A RESIDUAL BOUND FOR THE MIXING MATRIX IN ICA

Lieven De Lathauwer *Joos Vandewalle*
K.U.Leuven - E.E. Dept. - ESAT - SISTA/COSIC
Kard. Mercierlaan 94, B-3001 Leuven (Heverlee), Belgium
Tel: 32/16/321085 Fax: 32/16/321970
email: Lieven.DeLathauwer@esat.kuleuven.ac.be

ABSTRACT

In this paper, we derive a prewhitening-induced lower-bound on the Frobenius-norm of the difference between the original mixing matrix and its estimate in Independent Component Analysis. The derivation makes use of a lemma, stating that the sum of singular values of a matrix product cannot be larger than the sum of the products of the singular values of the distinct matrices.

1 INTRODUCTION

Let us use the following notation for the basic *Independent Component Analysis* (ICA) or *Blind Source Separation* (BSS) model:

$$Y = \mathbf{M}X + N = \tilde{Y} + N, \tag{1}$$

in which the observation vector Y, the noise vector N and the source vector X are zero-mean stochastic vectors with values in $\mathbb{R}$ or $\mathbb{C}$; the mixing matrix $\mathbf{M}$ is assumed to be regular; $\tilde{Y}$ is the signal part of the observations. The goal is to exploit the assumed mutual statistical independence of the source components to estimate the mixing matrix and/or the source signals from realizations of Y.

Many ICA-algorithms are prewhitening-based. An Eigenvalue Decomposition of the observed covariance allows to estimate the number of sources and to decorrelate them; the remaining rotational degree of freedom is fixed by resorting to the Higher-Order Statistics (HOS) of the observations. The prewhitening step has the disadvantage w.r.t. the higher-order step that its partial results are directly affected by additive Gaussian noise. The error introduced at this stage cannot be compensated by the higher-order step, and introduces an upper-bound to the performance of the ICA-algorithm. This has led to the development of higher-order-only ICA-procedures [1, 2, 6, 8].

In the literature, performance bound derivations focus on the quality of *separation*, in terms of the Inter-Symbol-Interference [4]. On the other hand, it would be natural to evaluate the *identification* accuracy in terms of the Frobenius-norm of the difference between the estimated and the true mixing matrix. This forms the topic of our paper.

In the next section, we will first prove a lemma stating that the sum of singular values of a matrix product cannot be larger than the sum of the products of the singular values of the distinct matrices. Although this lemma looks like a classical result, we have not been able to find it in the standard literature on matrix algebra. In particular it is stronger than what could be expected from the inequality in [10], Chapter 7, Sect. 3, Problem 18. Based on this lemma, we will derive a prewhitening-induced lower-bound on the Frobenius-norm of the difference between the original mixing matrix and its estimate. Section 3 illustrates the result by means of a number of simulations.

2 RESIDUAL BOUND

First we prove the following lemma:

Lemma 2.1 *Let the SVD's of* $\mathbf{A} \in \mathbb{C}^{I\times I}$, $\mathbf{B} \in \mathbb{C}^{I\times I}$ *and the product* $\mathbf{AB}$ *be given by:*

$$\mathbf{A} = \mathbf{U_A S_A V_A^H}, \tag{2}$$

$$\mathbf{B} = \mathbf{U_B S_B V_B^H}, \tag{3}$$

$$\mathbf{AB} = \mathbf{U_{AB} S_{AB} V_{AB}^H}, \tag{4}$$

and let their respective singular values be given by $\sigma_i(\mathbf{A})$, $\sigma_i(\mathbf{B})$ *and* $\sigma_i(\mathbf{AB})$ $(1 \leqslant i \leqslant I)$. *Then we have:*

$$\sum_i \sigma_i(\mathbf{AB}) \leqslant \sum_i \sigma_i(\mathbf{A})\sigma_i(\mathbf{B}). \tag{5}$$

The equality sign holds iff the matrix $\mathbf{V_A^H U_B}$ *is diagonal, containing only unit-norm scalars, in the case that all the pairs* $(\sigma_i(\mathbf{A}), \sigma_i(\mathbf{B}))$ *are mutually different; in the case that* $(\sigma_i(\mathbf{A}), \sigma_i(\mathbf{B})) = (\sigma_{i+1}(\mathbf{A}), \sigma_{i+1}(\mathbf{B})) = \ldots = (\sigma_j(\mathbf{A}), \sigma_j(\mathbf{B}))$, *the equality sign still holds if the corresponding diagonal block of* $\mathbf{V_A^H U_B}$ *is a unitary matrix.*

Proof: First we observe that the sum $\sum_i a_i b_i$, in which a_i and b_i are positive real numbers, is maximal if a_i and b_i are ordered in the same way as a function of

magnitude. This is easily verified in the case of two terms. The general case follows immediately, as any permutation (of e.g. the elements b_i) is a composition of transpositions (i.e. permutations of two elements).
Next, observe that the right-hand side of Eq. (5) can be written as the inner product $\langle \mathbf{S_A}, \mathbf{S_B} \rangle$. The left-hand side takes the form

$$\begin{aligned} &\langle \mathbf{AB}, \mathbf{U_{AB}} \mathbf{V_{AB}^H} \rangle \\ &= \langle \mathbf{S_A} (\mathbf{V_A^H} \mathbf{U_B}), (\mathbf{U_A^H} \mathbf{U_{AB}} \mathbf{V_{AB}^H} \mathbf{V_B}) \mathbf{S_B} \rangle \quad (6) \\ &\stackrel{\text{def}}{=} \langle \mathbf{S_A} \mathbf{P_1}, \mathbf{P_2} \mathbf{S_B} \rangle. \quad (7) \end{aligned}$$

We will consider this expression as a function f of the unitary matrices $\mathbf{P_1}$ and $\mathbf{P_2}$. It will be differentiated over the manifolds of unitary matrices, and the global optimum will be found among the "critical points" (i.e. where an infinitesimal variation of $\mathbf{P_1}$, or $\mathbf{P_2}$, does not influence the value of $f(\mathbf{P_1}, \mathbf{P_2})$).
The effect on f of a variation of e.g. $\mathbf{P_1}$ is investigated by giving $\mathbf{P_1}$ a "velocity" as a function of a "time" coordinate t. The derivative with respect to t is indicated by a dot: $\dot{\mathbf{P}}_\mathbf{1}$. The condition for $\mathbf{P_1}(t)$ to remain unitary, is that $\dot{\mathbf{P}}_\mathbf{1} = \mathbf{\Omega P_1}$, in which $\mathbf{\Omega}$ is skew-Hermitean. We have:

$$\begin{aligned} \dot{f}(\mathbf{P_1}, \mathbf{P_2}) &= \langle \mathbf{S_A} \dot{\mathbf{P}}_\mathbf{1}, \mathbf{P_2} \mathbf{S_B} \rangle \quad (8) \\ &= \langle \mathbf{\Omega}, \mathbf{S_A} \mathbf{P_2} \mathbf{S_B} \mathbf{P_1^H} \rangle. \quad (9) \end{aligned}$$

This inner product vanishes iff

$$\mathbf{S_A} \mathbf{P_2} \mathbf{S_B} \mathbf{P_1^H} = \mathbf{P_1} \mathbf{S_B} \mathbf{P_2^H} \mathbf{S_A}, \quad (10)$$

for which we define $\mathbf{C} = \mathbf{P_2} \mathbf{S_B} \mathbf{P_1^H}$. A similar derivation in terms of $\mathbf{P_2}$ yields:

$$\mathbf{S_A} \mathbf{C}^H = \mathbf{C} \mathbf{S_A}. \quad (11)$$

Eqs.(10,11) can be combined to yield:

$$\begin{aligned} \mathbf{S_A^2} (\mathbf{C} \mathbf{C}^H) &= (\mathbf{C} \mathbf{C}^H) \mathbf{S_A^2}, \quad (12) \\ \mathbf{S_A^2} (\mathbf{C}^H \mathbf{C}) &= (\mathbf{C}^H \mathbf{C}) \mathbf{S_A^2}, \quad (13) \\ \mathbf{S_B^2} (\mathbf{C} \mathbf{C}^H) &= (\mathbf{C} \mathbf{C}^H) \mathbf{S_B^2}, \quad (14) \\ \mathbf{S_B^2} (\mathbf{C}^H \mathbf{C}) &= (\mathbf{C}^H \mathbf{C}) \mathbf{S_B^2}. \quad (15) \end{aligned}$$

Assume that all the pairs $(\sigma_i(\mathbf{A}), \sigma_i(\mathbf{B}))$ are mutually different. Then the preceding commutation equations can only hold, if $\mathbf{P_1}$ is a matrix that contains exactly one, unit-modulus, entry in each column and row. Recalling the definition of $\mathbf{P_2}$, the equivalent condition for $\mathbf{P_2}$ is then automatically satisfied. Recalling the preliminary observation, the global optimum of f is reached when $\mathbf{P_1}$ is diagonal. In case that not all the singular value pairs are mutually different, the derivation can be generalized as specified by the lemma. □

Let the sample estimate of $\mathbf{C}_2^{\tilde{Y}}$ be given by $\hat{\mathbf{C}}_2^{\tilde{Y}}$, and let the symmetric EVD's of these matrices be given by

$$\begin{aligned} \mathbf{C}_2^{\tilde{Y}} &= \mathbf{E} \mathbf{D}^2 \mathbf{E}^H, \quad (16) \\ \hat{\mathbf{C}}_2^{\tilde{Y}} &= \hat{\mathbf{E}} \hat{\mathbf{D}}^2 \hat{\mathbf{E}}^H. \quad (17) \end{aligned}$$

For convenience we assume that the diagonal entries of $\mathbf{D} = \mathtt{diag}(\mathtt{d}_{11}, \mathtt{d}_{22}, \ldots)$ are strictly positive and mutually different (analogous for $\hat{\mathbf{D}} = \mathtt{diag}(\hat{\mathtt{d}}_{11}, \hat{\mathtt{d}}_{22}, \ldots)$). The derivation for the general case is analogous, but more cumbersome. With respect to the inherent indeterminacy of the ICA-solution (the mixing matrix can only be estimated up to a scaling and permutation of its columns), we assume that $\mathbf{M}$ and its estimate $\hat{\mathbf{M}}$ correspond to unit-variance sources and source estimates, respectively. The SVDs of $\mathbf{M}$ and $\hat{\mathbf{M}}$ are written as:

$$\begin{aligned} \mathbf{M} &= \mathbf{EDQ}, \quad (18) \\ \hat{\mathbf{M}} &= \hat{\mathbf{E}} \hat{\mathbf{D}} \hat{\mathbf{Q}}. \quad (19) \end{aligned}$$

Now we show that the quality of the estimation is bounded by the quality of the prewhitening in the following way:

Theorem 2.2 *The quality of the mixing matrix estimate is bounded by the quality of the prewhitening in the following way:*

$$\begin{aligned} \|\mathbf{M} - \hat{\mathbf{M}}\|^2 &\geqslant \sum_i (d_{ii}^2 + \hat{d}_{ii}^2 - 2 s_{ii}) \quad (20) \\ &\geqslant \sum_i (d_{ii} - \hat{d}_{ii})^2 \quad (21) \\ &\geqslant 0, \quad (22) \end{aligned}$$

in which s_{ii} is the ith singular value of $(\hat{\mathbf{C}}^{\tilde{Y}})^{H/2} \cdot (\mathbf{C}^{\tilde{Y}})^{1/2}$, involving arbitrary square roots of $\mathbf{C}^{\tilde{Y}}$ and $\hat{\mathbf{C}}^{\tilde{Y}}$. The first inequality reduces to an equality for an optimal choice of the unitary factor in the higher-order ICA-step. The bound is induced by the prewhitening in the following way: the second inequality vanishes iff $\mathbf{E} = \hat{\mathbf{E}} \mathbf{P}$, in which $\mathbf{P}$ is a column-wise permuted diagonal matrix, containing only unit-modulus entries, and the third inequality vanishes if the eigenvalues of $\mathbf{C}^{\tilde{Y}}$ are correctly estimated.

Proof: The minimization of $\|\mathbf{M} - \hat{\mathbf{M}}\|^2$ in terms of $\hat{\mathbf{Q}}$ is a unitary Procrustes problem ([9], p. 582). We have:

$$\begin{aligned} \|\mathbf{M} - \hat{\mathbf{M}}\|^2 &= \|\mathbf{M}\|^2 + \|\hat{\mathbf{M}}\|^2 - 2\mathtt{Re}(\langle \mathbf{M}, \hat{\mathbf{M}} \rangle) \quad (23) \\ &= \|\mathbf{D}\|^2 + \|\hat{\mathbf{D}}\|^2 \\ &\quad - 2\mathtt{Re}(\langle \hat{\mathbf{D}} \hat{\mathbf{E}}^{\mathtt{H}} \mathbf{ED}, \tilde{\mathbf{Q}} \rangle), \quad (24) \end{aligned}$$

in which $\tilde{\mathbf{Q}} \stackrel{\text{def}}{=} \hat{\mathbf{Q}}^H \mathbf{Q}$. If the SVD of $\hat{\mathbf{D}} \hat{\mathbf{E}}^H \mathbf{ED}$ is given by $\mathbf{USV}^H$, then the optimal $\tilde{\mathbf{Q}}$ takes the form of $\mathbf{UV}^H$. This proves the first inequality.
The second inequality,

$$\sum_i s_{ii} \leqslant \sum_i d_{ii} \hat{d}_{ii}, \quad (25)$$

is established by resorting to Lemma 2.1, in which $\mathbf{A} = \hat{\mathbf{D}} \hat{\mathbf{E}}^H$ and $\mathbf{B} = \mathbf{ED}$. □

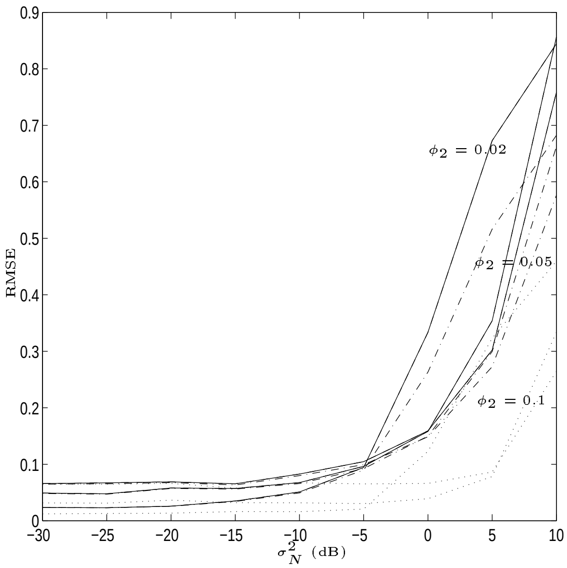

Figure 1: The RMSE between the true mixing matrix and its estimate. Effect of the SNR ($\sigma_1^2 = \sigma_2^2 = 0$dB) on the quality of the reconstruction. $\phi_1 = 0$. Solid: the achieved performance. Dashdotted: the first upper-bound of performance in Theor. 2.2. Dotted: the second upper-bound of performance in Theor. 2.2.

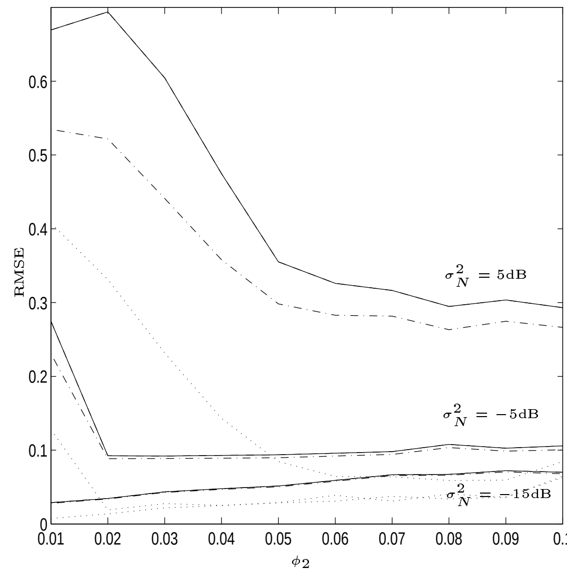

Figure 2: The RMSE between the true mixing matrix and its estimate. Effect of the difference in DOA ($\phi_1 = 0$) on the quality of the reconstruction. $\sigma_1^2 = \sigma_2^2 = 0$ (dB). Solid: the achieved performance. Dashdotted: the first upper-bound of performance in Theor. 2.2. Dotted: the second upper-bound of performance in Theor. 2.2.

3 SIMULATIONS

We consider two zero-mean complex-valued source signals, uniformly distributed over the circles with radius σ_1 and σ_2. Both signals impinge on a linear $\lambda/2$ equispaced array of 10 unit-gain omnidirectional sensors in the far field of the emitters. In the simulations the lengths of the columns of the estimated mixing matrix $\hat{\mathbf{M}}$ are normalized in the sense that the source estimates have unit power. Under this assumption, the theoretical values of the elements of the transfer matrix are given by $m_{pq} = \sigma_q e^{2j\pi p\phi_q}$, where ϕ_q equals the electrical angle of source q. The noise is Gaussian, with power σ_N^2. In each experiment the datalength $T = 100$ and the angle $\phi_1 = 0$. All curves are obtained by averaging over 500 Monte Carlo simulations. For the ICA, we used the algorithm described in [5], with the higher-order stage based on the fourth-order cumulant of the observations. In Figs. 1 and 2 we plot the Root Mean Square Error (RMSE) between the true mixing matrix and the one estimated with the maximal diagonality approach, both normalized in the same way. The dashdotted and dotted lines give the two upper-bounds of performance specified in Theor. 2.2.

4 ACKNOWLEDGEMENTS

This research was partially supported by the Flemish Government (the Concerted Research Action GOA-MIPS, the FWO (Fund for Scientific Research - Flanders) project G.0262.97, the FWO Research Communities ICCoS (Identification and Control of Complex Systems) and Advanced Numerical Methods for Mathematical Modelling, the IWT Action Programme on Information Technology (ITA/GBO/T23) - IT-Generic Basic Research), the Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs (the Interuniversity Poles of Attraction Programmes IUAP P4-02 and IUAP P4-24) and the European Commission (Human Capital and Mobility Network SIMONET, SCIENCE-ERNSI SC1-CT92-0779, ACTS project on Advanced Security for Personal Communications Technology (ASPECT), Keep In Touch (SYSIDENT-KIT124)). The scientific responsibility is assumed by its authors.

References

[1] J.-F. Cardoso, "Super-Symmetric Decomposition of the Fourth-Order Cumulant Tensor. Blind Identification of More Sources than Sensors." *Proc. ICASSP-91*, Vol. 5, pp. 3109-3112.

[2] J.-F. Cardoso, "A Tetradic Decomposition of 4th-order Tensors. Application to the Source Separation Problem." in: B. De Moor, M. Moonen (Eds.), *SVD and Signal Processing, III. Algorithms, Applications and Architectures*, Elsevier, Amsterdam, 1995, pp. 375-382.

[3] J.-F. Cardoso, A. Souloumiac, "Blind beamforming for non-Gaussian signals", *IEE Proceedings-F*, Vol. 140, No. 6, 1994, pp. 362-370.

[4] J.-F. Cardoso, "On the Performance of Orthogonal Source Separation Algorithms", *Proc. EUSIPCO-94*, Edinburgh, Scotland, U.K., Sept. 13-16, 1994, Vol. 2, pp. 776-779.

[5] P. Comon, "Independent Component Analysis, A New Concept?" *Signal Processing*, Special Issue *Higher Order Statistics*, Vol. 36, No. 3, April 1994, pp. 287-314.

[6] P. Comon, B. Mourrain, "Decomposition of Quantics in Sums of Powers of Linear Forms", *Signal Processing*, Special Issue *Higher Order Statistics*, Vol. 53, Nos. 2-3, Sept. 1996, pp. 93-108.

[7] L. De Lathauwer, B. De Moor, J. Vandewalle, "Blind Source Separation by Simultaneous Third-Order Tensor Diagonalization", *Proc. EUSIPCO-96*, Sept. 10-13, 1996, Trieste, Italy, Vol. 3, pp. 2089-2092.

[8] L. De Lathauwer, B. De Moor, J. Vandewalle, "Independent Component Analysis Based on Higher-Order Statistics Only", *Proc. 8th IEEE SP Workshop on Statistical Signal and Array Processing (SSAP-96)*, Corfu, Greece, June 24-26, 1996, pp. 356-359.

[9] G.H. Golub, C.F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, Maryland, 1991.

[10] R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, N.Y., 1991.