Optimality of the Johnson-Lindenstrauss Dimensionality Reduction for Practical Measures

Yair Bartal ✉
Hebrew University, Jerusalem, Israel

Ora Nova Fandina ✉
Aarhus University, Denmark

Kasper Green Larsen ✉
Aarhus University, Denmark

───── **Abstract** ─────

It is well known that the Johnson-Lindenstrauss dimensionality reduction method is optimal for worst case distortion. While in practice many other methods and heuristics are used, not much is known in terms of bounds on their performance. The question of whether the JL method is optimal for practical measures of distortion was recently raised in [8] (NeurIPS'19). They provided upper bounds on its quality for a wide range of practical measures and showed that indeed these are best possible in many cases. Yet, some of the most important cases, including the fundamental case of average distortion were left open. In particular, they show that the JL transform has $1 + \epsilon$ average distortion for embedding into k-dimensional Euclidean space, where $k = O(1/\epsilon^2)$, and for more general q-norms of distortion, $k = O(\max\{1/\epsilon^2, q/\epsilon\})$, whereas tight lower bounds were established only for large values of q via reduction to the worst case.

In this paper we prove that these bounds are best possible for any dimensionality reduction method, for any $1 \le q \le O(\frac{\log(2\epsilon^2 n)}{\epsilon})$ and $\epsilon \ge \frac{1}{\sqrt{n}}$, where n is the size of the subset of Euclidean space.

Our results also imply that the JL method is optimal for various distortion measures commonly used in practice, such as *stress, energy* and *relative error*. We prove that if any of these measures is bounded by ϵ then $k = \Omega(1/\epsilon^2)$, for any $\epsilon \ge \frac{1}{\sqrt{n}}$, matching the upper bounds of [8] and extending their tightness results for the full range moment analysis.

Our results may indicate that the JL dimensionality reduction method should be considered more often in practical applications, and the bounds we provide for its quality should be served as a measure for comparison when evaluating the performance of other methods and heuristics.

1 Introduction

Dimensionality reduction is a key tool in numerous fields of data analysis, commonly used as a compression scheme to enable reliable and efficient computation. In metric dimensionality reduction subsets of high-dimensional spaces are embedded into a low-dimensional space, attempting to preserve metric structure of the input. There is a large body of theoretical and applied research on such methods spanning a wide range of application areas such as online algorithms, computer vision, network design, machine learning, to name a few.
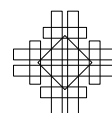
Metric embedding has been extensively studied by mathematicians and computer scientists over the past few decades (see [18, 25, 19] for surveys). developing a rich theory, and some original and elegant techniques have been developed and successfully applied in various fields of algorithmic research. See [27, 18, 34] for exposition of some applications.

The vast majority of these methods have been designed to optimize the worst-case distance error incurred by embedding. For metric spaces (X, d_X) and (Y, d_Y) an injective map $f : X \to Y$ is an embedding. It has (a worst-case) distortion $\alpha \geq 1$ if there is a positive constant c satisfying for all $u \neq v \in X$, $d_Y(f(u), f(v)) \leq c \cdot d_X(u, v) \leq \alpha \cdot d_Y(f(u), f(v))$. A cornerstone result in metric dimensionality reduction is the celebrated Johnson-Lindenstrauss lemma [21]. It states that any n-point subset of Euclidean space can be embedded, via a linear transform, into a $O(\log n/\epsilon^2)$-dimensional subspace with $1 + \epsilon$ distortion. It has been recently shown to be optimal in [24] and in [6] (improving upon [5]). Furthermore, it was shown in [26] that there are Euclidean pointsets in $\mathbb{R}^d$ for which any embedding into k-dimensions must have $n^{\Omega(1/k)}$ distortion, effectively ruling out dimensionality reduction into a constant number of dimensions with a constant worst-case distortion.

Metric embedding and in particular dimensionality reduction have also gained significant attention in applied community. Practitioners have frequently employed classic tools of metric embedding as well as have designed new techniques to cope with high-dimensional data. A large number of dimensionality reduction heuristics have been developed for a variety of practical settings, eg. [33, 28, 7, 36]. However, most of these heuristics have not been rigorously analyzed in terms of the incurred error. Recent papers [11] and [8] initiate the formal study of practically oriented analysis of metric embedding.

Practical distortion measures. In contrast to the worst case distortion the quality of practically motivated embedding is often determined by its average performance over all pairs, where an error per pair is measured as an additive error, a multiplicative error or a combination of both. There is a huge body of applied research investigating such notions of quality. For the list of citations and a more detailed account on the theoretical and practical importance of average distortion measures see [8].

In this paper we consider the most basic and commonly used in practical applications notions of average distortion, which we define in the following. Moment of distortion was defined in [4], and studied in various papers since then.

▶ **Definition 1** (ℓ_q-distortion). *For $u \neq v \in X$ let $expans_f(u, v) = d_Y(f(u), f(v))/d_X(u, v)$ and $contract_f(u, v) = d_X(u, v)/d_Y(f(u), f(v))$. Let $dist_f(u, v) = \max\{expans_f(u, v), contract_f(u, v)\}$. For any $q \geq 1$ the q-th moment of distortion is defined by*

$$\ell_q\text{-}dist(f) = \left(\frac{1}{\binom{|X|}{2}} \sum_{u \neq v \in X} (dist_f(u, v))^q \right)^{1/q}.$$

Additive average distortion measures are often used when a nearly isometric embedding is expected. Such notions as *energy*, *stress* and *relative error measure* (REM) are common in various statistic related applications. For a map $f : X \to Y$, for a pair $u \neq v \in X$ let $d_{u,v} := d_X(u, v)$ and let $\hat{d}_{uv} := d_Y(f(u), f(v))$. For all $q \geq 1$

▶ **Definition 2** (Additive measures).

$$Energy_q(f) = \left(\frac{1}{\binom{|X|}{2}} \sum_{u \neq v \in X} \left(\frac{|\hat{d}_{uv} - d_{uv}|}{d_{u,v}} \right)^q \right)^{\frac{1}{q}} = \left(\frac{1}{\binom{|X|}{2}} \sum_{u \neq v \in X} |expans_f(u, v) - 1|^q \right)^{\frac{1}{q}}.$$

$$Stress_q(f) = \left(\frac{\sum_{u \neq v \in X} |\hat{d}_{uv} - d_{uv}|^q}{\sum_{u \neq v \in X} (d_{uv})^q} \right)^{\frac{1}{q}}, \quad Stress^*{}_q(f) = \left(\frac{\sum_{u \neq v \in X} |\hat{d}_{uv} - d_{uv}|^q}{\sum_{u \neq v \in X} (\hat{d}_{uv})^q} \right)^{\frac{1}{q}}.$$

$$REM_q(f) = \left(\frac{1}{\binom{|X|}{2}} \sum_{u \neq v \in X} \left(\frac{|\hat{d}_{uv} - d_{uv}|}{\min\{\hat{d}_{uv}, d_{uv}\}} \right)^q \right)^{\frac{1}{q}}.$$

▷ **Claim 3** ([8]). For all $q \geq 1$, ℓ_q-$dist(f) - 1 \geq REM_q(f) \geq Energy_q(f)$.

In the full version we also address the machine learning motivated σ-distortion [12].

In [8] the authors rigorously analyzed dimensionality reduction for the above distortion measures. The central question they studied is: *What dimensionality reduction method is optimal for these quality measures and what are the optimal bounds achievable ? In particular, is the Johnson-Lindenstrauss (JL) transform also optimal for the average quality criteria?*

Their analysis of the Gaussian implementation of the JL embedding [20] shows that any Euclidean subset can be embedded with $1 + \epsilon$ *average* distortion (ℓ_1-$dist$) into $k = O(1/\epsilon^2)$ dimensions. And for more general case of the q-moment of distortion, the dimension is $k = O(\max\{1/\epsilon^2, q/\epsilon\})$. However, tight lower bounds were proved only for large values of q, leaving the question of optimality of the most important case of small q, and particularly the most basic case of $q = 1$, unresolved.

For the additive average measures (stress, energy and others) they prove that a bound of ϵ can be achieved in dimension $k = O(q/\epsilon^2)$. They showed a tight lower bound on the required dimension only for $q \geq 2$, leaving the basic case of $q = 1$ also unresolved.

In this paper, we prove that indeed the Johnson-Lindenstrauss bounds are best possible for any dimensionality reduction for the full range of $q \geq 1$, for all the average distortion measures defined in this paper. We believe that besides theoretical contribution this statement may have important implications for practical considerations. In particular, it may affect the way the JL method is viewed and used in practice, and the bounds we give may serve a basis for comparison for other methods and heuristics.

Our results. We show that the guarantees given by the Gaussian JL implementation are the best possible for the average distortion measures. In particular, we prove

▶ **Theorem 4.** *Given any integer n and $\Omega(\frac{1}{\sqrt{n}}) < \epsilon < 1$, there exists a $\Theta(n)$-point subset of Euclidean space such that any map f of it into ℓ_2^k with ℓ_1-$dist(f) \leq 1 + \epsilon$ requires $k = \Omega(1/\epsilon^2)$.*

▶ **Theorem 5.** *Given any integer n, and $\Omega(\frac{1}{\sqrt{n}}) < \epsilon < 1$, and $1 \leq q <= O(\log(\epsilon^2 n)/\epsilon)$, there exists a $\Theta(n)$-point subset of Euclidean space such that any embedding of it into ℓ_2^k with ℓ_q-distortion at most $1 + \epsilon$ requires $k = \Omega(q/\epsilon)$.*

As ℓ_q-distortion is monotonically increasing as a function of q, the theorems imply the lower bound of $k = \Omega\left(\max\left\{1/\epsilon^2, q/\epsilon\right\}\right)$. For the additive distortion measures we prove:

▶ **Theorem 6.** *Given any integer n and $\Omega(\frac{1}{\sqrt{n}}) < \epsilon < 1$, there exists a $\Theta(n)$-point subset of Euclidean space such that any embedding of it into ℓ_2^k with any of $Energy_1$, $Stress_1$, $Stress^*_1$, REM_1 or σ-distortion bounded above by ϵ requires $k = \Omega(1/\epsilon^2)$.*

Our main proof is of the lower bound for $Energy_1$ measure, which we show to imply the bound in Theorem 4 and for all measures in Theorem 6, with some small modifications for the stress measures. Furthermore, since all additive measures are monotonically increasing with

q the bounds hold for all $q \geq 1$. Therefore Theorems 4 and 5 together provide a tight bound of $\Omega(\max\{1/\epsilon^2, q/\epsilon\})$ for the ℓ_q-distortion. Additionally combined with the lower bounds of [8] for $q \geq 2$, Theorem 6 provides a tight bound of $\Omega(q/\epsilon^2)$ for all additive measures.

Techniques. The proofs of the lower bounds in all the theorems are based on counting argument, as in the lower bound for the worst case distortion proven by [24]. We extend the framework of [24] to the entire range of q moments of distortion, including the average distortion. As in the original proof we show that there exists a large family $\mathcal{P}$ of metric spaces that are quite different from each other so that if one can embed all of these into a Euclidean space with a small average distortion the resulting image spaces are different too. This implies that if the target dimension k is too small there is not enough space to accommodate all the different metric spaces from the family.

Let us first describe the framework of [24].[1] The main idea is to construct a large family of n-point subspaces $I \subseteq \ell_2^{\Theta(n)}$ so that each subspace in the family can be uniquely encoded using a small number of bits, assuming that each I can be embedded with $1 + \epsilon$ worst-case distortion in ℓ_2^k. The sets they construct are such that the information on the inner products between all the points in I, even if distorted by an additive error of $O(\epsilon)$, enables full reconstruction of the points in the set. In particular, each I consists of a zero vector together with the standard basis vectors E and an additional set of vectors denoted by Y. The set Y is defined in such a way that $\langle y, e \rangle \in \{0, c\epsilon\}$, for a constant $c > 1$, for all $(y, e) \in Y \times E$. The authors then show that a $1 + \epsilon$ distortion embedding f of I must map all the points into the ball of radius 2 while preserving all the inner products up to an additive error $\Theta(\epsilon)$, which enables to recover the vectors in Y. The next step is to show that all image points can be encoded using a small number of bits, while preserving the inner product information up to an $\Theta(\epsilon)$ additive error. This can be achieved by carefully discretizing the ball, and applying a map $\tilde{f}$ mapping every point to its discrete image approximation so that $\langle f(v), f(u) \rangle = \langle \tilde{f}(v), \tilde{f}(u) \rangle \pm \Theta(\epsilon)$. For this purpose one may use the method of [6] who showed[2] that randomly rounding the image points to the points in a small enough grid will preserve all the pairwise inner products within $\Theta(\epsilon)$ additive error with constant probability, and this in turn allows to derive a short binary encoding for each input point. This implies the lower bound on $k = \Omega(\log(\epsilon^2 n)/\epsilon^2)$, for $\epsilon = \Omega(1/\sqrt{n})$.

Let us now explain the challenges in applying this method to the case of bounded average distortion and q-moments. Assuming $f : I \to \ell_2^k$ has $1 + \epsilon$ average distortion neither implies that all images are in a ball of constant radius nor that f preserves all pairwise inner products. The bounded average distortion also does not guarantee the existence of a large subset of I with the properties above. We suggest the following approach to overcoming these issues. First, we add to I a large number of "copies" of 0 vectors which enables to argue that a large subset $\hat{I} \subseteq I$ will be mapped into a constant radius ball, such that the average additive distortion is $\Theta(\epsilon)$. The next difficulty is that if the images would be rounded to the points in a grid using a mapping which would preserve *all* pairwise inner products with $\Theta(\epsilon)$ additive error, then the resulting grid would be too large, which would't allow a sufficiently short encoding. We therefore round the images to a grid with $\Theta(\epsilon)$ additive approximation to the *average* of the inner products of $\hat{I}$ and thus reduce the size of the grid (and the encoding). The next step is showing that the above guarantees imply the existence of a large enough

---

[1] The description is based on combining the methods of [24, 6], and can be also viewed as our q-moments bound with $q = \Theta(\log(\epsilon^2 n)/\epsilon)$.

[2] The original proof of [24] uses a different elegant discretization argument.

subset of pairs $\mathcal{Z} \subseteq \binom{\text{I}}{2}$ of special structure, which allows to encode the *entire* set I with a few bits even if only the partial information about the inner products within $\mathcal{Z}$ is approximately preserved. In particular, we show that there is a large subset $\mathcal{Y}^G \subseteq Y$ such that for each point $y \in \mathcal{Y}^G$ there is a large enough subset $\mathcal{E}_y \subseteq E$ such that all pairwise inner products $\langle y, e \rangle$, where $y \in \mathcal{Y}^G$ and $e \in \mathcal{E}_y$, are additively preserved up to $\Theta(\epsilon)$ in the grid embedding, and therefore all the discretized images of these points have short binary encoding. The last step is to argue that this subset is sufficiently large so the knowledge of its approximate inner products possesses enough information in order to recover the entire point set I from our small size encoding. As this set still covers only a constant fraction of the pairs, and encoding the rest of the points is far more costly, this forces the dimension and number of points in our instance to be set to $d = \Theta(n) = \Theta(1/\epsilon^2)$, implying a lower bound of $k = \Omega(1/\epsilon^2)$. Finally, we prove that this can extend to arbitrary large subspaces via metric composition techniques. To extend these ideas to the general case of q-moments of distortion we prove that similar additive approximation distortion bounds hold with high probability of at least $1 - e^{-\Theta(\epsilon q)}$. This means that a smaller fraction of the pairs require a more costly encoding, and allows us to set $d = \Theta(n) = \Theta(1/\epsilon^2) \cdot e^{\Theta(\epsilon q)}$, implying a lower bound of $k = \Omega(q/\epsilon)$.

Related work. The study of "beyond the worst-case" distortion analysis of metric embedding initiated in [22] by introducing partial and scaling distortions. This has generated a rich line of follow up work, [1, 4, 2] just to name a few. The notions of average distortion and ℓ_q-distortion were introduced in [4] who gave bounds on embedding general metrics in normed spaces. Other notions of refined distortion analysis considered in the literature include such notions as Ramsey type embeddings [9], local distortion embeddings [3], terminal and prioritized distortion [15, 14], and recent works on distortion of the q-moments[3][29, 30, 23].

In applied community, various notions of average distortion are frequently used to infer the quality of heuristic methods [17, 16, 32, 13, 31, 35, 10].

However, the only work rigorously analyzing these notions we are aware of is that of [8]. They proved lower bounds of $k = \Omega(1/\epsilon)$ for the all additive measures average (1-norm) version, and for the average distortion measure (ℓ_1-distortion), which we improve here to the tight $\Omega(1/\epsilon^2)$ bound. For $q \geq 2$ they gave tight bounds of $\Omega(q/\epsilon^2)$ for all additive measures. For ℓ_q-*dist* they have shown that for $q = \Omega(\log(1/\epsilon)/\epsilon)$ the tight bound of $k = \Omega(q/\epsilon)$ follows from the black-box reduction to the lower bound on the worst case distortion.

2 Lower bound for average distortion and additive measures

In this section we prove Theorems 4 and Theorem 6. Using Claim 3, we may focus on proving the lower bound for $Energy_1(f)$ in order to obtain similar lower bounds for $REM_1(f)$ and ℓ_1-$dist(f)$. In full version of the paper we show how to change this proof in order to obtain lower bound on $Stress_1(f)$, and further show that the lower bounds for all the additive measures follow from the lower bounds on Energy and Stress.

We present here the proof of an existence of a bad metric space of size $\hat{n} = \Theta(1/\epsilon^2)$, while construction of a metric space of an arbitrary size $n \geq \hat{n}$, based on a similar technique appearing in [8] via metric composition [9], is given in the full version of the paper.

---

[3] The notion in these papers, also studied [4, 8], computes the ratio between the average of (or q-moments) of new distances to that of original distances, in contrast to the average distortion (or q-moments of distortion) measure in Definition 1, which measures the average (or q-moments) of pairwise distortions.

We construct a large family $\mathcal{P}$ of metric spaces, such that each $I \in \mathcal{P}$ can be completely recovered by computing the inner products between the points in I. For a given $\epsilon < 0$, let $l = \lceil \frac{1}{\gamma^2 \epsilon^2} \rceil$, for some large constant $\gamma > 1$ to be determined later. We will prove $k \geq \frac{c}{\gamma^2} \frac{1}{\epsilon^2}$, for $c < 1$, and so we may assume w.l.o.g. that $\epsilon \leq 1/\gamma$, otherwise the statement trivially holds. We construct point sets $I \subset \ell_2^d$, where $d = 2l$, each I of size $3d = 6l = \Theta(1/\epsilon^2)$.

Define a set $O = \{o_j\}_{j=1}^d$ of d arbitrary near zero vectors in ℓ_2^d, i.e., a set of d different vectors such that for all $o_j \in O$, $\|o_j\|_2 \leq \epsilon/100$. Let $E = \{e_1, e_2, \ldots, e_d\}$ denote the vectors of the standard basis of $\mathbb{R}^d$. For a set S of l indices from $\{1, 2, \ldots, d\}$, we define $y_S = \frac{1}{\sqrt{l}} \sum_{j \in S} e_j$. For a sequence of d index sets (possibly with repetitions) $S_1, S_2, \ldots, S_d$, let $Y[S_1, \ldots, S_d] = \{y_{S_1}, \ldots, y_{S_d}\}$. Each point set $I[S_1, \ldots, S_d] \in \mathcal{P}$ is defined as the union of the sets defined above[4], i.e., $I[S_1, \ldots, S_d] = O \cup E \cup Y[S_1, \ldots, S_d]$. The size of the family is $|\mathcal{P}| = \binom{d}{l}^d$. Note that each $I \in \mathcal{P}$ is contained in $B_2(1)$, the unit ball of ℓ_2^d, and has diameter $diam(I) = \sqrt{2}$. Additionally, for all $e_j \in E$ and $y_S \in Y$ the value of the inner product $\langle e_j, y_S \rangle$ determines whether $e_j \in \text{span}\{e_i | i \in S\}$. In particular, if $\langle e_j, y_S \rangle = 0$ then $j \notin S$, and if $\langle e_j, y_S \rangle = 1/\sqrt{l} \geq (1/2)\gamma\epsilon$ then $j \in S$.

Assume that for each $I \in \mathcal{P}$ there is an embedding $f : I \to \ell_2^k$, with $Energy_1(f) \leq \epsilon$. We prove that this implies that $k = \Omega(1/\epsilon^2)$. The strategy is to produce a unique binary encoding of each I in the family based on the embedding f. Let length(I) denote the length of the encoding for each I, we will show that $\text{length}(I) \lesssim l^2 + l \cdot k \log(\frac{1}{\epsilon k})$. Since the encoding defines an injective map from $\mathcal{P}$ to $\{0, 1\}^{\text{length}(I)}$, the number of different sets that can be recovered by decoding is at most $2^{\text{length}(I)}$. Now, because $|\mathcal{P}| = \binom{d}{l}^d \geq 2^{2l^2}$ we get that $k \log(\frac{1}{\epsilon k}) \gtrsim l$ and show that this implies the bound on $k \geq \Omega(l)$.

We are now set to describe the encoding for each I and to bound its length. First, in the following lemma, we show that there exists a large subset $\hat{I} \subseteq I$ that is mapped by f into a ball of a constant radius in k-dimensional space and that the average of the errors in the inner products incurred by f on the subset $\hat{I}$ is bounded by $\Theta(\epsilon)$.

▶ **Lemma 7.** *For any $I \in \mathcal{P}$ let $f : I \to \ell_2^k$ be an embedding with $Energy_1(f) \leq \epsilon$, with $\epsilon \leq 1/36$. Let $0 < \alpha \leq 1/16$ be a parameter. There is a subset $\hat{I} \subseteq I$ of size $|\hat{I}| \geq (1 - \alpha)|I|$ such that $f(\hat{I}) \subset B_2\left(1 + \frac{3.01\epsilon}{\alpha}\right)$, and $\frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v)\in\binom{\hat{I}}{2}} |\langle f(u), f(v)\rangle - \langle u, v\rangle| \leq (10 + \frac{1}{2\alpha})\epsilon$.*

Proof. By assumption we have that the following condition holds:

$$Energy_1(f) = \frac{1}{|\binom{I}{2}|} \sum_{(u,v)\in\binom{I}{2}} \left|expans_f(u,v) - 1\right| \leq \epsilon. \tag{1}$$

This bound implies that

$$\frac{1}{|O|(|I|-1)} \sum_{o_j \in O} \sum_{v \in I, v \neq o_j} \left|expans_f(o_j, v) - 1\right| \leq \frac{1}{|O|(|I|-1)} \sum_{u \neq v \in I} \left|expans_f(u,v) - 1\right|$$

$$\leq \frac{3d(3d-1)}{d(3d-1)}\epsilon = 3\epsilon.$$

From which follows that

$$\min_{o_j \in O} \frac{1}{|I|-1} \sum_{v \in I, v \neq o_j} \left|expans_f(o_j, v) - 1\right| \leq 3\epsilon. \tag{2}$$

---

[4] We will omit $[S_1, \ldots S_d]$ from notation for a fixed choice of the sets.

Let $\hat{o} \in O$ denote the point at which the minimum is obtained. We assume without loss of generality that $f(\hat{o}) = 0$. Let $\hat{I}$ be the set of all $v \in I$ such that $\left| expans_f(\hat{o}, v) - 1 \right| \leq \frac{3\epsilon}{\alpha}$. By Markov's inequality, $|\hat{I}| \geq (1-\alpha)|I|$. We have that for all $v \in \hat{I}$, $|expans_f(v, \hat{o}) - 1| = |\frac{\|f(v)\|_2}{\|v - \hat{o}\|_2} - 1| \leq \frac{3\epsilon}{\alpha}$, and using $\|v - \hat{o}\|_2 \leq \|v\|_2 + \|\hat{o}\|_2 \leq 1 + \epsilon/100$, so that $\|f(v)\|_2 \leq (1 + \frac{3\epsilon}{\alpha})(1 + \epsilon/100) \leq 1 + \frac{3.002\epsilon}{\alpha}$, implying that $f(v) \in B_2\left(1 + \frac{3.01\epsilon}{\alpha}\right)$.

For all $(u, v) \in \binom{\hat{I}}{2}$ we have:

$$
\begin{aligned}
\left| \langle f(u), f(v) \rangle - \langle u, v \rangle \right| \quad \leq \quad & \frac{1}{2}\left[\left| \|f(u)\|_2^2 - \|u\|_2^2 \right| + \left| \|f(v)\|_2^2 - \|v\|_2^2 \right| \right] \\
+ \quad & \frac{1}{2}\left[\left| \|f(u) - f(v)\|_2^2 - \|u - v\|_2^2 \right| \right].
\end{aligned}
$$

We can bound each term as follows:

$$
\begin{aligned}
\left| \|f(u)\|_2^2 - \|u\|_2^2 \right| = & \\
= \quad & \left| \|f(u) - f(\hat{o})\|_2^2 - \|u - \hat{o}\|_2^2 + \|u - \hat{o}\|_2^2 - \|u\|_2^2 \right| \\
\leq \quad & \left| \|f(u) - f(\hat{o})\|_2 - \|u - \hat{o}\|_2 \right| \cdot \left(\|f(u) - f(\hat{o})\|_2 + \|u - \hat{o}\|_2 \right) \\
+ \quad & \left| \|u - \hat{o}\|_2 - \|u\|_2 \right| \cdot \left(\|u - \hat{o}\|_2 + \|u\|_2 \right) \\
\leq \quad & \|u - \hat{o}\|_2 \cdot \left| expans_f(u, \hat{o}) - 1 \right| \cdot \left(\|f(u)\|_2 + \|u - \hat{o}\|_2 \right) + \|\hat{o}\|_2 \cdot \left(\|u - \hat{o}\|_2 + \|u\|_2 \right) \\
\leq \quad & \left(1 + \frac{\epsilon}{100}\right)\left| expans_f(u, \hat{o}) - 1 \right| \left(1 + \frac{3.002\epsilon}{\alpha} + 1 + \frac{\epsilon}{100}\right) + \frac{\epsilon}{100} \cdot \left(2 + \frac{\epsilon}{100}\right) \\
\leq \quad & \left(2 + \frac{3.01\epsilon}{\alpha}\right)\left| expans_f(u, \hat{o}) - 1 \right| + \frac{\epsilon}{40} \leq \left(2 + \frac{1}{9\alpha}\right)\left| expans_f(u, \hat{o}) - 1 \right| + \frac{\epsilon}{40},
\end{aligned}
$$

where we have used $\|\hat{o}\| \leq \epsilon/100$, $\|u - \hat{o}\|_2 \leq \|u\|_2 + \|\hat{o}\|_2 \leq 1 + \epsilon/100$, and the bound on the norms of the embedding within $\hat{I}$. Additionally, we have that

$$
\begin{aligned}
\left| \|f(u) - f(v)\|_2^2 - \|u - v\|_2^2 \right| = & \\
= \quad & \left| \|f(u) - f(v)\|_2 - \|u - v\|_2 \right| \left(\|f(u) - f(v)\|_2 + \|u - v\|_2 \right) \\
\leq \quad & \|u - v\|_2 \left| expans_f(u, v) - 1 \right| \left(\|f(u)\|_2 + \|f(v)\|_2 + \|u - v\|_2 \right) \\
\leq \quad & \sqrt{2}\left(2\left(1 + \frac{3.002\epsilon}{\alpha}\right) + \sqrt{2}\right)\left| expans_f(u, v) - 1 \right| \leq \left(5 + \frac{1}{4\alpha}\right)\left| expans_f(u, v) - 1 \right|,
\end{aligned}
$$

where the second to last inequality holds since $\|u - v\|_2 \leq diam(I) = \sqrt{2}$. It follows that:

$$
\frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v) \in \binom{\hat{I}}{2}} \left| \langle f(u), f(v) \rangle - \langle u, v \rangle \right| \leq \tag{3}
$$

$$
\begin{aligned}
\leq \quad & \left(2 + \frac{1}{9\alpha}\right) \cdot \frac{1}{|\binom{\hat{I}}{2}|}\left(\frac{|\hat{I}| - 1}{2}\right) \sum_{u \in \hat{I}, u \neq \hat{o}} \left| expans_f(u, \hat{o}) - 1 \right| \\
+ \quad & \frac{1}{2}\left(5 + \frac{1}{4\alpha}\right) \cdot \frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v) \in \binom{\hat{I}}{2}} \left| expans_f(u, v) - 1 \right| + \frac{\epsilon}{40}.
\end{aligned}
$$

By definition of $\hat{I}$, and using (2) we have that

$$
\begin{aligned}
\frac{1}{|\binom{\hat{I}}{2}|}\left(\frac{|\hat{I}| - 1}{2}\right) \sum_{u \in \hat{I}, u \neq \hat{o}} \left| expans_f(u, \hat{o}) - 1 \right| \quad = \quad & \frac{1}{|\hat{I}|} \sum_{u \in \hat{I}, u \neq \hat{o}} \left| expans_f(u, \hat{o}) - 1 \right| \\
\leq \quad & \frac{1}{|I|} \sum_{u \in I, u \neq \hat{o}} \left| expans_f(u, \hat{o}) - 1 \right| \leq 3\epsilon.
\end{aligned}
$$

Therefore (3) yields that

$$\frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v) \in \binom{\hat{I}}{2}} |\langle f(u), f(v) \rangle - \langle u, v \rangle| \leq$$

$$\leq \quad \left(2 + \frac{1}{9\alpha}\right) \cdot 3\epsilon + \frac{1}{2}\left(5 + \frac{1}{4\alpha}\right) \cdot \frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v) \in \binom{\hat{I}}{2}} |expans_f(u,v) - 1| + \frac{\epsilon}{40}$$

$$\leq \quad \frac{1}{2}\left(5 + \frac{1}{4\alpha}\right) \cdot \frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v) \in \binom{\hat{I}}{2}} |expans_f(u,v) - 1| + \left(7 + \frac{1}{3\alpha}\right)\epsilon.$$

Now, we have that

$$\frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v) \in \binom{\hat{I}}{2}} |(expans_f(u,v)) - 1| \leq \frac{6}{5}\frac{1}{|\binom{I}{2}|} \sum_{(u,v) \in \binom{I}{2}} |(expans_f(u,v)) - 1| \leq \frac{6}{5}\epsilon,$$

using $|\hat{I}| \geq (1-\alpha)|I|$, so that $\alpha \leq 1/16$ we have $|\binom{\hat{I}}{2}| \geq (1 - \frac{1}{3(1-\alpha)d})(1-\alpha)^2 \cdot |\binom{I}{2}| \geq \frac{5}{6}|\binom{I}{2}|$ and applying (1). Finally, we obtain

$$\frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v) \in \binom{\hat{I}}{2}} |\langle f(u), f(v) \rangle - \langle u, v \rangle| \leq \frac{6}{5} \cdot \frac{1}{2}\left(5 + \frac{1}{4\alpha}\right)\epsilon + \left(7 + \frac{1}{3\alpha}\right)\epsilon \leq \left(10 + \frac{1}{2\alpha}\right)\epsilon. \blacktriangleleft$$

We have shown thus far that for the large subset $\hat{I}$ of the set I, the average of the inner products between the images equals up to an additive factor $\Theta(\epsilon)$ to the average of the inner products between the original points. Moreover, all the images of $\hat{I}$ are in the constant radius ball. We next show that rounding these images to the (randomly chosen) points of the sufficiently small grid will not change the sum of the inner products too much, implying that instead of encoding the original images $f(\hat{I})$ we can encode its rounded counterpart. To show this, we use a technique of randomized rounding as proposed in [6].

▶ **Lemma 8.** *Let $X \subset \ell_2^k$ such that $X \subset B_2(r)$. For $\delta < r/\sqrt{k}$ let $G_\delta \subseteq B_2(r)$ denote the intersection of the δ-grid with $B_2(r)$. There is a mapping $g : X \rightarrow G_\delta$ such that $\frac{1}{|\binom{X}{2}|} \sum_{(u,v) \in \binom{X}{2}} |\langle g(u), g(v) \rangle - \langle u, v \rangle| \leq 3\delta r$, and the points of the grid can be represented using $L_{G_\delta} = k\log(4r/(\delta\sqrt{k}))$ bits.*

Proof. For each point $v \in X$ randomly and independently match a point $\tilde{v} = g(v)$ on the grid by rounding each of its coordinates v_i to one of the closets integral multiples of δ in such a way that $E[\tilde{v}_i] = v_i$. This distribution is given by assigning $\lceil \frac{v_i}{\delta} \rceil \delta$ with probability $p = (\frac{v_i}{\delta} - \lfloor \frac{v_i}{\delta} \rfloor)$, and assigning $\lfloor \frac{v_i}{\delta} \rfloor \delta$ with probability $1 - p$. For any $u \neq v \in X$ we have:

$$\begin{aligned}
E[|\langle \tilde{u}, \tilde{v} \rangle - \langle u, v \rangle|] &\leq E[|\langle \tilde{u} - u, v \rangle|] + E[|\langle \tilde{u}, \tilde{v} - v \rangle|] \\
&\leq (E[(\langle \tilde{u} - u, v \rangle)^2])^{1/2} + (E[(\langle \tilde{u}, \tilde{v} - v \rangle)^2])^{1/2},
\end{aligned}$$

where the last inequality is by Jensen's. We bound each term separately.

$$E[(\langle \tilde{u} - u, v \rangle)^2] = E\left[\left(\sum_{i=1}^{k} (\tilde{u}_i - u_i)v_i\right)^2\right] =$$

$$= \sum_{i=1}^{k} v_i^2 E[(\tilde{u}_i - u_i)^2] + 2\sum_{1 \leq i \neq j \leq k} v_i v_j E[\tilde{u}_i - u_i]E[\tilde{u}_j - u_j] \leq \delta^2 \|v\|_2^2$$

since $|\tilde{u}_i - u_i| \leq \delta$ and $E[\tilde{u}_i] = u_i$. Similarly, for the second term we have

$$E\left[(\langle\tilde{u}, \tilde{v} - v\rangle)^2\right] = E\left[\left(\sum_{i=1}^{k}\tilde{u}_i(\tilde{v}_i - v_i)\right)^2\right] \leq \sum_{i=1}^{k} E\left[\tilde{u}_i^2\right] E\left[(\tilde{v}_i - v_i)^2\right] \tag{4}$$

$$+ 2\sum_{1 \leq i \neq j \leq k} E[\tilde{u}_i\tilde{u}_j(\tilde{v}_i - v_i)]E[\tilde{v}_j - v_j] \leq \delta^2\sum_{i=1}^{k} E[\tilde{u}_i^2],$$

because the random variables $\tilde{u}_i$ and $\tilde{v}_i$ are independent. We also have that

$$\sum_{i=1}^{k} E[\tilde{u}_i^2] = \sum_{i=1}^{k} E[(u_i + (\tilde{u}_i - u_i))^2] = \sum_{i=1}^{k}\left(u_i^2 + 2u_i E[\tilde{u}_i - u_i] + E[(\tilde{u}_i - u_i)^2]\right) \leq \|u\|_2^2 + \delta^2 k.$$

Therefore, putting all together, $E[|\langle\tilde{u}, \tilde{v}\rangle - \langle u, v\rangle|] \leq \delta r + \delta(r^2 + \delta^2 k)^{1/2} \leq 2\delta r + \delta^2\sqrt{k} \leq 3\delta r$.

The bound on the average difference in inner product in the lemma follows by the linearity of expectation, and the implied existence of a mapping with bound at most the expectation. The upper bound on the representation of the grid points was essentially given in [6]: The ith coordinate of a point x on the grid is given by a sign and an absolute value $n_i\delta$, where $0 \leq n_i \leq r/\delta$ are integers satisfying $\sum_{1 \leq i \leq k} n_i^2 \leq (r/\delta)^2$. So can be represented by the sign and their binary representation of size at most $\sum_{i=1}^{k}(\log(n_i) + 1)$, which is maximized when all n_i^2's are equal, which gives the bound of $k\log(4r/(\delta\sqrt{k}))$. ◄

Combining the lemmas we obtain:

▶ **Corollary 9.** *For any $I \in \mathcal{P}$ let $f : I \to \ell_2^k$ be a map with $Energy_1(f) \leq \epsilon$, for $\epsilon \leq 1/36$. Let $0 < \alpha \leq 1/16$. There is $\hat{I} \subseteq I$ of size $|\hat{I}| \geq (1-\alpha)|I|$ such that there is a set $G \subset \ell_2^k$ and a map $g : \hat{I} \to G$ such that $\frac{1}{|\binom{\hat{I}}{2}|}\sum_{(u,v)\in\binom{\hat{I}}{2}}|\langle g(f(u)), g(f(v))\rangle - \langle u, v\rangle| \leq \left(13 + \frac{0.76}{\alpha}\right)\epsilon$, and the points in G can be uniquely represented by binary strings of length at most $L_G = k\log(4r/(\epsilon\sqrt{k}))$ bits, where $r < 1 + 0.09\frac{1}{\alpha}$.*

Proof. The corollary follows by applying Lemma 7 followed by Lemma 8 with $X = \hat{I}$ and $\delta = \epsilon$. Note that we may assume that $\epsilon = \delta < 1/\sqrt{k} < r/\sqrt{k}$, as otherwise we are done. ◄

We are now ready to obtain the main consequence which will imply the lower bound.

▶ **Corollary 10.** *For any $I \in \mathcal{P}$ let $f : I \to \ell_2^k$ be an embedding with $Energy_1(f) \leq \epsilon$, with $\epsilon \leq 1/36$. Let $0 < \alpha \leq 1/16$ and $0 < \beta$ be parameters. There is a subset of points G that satisfies the following: there is a subset $\mathcal{Y}^G \subseteq Y$ of size $|\mathcal{Y}^G| \geq (1 - 3\alpha - \frac{3}{\sqrt{2}}\beta)|Y|$ such that for each $y \in \mathcal{Y}^G$ there is a subset $\mathcal{E}_y^G \subseteq E$ of size $|\mathcal{E}_y^G| \geq (1 - 3\alpha - \frac{3}{\sqrt{2}}\beta)|E|$ such that for all pairs $(y, e) \in \mathcal{Y}^G \times \mathcal{E}_y^G$ we have: $|\langle g(f(y)), g(f(e))\rangle - \langle y, e\rangle| \leq \frac{1}{\beta^2}\left(13 + \frac{0.76}{\alpha}\right)\epsilon$, where $g : \mathcal{Y}^G \cup \{\mathcal{E}_y^G\}_{y\in\mathcal{Y}^G} \to G$. Moreover, the points in G can be uniquely represented by binary strings of length at most $L_G = k\log(4r/(\epsilon\sqrt{k}))$ bits, where $r < 1 + 0.09\frac{1}{\alpha}$.*

Proof. Applying Corollary 9 and Markov's inequality there are at most β^2 fraction of pairs $(u, v) \in \binom{\hat{I}}{2}$ such that $|\langle g(f(u)), g(f(v))\rangle - \langle u, v\rangle| > \frac{1}{\beta^2}\left(13 + \frac{0.76}{\alpha}\right)\epsilon$. It follows that the number of pairs in $Y \times E$ that are in $\binom{\hat{I}}{2}$ and have this property is at most $\beta^2 \cdot \frac{3d(3d-1)}{2} \leq \frac{9}{2}\beta^2 \cdot d^2$. Therefore there can be at most $\frac{3}{\sqrt{2}}\beta d$ points in $u \in Y$ such that there are more than $\frac{3}{\sqrt{2}}\beta d$ points in $v \in E$ with this property. Since there are at most $3\alpha d$ points in each of Y and E which may not be in $\hat{I}$ we obtain the stated bounds on the sizes of $|\mathcal{Y}^G|$ and $|\mathcal{E}_y^G|$. ◄

2.1   Encoding algorithm

Let $t = 8$. We set $\alpha = 1/(12t)$, $\beta = 1/(\sqrt{2}t)$, which implies that $r \leq 10$. Therefore, by Corollary 10, we can find a subset $G \subseteq B_2(10)$, and a mapping $g : f(I) \to G$, and a subset $\mathcal{Y}^G \subseteq Y$, with $|\mathcal{Y}^G| \geq \left(1 - \frac{1}{t}\right)|Y|$, where for all $y \in \mathcal{Y}^G$ we can find a susbet $\mathcal{E}_y^G \subseteq E$ with $|\mathcal{E}_y^G| \geq \left(1 - \frac{1}{t}\right)|E|$, such that for all pairs $(e, y) \in \mathcal{Y}^G \times \mathcal{E}_y^G$ the inner products $\left|\langle g(f(y)), g(f(e))\rangle - \langle y, e\rangle\right| \leq 12000\epsilon$. Moreover, each point in G can be uniquely encoded using at most $L_G = k\log(40/(\epsilon\sqrt{k}))$ bits.

We first encode all the points $Y \setminus \mathcal{Y}^G$. For each $y_S \in Y \setminus \mathcal{Y}^G$ we explicitly write down a bit for each $e_i \in E$ indicating whether $e_i \in S$. This requires d bits for each y_S and in total at most $\left(\frac{1}{t}\right)d^2$ bits for the subset $Y \setminus \mathcal{Y}^G$. The next step is to encode all the points in $\{\mathcal{E}_y^G\}_{y \in \mathcal{Y}^G}$ in a way tat will enable to recover all the vectors in the set together with the indeces. We can do that by writing an ordered list containing d strings (one for each vector in the set E, according to its order). Each string is of length L_G bits, where each point $e_i \in \{\mathcal{E}_y^G\}_{y \in \mathcal{Y}^G}$ is encoded by its representation in G, i.e., $g(f(e_i))$, and rest of points (if there are any) are encoded by zeros. This gives an encoding of total length dL_G bits.

Now we can encode the points in $\mathcal{Y}^G$. Each $y_S \in \mathcal{Y}^G$ is encoded by the encoding of $g(f(y_S))$ using L_G bits, and in addition we add the encoding of the set of indices of the points in $E \setminus \mathcal{E}_{y_S}^G$, using at most $\log\binom{d}{(1/t)d} \leq (1/t)d\log(et)$ bits. Note that this information is not enough in order to recover the vector y_S, as we can't deduce whether $i \in S$ for $e_i \in E \setminus \mathcal{E}_{y_S}^G$. So we add this information explicitly, by writing down whether $i \in S$ for each $e_i \in E \setminus \mathcal{E}_{y_S}^G$, using at most $(1/t)d$ bits. In total, it takes $L_G + (1/t)d\log(et) + (1/t)d$ bits per point in $\mathcal{Y}^G$.

Therefore, each instance $I \in \mathcal{P}$ can be encoded using at most

$$(1/t)d^2 + dL_G + |\mathcal{Y}^G| \cdot (L_G + d(1/t)\log(et) + (1/t)d) \leq (1/t)d^2(2 + \log(et)) + 2dL_G$$

bits, since $|\mathcal{Y}^G| \leq d$. For our choice of $t = 8$, this is at most $\frac{7}{8}d^2 + 2dL_G$.

2.2   Decoding algorithm

To recover a set I given the encoding it is enough to recover the set Y, as the sets O and E are the same in each I. We first recover $Y \setminus \mathcal{Y}^G$ in a straightforward way from its naive encoding. To recover a point $y_S \in \mathcal{Y}^G$ we need to know for each $e_i \in E$ whether $i \in S$. An important implication of Corollary 10 is that given $g(f(e_i))$ and $g(f(y_S))$ of any pair $(e_i, y_S) \in \mathcal{Y}^G \times \mathcal{E}_{y_S}^G$, we can decide whether $i \in S$ by computing $\langle g(f(e_i)), g(f(y_S))\rangle$. Recall that if $i \notin S$ then $\langle e_i, y_S\rangle = 0$, and if $i \in S$ then $\langle e_i, y_S\rangle \geq (1/2)\gamma\epsilon$. Therefore, by setting $\gamma = 48001$ we have that if $\langle g(f(e_i)), g(f(y_S))\rangle \leq 12000\epsilon$, then $i \notin S$, and $i \in S$ otherwise. We can recover each $g(f(y_S))$ for $y_S \in \mathcal{Y}^G$ from its binary representation. Next, we recover the set of indices of the points in $E \setminus \mathcal{E}_{y_S}^G$, from which we deduce the set of indices of the points $e_i \in \mathcal{E}_{y_S}^G$. This gives the information about the set $\{g(f(e_i))\}_{e_i \in \mathcal{E}_{y_S}^G}$. At this stage we have all the necessary information to compute the inner products $\langle g(f(y_S)), g(f(e_i))\rangle$ for all the pairs y_S and e_i that enable us to correctly decide whether $i \in S$. Finally, for the rest points $e \in E \setminus \mathcal{E}_{y_S}^G$ we have a naive encoding which explicitly states whether e is a part of y_S.

2.3   Deducing the lower bound

From the counting argument, the maximal number of different sets that can be recovered from the encoding of length at most $\rho = \frac{7}{8}d^2 + 2dL_G$ is at most 2^ρ. This implies $\frac{7}{8}d^2 + 2dL_G \geq \log|\mathcal{P}|$. On the other hand, the size of the family is $|\mathcal{P}| = \binom{d}{l}^d$. Recall that we have set $d = 2l$ so we have that $|\mathcal{P}| \geq \binom{2l}{l}^{2l} \geq \left(2^{(2l-1)}/\sqrt{l}\right)^{2l} \geq 2^{4l^2 - 2l\log l} \geq 2^{3.9l^2}$, where the last estimate

follows from our assumption on ϵ. Therefore, $\frac{7}{2}l^2 + 4lL_G \geq 3.9l^2$, implying $L_G \geq (1/10)l$, where $L_G = k\log(40/(\epsilon\sqrt{k})) = \frac{1}{2}k\log\left(16\left(\frac{10}{\epsilon}\right)^2\frac{1}{k}\right)$. This implies that $k\log\left(16\left(\frac{10}{\epsilon}\right)^2\frac{1}{k}\right) \geq (1/5)l \geq 1/(5\gamma^2 \cdot \epsilon^2)$. Setting $x = k \cdot (5\gamma^2 \cdot \epsilon^2)$ we have that

$$1 \leq x\log\left(\frac{0.5}{x} \cdot 2^{14}\gamma^2\right) = x\log(0.5/x) + x\log\left(2^{14}\gamma^2\right) \leq 1/2 + 2x(7 + \log\gamma),$$

where the last inequality we have used $x\log(0.5/x) \leq 0.5/(e\ln 2) < 1/2$ for all x. This implies the desired lower bound on the dimension: $k \geq 1/(20\gamma^2(7 + \log\gamma) \cdot \epsilon^2)$.

3   Lower bounds for q-moments of distortion

In this section we prove Theorem 5 which provides a lower bound for q-moments of distortion. Similarly, to the proof for ℓ_1-distortion in Section 2, we prove the theorem first for metric space of fixed size $\hat{n} = O(1/\epsilon^2) \cdot e^{O(\epsilon q)}$, which can be extended for metric spaces of size $\Theta(n)$ for any n via metric composition [9, 8], as described in the full version of the paper.

Assume w.l.o.g that $q \geq \frac{3}{\epsilon}$, otherwise the theorem follows from Theorem 4 by monotonicity of the ℓ_q-distortion. The proof strategy has exactly the same structure as in the proof of Section 2, however the sets I are constructed using different parameters. For a given $\epsilon < 0$, let $l = \lceil\frac{1}{\gamma^2\epsilon^2}\rceil$ be an integer for some large constant $\gamma > 1$ to be determined later. We construct point sets $I \subset \ell_2^d$, where $d = l\tau$, $\tau = e^{\epsilon q}$, and $|I| = 3d$. Assume that for all $I \in \mathcal{P}$ there is a map $f : I \to \ell_2^k$, with $\ell_q\text{-}dist(f) \leq 1 + \epsilon$. We show that this implies that $k = \Omega(q/\epsilon)$.

As before the strategy is to produce a unique binary encoding of I of length length(I). We will obtain that $|\mathcal{P}| = \binom{d}{l}^d \geq (d/l)^{ld}$, which will give that length$(I) \geq dl\log(d/l) = dl\log(\tau)$. We will show that this implies the bound on $k \geq \Omega(l\log(\tau)) = \Omega(1/\epsilon^2 \cdot \epsilon q) = \Omega(q/\epsilon)$.

As in the proof of Theorem 4, we can assume w.l.o.g. that $\epsilon \leq 1/\gamma$, which by the choice of γ later on implies $\epsilon < 1/36$.

▶ **Lemma 11.** *For any $I \in \mathcal{P}$ let $f : I \to \ell_2^k$ be an embedding with $\ell_q\text{-}dist(f) \leq 1 + \epsilon$, for $\epsilon < 1/36$. There is a subset $\hat{I} \subseteq I$ of size $|\hat{I}| \geq (1 - 3/\tau^4)|I|$ such that $f(\hat{I}) \subset B_2(1 + 6.02\epsilon)$, and for $1 - 2/\tau^4$ fraction of the pairs $(u, v) \in \binom{\hat{I}}{2}$ it holds that $|\langle f(u), f(v)\rangle - \langle u, v\rangle| \leq 32\epsilon$.*

Proof. By assumption we have $(\ell_q\text{-}dist(f))^q = \frac{1}{\left|\binom{I}{2}\right|}\sum_{(u,v)\in\binom{I}{2}}(dist_f(u,v))^q \leq (1 + \epsilon)^q$.

By the Markov inequality there are at least $1 - 1/\tau^4$ fraction of the pairs $(u, v) \in \left|\binom{I}{2}\right|$ such that $(dist_f(u,v))^q \leq \tau^4(1+\epsilon)^q \leq (1+\epsilon)^q \cdot e^{4\epsilon q}$, implying that $dist_f(u,v) \leq 1+6\epsilon$. Therefore, $|expans_f(u,v) - 1| \leq \max\{expans_f(u,v) - 1, 1/expans_f(u,v) - 1\} = dist_f(u,v) - 1 \leq 6\epsilon$.

For every $o_j \in O$, let F_j be the set of points $v \in I\backslash\{o_j\}$ such that $|expans_f(o_j, v) - 1| > 6\epsilon$. Then the total number of pairs $(u, v) \in \binom{I}{2}$ with the property that $|expans_f(u,v) - 1| > 6\epsilon$ is at least $\sum_{j=1}^d |F_j|/2$, implying that there must be a point $\hat{o} = o_{j*} \in O$ such that $|F_{j*}| \leq \frac{1}{\tau^4} \cdot \frac{3d(3d-1)}{d} \leq \frac{3}{\tau^4}(3d-1)$. Define $\hat{I} = I\backslash F_{j*}$ to be the complement of this set, so that $|\hat{I}| \leq (1 - \frac{3}{\tau^4})|I|$. We assume without loss of generality that $f(\hat{o}) = 0$. Let $\hat{O} = O \cap \hat{I}$. We have that $|expans_f(v, \hat{o}) - 1| = |\frac{\|f(v)\|_2}{\|v-\hat{o}\|_2} - 1| \leq 6\epsilon$, and using $\|v - \hat{o}\|_2 \leq \|v\|_2 + \|\hat{o}\|_2 \leq 1 + \epsilon/100$, so that $\|f(v)\|_2 \leq (1 + 6\epsilon)(1 + \epsilon/100) \leq 1 + 6.02\epsilon$, implying that $f(v) \in B_2(1 + 6.02\epsilon)$.

Denote by $\hat{G}$ the set of pairs $(u, v) \in \binom{\hat{I}}{2}$ satisfying that $|expans_f(u,v) - 1| \leq 6\epsilon$. To bound the fraction of these pairs from below, we can first bound $|\hat{I}| \geq (1 - \frac{3}{\tau^4})|I| \geq \frac{5}{2}d$ and $|\hat{I}| - 1 \geq 2d$, using that $\tau > 3$ by our assumption on q. Therefore, we have that the fraction of pairs $(u, v) \in \binom{\hat{I}}{2} \backslash \hat{G}$ is at most $\frac{1}{\tau^4} \cdot \frac{3d(3d-1)}{|\hat{I}|(|\hat{I}|-1)} \leq \frac{1}{\tau^4} \cdot \frac{9}{5} \leq \frac{2}{\tau^4}$.

Finally, to estimate the absolute difference in inner products over the set of pairs $\hat{G}$ we recall some of the estimates from the proof of Section 2. For all $(u,v) \in \hat{G}$ we have:

$$
\begin{aligned}
|\langle f(u), f(v)\rangle - \langle u, v\rangle| &\leq \frac{1}{2}\left[\left|\|f(u)\|_2^2 - \|u\|_2^2\right| + \left|\|f(v)\|_2^2 - \|v\|_2^2\right|\right] \\
&+ \frac{1}{2}\left[\left|\|f(u) - f(v)\|_2^2 - \|u - v\|_2^2\right|\right].
\end{aligned}
$$

We can bound each term as follows:

$$
\begin{aligned}
\left|\|f(u)\|_2^2 - \|u\|_2^2\right| &= \\
&= \left|\|f(u) - f(\hat{o})\|_2^2 - \|u - \hat{o}\|_2^2 + \|u - \hat{o}\|_2^2 - \|u\|_2^2\right| \\
&\leq \left|\|f(u) - f(\hat{o})\|_2 - \|u - \hat{o}\|_2\right| \cdot (\|f(u) - f(\hat{o})\|_2 + \|u - \hat{o}\|_2) \\
&+ \left|\|u - \hat{o}\|_2 - \|u\|_2\right| \cdot (\|u - \hat{o}\|_2 + \|u\|_2) \\
&\leq \|u - \hat{o}\|_2 \left|expans_f(u, \hat{o}) - 1\right| \cdot (\|f(u)\|_2 + \|u - \hat{o}\|_2) + \|\hat{o}\|_2 \cdot (\|u - \hat{o}\|_2 + \|u\|_2) \\
&\leq \left(1 + \frac{\epsilon}{100}\right)\left|expans_f(u, \hat{o}) - 1\right|\left(1 + 6.02\epsilon + 1 + \frac{\epsilon}{100}\right) + \frac{\epsilon}{100} \cdot \left(2 + \frac{\epsilon}{100}\right) \\
&\leq (2 + 6.06\epsilon)\left|expans_f(u, \hat{o}) - 1\right| + \frac{\epsilon}{40} \leq (2 + 6.06\epsilon) \cdot 6\epsilon + \frac{\epsilon}{40} \leq 14\epsilon,
\end{aligned}
$$

where we have used $\|\hat{o}\| \leq \epsilon/100$, $\|u - \hat{o}\|_2 \leq \|u\|_2 + \|\hat{o}\|_2 \leq 1 + \epsilon/100$, the bound on the norms of the embedding within $\hat{I}$, and the property of pairs in $\hat{G}$. Additionally, we have that

$$
\begin{aligned}
\left|\|f(u) - f(v)\|_2^2 - \|u - v\|_2^2\right| &= \\
&= \left|\|f(u) - f(v)\|_2 - \|u - v\|_2\right| \cdot (\|f(u) - f(v)\|_2 + \|u - v\|_2) \\
&\leq \|u - v\|_2 \left|expans_f(u, v) - 1\right| \cdot (\|f(u)\|_2 + \|f(v)\|_2 + \|u - v\|_2) \\
&\leq \sqrt{2}\left(2(1 + 6.02\epsilon) + \sqrt{2}\right)\left|expans_f(u, v) - 1\right| \leq 6\left|expans_f(u, v) - 1\right| \leq 36\epsilon,
\end{aligned}
$$

since $\|u - v\|_2 \leq diam(I) = \sqrt{2}$, and the last step follows using the property of pair in $\hat{G}$. We conclude that for all $(u,v) \in \hat{G}$: $|\langle f(u), f(v)\rangle - \langle u, v\rangle| \leq \frac{1}{2}(2 \cdot 14\epsilon + 36\epsilon) = 32\epsilon$.   ◀

As before, the goal is to encode the images of the embedding using a sufficiently small number of bits, by rounding them to the points of a grid of the Euclidean ball via the randomized rounding technique of [6] as to preserve the inner product gap. The following lemma provides the probability that this procedure fails.

▶ **Lemma 12.** *Let $X \subset \ell_2^k$ such that $X \subset B_2(r)$. For $\delta \leq r/\sqrt{k}$ let $G_\delta \subseteq B_2(r)$ denote the intersection of the δ-grid with $B_2(r)$. There is a mapping $g : X \to G_\delta$ such that for any $\eta \geq 1$, there is a $1 - 4e^{-\eta^2}$ fraction of the pairs $(u,v) \in \binom{X}{2}$ such that $|\langle g(u), g(v)\rangle - \langle u, v\rangle| \leq 3\sqrt{2}\eta\delta r$, and the points of the grid can be represented using $L_{G_\delta} = k \log(4r/(\delta\sqrt{k}))$ bits.*

Proof. For each point $v \in X$ randomly and independently match a point $\tilde{v}$ on the grid by rounding each of its coordinates v_i to one of the closest integral multiplies of δ in such a way that $E[\tilde{v}_i] = v_i$. For any $u \neq v \in X$ we have: $|\langle \tilde{u}, \tilde{v}\rangle - \langle u, v\rangle| \leq |\langle \tilde{u} - u, v\rangle| + |\langle \tilde{u}, \tilde{v} - v\rangle|$. Now, $E[\langle \tilde{u} - u, v\rangle] = \sum_{i=1}^k E[\tilde{u}_i - u_i]v_i = 0$ and $E[\langle \tilde{u}, \tilde{v} - v\rangle] = \sum_{i=1}^k E[\tilde{u}_i]E[\tilde{v}_i - v_i] = 0$. Next, we wish to make use of the Hoeffding bound. We therefore bound each of the terms $|(\tilde{u}_i - u_i)v_i| \leq \delta|v_i|$ and the sum $\sum_{i=1}^k \delta^2 v_i^2 = \delta^2 r$, and $|\tilde{u}_i(\tilde{v}_i - v_i)| \leq \delta(|u_i| + \delta)$, so that

$$
\sum_{i=1}^k \delta^2(v_i + \delta)^2 = \delta^2 \sum_{i=1}^k (v_i^2 + 2\delta v_i + \delta^2) \leq \delta^2(r + 2\delta\|v_i\|_1 + \delta^2 k) \leq \delta^2(r^2 + 2\delta r\sqrt{k} + \delta^2 k) \leq 4\delta^2 r^2.
$$

Applying the Hoeffding bound we get that $\Pr[|\langle \tilde{u} - u, v\rangle| > \sqrt{2}\eta\delta r] \leq 2e^{-\eta^2}$ and $\Pr[|\langle \tilde{u}, \tilde{v} - v\rangle| > 2\sqrt{2}\eta\delta r] \leq 2e^{-\eta^2}$, and therefore $\Pr[|\langle \tilde{u}, \tilde{v}\rangle - \langle u, v\rangle| > 3\sqrt{2}\eta\delta r] \leq 4e^{-\eta^2}$. This

probability also bounds the expected number of pairs with this property so there must exist an embedding to the grid where the bound stated in the lemma holds. The bound on the representation size is the same as in Lemma 8. ◀

Combining the lemmas we obtain:

▶ **Corollary 13.** *For any $I \in \mathcal{P}$ let $f : I \to \ell_2^k$ be an embedding with ℓ_q-dist$(f) \le 1 + \epsilon$, with $\epsilon \le 1/36$. There is a subset $\hat{I} \subseteq I$ of size $|\hat{I}| \ge (1 - 3/\tau^4)|I|$ such that for a fraction of at least $1 - 6/\tau^4$ of the pairs $(u, v) \in \binom{\hat{I}}{2}$ it holds that: $\left| \langle g(f(u)), g(f(v)) \rangle - \langle u, v \rangle \right| \le 42\epsilon$, where $g : \hat{I} \to G$. Moreover, the points in G can be uniquely represented by binary strings of length at most $L_G = k \log(5\sqrt{q/(\epsilon k)})$ bits.*

Proof. The corollary follows by applying Lemma 11 followed by Lemma 12 with $X = \hat{I}$ with $\delta = 2\sqrt{\epsilon/q}$ and $\eta = \sqrt{\ln(\tau)}$. Note that we may assume that indeed $2\sqrt{\epsilon/q} = \delta < 1/\sqrt{k} < r/\sqrt{k}$, since otherwise we are done. Therefore, the increase in the absolute difference of the inner products due to the grid embedding is at most: $3\sqrt{2}\eta\delta r = 6r\sqrt{2\ln(\tau)\epsilon/q} = 6r\sqrt{2(\epsilon q)\epsilon/q} \le 10\epsilon$. The bound on representation of the grid follows from Lemma 12: $L_G = k \log(4r/(\delta\sqrt{k})) = k \log(4r\sqrt{q/(\epsilon k)}) \le k \log(5\sqrt{q/(\epsilon k)})$. ◀

We are ready to obtain the main technical consequence which will imply the lower bound:

▶ **Corollary 14.** *For any $I \in \mathcal{P}$ let $f : I \to \ell_2^k$ be an embedding with ℓ_q-dist$(f) \le \epsilon$, with $\epsilon \le 1/36$. There is a subset of points G that satisfies the following: there is a subset $\mathcal{Y}^G \subseteq Y$ of size $|\mathcal{Y}^G| \ge (1 - 6/\tau^2)|Y|$ such that for each $y \in \mathcal{Y}^G$ there is a subset $\mathcal{E}_y^G \subseteq E$ of size $|\mathcal{E}_y^G| \ge (1 - 6/\tau^2)|E|$ such that for all pairs $(y, e) \in \mathcal{Y}^G \times \mathcal{E}_y^G$ we have: $\left| \langle g(f(y)), g(f(e)) \rangle - \langle y, e \rangle \right| \le 42\epsilon$, where $g : \mathcal{Y}^G \cup \{\mathcal{E}_y^G\}_{y \in \mathcal{Y}^G} \to G$. Moreover, the points in G can be uniquely represented by binary strings of length at most $L_G = k \log(5\sqrt{q/(\epsilon k)})$ bits.*

Proof. Applying Corollary 13 we have that there are at most $6/\tau^4$ pairs $(u, v) \in \binom{\hat{I}}{2}$ such that $\left| \langle g(f(u)), g(f(v)) \rangle - \langle u, v \rangle \right| > 42\epsilon$. It follows that the number of pairs in $Y \times E$ that are in $\binom{\hat{I}}{2}$ and have this property is at most $\frac{6}{\tau^4} \cdot \frac{3d(3d-1)}{2} \le \frac{27}{\tau^4} \cdot d^2$. Therefore there can be at most $\frac{3\sqrt{3}}{\tau^2} \cdot d$ points in $u \in Y$ such that there are more than $\frac{3\sqrt{3}}{\tau^2}d$ points in $v \in E$ with this property. Since there at most $\frac{3}{\tau^4} \cdot d < \frac{0.5}{\tau^2} \cdot d$ points in each of Y and E which may not be in $\hat{I}$ we obtain the stated bounds on the sizes of $|\mathcal{Y}^G|$ and $|\mathcal{E}_y^G|$. ◀

3.1 Encoding and decoding

For a set $I \in \mathcal{P}$ let $f : I \to \ell_2^k$ be a map with ℓ_q-dist$(f) = 1 + \epsilon$, where $\Omega\left(\frac{1}{\sqrt{n}}\right) \le \epsilon < 1/36$, and $q = O(\log(\epsilon^2 n)/\epsilon)$. Let $t = \tau^2/6$. Therefore, by Corollary 14, we can find a subset $G \subseteq B_2(2)$, and a mapping $g : f(I) \to G$, and a subset $\mathcal{Y}^G \subseteq Y$, with $|\mathcal{Y}^G| \ge \left(1 - \frac{1}{t}\right)|Y|$, where for all $y \in \mathcal{Y}^G$ we can find a susbet $\mathcal{E}_y^G \subseteq E$ with $|\mathcal{E}_y^G| \ge \left(1 - \frac{1}{t}\right)|E|$, such that for all pairs $(e, y) \in \mathcal{Y}^G \times \mathcal{E}_y^G$ the inner products $\left| \langle g(f(y)), g(f(e)) \rangle - \langle y, e \rangle \right| \le 42\epsilon$. Moreover, each point in G can be uniquely encoded using at most $L_G = k \log(5\sqrt{q/(\epsilon k)})$ bits.

The encoding is done according to the description in Section 2.1 so we similarly obtain the following bound on the bit length of the encoding: $(1/t)d^2(2 + \log(et)) + 2dL_G$.

The decoding works in the same way as before for an appropriate choice of $\gamma = 169$.

3.2   Deducing the lower bound

In this subsection we show that $k = \Omega(q/\epsilon)$, proving the desired lower bound for the case of $n = 3d = O(1/\epsilon^2) \cdot e^{O(\epsilon q)}$. From the counting argument, the maximal number of different sets that can be recovered from the encoding of length at most $\rho = (1/t)d^2(2 + \log(et)) + 2dL_G$ is at most 2^ρ. This implies $(1/t)d^2(2 + \log(et)) + 2dL_G \geq \log|\mathcal{P}|$. On the other hand, the size of the family is $|\mathcal{P}| = \binom{d}{l}^d \geq (d/l)^{ld} = \tau^{ld}$, so that $\log(|\mathcal{P}|) = ld\log(\tau)$. We therefore derive the following inequality

$$(1/t)d^2(2 + \log(et)) + 2dL_G \geq ld\log(\tau) \Rightarrow L_G \geq (1/4)l\log(\tau),$$

as $(1/t)d(2 + \log(et)) \leq d(2\log(\tau) + 4)/\tau^2 \leq d/(2\tau)\log(\tau) = l\log(\tau)/2$, using that $\log(\tau) > 4$.
Recall that $L_G = k\log(5\sqrt{q/(\epsilon k)}) = \frac{1}{2}k\log(25(q/(\epsilon k)))$. This implies that

$$k\log\left(25\left(\frac{q}{\epsilon k}\right)\right) \geq (1/2)l\log(\tau) \geq 1/(2\gamma^2 \cdot \epsilon^2) \cdot \epsilon q = 1/(2\gamma^2) \cdot q/\epsilon.$$

Setting $x = k \cdot (2\gamma^2 \cdot \epsilon/q)$ we have that

$$1 \leq x\log\left(\frac{0.5}{x} \cdot 100\gamma^2\right) = x\log(0.5/x) + x\log\left(100\gamma^2\right) \leq 1/2 + 2x\log(10\gamma),$$

where the last inequality we have used $x\log(0.5/x) \leq 0.5/(e\ln 2) < 1/2$ for all x. This implies the desired lower bound on the dimension: $k \geq 1/(8\gamma^2\log(10\gamma)) \cdot q/\epsilon$.

References

1   Ittai Abraham, Yair Bartal, T-H. Hubert Chan, Kedar Dhamdhere Dhamdhere, Anupam Gupta, Jon Kleinberg, Ofer Neiman, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '05, pages 83–100, USA, 2005. IEEE Computer Society. `doi:10.1109/SFCS.2005.51`.

2   Ittai Abraham, Yair Bartal, and Ofer Neiman. Embedding metrics into ultrametrics and graphs into spanning trees with constant average distortion. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, pages 502–511, USA, 2007. Society for Industrial and Applied Mathematics.

3   Ittai Abraham, Yair Bartal, and Ofer Neiman. Embedding metrics into ultrametrics and graphs into spanning trees with constant average distortion. In *Proceedings of the 18th annual ACM-SIAM symposium on Discrete algorithms*, SODA '07, pages 502–511, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics. URL: `http://portal.acm.org/citation.cfm?id=1283383.1283437`.

4   Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. *Advances in Mathematics*, 228(6):3026–3126, 2011. `doi:10.1016/j.aim.2011.08.003`.

5   Noga Alon. Perturbed identity matrices have high rank: Proof and applications. *Combinatorics, Probability & Computing*, 18(1-2):3–15, 2009. `doi:10.1017/S0963548307008917`.

6   Noga Alon and Bo'az Klartag. Optimal compression of approximate inner products and dimension reduction. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 639–650, 2017. `doi:10.1109/FOCS.2017.65`.

7   Ehsan Amid and Manfred K. Warmuth. Trimap: Large-scale dimensionality reduction using triplets, 2019. `arXiv:1910.00204`.

8   Yair Bartal, Nova Fandina, and Ofer Neiman. Dimensionality reduction: theoretical perspective on practical measures. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 10576–10588, 2019. URL: `https://proceedings.neurips.cc/paper/2019/hash/94f4ede62112b790c91d5e64fdb09cb8-Abstract.html`.

9   Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric ramsey-type phenomena. *Annals of Mathematics*, 162(2):643–709, 2005. URL: `http://www.jstor.org/stable/20159927`.

10 A. Censi and D. Scaramuzza. Calibration by correlation using metric embedding from nonmetric similarities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(10):2357–2370, October 2013. `doi:10.1109/TPAMI.2013.34`.

11 Leena Chennuru Vankadara and Ulrike von Luxburg. Measures of distortion for machine learning. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 4891–4900. Curran Associates, Inc., 2018.

12 Leena Chennuru Vankadara and Ulrike von Luxburg. Measures of distortion for machine learning. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL: `https://proceedings.neurips.cc/paper/2018/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf`.

13 Russ Cox, Frank Dabek, Frans Kaashoek, Jinyang Li, and Robert Morris. Practical, distributed network coordinates. *SIGCOMM Comput. Commun. Rev.*, 34(1):113–118, January 2004. `doi:10.1145/972374.972394`.

14 Michael Elkin, Arnold Filtser, and Ofer Neiman. Prioritized metric structures and embedding. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 489–498, New York, NY, USA, 2015. Association for Computing Machinery. `doi:10.1145/2746539.2746623`.

15 Michael Elkin, Arnold Filtser, and Ofer Neiman. Terminal Embeddings. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, volume 40 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 242–264, 2015.

16 Patrick J. F. Groenen, Rudolf Mathar, and Willem J. Heiser. The majorization approach to multidimensional scaling for minkowski distances. *Journal of Classification*, 12(1):3–19, 1995.

17 W. J Heiser. Multidimensional scaling with least absolute residuals. In *In H. H. Bock (Ed.) Classification and related methods*, pages 455–462. Amsterdam: NorthHolland, 1988a.

18 P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, FOCS '01, page 10, USA, 2001. IEEE Computer Society.

19 Piotr Indyk and Jiri Matousek. Low-distortion embeddings of finite metric spaces. In *in Handbook of Discrete and Computational Geometry*, pages 177–196. CRC Press, 2004.

20 Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, pages 604–613, New York, NY, USA, 1998. ACM. `doi:10.1145/276698.276876`.

21 William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, pages 189–206. American Mathematical Society, Providence, RI, 1984.

22 Jon Kleinberg, Aleksandrs Slivkins, and Tom Wexler. Triangulation and embedding using small sets of beacons. *J. ACM*, 56(6):32:1–32:37, September 2009. `doi:10.1145/1568318.1568322`.

23 Deepanshu Kush, Aleksandar Nikolov, and Haohua Tang. Near neighbor search via efficient average distortion embeddings. In *37th International Symposium on Computational Geometry, SoCG 2021, June 7-11, 2021, Buffalo, NY, USA (Virtual Conference)*, pages 50:1–50:14, 2021. `doi:10.4230/LIPIcs.SoCG.2021.50`.

24 Kasper Green Larsen and Jelani Nelson. Optimality of the johnson-lindenstrauss lemma. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 633–638, 2017. `doi:10.1109/FOCS.2017.64`.

25 N. Linial. Finite metric spaces- combinatorics, geometry and algorithms. In *Proceedings of the ICM*, 2002.

26 Jiří Matoušek. Bi-Lipschitz embeddings into low-dimensional Euclidean spaces. *Commentat. Math. Univ. Carol.*, 31(3):589–600, 1990.

27 Jiří Matoušek. *Lectures on Discrete Geometry*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002.

28 Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018. `doi:10.21105/joss.00861`.

29 Assaf Naor. Comparison of metric spectral gaps. *Analysis and Geometry in Metric Spaces*, 2:2:1–52, 2014.

30 Assaf Naor. An average John theorem. *Geometry and Topology*, 25(4):1631–1717, 2021. `doi:10.2140/gt.2021.25.1631`.

31 Puneet Sharma, Zhichen Xu, Sujata Banerjee, and Sung-Ju Lee. Estimating network proximity and latency. *Computer Communication Review*, 36(3):39–50, 2006. `doi:10.1145/1140086.1140092`.

32 Yuval Shavitt and Tomer Tankel. Big-bang simulation for embedding network distances in euclidean space. *IEEE/ACM Trans. Netw.*, 12(6):993–1006, December 2004. `doi:10.1109/TNET.2004.838597`.

33 Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. URL: `http://jmlr.org/papers/v9/vandermaaten08a.html`.

34 Santosh Srinivas Vempala. *The random projection method*, volume 65 of *DIMACS series in discrete mathematics and theoretical computer science*. Providence, R.I. American Mathematical Society, 2004. URL: `http://opac.inria.fr/record=b1101689`.

35 J. Fernando Vera, Willem J. Heiser, and Alex Murillo. Global optimization in any minkowski metric: A permutation-translation simulated annealing algorithm for multidimensional scaling. *J. Classif.*, 24(2):277–301, September 2007.

36 Yingfan Wang, Haiyang Huang, Cynthia Rudin, and Yaron Shaposhnik. Understanding how dimension reduction tools work: An empirical approach to deciphering t-sne, umap, trimap, and pacmap for data visualization, 2020. `arXiv:2012.04456`.