Restricted Canonical Correlations

Shubhabrata Das

Department of Mathematical Sciences
University of Montana
Missoula, Montana 59812

and

Pranab Kumar Sen

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina 27599

Submitted by George P. H. Styan

ABSTRACT

Given two random vectors $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ the first canonical correlation between them is defined as: sup{ Correlation$(\alpha'\mathbf{Y}^{(1)}, \beta'\mathbf{Y}^{(2)}) : \alpha \in \Re_p, \beta \in \Re_q\}$. However, in many practical situations (e.g. educational testing problems, neural networks), some natural restrictions on the coefficients α and β may arise which should be incorporated in this maximization procedure. The maximum correlation subject to such constraints is referred to as the restricted canonical correlation. This problem is treated here under the nonnegativity restriction on α and β. The analysis is extended to more general form of inequality constraints, and also when the restrictions are present only on some of the coefficients. Restricted versions of some other related measures are also discussed. This includes principal component analysis and different modifications of canonical correlations. Anomalies with higher order correlation are also described. Some properties of restricted canonical correlation, including its bounds, are studied.

---

1. INTRODUCTION

The usual concept of correlation originated in the work of Sir Francis Galton [5]. It was formulated mathematically by Pearson [8, 9]. Since then the concept of correlation has been generalized to include part, partial, and bipartial correlations. All these correlations are measures of dependence between two random variables.

655 Avenue of the Americas, New York, NY 10010 0024-3795/94/$7.00

As opposed to these three forms of correlation, multiple correlation measures dependence between a random variable, termed the *dependent variable*, and a random vector termed *independent variables*. This naturally extends to a study of dependence between two sets of random variables through *canonical correlation* (CC). Hotelling [6, 7] defined the CC as the maximum correlation between any two *representative* random variables of the two sets, where a *representative* of a set of random variables is simply any linear combination of its members. The idea of linear combination originated from the linearity of regression in multivariate normal distributions.

However, in many practical situations, there may be some natural restrictions on these linear combinations. In such a case, the traditional CC may overemphasize the dependence between the two sets, and it would be more appropriate to consider the maximization problem with the *representatives* restricted accordingly. Such a measure of dependence can be termed the *restricted canonical correlation* (RCC). The primary focus of this study is on the algebraic solutions of the RCC.

In Section 1.1, some of the needs of studying the RCC are discussed briefly. Various restrictions are of practical interest, and it seems unlikely that a unifying treatment of all such cases is feasible. Some of these different restrictions are described in Section 1.2; but explicit solutions have been obtained only for the restriction that the contributing coefficients of the original variables in their *representative* random variables be nonnegative. This is partly because one would expect the *representative* to be a convex linear combination of the original random variables, and since correlation is scale-invariant, this requires only nonnegativity of the coefficients. More importantly, this is a case where an exact analytic solution exists. Also, many other interesting cases may be reduced to this one using suitable transformations; this is shown in Section 1.3. In general, this class of restrictions is referred to as the *inequality* type. Finally, the notion of minimum RCC is introduced in the Section 1.4.

The calculation of the RCC is discussed in Section 2.1. In connection with its analytic solution, a relation between the RCC and the CCs of different subvectors is obtained in Section 2.2. In Section 2.3, several related topics are discussed briefly. The first of these is the partially restricted canonical correlation, which is the maximum correlation between linear combinations of two sets of random variables when only some of the coefficients are restricted while the others are not. Next, two approaches to restricted versions of higher order CCs are introduced. Canonical correlation analysis (CCA) is not the only area in multivariate analysis where these kinds of restrictions may be incorporated. The use of CC is particularly attractive because of its broad generality. However, restricted versions of some other multivariate methods may also be useful. A brief discussion of restricted principal component versions of part and bipartial canonical correlations are included at the end of this subsection. In Section 3, some properties of RCC are discussed. The last section consists of some general remarks on these problems.

1.1. *Whither RCC?*

If $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are p- and q-variate random vectors, then the first CC between them is defined as

$$\max_{\alpha,\beta} \text{Correlation}(\alpha'\mathbf{Y}^{(1)}, \beta'\mathbf{Y}^{(2)}).$$

If $\begin{bmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \end{bmatrix}$ has covariance matrix $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, then

$$\text{CC} = \max\{\alpha'\Sigma_{12}\beta : \alpha'\Sigma_{11}\alpha = 1 = \beta'\Sigma_{22}\beta,\ \alpha \in \Re_p,\ \beta \in \Re_q\},$$

where $\Re_p$ denotes the p-dimensional Euclidean space and $\alpha = ((\alpha_i))$ and $\beta = ((\beta_j))$ relate to the respective coefficients or weights of the original variables in their *representative* linear combinations, $\alpha'\mathbf{Y}^{(1)}$ and $\beta'\mathbf{Y}^{(2)}$. In this section, it will be explained why in many practical situations it may be necessary to incorporate some restrictions on these coefficients. The problem of restricted canonical correlation was encountered first by the present authors [3] in studying simultaneous spike trains in neurophysiology. In the following, an attempt will be made to provide further motivation of RCC through some scholastic studies.

Suppose from the scores of several students in different tests and homework assignments in two different subjects, one is interested in studying the relationship (correlation) of students' performance in these two subjects. A simple approach is to give "reasonable" weights for the different examinations and assignments to compute the composite scores in each subject and then compute the correlation between the composite scores. In contrast to having such predetermined weights, a CC approach would look for those weights which maximize the correlation between the composite scores. RCC fits somewhere in the middle. It also maximizes the correlation between possible composite scores; but instead of these weights being arbitrary as in the CC case, one may force them to be more reflective of their individual importance. For example, negative weights may be unrealistic and difficult to interpret. So the simple nonnegativity restriction on the weights seems rational. Another potential set of restrictions may be to impose an ordering of the weights. For instance, if there are three quizzes, a composite homework assignment score, two midterms, and a final in a subject, one may wish to impose the restriction

$$0 \le \alpha_{\text{Q1}} = \alpha_{\text{Q2}} = \alpha_{\text{Q3}} \le \alpha_{\text{H}} \le \alpha_{\text{MT1}} = \alpha_{\text{MT2}} \le \alpha_{\text{F}}, \tag{1}$$

where $\alpha_{\text{Q}i}$, α_{H}, $\alpha_{\text{MT}j}$, and α_{F} represent the weights of the ith quiz, the homework assignment, the jth midterm, and the final respectively. Or, if the course has only several tests, each of cumulative nature, then a reasonable restriction will be

$$0 \leq \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_p. \tag{2}$$

Both these restrictions can be reduced to the simple nonnegativity type, as shown later on.

1.2. *Different Types of Useful Restrictions*

Some of the useful restrictions arising in CCA, as introduced earlier, are :

(i) $\alpha_i \geq 0,\ \beta_j \geq 0,\ i = 1, \ldots, p,\ j = 1, \ldots, q$ (nonnegativity restriction).

(ii) $0 \leq \alpha_1 \leq \cdots \leq \alpha_p,\ 0 \leq \beta_1 \leq \cdots \leq \beta_q$ (monotone restriction).

(iii) $0 \leq \alpha_1 = \cdots = \alpha_{p_1} \leq \alpha_{p_1+1} = \cdots = \alpha_{p_1+p_2} \leq \cdots \leq \alpha_{p_1+\cdots+p_{g-1}+1} = \cdots = \alpha_{p_1+\cdots+p_g},\ 0 \leq \beta_1 = \cdots = \beta_{q_1} \leq \beta_{q_1+1} = \cdots = \beta_{q_1+q_2} \leq \cdots \leq \beta_{q_1+\cdots+q_{h-1}+1} = \cdots = \beta_{q_1+\cdots+q_h}$. (This can be called the *monotone layer restriction*, because coefficients corresponding to variables which are in successive layers, are monotone, while layers account for ties.)

(iv) In some situations, one may have partial restriction in the sense that additional restrictions are present only on some of the coefficients. For example, in the case of monotone restriction, if the smallest coefficients (α_1 and β_1) are not restricted to be nonnegative, this would come in handy after using the transformation which is described in the sequel. The term *partially restricted canonical correlation* may be used to denote the maximal correlation between two linear combinations of original variables, under a set of restrictions on some of the coefficients while the remaining ones may vary freely. The difference between this and the *restricted partial canonical correlation* may be noted here: the latter is the usual RCC obtained from the residual covariance matrix of two sets of variables after the linear effect of the third set has been removed. The calculation of PRCC is discussed also in Section 2.

(v) Professor Ley of INRA–Laboratoire de Biometrie, France, has encountered problems in genetic population dynamics where the first few coefficients in two groups should be the same, i.e.,

$$\alpha_r = \beta_r \quad \text{for } 1 \leq r \leq p_0, \qquad \text{where} \quad p_0 < \min(p, q).$$

It is conceivable that there can be other types of restrictions. For example:

(vi) Some coefficients are preassigned and are not flexible, i.e.,

$$\alpha_r = \alpha_r^0 \quad \text{for } 1 \leq r \leq p_1 < p, \qquad \beta_r = \beta_r^0 \quad \text{for } 1 \leq r \leq q_1 < q,$$

where α_r^0's and β_r^0's are fixed numbers.

(vii) The ratios of the first few coefficients are equal in the two groups, i.e.,

$$\alpha_1 : \alpha_2 : \cdots : \alpha_{p_0} = \beta_1 : \beta_2 : \cdots : \beta_{p_0},$$

where $p_0 < \min(p, q)$. A similar type of restriction is

$$\alpha_i : \beta_i = \text{constant} \qquad \text{for} \quad 1 \le r \le p_0.$$

1.3. Restrictions Which Can Be Reduced to Nonnegativity Type

Restrictions (ii) and (iii), can be reduced to the problem with *nonnegativity restrictions* (NNR) by using a transformation, which is discussed here.

First consider the monotone restriction (ii). Suppose

$$\rho^* = \sup\{\alpha'\Sigma_{12}\beta : \alpha'\Sigma_{11}\alpha = 1 = \beta'\Sigma_{22}\beta;\ 0 \le \alpha_1 \le \cdots \le \alpha_p,\ 0 \le \beta_1 \le \cdots \le \beta_q\}.$$

Define

$$\begin{aligned} \tilde{\alpha}_1 &= \alpha_1, & \tilde{\beta}_1 &= \beta_1,\\ \tilde{\alpha}_2 &= \alpha_2 - \alpha_1, & \tilde{\beta}_2 &= \beta_2 - \beta_1,\\ &\vdots & &\vdots\\ \tilde{\alpha}_p &= \alpha_p - \alpha_{p-1}, & \tilde{\beta}_q &= \beta_q - \beta_{q-1} \end{aligned}$$

or, equivalently,

$$\begin{aligned} \alpha_1 &= \tilde{\alpha}_1, & \beta_1 &= \tilde{\beta}_1,\\ \alpha_2 &= \tilde{\alpha}_2 + \tilde{\alpha}_1, & \beta_2 &= \tilde{\beta}_2 + \tilde{\beta}_1,\\ &\vdots & &\vdots\\ \alpha_p &= \tilde{\alpha}_p + \cdots + \tilde{\alpha}_1, & \beta_q &= \tilde{\beta}_q + \cdots + \tilde{\beta}_1. \end{aligned}$$

So $\alpha = \mathbf{L}_p\tilde{\alpha}$ and $\beta = \mathbf{L}_q\tilde{\beta}$, where $\mathbf{L}_r$ is a $r \times r$ lower triangular matrix with (i, j)th entries 1 for $i \ge j$. Thus

$$\rho^* = \sup\{\tilde{\alpha}'\tilde{\Sigma}_{12}\tilde{\beta} : \tilde{\alpha}'\tilde{\Sigma}_{11}\tilde{\alpha} = 1 = \tilde{\beta}'\tilde{\Sigma}_{22}\tilde{\beta};\ \tilde{\alpha}_i \ge 0;\ \tilde{\beta}_j \ge 0\},$$

where $\tilde{\Sigma}_{12} = \mathbf{L}_p'\Sigma_{12}\mathbf{L}_q$, $\tilde{\Sigma}_{11} = \mathbf{L}_p'\Sigma_{11}\mathbf{L}_p$, $\tilde{\Sigma}_{22} = \mathbf{L}_q'\Sigma_{22}\mathbf{L}_q$. This reduces the problem to one with NNR. Note that the (i, j)th element of $\tilde{\Sigma}_{kl}$ is nothing but the sum of all elements of Σ_{kl} except for those belonging to first $i - 1$ rows and $j - 1$ columns, i.e.,

$$\tilde{\sigma}_{kl}(i, j) = \sum_{m=i}^{p}\sum_{n=j}^{q} \sigma_{kl}(m, n), \qquad k, l = 1, 2.$$

Next consider the monotone layer restriction (iii), which is in fact a more general form of (ii). The transformation used in (ii) reduces this problem also to the one with nonnegativity restriction. Now one needs to work with the $g \times h$ covariance matrix

$$\check{\Sigma} = \begin{pmatrix} \check{\Sigma}_{11} & \check{\Sigma}_{12} \\ \check{\Sigma}_{21} & \check{\Sigma}_{11} \end{pmatrix},$$

where the (i, j)th element of $\check{\Sigma}_{kl}$ is

$$\check{\sigma}_{kl}(i, j) = \tilde{\sigma}_{kl}\left(1 + \sum_{t=1}^{i-1} p_t, 1 + \sum_{t=1}^{j-1} q_t\right).$$

1.4. Minimum Restricted Canonical Correlation

The traditional CC is always nonnegative. This is because if Correlation $(\alpha Y, \beta Z) = \rho < 0$ then Correlation$(-\alpha Y, \beta Z) = -\rho > 0$. However, since $-\alpha$ may not satisfy the restriction in a given problem, this is not true for RCC. In fact, in a restricted study, it may be sensible to consider the corresponding minimization problem rather than the maximization one. Consider, for example, a quasiintraclass correlation model where the covariance matrices are given by

$$\Sigma_{12} = \rho \mathbb{J}_{p \times q}, \qquad \Sigma_{11} = \rho_1 \mathbb{J}_{p \times p} + (1-\rho_1)\mathbb{I}_p, \quad \Sigma_{22} = \rho_2 \mathbb{J}_{q \times q} + (1-\rho_2)\mathbb{I}_q,$$

where $\mathbb{I}_p$ denotes the $p \times p$ identity matrix and $\mathbb{J}_{p \times q}$, the $p \times q$ matrix consisting of all ones. Let $\rho_1, \rho_2 \geq 0$, and $\rho < 0$. Then in restricted CCA with α, β nonnegative, the minimum RCC and the usual CC are related functionally.

More generally, such a situation arises if all elements of Σ_{12} are negative. However, the theory and calculation for finding the minimum RCC (the solution to the minimization problem), as well as its properties (sampling distribution), are very similar to those for the maximum RCC. Hence for most part of this work, only the maximum RCC will be considered.

2. EVALUATION OF RCC UNDER NNR, AND RELATED TOPICS

2.1. Incorporation of the KTL-Point Formula

In this section, the goal is to discuss the algebraic solution for the RCC when the contributing coordinates are restricted to be nonnegative. Note that

$$\text{RCC} = \sup\{\alpha' \Sigma_{12} \beta : \alpha' \Sigma_{11} \alpha = 1 = \beta' \Sigma_{22} \beta,\ \alpha \in \Re_p^+,\ \beta \in \Re_p^+\}. \tag{3}$$

where $\alpha = (\alpha_1, \ldots, \alpha_p)'$, $\beta = (\beta_1, \ldots, \beta_q)'$, and $\Re_t^+ = \{\mathbf{x} = ((x_i)) \in \Re_t : x_i \geq 0 \quad \forall i = 1, \ldots, t\}$. Equation (3) can also be written as

$$\text{RCC} = \sup \left\{ h(\mathbf{x}) = \frac{\mathbf{x}'\Sigma^*\mathbf{x}}{\sqrt{\mathbf{x}'\Sigma_{11}^*\mathbf{x}\mathbf{x}'\Sigma_{22}^*\mathbf{x}}} : x_i \geq 0,\ i = 1, \ldots, l \right\}, \tag{4}$$

where

$$\mathbf{x} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \qquad l = p + q,$$

$$\Sigma^* = \begin{bmatrix} \mathbf{0} & \frac{1}{2}\Sigma_{12} \\ \frac{1}{2}\Sigma_{21} & \mathbf{0} \end{bmatrix}, \qquad \Sigma_{11}^* = \begin{bmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \qquad \Sigma_{22}^* = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{bmatrix}.$$

A related problem is

$$\sup\{h(\mathbf{x}) : \mathbf{x} \in {}_0\Re_l^+\}, \tag{5}$$

where ${}_0\Re_l^+ = \{\mathbf{x} \in \Re_l : x_i > 0\}$. This corresponds to the RCC problem with strict positivity restriction.

LEMMA 1. *A maximum in the problem* (4) *always exists.*

Proof. Let $y^0 = \sup_{\mathbf{x} \in \Re_l^+} h(\mathbf{x})$. Since $h(\cdot)$ in scale-invariant, one also has $y^0 = \sup_{\mathbf{x} \in \mathcal{U}_l^+} h(\mathbf{x})$, where $\mathcal{U}_l^+ = \{\mathbf{x} \in \Re_l : 0 \leq \mathbf{x}_i \leq 1\}$. So $\exists\{\mathbf{x}_n\} \in \mathcal{U}_l^+$ such that $h(\mathbf{x}_n) \to y^0$. Since $\mathcal{U}_l^+$ is bounded, there exists a subsequence $\{n'\}$ and $\mathbf{x}^1 \in \mathcal{U}_l^+$ (since $\mathcal{U}_l^+$ is closed) s.t. $\mathbf{x}_{n'} \to \mathbf{x}^1$. Hence, by continuity of h, $h(\mathbf{x}^1) = \lim_{n' \to \infty} h(\mathbf{x}_{n'}) = \lim_{n \to \infty} h(\mathbf{x}_n) = y^0$. ■

LEMMA 2. *A maximum in the problem* (5) *may not exists*; *but if it does, then it is the same as the maximum in* (3).

Proof. The first part is true because ${}_0\Re_l^+$ is a open set. The second conclusion follows because ${}_0\Re_l^+$ is a dense subset of $\Re_l^+$ and h is continuous.

■

LEMMA 3. *The* RCC *or the maximum* (*supremum*) *in* (3) *must satisfy*

$$\begin{aligned} \Sigma_{12}\beta - \rho\Sigma_{11}\alpha + \lambda^{(p)} &= \mathbf{0}, \\ \Sigma_{21}\alpha - \rho\Sigma_{22}\beta + \lambda^{(q)} &= \mathbf{0}; \end{aligned} \tag{6}$$

$$\alpha_i \geq 0, \quad i = 1, \ldots, p, \qquad \beta_j \geq 0, \quad j = 1, \ldots, q; \tag{7}$$

$$\lambda_i \geq 0, \qquad i = 1, \ldots, n + 4; \tag{8}$$

$$\lambda_i\alpha_i = 0 = \lambda_{p+j}\beta_j, \qquad i = 1, \ldots, p, \quad j = 1, \ldots, q; \tag{9}$$

$$\alpha'\Sigma_{11}\alpha = 1 = \beta'\Sigma_{22}\beta, \tag{10}$$

where $\rho = \mathrm{RCC}$, $(\lambda^{(p)})' = (\lambda_1, \ldots, \lambda_p)$, *and* $(\lambda^{(q)})' = (\lambda_{p+1}, \ldots, \lambda_n)$ *are the usual slack/surplus variables.*

Proof. To solve (3), the Kuhn-Tucker Lagrangian theory is used. This states that the optimal solution of the problem

$$\text{Maximize } f(\mathbf{x}) \quad \text{subject to} \quad g_i(\mathbf{x}) \le b_i,\ i = 1, \ldots, m, \text{ and } \mathbf{x} \in \Re_n$$

must satisfy

$$\nabla f(\mathbf{x}) - \sum_1^m \lambda_i \nabla g_i(\mathbf{x}) = \mathbf{0}, \tag{11}$$

$$g_i(\mathbf{x}) \le b_i, \qquad i = 1, \ldots, m, \tag{12}$$

$$\lambda_i \ge 0, \qquad i = 1, \ldots, m, \tag{13}$$

$$\lambda_i\big[g_i(\mathbf{x}) - b_i\big] = 0, \qquad i = 1, \ldots, m. \tag{14}$$

This fits exactly into solving for (3) with

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{x}'\Sigma^*\mathbf{x}, \\ n &= l = p + q, \\ m &= l + 4, \\ g_i(\mathbf{x}) &= x_i, \qquad i = 1, \ldots, n, \\ g_{n+1}(\mathbf{x}) &= \mathbf{x}'\Sigma_{11}^*\mathbf{x} = -g_{n+2}(\mathbf{x}), \\ g_{n+3}(\mathbf{x}) &= \mathbf{x}'\Sigma_{22}^*\mathbf{x} = -g_{n+4}(\mathbf{x}), \\ b_i &= 0, \qquad i = 1, \ldots, n, \\ b_{n+1} = b_{n+3} &= 1 = -b_{n+2} = -b_{n+4}. \end{aligned}$$

Here,

$$\nabla g_i(\mathbf{x}) = (\underbrace{0, \ldots, 0}_{i-1}, -1, 0, \ldots, 0), \quad i = 1, \ldots, n; \qquad \nabla f(\mathbf{x}) = \begin{pmatrix} \Sigma_{12}\beta \\ \Sigma_{21}\alpha \end{pmatrix};$$

$$\begin{aligned} \nabla g_{n+1}(\mathbf{x}) &= \begin{pmatrix} 2\Sigma_{11}\alpha \\ \mathbf{0} \end{pmatrix} = -\nabla g_{n+2}(\mathbf{x}); \\ \nabla g_{n+3}(\mathbf{x}) &= \begin{pmatrix} \mathbf{0} \\ 2\Sigma_{22}\beta \end{pmatrix} = -\nabla g_{n+4}(\mathbf{x}), \end{aligned}$$

so that the optimum solution to (3) must satisfy

$$\Sigma_{12}\beta + 2(\lambda_{n+2} - \lambda_{n+1})\Sigma_{11}\alpha + \lambda^{(p)} = \mathbf{0}, \tag{15}$$

$$\Sigma_{21}\alpha + 2(\lambda_{n+4} - \lambda_{n+3})\Sigma_{22}\beta + \lambda^{(q)} = \mathbf{0}, \tag{16}$$

and the side restrictions (7) through (10).

Now (9) implies that $\alpha'\lambda^{(p)} = 0 = \beta'\lambda^{(q)}$. Using this and (10), one gets from premultiplying (15) and (16) by α' and β' respectively

$$2(\lambda_{n+2} - \lambda_{n+1}) = 2(\lambda_{n+4} - \lambda_{n+3}) = -\alpha'\Sigma_{12}\beta = -\rho,$$

where ρ stands for the RCC. Substituting this in (15) and (16), it follows that the optimal solution to (3) has to satisfy (6), with the additional restrictions as given. ■

LEMMA 4. *If the maximum for* (5) *exists, then it must satisfy*

$$\begin{pmatrix} -\rho\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\rho\Sigma_{12} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \mathbf{0}. \tag{17}$$

Proof. Since the maximum in (5) exists, by Lemma 2, it must satisfy (6). Further, since in this case $\alpha_i > 0$, $\beta_j > 0 \, \forall i, j$, one obtains from (9) that $\lambda_i = 0 \; \forall \; i = 1, \ldots, n$, i.e., $\lambda^{(p)} = \mathbf{0}$ and $\lambda^{(q)} = \mathbf{0}$. Hence, the maximum RCC (if it exists), under the strict positivity restriction on the coefficients, must satisfy

$$\Sigma_{12}\beta - \rho\Sigma_{11}\alpha = \mathbf{0},$$
$$\Sigma_{21}\alpha - \rho\Sigma_{22}\beta = \mathbf{0}$$

or, equivalently, (17). It should be also noted that this is the same equation as in the case of the usual CC. ■

Some notation is introduced in the following for convenience. Let p and q be fixed integers as previously.

Let $\mathbb{N}_k = \{1, 2, \ldots, k\}$ for k any positive integer.

$\mathbb{W}_k = \{\mathbf{a} : \emptyset \neq \mathbf{a} \subseteq \mathbb{N}_k$, with the elements in $\mathbf{a}$ written in natural order$\}$.

$|\mathbf{a}|$ = cardinality of $\mathbf{a}$.

For a p-component vector $\mathbf{X}$, and $\mathbf{a} \in \mathbb{W}_p$, let ${}_{\mathbf{a}}\mathbf{X}$ stand for the $|\mathbf{a}|$- component vector consisting of those components of $\mathbf{X}$ whose indices belong to $\mathbf{a}$. Similarly, for $p \times q$ matrix $\mathbf{S}$, $\mathbf{a} \in \mathbb{W}_p$, $\mathbf{b} \in \mathbb{W}_q$, let ${}_{\mathbf{a}\,:\,\mathbf{b}}\mathbf{S}$ represent the $|\mathbf{a}| \times |\mathbf{b}|$ submatrix of $\mathbf{S}$ consisting of those rows whose indices are in $\mathbf{a}$ and those columns whose indices are in $\mathbf{b}$. So if

$$\mathbf{X} = (X_1, \ldots, X_p)', \qquad \mathbf{S} = ((s_{ij})), \qquad \mathbf{a} = (i_1, \ldots, i_l),$$

and $\mathbf{b} = (j_1, \ldots, j_k)$,

then

$$_{\mathbf{a}}\mathbf{X} = \begin{pmatrix} X_{i_1} \\ \vdots \\ X_{i_l} \end{pmatrix}, \qquad _{\mathbf{a}:\mathbf{b}}\mathbf{S} = \begin{bmatrix} s_{i_1 j_1} & \cdots & s_{i_1 j_k} \\ \vdots & & \vdots \\ s_{i_l j_1} & \cdots & s_{i_l j_k} \end{bmatrix}.$$

Throughout this work, these submatrices will be referred to as "proper" submatrices, although this is more restrictive than the traditional use of the terminology. Clearly, there are $(2^p - 1) \times (2^q - 1) = \mathcal{P}\mathcal{Q}$ such "proper" submatrices of $\mathbf{S}$; where $\mathcal{P} = 2^p - 1$ and $\mathcal{Q} = 2^q - 1$.

2.2. *A Characterization of the RCC*

First note that

$$\text{RCC} = \sup_{\substack{\alpha \in \Re_p^+ \\ \beta \in \Re_q^+}} \text{Correlation}(\alpha' \mathbf{Y}^{(1)}, \beta' \mathbf{Y}^{(2)})$$

[here sup can be replaced by max]

$$= \max_{\substack{\mathbf{a} \in \mathbb{W}_p \\ \mathbf{b} \in \mathbb{W}_q}} \sup_{\substack{\alpha \in {}_0\Re^+_{|\mathbf{a}|} \\ \beta \in {}_0\Re^+_{|\mathbf{b}|}}} \text{Correlation}(\alpha'_{\mathbf{a}} \mathbf{Y}^{(1)}, \beta'_{\mathbf{b}} \mathbf{Y}^{(2)}) \tag{18}$$

$$= \max_{\substack{\mathbf{a} \in \mathbb{W}_p \\ \mathbf{b} \in \mathbb{W}_q}} \overset{*}{\max_{\substack{\alpha \in {}_0\Re^+_{|\mathbf{a}|} \\ \beta \in {}_0\Re^+_{|\mathbf{b}|}}}} \text{Correlation}(\alpha'_{\mathbf{a}} \mathbf{Y}^{(1)}, \beta'_{\mathbf{b}} \mathbf{Y}^{(2)}). \tag{19}$$

Now the supremum in (18) may not be attained for all $\mathbf{a}$ and $\mathbf{b}$ and hence cannot be replaced by a maximum in general. However, when the supremum is not attained, then necessarily there exist subvectors $\tilde{\mathbf{a}} \subseteq \mathbf{a}$, $\tilde{\mathbf{b}} \subseteq \mathbf{b}$ such that the supremum is attained for $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$, and this supremum is the same as that for $\mathbf{a}$ and $\mathbf{b}$. Thus it is all right to replace the supremum in (18) by the $\overset{*}{\max}$ in (19), which stands for the maximum when it exists, while ignoring the subgroup when it does not. When the maximum exists, then by Lemma 4 the maximal correlation ρ has to satisfy (17) with the coefficients satisfying (7) and (10), where Σ_{11}, Σ_{12}, and Σ_{22} are respectively replaced by $_{\mathbf{a}:\mathbf{a}}\Sigma_{11}$, $_{\mathbf{a}:\mathbf{b}}\Sigma_{12}$, and $_{\mathbf{b}:\mathbf{b}}\Sigma_{22}$. So for all "proper" submatrices of Σ, one needs to check whether there exist coefficients [corresponding to maximal correlation ρ, obtained from (17)] which satisfy the side restrictions (7) and (10). It may be noted here that usually (7) is the more difficult condition to satisfy. However, if ρ is the single root of the equation (17), i.e., it is an eigenvalue of multiplicity 1 (which is often the case with the sample covariance matrices), then this is easy to check. By (18), the RCC is the maximal such ρ [which has corresponding coefficients satisfying (7) and (10)].

Since a solution of (17) has to be one (first or higher) of the canonical correlations, the following lemma evolves from the above discussion.

LEMMA 5. *If the set of squared* CCs *corresponding to*

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

is represented by $\text{setcc}^2(\Sigma_{11}, \Sigma_{12}, \Sigma_{22})$, *then*

$$\text{RCC}^2(\Sigma_{11}, \Sigma_{12}, \Sigma_{22}) \in \bigcup_{\mathbf{a}\in\mathbb{W}_p,\ \mathbf{b}\in\mathbb{W}_q} \{\text{setcc}^2(\mathbf{a} : \mathbf{a}\Sigma_{11},\ \mathbf{a} : \mathbf{b}\Sigma_{12},\ \mathbf{b} : \mathbf{b}\Sigma_{22})\}.$$

Equivalently, the squared RCC *between* $\mathbf{Y}^{(1)}$ *and* $\mathbf{Y}^{(2)}$ *is equal to one of the squared CCs between* ${}_{\mathbf{a}}\mathbf{Y}^{(1)}$ *and* ${}_{\mathbf{b}}\mathbf{Y}^{(2)}$ *for some* $\mathbf{a} \in \mathbb{W}_p$, $\mathbf{b} \in \mathbb{W}_q$.

The proof follows from the above discussion. It is necessary to consider the squares of both RCC and CCs, because from (17), one can only solve for ρ^2, not ρ. This does not cause any concern for CC (since it is always nonnegative); but in the case of RCC, one needs to look at the coefficients to decide whether it should be the positive or the negative root. It is simple to characterize the situations when this added precaution is not necessary. Clearly, it is not needed when RCC is nonnegative, or equivalently, when at least one element of Σ_{12} is nonnegative.

Although Lemma 5 gives a clear guideline for finding coefficients corresponding to the RCC, the process can be made simpler by judicious successive steps in the search procedure. This is especially the case if the conjecture "RCC is the *first* CC between some subvectors" holds true (note that by Lemma 5, RCC must be *one of the* CC's between some subvectors). In such a case, a *branch and bound* principle of the following form should be implemented. The subvectors should be chosen in the following order:

Step I: $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$.

Step II: All $p - 1$-dimensional subvectors of $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$; and $\mathbf{Y}^{(1)}$ and all $q - 1$-dimensional subvectors of $\mathbf{Y}^{(2)}$.

Step III: All $p - 1$-dimensional subvectors of $\mathbf{Y}^{(1)}$ and all $q - 1$-dimensional subvectors of $\mathbf{Y}^{(2)}$.

$\vdots$

If at any step the canonical coefficients satisfy the NNR, then that leads to the RCC, and the search is stopped. If there are several of these within a step, then the one corresponding to the highest correlation gives the answer.

2.3. *Related Topics*

2.3.1. Partially Restricted Canonical Correlation. At the beginning of this section, the calculation of RCC was done *all* the canonical coefficients are restricted to be nonnegative. In this section, attention is devoted to the analysis when only *some* of the coefficients have the NNR. This is useful in the case of some other inequality restrictions as well. One example of this would be a problem with the restriction $\alpha_1 \leq \cdots \leq \alpha_p$, $\beta_1 \leq \cdots \leq \beta_q$. As opposed to the monotone restriction in Section 1, it is not required that $\beta_1 \geq 0$ or $\alpha_1 \geq 0$ here. In this case, the following will come in handy because $\tilde{\alpha}_1$ and $\tilde{\beta}_1$ (of Section 1) will be unrestricted in that case.

After appropriate renumbering, any partially restricted problem (with NNR) can be transformed to the situation where the first p_1 (≥ 0) coefficients of the first set and first q_1 (≥ 0) coefficients of the second set are restricted to be nonnegative, while the remaining coefficients are unrestricted. Let $p = p_1+p_2$ and $q = q_1+q_2$. It can be shown, following exactly the same type of calculations as in (6), that the optimal PRCC and the coefficients, under the NNR, must satisfy

$$\Sigma_{12}\beta - \rho\Sigma_{11}\alpha + \begin{pmatrix} \lambda^{(p_1)} \\ \mathbf{0} \end{pmatrix} = \mathbf{0},$$

$$\Sigma_{21}\alpha - \rho\Sigma_{22}\beta + \begin{pmatrix} \lambda^{(q_1)} \\ \mathbf{0} \end{pmatrix} = \mathbf{0}. \tag{20}$$

The calculation and argument proceed as in Sections 2.1 and 2.2; but instead of taking maximum over all "proper" subsets, now the maximum is over only those subsets which contain at least the last p_2 indices for the first set and the last q_2 indices of the second set. More formally,

$$\text{PRCC} = \max_{\substack{a\in\mathbb{W}_p^{p_1} \\ b\in\mathbb{W}_q^{q_1}}} \overset{*}{\max_{\alpha\beta}} \text{Correlation}(\alpha_a'\mathbf{Y}, \beta_b'\mathbf{Z}), \tag{21}$$

where $\mathbb{W}_k^l = \{a : \mathbb{N}_k \backslash \mathbb{N}_l \subseteq a \subseteq \mathbb{N}_k\}$, and $\overset{*}{\max}$ denotes the maximum (when it exists) over α, β satisfying

$$\alpha_i > 0 \quad \forall i \in a \cap \mathbb{N}_{p_1} \quad \text{and} \quad \beta_j > 0 \quad \forall j \in b \cap \mathbb{N}_{q_1}.$$

From this, PRCC can be calculated using the same formulae as before.

From (21) it is clear that if $p_1 = 0 = q_1$, then PRCC reduces to the usual CC; while if $p_1 = p$ and $q_1 = q \Leftrightarrow p_2 = 0 = q_2$, then one gets RCC. Thus, as one would expect, PRCC is a more general form of RCC and CC.

2.3.2. Higher Order RCC. In the traditional CCA, the higher order CCs play a very important role, especially when some of them are not negligible compared to the first CC. Naturally it is of interest to explore the extent to which the first RCC can be supplemented by its higher order analogues. Simple examples illustrate that a kth order RCC may not always exist (unlike the corresponding CC) even when $p \wedge q \geq k$, and no general statement can be made about number of higher order RCCs.

Evidently, one may or may not want to enforce the restriction that a canonical variable of a group of variables be orthogonal to the previous (lower order) canonical variables of the other group. It does not matter in the usual canonical analysis. However, there is a significant difference between the two approaches in the restricted case. The techniques described at the beginning of this section can be adopted to implement either approach. Unfortunately, they do not lead to explicit solutions under general circumstances in either of the two approaches, although simple solutions exist when some special conditions are met. The details of these are described in Das [2].

2.3.3. Restricted Principal Component. As opposed to canonical correlation, principal component analysis is an internal analysis. For example, to find the first principal component of a q-component random vector $\mathbf{X}$ we seek the normalized linear combination of components of $\mathbf{X}$ which has highest variance. That is, if Σ represents the covariance matrix of $\mathbf{X}$, then the problem is to

$$\text{maximize} \quad \beta'\Sigma\beta \quad \text{subject to} \quad \sum_{j=1}^{q} \beta_j^2 = 1. \tag{22}$$

It can be shown that the solution is the largest eigenvalue of Σ. The higher order principal components are defined as those with maximal variance subject to being orthogonal to the earlier components. It turns out that these maximal variances are also eigenvalues of Σ.

Similarly to the RCC, one can consider the restricted version of the problem (22). Next, the solution of the restricted principal component problem is discussed when the β_j's are restricted to be nonnegative. The solution is very similar to that for the corresponding RCC case. This is probably not very surprising, considering the fact that many people view canonical correlation theory as a generalization of principal component theory.

According to the Kuhn-Tucker Lagrangian theory, the solution ρ to the restricted problem must satisfy

$$\Sigma\beta - \rho\beta + \lambda = \mathbf{0}, \tag{23}$$

$$\lambda_j \beta_j = 0, \qquad j = 1, \ldots, q, \tag{24}$$

$$\sum_{j=1}^{q} \beta_j^2 = 1, \qquad \beta_j \geq 0, \quad j = 1, \ldots, q. \tag{25}$$

Thus, following the same sequence of arguments as in the RCC, the maximal variance of the principal component is the largest eigenvalue of any "proper" submatrix of Σ which has a corresponding eigenvector satisfying (25).

2.3.4. Restricted and Unrestricted Part, Partial, and Bipartial Canonical Correlations. Rao (1969) generalized the concept of the partial correlation coefficient to the partial canonical correlation between two sets of variables $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ with respect to a third set of variables $\mathbf{Y}^{(3)}$. This partial canonical correlation is defined as canonical correlation between $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$, after the effect of $\mathbf{Y}^{(3)}$ is removed from both of them. Essentially, this amounts to the change that in CCA,

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \text{is replaced by} \quad \begin{pmatrix} \Sigma_{11.3} & \Sigma_{12.3} \\ \Sigma_{21.3} & \Sigma_{22.3} \end{pmatrix}, \tag{26}$$

where

$$\Sigma_{ij.3} = \Sigma_{ij} - \Sigma_{i3}\Sigma_{33}^{-1}\Sigma_{3j}.$$

Timm and Carlson [11] extended this to part and bipartial canonical correlations. Both of these are natural extensions of (univariate) part and bipartial correlations and deal with conditional distributions. Unlike partial correlation, these are not symmetric measures between the two main variables. In the case of partial correlation, the effect of the third variable is removed from one of the principal variables, but not from the other. And in the case of bipartial correlation, the effect of the third variable is removed from one principal variable, while the effect of a fourth is removed from the other variable.

The extensions to CC versions of these are entirely natural and predictable. In the case of partial canonical correlation, there are three groups of random variables, $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$, and $\mathbf{Y}^{(3)}$, and one is interested in finding the CC between $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ after the effect of $\mathbf{Y}^{(3)}$ is removed from $\mathbf{Y}^{(1)}$. In bipartial canonical correlation, there is an additional group of variables $\mathbf{Y}^{(4)}$, whose effect is removed from $\mathbf{Y}^{(2)}$ before finding the canonical correlations. A detailed discussion of how to find these is given in Timm and Carlson (1976); but it again amounts to replacing Σ appropriately. For partial canonical correlation, the appropriate replacement is

$$\begin{pmatrix} \Sigma_{11.3} & \Sigma_{12.3} \\ \Sigma_{21.3} & \Sigma_{22} \end{pmatrix} \quad \text{in place of} \quad \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{27}$$

and for bipartial canonical correlation it is

$$\begin{pmatrix} \Sigma_{11.3} & \widehat{\Sigma}_{12} \\ \widehat{\Sigma}_{21} & \Sigma_{22.4} \end{pmatrix} \quad \text{in place of} \quad \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{28}$$

where $\widehat{\Sigma}_{12} = \Sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32} - \Sigma_{14}\Sigma_{44}^{-1}\Sigma_{42} + \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{34}\Sigma_{44}^{-1}\Sigma_{42}$, which is the covariance matrix between $\mathbf{Y}^{(1)} - \widehat{\mathbf{Y}}^{(1)}$ and $\mathbf{Y}^{(2)} - \widehat{\mathbf{Y}}^{(2)}$, where $\widehat{\mathbf{Y}}^{(1)}$ is obtained by regressing $\mathbf{Y}^{(1)}$ on $\mathbf{Y}^{(3)}$, and $\widehat{\mathbf{Y}}^{(2)}$ is obtained by regressing $\mathbf{Y}^{(2)}$ on $\mathbf{Y}^{(4)}$.

Naturally, one can consider the restricted versions of all these types of CC. This means, as before, that there may be some additional restrictions of the canonical coefficients of $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$. Since it has already been noted that the only difference (from the usual CC) in the methodology for finding these is in the adjustments of the covariance matrix, it is clear that to find the restricted versions (with NNR) of partial or bipartial CCs, one only needs to make appropriate adjustments as in (26), (27), and (28), and then follow the usual steps as in usual RCC, as described in Sections 2.1 and 2.2.

3. PROPERTIES OF RCC

When $p = q = 1$, RCC clearly reduces to the usual product-moment correlation. It may be noted here that this is not quite the case with traditional CC, since if Correlation$(X, Y) = \rho$ than CC$(X, Y) = |\rho|$. Thus, in a way RCC is a more natural generalization of correlation than CC.

3.1. *Upper and Lower Bounds*

Here, for simplicity, Σ is taken as the correlation matrix. Also, denote the largest element of Σ_{12} by $\sigma_{i_0 j_0}$.

LEMMA 6.

$$\sigma_{i_0 j_0} \leq \text{RCC} \leq \text{CC}.$$

Proof. The proof is trivial because maximum over a larger set is larger. ∎

NOTE. Of course, it is clear that the upper bounds is valid for any form of restrictions. So also is the lower bound as long as coefficients of the type $(0, \ldots, 0, 1, 0, \ldots, 0)$ are allowed in the restrictions.

Next, a brief study is done to explore when the bounds in Lemma 6 are attained. As usual, only the case with NNR is considered here. A matrix is called $\geq$ $(\leq)$ $\mathbf{0}$ if all elements of the matrix are nonnegative (nonpositive).

LEMMA 7. *If either*

(a) $\Sigma_{12} \leq \mathbf{0}$ *or*

(b) $\Sigma_{11} \geq \mathbf{0}$, $\Sigma_{22} \geq \mathbf{0}$, *and* $\sigma_{i_0 j_0}$ *is the only positive element of* Σ_{12}, *then the RCC attains the lower bound* $\sigma_{i_0 j_0}$.

Proof.

$$\text{Correlation}(\alpha' Y, \beta' Z) = \frac{\alpha' \Sigma_{12} \beta}{\sqrt{(\alpha' \Sigma_{11} \alpha)(\beta' \Sigma_{22} \beta)}} = (*)$$

and the numerator $\sum_{i=1}^{p} \sum_{j=1}^{q} \alpha_i \beta_j \sigma_{ij}^{12}$. So under (a),

$$(*) \leq \left(\frac{\sum \alpha_i}{\alpha' \Sigma_{11} \alpha} \right) \times \left(\frac{\sum \beta_j}{\beta' \Sigma_{22} \beta} \right) \times \sigma_{i_0 j_0} \qquad (\text{since } \sigma_{ij}^{12} \leq \sigma_{i_0 j_0}\ \forall i,\ j)$$
$$\leq \sigma_{i_0 j_0},$$

since $\left(\sum \alpha_i\right)^2 = \sum \sum \alpha_i \alpha_j \geq \sum \sum \alpha_i \alpha_j \sigma_{ij}^{11} = \alpha' \Sigma_{11} \alpha$, $\left(\sum \beta_j\right)^2 \geq \beta' \Sigma_{22} \beta$, and $\sigma_{i_0 j_0} \leq 0$. To prove the result under (b), it may be assumed, without loss of generality, that $\alpha_{i_0} = 1 = \beta_{j_0}$. Then

$$(*) \leq \frac{\sigma_{i_0 j_0}}{\sqrt{(\alpha' \Sigma_{11} \alpha)(\beta' \Sigma_{22} \beta)}} \qquad [\text{since } \sigma_{ij}^{12} \leq 0 \text{ for } (i,\ j) \neq (i_0,\ j_0)]$$
$$\leq \sigma_{i_0 j_0} \qquad (\text{since } \sigma_{i_0 j_0} \geq 0 \text{ and } \alpha' \Sigma_{11} \alpha \geq \sum \alpha_i^2 \geq \alpha_{i_0}^2 = 1).$$

■

NOTE. It seems that if one element $\sigma_{i_0 j_0}$ of Σ_{12} is sufficiently larger than the others (but Σ is still nonnegative definite), then RCC = $\sigma_{i_0 j_0}$.

LEMMA 8. *Suppose all interblock correlations are the same, say* ρ (> 0), *and all intrablock correlations are also the same, i.e.*,

$$\Sigma_{12} = \rho \mathbb{J}_{p \times q}, \qquad \Sigma_{11} = \rho_1 \mathbb{J}_{p \times p} + (1 - \rho_1)\mathbb{I}_p,$$
$$\Sigma_{22} = \rho_2 \mathbb{J}_{q \times q} + (1 - \rho_2)\mathbb{I}_q.$$

Then RCC $= \rho p q a b =$ CC, *where*

$$a = \frac{1}{\sqrt{\rho_1 p^2 + (1 - \rho_1) p}}, \qquad b = \frac{1}{\sqrt{\rho_2 q^2 + (1 \rho_2) q}}.$$

Proof. Simple calculation shows that the CC is $\rho p q a b$ and the canonical coefficients consist of equal weights a from the $Y_i^{(1)}$'s and b from the $Y_j^{(2)}$'s. The rest follows by noting that both a and b are positive. ■

NOTE 1. In this case min RCC $= \rho$. Similar results hold when $\rho < 0$.

NOTE 2. This simple proof illustrates an important (albeit trivial) point regarding finding RCC under any type of restrictions. That is, since the calculation of RCC is in general much more difficult than that of CC, it may be sensible to calculate the CC, and then check if the contributing coefficients satisfy the relevant restrictions.

NOTE 3. Neither Lemma 7 nor Lemma 8 provides necessary conditions for the bounds in Lemma 6 to be attained.

3.2. *Invariance under Nonsingular Transformation*

CC remains unchanged under any nonsingular linear transformation, but RCC does not. This can be illustrated by the following example:

EXAMPLE 1. Let $p = 2 = q$, and let $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ have the following covariance structure:

$$\Sigma_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \Sigma_{22} \quad \text{and} \quad \Sigma_{12} = \begin{bmatrix} 0.4 & 0 \\ 0 & -0.1 \end{bmatrix}$$

It is easy to verify that $\mathrm{CC}(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}) = 0.4 = \mathrm{RCC}(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)})$. Now consider the effect of multiplying $\mathbf{Y}^{(1)}$ by P and $\mathbf{Y}^{(2)}$ by Q, where

$$P = \mathbb{I}_2 = -Q.$$

Since both P and Q are nonsingular, the CC between $P\mathbf{Y}^{(2)}$ and $Q\mathbf{Y}^{(2)}$ remains the same as before. But the covariance matrix between them is $-\Sigma_{12}$, and the RCC is only 0.1.

Next, a closer look is taken at the characteristics of the matrices P and Q which keep the RCC invariant in the sense discussed above. Mathematically speaking, the goal is to find characteristics of P and Q such that $\mathrm{RCC}(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}) = \mathrm{RCC}(P\mathbf{Y}^{(1)}, Q\mathbf{Y}^{(2)})$. The only obvious way to ensure this is to have a P such that

$$\{\alpha \geq 0\} \iff \{P'\alpha \geq 0\} \tag{29}$$

(and a similar Q). Clearly (29) holds if the elements of P and P^{-1} are all nonnegative.

It may be noted that the RCC remains unchanged under multiplication by a permutation matrix. This is, of course, obvious even without the observation in the previous paragraph. Because the effect of multiplying the random variables by a permutation matrix is the same as that of relabeling, it does not change the problem or the solution. Also, since an orthonormal matrix P with all nonnegative elements ($\Leftrightarrow P^{-1} = P'$ with all nonnegative elements) has to be a permutation matrix, it is the only type of orthonormal matrix which *always* keeps RCC invariant.

4. CONCLUDING REMARKS

It may be somewhat surprising that the RCC (under nonnegativity restrictions) can be obtained from the unrestricted CCs of different subvectors as described in Section 2.2. However, it is intuitively clear why the subvectors appear in the picture: the only way the maximum in (5) may not be attained is if some of the optimal coefficients related to the supremum are zeros.

It has been noted that a general inequality type restriction can be transformed into NNR and hence can be solved as indicated in Section 2. However, there remain many other types of restrictions [viz., (v) and (vii) of Section 1] where it was not possible to find an exact analytic solution. The problem is to tackle the extraneous slack/surplus variables (similar to what has been experienced while considering higher order RCCs). Thus in many circumstances, it will be necessary to settle for approximate solutions. In such a case, numerical analysis and/or a neural network approach may be successful. It may be possible to find a unifying (over all possible restrictions) treatment for solving the RCC problem in those approaches, and further research is needed to achieve this goal. Yanai and Takane [12] took a different approach to the RCC problem with linear constraints.

One principal motivation for the general study of the RCC is the interpretability of the coefficients. In the usual CCA, often one can end up with very different sets of coefficients which give (at least approximately) maximal correlations. This is much less likely to happen with the RCC. Also, if it does, it should cause no concern to the experimenter because all the candidate coefficients must satisfy the "reasonable" constraints that have been built into the problem.

The sample RCC is obtained when, in the calculation of Section 2.1, Σ is replaced by the sample covariance matrix $\mathbf{S}_n$. The statistical properties of this sample RCC have been explored in Das and Sen [4].

The authors are grateful to the referee for his very careful reading of the manuscript and for pointing out the usefulness of a branch and bound algorithm in finding optimal coefficients.

REFERENCES

1 T. W. Anderson, *An Introduction of Multivariate Statistical Analysis*, 2nd ed. Wiley, 1984.

2 S. Das, Restricted Canonical Correlations, *Doctoral Dissertation*, Mimeo Ser. 2095, Inst. of Statistics, Univ. of North Carolina, Chapel Hill, 1993.

3 S. Das and P. K. Sen, Simultaneous spike trains and stochastic dependence, *submitted for publication.*

4 S. Das and P. K. Sen, Asymptotic distribution of restricted canonical correlations and relevant resampling methods, *submitted for publication.*

5 F. Galton, *Natural Inheritance*, Macmillan, London, 1889.
6 H. Hotelling, The most predictable criterion, *J. Educ. Psychol.* 26:139–142 (1935).
7 H. Hotelling, Relation between two sets of variates, *Biometrika* 28:321–377 (1936).
8 K. Pearson, Mathematical contributions to the theory of evolution *III*. Regression, heredity and panmixia, *Philos. Trans. Roy. Soc. London Ser. A* 187:253–318 (1896).
9 K. Pearson, Mathematical contributions to the theory of evolution V. On the Reconstruction of the stature of prehistoric races, *Philos. Trans. Roy. Soc. London Ser. A* 192:169–244 (1898).
10 B. R. Rao, Partial canonical correlations, *Trab. Estadist. Invest. Oper.* 20:211–219 (1969).
11 N. H. Timm and J. E. Carlson, Part and bipartial canonical correlation analysis, *Psychometrika* 41:159–176 (1976).
12 H. Yanai and Y. Takane, Canonical correlation analysis in linear constraints, *Linear Algebra Appl.* 176:75–89 (1992).

Received 20 April 1993; final manuscript accepted 15 March 1994