A MAXIMUM LIKELIHOOD APPROACH FOR 
INDEPENDENT VECTOR ANALYSIS OF GAUSSIAN DATA SETS

Javier Vía* Matthew Anderson† Xi-Lin Li† Tülay Adalı†

*Dept. of Communications Engineering, University of Cantabria. Spain
†Dept. of Computer Science and Electrical Engineering. University of Maryland, Baltimore County
e-mail: jvia@gtas.dicom.unican.es, matt.anderson@umbc.edu, lixilin@umbc.edu, adali@umbc.edu

ABSTRACT

This paper presents a novel algorithm for independent vector analysis (IVA) of Gaussian data sets. Following a maximum likelihood (ML) approach, we show that the cost function to be minimized by the proposed GML-IVA algorithm reduces to an estimate of the mutual information among the different sets of latent variables. The proposed method, which can be seen as a new generalization of canonical correlation analysis (CCA), is based on the sequential solution of different least squares problems obtained from the quadratic approximation of the non-convex IVA cost function. The convergence and performance of the proposed algorithm are illustrated by means of several simulation examples, including an application consisting in the joint blind source separation (J-BSS) of three color images.

Index Terms— Independent vector analysis (IVA), joint blind source separation (J-BSS), canonical correlation analysis (CCA), second-order statistics (SOS).

1. INTRODUCTION

In the last years, independent vector analysis (IVA) [1] has gained increasing attention due to its applications in joint blind source separation (J-BSS) problems, which, for example, arise in magnetic resonance imaging (fMRI) [2, 3] or blind deconvolution of speech signals [1]. The key difference between IVA and independent component analysis (ICA) [4] resides in the exploitation of the statistical dependencies among the available data sets. In other words, although the IVA model for one data set is identical to the ICA model, the different data sets are jointly processed by IVA, and therefore we can expect much more accurate results than those obtained by individually applying ICA techniques to each data set.

Although a general IVA algorithm could be based on all the statistical information provided by the observations, it has been recently proved that the IVA problem can be solved by exclusively exploiting the second-order statistics (SOS) of the data [2, 5], which allows us to obtain simple algorithms avoiding the lack of robustness associated to techniques based on estimates of the higher-order statistics (HOS). Thus, in this paper we focus on the fundamental case of Gaussian data, and present the maximum likelihood (ML) approach to the IVA problem. In particular, it is shown that the cost function to be minimized can be seen as an estimate of the mutual information among the sets of latent variables. In order to solve this non-convex problem, we propose an iterative algorithm based on the minimization of quadratic approximations of the cost function. In this way, the proposed GML-IVA algorithm amounts to the sequential solution of different least squares (LS) problems. Moreover, the proposed method can be seen as a new generalization of canonical correlation analysis (CCA) [6] for several data sets [2, 3, 7–9], and it also exhibits interesting connections with some approximate joint-diagonalization algorithms [10–12]. Finally, the convergence and performance of the GML-IVA algorithm are evaluated by means of simulations, and its practical utility is illustrated by an example consisting in the joint blind separation of three color images.

2. NOTATION AND DATA MODEL

In this paper we use bold-faced upper case letters to denote matrices, and bold-faced lower case letters for column vectors. Superscripts (·)T and (·)H denote transpose and Hermitian, respectively. The notation C ∈ Cn×m (or C ∈ Rn×m) means that C is an n × m matrix with complex (or real) entries. The trace, determinant, and Frobenius norm of a matrix C are denoted as Tr(C), det(C) and ∥C∥, respectively. I is the identity matrix of dimension n, and 0n×m is the n×m zero matrix. The Kronecker and Hadamard (element-wise) products are denoted by ⊗ and ⊙, respectively. The diagonal matrix with vector el along its diagonal is denoted by diag(ε). bldig(C) is the block-diagonal version (with K blocks) of matrix C, and offdiag(C) = C − bldig(C). Moreover, given a matrix C = [c1 · · · cK] we define the operator diagvec(C) as

\[
\text{diagvec}(C) = \begin{bmatrix} \text{diag}(c_1) \\ \vdots \\ \text{diag}(c_K) \end{bmatrix}.
\]

Finally, E is the expectation operator, and in general Ra,b is the cross-covariance matrix for random vectors a and b, i.e., Ra,b = Eab.

2.1. IVA Data Model

Let us consider K data sets \( x^{[k]} \) ∈ Cn×1 and the signal model

\[
x^{[k]} = A^{[k]} s^{[k]}, \quad k = 1, \ldots, K,
\]

where \( A^{[k]} \) ∈ Cn×N are the unknown non-singular mixing matrices and \( s^{[k]} \) ∈ Cn×1 are the unknown latent variables. Eq.

\[1\]We consider data sets of the same dimensionality for notational simplicity. The extension of the results to data sets with different dimensions is straightforward.
(1) can be written in compact form as \( \mathbf{x} = \mathbf{A}s \), with 
\[
\mathbf{x} = \begin{bmatrix} x_1^T, \ldots, x_K^T \end{bmatrix}^T, \quad s = \begin{bmatrix} s_1^T, \ldots, s_K^T \end{bmatrix}^T,
\]
and
\[
\mathbf{A} = \begin{bmatrix} \mathbf{A}^1 & 0_{N \times N} & \cdots & 0_{N \times N} \\
0_{N \times N} & \mathbf{A}^2 & \cdots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
0_{N \times N} & \cdots & 0_{N \times N} & \mathbf{A}^K \end{bmatrix}.
\]
Alternatively, introducing a permutation in the elements of \( \mathbf{x} \) and \( s \) we obtain
\[
\tilde{\mathbf{x}} = \tilde{\mathbf{A}} \tilde{s},
\]
where \( \tilde{\mathbf{x}}, \tilde{s} \in \mathbb{C}^{KN \times 1} \) are defined as
\[
\tilde{\mathbf{x}} = \mathbf{P} \mathbf{x} = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_N \end{bmatrix}, \quad \tilde{s} = \mathbf{P} s = \begin{bmatrix} \tilde{s}_1 \\ \vdots \\ \tilde{s}_N \end{bmatrix},
\]
and \( \mathbf{P} \in \mathbb{R}^{KN \times KN} \) is a permutation matrix. Furthermore, the vectors \( \tilde{x}_n \) and \( \tilde{s}_n \) are given by
\[
\tilde{x}_n = \begin{bmatrix} x_1^{[n]} \\ \vdots \\ x_K^{[n]} \end{bmatrix}, \quad \tilde{s}_n = \begin{bmatrix} s_1^{[n]} \\ \vdots \\ s_K^{[n]} \end{bmatrix},
\]
where \( x_k^{[n]} \) (respectively \( s_k^{[n]} \)) is the \( n \)th element in \( x_k \) (resp. \( s_k \)). Finally, the reordered extended mixing matrix is
\[
\tilde{\mathbf{A}} = \mathbf{P} \mathbf{A} \mathbf{P}^T = \begin{bmatrix} \tilde{\mathbf{A}}_{1,1} & \cdots & \tilde{\mathbf{A}}_{1,N} \\
\vdots & \ddots & \vdots \\
\tilde{\mathbf{A}}_{N,1} & \cdots & \tilde{\mathbf{A}}_{N,N} \end{bmatrix},
\]
where \( \tilde{\mathbf{A}}_{n,m} = \text{diag} (\tilde{a}_{n,m}) \), the vector \( \tilde{a}_{n,m} \in \mathbb{C}^{N \times 1} \) is defined as \( \tilde{a}_{n,m} = \begin{bmatrix} a_{1,m}^{[n]}, \ldots, a_{K,m}^{[n]} \end{bmatrix}^T \), and \( a_{k,m}^{[n]} \) denotes the element in the \( n \)th row and \( m \)th column of \( \mathbf{A}^k \).

### 3. PROBLEM FORMULATION

In this section, we formulate the IVA problem under the assumption of zero-mean complex and jointly-proper Gaussian sources.\(^2\) In particular, we consider the maximum-likelihood (ML) estimation of the mixing matrix \( \mathbf{A} \) and the covariance matrix of the latent variables \( \mathbf{R}_{\tilde{a},\tilde{a}} \). The key point for the solution of the IVA problem resides in the independence among the different sets of latent variables, that is, we know that \( \mathbf{R}_{\tilde{a}_{n,m}} = 0_{K \times K} \) for \( n \neq m \). In other words, \( \mathbf{R}_{\tilde{a},\tilde{a}} \) is a block-diagonal matrix with blocks of size \( K \times K \). Finally, we must mention that the identifiability conditions for the solution of the IVA problem from the SOS of the observations were analyzed in [5], whose principal result is summarized in the following lemma:

**Lemma 1 (J-BSS Identifiability Conditions)** [5] Given the data model in eq. (1), the sources and non-singular mixing matrices \( \mathbf{A}^k \) can be recovered from the SOS of the observations \( \mathbf{x} \) up to the trivial ambiguities (independent scale factors and a common permutation of the latent variables), and a non-trivial common mixture among the equivalently distributed sets of latent variables, i.e., the sets of latent variables \( \mathbf{s}_n, \mathbf{s}_m \) such that
\[
\mathbf{R}_{\tilde{s}_n,\tilde{s}_m} = \mathbf{A} \mathbf{R}_{\tilde{a}_{n,m}} \mathbf{A}^H,
\]
where \( \mathbf{A} \in \mathbb{C}^{K \times K} \) is a (non-singular) diagonal matrix.

### 3.1. Log-Likelihood Function

Under the assumption of zero-mean complex proper Gaussian data, the probability density function of the observations is
\[
p(\mathbf{x}) = \frac{1}{\pi^{KN} \det (\mathbf{R}_{\tilde{a},\tilde{a}})} \exp \left( - \frac{1}{2} \mathbf{x}^H \mathbf{R}_{\tilde{a},\tilde{a}}^{-1} \mathbf{x} \right).
\]
Thus, given \( T \) independent observations \( \tilde{x}_t \) (\( t = 0, \ldots, T-1 \), and after taking the logarithm, removing constant terms, and dividing by \( T \), we obtain the log-likelihood function
\[
\mathcal{L} = - \log \det (\mathbf{R}_{\tilde{a},\tilde{a}}) - \text{Tr} (\mathbf{R}_{\tilde{a},\tilde{a}}^{-1} \mathbf{R}_{\tilde{a},\tilde{a}}),
\]
where
\[
\mathbf{R}_{\tilde{a},\tilde{a}} = \frac{1}{T} \sum_{t=0}^{T-1} \tilde{x}_t \tilde{x}_t^H,
\]
is the sample covariance matrix.

### 3.2. ML Estimation Problem

Noting that \( \mathbf{R}_{\tilde{a},\tilde{a}} = \tilde{\mathbf{A}} \mathbf{R}_{\tilde{a},\tilde{a}} \tilde{\mathbf{A}}^H \), and adding the constant term \( \log \det (\mathbf{R}_{\tilde{a},\tilde{a}}) \) to the log-likelihood, it is easy to see that the ML estimates of \( \tilde{\mathbf{A}} \) and \( \mathbf{R}_{\tilde{a},\tilde{a}} \) are obtained as the solutions of
\[
\minimize_{\tilde{\mathbf{A}} \in \mathcal{B}_K, \mathbf{R}_{\tilde{a},\tilde{a}} \in \mathcal{D}_K} D (\mathbf{R}_{\tilde{a},\tilde{a}} \| \mathbf{R}_{\tilde{a},\tilde{a}}),
\]
where \( \mathcal{B}_K \) denotes the set of matrices with diagonal blocks of size \( K \times K \), \( \mathcal{D}_K \) is the set of semi-definite positive block-diagonal matrices with blocks of size \( K \times K \), and \( D (\mathbf{R}_{\tilde{a},\tilde{a}} \| \mathbf{R}_{\tilde{a},\tilde{a}}) \) denotes the Kullback-Leibler divergence between two zero-mean complex Gaussian distributions with covariance matrices \( \mathbf{R}_{\tilde{a},\tilde{a}} \) and \( \mathbf{R}_{\tilde{a},\tilde{a}} \), i.e.,
\[
D (\mathbf{R}_{\tilde{a},\tilde{a}} \| \mathbf{R}_{\tilde{a},\tilde{a}}) = \text{Tr} (\mathbf{R}_{\tilde{a},\tilde{a}}^{-1} \mathbf{R}_{\tilde{a},\tilde{a}}) - \log \det (\mathbf{R}_{\tilde{a},\tilde{a}}^{-1} \mathbf{R}_{\tilde{a},\tilde{a}}) - KN.
\]
Moreover, defining the separation matrix \( \mathbf{W} = \tilde{\mathbf{A}}^{-1} \), our estimation problem can be written as
\[
\minimize_{\mathbf{W} \in \mathcal{B}_K, \tilde{\mathbf{A}} \in \mathcal{D}_K} D (\tilde{\mathbf{A}} \| \tilde{\mathbf{A}}),
\]
where \( \tilde{\mathbf{A}} = \mathbf{WR}_{\tilde{a},\tilde{a}} \mathbf{W}^H \) can be seen as the sample covariance estimate for the recovered latent variables.

---

\(^2\)The results can be trivially extended to the case of real Gaussian vectors.

\(^3\)Obviously, this is equivalent to the assumption of diagonal cross-covariance matrices \( \mathbf{R}_{\tilde{a}_{k,l}} \).
3.3. Solution with respect to $R_{\hat{s}, \hat{s}}$
Let us start by writing the optimization problem in (2) as

$$\min_{W \in \mathcal{B}_K} \min_{R_{\hat{s}, \hat{s}} \in \mathcal{D}_K} D\left(R_{\hat{s}, \hat{s}} \| R_{\hat{s}, \hat{s}}\right).$$

Now, it is easy to prove that the solution with respect to $R_{\hat{s}, \hat{s}}$ is directly given by

$$R_{\hat{s}, \hat{s}} = \hat{D}_{\hat{s}, \hat{s}} = \text{blkdiag}_K\left(\hat{R}_{\hat{s}, \hat{s}}\right),$$

which reduces our estimation problem to

$$\min_{W \in \mathcal{B}_K} -\log \det \left(\Phi_{\hat{s}, \hat{s}}\right),$$

where $\Phi_{\hat{s}, \hat{s}} = \hat{D}_{\hat{s}, \hat{s}}^{1/2} \hat{R}_{\hat{s}, \hat{s}} \hat{D}_{\hat{s}, \hat{s}}^{-1/2}$ is defined as the coherence matrix associated to the recovered latent variables $\hat{s}_n[t]$, $(n = 1, \ldots, N, t = 0, \ldots, T - 1)$. Interestingly, $-\log \det \left(\Phi_{\hat{s}, \hat{s}}\right)$ can be seen as a measure of the mutual information among the sets of latent variables [4, 13]. Therefore, as expected, the IVA criterion reduces to minimizing the mutual information among the recovered sources.

4. GAUSSIAN ML IV A ALGORITHM

In this section, we present the GML-IVA algorithm, which amounts to solving the convex optimization problems obtained from quadratic approximations of the non-convex cost function in (3). In particular, in each iteration of the GML-IVA algorithm, the separation matrices are updated as

$$W \leftarrow \left(I_{KN} + \hat{\Delta}\right) W,$$

where $\hat{\Delta} \in \mathcal{B}_K$ is a sufficiently small matrix in order to ensure the invertibility of $W$.

4.1. Approximation of the cost function

Taking into account the structure of the updating rules, the cost function $J = -\log \det \left(\Phi_{\hat{s}, \hat{s}}\right)$ can be written as

$$J = -\log \det \left(\left[I_{KN} + \hat{\Delta}\right] R_{\hat{s}, \hat{s}} \left[I_{KN} + \hat{\Delta}^H\right]\right) + \log \det \left[\text{blkdiag}_K \left(\left[I_{KN} + \hat{\Delta}\right] R_{\hat{s}, \hat{s}} \left[I_{KN} + \hat{\Delta}^H\right]\right)\right]$$

$$= -\log \det \left(R_{\hat{s}, \hat{s}} + \Psi_{\hat{s}}\right) + \log \det \left[\text{blkdiag}_K \left(R_{\hat{s}, \hat{s}} + \Psi_{\hat{s}}\right)\right],$$

where $\Psi_{\hat{s}} = \hat{\Delta} R_{\hat{s}, \hat{s}} + R_{\hat{s}, \hat{s}} \hat{\Delta}^H + \hat{\Delta} R_{\hat{s}, \hat{s}} \hat{\Delta}^H$, and $R_{\hat{s}, \hat{s}}$ refers to the value of $WR_{\hat{s}, \hat{s}}W^H$ in the previous iteration. Now, left and right multiplying by $\hat{D}_{\hat{s}, \hat{s}}^{-1/2}$ we can write

$$J = -\log \det \left[\Phi_{\hat{s}, \hat{s}} + \Psi_{\Phi}\right] + \log \det \left[I_{KN} + \text{blkdiag}_K (\Psi_{\Phi})\right],$$

with $\Phi = \hat{D}_{\hat{s}, \hat{s}}^{-1/2} \hat{\Delta} \hat{D}_{\hat{s}, \hat{s}}^{1/2}$ and

$$\Psi_{\Phi} = \hat{\Omega} \Phi_{\hat{s}, \hat{s}} + \Phi_{\hat{s}, \hat{s}} \hat{\Omega}^H + \hat{\Omega} \Phi_{\hat{s}, \hat{s}} \hat{\Omega}^H.$$

In the next step we use the second order Taylor approximation for small $\Pi$

$$\log \det \left(I + \Pi\right) \simeq \text{Tr} \left(\Pi\right) - \frac{1}{2} \|\Pi\|^2,$$

which allows us to write

$$J \simeq \frac{1}{2} \left\|\Phi_{\hat{s}, \hat{s}} - I_{KN} + \Psi_{\Phi}\right\|^2 - \frac{1}{2} \left\|\text{blkdiag}_K (\Psi_{\Phi})\right\|^2,$$

or equivalently

$$J \simeq \frac{1}{2} \left\|\text{offdiag}_K \left(\Phi_{\hat{s}, \hat{s}} + \Psi_{\Phi}\right)\right\|^2.$$

Now, taking into account that $\|\Delta\| \ll 1$ implies $\|\hat{\Theta}\| \ll 1$, we can write $\Psi_{\Phi} \simeq \hat{\Theta} \Phi_{\hat{s}, \hat{s}} + \Phi_{\hat{s}, \hat{s}} \hat{\Theta}^H$, and approximate $J$ as

$$J \simeq \frac{1}{2} \left\|\text{offdiag}_K \left(\Phi_{\hat{s}, \hat{s}} + \hat{\Theta} \Phi_{\hat{s}, \hat{s}} + \Phi_{\hat{s}, \hat{s}} \hat{\Theta}^H\right)\right\|^2.$$

Moreover, writing $\Phi_{\hat{s}, \hat{s}} = I_{KN} + \text{offdiag}_K (\Theta_{\hat{s}, \hat{s}})$ and assuming that $\left\|\text{offdiag}_K \left(\Phi_{\hat{s}, \hat{s}}\right)\right\|^2 \ll 1$ (i.e. we are close to a solution of the IVA problem), we obtain

$$J \simeq \frac{1}{2} \left\|\text{offdiag}_K \left(\Phi_{\hat{s}, \hat{s}} + \hat{\Theta} + \hat{\Theta}^H\right)\right\|^2,$$

which can be finally rewritten as

$$J \simeq \sum_{n=1}^{N} \sum_{m=n+1}^{N} \left\|\Phi_{n,m} + \Theta_{n,m} + \Theta_{m,n}^H\right\|^2,$$

with

$$\Phi_{n,m} = \hat{R}_{n,m} - \hat{\delta}_{n,m} \hat{\delta}_{n,m}^H,$$

and where $\hat{R}_{n,m}$ can be seen as the sample covariance estimator of $R_{n,m}$, and $\hat{\delta}_{n,m} = \text{diag} \left(\hat{\delta}_{n,m}\right)$ denotes the $(n, m)$th $K \times K$ block of $\hat{\Delta}$.

4.2. GML-IVA Algorithm

Interestingly, the approximation of the cost function in (4) has the two following nice properties. Firstly, it is a quadratic cost function, whose solution (with respect to $\Delta$) can be obtained in closed-form. Secondly, the original non-convex optimization problem, with optimization variable $\Delta$, has been decomposed into $N(N-1)/2$ smaller convex problems, with optimization variables $\delta_{n,m}$ and $\delta_{m,n}$. Thus, in each step of the algorithm we have to solve the least squares problem

$$\min_{\delta_{n,m}, \delta_{m,n}} \left\|\Phi_{n,m} + \Theta_{n,m} + \Theta_{m,n}^H\right\|^2,$$

or, taking the column-wise vectorized version of the matrices in the previous problem

$$\min_{\delta_{n,m}, \delta_{m,n}} \left\|\phi_{n,m} + F_{n,m} G_{n,m} \delta_{m,n}\right\|^2,$$

where

$$\phi_{n,m} = \left(I_K \otimes \hat{R}_{n,m}^{-1/2}\right) \text{vec} \left(\hat{R}_{n,m} \hat{R}_{n,m}^{-1/2}\right),$$

$$F_{n,m} = \left(I_K \otimes \hat{R}_{n,m}^{-1/2}\right) \text{diagvec} \left(\hat{R}_{n,m}^{1/2}\right),$$

$$G_{n,m} = \left(I_K \otimes \hat{R}_{n,m}^{1/2}\right) \text{diagvec} \left(\hat{R}_{n,m}^{-1/2}\right).$$
Algorithm 1 Gaussian Maximum Likelihood Independent Vector Analysis (GML-IVA) Algorithm

\[\text{Algorithm 1 Gaussian Maximum Likelihood Independent Vector Analysis (GML-IVA) Algorithm} \]

**Input:** Sample covariance matrix \(R_{s,k}\) and threshold \(\mu\).

**Output:** Separation matrix \(\hat{W}\) and estimated residual cross-covariance \(\tilde{R}_{s,k}\).

**Initialize:** \(\hat{W} = I_{KN}, \tilde{R}_{s,k} = R_{s,k}\).

repeat

\[\text{Obtain} \quad \tilde{R}_{s,k}^{1/2} \tilde{R}_{s,k}^{-1/2} \quad \text{and} \quad \tilde{R}_{s,k}^{-1} \quad \text{for} \quad n = 1, \ldots, N.\]

\[\text{for} \quad n = 1, \ldots, N \quad \text{and} \quad m = n + 1, \ldots, N \quad \text{do} \]

\[\text{Obtain} \quad \delta_{n,m} \quad \text{and} \quad \delta_{n,n} \quad \text{from} \quad (9).\]

\[\text{end for}\]

\[\text{Obtain} \quad \Delta \quad \text{from} \quad \delta_{n,m} \quad (n, m = 1, \ldots, N).\]

\[\text{if} \quad \|\Delta\| > \mu, \text{then}\]

\[\text{Normalize} \quad \Delta = \frac{\mu}{\|\Delta\|}.\]

\[\text{end if}\]

\[\text{Update} \quad \hat{W} \leftarrow (I_{KN} + \Delta) \hat{W}.\]

\[\text{Update} \quad \tilde{R}_{s,k} \leftarrow (I_{KN} + \Delta) \tilde{R}_{s,k} (I_{KN} + \Delta)^H.\]

\[\text{until} \quad \text{Convergence}\]

With these definitions, it is clear that the solution of the LS problem in (5) is given by

\[\begin{bmatrix} \delta_{n,m} \\ \delta_{n,n} \end{bmatrix} = -U^{-1} V_2 1_{2K \times 1}, \quad (9)\]

where

\[U = \begin{bmatrix} \tilde{R}_{s,m,\beta_{s,n}}^T \odot \tilde{R}_{s,n,\beta_{s,m}}^{-1} I_K \\ \tilde{R}_{s,n,\beta_{s,m}}^{-T} \odot \tilde{R}_{s,m,\beta_{s,n}} \end{bmatrix},\]

\[V = \begin{bmatrix} \tilde{R}_{s,m,\beta_{s,n}}^{T/2} \odot \left( \tilde{R}_{s,n,\beta_{s,m}}^{-1/2} \tilde{R}_{s,m,\beta_{s,n}} \tilde{R}_{s,n,\beta_{s,m}}^{-1/2} \right) \\ \tilde{R}_{s,n,\beta_{s,m}}^{-T/2} \odot \left( \tilde{R}_{s,m,\beta_{s,n}} \tilde{R}_{s,n,\beta_{s,m}}^{-1/2} \right) \end{bmatrix},\]

and \(1_{2K \times 1}\) is the vector of ones of length \(2K\). Finally, the overall GML-IVA algorithm, which includes a normalization step to ensure the invertibility of \(\hat{W}\), is summarized in Algorithm 1.

4.3. Further Discussion

As we have seen, the GML-IVA algorithm amounts to minimizing the mutual information measure \(-\log \det (\hat{\Phi}_{s,k})\); i.e., it tries to block-diagonalize the matrix \(\hat{R}_{s,k}\). Equivalently, in terms of the sample (cross-)covariance matrices \(R_{s,k}\), the IVA problem can be seen as the problem of jointly-diagonalizing these \(K(K-1)/2\) matrices. Therefore, the proposed algorithm has some interesting connections with traditional joint-diagonalization techniques [10–12]. However, we must note that the IVA problem also has important particularities. For instance, IVA looks for \(K\) (and not only one) diagonalization matrices (the inverses of \(A^{[k]}\)), and it involves non-Hermitian sample cross-covariance matrices. In conclusion, conventional joint-diagonalization approaches cannot be applied to the IVA problem.

Interestingly, the IVA problem for Gaussian data is closely related to canonical correlation analysis (CCA) [6] and its generalization to several data sets [2, 3, 7–9]. In particular, the proposed algorithm reduces to conventional CCA in the case of \(K = 2\) data sets, and in this case a closed-form solution can be easily obtained. However, for \(K > 2\), the proposed method differs from the deflationary CCA generalizations proposed in [7, 8], and therefore, it can be seen as a new generalization of CCA.

Regarding the computational complexity of the proposed algorithm, the cost is dominated by the solution of (9), which is of order \(O(K^3)\). Since we have to solve \(N(N-1)/2\) instances of this problem in each iteration, the computational cost of the GML-IVA is \(O(N^2K^3)\) per iteration. Of course, the overall cost will depend on the number of iterations needed to achieve convergence, which depends on factors such as the accuracy of the assumption \(||\text{offdiag}_{\hat{\Phi}_{s,k}}||^2 \ll 1\). This implies that the convergence of the proposed algorithm will be slower at the beginning, although this problem could be partially avoided with a proper initialization approach [2].

5. EXPERIMENTAL RESULTS

In this section, the performance of the proposed algorithm is evaluated by means of some simulation examples. In all the experiments, the mixing matrices have been randomly generated with i.i.d. Gaussian entries of zero mean and unit variance. In the first set of simulations, the SOS \((R_{s,n,\beta_{s,n}})\) of the Gaussian latent variables have been generated as complex Wishart matrices with \(K\) degrees of freedom. The convergence of the proposed algorithm, for one realization of \(R_{s,n,\beta_{s,n}}\), and for \(K = 5\) data sets of dimension \(N = 5\), is illustrated in Fig. 1. The figure shows the evolution of the cost function for 200 different initializations and, as can be seen, the algorithm always converges to the same solution or, to be more specific, to solutions with the same value of the cost function. Additionally, Fig. 2 shows the cumulative density function of the number of iterations until convergence when the matrices \(R_{s,n,\beta_{s,n}}\) are randomly generated. As can be seen, the probability of convergence in less than 100 iterations is
We have presented a maximum-likelihood (ML) approach to the independent vector analysis (IVA) of Gaussian data sets. The proposed method, which is exclusively based on the second-order statistics (SOS) of the observations, amounts to minimizing an estimate of the mutual information among the sets of latent variables. In particular, the GML-IVA algorithm solves different least squares (LS) problems obtained from quadratic approximations of the non-convex cost function. The convergence and performance of the proposed technique have been illustrated by means of different numerical examples, which corroborate that the GML-IVA algorithm is a good candidate for practical IVA problems when the sources are expected to obey the SOS identifiability conditions.

6. CONCLUSIONS

We have presented a maximum-likelihood (ML) approach to the independent vector analysis (IVA) of Gaussian data sets. The proposed method, which is exclusively based on the second-order statistics (SOS) of the observations, amounts to minimizing an estimate of the mutual information among the sets of latent variables. In particular, the GML-IVA algorithm solves different least squares (LS) problems obtained from quadratic approximations of the non-convex cost function. The convergence and performance of the proposed technique have been illustrated by means of different numerical examples, which corroborate that the GML-IVA algorithm is a good candidate for practical IVA problems when the sources are expected to obey the SOS identifiability conditions.

7. REFERENCES


Fig. 4. Joint Blind Source Separation of three 512 × 512 color images. The mixtures in each color (RGB) channel is different.


