A Quadratic Divergence-Based Independence Measure Applied to Linear-Quadratic Mixtures

Denis G. Fantinato, Rafael A. Ando, Aline Neves, Leonardo T. Duarte, Christian Jutten and Romis Attux

Abstract—**In the context of the Blind Source Separation (BSS) problem, the use of Mutual Information (MI) as an independence measure can be very effective, even for certain types of nonlinear mixtures. However, in this case, it is generally necessary to estimate the joint and the marginal distributions associated with the random variables of interest. In this work, we consider the kernel methods for distribution estimation allied to a convenient metric as alternative to classical MI: the quadratic divergence. The proposed method is applied to the problem of Linear-Quadratic mixtures, using a recurrent network for separation and an evolutionary algorithm, the opt-aiNet, for parameter optimization. Its performance is analyzed and compared with the classical MI estimated via histogram in three different scenarios including synthetic and real data. The results are favorable to the proposal, especially for small data sets.**

Keywords— Blind Source Separation, Linear Quadratic, Quadratic Divergence, Kernel Methods

I. INTRODUCTION

In signal processing theory, the problem of Blind Source Separation (BSS) occupies a prominent place in view of its applicability in several real world problems, like geophysics, neuronal activity analysis, communications, audio/video data modeling, and sensor signals compensation [1,2]. From the classical perspective, the BSS problem consists of separating a set of signals by using only measurements of their mixtures and, possibly, some available *a priori* information.

Usually, it is assumed that the mixing process is linear and that the sources are statistically independent. Under these assumptions, one can successfully use techniques that are able to explore the independence prior, which establishes a connection with the so-called Independent Component Analysis (ICA) [1,2]. Undoubtedly, in this context, the entity known as Mutual Information (MI) plays a key role to quantify the degree of statistical dependence between the source estimates.

Although MI-based methods are shown to be very effective in the context of linear mixtures, their extension to the general nonlinear BSS problem is not straightforward, since ICA might not be sufficient in these cases [3,4]. In view of this, the problem is generally investigated for specific applications with constrained mixing models, among which can be included the Post-Nonlinear (PNL) [5] and the Linear-Quadratic (LQ) [6,7] cases.

The LQ model, in particular, is of special interest for us due to an interesting prerogative: though the mixtures are nonlinearly combined, independence among sources indicates to be sufficient for source separation [6]. The relevance of this model can also be justified in terms of its applicability [7-9] and of the theoretical line of investigation followed in works like [10], which, aside from being an important contribution *per se*, can be a path towards increasing the tractability of other nonlinear formulations.

Denis G. Fantinato, Rafael A. Ando and Romis Attux, School of Electrical and Computer Engineering, University of Campinas, Campinas-SP, Brazil, E-mails: {denisgf,assato,attux}@dca.fee.unicamp.br. Aline Neves, Engineering, Modeling and Applied Social Science Center, Federal University of ABC, Santo André-SP, Brazil, E-mail: aline.neves@ufabc.edu.br. Leonardo T. Duarte, School of Applied Sciences, University of Campinas, Limeira-SP, Brazil, E-mail: leonardo.duarte@fca.unicamp.br. Christian Jutten, GIPSA-Lab, Grenoble INP, CNRS, Grenoble, France, E-mail: christian.jutten@gipsa-lab.grenoble-inp.fr.

Within the LQ problem, it is customary to adopt the score functions as an indirect form of the MI concept [7]. However, if one considers the possibility of estimating the probabilistic densities functions (PDFs) from the available samples, the MI notion can be directly applied, e.g., via histogram or kernel methods [11]. In special, the last method – in this case, also referred to as Parzen window method – presents some advantages, like: (*i*) the ability of specifying more flexible window shapes than the rectangular window – used in the histogram method –, (*ii*) the improved convergence to the underlying density and (*iii*) its differentiability, allowing gradient methods to be applied [11,12]. Hence, in this work, we employ kernel methods for density estimation but, in addition to that, aim at an alternative measure of independence: the Quadratic Divergence (QD) between densities, a metric belonging to the field of Information Theoretic Learning (ITL) [12,13]. Although deeply related to MI, this measure presents the advantage of being a difference of PDFs, hence avoiding stability problems when the denominator PDF is zero, and can also be very convenient if combined with certain types of kernel functions. This approach, along with the use of a recurrent network and the opt-aiNet, an evolutionary optimization method [14], will be compared to the classical histogram MI estimate in synthetic and real data scenarios.

This work is organized as follows. In Section II, we describe the BSS problem and derive the QD separation criterion. In Section III, the linear quadratic model is presented. The invertibility of the nonlinear system and the optimization method are introduced in Section IV. A number of simulation results are analyzed in Section V. Finally, the conclusions are summarized in Section VI.

II. ICA AND THE MUTUAL INFORMATION

In the BSS problem, the main objective is to recover a finite set of N sources named $s_i(n)$, for $i = 1, \ldots, N$, from M observations of the mixtures, which were generated according to the model:

$$\boldsymbol{x}(n) = \mathbf{f}\big(\boldsymbol{s}(n)\big), \quad (1)$$

where $\boldsymbol{s}(n)$ is the column vector with all sources at the instant n, $\mathbf{f}(\cdot)$ is the mixing function, potentially nonlinear, and $\boldsymbol{x}(n)$

is the column vector containing the M mixtures (or observations) at the instant n. Therefore, it is desired to obtain $\boldsymbol{y}(n) = \mathbf{g}(\boldsymbol{x}(n))$, such that $\mathbf{g}(\cdot) = \mathbf{f}^{-1}(\cdot)$, up to scale and permutation factors [1].

In view of the assumption of statistically independent sources, the separation is usually performed within the ICA framework, being the mutual information [1,2] a canonical cost function to be minimized:

$$I(\boldsymbol{y}(n)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{y_1 \dots y_N}(\bar{y}_1, \dots, \bar{y}_N) \times log\left(\frac{f_{y_1 \dots y_N}(\bar{y}_1, \dots, \bar{y}_N)}{\prod_{i=1}^{N} f_{y_i}(\bar{y}_i)}\right) d\bar{y}_1 \dots d\bar{y}_N \quad (2)$$

where $f_{y_1 \dots y_N}(\bar{y}_1, \dots, \bar{y}_N)$ is the joint PDF of the output vector $\boldsymbol{y}(n)$ and $f_{y_i}(\bar{y}_i)$ the marginal PDF associated with the *i*-th recovered source.

For linear mixtures, Eq. (2) can be simplified, being necessary to compute only the marginal distributions [1,2]. Notwithstanding, in a more general nonlinear case, all PDFs, even the multivariate one, must be obtained/estimated. In that sense, there are distinct methods aimed at PDF estimation. In this work, we detach the kernel density estimators, which is a nonparametric approach capable of encompass multivariate and/or univariate signals [12,13].

Interestingly, the MI defined by Eq. (2) can be viewed from another perspective: since the minimum point is only achieved when the joint PDF $f_{y_1 \dots y_N}(\bar{y}_1, \dots, \bar{y}_N)$ is equal to the product of the marginal densities, i.e., $\prod_{i=1}^{N} f_{y_i}(\bar{y}_i)$, the MI cost can also be viewed as the matching of PDFs. From the ITL field, this notion – along with the kernel methods – was intensively studied from the standpoint of the Quadratic Divergence (QD) [12]. Although there are alternative ITL measures to the classical MI, like the α-mutual information (based on the Rényi's framework), the QD shows to be a simpler and more stable option [12,13].

A. The Quadratic Divergence

As an alternative to the classical MI, a very promising perspective is to employ the QD between estimates of the joint PDF and the marginal PDFs product of the separated sources, which can be expressed through the following relation:

$$J_{QD}(\boldsymbol{y}(n)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(f_{y_1 \dots y_N}(\bar{y}_1, \dots, \bar{y}_N) - \prod_{i=1}^{N} f_{y_i}(\bar{y}_i) \right)^2 d\bar{y}_1 \dots d\bar{y}_N \quad (3)$$

Note that the comparison between PDFs is carried out by a difference, which avoids some instability issues faced by MI when $\prod_{i=1}^{N} f_{y_i}(\bar{y}_i)$ (denominator) tends towards to zero. In Eq. (3), when both PDFs are coincident, the divergence is null, and the separated sources will be independent (analogously to the mutual information).

Henceforth, for the sake of exposition simplicity, we will consider the case in which there are only two sources to be separated, ($N = 2$), but, notwithstanding, the formulation can be extended to a generic case. Thus, with two sources y_1 and y_2, Eq. (3) reduces to

$$J_{QD}(\mathbf{W}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{y_1 y_2}^2(\bar{y}_1, \bar{y}_2)\, d\bar{y}_1 d\bar{y}_2 + \int_{-\infty}^{\infty} f_{y_1}^2(\bar{y}_1) d\bar{y}_1 \int_{-\infty}^{\infty} f_{y_2}^2(\bar{y}_2) d\bar{y}_2 - 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{y_1 y_2}(\bar{y}_1, \bar{y}_2) f_{y_1}(\bar{y}_1) f_{y_2}(\bar{y}_2)\, d\bar{y}_1 d\bar{y}_2 \quad (4)$$

The next step is to use the kernel estimates of the PDFs, based on Gaussian kernels. More specifically, the joint PDF estimate will be represented as

$$\hat{f}_{y_1 y_2}(\bar{y}_1, \bar{y}_2) = \frac{1}{L} \sum_{j=1}^{L} G_{\sigma^2}(\bar{y}_1 - y_1(j)) \times G_{\sigma^2}(\bar{y}_2 - y_2(j)) \quad (5)$$

and that of each marginal PDF as

$$\hat{f}_{y_i}(\bar{y}_i) = \frac{1}{L} \sum_{j=1}^{L} G_{\sigma^2}(\bar{y}_i - y_i(j)) \quad (6)$$

where

$$G_{\sigma^2}(\bar{y}_i - y_i(j)) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(\frac{-(\bar{y}_i - y_i(j))^2}{2\sigma^2}\right) \quad (7)$$

is the Gaussian kernel with kernel size σ, $y_i(j)$ is the *j*-th sample of the *i*-th separated source and L is the number of samples.

Hence, using (5) and (6) in (4), there results

$$J_{QD} = \sum_{j=1}^{L} \sum_{k=1}^{L} G_{2\sigma^2}(y_1(j) - y_1(k)) \times \left(\frac{1}{L^2} G_{2\sigma^2}(y_2(j) - y_2(k)) + \frac{1}{L^4} \sum_{l=1}^{L} \sum_{m=1}^{L} G_{2\sigma^2}(y_2(l) - y_2(m)) - \frac{2}{L^3} \sum_{l=1}^{L} G_{2\sigma^2}(y_2(j) - y_2(l)) \right) \quad (8)$$

where the following relation was used:

$$\int G_{\sigma^2}(\bar{y}_i - y_i(j)) G_{\sigma^2}(\bar{y}_i - y_i(k))\, d\bar{y}_i = G_{2\sigma^2}(y_i(j) - y_i(k)) \quad (9)$$

which is a convenient property of the Gaussian kernel – although other kernels are also employable.

We expect that the QD criterion – Eq. (8) –, as a kernel-based approach, will be able to engender reliable measures from a relatively low amount of data in comparison with other traditional methods in the field.

III. THE LINEAR-QUADRATIC MODEL

In the context of BSS problems, one specific nonlinear mixing model that can be considered of particular importance, both in theoretical and practical terms, is the Linear-Quadratic (LQ) [6]. While it is known that independence does not suffice for coping with general nonlinear mixtures, empirical results suggest that the ICA methods can be valid for the LQ model.

A. Description of the Model

The LQ mixture model can have any number of sources and mixtures, being described as follows:

$$x_i(n) = \sum_{j=1}^{N} a_{ij}s_j(n) + \sum_{j=1}^{N-1}\sum_{k=j+1}^{N} b_{ijk}s_j(n)s_k(n) \tag{10}$$

being $i = 1, \ldots, M$.Particularly, in the case of two sources and two mixtures, the system associated with this model can be described as:

$$\begin{aligned} x_1(n) &= a_{11}s_1(n) + a_{12}s_2(n) + b_1s_1(n)s_2(n) \\ x_2(n) &= a_{21}s_1(n) + a_{22}s_2(n) + b_2s_1(n)s_2(n) \end{aligned} \tag{11}$$

Unlike the case of linear mixtures, preprocessing cannot be carried out for general nonlinear mixtures, including the LQ case. Sources that differ by scaling factors and offsets still represent the same solution, but they cannot be easily modeled by a similar change in the mixtures due to the nonlinearities present in the model. Therefore, in the LQ case, we cannot perform even the simplest forms of preprocessing (such as normalization or offsetting the mean to zero).

IV. LQ Invertibility: Recurrent Network and Evolutionary Optimization

In order to completely separate the LQ mixtures, the use of a recurrent network might be required [10]. Indeed, it can be shown that, for the case of two sources and two mixtures, the following relation is suitable for solving the LQ problem [7]:

$$\begin{aligned} y_1(n+1) &= x_1(n) + L_1y_2(n) + Q_1y_1(n)y_2(n) \\ y_2(n+1) &= x_2(n) + L_2y_1(n) + Q_2y_1(n)y_2(n) \end{aligned} \tag{12}$$

where $L_i = -a_{ij}/a_{ii},\ i \neq j$, and $Q_i = -b_i/(a_{11}a_{22})$. Eq. (12) can be graphically represented by Fig. 1. If this idea is extended to more sources/mixtures, it is clear that more feedback connections will be necessary. However, since we are particularly concerned with the adopted criterion, we will focus our analysis in the 2 sources/2 mixtures case. It is important to mention that, from a theoretical point of view, the structure given by Eq. (12) can also establish connections with the Deville-Hosseini and the Hérault-Jutten recurrent networks [1,10].

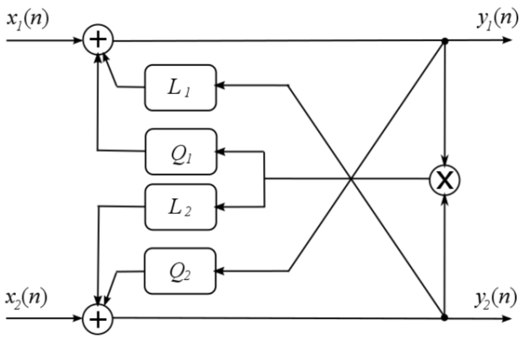

Fig. 1. Separating Structure: Recurrent network for LQ model.

Due to the ambiguity of permutation and scale factors in the solutions, it is expected that the cost function – Eq. (8) – be multimodal, i.e., that it present several null gradient points. Hence, to adjust the parameters L_1, L_2, Q_1 and Q_2 in the separation process, it becomes interesting to employ an evolutionary algorithm. Evolutionary algorithms can be defined, in simple terms, as populational metaheuristics that apply the biological principles of evolution (e.g. natural selection, random mutations and strategies of reproduction) as means of obtaining optimal "individuals" (i.e. solutions) according to a specific fitness function [14]. In this work, we used the opt-aiNet, a metaheuristic based on clonal selection principle and on the immune network theory. One of the advantages of the opt-aiNet is its good performance for optimization of multimodal functions [14], which can be very useful for problems with nonlinear/multivariate functions like ours.

In the opt-aiNet, each individual will be referred to as a network cell or simply as a cell, and will belong to a node. In each execution, the opt-aiNet must perform: cloning with mutation, exposition to the antigen (fitness evaluation), affinity measurement, selection of cells (suppression, by means of a threshold) and, finally, the introduction of random cells. In view of this, the following parameters must be adjusted: the number of initial nodes N_n; the number of clones per node N_c; the suppression threshold; the percentage of newcomers; the affinity scale β and the number of generations N_{gen}. The choice of the parameters values will be described in the following, where we also present the simulation results.

V. Simulation Results

In order to test the proposed criterion, we consider three simulation scenarios: two of them with artificially generated data – but with different distributions – and one real data problem (the show-through problem [7]). In all cases, the proposed criterion is compared to the classical MI criterion with PDFs estimated via histograms [15]. Both of them are provided with the same recurrent network and initial conditions for the opt-aiNet optimization method.

In all cases, the parameters of the opt-aiNet were chosen, after preliminary analyses, to be: $N_n = 20$ initial nodes; $N_c = 8$ clones per node; the suppression threshold and percentage of newcomers equal to 0.2; affinity scale $\beta = 10$ and number of generations $N_{gen} = 300$.

A. Synthetic Data: Scenario 1

Our first simulation for the LQ case was made with artificially-generated data, being one of the sources a positive-mean uniform distribution and the other one a combination of two Gaussians according to the model: $p\mathcal{N}(0.25, 0.01) + (1-p)\mathcal{N}(-0.25, 0.01)$, being $p = 0.5$ the occurrence probability of each Gaussian. The mixing parameters were randomly generated from a uniform distribution, resulting in $a_{11} = 0.8147$, $a_{12} = -0.6007$, $a_{21} = -0.1501$, $a_{22} = 0.9758$, $b_1 = -0.3041$ and $b_2 = -0.4954$; once generated, they remained fixed. The simulations were divided in four sets of samples, counting with 50, 100, 200 and 500 samples. Each criterion had access to the same set of data. It is important to mention that the histogram method used $\lceil \sqrt[3]{L} \rceil$ identical bins, being L the number of samples, and $\lceil \cdot \rceil$ the *ceil* operator; for the kernel method, the kernel size σ must be adjusted for each data set: through a linear search from 0.05 to 2, in intervals of 0.05, we found σ equal to 0.40, 0.1, 0.05 and 0.05 for the sets of 50, 100, 200 and 500 samples, respectively.

Since the results depend on the initial condition (which is random), as well as on the stochastic nature of the evolutionary algorithm it utilizes, our simulation was performed 10 times, being the average performance displayed. In Fig. 2 (upper plot), we present the Root Mean Square (RMS) values of the error for each set of samples for each criterion, i.e., the histogram MI method and the kernel QD method. As shown, the QD method presented better performance in all cases, especially for the sets of 100 and 200 samples, whose RMS values were approximately half of that obtained by the histogram MI method. The amount of 50 samples indicates to be a rather small number of samples, leading to an unsatisfactory performance in both cases. For 500 samples, the histogram MI method becomes more accurate and the performances of both methods are almost the same.

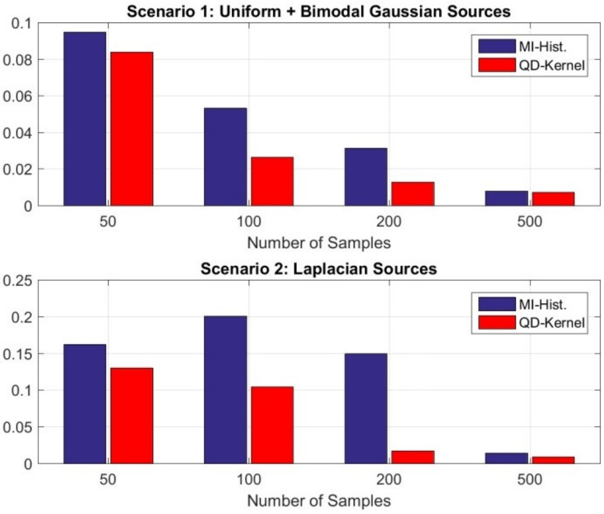

Fig. 2. RMS values for the QD-Kernel and MI-histogram criteria.

In the plots shown in Fig. 3, we can see the original sources, as well as the estimates obtained both by the histogram MI method and the QD method in one of the simulations with 200 samples. As we can see, both the kernel-based approach and the histogram estimate are able to separate the sources (the $s_i \times y_i$ plot tends to a diagonal line), but, as shown in Fig. 2, they lead to different levels of RMS values. It is also worth mentioning that, since the kernel method consists of multiple addition terms, its complexity is higher than that of the histogram approach and, in this case, for a set of 500 samples or more, the histogram method can be preferred. However, for a smaller set of data, the gain of performance brought by the QD is considerable.

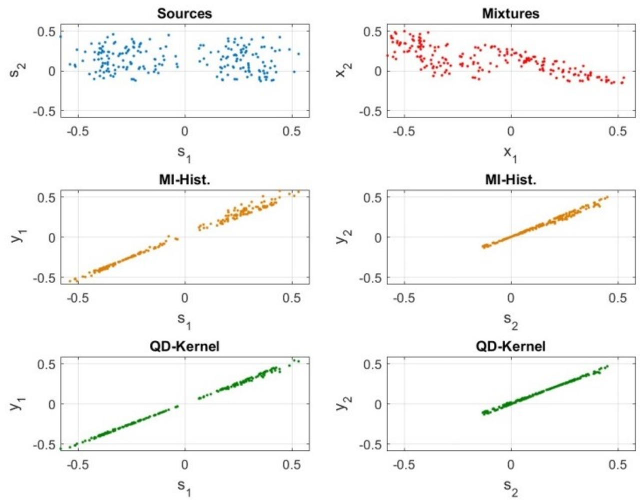

Fig. 3. Sources, mixtures and sources estimates obtained by MI and QD – 200 samples.

B. Synthetic Data: Scenario 2

In the second scenario, the sources are Laplacian distributed. In this case, the randomly generated mixing parameters were $a_{11} = 0.9058$, $a_{12} = -0.3168$, $a_{21} = 0.1006$, $a_{22} = -0.6324$, $b_1 = 0.1155$ and $b_2 = 0.1226$; we also used the sets of 50, 100, 200 and 500 samples, being the kernel sizes σ adjusted to 0.2, 0.05, 0.25 and 0.25 for each set of samples, respectively. It is worth mentioning that the kernel size can be more sensitive to a small number of samples.

By proceeding as in the previous scenario, we obtained the RMS values showed in Fig. 2 (bottom plot). It is possible to note that, for the QD, 50 and 100 samples were not sufficient for a good estimate, implying in a lower performance, however, for 200 samples and above, the method obtained good results. On the other hand, the histogram MI method presented more difficulties with this type of distribution, being necessary 500 samples to obtain a good performance.

In Fig. 4, we display one result the opt-aiNet for QD and MI, as well as the sources and mixtures (using the set of 200 samples). It is possible to observe that the QD method provided a better separation even for the data more distant from the origin.

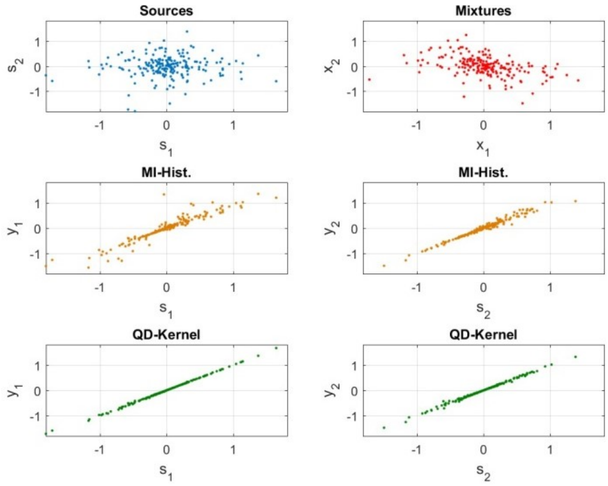

Fig. 4. Sources, mixtures and sources estimates obtained by MI and QD.

From a general point of view, we consider that the kernel-based QD method is an attractive approach when the number of samples is low, i.e., about less than 300 samples, otherwise the histogram MI method can be a less costly option with comparable performance.

C. Real Data: Show-through Problem

For the simulation with real data, we used images subject to the so-called show-through effect [7]. In these images, a thin paper containing grayscale data is digitally scanned on both sides, and for each scan, there is an interference originated from the image on the other side, as can be seen in Fig. 5. One of the images is mirrored to make it such that the two scans overlap the same images.

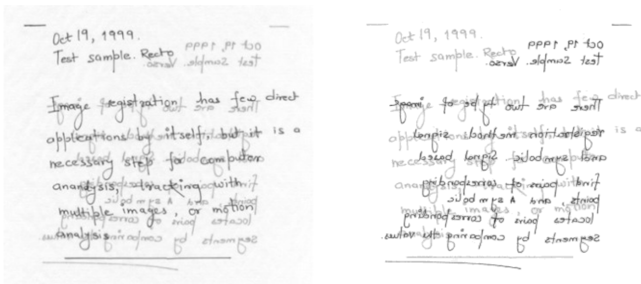

Fig. 5. Example of images with show-through effect.

If we define the original images on each side as the sources, then the scans are the mixtures and they can be modeled by an LQ model [6]. The reason why this mixture is not linear is because lighter pixels can be thought of as being "more vulnerable" to interference from the reverse image than darker ones. This nonlinear interference can be modeled by a cross-product of both pixels' brightness, which results in an LQ modeling of the problem [7].

The images considered in our simulations are two and have a resolution of 259x284 pixels, as illustrated in Fig. 5. However, to test the performance of the criterion, we considered a reduced set of samples: we randomly selected a single column in each image (the same column for both – more precisely, the column 96). From these samples, we performed the adjustment of the recurrent network by using, as before, the same parameters for the opt-aiNet and the two criteria: the QD ($\sigma = 1.2$) and the classical MI. The results are as shown in Fig. 6. Visually, we can say that the separation effectively occurred for both criteria, being the estimates very similar. However, if we look to the RMS values displayed in Tab. I, it is possible to note that the QD algorithm performed slightly better.

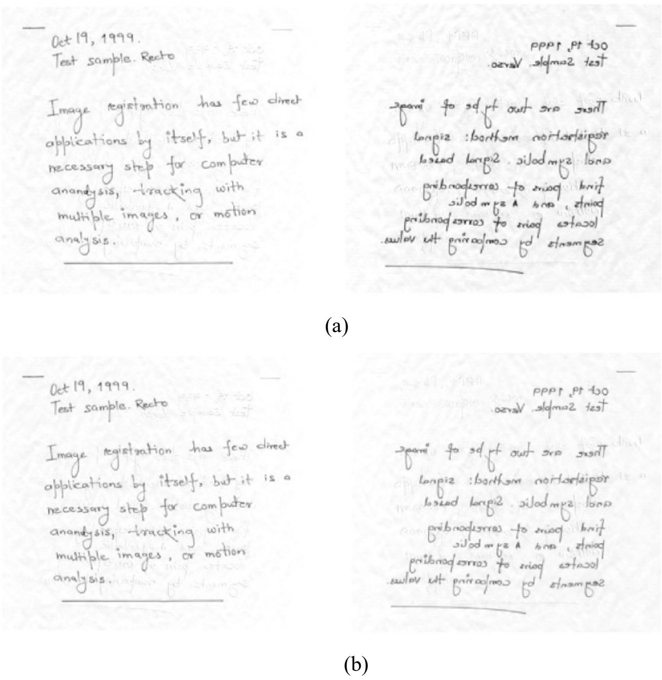

Fig. 6. Recovered Images for (a) the MI and (b) the QD method.

TABLE I. *RMS VALUES FOR THE KERNEL-QD AND HISTOGRAM-MI CRITERIA IN THE SHOW-THROUGH PROBLEM.*

No. Samples	MI	QD
259	0.002327	0.002168
518	0.002152	0.002134

In a second test, we increased the number of samples to two random columns, i.e., 518 samples and adapted the recurrent network. In this case, the performances of both criteria are very similar in terms of the RMS value, as shown in Tab. I. Note that for the histogram MI method, the addition of another column was necessary to increase the performance, while for the kernel-based QD method the performance was practically the same as before, meaning that it can provide a better performance from a lower number of samples.

VI. CONCLUSIONS

In this work, we have presented a kernel-based method to estimate the statistical independence among random variables through the metric know as quadratic distance. As shown, it can lead to simpler relations when combined with the kernel methods. This method was applied to the linear quadratic mixtures – a nonlinear BSS problem – and was compared to the histogram method for the estimation of the classical mutual information. The performance analysis took place in three different scenarios: two of them with synthetic data and the last one with real data (the show-through problem).

From the obtained results in the chosen scenarios, it is possible to state that the kernel-based quadratic distance method is able to estimate the independence between samples in a more accurate fashion than the histogram-based method, as long as the kernel size is correctly adjusted. However, the kernel size can be more sensitive for a small set of samples. The histogram method, on the other hand, becomes equally accurate when the number of samples is increased (for more than 500 samples, according to the simulations). Since the kernel method is computationally more costly than the histogram method, the proposed method is particularly attractive when the data set is small, when the performance can be twice as better as that of the other method. For future works, we intend to extend this analysis to other types of scenarios, including the ones with a larger number of sources and mixtures, as well as to underdetermined problems.

ACKNOWLEDGMENTS

The authors would like to thank FAPESP (2013/14185-2, 2015/23424-6), CNPq and CHESS for the financial support.

REFERENCES

[1] P Comon, C Jutten, Handbook of Blind Source Separation: Independent Component Analysis and Applications, Academic Press, 2010.

[2] J.M.T. Romano, R.R.F. Attux, C.C. Cavalcante, R. Suyama, Unsupervised Signal Processing: Channel Equalization and Source Separation. CRC Press, 2010.

[3] S. Hosseini, C. Jutten, On the Separability of Nonlinear Mixtures of Temporally Correlated Sources, IEEE Signal Processing Letters, vol. 10 (2), pp. 43-46, 2003.

[4] A. Taleb, C. Jutten, Source Separation in Post-Nonlinear Mixtures, IEEE Transactions on Signal Processing, vol. 47 (10), pp. 2807-2820, 1999.

[5] A. Hyvärinen, P. Pajunen. "Nonlinear Independent Component Analysis: Existence and Uniqueness Results." Neural Networks, vol. 12(3), pp. 429-439, 1999.

[6] S. Hosseini, Y. Deville, Blind Separation of Linear-Quadratic Mixtures of Real Sources Using a Recurrent Structure, Artificial Neural Nets Problem Solving Methods, pp. 241-248, Springer, 2003.

[7] F. Merrikh-Bayat, M. Babaie-Zadeh, C. Jutten, Linear-Quadratic Blind Source Separating Structure for Removing Show-Through in Scanned Documents. International Journal on Document Analysis and Recognition, Springer Verlag, vol. 14 (4), pp. 319-333, 2011.

[8] L. Almeida, Separating a Real-Life Nonlinear Image Mixture, Journal of Machine Learning Research, vol. 6, pp. 1199-1229, 2005.

[9] L. Duarte, C. Jutten, S. Moussaoui, A Bayesian Nonlinear Source Separation Method for Smart Ion-selective Electrode Array. IEEE Sensors Journal vol. 9 (12), pp. 1763-1771, 2009.

[10] Y. Deville, S. Hosseini, Recurrent Networks for Separating Extractable-Target Nonlinear Mixtures. Part I: Non-blind configurations, Signal Processing, vol. 89, pp. 378–393, 2009.

[11] Y. Moon, B. Rajagopalan, U. Lall, Estimation of Mutual Information Using Kernel Density Estimators, Physical Review E, vol. 52 (3), pp. 2318-2321, 1995.

[12] J. C. Principe, Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives, Springer, 2010.

[13] S. Achard, D. Pham, C. Jutten. "Quadratic Dependence Measure for Nonlinear Blind Sources Separation." International Conference on ICA and BSS. 2003

[14] L. De Castro, J. Timmis, An Artificial Immune Network for Multi-Modal Function Optimization. Proceedings of the IEEE Congress on Evolutionary Computation, vol. 1, pp. 699-704, 2002.

[15] D. W. Scott, Multivariate Density Estimation: Theory, Practice, and Visualization, Wiley, 2009.