ON LOWER BOUNDS FOR LOCAL VERSIONS OF METRIC EMBEDDINGS

VISHESH JAIN

Department of Mathematics, Statistics, and Computer Science, University of Illinois Chicago, Chicago, IL, 60607 USA

DUAN TU

Department of Mathematics, Statistics, and Computer Science, University of Illinois Chicago, Chicago, IL, 60607 USA

Abstract. We consider the problem of embedding points from an arbitrary finite metric space into a target metric space while preserving, up to a small distortion, only a subset of the pairwise distances specified by a bounded degree graph G. We provide a general reduction showing that, in many cases, this is no easier than embedding the points while approximately preserving *all* pairwise distances.

As an illustration of our general reduction, we show that there exists a Euclidean metric space X on n points along with a graph $G = (X, E)$ of maximum degree 3 such that any embedding of X into ℓ_2^m which only preserves distances specified by E up to a relative error of $(1 + \varepsilon)$ must satisfy $m = \Omega(\varepsilon^{-2} \log n)$, which matches the upper bound on the dimension coming from the Johnson-Lindenstrauss lemma (for approximately preserving all pairwise distances). Previously, such a lower bound was known only for the class of noncontracting embeddings [Schechtman-Shraibman, Discrete & Computational Geometry, 2009].

E-mail addresses: `visheshj@uic.edu, dtu4@uic.edu`.

V.J. was supported by NSF CAREER award DMS-2237646.

1. Introduction

Dimension reduction is a cornerstone of modern data analysis and theoretical computer science, providing methods to represent high-dimensional data in lower-dimensional spaces while preserving essential geometric properties. The Johnson-Lindenstrauss (JL) lemma [10] is a seminal result, establishing that any set of n points in Euclidean space can be embedded into $\ell_2^{O(\log n/\varepsilon^2)}$ (the Euclidean space in $O(\log n/\varepsilon^2)$ dimensions) while preserving all pairwise Euclidean distances up to a $(1+\varepsilon)$ distortion. It was shown by Larsen and Nelson that this bound is tight up to constant factors for $\varepsilon \in (n^{-0.499}, 1)$ (see [11, Theorem 2] for a more general statement). Another seminal result in the theory of metric embeddings is Bourgain's embedding theorem [7], which asserts that every metric space on n points can be embedded into Euclidean space with distortion $O(\log n)$; this is known to be tight up to constant factors (see, e.g., [12]). Here, we say that an embedding f from a metric space (X, d_X) to a metric space (Y, d_Y) has distortion at most $D \geq 1$ if there exists $r > 0$ such that

$$r \cdot d_X(x, y) \leq d_Y(f(x), f(y)) \leq D \cdot r \cdot d_X(x, y) \qquad \forall x, y \in X.$$

Local dimension reduction. In numerous applications, particularly those dealing with massive or inherently complex datasets like social networks or biological data, preserving the full distance information is often unnecessary and may be computationally expensive without incurring high distortion. Instead, maintaining the *local* structure of the data – the geometry and relationships of "neighboring" points – is of primary interest. This motivates the formal study of *local dimension reduction*, initiated by Abraham, Bartal, and Neiman [1] (see also the expanded journal version [3]). They proposed several novel formalisms [1, Definition 1] to capture the notion of local distortion that are based on the metric space's intrinsic structure, such as *k-local distortion*: an embedding f from a metric space (X, d_X) to a metric space (Y, d_Y) is said to have k-local distortion α if $d_Y(f(u), f(v)) \leq d_X(u, v)$ for all $u, v \in X$ and $d_Y(f(u), f(v)) \geq d_X(u, v)/\alpha$ for every u and every v which is among the k-nearest neighbors of u.

The results of [1] demonstrated that for their metric-based notions of local distortion, it is often possible to achieve embeddings with distortion or dimension dependent only on the locality parameter (e.g. k, in the notion of k-local distortion above) rather than the total number of points n. For instance, improving their result [1, Theorem 2], they showed in follow up work [2, Theorem 1] that any metric space (on n points) can be embedded into $\ell_p^{O(e^p \log^2 k)}$ with k-local distortion $O(\log k/p)$ (for any $k \leq n$, $1 \leq p \leq \log k$). In the low distortion setting of the JL lemma, they showed [2, Theorem 2] that for any $\varepsilon > 0$ and $p \geq 1$, an ultrametric space admits an embedding into $\ell_p^{O(\log k/\varepsilon^3)}$ with k-local distortion $(1+\varepsilon)$.

A natural question (asked in [1, Section 11]) is whether the ultrametricity assumption in the above result can be removed; specifically, is there a k-local version of the JL lemma, i.e. can any finite set of points in ℓ_2 be embedded into $\ell_2^{O(\log k/\varepsilon^2)}$ with distortion $(1+\varepsilon)$? This was answered in the negative by Schechtman and Shraibman [13], who constructed a set of $n+1$ points $X \subseteq \mathbb{R}^n$ such that any embedding $f : X \to \ell_2^m$ with 3-local distortion $(1+\varepsilon)$ (for a sufficiently small constant $\varepsilon > 0$) must have dimension $m = \Omega(\log n)$ [13, Theorem 9] (see also [13, Theorem 8] which obtains a lower bound with near optimal ε dependence under more requirements on the embedding).

Graphically local dimension reduction. The focus of this article is a different, graphical notion of local dimension reduction, which was studied by Schechtman and Shraibman [13, Section 5].

Definition 1.1. Let (X, d_X) and (Y, d_Y) be metric spaces. Let $G = (X, E)$ be a graph whose vertices are points of X. We say that a (not-necessarily injective) map $f : (X, d_X) \to (Y, d_Y)$ is a G-local D-embedding, where $D \geq 1$ is a real number, if there is a real number $r > 0$ such that

$$r \cdot d_X(x, y) \leq d_Y(f(x), f(y)) \leq D \cdot r \cdot d_X(x, y) \qquad \forall \{x, y\} \in E. \tag{1.1}$$

The infimum of the numbers $D \geq 1$ such that f is a G-local D-embedding is called the G-local distortion of f.

Remark. In the case that G is the complete graph on X, this reduces to the usual notion of (global) distortion. Also, k-local distortion essentially corresponds to the case when G is the k-nearest neighbour graph on X.

The natural choice of locality parameter in this setting is the maximum degree Δ of the graph G. In a similar spirit as the work of Abraham, Bartal, and Neiman [1], one can ask whether it is possible to achieve embeddings with G-local distortion or dimension dependent only on this locality parameter Δ. The possibility of this is suggested by the following example, highlighted in both [2] and [13].

Example 1.2. Let (X, d_X) be the set of points $X = \{e_1, \ldots, e_n\} \subseteq \mathbb{R}^n$ endowed with the Euclidean metric. Alon [4] showed that any embedding $f : (X, d_X) \to \ell_2^m$ with distortion $(1 + \varepsilon)$ must satisfy $m = \Omega(\log n/(\log(1/\varepsilon)\varepsilon^2))$. However, for any $G = (X, E)$ of maximum degree Δ, one may construct a G-local embedding $g : X \to \ell_2^{O(\log \Delta/\varepsilon^2)}$ with distortion $(1 + \varepsilon)$ as follows. First, since G has maximum degree at most Δ, we may greedily find a map $h : X \to \{e_1, \ldots, e_{\Delta+1}\} \subseteq \ell_2^{\Delta+1}$ such that $h(u) \neq h(v)$ if $\{u, v\} \in E$. The crucial property of h is that it is a G-local isometry i.e for any $\{u, v\} \in E$, $d_X(u, v) = \|h(u) - h(v)\|_2$. Now, we compose h with a Johnson-Lindenstrauss map from $\ell_2^{\Delta+1}$ to $\ell_2^{O(\log \Delta/\varepsilon^2)}$ to obtain the desired embedding g.

Since the set in Example 1.2 is a standard "hardness" example for the JL lemma, one may ask whether there is a G-local version of the JL lemma depending on the maximum degree Δ. For instance, this was asked in [9], where it was observed that if this were true, the embedding achieving this result would necessarily have to be non-linear. To the best of our knowledge, the only lower bound known for this problem is due to Schechtman and Shraibman. In [13, Theorem 11], they show that for X as in Example 1.2 and for any d-regular $G = (X, E)$ with $d \geq 3$ and second eigenvalue bounded by $d/2$, any G-local $(1 + \varepsilon)$-embedding $f : X \to \ell_2^m$ *which is non-contracting* (i.e. satisfies $\|f(x) - f(y)\|_2 \geq (1 - \varepsilon)\|x - y\|_2$ for *all* $x, y \in X$) must satisfy $m = \Omega(\log n)$. Of course, the non-contracting assumption precludes the embedding of Example 1.2, which achieves $m = O(\log \Delta/\varepsilon^2)$.

Our contribution. We show that in many settings of interest, including those of the JL lemma and Bourgain's embedding theorem, G-local embeddings do not provide *any* asymptotic saving in the dimension/distortion over global embeddings. This is in sharp contrast to metrically local embeddings (as in [1]) for which such savings are possible in high distortion regimes, as discussed above.

Our main technical contribution is a general reduction (Theorem 2.2), which takes as input a metric space (X, d_X) and constructs a metric extension (Z, d_Z) along with a graph $G = (Z, E)$ of maximum degree 3 such that any map $f : (Z, d_Z) \to (Y, d_Y)$ which (approximately) preserves distances for points in Z connected by an edge must necessarily (approximately) preserve *all* pairwise distances among points in X (in fact, there is a version of this statement for any distortion D). By applying this reduction with (X, d_X) being a known hard instance for some metric embedding problem, we are able to construct hard instances for G-local metric embeddings. We illustrate this with two applications: Corollary 2.6, which is a lower bound for G-local JL which is optimal up to a constant (unless ε is too small as a function of n) and Corollary 2.5, which is a lower bound for G-local embeddings of arbitrary finite metric spaces into Euclidean space, which is again optimal up to a constant. We also highlight some open problems.

2. Statements and proofs

In light of the equilateral space example (Example 1.2) and anticipating an open problem, we will track the *aspect ratio* of the metric spaces appearing in the statements of our results.

Definition 2.1. Let (X, d_X) be a finite metric space. The aspect ratio of X is defined to be the ratio of the maximum to minimum distance in X, i.e.

$$A_X := \frac{\max_{x \neq y} d_X(x, y)}{\min_{x \neq y} d_X(x, y)}.$$

Our main result is the following.

Theorem 2.2. *Let (X, d_X) be a metric space with $|X| = n$ and aspect ratio A_X.*

Suppose for a metric space (Y, d_Y), there exists a real number $D \geq 1$ such that any embedding $f : (X, d_X) \to (Y, d_Y)$ has distortion at least D. Then, for any $\delta \in (0, 1/100)$, there exists:

(1) a metric space (Z, d_Z) – with $|Z| = O(n^2)$ and aspect ratio $A_Z = O(A_X \cdot D^2 n/\delta)$ – and a graph $G = (Z, E)$ with maximum degree 3 such that for any $h : (Z, d_Z) \to (Y, d_Y)$, the G-local distortion of h is at least $D - \delta$.

(2) a metric space (Z, d_Z) – with $|Z| = O(n^2)$ and aspect ratio $A_Z = O(A_X \cdot D^2 \log n/\delta)$ – and a graph $G = (Z, E)$ with maximum degree 4 such that for any $h : (Z, d_Z) \to (Y, d_Y)$, the G-local distortion of h is at least $D - \delta$.

Moreover, if (X, d_X) is an ℓ_p-metric space for $p \in [1, \infty]$, then (Z, d_Z) can be taken to be an ℓ_p metric space as well.

Remark. We record several remarks about the optimality of this result.

(1) The bound of 3 on the maximum degree of G is the smallest possible. In Proposition 2.4, we show that for any finite metric space (X, d_X) and any graph $G = (X, E)$ of maximum degree at most 2, there exists a G-local isometric embedding $f : X \to \ell_p^2$. On the other hand, as discussed earlier, there are examples of (X, d_X) such that any embedding into ℓ_2 incurs distortion $\Omega(\log n)$.

(2) The bound $|Z| = O(n^2)$ cannot be improved in general. Indeed, Indyk and Wagner [8, Theorem 6.2] showed that $\Omega(n^2 \log(1/\varepsilon))$ bits are required to sketch finite metric spaces on n points for which all distances are between 1 and 2, up to distortion $(1 + \varepsilon)$ (see [8, Section 2] for formal definitions). Applying our theorem with $(Y, d_Y) = \ell_\infty$ (into which every metric embeds isometrically) and $\delta = \varepsilon$, it follows that storing the resulting metric (Z, d_Z) *restricted* to the edges of $G = (Z, E)$, up to relative error $(1 + \varepsilon)$, gives a sketch for (X, d_X) with distortion $(1 + 2\varepsilon)$. By rounding each of the $O(|Z|)$ distances we need to store to the nearest integer power of $(1 + \varepsilon)$ and storing the exponent (after appropriate scaling), this requires $O(|Z| \log(\log(n/\varepsilon)/\varepsilon)) = O(|Z| \log(1/\varepsilon) + |Z| \log \log(n/\varepsilon))$ bits, which contradicts the lower bound in [8] (say for $\varepsilon < 1/\log n$) unless $|Z| = \Omega(n^2)$.

(3) On the other hand, we do not know of any obstruction showing that the bounds on the aspect ratio A_Z are not improvable. In particular, it would be interesting if there is a construction with both $\Delta(G)$ and A_Z depending only on A_X, D, δ (in particular, independent of n). See the remark following Corollary 2.6.

Proof of Theorem 2.2. Since $|X| = n$, it follows that (X, d_X) can be isometrically embedded in ℓ_∞^n (e.g., using the Fréchet embedding, see [12]); in particular, we may identify X with a subset of points in ℓ_∞^n. By adding an arbitrary extra point to X if needed, we may (and will) assume without loss of generality that n is even. We will prove both conclusions of the theorem in a unified manner. Accordingly, our metric space and graph will be indexed by a parameter $j \in \{1, 2\}$, corresponding to the two different conclusions in the theorem statement.

Let $e_1, \ldots, e_n$ denote the standard basis of ℓ_∞^n and for $i \in [n]$, let $v_i^j := \alpha_j \cdot e_i$, where α_j will be chosen later. Consider the following set of points in ℓ_∞^n,

$$Z_j = X \cup \{x + v_i^j : x \in X, i \in [n-1]\}$$

and let (Z_j, d_Z^j) be the metric space induced by the ℓ_∞ metric on these points. Note that $|Z_j| = O(n^2)$, as claimed.

We will construct two graphs, $G_1 = (Z_1, E_1)$ and $G_2 = (Z_2, E_2)$ such that (Z_j, d_Z^j) and G_j will satisfy conclusion (j) of the theorem. Let K_n denote the complete graph on X (recall that $|X| = n$ is even, without loss of generality). By Baranyai's Theorem [6], the edges of K_n can be decomposed into a disjoint union of $(n-1)$ (perfect) matchings $M_1, \ldots, M_{n-1}$ (see Figure 1 for an illustration).

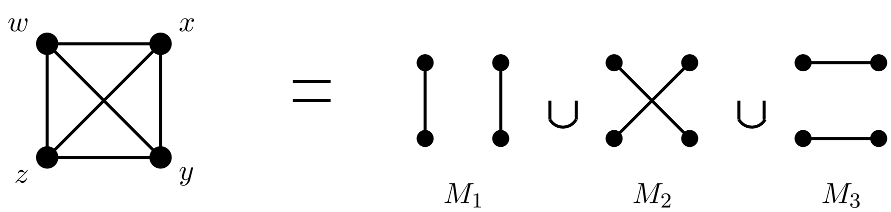

Figure 1. K_4 can be decomposed into three disjoint perfect matchings M_1, M_2, and M_3.

First, consider the collection of edges E'_j defined as follows:

$$E'_j = \{\{x + v_i^j, y + v_i^j\} : \{x, y\} \in M_i \text{ for some } i \in [n-1]\}.$$

In words, for each pair of points $\{x, y\}$ in X, let M_i be the unique matching in which the edge $\{x, y\}$ appears. Then, E'_j contains an edge between $x + v_i^j$ and $y + v_i^j$. Clearly, E'_j is itself a matching, i.e. has maximum degree 1.

The graphs $G_j = (Z_j, E_j)$ are obtained by augmenting E'_j as follows.

- To obtain E_1 from E'_1, we add the $|X|$ disjoint cycles of the form $x \to x + v_1^1 \to \cdots \to x + v_{n-1}^1 \to x$. Since the added edges constitute a 2-regular graph, it follows that $G_1 = (Z_1, E_1)$ is a graph of maximum degree 3. See Figure 2 for an illustration.
- To obtain E_2 from E'_2, we add $|X|$ disjoint arbitrary binary trees, rooted at $x \in X$, where the different binary trees span the points $\{x, x + v_1^2, \ldots, x + v_{n-1}^2\}$ for different choices of $x \in X$. Since the additional edges constitute a graph of maximum degree 3, it follows that $G_2 = (Z_2, E_2)$ has maximum degree 4. See Figure 3 for an illustration.

To complete our construction, it remains to specify our choice of α_j. Let $\gamma := \min_{x \neq y} d_X(x, y)$ be the minimum distance between two distinct points of X. We set $\alpha_1 = \delta\gamma/(8D^2 n)$ and $\alpha_2 = \delta\gamma/(8D^2 \log n)$. With this choice of parameters, the assertion about the aspect ratio of (Z_j, d_Z^j) is immediate. Additionally, we have the following.

Claim 2.3. *Suppose the map $h_j : (Z_j, d_Z^j) \to (Y, d_Y)$ is a G_j-local D'-embedding with $D' \leq D$ (i.e. satisfies (1.1) with some $r > 0$ and $1 \leq D' \leq D$). Then, for any $x \in X$ and $i \in [n-1]$,*

$$d_Y(h_j(x), h_j(x + v_i^j)) \leq r \cdot \delta / 8D \cdot \gamma$$

Proof. By construction, there is a path from x to $x + v_i^j$ of length at most $\lceil n/2 \rceil$ in E_1 and length at most $\lceil \log n \rceil$ in E_2; in either case, the edges of the path are of the form $\{x, x + v_k^j\}$ or $\{x + v_k^j, x + v_\ell^j\}$. Denoting the points along the path by $x = y_1^j, y_2^j, \ldots, y_{s_j}^j = x + v_i^j$, we have

$$d_Y(h_j(x), h_j(x + v_i^j)) \leq d_Y(h_j(y_1^j), h_j(y_2^j)) + \cdots + d_Y(h_j(y_{s_j - 1}^j), h_j(y_{s_j}^j))$$

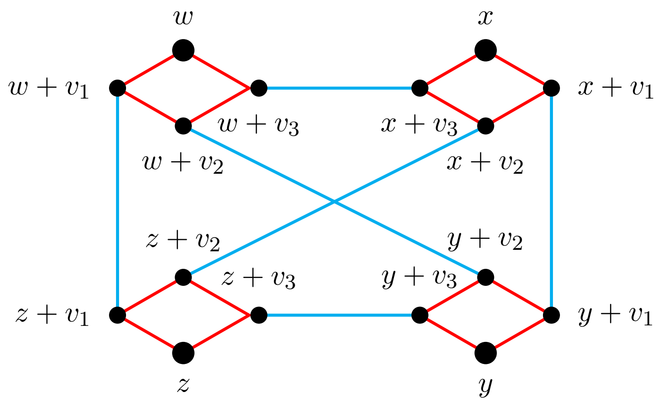

Figure 2. $G_1 = (Z_1, E_1)$ for $X = \{w, x, y, z\}$. The edges in E'_1, corresponding to the decomposition in Figure 1, are depicted in blue.

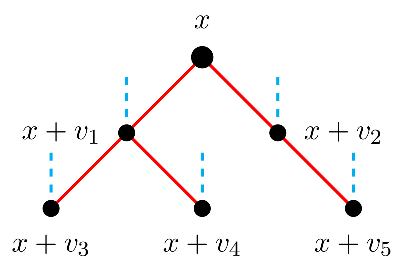

Figure 3. The binary tree gadget used in the construction of $G_2 = (Z_2, E_2)$ for $n = |X| = 6$.

$$
\begin{aligned}
&\leq r \cdot D \cdot \left(d_{Z_j}(y_1^j, y_2^j) + \cdots + d_{Z_j}(y_{s_j-1}^j, y_{s_j}^j) \right) \\
&\leq r \cdot D \cdot (s_j - 1) \cdot \alpha_j \\
&\leq r \cdot \delta / 8D \cdot \gamma;
\end{aligned}
$$

here, the first inequality is the triangle inequality; the second inequality uses that h_j is a G_j-local D'-embedding with $D' \leq D$; the third inequality uses $d_{Z_j}(y_k^j, y_{k+1}^j) = \alpha_j$, which is true by construction; and the last inequality uses our setting of α_j along with $s_1 \leq n, s_2 \leq \lceil \log n \rceil$. $\square$

With this claim in hand, we proceed with the proof of the theorem. Suppose for contradiction that there exists $h_j : (Z_j, d_Z^j) \to (Y, d_Y)$ which is G_j-local D'-embedding with $D' < D - \delta$. Let $f_j = (h_j)|_X : (X, d_X) \to (Y, d_Y)$ denote the restriction of h to X. We will show that the distortion of f_j is strictly less than D, thereby contradicting our assumption.

For $x, y \in X$, $x \neq y$, let $i \in [n-1]$ be such that $\{x + v_i^j, y + v_i^j\} \in E'_j \subseteq E_j$; recall that such a value of i is guaranteed to exist by construction. Then,

$$
\begin{aligned}
d_Y(f_j(x), f_j(y)) &= d_Y(h_j(x), h_j(y)) \\
&\leq d_Y(h_j(x), h_j(x + v_i^j)) + d_Y(h_j(x + v_i^j), h_j(y + v_i^j)) + d_Y(h_j(y + v_i^j), h_j(y)) \\
&\leq r \cdot (D - \delta) \cdot d_{Z_j}(x + v_i^j, y + v_i^j) + 2 \cdot r \cdot \delta / 8D \cdot \gamma \\
&= r \cdot (D - \delta) \cdot d_{Z_j}(x, y) + r \cdot \delta / 4D \cdot \gamma \\
&= r \cdot (D - \delta) \cdot d_X(x, y) + r \cdot \delta / 4D \cdot \gamma
\end{aligned}
$$

$$\leq r \cdot (D - \delta) \cdot (1 + \delta/4D) \cdot d_X(x, y);$$

here, the second line is the triangle inequality; the third line follows from Claim 2.3 and the assumption that h_j is a G_j-local $(D - \delta)$-embedding; the fourth line follows since the metric on Z_j is a norm; and the last line follows from the definition of γ.

Similarly, we have

$$\begin{aligned} d_Y(f_j(x), f_j(y)) &= d_Y(h_j(x), h_j(y)) \\ &\geq -d_Y(h_j(x), h_j(x + v_i^j)) + d_Y(h_j(x + v_i^j), h_j(y + v_i^j)) - d_Y(h_j(y + v_i^j), h_j(y)) \\ &\geq r \cdot d_{Z_j}(x + v_i^j, y + v_i^j) - 2 \cdot r \cdot \delta/8D \cdot \gamma \\ &= r \cdot d_X(x, y) - r \cdot \delta/4D \cdot \gamma \\ &\geq r(1 - \delta/4D) \cdot d_X(x, y). \end{aligned}$$

Combining the above two equations shows that the distortion of f is bounded above by $(D - \delta)(1 + \delta/4D)(1 - \delta/4D)^{-1}$, which is at most D by our assumption $\delta < 1/100$; this gives us the desired contradiction.

Finally, for the "moreover" part, if (X, d_X) is an ℓ_p metric space to start with, then we can view X as a subset of points in ℓ_p^N for $N = \binom{n}{2}$ (see, e.g. [12, Proposition 1.4.2]) and repeat the proof above. □

2.1. **Graphs of maximum degree two.** As remarked after the statement of Theorem 2.2, one cannot prove the theorem with graphs of maximum degree 2. This follows from the below proposition due to Sidhanth Mohanty, who has kindly agreed to let us include its proof here.

Proposition 2.4. *Let $X \subseteq \ell_p$ be a finite set of points. For any graph $G = (X, E)$ of maximum degree at most 2, there exists a G-local isometry (i.e. 1-embedding) $f : X \to \ell_p^2$.*

Proof. Since G has maximum degree 2, it is a disjoint union of paths and cycles. Therefore, it suffices to prove the proposition in the special case when G is a path and when G is a cycle. When G is a path, say $x_1 \to x_2 \to \cdots \to x_n$, one may even construct a G-local isometry $f : X \to \ell_p^1$ by setting $f(x_1) = 0$, and iteratively, setting $f(x_i) = f(x_{i-1}) + \|x_i - x_{i-1}\|_p$.

Next, consider the case when G is a cycle: $x_1 \to \ldots x_n \to x_1$. We will construct a G-local isometry $f : X \to \ell_p^2$. Initialize by setting $f(x_1) = 0$ and $f(x_n) = \|x_1 - x_n\|_p \cdot e_1$. We will construct $f(x_2), \ldots, f(x_{n-1})$ iteratively with the property that $\|f(x_i) - f(x_{i-1})\|_p = \|x_i - x_{i-1}\|_p$ and $\|f(x_i) - f(x_n)\|_p = \|x_i - x_n\|_p$ as follows. Let $S(x, r)$ denote the sphere in ℓ_p^2 of radius r, centered at x. Since $\|x_n - x_i\|_p + \|x_{i-1} - x_i\|_p \geq \|x_n - x_{i-1}\|_p$, it follows that $S(x_{i-1}, \|x_{i-1} - x_i\|_p) \cap S(x_n, \|x_n - x_i\|_p) \neq \emptyset$. To complete the iteration, let v_i denote an arbitrary point in this intersection and set $f(x_i) = v_i$. Figure 4 illustrates this construction for the case when G is a triangle: $x_1 \to x_2 \to x_3 \to x_1$. □

2.2. **Applications.** We conclude with two quick applications of Theorem 2.2. First, as discussed in the introduction, it was shown by Abraham, Bartal, and Neiman [2, Theorem 1] that any metric space on n points can be embedded into $\ell_p^{O(e^p \log^2 k)}$ with k-local distortion $O(\log k/p)$ (for any $k \leq n$ and $1 \leq p \leq \log k$). We show that for graphically local embeddings, this fails in a strong sense: below, we provide an example of an n point metric space and a graph G of maximum degree 3 such that any G-local embedding of this metric space into ℓ_2 incurs G-local distortion $\Omega(\log n)$ (recall that $O(\log n)$ *global* distortion is always achievable by Bourgain's theorem). We remark that a similar argument works for all ℓ_p spaces with fixed $p \geq 1$.

Corollary 2.5. *For any integer $n \geq 2$, there exists a metric space (Z, d_Z) with $|Z| = \Theta(n)$ and a graph $G = (Z, E)$ of maximum degree 3 such that any G-local embedding of (Z, d_Z) into ℓ_2 has G-local distortion $\Omega(\log n)$*

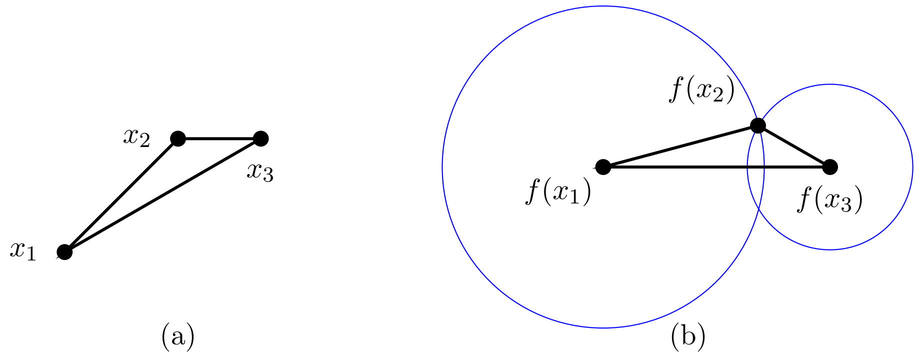

Figure 4. (a): G is a triangle: $x_1 \to x_2 \to x_3 \to x_1$.
(b): Construction of a G-local isometry to ℓ_p^2.

Proof. It is well known that there exists a metric space (X, d_X) on $\Theta(\sqrt{n})$ points such that any embedding of (X, d_X) into ℓ_2 incurs distortion $\Omega(\log n)$; for instance, one can take (X, d_X) to be the graph metric on a sufficiently good expander with $\Theta(\sqrt{n})$ vertices (see, e.g. [12, Theorem 3.5.3]). The statement now follows by applying part (1) of Theorem 2.2 to this with $D = 2\delta = \Theta(\log n)$. $\square$

Next, we show that that, in general, G-local dimension reduction is no easier than global dimension reduction, even for graphs of maximum degree 3. Previously, such a construction was only known under the additional restriction that the embedding is noncontracting (recall [13, Theorem 11] due to Schechtman and Shraibman).

Corollary 2.6. *For any integers $n, d \geq 2$, any $\varepsilon \in (\log^{0.51} n/\sqrt{\min(\sqrt{n}, d)}, 1/100)$ and $j \in \{1, 2\}$, there exist sets of points $Z_j \subseteq \mathbb{R}^d$ and graphs $G_j = (Z_j, E_j)$ of maximum degree $2 + j$ such that the following hold:*

- *$|Z_j| = O(n)$;*
- *the aspect ratio of the Euclidean metric on Z_1 is $O(\sqrt{n} \cdot \varepsilon^{-2})$ and on Z_2 is $O(\log n \cdot \varepsilon^{-2})$;*
- *for any map $f : Z_j \to \mathbb{R}^m$ satisfying*

$$\|x - y\|^2 \leq \|f(x) - f(y)\|_2 \leq (1 + \varepsilon)\|x - y\|_2 \qquad \forall \{x, y\} \in E_j,$$

  we must have

$$m = \Omega(\varepsilon^{-2} \log(\varepsilon^2 n)).$$

Remark. Our result leaves open the following questions.

(1) For global dimension reduction, Larsen and Nelson [11] are able to provide such a lower bound for every integer n, d and any $\varepsilon \in (\log^{0.51n}/\sqrt{\min(n, d)}, 1/100)$. Is the same lower bound also true for G-local dimension reduction for the entire range of ε? In our result, when $d \gg \sqrt{n}$, the range of ε is more limited than in [11].

(2) More interestingly, do there exist lower bound instances with bounded aspect ratio? Or is it the case that for every Euclidean metric space X on n points whose distances are all between 1 and Φ, and for every graph $G = (X, E)$ with maximum degree Δ, for every $\varepsilon > 0$, there exists a G-local embedding of X into ℓ_2^m with distortion $(1 + \varepsilon)$ and with $m = f(\varepsilon, \Delta, \Phi)$? Observe that this is true for the special case $\Phi = 1$ (Example 1.2). Also note that, since storing all distances appearing in G to $(1 + \varepsilon)$ multiplicative error only requires $O(n\Delta \log(1/\varepsilon) + n\Delta \log \log \Phi)$ bits, the existence of such an embedding cannot be ruled out simply by sketching arguments (for the JL lemma, the tight lower bound can be obtained from lower bounds on sketching, see [5]).

Proof. Let $n' := \lfloor \sqrt{n} \rfloor$. The main result of Larsen and Nelson [11] shows that for any $n', d \geq 2$ and $\varepsilon \in (\log^{0.51} n'/\sqrt{\min(n', d)}, 1/100)$, there exists a set of points $X \subseteq \mathbb{R}^d$ of size n' with aspect ratio $O(\varepsilon^{-1})$ such that for the Euclidean metric space (X, d_X), any embedding $f : (X, d_X) \to \mathbb{R}^m$ with distortion $(1 + 2\varepsilon)$ must satisfy $m = \Omega(\varepsilon^{-2} \log(\varepsilon^2 n))$. The result now follows by applying the "moreover" part of Theorem 2.2 to (X, d_X) and $(\mathbb{R}^m, \|\cdot\|_2)$ with $D = 1 + 2\varepsilon$ and $\delta = \varepsilon$. $\square$

References

[1] Ittai Abraham, Yair Bartal, and Ofer Neiman, *Local embeddings of metric spaces*, Proceedings of the thirty-ninth annual ACM Symposium on Theory of Computing, 2007, pp. 631–640. 1, 2

[2] Ittai Abraham, Yair Bartal, and Ofer Neiman, *On low dimensional local embeddings*, Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2009, pp. 875–884. 1, 2, 6

[3] Ittai Abraham, Yair Bartal, and Ofer Neiman, *Local embeddings of metric spaces*, Algorithmica **72** (2015), 539–606. 1

[4] Noga Alon, *Problems and results in extremal combinatorics—I*, Discrete Mathematics **273** (2003), 31–53. 2

[5] Noga Alon and Bo'az Klartag, *Optimal compression of approximate inner products and dimension reduction*, 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2017, pp. 639–650. 7

[6] Zsolt Baranyai, *On the factrization of the complete uniform hypergraphs*, Infinite and Finite Sets **1** (1974). 4

[7] Jean Bourgain, *On lipschitz embedding of finite metric spaces in Hilbert space*, Israel Journal of Mathematics **52** (1985), 46–52. 1

[8] Piotr Indyk and Tal Wagner, *Near-optimal (euclidean) metric compression*, Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2017, pp. 710–723. 3

[9] Vishesh Jain, Huy Pham, and Thuy-Duong Vuong, *Dimension reduction for maximum matchings and the fastest mixing Markov chain*, Comptes Rendus. Mathématique **361** (2023), 869–876. 2

[10] William B Johnson, Joram Lindenstrauss, et al., *Extensions of Lipschitz mappings into a Hilbert space*, Contemporary mathematics **26** (1984), 1. 1

[11] Kasper Green Larsen and Jelani Nelson, *Optimality of the Johnson-Lindenstrauss lemma*, 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), 2017, pp. 633–638. 1, 7, 8

[12] Jiri Matoušek, *Lecture notes on metric embeddings*, Tech. report, Technical report, ETH Zürich, 2013. 1, 3, 6, 7

[13] Gideon Schechtman and Adi Shraibman, *Lower bounds for local versions of dimension reductions*, Discrete & Computational Geometry **41** (2009), 273–283. 1, 2, 7