Taking advantage of source correlation in forensic analysis

Pedro Comesaña #1, Fernando Pérez-González #*2

Signal Theory and Communications Department, University of Vigo
E.E. Telecomunicación, Campus Lagoas-Marcosende, 36310 Vigo-SPAIN
[1] pcomesan@gts.uvigo.es
[2] fperez@gts.uvigo.es

* *Gradiant (Galician Research and Development Center in Advanced Telecommunications)*
36310 Vigo-SPAIN

***Abstract*—In a wide range of practical multimedia scenarios several correlated contents are available. The aim of this work is to quantify the gain that can be achieved in forensic applications by jointly considering those contents, instead of analyzing them separately. The used tool is the Kullback-Leibler Divergence between the distributions corresponding to different operators; the Maximum Likelihood estimator of the applied operator is also obtained, in order to illustrate how the correlation is exploited for estimation. Our detailed analysis is constrained to the Gaussian case (both for the input signal distribution and the processing randomness) and linear operators. Several practical scenarios are studied, and the relationships between the derived results are established. Finally, the links with Distributed Source Coding are highlighted.**

I. Introduction

In the last decades the number of multimedia contents and their impact in our lives has dramatically increased. A paradigmatic example of both the cost reduction and ubiquity of capture devices and the growth of digital networks where those contents can be published, shared and distributed, is the wide use of mobile devices (e.g., smart phones) that jointly offer the capturing and connectivity functionalities. Multimedia contents have been converted not only in valuable evidence of our personal evolution and social life, but also in a weapon that can be used to harm the public image of individuals and organizations. In fact, simultaneously with this growth, a huge number of editing tools available in applications for non-skilled users have proliferated, thus compromising the reliability of those contents, and strongly constraining their use in some applications, for example as court evidence. As a consequence, trust on multimedia contents has steadily decreased.

In this context, multimedia forensics, an area of multimedia security, has appeared as a possible solution to the decrease of confidence on multimedia contents. The target of multimedia forensics can be summarized as assessing the processing, coding and editing steps a content has gone through. Despite the large attention that multimedia forensics has deserved during the last years (see, for instance, [1] and the references therein), most of the previous works deal with single sources, i.e., they perform the forensic analysis of video, audio or still images, but they do not consider in a joint way several correlated instances of those media. However, examples of those correlated contents can be found in a number of practical situations, for example:

- Multimodal content: one of the most interesting cases are video files with audio tracks. For example, both the visual and audio contents provide environment information that should be coherent; otherwise, inconsistencies would indicate that at least one of the modalities was tampered with. This idea is explored in [2], where the volumetric characteristics of the capture environment are estimated both from the video and audio signals. Data from different multimodal sources are fusioned in [3] for the purposes of geo-location, and in Different nature contents are also explored
- Multitrack files: obviously, the left and right channels of stereo audio files are not independent; the correlation between them could be exploited for forensic purposes. The same idea is applicable to 3-D video, or multi-channel audio. A somewhat related strategy is exploited in [4], where the authors propose to complicate faker's task by recording a portion of the image preview.

Be aware that the common characteristic of those scenarios is that a number (typically 2) of correlated sources is considered. In this work we will try to measure, by taking a theoretical approach, the advantage of jointly considering these contents for performing the forensic analysis of the total multimedia contents; specifically, we will quantify the gain that can be achieved by considering them in a joint way. Both information-theoretic and estimation tools will be used.

The rest of the paper is organized as follows: Sect. II introduces the used notation and the goals of the detection and estimation forensic problems. The proposed target functions and general strategies are introduced in Sect. III, while they are particularized to the linear and Gaussian case in Sect. IV.

MMSP'13, Sept. 30 - Oct. 2, 2013, Pula (Sardinia), Italy.
???-?-????-????-?/10/$??.?? *©2013 IEEE.*

Numerical results are introduced in Sect. V, and conclusions are summarized in Sect. VI.

II. Notation and objectives

Random vectors will be denoted by capital bold letters (e.g., $\mathbf{Y}$), while their outcomes, and deterministic vectors in general, will use lower case bold letters (e.g., $\mathbf{y}$). $\Sigma_{\mathbf{X}}$ will be used for denoting the covariance matrix of random vector $\mathbf{X}$, and $\boldsymbol{\mu}_{\mathbf{X}}$ its mean. Subindices will be used for denoting the vector component at ith position (e.g., Y_i, or y_i); for the sake of notational simplicity $\left(\boldsymbol{\mu}_{\mathbf{X}}\right)_i = \mu_{X_i}$. The element at the ith row and jth column of a general matrix A will be denoted by $(A)_{i,j}$.

Let $X_1, X_2, \ldots, X_L$ denote L random variables, which model the correlated sources we consider; $\mathbf{X}$ will be used for denoting $(X_1, X_2, \ldots, X_L)$. Throughout this work we will assume the statistics (mean vector and covariance matrix) of $\mathbf{X}$ to be perfectly known at the detector/estimator.

We assume that each of those variables goes through a particular processing $Y_i = g_i(X_i)$, where $1 \leq i \leq L$, $g_i \in \mathcal{G}$, and $\mathcal{G}$ denotes the space of memoryless processing operators. In general, these operators can be randomized; this randomness will be modeled by variables Z_i, $1 \leq i \leq L$, where the statistics of $\mathbf{Z}$ will be assumed to be also perfectly known at the detector/estimator. For the sake of notational simplicity, we will define every $g_i \in \mathcal{G}$ by two sets of parameters, namely, φ_i and ϕ_i, so $g_i(\cdot) = g(\cdot, \varphi_i, \phi_i)$. These two sets of parameters are used for making the distinction between those that we want to estimate/detect, and those which we do not (typically known as unwanted or nuisance parameters [5]), respectively.

For the study of the Maximum Likelihood (ML) processing operator estimator, N independent observations of $\mathbf{Z}$ will be considered, i.e., we will assume each of those L sources and the corresponding processing to be memoryless. Each of those N observations of $\mathbf{Z}$ will be denoted by $\mathbf{Z}^i$, $1 \leq i \leq N$. In the information theoretic analysis, and due to the independence among the N observations, the obtained results will be proportional to N; consequently, and for the sake of notational simplicity, we will skip the superindex.

In this work we focus on two different problems:

- study of the distinguishability between g_i and h_i, $g_i \in \mathcal{G}$ and $h_i \in \mathcal{G}$. First, we analyze the case where only the marginal probability density function (pdf) of Y_i is considered, and then we compare it with its counterpart where the joint pdf of $\mathbf{Y}$ is exploited. Of course, one would expect that whenever the joint pdf is employed, the distinguishability is improved; in that sense, one of the main contributions of the current work is to consider several scenarios that model practical signal processing operations, and to quantify the improvement achieved by using the correlation between the sources (i.e., the joint pdf instead of the marginal). The block diagram of this scenario is plotted in Fig. 1 for the case $L = 2$.
- estimate the applied operator. Again, intuition says that the more data we consider, the better (or at least not worse) the estimation will be. We analyze how the correlation between sources is exploited by the processing operator estimation. The block diagram of this scenario is plotted in Fig. 2 for the case $L = 2$.

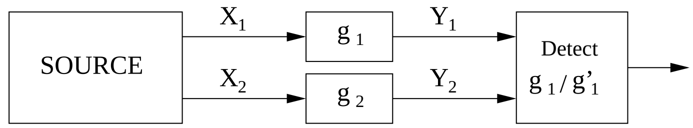

Fig. 1. Distinguishability problem framework for $L = 2$.

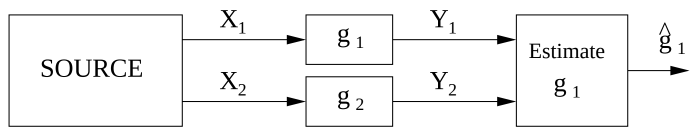

Fig. 2. Estimation problem framework for $L = 2$.

III. General case

Although already well-known in information theory, the Kullback-Leibler Divergence (KLD), also known as relative entropy, has been just recently proposed for distinguishing between different sources [6], and processing operators [7], [8] in multimedia forensics. The KLD for continuous L-dimensional random variables is defined as

$$D(f_0||f_1) = \int_{\mathbb{R}^L} f_0(\mathbf{x}) \log \left(\frac{f_0(\mathbf{x})}{f_1(\mathbf{x})} \right) d\mathbf{x},$$

where f_0 denotes the pdf under the null hypothesis, and f_1 under the alternative one (the two hypotheses under analysis). Its use is based on its asymptotical (when the dimensionality of the problem goes to infinity) optimality, since it is asymptotically equivalent to the Neyman-Pearson criterion, which is known to be the most powerful test for the binary hypothesis problem. Indeed, Chernoff-Stein's Lemma [9] states that the false positive probability error exponent achievable for a given non-null false negative probability asymptotically converges to the KLD between the pdfs under the null and alternative hypotheses (as long as the KLD takes a finite value) when the dimensionality of the problem goes to infinity.

In the case where we only want to distinguish between the values of some of the applied signal processing operator parameters (those that we have previously denoted by φ_i), but we are not interested in distinguishing between different values of the remaining ones (i.e., ϕ_i) we will follow a worst case approach. Specifically, given that we are interested in studying the distinguishability between the processing corresponding to φ_i and φ_i', we will look for those values of ϕ_i and ϕ_i' minimizing the relative entropy, i.e., to quantify the distinguishability between $f_{g(X,\varphi_i,\cdot)}$ and $f_{g(X,\varphi_i',\cdot)}$ we compute

$$\min_{\phi_i} \min_{\phi_i'} D(f_{g(X,\varphi_i,\phi_i)}||f_{g(X,\varphi_i',\phi_i')}).$$

This approach resembles the strategy which is typically followed in the literature for statistical detection theory with unwanted parameters (c.f., [5]), since it maximizes the performance of the system (by using the optimal distinguishability measure, the KLD) for the worst case scenario, ensuring the predicted performance. This strategy is also coherent with the approach proposed in [8] for quantifying the distinguishability between different classes of processing operators.

On the other hand, the ML estimate of processing g_i requires the calculation of $\hat{g}_i = \arg\max_{g_i \in \mathcal{G}} f_{\mathbf{Y}}(\mathbf{y}|g_i)$. Again, if we are interested in estimating only some of the parameters defining g_i, i.e. φ_i, then we must solve

$$\hat{\varphi}_i = \arg\max_{\varphi_i} \max_{\phi_i \in \Phi} f_{g(X,\varphi_i,\phi_i)},$$

where Φ is the feasible set of values of ϕ_i. This framework encompasses the case where ϕ_i is known to have a fixed value ϕ^*, as in such case $\Phi = \{\phi^*\}$. Note that, since in this case we are looking for the most probable operator, instead of a maxmin, a maxmax strategy will be followed; in other words, in the estimate problem it does not make sense to use a worst case approach, as one does not have to consider the probability of confusing with an alternative hypothesis.

Finally, we would like to mention that the improvement on the performance of the estimation of g_i could be also interpreted from an information-theoretic point of view. Indeed, if one considers G_i to be randomly chosen following a given distribution, then, based on fundamental properties of the entropy [9] we can bound $h(G_i|Y_i) \geq h(G_i|\mathbf{Y})$ (where $h(\cdot)$ stands for the differential entropy) i.e., the consideration of the output of the other processing branches will reduce (or at least not increase) the uncertainty about the processing undergone by X_i.

A. Links with Distributed Source Coding

In source coding, the exploitation of correlation between sources has been extensively used for improving the performance of the coding scheme in those scenarios where the coders do not share access to their input data, i.e., the Distributed Source Coding (DSC) problem [10], [11]. Indeed, this correlation is typically modeled as a virtual channel, and channel coding techniques are used for source coding purposes. Nevertheless, due to the differences in the target function between the current problem, where the processing undergone by the different sources is to be detected/estimated, and the DSC problem, where one wants to minimize the transmitted data, the translation of the channel-coding based techniques to the forensic application seems to be unfeasible.

Another related problem is the Distributed Hypothesis Testing [12], where one wants to determine how the data should be compressed in order to minimize the transmitted information when the goal is not the reproduction, but the inference from those data. Although in this case we have indeed a detection problem, in the forensic application we are not interested in reducing the transmitted data; consequently, the translation of the results in [12] to the current problem appears to be very difficult.

IV. Gaussian signals and linear operators

In order to provide close formulas that allow a clear comparison between the considered scenarios, we particularize the proposed framework to the case where Gaussian variables and linear processing is considered. Therefore, in this section we will consider the processing defined by $Y_i = a_i X_i + Z_i$, where a_i is a real constant, $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}}, \Sigma_{\mathbf{X}})$, and $Z_i \sim \mathcal{N}(\mu_{Z_i}, \sigma^2_{Z_i})$ is a Gaussian random variable independent of $\mathbf{X}$ and independent of Z_j, $1 \leq j \leq L$, $j \neq i$. Random variable Z_i might model the randomness of the processing, for example, the effect of quantizing the processed signal in a different domain, e.g., an image operator that scales the 8×8-block DCT coefficients depending on the frequency location, and then quantizes the image in the pixel domain; although the quantization error is not independent of the DCT coefficients, it is typically modeled as being so (see, for example, [13]), as a lot of different contributions are summed up when performing the DCT and IDCT.

Based on the definition of $\mathbf{Y}$, and the distributions of $\mathbf{X}$ and $\mathbf{Z}$, $\mathbf{Y}$ is also Gaussian, i.e., $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{Y}}, \Sigma_{\mathbf{Y}})$, where $\mu_{Y_i} = a_i \mu_{X_i} + \mu_{Z_i}$, and

$$\left(\Sigma_{\mathbf{Y}}\right)_{i,j} = a_i a_j \left(\Sigma_{\mathbf{X}}\right)_{i,j} + \sigma^2_{Z_i} \delta[i-j],$$

where $\delta[\cdot]$ stands for the Kronecker delta.

The main advantage of the Gaussian case, that drives us to consider this scenario with special detail, is the fact that closed formulas exist for the KLD of two Gaussian multivariate distributions. Indeed, if we consider $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{Y}}, \Sigma_{\mathbf{Y}})$ and $\mathbf{Y}' \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{Y}'}, \Sigma_{\mathbf{Y}'})$, then

$$\begin{aligned} D(f_{\mathbf{Y}}||f_{\mathbf{Y}'}) &= \frac{1}{2}\Bigg[\mathrm{tr}\left(\Sigma^{-1}_{\mathbf{Y}'}\Sigma_{\mathbf{Y}}\right) \\ &+ \left(\boldsymbol{\mu}_{\mathbf{Y}'} - \boldsymbol{\mu}_{\mathbf{Y}}\right)^T \Sigma^{-1}_{\mathbf{Y}'} \left(\boldsymbol{\mu}_{\mathbf{Y}'} - \boldsymbol{\mu}_{\mathbf{Y}}\right) \\ &- \log\left(\frac{|\Sigma_{\mathbf{Y}}|}{|\Sigma_{\mathbf{Y}'}|}\right) - L\Bigg], \end{aligned} \quad (1)$$

where $\mathrm{tr}(\cdot)$ is the trace operator, and $|\Sigma|$ is the determinant of matrix Σ.

Taking into account the form of Y_i considered in this section, g_i is entirely specified by a_i and $\sigma^2_{Z_i}$. In most practical scenarios we will be interested in estimating a_i, whereas $\sigma^2_{Z_i}$ is an unwanted parameter; therefore, following the notation introduced in the previous section, $\varphi_i = a_i$, and $\phi_i = \sigma^2_{Z_i}$. Consequently, the ML estimate of gain a_i requires the calculation of $\hat{a}_i = \arg\max_{a_i \in \mathbb{R}} \left(\max_{\sigma^2_{Z_i} \in \mathbb{R}^+} LLR(\mathbf{y}, a_i, \sigma^2_{Z_i})\right)$, where

$$LLR(\mathbf{y}, a_i, \sigma^2_{Z_i}) \triangleq \left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}}\right)^T \Sigma^{-1}_{\mathbf{Y}} \left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}}\right) + \log(|\Sigma_{\mathbf{Y}}|).$$

On the other hand, if the unwanted parameter is indeed known *a priori*, then that knowledge can be exploited in the estimation. Continuing with the estimate of a_i, but assuming that $\sigma^2_{Z_i}$ is known to be, say, $\left(\sigma^2_{Z_i}\right)^*$, we have that $\hat{a}_i = \arg\max_{a_i \in \mathbb{R}} LLR(\mathbf{y}, a_i, \left(\sigma^2_{Z_i}\right)^*)$.

In the following we consider 3 particular scenarios for $L = 2$ and different definitions of Y_2 (Sects. IV-B-IV-D), while keeping the same definition of Y_1. For all of them, we detail the theoretical results of both the ML estimator and the KLD. The target of this analysis is to illustrate how the knowledge of Y_2 helps to estimate/detect the processing undergone by Y_1 in comparison to the scenario where only Y_1 is available (Sect. IV-A). In order to keep the mathematical tractability, we will assume $\Sigma^2_{Z_i} = \mu_{X_i} = \mu_{Z_i} = 0$, for $i = 1, 2$. The noisy case (randomized processing operators) and non-zero mean will be considered in Sect. V by numerical results.

The proposed scenarios can be linked with real applications in the case of audio stereo files, where each audio channel goes through an equalization filter; the samples of each channel are windowed, frequency transformed, and then each frequency coefficient is subjected to a different scaling. This effect can be roughly modeled by a frequency dependent scaling, and the differences between this model and the real processing (encompassing, for example, the windowing effect, the lack of block periodicity, and the quantization of the filtered samples in the time domain) will be modeled by Z_i. Of course the frequency coefficients do not fit the theoretical model studied in this section, but the consideration of this application scenario showcases the power of the proposed methodology. We will particularize this illustrating application for each scenario.

A. Scenario $Y_1 = a_1 X_1 + Z_1$

Application Scenario: mono file, or only one of the stereo channels is considered for processing estimation/detection.

Under the hypotheses mentioned above,

$$\hat{a}_1 = \pm\sqrt{\frac{\sum_{i=1}^N (Y_1^i)^2}{N\left(\Sigma_{\mathbf{X}}\right)_{1,1}}}, \quad (2)$$

that is, the variance-based estimator, which in general is biased.

Concerning the KLD between $Y_1 = a_1 X_1 + Z_1$ and $Y_1' = b_1 X_1 + Z_1'$, one obtains

$$D(f_{Y_1}||f_{Y_1'}) = \frac{1}{2}\left(-1 + \frac{a_1^2}{b_1^2} - \log\left[\frac{a_1^2}{b_1^2}\right]\right). \quad (3)$$

B. Scenario $Y_1 = a_1 X_1 + Z_1$, $Y_2 = a_2 X_2 + Z_2, a_2 = a_1$

Application Scenario: stereo case, when we know that the same equalization is applied to both channels.

The ML estimator can be computed as

$$\hat{a}_1 = \pm\sqrt{\frac{\sum_{i=1}^N \left(Y_1^i\right)^2 (\Sigma_{\mathbf{X}})_{2,2} + \left(Y_2^i\right)^2 (\Sigma_{\mathbf{X}})_{1,1} - 2\left(Y_1^i Y_2^i\right)(\Sigma_{\mathbf{X}})_{1,2}}{2N\left[(\Sigma_{\mathbf{X}})_{1,1}(\Sigma_{\mathbf{X}})_{2,2} - (\Sigma_{\mathbf{X}})^2_{1,2}\right]}}.$$

Be aware that whenever X_1 and X_2 are independent, i.e., $\left(\Sigma_{\mathbf{X}}\right)_{1,2} = 0$, then the derived ML estimator is

$$\hat{a}_1 = \pm\sqrt{\frac{1}{2N}\left[\frac{\sum_{i=1}^N (Y_1^i)^2}{\left(\Sigma_{\mathbf{X}}\right)_{1,1}} + \frac{\sum_{i=1}^N (Y_2^i)^2}{\left(\Sigma_{\mathbf{X}}\right)_{2,2}}\right]}, \quad (4)$$

which is obvioulsy related to the ML estimator in (2).

Concerning the KLD between $(Y_1, Y_2) = (a_1 X_1 + Z_1, a_2 X_2 + Z_2)$ and $(Y_1', Y_2') = (b_1 X_1 + Z_1', b_2 X_2 + Z_2')$, we obtain

$$D(f_{(Y_1,Y_2)}||f_{(Y_1',Y_2')}) = -1 + \frac{a_1^2}{b_1^2} - \log\left[\frac{a_1^2}{b_1^2}\right], \quad (5)$$

which is nothing but twice (3). This result makes sense, since we are considering the same processing for both channels and the noiseless case, and consequently the correlation between X_1 and X_2 does not provide any additional information; therefore, from the KLD point of view one would expect to have the same result that is achieved when two independent realizations of Y_1 are available. This result also makes sense at the light of (4), although in the derivation of the latter we assumed $\left(\Sigma_{\mathbf{X}}\right)_{1,2} = 0$.

C. Scenario $Y_1 = a_1 X_1 + Z_1$, $Y_2 = a_2 X_2 + Z_2$. a_2 is known

Application Scenario: stereo, we know the equalization applied to one channel, and want to estimate the other one.

The ML estimator is

$$\begin{aligned}\hat{a}_1 &= \left\{\sum_{i=1}^N -a_2\left(\Sigma_{\mathbf{X}}\right)_{1,2} Y_1^i Y_2^i + \left[\left(\sum_{i=1}^N a_2\left(\Sigma_{\mathbf{X}}\right)_{1,2} Y_1^i Y_2^i\right)^2\right.\right.\\ &+ \left.\left. 4Na_2^4\left[\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2} - \left(\Sigma_{\mathbf{X}}\right)^2_{1,2}\right]\left(\Sigma_{\mathbf{X}}\right)_{2,2}\sum_{i=1}^N\left(Y_1^i\right)^2\right]^{1/2}\right\}\\ &\left[2Na_2^2\left(\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2} - \left(\Sigma_{\mathbf{X}}\right)^2_{1,2}\right)\right]^{-1}. \end{aligned} \quad (6)$$

Concerning the KLD between $(Y_1, Y_2) = (a_1 X_1 + Z_1, a_2 X_2 + Z_2)$ and $(Y_1', Y_2') = (b_1 X_1 + Z_1', b_2 X_2 + Z_2')$, we obtain

$$\begin{aligned}D(f_{(Y_1,Y_2)}||f_{(Y_1',Y_2')}) &= -1 - \tfrac{1}{2}\log\left(\tfrac{a_1^2 a_2^2}{b_1^2 b_2^2}\right)\\ &+ \tfrac{\left(a_2^2 b_1^2 + a_1^2 b_2^2\right)\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2} - 2a_1 a_2 b_1 b_2\left(\Sigma_{\mathbf{X}}\right)^2_{1,2}}{2b_1^2 b_2^2\left[\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2} - \left(\Sigma_{\mathbf{X}}\right)^2_{1,2}\right]}. \end{aligned} \quad (7)$$

Note that whenever $\left(\Sigma_{\mathbf{X}}\right)_{1,2} = 0$

$$\begin{aligned}D(f_{(Y_1,Y_2)}||f_{(Y_1',Y_2')}) &= \tfrac{1}{2}\left[-1 + \tfrac{a_1^2}{b_1^2} - \log\left(\tfrac{a_1^2}{b_1^2}\right)\right.\\ &\left. -1 + \tfrac{a_2^2}{b_2^2} - \log\left(\tfrac{a_2^2}{b_2^2}\right)\right], \end{aligned}$$

which also follows the intuition for the KLD of multivariate Gaussian distributions of diagonal covariance matrices.

The scenario $Y_1 = a_1 X_1 + Z_1$, $Y_2 = a_2 X_2 + a_3 X_1 + Z_2$ (so $Y_2 \neq g_2(X_2)$), where a_2 and a_3 are known, was also studied, although the obtained results are not shown here due to spatial constraints. Let only mention that it corresponds to the stereo case, where channel 2 is not only equalized, but edited by combining it with a filtered version of channel 1. Our target would be to estimate the equalizer undergone by output 1, that depends only on channel 1 input.

D. Scenario $Y_1 = a_1 X_1 + Z_1$, $Y_2 = a_2 X_2 + Z_2$. a_2 is not known

Application Scenario: stereo, we want to estimate the equalizer applied to one of the channels, but we do not know about the equalization applied to the other one.

In this framework, the value of a_2 (as a function of a_1) maximizing the ML target function is

$$\frac{-\xi + \sqrt{\xi^2 + 4Na_1\left[\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2} - \left(\Sigma_{\mathbf{X}}\right)^2_{1,2}\right]\sum_{i=1}^N a_1\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(Y_2^i\right)^2}}{2Na_1\left[\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2} - \left(\Sigma_{\mathbf{X}}\right)^2_{1,2}\right]},$$

where $\xi \triangleq \sum_{i=1}^{N} \left(\Sigma_{\mathbf{X}}\right)_{1,2} Y_1^i Y_2^i$, yielding the ML estimator

$$\hat{a}_1 = \pm \sqrt{\frac{\left(\Sigma_{\mathbf{X}}\right)_{2,2} \kappa - \sqrt{\frac{\left(\Sigma_{\mathbf{X}}\right)_{2,2}}{\left(\Sigma_{\mathbf{X}}\right)_{1,1}} \left(\Sigma_{\mathbf{X}}\right)_{1,2}^2 \kappa \left[\sum_{i=1}^{N} Y_1^i Y_2^i\right]^2}}{N\left[\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2} - \left(\Sigma_{\mathbf{X}}\right)_{1,2}^2\right] \sum_{i=1}^{N} \left(Y_2^i\right)^2}},$$

where $\kappa \triangleq \left[\sum_{i=1}^{N} \left(Y_1^i\right)^2\right] \left[\sum_{i=1}^{N} \left(Y_2^i\right)^2\right]$. Note that whenever $\left(\Sigma_{\mathbf{X}}\right)_{1,2} = 0$, then this estimator is equivalent to (2).

On the other hand, in the computation of the KLD, and given that we study the distinguishability between the processing corresponding to a_1 and b_1, we will look for those values of a_2 and b_2 minimizing the KLD. In the current scenario, the derivative of the KLD with respect to a_2 is

$$-\frac{1}{a_2} + \frac{b_1 a_2 \left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2} - a_1 b_2 \left(\Sigma_{\mathbf{X}}\right)_{1,2}^2}{b_1 b_2^2 \left[\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2} - \left(\Sigma_{\mathbf{X}}\right)_{1,2}^2\right]},$$

which has roots with respect to a_2 at $\frac{b_2\left(a_1\left(\Sigma_{\mathbf{X}}\right)_{1,2}^2 + \gamma\right)}{2b_1\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2}}$, where $\gamma \triangleq \sqrt{a_1^2 \left(\Sigma_{\mathbf{X}}\right)_{1,2}^4 + 4b_1^2 \left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2}\left[\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2} - \left(\Sigma_{\mathbf{X}}\right)_{1,2}^2\right]}$. If one replaces (IV-D) into the relative entropy, the result does not depend on b_2, so the minimization over that variable is indeed not necessary. The obtained value is

$$D(f_{(Y_1,Y_2)}||f_{(Y_1',Y_2')}) = -\tfrac{1}{2} + \frac{a_1^2\left[2\left(\Sigma_{\mathbf{X}}\right)_{1,1}^2\left(\Sigma_{\mathbf{X}}\right)_{2,2}^2 - \left(\Sigma_{\mathbf{X}}\right)_{1,2}^4\right] - a_1\left(\Sigma_{\mathbf{X}}\right)_{1,2}^2 \kappa}{4b_1^2\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2}\left[\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2} - \left(\Sigma_{\mathbf{X}}\right)_{1,2}^2\right]} - \tfrac{1}{2}\log\left(\frac{a_1^2\left[a_1\left(\Sigma_{\mathbf{X}}\right)_{1,2}^2 + \kappa\right]^2}{4b_1^4\left(\Sigma_{\mathbf{X}}\right)_{1,1}^2\left(\Sigma_{\mathbf{X}}\right)_{2,2}^2}\right).$$

An interesting scenario, is that where $\left(\Sigma_{\mathbf{X}}\right)_{1,2} = 0$; under that assumption, the derivative of the KLD with respect to a_2 is equal to $-\frac{1}{a_2} + \frac{a_2}{b_2^2}$, yielding the condition $a_2^2 = b_2^2$. Indeed, in that particular framework the KLD can be written as (check the obvious relationships with (3)) $D(f_{(Y_1,Y_2)}||f_{(Y_1',Y_2')}) = \frac{1}{2}\left(-1 + \frac{a_1^2}{b_1^2} - \log\left[\frac{a_1^2}{b_1^2}\right]\right) + \frac{1}{2}\left(-1 + \frac{a_2^2}{b_2^2} - \log\left[\frac{a_2^2}{b_2^2}\right]\right)$, and consequently we can minimize the KLD over $\frac{a_2^2}{b_2^2}$; straightforwardly, the achieved solution is $\frac{a_2^2}{b_2^2} = 1$, providing a null contribution to the total KLD, whose value will be

$$D(f_{(Y_1,Y_2)}||f_{(Y_1',Y_2')}) = \frac{1}{2}\left(-1 + \frac{a_1^2}{b_1^2} - \log\left[\frac{a_1^2}{b_1^2}\right]\right), \qquad (8)$$

i.e., the same value achieved in (3). The implications of this result are evident:

- Since X_1 and X_2 are independent, the consideration of Y_2 and Y_2' will not provide any knowledge that can improve the distinguishability between a_1 and b_1.
- So, if we look for those values of a_2 and b_2 minimizing the KLD, we will find that $a_2^2 = b_2^2$ (due to the symmetry obtained by assuming zero-mean random variables).
- Consequently, we go back to the framework studied in Sect. IV-A.

Another asymptotical scenario, that is also particularly interesting, is that where $\left(\Sigma_{\mathbf{X}}\right)_{1,2} \to \pm\sqrt{\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2}}$, i.e., if X_1 and X_2 are (almost) deterministically related. It can be checked that in that framework the KLD goes to infinity whenever $a_2 \neq \frac{a_1 b_2}{b_1}$. The intuition behind this result is also interesting: since X_1 and X_2 are related by a fixed factor, if we compute $\frac{Y_1}{Y_2} = \frac{a_1}{a_2}$ it will be trivial to distinguish that scenario from $\frac{Y_1'}{Y_2'} = \frac{b_1}{b_2}$ unless $\frac{a_1}{a_2} = \frac{b_1}{b_2}$.[1]

Therefore, in order to follow our worst case approach, we will choose a_2 to be $\frac{a_1 b_2}{b_1}$. By doing so, the resulting KLD value is $D(f_{(Y_1,Y_2)}||f_{(Y_1',Y_2')}) = -1 + \frac{a_1^2}{b_1^2} - \log\left[\frac{a_1^2}{b_1^2}\right]$, which is nothing but twice (3) (and therefore twice (8)), and exactly the same as (5), although in that case this value was obtained for a generic covariance matrix.

Again, this result illustrates what one would intuitively expect; the larger the correlation between the sources, the easier it will be to distinguish between the operators. Indeed, the two limit behaviors are also very enlightening:

- Whenever the considered sources are independent, the achieved KLD is equivalent to that where only Y_1 is considered. Of course in this framework Y_2 does not provide any knowledge on Y_1, and consequently we can just neglect that variable.
- Whenever the relationship between the sources is deterministic, the problem is equivalent to having two independent observations, coming from a single source.

V. Numerical results

In this section we will provide a glance at some of those Gaussian linear cases that have not been studied in the previous section due to their cumbersome mathematical expressions. First of all, we will consider the effect of the processing noise $\mathbf{Z}$. The solid lines in Fig. 3 show the results obtained when $a_2 = a_1$ (correspondingly, $b_2 = b_1$), i.e., the scenario studied in Sect. IV-B. Be aware that in that framework, and as it was previously discussed, the fact of considering twice the same processing operator helps to our estimation. Nevertheless, if the observations are noisy, the closer they are, the more difficult will be to appreciate the different information that each observation provides; indeed, when they get very close (i.e., when $\left(\Sigma_{\mathbf{X}}\right)_{1,2} \to \sqrt{\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2}}$) the distinguishability will be equivalent to having a single observation (the KLD decreases to half the value we have for $\left(\Sigma_{\mathbf{X}}\right)_{1,2} = 0$). This illustrates that a very high correlation between sources is not always positive for distinguishability.

On the other hand, Fig. 3 also contains the results when a_2 and b_2 are not known, i.e., the scenario considered in Sect. IV-D; the curve obtained for $\sigma_Z^2 = 0$ corresponds to the results derived there. As mentioned in Sect. III we have

[1] Take into account that we assume this equality to hold before taking the limit $\left(\Sigma_{\mathbf{X}}\right)_{1,2} \to \pm\sqrt{\left(\Sigma_{\mathbf{X}}\right)_{1,1}\left(\Sigma_{\mathbf{X}}\right)_{2,2}}$. In the limit the covariance matrix of $\mathbf{X}$ becomes singular (with computational problems arising when computing (1)), so it is important to note that by considering $\frac{a_1}{a_2} = \frac{b_1}{b_2}$ the KLD does not depend on the covariance matrix of $\mathbf{X}$.

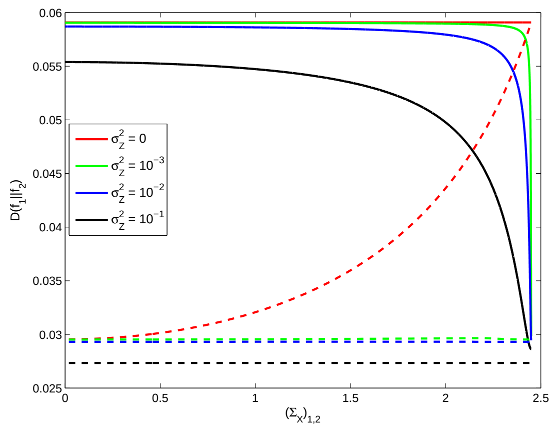

Fig. 3. KLD when $a_2 = a_1$ and $b_2 = b_1$ (solid lines) and when a_2 and b_2 are not known (dashed ones), for different values of $\sigma^2_{Z_1} = \sigma^2_{Z_2} = \sigma^2_Z$. $a_1 = 1$, $b_1 = 1.2$, $(\Sigma_{\mathbf{X}})_{1,1} = 2$, $(\Sigma_{\mathbf{X}})_{1,1} = 3$, $(\Sigma_{\mathbf{X}})_{1,2} \in \left[0, \sqrt{(\Sigma_{\mathbf{X}})_{1,1} (\Sigma_{\mathbf{X}})_{2,2}}\right]$, $\boldsymbol{\mu}_{\mathbf{X}} = \mathbf{0}$, $\boldsymbol{\mu}_{\mathbf{Z}} = \mathbf{0}$.

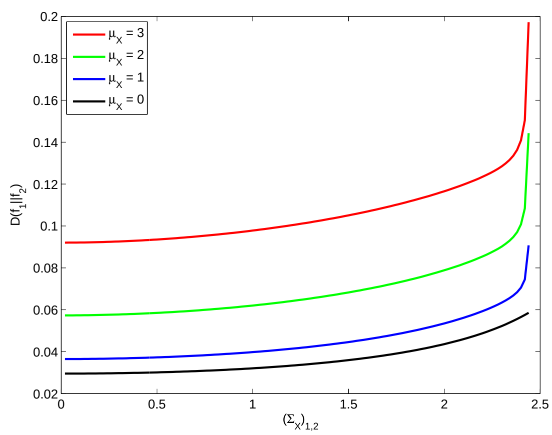

Fig. 4. KLD when a_2 and b_2 are not known, for $\sigma^2_{Z_1} = \sigma^2_{Z_2} = 0$ and different values of $\mu_{X_1} = \mu_{X_2} = \mu_X$. $a_1 = 1$, $b_1 = 1.2$, $(\Sigma_{\mathbf{X}})_{1,1} = 2$, $(\Sigma_{\mathbf{X}})_{1,1} = 3$, $(\Sigma_{\mathbf{X}})_{1,2} \in \left[0, \sqrt{(\Sigma_{\mathbf{X}})_{1,1} (\Sigma_{\mathbf{X}})_{2,2}}\right]$, $\boldsymbol{\mu}_{\mathbf{Z}} = \mathbf{0}$.

decided to follow a worst case approach for this scenario. Indeed, for the noisy case the values of a_2 and b_2 minimizing the KLD are $a_2 = 0$ and $b_2 = 0$; the intuitive idea behind this result is also clear: in the worst case we cannot trust the second observation, as it is only noise. In that case the KLD is $\frac{1}{2}\left[\frac{a_1^2 (\Sigma_{\mathbf{X}})_{1,1} + \sigma_Z^2}{b_1^2 (\Sigma_{\mathbf{X}})_{1,1} + \sigma_Z^2} - \log\left(\frac{a_1^2 (\Sigma_{\mathbf{X}})_{1,1} + \sigma_Z^2}{b_1^2 (\Sigma_{\mathbf{X}})_{1,1} + \sigma_Z^2}\right) - 1\right]$, which is independent of the correlation term, as expected.

Finally, the influence of the mean of the original signal on the distinguishability is illustrated in Fig. 4, which clearly shows that the larger the mean of the signal, the easier will be to distinguish the considered processing operators.

VI. Conclusions

In this work we have quantified the advantages of using the joint distribution of composite objects for improving the distinguishability between processing operators. Although for the sake of tractability we have focused on the linear Gaussian case, it is evident that the principles derived here would be applicable to more general frameworks. Among these principles, we can mention the behavior of the distinguishability measures in different scenarios, and how, for example, the case where the correlated sources are known to share their processing is equivalent to having independent sources. It also interesting to note that for the case where the unwanted processing is unknown, for the noiseless case the distinguishability achieved for deterministically correlated sources is double the one obtained for independent sources, but for the noisy case the obtained value is independent of that correlation, as the second observation is considered to be pure noise. Finally, we would like to mention the *a priori* striking result showing that a larger correlation between sources does not always imply a better distinguishability between operators.

Acknowledgments

Thanks are due to Prof. Mauro Barni for pointing out the links with [12]. This research was supported by the European Union under project REWIND (Grant Agreement Number 268478), the European Regional Development Fund (ERDF) and the Spanish Government under projects DYNACS (TEC2010-21245-C02-02/TCM) and COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010), and the Galician Regional Government under projects "Consolidation of Research Units" 2009/62, 2010/85 and SCALLOPS (10PXIB322231PR).

References

[1] "Special issue on digital forensics," *IEEE Signal Processing Magazing*, vol. 26, no. 2, March 2009.

[2] S. Milani, P. F. Piazza, M. Tagliasacchi, and S. Tubaro, "An audio-video compatibility test for multimedia tampering detection," submitted to ACM Information Hiding and Multimedia Security 2013.

[3] J. Choi, G. Friedland, V. Ekambaram, and K. Ramchandran, "Multimodal location estimation of consumer media: Dealing with sparse training data," in *Proc. of IEEE ICME*, July 2012, pp. 43–48.

[4] M. Kirchner, P. Winkler, and H. Farid, "Impeding forgers at photo inception," in *Proc. of the SPIE. Media Watermarking, Security and Forensics*, vol. 8665, February 2013.

[5] H. L. van Trees, *Detection, Estimation, and Modulation Theory.* John Wiley and Sons, 1968.

[6] M. Barni, "A game theoretic approach to source identification with known statistics," in *Proc. of IEEE ICASSP*, Kyoto, Japan, March 2012, pp. 1745–1748.

[7] R. Böhme and M. Kirchner, *Digital Image Forensics.* Springer, 2012, ch. Counter-Forensics: Attacking Image Forensics.

[8] P. Comesaña, "Detection and information theoretic measures for quantifying the distinguishability between multimedia operator chains," in *Proc. of IEEE WIFS*, Tenerife, Spain, December 2012, pp. 211–216.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley, 2006.

[10] D. Slepian and J. W. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[11] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, January 1976.

[12] M. S. Rahman and A. B. Wagner, "on the optimality of binning for distributed hypothesis testing," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6282–6303, October 2012.

[13] R. Neelamani, R. de Queiroz, Z. Fan, S. Dash, and R. G. Baraniuk, "JPEG compression history estimation for color images," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1365–1378, June 2006.