

Measure Transformed Canonical Correlation Analysis with Application to Financial Data

Koby Todros and Alfred O. Hero III

Dept. of Electrical Engineering and Computer Science, University of Michigan, Ann-Arbor 48105, MI, U.S.A

Email: ktodros@umich.edu, hero@eecs.umich.edu

Abstract—In this paper, a new nonlinear generalization of linear canonical correlation analysis (LCCA) is derived. This framework, called measure transformed canonical correlation analysis (MTCCA), applies LCCA to the considered pair of random vectors after transformation of their joint probability distribution. The proposed transform is structured by a pair of nonnegative functions called the MT-functions. It preserves statistical independence and maps the joint probability distribution into a set of joint probability measures on the joint observation space. Specification of MT-functions in the exponential family, leads to MTCCA, which, in contrast to LCCA, is capable of detecting nonlinear dependencies. In the paper, MTCCA is illustrated for recovery of a nonlinear system with known structure, and for construction of networks that analyze long-term associations between companies traded in the NASDAQ and NYSE stock markets.

Index Terms—Association analysis, multivariate data analysis, probability measure transform.

I. INTRODUCTION

Linear canonical correlation analysis (LCCA) [1] is a technique for multivariate data analysis that quantifies the linear associations between a pair of random vectors. In particular, LCCA generates a sequence of pairwise linear combinations of the considered random vectors that have the following statistical properties under their joint probability distribution: 1) each linear combination has unit variance, 2) the correlation coefficient between the elements of each pair is maximal, and 3) each pair is uncorrelated with its predecessors. The coefficients of these linear combinations give insight into the underlying linear relationships between the random vectors. They are easily obtained by solving a simple generalized eigenvalue decomposition (GEVD) problem, which only involves the covariance and cross-covariance matrices of the considered random vectors. LCCA has been applied to blind source separation [2], image set matching [3], direction-of-arrival estimation [4], data fusion [5], audio-video synchronization [6], among others.

In cases where the considered random vectors are statistically dependent yet uncorrelated, LCCA is not an informative tool. In order to overcome this limitation, several nonlinear generalizations of LCCA have been proposed in the literature. In [7], an information-theoretic approach to canonical correlation analysis was proposed that replaces the correlation coefficient with the mutual-information (MI). The MI approach [7] is sensitive to nonlinear dependencies. However, in contrast to LCCA, it does not reduce to a simple GEVD problem. Indeed, in [7] each pair of linear combinations is obtained separately via an iterative Newton-Raphson algorithm, which may converge to undesired local maxima. Moreover, each step of the algorithm involves re-estimation of the MI in a non-parametric manner at a potentially high computational cost.

Another nonlinear generalization of LCCA is kernel canonical correlation analysis (KCCA) [8]. KCCA applies LCCA to high-dimensional nonlinear transformations of the considered random vectors that map them into some reproducing kernel Hilbert spaces. Although the KCCA approach can be successful in extracting nonlinear relations, it is highly prone to over-fitting errors, and requires

regularization of the covariance matrices of the transformed random vectors. Moreover, the nonlinear mappings of the random vectors may mask the dependencies between their original coordinates.

In this paper, a new nonlinear generalization of LCCA is derived, which does not suffer from the limitations of the MI and KCCA approaches. This framework, called measure transformed canonical correlation analysis (MTCCA) applies LCCA to the considered random vectors after transformation of their joint probability distribution.

The proposed transform is structured by a pair of nonnegative functions, called the MT-functions. It preserves statistical independence and maps the joint probability distribution into a set of probability measures on the joint observation space. By modifying the MT-functions, the correlation coefficient under the transformed probability measure, called the MT-correlation coefficient, is modified, resulting in a new general framework for canonical correlation analysis. In MTCCA, the MT-correlation coefficients between the elements of each generated pair of linear combinations are called the MT-canonical correlation coefficients.

The MT-functions are selected from the exponential family parameterized by scaling parameter. Under this class, it is shown that nonlinear dependencies can be detected by MTCCA. The parameters of the MT-functions are selected via maximization of a lower bound on the largest MT-canonical correlation coefficient. We show that for these selected parameters, the corresponding largest MT-canonical correlation coefficient is a measure for statistical independence. Another variant of MTCCA that uses Gaussian MT-functions parameterized by location parameter is presented in [9].

In the paper we show the superiority of MTCCA over LCCA in recovery of a nonlinear system with known structure. Additionally, MTCCA is illustrated for construction of networks that analyze long-term associations between companies traded in the NASDAQ and NYSE stock markets. We show that MTCCA better associates companies in the same sector than does LCCA. In comparison to [9], in this paper we apply MTCCA to a different financial data set which includes more companies and more financial sectors.

The paper is organized as follows. In Section II, LCCA is generalized by applying a transform to the joint probability distribution. Selection of the MT-functions associated with the transform is discussed in section III. In Section IV, empirical implementation of MTCCA is obtained. In Section V, MTCCA is illustrated via simulation experiment. In Section VI, the main points of this contribution are summarized. The propositions and theorems stated throughout the paper are proved in [9].

II. MEASURE TRANSFORMED CANONICAL CORRELATION ANALYSIS

Let \mathbf{X} and \mathbf{Y} denote two random vectors, whose observation spaces are given by $\mathcal{X} \subseteq \mathbb{R}^p$ and $\mathcal{Y} \subseteq \mathbb{R}^q$, respectively. We define the measure space $(\mathcal{X} \times \mathcal{Y}, \mathcal{S}_{\mathcal{X} \times \mathcal{Y}}, P_{\mathbf{X}\mathbf{Y}})$, where $\mathcal{S}_{\mathcal{X} \times \mathcal{Y}}$ is a σ -algebra over $\mathcal{X} \times \mathcal{Y}$, and $P_{\mathbf{X}\mathbf{Y}}$ is the joint probability measure on $\mathcal{S}_{\mathcal{X} \times \mathcal{Y}}$.

In this section, LCCA is generalized by applying a transform to the joint probability measure $P_{\mathbf{X}\mathbf{Y}}$. First, a transform which maps $P_{\mathbf{X}\mathbf{Y}}$ into a set of joint probability measures $\{Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}\}$ on $\mathcal{S}_{\mathbf{X}\mathbf{Y}}$ is derived that have the property that they preserve statistical independence. The MTCCA method is then obtained by applying LCCA to \mathbf{X} and \mathbf{Y} under the transformed probability measure $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}$.

A. Transformation of the joint probability measure $P_{\mathbf{X}\mathbf{Y}}$

Definition 1. Given two nonnegative functions $u : \mathbb{R}^p \rightarrow \mathbb{R}$ and $v : \mathbb{R}^q \rightarrow \mathbb{R}$ satisfying $0 < \mathbb{E}[u(\mathbf{X})v(\mathbf{Y}); P_{\mathbf{X}\mathbf{Y}}] < \infty$, where $\mathbb{E}[\cdot; P_{\mathbf{X}\mathbf{Y}}]$ denotes the expectation under $P_{\mathbf{X}\mathbf{Y}}$, a transform on the joint probability measure $P_{\mathbf{X}\mathbf{Y}}$ is defined via the following relation

$$Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}(A) \triangleq \mathbb{T}_{u,v}[P_{\mathbf{X}\mathbf{Y}}](A) = \int_A \varphi_{u,v}(\mathbf{x}, \mathbf{y}) dP_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}), \quad (1)$$

where $A \in \mathcal{S}_{\mathbf{X}\mathbf{Y}}$, $\mathbf{x} \in \mathcal{X}$, $\mathbf{y} \in \mathcal{Y}$, and

$$\varphi_{u,v}(\mathbf{x}, \mathbf{y}) \triangleq u(\mathbf{x})v(\mathbf{y})/\mathbb{E}[u(\mathbf{X})v(\mathbf{Y}); P_{\mathbf{X}\mathbf{Y}}]. \quad (2)$$

The functions $u(\cdot)$ and $v(\cdot)$, associated with the transform $\mathbb{T}_{u,v}[\cdot]$, are called the MT-functions.

In [9] it is shown that $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}$ has the following properties: 1) $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}$ is a probability measure on $\mathcal{S}_{\mathbf{X}\mathbf{Y}}$ that preserves statistical independence, 2) $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}$ is absolutely continuous w.r.t. $P_{\mathbf{X}\mathbf{Y}}$, with Radon-Nikodym derivative [10] given by

$$dQ_{\mathbf{X}\mathbf{Y}}^{(u,v)}(\mathbf{x}, \mathbf{y})/dP_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = \varphi_{u,v}(\mathbf{x}, \mathbf{y}). \quad (3)$$

B. The MTCCA procedure

MTCCA generates a sequence of pairwise linear combinations $(\mathbf{a}_k^T \mathbf{X}, \mathbf{b}_k^T \mathbf{Y})$, $k = 1, \dots, r = \min(p, q)$ that have the following statistical properties under the transformed probability measure $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}$: 1) $\mathbf{a}_k^T \mathbf{X}$ and $\mathbf{b}_k^T \mathbf{Y}$ have unit variance, 2) the correlation between $\mathbf{a}_k^T \mathbf{X}$ and $\mathbf{b}_k^T \mathbf{Y}$ is maximal, and 3) $(\mathbf{a}_k^T \mathbf{X}, \mathbf{b}_k^T \mathbf{Y})$ are uncorrelated with $(\mathbf{a}_l^T \mathbf{X}, \mathbf{b}_l^T \mathbf{Y})$ for all $1 < l < k$. In MTCCA, the pairs $(\mathbf{a}_k, \mathbf{b}_k)$ and $(\mathbf{a}_k^T \mathbf{X}, \mathbf{b}_k^T \mathbf{Y})$ are called the k -th order MT-canonical directions and the k -th order MT-canonical variates, respectively. The correlation coefficient between $\mathbf{a}_k^T \mathbf{X}$ and $\mathbf{b}_k^T \mathbf{Y}$ under $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}$ is called the k -th order MT-canonical correlation coefficient.

The correlation coefficient between $\mathbf{a}^T \mathbf{X}$ and $\mathbf{b}^T \mathbf{Y}$ under $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}$ is given by

$$\text{Corr}[\mathbf{a}^T \mathbf{X}, \mathbf{b}^T \mathbf{Y}; Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}] = \frac{\mathbf{a}^T \Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)} \mathbf{b}}{\sqrt{\mathbf{a}^T \Sigma_{\mathbf{X}}^{(u,v)} \mathbf{a}} \sqrt{\mathbf{b}^T \Sigma_{\mathbf{Y}}^{(u,v)} \mathbf{b}}}, \quad (4)$$

where $\Sigma_{\mathbf{X}}^{(u,v)} \in \mathbb{R}^{p \times p}$, $\Sigma_{\mathbf{Y}}^{(u,v)} \in \mathbb{R}^{q \times q}$ and $\Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)} \in \mathbb{R}^{p \times q}$ denote the covariance matrix of \mathbf{X} under the marginal probability measure $Q_{\mathbf{X}}^{(u,v)}$, the covariance matrix of \mathbf{Y} under the marginal probability measure $Q_{\mathbf{Y}}^{(u,v)}$, and their cross-covariance matrix under $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}$, respectively. It is assumed that $\Sigma_{\mathbf{X}}^{(u,v)}$ and $\Sigma_{\mathbf{Y}}^{(u,v)}$ are non-singular.

Using (3) it can be shown that the measure transformed covariance cross-covariance matrices take the form

$$\begin{aligned} \Sigma_{\mathbf{X}}^{(u,v)} &= \mathbb{E}[\mathbf{X}\mathbf{X}^T \varphi_{u,v}(\mathbf{X}, \mathbf{Y}); P_{\mathbf{X}\mathbf{Y}}] - \boldsymbol{\mu}_{\mathbf{X}}^{(u,v)} \boldsymbol{\mu}_{\mathbf{X}}^{(u,v)T}, \\ \Sigma_{\mathbf{Y}}^{(u,v)} &= \mathbb{E}[\mathbf{Y}\mathbf{Y}^T \varphi_{u,v}(\mathbf{X}, \mathbf{Y}); P_{\mathbf{X}\mathbf{Y}}] - \boldsymbol{\mu}_{\mathbf{Y}}^{(u,v)} \boldsymbol{\mu}_{\mathbf{Y}}^{(u,v)T}, \\ \Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)} &= \mathbb{E}[\mathbf{X}\mathbf{Y}^T \varphi_{u,v}(\mathbf{X}, \mathbf{Y}); P_{\mathbf{X}\mathbf{Y}}] - \boldsymbol{\mu}_{\mathbf{X}}^{(u,v)} \boldsymbol{\mu}_{\mathbf{Y}}^{(u,v)T}, \end{aligned} \quad (5)$$

where the vectors $\boldsymbol{\mu}_{\mathbf{X}}^{(u,v)} = \mathbb{E}[\mathbf{X} \varphi_{u,v}(\mathbf{X}, \mathbf{Y}); P_{\mathbf{X}\mathbf{Y}}]$, and $\boldsymbol{\mu}_{\mathbf{Y}}^{(u,v)} = \mathbb{E}[\mathbf{Y} \varphi_{u,v}(\mathbf{X}, \mathbf{Y}); P_{\mathbf{X}\mathbf{Y}}]$. Equation (5) implies that $\Sigma_{\mathbf{X}}^{(u,v)}$, $\Sigma_{\mathbf{Y}}^{(u,v)}$ and $\Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)}$ are weighted covariance and cross-covariance matrices of \mathbf{X} and \mathbf{Y} under $P_{\mathbf{X}\mathbf{Y}}$, with weighting function $\varphi_{u,v}(\cdot, \cdot)$.

MTCCA solves the following constraint maximization sequentially over $k = 1, \dots, r$.

$$\rho_k^{(u,v)} = \max_{\mathbf{a}, \mathbf{b} \in E_k} \mathbf{a}^T \Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)} \mathbf{b}, \quad (6)$$

where $E_k = \{\mathbf{a}, \mathbf{b} : \mathbf{a}^T \Sigma_{\mathbf{X}}^{(u,v)} \mathbf{a} = \mathbf{b}^T \Sigma_{\mathbf{Y}}^{(u,v)} \mathbf{b} = 1, \mathbf{a}^T \Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)} \mathbf{b}_l = \mathbf{b}_l^T \Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)T} \mathbf{a}_l = \mathbf{a}_l^T \Sigma_{\mathbf{X}}^{(u,v)} \mathbf{a}_l = \mathbf{b}_l^T \Sigma_{\mathbf{Y}}^{(u,v)} \mathbf{b}_l = 0 \forall 1 < l < k\}$, and $\rho_k^{(u,v)}$ denotes the k -th order MT-canonical correlation coefficient.

In similar to LCCA [11], the constrained maximization problem in (6) reduces to the set of r distinct solutions of the following GEVD problem

$$\begin{bmatrix} \mathbf{0} & \Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)} \\ \Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)T} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \rho \begin{bmatrix} \Sigma_{\mathbf{X}}^{(u,v)} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{Y}}^{(u,v)} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \quad (7)$$

where $\rho = \rho_k^{(u,v)}$ is the k -th largest generalized eigenvalue of the pencil in (7), and $[\mathbf{a}^T, \mathbf{b}^T]^T = [\mathbf{a}_k^T, \mathbf{b}_k^T]^T$ is its corresponding generalized eigenvector.

By modifying the MT-functions $u(\cdot)$ and $v(\cdot)$, the joint probability measure $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}$ is modified, resulting in a family of canonical correlation analyses, generalizing LCCA. In particular, by choosing $u(\mathbf{x}) \equiv 1$ and $v(\mathbf{y}) \equiv 1$, then $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)} = P_{\mathbf{X}\mathbf{Y}}$, and the LCCA is obtained.

III. MTCCA WITH EXPONENTIAL MT-FUNCTIONS

In this subsection, we parameterize the MT-functions $u(\mathbf{x}; \mathbf{s})$ and $v(\mathbf{y}; \mathbf{t})$ with parameters $\mathbf{s} \in \mathbb{R}^p$ and $\mathbf{t} \in \mathbb{R}^q$ under the exponential family. This will result in the corresponding cross-covariance matrix $\Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)}(\mathbf{t}, \mathbf{s})$ gaining sensitivity to nonlinear relationships between \mathbf{X} and \mathbf{Y} . Optimal choice of the parameters \mathbf{s} and \mathbf{t} is also discussed.

Let $u(\cdot; \cdot)$ and $v(\cdot; \cdot)$ be defined as the parameterized functions

$$u(\mathbf{x}; \mathbf{s}) \triangleq \exp(\mathbf{s}^T \mathbf{x}) \quad \text{and} \quad v(\mathbf{y}; \mathbf{t}) \triangleq \exp(\mathbf{t}^T \mathbf{y}), \quad (8)$$

where $\mathbf{s} \in \mathbb{R}^p$ and $\mathbf{t} \in \mathbb{R}^q$. Using (2), (5) and (8) one can verify that the cross-covariance matrix of \mathbf{X} and \mathbf{Y} under $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}$ takes the form

$$\Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)}(\mathbf{s}, \mathbf{t}) = \partial^2 \log M_{\mathbf{X}\mathbf{Y}}(\mathbf{s}, \mathbf{t}) / \partial \mathbf{s} \partial \mathbf{t}^T, \quad (9)$$

where $M_{\mathbf{X}\mathbf{Y}}(\mathbf{s}, \mathbf{t}) \triangleq \mathbb{E}[\exp(\mathbf{s}^T \mathbf{X} + \mathbf{t}^T \mathbf{Y}); P_{\mathbf{X}\mathbf{Y}}]$ is the joint moment generating function of \mathbf{X}, \mathbf{Y} , and it is assumed that $M_{\mathbf{X}\mathbf{Y}}(\mathbf{s}, \mathbf{t})$ is finite in some open region in $\mathbb{R}^p \times \mathbb{R}^q$ containing the origin. We note that the quantity in (9) has been proposed in [12] for blind source separation.

In the following Theorem, which follows directly from (9) and the properties of $M_{\mathbf{X}\mathbf{Y}}(\mathbf{s}, \mathbf{t})$ [13], one sees that $\Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)}(\mathbf{s}, \mathbf{t})$ can capture nonlinear dependencies.

Theorem 1. Let U denote an arbitrary open region in $\mathbb{R}^p \times \mathbb{R}^q$ containing the origin, and assume that $M_{\mathbf{X}\mathbf{Y}}(\mathbf{s}, \mathbf{t})$ is finite on U . The random vectors \mathbf{X} and \mathbf{Y} are statistically independent under $P_{\mathbf{X}\mathbf{Y}}$ iff $\Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)}(\mathbf{s}, \mathbf{t}) = \mathbf{0} \forall (\mathbf{s}, \mathbf{t}) \in U$.

Therefore, if \mathbf{X} and \mathbf{Y} are statistically dependent, then there exist $\mathbf{a} \in \mathbb{R}^p$, $\mathbf{b} \in \mathbb{R}^q$, $\mathbf{s} \in \mathbb{R}^p$ and $\mathbf{t} \in \mathbb{R}^q$, such that $\mathbf{a}^T \Sigma_{\mathbf{X}\mathbf{Y}}^{(u,v)}(\mathbf{s}, \mathbf{t}) \mathbf{b} \neq 0$. Thus, (4) implies that if \mathbf{X} and \mathbf{Y} are statistically dependent then there exist linear combinations of the form $\mathbf{a}^T \mathbf{X}$ and $\mathbf{b}^T \mathbf{Y}$ which are correlated under $Q_{\mathbf{X}\mathbf{Y}}^{(u,v)}$.

The parameters \mathbf{s}, \mathbf{t} are chosen via maximization of a lower bound on the first-order MT-canonical correlation coefficient $\rho_1^{(u,v)}(\mathbf{s}, \mathbf{t})$. We show that the resultant first-order MT-canonical correlation coefficient is sensitive to dependence between \mathbf{X} and \mathbf{Y} .

Proposition 1. Define the following element-by-element average:

$$\psi^{(u,v)}(\mathbf{s}, \mathbf{t}) \triangleq \sqrt{\frac{1}{pq} \sum_{i=1}^p \sum_{j=1}^q \frac{[\Sigma_{\mathbf{XY}}^{(u,v)}(\mathbf{s}, \mathbf{t})]_{i,j}^2}{[\Sigma_{\mathbf{X}}^{(u,v)}(\mathbf{s}, \mathbf{t})]_{i,i} [\Sigma_{\mathbf{Y}}^{(u,v)}(\mathbf{s}, \mathbf{t})]_{j,j}}},$$

where $[\mathbf{A}]_{i,j}$ denotes the i, j -th entry of \mathbf{A} .

$$\psi^{(u,v)}(\mathbf{s}, \mathbf{t}) \leq \rho_1^{(u,v)}(\mathbf{s}, \mathbf{t}). \quad (10)$$

Proposition 1 suggests choosing the exponential MT-functions parameters by maximizing the lower bound in (10):

$$(\mathbf{s}^*, \mathbf{t}^*) = \arg \max_{(\mathbf{s}, \mathbf{t}) \in V} \psi^{(u,v)}(\mathbf{s}, \mathbf{t}), \quad (11)$$

where V a closed region in $\mathbb{R}^p \times \mathbb{R}^q$ containing the origin. The maximization problem in (11) can be solved numerically, e.g., using gradient ascent over the region V . The following theorem justifies the use of the first-order MT-canonical correlation coefficient as a measure of statistical independence.

Theorem 2. The random vectors \mathbf{X} and \mathbf{Y} are statistically independent under $P_{\mathbf{XY}}$ iff $\rho_1^{(u,v)}(\mathbf{s}^*, \mathbf{t}^*) = 0$.

IV. EMPIRICAL IMPLEMENTATION OF MTCCA

Given N i.i.d. samples of (\mathbf{X}, \mathbf{Y}) an empirical version of MTCCA can be implemented by replacing the measure transformed covariance matrices $\Sigma_{\mathbf{X}}^{(u,v)}$, $\Sigma_{\mathbf{Y}}^{(u,v)}$ and $\Sigma_{\mathbf{XY}}^{(u,v)}$ in (6), (7) and (11) with the following estimators.

$$\begin{aligned} \hat{\Sigma}_{\mathbf{X}}^{(u,v)} &\triangleq \frac{1}{N-1} \sum_{n=1}^N \mathbf{X}_n \mathbf{X}_n^T \hat{\varphi}_{u,v}(\mathbf{X}_n, \mathbf{Y}_n) - \frac{N}{N-1} \hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u,v)} \hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u,v)T}, \\ \hat{\Sigma}_{\mathbf{Y}}^{(u,v)} &\triangleq \frac{1}{N-1} \sum_{n=1}^N \mathbf{Y}_n \mathbf{Y}_n^T \hat{\varphi}_{u,v}(\mathbf{X}_n, \mathbf{Y}_n) - \frac{N}{N-1} \hat{\boldsymbol{\mu}}_{\mathbf{Y}}^{(u,v)} \hat{\boldsymbol{\mu}}_{\mathbf{Y}}^{(u,v)T}, \\ \hat{\Sigma}_{\mathbf{XY}}^{(u,v)} &\triangleq \frac{1}{N-1} \sum_{n=1}^N \mathbf{X}_n \mathbf{Y}_n^T \hat{\varphi}_{u,v}(\mathbf{X}_n, \mathbf{Y}_n) - \frac{N}{N-1} \hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u,v)} \hat{\boldsymbol{\mu}}_{\mathbf{Y}}^{(u,v)T}, \end{aligned} \quad (12)$$

where $(\mathbf{X}_n, \mathbf{Y}_n)$, $n = 1, \dots, N$ is a sequence of i.i.d. samples from the joint distribution $P_{\mathbf{XY}}$, $\hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u,v)} \triangleq \frac{\sum_{n=1}^N \mathbf{X}_n \hat{\varphi}_{u,v}(\mathbf{X}_n, \mathbf{Y}_n)}{N}$, $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}^{(u,v)} \triangleq \frac{\sum_{n=1}^N \mathbf{Y}_n \hat{\varphi}_{u,v}(\mathbf{X}_n, \mathbf{Y}_n)}{N}$, and $\hat{\varphi}_{u,v}(\mathbf{X}_n, \mathbf{Y}_n) \triangleq \frac{u(\mathbf{X}_n)v(\mathbf{Y}_n)}{\frac{1}{N} \sum_{n=1}^N u(\mathbf{X}_n)v(\mathbf{Y}_n)}$.

Under some mild assumptions it is shown in [9] that the estimators in (12) are strongly consistent.

V. NUMERICAL EXAMPLES

In this section, we illustrate the use of empirical MTCCA with the exponential MT-functions for recovery of a nonlinear system with known structure, and for measuring long-term associations between companies. The empirical MTCCA was performed via the procedure described in [9].

A. Recovery of a nonlinear system with known structure

We consider the random vectors $\mathbf{X} = [X_1, X_2, X_3, X_4, X_5]^T$ and $\mathbf{Y} = [Y_1, Y_2, Y_3]^T$ that satisfy the following nonlinear system:

$$\begin{aligned} \mathbf{b}_1^T \mathbf{Y} &= \sin(\mathbf{a}_1^T \mathbf{X}) + 0.01W_1, \\ \mathbf{b}_2^T \mathbf{Y} &= \cos(\mathbf{a}_2^T \mathbf{X}) + 0.01W_2, \end{aligned} \quad (13)$$

where X_i , $i = 1, \dots, 5$, Y_3 , and W_i , $i = 1, 2$, are mutually independent standard normal random variables, $\mathbf{a}_1 = [1, 0, 0, 0, 0]^T$, $\mathbf{b}_1 = [1, 0, 0]^T$, $\mathbf{a}_2 = [0, 1, 0.7, 0.5, 0.3]^T$, and $\mathbf{b}_2 = [0, 1, 0]^T$. In

this example there exist two independent pairs of linear combinations $(\mathbf{a}_k^T \mathbf{X}, \mathbf{b}_k^T \mathbf{Y})$, $k = 1, 2$, with maximal inter-dependencies. Note that while $(\mathbf{a}_1^T \mathbf{X}, \mathbf{b}_1^T \mathbf{Y})$ are correlated, $(\mathbf{a}_2^T \mathbf{X}, \mathbf{b}_2^T \mathbf{Y})$ are uncorrelated.

MTCCA and LCCA were applied for recovery of the system in (13) using $N = 1000$ i.i.d. samples of \mathbf{X} and \mathbf{Y} . Averaged estimates of the MT and linear canonical correlation coefficients and their corresponding averaged p -values, based on 1000 Monte-Carlo simulations, are given in Table I. Let $(\hat{\mathbf{a}}_k, \hat{\mathbf{b}}_k)$, $k = 1, 2$ denote the empirical canonical directions. The sample means and standard deviations of the absolute dot products of the pairs $(\mathbf{a}_k / \|\mathbf{a}_k\|_2, \hat{\mathbf{a}}_k / \|\hat{\mathbf{a}}_k\|_2)$ and $(\mathbf{b}_k / \|\mathbf{b}_k\|_2, \hat{\mathbf{b}}_k / \|\hat{\mathbf{b}}_k\|_2)$, $k = 1, 2$, based on 1000 Monte-Carlo simulations, are given in Table II. The absolute dot products should be equal to 1 when the estimated canonical directions $(\hat{\mathbf{a}}_k, \hat{\mathbf{b}}_k)$ are equal to $(\mathbf{a}_k, \mathbf{b}_k)$, $k = 1, 2$.

Observe that MTCCA detects the true dependencies between \mathbf{X} and \mathbf{Y} , and recovers $(\mathbf{a}_k, \mathbf{b}_k)$, $k = 1, 2$. As expected, the LCCA detects only the linearly dependent combinations $(\mathbf{a}_1^T \mathbf{X}, \mathbf{b}_1^T \mathbf{Y})$.

TABLE I

THE AVERAGED EMPIRICAL MT AND LINEAR CANONICAL CORRELATION COEFFICIENTS AND THEIR AVERAGED p -VALUES (IN PARENTHESIS). THE PROPOSED MTCCA METHOD CAPTURES THE HIGH CORRELATION ρ_2 THAT IS MISSED BY THE STANDARD LCCA.

	MTCCA	LCCA
$\hat{\rho}_1$	0.93 (0)	0.92 (0)
$\hat{\rho}_2$	0.76 (0)	0.08 (0.21)
$\hat{\rho}_3$	0.08 (0.2)	0.03 (0.35)

TABLE II

THE SAMPLE MEANS AND STANDARD DEVIATIONS (IN PARENTHESIS) OF $c(\mathbf{a}_k, \hat{\mathbf{a}}_k)$ AND $c(\mathbf{b}_k, \hat{\mathbf{b}}_k)$, $k = 1, 2$, WHERE $c(\mathbf{u}, \mathbf{v}) \triangleq \frac{|\mathbf{u}^T \mathbf{v}|}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2}$. THE PROPOSED MTCCA METHOD CAPTURES TRUE DEPENDENCY STRUCTURE BETTER THAN STANDARD LCCA.

	MTCCA	LCCA
$c(\mathbf{a}_1, \hat{\mathbf{a}}_1)$	0.99 ($2 \cdot 10^{-3}$)	0.99 ($2 \cdot 10^{-4}$)
$c(\mathbf{a}_2, \hat{\mathbf{a}}_2)$	0.99 ($8 \cdot 10^{-3}$)	0.51 (0.28)
$c(\mathbf{b}_1, \hat{\mathbf{b}}_1)$	0.99 ($2 \cdot 10^{-3}$)	0.99 ($1.6 \cdot 10^{-4}$)
$c(\mathbf{b}_2, \hat{\mathbf{b}}_2)$	0.99 ($7 \cdot 10^{-3}$)	0.79 (0.26)

B. Measuring long-term associations between companies

In this example, MTCCA is applied for measuring of long-term associations between pairs of companies traded on the NASDAQ and NYSE stock markets. The companies were selected from five sectors:

- Technology:** Microsoft (MSFT), Intel (INTC), Apple (AAPL), and International Business Machines Corp. (IBM).
- Pharmaceuticals:** Merck (MRK), Pfizer (PFE), Johnson and Johnson (JNJ), and Eli Lilly & Co. (LLY).
- Financials:** American express (AXP), JP Morgan (JPM), Bank of America (BAC), and U.S. Bancorp (USB).
- Energy:** Occidental Petroleum Corporation (OXY), Noble Corp. (NE), Apache Corp. (APA), and EOG Resources, Inc. (EOG).
- Industrial:** 3M Co. (MMM), United Technologies Corp. (UTX), Emerson Electric Co. (EMR), and General Electric (GE).

For each pair of companies, we considered the random vectors $\mathbf{X} = [X_1, X_2]^T$ and $\mathbf{Y} = [Y_1, Y_2]^T$. The variables X_1 and Y_1 are the log-ratios of two consecutive daily closing prices of a stock, called log-returns. The variables X_2 and Y_2 are the log-ratios of two consecutive daily trading volumes of a stock, called log-volume ratios. Consecutive daily measurements of \mathbf{X} and \mathbf{Y} from January 2,

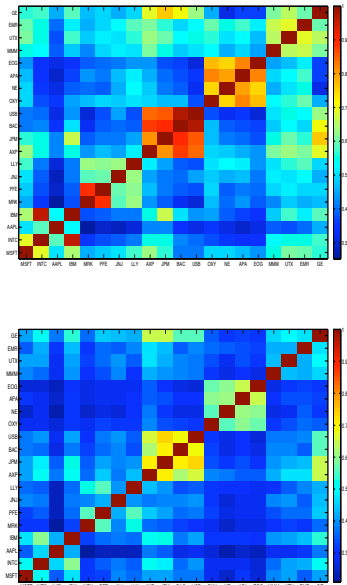


Fig. 1. **(Top)** MTCCA: empirical first-order MT-canonical correlation coefficients. **(Bottom)** LCCA: empirical first-order linear canonical correlation coefficients. Note the five blocks of mutually high canonical correlations revealed by MTCCA; MTCCA better clusters companies in similar sectors.

2001 to December 31, 2010, comprising 2514 samples, were obtained from the YAHOO finance database.

Fig. 1 displays the matrix of empirical first-order MT-canonical correlation coefficients, and the matrix of first-order linear canonical correlation coefficients, respectively. One can notice that MTCCA better clusters companies in the same sector. In this example, the p -values associated with all empirical first-order canonical correlation coefficients were less than 0.01.

The empirical first-order canonical correlation coefficients (MT and linear) were used for constructing graphical models in which the nodes represent the compared companies. The criterion for connecting a pair of nodes was set to empirical first-order canonical correlation coefficient greater than a threshold λ . Fig. 2 compares the graphical models selected by MTCCA and LCCA. In the first column we show the graphs selected by MTCCA for $\lambda = 0.46, 0.5, 0.59$. In the second column we show the corresponding graphs selected by LCCA by scanning λ over the interval $[0, 1]$ and finding the graph with minimum edit distance. The symmetric difference graphs are shown in the third column. The red lines in the symmetric difference graphs indicate edges found by MTCCA and not by LCCA, and vice-versa for the black lines. Note that for all of the threshold parameters λ investigated, the MTCCA graph shows a larger number of edges than the closest LCCA graph. This result suggests that MTCCA has captured more dependencies than LCCA. While there is no ground truth validation, the fact that MTCCA clusters together companies in similar sectors provides anecdotal support for the power and applicability of MTCCA.

VI. CONCLUSION

In this paper, LCCA was generalized by applying a transform to the joint probability distribution of \mathbf{X} and \mathbf{Y} . By modifying the functions associated with the transform, this generalization, called MTCCA, preserves independence and captures nonlinear dependencies. A class

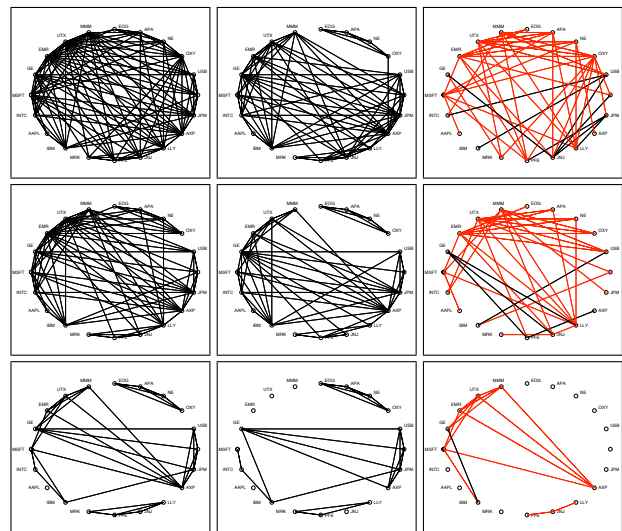


Fig. 2. **Left column:** The graphical models selected by MTCCA for threshold levels $\lambda = 0.46, 0.5, 0.59$. **Middle column:** The closest graphs, in the edit distance sense, selected by LCCA over the range of threshold levels $[0, 1]$. **Right column:** The symmetric difference graphs: the red lines in the symmetric difference graphs indicate edges found by MTCCA and not by LCCA, and vice-versa for the black lines. For these values of λ , MTCCA detects more dependencies than LCCA: the MTCCA graph has more edges than the closest LCCA graph.

of MTCCA was proposed based on specification of MT-functions in the exponential family. MTCCA was compared to LCCA for recovery of a nonlinear system with known structure, and for measuring long-term associations between companies traded on the NASDAQ and NYSE stock markets. Finding other classes of MT-functions that have a similar capability to accurately detect nonlinear dependencies is a topic for future research.

REFERENCES

- [1] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, pp. 321-377, 1936.
- [2] Y.O. Li, T. Adali, W. Wang, and V.D. Calhoun, "Joint Blind Source Separation by Multiset Canonical Correlation Analysis," *IEEE Trans. Signal Processing*, vol. 57, pp. 3918-3929, 2009.
- [3] T. K. Kim, J. Klitter, and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1005-1018, 2007.
- [4] G. Huang, L. Yang, and Z. He, "Canonical correlation analysis using for DOA estimation of multiple audio sources," *Computational Intelligence and Bioinspired Systems*, vol. 29, pp. 228-302, 2005.
- [5] N. M. Correa, T. Adali, and Y. O. Li, "Canonical correlation analysis for data fusion and group inferences," *IEEE Signal Processing Magazine*, vol. 27, pp. 39-50, 2010.
- [6] J. S. Lee, and T. Ebrahimi, "Audio-visual synchronization recovery in multimedia content," *Proc. of the ICASSP 2011*, pp. 2280 - 2283, 2011.
- [7] X. Yin, "Canonical correlation analysis based on information theory," *Journal of Multivariate Analysis*, vol. 91, pp. 161-176, 2004.
- [8] S. Akaho, "A kernel method for canonical correlation analysis," *Proc. of the IMPS 2001*,.
- [9] K. Todros, and A.O. Hero, "On Measure Transformed Canonical Correlation Analysis," available as arXiv:1111.6308.
- [10] G. B. Folland, *Real Analysis*. John Wiley and Sons, 1984.
- [11] T. W. Anderson, *An introduction to multivariate statistical analysis*, John Wiley and Sons, 2003.
- [12] A. Yeredor, "Blind Source Separation via the Second Characteristic Function," *Signal Processing*, vol. 80, pp. 897-902, 2000.
- [13] A. DasGupta, *Fundamentals of Probability: A First Course*. Springer Verlag, 2010.