HAL open science

Blind Source Separation of Overdetermined Linear-Quadratic Mixtures

Leonardo Tomazeli Duarte, Ricardo Suyama, Romis Attux, Yannick Deville, Joao M. T. Romano, Christian Jutten

To cite this version:

Leonardo Tomazeli Duarte, Ricardo Suyama, Romis Attux, Yannick Deville, Joao M. T. Romano, et al.. Blind Source Separation of Overdetermined Linear-Quadratic Mixtures. LVA/ICA 2010 - 9th International Conference on Latent Variable Analysis and Signal Separation, Sep 2010, Saint Malo, France. pp.263-270. hal-00525945

HAL Id: hal-00525945

https://hal.science/hal-00525945

Submitted on 13 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Blind Source Separation of Overdetermined Linear-quadratic Mixtures

Leonardo T. Duarte[1]*, Ricardo Suyama[2], Romis Attux[1], Yannick Deville[3], João M. T. Romano[1], and Christian Jutten[4]

[1] DSPCom Lab - Universtity of Campinas (Unicamp), Campinas, Brazil
{ltduarte,romano}@dmo.fee.unicamp.br, attux@dca.fee.unicamp.br
[2] Enginnering Modeling and Applied Social Sciences, UFABC, Santo André, Brazil
ricardo.suyama@ufabc.edu.br
[3] LATT, Université de Toulouse, CNRS, Toulouse, France
ydeville@ast.obs-mip.fr
[4] GIPSA-Lab, CNRS UMR-5216, Grenoble, and Institut Universitaire de France
christian.jutten@gipsa-lab.grenoble-inp.fr

Abstract. This work deals with the problem of source separation in overdetermined linear-quadratic (LQ) models. Although the mixing model in this situation can be inverted by linear structures, we show that some simple independent component analysis (ICA) strategies that are often employed in the linear case cannot be used with the studied model. Motivated by this fact, we consider the more complex yet more robust ICA framework based on the minimization of the mutual information. Special attention is given to the development of a solution that be as robust as possible to suboptimal convergences. This is achieved by defining a method composed of a global optimization step followed by a local search procedure. Simulations confirm the effectiveness of the proposal.

1 Introduction

An interesting extension of the classical Blind Source Separation (BSS) framework concerns the case in which the mixing model is nonlinear [1]. One of the motivations for studying nonlinear BSS comes from the observation that, in some applications, the mixing process is clearly nonlinear. This is common, for instance, in chemical sensor arrays [2, 3].

Nonlinear BSS, in its most general formulation, cannot be dealt with using independent component analysis (ICA) methods [1, 4]. Indeed, if no constraints are imposed, one can set up a nonlinear system that provides independent components that are still mixed versions of the sources [4]. This result suggests that, instead of searching for a general framework, nonlinear BSS should be treated on a case-by-case basis by focusing on relevant classes of nonlinear models. Having this in mind, we tackle in this work the problem of BSS in the so-called linear-quadratic (LQ) model [5]. This class of models is appealing both in a practical

* L. T. Duarte would like to thank FAPESP for the financial support.

context —for instance, it is used in the design of gas sensor arrays [3] —and in a theoretical standpoint —it paves the way for dealing with polynomial mixtures.

A major issue in the development of BSS methods for LQ mixtures concerns the definition of the separating system structure. In a determined case (equal number of sources and mixtures), this problem is indeed tricky due to the difficulty in expressing the inverse of the mixing mapping in an analytical form. Possible solutions to this problem can be found in the nonlinear recurrent networks proposed in [5, 6] or in the Bayesian approach of [7]. Moreover, in some particular cases —for instance when there are two sources and two mixtures —one can indeed find the inverse nonlinear mapping [5].

A second route for dealing with LQ mixtures relies on the following observation: when there are more mixtures than sources (overdetermined case), the inversion of the LQ mixing model becomes simpler as it can be performed using linear separating systems. Evidently, such a simplification opens the way for well-established ICA methods developed for the linear case. Furthermore, although we restrict our analysis to the LQ case, such a simplification is also interesting in the more general case of polynomial mixtures.

Even if the idea of separating LQ mixtures through linear ICA methods is not novel, the works that have exploited it focused on particular cases, such as sources in a finite alphabet [8] or circular sources [9]. In the present paper, however, we consider a more general framework, in which the only assumption made is that the sources are mutually statistically independent. The main difficulty here lies in the fact that, although overdetermined LQ models may admit a linear inverse, classical ICA strategies may not be able to separate LQ mixtures. Motivated by these difficulties, we develop an ICA method specially tailored for the considered problem.

2 Overdetermined linear-quadratic mixing model

Let us consider a problem with two sources[5] s_1 and s_2, which are assumed to be mutually statistically independent. In a LQ model, the i-th mixture is given by

$$x_i = a_{i1}s_1 + a_{i2}s_2 + a_{i3}s_1s_2, \ \forall i \in 1, \ldots, n_m, \tag{1}$$

where a_{ij} represents a mixing coefficient and n_m is the number of mixtures. The model (1) can be alternatively described through the following vector notation

$$\begin{bmatrix} x_1 \\ \vdots \\ x_{n_m} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ \vdots & \vdots & \vdots \\ a_{n_m1} & a_{n_m2} & a_{n_m3} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_1s_2 \end{bmatrix}. \tag{2}$$

This representation suggests an insightful interpretation of the LQ model: it can be seen as a special case of a linear mixing model, in which the sources are given by s_1, s_2 and $s_3 = s_1s_2$ and, therefore, are no longer independent.

[5] This scenario is representative in the design of gas sensor arrays as one usually has binary mixtures of gases.

When the number of LQ mixtures is $n_m = 2$ (determined), there is no advantage in expressing the original LQ problem in a linear formulation. In fact, besides the presence of dependent sources, the resulting dual linear problem is underdetermined (less mixtures than sources). Performing BSS in such a scenario is quite difficult and requires the incorporation of further information.

Conversely, if, for instance[6], $n_m = 3$ (overdetermined LQ model), the resulting mixing matrix in (2) becomes square and, thus, can be inverted as follows

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \tag{3}$$

where $\mathbf{y} = [y_1\ y_2\ y_3]^T$ represents the retrieved sources. That is, one can overcome the problem of how to define an LQ separating system by simply adding sensors into the array. Even better, the solution in this case is given by a matrix.

Of course, there remains the problem of how to find a separating matrix in the case of LQ mixtures. In a recent work, Castella [8] showed that, if the sources belong to a finite alphabet, cumulant-based ICA techniques, such as the JADE algorithm [10], can be used to adjust $\mathbf{W}$ in (3), despite the presence of mutually dependent sources in the linear formulation of Equation (2). The proposed approach in [8] is thus able to retrieve s_1, s_2 and $s_1 s_2$.

In the more general case of continuous sources, the presence of dependent sources in (2) does not allow one to apply ICA methods to adjust $\mathbf{W}$ in (3). Given that, instead of searching for the three sources s_1, s_2 and $s_3 = s_1 s_2$, we try to directly estimate s_1 and s_2 via a rectangular separating matrix, as follows

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{4}$$

Structurally speaking, this separating system is also able to retrieve s_1 and s_2, possibly permuted and/or scaled. Indeed, this is achieved for all $\mathbf{A}$

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix} = \mathbf{P} \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \end{bmatrix} \mathbf{A}^{-1}, \tag{5}$$

where $\mathbf{P}$ is a permutation matrix, and α and β are non-zero values representing a possible scaling of the retrieved signals. In the sequel, we discuss the use of ICA methods to adapt the rectangular matrix $\mathbf{W}$ in (4).

3 Toward a linear ICA algorithm for overdetermined LQ mixtures

ICA-based learning rules search for a matrix $\mathbf{W}$ that provides independent signals y_1 and y_2. At first glance, ICA techniques that are used in linear overdeter-

[6] In the rest of the paper, we restrict our analysis to the case of $n_m = 3$ mixtures.

mined models could be considered to separate LQ mixtures through (4). However, as it will be discussed in the sequel, the underlying nonlinear nature of the mixing process makes the application of some common ICA strategies difficult.

3.1 Limitations of ICA methods based on whitening as a pre-processing step

Often, ICA in overdetermined linear models is carried out in two steps. Firstly, the mixtures undergo a dimension reduction stage in order to obtain signals with dimension equal to the one of the sources —this is usually done via whitening[7] [10]. Then, ICA methods designed for determined models are applied.

For this two-step solution to work in the case of LQ mixtures, the process of dimension reduction should remove any trace of nonlinear mixing between the sources. Unfortunately, this cannot be achieved via whitening. To illustrate that, let us consider, as a working example, an LQ model ($n_m = 3$ mixtures and 2 uniformly distributed sources), where the mixing matrix (see the linear formulation of Equation (2)) is given by $\mathbf{A} = [1\ 0.7\ 0.3\ ;\ 0.6\ 1\ 0.5\ ;\ 0.5\ 0.5\ 0.6]$. We checked through simulations that the matrix $\mathbf{Q} = [8.67\ -1.21\ -9.28\ ;\ -1.21\ 8.47\ -8.39]$ provides a white two-dimensional signal. Yet, the combined system $\mathbf{QA}$ in this case is given by

$$\mathbf{QA} = \begin{bmatrix} 3.30 & 0.21 & -3.57 \\ -0.33 & 3.43 & -1.17 \end{bmatrix}, \tag{6}$$

that is, the nonlinear term $s_1 s_2$ remains in the whitened data.

3.2 Limitations of natural gradient learning

From the last section, a more reasonable approach is to consider overdetermined ICA methods that do not require a whitening step. A possibility in this case can be found in the natural gradient algorithm. Although originally developed to operate in linear determined models, this method also works in the overdetermined case [11]. The learning rule in this case is given by

$$\mathbf{W} \leftarrow \mathbf{W} + \mu(\mathbf{I} - E\{f(\mathbf{y})\mathbf{y}^T\})\mathbf{W}, \tag{7}$$

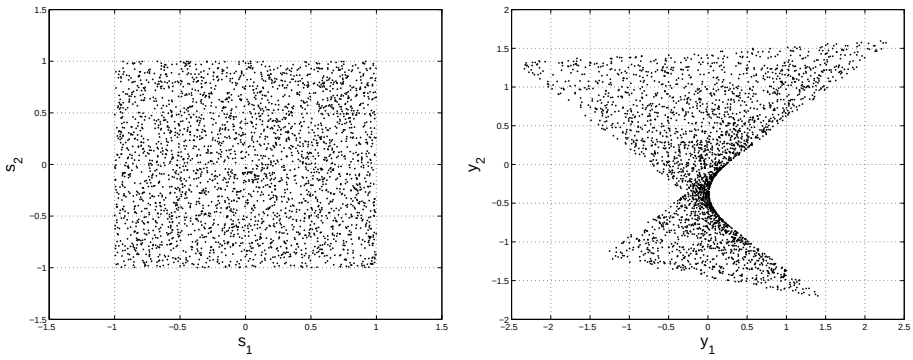
where μ is the step size, $\mathbf{I}$ represents the identity matrix and $f(\cdot)$ is a nonlinear function that should be previously defined based on the source distributions[8] [10]. Given that (7) converges when $E\{f(\mathbf{y})\mathbf{y}^T\} = \mathbf{I}$, this learning rule

[7] Whitening a vector $\mathbf{x}$ means finding a matrix $\mathbf{Q}$ that provides a vector $\mathbf{z} = \mathbf{Q}x$ whose covariance matrix is diagonal. Dimension reduction through whitening is based on the observation that the whitening matrix $\mathbf{Q}$ depends on the covariance matrix of $\mathbf{x}$, i.e. $\mathbf{R}_x$. Given that, one can have a lower dimensional vector $\mathbf{z}$ by only considering the eigenvectors associated with the largest eigenvalues of $\mathbf{R}_x$.

[8] Ideally, these functions should be as close as possible to the source score functions. However, even a rough approximation is enough to guarantee source separation in determined linear models.

is somehow trying to retrieve components that are nonlinearly decorrelated, a necessary but not sufficient condition for statistical independence (except if each component of $f(\mathbf{y})$ is a score functions of the related component of $\mathbf{y}$).

We tested (7) in the same working example as in the last section. The estimated matrix $\mathbf{W}$ in this case indeed provided nonlinearly decorrelated components satisfying $E\{f(\mathbf{y})\mathbf{y}^T\} = \mathbf{I}$ —we considered cubic functions $f(y_i) = y_i^3$, which are typically used for sub-Gaussian sources [10]. However, the mixtures were not separated. This is shown in Figure 1, which depicts the joint distribution of the original sources and of the retrieved signals. It is interesting to note here that, although nonlinearly decorrelated, the retrieved signals are not statistically independent. That is, unlike in the linear case, the nonlinear decorrelation is not a safe route for independence in overdetermined LQ models.



(a) Sources. (b) Recovered signals.

Fig. 1. Application of natural gradient algorithm: scatter plots.

3.3 Methods based on the minimization of the mutual information

We now consider a framework based on the minimization of the mutual information between the elements of $\mathbf{y}$, which is given by

$$I(\mathbf{y}) = H(y_1) + H(y_2) - H(\mathbf{y}), \tag{8}$$

where $H(\cdot)$ denotes the differential entropy [10]. Unlike in the nonlinear correlation, the mutual information offers a necessary but also sufficient condition for independence since it becomes null if and only if y_1 and y_2 are independent.

In [12], a framework to derive methods that minimize the mutual information[9] was introduced. Its application to linear models results in the following

[9] Usually, the derivation of methods based on the mutual information makes use of a common trick to avoid the estimation of the joint term $H(\mathbf{y})$. They express it in terms of $H(\mathbf{x})$ by using the entropy transformation law [10]. However, we cannot use this strategy because $\mathbf{W}$ is not invertible in our case.

learning rule

$$\mathbf{W} \leftarrow \mathbf{W} + \mu E\{\beta_{\mathbf{y}}(\mathbf{y})\mathbf{x}^T\}, \tag{9}$$

where the i-th element of $\beta_{\mathbf{y}}(\mathbf{y})$, the (opposite of the) so-called score function difference vector of $\mathbf{y}$, is given by $\beta_{y_i}(y_i) = (-\partial \log p(\mathbf{y})/\partial y_i) - (-d \log p(y_i)/dy_i)$. We applied the method proposed in [13] to estimate this vector.

After performing some tests, the algorithm (9) was able to recover the original sources in some runs. However, we also noticed that in many trials the method only provided poor estimation. One could give two reasons for such a bad performance: either the algorithm is getting trapped in spurious minima and, thus, it is an optimization issue, or the considered model is not *separable* in the sense of ICA, i.e. retrieving independent components does not assure source separation. Note that, while the first issue could be solved by developing algorithms robust to local convergence, the second one would pose a serious problem as any attempt to perform BSS through ICA would become questionable.

To gain more insight into that question, we performed a series of tests with (9). At the end of each run, we estimated the average signal-to-interference (SIR) ratio[10] and the mutual information between the retrieved signals y_1 and y_2 —we considered the estimator proposed in [14]. The results obtained after 20 realizations —with uniformly distributed sources, mixing coefficients drawn from a normal distribution and random initialization of the separating matrix $\mathbf{W}$ —are plotted in Figure 2(a), in which each mark represents one realization. Note that when a low SIR was observed, the retrieved signals were still dependent as their mutual information was not null. This is an indicator that bad convergence here comes from the optimization itself and not from a separability problem.

4 A robust ICA method for overdetermined LQ mixtures

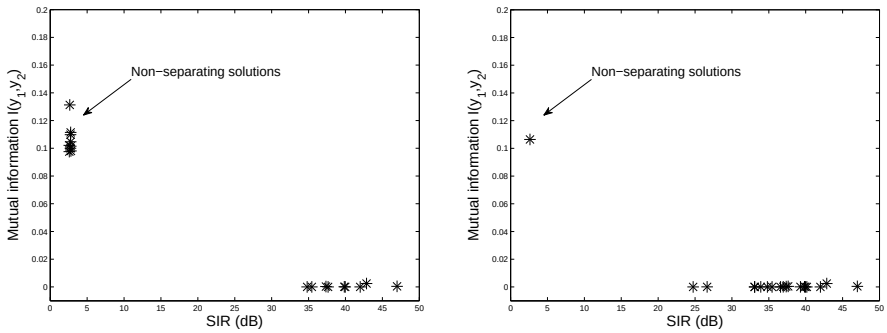
The results shown in Figure 2(a) revealed that the gradient-based learning rule of (9) may converge to local minima. A first possibility to deal with this problem is to consider global optimization methods such as evolutionary algorithms (EA). These methods are based on the notion of population, i.e. a set of possible candidate solutions (individuals) for the problem. At each iteration, new individuals are created from this population and, typically, the set of individuals that provides a better solution to the optimization problem is kept to the next iterations (selection). This population-based search gives EAs the ability of finding the global solution even when applied to multimodal cost functions.

The robustness to sub-optimal convergence in EAs comes at heavy computational burden. This is particularly problematic in the definition of an EA to perform ICA according to the minimum mutual information principle. Indeed, estimating the mutual information via accurate methods, such as the one presented in [14], is time demanding, and, since an EA performs many evaluations of the cost function during its execution, one may end up with a too slow method.

[10] The SIR associated with a source and its estimate is given by: $\mathrm{SIR}_i = 10 \log \left(E\{\hat{s}_i^2\}/E\{(\hat{s}_i - \hat{y}_i)^2\}\right)$, where $\hat{s}_i$ and $\hat{y}_i$ denote, respectively, the actual source and its corresponding estimate after mean, variance and sign normalization.

As an alternative to a direct application of an EA in our problem, we propose a hybrid scheme composed of two steps. Firstly, we indeed make use of an EA technique, the opt-aiNet algorithm (see [15] for details), to minimize the mutual information. However, instead of relying on a precise estimation of the cost function, we consider the rougher and thus simpler mutual information estimator proposed in [16]. Hence, this first step provides us with a coarse estimate of the sources. This coarse solution is then refined by the learning rule (9).

In order to assess the performance of proposed hybrid scheme, we conducted a set of simulations in the same scenario as considered in Section 3.3. In Figure 2(b), we show the results obtained after 20 runs. Whereas the simple application of (9) converged to a sub-optimal minimum in 9 out of 20 runs realizations, the proposed hybrid scheme was able to provide good estimates of the sources in 19 our of 20 realizations.



(a) Learning rule (9). (b) Proposed hybrid scheme.

Fig. 2. Analysis of the retrieved signals. Each mark corresponds to one realization.

5 Conclusions

In this work, we addressed the problem of BSS in overdetermined LQ mixtures. In this case, the mixing process can be inverted through linear structures. However, as illustrated by some examples, the application of common ICA strategies is not enough to perform source separation in the studied case. In view of this limitation, we introduced a hybrid scheme composed of a global optimization tool and of a gradient-based method for minimizing the mutual information between the retrieved signals. As checked via simulations, the proposed method is able to almost always avoid convergence to sub-optimal minima

In this first study, separability of overdetermined LQ models was only verified through simulations. As this approach is useful only for gaining some insight into this issue, a first perspective for future works is to study separability on a

theoretical basis. A second point that deserves further investigation is related to the transformation of the original overdetermined problem into a determined one. We saw that whitening cannot be used here. Nonetheless, we believe that such an approach is still valid when, for instance, additional prior information on the sources are taken into account. Finally, we intent to extend the results obtained here to scenarios in which the number of sources is larger than two and also to the more general case of polynomial mixtures.

References

1. Jutten, C., Karhunen, J.: Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures. International Journal of Neural Systems **14** (2004) 267–292
2. Duarte, L.T., Jutten, C., Moussaoui, S.: A Bayesian nonlinear source separation method for smart ion-selective electrode arrays. IEEE Sensors Journal **9**(12) (2009) 1763–1771
3. Bedoya, G.: Nonlinear blind signal separation for chemical solid-state sensor arrays. PhD thesis, Universitat Politecnica de Catalunya (2006)
4. Hyvärinen, A., Pajunen, P.: Nonlinear independent component analysis: existence and uniqueness results. Neural Networks **12** (1999) 429–439
5. Hosseini, S., Deville, Y.: Blind separation of linear-quadratic mixtures of real sources using a recurrent structure. In: Proc. of the IWANN. (2003) 289–296
6. Deville, Y., Hosseini, S.: Recurrent networks for separating extractable-target nonlinear mixtures. part i: Non-blind configurations. Signal Processing **89** (2009) 378–393
7. Duarte, L.T., Jutten, C., Moussaoui, S.: Bayesian source separation of linear-quadratic and linear mixtures through a MCMC method. In: Proc. of the IEEE MLSP. (2009)
8. Castella, M.: Inversion of polynomial systems and separation of nonlinear mixtures of finite-alphabet sources. IEEE Trans. on Sig. Proc. **56**(8) (2008) 3905–3917
9. Abed-Meraim, K., Belouchiani, A., Hua, Y.: Blind identification of a linear-quadratic mixture of independent components based on joint diagonalization procedure. In: Proceedings of the IEEE ICASSP 1996. Volume 5. (1996) 2718–272
10. Hyvarinen, A., Karhunen, J., Oja, E.: Independent component analysis. John Wiley & Sons (2001)
11. Zhang, L.Q., Cichocki, A., Amari, S.: Natural gradient algorithm for blind separation of overdetermined mixture with additive noise. IEEE Signal Processing Letters **6**(11) (2009) 293–295
12. Babaie-Zadeh, M., Jutten, C., Nayebi, K.: Differential of the mutual information. IEEE Signal Processing Letters **11**(1) (January 2004) 48–51
13. Pham, D.T.: Fast algorithm for estimating mutual information, entropies and score functions. In: Proceedings of the ICA. (2003) 17–22
14. Darbellay, G.A., Vajda, I.: Estimation of the information by an adaptive partitioning of the observation space. IEEE Trans. on Inf. Theory **45**(4) (1999) 1315–1321
15. de Castro, L.N., Timmis, J.: Artificial Immune Systems: A New Computational Approach. Springer-Verlag (2002)
16. Moddemeijer, R.: On estimation of entropy and mutual information of continuous distributions. Signal Processing **16**(3) (1989) 233–248