Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 70 (2007) 1207–1214
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Minimum range approach to blind partial simultaneous separation of bounded sources: Contrast and discriminacy properties ☆

Frédéric Vrins[a,*], Dinh-Tuan Pham[b]

[a] Université catholique de Louvain, UCL Machine Learning Group, Louvain-la-Neuve, Belgium
[b] Centre National de la Recherche Scientifique (CNRS), Laboratoire de Modélisation et Calcul, Grenoble, France

Available online 12 December 2006

Abstract

The blind source separation (BSS) problem is often solved by maximizing objective functions reflecting the statistical independence between outputs. Since global maximization may be difficult without exhaustive search, criteria for which all the local maxima correspond to an acceptable solution of the BSS problem are of interest. It is known that some BSS criteria used in a *deflation* procedure benefit from this property. More recently, the present authors have shown that the "spurious maximum free" property still holds for the minimum range approach to BSS in a *simultaneous* separation scheme. This paper extends the last result by showing that source demixing and *local* maximization of a range-based criterion are equivalent, even in a *partial* separation scheme, i.e. when $P \leqslant K$ sources are simultaneously extracted from K linearly independent mixtures of them.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Blind partial source separation; ICA; Range; Local maxima; Discriminant contrast function

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 [9]: $\mathbf{X} = \mathbf{AS}$, where $\mathbf{X} = [X_1, \ldots, X_K]^{\mathrm{T}}$ and $\mathbf{S} = [S_1, \ldots, S_K]^{\mathrm{T}}$ are respectively, the observed mixtures and the source vectors, both of dimension K, and $\mathbf{A}$ is the non-singular mixing matrix of order K.

In the blind context, no specific knowledge on the sources 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 a *contrast function*, which is a criterion being 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 [2]. The unmixing matrix can be recovered row by row (deflation), or globally, all rows at once (simultaneous extraction).

For instance, it is known [13] that if Q is a class II superadditive functional in the sense of Huber, i.e. if for any pair of random variables X, Y and two scalar number α, β:

$$\begin{cases} Q(\alpha X + \beta) = |\alpha| Q(X), \\ Q^2(X + Y) \geqslant Q^2(X) + Q^2(Y), \end{cases}$$

then any criterion of the form

$$f^{\square}(\mathbf{B}) \doteq \log|\det \mathbf{B}| - \sum_{i=1}^{K} \log Q(\mathbf{b}_i \mathbf{X}), \tag{1}$$

where $\mathbf{b}_i$ being the ith row of $\mathbf{B}$, is a contrast function for simultaneous separation.

Then, globally maximizing a contrast function such as $f^{\square}$ yields the recovering of the original sources. In practice, however, when no algebraic method is available for the contrast maximization, iterative approaches are needed, such as gradient-ascent techniques. Therefore, there is no guarantee, without exhaustive search, to reach a *global*

☆This work is an extension of the conference paper [16]. The authors have contributed equally to these articles.

*Corresponding author.
E-mail addresses: frederic.vrins@uclouvain.be (F. Vrins), dinh-tuan.pham@imag.fr (D.-T. Pham).

0925-2312/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2006.11.011

maximum. The problem is that the contrast property does not ensure that local maximization is equivalent to source separation (see [15,20] and references therein with $Q(\cdot)$ is chosen to be the entropy power function).

In order that the above useful equivalence holds, the contrast is needed to be *discriminant*: in this work, a contrast is said to be discriminant if its local maxima are attained only if **BA** is the product of a permutation and diagonal matrix. The discriminacy property has been proved for specific deflation contrasts; they are recalled in Section 2. This property has also recently been proved by the present authors for the minimum-range contrast used in a simultaneous separation framework [16].

This paper extends the last result to the blind partial simultaneous separation, which consists in simultaneously recovering $P \leqslant K$ sources from K linearly independent mixtures of K independent sources (including thus the simultaneous separation of all the K sources) [14]. After having proved that the criterion is a contrast that admits a local maximum point once the P outputs are proportional to P distinct sources (Section 3), it is proved that there does not exist any other local maximum point (Section 4). In other words, the criterion reaches a local maximum point *if and only if* the outputs are proportional to distinct sources.

2. Existing discriminant contrasts

The possible existence of *mixing 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 [4], even though at that time, the existence of such maxima was not yet established. In order to avoid this possible problem, the authors of [4] proposed to extract the sources sequentially, one by one, using a deflation approach. The maximum square kurtosis of the output $\kappa^2(Y_i)$, combined to additional constraints ensuring scale-invariance and decorrelation between outputs, has been suggested as a *deflation* objective function for BSS; for each $1 \leqslant k \leqslant K$, a kth source is extracted by maximizing $\kappa^2(\mathbf{b}_k\mathbf{X})$, ($\mathbf{b}_k$ being the kth row of $\mathbf{B}$) subjected to the aforementioned constraints. It is proved in [4] that each of the local maxima of this function is attained when the kth 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. In a same order of ideas, under the whitening constraint, a sinusoidal contrast function can be found when only two sources are considered [10]; the BSS problem reduces thus to a simple phase estimation problem. Similarly, the limit points of geometric ICA are shown to be the solutions of the BSS problem, again in the 2D case and for symmetric and unimodal densities [17].

Recently, Vrins et al. [19] proved that the minimum range criterion also possesses this interesting "discriminacy" property when separating bounded sources. We define the range $R(X)$ of a bounded random variable X as the difference between the upper bound and lower bound of the support $\Omega(X)$ of X:

$$R(X) \doteq \sup \Omega(X) - \inf \Omega(X). \tag{2}$$

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 same constraints as for the above kurtosis-based method. In addition to the aforementioned references [12,19] dealing with range-based BSS criteria, connections with information theory can be found in [3,18]. An extension to complex-valued signals is proposed by Erdogan in [5].

This work proposes a theoretical study of the true range. In practice however, the range has to be estimated, and the related criterion has to be maximized. In the framework of ICA, order statistics-based estimators of the range are most often considered because of their simplicity and computational efficiency (the *sort* operation has a complexity of $\mathcal{O}(N \log N)$ where N is the number of sample points). They should yield the set of theoretical solutions provided that the sample set is large enough if their optimization is managed efficiently. Regarding this last point, range-based BSS algorithms involving order statistics in the range estimation are evoked in [12] (simultaneous separation) and [21,22] (deflation); they are not usual gradient-ascent techniques as the gradient of the criterion does not exist at the desired solution points. Even though the algorithm in [12] has not been yet rigorously proved to converge almost surely to an exact solution, it works well in practice. Hence, in the following, we restrict our analysis to the theoretical behavior of the true range criterion in the framework of the blind partial simultaneous separation, as explained below.

3. Blind partial simultaneous extraction of bounded sources via output range minimization

In this section, a criterion for simultaneously extracting $P \leqslant K$ (called *partial simultaneous separation*) sources from K linearly independent mixtures of the K independent sources is presented. This criterion shall be proved to have a local maximum point when **BA** is the product of a permutation and gain matrices. For clarity, the proofs are relegated in the Appendix.

Let us note $\mathbf{A} \in \mathbb{R}^{K \times K}$ (with $\det \mathbf{A} \neq 0$), $\mathbf{B} \in \mathbb{R}^{P \times K}$, $\mathbf{W} = \mathbf{BA} \in \mathbb{R}^{P \times K}$ the mixing, demixing and transfer matrices, respectively. It is known from Pham [14] that if $Q(\cdot)$ is class II superadditive [8], any functional $f(\mathbf{B})$ of the form

$$f(\mathbf{B}) \doteq \frac{1}{2} \log \det(\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^{\mathrm{T}}) - \sum_{i=1}^{P} \log Q(\mathbf{b}_i\mathbf{X}), \tag{3}$$

where $\boldsymbol{\Sigma}_{\mathbf{X}}$ is the covariance matrix of $\mathbf{X} = \mathbf{AS}$, is a contrast function [2]; it reaches a global maximum point only if $\mathbf{W}$ is *non-mixing*. A $P \times K$ matrix $\mathbf{W}$ is said non-mixing (and noted $\mathbf{W} \in \mathscr{W}^{P \times K}$) if it has a single non-zero element per row and at most one per column. More concretely, we have the following theorem [14].

Theorem 1 (*Partial contrast property of f*). *If* $0<Q(S_i)<\infty$ *and Q is a class* II *superadditive functional, then* $f(\mathbf{B})$ *admits a global maximum point* $\mathbf{B}$ *over the set* $\mathscr{M}^{P\times K}$ *of full-row rank* $P\times K$ *matrices, only if* $\mathbf{BA}\in\mathscr{W}^{P\times K}$.

The above property is called *partial contrast* property because the global maximum of f may not be attained for any matrix $\mathbf{B}$ such that $\mathbf{BA}\in\mathscr{W}^{P\times K}$, but only for specific matrices in this set. It is easy to characterize this set: if we assume without loss of generality that the sources are zero-mean, unit-variance (i.e. $\mathbf{\Sigma_S}=\mathbf{I}$) and ordered according to $Q(\cdot)$ as

$$Q(S_1)\leqslant Q(S_2)\leqslant\cdots\leqslant Q(S_K),$$

for the sake of simplicity, one gets the following corollary [14].

Corollary 1 (*Characterization of global maximum point*). *Let us define* $P^m\doteq\min\{i\in\{1,\ldots,P\}:Q(S_i)=Q(S_P)\}-1$, *and* $P^M\doteq\max\{i\in\{P,\ldots,K\}:Q(S_i)=Q(S_P)\}$. *The global maximum points of f over the set* $\mathscr{M}^{P\times K}$ *are the matrices* $\mathbf{B}$ *such that* $\mathbf{BA}\in\mathscr{W}_P^{P\times K}$, *where* $\mathscr{W}_P^{P\times K}$ *is the set of matrices with at most one non-zero element per column and with* P^m *rows having a single non-zero element of column index in* $\{1,\ldots,P^m\}$ *and the remaining rows having a single non-zero element of column index in* $\{P^m+1,\ldots,P^M\}$.

An interesting example of functional $Q(\cdot)$ is the statistical range $R(\cdot)$: the strict superadditivity property of $R(\cdot)$ results directly from

$$R(X+Y)=R(X)+R(Y), \tag{4}$$

for any pair of independent bounded random variables X and Y. In [12], the contrast property of $f^{\square}(\mathbf{B})$ with $Q(\cdot)=R(\cdot)$ (noted $C^{\square}(\mathbf{B})$) was provided. From Theorem 1 and Corollary 1, the last result generalizes as follows:

Corollary 2. *If the sources are bounded, i.e. if* $0<R(S_i)<\infty$, $1\leqslant i\leqslant K$ *then the criterion*

$$C(\mathbf{B})\doteq\frac{1}{2}\log\det(\mathbf{B\Sigma_X B}^{\mathrm{T}})-\sum_{i=1}^{P}\log R(\mathbf{b}_i\mathbf{X}), \tag{5}$$

admits a global maximum point $\mathbf{B}$ *if and only if* $\mathbf{BA}\in\mathscr{W}_P^{P\times K}$, *where* $\mathscr{W}_P^{P\times K}$ *is defined in Corollary* 1.

Clearly maximizing $C^{\square}$ is equivalent to maximizing C with respect to $\mathbf{B}$ if $P=K$ since $\frac{1}{2}\log\det(\mathbf{B\Sigma_X B}^{\mathrm{T}})=\log|\det\mathbf{B}|+\mathrm{cst}$.

The criterion $C(\mathbf{B})$ can be rewritten as a function of the transfer matrix elements W_{ij}. Denoting by $W_{i1},\ldots,W_{iK}$ the components of $\mathbf{b}_i\mathbf{A}$, one has

$$R(\mathbf{b}_i\mathbf{X})=R(\mathbf{b}_i\mathbf{AS})=\sum_{j=1}^{K}|W_{ij}|R(S_j), \tag{6}$$

since $R(\alpha X)=|\alpha|R(X)$ for any real number α and any bounded random variable X. The above relation, combined with $\mathbf{\Sigma_X}=\mathbf{A\Sigma_S A}^{\mathrm{T}}=\mathbf{AA}^{\mathrm{T}}$, yields

$$C(\mathbf{B})=\frac{1}{2}\log\det(\mathbf{WW}^{\mathrm{T}})-\sum_{i=1}^{P}\log\left[\sum_{j=1}^{K}|W_{ij}|R(S_j)\right]. \tag{7}$$

Remark 1. Observe that the first term in the above criterion prevents a matrix $\mathbf{W}$ to have rank less than P unless it has a strictly null row. Indeed, since the rank of a product of two matrices cannot exceed that of each matrix factor, $\mathbf{W}$ having rank less than P means that $\det(\mathbf{WW}^{\mathrm{T}})=0$, hence the above criterion takes the value $-\infty$, since $\sum_{j=1}^{K}|W_{ij}|R(S_j)>0$ for all $i=1,\ldots,P$ if there are no null rows. Consequently, recovering twice a same source is not possible. The case where a row of $\mathbf{W}$ vanished identically is special as it leads to an indeterminate expression $\infty-\infty$. This case should be excluded, since it yields a null output (which cannot be a source). Therefore, in the following, we shall restrict $\mathbf{B}\in\mathscr{M}^{P\times K}$ ($\mathbf{B}=\mathbf{WA}^{-1}$ has the same rank as $\mathbf{W}$).

Clearly, when defining $\tilde{C}(\mathbf{W})\doteq C(\mathbf{B})$ where $\mathbf{W}=\mathbf{BA}$, maximizing $C(\mathbf{B})$ over the set $\mathscr{M}^{P\times K}$ is equivalent to maximizing $\tilde{C}(\mathbf{W})$ also over $\mathscr{M}^{P\times K}$. In particular, if $\widetilde{\operatorname{argmax}}_{\mathbf{B}}f(\mathbf{B})$ denotes the set of points *locally* maximizing f:

$$\widetilde{\operatorname*{argmax}}_{\mathbf{B}\in\mathscr{M}^{P\times K}}C(\mathbf{B})=\widetilde{\operatorname*{argmax}}_{\mathbf{B}:\mathbf{BA}\in\mathscr{M}^{P\times K}}\tilde{C}(\mathbf{BA}). \tag{8}$$

Any point $\mathbf{B}$ locally maximizing C is related to a point $\mathbf{W}$ locally maximizing $\tilde{C}$ by the relation $\mathbf{W}=\mathbf{BA}$. We do not claim that, generally speaking, $f(\mathbf{B})$ is locally maximized once $\mathbf{BA}\in\mathscr{W}^{P\times K}$, even under the class II superadditivity assumption on functional Q. Nevertheless, this result holds for the specific $Q(\cdot)=R(\cdot)$ case, as indicated by the following theorem.

Theorem 2 (*Non-mixing matrices are local maximum point of C*). *The criterion* $C(\mathbf{B})$ *admits a local maximum at any point* $\mathbf{B}$ *for which* $\mathbf{BA}\in\mathscr{W}^{P\times K}$.

Consequently, $C(\mathbf{B})$ reaches a global maximum if and only if $\mathbf{BA}\in\mathscr{W}_P^{P\times K}$ (Corollary 2) and a local maximum point if $\mathbf{BA}\in\mathscr{W}^{P\times K}$ (Theorem 2).

The local maximum points $\mathbf{W}\in\mathscr{W}^{P\times K}$ are called *non-mixing* because they correspond to non-mixing transfer matrix from $\mathbf{S}$ to $\mathbf{Y}$ and thus to the recovering of P distinct sources. By contrast, the discriminacy property of C is not yet established: the non-existence of *mixing* local maxima (i.e. the local maximum points verifying $\mathbf{W}\in\mathscr{M}^{P\times K}\backslash\mathscr{W}^{P\times K}$) remains to be proved. Such a property, addressed in the next section, ensures the equivalency between local maximization of $C(\mathbf{B})$ and partial source separation.

4. Discriminacy property of the simultaneous approach using minimum range

In order to analyze the possible existence of mixing maxima of $C(\mathbf{B})$ (i.e. of $\tilde{C}(\mathbf{W})$, with $\mathbf{W}=\mathbf{BA}$), we shall first compute the first two derivatives of $\log|\det(\mathbf{W}\mathbf{W}^{\mathrm{T}})|$ with respect to the elements W_{ij} of $\mathbf{W}$; they are provided in the following Lemma (some of useful mathematical relations involved in the proof are taken from [1,6,7,11]).

Lemma 1. *Let $\mathbf{W} \in \mathscr{M}^{P\times K}$ and denote by $\mathbf{W}^{+} \doteq \mathbf{W}^{\mathrm{T}}(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}$ its pseudoinverse. Then*

$$\frac{\partial \log|\det(\mathbf{W}\mathbf{W}^{\mathrm{T}})|}{\partial W_{ij}} = 2[(\mathbf{W}^{+})^{\mathrm{T}}]_{ij} = 2[\mathbf{W}^{+}]_{ji}$$

and

$$\frac{\partial^2 \log|\det(\mathbf{W}\mathbf{W}^{\mathrm{T}})|}{\partial W_{kl}\partial W_{ij}} = 2\{[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ki}(\delta_{jl} - [\mathbf{W}^{+}\mathbf{W}]_{lj}) - [\mathbf{W}^{+}]_{li}[\mathbf{W}^{+}]_{jk}\},$$

where δ_{jl} is the Kronecker delta. Remind that if $P = K$, $\mathbf{W}^{+} = \mathbf{W}^{-1}$ and $[\mathbf{W}^{+}\mathbf{W}]_{lj} = \delta_{lj}$.

Let us now use the above results for computing the first and second order derivatives of $\tilde{C}(\mathbf{W})$. The main problem is that $\tilde{C}$ is not everywhere differentiable on $\mathscr{M}^{P\times K}$, due to the absolute value in (7). To overcome this difficulty, we introduce the subsets $\mathscr{M}_I^{P\times K}$ of $\mathscr{M}^{P\times K}$, indexed by subsets I of $\mathbb{Z}^{P\times K} \doteq \{1,\ldots,P\} \times \{1,\ldots,K\}$, defined by

$$\mathscr{M}_I^{P\times K} \doteq \{\mathbf{W} \in \mathscr{M}^{P\times K} : W_{ij} \neq 0 \quad \text{if and only if } (i,j) \in I\}. \tag{9}$$

Due to the $\mathbf{W} \in \mathscr{M}^{P\times K}$ restriction, a subset $\mathscr{M}_I^{P\times K}$ may be empty for particular I. For example, if I is a subset of $\mathbb{Z}^{P\times K}$ such that its ith section $I_i \doteq \{j \in \{1,\ldots,K\}, (i,j) \in I\}$ is empty for some $i \in \{1,\ldots,P\}$, then any matrix $\mathbf{W}$ such that $W_{ij} = 0$ if $(i,j) \notin I$ (including all matrices $\mathbf{W}$ s.t. $W_{ij} \neq 0$ if and only if $(i,j) \in I$) satisfy $\operatorname{rank}(\mathbf{W}) < P$. Then, $\mathscr{M}_I^{P\times K}$ is empty because $\mathscr{M}_I^{P\times K} \subset \mathscr{M}^{P\times K}$ by definition.

Thus, we shall restrict ourselves to the collection $\mathscr{I}$ of distinct subsets I of $\mathbb{Z}^{P\times K}$ such that $\mathscr{M}_I^{P\times K}$ is not empty. Then the subsets $\mathscr{M}_I^{P\times K}$, $I \in \mathscr{I}$, form a partition of $\mathscr{M}^{P\times K}$, since they are clearly disjoint and their union equals $\mathscr{M}^{P\times K}$; in other words, for any matrix $\mathbf{W} \in \mathbb{R}^{P\times K}$, it exists one and only one subset I of $\mathbb{Z}^{P\times K}$ such that $\mathbf{W} \in \mathscr{M}_I^{P\times K}$. Therefore, any local maximum point of $\tilde{C}$ would belong to some $\mathscr{M}_I^{P\times K}$ with $I \in \mathscr{I}$ and is necessarily a local maximum point of the restriction of $\tilde{C}$ on $\mathscr{M}_I^{P\times K}$.

The key point is that the restriction of $\tilde{C}$ to $\mathscr{M}_I^{P\times K}$, $I \in \mathscr{I}$, is infinitely differentiable as a function of the non-zero elements of its matrix argument in $\mathscr{M}_I^{P\times K}$. Thus, one may look at the first and second derivatives of the restriction of $\tilde{C}$ to $\mathscr{M}_I^{P\times K}$ to identify its local maximum points. Noting that $\partial|W_{ij}|/\partial W_{ij} = \operatorname{sign}(W_{ij})$, the following result comes from Lemma 1 and the definition of $\tilde{C}$:

Lemma 2. *For $I \in \mathscr{I}$, the restriction of $\tilde{C}$ to $\mathscr{M}_I^{P\times K}$ admits the first and second partial derivatives*

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

$$\frac{\partial^2 \tilde{C}(\mathbf{W})}{\partial W_{ij}\partial W_{kl}} = [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ki}(\delta_{jl} - [\mathbf{W}^{+}\mathbf{W}]_{lj}) - [\mathbf{W}^{+}]_{li}[\mathbf{W}^{+}]_{jk},$$
$$(i,j) \in I, \ (k,l) \in I, \ k \neq i,$$

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

where $\operatorname{sign}(x) = \pm 1$ according to $x > 0$ or $x < 0$ (and can be either $+1$ or -1 if $x = 0$).

The above lemma allows one to characterize the stationary points of the restriction of $\tilde{C}$ to $\mathscr{M}_I^{P\times K}$, by setting its derivative to zero, yielding

$$[\mathbf{W}^{+}]_{ji} = \frac{\operatorname{sign}(W_{ij})R(S_j)}{\sum_{l=1}^{K}|W_{il}|R(S_l)} \quad (i,j) \in I. \tag{10}$$

Thus, one gets the following corollary.

Corollary 3. *Let $I \in \mathscr{I}$, then for any $\mathbf{W} \in \mathscr{M}_I^{P\times K}$ which is a stationary point of the restriction of $\tilde{C}$ on $\mathscr{M}_I^{P\times K}$:*

$$\{(i,j) \in \mathbb{Z}^{P\times K} : [\mathbf{W}^{+}]_{ji} \neq 0\} \supseteq I,$$

and

$$\frac{\partial^2 \tilde{C}(\mathbf{W})}{\partial W_{ij}\partial W_{il}} = [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ii}(\delta_{jl} - [\mathbf{W}^{+}\mathbf{W}]_{lj}) \quad (i,j) \in I, \ (i,l) \in I.$$

The first statement of the corollary results directly from the fact that if $\mathbf{W} \in \mathscr{M}_I^{P\times K}$ then $W_{ij} \neq 0$ if $(i,j) \in I$ and both sides of Eq. (10) are non-zero, too. The second claim is a consequence of Eq. (10) and Lemma 3.

Lemma 3. *Let $I \in \mathscr{I}$ such that either* (i) *the set $\bigcup_{i=1}^{P} I_i$ contains more than P elements, or* (ii) *there exists a pair of indices i,j in $\{1,\ldots,P\}$ for which the $I_i \cap I_j \neq \emptyset$. Then the restriction of $\tilde{C}$ in $\mathscr{M}_I^{P\times K}$ does not have a local maximum point.*

Lemma 3 allows one to eliminate subsets I in $\mathscr{I}$ for which the restriction of C in $\mathscr{M}_I^{P\times K}$ does not have a local maximum point. It can be proved that the only subsets I of $\mathscr{I}$ left are the ones such that all their ith sections, $i = 1,\ldots,P$, are distinct and reduce to a single point. This yields the discriminacy property of C over $\mathscr{M}^{P\times K}$, which ensures the source recovering via the *local* maximization of $C(\mathbf{B})$.

Theorem 3 (*Discriminacy of C*). *The local maximum points of C correspond to $\mathbf{B}$ such that $\mathbf{BA} \in \mathscr{W}^{P\times K}$.*

One concludes that for the specific $P = K$ case, $C(\mathbf{B}) = C^{\square}(\mathbf{B})$ admits a global maximum point $\mathbf{B}$ if and only if $\mathbf{BA} \in \mathscr{W}^{K\times K}$ (note that if $P = K$, $\mathbf{W}^{+} = \mathbf{W}^{-1}$), and one gets the main result of [16]. According to Theorem 3, this result still holds for $P \leqslant K$: C admits a local maximum point $\mathbf{B}$ if and only if $\mathbf{BA} \in \mathscr{W}^{P\times K}$ and a global maximum point $\mathbf{B}$ if and only if $\mathbf{BA} \in \mathscr{W}_P^{P\times K}$.

5. Conclusion

In this paper, a range-based criterion for the simultaneous separation of $P \leqslant K$ bounded sources is presented, which is shown to be a partial contrast function: its global maxima are reached if and only if the transfer matrix between $\mathbf{S}$ and $\mathbf{Y}$ belongs to the specific class of non-mixing matrices yielding the extraction of the P sources with the smallest range. Further, it is proved that the contrast C is discriminant: its local maximization is equivalent to separate $P \leqslant K$ distinct sources. In other words, there is no mixing maxima.

This property was already established for other *deflation* criteria but, to the authors' knowledge, the minimum output range method is the only one, up to now, that benefits from this property in the (possibly partial) simultaneous extraction scheme.

Acknowledgment

Frédéric Vrins is indebted to Prof. Marco Saerens from the Université catholique de Louvain (Information Systems Lab.) for having provided the useful monographs [1,6,7].

Appendix

Proof of Lemma 1. It is obvious that $\mathbf{W}\mathbf{W}^{\mathrm{T}} \in \mathbb{R}^{P\times P}$ is symmetric. It is also invertible since it is full-rank: $\mathrm{rank}(\mathbf{W}\mathbf{W}^{\mathrm{T}}) = \mathrm{rank}(\mathbf{W}) = P$.

Computation of first derivative:

$$\frac{\partial \log|\det(\mathbf{W}\mathbf{W}^{\mathrm{T}})|}{\partial W_{ij}} = \frac{1}{|\det(\mathbf{W}\mathbf{W}^{\mathrm{T}})|}\frac{\partial \det|(\mathbf{W}\mathbf{W}^{\mathrm{T}})|}{\partial W_{ij}}. \tag{11}$$

But, noting the *trace* operator by $\mathrm{Tr}(\cdot)$:

$$\frac{\partial \det|(\mathbf{W}\mathbf{W}^{\mathrm{T}})|}{\partial W_{ij}} = \det(\mathbf{W}\mathbf{W}^{\mathrm{T}})\,\mathrm{Tr}\left((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial(\mathbf{W}\mathbf{W}^{\mathrm{T}})}{\partial W_{ij}}\right). \tag{12}$$

Further, note that

$$\frac{\partial(\mathbf{W}\mathbf{W}^{\mathrm{T}})}{\partial W_{ij}} = \mathbf{W}\frac{\partial \mathbf{W}^{\mathrm{T}}}{\partial W_{ij}} + \frac{\partial \mathbf{W}}{\partial W_{ij}}\mathbf{W}^{\mathrm{T}} \tag{13}$$

yielding

$$\frac{\partial(\mathbf{W}\mathbf{W}^{\mathrm{T}})}{\partial W_{ij}} = \mathbf{W}\mathbf{J}^{ji} + \mathbf{G}^{ij}\mathbf{W}^{\mathrm{T}}. \tag{14}$$

In the above equality, $\mathbf{J}^{ji} \in \mathbb{Z}^{K\times P}$ and $\mathbf{G}^{ij} \in \mathbb{Z}^{P\times K}$ are *single-entry* matrices: $[\mathbf{J}^{ji}]_{kl} = [\mathbf{G}^{ij}]_{lk} = \delta_{kj}\delta_{li}, (k,l) \in \mathbb{Z}^{K\times P}$ (only the (i,j)th element of both $\mathbf{J}^{ij}$ and $\mathbf{G}^{ij}$ matrices is non-zero, and is set to one).

Observe that for any matrices $\mathbf{U}, \mathbf{V}$ with ad hoc size:

$$\mathrm{Tr}(\mathbf{V}\mathbf{J}^{ij}) = [\mathbf{V}^{\mathrm{T}}]_{ij} \tag{15}$$

and

$$\mathrm{Tr}(\mathbf{V}\mathbf{G}^{ij}\mathbf{U}) = [\mathbf{U}\mathbf{V}]_{ji}. \tag{16}$$

Then, one gets

$$\begin{aligned}
&\mathrm{Tr}\left((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial(\mathbf{W}\mathbf{W}^{\mathrm{T}})}{\partial W_{ij}}\right)\\
&\quad= \mathrm{Tr}((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{W}\mathbf{J}^{ji}) + \mathrm{Tr}((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{G}^{ij}\mathbf{W}^{\mathrm{T}})\\
&\quad= \Bigg[\underbrace{((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{W})^{\mathrm{T}}}_{\doteq \mathbf{W}^{+}}\Bigg]_{ji} + \Bigg[\underbrace{\mathbf{W}^{\mathrm{T}}(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}}_{\doteq \mathbf{W}^{+}}\Bigg]_{ji} &(17)\\
&\quad= 2[\mathbf{W}^{+}]_{ji} = 2[(\mathbf{W}^{+})^{\mathrm{T}}]_{ij}. \qquad \square &(18)
\end{aligned}$$

Computation of second derivative:

$$\begin{aligned}
\frac{\partial^2 \log|\det(\mathbf{W}\mathbf{W}^{\mathrm{T}})|}{\partial W_{kl}\partial W_{ij}} &= \frac{\partial}{\partial W_{kl}}\mathrm{Tr}\left((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial(\mathbf{W}\mathbf{W}^{\mathrm{T}})}{\partial W_{ij}}\right)\\
&= \mathrm{Tr}\left(\frac{\partial}{\partial W_{kl}}\left[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial(\mathbf{W}\mathbf{W}^{\mathrm{T}})}{\partial W_{ij}}\right]\right).
\end{aligned}$$

But, from (14)

$$\begin{aligned}
&\mathrm{Tr}\left(\frac{\partial}{\partial W_{kl}}\left[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial(\mathbf{W}\mathbf{W}^{\mathrm{T}})}{\partial W_{ij}}\right]\right)\\
&\quad= \mathrm{Tr}\left(\frac{\partial[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}(\mathbf{W}\mathbf{J}^{ji} + \mathbf{G}^{ij}\mathbf{W}^{\mathrm{T}})]}{\partial W_{kl}}\right)\\
&\quad= \mathrm{Tr}\left(\frac{\partial[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{W}\mathbf{J}^{ji}]}{\partial W_{kl}}\right)\\
&\qquad+ \mathrm{Tr}\left(\frac{\partial\Bigg[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\underbrace{\mathbf{G}^{ij}\mathbf{W}^{\mathrm{T}}}_{(\mathbf{W}\mathbf{J}^{ji})^{\mathrm{T}}}\Bigg]}{\partial W_{kl}}\right)\\
&\quad= \mathrm{Tr}\left(\frac{\partial[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]}{\partial W_{kl}}\mathbf{W}\mathbf{J}^{ji}\right) + \mathrm{Tr}\left(\frac{\partial[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]}{\partial W_{kl}}(\mathbf{W}\mathbf{J}^{ji})^{\mathrm{T}}\right)\\
&\qquad+ \mathrm{Tr}\left((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial[\mathbf{W}\mathbf{J}^{ji}]}{\partial W_{kl}}\right)\\
&\qquad+ \mathrm{Tr}\left((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial[(\mathbf{W}\mathbf{J}^{ji})^{\mathrm{T}}]}{\partial W_{kl}}\right).
\end{aligned}$$

But, noting $\mathbf{H}^{ki}$ the single-entry matrix in $\mathbb{R}^{P\times P}$:

$$\begin{aligned}
\mathrm{Tr}\left((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial[\mathbf{W}\mathbf{J}^{ji}]}{\partial W_{kl}}\right) &= \delta_{lj}\mathrm{Tr}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{H}^{ki}] &(19)\\
&= \delta_{lj}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ki}, &(20)
\end{aligned}$$

and similarly, $\mathrm{Tr}((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\partial[(\mathbf{W}\mathbf{J}^{ji})^{\mathrm{T}}]/\partial W_{kl}) = \delta_{lj}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ik}$. Consequently,

$$\begin{aligned}
&\mathrm{Tr}\left((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial[(\mathbf{W}\mathbf{J}^{ji})^{\mathrm{T}}]}{\partial W_{kl}}\right) + \mathrm{Tr}\left((\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial[\mathbf{W}\mathbf{J}^{ji}]}{\partial W_{kl}}\right)\\
&\quad= 2\delta_{lj}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ik}.
\end{aligned}$$

But, noting that

$$\frac{\partial \mathbf{F}^{-1}}{\partial W_{ij}} = -\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial W_{ij}}\mathbf{F}^{-1} \tag{21}$$

we can see that

$$\mathrm{Tr}\left(\frac{\partial[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]}{\partial W_{kl}}\mathbf{W}\mathbf{J}^{ji}\right) + \mathrm{Tr}\left(\frac{\partial[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]}{\partial W_{kl}}(\mathbf{W}\mathbf{J}^{ji})^{\mathrm{T}}\right)$$

equals

$$-\left\{\mathrm{Tr}\left[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial(\mathbf{W}\mathbf{W}^{\mathrm{T}})}{\partial W_{kl}}(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{W}\mathbf{J}^{ji}\right] + \mathrm{Tr}\left[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\frac{\partial(\mathbf{W}\mathbf{W}^{\mathrm{T}})}{\partial W_{kl}}(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}(\mathbf{W}\mathbf{J}^{ji})^{\mathrm{T}}\right]\right\}$$

that is from Eq. (14):

$$-\mathrm{Tr}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}[\mathbf{W}\mathbf{J}^{lk} + \mathbf{G}^{kl}\mathbf{W}^{\mathrm{T}}](\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}[\mathbf{W}\mathbf{J}^{ji} + \mathbf{G}^{ij}\mathbf{W}^{\mathrm{T}}]].$$

The last expression is no other than the following sum of four traces:

$$\mathrm{Tr}\left[\underbrace{(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{W}}_{(\mathbf{W}^+)^{\mathrm{T}}}\mathbf{J}^{lk}\underbrace{(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{W}}_{(\mathbf{W}^+)^{\mathrm{T}}}\mathbf{J}^{ji}\right] \tag{22}$$

$$+\,\mathrm{Tr}\left[\underbrace{(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{W}}_{(\mathbf{W}^+)^{\mathrm{T}}}\mathbf{J}^{lk}(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{G}^{ij}\mathbf{W}^{\mathrm{T}}\right] \tag{23}$$

$$+\,\mathrm{Tr}\left[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{G}^{kl}\underbrace{\mathbf{W}^{\mathrm{T}}(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}}_{\mathbf{W}^+}\mathbf{W}\mathbf{J}^{ji}\right] \tag{24}$$

$$+\,\mathrm{Tr}\left[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{G}^{kl}\underbrace{\mathbf{W}^{\mathrm{T}}(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}}_{\mathbf{W}^+}\mathbf{G}^{ij}\mathbf{W}^{\mathrm{T}}\right]. \tag{25}$$

To deal with the above traces, observe that

$$\mathrm{Tr}[\mathbf{A}\mathbf{J}^{pq}\mathbf{B}\mathbf{J}^{rs}] = [\mathbf{A}]_{sp}[\mathbf{B}]_{qr} \tag{26}$$

and

$$\mathrm{Tr}[\mathbf{A}\mathbf{J}^{pq}\mathbf{B}\mathbf{J}^{rs}\mathbf{C}] = \sum_{m=1}[\mathbf{A}]_{mp}[\mathbf{B}]_{qr}[\mathbf{C}]_{sm}. \tag{27}$$

This yields the following equalities:

$$\mathrm{Tr}[(\mathbf{W}^+)^{\mathrm{T}}\mathbf{J}^{lk}(\mathbf{W}^+)^{\mathrm{T}}\mathbf{J}^{ji}] = [\mathbf{W}^+]_{li}[\mathbf{W}^+]_{jk},$$
$$\mathrm{Tr}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{G}^{kl}\mathbf{W}^+\mathbf{W}\mathbf{J}^{ji}] = [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ik}[\mathbf{W}^+\mathbf{W}]_{lj},$$

$$\mathrm{Tr}[(\mathbf{W}^+)^{\mathrm{T}}\mathbf{J}^{lk}(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{G}^{ij}\mathbf{W}^{\mathrm{T}}] = \sum_m[\mathbf{W}^+]_{lm}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ki}[\mathbf{W}]_{mj} = [\mathbf{W}^+\mathbf{W}]_{lj}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ki},$$

$$\mathrm{Tr}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{G}^{kl}\mathbf{W}^+\mathbf{G}^{ij}\mathbf{W}^{\mathrm{T}}] = \sum_m[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{mk}[\mathbf{W}^+]_{li}[\mathbf{W}]_{mj} = [\mathbf{W}^+]_{li}[\mathbf{W}^+]_{jk}.$$

Finally, since the inverse of a symmetric matrix is symmetric, $[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ki} = [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ik}$, and it comes that

$$\frac{\partial^2 \log|\det(\mathbf{W}\mathbf{W}^{\mathrm{T}})|}{\partial W_{kl}\partial W_{ij}} = 2\{[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ki}(\delta_{jl} - [\mathbf{W}^+\mathbf{W}]_{lj}) - [\mathbf{W}^+]_{li}[\mathbf{W}^+]_{jk}\}. \qquad \square$$

Proof of Theorem 2. The proof of this theorem results from an adaptation to $P \leqslant K$ of the proof of Proposition 3 presented in [12]. As in this proof, in order to show that any matrix in $\mathscr{W}^{P\times K}$ is a local maximum point of $\tilde{C}$ it is sufficient to prove that for a small increment $\delta\mathbf{W}$ of $\mathbf{W} \in \mathscr{W}^{P\times K}$, the quantity

$$\sum_{i=1}^{P}\left\{\log\left[\sum_{j=1}^{K}|W_{ij} + \delta W_{ij}|R(S_j)\right] - \log\left[\sum_{j=1}^{K}|W_{ij}|R(S_j)\right]\right\}, \tag{28}$$

where W_{ij} and δW_{ij} denoting the general element of $\mathbf{W}$ and of $\delta\mathbf{W}$, is larger or equal than $\frac{1}{2}\{\log\det[(\mathbf{W}+\delta\mathbf{W})(\mathbf{W}+\delta\mathbf{W})^{\mathrm{T}}] - \log\det(\mathbf{W}\mathbf{W}^{\mathrm{T}})\}$, up to first order in $\delta\mathbf{W}$. But since $\mathbf{W} \in \mathscr{W}^{P\times K}$, there exists distinct indexes $j(1),\ldots,j(P)$ such that for $i = 1,\ldots,P$, $W_{ij} \neq 0$ if and only if $j = j(i)$. Thus, (28) reduces to

$$\sum_{i=1}^{P}\log\left[\left|1 + \frac{\delta W_{ij(i)}}{W_{ij(i)}}\right| + \frac{\sum_{j\neq j(i)}^{K}|\delta W_{ij}|R(S_j)}{|W_{ij(i)}|R(S_{j(i)})}\right],$$

which, for $|\delta W_{ij(i)}| < |W_{ij(i)}|$, equals

$$\sum_{i=1}^{P}\left[\frac{\delta W_{ij(i)}}{W_{ij(i)}} + \frac{\sum_{j\neq j(i)}^{K}|\delta W_{ij}|R(S_j)}{|W_{ij(i)}|R(S_{j(i)})}\right] + O(\|\delta\mathbf{W}\|^2).$$

On the other hand, the first order Taylor expansion of a multivariate function $f : \mathbb{R}^{P\times K} \to \mathbb{R}$ is $f(\mathbf{W}+\delta\mathbf{W}) = f(\mathbf{W}) + \langle\nabla f, \delta\mathbf{W}\rangle + \cdots$ where ∇f is the gradient of f and $\langle\cdot,\cdot\rangle$ denotes the dot product. Then, setting the log function for f in the last expansion, we have from Lemma 1 that

$$\langle\nabla f, \delta\mathbf{W}\rangle = 2\sum_{i,j}[(\mathbf{W}^+)^{\mathrm{T}}]_{ij}\delta W_{ij} = 2\,\mathrm{Tr}[\mathbf{W}^+\delta\mathbf{W}].$$

Note that due to the special form of $\mathbf{W}$, $\mathbf{W}\mathbf{W}^{\mathrm{T}}$ is diagonal with ith diagonal equal to $W^2_{ij(i)}$. Thus, $\mathbf{W}^+ = \mathbf{W}^{\mathrm{T}}(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}$ has ji element equal to 0 if $j \neq j(i)$ and to $1/W_{ij(i)}$ otherwise. Therefore,

$$\frac{1}{2}\{\log\det[(\mathbf{W}+\delta\mathbf{W})(\mathbf{W}+\delta\mathbf{W})^{\mathrm{T}}] - \log\det(\mathbf{W}\mathbf{W}^{\mathrm{T}})\} = \sum_{i=1}^{P}\frac{\delta W_{ij(i)}}{W_{ij(i)}} + O(\|\delta\mathbf{W}\|^2).$$

It follows that (28) is greater that the above left-hand side, up to a term of order $O(\|\delta\mathbf{W}\|^2)$. $\square$

Proof of Lemma 3. Let $\mathbf{W} \in \mathscr{M}_I^{P\times K}$ be a stationary point (if exists) of the restriction of $\tilde{C}$ to $\mathscr{M}_I^{P\times K}$, we shall show that it cannot realize a local maximum of this function.

Consider the case where $\bigcup_{i=1}^{P} I_i$ contains more than P elements. Then there must exists an index $j \in \bigcup_{i=1}^{P} I_i$ for which $\mathbf{e}_j$, the jth row of the identity matrix of order K, is not contained in the linear subspace spanned by the rows of $\mathbf{W}$, since this subspace is of dimension P. By definition, there exists $i \in \{1, \dots, P\}$ such that $(i,j) \in I$. Let $\tilde{\mathbf{W}}$ be a matrix differing from $\mathbf{W}$ only in the element W_{ij} by ε. Then by the Taylor expansion up to second order, noting that the first partial derivative of $\tilde{C}$ vanished at $\mathbf{W}$ and using Corollary 3,

$$\tilde{C}(\tilde{\mathbf{W}}) = \tilde{C}(\mathbf{W}) + \tfrac{1}{2}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ii}(1 - [\mathbf{W}^{+}\mathbf{W}]_{jj})\varepsilon^2 + O(|\varepsilon|^3), \tag{29}$$

as $\varepsilon \to 0$. It can be checked that $\mathbf{W}^{+}\mathbf{W}$ is idempotent (i.e. $(\mathbf{W}^{+}\mathbf{W})^2 = \mathbf{W}^{+}\mathbf{W}$) and symmetric, and hence the same can be seen to be true for $\mathbf{I} - \mathbf{W}^{+}\mathbf{W}$. Thus, the jth diagonal element of $\mathbf{I} - \mathbf{W}^{+}\mathbf{W}$, which is $1 - [\mathbf{W}^{+}\mathbf{W}]_{jj}$, is the same as the square norm of its jth row. Therefore, $1 - [\mathbf{W}^{+}\mathbf{W}]_{jj} \geqslant 0$ with equality if and only if the jth row of $\mathbf{I} - \mathbf{W}^{+}\mathbf{W}$ vanishes, or equivalently $\mathbf{e}_j = \mathbf{e}_j\mathbf{W}^{+}\mathbf{W}$. But since $\mathbf{e}_j$ is not in the linear subspace spanned by the rows of $\mathbf{W}$, this cannot happen. On the other hand, $\mathbf{W}\mathbf{W}^{\mathrm{T}}$ is symmetric and positive definite, implying that so is its inverse, and thus it exists a full-row rank matrix $\mathbf{P}$ such that $\mathbf{P}\mathbf{P}^{\mathrm{T}} = (\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}$. Consequently, each (i,i)th element of $(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}$, which is the square norm of the ith row of $\mathbf{P}$, is strictly positive. Hence, the second term of the right-hand side of (29) is strictly positive, yielding $\tilde{C}(\tilde{\mathbf{W}}) > \tilde{C}(\mathbf{W})$ for all $\varepsilon \neq 0$ and small enough; $\mathbf{W}$ is not a local maximum of $\tilde{C}$ on $\mathscr{M}_I^{P\times K}$.

Consider now the case $I_i \cap I_j \neq \emptyset$ for some $i \neq j$ in $\{1, \dots, P\}$. Let $k \in I_i \cap I_j$. By Lemma 2: $\partial^2\tilde{C}/\partial W_{ik}\partial W_{jk} = [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ji}(1 - [\mathbf{W}^{+}\mathbf{W}]_{kk}) - [\mathbf{W}^{+}]_{ki}[\mathbf{W}^{+}]_{kj}$. Also, by Corollary 3, $\partial^2\tilde{C}/(\partial W_{ik})^2 = [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ii}(1 - [\mathbf{W}^{+}\mathbf{W}]_{kk})$ and $\partial^2\tilde{C}/(\partial W_{jk})^2 = [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{jj}(1 - [\mathbf{W}^{+}\mathbf{W}]_{kk})$. Thus, let $\tilde{\mathbf{W}}$ be a matrix differing (slightly) from $\mathbf{W}$ only at the indexes (i,k) and (j,k): $\tilde{W}_{ik} = W_{ik} + \varepsilon$, $\tilde{W}_{jk} = W_{jk} + \eta$, then since the first partial derivatives of $\tilde{C}$ vanishes at $\mathbf{W}$, a second order Taylor expansion yields:

$$\begin{aligned}\tilde{C}(\tilde{\mathbf{W}}) = \tilde{C}(\mathbf{W}) &+ \frac{1 - [\mathbf{W}^{+}\mathbf{W}]_{kk}}{2}\\ &\times[\varepsilon\ \eta]\begin{bmatrix}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ii} & [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ij}\\ [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ji} & [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{jj}\end{bmatrix}\begin{bmatrix}\varepsilon\\ \eta\end{bmatrix}\\ &- \varepsilon\eta[\mathbf{W}^{+}]_{ki}[\mathbf{W}^{+}]_{kj} + O((|\varepsilon| + |\eta|)^3)\end{aligned} \tag{30}$$

as $\varepsilon, \eta \to 0$.

We have shown that $1 - [\mathbf{W}^{+}\mathbf{W}]_{kk} \geqslant 0$. Further, from the positive definiteness of $(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}$, one gets

$$[\varepsilon\ \eta]\begin{bmatrix}[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ii} & [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ij}\\ [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{ji} & [(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}]_{jj}\end{bmatrix}\begin{bmatrix}\varepsilon\\ \eta\end{bmatrix} \geqslant 0.$$

Therefore, $\tilde{C}(\tilde{\mathbf{W}}) \geqslant \tilde{C}(\mathbf{W}) - \varepsilon\eta[\mathbf{W}^{+}]_{ki}[\mathbf{W}^{+}]_{kj} + O((|\varepsilon| + |\eta|)^3)$ implying that $\tilde{C}(\tilde{\mathbf{W}}) > \tilde{C}(\mathbf{W})$ for $\varepsilon\eta$ having the opposite sign as that of $[\mathbf{W}^{+}]_{ki}[\mathbf{W}^{+}]_{kj}$ and $|\varepsilon| + |\eta| > 0$ and small enough. This proves that $\mathbf{W}$ cannot realize a local maximum of $\tilde{C}$. $\square$

Proof of Theorem 3. By Lemma 3, in order that the restriction of $\tilde{C}$ to $\mathscr{M}_I^{P\times K}$ admits a local maximum point, it is necessary that the sections $I_1, \dots, I_P$ of I are all disjoint and their union have at most P elements. On the other hand, none of these sections can be empty since otherwise $\mathscr{M}_I^{P\times K}$ would be empty. Therefore, these sections must be reduced to a single point: $I_i = \{(i, j(i))\}$, $i = 1, \dots, P$ where $j(1), \dots, j(P)$ are distinct indexes in $\{1, \dots, K\}$. By definition $j(i)$ denotes the column index of the unique non-zero elements of the ith row of $\mathbf{W}$. Thus, $\mathbf{W}$ has a single non-zero element per row and at most one non-zero element per column, meaning that $\mathbf{W} \in \mathscr{W}^{P\times K}$. Hence, a necessary condition for $\mathbf{W}$ being a local maximum point of $C(\mathbf{B})$ is that $\mathbf{BA} \in \mathscr{W}^{P\times K}$. This concludes the proof since, from Theorem 2, it is also a sufficient condition. $\square$

References

[1] D.S. Bernstein, Matrix Mathematics, Princeton University Press, Princeton, NJ, 1954.

[2] P. Comon, Independent component analysis, a new concept?, Signal Process. 36 (3) (1994) 287–314.

[3] S. Cruces, I. Duran, The minimum support criterion for blind source extraction: a limiting case of the strengthened Young's inequality, in: Proceedings of the International Conference Independent Component Analysis and Blind Signal Separation (ICA), Granada, Spain, Lecture Notes in Computer Science, vol. 3195, 2004, pp. 57–64.

[4] N. Delfosse, P. Loubaton, Adaptive blind separation of sources: a deflation approach, Signal Process. 45 (1995) 59–83.

[5] A.T. Erdogan, A simple geometric blind source separation method for bounded magnitude sources, IEEE Trans. Signal Process. 54 (2006) 438–447.

[6] F.A. Graybill, Matrices with Applications in Statistics, Duxbury, 1983.

[7] D.A. Harville, Matrix Algebra from a Statistician's Perspective, Springer, Berlin, 1997.

[8] P.J. Huber, Projection pursuit, Ann. Stat. 13 (2) (1985) 435–475.

[9] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.

[10] J. Murillo-Fuentes, F.J. Gonzalez-Serrano, A sinusoidal contrast function for the blind separation of statistically independent sources, IEEE Trans. Signal Process. 52 (12) (2004) 3459–3463.

[11] K.B. Petersen, M.S. Pedersen, The matrix cookbook, Technical Report, 2005.

[12] D.-T. Pham, Blind separation of instantenaous mixtures of sources based on order statistics, IEEE Trans. Signal Process. 48 (2) (2000) 363–375.

[13] D.-T. Pham, Contrast functions for blind separation and deconvolution of sources, in: Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation (ICA), San Diego, USA, 2001, pp. 37–42.

[14] D.-T. Pham, Blind partial separation of instantaneous mixtures of sources, in: Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation (ICA), Charleston, SC, USA, March 2006, Lecture Notes in Computer Science, vol. 3889, Springer, Berlin, 2006, pp. 868–875.

[15] D.-T. Pham, F. Vrins, Local minima of information-theoretic criteria in blind source separation, IEEE Signal Process. Lett. 12 (11) (2005) 788–791.

[16] D.-T. Pham, F. Vrins, Discriminacy of the minimum range approach to blind separation of bounded sources, in: Proceedings of European

Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium, 2006, pp. 377–382.
[17] F.J. Theis, A. Jung, C.G. Puntonet, E.W. Lang, Linear geometric ICA: fundamentals and algorithms, Neural Comput. 15 (2003) 419–439.
[18] F. Vrins, C. Jutten, D. Erdogmus, M. Verleysen, Zero-entropy minimization for blind extraction of bounded sources (bebs), in: Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation (ICA), Charleston, SC, USA, Lecture Notes in Computer Science, vol. 3889, Springer, Berlin, 2006, pp. 747–754.
[19] F. Vrins, C. Jutten, M. Verleysen, SWM : a class of convex contrasts for source separation, in: Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Philadelphia, USA, 2005, pp. V.161–V.164.
[20] F. Vrins, M. Verleysen, On the entropy minimization of a linear mixture of variables for source separation, Signal Process. 85 (5) (2005) 1029–1044.
[21] F. Vrins, M. Verleysen, Minimum support ICA using order statistics. Part I: performance analysis, in: Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation (ICA), Charleston, SC, USA, Lecture Notes in Computer Science, vol. 3889, Springer, Berlin, 2006, pp. 262–269.
[22] F. Vrins, M. Verleysen, Minimum support ICA using order statistics. Part II: performance analysis, in: Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation (ICA), Charleston, SC, USA, Lecture Notes in Computer Science, vol. 3889, Springer, Berlin, 2006, pp. 270–277.

Frédéric Vrins was born in Uccle, Belgium, in 1979. He received the M.S. degree in electromechanical engineering (mechatronics) and the D.E.A. degree in Applied Sciences from the Université catholique de Louvain (Belgium) in 2002 and 2004, respectively. He is currently working towards the Ph.D. degree in the UCL Machine Learning Group. His research interests are blind source separation, independent component analysis, Shannon and Renyi entropies, mutual information and information theory in adaptive signal processing. He was a member of the program committee of ICA 2006.

Dinh-Tuan Pham was born in Hanoï, VietNam, on February 10, 1945. He is graduated from the Engineering School of Applied Mathematics and Computer Science (ENSIMAG) of the Polytechnic Institute of Grenoble in 1968. He received the Ph.D. degree in Statistics in 1975 from the University of Grenoble. He was a Postdoctoral Fellow at Berkeley (Department of Statistics) in 1977–1978 and a Visiting Professor at Indiana University (Department of Mathematics) at Bloomington in 1979–1980. He is currently Director of Research at the French Centre National de la Recherche Scientifique (C.N.R.S). His researches include time series analysis, signal modelling, blind source separation, nonlinear (particle) filtering and biomedical signal processing.