

# Blind Source Separation and Channel Identification: Exploiting 2nd-Order Statistics in Bayesian Frameworks

João Xavier, Victor Barroso, Paulo Lopes

Instituto de Sistemas e Robótica – Instituto Superior Técnico, Portugal

*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.

## I. 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 (non-whitened) 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 II, 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 III, 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 IV, 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 V 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 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.

## II. DATA MODEL AND PREVIOUS WORK

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

$$x[n] = As[n] + w[n], \quad (1)$$

where  $x[n] = (x_1[n], \dots, x_M[n])^T$  represents the  $M$ -dimensional vector of observations,  $A \in \mathbb{R}^{M \times M}$  stands for the mixing matrix,  $s[n] = (s_1[n], \dots, s_M[n])^T$  contains the  $M$  source signals, and  $w[n] = (w_1[n], \dots, w_M[n])^T$  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,  $R_s = \mathbb{E}\{s[n]s[n]^T\} = I_M$ , and the mixing matrix is nonsingular,  $A \in \mathbb{GL}(M, \mathbb{R})$ . We assume that the process  $w[n]$  is zero-mean and wide-sense stationary with known correlation matrix  $R_w[0] = \mathbb{E}\{w[n]w[n]^T\}$ . It is well known that the 2nd order statistics of the observations can be exploited to partially solve for the unknown mixing matrix  $A$ . Here, we consider two alternative methods based on the polar and the  $LU$  decomposition of  $A$ , respectively. Both methods act on the so-called de-noised correlation matrix of  $x[n]$ ,

$$R = R_x[0] - R_w[0] = AA^T, \quad (2)$$

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

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

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

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

$$x[n] = PQs[n] + w[n] \quad (4)$$

or

$$x[n] = LUs[n] + w[n], \quad (5)$$

with  $P$  or  $L$  known, respectively. If the original mixing matrix has a prior,  $A \sim p(A)$ , then the unknown residual orthogonal matrices  $Q$  and  $U$  in (4) and (5), respectively, denote random objects. In [12], we investigated the distributions of  $Q$  or

$U$  for a given prior on  $A$ . 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.,  $\iota : \mathbb{O}(M) \rightarrow \mathbb{GL}(M, \mathbb{R})$ ,  $\iota(X) = X$ . Here,  $\mathbb{GL}(M, \mathbb{R})$  is identified with an open subset of the Euclidean space  $\mathbb{R}^{M^2}$  by interpreting a  $M \times M$  matrix as a  $M^2$ -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  $\Omega$ , 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  $\Omega$  over  $\mathbb{O}(M)$ , we have  $\Omega = f \Omega_{\mathbb{O}(M)}$  for a unique nonnegative smooth function  $f : \mathbb{O}(M) \rightarrow \mathbb{R}$ , we use the terminology *mass distribution* for either  $\Omega$  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  $p(A)$  denote the pdf (prior) on the mixing matrix  $A \in \mathbb{GL}(M, \mathbb{R})$ . Then, the factorization  $A = PQ$  induces the pdf on  $\mathbb{P}(M) \times \mathbb{O}(M)$  given, up to a normalizing constant, by

$$p(P, Q) = p(PQ)g(P), \quad (6)$$

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

$$p(L, U) = p(LU)h(L), \quad (7)$$

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

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

In this paper, we restrict ourselves to the scenario where  $A$  has a zero-mean matrix variate normal distribution with covariance matrix  $I_M \otimes \Psi$ , denoted  $A \sim \mathcal{N}(0, I_M \otimes \Psi)$ , where  $\Psi \in \mathbb{P}(M)$ , see [15]. This means that  $A \stackrel{d}{=} Z\Psi^{1/2}$  (equality in distribution), where  $Z$  denotes an  $M \times M$  random matrix whose entries are independent and identically distributed as zero-mean unit-variance Gaussian random variables. We notice that our assumption does not represent a restriction with respect to the

more general case where  $A \sim \mathcal{N}(0, \Sigma \otimes \Psi)$ ,  $\Sigma \in \mathbb{P}(M)$ , corresponding to  $A \stackrel{d}{=} \Sigma^{1/2} Z \Psi^{1/2}$ , because we can revert to our situation by pre-multiplying the observations  $x[n]$  in (1) with  $\Sigma^{-1/2}$ . The assumption  $A \sim \mathcal{N}(0, I_M \otimes \Psi)$  means that  $A$  is distributed over  $\mathbb{GL}(M, \mathbb{R})$  according to

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

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

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

and

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

respectively.

### III. 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  $A \sim \mathcal{N}(0, I_M \otimes \Psi)$ , where the dispersion matrix  $\Psi$  is assumed known at the receiver. Moreover, we let  $w[n]$  denote zero-mean, spatio-temporal white Gaussian distributed noise with known power  $\sigma^2$ , *i.e.*,

$$R_w[k] = \mathbb{E} \{w[n]w[n-k]^T\} = \sigma^2 I_M \delta[k], \quad (10)$$

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

$$X = AS + W, \quad (11)$$

where  $X = [x[1] \ x[2] \ \dots \ x[N]]$  denotes the data matrix containing the observations,  $S = [s[1] \ s[2] \ \dots \ s[N]]$  contains the information sequences sent by the sources (the  $m$ th row corresponds to the  $m$ th source) and  $W = [w[1] \ w[2] \ \dots \ w[N]]$  stands for the additive noise matrix. We are interested in estimating the binary matrix  $S$  in (11) from the available data matrix  $X$ , without knowing the mixing channel matrix  $A$ . A natural approach in this Bayesian setting is to look for the maximum a posteriori (MAP) estimates of the transmitted bits,

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

where  $\mathcal{B}_{M \times N}$  stands for the discrete set of  $M \times N$  binary matrices. With the prior  $A \sim \mathcal{N}(0, I_M \otimes \Psi)$ , it is easily seen (after

some straightforward computations) that problem (12) is equivalent to

$$\widehat{S}_{\text{MAP}} = \arg \max_{S \in \mathcal{B}_{M \times N}} \frac{1}{\sigma^2} \operatorname{tr}(XS^T \Delta_S^{-1} SX^T) - M \log(\det \Delta_S), \quad (13)$$

where

$$\Delta_S = SS^T + \sigma^2 \Psi^{-1}. \quad (14)$$

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.*,

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

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

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

for  $k = 0, 1, 2, \dots$  until a fixed-point is attained. Given our statistical assumptions, we have, after some calculus, the iterates

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

$$A^{(k+1)} = X S^{(k+1)T} \Delta_{S^{(k+1)}}^{-1}. \quad (17)$$

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

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

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

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

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  $A^{(0)}$  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  $A$  as explained in

section II. Namely, consider the sample-mean estimate of the correlation matrix of the observed data  $x[n]$  given in (3). The denoised correlation matrix  $R$  in (2) is estimated as

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

where  $\beta$  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  $R_w[0] = \sigma^2 I_M$  from  $\widehat{R}_x[0]$  as equation (2) suggests, because, for finite datasets ( $N < \infty$ ), that method does not guarantee a positive-definite matrix  $\widehat{R}$  (which is essential for the remaining processing). Let

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

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

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

where

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

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

$$\widehat{Q} = \arg \min_{Q \in \mathbb{O}(M)} \text{tr} \left( Q \Psi^{-1} Q^T \widehat{P}^2 \right). \quad (23)$$

A closed-form solution for (23) is available [16], and can be computed as follows. Let  $\Psi = Z D Z^T$  denote an eigendecomposition of  $\Psi$ , where  $Z \in \mathbb{O}(M)$  and  $D = \text{diag}(d_1, d_2, \dots, d_M)$  denotes a diagonal matrix with its diagonal entries sorted in decreasing order,  $d_1 \geq d_2 \geq \dots \geq d_M$ . Then,  $\widehat{Q} = V Z^T$ . Notice that the determination of  $\widehat{Q}$  does not involve any significant extra computational burden:  $V$  is already available from the step determining  $\widehat{P}$ , see (20), and  $Z$  can be computed off-line (it does not depend on the received data, only on  $\Psi$ ). As a final remark, perhaps a more defensible choice for  $A^{(0)}$  would be

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

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

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

Certainly, both these approaches incorporate more information, in fact, all the available data  $X$ , than our simple method in (21), which makes use of only the SOS of the observations. However, it 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  $A$  is  $\mathcal{N}(0, I_M \otimes \Psi)$ , where

$$\Psi = \begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix}. \quad (24)$$

This models a scenario where one user strongly dominates the other, in terms of received power (the channel is not well conditioned). We varied the signal-to-noise ratio (SNR) from  $\text{SNR}_{\min} = 5$  dB to  $\text{SNR}_{\max} = 20$  dB, in steps of  $\Delta = 2.5$  dB. The SNR is defined as  $\text{SNR} = \text{E} \|A s[n]\|^2 / \text{E} \|w[n]\|^2 = \|A\|^2 / M \sigma^2$ . For each SNR, 5000 statistically independent Monte-Carlo runs were performed. Each Monte-Carlo run consists in generating a realization of  $A$ ,  $S$  and  $W$ , 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  $A^{(0)}$  in (21). For comparison, we also performed  $I = 1$  iteration starting from a random initialization  $A^{(0)} \sim \mathcal{N}(0, I_M \otimes \Psi)$ , that is, an independent realization of the channel model. Figures 1 and 2 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 denotes

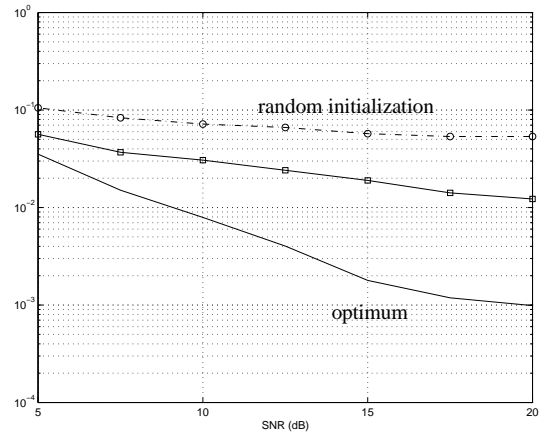


Fig. 1. BER of user 1 versus SNR ( $I = 1$  iteration)

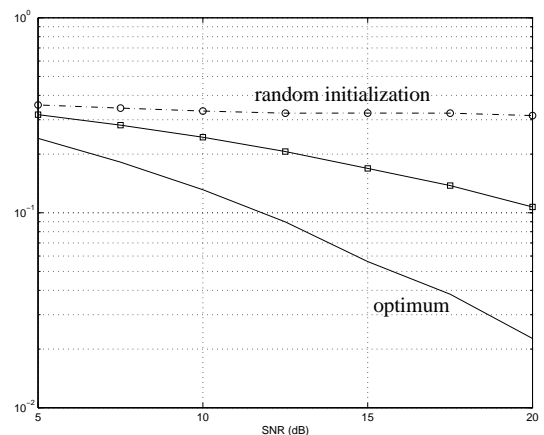


Fig. 2. BER of user 2 versus SNR ( $I = 1$  iteration)

a bound (maximum likelihood bit decoding with the channel  $A$

known), the solid line with squares refers to our proposed initialization, and the dashed lined with circles corresponds to the random initialization. As can be seen, our educated guess permits to outperform the random initialization. In figures 3 and 4, we plot the results of similar simulations, but allow for  $I = 2$  loops of the iterative algorithm (16) and (17). As expected,

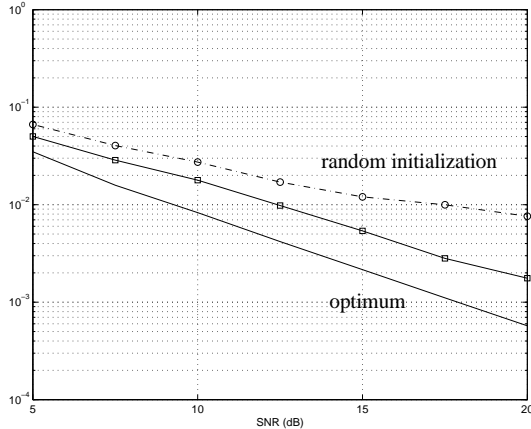


Fig. 3. BER of user 1 versus SNR ( $I = 2$  iterations)

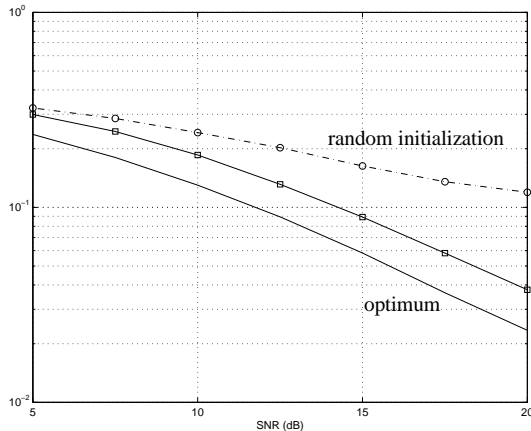


Fig. 4. BER of user 2 versus SNR ( $I = 2$  iterations)

allowing for more flops improves the 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 5 and 6 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 conducted a set of similar computer simulations, but using the LU factorization method. The performance was identical to the PQ factorization.

#### IV. 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 III, 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 III. Moreover, we maintain the prior on the

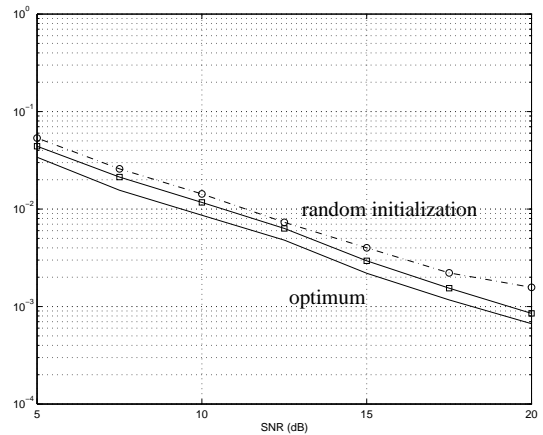


Fig. 5. BER of user 1 versus SNR ( $I = 3$  iterations)

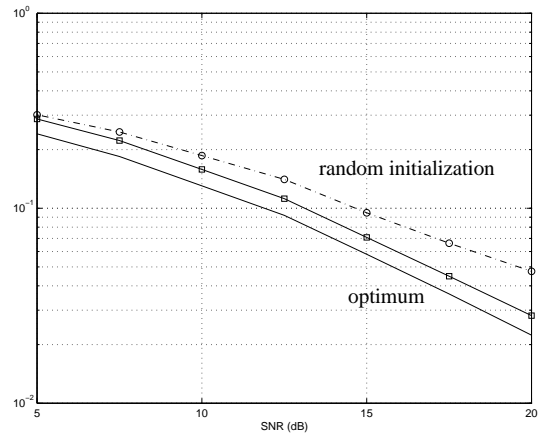


Fig. 6. BER of user 2 versus SNR ( $I = 3$  iterations)

mixing matrix,  $A \sim \mathcal{N}(0, I_M \otimes \Psi)$ . We consider a trained-based channel identification scenario. We assume that  $P$  of the  $N$  emitted symbols by the  $M$  sources, say, the sources' packet header  $S = [s[1] s[2] \dots s[P]]$ , 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  $A$  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  $x[P+1], x[P+2], \dots, x[N]$ . A possible channel identification strategy is

$$\hat{A}_{\text{MAP}} = \arg \max_{A \in \mathbb{GL}(M, \mathbb{R})} p(A | \mathcal{X}), \quad (25)$$

where  $\mathcal{X} = [x[1] x[2] \dots x[P]]$  denotes the observed packet header. That is,  $\hat{A}_{\text{MAP}}$  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 III (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). It is straightforward to check that, under our statistical framework, we have  $\hat{A}_{\text{MAP}} = \mathcal{X}^S T \Delta_S^{-1}$ , where  $\Delta_S$  is defined in (14). 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  $\hat{P}$  denote the estimate of the  $P$  factor of  $A = PQ$ , computed from the received data  $X$  as explained in section III. We propose to estimate the channel as

$$\hat{A} = \hat{P}\hat{Q}, \quad (26)$$

where

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

Thus,  $\hat{Q}$  denotes the most probable realization of the  $Q$  factor of the mixing matrix  $A$ , given that its  $P$  factor is  $\hat{P}$  and the available packet header  $\mathcal{X}$ . 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

$$\hat{Q} = \arg \min_{Q \in \mathbb{O}(M)} \text{tr} \left( Q^T \hat{P}^2 Q \Delta_S \right) - 2 \text{tr} \left( Q^T \hat{P} \mathcal{X} S^T \right). \quad (27)$$

Problem (27) 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  $\mathbb{O}(M)$ . These algorithms are beyond the scope of this paper and are discussed in [13].

We carried out some computer simulations to assess the accuracy of both channel identification strategies, *i.e.*, (25) and (26). The prior on the channel is unchanged, see (24). We varied the SNR between  $\text{SNR}_{\min} = 0$  dB and  $\text{SNR}_{\max} = 15$  dB in steps of  $\Delta = 2.5$  dB. For each SNR, 1000 statistically independent Monte-Carlo runs were performed. Each Monte-Carlo involves realizing  $A$ ,  $S$  and  $W$  (11). The packet length is  $N = 200$ , and we assume that the training header has length  $P = 10$ . Both channel estimators  $\hat{A}_{\text{MAP}}$  (25) and  $\hat{A}$  (26) are implemented, and the respective squared channel errors  $\|\hat{A}_{\text{MAP}} - A\|^2$  and  $\|\hat{A} - A\|^2$  are computed. Figure 7 shows the mean-square errors (MSE) obtained for both channel estimates. The dashed line

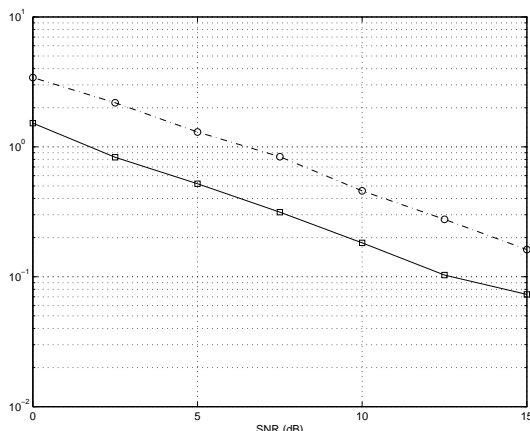


Fig. 7. MSE of channel estimate: non-SOS (dashed) and SOS (solid)

with circles refer to the non 2nd order statistics based channel

estimator  $\hat{A}_{\text{MAP}}$ , while the solid line corresponds to the SOS-based channel estimator  $\hat{A}$ . We can see that the SOS based estimator achieves the best performance overall the SNRSs considered.

## V. CONCLUSIONS

We study how 2nd order statistics (SOS) can be exploited in Bayesian setups for improving the performance of non-SOS based estimators. We addressed two problems: blind separation of co-channel binary sources and multi-user channel identification with training sequences. A prior is assumed on the mixing channel matrix. The SOS of the observations convey information about the unknown underlying channel. They permit to resolve the channel, modulo an orthogonal ambiguity factor, which becomes a random object under the Bayesian framework. We exploited the distribution of this residual mixing matrix for improving the performance of non-SOS based estimation schemes.

## 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," to appear in *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.