International Journal of Wavelets, Multiresolution and Information Processing

WAVELET TRANSFORM ON ULTRAMETRIC AND OTHER TREE TOPOLOGIES

F. MURTAGH

School of Computer Science, Queen's University Belfast
Belfast BT7 1NN, Northern Ireland, UK
fmurtagh@acm.org

Received (Day Month Year)
Revised (Day Month Year)
Communicated by (xxxxxxxxxx)

We consider the wavelet transform of a rooted, node-ranked, p-way tree. We focus on the case of binary trees, and a variant of the Haar wavelet transform. We distinguish between two cases: firstly, where the binary tree represents a hierarchy of embedded, non-overlapping subsets of a given set; and secondly, where the binary tree represents an ultrametrically related set of points. Wavelet transforms allow for multiresolution analysis through translation and dilation of a wavelet function. For a p-way tree translation becomes trivial. For dilation we define and comprehensively describe in this context the use of an expansive automorphism, as introduced by Benedetto and Benedetto.

Keywords: Haar wavelet transform; binary tree; ultrametric topology; p-adic numbers; hierarchical clustering; data mining; local field; abelian group.

AMS Subject Classification: 22E46, 53C35, 57S20

1. Introduction

1.1. *Motivation*

The wavelet transform, developed for signal and image processing, has been extended for use on relational data tables and multidimensional data sets [25,10] for data summarization (micro-aggregation) with the goal of anonymization (or statistical disclosure limitation) and macrodata generation; and data summarization with the goal of computational efficiency, especially in query optimization. There are problems, however, in doing this with direct application of a wavelet transform. Essentially, a relational table is treated in the same way as a 2-dimensional pixelated image, although the former case is invariant under row and column permutation, whereas the latter case is not [14]. Therefore there are immediate problems related to non-uniqueness, and data order dependence. In addition, for very small dimensions, for example a small number of attributes in a relational data table, a classical application of a wavelet transform is troublesome, and in addition if table dimensionality equal to an integer power of 2 is required, the procedure is burdensome

to the point of being counter-productive. Sparse tabular data cannot be interpolated in the same way as sparse pixelated data (e.g. [24]) if only because row/column permutation invariance causes the outcome to be dominated by sparsity-induced effects.

Our work allows for a very different way to wavelet transform tabular data. An intermediate hierarchical representation though is used. Once this is done, the hierarchy is wavelet transformed. The approach is a natural and "integral" one. A hierarchy may be constructed through use of any constructive, hierarchical clustering algorithm [3,11,12,13].

A second motivation for the development of wavelet transforms in ultrametric or hierarchical data structures is to cater for "naturally" or enforced ultrametric data. An example is that of questionnaire results data with embedded question sets. Another very significant example of such a situation is that of data with already strong ultrametric tendency such as high-dimensional data in speech analysis, genomics and proteomics, and other fields [17,18,15], and complete disjunctive form and other alternative forms of data coding in correspondence analysis [4].

1.2. *Local Fields*

Wavelet transform analysis is the determining of a "useful" basis for $L^2(\mathbb{R}^m)$ with the following properties:

- induced from a discrete subgroup of $\mathbb{R}^m$,
- using translations on this subgroup, and
- dilations of the basis functions.

Classically,[23,7,5] the wavelet transform avails of a wavelet function $\psi(x) \in L^2(\mathbb{R})$, where the latter is the space of all square integrable function. Wavelet transforms are bases on $L^2(\mathbb{R}^m)$, and the discrete lattice subgroup $\mathbb{Z}^m$ is used to allow discrete groups of dilated translation operators to be induced on $\mathbb{R}^m$.

To consider the wavelet transform approach not in a Hilbert space but rather in locally-defined and discrete spaces we have to change the specification of a wavelet function in $L^2(\mathbb{R})$ and instead use $L^2(G)$ (see Benedetto and Benedetto[1,2]). The group G is a locally compact abelian group. Analogous to the integer grid, a compact subgroup is used to allow a discrete group of operators to be defined on $L^2(G)$.

The property of locally compact (essentially: finite and free of edges) abelian (viz., commutative) groups that is most important is the existence of the Haar measure [26]. The Haar measure allows integration, and definition of a topology on the algebraic structure of the group.

1.3. *Haar Wavelet Transform*

Classically, the Haar wavelet function basis for analysis of $L^2(\mathbb{R}^m)$ is determined by inducing the basis from an m-dimensional pixel (time step, voxel, etc.) grid, $\mathbb{Z}^m$.

Basis functions of a space denoted by V_j are defined from a scaling function ϕ as follows[22]:

$$\phi_{j,i}(x) = \phi(2^{-j}x - i) \quad i = 0, \ldots, 2^j - 1 \qquad \text{with } \phi(x) = \begin{cases} 1 \text{ for } 0 \leq x < 1 \\ 0 \text{ otherwise} \end{cases} \tag{1.1}$$

The dimensionality of space V_j directly leads to a dyadic analysis. The functions ϕ are all box functions, defined on the interval $[0, 1)$ and are piecewise constant on 2^j subintervals. We can approximate any function in spaces V_j associated with basis functions ϕ_j, in a very fine manner for V_0 (in this case of V_0, all values), more crudely for V_{j+1} and so on. We consider the nesting of spaces, $\ldots V_{j+1} \subset V_j \subset V_{j-1} \ldots \subset V_0$.

Next we consider the orthogonal complement of V_{j+1} in V_j, and call it W_{j+1}. The basis functions for W_j are derived from the Haar wavelet. We find

$$\psi_{j,i}(x) = \psi(2^{-j}x - i) \quad i = 0, \ldots, 2^j - 1 \text{ with } \psi(x) = \begin{cases} 1 \; 0 \leq x < \frac{1}{2} \\ -1 \; \frac{1}{2} \leq x < 1 \\ 0 \text{ otherwise} \end{cases} \tag{1.2}$$

This leads to the basis for V_j as being equal to: the basis for V_{j+1} together with the basis for W_{j+1}. In practice we use this finding like this: we write a given function in terms of basis functions in V_j; then we rewrite in terms of basis functions in V_{j+1} and W_{j+1}; and then we rewrite the former to yield, overall, an expression in terms of basis functions in V_{j+2}, W_{j+2} and W_{j+1}. The wavelet parts provide the detail part, and the space V_{j+2} provides the smooth part.

For the definitions of scaling function and wavelet function in the case of the Haar wavelet transform, proceeding from the given signal, the spaces V_j are formed by averaging of pairs of adjacent values, and the spaces W_j are formed by differencing of pairs of adjacent values. Proceeding in this direction, from the given signal, we see that application of the scaling or wavelet functions involves downsampling of the data. The low-pass filter is a moving average. The high-pass filter is a moving difference. Other low- and high-pass filters are alternatively used to yield other wavelet transforms.

1.4. *Hierarchy, Binary Tree and Ultrametric Topology*

A hierarchy, H, is defined as a binary, rooted, unlabeled, node-ranked tree, also termed a dendrogram[3,11,12,13]. A hierarchy defines a set of embedded subsets of a given set, I. However these subsets are totally ordered by an index function ν, which is a stronger condition than the partial order required by the subset relation. A bijection exists between a hierarchy and an ultrametric space.

Let us show these equivalences between embedded subsets, hierarchy, and binary tree, through the constructive approach of inducing H on a set I.

Hierarchical agglomeration on n observation vectors, $i \in I$, involves a series of $1, 2, \ldots, n-1$ pairwise agglomerations of observations or clusters, with the following properties. A hierarchy $H = \{q | q \in 2^I\}$ such that (i) $I \in H$, (ii) $i \in H \; \forall i$, and (iii) for each $q \in H, q' \in H : q \cap q' \neq \emptyset \implies q \subset q'$ or $q' \subset q$. Here we have denoted

the power set of set I by 2^I. An indexed hierarchy is the pair (H, ν) where the positive function defined on H, i.e., $\nu : H \rightarrow \mathbb{R}^+$, satisfies: $\nu(i) = 0$ if $i \in H$ is a singleton; and (ii) $q \subset q' \Longrightarrow \nu(q) < \nu(q')$. Here we have denoted the positive reals, including 0, by $\mathbb{R}^+$. Function ν is the agglomeration level. Take $q \subset q'$, let $q \subset q''$ and $q' \subset q''$, and let q'' be the lowest level cluster for which this is true. Then if we define $D(q, q') = \nu(q'')$, D is an ultrametric. In practice, we start with a Euclidean or other dissimilarity, use some criterion such as minimizing the change in variance resulting from the agglomerations, and then define $\nu(q)$ as the dissimilarity associated with the agglomeration carried out.

1.5. *Wavelet Transform on Hierarchy of Embedded Subsets, and on Ultrametric Topology*

Our objectives are to explore the foundations of two distinct approaches. Both seek a Haar wavelet basis. There two approaches are as follows.

- Carry out the analysis in $L^2(\mathbb{R}^m)$, inducing the Haar basis from the hierarchy H that expresses the relationships in a set of ultrametrically related points, I.
- Carry out the analysis in $L^2(I)$, inducing the Haar basis from the hierarchy H defining a set of subsets of I.

In the ultrametric case, each point $i \in I$ defines an m-dimensional vector: $i \in \mathbb{R}^m$. For notational convenience therefore i is either the index, or a vector.

In the set of subsets case, each point $i \in I$ can be defined as an n-dimensional index vector. Thus for example the second point is defined as $(0, 1, 0, 0, \ldots, 0)$.

We need to consider translation and dilation operations. Translation for a p-way tree is defined straightforwardly. For sufficiently large p, translation is implemented in the usual manner. In practice we are more interested in small p (2, in this article, or 3). In this case, the support of the wavelet function is of the same size at all levels, and translation is not needed.

Dilation is considered as an expansive automorphism[1]. Let I be a locally compact abelian group with compact open subgroup H. Note that ultrametric sets are simultaneously open and closed (they are *clopen*). Let A be an automorphism of I. Then A is expansive with respect to H if both of the following conditions hold:

Definition 1.1 (Expansive automorphism).

(1) $H \subset AH$ and $H \neq AH$
(2) $\cup_{j \geq 0} A^j H = \{0\}$

In the next section we will introduce a notational framework in order to show how these properties hold. The contribution of this paper is to show the validity and the applicability of the expansive automorphism operator. To achieve this we require a special encoding of the domain under investigation. For this we use a special p-adic number representation.

We begin by noting that a group has an addition operation, and we do not have a natural operation of this sort. Therefore we will introduce an addition operation on dendrograms, based on a particular p-adic representation. We then develop a wavelet transform approach for ranked binary rooted trees, representing either (case 1) a hierarchy of sets, or (case 2) an ultrametric topology.

2. H Expressed P-adically: P-adic Encoding of a Dendrogram

A terminal-to-root traversal in a dendrogram or binary rooted tree is defined as follows. We use the path $x \subset q \subset q' \subset q'' \subset \ldots q_{n-1}$, where x is a given object specifying a given terminal, and $q, q', q'', \ldots$ are the embedded classes along this path, specifying nodes in the dendrogram. The root node is specified by the class q_{n-1} comprising all objects.

A terminal-to-root traversal is the shortest path between the given terminal node and the root node, assuming we preclude repeated traversal (backtrack) of the same path between any two nodes.

By means of terminal-to-root traversals, we define the following p-adic encoding of terminal nodes, and hence objects, in Figure 1.

$$
\begin{aligned}
x_1 &: +1 \cdot p^1 + 1 \cdot p^2 + 1 \cdot p^5 + 1 \cdot p^7 \\
x_2 &: -1 \cdot p^1 + 1 \cdot p^2 + 1 \cdot p^5 + 1 \cdot p^7 \\
x_3 &: \quad -1 \cdot p^2 + 1 \cdot p^5 + 1 \cdot p^7 \\
x_4 &: +1 \cdot p^3 + 1 \cdot p^4 - 1 \cdot p^5 + 1 \cdot p^7 \\
x_5 &: -1 \cdot p^3 + 1 \cdot p^4 - 1 \cdot p^5 + 1 \cdot p^7 \\
x_6 &: \quad -1 \cdot p^4 - 1 \cdot p^5 + 1 \cdot p^7 \\
x_7 &: \quad +1 \cdot p^6 - 1 \cdot p^7 \\
x_8 &: \quad -1 \cdot p^6 - 1 \cdot p^7
\end{aligned}
$$

If we choose $p = 2$ the resulting decimal equivalents could be identical: cf. contributions based on $+1 \cdot p^1$ and $-1 \cdot p^1 + 1 \cdot p^2$. Given that the coefficients of the p^j terms ($1 \leq j \leq 7$) are in the set $\{-1, 0, +1\}$ (implying for x_1 the additional terms: $+0 \cdot p^3 + 0 \cdot p^4 + 0 \cdot p^6$), the coding based on $p = 3$ is required to avoid ambiguity among decimal equivalents.

A few general remarks on this encoding follow. For the labeled ranked binary trees that we are considering, we require the labels $+1$ and -1 for the two branches at any node. Of course we could interchange these labels, and have these $+1$ and -1 labels reversed at any node. By doing so we will have different p-adic codes for the objects, x_i.

The following properties hold: (i) *Unique encoding:* the decimal codes for each x_i (lexicographically ordered) are unique for $p \geq 3$; and (ii) *Reversibility:* the dendrogram can be uniquely reconstructed from any such set of unique codes.

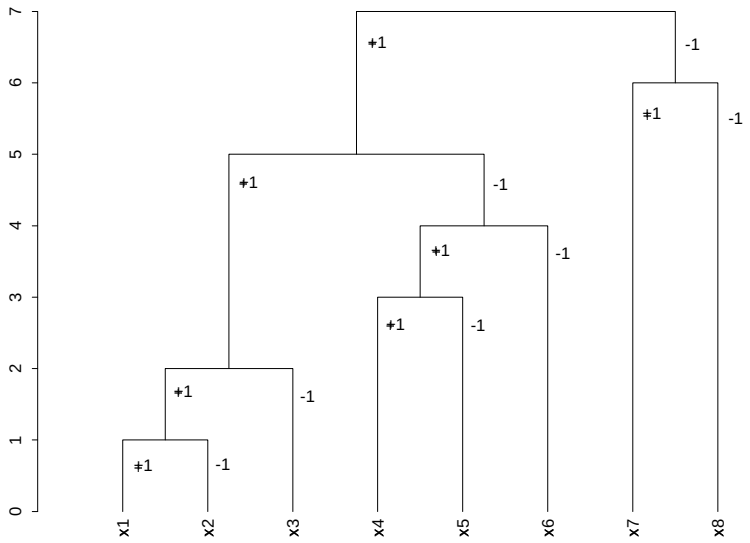

Fig. 1. Labeled, ranked dendrogram on 8 terminal nodes, $x_1, x_2, \ldots, x_8$. Branches are labeled $+1$ and -1. Clusters are: $q_1 = \{x_1, x_2\}, q_2 = \{x_1, x_2, x_3\}, q_3 = \{x_4, x_5\}, q_4 = \{x_4, x_5, x_6\}, q_5 = \{x_1, x_2, x_3, x_4, x_5, x_6\}, q_6 = \{x_7, x_8\}, q_7 = \{x_1, x_2, \ldots, x_7, x_8\}$.

The p-adic encoding defined for any object set above can be expressed as follows for any object x associated with a terminal node:

$$x = \sum_{j=1}^{n-1} c_j p^j \text{ where } c_j \in \{-1, 0, +1\} \tag{2.3}$$

In greater detail we have:

$$x_i = \sum_{j=1}^{n-1} c_{ij} p^j \text{ where } c_{ij} \in \{-1, 0, +1\} \tag{2.4}$$

Here j is the level or rank (root: $n-1$; terminal: 1), and i is an index of each possible branch (left, not present, right).

In our examples we have used: $a_j = +1$ for a left branch (in the sense of Figure 1), $= -1$ for a right branch, and $= 0$ when the node is not on the path from that particular terminal to the root.

A matrix form of this encoding is as follows, where $\{\cdot\}^t$ denotes the transpose of the vector.

Let $\mathbf{x}$ be the column vector $\{x_1\ x_2\ \ldots x_n\}^t$.

Let $\mathbf{p}$ be the column vector $\{p^1\ p^2\ \ldots p^{n-1}\}^t$.

Define a characteristic matrix C of the branching codes, $+1$ and -1, and an absent or non-existent branching given by 0, as a set of values c_{ij} where $i \in I$, the indices of the object set; and $j \in \{1, 2, \ldots, n-1\}$, the indices of the dendrogram levels or nodes ordered increasingly. For Figure 1 we therefore have:

$$C = \{c_{ij}\} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ -1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{pmatrix} \tag{2.5}$$

For given level j, $\forall i$, the absolute values $|c_{ij}|$ give the membership function either by node, j, which is therefore read off columnwise; or by object index, i which is therefore read off rowwise.

The matrix form of the p-adic encoding is:

$$\mathbf{x} = C\mathbf{p} \tag{2.6}$$

Here, $\mathbf{x}$ is the decimal encoding, C is the matrix with dendrogram branching codes and $\mathbf{p}$ is the vector of powers of a fixed integer (usually, more restrictively, fixed prime) p.

The encoding of the tree shown in Figure 1 with labels (or coefficients in equations 2.3 or 2.4, 2.5 or 2.6) $+1$ and -1 is not commonly used: zero and one labels are more common. We required the ± 1 labels, however, to fully cater for the ranked nodes (i.e. the total order, as opposed to a partial order, on the nodes).

We can consider the objects that we are dealing with to have equivalent integer values. To show that, all we must do is work out decimal equivalents of the p-adic expressions used above for $x_1, x_2, \ldots$. As noted in Gouvêa[9], we have equivalence between: a p-adic number; a p-adic expansion; and an element of $\mathbb{Z}_p$ (the p-adic integers). The coefficients used to specify a p-adic number, Gouvêa[9] notes (p. 69) "must be taken in a set of representatives of the class modulo p. The numbers between 0 and $p-1$ are only the most obvious choice for these representatives. There are situations, however, where other choices are expedient."

3. P-adic Dendrogram Addition and Multiplication

As noted already the wavelet basis on $L^2(\mathbb{R}^m)$ is often induced from the discrete subgroup, $\mathbb{Z}^m$. Now for a discrete subgroup we use the dendrogram, H. The addition operation on the group H will now be explored.

In order to define a group structure on the p-adic encoded objects, we require an addition operation. We do not "carry and add" in the traditional way because this does not make sense in this context. Instead we define the following "average

and threshold" operation for any coefficients (of values of **p**, as used in equations 2.5 or 2.6). We define the following compositions for such coefficients.

$$\begin{array}{rcl} +1+1 & \longrightarrow & +1 \\ -1-1 & \longrightarrow & -1 \\ +1-1 & \longrightarrow & 0 \\ -1+1 & \longrightarrow & 0 \\ +1 \pm 0 & \longrightarrow & 0 \\ -1 \pm 0 & \longrightarrow & 0 \end{array} \tag{3.7}$$

Examples from the encoding defined above for $x_1, x_2, \ldots$ (again with reference to Figure 1, and equations 2.3 or 2.4, 2.5 or 2.6) follows.

$x_1 + x_2 = +1 \cdot p^2 + 1 \cdot p^5 + 1 \cdot p^7$
$x_1 + x_3 = +1 \cdot p^5 + 1 \cdot p^7$
$x_1 + x_7 = 0$
$x_3 + x_6 = +1 \cdot p^7$
$x_5 + x_8 = 0$

Informally: in the tree, this addition operation only retains non-zero terms for nodes in the tree strictly *above* the first (i.e. lowest level) cluster within which the two objects find themselves. This means that if the two objects only find themselves together for the first time in the same cluster that contains all objects then the result of the addition operation is 0.

Let us use our "average and threshold" operation, which we are using as a customized addition, to define clusters. We will do so by example, taking Figure 1 as our case study. We will call the clusters, ranked by increasing node level, $q_1, q_2, \ldots$ as used in the caption of Figure 1.

$q_1 = x_1 + x_2 = +1 \cdot p^2 + 1 \cdot p^5 + 1 \cdot p^7$
$q_2 = q_1 + x_3 = +1 \cdot p^5 + 1 \cdot p^7$
$q_3 = x_4 + x_5 = +1 \cdot p^4 - 1 \cdot p^5 + 1 \cdot p^7$
$q_4 = q_3 + x_6 = -1 \cdot p^5 + 1 \cdot p^7$
$q_5 = q_2 + q_4 = +1 \cdot p^7$
$q_6 = x_7 + x_8 = -1 \cdot p^7$
$q_7 = 0$

The trivial cluster containing all n objects, q_{n-1}, is of value 0 in this representation.

Definition 3.1 (Null element). On the dendrogram H, the set $q_{n-1} = I$ is the null element when using our p-adic encoding (given in 2.4 and 2.6) and addition operation (3.7).

Defining p-adic notation for clusters in this way allows us to define norms of clusters; or to define p-adic distances between clusters; or indeed to define p-adic distances between clusters and objects (singletons, terminals). We will look at these below.

For completeness we will provide a definition of p-adic dendrogram multiplication. Take $x = \sum_j c_j p^j$ and let $y = \sum_j c'_j p^j$. The product operation is defined on the formal (Laurent) power series as:

$$xy = \left(\sum_j c_j p^j \right) \left(\sum_{j'} c'_{j'} p^{j'} \right) = \sum_{jj'} c_j c'_{j'} p^{j+j'} \tag{3.8}$$

With restriction to the term in p^{n-1}, we find that $x_1 x_2 = -1 \cdot p^2$; and $x_1 x_3 = -1 \cdot p^3 - 1 \cdot p^4 + 1 \cdot p^6$. P-adic dendrogram multiplication will be used below in the definition of the expansive operator: this is multiplication by $1/p$.

4. P-adic Distance and Norm on a Dendrogram

Thus far, we have been concerned with an analytic framework. Now we will induce a metric topology on H.

To find the p-adic distance, we look for the term p^r in the p-adic codes of the two objects, where r is the lowest level such that the absolute values of the coefficients of p^r are equal.

Let us look at the set of p-adic codes for $x_1, x_2, \ldots$ above (Figure 1), to give some examples of this.

For x_1 and x_2, we find the term we are looking for to be p^1, and so $r = 1$.
For x_1 and x_5, we find the term we are looking for to be p^5, and so $r = 5$.
For x_5 and x_8, we find the term we are looking for to be p^7, and so $r = 7$.

Having found the value r, the distance is defined as p^{-r}.

(See, inter alia, Benzécri[3], and Gouvêa[9] for this definition of ultrametric distance.)

Examples based on Figure 1:

$|x_1 - x_2|_p = |x_2 - x_1|_p = p^{-1}$ since $r = 1$.
$|x_1 - x_4|_p = |x_4 - x_1|_p = p^{-5}$ since $r = 5$.
$|x_3 - x_6|_p = |x_6 - x_3|_p = p^{-5}$ since $r = 5$.

Examples for clusters from Figure 1:

$|q_1 - q_3|_p = |q_3 - q_1|_p = p^{-5}$.
$|q_2 - q_6|_p = |q_6 - q_2|_p = p^{-7}$.

Defining the norm as $|x - 0|_p$, we have always for a singleton object $r = 0$, and so the norm of an object is always 1. We therefore define the p-adic norm, $|.|_p$, of an object corresponding to a terminal node in the following way: for any object, x, $|x|_p = 1$.

The norm of a non-singleton cluster is different from that of a singleton. It is seen to be strictly smaller. We have: $|q_2|_p = |q_2 - 0|_p = p^{-2}$; $|q_4|_p = |q_4 - 0|_p = p^{-4}$.

For the expansive operator that takes the place of dilation in our analysis of local fields, it will be useful for us to consider product with $1/p$. The norm associated with this operator is seen to be $|1/p - 0|_p = |p^{-1}|_p = p^{-(-1)} = p$.

The expansive map given by multiplication by $1/p$ therefore has norm or modulus p.

The p-adic norm, or p-adic valuation, satisfies the following properties:[21]

(1) $|x|_p \geq 0; |x|_p = 0$ iff $x = 0$
(2) $|x + y|_p \leq \max(|x|_p, |y|_p)$
(3) $|xy|_p = |x|_p |y|_p$

We also have: $|q|_p \leq 1$ with equality only if q is a singleton.

5. The Expansive Automorphism

Let us call product with $1/p$ the operator A.

A is an automorphism of I. That is, A is an algebraic isomorphism from I to I such that A is also a topological homeomorphism.

Theorem 5.1 (Expansive automorphism). The operator A satisfies:

(1) $q \subset Aq$ for all $q \in H$
(2) $\cup_{j \geq 0, q \in H} A^j q = \{0\}$

Now consider the set $\{x_i | i \in I\}$ with its p-adic coding considered above. Take $p = 2$. (Non-uniqueness of corresponding decimal codes is not of concern to us now.) Multiplication of $x_1 = +1 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^5 + 1 \cdot 2^7$ by $1/p = 1/2$ gives: $+1 \cdot 2^1 + 1 \cdot 2^4 + 1 \cdot 2^6$. Each level has decreased by one, and the lowest level has been lost. By carrying out the multiplication-by-$1/p$ operation on all objects, it is seen that the effect is to rise in the hierarchy by one level. Therefore the subset relation stated in part (1) of Theorem 5.1 is exemplified by the expansive map, A.

Repeated application of the operator A gives Aq, A^2q, A^3q, ... Starting with any singleton, $i \in I$, this gives a path from the terminal to the root node in the tree. Each such path ends with the null element, as a result of the Null Element Definition, Definition 3.1. Therefore the intersection of the paths equals the null element. Therefore we have proved part (2) of Theorem 5.1.

Some implications of Theorem 5.1 follow.

For any q, taking $q, Aq, A^2q, \ldots$ as a sequence of open subgroups of I, with $q \subset Aq \subset A^2q \subset \ldots$, and $I = \cup\{q, Aq, A^2q, \ldots\}$, this is termed an inductive sequence of I, and I itself is the inductive limit [19].

Each path defined by application of the expansive automorphism defines a spherically complete system [21,8], which is a formalization of well-defined subset embeddedness.

6. Implementation Using a Haar Wavelet Function

6.1. *The Forward Transform*

Consider any hierarchical clustering, H, represented as a binary rooted tree. For each cluster q'' with offspring nodes q and q', we define $s(q'')$ through application

of the low-pass filter $\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$:

$$s(q'') = \frac{1}{2}\left(s(q) + s(q')\right) = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}^t \begin{pmatrix} s(q) \\ s(q') \end{pmatrix} \tag{6.9}$$

The application of the low-pass filter is carried out in order of increasing node number (i.e., from the smallest non-terminal node, through to the root node). For a terminal node, $s(i)$ is just the given vector, and this issue is addressed further below.

Next for each cluster q'' with offspring nodes q and q', we define detail coefficients $d(q'')$ through application of the band-pass filter $\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$:

$$d(q'') = \frac{1}{2}(s(q) - s(q')) = \begin{pmatrix} 0.5 \\ -0.5 \end{pmatrix}^t \begin{pmatrix} s(q) \\ s(q') \end{pmatrix} \tag{6.10}$$

Again, increasing order of node number is used for application of this filter.

The scheme followed is illustrated in Figure 2, which shows the hierarchy (constructed by the median agglomerative method, although this plays no role here), using for display convenience the first 8 observation vectors in Fisher's iris data [6].

6.2. *The Ultrametric Case*

We now return to the issue of how we start this scheme, i.e. how we define $s(i)$, or the "smooth" of a terminal node. We have distinguished above between:

(1) H as representing an ultrametric set of relations,
(2) H as representing an embedded set of sets.

For case (1) we take $s(i)$ as the m-dimensional observation vector corresponding to i. So, taking all n vectors $s(i)$ we have the initial data matrix X of dimensions $n \times m$.

Then for our set of n points in $\mathbb{R}^m$ given in the form of matrix X we have:

$$X = CD + S_{n-1} \tag{6.11}$$

where D is the matrix collecting all wavelet projections or detail coefficients, d. The dimensions of C are: $n \times (n-1)$. The dimensions of D are $(n-1) \times m$.

If s_{n-1} is the final data smooth, in the limit for very large n a constant-valued m-component vector, then let S_{n-1} be the $n \times m$ matrix with s_{n-1} repeated on each of the n rows.

Consider the jth coordinate of the m-dimensional observation vector corresponding to i. For any $d(q_j)$ we have: $\sum_k d(q_j)_k = 0$, i.e. the detail coefficient vectors are each of zero mean.

To recapitulate we have:

X is of dimensions $n \times m$.

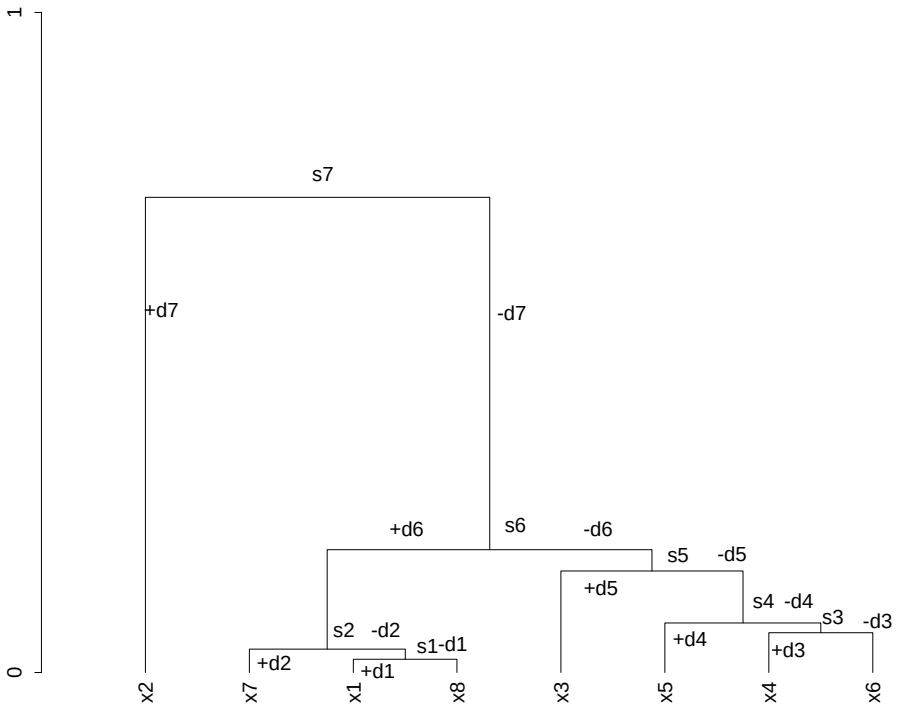

Fig. 2. Dendrogram on 8 terminal nodes constructed from first 8 values of Fisher's iris data. The median agglomerative method was used in this case (which has no direct bearing on the wavelet transform).

C is of dimensions $n \times (n-1)$.
D is of dimensions $(n-1) \times m$.
S_{n-1} is of dimensions $n \times m$.

6.3. *The Case of Embedded Set of Sets*

We have distinguished between

(1) H as representing an ultrametric set of relations,
(2) H as representing an embedded set of sets.

We now turn attention to the latter.

In this case we take $s(i)$ as an n-dimensional indicator vector corresponding to i. So, taking all n vectors $s(i)$ we have the initial data matrix X which is none other than the $n \times n$ dimensional identity matrix. We will write X_{ind} for this identity

Filtering threshold	% coefficients set to zero	mean square error
0	16.95	0
0.1	70.13	0.0098
0.2	91.95	0.0487
0.3	97.15	0.0837
0.4	97.82	0.1040

Table 1. Hierarchical Haar filtering results for Fisher's 150×4 iris data. Filtering is carried out by setting small (less than the threshold) wavelet coefficient values to zero. The thresholds are not reported here. The data is then reconstructed. The quality of reconstruction between the input data matrix, and the reconstructed data matrix, is measured using mean square error. In the 1st row of column 2, 16.95% of coefficients were from the start 0-valued.

matrix.

The wavelet transform in this case is: $X_{\text{ind}} = CD + S_{n-1}$.

X_{ind} is of dimensions $n \times n$.

C, exactly as in case 1 (ultrametric case) is of dimensions $n \times (n-1)$.

D, of necessity different in values from case 1, is of dimensions $(n-1) \times n$.

S_{n-1}, of necessity different in values from case 1, is of dimensions $n \times n$.

6.4. *The Inverse Transform*

In both cases considered (viz., ultrametric, and set of sets) the forward and inverse transforms are performed in the same way. The algorithms are identical – the inputs alone differ.

The inverse transform allows exact reconstruction of the input data. We begin with s_{n-1}. If this root node has subnodes q and q', we use $d(q)$ and $d(q')$ to form $s(q)$ and $s(q')$.

6.5. *Wavelet Filtering*

Setting wavelet coefficients to zero and then reconstructing the data is referred to as hard thresholding (in wavelet space) and this is also termed wavelet smoothing or regression. Table 1 illustrates results obtained for Fisher's iris data. Further results on the energy compacting or compression properties of this new ultrametric Haar transform are given in Murtagh[16], together with R code for this new transform and its inverse.

6.6. *Hierarchic Wavelet Transform and p-Adic Encoding*

We will look at the ultrametric case. The matrix generalization of equation 2.6 is:

$$X = CP \tag{6.12}$$

Matrix P is formed from the vectors $\mathbf{p}$ of equation 2.6 by replicating rows.

Now the wavelet transform gives us: $X = CD + S_{n-1}$. Each (replicated) row of matrix S_{n-1} is a particular measure of central tendency.

Centering X relative to this gives:

$$X - S_{n-1} = CD \tag{6.13}$$

We conclude from the formal similarity of expressions 6.12 and 6.13: the initial p-adic encoding of our data vectors has been mapped into a *wavelet encoding* by the wavelet transform.

7. Conclusions

Generalization to regular p-way trees, for $p > 2$, may also be considered. For $p = 3$ a natural wavelet function is derived from the triangle scaling [22] function, which is itself a convolution of a box function (the scaling function defining the Haar transform, used in this article) with itself. The Haar scaling function used above was $(\frac{1}{2}, \frac{1}{2})$. Convolving this with itself gives then the scaling function $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Convolving the box function again with the triangle function gives the B_3 spline scaling function, $(\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$, which is particularly natural for the analysis of a 5-way, $p = 5$, tree.

A remark on implementation follows: the 3-way tree is unfolded at each node into two 2-way trees. More generally any regular p-way tree is unfolded at each node into $p - 1$ two-way branchings. The wavelet transform algorithm described previously is then directly applied.

References

1. R.L. Benedetto, *Examples of wavelets for local fields*, in C. Heil, P. Jorgensen, D. Larson, eds., *Wavelets, Frames, and Operator Theory, Contemporary Mathematics Vol. 345* (2004).
2. J.J. Benedetto and R.L. Benedetto, *A wavelet theory for local fields and related groups*, *The Journal of Geometric Analysis*, in press (2004).
3. J.P. Benzécri, *La Taxinomie*, 2nd ed. (Dunod, 1979).
4. J.P. Benzécri, transl. T.K. Gopalan, *Correspondence Analysis Handbook* (Marcel Dekker, 1992).
5. L. Debnath and P. Mikusiński, *Introduction to Hilbert Spaces with Applications*, 2nd edn. (Academic Press, 1999).
6. R.A. Fisher, *The use of multiple measurements in taxonomic problems*, *The Annals of Eugenics* **7** (1936), 179–188.
7. M.W. Frazier, *An Introduction to Wavelets through Linear Algebra* (Springer, 1999).
8. L. Gajić, *On ultrametric space*, *Novi Sad Journal of Mathematics* **31** (2001), 69–71.
9. F.Q. Gouvêa, *P-Adic Numbers* (Springer-Verlag, 2003).
10. M.J. Joe, K.-Y. Whang and S.-W. Kim, *Wavelet transformation-based management of integrated summary data for distributed query processing*, *Data and Knowledge Engineering* **39** (2001), 293–312.
11. S.C. Johnson, *Hierarchical clustering schemes*, Psychometrika **32** (1967), 241–254.
12. I.C. Lerman, *Classification et Analyse Ordinale des Données* (Dunod, 1981).
13. F. Murtagh, *Multidimensional Clustering Algorithms* (Physica-Verlag, 1985).

14. F. Murtagh, J.-L. Starck and M. Berry, *Overcoming the curse of dimensionality in clustering by means of the wavelet transform*, *The Computer Journal* **43** (2000), 107–120.
15. F. Murtagh, *On ultrametricity, data coding, and computation*, *Journal of Classification*, in press (2004).
16. F. Murtagh, *Hierarchical or ultrametric Haar wavelet transform in multivariate data analysis and data mining*, preprint (2004).
17. R. Rammal, J.C. Angles d'Auriac and B. Doucot, *On the degree of ultrametricity*, *Le Journal de Physique – Lettres* **46** (1985) L-945 – L-952.
18. R. Rammal, G. Toulouse and M.A. Virasoro, *Ultrametricity for physicists*, *Reviews of Modern Physics* **58** (1986) 765–788.
19. H. Reiter and J.D. Stegeman, *Classical Harmonic Analysis and Locally Compact Groups*, 2nd edition (Oxford University Press, 2000). (Definition 4.1.16, p. 131.)
20. A. Said and W.A. Pearlman, *A new fast and efficient image codec based on set partitioning in hierarchical trees*, *IEEE Transactions on Circuits and Systems for Video Technology* **6** (1996), 243–250.
21. W.H. Schikhof, *Ultrametric Calculus*, (Cambridge University Press, 1984). (Chapters 18, 19, 20, 21.)
22. J.-L. Starck, F. Murtagh and A. Bijaoui, *Image and Data Analysis: The Multiscale Approach* (Cambridge University Press, 1998).
23. G. Strang and T. Nguyen, *Wavelets and Filter Banks* (Wellesley-Cambridge Press, 1996).
24. Wenchang Sun and Xingwei Zhou, *Sampling theorem for wavelet subspaces: error estimate and irregular sampling theorem*, *IEEE Transactions on Signal Processing* **48** (2000), 223–226.
25. J.S. Vitter and M. Wang, *Approximate computation of multidimensional aggregates of sparse data using wavelets*, in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 193–204, 1999.
26. T. Ward, *Entropy of Compact Group Automorphisms*, http://www.mth.uea.ac.uk/∼h720/lecture_notes, 1994.