EXPLOITING 2ND-ORDER STATISTICS IN BAYESIAN SIGNAL RECONSTRUCTION PROBLEMS

Paulo Lopes, João Xavier, Victor Barroso

Instituto de Sistemas e Robótica - Instituto Superior Técnico, Portugal

ABSTRACT

We study signal reconstruction with the aid of second order statistics (SOS) in some Bayesian multi-user setups. Under a Bayesian framework, the unknown mixing channel matrix is a random object. Furthermore, it is well known that the SOS of the received data can solve the channel matrix up to an orthogonal factor, which now becomes a random object. That is, the prior for the original channel matrix contracts to another prior for the residual orthogonal mixing matrix. This paper shows how to exploit this new prior over the orthogonal group in two different applications: blind separation of binary users and semi-blind channel identification. Computer simulations assess the benefits of our proposed solutions over some classical "textbook" solutions.

1. INTRODUCTION

In the past years, blind source separation (BSS) has been subject to intense research activity [1]. A possible application is found in Space Division Multiple Access (SDMA) networks for wireless communications. When the mixing space-time channel is unknown at the receiver, BSS techniques are needed to reconstruct each of the transmitted signals [2, 3, 4, 5]. Commonly, the first step to take is to use the SOS of the received data to solve the unknown channel matrix up to an orthogonal factor [1, 6, 7]. Some alternatives which solve for the residual mixing matrix can be found in [2, 3, 5, 6, 7]. These methods, however, are not based on Bayesian frameworks.

Contribution. Under a Bayesian framework, the original channel matrix is modeled as a random object with a known pdf. Still, the SOS can only solve the unknown mixing matrix up to an orthogonal factor. But, now, the unknown orthogonal factor obeys a statistical model: the original prior contracts to a pdf over the group of orthogonal matrices. We show how to exploit this information to improve the performance of some BSS techniques.

Paper organization. In section 2, we introduce the data model and discuss the choice of the prior pdf for the channel matrix. We analyze how this prior contracts to a pdf in the group of orthogonal matrices, through two distinct channel pre-whitening methods: PQ (polar) and LU. In section 3, we study an application for the pdfs obtained in section 2 based on blind source separation of binary users. In section 4, we explore another application, semi-blind channel identification. Section 5 concludes our paper.

All Authors are with the Instituto de Sistemas e Robótica, Instituto Superior Técnico, Torre Norte, Piso 7, Av. Rovisco Pais, 1049-001, Fax: +351 21 841 8291. E-mails: {jxavier,vab}@isr.ist.utl.pt. paulomiguel@yahoo.com. 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

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 (boldface) and vectors in lowercase (boldface). The symbols $(\cdot)^T$, $\det(\cdot)$, $\otimes$ and $\boldsymbol{I}_n$ denote the transpose operator, the determinant, the Kronecker product and the $n \times n$ identity matrix, respectively. The notation $\mathbb{GL}(n, \mathbb{R})$, $\mathbb{O}(n) = \left\{\boldsymbol{Q} : \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_n\right\}$ 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 ASSUMPTIONS

We work under the following data model for an instantaneous mixture of M signals:

$$\boldsymbol{x}[n] = \boldsymbol{A}\boldsymbol{s}[n] + \boldsymbol{w}[n]. \tag{1}$$

Here, $\boldsymbol{x}[n] = (x_1[n], x_2[n], \ldots, x_M[n])^T$ represents the vector of observations, $\boldsymbol{A}$ is the $M \times M$ unknown mixing matrix, $\boldsymbol{s}[n] = (s_1[n], s_2[n], \ldots, s_M[n])^T$ is the vector of M source signals and $\boldsymbol{w}[n]$ denotes additive observation noise. We also take the following standard assumptions. The channel matrix $\boldsymbol{A}$ is assumed to be invertible, i.e., $\boldsymbol{A} \in \mathbb{GL}(M, \mathbb{R})$. We assume that $\boldsymbol{s}[n]$ denotes an uncorrelated wide sense stationary process, with a correlation matrix given by

$$\boldsymbol{R}_{\boldsymbol{s}}[k] = \mathrm{E}\left\{\boldsymbol{s}[n]\boldsymbol{s}[n-k]^T\right\} = \boldsymbol{I}_M\delta[k], \tag{2}$$

where $\delta[k]$ denotes the discrete time Kronecker delta. We also assume that the noise correlation matrix $\boldsymbol{R}_{\boldsymbol{w}} = \mathrm{E}\left\{\boldsymbol{w}[n]\boldsymbol{w}[n]^T\right\}$ is known. Under a Bayesian framework, $\boldsymbol{A}$ is regarded as a random object with an associated pdf reflecting our a priori knowledge of the mixing channel.

It is well known that the SOS of $\boldsymbol{x}[n]$ partially solve the channel matrix $\boldsymbol{A}$ through a prewhitening mechanism. Consider the denoised correlation matrix of the observations:

$$\boldsymbol{R} = \boldsymbol{R}_{\boldsymbol{x}} - \boldsymbol{R}_{\boldsymbol{w}} = \mathrm{E}\left\{\boldsymbol{x}[n]\boldsymbol{x}[n]^T\right\} - \boldsymbol{R}_{\boldsymbol{w}} = \boldsymbol{A}\boldsymbol{A}^T. \tag{3}$$

Notice that $\boldsymbol{R}$ is available because $\boldsymbol{R}_{\boldsymbol{x}}$ can be estimated through its sample mean estimator

$$\boldsymbol{R}_{\boldsymbol{x}}[0] = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}[n]\boldsymbol{x}[n]^T \tag{4}$$

and $\boldsymbol{R}_{\boldsymbol{w}}$ was assumed to be known.

Throughout the paper, we consider pre-whitening mechanisms based on two distinct matricial decompositions. The first is given by the PQ (polar) decomposition of $\boldsymbol{A}$. Write $\boldsymbol{A} = \boldsymbol{PQ}$, where $\boldsymbol{P} \in \mathbb{P}(M)$ and $\boldsymbol{Q} \in \mathbb{O}(M)$. Using (3), we have $\boldsymbol{R} = \boldsymbol{P}^2$. Thus, $\boldsymbol{R}$ reveals $\boldsymbol{P}$. The other decomposition is the LU decomposition. Write $\boldsymbol{A} = \boldsymbol{LU}$, where $\boldsymbol{L} \in \mathbb{L}(M)$ and $\boldsymbol{U} \in \mathbb{O}(M)$. We have $\boldsymbol{R} = \boldsymbol{LL}^T$ and $\boldsymbol{L}$ is revealed through $\boldsymbol{R}$ as its Cholesky factor. An interpretation for this algebra is based on the degrees of freedom contained in the problem. The channel $\boldsymbol{A}$ has $\dim \mathbb{GL}(M, \mathbb{R}) = M^2$ unknowns. Through either the PQ or LU decomposition, exactly $\dim \mathbb{P}(M) = \dim \mathbb{L}(M) = M(M+1)/2$ unknowns are revealed by $\boldsymbol{R}$. In either case, $\dim \mathbb{O}(M) = M(M-1)/2$ unknowns are left unsolved. But, since in a Bayesian setup $\boldsymbol{A}$ is random, both the missing factors $\boldsymbol{Q}$ or $\boldsymbol{U}$ are random also. It is important to know their associated pdfs for subsequent optimum signal processing. From particular results in [8], it can be shown that

$$p(\boldsymbol{P}, \boldsymbol{Q}) = \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{Q}\boldsymbol{\Psi}^{-1}\boldsymbol{Q}^T\boldsymbol{P}^2\right) g(\boldsymbol{P}) \tag{5}$$

$$p(\boldsymbol{L}, \boldsymbol{U}) = \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{U}\boldsymbol{\Psi}^{-1}\boldsymbol{U}^T\boldsymbol{L}^T\boldsymbol{L}\right) h(\boldsymbol{L}) \tag{6}$$

where $\operatorname{etr}\{\boldsymbol{Y}\} = \exp\{\operatorname{tr}(\boldsymbol{Y})\}$, for some functions $g : \mathbb{P}(M) \to \mathbb{R}$ and $h : \mathbb{L}(M) \to \mathbb{R}$. In the sequel, we discuss some applications where this knowledge can be exploited.

3. BLIND SOURCE SEPARATION OF BINARY USERS

Here, we consider the problem of blindly separating binary users mixed by an random channel matrix from a given set of observations. We work under the data model (1), and $\boldsymbol{s}[n]$ is a vector of M binary i.i.d. sources. The prior on $\boldsymbol{A}$ is based on the independent Rayleigh fading assumption, $\boldsymbol{A} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_M \otimes \boldsymbol{\Psi})$, where the dispersion matrix $\boldsymbol{\Psi}$ is known at the receiver. Notice that the Rayleigh assumption is commonly adopted in multi-antenna setups [9]. For simplicity, $\boldsymbol{w}[n]$ is considered to be spatio-temporal white Gaussian noise, $\boldsymbol{w}[n] \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_M)$.

The problem at hand consists in estimating the source signals $\boldsymbol{S} = [\boldsymbol{s}[1]\ \boldsymbol{s}[2]\ \cdots\ \boldsymbol{s}[N]]$ given the observation matrix $\boldsymbol{X} = [\boldsymbol{x}[1]\ \boldsymbol{x}[2]\ \cdots\ \boldsymbol{x}[N]]$, where N is the number of available data samples, in the matricial data model

$$\boldsymbol{X} = \boldsymbol{AS} + \boldsymbol{W}. \tag{7}$$

The classical solution is given by the maximum a posteriori (MAP) estimator

$$\widehat{\boldsymbol{S}}_{\mathrm{MAP}} = \underset{\boldsymbol{S} \in \mathcal{B}_{M \times N}}{\arg\max}\ p\left(\boldsymbol{S} \mid \boldsymbol{X}\right), \tag{8}$$

where $\mathcal{B}_{M \times N}$ denotes the set of 2^{MN} binary matrices. This solution leads to

$$\widehat{\boldsymbol{S}}_{\mathrm{MAP}} = \underset{\boldsymbol{S} \in \mathcal{B}_{M \times N}}{\arg\max}\ f(\boldsymbol{S}), \tag{9}$$

where we have the equality

$$f(\boldsymbol{S}) = \frac{1}{\sigma^2}\operatorname{tr}\left(\boldsymbol{X}\boldsymbol{S}^T\boldsymbol{\Delta}_{\boldsymbol{S}}^{-1}\boldsymbol{S}\boldsymbol{X}^T\right) - M\log\left(\det \boldsymbol{\Delta}_{\boldsymbol{S}}\right) \tag{10}$$

and $\boldsymbol{\Delta}_{\boldsymbol{S}} = \boldsymbol{SS}^T + \sigma^2\boldsymbol{\Psi}^{-1}$. Solving for $\widehat{\boldsymbol{S}}_{\mathrm{MAP}}$ is, however, infeasible due to the exhaustive search required over a set of cardinality $\#\mathcal{B}_{M \times N} = 2^{MN}$. An alternative approach consists in estimating both $\boldsymbol{A}$ (nuisance parameter) and $\boldsymbol{S}$. This leads to a computationally feasible scheme. We have

$$\widehat{(\boldsymbol{A}, \boldsymbol{S})}_{\mathrm{MAP}} = \underset{\boldsymbol{A} \in \mathbb{GL}(M,\mathbb{R}),\, \boldsymbol{S} \in \mathcal{B}_{M \times N}}{\arg\max}\ p\left(\boldsymbol{A}, \boldsymbol{S} \mid \boldsymbol{X}\right). \tag{11}$$

This leads to the following locally-convergent iterative algorithm: given an initial estimate $\boldsymbol{A}^{(0)}$, let

$$\boldsymbol{S}^{(k+1)} = \underset{\boldsymbol{S} \in \mathcal{B}_{M \times N}}{\arg\max}\ p(\boldsymbol{A}^{(k)}, \boldsymbol{S} \mid \boldsymbol{X})$$

$$\boldsymbol{A}^{(k+1)} = \underset{\boldsymbol{A} \in \mathbb{GL}(M,\mathbb{R})}{\arg\max}\ p(\boldsymbol{A}, \boldsymbol{S}^{(k+1)} \mid \boldsymbol{X}).$$

After some calculus, we have the feasible iterations:

$$\boldsymbol{S}^{(k+1)} = \underset{\boldsymbol{S} \in \mathcal{B}_{M \times N}}{\arg\min}\ \left\|\boldsymbol{X} - \boldsymbol{A}^{(k)}\boldsymbol{S}\right\|^2 \tag{12}$$

$$\boldsymbol{A}^{(k+1)} = \boldsymbol{X}\boldsymbol{S}^{(k+1)^T}\boldsymbol{\Delta}^{-1}_{\boldsymbol{S}^{(k+1)}}. \tag{13}$$

Notice that computing $\boldsymbol{S}^{(k+1)}$ does not require a search over a set of high cardinality, since it can be obtained columnwise. The main problem is to provide accurate initial points $\boldsymbol{A}^{(0)}$ for the algorithm defined by (12) and (13). One possibility would be to try the maximum a priori estimate of $\boldsymbol{A}$, $\boldsymbol{A}^{(0)} = \arg\max_{\boldsymbol{A}} p\left(\boldsymbol{A}\right)$. But this is useless since it leads to $\boldsymbol{A}^{(0)} = \boldsymbol{0}$. We could try to use $\boldsymbol{A}^{(0)} = \arg\max_{\boldsymbol{A}} p(\boldsymbol{A}|\boldsymbol{X})$ but this leads to an intractable problem basically for the same reasons involving the solution of (10). In the sequel, we consider two feasible alternatives. The first is a random initialization taken from the prior $p(\boldsymbol{A})$. Our own alternative puts the SOS into use. More precisely, we propose to initialize (12) with $\boldsymbol{A}^{(0)} = \widehat{\boldsymbol{P}}\widehat{\boldsymbol{Q}}$, where $\widehat{\boldsymbol{P}}$ is revealed through the SOS of the received data and where

$$\widehat{\boldsymbol{Q}} = \underset{\boldsymbol{Q} \in \mathbb{O}(M)}{\arg\max}\ p(\boldsymbol{Q} \mid \boldsymbol{P} = \widehat{\boldsymbol{P}}). \tag{14}$$

That is, we construct the missing orthogonal factor as the most probable realization of $\boldsymbol{Q}$ given that $\boldsymbol{P} = \widehat{\boldsymbol{P}}$. Using (5) we have

$$\widehat{\boldsymbol{Q}} = \underset{\boldsymbol{Q} \in \mathbb{O}(M)}{\arg\min}\ \operatorname{tr}\left(\boldsymbol{Q}\boldsymbol{\Psi}^{-1}\boldsymbol{Q}^T\widehat{\boldsymbol{P}}^2\right), \tag{15}$$

which can be solved in closed-form as follows. If $\widehat{\boldsymbol{R}} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$ is the eigenvalue decomposition of $\widehat{\boldsymbol{R}}$ (eigenvalues sorted in increasing order) and $\boldsymbol{\Psi} = \boldsymbol{ZDZ}^T$ is the eigendecomposition of $\boldsymbol{\Psi}$ (eigenvalues sorted in decreasing order), then $\widehat{\boldsymbol{Q}} = \boldsymbol{VZ}^T$ [10].

We conducted some computer simulations to test the efficiency of the two different initializations: random and SOS-based. We consider $M = 2$ binary users and packets with $N = 200$ data samples. We performed 5000 statistically independent Monte Carlo runs of the algorithm for SNR's ranging from 5 to 15 dB with a spacing of 2.5 dB. The SNR is given by the expression SNR = $\mathrm{E}\left\{\|\boldsymbol{As}[n]\|^2\right\}/\mathrm{E}\left\{\|\boldsymbol{w}[n]\|^2\right\} = \|\boldsymbol{A}\|^2/M\sigma^2$. An independent realization of $\boldsymbol{A}$, $\boldsymbol{S}$ and $\boldsymbol{W}$ was generated for each Monte Carlo run. The dispersion matrix $\boldsymbol{\Psi}$ was fixed at the outset as

$$\boldsymbol{\Psi} = \boldsymbol{C}\begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}\boldsymbol{C}^T, \tag{16}$$

where

$$\boldsymbol{C} = \begin{bmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{bmatrix}. \tag{17}$$

Figures 1 and 2 represent the BER of user 1 and user 2, respectively, as a function of the SNR. The solid lines represent the results after $I = 1$ iteration of the proposed algorithm, whereas the dashed lines correspond to $I = 2$ iterations. The lines with diamonds correspond to the random initialization. The line with circles corresponds to our proposed initialization using the PQ decomposition and the one with squares to the LU decomposition (the two curves coincide). The line with stars denotes the performance of the maximum likelihood (ML) decoder, assuming that the channel matrix is known.

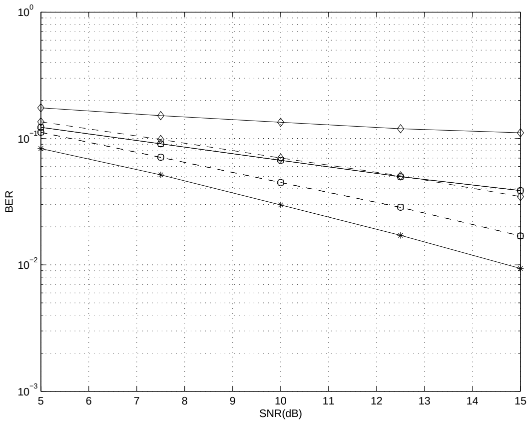

Fig. 1. BER of user 1 versus SNR

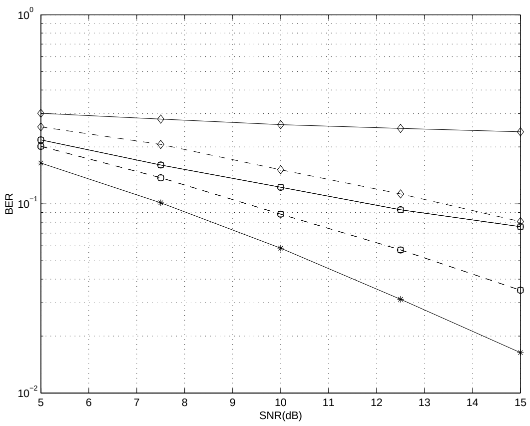

Fig. 2. BER of user 2 versus SNR

Figure 3 represents the mean square error (MSE) of the channel matrix estimate as a function of the SNR. The lines follow the same definitions, except that the bound corresponds to the stochastic Cramér-Rao bound for the MSE of the channel matrix. It is given by

$$\mathrm{MSE}(\widehat{\boldsymbol{A}}) \geq \sigma^2 M \mathrm{tr}\left(\boldsymbol{\Delta_{\mathcal{S}}}^{-1}\right), \tag{18}$$

where $\widehat{\boldsymbol{A}}$ is an estimate for the channel matrix and $\boldsymbol{\Delta_{\mathcal{S}}}^{-1}$ is defined as in (10). We can clearly see that our solution outperforms the random initialization always with significant gains. The results also improve as the number of iterations increases.

4. SEMI-BLIND CHANNEL IDENTIFICATION

As our second application, we study channel identification based on known data preambles (pilot symbols). The data model (1) and

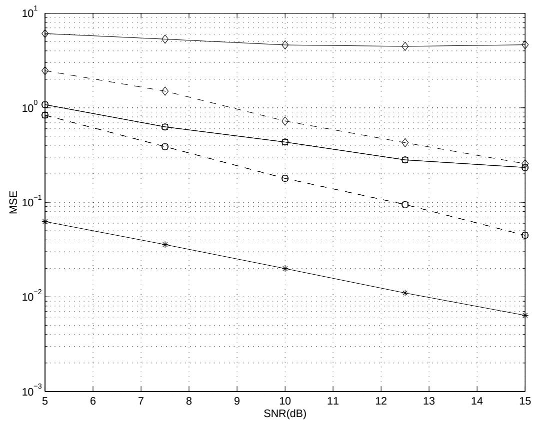

Fig. 3. MSE of channel estimate versus SNR

the assumptions taken in section 3 are maintained. The difference now is that the receiver knows the first P symbols in the sequence sent by each source. We shall denote the matrix composed by these symbols by $\boldsymbol{\mathcal{S}}$. The problem now is to identify the channel matrix $\boldsymbol{A}$ given all the available observations. A classical approach would be to find the MAP estimator (which in this case would be the same as the MMSE estimator since all the variables are Gaussian):

$$\widehat{\boldsymbol{A}}_{\mathrm{MAP}} = \underset{\boldsymbol{A} \in \mathbb{GL}(M, \mathbb{R})}{\arg\max} \; p(\boldsymbol{A} \mid \boldsymbol{\mathcal{X}}), \tag{19}$$

where $\boldsymbol{\mathcal{X}}$ is the set of observations which correspond to the known packet header. After some calculations, we can verify that this estimator is given by the expression

$$\widehat{\boldsymbol{A}}_{\mathrm{MAP}} = \boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{S}}^T \boldsymbol{\Delta_{\mathcal{S}}}^{-1}, \tag{20}$$

with $\boldsymbol{\Delta_{\mathcal{S}}}$ defined in (10). Notice that considering a MAP estimator that takes into account all observations leads to an infeasible problem (computationally intractable). This is due to the fact that the prior on the channel matrix would have to be integrated against all possible source signal sequences. Our own solution again exploits the SOS of the received data. We estimate $\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{P}}\widehat{\boldsymbol{Q}}$ where $\widehat{\boldsymbol{P}}$ is recovered through the SOS and $\widehat{\boldsymbol{Q}}$ is given by

$$\widehat{\boldsymbol{Q}} = \underset{\boldsymbol{Q} \in \mathbb{O}(M)}{\arg\max} \; p(\boldsymbol{Q} \mid \boldsymbol{\mathcal{X}}, \boldsymbol{P} = \widehat{\boldsymbol{P}}). \tag{21}$$

That is, we reconstruct the missing orthogonal factor as the most probable realization of $\boldsymbol{Q}$ given the header of observations $\boldsymbol{\mathcal{X}}$ and that $\boldsymbol{P} = \widehat{\boldsymbol{P}}$. This leads to

$$\widehat{\boldsymbol{Q}} = \underset{\boldsymbol{Q} \in \mathbb{O}(M)}{\arg\min} \; \mathrm{tr}\left(\boldsymbol{Q}^T \widehat{\boldsymbol{P}}^2 \boldsymbol{Q} \boldsymbol{\Delta_{\mathcal{S}}}\right) - 2\,\mathrm{tr}\left(\boldsymbol{Q}^T \widehat{\boldsymbol{P}} \boldsymbol{\mathcal{X}} \boldsymbol{\mathcal{S}}^T\right). \tag{22}$$

This problem has no closed-form solution but allows for fast, efficient solvers, e.g., iterative geodesic descent methods. Due to paper length constraints, details are omitted.

To assess the efficiency of our solution, we carried out some computer simulations. Figure 4 shows a scenario with $M = 2$ users, $N = 200$ data samples and $P = 8$ known header symbols. The matrix $\boldsymbol{\Psi}$ is the same as in (16). The MSE of both channel estimates is plotted against SNR's ranging from 0 to 20 dB in steps

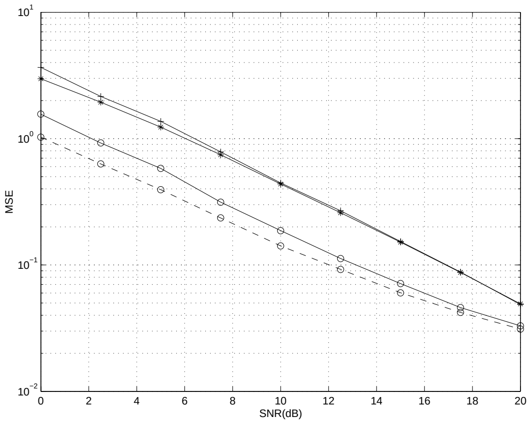

Fig. 4. MSE of channel estimate for $M = 2$ users

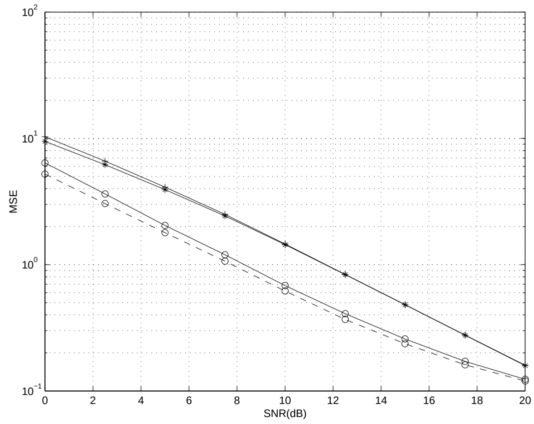

Fig. 5. MSE of channel estimate for $M = 3$ users

of 2.5 dB. For each SNR, 5000 statistically independent Monte Carlo runs were performed. The solid line with plus signs is obtained with the estimator (20). The solid line with stars is a bound calculated considering that we know the header $\boldsymbol{\mathcal{S}}$. The two curves are almost equivalent because (19) is precisely the estimator that attains this bound. The dashed line with circles corresponds to our method of estimating $\boldsymbol{A}$ with the iterative algorithm initialized near the optimum point. The solid line with circles corresponds to the iterative algorithm initialized under the assumption that the SNR is not low. This means that $\boldsymbol{\Delta}_{\boldsymbol{\mathcal{S}}} \simeq \boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{S}}^T$ and allows to alter (22) to

$$\boldsymbol{Q}^{(0)} = \underset{\boldsymbol{Q} \in \mathbb{O}(M)}{\arg\max} \ \operatorname{tr}\left(\boldsymbol{Q}^T \widehat{\boldsymbol{P}} \boldsymbol{\mathcal{X}} \boldsymbol{\mathcal{S}}^T\right), \quad (23)$$

which has a closed form solution (not shown here). We can see that our algorithm always performs better than the classical solution. This is due to the fact that the set of all observations are taken into account through the use of SOS to obtain the $\boldsymbol{P}$ factor, whereas the other solution does not take all observations into account because, as was mentioned, it leads to a computationally infeasible scheme. We can also see that our heuristic initialization method (23) presents reasonably similar results to the optimum initialization. Figure 5 shows the same results but for $M = 3$ users. The matrix $\boldsymbol{\Psi}$ is now considered to be

$$\boldsymbol{\Psi} = \boldsymbol{D} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{D}^T, \quad (24)$$

where $\boldsymbol{D} \in \mathbb{O}(M)$ was randomly generated. We can make the same conclusions 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 data samples.

5. CONCLUSIONS AND FUTURE WORK

We studied two different applications where SOS improved the performance of non-SOS based estimators in some Bayesian contexts: blind source separation of binary users and semi-blind channel identification. SOS solve the channel matrix up to an orthogonal factor. We have exploited the distribution of this residual matrix over the Lie group of orthogonal matrices to improve certain standard signal processing schemes.

6. 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, 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.

[6] 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.

[7] 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.

[8] J. Xavier, V. Barroso, Paulo Lopes, and Tiago Patrão, "Prewhitening of Mixing Matrices: Impact on the Contraction of Their Probability Mass Priors," *2002 IEEE International Symposium on Information Theory (ISIT'02)*, Lausanne, Switzerland, 2002.

[9] I. Telatar, "Capacity of multi-antenna Gaussian channels," Technical Memorandum, AT&T Bell Laboratories, 1995.

[10] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press.