Local versus Global Properties of Metric Spaces
Extended abstract

Sanjeev Arora* László Lovász† Ilan Newman‡ Yuval Rabani§ Yuri Rabinovich¶
Santosh Vempala‖

Abstract

Motivated by applications in combinatorial optimization, we initiate a study of the extent to which the global properties of a metric space (especially, embeddability in ℓ_1 with low distortion) are determined by the properties of small subspaces. We note connections to similar issues studied already in Ramsey theory, complexity theory (especially PCPs), and property testing. We prove both upper bounds and lower bounds on the distortion of embedding locally constrained metrics into various target spaces.

1 Introduction

Suppose that we are given a finite metric space (X, d) and we are told that the induced metric on every small subset of X embeds isometrically into ℓ_1. What can we say about the distortion of embedding the entire metric into ℓ_1? In this paper we initiate the study of this question and similar questions.

One reason to study such problems is that certain embedding questions are intimately related to problems in combinatorial optimization. In particular, finite ℓ_1 metrics correspond exactly to combinations of cuts (see the book [10]). Approximation algorithms for NP-hard cut problems such as *sparsest cut* are derived by embedding general metric spaces into ℓ_1 [18, 5], and more recently, negative type metrics into ℓ_1 [4, 7, 3]. Conceivably, the relaxations underlying the results for sparsest cut could be tightened by restricting them to metrics that have the property that every subset of size k embeds isometrically into ℓ_1 (where k is either a constant or a slowly growing function of the input size). Interestingly, similar constraints arise when applying k rounds of a lift-and-project method such as Lovasz-Schrijver or Sherali-Adams. These relaxations can be computed in $n^{O(k)}$ time, where n is the number of vertices of the input graph. (Such observations were made in a recent paper [2], where it was observed that studying such questions leads to the study of local versus global structure. That paper restricted attention to vertex cover, however.)

We show (Theorem 3.1) that if every subset of size $\frac{n}{c}$ of an n-point metric space embeds isometrically into ℓ_1, then the entire space embeds into ℓ_1 with distortion $O(c^2)$. (The result also holds if the isometric embedding of subsets is replaced by low distortion embedding.) By the above discussion, for any $c = c(n) \to \infty$, we get a $2^{o(n)}$-time $O(c^2)$-approximation algorithm for sparsest cut. (Recent reductions [8] show that such approximation is hard, assuming the unique games conjecture [19].)

On the other hand, we construct (Theorem 3.5) metric spaces where every constant sized subset embeds

*Computer Science Department, Princeton University. Supported by David and Lucile Packard Fellowship and NSF grant CCR 0205594. Email: arora@cs.princeton.edu

†Microsoft Research and Eötvös Loránd University. Email: lovasz@microsoft.com

‡Computer Science Department, University of Haifa, Haifa 31905, Israel. Email: ilan@cs.haifa.ac.il.

§Computer Science Department, Technion, Haifa 32000, Israel. Email: rabani@cs.technion.ac.il.

¶Computer Science Department, University of Haifa, Haifa 31905, Israel. Supported in part by a grant ISF-247-02-10.5. Email: yuri@cs.haifa.ac.il.

‖Mathematics Department, MIT. Supported in part by NSF CCR-0312339 and a Sloan foundation fellowship. Email: vempala@math.mit.edu

isometrically into ℓ_1, yet the entire space incurs large distortion when embedded into ℓ_1. In fact, in our proof the distortion remains super-constant as long as the subset size remains $o(\log\log n)$. This restricts the possibility of designing polynomial time approximation algorithms for sparsest cut with significantly better guarantees using some lift-and-project methods.

Thus far we were insisting that subsets have to be *isometrically* embeddable into ℓ_1, and in this setting there is a large gap between our upper bounds and our lower bounds[1]. However, it is also interesting to consider the case where subsets are embeddable with constant distortion into ℓ_1. After all, many plausible ideas for inferring global structure from local structure would try to relate the distortion of the entire space to the distortion of (the metric induced on) small subsets. Upon thus relaxing the question, the gap between upper and lowerbounds nearly vanishes. We construct (Theorem 3.4) n-point metrics that require $\Omega(\log n)$ distortion to embed into ℓ_1, but whose every subset of size $n^{1-\epsilon}$ embeds into ℓ_1 (or even into restricted subclasses of ℓ_1) with distortion $O(1/\epsilon^2)$. We note that such constructions are nontrivial precisely because we need a fairly strong property to hold for *every* subset. A significant contribution of our work is a new insight into shortest-path metrics derived from random graphs of bounded degree, which are used in most of our lower bound results. These metrics were shown to be extremal for many metric-theoretic properties in the past. Surprisingly, their local structure turns out to be rather simple, even when the size of the sub-metrics is as large as $n^{1-\epsilon}$ (see Section 3.2).

Our results are also related to work on Ramsey phenomena in metric spaces, a line of work motivated both by applications to lower bounds in on-line computation and by deep questions about the local theory of metric spaces (see [6] and the references therein) and our lower-bound of Section 3.2 answers an open question of [6]. Ramsey theory in general shows that in the midst of global "disorder" there is always a significant subset exhibiting "order." In this phrasing, our upper bounds trivially imply an upper bound on the size of the smallest "disordered" subset, assuming global "disorder." For example, if a metric space does not embed into ℓ_1 with distortion at most α, there must be a subspace of cardinality $O(n\sqrt{\beta/\alpha})$ that does not embed into ℓ_1 with distortion less than β. In fact, we show that there are many such subspaces, and one can construct, independently of the metric, a poly(n) sized set of candidate subspaces to check.

[1] By the terms "upper bound" or "lower bound" we mean an upper bound or a lower bound on the distortion of embedding metrics of a certain class into some target space.

Local versus global questions also play an important role in areas such as the construction of probabilistically checkable proofs, program checking, property testing, etc. In those settings one has to infer a global property from the knowledge that the "local" property only has to hold for *many* local neighborhoods. Our study has a very similar feel, except we are interested in inferring global properties when the local property holds for *all* local neighborhoods, not just most.

The results discussed so far carry over (with small changes) to ℓ_2 as well. The upper bounds hold, in fact, not only for ℓ_1 but also for many other classes of metrics, such as polygonal metrics, hypermetrics, and negative type metrics. We introduce the notion of a *baseline* class of metrics, obtained by postulating some properties shared by the above examples. These properties are sufficient for proving the upper bounds. We show (Theorem 5.2) that the class of all metrics is "far" from any non-trivial class of baseline metrics, in the sense that there are metrics that do not embed into such a class with bounded distortion. Our proof uses the notion of a *forbidden sub-metric*, akin in some respects to the topological notion of a forbidden minor. For every non-trivial class of baseline metrics there is a fixed size metric that does not embed into any member of the class with distortion below some constant. This is the starting point of our asymptotic bound. (See also [20].) On the other hand, our lower bounds on the distortion of embedding into ℓ_1 imply that polygonal metrics are also "far" (in the same sense) from ℓ_1. Our results motivate future investigation into the potential use of baseline metrics in approximation algorithms.

Finally, we also study ultrametrics, a class of metrics that is used in hierarchical clustering and metric Ramsey theory (see [13, 12, 6] and the references therein). These are metrics that satisfy the condition $\forall x, y, z\ d(x, z) = \max\{d(x, y), d(z, y)\}$. In particular, they are a (very restricted) subset of ℓ_2 metrics. By definition, if every subset of size three is an ultrametric, then so is the whole space. On the other hand, we show (Theorem 4.2) that the situation changes dramatically if we relax the requirement on the subsets. We construct for every c and ϵ metric spaces on n points such that every subset of cardinality n^ϵ embeds into an ultrametric with distortion bounded by c, yet the entire metric space does not embed into an ultrametric with distortion less than $c^{1/\epsilon}$. We show that this bound is tight by a establishing a matching upper bound on the distortion.

It should be stressed that the most important open problem arising from our work is to construct metrics that require large distortion to embed into ℓ_1, but where every subset of size at most, say, n^ϵ (or even $\Omega(\log n)$)

embeds *isometrically* (as opposed to embedding with low distortion) into ℓ_1. Our work on ultrametrics indicates that the two conditions may vary considerably in their behavior. Thus the possibility remains of improving the recent $O(\sqrt{\log n})$ approximation guarantees for sparsest cut and other problems via the lift-and-project approach outlined above. Our results do not rule out, say, an $n^{\log n}$ time algorithm.

2 Preliminaries

We use $\mathsf{dist}(d, d')$ to denote the *distortion* between two distance functions d and d' on the same set of points. For a class $\mathcal{C}$ of distance functions, we use $\mathsf{dist}(d \hookrightarrow \mathcal{C})$ to denote the minimum distortion between d and $d' \in \mathcal{C}$. (This assumes, of course, that $\mathcal{C}$ contains distance functions on the same set of points as d.)

Let d be a distance function (on an underlying point set P), and let $f : \mathbb{R} \to \mathbb{R}$ be a monotonically non-decreasing function with $f(0) = 0$. We denote by $f(d)$ the distance function where $\forall p, q \in P$, $f(d)(p, q) = f(d(p, q))$. Notice that if d is a metric and f is concave, then $f(d)$ is a metric. The *power scale* $f(x) = x^c$, $c \in [0, 1]$ plays an important role in this paper. It is worth noting the following simple fact: $\mathsf{dist}(d^c, (d')^c) = (\mathsf{dist}(d, d'))^c$.

Let d be a distance function, and let Q be a subset of the points on which d is defined. We use d^Q to denote the restriction of d to the pairs of points in Q.

A set of metrics $\mathcal{C}$ is called *baseline* if it has the following properties, shared e.g., by the classes NEG, HYP, and $\mathcal{M}_k$ to be discussed later (see Section 3.2):

1. It is symmetric, i.e., for every $d \in \mathcal{C}$, any metric d' derived from d by permuting the underlying set of points is also in $\mathcal{C}$.

2. It is a closed cone, i.e., for every $d, d' \in \mathcal{C}$ on the same set of points, for every $a, a' \geq 0$, also $a \cdot d + a' \cdot d' \in \mathcal{C}$.

3. It is hereditary, i.e., for every $d \in \mathcal{C}$, for every subset of points Q on which d is defined, also $d^Q \in \mathcal{C}$.

4. For every $d \in \mathcal{C}$, consider a metric d', obtained from d by performing the following *cloning* operation: Pick a point p, add a "clone" q, and set $d'(q, x) = d(p, x)$ for all points x ($d(p, q) = 0$). Then, $d' \in \mathcal{C}$.

Observe that every baseline set of metrics includes all cut metrics, and therefore all metrics that embed isometrically in ℓ_1. Further notice that if $\mathcal{C}$ is a baseline set of metrics, then for every $\gamma \geq 1$, the set of metrics $\mathcal{C}_\gamma = \{d : \mathsf{dist}(d \hookrightarrow \mathcal{C}) \leq \gamma\}$ is also baseline .

3 Baseline sets of metrics

3.1 Upper Bounds. This section is devoted to the proof of the following theorem and its consequences.

THEOREM 3.1. *Let $m, n \in \mathbb{N}$, $m \leq n$, let $c \geq 1$, and let $\mathcal{C}$ be a baseline set of metrics. Let d be a metric on n points such that every m-point subspace Q has* $\mathsf{dist}(d^Q \hookrightarrow \mathcal{C}) \leq \gamma$. *Then,*

$$\mathsf{dist}(d \hookrightarrow \mathcal{C}) = O\left(\gamma \cdot \left(\frac{n}{m}\right)^2\right).$$

We require a definition. Let (X, d) be a metric space. A *tree-like extension* of (X, d) is a metric space obtained from (X, d) by repeatedly performing the following *attachment* operation: Pick a point $p \in X$ and a weight $w \geq 0$, and "attach" to p a new point $q \notin X$ by an edge of weight w, i.e., set $d'(q, x) = d'(p, x) + w$ for all points $x \in X$, and augment X by q.

LEMMA 3.1. *Let $\mathcal{C}$ be a baseline set of metrics, let $d \in \mathcal{C}$, and let d' be a tree-like extension of d. Then $d' \in \mathcal{C}$.*

Proof. Clearly, it suffices to prove this for a single attachment operation. Let d_p be the metric obtained from d by adding a clone q of a point p. Let δ be the cut metric defined by $\delta(x, y) = 1$ if exactly one of the points x, y is q, and $\delta(x, y) = 0$ otherwise. Both d_p and δ are in $\mathcal{C}$ (the former by definition, the latter because $\mathcal{C}$ must contain all cut metrics). Attaching q to p at distance w gives the metric $d' = d_p + w \cdot \delta$. As $\mathcal{C}$ is a closed cone, $d' \in \mathcal{C}$. $\square$

Next, we introduce a construction that will be used in the proof of Theorem 3.1. Let d be a metric on the finite set of points $P = \{p_1, p_2, \ldots, p_n\}$. It will be assumed w.l.o.g. that for any $p_i \in X$, the distances between p_i and the other points in X are all distinct. Let $\sigma \in S_n$ be a permutation on $\{1, 2, \ldots, n\}$. The metric d^σ is defined as follows. We start with restriction of d to $\{p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\}$. Then, for $i = m + 1, \ldots, n$, the new point $p_{\sigma(i)}$ is attached to $p_{\sigma(i^*)}$ at distance w_i^σ, where $i^* \in \{1, 2, \ldots, i - 1\}$ minimizes $d(p_{\sigma(i)}, p_{\sigma(i*)})$, and $w_i = d(p_{\sigma(i)}, p_{\sigma(i^*)})$. Finally, we define d^* on P as the average of all d^σ's: For every $p, q \in P$,

$$d^*(p, q) = \frac{1}{n!} \cdot \sum_{\sigma \in S_n} d^\sigma(p, q).$$

Proof. (of Theorem 3.1.) Throughout the proof m is fixed, whereas n and d vary. Let $T_{n,m}$, $n \geq m$ denote the supremum over all n-point metrics d of $\mathsf{dist}(d, d^*)$. Clearly, $T_{m,m} = 1$. Notice that for every $p, q \in P$, $d^\sigma(p, q) \geq d(p, q)$, and therefore d^* dominates d. To bound the stretch, observe that

$$
\begin{aligned}
d^*(p,q) &= \mathsf{E}_\sigma[d^\sigma(p,q)] \\
&= \frac{2}{n}\cdot \mathsf{E}_\sigma\left[d^\sigma(p,q)\,|\,\sigma(n)\in\{p,q\}\right] \\
&\quad + \left(1-\frac{2}{n}\right)\cdot \mathsf{E}_\sigma\left[d^\sigma(p,q)\,|\,\sigma(n)\notin\{p,q\}\right].
\end{aligned}
$$

Notice that $\mathsf{E}_\sigma\left[d^\sigma(p,q)\,|\,\sigma(n)\notin\{p,q\}\right] \le T_{n-1,m}\cdot d(p,q)$.

For the case $\sigma(n)=p$ let $p^*\in P$ be the point in P that is closest to p. By our assumptions, p^* is unique and hence it will be the point to which p will be attached. As $d(p,p^*)\le d(p,q)$ the triangle inequality implies that $d(p^*,q)\le 2d(p,q)$ and thus,

$$
\begin{aligned}
\mathsf{E}_\sigma[d^\sigma(p,q)|\,\sigma(n)=p] &= d(p,p^*)+\mathsf{E}_\sigma[d^*(p^*,q)|\,\sigma(n)=p] \\
&\le d(p,q)+T_{n-1,m}\cdot 2d(p,q).
\end{aligned}
$$

The case when $\sigma(n)=q$ is analogous. Therefore,

$$
\begin{aligned}
T_{n,m} &\le \left(1-\frac{2}{n}\right)\cdot T_{n-1,m}+\frac{2}{n}\cdot(2T_{n-1,m}+1) \\
&= \left(1+\frac{2}{n}\right)\cdot T_{n-1,m}+\frac{2}{n}.
\end{aligned}
$$

Solving the recurrence, we get that $T_{n,m}=O\left(\left(\frac{n}{m}\right)^2\right)$.

Next, recall that $\mathcal{C}_\gamma$ also is a baseline set of metrics. By the conditions of the theorem, for every m-point subset Q, $d^Q\in\mathcal{C}_\gamma$. Therefore, by Lemma 3.1, for every permutation σ, $d^\sigma\in\mathcal{C}_\gamma$. As $\mathcal{C}_\gamma$ is a closed cone, also $d^*\in\mathcal{C}_\gamma$. As $\mathsf{dist}(d,d^*)=O((\frac{n}{m})^2)$, the theorem now follows. □

THEOREM 3.2. *A metric $\tilde{d}\in\mathcal{C}$, which is an embedding of d satisfying the statement of Theorem 3.1, can be computed in probabilistic polynomial time.*

Proof. The construction of $\tilde{d}$ is based on the construction of d^*, and we use the same terminology as in the proof of Theorem 3.1. Let $K\subseteq S_n$ be a subset of permutations, with $|K|/|S_n|=\kappa\le 1$. Extending the definition of d^*, let $d^*_K=\mathsf{E}_\sigma[d^\sigma\,|\sigma\in K]$. Observe that

$$
d\ \le d^*_K\ \le\ \kappa^{-1}d^*\ . \tag{3.1}
$$

The first inequality holds since each d^σ dominates d, the second since $d^* = d^*_K\cdot\kappa+d^*_{\bar{K}}\cdot(1-\kappa)$.

Let $G=\{\sigma\,|\,d^\sigma\le 10n^2T_{n,m}\cdot d\}\subseteq S_n$. By Theorem 3.1, the expected stretch of d^* with respect to any pair of points in the space is $\le T_{n,m}$. Therefore, using Markov's Inequality, $|G|\ge 0.9|S_n|$, and $d^*_G\le 1.1d^*$. Next, let $\tilde{G}$ be a random sample from G of size N. By Hoeffding's large deviation bound, the expected stretch of $d^*_{\tilde{G}}$ with respect to any pair of points x,y in the space is

$$
\Pr\left[d^*_{\tilde{G}}(x,y)\ge(1+\delta)\cdot d^*_G(x,y)\right]\ \le\ \left(\frac{e}{1+\delta}\right)^{\frac{\delta N}{10n^2T_{n,m}}}
$$

Thus, choosing $N=10n^2T_{n,m}\log_2 n$, we conclude that $d^*_{\tilde{G}}\le 3.5d^*_G\le 4d^*$ with probability close to 1. Finally, to create a random sample $\tilde{G}$, we sample permutations σ from S_n, construct d^σ, and discard it if any of the edges is stretched by more than $10n^2T_{n,m}$. Since 0.9 fraction of the permutations in S_n are in G, this gives a polynomial time probabilistic algorithm. □

Theorem 3.2 has the following interesting structural implication.

COROLLARY 3.1. *The assumption of Theorem 3.1 that* all *size-m subspaces are γ-close to $\mathcal{C}$ can be replaced by a weaker assumption that only a κ-fraction of the size-m subspaces have this property, at the cost of an additional multiplicative factor of κ^{-1} in the upper bound.*

Indeed, the permutations in S_n whose m-prefix corresponds to a good size-m subset constitute a κ-fraction of all permutations, and (3.1) applies.

We now show that Theorems 3.1 and 3.2 imply a sub-exponential time algorithm for approximating sparsest cut to within any super-constant factor.

THEOREM 3.3. *Sparsest cut can be approximated to within a factor of $O(c^2)$ in time $\exp\left(\frac{n\log c}{c}\right)$, where n is the number of nodes in the input graph.*

Proof. Let (G,w,T,h) be an instance of sparsest cut. Here $G=(V,E)$ is an undirected graph, $w:E\to\mathbb{N}$ is a weight function on the edges of G, $T=\{(s_1,t_1),(s_2,t_2),\dots,(s_k,t_k)\}$ is a set of pairs of nodes of G (called terminals), and $h:T\to\mathbb{N}$ is the demand function. Let D be the set of semi-metrics d on V, such that for every $U\subset V$ with $|U|\le\frac{1}{c}|V|$, the restriction of d to U embeds isometrically in ℓ_1. By Theorem 3.1,

$$
z^*=\min\left\{\frac{\sum_{e\in E}w(e)d(e)}{\sum_{(s,t)\in T}h(s,t)d(s,t)}\ :\ d\in D\right\} \tag{3.2}
$$

is achieved at a semi-metric d that embeds into ℓ_1 with distortion $O(c^2)$. By the results of [18, 5], given the embedding of d into ℓ_1, one can find a cut $(S,V\setminus S)$ in G such that

$$
\frac{\sum_{e\in E:\ |e\cap S|=1}w(e)}{\sum_{(s,t)\in T:\ |\{s,t\}\cap S|=1}h(s,t)}=O(c^2z^*). \tag{3.3}
$$

It is known that if we replace $d\in D$ with $d\in\ell_1$ in equation (3.2) we get the value of the sparsest cut.

Hence, z^* is a lower bound on the value of the sparsest cut and thus the cut for which equation (3.3) holds approximates the sparsest cut as claimed. Now z^* can be computed by a linear program with $\binom{|V|}{2}$ variables and $\binom{|V|}{\frac{1}{c}|V|} \cdot 2^{O(\frac{1}{c}|V|)}$ constraints. To find the cut we need to compute the embedding of d into ℓ_1. By Theorem 3.2, it is sufficient to compute the (isometric) embedding of poly($|V|$) tree-like extensions of subsets U of size $|V|/c$ (at the cost of an extra $O(1)$ factor in the approximation guarantee). The embedding of a tree-like extension of a subset U is trivial to compute, given the embedding of U. The latter can be computed through a linear program with $2^{|U|-2}$ variables (one for each possible cut in U) and $\binom{|U|}{2}$ constraints. □

3.2 Lower Bounds. The main result in this section is a nearly tight counterpart to some of the upper bounds from Section 3.1. The next lemma will play a key role in the proof.

Lemma 3.2. *Let d be the shortest path metric of an n-node graph $G = (V, E)$ with girth at least p, diameter D and such that the subgraph induced by any subset $S \subset V$ of size at most $n^{1-\varepsilon/2}$ has at most $|S|(1 + 1/p)$ edges. Then, for every $S \subseteq V$ with $|S| \leq n^{1-\varepsilon}$, the corresponding d^S can be embedded in ℓ_1 (in fact, into a distribution over dominating tree-metrics) with distortion $O(1/\varepsilon) \cdot (D + 1)/(p + 1)$.*

Theorem 3.4. *For every $\varepsilon > 0$ and for every integer $n \geq 2$, the following statements hold.*

1. *There is an n-point metric d such that for every $n^{1-\varepsilon}$-point subspace Q, $\mathsf{dist}(d^Q \hookrightarrow \ell_1) = O(1/\varepsilon^2)$ yet $\mathsf{dist}(d \hookrightarrow \ell_1) = \Omega(\log n)$.* [2]

2. *There is an n-point metric d such that for every $n^{1-\varepsilon}$-point subspace Q, $\mathsf{dist}(d^Q \hookrightarrow \ell_2) = O(1/\varepsilon)$, yet $\mathsf{dist}(d \hookrightarrow \ell_2) = \Omega(\sqrt{\log n})$.*

Proof. We start with a random 3-regular graph and delete $o(n)$ vertices so that the girth of the resulting graph G is $p = \Theta(\varepsilon' \log n)$ and G is an expander (for convenience, let n be the number of vertices after the deletion). We will apply Lemma 3.2 to G. To do this, we note that $D = O(\log n)$ for an expander and by Lemma 3.3 below we get $p = \Theta(\varepsilon \log n)$, which gives an upper bound on the distortion of $O(1/\varepsilon^2)$.

It is well-known that the distortion of embedding the shortest path metric of an n-node bounded degree expander into ℓ_1 is $\Omega(\log n)$, so the metric d induced by G satisfies the first statement. For the second statement we use the metric $\sqrt{d}$, keeping in mind that the square root of an ℓ_1 metric is an ℓ_2 metric. □

Lemma 3.3. **[2]** *In a random 3-regular graph, with high probability, the subgraph induced by any subset S of size at most $n^{1-\varepsilon}$ has at most $(1+\frac{c}{\varepsilon \log n})(|S|-1)$ edges where c is an absolute constant.*

Proof. (of Lemma 3.2) Fix a subset S with at most $n^{1-\varepsilon}$ vertices, and consider the corresponding induced metric d^S. Recall the definition of a *spanner* of (S, d^S): it is a graph $Y = (S, L)$ on the vertex set S, such that the weight of an edge $(i, j) \in L$ is $d^S(i, j)$. By [1], there exists a *spanner* Y of (S, d^S) such that $|L(Y)| \leq n^{1-\varepsilon} \cdot n^{\varepsilon/2} = n^{1-\varepsilon/2}$, and the shortest-path metric of Y approximates d^S up to a factor $O(1/\varepsilon)$.

Let $H = (U, F)$ be the subgraph obtained by including all the edges on the shortest paths between pairs of vertices $i, j \in S$, such that $(i, j) \in L(Y)$. We will show that the shortest-path metric d^H induced by H can be embedded into a distribution of dominating tree metrics with distortion $O((D+1)/(p+1))$. Hence, d^S embeds into such a distribution with distortion $O(1/\varepsilon) \cdot (D + 1)/(p + 1)$.

Let $H = (U, F)$ be a subgraph of G with $O(n^{1-\varepsilon'})$ vertices. The lemma follows from the following two claims.

Claim 3.1. *There is a probability distribution on spanning trees of H such that each edge of H occurs with probability at least $p/(p + 1)$.*

It is well-known that metrics induced by trees are isometrically embeddable in ℓ_1. The second claim is that *truncated* tree metrics are embeddable with a constant distortion.

Claim 3.2. *Given a tree metric t and a number $M \geq 0$, let $t'_{ij} = \min\{t_{ij}, M\}$. Then t' can be embedded into ℓ_1* [3] *with constant distortion.*

Applying Claim 3.1 to H, we conclude that there is a probability distribution on spanning trees $\{T_i\}$ of H such that each edge of H occurs with probability at least $\alpha = p/(p + 1)$. Let D be the diameter of H. For each T_i in the distribution, consider the corresponding metric $t_i = \min\{D, d_{T_i}\}$. Define a metric $t = \sum w_i t_i$, where w_i is the weight of T_i in the distribution. Clearly, t dominates d_H. To upper-bound $\mathsf{dist}(t, d^H)$, consider an edge of H. It's $t-$length is at most

$$1 \cdot \alpha + D \cdot (1 - \alpha) \;\leq\; \frac{1 + D}{1 + p}.$$

[2] In fact, the lower bound holds even for embedding into NEG, the class of negative type metrics (for definition — see below).

[3] In fact, ℓ_1 can be replaced by a distribution of dominating tree-metrics, a more restricted class of metrics.

Finally, by Claim 3.2, every t_i, and hence t can be embedded in a distribution of H-dominating tree metrics with constant distortion.

It remains to prove the claims. For the first, we define two polytopes in $\mathbb{R}^{|F|}$. The first polytope, P, will be the *spanning tree polytope* of H, i.e., the set of all vectors that are convex combinations of incidence vectors of spanning trees of H. The second polytope, B, will be the following axis-parallel box with one corner being the vector of all 1's.

$$B_\alpha = \{v \in \mathbb{R}^{|F|} | \forall e : \alpha \leq v_e \leq 1\}. \tag{3.4}$$

The claim is that for $\alpha \leq p/(p+1)$, the intersection of P and B_α is nonempty.

By Farkas' Lemma, it suffices to show that for any $w \in \mathbb{R}^{|F|}$, there exists a vector $v \in B_\alpha$ such that

$$\max_{x \in P} w \cdot x \geq w \cdot v. \tag{3.5}$$

Note that since the extreme points of P are spanning trees of H, the LHS is always maximized by the incidence vector of some spanning tree. We consider two extreme cases:

1. $w \leq 0$: In this case we set $v_{ij} = 1$ for every edge (i, j). The inequality follows.

2. $w \geq 0$: In this case we set $v_{ij} = \alpha$ for every edge. Suppose the LHS of (3.5) is maximized by the spanning tree T. We will prove that the total weight of all the edges in H is only slightly larger than the weight of T. For this consider the following bipartite graph. The left side of the bipartition has a point corresponding to each edge of T. The right side has a point for each edge of H that is not in T. There is an edge (e, f) if $e \in T$ belongs to the fundamental cycle of $f \notin T$. Note that the optimality of T implies that $w_e \geq w_f$. Let the girth of G be g. Recall $g \geq p$. Thus the degree of each vertex in $\overline{T}$ is at least p. We claim that this bipartite graph has a p-matching: a subgraph with degree 1 for points on the left and degree p for points on the right. Suppose not. Then there is some minimal subset X on the right side whose neighborhood $N(X)$ has size $|N(X)| < |X|p$. Now consider the subtree of T induced by $N(X)$ (if the edges corresponding to $N(X)$ do not form a connected component, then X is not minimal). This subtree has $|N(X)| + 1$ vertices and the subgraph of H induced by these vertices has at least $|N(X)|+|X| > |N(X)|(1+1/p)$ edges. But this contradicts Lemma 3.3.

   The existence of the p-matching implies that the edges of T can be partitioned into p subsets such that the weight of each subset is more than the weight of all the edges not in T. Thus,

   $$\sum_{ij} w_{ij} \leq \left(1 + \frac{1}{p}\right) \sum_{ij \in T} w_{ij}.$$

   This implies that inequality (3.5) holds for any $\alpha \leq p/(p+1)$.

For the general case, take an arbitrary vector w and set $c_{ij} = 1$ for $w_{ij} \leq 0$ and $c_{ij} = \alpha$ if $w_{ij} > 0$. Consider the connected components induced by the nonnegative edges. For each component the inequality is implied separately by the second case above. Now shrink all the components to single vertices. The inequality on the induced graph follows from the first case. Summing up, (3.5) is proved.

We conclude with the proof of the second claim. Let T be the tree corresponding to t. Build a (weighted) graph T' by introducing a new vertex u, and connecting it to every vertex of T by an edge of length $M/2$. Observe that the shortest-path metric of T' restricted to $V(T)$ is precisely t, and that T' is 2-outerplanar. By [9], this implies that t', and hence t, can be embedded into a distribution of dominating tree metrics (and thus into ℓ_1) with constant distortion. $\square$

We now turn our attention to the case where subspaces embed isometrically into an interesting class of metrics. Our lower bounds in this case are much weaker. We need the following definitions.

A distance function d is *k-gonal* iff for every two sequences of points $p_1, p_2, \ldots, p_{\lfloor k/2 \rfloor}$ and $q_1, q_2, \ldots, q_{\lceil k/2 \rceil}$ (where points are allowed to appear multiple times in each sequence) the following inequality holds:

$$\sum_{i=1}^{\lfloor k/2 \rfloor} \sum_{j=1}^{\lceil k/2 \rceil} d(p_i, q_j) \geq$$

$$\sum_{i=1}^{\lfloor k/2 \rfloor} \sum_{i'=1}^{\lfloor k/2 \rfloor} d(p_i, p_{i'}) + \sum_{j=1}^{\lceil k/2 \rceil} \sum_{j'=1}^{\lceil k/2 \rceil} d(q_j, q_{j'}).$$

We use $\mathcal{M}_k$ to denote the class of all k-gonal distance functions. Clearly, $\mathcal{M}_3$ is simply all metrics. Also, for every $k \in \mathbb{N}$, $k \geq 2$, $\mathcal{M}_{k+2} \subset \mathcal{M}_k$ and $\mathcal{M}^n_{2k-1} \subset \mathcal{M}^n_{2k}$. On the other hand, for every $k \in \mathbb{N}$, $k \geq 1$, distance functions in $\mathcal{M}^n_{2k}$ are not necessarily metrics. The class of all *negative type* distance functions is

$$\text{NEG} = \bigcap_{k=2}^{\infty} \mathcal{M}_{2k}.$$

Schoenberg showed that $d \in$ NEG iff $\sqrt{d}$ embeds isometrically into ℓ_2. The class of all hypermetrics is

$$\text{HYP} = \bigcap_{k=2}^{\infty} \mathcal{M}_{2k-1}.$$

Thus, all hypermetrics are negative type metrics. It is known that all ℓ_1 metrics are hypermetrics. All classes of metrics discussed above (except for ℓ_2 metrics) are baseline .

A theorem in [11], combined with an argument similar to the proof of Theorem 3.4 gives the following theorem. The proof is omitted from this extended abstract.

THEOREM 3.5. *For every integer $n \geq 2$ and for every $k \in \mathbb{N}$, $k \leq n$, the following statements are true:*

1. *There exists an n-point k-gonal metric d such that* $\mathsf{dist}(d \hookrightarrow \text{NEG}) = \Omega\left((\log n)^{\log_2(1+1/(\lceil k/2\rceil-1))}\right)$.
2. *There exists an n-point metric d such that every k-point subspace is hypermetric, yet* $\mathsf{dist}(d \hookrightarrow \text{NEG}) = \Omega\left((\log n)^{\log_2(1+1/(k-1))}\right)$.
3. *There exists an n-point metric d such that every k-point subspace embeds isometrically in ℓ_2, yet* $\mathsf{dist}(d \hookrightarrow \text{NEG}) = \Omega\left((\log n)^{\frac{1}{2}\log_2(1+1/(k-1))}\right)$.

COROLLARY 3.2. *For $k = o(\log\log n)$, there exist an n-point metric d such that every k-point subspace is in ℓ_2, yet $\mathsf{dist}(d \hookrightarrow \text{NEG}) = \omega(1)$. (Recall that $\ell_2 \subset \ell_1 \subset$ NEG).*

4 Ultrametrics

The set of ultrametrics is the set of metrics ULT $= \{d : d(p,q) \leq \max\{d(p,r), d(q,r)\}, \forall p,q,r\}$. All ultrametrics embed isometrically into ℓ_2. Notice that ULT is not baseline , so the results from the previous section do not apply to this set. Consider an ultrametric d. Given two points x, y, an xy-path P is a sequence of points $(x = p_0, p_1, p_2, \ldots, p_m = y)$ of arbitrary length. We say that $pq \in P$ iff there exists $j \in \{1, 2, \ldots, m\}$ such that $p = p_{j-1}$ and $q = p_j$. For every two points x, y put

$$u(x,y) = \min_{xy\text{-paths } P} \{\max\{d(p,q) : pq \in P\}\}.$$

THEOREM 4.1. (FARACH-COLTON [13]) *The distance function u is an ultrametric which is dominated by d (i.e., $u(x,y) \leq d(x,y)$, for every $x, y \in X$). Moreover, every ultrametric u' that is dominated by d is also dominated by u.*

As an immediate corollary we get the following criterion.

COROLLARY 4.1. *Let $c \leq 1$ be the maximum value such that for every $x, y \in X$, every xy-path P contains $pq \in P$ such that $d(p,q) \geq c \cdot d(x,y)$. Then,* $\mathsf{dist}(d \hookrightarrow \text{ULT}) = c^{-1}$.

Using this criterion we establish the following theorem.

THEOREM 4.2. *Let $c \geq 1$, and let $\epsilon > 0$. Let d be an n-point metric such that for every n^ϵ-point subspace Q,* $\mathsf{dist}(d^Q \hookrightarrow \text{ULT}) \leq c$. *Then,* $\mathsf{dist}(d \hookrightarrow \text{ULT}) = c^{\lceil 1/\epsilon\rceil}$. *This bound is essentially tight.*

Proof. For the upper bound, it suffices to show that for $n = m^k - m^{k-1} + 1$, $k \in \mathbb{N}$, it holds that $\mathsf{dist}(d \hookrightarrow \text{ULT}) \leq c^k$. The proof is by induction on k. For $k = 1$ the theorem is trivially true. For $k > 1$, by Corollary 4.1, it suffices to show that for every $x, y \in X$, any simple xy-path P contains $pq \in P$ with $d(p,q) \geq d(x,y)/c^k$. Let $P = (x = v_1, v_2, \ldots, v_r = y)$, $r \leq n$, be such a path. Consider the xy-path $P' = (v_1, v_m, v_{2m-1}, v_{3m-2}, \ldots, v_r)$. As P' has at most $1 + \frac{n-1}{m} = m^{k-1} - m^{k-2} + 1$ points, the induction hypothesis implies that there exists $v_{j-1}v_j \in P'$ with $d(v_{j-1}, v_j) \geq d(x,y)/c^{k-1}$. Consider the segment of P $(v_{j-1}, \ldots, v_j)$ containing at most m points. By the base case of the induction, there exists $pq \in (v_{j-1}, \ldots, v_j)$ such that $d(p,q) \geq d(v_{j-1}, v_j)/c \geq d(x,y)/c^k$.

For the lower bound, consider the metrics d_n^c, where d_n is the shortest path metric of the n-node cycle and $c \in [0,1]$. The reader can verify easily using Corollary 4.1 that $\mathsf{dist}(d_n^c \hookrightarrow \text{ULT}) = \Omega(n^c)$, whereas for every subspace on n^ϵ points, the restriction d' of d_n^c to this subspace has $\mathsf{dist}(d' \hookrightarrow \text{ULT}) = O(n^{\epsilon c})$. □

REMARK 4.1. *The metrics d_n^c are, in fact, $\Omega(n^c)$ far from the more general set of tree metrics (by the argument from [22, Corollary 5.3]). Hence, the lower bounds hold for tree metrics as well.*

5 Separating a baseline Metric Class from ℓ_∞

Let $\mathcal{C}$ be a baseline metric class that excludes some metric. How well can the metrics from $\mathcal{C}$ approximate general metrics? The following purely existential result of Matousek [20] implies a separation between the class of all metrics and $\mathcal{C}$, i.e., for every $\gamma > 1$ there exists a metric D such that $\mathsf{dist}(D \hookrightarrow \mathcal{C}) \geq \gamma$.

THEOREM 5.1. *For every finite metric μ and any constants $\varepsilon > 0, \gamma > 1$, there exists a (larger) finite metric D such that, for any metric M on the same set of points as D, if $\mathsf{dist}(D, M) \leq \gamma$, then M contains a submetric μ' with* $\mathsf{dist}(\mu, \mu') \leq (1 + \varepsilon)$.

We conjecture that a much stronger separation holds.

Conjecture 5.1. *For any $n \in \mathbb{N}$, there exists an n-point metric d_n such that* $\mathsf{dist}(d_n \hookrightarrow \mathcal{C}) \geq \Omega(\log^\alpha n)$ *for some constant $\alpha > 0$.*

In what follows, we produce a supporting evidence for this conjecture by proving its analogue for normed spaces. Unlike in the rest of the paper, we assume here that $\mathcal{C}$ contains not only finite metrics, but also metrics whose underlying space is the entire $\mathbb{R}^n$ or $\mathbb{Z}^n$, and, in particular $\{\ell_1^n\}_{n=1}^\infty \in \mathcal{C}$.

Theorem 5.2. *Let $\mathcal{C}$ be a baseline metric class, and assume that there exists a metric μ_k on k points such that* $\mathsf{dist}(\mu_k \hookrightarrow \mathcal{C}) = \beta > 1$. *Then, for any $\mathcal{C}$-metric d on $\mathbb{R}^n$, it holds*

$$\mathsf{dist}(d, \ell_\infty^n) \;=\; \Omega(n^\alpha), \quad \textit{where} \quad \alpha \approx \frac{1}{2}\frac{\beta-1}{\beta+1}\frac{1}{\ln k} \;.$$

Observe that the gap between the two may not exceed $\sqrt{n}$, the gap between ℓ_∞^n and $\ell_2^n \subset \mathcal{C}$.

The proof of the theorem uses the following lemma.

Lemma 5.1. *For any $d \in \mathcal{C}$ on $\mathbb{R}^n$, there exists a* norm $\| * \| \in \mathcal{C}$ *on $\mathbb{R}^n$, such that*

$$\mathsf{dist}(\ell_\infty^n, \| * \|) \;\leq\; \mathsf{dist}(\ell_\infty^n, d) \;. \tag{5.6}$$

The proof of this lemma appears in the Appendix.

Next, we need the following quantitative version of a theorem by James [16], communicated to us, together with an outline of its proof, by W.B. Johnson and G. Schechtman:

Theorem 5.3. *Assume that $\gamma = (1+\delta)^{2^r}$, and $n \geq k^{2^r}$, where $r, k \in \mathbb{N}$, and $0 \leq \delta < 1$. Then, if an n-dimensional norm $\|*\|$ is γ-close (in the sense of metric distortion) to ℓ_∞^n, then there exists a subspace L of $\mathbb{R}^n$ of dimension* $\dim(L) = k$, *such that the restriction of $\| * \|$ to L is $\frac{1+\delta}{1-\delta}$-close to an ℓ_∞ norm on L.*

The theorem as stated follows from a lemma from [21], pp.74-75, which establishes L of dimension k, such that the restriction of $\| * \|$ to L satisfies $\| \sum_i \alpha_i v_i \| \leq (1+\delta) \cdot \max_i |\alpha_i| \cdot \|v_i\|$, together with a simple claim [17] that,

$$\| \sum_i \alpha_i v_i \| \leq (1+\delta) \cdot \max_i |\alpha_i| \cdot \|v_i\|$$
$$\Rightarrow \| \sum_i \alpha_i v_i \| \geq (1-\delta) \cdot \max_i |\alpha_i| \cdot \|v_i\| \;.$$

Finally, we prove Theorem 5.2. Assume for simplicity that n is of the form $n = k^{2^r}$. The metric $\mu_k \notin \mathcal{C}$, being a metric on k points, isometrically embeds into ℓ_∞^k. We conclude by Theorem 5.3 that for any $\mathcal{C}$-norm $\| * \|$ on $\mathbb{R}^n$ it holds

$$\mathsf{dist}(\| * \|, \ell_\infty^n) \;\geq\; \left(1 + \frac{\beta-1}{\beta+1}\right)^{2^r} \;.$$

The same estimate holds, by Lemma 5.1, for any metric $d \in \mathcal{C}$ on $\mathbb{R}^n$. Thus, for such n, the theorem holds with constant $\alpha = \log_k \left(1 + \frac{\beta-1}{\beta+1}\right)$. If n is not of the form k^{2^r}, take the largest such power $\leq n$, at the cost of paying an extra factor $1/2$ in the above α. This concludes the proof of Theorem 5.2.

6 Concluding remarks

We already mentioned the main open problem in the introduction, namely, to understand metrics whose small sets embed isometrically into ℓ_1. Theorems 3.1 and 3.5 provide a starting point for a further research.

Approximating general (or special) metrics by metrics from some nontrivial baseline class $\mathcal{C}$ may have interesting algorithmic applications. We conjecture that for any $n \in \mathbb{N}$, there exists an n-point metric d_n such that $\mathsf{dist}(d_n \hookrightarrow \mathcal{C}) \geq \Omega(\log_n^\alpha)$ for some constant $\alpha > 0$ depending on $\mathcal{C}$. A corresponding upper bound with $\alpha < 1$ would be most interesting. Regarding special metrics, it would be interesting to show, e.g., that any planar metric can be approximated by a metric in $\mathcal{M}_6$ with *constant* distortion. This is closely related to the famous question about ℓ_1-embeddability of planar metrics (see, e.g., [15]). Gupta [14] showed that planar metrics embed with constant distortion into NEG, and hence into $\mathcal{M}_{2k}$.

It might be of interest to study the implications of a local property on a different global property. For example, the extremal metrics constructed in the proof of Theorem 4.2, while far from being ultrametrics, are essentially very simple metrics. In particular, they are outerplanar and, up to a factor of $\pi/2$, Euclidean. It makes sense to ask, for example, if metrics that are locally almost ultrametric are globally almost ℓ_1 metrics.

The findings of this paper and other results indicate that the shortest path metrics of random k-regular graphs have a surprisingly simple local structure. Further research leading to a better understanding of this local structure, may prove useful for constructing lower bounds.

Acknowledgments

We are deeply grateful to G. Schechtman and B.W. Johnson for their explanation of Theorem 5.3. We also thank J. Matousek for pointing out important references.

References

[1] I. Althoefer, G. Das,D. Dobkin, D. Joseph, and J. Soares. On generating sparse spanners for weighted graphs. *Discrete and Computational Geometry* 9:81–100, 1993.

[2] S. Arora, B. Bollobas, and L. Lovász. Proving integrality gaps without knowing the linear program. In *Proc. of the 43rd Ann. IEEE Symp. on Foundations of Computer Science*, 2002.

[3] S. Arora, J. Lee, and A. Naor. Euclidean distortion and sparsest cut. *Manuscript, Nov 2004.*

[4] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *Proc. of the 36th Ann. Symp. on Theory of Computing*, pages 222–231, 2004.

[5] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. SIAM Journal on Computing, 27(1):291–301, 1998.

[6] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric Ramsey type phenomena. *Annals of Mathematic*, to appear. Preliminary version appeared in *Proc. STOC '03.*

[7] S. Chawla, A. Gupta, and H. Räcke. An improved approximation to sparsest cut. In *Proc. of the 16th Ann. ACM-SIAM Symp. on Discrete Algorithms*, January 2005.

[8] S. Chawla, R. Kumar, R. Krauthgamer, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest cut. To appear in *Proc. of the 20th Ann. IEEE Conference on Computational Complexity.*

[9] C. Chekuri, A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Embedding k-outerplanar graphs into l_1. In *Proc. of the 14th Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 527–536, 2003.

[10] M.M. Deza and M. Laurent. *Geometry of Cuts and Metrics.* Springer Verlag, 1997.

[11] M.M. Deza and H. Maehara. Metric transform and Euclidean embeddings. *Transactions of the American Mathematical Society*, 317:661–671, 1990.

[12] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proc. of the 35th Ann. ACM Symp. on Theory of Computing*, pages 448–455, June 2003.

[13] M. Farach-Colton. (Nearly) optimal embeddings into additive and ultrametric trees. *DIMACS Workshop on Discrete Metric Embeddings*, Princeton, August 2003.

[14] A. Gupta. *Embeddings of Finite Metrics.* Ph.D. thesis, University of California, Berkeley. August 2000.

[15] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and ℓ_1-embeddings of graphs. In *Proc. of the 40th Ann. IEEE Symp. on Foundations of Computer Science*, pages 399–408, 1999.

[16] R.C. James. Uniformly non-square Banach spaces. *Annals of Mathematics*, 80:542–550, 1964.

[17] W.B. Johnson and G. Schechtman. Private communication, 2003.

[18] N. Linial, E. London, and Yu. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[19] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. of the 34th Ann. ACM Symp. on Theory of Computing*, pages 767–775, 2002.

[20] J. Matousek. Ramsey-like properties for bi-Lipschitz mappings of finite metric spaces. *Comment. Math. Univ. Carolinae*, 33:451–463, 1992.

[21] V. Milman and G. Schechtman. *Asymptotic Theory of Finite Dimensional Spaces.* LNM 1200, Springer-Verlag 1986.

[22] Y. Rabinovich and R. Raz. Lower bounds on distortion of embedding finite metric spaces in graphs. *Discrete and Computational Geometry*, 19:79–94, 1998.

A Proof of Lemma 5.1:

W.l.o.g., in what follows we restrict our attention to d's dominating ℓ_∞^n, and respectively, speak of (supremum) stretch incurred by d instead of speaking of distortion. It will be convenient to bring the discussion back to the realm of discrete metric spaces. Instead of proving (5.6) for $\mathbb{R}^n$, we shall prove it for $\mathbb{Z}^n$. Clearly, this is a fully equivalent statement. Observe that a norm on $\mathbb{Z}^n$ is just a translation-invariant scalable metric.

First, we construct a translation-invariant metric $d_* \in \mathcal{C}$ on $\mathbb{Z}^n$, such that the stretch incurred by d_* is no more than that of d. The construction is as follows. Given d and a point $p \in \mathbb{Z}^n$, define a metric d^{+p} on $\mathbb{Z}^n$ by

$$d^{+p}(x,y) \;=\; d(x+p, y+p)\,.$$

Observe that by the symmetry of $\mathcal{C}$, $d^{+p}(x,y) \in \mathcal{C}$. Moreover, it dominates $\mathbb{Z}^n$ equipped with the ℓ_∞^n metric, and has the same stretch as d.

For an integer i let $[-i,i] = \{-i,\ldots,i\}$. Let $[-i,i]^n \subseteq \mathbb{Z}^n$ denote the corresponding discrete cube. For a point $x \in \mathbb{Z}^n$ let $[-i,i]^n - x = \{y-x \mid y \in [-i,i]^n\}$ denote the shifted cube. Consider a sequence of metrics $d = d_0, d_1, d_2, \ldots$ defined by:

$$d_i \;=\; \frac{1}{|[-i,i]^n|} \sum_{p\in[-i,i]^n} d^{+p}\,.$$

Clearly d_i belongs to $\mathcal{C}$, it dominates the ℓ_∞^n metric, and the stretch incurred by d_i is no more than that incurred by d. Observe also that For every $x, y \in \mathbb{Z}^n$ we have

$$\begin{aligned}
& \lim_{i\to\infty} |d_i(x,y) - d_i(0, y-x)| \\
= \;& \lim_{i\to\infty} \mid \frac{1}{(2i+1)^n} \sum_{p\in[-i,i]^n} d(x+p, y+p) \\
& - \frac{1}{(2i+1)^n} \sum_{p\in[-i,i]^n} d(p, y-x+p) \mid \\
\leq \;& \lim_{i\to\infty} \frac{1}{(2i+1)^n} \sum_{p\in\; [-i,i]^n \triangle\, ([-i,i]^n - x)} d(x+p, y+p) \\
\leq \;& \lim_{i\to\infty} \frac{1}{(2i+1)^n} \cdot 2n \cdot \|x\|_\infty \cdot (2i+1)^{n-1} \cdot \\
& \cdot \mathsf{dist}(\ell_\infty^n, d) \|x-y\|_\infty \;=\; 0.
\end{aligned}$$

Next, we employ the following standard procedure. Order all vectors of $\mathbb{Z}^n$ in some order $v_1, v_2, v_3, \ldots$. Consider an infinite subsequence of $\{d_i\}$ such that the value of $d_i(0, v_1)$ converges on it; call this limit $\nu(v_1)$. Do the same with the latter subsequence to obtain $\nu(v_2)$ and a sub-subsequence, and continue in the same manner ad infinitum. Finally, for each $x, y \in \mathbb{Z}^n$, define

$$d_*(x,y) \;=\; \nu(y-x)\,.$$

The above observation implies that d_* is indeed a translation-invariant metric. Clearly, $d_* \in \mathcal{C}$, it is ℓ_∞^n-dominating, and the stretch incurred by it is bounded by the stretch incurred by d.

Second, we use d_* to construct $d_{**} \in \mathcal{C}$ with the same properties, which is not only translation-invariant, but also scalable. The construction is similar to the previous one, but is a bit simpler. Consider a sequence of translation-invariant metrics $d^{(0)}, d^{(1)}, d^{(2)}, \ldots$ defined as follows:

$$d^{(r)}(x,y) \;=\; 2^{-r}\, d_*(2^r \cdot x,\, 2^r \cdot y)\,.$$

Observe that $d^{(r)}$'s are (pointwise) monotone non-increasing with r, since for any $a \in \mathbb{N}^+$, and for $a = 2$ in particular, $d^*(ax, ay) \;\leq\; a d_*(x,y)$ due to translation-invariance of d^*.

Taking the limit of $d^{(r)}$'s we obtain the desired d_{**}. It is easy to check that d_{**} has all the required properties. E.g., the scalability holds, since, by the previous observation, the limit $\lim_{r\to\infty} a^{-1} d^{(r)}(ax, ay)$ exists for every natural a. Therefore d_{**} is scalable with respect to all $a \in \mathbb{N}^+$, and hence with respect to all $a \in \mathbb{Q}^+$, as required. $\square$