Abstract

Canonical correlation analysis is a statistical tool for investigating the presence of any patterns that simultaneously exist in two different data sets and compute the correlation between associated patterns. In this paper, dynamical systems for computing the canonical correlations and canonical variates between two data sets are derived. These systems are developed by optimizing constrained and unconstrained merit functions. To extract the actual individual canonical variates, some of these systems are weighted by a positive definite diagonal matrix, while others employed upper triangular matrices.

Keywords: Two-set canonical correlation analysis, constrained optimization, Rayleigh quotient, polynomial dynamical systems

1 Introduction

Canonical Correlation Analysis (CCA) is one of a family of correlation techniques (Product moment correlation, multiple regression analysis, etc.). It is the most general form of correlation. Despite certain conceptual and terminological similarities, CCA and principal component analysis (PCA) are different in that CCA is a method used to investigate the intercorrelation between two sets of variables, whereas PCA identifies the pattern of relationship within one set of data.

Canonical correlation analysis is currently being used in fields like chemistry, biology, meteorology, demography, artificial intelligence, cognitive science, political science, sociology, psychometrics, educational research, economics, and management science to analyze multidimensional relations between multiple independent and multiple dependent variables.

Two-set canonical correlations analysis was first introduced in [1, 2] to analyze linear relations between two sets of variables. It was later extended to several data sets by Kettingring [3]. The technique is described in most standard textbooks on multivariate statistics, e.g., [4] and [5]. Work on nonlinear canonical correlations analysis is dealt with in [6]-[7]. Recently, several dynamical systems for CCA computation have been proposed by the author [8]-[9].

The following notation will be used throughout. The symbols \( \mathbb{R} \) denotes the set of real numbers. The transpose of a real matrix \( x \) is denoted by \( x^T \), and the derivative of \( x \) with respect to time is written as \( \dot{x} \). The identity matrix of appropriate dimension is expressed with the symbol \( I \). If \( x \) is a square matrix, the notation \( tr(x) \) and \( det(x) \) denote the trace and determinant of \( x \), respectively. Finally, if \( X \) is positive definite matrix, the notation \( X^{\frac{1}{2}} \) represents the unique principal square root of \( X \).

2 Formulations of the CCA Problem

For the readers convenience, we first offer a brief review of the main facts about the CCA problem. Understanding different formulations for the CCA problem helps in devising efficient computational systems which have better convergence properties. For simplicity, we state some results in equivalent form of formulating the CCA problem. Constrained and unconstrained optimization techniques are essentially the two main approaches for formulating the CCA problem. The motivation for using different formulations is that each leads to different dynamical system with distinct convergence properties. Additionally, some formulations are more suited than others for generalization to the multiset case.

In canonical correlation analysis the goal is to maximize correlations between objects that are represented with two data sets. Let these data sets be \( Y \) and \( Z \), of dimensions \( n \times p \) and \( m \times p \), respectively. Sometimes the data in \( Y \) and \( Z \) are called the dependent and the independent data, respectively. The maximum number of correlations that can be found is then equal to the minimum of the column dimensions \( m \) and \( n \). Let the directions of optimal correlations for the \( Y \) and \( Z \) data sets be given by the matrices \( \alpha \) and \( \beta \), respectively. When the data is projected on these directions, we obtain two new vectors \( Y_{\alpha} \) and \( Z_{\beta} \).

The main objective is to determine \( \alpha \) and \( \beta \) so that \( Y_{\alpha} \) and \( Z_{\beta} \) are maximally correlated. When \( \alpha \in \mathbb{R}^{m \times 1} \) and \( \beta \in \mathbb{R}^{m \times 1} \), the vectors \( \alpha \) is a solution to the eigenvalue problem:

\[
(Y^TY)^{-1}Y^TZ(Z^TZ)^{-1}Z^TY \alpha = \mu^2 \alpha, \quad (1a)
\]

while \( \beta \) is a solution to the eigenvalue problem:

\[
(Z^TZ)^{-1}Z^TY(Y^TY)^{-1}Y^TZ \beta = \mu^2 \beta, \quad (1b)
\]

Here and in the following development, it is assumed that \( Y \) and \( Z \) are full rank. The maximum eigenvalue \( \mu^2 \) is called the correlation coefficient and it represents squared correlation between the corresponding canonical variates. It is clear that \( \mu^2 \) is the largest number that makes the following determinant zero:

\[
det((Y^TY)^{-1}Y^TZ(Z^TZ)^{-1}Z^TY - \mu^2 I) = 0, \quad (2a)
\]

or

\[
det((Z^TZ)^{-1}Z^TY(Y^TY)^{-1}Y^TZ - \mu^2 I) = 0. \quad (2b)
\]

If \( Y \) and \( Z \) are one-dimensional, the maximum eigenvalue \( \mu^2 \) is called the correlation coefficient. It only provides a measure of the linear association between the two variables. For example, when the two variables are uncorrelated, i.e., when their correlation coefficient is zero, this only means that no linear function describes their relationship. A quadratic relationship or some other non-linear relationship may still exist.

Some additional properties of the above matrices (1a) (1b) are given in the following remarks.

Remark 1: The matrices \( (Z^TZ)^{-1}Z^TY(Y^TY)^{-1}Y^TZ \) and \( (Y^TY)^{-1}Y^TZ(Z^TZ)^{-1}Z^TY \) are invariant under nonsingular transformations, i.e., if \( Z_1 = ZP \) and \( Y_1 = YQ \), where \( P \) and \( Q \) are nonsingular matrices then

\[
(Z_1^TZ_1)^{-1}Z_1^TY_1(Y_1^TY_1)^{-1}Y_1^TZ_1 = (Z^TZ)^{-1}Z^TY(Y^TY)^{-1}Y^TZ, \quad (3a)
\]

\[
(Y_1^TY_1)^{-1}Y_1^TY_1(Z_1^TZ_1)^{-1}Z_1^TY_1 = (Y^TY)^{-1}Y^TZ(Z^TZ)^{-1}Z^TY. \quad (3b)
\]

Remark 2: Given a square matrix \( A \in \mathbb{R}^{m \times m} \) and a full rank rectangular matrix \( P \in \mathbb{R}^{n \times m} \) where \( n \geq m \). The nonzero...
eigenvalues of $A$ and $PA \ P^+$ can be shown to be the same, where $P^+ = (P^T P)^{-1} P^T$ is the Moore-Penrose generalized inverse of $P$. To prove this assertion, let $P = QR$ be the qr factorization of $P$, where $Q$ is an $m \times m$ orthogonal matrix and $R = \begin{bmatrix} R_1 & 0 \\ 0 & 0 \end{bmatrix}$, where $R_1 \in \mathbb{R}^{m \times m}$ is upper triangular nonsingular matrix. Then $R^+ = \begin{bmatrix} R_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}$. Therefore,

$$PAP^+ = Q \begin{bmatrix} R_1 & 0 \\ 0 & 0 \end{bmatrix} A \begin{bmatrix} R_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q^T = Q \begin{bmatrix} R_1 A R_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q^T = Q R_1 A R_1^{-1} Q^T.$$ 

The conclusion follows from the fact that $Q$ is orthogonal, and that $R_1 A R_1^{-1}$ and $A$ are similar.

**Remark 3:** Let $A$ and $B$ be any two full rank matrices such that $AB$ is a square matrix. Clearly, $BA$, $A^T B^T$, and $B^T A^T$ are also square matrices. Moreover, it can be shown that the nonzero eigenvalues of the matrices $AB$, $BA$, $A^T B^T$, and $B^T A^T$ are all the same. This follows from the observation that if $(B^T B)^{-1}$ exists, then

$$(B^T B)^{-1} B^T (BA) B = AB.$$ Similarly if $(A^T A)^{-1}$ exists, then

$$(A^T A)^{-1} A (AB) A = BA.$$ If $A^T A$ is singular, then $AA^T$ is invertible and hence,

$$(AA^T)^{-1} A (B^T B) A^T = B^T A^T,$$

while $B^T B$ is singular, then $BB^T$ is invertible and hence,

$$(BB^T)^{-1} B (AB) B^T = A^T B^T.$$ Thus the matrices (1a) and (1b) have the same nonzero eigenvalues of the positive semi-definite matrices $P_1 = (Z^T Z)^{-\frac{1}{2}} Z^T Y (Y^T Y)^{-\frac{1}{2}} Z (Z^T Z)^{-\frac{1}{2}}$ and $P_2 = (Y^T Y)^{-\frac{1}{2}} Y^T Z (Z^T Z)^{-\frac{1}{2}} Z^T Y (Y^T Y)^{-\frac{1}{2}}$. Here $(Y^T Y)^{-\frac{1}{2}}$ and $(Z^T Z)^{-\frac{1}{2}}$ denote the unique principal square roots of $(Y^T Y)^{-1}$ and $(Z^T Z)^{-1}$, respectively. Although it is much more convenient to deal with symmetric semi-definite matrices, the square roots involved in the matrices $P_1$ and $P_2$ make the computations much more complicated. Nonetheless, one can use the same methods to formulate maximization problems for computing the largest eigenvalues follows:

Maximize$_u tr \{ a^T (Y^T Y)^{-\frac{1}{2}} Y^T Z (Z^T Z)^{-1} Z^T Y (Y^T Y)^{-\frac{1}{2}} a \}$, subject to $a^T a = I$, 

or equivalently,

Maximize$_\beta \ tr \{ \beta^T (Z^T Z)^{-\frac{1}{2}} Z^T Y (Y^T Y)^{-\frac{1}{2}} Y^T Z (Z^T Z)^{-\frac{1}{2}} \}$, subject to $\beta^T \beta = I$. 

By using the substitution $u = (Y^T Y)^{-\frac{1}{2}} a$ and $v = (Z^T Z)^{-\frac{1}{2}} \beta$, the above optimization problems transform into:

Maximize$_u \ tr \{ u^T Y^T Z (Z^T Z)^{-1} Z^T Y u \}$, subject to $u^T Y^T Y u = I$, 

Maximize$_u \ tr \{ u^T Y^T Z (Z^T Z)^{-1} Z^T Y u \}$, subject to $u^T Z^T Z v = I$, 

respectively.

When $\alpha, \beta, u, v$ are one-dimensional, the equivalent Rayleigh quotient formulation of the above maximization problems (5),(6),(7), are:

Maximize$_{\alpha \neq 0} \ tr \{ \alpha^T (Y^T Y)^{-\frac{1}{2}} Y^T Z (Z^T Z)^{-1} Z^T Y (Y^T Y)^{-\frac{1}{2}} \}$, subject to $\alpha^T \alpha = I$. 

Maximize$_{\beta \neq 0} \ tr \{ \beta^T (Z^T Z)^{-\frac{1}{2}} Z^T Y (Y^T Y)^{-\frac{1}{2}} Y^T Z (Z^T Z)^{-\frac{1}{2}} \}$, subject to $\beta^T \beta = I$. 

3 Unconstrained Cost Functions

Given two data sets $Y, Z$, with dimensions $m \times p$ and $n \times p$, the objective is to search for directions $\alpha$ and $\beta$ that maximize the correlation between $Y \alpha$ and $Z \beta$ under certain constraints. In this section, we consider three dynamical systems for deriving gradient dynamical systems for approximating low rank canonical correlation analysis.

3.1 A Quartic Cost Function

Consider a quartic cost function given by:

$$F_1(x) = tr \{ x^T A y - \frac{1}{4} (x^T B z)^2 - \frac{1}{4} (y^T C y)^2 \},$$

(17)
where $A, B, C$ are scaled covariance matrices defined by

\[ B = \eta C_{yy} = (Y - \bar{Y})(Y - \bar{Y})^T, \]
\[ C = \eta C_{zz} = (Z - \bar{Z})(Z - \bar{Z})^T, \]
\[ A = \eta C_{xy} = (Y - \bar{Y})(Z - \bar{Z})^T, \]
\[ A^T = \eta C_{yx} = (Z - \bar{Z})(Y - \bar{Y})^T. \]

Here $\bar{Y}$ and $\bar{Z}$ are the sample means of $Y$ and $Z$, and $\eta$ is some positive number.

To understand the nature of the equilibrium points of the proposed systems, in the next example we consider the least dimensional problem.

**Example:** Assume that $a, b, c$ are nonzero real numbers and assume that $b, c > 0$. Consider the maximization problem:

\[ \text{Maximize } F_2(x, y) = axy - \frac{bx^4}{4} - \frac{cy^4}{4}, \quad (19) \]

The gradient and the Hessian matrix of $F_2$ are respectively given by

\[ \nabla F_2 = \begin{bmatrix} a & -bx^3 \\ 0 & a \end{bmatrix}, \quad \nabla^2 F_2 = \begin{bmatrix} 0 & a \\ -3ax^2 & 0 \end{bmatrix}. \]

The set of points for which $\nabla F_2 = 0$ includes:

\{(0, 0), (r, s), (r, -s), (-r, s), (-r, -s)\} where $r = \frac{|a|^2 - bx^3}{a}$ and $s = \frac{|a|^2 - cy^3}{a}$. Clearly if $(x, y)$ is a solution of $\nabla F_2 = 0$, then

\[ bx^4 = cy^4, \]
\[ |a|^2 xy = bc(xy)^3. \]

Thus the Hessian matrix is negative definite at any nonzero equilibrium point since $-3ax^2 < 0$ and the determinant is

\[ 9bc(xy)^2 - |a|^2 = 8|a|^2 > 0. \]

The zero solution $(0, 0)$ is neither a maximizer nor a minimizer of $F_2$ since the Hessian matrix evaluated at $(0, 0)$ is indefinite matrix:

\[ \nabla^2 F_2(0, 0) = \begin{bmatrix} 0 & a \\ a & 0 \end{bmatrix}. \]

Finally, the maximum of the function $F_2$ at $(0, 0)$ is undefined.

This shows that any nonzero critical values of $(19)$ is a maximizer.

This behavior can be shown to be the same for higher order systems.

In what follows, it will be assumed that

\[ \bar{A} = B^{-1}AC^{-1} = \sum_{k=1}^{\min(m, n)} \sigma_k u_k v_k^T, \quad (21) \]

is the singular value decomposition of the coherence matrix $\bar{A}$.

Here $\sigma_1 \geq \sigma_2 \geq \cdots$ and $u_k^T u_k = \delta(i, j), v_k^T v_k = \delta(i, j)$. Thus if $(u, \Sigma, v)$ represents the canonical triplet corresponding to the largest $r$ canonical correlations, then

\[ Av = B^{-1} u \Sigma, \]
\[ A^T u = C^{-1} v \Sigma. \]

Here $u = [u_1, \ldots, u_r], v = [v_1, \ldots, v_r], \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$.

The gradient of $F_1$ with respect to $x$ and $y$ is

\[ \nabla F_1 = \begin{bmatrix} Ay - Bx^T Bx \\ A^T x - Cy y^T C y \end{bmatrix}. \]

This yields the following dynamical system

\[ x' = Ay - Bx^T Bx, \]
\[ y' = A^T x - Cy y^T C y. \]

Lagrange stability theory can be applied to provide convergence and stability analysis of this system. Let $V = \frac{1}{2} \text{tr}(x^T x + y^T y)$, then

\[ \dot{V} = \text{tr}\{x^T Ay + y^T A^T x - (x^T Bx)^2 - (y^T Cy)^2\}. \]

Since $B$ and $C$ are assumed to be positive definite matrices, one can show that there is a bounded set $W$ such that $V > 0$ for each $(x, y) \in W$ and $V \leq 0$ for each $(x, y) \in W$. Additionally, $V(x, y) \to \infty$ as $x^T x + y^T y \to \infty$. Thus Theorem 7 (see Appendix) implies that every solution of (23) is bounded for $t \geq 0$.

The properties of the solutions of (23) as $t \to \infty$ are described in the following result.

**Theorem 1.** Consider the dynamical system (23). Let $(x(t), y(t))$ be a solution of (23) in the interval $[0, \infty)$, where $x(0), y(0)$ are given full rank matrices having nonzero components along the vectors $(u_1, \ldots, u_r)$, and $(v_1, \ldots, v_r)$, respectively. Define $\bar{B} = \lim_{t \to \infty} x(t)^T B x(t), \bar{A} = \lim_{t \to \infty} x(t)^T A y(t), \bar{C} = \lim_{t \to \infty} y(t)^T C y(t)$. Let $\bar{x}(t) = \lim_{t \to \infty} x(t), \bar{y} = \lim_{t \to \infty} y(t)$. Then $B = \bar{C}, A = \bar{A} = \bar{B} = \bar{B} = \bar{C}^2$ are positive definite. Additionally, the nonzero equilibrium points of (23) are

\[ \bar{x} = B^{\frac{1}{4}} u_P, \]
\[ \bar{y} = C^{\frac{1}{4}} v_P, \]

where $P$ is a nonsingular matrix such that

\[ PP^T = \Sigma, \]
\[ P^T P = B = \bar{C}. \]

The matrix $P$ satisfies

\[ \bar{B} = \bar{C} = \lim_{t \to \infty} x(t)^T B x(t). \]

Another parameterization of $\bar{x}$ and $\bar{y}$ is

\[ \bar{x} = B^{\frac{1}{4}} u P_2 \Sigma_1 u^T, \]
\[ \bar{y} = C^{\frac{1}{4}} v P_2 \Sigma_2 v^T, \]

where $P_2$ is a permutation matrix, $u$ is an orthogonal matrix that diagonalizes $B$ ($u^T B u$ is diagonal), and $D_3$ is a diagonal matrix satisfying $D_3^2 = I$.

**Proof:** Let $(x(t), y(t))$ be a solution of (23) in the interval $[0, \infty)$, and let $\bar{x} = \lim_{t \to \infty} x(t), \bar{y} = \lim_{t \to \infty} y(t)$, then rank$(x(t)) =$ rank$(x(0))$ and rank$(y(t)) =$ rank$(y(0))$ for $t \in [0, \infty)$. Now,

\[ A \bar{y} = B \bar{x}^T B \bar{x}, \]
\[ A^T \bar{x} = C \bar{y} \bar{y}^T C \bar{y}. \]

Premultiplying both sides of the previous equations by $\bar{x}^T$ and $\bar{y}^T$ we obtain

\[ \bar{A} = B^2, \]
\[ \bar{A}^T \bar{x} = \bar{C} \bar{y}. \]

Hence $\bar{A} = \bar{A}^T$ and $\bar{A}$ is positive semi-definite. It also follows that $\bar{C} = \bar{B}$. From the relations

\[ A \bar{y} = B \bar{x} \bar{B}, \]
\[ A^T \bar{x} = C \bar{y} \bar{C}, \]

and

\[ A \bar{u} = B \bar{v} \Sigma, \]
\[ A^T \bar{v} = C \bar{u} \Sigma, \]

it follows that

\[ \bar{x} = B^{\frac{1}{4}} u_P, \]

4871
and
\[ \bar{y} = C^{-1/2} vQ, \]
where \( P \) and \( Q \) are nonsingular matrices. Consequently, \( \bar{A} = P^T \Sigma Q, P^T P = B \) and \( Q^T Q = C \). Equations (28) imply
\[ \bar{A} = P^T \Sigma Q = P^T P P^T P, \]
\[ \bar{A}^T = Q^T \Sigma P = Q^T Q Q^T Q, \]
or equivalently (since \( \bar{B} = \bar{C} = P^T P = Q^T Q \)),
\[ \Sigma Q = P^T P = PQ^T Q, \]
\[ \Sigma P = Q^T Q = QP^T P, \]
This shows that
\[ \Sigma = P Q^T Q = P P^T, \]
or
\[ Q = \Sigma P^{-1}. \]
Substituting \( Q \) in the first equation of (30) yields:
\[ \Sigma P^{-1} = P P^T P, \]
\[ \Sigma^2 = (P P^T)^2, \]
Since \( \Sigma \) is positive definite
\[ \Sigma = P P^T. \]
Similarly,
\[ \Sigma = QQ^T. \]
Now, Equation (30) yields
\[ \Sigma Q = P P^T P = \Sigma P. \]
Hence \( Q = P \). Thus we have the following two equations:
\[ P P^T = \Sigma, \]
\[ P^T P = B = C. \]
Additionally, since
\[ P P^T P = \Sigma P, \]
it follows
\[ P B = \Sigma P, \]
or
\[ B = P^{-1} \Sigma P. \]
This shows that each column of \( P^{-1} \) is an eigenvector of \( B \). A more precise description of \( P \) is given next. Let \( P = z \Sigma_1 w^T \) be the singular value decomposition of \( P \), i.e., \( z^T z = I \), \( w^T w = I \) and \( \Sigma_1 \) is diagonal. Then
\[ \Sigma = P P^T = z \Sigma_1 z^T, \]
and
\[ \bar{B} = P^T P = w \Sigma_1 w^T. \]
Since \( \Sigma \) and \( \Sigma_1 \) are diagonal, there exists a permutation matrix \( P_1 \) so that
\[ z = P_1 D_1 \]
where \( D_1^2 = I \), and \( D_1 \) is diagonal.
Hence the final solution can be expressed as
\[ \bar{x} = B^{-1/2} u P \]
\[ \bar{y} = C^{-1/2} v P \]
and
\[ w \] is an orthogonal matrix such that \( B = w \Sigma_2 w^T \).

### 3.2 The Use of Triangular Matrices

Here we consider a modified version of the system (23):
\[ x' = Ay - Bz \begin{bmatrix} x \end{bmatrix} Bx, \]
\[ y' = A^T x - Cy \begin{bmatrix} y \end{bmatrix} Cy, \]
where the notation \( UT(X) \) represents the upper triangular part of \( X \). The analysis below will also apply if the function \( UT() \) in the system (33) is replaced with \( LT() \), which denotes the lower triangular part of \( () \).

The properties of the solutions of (33) as \( t \to \infty \) are described in the following result.

**Theorem 2.** Let \( (x(t), y(t)) \) be a solution of (33) in the interval \([0, \infty)\), where \( x(0), y(0) \) are given full rank matrices. Let \( \bar{x}, \bar{y}, \bar{B}, \bar{C}, \bar{A} \) be as defined in Theorem 1, and let \( u, v, \Sigma \) be as defined in (21) and (22). Then \( \bar{B}, \bar{C}, \bar{A} \) are diagonal such that \( \bar{B} = \bar{C} = \Sigma \), and \( \bar{A} = \Sigma^2 \). Moreover,
\[ \bar{x} = B^{-1/2} u \Sigma_1^2 D_1, \]
\[ \bar{y} = C^{-1/2} v \Sigma_2^2 D_1, \]
where \( D_1^2 = I \), and \( D_1 \) is diagonal.

**Proof:** Let \( (x(t), y(t)) \) be a solution of (33) in the interval \([0, \infty)\), where \( x(0), y(0) \). We show first that \( \bar{B} = \lim_{\infty} x(t)^T B x(t), \bar{A} = \lim_{\infty} x(t)^T A y(t), \) and \( \bar{C} = \lim_{\infty} y(t)^T C y(t) \) are diagonal. Assume that
\[ \bar{B} = U_1 + L_1, \]
\[ \bar{C} = U_2 + L_2, \]
where \( U_1 \) and \( U_2 \) are upper triangular matrices and \( L_1 \) and \( L_2 \) are strictly lower triangular matrices whose diagonal elements are zeros. From the relations:
\[ A \bar{y} = B z U T (x^T B x), \]
\[ A^T \bar{x} = C y U T (y^T C y), \]
the following relations hold:
\[ \bar{A} = B U_1, \]
\[ A^T = C U_2. \]
Solving (35) for \( U_1 \) and \( U_2 \) and substituting in (37) yield:
\[ \bar{A} = B (B - L_1) = B^2 - B L_1, \]
\[ A^T = C U_2 = C (C - L_2) = C^2 - C L. \]
Consequently,
\[ \bar{B}_2 - B L_1 = C^2 - L_2 C, \]
\[ B^2 - C^2 = B L_1 - L_2^T C = L_1^T B - C L_2, \]
\[ B L_1 + C L_2 = L_1^T C + L_1^T B, \]
\[ (U_1^T + L_1^T) L_1 + (U_2^T + L_2^T) L_2 = L_1^T (U_2 + L_2) + L_1^T (U_1 + L_1). \]
Thus,
\[ U_1^T L_1 + U_2^T L_2 = L_1^T U_2 + L_1^T U_1, \]
Since \( U_1^T L_1 + U_2^T L_2 \) and \( L_1^T U_2 + L_1^T U_1 \) are strictly lower and upper triangular matrices, respectively, then
\[ U_1^T L_1 + U_2^T L_2 = L_1^T U_2 + L_1^T U_1 = D_2 : \text{diagonal} \]
Since \( L_1 \) and \( L_2 \) are strictly lower triangular, it follows that \( D_2 = 0 \), i.e.,
\[ U_1^T L_1 + U_2^T L_2 = L_1^T U_2 + L_1^T U_1 = 0. \]
If it is assumed that \( \bar{B} = [b_{ij}] \) and \( \bar{C} = [c_{ij}] \) are diagonally dominant in the sense that if \( k \neq j \), then \( |b_{kj}| > |b_{kj}| \), and \( |c_{kk}| > |c_{kj}| \), then \( L_1 = 0 \) and \( L_2 = 0 \). Hence
\[ B = U_1, \]
\[ \bar{C} = U_2. \]
Since \( \bar{B} \) and \( \bar{C} \) are symmetric, it follows \( \bar{B}, \bar{C}, \bar{A} \) are diagonal:
\[ \bar{B} = D_3, \]
\[ \bar{C} = D_3, \]
\[ \bar{A} = D_3^2, \]
are diagonal.

Since \( \bar{A} = B^2 = C^2 \), then \( \Sigma = B^{-1} A \bar{C}^{-1} = B = C \), and \( \bar{A} = \Sigma^2 \).

Next we solve for \( \bar{x} \) and \( \bar{y} \). From the equations

\[ 4872 \]
\[ A\ddot{y} = B\ddot{x}D_3 = B\ddot{x}\Sigma, \]
\[ A^T\ddot{x} = C\ddot{y}D_3 = C\ddot{y}\Sigma, \]
and the equations
\[ Au = Bu\Sigma, \]
\[ A^Tv = Cu\Sigma, \]
it follows that
\[ \ddot{x} = B^{-\frac{1}{2}}uP, \]
and
\[ \ddot{y} = C^{-\frac{1}{2}}vQ, \]
where \( P \) and \( Q \) are nonsingular matrices. Consequently, \( \hat{A} = P^T\Sigma Q \), \( P^TP = B = \Sigma \) and \( Q^TQ = \hat{C} = \Sigma \). Equations (39) imply
\[ \Sigma Q = P\Sigma, \]
\[ \Sigma P = Q\Sigma. \]
(40)

Adding these equations yields
\[ \Sigma(P + Q) = (P + Q)\Sigma. \]

Proposition 5 (see Appendix) guarantees that \( P + Q \) is diagonal. Assume that \( P + Q = D_1 \), where \( D_1 \) is diagonal. By substituting the relation \( Q = D_1 - P \) in (40) we obtain
\[ \Sigma P + P\Sigma = D_2\Sigma. \]

Proposition 6 (see Appendix) implies that \( P \) is diagonal, and therefore \( Q = P = D_1 \). Now, \( \hat{A} = \hat{x}^T\hat{A}\ddot{y} = (\hat{x}^TB\hat{x})^2 \), and hence \( \hat{A} = P^T\Sigma Q = (P^TP)^2 = \Sigma^2 = P^T\Sigma P = P^2\Sigma. \) Thus,
\[ P^2 = \Sigma, \]
or \( P = D_2\sqrt{\Sigma} \) where \( D_1 \) is diagonal matrix such that \( D_1^2 = I \). Therefore,
\[ \ddot{x} = B^{-\frac{1}{2}}u\Sigma^\frac{1}{2}D_1, \]
\[ \ddot{y} = C^{-\frac{1}{2}}v\Sigma^\frac{1}{2}D_1. \]
This shows that the equilibrium solution is unique up to the multiplication of each component of \( \ddot{x} \) and \( \ddot{y} \) by \( \pm 1 \).

### 3.3 Weighted Quartic Cost Function

The cost function (17) may be modified as follows:
\[ F_3(x) = \text{tr}(x^TAyD - \frac{1}{4}(x^TBx)^2 - \frac{1}{4}(y^TCy)^2), \]
(41)
where \( D \) is a diagonal matrix with distinct positive diagonal elements. The gradient dynamical system based on \( F_3 \) is
\[ x' = AyD - Bxx^TBx, \]
\[ y' = A^Ttx - Cyy^TCy. \]
(42)

#### 3.3.1 Analysis of Stationary Points

The properties of the solutions of (42) as \( t \to \infty \) are described in the following result.

**Theorem 3.** Let \((x(t), y(t))\) be a solution of (42) in the interval \([0, \infty)\), where \(x(0), y(0)\) are given full rank matrices. Let \( \ddot{x}, \ddot{y}, B, C, A \) be as defined in Theorem 1. Then \( B, C, A \) are diagonal such that \( B = C = \Sigma D \), and \( A = \Sigma^2 D \). Moreover,
\[ \ddot{x} = B^{-\frac{1}{2}}uPD_2\sqrt{\Sigma}, \]
\[ \ddot{y} = C^{-\frac{1}{2}}vPD_2\sqrt{\Sigma}. \]
(43)

\( D_2 \) is a diagonal matrix such that \( D_2^2 = I \). Moreover \( \ddot{x}^TA\ddot{y} \) converges to a diagonal matrix \( \Sigma \) whose diagonal elements have the same ordering as those of \( D \).

**Proof:** Let \((x(t), y(t))\) be a solution of (42) in the interval \([0, \infty)\), where \(x(0), y(0)\) are given full rank. Assume that \( \ddot{x} = \lim_{t \to \infty} x(t) \) and \( \ddot{y} = \lim_{t \to \infty} y(t) \) are full rank. As \( t \to \infty \), equations (42) imply that
\[ \hat{A}D = B^2, \]
\[ \hat{A}^TD = C^2. \]
(44)

Thus \( \hat{A}D \) and \( \hat{A}^TD \) are symmetric:
\[ \hat{A}D = D\hat{A}^T, \]
\[ \hat{A}^TD = D\hat{A}. \]
(45)

By adding these equations we obtain
\[ (\hat{A} + \hat{A}^T)D = D(\hat{A} + \hat{A}^T). \]

Proposition 5 (see Appendix) implies that
\[ \hat{A} + \hat{A}^T = D_3, \]
(46)
for some diagonal matrix \( D_3 \). Solving (46) for \( \hat{A}^T \) and substituting in (45) yield
\[ \hat{A}D = (D_3 - \hat{A})D, \]
or
\[ \hat{A}D + D\hat{A} = DD_3. \]

Proposition 6 (see Appendix) implies that \( \hat{A} \) is diagonal and \( \hat{A} = \frac{D_1}{2} \), and thus
\[ B^2 = C^2 = \frac{D_1D_2}{2}, \]
\[ B = C = \sqrt{\frac{D_1D_2}{2}}. \]

Now, since \( \hat{A}, B, C \) are diagonal, it follows that
\[ \Sigma = B^{-\frac{1}{2}}AC^{-\frac{1}{2}}D = D^{-1}B. \]

Hence \( B = C = \Sigma D \), and \( \hat{A} = \Sigma^2 D \). To determine \( \ddot{x} \) and \( \ddot{y} \), the relations
\[ Au = Bu\Sigma, \]
\[ A^Tv = Cu\Sigma, \]
imply that
\[ \ddot{x} = B^{-\frac{1}{2}}uP, \]
and
\[ \ddot{y} = C^{-\frac{1}{2}}vQ. \]
where \( P \) and \( Q \) are nonsingular matrices. Since \( \ddot{x}^TB\ddot{x} = D\Sigma \) and \( \ddot{y}^TC\ddot{y} = D\Sigma \), then \( P^TP = D\Sigma \) and \( Q^TQ = D\Sigma \). This implies that \( P = u_1\sqrt{D\Sigma} \) and \( Q = v_1\sqrt{D\Sigma} \) for some orthogonal matrices \( u_1 \) and \( v_1 \). Substituting these relations in (46) yields
\[ AC^{-\frac{1}{2}}vQD = BuPP^TP, \]
and
\[ A^TvPD_2\sqrt{\Sigma} = Cu_1\sqrt{D\Sigma}, \]
\[ u_SQD = vPP^TP, \]
\[ v\Sigma PD = u_1\sqrt{D\Sigma}, \]
or
\[ \Sigma QD = PP^TP, \]
\[ \Sigma PD = QQ^TQ. \]

From the relations \( P^TP = \Sigma D \) and \( Q^TQ = \Sigma D \) we obtain:
\[ \Sigma QD = PD\Sigma, \]
\[ \Sigma PD = Q\Sigma D, \]
and consequently,
\[ \Sigma Q = \Sigma D, \]
\[ \Sigma P = \Sigma D. \]

Thus by adding the last two equations and invoking Propositions 5 and 6, it follows that \( P = Q \) is diagonal. Hence \( P^TP = Q^TQ = \Sigma D \).
\[ P^2 = Q^2 = D\Sigma \text{ and therefore, } P = Q = \sqrt{D\Sigma D_2}, \text{ where } D_2 \text{ is a diagonal matrix such that } D_2^2 = I. \]

Therefore,
\[
\bar{x} = B^{-1} uP = B^{-1} uP_1 D_2 \sqrt{D\Sigma}, \\
\bar{y} = C^{-1} vP = C^{-1} vP_1 D_2 \sqrt{D\Sigma}.
\]

The maximum of \( F_q \) is \( \frac{1}{2} \text{tr}(D^2 \Sigma^2) \), where \( \Sigma \) is diagonal having the same diagonal elements which are a permutation of the diagonal elements of \( \Sigma \). Since both \( D \) and \( \Sigma \) are diagonal, \( \text{tr}(D^2 \Sigma^2) \) is maximum if the diagonal elements of \( \Sigma \) have the same ordering as those of \( D \). In other words, if \( d_i < d_j \), then \( \sigma_i < \sigma_j \). In particular, if the diagonal elements of \( D \) are in decreasing or increasing order, then the diagonal elements of \( \Sigma \) are in decreasing or increasing order, respectively. Thus the final solution is
\[
\bar{x} = B^{-1} uP = B^{-1} P_1 uP_1 D_2 \sqrt{D\Sigma}, \\
\bar{y} = C^{-1} vP = C^{-1} P_1 vD_2 \sqrt{D\Sigma}.
\]

where \( P_1 \) is a permutation matrix such that \( \text{tr}((P_1 \Sigma P_1^T)^2 D^2) \) is maximum.

4 Conclusion

We have provided derivations of various cost functions for computing the CCA problem. Most of the development is devoted to the analysis of several variations of a dynamical system derived from the gradient of a quartic cost function. The equilibrium points of these systems are completely described. Global convergence and stability are established using Lagrange stability. Preliminary simulations, which are not shown here due to space limits, have shown that these systems converge from a wide range of initial conditions and demonstrating global convergence behavior. It is also observed that convergence may be accelerated by incorporating penalty terms in the cost functions. By incorporating diagonal or triangular matrices, the actual canonical variates are shown to be uniquely determined up to a permutation and multiplication by \( \pm 1 \).

5 Appendix

Finally, we state a few results which are essential for the derivations of the proposed methods.

5.1 Gradient and Hessian Matrices

The gradient and Hessian matrices can be obtained from first and second order differentials as the following lemma [10].

**Lemma 4.** Let \( \phi \) be a twice differentiable real-valued function of an \( n \times p \) matrix. Then, the following relationships hold:

\[
\begin{align*}
\nabla \phi(\bar{x}) &= \text{vec}(A^T d\bar{x}) \iff \nabla \phi(\bar{x}) = A \\
\nabla^2 \phi(\bar{x}) &= \text{vec}(B(4\bar{x}^T C d\bar{x}) + H(\bar{x})) = \frac{1}{2} (B^T \otimes C + B \otimes C^T) \\
\n\nabla^3 \phi(\bar{x}) &= \text{vec}(B(d(4\bar{x}^T C d\bar{x}) + H(\bar{x}))) = \frac{1}{2} K_{mn}(B^T \otimes C + C^T \otimes B)
\end{align*}
\]

where \( d \) denotes the differential, and \( A, B, C \) are matrices, each of which may be a function of \( X \). The gradient of \( \phi \) with respect to \( X \) and the Hessian matrix of \( \phi \) at \( \bar{X} \) are defined as

\[
\begin{align*}
\nabla \phi(\bar{X}) &= \frac{\partial \phi(\bar{X})}{\partial \bar{X}} \\
H(\bar{X}) &= \frac{\partial}{\partial \text{vec}(X)^T} \left( \frac{\partial \phi(\bar{X})}{\partial \text{vec}(X)^T} \right)^T.
\end{align*}
\]

where vec is the vector operator and stands for the operation of stacking the columns of a matrix into one column, and \( \otimes \) denotes the Kronecker product. The matrix \( K_{mn} \) denotes the \( mn \times mn \) commutation matrix; \( K_{mn} = K_{pn} = K_{np} \) and \( K_{mn}(A \otimes C) = (C \otimes A)K_{mn} \), where \( A \in \mathbb{R}^{m \times n} \) and \( C \in \mathbb{R}^{r \times q} \).

**Proposition 5** [11]. Let \( D, C \in \mathbb{R}^{m \times n} \) such that \( D \) is diagonal having distinct eigenvalues. If \( CD = DC \), then \( C \) is diagonal.

**Proposition 6** [11]. Let \( B, D \in \mathbb{R}^{p \times p} \) and assume that \( D \) is diagonal and all eigenvalues of \( D \) are distinct. If \( BD + DB \) is diagonal, then \( B \) is diagonal.

We state next a well known result about Lagrange stability. A dynamical system is Lagrange stable if the continuous state remains bounded from any initial condition. For example, if the continuous state converges to a stationary set, the dynamical system is Lagrange stable.

**Theorem 7** (Lagrange Stability Theorem) [12]. Consider the system \( x' = g(x) \) where \( g : \mathbb{R}^n \to \mathbb{R}^n \), \( N \) is a positive integer, is continuously differentiable function. Let \( W \) be a bounded neighborhood of the origin and let \( W' \) be its complement \( W' \) is the set of all points outside \( W \). Assume that \( V(x) \) is a scalar function with continuous first partial derivatives in \( W' \) and satisfying:

1. \( V(x) > 0 \) for all \( x \in W' \),
2. \( V(x) \leq 0 \) for all \( x \in W' \),
3. \( V(x) \to \infty \) as \( ||x|| \to \infty \).

Then each solution of \( x' = g(x) \), is bounded for all \( t > 0 \).

**Theorem 8** [13] (The Inequalities of Amir-Moez). Let \( A \) and \( B \) be positive definite (strictly positive) operators on an \( n \)-dimensional Hilbert space \( \alpha_1, \gamma_1 \), be the \( i \)-th eigenvalues of \( A, B \) and \( AB \) respectively in a descending enumeration, i.e., \( \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \), \( \beta_1 \geq \beta_2 \geq \cdots \geq \beta_n \), and \( \gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_n \). Let \( i + j \leq n + 1 \), then \( \gamma_{i+j} \leq \alpha_i \beta_j \). If \( i + j \geq n + 1 \), then \( \gamma_{i+j} \geq \alpha_i \beta_j \).

**References**


