where $P = (p_{ij}) = WA$, is computed in Fig. 1 for the SNG and the SMG with activation function $\varphi(y) = y^5$. Besides, in Fig. 2 we evaluate condition (5). Note that at $N = 1.510^1$ and $N = 6.510^1$, condition (5) does not hold and the SNG becomes unstable.

5 Conclusions

The “natural gradient” was presented in [1] as the steepest descent direction. However, this result does not analyse the stochastic case. It just use the instantaneous value. We exploit the invariance respect to scale changes in the BSS problem to derive a new gradient: the “stochastic median gradient”. For symmetric distributions, we obtain less stringent stability conditions.
Figure 1: Index $Q$ for a mixture of a symmetric exponential and an uniform distribution. \ldots SNG \ldots SMG

Figure 2: Value of $\sigma_i^2 \sigma_j^2 k_{ij}$ for the SNG case in Fig. 1.

References


**Authors’ affiliations:**
J.J. Murillo-Fuentes and F.J. González-Serrano (Area de Teoría de la Señal y Comunicaciones, Universidad Carlos III, C/Butarque 15, 28911 Leganés, Madrid, Spain, e-mail: murillo@tsc.uc3m.es and fran@tsc.uc3m.es)