RENATO LEONI

Principal Component Analysis

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

1 INTRODUCTION

Researchers are frequently faced with the task of analysing a data collection concerning a large number of quantitative variables measured on many individuals (units) and usually displayed in tabular form. The aim of the analysis is often to find out patterns of interrelationships which may exist among variables or individuals. The problem is that, given the data volume, this aim is not readily achieved.

The focus of principal component analysis (PCA) is on the study of a large data collection of the type mentioned above from the point of view of the interrelationships which may subsist among variables or individuals, providing at the same time the researcher with a graphical representation of results on a subspace of low dimension (usually one or two).

In this paper, without making any assumption about an underlying probabilistic model, we will present the main features of PCA.

The contents of the paper can be summarized as follows.

In Section 2, the basic data and their algebraic structure are set out. In Section 3 and 4, privileging a geometrical language, some concepts which will be used extensively during the paper are introduced. Section 5 is devoted to a presentation of an approach to PCA. In Section 6, rules for a graphical representation of results are given. Finally, in Section 7, other approaches to PCA are set out (1).

(1) Numerical examples, based both on fictitiuos and real data, are provided apart. Relevant algebraic concepts are stated in [20].

2 BASIC DATA AND THEIR ALGEBRAIC STRUCTURE

2.1 RAW DATA MATRIX

Consider the matrix (*raw data 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.

Although in practical applications the number n of individuals is often strictly greater of the number p of variables, that assumption is not necessary in performing PCA and will be dropped; in other words, we will suppose that it may be $n \gtreqless p$.

Notice that, setting $(i = 1, \dots, n)$

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

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

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

we can write

$$\mathbf{X}' = \begin{bmatrix} \boldsymbol{x}_1 \cdots \boldsymbol{x}_n \end{bmatrix}$$

and

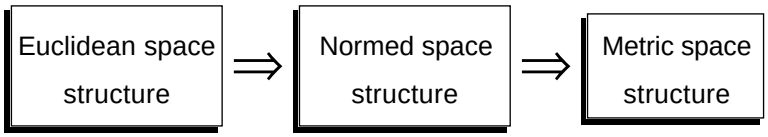
$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \cdots \mathbf{x}_p \end{bmatrix}.$$

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

2.2 ALGEBRAIC STRUCTURE

Regarding $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ and $\mathbf{x}_1, \ldots, \mathbf{x}_p$ as elements of R^p and R^n, respectively, R^p (*individual space*) and R^n (*variable space*) are equipped with a Euclidean metric.

Obviously, the introduction of a Euclidean metric allows us to calculate in R^p and R^n, in addition to the inner product between vectors, both the length of vectors and the distance between vectors.



2.2.1 EUCLIDEAN METRIC IN THE INDIVIDUAL SPACE

In R^p the matrix (symmetric and positive definite (p.d.)) of the Euclidean metric – with respect to the basis consisting of the p canonical vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$ – is generally of the form

$$\mathbf{Q} = \mathrm{diag}\,(q_1, \ldots, q_p)$$

where $q_j > 0$ $(j = 1, \ldots, p)$ represents the *weight* given to the jth variable and denotes its «importance» in the set of the p variables [(2)].

The choice of the weights $q_1, \ldots, q_p$ generally depends on the measurement units and/or the variances of the variables $\mathbf{x}_1, \ldots, \mathbf{x}_p$.

The situations which may occur are:

- the variables $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are expressed in the same measurement unit and present approximatively the same variance;
- the variables $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are expressed in the same measurement unit but present considerably different variances;
- the variables $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are expressed in different measurement units.

(2) In contexts differing from PCA (e.g., canonical correlation analysis), the metric is specified in other ways.

In the first case, the weights $q_1, \dots, q_p$ are usually chosen setting $q_1 = \dots = q_p = 1$ – namely, assuming that each variable has the same importance of all the others – and thus $\mathbf{Q} = \mathbf{I}_p$.

In the remaining two cases, we often choose the weights $q_1, \dots, q_p$ as the reciprocals of the variances of the variables $\mathbf{x}_1, \dots, \mathbf{x}_p$. The meaning of this choice will be explained below (Section 5.3).

2.2.2 EUCLIDEAN METRIC IN THE VARIABLE SPACE

In R^n the matrix (symmetric and p.d.) of the Euclidean metric – 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.

Whenever we do not have sufficient indications about the differing importance of the n individuals, we can set $m_1 = \dots = m_n = m^*$ from which – taking into account the condition $\Sigma_i m_i = 1$ – we obtain $m^* = \frac{1}{n}$ and thus

$$\mathbf{M} = \mathrm{diag}\left(\frac{1}{n}, \dots, \frac{1}{n}\right).$$

2.3 CENTRED DATA MATRICES

2.3.1 GENERAL CENTRED DATA MATRIX

Given any vector

$$\mathbf{c}_* = \begin{bmatrix} c_{*1} \\ \vdots \\ c_{*p} \end{bmatrix},$$

consider the matrix (*general centred data matrix*)

$$\mathbf{Z} = \mathbf{X} - \mathbf{u}\,\mathbf{c}_*' = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \cdots & \cdots & \cdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} - \begin{bmatrix} c_{*1} & \cdots & c_{*p} \\ \cdots & \cdots & \cdots \\ c_{*1} & \cdots & c_{*p} \end{bmatrix} = \begin{bmatrix} x_{11} - c_{*1} & \cdots & x_{1p} - c_{*p} \\ \cdots & \cdots & \cdots \\ x_{n1} - c_{*1} & \cdots & x_{np} - c_{*p} \end{bmatrix}$$

where **u** is a column vector of order n with elements all equal to 1.

Then, setting (i = 1, ... , n)

$$\mathbf{z}_i = \begin{bmatrix} x_{i1} - c_{*1} \\ \vdots \\ x_{ip} - c_{*p} \end{bmatrix} = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix} - \begin{bmatrix} c_{*1} \\ \vdots \\ c_{*p} \end{bmatrix} = \boldsymbol{x}_i - \mathbf{c}_*$$

and (j = 1, ... , p)

$$\mathbf{z}_j = \begin{bmatrix} x_{1j} - c_{*j} \\ \vdots \\ x_{nj} - c_{*j} \end{bmatrix} = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{bmatrix} - \begin{bmatrix} c_{*j} \\ \vdots \\ c_{*j} \end{bmatrix} = \mathbf{x}_j - \bar{\mathbf{c}}_{*j} ,$$

we can write

$$\mathbf{Z}' = \begin{bmatrix} \boldsymbol{z}_1 \cdots \mathbf{z}_n \end{bmatrix}$$

and

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 \cdots \mathbf{z}_p \end{bmatrix} .$$

2.3.2 MEAN CENTRED DATA MATRIX

Let

$$\mathbf{g} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix}$$

where $\bar{x}_j = \sum_i m_i x_{ij}$ is the (weighted) arithmetic mean of the variable $\mathbf{x}_j$.

Notice that we can write

$$\mathbf{g} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \sum_i m_i x_{i1} \\ \vdots \\ \sum_i m_i x_{ip} \end{bmatrix} = \begin{bmatrix} x_{11} \cdots x_{n1} \\ \cdots \cdots \cdots \\ x_{1p} \cdots x_{np} \end{bmatrix} \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{x}_1 \cdots \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_p' \end{bmatrix} \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix} = \mathbf{X}'\mathbf{M}\mathbf{u} .$$

The vector **g** is called the *barycentre* (*centroid*) of the n individuals

$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ or the *mean vector* of the p variables $\mathbf{x}_1, \ldots, \mathbf{x}_p$.

Next, consider the matrix (*mean centred data matrix*)

$$\mathbf{Y} = \mathbf{X} - \mathbf{u}\,\mathbf{g}' = \begin{bmatrix} x_{11} \cdots x_{1p} \\ \cdots \cdots \cdots \\ x_{n1} \cdots x_{np} \end{bmatrix} - \begin{bmatrix} \overline{x}_1 \cdots \overline{x}_p \\ \cdots \cdots \cdots \\ \overline{x}_1 \cdots \overline{x}_p \end{bmatrix} = \begin{bmatrix} x_{11} - \overline{x}_1 \cdots x_{1p} - \overline{x}_p \\ \cdots \quad \cdots \quad \cdots \\ x_{n1} - \overline{x}_1 \cdots x_{n1} - \overline{x}_p \end{bmatrix}.$$

Then, setting $(i = 1, \ldots, n)$

$$\boldsymbol{y}_i = \begin{bmatrix} x_{i1} - \overline{x}_1 \\ \vdots \\ x_{ip} - \overline{x}_p \end{bmatrix} = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix} - \begin{bmatrix} \overline{x}_1 \\ \vdots \\ \overline{x}_p \end{bmatrix} = \boldsymbol{x}_i - \mathbf{g}$$

and $(j = 1, \ldots, p)$

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

we can write

$$\mathbf{Y}' = \begin{bmatrix} \boldsymbol{y}_1 \cdots \boldsymbol{y}_n \end{bmatrix}$$

and

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \cdots \mathbf{y}_p \end{bmatrix}.$$

Taking into account the notation just introduced, we say that $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ and $\mathbf{y}_1, \ldots, \mathbf{y}_p$ represent, respectively, the n individuals and the p variables (measured in terms of deviations from the means).

Of course, the (weighted) arthmetic mean of each $\mathbf{y}_j$ $(j = 1, \ldots, p)$ is zero.

REMARK 1. Notice that

$$\begin{aligned} \mathbf{Z} = \mathbf{X} - \mathbf{u}\,\mathbf{c}_*' \qquad\qquad & = \mathbf{X} - \mathbf{u}\,(\mathbf{c}_* + \mathbf{g} - \mathbf{g})' \\ = \mathbf{X} - \mathbf{u}\,\mathbf{g}' - \mathbf{u}\,(\mathbf{c}_* - \mathbf{g})' & = \mathbf{Y} - \mathbf{u}\,(\mathbf{c}_* - \mathbf{g})'\,. \end{aligned}$$

3 PRELIMINARY CONCEPTS IN THE INDIVIDUAL SPACE

3.1 INERTIA RELATIVE TO A VECTOR

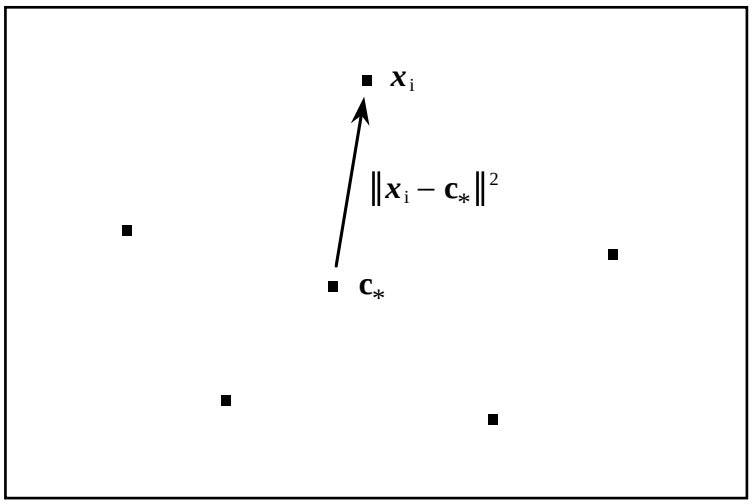
Consider the n individuals $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathrm{R}^p$ with weights given, respectively, by $\mathrm{m}_1, \ldots, \mathrm{m}_n$ and a generic vector $\mathbf{c}_* \in \mathrm{R}^p$.

The quantity

$$\mathrm{I}_{\mathbf{c}_*} = \Sigma_i \, \mathrm{m}_i \left\| \boldsymbol{x}_i - \mathbf{c}_* \right\|^2 = \Sigma_i \, \mathrm{m}_i (\boldsymbol{x}_i - \mathbf{c}_*)' \mathbf{Q} (\boldsymbol{x}_i - \mathbf{c}_*)$$

is called the *inertia of* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ *relative to* $\mathbf{c}_*$ and represents a (weighted) dispersion measure of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ with respect to $\mathbf{c}_*$ (Fig. 1) [3].

Fig. 1



In turn, the quantity

$$\mathrm{I}_{\mathbf{g}} = \Sigma_i \, \mathrm{m}_i \left\| \boldsymbol{x}_i - \mathbf{g} \right\|^2 = \Sigma_i \, \mathrm{m}_i (\boldsymbol{x}_i - \mathbf{g})' \mathbf{Q} (\boldsymbol{x}_i - \mathbf{g})$$

is called the *inertia of* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ *relative to the barycentre*.

Notice that, taking into account the notations introduced above (Sections 2.3.1 and 2.3.2), we can write

(3) Some of the concepts introduced in this Section 3 from the point of view of the individual space will be reinterpreted in the following Section 4 from the point of view of the variable space.

$$
\begin{aligned}
\mathrm{I}_{\mathbf{c}_*} &= \Sigma_{\mathrm{i}}\, \mathrm{m}_{\mathrm{i}}\, \boldsymbol{z}_{\mathrm{i}}'\mathbf{Q}\boldsymbol{z}_{\mathrm{i}} && = \mathrm{tr}\left\{\begin{bmatrix} \mathrm{m}_1\, \boldsymbol{z}_1'\mathbf{Q}\boldsymbol{z}_1 & \cdots & \mathrm{m}_1\, \boldsymbol{z}_1'\mathbf{Q}\boldsymbol{z}_{\mathrm{n}} \\ \cdots & \cdots & \cdots \\ \mathrm{m}_{\mathrm{n}}\, \boldsymbol{z}_{\mathrm{n}}'\mathbf{Q}\boldsymbol{z}_1 & \cdots & \mathrm{m}_{\mathrm{n}}\, \boldsymbol{z}_{\mathrm{n}}'\mathbf{Q}\boldsymbol{z}_{\mathrm{n}} \end{bmatrix}\right\} \\
&= \mathrm{tr}\left\{\mathbf{M}\begin{bmatrix} \boldsymbol{z}_1'\mathbf{Q}\boldsymbol{z}_1 & \cdots & \boldsymbol{z}_1'\mathbf{Q}\boldsymbol{z}_{\mathrm{n}} \\ \cdots & \cdots & \cdots \\ \boldsymbol{z}_{\mathrm{n}}'\mathbf{Q}\boldsymbol{z}_1 & \cdots & \boldsymbol{z}_{\mathrm{n}}'\mathbf{Q}\boldsymbol{z}_{\mathrm{n}} \end{bmatrix}\right\} && = \mathrm{tr}\left\{\mathbf{M}\begin{bmatrix} \boldsymbol{z}_1' \\ \vdots \\ \boldsymbol{z}_{\mathrm{n}}' \end{bmatrix}\mathbf{Q}\begin{bmatrix} \boldsymbol{z}_1 \cdots \boldsymbol{z}_{\mathrm{n}} \end{bmatrix}\right\} \\
&= \mathrm{tr}\left\{\mathbf{M}\mathbf{Z}\mathbf{Q}\mathbf{Z}'\right\} && = \mathrm{tr}\left\{\mathbf{Z}'\mathbf{M}\mathbf{Z}\mathbf{Q}\right\} \\
&= \mathrm{tr}\left\{\mathbf{V}_{\mathbf{c}_*}\mathbf{Q}\right\}
\end{aligned}
$$

and, analogously,

$$
\begin{aligned}
\mathrm{I}_{\mathbf{g}} &= \Sigma_{\mathrm{i}}\, \mathrm{m}_{\mathrm{i}}\, \boldsymbol{y}_{\mathrm{i}}'\mathbf{Q}\boldsymbol{y}_{\mathrm{i}} = \mathrm{tr}\left\{\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{Q}\right\} \\
&= \mathrm{tr}\left\{\mathbf{V}_{\mathbf{g}}\mathbf{Q}\right\}.
\end{aligned}
$$

The matrices

$$
\mathbf{V}_{\mathbf{c}_*} = \begin{bmatrix} \boldsymbol{z}_1 \cdots \boldsymbol{z}_{\mathrm{n}} \end{bmatrix}\mathbf{M}\begin{bmatrix} \boldsymbol{z}_1' \\ \vdots \\ \boldsymbol{z}_{\mathrm{n}}' \end{bmatrix} = \mathbf{Z}'\mathbf{M}\mathbf{Z} \quad , \quad \mathbf{V}_{\mathbf{g}} = \begin{bmatrix} \boldsymbol{y}_1 \cdots \boldsymbol{y}_{\mathrm{n}} \end{bmatrix}\mathbf{M}\begin{bmatrix} \boldsymbol{y}_1' \\ \vdots \\ \boldsymbol{y}_{\mathrm{n}}' \end{bmatrix} = \mathbf{Y}'\mathbf{M}\mathbf{Y}
$$

denote the so-called *inertia matrices of* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\mathrm{n}}$ *relative,* respectively, *to* $\mathbf{c}_*$ and $\mathbf{g}$.

3.1.1 HUIGHENS THEOREM

Taking into account that $\mathbf{Z} = \mathbf{Y} - \mathbf{u}(\mathbf{c}_* - \mathbf{g})'$ (Remark 1) and that, as can easily be verified, $\mathbf{Y}'\mathbf{M}\mathbf{u} = \mathbf{0}$, we have

$$
\begin{aligned}
\mathbf{V}_{\mathbf{c}_*} &= \mathbf{Z}'\mathbf{M}\mathbf{Z} \\
&= (\mathbf{Y} - \mathbf{u}(\mathbf{c}_* - \mathbf{g})')'\mathbf{M}(\mathbf{Y} - \mathbf{u}(\mathbf{c}_* - \mathbf{g})') \\
&= \mathbf{Y}'\mathbf{M}\mathbf{Y} - \mathbf{Y}'\mathbf{M}\mathbf{u}(\mathbf{c}_* - \mathbf{g})' - (\mathbf{c}_* - \mathbf{g})\,\mathbf{u}'\mathbf{M}\mathbf{Y} + (\mathbf{c}_* - \mathbf{g})\,\mathbf{u}'\mathbf{M}\mathbf{u}(\mathbf{c}_* - \mathbf{g})' \\
&= \mathbf{V}_{\mathbf{g}} + (\mathbf{c}_* - \mathbf{g})(\mathbf{c}_* - \mathbf{g})'.
\end{aligned}
$$

Thus, we get (*Huighens theorem*)

$$I_{\mathbf{c}_*} = \mathrm{tr}\left\{\mathbf{V}_{\mathbf{c}_*}\mathbf{Q}\right\} = \mathrm{tr}\left\{[\mathbf{V}_{\mathbf{g}} + (\mathbf{c}_* - \mathbf{g})(\mathbf{c}_* - \mathbf{g})']\mathbf{Q}\right\}$$
$$= \mathrm{tr}\left\{\mathbf{V}_{\mathbf{g}}\mathbf{Q}\right\} + (\mathbf{c}_* - \mathbf{g})'\mathbf{Q}(\mathbf{c}_* - \mathbf{g}) = I_{\mathbf{g}} + \left\|\mathbf{c}_* - \mathbf{g}\right\|^2 .$$

In other words, the inertia $I_{\mathbf{c}_*}$ of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ relative to $\mathbf{c}_*$ may be split up into the sum of two addenda:

- $I_{\mathbf{g}}$ which represents the inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ relative to $\mathbf{g}$;
- $\left\|\mathbf{c}_* - \mathbf{g}\right\|^2$ which represents the square distance between $\mathbf{c}_*$ and $\mathbf{g}$.

$I_{\mathbf{c}_*}$ inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ relative to $\mathbf{c}_*$	=	$I_{\mathbf{g}}$ inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ relative to $\mathbf{g}$	+	$\left\|\mathbf{c}_* - \mathbf{g}\right\|^2$ square distance between $\mathbf{c}_*$ and $\mathbf{g}$

REMARK 2. Notice that $I_{\mathbf{c}_*}$ reaches the minimum $I_{\mathbf{g}}$ when $\mathbf{c}_* = \mathbf{g}$.

3.2 INERTIA ALONG A LINEAR VARIETY

Consider in R^p a subspace C_k of dimension k ($1 \le k < p$), and its orthogonal complement $C_k^{\perp}$.

Denote the orthogonal projection matrices on C_k and $C_k^{\perp}$, respectively, by $\mathbf{P}$ and $\mathbf{I}_p - \mathbf{P}$.

Of course, both these matrices are idempotent and selfadjoint, namely

$$\mathbf{P}^2 = \mathbf{P} \quad , \quad \mathbf{P}'\mathbf{Q} = \mathbf{Q}\mathbf{P}$$
$$(\mathbf{I}_p - \mathbf{P})^2 = (\mathbf{I}_p - \mathbf{P}) \quad , \quad (\mathbf{I}_p - \mathbf{P})'\mathbf{Q} = \mathbf{Q}(\mathbf{I}_p - \mathbf{P}) .$$

Successively, consider the linear variety $\mathbf{c}_* + C_k$ of direction C_k and translation $\mathbf{c}_*$.

Clearly, the vector $\widehat{\boldsymbol{x}}_i + (\mathbf{c}_* - \widehat{\mathbf{c}}_*)$ – where $\widehat{\boldsymbol{x}}_i = \mathbf{P}\boldsymbol{x}_i$ and $\widehat{\mathbf{c}}_* = \mathbf{P}\mathbf{c}_*$ denote the orthogonal projections, respectively, of $\boldsymbol{x}_i$ and $\mathbf{c}_*$ on C_k – is the orthogonal projection of $\boldsymbol{x}_i$ on $\mathbf{c}_* + C_k$ (Fig. 2).

The quantity

$$I_{\mathbf{c}_*+C_k} = \Sigma_i\, m_i \left\| (\widehat{\boldsymbol{x}}_i + (\mathbf{c}_* - \widehat{\mathbf{c}}_*)) - \mathbf{c}_* \right\|^2 = \Sigma_i\, m_i \left\| \widehat{\boldsymbol{x}}_i - \widehat{\mathbf{c}}_* \right\|^2$$

is called the *inertia of* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ *along* $\mathbf{c}_* + C_k$ or *explained by* $\mathbf{c}_* + C_k$ and represents a (weighted) dispersion measure of the projected vectors $\widehat{\boldsymbol{x}}_1 + (\mathbf{c}_* - \widehat{\mathbf{c}}_*)$, ... , $\widehat{\boldsymbol{x}}_n + (\mathbf{c}_* - \widehat{\mathbf{c}}_*)$ with respect to $\mathbf{c}_*$.

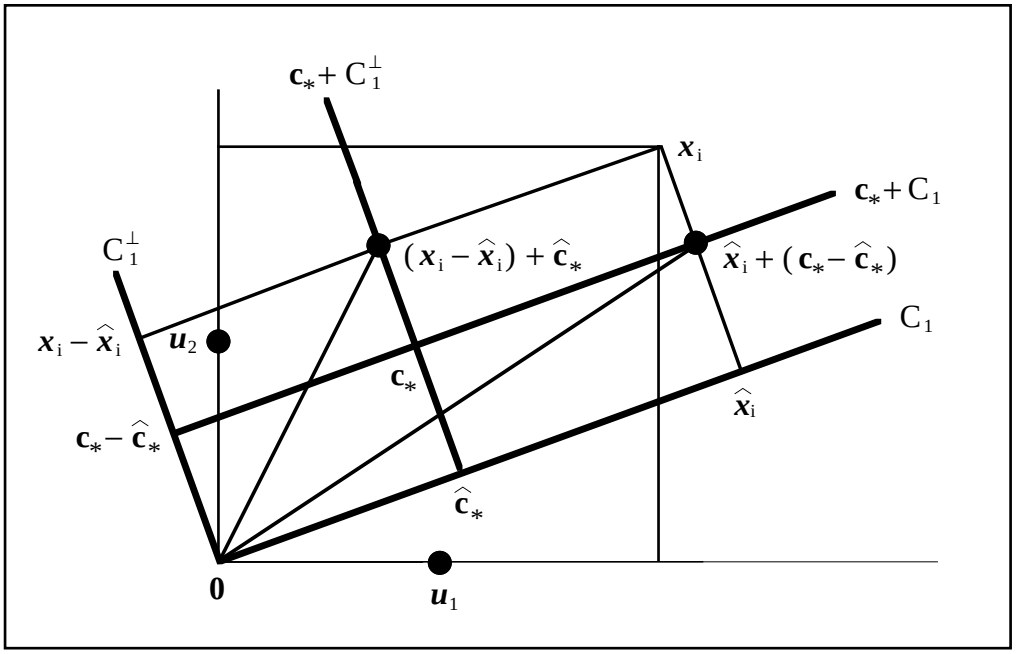
Analogously, the vector $(\boldsymbol{x}_i - \widehat{\boldsymbol{x}}_i) + \widehat{\mathbf{c}}_*$ is the orthogonal projection of $\boldsymbol{x}_i$ on the linear variety $\mathbf{c}_* + C_k^{\perp}$ of direction $C_k^{\perp}$ and translation $\mathbf{c}_*$.

The quantity

$$I_{\mathbf{c}_*+C_k^{\perp}} = \Sigma_i\, m_i \left\| ((\boldsymbol{x}_i - \widehat{\boldsymbol{x}}_i) + \widehat{\mathbf{c}}_*) - \mathbf{c}_* \right\|^2 = \Sigma_i\, m_i \left\| \boldsymbol{x}_i - (\widehat{\boldsymbol{x}}_i + (\mathbf{c}_* - \widehat{\mathbf{c}}_*)) \right\|^2$$

is called the *inertia of* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ *along* $\mathbf{c}_* + C_k^{\perp}$ or *not explained by* $\mathbf{c}_* + C_k$ and represents a (weighted) dispersion measure of the projected vectors $(\boldsymbol{x}_1 - \widehat{\boldsymbol{x}}_1) + \widehat{\mathbf{c}}_*$, ... , $(\boldsymbol{x}_n - \widehat{\boldsymbol{x}}_n) + \widehat{\mathbf{c}}_*$ with respect to $\mathbf{c}_*$.

Fig. 2



Now, consider the linear variety $\mathbf{g} + C_k$ of direction C_k and translation $\mathbf{g}$.

The vector $\widehat{\boldsymbol{x}}_i + (\mathbf{g} - \widehat{\mathbf{g}})$ – where $\widehat{\boldsymbol{x}}_i = \mathbf{P}\boldsymbol{x}_i$ and $\widehat{\mathbf{g}} = \mathbf{P}\mathbf{g}$ denote the orthogonal projections, respectively, of $\boldsymbol{x}_i$ and $\mathbf{g}$ on C_k – is the orthogonal

projection of $\boldsymbol{x}_i$ on $\mathbf{g} + C_k$ (Fig. 3).

The quantity

$$I_{\mathbf{g}+C_k} = \Sigma_i \, m_i \left\| (\hat{\boldsymbol{x}}_i + (\mathbf{g} - \hat{\mathbf{g}})) - \mathbf{g} \right\|^2 = \Sigma_i \, m_i \left\| \hat{\boldsymbol{x}}_i - \hat{\mathbf{g}} \right\|^2$$

is called the *inertia of* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ *along* $\mathbf{g} + C_k$ or *explained by* $\mathbf{g} + C_k$ and represents a (weighted) dispersion measure of the projected vectors $\hat{\boldsymbol{x}}_1 + (\mathbf{g} - \hat{\mathbf{g}}), \ldots, \hat{\boldsymbol{x}}_n + (\mathbf{g} - \hat{\mathbf{g}})$ with respect to $\mathbf{g}$.

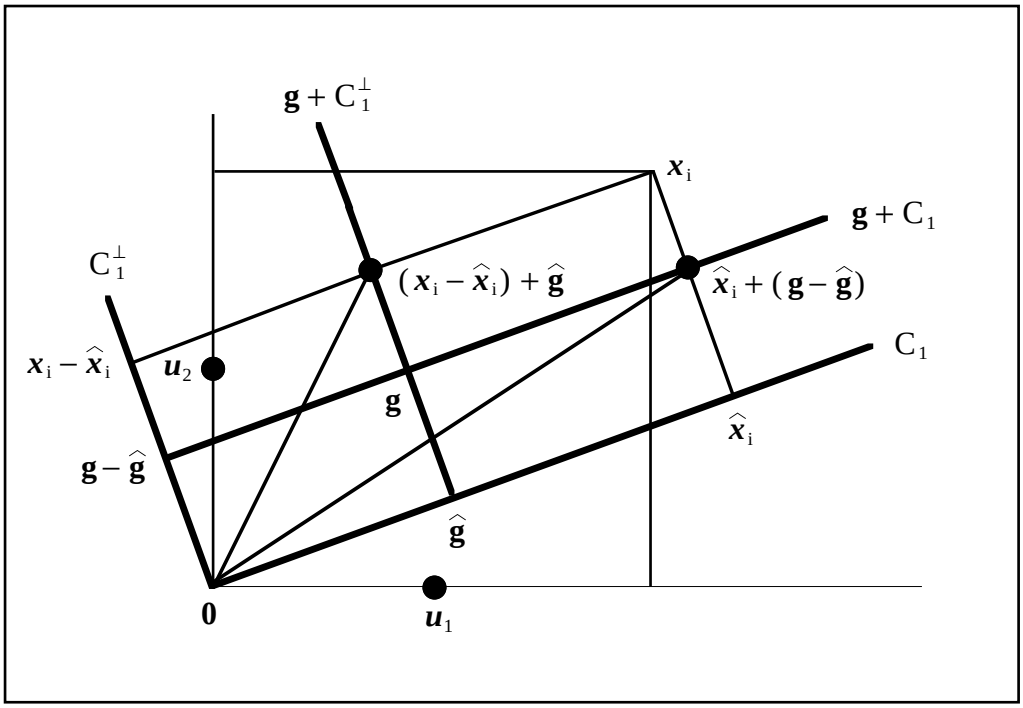
Analogously, the vector $(\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i) + \hat{\mathbf{g}}$ is the orthogonal projection of $\boldsymbol{x}_i$ on the linear variety $\mathbf{g} + C_k^{\perp}$ of direction $C_k^{\perp}$ and translation $\mathbf{g}$.

The quantity

$$I_{\mathbf{g}+C_k^{\perp}} = \Sigma_i \, m_i \left\| ((\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i) + \hat{\mathbf{g}}) - \mathbf{g} \right\|^2 = \Sigma_i \, m_i \left\| \boldsymbol{x}_i - (\hat{\boldsymbol{x}}_i + (\mathbf{g} - \hat{\mathbf{g}})) \right\|^2$$

is called the *inertia of* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ *along* $\mathbf{g} + C_k^{\perp}$ or *not explained by* $\mathbf{g} + C_k$ and represents a (weighted) dispersion measure of the projected vectors $(\boldsymbol{x}_1 - \hat{\boldsymbol{x}}_1) + \hat{\mathbf{g}}, \ldots, (\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n) + \hat{\mathbf{g}}$ with respect to $\mathbf{g}$.

Fig. 3



3.2.1 A DECOMPOSITION OF THE INERTIA RELATIVE TO A VECTOR

Firstly, notice that we can write

$$\begin{aligned} I_{\mathbf{c}_*} &= \mathrm{tr}\left\{\mathbf{V}_{\mathbf{c}_*}\mathbf{Q}\right\} && = \mathrm{tr}\left\{\mathbf{V}_{\mathbf{c}_*}\mathbf{Q}\mathbf{P} + \mathbf{V}_{\mathbf{c}_*}\mathbf{Q}(\mathbf{I}_p - \mathbf{P})\right\} \\ &= \mathrm{tr}\left\{\mathbf{V}_{\mathbf{c}_*}\mathbf{Q}\mathbf{P}\right\} + \mathrm{tr}\left\{\mathbf{V}_{\mathbf{c}_*}\mathbf{Q}(\mathbf{I}_p - \mathbf{P})\right\} . \end{aligned}$$

Thus – since we have ($\mathbf{P}'\mathbf{Q}\mathbf{P} = \mathbf{Q}\mathbf{P}$)

$$\begin{aligned} I_{\mathbf{c}_*+C_k} &= \Sigma_i\, m_i \left\|\mathbf{P}\boldsymbol{x}_i - \mathbf{P}\mathbf{c}_*\right\|^2 && = \Sigma_i\, m_i \left\|\mathbf{P}(\boldsymbol{x}_i - \mathbf{c}_*)\right\|^2 \\ &= \Sigma_i\, m_i \left\|\mathbf{P}\boldsymbol{z}_i\right\|^2 && = \Sigma_i\, m_i\, \boldsymbol{z}_i'\mathbf{P}'\mathbf{Q}\mathbf{P}\,\boldsymbol{z}_i \\ &= \mathrm{tr}\left\{ \begin{bmatrix} m_1\, \boldsymbol{z}_1'\mathbf{Q}\mathbf{P}\,\boldsymbol{z}_1 & \cdots & m_1\, \boldsymbol{z}_1'\mathbf{Q}\mathbf{P}\,\boldsymbol{z}_n \\ \cdots & \cdots & \cdots \\ m_n\, \boldsymbol{z}_n'\mathbf{Q}\mathbf{P}\,\boldsymbol{z}_1 & \cdots & m_n\, \boldsymbol{z}_n'\mathbf{Q}\mathbf{P}\,\boldsymbol{z}_n \end{bmatrix} \right\} && = \mathrm{tr}\left\{ \mathbf{M} \begin{bmatrix} \boldsymbol{z}_1' \\ \vdots \\ \boldsymbol{z}_n' \end{bmatrix} \mathbf{Q}\mathbf{P} \begin{bmatrix} \boldsymbol{z}_1 \cdots \boldsymbol{z}_n \end{bmatrix} \right\} \\ &= \mathrm{tr}\left\{\mathbf{M}\mathbf{Z}\mathbf{Q}\mathbf{P}\mathbf{Z}'\right\} && = \mathrm{tr}\left\{\mathbf{Z}'\mathbf{M}\mathbf{Z}\mathbf{Q}\mathbf{P}\right\} \\ &= \mathrm{tr}\left\{\mathbf{V}_{\mathbf{c}_*}\mathbf{Q}\mathbf{P}\right\} \end{aligned}$$

and, analogously,

$$I_{\mathbf{c}_*+C_k^{\perp}} = \mathrm{tr}\left\{\mathbf{V}_{\mathbf{c}_*}\mathbf{Q}(\mathbf{I}_p - \mathbf{P})\right\}$$

– we obtain the decomposition

$$I_{\mathbf{c}_*} = I_{\mathbf{c}_*+C_k} + I_{\mathbf{c}_*+C_k^{\perp}}$$

Namely, the inertia $I_{\mathbf{c}_*}$ of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ relative to $\mathbf{c}_*$ may be split up into the sum of two addenda:

- $I_{\mathbf{c}_*+C_k}$ which is the inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ explained by $\mathbf{c}_* + C_k$;
- $I_{\mathbf{c}_*+C_k^{\perp}}$ which is the inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ not explained by $\mathbf{c}_* + C_k$.

$I_{\mathbf{c}_*}$ inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ relative to $\mathbf{c}_*$	=	$I_{\mathbf{c}_*+C_k}$ inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ explained by $\mathbf{c}_* + C_k$	+	$I_{\mathbf{c}_*+C_k^{\perp}}$ inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ not explained by $\mathbf{c}_* + C_k$

Of course, for $\mathbf{c}_* = \mathbf{g}$, we get

$$I_{\mathbf{g}} = I_{\mathbf{g}+C_k} + I_{\mathbf{g}+C_k^{\perp}}$$

where $(\widehat{\mathbf{g}} = \mathbf{P}\mathbf{g})$

$$I_{\mathbf{g}+C_k} = \Sigma_i\, m_i \left\| \widehat{x}_i - \widehat{\mathbf{g}} \right\|^2 = \mathrm{tr}\left\{ \mathbf{V}_{\mathbf{g}}\mathbf{Q}\mathbf{P} \right\},$$
$$I_{\mathbf{g}+C_k^{\perp}} = \Sigma_i\, m_i \left\| (x_i - \widehat{x}_i) + \widehat{\mathbf{g}} - \mathbf{g} \right\|^2 = \mathrm{tr}\left\{ \mathbf{V}_{\mathbf{g}}\mathbf{Q}(\mathbf{I}_p - \mathbf{P}) \right\}.$$

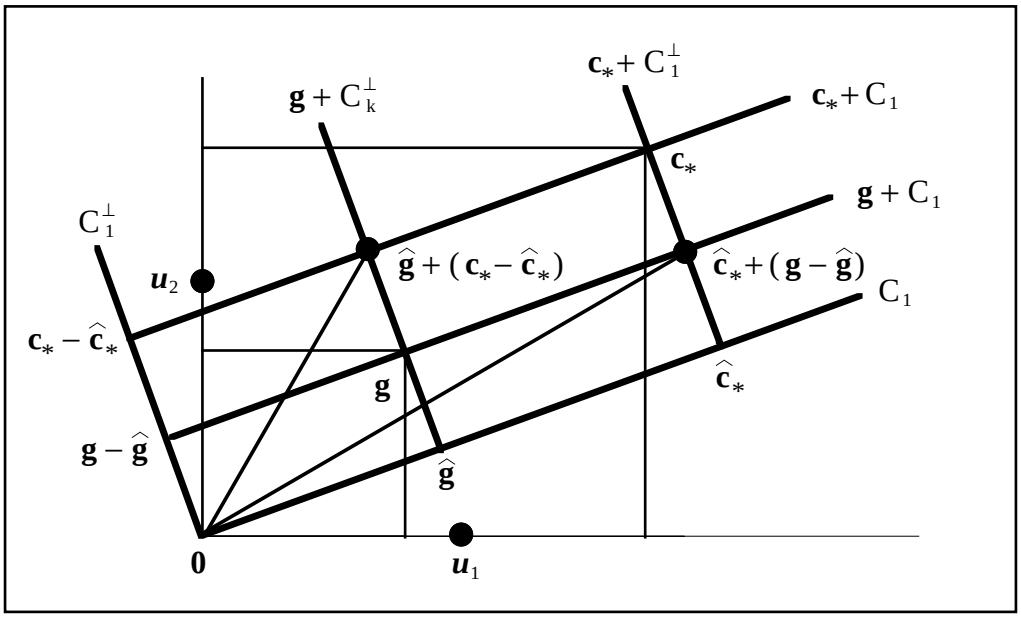
$I_{\mathbf{g}}$ inertia of $x_1, \ldots, x_n$ relative to $\mathbf{g}$	=	$I_{\mathbf{g}+C_k}$ inertia of $x_1, \ldots, x_n$ explained by $\mathbf{g} + C_k$	+	$I_{\mathbf{g}+C_k^{\perp}}$ inertia of $x_1, \ldots, x_n$ not explained by $\mathbf{g} + C_k$

3.2.2 A DECOMPOSITION OF THE INERTIA ALONG A LINEAR VARIETY

Firstly, consider the linear varieties $\mathbf{c}_* + C_k$ and $\mathbf{g} + C_k$. The vector $\widehat{\mathbf{c}}_* + (\mathbf{g} - \widehat{\mathbf{g}})$ is the orthogonal projection of $\mathbf{c}_*$ on the linear variety $\mathbf{g} + C_k$ (Fig. 4). Hence, the square distance between $\mathbf{c}_* + C_k$ and $\mathbf{g} + C_k$ is

$$\left\| \mathbf{c}_* - \widehat{\mathbf{c}}_* - (\mathbf{g} - \widehat{\mathbf{g}}) \right\|^2 = \left\| (\mathbf{c}_* - \mathbf{g}) - (\widehat{\mathbf{c}}_* - \widehat{\mathbf{g}}) \right\|^2.$$

Fig. 4



Successively, consider the linear varieties $\mathbf{c}_* + \mathrm{C}_\mathrm{k}^\perp$ and $\mathbf{g} + \mathrm{C}_\mathrm{k}^\perp$. The vector $\widehat{\mathbf{g}} + (\mathbf{c}_* - \widehat{\mathbf{c}}_*)$ is the orthogonal projection of $\mathbf{c}_*$ on the linear variety $\mathbf{g} + \mathrm{C}_\mathrm{k}^\perp$. Therefore, the square distance between $\mathbf{c}_* + \mathrm{C}_\mathrm{k}^\perp$ and $\mathbf{g} + \mathrm{C}_\mathrm{k}^\perp$ is

$$\left\| \mathbf{c}_* - \widehat{\mathbf{g}} - (\mathbf{c}_* - \widehat{\mathbf{c}}_*) \right\|^2 = \left\| \widehat{\mathbf{c}}_* - \widehat{\mathbf{g}} \right\|^2 .$$

Now, notice that – since

$$\mathrm{I}_{\mathbf{c}_* + \mathrm{C}_\mathrm{k}} = \mathrm{I}_{\mathbf{c}_*} - \mathrm{I}_{\mathbf{c}_* + \mathrm{C}_\mathrm{k}^\perp} \quad , \quad \mathrm{I}_{\mathbf{g} + \mathrm{C}_\mathrm{k}} = \mathrm{I}_{\mathbf{g}} - \mathrm{I}_{\mathbf{g} + \mathrm{C}_\mathrm{k}^\perp} \quad , \quad \mathbf{V}_{\mathbf{c}_*} = \mathbf{V}_\mathbf{g} + (\mathbf{c}_* - \mathbf{g})(\mathbf{c}_* - \mathbf{g})'$$

– we get ($\mathbf{P}'\mathbf{Q}\mathbf{P} = \mathbf{Q}\mathbf{P}$)

$$\begin{aligned}
\mathrm{I}_{\mathbf{c}_* + \mathrm{C}_\mathrm{k}} &= \mathrm{tr}\left\{ [\mathbf{V}_\mathbf{g} + (\mathbf{c}_* - \mathbf{g})(\mathbf{c}_* - \mathbf{g})'] \mathbf{Q} \right\} - \mathrm{tr}\left\{ [\mathbf{V}_\mathbf{g} + (\mathbf{c}_* - \mathbf{g})(\mathbf{c}_* - \mathbf{g})'] \mathbf{Q}(\mathbf{I} - \mathbf{P}) \right\} \\
&= \mathrm{tr}\left\{ \mathbf{V}_\mathbf{g} \mathbf{Q} \right\} - \mathrm{tr}\left\{ \mathbf{V}_\mathbf{g} \mathbf{Q}(\mathbf{I}_\mathrm{p} - \mathbf{P}) \right\} + \mathrm{tr}\left\{ (\mathbf{c}_* - \mathbf{g})(\mathbf{c}_* - \mathbf{g})' \mathbf{Q} \right\} \\
&\quad - \mathrm{tr}\left\{ (\mathbf{c}_* - \mathbf{g})(\mathbf{c}_* - \mathbf{g})' \mathbf{Q}(\mathbf{I}_\mathrm{p} - \mathbf{P}) \right\} \\
&= \mathrm{tr}\left\{ \mathbf{V}_\mathbf{g} \mathbf{Q}\mathbf{P} \right\} + \mathrm{tr}\left\{ (\mathbf{c}_* - \mathbf{g})(\mathbf{c}_* - \mathbf{g})' \mathbf{Q}\mathbf{P} \right\} \\
&= \mathrm{I}_{\mathbf{g} + \mathrm{C}_\mathrm{k}} + \mathrm{tr}\left\{ (\mathbf{c}_* - \mathbf{g})(\mathbf{c}_* - \mathbf{g})' \mathbf{P}'\mathbf{Q}\mathbf{P} \right\} \\
&= \mathrm{I}_{\mathbf{g} + \mathrm{C}_\mathrm{k}} + (\mathbf{c}_* - \mathbf{g})' \mathbf{P}'\mathbf{Q}\mathbf{P} (\mathbf{c}_* - \mathbf{g}) \\
&= \mathrm{I}_{\mathbf{g} + \mathrm{C}_\mathrm{k}} + (\mathbf{P}\mathbf{c}_* - \mathbf{P}\mathbf{g})' \mathbf{Q} (\mathbf{P}\mathbf{c}_* - \mathbf{P}\mathbf{g}) \\
&= \mathrm{I}_{\mathbf{g} + \mathrm{C}_\mathrm{k}} + (\widehat{\mathbf{c}}_* - \widehat{\mathbf{g}})' \mathbf{Q} (\widehat{\mathbf{c}}_* - \widehat{\mathbf{g}}) \\
&= \mathrm{I}_{\mathbf{g} + \mathrm{C}_\mathrm{k}} + \left\| \widehat{\mathbf{c}}_* - \widehat{\mathbf{g}} \right\|^2 .
\end{aligned}$$

Namely, the inertia $\mathrm{I}_{\mathbf{c}_* + \mathrm{C}_\mathrm{k}}$ of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\mathrm{n}$ explained by $\mathbf{c}_* + \mathrm{C}_\mathrm{k}$ may be split up into the sum of two addenda:

- $\mathrm{I}_{\mathbf{g} + \mathrm{C}_\mathrm{k}}$ which is the inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\mathrm{n}$ explained by $\mathbf{g} + \mathrm{C}_\mathrm{k}$;
- the square distance between the linear varieties $\mathbf{c}_* + \mathrm{C}_\mathrm{k}^\perp$ and $\mathbf{g} + \mathrm{C}_\mathrm{k}^\perp$.

$\mathrm{I}_{\mathbf{c}_* + \mathrm{C}_\mathrm{k}}$ inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\mathrm{n}$ explained by $\mathbf{c}_* + \mathrm{C}_\mathrm{k}$	=	$\mathrm{I}_{\mathbf{g} + \mathrm{C}_\mathrm{k}}$ inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\mathrm{n}$ explained by $\mathbf{g} + \mathrm{C}_\mathrm{k}$	+	square distance between $\mathbf{c}_* + \mathrm{C}_\mathrm{k}^\perp$, $\mathbf{g} + \mathrm{C}_\mathrm{k}^\perp$

Analogously, as can easily be verified, we obtain the decomposition

$$I_{\mathbf{c}_*+C_k^{\perp}} = I_{\mathbf{g}+C_k^{\perp}} + \| \mathbf{c}_* - \hat{\mathbf{c}}_* - (\mathbf{g} - \hat{\mathbf{g}}) \|^2$$

which shows that the inertia $I_{\mathbf{c}_*+C_k^{\perp}}$ of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ not explained by $\mathbf{c}_* + C_k$ may be split up into the sum of two addenda:

- $I_{\mathbf{g}+C_k^{\perp}}$ which is the inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ not explained by $\mathbf{g} + C_k$;
- the square distance between the linear varieties $\mathbf{c}_* + C_k$ and $\mathbf{g} + C_k$.

$I_{\mathbf{c}_*+C_k^{\perp}}$ inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ not explained by $\mathbf{c}_* + C_k$	=	$I_{\mathbf{g}+C_k^{\perp}}$ inertia of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ not explained by $\mathbf{g} + C_k$	+	square distance between $\mathbf{c}_* + C_k$, $\mathbf{g} + C_k$

REMARK 3. Notice that both the inertias $I_{\mathbf{c}_*+C_k}$ and $I_{\mathbf{c}_*+C_k^{\perp}}$ of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are minimized whenever the linear varieties $\mathbf{c}_* + C_k$ and $\mathbf{c}_* + C_k^{\perp}$ pass through the barycentre $\mathbf{g}$.

REMARK 4. It is immediately apparent that, if we consider the n individuals $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ (measured in terms of deviations from the means) instead of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, we can interpret:

- $I_{\mathbf{g}}$ as the inertia of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ relative to $\mathbf{0}$;
- $I_{\mathbf{g}+C_k}$ as the inertia of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ explained by the subspace C_k;
- $I_{\mathbf{g}+C_k^{\perp}}$ as the inertia of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ not explained by the subspace C_k.

4 PRELIMINARY CONCEPTS IN THE VARIABLE SPACE

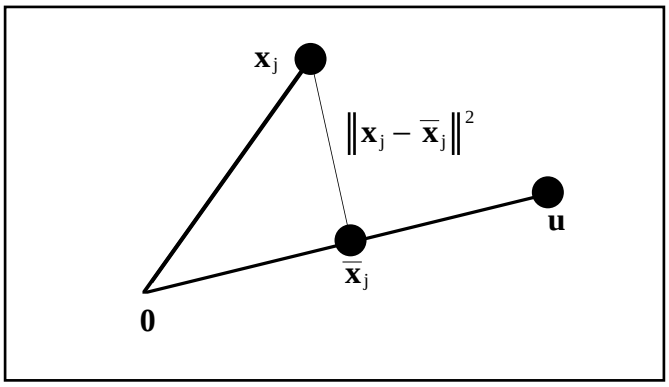
4.1 VARIANCES AND COVARIANCES

Consider the p variables $\mathbf{x}_1, \dots, \mathbf{x}_p \in R^n$ with weights given, respectively, by $q_1, \dots, q_p$.

The orthogonal projection $\overline{\mathbf{x}}_j$ of $\mathbf{x}_j$ on the subspace (of dimension 1) spanned by the vector $\mathbf{u} \in R^n$ with elements all equal to 1 is (Fig. 5)

$$\overline{\mathbf{x}}_j = \mathbf{u}\,(\mathbf{u}'\mathbf{M}\mathbf{u})^{-1}\,\mathbf{u}'\mathbf{M}\mathbf{x}_j = \mathbf{u}\,\mathbf{u}'\mathbf{M}\mathbf{x}_j = \mathbf{u}\,\overline{x}_j\,.$$

Fig. 5



The quantity

$$\sigma_j^2 = \|\mathbf{x}_j - \overline{\mathbf{x}}_j\|^2 = (\mathbf{x}_j - \overline{\mathbf{x}}_j)'\mathbf{M}(\mathbf{x}_j - \overline{\mathbf{x}}_j) = \mathbf{y}_j'\mathbf{M}\mathbf{y}_j$$

is the *variance* of $\mathbf{x}_j$ or $\mathbf{y}_j$ and the quantity $(j, t = 1, \dots, p)$

$$\sigma_{jt} = (\mathbf{x}_j - \overline{\mathbf{x}}_j)'\mathbf{M}(\mathbf{x}_t - \overline{\mathbf{x}}_t) = \mathbf{y}_j'\mathbf{M}\mathbf{y}_t$$

is the *covariance* between $\mathbf{x}_j$ and $\mathbf{x}_t$ or between $\mathbf{y}_j$ and $\mathbf{y}_t$.

In turn, the (symmetric) matrix

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1p} \\ \cdots & \cdots & \cdots \\ \sigma_{p1} & \cdots & \sigma_p^2 \end{bmatrix}$$

denotes the so-called *covariance matrix* of $\mathbf{x}_1, \dots, \mathbf{x}_p$ or $\mathbf{y}_1, \dots, \mathbf{y}_p$.

Notice that we can write

$$\mathbf{V} = \begin{bmatrix} \mathbf{y}_1'\mathbf{M}\mathbf{y}_1 & \cdots & \mathbf{y}_1'\mathbf{M}\mathbf{y}_p \\ \cdots & \cdots & \cdots \\ \mathbf{y}_p'\mathbf{M}\mathbf{y}_1 & \cdots & \mathbf{y}_p'\mathbf{M}\mathbf{y}_p \end{bmatrix} = \mathbf{Y}'\mathbf{M}\mathbf{Y}\,.$$

Hence, the covariance matrix $\mathbf{V}$ is nothing other than the inertia matrix $\mathbf{V_g}$ defined above (Section 3.1).

Moreover, we immediately realize that $\mathbf{V}$ is the Gram matrix of $\mathbf{y}_1, \ldots, \mathbf{y}_p$ and hence $\mathbf{V}$ is positive definite or positive semi-definite according to $\mathbf{y}_1, \ldots, \mathbf{y}_p$ are linearly independent or dependent [4].

Of course,

$$r(\mathbf{Y}'\mathbf{M}\mathbf{Y}) = r(\mathbf{Y}'\mathbf{M}^{1/2}\mathbf{M}^{1/2}\mathbf{Y}) = r(\mathbf{M}^{1/2}\mathbf{Y}) = r(\mathbf{Y})\,.$$

Finally, the quantity

$$J_p = \Sigma_j\, q_j \sigma_j^2 = \Sigma_j\, q_j \mathbf{y}_j'\mathbf{M}\mathbf{y}_j$$

denotes the so-called *global variability* of $\mathbf{x}_1, \ldots, \mathbf{x}_p$ or $\mathbf{y}_1, \ldots, \mathbf{y}_p$.

4.2 CORRELATIONS

The cosine of the angle formed by the vectors $\mathbf{y}_j = \mathbf{x}_j - \overline{\mathbf{x}}_j$ and $\mathbf{y}_t = \mathbf{x}_t - \overline{\mathbf{x}}_t$ $(j,t = 1, \ldots, p;\ \mathbf{y}_j, \mathbf{y}_t \neq \mathbf{0})$ is, as can easily be verified, the *linear correlation coefficient* r_{jt} between $\mathbf{x}_j$ and $\mathbf{x}_t$ or between $\mathbf{y}_j$ and $\mathbf{y}_t$; namely, we have

$$\cos(\mathbf{y}_j, \mathbf{y}_t) = \frac{(\mathbf{x}_j - \overline{\mathbf{x}}_j)'\mathbf{M}(\mathbf{x}_t - \overline{\mathbf{x}}_t)}{\sigma_j \sigma_t} = \frac{\sigma_{jt}}{\sigma_j \sigma_t} = r_{jt}\,.$$

In turn, the matrix

$$\mathbf{R} = \mathbf{Q}_{1/\sigma}\,\mathbf{Y}'\mathbf{M}\mathbf{Y}\,\mathbf{Q}_{1/\sigma}\,,$$

where

(4) Notice that, in general, $r(\mathbf{Y}) \neq r(\mathbf{X})$.

$$\mathbf{Q}_{1/\sigma} = \mathrm{diag}\left(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_p}\right),$$

represents the so-called *correlation matrix* of $\mathbf{x}_1, \dots, \mathbf{x}_p$ or $\mathbf{y}_1, \dots, \mathbf{y}_p$.

Finally, consider a generic variable $\mathbf{x} \in \mathrm{R}^n$ and the corresponding vector $\mathbf{y} = \mathbf{x} - \mathbf{u}\overline{x}$, where $\overline{x} = \Sigma_i m_i x_i$ is the (weighted) arithmetic mean of $\mathbf{x}$.

The orthogonal projection $\widehat{\mathbf{y}}$ of $\mathbf{y}$ on the subspace spanned by the p vectors $\mathbf{y}_1, \dots, \mathbf{y}_p$, assuming that $r(\mathbf{Y}) = p$, is given by

$$\widehat{\mathbf{y}} = \mathbf{Y}(\mathbf{Y}'\mathbf{M}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{M}\mathbf{y} = \mathbf{P}_{\mathbf{Y}}\mathbf{y}.$$

The square cosine of the angle formed by the vectors $\mathbf{y}$ and $\widehat{\mathbf{y}}$ denotes the *square multiple linear correlation coefficient* (*linear determination coefficient*) ρ between $\mathbf{x}$ and $\mathbf{x}_1, \dots, \mathbf{x}_p$ or between $\mathbf{y}$ and $\mathbf{y}_1, \dots, \mathbf{y}_p$.

In fact, since ($\mathbf{M}\mathbf{P}_{\mathbf{Y}} = \mathbf{P}_{\mathbf{Y}}'\mathbf{M}$)

$$\mathbf{y}'\mathbf{M}\widehat{\mathbf{y}} = \mathbf{y}'\mathbf{M}\mathbf{P}_{\mathbf{Y}}\mathbf{y} = \mathbf{y}'\mathbf{M}\mathbf{P}_{\mathbf{Y}}\mathbf{P}_{\mathbf{Y}}\mathbf{y} = \mathbf{y}'\mathbf{P}_{\mathbf{Y}}'\mathbf{M}\mathbf{P}_{\mathbf{Y}}\mathbf{y} = \widehat{\mathbf{y}}'\mathbf{M}\widehat{\mathbf{y}},$$

we can write

$$\cos^2(\mathbf{y}, \widehat{\mathbf{y}}) = \frac{(\mathbf{y}'\mathbf{M}\widehat{\mathbf{y}})^2}{(\mathbf{y}'\mathbf{M}\mathbf{y})(\widehat{\mathbf{y}}'\mathbf{M}\widehat{\mathbf{y}})} = \frac{(\widehat{\mathbf{y}}'\mathbf{M}\widehat{\mathbf{y}})^2}{(\mathbf{y}'\mathbf{M}\mathbf{y})(\widehat{\mathbf{y}}'\mathbf{M}\widehat{\mathbf{y}})} = \frac{\widehat{\mathbf{y}}'\mathbf{M}\widehat{\mathbf{y}}}{\mathbf{y}'\mathbf{M}\mathbf{y}} = \rho.$$

4.3 INTERPRETATION OF SOME CONCEPTS OF INERTIA

4.3.1 INERTIA RELATIVE TO THE BARYCENTRE

We want to show that the inertia $I_{\mathbf{g}}$ (Section 3.1) is nothing other that the global variability J_p defined above (Section 4.1).

In fact,

$$\begin{aligned} I_{\mathbf{g}} &= \mathrm{tr}\{\mathbf{V}\mathbf{Q}\} && = \mathrm{tr}\{\mathbf{Q}\mathbf{V}\} \\ &= \mathrm{tr}\left\{\mathbf{Q}\begin{bmatrix} \mathbf{y}_1'\mathbf{M}\mathbf{y}_1 & \cdots & \mathbf{y}_1'\mathbf{M}\mathbf{y}_p \\ \cdots & \cdots & \cdots \\ \mathbf{y}_p'\mathbf{M}\mathbf{y}_1 & \cdots & \mathbf{y}_p'\mathbf{M}\mathbf{y}_p \end{bmatrix}\right\} &&= \mathrm{tr}\left\{\begin{bmatrix} q_1\,\mathbf{y}_1'\mathbf{M}\mathbf{y}_1 & \cdots & q_1\,\mathbf{y}_1'\mathbf{M}\mathbf{y}_p \\ \cdots & \cdots & \cdots \\ q_p\,\mathbf{y}_p'\mathbf{M}\mathbf{y}_1 & \cdots & q_p\,\mathbf{y}_p'\mathbf{M}\mathbf{y}_p \end{bmatrix}\right\} \\ &= \Sigma_j\, q_j\,\mathbf{y}_j'\mathbf{M}\mathbf{y}_j && = J_p. \end{aligned}$$

4.3.2 INERTIA ALONG A LINEAR VARIETY THROUGH THE BARYCENTRE

Consider again the n individuals $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathrm{R}^p$ and the orthogonal projection matrix $\mathbf{P}$ on C_k.

Let

$$\mathbf{P}\mathbf{X}' = \mathbf{P}\left[\boldsymbol{x}_1 \cdots \boldsymbol{x}_n\right] = \left[\mathbf{P}\boldsymbol{x}_1 \cdots \mathbf{P}\boldsymbol{x}_n\right] = \left[\hat{\boldsymbol{x}}_1 \cdots \hat{\boldsymbol{x}}_n\right] = \begin{bmatrix} \hat{x}_{11} \cdots \hat{x}_{n1} \\ \cdots \cdots \cdots \\ \hat{x}_{1p} \cdots \hat{x}_{np} \end{bmatrix}.$$

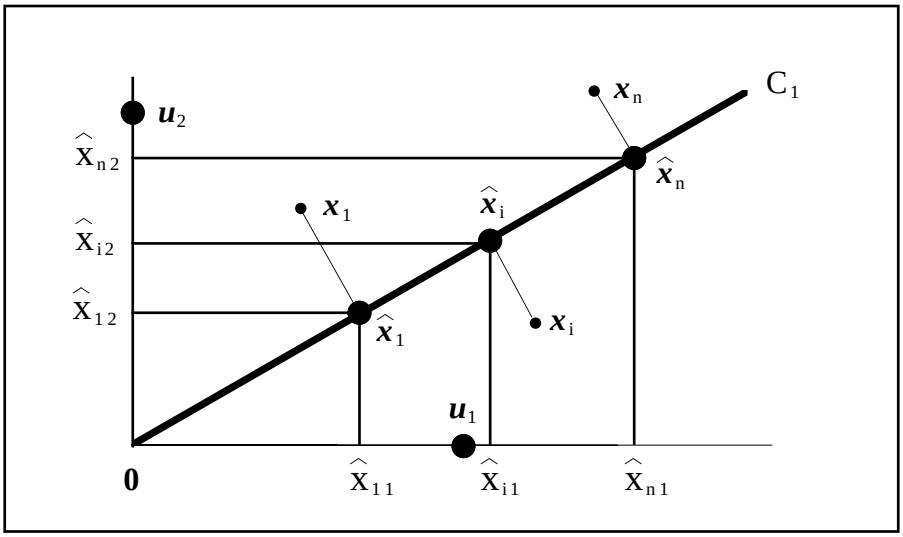
Notice that $\hat{x}_{i1}, \ldots, \hat{x}_{ip}$ represent the co-ordinates of $\hat{\boldsymbol{x}}_i$ (Fig. 6).

Moreover, we can write

$$\mathbf{X}\mathbf{P}' = \begin{bmatrix} \hat{x}_{11} \cdots \hat{x}_{1p} \\ \cdots \cdots \cdots \\ \hat{x}_{n1} \cdots \hat{x}_{np} \end{bmatrix} = \begin{bmatrix} x_{11} \cdots x_{1p} \\ \cdots \cdots \cdots \\ x_{n1} \cdots x_{np} \end{bmatrix} \begin{bmatrix} p_{11} \cdots p_{p1} \\ \cdots \cdots \cdots \\ p_{1p} \cdots p_{pp} \end{bmatrix} = \left[\mathbf{X}\mathbf{p}_1 \cdots \mathbf{X}\mathbf{p}_p\right]$$

where $\mathbf{p}_j$ $(j = 1, \ldots, p)$ denotes the jth column of $\mathbf{P}'$.

Fig. 6



Now, consider the variable $(j = 1, \ldots, p)$

$$\hat{\mathbf{x}}_j = \mathbf{X}\mathbf{p}_j = \begin{bmatrix} \hat{x}_{1j} \\ \vdots \\ \hat{x}_{nj} \end{bmatrix}$$

and notice that its (weighted) arithmetic mean and variance are, respectively,

$$\overline{\widehat{x}}_j = [\,m_1 \cdots m_n\,] \begin{bmatrix} \widehat{x}_{1j} \\ \vdots \\ \widehat{x}_{nj} \end{bmatrix} = \mathbf{u}'\mathbf{M}\widehat{\mathbf{x}}_j = \mathbf{u}'\mathbf{M}\mathbf{X}\mathbf{p}_j = \mathbf{g}'\mathbf{p}_j$$

and

$$\begin{aligned}
\widehat{\sigma}_j^2 &= (\mathbf{X}\mathbf{p}_j - \mathbf{u}\,\mathbf{g}'\mathbf{p}_j)'\mathbf{M}(\mathbf{X}\mathbf{p}_j - \mathbf{u}\,\mathbf{g}'\mathbf{p}_j) \\
&= \mathbf{p}_j'\,(\mathbf{X} - \mathbf{u}\,\mathbf{g}')'\mathbf{M}(\mathbf{X} - \mathbf{u}\,\mathbf{g}')\,\mathbf{p}_j \\
&= \mathbf{p}_j'\,([\,\mathbf{x}_1 \cdots \mathbf{x}_p\,] - \mathbf{u}[\,\overline{x}_1 \cdots \overline{x}_p\,])'\,\mathbf{M}\,([\,\mathbf{x}_1 \cdots \mathbf{x}_p\,] - \mathbf{u}[\,\overline{x}_1 \cdots \overline{x}_p\,])\,\mathbf{p}_j \\
&= \mathbf{p}_j'\,[\,\mathbf{x}_1 - \mathbf{u}\,\overline{x}_1 \cdots \mathbf{x}_p - \mathbf{u}\,\overline{x}_p\,]'\,\mathbf{M}\,[\,\mathbf{x}_1 - \mathbf{u}\,\overline{x}_1 \cdots \mathbf{x}_p - \mathbf{u}\,\overline{x}_p\,]\,\mathbf{p}_j \\
&= \mathbf{p}_j'\,[\,\mathbf{y}_1 \cdots \mathbf{y}_p\,]'\,\mathbf{M}\,[\,\mathbf{y}_1 \cdots \mathbf{y}_p\,]\,\mathbf{p}_j \\
&= \mathbf{p}_j'\,\mathbf{V}\,\mathbf{p}_j\,.
\end{aligned}$$

The quantity

$$\widehat{J}_p = \Sigma_j\, q_j\, \widehat{\sigma}_j^2 = \Sigma_j\, q_j\, \mathbf{p}_j'\,\mathbf{V}\,\mathbf{p}_j$$

denotes the *global variability* of $\widehat{\mathbf{x}}_1, \ldots, \widehat{\mathbf{x}}_p$.

We want to show that the inertia $I_{\mathbf{g}+C_k}$ of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ explained by the linear variety $\mathbf{g} + C_k$ is nothing other that the global variability $\widehat{J}_p$.

In fact,

$$\begin{aligned}
I_{\mathbf{g}+C_k} &= \mathrm{tr}\{\mathbf{V}\mathbf{Q}\mathbf{P}\} && = \mathrm{tr}\{\mathbf{V}\mathbf{P}'\mathbf{Q}\mathbf{P}\} \\
&= \mathrm{tr}\{\mathbf{Q}\mathbf{P}\mathbf{V}\mathbf{P}'\} && = \mathrm{tr}\left\{\mathbf{Q}\begin{bmatrix} \mathbf{p}_1' \\ \vdots \\ \mathbf{p}_p' \end{bmatrix}\mathbf{V}[\,\mathbf{p}_1 \cdots \mathbf{p}_p\,]\right\} \\
&= \mathrm{tr}\left\{\mathbf{Q}\begin{bmatrix} \mathbf{p}_1'\mathbf{V}\,\mathbf{p}_1 \cdots \mathbf{p}_1'\mathbf{V}\,\mathbf{p}_p \\ \cdots \quad \cdots \quad \cdots \\ \mathbf{p}_p'\mathbf{V}\,\mathbf{p}_1 \cdots \mathbf{p}_p'\mathbf{V}\,\mathbf{p}_p \end{bmatrix}\right\} && = \mathrm{tr}\left\{\begin{bmatrix} q_1\,\mathbf{p}_1'\mathbf{V}\,\mathbf{p}_1 \cdots q_1\,\mathbf{p}_1'\mathbf{V}\,\mathbf{p}_p \\ \cdots \quad \cdots \quad \cdots \\ q_p\,\mathbf{p}_p'\mathbf{V}\,\mathbf{p}_1 \cdots q_p\,\mathbf{p}_p'\mathbf{V}\,\mathbf{p}_p \end{bmatrix}\right\} \\
&= \Sigma_j\, q_j\, \mathbf{p}_j'\,\mathbf{V}\,\mathbf{p}_j && = \widehat{J}_p\,.
\end{aligned}$$

5 AN APPROACH TO PCA

5.1 PRINCIPAL VECTORS, PRINCIPAL COMPONENTS

Consider again the n individuals $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathrm{R}^p$ with weights given, respectively, by $m_1, \ldots, m_n$ and the linear varieties $\mathbf{c}_* + \mathrm{C}_k$ and $\mathbf{c}_* + \mathrm{C}_k^{\perp}$.

The vectors $\widehat{\boldsymbol{x}}_1 + (\mathbf{c}_* - \widehat{\mathbf{c}}_*), \ldots, \widehat{\boldsymbol{x}}_n + (\mathbf{c}_* - \widehat{\mathbf{c}}_*)$ may be interpreted as the «images» of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ on $\mathbf{c}_* + \mathrm{C}_k$.

If we want such images to be, on the whole, the most representative of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, a criterion may consist in maximizing the inertia explained by $\mathbf{c}_* + \mathrm{C}_k$ with respect to $\mathbf{c}_*$ and C_k [5].

This problem can be solved in two steps: at the first step, taking into account Remark 3, we force the linear variety $\mathbf{c}_* + \mathrm{C}_k$ to pass through the barycentre $\mathbf{g}$; at the second step, we maximize the inertia explained by $\mathbf{g} + \mathrm{C}_k$ with respect to C_k.

As regards this last problem, first notice that we may suppose that the orthogonal subspaces C_k and $\mathrm{C}_k^{\perp}$ of R^p are spanned, respectively, by the orthonormal vectors $\mathbf{c}_1, \ldots, \mathbf{c}_k$ and $\mathbf{c}_{k+1}, \ldots, \mathbf{c}_p$.

Thus – setting

$$\mathbf{C}_k = [\mathbf{c}_1 \cdots \mathbf{c}_k] \quad , \quad \mathbf{C}_{p-k} = [\mathbf{c}_{k+1} \cdots \mathbf{c}_p] \quad , \quad \mathbf{C}_p = [\mathbf{C}_k \;\; \mathbf{C}_{p-k}]$$

– the orthogonal projection matrices on C_k and $\mathrm{C}_k^{\perp}$ become, respectively,

$$\mathbf{P} = \mathbf{C}_k (\mathbf{C}_k' \mathbf{Q} \mathbf{C}_k)^{-1} \mathbf{C}_k' \mathbf{Q} = \mathbf{C}_k \mathbf{C}_k' \mathbf{Q}$$

$$\mathbf{I}_p - \mathbf{P} = \mathbf{C}_{p-k} (\mathbf{C}_{p-k}' \mathbf{Q} \mathbf{C}_{p-k})^{-1} \mathbf{C}_{p-k}' \mathbf{Q} = \mathbf{C}_{p-k} \mathbf{C}_{p-k}' \mathbf{Q} \, .$$

Moreover,

$$\mathrm{I}_{\mathbf{g}+\mathrm{C}_k} = \mathrm{tr}\left\{\mathbf{VQP}\right\} = \mathrm{tr}\left\{\mathbf{VQC}_k \mathbf{C}_k' \mathbf{Q}\right\} = \mathrm{tr}\left\{\mathbf{C}_k' \mathbf{QVQC}_k\right\} \, .$$

Hence, our problem lies in finding out

---

(5) Of course, given that the inertia relative to $\mathbf{c}_*$ is a fixed quantity, this criterion is equivalent to minimizing the inertia not explained by $\mathbf{c}_* + \mathrm{C}_k$.

$$\underset{\mathbf{C}_k}{\text{Max}}\ \text{tr}\left\{\mathbf{C}_k'\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{C}_k\right\}\ ,\ \mathbf{C}_k'\mathbf{Q}\mathbf{C}_k = \mathbf{I}_k\ .$$

In order to solve the problem at hand, consider the Lagrange function

$$\mathrm{L}(\mathbf{C}_k, \mathbf{L}) = \text{tr}\left\{\mathbf{C}_k'\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{C}_k\right\} - \text{tr}\left\{(\mathbf{C}_k'\mathbf{Q}\mathbf{C}_k - \mathbf{I}_k)\mathbf{L}\right\}$$

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

At $(\tilde{\mathbf{C}}_k, \tilde{\mathbf{L}})$ where $\mathrm{L}(\mathbf{C}_k, \mathbf{L})$ has a maximum, it must be

$$\left.\frac{\partial \mathrm{L}(\mathbf{C}_k, \mathbf{L})}{\partial \mathbf{C}_k}\right|_{(\tilde{\mathbf{C}}_k, \tilde{\mathbf{L}})} = 2\mathbf{Q}(\mathbf{V}\mathbf{Q}\tilde{\mathbf{C}}_k - \tilde{\mathbf{C}}_k\tilde{\mathbf{L}}) = \mathbf{O}_{(p,k)}$$

$$\left.\frac{\partial \mathrm{L}(\mathbf{C}_k, \mathbf{L})}{\partial \mathbf{L}}\right|_{(\tilde{\mathbf{C}}_k, \tilde{\mathbf{L}})} = -(\tilde{\mathbf{C}}_k'\mathbf{Q}\tilde{\mathbf{C}}_k - \mathbf{I}_k) \quad = \mathbf{O}_{(k,k)}$$

which gives

$$\mathbf{V}\mathbf{Q}\tilde{\mathbf{C}}_k = \tilde{\mathbf{C}}_k\tilde{\mathbf{L}}\ ,\ \tilde{\mathbf{C}}_k'\mathbf{Q}\tilde{\mathbf{C}}_k = \mathbf{I}_k\ .$$

Therefore, we must look for solutions of the system

$$(*) \qquad \mathbf{V}\mathbf{Q}\mathbf{C}_k = \mathbf{C}_k\mathbf{L}\ ,\ \mathbf{C}_k'\mathbf{Q}\mathbf{C}_k = \mathbf{I}_k$$

in the unknowns $\mathbf{C}_k$ and $\mathbf{L}$.

To this end, consider the equation

$$\mathbf{V}\mathbf{Q}\mathbf{c} = \lambda\mathbf{c}$$

in the unknowns $\mathbf{c}$ and λ.

This equation possesses p orthonormal eigenvectors $\tilde{\mathbf{c}}_1, \ldots, \tilde{\mathbf{c}}_k, \tilde{\mathbf{c}}_{k+1}, \ldots, \tilde{\mathbf{c}}_p$ corresponding to the p (real) eigenvalues $\tilde{\lambda}_1 \geq \ldots \geq \tilde{\lambda}_k \geq \tilde{\lambda}_{k+1} \geq \ldots \geq \tilde{\lambda}_p$.

Moreover, since $(j = 1, \ldots, p)$

$$\mathbf{V}\mathbf{Q}\tilde{\mathbf{c}}_j = \tilde{\lambda}_j\tilde{\mathbf{c}}_j\ ,$$

premultiplying both members by $\widetilde{\mathbf{c}}_j'\mathbf{Q}$, we get

$$\widetilde{\mathbf{c}}_j'\mathbf{QVQ}\widetilde{\mathbf{c}}_j = \widetilde{\lambda}_j\widetilde{\mathbf{c}}_j'\mathbf{Q}\widetilde{\mathbf{c}}_j = \widetilde{\lambda}_j .$$

On the other hand, as $\mathbf{V}$ is positive definite or positive semi-definite and $\mathbf{Q} = \mathbf{Q}'$, $\mathbf{QVQ}$ is also positive definite or positive semi-definite – with $r(\mathbf{QVQ}) = r(\mathbf{V}) = r(\mathbf{Y})$ – and hence $\widetilde{\lambda}_j \geq 0$ $(j = 1, \ldots, p)$.

Thus – setting

$$\widetilde{\mathbf{C}}_k = [\widetilde{\mathbf{c}}_1 \cdots \widetilde{\mathbf{c}}_k] \quad , \quad \widetilde{\mathbf{C}}_{p-k} = [\widetilde{\mathbf{c}}_{k+1} \cdots \widetilde{\mathbf{c}}_p] \quad , \quad \widetilde{\mathbf{C}}_p = [\widetilde{\mathbf{c}}_1 \cdots \widetilde{\mathbf{c}}_p]$$
$$\widetilde{\mathbf{D}}_k = \mathrm{diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_k) \quad , \quad \widetilde{\mathbf{D}}_{p-k} = \mathrm{diag}(\widetilde{\lambda}_{k+1}, \ldots, \widetilde{\lambda}_p) \quad , \quad \widetilde{\mathbf{D}}_p = \mathrm{diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_p)$$

– solutions of the system (*) are provided by $\widetilde{\mathbf{C}}_k = [\widetilde{\mathbf{c}}_1 \cdots \widetilde{\mathbf{c}}_k]$ and $\widetilde{\mathbf{L}} = \widetilde{\mathbf{D}}_k$.

Summing up, first we have

$$\begin{aligned}
\mathbf{VQ}\widetilde{\mathbf{C}}_k &= \widetilde{\mathbf{C}}_k\widetilde{\mathbf{D}}_k &,\quad \widetilde{\mathbf{C}}_k'\mathbf{Q}\widetilde{\mathbf{C}}_k &= \mathbf{I}_k ,\\
\mathbf{VQ}\widetilde{\mathbf{C}}_{p-k} &= \widetilde{\mathbf{C}}_{p-k}\widetilde{\mathbf{D}}_{p-k} &,\quad \widetilde{\mathbf{C}}_{p-k}'\mathbf{Q}\widetilde{\mathbf{C}}_{p-k} &= \mathbf{I}_{p-k} ,\\
\mathbf{VQ}\widetilde{\mathbf{C}}_p &= \widetilde{\mathbf{C}}_p\widetilde{\mathbf{D}}_p &,\quad \widetilde{\mathbf{C}}_p'\mathbf{Q}\widetilde{\mathbf{C}}_p &= \mathbf{I}_p
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{\mathbf{C}}_k'\mathbf{QVQ}\widetilde{\mathbf{C}}_k &= \widetilde{\mathbf{D}}_k &,\quad \widetilde{\mathbf{C}}_k'\mathbf{Q}\widetilde{\mathbf{C}}_k &= \mathbf{I}_k ,\\
\widetilde{\mathbf{C}}_{p-k}'\mathbf{QVQ}\widetilde{\mathbf{C}}_{p-k} &= \widetilde{\mathbf{D}}_{p-k} &,\quad \widetilde{\mathbf{C}}_{p-k}'\mathbf{Q}\widetilde{\mathbf{C}}_{p-k} &= \mathbf{I}_{p-k} ,\\
\widetilde{\mathbf{C}}_p'\mathbf{QVQ}\widetilde{\mathbf{C}}_p &= \widetilde{\mathbf{D}}_p &,\quad \widetilde{\mathbf{C}}_p'\mathbf{Q}\widetilde{\mathbf{C}}_p &= \mathbf{I}_p .
\end{aligned}$$

Then, we have ($\widetilde{\mathbf{P}} = \widetilde{\mathbf{C}}_k\widetilde{\mathbf{C}}_k'\mathbf{Q}$)

$$\begin{aligned}
I_{\mathbf{g}+\widetilde{C}_k} &= \mathrm{tr}\left\{\mathbf{VQ}\widetilde{\mathbf{P}}\right\} &&= \mathrm{tr}\left\{\widetilde{\mathbf{C}}_k'\mathbf{QVQ}\widetilde{\mathbf{C}}_k\right\}\\
&= \mathrm{tr}\left\{\widetilde{\mathbf{D}}_k\right\} &&= \widetilde{\lambda}_1 + \ldots + \widetilde{\lambda}_k ,\\
I_{\mathbf{g}+\widetilde{C}_k^{\perp}} &= \mathrm{tr}\left\{\mathbf{VQ}(\mathbf{I} - \widetilde{\mathbf{P}})\right\} &&= \mathrm{tr}\left\{\widetilde{\mathbf{C}}_{p-k}'\mathbf{QVQ}\widetilde{\mathbf{C}}_{p-k}\right\}\\
&= \mathrm{tr}\left\{\widetilde{\mathbf{D}}_{p-k}\right\} &&= \widetilde{\lambda}_{k+1} + \ldots + \widetilde{\lambda}_p ,\\
I_{\mathbf{g}} &= \mathrm{tr}\left\{\mathbf{VQ}\right\} &&= \mathrm{tr}\left\{\widetilde{\mathbf{C}}_p'\mathbf{QVQ}\widetilde{\mathbf{C}}_p\right\}\\
&= \mathrm{tr}\left\{\widetilde{\mathbf{D}}_p\right\} &&= \widetilde{\lambda}_1 + \ldots + \widetilde{\lambda}_p .
\end{aligned}$$

Moreover, the ratio

$$\text{GQRI} = \frac{I_{\mathbf{g}+\widetilde{C}_k}}{I_{\mathbf{g}}} = \frac{\widetilde{\lambda}_1 + ... + \widetilde{\lambda}_k}{\widetilde{\lambda}_1 + ... + \widetilde{\lambda}_p},$$

which denotes that part of the inertia explained by the linear variety $\mathbf{g} + \widetilde{C}_k$, may be used as an index measuring the global quality of representation of $\boldsymbol{x}_1 + ... + \boldsymbol{x}_n$ on $\mathbf{g} + \widetilde{C}_k$.

The eigenvector $\widetilde{\mathbf{c}}_j$ $(j = 1, ..., p)$ is called the jth *principal vector*, while the vectors $\mathbf{Q}\widetilde{\mathbf{c}}_j$ and $\widetilde{\mathbf{y}}_j = \mathbf{Y}\mathbf{Q}\widetilde{\mathbf{c}}_j$ are called, respectively, the jth *principal factor* and the jth *principal component*.

REMARK 5. The solutions $\widetilde{\mathbf{C}}_k$ and $\widetilde{\mathbf{D}}_k$ of the system (*) are not unique.

All other solutions are obtained by the transformations

$$\widetilde{\mathbf{C}}_k \to \widetilde{\mathbf{C}}_k \mathbf{T} \quad , \quad \widetilde{\mathbf{D}}_k \to \mathbf{T}' \widetilde{\mathbf{D}}_k \mathbf{T}$$

where $\mathbf{T}$ is an orthogonal matrix of order (k, k).

In fact, from

$$\mathbf{V}\mathbf{Q}\widetilde{\mathbf{C}}_k = \widetilde{\mathbf{C}}_k \widetilde{\mathbf{D}}_k \quad , \quad \widetilde{\mathbf{C}}_k' \mathbf{Q} \widetilde{\mathbf{C}}_k = \mathbf{I}_k$$

we get $(\mathbf{T}'\mathbf{T} = \mathbf{I}_k = \mathbf{T}\mathbf{T}')$

$$\mathbf{V}\mathbf{Q}(\widetilde{\mathbf{C}}_k \mathbf{T}) = (\widetilde{\mathbf{C}}_k \mathbf{T})(\mathbf{T}' \widetilde{\mathbf{D}}_k \mathbf{T}) \quad , \quad (\mathbf{T}' \widetilde{\mathbf{C}}_k') \mathbf{Q} (\widetilde{\mathbf{C}}_k \mathbf{T}) = \mathbf{T}'\mathbf{T} = \mathbf{I}_k$$

and inversely.

Notice that, in this case, $\mathbf{T}' \widetilde{\mathbf{D}}_k \mathbf{T}$ is symmetric but not more diagonal.

Moreover,

$$\text{tr}(\mathbf{T}' \widetilde{\mathbf{C}}_k' \mathbf{Q}\mathbf{V}\mathbf{Q} \widetilde{\mathbf{C}}_k \mathbf{T}) = \text{tr}(\mathbf{T}' \widetilde{\mathbf{D}}_k \mathbf{T}) = \text{tr}(\mathbf{T}\mathbf{T}' \widetilde{\mathbf{D}}_k) = \text{tr}(\widetilde{\mathbf{D}}_k) \, .$$

The solution $\widetilde{\mathbf{C}}_k$ is chosen because, as will become apparent below (Section 5.2), it allows us to build up uncorrelated principal components.

It can also be shown that $\widetilde{\mathbf{C}}_k$ may be obtained by a step by step procedure that maximizes the inertia explained by a linear variety of increasing dimension (from 1 to k).

5.2 MAIN PROPERTIES OF PRINCIPAL COMPONENTS

1. $\widetilde{\mathbf{y}}_j$ $(j = 1, \ldots, p)$ is a linear combination of the p vectors $\mathbf{y}_1, \ldots, \mathbf{y}_p \in R^n$ the coefficients of which are represented by the elements of the principal factor $\mathbf{Q}\widetilde{\mathbf{c}}_j$.

Moreover,

$$[\, m_1 \cdots m_n \,] \begin{bmatrix} \widetilde{y}_{1j} \\ \vdots \\ \widetilde{y}_{nj} \end{bmatrix} = \mathbf{u}'\mathbf{M}\widetilde{\mathbf{y}}_j = \mathbf{u}'\mathbf{M}\mathbf{Y}\mathbf{Q}\widetilde{\mathbf{c}}_j = 0 \,,$$

namely the (weighted) arthmetic mean of $\widetilde{\mathbf{y}}_j$ is zero.

2. Setting

$$\widetilde{\mathbf{Y}} = [\, \widetilde{\mathbf{y}}_1 \cdots \widetilde{\mathbf{y}}_p \,] = \mathbf{Y}\mathbf{Q}\widetilde{\mathbf{C}}_p \,,$$

it follows

$$\widetilde{\mathbf{Y}}'\mathbf{M}\widetilde{\mathbf{Y}} = \widetilde{\mathbf{C}}_p' \mathbf{Q}\mathbf{V}\mathbf{Q}\,\widetilde{\mathbf{C}}_p = \widetilde{\mathbf{D}}_p \,.$$

This last expression indicates the covariance matrix of principal components. It shows that principal components are uncorrelated (orthogonal) with variances given, respectively, by $\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_p$.

Notice that

$$r(\widetilde{\mathbf{Y}}) = r(\widetilde{\mathbf{D}}_p) = r(\mathbf{Y})$$

and

$$tr\,(\widetilde{\mathbf{Y}}'\mathbf{M}\widetilde{\mathbf{Y}}) = tr\,(\widetilde{\mathbf{D}}_p) = tr\,(\mathbf{V}\mathbf{Q}) \,.$$

3. Considering the relation $(h = 1, \ldots, r;\ r = r(\widetilde{\mathbf{D}}_p) = r(\mathbf{Y}))$

$$\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{Q}\widetilde{\mathbf{c}}_h = \widetilde{\lambda}_h \widetilde{\mathbf{c}}_h$$

and premultiplying both members by $\mathbf{YQ}$, we obtain

$$\mathbf{Y}\mathbf{Q}\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{Q}\widetilde{\mathbf{c}}_h = \widetilde{\lambda}_h \mathbf{Y}\mathbf{Q}\widetilde{\mathbf{c}}_h$$

or

$$\mathbf{YQY'M}\,\widetilde{\mathbf{y}}_h = \widetilde{\lambda}_h\,\widetilde{\mathbf{y}}_h\,.$$

This relation shows that $\widetilde{\mathbf{y}}_h$ represents an eigenvector of the matrix $\mathbf{YQY'M}$, obtained for the eigenvalue $\widetilde{\lambda}_h > 0$.

Of course, the eigenvalues different from zero of the matrices $\mathbf{YQY'M}$ and $\mathbf{VQ}$ are the same with the same multiplicities.

4. Setting $(h = 1, \dots, r)$

$$\widetilde{\widetilde{\mathbf{y}}}_h = \frac{1}{\sqrt{\widetilde{\lambda}_h}}\,\widetilde{\mathbf{y}}_h = \frac{1}{\sqrt{\widetilde{\lambda}_h}}\,\mathbf{YQ}\,\widetilde{\mathbf{c}}_h$$

where $\widetilde{\widetilde{\mathbf{y}}}_h$ denotes the hth standardized principal component, premultiplying both members by $\mathbf{Y'M}$, we obtain

$$\widetilde{\mathbf{c}}_h = \frac{1}{\sqrt{\widetilde{\lambda}_h}}\,\mathbf{Y'M}\,\widetilde{\widetilde{\mathbf{y}}}_h\,.$$

These two relations, called *transition formulas*, allow us to pass from $\widetilde{\mathbf{c}}_h$ to $\widetilde{\widetilde{\mathbf{y}}}_h$ and vice versa.

5. Since

$$\mathbf{YQ}\,\widetilde{\mathbf{C}}_p = \widetilde{\mathbf{Y}}\,,$$

postmultiplying both members by $\widetilde{\mathbf{C}}'_p$ we obtain

$$\mathbf{YQ}\,\widetilde{\mathbf{C}}_p\,\widetilde{\mathbf{C}}'_p = \widetilde{\mathbf{Y}}\,\widetilde{\mathbf{C}}'_p$$

from which it follows ($\widetilde{\mathbf{C}}_p\,\widetilde{\mathbf{C}}'_p = \mathbf{Q}^{-1}$)

$$\mathbf{Y} = \widetilde{\mathbf{Y}}\,\widetilde{\mathbf{C}}'_p = \left[\,\widetilde{\mathbf{y}}_1 \cdots \widetilde{\mathbf{y}}_p\,\right]\begin{bmatrix}\widetilde{\mathbf{c}}'_1\\ \vdots\\ \widetilde{\mathbf{c}}'_p\end{bmatrix} = \Sigma_j\,\widetilde{\mathbf{y}}_j\,\widetilde{\mathbf{c}}'_j = \Sigma_h\,\sqrt{\widetilde{\lambda}_h}\,\widetilde{\widetilde{\mathbf{y}}}_h\,\widetilde{\mathbf{c}}'_h\,,$$

the so-called *reconstitution formula* or *singular value decomposition* of the matrix $\mathbf{Y}$.

Of course, if the summation is limited to the first $h^* < r$ terms, we obtain

an approximated reconstitution of $\mathbf{Y}$, namely ($h = 1, \ldots, h^*$)

$$\mathbf{Y} \cong \sum_h^{h^*} \tilde{\mathbf{y}}_h \tilde{\mathbf{c}}_h' = \sum_h^{h^*} \sqrt{\tilde{\lambda}_h}\, \tilde{\tilde{\mathbf{y}}}_h \tilde{\mathbf{c}}_h' .$$

6. The cosine of the angle formed by the vectors $\mathbf{y}_j$ ($j = 1, \ldots, p$) and $\tilde{\mathbf{y}}_h$ ($h = 1, \ldots, r$) – the linear correlation coefficient r_{jh} – is given by

$$\cos(\mathbf{y}_j, \tilde{\mathbf{y}}_h) = \frac{\mathbf{y}_j' \mathbf{M} \tilde{\mathbf{y}}_h}{\sigma_j \sqrt{\tilde{\lambda}_h}} = \frac{\boldsymbol{u}_j' \mathbf{Y}' \mathbf{M} \mathbf{Y} \mathbf{Q} \tilde{\mathbf{c}}_h}{\sigma_j \sqrt{\tilde{\lambda}_h}} = \frac{\boldsymbol{u}_j' \mathbf{V} \mathbf{Q} \tilde{\mathbf{c}}_h}{\sigma_j \sqrt{\tilde{\lambda}_h}}$$
$$= \frac{\boldsymbol{u}_j' \tilde{\lambda}_h \tilde{\mathbf{c}}_h}{\sigma_j \sqrt{\tilde{\lambda}_h}} = \frac{\boldsymbol{u}_j' \sqrt{\tilde{\lambda}_h}\, \tilde{\mathbf{c}}_h}{\sigma_j} \qquad = \frac{\sqrt{\tilde{\lambda}_h}\, \tilde{\mathbf{c}}_h' \boldsymbol{u}_j}{\sigma_j} = r_{jh} .$$

7. The orthogonal projection of $\mathbf{y}_j$ ($j = 1, \ldots, p$) on the subspace spanned by the principal component $\tilde{\mathbf{y}}_h$ ($h = 1, \ldots, r$) is given by

$$\tilde{\mathbf{y}}_h (\tilde{\mathbf{y}}_h' \mathbf{M} \tilde{\mathbf{y}}_h)^{-1} \tilde{\mathbf{y}}_h' \mathbf{M} \mathbf{y}_j = \tilde{\mathbf{y}}_h \frac{1}{\tilde{\lambda}_h} \tilde{\mathbf{c}}_h' \mathbf{Q} \mathbf{Y}' \mathbf{M} \mathbf{Y} \boldsymbol{u}_j = \tilde{\mathbf{y}}_h \frac{1}{\tilde{\lambda}_h} \tilde{\mathbf{c}}_h' \mathbf{Q} \mathbf{V} \boldsymbol{u}_j$$
$$= \tilde{\mathbf{y}}_h \frac{1}{\tilde{\lambda}_h} \tilde{\lambda}_h \tilde{\mathbf{c}}_h' \boldsymbol{u}_j \qquad = \tilde{\mathbf{y}}_h \tilde{\mathbf{c}}_h' \boldsymbol{u}_j \qquad\qquad = \tilde{\mathbf{y}}_h \frac{\sigma_j}{\sqrt{\tilde{\lambda}_h}} r_{jh} .$$

8. The orthogonal projection of $\boldsymbol{y}_i$ ($i = 1, \ldots, n$) on the subspace spanned by the principal vector $\tilde{\mathbf{c}}_j$ ($j = 1, \ldots, p$) is given by

$$\tilde{\mathbf{c}}_j (\tilde{\mathbf{c}}_j' \mathbf{Q} \tilde{\mathbf{c}}_j)^{-1} \tilde{\mathbf{c}}_j' \mathbf{Q} \boldsymbol{y}_i = \tilde{\mathbf{c}}_j \tilde{\mathbf{c}}_j' \mathbf{Q} \mathbf{Y}' \mathbf{u}_i$$
$$= \tilde{\mathbf{c}}_j \tilde{\mathbf{y}}_j' \mathbf{u}_i \qquad\qquad = \tilde{\mathbf{c}}_j \tilde{y}_{ij} .$$

5.3 CHOICE OF THE EUCLIDEAN METRIC IN THE INDIVIDUAL SPACE

The choice of the Euclidean metric in the individual space, the matrix $\mathbf{Q}$, is probably one of the most delicate problem in PCA.

As we have said above (Section 2.2.1), this choice generally depends on the measurement units and/or the variances of the variables $\mathbf{x}_1, \ldots, \mathbf{x}_p$.

First, suppose that the variables $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are expressed in the same measurement unit and present approximatively the same variance.

In this case, the metric is usually chosen as $\mathbf{Q} = \mathbf{I}_p$, which is equivalent

to perform a PCA on the basis of the covariance matrix $\mathbf{V}$.

Notice that, if we represent a change in the measurement unit by a constant $s > 0$, we get

$$\begin{aligned} [\mathbf{x}_1 \cdots \mathbf{x}_p] = \mathbf{X} &\rightarrow [s\mathbf{x}_1 \cdots s\mathbf{x}_p] = \mathbf{X}s \\ [\mathbf{y}_1 \cdots \mathbf{y}_p] = \mathbf{Y} &\rightarrow [s\mathbf{y}_1 \cdots s\mathbf{y}_p] = \mathbf{Y}s \\ \mathbf{V}\tilde{\mathbf{c}} = \tilde{\lambda}\tilde{\mathbf{c}} &\rightarrow (s^2\mathbf{V})\tilde{\mathbf{c}} = (s^2\tilde{\lambda})\tilde{\mathbf{c}} \end{aligned}$$

and, thus,

$$\begin{aligned} [\tilde{\mathbf{y}}_1 \cdots \tilde{\mathbf{y}}_p] = \tilde{\mathbf{Y}} &\rightarrow [s\tilde{\mathbf{y}}_1 \cdots s\tilde{\mathbf{y}}_p] = \tilde{\mathbf{Y}}s \\ \tilde{\mathbf{D}}_p &\rightarrow \tilde{\mathbf{D}}_p s^2 . \end{aligned}$$

In other words, each new principal component is s times the corresponding old principal component and has a variance s^2 times the corresponding old variance.

Second, suppose again that the variables $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are expressed in the same measurement unit but present considerably different variances.

In this case, besides the effects of a change in the measurement unit mentioned above, if we perform a PCA on the basis of the covariance matrix $\mathbf{V}$, those variables whose variances are largest tend to dominate the first few principal components.

To illustrate the point in the simplest way, suppose that we have two variables whose covariance matrix is

$$\mathbf{V} = \begin{bmatrix} 9 & 0.5 \\ 0.5 & 1 \end{bmatrix} .$$

Simple calculations show that $\tilde{\mathbf{y}}_1 = \mathbf{y}_1 0.9981 + \mathbf{y}_2 0.0621$, namely that the first principal component is almost identified with the first variable, the variable with the largest variance.

A solution may be found in choosing the metric

$$\mathbf{Q} = \mathbf{Q}_{1/\sigma^2} = \operatorname{diag}\left(\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_p^2}\right)$$

which is equivalent to standardizing the variables $\mathbf{x}_1, \ldots, \mathbf{x}_p$ and to perfor-

ming the analysis on the correlation matrix $\mathbf{R}$.

Yet, results of PCA based on $\mathbf{R}$ are generally different from the corresponding results based on $\mathbf{V}$.

Third, suppose that the variables $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are expressed in different measurement units.

In this case, it does not make sense to perform a PCA on the basis of the covariance matrix $\mathbf{V}$, because operations involving the trace of that matrix have no meaning.

Moreover, setting

$$\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_p)$$

where the constants $s_1 > 0, \ldots, s_p > 0$ represent a change in the measurement units, we get

$$\begin{aligned}
&[\mathbf{x}_1 \cdots \mathbf{x}_p] = \mathbf{X} \;\rightarrow\; [s_1\mathbf{x}_1 \cdots s_p\mathbf{x}_p] = \mathbf{XS} \\
&[\mathbf{y}_1 \cdots \mathbf{y}_p] = \mathbf{Y} \;\rightarrow\; [s_1\mathbf{y}_1 \cdots s_p\mathbf{y}_p] = \mathbf{YS} \\
&\mathbf{V}\tilde{\mathbf{c}} = \tilde{\lambda}\tilde{\mathbf{c}} \;\rightarrow\; (\mathbf{SVS})\tilde{\mathbf{c}}^* = \tilde{\lambda}^*\tilde{\mathbf{c}}^*
\end{aligned}$$

and, generally, $\tilde{\lambda}^* \neq \tilde{\lambda}$ and $\tilde{\mathbf{c}}^* \neq \tilde{\mathbf{c}}$.

Again, a solution may be found in standardizing the variables $\mathbf{x}_1, \ldots, \mathbf{x}_p$ and in performing the analysis on the correlation matrix $\mathbf{R}$.

6 GRAPHICAL REPRESENTATION OF INDIVIDUALS AND VARIABLES

6.1 GRAPHICAL REPRESENTATION OF INDIVIDUALS

Assuming that $r = r(\mathbf{Y}) \geq 2$, a graphical representation of the n individuals $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ (measured in terms of deviations from the means) is usually obtained by their orthogonal projections on the subspace $\tilde{C}_2$ spanned by the first two principal vectors $\tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2$ (*principal plane*).

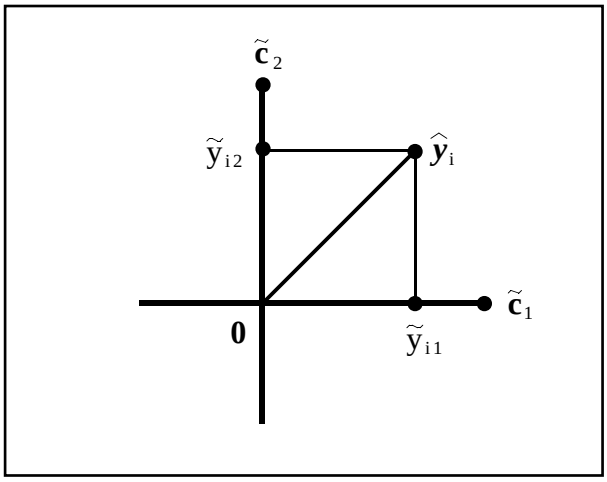
Taking into account what was mentioned above (Section 5.2.*8*) and denoting by $\hat{\boldsymbol{y}}_i$ the orthogonal projection of $\boldsymbol{y}_i$ $(i = 1, \ldots, n)$ on the principal plane, we have

$$\hat{\boldsymbol{y}}_i = \tilde{\mathbf{c}}_1 \, \tilde{y}_{i1} + \tilde{\mathbf{c}}_2 \, \tilde{y}_{i2}$$

where $\tilde{y}_{ij}$ $(j = 1, 2)$ denotes the ith element of the principal component $\tilde{\mathbf{y}}_j$.

Thus, the co-ordinates of $\hat{\boldsymbol{y}}_i$ relative to $\tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2$ are $(\tilde{y}_{i1}, \tilde{y}_{i2})$ (Fig. 7).

Fig. 7



A measure of the global quality of representation of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ on the principal plane is given by the index

$$\text{GQRI} = \frac{\tilde{\lambda}_1 + \tilde{\lambda}_2}{\tilde{\lambda}_1 + \ldots + \tilde{\lambda}_p}$$

which may be interpreted as that part of inertia of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ explained by the subspace $\widetilde{C}_2$ (Remark 4).

REMARK 6. Generally, the representation of the individuals on the principal plane is judged to be adequate if GQRI equals or exceeds a predetermined threshold (for example, 0.7). ∎

However, as the representation of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ may be good even if some individual $\boldsymbol{y}_i$ is far from its orthogonal projection $\widehat{\boldsymbol{y}}_i$, it is necessary to consider the quality of representation of each $\boldsymbol{y}_i$ $(i = 1, \ldots, n)$

An index which may serve this purpose is given by the square cosine of the angle formed by $\boldsymbol{y}_i$ and $\widehat{\boldsymbol{y}}_i$, that is to say by

$$QR(i; \widetilde{\mathbf{c}}_1, \widetilde{\mathbf{c}}_2) = \frac{(\boldsymbol{y}_i' \mathbf{Q} \widehat{\boldsymbol{y}}_i)^2}{(\boldsymbol{y}_i' \mathbf{Q} \boldsymbol{y}_i)(\widehat{\boldsymbol{y}}_i' \mathbf{Q} \widehat{\boldsymbol{y}}_i)}.$$

A high $QR(i; \widetilde{\mathbf{c}}_1, \widetilde{\mathbf{c}}_2)$ – for example, $QR(i; \widetilde{\mathbf{c}}_1, \widetilde{\mathbf{c}}_2) \geq 0.7$ – means that $\boldsymbol{y}_i$ is well represented by $\widehat{\boldsymbol{y}}_i$; on the contrary, a low $QR(i; \widetilde{\mathbf{c}}_1, \widetilde{\mathbf{c}}_2)$ means that the representation of $\boldsymbol{y}_i$ by $\widehat{\boldsymbol{y}}_i$ is poor.

Notice that an explicit expression of $QR(i; \widetilde{\mathbf{c}}_1, \widetilde{\mathbf{c}}_2)$ may be obtained taking into account that we have the following identities

$$\boldsymbol{y}_i' \mathbf{Q} \widehat{\boldsymbol{y}}_i = \boldsymbol{y}_i' \mathbf{Q} \mathbf{P}_{\widetilde{C}_2} \boldsymbol{y}_i = \boldsymbol{y}_i' \mathbf{Q} \mathbf{P}_{\widetilde{C}_2} \mathbf{P}_{\widetilde{C}_2} \boldsymbol{y}_i = \boldsymbol{y}_i' \mathbf{P}_{\widetilde{C}_2}' \mathbf{Q} \mathbf{P}_{\widetilde{C}_2} \boldsymbol{y}_i = \widehat{\boldsymbol{y}}_i' \mathbf{Q} \widehat{\boldsymbol{y}}_i$$

where $\mathbf{P}_{\widetilde{C}_2}$ $(\mathbf{Q}\mathbf{P}_{\widetilde{C}_2} = \mathbf{P}_{\widetilde{C}_2}' \mathbf{Q})$ denotes the orthogonal projection matrix on the subspace $\widetilde{C}_2$, and

$$\widehat{\boldsymbol{y}}_i' \mathbf{Q} \widehat{\boldsymbol{y}}_i = (\widetilde{\mathbf{c}}_1 \widetilde{y}_{i1} + \widetilde{\mathbf{c}}_2 \widetilde{y}_{i2})' \mathbf{Q} (\widetilde{\mathbf{c}}_1 \widetilde{y}_{i1} + \widetilde{\mathbf{c}}_2 \widetilde{y}_{i2}) = \widetilde{y}_{i1}^2 + \widetilde{y}_{i2}^2,$$

$$\begin{aligned} \boldsymbol{y}_i' \mathbf{Q} \boldsymbol{y}_i &= \mathbf{u}_i' \mathbf{Y} \mathbf{Q} \mathbf{Y}' \mathbf{u}_i = \mathbf{u}_i' \widetilde{\mathbf{Y}} \widetilde{\mathbf{C}}_p' \mathbf{Q} \widetilde{\mathbf{C}}_p \widetilde{\mathbf{Y}}' \mathbf{u}_i = \mathbf{u}_i' \widetilde{\mathbf{Y}} \widetilde{\mathbf{Y}}' \mathbf{u}_i \\ &= \begin{bmatrix} \widetilde{y}_{i1} \cdots \widetilde{y}_{ip} \end{bmatrix} \begin{bmatrix} \widetilde{y}_{i1} \\ \vdots \\ \widetilde{y}_{ip} \end{bmatrix} = \widetilde{y}_{i1}^2 + \ldots + \widetilde{y}_{ip}^2. \end{aligned}$$

Thus,

$$QR(i;\tilde{\mathbf{c}}_1,\tilde{\mathbf{c}}_2)=\frac{(\boldsymbol{y}_i'\mathbf{Q}\hat{\boldsymbol{y}}_i)^2}{(\boldsymbol{y}_i'\mathbf{Q}\boldsymbol{y}_i)(\hat{\boldsymbol{y}}_i'\mathbf{Q}\hat{\boldsymbol{y}}_i)}=\frac{\hat{\boldsymbol{y}}_i'\mathbf{Q}\hat{\boldsymbol{y}}_i}{\boldsymbol{y}_i'\mathbf{Q}\boldsymbol{y}_i}=\frac{\tilde{y}_{i1}^2+\tilde{y}_{i2}^2}{\tilde{y}_{i1}^2+\ldots+\tilde{y}_{ip}^2}.$$

Moreover, since we can write

$$QR(i;\tilde{\mathbf{c}}_1,\tilde{\mathbf{c}}_2)=\frac{\tilde{y}_{i1}^2}{\boldsymbol{y}_i'\mathbf{Q}\boldsymbol{y}_i}+\frac{\tilde{y}_{i2}^2}{\boldsymbol{y}_i'\mathbf{Q}\boldsymbol{y}_i}=QR(i;\tilde{\mathbf{c}}_1)+QR(i;\tilde{\mathbf{c}}_2)$$

where $QR(i;\tilde{\mathbf{c}}_j)$ $(j=1,2)$ denotes the square cosine of the angle formed by $\boldsymbol{y}_i$ and its orthogonal projection on the subspace spanned by $\tilde{\mathbf{c}}_j$ and is a measure of the quality of representation of $\boldsymbol{y}_i$ on that subspace, we are able to attribute to each axis the due part of $QR(i;\tilde{\mathbf{c}}_1,\tilde{\mathbf{c}}_2)$.

After having examined the quality of representation of each individual by means of the index $QR(i;\tilde{\mathbf{c}}_1,\tilde{\mathbf{c}}_2)$, we are in a position to correctly judge proximities among their orthogonal projections on the principal plane: if two individuals $\boldsymbol{y}_i,\boldsymbol{y}_{i^*}$ are close means that $\hat{\boldsymbol{y}}_i,\hat{\boldsymbol{y}}_{i^*}$ are close too, provided they are well represented.

In interpreting results of the analysis, it is also important to examine the contribution of each individual $\boldsymbol{y}_i$ to the inertia $\tilde{\lambda}_j$ explained by $\tilde{\mathbf{c}}_j$.

Since

$$\tilde{\lambda}_j=\tilde{\mathbf{c}}_j'\mathbf{QY'MYQ}\tilde{\mathbf{c}}_j=\tilde{\mathbf{y}}_j'\mathbf{M}\tilde{\mathbf{y}}_j=\Sigma_i\, m_i\tilde{y}_{ij}^2,$$

an index often considered is

$$C(i;\tilde{\mathbf{c}}_j)=\frac{m_i\tilde{y}_{ij}^2}{\tilde{\lambda}_j}.$$

The usefulness of examining these contributions may be pointed out first noting that, on the graph, only the co-ordinates of $\hat{\boldsymbol{y}}_1,\ldots,\hat{\boldsymbol{y}}_n$ relative to $\tilde{\mathbf{c}}_j$ are represented; thus, our attention falls on points with a high $|\tilde{y}_{i1}|$, even if some of these may have a small weight.

On the contrary, taking up the examination of contributions – as $C(i;\tilde{\mathbf{c}}_j)$ depends both on m_i and $\tilde{y}_{i1}^2$ – allows us to detect those individuals which have contributed most to the inertia explained by the subspace under con-

sideration, namely the individuals characterizing that subspace.

Of course, if the individuals all have the same importance – namely, if $\mathbf{M} = \text{diag}(\frac{1}{n}, \ldots, \frac{1}{n})$ – examination of the co-ordinates relative to $\tilde{\mathbf{c}}_j$ suffices.

Moreover, it may happen that the contribution of an individual relative to the others is very high; in that case, it is advisable to perform the analysis again after its exclusion from the data set and to reintroduce it as a «supplementary» individual (Section 6.3). This allows us to appreciate differences among remaining individuals, differences which might otherwise be difficult to visualize on the graph since the point scatter is strongly conditioned by the presence of an atypical individual.

6.2 GRAPHICAL REPRESENTATION OF VARIABLES

Assuming that $r = r(\mathbf{Y}) \geq 2$, 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 $S(\tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2)$ spanned by the first two standardized principal components $\tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2$.

Taking into account what we said above (Section 5.2.*7*) and denoting by $\hat{\mathbf{y}}_j$ the orthogonal projection of $\mathbf{y}_j$ $(j = 1, \ldots, p)$ on $S(\tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2)$, we have

$$\hat{\mathbf{y}}_j = \tilde{\tilde{\mathbf{y}}}_1 \, \sigma_j \, r_{j1} + \tilde{\tilde{\mathbf{y}}}_2 \, \sigma_j \, r_{j2}$$

where r_{j1} and r_{j2} denote, respectively, the linear correlation coefficients of $\mathbf{y}_j$ with $\tilde{\tilde{\mathbf{y}}}_1$ and $\tilde{\tilde{\mathbf{y}}}_2$.

Thus, the co-ordinates of $\hat{\mathbf{y}}_j$ relative to $\tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2$ are $(\sigma_j r_{j1}, \sigma_j r_{j2})$.

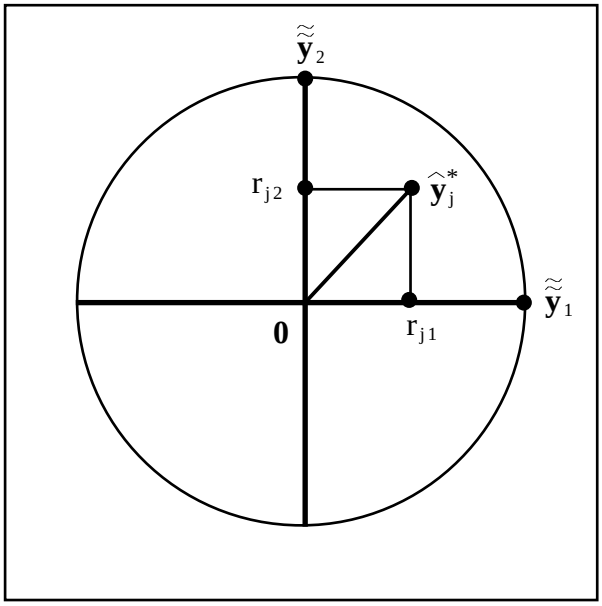
However, since we are mainly interested in representing linear correlations between pairs of variables or between a variable and a principal component and linear correlations are invariant if each variable is scaled by its standard deviation, it is more suitable to work with standardized variables.

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 $S(\tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2)$ is given by

$$\hat{\mathbf{y}}_j^* = \tilde{\tilde{\mathbf{y}}}_1 \, r_{j1} + \tilde{\tilde{\mathbf{y}}}_2 \, r_{j2}$$

so that the co-ordinates of $\widehat{\mathbf{y}}_j^*$ relative to $\tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2$ are (r_{j1}, r_{j2}) (Fig. 8) and hence it is very easy to distinguish those variables which are the most correlated with a principal component and which play a significant role in its interpretation.

Fig. 8



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

Moreover, the quality of representation of each variable on $S(\tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2)$ can be judged by means of the square cosine of the angle formed by $\mathbf{y}_j^*$ and $\widehat{\mathbf{y}}_j^*$ which is given by $((\mathbf{y}_j^*)'\mathbf{M}(\mathbf{y}_j^*) = 1)$

$$QR(j; \tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2) = \frac{[(\mathbf{y}_j^*)'\mathbf{M}(\widehat{\mathbf{y}}_j^*)]^2}{[(\mathbf{y}_j^*)'\mathbf{M}(\mathbf{y}_j^*)]\,[(\widehat{\mathbf{y}}_j^*)'\mathbf{M}(\widehat{\mathbf{y}}_j^*)]} = \frac{[(\mathbf{y}_j^*)'\mathbf{M}(\widehat{\mathbf{y}}_j^*)]^2}{(\widehat{\mathbf{y}}_j^*)'\mathbf{M}(\widehat{\mathbf{y}}_j^*)}.$$

A high $QR(j; \tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2)$ – for example, $QR(j; \tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2) \geq 0.7$ – means that $\mathbf{y}_j^*$ is well represented by $\widehat{\mathbf{y}}_j^*$; on the contrary, a low $QR(j; \tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2)$ means that the representation of $\mathbf{y}_j^*$ by $\widehat{\mathbf{y}}_j^*$ is poor.

Notice that another expression of $QR(j; \tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2)$ may be obtained taking

into account that

$$(\widehat{\mathbf{y}}_j^*)'\mathbf{M}(\widehat{\mathbf{y}}_j^*) = (\tilde{\tilde{\mathbf{y}}}_1 r_{j1} + \tilde{\tilde{\mathbf{y}}}_2 r_{j2})'\mathbf{M}(\tilde{\tilde{\mathbf{y}}}_1 r_{j1} + \tilde{\tilde{\mathbf{y}}}_2 r_{j2}) = r_{j1}^2 + r_{j2}^2$$

and (Section 5.2.*6*)

$$(\mathbf{y}_j^*)'\mathbf{M}(\widehat{\mathbf{y}}_j^*) = (\mathbf{y}_j^*)'\mathbf{M}(\tilde{\tilde{\mathbf{y}}}_1 r_{j1} + \tilde{\tilde{\mathbf{y}}}_2 r_{j2}) \quad = \frac{\mathbf{y}_j'}{\sigma_j}\mathbf{M}\left(\frac{\tilde{\mathbf{y}}_1}{\sqrt{\tilde{\lambda}_1}} r_{j1} + \frac{\tilde{\mathbf{y}}_2}{\sqrt{\tilde{\lambda}_2}} r_{j2}\right)$$

$$= \frac{\mathbf{y}_j'\mathbf{M}\tilde{\mathbf{y}}_1}{\sigma_j\sqrt{\tilde{\lambda}_1}} r_{j1} + \frac{\mathbf{y}_j'\mathbf{M}\tilde{\mathbf{y}}_2}{\sigma_j\sqrt{\tilde{\lambda}_2}} r_{j2} = r_{j1}^2 + r_{j2}^2 .$$

Thus,

$$QR(j; \tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2) = r_{j1}^2 + r_{j2}^2 .$$

On the other hand, since $QR(j; \tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2)$ also denotes the square distance of $\widehat{\mathbf{y}}_j^*$ from the correlation circle centre, we can see that well-represented points 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 principal component;
- which variables are uncorrelated (orthogonal) among themselves and with each principal component.

6.3 SUPPLEMENTARY INDIVIDUALS AND VARIABLES

In applying PCA, it often happens that additional information is available besides that contained in the data matrix **Y**.

For example, we may have m additional individuals (measured in terms of deviations from the means)

$$\boldsymbol{y}'_{n+1} = \left[\, y_{n+1,1} \cdots y_{n+1,p} \,\right] \quad , \ldots , \quad \boldsymbol{y}'_{n+m} = \left[\, y_{n+m,1} \cdots y_{n+m,p} \,\right]$$

which belong to a control group and which cannot therefore be included in **Y**.

Analogously, we might have q additional variables (measured in terms of deviations from the means)

$$\mathbf{y}'_{p+1} = \left[y_{1,p+1} \cdots y_{n,p+1} \right] \quad , \ldots , \quad \mathbf{y}'_{p+q} = \left[y_{1,p+q} \cdots y_{n,p+q} \right]$$

which are of a different nature with respect to the variables contained in $\mathbf{Y}$ and which we do not wish to incorporate in $\mathbf{Y}$.

After having obtained – on the basis of the matrix $\mathbf{Y}$ – principal vectors and principal components and represented the initial individuals and variables, we would like to place the m additional individuals and the q additional variables on the respective graphs.

The procedure for doing this consists of positioning the supplementary individuals and variables on the graphs.

Of course, as before, the orthogonal projection matrix of a supplementary individual on the subspace spanned by $\tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2$ is

$$\left[\tilde{\mathbf{c}}_1 \quad \tilde{\mathbf{c}}_2 \right] \left[\tilde{\mathbf{c}}_1 \quad \tilde{\mathbf{c}}_2 \right]' \mathbf{Q} ,$$

and the orthogonal projection matrix of a supplementary variable on the subspace spanned by the first two standardized principal components $\tilde{\tilde{\mathbf{y}}}_1, \tilde{\tilde{\mathbf{y}}}_2$ is

$$\left[\tilde{\tilde{\mathbf{y}}}_1 \quad \tilde{\tilde{\mathbf{y}}}_2 \right] \left[\tilde{\tilde{\mathbf{y}}}_1 \quad \tilde{\tilde{\mathbf{y}}}_2 \right]' \mathbf{M} .$$

7 OTHER APPROACHES TO PCA

7.1 THE APPROACH IN TERMS OF WEIGHTED SUM OF SQUARE DISTANCES BETWEEN ANY PAIR OF INDIVIDUALS

Consider the n individuals $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ (measured in terms of deviations from the means) with weights given, respectively, by $m_1, \ldots, m_n$.

Let

$$d^2_{i,i^*} = (\boldsymbol{y}_i - \boldsymbol{y}_{i^*})' \mathbf{Q} (\boldsymbol{y}_i - \boldsymbol{y}_{i^*})$$

the square distance between a pair of individuals $(\boldsymbol{y}_i, \boldsymbol{y}_{i^*})$ $(i, i^* = 1, \ldots, n)$ and

$$\mathbf{D}^2 = \begin{bmatrix} d^2_{1,1} \cdots d^2_{1,n} \\ \cdots \cdots \cdots \\ d^2_{n,1} \cdots d^2_{n,n} \end{bmatrix}$$

the corresponding square distance matrix (6).

Setting

$$\mathbf{d}^2\mathbf{u}' = \begin{bmatrix} \boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_n \end{bmatrix} \mathbf{u}' = \begin{bmatrix} \boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_1 \cdots \boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_1 \\ \cdots \quad \cdots \quad \cdots \\ \boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_n \cdots \boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_n \end{bmatrix},$$

$$\mathbf{u}\,\mathbf{d}^{2\prime} = \mathbf{u}\left[\boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_1 \cdots \boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_n\right] = \begin{bmatrix} \boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_1 \cdots \boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_n \\ \cdots \quad \cdots \quad \cdots \\ \boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_1 \cdots \boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_n \end{bmatrix},$$

$$\mathbf{Y}\mathbf{Q}\mathbf{Y}' = \begin{bmatrix} \boldsymbol{y}_1' \\ \vdots \\ \boldsymbol{y}_n' \end{bmatrix} \mathbf{Q} \left[\boldsymbol{y}_1 \cdots \boldsymbol{y}_n\right] = \begin{bmatrix} \boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_1 \cdots \boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_n \\ \cdots \quad \cdots \quad \cdots \\ \boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_1 \cdots \boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_n \end{bmatrix},$$

it can easily be shown that

(6) Of course, this matrix is symmetric and has the elements on the principal diagonal all equal to zero.

$$\mathbf{D}^2 = \mathbf{d}^2\mathbf{u}' + \mathbf{u}\mathbf{d}^{2\prime} - 2\,\mathbf{Y}\mathbf{Q}\mathbf{Y}' .$$

Now, consider the weighted sum of square distances between any pair of individuals, each square distance weighted by $m_i\, m_{i^*}$.

We want to show that this last quantity, which can be written as $\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{D}^2)$, is nothing other than twice $I_{\mathbf{g}}$.

In fact, first we have $(\mathbf{u}'\mathbf{M}\mathbf{Y} = \mathbf{0}_{(1,p)})$

$$\begin{aligned}\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{D}^2) &= \mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}(\mathbf{d}^2\mathbf{u}' + \mathbf{u}\mathbf{d}^{2\prime} - 2\mathbf{Y}\mathbf{Q}\mathbf{Y}')) \\ &= \mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{d}^2\mathbf{u}') + \mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{u}\mathbf{d}^{2\prime}) - 2\,\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{Y}\mathbf{Q}\mathbf{Y}') \\ &= \mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{d}^2\mathbf{u}') + \mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{u}\mathbf{d}^{2\prime}) .\end{aligned}$$

Then – since

$$\begin{aligned}&\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{d}^2\mathbf{u}') && = \mathrm{tr}(\mathbf{d}^2\mathbf{u}'\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}) \\ &= \mathrm{tr}(\mathbf{d}^2\mathbf{u}'\mathbf{M}) && = \mathrm{tr}\begin{bmatrix} m_1\,\boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_1 & \cdots & m_n\,\boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_1 \\ \cdots & \cdots & \cdots \\ m_1\,\boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_n & \cdots & m_n\,\boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_n \end{bmatrix} \\ &= \mathrm{tr}\begin{bmatrix} m_1\,\boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_1 & \cdots & m_1\,\boldsymbol{y}_1'\mathbf{Q}\boldsymbol{y}_n \\ \cdots & \cdots & \cdots \\ m_n\,\boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_1 & \cdots & m_n\,\boldsymbol{y}_n'\mathbf{Q}\boldsymbol{y}_n \end{bmatrix} && = \mathrm{tr}(\mathbf{Y}\mathbf{Q}\mathbf{Y}'\mathbf{M}) \\ &= \mathrm{tr}(\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{Q}) && = \mathrm{tr}(\mathbf{V}\mathbf{Q}) \\ &= I_{\mathbf{g}}\end{aligned}$$

and, analogously,

$$\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{u}\mathbf{d}^{2\prime}) = I_{\mathbf{g}} ,$$

– we get

$$\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{D}^2) = 2I_{\mathbf{g}} .$$

Next, let

$$\hat{d}^2_{i,i^*} = (\mathbf{P}\boldsymbol{y}_i - \mathbf{P}\boldsymbol{y}_{i^*})'\mathbf{Q}(\mathbf{P}\boldsymbol{y}_i - \mathbf{P}\boldsymbol{y}_{i^*})$$

the square distance between a pair of projected individuals $(\mathbf{P}\boldsymbol{y}_i, \mathbf{P}\boldsymbol{y}_{i^*})$ on C_k and

$$\widehat{\mathbf{D}}^2 = \begin{bmatrix} \widehat{d}^2_{1,1} \cdots \widehat{d}^2_{1,n} \\ \cdots \cdots \cdots \\ \widehat{d}^2_{n,1} \cdots \widehat{d}^2_{n,n} \end{bmatrix}$$

the corresponding square distance matrix.

Setting

$$\widehat{\mathbf{d}}^2\mathbf{u}' = \begin{bmatrix} \boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n \end{bmatrix}\mathbf{u}' = \begin{bmatrix} \boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 \cdots \boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 \\ \cdots \quad \cdots \quad \cdots \\ \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n \cdots \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n \end{bmatrix},$$

$$\mathbf{u}\widehat{\mathbf{d}}^{2\prime} = \mathbf{u}\left[\boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 \cdots \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n\right] = \begin{bmatrix} \boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 \cdots \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n \\ \cdots \quad \cdots \quad \cdots \\ \boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 \cdots \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n \end{bmatrix},$$

$$\mathbf{Y}\mathbf{P}'\mathbf{Q}\mathbf{P}\mathbf{Y}' = \begin{bmatrix} \boldsymbol{y}_1'\mathbf{P}' \\ \vdots \\ \boldsymbol{y}_n'\mathbf{P}' \end{bmatrix}\mathbf{Q}\left[\mathbf{P}\boldsymbol{y}_1 \cdots \mathbf{P}\boldsymbol{y}_n\right] = \begin{bmatrix} \boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 \cdots \boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n \\ \cdots \quad \cdots \quad \cdots \\ \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 \cdots \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n \end{bmatrix},$$

it can easily be shown that

$$\widehat{\mathbf{D}}^2 = \widehat{\mathbf{d}}^2\mathbf{u}' + \mathbf{u}\widehat{\mathbf{d}}^{2\prime} - 2\,\mathbf{Y}\mathbf{P}'\mathbf{Q}\mathbf{P}\mathbf{Y}'.$$

Now, consider the weighted sum of square distances between any pair of projected individuals, each square distance weighted by $m_i m_{i^*}$.

We want to show that this last quantity, which can be written as $\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\widehat{\mathbf{D}}^2)$, is nothing other than twice $I_{\mathbf{g}+C_k}$.

In fact, first we have ($\mathbf{u}'\mathbf{M}\mathbf{Y} = \mathbf{0}_{(1,p)}$)

$$\begin{aligned} \mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\widehat{\mathbf{D}}^2) &= \mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}(\widehat{\mathbf{d}}^2\mathbf{u}' + \mathbf{u}\widehat{\mathbf{d}}^{2\prime} - 2\,\mathbf{Y}\mathbf{P}'\mathbf{Q}\mathbf{P}\mathbf{Y}')) \\ &= \mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\widehat{\mathbf{d}}^2\mathbf{u}') + \mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{u}\widehat{\mathbf{d}}^{2\prime}) - 2\,\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{Y}\mathbf{P}'\mathbf{Q}\mathbf{P}\mathbf{Y}') \\ &= \mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\widehat{\mathbf{d}}^2\mathbf{u}') + \mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{u}\widehat{\mathbf{d}}^{2\prime}). \end{aligned}$$

Then – since

$$\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\widehat{\mathbf{d}}^2\mathbf{u}') = \mathrm{tr}(\widehat{\mathbf{d}}^2\mathbf{u}'\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M})$$

$$= \mathrm{tr}(\widehat{\mathbf{d}}^2\mathbf{u}'\mathbf{M}) \qquad = \mathrm{tr}\begin{bmatrix} m_1 \boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 & \cdots & m_n \boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 \\ \cdots & \cdots & \cdots \\ m_1 \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n & \cdots & m_n \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n \end{bmatrix}$$

$$= \mathrm{tr}\begin{bmatrix} m_1 \boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 & \cdots & m_1 \boldsymbol{y}_1'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n \\ \cdots & \cdots & \cdots \\ m_n \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_1 & \cdots & m_n \boldsymbol{y}_n'\mathbf{P}'\mathbf{Q}\mathbf{P}\boldsymbol{y}_n \end{bmatrix} = \mathrm{tr}(\mathbf{Y}\mathbf{P}'\mathbf{Q}\mathbf{P}\mathbf{Y}'\mathbf{M})$$

$$= \mathrm{tr}(\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{Q}\mathbf{P}) \qquad = \mathrm{tr}(\mathbf{V}\mathbf{Q}\mathbf{P})$$

$$= I_{\mathbf{g}+C_k}$$

and, analogously,

$$\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\mathbf{u}\widehat{\mathbf{d}}^{2\prime}) = I_{\mathbf{g}+C_k},$$

– we get

$$\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}\widehat{\mathbf{D}}^2) = 2I_{\mathbf{g}+C_k}.$$

Summing up, since we have shown that the weighted sum of square distances between any pair of individuals is twice $I_{\mathbf{g}}$ and that the weighted sum of square distances between any pair of projected individuals is twice $I_{\mathbf{g}+C_k}$, finding out the subspace maximizing the inertia explained by $I_{\mathbf{g}+C_k}$ (Section 5.1) is equivalent to looking for the subspace maximizing the weighted sum of square distances between any pair of projected individuals.

In other words, with this interpretation the criterion consists of finding out the subspace modifying as little as possible the weighted sum of square distances between any pair of individuals when passing to C_k.

7.2 THE APPROACH IN TERMS OF GLOBAL VARIABILITY AND GENERALIZED VARIANCE

Consider the problem of finding out (Section 5.1)

$$\underset{\mathbf{C}_k}{\mathrm{Max}}\ \mathrm{tr}\left\{\mathbf{C}_k'\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{C}_k\right\} = \underset{\mathbf{C}_k}{\mathrm{Max}}\ \mathrm{tr}\left\{\mathbf{C}_k'\mathbf{Q}\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{Q}\mathbf{C}_k\right\} \quad , \quad \mathbf{C}_k'\mathbf{Q}\mathbf{C}_k = \mathbf{I}_k .$$

Since $\mathbf{c}_h' \mathbf{Q}\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{Q}\mathbf{c}_h$ is the variance of the hth ($h = 1, \dots, k$) linear combination $\mathbf{Y}\mathbf{Q}\mathbf{c}_h$ of the p variables $\mathbf{y}_1, \dots, \mathbf{y}_p$ (measured in terms of deviations from the means), we realize that the above-mentioned problem is equivalent to maximizing the *global variability* of the k linear combinations $\mathbf{Y}\mathbf{Q}\mathbf{c}_1, \dots, \mathbf{Y}\mathbf{Q}\mathbf{c}_k$ of $\mathbf{y}_1, \dots, \mathbf{y}_p$ under the constraint $\mathbf{C}_k'\mathbf{Q}\mathbf{C}_k = \mathbf{I}_k$.

Of course, $\tilde{\mathbf{C}}_k = [\tilde{\mathbf{c}}_1 \cdots \tilde{\mathbf{c}}_k]$ is a solution of this problem and

$$\operatorname{tr}\left\{\tilde{\mathbf{C}}_k'\mathbf{V}\tilde{\mathbf{C}}_k\right\} = \Sigma_h \tilde{\lambda}_h .$$

Instead of considering $\operatorname{tr}\left\{\mathbf{C}_k'\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{C}_k\right\}$ as a measure of the variability of the k linear combinations $\mathbf{Y}\mathbf{Q}\mathbf{c}_1, \dots, \mathbf{Y}\mathbf{Q}\mathbf{c}_k$ of $\mathbf{y}_1, \dots, \mathbf{y}_p$, we also may refer to $\det\left\{\mathbf{C}_k'\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{C}_k\right\}$, namely to the so-called *generalized variance*.

In this case, the problem becomes

$$\operatorname*{Max}_{\mathbf{C}_k} \det\left\{\mathbf{C}_k'\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{C}_k\right\} \quad , \quad \mathbf{C}_k'\mathbf{Q}\mathbf{C}_k = \mathbf{I}_k .$$

It can be shown (7) that $\tilde{\mathbf{C}}_k = [\tilde{\mathbf{c}}_1 \cdots \tilde{\mathbf{c}}_k]$ is a solution and that

$$\det\left\{\tilde{\mathbf{C}}_k'\mathbf{V}\tilde{\mathbf{C}}_k\right\} = \Pi_h \tilde{\lambda}_h .$$

7.3 THE APPROACH IN TERMS OF SUM OF SQUARE LINEAR CORRELATION COEFFICIENTS BETWEEN A NORMALIZED LINEAR COMBINATION OF THE ORIGINAL VARIABLES AND EACH ORIGINAL VARIABLE

Consider the problem of finding out a normalized variable $\tilde{\tilde{\mathbf{y}}}_{(1)}$, linear combination of $\mathbf{y}_1, \dots, \mathbf{y}_p$, maximizing the sum of the square linear correlation coefficients between $\tilde{\tilde{\mathbf{y}}}_{(1)}$ and each $\mathbf{y}_j$ ($j = 1, \dots, p$).

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

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

(7) A proof is given in Jolliffe on pp.15-16.

$$\Sigma_j \cos^2(\tilde{\tilde{\mathbf{y}}}, \mathbf{y}_j) = \Sigma_j \frac{(\tilde{\tilde{\mathbf{y}}}'\mathbf{M}\mathbf{y}_j)^2}{\sigma_j^2}$$

$$= \Sigma_j \frac{\tilde{\tilde{\mathbf{y}}}'\mathbf{M}\mathbf{y}_j\mathbf{y}_j'\mathbf{M}\tilde{\tilde{\mathbf{y}}}}{\sigma_j^2} = \tilde{\tilde{\mathbf{y}}}'\mathbf{M}\,(\Sigma_j \frac{\mathbf{y}_j\mathbf{y}_j'}{\sigma_j^2})\,\mathbf{M}\tilde{\tilde{\mathbf{y}}}$$

$$= \tilde{\tilde{\mathbf{y}}}'\mathbf{M}\,(\begin{bmatrix}\mathbf{y}_1 \cdots \mathbf{y}_p\end{bmatrix}\mathbf{Q}_{1/\sigma^2}\begin{bmatrix}\mathbf{y}_1' \\ \vdots \\ \mathbf{y}_p'\end{bmatrix})\,\mathbf{M}\tilde{\tilde{\mathbf{y}}} = \tilde{\tilde{\mathbf{y}}}'\mathbf{M}\mathbf{Y}\mathbf{Q}_{1/\sigma^2}\mathbf{Y}'\mathbf{M}\tilde{\tilde{\mathbf{y}}},$$

we must find out

$$\underset{\tilde{\tilde{\mathbf{y}}}}{\text{Max}}\ \tilde{\tilde{\mathbf{y}}}'\mathbf{M}\mathbf{Y}\mathbf{Q}_{1/\sigma^2}\mathbf{Y}'\mathbf{M}\tilde{\tilde{\mathbf{y}}}\ ,\ \tilde{\tilde{\mathbf{y}}}'\mathbf{M}\tilde{\tilde{\mathbf{y}}} = 1\ .$$

As can easily be seen, a solution of this problem is given by the normalized eigenvector $\tilde{\tilde{\mathbf{y}}}_{(1)}$ of the matrix $\mathbf{Y}\mathbf{Q}_{1/\sigma^2}\mathbf{Y}'\mathbf{M}$ associated with the eigenvalue $\hat{\lambda}_1$, and that $\Sigma_j \cos^2(\tilde{\tilde{\mathbf{y}}}, \mathbf{y}_j) = \hat{\lambda}_1$.

Thus, assuming that $\mathbf{Q} = \mathbf{Q}_{1/\sigma^2}$, $\tilde{\tilde{\mathbf{y}}}_{(1)}$ equals $\tilde{\tilde{\mathbf{y}}}_1$, the first standardized principal component.

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

7.4 THE APPROACH IN TERMS OF THE MULTIVARIABLE LINEAR MODEL

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

To this end, first remember that this latter can be expressed in the form

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

where

- $\mathbf{Y}_1$ is the matrix, of order (n, p_1), of the observed values of p_1 dependent variables, measured in terms of deviations from the means;
- $\mathbf{Y}_2$ is the matrix, of order (n, p_2), of the observed values of p_2 independent variables, measured in terms of deviations from the means;

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

In order to determine the matrix $\mathbf{H}_2$, we can choose a least squares criterion, which means finding out

$$\underset{\mathbf{H}_2}{\text{Min}} \ \text{tr}\{(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{H}_2)'\mathbf{M}(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{H})\mathbf{Q}\} .$$

Of course, assuming that $r(\mathbf{Y}_2) = p_2$, the best solution is given by $\widehat{\mathbf{H}}_2 = (\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_2)^{-1}\mathbf{Y}_2'\mathbf{M}\mathbf{Y}_1$.

Now, in case $\mathbf{Y}_1 = \mathbf{Y}_2 = \mathbf{Y}$, consider the model

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

from which is clear that, without any assumption regarding the matrix $\mathbf{H}$, of order (p, p), the best solution is trivially given by $\widetilde{\mathbf{H}} = \mathbf{I}_p$.

Then, assume that $\mathbf{H}$ has rank $h^* < r = r(\mathbf{Y})$, so that it may be written in the form ($\mathbf{F}$ and $\mathbf{G}$ of order, respectively, (p, h^*) and (h^*, p))

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

with $r(\mathbf{F}) = r(\mathbf{G}) = h^*$.

Our model becomes

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

and we propose to find out

$$\underset{\mathbf{F},\mathbf{G}}{\text{Min}} \ \text{tr}\{(\mathbf{Y} - \mathbf{Y}\mathbf{F}\mathbf{G})'\mathbf{M}(\mathbf{Y} - \mathbf{Y}\mathbf{F}\mathbf{G})\mathbf{Q}\} \quad , \quad \mathbf{F}'\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{F} = \mathbf{I}_{h^*} .$$

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

$$\underset{\mathbf{F},\mathbf{G}}{\text{Min}} \ \{\text{tr}\{\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{Q}\} - 2\,\text{tr}\{\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{F}\mathbf{G}\mathbf{Q}\} + \text{tr}\{\mathbf{G}'\mathbf{G}\mathbf{Q}\}\} \quad , \quad \mathbf{F}'\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{F} = \mathbf{I}_{h^*}$$

or, equivalently,

$$\underset{\mathbf{F},\mathbf{G}}{\text{Max}} \ \{2\text{tr}\{\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{F}\mathbf{G}\mathbf{Q}\} - \text{tr}\{\mathbf{G}'\mathbf{G}\mathbf{Q}\}\} \quad , \quad \mathbf{F}'\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{F} = \mathbf{I}_{h^*} .$$

Now, consider the Lagrange function

$$L(\mathbf{F},\mathbf{G},\mathbf{L}) = 2\text{tr}\{\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{F}\mathbf{G}\mathbf{Q}\} - \text{tr}\{\mathbf{G}'\mathbf{G}\mathbf{Q}\}\} - \text{tr}\{(\mathbf{F}'\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{F} - \mathbf{I}_{h^*})\,\mathbf{L}\}$$

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

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

$$\begin{aligned}
\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{Q}\tilde{\mathbf{G}}' &= \mathbf{Y}'\mathbf{M}\mathbf{Y}\tilde{\mathbf{F}}\tilde{\mathbf{L}} \\
\tilde{\mathbf{F}}'\mathbf{Y}'\mathbf{M}\mathbf{Y} &= \tilde{\mathbf{G}} \\
\tilde{\mathbf{F}}'\mathbf{Y}'\mathbf{M}\mathbf{Y}\tilde{\mathbf{F}} &= \mathbf{I}_{h^*} .
\end{aligned}$$

Therefore, we must find out solutions of the system

$$\begin{aligned}
\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{Q}\mathbf{G}' &= \mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{F}\mathbf{L} \\
\mathbf{F}'\mathbf{Y}'\mathbf{M}\mathbf{Y} &= \mathbf{G} \\
\mathbf{F}'\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{F} &= \mathbf{I}_{h^*}
\end{aligned}$$

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

But, premultiplying the first equation by $\mathbf{F}'$ and taking into account the remaining equations, we obtain

$$\mathbf{F}'\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{Q}\mathbf{Y}'\mathbf{M}\mathbf{Y}\mathbf{F} = \mathbf{L} .$$

This matrix has n eigenvalues $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ of which $r = r(\mathbf{Y})$ are positive, the remainder zero.

Then, associate to the first h^* positive eigenvalues $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{h^*}$ the h^* orthonormal eigenvectors $\tilde{\tilde{\mathbf{y}}}_1, \ldots, \tilde{\tilde{\mathbf{y}}}_{h^*}$.

Setting

$$\tilde{\mathbf{D}}_{h^*} = \text{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{h^*}) \quad , \quad \tilde{\tilde{\mathbf{Y}}}_{h^*} = \left[\tilde{\tilde{\mathbf{y}}}_1 \cdots \tilde{\tilde{\mathbf{y}}}_{h^*}\right],$$

we can write

$$\mathbf{Y}\mathbf{Q}\mathbf{Y}'\mathbf{M}\tilde{\tilde{\mathbf{Y}}}_{h^*} = \tilde{\tilde{\mathbf{Y}}}_{h^*}\tilde{\mathbf{D}}_{h^*} \quad , \quad \tilde{\tilde{\mathbf{Y}}}'_{h^*}\mathbf{M}\tilde{\tilde{\mathbf{Y}}}_{h^*} = \mathbf{I}_{h^*}$$

and also

$$\tilde{\tilde{\mathbf{Y}}}'_{h^*}\mathbf{MYQY'M}\tilde{\tilde{\mathbf{Y}}}_{h^*} = \tilde{\mathbf{D}}_{h^*} .$$

Thus, $\tilde{\tilde{\mathbf{y}}}_1, \ldots, \tilde{\tilde{\mathbf{y}}}_{h^*}$ are the first h^* standardized principal components of $\mathbf{VQ}$.

Thus,we immediately realize that

$$\tilde{\mathbf{F}} = \mathbf{Q}\tilde{\mathbf{C}}_{h^*}\tilde{\mathbf{D}}_{h^*}^{-1/2} \quad , \quad \tilde{\mathbf{G}} = \tilde{\mathbf{F}}'\mathbf{Y'MY} = \tilde{\tilde{\mathbf{Y}}}'_{h^*}\mathbf{MY} \quad , \quad \tilde{\mathbf{L}} = \tilde{\mathbf{D}}_{h^*}$$

represent a solution of our problem.

Summing up, both $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{G}}$ can be interpreted in terms of principal factors and standardized principal components.

In fact, while $\tilde{\mathbf{F}}$ is linked to the matrix $\mathbf{Q}\tilde{\mathbf{C}}_{h^*}$ of the principal factors by means of the matrix $\tilde{\mathbf{D}}_{h^*}^{-1/2}$, $\tilde{\mathbf{G}}$, as is at once apparent, represents the matrix of the coefficients of the orthogonal projection of $\mathbf{Y}$ on the subspace spanned by the column vectors of the matrix

$$\tilde{\tilde{\mathbf{Y}}}_{h^*} = \mathbf{Y}\ \mathbf{Q}\tilde{\mathbf{C}}_{h^*}\tilde{\mathbf{D}}_{h^*}^{-1/2} = \mathbf{Y}\tilde{\mathbf{F}} .$$

Notice that, since ($\mathbf{VQ}\tilde{\mathbf{C}}_{h^*} = \tilde{\mathbf{C}}_{h^*}\tilde{\mathbf{D}}_{h^*}$)

$$\begin{aligned}\tilde{\mathbf{H}} = \tilde{\mathbf{F}}\tilde{\mathbf{G}} \qquad\qquad &= \mathbf{Q}\tilde{\mathbf{C}}_{h^*}\tilde{\mathbf{D}}_{h^*}^{-1/2}\tilde{\mathbf{D}}_{h^*}^{-1/2}\tilde{\mathbf{C}}'_{h^*}\mathbf{QV} \\ = \mathbf{Q}\tilde{\mathbf{C}}_{h^*}\tilde{\mathbf{D}}_{h^*}^{-1}\tilde{\mathbf{D}}_{h^*}\tilde{\mathbf{C}}'_{h^*} &= \mathbf{Q}\tilde{\mathbf{C}}_{h^*}\tilde{\mathbf{C}}'_{h^*} ,\end{aligned}$$

we get

$$\mathbf{Y}\tilde{\mathbf{H}} = \mathbf{YQ}\tilde{\mathbf{C}}_{h^*}\tilde{\mathbf{C}}'_{h^*} = \tilde{\mathbf{Y}}_{h^*}\tilde{\mathbf{C}}'_{h^*} = \Sigma_h^{h^*}\tilde{\mathbf{y}}_h\tilde{\mathbf{c}}'_h .$$

Namely, $\mathbf{Y}\tilde{\mathbf{H}}$ is an approximated reconstitution of $\mathbf{Y}$ (Section 5.2.5).

REFERENCES

[l] 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] Bolasco, S., *Analisi multidimensionale dei dati*, Carocci, Roma, 1999.

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

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

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

[8] De Antoni, F., (a cura di), *I fondamenti dell'analisi dei dati*, Istituto di Statistica e Ricerca Sociale "C. Gini", Roma, 1982.

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

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

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

[12] Jackson, J.E., *A User's Guide to Principal Components*, John Wiley and Sons, New York, 1991.

[13] Jolliffe, I.T., *Principal Component Analysis*, Springer-Verlag, New

York, 2002.

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

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

[16] Lebart, L., Morineau, A., Warwick, K.M., *Multivariate Descriptive Analysis*, John Wiley and Sons, New York, 1984.

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

[18] Leoni, R., (a cura di) *Alcuni lavori di analisi statistica multivariata*, SIS, Firenze, 1982.

[19] Leoni, R., *Principal Component Analysis*, in «Methods for Multidimensional Data Analysis», European Courses in Advanced Statistics, Anacapri, 1987.

[20] 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*).

[21] 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.

[22] Marchetti, G., *Analisi in componenti principali e approssimazioni di matrici*, Dipartimento Statistico, Firenze, 1984.

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

[24] Mignani, S., Montanari, A., *Appunti di analisi statistica multivariata*, Società Editrice Esculapio, Bologna, 1998.

[25] Rao, C.R., *The Use and Interpretation of Principal Component*

Analysis in Applied Research, Sankyā A, 26, 1964.

[26] Rao, C.R., *Matrix Approximations and Reduction of Dimensionality in Multivariate Statistical Analysis*, in «Multivariate Analysis-V» (Krishnaiah, P.R., ed.), North-Holland Publishing Company, Amsterdam, 1980.

[27] Rencher, A.C., *Methods of Multivariate Analysis*, John Wiley & Sons, New York, 1995.

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

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

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

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

[32] Zani, S., *Analisi dei dati statistici II*, Giuffrè Editore, Milano, 2000.