From Data to the Physics Using Ultrametrics: New Results in High Dimensional Data Analysis

Fionn Murtagh

Department of Computer Science
Royal Holloway, University of London
Egham TW20 0EX, England
fmurtagh@acm.org

Abstract. We begin with pervasive ultrametricity due to high dimensionality and/or spatial sparsity. How extent or degree of ultrametricity can be quantified leads us to the discussion of varied practical cases when ultrametricity can be partially or locally present in data. We show how the ultrametricity can be assessed in text or document collections, in time series signals, and in other areas. We conclude with a discussion of ultrametricity in astrophysics, relating to observational cosmology.

1. INTRODUCTION

P-adic algebraic or ultrametric topological representation is important for observational data analysis because of the insights and benefits that such an approach provides.

Firstly ultrametricity is a pervasive property of observational data. It arises as a limit case when data dimensionality or sparsity grows. More strictly such a limit case can be a regular lattice structure. Ultrametricity is one possible representation for it. Notwithstanding alternative representations, ultrametricity offers computational efficiency (related to tree depth/height being logarithmic in number of terminal nodes), linkage with dynamical or related functional properties (phylogenetic interpretation), and well understood processing tools based on well known p-adic or ultrametric theory (example: ultrametric wavelet transform).

Secondly, practical data sets (derived from, or observed in, databases and data spaces) present some but not exclusively ultrametric characteristics. This can be used for forensic data exploration (fingerprinting data sets, as we discuss below). Or, it can be used to expedite search and discovery in information spaces. Indeed we would like to go a lot further, and gain new insights into data (and observed phenomena and events) through ultrametric or p-adic representations. We see this as a program of work for the near future.

2. ULTRAMETRICITY AND DIMENSIONALITY

2.1. Very High Dimensions are Naturally Ultrametric

Bellman's [2] "curse of dimensionality" relates to exponential growth of hypervolume as a function of dimensionality. Problems become tougher as dimensionality increases. In particular problems related to proximity search in high-dimensional spaces tend to become intractable.

But in very high dimensions the situation is quite different. In very high dimensions there is no longer a "curse of dimensionality" as we will see in sections 2.2 and 2.3 below. High dimensionality implies relative sparsity. So our results are equally valid (i) for high spatial dimensionality, and (ii) for spatially sparse clouds of data points.

In a way, a "trivial limit" (Treves [27]) case is reached as dimensionality increases. This makes high dimensional proximity search very different, and given an appropriate data structure – such as a binary hierarchical clustering tree – we can find nearest neighbors in worst case $O(1)$ or constant computational time [19]. The proof is simple: the tree data structure affords a constant number of edge traversals.

The fact that limit properties are "trivial" makes them no less interesting to study. Let us refer to such "trivial" properties as (structural or geometrical) regularity properties (e.g. all points lie on a regular lattice). First of all, the symmetries of regular structures in our data may be of importance. Secondly, "islands" or clusters in our data, where each "island" is of regular structure, may be exploitable. Thirdly, the mention of exploitability points to the application areas targeted: in this paper, we focus on search and matching and show some ways in which ultrametric regularity can be exploited in practice. Fourthly, and finally, regularity by no means implies complete coverage (e.g., existence of all pairwise linkages) so that interesting or revealing structure will be present in real data sets.

2.2. Distance Properties in Very Sparse Spaces

In Murtagh [19] we discussed the finding that ultrametricity is a natural property of spaces in the (spatially) sparse limit. By this we mean that as spatial dimensionality and spatial sparseness increase, so too do ultrametric properties.

The triangular inequality holds for a metric space: $d(x,z) \leq d(x,y) + d(y,z)$ for any triplet of points x,y,z. In addition the properties of symmetry and positive definiteness are respected. The "strong triangular inequality" or ultrametric inequality is: $d(x,z) \leq \max \{d(x,y), d(y,z)\}$ for any triplet x,y,z. An ultrametric space implies respect for a range of stringent properties. For example, the triangle formed by any triplet is necessarily isosceles, with the two large sides equal; or is equilateral.

Murtagh [19], and earlier work by Rammal et al. [24, 25], has demonstrated the pervasiveness of ultrametricity, by focusing on the fact that sparse high-dimensional data tend to be ultrametric. One reason for this is as follows.

As dimensionality grows, so too do distances (or indeed dissimilarities, if they do not satisfy the triangular inequality). The least change possible for dissimilarities to

become distances has been formulated in terms of the smallest additive constant needed, to be added to all dissimilarities [26, 5, 6, 23]. Adding a sufficiently large constant to all dissimilarities transforms them into a set of distances. Through addition of a larger constant, it follows that distances become approximately equal, thus verifying a trivial case of the ultrametric or "strong triangular" inequality. Adding to dissimilarities or distances may be a direct consequence of increased dimensionality.

For a close fit or good approximation, the situation is not as simple for taking dissimilarities, or distances, into ultrametric distances. If we want a close fit to the given dissimilarities then a good choice would avail either of the maximal inferior, or subdominant, ultrametric; or the minimal superior ultrametric. Stepwise algorithms for these are commonly known as, respectively, single linkage hierarchical clustering; and complete link hierarchical clustering. (See [3, 16, 17] and other texts on hierarchical clustering.)

2.3. Quantifying Degree of Ultrametricity

Summarizing a full description in Murtagh [19] we explored two measures quantifying how ultrametric a data set is.

Firstly, Rammal et al. [25] used discrepancy between each pairwise distance and the corresponding subdominant ultrametric. Now, the subdominant ultrametric is also known as the ultrametric distance resulting from the single linkage agglomerative hierarchical clustering method. Closely related graph structures include the minimal spanning tree, and graph (connected) components. While the subdominant provides a good fit to the given distance (or indeed dissimilarity), it suffers from the "friends of friends" or chaining effect.

Secondly, Lerman [16] developed a measure of ultrametricity, termed H-classifiability, using ranks of all pairwise given distances (or dissimilarities). The isosceles (with small base) or equilateral requirements of the ultrametric inequality impose constraints on the ranks. The interval between median and maximum rank of every set of triplets must be empty for ultrametricity. We have used extensively Lerman's measure of degree of ultrametricity in a data set. Taking ranks provides scale invariance. But the limitation of Lerman's approach, we find, is that it is not reasonable to study ranks of real-valued distances defined on a large set of points.

Thirdly, our own measure of extent of ultrametricity [19] can be described algorithmically. We assume a Euclidean metric. We examine triplets of points (exhaustively if possible, or otherwise through sampling), and determine the three angles formed by the associated triangle. We select the smallest angle formed by the triplet points. Then we check if the other two remaining angles are approximately equal. If they are equal then our triangle is isosceles with small base, or equilateral (when all triangles are equal). The approximation to equality is given by 2 degrees (0.0349 radians). Our motivation for the approximate ("fuzzy") equality was that it makes our approach robust and independent of measurement precision.

Studies are discussed in Murtagh [19] showing how numbers of points in our clouds of data points are irrelevant; but what counts is the ambient spatial dimensionality. Among cases looked at are statistically uniformly (hence "unclustered", or without structure in a

certain sense) distributed points, and statistically uniformly distributed hypercube points (so the latter are random 0/1 valued vectors). Using our ultrametricity measure, there is a clear tendency to ultrametricity as the spatial dimensionality (hence spatial sparseness) increases.

3. LOCAL OR PARTIAL ULTRAMETRICITY

3.1. Approximating Local Ultrametricity

Now we look at data where some triangles are consistent with ultrametric properties, while others are not.

It has long been known [8, 28] that forms of data structuring, and more particularly data clustering, can be used to expedite search problems in high dimensions. Some of the work of Chávez and Navarro and their colleagues provides an explanation as to why and how clustering can be exploited for high dimensional proximity search.

In large data sets, i.e. large n or number of observations, a clever way to expedite proximity searching (in particular nearest neighbor finding) in metric spaces is as follows. The metric property implies that the triangular inequality holds. We have a given point and we are looking for its nearest neighbor. We use a third point, called a pivot point. Such a pivot point is carefully selected at the start of the processing, and all necessary distances to it are stored. Through the triangular inequality, we then form a bound on the best potential nearest neighbor distance. Thereby we limit the region within which the search is carried out. See [7, 8, 9, 4]. As pointed out in Murtagh [19], the bounding rule, or rejection rule, that ensues, is forcing retained triangles to be isosceles. This is interesting because it can be viewed as finding locally ultrametric relationships.

In [7, 8], the ambient spatial dimension is termed the "representational dimension", or embedding dimension, m. (This is dimensionality, m: we have for example $x \in \mathbb{R}^m$.) Search is subject to the curse of dimensionality when addressed in all generality in $\mathbb{R}^m$. However there is often a smaller "intrinsic dimensionality", or average local dimensionality (e.g. when the data are clustered, or lie on a surface of dimension $< m$). This can be exploited to provide fast proximity searching opportunities. However it is difficult in general to define the intrinsic dimensionality.

These authors define intrinsic dimensionality of a metric space as: $\rho = \frac{\mu^2}{2\sigma^2}$ where μ and σ^2 are, respectively, the mean and variance of the distances.

So, firstly, the intrinsic dimensionality grows with the mean distance. We have observed that ultrametricity increases with average distance both by simulations in Murtagh [19], and also through the argument of a simple additive transformation (in section 2.2 above). Secondly, the intrinsic dimensionality grows with inverse variance. Small variance of distances implies equilateral triangles between point triplets, and therefore implies ultrametricity.

We see therefore that the intrinsic dimensionality of [7, 9] affords another definition of ultrametricity. We have already observed how their fast, pivot-based proximity rule can be interpreted as local enforcement of the ultrametric inequality. We conclude from these observations that local or global ultrametricity (i.e., high values of Chávez and

Navarro's ρ, or high local contributions to ρ) permit fast proximity search.

3.2. The Case of High Dimensional Gaussian Clouds

We have already seen that high dimensions lead to "natural" ultrametricity. In this section we discuss the insightful work of [15], viz.

- algorithms for registering (finding an exact match between) two high dimensional Gaussian point sets; and
- an algorithm for partitioning a high dimensional Gaussian point set into two subsets.

What we have discussed in sections 2.1 and 2.2 above allows us to accept that a "trivial limit" of similar distances may be arrived at in high dimensional spaces. But this opens up the intriguing perspective on more than one high dimensional point cloud, each of which has simple simplex structure, but different clouds have different such structures.

Consider two clouds of points in the same space, of typical values x and y, respectively. We consider the following:

- The average variances of x and y points, over all dimensions, are respectively constant σ^2 and τ^2.
- The average over dimensions of the distance squared between the means of x points, and of y points, on each dimension is constant μ^2.

From the first bullet point here, when rescaled by $m^{-1/2}$, points in the x and y clouds are asymptotically located at the vertices of regular simplexes with edge lengths, respectively, $2\sigma^{1/2}$ and $2\tau^{1/2}$.

From the above two points, the independence of the two clouds, and the Law of Large Numbers, the following result holds.

$$\left[\frac{1}{m}\sum_{k=1}^{m}(x_{ik}-y_{jk})^2\right]^{1/2} \xrightarrow{p} (\sigma^2+\tau^2+\mu^2)^{1/2} \tag{1}$$

Interpretation:

All points x are equidistant with distance $2^{1/2}\sigma$.

All points y are equidistant with distance $2^{1/2}\tau$.

Points x tend in probability to be all the same distance from *all* points in y with distance $(\sigma^2+\tau^2+\mu^2)^{1/2}$, from result (1) above.

For the matching or registering of two Gaussian point sets, we require that: $\sigma = \tau$, and $\mu = 0$, so that we have as a result just one point set.

For the partitioning of a high dimensional Gaussian point cloud into two clouds, we require the following:

- $\frac{1}{2}n_1(n_1-1)$ distances equal to $2^{1/2}\sigma$
- $\frac{1}{2}n_2(n_2-1)$ distances equal to $2^{1/2}\tau$

- $\frac{1}{2}(n_1 + n_2)(n_1 + n_2 - 1) - \frac{1}{2}n_1(n_1 - 1) - \frac{1}{2}n_2(n_2 - 1) = n_1 n_2$ distances equal to $(\sigma^2 + \tau^2 + \mu^2)^{1/2}$.

An algorithm based on Gaussian mixture modeling is then as follows. We seek clusters of distances of (possibly very) small variance to define the $x, y, \ldots$ clouds. All other distances can be considered as (by assumption, Poisson) background noise. Model-based clustering [14] can accommodate such a model (although the computational aspect in high dimensions may be troublesome), and provide Bayes factor support for the inherent number of clusters present.

Full exploitation of high dimensional Gaussian cloud merging or subdivision has yet to be undertaken.

4. INCREASING ULTRAMETRICITY THROUGH DATA RECODING

4.1. Ultrametricity of Time Series

In Murtagh [20] we use the following coding to show that chaotic time series are less ultrametric than, say, financial, biomedical or meteorological time series; random generated (uniformly distributed) time series data are remarkably similar in their ultrametric properties; and ultrametricity can be used to distinguish various types of biomedical (EEG) signals.

A time series can be easily embedded in a space of dimensionality m, by taking successive intervals of length m, or a delay embedding of order m. Thus we define points

$$\mathbf{x}_r = (x_{r-m+1}, x_{r-m+2}, \ldots, x_{r-1}, x_r)^t \in \mathbb{R}^m$$

where t denotes vector transpose. Based on previous results we expect that as the dimension m grows, then the point set in $\mathbb{R}^m$ becomes more ultrametric. This finding is borne out below.

Given any $\mathbf{x}_r = (x_{r-m+1}, x_{r-m+2}, \ldots, x_{r-1}, x_r)^t \in \mathbb{R}^m$, let us consider the set of s such contiguous intervals determined from the time series of overall size n. For convenience we will take $s = \lfloor n/m \rfloor$ where $\lfloor . \rfloor$ is integer truncation. The contiguous intervals could be overlapping but for exhaustive or near-exhaustive coverage it is acceptable that they be non-overlapping. In our work, the intervals were non-overlapping. The quantification of the ultrametricity of the overall time series is provided by the aggregate over s time intervals of the ultrametricity of each $\mathbf{x}_r$, $1 \leq r \leq s$.

We seek to directly quantify the extent of ultrametricity in time series data. In [25, 19] it was shown how increase in ambient spatial dimensionality leads to greater ultrametricity. However it is not satisfactory from a practical point of view to simply increase the embedding dimensionality m insofar as short memory relationships are of greater practical relevance (especially for prediction). The greatest possible value of $m > 1$ is the total length of the time series, n. Instead we will look for an ultrametricity measurement approach for given and limited sized dimensionalities m. Our experimental results for real and for random data sets are for "window" lengths $m = 5, 10, \ldots, 105, 110$.

We seek local ultrametricity, i.e. hierarchical structure, by studying the following: Euclidean distance squared, $d_{jj'} = (x_{rj} - x_{rj'})^2$ for all $1 \leq j, j' \leq m$ in each time window, $\mathbf{x}_r$.

We enforce sparseness [24, 25, 19] on our given distance values, $\{d_{jj'}\}$. We do this by linearly approximating each value $d_{jj'}$, in the range $\max_{jj'} d_{jj'} - \min_{jj'} d_{jj'}$, by an integer in $1, 2, \ldots p$. Note that the range is chosen with reference to the currently considered time series window, $1 \leq j, j' \leq m$. Note too that the value of p must be specified. In our work we set $p = 2$. Thus far, the recoded value, $d'_{jj'}$ is not necessarily a distance. With the extra requirement that $d'_{jj'} \longrightarrow 0$ whenever $j = j'$ it can be shown that $d'_{jj'}$ is a metric.

To summarize, in our coding, a small pairwise transition is mapped onto a value of 1; and an exceptionally large pairwise transition is mapped onto a value of 2. A pairwise transition is defined not just for data values that are successive in time but for any pair of data values in the window considered.

4.2. Ultrametricity of Text

In [21], (in principle all, but in practice a set of the few hundred most frequent) words appearing in a text are used to "fingerprint" it. Rare words in a text corpus may be appropriate for querying the corpus for relevant texts, but such words are of little help for inter-text characterization and comparison. We also use entire words, with no stemming or other preprocessing. A full justification for such an approach to textual data analysis can be found in Murtagh [22].

So our methodology for studying a set of texts is to characterize each text with numbers of terms appearing in the text, for a set of terms. The χ^2 distance is an appropriate weighted Euclidean distance for use with such data [3, 21]. Consider texts i and i' crossed by words j. Let k_{ij} be the number of occurrences of word j in text i. Then, omitting a constant, the χ^2 distance between texts i and i' is given by $\sum_j 1/k_j (k_{ij}/k_i - k_{i'j}/k_{i'})^2$. The weighting term is $1/k_j$. The weighted Euclidean distance is between the *profile* of text i, viz. k_{ij}/k_i for all j, and the analogous *profile* of text i'.

Correspondence analysis allows us to project the space of documents (we could equally well explore the terms in the *same* projected space) into a Euclidean space. It maps the all pairs χ^2 distance into the corresponding Euclidean distance. In the resulting factor space, we use our triangle-based approach for quantifying how ultrametric the data are.

We did this for a large number of texts (novels – Jane Austen, James Joyce, technical reports – airline accident reports, fairy tales – Brothers Grimm, dream reports, Aristotle's *Categories*, etc.), finding consistent degree of ultrametricity results over texts of the same sort.

Some very intriguing ultrametricity characterizations were found in our work. For example, we found that the technical vocabulary of air accidents did not differ greatly in terms of inherent ultrametricity compared to the Brothers Grimm fairy tales. Secondly we found that novelist Austen's works were distinguishable from the Grimm fairy tales. Thirdly we found dream reports to be have higher ultrametricity level than the other text collections.

4.3. Data Recoding in the Correspondence Analysis Tradition

If the χ^2 distance (see above, section 4.2) is used on data tables with constant marginal sums then it becomes a weighted Euclidean distance. This is important for us, because it means that we can directly influence the analysis by equi-weighting, say, the table rows in the following way: we double the row vector values by including an absence (0 value) whenever there is a presence (1 value) and vice versa. Or for a table of percentages, we take both the original value x and $100 - x$. In the correspondence analysis tradition [3, 20] this is known as *doubling* (*dédoublement*).

More generally, booleanizing, or making qualitative, data in this way, for a varying (value-dependent) number of target value categories (or modalities) leads to the form of coding known as *complete disjunctive form*.

Such coding increases the embedding dimension, and data sparseness, and thus may encourage degree of ultrametricity. That it can do more we will now show.

The iris data has been very widely used as a toy data set since Fisher used it in 1936 ([10], taking from a 1935 article by Anderson) to exemplify discriminant analysis. It consists of 150 iris flowers, each characterized by 4 petal and sepal, width and breadth, measurements. On the one hand, therefore, we have the 150 irises in $\mathbb{R}^4$. Next, each variable value was recoded to be a rank (all ranks of a given variable considered) and the rank was boolean-coded (viz., for the top rank variable value, $1000\ldots$, for the second rank variable value, $0100\ldots$, etc.). Following removal of zero total columns, the second data sets defined the 150 irises in $\mathbb{R}^{123}$. Actually, this definition of the 150 irises is in fact in $\{0,1\}^{123}$.

Our triangle-based measure of the degree of ultrametricity in a data set (here the set of irises), with 0 = no ultrametricity, and 1 = every triangle an ultrametric-respecting one, gave the following: for irises in $\mathbb{R}^4$, 0.017; and for irises in $\{0,1\}^{123}$: 0.948.

This provides a nice illustration of how recoding can dramatically change the picture provided by one's data. Furthermore it provides justification for data recoding if the ultrametricity can be instrumentalized by us in some way.

It is with such instrumentalization of local ultrametricity that we are now working, applied to chemical structure databases. The clustering of chemical compounds, based on chemical descriptors or representations, is important in the pharmaceutical and chemical sectors. It is used for screening and knowledge discovery in large databases of chemical compounds. A chemical compound is encoded as a bit string (i.e. a set of boolean or 0/1 values). We have started to look at a set of 1.2 million chemical compounds, each characterized (in a given descriptor or coding system) by 1052 variables. Through (i) normalizing to take variable colum sums into account, and then (ii) using the Euclidean distance, we find the following: (i) many resultant distances are zero, pointing to membership in the same cluster; and (ii) of non-zero distances, triangles that are isosceles with small base predominate. In other words, we have a simple but most effective way to embed our data in an ultrametric space.

5. ULTRAMETRICITY IN OBSERVATIONAL COSMOLOGY

In [1], the authors suggest that for large scale galactic structures the ancestor distance – tree distance and hence ultrametric – is important in addition to the distance measured by traveling light waves. In this section we will discuss how this can be instrumentalized.

Clearly one way to construct a hierarchical classification of galaxies is to have observed physical descriptors for each galaxy (related to morphology, motion, photometry, etc.). In such a parameter space it is straightforward to carry out a hierarchical clustering of the galaxies (see some references in [18]). This gives us a hierarchy from which we can read off an ultrametric topology on our galaxy set. But to infer ancestor relations is quite another matter, and such an approach is not suitable for such an objective.

In order to more directly derive hierarchical relationships from positional data, we took data on just over 118,000 galaxies from the Sloan Digital Sky Survey (SDSS), using right ascension (RA), declination (dec) and photometric redshift. We have an immediate issue to address in regard to data encoding: the redshift is far less precisely known, it is substantively different, and is error-prone. Our processing procedure started with binning the data into a regular 3-dimensional grid of dimensions $50 \times 50 \times 50$ (for separate but in each case contiguous parts of the sky that we analyzed). Then we used Lerman's H-classifiability ultrametricity measure, rather than our own, in view of the possible benefit of rank order information in this case. For three parts of the sky, using respectively 589, 554 and 715 galaxies, we found H-classifiability values of 0.15, 0.12, 0.15. For comparison, corresponding (statistically) uniformly distributed positions and redshifts were also used, and processed in the same way. The H-classifiability values (0 = fully ultrametric, 1 = non-ultrametric) were, respectively: 0.19, 0.18 and 0.17. At least we see that we find greater ultrametricity than random values.

Our methodology has not done justice to the data although it must be admitted that data analysis that involves redshifts (in particular) is problematic. Perhaps a better way to explore observational data on galaxies is being pursed by Fraix-Burnet [11, 12, 13]. This work takes observational data (or simulated data, to validate the methodology) on galaxy morphologies. Then a taxonomical approach is pursued where presence or absence of features are processed in order to build a graph of relationships. This procedure yields a robust hierarchy. The stated aim of the work, "to evaluate astrocladistics in reconstructing phylogenies of galaxies", in order "to formalize the concept of galaxy formation and to identify the processes of diversification", has led (in this early stage of the work) to strong evidence for hierarchical branching processes.

6. CONCLUSIONS

It has been our aim in this work to link observed data with an ultrametric topology for such data. The traditional approach in data analysis, of course, is to impose structure on the data. This is done, for example, by using some agglomerative hierarchical clustering algorithm. We can always do this (modulo distance or other ties in the data). Then we can assess the degree of fit of such a (tree or other) structure to our data.

For our purposes, here, this is unsatisfactory.

Firstly, our aim was to show that ultrametricity can be naturally present in our data,

globally or locally. We did not want any "measuring tool" such as an agglomerative hierarchical clustering algorithm to overly influence this finding. (Unfortunately [25] suffers from precisely this unhelpful influence of the "measuring tool" of the subdominant ultrametric.)

Secondly, let us assume that we did use hierarchical clustering, and then based our discussion around the goodness of fit. This again is a traditional approach used in data analysis, and in statistical data modeling. But such a discussion would have been unnecessary and futile. For, after all, if we have ultrametric properties in our data then many of the widely used hierarchical clustering algorithms will give precisely the same outcome, and furthermore the fit is by definition exact.

In linking data with an ultrametric view of it we have, in this article, proceeded a little in the direction of exploiting this achievement. Processing tools available to us include, for example, the ultrametric wavelet transform studied by S.V. Kozyrev, A.Yu. Khrennikov and others (including implementation and case studies for filtering of tabular data by ourselves). The full unleashing of the potential of such data handling and processing tools still remains to be fully investigated.

While some applications, like discrimination between time series signals, or texts, have been covered here, other applications like chemical compound database search and discovery, and analysis of large scale cosmological structures, have just been opened up. In the distance there looms the challenge of analysis of networks of enormous size (internet, or biological). There is a great deal of work to be accomplished.

REFERENCES

1. M.V. Altaisky and B.G. Sidharth, "p-Adic physics below and above Planck scales", *Chaos, Solitons and Fractals*, **10**, 167–176, 1999.
2. R. Bellman, *Adaptive Control Processes: A Guided Tour*, Princeton University Press, 1961.
3. J.P. Benzécri, *La Taxinomie*, 2nd ed., Paris, Dunod, 1979.
4. D. Bustos, G. Navarro and E. Chávez, "Pivot selection techniques for proximity searching in metric spaces", *Pattern Recognition Letters*, **24**, 2357–2366, 2003.
5. F. Cailliez and J.P. Pagès, *Introduction à l'Analyse de Données*, SMASH (Société de Mathématiques Appliquées et de Sciences Humaines), Paris, 1976.
6. F. Cailliez, "The analytical solution of the additive constant problem", *Psychometrika*, **48**, 305–308, 1983.
7. E. Chávez and G. Navarro, "Measuring the dimensionality of general metric spaces", Technical Report TR/DCC-00-1, Department of Computer Science, University of Chile, 2000.
8. E. Chávez, G. Navarro, R. Baeza-Yates and J.L. Marroquín, "Proximity searching in metric spaces", *ACM Computing Surveys*, **33**, 273–321, 2001.
9. E. Chávez and G. Navarro, "Probabilistic proximity search: fighting the curse of dimensionality in metric spaces", *Information Processing Letters*, **85**, 39–56, 2003.
10. R.A. Fisher, "The use of multiple measurements in taxonomic problems", *The Annals of Eugenics*, **7**, 179–188, 1936.
11. D. Fraix-Burnet, Astrocladistics web page, www-laog.obs.ujf-grenoble.fr/ ∼fraix/astroclad.htm
12. D. Fraix-Burnet, P. Choler, E. Douzery and A. Verhamme "Astrocladistics: a phylogenetic analysis of galaxy evolution. I. Character evolutions and galaxy histories", *Journal of Classification*, submitted, 2005.
13. D. Fraix-Burnet, E. Douzery, P. Choler and A. Verhamme "Astrocladistics: a phylogenetic analysis of galaxy evolution. II. Formation and diversification of galaxies", *Journal of Classification*, submitted, 2005.

14. C. Fraley and A.E. Raftery, "How many clusters? Which clustering method? Answers via model-based cluster analysis", *The Computer Journal*, **41**, 578–588, 1998.
15. P. Hall, J.S. Marron and A. Neeman, "Geometric representation of high dimension low sample size data", *Journal of the Royal Statistical Society B*, **67**, 427–444, 2005.
16. I.C. Lerman, *Classification et Analyse Ordinale des Données*, Paris, Dunod, 1981.
17. F. Murtagh, *Multidimensional Clustering Algorithms*, Physica-Verlag, 1985.
18. F. Murtagh and A. Heck, *Multivariate Data Analysis*, Kluwer, 1987.
19. F. Murtagh, "On ultrametricity, data coding, and computation", *Journal of Classification*, **21**, 167–184, 2004.
20. F. Murtagh, "Identifying the ultrametricity of time series", *European Physical Journal B*, **43**, 573–579, 2005.
21. F. Murtagh, "A note on local ultrametricity in text", *Literary and Linguistic Computing*, submitted, 2005.
22. F. Murtagh, *Correspondence Analysis and Data Coding with R and Java*, Chapman & Hall/CRC, 2005.
23. E. Neuwirth and L. Reisinger, "Dissimilarity and distance coefficients in automation-supported thesauri", *Information Systems*, **7**, 47–52, 1982.
24. R. Rammal, J.C. Angles d'Auriac, and B. Doucot, "On the degree of ultrametricity", *Le Journal de Physique – Lettres*, **46**, L-945–L-952, 1985.
25. R. Rammal, G. Toulouse, and M.A. Virasoro, "Ultrametricity for physicists", *Reviews of Modern Physics*, **58**, 765–788, 1986.
26. W.S. Torgerson, *Theory and Methods of Scaling*, Wiley, 1958.
27. A. Treves, "On the perceptual structure of face space", *BioSystems*, **40**, 189–196 (1997).
28. C.J. van Rijsbergen, *Information Retrieval*, 2nd ed. Butterworths, 1979.