

Research Article

Stan Lipovetsky*

Orthonormal Canonical Correlation Analysis

<https://doi.org/10.1515/stat-2020-0104>

Received Jun 07, 2021; accepted Aug 30, 2021

Abstract: Complex managerial problems are usually described by datasets with multiple variables, and in lack of a theoretical model, the data structures can be found by special multivariate statistical techniques. For two datasets, the canonical correlation analysis and its robust version are known as good working research tools. This paper presents their further development via the orthonormal approximation of data matrices which corresponds to using singular value decomposition in the canonical correlations. The features of the new method are described and applications considered. This type of multivariate analysis is useful for solving various practical problems of applied statistics requiring operating with two data sets, and can be helpful in managerial estimations and decision making.

Keywords: canonical correlations, multicollinearity, robustness, singular value decomposition, orthonormal approximation, managerial decisions.

2020 Mathematics Subject Classification: 15A09, 15A18, 62H20

1 Introduction

Multivariate statistical analysis presents a natural approach to description and modeling of big data available nowadays on various real managerial objects and problems. In a lack of an adequate theoretical model, which is a common situation for any practical research with complex social-economic data, a proxy structure can be identified by special multivariate statistical techniques. For two sets of variables of different designation, the Canonical Correlation Analysis (CCA) is the main tool for finding parameters for their aggregation into the total scores with maximum correlation (Dillon and Goldstein, 1984; Hardle and Hlavka, 2007; Izenman, 2008; Hardle and Simar, 2012). CCA serves to find how one group of the predictor variables can influence another group of the outcome variables.

The CCA technique was originated by H. Hotelling in 1936 for studying correlations between the aggregated variables of two datasets, and has been developed in multiple aspects, including nonlinear analysis and several data sets (Horst, 1961; Fornell, 1982; Tishler and Lipovetsky, 1996; Kessy *et al.*, 2018). Its applications are known in different fields, from management and information systems to machine learning and biometrics – see examples in (Ahituv *et al.*, 1998; Hair *et al.*, 1998; Hardoon *et al.*, 2004; Andrew *et al.*, 2013; Adrover and Donato, 2015; Cao *et al.*, 2015; Wilms and Croux, 2015; Jendoubi and Strimmer, 2019).

The Robust Canonical Analysis (RCA), a modification of classical CCA, was developed in (Tishler and Lipovetsky, 2000). The RCA in other formulations corresponds to the Partial Least Squares (PLS) technique (Bookstein *et al.*, 1990; Varmuza and Filzmoser, 2009), specifically, its C_2 algorithm (Wold, 1975), and the inter-battery factor analysis (Tucker, 1958). Further extensions of CCA and RCA, related to canonical covariance and restricted eigenvector analysis, have been considered for solving various socio-economic problems in (Tishler *et al.*, 1996; Lipovetsky *et al.*, 2002; Tam, 2003; Shenhar *et al.*, 2002; Dvir *et al.*, 2003; Carmeli and Tishler, 2004; Lipovetsky, 2012).

*Corresponding Author: Stan Lipovetsky; Independent consultant, Minneapolis 55305 MN, United States of America; Email: stan.lipovetsky@gmail.com

The RCA method can be also described as the Singular Value Decomposition (SVD) of the cross-correlations matrix between the variables of two data sets, and in contrast to the regular CCA it does not contain the matrices inversion, therefore, has no impact from the matrices' ill-conditioning, so its results are not prone to multicollinearity among the variables within each set of variables. Classical SVD method is widely applied in multivariate statistics – see, for example, in (Horn and Johnson, 2013; Gentle, 2017; Mair, 2018; Demidenko, 2020; Irizarry, 2020). Various modifications of SVD have been introduced for specific solutions (Lipovetsky and Tishler, 1994; Lipovetsky and Conklin, 2005a; Lipovetsky, 2009, 2016).

A special application of SVD to finding the predictors' relative importance in multiple regression was described in (Lipovetsky and Conklin, 2015). The approach taken in that work is based on the so-called orthonormal matrix approximation to a dataset, originated in the works (Gibson, 1962; Johnson, 1963, 1966). Applying this technique to the canonical correlation matrices yields a new technique of the ORthonormal Canonical Analysis (ORCA), which holds the good properties of both CCA and RCA. The new method and its properties are described and applications considered. This type of modern multivariate analysis can serve in solving various practical problems and can be helpful to managerial estimations and decision making.

The paper is organized as follows. Regular CCA and RCA are described in Section II, the SVD and matrix orthonormal approximation are presented in Section III, and the ORCA method is introduced in Section IV. Relation to multiple regression is shown in Section V, numerical examples and comparisons are given in Section VI, and Section VII summarizes.

2 CCA and RCA techniques

Let us briefly describe canonical correlation analysis, or CCA, and its modification in the robust canonical analysis, or RCA. Consider two data sets presented in the matrices X and Y of the orders $N \times n$ and $N \times m$, where N is a number of observations, and n and m are numbers of variables, or columns in X and Y , respectively. Assume for convenience that the variables are centered and normalized by their standard deviations. The scores of each data set aggregation variables are:

$$\xi = Xa, \quad \eta = Yb, \quad (1)$$

where ξ and η are vector-columns both of the N -th order, and a and b are the vector-columns of parameters of the order n and m , respectively. Canonical correlation between these data sets, or the pair correlation of their scores ξ and η , measures a connection between these sets:

$$\rho = \frac{\text{cov}(\xi, \eta)}{\sqrt{\text{var}(\xi)\text{var}(\eta)}} = \frac{\xi' \eta}{\sqrt{(\xi' \xi)(\eta' \eta)}} = \frac{a' X' Y b}{\sqrt{(a' X' X a)(b' Y' Y b)}} = \frac{a' R_{xy} b}{\sqrt{(a' R_{xx} a)(b' R_{yy} b)}} \quad (2)$$

where cov and var denote sample covariance and variance, the prime denotes a matrix transposition, and the correlation matrices are denoted as follows:

$$R_{xx} = X'X, \quad R_{yy} = Y'Y, \quad R_{xy} = X'Y, \quad R_{yx} = Y'X. \quad (3)$$

CCA estimates the vectors a and b via maximizing the correlation (2) by the conditional Lagrange objective

$$L = a' R_{xy} b - \frac{\lambda}{2} (a' R_{xx} a - 1) - \frac{\mu}{2} (b' R_{yy} b - 1) \quad (4)$$

where λ and μ are Lagrange terms. Differentiating L (4) with respect to the vectors a and b and putting these differentials to zero yields a system of equations:

$$R_{xy} b = \lambda R_{xx} a, \quad R_{yx} a = \mu R_{yy} b \quad (5)$$

Multiplying vector-rows and by the first and second equations (5), respectively, and using conditions (4) yields the equality of the Lagrange multipliers to the bilinear form:

$$\lambda = \mu = a' R_{xy} b. \quad (6)$$

Solving the second equation (5) for the vector b and substituting it into the first one, also symmetrically solving the first equation (5) for the vector a and substituting it into the second one, produces the expressions:

$$R_{xy}R_{yy}^{-1}R_{yx}a = \lambda^2 R_{xx}a, \quad R_{yx}R_{xx}^{-1}R_{xy}b = \lambda^2 R_{yy}b. \quad (7)$$

Finally, the equations (7) can be rewritten in the regular CCA form:

$$R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}a = \lambda^2 a, \quad R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy}b = \lambda^2 b. \quad (8)$$

The eigenproblems (8) present classical CCA solution with the eigenvalues for the squared canonical correlations λ^2 and the corresponding pairs of eigenvectors a and b .

In difference to CCA, the RCA considers the covariance of the scores by the standardized data, and estimates the normalized vectors a and b which maximize the numerator in (2) when the objective (4) becomes:

$$L = a'R_{xy}b - \frac{\lambda}{2}(a'a - 1) - \frac{\mu}{2}(b'b - 1). \quad (9)$$

Then (5) reduces to the equations with identity matrices instead of the R_{xx} and R_{yy} matrices:

$$R_{xy}b = \lambda a, \quad R_{yx}a = \mu b. \quad (10)$$

The relation (6) can be obtained from (10) as well, and (10) yields the eigenproblems:

$$R_{xy}R_{yx}a = \lambda^2 a, \quad R_{yx}R_{xy}b = \lambda^2 b. \quad (11)$$

The eigenproblems formulation (10)–(11) correspond to the singular value decomposition of the inter-correlations' matrix R_{xy} between the variables in two data sets, and the eigenvectors a and b suggest the solution for the maximum covariance between the aggregates (1). In contrast to CCA (8), the RCA solution (11) does not contain any matrix inversions and stays robust in multicollinearity among the variables in any set.

3 SVD and matrix orthonormal approximation

Let us briefly consider singular value decomposition on the example of the matrix X , remembering that it can be repeated for another matrix Y as well. SVD can be described as a matrix approximation by the outer product of the vectors p and q of the N -th and n -th order, respectively:

$$X = \Delta p q' + \epsilon, \quad (12)$$

where Δ is a constant term, and ϵ denotes the matrix of residuals. In the least squares (LS) objective of the squared error minimization

$$S^2 = \|\epsilon\|^2 = \|X - \Delta p q'\|^2, \quad (13)$$

from the first order conditions for the extremum $\partial S^2 / \partial p = 0$, $\partial S^2 / \partial q = 0$, taking into account the vectors normalization $p'p = 1$, $q'q = 1$, we obtain a system of equations:

$$Xq = \Delta p, \quad X'p = \Delta q. \quad (14)$$

Substituting one of these equations into another produces the eigenproblems

$$(X'X)q = \Delta^2 q, \quad (XX')p = \Delta^2 p, \quad (15)$$

where Δ^2 are the eigenvalues. They yield the matrix spectral decomposition

$$X = \Delta_1 p_1 q_1' + \Delta_n p_n q_n', \quad (16)$$

where the matrix rank equals the number of predictors. Combining the eigenvectors p and q into the columns of the matrices P and Q , and all the terms Δ (square roots of the eigenvalues (15) known as singular values) into the diagonal matrix

$$D = \text{diag}(\Delta_1, \dots, \Delta_n), \quad (17)$$

we can represent the matrix X in the SVD decomposition

$$X = PDQ'. \quad (18)$$

The SVD formulation (18) presents the matrix X as product of the orthonormal matrix P , diagonal matrix D , and orthogonal matrix Q , which have the following properties:

$$P'P = I, \quad Q'Q = QQ' = I, \quad (19)$$

where I is the n -th order identity matrix.

By (18) and (19), the correlation matrix equals:

$$X'X = QDP'PDQ' = QD^2Q', \quad (20)$$

where D^2 is defined as the diagonal matrix (17) of the singular values squared, that is the matrix of eigenvalues $D^2 = \text{diag}(\Delta_1^2, \dots, \Delta_n^2)$. Multiplying the equation (20) from the right-hand side by the matrix Q yields the relation

$$(X'X)Q = QD^2, \quad (21)$$

which corresponds to the first eigenproblem in (15) presented in the matrix form. It is nothing else but the principal component analysis (PCA) eigenproblem for the correlation matrix $X'X$, with the eigenvalues in the matrix D^2 and the eigenvectors in columns of the matrix Q . The second eigenproblem in (15) defines the PCA scores, due to the resulting from (14) relation $p = Xq/\Delta$.

A function of a square matrix is defined by its application to the eigenvalues in the matrix decomposition by eigenvectors, so the square root of the correlation matrix (20) and its inversion are defined as follows:

$$(X'X)^{1/2} = QDQ', \quad (X'X)^{-1/2} = QD^{-1}Q'. \quad (22)$$

Indeed, the product of two square root matrices, taking into account the properties of the orthogonal Q matrix, is

$$(X'X)^{1/2}(X'X)^{1/2} = QDQ'QDQ' = QD^2Q' = X'X, \quad (23)$$

which returns to the same correlation matrix (20). Also, the product of both matrices in (22) leads, as it should be, to the identity matrix:

$$(X'X)^{1/2}(X'X)^{-1/2} = QDQ'QD^{-1}Q' = I. \quad (24)$$

Now let us describe the so-called a matrix X orthonormal approximation by its SVD vectors. For this aim, using (18)–(19) and (22) consider the following matrix form:

$$(X'X)^{-1/2}X' = QD^{-1}Q'QDP' = QP' \equiv Z', \quad (25)$$

where the matrix Z is defined as

$$Z = PQ'. \quad (26)$$

The matrix Z was introduced in (Gibson, 1962; Johnson, 1966) and used for regression modeling in (Lipovetsky and Conklin, 2015). With relations (19), this matrix has the property of orthonormality:

$$Z'Z = QP'PQ' = QQ' = I. \quad (27)$$

Due to (18) and (26), projections of X by Z , and Z by X , produce the same symmetric matrix:

$$Z'X = X'Z = QDQ' = (X'X)^{1/2}. \quad (28)$$

The matrix Z (26) differs from the matrix X (18) by changing the diagonal matrix D of singular values (17) to the identity matrix I . Thus, the matrix Z does not depend on singular values in (17), and it is not prone to inversion problems and multicollinearity in data.

The matrix Z presents the best orthonormal approximation to the data matrix X , and as shown in (Lipovetsky and Conklin, 2015) it is possible to improve this approximation to a better fit of X by Z . For this aim, consider a constant parameter t with Z matrix, when the LS objective is:

$$\begin{aligned} LS &= \sum_{i=1}^N \sum_{k=1}^n (X - tZ)_{ik}^2 = \text{Tr}((X - tZ)'(X - tZ)) = \text{Tr}(X'X - 2tQDQ' + t^2I) \\ &= n - 2t\text{Tr}(QDQ') + t^2n = n - 2t\text{Tr}(DQ'Q) + t^2n = n - 2t\text{Tr}(D) + t^2n = n(1 - 2t\bar{\Delta} + t^2), \end{aligned} \quad (29)$$

where (27)–(28) are accounted, the trace of correlation matrix equals its order $\text{Tr}(X'X) = n$, and $\bar{\Delta}$ denotes the mean value of the singular values (17):

$$\bar{\Delta} = \text{Tr}(D)/n = \frac{1}{n} \sum_{k=1}^n \Delta_k. \quad (30)$$

The convex function (29) reaches its minimum at $t = \bar{\Delta}$, thus, a higher precision in approximation of X by the orthonormal matrix is $X = \bar{\Delta}PQ'$. All the relations (12)–(30) can be rewritten for Y matrix as well.

4 ORCA method

Let us apply the considered orthonormal presentation technique to the CCA problem, where are two data matrices X and Y , and for each of them there is the SVD decomposition (18) which can be rewritten more specifically for each matrix:

$$X = P_x D_x Q_x', \quad Y = P_y D_y Q_y'. \quad (31)$$

The relation (25) can be represented for both matrices (31) as follows

$$R_{xx}^{-1/2} X' = (X'X)^{-1/2} X' = Q_x P_x' = Z_x', \quad R_{yy}^{-1/2} Y' = (Y'Y)^{-1/2} Y' = Q_y P_y' = Z_y' \quad (32)$$

where two matrices Z are defined similarly to (26).

In regression modeling, the multicollinearity effect occurs when correlation matrix is close to a singular matrix, or some its eigenvalues are close to zero. Then inversion of such a matrix corresponds to division by a value close to zero, which inflates the model coefficients to big values of both signs. The multicollinearity within anyone of two data sets can impact the CCA estimations (8) deteriorating the solution.

Let us describe how to escape these undesirable effects with help of the orthonormal canonical analysis, or in the ORCA approach. Consider the first of the CCA relations (8) which can be rewritten in the following structure, taking into account the definitions (3)

$$R_{xx}^{-1/2} X' Y R_{yy}^{-1/2} R_{yy}^{-1/2} Y' X R_{xx}^{-1/2} R_{xx}^{1/2} a = \lambda^2 R_{xx}^{1/2} a. \quad (33)$$

Using the constructs (32) we can represent the formula (33) via the orthonormal Z matrices:

$$Z_x' Z_y Z_y' Z_x R_{xx}^{1/2} a = \lambda^2 R_{xx}^{1/2} a. \quad (34)$$

A similar relation, up to interchange of x and y subscripts, can be derived from the second equation (8) for the vector b . Introducing two new vectors alpha and beta defined by the transformations

$$\alpha = R_{xx}^{1/2} a, \quad \beta = R_{yy}^{1/2} b, \quad (35)$$

we can rewrite the expression (34) and its counterpart for another vector as follows:

$$Z'_x Z_y Z'_y Z_x \alpha = \lambda^2 \alpha, \quad Z'_y Z_x Z'_x Z_y \beta = \lambda^2 \beta. \quad (36)$$

These eigenproblems present the ORCA solution expressing the original CCA (8) via the orthonormal matrices (26) for both original matrices as in (32):

$$Z_x = P_x Q'_x, \quad Z_y = P_y Q'_y. \quad (37)$$

Such matrices contain the SVD vectors but not the singular values, so they are neither ill-conditioned nor prone to collinearity in data, thus, the ORCA problems (36) yield the robust solutions for the eigenvectors.

Similar to the matrices $R_{xy} = X'Y$ and $R_{yx} = Y'X$ built in (3) by the original data X and Y , consider the cross-product matrices S_{xy} and S_{yx} via the orthonormal matrices Z_x and Z_y :

$$S_{xy} = Z'_x Z_y, \quad S_{yx} = Z'_y Z_x. \quad (38)$$

Then the eigenproblems (36) can be simplified as follows:

$$S_{xy} S_{yx} \alpha = \lambda^2 \alpha, \quad S_{yx} S_{xy} \beta = \lambda^2 \beta, \quad (39)$$

that reminds the RCA technique described in (11). Thus, the ORCA solution presents the SVD (39) applied to the products (38) of the orthonormal approximation matrices (38).

With the eigenvectors from (39), the original vectors of CCA (8) can be found from the relations (35) inverted:

$$a = R_{xx}^{-1/2} \alpha = Q_x D_x^{-1} Q'_x \alpha, \quad b = R_{yy}^{-1/2} \beta = Q_y D_y^{-1} Q'_y \beta. \quad (40)$$

Each one of the original CCA eigenproblems (8) contains two inverted matrices R_{xx}^{-1} and R_{yy}^{-1} , that produces an impression that both vectors a and b are prone to multicollinearity in any original group of variables. However, the results (40) prove that vectors a and b depend on possible ill-conditioning only in one matrix of X or Y data which it aggregates (1). For example, if the matrix D_x contains singular values close to zero, but the matrix D_y does not have small values, then only the vector a is impacted by multicollinearity.

It is possible to substitute the relations (37) into (36) to represent ORCA eigenproblems in a deeper data structure as follows:

$$Q_x P'_x P_y Q'_y Q_y P'_y P_x Q'_x \alpha = \lambda^2 \alpha, \quad Q_y P'_y P_x Q'_x Q_x P'_x P_y Q'_y \beta = \lambda^2 \beta. \quad (41)$$

For the orthogonal matrices Q , with properties (19) and $Q^{-1} = Q'$ we can transform (41) to:

$$P'_x P_y P'_y P_x Q'_x \alpha = \lambda^2 Q'_x \alpha, \quad P'_y P_x P'_x P_y Q'_y \beta = \lambda^2 Q'_y \beta. \quad (42)$$

Introducing new vectors by the transformations

$$\gamma = Q'_x \alpha = Q'_x R_{xx}^{1/2} a, \quad \delta = Q'_y \beta = Q'_y R_{yy}^{1/2} b, \quad (43)$$

we can reduce the relations (42) to the expressions

$$P'_x P_y P'_y P_x \gamma = \lambda^2 \gamma, \quad P'_y P_x P'_x P_y \delta = \lambda^2 \delta. \quad (44)$$

These eigenproblems reveal the simple CCA structure given via the orthonormal P matrices of the SVD solutions (31).

Similar to the matrices S_{xy} and S_{yx} built from the orthonormal matrices Z_x and Z_y in (38), we can introduce cross-product matrices T_{xy} and T_{yx} from the orthonormal matrices in (44):

$$T_{xy} = P'_x P_y, \quad T_{yx} = P'_y P_x. \quad (45)$$

Then the eigenproblems (44) reduce to the following forms:

$$T_{xy} T_{yx} \gamma = \lambda^2 \gamma, \quad T_{yx} T_{xy} \delta = \lambda^2 \delta, \quad (46)$$

that corresponds to the RCA techniques (11) and (39). Therefore, the ORCA solution presents the combined SVD (46) applied to the product (45) of the separated SVD scores P_x and P_y (31). When the problems (46) are solved, the original vectors of CCA (8) can be found, similarly to (40), from the relations (43) inverted:

$$a = R_{xx}^{-1/2} Q_x \gamma = Q_x D_x^{-1} \gamma, \quad b = R_{yy}^{-1/2} Q_y \delta = Q_y D_y^{-1} \delta. \quad (47)$$

The ORCA eigenproblems (39) or (46) have the same canonical correlations values λ . Their eigenvectors depend on the normalization of the matrices used, but these vectors can be transformed by (40) or (47) back to the original CCA. The numerical results show that ORCA solution (39) presents results close both to the CCA and RCA, so its results are convenient for interpretation and robust.

5 Relation to multiple regression

As it was discussed by the relations (40), and we can see it by (47) too, the vectors a and b depend on ill-conditioning only in the corresponding matrix X or Y . The extent at which the vectors are impacted by multicollinearity depend on the inverted singular values in the matrices D_x^{-1} or D_y^{-1} , respectively. It means that the level of multicollinearity is rather measured by the square root of the inverted matrices, $R_{xx}^{-1/2}$ and $R_{yy}^{-1/2}$, not by the inverted correlation matrices R_{xx}^{-1} and R_{yy}^{-1} themselves. This conclusion can be related to measuring a degree of multicollinearity in regression modeling as well.

Indeed, if in one data set (1) there is only one variable y , then the CCA problem reduces to the multiple regression model. With one variable y , the vector b is degenerating to a scalar, the matrix R_{xy} is reducing to the vector r_{xy} of the regressors x correlations with the outcome. Then the first equation (5), taking the constant terms into the vector of parameters a , reduces to the expression:

$$r_{xy} = R_{xx} a, \quad (48)$$

that is nothing else but the normal system of equations for the regression model of y by x s.

It is interesting to note that the second relation (5) reduces at the same time to the scalar product of the vector r_{xy} and the vector of regression parameters a , which defines the coefficient of multiple determination $r'_{xy} a = R^2$ of the quality of model fit. Solution of the system (48) is given via the matrix inversion as follows:

$$a = R_{xx}^{-1} r_{xy} = (X'X)^{-1} X'y. \quad (49)$$

The expression (49) produces an impression that the impact of multicollinearity can be measured by the inverted correlation matrix. The so-called variance inflation factor defined by the diagonal elements of this matrix is often used as a gauge for each j -th predictor influence on the effect of multicollinearity (Lipovetsky and Conklin, 2005b):

$$VIF_j = [(X'X)^{-1}]_{jj}. \quad (50)$$

The elements of the inverted matrix in (49)–(50), as we can see by (20)–(22), are defined with the squared matrix of the inverted singular values, $D^{-2} = \text{diag}(\Delta_1^{-2}, \dots, \Delta_n^{-2})$, or the inverted eigenvalues of the correlation matrix. However, the solution (49) is defined via the non-squared inverted singular values. Indeed, using SVD results from (18) and (20) in (49), we can re-write it as follows:

$$a = (X'X)^{-1} X'y = QD^{-2} Q' QDP'y = QD^{-1} P'y. \quad (51)$$

The expression (51) reveals that the solution (49) for the coefficients of multiple regression depends not on the eigenvalues Δ_j^{-2} , but on the singular values Δ_j^{-1} in the matrix D^{-1} . It means that in measuring the multicollinearity impact on the regression parameters, instead of the scale of VIF -factors (50) which depends on D^{-2} , it makes sense to use a more adequate to (51) scale related on D^{-1} . For example, in place of the indicator (50), it could be better to apply such a measure as the square root of it. In another evaluation, taking (22) in account, the VIF -adjusted can be rather evaluated as following:

$$VIF.adj_j = [(X'X)^{-1/2}]_{jj} = [QD^{-1} Q']_{jj}. \quad (52)$$

The expression (52) can be also used as the measure of multicollinearity for CCA and ORCA problems.

Another important note can be made by the expression (51) which actually presents the so-called pseudo-inverse, or a generalized matrix inverse

$$A^+ = (X'X)^{-1}X', \tag{53}$$

also known as the Moore-Penrose inverse. It presents the least squares solution (49) of the system of linear equations (48) expressed via the SVD matrices (18) or (31). This solution can be obtained directly from the linear model $y = Xa$ written via the orthonormal P , diagonal D , and orthogonal Q matrices: which have the following properties

$$y = PDQ'a. \tag{54}$$

Multiplying this relation from the left-hand side subsequently by the operators P' , D^{-1} , and Q , and using the properties (19) yields the equation

$$QD^{-1}P'y = a, \tag{55}$$

which coincides with the solution (51) for the linear regression parameters. Therefore, the regression model as well as the CCA and its modifications can be expressed in the terms of the matrices generalized inversion.

6 Numerical examples

Consider a data set on a financial problem in the file *LifeCycleSavings* used in the R statistical package as example for canonical correlation analysis performed by the function *cancor*. The first matrix X contains two variables (x_1 – pop15, x_2 – pop 75, or percentage of population under 15 and over 75 years old, respectively), and the second matrix Y consists of three variables (y_1 – sr, or the saving ratio, presented by the aggregate personal saving divided by the disposable income; y_2 – dpi, or real per-capita disposable income; y_3 – ddpi, the percentage rate of change in per-capita disposable income). The data is given by 50 countries, with characteristics averaged over the decade to remove the business cycle or other short-term fluctuations, and it is described in (Belsley *et al.*, 1980).

The outcome of the function *cancor*(X, Y) is presented in three numerical columns at the left-hand side in Table 1 as the CCA original solution built by the covariance matrices. The same data centered and normalized by the square root of sum of squares (SRSS, given in the central column in Table 1) correspond to using correlation matrices in the CCA problems (8) which yield results shown in three columns at the right-hand side in Table 1. The CCA solutions, similarly to regression model parameters, are invariant to the scaling transformation, so

Table 1: CCA for financial data-1.

variables	CCA by covariance matrices			SRSS	CCA by correlation matrices		
X	a_1	a_2			a_1	a_2	
x_1	-0.0091	-0.0362		64.0621	-0.5837	-2.3205	
x_2	0.0486	-0.2603		9.0354	0.4395	-2.3520	
norm a					0.7307	3.3040	
Y	b_1	b_2	b_3		b_1	b_2	b_3
y_1	0.0085	0.0334	-0.0052	31.3628	0.2657	1.0469	-0.1617
y_2	0.0001	-0.0001	0.0000	6936.0822	0.9068	-0.5263	0.0315
y_3	0.0042	-0.0123	0.0519	20.0891	0.0838	-0.2465	1.0423
norm b					0.9486	1.1974	
Canonical correlation	0.8248	0.3653	0		0.8248	0.3653	0

multiplying each component of a CCA vector in the original units by the corresponding SRSS terms produces parameters of the standardized variables in comparable units. Essentially, those are the same solutions, in the original and in the standardized variables, so we can consider the CCA in the correlation matrices (8), which solution is more convenient for comparison of the vector parameters and finding more influential of them by their bigger absolute value.

There are two variables in the first matrix X , so two CCA vectors a with two elements, while for the second matrix Y with three variables there are three vectors b , each with three elements. Each pair of vectors a and b defines aggregated scores (1) and the corresponding canonical correlation (2), shown at the bottom line, below each column of vectors. The rank of the matrix used in the eigenproblem (8) equals 2 because it is the minimum size in the matrices entering to that combination. Therefore, there are two positive eigenvalues and one of zero value. The maximum correlation (square root of the eigenvalue λ^2 in (8), where the squared canonical correlations are used) equals 0.8248, and the next one equals 0.3653. Below the vectors a and the vectors b in the standardized units, there are norms of the vectors (square root of the total of vector's element squared).

Table 2 presents several solutions. The first solution is shown in two columns of the CCA (8) given by the vectors divided by their norms, so a new norm equals one, as is shown under each vector in this table. The third vector b is not useful for canonical correlation analysis because it is related to zero correlation (as in Table 1), so it is omitted in Table 2. The canonical correlations shown in the last row of Table 2 are the same as in Table 1 because they do not depend on the vectors' normalization.

Table 2: CCA, RCA, and ORCA for financial data-1.

variables	CCA (8), normalized		RCA, or canonical covariance (11)		ORCA, or SVD (39)	
X	a_1	a_2	a_1	a_2	$alpha_1$	$alpha_2$
x_1	-0.7988	-0.7023	-0.7216	-0.6923	-0.7286	-0.6850
x_2	0.6016	-0.7119	0.6923	-0.7216	0.6850	-0.7286
norm a	1	1	1	1	1	1
Y	b_1	b_2	b_1	b_2	$beta_1$	$beta_2$
y_1	0.2801	0.8743	0.4485	0.8805	0.3807	0.9246
y_2	0.9559	-0.4396	0.8928	-0.4495	0.9230	-0.3792
y_3	0.0883	-0.2058	0.0426	0.1502	0.0568	-0.0359
norm b	1	1	1	1	1	1
Canonical correlation	0.8248	0.3653	0.8147	0.3382	0.8248	0.3653

The next two columns in Table 2 present the robust canonical correlations, RCA, obtained by the objective of canonical covariance (11). Taking its eigenvectors for the aggregated scores (1) and calculating with them the correlations yield the values given under these vectors in the last row of Table 2. We see that these correlations are slightly smaller than those of CCA solution, but the eigenvectors are more robust, as it was discussed for the solution (11). The last two columns in Table 2 show the ORCA solution (39). It comprises the best features from the CCA and RCA – on one hand, it has the same maximum possible correlations of CCA, and on the other hand, it is built on the orthonormal approximation of the data matrices, which do not contain the matrices inversion, thus, are robust to the detrimental effects of possible multicollinearity in the datasets.

It is useful to note that the solutions by each vector are very similar across all three discussed methods of CCA, RCA, and ORCA in Table 2. By all solutions for the first pair of the vectors, the predictors x_1 and x_2 enters into the vector a with opposite signs and are connected mostly with the outcome y_2 in the vector b . In the second pair of the vectors, the predictors x_1 and x_2 enters into the vector a with the same sign and are connected mostly with the outcome y_1 of the opposite sign in the vector b . Note that although the eigenvectors

are defined up to an arbitrary constant term, but if to change a sign of one vector (for instance, in the vector a) then the sign should be changed in the paired vector (vector b) as well.

Consider another example from a real marketing research project on a specific pharmaceutical product. The opinions of 420 respondents were measured in the Likert 5-point scale by the following characteristics. The first data set X consists of the predictors: x_1 – quick relief; x_2 – safe to use; x_3 – safe with a specific disease; x_4 – few side effects; x_5 – long lasting relief; x_6 – effective for children; x_7 – provides the most symptoms relief; x_8 – provides highest patient satisfaction; x_9 – provides highest quality of life. The second data set Y contains the outcome variables: y_1 – low cost; y_2 – provide most value; y_3 – patient satisfaction; y_4 – likelihood of prescription.

Table 3 presents the CCA results similarly to Table 1: original solution for four vectors built by the covariance matrices, the column of SRSS, and CCA by the correlation matrices. There are norms of the vectors below each of a and the b in the standardized units. The canonical correlations (2) of the predictors’ aggregates (1) are given in the last line of Table 3, where the maximum correlation reaches 0.692, and the others are noticeably smaller. It is interesting to mention that the pair correlations between the original variables of two datasets are mostly at the level 0.1-0.2, and all of them are below the level 0.5. Therefore, the first pair of the canonical variables presents the best combinations of the variables of each dataset. Elements of the vectors make a sense similar to the loadings in the factor analysis, so they correspond to the most influential variables in the connection between two data sets.

Table 3: CCA for a pharmaceutical product data-2.

variables	CCA by covariance matrices				SRSS	CCA by correlation matrices			
X	a1	a2	a3	a4		a1	a2	a3	a4
x1	0.001	-0.021	-0.028	0.026	14.856	0.016	-0.314	-0.420	0.384
x2	0.010	-0.011	-0.025	0.000	16.898	0.169	-0.179	-0.414	0.008
x3	0.002	0.009	-0.030	-0.006	17.632	0.032	0.150	-0.530	-0.114
x4	0.012	0.034	-0.012	-0.006	16.802	0.194	0.568	-0.209	-0.095
x5	0.011	-0.017	0.027	-0.056	14.427	0.159	-0.239	0.397	-0.810
x6	0.005	0.018	-0.004	0.010	30.149	0.137	0.547	-0.131	0.303
x7	0.014	-0.031	0.005	0.044	15.093	0.205	-0.475	0.081	0.658
x8	0.022	-0.012	-0.001	-0.037	15.437	0.344	-0.192	-0.019	-0.564
x9	0.018	0.029	0.046	0.028	15.173	0.279	0.434	0.705	0.418
norm a						0.592	1.133	1.162	1.357
Y	b1	b2	b3	b4		b1	b2	b3	b4
y1	0.004	0.022	-0.040	0.025	20.224	0.078	0.446	-0.803	0.501
y2	0.032	0.012	0.053	0.030	16.977	0.548	0.208	0.896	0.511
y3	0.024	-0.069	-0.034	0.006	13.259	0.319	-0.917	-0.449	0.079
y4	0.023	0.017	-0.016	-0.048	18.909	0.429	0.322	-0.311	-0.901
norm b						0.770	1.090	1.321	1.153
Canonical correlation	0.692	0.248	0.206	0.097		0.692	0.248	0.206	0.097

The next Table 4 is organized similarly to Table 2. It presents three solutions, each shown in four columns of the vectors normalized by one as indicated under each vector: the CCA (8); the robust canonical correlations RCA (11); and the ORCA orthonormal results of (39). Again, all these solutions are very similar across three discussed methods. The canonical correlations given the last row of Table 4 are the same for CCA and ORCA and they equal those in Table 3. For the RCA, using its eigenvectors for the aggregated scores (1) and calculating with them the correlations (2) yield slightly lesser correlations than those of CCA, but the RCA eigenvectors are more robust.

Table 4: CCA, RCA, and ORCA for a pharmaceutical product data-2.

X	CCA (8), normalized				RCA, or canonical covariance (11)				ORCA, or SVD (39)			
	a1 norm	a2 norm	a3 norm	a4 norm	a1 norm	a2 norm	a3 norm	a4 norm	alpha 1	alpha 2	alpha 3	alpha 4
x1	0.027	-0.277	-0.361	0.283	0.292	-0.311	-0.309	0.257	0.208	-0.338	-0.293	0.294
x2	0.286	-0.158	-0.356	0.006	0.330	0.114	-0.380	-0.055	0.319	-0.071	-0.400	-0.007
x3	0.054	0.132	-0.456	-0.084	0.241	0.115	-0.530	-0.209	0.178	0.064	-0.526	-0.133
x4	0.328	0.501	-0.179	-0.070	0.287	0.590	0.027	-0.160	0.310	0.538	-0.178	-0.126
x5	0.269	-0.211	0.341	-0.596	0.365	-0.221	0.347	-0.496	0.344	-0.199	0.321	-0.551
x6	0.232	0.483	-0.113	0.223	0.214	0.513	-0.054	0.396	0.220	0.490	-0.141	0.291
x7	0.347	-0.419	0.070	0.485	0.368	-0.415	-0.046	0.497	0.365	-0.439	0.071	0.518
x8	0.581	-0.170	-0.016	-0.415	0.424	-0.170	-0.019	-0.373	0.477	-0.189	-0.042	-0.383
x9	0.472	0.383	0.606	0.308	0.413	0.125	0.594	0.271	0.447	0.282	0.563	0.276
norm a	1	1	1	1	1	1	1	1	1	1	1	1
Y	b1 norm	b2 norm	b3 norm	b4 norm	b1 norm	b2 norm	b3 norm	b4 norm	beta 1	beta 2	beta 3	beta 4
y1	0.101	0.409	-0.608	0.434	0.256	0.599	-0.624	0.431	0.192	0.457	-0.693	0.523
y2	0.712	0.191	0.678	0.443	0.638	0.039	0.617	0.460	0.673	0.169	0.599	0.400
y3	0.415	-0.842	-0.340	0.068	0.481	-0.733	-0.480	0.038	0.447	-0.821	-0.340	0.103
y4	0.558	0.295	-0.235	-0.781	0.544	0.320	-0.005	-0.776	0.558	0.298	-0.212	-0.745
norm b	1	1	1	1	1	1	1	1	1	1	1	1
Canon. Correl.	0.692	0.248	0.206	0.097	0.678	0.237	0.199	0.093	0.692	0.248	0.206	0.097

For the sake of a convenient comparison, the signs in several paired vectors a and b in RCA, and $alpha$ and $beta$ in ORCA, were changed to adjust them to the basic CCA vectors. The first vectors' solution of the maximum canonic correlation suggests the main loadings for the variables' aggregation in two datasets. As we see by the vector elements, the main role in the first group is played by the variables x_8 and x_9 , and in the second group the main variables are y_2 and y_4 . The described closeness across the CCA, RCA, and ORCA results was observed by different studied datasets as well.

7 Summary

The paper considered canonical correlation analysis, or CCA technique, together with the earlier developed robust canonical analysis, or RCA, and the newly introduced orthonormal canonical analysis, ORCA, which as the other problems can also be presented as a specific eigenproblem. The ORCA comprises the best features of the CCA and RCA – on one hand, it has the same maximum possible correlations of CCA, and on the other hand, it is built on the orthonormal approximation of the data matrices, which do not contain the matrices inversion, thus, are robust to the detrimental effects of possible multicollinearity in the datasets. It is also shown that the CCA formulation with two matrices inverted in the eigenproblem of each vector gives an impression that both vectors can be prone to multicollinearity in any original group of variables. However, the results of ORCA derivation prove that each vector depends on possible ill-conditioning only in the matrix of data which it aggregates. Thus, the new solutions can be used for finding a better and adequate interpretation of the variables' integration into their aggregates. The relations with multiple regression modeling and description of the variance inflation factor and its adjustment to CCA problems are also given.

For future studies, the orthonormal canonical correlations approach can be extended to the problems with many data sets. Also, more research is needed for improving measures of multicollinearity in the CCA problems and their modifications. The numerical results demonstrate a close similarity of the considered techniques, so the new approach can be successfully applied to the problems of aggregation in complex data. This type of modern multivariate analysis can serve in solving various practical problems and can be helpful to managerial estimations and decision making.

Acknowledgement: I am grateful to three reviewers for their comments and suggestions which helped to improve the paper.

Conflict of interest: The author states no conflict of interest.

References

- Adrover, J., & Donato, S. (2015), A robust predictive approach for canonical correlation analysis, *Journal of Multivariate Analysis*, 133, 356–376.
- Ahituv, N., Lipovetsky, S., & Tishler, A. (1998) The relationship between firm's information systems policy and business performance: a multivariate approach. In: Mahmood, M.A., & Szewizak, E.J., editors. *Measuring information technology investment payoff: contemporary approach*, 62-82. Idea Group, Hershey, PA.
- Andrew, G., Arora, R., Bilmes, J., & Livescu, K. (2013) Deep canonical correlation analysis, in *International Conference on Machine Learning (ICML)*, 1247–1255.
- Belsley, D.A., Kuh, E., & Welsch, R.E. (1980) *Regression Diagnostics*. Wiley, New York.
- Bookstein, F.L., Sampson, P.D., Streissguth, A.P., & Barr, H.M. (1990) Measuring Dose and Response with Multivariate Data Using Partial Least Squares Techniques, *Communication in Statistics: Theory and Methods*, 19(3), 765-804.
- Carmeli, A., & Tishler, A. (2004) The relationship between intangible organizational elements and organizational performance, *Strategic Management Journal*, 25, 1257-1278.
- Cao, D.-S., Liu, S., Zeng, W.-B., & Liang, Y.-Z. (2015) Sparse canonical correlation analysis applied to -omics studies for integrative analysis and biomarker discovery. *J. Chemometrics*, 29, 371–378
- Demidenko, E. (2020) *Advanced Statistics with Applications in R*, Wiley, Hoboken, NJ.
- Dillon, W.R., & Goldstein, M. (1984) *Multivariate Analysis, Methods and Applications*. Wiley, New York.
- Dvir, D., Lipovetsky, S., Shenhar, A., & Tishler, A. (2003) What is really important for project success: A refined, multivariate, comprehensive analysis, *International Journal of Management and Decision Making*, 4, 382-404.
- Fornell, C., editor (1982) *A Second Generation of Multivariate Analysis; vol. 1 – methods, vol. 2 – measurement and evaluation*. Praeger, New York.
- Gentle, J.E. (2017) *Matrix Algebra*, Springer, Cham, Switzerland.
- Gibson, W.A. (1962) Orthogonal predictors: A possible resolution of the Hoffman-Ward controversy, *Psychological Reports*, 11, 32-34.
- Hair Jr., J.F., Anderson, R.E., Tatham, R.L., & Black, W.C. (1998). *Multivariate Data Analysis*, 5th edition, Prentice Hall, Inc. Upper Saddle River, NJ.
- Hardle, W., & Hlavka, Z. (2007) *Multivariate Statistics: Exercises and Solutions*, Springer, New York.
- Hardle, W., & Simar, L. (2012) *Applied Multivariate Statistical Analysis*, Springer, New York.
- Hardoon, D., Szedmak, S., & Shawe-Taylor, J. (2004) Canonical correlation analysis: An overview with application to learning methods, *Neural Computation*, 16, 12, 2639–64.
- Horn, R.A., & Johnson, C.R. (2013) *Matrix Analysis*, Cambridge University Press, New York.
- Horst P. (1961) Relations among m sets of measures. *Psychometrika* 26, 129-149.
- Hotelling, H. (1936) Relations between two sets of variates, *Biometrika*, 28, 3–4, 321–377.
- Irizarry, R.A. (2020) *Introduction to Data Science*, CRC Press, Boca Raton, FL.
- Izenman, A.J. (2008) *Modern Multivariate Statistical Techniques*, Springer, New York.
- Jendoubi, T., & Strimmer K. (2019) A whitening approach to probabilistic canonical correlation analysis for omics data integration, *BMC Bioinformatics*, 20, 15.
- Johnson, R.M. (1963) On the theorem stated by Eckart and Young, *Psychometrika*, 28, 259-263.
- Johnson, R.M. (1966) The minimal transformation to orthonormality, *Psychometrika*, 31, 61-66.
- Kessy, A., Lewin, A., & Strimmer, K. (2018) Optimal whitening and decorrelation. *The American Statistician*, 72, 309–314.
- Lipovetsky, S. (2009) PCA and SVD with Nonnegative Loadings, *Pattern Recognition*, 42, 68-76.
- Lipovetsky, S. (2012) Dual PLS Analysis, *International Journal of Information Technology and Decision Making*, 11, 879-891.
- Lipovetsky, S. (2016) Generalized Singular Value Decomposition with Additive Components, *Journal of Modern Applied Statistical Methods*, 15(1), 584-599.
- Lipovetsky, S., & Conklin, M. (2005a) Singular Value Decomposition in Additive, Multiplicative, and Logistic Forms, *Pattern Recognition*, 38, 1099-1110.
- Lipovetsky, S., & Conklin, M. (2005b) Incremental Net Effects in Multiple Regression, *International J. of Mathematical Education in Science and Technology*, 36, 4, 361–373.
- Lipovetsky, S., & Conklin, M. (2015) Predictor Relative Importance and Matching Regression Parameters, *Journal of Applied Statistics*, 42(5), 1017-1031.

- Lipovetsky, S., & Tishler, A. (1994) Linear Methods in Multimode Data Analysis for Decision Making, *Computers & Operations Research*, 21, 169-183.
- Lipovetsky, S., Tishler, A., and Conklin, M. (2002) Multivariate Least Squares and its Relation to Other Multivariate Techniques, *Applied Stochastic Models in Business and Industry*, 18, 347-356.
- Mair, P. (2018) *Modern Psychometrics with R*, Springer, Cham, Switzerland.
- Shenhar, A.J., Tishler, A., Dvir, D., Lipovetsky, S., & Lechler, T. (2002) Refining the Search for Project Success Factors: A Multivariate, Typological Approach, *R&D Management*, 32, 111-126.
- Tam, A., (2003) *Robust Canonical Analysis (RCA): Identifying the Most Important Variables in a Two-Set Multivariate Problem*. Ph.D. Dissertation, Tel Aviv University.
- Tishler, A., Dvir, D., Shenhar, A., & Lipovetsky, S. (1996) Identifying Critical Factors of Defense Development Projects: A Multivariate Analysis, *Technological Forecasting and Social Change*, 51, 151-171.
- Tishler, A., & Lipovetsky, S. (1996) Canonical Correlation Analyses for Three Data Sets: A Unified Framework with Application to Management, *Computers & Operations Research*, 23, 667-679.
- Tishler, A., & Lipovetsky, S. (2000) Modeling and Forecasting with Robust Canonical Analysis: Method and Application, *Computers and Operations Research*, 27, 217-232.
- Tucker, L.R. (1958) Inter-Battery Method of Factor Analysis, *Psychometrika*, 23, 111-136.
- Varmuza, K., & Filzmoser, P. (2009) *Introduction to Multivariate Statistical Analysis in Chemometrics*, CRC Press, Boca Raton, FL.
- Wilms, I., & Croux, C. (2015) Sparse canonical correlation analysis from a predictive point of view. *Biometrics Journal*, 57, 834–851.
- Wold, H. (1975) Path Models with Latent Variables: The NIPALS Approach. In Blalock H.M. et al. (eds), *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling*, 307-357, Academic Press, New York.