A novel formulation of Independence Detection based on the Sample Characteristic Function

Ferran de Cabrera, *Student Member, IEEE* and Jaume Riba, *Senior Member, IEEE*

Signal Theory and Communications Department, Technical University of Catalonia (UPC)

{ferran.de.cabrera, jaume.riba}@upc.edu

Abstract—A novel independence test for continuous random sequences is proposed in this paper. The test is based on seeking for coherence in a particular fixed-dimension feature space based on a uniform sampling of the sample characteristic function of the data, providing significant computational advantages over kernel methods. This feature space relates uncorrelation and independence, allowing to analyze the second order statistics as it is encountered in traditional signal processing. As a result, the possibility of utilizing well known correlation tools arises, motivating the usage of Canonical Correlation Analysis as the main tool for detecting independence. Comparative simulation results are provided using a model based on fading AWGN channels.

Index Terms—Independence Detection, Characteristic Function, Mutual Information, ITL, Kernel, HSIC, CCA.

I. INTRODUCTION

Methods for measuring and/or detecting statistical dependence, a problem that has a long history, has recently found applications in a wide variety of areas ([10][19] and references therein). Some important examples from the broad literature are the Maximal Information Coefficient (MIC) [11], the distance Correlation (dCor) [16], the Mutual Information Dimension (MID) [15] and the Hilbert-Schmidt Independence Criterion (HSIC) [5]. Detecting dependence has a close link with machine learning, which, in general terms, tries to discover structure in data. In both cases the objective is to construct a non-linear mapping from the data onto a higher-dimensional space (referred to as feature space) in which, in some sense, the original problem becomes simpler, resorting the Cover's theorem [2]. For the specific problem of detecting statistical dependence, the objective of the mapping is to move the original problem into a simpler problem of measuring second order statistics.

Kernel methods ([17][6][1]), which are based on the rigorous mathematical framework of reproducing kernel Hilbert spaces (RKHS), provide an elegant solution for this mapping known as kernel trick: replacing inner products by positive-definite kernels is equivalent to implicitly processing the data on an extremely high dimensional feature space. However, apart from the need of regularization to control the smoothness of the results, an important drawback of kernel methods is their compromised computational complexity for large data sets, which requires the use of some special techniques to

avoid the need of accessing to all pairs of input data points, such as the Incomplete Cholesky decomposition [13].

This paper proposes an alternative to the kernel trick based on a particular feature space. Following the rationale of the Information Theoretic Learning (ITL) issued in [14], the objective is to define a detector based on quadratic measures on the space of the Characteristic Function (CF), and we will show that we are capable of defining an independence test based solely on first and second order statistics in this space.

The mapping is based on uniformly sampling the CF whose fixed-dimension is independent of the data size, allowing to ameliorate the complexity issues raised by the kernels. This feature space resembles an analog memory-less frequency-modulation of the data, giving rise to second-order auto-correlation matrices, which, jointly with the link between uncorrelation and independence, allows to expand the study to the Canonical Correlation Analysis (CCA), which is shown to avoid the well-known overfitting problem that arises in Kernel CCA (K-CCA) [7]. The rationale is similar to the study of independent component analysis under a CF approach present in [3], in which the empirical joint CF is compared to the marginal ones. However, while their work is based on expanding the analysis of the CF through its cumulants from the Taylor expansion around zero, we are more interested on studying the sampling of the feature space in order to handle the problem with classical signal processing.

Finally, we provide a comparative analysis of the method with respect to other known methods, in addition with a genie-aided detector based on the sample Mutual Information (MI) estimation with known Probability Density Function (PDF). To this end, three data models based on Gaussian mixtures are defined at different Signal to Noise Ratio (SNR) regimes, which are inspired on fading AWGN communications channels.

II. SAMPLE CHARACTERISTIC FUNCTION AS A FEATURE SPACE

Let $f_{x,y}(x,y)$ be the joint PDF of the random variables $x, y \in \mathbb{R}$. The joint CF is

$$\phi_{x,y}(\omega_1; \omega_2) = \int \int_{\mathbb{R}} f_{x,y}(x,y) e^{j\omega_1 x + j\omega_2 y} dx dy$$
$$= E\left[e^{j\omega_1 x + j\omega_2 y}\right] = E\left[e^{j\omega_1 x} e^{j\omega_2 y}\right] \quad (1)$$

for any $\omega_1, \omega_2 \in \mathbb{R}$. A necessary and sufficient condition for independence is the separability of the CF $\phi_{x,y}(\omega_1; \omega_2) =$

$\phi_x(\omega_1)\phi_y(\omega_2) = E\left[e^{j\omega_1 x}\right] E\left[e^{j\omega_2 y}\right]$. Considering the random variables $z_1 = e^{j\omega_1 x}$ and $z_2 = e^{j\omega_2 y}$ the condition for independence can be written as

$$E\left[(z_1 - E[z_1])(z_2 - E[z_2])^*\right] = 0 \qquad (2)$$

which constitutes the requisite for uncorrelation. This property succeeds in moving the independence condition to an equivalent uncorrelation one, which will be used as the basis of independence criteria.

A. Sample characteristic function analysis

Consider $\{x_i\}_{i=1...N}$ a N-length sequence composed by i.i.d. data observations drawn from the random variable x with PDF $f_x(x)$ and associated CF $\phi_x(\omega) = \int_{-\infty}^{\infty} f_x(x)e^{jx\omega}dx$. Since the CF exists at all values of ω, we are interested in studying under which conditions a finite dimensional feature space is sufficient for detecting independence.

Taking a specific sample x_i, the windowed sample CF is

$$\phi_{x_i}(\omega) = e^{j\omega x_i}G(\omega) \qquad (3)$$

being $G(\omega)$ an even and positive window function with $G(0) = 1$ and $\int |G(\omega)|\, d\omega < \infty$. The function $G(\omega)$ assumes that most of the informative part of the CF is concentrated around the window bandwidth σ. Let us consider an additional mapping based on the Fourier Transform (FT):

$$\Phi_i(\beta) = \int_{-\infty}^{\infty} \phi_{x_i}(\omega)e^{-j\beta\omega}d\omega = \int_{-\infty}^{\infty} G(\omega)e^{-j\omega(\beta-x_i)}d\omega \quad (4)$$

Defining $g(\beta) = \int G(\omega)e^{-j\omega\beta}d\omega$ as the FT of $G(\omega)$, then we have

$$\Phi_i(\beta) = g(\beta - x_i) \qquad (5)$$

corresponding to a real function $g(\beta)$ centered around the data value x_i.

Let us now define the first and second order statistics. The expectation over the data samples is

$$\mu_\Phi(\beta) = E[\Phi_i(\beta)] = \int_{-\infty}^{\infty} g(\beta - x) f_x(x)\, dx$$

$$= (f_x * g)(\beta) \qquad (6)$$

and its empirical version is given by

$$\hat{\mu}_\Phi(\beta) = \frac{1}{N}\sum_{i=1}^{N} g(\beta - x_i) \qquad (7)$$

which is in fact a kernel density estimate [9] given $g(\beta)$ is a unitary area function. This result ensures that the expected value of the FT of the sample CF is a smoothed estimate of the original PDF.

We define the second order statistic as:

$$r_\Phi(\beta_1, \beta_2) = E[\Phi_i(\beta_1)\Phi_i^*(\beta_2)]$$

$$= \int_{-\infty}^{\infty} g(\beta_1 - x) g^*(\beta_2 - x) f_x(x)\, dx \qquad (8)$$

It is worth noting that for $g(\beta)$ narrow enough, a data association phenomenon occurs such that

$$r_\Phi(\beta_1, \beta_2) \approx 0 \quad \forall |\beta_1 - \beta_2| > \varepsilon \qquad (9)$$

for any small positive $\varepsilon > 0$, whose insight is to measure closeness between data, and resembling the pairwise sample interactions present in [10]. If we restrict Eq. (8) to lay in the same β, then we can express it as

$$r_\Phi(\beta) = \int_{-\infty}^{\infty} h(\beta - x) f_x(x)\, dx = (f_x * h)(\beta) \qquad (10)$$

and its empirical estimate as

$$\hat{r}_\Phi(\beta) = \frac{1}{N}\sum_{i=1}^{N} h(\beta - x_i) \qquad (11)$$

with $h(\beta) = |g(\beta)|^2$. Since the window $G(\omega)$ is expected to be wide enough in order to cover most of the informative part of the CF, then the support of its squared FT $h(\beta)$ tends to decrease. If we approximate $h(\beta)$ as a delta-like function such that $h(\beta) \simeq \delta(\beta)$ with $\int \delta(\beta)\, d\beta = k \gg 1$, then

$$r_\Phi(\beta) \simeq (f_x * \delta)(\beta) = k f_x(\beta) \qquad (12)$$

which is a scaled version of the original PDF.

Let us assume a new set of i.i.d. data $\{y_i\}_{i=1...N}$ extracted from the random variable y with PDF $f_y(y)$. The cross-correlation between the two data sets evaluated in the feature space can now be expressed in terms of (6) and (8). To this end, consider $d(x_i) = g(\beta_1 - x_i)$ and $d(y_i) = g(\beta_2 - y_i)$ and formulate the squared Pearson coefficient as follows:

$$\rho^2 = \frac{(E[d(x_i)d(y_i)] - E[d(x_i)]E[d(y_i)])^2}{(E[d^2(x_i)] - E^2[d(x_i)])(E[d^2(y_i)] - E^2[d(y_i)])} \qquad (13)$$

where $E[d(x_i)d(y_i)]$ is an approximate of the joint PDF $f_{xy}(x,y)$:

$$E[d(x_i)d(y_i)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(\beta_1 - x) g(\beta_2 - y)$$
$$\times f_{x,y}(x,y)\, dxdy \quad \simeq f_{x,y}(\beta_1, \beta_2) \qquad (14)$$

and the expectation $E[d(x_i)]$ is the same as in Eq. (6), which is approximately $f_x(\beta_1)$, and analogously $E[d(y_i)] \simeq f_y(\beta_2)$. By analyzing Eq. (13) with the same rationale and using Eq. (12) we get:

$$\rho^2 \rightarrow$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{(f_{x,y}(\beta_1,\beta_2) - f_x(\beta_1)f_y(\beta_2))^2}{(kf_x(\beta_1) - f_x^2(\beta_1))(kf_x(\beta_2) - f_x^2(\beta_2))} d\beta_1 d\beta_2 \qquad (15)$$

Interestingly, if we let k to increase without limit, then the squared Pearson coefficient tends to

$$\rho^2 \xrightarrow{k\to\infty} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{(f_{x,y}(\beta_1,\beta_2) - f_x(\beta_1)f_y(\beta_2))^2}{k^2 f_x(\beta_1) f_y(\beta_2)} d\beta_1 d\beta_2$$

$$= \frac{1}{k^2} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \left(\frac{f_{x,y}(x,y)}{f_x(x) f_y(y)} - 1 \right)^2 f_x(x) f_y(y) \, dx dy \quad (16)$$

which results in the kernel-free expression of the Normalized Cross-Covariance Operator (NOCCO) developed by Fukumizu et al. ([4]) that studies the dependence structure based on normalized cross-covariance operators.

Eq. (16) constitutes an indicator of independence between random variables through the joint PDF and the product of the marginal ones, and compiles it as a correlation measure based on the squared Pearson coefficient. The possibility of drawing correlation metrics from the first and second order statistics in the proposed feature space, as well as the capability of linking uncorrelation and independence in terms of the CF, motivates us to explore the CCA from the uniform sampling of the CF, as we will see in Section III.

B. Uniform sampling

In order to operate with the CF from data samples, we will proceed by applying a uniform sampling of the sample feature space $\phi_{x_i}(\omega_1)$, and equivalently for $\phi_{y_i}(\omega_2)$, by evaluating ω_1 and ω_2 for $n\alpha$ being $\alpha \in \mathbb{R}$ the sampling period and $n = -(N_z - 1)/2, ..., (N_z - 1)/2$ with N_z any real number. Then, we construct the mapped data matrices $\boldsymbol{Z}_x = [\mathbf{z}_x(1), \mathbf{z}_x(2), ..., \mathbf{z}_x(N)] \in \mathbb{C}^{N_z \times N}$ and $\boldsymbol{Z}_y = [\mathbf{z}_y(1), \mathbf{z}_y(2), ..., \mathbf{z}_y(N)] \in \mathbb{C}^{N_z \times N}$ containing in their i-th column the mapping of the CF for a given sample[1]:

$$[\mathbf{z}_x(i)]_n = e^{jn\alpha x_i}$$
$$[\mathbf{z}_y(i)]_n = e^{jn\alpha y_i} \quad (17)$$

This sampling assumes that the relevant CF space is contained in a support of $N_z\alpha$, and so the feature space dimension N_z is sufficiently large for the detection task but still lower than the data dimensionality ($N_z \ll N$) in order to reduce the computational complexity from $O(N^2)$ present in kernel methods to $O(N_z N)$. Moreover, this assumption also minimizes the aliasing induced by the sampling and the Inverse Fourier Transform intrinsic in the CF. Considering the space between replicas to be $1/\alpha$, the inverse of the sampling period, then we must ensure that $\alpha < 1/\sigma$, with σ corresponding to the window bandwidth. Note also that it is possible to sample asymmetrically with $n = 1, ..., N_z$ since the CF is Hermitian.

The sample mean vectors evaluated in the feature map are then $\bar{\mathbf{z}}_x = \boldsymbol{Z}_x \mathbf{1}_N / N$ and $\bar{\mathbf{z}}_y = \boldsymbol{Z}_y \mathbf{1}_N / N$, auto-covariance matrices $\hat{\boldsymbol{C}}_x = \boldsymbol{Z}_x \boldsymbol{P} \boldsymbol{Z}_x^H / N$ and $\hat{\boldsymbol{C}}_y = \boldsymbol{Z}_y \boldsymbol{P} \boldsymbol{Z}_y^H / N$, and the cross-covariance matrices $\hat{\boldsymbol{C}}_{xy} = \boldsymbol{Z}_x \boldsymbol{P} \boldsymbol{Z}_y^H / N$ and $\hat{\boldsymbol{C}}_{yx} = \hat{\boldsymbol{C}}_{xy}^H$ being $\mathbf{1}_N$ a $N \times 1$ column vector of ones, $\boldsymbol{P} = \boldsymbol{I}_N - \mathbf{1}_N \mathbf{1}_N^T / N$ the projection matrix and $\boldsymbol{I}_N$ the $N \times N$ identity matrix.

III. CANONICAL CORRELATION ANALYSIS

Given two data matrices $\boldsymbol{Z}_x, \boldsymbol{Z}_y \in \mathbb{C}^{N_z \times N}$, the CCA problem is based on finding the canonical variables $\boldsymbol{u}$ and $\boldsymbol{v}$

[1]Both even and odd N_z are permitted. Although an odd real value induces bias given by $[\mathbf{z}(i)]_{n=0} = 1$, it does not impact the performance of the detector since the bias is known and can be fixed.

Algorithm 1 Inversion algorithm pseudocode.

Input: $\boldsymbol{Z}$ matrix of dimension $N_z \times N$
Initialize: $\boldsymbol{R}^{-1}(0) = \delta \boldsymbol{I}_{N_z}$ with $\delta \gg 1$
for $i = 0, 1, ..., N - 1$ **do**
    $g = 1 + \mathbf{z}^H(i) \boldsymbol{R}^{-1}(i) \mathbf{z}(i)$
    $\boldsymbol{R}^{-1}(i+1) \leftarrow \boldsymbol{R}^{-1}(i) \mathbf{z}(i) \mathbf{z}^H(i) \boldsymbol{R}^{-1}(i) / g$
end

---

that maximize the correlation between the canonical variates $\boldsymbol{w}_x = \boldsymbol{u}^T \boldsymbol{Z}_x$ and $\boldsymbol{w}_y = \boldsymbol{v}^T \boldsymbol{Z}_y$, i.e.

$$\arg\max_{\boldsymbol{u},\boldsymbol{v}} \rho^2 = \left| \frac{\boldsymbol{u}^T \hat{\boldsymbol{C}}_{xy} \boldsymbol{v}}{\sqrt{\left(\boldsymbol{u}^T \hat{\boldsymbol{C}}_x \boldsymbol{u}\right) \left(\boldsymbol{v}^T \hat{\boldsymbol{C}}_y \boldsymbol{v}\right)}} \right|^2 \quad (18)$$

where ρ^2 corresponds to the squared maximum canonical correlation. The solution to this problem, and thereby the canonical variates, is given by the Singular Value Decomposition (SVD) of the Coherence matrix, a whitened version of the cross-covariance matrix, defined as

$$\Delta = \hat{\boldsymbol{C}}_x^{-1/2} \hat{\boldsymbol{C}}_{xy} \hat{\boldsymbol{C}}_y^{-1/2} \quad (19)$$

The eigenvalues of Δ are the non-negative roots of the eigenvalues $\lambda_{m=1,...,N_z}$ of the squared Coherence matrix $\boldsymbol{K} = \Delta \Delta^H$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_z} \geq 0$ ([12]), corresponding to all the canonical correlations ρ_m^2 with the highest being ρ^2. Thus, the maximum squared canonical correlation ρ^2 is given by the largest eigenvalue of $\boldsymbol{K}$, and so we can evaluate the main solution as:

$$\lambda_1 = \lambda_{max} \left(\hat{\boldsymbol{C}}_x^{-1/2} \hat{\boldsymbol{C}}_{xy} \hat{\boldsymbol{C}}_y^{-1} \hat{\boldsymbol{C}}_{xy}^H \hat{\boldsymbol{C}}_x^{-1/2} \right)$$
$$= \lambda_{max} \left(\hat{\boldsymbol{C}}_x^{-1} \hat{\boldsymbol{C}}_{xy} \hat{\boldsymbol{C}}_y^{-1} \hat{\boldsymbol{C}}_{yx} \right) \quad (20)$$

However, the auto-covariance matrices $\hat{\boldsymbol{C}}_x$ and $\hat{\boldsymbol{C}}_y$ are ill-conditioned, making the inversion task difficult under real simulations. In this case, the inversion of the auto-covariance matrices can be solved through Algorithm 1 by means of the Woodbury matrix identity ([8]).

Finally, we can build the Characteristic Canonical Independence Test (CCIT) as the maximum canonical correlation:

$$CCIT(\boldsymbol{x}, \boldsymbol{y}, N_z, \alpha) = \lambda_{max} \left(\hat{\boldsymbol{C}}_x^{-1} \hat{\boldsymbol{C}}_{xy} \hat{\boldsymbol{C}}_y^{-1} \hat{\boldsymbol{C}}_{yx} \right) \quad (21)$$

Then, the proposed detector is

$$CCIT(\boldsymbol{x}, \boldsymbol{y}, N_z, \alpha) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma \quad (22)$$

being γ the detector threshold, $\mathcal{H}_0$ the independence hypothesis and $\mathcal{H}_1$ the dependence hypothesis.

Note that we can also evaluate the trace of $\boldsymbol{K}$ as a detector such as $CCIT_2(\boldsymbol{x}, \boldsymbol{y}, N_z, \alpha) = tr(\boldsymbol{K})$ since it contains all the subsequent canonical correlations ρ_m^2 from the CCA and permitting to avoid the SVD of $\boldsymbol{K}$. But, this test contributes with a bias given the addition of the lesser eigenvalues. However, since the canonical correlations are concentrated at the largest eigenvalues, the impact to the test is minimal as we will see in the numerical results.
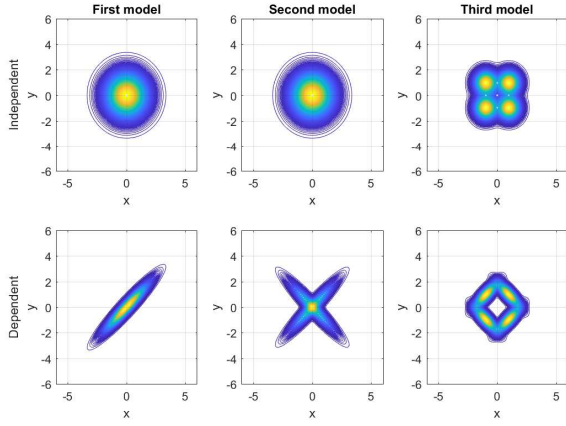
Figure 1. Contour plots of the joint probability functions with $SNR = 8$ and $p = 0.95$ in the third model.

Figure 2. True MI vs. SNR of the three models obtained by the unbiased genie-aided estimator given in Eq. (24) with $N = 10^6$.

IV. NUMERICAL RESULTS

In order to evaluate the detector, the data is modeled as:

$$x(i) = h_x(i)\left(-\sqrt{p} + \sqrt{1-p}z_x(i)\right) + w_x(i)$$
$$y(i) = h_y(i)\left(\sqrt{p} + \sqrt{1-p}z_y(i)\right) + w_y(i) \qquad (23)$$

corresponding to fast fading AWGN channels, with $z_x, z_y \sim \mathcal{N}(0,1)$ independent from each other under independence hypothesis $\mathcal{H}_0$ and equal under dependence hypothesis $\mathcal{H}_1$, and the noise $w_x, w_y \sim \mathcal{N}(0, \sigma_w^2)$ mutually independent for both hypothesis. The intention is to generate data with different kinds of associations through the parameter p and the channels h_x and h_y, following the examples in [11], while maintaining a close relation with a communication channel. Thus, we are able to relate the difficulty of detecting independence with the SNR. Additionally, each one of the three models, observed in Figure 1, holds the same marginal PDFs for both independence and dependence hypothesis.

The basic and first model consists on the presence of correlation and dependence, represented by a linear structure of the data. For this model, the parameters are $p = 0$, $h_x(i) = h_y(i) = 1$ and its MI is given by $I(x,y) = \frac{1}{2}\ln\left(1 + 4SNR^2/(1+4SNR)\right)$ with $SNR = 1/(2\sigma_w^2)$, which for high SNR is asymptotically the MI of an AWGN channel $I(x,y) = \frac{1}{2}\ln(1 + SNR)$.

The second model is based on the dependence of variables while being uncorrelated. The intention is to evaluate if the CCIT can properly detect dependence under non-linear associations of the data. For this model, $p = 0$ and $h_x(i)$, $h_y(i)$ are binary distributed with equal probability $[-1,1]$ and independent to each other, which act as fast fading channels.

Lastly, the third model provides uncorrelation but with an increased difficulty of detection. For this model $h_x(i)$ and $h_y(i)$ are distributed as the second model but the parameter $0 < p < 1$ determines the difficulty of detection, being $p = 0$ akin to the second model and $p = 1$ nullifies the dependence.

The generated data is shaped as a Gaussian Mixture Model (GMM), allowing an explicit expression of the PDFs and
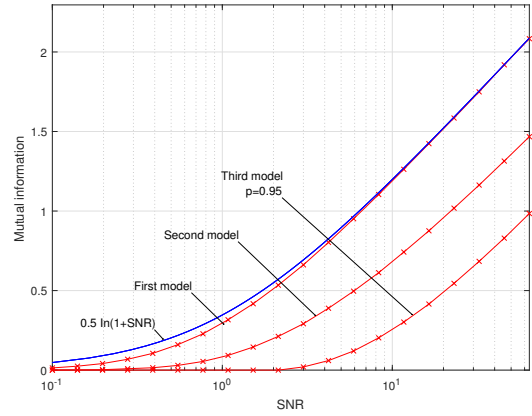
ensuring a consistent estimate of the MI based on the sample mean as

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N}\ln\frac{f_{x,y}(x_i, y_i)}{f_x(x_i) f_x(y_i)} \qquad (24)$$

We will compare the performance of the detector with a benchmark based on a threshold from the genie-aided statistic in Eq. (24), whose values for each model as can be seen in Figure 2. Note also that the proposed models provide an increased difficulty for detecting dependence, as appreciated by a higher threshold effect of MI vs. SNR.

Figure 3 provides a comparative analysis of the Probability of Detection P_D for multiple detectors for a fixed false alarm probability of 5%. The results were obtained with 400 Monte Carlo simulations and the parameters optimized for each case. For the HSIC, a Gaussian kernel was used with the bandwidth recommended in [5], which is the heuristic for these cases. Regarding the parameters of the CCIT, $N_z = 21$ and the sampling period $\alpha = 0.001$ were chosen, providing a trade-off between computational complexity and leakage in the form of aliasing in the FT.

Figure 3a shows the performance of the first model, focusing on the low SNR scenario. It can be seen that for linear associations of the data, HSIC and dCor slightly outperform the CCIT with all of them presenting a high robustness to noise, while MIC and MID are very sensitive to it. On the other hand, Figure 3b and 3c show that CCIT improves the detection rate with respect to other detectors for uncorrelated data. Specifically, CCIT provides a higher convergence for low SNR values, whose SNR threshold for a P_D close to 100% is near to that of the genie-aided estimate.

The underperformance of the CCIT for linear dependencies is caused by an increase of the condition number of the auto-covariance matrices given precisely by this linear relation, since the data mapped in the feature space is confined in a small dimensional subspace. Under non-linear dependencies, the feature space spreads the subspace dimension and the inverse of the matrices can be better estimated. This affectation can be solved by reducing the sampling period α, thus limiting the mapping into the desired subspace. Then, the CCIT is
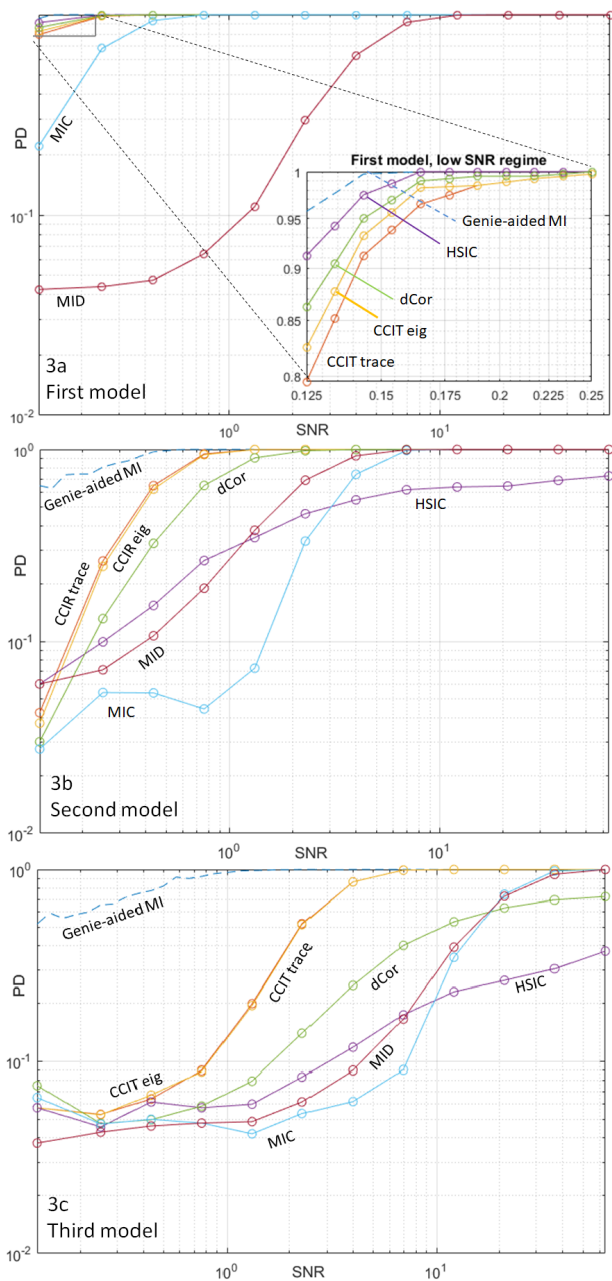
Figure 3. Probability of Detection for $P_{FA} = 5\%$ vs SNR from 0.125 to 64 of the three models. In dashed blue, the P_D of Eq. (24) estimate. ($N = 300, p = 0.95, N_z = 21, \alpha = 0.001$)

more aware of the linear dependence and the P_D is improved. However, in the figures a fixed value of α is chosen due to the seek of universality of the detector ([19]), and so to be fair in the comparison with the case of non-linear dependencies. Alternatively, the detection rate could be improved by avoiding the inverse process and measuring the canonical correlation through adaptive methods instead of CCA [18].

V. CONCLUSIONS

In this paper we have presented a feature space based on the uniform sampling of the marginal sample CFs. This mapping allowed us to take advantage of the relation between

uncorrelation and independence from the chosen feature space, and we have used a correlation measure to build the detector. We have shown that by applying a specific condition on the sample characteristic function we can relate, at least asymptotically, the proposed CCIT with other quadratic measures of information. The numerical simulations proved that this method is capable of improving the detection rate with respect to others in the literature, specifically under non-linear associations, and with a clear computational advantage over kernel signal processing. However, an exhaustive analysis on adaptive methods is needed in order to solve the numerical problems when inverting the ill-conditioned matrices. Moreover, the study of dependence under the proposed feature space for non i.i.d. data provides an interesting future research line.

REFERENCES

[1] F.R. Bach and M.I. Jordan. Kernel independent component analysis. *Journal of Machine Learning*, 3, pages 1-48, 2002.

[2] T.M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers, Volume: EC-14, Issue: 3, pages 326-334*, 1965.

[3] J. Eriksson and V. Koivunen. Characteristic-function-based independent component analysis. *Signal Processing*, 83(10):2195 – 2208, 2003.

[4] K. Fukumizu, A. Gretton, X. Sun, and B. Schölkopf. Kernel measures of conditional dependence. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 489–496. Curran Associates, Inc., 2008.

[5] A. Gretton, A. Smola, O. Bousquet, and B. Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. Technical report, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, 2005.

[6] A. Gretton, A. Smola, R. Herbrich, O. Bousquet, and B. Schölkopf. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 2005.

[7] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural Comput.*, 16(12):2639–2664, 2004.

[8] Monson H. Hayes. *Statistical Digital Signal Processing and Modeling*. Number 541-553. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1996, pp 541-553.

[9] E. Parzen. *On a Estimation of a Probability Density Function and Mode*, volume 33, 1065-1067. Ann. Math. Statist., 1962.

[10] J.C. Príncipe. *Information Theoretic Leraning*. Springer, 2010.

[11] D.N. Reshef, Y.A. Reshef, H.K. Finucane, S.R. Grossman, G. McVean, P.J. Turnbaugh, E.S. Lander, M. Mitzenmacher, and P.C. Sabeti. Detecting novel associations in large data sets. *Science*, 334(6062):1518–1524, December 2011.

[12] P.J. Schreier. A unifying discussion of correlation analysis for complex random vectors. *IEEE Transactions on Signal Processing*, 56(4), 2008.

[13] S. Seth and J.C. Príncipe. On speeding up computation in information theoretic learning. In *International Joint Conference on Neural Networks, IJCNN*, 2009.

[14] S. Seth, M. Rao, I. Park, and J.C. Príncpie. A unified framework for quadratic measures of independence. *IEEE Transactions on Signal Processing*, 59(8), 2011.

[15] M. Sugiyama and K.M Borgwardt. Measuring statistical dependence via the mutual information dimension. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)*, 2013.

[16] G.J Székely and M.L Rizzo. Brownian distance covariance. *Annals of Statistics*, 2010.

[17] S.V. Vaerenbergh. *Kernel Methods for Nonlinear Identification, Equalization and Separation of Signals*. PhD thesis, University of Cantabria, Department of Communications Engineering, 2009.

[18] J. Vía, I. Santamaría, and J. Pérez. A robust algorithm for adaptive canonical correlation analysis. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2005.

[19] Q. Wang, S.R.Kulkarni, and S.Verdú. *Universal Estimation of Information Measures for Analog Sources*. Now publishers, 2009.