Chapter 3

Euclidean Cloud

(*'CLOUD'*)

A cloud without metric is a cloud without shape.
J.P. Benzécri

Introduction

In this chapter, we study the central object of Geometric Data Analysis (GDA), namely, the *Euclidean cloud,* that is, a family of statistical observations conceptualized as points in a multidimensional Euclidean space[1].

Ready–made Euclidean clouds occur whenever observations are points in a plane (two–dimensional space) or in a physical three–dimensional space, structured as classical geometric Euclidean spaces, see chapter 10 (*MATH.BASES*) p.439. For instance, the impacts of bullets on a target, or the positions of the bees in a swarm define Euclidean clouds. In GDA, Euclidean clouds are constructed from numerical data sets, such as contingency tables in Correspondence Analysis, dissimilarity tables in metric or nonmetric MDS, Individuals × Variables tables in Principal Component Analysis for numerical variables and Multiple Correspondence Analysis for categorized ones.

The chapter is organized as follows. First of all, we define basic statistics (§3.1), followed by the orthogonal projection of a cloud onto a subspace and its breaking down into projected clouds (§3.2). We study the principal directions of a cloud, we present Specific Analysis (§3.3) and introduce principal hyperellipsoids (§3.4). Then we study the partition of a cloud into

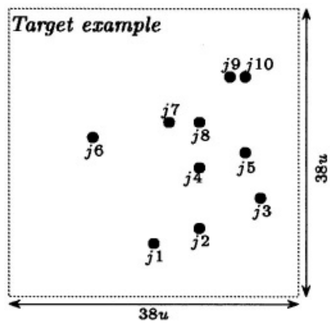
[1]In this chapter, by a space (without qualification) we always mean a multidimensional geometric space whose elements are points, as opposed to a vector space whose elements are vectors.

subclouds, with the between–within breakdown (§3.5), and introduce Euclidean classification, that is, Agglomerative Hierarchical Clustering (AHC) with aggregation according to variance (§3.6). Finally, we move on "from points to numbers", choosing a Cartesian frame and working on coordinate variables (§3.7).

The Mathematical Bases relevant to this chapter are multidimensional geometry (*MATH.BASES,* §10.4) and the theorem of spectral analysis (§10.5).

3.1 Basic Statistics

In this chapter, we will study a Euclidean cloud, in the line of pure geometry, that is, using a coordinate–free approach. Without loss of generality, this study will be conducted on a cloud of ten points in a plane, the *Target example,* as represented on the opposite figure. When it comes to numerical computations, we will take a unit length u such that the frame of the picture is a $38u \times 38u$ square.



Definition 3.1 (Euclidean cloud). *Given a Euclidean space $\mathcal{U}$ of dimensionality K, a Euclidean cloud (in brief a cloud) is a protocol whose observations $(\mathrm{M}^j)_{j\in J}$ are points in $\mathcal{U}$. The points are weighted by masses $(\varpi_j)_{j\in J}$ (with $\forall j \in J : \varpi_j > 0$).*

Denoting J the set–theoretic support of the cloud, one has:

$$\begin{array}{llll} \mathrm{M}^J : & J & \longrightarrow \mathcal{U} & \text{(Euclidean space)} \\ & j & \longmapsto \mathrm{M}^j & \text{(point)} \end{array} \qquad \begin{array}{llll} \varpi_J : & J & \longrightarrow \mathbb{R} & \\ & j & \longmapsto \varpi_j & \text{(mass)} \end{array}$$

According to the *duality notation* (*CORRESPONDENCE,* p.26), points are indexed by upper indices and *weights* (or masses) by lower indices.

Weights are usually derived from an absolute frequency measure n_J (possibly elementary) with its relative frequency measure $f_j = n_j/n$, letting $n = \sum\limits_{j\in J} n_j$ (total absolute frequency of cloud). The affine *support* of the cloud (in brief support), denoted $\mathcal{M}\subseteq\mathcal{U}$, is the smallest subspace containing the points of the cloud; the *dimensionality of the cloud,* denoted L, is by definition the dimensionality of its support with $L \leq K$ ($\dim \mathcal{M} = L$). By *plane cloud* we mean a two–dimensional cloud ($L = 2$).

The basic statistics of a Euclidean cloud (mean point, sum of squares, variance and contributions) are the multidimensional extensions of the basic statistics of a numerical protocol.

Geometric notation. *Points* are denoted by capital letters: M, P, A, etc. *Geometric vectors,* elements of the vector space $\mathcal{V}$ underlying $\mathcal{U}$, are "arrowed": $\overrightarrow{\varepsilon}$, $\overrightarrow{\alpha}$, $\overrightarrow{0}$ (null–vector); the vector associated with the bipoint (P,M) is denoted $\overrightarrow{\mathrm{PM}}$, or M − P as the *deviation* of M from P ("endpoint – origin"). The *scalar product* on $\mathcal{V}$ is denoted $\langle\cdot|\cdot\rangle$ and the *Euclidean norm* is denoted $\|\cdot\|$; the *Euclidean distance* between two points M and P is denoted MP (or PM).

3.1.1 Mean Point

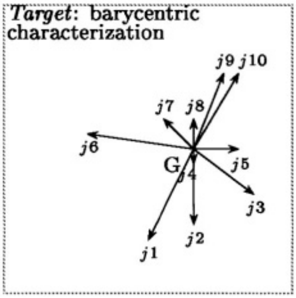
Definition 3.2 (Mean point). *Let* P *be a point in* $\mathcal{U}$; *the point* G *such that* $\overrightarrow{\mathrm{PG}} = \sum_{j\in J} f_j \overrightarrow{\mathrm{PM}^j}$ *does not depend on point* P; *it is called the mean point of the cloud and can be written* $\mathrm{G} = \sum_{j\in J} f_j\, \mathrm{M}^j$ (*MATH.BASES,* p.436).

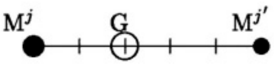
Property 3.1 (Barycentric property). *The following equations define the barycentric characterization of the mean point.*

$$\sum_{j\in J} f_j \overrightarrow{\mathrm{GM}^j} = \overrightarrow{0} \quad \textit{or} \quad \sum_{j\in J} n_j \overrightarrow{\mathrm{GM}^j} = \overrightarrow{0}$$

Proof. If we let P = G in the preceding definition, we obtain the property. ◁

For a *cloud of two points* (M^j, n_j) and $(\mathrm{M}^{j'}, n_{j'})$, the mean point G belongs to the segment. $\mathrm{M}^j\mathrm{M}^{j'}$: it is such that $n_j\overrightarrow{\mathrm{GM}^j} + n_{j'}\overrightarrow{\mathrm{GM}^{j'}} = \overrightarrow{0}$ (cf. opposite figure with $n_j = 3$, $n_{j'} = 2$; $\mathrm{GM}^j/\mathrm{GM}^{j'} = n_{j'}/n_j = 2/3$).
For $n_j = n_{j'}$, the mean point is the midpoint of the segment $\mathrm{M}^j\mathrm{M}^{j'}$.





3.1.2 Inertia, Sum of Squares, Variance and Contributions

Definition 3.3 (Inertia). *Given a weighting measure* ϖ_J *and a reference point P, the inertia of the cloud with respect to point P is the* ϖ_J*–weighted sum of the squares of the distances of its points from point* P, *and is denoted* $\mathrm{In}^{\mathrm{P}}\mathrm{M}^J$, *with:*

$$\mathrm{In}^{\mathrm{P}}\mathrm{M}^J = \sum_{j\in J} \varpi_j\, (\mathrm{PM}^j)^2$$

The part of this sum accounted for by the point M^j is denoted $\mathrm{In}_j^{\mathrm{P}}$, with $\mathrm{In}_j^{\mathrm{P}} = \varpi_j(\mathrm{PM}^j)^2$.

The inertias most commonly used in GDA are the n_J**–inertia** or *Sum of Squares,* and the f_J**–inertia** or *Mean of Squares.*

Definition 3.4 (Variance). *The variance of the cloud* M^J *is the weighted mean of the squares of the distances of its points from the mean point* G.

$$\operatorname{Var} \mathrm{M}^J = \sum_{j \in J} f_j (\mathrm{GM}^j)^2$$

The variance of the cloud M^J can also be expressed as a function of the distances between pairs of points (cf. Exercise 3.1, p.121), that is:

$$\operatorname{Var} \mathrm{M}^J = \frac{1}{2} \sum_{j \in J} \sum_{j' \in J} f_j f_{j'} (\mathrm{M}^j \mathrm{M}^{j'})^2 = \sum_{jj' \in \mathcal{P}_2(J)} f_j f_{j'} (\mathrm{M}^j \mathrm{M}^{j'})^2 \qquad (3.1)$$

where $\mathcal{P}_2(J)$ denotes the set of $J(J-1)/2$ pairs of elements of J.

Definition 3.5 (Absolute contribution). *The absolute contribution of the point* M^j *to the variance of the cloud, denoted* $\mathbf{Cta}_j$ *(*"Ct" *for Contribution,* "a" *for absolute), is the part of variance accounted for by this point:*

$$\mathbf{Cta}_j = f_j\,(\mathrm{GM}^j)^2$$

The breakdown of variance according to points is therefore:

$$\operatorname{Var} \mathrm{M}^J = \sum_{j \in J} \mathbf{Cta}_j$$

Definition 3.6 (Relative contribution). *The relative contribution of the point* M^j *to the variance of the cloud, denoted* $\mathbf{Ctr}_j$ *(*"Ct" *for Contribution,* "r" *for relative), is the proportion of variance accounted for by the point* M^j*:*

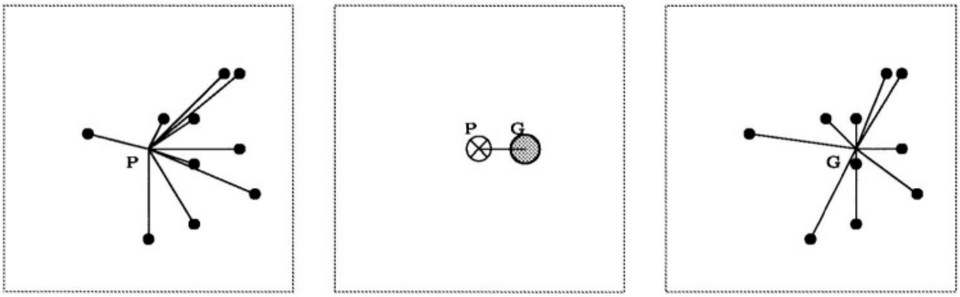
$$\mathbf{Ctr}_j = \frac{\mathbf{Cta}_j}{\operatorname{Var} \mathrm{M}^J} = \frac{n_j (\mathrm{GM}^j)^2}{\sum_{j \in J} n_j (\mathrm{GM}^j)^2}$$

Contributions add up: According to the *duality notation,* index j is written below (subscript) in $\mathbf{Cta}_j$ and $\mathbf{Ctr}_j$.

Target example. With the distance unit previously specified, the sum of squares with respect to the target center P is equal to 1280; the centered sum of squares is equal to 920, hence the variance is equal to 92.

Theorem 3.1 (First Huyghens' theorem). *The mean of the squares of the distances of the points of a cloud from a reference point* P *is equal to the variance of the cloud plus the square of the distance from the mean point of the cloud to the reference point* P.

$$\forall \mathrm{P} \in \mathcal{U} \ : \ \sum_{j\in J} f_j(\mathrm{PM}^j)^2 = (\mathrm{PG})^2 + \sum_{j\in J} f_j(\mathrm{GM}^j)^2$$



***Target example*: illustration of Huyghens' theorem.**

Proof. From $\overrightarrow{\mathrm{PM}^j} = \overrightarrow{\mathrm{PG}} + \overrightarrow{\mathrm{GM}^j}$, one gets $(\mathrm{PM}^j)^2 = \mathrm{PG}^2 + (\mathrm{GM}^j)^2 + 2\langle\overrightarrow{\mathrm{PG}}|\overrightarrow{\mathrm{GM}^j}\rangle$, and $\sum\limits_{j\in J} f_j\,(\mathrm{PM}^j)^2 = \mathrm{PG}^2 + \sum\limits_{j\in J} f_j\,(\mathrm{GM}^j)^2 + 2\langle\overrightarrow{\mathrm{PG}}|\sum\limits_{j\in J} f_j\,\overrightarrow{\mathrm{GM}^j}\rangle$. One has $\sum\limits_{j\in J} f_j\overrightarrow{\mathrm{GM}^j} = \overrightarrow{0}$ (barycentric property), hence the theorem. ◁

From Huyghens' theorem, the *metric characterization of the mean point* follows: If to each point $\mathrm{P} \in \mathcal{U}$ there is assigned the quantity $\sum\limits_{j\in J} f_j(\mathrm{PM}^j)^2$, this quantity is minimum when P coincides with the mean point G. In other words, the mean point is the *point of least squares;* among the clouds that reduce to a single point (i.e. zero–dimensional clouds), the one which is concentrated in point G is the one which best fits the observed cloud in the sense of least squares.

3.2 Projected Clouds

Let (M^J, n_J) be a Euclidean cloud, and let $\mathcal{H}$ be a subspace of $\mathcal{U}$. Let H^j be the orthogonal projection of point M^j onto $\mathcal{H}$ (*MATH.BASES,* p.439).

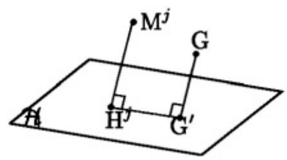


Figure 3.1: $\dim \mathcal{U} = 3, \dim \mathcal{H} = 2$

Definition 3.7 (Projected cloud). *The orthogonal projection onto* $\mathcal{H}$ *of the cloud* M^J*, or projected cloud, is the cloud* H^J *whose points are the orthogonal projections of points of the cloud* M^J *onto* $\mathcal{H}$.

Property 3.2. *The mean point of the cloud projected onto $\mathcal{H}$ is the projection onto $\mathcal{H}$ of the mean point of the cloud M^J.*

Proof. As an affine mapping, projection preserves barycenters, that is, from $\mathrm{M}^j \mapsto \mathrm{H}^j$ one gets $\sum\limits_{j\in J} f_j\,\mathrm{M}^j \mapsto \sum\limits_{j\in J} f_j\,\mathrm{H}^j$, therefore $\mathrm{G} \mapsto \mathrm{G}'$ (G' is the mean point of the projected cloud H^J). ◁

— *Remark.* In this chapter, all projections are orthogonal ones; consequently, we will often omit "orthogonal".

3.2.1 Variance in a Direction

Property 3.3 (Variance in a direction). *Projected clouds onto parallel subspaces of same dimensionalities have the same variances; their common value is called the variance of cloud in the direction of these subspaces.*

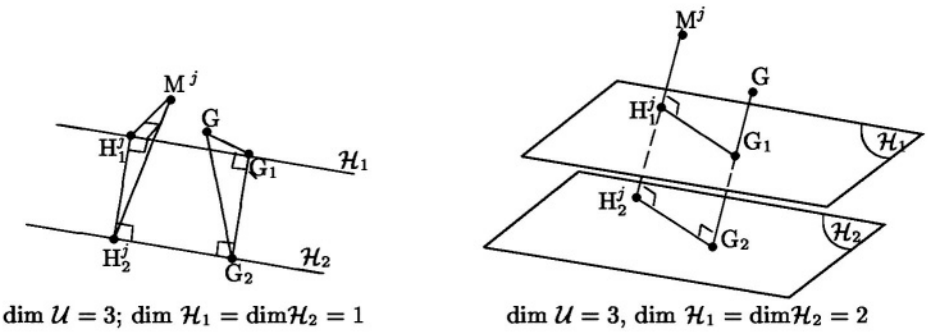


Figure 3.2. Projections onto parallel subspaces (of same dimensionalities)

Proof. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two parallel subspaces (of same dimensionality); let H_1^J and H_2^J be the clouds projected onto $\mathcal{H}_1$ and $\mathcal{H}_2$, and G_1 and G_2 their mean points.

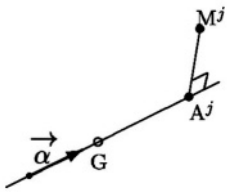
The cloud H_2^J is obtained from the cloud H_1^J by the translation of vector $\overrightarrow{\mathrm{G}_1\mathrm{G}_2}$; one has $\forall j \in J\ :\ \mathrm{H}_2^j = \mathrm{H}_1^j + \overrightarrow{\mathrm{G}_1\mathrm{G}_2}$ (Figure 3.2) and $\overrightarrow{\mathrm{G}_1\mathrm{H}}{}_1^j = \overrightarrow{\mathrm{G}_2\mathrm{H}}{}_2^j$, hence $\sum\limits_{j\in J} f_j\,(\mathrm{G}_1\mathrm{H}_1^j)^2 = \sum\limits_{j\in J} f_j\,(\mathrm{G}_2\mathrm{H}_2^j)^2$. ◁

A *line* (one–dimensional subspace) defined by a point and a vector $\overrightarrow{\alpha}$ ("direction vector") is called an *axis;* the variance of the projected cloud onto the line is called *variance in direction* $\overrightarrow{\alpha}$, or *variance of axis* (even though it does not depend on orientation).

Property 3.4. *The variance of the cloud* M^J *in direction* $\overrightarrow{\alpha}$ *is equal to:*

$$\sum_{j\in J} f_j \frac{\langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha} \rangle^2}{\|\overrightarrow{\alpha}\|^2}$$

Proof. The projected point of M^j onto the axis $(\mathrm{G}, \overrightarrow{\alpha})$ is the point A^j defined by the formula (*MATH.BASES*, p.439): $\overrightarrow{\mathrm{GA}}^j = \frac{\langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha} \rangle}{\|\overrightarrow{\alpha}\|^2} \overrightarrow{\alpha}$; hence:
$\mathrm{Var}\,\mathrm{A}^J = \sum_{j\in J} f_j (\mathrm{GA}^j)^2 = \sum_{j\in J} f_j \frac{\langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha} \rangle^2}{\|\overrightarrow{\alpha}\|^2}$. ◁



Definition 3.8 (Spherical cloud). *A cloud is spherical if its variance is the same in all the directions of its support.*

3.2.2 Residual Square Mean

Definition 3.9 (Residual square mean). *The mean of residual squares (or residual square mean) of a cloud with respect to a subspace is the weighted mean of the squares of the distances from the points of the cloud to this subspace.*

Let H^j be the orthogonal projection of M^j onto $\mathcal{H}$, hence the distance $\mathrm{M}^j\mathrm{H}^j$ of M^j to $\mathcal{H}$. The residual square mean of M^J with respect to $\mathcal{H}$ is equal to $\sum_{j\in J} f_j\, (\mathrm{M}^j\mathrm{H}^j)^2$.

Theorem 3.2 (General Huyghens' theorem). *The mean of residual squares of a cloud with respect to a subspace is equal to its residual square mean with respect to the parallel subspace of same dimensionality going through the mean point, plus the square of the distance from the mean point to this subspace.*

Let $\mathcal{A}$ be the subspace parallel to $\mathcal{H}$ going through the mean point G. Denoting A^j the projection of the point M^j onto $\mathcal{A}$, and denoting H^j and G' the respective orthogonal projections of points M^j and G onto $\mathcal{H}$ (cf. Figure 3.3, p.82), the theorem reads:

$$\sum_{j\in J} f_j\, (\mathrm{M}^j\mathrm{H}^j)^2 = \sum_{j\in J} f_j (\mathrm{M}^j\mathrm{A}^j)^2 + (\mathrm{GG}')^2$$

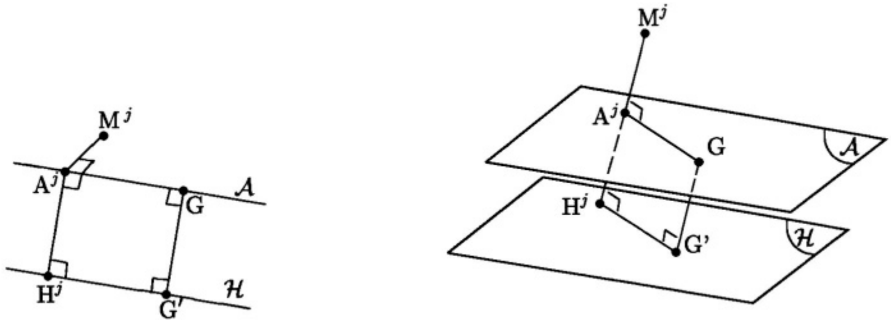


Figure 3.3. $\dim \mathcal{U} > \dim\mathcal{H} = 1$ **and** $\dim \mathcal{U} > \dim\mathcal{H} = 2$

Proof. One has: $\overrightarrow{M^jH^j} = \overrightarrow{M^jA^j} + \overrightarrow{A^jH^j}$, with $\overrightarrow{A^jH^j} = \overrightarrow{GG'}$, therefore: $(M^jH^j)^2 = (M^jA^j)^2 + (GG')^2 + 2\langle\overrightarrow{M^jA^j}|\overrightarrow{GG'}\rangle$. From barycentric property (p.77) one has $\sum\limits_{j\in J} f_j \overrightarrow{GM}^j = \overrightarrow{0} = \sum\limits_{j\in J} f_j \overrightarrow{GA}^j$, hence $\sum\limits_{j\in J} f_j \overrightarrow{M^jA^j} = \overrightarrow{0}$; it follows that $\sum\limits_{j\in J} f_j (M^jH^j)^2 = \sum\limits_{j\in J} f_j(M^jA^j)^2 + (GG')^2$. ◁

3.2.3 Fitted and Residual Clouds

In the fitting perspective, whenever one replaces the cloud M^J by a cloud of lower dimensionality, the projected cloud A^J is called a *fitted cloud,* and $\forall j \in J$ the vector $\overrightarrow{A^jM}^j$ is called the *residual deviation,* where:

$$\forall j \in J \quad M^j = A^j + \overrightarrow{A^jM}^j \quad \text{with} \quad \overrightarrow{A^jM}^j \perp \mathcal{A}$$

observed point = fitted point + residual deviation

Property 3.5. *If* G' *and* H^j *are the orthogonal projections of points* G *and* M^j *onto the subspace* $\mathcal{H}$, *the cloud* R^J *defined by* $R^j = G' + \overrightarrow{H^jM}^j$ $(j \in J)$ *only depends on the direction of* $\mathcal{H}$.

Proof.
$$\begin{aligned} R^j &= G' + \overrightarrow{H^jM}^j \\ &= (G' + \overrightarrow{G'G}) + \overrightarrow{GG'} + \overrightarrow{H^jM}^j;\ \text{now}\quad \overrightarrow{GG'} = \overrightarrow{A^jH}^j \\ &= \quad G \quad + \overrightarrow{A^jH^j} + \overrightarrow{H^jM^j} = G + \overrightarrow{A^jM}^j \end{aligned}$$
◁

Definition 3.10 (Residual cloud). *The residual cloud is defined by:*

$$\forall j \in J\ :\ R^j = G + \overrightarrow{A^jM}^j$$

residual point = mean point + residual deviation

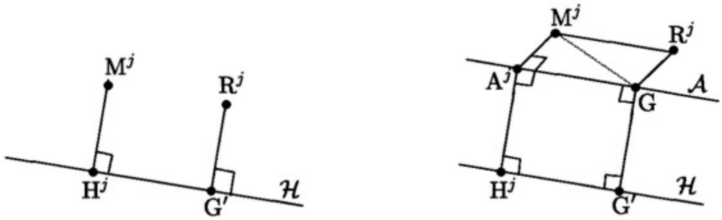


Figure 3.4. Residual cloud ($\dim \mathcal{H} = 1$, $\dim \mathcal{U} > 1$)

By construction, G is the mean point of the residual cloud R^J; the support of R^J is orthogonal to $\mathcal{H}$; hence the orthogonal breakdown:

$$\overrightarrow{\mathrm{GM}}^j = \overrightarrow{\mathrm{GA}}^j + \overrightarrow{\mathrm{GR}}^j \quad \text{with } \overrightarrow{\mathrm{GA}}^j \perp \overrightarrow{\mathrm{GR}}^j \text{ (cf. figure 3.4)}$$

The variance of the residual cloud is equal to the residual square mean of the cloud M^J with respect to $\mathcal{A}$: $\mathrm{Var}\,\mathrm{R}^J = \sum\limits_{j\in J} f_j(\mathrm{GR}^j)^2 = \sum\limits_{j\in J} f_j(\mathrm{A}^j\mathrm{M}^j)^2$.

Property 3.6 (Breakdown of variance of cloud). *The variance of a cloud is equal to its variance in the direction of $\mathcal{H}$ (variance of the projected cloud), plus the variance of the residual cloud.*

$$\mathrm{Var}\,\mathrm{M}^J = \mathrm{Var}\,\mathrm{A}^J + \mathrm{Var}\,\mathrm{R}^J \tag{3.2}$$

Total variance = fitted variance + residual variance

Proof. $\overrightarrow{\mathrm{GM}}^j = \overrightarrow{\mathrm{GA}}^j + \overrightarrow{\mathrm{GR}}^j$, with $\overrightarrow{\mathrm{GA}}^j \perp \overrightarrow{\mathrm{GR}}^j$, now $\mathrm{GR}^j = \mathrm{A}^j\mathrm{M}^j$, hence: $(\mathrm{GM}^j)^2 = (\mathrm{GA}^j)^2 + (\mathrm{A}^j\mathrm{M}^j)^2$. ◁

3.2.4 Variables Attached to an Axis

Let $(\mathrm{G}, \overrightarrow{\alpha})$ be an axis, one defines the projected cloud $\mathrm{A}^J = (\mathrm{A}^j)_{j\in J}$ onto this axis and the following variables.

The *covariant variable* $\alpha^J = (\alpha^j)_{j\in J}$ with $\alpha^j = \langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha} \rangle$; α^j is the *covariant coordinate* of the point M^j with respect to $(\mathrm{G}, \overrightarrow{\alpha})$. The α^J variable is centered and its variance is $\mathrm{Var}\,\mathrm{A}^J \,\|\overrightarrow{\alpha}\|^2$.

The *calibrated variable* y^J with $y^j = \alpha^j / \|\overrightarrow{\alpha}\|$. The variable y^J is centered and its variance is $\mathrm{Var}\,\mathrm{A}^J$.

The *standard variable* z^J with $z^j = y^j / \mathrm{SD}\, y^J = y^j / \mathrm{SD}\,\mathrm{A}^J$ (Mean $z^J = 0$ and $\mathrm{Var}\, z^J = 1$).

The *axial variable* t^J with $t^j = \langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha} \rangle / \|\overrightarrow{\alpha}\|^2$. The variable t^J is centered and its variance is $\mathrm{Var}\, t^J = \mathrm{A}^J / \|\overrightarrow{\alpha}\|^2$; t^j is the coordinate of A^j on axis $(\mathrm{G}, \overrightarrow{\alpha})$ (*MATH.BASES*, p.440).

If $\overrightarrow{\alpha}$ is multiplied by a coefficient κ, the covariant coordinate α^j of point M^j is multiplied by κ, hence the term "covariant" (nothing to do with the covariance of two variables). The calibrated variable y^J is the covariant variable associated with the unit–norm vector $\overrightarrow{\alpha}/\|\overrightarrow{\alpha}\|$. The covariant variable α^J is mainly of mathematical interest; the calibrated and the standard variables are *intrinsic* (up to orientation) and more directly interpretable statistically.

Dimensional considerations (*MATH.BASES*, p.442). The norm of a geometric vector is assimilable to a *length,* that is, it is of dimension 1; the scalar product of two vectors is of dimension 2. The covariant variable is therefore of dimension 2, the calibrated variable is of dimension 1, the standard variable is "without dimension". The properties of Euclidean clouds — being pure geometrical properties — are *homogeneous,* that is, they do not depend on unit length.

3.2.5 Linear Formalization

We now introduce the linear mappings that will intervene in the determination of principal directions; these mappings are defined in terms of the following vector spaces: the Euclidean vector space $\mathcal{V}$ underlying the geometric space $\mathcal{U}$; the vector space $\mathbb{R}^J$ of variables over J; and the vector space $\mathbb{R}_J$ of measures over J.

The linear algebra notions used in the rest of this chapter are recalled in chapter 10 (*MATH.BASES*).

Definition 3.11 (Linear mapping $\boldsymbol{Vac}$). *Given a point* $\mathrm{P}\in\mathcal{U}$, *the mapping* $\boldsymbol{Vac}_\mathrm{P}$ *("**Va**" for **Va**riable, "**c**" for **c**ovariant) is the linear mapping (homomorphism) (depending on P) such that:*

$$\begin{aligned} \boldsymbol{Vac}_\mathrm{P} : \mathcal{V} &\longrightarrow \mathbb{R}^J \\ \overrightarrow{\alpha} &\longmapsto \alpha^J_\mathrm{P} \quad \textbf{with}\ \alpha^J_\mathrm{P} = \left(\langle\overrightarrow{\mathrm{PM}}^j|\overrightarrow{\alpha}\rangle\right)_{j\in J} \end{aligned}$$

Let $\boldsymbol{Vac}$ denote $\boldsymbol{Vac}_\mathrm{P}$ when P = G. From $\overrightarrow{\mathrm{PM}}^j = \overrightarrow{\mathrm{PG}} + \overrightarrow{\mathrm{GM}}^j$ we deduce $\langle\overrightarrow{\mathrm{PM}}^j|\overrightarrow{\alpha}\rangle = \langle\overrightarrow{\mathrm{PG}}\,|\overrightarrow{\alpha}\rangle + \langle\overrightarrow{\mathrm{GM}}^j|\overrightarrow{\alpha}\rangle$, that is, $\alpha^j_\mathrm{P} = \langle\overrightarrow{\mathrm{PG}}\,|\overrightarrow{\alpha}\rangle + \alpha^j_\mathrm{G}$, hence:

$$\forall\overrightarrow{\alpha}\in\mathcal{V}\quad \boldsymbol{Vac}_\mathrm{P}(\overrightarrow{\alpha}) = \langle\overrightarrow{\mathrm{PG}}\,|\overrightarrow{\alpha}\rangle 1^J + \boldsymbol{Vac}(\overrightarrow{\alpha}) \tag{3.3}$$

Definition 3.12 (Linear mapping $\boldsymbol{Eff}$). *Given a point* $\mathrm{P}\in\mathcal{U}$, *the mapping* $\boldsymbol{Eff}^\mathrm{P}$ *($\boldsymbol{Eff}$ for* **Eff***ect) is the linear mapping such that:*

$$\begin{aligned} \boldsymbol{Eff}^\mathrm{P} : \mathbb{R}_J &\longrightarrow \mathcal{V} \\ u_J &\longmapsto \sum_{j\in J} u_j\overrightarrow{\mathrm{PM}}^j \end{aligned}$$

Let $u = \sum\limits_{j \in J} u_j$. If mbox$u \neq 0$, $\mathit{Eff}^{\mathrm{P}}$ depends on P and $\mathit{Eff}^{\mathrm{P}}(u_J) = u\,\overrightarrow{\mathrm{PG}}$. If $u = 0$ (u_J is a contrast), the vector $\sum\limits_{j \in J} u_j \overrightarrow{\mathrm{PM}}^j$ does not depend on P, it is denoted $\sum\limits_{j \in J} u_j \mathrm{M}^j$ (*MATH.BASES*, p.436) and called *vector–effect.*

From the orthogonal breakdown of u_J as the contrast $u_J^0 = (u_j - u f_j)_{j \in J}$ and the measure $u\, f_J$, one gets $\mathit{Eff}^{\mathrm{P}}(u_J) = \mathit{Eff}^{\mathrm{P}}(u_J^0) + u\,\mathit{Eff}^{\mathrm{P}}(f_J)$.

Letting Eff be the Effect mapping with G as a reference point, we have:
$\forall u_j \in \mathbb{R}_J$: $\mathit{Eff}(u_J) = \sum\limits_{j \in J} u_j \overrightarrow{\mathrm{GM}}^j = \sum\limits_{j \in J} u_j^0 \mathrm{M}^j = \mathit{Eff}(u_J^0)$, hence:

$$\forall u_J \in \mathbb{R}_J \quad \mathit{Eff}^{\mathrm{P}}(u_J) = u\,\mathit{Eff}^{\mathrm{P}}(f_J) + \mathit{Eff}(u_J) \quad \text{with } \mathit{Eff}^{\mathrm{P}}(f_J) = \overrightarrow{\mathrm{PG}} \qquad (3.4)$$

— *Remark.* In Analysis of Variance, when u_J is a contrast, the vector $\sum u_j \mathrm{M}^j$ is interpreted as the *Effect* of the contrast on the cloud; hence the name *Effect* that we use for this mapping.

Duality scheme. Let us now recall (*CORRESPONDENCE,* p.27) that with each variable x^J the isomorphism $f_J : \mathbb{R}^J \to \mathbb{R}_J$ associates the measure $x_J = (f_j\, x^j)_{j \in J}$ such that x^J is the f_J–density of x_J. We are thus led to the following *duality scheme:*

$$\begin{tikzcd} & \mathbb{R}_J \arrow[dl, "\mathit{Eff}^{\mathrm{P}}"'] \arrow[dd, shift right, "f_J^{-1}"'] \\ \mathcal{V} \arrow[dr, "\mathit{Vac}_{\mathrm{P}}"'] & \\ & \mathbb{R}^J \arrow[uu, shift right, "f_J"'] \end{tikzcd}$$

In connection with the duality scheme, there are the homomorphism $\mathit{Vac}^*_{\mathrm{P}} : \mathbb{R}^J \to \mathcal{V}$, adjoint of $\mathit{Vac}_{\mathrm{P}}$, as well as the two symmetric and positive homomorphisms $\mathit{Som}_{\mathrm{P}} : \mathcal{V} \to \mathcal{V}$ and $\mathit{Tom}_{\mathrm{P}} : \mathbb{R}^J \to \mathbb{R}^J$ (*MATH.BASES*, p.433).

Property 3.7. *The homomorphism* $\mathit{Eff}^{\mathrm{P}} \circ f_J$ *is the adjoint of* $\mathit{Vac}_{\mathrm{P}}$.

Proof. One has:

$$\begin{tikzcd} & \mathbb{R}_J \arrow[dl, "\mathit{Eff}^{\mathrm{P}}"'] \\ \mathcal{V} & \\ & \mathbb{R}^J \arrow[uu, "f_J"'] \arrow[ul, "\mathit{Vac}^*_{\mathrm{P}}"] \end{tikzcd}$$

$\langle \mathit{Vac}_{\mathrm{P}}(\overrightarrow{\alpha}) | x^J \rangle = \sum\limits_{j \in J} f_j x^j \alpha^j_{\mathrm{P}} = \sum\limits_{j \in J} f_j x^j \langle \overrightarrow{\mathrm{PM}}^j | \overrightarrow{\alpha} \rangle$
and $(\mathit{Eff}^{\mathrm{P}} \circ f_J)(x^J) = \sum\limits_{j \in J} f_j x^j \overrightarrow{\mathrm{PM}}^j$. Hence
$\forall \overrightarrow{\alpha} \in \mathcal{V}\ \forall x^J \in \mathbb{R}^J$: $\langle \mathit{Vac}_{\mathrm{P}}(\overrightarrow{\alpha}) | x^J \rangle = \langle \overrightarrow{\alpha} | \mathit{Eff}^{\mathrm{P}} \circ f_J(x^J) \rangle$
and the property of adjoint homomorphisms (*MATH.BASES*, Theorem 10.3, p.432). ◁

Linear mapping Vac^*. Let Vac^*_P be the homomorphism $Eff^P \circ f_J$:

$$\begin{aligned} Vac^*_P : \mathbb{R}^J &\longrightarrow \mathcal{V} \\ x^J &\longmapsto \sum_{j\in J} f_j\, x^j \overrightarrow{PM}^j \end{aligned}$$

For P = G, we denote Vac the linear mapping. From $\overrightarrow{PM}^j = \overrightarrow{PG} + \overrightarrow{GM}^j$ one deduces: $\sum\limits_{j\in J} f_j x^j \overrightarrow{PM}^j = \overline{x}\overrightarrow{PG} + \sum\limits_{j\in J} f_j x^j \overrightarrow{GM}^j$ with $\overrightarrow{PG} = Vac^*_P(1^J)$ and $\overline{x} = \sum\limits_{j\in J} f_j x^j$, hence:

$$\forall x^J \in \mathbb{R}^J:\ Vac^*_P(x^J) = \overline{x}\ Vac^*_P(1^J) + Vac^*(x^J) \tag{3.5}$$

Definition 3.13 (Endomorphism Som). *$Eff^P \circ f_J \circ Vac_P$ is the symmetric endomorphism on $\mathcal{V}$ associated with Vac_P and denoted Som_P.*

$$\begin{aligned} Som_P : \mathcal{V} &\longrightarrow \mathcal{V} \\ \vec{\alpha} &\longmapsto \sum_{j\in J} f_j \langle \overrightarrow{PM}^j | \vec{\alpha} \rangle \overrightarrow{PM}^j \end{aligned}$$

$$\begin{tikzcd} \sum\limits_{j\in J} f_j \alpha^j_P \overrightarrow{PM}^j & (f_j\alpha^j_P)_{j\in J} \arrow[l, mapsto, "Eff^P"'] \\ \vec{\alpha} \arrow[u, mapsto, "Som_P"] \arrow[r, mapsto, "Vac"] & \alpha^J_P \arrow[u, mapsto, "f_J"'] \end{tikzcd}$$

$Som_P = Vac^*_P \circ Vac_P$ is a positive endomorphism of $\mathcal{V}$ (MATH.BASES, p.433).

Denoting Som the endomorphism with G as reference point, one has:

$$\forall \vec{\alpha} \in \mathcal{V} : Som_P(\vec{\alpha}) = \langle \overrightarrow{PG} | \vec{\alpha} \rangle \overrightarrow{PG} + Som(\vec{\alpha}) \tag{3.6}$$

Property 3.8. *The variance of cloud in direction $\vec{\alpha}$ is given by:*

$$\frac{\| Vac(\vec{\alpha})\|^2}{\|\vec{\alpha}\|^2} = \frac{\langle\ Som(\vec{\alpha}) | \vec{\alpha} \rangle}{\|\vec{\alpha}\|^2}$$

Proof. One has: $\| Vac(\vec{\alpha})\|^2 = \mathrm{Var}\,\alpha^J = \mathrm{Var}\,\mathrm{A}^J\, \|\vec{\alpha}\|^2$ (cf. §3.2.4, p.83), and $\langle Som(\vec{\alpha}) | \vec{\alpha} \rangle = \| Vac(\vec{\alpha})\|^2$. ◁

Definition 3.14 (Endomorphism Tom). *$Vac_P \circ Eff^P \circ f_J$ is the symmetric endomorphism on $\mathbb{R}^J$ associated with Vac_P and denoted Tom_P.*

$$\begin{aligned} Tom_P : \mathbb{R}^J &\to \mathbb{R}^J \\ x^J &\mapsto \Big(\sum_{j'\in J} f_{j'} x^{j'} \langle \overrightarrow{PM}^{j'} | \overrightarrow{PM}^j \rangle \Big)_{j\in J} \end{aligned}$$

$$\begin{tikzcd} \sum\limits_{j\in J} f_j x^j \overrightarrow{PM}^j \arrow[d, mapsto, "Vac_P"'] & (f_j x^j)_{j\in J} \arrow[l, mapsto, "Eff^P"'] \\ Vac_P(\sum\limits_{j\in J} f_j x^j \overrightarrow{PM}^j) & x^J \arrow[l, mapsto, "Tom_P"'] \arrow[u, mapsto, "f_J"'] \end{tikzcd}$$

Bilinear and quadratic forms. We now take the mean point G as reference point.

- The bilinear form connected with Vac is the mapping $\mathcal{V}\times \mathbb{R}^J \to \mathbb{R}$ that associates with $(\overrightarrow{\alpha}, x^J)$ the scalar product $\langle Vac(\overrightarrow{\alpha})|x^J\rangle$, i.e. the covariance between the covariant variable α^J and the variable x^J:
$$\langle Vac(\overrightarrow{\alpha})|x^J\rangle = \sum_{j\in J} f_j x^j \langle \overrightarrow{\mathrm{GM}}^j|\overrightarrow{\alpha}\rangle = \sum_{j\in J} f_j \alpha^j x^j = \mathrm{Cov}(\alpha^J|x^J)$$
(since the covariant variable α^J is centered).

- The bilinear form connected with Som is the mapping $\mathcal{V}\times\mathcal{V}\to\mathbb{R}$ that associates with $(\overrightarrow{\alpha}, \overrightarrow{\beta})$ the scalar product $\langle \overrightarrow{\alpha}|Som(\overrightarrow{\beta})\rangle$, i.e. the covariance of the covariant variables associated with the vectors $\overrightarrow{\alpha}$ and $\overrightarrow{\beta}$.
Indeed, $\langle \overrightarrow{\alpha}|Som(\overrightarrow{\beta})\rangle = \langle \overrightarrow{\alpha}|\sum_{j\in J} f_j \langle \overrightarrow{\mathrm{GM}}^j|\overrightarrow{\beta}\rangle \overrightarrow{\mathrm{GM}}^j\rangle = \mathrm{Cov}(\alpha^J|\beta^J)$.
The quadratic form connected with Som is the mapping $\mathcal{V}\to\mathbb{R}$ that associates with $\overrightarrow{\alpha}\in\mathcal{V}$ the variance of α^J: $\langle Som(\overrightarrow{\alpha})|\overrightarrow{\alpha}\rangle = \mathrm{Var}\,\alpha^J$.

- The bilinear form connected with Tom is the mapping $\mathbb{R}^J\times\mathbb{R}^J\to\mathbb{R}$ that associates with the pair of variables (x^J, y^J) the scalar product $\langle x^J|Tom(y^J)\rangle = \langle Eff(x_J)|Eff(y_J)\rangle$ (with $x_J = (f_j x^j)_{j\in J}$ and $y_J = (f_j y^j)_{j\in J}$). The quadratic form connected with Tom is the mapping form $\mathbb{R}^J\to\mathbb{R}$ that associates with the variable x^J the square of the norm of the effect of x_J, that is, $\|\sum_{j\in J} x_j \overrightarrow{\mathrm{GM}}^j\|$.

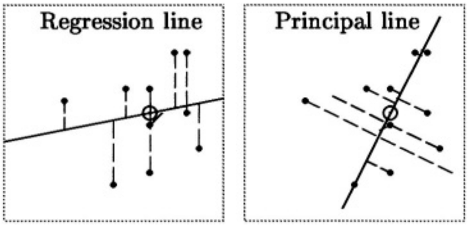
3.3 Principal Directions of a Cloud

The problem of principal directions. Given a high–dimensional cloud, one seeks to approximate it by a cloud in a lower dimensional space, such as a one–dimensional cloud (line) or a two–dimensional cloud (plane), etc. such that the approximated cloud is intuitively "of greatest elongation". When operationalized in terms of *variance,* this requirement leads to the *problem of principal directions.* For any given L' (with $1\le L'\le L$), among the L'–dimensional projected clouds, we search the projected cloud having maximum variance. From the variance breakdown property (p.83), it is equivalent to search for the L'–dimensional subspace going through the mean point G for which the residual square mean is minimum. This subspace is called the *first L'–dimensional principal subspace,* and its direction, the *first L'–dimensional principal direction.* As will be seen, for any given

L', the first principal direction always exists and is generally unique. Thus there is in general a unique first principal line ($L' = 1$), then a unique first principal plane ($L' = 2$), etc., hence a *hierarchy of principal directions.* Furthermore, this hierarchy possesses the *heredity property,* that is, the first principal plane contains the first principal line, etc.

— *Remarks.* The "principal concept" extends to $L' = 0$, the mean point is the "principal zero–dimensional subspace" (first Huyghens' theorem, p.79).

The problem of principal directions is akin to that of *regression.* In both problems we search a fitting which is optimal in the sense of least squares. In regression, the least squares are *parallel* to a given direction and only involve affine structures; whereas in principal directions, they are "orthogonal least squares" (perpendicular to the sought direction) and involve the Euclidean distance.



The case of a *plane cloud* is thoroughly treated by elementary algebra in Exercise 3.3 (p.123); the results are applied to the *Target example* in Exercise 3.4 (p.125).

3.3.1 Principal Axes

For the first principal line, the problem reads: Among all one–dimensional directions, find one along which the variance of cloud is maximum.

From Property 3.8 (p.86) the variance of the projected cloud on a line directed by $\vec{\alpha}$ is $(\|\mathbf{Vac}(\vec{\alpha})\|/\|\vec{\alpha}\|)^2$. Therefore one looks for $\vec{\alpha}$ such that the ratio $\|\mathbf{Vac}(\vec{\alpha})\|/\|\vec{\alpha}\|$ is maximum.

The solution is provided by the *spectral analysis* of the linear mapping $\mathbf{Vac}$ or of the endomorphism $\mathbf{Som}$, which leads to the eigendirection and eigenvalue equation $\mathbf{Som}(\vec{\alpha}) = \lambda\,\vec{\alpha}$. The maximum of $\|\mathbf{Vac}(\vec{\alpha})\|/\|\vec{\alpha}\|$ is reached for any eigenvector of $\mathbf{Som}$ associated with its maximum eigenvalue λ_1 (MATH.BASES, p.446), and is equal to the maximum singular value $\xi_1 = \sqrt{\lambda_1}$.

Theorem 3.3 (Principal direction equation). *The direction vectors of principal lines are eigenvectors of the endomorphism* $\mathbf{Som}$.

$$\forall \ell = 1, \dots L \;:\; \mathbf{Som}(\vec{\alpha_\ell}) = \lambda_\ell \vec{\alpha_\ell}$$

Proof. Any eigenvector $\overrightarrow{\alpha_1^*}$ of the endomorphism $\boldsymbol{Som}$ associated with the greatest eigenvalue λ_1 defines a *first principal one–dimensional direction* (or line direction); $\overrightarrow{\alpha_1^*}$ is called *first principal vector.*

If the eigenvalue λ_1 is simple, the first principal line direction is unique, the axis defined by $(\mathrm{G}, \overrightarrow{\alpha_1^*})$ is the first principal axis of the cloud M^J; the projected cloud A_1^J onto this axis is the first one–dimensional principal cloud, its variance (i.e. the variance along direction $\overrightarrow{\alpha_1^*}$) is equal to λ_1.

If the eigenvalue λ_1 is multiple of order p, there is a p–dimensional subspace for which every line direction is principal; a first principal line direction will be any line direction arbitrarily chosen among them, yielding a first one–dimensional principal cloud of variance λ_1.

With a first one–dimensional principal cloud A_1^J there is associated a residual cloud R_1^J such that $\overrightarrow{\mathrm{GR}}_1^j = \overrightarrow{\mathrm{GM}}_1^j - \overrightarrow{\mathrm{GA}}_1^j$ $(j \in J)$ (cf. Definition 3.10, p.82). If the residual cloud R_1^J is concentrated in the point G, that is, if M^J is one–dimensional, the search of principal direction is completed. If this is not the case, one determines the first principal direction of the residual cloud R_1^J, which leads to the *second principal line direction* of the cloud M^J. Let $\boldsymbol{Vac'} : \mathcal{V} \to \mathbb{R}^J$ be the linear mapping associated with the residual cloud R^J defined by $\boldsymbol{Vac'}(\overrightarrow{\alpha}) = (\langle \overrightarrow{\mathrm{GR}}_1^j | \overrightarrow{\alpha} \rangle)_{j \in J}$. One has $\boldsymbol{Vac}(\overrightarrow{\alpha}) = (\langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha} \rangle)_{j \in J}$ and $\boldsymbol{Vac}_1(\overrightarrow{\alpha}) = (\langle \overrightarrow{\mathrm{GA}}_1^j | \overrightarrow{\alpha} \rangle)_{j \in J}$, therefore $\boldsymbol{Vac'} = \boldsymbol{Vac} - \boldsymbol{Vac}_1$. Let likewise $\boldsymbol{Vac'^*} = \boldsymbol{Vac^*} - \boldsymbol{Vac_1^*} : \mathbb{R}^J \to \mathcal{V}$ such that $\boldsymbol{Vac}_1^*(x^J) = \sum_{j \in J} f_j x^j \overrightarrow{\mathrm{GA}}_1^j$. It is easily proved that $\boldsymbol{Som'}$ $(= \boldsymbol{Vac'^*} \circ \boldsymbol{Vac'})$ is equal to $\boldsymbol{Som} - \boldsymbol{Som}_1$, with $\boldsymbol{Som}_1 = \boldsymbol{Vac}_1^* \circ \boldsymbol{Vac}_1$. Now $\boldsymbol{Som'}$ has the same eigenvectors as $\boldsymbol{Som}$ (*MATH.BASES,* Property 10.28, p.443), and $\overrightarrow{\alpha_1^*}$ is an eigenvector of $\boldsymbol{Som'}$ associated with the eigenvalue 0, and the greatest eigenvalue of $\boldsymbol{Som'}$ is λ_2 (possibly equal to λ_1, when λ_1 is not simple). As a result, the maximum of $\| \boldsymbol{Vac'}(\overrightarrow{\alpha}) \| / \| \overrightarrow{\alpha} \|$ is ξ_2 and it is attained by $\overrightarrow{\alpha_2^*}$, an eigenvector of $\boldsymbol{Som}$ associated with the second eigenvalue λ_2, that is, $\boldsymbol{Som}(\overrightarrow{\alpha_2^*}) = \lambda_2 \overrightarrow{\alpha_2^*}$ and $\overrightarrow{\alpha_2^*} \perp \overrightarrow{\alpha_1^*}$ since the mapping $\boldsymbol{Som}$ is symmetric (*MATH.BASES*, Property 10.26, p.443).

From the residual cloud $(\overrightarrow{\mathrm{GM}}^j - \overrightarrow{\mathrm{GA}}_1^j - \overrightarrow{\mathrm{GA}}_2^j)_{j \in J}$, one proceeds in the same way to determine a *third principal direction,* and so on, until the determination of an L-th one–dimensional principal cloud, with a zero–dimensional residual cloud. ◁

The construction of the L line directions entails the following property.

Property 3.9. *The principal vectors* $(\overrightarrow{\alpha_\ell^*})_{\ell=1,\ldots L}$ *are pairwise orthogonal.*

The line defined by $(\mathrm{G}; \overrightarrow{\alpha_\ell})$ is called the ℓ–th principal axis, or in brief *Axis* ℓ. The projected cloud A^J_ℓ onto Axis ℓ is called the *ℓ–thprincipal one–dimensional cloud.* The variance of the cloud A^J_ℓ is equal to λ_ℓ and is called variance of Axis ℓ.
Notation. The index ℓ is here a mere enumeration index, arbitrarily put as a lower index without reference to duality; as opposed to its position in the notation of contributions (cf. later §3.3.3)

Property 3.10 (Principal breakdown of variance). *The variance of the cloud is equal to the sum of eigenvalues:*

$$\mathrm{Var}\,\mathrm{M}^J = \sum_{\ell=1}^{L} \lambda_\ell$$

The breakdown of the variance of a cloud follows from the breakdown of the cloud into L one–dimension orthogonal principal clouds:

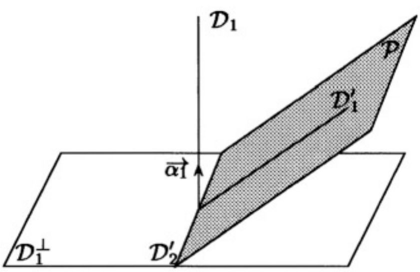
$$\forall j \in J : \mathrm{M}^j = \mathrm{G} + \sum_{\ell=1}^{L} \overrightarrow{\mathrm{GA}}{}^j_\ell \qquad (\ell \neq \ell' \Rightarrow \forall j \in J \; : \; \overrightarrow{\mathrm{GA}}{}^j_\ell \perp \overrightarrow{\mathrm{GA}}{}^j_{\ell'})$$

Theorem 3.4. *A first L'–dimensional principal direction is spanned by L' first principal linearly independent vectors $(1 \leq L' \leq L)$.*

Property 3.11 (Heredity property). *If $\lambda_{L'} > \lambda_{L'-1}$, a first principal L'–dimensional direction contains all the first principal L''–dimensional directions with $L'' < L'$.*

Thus, a first principal plane contains the first principal line, and so on. The theorem can be proved recursively, by showing that the L–dimensional subspace contains the $L-1$ one, then that it contains the principal line $\mathcal{D}_L$. Hereafter we sketch the proof by showing that the first principal plane is spanned by the first two principal lines (assuming $\lambda_1 > \lambda_2 > \lambda_3$).

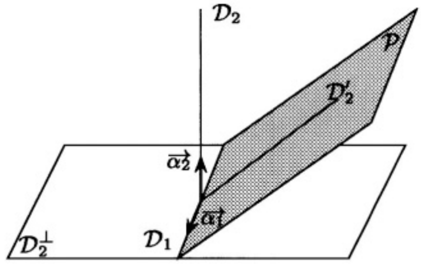
Proof. Denoting $\mathcal{P}$ the first principal plane and $\mathcal{D}_1$ the first principal line, suppose that $\mathcal{D}_1 \not\subset \mathcal{P}$ (proof by contradiction).
Let $\mathcal{D}'_2 = \mathcal{P} \cap \mathcal{D}_1^\perp$ be the intersection of the first principal plane $\mathcal{P}$ with the hyperplane orthogonal to $\mathcal{D}_1$, and $\mathcal{D}'_1$ be the line of $\mathcal{P}$ orthogonal to $\mathcal{D}'_2$ ($\mathcal{D}'_1 \in \mathcal{P}$ and $\mathcal{D}'_1 \perp \mathcal{D}'_2$). If A'^J_1 and A'^J_2 are respectively the projected clouds onto $\mathcal{D}'_1$ and $\mathcal{D}'_2$, the variance of the projected cloud onto $\mathcal{P}$ is equal to $\mathrm{Var}\,\mathrm{A}'^J_1 + \mathrm{Var}\,\mathrm{A}'^J_2$.



dim $\mathcal{U} = 3$

Since $\mathcal{D}_1$ is the first principal line, one has $\operatorname{Var} \mathrm{A'}_1^J \leq \operatorname{Var} \mathrm{A}_1^J = \lambda_1$, hence the cloud projected onto the plane spanned by $(\mathcal{D}_1, \mathcal{D}'_2)$ has a greater variance than the cloud projected onto the plane $\mathcal{P}$, which contradicts the hypothesis. Therefore the plane $\mathcal{P}$ contains the line $\mathcal{D}_1$.

Suppose now that $\mathcal{D}_2 \not\subset \mathcal{P}$. The intersection of the hyperplane orthogonal to $\mathcal{D}_2$ with $\mathcal{P}$ is $\mathcal{D}_1$; let $\mathcal{D}'_2$ denote the line of $\mathcal{P}$ orthogonal to $\mathcal{D}_1$. If A_1^J and $\mathrm{A}_2'^J$ are respectively the projected clouds onto $\mathcal{D}_1$ and $\mathcal{D}'_2$, the variance of the projected cloud onto the plane $\mathcal{P}$ is equal to $\operatorname{Var} \mathrm{A}_1^J + \operatorname{Var} \mathrm{A'}_2^J = \lambda_1 + \operatorname{Var} \mathrm{A'}_2^J$.



$\dim \mathcal{U} = 3$

Since $\mathcal{D}'_2 \perp \mathcal{D}_1$, $\operatorname{Var} \mathrm{A'}_2^J \leq \lambda_2$, the variance of the projected cloud onto the plane $(\mathcal{D}_1, \mathcal{D}_2)$ is equal to $\lambda_1 + \lambda_2$; hence the variance of the projected cloud onto the plane $\mathcal{P}$ is lower than $\lambda_1 + \lambda_2$: the plane $\mathcal{P}$ contains $\mathcal{D}_2$. ◁

Multiple eigenvalues. The approach remains valid when some eigenvalues are multiple. With each multiple eigenvalue there is associated a principal subspace for which each direction is principal, and within which it is always possible to choose orthonormal directions.

Null eigenvalues. In the supplementary orthogonal subspace of the support of the cloud every direction is a principal one associated with a null eigenvalue; the corresponding principal clouds are concentrated in the point G and the associated principal variables are identically null.

3.3.2 Principal Variables

Definition 3.15 (Principal coordinates). *The coordinate* y_ℓ^j *on Axis* ℓ *in frame* $(\mathrm{G}, \overrightarrow{\alpha_\ell}/\|\overrightarrow{\alpha_\ell}\|)$ *of the projected point* A_ℓ^j *of the point* M^j *onto Axis* ℓ *is called* ℓ*–th principal coordinate of the point* M^j.

$$y_\ell^j = \langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha_\ell} \rangle / \|\overrightarrow{\alpha_\ell}\|$$

Property 3.12 (Principal breakdown of distances).

$$(\mathrm{GM}^j)^2 = \sum_{\ell=1}^{L} (y_\ell^j)^2 \qquad (\mathrm{M}^j\mathrm{M}^{j'})^2 = \sum_{\ell=1}^{L} (y_\ell^j - y_\ell^{j'})^2$$

Proof. $\overrightarrow{\mathrm{GM}}^j = \sum_{\ell=1}^{L} \overrightarrow{\mathrm{GA}}_\ell^j = \sum_{\ell=1}^{L} y_\ell^j \, \overrightarrow{\alpha_\ell}/\|\overrightarrow{\alpha_\ell}\|$ where $(\overrightarrow{\alpha_\ell})_{\ell=1,\dots L}$ is an orthogonal basis of the support of the cloud, hence the property. ◁

With the one–dimensional principal cloud referred to $(\mathrm{G}, \overrightarrow{\alpha_\ell})$ are attached the following two (centered and proportional) *principal variables*.
— The first *calibrated principal variable* y_1^J, in brief *principal variable*, is defined by $y_1^j = \langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha_1} \rangle / \| \overrightarrow{\alpha_1} \|$, i.e. the coordinate of point A_1^j in frame $(\mathrm{G}, \overrightarrow{\alpha_1}/\|\overrightarrow{\alpha_1}\|)$; one has $\mathrm{Var}\, y^J = \lambda_1$.
— The first *standard principal variable* z_1^J is defined by $z_1^j = y_1^j/\xi_1$, i.e. the coordinate of point A_1^j in frame $(\mathrm{G}, \xi_1 \overrightarrow{\alpha_1}/\|\overrightarrow{\alpha_1}\|)$; one has $\mathrm{Var}\, z^J = 1$.

Property 3.13 (Passage formulas). *Between the principal vector $\overrightarrow{\alpha_\ell}$ and the standard principal variable z_ℓ^J, one has:*

$$\begin{cases} Vac(\overrightarrow{\alpha_\ell}/\|\overrightarrow{\alpha_\ell}\|) & = \xi_\ell\, z_\ell^J \\ Vac^*(z_\ell^J) & = \xi_\ell\, \overrightarrow{\alpha_\ell}/\|\overrightarrow{\alpha_\ell}\| \end{cases} \tag{3.7}$$

The passage formulas allow one to go from a principal vector to a principal variable and vice versa (*MATH.BASES*, Theorem 10.6, p.445).

Theorem 3.5. *The principal variables are eigenvectors of the endomorphism* Tom.

$$\forall \ell = 1, \dots L \quad Tom(z_\ell^J) = \lambda_\ell\, z_\ell^J \tag{3.8}$$

Proof. From passage formulas, it follows that $Tom(z_\ell^J) = Vac \circ Vac^*(z_\ell^J) = \xi_\ell\, Vac(\overrightarrow{\alpha_\ell})/\|\overrightarrow{\alpha_\ell}\| = \lambda_\ell z_\ell^J$. ◁

Property 3.14. *The ℓ–th and ℓ'–th principal variables $(\xi_\ell \neq \xi_{\ell'})$ are uncorrelated.*

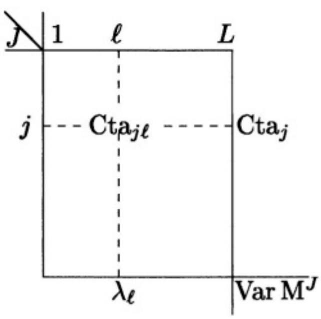
This property follows from the symmetry of the endomorphism Tom.

3.3.3 Contributions

Definition 3.16 (Absolute contribution of point and axis). *The absolute contribution of point* M^j *and Axis* ℓ*, denoted* $\mathrm{Cta}_{j\ell}$*, is the absolute contribution of* A_ℓ^j *to the variance of the projected cloud* A_ℓ^J *on Axis* ℓ*:*

$$\mathrm{Cta}_{j\ell} = f_j(\mathrm{GA}_\ell^j)^2 = f_j(y_\ell^j)^2$$

Since $\overrightarrow{\mathrm{GA}}_\ell^j = y_\ell^j \overrightarrow{\alpha_\ell}$, one has $\mathrm{Cta}_{j\ell} = f_j(y_\ell^j)^2$, $\sum\limits_{j \in J} \mathrm{Cta}_{j\ell} = \lambda_\ell$ (variance of Axis ℓ), and $\sum\limits_{\ell=1}^{L} \mathrm{Cta}_{j\ell} = f_j \sum\limits_{\ell=1}^{L} (\mathrm{GA}_\ell^j)^2 = f_j(\mathrm{GM}^j)^2 = \mathrm{Cta}_j$ (absolute contribution of j to the variance of the cloud, cf. definition 3.5 p.78).



The *double breakdown of the variance of the cloud according to points and axes* follows:

$$\operatorname{Var} M^J = \sum_{j\in J}\sum_{\ell=1}^{L} \mathrm{Cta}_{j\ell}$$

Absolute contributions add up; the lower positions of j and ℓ are meaningful in terms of duality notation. From absolute contributions, the following two relative contributions are defined.

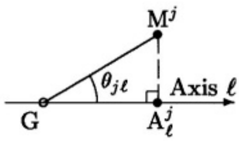
Definition 3.17 (Relative contribution of point to axis).
The relative contribution of the point M^j *to Axis* ℓ*, denoted* Ctr_j^ℓ*, is the proportion of the variance of Axis* ℓ *accounted for by the point* M^j*:*

$$\mathrm{Ctr}_j^\ell = \mathrm{Cta}_{j\ell}/\lambda_\ell = f_j(y_\ell^j)^2/\lambda_\ell$$

More generally, the relative contribution of the point M^j to the L'–principal direction is defined by $\sum_{\ell=1}^{L'} \mathrm{Cta}_{j\ell}/(\sum_{\ell=1}^{L'}\lambda_\ell)$, which is the weighted mean (the weights being $(\lambda_\ell)_{\ell=1,\dots L'}$) of the relative contributions of M^j to axes 1 to L'. For $L'=L$, $\mathrm{Ctr}_j^L = f_j(\mathrm{GM}^j)^2/\operatorname{Var} M^J$ is the relative contribution of the point M^j to the variance of the cloud (denoted Ctr_j, cf. p.78).

Definition 3.18 (Relative contribution of axis to point).
The relative contribution of Axis ℓ *to the point* M^j*, or quality of representation of* M^j *on Axis* ℓ*, denoted* Ctr_ℓ^j*, is equal to* $\mathrm{Cta}_{j\ell}/\mathrm{Cta}_j$.

$\mathrm{Ctr}_\ell^j = (y_\ell^j)^2/(\mathrm{GM}^j)^2 = (\mathrm{GA}_\ell^j)^2/(\mathrm{GM}^j)^2$ is equal to the square of the cosine of the (acute) angle $\theta_{j\ell}$ between $\overrightarrow{\mathrm{GM}}^j$ with its projection onto Axis ℓ.



This contribution can be interpreted as the *quality of representation* of the point M^j on Axis ℓ: The closer to 1 its value is, the better the point is represented on axis. The quality of representation of the point M^j on the L'–principal space is the sum of the qualities of representation of this point on axes 1 through L': $\sum_{\ell=1}^{L'} \mathrm{Ctr}_\ell^j$. For $L'=L$ (complete reconstitution of the cloud), the quality of representation is equal to 1.

The duality notation applies here like for the conditional frequencies of a contingency table (§2.2.1, p.32): ℓ is an upper index in $\mathrm{Ctr}_j^\ell = \mathrm{Cta}_{j\ell}/\lambda_\ell$, and j is an upper index in Ctr_ℓ^j.

3.3.4 Specific Analysis of a Cloud

The *specific analysis* of a cloud consists in determining its principal axes under the constraint that they must belong to a given prespecified subspace.

Let a subspace $\mathcal{A} \subset \mathcal{U}$ going through the point G. Let us denote A^j the orthogonal projection of the point M^j onto $\mathcal{A}$ and B^j its projection onto $\mathcal{A}^{\perp}$, orthogonal supplementary subspace in $\mathcal{A}$ going through G. One has:

$$\forall j \in J \ : \ \overrightarrow{\mathrm{GM}}^j = \overrightarrow{\mathrm{GA}}^j + \overrightarrow{\mathrm{GB}}^j \quad \text{with } \overrightarrow{\mathrm{GA}}^j \perp \overrightarrow{\mathrm{GB}}^j$$

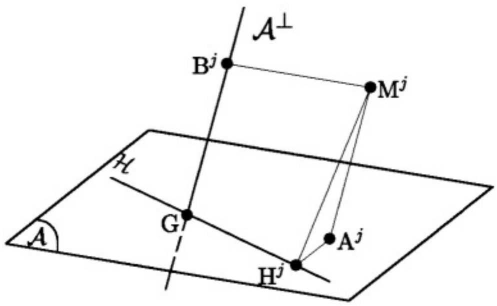


Figure 3.5. Specific analysis ($\dim\mathcal{U} = 3$, $\dim\mathcal{A} = 2$)

Let $\mathcal{H}$ be a subspace of $\mathcal{A}$. the point H^j, the orthogonal projection of M^j onto $\mathcal{H}$, is also the orthogonal projection of A^j onto $\mathcal{H}$, hence the relation: $(\mathrm{M}^j\mathrm{H}^j)^2 = (\mathrm{M}^j\mathrm{A}^j)^2 + (\mathrm{A}^j\mathrm{H}^j)^2$, and

$$\sum_{j\in J} f_j d^2(\mathrm{M}^j, \mathcal{H}) = \sum_{j\in J} f_j d^2(\mathrm{M}^j, \mathcal{A}) + \sum_{j\in J} f_j d^2(\mathrm{A}^j, \mathcal{H})$$

Therefore, the subspace $\mathcal{H}$ of $\mathcal{A}$ such that the residual square mean of the cloud M^J is minimum is the first principal subspace associated with the cloud A^J; the specific coordinates of the point M^j are therefore equal to the principal coordinates of the point A^j.

The projection of the bipoint $(\mathrm{G},\mathrm{M}^j)$ is $(\mathrm{G},\mathrm{A}^j)$, hence $\mathrm{GA}^j \leq \mathrm{GM}^j$: The variance of the specific cloud A^J is always less than or equal to that of cloud M^J. More precisely, $(\lambda'_\ell)_{\ell=1,\dots L'}$ denoting the eigenvalues of the specific cloud $(L' \leq L)$, then one has the property:

$$\sum_{\ell=1}^{L''} \lambda'_\ell \leq \sum_{\ell=1}^{L''} \lambda_\ell \quad (1 \leq L'' \leq L')$$

As will be seen in next chapters, subspaces involve in a specific analysis are usually supports of subcloud or derived cloud of methodological interest: e.g. the between–cloud associated with a partition, or, in MCA, a subcloud determined by some modalities of interest (*MULTIPLE,* §5.2).

3.4 Principal or Inertia Hyperellipsoids

This section is devoted to geometric characterizations that are intrinsically related to the principal directions of a cloud.

In the Euclidean space $\mathcal{U}$, we consider the geometric support of the cloud $\mathcal{M} \subseteq \mathcal{U}$ (see p.76), with the underlying vector subspace $\mathcal{L} \subseteq \mathcal{V}$ with its two orthogonal principal bases: the orthonormal one $(\overrightarrow{\alpha_\ell})_{\ell=1,\ldots L}$, and the "orthocalibrated" one $(\overrightarrow{\beta_\ell})_{\ell=1,\ldots L}$ with $\overrightarrow{\beta_\ell} = \xi_\ell \overrightarrow{\alpha_\ell}$.

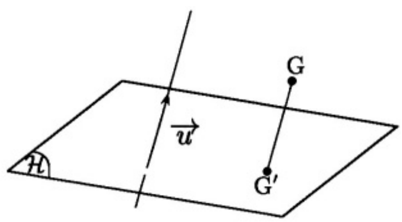
In the space of variables, we consider the principal subspace $\mathcal{X}_0 = Vac(\mathcal{L}) \subseteq \mathbb{R}_0^J$ (subspace of centered variables over J), also with its two orthogonal principal bases: the orthonormal one $(z_\ell^J)_{\ell=1,\ldots L}$, and the orthocalibrated one $(y_\ell^J)_{\ell=1,\ldots L}$.

Then we consider the bijections $\mathcal{L} \to \mathcal{X}_0$, $\mathcal{X}_0 \to \mathcal{L}$, and $\mathcal{L} \to \mathcal{L}$ respectively associated with the homomorphisms Vac, Vac^* and Som, denoted with the same symbols and defined by

$$Vac\,(\overrightarrow{\alpha_\ell}) = y_\ell^J = \xi_\ell\, z_\ell^J \text{ and } Vac^*\,(z_\ell^J) = \overrightarrow{\beta_\ell} = \xi_\ell\, \overrightarrow{\alpha_\ell} \quad (\ell = 1, \ldots L)$$
$$Som\,(\overrightarrow{\alpha_\ell}) = Vac^* \circ Vac\,(\overrightarrow{\alpha_\ell}) = \lambda_\ell\, \overrightarrow{\alpha_\ell} \quad (\ell = 1, \ldots L)$$

3.4.1 Definitions and Properties

Let $\mathcal{H} \subset \mathcal{M}$ be the hyperplane of equation (in the orthonormal basis) $\sum_{\ell=1}^{L} u_\ell y_\ell = \pm h$, with $\sum_{\ell=1}^{L} (u_\ell)^2 = 1$; the unit–norm vector $\overrightarrow{u} = \sum_{\ell=1}^{L} u_\ell \overrightarrow{\alpha_\ell}$ is a direction vector of the normal (perpendicular line) to $\mathcal{H}$, and h (> 0) is the distance of the mean point G of the cloud to hyperplane $\mathcal{H}$ (*MATH.BASES*, p.440).



$\dim \mathcal{M} = 3$, $d(\mathrm{G}, \mathcal{H}) = \mathrm{GG}' = h$

Let σ be the standard deviation (square root of variance) of the cloud along the direction $\overrightarrow{u}$ with $\sigma^2 = \| Vac(\overrightarrow{u}) \|^2 = \sum_{\ell=1}^{L} \lambda_\ell u_\ell^2$; and let $\kappa = h/\sigma$.

Definition 3.19. *The hyperellipsoid $\mathcal{E}_\kappa$ with principal axes of half–lengths $(\kappa\,\xi_\ell)_{\ell=1,\ldots L}$ is called the* **κ–principal,** *or* **κ–inertia** *hyperellipsoid of the cloud. For $\kappa = 1$, it is called the indicator hyperellipsoid.*

Relative to the orthonormal frame $(\mathrm{G}, (\overrightarrow{\alpha_\ell})_{\ell=1,\ldots L})$ of $\mathcal{M}$, the hyperellipsoid $\mathcal{E}_\kappa$ has Cartesian equation $\sum_{\ell=1}^{L} y_\ell^2/\lambda_\ell = \kappa^2$, and tangential equation $\sum_{\ell=1}^{L} \lambda_\ell u_\ell^2 = h^2/\kappa^2$. In any orthonormal basis $(\overrightarrow{\delta_k})_{k\in K}$ of $\mathcal{U}$, the hyperellipsoid $\mathcal{E}_\kappa$ has tangential equation $\sum_{k\in K}\sum_{k'\in K} v^{kk'} u_k u_{k'} = h^2/\kappa^2$, where $v^{kk'} = \langle Som(\overrightarrow{\delta^k})|\overrightarrow{\delta^{k'}}\rangle$ denotes the covariance of the covariant variables associated with $\overrightarrow{\delta_k}$ and $\overrightarrow{\delta_{k'}}$ (see p.87), and $(u_k)_{k\in K}$ the coordinates of the unit–norm vector $\overrightarrow{u}$ (*MATH.BASES*, p.441)[2].

For the matrix formulations of those equations, see p.121.

Property 3.15. *The principal κ–**hyperellipsoid** is the envelope of the hyperplanes that are at the distance $\kappa\,\sigma$ of the mean point of the cloud.*

Proof. $h^2 = \kappa^2\sigma^2 \iff h^2 = \kappa^2 \sum_{\ell=1}^{L} \lambda_\ell u_\ell^2$, that is, $\sum_{\ell=1}^{L} \lambda_\ell u_\ell^2 = h^2/\kappa^2$ (tangential equation of **κ-principal** hyperellipsoid). ◁

Principal hyperellipsoids — also named *inertia hyperellipsoids* — are intrinsically attached to the cloud for which they provide a geometric summary. The distance between the two hyperplanes orthogonal to a given direction and tangent to the **κ-hyperellipsoid** is equal to 2κ times the standard deviation of the cloud along that direction. The contact points of the hyperplanes normal to a principal direction and tangent to the hyperellipsoid are the edges of the hyperellipsoid.

Property 3.16. *The indicator hyperellipsoid is the image by the mapping Vac* of the set of the standard variables (unit hypersphere) of $\mathcal{X}_0$.*

Proof. Let $x^J = \sum_{\ell=1}^{L} x_\ell z_\ell^J$ be a standard variable of $\mathcal{X}_0$ ($\sum_{\ell=1}^{L} x_\ell^2 = 1$). One has $Vac^*(x^J) = \sum_{\ell=1}^{L} x_\ell \xi_\ell \overrightarrow{\alpha}_\ell$. Letting $y_\ell = x_\ell \xi_\ell$, one has $\sum_{\ell=1}^{L} y_\ell^2/\lambda_\ell = 1$. When x^J goes through the hypersphere of unit radius of $\mathcal{X}_0$, its image $Vac^*(x^J)$ goes through the indicator hyperellipsoid. ◁

Property 3.17 (Projection of a hyperellipsoid). *If a cloud is orthogonally projected onto a subspace, the κ–**hyperellipsoid** of the projected cloud is the projection of the κ–**hyperellipsoid** of the cloud onto this subspace.*

---

[2] As a matter of fact, the tangential equation remains valid for a nonorthonormal basis, that is, κ is an affine index (independent from the initial metric on the whole space); of course, this invariance property does not hold if index κ is restricted to a strict principal subspace.

Proof. Let us refer the cloud to an orthonormal basis $(\overrightarrow{\delta_k})_{k\in K}$ such that $(\overrightarrow{\delta_k})_{k\in K'}$ is a basis of the subspace $\mathcal{L}'$ of $\mathcal{L}$ onto which the cloud is projected. The κ–hyperellipsoid $\mathcal{E}_\kappa$ has tangential equation $\sum\limits_{k\in K}\sum\limits_{k'\in K} u_k u_{k'} v^{kk'} = \frac{h^2}{\kappa^2}$. The cloud projected onto $\mathcal{L}'$ has variances and covariances $(v^{kk'})_{k\in K',k'\in K'}$, that is, the restriction to K' of the variances and covariances of the initial cloud; the κ–hyperellipsoid $\mathcal{E}'_\kappa$ of the projected cloud has therefore the equation in $\mathcal{L}'$: $\sum\limits_{k\in K'}\sum\limits_{k'\in K'} u_k u_{k'} v^{kk'} = h^2/\kappa^2$. On the other hand, the projection of $\mathcal{E}_\kappa$ onto $\mathcal{L}'$ is defined by the tangential equation of $\mathcal{E}_\kappa$ restricted to the hyperplanes that are orthogonal to $\mathcal{L}'$, which therefore verify $\forall k \in K - K' : u_k = 0$; which indeed comes back to $\mathcal{E}'_\kappa$. ◁

3.4.2 Hypersphere of the Contrast Space

Let $\mathcal{X}^0 = (f_J \circ Vac)(\mathcal{M})$ be the subspace of $\mathbb{R}_J$ spanned by the principal contrasts $(z_{J\ell})_{\ell=1,\ldots L}$, with $z_{J\ell} = (f_j\, z^j_\ell)_{j\in J}$. On the one hand, the contrast $c^j_J = \sum\limits_{\ell=1}^{L} z^j_\ell z_{J\ell}$ is the projection of the measure δ^j_J onto $\mathcal{X}^0$; on the other hand, the effect of the contrast c^j_J on the cloud is $\overrightarrow{\mathrm{GM}}^j$. In $\mathcal{X}^0$, vector $\overrightarrow{\mathrm{GM}}^j$ will be represented by contrast c^j_J, and the cloud M^J will therefore be represented by a spherical cloud of unit variance in all directions. The image by the mapping $E\!f\!f^{-1} : \mathcal{L} \to \mathcal{X}^0$ of the κ–hyperellipsoid of equation $\sum\limits_{\ell=1}^{L} (y^j_\ell)^2/\lambda_\ell = \kappa^2$ is the κ–hypersphere of equation $\sum\limits_{\ell=1}^{L} (z^j_\ell)^2 = \kappa^2$.

3.4.3 Plane Cloud

Relative to the principal orthonormal basis $(\overrightarrow{\alpha_1}, \overrightarrow{\alpha_2})$, the principal κ–ellipse has equation:

$$y_1^2/\lambda_1 + y_2^2/\lambda_2 = \kappa^2 \tag{3.9}$$

Relative to any orthonormal basis $(\overrightarrow{\varepsilon}, \overrightarrow{\varepsilon}')$, where $v = \langle Som(\overrightarrow{\varepsilon})|\overrightarrow{\varepsilon}\rangle$, $v' = \langle Som(\overrightarrow{\varepsilon}')|\overrightarrow{\varepsilon}'\rangle$ and $c = \langle Som(\overrightarrow{\varepsilon})|\overrightarrow{\varepsilon}'\rangle$ (variances and covariance in the directions $\overrightarrow{\varepsilon}$, $\overrightarrow{\varepsilon}'$) and $r = c/\sqrt{v\,v'}$, the principal κ–ellipse has equation:

$$\frac{1}{1-r^2}\left(\frac{x^2}{v} - 2c\frac{xx'}{vv'} + \frac{x'^2}{v'}\right) = \kappa^2 \text{ or } \frac{1}{vv'-c^2}(v'x^2 - 2cxx' + vx'^2) = \kappa^2$$

For $-\pi \leq \varphi \leq +\pi$, let us consider the standard variable $t^J(\varphi)$ defined by $t^J(\varphi) = z^J_1 \cos\varphi + z^J_2 \sin\varphi$. When φ varies, the end of the vector $t^J(\varphi)$ goes through a unit circle.

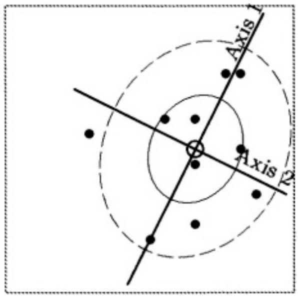
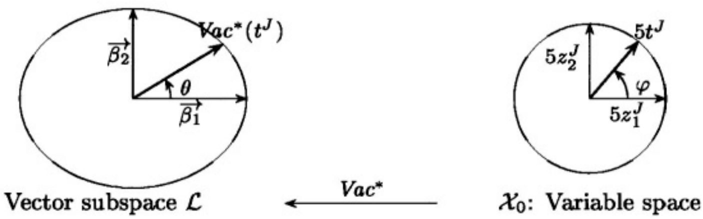


Figure 3.6. *Target example*: $(v = 56,\ v' = 36,\ c = 8)$. Indicator ellipse $(\kappa = 1$, solid line) and concentration ellipse $(\kappa = 2$, dashed line).

One has: $Vac^*(t^J(\varphi)) = \cos\varphi\,\overrightarrow{\beta_1} + \sin\varphi\,\overrightarrow{\beta_2} = \cos\theta\,\overrightarrow{\alpha_1} + \sin\theta\,\overrightarrow{\alpha_2}$, with $\tan\theta = (\xi_2/\xi_1)\tan\varphi$; the indicator ellipse is the image of the unit radius circle of the variable space by the mapping *Vac** (see figure hereafter).



The standard deviation σ in direction $\overrightarrow{u}$ is equal to the length of segment GA (see figure 3.7), where A is the point where line $(\mathrm{G}, \overrightarrow{u})$ intersects one of the tangents to the indicator ellipse perpendicular to this line. When $\overrightarrow{u}$ varies, the point A goes through a quartic curve (the "pedal curve" of ellipse); this curve has polar equation (in frame $(\mathrm{G}, \overrightarrow{\alpha_1}))$ $\rho^2(\theta) = \lambda_1 \cos^2\theta + \lambda_2 \sin^2\theta$, and Cartesian equation (in frame $(\mathrm{G}, \overrightarrow{\alpha_1}, \overrightarrow{\alpha_2}))$ $(y_1^2 + y_2^2)^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2$. When the values ξ_1 and ξ_2 are clearly separated (which is not the case for the *Target example*), the pedal curve stands out well distinct from the ellipse (see figure 3.7 with $\xi_2/\xi_1 = 2/5$).

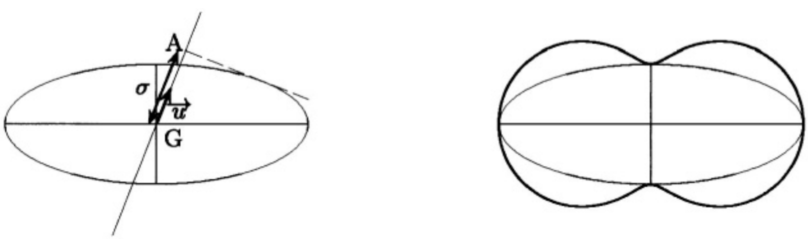


Figure 3.7. Standard deviation in a direction and pedal curve.

In any (principal or not) orthonormal basis $(O, \vec{\varepsilon}, \vec{\varepsilon}')$, for any principal ellipse, the conjugate axis of $(O, \vec{\varepsilon}')$ (ordinate axis x') is the *regression line* of x' on x, similarly the conjugate axis of $(O, \vec{\varepsilon})$ (abscissa axis x) is the regression line of x on x' (for conjugate axes, cf. *MATH.BASES*, Definition 10.22, p. 441). The angles of those regression lines with the abscissa axis are θ and θ' defined by $\tan\theta = c/v$ and $\tan\theta' = v'/c$ respectively. The angle θ_1 between the first principal axis and the abscissa axis is such that $\tan\theta_1 = (v' - v + \sqrt{(v-v')^2 + 4c^2})/2c$. The first principal axis lies between the two regression lines: When $c > 0$, one has $\theta \leq \theta_1 \leq \theta'$, that is, $c/v \leq (v' - v + \sqrt{(v-v')^2 + 4c^2})/2c \leq v'/c$, a relation which follows from $c^2 \leq vv'$ i.e. $|r| \leq 1$.

When $v = v'$, the first principal axis is the bissector of axes, and the two regression lines are symmetrical with respect to this bissector.

Target example (Figure 3.8). One has: $\tan\theta' = 6.5$ ($\theta' = 81°3$), $\tan\theta_1 = 2$ ($\theta_1 = 63°4$), $\tan\theta = 0.20$ ($\theta = 11°3$).

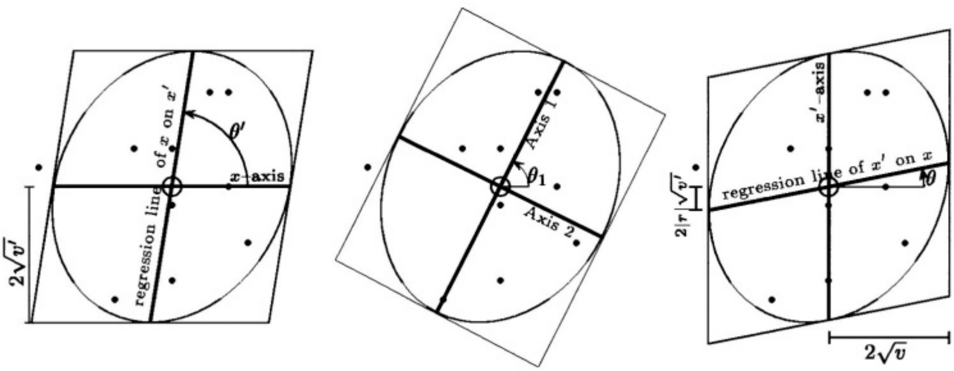


Figure 3.8. Target Example: Concentration ellipse and regression lines.

3.4.4 Concentration Hyperellipsoids

The concept of principal hyperellipsoids applies to *continuous distributions*. For the multidimensional normal and related distributions (such as the **t–distribution**), principal hyperellipsoids are equal density contours, and the density for every point inside a hyperellipsoid is higher than the density for every outside point. Given a cloud together with its **κ–hyperellipsoid**, one may consider a distribution of uniform density inside the hyperellipsoid; it can be shown that this distribution has the same principal axes as the cloud and that its eigenvalues are equal to $\left(\lambda_\ell \kappa^2/(L+2)\right)_{\ell=1,\dots L}$. Consequently, a uniform distribution on the κ–hyperellipsoid with $\kappa^2 = L+2$ has the same eigenvalues $(\lambda_\ell)_{\ell=1,\ldots L}$ as the cloud; this hyperellipsoid is called the *concentration hyperellipsoid* of the cloud (Cramér, 1946, p.283; Anderson, 1958, p.44; Malinvaud, 1980).

For the normal L–dimensional distribution, the proportion of the distribution inside the κ–hyperellipsoid is given by $P(\chi^2_L \leq \kappa^2)$, where χ^2_L denotes the classical variable χ^2 with L degrees of freedom. For a given value of κ, the higher the value of L, the smaller the corresponding proportion. For $L = 1$, the interval of half–length σ $(\kappa = 1)$ ("indicator interval") contains 68.27% of distribution; that of half–length 2σ ("concentration interval") contains 95.45% of distribution.

For $L = 2$, the *indicator ellipse* $(\kappa = 1)$ contains 39.35% of distribution; the *concentration ellipse* $(\kappa = 2)$ contains 86.47% of distribution.

Magnitude of deviation. For the point M^j (with principal coordinates $(y^j_\ell)_{\ell=1,\ldots L}$), the coefficient $\kappa = \left(\sum\limits_{\ell=1}^{L} (y^j_\ell)^2/\lambda_\ell\right)^{1/2}$ is an index of the *magnitude of the deviation* $\overrightarrow{\mathrm{GM}^j} = \mathrm{M}^j - \mathrm{G}$ that takes into account the direction of deviation. The κ–norm of $\overrightarrow{\mathrm{GM}^j}$, denoted $|\overrightarrow{\mathrm{GM}^j}|_\kappa$, is the multidimensional extension of standard score for numerical variables.

When the fit of the cloud by a distribution close to normal is acceptable, the κ–norm can be used as an *extremality index* of point M^j among the points of the cloud: the higher κ is, the more extreme the point is[3].

Target example. The two points M^{j2} and M^{j3} are both at the same geometric distance 10 from the mean point. Now for M^{j2}, one has $\kappa = 1.41$ and for M^{j3}, one has $\kappa = 1.66$; therefore the point M^{j3} (which is closer to Axis 2) is less extreme than M^{j2} (which is closer to Axis 1).

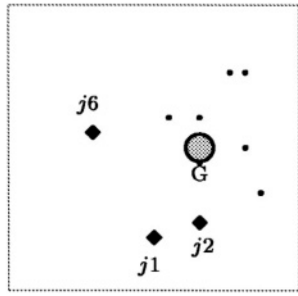
3.5 Between and Within Clouds

3.5.1 Subcloud

With a subset $J\langle c \rangle$ of J, there is associated the *subcloud* of points M^j such that $j \in J\langle c \rangle$. This subcloud will be denoted $\mathrm{m}^{J\langle c \rangle}$; its support $J\langle c \rangle$ is weighted by the restriction to $J\langle c \rangle$ of the absolute frequency–measure n_J, and its weight is $n_c = \sum\limits_{j \in J\langle c \rangle} n_j$.

[3]The κ–index, applied to mean points of subclouds obtained by sampling, will be fundamental in statistical inference (chapter INDUCTIVE, §8.5).

Target example (opposite picture). Subcloud of three points $\{j1, j2, j6\}$ (diamonds).



The *mean point of subcloud* is such that: $\mathrm{M}^c = \sum\limits_{j \in J<c>} n_j \mathrm{M}^j / n_c$ (Definition 3.2, p.77), that is, letting $f_j^c = n_j/n_c$ ($j \in J<c>$) (duality notation), $\mathrm{M}^c = \sum\limits_{j \in J<c>} f_j^c \mathrm{M}^j$. This point, called *center of class* c, is weighted by n_c.

The *variance of subcloud* is equal to $\sum\limits_{j \in J<c>} f_j^c (\mathrm{M}^c\mathrm{M}^j)^2$ (Definition 3.4, p.78).

We will study statistics relevant to a subcloud by taking the cloud M^J as *reference cloud* and consequently its mean point G as *reference point* and its weight n as *reference weight.* In particular, the absolute contribution of M^c to the variance of M^J is equal to $\mathrm{Cta}_c = f_c(\mathrm{GM}^c)^2$, with $f_c = n_c/n$.

Definition 3.20. *The absolute contribution of the subcloud* $\mathrm{M}^{J<c>}$, *denoted* $\mathrm{Cta}_{J<c>}$, *is equal to the sum of the absolute contributions of its points:*

$$\mathrm{Cta}_{J<c>} = \sum_{j \in J<c>} \mathrm{Cta}_j = \sum_{j \in J<c>} f_j (\mathrm{GM}^j)^2$$

Definition 3.21. *The absolute within–contribution of the subcloud* $\mathrm{M}^{J<c>}$, *denoted* $\mathrm{Cta}_{J(c)}$, *is the* f_j*–weighted sum of the squares of distances of the points of the subcloud from its mean point.*

$$\mathrm{Cta}_{J(c)} = \sum_{j \in <c>} f_j (\mathrm{M}^c\mathrm{M}^j)^2$$

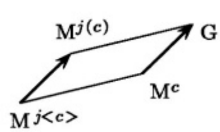
Within notation. In $J(c)$, "c" is between parentheses, not between chevrons ($<>$): $J(c)$ should be read "J within–c" (not "J in c").

Property 3.18. *The absolute within–contribution of a subcloud is equal to the product of its relative weight by its variance:*

$$\mathrm{Cta}_{J(c)} = f_c \operatorname{Var} \mathrm{M}^{J<c>}$$

Proof. $f_c \operatorname{Var} \mathrm{M}^{J<c>} = \sum\limits_{j \in J<c>} f_c f_j^c (\mathrm{M}^c\mathrm{M}^j)^2 = \sum\limits_{j \in J<c>} f_j (\mathrm{M}^c\mathrm{M}^j)^2 = \mathrm{Cta}_{J(c)}$. ◁

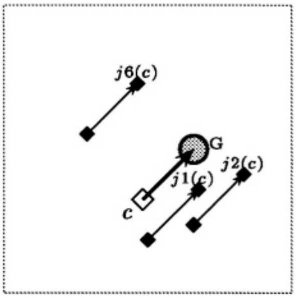
Definition 3.22. *The within–c cloud, denoted* $\mathrm{M}^{J(c)}$, *is obtained from the subcloud* $\mathrm{M}^{J<c>}$ *by the translation of vector* $\overrightarrow{\mathrm{M}^c\mathrm{G}}$.



$$\forall j \in J{<}c{>} : \; \mathrm{M}^{j(c)} = \mathrm{M}^{j{<}c{>}} + \overrightarrow{\mathrm{M}^c\mathrm{G}}$$

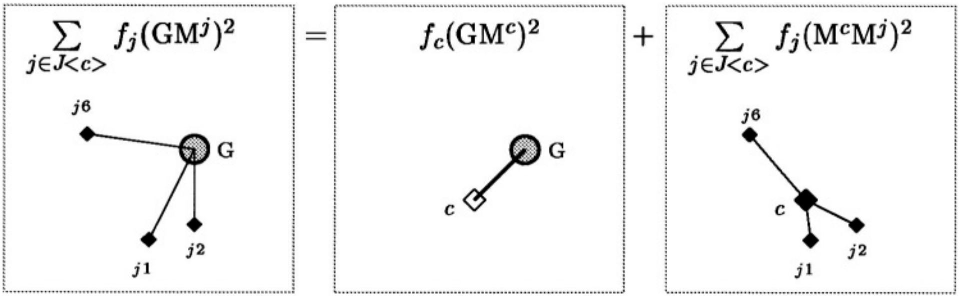
Point $\mathrm{M}^{j(c)}$ is weighted by n_j. By construction, the **within–c** cloud has mean point G, and its variance is the variance of subcloud $\mathrm{M}^{J<c>}$.

Target example. Construction of **within–c** cloud, see opposite picture.



Property 3.19. *The absolute contribution of a subcloud is the sum of its mean point contribution and its within–contribution:*

$$\mathrm{Cta}_{J<c>} = \mathrm{Cta}_c + \mathrm{Cta}_{J(c)}$$



This property is obtained by applying the first Huyghens' theorem (p.79) to the subcloud taking G as reference point.

Property 3.20. *The variance and the absolute within–contribution of a subcloud of two points* $\{\mathrm{M}^j, \mathrm{M}^{j'}\}$ *are respectively given by*

$$\mathrm{Var}_{\{j,j'\}} = \frac{f_j f_{j'}}{(f_j + f_{j'})^2}(\mathrm{M}^j\mathrm{M}^{j'})^2; \qquad \mathrm{Cta}_{(j,j')} = \frac{(\mathrm{M}^j\mathrm{M}^{j'})^2}{\frac{1}{f_j} + \frac{1}{f_{j'}}}$$

Definition 3.23 (Dipole). *Dipole jj' is the bipoint* $(\mathrm{M}^j, \mathrm{M}^{j'})$ *weighted by* $\widetilde{n}_{jj'} = 1/(\frac{1}{n_j} + \frac{1}{n_{j'}}) = n_j n_{j'}/(n_j + n_{j'})$.

Property 3.21 (Dipole contribution). *The absolute contribution of the dipole jj' is equal to the absolute within–contribution of the subcloud of the two points* $\{\mathrm{M}^j, \mathrm{M}^{j'}\}$*:*

$$\mathrm{Cta}_{jj'} = \frac{\widetilde{n}_{jj'}}{n}(\mathrm{M}^j\mathrm{M}^{j'})^2 = \frac{(\mathrm{M}^j\mathrm{M}^{j'})^2}{\frac{1}{f_j} + \frac{1}{f_{j'}}} = \mathrm{Cta}_{(j,j')}$$

With the subcloud $\mathrm{M}^{J<c>}$, there are associated the three endomorphisms on $\mathcal{V}$ defined as follows:

$$\forall \overrightarrow{\alpha} \in \mathcal{V} : Som_{J<c>}(\overrightarrow{\alpha}) = \sum_{j \in J<c>} f_j \langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha} \rangle \overrightarrow{\mathrm{GM}}^j \; (f_j = n_j/n) \quad (3.10)$$

$$Som_{J(c)}(\overrightarrow{\alpha}) = \sum_{j \in J(c)} f_j^c \langle \overrightarrow{\mathrm{M}^c\mathrm{M}^j} | \overrightarrow{\alpha} \rangle \overrightarrow{\mathrm{M}^c\mathrm{M}^j} \; (f_j^c = n_j/n_c) \quad (3.11)$$

$$Som_c(\overrightarrow{\alpha}) = f_c \langle \overrightarrow{\mathrm{GM}}^c | \overrightarrow{\alpha} \rangle \overrightarrow{\mathrm{GM}}^c \qquad (f_c = n_c/n) \quad (3.12)$$

Lemma 3.1. *Between these three endomorphisms, one has the relations:*

$$Som_{J<c>} = Som_c + f_c Som_{J(c)}$$

Proof. From relation 3.6 (p.86), applied to subcloud $\mathrm{M}^{J<c>}$, one has: $\sum_{j \in J<c>} f_j^c \langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha} \rangle \overrightarrow{\mathrm{GM}}^j = \langle \overrightarrow{\mathrm{GM}}^c | \overrightarrow{\alpha} \rangle \overrightarrow{\mathrm{GM}}^c + \sum_{j \in J<c>} f_j^c \langle \overrightarrow{\mathrm{M}^c\mathrm{M}}^j | \overrightarrow{\alpha} \rangle \overrightarrow{\mathrm{M}^c\mathrm{M}}^j$ $(\overrightarrow{\alpha} \in \mathcal{V})$ and thus, by multiplying f_c, the lemma. ◁

3.5.2 Partition of a Cloud, Between and Within Clouds

Let a partition of J into C classes (subsets of J) be denoted $J<C>$; hence the partition of M^J into subclouds (classes) $\mathrm{M}^{J<C>} = (\mathrm{M}^{J<c>})_{c \in C}$.

Property 3.22. *The variance of the cloud is the sum of the contributions of the classes (subclouds) of the partition:*

$$\mathrm{Var}\,\mathrm{M}^J = \sum_{c \in C} \mathrm{Cta}_{J<c>}$$

Definition 3.24 (Between–cloud and variance). *The between–cloud is the derived cloud of the mean points of subclouds $(\mathrm{M}^{J<c>})_{c \in C}$; its variance is called between–variance.*

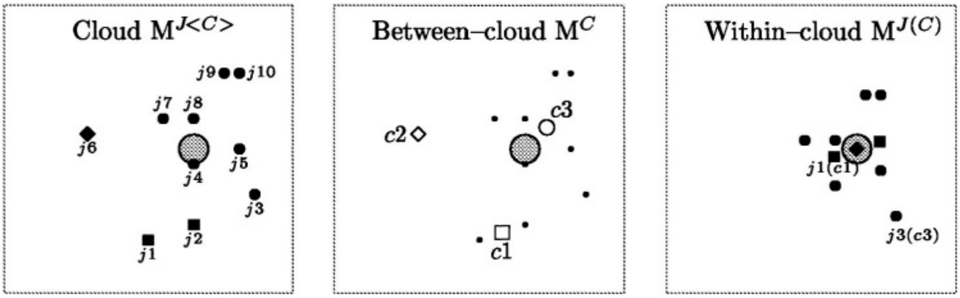
The between–cloud is denoted M^C. Its mean point is the mean point G of the total cloud and its variance is equal to $\sum_{c \in C} f_c (\mathrm{GM}^c)^2$.

Definition 3.25 (Within–cloud and variance). *The within–cloud, denoted $\mathrm{M}^{J(C)}$, is the union of the* **within–c** *clouds; its variance is called within–variance.*

Each point $\mathrm{M}^{j(c)}$ is weighted by n_j. The variance of the within–cloud is equal to the sum of the **within–c** contributions, or equivalently to the weighted mean of the variances of the subclouds $(\mathrm{M}^{J<c>})_{c \in C}$.

$$\mathrm{Var}\,\mathrm{M}^{J(C)} = \sum_{c \in C} \mathrm{Cta}\,\mathrm{M}^{J(c)} = \sum_{c \in C} f_c\, \mathrm{Var}\,\mathrm{M}^{J<c>} \qquad (3.13)$$

Target example. Consider the three–class partition: $J{<}c1{>} = \{j1, j2\}$ (■), $J{<}c2{>} = \{j6\}$ (◆) and $J{<}c3{>} = \{j3, j4, j5, j7, j8, j9, j10\}$ (•), hence the between–cloud M^C (□, ◇, ○) and the within–cloud $\mathrm{M}^{J(C)}$.



Property 3.23 (Between–within breakdown of variance). *The total variance (variance of the cloud) is the sum of the between and the within–variances:*

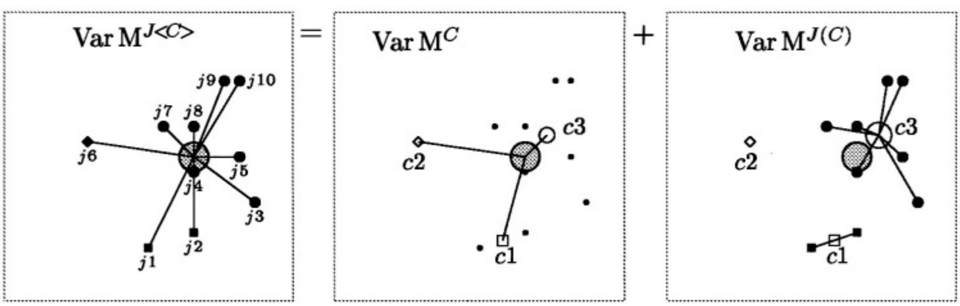
$$\mathrm{Var}\,\mathrm{M}^{J<C>} = \mathrm{Var}\,\mathrm{M}^C + \mathrm{Var}\,\mathrm{M}^{J(C)}$$

This property follows from Property 3.19 (p.102), by summing over all classes.

Definition 3.26. *The η^2-**coefficient** (or correlation ratio) of the partition is the ratio of the between–variance to the total variance.*

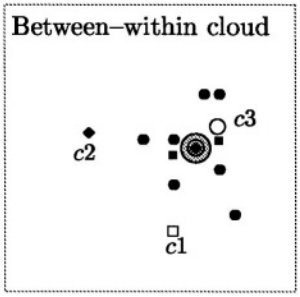
$$\eta^2 = \mathrm{Var}\,\mathrm{M}^C / \mathrm{Var}\,\mathrm{M}^J$$

Target example. One has: $\mathrm{Var}\,\mathrm{M}^J = 92$; $\mathrm{Var}\,\mathrm{M}^C = 57.429$; $\mathrm{Var}\,\mathrm{M}^{J(C)} = 34.571$, that is, 92 = 57.429 + 34.571, and $\eta^2 = 0.624$.



Definition 3.27 (Between–within cloud). *The between-within cloud is the union of the between and within–clouds.*

The between–within cloud has mean point G, its weight is equal to $2n$, therefore its variance is half of the variance of cloud M^J.



Theorem 3.6. *Let Som be the endomorphism of $\mathcal{V}$ for cloud M^J, Som_C the endomorphism for the between–cloud, and $Som_{J(C)}$ the endomorphism for the within–cloud:*

$$Som = Som_C + Som_{J(C)}$$

Corollary 3.1. *The between–within cloud has the same principal directions as the cloud M^J associated with eigenvalues $\lambda_\ell/2$ $(\ell = 1, \ldots L)$.*

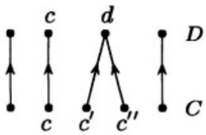
Property 3.24 (Two–class partition). *For a partition with two classes $\{c, c'\}$, the following relation holds:*

$$\mathrm{G} = \frac{n_c}{n}\mathrm{M}^c + \frac{n_{c'}}{n}\mathrm{M}^{c'}$$

One deduces the relations: $\overrightarrow{\mathrm{M}^c\mathrm{G}} = (n_{c'}/n)\overrightarrow{\mathrm{M}^c\mathrm{M}}^{c'} = (n_{c'}/n_c)\overrightarrow{\mathrm{GM}}^{c'}$.

Property 3.25 (Grouping property). *If two classes of a partition are grouped together, the between–variance decreases from an amount equal to the contribution of the dipole defined by the centers of the grouped classes.*

Proof. Let $C\langle d\rangle = \{c', c''\}$ be a pair of elements of C, and D the set obtained from C by grouping c' and c'' in a class d. From the between–within breakdown of variance (p.104), one deduces:



$\mathrm{Var}\,\mathrm{M}^D = \mathrm{Var}\,\mathrm{M}^C - \mathrm{Var}\,\mathrm{M}^{C(D)}$;
$\mathrm{Var}\,\mathrm{M}^{C(D)} = \sum_{\substack{c \in C \\ c \neq c';\, c \neq c''}} f_c \times 0 + (f_{c'} + f_{c''})\,\mathrm{Var}_{\{c', c''\}}$;

hence $\mathrm{Var}\,\mathrm{M}^{C(D)} = \mathrm{Cta}_{c'c''}$ which is the contribution of dipole $c'c''$, (cf. Property 3.21, p.102); hence $\mathrm{Var}\,\mathrm{M}^D = \mathrm{Var}\,\mathrm{M}^C - \mathrm{Cta}_{c'c''}$ ◁

The grouping property is fundamental for Euclidean classification.

Target example. For the three–class partition (p.104), if the two classes $c1$ and $c2$ are grouped, one has $(\mathrm{M}^{c1}\mathrm{M}^{c2})^2 = 290$, hence $\mathrm{Cta}_{(c1,c2)} = 290 \times (\frac{2}{10} \times \frac{1}{10})/(\frac{2}{10} + \frac{1}{10}) = 19.333$; the new partition in two classes $d1 = \{c1, c2\}$ and $d2 = \{c3\}$ has between–variance 38.095, that is, the between–variance of the three–class partition (57.429) minus the contribution of the dipole $(c1, c2)$ (19.333).

3.6 Euclidean Classification

Automatic Classification — alias *Cluster analysis* — is a world in itself. Classification will be tackled here only within the framework of GDA, where the objects to classify are points of Euclidean clouds.

The overall objective of any classification method is intuitive. What is sought is to construct clusters of objects — in GDA subclouds of points — so that the objects within a same cluster are as close together as possible whereas those belonging to different clusters are as remote from one another as possible — even though, in some situations these two demands may not be easy to reconcile (see e.g. Benzécri's counterexample, 1992, p.572). At an early stage of the classificatory process, one may be content to delineate *clusters,* i.e. classes of points corresponding to high density zones, with the twofold requirement of class *separability* and of class *compactness,* in other words, of heterogeneity between classes and homogeneity within classes. At a more elaborate stage, one can seek a set of classes that constitute a *partition* of the cloud, that is, such that every point belongs to one and only one class. At an even more elaborate stage, what is sought is a *hierarchical classification,* that is, a *system of nested classes,* after the pattern of natural science, where animals are divided into vertebrates, molluscs, etc., the vertebrates being in turn subdivided into mammals, birds, etc.

Classification methods developed less rapidly than GDA, mainly owing to technical obstacles (the "algorithmic wall"!) that were only gradually overcome. Yet, Volume 1 of Benzécri & coll (1973)was already entirely devoted to classification ("taxinomy"). As for AHC, one can also refer to part V of Benzécri's book (1992); to chapter 4 of Lebart & al (1984), chapter 2 of Lebart & al (1995), and to the book by Jambu & Lebeaux (1983), etc.

The present section will be mainly devoted to the *agglomerative (or ascending) hierarchical classification* (AHC) method with variance as aggregation index; we call this method *Euclidean classification,* as it is the method which has definitely emerged as the one most in harmony with the mathematical structures of GDA.

3.6.1 Hierarchy, Hierarchical Tree

Definition 3.28. *Let J be a nonempty finite set; a set $\mathcal{H}$ of classes (nonempty subsets of J) is called a hierarchy on J if the classes of $\mathcal{H}$ possess the two following properties:*

P1 *(intersection): for any pair of classes of $\mathcal{H}$: either the two classes are disjoint, or one is included in the other.*

P2 *(union): every class of $\mathcal{H}$ is the union of the classes included in this class.*

In brief, hierarchy classes are obtained by grouping and are represented by a *hierarchical tree.*

Figure 3.9 shows such a tree for the *Target example.* At the basis of the tree there are 10 one–element classes, indexed from $\boldsymbol{\ell 1}$ to $\boldsymbol{\ell 10}$; thus class $\boldsymbol{\ell 1}$ has $\boldsymbol{j1}$ as a single element, etc. Classes with several elements correspond to the *nodes* of the tree and are indexed from $\boldsymbol{\ell 11}$ to $\boldsymbol{\ell 19}$. The hierarchy is *dichotomous* (and represented by a binary tree), in the sense that every class indexed by a node is the union of two classes. For example, class $\boldsymbol{\ell 13}$ is the union of classes $\boldsymbol{\ell 1}$ and $\boldsymbol{\ell 2}$; class $\boldsymbol{\ell 18}$ is the union of classes $\boldsymbol{\ell 13}$ and $\boldsymbol{\ell 6}$, etc. A dichotomous hierarchy of J objects has $J-1$ nodes[4] (here 9 nodes numbered from $\boldsymbol{\ell 11}$ to $\boldsymbol{\ell 19}$). By "cutting" the tree at any level, one determines a partition of the set. For instance, cutting the tree between nodes $\boldsymbol{\ell 17}$ and $\boldsymbol{\ell 18}$ (Figure 3.9, p.107) generates the partition into three classes (cf. p.104) corresponding to the three nodes $\boldsymbol{\ell 6}$, $\boldsymbol{\ell 13}$ and $\boldsymbol{\ell 17}$.

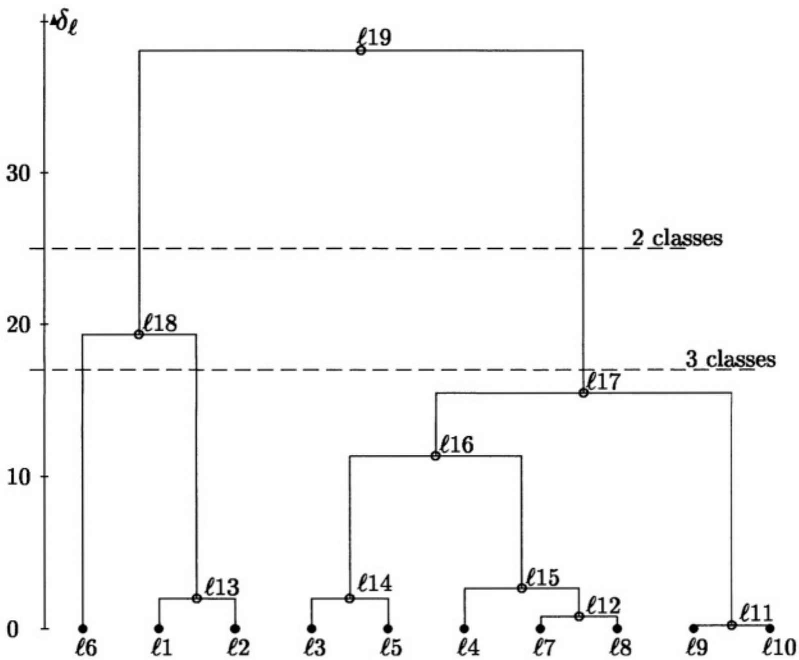


Figure 3.9. ***Target example*: Hierarchical tree.**

[4]Let us recall that we use capital letters to denote both finite sets and their cardinalities: e.g. the cardinality of the set J minus 1 is written $J-1$, etc.

3.6.2 Divisive vs Agglomerative Classification

In order to construct a hierarchy there are two main ways of : the divisive one and the agglomerative one.

In a *divisive classification* (or segmentation, or descending classification), one starts with the objects to be classified, regarded as a single class which is divided into classes; those classes are in turn subdivided, etc. A well–known example of divisive classification is the CART method (Classification And Regression Tree, Breiman & al., 1984).

In an *agglomerative (or ascending) classification,* one starts with the objects regarded as a set of one–element classes, from which one proceeds to successive aggregations, until all objects are grouped within a single class. At each step of the construction, one starts with a partition and groups two classes of this partition. In what follows, we will deal in detail with Agglomerative Hierarchical Classification (AHC).

3.6.3 Aggregation Indices

To render an AHC method operational, an aggregation index must be specified to select the classes to be aggregated at each step.

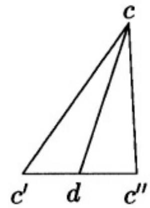
Given a set C of classes, an aggregation index $\delta(c, c')$ is an index defined on the two disjoint classes c and c' of C that can be interpreted either as a direct index of separability between classes c and c', or as an inverse index of compactness of the class grouping those classes. Several aggregation indices are commonly used in AGD. They are all defined from the geometric distance, and sometimes called "distance indices", even though most of them do not verify the triangle inequality: It may happen that two classes are (in the sense of the aggregation index) close to a same third one, while at the same time being distant from each other.

In any AHC method, one requirement must be satisfied, that is, the aggregation of two classes at some stage should not call into question previous aggregations. Specifically, suppose classes c' and c'' have already been aggregated into the new class d, which implies that $\forall c \in C - \{c', c''\}$: $\delta(c', c'') < \delta_{\min} = \min\{\delta(c, c') \,;\, \delta(c, c'')\}$. What we do not want is a class c such that $\delta(c, d) < \delta_{\min}$, since this would imply that class c should have been aggregated with d before aggregating c' and c''. Therefore the aggregation index must satisfy the following median inequality.

Definition 3.29 (Median inequality). *Aggregation index δ is said to verify the median inequality, if one has:*

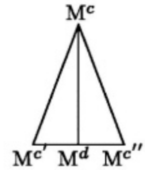
$$\forall c \in C - \{c', c''\} : \delta(c', c'') \leq \delta_{\min} \Longrightarrow \delta(c, d) \geq \delta_{\min}$$

letting $\delta_{\min} = \min\{\delta(c,c')\,;\,\delta(c,c'')\}$ *and* d *the pooled class of* c' *and* c''.



If c, c', c'' are pictured as vertices of a triangle, the median (c,d) opposed to the smallest side (c',c'') must be greater than (or equal to) the smaller of the other two sides; hence the name "median inequality". However, this pictorial figuration is metaphorical because the δ index is not a genuine distance.

Counterexample. As a separability index between classes, one might think of simply taking the geometric distance between the centers M^c and $\mathrm{M}^{c'}$ of classes c and c'. If classes c' and c'' are aggregated, it may happen that the center M^d of the new class d (lying between $\mathrm{M}^{c'}$ and $\mathrm{M}^{c''}$) is such that $\mathrm{M}^d\mathrm{M}^c < \min\{\mathrm{M}^c\mathrm{M}^{c'}\,;\,\mathrm{M}^c\mathrm{M}^{c''}\}$, which contradicts the median inequality (see opposite figure).



All aggregation indices used in GDA verify the median inequality.

Definition 3.30 (Variance index). *Given two classes* c *and* c', *the variance index is the contribution of the dipole (cf. p.102) of the class centers* M^c *and* $\mathrm{M}^{c'}$, *denoted* $\delta(c,c')$ *with*

$$\delta(c,c') = \frac{f_c f_{c'}}{f_c + f_{c'}}(\mathrm{M}^c\mathrm{M}^{c'})^2$$

This index is also called *inertia index,* or centered moment of order 2, or *Ward index* (see Ward, 1963).

Property 3.26 (Recurrence formula). *If class* d *is obtained by grouping the two classes* c' *and* c'', *the aggregation index verifies the formula:*
$\forall c \in C, c' \in C$ and $\forall c \in C - \{c', c''\}$:

$$\delta(c,d) = \frac{(n_c + n_{c'})\delta(c,c') + (n_c + n_{c''})\delta(c,c'') - n_c\delta(c',c'')}{n_c + n_{c'} + n_{c''}} \tag{3.14}$$

Proof. The sum of squares of the cloud made up of two points $\mathrm{M}^{c'}$ and $\mathrm{M}^{c''}$ with respect to point M^c is $n_{c'}(\mathrm{M}^c\mathrm{M}^{c'})^2 + n_{c''}(\mathrm{M}^c\mathrm{M}^{c''})^2$. From Huyghens' theorem (p.79), this sum of squares can be written $n_d(\mathrm{M}^c\mathrm{M}^d)^2 + n\,\delta(c',c'')$, where $\mathrm{M}^d = (n_{c'}\mathrm{M}^{c'} + n_{c''}\mathrm{M}^{c''})/n_d$ denotes the mean point of subcloud $\{\mathrm{M}^{c'}, \mathrm{M}^{c''}\}$ and $n_d = n_{c'} + n_{c''}$ its weight.
Therefore $n_{c'}(\mathrm{M}^c\mathrm{M}^{c'})^2 + n_{c''}(\mathrm{M}^c\mathrm{M}^{c''})^2 = n_d(\mathrm{M}^c\mathrm{M}^d)^2 + n\,\delta(c',c'')$. Replacing in this equality the squares of distances by their expression in terms of the aggregation index (cf. Definition 3.30), one gets $\frac{n_c+n_{c'}}{n_c}\delta(c,c') + \frac{n_c+n_{c''}}{n_c}\delta(c,c'') = \frac{n_c+n_d}{n_c}\delta(c,d) + \delta(c',c'')$, hence the formula. ◁

Corollary 3.2. *The variance index verifies the median inequality.*

Proof. This property follows from the recurrence formula. Indeed, by minoring $\delta(c,c')$ and $\delta(c,c'')$ by $\delta_{\min}$, and by majoring $\delta(c',c'')$ by $\delta_{\min}$

$$\delta(c,d) \geq \frac{n_c+n_{c'}}{n_c+n_d}\delta_{\min} + \frac{n_c+n_{c''}}{n_c+n_d}\delta_{\min} - \frac{n_c}{n_c+n_d}\delta_{\min} = \delta_{\min} \quad \triangleleft$$

The variance index does not satisfy the triangle inequality, as the following *counterexample* shows. Let three classes c, c', c'' with equidistant centers, with $f_c = \epsilon$, $f_{c'} = f_{c''} = (1-\epsilon)/2$. Hence $\delta(c,c') = \delta(c,c'')$ and $(\delta(c,c') + \delta(c,c''))/\delta(c',c'') = 8\epsilon/(1+\epsilon)$. If $\epsilon < 1/7$, one has $\delta(c,c') + \delta(c',c'') < \delta(c',c'')$. Class c is close to classes c' and c'' which are distant from each other.

In *Euclidean classification,* one takes as an aggregation index the *variance index.* Euclidean classification tends to aggregate firstly classes with small numbers of observations: In brief, the lighter classes are, the earlier they are aggregated.

Local and global viewpoints. In Euclidean classification, it follows from the *grouping property* (p.105) that at each stage the aggregated classes are those which lead to the minimal decrease of the between–variance of the partition, or equivalently to the minimal increase of the within–variance. Thus the variance index which, from a local point of view, is just one measure of deviation between two classes, among others, has good properties from a global viewpoint, since the two requirements of compactness within classes and separability between classes come down to the same thing.

Other aggregation indices in AHC. Beside the variance index, three AHC indices are also used in GDA. Let $j \in J{<}c{>}$ and $j' \in J{<}c'{>}$.

• *Minimal jump.* Let $D_{\min}(c,c') = \min\limits_{jj', j \neq j'} \mathrm{M}^j\mathrm{M}^{j'}$. Taking $D_{\min}(c,c')$ as aggregation index defines the minimal jump, or *single linkage clustering.*

• *Maximal jump.* Let $D_{\max}(c,c') = \max\limits_{jj'} \mathrm{M}^j\mathrm{M}^{j'}$. Taking $D_{\max}(c,c')$ as aggregation index defines the maximal jump, or *diameter index,* or *complete linkage clustering.*

• *Mean distance.* Let $D_{\text{mean}}(c,c') = \sum\limits_{j\in J<c>} \sum\limits_{j'\in J<c'>} f_j f_{j'} \mathrm{M}^j\mathrm{M}^{j'}/(f_c f_{c'})$ (weighted mean of distances between the points of classes c and c'). Taking $D_{\text{mean}}(c,c')$ as aggregation index defines the mean distance index, or *average linkage clustering.*

All three indices verify the median inequality. Only D_{mean} verifies the triangle inequality. The D_{min} index produces clusters that are clearly separated from each other but tend to induce "chaining effects"; the D_{max} index has the opposite features. The D_{mean} index is a compromise between D_{min} and D_{max}, since it takes into account all segments of the classes (the smallest as well as the biggest) like the variance index. For this reason, the D_{mean} index and the variance index are usually preferred; the latter has the global property stated earlier, which also entails technical advantages in term of memory space and algorithmic complexity.

3.6.4 Basic and Reciprocal Neighbor Algorithms

Basic algorithm. Once an aggregation index has been chosen, the *basic algorithm* of AHC is as follows:

Step 1. From the table of distances between the J points, calculate the aggregation index for the $J(J-1)/2$ pairs of one–element classes, then aggregate a pair of classes for which the index is minimum: hence a partition into $J-1$ classes.

Step 2. Calculate the aggregation indices between the new class and the $J-2$ others, and aggregate a pair of classes for which the index is minimum; this gives a second partition into $J-2$ classes in which the first partition is nested.

Step 3. Iterate the procedure until a single class is reached.

In *Euclidean classification,* at step ℓ, the partition C_ℓ is the between–cloud M^{C_ℓ}; the minimum of the aggregation index is called *level index* and denoted δ_ℓ, with $\delta_\ell = \mathrm{Var}\,\mathrm{M}^{C_{\ell-1}} - \mathrm{Var}\,\mathrm{M}^{C_\ell}$. At each step, it is sufficient to consider the between–cloud, without considering the points inside classes. In brief, "one thinks about classes, but one works on class centers".

Reciprocal neighbor algorithm. The calculation time for the basic algorithm is of the order of J^3, since at each step one seeks the smallest aggregation index of all node pairs. There are algorithms which lessen calculation time; for instance, the reciprocal neighbor algorithm, whose principle goes back to McQuitty (1966), introduces — without complicating the basic algorithm — an acceleration principle by aggregating at each step the node pairs that are nearest *reciprocal neighbors.*

Definition 3.31. *Two points* M^c *and* $\mathrm{M}^{c'}$ *are said to be reciprocal neighbors if* $(\delta(c,c''))_{\substack{c''\in C\\ c''\neq c}}$ *is minimum for* $c''=c'$ *and if* $(\delta(c',c''))_{\substack{c''\in C\\ c''\neq c'}}$ *is minimum for* $c''=c$.

In the basic algorithm, indeterminations may arise when pairs of nodes have equal aggregation indices. The reciprocal neighbor algorithm often enables one to remove such indeterminations. When there is no indetermination, the reciprocal neighbor algorithm yields the same result as the basic algorithm for any aggregation index that verifies the median inequality (see Benzécri, 1982b). The reciprocal neighbor algorithm will be illustrated on the *Target example.*

δ	$j1$	$j2$	$j3$	$j4$	$j5$	$j6$	$j7$	$j8$	$j9$
$j2$	**2**								
$j3$	11.6	4							
$j4$	6.8	3.2	4						
$j5$	14.4	6.8	**2**	2					
$j6$	13	17	27.4	10.6	20.2				
$j7$	13	10.6	12.2	2.6	5.8	5.2			
$j8$	14.6	9.8	8.2	1.8	2.6	10	**0.8**		
$j9$	29.2	20.8	13.6	8	5.2	19.4	5	2.6	
$j10$	31.4	21.8	13	9	5	23.2	6.8	3.6	**0.2**

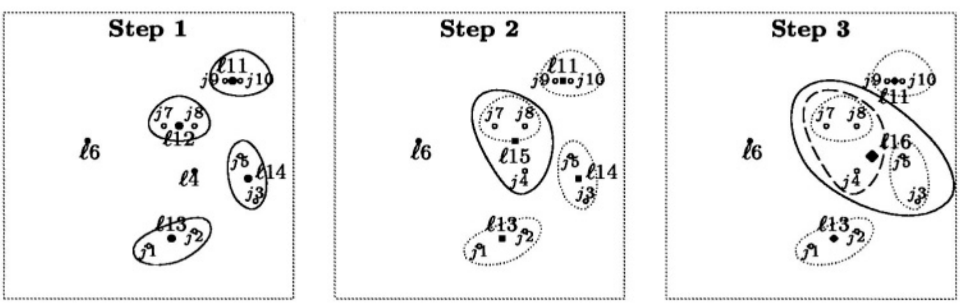
Table 3.1. ***Target example*: contributions of the $9 \times 10/2 = 45$ dipoles**

• **Step 1.** From the distances between pairs of points, calculate the contributions of the 45 dipoles (Table 3.1). Four pairs of reciprocal neighbors are found with aggregation indices $\delta(\ell9, \ell10) = 0.2$ (class $\ell11$); $\delta(\ell7, \ell8) = 0.8$ (class $\ell12$); $\delta(\ell1, \ell2) = 2 = \delta(\ell3, \ell5)$ (classes $\ell13$ and $\ell14$).

• **Step 2.** Applying the recurrence formula, calculate the aggregation indices between the 4 new classes $\ell11, \ell12, \ell13, \ell14$ and $\ell4$, then between these 4 new classes and $\ell6$, then between the 4 new classes (opposite Table).

δ	$\ell4$	$\ell6$	$\ell11$	$\ell12$	$\ell13$
$\ell6$	10.6				
$\ell11$	11.267	28.334			
$\ell12$	**2.667**	9.867	8.5		
$\ell13$	6	19.334	50.5	22.6	
$\ell14$	3.334	31.067	17.3	13	16.4

Classes $\ell4$ and $\ell12$ are reciprocal neighbors hence the new class $\ell15 = \{j4, j7, j8\}$ at aggregation level 2.667.



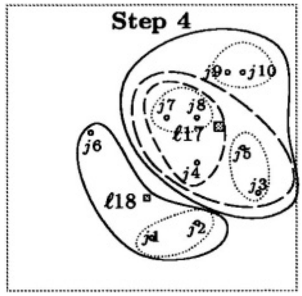
• **Step 3.** Calculate the aggregation indices between class $\ell 15$ and the 4 classes $\ell 6$, $\ell 11$, $\ell 13$, and $\ell 14$, hence the opposite table.

δ	$\ell 6$	$\ell 11$	$\ell 13$	$\ell 14$
$\ell 11$	28.334			
$\ell 13$	19.334	50.5		
$\ell 14$	31.067	17.3	16.4	
$\ell 15$	12.033	12.493	20.613	**11.333**

The two classes $\ell 14$ and $\ell 15$ are aggregated into the class $\ell 16$ at level 11.333.

• **Step 4.** Calculate the three aggregation indices between $\ell 16$ and $\ell 6$ $\ell 11$ and $\ell 13$. Hence the following table:

δ	$\ell 6$	$\ell 11$	$\ell 13$
$\ell 11$	28.333		
$\ell 13$	**19.333**	50.5	
$\ell 16$	20.666	**15.571**	20.857



There are two pairs of reciprocal neighbors $(\ell 11, \ell 16)$ and $(\ell 6, \ell 13)$; they are respectively aggregated at level 15.572 into class $\ell 17 = \{j3, j4, j5, j7, j8, j9, j10\}$ and at level 19.333 into class $\ell 18 = \{j1, j2, j6\}$).

• **Step 5.** The aggregation of the two nodes $\ell 17$ and $\ell 18$ at level 38.095 (node $\ell 19$) achieves the classification.

Property 3.27 (Breakdown of variance for a class hierarchy). *The sum of all level indices is equal to the variance of the cloud.*

Proof. $$\sum_{\ell=J+1}^{2J-1} \delta_\ell = \sum_{\ell=J+1}^{2J-1} \mathrm{Var}\, M^{C_{\ell-1}} - \sum_{\ell=J+1}^{2J-1} \mathrm{Var}\, M^{C_\ell} = \mathrm{Var}\, M^J \quad \triangleleft$$

For each grouping, the level index is equal to the decrease of the between–variance, or else to the increase of the within–variance: When the nodes $\ell 11$ and $\ell 16$ (step 4) are grouped, the between–variance decreases by 15.571.

As one ascends the construction, the level indices form an *increasing sequence*; the higher in the hierarchy, the higher the heterogenity level where aggregation is made; this property follows from median inequality.

At each level, the between–variance of the corresponding partition is equal to the sum of the indices of higher levels (above), and the within–variance is equal to the sum of this level and those of lower levels (below). For instance, for the partition into three classes $\ell 6$, $\ell 13$ and $\ell 17$, the between–variance is the sum of the $C - 1 = 2$ level indices running from $\ell = 2J - C + 1 = 18$ to $2J - 1 = 19$, that is $19.33 + 38.10 = 57.43$; the between–variance is equal to $0.2 + 0.8 + 2 + 2 + 2.67 + 11.33 + 15.67$, that is, 34.57. The sum of the nine level indices is equal to 92, i.e. the total variance. Table 3.2 (p.114) summarizes the 9 levels of the classification of the *Target example* (the classification process goes from bottom to top).

ℓ	δ_ℓ	classes		n	class description
$\ell 19$	38.095	$\ell 18$	$\ell 17$	10	$j9\,j10\,j3\,j5\,j4\,j7\,j8\,j6\,j1\,j2$
$\ell 18$	19.333	$\ell 13$	$\ell 6$	3	$j6\,j1\,j2$
$\ell 17$	15.571	$\ell 16$	$\ell 11$	7	$j9\,j10\,j3\,j5\,j4\,j7\,j8$
$\ell 16$	11.333	$\ell 15$	$\ell 14$	5	$j3\,j5\,j4\,j7\,j8$
$\ell 15$	2.667	$\ell 12$	$\ell 4$	3	$j4\,j7\,j8$
$\ell 14$	2.	$\ell 5$	$\ell 3$	2	$j3\,j5$
$\ell 13$	2.	$\ell 2$	$\ell 1$	2	$j1\,j2$
$\ell 12$	0.8	$\ell 8$	$\ell 7$	2	$j7\,j8$
$\ell 11$	0.2	$\ell 10$	$\ell 9$	2	$j9\,j10$

Between Var		η_ℓ^2
$\ell 19$	38.10	.414
$\ell 18$	57.43	.624
$\ell 17$	73.00	.793
$\ell 16$	84.33	.917
$\ell 15$	87.00	.957
$\ell 14$	89.00	.967
$\ell 13$	91.90	.989
$\ell 12$	91.80	.998
$\ell 11$	92.00	1

Table 3.2. ***Target example*: steps of the AHC.**

3.6.5 Reading a Classification, Aids to Interpretation

The hierarchical tree, once constructed, can be read in one of two ways, ascending or descending. The *ascending reading* allows one to single out, at the basis of the tree, the objects or classes of objects that are nearest; the *descending reading* is a better way for the overall interpretation, according to the maxim: "One constructs the hierarchy by ascending, one interprets it by descending".

When what is sought is a partition with a number of classes specified in advance, descending reading yields such a partition by cutting the tree as already shown in the *Target example* (Figure 3.9, p.107). More often, a hierarchy is sought with partitions whose numbers of classes are not fixed in advance. One will then look for a compromise between the number of classes — preferably small — and the proportion of variance — preferably high — accounted for by the finest partition of the retained hierarchy. The further one goes down the tree, the more classes there are, and so the accounted part of variance increases, but this increase gets smaller and smaller as level indices decrease. One will end the procedure at a degree of fineness beyond which a further subdividing of classes would not bring any appreciable and/or interpretable variance increase.

Target example. Cutting the tree between levels $\ell 19$ and $\ell 18$ generates the partition into two classes $\ell 18$ and $\ell 17$, with $\ell 18 = \{j6, j1, j2\}$ and $\ell 17 = \{j3, j5, j4, j7, j8, j9, j10\}$. This partition only accounts for 41.4% of the variance of the cloud; one will therefore cut the tree at a lower level. As a rule of thumb, the between–variance should be greater than the within–variance $(\eta^2 > 0.5)$. Cutting between levels $\ell 18$ and $\ell 17$ determines the hierarchy with the previous three–class partition accounting for 62.4% of variance. Cutting the tree between levels $\ell 17$ and $\ell 16$ would generate a still finer hierarchy, with a four–class partition accounting for 79.3% of variance. Which of the foregoing hierarchies (with three or four–class partitions) is to be eventually retained will be a matter of appreciation. With real data, "one of the major roles of interpretation is to specify among the dichotomies of an AHC those which correspond to divisions that are geometrically well–defined and conceptually interpretable." (Benzécri, 1992, p. 573).

> *To sum up*: The end–product of the classification procedure is a hierarchy represented by a tree whose classes and partitions are interpreted.

3.6.6 Other Clustering Methods and Issues

Optimal partitions

An attractive feature of AHC is that it generates a set of nested partitions which permit differentiated interpretations. On the other hand, the AHC method does not possess the desirable property that each partition produced by the hierarchy has minimum variance among all the partitions with the same number of classes.

Target example. The η^2 of the partition into three classes $\{i1, i2, i3\}$, $\{i6\}$, $\{i4, i5, i7, i8, i9, i10\}$ is equal to 0.651, therefore greater than the η^2 $(= 0.624)$ of the three–class partition of the AHC.

Nevertheless, starting with a partition produced by AHC, one can always try to improve it in order to get (at least locally) an optimal partition with a fixed number of classes.

Among such methods let us mention the *method of moving centers,* that of *dynamic clustering* ("Nuées dynamiques"), and that of *stable clusterings* (Diday, 1971), which consists in constructing several partitions starting with different sets of centers, and looking for stable clusters, i.e. sets of individuals that remain in the same clusters throughout the process.

Let us also mention the *transfer algorithm,* which consists in moving one element from a class to another. Transferring element j from class c to class c' entails a variation of the between–variance equal to $(f_c f_j/(f_c - f_j)) \times (\mathrm{M}^c\mathrm{M}^j)^2 - (f_{c'} f_j/(f_{c'} + f_j)) \times (\mathrm{M}^{c'}\mathrm{M}^j)^2$. Given a partition, for each possible transfer we calculate the variation of between–variance, and choose the transfer for which the variance increase is greatest, then the procedure is iterated until the variance increase becomes negligible.

Joint use of GDA and Classification

In the analysis of a complex data set, it is recommended, as a rule, to use GDA and classification jointly. The two methods are complementary, especially at the stage of interpretation. In this connection, there is an issue that must be dealt with in each situation, and that is *overall vs principal classification.* Since in GDA only a principal cloud is retained, the question arises whether the classification should be performed either on the full multidimensional cloud or on its principal projection.

3.7 From Points to Numbers: Matrix Formulas

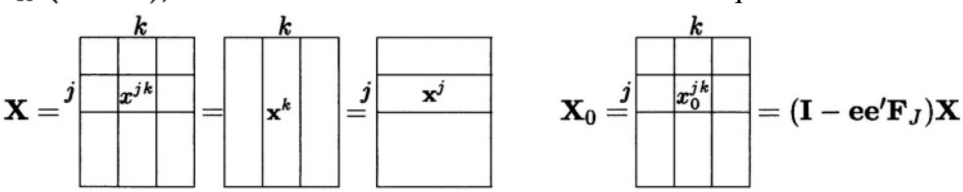
By choosing a Cartesian frame and working on the coordinates of points, we will now go "from points to numbers", and get "numerical principal formulas", with their matrix expressions, for the direction and eigenvalues equations, and the passage and reconstitution formulas.

Letting $(\mathrm{O}, (\overrightarrow{\delta_k})_{k\in K})$ be a Cartesian frame of $\mathcal{U}$ with $q_{kk'} = \langle\overrightarrow{\delta_k}|\overrightarrow{\delta_{k'}}\rangle$, we denote $(x^{jk})_{k\in K}$ the coordinates of point M^j, that is: $\overrightarrow{\mathrm{OM}}^j = \sum\limits_{k\in K} x^{jk}\overrightarrow{\delta_k}$. The family $x^{jK} = (x^{jk})_{k\in K}$ is called the *initial profile* of point M^j ("initial" is opposed to "principal"). The family $x^{Jk} = (x^{jk})_{j\in J}$ defines the k–th *initial variable*; we denote $\overline{x}^k = \sum\limits_{j\in J} f_j\, x^{jk}$ its mean, $v^{kk} = \sum\limits_{j\in J} f_j(x^{jk} - \overline{x}^k)^2$ its variance, and $v^{kk'}$ the covariance between the initial variables x^{Jk} and $x^{Jk'}$. The K variables $(x^{Jk})_{k\in K}$ define the K–numerical protocol x^{JK}.

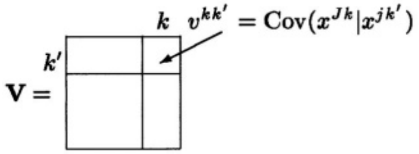
When individuals j are grouped, coordinates average up, that is, x^{Jk} is indeed a variable on J, therefore the upper position of index j is meaningful in terms of duality notation. On the other hand, the upper position of index k is here arbitrary.

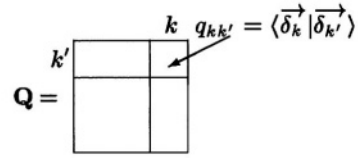
When the mean point G of the cloud is taken as the origin, the initial variables are *centered*: $(x_0^{Jk})_{k\in K}$, with $x_0^{jk} = x^{jk} - \overline{x}^k$.

Matrix notation. We denote $\mathbf{X}$ the table ($J \times K$) of protocol x^{JK}, $\mathbf{x}^j$ the K–row representing the profile of j and $\mathbf{x}^k$ the J–column representing variable x^{Jk}, $\mathbf{X}_0$ the centered protocol x_0^{JK}; we denote $\mathbf{f}_J$ the J–column of $(f_j)_{j\in J}$, and $\mathbf{F}_J$ the diagonal matrix with $(f_j)_{j\in J}$ diagonal terms. The following auxiliary matrices is used: the identity matrices $\mathbf{I}_J$ $(J \times J)$ and $\mathbf{I}_K$ $(K \times K)$, as well as the J–column $\mathbf{e}$ with all terms equal to 1.



We denote $\mathbf{V}$ the $(K \times K)$ matrix of covariances between initial variables $(x^{Jk})_{k\in K}$, with $\mathbf{V} = \mathbf{X}_0'\mathbf{F}_J\mathbf{X}_0$, and $\mathbf{Q}$ the $(K \times K)$ matrix of scalar products between the basis vectors $(\overrightarrow{\delta_k})_{k\in K}$.





3.7.1 Basic Statistics

Property 3.28. *The Cartesian coordinates of the mean point of a cloud are the means of the initial variables.*

Proof. $\overrightarrow{\mathrm{OG}} = \sum\limits_{j\in J} f_j \overrightarrow{\mathrm{OM}}^j = \sum\limits_{j\in J} f_j \sum\limits_{k\in K} x^{jk}\overrightarrow{\delta_k} = \sum\limits_{k\in K}(\sum\limits_{j\in J} f_j\, x^{jk})\overrightarrow{\delta_k}.$ ◁

If $\overline{\mathbf{x}}^k$ denotes the K–row of means $(\overline{x}^k)_{k\in K}$, one has $\overline{\mathbf{x}}^k = \mathbf{f}_J'\mathbf{X}$.

Property 3.29. *The variance of cloud is equal to* $\sum\limits_{k\in K}\sum\limits_{k'\in K} v^{kk'} q_{kk'}$.

Proof. From $(\mathrm{GM}^j)^2 = \langle \sum\limits_{k\in K}(x^{jk} - \overline{x}^k)\overrightarrow{\delta_k} | \sum\limits_{k'\in K}(x^{jk'} - \overline{x}^{k'})\overrightarrow{\delta_{k'}}\rangle$, it follows:
$\sum\limits_{j\in J} f_j\,(\mathrm{GM}^j)^2 = \sum\limits_{k\in K}\sum\limits_{k'\in K}\big(\sum\limits_{j\in J} f_j(x^{jk} - \overline{x}^k)(x^{jk'} - \overline{x}^{k'})\big)\,\langle\overrightarrow{\delta_k}|\overrightarrow{\delta_{k'}}\rangle.$ ◁

Matrix notation. $(\mathrm{GM}^j)^2 = \mathbf{x}_0'\mathbf{Q}\mathbf{x}_0^j$, $\mathrm{Var}\,\mathrm{M}^J = \mathbf{tr}\,\mathbf{VQ}$ (trace of $\mathbf{VQ}$). If the frame is orthonormal, one has $\mathbf{Q} = \mathbf{I}_K$, and the variance of cloud is equal to $\mathbf{tr}\mathbf{V} = \sum\limits_{k\in K}\mathrm{Var}\,x^{jk}$ (sum of the variances of initial variables).

3.7.2 Principal Formulas

To get the principal formulas in matrix form, one starts from the linear mappings defined in §3.2.5 (p.84), and writes their matrices in the basis $(\overrightarrow{\delta_k})_{k\in K}$ of $\mathcal{V}$ and in the canonical bases $(\delta_j^J)_{j\in J}$ of $\mathbb{R}^J$ and $(\delta_J^j)_{j\in J}$ of $\mathbb{R}_J$.

Property 3.30. *The $(J \times K)$ matrix of Vac is* $\mathbf{X}_0\mathbf{Q}$.

Proof. From Definition 3.11 (p.84) $Vac(\overrightarrow{\delta_k}) = \sum_{j \in J} \langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\delta_k} \rangle \, \delta^J_j$. One has $\langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\delta_k} \rangle = \sum_{k' \in K} x_0^{jk} \langle \overrightarrow{\delta_{k'}} | \overrightarrow{\delta_k} \rangle$; hence the (j, k)–term of the matrix of *Vac* is equal to $\langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\delta_k} \rangle = \sum_{k' \in K} q_{kk'} x_0^{jk'}$. ◁

Corollary 3.3. *The variance of cloud in direction* $\overrightarrow{\alpha} = \sum_{k \in K} a^k \overrightarrow{\delta_k}$ *is equal to* $(\mathbf{a}'\mathbf{QVQa})/(\mathbf{a}'\mathbf{Qa})$, *where* $\boldsymbol{a}$ *denotes the column of coefficients* $(a^k)_{k \in K}$.

Proof. From property 3.8 (p.86), the variance in direction $\overrightarrow{\alpha}$ is equal to $\|Vac(\overrightarrow{\alpha})\|^2 / \|\overrightarrow{\alpha}\|^2$. The numerator is equal to $(\mathbf{X}_0\mathbf{Qa})'\mathbf{F}_J(\mathbf{X}_0\mathbf{Q})\mathbf{a} = \mathbf{a}'\mathbf{Q}(\mathbf{X}_0'\mathbf{F}_J\mathbf{X}_0)\mathbf{Qa}$, and the denominator to $\mathbf{a}'\mathbf{Qa}$, hence the formula. ◁

Property 3.31. *Let* $\mathbf{V}_L$ *be the* $(L \times L)$ *matrix of covariances of the L variables associated with a basis of the support of cloud, and* $\mathbf{Q}_L$ *the* $(L \times L)$ *matrix of the scalar products of the L vectors of this basis; if one has the property* $\mathbf{V}_L = \lambda \mathbf{Q}_L^{-1}$, *the cloud is spherical.*

Property 3.32. *The* $(K \times J)$ *matrix of Eff is* $\mathbf{X}_0'$.

Proof. $\mathit{Eff}(\delta^j_J) = \sum_{j' \in J} \delta^j_{j'} \overrightarrow{\mathrm{GM}}^{j'} = \overrightarrow{\mathrm{GM}}^j = \sum_{k \in K} x_0^{jk} \overrightarrow{\delta_k}$ (Definition 3.12, p.84).
The (k, j)–term of the *Eff* matrix is therefore x_0^{jk}. ◁

The preceding two properties entail the following three corollaries.

Corollary 3.4. *The* $(K \times J)$ *matrix of* $Vac^\star = \mathit{Eff} \circ f_J$ *is* $\mathbf{X}_0'\mathbf{F}_J$.

Corollary 3.5. *The* $(K \times K)$ *matrix of* $Som = Vac^\star \circ Vac$ *is*

$$\mathbf{X}_0'\mathbf{F}_J\mathbf{X}_0\mathbf{Q} = \mathbf{VQ}$$

(Recall that the (k, k')–term of $\mathbf{V}$ is $v^{kk'} = \sum_j f_j x_0^{jk} x_0^{jk'}$).

Corollary 3.6. *The matrix of* $Tom = Vac \circ Vac^\star$ *is the* $(J \times J)$ *matrix* $\mathbf{X}_0\mathbf{QX}_0'\mathbf{F}_J = \mathbf{WF}_J$, *with* $\mathbf{W} = \mathbf{X}_0\mathbf{QX}_0'$, *with* $w_{jj'} = \langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\mathrm{GM}}^{j'} \rangle$.

Relative to the orthonormal principal bases, the matrices of $Vac : \mathcal{L} \to \mathcal{X}_0$ and $Vac^* : \mathcal{X}_0 \to \mathcal{L}$ are both equal to the diagonal matrix $\boldsymbol{\Xi}$ of singular values $(\xi_\ell)_{\ell=1,\ldots L}$; the matrices of *Som* and *Tom* are diagonal and both equal to $\boldsymbol{\Lambda} = \boldsymbol{\Xi}^2$.

Property 3.33. *In matrix terms, the eigendirection and eigenvalue equation is* $\mathbf{VQa}_\ell = \lambda_\ell \mathbf{a}_\ell$, *with* $\mathbf{a}_\ell'\mathbf{Qa}_{\ell'} = 0$ *for* $\ell \neq \ell'$.

Proof. From Theorem 3.3 (p.88), one has: $\sum\limits_{k'\in K}\Big(\sum\limits_{k''\in K} v^{kk''} q_{k'k''}\Big) a^{k'} = \lambda\, a^k$; hence $\mathbf{VQa} = \lambda\mathbf{a}$ ◁

Le us denote $\mathbf{A}$ the *(K × L)* matrix of the L columns $(\mathbf{a}_\ell)_{\ell=1,\dots L}$, and $\mathbf{\Lambda}$ the *(L × L)* diagonal matrix of eigenvalues $(\lambda_\ell)_{\ell=1,\dots L}$, and $\mathbf{I}_L$ the *(L × L)* identity matrix; looking for unit–norm eigenvectors one has:

$$\mathbf{VQA} = \mathbf{A\Lambda} \quad \text{with } \mathbf{A'QA} = \mathbf{I}_L \tag{3.15}$$

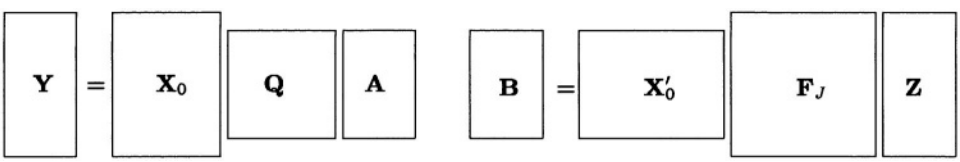
Property 3.34 (Passage formulas). *In matrix terms, passage formulas between the* ℓ*–th principal direction (K–column* $\mathbf{a}_\ell$*) and the* ℓ*–th standard principal variable (J–column* $\mathbf{z}_\ell$*) read:*

$$\begin{cases} \mathbf{X}_0\mathbf{Q}\mathbf{a}_\ell &= \xi_\ell \mathbf{z}_\ell \quad (Vac) \\ \mathbf{X}_0{}'\mathbf{F}_J\mathbf{z}_\ell &= \xi_\ell \mathbf{a}_\ell \quad (Vac^\star) \end{cases} \quad \text{with} \quad \begin{cases} \mathbf{a}_\ell{}'\mathbf{Q}\mathbf{a}_\ell &= 1 \\ \mathbf{z}_\ell{}'\mathbf{F}_J\mathbf{z}_\ell &= 1 \end{cases} \tag{3.16}$$

In the *profile space,* the first equation corresponds to the passage from the initial profile x_0^{jK} of j to its principal profile $(y_\ell^j = \sum\limits_{k\in K} a_{k\ell} x_0^{jk})_{\ell=1,\dots L}$, letting $a_{k\ell} = \sum\limits_{k'\in K} q_{kk'} a_\ell^{k'}$. In the *variable space,* the second equation corresponds to the passage from the k–th initial variable x^{Jk} to its regression coefficient $b_\ell^k = \sum\limits_{j\in J} f_j z_\ell^j x_0^{jk}$ with the ℓ–th standard principal variable.

If we denote $\mathbf{Z}$ the $(J \times L)$ matrix of the L standard principal variables $(z_\ell^J)_{\ell=1,\dots L}$, and $\mathbf{\Xi}$ the *(L × L)* diagonal matrix of the singular values $(\xi_\ell)_{\ell=1,\dots L}$, the passage formulas read:

$$\begin{cases} \mathbf{X}_0\mathbf{QA} &= \mathbf{Z\Xi} \quad (Vac) \\ \mathbf{X}_0{}'\mathbf{F}_J\mathbf{Z} &= \mathbf{A\Xi} \quad (Vac^\star) \end{cases} \quad \text{with} \quad \begin{cases} \mathbf{A'QA} &= \mathbf{I}_L \\ \mathbf{Z'F}_J\mathbf{Z} &= \mathbf{I}_L \end{cases} \tag{3.17}$$

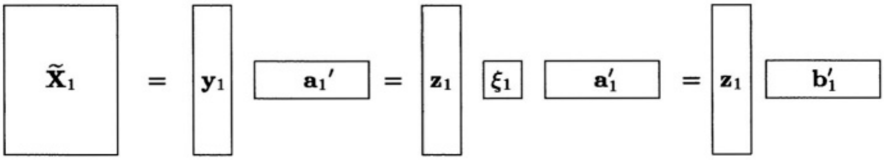


Corollary 3.7 (Principal variable equation).
$\mathbf{X}_0\mathbf{QX'}_0\mathbf{F}_J\mathbf{z}_\ell = \lambda_\ell\mathbf{z}_\ell$ *with* $\mathbf{z}'_\ell\mathbf{F}_J\mathbf{z}_\ell = 1$, *that is* $\mathbf{WF}_J\mathbf{z}_\ell = \lambda_\ell\mathbf{z}_\ell$

Property 3.35 (Reconstitution of order 1). *With the first principal one–dimensional cloud there is associated the reconstitution* $\widetilde{\mathbf{X}}_1$ *of order 1 of* $\mathbf{X}_0$*:*

$$\widetilde{\mathbf{X}}_1 = \mathbf{y}_1\mathbf{a}'_1 = \xi_1\mathbf{z}_1\mathbf{a}_1{}' = \mathbf{z}_1\mathbf{b}'_1$$

Proof. $\overrightarrow{\mathrm{GA}}^j_\ell = y^j_\ell \overrightarrow{\alpha_\ell} = y^j_\ell \sum\limits_{k\in K} a^k_\ell \overrightarrow{\delta_k} = \sum\limits_{k\in K} \widetilde{x}^{jk}_\ell \overrightarrow{\delta_k}$; hence $\widetilde{x}^{Jk}_\ell = a^k_\ell y^J_\ell$. ◁



More generally, with the principal L'–dimensional cloud there *is* associated the *reconstitution of order* L':

$$x^{Jk}_{L'} = \sum_{\ell=1}^{L'} a^k_\ell y^J_\ell = \sum_{\ell=1}^{L'} b^k_\ell z^J_\ell \tag{3.18}$$

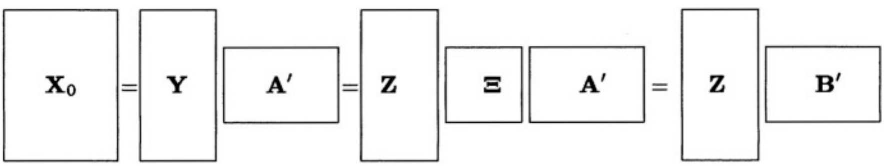
hence the reconstitution of $\widetilde{\mathbf{X}}_{L'}$ of order L' of $\mathbf{X}_0$:

$$\widetilde{\mathbf{X}}_{L'} = \sum_{\ell=1}^{L'} \mathbf{y}_\ell \mathbf{a}_{\ell'} = \sum_{\ell=1}^{L'} \xi_\ell \mathbf{z}_\ell \mathbf{a}'_\ell \quad \text{with} \quad \mathbf{a}'_\ell \mathbf{Q} \mathbf{a}_\ell = 1 \text{ and } \mathbf{z}'_\ell \mathbf{F}_J \mathbf{z}_\ell = 1$$

Proof. $\overrightarrow{\mathrm{GA}}^j_{L'} = \sum\limits_{\ell=1}^{L'} \overrightarrow{\mathrm{GA}}^j_\ell = \sum\limits_{k\in K} \Big(\sum\limits_{\ell=1}^{L'} y^j_\ell a^k_\ell\Big) \overrightarrow{\delta_k} = \sum\limits_{k\in K} \widetilde{x}^{jk}_{L'} \overrightarrow{\delta_k}$. ◁

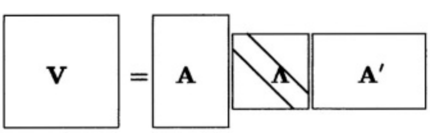
For $L' = L$, one gets the *formula of complete reconstitution:*

$$\mathbf{X}_0 = \mathbf{Y}\mathbf{A}' = \mathbf{Z}\boldsymbol{\Xi}\mathbf{A}' = \mathbf{Z}\mathbf{B}'$$



From the reconstitution formula, the other principal formulas can be derived. For instance, postmultiplying both sides of the equation $\mathbf{X}_0 = \mathbf{Z}\boldsymbol{\Xi}\mathbf{A}'$ by $\mathbf{QA}$, one gets: $\mathbf{X}_0\mathbf{QA} = \mathbf{Z}\boldsymbol{\Xi}\mathbf{A}'\mathbf{QA} = \mathbf{Z}\boldsymbol{\Xi}$, etc. To the reconstitution of $\mathbf{X}_0$ there is attached the *reconstitution of matrix* $\mathbf{V}$ *of covariances:* that of order 1 is $\lambda_1 \mathbf{a}_1 \mathbf{a}'_1$; that of order L' (with $L' \leq L$) is such that $\sum\limits_{\ell=1}^{L'} \lambda_\ell \mathbf{a}_\ell \mathbf{a}'_\ell$.

The complete reconstitution of matrix $\mathbf{V}$ is $\sum\limits_{\ell=1}^{L} \lambda_\ell \mathbf{a}_\ell \mathbf{a}'_\ell$, that is $\mathbf{V} = \mathbf{A}\boldsymbol{\Lambda}\mathbf{A}'$ with $\mathbf{A}'\mathbf{QA} = \mathbf{I}_L$.



If the initial frame is orthonormal $(\mathbf{Q} = \mathbf{I}_K)$, letting $d_1^{kk'} = v^{kk'} - \lambda_1 a_1^k a_1^{k'}$ be the residual deviation, and $\mathbf{D}_1$ the $K \times K$ matrix of residual deviations, one has: $\sum\limits_{k \in K} \sum\limits_{k' \in K} (d_1^{kk'})^2 = \operatorname{tr} \mathbf{D}_1' \mathbf{D}_1 = \sum\limits_{\ell=2}^{L} \lambda_\ell^2$.

The reconstitution of order L' of the variance of the cloud is $\sum\limits_{\ell=1}^{L'} \lambda_\ell$; the complete reconstitution is $\operatorname{tr} \mathbf{VQ} = \sum\limits_{\ell=1}^{L} \lambda_\ell = \operatorname{tr} \mathbf{\Lambda} = \operatorname{Var} \mathrm{M}^J$.

3.7.3 Principal Hyperellipsoids

In any basis, the κ–**hyperellipsoid** has Cartesian equation $\mathbf{x}'\mathbf{V}^{-1}\mathbf{x} = \kappa^2$ and tangential equation $\mathbf{u}'\mathbf{V}\mathbf{u} = 1/\kappa^2$. With respect to principal basis, those equations become $\mathbf{y}'\mathbf{\Lambda}^{-1}\mathbf{y} = \kappa^2$ and $\mathbf{u}'\Lambda\mathbf{u} = 1/\kappa^2$ respectively.

Exercises of Chapter 3

Exercise 3.1 (Analysis of a distance table)
Let M^J a Euclidean cloud on weighted support (J, n_J) with mean point G. One lets $\Delta^{jj'} = (\mathrm{M}^j\mathrm{M}^{j'})^2$ the square of distance between M^j and $\mathrm{M}^{j'}$; $\Delta^j = \sum\limits_{j' \in J} f_{j'} \Delta^{jj'}$; $\Delta = \sum\limits_{j \in J} \sum\limits_{j' \in J} f_j f_{j'} \Delta^{jj'}$ and $w^{jj'} = \langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\mathrm{GM}}^{j'} \rangle$.

1. Prove that $\Delta^j = (\mathrm{GM}^j)^2 + \operatorname{Var} \mathrm{M}^J$. Deduce: $\Delta = 2 \operatorname{Var} \mathrm{M}^J$.

2. Prove that $\Delta^{jj'} = (\mathrm{GM}^j)^2 + (\mathrm{GM}^{j'})^2 - 2w^{jj'}$.

3. From the two preceding questions deduce *Torgerson's formula:*
$$w^{jj'} = -(\Delta^{jj'} - \Delta^j - \Delta^{j'} + \Delta)/2$$

4. For the elementary cloud of 10 points of the *Target example,* one gives the table of the squares of distances $(\Delta^{jj'})$ between the 45 pairs of points. Deduce $\operatorname{Var} \mathrm{M}^J$.
Given OG = 6, calculate the mean of the squares of distances from the points of the cloud to the point O.

Δ	$j1$	$j2$	$j3$	$j4$	$j5$	$j6$	$j7$	$j8$	$j9$
$j2$	40								
$j3$	232	80							
$j4$	136	64	80						
$j5$	288	136	40	40					
$j6$	260	340	548	212	404				
$j7$	260	212	244	52	116	104			
$j8$	292	196	164	36	52	200	16		
$j9$	584	416	272	160	104	388	100	52	
$j10$	628	436	260	180	100	464	136	72	4

5. For the partition into 3 classes with $J\langle c1\rangle = \{j1, j2\}$, $J\langle c2\rangle = \{j6\}$ and $J\langle c3\rangle = \{j3, j4, j5, j7, j8, j9, j10\}$, one gives the squares of distances from the 3 class centers $\mathrm{M}^{c1}, \mathrm{M}^{c2}, \mathrm{M}^{c3}$ to point G: $(\mathrm{GM}^{c1})^2 = 130$, $(\mathrm{GM}^{c2})^2 = 200$ and $(\mathrm{GM}^{c3})^2 = 16.32654$. Deduce the between–variance then the within–

variance. Using the above table, calculate the variances of the 3 subclouds and deduce the contributions to the variance of the 3 subclouds.

6. From the 10×10 table of question 4, deduce (using Torgerson's formula) the matrix $\mathbf{W}$ of scalar products. Verify numerically that the variables $(y_\ell^J)_{\ell=1,2}$ found in the *Target example* exercise (question 3, p.126) are solutions of the principal variables equation $\mathbf{W}\mathbf{F}_J\mathbf{y}_\ell = \lambda_\ell \mathbf{y}_\ell$ (p.119).
7. *Application to* CA. Given a contingency table n_{JK}, consider the $J \times J$ table of square distances $(\mathrm{M}^j\mathrm{M}^{j'})^2 = \sum\limits_{k\in K} \frac{(f_k^j - f_k^{j'})^2}{f_k} = \delta^{jj'}$. Prove that the principal analysis of the table of distances yields the principal coordinates $(y_\ell^J)_{\ell\in L}$ of CA. Verify the property on the *Eye & Hair Color* (CORRESPONDENCE), by starting with the square distances $\delta^{jj'}$, then constructing the scalar products $w^{jj'}$.

Solution

1. From Huyghens' theorem (p.79) taking M^j as a reference point, one has, : $\sum\limits_{j'\in J} f_j(\mathrm{M}^j\mathrm{M}^{j'})^2 = (\mathrm{GM}^j)^2 + \mathrm{Var}\,\mathrm{M}^J$; hence Δ^j and $\Delta = \sum\limits_{j\in J} f_j\Delta^j$.
2. $(\mathrm{M}^j\mathrm{M}^{j'})^2 = \|\overrightarrow{\mathrm{GM}}^{j'} - \overrightarrow{\mathrm{GM}}^j\|^2 = (\mathrm{GM}^j)^2 + (\mathrm{GM}^{j'})^2 - 2\langle\overrightarrow{\mathrm{GM}}^j|\overrightarrow{\mathrm{GM}}^{j'}\rangle$; hence $\Delta^{jj'}$.
3. From question 1: $(\mathrm{GM}^j)^2 + (\mathrm{GM}^{j'})^2 = \Delta^j + \Delta^{j'} - \Delta$, hence $w^{jj'}$ by replacing in the relation obtained in question 2.
4. The sum of squares of the distances between the 45 pairs of points is equal to 9 200, hence the variance $\frac{1}{10} \times \frac{1}{10} \times 9\,200 = 92$. From Huyghens' theorem (p.79), one has: $\sum\limits_{j\in J} f_j(\mathrm{OM}^j)^2 = 6^2 + 92 = 128$.
5. Between–variance: $\frac{2}{10} \times 130 + \frac{1}{10} \times 200 + \frac{7}{10} \times 16.32654 = 57.429$. From the between–within breakdown of the variance (property 3.23, p.104), one deduces the within–variance: $\mathrm{Var}\,\mathrm{M}^{J(C)} = \mathrm{Var}\,\mathrm{M}^J - \mathrm{Var}\,\mathrm{M}^C = 92 - 57.429 = 34.571$. One has $\mathrm{Var}\,\mathrm{M}^{J<c1>} = \frac{1}{2} \times \frac{1}{2} \times 40 = 10$, hence $\mathrm{Cta}_{J(c1)} = \frac{2}{10} \times \mathrm{Var}\,\mathrm{M}^{J<c1>} = 2$, and $\mathrm{Cta}_{c1} = \frac{2}{10} \times 130 = 26$; from the property 3.19 (p.102), one deduces $\mathrm{Cta}_{J<c1>} = 26 + 2 = 28$. Likewise, one has $\mathrm{Cta}_{J<c2>} = 20$, and $\mathrm{Cta}_{J<c3>} = 44$.

Comment

Analysis of distance tables. When a data set is a $J \times J$ table of distances between J objects, together with a weighting f_J, an Euclidean cloud can be constructed such that the distances between points are reconstituted, at least approximately, using Torgerson formula then the principal variable equation. The analysis is also known as "principal coordinate analysis". It amounts to *metric* MDS.

Analysis of dissimilarity tables. When a data set is a *dissimilarity* table[5], approximate Euclidean clouds can be still be constructed. One method is the *additive constant,* where a constant is added to each dissimilarity, such that the resulting table is a distance table. Another method is *nonmetric* MDS, which produces

[5] A dissimilarity index satisfies the positiveness and symmetry axioms, but not necessarily the triangle axiom.

an approximate multidimensional representation from the mere rank ordering of dissimilarities (Shepard, 1962, 1966). Nonmetric as well as metric MDS are GDA methods, since they produce Euclidean clouds[6]. For further discussion, see Benzécri (1964, 1965), Benzécri & al (1973, Vol. 2, p. 93-95), Saporta (1990).

Exercise 3.2 (Euclidean cloud)

One considers a cloud of 4 points O, A, B, C, such that $\mathrm{OA} = \mathrm{OB} = \mathrm{OC} = r$ and $\mathrm{AB} = \mathrm{AC} = \mathrm{BC} = a$; point O has weight $1-m$, and each of the points A, B and C has weight $m/3$, with $0 < m < 1$. Let O' be the mean point of the subcloud of the three points A, B, C.

1. Express distances $\mathrm{O'A}$, $\mathrm{O'B}$ and $\mathrm{O'C}$ as a function of a. Which relation a and r must verify so that the four points A, B, C and O define a Euclidean cloud? Express distance $\mathrm{OO'}$ as a function of a and r. What is the dimensionality of the cloud? In which case does the cloud reduce to a plane cloud?

   Prove that the mean point G of the cloud is such that $\overrightarrow{\mathrm{OG}} = m\overrightarrow{\mathrm{OO'}}$.

2. Express the variance of the subcloud of the three points A, B, C as a function of a; deduce that the within–contribution of this subcloud to the cloud of the 4 points is equal to $ma^2/3$.

3. One considers the partition in two classes $\mathcal{O} = \{\mathrm{O}\}$, and $\mathcal{A} = \{\mathrm{A,B,C}\}$. Express the *within–variance* as a function of a and m.

   What are the two points of the *between–cloud* associated with this partition, and what are their weights? Deduce the *between–variance* and the *variance* of the 4–point cloud as a functions of r, a and m.

4. What is the variance of the cloud in the direction $\mathrm{OO'}$? of the plane ABC?

5. Show that $\mathrm{OO'}$ is the principal line of the 4–point cloud associated with the eigenvalue $m(1-m)(r^2-a^2/3)$, and that every line going through G parallel to plane ABC is principal with the double eigenvalue $ma^2/6$.

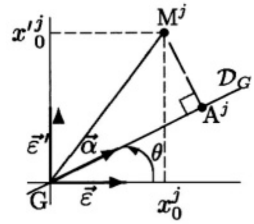
Exercise 3.3 (Principal directions of a plane cloud)

Let (M^J, n_J) be a plane cloud referred to an orthonormal Cartesian frame with origin G (mean point of the cloud) and basis vectors $(\vec{\varepsilon}, \vec{\varepsilon}\,')$.
One lets $\overrightarrow{\mathrm{GM}}^j = x_0^j \vec{\varepsilon} + x'^j_0 \vec{\varepsilon}\,'$. One denotes respectively v, v' the variances of centered variables x_0^J and x'^J_0, c their covariance and r their correlation.[7]

1. Express the variance of the cloud M^J as a function of v and v'.

[6] MDS algorithms do not satisfy the heredity property, that is, the principal coordinates of the one–dimensional representation may differ from the first principal coordinates in the two–dimensional representation, etc.

[7] In this exercise, one solves, by elementary algebra, the problem of principal directions and variables of a plane cloud.



2. Let $\vec{\alpha} = \cos\theta\, \vec{\varepsilon} + \sin\theta\, \vec{\varepsilon}\,'$. Show that the variance of the cloud in direction $\vec{\alpha}$ can be written:
   $V(\theta) = v\cos^2\theta + v'\sin^2\theta + 2c\cos\theta\sin\theta$.
3. Show that, in general, $V(\theta)$ is maximum for $\theta = \theta_1$ and minimum for $\theta = \theta_2$ (determined up to a $k\pi$ factor)verifying equation $c\tan^2\theta + (v - v')\tan\theta - c = 0$.

   Show that $\theta_2 = \theta_1 \pm \pi/2$. Write the solutions $\tan\theta_1$ and $\tan\theta_2$ of this equation and discuss the sign of $\tan\theta_1$ in relation to the sign of c (particular cases $c = 0$ and $\theta = \pm\pi/2$ will be studied in question 8).

   Express unit–norm principal vectors $\vec{\alpha_1}$ and $\vec{\alpha_2}$ as a function of θ_1, $\vec{\varepsilon}$ and $\vec{\varepsilon}\,'$ $(-\pi/2 \leq \theta_1 \leq +\pi/2,\ \theta_2 = \theta_1 + \pi/2)$.
4. For $\ell = 1, 2$, one lets: $\lambda_\ell = v + c\tan\theta_\ell$. Show that θ_ℓ and λ_ℓ (ℓ–th eigenvalue) are solutions of the following system of two equations:
   $v\cos\theta + c\sin\theta = \lambda\cos\theta$ and $c\cos\theta + v'\sin\theta\lambda\sin\theta$
5. Write the two solutions of equation $\lambda^2 - (v + v')\lambda + vv' - c^2 = 0$.
6. Show that $\lambda_\ell = v + c\tan\theta_\ell$ is the variance of the cloud in the direction of angle θ_ℓ $(\ell = 1, 2)$.
7. Prove that if $\lambda_1 \neq \lambda_2$ the 2 principal variables are uncorrelated.
8. Study the following particular cases: $c = 0$ and $v < v'$; $c = 0$ and $v > v'$; $v = v'$ and $c \neq 0$; $c/\sqrt{vv'} = r = \pm 1$.

Solution

1. $\mathrm{Var}\, M^J = \sum_{j\in J} f_j (\mathrm{GM}^j)^2 = \sum_{j\in J} f_j \big((x^j_0)^2 + (x'^j_0)^2\big) = v + v'$.

2. Point A^j, projection of point M^j on line $\mathcal{D}_G$, is such that $\overrightarrow{\mathrm{GA}}^j = y^j\,\vec{\alpha}$, with $y^j = x^j_0\cos\theta + x'^j_0\sin\theta$, hence the variance of cloud A^J is written:

$$V(\theta) = \sum_{j\in J} f_j (x^j_0\cos\theta + x'^j_0\sin\theta)^2 = v\cos^2\theta + v'\sin^2\theta + 2c\sin\theta\cos\theta \qquad (3.19)$$

3. $V(\theta)$ can be derived everywhere on interval $(-\pi/2\,;\pi/2)$ and has derivative: $V'(\theta) = 2(v' - v)\cos\theta\sin\theta + 2c(\cos^2\theta - \sin^2\theta)$; or else (for $\theta \neq \pm\pi/2$): $V'(\theta) = -2\cos^2\theta\big(c\tan^2\theta + (v - v')\tan\theta - c\big)$. The angles θ for which $V(\theta)$ is extremum are the solutions of $V'(\theta) = 0$. Hence the *principal direction equation:*

$$c\tan^2\theta + (v - v')\tan\theta - c = 0 \qquad (\text{with} \quad -\pi/2 < \theta < +\pi/2) \qquad (3.20)$$

This equation ofunknown $\tan\theta$ has two solutions verifying:

$$\tan\theta = (v' - v \pm \sqrt{(v - v')^2 + 4c^2})/2c \qquad (c \neq 0)$$

(for cases $v = v'$, $c = 0$ and $\theta = \pm\pi/2$, cf. question 8).

The study of the sign of $V'(\theta)$ $(-\pi/2 \leq \theta \leq \pi/2)$ shows that $V(\theta)$ is maximum for $\tan\theta_1 > 0$ if $c > 0$, and for $\tan\theta_1 < 0$ if $c < 0$. The angle θ_1 for which $V(\theta)$

is maximum has the sign of c, and it is given by taking the + sign in the above formula. The angle θ_2 rendering $V(\theta)$ minimum is such that $\theta_2 = \theta_1 \pm \pi/2$ since $\tan\theta_1 \times \tan\theta_2 = -1$ (product of the roots of Equation (3.20)). Hence unit–norm principal vectors are $\overrightarrow{\alpha_1} = \cos\theta_1 \overrightarrow{\epsilon} + \sin\theta_1 \overrightarrow{\epsilon}'$ and $\overrightarrow{\alpha_2} = -\sin\theta_1 \overrightarrow{\epsilon} + \cos\theta_1 \overrightarrow{\epsilon}'$.

—*Remark.* $\tan 2\theta = 2\tan\theta/(1-\tan^2\theta)$, then Equation (3.20) can be written $\tan 2\theta = 2c/(v-v')$.

4. The relation $v + c\tan\theta = \lambda$ can be written: $v\cos\theta + c\sin\theta = \lambda\cos\theta$. In Equation (3.20), if one replaces $v + c\tan\theta$ by λ, one has: $\tan\theta(\lambda - v') = c$, hence $c\cos\theta + v'\sin\theta = \lambda\sin\theta$.

Let $\mathbf{a}' = \boxed{\cos\theta \quad \sin\theta}$, and $\mathbf{V}$ the variance–covariance matrix, these equations can be written $\mathbf{Va} = \lambda\mathbf{a}$ (diagonalization of $\mathbf{V}$, cf. *MATH.BASES,* p.443).

5. In Equation (3.20), if one replaces $\tan\theta$ by $(\lambda - v)/c$ $(c \neq 0)$, one obtains the *eigenvalue equation,* whose solutions are:

$$\lambda_1 = \tfrac{v+v'}{2} + \tfrac{1}{2}\sqrt{(v-v')^2 + 4c^2} \qquad \text{and} \qquad \lambda_2 = \tfrac{v+v'}{2} - \tfrac{1}{2}\sqrt{(v-v')^2 + 4c^2}$$

6. If one replaces $v\cos\theta_\ell + c\sin\theta_\ell$ by $\lambda_\ell\cos\theta_\ell$ then $c\cos\theta_\ell + v'\sin\theta_\ell$ by $\lambda_\ell\sin\theta_\ell$ in Equation 3.19, one obtains: $V(\theta_\ell) = \lambda_\ell$.

7. One has:

$$\begin{aligned} \alpha_1^j &= \langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha_1} \rangle \quad \text{therefore} \quad y_1^j = x_0^j \cos\theta_1 + x'^j_0 \sin\theta_1 \\ \alpha_2^j &= \langle \overrightarrow{\mathrm{GM}}^j | \overrightarrow{\alpha_2} \rangle \quad \text{therefore} \quad y_2^j = -x_0^j \sin\theta_1 + x'^j_0 \cos\theta_1 \end{aligned}$$

One has $\mathrm{Cov}(\alpha_1^J|\alpha_2^J) = (v'-v)\cos\theta_1\sin\theta_1 + c(\cos^2\theta_1 - \sin^2\theta_1) = \frac{1}{2}V'(\theta_1)$, hence (question 3): $V'(\theta_1) = 0$, so $\mathrm{Cov}(\alpha_1^J|\alpha_2^J) = 0$.

8. If $c = 0$ and $v > v'$ one has $\theta_1 = 0$ and $\theta_2 = \pi/2$, if $c = 0$ and $v < v'$ one has $\theta_1 = \pi/2$ and $\theta_2 = 0$: initial axes are principal.

If $v = v'$ and $c \neq 0$, then $\tan^2\theta_\ell = 1$, hence $\theta_\ell = \pm\pi/4$ (the principal lines are the bissectors of the plane) and $\lambda = v(1 \pm r)$. If $v = v'$ and $c = 0$, $V(\theta)$ does not depend on θ, then every direction of the plane is principal, and is associated with the eigenvalue of multiplicity 2: $\lambda_1 = \lambda_2 = v = \mathrm{Var}\,\mathrm{M}^J/2$.

For a *one–dimensional cloud,* one has $vv' = c^2$ (or $r = \pm 1$), $\lambda_1 = v + v'$ and $\lambda_2 = 0$.

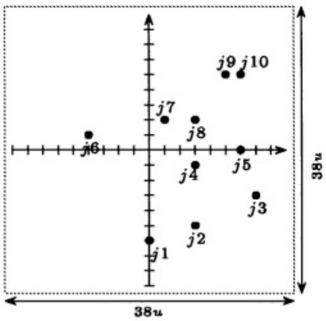
If $c > 0$ and $v > v'$, then $0 < \theta_1 < \pi/4$. If $c > 0$ and $v' > v$, then $\pi/4 < \theta_1 < \pi/2$.

If $c < 0$ and $v > v'$, then $-\pi/4 < \theta_1 < 0$. If $c < 0$ and $v' > v$,

$-\pi/2 < \theta_1 < -\pi/4$. That is, to sum up: $|\theta_1| < \pi/4 \iff v > v'$; $|\theta_1| = \pi/4 \iff v = v'$; $|\theta_1| > \pi/4 \iff v < v'$.

Exercise 3.4 (Target example)

If one refers the plane (p.76) to two rectangular axes (horizontal and vertical) going through point O (center of the target), the abscissas and ordinates of points M^j are shown in the following table (the graduations of the figure below correspond to $2u$). Let $(\mathrm{O}, \overrightarrow{\varepsilon_1}, \overrightarrow{\varepsilon_2})$ be the orthonormal Cartesian frame and x^{Jk1} and x^{Jk2} the two initial variables.



	x^{Jk1}	x^{Jk2}
$j1$	0	−12
$j2$	6	−10
$j3$	14	−6
$j4$	6	−2
$j5$	12	0
$j6$	−8	2
$j7$	2	4
$j8$	6	4
$j9$	10	10
$j10$	12	10

1. Calculate the coordinates of the mean point; deduce the table $\mathbf{X_0}$ of coordinates of points in frame $(\mathrm{G}, \vec{\varepsilon_1}, \vec{\varepsilon_2})$. Write the variance–covariance matrix (denoted $\mathbf{V}$). Deduce the variance of the cloud.

2. Diagonalize covariance matrix V by solving the matrix equation $\mathbf{Va} = \lambda\,\mathbf{a}$, with $\mathbf{a}' = \boxed{\cos\theta \quad \sin\theta}$ and write (2×2) matrix $\mathbf{A}$ verifying $\mathbf{VA} = \mathbf{A\Lambda}$ (with $\mathbf{AA}' = \mathbf{I}$). Draw the principal lines and the two principal clouds.

3. Write the passage formula from the centered initial variables to the calibrated principal variables; calculate the calibrated principal variables and present the matrix diagram going with them. Plot the projected clouds on principal axes 1 and 2 (pictured both horizontally), as well as the plane cloud referred to principal axes (Axis 1 horizontal, Axis 2 vertical).

4. Calculate the regression coefficients (b_1^{k1}, b_2^{k1}) of initial variable x^{Jk1} on the two standard principal variables z_1^J and z_2^J, and the correlation coefficients (r_1^{k1}, r_2^{k2}). Calculate the regression and correlation coefficients for x^{Jk2}. Express the centered initial variables as a function of the standard principal variables; write the reconstitution of matrix $\mathbf{X_0}$. Draw the centered initial variables in the principal plane.

5. Write the Cartesian equations of the concentration ellipse in the initial and the principal frames and plot the ellipse.

Solution

1. Coordinates of the mean point: $(6, 0)$.

$\mathrm{Var}\,x^{Jk1} = 40, \mathrm{Var}\,x^{Jk2} = 52$ and $\mathrm{Cov}(x^{Jk1}|x^{Jk2}) = +8$.

Variance–covariance matrix: $\mathbf{V} = \boxed{\begin{matrix} 40 & 8 \\ 8 & 52 \end{matrix}}$

Variance of the cloud: $\mathrm{Var}\,\mathrm{M}^J = 40 + 52 = 92$.

$$\mathbf{X}_0 = \boxed{\begin{matrix} -6 & -12 \\ 0 & -10 \\ 8 & -6 \\ 0 & -2 \\ 6 & 0 \\ -14 & 2 \\ -4 & 4 \\ 0 & 4 \\ 4 & 10 \\ 6 & 10 \end{matrix}}$$

2. Direction and eigenvalue equations ($\mathbf{Va} = \lambda\mathbf{a}$):

$$\begin{cases} 40\cos\theta + 8\sin\theta = \lambda\cos\theta \\ 8\cos\theta + 52\sin\theta = \lambda\sin\theta \end{cases} \text{ or } \begin{cases} 8\tan\theta = \lambda - 40 \\ (\lambda - 52)\tan\theta = 8 \end{cases}$$

From $\frac{\lambda-40}{8} = \frac{8}{\lambda-52}$, one deduces: $\lambda^2 - 92\lambda + 2016 = 0$.

The solutions of this equation are $\lambda_1 = 56$ and $\lambda_2 = 36$. From the preceding system of two equations, one deduces: $\tan\theta_1 = 2$, hence $\theta_1 = +63°4$ and $\theta_2 = -26°6$.

Thus one has: $\cos^2\theta_1 = 1/(1+\tan^2\theta) = 1/5$, hence $\cos\theta_1 = 1/\sqrt{5}$, $\sin\theta_1 = 2/\sqrt{5}$; and $\cos\theta_2 = -2/\sqrt{5}$, $\sin\theta_2 = 1/\sqrt{5}$.

Passage matrix **A** from initial basis vectors $(\overrightarrow{\varepsilon_1}, \overrightarrow{\varepsilon_2})$ to principal basis vectors $(\overrightarrow{\alpha_1}, \overrightarrow{\alpha_2})$ is equal to:

$$\mathbf{A} = \begin{array}{|c|c|}\hline \cos\theta_1 & \cos\theta_2 \\ \sin\theta_1 & \sin\theta_2 \\ \hline\end{array} = \begin{array}{|c|c|}\hline 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \\ \hline\end{array} = \begin{array}{|c|c|}\hline \mathbf{a}_1 & \mathbf{a}_2 \\ \hline\end{array}$$

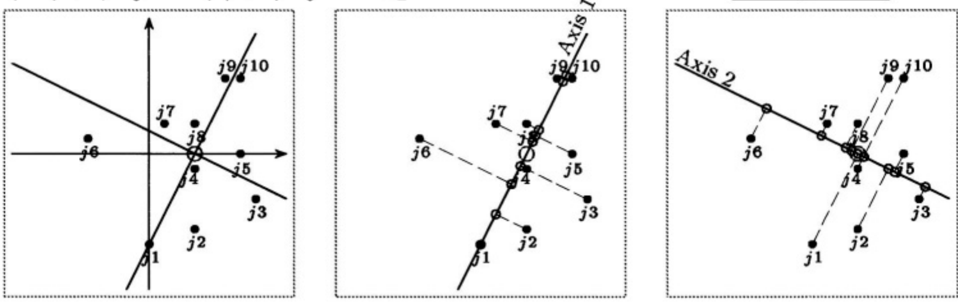
Principal vectors:

$$\overrightarrow{\alpha_1} = (\overrightarrow{\varepsilon_1} + 2\overrightarrow{\varepsilon_2})\sqrt{5}$$
$$\overrightarrow{\alpha_2} = (-2\overrightarrow{\varepsilon_1} + \overrightarrow{\varepsilon_2})/\sqrt{5}$$

3. Passage formula from centered initial variables to calibrated principal variables (cf. opposite scheme):

$$\begin{array}{ccc} & y_1^J & y_2^J \\ \nearrow & \overrightarrow{\alpha_1} & \overrightarrow{\alpha_2} \\ x_0^{Jk1}\ \overrightarrow{\varepsilon_1} & 1/\sqrt{5} & -2/\sqrt{5} \\ x_0^{Jk2}\ \overrightarrow{\varepsilon_2} & 2/\sqrt{5} & 1/\sqrt{5} \end{array} = \mathbf{A}$$

$$(1/\sqrt{5})\,x_0^{Jk1} + (2/\sqrt{5})\,x_0^{Jk2} = y_1^J$$
$$(-2/\sqrt{5})\,x_0^{Jk1} + (1/\sqrt{5})\,x_0^{Jk2} = y_2^J$$

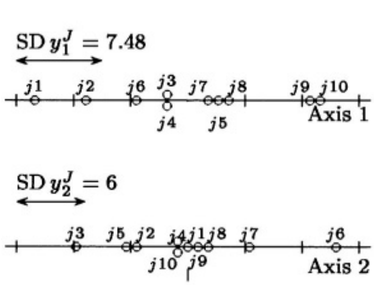


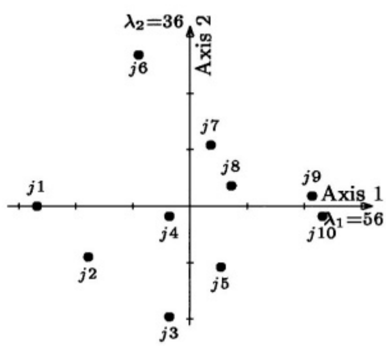
Graphical display of the principal lines and the principal clouds (empty circles)

Matrix formula: $\mathbf{X}_0\mathbf{A} = \mathbf{Y}$ (cf. p 119).

$$\begin{array}{|rr|}\hline -6 & -12 \\ 0 & -10 \\ 8 & -6 \\ 0 & -2 \\ 6 & 0 \\ -14 & 2 \\ -4 & 4 \\ 0 & 4 \\ 4 & 10 \\ 6 & 10 \\ \hline\end{array} \begin{array}{|cc|}\hline \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \hline\end{array} = \frac{1}{\sqrt{5}} \begin{array}{|rr|}\hline -30 & 0 \\ -20 & -10 \\ -4 & -22 \\ -4 & -2 \\ 6 & -12 \\ -10 & 30 \\ 4 & 12 \\ 8 & 4 \\ 24 & 2 \\ 26 & -2 \\ \hline\end{array} = \begin{array}{|rr|}\hline -13.4 & 0 \\ -8.9 & -4.5 \\ -1.8 & -9.8 \\ -1.8 & -0.9 \\ 2.7 & -5.4 \\ -4.5 & 13.4 \\ 1.8 & 5.4 \\ 3.6 & 1.8 \\ 10.7 & 0.9 \\ 11.6 & -0.9 \\ \hline\end{array}$$

Diagrams with graph scale = $1.5u$ (graduations every $5u$).





4. Regression coefficients:

$b_1^{k1} = \sqrt{56} \times \frac{1}{\sqrt{5}} = 3.35 \quad b_1^{k2} = \sqrt{56} \times \frac{2}{\sqrt{5}} = 6.69$

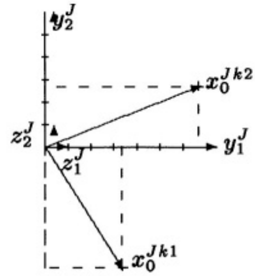
$b_2^{k1} = \sqrt{36} \times \frac{-2}{\sqrt{5}} = -5.37 \quad b_2^{k2} = \sqrt{36} \times \frac{1}{\sqrt{5}} = 2.68$

$x_0^{Jk1} = 3.35\, z_1^J - 5.37\, z_2^J$ and $x_0^{Jk2} = 6.69\, z_1^J + 2.68\, z_2^J$

Correlation coefficients:

$r_1^{k1} = .529,\ r_2^{k1} = -.849,\ r_1^{k2} = .928,\ r_2^{k2} = .372.$

Graphical display in variable space (see opposite figure, with graphical scale equal to $3u$); the coordinates of the initial variables are equal to the regression coefficients.



Reconstitution of centered protocol:

$$
\begin{aligned}
x_0^{Jk1} &= (1/\sqrt{5})\, y_1^J - (2/\sqrt{5})\, y_2^J \\
x_0^{Jk2} &= (2/\sqrt{5})\, y_1^J + (1/\sqrt{5})\, y_2^J
\end{aligned}
$$

$$
\begin{array}{cc|cc|}
 & \swarrow & \begin{matrix} y_1^J \\ \overrightarrow{\alpha_1} \end{matrix} & \begin{matrix} y_2^J \\ \overrightarrow{\alpha_2} \end{matrix} \\
\hline
x_0^{Jk1} & \overrightarrow{\epsilon_1} & 1/\sqrt{5} & -2/\sqrt{5} \\
\hline
x_0^{Jk2} & \overrightarrow{\epsilon_2} & 2/\sqrt{5} & 1/\sqrt{5} \\
\hline
\end{array}
$$

Matrix formulation: $\mathbf{X_0} = \mathbf{YA'}$ (cf. Equations (3.17), p 119).

$$
\begin{bmatrix}
-6 & -12 \\ 0 & -10 \\ 8 & -6 \\ 0 & -2 \\ 6 & 0 \\ -14 & 2 \\ -4 & 4 \\ 0 & 4 \\ 4 & 10 \\ 6 & 10
\end{bmatrix}
= \frac{1}{\sqrt{5}}
\begin{bmatrix}
-30 & 0 \\ -20 & -10 \\ -4 & -22 \\ -4 & -2 \\ 6 & -12 \\ -10 & +30 \\ 4 & +12 \\ 8 & +4 \\ 24 & +2 \\ 26 & -2
\end{bmatrix}
\times
\begin{bmatrix}
\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} & \frac{1}{\sqrt{5}}
\end{bmatrix}
=
\begin{bmatrix}
-6 & -12 \\ -4 & -8 \\ -0.8 & -1.6 \\ -0.8 & -1.6 \\ 1.2 & 2.4 \\ -2 & -4 \\ 0.8 & 1.6 \\ 1.6 & 3.2 \\ 4.8 & 9.6 \\ 5.2 & 10.4
\end{bmatrix}
+
\begin{bmatrix}
0 & 0 \\ 4 & -2 \\ 8.8 & -4.4 \\ 0.8 & -0.4 \\ 4.8 & -2.4 \\ -12 & 6 \\ -4.8 & 2.4 \\ -1.6 & 0.8 \\ -0.8 & 0.4 \\ 0.8 & -0.4
\end{bmatrix}
$$

5. Cartesian equation of the concentration ellipse (cf. §3.4.3, p.97):

in frame $(O, \overrightarrow{\epsilon_1}, \overrightarrow{\epsilon_2})$: $\frac{1}{1-0.175}\left(\frac{(x^{k1})^2}{40} - 6\frac{x^{k1}x^{k2}}{2080} + \frac{(x^{k2})^2}{52}\right) = 4;$

in frame $(O, \overrightarrow{\alpha_1}, \overrightarrow{\alpha_2})$: $(y_1)^2/56 + (y_2)^2/36 = 4.$

Graphical display: cf. p.98.