ULTRAMETRIC SKELETONS

MANOR MENDEL AND ASSAF NAOR

ABSTRACT. We prove that for every $\varepsilon \in (0,1)$ there exists $C_\varepsilon \in (0,\infty)$ with the following property. If (X,d) is a compact metric space and μ is a Borel probability measure on X then there exists a compact subset $S \subseteq X$ that embeds into an ultrametric space with distortion $O(1/\varepsilon)$, and a probability measure ν supported on S satisfying $\nu\left(B_d(x,r)\right) \leqslant \left(\mu(B_d(x,C_\varepsilon r)\right)^{1-\varepsilon}$ for all $x \in X$ and $r \in (0,\infty)$. The dependence of the distortion on ε is sharp. We discuss an extension of this statement to multiple measures, as well as how it implies Talagrand's majorizing measure theorem.

1. INTRODUCTION

Our main result is the following theorem.

Theorem 1.1. *For every $\varepsilon \in (0,1)$ there exists $C_\varepsilon \in (0,\infty)$ with the following property. Let (X,d) be a compact metric space and let μ be a Borel probability measure on X. Then there exists a compact subset $S \subseteq X$ satisfying*

(1) S embeds into an utrametric space with distortion $O(1/\varepsilon)$.
(2) There exists a Borel probability measure supported on S satisfying

$$\nu\left(B_d(x,r)\right) \leqslant \left(\mu(B_d(x,C_\varepsilon r)\right)^{1-\varepsilon} \tag{1}$$

for all $x \in X$ and $r \in [0,\infty)$.

Recall that an ultrametric space is a metric space (U,ρ) satisfying the strengthened triangle inequality $\rho(x,y) \leqslant \max\{\rho(x,z),\rho(y,z)\}$ for all $x,y,z \in U$. Saying that (S,d) embeds with distortion $D \in [1,\infty)$ into an ultrametric space means that there exists an ultrametric space (U,ρ) and an injection $f : S \to U$ satisfying $d(x,y) \leqslant \rho(f(x),f(y)) \leqslant Dd(x,y)$ for all $x,y \in S$. In the statement of Theorem 1.1, and in the rest of this paper, given a metric space (X,d), a point $x \in X$ and a radius $r \in [0,\infty)$, the corresponding closed ball is denoted $B_d(x,r) = \{y \in X : \ d(y,x) \leqslant r\}$, and the corresponding open ball is denoted $B_d^\circ(x,r) = \{y \in X : \ d(y,x) < r\}$. (We explicitly indicate the underlying metric since the ensuing discussion involves multiple metrics on the same set.)

We call the metric measure space (S,d,ν) from Theorem 1.1 an *ultrametric skeleton* of the metric measure space (X,d,μ). The literature contains several theorems about the existence of "large" ultrametric subsets of metric spaces; some of these results will be mentioned below. As we shall see, the subset S of Theorem 1.1 must indeed be large, but it is also geometrically "spread out" with respect to the initial probability measure μ. For example, if μ assigns positive mass to two balls $B_d(x,r)$ and $B_d(y,r)$, where $x,y \in X$ satisfy $d(x,y) > (C_\varepsilon+1)r$, then the probability measure ν, which is supported on S, cannot assign full mass to any one of these balls. This is one reason why (S,d,ν) serves as a "skeleton" of (X,d,μ).

More significantly, we call (S,d,μ) an ultrametric skeleton because it can be used to deduce global information about the entire initial metric measure space (X,d,μ). Examples of such global

M. M. was partially supported by ISF grants 221/07 and 93/11, BSF grants 2006009 and 2010021, and a gift from Cisco Research Center. A. N. was partially supported by NSF grant CCF-0832795, BSF grants 2006009 and 2010021, and the Packard Foundation. Part of this work was completed when M. M. was a visiting researcher at Microsoft Research and A. N. was visiting the Quantitative Geometry program at the Mathematical Sciences Research Institute.

applications of statements that are implied by Theorem 1.1 are described in [13, 14], and an additional example will be presented below. As a qualitative illustration of this phenomenon, consider a stochastic process $\{Z_t\}_{t\in T}$, assuming for simplicity that the index set T is finite and that each random variable Z_t has finite second moment. Equip T with the metric $d(s,t)=\sqrt{\mathbb{E}\left[(Z_s-Z_t)^2\right]}$. Assume that there exists a unique (random) point $\tau\in T$ satisfying $Z_\tau=\max_{t\in T} Z_t$. Let μ be the law of τ, and apply Theorem 1.1, say, with $\varepsilon=1/2$, to the metric measure space (X,d,μ). One obtains a subset $S\subseteq X$ that embeds into an ultrametric space with distortion $O(1)$, and a probability measure ν that is supported on S and satisfies (1) (with $\varepsilon=1/2$). If $\sigma\in S$ is a random point of S whose law is ν, then it follows that for every $x\in T$, $p\in(0,1)$ and $r\in[0,\infty)$, if σ falls into $B_d(x,r)$ with probability at least p, then the global maximum τ falls into $B_d(x,O(r))$ with probability at least p^2. One can therefore always find a subset of T that is more structured due to the fact that it is approximately an ultrametric space (e.g., such structure can be harnessed for chaining-type arguments), yet this subset reflects the location of the global maximum of $\{Z_t\}_{t\in T}$ in the above distributional/geometric sense. A quantitatively sharp variant of the above qualitative interpretation of Theorem 1.1 is discussed in Section 1.1.1 below.

1.1. **Nonlinear Dvoretzky theorems.** Nonlinear Dvoretzky theory, as initiated by Bourgain, Figiel and Milman [4], asks for theorems asserting that any "large" metric space contains a "large" subset that embeds with specified distortion into Hilbert space. We will see below examples of notions of "largeness" of a metric space for which a nonlinear Dvoretzky theorem can be proved. For an explanation of the relation of such problems to the classical Dvoretzky theorem [5], see [4, 1, 14]. Most known nonlinear Dvoretzky theorems actually obtain subsets that admit a low distortion embedding into an ultrametric space. Since ultrametric spaces admit an isometric embedding into Hilbert space [17], such a result falls into the Bourgain-Figiel-Milman framework. Often (see [4, 1]) one can prove an asymptotically matching impossibility result which shows that all subsets of a given metric space that admit a low distortion embedding into Hilbert space must be "small". Thus, in essence, it is often the case that the best way to find an almost Hilbertian subset is actually to aim for a subset satisfying the seemingly more stringent requirement of being almost ultrametric.

Apply Theorem 1.1 to an n-point metric space (X,d), with $\mu(\{x\})=1/n$ for all $x\in X$. Since ν is a probability measure on S, there exists $x\in X$ with $\nu(\{x\})\geqslant 1/|S|$. An application of (1) with $r=0$ shows that $1/|S|\leqslant \mu(\{x\})^{1-\varepsilon}=1/n^{1-\varepsilon}$, or $|S|\geqslant n^{1-\varepsilon}$. Since S embeds into an ultrametric space with distortion $O(1/\varepsilon)$, this shows that Theorem 1.1 implies the sharp solution of the Bourgain-Figiel-Milman nonlinear Dvoreztky problem that was first obtained in [13]. Sharpness in this context means that, as shown in [1], there exists a universal constant $c\in(0,\infty)$ and for every $n\in\mathbb{N}$ there exists an n-point metric space X_n such that every $S\subseteq X_n$ with $|S|\geqslant n^{1-\varepsilon}$ incurs distortion at least c/ε in any embedding into Hilbert space. Thus, the distortion bound in Theorem 1.1 cannot be improved (up to constants), even if we only require that S embeds into Hilbert space.

Assume that (X,d) is a compact metric space of Hausdorff dimension greater than $\alpha\in(0,\infty)$. Then there exists [9, 11] an α-Frostman measure on (X,d), i.e., a Borel probability measure μ satisfying $\mu(B_d(x,r))\leqslant Kr^\alpha$ for every $x\in X$ and $r\in(0,\infty)$, where K is a constants that may depend on X and α but not on x and r. An application of Theorem 1.1 to (X,d,μ) yields a compact subset $S\subseteq X$ that embeds into an ultrametric space with distortion $O(1/\varepsilon)$, and a Borel probability measure ν supported on S satisfying $\nu(B_d(x,r))\leqslant\mu(B_d(x,C_\varepsilon r))^{1-\varepsilon}\leqslant K^{1-\varepsilon}C_\varepsilon^{(1-\varepsilon)\alpha}r^{(1-\varepsilon)\alpha}$ for all $x\in X$ and $r\in(0,\infty)$. Hence ν is a $(1-\varepsilon)\alpha$-Frostman measure on S, implying [11] that S has Hausdorff dimension at least $(1-\varepsilon)\alpha$. Thus Theorem 1.1 implies the sharp solution of the nonlinear Dvoretzky problem for Hausdorff dimension that was first obtained in [14].

More generally, the following result was proved in [14] as the main step towards the solution of a question of Tao concerning the nonlinear Dvoretzky problem for Hausdorff dimension.

Theorem 1.2. *For every $\varepsilon \in (0,1)$ there exists $c_\varepsilon = e^{O(1/\varepsilon^2)} \in (0,\infty)$ with the following property. Let (X,d) be a compact metric space and let μ be a Borel probability measure on X. Then there exists a compact subset $S \subseteq X$ satisfying*

(1) S embeds into an utrametric space with distortion $O(1/\varepsilon)$.

(2) If $\{x_i\}_{i\in I} \subseteq X$ and $\{r_i\}_{i\in I} \subseteq [0,\infty)$ satisfy $\bigcup_{\in I} B_d(x_i,r_i) \supseteq S$ then

$$\sum_{i\in I} \mu\left(B_d(x_i,c_\varepsilon r_i)\right)^{1-\varepsilon} \geqslant 1. \tag{2}$$

Theorem 1.2 is a consequence of Theorem 1.1. Indeed, if $S \subseteq X$ and ν are the subset and probability measure from Theorem 1.1, then $1 = \nu(S) \leqslant \sum_{i\in I} \nu(B_d(x_i,r_i)) \leqslant \sum_{i\in I} \mu(B_d(x_i,C_\varepsilon r_i))^{1-\varepsilon}$ whenever $\bigcup_{i\in I} B_d(x_i,r_i) \supseteq S$. But, Theorem 1.2 is the main reason for the validity of the phenomenon described in Theorem 1.1: here we show how to formally deduce Theorem 1.1 from Theorem 1.2, with $C_\varepsilon = O(c_\varepsilon/\varepsilon) = e^{O(1/\varepsilon^2)}$. Alternatively, with more work, one can repeat the proof of Theorem 1.2 in [14] while making changes to several lemmas in order to prove Theorem 1.1 directly, and obtain $C_\varepsilon = c_\varepsilon$. Since the proof of Theorem 1.2 in [14] is quite involved, we believe that it is instructive to establish Theorem 1.1 via the argument described here.

1.1.1. *Majorizing measures and stochastic processes.* Theorem 1.1 makes it possible to relate the nonlinear Dvoretzky framework of [14] to Talagrand's nonlinear Dvoretzky theorem [16], and consequently to Talagrand's majorizing measure theorem [16]. Given a metric space (X,d) let $\mathscr{P}_X$ be the Borel probability measures on X. The Fernique-Talagrand γ_2 functional is defined as follows.

$$\gamma_2(X,d) = \inf_{\mu\in\mathscr{P}_X} \sup_{x\in X} \int_0^\infty \sqrt{\log\left(\frac{1}{\mu(B_d(x,r))}\right)}dr. \tag{3}$$

A nonlinear Dvoretzky-like theorem due to Talagrand [16] asserts that every finite metric space (X,d) has a subset $S \subseteq X$ that embeds into an ultrametric space with distortion $O(1)$ and[1] $\gamma_2(S,d) \gtrsim \gamma_2(X,d)$. Talagrand proved this nonlinear version of Dvoretzky's theorem in order to prove his celebrated majorizing measure theorem, which asserts that if $\{G_x\}_{x\in X}$ is a centered Gaussian process and for $x,y \in X$ we set $d(x,y) = \sqrt{\mathbb{E}\left[(G_x - G_y)^2\right]}$, then $\mathbb{E}\left[\sup_{x\in X} G_x\right] \gtrsim \gamma_2(X,d)$. There is also a simpler earlier matching upper bound on $\mathbb{E}\left[\sup_{x\in X} G_x\right]$ due to Fernique [6], so $\mathbb{E}\left[\sup_{x\in X} G_x\right] \asymp \gamma_2(X,d)$. The fact that Talagrand's nonlinear Dvoretzky theorem implies the majorizing measure theorem is simple; see [16, Prop. 13] and also the discussion in [14, Sec. 1.3].

To understand the link between Theorem 1.1 and Talagrand's nonlinear Dvoretzky theorem, consider the following quantity, associated to every compact metric space (X,d).

$$\delta_2(X,d) = \sup_{\mu\in\mathscr{P}_X} \inf_{x\in X} \int_0^\infty \sqrt{\log\left(\frac{1}{\mu(B_d(x,r))}\right)}dr. \tag{4}$$

Intuitively, $\gamma_2(X,d)$ should be viewed as a multi-scale version of a covering number, while $\delta_2(X,d)$ should be viewed as a multi-scale version of a packing number. It is therefore not surprising that $\gamma_2(X,d) \asymp \delta_2(X,d)$. In fact, in Section 3 we note that $\gamma_2(U,\rho) = \delta_2(U,\rho)$ for every finite ultrametric space (U,ρ), and $\delta_2(X,d) \geqslant \gamma_2(X,d)$ for every finite metric space (X,d) (the latter inequality is an improvement of our original bound $\delta_2(X,d) \gtrsim \gamma_2(X,d)$, due to an elegant argument of Witold Bednorz [2]). The remaining estimate $\delta_2(X,d) \lesssim \gamma_2(X,d)$ will not be needed here, though it follows from our discussion (see Remark 1.3), and it also has a simpler direct proof.

Let $\mu \in \mathscr{P}_X$ satisfy $\delta_2(X,d) = \inf_{x\in X} \int_0^\infty \sqrt{\log\left(1/\mu(B_d(x,r))\right)}dr$. Theorem 1.1 applied to (X,d,μ) yields $S \subseteq X$ and an ultrametric $\rho : S\times S \to [0,\infty)$ satisfying $d(x,y) \leqslant \rho(x,y) \leqslant K d(x,y)$

[1] Here, and in what follows, the relations $\lesssim, \gtrsim$ indicate the corresponding inequalities up to factors which are universal constants. The relation $A \asymp B$ stands for $(A \lesssim B) \wedge (A \gtrsim B)$.

for all $x, y \in S$. There also exists $\nu \in \mathscr{P}_S$ satisfying $\nu\left(B_d(x,r)\right) \leqslant \sqrt{\mu(B_d(x,Kr))}$ for all $x \in X$ and $r \in [0,\infty)$. Here $K \in (0,\infty)$ is a universal constant. Since $B_\rho(x,r) \subseteq B_d(x,r)\cap S \subseteq B_\rho(x,Kr)$ for all $x \in S$ and $r \in [0,\infty)$, we have $\delta_2(S,d) \leqslant \delta_2(S,\rho) = \gamma_2(S,\rho) \leqslant K\gamma_2(S,d)$, where we used the fact that $\gamma_2(\cdot)$ and $\delta_2(\cdot)$ coincide for ultrametrics and a simple change of variable argument. Hence,

$$K\gamma_2(S,d) \geqslant \delta_2(S,d) \geqslant \inf_{x\in S}\int_0^\infty \sqrt{\log\left(\frac{1}{\nu(B_d(x,r))}\right)}dr \geqslant \inf_{x\in S}\int_0^\infty \sqrt{\log\left(\frac{1}{\sqrt{\mu(B_d(x,Kr))}}\right)}dr$$
$$= \frac{1}{K\sqrt{2}}\inf_{x\in S}\int_0^\infty \sqrt{\log\left(\frac{1}{\mu(B_d(x,r))}\right)}dr \geqslant \frac{\delta_2(X,d)}{K\sqrt{2}} \geqslant \frac{\gamma_2(X,d)}{K\sqrt{2}}. \quad (5)$$

This completes the deduction of Talagrand's nonlinear Dvoretzky theorem from Theorem 1.1.

Remark 1.3. It is easy to check (see [16, Lem. 6]) that $\gamma_2(S,d) \leqslant 2\gamma_2(X,d)$ for every $S \subseteq X$ (and, even more trivially, $\delta_2(S,d) \leqslant \delta_2(X,d)$). Thus, it follows from (5) that $\gamma_2(X,d) \gtrsim \delta_2(X,d)$ for every finite metric space (X,d).

Since the original 1987 publication of Talagrand's majorizing measure theorem, this theorem has been reproved and simplified in several subsequent works, yielding important applications and generalizations (mainly due to Talagrand himself). These proofs are variants of the same basic idea: a greedy top-down construction, in which one looks at a given scale for a ball on which a certain functional is maximized, removes a neighborhood of this ball, and iterates this step on the remainder of the metric space. It seems that the framework described here is genuinely different. The proof of Theorem 1.2 in [14] has two phases. One first constructs a nested family of partitions in a bottom-up fashion: starting with singletons one iteratively groups the points together based on a gluing rule that is tailor-made in anticipation of the ensuing "pruning" or "sparsification" step. This second step is a top-down iterative removal of appropriately "sparse" regions of the partitions that were constructed in the first step; here one combines an analytic argument with the pigeonhole principle to show that there are sufficiently many potential pruning locations so that successive iterations of this step can be made to align appropriately. Our new approach has the advantage that it yields distributional statements such as (1), the majorizing measure theorem itself being a result of integrating these pointwise estimates.

1.2. **Multiple measures.** In anticipation of further applications of ultrametric skeletons, we end by addressing what is perhaps the simplest question that one might ask about these geometric objects: to what extent is the union of two ultrametric skeletons also an ultrametric skeleton?

We show in Remark 4.3 that for arbitrarily large $D_1, D_2 \in [1,\infty)$, one can find a finite metric space (X,d), and two disjoint subsets $U_1, U_2 \subseteq X$, such that each U_i embeds into an ultrametric space with distortion D_i, yet any embedding of $U_1 \cup U_2$ into an ultrametric space incurs distortion at least $(D_1+1)(D_2+1)-1$. In Section 4 we prove the following geometric result of independent interest (which, as explained above, is sharp up to lower order terms).

Theorem 1.4. *Fix $D_1, D_2 \in [1,\infty)$. Let (X,d) be a metric space and $U_1, U_2 \subseteq X$. Assume that (U_1,d) embeds with distortion D_1 into an ultrametric space and that (U_2,d) embeds with distortion D_2 into an ultrametric space. Then the metric space $(U_1\cup U_2,d)$ embeds with distortion at most $(D_1+2)(D_2+2)-2$ into an ultrametric space.*

Consequently, one can always find an ultrametric skeleton that is "large" with respect to any finite list of probability measures.

Corollary 1.5. *For every $\varepsilon \in (0,1)$ let C_ε be as in Theorem 1.1. Let (X,d) be a compact metric space, and let $\mu_1,\ldots,\mu_k$ be Borel probability measures on X. Then there exists a compact subset*

$S \subseteq X$, *and Borel probability measures* $\nu_1, \ldots, \nu_k$ *supported on* S, *such that* S *emebds into an ultrametric space with distortion at most* $(O(1)/\varepsilon)^k$ *and for every* $x \in X$ *and* $r \in [0, \infty)$ *we have* $\nu_i\left(B_d(x, r)\right) \leqslant \left(\mu_i(B_d(x, C_\varepsilon r)\right)^{1-\varepsilon}$ *for all* $i \in \{1, \ldots, k\}$.

2. Proof of Theorem 1.1

A *submeasure* on a set X is a function $\xi : 2^X \to [0, \infty)$ satisfying the following conditions.

(a) $\xi(\emptyset) = 0$,
(b) $A_1 \subseteq A_2 \subseteq X \implies \xi(A_1) \leqslant \xi(A_2)$,
(c) $\{A_i\}_{i\in I} \subseteq X \implies \xi\left(\bigcup_{i\in I} A_i\right) \leqslant \sum_{i\in I} \xi(A_i)$.

If in addition $\xi(X) = 1$ we call ξ a *probability submeasure.*

Lemma 2.1. *Let* (U, ρ) *be a compact ultrametric space, and let* $\xi : 2^U \to [0, \infty)$ *be a probability submeasure. Then there exists a Borel probability measure* ν *on* U *satisfying* $\nu(B_\rho(x, r)) \leqslant \xi(B_\rho(x, r))$ *for all* $x \in X$ *and* $r \in [0, \infty)$.

Remark 2.2. It is known [8, 15] that there exist probability submeasures that do not dominate any nonzero measure (in the literature such measures are called pathological submeasures). Lemma 2.1 shows that probability submeasures on ultrametric spaces always dominate *on all balls* some probability measure.

Assuming the validity of Lemma 2.1 for the moment, we prove Theorem 1.1.

Proof of Theorem 1.1. Let (X, d) be a compact metric space and μ a Borel probability measure on X. By Theorem 1.2 there exists a compact subset $S \subseteq X$ satisfying the covering estimate (2), and an ultrametric $\rho : S \times S \to [0, \infty)$ satisfying $d(x, y) \leqslant \rho(x, y) \leqslant \frac{K}{\varepsilon} d(x, y)$ for all $x, y \in S$, where K is a universal constant.

For every $A \subseteq S$ define

$$\xi(A) = \inf \left\{ \sum_{i\in I} \mu\left(B_d(x_i, c_\varepsilon r_i)\right)^{1-\varepsilon} : \ \{(x_i, r_i)\}_{i\in I} \subseteq X \times [0, \infty) \ \wedge \ \bigcup_{i\in I} B_d(x_i, r_i) \supseteq A \right\}. \tag{6}$$

In (6) the index set I can be countably infinite or finite, with the convention that an empty sum vanishes. One checks that $\xi : 2^S \to [0, \infty)$ is a submeasure on S. Moreover, for every $x \in X$ and $r \in [0, \infty)$, by considering $B_d(x, r)$ as covering itself, we deduce from (6) that

$$\xi\left(S \cap B_d(x, r)\right) \leqslant \mu\left(B_d(x, c_\varepsilon r)\right)^{1-\varepsilon}. \tag{7}$$

Since μ is a probability measure and X has bounded diameter, it follows from (7) that $\xi(S) \leqslant 1$. The covering estimate (2) implies that $\xi(S) \geqslant 1$, so in fact ξ is a probability submeasure on S.

An application of Lemma 2.1 to (S, ρ, ξ) yields a Borel probability measure ν supported on S and satisfying $\nu(B_\rho(y, r)) \leqslant \xi(B_\rho(y, r))$ for all $y \in S$ and $r \in [0, \infty)$. Fix $x \in X$ and $r \in [0, \infty)$. The desired estimate (1) holds trivially if $B_d(x, r) \cap S = \emptyset$, so we may assume that there exists $y \in S$ with $d(x, y) \leqslant r$. Thus

$$S \cap B_d(x, r) \subseteq S \cap B_d(y, 2r) \subseteq B_\rho\left(y, \frac{2K}{\varepsilon} r\right) \subseteq S \cap B_d\left(y, \frac{2K}{\varepsilon} r\right) \subseteq S \cap B_d\left(x, \left(1 + \frac{2K}{\varepsilon}\right) r\right).$$

It follows that

$$\begin{aligned}\nu(B_d(x, r)) &\leqslant \nu\left(B_\rho\left(y, \frac{2K}{\varepsilon} r\right)\right) \leqslant \xi\left(B_\rho\left(y, \frac{2K}{\varepsilon} r\right)\right) \\ &\leqslant \xi\left(S \cap B_d\left(x, \left(1 + \frac{2K}{\varepsilon}\right) r\right)\right) \leqslant \mu\left(B_d\left(x, c_\varepsilon\left(1 + \frac{2K}{\varepsilon}\right) r\right)\right)^{1-\varepsilon}.\end{aligned}$$

This completes the deduction of Theorem 1.1 from Theorem 1.2 and Lemma 2.1. $\square$

Prior to proving Lemma 2.1, we review some basic facts about compact ultrametric spaces; see [10] for an extended and more general treatment of this topic. Fix a compact ultrametric space (U,ρ). For every $r\in(0,\infty)$ we have $|\{B_\rho(x,s):\ (x,s)\in U\times[r,\infty)\}|<\infty$, i.e., there are only finitely many closed balls in U of radius at least r. Indeed, by compactness U contains only finitely many disjoint closed balls of radius at least r. Since $B_\rho(x,s)\cap B_\rho(y,t)\in\{\emptyset,B_\rho(x,s),B_\rho(y,t)\}$ for every $x,y\in U$ and $s,t\in[0,\infty)$, assuming for contradiction that $\{B_\rho(x,s):\ (x,s)\in U\times[r,\infty)\}$ is infinite, we deduce that there exist $\{(x_i,s_i)\}_{i=1}^\infty\subseteq U\times[r,\infty)$ satisfying $B_\rho(x_i,s_i)\subsetneq B_\rho(x_{i+1},s_{i+1})$ for all $i\in\mathbb{N}$. Fix $y_i\in B_\rho(x_{i+1},s_{i+1})\smallsetminus B_\rho(x_i,s_i)$. If $i<j$ then $y_j\notin B_\rho(x_j,s_j)\supseteq B_\rho(x_{i+1},s_{i+1})$ and $y_i\in B_\rho(x_{i+1},s_{i+1})$. Hence $s_{i+1}<\rho(y_j,x_{i+1})\leqslant\max\{\rho(y_j,y_i),\rho(y_i,x_{i+1})\}\leqslant\max\{\rho(y_j,y_i),s_{i+1}\}$. It follows that $\rho(y_i,y_j)>s_{i+1}\geqslant r$ for all $j>i$, contradicting the compactness of (U,ρ).

A consequence of the above discussion is that for every $x\in U$ and $r\in(0,\infty)$ there exists $\varepsilon\in(0,\infty)$ such that $B_\rho(x,r)=B_\rho(x+\varepsilon)$. Therefore $B_\rho(x,r)=B_\rho^\circ(x,r+\varepsilon/2)$. Similarly, since $B^\circ_\rho(x,r)=\bigcup_{\delta\in(0,r/2]}B_\rho(x,r-\delta)$, where there are only finitely many distinct balls appearing in this union, there exists $\delta\in(0,r/2]$ such that $B^\circ_\rho(x,r)=B_\rho(x,r-\delta)$. Thus every open ball in U of positive radius is also a closed ball, and every closed ball in U of positive radius is also an open ball. Consider the equivalence relation on U given by $x\sim y\iff\rho(x,y)<\mathrm{diam}_\rho(U)$. This is indeed an equivalence relation since ρ is an ultrametric. The corresponding equivalence classes are all of the form $B^\circ_\rho(x,\mathrm{diam}_\rho(U))$ for some $x\in U$. Being open sets that cover U, there are only finitely many such equivalence classes, say, $\{B_1^1,B^2,\ldots,B_1^{k_1}\}$. By the above discussion, each of the open balls B_1^i is also a closed ball, and hence (B_1^i,ρ) is a compact ultrametric space. We can therefore continue the above construction iteratively, obtaining a sequence $\{P_j\}_{j=0}^\infty$ of partitions of U with the following properties.

(1) $P_0=\{U\}$.
(2) P_j is finite for all j.
(3) P_{j+1} is a refinement of P_j for all j.
(4) Every $C\in P_j$ is of the form $B^\circ_\rho(x,r)$ for some $x\in U$ and $r\in[0,\infty)$.
(5) For every j, if $C\in P_j$ is not a singleton then there exists $x_1,\ldots,x_k\in U$ such that $\{B^\circ_\rho(x_i,\mathrm{diam}_\rho(C))\}_{i=1}^k\subseteq P_{j+1}$, the open balls $\{B^\circ_\rho(x_i,\mathrm{diam}_\rho(C))\}_{i=1}^k$ are disjoint, and $C=\bigcup_{i=1}^k B^\circ_\rho(x_i,\mathrm{diam}_\rho(C))$.
(6) $\lim_{j\to\infty}\max_{C\in P_j}\mathrm{diam}_\rho(C)=0$.
(7) For every $x\in U$ and $r\in(0,\infty)$ there exists j such that $B^\circ_\rho(x,r)\in P_j$.

The first five items above are valid by construction. The sixth item follows from the fact that for all $j\in\mathbb{N}$ either P_{j-1} consists of singletons or $\max_{C\in P_j}\mathrm{diam}_\rho(C)<\max_{C\in P_{j-1}}\mathrm{diam}_\rho(C)$. Since for every $r\in(0,\infty)$ there are only finitely many balls of radius at least r in U, necessarily $\lim_{j\to\infty}\max_{C\in P_j}\mathrm{diam}_\rho(C)=0$. To prove the seventh item above, assume for contradiction that $(x,r)\in U\times(0,\infty)$ is such that $B^\circ_\rho(x,r)\notin P_j$ for all j. Since the set $\{B^\circ_\rho(x,s)\}_{s\geqslant r}$ is finite, and $B^\circ_\rho(x,\mathrm{diam}(U)+1)=U\in P_0$, by replacing, if necessary, $B^\circ_\rho(x,r)$ with the largest open ball centered at x that is not contained in $\bigcup_{j=0}^\infty P_j$ we may assume without loss of generality that $B^\circ_\rho(x,s)\in\bigcup_{j=0}^\infty P_j$ for all $s\in(r,\infty)$. But since $B_\rho(x,r)=B^\circ_\rho(x,s)$ for some $s\in(r,\infty)$, it follows that $B_\rho(x,r)\in P_j$ for some j. In particular, $B^\circ_\rho(x,r)\neq B_\rho(x,r)$, implying that $\mathrm{diam}_\rho(B_\rho(x,r))=r$. Therefore by construction $B^\circ_\rho(x,r)\in P_{j+1}$, a contradiction.

Proof of Lemma 2.1. Let $\{P_j\}_{j=0}^\infty$ be the sequence of partitions of U that was constructed above. We will first define ν on $\mathcal{S}=\bigcup_{j=0}^\infty P_j\cup\{\emptyset\}$, which is the set of all open balls in U (allowing the radius to vanish, in which case the corresponding open ball is empty). Setting $\nu(X)=1$ and $\nu(\emptyset)=0$, assume inductively that ν has been defined on P_j. For $C\in P_{j+1}$ let $D\in P_j$ be the unique set satisfying $C\subseteq D$. There exist disjoint sets $C_1,\ldots,C_k\in P_{j+1}$, with $C\in\{C_1,\ldots,C_k\}$, such that $D = C_1 \cup \dots \cup C_k$. Define

$$\nu(C) = \frac{\xi(C)}{\sum_{i=1}^k \xi(C_i)} \cdot \nu(D). \tag{8}$$

This completes the inductive definition of $\nu : \mathcal{S} \to [0,\infty)$.

We claim that one can apply the Carathéodory extension theorem to extend ν to a Borel measure on U. To this end, note that $\mathcal{S}$ is a semi-ring of sets. Indeed, $\mathcal{S}$ is closed under intersection since $C \cap D \in \{\emptyset, C, D\}$ for all $C, D \in \mathcal{S}$. We therefore need to check that for every $C, D \in \mathcal{S}$, the set $D \smallsetminus C$ is a finite disjoint union of elements in $\mathcal{S}$. For this purpose we may assume that $D \smallsetminus C \neq \emptyset$, implying that $C \subsetneq D$. Assume that $C \in P_j$ and $D \in P_i$ for $i < j$. Let $C_1, \dots, C_k \in P_j$ be the distinct elements of P_j that are contained in D, enumerated so that $C = C_1$. Then $D \smallsetminus C = C_2 \cup \dots \cup C_k$, and this union is disjoint, as required.

In order to apply the Carathéodory extension theorem, it remains to check that if $\{A_i\}_{i=1}^\infty \subseteq \mathcal{S}$ are pairwise disjoint and $\bigcup_{i=1}^\infty A_i \in \mathcal{S}$, then $\nu\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \nu(A_i)$. Since all the elements of $\mathcal{S}$ are both open and closed, compactness implies that it suffices to show that if $A_1, \dots, A_m \in \mathcal{S}$ are pairwise disjoint and $A_1 \cup \dots \cup A_m \in \mathcal{S}$, then $\nu(A_1 \cup \dots \cup A_m) = \sum_{i=1}^m \nu(A_i)$. We proceed by induction on m, the case $m = 1$ being vacuous. For every $i \in \{1, \dots, m\}$ there is a unique $k_i \in \mathbb{N}$ such that $A_i \in P_{k_i - 1} \smallsetminus P_{k_i}$. Define $k = \max\{k_1, \dots, k_m\}$. If $k = 1$ then necessarily $m = 1$ and $A_1 = U$. Assume that $k > 1$ and fix $j \in \{1, \dots, m\}$ satisfying $k_j = k$. Let $D \in P_{k-2}$ be the unique element of P_{k-2} containing A_j, and let $C_1, \dots, C_\ell \in P_{k-1}$ be the distinct elements of P_{k-1} contained in D. Since $A_1 \cup \dots \cup A_m$ is a ball containing $A_j \subseteq D$, we have $A_1 \cup \dots \cup A_m \supseteq D = C_1 \cup \dots \cup C_\ell$. By maximality of k it follows that $A_j \in \{C_1, \dots, C_\ell\} \subseteq \{A_1, \dots, A_m\}$. For $i \in \{1, \dots, \ell\}$ let $n_i \in \{1, \dots, m\}$ be such that $C_i = A_{n_i}$. Since $\left(\bigcup_{i \in \{1,\dots,m\} \smallsetminus \{n_1,\dots,n_\ell\}} A_i\right) \bigcup D = \bigcup_{i=1}^m A_i$, the inductive hypothesis implies that $\sum_{i \in \{1,\dots,m\} \smallsetminus \{n_1,\dots,n_\ell\}} \nu(A_i) + \nu(D) = \nu(A_1 \cup \dots \cup A_m)$. But by our definition (8) we have $\nu(D) = \nu(A_{n_1}) + \dots + \nu(A_{n_\ell})$, so that indeed $\nu(A_1 \cup \dots \cup A_m) = \sum_{i=1}^m \nu(A_i)$.

Having defined the Borel probability measure ν, it remains to check by induction on j that if $C \in P_j$ then $\nu(C) \leqslant \xi(C)$. If $j = 0$ then $C = U$ and $\nu(U) = \xi(U) = 1$. If $j \geqslant 1$ then let $D \in P_{j-1}$ satisfy $C \subseteq D$. There exist disjoint sets $C_1, \dots, C_k \in P_{j+1}$, with $C \in \{C_1, \dots, C_k\}$, such that $D = C_1 \cup \dots \cup C_k$. Since ξ is a submeasure, $\xi(D) \leqslant \xi(C_1) + \dots + \xi(C_k)$. By the inductive hypothesis $\nu(D) \leqslant \xi(D)$. Our definition (8) now implies that $\nu(C) \leqslant \xi(C)$. The proof of Lemma 2.1 is complete. $\square$

3. $\gamma_2(X,d)$ and $\delta_2(X,d)$

Let (X,d) be a finite metric space. For every measurable $\phi : (0,\infty) \to [0,\infty)$ define

$$\gamma_\phi(X,d) = \inf_{\mu \in \mathscr{P}_X} \sup_{x \in X} \int_0^\infty \phi(\mu(B_d(x,r)))dr,$$

and

$$\delta_\phi(X,d) = \sup_{\mu \in \mathscr{P}_X} \inf_{x \in X} \int_0^\infty \phi(\mu(B_d(x,r)))dr.$$

Thus $\gamma_2(\cdot) = \gamma_\phi(\cdot)$ and $\delta_2(\cdot) = \delta_\phi(\cdot)$ for $\phi(x) = \sqrt{\log(1/x)}$. The following lemma is a variant of an argument of Bednorz [2, Lem. 4]. The elegant proof below was shown to us by Keith Ball; it is a generalization and a major simplification of our original proof of the estimate $\delta_2(X,d) \gtrsim \gamma_2(X,d)$.

Lemma 3.1. *Assume that $\phi : (0,\infty) \to (0,\infty)$ is continuous and $\lim_{x \to 0^+} \phi(x) = \infty$. Then $\delta_\phi(X,d) \geqslant \gamma_\phi(X,d)$.*

Proof. Write $X = \{x_1, \dots, x_n\}$. Thus $\mathscr{P}_X$ can be identified with the $(n-1)$-dimensional simplex $\Delta_{n-1} = \{(\mu_1, \dots, \mu_n) \in [0,1]^n;\ \mu_1 + \dots + \mu_n = 1\}$ (by setting $\mu\{x_i\}) = \mu_i$).

Define $f_1,\ldots,f_n:\Delta_{n-1}\to[0,\infty)$ by

$$f_i(\mu)=\begin{cases}0 & \text{if } \mu_i=0,\\ \left(1+\int_0^\infty \phi(\mu(B(x_i,r)))dr\right)^{-1} & \text{if } \mu_i>0.\end{cases}$$

Writing $S(\mu)=\sum_{i=1}^n f_i(\mu)$, we define $F:\Delta_{n-1}\to\Delta_{n-1}$ by $F(\mu)=(f_1(\mu),\ldots,f_n(\mu))/S(\mu)$. Since ϕ is continuous and $\lim_{x\to0^+}\phi(x)=\infty$, all the f_i are continuous on Δ_{n-1}. Since each $\mu\in\Delta_{n-1}$ has at least one positive coordinate, $S(\mu)>0$. Thus F is continuous. Note that by definition F maps each face of Δ_{n-1} into itself. By a standard reformulation of the Brouwer fixed point theorem (see, e.g., [7, Sec. $4.29\frac{1}{2}{}_+$]), it follows that $f(\Delta_{n-1})=\Delta_{n-1}$. In particular, there exists $\mu\in\Delta_{n-1}$ for which $F(\mu)=(1/n,\ldots,1/n)$. In other words, there exists $\mu\in\mathscr{P}_X$ such that $\int_0^\infty\phi(\mu(B_d(x,r)))dr$ does not depend on $x\in X$. Hence,

$$\delta_\phi(X,d)\geqslant\inf_{x\in X}\int_0^\infty\phi(\mu(B_d(x,r)))dr=\sup_{x\in X}\int_0^\infty\phi(\mu(B_d(x,r)))dr\geqslant\gamma_\phi(X,d).\qquad\square$$

Lemma 3.2. *Assume that $\phi:(0,\infty)\to[0,\infty)$ is non-increasing. Let (U,ρ) be a finite ultrametric space. Then $\delta_\phi(U,\rho)\leqslant\gamma_\phi(U,\rho)$.*

Proof. We claim that if μ,ν are nonnegative measures on U satisfying $\mu(U)\leqslant\nu(U)$ then there exists $a\in U$ satisfying $\mu(B_\rho(a,r))\leqslant\nu(B_\rho(a,r))$ for all $r\in(0,\infty)$. This would imply the desired estimate since if $\mu,\nu\in\mathscr{P}_X$ are chosen so that $\sup_{x\in X}\int_0^\infty\phi(\mu(B_\rho(x,r)))dr=\gamma_\phi(U,\rho)$ and $\inf_{x\in X}\int_0^\infty\phi(\mu(B_\rho(x,r)))dr=\delta_\phi(U,\rho)$, then

$$\gamma_\phi(U,\rho)\geqslant\int_0^\infty\phi(\mu(B_\rho(a,r)))dr\geqslant\int_0^\infty\phi(\nu(B_\rho(a,r)))dr\geqslant\delta_\phi(U,\rho).$$

The proof of the existence of $a\in U$ is by induction on $|U|$. If $|U|=1$ there is nothing to prove. Otherwise, as explained in Section 2, there exist $x_1,\ldots,x_k\in U$ such that the balls $\{B_\rho^\circ(x_i,\operatorname{diam}_\rho(U))\}_{i=1}^k$ are nonempty, pairwise disjoint, and $\bigcup_{i=1}^k B_\rho^\circ(x_i,\operatorname{diam}_\rho(U))=U$. It follows that $\sum_{i=1}^k\mu(B_\rho^\circ(x_i,\operatorname{diam}_\rho(U)))=\mu(U)\leqslant\nu(U)=\sum_{i=1}^k\nu(B_\rho^\circ(x_i,\operatorname{diam}_\rho(U)))$. Consequently there exists $i\in\{1,\ldots,k\}$ such that $\mu(B_\rho^\circ(x_i,\operatorname{diam}_\rho(U)))\leqslant\nu(B_\rho^\circ(x_i,\operatorname{diam}_\rho(U)))$. By the inductive hypothesis there exists $a\in B_\rho^\circ(x_i,\operatorname{diam}_\rho(U))$ satisfying $\mu(B_\rho(a,r))\leqslant\nu(B_\rho(a,r))$ for all $r<\operatorname{diam}_\rho(U)$. Since for $r\geqslant\operatorname{diam}_\rho(U)$ we have $B_\rho(a,r)=U$, the proof is complete. $\square$

A combination of Lemma 3.1 and Lemma 3.2 yields the following corollary.

Corollary 3.3. *If $\phi:(0,\infty)\to[0,\infty)$ is non-increasing, continuous, and $\lim_{x\to0^+}\phi(x)=\infty$, then $\delta_\phi(U,\rho)=\gamma_\phi(U,\rho)$ for all finite ultrametric spaces (U,ρ).*

Remark 3.4. Consider the star metric d_n on $\{0,1,\ldots,n\}$, i.e., $d_n(0,i)=1$ for all $i\in\{1,\ldots,n\}$ and $d_n(p,q)=2$ for all distinct $p,q\in\{1,\ldots,n\}$. The measure ν on $\{0,1,\ldots,n\}$ given by $\nu(\{0\})=0$ and $\nu(\{i\})=1/n$ for $i\in\{1,\ldots,n\}$, shows that $\delta_2(\{0,1,\ldots,n\},d_n)\geqslant2\sqrt{\log n}$. At the same time, the measure μ on $\{0,1,\ldots,n\}$ given by $\mu(\{0\})=1/2$ and $\mu(\{i\})=1/(2n)$ for $i\in\{1,\ldots,n\}$, shows that $\gamma_2(\{0,1,\ldots,n\},d_n)\leqslant\sqrt{\log(2n)}+\sqrt{\log(2n/(n+1))}\leqslant(1/2+o(1))\delta_2(\{0,1,\ldots,n\},d_n)$. Thus, unlike the case of ultrametric spaces, for general metric spaces it is not always true that $\gamma_2(X,d)=\delta_2(X,d)$. Of course, due to Lemma 3.1 and Remark 1.3 we know that $\gamma_2(X,d)\asymp\delta_2(X,d)$. It seems plausible that always $\delta_2(X,d)\leqslant2\gamma_2(X,d)$, but we do not investigate this here.

4. Unions of approximate ultrametrics

In this section we prove Theorem 1.4 and present some related examples. Below, given a partition P of a set X, for $x\in X$ we denote by $P(x)$ the element of P to which x belongs.

Lemma 4.1. *Fix $D_1, D_2 \geqslant 1$. Let (X,d) be a metric space and let $U_1, U_2 \subseteq X$ be two bounded subsets of X. Assume that (U_1,d) embeds with distortion D_1 into an ultrametric space and that (U_2,d) embeds with distortion D_2 into an ultrametric space. Then for every $\varepsilon \in (0,1)$ there is a partition P of $U_1 \cup U_2$ with the following properties.*

- *For every $C \in P$,*

$$\operatorname{diam}_d(C) \leqslant (1-\delta)\operatorname{diam}_d(U_1 \cup U_2), \tag{9}$$

where

$$\delta \stackrel{\mathrm{def}}{=} \frac{2\varepsilon D_2}{(D_1D_2+2D_1+2D_2+2)(D_1D_2+2D_1+2D_2+2+\varepsilon)}. \tag{10}$$

- *For every distinct $C_1, C_2 \in P$,*

$$d(C_1,C_2) \geqslant \frac{\operatorname{diam}_d(U_1 \cup U_2)}{D_1D_2+2D_1+2D_2+2+\varepsilon}. \tag{11}$$

Proof. By rescaling we may assume that $\operatorname{diam}_d(U_1 \cup U_2) = 1$. Let ρ_1 be an ultrametric on U_1 satisfying $d(x,y) \leqslant \rho_1(x,y) \leqslant D_1 d(x,y)$ for all $x,y \in U_1$. Define

$$a \stackrel{\mathrm{def}}{=} \frac{D_1D_2+2D_1}{D_1D_2+2D_1+2D_2+2}, \tag{12}$$

and consider the equivalence relation on U_1 given by $x \sim_1 y \iff \rho_1(x,y) \leqslant a$ (this is an equivalence relation since ρ_1 is an ultrametric). Let $\{E_i\}_{i\in I} \subseteq 2^{U_1}$ be the corresponding equivalence classes. Thus

$$\operatorname{diam}_d(E_i) \leqslant \operatorname{diam}_{\rho_1}(E_i) \leqslant a \tag{13}$$

for all $i \in I$, and for distinct $i,j \in I$ we have

$$d(E_i,E_j) \geqslant \frac{\rho_1(E_i,E_j)}{D_1} \geqslant \frac{a}{D_1}. \tag{14}$$

Let ρ_2 be an ultrametric on U_2 satisfying $d(x,y) \leqslant \rho_2(x,y) \leqslant D_2 d(x,y)$ for all $x,y \in U_2$. Define

$$b \stackrel{\mathrm{def}}{=} \frac{D_2}{D_1D_2+2D_1+2D_2+2+\varepsilon}, \tag{15}$$

and consider similarly the equivalence relation on U_2 given by $x \sim_2 y \iff \rho_2(x,y) \leqslant b$. The corresponding equivalence classes will be denoted $\{F_j\}_{j\in J} \subseteq 2^{U_2}$. Thus

$$\operatorname{diam}_d(F_i) \leqslant \operatorname{diam}_{\rho_1}(F_i) \leqslant b \tag{16}$$

for all $i \in J$, and for distinct $i,j \in J$ we have

$$d(F_i,F_j) \geqslant \frac{\rho_2(F_i,F_j)}{D_2} \geqslant \frac{b}{D_2}. \tag{17}$$

For every $i \in I$ denote

$$J_i \stackrel{\mathrm{def}}{=} \{j \in J:\ d(E_i,F_j) \leqslant c\}, \tag{18}$$

where

$$c \stackrel{\mathrm{def}}{=} \frac{1}{D_1D_2+2D_1+2D_2+2}. \tag{19}$$

Note that for every $j \in J$ there is at most one $i \in I$ for which $j \in J_i$. Indeed, if $j \in J_i \cap J_\ell$, where $i \neq \ell$, then

$$\begin{aligned} d(E_i, E_\ell) &\leqslant d(E_i, F_j) + \operatorname{diam}_d(F_j) + d(F_j, E_\ell) \\ &\overset{(16)\wedge(18)}{\leqslant} 2c + b \\ &\overset{(15)\wedge(19)}{=} \frac{2}{D_1D_2 + 2D_1 + 2D_2 + 2} + \frac{D_2}{D_1D_2 + 2D_1 + 2D_2 + 2 + \varepsilon} \\ &< \frac{2 + D_2}{D_1D_2 + 2D_1 + 2D_2 + 2} \overset{(12)}{=} \frac{a}{D_1}, \end{aligned}$$

Contradicting (14).

Consider the partition P of $U_1 \cup U_2$ consisting of the sets

$$\left\{ E_i \cup \left(\bigcup_{j \in J_i} F_j \right) \right\}_{i \in I} \quad \text{and} \quad \{F_j \smallsetminus U_1\}_{j \in J \smallsetminus \left(\bigcup_{i \in I} J_i\right)}.$$

It follows from (14), (17) and (18) that for every distinct $C_1, C_2 \in P$,

$$d(C_1, C_2) \geqslant \min\left\{ \frac{a}{D_1}, \frac{b}{D_2}, c \right\} = \frac{1}{D_1D_2 + 2D_1 + 2D_2 + 2 + \varepsilon}.$$

Thus the partition P satisfies (11). It also follows from (13), (16) and (18) that for every $C \in P$,

$$\operatorname{diam}_d(C) \leqslant a + 2b + 2c = (1 - \delta),$$

where we used the definitions (10), (12), (15), (19). Thus the partition P satisfies (9), completing the proof of Lemma 4.1. □

Proof of Theorem 1.4. Assume first that U_1, U_2 are bounded. Define a sequence $\{P_k\}_{k=0}^\infty$ of partitions of $U_1 \cup U_2$ as follows. Start with the trivial partition $P_0 = \{U_1 \cup U_2\}$, and having defined P_k, the partition P_{k+1} is obtained by applying Lemma 4.1 to the sets $U_1 \cap C$ and $U_2 \cap C$ for each $C \in P_k$. Then the partitions $\{P_k\}_{k=0}^\infty$ have the following properties.

- P_{k+1} is a refinement of P_k,
- for every $C \in P_k$ we have
$$\operatorname{diam}_d(C) \leqslant (1 - \delta)^k \operatorname{diam}_d(U_1 \cup U_2), \tag{20}$$
- for every distinct $C_1, C_2 \in P_{k+1}$ such that $C_1, C_2 \subseteq C$ for some $C \in P_k$, we have
$$d(C_1, C_2) \geqslant \frac{\operatorname{diam}_d(C)}{D_1D_2 + 2D_1 + 2D_2 + 2 + \varepsilon}. \tag{21}$$

It follows from (20) that for every distinct $x, y \in U_1 \cup U_2$ we have $P_k(x) \neq P_k(y)$ for $k \geqslant 0$ large enough. Thus for distinct $x, y \in U_1 \cup U_2$ let $k(x, y)$ denote the largest integer $k \geqslant 0$ such that $P_k(x) = P_k(y)$. Define

$$\rho(x, y) = \begin{cases} \operatorname{diam}_d\left(P_{k(x,y)}(x)\right) & x \neq y, \\ 0 & x = y. \end{cases}$$

Then ρ is an ultrametric on $U_1 \cup U_2$. Indeed, for distinct $x, y, z \in U_1 \cup U_2$ let $k \geqslant 0$ be the largest integer such that $P_k(x) = P_k(y) = P_k(z)$. Then $k = \min\{k(x, z), k(y, z)\}$ and $P_k(x) \supseteq P_{k(x,y)}(x)$, implying that $\rho(x, y) = \operatorname{diam}_d\left(P_{k(x,y)}(x)\right) \leqslant \operatorname{diam}_d\left(P_k(x)\right) = \max\{\rho(x, z), \rho(y, z)\}$. For distinct

$x, y \in U_1 \cup U_2$, since $x, y \in P_{k(x,y)}(x)$, we have $\rho(x,y) = \mathrm{diam}_d\left(P_{k(x,y)}(x)\right) \geqslant d(x,y)$, while since $P_{k(x,y)+1}(x) \neq P_{k(x,y)+1}(y)$, we deduce from (21) that

$$d(x,y) \geqslant d\left(P_{k(x,y)+1}(x), P_{k(x,y)+1}(y)\right) \geqslant \frac{\mathrm{diam}_d\left(P_{k(x,y)}(x)\right)}{D_1D_2 + 2D_1 + 2D_2 + 2 + \varepsilon} = \frac{\rho(x,y)}{D_1D_2 + 2D_1 + 2D_2 + 2 + \varepsilon}.$$

The above argument shows that if U_1, U_2 are bounded, the metric space $(U_1 \cup U_2, d)$ embeds with distortion $D_1D_2 + 2D_1 + 2D_2 + 2 + \varepsilon$ into an ultrametric space for every $\varepsilon \in (0,1)$.

For possibly unbounded $U_1, U_2 \subseteq X$, fix $x_0 \in X$, and for every $n \in \mathbb{N}$ let ρ_n be an ultrametric on $(U_1 \cap B_d(x_0,n)) \cup (U_2 \cap B_d(x_0,n))$ satisfying

$$d(x,y) \leqslant \rho_n(x,y) \leqslant \left(D_1D_2 + 2D_1 + 2D_2 + 2 + \frac{1}{n}\right) d(x,y)$$

for all $x, y \in (U_1 \cap B_d(x_0,n)) \cup (U_2 \cap B_d(x_0,n))$. Define also $\rho_n(x,y) = 0$ if $\{x,y\}$ is not contained in $(U_1 \cap B_d(x_0,n)) \cup (U_2 \cap B_d(x_0,n))$. Let $\mathscr{U}$ be a free ultrafilter on $\mathbb{N}$ and set

$$\rho_\infty(x,y) \stackrel{\mathrm{def}}{=} \lim_{n\to\mathscr{U}} \rho_n(x,y).$$

Then ρ_∞ is an ultrametric on $U_1 \cup U_2$ satisfying $d \leqslant \rho_\infty \leqslant (D_1D_2 + 2D_1 + 2D_2 + 2)d$. $\square$

Remark 4.2. There are several interesting variants of the problem studied in Theorem 1.4. For example, answering our initial question, Konstantin Makarychev and Yury Makarychev proved (private communication) that if (X,d) is a metric space and $E_1, E_2 \subseteq X$ embed into Hilbert space with distortion $D \geqslant 1$ then $(E_1 \cup E_2, d)$ embeds into Hilbert space with distortion $f(D)$. Their argument crucially uses Hilbert space geometry, and therefore the following natural question remains open: if (X,d) is a metric space and $E_1, E_2 \subseteq X$ embed into L_1 with distortion $D \geqslant 1$, does it follow that $(E_1 \cup E_2, d)$ embeds into L_1 with distortion $f(D)$? Regarding unions of more than two subsets, perhaps even the following (ambitious) question has a positive answer: if $E_1, \ldots, E_n \subseteq X$ embed into Hilbert space with distortion $D \geqslant 1$, does it follow that $(E_1 \cup \ldots \cup E_n, d)$ embeds into Hilbert space with distortion $O(\log n) f(D)$? If true, this statement (in the isometric case $D = 1$) would yield a very interesting strengthening of Bourgain's embedding theorem [3], which asserts that any n-point metric space embeds into Hilbert space with distortion $O(\log n)$.

Remark 4.3. The following example shows that Theorem 1.4 is sharp up to lower order terms. Fix two integers $M, N \geqslant 2$, and write $MN = K(M+N)+L$, where $K \in \mathbb{N}$ and $L \in \{0, 1, \ldots, M+N-1\}$. Consider the following two subsets of the real line:

$$U_1 \stackrel{\mathrm{def}}{=} \bigcup_{i=0}^{K-1} \{i(M+N), i(M+N)+1, \ldots, i(M+N)+M-1\},$$
$$U_2 \stackrel{\mathrm{def}}{=} \bigcup_{i=0}^{K-1} \{i(M+N)+M, i(M+N)+M+1, \ldots, (i+1)(M+N)-1\}.$$

For $i \in \{1,2\}$, let $D_i \geqslant 1$ be the best possible distortion of U_i (with the metric inherited from $\mathbb{R}$) in an ultrametric space. For $i_1, i_2 \in \{0, \ldots, K-1\}$ and $j_1, j_2 \in \{0, \ldots, M-1\}$ define

$$\rho_1\left(i_1(M+N)+j_1, i_2(M+N)+j_2\right) \stackrel{\mathrm{def}}{=} \begin{cases} 0 & i_1 = i_2 \ \wedge \ j_1 = j_2, \\ M-1 & i_1 = i_2 \ \wedge \ j_1 \neq j_2, \\ (K-1)(M+N)+M-1 & i_1 \neq i_2. \end{cases}$$

Then ρ_1 is an ultrametric on U_1 satisfying $|x-y| \leqslant \rho_1(x,y) \leqslant (M-1)|x-y|$ for all $x,y \in U_1$. Hence $D_1 \leqslant M-1$. Similarly, for $i_1, i_2 \in \{0,\ldots,K-1\}$ and $j_1, j_2 \in \{M,\ldots,M+N-1\}$ define

$$\rho_2\left(i_1(M+N)+j_1, i_2(M+N)+j_2\right) \stackrel{\mathrm{def}}{=} \begin{cases} 0 & i_1 = i_2 \ \wedge \ j_1 = j_2, \\ N-1 & i_1 = i_2 \ \wedge \ j_1 \neq j_2, \\ (K-1)(M+N)+N-1 & i_1 \neq i_2. \end{cases}$$

Then ρ_2 is an ultrametric on U_2 satisfying $|x-y| \leqslant \rho_1(x,y) \leqslant (N-1)|x-y|$ for all $x,y \in U_2$. Hence $D_2 \leqslant N-1$. But $U_1 \cup U_2 = \{0,1,\ldots,K(M+N)-1\}$, and hence any embedding of $U_1 \cup U_2$ into an ultrametric space incurs distortion at least $K(M+N)-1$ (see for example [12, Lem. 2.4]). Observe that this lower bound on the distortion equals $MN-L-1 \geqslant (M-1)(N-1)-1 \geqslant D_1D_2-1$. When $L=0$ (e.g., when $M=N=2S$ or $M=2N=6S$ for some $S \in \mathbb{N}$), the above distortion lower bound becomes $MN-L-1 \geqslant (D_1+1)(D_2+1)-1$. Thus one cannot improve the bound in Theorem 1.4 to D_1D_2, i.e., additional lower order terms are necessary.

Acknowledgements. We are grateful to the following people for helpful suggestions: Keith Ball, Witold Bednorz, Rafał Latała, Gilles Pisier, Gideon Schechtman, Michel Talagrand.

References

[1] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric Ramsey-type phenomena. *Ann. of Math. (2)*, 162(2):643–709, 2005.
[2] W. Bednorz. The measure approach to sample boundedness. Preprint, 2011.
[3] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.
[4] J. Bourgain, T. Figiel, and V. Milman. On Hilbertian subsets of finite metric spaces. *Israel J. Math.*, 55(2):147–152, 1986.
[5] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, pages 123–160. Jerusalem Academic Press, Jerusalem, 1961.
[6] X. Fernique. Évaluations de processus gaussiens composés. In *Probability in Banach spaces (Proc. First Internat. Conf., Oberwolfach, 1975)*, pages 67–83. Lecture Notes in Math., Vol. 526. Springer, Berlin, 1976.
[7] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Modern Birkhäuser Classics. Birkhäuser Boston Inc., Boston, MA, english edition, 2007. Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates.
[8] W. Herer and J. P. R. Christensen. On the existence of pathological submeasures and the construction of exotic topological groups. *Math. Ann.*, 213:203–210, 1975.
[9] J. D. Howroyd. On dimension and on the existence of sets of finite positive Hausdorff measure. *Proc. London Math. Soc. (3)*, 70(3):581–604, 1995.
[10] B. Hughes. Trees and ultrametric spaces: a categorical equivalence. *Adv. Math.*, 189(1):148–191, 2004.
[11] P. Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Fractals and rectifiability.
[12] M. Mendel and A. Naor. Euclidean quotients of finite metric spaces. *Adv. Math.*, 189(2):451–494, 2004.
[13] M. Mendel and A. Naor. Ramsey partitions and proximity data structures. *J. Eur. Math. Soc.*, 9(2):253–275, 2007.
[14] M. Mendel and A. Naor. Ultrametric subsets with large Hausdorff dimension. To appear in Invent. Math., preprint available at `http://arxiv.org/abs/1106.0879`, 2011.
[15] M. Talagrand. A simple example of pathological submeasure. *Math. Ann.*, 252(2):97–102, 1979/80.
[16] M. Talagrand. Regularity of Gaussian processes. *Acta Math.*, 159(1-2):99–149, 1987.
[17] I. A. Vestfrid and A. F. Timan. A universality property of Hilbert spaces. *Dokl. Akad. Nauk SSSR*, 246(3):528–530, 1979.

Mathematics and Computer Science Department, Open University of Israel, 1 University Road, P.O. Box 808 Raanana 43107, Israel
E-mail address: `mendelma@gmail.com`

Courant Institute, New York University, 251 Mercer Street, New York NY 10012, USA
E-mail address: `naor@cims.nyu.edu`