

BLIND SOURCE SEPARATION AND CHANNEL IDENTIFICATION: EXPLOITING 2ND-ORDER STATISTICS IN BAYESIAN FRAMEWORKS

Paulo Lopes, João Xavier, and Victor Barroso

Abstract - We study how 2nd order statistics (SOS) can be exploited in two signal processing problems, blind separation of binary sources and trained-based multi-user channel identification, in a Bayesian context where a prior on the mixing channel matrix is available. It is well known that the SOS of the received data permit to resolve the unknown mixing matrix, up to an orthogonal factor. In a Bayesian framework, this residual orthogonal mixing matrix becomes a random object in its own right, with an associated distribution over the group of orthogonal matrices. This distribution is induced by the prior on the mixing matrix, and must be known for optimum statistical processing. We rely on a previous theoretical work to provide these answers, and discuss applications for this induced probability density function (pdf) over the orthogonal group, in the two aforementioned signal processing problems. Preliminary results, obtained through computer simulations, demonstrate the effectiveness of incorporating this induced distribution associated with the residual orthogonal matrix into the design of several estimators.

Keywords: Second-Order Statistics, Bayesian Frameworks, Blind Source Separation, Random Matrix Theory, Blind Channel Identification

Resumo - Estudamos a forma como estatísticas de segunda ordem podem ser exploradas em dois problemas de processamento de sinal, separação cega de fontes binárias e identificação de canal multi-utilizador baseada em treino prévio, num contexto Bayesiano em que se tem acesso a um modelo probabilístico da matriz de mistura do canal. É sabido que as estatísticas de segunda ordem dos dados recebidos permitem resolver a matriz de mistura incógnita a menos de um factor ortogonal. Num cenário Bayesiano, esta matriz ortogonal residual de mistura torna-se um objecto aleatório em si mesmo, com uma distribuição associada sobre o grupo das matrizes ortogonais. Esta distribuição é induzida pelo modelo probabilístico da matriz de mistura, e tem de ser conhecida para efeitos de processamento estatístico óptimo. Baseamo-nos em trabalho teórico prévio para fornecer estas respostas, e discutimos aplicações para esta função densidade de probabilidade (pdf) sobre o grupo ortogonal, nos dois problemas de processamento de sinal

This work was supported by the FCT Programa Operacional Sociedade de Informação (POSI) in the frame of QCA III, under contract POSI/2001/CPS/38775

All authors are with the Instituto de Sistemas e Robótica, Instituto Superior Técnico, Torre Norte. Piso 7, Av. Rovisco Pais, 1049-001, Fax: +351 21 842 8291. E-mails: paulomiguel@yahoo.com, jxavier,vab@isr.ist.utl.pt

anteriormente mencionados. Resultados preliminares, obtidos através de simulações em computador, demonstram a relevância da incorporação desta distribuição induzida associada com a matriz ortogonal residual na concepção de diversos estimadores.

Palavras-chave: Estatísticas de Segunda Ordem, Contextos Bayesianos, Separação Cega de Fontes, Teoria das Matrizes Aleatórias, Identificação Cega de Canal

1. INTRODUCTION

Blind source separation (BSS) has been an active area of research over the past few years [1]. It finds direct application in the exploding field of wireless multi-user communications with spatial diversity, e.g., Space Division Multiple Access (SDMA) networks. In these wireless systems, unknown space-time channels mix the co-channel user signals prior to base station reception. Blind signal separation techniques are needed at the receiver to reconstruct the source signals from the antenna array observations [2, 3, 4, 5, 6].

A common first step in BSS techniques consists in exploiting the 2nd order statistics (SOS) of the observations to partially resolve the unknown mixing matrix [1, 7, 8, 9, 10, 11]. Usually, the SOS of the received data are used to turn the unknown mixing matrix into an unknown rotation mixing matrix. This simplifies the remaining processing as the algebraic constraints of the orthogonal group can be efficiently exploited for algorithmic purposes. Notice that the original (nonwhitened) mixing matrix often lacks any interesting structure. The residual unknown rotation matrix can be solved under several identification strategies depending on the source characteristics, number of available data samples, etc. Some options include: iterative joint diagonalization of several cumulant matrices for non-Gaussian signals [7], iterative joint diagonalization of several covariance matrices for instantaneously mixed stationary sources with sufficiently diverse but unknown 2nd order spectra [8], closed-form isometry fitting for convolutively mixed stationary sources with sufficiently diverse and known 2nd order spectra [6], analytical signal separation for constant-modulus sources [2], iterative demodulation of finite-alphabet sources [3], globally convergent iterative separation of independent and identically distributed sources by kurtosis-based criteria [11].

In many scenarios, the mixing matrix can be modeled as a random object with given prior probability density function (pdf). In [12], we studied how a given pdf on the set of non-whitened mixing matrices contracts to a pdf on the lower-dimensional orthogonal group containing the whitened

mixing matrices. The contribution of this paper consists in providing signal processing applications for the theoretical framework developed in [12]. We show how the derived priors over the orthogonal group can be exploited for improving performance in two problems: blind separation of co-channel binary sources and trained-based channel identification in multi-antenna systems. Further applications can be found in [13].

Our paper is organized as follows. In section 2, we introduce our data model and briefly review the work in [12]. We assume that the non-whitened random mixing matrix has a zero-mean matrix variate normal distribution with given dispersion matrix. A particular case of this prior is commonplace in works with multiple-antenna systems where it is known as the independent Rayleigh fading assumption [14]. We handle a more general model allowing correlation between the entries of the mixing matrix. We present the results of [12] in which we examined how this prior on the non-whitened matrices contracts to a pdf over the group of orthogonal matrices, under the action of two distinct prewhitening methods. The two prewhitening methods considered are based on the polar and the LU decomposition of the non-whitened mixing matrix, respectively. In section 3, we address the problem of blind source separation when the mixing matrix is drawn from a known zero-mean Gaussian prior. We apply the results in [12] to find educated guesses for initializing a locally convergent source separation algorithm. In section 4, we address the problem of channel identification in the context of multi-antenna systems. We show how the pdfs derived in [12] can be exploited to improve the accuracy of trained-based channel estimators. Section 5 contains the main conclusions of our work.

Throughout the paper, we use the following notation. The set of $n \times n$ matrices with real entries is denoted by $\mathbb{R}^{n \times n}$. Matrices are written in bold uppercase. The symbols $(\cdot)^T$, $\det(\cdot)$, \otimes and I_n denote the transpose operator, the determinant, the Kronecker product and the $n \times n$ identity matrix, respectively. The notation $\mathbb{GL}(n, \mathbb{R})$, $\mathbb{O}(n) = \{Q : Q^T Q = I_n\}$ and $\mathbb{L}(n)$, stand for the groups of $n \times n$ non-singular, orthogonal and lower triangular matrices with positive diagonal entries, respectively. The cone of positive definite matrices of size $n \times n$ is represented by $\mathbb{P}(n)$. Additional notation is introduced as needed.

2. DATA MODEL AND PREVIOUS WORK

We adopt the standard discrete-time instantaneous linear mixture data model, e.g., see [1, 8],

$$\mathbf{x}[n] = \mathbf{A}s[n] + \mathbf{w}[n], \quad (1)$$

where $\mathbf{x}[n] = (x_1[n], \dots, x_M[n])^T$ represents the M -dimensional vector of observations, $\mathbf{A} \in \mathbb{R}^{M \times M}$ stands for the mixing matrix, $\mathbf{s}[n] = (s_1[n], \dots, s_M[n])^T$ contains the M source signals, and $\mathbf{w}[n] = (w_1[n], \dots, w_M[n])^T$ models observation noise. Here, for simplicity, all data objects take values in the field of real numbers. As usual, the sources are assumed to be zero-mean, 2nd order stationary and uncorrelated, $\mathbf{R}_s = E\{s[n]s[n]^T\} = I_M$, and the mixing matrix is

non-singular, $\mathbf{A} \in \mathbb{GL}(M, \mathbb{R})$. We assume that the process $w[n]$ is zero-mean and wide-sense stationary with known correlation matrix $\mathbf{R}_w[0] = E\{w[n]w[n]^T\}$. It is well known that the 2nd order statistics of the observations can be exploited to partially solve for the unknown mixing matrix \mathbf{A} . Here, we consider two alternative methods based on the polar and the LU decomposition of \mathbf{A} , respectively. Both methods act on the so-called denoised correlation matrix of $\mathbf{x}[n]$,

$$\mathbf{R} = \mathbf{R}_x[0] - \mathbf{R}_w[0] = \mathbf{A}\mathbf{A}^T, \quad (2)$$

where $\mathbf{R}_x[0] = E\{\mathbf{x}[n]\mathbf{x}[n]^T\}$. In practice, $\mathbf{R}_x[0]$ can be replaced by its sample-mean estimator

$$\widehat{\mathbf{R}}_x[0] = \frac{1}{N} \sum_{n=1}^N \mathbf{x}[n]\mathbf{x}[n]^T, \quad (3)$$

where N denotes the number of available data samples. **i)** Focusing first on the polar decomposition, write $\mathbf{A} = \mathbf{P}\mathbf{Q}$ where $\mathbf{P} \in \mathbb{P}(M)$ and $\mathbf{Q} \in \mathbb{O}(M)$. This factorization exists and it is unique for any $\mathbf{A} \in \mathbb{GL}(M, \mathbb{R})$ [16]. Substituting in (2) yields $\mathbf{R} = \mathbf{P}^2$. Thus, \mathbf{P} can be obtained from the available \mathbf{R} as its square-root, $\mathbf{P} = \mathbf{R}^{1/2}$. Thus, the SOS of the data permit the receiver to recover the factor \mathbf{P} of the mixing matrix $\mathbf{A} = \mathbf{P}\mathbf{Q}$. The factor \mathbf{Q} is not resolved. **ii)** Using the LU decomposition, we can write also uniquely $\mathbf{A} = \mathbf{L}\mathbf{U}$, where $\mathbf{L} \in \mathbb{L}(M)$ and $\mathbf{U} \in \mathbb{O}(M)$. Thus, $\mathbf{R} = \mathbf{L}\mathbf{L}^T$, meaning that \mathbf{L} may be obtained from \mathbf{R} as its unique Cholesky factor. Again, the mixing matrix $\mathbf{A} = \mathbf{L}\mathbf{U}$ is partially resolved. The factor \mathbf{L} is revealed by the correlation matrix of the observations, but \mathbf{U} remains unknown.

After either the PQ or LU pre-processing step is performed, our original data model (1) switches to

$$\mathbf{x}[n] = \mathbf{P}\mathbf{Q}\mathbf{s}[n] + \mathbf{w}[n] \quad (4)$$

or

$$\mathbf{x}[n] = \mathbf{L}\mathbf{U}\mathbf{s}[n] + \mathbf{w}[n], \quad (5)$$

with \mathbf{P} or \mathbf{L} known, respectively. If the original mixing matrix has a prior, $\mathbf{A} \sim p(\mathbf{A})$, then the unknown residual orthogonal matrices \mathbf{Q} and \mathbf{U} in (4) and (5), respectively, denote random objects. In [12], we investigated the distributions of \mathbf{Q} or \mathbf{U} for a given prior on \mathbf{A} . The results derived in [12] are expressed in the setting of Riemannian geometry [17, 18]. This viewpoint is natural and almost mandatory since we are dealing with distributions over lower-dimensional submanifolds of Euclidean spaces, e.g., the orthogonal group $\mathbb{O}(M)$, the cone of positive definite matrices $\mathbb{P}(M)$, etc. In the sequel, we assume the reader to be acquainted with such differential-geometric concepts. We try to use notation compatible with [17]. We regard $\mathbb{O}(M)$, $\mathbb{P}(M)$ and $\mathbb{L}(M)$ as differentiable manifolds taking their Riemannian structure from the corresponding embedding, e.g., $\iota : \mathbb{O}(M) \rightarrow \mathbb{GL}(M, \mathbb{R})$, $\iota(\mathbf{X}) = \mathbf{X}$. Here, $\mathbb{GL}(M, \mathbb{R})$ is identified with an open subset of the Euclidean space \mathbb{R}^{M^2} by interpreting an $M \times M$ matrix as an M^2 -dimensional vector. All these manifolds are orientable and we let $\Omega_{\mathbb{GL}(M, \mathbb{R})}$, $\Omega_{\mathbb{O}(M)}$, $\Omega_{\mathbb{P}(M)}$ and $\Omega_{\mathbb{L}(M)}$ denote the corresponding volume elements (the particular choice of orientation is not important to us) derived from their Riemannian metrics. Moreover,

when taking the Cartesian product of manifolds, we implicitly assume the canonical construction for the product metric, hence, for the volume element of the product manifold. In this Riemannian context, a mass distribution or pdf over any of these manifolds is a non-negatively oriented exterior form. As an example, a mass distribution Ω , say, over the orthogonal group, belongs to the bundle of alternating tensors $\bigwedge^m(\mathbb{O}(M))$, where $m = \dim \mathbb{O}(M) = M(M-1)/2$. Since for any given distribution Ω over $\mathbb{O}(M)$, we have $\Omega = f \Omega_{\mathbb{O}(M)}$ for an unique nonnegative smooth function $f : \mathbb{O}(M) \rightarrow \mathbb{R}$, we use the terminology *mass distribution* for either Ω or f . This also applies to the other manifolds considered in this paper.

In the sequel, we shall make use of the following results from [12]. Let $p(\mathbf{A})$ denote the pdf (prior) on the mixing matrix $\mathbf{A} \in \mathbb{GL}(M, \mathbb{R})$. Then, the factorization $\mathbf{A} = \mathbf{P}\mathbf{Q}$ induces the pdf on $\mathbb{P}(M) \times \mathbb{O}(M)$ given, up to a normalizing constant, by

$$p(\mathbf{P}, \mathbf{Q}) = p(\mathbf{P}\mathbf{Q})g(\mathbf{P}), \quad (6)$$

where the function $g : \mathbb{P}(M) \rightarrow \mathbb{R}$ satisfies $g(I_V(\mathbf{P})) = g(\mathbf{P})$, where $I_V : \mathbb{P}(n) \rightarrow \mathbb{P}(n)$ denotes conjugation by $V \in \mathbb{O}(n)$, $I_V(\mathbf{P}) = \mathbf{V}\mathbf{P}\mathbf{V}^T$. Likewise, the factorization $\mathbf{A} = \mathbf{L}\mathbf{U}$ induces a pdf on $\mathbb{L}(M) \times \mathbb{O}(M)$ given, up to a normalizing constant, by

$$p(\mathbf{L}, \mathbf{U}) = p(\mathbf{L}\mathbf{U})h(\mathbf{L}), \quad (7)$$

where $h : \mathbb{L}(M) \rightarrow \mathbb{R}$ is given by $h(\mathbf{L}) = \text{pm}(\mathbf{L})/\det(\mathbf{L})$. Here, the function $\text{pm} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ is defined, for $\mathbf{X} \in \mathbb{R}^{n \times n}$, as

$$\text{pm}(\mathbf{X}) = \prod_{m=1}^n \det \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mm} \end{pmatrix}.$$

In this paper, we restrict ourselves to the scenario where \mathbf{A} has a zero-mean matrix variate normal distribution with covariance matrix $\Sigma \otimes \Psi$, denoted $\mathbf{A} \sim \mathcal{N}(\mathbf{0}, \Sigma \otimes \Psi)$, where $\Sigma, \Psi \in \mathbb{P}(M)$, see [15]. This means that $\mathbf{A} \stackrel{d}{=} \Sigma^{1/2} \mathbf{Z} \Psi^{1/2}$ (equality in distribution), where \mathbf{Z} denotes an $M \times M$ random matrix whose entries are independent and identically distributed as zero-mean unit-variance Gaussian random variables. The assumption $\mathbf{A} \sim \mathcal{N}(\mathbf{0}, \Sigma \otimes \Psi)$ means that \mathbf{A} is distributed over $\mathbb{GL}(M, \mathbb{R})$ according to

$$p(\mathbf{A}) = \alpha \text{etr} \left\{ -\frac{1}{2} \Sigma^{-1} \mathbf{A} \Psi^{-1} \mathbf{A}^T \right\},$$

where $\alpha = (2\pi)^{-\frac{1}{2}M^2} \det(\Sigma\Psi)^{-\frac{M}{2}}$ and $\text{etr}\{\mathbf{X}\} = \exp\{\text{tr}(\mathbf{X})\}$ for a generic matrix \mathbf{X} , see [15]. Using (6) and (7) for this particular choice of the prior on \mathbf{A} , we have, up to a constant, the joint distributions on the pairs (\mathbf{P}, \mathbf{Q}) and (\mathbf{L}, \mathbf{U}) ,

$$p(\mathbf{P}, \mathbf{Q}) = \text{etr} \left(-\frac{1}{2} \Sigma^{-1} \mathbf{P}\mathbf{Q}\Psi^{-1}\mathbf{Q}^T\mathbf{P} \right) g(\mathbf{P}) \quad (8)$$

and

$$p(\mathbf{L}, \mathbf{U}) = \text{etr} \left(-\frac{1}{2} \Sigma^{-1} \mathbf{L}\mathbf{U}\Psi^{-1}\mathbf{U}^T\mathbf{L}^T \right) h(\mathbf{L}), \quad (9)$$

respectively.

3. BLIND SEPARATION OF BINARY SOURCES

In this section, we present a possible application for the results in [12], more specifically, for the induced pdfs in (8) and (9). We take the data model (1) along with the already discussed prior on the unknown mixing matrix $\mathbf{A} \sim \mathcal{N}(\mathbf{0}, \Sigma \otimes \Psi)$, where the dispersion matrices Σ and Ψ are assumed known at the receiver. Moreover, we let $\mathbf{w}[n]$ denote zero-mean, spatio-temporal white Gaussian distributed noise with known power σ^2 , i.e.,

$$\mathbf{R}_{\mathbf{w}}[k] = \mathbb{E} \{ \mathbf{w}[n] \mathbf{w}[n-k]^T \} = \sigma^2 \mathbf{I}_M \delta[k], \quad (10)$$

where $\delta[\cdot]$ denotes the discrete-time Kronecker delta ($\delta[0] = 1$ and $\delta[k] = 0$ for nonzero k). We consider that $\mathbf{s}[n] = (s_1[n], \dots, s_M[n])^T$ denotes a vector of M independent binary sources. We assume that each source emits independent and identically distributed symbols, $\text{Prob}\{s_m[n] = 1\} = \text{Prob}\{s_m[n] = -1\} = 1/2$, for $m = 1, 2, \dots, M$. See [2, 3, 4, 5, 9] for closely related, although non-Bayesian (no prior is assumed on \mathbf{A}), wireless communication scenarios. Assuming that N data samples are available, we have the matricial data model

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{W}, \quad (11)$$

where $\mathbf{X} = [x[1] x[2] \cdots x[N]]$ denotes the data matrix containing the observations, $\mathbf{S} = [s[1] s[2] \cdots s[N]]$ contains the information sequences sent by the sources (the m th row corresponds to the m th source) and $\mathbf{W} = [w[1] w[2] \cdots w[N]]$ stands for the additive noise matrix. Figure 1 illustrates a block diagram which corresponds to the matricial model in (11). We are interested in estimating the

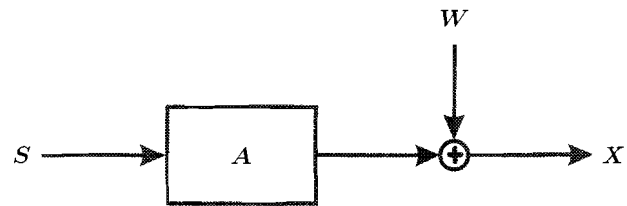


Figure 1. Block diagram corresponding to the data model (11)

binary matrix \mathbf{S} in (11) from the available data matrix \mathbf{X} , without knowing the mixing channel matrix \mathbf{A} . A natural approach in this Bayesian setting is to look for the maximum a posteriori (MAP) estimates of the transmitted bits,

$$\hat{\mathbf{S}}_{\text{MAP}} = \arg \max_{\mathbf{S} \in \mathcal{B}_{M \times N}} p(\mathbf{S} | \mathbf{X}), \quad (12)$$

where $\mathcal{B}_{M \times N}$ stands for the discrete set of $M \times N$ binary matrices. With the prior $\mathbf{A} \sim \mathcal{N}(\mathbf{0}, \Sigma \otimes \Psi)$, it is easily seen (after some straightforward computations) that problem (12) is equivalent to

$$\hat{\mathbf{S}}_{\text{MAP}} = \arg \min_{\mathbf{S} \in \mathcal{B}_{M \times N}} \log \det(\mathbf{C}) + \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} \quad (13)$$

where

$$\mathbf{C} = \mathbf{S}^T \Psi \mathbf{S} \otimes \Sigma + \sigma^2 \mathbf{I}_{MN} \quad (14)$$

and $\mathbf{x} = \text{vec}(\mathbf{X})$ corresponds to the columns of \mathbf{X} stacked into the vector \mathbf{x} . Problem (13) is an integer optimization problem over a set with cardinality $\#\mathcal{B} = 2^{MN}$. Thus, solving (13) is infeasible due to the required high-dimensional exhaustive search. An alternative approach, which leads to a feasible computational scheme, consists in estimating the most probable realization of the joint channel-source matrix pair given the available data, *i.e.*,

$$\widehat{(\mathbf{A}, \mathbf{S})}_{\text{MAP}} = \arg \max_{\mathbf{A} \in \text{GL}(M, \mathbb{R}), \mathbf{S} \in \mathcal{B}_{M \times N}} p(\mathbf{A}, \mathbf{S} | \mathbf{X}). \quad (15)$$

Problem (15) can be solved by the following locally-convergent iterative algorithm: given an initial estimate $\mathbf{A}^{(0)}$ for the channel matrix, let

$$\begin{aligned} \mathbf{S}^{(k+1)} &= \arg \max_{\mathbf{S} \in \mathcal{B}_{M \times N}} p(\mathbf{A}^{(k)}, \mathbf{S} | \mathbf{X}) \\ \mathbf{A}^{(k+1)} &= \arg \max_{\mathbf{A} \in \text{GL}(M, \mathbb{R})} p(\mathbf{A}, \mathbf{S}^{(k+1)} | \mathbf{X}) \end{aligned}$$

for $k = 0, 1, 2, \dots$ until a fixed-point is attained. Given our statistical assumptions, and having defined $\mathbf{a}^{(k+1)} = \text{vec}(\mathbf{A}^{(k+1)})$, we have, after some calculus, the iterates

$$\mathbf{S}^{(k+1)} = \arg \min_{\mathbf{S} \in \mathcal{B}_{M \times N}} \left\| \mathbf{X} - \mathbf{A}^{(k)} \mathbf{S} \right\|^2 \quad (16)$$

$$\mathbf{a}^{(k+1)} = \mathbf{E}^{-1} \text{vec}(\mathbf{F}), \quad (17)$$

where

$$\mathbf{E} = \Psi \mathbf{S}^{(k+1)} \mathbf{S}^{(k+1)T} \otimes \mathbf{I}_M + \mathbf{I}_M \otimes \sigma^2 \Sigma^{-1}$$

and

$$\mathbf{F} = \mathbf{X} \mathbf{S}^{(k+1)T} \Psi.$$

Solving problem (16) does not require a search over $\mathcal{B}_{M \times N}$. In fact, since

$$\left\| \mathbf{X} - \mathbf{A}^{(k)} \mathbf{S} \right\|^2 = \sum_{n=1}^N \left\| \mathbf{x}[n] - \mathbf{A}^{(k)} \mathbf{s}[n] \right\|^2,$$

the optimization problem decouples in N independent sub-problems. The n th subproblem only involves the n th column of \mathbf{S} . Thus, the n th column of $\mathbf{S}^{(k+1)}$, written $\mathbf{s}^{(k+1)}[n]$ can be found by solving

$$\mathbf{s}^{(k+1)}[n] = \arg \min_{\mathbf{s} \in \mathcal{B}_M} \left\| \mathbf{x}[n] - \mathbf{A}^{(k)} \mathbf{s} \right\|^2, \quad (18)$$

where \mathcal{B}_M denotes the set of M -dimensional binary vectors. Problem (18) requires a search over a much smaller set, $\#\mathcal{B}_M = 2^M$, and is easily implemented with parallel processors. The main drawback of the iterative algorithm in (16) and (17) is its lack of global convergence. Accurate initial points $\mathbf{A}^{(0)}$ are required to obtain a good performance in practice. In the sequel, we delineate a method which exploits the SOS of the received data and the results in (8) and (9) to find educated guesses for starting the iterations. We present our method only for the PQ factorization. The extension to

the LU factorization is straightforward. We start by partially solving for \mathbf{A} as explained in section 2. Namely, consider the sample-mean estimate of the correlation matrix of the observed data $\mathbf{x}[n]$ given in (3). The denoised correlation matrix $\widehat{\mathbf{R}}$ in (2) is estimated as

$$\widehat{\mathbf{R}} = \widehat{\mathbf{R}}_x[0] - \beta \sigma^2 \mathbf{I}_M, \quad (19)$$

where β denotes the maximum number in $\{0, 0.1, 0.2, \dots, 0.9, 1\}$ which makes the right-hand side of (19) positive-definite. We do not simply subtract $\widehat{\mathbf{R}}_w[0] = \sigma^2 \mathbf{I}_M$ from $\widehat{\mathbf{R}}_x[0]$ as equation (2) suggests, because, for finite data sets ($N < \infty$), that method does not guarantee a positive-definite matrix $\widehat{\mathbf{R}}$ (which is essential for the remaining processing). Let

$$\widehat{\mathbf{R}} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \quad (20)$$

denote an eigenvalue decomposition of $\widehat{\mathbf{R}}$. That is, $\mathbf{V} \in \mathbb{O}(M)$ and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$ denotes a diagonal matrix with positive diagonal entries. In (20), we assume that the diagonal of $\mathbf{\Lambda}$ is sorted in increasing order, $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_M$. From (20), the \mathbf{P} factor of $\mathbf{A} = \mathbf{P}\mathbf{Q}$ is estimated as $\widehat{\mathbf{P}} = \mathbf{V} \mathbf{\Lambda}^{1/2} \mathbf{V}^T$. We propose to initialize the aforementioned iterative algorithm with

$$\mathbf{A}^{(0)} = \widehat{\mathbf{P}} \widehat{\mathbf{Q}}, \quad (21)$$

where

$$\widehat{\mathbf{Q}} = \arg \max_{\mathbf{Q} \in \mathbb{O}(M)} p(\mathbf{Q} | \mathbf{P} = \widehat{\mathbf{P}}). \quad (22)$$

That is, $\widehat{\mathbf{Q}}$ denotes the most probable realization of the \mathbf{Q} factor of the mixing matrix \mathbf{A} , given that its \mathbf{P} factor is $\widehat{\mathbf{P}}$. Given the joint (\mathbf{P}, \mathbf{Q}) pdf in (8), we have

$$\widehat{\mathbf{Q}} = \arg \min_{\mathbf{Q} \in \mathbb{O}(M)} \text{tr} \left(\mathbf{Q} \Psi^{-1} \mathbf{Q}^T \widehat{\mathbf{P}} \Sigma^{-1} \widehat{\mathbf{P}} \right). \quad (23)$$

A closed-form solution for (23) is available [16], and can be computed as follows. Let $\Psi^{-1} = \mathbf{Z} \mathbf{D} \mathbf{Z}^T$ denote an eigendecomposition of Ψ^{-1} , where $\mathbf{Z} \in \mathbb{O}(M)$ and $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_M)$ denotes a diagonal matrix with its diagonal entries sorted in decreasing order, $d_1 \geq d_2 \geq \dots \geq d_M$. Also, let $\widehat{\mathbf{P}} \Sigma^{-1} \widehat{\mathbf{P}} = \mathbf{Y} \mathbf{\Phi} \mathbf{Y}^T$ denote an eigendecomposition of $\widehat{\mathbf{P}} \Sigma^{-1} \widehat{\mathbf{P}}$, where $\mathbf{Y} \in \mathbb{O}(M)$ and $\mathbf{\Phi} = \text{diag}(\phi_1, \phi_2, \dots, \phi_M)$ denotes a diagonal matrix with its diagonal entries sorted in increasing order, $\phi_1 \leq \phi_2 \leq \dots \leq \phi_M$. Then, $\widehat{\mathbf{Q}} = \mathbf{Y} \mathbf{Z}^T$. Notice that the determination of $\widehat{\mathbf{Q}}$ does not involve any significant extra computational burden since \mathbf{Z} can be computed off-line (it does not depend on the received data, only on Ψ). As a final remark, perhaps a more defensible choice for $\mathbf{A}^{(0)}$ would be

$$\mathbf{A}^{(0)} = \arg \max_{\mathbf{A} \in \text{GL}(M, \mathbb{R})} p(\mathbf{A} | \mathbf{X}),$$

or $\mathbf{A}^{(0)} = \widehat{\mathbf{P}} \widetilde{\mathbf{Q}}$, where

$$\widetilde{\mathbf{Q}} = \arg \max_{\mathbf{Q} \in \mathbb{O}(M)} p(\mathbf{Q} | \mathbf{X}).$$

Certainly, both these approaches incorporate more information, in fact, all the available data \mathbf{X} , than our simple method in (21), which makes use of only the SOS of the observations. However, it is easily checked that both these alternative approaches lead to computationally untractable problems.

We conducted some computer simulations to assess the effectiveness of our initialization scheme. We considered a scenario with $M = 2$ binary users. The prior on the mixing matrix \mathbf{A} is $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{\Psi})$, where

$$\mathbf{\Psi} = \mathbf{G} \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{G}^T \quad (24)$$

with

$$\mathbf{G} = \begin{bmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{bmatrix}.$$

and

$$\mathbf{\Sigma} = \mathbf{H} \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{H}^T \quad (25)$$

with

$$\mathbf{H} = \begin{bmatrix} \cos(\pi/8) & -\sin(\pi/8) \\ \sin(\pi/8) & \cos(\pi/8) \end{bmatrix}.$$

We varied the signal-to-noise ratio (SNR) from $\text{SNR}_{\min} = 5$ dB to $\text{SNR}_{\max} = 15$ dB, in steps of $\Delta = 2.5$ dB. The SNR is defined as $\text{SNR} = \mathbb{E} \|\mathbf{A}\mathbf{s}[n]\|^2 / \mathbb{E} \|\mathbf{w}[n]\|^2 = \text{tr}(\mathbf{\Sigma})\text{tr}(\mathbf{\Psi}) / M\sigma^2$. For each SNR, 5000 statistically independent Monte-Carlo runs were performed. Each Monte-Carlo run consists in generating a realization of \mathbf{A} , \mathbf{S} and \mathbf{W} , see (11), for a data packet length of $N = 200$. Next, $I = 1$ iteration of the iterative algorithm in (16) and (17) is performed starting from the educated guess $\mathbf{A}^{(0)}$ in (21). For comparison, we also performed $I = 1$ iteration starting from a random initialization $\mathbf{A}^{(0)} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{\Psi})$, that is, an independent realization of the channel model. Figures 2 and 3 present the bit error rate (BER), averaged over the Monte-Carlos, for user 1 and user 2, respectively, as a function of the SNR. The solid line with stars denotes a bound (max-

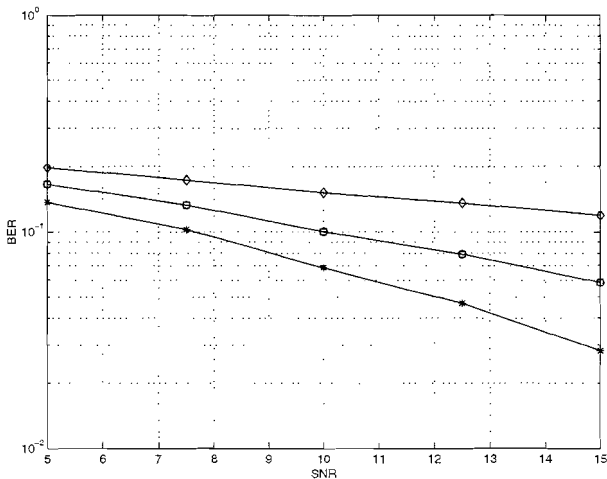


Figure 2. BER of user 1 versus SNR ($I = 1$ iteration)

imum likelihood bit decoding with the channel \mathbf{A} known). The solid line with squares refers to our proposed initialization using the LU factorization and the solid line with circles

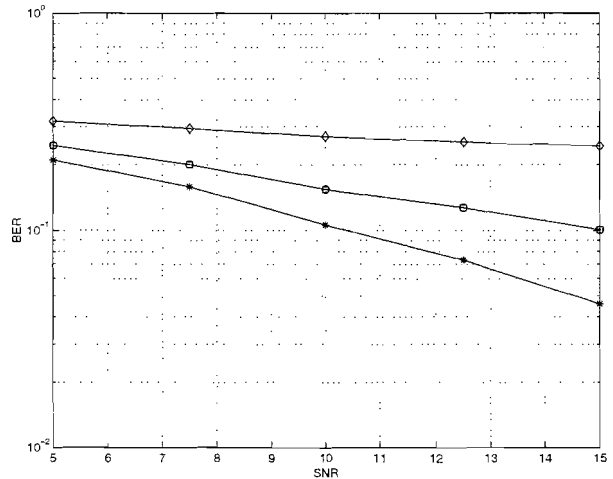


Figure 3. BER of user 2 versus SNR ($I = 1$ iteration)

refers to our proposed initialization using the PQ factorization (these two curves coincide). The solid line with diamonds corresponds to the random initialization. As can be seen, our educated guess permits to outperform the random initialization. Figure 4 presents the mean square error (MSE) of the channel estimate, averaged over the Monte Carlos, as a function of the SNR. The lines follow the same definitions,

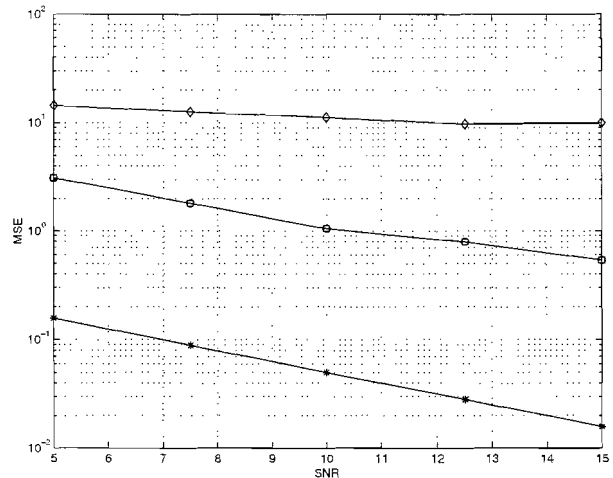


Figure 4. MSE of channel estimate versus SNR ($I = 1$ iteration)

except that the bound corresponds to the stochastic Crámer-Rao bound for the MSE of the channel matrix. This bound is obtained as follows. We suppose that \mathbf{S} is known. If we apply the $\text{vec}(\cdot)$ operator to the data model (11), we have

$$\begin{aligned} \text{vec}(\mathbf{X}) &= \text{vec}(\mathbf{AS}) + \text{vec}(\mathbf{W}) \\ &= (\mathbf{S}^T \otimes \mathbf{I}_M) \text{vec}(\mathbf{A}) + \text{vec}(\mathbf{W}). \end{aligned}$$

Let $\mathbf{x} = \text{vec}(\mathbf{X})$, $\mathbf{a} = \text{vec}(\mathbf{A})$ and $\mathbf{\Gamma} = \mathbf{S}^T \otimes \mathbf{I}_M$. The pdf $p(\mathbf{x} | \mathbf{a})$ will follow the same distribution as the noise, but with mean $\mathbf{\Gamma}\mathbf{a}$. We then have:

$$p(\mathbf{x} | \mathbf{a}) = \beta \exp\left(-\frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{\Gamma}\mathbf{a})^T (\mathbf{x} - \mathbf{\Gamma}\mathbf{a})\right)$$

with $\beta = ((2\pi)^{\frac{MN}{2}} \sigma^{MN})^{-1}$. We also know that \mathbf{a} denotes a zero-mean Gaussian vector with covariance matrix $\mathbf{R}_a = \Psi \otimes \Sigma$. This means that the pdf $p(\mathbf{a})$ can be expressed in the following way:

$$p(\mathbf{a}) = \rho \exp\left(-\frac{1}{2} \mathbf{a}^T (\Psi^{-1} \otimes \Sigma^{-1}) \mathbf{a}\right). \quad (26)$$

with $\rho = (2\pi)^{-\frac{M^2}{2}} (\det(\Psi \otimes \Sigma))^{-\frac{1}{2}}$. From [19], we can see that the Fisher information matrix $\mathbf{J}_T(\mathbf{a})$ of the parameter to be estimated (in this case, $\mathbf{a} = \text{vec}(\mathbf{A})$) can be obtained in the following way:

$$\mathbf{J}_T(\mathbf{a}) = \mathbf{J}_D(\mathbf{a}) + \mathbf{J}_P(\mathbf{a}).$$

The matrix $\mathbf{J}_D(\mathbf{a})$ represents the information obtained from the observations and is given by:

$$\begin{aligned} \mathbf{I}_D(\mathbf{a}) &= -\mathbb{E} \left[\nabla_{\mathbf{a}} \left(\left\{ \nabla_{\mathbf{a}} [\log p(\mathbf{x} | \mathbf{a})] \right\}^T \right) \right] \\ &= \frac{1}{\sigma^2} \left(\mathbf{S}\mathbf{S}^T \otimes \mathbf{I}_M \right). \end{aligned}$$

The matrix $\mathbf{J}_P(\mathbf{a})$ represents the information contained in the prior for \mathbf{A} . This matrix is given by

$$\begin{aligned} \mathbf{J}_P(\mathbf{a}) &= -\mathbb{E} \left[\nabla_{\mathbf{a}}^2 \log p(\mathbf{a}) \right] \\ &= (\Psi^{-1} \otimes \Sigma^{-1}). \end{aligned}$$

Thus, we know the matrix $\mathbf{J}_T(\mathbf{a})$. As is known, the Crámer-Rao inequality is given by

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \succeq \mathbf{J}_T^{-1}(\boldsymbol{\theta}), \quad (27)$$

where $\boldsymbol{\theta}$ corresponds to the vector of parameters to be estimated, $\hat{\boldsymbol{\theta}}$ represents an unbiased estimator for $\boldsymbol{\theta}$ and $\text{Cov}(\hat{\boldsymbol{\theta}}) = \mathbb{E} \left\{ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \right\}$ denotes the covariance matrix of the error $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$. From (27), we have

$$\begin{aligned} \text{tr} \left(\mathbb{E} \left\{ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \right\} \right) &\geq \text{tr}(\mathbf{J}_T^{-1}(\boldsymbol{\theta})) \\ \Leftrightarrow \mathbb{E} \left\{ \text{tr} \left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \right) \right\} &\geq \text{tr}(\mathbf{J}_T^{-1}(\boldsymbol{\theta})) \\ \Leftrightarrow \mathbb{E} \left\{ \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \right\} &\geq \text{tr}(\mathbf{J}_T^{-1}(\boldsymbol{\theta})). \end{aligned}$$

This expression gives us a minimum bound for the MSE of any unbiased estimator of $\boldsymbol{\theta}$. In our case, $\boldsymbol{\theta} = \mathbf{a}$ and given the Fisher information matrix $\mathbf{J}_T(\mathbf{a})$, we have

$$\mathbf{J}_T^{-1}(\mathbf{a}) = \left(\frac{1}{\sigma^2} (\mathbf{S}\mathbf{S}^T \otimes \mathbf{I}_M) + (\Psi^{-1} \otimes \Sigma^{-1}) \right)^{-1}.$$

and

$$\mathbb{E} \left\{ \|\hat{\mathbf{A}} - \mathbf{A}\|^2 \right\} \geq \text{tr}(\mathbf{J}_T^{-1}(\mathbf{a})), \quad (28)$$

which is the expression for the bound for the MSE used in the simulations.

In figures 5, 6 and 7 we plot the results of similar simulations, but allow for $I = 2$ loops of the iterative algorithm (16) and (17). As expected, allowing for more flops improves the

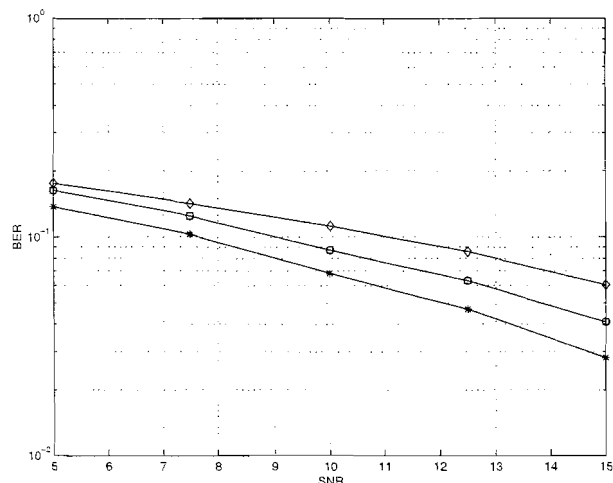


Figure 5. BER of user 1 versus SNR ($I = 2$ iterations)

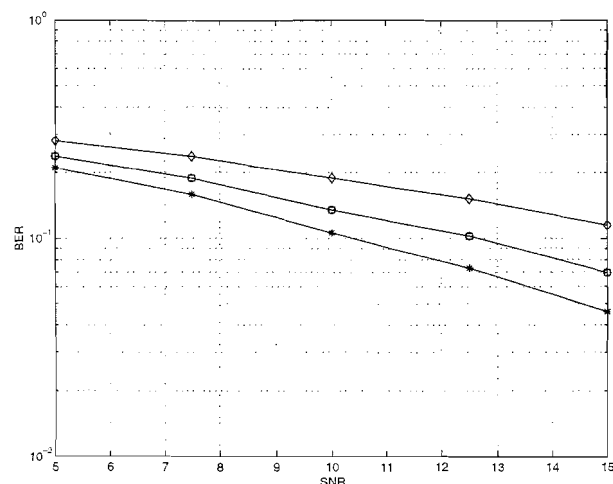


Figure 6. BER of user 2 versus SNR ($I = 2$ iterations)

BER for both users, irrespective of the initialization method. However, the random initialization is still outperformed by our approach over the entire range of SNRs simulated. Figures 8, 9 and 10 show the results corresponding to $I = 3$ iterations of the algorithm in (16) and (17). We can draw conclusions similar to the previous ones. We stress that in terms of performance, the LU factorization was identical to the PQ factorization.

4. TRAINED-BASED CHANNEL IDENTIFICATION

In this section, we discuss another application for the results in [12], namely the pdfs in (8) and (9). The data model is as in (11), and we assume the additive observation noise to have the same statistics as in section 3, see (10). Although not necessary for the method to be discussed, we also let the sources be binary and follow the same statistical characterization detailed in section 3. Moreover, we maintain the prior on the mixing matrix, $\mathbf{A} \sim \mathcal{N}(\mathbf{0}, \Sigma \otimes \Psi)$. We consider a trained-based channel identification scenario. We assume that P of the N emitted symbols by the M sources,

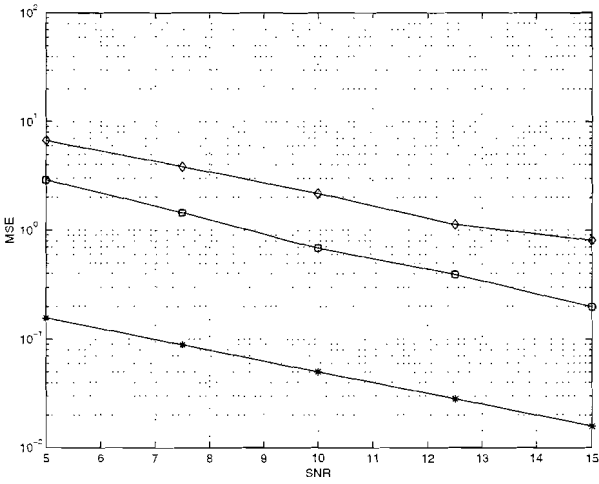


Figure 7. MSE of channel estimate versus SNR ($I = 2$ iterations)

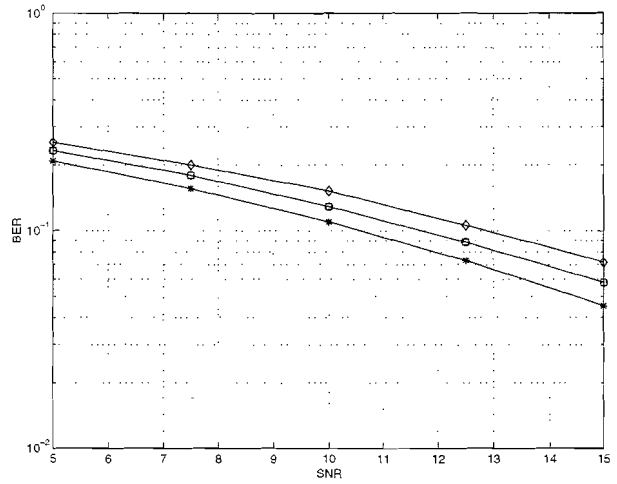


Figure 9. BER of user 2 versus SNR ($I = 3$ iterations)

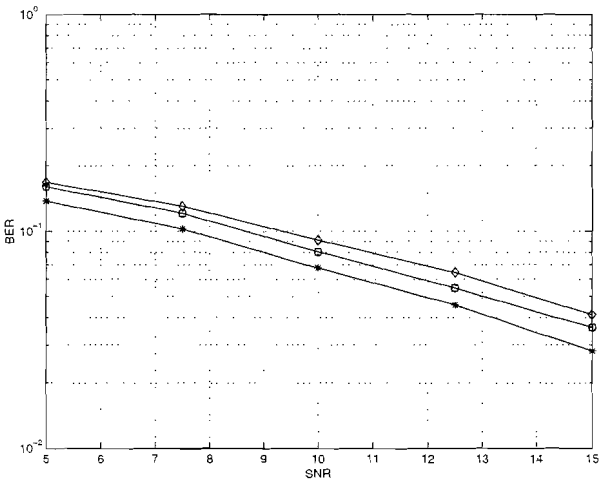


Figure 8. BER of user 1 versus SNR ($I = 3$ iterations)

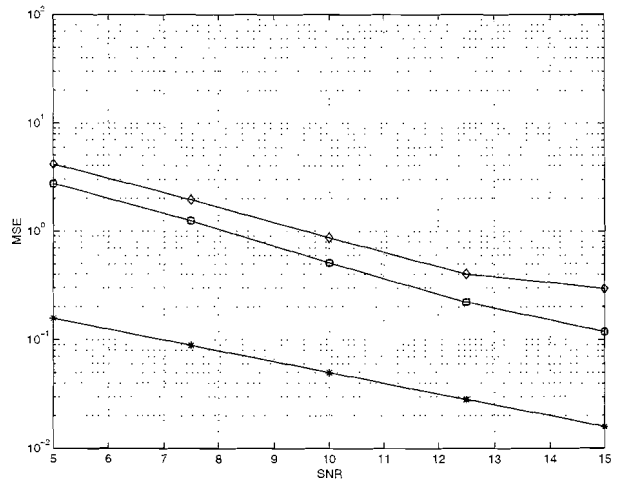


Figure 10. MSE of channel estimate versus SNR ($I = 3$ iterations)

say, the sources' packet header $\mathcal{S} = [s[1] s[2] \cdots s[P]]$, is known by the receiver. This preamble is included by the sources in order to assist the receiver in acquiring or estimating the channel. Once the channel \mathbf{A} is estimated, it can be used to decode the remaining information symbols in $s[P+1], s[P+2], \dots, s[N]$ from the observations $\mathbf{x}[P+1], \mathbf{x}[P+2], \dots, \mathbf{x}[N]$. A possible channel identification strategy is

$$\hat{\mathbf{A}}_{\text{MAP}} = \arg \max_{\mathbf{A} \in \text{GL}(M, \mathbb{R})} p(\mathbf{A} | \mathcal{X}), \quad (29)$$

where $\mathcal{X} = [\mathbf{x}[1] \mathbf{x}[2] \cdots \mathbf{x}[P]]$ denotes the observed packet header. That is, $\hat{\mathbf{A}}_{\text{MAP}}$ denotes the most probable channel realization given the available header of data observations. Notice that this approach does not take into account all the received data, only the header. It can be verified that processing all observations would lead to a computationally infeasible solution, for basically the same reasons exposed in section 3 (beyond the time instant $n = P$, the transmitted data is unknown, representing $2^{M(N-P)}$ bits, and the prior must be integrated against all possible source sequences).

This estimate of the channel matrix is obtained as follows. From Bayes' rule, we know that

$$\begin{aligned} \hat{\mathbf{A}}_{\text{MAP}} &= \arg \max_{\mathbf{A} \in \text{GL}(M, \mathbb{R})} p(\mathbf{A} | \mathcal{X}) \\ &= \arg \max_{\mathbf{A} \in \text{GL}(M, \mathbb{R})} p(\mathcal{X} | \mathbf{A})p(\mathbf{A}) \end{aligned}$$

We base ourselves on the data model

$$\mathcal{X} = \mathbf{A}\mathcal{S} + \mathcal{W}, \quad (30)$$

where $\mathbf{A} \sim \mathcal{N}(0, \Sigma \otimes \Psi)$ and $\mathcal{W} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{MP})$. This means that $\mathbf{a} = \text{vec}(\mathbf{A})$ and $\mathbf{w} = \text{vec}(\mathcal{W})$ are zero-mean Gaussian vectors with correlation matrices $\mathbf{R}_a = \Psi \otimes \Sigma$ e $\mathbf{R}_w = \sigma^2 \mathbf{I}_{MP}$, respectively. The pdf $p(\mathbf{a})$ can then be expressed as in (26). We now need to find the expression for $p(\mathbf{x} | \mathbf{A})$, where $\mathbf{x} = \text{vec}(\mathcal{X})$. Since the header of the data sent is known, that is, the matrix \mathcal{S} , and that the channel matrix is given, the random vector $\mathbf{x} | \mathbf{A}$ has a Gaussian distribution with mean $\boldsymbol{\mu} = \text{vec}(\mathbf{A}\mathcal{S})$ and covariance equal to the

noise's, that is, $\mathbf{R}_x = \mathbf{R}_w = \sigma^2 \mathbf{I}_{MP}$. We have

$$p(\mathbf{x} | \mathbf{A}) = \frac{1}{(2\pi)^{MP/2} \sigma^{MP}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \boldsymbol{\mu}\|^2\right)$$

Using these two pdfs and further straightforward calculations we come to the conclusion

$$\hat{\mathbf{A}}_{\text{MAP}} = \arg \max_{\mathbf{A} \in \text{GL}(M, \mathbb{R})} \phi(\mathbf{a})$$

where

$$\phi(\mathbf{a}) = -\frac{1}{2\sigma^2} \|\mathbf{x} - \boldsymbol{\mu}\|^2 - \frac{1}{2} \mathbf{a}^T (\boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{a}.$$

Since this function is concave, its maximum is attained at the stationary point $\nabla \phi(\mathbf{a}) = \mathbf{0}$. It can be easily verified that this stationary point corresponds to

$$\hat{\mathbf{a}}_{\text{MAP}} = \left(\boldsymbol{\Psi} \mathbf{S} \mathbf{S}^T \otimes \mathbf{I}_M + \mathbf{I}_M \otimes \sigma^2 \boldsymbol{\Sigma}^{-1} \right)^{-1} \text{vec} \left(\boldsymbol{\chi} \mathbf{S}^T \boldsymbol{\Psi} \right)$$

and this is the solution of (29).

We propose an alternative channel identification strategy, exploiting the SOS of the received data. Again, we present our results based only on the PQ factorization (the extension to the LU factorization is similar). Let $\hat{\mathbf{P}}$ denote the estimate of the \mathbf{P} factor of $\mathbf{A} = \mathbf{P}\mathbf{Q}$, computed from the received data \mathbf{X} as explained in section 3. We propose to estimate the channel as

$$\hat{\mathbf{A}} = \hat{\mathbf{P}} \hat{\mathbf{Q}}, \quad (31)$$

where

$$\hat{\mathbf{Q}} = \arg \max_{\mathbf{Q} \in \mathbb{O}(M)} p(\mathbf{Q} | \boldsymbol{\chi}, \mathbf{P} = \hat{\mathbf{P}}).$$

Thus, $\hat{\mathbf{Q}}$ denotes the most probable realization of the \mathbf{Q} factor of the mixing matrix \mathbf{A} , given that its \mathbf{P} factor is $\hat{\mathbf{P}}$ and the available packet header $\boldsymbol{\chi}$. This strategy makes the totality of the received data participate in the channel estimate, through its 2nd-order statistics. Using the Bayes rule and the identity in (8) yields

$$\hat{\mathbf{Q}} = \arg \min_{\mathbf{Q} \in \mathbb{O}(M)} f(\mathbf{Q}), \quad (32)$$

with

$$f(\mathbf{Q}) = \text{tr} \left(\mathbf{Q}^T \mathbf{A} \mathbf{Q} \mathbf{B} \right) + \sigma^2 \text{tr} \left(\mathbf{Q}^T \mathbf{C} \mathbf{Q} \mathbf{D} \right) - 2 \text{tr} \left(\boldsymbol{\varepsilon}^T \mathbf{Q} \right),$$

where

$$\begin{aligned} \mathbf{A} &= \hat{\mathbf{P}}^2, \\ \mathbf{B} &= \mathbf{S} \mathbf{S}^T, \\ \mathbf{C} &= \hat{\mathbf{P}} \boldsymbol{\Sigma}^{-1} \hat{\mathbf{P}}, \\ \mathbf{D} &= \boldsymbol{\Psi}^{-1}, \\ \boldsymbol{\varepsilon} &= \hat{\mathbf{P}} \boldsymbol{\chi} \mathbf{S}^T. \end{aligned}$$

Problem (32) does not afford, in general, a closed-form solution. However, due to the special structure of the constraints, it allows for efficient low-complexity solvers exploiting the curvature of the Lie group $\mathbb{O}(M)$.

We now present a locally convergent iterative algorithm which allows to solve (32). This algorithm is based on the

first order geodesic descent method. The structure of the algorithm for a generic optimization problem is the following. Consider the constrained optimization problem

$$\begin{aligned} \min f(\mathbf{x}) \\ \mathbf{x} \in \mathcal{M}, \end{aligned}$$

where \mathcal{M} is a manifold in the Euclidean space \mathbb{R}^n (a sphere, for example). If we are at a point $\mathbf{x}_k \in \mathcal{M}$, the usual gradient descent solution (for unconstrained problems) consists in searching for \mathbf{x}_{k+1} along the line starting at \mathbf{x}_k with direction $-\nabla f(\mathbf{x}_k)$. This means that in the traditional solution, we have $\mathbf{x}_{k+1} = \gamma(t^*)$, for some $t^* \geq 0$, where $\gamma(t) = \mathbf{x}_k - t \nabla f(\mathbf{x}_k)$. However, in the presence of constraints, such a solution is not, in general, feasible since $\mathbf{x}_{k+1} \notin \mathcal{M}$. We can, however, move along a curve $\gamma(t)$ which has the same direction as the negative gradient (projected onto the tangent plane to \mathcal{M} at the point \mathbf{x}_k) but stays on the manifold \mathcal{M} . This curve $\gamma(t)$ which we move along is a geodesic. A geodesic constitutes a generalization of straight lines to surfaces with curvature. It is a regular curve which (locally) minimizes the distance between two points belonging to one surface. In this algorithm, we calculate the geodesic curve which starts at \mathbf{x}_k with an initial direction identical to the negative gradient projected onto the tangent plane to \mathcal{M} at the point \mathbf{x}_k . We move along this geodesic to a new point which corresponds to a lower cost of f . Given a feasible point \mathbf{x}_k we can then resume the algorithm in the following steps:

- i) Calculate the projection \mathbf{d}_k of $-\nabla f(\mathbf{x}_k)$ onto the tangent plane to \mathcal{M} at \mathbf{x}_k ;
- ii) Determine the geodesic, $\gamma(t)$, $t \geq 0$, of the constraint surface which has $\gamma(0) = \mathbf{x}_k$ and $\frac{d}{dt} \gamma(0) = \mathbf{d}_k$;
- iii) Minimize $f(\gamma(t))$ with respect to $t \geq 0$, obtaining t^* and $\mathbf{x}_{k+1} = \gamma(t^*)$.

Figure 11 presents a geometric interpretation of the algorithm. More details about it can be found in [20].

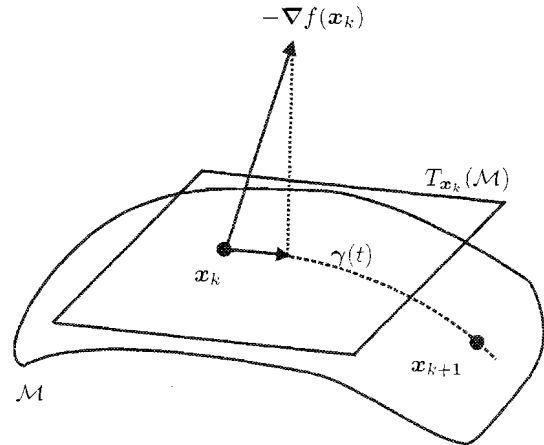


Figure 11. First order geodesic descent method.

We carried out some computer simulations to assess the accuracy of both channel identification strategies, *i.e.*, (29) and (31). The prior on the channel is unchanged, see (24)

and (25). We varied the SNR between $\text{SNR}_{\min} = 0$ dB and $\text{SNR}_{\max} = 20$ dB in steps of $\Delta = 2.5$ dB. For each SNR, 5000 statistically independent Monte-Carlo runs were performed. Each Monte-Carlo involves realizing \mathbf{A} , \mathbf{S} and \mathbf{W} (11). The packet length is $N = 200$, and we assume that the training header has length $P = 8$. Both channel estimators $\hat{\mathbf{A}}_{\text{MAP}}$ (29) and $\hat{\mathbf{A}}$ (31) are implemented, and the respective squared channel errors $\|\hat{\mathbf{A}}_{\text{MAP}} - \mathbf{A}\|^2$ and $\|\hat{\mathbf{A}} - \mathbf{A}\|^2$ are computed. Figure 12 shows the mean-square errors (MSE) obtained for both channel estimates. The solid

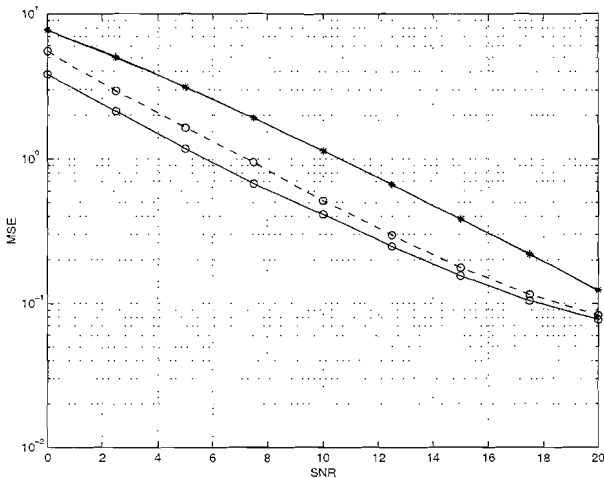


Figure 12. MSE of channel estimate versus SNR

line with plus signs corresponds to the non-SOS based estimator. This line is coincidental with the solid line with stars which corresponds to a bound calculated considering that we know the header \mathbf{S} . The expression for this bound is obtained using the same theory behind (28), only now we have

$$\mathbf{J}_T^{-1}(\mathbf{a}) = \left(\frac{1}{\sigma^2} (\mathbf{S}\mathbf{S}^T \otimes \mathbf{I}_M) + (\Psi^{-1} \otimes \Sigma^{-1}) \right)^{-1}.$$

The two curves coincide because (29) is precisely the estimator that attains this bound. The solid line with circles corresponds to the SOS-based iterative algorithm initialized near the optimum. The dashed line with circles corresponds to the SOS-based iterative algorithm for estimating \mathbf{A} initialized under the assumption that the SNR is not low. If this is so, then (32) can be reduced to

$$\hat{\mathbf{Q}}^{(0)} = \arg \min_{\mathbf{Q} \in \mathbb{O}(M)} \text{tr}(\mathbf{Q}^T \hat{\mathbf{P}}^2 \mathbf{Q} \mathbf{S} \mathbf{S}^T) - 2 \text{tr}(\mathbf{S} \mathbf{X}^T \hat{\mathbf{P}} \mathbf{Q}). \quad (33)$$

Supposing the headers used by each transmitter are orthogonal to each other, we have $\mathbf{S}\mathbf{S}^T = \mathbf{I}_M$. This reduces (33) to

$$\hat{\mathbf{Q}}^{(0)} = \arg \max_{\mathbf{Q} \in \mathbb{O}(M)} \text{tr}(\mathbf{Q}^T \hat{\mathbf{P}} \mathbf{X} \mathbf{S}^T), \quad (34)$$

which has a closed form solution. Let $\hat{\mathbf{P}} \mathbf{X} \mathbf{S}^T = \mathbf{U} \mathbf{S} \mathbf{V}^T$ be a singular value decomposition of $\hat{\mathbf{P}} \mathbf{X} \mathbf{S}^T$. The solution to (34) is $\hat{\mathbf{Q}}^{(0)} = \mathbf{U} \mathbf{V}^T$. We can see that the SOS-based estimator achieves the best performance over all the SNRs

considered. The initialization proposed in (34) is also close to the optimum. Figure 13 shows the same results but for $M = 3$ users. The computer simulations were conducted

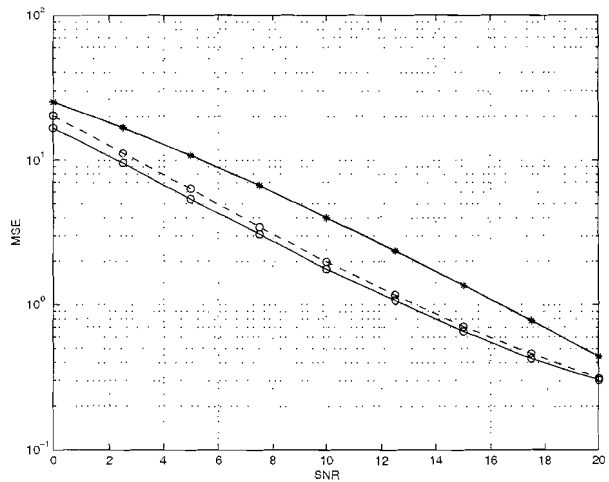


Figure 13. MSE of channel estimate versus SNR

under the same conditions except that the matrices Ψ and Σ were taken to be

$$\Psi = \mathbf{G} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{G}^T$$

and

$$\Sigma = \mathbf{H} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{H}^T$$

with \mathbf{G} and $\mathbf{H} \in \mathbb{O}(M)$ randomly generated. We can make the same conclusion as before, although the MSE increases for all estimates. This is due to the fact that we are estimating more variables based on the same amount of samples.

5. CONCLUSION

Two distinct problems were discussed in which the use of 2nd order statistics (SOS) was applied: blind source separation and trained-based channel identification. In the first case, the emphasis was set upon the initialization of a locally convergent iterative algorithm. The solution proposed by us outperforms a standard solution, the comparison being made by running computer simulations. In the second case, we created a locally convergent iterative algorithm which when initialized with the application of SOS also outperforms a non-SOS based solution. Thus, in both cases, by solving the channel matrix up to an unknown orthogonal factor, the devised methods improve the performance of non-SOS based estimation schemes. In future work, we expect to study cases in which the channel matrix is not square and it bears complex entries.

REFERENCES

- [1] J. F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, October 1998.

- [2] A. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1136–1155, May 1996.
- [3] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array – part I: algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1184–1197, May 1996.
- [4] V. Barroso, J. M. F. Moura, and J. Xavier, "Blind array channel division multiple access (AChDMA) for mobile communications," *IEEE Transactions on Signal Processing*, vol. 46, pp. 737–752, March 1998.
- [5] J. Xavier and V. Barroso, "Blind source separation, ISI cancellation and carrier phase recovery in SDMA systems for mobile communications," *Wireless Personal Communications*, special issue on *Wireless Broadband Communications*, vol. 10, pp. 36–76, Kluwer Academic Publishers, June 1999.
- [6] J. Xavier, V. Barroso, and José M. F. Moura, "Closed-form correlative coding (CFC2) blind identification of MIMO channels: isometry fitting to second order statistics," *IEEE Transactions on Signal Processing*, vol. 49, no. 5, pp. 1073–1086, May 2001.
- [7] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proc.-F, Radar & Signal Processing*, vol. 140, pp. 362–370, December 1993.
- [8] A. Belouchrani, K. Abed-Meraim, J-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, February 1997.
- [9] J. Xavier and V. Barroso, "Blind SDMA with a signal space geometry matched algorithm," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)*, vol. 5, pp. 3437–3440, Munich, Germany, April 1997.
- [10] J. Xavier and V. Barroso, "Semi-blind maximum likelihood separation of linear convolutive mixtures," in *Proceedings of the 10th IEEE Workshop on Statistical Signal and Array Processing*, Pennsylvania, August 2000.
- [11] C. Papadias, "Blind source separation based on multi-user kurtosis criteria," on source correlative filtering," in *Proceedings of 2000 International Symposium on Information Theory, ISIT'2000*, Sorrento, Italy, June 2000.
- [12] J. Xavier, V. Barroso, Paulo Lopes, and Tiago Patrão, "Prewhitening of Mixing Matrices: Impact on the Contraction of Their Probability Mass Priors," in *Proceedings of 2002 IEEE International Symposium on Information Theory (ISIT'02)*, Lausanne, 2002.
- [13] J. Xavier, V. Barroso, and P. Lopes, "Blind identification of orthogonal mixing matrices under a Bayesian framework: algorithms and performance analysis," in preparation.
- [14] I. Telatar, "Capacity of multi-antenna Gaussian channels," Technical Memorandum, AT&T Bell Laboratories, 1995.
- [15] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Chapman & Hall, CRC.
- [16] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press.
- [17] W. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. New York: Academic Press.
- [18] Manfredo P. Carmo. *Geometria Riemanniana*, 2nd ed., Instituto de Matemática Pura e Aplicada.
- [19] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory : PART I*, John Wiley and Sons.
- [20] A. Edelman, T. A. Arias, S. T. Smith. *The Geometry of Algorithms with Orthogonality Constraints*. SIAM J. Matrix Anal. Appl., Vol. 20 (1998), No. 2, pp. 303-353.

Paulo Lopes was born in Lisbon, Portugal, on June 6, 1979. He received the E.C.E. degree in 2002 from the Instituto Superior Técnico (IST), Lisbon, Portugal. Presently, he is pursuing the M.Sc. degree at IST.

In 2001 he joined the Signal and Image Processing Group of the Institute for Systems and Robotics (ISR). His current research interests are in the area of statistical signal processing with applications to mobile wireless digital communications.

João Xavier (S'97) received the E.C.E, the M.Sc., and the Ph.D. degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), Lisbon, Portugal, respectively in 1995, 1997 and 2002.

In 1997, he joined the faculty of the Department of Electrical and Computer Engineering at IST, where he is with the Signal and Image Processing Group of the Institute for Systems and Robotics (ISR). His current research interests are in the area of statistical signal processing with applications to mobile wireless digital communications.

Victor A. N. Barroso (IEEE M'89-SM'00) was born in Lisbon, Portugal, on December 22, 1952. He received the E.E. degree in 1976 and the Ph.D. degree in electrical and computer engineering in 1990, both from Instituto Superior Técnico (IST), Lisbon, Portugal.

In 1976, he joined the faculty of the Department of Electrical and Computer Engineering at IST. He was promoted to Assistant Professor in 1990 and to Associate Professor in 1994. He has taught courses in systems and signal theory, control systems, signal and array processing, computer networks, and communications. Currently, he is the Vice President of IST's Scientific Council for research and post-graduation activities. He is also a Researcher at the Institute for Systems and Robotics (ISR), Lisbon, Portugal, where he is the coordinator of the Signal and Image Processing Group. His research interests are in statistical signal/image and array processing, wireless communications, and underwater acoustics. He has published over 80 journal and conference papers.

Dr. Barroso was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1996 to 1999 and the representative of the IEEE Signal Processing Society at the IEEE Wireless Periodicals Working Group in 2000-01. Presently he is Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, and member of the SENSOR ARRAY AND MULTICHANNEL Technical Committee and of the steering committee of the IEEE TRANSACTIONS ON MOBILE COMPUTING. Dr. Barroso is the Chair of the Portuguese IEEE Signal Processing Chapter. He is affiliated with several IEEE societies.