Optimal terminal dimensionality reduction in Euclidean space

Shyam Narayanan* Jelani Nelson†

October 21, 2018

Abstract

Let $\varepsilon \in (0,1)$ and $X \subset \mathbb{R}^d$ be arbitrary with $|X|$ having size $n > 1$. The Johnson-Lindenstrauss lemma states there exists $f : X \to \mathbb{R}^m$ with $m = O(\varepsilon^{-2} \log n)$ such that

$$\forall x \in X\ \forall y \in X, \|x - y\|_2 \le \|f(x) - f(y)\|_2 \le (1 + \varepsilon)\|x - y\|_2.$$

We show that a strictly stronger version of this statement holds, answering one of the main open questions of [MMMR18]: "$\forall y \in X$" in the above statement may be replaced with "$\forall y \in \mathbb{R}^d$", so that f not only preserves distances within X, but also distances *to* X from the rest of space. Previously this stronger version was only known with the worse bound $m = O(\varepsilon^{-4} \log n)$. Our proof is via a tighter analysis of (a specific instantiation of) the embedding recipe of [MMMR18].

1 Introduction

Metric embeddings may play a role in algorithm design when input data is geometric, in which the technique is applied as a pre-processing step to map input data living in some metric space $(\mathcal{X}, d_{\mathcal{X}})$ into some target space $(\mathcal{Y}, d_{\mathcal{Y}})$ that is algorithmically friendlier. One common approach is that of *dimensionality reduction*, in which $\mathcal{X}$ and $\mathcal{Y}$ are both subspaces of the same normed space, but where $\mathcal{Y}$ is of much lower dimension. Working with lower-dimensional embedded data then typically results in efficiency gains, in terms of memory, running time, and/or other resources.

A cornerstone result in this area is the Johnson-Lindenstrauss (JL) lemma [JL84], which provides dimensionality reduction for Euclidean space.

Lemma 1.1. *[JL84] Let $\varepsilon \in (0,1)$ and $X \subset \mathbb{R}^d$ be arbitrary with $|X|$ having size $n > 1$. There exists $f : X \to \mathbb{R}^m$ with $m = O(\varepsilon^{-2} \log n)$ such that*

$$\forall x, y \in X, \|x - y\|_2 \le \|f(x) - f(y)\|_2 \le (1 + \varepsilon)\|x - y\|_2. \tag{1}$$

It has been recently shown that the dimension m of the target Euclidean space achieved by the JL lemma is best possible, at least for $\varepsilon \gg 1/\sqrt{\min(n, d)}$ [LN17] (see also [AK17]).

The multiplicative factor on the right hand side of Eqn. (1), in this case $1 + \varepsilon$, is referred to as the *distortion* of the embedding f. Recent work of Elkin et al. [EFN17] showed a stronger form

*Harvard University. shyamnarayanan@college.harvard.edu. Supported by a PRISE fellowship and a Herchel-Smith Fellowship.

†Harvard University. minilek@seas.harvard.edu. Supported by NSF CAREER award CCF-1350670, NSF grant IIS-1447471, ONR grant N00014-18-1-2562, ONR DORECG award N00014-17-1-2127, an Alfred P. Sloan Research Fellowship, and a Google Faculty Research Award.

of Euclidean dimensionality reduction for the case of constant distortion. Namely, they showed that in Eqn. (1), whereas x is taken as an arbitrary point in X, y may be taken as an arbitrary point in $\mathbb{R}^d$. They called such an embedding a *terminal embedding*[1]. Though rather than achieving terminal distortion $1+\varepsilon$, their work only showed how to achieve constant terminal distortion with $m = O(\log n)$ for a constant that could be made arbitrarily close to $\sqrt{10}$ (see [EFN17, Theorem 1]). Terminal embeddings can be useful in static high-dimensional computational geometry data structural problems. For example, consider nearest neighbor search over some finite database $X \subset \mathbb{R}^d$. If one builds a data structure over $f(X)$ for a terminal embedding f, then *any* future query is guaranteed to be handled correctly (or, at least, the embedding will not be the source of failure). Contrast this with the typical approach where one uses a randomized embedding oblivious to the input (e.g. random projections) that preserves the distance between any fixed pair of vectors with probability $1 - 1/\mathit{poly}(n)$. One can verify that the embedding preserves distances *within* X during pre-processing, but for any later query there is some non-zero probability that the embedding will fail to preserve the distance between the query point q and some points in X.

Subsequent to [EFN17], work of Mahabadi et al. [MMMR18] gave a construction for terminal dimensionality reduction in Euclidean space achieving terminal distortion $1+\varepsilon$, with $m = O(\varepsilon^{-4}\log n)$. They asked as one of their main open questions (see [MMMR18, Open Problem 3]) whether it is possible to achieve this terminal embedding guarantee with $m = O(\varepsilon^{-2}\log n)$, which would be optimal given the JL lower bound of [LN17]. Our contribution in this work is to resolve this question affirmatively; the following is our main theorem.

Theorem 1.1. *Let $\varepsilon \in (0,1)$ and $X \subset \mathbb{R}^d$ be arbitrary with $|X|$ having size $n > 1$. There exists $f : X \to \mathbb{R}^m$ with $m = O(\varepsilon^{-2}\log n)$ such that*

$$\forall x \in X, \forall y \in \mathbb{R}^d,\ \ \|x-y\|_2 \le \|f(x) - f(y)\|_2 \le (1+\varepsilon)\|x-y\|_2.$$

The embedding we analyze in this work is in fact one that fits within the family introduced in [MMMR18]; our contribution is to provide a sharper analysis.

We note that unlike in [MMMR18], which provided a deterministic polynomial time construction of the terminal embedding, we here only provide a Monte Carlo polynomial time construction algorithm. Our error probability though only comes from mapping the points in X. If the points in X are mapped to $\mathbb{R}^m$ well (with low "convex hull distortion", which we define later), which happens with high probability, then our final terminal embedding is guaranteed to have low terminal distortion as map from all of $\mathbb{R}^d$ to $\mathbb{R}^m$.

Remark. Unlike in the JL lemma for which the embedding f may be linear, the terminal embedding we analyze here (as was also in the case in [EFN17, MMMR18]) is nonlinear, as it must be. To see that it must be nonlinear, consider that any linear embedding with constant terminal distortion, $x \mapsto \Pi x$ for some matrix $\Pi \in \mathbb{R}^{m\times d}$, must have $\Pi(x-y) \neq 0$ for any $x \in X$ and any $y \in \mathbb{R}^d$ on the unit sphere centered at x. In other words, one needs $\mathit{ker}(\Pi) \neq \emptyset$, which is impossible unless $m \ge d$.

1.1 Overview of approach

We outline both the approach of previous works as well as our own. In all these approaches, one starts with an embedding $f : X \to \ell_2^m$ with good distortion, then defines an *outer extension* f_{Ext} as introduced in [EFN17] and defined explicitly in [MMMR18].

[1]More generally, a terminal embedding from $(\mathcal{X}, d_{\mathcal{X}})$ into $(\mathcal{Y}, d_{\mathcal{Y}})$ with terminal set $K \subset \mathcal{X}$ and *terminal distortion* α is a map $f : \mathcal{X} \to \mathcal{Y}$ s.t. $\exists\ c > 0$ satisfying $d_{\mathcal{X}}(u,w) \le c \cdot d_{\mathcal{Y}}(f(u), f(w)) \le \alpha \cdot d_{\mathcal{X}}(u,w)$ for all $u \in K, w \in \mathcal{X}$.

Definition 1.1. For $f : X \to \mathbb{R}^m$ and $Z \supsetneq X$, we say $g : Z \to \mathbb{R}^{m'}$ is an *outer extension* of f if $m' \geq m$, and $g(x)$ for $x \in X$ has its first m entries equal to $f(x)$ and last $m' - m$ entries all 0.

In [EFN17, MMMR18] and the current work, a terminal embedding is obtained by, for each $u \in \mathbb{R}^d \backslash X$, defining an outer extension $f_{\text{Ext}}^{(u)}$ for $Z = X \cup \{u\}$ with $m' = m + 1$. Since $f_{\text{Ext}}^{(u)}$ and $f_{\text{Ext}}^{(u')}$ act identically on X for any $u, u' \in \mathbb{R}^d$, we can then define our final terminal embedding by $\tilde{f}(u) = (f(u), 0)$ for $u \in X$ and $\tilde{f}(u) = f_{\text{Ext}}^{(u)}(u)$ for $u \in \mathbb{R}^d \backslash X$, which will have terminal distortion $\sup_{u \in \mathbb{R}^d \backslash X} \mathsf{Dist}(f_{\text{Ext}}^{(u)})$, where $\mathsf{Dist}(g)$ denotes the distortion of g. The main task is thus creating an outer extension with low distortion for a set $Z = X \cup \{u\}$ for some u. In all the cases that follow, we will have $f_{\text{Ext}}(x) = (f(x), 0)$ for $x \in X$, and we will then specify how to embed points $u \notin X$.

The construction of Elkin et al. [EFN17], the "EFN extension", is as follows. Suppose $f : X \to \ell_p^m$ is an α-distortion embedding. The EFN extension $f_{\text{EFN}} : X \cup \{u\} \to \ell_p^{m+1}$ is then defined by $f_{\text{EFN}}(u) = (f(\rho_{\ell_p}(u)), d(\rho_{\ell_p}(u), u))$, where $\rho_d(u) = argmin_{x \in X}\, d(x, u)$ for some metric d. In the case that we view $X \cup \{u\}$ as living in the metric space ℓ_p^d, Elkin et al. showed that the EFN extension has terminal distortion at most $2^{(p-1)/p} \cdot ((2\alpha)^p + 1)^{1/p}$ [EFN17, Theorem 1]. If $p = 2$ and f is obtained via the JL lemma to have distortion $\alpha \leq 1 + \varepsilon$, this implies the EFN extension would have terminal distortion at most $\sqrt{10} + O(\varepsilon)$.

The EFN extension does not in general achieve terminal distortion $1+\varepsilon$, even if starting with f a perfect isometry. In fact, the bound of [EFN17] showing distortion at least $\sqrt{10}$ is sharp. Consider for example $X = \{-1, 0, 2\} \subset \mathbb{R}$. Consider the identity map $f(x) = x$, which has distortion 1 as a map from (X, ℓ_2) to $(\mathbb{R}, \ell_2)$. Then $f_{\text{EFN}}(-1) = (-1, 0)$, $f_{\text{EFN}}(0) = (0, 0)$, $f_{\text{EFN}}(2) = (2, 0)$, and $f_{\text{EFN}}(1) = (0, 1)$, and thus $\|f_{\text{EFN}}(1) - f_{\text{EFN}}(-1)\|_2 = \sqrt{2}$ (distance shrunk by a $\sqrt{2}$ factor) and $\|f_{\text{EFN}}(1) - f_{\text{EFN}}(2)\|_2 = \sqrt{5}$ (distance increased by a $\sqrt{5}$ factor). Thus f_{EFN} has terminal distortion at least (in fact exactly equal to) $\sqrt{10}$. This example in fact shows sharpness for ℓ_p for all $p \geq 1$.

Thus to achieve terminal distortion $1 + \varepsilon$, the work of [MMMR18] had to develop a new outer extension, the "MMMR extension", which they based on the following core lemma.

Lemma 1.2 ([MMMR18, Lemma 3.1 (rephrased)]). *Let X be a finite subset of ℓ_2^d, and suppose $f : X \to \ell_2^m$ has distortion $1 + \gamma$. Fix some $u \in \mathbb{R}^d$, and define $x_0 := \rho_{\ell_2}(u)$. Then $\exists u' \in \mathbb{R}^m$ s.t.*

- $\|u' - f(x_0)\|_2 \leq \|u - x_0\|_2$, *and*
- $\forall x \in X$, $|\langle u' - f(x_0), f(x) - f(x_0)\rangle - \langle u - x_0, x - x_0\rangle| \leq 3\sqrt{\gamma}(\|x - x_0\|_2^2 + \|u - x_0\|_2^2)$

Mahabadi et al. then used the u' promised by Lemma 1.2 as part of a construction that takes a $(1 + \gamma)$-distortion embedding $f : X \to \ell_2^m$ and uses it in a black box way to construct an outer extension $f_{\text{MMMR}} : X \cup \{u\} \to \ell_2^{m+1}$. In particular, they define $f_{\text{MMMR}}(u) = (u', \sqrt{\|u - x_0\|_2^2 - \|u' - f(x_0)\|_2^2})$. It is clear this map perfectly preserves the distance from u to x_0; in [MMMR18, Theorem 1.5], it is furthermore shown that f_{MMMR} preserves distances from u to all of X up to a $1 + O(\sqrt{\gamma})$ factor. Thus one should set $\gamma = \Theta(\varepsilon^2)$ so that f_{MMMR} has distortion $1 + \varepsilon$, which is achieved by starting with an f guaranteed by the JL lemma with $m = \Theta(\varepsilon^{-4} \log n)$.

They then showed that this loss is *tight*, in the sense that there exist X, u, $f : X \to \mathbb{R}^m$, where f has distortion $1+\gamma$, such that *any* outer extension f_{Ext} to domain $X \cup \{u\}$ has distortion $1 + \Omega(\sqrt{\gamma})$ [MMMR18, Section 3.2]. Thus, seemingly a new approach is needed to achieve $m = O(\varepsilon^{-2} \log n)$.

One may be discouraged by the above-mentioned tightness of the $\gamma \to \Omega(\sqrt{\gamma})$ loss, but in this work we show that, in fact, the MMMR extension can be made to provide $1 + \varepsilon$ distortion with the optimal $m = O(\varepsilon^{-2} \log n)$! The tightness result mentioned in the last paragraph is only an obstacle to using the low-distortion property of f *in a black box way*, as it is only shown that there *exist* f where the $\gamma \to \Omega(\sqrt{\gamma})$ loss is necessary. However, the f we are using is not an arbitrary f, but rather is the f obtained via the JL lemma. A standard way of proving the JL lemma is to choose $\Pi \in \mathbb{R}^{m\times d}$ with i.i.d. subgaussian entries, scaled by $1/\sqrt{m}$ for $m = \Theta(\varepsilon^{-2}\log n)$ [Ver18, Exercise 5.3.3]. The low-distortion embedding is then $f(x) = \Pi x$. We show in this work that by not just using that this f is a low-distortion embedding for X, but rather that it satisfies a stronger property we dub *convex hull distortion* (which we show $x \mapsto \Pi x$ does satisfy with high probability), one can achieve the desired terminal embedding result with optimal m.

Definition 1.2. For $T \subset S^{d-1}$ a subset of the unit sphere in $\mathbb{R}^d$, and $\varepsilon \in (0,1)$, we say for $\Pi \in \mathbb{R}^{m\times d}$ that Π *provides ε-convex hull distortion for T* if

$$\forall x \in \mathrm{conv}(T),\ |\|\Pi x\|_2 - \|x\|_2| < \varepsilon$$

where $\mathrm{conv}(T) := \{\sum_i \lambda_i t_i : \forall i\ t_i \in T, \lambda_i \geq 0, \sum_i \lambda_i = 1\}$ denotes the convex hull of T.

We show that a random Π with subgaussian entries provides ε-convex hull distortion for T with probability at least $1-\delta$ as long as $m = \Omega(\varepsilon^{-2}\log(|T|/(\varepsilon\delta)))$. We then replace Lemma 1.2 with a new lemma that shows that as long as $f(x) = \Pi x$ does not just have $1+\gamma$ distortion for X, but rather provides γ-convex hull distortion for $T = \{(x-y)/\|x-y\|_2 : x, y \in X\}$, then the $3\sqrt{\gamma}$ term on the RHS of the second bullet of Lemma 1.2 can be replaced with $O(\gamma)$ (plus one other technical improvement; see Lemma 3.1). Note $|T| = \binom{n}{2}$, so $\log|T| = O(\log n)$. We then show how to use the modified lemma to achieve an outer extension with $m = O(\varepsilon^{-2}\log(n/\varepsilon)) = O(\varepsilon^{-2}\log n)$. This last equality holds since we may assume $\varepsilon = \Omega(1/\sqrt{n})$, since otherwise there is a trivial terminal embedding with $m = n = O(\varepsilon^{-2})$ with no distortion: if $d \leq n$, take the identity map. Else, translate X so $0 \in X$; then $E := span(X)$ has $dim(E) \leq n-1$. By rotation, we can assume $E = span\{e_1, \ldots, e_{n-1}\}$ so that every $x \in E$ can be written as $\sum_{i=1}^{n-1} \alpha(x)_i e_i$ for some vector $\alpha(x) \in \mathbb{R}^{n-1}$. We can then define a terminal embedding $\tilde{f} : \mathbb{R}^d \to \mathbb{R}^n$ with $\tilde{f}(x) = (\alpha(proj_E(x)), \|proj_{E^\perp}(x)\|_2)$ for all $x \in \mathbb{R}^d$. Here $proj_E$ denotes orthogonal projection onto E.

2 Preliminaries

For our optimal terminal embedding analysis, we rely on two previous results. The first result is the von Neumann Minimax theorem [vN28], which was also used in the terminal embedding analysis in [MMMR18]. The theorem states the following:

Theorem 2.1. *Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be compact convex sets. Suppose that $f : X \times Y \to \mathbb{R}$ is a continuous function that satisfies the following properties:*

1. *$f(\cdot, y) : X \to \mathbb{R}$ is convex for any fixed $y \in Y$,*
2. *$f(x, \cdot) : Y \to \mathbb{R}$ is concave for any fixed $x \in X$.*

Then, we have that

$$\min_{x\in X}\max_{y\in Y} f(x,y) = \max_{y\in Y}\min_{x\in X} f(x,y).$$

The second result is a result of Dirksen [Dir15, Dir16] that provides a uniform tail bound on subgaussian empirical processes. To explain the result, we first make the following definitions:

Definition 2.1. A semi-metric d on a space X is a function $X \times X \to \mathbb{R}_{\geq 0}$ such that $d(x,x) = 0$, $d(x,y) = d(y,x)$, and $d(x,y) + d(y,z) \geq d(x,z)$ for all $x, y, z \in X$.

Note a semi-metric may have $d(x,y) = 0$ for $x \neq y$.

Definition 2.2. Given a semi-metric d on T, and subset $S \subset T$, define $d(t,S) = \inf_{s \in S} d(t,s)$ for any point $t \in T$.

Definition 2.3. Given a semi-metric d on T, define

$$\gamma_2(T,d) = \inf_{\{S_r\}_{r=0}^{\infty}} \sup_{t \in T} \sum_{r \geq 0} 2^{r/2} d(t, S_r),$$

where the first infimum runs over all subsequences $S_0 \subset S_1 \subset \ldots \subset T$ where $|S_0| = 1$ and $|S_r| \leq 2^{2^r}$.

Definition 2.4. For any random variable X, we define the *subgaussian norm* of X as

$$\|X\|_{\psi_2} = \inf \left\{ C \geq 0 : \mathbb{E}\left(e^{X^2/C^2} \right) \leq 2 \right\}.$$

It is well known that the subgaussian norm as defined above indeed defines a norm [BLM13]. We also note the following proposition:

Proposition 2.1. *[BLM13] There exists some constant k such that if $X_1, ..., X_n$ are i.i.d. subgaussians with mean 0 and subgaussian norm C, then for any $a_1, ..., a_n \in \mathbb{R}$, $a_1 X_1 + ... + a_n X_n$ is subgaussian with subgaussian norm $\leq k\|a\|_2 C$, where $\|a\|_2 = \sqrt{a_1^2 + ... + a_n^2}$.*

Theorem 2.2. *[Dir16, Theorem 3.2] Let T be a set, and suppose that for every $t \in T$ and every $1 \leq i \leq m$, $X_{t,i}$ is a random variable with finite expected value and variance. For each $t \in T$, let*

$$A_t = \frac{1}{m} \sum_{i=1}^{m} \left(X_{t,i}^2 - \mathbb{E} X_{t,i}^2 \right).$$

Consider the following semi-metric on T: for $s, t \in T$, define

$$d_{\psi_2}(s,t) := \max_{1 \leq i \leq m} \|X_{s,i} - X_{t,i}\|_{\psi_2}$$

and define

$$\overline{\Delta}_{\psi_2}(T) := \sup_{t \in T} \max_{1 \leq i \leq m} \|X_{t,i}\|_{\psi_2}.$$

Then, there exist constants $c, C > 0$ such that for all $u \geq 1$,

$$\mathbb{P}\left(\sup_{t \in T} |A_t| \geq C \left(\frac{1}{m} \gamma_2^2(T, d_{\psi_2}) + \frac{1}{\sqrt{m}} \overline{\Delta}_{\psi_2}(T) \gamma_2(T, d_{\psi_2}) \right) + c \left(\sqrt{u} \frac{\overline{\Delta}_{\psi_2}^2(T)}{\sqrt{m}} + u \frac{\overline{\Delta}_{\psi_2}^2(T)}{m} \right) \right) \leq e^{-u}.$$

3 Construction of the Terminal Embedding

3.1 Universal Dimensionality Reduction with an additional ℓ_1 condition

Here we show that for all $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ a set of unit norm vectors, there exists $\Pi \in \mathbb{R}^{m \times d}$ for $m = O(\varepsilon^{-2} \log n)$ providing ε-convex hull distortion for X, as defined in Definition 1.2.

If $\varepsilon^{-2} > n$, this construction follows by projecting onto a spanning subspace of $x_1, \dots, x_n$ and choosing an orthonormal basis. For $\varepsilon^{-2} < n$, our goal is to show that if $\Pi \in \mathbb{R}^{m \times d}$ is a normalized random matrix with i.i.d. subgaussian entries (normalized by $1/\sqrt{m}$), then Π provides ε-convex hull distortion with high probability.

Define $T = \operatorname{conv}(X)$. We apply Theorem 2.2 as follows. For some m which we will choose later, let Π^0 be a matrix in $\mathbb{R}^{m \times d}$ with i.i.d. subgaussian entries with mean 0, variance 1, and some subgaussian norm C_1. Let Π be the scaled matrix, i.e. $\frac{1}{\sqrt{m}} \cdot \Pi^0$. Let Π_i denote the ith row of Π^0 and for any $t \in \mathbb{R}^d$, let $X_{t,i} = \Pi_i t$. Finally, let T_k be the subset of T of points with norm at most 2^{-k}, i.e., $T_k = \{x \in T : \|x\|_2 \le 2^{-k}\}$.

First, note that

$$A_t := \frac{1}{m} \sum_{i=1}^{m} \left((\Pi_i t)^2 - \mathbb{E}(\Pi_i t)^2 \right) = \frac{1}{m} \left(\|\Pi^0 t\|_2^2 - m \cdot \|t\|_2^2 \right) = \|\Pi t\|_2^2 - \|t\|_2^2.$$

For any $t \in T$ and $1 \le i \le m$, define $X_{t,i} = \Pi_i t$. Then, A_t corresponds to the definition in Theorem 2.2. Note that for any $s, t \in T$ and for any $1 \le i \le m$, $X_{s,i} - X_{t,i} = X_{s-t,i}$, which is a subgaussian with mean 0, variance $\|s - t\|_2^2$, and subgaussian norm at most $k\|s - t\|_2 C_1$ by Proposition 2.1. Therefore, if we define $d_{\psi_2}(s,t) = \max_{1 \le i \le m} \|X_{s,i} - X_{t,i}\|_{\psi_2}$, as in Theorem 2.2, $d_{\psi_2}(s,t) \le k\|s - t\|_2 C_1$. As a result, we have the following:

Proposition 3.1. $\overline{\Delta}_{\psi_2}(T_k) = O(2^{-k})$.

Proof. Since any point $t \in T_k$ has Euclidean norm at most 2^{-k}, the conclusion is immediate. □

We also note the following:

Proposition 3.2. $\gamma_2(T_k, \ell_2) = O(\sqrt{\log n})$, *where ℓ_2 represents the standard Euclidean distance.*

Proof. The Majorizing Measures theorem [Tal14] gives $\gamma_2(T_k, \ell_2) = \Theta\left(\mathbb{E}_g(\sup_{x \in T_k} \langle g, x \rangle)\right)$, where g is a d-dimensional vector of i.i.d. standard normal random variables. Also, $\mathbb{E}_g(\sup_{x \in T_k} \langle g, x \rangle) \le \mathbb{E}_g(\sup_{x \in T} \langle g, x \rangle)$ because $T_k \subset T$. However, since $T = \operatorname{conv}(X)$ and since $\langle g, x \rangle$ is a linear function of x, we have $\sup_{x \in T} \langle g, x \rangle = \sup_{x \in X} \langle g, x \rangle$. Since the x_i's are unit norm vectors, then each $g_i := \langle g, x_i \rangle$ is standard normal, which implies (even if they are dependent) that $\mathbb{E} \sup_i |g_i| = O(\sqrt{\log n})$. This means that $\gamma_2(T_k, \ell_2) = O(\mathbb{E}_g(\sup_{x \in T} \langle g, x \rangle)) = O(\sqrt{\log n})$, so the proof is complete. □

This allows us to bound $\gamma_2(T, d_{\psi_2})$ as follows.

Corollary 3.1. $\gamma_2(T_k, d_{\psi_2}) = O(\sqrt{\log n})$.

Proof. As $d_{\psi_2}(s,t) = O(\|s - t\|_2)$ for any points s, t, the result follows from Proposition 3.2. □

Therefore, using $T = T_k$ in Theorem 2.2 and varying k gives us the following.

Theorem 3.1. *Suppose that $\varepsilon, \delta < 1$ and $m = \Theta\left(\frac{1}{\varepsilon^2}\log\frac{n\log(2/\varepsilon)}{\delta}\right)$. Then, there exists some constant C_2 such that for all $k \geq 0$,*

$$\mathbb{P}\left(\sup_{t\in T_k}\left|\|\Pi t\|_2^2 - \|t\|_2^2\right| \geq C_2\left(\varepsilon^2 + \varepsilon\cdot 2^{-k}\right)\right) \leq \frac{\delta}{n\log(2/\varepsilon)}.$$

Consequently, with probability at least $1-\delta/n$, we have that for all $t \in T$,

$$|\|\Pi t\|_2 - \|t\|_2| = O(\varepsilon).$$

Proof. The first part follows from Theorem 2.2 and the previous propositions. Note that

$$\sup_{t\in T_k}\left|\|\Pi t\|_2^2 - \|t\|_2^2\right| = \sup_{t\in T_k}|A_t|.$$

Moreover, $\frac{1}{m}\gamma_2^2(T_k, d_{\psi_2}) = O(\varepsilon^2)$ and $\frac{1}{\sqrt{m}}\overline{\Delta}_{\psi_2}(T_k)\gamma_2(T_k, d_{\psi_2}) = O(\varepsilon\cdot 2^{-k})$. If we let $u = \ln\frac{n\log(2/\varepsilon)}{\delta}$, then $\sqrt{u}\frac{\overline{\Delta}^2_{\psi_2}(T_k)}{\sqrt{m}} = O(\varepsilon\cdot 2^{-2k})$, and $u\frac{\overline{\Delta}^2_{\psi_2}(T)}{m} = O(\varepsilon^2\cdot 2^{-2k})$. The first result now follows immediately from Theorem 2.2, as $e^{-u} = \frac{\delta}{n\log(2/\varepsilon)}$. Note that it is possible for T_k to be empty, but in this case we see that $\sup_{t\in T_k}|A_t| = 0$ so the first result is immediate.

For the second part, assume WLOG that $\varepsilon^{-1} = 2^\ell$ for some ℓ. Define $T'_0, T'_1, \ldots, T'_\ell$ such that $T'_\ell = T_\ell$ and for all $k < \ell$, $T'_k = T_k\backslash T_{k+1}$. Then, $T'_0, \ldots, T'_\ell$ forms a partition of T, since $\|x\|_2 \leq 1$ if $x \in T$. Note that if $T \in T_\ell$, then with probability at least $1 - \frac{\delta}{n\log(2/\varepsilon)}$, $\left|\|\Pi t\|_2^2 - \|t\|_2^2\right| = O(\varepsilon^2)$ and thus $\|\Pi t\|_2^2 = O(\varepsilon^2)$ and $|\|\Pi t\|_2 - \|t\|_2| = O(\varepsilon)$. If $T \in T'_k$ for some $k < \ell$, then with probability at least $1 - \frac{\delta}{n\log(2/\varepsilon)}$, $\left|\|\Pi t\|_2^2 - \|t\|_2^2\right| = |\|\Pi t\|_2 + \|t\|_2|\cdot|\|\Pi t\|_2 - \|t\|_2| = O(\varepsilon\cdot 2^{-k})$. This means that since $|\|\Pi t\|_2 + \|t\|_2| \geq 2^{-(k+1)}$, we must have that $|\|\Pi t\|_2 - \|t\|_2| = O(\varepsilon)$.

Therefore, by union bounding over all $0 \leq k \leq \ell$, with probability at least

$$1 - (\ell+1)\frac{\delta}{n\log(2/\varepsilon)} = 1 - \frac{(\ell+1)\delta}{n(\ell+1)} = 1 - \frac{\delta}{n},$$

for all $t \in T$, $|\|\Pi t\|_2 - \|t\|_2| = O(\varepsilon)$. Thus, we are done. ☐

We therefore have the following immediate corollary.

Corollary 3.2. *For $1 \leq \varepsilon^{-2} < n$ and for any $X = \{x_1, ..., x_n\} \subset S^{d-1}$, with probability at least $1 - poly(n)^{-1}$, a randomly chosen Π with $m = \Omega(\varepsilon^{-2}\log n)$ provides ε-convex hull distortion for X.*

3.2 Completion of the Terminal Embedding

We note that the methods for completing the terminal embedding in this section are very similar to those in [MMMR18]. Specifically, our proofs for Lemmas 3.1 and 3.2 are based on the proofs of [MMMR18, Lemma 3.1] and [MMMR18, Theorem 1.5], respectively.

Lemma 3.1. *Let $x_1, ..., x_n$ be nonzero points in $\mathbb{R}^d$ and let $v_i = \frac{x_i}{\|x_i\|_2}$. Suppose that Π provides ε-convex hull distortion for $V = \{v_1, -v_1, ..., v_n, -v_n\}$. Then, for any $u \in \mathbb{R}^d$, there exists a $u' \in \mathbb{R}^m$ such that $\|u'\|_2 \leq \|u\|_2$ and $|\langle u', \Pi x_i\rangle - \langle u, x_i\rangle| \leq \varepsilon\|u\|_2\cdot\|x_i\|_2$ for every x_i.*

Proof. This statement is trivial for for $u = 0$, so assume $u \neq 0$. It suffices to show that there always exists a u' such that $\|u'\|_2 \leq \|u\|_2$ and $|\langle u', \Pi v_i\rangle - \langle u, v_i\rangle| \leq \varepsilon\|u\|_2$ for all v_i. We will have that $|\langle u', \Pi x_i\rangle - \langle u, x_i\rangle| \leq \varepsilon\|u\|_2 \cdot \|x_i\|_2$ by scaling. Now, let B be the ball in $\mathbb{R}^m$ of radius $\|u\|_2$ and Λ be the the unit ℓ_1-ball in $\mathbb{R}^n$, where we write $\lambda \in \Lambda$ as $(\lambda_1, \ldots, \lambda_n)$. Furthermore, define for $u' \in B, \lambda \in \Lambda$,

$$\Phi(u', \lambda) := \sum_{i=1}^{n} \left(\lambda_i(\langle u, v_i\rangle - \langle u', \Pi v_i\rangle) - \varepsilon|\lambda_i| \cdot \|u\|_2\right).$$

We wish to show that there exists a $u' \in B$ such that for all $\lambda \in \Lambda$, $\Phi(u', \lambda) \leq 0$. This clearly suffices by looking at $\lambda = \pm v_j$ for $1 \leq j \leq n$.

Note that Φ is linear in u' and concave in λ, and that B, Λ are compact and convex. Then, by the von Neumann minimax theorem [vN28],

$$\min_{u' \in B} \max_{\lambda \in \Lambda} \Phi(u', \lambda) = \max_{\lambda \in \Lambda} \min_{u' \in B} \Phi(u', \lambda).$$

Thus it suffices to show the right hand side is nonpositive, i.e. for any u, λ, there exists $u' \in B$ s.t. $\Phi(u', \lambda) \leq 0$. Defining $P = \sum \lambda_i v_i$, then for $u' = \|u\|_2 \cdot \frac{\Pi P}{\|\Pi P\|_2}$, it suffices to show for all u, λ

$$\langle u, P\rangle - \langle u', \Pi P\rangle = \langle u, P\rangle - \|u\|_2 \cdot \|\Pi P\|_2 \leq \varepsilon\|\lambda\|_1\|u\|_2.$$

But in fact $\langle u, P\rangle \leq \|u\|_2 \cdot \|P\|_2$, so having that $\|P\|_2 - \|\Pi P\|_2 \leq \varepsilon\|\lambda\|_1$ for all $\lambda \in \Lambda$ is sufficient. For $\|\lambda\|_1 = 1$ this follows from Π providing ε-convex hull distortion for V, and for $\|\lambda\|_1 < 1$, it follows because we can scale λ so that $\|\lambda\|_1 = 1$. $\square$

Lemma 3.2. *Let $x_1, ..., x_n \in \mathbb{R}^d$ be distinct. Let $Y = \left\{\frac{x_i - x_j}{\|x_i - x_j\|_2} : i \neq j\right\}$. Moreover, suppose that $\Pi \in \mathbb{R}^{m \times d}$ provides ε-convex hull distortion for Y. Then, for any $u \in \mathbb{R}^d$, there exists an outer extension $f : \{x_1, \ldots, x_n, u\} \to \mathbb{R}^{m+1}$ with distortion $1 + O(\varepsilon)$, where $f(x_i) = \Pi x_i$.*

Proof. Note that the map $x \mapsto \Pi x$ yields a $1 + \varepsilon$-distortion embedding of $x_1, \ldots, x_n$ into $\mathbb{R}^m$ as it approximately preserves the norm of all points in Y. Therefore, we just have to verify that there exists $f(u) \in \mathbb{R}^{m+1}$ such that $\|f(u) - \Pi x_i\|_2 = (1 \pm O(\varepsilon))\|u - x_i\|_2$ for all i and any $u \in \mathbb{R}^d$. Fix $u \in \mathbb{R}^d$, and let x_k be the closest point to u among $\{x_1, \ldots, x_n\}$. By Lemma 3.1, there exists $u' \in \mathbb{R}^m$ such that $\|u'\|_2 \leq \|u - x_k\|_2$ and for all i,

$$|\langle u', \Pi(x_i - x_k)\rangle - \langle u - x_k, x_i - x_k\rangle| \leq \varepsilon\|u - x_k\|_2\|x_i - x_k\|_2.$$

Next, let $f(u) \in \mathbb{R}^{m+1}$ be the point $(\Pi x_k + u', \sqrt{\|u - x_k\|^2 - \|u'\|_2^2})$. If $w_i := x_i - x_k$, then

$$\|f(u) - f(x_i)\|_2^2 = \|u - x_k\|_2^2 - \|u'\|_2^2 + \|u' - \Pi w_i\|_2^2 = \|u - x_k\|_2^2 + \|\Pi w_i\|_2^2 - 2\langle u', \Pi w_i\rangle \tag{2}$$

and

$$\|u - x_i\|_2^2 = \|u - x_k\|_2^2 + \|w_i\|_2^2 - 2\langle u - x_k, w_i\rangle. \tag{3}$$

Since $\|u - x_i\|_2^2 \geq \|u - x_k\|_2^2$ and $\|u - x_i\|_2^2 \geq (\|w_i\|_2 - \|u - x_k\|_2)^2$, we have that $\|u - x_i\|_2^2 \geq (\|u - x_k\|_2^2 + \|w_i\|_2^2)/5$. This follows from the fact that $\max(1, (x-1)^2) \geq (x^2 + 1)/5$ for all $x \geq 0$. Since $\|\Pi(x_i - x_k)\|_2 = (1 \pm \varepsilon)\|x_i - x_k\|_2$, we also have that

$$\left|\|\Pi w_i\|_2^2 - \|w_i\|_2^2\right| = \left|\|\Pi(x_i - x_k)\|_2^2 - \|x_i - x_k\|_2^2\right| \leq 3\varepsilon\|x_i - x_k\|_2^2 = 3\varepsilon\|w_i\|_2^2$$

for all i, j, assuming $\varepsilon \leq 1$. Therefore, by subtracting Eq. (3) from Eq.(2), we have that

$$\left|\|f(u) - f(x_i)\|_2^2 - \|u - x_i\|_2^2\right| \leq 3\varepsilon\|w_i\|_2^2 + 2\left|\langle u', \Pi w_i\rangle - \langle u - x_k, w_i\rangle\right|$$

$$\leq 3\varepsilon\|w_i\|_2^2 + 2\varepsilon\|u - x_k\|_2\|w_i\|_2 \leq 4\varepsilon\left(\|w_i\|_2^2 + \|u - x_k\|_2^2\right) \leq 20\varepsilon\|u - x_i\|_2^2,$$

as desired. ☐

We summarize the previous results, which allows us to prove our main result.

Theorem 3.2. *For all $\varepsilon < 1$ and $x_1, \ldots, x_n \in \mathbb{R}^d$, there exists $m = O(\varepsilon^{-2}\log n)$ and a (nonlinear) map $f : \mathbb{R}^d \to \mathbb{R}^m$ such that for all $1 \leq i \leq n$ and $u \in \mathbb{R}^d$,*

$$(1 - O(\varepsilon))\|u - x_i\|_2 \leq \|f(u) - f(x_i)\|_2 \leq (1 + O(\varepsilon))\|u - x_i\|_2.$$

Proof. By Corollary 3.2, there exists a $\Pi \in \mathbb{R}^{m \times d}$ with $m = \Theta(\varepsilon^{-2}\log n)$ that provides ε-convex hull distortion for Y, where Y is defined in Lemma 3.2. By Lemma 3.2, for each $u \in \mathbb{R}^d$ there exists an outer extension $f^{(u)} : \{x_1, \ldots, x_n, u\} \to \mathbb{R}^{m+1}$ with distortion $1 + O(\varepsilon)$ sending x_i to $(\Pi x_i, 0)$. Therefore, we map $x_i \mapsto (\Pi x_i, 0)$ and map each $u \notin \{x_1, ..., x_n\}$ to $f^{(u)}(u)$, which gives us a terminal embedding to $m = O(\varepsilon^{-2}\log n)$ dimensions with distortion $1 + O(\varepsilon)$. ☐

3.3 Algorithm to Construct Terminal Embedding

We briefly note how one can produce a terminal embedding into $\mathbb{R}^m$ with a Monte Carlo randomized polynomial time algorithm, where $m = O(\varepsilon^{-2}\log n)$. By choosing a random Π from Subsection 3.1, we get with at least $1 - n^{-\Theta(1)}$ probability a matrix providing ε-convex hull distortion for our set Y in Lemma 3.2. To map any point in $\mathbb{R}^d$ into $\mathbb{R}^{m+1}$ dimensions, for any $u \in \mathbb{R}^d$, it suffices to find a $u' \in \mathbb{R}^m$ such that if x_k is the point in X closest to u, $||u'||_2 \leq ||u - x_k||_2$ and for all i,

$$|\langle u', \Pi(x_i - x_k)\rangle - \langle u - x_k, x_i - x_k\rangle| \leq \varepsilon\|u - x_k\|_2\|x_i - x_k\|_2.$$

Assuming that Π provides an ε-convex hull distortion for Y, such a u' exists for all u, which means that u' can be found with semidefinite programming in polynomial time, as noted in [MMMR18].

References

[AK17] Noga Alon and Bo'az Klartag. Optimal compression of approximate inner products and dimension reduction. In *Proceedings of the 58th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 639–650, 2017.

[BLM13] Stephane Boucheron, Gabor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 1st edition, 2013.

[Dir15] Sjoerd Dirksen. Tail bounds via generic chaining. *Electron. J. Probab.*, 20(53):1–29, 2015.

[Dir16] Sjoerd Dirksen. Dimensionality reduction with subgaussian matrices: a unified theory. *Found. Comp. Math.*, 16(5):1367–1396, 2016.

[EFN17] Michael Elkin, Arnold Filtser, and Ofer Neiman. Terminal embeddings. *Theoretical Computer Science*, 697:1–36, 2017.

[JL84] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[LN17] Kasper Green Larsen and Jelani Nelson. Optimality of the Johnson-Lindenstrauss lemma. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2017.

[MMMR18] Sepideh Mahabadi, Konstantin Makarychev, Yury Makarychev, and Ilya P. Razenshteyn. Nonlinear dimension reduction via outer Bi-Lipschitz extensions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1088–1101, 2018.

[Tal14] Michel Talagrand. *Upper and lower bounds for stochastic processes: modern methods and classical problems*. Springer, 2014.

[Ver18] Roman Vershynin. *High-dimensional probability: an introduction with applications in data science*. Cambridge University Press, 2018.

[vN28] John von Neumann. Zur theorie der Gesellschaftsspiele. *Math. Ann.*, 100:295–320, 1928.