Theoretical Mathematics & Applications, vol.6, no.2, 2016, 99-113
ISSN: 1792-9687 (print), 1792-9709 (online)
Scienpress Ltd, 2016

A Predictive Model for Canonical Correlation Analysis with Implications for the Simple and Multiple Correlations

Ayodeji Idowu Oluwasayo[1] **and T.O. Obilade**[2]

Abstract

This study presents a new formulation for canonical correlation analysis as an equivalent multiple regression model for two Gaussian random vectors. In addition, implications for the simple and multiple correlations are discussed. The new model can also be extended to measure nonlinear associations between sets of variables.

Mathematics Subject Classification: 62H20; 62J05; 62J12
Keywords: canonical correlation; multivariate analysis; simple correlation; multiple correlation

[1] Department of Mathematics, Obafemi Awolowo University, 220005, Ile-Ife, Nigeria.
E-mail: idowu.sayo@yahoo.com
[2] Department of Mathematics, Obafemi Awolowo University, 220005, Ile-Ife, Nigeria.
E-mail: tobilade@yahoo.com

Article Info: *Received* : November 20, 2015. *Revised* : January 12, 2016.
Published online : April 5, 2016.

1 Introduction

Canonical correlation analysis (CCA) was introduced by Hotelling [1] to measure relationships between sets of dependent variates. It is especially useful in instances where one may be interested in the level of association between sets of variables or relationship within a single set. The system of classical CCA is described in Figure 1.

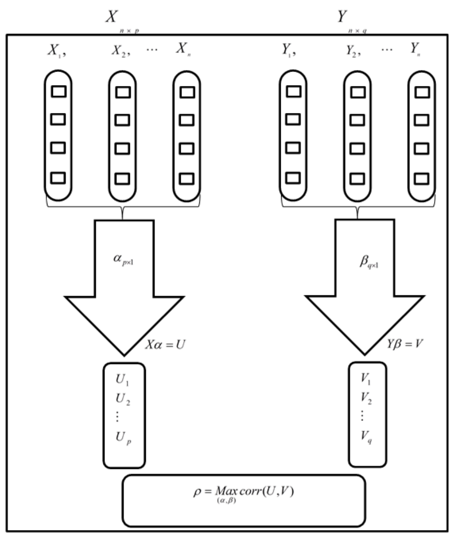

Figure 1: The System of Classical CCA

Given two sets of variables X and Y, it is customary to refer to one as independent variables (IV), and the other as dependent variables (DV), though the two sets can play symmetrical roles in this context. Canonical correlation analysis forms a linear composite, that is, canonical variate $U = X\alpha$ and $V = Y\beta$ from each set, then develops a function that maximizes the canonical correlation coefficient ρ between the two canonical variates.

Canonical correlation analysis has several advantages: (i) it limits the probability to reject the null hypothesis (H_0) when it is true. In other words, it reduces the probability of committing Type I error. The risk of a Type I error is related to the likelihood of finding a statistically significant result when it does not exist. Increased risk of Type I error results from when the same variables in a dataset are used for too many statistical tests. Suppose a researcher is interested in predicting four DVs using three IVs through multiple regression, then a series of four regression equations are required (i.e., one for each DV).

However, conducting separate statistical tests for each equation substantially increases the risk of Type I error. CCA can investigate these relationships in a single equation rather than using separate equations for each DV (Anderson [2]); (ii) it can identify two or more unique relationships, if they exist: CCA develops multiple canonical functions; each function is orthogonal with respect to the other functions so that they depict different relations among variables. Thus CCA is a well designed technique for analyzing data involving multiple DVs and IVs and its theory is consistent with that purpose; and (iii) it provides a platform to analyze relationships in accordance with the realities of life: Often times, the complexity of research studies involving human or organizational behavior may suggest multiple DVs that interract and thus create problems (such as misspecification) when the variables are examined separately. Accordingly, canonical correlation would represent a relationship between the sets of variables rather than individual variables.

Though CCA generalizes most linear models, its use in practice has often been limited to dimension reduction. Further, the issue of interpretation is another major limitation. "The standard derivation of canonical solutions is mathematically elegant but uninterpretable" (Tabachnick and Fidell [3]). While the solutions of other related procedures like Factor Analysis and Principal Components Analysis can be rotated in an attempt to improve interpretability, that of CCA cannot because "rotation destroys the optimality of the canonical correlations and also introduces correlations among succeeding canonical variates" (Rencher [4]).

In view of these limitations, the present study presents a probabilistic formulation for CCA using a novel combination of the techniques of multiple regression and canonical correlation. Such formulation enhances the understanding of CCA as a model-based method that is useful in modeling and prediction. It also suggests generalization of CCA to distributions other than the Gaussian distribution. What is more, the formulation can be exploited to learn nonlinear associations between sets of variables. In addition, since multiple regression is a very popular method, well understood by many, and frequently used by practitioners, consequently, such representation will provide simple, easy-to-understand and straightforward derivation of CCA which will facilitate ease of interpretation of the variates and the correlations.

The rest of the paper is organized as follows: Section 2 reviews the previous formulations in canonical correlation analysis. Section 3 provides the derivations of canonical functions. The next section presents an equivalent probabilistic expression for CCA together with the canonical solutions. Section 5 contains the conclusion.

2 Literature Review

Canonical Correlation Analysis (CCA) usually referred to as the classical CCA originated from Hotelling [1], [5], who applied the technique to a data set in which one set of variables consisted of mental tests and the other involved physical measurements. Canonical correlation (ρ) between two sets of multivariate random variables X and Y is the covariance (Cov) between the two variables normalized by the geometric mean of the variances (Var) of $X\alpha$ and $Y\beta$;

$$\rho = \frac{Cov(X\alpha, Y\beta)}{\sqrt{Var(X\alpha)Var(Y\beta)}}; \tag{1}$$

where α and β are vectors of appropriate dimensions.

Initially, CCA did not receive much attention from practitioners due to little understanding of the concept and absence of computer programs. According to Cramer [6], the "complicated way" in which canonical correlation equations are derived in standard texts like Anderson [2] or Morrison [7] contributed to the reason it is not well understood. Consequently, the study proposed a particular simple derivation which follows directly from the relation between multiple regression analysis and multiple correlation. Though the attempt did not yield the exact canonical variates as Hotelling's, yet it achieved a simple derivation of the technique and also suggested that CCA could be cast in a regression model.

An alternative approach which presented CCA as a least squares problem was first discussed in Muller [8]. A multivariate multiple regression representation of CCA amounts to finding an estimate of β, α and $D(\rho_k)$ in the following model equation:

$$Y\beta = X\alpha D(\rho_k) + E; \tag{2}$$

where β is a $q \times d$ matrix, with the kth column being the canonical weights for the set for the kth canonical variate pair. $D(\rho_k)$ is a $d \times d$ diagonal matrix of canonical correlation. α is a $p \times d$ matrix, with the kth column being the canonical weights for the X set for the kth canonical variate pair. The matrices β, α, and $D(\rho_k)$ must correspond in the sense that the kth columns of β and α provide the linear combinations that are correlated ρ_k , which is the (k, k) element of $D(\rho_k)$. E is a $n \times d$ matrix of errors.

Overall, this is a novel and clearer approach; however, the multivariate formulation introduced some "greater complication" which needs "extra care to deal with" (Muller [8]). For instance, the equivalence of β, α, and $D(\rho_k)$ in the standard statement of CCA are vectors (not matrices). Hence, "the standard statement of canonical correlation has more in common with the univariate statement than with the multivariate (that was developed in the study)" (Muller [8]).

The least squares approach now popularly referred to as 'The least squares formulation' facilitated better understanding of canonical correlation. Consequently, several limitations of CCA were identified: (i) it is limited to linear association; (ii) it is sensitive to outliers due to the normal density assumption. Hence, two other basic reformulations of CCA were developed: (i) Kernel CCA; and (ii) probabilistic graphical approach.

Kernel CCA was formulated by Fyfe and Lai [9] to address the nonlinear associations between sets of variables. Given the pair of multivariate random variables (X, Y), kernel CCA maximizes the equation

$$\rho_K = \max_{\alpha,\beta} \alpha^{'} K_X K_Y \beta \tag{3}$$

$$s.t. \quad \alpha^{'} K_X^2 \alpha = \beta^{'} K_Y^2 \beta = 1;$$

where $\alpha^{'} = (\alpha_1, \alpha_2, \cdots, \alpha_n)$, $\beta^{'} = (\beta_1, \beta_2, \cdots, \beta_n)$, and K_x and K_y are the Gram matrices $[K_X]_{n\times n} = k(X_i, X_j)$ and $[K_Y]_{n\times n} = k(Y_i, Y_j)$ calculated from the sample.

KCCA though nonparametric, generalizes the classical CCA to a nonlinear setting. It is mostly used in multimedia applications. KCCA first transforms the data to a higher (or infinite) dimensional nonlinear space, called the reproducing kernel Hilbert space, and then assumes that there exists a linear relationship between the variables in the transformed space. In other words, the method computes linear correlation between a linear combination of (nonlinearly) transformed variables in one set and a linear combination of (nonlinearly) transformed variables in the other set. More details on KCCA can be found in Hardoon, Szedmak and Shawe-Taylor [10].

The most recent reformulation is the probabilistic graphical approach introduced by Bach and Jordan [11]. The model is presented as

$$\begin{array}{rcll} z & \sim & N(0, I_d), & min(p,q) \geq d \geq 1, \\ X|z & \sim & N(\alpha z, \psi_1) & \alpha \in \mathbb{R}^{p\times d}, \quad \psi_1 \geq 0, \\ Y|z & \sim & N(\beta z, \psi_2) & \beta \in \mathbb{R}^{q\times d}, \quad \psi_2 \geq 0. \end{array} \tag{4}$$

The probabilistic interpretation opens doors to several extension of CCA: first, it demonstrates that Hotelling's canonical variates can be obtained through maximum likelihood estimation; second, since graphical models are naturally viewed as exponential families (Wainwright and Jordan [12]), the approach has potential to extending the scope of CCA from the assumption of normal density to the exponential family.

The new model proposed in the study is a hybrid of the least squares and Bach-Jordan formulations. Along the line of Bach-Jordan, it is a probabilistic model though not graphical; and along the least squares, it is a multiple regression representation though not multivariate multiple. The model combines the advantages of (i) mathematical simplicity; and (ii) linear and nonlinear relations between variables from the least squares approach, and the probabilistic advantage of Bach-Jordan. In addition, it enhances ease of interpretation and allows easy and straight-forward generalizing of CCA to nonnormal cases.

3 The Classical Correlation Analysis

Consider N observations on two sets of standardized variables X and Y

with

$$X = \{x_{ij}\}; \quad i = 1, 2, \cdots, N, j = 1, 2, \cdots, p,$$
$$Y = \{y_{ij}\}; \quad i = 1, 2, \cdots, N, j = 1, 2, \cdots, q,$$

The matrices of correlations R among the X variables and the Y variables, and between the two sets of variables are R_{XX}, R_{YY} and R_{XY}, respectively.

CCA seeks to maximize the correlation (ρ) between $X\alpha$ and $Y\beta$; ($\alpha_{p\times 1}$ and $\beta_{q\times 1}$) $\ni$ the variance, $Var(X\alpha) = Var(Y\beta) = 1$:

$$\rho = \frac{Cov(X\alpha, Y\beta)}{\Big(Var(X\alpha), Var(Y\beta)\Big)^{\frac{1}{2}}}. \tag{5}$$

Of course, Cov is the covariance between U and V.

(5) can be equivalently expressed as,

$$\rho = \frac{\alpha' R_{XY}\beta}{\Big((\alpha' R_{XX}\alpha)(\beta' R_{YY}\beta)\Big)^{\frac{1}{2}}} \tag{6}$$

$$s.t. \quad \alpha' R_{XX}\alpha = \beta' R_{YY}\beta = 1.$$

Maximize the quantity $\alpha' R_{XY}\beta$ subject to the constraints $\alpha' R_{XX}\alpha = \beta' R_{YY}\beta = 1$: Introduce Lagrangian multipliers, then compute matrix derivatives, set them to zero and simplify. The solution leads to the normal equations

$$R_{XY}\beta - KR_{XX}\alpha = 0 \tag{7}$$
$$R_{YX}\alpha - KR_{YY}\beta = 0 \tag{8}$$

The values of K are obtained by solving the multivariate eigenvalue problem

$$\left|R_{XX}^{-1}R_{XY}R_{YY}^{-1}R_{YX} - K^2 I\right| = 0 \tag{9}$$

or

$$\left|R_{YY}^{-1}R_{YX}R_{XX}^{-1}R_{XY} - K^2 I\right| = 0. \tag{10}$$

The positive square root of the largest eigenvalue gives the largest correlation (Rencher [4]).

Further, the values of α and β can be obtained from Equations (7) and (8) as follows:

$$\beta = \frac{1}{\rho} R_{XX}^{-1} R_{YX} \alpha \tag{11}$$

Substituting (11) in Equation (8) and rearranging terms gives

$$\left(R_{XX}^{-1} R_{XY} \Sigma_{YY}^{-1} R_{YX} - K^2 I\right) \alpha = 0 \tag{12}$$

Similarly,

$$\left(R_{YY}^{-1} R_{YX} R_{XX}^{-1} R_{XY} - K^2 I\right) \beta = 0 \tag{13}$$

4 Main Result: A Predictive Model for Canonical Correlation Analysis

Consider N observations of two sets of standardized variables with

$$\begin{aligned} X &= \{x_{ij}\}; \quad i = 1, 2, \cdots, N, \quad j = 1, 2, \cdots, p; \\ Y &= \{y_{ij}\}; \quad i = 1, 2, \cdots, N, \quad j = 1, 2, \cdots, q. \end{aligned}$$

The matrices of correlations among the X variables, among the Y variables, and between the two sets are $R_{XX} = \frac{1}{N} X' X$, $R_{XX} = \frac{1}{N} X' X$, $R_{YY} = \frac{1}{N} Y' Y$, $R_{XY} = \frac{1}{N} X' Y$, respectively.

Recall Equations (5) and (6). The problem of CCA is to choose constants β in such a way that $y = Y\beta$ will have the greatest possible multiple correlation with $X's$; this is equivalent to the requirement that $y = Y\beta$ be estimated from a regression equation with the smallest possible mean square error (in the sense of least squares) (Hotelling [1]):

$$\max_{\alpha,\beta} \alpha' R_{XY} \beta \equiv \max_{\alpha,\beta} \varsigma; \tag{14}$$

where $\varsigma = (Y\beta - X\alpha)'(Y\beta - X\alpha)$;

subject to $(X\alpha)'(X\alpha) = 1$ and $(Y\beta)'(Y\beta) = 1$.
Equation (14) may be equivalently expressed as

$$\max_{\alpha,\beta} \alpha' R_{XY} \beta \equiv \max_{\alpha,\beta} l; \tag{15}$$

where $l = -\frac{N}{2}\log 2\pi - \frac{1}{2}(Y\beta - X\alpha)'(Y\beta - X\alpha)$.

Thus the estimates of α and β may be obtained from the regression model

$$Y\beta = X\alpha + \epsilon; \quad \epsilon \sim N(0,1). \tag{16}$$

From this viewpoint, canonical correlation analysis can be seen as a straightforward generalization of multiple regression model in which several Y variables are simultaneously related to several X variables. In this view, CCA seeks a vector β such that the composite variable $Y\beta$ is most predictable from the variables of X.

Remark 4.1. *Clearly, Model (16) allows a straightforward extension of the scope of CCA beyond the normal distribution. It can, as a matter of fact, be generalized to any distribution f :*

$$Y\beta = X\alpha + \epsilon; \qquad \epsilon \sim f. \tag{17}$$

Remark 4.2. *Following Davidson and Mckinnon [13], the next expression generalizes Equation (17) to measuring nonlinear associations:*

$$Y\beta = g\,(X;\alpha) + \epsilon; \qquad \epsilon \sim f. \tag{18}$$

4.1 Solutions to CCA Model Parameters

Recall

$$Y\beta = X\alpha + \epsilon; \quad \epsilon \sim N(0,1). \tag{16}$$

Model (16) relates to the case when the variables X and Y are linearly related and normally distributed. Recall that these were the assumptions of Hotelling [1].

The following results demonstrate a maximum likelihood approach to obtaining canonical solutions that are identical to Hotelling's:

Theorem 4.3. *Given Model (16) subject to* $(X\alpha)^{'}(X\alpha) = 1$ *and* $(Y\beta)^{'}(Y\beta) = 1$, *the maximum likelihood estimation of* α *and* β *yields the following normal equations*

$$\begin{aligned} R_{XY}\beta - \rho R_{XX}\alpha &= 0, & (19)\\ R_{YX}\alpha - \rho R_{YY}\beta &= 0; & (20) \end{aligned}$$

where $R_{XX} = \frac{1}{N}X^{'}X$, $R_{XX} = \frac{1}{N}X^{'}X$, $R_{YY} = \frac{1}{N}Y^{'}Y$, $R_{XY} = \frac{1}{N}X^{'}Y$; *and* $\rho = corr(X\alpha, Y\beta)$.

Proof. Let $y = Y\beta$, then Model (16) can be re-written as

$$y = X\alpha + \epsilon,$$

which is the familiar multiple regression expression. The loglikelihood equation denoted l is readily written down:

$$l = -\frac{N}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{N}(y_i - X_i\alpha)^2. \tag{21}$$

The observations are assumed to be independent hence the loglikelihood function l is just the sum of these contributions over all i, or

$$\begin{aligned} l &= -\frac{N}{2}\log 2\pi - \frac{1}{2}(y - X\alpha)^{'}(y - X\alpha)\\ &= -\frac{N}{2}\log 2\pi - \frac{1}{2}(Y\beta - X\alpha)^{'}(Y\beta - X\alpha). \end{aligned} \tag{22}$$

In line with the conditions of CCA, we wish to find a pair of vectors (α, β) which yields sets of composite variables $X\alpha$ and $Y\beta$ such that $Y\beta$ is most predictable from the variables of X. In other words, we seek a pair of vectors (α, β) that maximizes the function Q such that

$$Q = -\frac{N}{2}\log 2\pi - \frac{1}{2}(Y\beta - X\alpha)^{'}(Y\beta - X\alpha) - \frac{K}{2}\{(X\alpha)^{'}(X\alpha) - 1\} - \frac{L}{2}\{(Y\beta)^{'}(Y\beta) - 1\}. \tag{23}$$

Then the first order conditions are

$$\begin{aligned} \frac{\delta}{\delta\alpha}Q &= X^{'}Y\beta - (1+K)X^{'}X\alpha,\\ \frac{\delta}{\delta\beta}Q &= Y^{'}X\alpha - (1+L)Y^{'}Y\beta. \end{aligned}$$

The solution leads to the following stationary equations

$$R_{XY}\beta - AR_{XX}\alpha = 0, \tag{24}$$
$$R_{YX}\alpha - BR_{YY}\beta = 0; \tag{25}$$

where $A = 1 + K$ and $B = 1 + L$.

Multiplying Equation (24) by $\alpha^{'}$ and Equation (25) by $\beta^{'}$ shows that

$$A = B = \alpha^{'} R_{XY}\beta = \rho.$$

Consequently,

$$R_{XY}\beta - \rho R_{XX}\alpha = 0, \tag{19}$$

$$R_{YX}\alpha - \rho R_{YY}\beta = 0. \tag{20}$$

$\square$

Theorem 4.4. *The solutions of the stationary equations (19) and (20) yield the canonical correlations ρ and canonical variates α, β that are defined in $(R_{XX}^{-1}R_{XY}R_{YY}^{-1}R_{YX} - \rho^2 I)\alpha$ or $(R_{YY}^{-1}R_{YX}R_{XX}^{-1}R_{XY} - \rho^2 I)\beta$.*

Proof. ¿From Equations (19) and (20),

$$MX = 0; \tag{26}$$

where $M = \begin{pmatrix} -\rho R_{XX} & R_{XY} \\ R_{YX} & -\rho R_{YY} \end{pmatrix}$ and $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$.

The homogeneous system (26) has a non-zero solution iff $|M| = 0$ (Lipschutz [14]).

We thus seek ρ so that

$$|M| = \begin{vmatrix} -\rho R_{XX} & R_{XY} \\ R_{YX} & -\rho R_{YY} \end{vmatrix} = 0.$$

By the determinant property of block matrices,

$$|M| = 0 \Rightarrow |R_{XX}||R_{YY}|\left|R_{YY}^{-1}R_{YX}R_{XX}^{-1}R_{XY} - \rho^2 I\right| = 0. \tag{27}$$

Alternatively,

$$|M| = 0 \Rightarrow |R_{YY}||R_{XX}| \left| R_{XX}^{-1} R_{XY} R_{YY}^{-1} R_{YX} - \rho^2 I \right| = 0. \tag{28}$$

The values of ρ^2 is determined by solving the multivariate eigenvalue problem (MVEP)

$$\left| R_{YY}^{-1} R_{YX} R_{XX}^{-1} R_{XY} - \rho^2 I \right| = 0 \tag{29}$$

or

$$\left| R_{XX}^{-1} R_{XY} R_{YY}^{-1} R_{YX} - \rho^2 I \right| = 0. \tag{30}$$

To obtain the solutions of α and β, we revert again to Equation (19), which implies

$$\beta = \frac{1}{\rho} R_{YY}^{-1} R_{YX} \alpha. \tag{31}$$

Substituting Equation (31) in (20) and re-arranging the term gives

$$\left(R_{XX}^{-1} R_{XY} R_{YY}^{-1} R_{YX} - \rho^2 I \right) \alpha = 0. \tag{32}$$

Similarly,

$$\left(R_{YY}^{-1} R_{YX} R_{XX}^{-1} R_{XY} - \rho^2 I \right) \beta = 0. \tag{33}$$

□

4.2 Implications for the Simple and Multiple Correlations

The following results demonstrates that the multiple and the simple correlations can also be obtained from the new formulation - Model (16).

Corollary 4.5. *The multiple correlation coefficient corresponds to the square root of the case $p = p$, $q = 1$ in Model (16).*

Proof. Given Model (16), let $p = p$ and $q = 1$. Then Equation (29) becomes

$$R^2 = \frac{\vec{R}_{YX} R_{XX}^{-1} \vec{R}_{XY}}{r_{YY}} = \vec{R}_{YX} R_{XX}^{-1} \vec{R}_{XY}. \tag{34}$$

Since canonical correlations are invariant to any full rank linear transformations, it follows that

$$R^2 = \frac{\vec{S}_{YX} S_{XX}^{-1} \vec{S}_{XY}}{S_{YY}}; \tag{35}$$

where R^2 is the multiple coefficient of determination and $\hat{\Sigma} = S$. Of course, the multiple correlation coefficient $R = \sqrt{R^2}$. □

Corollary 4.6. *The simple correlation coefficient corresponds to the square root of the case* $p = 1$, $q = 1$ *in Model (16).*

Proof. Given Model (16), let $p = 1$ and $q = 1$. Then Equation (29) becomes

$$r^2 = \frac{r_{YX} r_{XY}}{r_{XX} r_{YY}} = r_{YX} r_{XY} = r_{XY}^2. \tag{36}$$

The last equality follows by noting thar $r_{XY} = r_{YX} = r$; i.e. symmetric property of ρ.

Again, since canonical correlations are invariant to any full rank linear transformations, it follows that

$$r^2 = \frac{S_{YX} S_{XY}}{S_{YY} S_{XX}} = \frac{S_{XY}^2}{S_{YY} S_{XX}}; \tag{37}$$

where r^2 is the coefficient of determination and $\hat{\Sigma} = S$. Of course, the simple correlation coefficient $r = \pm\sqrt{r^2}$. □

5 Conclusion

The study formulated a predictive model for canonical correlation analysis and discussed the implementation of the simple and multiple correlation coefficients under the formulated model. This new formulation enhances the understanding of canonical correlation as a model-based method. In addition, it aids ease of interpretation as canonical variates can be interpreted as the familiar beta weights.

References

[1] Hotelling, H., The Most Predictable Criterion, *Journal of Experimental Psychology*, **26**, (1935), 139 - 142.

[2] Anderson, T., *An Introduction to Multivariate Analysis*, Wiley, New York, 1958.

[3] Tabachnick, B. and Fidell, L., *Using Multivariate Statistics*, 5th ed., Pearson Education. Inc., 2007.

[4] Rencher, A., *Multivariate Statistical Inference and Applications*, New York, Wiley, 1998.

[5] Hotelling, H., Relations between Two Sets of Variates, *Biometrica*, **28**, (1936), 312-377.

[6] Cramer, E., A Simple Derivation of the Canonical Correlation Equations, *Biometrics*, **29**(2), (1973), 379 - 380.

[7] Morrison, D., *Multivariate Statistical Methods*, McGraw-Hill, New York, 1957.

[8] Muller, K., Understanding Canonical Correlation Analysis through the General Linear Model and Principal Components, *The American Statistician*, **36**(4), (1982), 342-346.

[9] Fyfe, C. and Lai, P., Kernel and Nonlinear Canonical Correlation Analysis, *International Journal of Neural Systems*, (2001).

[10] Hardoon, D., Szedmak, S. and Shawe-Taylor, J., *Canonical Correlation Analysis; An Overview with Application to Learning Methods*, Technical Report, Department of Computer Science, Royal Holloway, University of London. CSD-TR-03-02, 2003.

[11] Bach, F. and Jordan, M., *A Probabilistic Interpretation of Canonical Correlation Analysis*, Tech. Rep. 688, Department of Statistics, University of California, Berkeley, 2005.

[12] Wainwright, M. and Jordan, M., Graphical Models, Exponential Families, and Variational Inference, *Foundations and Trends in Machine Learning*, **1**, (2008), 1 - 305.

[13] Davidson, R. and McKinnon, J., *Econometric Theory and Methods*, Oxford Univ. Press, 1999.

[14] Lipschutz, S., *Theory and Problems of Linear Algebra*, McGraw-Hill, Inc., 1968.