ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

Discriminacy of the minimum range approach to blind separation of bounded sources

Dinh-Tuan Pham[1] and Frédéric Vrins[2] *

1- Laboratoire de Modélisation et Calcul (IMAG) and C.N.R.S
51 rue des Mathématiques, Grenoble (France) - dinh-tuan.pham@imag.fr

2- Machine Learning Group - Université catholique de Louvain
3 Place du Levant, Louvain-la-Neuve (Belgium) - frederic.vrins@uclouvain.be

Abstract. The Blind Source Separation (BSS) problem is often solved by maximizing objective functions reflecting the statistical dependency between outputs. Since global maximization may be difficult without exhaustive search, criteria for which it can be proved that all the local maxima correspond to an acceptable solution of the BSS problem have been developed. These criteria are used in a deflation procedure. This paper shows that the "spurious maximum free" property still holds for the minimum range approach when the sources are extracted simultaneously.

1 Introduction

Blind Source Separation (BSS) aims at recovering source signals from mixtures of them only based on mild assumptions on the sources and on the mixing scheme, justifying the "blind" term. We are interested here in the basic and most common mixture model[1]: $\mathbf{X} = \mathbf{AS}$, where $\mathbf{X} = [X_1, \cdots, X_K]^{\mathrm{T}}$ and $\mathbf{S} = [S_1, \cdots, S_K]^{\mathrm{T}}$ are respectively the observed mixtures and the source vectors, both of dimension K, and $\mathbf{A}$ is the nonsingular mixing matrix of order K.

In the blind context, no specific knowledge on the source is available except the basic assumption of their independence. Thus, one looks for an unmixing matrix $\mathbf{B}$ such that the extracted sources, which are the components of $\mathbf{Y} = \mathbf{BX}$ are the most independent in some sense. This approach often leads to the maximization of an objective function which possesses the contrast property according to Comon [2]: it is maximized if and only if its argument $\mathbf{B}$ equals $\mathbf{A}^{-1}$ up to a left multiplication by a diagonal and a permutation matrices.

Recently, both simulation and theoretical approaches have indicated that entropy based BSS contrast functions suffer from the existence of spurious maxima (see [8] and references therein); these contrasts are not *discriminant*, in the sense that each of their local maximum does not necessarily correspond to a non-mixing solution of the BSS problem. Hence, adaptive methods such as gradient ascent can yield unmixing matrix $\mathbf{B}$ such that $\mathbf{BA}$ is not the product of a diagonal and a permutation matrix. On the other hand, it also exists contrasts for which the discriminacy property has been proved, but only when the sources are extracted one by one, as recalled in the next section. In section 3, we show

*The authors thank the Scientific Cooperation between *Communauté Wallonie-Bruxelles* and France.

that this property still holds for the minimum range approach even when the sources are separated simultaneously.

2 Discriminant contrasts for deflation

The possible existence of "spurious maxima" of a contrast is a critical issue in BSS, which often involves iterative optimization algorithms. This point has motivated the work of Delfosse and Loubaton in 1995, even though at that time, the existence of such maxima was not established. In order to avoid this possible problem, the authors of [3] proposed to extract the sources sequentially, one by one, using a deflation approach. Assuming that $k-1$ sources have already been extracted, a k-th source can be extracted by maximizing a non-Gaussianity index of $\mathbf{b}_k\mathbf{X}$, with respect to the k-th row $\mathbf{b}_k$ of $\mathbf{B}$, subjected to the constraint that i) $\mathbf{b}_k\mathbf{X}$ has unit variance and ii) is non correlated with $\mathbf{b}_j\mathbf{X}$ for $j < k$. Many non-Gaussianity indexes possess the contrast property in the sense that they can be maximized if and only if their argument $\mathbf{b}_k\mathbf{X}$ is proportional to the source with the highest non-Gaussianity index (among the $K+1-k$ sources which have not yet been extracted). The maximum square kurtosis $\kappa^2(\mathbf{b}_k\mathbf{X})$ has been suggested as a *deflation* objective function for BSS, where $\kappa(y)$ is the kurtosis of y. It is proved in [3] that all the local maxima of this function are attained when the k-th output is proportional to one source; the contrast is discriminant. The global maximization thus reduces to a local maximization, which is much simpler and can be achieved by using gradient-ascent methods.

Recently, Vrins et al. [4] proved that the minimum range criterion also possesses this interesting "discriminacy" property when separating bounded sources. We define the range $R(Y)$ of a bounded random variable Y as the difference between the upper bound and lower bound of the support of Y. The minimum range deflation approach consists in minimizing successively $R(\mathbf{b}_k\mathbf{X})$, $k = 1, \ldots, K$, with respect to the vector $\mathbf{b}_k$, subjected to the aforementioned constraints i) and ii).

However, all deflation methods share a same specific performance problem. In practice, the contrast function must be estimated by some empirical contrast function. Therefore, the sources cannot be exactly extracted and some statistical error are unavoidable. Due to the non correlation constraint, the error committed in the previously extracted sources will propagate to subsequent sources. Moreover, the latter constraint must be enforced empirically through sample correlation, resulting in further errors. When many sources have to be extracted, these errors are cumulated. Hence, the "quality extraction" of the last extracted sources is much worst than the quality of the first extracted ones. From this viewpoint, methods that extract all sources at the same time are preferable. Unfortunately, to the authors knowledge, there is no result of discriminacy property of any contrast in such a simultaneous approach. In this paper, it is proved that the discriminacy property is not a specificity of the sequential approach. The simultaneous extraction method based on the sum of the output range also possesses this property, as shown in the next section.

3 Discriminacy property of the simultaneous approach using minimum range

The minimum range approach for the simultaneous extraction of bounded sources has been first introduced by Pham in [6]. The following criterion C was introduced, which can be viewed as the simultaneous counterpart to $\log R(\mathbf{b}_k\mathbf{X})$:

$$C(\mathbf{B}) = \sum_{i=1}^{K} \log R(\mathbf{b}_i\mathbf{X}) - \log|\det \mathbf{B}|.$$

This criterion is to be minimized with respect to $\mathbf{B}$. Clearly, it prevents $\mathbf{B}$ to be singular, therefore we shall focus to $\mathbf{B} \in \mathcal{M}(K)$, the set of all nonsingular matrices of order K.

It has been shown in [5, 6] that $-C$ is a contrast in the sense of Comon [2]. The proof in [5] of this result relies on the fact that the range functional is strictly superadditive in the sense of Huber (see [5]), that is $R^2(X+Y) > R^2(X) + R^2(Y)$ for any pair of independent bounded random variables X and Y. In fact the range functional possesses a stronger property which implies the strict superadditivity:

$$R(X+Y) = R(X) + R(Y)$$

for any pair of independent bounded random variables X and Y (see [4, 6]). This equality plays a crucial role in proving the discriminacy property of the contrast $-C$. In particular, denoting by $W_{i1}, \ldots, W_{iK}$ the components of $\mathbf{b}_i\mathbf{A}$, one has

$$R(\mathbf{b}_i\mathbf{X}) = R(\mathbf{b}_i\mathbf{A}\mathbf{S}) = \sum_{j=1}^{K} |W_{ij}| R(S_j),$$

since $R(\alpha X) = |\alpha| R(X)$ for any real number α and any bounded random variable X. The above relation shows that it is of interest to rewrite C in terms of the matrix $\mathbf{W} = \mathbf{BA}$ which has general element W_{ij}, which yields

$$C(\mathbf{B}) = \sum_{i=1}^{K} \log\Big[\sum_{j=1}^{K} |W_{ij}| R(S_j)\Big] - \log|\det \mathbf{W}| + \log|\det \mathbf{A}|.$$

Clearly, minimizing $C(\mathbf{B})$ over $\mathcal{M}(K)$ is equivalent to minimizing

$$\tilde{C}(\mathbf{W}) = \sum_{i=1}^{K} \log\Big[\sum_{j=1}^{K} |W_{ij}| R(S_j)\Big] - \log|\det \mathbf{W}| \qquad (1)$$

also over $\mathcal{M}(K)$. The point $\mathbf{B}$ minimizing C is related to the point $\mathbf{W}$ minimizing $\tilde{C}$ by the relation $\mathbf{W} = \mathbf{BA}$. The contrast property of $-C$ means that $\tilde{C}$ attains its global minimum at and only at matrices $\mathbf{W}$ which are products of a permutation and a diagonal matrix. In the remaining part of this paper, it is shown that there is no other local minimum of this criterion.

The function $\tilde{C}$ is not everywhere differentiable on $\mathcal{M}(K)$, due to the absolute value in (1). To overcome this difficulty, we introduce the subsets $\mathcal{M}_I(K)$ of $\mathcal{M}(K)$, indexed by subsets I of $\{1,\dots,K\}^2$, defined by

$$\mathcal{M}_I(K) = \{\mathbf{W} \in \mathcal{M}(K) : W_{ij} \neq 0 \text{ if and only if } (i,j) \in I\}.$$

Due to the non singularity condition, a subset $\mathcal{M}_I(K)$ may be empty for particular I. Thus we shall restrict ourselves to the collection $\mathcal{I}$ of distinct subsets I of $\{1,\dots,K\}^2$ such that $\mathcal{M}_I(K)$ is not empty. Then the subsets $\mathcal{M}_I(K)$, $I \in \mathcal{I}$, form a partition of $\mathcal{M}(K)$, since they are clearly disjoint and their union equals $\mathcal{M}(K)$. Therefore any local minimum point of $\tilde{C}$ would belong to some $\mathcal{M}_I(K)$ with $I \in \mathcal{I}$ and is necessarily a local minimum point of the restriction of $\tilde{C}$ on $\mathcal{M}_I(K)$.

The key point is that the restriction of $\tilde{C}$ to $\mathcal{M}_I(K)$, $i \in \mathcal{I}$, is infinitely differentiable as a function of the nonzero elements of its matrix argument in $\mathcal{M}_I(K)$. Thus, one may look at the first and second derivatives of the restriction of $\tilde{C}$ to $\mathcal{M}_I(K)$ to identify its local minimum points.

Lemma 1 *For $I \in \mathcal{I}$, the restriction of $\tilde{C}$ to $\mathcal{M}_I(K)$ admits the first and second partial derivatives*

$$\frac{\partial \tilde{C}(\mathbf{W})}{\partial W_{ij}} = \frac{\operatorname{sign}(W_{ij}) R(S_j)}{\sum_{l=1}^K |W_{il}| R(S_l)} - W^{ij}, \quad (i,j) \in I \tag{2}$$

$$\frac{\partial^2 \tilde{C}(\mathbf{W})}{\partial W_{ij} \partial W_{kl}} = W^{kj} W^{il}, \qquad (i,j) \in I, (k,l) \in I,\ k \neq i,$$

$$\frac{\partial^2 \tilde{C}(\mathbf{W})}{\partial W_{ij} \partial W_{il}} = -\frac{\operatorname{sign}(W_{ij}) \operatorname{sign}(W_{il}) R(S_j) R(S_l)}{[\sum_{k=1}^K |W_{ik}| R(S_k)]^2} + W^{ij} W^{il}, \qquad (i,j) \in I, (i,l) \in I,$$

where sign$(x) = \pm 1$ according to $x > 0$ or $x < 0$ (and can be either +1 or -1 if $x = 0$) and W^{ij} denote the (j,i) element of $\mathbf{W}^{-1}$.

The above Lemma allows one to characterize the stationary points of the restriction of $\tilde{C}$ to $\mathcal{M}_I(K)$, by setting its derivative to zero, yielding

$$W^{ij} = \frac{\operatorname{sign}(W_{ij}) R(S_j)}{\sum_{l=1}^K |W_{il}| R(S_l)}, \quad (i,j) \in I.$$

Thus one get the following corollary.

Corollary 1 *Let $I \in \mathcal{I}$, then for any $\mathbf{W} \in \mathcal{M}_I(K)$ which is a stationary point of the restriction of $\tilde{C}$ on $\mathcal{M}_I(K)$:*

$$\{(i,j) \in \{1,\dots,K\}^2 : W^{ij} \neq 0\} \supseteq I$$

and

$$\frac{\partial^2 \tilde{C}(\mathbf{W})}{\partial W_{ij} \partial W_{il}} = 0, \qquad (i,j) \in I, (i,l) \in I.$$

The above corollary is the key point for proving the following Lemma.

Lemma 2 *Let $I \in \mathcal{I}$ such that there exists a pair of indices i, j in $\{1, \ldots, K\}$ for which the i-th and the j-th sections of I are not disjoint (the i-th section of I is the set $\{k \in \{1, \ldots, K\} : (i, k) \in I\}$). Then the restriction of $\tilde{C}$ in $\mathcal{M}_I(K)$ does not have a local minimum point.*

Lemma 2 allows one to eliminate subsets I in $\mathcal{I}$ for which the restriction of $\tilde{C}$ in $\mathcal{M}_I(K)$ does not have a local minimum point. It can be proved that the only subset left is the one such that its i-th section reduces to a single point, for all $i = 1, \ldots, K$. This yields the discriminacy property of $\tilde{C}$.

Proposition 1 *The only local minimum points of C are the matrices $\mathbf{W}$ which are the product of a (nonsingular) diagonal and a permutation matrix. (They are also the global minimum points.)*

4 Conclusion

This work has focussed on the non existence of local maxima of BSS contrasts. Existing results have established that the square kurtosis and minus the output ranges both possess this property, when they are used in a deflation approach. This is a rather restrictive constraint though, since one often prefers a simultaneous separation in order to avoid errors cumulation. In this paper, it is shown that the minimum range criterion, even used in a *simultaneous* approach to the blind separation of bounded sources also possesses this "discriminacy" property. To the authors knowledge, this contrast is the only one for which this property has been proved up to now in the simultaneous approach.

5 Appendix: proofs of lemmas

Proof of Lemma 1

To compute the partial derivatives of $\tilde{C}$ given by (1), we note that

$$d|W_{ij}|/dW_{ij} = \operatorname{sign}(W_{ij}) \quad \text{if } W_{ij} \neq 0.$$

and that from [7]

$$\frac{\partial \log\det \mathbf{W}}{\partial W_{ij}} = \left[\frac{\partial \log\det \mathbf{W}}{\partial \mathbf{W}}\right]_{ij} = \left[\mathbf{W}^{-1\mathrm{T}}\right]_{ij} = W^{ij}.$$

Let us compute the partial derivative of W^{ij} with respect to W_{kl}. We note that $W^{ij} = \operatorname{tr}(\mathbf{E}_{ij}\mathbf{W}^{-1})$ where $\mathbf{E}_{ij}$ is the matrix with only one nonzero element at the (i, j) place which equals 1 and tr denotes the trace, hence [7]

$$\frac{\partial \operatorname{tr}(\mathbf{E}_{ij}\mathbf{W}^{-1})}{\partial W_{kl}} = -\left[(\mathbf{W}^{-1}\mathbf{E}_{ij}\mathbf{W}^{-1})^{\mathrm{T}}\right]_{kl} = -\left[\mathbf{W}^{-1}\mathbf{E}_{ij}\mathbf{W}^{-1}\right]_{lk} = -W^{il}W^{kj}.$$

This yields the formula for the second partial derivatives of restriction of $\tilde{C}$ to $\mathcal{M}_I(K)$ as given by the Lemma.

Proof of Lemma 2

Let $\mathbf{W} \in \mathcal{M}_I(K)$ be a stationary point (if exists) of the restriction of $\tilde{C}$ to $\mathcal{M}_I(K)$, we shall show that it cannot realize a local minimum of this function. By assumption, there exists i, j, k in $\{1, \ldots, K\}$ and $i \neq j$, such that (i, k) and (j, k) are both in I. Therefore, by Corollary 1, W^{ik} and W^{jk} are non-zero, hence by Lemma 1: $\partial^2\tilde{C}/\partial W_{ik}\partial W_{jk} = W^{ik}W^{jk} \neq 0$. Also, by the same corollary, $\partial^2\tilde{C}/(\partial W_{ik})^2 = \partial^2\tilde{C}/(\partial W_{jk})^2 = 0$. Thus, let $\tilde{W}$ be a matrix differing (slightly) from $\mathbf{W}$ only at the indexes (i, k) and (j, k): $\tilde{W}_{ik} = W_{ik} + \epsilon$, $\tilde{W}_{jk} = W_{jk} + \eta$, then since the first partial derivatives of $\tilde{C}$ vanishes at $\mathbf{W}$, a second order Taylor expansion yields:

$$\tilde{C}(\tilde{\mathbf{W}}) = \tilde{C}(\mathbf{W}) + W^{ik}W^{jk}\epsilon\eta + O((|\epsilon| + |\eta|)^3)$$

as $\epsilon, \eta \to 0$. Therefore, $\tilde{C}(\tilde{\mathbf{W}}) < \tilde{C}(\mathbf{W})$ if ϵ and η both are small enough and their product have the same sign as $W^{ik}W^{jk}$. This shows that $\mathbf{W}$ cannot realize a local minimum of $\tilde{C}$ in $\mathcal{M}_I(K)$.

Proof of Proposition 1

By Lemma 2, in order that the restriction of $\tilde{C}$ to $\mathcal{M}_I(K)$ admits a local minimum point, it is necessary that the sections $I_1, \ldots, I_K$ of I are all disjoint. On the other hand, none of these sections can be all empty since otherwise $\mathcal{M}_I(K)$ would be empty. Therefore these sections must be reduced to a single point: $I_i = \{(i, j_i)\}$, $i = 1, \ldots, K$ where $j_1, \ldots, j_K$ are indexes in $\{1, \ldots, K\}$. These indexes must be distinct since otherwise $\mathcal{M}_I(K)$ would be empty. But a matrix in $\mathcal{M}_I(K)$ where $I = \{(1, j_1), \ldots, (i, j_K)\}$ with $j_1, \ldots, j_K$ being a permutation of $1, \ldots, K$, is simply a product of a diagonal and a permutation matrix. Such matrix is already known to realize the global minimum of $\tilde{C}$. This completes the proof.

References

[1] A. Hyvärinen and J. Karhunen and E. Oja. *Independent Component Analysis*, John Willey and Sons, New York, 2001.

[2] P. Comon. Independent Component Analysis, a new concept? *Signal Processing*, 36(3):287-314, 1994.

[3] N. Delfosse and P. Loubaton. Adaptive blind separation of sources: A deflation approach, *Signal Processing*, 45:59-83, 1995.

[4] F. Vrins, C. Jutten and M. Verleysen. SWM : A class of convex contrasts for source separation. In Proc. ICASSP 2005, Philadelphia (USA), pp. V.161-V.164, 2005.

[5] D. T. Pham. Contrast functions for blind separation and deconvolution of sources. In T.-W. Lee, T.-P. Jung, S. Makeig and T. Sejnowski editor, proc. of ICA 2001, San-Diego, December 9-12, 2001.

[6] D. T. Pham. Blind Separation of Instantenaous Mixtrures of Sources Based on Order Statistics. *IEEE Trans. Signal Processing*, 48(2):363-375, 2000.

[7] K. Petersen and M. Pedersen. The Matrix Cookbook, October 3, 2005.

[8] D. T. Pham and F. Vrins. Local Minima of Information-Theoretic Contrasts in Blind Source Separation. *IEEE Signal Process. Lett.* 12(11), pp. 788-791, 2005.