МАТЕМАТИКА ТА МЕХАНІКА

Алгебра та геометрія

УДК 515.124
MSC 2020: 54E35
DOI: 10.31471/2304-7399-2026-22(83)-67-75

SPARSIFICATION OF COMPACT ULTRAMETRICS

O.R. Nykyforchyn*, **V.M. Penhryn**
Vasyl Stefanyk Carpathian National University;
76018, 57 Shevchenka street, Ivano-Frankivsk, Ukraine
e-mail: oleh.nykyforchyn@cnu.edu.ua, volodymyrpenhryn@gmail.com

We introduce and study a relation of refinement between compact ultrametrics. Efficient methods to determine ultrametrics by functions on binary trees and by symmetric bilinear forms that attain a diagonal form in a basis of Haar-like wavelets, are also proposed.

Key words: *Ultrametric, binary tree, Haar basis.*

Introduction

Ultrametrics are extremely useful in problems of classification [2] as well as in theoretical computer science [5]. Sets of ultrametrics are important objects of study on their own [4].

Ultrametric spaces can be large, even uncountable, therefore approximate methods may be of interest here. This motivates to develop computationally efficient ways of representation of ultrametrics.

We generalize a method for sparsification of ultrametric matrices proposed by Gorman and Lladser [1].

*Corresponding author

1. Trees representing compact ultrametrics

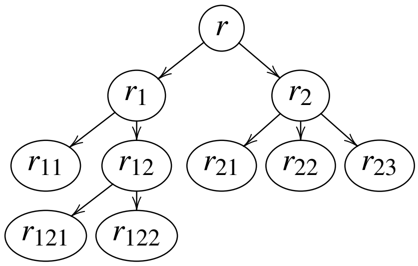
For a compact ultrametric d on a set X, let V_d be the set of all closed balls $\bar{B}_r(x) = \{y \in X \mid d(x,y) \leqslant r\}$, for $x \in X$, $r \geq 0$. Observe that $\bar{B}_r(x) = \bar{B}_{r'}(x)$ is possible for $r \neq r'$, hence for a ball we consider the least possible radius, which coincides with the diameter (the least upper bound of pairwise distances) then. A closed ball $\bar{B}_r(x)$ can be a singleton for $r > 0$ if the point x is isolated, then it is considered as a closed ball or radius 0.

Clearly, for all $b, b' \in V_d$ exactly one of the following: $b = b'$, $b \supsetneq b'$, $b' \supsetneq b$, $b \cap b' = \varnothing$ is valid. If $b \supsetneq b'$, then a unique finite sequence $b_0, b_1, \dots, b_k$ in V exists such that

$$b = b_0 \supsetneq b_1 \supsetneq \dots \supsetneq b_{k-1} \supsetneq b_k = b',$$

and $b_{i-1} \supsetneq b'' \supsetneq b_i$ is impossible for $b'' \in V$.

Consider the oriented graph T_d with the vertices V_d and the edges (b, b'), where $b, b' \in V_d$, $b \supsetneq b'$, and there is no $b'' \in V_d$ such that $b \supsetneq b'' \supsetneq b'$. It is an out-rooted tree such that each vertex has a finite number of children, and there is no vertex with one child. Here is an example for a finite ultrametric space, each vertex (a ball) is labeled with its "true" radius, which is strictly less for a descendant than for an ancestor, e.g., $r > r_1 > r_{12} > r_{121}$:



Kuratowski-Zorn Lemma implies that every simple path beginning in the root (which is the entire X) is contained in a maximal (non-expandable) path, which can be finite or infinite. It consists of all $b \in V_d$ that contain a certain point x, which is unique for the path.

Hence points x of X correspond one-to-one to maximal (non-expandable) paths in T_d. We consider x as a leaf (outer to T_d) that ends the respective path, which therefore goes from the root to x (this "extended" path is not a path in the sense of graph theory if it is infinite). The sets U_b of all points of X paths to which go through b, for all $b \in V_d$, form a base of the topology on X, hence it can be uniquely recovered from T_d.

On the other hand, each out-rooted tree T, with finite numbers of children not equal to 1 for all vertices, represents a compact ultrametric in this manner.

Example 1. *Let X be the set of all maximal paths in T, and, for paths x,y in X, let $d_T(x,y)$ be 0 if $x=y$, or $d_T(x,y)=\frac{1}{k}$ if $x\neq y$ have exactly k common vertices. The ultrametric d_T is a required one. It is not unique, but, by the above, all other compact ultrametrics that determine the same tree T are topologically equivalent.*

2. Refinements

Definition 2.1. *We say that a point z is* between *points x and y w.r.t. an ultrametric d if $d(z,x)\leqslant d(x,y)$ (or, equivalently, $d(z,y)\leqslant d(x,y)$).*

It means that z is not farther from any of the points x and y than they are from each other.

In terms of the tree T_d, it can be said equivalently that the leaf z is a descendant of each common ancestor of the leaves x and y in V_d (in particular, of their lowest common ancestor, if $x\neq y$).

Definition 2.2. *A bijection f between compact ultrametric spaces (X,d) and (X',d') is called a* refinement *if, for all $x,y,z\in X$, the inequality $d(x,z)\leqslant d(x,y)$ implies $d'(f(x),f(z))\leqslant d'(f(x),f(y))$.*

In other words, if z is between x and y w.r.t. d, then $f(z)$ is between $f(x)$ and $f(y)$ w.r.t. d'. It is easy to see that a refinement is a homeomorphism.

If $X=X'$ and f is the identity mapping on X, we obtain an important particular case.

Definition 2.3. *For compact ultrametrics d,d' on a set X we say that d* refines *d' and write $d\succcurlyeq d'$ if, for all $x,y,z\in X$ such that z is between x,y w.r.t. d the point z is between x,y w.r.t. d' as well.*

Clearly d and d' are topologically equivalent then.

Proposition 2.1. *A bijection f between compact ultrametric spaces (X,d) and (X',d') is a refinement if and only if the preimage under f of each closed ball in (X',d') is a closed ball in (X,d).*

In particular, for compact ultrametrics d,d' on a set X, the ultrametric d refines the ultrametric d' if and only if each closed ball w.r.t. d' is a closed ball w.r.t. d'.

We omit a straightforward proof.

If we interpret closed balls w.r.t. a compact ultrametric as classes in a hierarchical classification, then $d\succcurlyeq d'$, $d\neq d'$ means that d provides more detailed classification than d' (there are additional classes intermediate w.r.t. inclusion).

Theorem 2.1. *Let (X,d) and (X',d') be compact ultrametric spaces. There is a one-to-one correspondence between refinements $f:(X,d)\to(X',d')$ and injective mappings $F:V_{d'}\to V_d$ between the sets of closed balls in the spaces (X',d') and (X,d) respectively such that:*

(1) $F(X')=X$;

(2) if $b'_1\subset b'_2$, $b'_1,b'_2\in V_{d'}$, then $F(b'_1)\subset F(b'_2)$ (monotonicity);

(3) for all $b\in V_d$ there is $b'\in V_{d'}$ such that $F(b')\subset b$ (cofinality).

Proof. For a given f, the required F simply takes each closed ball in X' to its preimage under f. If $F:V_{d'}\to V_d$ satisfies (1)–(3), then, for all $x\in X$, the value $f(x)\in X$ is the unique common point x' of all closed balls $b'\in V_{d'}$ such that $F(b')\ni x$. □

If the closed balls in (X,d) and (X',d') are considered as vertices of the trees T and T', as has been described above, then the mentioned F is an injective mapping that takes the vertices of the second tree to vertices of the first one, and:

(1) preserves the root;

(2) preserves being a descendant;

(3) has a cofinal image (each vertex of T is an image under F of a vertex of T', or has a descendant being an image a vertex of T').

We call any mapping F from the set of vertices of a tree T' to the set of vertices of a tree T that satisfies (1)–(3) a *refinement* from the tree T to the tree T' (note the order!), and say that T *refines* T', or T' *is refined by* T.

3. Representation by binary trees

By the last theorem, refinements between compact ultrametric spaces correspond to refinements between out-rooted trees with finite number of children not equal to 1, for all vertices. It is easy to show that every such tree T is refined by an *out-rooted binary tree* (ORB) T_0, in which the vertices have either two, or zero children: each vertex with >2 children is replaced by a binary caterpillar, see [1, Definition 3.7] for details. This replacement is not unique, therefore many ORBs T_0 can refine the same tree T.

If a refinement F from an ORB T_0 to T (i.e., a mapping from the set of the vertices of T to the set of the vertices of T_0 that satisfies the above conditions) is fixed, and T is the tree T_d for a compact ultrametric d on X, then we make T_0 weighted. For any vertex v of T_0 take the closest non-strict ancestor of the form $F(b)$, $b\in V_d$, and put $c(v)=\operatorname{diam} b$. Observe that, for any vertex v of T_0, the greatest lower bound of all values $c(v')$ for the descendants v' of v is equal to 0.

For an edge (v_1,v_2) of T_0 clearly $c(v_2) \leqslant c(v_1)$, and if v_2 has no preimage under F, then $c(v_1) = c(v_2)$. Let this edge have the weight $w(v_1,v_2) = c(v_1) - c(v_2)$, then, in particular, $w(v_1,v_2) = 0$ if v_2 has no preimage under F.

Thus in the weighted tree T_0 the distance between the root and a vertex v is equal to $\operatorname{diam} X - c(v)$, in particular, the path from the root of T_0 to $F(b)$ has the total length $\operatorname{diam} X - \operatorname{diam} b$. Hence all maximal (non-expandable) paths in T that start at $F(b)$ have the same total length $\operatorname{diam} b$.

Fix an encoding of the vertices of T_0 by finite sequences of 0s and 1s so that the empty sequence $\varepsilon = ()$ corresponds to the root, and two children of a vertex v that corresponds to $(x_1,x_2,\ldots,x_n)$ (if there are any) are encoded by $(x_1,x_2,\ldots,x_n,0)$ and $(x_1,x_2,\ldots,x_n,1)$. From now on we identity v with the sequence $(x_1,x_2,\ldots,x_n)$ (writing $c(x_1,x_2,\ldots,x_n)$ for $c(v)$) and denote $v|0 = (x_1,x_2,\ldots,x_n,0)$ and $v|1 = (x_1,x_2,\ldots,x_n,1)$.

If $v = (x_1,x_2,\ldots,x_n,0)$ has no children, then it is the end of a maximal path from the root, and we denote this path by $(x_1,x_2,\ldots,x_n,0,0,\ldots)$. Any infinite maximal path from the root in T_0 is determined by a unique sequence $(x_1,x_2,\ldots,x_n,x_{n+1},\ldots)$ such that the vertices if this path are ε, (x_1), (x_1,x_2), $\ldots$, $(x_1,x_2,\ldots,x_n)$, $\ldots$.

The set $\bar{X}$ of all sequences that encode maximal paths in T_0 is a non-empty compact subspace of D^ω, where D is $\{0,1\}$ with the discrete topology. We obtain a bijection $\bar{F} : X \to \bar{X}$, namely for each $x \in X$ all balls $b \in V_d$ that contain x, are taken, then the sequence of $F(b)$ is uniquely extended to a maximal path in T_0, which is encoded by $\bar{F}(x)$. Clearly $\bar{F}$ is homeomorphism.

For all $n \in \mathbb{N} = \{0,1,2,\ldots\}$, let pr_n be the projection $D^\omega \to D^n$ to the first n coordinates. Denote $X_n = \mathrm{pr}_n(\bar{X}) \subset D^n$, and $X_\omega = X_0 \sqcup X_1 \sqcup X_2 \sqcup \ldots$. Observe that $X_0 = D^0 = \{\varepsilon\}$.

For any $x = (x_1,x_2,\ldots,x_n) \in D^n$ denote $\widehat{x} = \{\bar{x} \in \bar{X} \mid \mathrm{pr}_n(\bar{x}) = x\}$. Clearly $\widehat{x} \neq \varnothing$ for $x \in D^n$ if and only if $x \in X_n$. For all $x \in X_\omega$ the set $\widehat{x|0}$ in non-empty, but the set $X'' = \{x \in X_\omega \mid \widehat{x|1} \neq \varnothing\}$ consists exactly of all bifurcation vertices of the tree T_0. The set

$$X' = \{v \in X_\omega \mid v = x|0 \;\text{ or }\; v = x|1,\, x \in X'' \;\text{ but }\; v \notin X''\}$$

contains all leaves (terminal nodes) of T_0, and the sequences in $T_\omega \setminus (T' \cup T'')$ are “continuations” of terminal nodes.

Observe that $c(v) = 0$ for all $v \in X'$ and extend the function c to X_ω by putting $c(v) = 0$ for all $v \notin X''$. Then:

(1) $c(v|0) \leqslant c(v)$, $c(v|1) \leqslant c(v)$ if the respective elements exist in X_ω;

(2) $\lim_{n\to\infty} c(x_1,x_2,\dots,x_n) = 0$ for all $(x_1,x_2,\dots,x_k,\dots) \in D^\omega$;
(3) $c(x_1,x_2,\dots,x_n) = 0$ if and only if $(x_1,x_2,\dots,x_n) \in X_\omega \setminus X''$.

It is easy to recover the unique ultrametric $\bar{d}$ on $\bar{X}$ such that $\bar{F} : X \to \bar{X}$ is an isometry. For all distinct $x,x' \in \bar{X}$ let v be the common initial part (the longest common prefix) of the sequences x and x', then $\bar{d}(x,x') = c(v)$.

Moreover, any function c on X_ω (in fact, on T_0) that satisfies (1)–(3) determines this way an ultrametric $\bar{d}$ on $\bar{X}$. If T_0 refines the tree T_d for an ultrametric d, then $\bar{d}$ refines d in the obvious sense.

For example, let X_ω be built upon an ORB T_0, and put

$$c(x_1,x_2,\dots,x_n) = \begin{cases} \frac{1}{k}, & (x_1,x_2,\dots,x_n) \in X'', \\ 0, & (x_1,x_2,\dots,x_n) \notin X'', \end{cases} \quad c(x_1,x_2,\dots,x_n) \in X_\omega.$$

Then the induced topology on $\bar{X} \subset D^\omega$ is metrizable by the ultrametric

$$d_{T_0}\big((x_1,x_2,x_3,\dots),(y_1,y_2,y_3,\dots)\big) = \frac{1}{k},$$
$$\text{where } k = \max\{i \in \{1,2,3,\dots\} \mid (x_1,x_2,\dots,x_i) = (y_1,y_2,\dots,y_i)\},$$

here we assume $\max\varnothing = 0$. It is a particular case of the ultrametric d_T constructed in Example 1.

4. Sparsification via a Haar-like basis

We have described a way to encode each ultrametric on X that is refined by d_{T_0} for a fixed bijection $\bar{X} \to X$, by a function $c : T_0 \to \mathbb{R}$ that satisfies (1)–(3). This representation is quite redundant, especially when we take into account that many values of c may coincide.

We propose a method to reduce this redundancy which develops a method proposed by Gorman and Lladser [1]. Recall that $\bar{X}$ is a compact Hausdorff space.

Theorem 4.1. *There is a unique regular measure μ on $\bar{X}$ that satisfies the conditions:*

- $\mu(\bar{X}) = 1$ *(i.e., μ is a probability measure);*
- *if $x = (x_1,\dots,x_n) \in X''$, then $\mu(x|0) = \mu(x|1)$ (each bifurcation vertex divides a measure into halves).*

For all $x \in X_\omega$ the inequality $\mu(\widehat{x}) > 0$ is valid.

We define an inner product on the space $L_2(\bar{X},\mu)$ of square integrable w.r.t. μ real-valued functions on $\bar{X}$ by the formula

$$f\cdot g=\int_{\bar{X}} f(\bar{x})g(\bar{x})\,d\mu(\bar{x}).$$

Clearly this space contains the vector space $C(\bar{X},\mathbb{R})$ of all continuous functions on $\bar{X}$.

For any ultrametric d on $\bar{X}$ refined by $\bar{d}$, the formula

$$S_d(f,g)=\int_{\bar{X}\times\bar{X}} d(x,y)f(x)g(y)\,d\mu\times\mu(x,y)$$

determines a symmetric bilinear form on $L_2(\bar{X},\mu)$. It can be used to compute distances d between all distinct points $\bar{x},\bar{y}\in\bar{X}$. The approach is somewhat similar to matrix methods used in [6] for finite spaces.

Consider the functions $u_x:\bar{X}\to\mathbb{R}$ for all $x\in X'\cup X''$ (i.e., for all vertices of T_0):

$$u_x(\bar{x})=\begin{cases}1, & \bar{x}\in\widetilde{x},\\ 0, & \bar{x}\notin\widetilde{x}.\end{cases}$$

Obviously $u_x\cdot u_y=0$ of and only if neither of x and y is a non-strict descendant of the other one. For $\bar{x}\neq\bar{y}$ in $\bar{X}$ clearly there are $x,y\in X'\cup X''$ such that $u_x(\bar{x})\neq 0$, $u_y(\bar{y})\neq 0$, $u_x\cdot u_y=0$. Then

$$d(\bar{x},\bar{y})=\frac{S_d(u_x,u_y)}{(u_x\cdot u_x)(u_y\cdot u_y)}.$$

Unfortunately the system of the functions u_x is not linearly independent, hence is not a basis of $L_2(\bar{X},\mu)$. We propose another system for this purpose.

Theorem 4.2. *The functions $w_x:\bar{X}\to\mathbb{R}$ defined for all $x\in X''$ as follows:*

$$w_x(\bar{x})=\begin{cases}\dfrac{1}{\sqrt{\mu(\widetilde{x})}}, & \bar{x}\in\widetilde{x|0},\\[2mm] \dfrac{-1}{\sqrt{\mu(\widetilde{x})}}, & \bar{x}\in\widetilde{x|1},\\[2mm] 0 & \textit{otherwise},\end{cases}$$

together with the function $u_\varepsilon\equiv 1$ form an orthonormal system with respect to the above scalar product, which is a basis in $L_2(\bar{X},\mu)$.

Proof is straightforward. These functions are analogues of *Haar-like wavelets* [3], and the following observation generalizes Theorem 2.3 [1].

Theorem 4.3. *For all $x \neq y$ in X'' the equalities $S_d(u_\varepsilon, w_x) = 0$ and $S_d(w_x, w_y) = 0$ are valid.*

This means that a countable system of numbers $S_d(u_\varepsilon, u_\varepsilon)$ and $S_d(w_x, w_x)$ for all $x \in X''$ completely describes the ultrametric. Recall that the values $c(v)$ in the previous section were set for all $v \in X' \cup X''$, i.e., this representation is formally more efficient.

5. Conclusions and future research

The above efficiency improvement for encoding of ultrametrics comes at the cost of difficulty of the inverse transformation. In particular, it is unknown what values of $S_d(u_\varepsilon, u_\varepsilon)$ and $S_d(w_x, w_x)$ for all $x \in X''$ are possible for a given T_0, to determine a valid symmetric bilinear form S_d and hence a compact ultrametric d. This will be a topic of a future publication.

We also plan to study distances between ultrametrics based on their representations by trees.

References

1. E. Gorman, M.E. Lladser, *Sparsification of large ultrametric matrices: insights into the microbial Tree of Life*, Proc. R. Soc. A, 479: 20220847 (2023). doi:10.1098/rspa.2022.0847
2. M. Krötzsch, *Generalized ultrametric upaces in quantitative domain theory*, Theor. Comput. Sci. **368** (1-2) (2006), 30-49. doi:10.1016/j.tcs.2006.05.037
3. S.G.Mallat, *A wavelet tour of signal processing: the sparse way.* Orlando, FL: Elsevier/Academic Press, 2009. https://www.di.ens.fr/ mallat/papiers/WaveletTourChap1-2-3.pdf
4. S. Nykorovych, O. Nykyforchyn, *Metric and Topology on the Poset of Compact Pseudoultrametrics*, Carpathian Math. Publ. **15** (2) (2023), 321–330. doi:10.15330/cmp.15.2.321-330
5. N. Uglešić, *On ultrametrics and equivalence relations — duality*, International Mathematical Forum. **5** (21) (2010), 1037–1048(2010). https://www.m-hikari.com/imf-2010/21-24-2010/uglesicIMF21-24-2010.pdf
6. R.S. Varga, R. Nabben, *On Symmetric Ultrametric Matrices*, in: Numerical Linear Algebra, Berlin, New York: De Gruyter, pp. 193–200 (1993). doi:10.1515/9783110857658.193

Received by Editorial Board December 16, 2025
Accepted by Editorial Board March 10, 2026

РОЗРІДЖЕННЯ КОМПАКТНИХ УЛЬТРАМЕТРИК

О.Р. Никифорчин*, **В.М. Пенгрин**
Карпатський національний університет імені Василя Стефаника;
76018, вул. Шевченка 57, м. Івано-Франківськ, Україна
e-mail: oleh.nykyforchyn@cnu.edu.ua, volodymyrpenhryn@gmail.com

Введено і досліджено відношення уточнення між компактними ультраметриками. Запропоновано економні методи задання ультраметрик функціями на бінарних деревах та симетричними білінійними формами, які набувають нескінченної діагональної форми у базі, складеній вейвлетами типу Гаара.

Ключові слова: *Ультраметрика, бінарне дерево, база Гаара.*