QÜESTIIÓ, Vol 11, nº 3, pp. 37-57, 1987

EIGENANALYSIS AND METRIC MULTIDIMENSIONAL SCALING ON HIERARCHICAL STRUCTURES

C.M. CUADRAS and J.M. OLLER

University of Barcelona.

The known hierarchical clustering scheme is equivalent to the concept of ultrametric distance. All distance can be represented in a spatial model using multidimensional scaling. We relate both clases of representations of proximity data in an algebraic way, obtaining some results and relations on clusters and the eigenvalues of the inner product matrix for an ultrametric distance. Principal coordinate analysis on an ultrametric distance gives two classes of independent coordinates, describing compact clusters and representing objects inside every cluster.

Keywords: Ultrametric tree, clustering, principal coordiante analysis, latent vectors, proximity data, Euclidean space.

1. INTRODUCTION

Holman (1972) has established the relationship between the Euclidean spatial model and the hierarchical clustering scheme, both frequently used in representation of proximity data. He also proves that the Euclidean dimension required to represent n objects which constitute a hierarchical system is at least n–1. Thus Euclidean representation and network representation seem to be in oppositon.

In this paper we discuss some properties connected with the different Euclidean dimensions that are obtained via principal coordinate analysis from a ultrametric distance matrix, thereby showing that some maximal clusters are well reflected in spatial models.

–C.M. Cuadras i J.M. Oller - Universitat de Barcelona - Facultat de Biologia - Diagonal, 645 - 08028 Barcelona.

–Article rebut el desembre de 1987.

Let $E = \{e_1, \ldots, e_n\}$ be a finite set with $n = |E| \geq 3$ objects, and let δ be a distance on E. Define $\delta_{ij} = \delta(e_i, e_j)$, and thus $\delta_{ij} \geq 0$ with equality if and only if $i = j (i, j = 1, \ldots, n)$.

Let I_n be the $n \times n$ identity matrix and consider the $n \times n$ matrix A whose ij–element is

$$a_{ij} = \left(-1/2 \cdot \delta_{ij}^2\right)$$

Define

$$H = I_n - \frac{1}{n} 1.1'$$

where

$$1 = \underbrace{(1, \ldots, 1)'}_{n}$$

and considerer the iner product matrix

$$B = HAH$$

It is well–known (see, for example, Mardia *et al.*, 1979) that δ has an m–dimensional Euclidean representation, i.e., (E, δ) can be imbedded in (R^m, d), where d is the usual Euclidean distance, if B is semi positive–definite and m =rank (B).

The above Euclidean model identifies each object e_i of E as a point in R^m with coordinates $(x_{i1}, ..., x_{im})$ such that the columns of the $n \times m$ matrix $X = (x_{ij})$ are the eigenvectors of B. Thus

$$\left.\begin{array}{l} X.X' = B \\ X'.X = D \end{array}\right\} \qquad (1)$$

where D is a $m \times m$ diagonal matrix whose diagonal entries are the positive eigenvalues of B. The method based on this identification is known as metric multidimensional scaling,or principal coordinate analysis.

It is also well–known that if δ satisfies the so called ultrametric inequality

$$\delta_{ij} \leq \max(\delta_{ik}, \delta_{jk}) \qquad (i, j, k = 1, \ldots, n)$$

then E can be identified with an ultrametric tree, usually represented as a dendrogram. Moreover, in this case E can be analyzed via the hierarchical clustering scheme (Johnson, 1967), i. e. it is possible to construct a hierarchical structure $\mathcal{C}$ of nested clusters and a level function α, defined on clusters, which is closely related to δ.

Both models have been widely used and applied to the same data. Less frequent, however, are theoretical studies on the relationship between spatial and hierarchical models. Gower (1971) conjectured that an ultrametric distance is always Euclidean. This conjecture was proved by Holman (1972) who seems to be the first investigator who obtained theoretical results on the subject. Other proofs of Gower's conjecture have been obtained by Gower and Banfield (1975), and Cuadras and Carmona (1983). Ohsumi and Nakamura (1981) studied (with examples) a relationship between the eigenvalues of the inner product matrix B and the cluster formations on the hierarchical structure $(\mathcal{C}, \alpha)$ associated with and ultrametric distance δ. Carroll (1976) introduced tree structures as intermediate models between spatial models and the hierarchical clustering scheme, and described an algorithm to fit a tree structure to proximity data. Pruzansky *et al.* (1982) studied and compared two dimensional Euclidean planes and additive tree representations, and proposed two indices (skewness and elongation) of goodness–of–fit and some criteria for deciding which model is more appropiate. Critchley (1985) proves that a low dimensional scaling of a dendrogram may have a very good fit and gives some useful results to elucidate the Holman's theorem consequences.

2. EIGENANALYSIS OF ULTRAMETRIC PROXIMITY MATRICES

Let δ be an ultrametric distance defined on a finite set $E = \{e_1, \ldots, e_n\}$ and consider the $n \times n$ matrix $\Delta = (\delta_{ij})$ with $\delta_{ij} = \delta(e_i, e_j)(i, j = 1, \ldots, n)$. Most of the properties of the present study may be read off directly from the Δ matrix, but we prefer an alternative approach of first studying a proximity matrix, and then connecting the resulting consequences with the Δ matrix.

Let $P = (p_{ij})$ be a $n \times n$ symmetric proximity matrix, i.e. $p_{ij} = \pi(e_i, e_j) \geq 0$ satisfying:

$$p_{ii} \geq p_{ij} = p_{ji} \qquad (i, j = 1, \ldots, n)$$

If P also satisfies:

$$(2) \qquad p_{ii} = x \geq p_{ij} \geq \min(p_{ik}, p_{jk}) \quad (i, j, k = 1, \ldots, n)$$

then P is called a x–ultrametric proximity matrix, and is denoted by $P^{(x)}$. If in addition,

$$x > p_{ij} > 0 \qquad (i, j = 1, \ldots, n; i \neq j)$$

then $P^{(x)}$ is said to be a non–singular x–ultrametric proximity matrix.

Theorem 1. Let $P^{(x)}$ be a non–singular x–ultrametric proximity matrix, and let $y \geq 0$. Then:

a) $P^{(x)} + yI_n$ is a non–singular $(x + y)$ – ultrametric proximity matrix.

b) $P^{(x)}$ is positive–definite;

c) the smallest eigenvalue λ of $P^{(x)}$ is given by

$$\lambda = x - \max\{p_{ij} : i \neq j\}$$

Proof. To prove a), we observe that:

$$p_{ii} + y = x + y \geq x \geq \min(p_{ik}, p_{jk}) \quad (i, j, k = 1, \ldots, n)$$

and thus $P^{(x)} + yI_n = P^{(x+y)}$.

To prove part b), we show by induction that $\det\left(P^{(x)}\right) > 0$ and that all principal minors of $P^{(x)}$ are positive. Let $a = \max\{p_{ij} : i \neq j\}$. If $p_{ij} = p_{ji} = a$, then by (2)

$$p_{ik} \geq \min\{p_{ij}, p_{kj}\} = p_{kj} \qquad k \neq j$$

$$p_{jk} \geq \min\{p_{ji}, p_{ki}\} = p_{ki} \qquad k \neq i$$

and since $P^{(x)}$ is symmetric, we obtain

$$p_{ik} = p_{jk} \qquad k = 1, \ldots, n \qquad (k \neq i, j)$$

Thus, taking $x = a$, $P^{(a)}$ has two equal rows, and hence $\det\left(P^{(a)}\right) = 0$. For $n = 1$, $P^{(x)}$ is trivially positive definite. We assume that $P^{(x)}$ is positive definite for $n = k$. Consider the function $f : [0, \infty) \longrightarrow R$ defined by

$$f(z) = \det\left(P^{(a+z)}\right) = \det\left(P^{(a)} + zI_{k+1}\right)$$

(Notice that all $P^{(a+z)}$ matrices can be expressed in the form $P^{(a)} + zI_n$). Then

$$f(0) = 0$$

$$f'(z) = \sum_{i=1}^{k+1} \det\left(P_i^{(a)} + zI_k\right) = \sum_{i=1}^{k+1} \det\left(P_i^{(a+z)}\right)$$

where $P_i^{(a+z)}$ is the $k \times k$ matrix obtained after deleting the $i-th$ row and the $i-th$ column of $P^{(a+z)}$. Since $P_i^{(a+z)}$ is an $(a+z)$–ultrametric proximity matrix, the induction hypothesis gives

$$\det\left(P_i^{(a+z)}\right) > 0$$

whenever $z > 0$. Therefore $f'(z) > 0$ for $z < 0$ and $f(z)$ is a strictly increasing function. In particular, taking $z = x - a$, we have that $\det\left(P^{(x)}\right) > 0$ and since all principal minors of $P^{(x)}$ are positive, $P^{(x)}$ is positive definite. It follows, by continuity, that $P^{(a)}$ is semi positive–definite, and part b) follows.

We now prove c). Let u be an eigenvector of $P^{(a)}$ with eigenvalue α. Then

$$P^{(x)}u = \left(P^{(a)} + (x-a)I_n\right)u = (\alpha + (x-a))\,u,$$

and so u is also an eigenvector of $P^{(x)}$ with eigenvalue $\lambda = \alpha + (x-a)$. Since $P^{(a)}$ is positive semidefinite, 0 is its smallest eigenvalue and hence $\lambda = (x-a)$ is the smallest eigenvalue of $P^{(x)}$. This concludes the proof.

By reordering the finite set $E = \{e_1, \ldots, e_n\}$ into $E = \{e_{i_1}, \ldots e_{i_n}\}$, one is able to construct an adequate form of the matrix $P^{(x)}$ as follows:

Let $1 = (1, \ldots, 1)', J = 11'$ and define

$$\begin{aligned} a_1 &= \max\{p_{ij} : e_i, e_j \in E, \quad i \neq j\} \\ E_1 &= \left\{e_{i_1}, \ldots, e_{i_{n_1}} : p_{i_h i_k} = a_1, \quad i_h \neq i_k\right\} \end{aligned} \tag{3}$$

Then $|E_1| = n_1 > 1$. The sub–matrix P_1 of $P^{(x)}$ corresponding to E_1 is the $n_1 \times n_1$ matrix defined by $P_1 = a_1 J + (x - a_1)\, I$, where the matrices J and I have the appropiate order.

In the same way, we define

$$\begin{aligned} a_2 &= \max\{p_{ij} : e_i, e_j \in E - E_1, \qquad i \neq j\} \\ E_2 &= \left\{e_{i_{n_1+1}}, \ldots, e_{i_{n_1+n_2}} : p_{i_h i_k} = a_2, i_h \neq i_k,\right. \\ &\qquad \left.\text{and } p_{i_h j} = p_{i_k j} \text{ for all } j \neq i_{n_1+1}, \ldots, i_{n_1+n_2}\right\} \end{aligned} \tag{4}$$

Then $|E_2| = n_2 \geq 1$. Proceeding in this way we obtain a partition of E

$$(5) \qquad E = E_1 + \cdots + E_r + E_{r+1} + \cdots + E_m \qquad E_i \cap E_j = \phi \quad i \neq j$$

where

$$|E_i| = n_i > 1 \qquad \text{if } 1 \leq i \leq r,$$

$$|E_i| = 1 \qquad \text{if } i > r.$$

This partition is formed by r maximal clusters of equidistan objects, in relation to $P^{(x)}$, and $(m-r)$ isolated objects. Note that $r \geq 1$ and $r = m$ are possible.

Furthermore

$$p_{ij} = p_{kj} \qquad \text{if} \qquad e_i, e_k \in E_q \; e_j \notin E_q,$$

i.e., all the elements of E_q are equidistant to the other elements of E. Notice that the equivalence relation $\sim$ on E defined by $e_i \sim e_k$ if and only if $p_{ij} = p_{kj}$ for all $e_j \in E, j \neq i, k$, gives the partition (5), i. e. the E_q are the equivalence classes.

Henceforth we assume that E has been arranged in accordance with (3), (4) and the partition (5). One may refer to this arrangement as the arrangement of E induced by $P^{(x)}$.

Theorem 2. Let $P^{(x)}$ be a x–ultrametric no singular proximity matrix. Then

$$\lambda_1 = x - a_1 \leq \lambda_2 = x - a_2 \leq \cdots \leq \lambda_r = x - a_r$$

are the eigenvalues of $P^{(x)}$.The multiplicity of λ_i is $(n_i - 1)$ and the corresponding eigenspace is generated by the $n_i - 1$ orthogonal eigenvectors:

$$u_1^{(i)} = (\underline{0}; 1, -1, 0, ..., 0; \bar{0})', u_2^{(i)} = (\underline{0}, 1, 1, -2, 0, ..., 0; \bar{0})'$$
$$, ..., u_{n_i-1}^{(i)} = (\underline{0}; 1, ..., 1, -(n_i - 1); \bar{0})'$$

where $\underline{0}$ and $\bar{0}$ denote null vectors of orders $\sum_{j=1}^{i-1} n_j$ and $n - \sum_{j=1}^{i} n_j$,

respectively.

Proof. The sub–matrix of $P^{(x)}$ corresponding to E_i is $P_i = a_i J + (x - a_i)I$. Let

$$v_1^{(i)} = (1, -1, 0, \ldots, 0)', \ldots, v_{n_i-1} = (1, \ldots, 1, -(n_i - 1))'$$

Then

$$P_i v_j^{(i)} = (x - a_i)\, v_j^{(i)} \qquad j = 1, \ldots, n_i - 1,$$

and so $v_j^{(i)}$ is an eigenvector of P_i. Then it is easy to see that $u_1^{(i)}, \ldots, u_{n_i-1}^{(i)}$ are orthogonal eigenvectors of $P^{(x)}$ whose eigenvalue is $(x - a_i)$. Moreover, $\lambda_1 = x - a_1$ is the smallest eigenvalue of $P^{(x)}$, by Theorem 1, and the proof is complete.

3. EIGENANALYSIS OF INNER PRODUCT MATRIX FOR AN ULTRAMETRIC DISTANCE

Let $\Delta = (\delta_{ij})$ be the previously considered matrix determined by an ultrametric distance δ on E. Suppose that $\delta_{ij} > 0$ for all $i \neq j$. Define the x–ultrametric proximity matrix $P^{(x)} = (p_{ij})$, where

$$(6) \qquad p_{ii} = x \qquad p_{ij} = p_{ji} = x - 1/2.\delta_{ij}^2 > 0 \qquad i \neq j$$

and suppose E is with the arrangement induced by $P^{(x)}$. Consider also the partition (5) and let h_q be the common distance between all the pairs of different objects in E_q,

$$(7) \qquad h_q = \delta_{ij} \qquad e_i,\, e_j \in E_q \qquad i \neq j \qquad 1 \leq q \leq r$$

The next theorem points out certain properties of the eigenvalues and eigenvectors of the inner product matrix associated with Δ. Later on, we discuss some geometric interpretations of this theorem.

Theorem 3. Let $\Delta = (\delta_{ij})$ be an ultrametric distance matrix and let $B = HAH$, where $A = \left(-1/2\delta_{ij}^2\right)$, $H = I_n - \frac{1}{n}.J$, $J = 11'$ and $1 = (1, ..., 1)'$. Let μ be the largest eivenvalue of B. Then:

$$(8) \qquad \text{a)} \quad \lambda_1 = \tfrac{1}{2}h_1^2 \leq \ldots \leq \lambda_r = \tfrac{1}{2}h_r^2 \leq \mu$$

are the eigenvalues of B. The multiplicity of λ_q is $(n_q - 1)$ and the corresponding orthogonal eigenvectors are the same as those of Theorema 2. Furthermore, λ_1 is the smallest eigenvalue of B.

b) $\mu = \lambda_r$ if and only if $\delta_{ij} = \delta_{ik} = \delta_{jk}$ for every e_i, e_j, e_k, $\in E$, i.e. all elements of E are equidistant.

Proof. Considerer the x–ultrametric proximity matrix $P^{(x)}$ defined in (6). Then $\lambda_1 = x - a_1$, where a_1 has been defined in (3), is the smallest eigenvalue of $P^{(x)}$ (Theorem 1). Moreover since $P^{(x)} = xJ + A$, then

$$B = HAH = HP^{(x)}H \qquad \text{and} \quad \text{rank}(B) = n - 1,$$

because $HJ = 0$ and $\text{rank}\left(P^{(x)}\right) = n$.

Furthermore,

$$Bu_i^{(1)} = \lambda_1 u_i^{(1)} \qquad i = 1, ..., n_1 - 1$$

where $u_i^{(1)}$ is defined as in Theorem 2, because $Hu_i^{(1)} = u_i^{(1)}$. Thus λ_1 is an eigenvalue of B. In a similar manner, one can prove that $\lambda_2 \leq ... \leq \lambda_r$ are eigenvalues of B with eigenvectors given by Theorem 2. Since $P^{(x)}$ is positive definite, we have that $P^{(x)} = TT'$, $\text{rank}(T) = n$, for some matrix T. It follows that $P^{(x)}$ and $Q = T'T$ have the same eigenvalues. Similarly, $S = T'HHT = T'HT$ and $B = HP^{(x)}H = HTT'H$ have the same eigenvalues, and

$$S = T'T - \frac{1}{n}T'JT = Q - C,$$

with $\text{rank}(S) = n - 1$ and $\text{rank}(C) = 1$. Thus, $\mu_1(C) > 0$ and $\mu_i(C) = 0$ if $i > 1$.

On the other hand, let $\mu_1(M) \geq ... \geq \mu_n(M)$ be the eigenvalues of any symetric positive semidefinite matrix M. Then (Okamoto, 1969), if N and $M - N$ are semi positive–definite matrices, and $\text{rank}(N) = k \leq n$, the following inequalities are satisfied:

$$\mu_i(M - N) \geq \mu_{k+i}(M) \tag{9}$$

where $\mu_j(M) = 0$ for all $j > n$.

Therefore, from (9), we obtain:

$$\mu_1(Q-C) = \mu_1(B) \geq \mu_2(Q) = \mu_2\left(P^{(x)}\right)$$

Moreover, $\mu_1\left(P^{(x)}\right) > \lambda_r$, because $P^{(x)}$ is positive definite and an eigenvector of the largest eigenvalue must have positive components (Perron's theorem). It follows that $\mu_2\left(P^{(x)}\right) \geq \lambda_r$ and thus $\mu = \mu_1(B) \geq \lambda_r$ and (8) follows.

We now let $a_1 = \max\left(p_{ij}\right) = x - \min_{i \neq j}\left(1/2.\delta_{ij}^2\right) = x - \lambda_1$. By Theorem 1, we have $\mu_n\left(P^{(x)}\right) = \lambda_1$, and taking again (9) into account, we obtain

$$\mu_1(B) \geq \ldots \geq \mu_{n-1}(B) \geq \mu_n\left(P^{(x)}\right) = \lambda_1 > \mu_n(B) = 0$$

But λ_1 is also an eigenvalue of B, so $\mu_{n-1}(B) = \lambda_1 = \frac{1}{2}h_1^2 > 0$ is the smallest eigenvalue of B, and part a) follows.

Finally, suppose $\mu = \lambda_r$. Then $\mu_1(B) = \mu_2(P)$, i.e., $\mu_1(Q-C) = \mu_2(Q)$. This implies that $C = \mu_1(Q).v_1v_1'$, where v_1 is an eigenvector corresponding to the largest eivenvalue of Q with $||v_1|| = 1$ (Okamoto, 1969). Then u_1 defined by $v_1 = T'u_1$, is an eigenvector of the largest eigenvalue of $P^{(x)}$. It follows that

$$\frac{1}{n}T'11'T = \mu_1(Q)T'u_1u_1'T$$

and so $u_1 = n^{-1/2}1$. Therefore, 1 is an eigenvector of the ultrametric proximity matrix $P^{(x)}$ and it is easy to see that $p_{ij} = p_{ik} = p_{jk}$ for all $i \neq j \neq k$. It follows from (6) that $\delta_{ij} = \delta_{ik} = \delta_{jk}$. Then $E = E_1$, $n = n_1$. This concludes the proof.

Corollary. A non singular ultrametric distance δ for a finite set E with n objects has a $n-1$ dimensional Euclidean representation.

Proof. This follows since B is semi positive-definite and $\text{rank}(B) = n-1$.

The smallest eigenvalue λ_1 has been obtained by Cuadras (1981). See also Ohsumi and Nakamura (1981), Cuadras and Carmona (1983) and Cuadras (1983).

It is not feasible to obtain all the eigenvalues and eigenvectors of B, but through Theorema 3 it is posible to give some partial results. Let $\mu_1 > \mu_2 \geq \mu_3 \geq \ldots$ be the eigenvalues not equal to $\lambda_r \geq \ldots \geq \lambda_1$. Except for the greatest eigenvalue μ_1, every other eigenvalue can be greater or less than λ_i $(\lambda_i \neq \lambda_1)$.

In some special cases, the eigenvalues μ_i can be obtained explicitly. For instance, suppose:

$$|E_b| = |E_{b+1}| = ... = |E_c| = n_b$$
$$H_b = \delta_{ij} = \delta_{ik} = \delta_{jk} \quad \text{if} \quad e_i \in E_{t_1},\ e_j \in E_{t_2},\ e_k \in E_{t_3}$$
$$b \leq t_1 < t_2 < t_3 \leq c$$

i.e., $E_b, ..., E_c$ ere equidistant clusters. Then it is possible to prove that:

$$\tilde{\mu} = 1/2\left(n_b H_b^2 - (n_b - 1)\, h_b^2\right) \tag{10}$$

is an eigenvalue of B of multiplicity $(c-b)-1$. The corresponding eigenvectors are:

$$\left(\underline{0}; 1, \overset{(n_b}{\ldots}, 1, -1, \overset{(n_b}{\ldots}, -1, 0, ..., 0; \bar{0}\right),\ \left(\underline{0}; 1, \overset{(n_b}{\ldots}, 1, 1, \overset{(n_b}{\ldots} 1, -2, \overset{(n_b}{\ldots}, -2, 0, ..., 0; \bar{0}\right), ...$$

i.e., with similar structure of the eigenvectors obtained in Theorem 2.

We shall now describe a result concerning the variation of μ_i after a perturbation of λ_i. We know (see, for exemple, Mardia *at al.*, 1979) that

$$\sum_{i<j} \delta_{ij}^2 = n\left(\sum \mu_k + \sum \lambda_i\right)$$

For the same reason, we have from (7)

$$\sum_{q=1}^{r} n_q\,(n_q - 1)\, h_q^2/2 + D = n\left[\sum \mu_k + \sum_{q=1}^{r} (n_q - 1)\, h_q^2/2\right] \tag{11}$$

where $n_q\,(n_q - 1)\, h_q^2/2$ is the sum of the intradistances between all the pairs of different objects in E_q, and D is the sum of the interdistances between objects of different clusters.

Suppose that δ_{ij} is perturbed within the cluster E_q

$$\tilde{\delta}_{ij}^2 = \delta_{ij}^2 + \varepsilon_q = h_q^2 + \varepsilon_q \qquad e_i,\ e_j \in E_q \qquad 1 \leq q \leq r$$

but with the condition that the interdistances between different clusters will not be altered, i.e., $\tilde{\delta}_{ij} = \delta_{ij}$, for $e_i \in E_{q_1},\ e_j \in E_{q_2},\ q_1 \neq q_2$.

Then D does not vary and we have

$$(12)\quad \sum_{q=1}^{r} n_q (n_q - 1)\left(h_q^2 + \varepsilon_q\right)/2 + D = n\left[\sum \tilde{\mu}_k + \sum_{q=1}^{r} (n_q - 1)\left(h_q^2 + \varepsilon_q\right)/2\right]$$

where $\tilde{\mu}_k$ are the corresponding eigenvalues resulting from $\left(\tilde{\delta}_{ij}\right)$. From (11) and (12) we deduce that

$$(13)\qquad \sum \tilde{\mu}_k = \sum \mu_k - \sum_{q=1}^{r} (n - n_q)(n_q - 1).\varepsilon_q/2n$$

The last formula gives the variation of the eigenvalues μ_k which decrease if the h_q distances are increased. It can be also proved that the order of the eigenvalues of B can be modified if we perform a monotone transformation to δ_{ij}. Then a clear relation does not exist between the order of the eigenvalues and the hierarchical clustering scheme constructed using δ_{ij} (see the exemple below).

4. METRIC MULTIDIMENSIONAL SCALING ON AN ULTRAMETRIC DISTANCE

We shall exhibit some geometrical interpretations on the principal coordinates deduced from a ultrametric distance δ_{ij} on a set E. Let us suppose E is ordered according to the ordering induced by $\Delta = (\delta_{ij})$.

Let X be the $n \times (n-1)$ matrix satisfying (1). Let us arrange X as follows:

$$(14)\qquad X = (X_0, X_1, ..., X_r)$$

where X_q is the $n \times (n_q - 1)$ matrix with the principal coordinates associated with the $n_q - 1$ eigenvectors with eigenvalue $\lambda_q = \frac{1}{2}h_q^2\,(1 \leq q \leq r)$, and X_0 contains the remaing coordinates of X. Through each $X_t (0 \leq t \leq r)$ we can project E in a $(n_t - 1)$–dimensional Euclidean space with a Euclidean distance $d_t(i, j)$, resulting in

$$n_t = |E_t| \qquad t = 1, ..., r$$

$$n_0 = n - 1 - \sum_{t=1}^{m} (n_t - 1) = m - 1$$

The next theorem describes this reduction in the dimensionality.

Theorem 4. Let $\Delta = (\delta_{ij})$ be an ultrametric distance on E, and assume that E is ordered according to the ordering induced by Δ. Considerer the partition $E_1, ..., E_m$ and the principal coordinates (14). Let $d_t(e_i, e_j)$ be the Euclidean distance obtained by using only the principal coordinates $X_t (t = 0, 1, ..., r)$.

$$
\begin{array}{lllll}
\text{a)} & d_q(e_i, e_j) = h_q > 0 & \text{if} & e_i, e_j \in E_q,\ e_i \neq e_j, & 1 \leq q \leq r, \\
 & d_q(e_i, e_j) = 0 & \text{if} & e_i, e_j \notin E_q, & 1 \leq q \leq m, \\
 & \text{where } h_q^2 = 2\lambda_q. & & & \\
\text{b)} & d_0(e_i, e_j) = 0 & \text{if} & e_i, e_j \in E_q, & 1 \leq q \leq m, \\
 & d_0(e_i, e_j) > 0 & \text{if} & e_i \in E_p,\ e_j \in E_q, & 1 \leq p \neq q \leq m.
\end{array}
$$

Proof: As a consequence of Theorem 3, for $1 \leq q \leq r$, the columns of X_q are

$$(\lambda_q/2)^{1/2} . (\underline{0}; 1, -1, 0, ..., 0; \bar{0})', ..., (\lambda_q/\beta)^{1/2} . (\underline{0}; 1, ..., 1, -(n_q - 1); \bar{0})'$$

where $\beta = n_q^2 - n_q$. We observe that the square module of each vector is λ_q, and that the Euclidean distance, using non null rows, of X_q is

$$d_q(e_i, e_j) = (2\lambda_q)^{1/2} = h_q$$

Obviously the other distances are zero, and a) follows.

As the columns of X_0 are orthogonal to the columns of each $X_q (1 \leq q \leq r)$, X_0 must have its first n_1 rows equal to a vector row $(\alpha_1, \alpha_2, ...)$, its following n_2 vector rows equal to a second row $(\beta_1, \beta_2, ...)$, etc. Then $d_0(e_i, e_j) = 0$ if $e_i, e_j \in E_q$.

Let us now show that $\alpha_t \neq \beta_t$ for some t. If we were to suppose the opposite, then we would have

$$\delta_{ij} = \left[\left(\frac{n_1 - 1}{n_1} \right) \alpha_1 + \left(\frac{n_2 - 1}{n_2} \right) \alpha_2 \right]^{1/2} < (\alpha_1 + \alpha_2)^{1/2} \leq (2\lambda_2) = h_2 \quad \begin{array}{l} e_i \in E_1 \\ e_j \in E_2 \end{array}$$

i.e., the inter–cluster distance δ_{ij} would be less than the intra–cluster distance h_2, which is incompatible with the ultrametric property.

Generally, we can affirm that given two different clusters E_p and E_q, n_p rows of X_0 that describe E_p, are equal between them but different from n_q rows that describe E_q. Then $d_0(e_i, e_j) \neq 0$ for $e_i \in E_p$, $e_j \in E_q$. And this concludes the proof.

We conclude this section with some results concerning the Euclidean distance d_0, its relation to δ_{ij} and the representation of clustering $C_1 =< E_1, ..., E_m >$.

Since δ_{ij} is ultrametric.

$$\delta_{pq}^{(1)} = \delta_{ij} \qquad \text{if} \quad e_i \in E_p\,,\ e_j \in E_q\,,\ p \neq q\,,$$

is also an ultrametric distance on C_1. Thus we can represent C_1 by using a dendrogram. Moreover, we can apply the previous results to the ultrametric distance $\delta^{(1)}$ on C_1.

Thus grouping equidistant clusters in C_1, we can define a new clustering C_2, and applying Theorem 3, we can obtain certain eigenvalues $\lambda_1^{(1)}, ..., \lambda_{r_1}^{(1)}$. Calculated directly from $\delta^{(1)}$. Repeating this operation as many times as necessary, we can build a hierarchy of clusterings

$$C_0, C_1, C_2, ..., C_k$$

being $C_0 =< \{e_1\}, \{e_n\}, ..., \{e_n\} >, C_k = E$. The clustering C_{i+1} is built by gathering equidistant clusters in C_i. Each one of the clusters of the hierarchical structure $\mathcal{C}$ obtained by applying the hierarchical clustering scheme on the initial ultrametric distance δ, belongs to some of the cluster C_i. Theorem 3 tells us that starting from the distances $\delta, \delta^{(1)}, ...$, on $C_0, C_1, ...$, we can build a succession of eigenvalues:

$$(15) \quad \lambda_1 \leq ... \leq \lambda_r \leq \lambda_1^{(1)} \leq ... \leq \lambda_{r_1}^{(1)} \leq ... \leq \lambda_1^{(k-2)} \leq ... \leq \lambda_{r_{k-2}}^{(k-2)} \leq \lambda^{(k-1)}$$

Each eigenvalue (15) can be calculated directly from the original distance δ. In particular,

$$\lambda_1 = \min \{\frac{1}{2}.\delta_{ij}^2 : e_i \neq e_j \in E\} \qquad \lambda^{(k-1)} = \max \{\tfrac{1}{2}.\delta_{ij}^2 : e_i, e_j \in E\}$$

Let us note that the number of independent eigenvalues in (15) is equal to the number of nodes in the dendrogram representing $\mathcal{C}$, that is to say, to the number of clusters in $\mathcal{C}$ different from $\{e_1\}, ..., \{e_n\}$.

As a consequence of Theorem 4 the X_0–coordinates allow a $(m-1)$ dimensional Euclidean representation of C_1. Let us relate the Euclidean distance $d_0(e_i, e_j)$ with $\delta^{(1)}$. It is easy to prove that

$$(16) \qquad d_0^2(e_i, e_j) = \left(\delta_{pq}^1\right)^2 - \left[(n_p - 1)/n_p\right]\lambda_p - \left[(n_q - 1)/n_q\right]\lambda_q$$

where λ_p is defined in Theorem 3 for $n_p > 1$. If $n_p = 1$, we take an arbitrary value for λ_p in (16). Although the Euclidean distance $d_0^{(1)}$, defined on E, that is to say, $d_0^{(1)}(E_p, E_q) = d_0(e_i, e_j)$ if $e_i \in E_p$, $e_j \in E_q$, is not ultrametric, (16) assures us that $\left(d_0^{(1)}\right)^2$ satisfies the additive inequality (or the condition of the four points). Then C_1, with the metric $\left(d_0^{(1)}\right)^2$, can be represented using an additive tree, (see Sattah and Tversky, 1977).

Finally, let us suppose that $E_b, E_{b+1}, ..., E_c$ are equidistant clusters with a common distance H_b and such that $|E_b| = ... = |E_c|$. Then there exist, and can be calculated, some eigenvalues $\tilde{\mu}$ (see(10)) of multiplicity $(c-b)-1$.

Let $X_0 = (X_\alpha, X_{\tilde{\mu}})$ be, where $X_{\tilde{\mu}}$ contains the principal coordinates associated with $\tilde{\mu}$, and X_α contains the remaining coordinates. Then it can be proved that $\mu > \tilde{\mu}$ and that

$$(17) \quad \begin{aligned} &d_\alpha(e_i, e_j) = 0 \quad d_{\tilde{\mu}}(e_i, e_j) = H_b \quad \text{if} \quad e_i \in E_p,\ e_j \in E_q \quad b \le p \ne q \le c \\ &d_\alpha(e_i, e_j) = d_{\tilde{\mu}}(e_i, e_j) = 0 \qquad\qquad \text{if} \quad e_i, e_j \in E_q, \quad b \le q \le c. \end{aligned}$$

A consequence of (17) is that the α–coordinates may be inadequate for "embedding" a hierarchical clustering scheme in a Euclidean space. For example, in a representation in a plane, instead of using the principal coordinates associated with the eigenvalues μ_1, μ_2, it is more convenient to use the coordinates associated with μ_1, $\tilde{\mu}$, where $\tilde{\mu}$ could be the largest eigenvalue satisfying (10).

An example. Frömmel and Holzhütter (1985) present a method for estimating quantitatively the influence of point mutations and selection on the frequencies of codons and amino acids. They study an asymmetric mutation matrix $M = (m_{ij})$, where m_{ij} is the probability of an amino acid replacement $A_i \longrightarrow A_j$ caused by a point mutation occurring per time unit. For a more systematic classification of amino acids, they introduce a genetic distance d_{ij} between two amino acids A_i and A_j. The distance d_{ij} is proportional to the inverse probability of mutual replacement. Then a cluster analysis is performed by means of the unweighted pair–group method. They obtain a dendrogram

defining two main groups representing the hydrophobic residues and the polar ones.

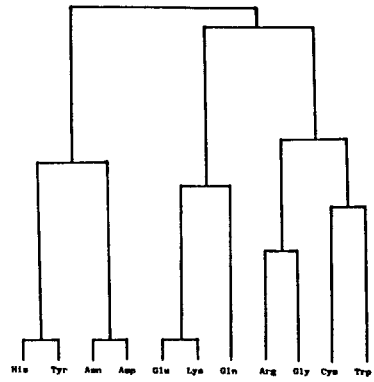

FIGURE 1. Dendrogram obtained using a genetic distance according to Frömmel and Holzhütter (1985).

We use in this exemple the 11 amino acids (histidine, tyrosine, asparagine, aspastic acid, glutamatic acid, lysine, glutamine, arginine, glycine, cysteine and tryptophan) belonging to the polar group. Figure 1 represents the corresponding dendrogram. Table 1 contains the ultrametric distance computed from the dendrogram but making the minimum distance equal to 1. Table 2 contains the related principal coordinates and eigenvalues.

TABLE 1.

Distances among the 11 amino acids belonging to the polar group.

	His	Tyr	Asn	Asp	Glu	Lys	Gln	Arg	Cly	Cys	Trp
His	0										
Tyr	1	0									
Asn	10.7	10.7	0								
Asp	10.7	10.7	1	0							
Glu	19.1	19.1	19.1	19.1	0						
Lys	19.1	19.1	19.1	19.1	1	0					
Gln	19.1	19.1	19.1	19.1	9.6	9.6	0				
Arg	19.1	19.1	19.1	19.1	18	18	18	0			
Cly	19.1	19.1	19.1	19.1	18	18	18	6	0		
Cys	19.1	19.1	19.1	19.1	18	18	18	12	12	0	
Trp	19.1	19.1	19.1	19.1	18	18	18	12	12	8.4	0

TABLE 2.

Principal coordinates obtained from the distance matrix presented in Table 1.

	μ_1	μ_2	μ_3	$\mu_4 = \tilde{\mu}$	μ_5	λ_5	λ_4	λ_3	λ_2	λ_1
	610.03	446.28	117.23	114	60.44	35.28	18	0.5	0.5	0.5
His	9.8	−0.803	−0.032	−5.34	0.05	0	0	0.5	0	0
Tyr	9.8	−0.803	−0.032	−5.34	0.05	0	0	−0.5	0	0
Asn	9.8	−0.803	−0.032	−5.34	0.05	0	0	0	0.5	0
Asp	9.8	−0.803	−0.032	−5.34	0.05	0	0	0	−0.5	0
Glu	−4.7	10.00	−0.076	0	3.00	0	0	0	0	0.5
Lys	−4.7	10.00	−0.076	0	3.00	0	0	0	0	−0.5
Gln	−4.31	8.86	−0.027	0	−6.52	0	0	0	0	0
Arg	−6.49	−6.58	−5.33	0	0.07	0	3	0	0	0
Cly	−6.49	−6.58	−5.33	0	0.07	0	−3	0	0	0
Cys	−6.26	−6.25	5.50	0	0.09	4.2	0	0	0	0
Trp	−6.26	−6.25	5.50	0	0.09	−4.2	0	0	0	0

The obtained theoretical results may be illustrated as follows:

1) The partition (5) is

(18)

$$E = \{\text{His, Tyr}\} + \{\text{Asn, Asp}\} + \{\text{Glu, Lys}\} + \{\text{Arg, Gly}\} + \{\text{Cys, Trp}\} + \{\text{Gln}\}$$

Thus $r = 5$, $m = 6$ and

$$n_1 = \ldots = n_5 = 2, \quad n_6 = 1$$

Moreover

$$h_1 = h_2 = h_3 = 1 < h_4 = 6 < h_5 = 8.4,$$
$$\tilde{\mu} = \left[2 \times (10.7)^2 - (2-1) \times 1^2\right]/2 = 114$$

Looking at Table 2, we observe that theorem 3 and 4 are satisfied and the $\tilde{\mu}$–coordinates satisfy (10).

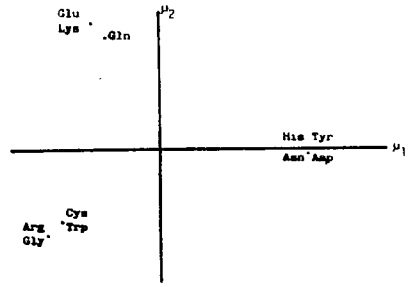

FIGURE 2. Two dimensional representation of dendrogram displayed in Figure 1 using the first and second principal axes.

2) Figure 2 is a two–dimensional representation of the 11–aminoacids using the first and second principal coordinates. Note that His, Tyr, Asn and Asp lie on the same point, so the partition (18) is not well represented.

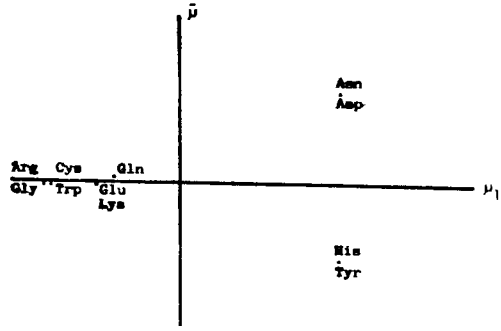

FIGURE 3. Two dimensional representation of dendrogram displayed in Figure 1 using the first and fourth principal axes.

3) Figure 3 is the representation using the first and fourth principal coordinates. As {His, Tyr} and {Asn, Asp} are locate separately, this representation is more suitable than the previous one.

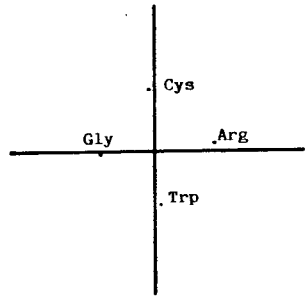

FIGURE 4. Two dimensional representation of amino acids inside the clusters Arg, Gly and Cys, Trp.

4) The amino acids inside every cluster are described through the principal coordinates related to λ–eigenvalues. For exemple, Figure 4 contains the representation of the amino acids {Arg, Gly}, {Cys, Trp}. On the other hand, note that we can represent Gln only through μ–coordinates.

5) By gathering equidistant clusters we obtain the following clusterings:

$C_0 = \{[\text{His}], [\text{Tyr}], [\text{Asn}], [\text{Asp}], ..., [\text{Cys}], [\text{Trp}]\}$
$C_1 = \{[\text{His, Tyr}], [\text{Asn, Asp}], [\text{Glu, Lys}], [\text{Gln}], [\text{Arg, Gly}], [\text{Cys, Trp}]\}$
$C_2 = \{[\text{His, Tyr, Asn, Asp}], [\text{Glu, Lys, Gln}], [\text{Arg, Gly, Cys, Trp}]\}$
$C_3 = \{[\text{His, Tyr, Asn, Asp}], [\text{Glu, Lys, Gln, Arg, Gly, Cys, Trp}]\}$
$C_4 = \{[\text{His, Tyr, Asn, Asp,..., Cys, Trp}]\}$

The succession of eigenvalues (15) is

$$\lambda_1 = \lambda_2 = \lambda_3 = 0.5 < \lambda_4 = 18 < \lambda_5 = 35.28$$
$$< \lambda_1^{(1)} = 46 < \lambda_2^{(1)} = 57.24 < \lambda_3^{(1)} = 72 < \lambda_1^{(2)} = 162 < \lambda^{(3)} = 182.4$$

Note that these eigenvalues can be computed directly from the ultrametric distance matrix on Table 1.

6) The X_0 coordinates allow a five dimensional representation of C_1. The Euclidean distance d_0 verifies (16), i.e. d_0^2 satisfy the additive inequality. Figure 5 is an additive tree in parallel form, representing C_1 with the metric $\left(d_0^{(1)}\right)^2$.

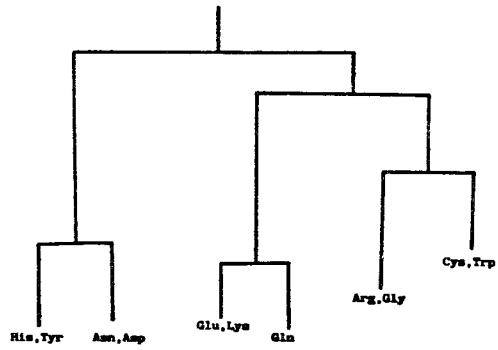

FIGURE 5. Additive tree in parellel form representing the clustering C_1.

7) We are also able to represent C_2, C_3 and C_4 using the coordinates related to the eigenvalues $\lambda_1^{(1)}$, $\lambda_2^{(1)}$, $\lambda_3^{(1)}$, $\lambda_1^{(2)}$, $\lambda^{(3)}$. As C_2 contains only three clusters, the graphical display is omitted in this case.

8) Finally, we now consider a monotone transformation of the distance on Table 1. We put 17.7 instead of 10.7 and 17 instead of 12. Then the hierarchical clustering scheme does not vary, but the new eigenvalues are

μ_1	μ_2	$\mu_3 = \tilde{\mu}$	μ_4	λ_5	λ_4	...,
481.8	378	278.38	261.96	35.28	18	...,

i.e., the order is different from Table 2.

5. SUMMARY AND CONCLUSSIONS

In this paper we investigate some geometric and algebraic properties on representations of ultrametric distances using metric multidimensional scaling or principal coordinate analysis. We also obtain some results about eigenstructure of the inner product matrix for an ultrametric distance on a finite set E with n objects.

The main conclussions obtained in this paper are:

a) The eigenanalysis on $\Delta = (\delta_{ij})$ can be done directly or on a matrix of proximities P, relating P and Δ using (6). The matrix P has the advantage that is positive definite.

b) Let B be the inner product matrix obtained from Δ. There exist certain maximal clusters $E_1, ..., E_m$ that define a partition in E. Each E_i contains equidistant objects. The clusters such that $|E_i| = n_i > 1$ are related to certain eigenvalues λ_i of B that can be obtained directly from Δ. In some special cases other eigenvalues, using elemental expressions, can be obtained. However, it is not possible to obtain a global solution relating to the eigenstructure of B.

c) Let us consider the spectral descomposition $B = \Gamma D \Gamma' = XX'$. We can express the main coordinates as $X = (X_0, X_r, ..., X_1)$, where $X_i (i > 0)$ contains the eigenvectors with eigenvalue λ_i. Then X_0 allows Euclidean representation of $E_1, ...E_m$, while each $X_i (i = 1, ...r)$ allows an Euclidean representation of those E_i such that $|E_i| > 1$. If $|E_i| = 1$, the Euclidean representation of E_i is obtained through the X_0–coordinates. If $|E_i| > 1$, the objects contained in E_i, do not remain separated if we project through the X_j–coordinates $(j \neq i)$. Moreover, the coordinates $X_0, X_1, ...X_r$ define orthogonal spaces.

d) Even though the number of Euclidean dimensions needed to represent E is $(n - 1)$ (Holman, 1972), in fact it would be enough to project an $(m - 1)$ dimensional space.

e) The Euclidean distance d_0 defined through X_0 can be expressed through δ_{ij} and λ_k. Moreover d_0^2 satisfies the additive inequality, or the condition of the four points.

f) The perturbations of the eigenvalues of B after certain monotonic transformations of δ_{ij} has been studied, and some conclusions have been obtained

about the order of the eigenvalues and the hierarchical clustering scheme constructed on E.

REFERENCES

- **Carroll, J.D.**"Spatial, non–spatial and hybrid models for scaling". Psychometrika, 1976, 41, 439-463.
- **Critchley, F.** "Dimensionality theorems in multidimensional scaling and hierarchical cluster analysis". Fourth Int. Symp. Data Analysis and Informatics, Vol. 1, Versailles: Inst. Nat. de Recherche en Inform. et en Autom., 85-110, 1985.
- **Cuadras, C.M.** "Métodos de Análisis Multivariante". Barcelona, Eunibar, 1981.
- **Cuadras, C.M.** "Análisis algebraico sobre distancias ultramétricas". ISI, 44th Session, Contributed Papers, 1983, 2, 554-557.
- **Cuadras, C.M.** and **Carmona, F.** "Euclidean dimensionality of ultrametric distances" (in catalan). Qüestiió, 1983, 7, 353-358.
- **Frömmel, C.** and Holzhütter, H.G. "An stimate of the effect of point mutation and natural selection on the rate of amino acid replacement in proteins". J. of Molecular Evolution, 1985, 21, 233-257.
- **Gower, J.C.** "Statistical methods of comparing different multivariate analysis of the same data." In F.R. Hodson, D.G. Kendall and P. Tautu (Eds.), Mathematics in Archaelogical and Historical Sciences. Edinburgh: Edinburgh University Press, 1971.
- **Gower, J.C.** and **Banfield, C.F.** Goodness–of–fit criteria for hierarchical classification and their empirical distributions. In L.C.A. Corsten and T. Postelnicu, (Eds.), Proceed. 8 th. Intern. Biometric Conference. Constanta: Ed., Acad. Rep. Soc. Romania, 1975.
- **Holman, E.W.** "The relation between hierarchical and Euclidean models for psychological distances". Psychometrika, 1972, 37, 417-423.
- **Johnson, S.C.** "Hierarchical clustering schemes". Psychometrika, 1967, 32, 241-254.
- **Mardia, K.V.**; **Kent, J.T.** and **Bibby, J.M.** "Multivariate Analysis". London: Academic Press, 1979.
- **Ohsumi, N.** and **Nakamura, T.** "Some properties of monotone hierarchical dendrogram in numerical classification" (in japanese). Proc. Inst. Statist. Mathem. (Tokyo), 1981, 28, 117-133.

- **Okamoto, M.** "Optimality of principal components". In P.R. Krishnaiah (Ed.), Multivariate Analysis–II. New York: Academic Press, 1969.
- **Pruzansky, S.; Tversky, A.** and **Carroll, J.D.** "Spatial versus tree representations of proximity data". Psychometrika, 1982, 47, 3-24.
- **Sattah, S.** and **Tversky, A.** "Additive similarity trees". Psychometrika, 1977, 42, 319-345.