RENATO LEONI

Canonical Correlation Analysis

UNIVERSITY OF FLORENCE
DEPARTMENT OF STATISTICS "G. PARENTI"
FLORENCE, 2007

1 INTRODUCTION

In analysing a data matrix such as that considered in PCA, it is often possible to recognize homogeneous sets of variables (e.g., economic, demographic, social), each set representing a certain aspect of the phenomenon under study.

The focus of canonical correlation analysis (CCA) is on the study of relationships among these sets of variables by an appropriate synthesis of the original variables of each set, providing at the same time the researcher with a graphical representation of results on a subspace of low dimension (usually one or two).

It may be worth adding that, although the applications of CCA are rather limited since the interpretation of results is often difficult, CCA provides a general framework for many multidimensional methods like regression, discriminant analysis and correspondence analysis, which are all special cases of CCA.

Unlike other approaches to CCA, in this paper no reference to an underlying probabilistic model will be done. Moreover, we will confine ourselves to considering the case of two sets of variables.

The contents of the paper can be summarized as follows.

In Section 2, the basic data and their algebraic structure are introduced. In Section 3, an approach to CCA is presented. In Section 4, rules for a graphical representation of results are given. Finally, in Section 5, other approaches to CCA are set out [(1)].

(1) Numerical examples, based both on fictitiuos and real data, are provided apart. Relevant algebraic concepts are stated in [17].

2 BASIC DATA AND THEIR ALGEBRAIC STRUCTURE

2.1 BASIC DATA

Consider the matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \cdots & \cdots & \cdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

where x_{ij} $(i = 1, \dots, n; j = 1, \dots, p)$ denotes the value of the jth quantitative variable observed on the ith individual (2).

Setting $(j = 1, \dots, p)$

$$\mathbf{x}_j = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{bmatrix}$$

and $(i = 1, \dots, n)$

$$\boldsymbol{x}_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix},$$

we can write

$$\mathbf{X} = [\, \mathbf{x}_1 \cdots \mathbf{x}_p \,]$$

and

$$\mathbf{X}' = [\, \boldsymbol{x}_1 \cdots \boldsymbol{x}_n \,] \,.$$

Considering the notation just introduced, we say that $\mathbf{x}_1, \dots, \mathbf{x}_p$ and $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ represent, respectively, the p variables and the n individuals.

Moreover, suppose that the p variables $\mathbf{x}_1, \dots, \mathbf{x}_p$ are partitioned into two sets – the first consisting of the p_1 variables $\mathbf{x}_1, \dots, \mathbf{x}_{p_1}$, the second of the $p_2 \geq p_1$ variables $\mathbf{x}_{p_1+1}, \dots, \mathbf{x}_{p_1+p_2}$ $(p_1 + p_2 = p)$.

(2) In what follows we consider acquired the main concepts and definitions introduced in [18], partly summarized here.

Then, we can write

$$\mathbf{X} = \left[\mathbf{x}_1 \cdots \mathbf{x}_{p_1} \mid \mathbf{x}_{p_1+1} \cdots \mathbf{x}_{p_1+p_2}\right] = \left[\mathbf{X}_1 \quad \mathbf{X}_2\right].$$

2.2 ALGEBRAIC STRUCTURE

With reference to the variables, regarding them as elements of R^n, R^n (variable space) is equipped with a Euclidean metric in the following way.

As in PCA, the matrix (symmetric and positive definite (p.d.)) of the Euclidean metric in R^n – with respect to the basis consisting of the n canonical vectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ – is

$$\mathbf{M} = \mathrm{diag}\,(m_1, \dots, m_n)$$

where $m_i > 0$ $(i = 1, \dots, n)$, $\Sigma_i m_i = 1$, represents the weight given to the ith individual and denotes its «importance» in the set of the n individuals.

Since the matrix $\mathbf{Y}$ of the p variables measured in terms of deviations from the means – partitioned in the same way as $\mathbf{X}$ – becomes

$$\mathbf{Y} = \left[\mathbf{y}_1 \cdots \mathbf{y}_{p_1} \mid \mathbf{y}_{p_1+1} \cdots \mathbf{y}_{p_1+p_2}\right] = \left[\mathbf{Y}_1 \quad \mathbf{Y}_2\right],$$

the covariance matrix $\mathbf{V}$ of the p variables can be written as

$$\mathbf{V} = \mathbf{Y}'\mathbf{M}\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1'\mathbf{M}\mathbf{Y}_1 & \mathbf{Y}_1'\mathbf{M}\mathbf{Y}_2 \\ \mathbf{Y}_2'\mathbf{M}\mathbf{Y}_1 & \mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}.$$

Notice that – assuming that $r(\mathbf{Y}_1) = p_1$ and $r(\mathbf{Y}_2) = p_2$ – is

$$r(\mathbf{V}_{11}) = p_1 \quad , \quad r(\mathbf{V}_{22}) = p_2 .$$

Moreover,

$$r(\mathbf{V}_{12}) = r(\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_2) = r(\mathbf{Y}_1'\mathbf{M}^{1/2}\mathbf{M}^{1/2}\mathbf{Y}_2) \le \min\{r(\mathbf{Y}_1), r(\mathbf{Y}_2)\} = p_1 .$$

Of course, $r(\mathbf{V}_{12}) = r(\mathbf{V}_{21})$.

We will suppose that $r(\mathbf{V}_{12}) = k > 0$. The reasons of this assumption will become apparent from what follows.

With reference to the individuals, regarding them as elements of R^p, R^p (individual space) is equipped with a Euclidean metric in the following way.

The matrix of the Euclidean metric in R^p – with respect to the basis consisting of the p canonical vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$ – is (*block Mahalanobis metric*)

$$\mathbf{Q} = \mathrm{diag}\,(\mathbf{V}_{11}^{-1}, \mathbf{V}_{22}^{-1}) \ .$$

Clearly, as $\mathbf{V}_{11}^{-1}$ and $\mathbf{V}_{22}^{-1}$ are symmetric and p.d., $\mathbf{Q}$ is symmetric and p.d. too.

The choice of this metric is a natural extension of the choice made in PCA, where the scope of obtaining homogeneus variances was reached setting $\mathbf{Q} = \mathbf{Q}_{1/\sigma^2}$. Obviously, in the present case, instead of the inverses of the variances of the variables, it is necessary to consider the inverses of the covariance matrices of the two sets of variables.

As will be afterwards shown, this choice is coherent with other ways of presenting CCA (Section 5.2).

3 AN APPROACH TO CCA

3.1 CANONICAL FACTORS, CANONICAL VARIABLES, AND CANONICAL CORRELATION COEFFICIENTS

3.1.1 THE FIRST STEP

The first step of the approach to CCA we are considering consists in determining a linear combination $\tilde{\mathbf{z}}_{(1)1}$ of $\mathbf{y}_1, \dots, \mathbf{y}_{p_1}$ and a linear combination $\tilde{\mathbf{z}}_{(2)1}$ of $\mathbf{y}_{p_1+1}, \dots, \mathbf{y}_{p_1+p_2}$ such that the cosine of the angle they form (the linear correlation coefficient) $\cos(\tilde{\mathbf{z}}_{(1)1}, \tilde{\mathbf{z}}_{(2)1})$ is a maximum (3).

Setting

$$\mathbf{z}_{(1)1} = \mathbf{y}_1 a_{(1)1} + \dots + \mathbf{y}_{p_1} a_{(1)p_1} = \left[\mathbf{y}_1 \cdots \mathbf{y}_{p_1}\right] \begin{bmatrix} a_{(1)1} \\ \vdots \\ a_{(1)p_1} \end{bmatrix} = \mathbf{Y}_1 \mathbf{a}_{(1)1} \, ,$$

$$\mathbf{z}_{(2)1} = \mathbf{y}_{p_1+1} a_{(2)p_1+1} + \dots + \mathbf{y}_{p_1+p_2} a_{(2)p_1+p_2} = \left[\mathbf{y}_{p_1+1} \cdots \mathbf{y}_{p_1+p_2}\right] \begin{bmatrix} a_{(2)p_1+1} \\ \vdots \\ a_{(2)p_1+p_2} \end{bmatrix} = \mathbf{Y}_2 \mathbf{a}_{(2)1} \, ,$$

$$\cos(\mathbf{z}_{(1)1}, \mathbf{z}_{(2)1}) = \cos(\mathbf{Y}_1 \mathbf{a}_{(1)1}, \mathbf{Y}_2 \mathbf{a}_{(2)1})$$

$$= \frac{\mathbf{a}'_{(1)1} \mathbf{Y}'_1 \mathbf{M} \mathbf{Y}_2 \mathbf{a}_{(2)1}}{\{(\mathbf{a}'_{(1)1} \mathbf{Y}'_1 \mathbf{M} \mathbf{Y}_1 \mathbf{a}_{(1)1})(\mathbf{a}'_{(2)1} \mathbf{Y}'_2 \mathbf{M} \mathbf{Y}_2 \mathbf{a}_{(2)1})\}^{1/2}} = \frac{\mathbf{a}'_{(1)1} \mathbf{V}_{12} \mathbf{a}_{(2)1}}{\{(\mathbf{a}'_{(1)1} \mathbf{V}_{11} \mathbf{a}_{(1)1})(\mathbf{a}'_{(2)1} \mathbf{V}_{22} \mathbf{a}_{(2)1})\}^{1/2}} \, ,$$

we have to find out

$$(1) \qquad \underset{\mathbf{a}_{(1)1}, \mathbf{a}_{(2)1}}{\text{Max}} \frac{\mathbf{a}'_{(1)1} \mathbf{V}_{12} \mathbf{a}_{(2)1}}{\{(\mathbf{a}'_{(1)1} \mathbf{V}_{11} \mathbf{a}_{(1)1})(\mathbf{a}'_{(2)1} \mathbf{V}_{22} \mathbf{a}_{(2)1})\}^{1/2}} \, .$$

In order to facilitate the solution of this problem, first it is useful to notice that $\cos(\mathbf{z}_{(1)1}, \mathbf{z}_{(2)1})$ is invariant when $\mathbf{z}_{(1)1}$ (or $\mathbf{a}_{(1)1}$) is multiplied by c_1 and $\mathbf{z}_{(2)1}$ (or $\mathbf{a}_{(2)1}$) is multiplied by c_2, where c_1, c_2 are different from zero and of the same sign.

---

(3) Coherently with the aim of CCA, principally directed to the study of relationships among sets of variables, we will mainly refer to the solution in the variable space.

Therefore, we may consider $\mathbf{z}_{(1)1}$ and $\mathbf{z}_{(2)1}$ as vectors of unitary square length (variance), so (1) is simplified in the following way

$$(1') \quad \underset{\mathbf{a}_{(1)1},\,\mathbf{a}_{(2)1}}{\text{Max}} \; \mathbf{a}'_{(1)1}\mathbf{V}_{12}\,\mathbf{a}_{(2)1} \quad , \quad \mathbf{a}'_{(1)1}\mathbf{V}_{11}\,\mathbf{a}_{(1)1} = 1 \quad , \quad \mathbf{a}'_{(2)1}\mathbf{V}_{22}\,\mathbf{a}_{(2)1} = 1 \; .$$

To solve the problem of constrained maximization set in (1'), consider the Lagrange function

$$\begin{aligned} \mathrm{L}(\mathbf{a}_{(1)1}, \mathbf{a}_{(2)1}, \mu_1, \mu_2) = \mathbf{a}'_{(1)1}\mathbf{V}_{12}\,\mathbf{a}_{(2)1} - \frac{1}{2}\mu_1(\mathbf{a}'_{(1)1}\mathbf{V}_{11}\,\mathbf{a}_{(1)1} - 1) \\ - \frac{1}{2}\mu_2(\mathbf{a}'_{(2)1}\mathbf{V}_{22}\,\mathbf{a}_{(2)1} - 1) \end{aligned}$$

where μ_1, μ_2 are Lagrange multipliers.

At $(\tilde{\mathbf{a}}_{(1)1}, \tilde{\mathbf{a}}_{(2)1}, \tilde{\mu}_1, \tilde{\mu}_2)$ where $\mathrm{L}(\mathbf{a}_{(1)1}, \mathbf{a}_{(2)1}, \mu_1, \mu_2)$ has a maximum, it must be (4)

$$\begin{aligned} \left.\frac{\partial \mathrm{L}}{\partial \mathbf{a}_{(1)1}}\right|_{(\tilde{\mathbf{a}}_{(1)1}, \tilde{\mathbf{a}}_{(2)1}, \tilde{\mu}_1, \tilde{\mu}_2)} &= \mathbf{V}_{12}\tilde{\mathbf{a}}_{(2)1} - \tilde{\mu}_1\mathbf{V}_{11}\tilde{\mathbf{a}}_{(1)1} = \mathbf{0} \\ \left.\frac{\partial \mathrm{L}}{\partial \mathbf{a}_{(2)1}}\right|_{(\tilde{\mathbf{a}}_{(1)1}, \tilde{\mathbf{a}}_{(2)1}, \tilde{\mu}_1, \tilde{\mu}_2)} &= \mathbf{V}_{21}\tilde{\mathbf{a}}_{(1)1} - \tilde{\mu}_2\mathbf{V}_{22}\tilde{\mathbf{a}}_{(2)1} = \mathbf{0} \\ \left.\frac{\partial \mathrm{L}}{\partial \mu_1}\right|_{(\tilde{\mathbf{a}}_{(1)1}, \tilde{\mathbf{a}}_{(2)1}, \tilde{\mu}_1, \tilde{\mu}_2)} &= \tilde{\mathbf{a}}'_{(1)1}\mathbf{V}_{11}\tilde{\mathbf{a}}_{(1)1} = 1 \\ \left.\frac{\partial \mathrm{L}}{\partial \mu_2}\right|_{(\tilde{\mathbf{a}}_{(1)1}, \tilde{\mathbf{a}}_{(2)1}, \tilde{\mu}_1, \tilde{\mu}_2)} &= \tilde{\mathbf{a}}'_{(2)1}\mathbf{V}_{22}\tilde{\mathbf{a}}_{(2)1} = 1 \end{aligned}$$

from which we immediately deduce that

$$\tilde{\mathbf{a}}'_{(1)1}\mathbf{V}_{12}\tilde{\mathbf{a}}_{(2)1} = \tilde{\mu}_1 \quad , \quad \tilde{\mathbf{a}}'_{(2)1}\mathbf{V}_{21}\tilde{\mathbf{a}}_{(1)1} = \tilde{\mu}_2$$

and hence that

$$\tilde{\mathbf{a}}'_{(1)1}\mathbf{V}_{12}\tilde{\mathbf{a}}_{(2)1} = \tilde{\mu}_1 = \cos(\tilde{\mathbf{z}}_{(1)1}, \tilde{\mathbf{z}}_{(2)1}) = \tilde{\mu}_2 = \tilde{\mathbf{a}}'_{(2)1}\mathbf{V}_{21}\tilde{\mathbf{a}}_{(1)1} \; .$$

Therefore − since it must be

(4) Here and in what follows **0** denotes a zero column vector of appropriate order.

$$\mathbf{V}_{12}\,\tilde{\mathbf{a}}_{(2)1} = \cos(\tilde{\mathbf{z}}_{(1)1},\tilde{\mathbf{z}}_{(2)1})\,\mathbf{V}_{11}\,\tilde{\mathbf{a}}_{(1)1} \quad , \quad \mathbf{V}_{21}\,\tilde{\mathbf{a}}_{(1)1} = \cos(\tilde{\mathbf{z}}_{(1)1},\tilde{\mathbf{z}}_{(2)1})\,\mathbf{V}_{22}\,\tilde{\mathbf{a}}_{(2)1}$$
$$\tilde{\mathbf{a}}'_{(1)1}\mathbf{V}_{11}\,\tilde{\mathbf{a}}_{(1)1} = 1 \quad , \quad \tilde{\mathbf{a}}'_{(2)1}\mathbf{V}_{22}\,\tilde{\mathbf{a}}_{(2)1} = 1$$

– we realize that $\cos(\tilde{\mathbf{z}}_{(1)1},\tilde{\mathbf{z}}_{(2)1})$ and $\tilde{\mathbf{a}}_{(1)1},\tilde{\mathbf{a}}_{(2)1}$ must be found among the solutions of the system

$$(2) \quad \begin{bmatrix} -r\,\mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -r\,\mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{(1)1} \\ \mathbf{a}_{(2)1} \end{bmatrix} = \mathbf{0} \;\; , \;\; \mathbf{a}'_{(1)1}\mathbf{V}_{11}\,\mathbf{a}_{(1)1} = 1 \;\; , \;\; \mathbf{a}'_{(2)1}\mathbf{V}_{22}\,\mathbf{a}_{(2)1} = 1$$

in the unknowns $r\,,\mathbf{a}_{(1)1},\mathbf{a}_{(2)1}$.

To this end, pay attention to the system

$$(3) \qquad \begin{bmatrix} -r\,\mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -r\,\mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{(1)1} \\ \mathbf{a}_{(2)1} \end{bmatrix} = \mathbf{0}$$

and ask for which values of r it admits non-trivial solutions with respect to the unknowns $\mathbf{a}_{(1)1},\mathbf{a}_{(2)1}$ (5).

In order for it to possibly happen it is necessary and sufficient that

$$(4) \qquad \det \begin{bmatrix} -r\,\mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -r\,\mathbf{V}_{22} \end{bmatrix} = 0 \; .$$

Assuming for the sake of simplicity that the solutions different from zero of this latter equation are all distinct (6), it can be seen that the $p = p_1 + p_2$ real values of r which become available are of the kind:

- $k = r(\mathbf{V}_{12}) = r(\mathbf{V}_{21})$ positive values $\tilde{r}_1 > \ldots > \tilde{r}_k$;
- $p - 2k$ zero values $\tilde{r}_0$;
- k negative values $-\tilde{r}_k > \ldots > -\tilde{r}_1$.

Let us consider the value $\tilde{r}_1$.

Notice that $\tilde{r}_1$ can also be obtained as square root of the first largest eigenvalue of the equation

(5) On account of the constraints of normalization in (2), it is necessary to consider only the non-trivial solutions of the system (3).

(6) The case in which the solutions different from zero are not all distinct does not present any difficulty and is considered in [15].

(5) $$\det(-r^2\mathbf{I}_{p_1} + \mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}) = 0$$

where $\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$ may be interpreted as the matrix of a selfadjoint transformation in the metric represented by $\mathbf{V}_{11}$ $((\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})'\mathbf{V}_{11} = \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} = \mathbf{V}_{11}(\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}))$.

Now, consider the system

(6) $$\begin{bmatrix} -\tilde{r}_1\mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -\tilde{r}_1\mathbf{V}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{a}_{(1)1} \\ \mathbf{a}_{(2)1} \end{bmatrix} = \mathbf{0} \;\;,\;\; \mathbf{a}'_{(1)1}\mathbf{V}_{11}\mathbf{a}_{(1)1} = 1 \;\;,\;\; \mathbf{a}'_{(2)1}\mathbf{V}_{22}\mathbf{a}_{(2)1} = 1$$

in the unknowns $\mathbf{a}_{(1)1}, \mathbf{a}_{(2)1}$, obtained by setting $r = \tilde{r}_1$ in (2).

Pay attention to the system represented by the first equation in (6), namely

(6') $$\begin{bmatrix} -\tilde{r}_1\mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -\tilde{r}_1\mathbf{V}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{a}_{(1)1} \\ \mathbf{a}_{(2)1} \end{bmatrix} = \mathbf{0}\ .$$

Premultiplying both members of (6') by the matrix

(7) $$\begin{bmatrix} \tilde{r}_1\mathbf{I}_{p_1} & \mathbf{V}_{12}\mathbf{V}_{22}^{-1} \\ \mathbf{O}_{(p_2,p_1)} & (1/\tilde{r}_1)\,\mathbf{V}_{22}^{-1} \end{bmatrix},$$

we get the system

(8) $$(-\tilde{r}_1^2\mathbf{I}_{p_1} + \mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{a}_{(1)1} = \mathbf{0} \;\;,\;\; \mathbf{a}_{(2)1} = \frac{1}{\tilde{r}_1}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{a}_{(1)1}$$

and, as the matrix (7) is nonsingular, the systems (6') and (8) are equivalent.

Clearly, the first equation in (8) admits the eigenvector $\bar{\mathbf{a}}_{(1)1}$ such that $\bar{\mathbf{a}}'_{(1)1}\mathbf{V}_{11}\bar{\mathbf{a}}_{(1)1} = 1$, corresponding to the eigenvalue $\tilde{r}_1^2$.

In turn, the second equation in (8), for $\mathbf{a}_{(1)1} = \bar{\mathbf{a}}_{(1)1}$, gives the vector

$$\bar{\mathbf{a}}_{(2)1} = \frac{1}{\tilde{r}_1}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\bar{\mathbf{a}}_{(1)1}$$

such that $\bar{\mathbf{a}}'_{(2)1}\mathbf{V}_{22}\bar{\mathbf{a}}_{(2)1} = 1$.

In fact, taking into account that

$$\{(-\tilde{r}_1^2 \mathbf{I}_{p_1} + \mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})\bar{\mathbf{a}}_{(1)1} = \mathbf{0}\} \Leftrightarrow \{\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\bar{\mathbf{a}}_{(1)1} = \tilde{r}_1^2\,\bar{\mathbf{a}}_{(1)1}\}$$
$$\Leftrightarrow \{\bar{\mathbf{a}}'_{(1)1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\bar{\mathbf{a}}_{(1)1} = \tilde{r}_1^2\}\ ,$$

we have

$$\bar{\mathbf{a}}'_{(2)1}\mathbf{V}_{22}\bar{\mathbf{a}}_{(2)1} = \frac{1}{\tilde{r}_1^2}\,\bar{\mathbf{a}}'_{(1)1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{22}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\bar{\mathbf{a}}_{(1)1} = \frac{1}{\tilde{r}_1^2}\,\bar{\mathbf{a}}'_{(1)1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\bar{\mathbf{a}}_{(1)1}$$
$$= \frac{1}{\tilde{r}_1^2}\,\tilde{r}_1^2 = 1\ .$$

Thus, we conclude that $\bar{\mathbf{a}}_{(1)1}, \bar{\mathbf{a}}_{(2)1}$, solutions of the system (8) with the properties mentioned above, are also solutions of the system (6).

Before maintaining that $\tilde{\mathbf{a}}_{(1)1} = \bar{\mathbf{a}}_{(1)1}, \tilde{\mathbf{a}}_{(2)1} = \bar{\mathbf{a}}_{(2)1}$ solve the problem set in (1) it is necessary to verify that $\tilde{r}_1^2 \le 1$, which implies that the solutions of equation (4) fall within the (closed) interval $[-1\ ,+1]$.

Actually, we can write ($\bar{\mathbf{a}}'_{(1)1}\mathbf{V}_{11}\bar{\mathbf{a}}_{(1)1} = 1$)

$$\begin{aligned}
\tilde{r}_1^2 &= \frac{\bar{\mathbf{a}}'_{(1)1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\bar{\mathbf{a}}_{(1)1}}{\bar{\mathbf{a}}'_{(1)1}\mathbf{V}_{11}\bar{\mathbf{a}}_{(1)1}} \\
&= \frac{\bar{\mathbf{a}}'_{(1)1}\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_2(\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2)^{-1}\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1}}{\bar{\mathbf{a}}'_{(1)1}\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1}} \\
&= \frac{\bar{\mathbf{a}}'_{(1)1}\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_2(\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2)^{-1}\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2(\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2)^{-1}\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1}}{\bar{\mathbf{a}}'_{(1)1}\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1}} \\
&= \frac{(\mathbf{P}_2\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1})'\mathbf{M}(\mathbf{P}_2\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1})}{\bar{\mathbf{a}}'_{(1)1}\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1}} \\
&= \frac{\|\mathbf{P}_2\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1}\|^2}{\|\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1}\|^2}
\end{aligned}$$

where $\mathbf{P}_2 = \mathbf{Y}_2(\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2)^{-1}\mathbf{Y}_2'\mathbf{M}$ denotes the orthogonal projection matrix on the subspace $S(\mathbf{Y}_2)$ of R^n spanned by the column vectors of $\mathbf{Y}_2$.

Therefore, as the square length of the orthogonal projection $\mathbf{P}_2\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1}$ of $\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1}$ on $S(\mathbf{Y}_2)$ is not greater than the square length of $\mathbf{Y}_1\bar{\mathbf{a}}_{(1)1}$ (Pythagoras theorem), we have that $\tilde{r}_1^2 \le 1$.

The vectors $\tilde{\mathbf{a}}_{(1)1}$ and $\tilde{\mathbf{a}}_{(2)1}$, such that

$$\tilde{\mathbf{a}}'_{(1)1}\mathbf{V}_{11}\tilde{\mathbf{a}}_{(1)1} = 1 \quad , \quad \tilde{\mathbf{a}}'_{(2)1}\mathbf{V}_{22}\tilde{\mathbf{a}}_{(2)1} = 1 \quad , \quad \tilde{\mathbf{a}}'_{(1)1}\mathbf{V}_{12}\tilde{\mathbf{a}}_{(2)1} = \tilde{\mathbf{a}}'_{(2)1}\mathbf{V}_{21}\tilde{\mathbf{a}}_{(1)1} = \tilde{r}_1\ ,$$

are called (the first two) *canonical factors*.

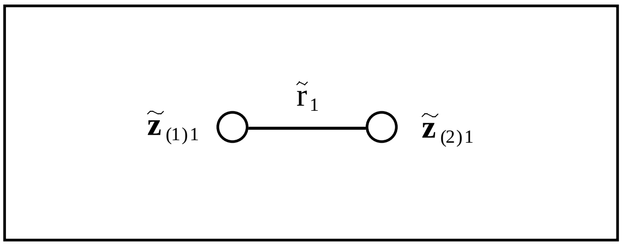
The vectors $\tilde{\mathbf{z}}_{(1)1} = \mathbf{Y}_1 \tilde{\mathbf{a}}_{(1)1}$ and $\tilde{\mathbf{z}}_{(2)1} = \mathbf{Y}_2 \tilde{\mathbf{a}}_{(2)1}$, such that

$$\tilde{\mathbf{z}}'_{(1)1}\mathbf{M}\tilde{\mathbf{z}}_{(1)1} = 1 \quad , \quad \tilde{\mathbf{z}}'_{(2)1}\mathbf{M}\tilde{\mathbf{z}}_{(2)1} = 1 \quad , \quad \tilde{\mathbf{z}}'_{(1)1}\mathbf{M}\tilde{\mathbf{z}}_{(2)1} = \tilde{\mathbf{z}}'_{(2)1}\mathbf{M}\tilde{\mathbf{z}}_{(1)1} = \tilde{\mathrm{r}}_1 \ ,$$

are called (the first two) *canonical variables* or *canonical vectors*.

In turn, the cosine of the angle formed by $\tilde{\mathbf{z}}_{(1)1}$ and $\tilde{\mathbf{z}}_{(2)1}$, namely $\tilde{\mathrm{r}}_1$, is called (the first one) *canonical correlation coefficient* (Fig. 1).

Fig. 1



3.1.2 THE SECOND STEP

The second step consists in determining a linear combination $\tilde{\mathbf{z}}_{(1)2}$ of $\mathbf{y}_1, \dots, \mathbf{y}_{p_1}$, orthogonal to the subspace spanned by $\tilde{\mathbf{z}}_{(1)1}$, and a linear combination $\tilde{\mathbf{z}}_{(2)2}$ of $\mathbf{y}_{p_1+1}, \dots, \mathbf{y}_{p_1+p_2}$, orthogonal to the subspace spanned by $\tilde{\mathbf{z}}_{(2)1}$, such that the cosine of the angle they form (the linear correlation coefficient) $\cos(\tilde{\mathbf{z}}_{(1)1}, \tilde{\mathbf{z}}_{(2)1})$ is a maximum.

Setting

$$\mathbf{z}_{(1)2} = \mathbf{Y}_1 \mathbf{a}_{(1)2} \quad , \quad \mathbf{z}_{(2)2} = \mathbf{Y}_2 \mathbf{a}_{(2)2}$$

and

$$\cos(\mathbf{z}_{(1)2}, \mathbf{z}_{(2)2}) = \frac{\mathbf{a}'_{(1)2}\mathbf{V}_{12}\mathbf{a}_{(2)2}}{\{(\mathbf{a}'_{(1)2}\mathbf{V}_{11}\mathbf{a}_{(1)2})(\mathbf{a}'_{(2)1}\mathbf{V}_{22}\mathbf{a}_{(2)2})\}^{1/2}} \ ,$$

we have to find out

$$(9) \qquad \underset{\mathbf{a}_{(1)2},\, \mathbf{a}_{(2)2}}{\mathrm{Max}} \ \frac{\mathbf{a}'_{(1)2}\mathbf{V}_{12}\mathbf{a}_{(2)2}}{\{(\mathbf{a}'_{(1)2}\mathbf{V}_{11}\mathbf{a}_{(1)2})(\mathbf{a}'_{(2)2}\mathbf{V}_{22}\mathbf{a}_{(2)2})\}^{1/2}}$$

under the constraints

$$\mathbf{z}'_{(1)2}\mathbf{M}\widetilde{\mathbf{z}}_{(1)1} = \mathbf{a}'_{(1)2}\mathbf{Y}'_1\mathbf{M}\mathbf{Y}_1\widetilde{\mathbf{a}}_{(1)1} = \mathbf{a}'_{(1)2}\mathbf{V}_{11}\widetilde{\mathbf{a}}_{(1)1} = 0$$

(10)

$$\mathbf{z}'_{(2)2}\mathbf{M}\widetilde{\mathbf{z}}_{(2)1} = \mathbf{a}'_{(2)2}\mathbf{Y}'_2\mathbf{M}\mathbf{Y}_2\widetilde{\mathbf{a}}_{(2)1} = \mathbf{a}'_{(2)2}\mathbf{V}_{22}\widetilde{\mathbf{a}}_{(2)1} = 0 \,.$$

Equivalently, assuming that $\mathbf{z}_{(1)2}$ and $\mathbf{z}_{(2)2}$ are vectors of unitary square length, we have to look for

(9') $$\underset{\mathbf{a}_{(1)2},\,\mathbf{a}_{(2)2}}{\text{Max}} \; \mathbf{a}'_{(1)2}\mathbf{V}_{12}\mathbf{a}_{(2)2} \quad , \quad \mathbf{a}'_{(1)2}\mathbf{V}_{11}\mathbf{a}_{(1)2} = 1 \quad , \quad \mathbf{a}'_{(2)2}\mathbf{V}_{22}\mathbf{a}_{(2)2} = 1$$

under the constraints in (10).

To solve the problem of constrained maximization set in (9') and (10), consider the Lagrange function

$$\mathrm{L}(\mathbf{a}_{(1)2}, \mathbf{a}_{(2)2}, \kappa_1, \kappa_2, \kappa_3, \kappa_4) = \mathbf{a}'_{(1)2}\mathbf{V}_{12}\mathbf{a}_{(2)2} - \frac{1}{2}\kappa_1(\mathbf{a}'_{(1)2}\mathbf{V}_{11}\mathbf{a}_{(1)2} - 1)$$
$$- \frac{1}{2}\kappa_2(\mathbf{a}'_{(2)2}\mathbf{V}_{22}\mathbf{a}_{(2)2} - 1) - \frac{1}{2}\kappa_3(\mathbf{a}'_{(1)2}\mathbf{V}_{11}\widetilde{\mathbf{a}}_{(1)1}) - \frac{1}{2}\kappa_4(\mathbf{a}'_{(2)2}\mathbf{V}_{22}\widetilde{\mathbf{a}}_{(2)1})$$

where $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ are Lagrange multipliers.

At $(\widetilde{\mathbf{a}}_{(1)1}, \widetilde{\mathbf{a}}_{(2)1}, \widetilde{\kappa}_1, \widetilde{\kappa}_2, \widetilde{\kappa}_3, \widetilde{\kappa}_4)$ where $\mathrm{L}(\mathbf{a}_{(1)2}, \mathbf{a}_{(2)2}, \kappa_1, \kappa_2, \kappa_3, \kappa_4)$ has a maximum, it must be

$$\left.\frac{\partial \mathrm{L}}{\partial \mathbf{a}_{(1)2}}\right|_{(\widetilde{\mathbf{a}}_{(1)1}, \widetilde{\mathbf{a}}_{(2)1}, \widetilde{\kappa}_1, \widetilde{\kappa}_2, \widetilde{\kappa}_3, \widetilde{\kappa}_4)} = \mathbf{V}_{12}\widetilde{\mathbf{a}}_{(2)2} - \widetilde{\kappa}_1\mathbf{V}_{11}\widetilde{\mathbf{a}}_{(1)2} = \mathbf{0}$$

$$\left.\frac{\partial \mathrm{L}}{\partial \mathbf{a}_{(2)2}}\right|_{(\widetilde{\mathbf{a}}_{(1)1}, \widetilde{\mathbf{a}}_{(2)1}, \widetilde{\kappa}_1, \widetilde{\kappa}_2, \widetilde{\kappa}_3, \widetilde{\kappa}_4)} = \mathbf{V}_{21}\widetilde{\mathbf{a}}_{(1)2} - \widetilde{\kappa}_2\mathbf{V}_{22}\widetilde{\mathbf{a}}_{(2)2} = \mathbf{0}$$

$$\left.\frac{\partial \mathrm{L}}{\partial \kappa_1}\right|_{(\widetilde{\mathbf{a}}_{(1)1}, \widetilde{\mathbf{a}}_{(2)1}, \widetilde{\kappa}_1, \widetilde{\kappa}_2, \widetilde{\kappa}_3, \widetilde{\kappa}_4)} = \widetilde{\mathbf{a}}'_{(1)2}\mathbf{V}_{11}\widetilde{\mathbf{a}}_{(1)2} = 1$$

$$\left.\frac{\partial \mathrm{L}}{\partial \kappa_2}\right|_{(\widetilde{\mathbf{a}}_{(1)1}, \widetilde{\mathbf{a}}_{(2)1}, \widetilde{\kappa}_1, \widetilde{\kappa}_2, \widetilde{\kappa}_3, \widetilde{\kappa}_4)} = \widetilde{\mathbf{a}}'_{(2)2}\mathbf{V}_{22}\widetilde{\mathbf{a}}_{(2)2} = 1$$

$$\left.\frac{\partial \mathrm{L}}{\partial \kappa_3}\right|_{(\widetilde{\mathbf{a}}_{(1)1}, \widetilde{\mathbf{a}}_{(2)1}, \widetilde{\kappa}_1, \widetilde{\kappa}_2, \widetilde{\kappa}_3, \widetilde{\kappa}_4)} = \mathbf{a}'_{(1)2}\mathbf{V}_{11}\widetilde{\mathbf{a}}_{(1)1} = 0$$

$$\left.\frac{\partial \mathrm{L}}{\partial \kappa_4}\right|_{(\widetilde{\mathbf{a}}_{(1)1}, \widetilde{\mathbf{a}}_{(2)1}, \widetilde{\kappa}_1, \widetilde{\kappa}_2, \widetilde{\kappa}_3, \widetilde{\kappa}_4)} = \mathbf{a}'_{(2)2}\mathbf{V}_{22}\widetilde{\mathbf{a}}_{(2)1} = 0$$

from which we immediately deduce that

$$\tilde{\mathbf{a}}'_{(1)2}\mathbf{V}_{12}\tilde{\mathbf{a}}_{(2)2} = \tilde{\kappa}_1 \quad , \quad \tilde{\mathbf{a}}'_{(2)2}\mathbf{V}_{21}\tilde{\mathbf{a}}_{(1)2} = \tilde{\kappa}_2$$

and hence that

$$\tilde{\mathbf{a}}'_{(1)2}\mathbf{V}_{12}\tilde{\mathbf{a}}_{(2)2} = \tilde{\kappa}_1 = \cos(\tilde{\mathbf{z}}_{(1)2}, \tilde{\mathbf{z}}_{(2)2}) = \tilde{\kappa}_2 = \tilde{\mathbf{a}}'_{(2)2}\mathbf{V}_{21}\tilde{\mathbf{a}}_{(1)2} \, .$$

Therefore – since it must be

$$\begin{array}{ccc} \mathbf{V}_{12}\tilde{\mathbf{a}}_{(2)2} = \cos(\tilde{\mathbf{z}}_{(1)2}, \tilde{\mathbf{z}}_{(2)2})\mathbf{V}_{11}\tilde{\mathbf{a}}_{(1)2} & , & \mathbf{V}_{21}\tilde{\mathbf{a}}_{(1)2} = \cos(\tilde{\mathbf{z}}_{(1)2}, \tilde{\mathbf{z}}_{(2)2})\mathbf{V}_{22}\tilde{\mathbf{a}}_{(2)2} \\ \tilde{\mathbf{a}}'_{(1)2}\mathbf{V}_{11}\tilde{\mathbf{a}}_{(1)2} = 1 & , & \tilde{\mathbf{a}}'_{(2)2}\mathbf{V}_{22}\tilde{\mathbf{a}}_{(2)2} = 1 \\ \tilde{\mathbf{a}}'_{(1)2}\mathbf{V}_{11}\tilde{\mathbf{a}}_{(1)1} = 0 & , & \tilde{\mathbf{a}}'_{(2)2}\mathbf{V}_{22}\tilde{\mathbf{a}}_{(2)1} = 0 \end{array}$$

– we realize that $\cos(\tilde{\mathbf{z}}_{(1)2}, \tilde{\mathbf{z}}_{(2)2})$ and $\tilde{\mathbf{a}}_{(1)2}, \tilde{\mathbf{a}}_{(2)2}$ must be found among the solutions of the system

$$(11) \quad \begin{bmatrix} -\mathrm{r}\mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -\mathrm{r}\mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{(1)2} \\ \mathbf{a}_{(2)2} \end{bmatrix} = \mathbf{0} \; , \; \begin{array}{ccc} \mathbf{a}'_{(1)2}\mathbf{V}_{11}\mathbf{a}_{(1)2} = 1 & , & \mathbf{a}'_{(2)2}\mathbf{V}_{22}\mathbf{a}_{(2)2} = 1 \\ \mathbf{a}'_{(1)2}\mathbf{V}_{11}\tilde{\mathbf{a}}_{(1)1} = 0 & , & \mathbf{a}'_{(2)2}\mathbf{V}_{22}\tilde{\mathbf{a}}_{(2)1} = 0 \end{array}$$

in the unknowns $\mathrm{r}, \mathbf{a}_{(1)2}, \mathbf{a}_{(2)2}$.

To this end, consider the system

$$(12) \qquad \begin{bmatrix} -\mathrm{r}\mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -\mathrm{r}\mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{(1)2} \\ \mathbf{a}_{(2)2} \end{bmatrix} = \mathbf{0}$$

and ask for which values of r it admits non-trivial solutions with respect to the unknowns $\mathbf{a}_{(1)2}, \mathbf{a}_{(2)2}$ [(7)].

This system – similar to that one written in (3) – admits the solution $\tilde{\mathrm{r}}_2$ for the unknown r.

This solution can also be obtained as square root of the second largest eigenvalue of the equation (5).

Now, consider the system

(7) On account of the constraints of normalization in (11), it is necessary to consider only the non-trivial solutions of the system (12).

(13) $\begin{bmatrix} -\tilde{r}_2 \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -\tilde{r}_2 \mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{(1)2} \\ \mathbf{a}_{(2)2} \end{bmatrix} = \mathbf{0}$, $\begin{matrix} \mathbf{a}'_{(1)2} \mathbf{V}_{11} \mathbf{a}_{(1)2} = 1 & , & \mathbf{a}'_{(2)2} \mathbf{V}_{22} \mathbf{a}_{(2)2} = 1 \\ \mathbf{a}'_{(1)2} \mathbf{V}_{11} \widetilde{\mathbf{a}}_{(1)1} = 0 & , & \mathbf{a}'_{(2)2} \mathbf{V}_{22} \widetilde{\mathbf{a}}_{(2)1} = 0 \end{matrix}$

in the unknowns $\mathbf{a}_{(1)2}$, $\mathbf{a}_{(2)2}$, obtained by setting $r = \tilde{r}_2$ in (11).

Pay attention to the system represented by the first equation in (13), namely

(13') $$\begin{bmatrix} -\tilde{r}_2 \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -\tilde{r}_2 \mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{(1)2} \\ \mathbf{a}_{(2)2} \end{bmatrix} = \mathbf{0} .$$

Premultiplying both members of (13') by the matrix

(14) $$\begin{bmatrix} \tilde{r}_2 \mathbf{I}_{p_1} & \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \\ \mathbf{O}_{(p_2, p_1)} & (1/\tilde{r}_2)\, \mathbf{V}_{22}^{-1} \end{bmatrix},$$

we get the system

(15) $$(-\tilde{r}_2^2 \mathbf{I}_{p_1} + \mathbf{V}_{11}^{-1} \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21}) \mathbf{a}_{(1)2} = \mathbf{0} \quad , \quad \mathbf{a}_{(2)2} = \frac{1}{\tilde{r}_2} \mathbf{V}_{22}^{-1} \mathbf{V}_{21} \mathbf{a}_{(1)2}$$

and, as the matrix (14) is nonsingular, the systems (13') and (15) are equivalent.

Clearly, the first equation in (15) admits the eigenvector $\overline{\mathbf{a}}_{(1)2}$ such that $\overline{\mathbf{a}}'_{(1)2} \mathbf{V}_{11} \overline{\mathbf{a}}_{(1)2} = 1$ and $\overline{\mathbf{a}}'_{(1)2} \mathbf{V}_{11} \widetilde{\mathbf{a}}_{(1)1} = 0$, corresponding to the eigenvalue $\tilde{r}_2^2$.

In turn, the second equation in (15), for $\mathbf{a}_{(1)2} = \overline{\mathbf{a}}_{(1)2}$, gives the vector

$$\overline{\mathbf{a}}_{(2)2} = \frac{1}{\tilde{r}_2} \mathbf{V}_{22}^{-1} \mathbf{V}_{21} \overline{\mathbf{a}}_{(1)2}$$

such that $\overline{\mathbf{a}}'_{(2)2} \mathbf{V}_{22} \overline{\mathbf{a}}_{(2)2} = 1$ and $\overline{\mathbf{a}}'_{(2)2} \mathbf{V}_{22} \widetilde{\mathbf{a}}_{(2)1} = 0$.

In fact, taking into account that

$$\{(-\tilde{r}_2^2 \mathbf{I}_{p_1} + \mathbf{V}_{11}^{-1} \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21}) \overline{\mathbf{a}}_{(1)2} = \mathbf{0}\} \Leftrightarrow \{\mathbf{V}_{11}^{-1} \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21} \overline{\mathbf{a}}_{(1)2} = \tilde{r}_2^2\, \overline{\mathbf{a}}_{(1)2}\}$$
$$\Leftrightarrow \{\overline{\mathbf{a}}'_{(1)2} \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21} \overline{\mathbf{a}}_{(1)2} = \tilde{r}_2^2\} ,$$

we have

$$\bar{\mathbf{a}}'_{(2)2}\mathbf{V}_{22}\,\bar{\mathbf{a}}_{(2)2} = \frac{1}{\tilde{r}_2^2}\,\bar{\mathbf{a}}'_{(1)2}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{22}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\bar{\mathbf{a}}_{(1)2} = \frac{1}{\tilde{r}_2^2}\,\bar{\mathbf{a}}'_{(1)2}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\bar{\mathbf{a}}_{(1)2}$$
$$= \frac{1}{\tilde{r}_2^2}\,\tilde{r}_2^2 \qquad = 1 .$$

Further, since $\mathbf{V}_{12}\,\tilde{\mathbf{a}}_{(2)1} = \tilde{r}_1\,\mathbf{V}_{11}\,\tilde{\mathbf{a}}_{(1)1}$ (Section 3.1.1), we get

$$\bar{\mathbf{a}}'_{(2)2}\mathbf{V}_{22}\,\tilde{\mathbf{a}}_{(2)1} = \frac{1}{\tilde{r}_2}\,\bar{\mathbf{a}}'_{(1)2}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{22}\,\tilde{\mathbf{a}}_{(2)1} = \frac{1}{\tilde{r}_2}\,\bar{\mathbf{a}}'_{(1)2}\mathbf{V}_{12}\,\tilde{\mathbf{a}}_{(2)1}$$
$$= \frac{\tilde{r}_1}{\tilde{r}_2}\,\bar{\mathbf{a}}'_{(1)2}\mathbf{V}_{11}\,\tilde{\mathbf{a}}_{(1)1} \qquad = 0 .$$

Thus, we immediately conclude that $\bar{\mathbf{a}}_{(1)2}, \bar{\mathbf{a}}_{(2)2}$, solutions of the system (15) with the properties mentioned above, are also solutions of the system (13).

Of course, $\tilde{\mathbf{a}}_{(1)2} = \bar{\mathbf{a}}_{(1)2}, \tilde{\mathbf{a}}_{(2)2} = \bar{\mathbf{a}}_{(2)2}$ represent a solution of the problem set in (9) and (10).

Finally, it can easily be shown that

$$\tilde{\mathbf{a}}'_{(1)1}\mathbf{V}_{12}\,\tilde{\mathbf{a}}_{(2)2} = \tilde{\mathbf{a}}'_{(2)1}\mathbf{V}_{21}\,\tilde{\mathbf{a}}_{(1)2} = 0 .$$

The vectors $\tilde{\mathbf{a}}_{(1)2}$ and $\tilde{\mathbf{a}}_{(2)2}$, such that

$$\tilde{\mathbf{a}}'_{(1)2}\mathbf{V}_{11}\,\tilde{\mathbf{a}}_{(1)2} = 1 \;,\; \tilde{\mathbf{a}}'_{(2)2}\mathbf{V}_{22}\,\tilde{\mathbf{a}}_{(2)2} = 1 ,$$
$$\tilde{\mathbf{a}}'_{(1)2}\mathbf{V}_{11}\,\tilde{\mathbf{a}}_{(1)1} = \tilde{\mathbf{a}}'_{(2)2}\mathbf{V}_{22}\,\tilde{\mathbf{a}}_{(2)1} = 0 \quad , \quad \tilde{\mathbf{a}}'_{(1)1}\mathbf{V}_{12}\,\tilde{\mathbf{a}}_{(2)2} = \tilde{\mathbf{a}}'_{(2)1}\mathbf{V}_{21}\,\tilde{\mathbf{a}}_{(1)2} = 0 ,$$
$$\tilde{\mathbf{a}}'_{(1)2}\mathbf{V}_{12}\,\tilde{\mathbf{a}}_{(2)2} = \tilde{\mathbf{a}}'_{(2)2}\mathbf{V}_{21}\,\tilde{\mathbf{a}}_{(1)2} = \tilde{r}_2 ,$$

are called (the second two) *canonical factors*.

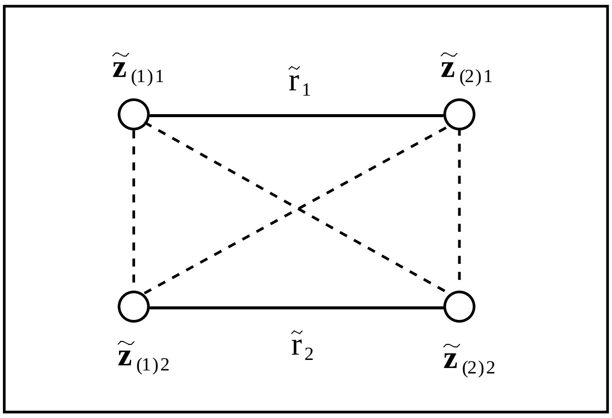
The vectors $\tilde{\mathbf{z}}_{(1)2} = \mathbf{Y}_1\tilde{\mathbf{a}}_{(1)2}$ and $\tilde{\mathbf{z}}_{(2)2} = \mathbf{Y}_2\tilde{\mathbf{a}}_{(2)2}$, such that

$$\tilde{\mathbf{z}}'_{(1)2}\mathbf{M}\tilde{\mathbf{z}}_{(1)2} = 1 \;,\; \tilde{\mathbf{z}}'_{(2)2}\mathbf{M}\tilde{\mathbf{z}}_{(2)2} = 1 ,$$
$$\tilde{\mathbf{z}}'_{(1)2}\mathbf{M}\tilde{\mathbf{z}}_{(1)1} = \tilde{\mathbf{z}}'_{(2)2}\mathbf{M}\tilde{\mathbf{z}}_{(2)1} = 0 \quad , \quad \tilde{\mathbf{z}}'_{(1)1}\mathbf{M}\tilde{\mathbf{z}}_{(2)2} = \tilde{\mathbf{z}}'_{(2)1}\mathbf{M}\tilde{\mathbf{z}}_{(1)2} = 0 ,$$
$$\tilde{\mathbf{z}}'_{(1)2}\mathbf{M}\tilde{\mathbf{z}}_{(2)2} = \tilde{\mathbf{z}}'_{(2)2}\mathbf{M}\tilde{\mathbf{z}}_{(1)2} = \tilde{r}_2 ,$$

are called (the second two) *canonical variables* or *canonical vectors*.

The cosine of the angle formed by $\tilde{\mathbf{z}}_{(1)2}$ and $\tilde{\mathbf{z}}_{(2)2}$, namely $\tilde{r}_2$, is called (the second one) *canonical correlation coefficient* (Fig. 2).

Fig. 2



3.1.3 THE FOLLOWING STEPS

The procedure described in the preceding pages may be iterated for $s = 3, \dots, k$.

At the sth step the problem lies in finding a linear combination $\tilde{\mathbf{z}}_{(1)s}$ of $\mathbf{y}_1, \dots, \mathbf{y}_{p_1}$, orthogonal to the subspace spanned by $\tilde{\mathbf{z}}_{(1)1}, \dots, \tilde{\mathbf{z}}_{(1)s-1}$, and a linear combination $\tilde{\mathbf{z}}_{(2)s}$ of $\mathbf{y}_{p_1+1}, \dots, \mathbf{y}_{p_1+p_2}$, orthogonal to the subspace spanned by $\tilde{\mathbf{z}}_{(2)1}, \dots, \tilde{\mathbf{z}}_{(2)s-1}$, such that the cosine of the angle they form (the linear correlation coefficient) $\cos(\tilde{\mathbf{z}}_{(1)s}, \tilde{\mathbf{z}}_{(2)s})$ is a maximum.

Setting

$$\mathbf{z}_{(1)s} = \mathbf{Y}_1 \mathbf{a}_{(1)s} \quad , \quad \mathbf{z}_{(2)s} = \mathbf{Y}_2 \mathbf{a}_{(2)s}$$

and

$$\cos(\mathbf{z}_{(1)s}, \mathbf{z}_{(2)s}) = \frac{\mathbf{a}'_{(1)s} \mathbf{V}_{12} \mathbf{a}_{(2)s}}{\{(\mathbf{a}'_{(1)s} \mathbf{V}_{11} \mathbf{a}_{(1)s})(\mathbf{a}'_{(2)s} \mathbf{V}_{22} \mathbf{a}_{(2)s})\}^{1/2}} ,$$

we have to find out

$$\text{(16)} \qquad \underset{\mathbf{a}_{(1)s}, \mathbf{a}_{(2)s}}{\text{Max}} \frac{\mathbf{a}'_{(1)s} \mathbf{V}_{12} \mathbf{a}_{(2)s}}{\{(\mathbf{a}'_{(1)s} \mathbf{V}_{11} \mathbf{a}_{(1)s})(\mathbf{a}'_{(2)s} \mathbf{V}_{22} \mathbf{a}_{(2)s})\}^{1/2}}$$

under the constraints ($3 \leq s \leq k$; $t = 1, \ldots, s-1$)

$$\mathbf{z}'_{(1)s}\mathbf{M}\widetilde{\mathbf{z}}_{(1)t} = \mathbf{a}'_{(1)s}\mathbf{Y}'_1\mathbf{M}\mathbf{Y}_1\widetilde{\mathbf{a}}_{(1)t} = \mathbf{a}'_{(1)s}\mathbf{V}_{11}\widetilde{\mathbf{a}}_{(1)t} = 0$$

(17)

$$\mathbf{z}'_{(2)s}\mathbf{M}\widetilde{\mathbf{z}}_{(2)t} = \mathbf{a}'_{(2)s}\mathbf{Y}'_2\mathbf{M}\mathbf{Y}_2\widetilde{\mathbf{a}}_{(2)t} = \mathbf{a}'_{(2)s}\mathbf{V}_{22}\widetilde{\mathbf{a}}_{(2)t} = 0\,.$$

Equivalently, assuming that $\mathbf{z}_{(1)s}$ and $\mathbf{z}_{(2)s}$ are vectors of unitary square length, we have to look for

(16') $$\underset{\mathbf{a}_{(1)s},\,\mathbf{a}_{(2)s}}{\text{Max}}\ \mathbf{a}'_{(1)s}\mathbf{V}_{12}\mathbf{a}_{(2)s} \quad , \quad \mathbf{a}'_{(1)s}\mathbf{V}_{11}\mathbf{a}_{(1)s} = 1 \quad , \quad \mathbf{a}'_{(2)s}\mathbf{V}_{22}\mathbf{a}_{(2)s} = 1$$

under the constraints in (17).

Solving the problem of maximization set in (16') and (17) by the Lagrange method – since at the maximum it must be [8]

$$\begin{array}{lcl} \mathbf{V}_{12}\widetilde{\mathbf{a}}_{(2)s} = \cos(\widetilde{\mathbf{z}}_{(1)s},\widetilde{\mathbf{z}}_{(2)s})\,\mathbf{V}_{11}\widetilde{\mathbf{a}}_{(1)s} & , & \mathbf{V}_{21}\widetilde{\mathbf{a}}_{(1)s} = \cos(\widetilde{\mathbf{z}}_{(1)s},\widetilde{\mathbf{z}}_{(2)s})\,\mathbf{V}_{22}\widetilde{\mathbf{a}}_{(2)s} \\ \widetilde{\mathbf{a}}'_{(1)s}\mathbf{V}_{11}\widetilde{\mathbf{a}}_{(1)s} = 1 & , & \widetilde{\mathbf{a}}'_{(2)s}\mathbf{V}_{22}\widetilde{\mathbf{a}}_{(2)s} = 1 \\ \widetilde{\mathbf{a}}'_{(1)s}\mathbf{V}_{11}\widetilde{\mathbf{a}}_{(1)t} = 0 & , & \widetilde{\mathbf{a}}'_{(2)s}\mathbf{V}_{22}\widetilde{\mathbf{a}}_{(2)t} = 0 \end{array}$$

– we realize that $\cos(\widetilde{\mathbf{z}}_{(1)s},\widetilde{\mathbf{z}}_{(2)s})$ and $\widetilde{\mathbf{a}}_{(1)s},\widetilde{\mathbf{a}}_{(2)s}$ has to be found among the solutions of the system

(18) $$\begin{bmatrix} -r\mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -r\mathbf{V}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{a}_{(1)s} \\ \mathbf{a}_{(2)s} \end{bmatrix} = \mathbf{0} \ , \ \begin{array}{lcl} \mathbf{a}'_{(1)s}\mathbf{V}_{11}\mathbf{a}_{(1)s} = 1 & , & \mathbf{a}'_{(2)s}\mathbf{V}_{22}\mathbf{a}_{(2)s} = 1 \\ \mathbf{a}'_{(1)s}\mathbf{V}_{11}\widetilde{\mathbf{a}}_{(1)t} = 0 & , & \mathbf{a}'_{(2)s}\mathbf{V}_{22}\widetilde{\mathbf{a}}_{(2)t} = 0 \end{array}$$

in the unknowns $r, \mathbf{a}_{(1)s}, \mathbf{a}_{(2)s}$.

To this end, consider the system

(19) $$\begin{bmatrix} -r\mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -r\mathbf{V}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{a}_{(1)s} \\ \mathbf{a}_{(2)s} \end{bmatrix} = \mathbf{0}$$

and ask for which values of r it admits non-trivial solutions with respect to

(8) Details are left to the reader.

the unknowns $\mathbf{a}_{(1)s}$, $\mathbf{a}_{(2)s}$ (9).

This system − similar to that one written in (3) − admits the solution $\tilde{r}_s > 0$ for the unknown r.

This solution can also be obtained as square root of the sth largest eigenvalue of the equation (5).

Now, consider the system

$$(20) \quad \begin{bmatrix} -\tilde{r}_s \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -\tilde{r}_s \mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{(1)s} \\ \mathbf{a}_{(2)s} \end{bmatrix} = \mathbf{0} \ , \quad \begin{matrix} \mathbf{a}'_{(1)s} \mathbf{V}_{11} \mathbf{a}_{(1)s} = 1 & , & \mathbf{a}'_{(2)s} \mathbf{V}_{22} \mathbf{a}_{(2)s} = 1 \\ \mathbf{a}'_{(1)s} \mathbf{V}_{11} \widetilde{\mathbf{a}}_{(1)t} = 0 & , & \mathbf{a}'_{(2)s} \mathbf{V}_{22} \widetilde{\mathbf{a}}_{(2)t} = 0 \end{matrix}$$

in the unknowns $\mathbf{a}_{(1)s}$, $\mathbf{a}_{(2)s}$, obtained by setting $r = \tilde{r}_s$ in (18).

Pay attention to the system represented by the first equation in (20), namely

$$(20') \quad \begin{bmatrix} -\tilde{r}_s \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -\tilde{r}_s \mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{(1)s} \\ \mathbf{a}_{(2)s} \end{bmatrix} = \mathbf{0} \ .$$

Premultiplying both members of (20') by the matrix

$$(21) \quad \begin{bmatrix} \tilde{r}_s \mathbf{I}_{p_1} & \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \\ \mathbf{O}_{(p_2, p_1)} & (1/\tilde{r}_s)\, \mathbf{V}_{22}^{-1} \end{bmatrix},$$

we get the system

$$(22) \quad (-\tilde{r}_s^2 \mathbf{I}_{p_1} + \mathbf{V}_{11}^{-1} \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21}) \mathbf{a}_{(1)s} = \mathbf{0} \quad , \quad \mathbf{a}_{(2)s} = \frac{1}{\tilde{r}_s} \mathbf{V}_{22}^{-1} \mathbf{V}_{21} \mathbf{a}_{(1)s}$$

and, as the matrix (21) is nonsingular, the systems (20') and (22) are equivalent.

Clearly, the first equation in (22) admits the eigenvector $\overline{\mathbf{a}}_{(1)s}$ such that $\overline{\mathbf{a}}'_{(1)s} \mathbf{V}_{11} \overline{\mathbf{a}}_{(1)s} = 1$ and $\overline{\mathbf{a}}'_{(1)s} \mathbf{V}_{11} \widetilde{\mathbf{a}}_{(1)t} = 0$, corresponding to the eigenvalue $\tilde{r}_s^2$.

In turn, the second equation in (22), for $\mathbf{a}_{(1)s} = \overline{\mathbf{a}}_{(1)s}$, gives the vector

(9) On account of the constraints of normalization in (17), it is necessary to consider only the non-trivial solutions of the system (19).

$$\bar{\mathbf{a}}_{(2)s} = \frac{1}{\tilde{r}_s} \mathbf{V}_{22}^{-1} \mathbf{V}_{21} \bar{\mathbf{a}}_{(1)s}$$

such that $\bar{\mathbf{a}}'_{(2)s} \mathbf{V}_{22} \bar{\mathbf{a}}_{(2)s} = 1$ and $\bar{\mathbf{a}}'_{(2)s} \mathbf{V}_{22} \tilde{\mathbf{a}}_{(2)t} = 0$.

Thus, we conclude that $\bar{\mathbf{a}}_{(1)s}, \bar{\mathbf{a}}_{(2)s}$, solutions of the system (22) with the properties mentioned above, are also solutions of the system (20).

Of course, $\tilde{\mathbf{a}}_{(1)s} = \bar{\mathbf{a}}_{(1)s}, \tilde{\mathbf{a}}_{(2)s} = \bar{\mathbf{a}}_{(2)s}$ represent a solution of the problem set in (16) and (17).

Finally, it can easily be shown that

$$\tilde{\mathbf{a}}'_{(1)t} \mathbf{V}_{12} \tilde{\mathbf{a}}_{(2)s} = \tilde{\mathbf{a}}'_{(2)t} \mathbf{V}_{21} \tilde{\mathbf{a}}_{(1)s} = 0 \ .$$

The vectors $\tilde{\mathbf{a}}_{(1)s}$ and $\tilde{\mathbf{a}}_{(2)s}$, such that

$$\begin{aligned}
&\tilde{\mathbf{a}}'_{(1)s} \mathbf{V}_{11} \tilde{\mathbf{a}}_{(1)s} = 1 \ , \ \tilde{\mathbf{a}}'_{(2)s} \mathbf{V}_{22} \tilde{\mathbf{a}}_{(2)s} = 1 \ , \\
&\tilde{\mathbf{a}}'_{(1)s} \mathbf{V}_{11} \tilde{\mathbf{a}}_{(1)t} = \tilde{\mathbf{a}}'_{(2)s} \mathbf{V}_{22} \tilde{\mathbf{a}}_{(2)t} = 0 \quad , \quad \tilde{\mathbf{a}}'_{(1)t} \mathbf{V}_{12} \tilde{\mathbf{a}}_{(2)s} = \tilde{\mathbf{a}}'_{(2)t} \mathbf{V}_{21} \tilde{\mathbf{a}}_{(1)s} = 0 \ , \\
&\tilde{\mathbf{a}}'_{(1)s} \mathbf{V}_{12} \tilde{\mathbf{a}}_{(2)s} = \tilde{\mathbf{a}}'_{(2)s} \mathbf{V}_{21} \tilde{\mathbf{a}}_{(1)s} = \tilde{r}_s \ ,
\end{aligned}$$

are called (the sth two) *canonical factors*.

The vectors $\tilde{\mathbf{z}}_{(1)s} = \mathbf{Y}_1 \tilde{\mathbf{a}}_{(1)s}$ and $\tilde{\mathbf{z}}_{(2)s} = \mathbf{Y}_2 \tilde{\mathbf{a}}_{(2)s}$, such that

$$\begin{aligned}
&\tilde{\mathbf{z}}'_{(1)s} \mathbf{M} \tilde{\mathbf{z}}_{(1)s} = 1 \ , \ \tilde{\mathbf{z}}'_{(2)s} \mathbf{M} \tilde{\mathbf{z}}_{(2)s} = 1 \ , \\
&\tilde{\mathbf{z}}'_{(1)s} \mathbf{M} \tilde{\mathbf{z}}_{(1)t} = \tilde{\mathbf{z}}'_{(2)s} \mathbf{M} \tilde{\mathbf{z}}_{(2)t} = 0 \quad , \quad \tilde{\mathbf{z}}'_{(1)t} \mathbf{M} \tilde{\mathbf{z}}_{(2)s} = \tilde{\mathbf{z}}'_{(2)t} \mathbf{M} \tilde{\mathbf{z}}_{(1)s} = 0 \ , \\
&\tilde{\mathbf{z}}'_{(1)s} \mathbf{M} \tilde{\mathbf{z}}_{(2)s} = \tilde{\mathbf{z}}'_{(2)s} \mathbf{M} \tilde{\mathbf{z}}_{(1)s} = \tilde{r}_s \ ,
\end{aligned}$$

are called (the sth two) *canonical variables* or *canonical vectors*.

The cosine of the angle formed by $\tilde{\mathbf{z}}_{(1)s}$ and $\tilde{\mathbf{z}}_{(2)s}$, namely $\tilde{r}_s$, is called (the sth one) *canonical correlation coefficient*.

3.2 FUNDAMENTAL PROPERTIES OF CCA

Writing

$$\tilde{\mathbf{A}}_{(1)} = [\, \tilde{\mathbf{a}}_{(1)1} \cdots \tilde{\mathbf{a}}_{(1)k} \,] \ , \ \tilde{\mathbf{A}}_{(2)} = [\, \tilde{\mathbf{a}}_{(2)1} \cdots \tilde{\mathbf{a}}_{(2)k} \,] \ , \ \tilde{\mathbf{R}} = \mathrm{diag}\,(\tilde{r}_1, \ldots, \tilde{r}_k)$$

and

$$\widetilde{\mathbf{Z}}_{(1)} = [\,\widetilde{\mathbf{z}}_{(1)1} \;\cdots\; \widetilde{\mathbf{z}}_{(1)k}\,] = \mathbf{Y}_1 \widetilde{\mathbf{A}}_{(1)} \quad , \quad \widetilde{\mathbf{Z}}_{(2)} = [\,\widetilde{\mathbf{z}}_{(2)1} \;\cdots\; \widetilde{\mathbf{z}}_{(2)k}\,] = \mathbf{Y}_2 \widetilde{\mathbf{A}}_{(2)} \quad ,$$

some fundamental properties of CCA can be pointed out.

1. We have

(23) $$\widetilde{\mathbf{Z}}'_{(1)}\mathbf{M}\widetilde{\mathbf{Z}}_{(1)} = \widetilde{\mathbf{A}}'_{(1)}\mathbf{V}_{11}\widetilde{\mathbf{A}}_{(1)} = \mathbf{I}_k \quad , \quad \widetilde{\mathbf{Z}}'_{(2)}\mathbf{M}\widetilde{\mathbf{Z}}_{(2)} = \widetilde{\mathbf{A}}'_{(2)}\mathbf{V}_{22}\widetilde{\mathbf{A}}_{(2)} = \mathbf{I}_k \ .$$

In other words, the canonical variables $\widetilde{\mathbf{z}}_{(1)1}, \dots, \widetilde{\mathbf{z}}_{(1)k}$ of the first set are uncorrelated and with unitary variance; the same is true for the canonical variables $\widetilde{\mathbf{z}}_{(2)1}, \dots, \widetilde{\mathbf{z}}_{(2)k}$ of the second set.

2. We have

(24) $$\widetilde{\mathbf{Z}}'_{(1)}\mathbf{M}\widetilde{\mathbf{Z}}_{(2)} = \widetilde{\mathbf{A}}'_{(1)}\mathbf{V}_{12}\widetilde{\mathbf{A}}_{(2)} = \widetilde{\mathbf{R}} \ .$$

Namely, each canonical variable of the first set presents correlation $\tilde{r}_h > 0$ $(h = 1, \dots, k)$ with the corresponding canonical variable of the second set, while each of the remaining canonical variables of the first set is uncorrelated with each of the remaining canonical variables of the second set.

REMARK 1. Since canonical variables and canonical correlation coefficients are, as can easily be shown, invariant with respect to scale changes, CCA is very often performed after standardization of each original variable, which leads us to work with correlation matrices rather than with covariance matrices.

REMARK 2. In order to compute the canonical correlation coefficients $\tilde{r}_1, \dots, \tilde{r}_k$, instead of the equation

(*i*) $$\det(-r^2\,\mathbf{I}_{p_1} + \mathbf{V}_{11}^{-1}\mathbf{V}_{12}\,\mathbf{V}_{22}^{-1}\,\mathbf{V}_{21}) = 0 \ ,$$

we could employ the equation

(*ii*) $$\det(-r^2\,\mathbf{I}_{p_2} + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\,\mathbf{V}_{11}^{-1}\,\mathbf{V}_{12}) = 0$$

where $\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\,\mathbf{V}_{11}^{-1}\,\mathbf{V}_{12}$ may be interpreted as the matrix of a selfadjoint transformation in the metric represented by $\mathbf{V}_{22}$ $((\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12})'\mathbf{V}_{22} = \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} = \mathbf{V}_{22}(\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}))$.

Analogously, to the end of computing the canonical factors $\tilde{\mathbf{a}}_{(1)1}, \dots, \tilde{\mathbf{a}}_{(1)k}$ and $\tilde{\mathbf{a}}_{(2)1}, \dots, \tilde{\mathbf{a}}_{(2)k}$, instead of the equations $(h = 1, \dots, k)$

$$(iii) \quad (-\tilde{r}_h^2 \mathbf{I}_{p_1} + \mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{a}_{(1)h} = \mathbf{0} \quad , \quad \mathbf{a}_{(2)h} = \frac{1}{\tilde{r}_h}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{a}_{(1)h} \, .$$

we could employ the equations $(h = 1, \dots, k)$

$$(iv) \quad (-\tilde{r}_h^2 \mathbf{I}_{p_2} + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12})\mathbf{a}_{(2)h} = \mathbf{0} \quad , \quad \mathbf{a}_{(2)h} = \frac{1}{\tilde{r}_h}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{a}_{(2)h} \, .$$

However, from the computation point of view, employing (*ii*) and (*iv*) instead of (*i*) and (*iii*) is the same if $p_1 = p_2$, and is not convenient if $p_1 < p_2$.

REMARK 3. Since $(h = 1, \dots, k)$

$$\mathbf{V}_{12}\tilde{\mathbf{a}}_{(2)h} = \tilde{r}_h \mathbf{V}_{11}\tilde{\mathbf{a}}_{(1)h} \, ,$$

we have

$$\{\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_2\tilde{\mathbf{a}}_{(2)h} = \tilde{r}_h \mathbf{Y}_1'\mathbf{M}\mathbf{Y}_1\tilde{\mathbf{a}}_{(1)h}\} \Leftrightarrow \{\mathbf{Y}_1(\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_1)^{-1}\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_2\tilde{\mathbf{a}}_{(2)h} = \tilde{r}_h \mathbf{Y}_1\tilde{\mathbf{a}}_{(1)h}\}$$
$$\Leftrightarrow \{\mathbf{P}_1\tilde{\mathbf{z}}_{(2)h} = \tilde{r}_h \tilde{\mathbf{z}}_{(1)h}\}$$

where $\mathbf{P}_1 = \mathbf{Y}_1(\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_1)^{-1}\mathbf{Y}_1'\mathbf{M}$ represents the orthogonal projection matrix on the subspace $S(\mathbf{Y}_1)$ of R^n spanned by the column vectors of $\mathbf{Y}_1$.

Analogously, since $(h = 1, \dots, k)$

$$\mathbf{V}_{21}\tilde{\mathbf{a}}_{(1)h} = \tilde{r}_h \mathbf{V}_{22}\tilde{\mathbf{a}}_{(2)h} \, ,$$

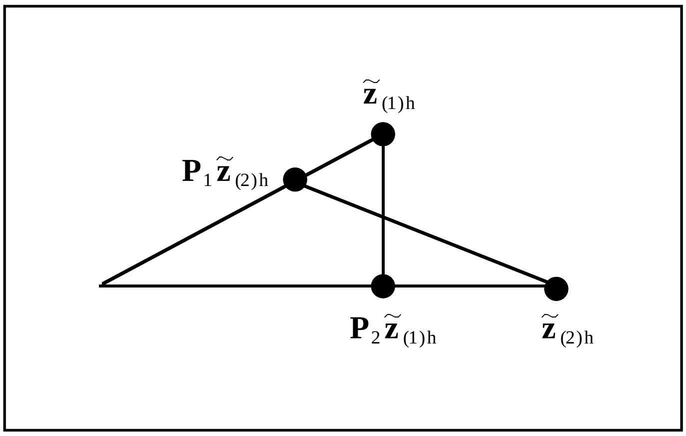
we have

$$\{\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_1\tilde{\mathbf{a}}_{(1)h} = \tilde{r}_h \mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2\tilde{\mathbf{a}}_{(2)h}\} \Leftrightarrow \{\mathbf{Y}_2(\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2)^{-1}\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_1\tilde{\mathbf{a}}_{(1)h} = \tilde{r}_h \mathbf{Y}_2\tilde{\mathbf{a}}_{(2)h}\}$$
$$\Leftrightarrow \{\mathbf{P}_2\tilde{\mathbf{z}}_{(1)h} = \tilde{r}_h \tilde{\mathbf{z}}_{(2)h}\}$$

where $\mathbf{P}_2 = \mathbf{Y}_2(\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2)^{-1}\mathbf{Y}_2'\mathbf{M}$ represents the orthogonal projection matrix on the subspace $S(\mathbf{Y}_2)$ of R^n spanned by the column vectors of $\mathbf{Y}_2$.

Hence, the orthogonal projection of $\tilde{\mathbf{z}}_{(2)h}$ on $S(\mathbf{Y}_1)$ is homothetic with $\tilde{\mathbf{z}}_{(1)h}$ and, similarly, the orthogonal projection of $\tilde{\mathbf{z}}_{(1)h}$ on $S(\mathbf{Y}_2)$ is homothetic with $\tilde{\mathbf{z}}_{(2)h}$ (Fig. 3).

Fig. 3



REMARK 4. Taking into account what was mentioned above, is immediately apparent that we can write

$$\mathbf{P}_1\mathbf{P}_2\tilde{\mathbf{z}}_{(1)h} = \tilde{r}_h^2\,\tilde{\mathbf{z}}_{(1)h} \quad , \quad \mathbf{P}_2\mathbf{P}_1\tilde{\mathbf{z}}_{(2)h} = \tilde{r}_h^2\,\tilde{\mathbf{z}}_{(2)h} \; .$$

Thus, square canonical correlation coefficients and canonical variables may also be interpreted as eigenvalues and eigenvectors of the linear transformations corresponding to the matrices $\mathbf{P}_1\mathbf{P}_2$ and $\mathbf{P}_2\mathbf{P}_1$ [(10)].

REMARK 5. Some particular aspects of CCA should be noticed.

a. If $p_1 = 1 = p_2$, the equation (5) can be written as

$$\det(-r^2 + \frac{\sigma_{12}^2}{\sigma_1^2\,\sigma_2^2}) = 0 \; .$$

Then,

$$\tilde{r}_1^2 = \frac{\sigma_{12}^2}{\sigma_1^2\,\sigma_2^2} \, ,$$

(10) As can easily be verified, the restrictions of these linear transformations, respectively, to $S(\mathbf{Y}_1)$ and $S(\mathbf{Y}_2)$ are selfadjoint.

the square linear correlation coefficient between the variables $\mathbf{y}_1$ and $\mathbf{y}_2$.

b. If $p_1 = 1$ and $p_2 > 1$, the equation (5) can be written as

$$\det(-r^2 + \frac{\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}}{\sigma_1^2}) = 0 \ .$$

Then,

$$\tilde{r}_1^2 = \frac{\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}}{\sigma_1^2} \ ,$$

the square multiple linear correlation coefficient between the variables $\mathbf{y}_1$ and $\mathbf{y}_2 , \dots , \mathbf{y}_{1+p_2}$.

c. Denote by $\rho(\tilde{\mathbf{z}}_{(1)h}, \mathbf{Y}_2)$ $(h = 1, \dots, k)$ the square multiple linear correlation coefficient between the variables $\tilde{\mathbf{z}}_{(1)h} = \mathbf{Y}_1 \tilde{\mathbf{a}}_{(1)h}$ and $\mathbf{y}_{p_1+1}, \dots , \mathbf{y}_{p_1+p_2}$.

Moreover, denote by

$$\hat{\tilde{\mathbf{z}}}_{(1)h\,,\,\mathbf{Y}_2} = \mathbf{Y}_2(\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2)^{-1}\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_1\tilde{\mathbf{a}}_{(1)h}$$

the orthogonal projection of $\tilde{\mathbf{z}}_{(1)h}$ on the subspace spanned by the variables $\mathbf{y}_{p_1+1}, \dots , \mathbf{y}_{p_1+p_2}$.

Then, as can easily be verified, it results

$$\rho(\tilde{\mathbf{z}}_{(1)h}, \mathbf{Y}_2) = \cos^2(\tilde{\mathbf{z}}_{(1)h}, \hat{\tilde{\mathbf{z}}}_{(1)h\,,\,\mathbf{Y}_2}) = \tilde{\mathbf{a}}'_{(1)h}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\tilde{\mathbf{a}}_{(1)h} = \tilde{r}_h^2 \ .$$

Analogously, denote by $\rho(\tilde{\mathbf{z}}_{(2)h}, \mathbf{Y}_1)$ $(h = 1, \dots, k)$ the square multiple linear correlation coefficient between the variables $\tilde{\mathbf{z}}_{(2)h} = \mathbf{Y}_2 \tilde{\mathbf{a}}_{(2)h}$ and $\mathbf{y}_1, \dots , \mathbf{y}_{p_1}$.

Moreover, denote by

$$\hat{\tilde{\mathbf{z}}}_{(2)h\,,\,\mathbf{Y}_1} = \mathbf{Y}_1(\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_1)^{-1}\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_2\tilde{\mathbf{a}}_{(2)h}$$

the orthogonal projection of $\tilde{\mathbf{z}}_{(2)h}$ on the subspace spanned by the variables $\mathbf{y}_1, \dots , \mathbf{y}_{p_1}$.

Then, it results

$$\rho(\tilde{\mathbf{z}}_{(2)h}, \mathbf{Y}_1) = \cos^2(\tilde{\mathbf{z}}_{(2)h}, \hat{\tilde{\mathbf{z}}}_{(2)h\,,\,\mathbf{Y}_1}) = \tilde{\mathbf{a}}'_{(2)h}\mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\tilde{\mathbf{a}}_{(2)h} = \tilde{r}_h^2 \ .$$

REMARK 6. Assuming that $k < p_1 \le p_2$, the k canonical variables

$$\widetilde{\mathbf{z}}_{(1)1}, \dots, \widetilde{\mathbf{z}}_{(1)k}$$

form an orthonormal basis of a (proper) subspace of the space $S(\mathbf{Y}_1)$ (of dimension p_1) spanned by the column vectors of $\mathbf{Y}_1$ and, analogously, the k canonical variables

$$\widetilde{\mathbf{z}}_{(2)1}, \dots, \widetilde{\mathbf{z}}_{(2)k}$$

form an orthonormal basis of a (proper) subspace of the space $S(\mathbf{Y}_2)$ (of dimension p_2) spanned by the column vectors of $\mathbf{Y}_2$.

In order to complete these bases, we can proceed as follows.

Firstly, find $p_1 - k$ *canonical factors* $\widetilde{\mathbf{a}}_{(1)k+1}, \dots, \widetilde{\mathbf{a}}_{(1)p_1}$ and $p_2 - k$ *canonical factors* $\widetilde{\mathbf{a}}_{(2)k+1}, \dots, \widetilde{\mathbf{a}}_{(2)p_2}$ – which are solutions, respectively, of the equations $(u = k+1, \dots, p_1;\ v = k+1, \dots, p_2)$

$$\mathbf{V}_{21}\mathbf{a}_{(1)u} = \mathbf{0} \ \text{ or } \ (\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{a}_{(1)u} = \mathbf{0}$$

and

$$\mathbf{V}_{12}\mathbf{a}_{(2)v} = \mathbf{0} \ \text{ or } \ (\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12})\mathbf{a}_{(2)v} = \mathbf{0}$$

– such that, setting

$$\widetilde{\mathbf{A}}_{(1)+} = \left[\, \widetilde{\mathbf{a}}_{(1)k+1} \ \cdots \ \widetilde{\mathbf{a}}_{(1)p_1} \right] \ , \ \widetilde{\mathbf{A}}_{(2)+} = \left[\, \widetilde{\mathbf{a}}_{(2)k+1} \ \cdots \ \widetilde{\mathbf{a}}_{(2)p_2} \right],$$

we have ($\mathbf{I}$, $\mathbf{O}$ of appropriate order)

$$\begin{aligned}
&\widetilde{\mathbf{A}}'_{(1)+}\mathbf{V}_{11}\widetilde{\mathbf{A}}_{(1)+} = \mathbf{I} \quad , \quad \widetilde{\mathbf{A}}'_{(2)+}\mathbf{V}_{22}\widetilde{\mathbf{A}}_{(2)+} = \mathbf{I}\\
&\widetilde{\mathbf{A}}'_{(1)}\mathbf{V}_{12}\widetilde{\mathbf{A}}_{(2)+} = \mathbf{O} \quad , \quad \widetilde{\mathbf{A}}'_{(2)}\mathbf{V}_{21}\widetilde{\mathbf{A}}_{(1)+} = \mathbf{O}\\
&\widetilde{\mathbf{A}}'_{(2)+}\mathbf{V}_{21}\widetilde{\mathbf{A}}_{(1)+} = \mathbf{O} \quad , \quad \widetilde{\mathbf{A}}'_{(1)}\mathbf{V}_{11}\widetilde{\mathbf{A}}_{(1)+} = \mathbf{O}\\
&\widetilde{\mathbf{A}}'_{(2)}\mathbf{V}_{22}\widetilde{\mathbf{A}}_{(2)+} = \mathbf{O}\,.
\end{aligned}$$

Successively, define $p_1 - k$ *canonical variables*

$$\widetilde{\mathbf{z}}_{(1)k+1} = \mathbf{Y}_1\widetilde{\mathbf{a}}_{(1)k+1}, \dots, \widetilde{\mathbf{z}}_{(1)p_1} = \mathbf{Y}_1\widetilde{\mathbf{a}}_{(1)p_1}$$

and $p_2 - k$ *canonical variables*

$$\widetilde{\mathbf{z}}_{(2)k+1} = \mathbf{Y}_2 \widetilde{\mathbf{a}}_{(2)k+1}, \dots, \widetilde{\mathbf{z}}_{(2)p_2} = \mathbf{Y}_2 \widetilde{\mathbf{a}}_{(2)p_2}$$

such that, setting

$$\widetilde{\mathbf{Z}}_{(1)+} = \left[\, \widetilde{\mathbf{z}}_{(1)k+1} \ \cdots \ \widetilde{\mathbf{z}}_{(1)p_1} \right] \ , \ \widetilde{\mathbf{Z}}_{(2)+} = \left[\, \widetilde{\mathbf{z}}_{(2)k+1} \ \cdots \ \widetilde{\mathbf{z}}_{(2)p_2} \right],$$

we have ($\mathbf{I}$, $\mathbf{O}$ of appropriate order)

$$\begin{aligned}
&\widetilde{\mathbf{Z}}'_{(1)+}\mathbf{M}\widetilde{\mathbf{Z}}_{(1)+} = \mathbf{I} \quad , \quad \widetilde{\mathbf{Z}}'_{(2)+}\mathbf{M}\widetilde{\mathbf{Z}}_{(2)+} = \mathbf{I} \\
&\widetilde{\mathbf{Z}}'_{(1)}\mathbf{M}\widetilde{\mathbf{Z}}_{(2)+} = \mathbf{O} \quad , \quad \widetilde{\mathbf{Z}}'_{(2)}\mathbf{M}\widetilde{\mathbf{Z}}_{(1)+} = \mathbf{O} \\
&\widetilde{\mathbf{Z}}'_{(2)+}\mathbf{M}\widetilde{\mathbf{Z}}_{(1)+} = \mathbf{O} \quad , \quad \widetilde{\mathbf{Z}}'_{(1)}\mathbf{M}\widetilde{\mathbf{Z}}_{(1)+} = \mathbf{O} \\
&\widetilde{\mathbf{Z}}'_{(2)}\mathbf{M}\widetilde{\mathbf{Z}}_{(2)+} = \mathbf{O}\,.
\end{aligned}$$

Notice that, writing

$$\begin{aligned}
&\widetilde{\mathbf{A}}_{(1)++} = \left[\, \widetilde{\mathbf{A}}_{(1)} \quad \widetilde{\mathbf{A}}_{(1)+} \right] \ , \ \widetilde{\mathbf{A}}_{(2)++} = \left[\, \widetilde{\mathbf{A}}_{(2)} \quad \widetilde{\mathbf{A}}_{(2)+} \right] , \\
&\widetilde{\mathbf{Z}}_{(1)++} = \left[\, \widetilde{\mathbf{Z}}_{(1)} \quad \widetilde{\mathbf{Z}}_{(1)+} \right] \ , \ \widetilde{\mathbf{Z}}_{(2)++} = \left[\, \widetilde{\mathbf{Z}}_{(2)} \quad \widetilde{\mathbf{Z}}_{(2)+} \right] ,
\end{aligned}$$

it results ($\mathbf{I}$, $\mathbf{O}$ of appropriate order)

$$\begin{aligned}
&\widetilde{\mathbf{A}}'_{(1)++}\mathbf{V}_{11}\widetilde{\mathbf{A}}_{(1)++} = \mathbf{I} \quad , \quad \widetilde{\mathbf{A}}'_{(2)++}\mathbf{V}_{22}\widetilde{\mathbf{A}}_{(2)++} = \mathbf{I} \\
&\widetilde{\mathbf{A}}'_{(1)++}\mathbf{V}_{12}\widetilde{\mathbf{A}}_{(2)++} = \operatorname{diag}(\widetilde{\mathbf{R}}, \mathbf{O})
\end{aligned}$$

and

$$\begin{aligned}
&\widetilde{\mathbf{Z}}'_{(1)++}\mathbf{M}\widetilde{\mathbf{Z}}_{(1)++} = \mathbf{I} \quad , \quad \widetilde{\mathbf{Z}}'_{(2)++}\mathbf{M}\widetilde{\mathbf{Z}}_{(2)++} = \mathbf{I} \\
&\widetilde{\mathbf{Z}}'_{(1)++}\mathbf{M}\widetilde{\mathbf{Z}}_{(2)++} = \operatorname{diag}(\widetilde{\mathbf{R}}, \mathbf{O})\,.
\end{aligned}$$

REMARK 7. As can easily be verified, from the relations

$$\widetilde{\mathbf{Z}}_{(1)++} = \mathbf{Y}_1 \widetilde{\mathbf{A}}_{(1)++} \quad , \quad \widetilde{\mathbf{Z}}_{(2)++} = \mathbf{Y}_2 \widetilde{\mathbf{A}}_{(2)++} ,$$

we get the so-called *reconstitution formulas*

$$\mathbf{Y}_1 = \widetilde{\mathbf{Z}}_{(1)}\widetilde{\mathbf{A}}'_{(1)}\mathbf{V}_{11} + \widetilde{\mathbf{Z}}_{(1)+}\widetilde{\mathbf{A}}'_{(1)+}\mathbf{V}_{11} \quad , \quad \mathbf{Y}_2 = \widetilde{\mathbf{Z}}_{(2)}\widetilde{\mathbf{A}}'_{(2)}\mathbf{V}_{22} + \widetilde{\mathbf{Z}}_{(2)+}\widetilde{\mathbf{A}}'_{(2)+}\mathbf{V}_{22}\,.$$

4 GRAPHICAL REPRESENTATION OF VARIABLES AND INDIVIDUALS

4.1 GRAPHICAL REPRESENTATION OF VARIABLES

A graphical representation of the p variables $\mathbf{y}_1, \ldots, \mathbf{y}_p$ (measured in terms of deviations from the means) is usually obtained by their orthogonal projections on the subspace spanned by the first canonical variable (*canonical axis*) or the first two canonical variables (*canonical plane*), belonging to $S(\mathbf{Y}_1)$ or $S(\mathbf{Y}_2)$.

In $S(\mathbf{Y}_1)$, for example, the orthogonal projection $\hat{\mathbf{y}}_j$ of $\mathbf{y}_j$ $(j = 1, \ldots, p)$ on the canonical plane $S(\tilde{\mathbf{z}}_{(1)1}, \tilde{\mathbf{z}}_{(1)2})$ is given by

$$\hat{\mathbf{y}}_j = \tilde{\mathbf{z}}_{(1)1}\tilde{\mathbf{z}}'_{(1)1}\mathbf{M}\mathbf{y}_j + \tilde{\mathbf{z}}_{(1)2}\tilde{\mathbf{z}}'_{(1)2}\mathbf{M}\mathbf{y}_j = \tilde{\mathbf{z}}_{(1)1}\,\sigma_j\, r_{1j} + \tilde{\mathbf{z}}_{(1)2}\,\sigma_j\, r_{2j}$$

where r_{1j} and r_{2j} denote, respectively, the linear correlation coefficients of $\tilde{\mathbf{z}}_{(1)1}$ and $\tilde{\mathbf{z}}_{(1)2}$ with $\mathbf{y}_j$.

However, since we are mainly interested in representing linear correlations between pairs of variables or between a variable and a canonical variable, it is more suitable to work with standardized variables.

In that case, the orthogonal projection $\hat{\mathbf{y}}^*_j$ of the standardized variable $\mathbf{y}^*_j = \mathbf{y}_j / \sigma_j$ $(j = 1, \ldots, p)$ on the canonical plane $S(\tilde{\mathbf{z}}_{(1)1}, \tilde{\mathbf{z}}_{(1)2})$ is given by

$$\hat{\mathbf{y}}^*_j = \tilde{\mathbf{z}}_{(1)1}\tilde{\mathbf{z}}'_{(1)1}\mathbf{M}\mathbf{y}^*_j + \tilde{\mathbf{z}}_{(1)2}\tilde{\mathbf{z}}'_{(1)2}\mathbf{M}\mathbf{y}^*_j = \tilde{\mathbf{z}}_{(1)1} r_{1j} + \tilde{\mathbf{z}}_{(1)2} r_{2j}\,.$$

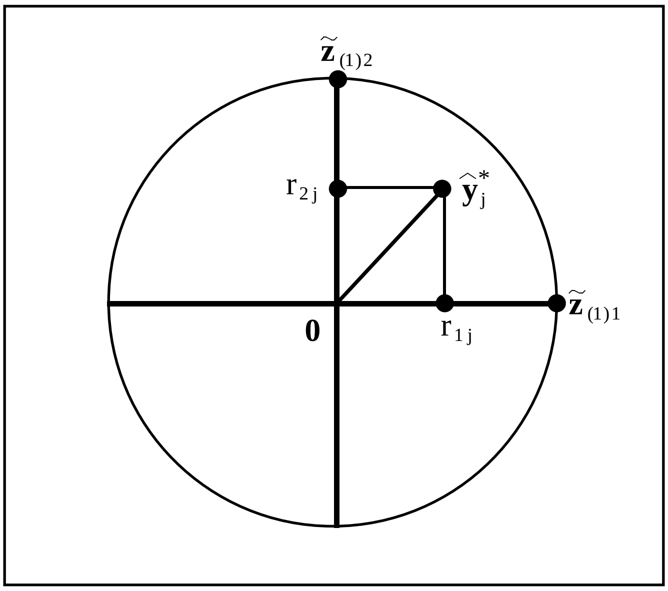
Thus, the co-ordinates of $\hat{\mathbf{y}}^*_j$ relative to $\tilde{\mathbf{z}}_{(1)1}, \tilde{\mathbf{z}}_{(1)2}$ are (r_{1j}, r_{2j}) (Fig. 4).

Of course, each $\hat{\mathbf{y}}^*_j$ $(j = 1, \ldots, p)$ lies inside a circle of centre $\mathbf{0}$ and radius 1 (the so-called *correlation circle*).

Moreover, the quality of representation of $\mathbf{y}^*_j$ on $S(\tilde{\mathbf{z}}_{(1)1}, \tilde{\mathbf{z}}_{(1)2})$ can be judged by means of the square cosine of the angle formed by $\mathbf{y}^*_j$ and $\hat{\mathbf{y}}^*_j$ which is given by $((\mathbf{y}^*_j)'\mathbf{M}(\mathbf{y}^*_j) = 1)$

$$QR(j\,; \tilde{\mathbf{z}}_{(1)1}, \tilde{\mathbf{z}}_{(1)2}) = \frac{[(\mathbf{y}^*_j)'\mathbf{M}(\hat{\mathbf{y}}^*_j)]^2}{[(\mathbf{y}^*_j)'\mathbf{M}(\mathbf{y}^*_j)]\,[(\hat{\mathbf{y}}^*_j)'\mathbf{M}(\hat{\mathbf{y}}^*_j)]} = \frac{[(\mathbf{y}^*_j)'\mathbf{M}(\hat{\mathbf{y}}^*_j)]^2}{(\hat{\mathbf{y}}^*_j)'\mathbf{M}(\hat{\mathbf{y}}^*_j)}\,.$$

Fig. 4



A high $QR(j;\widetilde{\mathbf{z}}_{(1)1},\widetilde{\mathbf{z}}_{(1)2})$ − for example, $QR(j;\widetilde{\mathbf{z}}_{(1)1},\widetilde{\mathbf{z}}_{(1)2}) \geq 0.7$ − means that $\mathbf{y}_j^*$ is well represented by $\widehat{\mathbf{y}}_j^*$; on the contrary, a low $QR(j;\widetilde{\mathbf{z}}_{(1)1},\widetilde{\mathbf{z}}_{(1)2})$ means that the representation of $\mathbf{y}_j^*$ by $\widehat{\mathbf{y}}_j^*$ is poor.

Notice that another expression of $QR(j;\widetilde{\mathbf{z}}_{(1)1},\widetilde{\mathbf{z}}_{(1)2})$ may be obtained taking into account that

$$(\mathbf{y}_j^*)'\mathbf{M}(\widehat{\mathbf{y}}_j^*) = (\mathbf{y}_j^*)'\mathbf{M}(\widetilde{\mathbf{z}}_{(1)1} r_{1j} + \widetilde{\mathbf{z}}_{(1)2} r_{2j}) = r_{1j}^2 + r_{2j}^2$$

and

$$(\widehat{\mathbf{y}}_j^*)'\mathbf{M}(\widehat{\mathbf{y}}_j^*) = (\widetilde{\mathbf{z}}_{(1)1} r_{1j} + \widetilde{\mathbf{z}}_{(1)2} r_{2j})'\mathbf{M}(\widetilde{\mathbf{z}}_{(1)1} r_{1j} + \widetilde{\mathbf{z}}_{(1)2} r_{2j}) = r_{1j}^2 + r_{2j}^2 \,.$$

Thus,

$$QR(j;\widetilde{\mathbf{z}}_{(1)1},\widetilde{\mathbf{z}}_{(1)2}) = r_{1j}^2 + r_{2j}^2 \,.$$

On the other hand, since $QR(j;\widetilde{\mathbf{z}}_{(1)1},\widetilde{\mathbf{z}}_{(1)2})$ also denotes the square distance of $\widehat{\mathbf{y}}_j^*$ from the correlation circle centre, we can see that well-represented variables lie near the circumference of the correlation circle.

Concluding, for well-represented variables we can visualize on the correlation circle:

- which variables are correlated among themselves and with each canonical variable;
- which variables are uncorrelated (orthogonal) among themselves and with each canonical variable.

Of course, an analogous representation may be carried out on the canonical plane $S(\widetilde{\mathbf{z}}_{(2)1}, \widetilde{\mathbf{z}}_{(2)2})$.

These two representations refer to different ways of visualization and are very similar provided that the canonical correlation coefficients between each pair of corresponding canonical variables are close to 1.

4.2 GRAPHICAL REPRESENTATION OF INDIVIDUALS

Now, let us consider the n column vectors (individuals) $\boldsymbol{y}_{(1)1}, \ldots, \boldsymbol{y}_{(1)n}$ of $\mathbf{Y}_1'$ and the n column vectors (individuals) $\boldsymbol{y}_{(2)1}, \ldots, \boldsymbol{y}_{(2)n}$ of $\mathbf{Y}_2'$.

Suppose that these vectors belong, respectively, to the vector spaces R^{p_1} and R^{p_2}.

Moreover, assume that $\mathbf{V}_{11}^{-1}$ and $\mathbf{V}_{22}^{-1}$ are the matrices of the scalar product in R^{p_1} and R^{p_2}, relative to their corresponding canonical bases.

Setting $(h = 1, \ldots, k)$

$$\widetilde{\widetilde{\mathbf{c}}}_{(1)h} = \mathbf{V}_{11}\widetilde{\mathbf{a}}_{(1)h} \quad , \quad \widetilde{\widetilde{\mathbf{c}}}_{(2)h} = \mathbf{V}_{22}\widetilde{\mathbf{a}}_{(2)h}$$

it is immediately apparent that

$$[\widetilde{\widetilde{\mathbf{c}}}_{(1)1} \cdots \widetilde{\widetilde{\mathbf{c}}}_{(1)k}]'\mathbf{V}_{11}^{-1}[\widetilde{\widetilde{\mathbf{c}}}_{(1)1} \cdots \widetilde{\widetilde{\mathbf{c}}}_{(1)k}] = \mathbf{I}_k \ , \ [\widetilde{\widetilde{\mathbf{c}}}_{(2)1} \cdots \widetilde{\widetilde{\mathbf{c}}}_{(2)k}]'\mathbf{V}_{22}^{-1}[\widetilde{\widetilde{\mathbf{c}}}_{(2)1} \cdots \widetilde{\widetilde{\mathbf{c}}}_{(2)k}] = \mathbf{I}_k \ .$$

In R^{p_1}, a graphical representation of the n individuals $\boldsymbol{y}_{(1)1}, \ldots, \boldsymbol{y}_{(1)n}$ is usually obtained by their orthogonal projections on the subspace $S(\widetilde{\widetilde{\mathbf{c}}}_{(1)1})$ spanned by $\widetilde{\widetilde{\mathbf{c}}}_{(1)1}$ or on the subspace $S(\widetilde{\widetilde{\mathbf{c}}}_{(1)1}, \widetilde{\widetilde{\mathbf{c}}}_{(1)2})$ spanned by $\widetilde{\widetilde{\mathbf{c}}}_{(1)1}, \widetilde{\widetilde{\mathbf{c}}}_{(1)2}$.

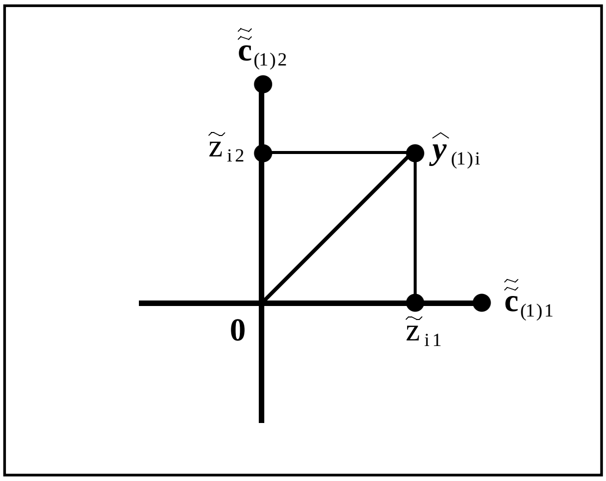
Confining ourselves to considering this last type of representation, we notice that the orthogonal projection $\widehat{\boldsymbol{y}}_{(1)i}$ of $\boldsymbol{y}_{(1)i}$ $(i = 1, \ldots, n)$ on $S(\widetilde{\widetilde{\mathbf{c}}}_{(1)1}, \widetilde{\widetilde{\mathbf{c}}}_{(1)2})$ is given by

$$\begin{aligned}\widehat{\boldsymbol{y}}_{(1)\mathrm{i}} &= \left[\tilde{\tilde{\mathbf{c}}}_{(1)1} \;\; \tilde{\tilde{\mathbf{c}}}_{(1)2}\right]\left[\tilde{\tilde{\mathbf{c}}}_{(1)1} \;\; \tilde{\tilde{\mathbf{c}}}_{(1)2}\right]' \mathbf{V}_{11}^{-1} \boldsymbol{y}_{(1)\mathrm{i}} \\ &= \tilde{\tilde{\mathbf{c}}}_{(1)1} \tilde{\tilde{\mathbf{c}}}'_{(1)1} \mathbf{V}_{11}^{-1} \boldsymbol{y}_{(1)\mathrm{i}} + \tilde{\tilde{\mathbf{c}}}_{(1)2} \tilde{\tilde{\mathbf{c}}}'_{(1)2} \mathbf{V}_{11}^{-1} \boldsymbol{y}_{(1)\mathrm{i}} \\ &= \tilde{\tilde{\mathbf{c}}}_{(1)1} \tilde{\mathbf{a}}'_{(1)1} \boldsymbol{y}_{(1)\mathrm{i}} + \tilde{\tilde{\mathbf{c}}}_{(1)2} \tilde{\mathbf{a}}'_{(1)2} \boldsymbol{y}_{(1)\mathrm{i}} \\ &= \tilde{\tilde{\mathbf{c}}}_{(1)1} \tilde{\mathrm{z}}_{\mathrm{i}1} + \tilde{\tilde{\mathbf{c}}}_{(1)2} \tilde{\mathrm{z}}_{\mathrm{i}2}\end{aligned}$$

where $\tilde{\mathrm{z}}_{\mathrm{ij}}$ $(\mathrm{j} = 1, 2)$ is the ith element of the canonical vector $\tilde{\mathbf{z}}_{(1)\mathrm{j}}$.

Thus, the co-ordinates of $\widehat{\boldsymbol{y}}_{(1)\mathrm{i}}$ relative to $\tilde{\tilde{\mathbf{c}}}_{(1)1}, \tilde{\tilde{\mathbf{c}}}_{(1)2}$ are $(\tilde{\mathrm{z}}_{\mathrm{i}1}, \tilde{\mathrm{z}}_{\mathrm{i}2})$ (Fig. 5).

Fig. 5



Moreover, the quality of representation of each $\boldsymbol{y}_{(1)\mathrm{i}}$ $(\mathrm{i} = 1, \ldots, \mathrm{n})$ on $\mathrm{S}(\tilde{\tilde{\mathbf{c}}}_{(1)1}, \tilde{\tilde{\mathbf{c}}}_{(1)2})$ can be judged by means of the square cosine of the angle formed by $\boldsymbol{y}_{(1)\mathrm{i}}$ and $\widehat{\boldsymbol{y}}_{(1)\mathrm{i}}$ which is given by

$$\mathrm{QR}(\mathrm{i}; \tilde{\tilde{\mathbf{c}}}_{(1)1}, \tilde{\tilde{\mathbf{c}}}_{(1)2}) = \frac{(\boldsymbol{y}'_{(1)\mathrm{i}} \mathbf{V}_{11}^{-1} \widehat{\boldsymbol{y}}_{(1)\mathrm{i}})^2}{(\boldsymbol{y}'_{(1)\mathrm{i}} \mathbf{V}_{11}^{-1} \boldsymbol{y}_{(1)\mathrm{i}})(\widehat{\boldsymbol{y}}'_{(1)\mathrm{i}} \mathbf{V}_{11}^{-1} \widehat{\boldsymbol{y}}_{(1)\mathrm{i}})}\,.$$

A high $\mathrm{QR}(\mathrm{i}; \tilde{\tilde{\mathbf{c}}}_{(1)1}, \tilde{\tilde{\mathbf{c}}}_{(1)2})$ – for example, $\mathrm{QR}(\mathrm{i}; \tilde{\tilde{\mathbf{c}}}_{(1)1}, \tilde{\tilde{\mathbf{c}}}_{(1)2}) \geq 0.7$ – means that $\boldsymbol{y}_{(1)\mathrm{i}}$ is well represented by $\widehat{\boldsymbol{y}}_{(1)\mathrm{i}}$; on the contrary, a low $\mathrm{QR}(\mathrm{i}; \tilde{\tilde{\mathbf{c}}}_{(1)1}, \tilde{\tilde{\mathbf{c}}}_{(1)2})$ means that the representation of $\boldsymbol{y}_{(1)\mathrm{i}}$ by $\widehat{\boldsymbol{y}}_{(1)\mathrm{i}}$ is poor.

Of course, the procedure described above may be applied for the representation of $\boldsymbol{y}_{(2)1}, \ldots, \boldsymbol{y}_{(2)\mathrm{n}}$ on the subspace $\mathrm{S}(\tilde{\tilde{\mathbf{c}}}_{(2)1}, \tilde{\tilde{\mathbf{c}}}_{(2)2})$ spanned by $\tilde{\tilde{\mathbf{c}}}_{(2)1}, \tilde{\tilde{\mathbf{c}}}_{(2)2}$.

5 OTHER APPROACHES TO CCA

5.1 THE APPROACH IN TERMS OF THE MULTIVARIABLE LINEAR MODEL

The approach we would like to mention is based on the multivariable linear model.

Firstly, consider the model

$$\mathbf{Y}_1 = \mathbf{Y}_2\mathbf{H}_2 + \mathbf{E}_2$$

where $\mathbf{E}_2$ is a matrix of «residuals» and $\mathbf{H}_2$ is a matrix, of order (p_2, p_1), of unknown coefficients.

In order to determine the matrix $\mathbf{H}_2$, we can choose a least square criterion.

However, without any assumption regarding the rank of $\mathbf{H}_2$, the best solution is trivially given by

$$\widehat{\mathbf{H}}_2 = (\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2)^{-1}\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_1$$

where $r(\widehat{\mathbf{H}}_2) \leq p_1 = r(\mathbf{Y}_1)$.

Then, assume that $\mathbf{H}_2$ has rank $k^* < p_1$, so that it may be written in the form ($\mathbf{F}_2$ and $\mathbf{G}_2$ of order, respectively, (p_2, k^*) and (k^*, p_1))

$$\mathbf{H}_2 = \mathbf{F}_2\mathbf{G}_2$$

where $r(\mathbf{F}_2) = r(\mathbf{G}_2) = k^*$.

In this case, our model becomes

$$\mathbf{Y}_1 = \mathbf{Y}_2\mathbf{F}_2\mathbf{G}_2 + \mathbf{E}_2$$

and we propose to find out

$$(25) \quad \underset{\mathbf{F}_2, \mathbf{G}_2}{\text{Min}} \quad \text{tr}\{(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{F}_2\mathbf{G}_2)'\mathbf{M}(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{F}_2\mathbf{G}_2)\mathbf{V}_{11}^{-1}\} \quad , \quad \mathbf{F}_2'\mathbf{V}_{22}\mathbf{F}_2 = \mathbf{I}_{k^*} .$$

To this end, first notice that, taking into account the constraint on the matrix $\mathbf{F}_2$, we can write

$$\begin{aligned} & \operatorname{tr}\{(\mathbf{Y}_1-\mathbf{Y}_2\mathbf{F}_2\mathbf{G}_2)'\mathbf{M}(\mathbf{Y}_1-\mathbf{Y}_2\mathbf{F}_2\mathbf{G}_2)\mathbf{V}_{11}^{-1}\} \\ &= \operatorname{tr}\{\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_1\mathbf{V}_{11}^{-1}\} - \operatorname{tr}\{\mathbf{Y}_1'\mathbf{M}\mathbf{Y}_2\mathbf{F}_2\mathbf{G}_2\mathbf{V}_{11}^{-1}\} \\ &\quad - \operatorname{tr}\{\mathbf{G}_2'\mathbf{F}_2'\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_1\mathbf{V}_{11}^{-1}\} + \operatorname{tr}\{\mathbf{G}_2'\mathbf{F}_2'\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2\mathbf{F}_2\mathbf{G}_2\mathbf{V}_{11}^{-1}\} \\ &= \operatorname{tr}\{\mathbf{I}_{k*}\} - 2\operatorname{tr}\{\mathbf{V}_{12}\mathbf{F}_2\mathbf{G}_2\mathbf{V}_{11}^{-1}\} + \operatorname{tr}\{\mathbf{G}_2'\mathbf{G}_2\mathbf{V}_{11}^{-1}\} \;. \end{aligned}$$

Thus, our problem lies in finding out

$$(25') \qquad \underset{\mathbf{F}_2,\mathbf{G}_2}{\operatorname{Max}} \; \{2\operatorname{tr}\{\mathbf{V}_{12}\mathbf{F}_2\mathbf{G}_2\mathbf{V}_{11}^{-1}\} - \operatorname{tr}\{\mathbf{G}_2'\mathbf{G}_2\mathbf{V}_{11}^{-1}\}\} \quad , \quad \mathbf{F}_2'\mathbf{V}_{22}\mathbf{F}_2 = \mathbf{I}_{k*} \;.$$

Now, consider the function

$$\mathrm{L}(\mathbf{F}_2,\mathbf{G}_2,\mathbf{L}_2) = 2\operatorname{tr}\{\mathbf{V}_{12}\mathbf{F}_2\mathbf{G}_2\mathbf{V}_{11}^{-1}\} - \operatorname{tr}\{\mathbf{G}_2'\mathbf{G}_2\mathbf{V}_{11}^{-1}\}\} - \operatorname{tr}\{(\mathbf{F}_2'\mathbf{V}_{22}\mathbf{F}_2 - \mathbf{I}_{k*})\,\mathbf{L}_2\}$$

where $\mathbf{L}_2 = \mathbf{L}_2'$ is a matrix of Lagrange multipliers of order (k^*, k^*).

At $(\tilde{\mathbf{F}}_2,\tilde{\mathbf{G}}_2,\tilde{\mathbf{L}}_2)$ where $\mathrm{L}(\mathbf{F}_2,\mathbf{G}_2,\mathbf{L}_2)$ has a maximum, as can easily be verified, it must be

$$\begin{aligned} \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\tilde{\mathbf{G}}_2' &= \mathbf{V}_{22}\tilde{\mathbf{F}}_2\tilde{\mathbf{L}}_2 \\ \tilde{\mathbf{F}}_2'\mathbf{V}_{21} &= \tilde{\mathbf{G}}_2 \\ \tilde{\mathbf{F}}_2'\mathbf{V}_{22}\tilde{\mathbf{F}}_2 &= \mathbf{I}_{k*} \;. \end{aligned}$$

Therefore, we must find out solutions of the system

$$\begin{aligned} \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{G}_2' &= \mathbf{V}_{22}\mathbf{F}_2\mathbf{L}_2 \\ \mathbf{F}_2'\mathbf{V}_{21} &= \mathbf{G}_2 \\ \mathbf{F}_2'\mathbf{V}_{22}\mathbf{F}_2 &= \mathbf{I}_{k*} \end{aligned}$$

in the unknowns $\mathbf{F}_2, \mathbf{G}_2, \mathbf{L}_2$.

Clearly,

$$\tilde{\mathbf{F}}_2 = \tilde{\mathbf{A}}^*_{(2)} = \left[\tilde{\mathbf{a}}_{(2)1}\cdots\tilde{\mathbf{a}}_{(2)k*}\right] \;,\; \tilde{\mathbf{G}}_2 = (\tilde{\mathbf{A}}^*_{(2)})'\mathbf{V}_{21} \;,\; \tilde{\mathbf{L}}_2 = (\tilde{\mathbf{R}}^*)^2 = \operatorname{diag}(\tilde{r}_1^2,\ldots,\tilde{r}_{k*}^2)$$

is a solution of our problem.

Successively, consider the model

$$\mathbf{Y}_2 = \mathbf{Y}_1\mathbf{H}_1 + \mathbf{E}_1$$

where $\mathbf{E}_1$ is a matrix of «residuals» and $\mathbf{H}_1$ is a matrix, of order (p_1, p_2), of unknown coefficients.

Assume that $\mathbf{H}_1$ has rank $k^* < p_1$, so that it may be written in the form ($\mathbf{F}_1$ and $\mathbf{G}_1$ of order, respectively, (p_1, k^*) and (k^*, p_2))

$$\mathbf{H}_1 = \mathbf{F}_1 \mathbf{G}_1$$

where $r(\mathbf{F}_1) = r(\mathbf{G}_1) = k^*$.

Then, our model becomes

$$\mathbf{Y}_2 = \mathbf{Y}_1 \mathbf{F}_1 \mathbf{G}_1 + \mathbf{E}_1$$

and we propose to find out

$$(26) \quad \underset{\mathbf{F}_1, \mathbf{G}_1}{\text{Min}} \ \text{tr}\{(\mathbf{Y}_2 - \mathbf{Y}_1 \mathbf{F}_1 \mathbf{G}_1)'\mathbf{M}(\mathbf{Y}_2 - \mathbf{Y}_1 \mathbf{F}_1 \mathbf{G}_1)\mathbf{V}_{22}^{-1}\} \quad , \quad \mathbf{F}_1' \mathbf{V}_{11} \mathbf{F}_1 = \mathbf{I}_{k^*} .$$

Reasoning as above, our problem lies in finding out

$$(26') \quad \underset{\mathbf{F}_1, \mathbf{G}_1}{\text{Max}} \ \{2\,\text{tr}\{\mathbf{V}_{21} \mathbf{F}_1 \mathbf{G}_1 \mathbf{V}_{22}^{-1}\} - \text{tr}\{\mathbf{G}_1' \mathbf{G}_1 \mathbf{V}_{22}^{-1}\}\} \quad , \quad \mathbf{F}_1' \mathbf{V}_{11} \mathbf{F}_1 = \mathbf{I}_{k^*} .$$

Therefore, we must find out solutions of the system

$$\begin{aligned} \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{G}_1' &= \mathbf{V}_{11} \mathbf{F}_1 \mathbf{L}_1 \\ \mathbf{F}_1' \mathbf{V}_{12} &= \mathbf{G}_1 \\ \mathbf{F}_1' \mathbf{V}_{11} \mathbf{F}_1 &= \mathbf{I}_{k^*} \end{aligned}$$

in the unknowns $\mathbf{F}_1, \mathbf{G}_1, \mathbf{L}_1$ ($\mathbf{L}_1 = \mathbf{L}_1'$: matrix of Lagrange multipliers of order (k^*, k^*).

Clearly,

$$\tilde{\mathbf{F}}_1 = \tilde{\mathbf{A}}^*_{(1)} = \left[\tilde{\mathbf{a}}_{(1)1} \cdots \tilde{\mathbf{a}}_{(1)k^*}\right] \ , \ \tilde{\mathbf{G}}_1 = (\tilde{\mathbf{A}}^*_{(1)})' \mathbf{V}_{12} \ , \ \tilde{\mathbf{L}}_1 = (\tilde{\mathbf{R}}^*)^2 = \text{diag}\,(\tilde{r}_1^2, \ldots, \tilde{r}_{k^*}^2)$$

is a solution of our problem.

REMARK 8. As we have pointed out, $\tilde{\mathbf{F}}_1$ and $\tilde{\mathbf{F}}_2$ are the matrices of the canonical factors corresponding to the first k^* canonical correlation coefficients.

In turn, $\tilde{\mathbf{G}}_1$ and $\tilde{\mathbf{G}}_2$ − taking into account that $\tilde{\mathbf{Z}}^*_{(1)} = \mathbf{Y}_1\tilde{\mathbf{F}}_1$ and $\tilde{\mathbf{Z}}^*_{(2)} = \mathbf{Y}_2\tilde{\mathbf{F}}_2$ are the matrices of the first k∗ canonical variables − can be interpreted as the matrices of the orthogonal projection coefficients of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ on the sub-spaces spanned by those canonical variables.

REMARK 9. Notice the relations

$$\tilde{\mathbf{F}}_1 = \mathbf{V}_{11}^{-1}\mathbf{V}_{12}\tilde{\mathbf{F}}_2(\tilde{\mathbf{R}}^*)^{-1} = \mathbf{V}_{11}^{-1}\tilde{\mathbf{G}}_2'(\tilde{\mathbf{R}}^*)^{-1} \;,\; \tilde{\mathbf{F}}_2 = \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\tilde{\mathbf{F}}_1(\tilde{\mathbf{R}}^*)^{-1} = \mathbf{V}_{22}^{-1}\tilde{\mathbf{G}}_1'(\tilde{\mathbf{R}}^*)^{-1} \;.$$

Alternatively, consider the model

$$\mathbf{Y}_1\mathbf{A}_1 = \mathbf{Y}_2\mathbf{A}_2 + \mathbf{E}_3$$

where $\mathbf{E}_3$ is a matrix of «residuals», $\mathbf{A}_1$ and $\mathbf{A}_2$ are matrices of unknown coefficients of order, respectively, $(p_1, k*)$ and $(p_2, k*)$, such that $r(\mathbf{A}_1) = r(\mathbf{A}_2) = k*$.

We propose to find out

$$(27)\qquad \underset{\mathbf{A}_1,\mathbf{A}_2}{\text{Min}}\ \text{tr}\{(\mathbf{Y}_1\mathbf{A}_1 - \mathbf{Y}_2\mathbf{A}_2)'\mathbf{M}(\mathbf{Y}_1\mathbf{A}_1 - \mathbf{Y}_2\mathbf{A}_2)\} \quad,\quad \mathbf{A}_1'\mathbf{V}_{11}\mathbf{A}_1 = \mathbf{I}_{k*}\;.$$

To this end, first notice that, taking into account the constraint on the matrix $\mathbf{A}_1$, our problem lies in finding out

$$(27')\qquad \underset{\mathbf{A}_1,\mathbf{A}_2}{\text{Max}}\ 2\,\text{tr}\{\mathbf{A}_1'\mathbf{V}_{12}\mathbf{A}_2\} - \text{tr}\{\mathbf{A}_2'\mathbf{V}_{22}\mathbf{A}_2\} \quad,\quad \mathbf{A}_1'\mathbf{V}_{11}\mathbf{A}_1 = \mathbf{I}_{k*}\;.$$

Now, consider the function

$$L(\mathbf{A}_1,\mathbf{A}_2,\mathbf{L}_3) = 2\,\text{tr}\{\mathbf{A}_1'\mathbf{V}_{12}\mathbf{A}_2\} - \text{tr}\{\mathbf{A}_2'\mathbf{V}_{22}\mathbf{A}_2\}\} - \text{tr}\{(\mathbf{A}_1'\mathbf{V}_{11}\mathbf{A}_{(1)} - \mathbf{I}_{k*})\,\mathbf{L}_3\}$$

where $\mathbf{L}_3 = \mathbf{L}_3'$ is a matrix of Lagrange multipliers of order $(k*, k*)$.

At $(\tilde{\mathbf{A}}_1,\tilde{\mathbf{A}}_2,\tilde{\mathbf{L}}_3)$ where $L(\mathbf{A}_1,\mathbf{A}_2,\mathbf{L}_3)$ has a maximum, it must be

$$\begin{aligned} \mathbf{V}_{12}\tilde{\mathbf{A}}_2 &= \mathbf{V}_{11}\tilde{\mathbf{A}}_1\tilde{\mathbf{L}}_3 \\ \mathbf{V}_{21}\tilde{\mathbf{A}}_1 &= \mathbf{V}_{22}\tilde{\mathbf{A}}_2 \\ \tilde{\mathbf{A}}_1'\mathbf{V}_{11}\tilde{\mathbf{A}}_1 &= \mathbf{I}_{k*}\;. \end{aligned}$$

Therefore, we must find out solutions of the system

$$\mathbf{V}_{12}\mathbf{A}_2 = \mathbf{V}_{11}\mathbf{A}_1\mathbf{L}_3$$
$$\mathbf{V}_{21}\mathbf{A}_1 = \mathbf{V}_{22}\mathbf{A}_2$$
$$\mathbf{A}_1'\mathbf{V}_{11}\mathbf{A}_1 = \mathbf{I}_{k*}$$

in the unknowns $\mathbf{A}_1, \mathbf{A}_2, \mathbf{L}_3$.

Clearly,

$$\tilde{\mathbf{A}}_1 = \tilde{\mathbf{F}}_1 \quad , \quad \tilde{\mathbf{A}}_2 = \tilde{\mathbf{F}}_2\tilde{\mathbf{R}}^* \quad , \quad \tilde{\mathbf{L}}_3 = (\tilde{\mathbf{R}}^*)^2$$

is a solution of our problem.

Analogously, consider the model

$$\mathbf{Y}_2\mathbf{B}_2 = \mathbf{Y}_1\mathbf{B}_1 + \mathbf{E}_4$$

where $\mathbf{E}_4$ is a matrix of «residuals», $\mathbf{B}_2$ and $\mathbf{B}_1$ are matrices of unknown coefficients of order, respectively, $(p_2, k*)$ and $(p_1, k*)$, such that $r(\mathbf{B}_2) = r(\mathbf{B}_1) = k*$.

It can easily be shown that

$$\tilde{\mathbf{B}}_2 = \tilde{\mathbf{F}}_2 \quad , \quad \tilde{\mathbf{B}}_1 = \tilde{\mathbf{F}}_1\tilde{\mathbf{R}}^* \ .$$

REMARK 10. Notice that

$$\tilde{\mathbf{B}}_1 = \tilde{\mathbf{A}}_1\tilde{\mathbf{R}}^* \quad , \quad \tilde{\mathbf{B}}_2 = \tilde{\mathbf{A}}_2(\tilde{\mathbf{R}}^*)^{-1} \ .$$

5.2 THE APPROACH IN TERMS OF PCA

Consider again the fundamental equation of CCA, namely the equation

$$\begin{bmatrix} -r\mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & -r\mathbf{V}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{a}_{(1)} \\ \mathbf{a}_{(2)} \end{bmatrix} = \mathbf{0} \ .$$

Setting

$$-r = 1 - \lambda \ ,$$

we can write ($\mathbf{Q} = \text{diag}(\mathbf{V}_{11}^{-1}, \mathbf{V}_{22}^{-1})$; $\mathbf{O}$ of appropriate order)

$$\{\begin{bmatrix} (1-\lambda)\,\mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & (1-\lambda)\,\mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{(1)} \\ \mathbf{a}_{(2)} \end{bmatrix} = \mathbf{0}\}$$
$$\Leftrightarrow \{(\begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} - \begin{bmatrix} \lambda\,\mathbf{V}_{11} & \mathbf{O} \\ \mathbf{O} & \lambda\,\mathbf{V}_{22} \end{bmatrix})\begin{bmatrix} \mathbf{a}_{(1)} \\ \mathbf{a}_{(2)} \end{bmatrix} = \mathbf{0}\}$$
$$\Leftrightarrow \{(\begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{V}_{11}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{V}_{22}^{-1} \end{bmatrix} - \lambda\,\mathbf{I}_p)\begin{bmatrix} \mathbf{V}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{V}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{a}_{(1)} \\ \mathbf{a}_{(2)} \end{bmatrix} = \mathbf{0}\}$$
$$\Leftrightarrow \{(\mathbf{V}\mathbf{Q} - \lambda\,\mathbf{I}_p)\,\mathbf{Q}^{-1}\begin{bmatrix} \mathbf{a}_{(1)} \\ \mathbf{a}_{(2)} \end{bmatrix} = \mathbf{0}\}$$
$$\Leftrightarrow \{(\mathbf{V}\mathbf{Q} - \lambda\,\mathbf{I}_p)\,\mathbf{Q}^{-1}\begin{bmatrix} \mathbf{a}_{(1)} \\ \mathbf{a}_{(2)} \end{bmatrix}\frac{1}{\sqrt{2}} = \mathbf{0}\}\ .$$

Thus, the canonical correlation coefficients and the canonical factors are linked to the eigenvalues and to the eigenvectors (appropriately normalized) of the fundamental equation of PCA

$$(\mathbf{V}\mathbf{Q} - \lambda\mathbf{I}_p)\,\mathbf{c} = \mathbf{0}\ ,$$

by means of the relations $(h = 1, \dots, k)$

$$-\tilde{r}_h = 1 - \tilde{\lambda}_h \quad , \quad \begin{bmatrix} \tilde{\mathbf{a}}_{(1)h} \\ \tilde{\mathbf{a}}_{(2)h} \end{bmatrix} = \sqrt{2}\,\mathbf{Q}\,\tilde{\mathbf{c}}_h\ .$$

Moreover, as is easily seen − for $0 < \tilde{r}_h \le 1$, namely for $1 < \tilde{\lambda}_h \le 2$ − the principal component $\tilde{\mathbf{y}}_h$ is linked to the canonical variables $\tilde{\mathbf{z}}_{(1)h}, \tilde{\mathbf{z}}_{(2)h}$ by means of the formula

$$(28) \quad \tilde{\mathbf{y}}_h = [\,\mathbf{Y}_1 \;\; \mathbf{Y}_2\,]\,\mathbf{Q}\,\tilde{\mathbf{c}}_h = [\,\mathbf{Y}_1 \;\; \mathbf{Y}_2\,]\begin{bmatrix} \tilde{\mathbf{a}}_{(1)h} \\ \tilde{\mathbf{a}}_{(2)h} \end{bmatrix}\frac{1}{\sqrt{2}} = \tilde{\mathbf{z}}_{(1)h}\frac{1}{\sqrt{2}} + \tilde{\mathbf{z}}_{(2)h}\frac{1}{\sqrt{2}}\ .$$

Finally, since

$$(29) \quad \mathbf{P}_1\,\tilde{\mathbf{y}}_h = \mathbf{P}_1\,\tilde{\mathbf{z}}_{(1)h}\frac{1}{\sqrt{2}} + \mathbf{P}_1\,\tilde{\mathbf{z}}_{(2)h}\frac{1}{\sqrt{2}} \quad = \tilde{\mathbf{z}}_{(1)h}\frac{1}{\sqrt{2}} + \tilde{r}_h\,\tilde{\mathbf{z}}_{(1)h}\frac{1}{\sqrt{2}}$$
$$= \tilde{\mathbf{z}}_{(1)h}\frac{1}{\sqrt{2}} + (\tilde{\lambda}_h - 1)\,\tilde{\mathbf{z}}_{(1)h}\frac{1}{\sqrt{2}} = \tilde{\mathbf{z}}_{(1)h}\frac{1}{\sqrt{2}} + \tilde{\lambda}_h\,\tilde{\mathbf{z}}_{(1)h}\frac{1}{\sqrt{2}} - \tilde{\mathbf{z}}_{(1)h}\frac{1}{\sqrt{2}}$$
$$= \tilde{\lambda}_h\,\tilde{\mathbf{z}}_{(1)h}\frac{1}{\sqrt{2}}$$

and, analogously,

(30) $$\mathbf{P}_2\,\tilde{\mathbf{y}}_h = \tilde{\lambda}_h\,\tilde{\mathbf{z}}_{(2)h}\,\frac{1}{\sqrt{2}}$$

we also find that

(31) $$\begin{aligned}(\mathbf{P}_1+\mathbf{P}_2)\,\tilde{\mathbf{y}}_h &= \tilde{\lambda}_h\,\tilde{\mathbf{z}}_{(1)h}\,\frac{1}{\sqrt{2}} + \tilde{\lambda}_h\,\tilde{\mathbf{z}}_{(2)h}\,\frac{1}{\sqrt{2}}\\ &= \tilde{\lambda}_h\,(\tilde{\mathbf{z}}_{(1)h}\,\frac{1}{\sqrt{2}} + \tilde{\mathbf{z}}_{(2)h}\,\frac{1}{\sqrt{2}})\\ &= \tilde{\lambda}_h\,\tilde{\mathbf{y}}_h\,.\end{aligned}$$

In other words, $\tilde{\mathbf{y}}_h$ is an eigenvector of the matrix $(\mathbf{P}_1+\mathbf{P}_2)$ corresponding to the eigenvalue $\tilde{\lambda}_h$ (11).

REMARK 11. Notice that from (29) and (30) we get the relations

$$\tilde{\mathbf{z}}_{(1)h} = \sqrt{2}\;\mathbf{P}_1\,\frac{\tilde{\mathbf{y}}_h}{\tilde{\lambda}_h}\quad,\quad \tilde{\mathbf{z}}_{(2)h} = \sqrt{2}\;\mathbf{P}_2\,\frac{\tilde{\mathbf{y}}_h}{\tilde{\lambda}_h}\,.$$

REMARK 12. Whenever $n > p_1 + p_2 = p$, as often happens in practical applications, it is not convenient, from the computation point of view, to use the equation $(\mathbf{P}_1+\mathbf{P}_2)\mathbf{y} = \lambda\mathbf{y}$ to obtain first the eigenvalues $\tilde{\lambda}_1, \dots, \tilde{\lambda}_k$ and the principal components $\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_k$, then the canonical variables $\tilde{\mathbf{z}}_{(1)1}, \tilde{\mathbf{z}}_{(2)1}, \dots, \tilde{\mathbf{z}}_{(1)k}, \tilde{\mathbf{z}}_{(2)k}$.

Rather, it is more suitable to perform a PCA to obtain first the eigenvalues and the principal components, then the canonical variables.

It is important to point out the statistical criterion underlying (31).

To this end, suppose we want to find a normalized linear combination $\tilde{\tilde{\mathbf{y}}}_{(1)}$ of $\mathbf{y}_1, \dots, \mathbf{y}_p$ maximizing the sum of square multiple linear correlation coefficients $\rho_1 + \rho_2$ between $\tilde{\tilde{\mathbf{y}}}_{(1)}$ and the column vectors of $\mathbf{Y}_1$ and $\mathbf{Y}_2$.

Denote by $\tilde{\tilde{\mathbf{y}}}$ a generic normalized linear combination of $\mathbf{y}_1, \dots, \mathbf{y}_p$.

Since we have $(\tilde{\tilde{\mathbf{y}}}'\mathbf{M}\tilde{\tilde{\mathbf{y}}} = 1)$

$$\rho_1 = \cos^2(\tilde{\tilde{\mathbf{y}}}, \mathbf{P}_1\tilde{\tilde{\mathbf{y}}}) = \tilde{\tilde{\mathbf{y}}}'\mathbf{M}\mathbf{P}_1\tilde{\tilde{\mathbf{y}}}\quad,\quad \rho_2 = \cos^2(\tilde{\tilde{\mathbf{y}}}, \mathbf{P}_2\tilde{\tilde{\mathbf{y}}}) = \tilde{\tilde{\mathbf{y}}}'\mathbf{M}\mathbf{P}_2\tilde{\tilde{\mathbf{y}}}\,,$$

---

(11) As can easily be verified, the linear transformation associated to $(\mathbf{P}_1 + \mathbf{P}_2)$ is selfadjoint.

we must find out

$$\operatorname*{Max}_{\tilde{\tilde{\mathbf{y}}}} \; (\rho_1 + \rho_2) = \operatorname*{Max}_{\tilde{\tilde{\mathbf{y}}}} \; (\tilde{\tilde{\mathbf{y}}}'\mathbf{M}(\mathbf{P}_1 + \mathbf{P}_2)\tilde{\tilde{\mathbf{y}}}) \quad , \quad \tilde{\tilde{\mathbf{y}}}'\mathbf{M}\tilde{\tilde{\mathbf{y}}} = 1 \; .$$

This problem of constrained maximization can be solved very easily.

It results that $\tilde{\tilde{\mathbf{y}}}_{(1)}$ is given by the normalized eigenvector of $(\mathbf{P}_1 + \mathbf{P}_2)$ associated with the eigenvalue $\tilde{\lambda}_1$.

In other words, $\tilde{\tilde{\mathbf{y}}}_{(1)} = \tilde{\tilde{\mathbf{y}}}_1$, the first standardized principal component.

Of course, an analogous meaning may be attributed to each of the subsequent standardized principal components.

REFERENCES

[1] Anderson, T.W., *Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.

[2] Basilevsky, A., *Statistical Factor Analysis and Related Methods*, John Wiley and Sons, New York, 1994.

[3] Bertier, P., Bouroche, J.M., *Analyse des données multidimensionnelles*, PUF, Paris, 1977.

[4] Bouroche, J.M., Saporta, G., *L'analisi dei dati*, CLU, Napoli, 1983.

[5] Cailliez, F., Pages, G.P., *Introduction à l'analyse des données*, Smash, Paris, 1976.

[6] Carroll, J.D., *A Generalization of Canonical Correlation to Three or More Sets of Variables*, Proc. 76th Conv. Amer. Psych. Ass., 1968.

[7] Coppi, R., *Appunti di statistica metodologica: analisi lineare dei dati*, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Roma, 1986.

[8] Delvecchio, F., *Analisi statistica di dati multidimensionali*, Cacucci Editore, Bari, 1992.

[9] Diday, E., Lemaire, J., Pouget, J., Testu, F., *Eléments d'analyse des données*, Dunod, Paris, 1982.

[10] Fabbris, L., *Analisi esplorativa di dati multidimensionali*, cleup editore, Padova, 1990.

[11] Kettenring, R.J., *Canonical Analysis of Several Sets of Variables*, Biometrika, 1971.

[12] Krzanowski, W.J., *Principles of Multivariate Analysis*, Oxford University Press, Oxford, 2000.

[13] Kshirsagar, A.M., *Multivariate Analysis*, Marcel Dekker, Inc., New York, 1972.

[14] Leoni, R., *Alcuni argomenti di analisi statistica multivariata*, Dipartimento Statistico, Firenze, 1978.

[15] Leoni, R., *Canonical Correlation Analysis*, in «Methods for Multidimensional Data Analysis», European Courses in Advanced Statistics, Anacapri, 1987.

[16] Leoni, R., *Modello lineare multivariato e analisi statistica multidimensionale*, in «Conferenze di statistica nell'anno del 750° anniversario dell'Università degli Studi di Siena», Dipartimento di Metodi Quantitativi, Siena, 1994.

[17] Leoni, R., *Algebra lineare per le applicazioni statistiche*, Dipartimento di Statistica "G. Parenti", Firenze, 2007 (sta in <http://www.ds.unifi.it> alla voce *Materiale Didattico*).

[18] Leoni, R., *Principal Component Analysis*, Department of Statistics "G. Parenti", Florence, 2007.

[19] Mardia, K.V., Kent, I.T., Bibby, J.M., *Multivariate Analysis*, Academic Press, London, 1979.

[20] Rizzi, A., *Analisi dei dati*, NIS, Roma, 1985.

[21] Saporta, G., *Probabilités*, *Analyse des données et Statistique*, Èditions Technip, Paris, 1990.

[22] Seber, G.A.F., *Multivariate Observations*, John Wiley and Sons, New York, 1984.

[23] Volle, M., *Analyse des données*, Economica, Paris, 1981.