Optimality of the Johnson-Lindenstrauss Dimensionality Reduction for Practical Measures

Yair Bartal [*] Ora Nova Fandina [†] Kasper Green Larsen [‡]

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 [BFN19] (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 $1 \leq q \leq O(\frac{\log(2\epsilon^2 n)}{\epsilon})$ and $\epsilon \geq \frac{1}{\sqrt{n}}$, 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 $\epsilon \geq \frac{1}{\sqrt{n}}$, matching the upper bounds of [BFN19] 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.

[*] Hebrew University. Supported in part by a grant from the Israeli Science Foundation (1817/17). yair@cs.huji.ac.il
[†] Aarhus University. fandina@cs.au.dk
[‡] Aarhus University. larsen@cs.au.dk

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 [Ind01a, Lin02, IM04] for surveys). In addition to the beautiful theory, a plethora of original and elegant techniques have been developed and successfully applied in various fields of algorithmic research, e.g., clustering, nearest-neighbor, distance oracle. See [Mat02, Ind01b, Vem04] 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 [JL84]. 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 [LN17] and in [AK17] (improving upon [Alo09]). On the other hand, it was shown [Mat90] that there are Euclidean pointsets for which any embedding into k-dimensions must have $n^{\Omega(k)}$ 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, within such fields of computational research as biology, neuroscience, medicine, psychology, meteorology, archeology and musicology, among others. Practitioners have frequently employed classic tools of metric embedding theory, 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. Following are some of the most recent techniques [vdMH08, MHSG18, AW19, WHRS20]. However, most of these heuristics have not been rigorously analyzed in terms of absolute bounds. Recent papers [CVvL18a] and [BFN19] initiate the formal study of practically oriented analysis of metric embedding.

In contrast to the worst case-analysis, the quality of practically motivated embedding is determined by the average performance over all pairs, where an error per pair is measured as an additive error, a multiplicative error or a combination of both. In [BFN19], authors analyzed dimensionality reduction for various basic and widely used in practice average distortion measures.

Dimensionality reduction for average distortions Following are some of the most basic average quality measurement criteria, which are commonly used by practitioners to infer embedding's quality.

Definition 1.1 (ℓ_q-distortion). *Let $expans_f(u, v) = d_Y(f(u), f(v))/d_X(u, v)$, $contract_f(u, v) = 1/expans_f(u, v)$, and $dist_f(u, v) = \max\{expans_f(u, v), contract_f(u, v)\}$. The q-th moment of distortion is defined by*

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

for any $q \geq 1$ and an n point X.

Definition 1.2 (Energy)**.** *Let $d_{uv} := d_X(u,v)$ and $\hat{d}_{uv} := d_Y(f(u), f(v))$. For any $q \geq 1$,*

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

There are several additional additive measures that are commonly used in practice such as *stress*, and *relative error measure*. These are defined and further discussed and analyzed in Section 2 and Appendix B.

In [BFN19] the authors raised the question: *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 [IM98] shows that any Euclidean subset can be embedded with $1+\epsilon$ average distortion 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 $q = 1$ (the case of the *average*) unresolved. For all the additive measures they prove that a bound of ϵ can be achieved in dimension $k = O(q/\epsilon^2)$.

In this paper, we close the gap proving that indeed the Johnson-Lindenstrauss bounds are best possible for any dimensionality reduction for the full range of $q \geq 1$.

Our results effectively show that the JL dimensionality reduction is optimal for all average distortion measures, for all values of q-th moment. 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 Contribution In this paper we prove the following lower bounds:

Theorem 1.1. *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 average distortion at most $1+\epsilon$ requires $k = \Omega(1/\epsilon^2)$.*

For the more general case of large values of q, we show

Theorem 1.2. *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 the following theorem:

Theorem 1.3. *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 $k = \Omega(1/\epsilon^2)$.*

In fact, our main proof is of the lower bound for $Energy_1$ quality measure, which we show to imply the bound in Theorem 1.1 and for all measures in Theorem 1.3, with some small modifications for the stress measures. Furthermore, since all additive measures are also monotonically increasing with q the bounds hold for all $q \geq 1$. Therefore Theorems 1.1 and 1.2 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 [BFN19] for $q \geq 2$, Theorem 1.3 provides a tight bound of $\Omega(q/\epsilon^2)$ for all additive measures.

Our Techniques. Our proof of the lower bound is based on the counting argument, as in [LN17]. Essentially, we extend the framework developed in [LN17] for the worst case distortion to work in regime of the average distortion. Particularly, as in [LN17], 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. We then show that there is not enough space to accommodate all the different metric spaces from the family if the target dimension is too low, using the random discretization argument of [AK17]. The main difficulty is showing that the weaker guaranties on the *average* or *q-moments* distortions, rather than on the worst case, still suffice to maintain the necessary conditions for obtaining the desired bounds.

Related work. Beyond the worst-case distortion analysis of metric embedding was suggested in [KSW09], where partial and scaling distortions were defined. Other refined notions have been considered as well, including Ramsey type embeddings [BLMN05] and prioritized distortion [EFN15], among others.

The study of average distance and ℓ_q-distortion was initiated in [ABN11] where bounds were given for embedding arbitrary metrics in to Euclidean space and other normed spaces.

Average additive distortion measures are commonly used and explored in applied community [Hei8a, GMH95, ST04, CDK$^+$04, SXBL06, VHM07, CS13].

Finally, [BFN19] have initiated the first rigorous theoretical study providing analysis of the practically motivated average distortion measures, which are further discussed in this paper. 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. They have also shown that for $q = \Omega(\log(1/\epsilon)/\epsilon)$ the tight bound of $\Omega(q/\epsilon)$ follows from reduction to the worst case.

2 Preliminaries

For metric spaces (X, d_X), (Y, d_Y) and an embedding $f : X \to Y$, for a pair of points $u \neq v \in X$, $expans_f(u, v) = d_Y(f(u), f(v))/d_X(u, v)$, $contr_f(u, v) = d_X(u, v)/d_Y(f(u), f(v))$. Distortion of f, denoted $dist(f)$, is defined as $dist(f) = \sup_{u \neq v \in X}\{expans_f(u, v)\} \cdot \sup_{u \neq v \in X}\{contr_f(u, v)\}$. Distortion of a pair $u \neq v \in X$ is defined by $dist_f(u, v) = \max\{expans_f(u, v), contract_f(u, v)\}$.

Additive Distortions In [BFN19], authors considered a set of average distortion measures that are very common in applied research. We define here the relative error measure, which is closely related to ℓ_q-distortion. The rest of the measures are defined and analyzed in Appendix B. For an embedding $f : X \to Y$, for a pair $u \neq v \in X$ denote $d_{u,v} := d_X(u,v)$ and $\hat{d}_{u,v}$.

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

Following are some basic relations between some of the distortion measures observed in [BFN19] which we will use in the paper:

Claim 2.1 (ℓ_q-distortion, $Energy_q$ and REM_q)**.** *For all* $1 \leq q \leq \infty$ *the following relations hold:*

$$\ell_q\text{-}dist(f) - 1 \geq REM_q(f) \geq Energy_q(f).$$

Proof. The relation between Energy and REM is obvious by definition. Additionally,

$$(REM_q(f))^q + 1 = \mathrm{E}\left[(|dist_f(u,v) - 1|)^q + 1\right] \leq \mathrm{E}\left[(dist_f(u,v))^q\right] = \ell_q\text{-}dist(f),$$

using the inequality: $x^q + 1 \leq (x+1)^q \leq 2^q(x^q + 1)$ for $x \geq 0$, applied for $x = dist_f(u,v) - 1$. □

3 Lower Bound for Average Distortion and Additive Measures

In this section we prove Theorems 1.1 and Theorem 1.3. Using Claim 2.1, we may focus on proving the lower bound for $Energy_1(f)$ in order to obtain similar lower bounds for $REM_1(f)$ and $\ell_1\text{-}dist(f)$. In Appendix B, using relations between the additive measures, we conclude that the lower bound for all the additive measures follow from the lower bound on Energy and on Stress.

We first present the proof for $\hat{n} = \Theta(1/\epsilon^2)$ and later show how to extend it for any $n \geq \hat{n}$. We construct a large family $\mathcal{P}$ of metric spaces, such that each $\mathrm{I} \in \mathcal{P}$ can be completely recovered by computing the inner products between the points in I. For a given $\epsilon < 0$, let $l = \lceil \frac{1}{\gamma^2 \epsilon^2} \rceil$ be an integer for some large constant $\gamma > 1$ to be determined later. We will construct point sets $\mathrm{I} \subset \ell_2^d$, where $d = 2l$.

Let $O = \{o_j\}_{j=1}^d$ denote the set of near zero vectors in ℓ_2^d, i.e., all $\|o_j\|_2 \leq \epsilon/6$. We assume that $o_1 = 0$. 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 $\mathrm{Y}[S_1, \ldots, S_d] = \{y_{S_1}, \ldots, y_{S_d}\}$. Each point set $\mathrm{I}[S_1, \ldots, S_d] \in \mathcal{P}$ is defined as the union of the sets defined above, i.e., $\mathrm{I}[S_1, \ldots, S_d] = O \cup E \cup \mathrm{Y}[S_1, \ldots, S_d]$[1]. The size of the family is $|\mathcal{P}| = \binom{d}{l}^d$. Note that for each $\mathrm{I} \in \mathcal{P}$, $|I| = 3d = 6l$, I is contained in the unit ball of ℓ_2: $I \in B_2(1)$, and its diameter is bounded by $diam(I) = \sqrt{2}$. 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 $e \in \mathrm{span}\{e_i | i \in S\}$. In particular, if $\langle e_j, y_S \rangle = 0$ then $j \notin S$, and if $\langle e_j, y_S \rangle = 1/\sqrt{l} = \gamma\epsilon$ then $j \in S$.

To obtain the lower bound, we assume that for all $\mathrm{I} \in \mathcal{P}$ 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 I of length $\mathrm{length(I)}$. Clearly, the length of the encoding is a function of the target dimension k (in addition to ϵ, and the size of I). The encoding is essentially an injective map from $\mathcal{P}$ to $\{0,1\}^{\mathrm{length(I)}}$. Therefore, the number of different sets that can be recovered by decoding is at most $2^{\mathrm{length(I)}}$. Now, because $|\mathcal{P}| = \binom{d}{l}^d \geq \left(2^{(2l-1)}/\sqrt{l}\right)^{2l}$ we get that $\mathrm{length(I)} \geq 3.9 \cdot l^2$. We will show that this implies the bound on $k \geq \Omega(l)$.

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

Let I $\in \mathcal{P}$ be a metric space, and let $f : I \to \ell_2^k$ be an embedding with $Energy_1(f) \leq \epsilon$. In the following lemma we show that there exists a large subset of I which is contained in a ball of constant radius for which the average absolute difference between the inner products of the images of the embedding and the original space is bounded by $O(\epsilon)$ additive error.

Lemma 3.1. *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 $\left|\hat{I}\right| \geq (1-\alpha)|I|$ such that $f(\hat{I}) \subset B_2\left(1 + \frac{3.01\epsilon}{\alpha}\right)$, and*

$$\frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v)\in\binom{\hat{I}}{2}} \left|\langle f(u), f(v)\rangle - \langle u, v\rangle\right| \leq (20 + \frac{1}{\alpha})\epsilon.$$

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

$$Energy_1(f) = \frac{1}{|\binom{I}{2}|} \sum_{(u,v)\in\binom{I}{2}} \left|expans_f(u,v) - 1\right| \leq \epsilon. \tag{3.0.1}$$

This bound implies that

$$\frac{1}{|O||I|}\left(\sum_{o_j\in O}\sum_{v\in I, v\neq o_j} \left|expans_f(o_j,v) - 1\right|\right) \leq \frac{1}{|O||I|}\sum_{u\neq v\in I}\left|expans_f(u,v) - 1\right| \leq \frac{3d(3d-1)}{d(3d-1)}\epsilon = 3\epsilon. \tag{3.0.2}$$

From (3.0.2) it follows that

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

Let $\hat{o} \in O$ denote the point at which the minimum is obtained. We assume without loss of generality that $f(\hat{o}) = 0$. By Markov's inequality, there is a subset I_0 of size $|\hat{I}| \geq (1-\alpha)\left|I\right|$ such that for all $v \in \hat{I}$, $\left|expans_f(\hat{o},v) - 1\right| \leq \frac{3\epsilon}{\alpha}$. Let $\hat{O} = O \cap \hat{I}$. We have that $|expans_f(v,\hat{o}) - 1| = |\frac{\|f(v)\|_2}{\|v-\hat{o}\|_2} - 1| \leq \frac{3\epsilon}{\alpha}$, and using $\|v - \hat{o}\|_2 \leq \|v\|_2 + \|\hat{o}\|_2 \leq 1 + \epsilon/6$, so that $\|f(v)\|_2 \leq (1 + \frac{3\epsilon}{\alpha})(1 + \epsilon/6) \leq 1 + \frac{3.01\epsilon}{\alpha}$, implying that $f(v) \in B_2\left(1 + \frac{3.01\epsilon}{\alpha}\right)$.

For all $(u,v) \in \binom{I}{2}$ and $1 \leq j \leq d$ we have:

$$\left|\langle f(u), f(v)\rangle - \langle u, v\rangle\right| \leq \frac{1}{2}\left[\left|\|f(u)\|_2^2 - \|u\|_2^2\right| + \left|\|f(v)\|_2^2 - \|v\|_2^2\right| + \left|\|f(u) - f(v)\|_2^2 - \|u - v\|_2^2\right|\right] \tag{3.0.3}$$

$$= \frac{1}{2}\left[|\|f(u) - f(o_j)\|_2^2 - \|u - 0\|_2^2| + |\|f(v) - f(o_j)\|_2^2 - \|v - 0\|_2^2| + \left|\|f(u) - f(v)\|_2^2 - \|u - v\|_2^2\right|\right]$$

$$\leq \frac{1}{2}\left[(|(expans_f(u,o_j))^2 - 1| + \epsilon/2) + (|(expans_f(v,o_j))^2 - 1| + \epsilon/2) + \|u - v\|_2^2\,|(expans_f(u,v))^2 - 1|\right]$$

$$\leq \frac{1}{2}|(expans_f(u,o_j))^2 - 1| + \frac{1}{2}|(expans_f(v,o_j))^2 - 1| + |(expans_f(u,v))^2 - 1| + \epsilon/2,$$

where the second inequality is due to $|\,\|x - 0\|_2 - \|x - o_j\|\,| \leq \|o_j\| \leq \epsilon/6$, for all $x \in I$ and $o_j \in O$, and the last inequality holds since $\|u - v\| \leq diam(I) = \sqrt{2}$.

Therefore:

$$\sum_{(u,v)\in\binom{\hat{I}}{2}} \left|\langle f(u), f(v)\rangle - \langle u,v\rangle\right| \leq \tag{3.0.4}$$

$$\leq \sum_{(u,v)\in\binom{\hat{I}}{2}} \left[\frac{1}{2}|(expans_f(u,o_j))^2-1| + \frac{1}{2}|(expans_f(v,o_j))^2-1| + |(expans_f(u,v))^2-1| + \epsilon/2\right].$$

We have that

$$\begin{aligned}
\sum_{(u,v)\in\binom{\hat{I}}{2}} |(expans_f(u,o_j))^2-1| &= \frac{|\hat{I}|(|\hat{I}|-1)/2}{|\hat{O}|(|\hat{I}|-1)} \sum_{o_j\in\hat{O}, u\neq o_j\in\hat{I}} |(expans_f(u,o_j))^2-1| \\
&\leq \frac{3d/2}{(1-3\alpha)d} \sum_{o_j\in\hat{O}, u\neq o_j\in\hat{I}} |(expans_f(u,o_j))^2-1| \leq 2 \sum_{o_j\in\hat{O}, u\neq o_j\in\hat{I}} |(expans_f(u,o_j))^2-1| \\
&\leq 2\sum_{u\neq v\in\hat{I}} |(expans_f(u,v))^2-1| = 4 \sum_{(u,v)\in\binom{\hat{I}}{2}} |(expans_f(u,v))^2-1|,
\end{aligned}$$

where we have used $|\hat{O}| \geq |O| - \alpha|\hat{I}| \geq d - 3\alpha d = (1-3\alpha)d$, and $\alpha \leq 1/16$. Therefore (3.0.4) yields that

$$\frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v)\in\binom{\hat{I}}{2}} \left|\langle f(u), f(v)\rangle - \langle u,v\rangle\right| \leq \frac{5}{|\binom{\hat{I}}{2}|} \sum_{(u,v)\in\binom{\hat{I}}{2}} |(expans_f(u,v))^2-1| + \epsilon/2.$$

Now we have that for all $u, v \in \hat{I}$:

$$\begin{aligned}
|(expans_f(u,v))^2-1| &= |(expans_f(u,v))-1||(expans_f(u,v))+1| \\
&= |(expans_f(u,v))-1|(1+\frac{\|f(u)-f(v)\|}{\|u-v\|}) \\
&\leq |(expans_f(u,v))-1|(1+\frac{\|f(u)\|+\|f(v)\|}{\|u-v\|}) \\
&\leq |(expans_f(u,v))-1|(1+2(1+\frac{3\epsilon}{\alpha}+\epsilon/9) \\
&= |(expans_f(u,v))-1|(3+1/27+\frac{1}{6\alpha}).
\end{aligned}$$

Therefore

$$\frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v)\in\binom{\hat{I}}{2}} \left|\langle f(u), f(v)\rangle - \langle u,v\rangle\right| \leq (16+\frac{5}{6\alpha})\frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v)\in\binom{\hat{I}}{2}} |(expans_f(u,v))-1| + \epsilon/2.$$

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 $|I'| \geq (1-\alpha)|I|$, so that $\alpha \leq 1/16$ it holds that $|\binom{\hat{I}}{2}| \geq (1 - \frac{1}{3(1-\alpha)d})(1-\alpha)^2 \cdot |\binom{I}{2}| \geq \frac{5}{6}|\binom{I}{2}|$, and applying (3.0.1). Finally, we obtain

$$\frac{1}{|\binom{\hat{I}}{2}|} \sum_{(u,v)\in\binom{\hat{I}}{2}} \big|\langle f(u), f(v)\rangle - \langle u,v\rangle\big| \leq \frac{6}{5}(16 + \frac{5}{6\alpha})\epsilon + \epsilon/2 \leq (20 + \frac{1}{\alpha})\epsilon,$$

□

We will later show that an upper bound on the average absolute difference of inner products implies that there is a large subset of the set Y of the instance I that can be roughly recovered from the images $f(e_i)$ and $f(y_{S_j})$ by computing the inner products $\langle f(e_i), f(y_{S_j})\rangle$ and observing the gap. The goal is then to encode the images using a sufficiently small number of bits. For that end we carefully discretize the Euclidean ball containing the images of the points that preserve the inner product gap. For this aim we apply a technique of randomized rounding to the points on a grid, proposed in [AK17]. The following lemma shows that the average absolute difference of inner products is preserved under this embedding to the grid, which may be of independent interest.

Lemma 3.2. *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 $g : X \to G_\delta$ such that*

$$\frac{1}{|\binom{X}{2}|} \sum_{(u,v)\in\binom{X}{2}} |\langle g(u), g(v)\rangle - \langle u,v\rangle| \leq 3\delta r.$$

The points of the grid can be represented using a number of bits bounded by: $L_{G_\delta} = 2k + \log(r/(\delta\sqrt{k}))$.

Proof. We randomly round the points in X to the δ-grid of $B_2(r)$, as proposed in [AK17], and show that the expected value of the sum is bounded, which implies the lemma. For each point $v \in X$ we randomly and independently match a point $\tilde{v}$ 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 $E[\tilde{v}_i] = v_i$. In particular, this distribution is given by assigning the value $\lceil \frac{v_i}{\delta} \rceil \delta$ with probability $p = \left(\frac{v_i}{\delta} - \lfloor \frac{v_i}{\delta} \rfloor\right)$, and assigning the value $\lfloor \frac{v_i}{\delta} \rfloor \delta$ with probability $1-p$. For any $u \neq v \in X$ we have:

$$\mathrm{E}\left[|\langle \tilde{u},\tilde{v}\rangle - \langle u,v\rangle|\right] \leq \mathrm{E}\left[|\langle \tilde{u}-u, v\rangle|\right] + \mathrm{E}\left[|\langle \tilde{u}, \tilde{v}-v\rangle|\right] \leq \left(\mathrm{E}\left[(\langle \tilde{u}-u, v\rangle)^2\right]\right)^{1/2} + \left(\mathrm{E}\left[(\langle \tilde{u}, \tilde{v}-v\rangle)^2\right]\right)^{1/2},$$

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

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

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

$$\begin{aligned}\mathrm{E}\left[(\langle \tilde{u}, \tilde{v}-v\rangle)^2\right] = \mathrm{E}\left[\left(\sum_{i=1}^{k} \tilde{u}_i(\tilde{v}_i - v_i)\right)^2\right] &\leq \sum_{i=1}^{k} \mathrm{E}\left[\tilde{u}_i^2\right] \mathrm{E}\left[(\tilde{v}_i - v_i)^2\right] + 2\sum_{1\leq i\neq j\leq k} \mathrm{E}[\tilde{u}_i\tilde{u}_j(\tilde{v}_i - v_i)]\mathrm{E}[\tilde{v}_j - v_j] \\ &\leq \delta^2 \sum_{i=1}^{k} \mathrm{E}[\tilde{u}_i^2],\end{aligned}$$

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

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

Therefore, putting all together,

$$\mathrm{E}\left[|\langle \tilde{u}, \tilde{v}\rangle - \langle u, v\rangle|\right] \le \delta r + \delta(r^2 + \delta^2 k)^{1/2} \le 2\delta r + \delta^2\sqrt{k} \le 3\delta r.$$

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 [AK17]: The ith coordinated of point x on the grid are given by a sign and absolute value $n_i\delta$, where $0 \le n_i \le r/\delta$ are integers satisfying $\sum_{1\le i\le k} n_i^2 \le (r/\delta)^2$. 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 3.1. *For any $I \in \mathcal{P}$ let $f : I \to \ell_2^k$ be an embedding with $Energy_1(f) \le \epsilon$, with $\epsilon \le 1/36$. Let $0 < \alpha \le 1/16$ and $0 < \beta$ be a parameters. There is a subset $\hat{I} \subseteq I$ of size $|\hat{I}| \ge (1-\alpha)|I|$ that satisfies the following:*

$$\sum_{(u,v)\in\binom{\hat{I}}{2}} \left|\langle g(f(u)), g(f(v))\rangle - \langle u, v\rangle\right| \le \left(23 + \frac{1.26}{\alpha}\right)\epsilon,$$

where $g : \hat{I} \to G$. Moreover, the points in G can be uniquely represented by binary strings of length at most $L_G = k\log(4r/(\epsilon\sqrt{k}))$ bits.

Proof. The corollary follows by applying Lemma 3.1 followed by Lemma 3.2 with $X = \hat{I}$, with $\delta = \epsilon$. Note that we may assume that indeed $\epsilon = \delta < 1/\sqrt{k} < r/\sqrt{k}$, since otherwise we are done. □

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

Corollary 3.2. *For any $I \in \mathcal{P}$ let $f : I \to \ell_2^k$ be an embedding with $Energy_1(f) \le \epsilon$, with $\epsilon \le 1/36$. Let $0 < \alpha \le 1/16$ 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 $|\mathcal{Y}^G| \ge (1 - 3\alpha - \frac{3}{\sqrt{2}}\beta)|Y|$ such that for each $y \in \mathcal{Y}_G$ there is a subset $\mathcal{E}_y^G \subseteq E$ of size $|\mathcal{E}_y^G| \ge (1 - 3\alpha - \frac{3}{\sqrt{2}}\beta)|E|$ such that for all pairs $(y, e) \in \mathcal{Y}^G \times \mathcal{E}_y^G$ we have:*

$$\left|\langle g(f(y)), g(f(e))\rangle - \langle y, e\rangle\right| \le \frac{1}{\beta^2}\left(23 + \frac{1.26}{\alpha}\right)\epsilon,$$

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.

Proof. Applying Corollary 3.1 and Markov's inequality we have that there are at most β^2 pairs $(u,v) \in \binom{\hat{I}}{2}$ such that $\left|\langle g(f(u)), g(f(v))\rangle - \langle u, v\rangle\right| > \frac{1}{\beta^2}\left(23 + \frac{1.26}{\alpha}\right)\epsilon$. 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 $\beta^2 \cdot \frac{3d(3d-1)}{2} \le \frac{9}{2}\beta^2 \cdot d^2$. 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 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 $|\mathcal{E}_y^G|$. □

This corollary suggests that the gap between the inner products $\langle e_i, y_S \rangle$ is preserved up to a small additive error between the images of the embedding rounded to the points of the grid, for the large fraction of these points in the instance. Therefore, it should be the case that encoding the points on the grid is enough to uniquely recover the large part of the original instance I. In the following we show that the whole instance I can be encoded using sufficiently small number of bits.

3.1 Encoding Algorithm

For an instance $\mathrm{I} \in \mathcal{P}$ let $f : \mathrm{I} \to \ell_2^k$ be an embedding with $Energy_1(f) \le \epsilon$, where $\Omega\left(\frac{1}{\sqrt{n}}\right) \le \epsilon < 1/36$. Let $t = 8$. We set $\alpha = 1/(12t)$, $\beta = 1/(\sqrt{2}t)$. Therefore, by Corollary 3.2, we can find a subset $G \subseteq B_2(9)$, 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 $\left|\langle g(f(y)), g(f(e))\rangle - \langle y, e\rangle\right| \le 20000\epsilon$. Moreover, each point in G can be uniquely encoded using at most $L_G = k\log(4r/(\epsilon\sqrt{k}))$ 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 each of length L_G bits, where each point $e_i \in \{\mathcal{E}_y^G\}_{y \in \mathcal{Y}^G}$ is encoded by the index of $g(f(e_i))$ and rest of points are encoded by zeros.

Now we can encode the points in $\mathcal{Y}^G$. For each $y_S \in \mathcal{Y}^G$ we write the index of the grid point $g(f(y_S))$, using L_G bits and the encoding of the set of indices of the points in $E \setminus \mathcal{E}_{y_S}^G$, using at most $\log\binom{d}{(1/t)d} \le (1/t)d\log(et)$. In addition, for each $e_i \in E \setminus \mathcal{E}_{y_S}^G$, we explicitly write down whether $i \in S$ 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 $\mathrm{I} \in \mathcal{P}$ can be encoded using at most

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

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

3.2 Decoding Algorithm

To recover the instance I from the encoding it is enough to recover the vectors Y, since the set of points O and E is the same in each I. We first recover the set $Y \setminus \mathcal{Y}^G$ in a straightforward way from its naive encoding.

To recover a point $y_{S_j} \in \mathcal{Y}^G$ we need to know for each $e_i \in E$ whether $i \in S_j$. An important implication of Corollary 3.2 is that given $g(f(e_i))$ and $g(f(y_{S_j}))$ of any pair $(e_i, y_{S_j}) \in \mathcal{Y}^G \times \mathcal{E}_{y_{S_j}}^G$, we can decide whether $i \in S_j$ by computing $\langle g(f(e_i)), g(f(y_{S_j}))\rangle$. Recall that if $i \notin S_j$ then $\langle e_i, y_{S_j}\rangle = 0$, and if $i \in S_j$ then $\langle e_i, y_{S_j}\rangle = \gamma\epsilon$. Therefore, by setting $\gamma = 20001$ we have that if $\langle g(f(e_i)), g(f(y_{S_j}))\rangle \le 20000\epsilon$, then $i \notin S_j$, and $i \in S_j$ otherwise. We can recover each $y_{S_j} \in \mathcal{Y}^G$ from its binary representation, and we can recover the set of indices of the points in the set $E \setminus \mathcal{E}_{y_{S_j}}^G$ from which we can deduce the set of indices of the points $e_i \in \mathcal{E}_{y_{S_j}}^G$. Afterward, we can look up the binary representations of $\{g(e_i)\}_{e_i \in \mathcal{E}_{y_{S_j}}^G}$ and recover them, which will enable to compute the inner products and deduce the answer. Finally, for each point $e \in E \setminus \mathcal{E}_{y_{s_j}}^G$ we have a naive encoding which explicitly says whether e is a part of y_{S_j}.

3.3 Deducing The Lower Bound

In this subsection we show that $k = \Omega(1/\epsilon^2)$, proving the desired lower bound for the case of $n = 6l$. 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 following from our assumption on ϵ. We therefore derive the following inequality

$$\frac{7}{2}l^2 + 4lL_G \geq 3.9l^2 \Rightarrow L_G \geq (1/10)l,$$

where $L_G = k\log(4r/(\epsilon\sqrt{k})) = \frac{1}{2}k\log\left(16(\frac{9}{\epsilon})^2\frac{1}{k}\right)$. This implies that

$$k\log\left(16\left(\frac{9}{\epsilon}\right)^2\frac{1}{k}\right) \geq (1/5)l \geq 1/(5\gamma^2 \cdot \epsilon^2).$$

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:

$$k \geq 1/(20\gamma^2(7 + \log\gamma) \cdot \epsilon^2).$$

3.4 Metric Spaces of Arbitrary Size

In order to extend the lower bound for the input metrics of an arbitrary size $n \geq \hat{n} = \Theta(1/\epsilon^2)$, we use the notion of the metric composition proposed in [BLMN05]. Given a metric space X, we compose it with another metric space Y of size $n/|X|$ by substituting each point in X with a copy of Y. The first observation is that in such composition pairs of the points from different copies constitute a constant fraction of all the points in the space. The second observation is that, loosely speaking, the average error over these pairs is, up to a constant, the average of average errors over different "copies" of X in the composition.

The following lemma is a variant of a lemma that appeared in [BFN19]:

Lemma 3.3. *Let (X, d_X) be a metric space of size $|X| = \hat{n} > 1$, and let (Y, d_Y) be a metric space. Assume that $\alpha > 0$ is such that for any embedding $f : X \to Y$ it holds that $Energy_q(f) \geq \alpha$. For any $n \geq \hat{n}$ there is a metric space Z of size $|Z| = \Theta(n)$ such that any embedding $F : Z \to Y$ has $Energy_q(F) \geq \alpha/2$.*

Moreover, if X is a Euclidean subset then there is an embedding from Z into a finite dimensional Euclidean space with distortion $1 + \delta$ for any $\delta > 0$.

We prove the lemma for the completeness in Appendix A. The lemma implies that to obtain a family of metric spaces of any size $\Theta(n)$ it is enough to compose the metric spaces I in the family $\mathcal{P}$, of size $|I| = 6l = \hat{n}$ with, for example, an equilateral metric space on $\lceil n/6l \rceil$ points.

4 Lower Bounds for q-Moments of Distortion

In this section we prove Theorem 1.2 which provides a lower bound for q-moments of distortion. Similarly, to the proof for ℓ_1-distortion in Section 3, we prove the theorem first for metric space of fixed size $\hat{n} = O(1/\epsilon^2) \cdot e^{O(\epsilon q)}$ and show that it can be extended for metric spaces of size $\Theta(n)$, for any n.

To conclude the theorem for metric spaces of an arbitrary size $\Theta(n)$, we use the following variation of lemma proved in [BFN19]:

Lemma 4.1. *Let (X, d_X) be a metric space of size $|X| = n > 1$, and let (Y, d_Y) be a metric space. Assume that for any embedding $f : X \to Y$ it holds that ℓ_q-$dist(f) \geq 1 + \epsilon$. For any $n \geq \hat{n}$ there is a metric space Z of size $|Z| = \Theta(n)$ such that any embedding $F : Z \to Y$ has ℓ_q-$dist(F) \geq 1 + \epsilon/2$.*

Moreover, if X is a Euclidean subset then there is an embedding from Z into a finite dimensional Euclidean space with distortion $1 + \delta$ for any $\delta > 0$.

For simplicity we may assume w.l.o.g that $q \geq \frac{3}{\epsilon}$, otherwise the theorem follows from Theorem 1.1 by monotonicity of the ℓ_q-distortion. The construction has exactly the same structure as in the proof of Section 3. We construct a large family $\mathcal{P}$ of metric spaces, such that each $\mathrm{I} \in \mathcal{P}$ can be completely recovered by computing the inner products between the points in I. For a given $\epsilon < 0$, let $l = \lceil \frac{1}{\gamma^2 \epsilon^2} \rceil$ be an integer for some large constant $\gamma > 1$ to be determined later. We will construct point sets $\mathrm{I} \subset \ell_2^d$, where $d = l\tau$, where $\tau = e^{\epsilon q}$, and $|\mathrm{I}| = 3d$.

To obtain the lower bound, we assume that for all $\mathrm{I} \in \mathcal{P}$ there is an embedding $f : I \to \ell_2^k$, with ℓ_q-$dist(f) \leq 1 + \epsilon$. 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 $\mathrm{length}(\mathrm{I})$. Clearly, the length of the encoding is a function of the target dimension k (in addition to ϵ, q, and the size of I). The encoding is essentially an injective map from $\mathcal{P}$ to $\{0,1\}^{\mathrm{length}(\mathrm{I})}$. Therefore, the number of different sets that can be recovered by decoding is at most $2^{\mathrm{length}(\mathrm{I})}$. Now, because $|\mathcal{P}| = \binom{d}{l}^d \geq (d/l)^{ld}$ we get that $\mathrm{length}(\mathrm{I}) \geq dl \log(d/l) = dl \log(\tau)$. We will show that this implies the bound on $k \geq \Omega(l \log(\tau)) = \Omega(1/\epsilon^2 \cdot \epsilon q) = \Omega(q/\epsilon)$.

Let $\mathrm{I} \in \mathcal{P}$ be a metric space, and let $f : I \to \ell_2^k$ be an embedding with ℓ_q-$dist(f) \leq \epsilon$. In the following lemma we show that there exists a large subset of I which is contained in a ball of constant radius for which there is a large subset of the pairs for which absolute difference between the inner products of the images of the embedding and the original space is bounded by $O(\epsilon)$ additive error.

For convenience of the analysis we assume w.l.o.g that $\epsilon < 1/36$, otherwise the theorem follows from lower bounds proved in [BFN19].

Lemma 4.2. *For any $I \in \mathcal{P}$ let $f : I \to \ell_2^k$ be an embedding with ℓ_q-$dist(f) \leq \epsilon$, for $\epsilon < 1/36$. There is a subset $\hat{I} \subseteq I$ of size $|\hat{I}| \geq (1 - 3/\tau^4)|I|$ such that $f(\hat{I}) \subset B_2(1 + 6.01\epsilon)$, and for $1 - 2/\tau^4$ fraction of the pairs $(u, v) \in \binom{\hat{I}}{2}$ it holds that*

$$\left|\langle f(u), f(v)\rangle - \langle u, v\rangle\right| \leq 41\epsilon.$$

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

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

By the Markov inequality we have that there at 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

$$\left|expans_f(u, v) - 1\right| \leq \max\{expans_f(u, v) - 1, 1/expans_f(u, v) - 1\} = dist_f(u, v) - 1 \leq 6\epsilon.$$

For every point $o_j \in O$, let F_j be the set of points $v \in I \setminus \{o_j\}$ such that $\left|expans_f(\hat{o}, v) - 1\right| > 6\epsilon$. Then we have that the total number of pairs with this property it 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 $\hat{I} = I \setminus F_{j*}$ to be 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 $\|v - \hat{o}\|_2 \leq \|v\|_2 + \|\hat{o}\|_2 \leq 1 + \epsilon/6$, so that $\|f(v)\|_2 \leq (1 + 6\epsilon)(1 + \epsilon/6) \leq 1 + 6.01\epsilon$, implying that $f(v) \in B_2\left(1 + 6.01\epsilon\right)$.

Denote by $\hat{G}$ the set of pairs $(u, v) \in \binom{\hat{I}}{2}$ satisfying that $\left|expans_f(u, v) - 1\right| \leq 6\epsilon$. 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 $(u, v) \in \binom{\hat{I}}{2} \setminus \hat{G}$ 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 3. For all $(u, v) \in \binom{I}{2}$ and $1 \leq j \leq d$ we have:

$$
\begin{aligned}
&\left|\langle f(u), f(v)\rangle - \langle u, v\rangle\right| \leq \frac{1}{2}\left[\left|\|f(u)\|_2^2 - \|u\|_2^2\right| + \left|\|f(v)\|_2^2 - \|v\|_2^2\right| + \left|\|f(u) - f(v)\|_2^2 - \|u - v\|_2^2\right|\right] \\
&= \frac{1}{2}\left[|\|f(u) - f(\hat{o})\|_2^2 - \|u - 0\|_2^2| + |\|f(v) - f(\hat{o})\|_2^2 - \|v - 0\|_2^2| + \left|\|f(u) - f(v)\|_2^2 - \|u - v\|_2^2\right|\right] \\
&\leq \frac{1}{2}\left[(|(expans_f(u, \hat{o}))^2 - 1| + \epsilon/2) + (|(expans_f(v, \hat{o}))^2 - 1| + \epsilon/2) + \|u - v\|_2^2\,|(expans_f(u, v))^2 - 1|\right] \\
&\leq \frac{1}{2}|(expans_f(u, \hat{o}))^2 - 1| + \frac{1}{2}|(expans_f(v, \hat{o}))^2 - 1| + |(expans_f(u, v))^2 - 1| + \epsilon/2,
\end{aligned}
$$

where the second inequality is due to $|\,\|x - 0\|_2 - \|x - \hat{o}\|\,| \leq \|\hat{o}\| \leq \epsilon/6$, for all $x \in I$ and $o_j \in O$, and the last inequality holds since $\|u - v\| \leq diam(I) = \sqrt{2}$.

Now, we have that for all $(u, v) \in \hat{G}$:

$$
\begin{aligned}
|(expans_f(u, v))^2 - 1| &= |(expans_f(u, v)) - 1||(expans_f(u, v)) + 1| \\
&= |(expans_f(u, v)) - 1|(1 + \frac{\|f(u) - f(v)\|}{\|u - v\|}) \\
&\leq |(expans_f(u, v)) - 1|(1 + \frac{\|f(u)\| + \|f(v)\|}{\|u - v\|}) \\
&\leq |(expans_f(u, v)) - 1|(1 + 2(1 + 6.01\epsilon)) \\
&= |(expans_f(u, v)) - 1|(3 + 6.02\epsilon) \leq 6\epsilon(3 + 6.02\epsilon) \leq 20\epsilon,
\end{aligned}
$$

by our assumption on ϵ. Therefore

$$
\left|\langle f(u), f(v)\rangle - \langle u, v\rangle\right| \leq 2 \cdot 20\epsilon + \epsilon/2 \leq 41\epsilon.
$$

□

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 [AK17] as to preserve the inner product gap. The following lemma provides the probability that this procedure fails.

Lemma 4.3. *Let $X \subset \ell_2^k$ such that $X \subset B_2(r)$. For $\delta \le r/\sqrt{k}$ 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 $\eta \ge 1$, there is a $1 - 4e^{-\eta^2}$ fraction of the pairs $(u, v) \in \binom{X}{2}$ such that*

$$|\langle g(u), g(v)\rangle - \langle u, v\rangle| \le 3\sqrt{2}\eta\delta r.$$

The points of the grid can be represented using a number of bits bounded by: $L_{G_\delta} = k\log(4r/(\delta\sqrt{k}))$.

Proof. We randomly round the points in X to the δ-grid of $B_2(r)$, as proposed in [AK17], and bound the probability the inner product is not well preserved. For each point $v \in X$ we randomly and independently match a point $\tilde{v}$ 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 $E[\tilde{v}_i] = v_i$. In particular, this distribution is given by assigning the value $\left\lceil \frac{v_i}{\delta} \right\rceil \delta$ with probability $p = \left(\frac{v_i}{\delta} - \left\lfloor \frac{v_i}{\delta} \right\rfloor\right)$, and assigning the value $\left\lfloor \frac{v_i}{\delta} \right\rfloor \delta$ with probability $1 - p$. For any $u \ne v \in X$ we have:

$$|\langle \tilde{u}, \tilde{v}\rangle - \langle u, v\rangle| \le |\langle \tilde{u} - u, v\rangle| + |\langle \tilde{u}, \tilde{v} - v\rangle|.$$

Now, we have that $\mathrm{E}[\langle \tilde{u} - u, v\rangle] = \sum_{i=1}^k \mathrm{E}[\tilde{u}_i - u_i]v_i = 0$ and $\mathrm{E}[\langle \tilde{u}, \tilde{v} - v\rangle] = \sum_{i=1}^k \mathrm{E}[\tilde{u}_i]\mathrm{E}[\tilde{v}_i - v_i] = 0$. Next, we wish to make use of the Hoeffding bound. We therefore bound each of the terms $|(\tilde{u}_i - u_i)v_i| \le \delta|v_i|$ and the sum $\sum_{i=1}^k \delta^2 v_i^2 = \delta^2 r$, and $|\tilde{u}_i(\tilde{v}_i - v_i)| \le \delta(|u_i| + \delta)$, so that

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

Applying the Hoeffding bound we get that $\Pr[|\langle \tilde{u} - u, v\rangle| > \sqrt{2}\eta\delta r] \le 2e^{-\eta^2}$. and $\Pr[|\langle \tilde{u}, \tilde{v} - v\rangle| > 2\sqrt{2}\eta\delta r] \le 2e^{-\eta^2}$, and therefore $\Pr[|\langle \tilde{u}, \tilde{v}\rangle - \langle u, v\rangle| > 3\sqrt{2}\eta\delta r] \le 4e^{-\eta^2}$. This probability also bounds the expected number of pairs with this property so there must exist an embedding to the grid where this bound holds. The bound on the representation size is the same as in Lemma 3.2. ☐

Combining the lemmas we obtain:

Corollary 4.1. *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 $\hat{I} \subseteq I$ 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:*

$$\left|\langle g(f(u)), g(f(v))\rangle - \langle u, v\rangle\right| \le 51\epsilon,$$

where $g : \hat{I} \to G$. Moreover, the points in G can be uniquely represented by binary strings of length at most $L_G = k\log(4r\sqrt{q/(\epsilon k)})$ bits.

Proof. The corollary follows by applying Lemma 4.2 followed by Lemma 4.3 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. It follows that 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 4.3:

$$L_G = k\log(4r/(\delta\sqrt{k})) = k\log(4r\sqrt{q/(\epsilon k)}).$$

☐

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

Corollary 4.2. *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 $\left|\mathcal{Y}^G\right| \ge (1 - 6/\tau^2)|Y|$ such that for each $y \in \mathcal{Y}_G$ there is a subset $\mathcal{E}_y^G \subseteq E$ of size $\left|\mathcal{E}_y^G\right| \ge (1 - 6/\tau^2)\left|E\right|$ such that for all pairs $(y, e) \in \mathcal{Y}^G \times \mathcal{E}_y^G$ we have:*

$$\left|\langle g(f(y)), g(f(e))\rangle - \langle y, e\rangle\right| \le 51\epsilon,$$

where $g : \mathcal{Y}^G \cup \{\mathcal{E}_y^G\}_{y \in \mathcal{Y}^{\mathcal{G}}} \to G$. Moreover, the points in G can be uniquely represented by binary strings of length at most $L_G = k\log(4r\sqrt{q/(\epsilon k)})$ bits.

Proof. Applying Corollary 4.1 we have that there are at most $6/\tau^4$ pairs $(u, v) \in \binom{\hat{I}}{2}$ such that $\left|\langle g(f(u)), g(f(v))\rangle - \langle u, v\rangle\right| > 51\epsilon$. 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 $|\mathcal{E}_y^G|$. □

As before it we proceed to show that the whole instance I can be encoded using sufficiently small number of bits.

4.1 Encoding and Decoding

For an instance I $\in \mathcal{P}$ let f : I $\to \ell_2^k$ be an embedding 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 4.2, 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 $\left|\langle g(f(y)), g(f(e))\rangle - \langle y, e\rangle\right| \le 51\epsilon$. Moreover, each point in G can be uniquely encoded using at most $L_G = k\log(4r\sqrt{q/(\epsilon k)})$ bits.

The encoding is done according to the description in Section 3.1 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 $\gamma = 52$.

4.2 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 $(1/t)d^2(2+\log(et)) + 2dL_G \ge \log|\mathcal{P}|$. On the other hand, the size of the family is $|\mathcal{P}| = \binom{d}{l}^d \ge (d/l)^{ld} = \tau^{ld}$, so that $\log(|\mathcal{P}|) = ld\log(\tau)$. We therefore derive the following inequality

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

as $(1/t)d(2 + \log(et)) \le d(2\log(\tau) + 4)/\tau^2 \le d/(2\tau)\log(\tau) = l\log(\tau)/2$, using that $\log(\tau) > 4$.

Recall that $L_G = k\log(4r\sqrt{q/(\epsilon k)}) = \frac{1}{2}k\log\left(64(q/(\epsilon k))\right)$. This implies that

$$k\log\left(64\left(\frac{q}{\epsilon k}\right)\right) \ge (1/2)l\log(\tau) \ge 1/(2\gamma^2 \cdot \epsilon^2) \cdot \epsilon q = 1/(2\gamma^2) \cdot q/\epsilon.$$

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

$$1 \le x \log\left(\frac{0.5}{x} \cdot 256\gamma^2\right) = x\log(0.5/x) + x\log\left(256\gamma^2\right) \le 1/2 + 2x(4 + \log\gamma),$$

where the last inequality we have used $x\log(0.5/x) \le 0.5/(e\ln 2) < 1/2$ for all x. This implies the desired lower bound on the dimension:

$$k \ge 1/(8\gamma^2(4 + \log\gamma)) \cdot q/\epsilon.$$

References

[ABN11] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. *Advances in Mathematics*, 228(6):3026 – 3126, 2011.

[AK17] 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.

[Alo09] Noga Alon. Perturbed identity matrices have high rank: Proof and applications. *Combinatorics, Probability & Computing*, 18(1-2):3–15, 2009.

[AW19] Ehsan Amid and Manfred K. Warmuth. Trimap: Large-scale dimensionality reduction using triplets, 2019.

[BFN19] Y. Bartal, Nova Fandina, and Ofer Neiman. Dimensionality reduction: theoretical perspective on practical measures. In *NeurIPS*, 2019.

[BLMN05] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric ramsey-type phenomena. *Annals of Mathematics*, 162(2):643–709, 2005.

[CDK+04] 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.

[CS13] 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, Oct. 2013.

[CVvL18a] 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.

[CVvL18b] 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.

[EFN15] 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.

[GMH95] 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.

[Hei8a] 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.

[IM98] 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.

[IM04] 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.

[Ind01a] Piotr Indyk. Algorithmic applications of low-distortion geometric embeddings. In *Proceedings of the 42nd IEEE symposium on Foundations of Computer Science*, FOCS '01, pages 10–33, Washington, DC, USA, 2001. IEEE Computer Society.

[Ind01b] Piotr Indyk. Algorithmic applications of low-distortion geometric embeddings. In *FOCS*, pages 10–33. IEEE Computer Society, 2001.

[JL84] 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.

[KSW09] 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.

[Lin02] N. Linial. Finite metric spaces- combinatorics, geometry and algorithms. In *Proceedings of the ICM*, 2002.

[LN17] 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.

[Mat90] Jiří Matoušek. Bi-Lipschitz embeddings into low-dimensional Euclidean spaces. *Commentat. Math. Univ. Carol.*, 31(3):589–600, 1990.

[Mat02] Jiří Matoušek. *Lectures on Discrete Geometry*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002.

[MHSG18] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018.

[ST04] 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.

[SXBL06] Puneet Sharma, Zhichen Xu, Sujata Banerjee, and Sung-Ju Lee. Estimating network proximity and latency. *Computer Communication Review*, 36(3):39–50, 2006.

[vdMH08] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

[Vem04] 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.

[VHM07] 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.

[WHRS20] 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.

A Proof of Lemma 3.3

We provide the proof of the lemma for the completeness. We start with definition of the composition of metric spaces given in [BLMN05]:

Definition A.1. *Let (S, d_S), (T, d_T) be finite metric spaces. For any $\beta \geq 1/2$, the β-composition of S with T, denoted by $Z = S_\beta[T]$, is a metric space of size $|Z| = |S| \cdot |T|$ constructed in the following way. Each point $u \in X$ is substituted with a copy of the metric space T, denoted by $T^{(u)}$. Let $u, v \in S$, and $z_i \neq z_j \in Z$, such that $z_i \in T^{(u)}$, and $z_j \in T^{(v)}$, then*

$$d_Z(z_i, z_j) = \begin{cases} \frac{1}{\gamma}\frac{1}{\beta} \cdot d_T(z_i, z_j), & u = v \\ d_S(u, v), & u \neq v \end{cases}$$

where $\gamma = \frac{\max_{t \neq t' \in T}\{d_T(t,t')\}}{\min_{s \neq s' \in s}\{d_S(s,s')\}}$.

Proof. Given any $n \geq \hat{n}$ let $m = \lceil n/\hat{n} \rceil$, and let T be any m-point metric space. For any $\beta \geq 1/2$ let Z be the β-metric composition of X with T (note that the choice of T is arbitrary), and let $N = |Z| = \hat{n}m = \Theta(n)$. Let $F : Z \to Y$ be any embedding, and consider the set of pairs $B \subseteq \binom{Z}{2}$, $B = \{(z_i, z_j) | z_i \in T^{(u)}, z_j \in T^{(v)}, \forall u \neq v \in X\}$. Then, $|B| = m^2 \cdot \binom{\hat{n}}{2}$. Let $q \geq 1$, and note that for all $z_i \neq z_j \in Z$ it holds that $|expans_F(z_i, z_j) - 1| \geq 0$. Then

$$(Energy_q(F))^q \geq \frac{1}{\binom{N}{2}} \sum_{z_i \neq z_j \in B} |expans_F(z_i, z_j) - 1|^q \geq \frac{1}{2} \cdot \frac{1}{m^2\binom{\hat{n}}{2}} \sum_{z_i \neq z_j \in B} |expans_F(z_i, z_j) - 1|^q.$$

Let $\mathcal{X}$ denote the family of all possible n-point subsets $\bar{X} \subset Z$, where each point of $\bar{X}$ is chosen from exactly one of the copies $T^{(x_1)}, T^{(x_2)}, \ldots T^{(x_{\hat{n}})}$. Namely, each $\bar{X}$ is a metric space isometric to X. Let $F|_{\bar{X}}$ denote the embedding F restricted to the points of $\bar{X}$. The size of the family $|\mathcal{X}| = m^{\hat{n}}$, and it holds that

$$\frac{1}{m^{\hat{n}}}\sum_{\bar{X}\in\mathcal{X}}(Energy_q(F|_{\bar{X}})^q = \frac{1}{m^{\hat{n}}}\sum_{\bar{X}\in\mathcal{X}}\frac{1}{\binom{\hat{n}}{2}}\sum_{u,v\in\bar{X}}|expans_F(u,v)-1|^q$$
$$= \frac{1}{m^{\hat{n}}}\frac{1}{\binom{\hat{n}}{2}}\sum_{z_i\neq z_j\in B} m^{\hat{n}-2}\cdot|expans_F(u,v)-1|^q$$
$$= \frac{1}{\binom{\hat{n}}{2}m^2}\cdot\sum_{z_i\neq z_j\in B}|expans_F(u,v)-1|^q.$$

By the assumption it holds that $Energy_q(F|_{\bar{X}})^q \geq \alpha^q$, implying that $Energy_q(F) \geq \alpha$.

Note that the bound on Energy does not depend on the value of β. It was shown in [BLMN05] (Proposition 2.12) that if X and T are both Euclidean subsets, then their composition $Z = X_\beta[T]$ embeds into a finite dimensional Euclidean space with distortion $(1+\epsilon)$, for $\beta = O(1/\epsilon)$. This completes the proof of the lemma. □

B More Additive Distortion Measures

In this section we prove Theorem 1.3. We first define all the notions appearing in the theorem. In addition to ℓ_q-distortion, REM and Energy, several more practical quality measurement criteria were analyzed in [BFN19]. In particular, Stress (and its variant) is widely used in multidimensional scaling community:

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

Finally, [CVvL18b] defined σ-distortion and showed it to be particularly useful for application of metric embedding in machine learning. For $r \geq 1$, let $\ell_r\text{-}expans(f) = (\binom{n}{2}^{-1}\sum_{u\neq v}(expans_f(u,v))^r)^{1/r}$.

$$\sigma_{q,r}(f) = \left(\frac{1}{\binom{|X|}{2}}\left|\frac{expans_f(u,v)}{\ell_r\text{-}expans(f)}-1\right|^q\right)^{1/q}.$$

Following are some of the observations made in [BFN19] about basic relations between the measures:

Claim B.1. *For an embedding $f : X \to Y$, for any $r \geq 1$, there is an embedding $f' : X \to Y$ such that $\sigma_{1,r}(f) = Energy_1(f')$.*

Claim B.2. *For any embedding $f : X \to Y$ there is an embedding $f' : X \to Y$ such that* $\mathrm{Stress}_1(f') \leq 4\cdot\mathrm{Stress}^*{}_1(f)$.

Together with Claim 2.1, these imply that

Corollary B.1. *In order to show a lower bound for ℓ_1-distortion, REM_1 and $\sigma_{1,r}$ it is enough to lower bound $Energy_1$. In order to show a lower bound for* $\mathrm{Stress}^*{}_1$ *it is enough to lower bound* Stress_1.

We are ready now to prove Theorem 1.3. We restate it here for convenience:

Theorem 1.3. *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 $k = \Omega(1/\epsilon^2)$.*

Proof. We already proved the theorem for $Energy_1$, therefore by Corollary B.1 it remains to prove it for Stress_1. First, not that for any embedding $f : X \to Y$

$$Stress_1(f) = \frac{\sum_{u \neq v \in X} |\hat{d}_{uv} - d_{uv}|}{S[X]} = \frac{1}{S[X]/\binom{|X|}{2}} \frac{\sum_{u \neq v \in X} d_{uv}|expans_f(u,v) - 1|}{\binom{|X|}{2}},$$

where $S[X] = \sum_{u \neq v \in X} d_{uv}$. We define

$$\overline{Stress}_1(f) := \frac{\sum_{u \neq v \in X} d_{uv}|expans_f(u,v) - 1|}{\binom{|X|}{2}}.$$

Observe that if X is such that $S[X]/\binom{|X|}{2} \leq c$ for a constant $c > 0$ then $Stress_1(f) = \Omega(\overline{Stress}_1(f))$. Therefore, it is enough to show that there is a metric space (of an arbitrary size $\Theta(n)$) with at most constant average distance on which the lower bound is obtained for $\overline{Stress}_1$. We note that the composition Lemma 3.3 also works for constructing arbitrary size metrics for $\overline{Stress}$ notion.

Therefore, we show that there is a metric space I of size $\Theta(1/\epsilon^2)$ such that if any embedding $f : I \to \ell_2^k$ has $\overline{Stress}_1(f) \leq \epsilon$ then $k = \Omega(1/\epsilon^2)$. In addition, I is such that $S[I]/\binom{|I|}{2} \leq \sqrt{2}$. Since a metric space obtained by the composition in Claim 3.3 has the same diameter as the base space (I in our case), this will complete the proof, and so its embedding in Euclidean space will at most increase this bound further by extra $1 + \delta$, for arbitrary $\delta > 0$.

We argue that a slight variation on the proof in Section 3 for $Energy_1$ works for $\overline{Stress}_1$ as well. The main change is that in the estimations in (3.0.3) in Lemma 3.1we need to modify the last inequality as follows:

$$\begin{aligned}
&\left|\langle f(u), f(v)\rangle - \langle u, v\rangle\right| \leq \frac{1}{2}\left[\left|\|f(u)\|_2^2 - \|u\|_2^2\right| + \left|\|f(v)\|_2^2 - \|v\|_2^2\right| + \left|\|f(u) - f(v)\|_2^2 - \|u - v\|_2^2\right|\right] \\
&= \frac{1}{2}\left[|\|f(u) - f(o_j)\|_2^2 - \|u - 0\|_2^2| + |\|f(v) - f(o_j)\|_2^2 - \|v - 0\|_2^2| + \left|\|f(u) - f(v)\|_2^2 - \|u - v\|_2^2\right|\right] \\
&\leq \frac{1}{2}\left[\|u\|_2^2\left(|(expans_f(u, o_j))^2 - 1| + \epsilon/2\right) + \left(\|v\|_2^2\,|(expans_f(v, o_j))^2 - 1| + \epsilon/2\right) + \|u - v\|_2^2\,|(expans_f(u, v))^2 - 1|\right] \\
&\leq \frac{1}{2}\|u\|_2\,|(expans_f(u, o_j))^2 - 1| + \frac{1}{2}\|v\|_2\,|(expans_f(v, o_j))^2 - 1| + \sqrt{2}\,\|u - v\|_2\,|(expans_f(u, v))^2 - 1| + \epsilon/2.
\end{aligned}$$

Then we continue as in Lemma 3.1 to bound the terms of the form $\|u - v\|_2\,|(expans_f(u, v))^2 - 1|$ (for any $(u, v) \in \hat{I}$ by some constant times $\|u - v\|\,|expans_f(u, v) - 1|$, which when summed over all pairs is a constant times $\overline{Stress}_1(f)$. This brings the bound on the absolute value of the inner products difference to be in the form of the assumption: ($\overline{Stress}_1(f) \leq \epsilon$). The rest of the proof carries on exactly the same as before with appropriate adjustment of the constants. Recall that each $\mathrm{I} \in \mathcal{P}$ has $\mathrm{diam}(\mathrm{I}) \leq \sqrt{2}$, so that this bound applied to $S[I]/\binom{|I|}{2}$ as well, which completes the proof of the theorem.

□