

# Embeddings of Surfaces, Curves, and Moving Points in Euclidean Space\*

Pankaj K. Agarwal<sup>†</sup>      Sariel Har-Peled<sup>‡</sup>      Hai Yu<sup>§</sup>

July 15, 2008

## Abstract

In this paper we show that dimensionality reduction (i.e., Johnson-Lindenstrauss lemma) preserves not only the distances between static points, but also between moving points, and more generally between low-dimensional flats, polynomial curves, curves with low winding degree, and polynomial surfaces. We also show that surfaces with bounded doubling dimension can be embedded into low dimension with small additive error. Finally, we show that for points with polynomial motion, the radius of the smallest enclosing ball can be preserved under dimensionality reduction.

## 1 Introduction

In recent years there has been a flurry of research on algorithmic problems dealing with high-dimensional data sets such as text and image data [Kle97, IM98, KOR00]. Dimensionality reduction is a powerful tool in tackling these problems, as it overcomes the challenge of high dimensionality by mapping the data into a low-dimensional space while preserving essential properties of the data, thus enabling the problem to be solved in that low-dimensional space. There are several effective dimensionality reduction methods such as singular value decomposition (SVD) and random projections. They are widely used for data compression, information retrieval, machine learning, and many other applications.

The *random-projection method* is a popular approach for dimensionality reduction. In this method, one chooses a random low-dimensional linear subspace and projects the data onto that subspace. The Johnson-Lindenstrauss flattening lemma [JL84] states that pairwise distances are  $(1+\varepsilon)$ -preserved when one projects a set  $P$  of  $n$  points into a random linear subspace of  $O(\varepsilon^{-2} \log n)$  dimensions. The dimension of the host space cannot be (significantly)

---

\*P.A. and H.Y. are supported by NSF under grants CCR-00-86013, EIA-01-31905, CCR-02-04118, and DEB-04-25465, by an ARO grant W911NF-04-1-0278, and by a grant from the U.S.–Israel Binational Science Foundation. S.H.-P. is supported by a NSF CAREER award CCR-0132901.

<sup>†</sup>Department of Computer Science, Duke University, Durham, NC 27708, USA; [pankaj@cs.duke.edu](mailto:pankaj@cs.duke.edu)

<sup>‡</sup>Department of Computer Science, University of Illinois, Urbana, IL 61801, USA; [sariel@uiuc.edu](mailto:sariel@uiuc.edu)

<sup>§</sup>Department of Computer Science, Duke University, Durham, NC 27708, USA; [fishhai@cs.duke.edu](mailto:fishhai@cs.duke.edu)

further reduced in general [Alo03]. Subsequently, it has been shown that the random projection process can be performed with different projection matrices [Ach01, DG03, FM88, IM98], which either simplifies the proof of the JL lemma or is computationally more convenient.

The random-projection method had wide success in solving high-dimensional algorithmic problems, including nearest neighbor search, learning Gaussian mixtures, clustering, machine learning, and processing image and text data [Das99, AV99, BM01]. See also the book [Vem04] and the surveys [Lin02, Ind01, IM04]. Its versatility raises the possibility that random projection preserves much more information besides pairwise distances. In particular, Magen [Mag02] showed that it preserves angles induced by subsets of points of  $P$ .

In this paper, we investigate what further information is preserved under random projections. Consider the problem of dimensionality reduction for high-dimensional time-varying data. This problem arises in a number of practical applications, such as video databases and complex motion databases (e.g., human motion or protein folding process, which has many degrees of freedom). The disadvantage of using SVD in dimensionality reduction for high-dimensional time-varying data is that if one performs SVD on each snapshot of the data independently, the resulting low-dimensional data may lose the inherent smoothness of the original data along the time dimension. There has been work on how to compute a smooth SVD for time-varying matrices [BH03, DE99]; see also [BN03, HPP01] for other similar spectral analytic approaches for reducing the dimensionality of time-varying data. However, they involve expensive analytical machineries such as solving complex differential equations. In contrast, the random projection method applied to time-varying data appears much simpler and more intuitive, and naturally retains the smoothness of the data. However, while it is well-known that this method works for static data, it is not immediately clear whether it still works for time-varying data. This is because a projection that works for the initial configuration of the data does not necessarily provide a low-distortion embedding for future configurations. A natural question is whether there exist a low-dimensional linear subspace  $\mathfrak{h} \subseteq \mathbb{R}^d$  such that the projection of the data onto  $\mathfrak{h}$  is always a low-distortion embedding at all times. In this paper we answer this question in the affirmative. We also study the problem of projecting a family of curves and surfaces so that pairwise distances between points on them are preserved.

**Our results.** We first show in Section 3 that random projection preserves the structure of curves, namely, for any two points on these curves, their distances are preserved. In fact, we prove an even more general result: given a collection of  $n$  surfaces of linearization dimension  $\tau$  (see Section 2 for the definitions), if one embeds them into  $O(\tau\varepsilon^{-2} \log(n\tau/\varepsilon))$  dimensions, the projected surfaces preserve their structure in the sense that for any pair of points on these surfaces their distances are preserved. A similar observation for the linear case is implicit in the work of Magen [Mag02] and explicit in the work of Sarlós [Sar06]. This observation indicates that the JL lemma and dimensionality reduction do not only work for points, but also work for low-dimensional curves, flats or surfaces, and therefore preserve much more structure than was realized before.

In Section 4, we show that one can embed a surface of low doubling dimension into low dimension so that the distance between any pair of points on the surface is preserved with an *additive* error. Intuitively, a surface has a low doubling dimension if its geodesic metric

behaves like a low dimensional metric. Baraniuk and Wakin [BW06] showed a stronger result (with multiplicative error) but under considerably stronger assumptions

Next, we study the problem of dimensionality reduction for high-dimensional curves whose central projection on the unit sphere, centered at the origin, has (finite) length  $\omega$ . We show that for  $k = O(\varepsilon^{-2} \log(\omega/\varepsilon))$ , the projection of such a curve into a random  $k$ -dimensional linear subspace preserves pairwise distances of the points on the curve. This result implies an extension of the JL lemma to high-dimensional moving points: for  $k = O(\varepsilon^{-2} \log(n\omega/\varepsilon))$ , the projection of a set of  $n$  moving points in  $\mathbb{R}^d$  into a random (but fixed)  $k$ -dimensional linear subspace preserves pairwise distances of the moving points at *all* times. Note that even if  $\omega$  is polynomial in  $n$ , the above bound is still about  $O(\varepsilon^{-2} \log(n/\varepsilon))$ , which is close to the bound in the JL lemma.

Finally, in Section 6 we show that, besides preserving pairwise distances of the moving points, random projection is able to preserve the radius of the smallest enclosing ball of the moving points at *every* moment of the motion. While a similar result holds easily for random projections of static point sets, more careful arguments that borrow ideas from coresets [BC03] are needed to establish this fact for embeddings of moving points. This result is useful for certain applications in which, in addition to preserving pairwise distances, one may also wish to preserve the radius of each cluster in the time-varying data, so as to retain meaningful clustering information when projecting high-dimensional data into a low-dimensional space.

## 2 Preliminaries

**Embeddings.** We use  $\|\cdot\|$  to denote the  $L_2$ -norm throughout the paper. Let  $P$  be a set of points in  $\mathbb{R}^d$ . A map  $f : P \rightarrow \mathbb{R}^k$  is called a  $D$ -embedding (or *embedding with distortion  $D$* ) of  $P$  if there is a number  $c > 0$ , referred to as the *scaling parameter* of the embedding, such that for every  $p, q \in P$ ,

$$\|p - q\| \leq c \cdot \|f(p) - f(q)\| \leq D \cdot \|p - q\|.$$

We call  $f$  a  $D$ -draft of  $P$  if for every  $p \in P$ ,

$$\|p\| \leq c \cdot \|f(p)\| \leq D \cdot \|p\|.$$

Namely, a  $D$ -draft preserves the distances of the points in  $P$  to the origin. Note that a linear  $D$ -draft of  $\{p - q \mid p, q \in P\}$  is a  $D$ -embedding of  $P$ .

Let  $\mathfrak{h}$  be a linear subspace of  $\mathbb{R}^d$ . Let  $\downarrow_{\mathfrak{h}} : \mathbb{R}^d \rightarrow \mathfrak{h}$  denote the projection of  $\mathbb{R}^d$  into  $\mathfrak{h}$ . For a subset  $\mathbb{X} \subseteq \mathbb{R}^d$ , we use  $\downarrow_{\mathfrak{h}}(\mathbb{X})$  to denote the image of  $\mathbb{X}$  in  $\mathfrak{h}$ .

**Lemma 2.1 (JL Lemma, [JL84])** *Let  $P$  be a set of  $n$  points in  $\mathbb{R}^d$ , let  $k \leq d$  be an integer, let  $0 < \varepsilon < 1$  be a parameter, and let  $\mathfrak{h}$  be a random linear subspace of dimension  $k$ . Then  $\downarrow_{\mathfrak{h}}(P)$  is a  $(1 + \varepsilon)$ -embedding of  $P$  with probability at least  $1 - n^2 \exp(-\varepsilon^2 k / c_{\text{JL}})$ , where  $c_{\text{JL}}$  is a constant. Alternatively, for any  $\beta > 1$ , with probability at least  $1 - n^2 \exp(-(\beta - 1)^2 k / c_{\text{JL}})$ , the following holds for every pair  $p, q \in P$ :*

$$\|p - q\| \leq c \|\downarrow_{\mathfrak{h}}(p) - \downarrow_{\mathfrak{h}}(q)\| \leq \beta \|p - q\|,$$

where  $c$  is a scaling parameter.

The JL Lemma implies that for  $k = O(\varepsilon^{-2} \log n)$ , the projection into a random linear subspace of dimension  $k$  is a  $(1+\varepsilon)$ -embedding of  $P$  with high probability. We now introduce the notion of embeddings of moving points.

**Definition 2.2 (Embedding of moving points)** Let  $P$  be a set of points moving in  $\mathbb{R}^d$ , and let  $S$  be another set of points moving in  $\mathbb{R}^k$ . A bijection  $f : P \rightarrow S$  is called a *D-embedding* of  $P$  if there is a scaling parameter  $c > 0$  such that for every  $p, q \in P$  and for every  $t \in \mathbb{R}$ ,

$$\|p(t) - q(t)\| \leq c \cdot \|\tilde{p}(t) - \tilde{q}(t)\| \leq D \cdot \|p(t) - q(t)\|,$$

where  $\tilde{p} = f(p)$  and  $\tilde{q} = f(q)$ .

**Curves and winding degree.** In the following, let  $\mathbb{S}^{d-1}$  denote the unit sphere in  $\mathbb{R}^d$ . For a set  $\mathbb{X} \subseteq \mathbb{R}^d$ , let  $\mathbf{S}(\mathbb{X}) \subseteq \mathbb{S}^{d-1}$  denote the *central projection* of  $\mathbb{X}$  on  $\mathbb{S}^{d-1}$ , i.e.,  $\mathbf{S}(\mathbb{X}) = \left\{ x/\|x\| \mid x \in \mathbb{X}, x \neq 0 \right\}$ . Let  $\xi : \mathbb{R} \rightarrow \mathbb{R}^d$  be a curve in  $\mathbb{R}^d$ , and let  $|\xi|$  denote its arclength. We refer to  $|\mathbf{S}(\xi)|$  as the *winding degree* of a curve  $\xi$ .

A curve  $\xi(t) = (\xi_1(t), \dots, \xi_d(t))$ , for  $t \in \mathbb{R}$ , is called a *polynomial curve of degree  $\omega$*  if each  $\xi_i$  is a polynomial of  $t$  of degree at most  $\omega$ . A variant of the well-known Cauchy-Crofton formula [San53] states that, for any curve  $\xi$  in  $\mathbb{R}^d$ ,

$$|\mathbf{S}(\xi)| = \pi \int_{\mathbf{h} \in \mathbb{L}} \#(\xi \cap \mathbf{h}) \, d\mathbf{h},$$

where  $\mathbb{L}$  is the set of all linear hyperplanes in  $\mathbb{R}^d$  and  $\#(\xi \cap \mathbf{h})$  counts the number of intersections between  $\xi$  and  $\mathbf{h} \in \mathbb{L}$ . Since every linear hyperplane  $\mathbf{h} \subset \mathbb{R}^d$  intersects a polynomial curve  $\xi$  of degree  $\omega$  at most  $\omega$  times, the winding degree of  $\xi$  is at most  $\pi\omega$ . Hence, the winding degree can be viewed as an extension of polynomial degree.

Let  $P$  be a set of moving points in  $\mathbb{R}^d$ . The motion of  $P$  is *polynomial of degree  $\omega$*  if the trajectory of each point in  $P$  is a polynomial curve of degree at most  $\omega$ . The motion of  $P$  has *winding degree  $\omega$*  if for each pair  $p, q \in P$ , the winding degree of the curve  $p(t) - q(t)$  is at most  $\omega$ . That is, from each point's viewpoint, the trajectory of every other point is a curve of winding degree at most  $\omega$ . Motion arising in many real-world applications has a small winding degree.

**Surfaces, linearization, and doubling dimension.** Some of our results on curves and moving points extend to the more general surfaces. A  *$\psi$ -dimensional (parametric) surface* is a map  $\mathbb{M} : \mathbb{R}^\psi \rightarrow \mathbb{R}^d$ , where  $\mathbb{M}(x) = \{p_1(x), \dots, p_d(x)\}$  and each  $p_i$  is a  $\psi$ -variate function  $p_i : \mathbb{R}^\psi \rightarrow \mathbb{R}$ . We say that  $\mathbb{M}$  is a *polynomial surface of degree  $\omega$*  if each  $p_i$  is a polynomial of degree at most  $\omega$ .

Let  $d_{\mathbb{M}}(\cdot, \cdot)$  denote the *geodesic distance* on  $\mathbb{M}$ . For a point  $p \in \mathbb{M}$  and a number  $r \geq 0$ , let  $\mathbb{B}_{\mathbb{M}}(p, r) = \left\{ q \in \mathbb{M} \mid d_{\mathbb{M}}(p, q) \leq r \right\}$  denote the (geodesic) ball of radius  $r$  centered at  $p$  on  $\mathbb{M}$ . The *doubling constant*  $\lambda$  of  $\mathbb{M}$ , defined as the minimum over  $m \in \mathbb{N}$  such that every ball in  $\mathbb{M}$  can be covered by at most  $m$  balls of at most half the radius (note that the balls are, again, geodesic balls). The *doubling dimension* of  $\mathbb{M}$  is defined as  $\tau = \lceil \log_2 \lambda \rceil$ . Intuitively,

the doubling dimension is an extension of the standard Euclidean dimension, and a surface with low doubling dimension can be thought of as having low dimension.

Let  $\mathcal{F} = \{f_1(x), \dots, f_n(x)\}$  be a family of  $d$ -variate functions over  $x \in \mathbb{R}^d$ . Suppose we can express each  $f_i(x)$  as

$$f_i(x) = v_i^{(0)} + v_i^{(1)}\varphi_1(x) + \dots + v_i^{(\tau)}\varphi_\tau(x),$$

where  $v_i^{(0)}, \dots, v_i^{(\tau)}$  are constants and  $\varphi_1(x), \dots, \varphi_\tau(x)$  are  $d$ -variate functions over  $x \in \mathbb{R}^d$ . We call the map  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^\tau$ , where  $\varphi(x) = (\varphi_1(x), \dots, \varphi_\tau(x))$ , a *linearization* of  $\mathcal{F}$ , as each  $f_i \in \mathcal{F}$  maps to a  $\tau$ -variate linear function

$$h_i(y_1, \dots, y_\tau) = v_i^{(0)} + v_i^{(1)}y_1 + \dots + v_i^{(\tau)}y_\tau,$$

in the sense that for any  $x \in \mathbb{R}^d$ ,  $f_i(x) = h_i(\varphi(x))$ . We refer to  $\tau$  as the *linearization dimension* of  $\mathcal{F}$ , and say that  $\mathcal{F}$  admits a linearization of dimension  $\tau$ . Agarwal and Matoušek [AM94] described an algorithm that computes a linearization of the smallest dimension under certain weak assumptions. We say that a surface  $\mathbb{M}(x) = (p_1(x), \dots, p_d(x))$  has linearization dimension  $\tau$  if the set  $\Pi(\mathbb{M}) = \{p_1, \dots, p_d\}$  of functions admits a linearization of dimension  $\tau$ . We say that a family of surfaces  $\mathbb{M}_1, \dots, \mathbb{M}_n$  has (combined) linearization dimension  $\tau$  if the set  $\bigcup_{1 \leq i \leq n} \Pi(\mathbb{M}_i)$  admits a linearization of dimension  $\tau$ .

**Lemma 2.3** *If a surface  $\mathbb{M}$  has linearization dimension  $\tau$ , then  $\mathbb{M}$  is contained inside an affine subspace of dimension  $\tau$ .*

*Proof:* Let  $\mathbb{M}(x) = (p_1(x), \dots, p_d(x))$  be a surface with linearization dimension  $\tau$  in  $\mathbb{R}^d$ . Then by definition there is a map  $\varphi(x) = (\varphi_1(x), \dots, \varphi_\tau(x))$  such that each  $p_i$ , for  $1 \leq i \leq d$ , can be rewritten as a linear function  $h_i(y_1, \dots, y_\tau) = v_i^{(0)} + \sum_{j=1}^{\tau} v_i^{(j)} \cdot y_j$  in the sense that for all  $x$ ,  $p_i(x) = h_i(\varphi(x))$ . Let  $v_j = (v_1^{(j)}, \dots, v_d^{(j)})$  be a vector in  $\mathbb{R}^d$ , for  $0 \leq j \leq \tau$ . Clearly, for every  $x$ ,

$$\mathbb{M}(x) = (p_1(x), \dots, p_d(x)) = v_0 + \sum_{j=1}^{\tau} v_j \cdot \varphi_j(x) \in v_0 + \text{span}(v_1, \dots, v_\tau),$$

thereby implying that  $\mathbb{M}$  is contained in an affine subspace of dimension  $\tau$ . ■

**Definition 2.4** Let  $\mathbb{M} : \mathbb{R}^\psi \rightarrow \mathbb{R}^d$  be a  $\psi$ -dimensional surface in  $\mathbb{R}^d$ . A surface  $\mathbb{U} : \mathbb{R}^\psi \rightarrow \mathbb{R}^k$  is a *D-draft* of  $\mathbb{M}$  if there exists a scaling parameter  $c > 0$  such that for all  $x \in \mathbb{R}^\psi$ ,

$$\|\mathbb{M}(x)\| \leq c \cdot \|\mathbb{U}(x)\| \leq D \cdot \|\mathbb{M}(x)\|.$$

As in the JL lemma, we can embed a surface  $\mathbb{M} \subset \mathbb{R}^d$  into  $\mathbb{R}^k$  by randomly choosing a linear subspace  $\mathfrak{h}$  of  $\mathbb{R}^d$  and then projecting  $\mathbb{M}$  onto  $\mathfrak{h}$ . The following lemma states that we can focus on the distortion of  $\mathbb{S}(\mathbb{M})$  when we embed  $\mathbb{M}$  via  $\downarrow_{\mathfrak{h}}$ .

**Lemma 2.5** *Let  $\mathfrak{h} \subseteq \mathbb{R}^d$  be a linear subspace and  $\mathbb{M}$  be a surface in  $\mathbb{R}^d$ . Then  $\downarrow_{\mathfrak{h}}(\mathbb{M})$  is a *D-draft* of  $\mathbb{M}$  if and only if  $\downarrow_{\mathfrak{h}}(\mathbb{S}(\mathbb{M}))$  is a *D-draft* of  $\mathbb{S}(\mathbb{M})$ .*

*Proof:* Since  $\downarrow_{\mathbf{h}}$  is a linear operator, for any point  $p \in \mathbb{M}$ ,

$$\frac{\downarrow_{\mathbf{h}}(p)}{\|p\|} = \downarrow_{\mathbf{h}}\left(\frac{p}{\|p\|}\right) = \downarrow_{\mathbf{h}}(\mathbf{S}(p)). \quad (1)$$

Now, if  $\downarrow_{\mathbf{h}}(\mathbb{M})$  is a  $D$ -draft of  $\mathbb{M}$ , then there is a constant  $c$ , such that for any point  $p \in \mathbb{M}$ , we have  $\|p\| \leq c \|\downarrow_{\mathbf{h}}(p)\| \leq D \|p\|$ , which is equivalent to  $1 \leq \left\| \downarrow_{\mathbf{h}}(\mathbf{S}(p)) \right\| \leq D$  by (1). Since  $\|\mathbf{S}(p)\| = 1$ , we obtain

$$\|\mathbf{S}(p)\| \leq c \left\| \downarrow_{\mathbf{h}}(\mathbf{S}(p)) \right\| \leq D \|\mathbf{S}(p)\|,$$

as required. ■

### 3 Embedding of Surfaces with Bounded Linearization Dimensions

In this section, we prove embedding bounds for surfaces with bounded linearization dimension, which can be used to obtain a weaker bound on the embedding of moving points. Let  $e_i$  be the intersection point of  $\mathbb{S}^{d-1}$  with the  $(+x_i)$ -axis, i.e., the unit vector in direction along the  $(+x_i)$ -axis. For  $k \leq d$ , we define the  $k$ -diamond, denoted by  $\diamond_k$ , to be  $\text{conv}(\pm e_1, \dots, \pm e_k)$ , which is the “unit ball” in  $\mathbb{R}^k$  centered at the origin under the  $L_1$ -norm. Note that  $\diamond_k$  contains a (Euclidean) ball of radius  $1/\sqrt{k}$  centered at the origin. In our applications, we will be working in a random  $k$ -dimensional linear subspace  $\mathbf{h}$  (instead of  $\mathbb{R}^k$ ). We assume that we have an orthonormal basis in  $\mathbf{h}$ , and  $\diamond_k$  is defined with respect to this basis.

**Lemma 3.1** *Fix a positive integer  $k \leq d$ . Let  $p_0 \in \mathbb{S}^{d-1}$ , let  $\sigma = p_0 + \varepsilon \cdot \diamond_k$ , where  $\varepsilon \leq 1/10$  is a parameter, and let  $\mathbf{h}$  be a linear subspace in  $\mathbb{R}^d$ . Let  $V = \{e_1, \dots, e_k\} \cup \{p_0\}$ . If  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon)$ -draft of  $V$  (with scaling parameter  $c$ ), then  $\downarrow_{\mathbf{h}}$  is a  $(1 + 8\varepsilon)$ -draft of  $\sigma$  (with scaling parameter  $c/(1 - 2\varepsilon)$ ).*

*Proof:* Let  $p_1, \dots, p_{2k}$  be the vertices of  $\sigma$ . Let  $p$  be any point of  $\sigma$ . By convexity, we can write

$$\begin{aligned} c \|\downarrow_{\mathbf{h}}(p_0 - p)\| &\leq \max(c \|\downarrow_{\mathbf{h}}(p_0 - p_1)\|, \dots, c \|\downarrow_{\mathbf{h}}(p_0 - p_{2k})\|) \\ &= \varepsilon \cdot \max(c \|\downarrow_{\mathbf{h}}(e_1)\|, \dots, c \|\downarrow_{\mathbf{h}}(e_k)\|) \leq (1 + \varepsilon)\varepsilon. \end{aligned}$$

Hence, by the triangle inequality,

$$c \|\downarrow_{\mathbf{h}}(p)\| \geq c \|\downarrow_{\mathbf{h}}(p_0)\| - c \|\downarrow_{\mathbf{h}}(p_0 - p)\| \geq \|p_0\| - (1 + \varepsilon)\varepsilon.$$

Similarly,

$$c \|\downarrow_{\mathbf{h}}(p)\| \leq c \|\downarrow_{\mathbf{h}}(p_0)\| + c \|\downarrow_{\mathbf{h}}(p_0 - p)\| \leq (1 + \varepsilon) \|p_0\| + (1 + \varepsilon)\varepsilon.$$

Furthermore, observe that  $\|p_0\| - \varepsilon \leq \|p\| \leq \|p_0\| + \varepsilon$ . As such, we obtain

$$c \|\downarrow_{\mathbf{h}}(p)\| = \frac{c \|\downarrow_{\mathbf{h}}(p)\|}{\|p\|} \|p\| \geq \frac{\|p_0\| - (1 + \varepsilon)\varepsilon}{\|p_0\| + \varepsilon} \|p\| \geq (1 - 2\varepsilon) \|p\|,$$

and

$$c \|\downarrow_{\mathbf{h}}(p)\| = \frac{c \|\downarrow_{\mathbf{h}}(p)\|}{\|p\|} \|p\| \leq \frac{(1 + \varepsilon) \|p_0\| + (1 + \varepsilon)\varepsilon}{\|p_0\| - \varepsilon} \cdot \|p\| \leq (1 + 4\varepsilon) \|p\|,$$

since  $\|p_0\| = 1$  and  $\varepsilon \leq 1/10$ . This implies that the mapping is a  $D$ -draft of  $p$  for

$$D = \frac{1 + 4\varepsilon}{1 - 2\varepsilon} \leq 1 + 8\varepsilon$$

with scaling parameter  $c/(1 - 2\varepsilon)$ . ■

Let  $\mathbb{X} \subseteq \mathbb{R}^d$ . For a parameter  $\alpha > 0$ , a set  $N \subseteq \mathbb{X}$  is an  $\alpha$ -net of  $\mathbb{X}$  if for any  $x \in \mathbb{X}$ , there exists  $y \in N$ , such that  $\|x - y\| \leq \alpha$ .

**Lemma 3.2** *Let  $\mathbb{M} \subseteq \mathbb{R}^d$  be a surface with linearization dimension  $\tau$ . Let  $\varepsilon, \delta > 0$  be parameters, let  $k = O(\tau\varepsilon^{-2} \log(\tau/\varepsilon\delta))$ , and let  $\mathbf{h}$  be a random linear subspace of dimension  $k$ . Then  $\downarrow_{\mathbf{h}}(\mathbb{M})$  is a  $(1 + \varepsilon)$ -draft as well as a  $(1 + \varepsilon)$ -embedding of  $\mathbb{M}$  with probability at least  $1 - \delta$ .*

*Proof:* Since  $\mathbb{M}$  is a surface with linearization dimension  $\tau$ , by Lemma 2.3, there exists a linear subspace  $\mathbf{g} \subseteq \mathbb{R}^d$  of dimension  $\tau + 1$  that contains  $\mathbb{M}$  (if  $\tau + 1 \geq d$ , then  $\mathbf{h}$  is  $\mathbb{R}^d$  itself). The central projection  $\mathbf{S}(\mathbb{M})$  of  $\mathbb{M}$  is contained in  $\mathbb{X} = \mathbf{g} \cap \mathbb{S}^{d-1}$ . Note that  $\mathbb{X}$  is a  $\tau$ -hypersphere. Let  $N$  be an  $\alpha$ -net of  $\mathbb{X}$ , where  $\alpha = \varepsilon/(8\sqrt{\tau + 1})$ ;  $|N| = O((4/\alpha)^\tau)$  (see, e.g., [Mat02]). The balls of radius  $\alpha$  centered at the points of  $N$  cover  $\mathbb{X}$ . In particular, the set

$$\Sigma = \left\{ \sigma_p = p + (\varepsilon/8) \cdot \diamond_{(\tau+1)} \mid p \in N \right\}$$

covers  $\mathbb{X}$ .

Next we let  $P = N \cup \{v_1, \dots, v_{\tau+1}\}$ , where  $v_1, \dots, v_{\tau+1}$  are the intersection points of  $\mathbb{S}^{d-1}$  with an orthonormal basis in  $\mathbf{g}$  that is being used to define  $\diamond_{\tau+1}$ . Clearly,  $|P| = (\tau/\varepsilon)^{O(\tau)}$ . By the JL Lemma, for  $k = O(\varepsilon^{-2} \ln(|P|/\delta)) = O(\varepsilon^{-2} \tau \ln(\tau/\varepsilon\delta))$ , the projection into a  $k$ -dimensional random linear subspace  $\mathbf{h}$  is a  $(1 + \varepsilon/8)$ -draft of  $P$  with probability at least  $1 - \delta$ . By Lemma 3.1, the projection  $\downarrow_{\mathbf{h}}$  is then a  $(1 + \varepsilon)$ -draft of  $\mathbb{X}$ . Since  $\mathbf{S}(\mathbb{M}) \subseteq \mathbb{X}$ , Lemma 2.5 implies that  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon)$ -draft of  $\mathbb{M}$  as well as  $\mathbf{g}$ .

Moreover, observe that  $p - q \in \mathbf{g}$  for any pair of points  $p, q \in \mathbb{M} \subseteq \mathbf{g}$ . Since the projection into  $\mathbf{h}$  is a  $(1 + \varepsilon)$ -draft of all points in  $\mathbf{g}$ , it is also a  $(1 + \varepsilon)$ -draft of  $\{p - q \mid p, q \in \mathbb{M}\}$ , thereby implying that this projection is a  $(1 + \varepsilon)$ -embedding of  $\mathbb{M}$  as well. ■

Next, let  $\mathbb{M}_1$  and  $\mathbb{M}_2$  be two surfaces that have (combined) linearization dimension  $\tau$ . Then by definition, the surface  $\mathbb{M}$  defined as  $\mathbb{M}(x, y) = \mathbb{M}_1(x) - \mathbb{M}_2(y)$  also has linearization dimension  $\tau$ . By Lemma 3.2, the random projection into  $k$  dimensions is a  $(1 + \varepsilon)$ -draft of  $\mathbb{M}$  with probability at least  $1 - \delta$ , for  $k = O(\tau\varepsilon^{-2} \log(\tau/\varepsilon\delta))$ . In other words, not only this projection preserves the pairwise distances within  $\mathbb{M}_1$  and  $\mathbb{M}_2$ , but it also preserves the pairwise distances *between* points of  $\mathbb{M}_1$  and  $\mathbb{M}_2$ . This claim can be easily extended to  $n$  surfaces. We thus summarize:

**Theorem 3.3** *Let  $\mathbb{M}_1, \dots, \mathbb{M}_n$  be  $n$  surfaces with (combined) linearization dimension  $\tau$ . Let  $\varepsilon, \delta > 0$  be parameters, let  $k = O(\tau\varepsilon^{-2} \ln(n\tau/\varepsilon\delta))$ , and let  $\mathbf{h}$  be a random linear subspace of dimension  $k$ . Then  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon)$ -embedding of the set  $\bigcup_{i=1}^n \mathbb{M}_i$  with probability at least  $1 - \delta$ .*

Let  $P$  be a set of  $n$  points moving in  $\mathbb{R}^d$  with polynomial motion of degree  $\omega$ . For every pair of points  $p, q \in P$ , the point  $p(t) - q(t)$  traces a curve with linearization dimension  $\omega$ . Using Theorem 3.3, we then obtain that a projection into a random linear subspace of dimension  $k = O(\omega\varepsilon^{-2} \ln(n\omega/\varepsilon\delta))$  is a  $(1 + \varepsilon)$ -embedding of  $P$  with probability at least  $1 - \delta$ . This result will be further generalized and improved in Section 5.

## 4 Embedding of Surfaces with Bounded Doubling Dimensions

The following theorem shows that for a surface with bounded doubling dimension and geodesic diameter, one can embed it into low dimensions and preserve distances with additive error.

**Theorem 4.1** *Let  $\mathbb{M} \subseteq \mathbb{R}^d$  be a surface with doubling dimension  $\tau$  and geodesic diameter  $\Delta$ , let  $\varepsilon, \delta > 0$  be parameters, let  $k = O(\tau\varepsilon^{-2} \log(1/(\delta\varepsilon)))$ , and let  $\mathbf{h}$  be a random linear subspace of dimension  $k$ . Then, with probability at least  $1 - \delta$ , for every  $p, q \in \mathbb{M}$ , it holds*

$$\|p - q\| - \varepsilon\Delta \leq c \|\downarrow_{\mathbf{h}}(q) - \downarrow_{\mathbf{h}}(p)\| \leq \|p - q\| + \varepsilon\Delta,$$

and

$$(1 - \varepsilon) \|p\| - \varepsilon\Delta \leq c \|\downarrow_{\mathbf{h}}(p)\| \leq (1 + \varepsilon) \|p\| + \varepsilon\Delta.$$

Here  $c$  is a scaling parameter.

*Proof:* The proof is somewhat similar to the proof of Indyk and Naor [IN07]. Let  $b$  be a sufficiently large constant to be specified shortly, and let

$$\tau_i = \frac{\varepsilon\Delta}{b8^i}$$

for  $i \geq 0$ . Let  $N_i$  be a  $\tau_i$ -net of  $\mathbb{M}$ . By the doubling dimension property,

$$n_i = |N_i| \leq (b8^i/\varepsilon)^\tau.$$

For any point  $p \in \mathbb{M}$ , let  $u_i(p)$  be the nearest neighbor of  $p$  in  $N_i$ .

By the JL Lemma, with probability at least  $1 - \delta/2$ , the mapping  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon/10)$ -embedding of  $N_0$ . For simplicity, let  $T = c \downarrow_{\mathbf{h}}$ , where recall  $c$  is the scaling parameter of  $\downarrow_{\mathbf{h}}$ . Observe that for any  $p, q \in \mathbb{M}$ , it holds

$$\begin{aligned} \|T(u_0(p)) - T(u_0(q))\| &\leq (1 + \varepsilon/10) \|u_0(p) - u_0(q)\| \\ &\leq (1 + \varepsilon/10)(\|u_0(p) - p\| + \|p - q\| + \|q - u_0(q)\|) \\ &\leq (1 + \varepsilon/10)(\|p - q\| + 2\tau_0) \\ &\leq (1 + \varepsilon/10) \|p - q\| + 3\tau_0, \end{aligned} \tag{2}$$

by the triangle inequality. Similarly,

$$\|T(u_0(p)) - T(u_0(q))\| \geq \|p - q\| - 2\tau_0.$$



Assume, for the time being, that the distance between  $u_i(p)$  and  $u_{i+1}(p)$  gets expanded by at most a factor of  $4^{i+1}$  by the embedding  $T$ , for any  $p \in \mathbb{M}$  and  $i \geq 0$ . Observe that  $\lim_{i \rightarrow \infty} u_i(p) = p$ , for any  $p \in \mathbb{M}$ . Furthermore, by the triangle inequality, for any  $p \in \mathbb{M}$ , we have

$$\|u_i(p) - u_{i+1}(p)\| \leq \|u_i(p) - p\| + \|p - u_{i+1}(p)\| \leq \tau_i + \tau_{i+1}. \quad (3)$$

Therefore, for any  $p, q \in \mathbb{M}$ ,

$$\begin{aligned} \|T(p) - T(q)\| &\leq \|T(u_0(p)) - T(u_0(q))\| + \sum_{i=0}^{\infty} \|T(u_i(p)) - T(u_{i+1}(p))\| \\ &\quad + \sum_{i=0}^{\infty} \|T(u_i(q)) - T(u_{i+1}(q))\| \\ &\leq (1 + \varepsilon/10) \|p - q\| + 3\tau_0 + 2 \sum_{i=0}^{\infty} 4^{i+1}(\tau_i + \tau_{i+1}) \quad (\text{by (2) and (3)}) \\ &\leq \|p - q\| + \varepsilon\Delta/10 + 3\tau_0 + 16 \sum_{i=0}^{\infty} 4^i \cdot \tau_i \\ &\leq \|p - q\| + \varepsilon\Delta, \end{aligned}$$

provided  $b \geq 40$ . A similar argument shows that

$$\|T(p) - T(q)\| \geq \|p - q\| - \varepsilon\Delta.$$

By the JL Lemma, the probability that the distance between any pair of points of  $N_i \cup N_{i+1}$  get expanded by a factor larger than  $4^{i+1}$  is at most

$$(n_i + n_{i+1})^2 \exp\left(-\frac{k(4^{i+1} - 1)^2}{c_{\text{JL}}}\right) \leq 4 \left(\frac{b8^{i+1}}{\varepsilon}\right)^{2\tau} \exp\left(-\frac{k(4^{i+1} - 1)^2}{c_{\text{JL}}}\right) \leq \frac{\delta}{2^{i+2}}$$

for  $k = \Omega(\tau \ln(1/\varepsilon) + \ln(1/\delta))$ . In particular, the probability that this embedding fails for any pair in  $N_i \times N_{i+1}$ , for  $i = 1, \dots, \infty$ , is bounded by  $\sum_{i=0}^{\infty} (\delta/2^{i+2}) \leq \delta/2$ .

The second claim can be proved in a similar manner.  $\blacksquare$

In particular, observe that any curve (of finite length) in  $\mathbb{R}^d$  has doubling dimension 2 if we consider the metric to be the distance along the curve (note that if instead we use the Euclidean distance, then the doubling dimension of a curve might be  $\Theta(d)$ ).

**Corollary 4.2** *Let  $\xi \subseteq \mathbb{R}^d$  be a curve of length  $\ell$ . Let  $\varepsilon, \delta > 0$  be parameters, let  $k = O(\varepsilon^{-2} \log(1/(\delta\varepsilon)))$ , and let  $\mathfrak{h}$  be a random linear subspace of dimension  $k$ . Then, with probability at least  $1 - \delta$ , for every  $p, q \in \xi$ , we have*

$$\|p - q\| - \varepsilon\ell \leq c \|\Downarrow_{\mathfrak{h}}(q) - \Downarrow_{\mathfrak{h}}(p)\| \leq \|p - q\| + \varepsilon\ell,$$

and

$$(1 - \varepsilon) \|p\| - \varepsilon\ell \leq c \|\Downarrow_{\mathfrak{h}}(p)\| \leq (1 + \varepsilon) \|p\| + \varepsilon\ell,$$

where  $c$  is a scaling parameter.

**Remark 4.3** Corollary 4.2 cannot be strengthened to hold for multiplicative error. Indeed, consider a  $\alpha$ -net  $S = \{p_1, \dots, p_n\}$  on the sphere  $\mathbb{S}^{d-1}$ , for a sufficiently small  $\alpha > 0$ . Consider the polygonal curve  $\xi$ , starting at the origin and having  $\sum_{j=1}^i p_j$  as its  $i$ th vertex. Clearly,  $\xi$  is multiplicatively preserved by an embedding only if the distance of every point of  $S$  to the origin is preserved. However, if  $k < d$ , then any linear map of  $\mathbb{R}^d$  into  $\mathbb{R}^k$  has a linear subspace of dimension  $n - k$  that is mapped to the origin (this is the kernel of the mapping). Namely, it has arbitrarily bad distortion, and picking  $\alpha$  small enough, we can come arbitrarily close to this region of singularity (as far as distortion is concerned) in the embedding.

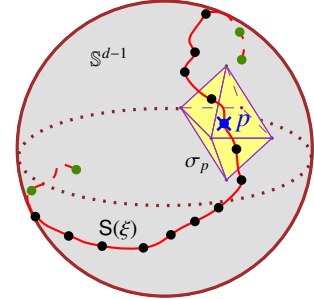
## 5 Better Embedding of Curves and Moving Points

Using results of the previous section, we are now able to prove our main result for embeddings of curves and moving points with bounded winding degree.

**Lemma 5.1** *Let  $\xi \subseteq \mathbb{R}^d$  be a curve with winding degree  $\omega$ , let  $\varepsilon, \delta > 0$  be parameters,  $k = O(\varepsilon^{-2} \log(\omega d / \varepsilon \delta))$ , and let  $\mathbf{h}$  be a random linear subspace of dimension  $k$ . Then  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon)$ -draft of  $\xi$  with probability at least  $1 - \delta$ .*

*Proof:* The proof is similar to that of Lemma 3.2. By Lemma 2.5, it suffices to prove the lemma for  $\mathcal{S}(\xi)$ . Set  $\alpha = \varepsilon / (8\sqrt{d})$ , and let  $Q$  be an  $\alpha$ -net for the curve  $\mathcal{S}(\xi)$ , where  $|Q| \leq \omega / \alpha = 8\omega\sqrt{d} / \varepsilon$ . Let  $\Sigma = \left\{ \sigma_p = p + (\varepsilon/8) \cdot \diamond_d \mid p \in Q \right\}$ . Since each simplex  $\sigma_p \in \Sigma$  contains a ball of radius  $\varepsilon / (8\sqrt{d}) = \alpha$  centered at  $p$ , the set  $\Sigma$  covers  $\mathcal{S}(\xi)$ .

Let  $P = Q \cup \{e_1, \dots, e_d\}$ ;  $|P| = |Q| + d = O(\omega d / \varepsilon)$ . Set  $k = O(\varepsilon^{-2} \log(|P| / \delta)) = O(\varepsilon^{-2} \log(\omega d / \delta))$ . By the JL Lemma, the projection  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon/8)$ -draft of  $P$  with probability at least  $1 - \delta$ . By Lemma 3.1 (and noting that the scaling parameter for each  $\sigma \in \Sigma$  is the same in the lemma), the projection  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon)$ -draft of  $\bigcup_{\sigma \in \Sigma} \sigma \supseteq \mathcal{S}(\xi)$ . ■



Note that if  $\xi$  is a curve of length  $\ell$  such that  $\|p\| \geq 1/2$  for all  $p \in \xi$ , then  $\xi$  is of winding degree  $O(\ell)$ . Therefore Lemma 5.1 implies the following.

**Corollary 5.2** *Let  $\xi \subseteq \mathbb{R}^d$  be a curve of length  $\ell$  such that  $\|p\| \geq 1/2$  for all  $p \in \xi$ , let  $\varepsilon, \delta > 0$  be parameters, let  $k = O(\varepsilon^{-2} \log(\ell d / \varepsilon \delta))$ , and let  $\mathbf{h}$  be a random  $k$ -dimensional linear subspace. Then  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon)$ -draft of  $\xi$  with probability at least  $1 - \delta$ .*

**Lemma 5.3** *Let  $\xi \subseteq \mathbb{R}^d$  be a curve,  $\varepsilon > 0$  a parameter,  $k = \Omega(1/\varepsilon^2)$ , and let  $\mathbf{h}$  be a random linear subspace of dimension  $k$ . The expected length of  $\downarrow_{\mathbf{h}}(\xi)$  is bounded by  $(1 + b\varepsilon) |\xi| / c$ , where  $c$  is the scaling parameter and  $b > 0$  is an absolute constant.*

*Proof:* This follows from the JL lemma. Indeed, consider the approximation to  $\xi$  by a polygonal line having  $m$  (equally spaced) vertices. Bound the expected length of each edge of

this polygonal line in the projection using the JL lemma. Now, using linearity of expectation, and by taking the limit (over  $m$ ), the claim follows.  $\blacksquare$

**Theorem 5.4** *Let  $\xi$  be a curve of winding degree  $\omega$  in  $\mathbb{R}^d$ . Let  $\varepsilon, \delta > 0$  be parameters,  $k = O(\varepsilon^{-2} \log(\omega/\varepsilon\delta))$ , and let  $\mathbf{h}$  be a random linear subspace of dimension  $k$ . Then  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon)$ -draft of  $\xi$  with probability at least  $1 - \delta$ .*

*Proof:* Set  $\mu = \varepsilon/(8\omega)$  and  $m = O(\mu^{-2} \log(1/(\delta\mu))) = O((\omega/\varepsilon)^2 \log(\omega/(\varepsilon\delta)))$ . Let  $\mathbf{g}$  be a random linear subspace of dimension  $m$ . By Corollary 4.2, with probability at least  $1 - \delta/3$ , for every  $p \in \mathbf{S}(\xi)$ , we have

$$c \|\downarrow_{\mathbf{g}}(p)\| \leq (1 + \mu) + \mu\omega \leq (1 + \varepsilon/10) \|p\|$$

and

$$c \|\downarrow_{\mathbf{g}}(p)\| \geq (1 - \mu) - \mu\omega \geq (1 - \varepsilon/10) \|p\|,$$

where  $c$  is a scaling parameter. Choosing  $c' = c/(1 - \varepsilon/10)$ , we obtain

$$\|p\| \leq c' \|\downarrow_{\mathbf{g}}(p)\| \leq \frac{1 + \varepsilon/10}{1 - \varepsilon/10} \|p\| \leq (1 + \varepsilon/3) \|p\|$$

for every  $p \in \mathbf{S}(\xi)$ . Hence  $\downarrow_{\mathbf{g}}$  is a  $(1 + \varepsilon/3)$ -draft of  $\mathbf{S}(\xi)$  with probability at least  $1 - \delta/3$  (and with scaling parameter  $c'$ ).

Let  $\xi' = \downarrow_{\mathbf{g}}(\mathbf{S}(\xi))$ , and set  $\eta = c \cdot \xi'$ . Observe that for any  $p \in \eta$ ,  $\|p\| \geq 1 - \varepsilon/10 \geq 1/2$ . Moreover, Lemma 5.3 implies that

$$\mathbf{E}[|\eta|] = c \mathbf{E}[|\xi'|] = O(|\mathbf{S}(\xi)|) = O(\omega).$$

By Markov's inequality,  $\Pr[|\eta| \geq b\omega/\delta] \leq \delta/3$ , where  $b$  is some constant.

Let  $\mathbf{h}$  be a random linear subspace of  $\mathbf{g}$  of dimension

$$k = O(\varepsilon^{-2} \log(|\eta|m/\varepsilon\delta)) = O(\varepsilon^{-2} \log(\omega/(\varepsilon\delta))).$$

By Corollary 5.2,  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon/3)$ -draft of  $\eta$ , and thus also a  $(1 + \varepsilon/3)$ -draft of  $\xi'$ , with probability at least  $1 - \delta/3$ .

Hence, with probability at least  $1 - \delta$ , the composition  $\downarrow_{\mathbf{h}} \circ \downarrow_{\mathbf{g}}$  is a  $(1 + \varepsilon/3)^2$ -draft, or  $(1 + \varepsilon)$ -draft, of  $\mathbf{S}(\xi)$ . Since  $\mathbf{g}$  is a random  $m$ -dimensional linear subspace of  $\mathbb{R}^d$  and  $\mathbf{h}$  is a random  $k$ -dimensional linear subspace of  $\mathbf{g}$ , it follows that  $\mathbf{h}$  is a random  $k$ -dimensional linear subspace of  $\mathbb{R}^d$ . Using Lemma 2.5, we conclude that the projection of  $\mathbb{R}^d$  onto  $\mathbf{h}$  is a  $(1 + \varepsilon)$ -draft of  $\xi$  with probability at least  $1 - \delta$ .  $\blacksquare$

The existence of an embedding of moving points is a direct consequence of Theorem 5.4.

**Theorem 5.5** *Let  $P$  be a set of  $n$  moving points in  $\mathbb{R}^d$  whose motion is of winding degree  $\omega$ . Let  $\varepsilon, \delta > 0$  be given parameters, let  $k = O(\varepsilon^{-2} \log(n\omega/\varepsilon\delta))$ , and let  $\mathbf{h}$  be a random linear subspace of dimension  $k$ . With probability at least  $1 - \delta$ , the projection  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon)$ -embedding of  $P$ .*

*Proof:* For each  $p, q \in P$ , the projection  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon)$ -draft of the curve  $p(t) - q(t)$ , with probability at least  $\delta/n^2$ , by Theorem 5.4. Consequently, with probability at least  $1 - \delta$ , a projection into a random linear subspace of dimension  $k$  is a  $(1 + \varepsilon)$ -embedding of  $P$ , as desired.  $\blacksquare$

## 6 Embedding Preserves Radius of Smallest Enclosing Ball

Let  $P$  be a set of  $n$  points in  $\mathbb{R}^d$ . Let  $\mathbb{B}(P)$  denote the smallest enclosing ball of  $P$ , and let  $\text{rad}(P)$  denote the radius of  $\mathbb{B}(P)$ . In this section we study to what extent  $\text{rad}(P(t))$  is preserved when we project a set  $P$  of moving points into low dimensions; here  $P(t)$  denotes the configuration of  $P$  at time  $t$ . This is motivated by certain applications in which, in addition to preserving pairwise distances, one may also wish to preserve the radius of each cluster in the time-varying data, so as to retain meaningful clustering information when projecting high-dimensional data into a low-dimensional space.

Intuitively, if the pairwise distances within  $P$  are scaled by a factor of  $c$  after embedding, then it is natural to ask whether the radius  $\text{rad}(P)$  is also scaled by a factor of  $c$  after embedding. Given an embedding with this property holding at all times, an immediate consequence is that one can simply compute an approximation to  $\text{rad}(P(t))$  at any time  $t$  in a low-dimensional space instead of in the original space  $\mathbb{R}^d$ .

### 6.1 Embedding of static points

For a static point set  $P$ , it is easy to show that a random projection of  $P$  does preserve  $\text{rad}(P)$ . This is probably a folklore, but we include a proof here for the sake of completeness.

**Lemma 6.1** *Let  $P$  be a set of  $n$  points in  $\mathbb{R}^d$ , and let  $\varepsilon, \delta > 0$  be parameters. Let  $\mathbf{h}$  be a random linear subspace of dimension  $k = O(\varepsilon^{-2} \log(n/\delta))$ . Then with probability at least  $1 - \delta$ , it holds*

$$\text{rad}(P) \leq c \text{rad}(\downarrow_{\mathbf{h}}(P)) \leq (1 + \varepsilon) \text{rad}(P),$$

where  $c$  is a scaling parameter.

*Proof:* Let  $z$  be the center of the ball  $\mathbb{B}(P)$ . By the JL Lemma (Lemma 2.1), with probability at least  $1 - \delta$ , the projection  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon)$ -draft of  $\{p - z \mid p \in P\}$ , which we shall assume. Further, let  $Q = P \cap \partial\mathbb{B}(P)$ . Clearly,  $z$  lies inside the convex hull of  $Q$  [GIV01]. As such,  $\hat{z} = \downarrow_{\mathbf{h}}(z)$  is also inside the convex hull of  $\downarrow_{\mathbf{h}}(Q)$  in  $\mathbf{h}$ . Let  $\hat{w} \in \mathbf{h}$  be the center of  $\mathbb{B}(\downarrow_{\mathbf{h}}(P))$ . If  $\hat{w} = \hat{z}$ , then for every  $p \in P$ ,

$$\|p - z\| \leq c \|\downarrow_{\mathbf{h}}(p) - \hat{z}\| = c \|\downarrow_{\mathbf{h}}(p) - \hat{w}\| \leq (1 + \varepsilon) \|p - z\|,$$

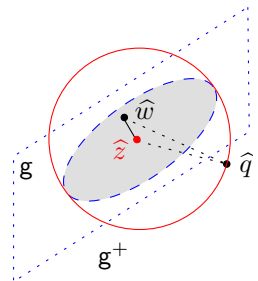
implying the lemma.

Hence, assume that  $\hat{w} \neq \hat{z}$ . Let  $\mathbf{g} \subseteq \mathbf{h}$  be the hyperplane passing through  $\hat{z}$  and orthogonal to the vector  $\hat{w} - \hat{z}$ , and let  $\mathbf{g}^+$  be the closed halfspace bounded by  $\mathbf{g}$  and not containing  $\hat{w}$ . Then there exists at least one point  $q \in Q$  such that  $\hat{q} = \downarrow_{\mathbf{h}}(q) \in \mathbf{g}^+$ . Consequently,

$$c \text{rad}(\downarrow_{\mathbf{h}}(P)) \geq c \|\hat{q} - \hat{w}\| \geq c \|\hat{q} - \hat{z}\| \geq \|q - z\| = \text{rad}(P).$$

On the other hand, it is clear that

$$c \text{rad}(\downarrow_{\mathbf{h}}(P)) \leq (1 + \varepsilon) \text{rad}(P),$$



because  $c \|\downarrow_{\mathbf{h}}(p) - \hat{z}\| \leq (1 + \varepsilon) \|p - z\|$  for all  $p \in P$ . Putting things together, we obtain the claimed result.  $\blacksquare$

## 6.2 Embedding of moving points

We next show a similar result for embeddings of moving points with polynomial motion (we cannot obtain a similar result for the more general motion of bounded winding degree). Note that a direct extension of the proof of Lemma 6.1 does not work because the trajectory of the center of  $\mathbb{B}(P(t))$  may have high descriptive complexity (the best known bound is  $O(n^{d+1})$ ).

Let  $\varepsilon > 0$  be a parameter and  $P$  be a set of  $n$  points in  $\mathbb{R}^d$ . Consider the following algorithm by Bădoiu and Clarkson [BC03] for computing an  $\varepsilon$ -approximation of  $\mathbb{B}(P)$ . We will refer to this algorithm as **aprxEncBall**. Pick an arbitrary point  $p_0 \in P$  and let  $z_0 = p_0$ . In the  $i$ th iteration, let  $p_i \in P$  be the point farthest away from  $z_{i-1}$  and let  $z_i = (1/(i+1)) \sum_{j=0}^i p_j$  be the centroid of  $p_0, \dots, p_i$ . (Note that a point may appear multiple times in the sequence, in which case it is considered with multiplicity in the computation of  $z_i$ .) Repeat this process  $\alpha = O(1/\varepsilon^2)$  times. It was shown in [BC03] that the distance between the final output  $z_\alpha$  and the center of  $\mathbb{B}(P)$  is at most  $(\varepsilon/3) \text{rad}(P)$ . Their argument can be adapted to show that it suffices to choose  $p_i$  to be a  $(1 - \varepsilon^2/b)$ -approximate farthest point of  $z_i$  in each step, for some sufficiently large constant  $b$ .

For a set  $\mathbb{X} \subseteq \mathbb{R}^d$  and a point  $p \in \mathbb{R}^d$ , let  $F_{\mathbb{X}}(p) = \max_{q \in \mathbb{X}} \|p - q\|$  denote the distance between  $p$  and its farthest neighbor in  $\mathbb{X}$ . A sequence  $\sigma = \langle p_0, p_1, \dots, p_\alpha \rangle$  of points from  $P$  is  $(1 - \varepsilon^2/b)$ -faithful with respect to  $P$  if for every  $0 \leq i \leq \alpha$ , the point  $p_i$  is at distance at least  $(1 - \varepsilon^2/b)F_P(z_{\sigma, i-1})$  from  $z_{\sigma, i-1}$ , where  $z_{\sigma, i-1} = (1/i) \sum_{j=0}^{i-1} p_j$ . Namely, a  $(1 - \varepsilon^2/b)$ -faithful sequence can be interpreted as a trace of an execution of the **aprxEncBall** algorithm which would output a  $(1 + \varepsilon/3)$ -approximate smallest enclosing ball to the given points.

**Theorem 6.2** *Let  $P$  be a set of  $n$  points in  $\mathbb{R}^d$  moving with degree of motion  $\omega$ , and  $\varepsilon, \delta > 0$  be parameters. Let  $\mathbf{h}$  be a random linear subspace of dimension  $k = O((1/\varepsilon^6) \log(n\omega/\varepsilon\delta))$ . Then with probability at least  $1 - \delta$ , for every time  $t \in \mathbb{R}$ ,*

$$\text{rad}(P(t)) \leq c \text{rad}(\downarrow_{\mathbf{h}}(P(t))) \leq (1 + \varepsilon) \text{rad}(P(t)),$$

where  $c$  is a scaling parameter.

*Proof:* Let  $V$  be the set created from  $P$  by adding to it all the centroids of multisets of  $P$  of size at most  $O(1/\varepsilon^2)$ . Since the point set  $V$  has size  $n^{O(1/\varepsilon^2)}$ , and all the points in it have degree of motion  $\omega$ , one can apply Theorem 5.5 to  $V$ , with distortion  $1 + \varepsilon^2/b$ . The resulting random linear subspace  $\mathbf{h}$  has dimension

$$k = O\left(\frac{1}{\varepsilon^4} \log \frac{|V|\omega}{\varepsilon\delta}\right) = O\left(\frac{1}{\varepsilon^6} \log \frac{n\omega}{\varepsilon\delta}\right),$$

and with probability at least  $1 - \delta$ , the mapping  $\downarrow_{\mathbf{h}}$  is a  $(1 + \varepsilon^2/b)$ -embedding of  $V$  at any point in time.

Running the algorithm **aprxEncBall** on the points of  $\downarrow_{\mathbf{h}}(V(t))$ , at time  $t$ , yields a sequence  $\sigma$  of points. The sequence can only be made out of the points that are in  $\downarrow_{\mathbf{h}}(P(t))$ .

Indeed, any other point of  $\downarrow_h(V(t))$  is a convex combination of points of  $\downarrow_h(P(t))$  and cannot be the farthest neighbor of any point in the sequence, implying that  $\sigma \subseteq \downarrow_h(P(t))$ .

Let  $\hat{\sigma}$  be the sequence of original points of  $P$  that corresponds to the points of  $\sigma$ . Observe that the centroid of any prefix of  $\hat{\sigma}$  is in  $V$ , by definition. Since  $\downarrow_h(V)$  is a  $(1 + \varepsilon^2/b)$ -embedding, it follows that the sequence  $\hat{\sigma}$  is  $(1 - \varepsilon^2/b)$ -faithful.

Let  $z$  be the centroid of  $\hat{\sigma}$ . Let  $R$  be the radius of the smallest enclosing ball of  $P(t)$  centered at  $z$ , and  $r$  be the radius of the smallest enclosing ball of  $\downarrow_h(P(t))$  centered at  $\downarrow_h(z)$ . We have

$$\text{rad}(P(t)) \leq R \leq (1 + \varepsilon/3) \text{rad}(P(t)),$$

since  $\hat{\sigma}$  is  $(1 - \varepsilon^2/b)$ -faithful and by the correctness of **aprxEncBall** when applied to  $\hat{\sigma}$ . Similarly, we have

$$\text{rad}(\downarrow_h(P(t))) \leq r \leq (1 + \varepsilon/3) \text{rad}(\downarrow_h(P(t))),$$

by the correctness of **aprxEncBall** when applied to  $\sigma$ . On the other hand, since  $z \in V$  and  $\downarrow_h$  is  $(1 + \varepsilon^2/b)$ -embedding, it follows that

$$R \leq c_h r \leq (1 + \varepsilon^2/b)R,$$

where  $c_h$  is the scaling parameter of the embedding  $\downarrow_h$ . By combining the above three inequalities, we obtain

$$(1 + \varepsilon/3)^{-1} \text{rad}(P(t)) \leq c_h \text{rad}(\downarrow_h(P(t))) \leq (1 + \varepsilon/3)(1 + \varepsilon^2/b) \text{rad}(P(t)).$$

Setting  $c = c_h(1 + \varepsilon/3)$  and noting that  $(1 + \varepsilon/3)^2(1 + \varepsilon^2/b) \leq 1 + \varepsilon$  for a sufficiently large constant  $b$ , we obtain the claimed theorem.  $\blacksquare$

### 6.3 Coresets for high-dimensional moving points

In this section we provide an extension of Bădoiu and Clarkson [BC03] result to moving points in high dimensions, which may be of independent interest.

**Lemma 6.3** *Let  $P$  be a set of  $n$  points with polynomial motion of degree  $\omega$  in  $\mathbb{R}^d$  and  $b > 0$  be a prespecified constant. Let  $q \in \mathbb{R}^d$  be another point that also moves along a polynomial trajectory of degree  $\omega$ . Then for any  $\varepsilon > 0$ , there is a subset  $Q \subseteq P$  of  $O(\omega^2/\varepsilon^2)$  points so that, at any time  $t \in \mathbb{R}$ , one of the points in  $Q$  is a  $(1 - \varepsilon^2/b)$ -approximate farthest neighbor of  $q(t)$  in  $P(t)$ , i.e., for all  $t \in \mathbb{R}$ , we have*

$$\max_{p \in Q} \|p(t) - q(t)\| \geq (1 - \varepsilon^2/b) \cdot \max_{p \in P} \|p(t) - q(t)\|.$$

*Proof:* For a moving point  $p$ , we define  $f_p(t) = \|p(t) - q(t)\|$ . Set  $\mathcal{F} = \{f_p \mid p \in P \cup \{q\}\}$ . Each function  $f_p(t) \in \mathcal{F}$  is the square root of a (positive) polynomial of degree at most  $2\omega$ . Let  $\mathcal{G} \subseteq \mathcal{F}$  be a  $\varepsilon^2/b$ -coreset of  $\mathcal{F}$  of size  $O(\omega^2/\varepsilon^2)$  [AHV04], in the sense that for any  $t \in \mathbb{R}$ ,

$$\max_{f_p \in \mathcal{G}} f_p(t) - \min_{f_p \in \mathcal{G}} f_p(t) \geq (1 - \varepsilon^2/b) \cdot \left( \max_{f_p \in \mathcal{F}} f_p(t) - \min_{f_p \in \mathcal{F}} f_p(t) \right).$$

Observe that  $\min_{f_p \in \mathcal{F}} f_p(t) = f_q(t) = 0$  for all  $t \in \mathbb{R}$ . This implies  $\max_{f_p \in \mathcal{G}} f_p(t) \geq (1 - \varepsilon^2/b) \max_{f_p \in \mathcal{F}} f_p(t)$ . Therefore, if we let  $Q = \{p \in P \mid f_p \in \mathcal{G}\}$ , then

$$\max_{p \in Q} \|p(t) - q(t)\| \geq (1 - \varepsilon^2/b) \cdot \max_{p \in P} \|p(t) - q(t)\|,$$

which implies  $Q$  is the desired subset we seek. Note that the size of  $Q$  is independent of the dimension  $d$  of the ambient space.  $\blacksquare$

**Lemma 6.4** *Let  $P$  be a set of  $n$  points with polynomial motion of degree  $\omega$  in  $\mathbb{R}^d$ , and let  $\varepsilon, \delta > 0$  be parameters. Let  $\alpha = O(1/\varepsilon^2)$ . One can find a set  $\Sigma$  of  $(O(\omega^2/\varepsilon^2))^\alpha = \exp(O(\varepsilon^{-2} \log(\omega/\varepsilon)))$  sequences of points of  $P$ , such that at any time  $t \in \mathbb{R}$ , at least one sequence  $\sigma \in \Sigma$  is  $(1 - \varepsilon^2/b)$ -faithful with respect to  $P(t)$ .*

*Proof:* For  $1 \leq i \leq \alpha$ , we construct a collection  $\Sigma_i$  of sequences of  $i$  points from  $P$  in the following way. For  $i = 1$ , we pick an arbitrary point  $p \in P$  and set  $\Sigma_1 = \{\langle p \rangle\}$ . To construct  $\Sigma_i$  for  $i \geq 2$ , we proceed as follows. For each sequence  $\sigma \in \Sigma_{i-1}$ , the centroid  $z_\sigma$  of the sequence moves with polynomial degree  $\omega$ . By Lemma 6.3, we can find a subset  $Q_\sigma \subseteq P$  of  $O(\omega^2/\varepsilon^2)$  points so that at any time  $t \in \mathbb{R}$ ,  $Q_\sigma(t)$  contains a  $(1 - \varepsilon^2/b)$ -approximate farthest neighbor of  $z_\sigma(t)$  in  $P(t)$ . We set

$$\Sigma_i = \bigcup_{\sigma \in \Sigma_{i-1}} \left\{ \sigma \circ q \mid q \in Q_\sigma \right\},$$

where  $\sigma \circ q$  denotes the sequence obtained by appending  $q$  to  $\sigma$ . Note that  $|\Sigma_\alpha| = (O(\omega^2/\varepsilon^2))^\alpha$ , and by construction at any time at least one sequence in  $\Sigma_\alpha$  is  $(1 - \varepsilon^2/b)$ -faithful. Setting  $\Sigma = \Sigma_\alpha$ , we obtain the lemma.  $\blacksquare$

By letting  $Q = \bigcup_{\sigma \in \Sigma} \{q \mid q \in \sigma\}$ , and using the argument in Bădoiu and Clarkson [BC03], we obtain the following extension of their result for moving points.

**Theorem 6.5** *For any set  $P$  of  $n$  points in  $\mathbb{R}^d$  with polynomial motion of degree  $\omega$ , there exists an  $\varepsilon$ -coreset  $Q \subset P$  of size  $\exp(O(\varepsilon^{-2} \log(\omega/\varepsilon)))$ , in the sense that for any time  $t \in \mathbb{R}$ ,*

$$\text{rad}(Q(t)) \leq \text{rad}(P(t)) \leq (1 + \varepsilon) \text{rad}(Q(t)).$$

## 7 Conclusions

In this paper we presented some progress towards the problem of dimensionality reduction for time-varying data. Our results indicate that the JL lemma and dimensionality reduction in fact preserve much more structure than was realized before. In particular, they work for more general low-dimensional objects including curves, flats and surfaces. Some of our results can be further extended to the noisy setting. For example, although in Theorem 5.4 we have assumed the curve to be of bounded winding degree, the proof in fact suggests that as long as a curve  $\xi$  can be approximated by another curve  $\eta$  with winding degree  $\omega$ , in the

sense that the Hausdorff distance between  $S(\xi)$  and  $S(\eta)$  is  $O(1/\sqrt{d})$ , then a similar result remains true for  $\xi$ .

Besides embedding time-varying data in Euclidean space, a similar problem is embedding time-varying metrics into normed spaces. A celebrated result by Bourgain [Bou85] states that every  $n$ -point metric space  $(\mathbb{X}, d)$  can be embedded into  $\ell_p$  ( $p \geq 1$ ) with  $O(\log n)$  distortion. Furthermore, one can choose the dimension of the host space to be  $O(\log n)$  [ABN06]. Now consider the case in which the distances between objects in a metric space  $(\mathbb{X}, d)$  can change over time. More precisely, if the distance of any two objects  $u, v \in \mathbb{X}$  is a function of time  $t \in \mathbb{R}$ , then  $(\mathbb{X}, d)$  is called a *time-varying metric space*. For example, if the lengths of the edges in a graph are functions of time  $t$ , then the shortest path metric on the graph induces a time-varying metric space. We can obtain a similar result for embedding time-varying metrics into normed spaces as follows. Let  $(\mathbb{X}, d)$  be a time-varying metric such that the distance of any pair of objects in  $\mathbb{X}$  is a polynomial of  $t$  of degree at most  $\omega$ . Then there exists a set  $P(t)$  of  $n$  continuously moving points in  $\ell_p^k$  where  $k = O(\log(n\omega))$  for  $p \in [1, 2]$  and  $k = O(\log n \log(n\omega))$  for  $p > 2$ , such that for any  $t \in \mathbb{R}$ ,  $P(t)$  is an embedding of  $(\mathbb{X}, d)$  at time  $t$  into  $\ell_p^k$  with distortion  $O(\log n)$ . The claim follows by adapting of the proof of Linial, London, and Rabinovich [LLR95] and using Theorem 5.5; the result can be extended to other well-behaved functions besides polynomials.

## Acknowledgments.

The authors thank Sayan Mukherjee, Piotr Indyk, Assaf Naor, Kasturi Varadarajan, and Yusu Wang for helpful discussions concerning the problems studied in this paper.

## References

- [ABN06] I. Abraham, Y. Bartal, and O. Neimany. Advances in metric embedding theory. In *Proc. 38th Annu. ACM Sympos. Theory Comput.*, pages 271–286, 2006.
- [Ach01] D. Achlioptas. Database-friendly random projections. In *Proc. 20th ACM Sympos. Principles Database Syst.*, pages 274–281, 2001.
- [AHV04] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan. Approximating extent measures of points. *J. Assoc. Comput. Mach.*, 51(4):606–635, 2004.
- [Alo03] N. Alon. Problems and results in extremal combinatorics I. *Discrete Math.*, 273:31–53, 2003.
- [AM94] P. K. Agarwal and J. Matoušek. On range searching with semialgebraic sets. *Discrete Comput. Geom.*, 11:393–418, 1994.
- [AV99] R. Arriaga and S. Vempala. An algorithmic theory of learning: robust concepts and random projection. In *Proc. 40th Annu. IEEE Sympos. Found. Comput. Sci.*, pages 616–623, 1999.
- [BC03] M. Bădoiu and K. Clarkson. Smaller coresets for balls. In *Proc. 14th ACM-SIAM Sympos. Discrete Algorithms*, pages 801–802, 2003.



- [BH03] M. Baumann and U. Helmke. Singular value decomposition of time-varying matrices. *Future Gen. Comp. Sys.*, 19:353–361, 2003.
- [BM01] E. Bingham and H. Mannila. Random projection in dimensionality reduction: applications to image and text data. In *Proc. 7th ACM SIGKDD Intl. Conf. Knowl. Disc. Data Mining*, pages 245–250, 2001.
- [BN03] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.
- [Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.
- [BW06] R. Baraniuk and M. Wakin. Random projections of smooth manifolds. In *Proc. IEEE Intl. Conf. Acoustics, Speech and Signal Processing*, pages 941–944, 2006.
- [Das99] S. Dasgupta. Learning mixtures of Gaussians. In *Proc. 40th Annu. IEEE Sympos. Found. Comput. Sci.*, pages 634–644, 1999.
- [DE99] L. Dieci and T. Eirola. On smooth decompositions of matrices. *SIAM J. Matrix Anal. Appl.*, 20:800–819, 1999.
- [DG03] S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Rand. Struct. Alg.*, 22(3):60–65, 2003.
- [FM88] P. Frankl and H. Maehara. The johnson-lindenstrauss lemma and the sphericity of some graphs. *J. Comb. Theory B*, 44:355–362, 1988.
- [GIV01] A. Goel, P. Indyk, and K. R. Varadarajan. Reductions among high dimensional proximity problems. In *Proc. 12th ACM-SIAM Sympos. Discrete Algorithms*, pages 769–778, 2001.
- [HPP01] P. Hall, D. Poskitt, and B. Presnell. A functional data-analytic approach to signal discrimination. *Technometrics*, 43(1):1–9, 2001.
- [IM98] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proc. 30th Annu. ACM Sympos. Theory Comput.*, pages 604–613, 1998.
- [IM04] P. Indyk and J. Matoušek. Low-distortion embeddings of finite metric spaces. In *Handbook of Discrete and Computational Geometry*, pages 177–196. CRC Press, 2004.
- [IN07] P. Indyk and A. Naor. Nearest neighbor preserving embeddings. *ACM Trans. Algo.*, 2007. To appear.
- [Ind01] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *Proc. 42nd Annu. IEEE Sympos. Found. Comput. Sci.*, pages 10–31, 2001. Tutorial.

- [JL84] W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz mapping into hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.
- [Kle97] J. Kleinberg. Two algorithms for nearest-neighbor search in high dimensions. In *Proc. 29th Annu. ACM Sympos. Theory Comput.*, pages 599–608, 1997.
- [KOR00] E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM J. Comput.*, 2(30):457–474, 2000.
- [Lin02] N. Linial. Finite metric spaces – combinatorics, geometry and algorithms. In *Proc. Int. Cong. Math. III*, pages 573–586, 2002.
- [LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.
- [Mag02] A. Magen. Dimensionality reductions that preserve volumes and distance to affine spaces, and their algorithmic applications. In *The 6th Intl. Work. Rand. Appr. Tech. Comp. Sci.*, pages 239–253, 2002.
- [Mat02] J. Matoušek. *Lectures on Discrete Geometry*. Springer, 2002.
- [San53] L. Santalo. *Introduction to Integral Geometry*. Paris, Hermann, 1953.
- [Sar06] T. Sarlos. Improved approximation algorithms for large matrices via random projections. In *Proc. 47th Annu. IEEE Sympos. Found. Comput. Sci.*, pages 143–152, 2006.
- [Vem04] S. Vempala. *The Random Projection Method*, volume 65 of *Discret. Math. and Theoret. Comput. Sci.* American Mathematical Society, 2004.