Blind spectral unmixing by local maximization of non-Gaussianity

C.F. Caiafa\textsuperscript{a,*}, E. Salerno\textsuperscript{b}, A.N. Proto\textsuperscript{a,c}, L. Fiumi\textsuperscript{d}

\textsuperscript{a}Laboratorio de Sistemas Complejos, Facultad de Ingenier\'ia - UBA, Av. Paseo Col\'on 850 4to. piso, Ala sur (1063) Cap. Fed., Argentina
\textsuperscript{b}Istituto di Scienza e Tecnologie dell'Informazione - CNR, Via Moruzzi, 1, I-56124 Pisa, Italy
\textsuperscript{c}Comisi\'on de Investigaciones Cient\'ificas PBA (CIC), Av. Paseo Col\'on 751 (1063), Buenos Aires, Argentina
\textsuperscript{d}Lab. Aereo Ricerche Ambientali, Istituto sull'Inquinamento Atmosferico - CNR, Via del Fosso del Cavaliere, 100 00133 Roma, Italy

Received 21 December 2006; received in revised form 29 June 2007; accepted 2 July 2007
Available online 21 July 2007

Abstract

We approach the estimation of material percentages per pixel (endmember fractional abundances) in hyperspectral remote-sensed images as a blind source separation problem. This task is commonly known as spectral unmixing. Classical techniques require the knowledge of the existing materials and their spectra, which is an unrealistic situation in most cases. In contrast to recently presented blind techniques based on independent component analysis, we implement here a dependent component analysis strategy, namely the MaxNG (maximum non-Gaussianity) algorithm, which is capable to separate even strongly dependent signals. We prove that, when the abundances verify a separability condition, they can be extracted by searching the local maxima of non-Gaussianity. We also provide enough theoretical as well as experimental facts that indicate that this condition holds true for endmember abundances. In addition, we discuss the implementation of MaxNG in a noisy scenario, we introduce a new technique for the removal of scale ambiguities of estimated sources, and a new fast algorithm for the calculation of a Parzen windows-based NG measure. We compare MaxNG to commonly used independent component analysis algorithms, such as FastICA and JADE. We analyze the efficiency of MaxNG in terms of the number of sensor channels, the number of available samples and other factors, by testing it on synthetically generated as well as real data. Finally, we present some examples of application of our technique to real images captured by the MIVIS airborne imaging spectrometer. Our results show that MaxNG is a good tool for spectral unmixing in a blind scenario.

\textcopyright 2007 Elsevier B.V. All rights reserved.

Keywords: Blind source separation (BSS); Dependent component analysis (DCA); Hyperspectral remote sensed images; Independent component analysis (ICA); Spectral unmixing

0. Introduction

Modern airborne and satellite-borne optical sensors are able to provide a huge amount of data with high resolution in the spatial domain as well as in the spectral domain. Multispectral images, i.e., images with less than 10 spectral channels, have
allowed whole pixels to be classified [1], but the analysis of the individual substances that constitute each pixel is limited by the small number of available channels [2]. On the other hand, in hyperspectral images, where hundreds of channels are normally available, it is possible to obtain an estimation of the material (endmember) percentages per pixel. This task is called spectral unmixing and is a new and fascinating field of research.

Under certain conditions (see [2,3] and all the references therein for details), the spectral radiance upon the sensor location can be assumed to be well approximated by a linear mixture of the endmember radiances weighted by the corresponding fractional abundances, plus additive sensor noise. This is called the linear mixing model and, basically, it is valid when multiple scattering among distinct endmembers is negligible and the pixel surface is partitioned according to the fractional abundances [2].

Many spectral unmixing algorithms have been developed by exploiting this linear mixing model. The classical approaches require some knowledge of the existing materials and their spectra. If the endmember spectra are known, then the estimation of the abundances is an inverse problem that has been solved, for example, by least squares methods [2] and other approaches [4]. In any case, endmember determination is not an easy task, being usually achieved through an educated trial-and-error approach, and the analyst should have some knowledge of the field. For this reason, during the last years, many efforts were made to derive blind spectral unmixing techniques that may not require users to have any knowledge about the endmembers.

This kind of problem (blind source separation, or BSS) was deeply studied during the last 10 years for the special case where the sources are mutually independent (independent component analysis, or ICA [5,6]). Some authors applied ICA to the spectral unmixing problem and obtained promising results [7–10]. At the same time, however, ICA algorithms did not prove to be well suited for this problem, and many issues still need to be solved. The main reason why ICA is not a good solution is that, in the spectral unmixing framework, the endmember abundances (also referred to as sources in the BSS-ICA context) are not independent; moreover, they are constrained to add up to one [3]. Additionally, any BSS algorithm obtains the source estimates up to a permutation and scale indeterminacy, which represents a problem for the interpretation of the results.

The separation of dependent sources, or dependent component analysis (DCA) [11,12], was attempted by few researchers in the literature. In [13], a new algorithm called MAXNG was introduced for the separation of dependent sources, based on a non-Gaussianity measure and using the $L^2$-distance between probability distributions calculated via the Parzen windows technique.

In this paper, we propose to approach the spectral unmixing problem through the MAXNG algorithm and solve the problem of scale indeterminacy. This paper is organized as follows: in Section 1, the linear mixing model for hyperspectral images and a probabilistic model for the sources are presented; in Section 2, spectral unmixing is presented as a DCA problem and a theoretical explanation is provided to justify MAXNG as a tool for endmember abundance extraction; in Section 3, a new technique for the removal of scale ambiguities is presented; in Section 4, the details for the implementation of MAXNG are provided including a novel fast algorithm for the calculation of the Parzen windows-based non-Gaussianity measure; in Section 5, experimental results are presented with synthetic and real data sets demonstrating the efficiency of MAXNG and including a comparison with FastICA and JADE; finally, in Section 6, our main conclusions are outlined.

1. Hyperspectral imagery: a probabilistic model

1.1. Sensor signal model (mixtures)

At any fixed pixel, a simplified linear mixing model for $M$ sensors (channels) and $P$ endmembers (usually $P \ll M$ in hyperspectral images) can be written as follows [3]:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n},$$

where $\mathbf{x} = [x_0 \ x_1 \ldots x_{M-1}]^\top$ is an $M \times 1$ vector which contains the random variables assigned to the sensor measurements at different center frequencies, $\mathbf{A}$ is the mixing matrix ($M \times P$) containing the spectral signatures of the endmembers existing in the covered area, $\mathbf{s} = [s_0 \ s_1 \ldots s_{P-1}]^\top$ is an $P \times 1$ vector containing the endmember fractional abundances $\mathbf{a}$, generally written as $\mathbf{a} = [a_0 \ a_1 \ldots a_{M-1}]^\top$, where $a_i$ represents the fractional contribution of the $i$-th endmember.

\textsuperscript{1}The phrase fractional abundance, widely used in spectral unmixing papers (see [3]), denotes the percent contribution of an endmember to a pixel.
abundances and \( n = [n_0 \ n_1 \ldots n_{M-1}]^T \) is an \( M \times 1 \) vector containing the system noise, which is assumed to be zero-mean Gaussian. Note that some effects that are often important, such as the ones derived from the topography or from the spectral variability, are not included in this simple model.

The task of extracting the endmember fractional abundances from the available measurements is known as spectral unmixing. We can think about equation (1) as a linear, instantaneous data model for a source separation problem where the abundances, \( s_0, s_1, \ldots, s_{P-1} \), are the sources and the measurements, \( x_0, x_1, \ldots, x_{M-1} \), are the mixtures.

### 1.2. Endmember fractional abundances model (sources)

Now, we analyze the properties of sources when they are considered as random variables. In this model, all the sources in a pixel are greater or equal to zero \( (s_i \geq 0 \text{ for every } i = 0, 1, \ldots, P-1) \) and constrained to sum up to one, since they can be interpreted as the percentage contribution of each endmember to the pixel:

\[
\sum_{i=0}^{P-1} s_i = 1. \tag{2}
\]

It is easy to see that the constraint (2) imposes dependence among random variables; in fact, every source can be expressed as a linear combination of the others, and this makes them dependent. Furthermore, in a real scenario, the materials can be strongly related to each other, thus imposing additional dependences.

### 1.3. Notation and basic definitions

- We denote by \( \bar{s}_i \) and \( s_i \) the mean values of the source \( s_i \) and the mixture \( x_i \), i.e., \( \bar{s}_i = E[s_i] \) and \( s_i = E[x_i] \). We also apply this notation to vectors, for example: \( \bar{s} = E[s] \) and \( \bar{x} = E[x] \).

- We define the centered source \( s'_i \) and the centered mixture \( x'_i \) by subtracting from them their mean values: \( s'_i = s_i - \bar{s}_i \) and \( x'_i = x_i - \bar{x}_i \). Their vector notation forms are \( s' = s - \bar{s} \) and \( x' = x - \bar{x} \).

- For convenience, we also define the normalized sources: \( u_i = \frac{s'_i}{\bar{s}_i} = \frac{s_i - \bar{s}_i}{\bar{s}_i} \), where \( \sigma_i = \sqrt{E[(s_i - \bar{s}_i)^2]} \) is the standard deviation of source \( s_i \). Note that \( E[u_i] = 0 \) and \( E[u_i^2] = 1 \).

It is also easy to see that condition (2) can be expressed in terms of the mean values and the centered sources as follows:

\[
\sum_{i=0}^{P-1} \bar{s}_i = 1 \quad \text{and} \quad \sum_{i=0}^{P-1} s'_i = 0. \tag{3}
\]

### 2. Spectral unmixing as a dependent component analysis problem

If \( A \) is known, an estimate of the sources can be obtained by using the Moore–Penrose pseudoinverse \( A^+ = (A^T A)^{-1} A^T \) [14], i.e., \( \hat{s} = A^+ \bar{x} \). Note that, in this case, we obtain the sources plus a noise component: \( \hat{s} = s + v \), where \( v = A^+ \bar{n} \) remains Gaussian. Usually, we do not have access to the endmember spectra and we need to proceed with the separation in a blind fashion, i.e., without knowing matrix \( A \), which makes the task more complicated.

Many authors have approached this problem by using ICA [3,8–10,15,16] and some promising results were obtained for the case of having many endmembers. However, the assumption of independence among endmember abundances is always violated in real situations, and this makes ICA methods not well suited to spectral unmixing, especially when the number of sources is small. These methods enforce minimum mutual information between the reconstructed sources. We observed elsewhere [13] that the minimum mutual information and the maximum NG criteria are only equivalent for independent sources, and conjectured that enforcing maximum NG without constraining the outputs to be orthogonal could be a good strategy to separate dependent sources. This fact motivated us to investigate a DCA technique such as MAXNG in order to improve the results.

As usual, we will work with centered signals \( x' \) to make the separation easier. Our objective is to find a separation matrix \( D \) that provides an estimation \( \hat{s}' \) of the centered sources based on the available centered mixtures: \( \hat{s}' = D x' \). Once \( \hat{s}' \) is obtained, the corresponding source estimates are given by

\[
\hat{s} = D x' + D \bar{s} \tag{4}
\]
It is also important to note that, as usual in BSS algorithms, the sources are obtained up to a permutation and scaling transformation. The permutation indeterminacy is not critical for our purpose, but the scaling factors are essential for a good interpretation of the results. In Section 3, we show that it is possible to solve the scaling indeterminacy by using the source constraint (3).

2.1. Separation of dependent sources by local maximization of non-Gaussianity

Let us now introduce the theoretical reason that makes MaxNG a powerful DCA technique. In order to make the mathematical analysis simpler, we will present it for the case of having zero-noise (\( n = 0 \)) (in Section 4, two solutions for the noisy case are also presented which are tested through experiments in Section 5). In this case, the centered mixtures are obtained as follows:

\[
x' = As
\]

and we are interested in analyzing how to obtain a separation matrix \( D \) such that \( \hat{u} = Dx' \) is a normalized and possibly permuted version of the centered sources \( s' \).

MaxNG proposes to build the rows of matrix \( D \) so as to make the source estimates as non-Gaussian as possible, according to the NG measure defined below.

**Definition 1.** Given a zero-mean and unit-variance random variable \( z \), we define its non-Gaussianity (NG) measure \( \Gamma_z \), as the squared \( L^2 \)-distance between its pdf \( f_z(z) \) and the Gaussian pdf with zero-mean and unit-variance:

\[
\Gamma_z = \int_{-\infty}^{+\infty} [f_z(z) - \Phi(z)]^2 dz \quad \text{where} \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)z^2}.
\]

2.2. Separability condition

In this Section, a sufficient condition to assure source separability is introduced. For analytic tractability we restrict the problem to the case \( P = 3 \). Note that we can write \( \hat{u} = DA s' \) which means that every normalized source estimate is a linear combination of the three centered sources \( s'_0, s'_1 \) and \( s'_2 \) as follows:

\[
\hat{u}_i = \alpha_{i,0}s'_0 + \alpha_{i,1}s'_1 + \alpha_{i,2}s'_2,
\]

where the parameters \( \alpha_{i,j} \) are the elements of matrix \( DA \), i.e., \( \alpha_{i,j} = [DA]_{i,j} \).

Now, let us look at a particular normalized source estimate \( \hat{u}_i \). By using (3) and the definition of the normalized sources \( \sigma_i u_i = s'_i \), we can express \( \hat{u}_i \) as a linear combination of only two normalized sources, for example, in terms of \( u_0 \) and \( u_1 \) we can write

\[
\hat{u}_i = \alpha u_0 + \beta u_1,
\]

where \( \alpha = (\alpha_{i,0} - \alpha_{i,2})\sigma_0 \) and \( \beta = (\alpha_{i,1} - \alpha_{i,2})\sigma_1 \) are scalars and, obviously, a perfect reconstruction is obtained by \( (\alpha, \beta) = (\pm 1, 0) \) for \( u_0 \) and \( (\alpha, \beta) = (0, \pm 1) \) for \( u_1 \). Now we observe that, if we force the variance of the source estimate to one, we have

\[
E[\hat{u}_i^2] = \alpha^2 + \beta^2 + 2\alpha\beta\rho = 1,
\]

where \( \rho = E[u_0u_1] \) is the cross-correlation coefficient. In other words, enforcing the source estimate variance to be equal to one is equivalent to restricting the parameters \( (\alpha, \beta) \) to lie on one of the ellipses shown in Fig. 1. Our objective is to find a condition that guarantees that the NG measure will have local maxima at points \( (\alpha, \beta) = \{(1, 0), (0, 1), (-1, 0), (0, -1)\} \) when the parameter domain is restricted to the ellipse. Accordingly, we introduce the following definition of separability.

**Definition 2.** Given two normalized sources \( u_0, u_1 \), we say that source \( u_1 \) is separable from \( u_0 \) if the NG measure \( \Gamma_z(\alpha, \beta) \) of the linear combination \( z = \alpha u_0 + \beta u_1 \) has a local maximum (or minimum) at \( (\alpha, \beta) = (0, 1) \) when the parameters are restricted to lie on the ellipse \( \alpha^2 + \beta^2 + 2\alpha\beta\rho = 1 \).

The following theorem provides a condition for which the separability of sources is guaranteed.

**Theorem 3.** Given two dependent sources \( s_0 \) and \( s_1 \), if the conditional expectation \( E[s_0|s_1] \) is linear in \( s_1 \), that

![Fig. 1. Set of allowed values for parameters \( \alpha \) and \( \beta \) for \( \rho = 0.0, 0.5 \) and 0.9.](image)
is, \( E[s_0|s_1] = as_1 + b \) for some values of the parameters \( a \) and \( b \), then the following properties hold true:

(I) Parameters \( a \) and \( b \) can be calculated as \( a = \frac{s_0 - \bar{s}}{\sigma_1} \) and \( b = \bar{s} - \frac{s_0}{\sigma_1} \), where \( \rho \) is the correlation coefficient \( \rho = E[u_0u_1] \), \( \sigma_0 \) and \( \sigma_1 \) are the standard deviations, and \( \bar{s}_0, \bar{s}_1 \) are the mean values.

(II) The conditional expectation of the normalized sources is \( E[u_0|u_1] = \rho u_1 \).

(III) The normalized source \( u_1 \) is separable from \( u_0 \).

**Proof.** See Appendix A.

In our case, the main consequence of this theorem is that, if we are able to verify that the fractional abundances have linear conditional expectations, then all the sources are separable and can be detected by finding the local extrema of the NG measure. In Section 5, we show several examples where the separability condition is theoretically or experimentally verified.

3. Scale disambiguation of unmixed sources

Like most BSS algorithms, MAXNG provides an estimate of the sources up to a scaling and permutation transformation \[6\]. The permutation indeterminacy is not important for our purposes, but scaling is critical for allowing us to make a correct interpretation of the results, taking into account that, in general, different scale factors can be observed for different sources. Suppose that we have found the proper separating matrix such that \( x' = Dx' = DAš' + v \), where \( v = Dn \) is a Gaussian noise vector. In this case, for each entry of vector \( x' \), we have

\[
  z'_i = h_iš'_i + v_i,
\]

where no permutation was assumed and \( h_0, h_1, \ldots, h_{P-1} \) are defined as the scale factors. In order to determine the proper scale factors, we exploit the property for which the fractional end-member abundances must sum up to 1 at each pixel (Eq. (2)). If we set \( q_i = 1/h_i \), then the constraint on the centered sources (3) becomes

\[
  \sum_{i=0}^{P-1} q_i(z'_i - v_i) = 0.
\]

Of course, this cannot be enforced directly, but we can try to minimize its mean squared error (MSE):

\[
  e^2 = E \left[ \sum_{i=0}^{P-1} q_i(z'_i - v_i)^2 \right]
  = \sum_{i=0}^{P-1} \sum_{j=0}^{P-1} q_i E[z'_i z'_j] - E[v_i v_j] q_j,
\]

where we exploited the independence of source \( s'_i \) and noise \( n_i \), and the fact that, as is easy to see, \( E[z'_i v_j] = E[z'_j v_i] = E[v_i v_j] \). The equation above can also be written in the following matrix form:

\[
  e^2 = q^T (R_{x'^2} - R_v) q = q^T R q,
\]

where \( q^T = [q_0 q_1 \ldots q_{P-1}] \), \( R_{x'^2} = E[x'x'^T] \) and \( R_v = E[vv^T] \). We need to minimize Eq. (13) under the constraint \( \|q\| \neq 0 \); from elementary matrix theory, it is clear that the error will be minimum when \( q = γe \) where \( e \) is the eigenvector of matrix \( R \) associated with the minimum eigenvalue \( λ_{\text{min}} \). It is interesting to note that the obtained MSE is determined by the value \( q^T R q = γ^2 e^T Re = γ^2 λ_{\text{min}} \).

In order to complete the scale factor vector determination, we need to calculate the global constant \( γ \) and, to this end, we use the constraint (3) on the source means in the following way. Rewriting the constraint on the mean values we have

\[
  \bar{z}_i = (D\bar{s})_i = h_i\bar{s}_i = \frac{\bar{s}_i}{γ e_i} \quad \text{then} \quad γ e_i (D\bar{s})_i = \bar{s}_i, \quad (14)
\]

and taking the sum over \( i = 0, 1, \ldots, P - 1 \) and using (3) we arrive at

\[
  q_i = γe_i \quad \text{with} \quad γ = \frac{1}{\sum_{i=0}^{P-1} e_i (D\bar{s})_i}.
\]

While Eq. (15) provides a complete solution to the scale factor indeterminacy, in practice we can use other considerations in order to validate the results.

In particular, we know that all the sources’ values lie in the range \([0, 1]\), then if we assume that there are pixels where some specific material is not present \( (s_k = 0 \text{ for some } k) \) or if we have additional information about some pixels where only one material is present \( (s_k = 1 \text{ for some } k) \), we can use these pixels as control points to validate the results obtained through Eq. (15).

\[1\] In other words, we choose \( q = γe \) with \( γ \neq 0 \), \( Re = λ_{\text{min}}e \), and \( \|e\| = 1 \).
4. MaxNG algorithm implementation

4.1. NG measure in a noisy scenario

The MaxNG algorithm, originally developed for a noiseless case, proposes to detect the local maxima of the NG measure of a variable defined by a linear combination of the mixtures by fixing the variance of the source estimates to one [13], more specifically, we need to find the local maxima of the random variable \( z \) defined by

\[
\mathbf{z} = \mathbf{d}^T \mathbf{x}' = \sum_{i=0}^{M-1} d_i x'_i \quad \text{with} \quad E[z^2(\mathbf{d})] = 1 \quad (16)
\]

with the vector of parameters defined as \( \mathbf{d}^T = [d_0 d_1 \ldots d_{M-1}] \). When applied to the noisy data (1), this relation gives

\[
\mathbf{z} = \mathbf{d}^T \mathbf{x} + \mathbf{d}^T \mathbf{n} = \mathbf{y} + \mathbf{v}, \quad (17)
\]

where \( \mathbf{y} \) is a linear combination of the real sources \( \mathbf{s}_i \), and \( \mathbf{v} \) is Gaussian since it is a linear combination of Gaussian variables. Note that variable \( z \) is a sum of two independent variables and, therefore, its pdf can be written as a convolution:

\[
f_z(z) = (f_y f_v)(z). \quad (18)
\]

There are two ways of approaching the separation by the MaxNG principle:

Maximizing the NG of the noisy signal pdf \( f_z(z) \): In this case, we assume that the effect of noise on the resulting pdf \( f_z(z) \) will not affect too much the positions of local maxima. For \( f_z(z) \) we adopt the same strategy adopted in noiseless MaxNG, namely, the non-parametric approximation based on Parzen windows [17,18] with Gaussian kernels by using a set of \( N \) samples \( z(0), z(1), \ldots, z(N-1) \):

\[
f_z(z) \approx \frac{1}{Nh} \sum_{i=0}^{N-1} \phi \left( \frac{z - z(i)}{h} \right), \quad (19)
\]

where \( h \) (aperture parameter) is the standard deviation of the Gaussian kernels used as the window functions. The theory of density estimation [18] establishes criteria to choose the aperture parameter. In Section 5, many experimental results are presented which confirm this hypothesis for a wide range of the signal-to-noise ratios (SNRs) (higher than -10 dB). This allows us to apply MaxNG directly on noisy data, as in the zero noise case, even when we do not have information about the noise covariances.

Maximizing the NG of the deconvolved pdf \( f_y(y) \): In the case that we already know the noise covariance matrix, we can use this information to measure the NG of variable \( y \). We need to estimate the pdf \( f_y(y) \) in terms of the pdf \( f_z(z) \). The idea is to modify the NG measure \( \Gamma_z \) of Eq. (6) in order to make it as insensitive to noise as possible. Taking the Fourier transform of (18) we obtain

\[
\tilde{f}_z(\omega) = \frac{\tilde{f}_y(\omega)}{f_z(\omega)}, \quad (20)
\]

from which the desired pdf \( f_y(y) \) can be obtained by inverse transformation. For \( f_y(y) \), it is easy to see that it is zero-mean Gaussian with variance \( \sigma_y^2 = \mathbf{d}^T \mathbf{R}_n \mathbf{d} \), where \( \mathbf{R}_n = E[\mathbf{n}\mathbf{n}^T] \) is the noise covariance matrix and then its pdf is

\[
f_y(y) = \frac{1}{\sigma_y} \Phi \left( \frac{y}{\sigma_y} \right). \quad (21)
\]

Now, using the Fourier transforms of (19) and (21), putting them into (18) and taking the inverse Fourier transform we obtain

\[
f_y(y) \approx \frac{1}{Nh^2} \sum_{i=0}^{N-1} \phi \left( \frac{z - z(i)}{h^2} \right) \quad (22)
\]

with

\[
h^* = \sqrt{h^2 - \sigma_y^2} = \sqrt{h^2 - \mathbf{d}^T \mathbf{R}_n \mathbf{d}}. \quad (23)
\]

Comparing Eqs. (22) and (19), we conclude that, in the case of having noise included in the model, we can still rely on a Parzen windows pdf estimation in evaluating the NG measure, provided that we use a modified aperture parameter \( h^* \).

It is important to note what is the role of the modified aperture parameter given by (23). Reducing \( h^2 \) by \( \sigma_y^2 \) means to try to recover the smoothing caused by the noise on the pdf \( f_z(z) \). If the noise is very strong, however, the estimation of \( f_y(y) \) (22) becomes too “spiky”, due to the excessively narrow windows used. In the limit, when noise becomes so strong that \( \sigma_y^2 > h^2 \), Eq. (23) becomes meaningless. Parameter \( h \) thus plays the role of a regularization parameter [25]. It establishes a compromise between an assumed smoothness of the source pdf and the oversmoothing caused by the presence of noise. One disadvantage of this method is that the effective aperture parameter \( h^* \) depends on vector \( \mathbf{d} \), and the derivatives of the resulting NG measure in terms of \( \mathbf{d} \) have more complicated forms than in the previous
method (maximizing NG of noisy signal pdf), where the aperture parameter is fixed.

In Section 5.2 we present experimental results showing the noise effect on the NG maxima and comparing the two approaches considering the noisy signal pdf or the deconvolved one.

4.2. Whitening of data

As proposed in [13] for the case of zero noise, and widely used in ICA methods, a first step to solving the data (useful to both obtaining unit-norm uncorrelated eigenvectors and eigenvalues of matrix $R_{xx}$), and this can be done by considering the whitening filter matrix, i.e., $\mathbf{d} = \mathbf{d} A^{-1/2} V^T$.

## 4.3. Fast calculation of Parzen windows-based NG measure and its derivatives

In [13], a technique based on Parzen windows was presented for the calculation of the NG measure $\Gamma_z$ and its gradient $\nabla_d \Gamma_z$ by using a set of $N$ samples $z(0), z(1), \ldots, z(N-1)$. These estimates have a very high complexity of calculation since they both involve double sums over the set of samples ($O(N^2)$).

It is well known that estimators based on Parzen windows can be implemented in an optimized way if samples are arranged on a regular grid since the involved sums become convolutions allowing them to be calculated through the fast Fourier transform (FFT) [18]. This idea was used in the past for estimations of densities [18] and also in the estimation of mutual information for ICA [19]. In this paper we present a new fast algorithm for the calculation of $\Gamma_z$ and $\nabla_d^T \Gamma_z$. Let us first study a simple case of a general Parzen windows-based estimator given by

$$S(t) = \sum_{i=0}^{N-1} \Psi(t - z(i)).$$

where $\Psi$ is a certain continuous function $\Psi(t)$ (a Gaussian window or a first derivative of a Gaussian, for example) which tends to zero at infinity ($|\Psi(t)| \to 0$ when $t \to \pm \infty$). Given a data set with $N$ samples: $z(0), z(1), \ldots, z(N-1)$, we divide the continuous range of the variable $z$ in $N_A$ bins. If the samples are in the segment $[a, b)$, then a regular grid is defined by taking each bin with a fixed width given by $\Delta = (b - a)/N_A$ and each bin is defined by $\delta(n) = a + n \Delta/N_A$, $a + (n + 1) \Delta/N_A$ for $n = 0, 1, \ldots, N_A - 1$. Then, the following approximation of a sum $S(t)$ can be used:

$$S(t) \approx \sum_{n=0}^{N_A-1} f(n) \Psi(t - \delta(n)).$$

where $f(n)$ is the number of samples in the $n$th bin, and $\delta(n) = a + (n + \frac{1}{2}) \Delta/N_A$ corresponds to the center position of the $n$th bin. In Fig. 2 this approximation is explained through a simple example. It is obvious that the above approximation is more accurate for a large number of cells but, in practice with $N_A \ll N$, we usually obtain good results without introducing a considerable error. Another good property of Eq. (27) is that, since it is a convolution, it can be calculated through the FFT algorithm, which

---

4 Note that, in the noiseless case, the actual dimensionality of the data space is $P = 1$, because the source constraint reduces by one the effective dimensionality.

5 The vector $\mathbf{d}$ in (16) is related with the vector $\mathbf{d}$ through the whitening filter matrix, i.e., $\mathbf{d} = \mathbf{d} A^{-1/2} V^T$.

6 Notice that function $f(n)$ is proportional to the histogram function, defined by $f(n)/N$. 

---

Finally, applying approximation (27), we obtain the following equations for the NG measure and its derivatives (see Appendix B for a proof):

$$\Gamma_z = \frac{1}{2\sqrt{\pi}} + \Gamma_1 + \Gamma_2,$$

(28)

$$\Gamma_1 \approx \frac{1}{N} \sum_{n=0}^{N-1} f(n)\Phi_1(n)$$ and

$$\Gamma_2 \approx \frac{1}{N} \sum_{n=0}^{N-1} f(n)(f * \Phi_2)(n),$$

with $\Phi_1(n) = \frac{-2}{\sqrt{h^2+1}} \Phi(\frac{3(n)}{2\sqrt{h^2+1}})$, $\Phi_2(n) = \frac{1}{N\sqrt{2h}} \Phi(\frac{nA}{\sqrt{2h}})$ and where $(f * \Phi_2)(n)$ denotes the convolution of functions $f$ and $\Phi_2$.

$$\frac{\partial \Gamma_z}{\partial d_k} = \frac{\partial \Gamma_1}{\partial d_k} + \frac{\partial \Gamma_2}{\partial d_k},$$

(29)

$$\frac{\partial \Gamma_1}{\partial d_k} \approx \frac{2}{N} \sum_{n=0}^{N-1} g_k(n)\Phi_3(n)$$ and

$$\frac{\partial \Gamma_2}{\partial d_k} \approx -\frac{1}{N} \sum_{n=0}^{N-1} g_k(n)(f \Phi_4)(n),$$

(30)

where $\Phi_3(n) = \frac{3(n)}{N(h^2+1)} \Phi(\frac{3(n)}{2\sqrt{h^2+1}})$, $\Phi_4(n) = \frac{nA}{N\sqrt{2h}} \Phi(\frac{nA}{\sqrt{2h}})$, and $g_k(n)$ considers the sum of all whitened mixture samples $\tilde{z}_k(i)$ such that their corresponding $z(i)$ are in the cluster $\delta(n)$:

$$g_k(n) = \frac{1}{\sqrt{h^2+1}} \sum_{i/z(i)\in\delta(n)} \tilde{z}_k(i).$$

(31)

Note that the complexity for the calculation of functions $f(n)$ and $g_k(n)$ is $\mathcal{O}(N)$ and the heavy part of the calculations are associated with the convolutions having a complexity of $\mathcal{O}(N_A \log N_A)$. In practice, very accurate calculations are obtained with $N_A = 512$ or $N_A = 1024$ which represents a huge amount of saved time if we take into consideration that a very large number of samples $N$ are usually available (for example, $N = 22500$ for a small image of $150 \times 150$ pixels). It is important to note that the original computational complexity for the calculation of the NG measure and its derivatives is $\mathcal{O}(N^2)$, which makes it unpractical in most cases if we do not consider the optimization method presented here.

### 4.4 Parzen windows deflation technique

In order to avoid finding the same local maxima repeatedly along the different iterations of the algorithm, a deflation technique needs to be implemented. The idea is to remove the non-Gaussian structure found at the position of a local maximum once it was detected. This technique was first proposed by Friedman [20] in the exploratory projection pursuit context and consisted in enforcing the data to become Gaussian at a particular projection direction. Suppose a local maximum is found along direction $d^*$. This means that $z^* = d^{*T}\tilde{x}$ is a non-Gaussian variable. It is easy to see that, if we consider the transformation:

$$w = G^{-1}(F_{z^*}(z^*))$$

(32)

with $G^{-1}$ being the inverse of the Gaussian cumulative distribution function (cdf) $G(x) = \int_{-\infty}^{x} \Phi(u)\,du$ and $F_{z^*}$ is the cdf of $z^*$, i.e., $F_{z^*}(x) = \int_{-\infty}^{x} f_{z^*}(u)\,du$, then, the transformed variable $w$ has a Gaussian pdf $f_w(w) = \Phi(w)$. Using this

---

Fig. 2. Example of approximation of a Parzen-based estimator. (a) Original sum, (b) approximated sum.
basic fact, Friedman proposed the following transformation:
\[ w = U^T \Theta(Ux), \]  
(33)
where \( U^T \) is an orthogonal matrix whose first row is given by the vector \( d^T \) and where \( \Theta \) is a non-linear operator defined by
\[ \Theta(u^T) = \Theta([u_0 \ u_1 \ldots u_{Q-1}]^T) \]
\[ = [G^{-1}(F_{u_0}(u_0)) \ u_1 \ldots u_{Q-1}]^T. \]  
(34)

The vector transformation \( \Theta(u^T) \) applies the transformation (32) to the first vector component \( u_0 \) and leaves the rest of the components unchanged.

Friedman demonstrated that the transformation defined by (33) produces a new data set \( w \) whose distribution is as close as possible to that of \( x \) under the constraint that its marginal distribution along \( d^T \) is Gaussian. We introduce here a technique for estimating the cdf of interest \( F_{x^*}(z^*) \) by using Parzen windows approach. Taking the integral on Eq. (19) we arrive at
\[ F_{x^*}(z^*) \approx \frac{1}{N} \sum_{i=0}^{N-1} G\left(\frac{z - z(i)}{h}\right), \]  
(35)
and then we provide a practical way of implementing the Friedman non-linear transformation for removing non-Gaussian structures at a desired projection direction.

5. Experimental results

In this section, we analyze through extensive experiments the efficiency of MAXNG as a tool for blind spectral unmixing of hyperspectral images. We present several experimental results based on three different data sets:

(a) Data set A: synthetically generated sources and uniformly distributed mixing matrix \( A \): We generate sources that verify constraint (2) by doing the following: first, we generate \( P \) independent random variables, let us say, \( w_0, w_1, \ldots, w_{P-1} \) using a common pdf \( f_w(w) \); after that, we calculate our sources as \( s_k = w_k/\sum_{i=0}^{P-1} w_i \), and this guarantees that the properties of the model are verified. The domain of the pdf \( f_w(w) \) must be restricted to a bounded interval in \([0, +\infty)\). Here, we considered uniformly distributed variables in \([0, 1]\). It is easy to see that all variables \( s_k \) have the same pdfs and all the conditional expectations and the correlation coefficients are the same. It can be shown that the separability condition holds; more specifically,
\[ \rho = E[u_iu_j] = -\frac{1}{P-1}, \]  
(36)
\[ E[u_i|u_j] = -\frac{1}{P-1} u_j = \rho u_j, \]
where \( u_i \) and \( u_j \) are the corresponding normalized sources as defined in Section 1.3.

In this case, the mixing matrix entries are randomly generated from a uniform distribution in a positive interval.

(b) Data set B: real hyperspectral image: A real, radiometrically corrected hyperspectral image of a Rome city area was provided by the Airborne Laboratory for Environmental Research at IIA-CNR in Rome, Italy (see Fig. 3). This 540 x 337 pixels original image comes from the MIVIS sweepbroom airborne imaging spectrometer and contains 102 spectral channels from four independent sensors in different bands: channels 1–20 (0.43–0.83 \( \mu \)m); channels 21–28 (1.15–1.55 \( \mu \)m); channels 29–92 (2.00–2.50 \( \mu \)m) and 93–102 (8.20–12.70 \( \mu \)m). A pixel classification containing

Fig. 3. Top: RGB channels of the original hyperspectral image (540 x 337 pixels). Bottom: ground truth image; nine classes are provided.
nine classes has been also provided (Fig. 3 (bottom)), which was obtained by the standard spectral angle mapper method [21] integrated by a series of color stereo images and direct field inspections. The extracted classes are bricks (red), grit (yellow), other surfaces (grey), infrastructures (brown), trees (dark green), bush (green), meadows (light green), water (blue) and unclassified (black). Of course, we cannot take this map as ground truth, since we are looking for subpixel details; however, it will help evaluating the results with this data set and, on the other hand, we will use this map to build the simulated data described below.

(c) Data set C: simulated endmember abundances and approximated real spectra: In order to have a more realistic set of sources for testing our method, we propose to generate endmember abundances and their spectra from the data set B as follows: we assign an endmember to each class in Fig. 3 (bottom), and, for the estimation of each endmember spectrum, we average the spectra of pixels belonging to each class. Now, by using the original classification image, we generate a simulated low resolution image by mixing the spectra of existing materials for each $8 \times 8$ pixel subarea. In this way, we were able to calculate the material abundances as the percentages of the classes contained in each subarea. In Fig. 4, we show the mean estimated spectra and their standard deviation for three classes: red, yellow and light green.

5.1. Experiment 1: verification of the separability condition

As it was proved, when the conditional expectations are linear, the sources are separable and the MAXNG algorithm is able to extract them correctly. In the data set A, the separability condition is theoretically verified by Eq. (36), while, for the data sets B and C, it can be verified experimentally. In Fig. 5, three selected conditional expectation estimates are shown for sources corresponding to data sets A, B and C. This experiment explicitly shows the dependence among sources in all the cases. Note that the estimated conditional expectations are very close to the theoretical linear form.

5.2. Experiment 2: the effect of noise on the NG local maxima

The idea of this subsection experiments is to study the robustness of the MAXNG algorithm against noise and to compare the two strategies as presented in Section 4.1. We have considered different SNRs for a set of $P = 3$ sources generated according to the model of the data set A with $N = 4096$. A total number of channels $M = 102$ was used and a randomly generated mixing matrix $A$ was considered. Mixtures $x_i$ were generated according to the linear model (1) and uncorrelated Gaussian noise was introduced in each channel maintaining the...
SNR fixed for all the channels. We have run simulations for SNR’s in the range $-15$ to $+50$ dB and the NG measure was calculated within the entire parameter space for two cases: (a) NG of the noisy signal pdf; and (b) NG of the deconvolved pdf. Note that different whitening filters are required in the two cases (see Section 4.2). In all the cases, we have considered the first two eigenvalues as the dominant ones. Note that we need to calculate the NG measure of variable $z(d) = ^T \tilde{x}$, where $\tilde{x}$ is the whitened random vector and $d$ is a unit-norm vector that can be written in terms of a single parameter $\theta$ as follows: $d = [\cos \theta \sin \theta]$. In order to cover the whole parameter space, we have calculated the NG measures for cases (a) and (b) with $\theta$ in the range of $0^\circ$–$180^\circ$ by using Eq. (19) for case (a) and Eq. (22) for case (b). Notice that some criterion needs to be used for the selection of the proper aperture parameter $h$. For example, we can use $h = 1.06 \times N^{-0.2}$, which is the optimum value for a Gaussian kernel and a Gaussian pdf, in the sense that it minimizes the mean integrated square error (MISE) [18].

In Fig. 6, the NG measures for cases (a) and (b) are compared at different SNR levels for an aperture parameter $h = 0.2$ (Fig. 6(top)), which is the value that provides an optimum MISE for $N = 4096$; and for $h = 0.65$ (Fig. 6 (bottom)). For SNR $= 0$ dB and $h = 0.2$, the measure based on the deconvolved pdf introduces irregularities (see Fig. 6(b.1)) because the effective aperture parameter $h^*$ becomes too small. For SNRs lower than $0$ dB, $h^*$ becomes meaningless (see Eq. (23)). With $h = 0.65$, we were able to calculate the NG measure of the deconvolved pdf for SNRs ranging from $-10$ to $50$ dB. In all the cases ((a.1), (a.2), (b.1) and (b.2)) we compared the obtained local maxima to their theoretically expected positions.\(^{10}\) It is remarkable that, with $h = 0.2$, the obtained local maxima of the NG for the noisy signal pdf were accurate estimations of the real ones for the SNR range from $-5$ to $50$ dB. Also a close estimate was obtained for a very low SNR ($-10$ dB) and, for the extreme case of SNR $= -15$ dB, the noise caused local maxima to move apart from the theoretical positions. The most important conclusion from this study is that, in the case where we do not have any information about the noise covariance, we can apply the MaxNG algorithm directly to the noisy data without introducing significant errors.

5.3. Experiment 3: comparison with ICA algorithms

In order to compare our MaxNG approach to some methods already used in blind spectral unmixing [3], we have considered many cases of separation for $P = 3$ sources generated by using data set A as explained in the previous subsection. Given an estimation of the normalized source $\hat{u}$, we

\(^{9}\)Values in the range of $180^\circ$–$360^\circ$ provide the same results as in the range of $0^\circ$–$180^\circ$ but inverted (multiplied by a factor $-1$).

\(^{10}\)Note that, as expected, there are exactly three local maxima of the NG measure, in the range $0^\circ < \theta < 180^\circ$, corresponding to the $P = 3$ original sources.
evaluate the signal-to-interference ratio (SIR) as it is usually adopted in the literature for unit-norm and zero-mean signals, that is, for the $i$th normalized source: 

$$SIR_i = 10 \log_{10} \left( \frac{\text{var}(u_i)}{\text{var}(\hat{u}_i)} \right)$$

In Fig. 7, a comparison among the results of applying MaxNG (maximizing the NG of the noisy signal pdf), FastICA [6,22] and JADE [6,23] algorithms is shown. The mean SIR values were calculated for the SNR in the range of -25 dB to 60 dB over a set of 60 source estimations for each SNR level. For the implementation of ICA algorithms we have used the ICALAB software package [24]. Usually, SIR levels below a threshold of 5 dB or 8 dB are considered as false source detections [19], then for SNR $< -5$ dB, the three methods failed. As a reference, we compare these results with the case of using a perfect separating matrix $D$ such that $DA = I$. In this case, the error of estimation is only caused by noise. It is easy to see that, the obtained SIR with a perfect separating matrix, for the case of having
\[ R_{\text{sn}} = \sigma_n^2 I, \] is given by \( \text{SIR}_i = -10 \log_{10}(E[(D^T A^{-1/2}V^T I)^2]) \), which can be written in terms of the SNR and the eigenvalues \( \lambda_i \) as follows: \( \text{SIR}_i = \text{SNR} - 10 \log_{10}(\sum_{i=0}^{Q-1} \frac{1}{\lambda_i}) \). Notice that the MaxNG performance is very close to the case of using a perfect separating matrix in the range of \(-25\) to \(25\) dB. For SNR levels higher than \(25\) dB the MaxNG SIR tends to a constant value of \(38\) dB which is a very high value.

We also compared MaxNG (maximizing the NG of the noisy signal pdf), FASTICA and JADE when applied to \( P = 9 \) simulated sources of the data set \( C \) for a case of zero noise and having a data size of \( N = 2814 \). In Table 1, the results are shown for each of the nine simulated sources. Note that the dimensionality of the data is \( Q = 8 \) and FASTICA and JADE provide eight source estimations.

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>SIR 0</th>
<th>SIR 1</th>
<th>SIR 2</th>
<th>SIR 3</th>
<th>SIR 4</th>
<th>SIR 5</th>
<th>SIR 6</th>
<th>SIR 7</th>
<th>SIR 8</th>
<th>Mean SIR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max NG</td>
<td>19.5</td>
<td>17.6</td>
<td>32.4</td>
<td>15.1</td>
<td>25.3</td>
<td>17.1</td>
<td>20.1</td>
<td>14.4</td>
<td>34.0</td>
<td>21.73</td>
</tr>
<tr>
<td>Fast ICA</td>
<td>27.9</td>
<td>10.2</td>
<td>10.2</td>
<td>7.1</td>
<td>7.5</td>
<td>7.8</td>
<td>5.2</td>
<td>9.4</td>
<td>10.66</td>
<td></td>
</tr>
<tr>
<td>JADE</td>
<td>10.2</td>
<td>18.4</td>
<td>7.4</td>
<td>10.2</td>
<td>8.9</td>
<td>9.5</td>
<td>13.4</td>
<td>3.0</td>
<td>10.15</td>
<td></td>
</tr>
</tbody>
</table>

5.4. **Experiment 4: MaxNG efficiency evaluation**

There are several factors that may affect the efficiency of MaxNG for the case of maximizing the NG of the noisy signal pdf. In this section, we study how the number of pixels (\( N \)), the number of sensor channels (\( M \)) and the Parzen aperture parameter (\( h \)) impact the resulting mean SIR. In Fig. 8, the obtained mean SIRs versus \( N, M \) and \( h \) are shown for different SNRs. As more samples are taken into account, i.e., by increasing \( N \), the obtained mean SIR is also increased (for a fixed SNR) but it is remarkable that for \( N > 2000 \), the improvement of the efficiency is less perceptible, especially for \( \text{SNR} = 0\) dB and \(10\) dB. Also, the effect of increasing the number of sensor channels \( M \) (for a fixed SNR), is to improve the efficiency in terms of the obtained mean SIR. Regarding the aperture parameter \( h \), it was verified that the optimum value for a Gaussian pdf, given by \( h = 1.06 \times N^{-0.2} = 0.2 \) for \( N = 4096 \) [18], is within the range where also a good solution in terms of the obtained mean SIR (0.1 \( h < 0.5 \)) is reached. In order to study the MaxNG efficiency with different types of sources and spectra, the following cases were considered, where combinations of different types of sources and spectra were taken into account: (I) Mixing matrix and sources from the data set C; (II) mixing matrix from the data set C and sources from the data set A; (III) mixing matrix from the data set A and sources from the data set C and (IV) mixing matrix and sources from the data set A.
In Fig. 9, the Mean SIRs obtained for these four cases are shown. It is demonstrated that, when we use the approximated real spectra (cases I and II) the efficiency is deteriorated. An important feature of real spectra is that they give a very little contribution to most channels (see, for example, the contribution of each material in the channels 40–102 in Fig. 4). As a result, the effective number of channels is lower than the number \( M \) of sensor outputs, and this reduces the efficiency of separation (see Fig. 8 (middle)). We think this is the main reason why our approximated spectra (data set C) caused a reduced separation efficiency. Additionally, and this is a known problem \[3\], the effect of averaging spectra tends to create similar spectra for different materials.

5.5. Experiment 5: real hyperspectral data

As final experiments, we have applied MaxNG to small patches of the real hyperspectral image shown in Fig. 3. Since our data set was not accompanied by any information on instrumental sensitivities, we have not been able to take noise into account, and proceeded by maximizing the NG of the noisy output pdfs. It is important to note that, the more pixels are considered, the better is the efficiency expected (see Fig. 8) but, in real images, there is spectral variability to take into consideration: it is not realistic to assume that the spectrum of a material is the same in all the pixels. The main reasons that make an endmember spectrum fluctuate around an average curve are related to the illumination-scattering geometry and to local topography \[2,3\]. For example, adjacent pixels made from the same material may be illuminated differently, and their spectra can be scaled differently. Other reasons, such as local differences in atmospheric absorption, can be partly corrected and, normally, do not affect much the data, when they are collected in routine operations. In other cases, these errors can be hard to correct. Whatever the reason, it is expected that spectral variability is more critical for data images representing large areas.

We can evaluate the accuracy of a separation through the mean squared difference between the sum of the reconstructed sources and the constant 1 which we denote as mean square error: 
\[
\text{MSE} = E[(\sum_{i=0}^{P-1} \hat{s}_i - 1)^2].
\]
If the separation is successful, then this residual should be small; on the other hand, a big residual suggests us a poor separation.

We have applied MaxNG to two subimages of sizes \( 50 \times 50 \) and \( 75 \times 75 \)-pixel, shown in Figs. 10(a) and in 11(a), respectively. The eigenvalue analysis gave an effective dimensionality \( Q = 5 \) in both cases. This was used to implement the whitening filter. The good results obtained indicate that noise is not a critical issue in these images, and this is probably due to the huge number of channels available.

In some cases, especially when the data set contains a small number of pixels, the NG measure can show a few local maxima that do not correspond to actual sources (we call them “false” local maxima). To avoid wrong results, we need some criterion to reject them. One way to do this is to consider every subset of local maxima and to evaluate the mean squared residual in each case. It is reasonable to choose that subset of estimated sources that yields the minimum residual, since they verify constraints (1) and (2) more accurately.

In Fig. 10 (“Casino del bel respiro” and hidden garden at Villa Doria-Pamphilj, Rome), the results of applying MaxNG to a piece of \( 50 \times 50 \) pixels are shown together with its ground truth as a reference. It is observed that the extracted sources 0 and 1 match with yellow and red classes, respectively, and the third extracted source contains the green part plus other minor classes as grey, black, brown and blue. It is verified that MaxNG was not able to separate, for example, the class colored in blue because there are only a few pixels containing that material and the statistics are not enough for its identification.

![Fig. 9. Mean SIR versus SNR for the cases (I)-(IV) with a data size of \( N = 2814 \) and \( P = 9 \) sources.](image-url)
In Fig. 11 (Palazzo Corsini alla Lungara and Rome botanic garden), the results for a piece of 75 x 75 pixels are presented. In this case, four classes were identified: source 0, source 1 and source 2 match very well the yellow, green and red classes; source 3 contains the black class plus the brown and grey classes together. Finally, we compare the MSEs obtained for images of different sizes. In Table 2, the minimum MSE values are shown for the examples shown here and for a case with a 150 x 150 pixel image. Note that, for the first and second cases, the obtained MSEs are small and can be interpreted as good source separation results. On the other hand, when an image of 150 x 150 is considered, a high MSE is obtained and this is probably caused by the high spectral variability.

6. Conclusion and discussion

This work provides a method for the solution to the problem of blind spectral unmixing, which we already treated in [26]. This is a particular instance of blind source separation, where the sources to be separated are the fractional abundances of the endmember materials remotely sensed by hyperspectral systems. Since, for several reasons, our sources are not independent of one another, using blind separation methods based on independent component analysis may not be adequate to our purposes. The MaxNG dependent component analysis algorithm was verified to be more appropriate than other currently used techniques, such as FASTICA and JADE, to extract dependent sources from linear mixtures thereof [13]. In this paper, we extend the data model by introducing additive Gaussian noise and provide some theoretical grounds for MaxNG. In particular, we establish a separability condition that proves to be satisfied in spectral unmixing. Moreover, we show that the source constraint that characterizes spectral unmixing enables us to overcome the typical scaling ambiguity that affects all the blind separation problems.

To validate the theory, we tested MaxNG against several data sets, coming in part from actual measurements and in part from simulations, and obtained both a confirmation of our theoretical treatment and promising results for practical applications. An important drawback of the present algorithm, however, is that a space-invariant mixing
matrix is assumed. This is a very rough approximation of reality and is relevant to all the real-world applications in remote sensing. At present, we are facing this difficulty by analyzing small terrain patches, but this does not solve all the problems, since there is a risk of losing statistical significance. In fact, as our results indicate, those materials that contribute to a small number of pixels are not extracted individually, and remain combined with other sources in the final results (see, for example, “brown”, “black” and “grey” classes in Figs. 10 and 11). We will analyze this aspect extensively in the future. Our activity will also include tests on real hyperspectral data sets equipped with ground-truth information.

Acknowledgments

C. Caiafa acknowledges financial support from Facultad de Ingeniería, Universidad de Buenos Aires, Argentina (Beca Peruilh). This research has been partially supported by the Italian Ministry of Foreign Affairs in the framework of the Italian–Argentinian earth-observation agreement SIASGE. Partial support by the European Commission, FP6-507752-MUSCLE, is also acknowledged.

Appendix A. Proof of Theorem 3

(I) Let us first calculate the conditional expectation of normalized sources $E[u_0|u_1]$ by using the linearity condition $E[u_0|s_1] = ax_1 + b$. By using basic properties of the expectation operator and the
Now we will relate these results to the second order terms of a single parameter $t$:
\[
\alpha(t) = t, \\
\beta(t) = -tp \pm \sqrt{t^2(\rho^2 - 1) + 1}. 
\] (42)

In order to simplify the notation we will consider the source variables $x = u_0$, $y = u_1$ and their mixture $z = ax + by$. We note that the pdf of the mixture $z$ can be written in an integral form in terms of the joint pdf $f_{xy}(x,y)$ and the mixing coefficients $\alpha$ and $\beta$ (for $\beta \neq 0$):
\[
f_z(z; \alpha, \beta) = \frac{1}{\beta} \int f_{xy}(x, \frac{z - ax}{\beta}) \, dx. 
\] (43)

Note that, in the case of independent variables, Eq. (43) is reduced to the convolution of marginal pdfs.

First, we will show that the derivative of the pdf is zero at $t = 0$, i.e., $f'_z(z; \alpha(0), \beta(0)) = 0$. Observe that, by the chain rule, we have
\[
f'_z(z; \alpha(t), \beta(t)) = \frac{\partial}{\partial \alpha} f_z(z; \alpha, \beta) \alpha'(t) + \frac{\partial}{\partial \beta} f_z(z; \alpha, \beta) \beta'(t). 
\] (44)

Now, assuming that we can calculate derivatives with respect to $\alpha$ and $\beta$ by introducing the differentiation inside the integral in (43), after some algebraic work, we obtain the following (see Appendix A.1):
\[
\frac{\partial}{\partial \alpha} f_z(z = y; \alpha = 0, \beta = 1) = -\frac{d}{dy} (f_y(y)E[x|y]), \\
\frac{\partial}{\partial \beta} f_z(z = y; \alpha = 0, \beta = 1) = -\frac{d}{dy} (vf_y(y)). 
\] (45) (46)

The derivatives of the parameters $\alpha$ and $\beta$ at $t = 0$ are easily obtained from Eq. (42):
\[
\alpha'(0) = 1, \\
\beta'(0) = -\rho. 
\] (47)

Now, substituting Eqs. (45)–(47) in (44), we arrive at the following expression:
\[
f'_z(z = y; \alpha(0), \beta(0)) = -\frac{d}{dy} (f_y(y)E[x|y]) + \frac{d}{dy} (yf_y(y))\rho, 
\] (48)
derivative of expression (6):
\[
I'_z(t) = \int_{-\infty}^{+\infty} 2[f_z(z; \alpha(t), \beta(t)) - \Phi(z)]f'_z(z; \alpha(t), \beta(t)) \, dz
\]
(49)
and when evaluating this at \( t = 0 \) by using the fact that \( f'_z(z; 0, 0) = 0 \), we obtain the desired result.

A.1. Proof of Eqs. (45) and (46)

Let us first calculate the derivative of Eq. (43) with respect to \( \alpha \):
\[
\frac{\partial}{\partial \alpha} f_z(x; \alpha, \beta) = -\frac{1}{\beta^2} \int \frac{\partial}{\partial \alpha} f_{xy}(x, \frac{z - \alpha x}{\beta}) \, dx,
\]
(50)
when we evaluate it at \( \alpha = 0, \beta = 1 \), we obtain
\[
\frac{\partial}{\partial \alpha} f_z(x; 0, 1) = -\int \frac{\partial}{\partial \alpha} f_{xy}(x, \alpha x) \, dx
\]
\[
= -\frac{1}{\beta^2} \int f_{xy}(x, \alpha x) \, dx
\]
\[
= -\frac{1}{\beta^2} (f_x(\alpha) E[x|y]).
\]
(51)
in a similar way, we calculate the derivative of Eq. (43) with respect to \( \beta \):
\[
\frac{\partial}{\partial \beta} f_z(x; \alpha, \beta) = -\frac{1}{\beta^2} \int \frac{\partial}{\partial \beta} f_{xy}(x, \frac{z - \alpha x}{\beta}) \, dx
\]
\[
= -\frac{d}{dy} \int f_{xy}(x, \alpha x) \, dx
\]
\[
= -\frac{d}{dy} (f_x(\alpha) E[x|y]).
\]
(52)
Now, when we evaluate the above expression at \( \alpha = 0, \beta = 1 \), we have
\[
\frac{\partial}{\partial \beta} f_z(x; 0, 1) = -\int \frac{\partial}{\partial \beta} f_{xy}(x, y) y \, dx
\]
\[
= -\frac{d}{dy} \int f_{xy}(x, y) y \, dx = -\frac{d}{dy} (f_y(y)).
\]
(53)

Appendix B. Approximations to the NG measure and its derivatives

In [13], the following Parzen windows-based equations were obtained for \( \Gamma_1 \) and \( \Gamma_2 \):
\[
\Gamma_1 = \frac{-2}{N\sqrt{h^2 + 1}} \sum_{i=0}^{N-1} \Phi \left( \frac{y(i)}{\sqrt{h^2 + 1}} \right),
\]
(54)
\[
\Gamma_2 = \frac{1}{N^2 \sqrt{2h}} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \Phi \left( \frac{y(j) - y(i)}{\sqrt{2h}} \right).
\]
(55)
Let us first work with Eq. (54). Letting \( \Psi = \Phi \), taking \( t = 0 \) and \( z(i) = \frac{y(i)}{\sqrt{h^2 + 1}} \) in Eq. (26), we are able to apply the approximation in (27) reaching to
\[
\Gamma_1 \approx \frac{1}{N} \sum_{n=0}^{N-1} f(n) \Phi_1(n)
\]
(56)
with \( \Phi_1(n) = \frac{-2}{\sqrt{h^2 + 1}} \Phi \left( \frac{y(i)}{\sqrt{h^2 + 1}} \right) \). For the approximation in Eq. (55), we first identify \( \Psi = \Phi, t = \frac{y(i)}{\sqrt{2h}} \) and \( z(i) = \frac{y(i)}{\sqrt{2h}} \), which provides the following approximation:
\[
\Gamma_2 \approx \frac{1}{N^2 \sqrt{2h}} \sum_{i=0}^{N-1} \sum_{n=0}^{N-1} f(n) \Phi \left( \frac{\sqrt{h^2 + 1}}{\sqrt{2h}} n - m \right)
\]
(57)
and, by applying again approximation (27), we finally obtain
\[
\Gamma_2 \approx \frac{1}{N} \sum_{n=0}^{N-1} f(n) \Phi_2(n)
\]
(58)
with \( \Phi_2(n) = \frac{1}{N \sqrt{2h}} \Phi \left( \frac{y(i)}{\sqrt{2h}} \right) \) and where \( (f * \Phi_2)(n) \) denotes the convolution of functions \( f \) and \( \Phi_2 \).

Following an equivalent procedure, we obtain the approximation of the derivatives as presented in Eq. (29).

References


