Near Optimal Dimensionality Reductions That Preserve Volumes

Avner Magen and Anastasios Zouzias

Department of Computer Science,
University of Toronto

Abstract. Let P be a set of n points in Euclidean space and let $0 < \varepsilon < 1$. A well-known result of Johnson and Lindenstrauss states that there is a projection of P onto a subspace of dimension $O(\varepsilon^{-2}\log n)$ such that distances change by at most a factor of $1+\varepsilon$. We consider an extension of this result. Our goal is to find an analogous dimension reduction where not only pairs but all subsets of at most k points maintain their volume approximately. More precisely, we require that sets of size $s \leq k$ preserve their volumes within a factor of $(1+\varepsilon)^{s-1}$. We show that this can be achieved using $O(\max\{\frac{k}{\varepsilon}, \varepsilon^{-2}\log n\})$ dimensions. This in particular means that for $k = O(\log n/\varepsilon)$ we require no more dimensions (asymptotically) than the special case $k = 2$, handled by Johnson and Lindenstrauss. Our work improves on a result of Magen (that required as many as $O(k\varepsilon^{-2}\log n)$ dimensions) and is tight up to a factor of $O(1/\varepsilon)$. Another outcome of our work is an alternative and greatly simplified proof of the result of Magen showing that all distances between points and affine subspaces spanned by a small number of points are approximately preserved when projecting onto $O(k\varepsilon^{-2}\log n)$ dimensions.

1 Introduction

A classical result of Johnson and Lindenstrauss [12] shows that a set of n points in the Euclidean space can be projected onto $O(\varepsilon^{-2}\log n)$ dimensions so that all distances are changed by at most a factor of $1+\varepsilon$. Many important works in areas such as computational geometry, approximation algorithms and discrete geometry build on this result in order to achieve a computation speed-up, reduce space requirements or simply exploit the added simplicity of a low dimensional space.

However, the rich structure of Euclidean spaces gives rise to many many geometric parameters other than distances between points. For example, we could care about the centre of gravity of sets of points, angles and areas of triangles of triplets of points among a fixed set of points P, and more generally, the volume spanned by some subsets of P or the volume of the smallest ellipsoid containing them. The generalization of the Johnson-Lindenstauss lemma to the geometry of subsets of bounded size was considered in [14] where it was shown that it is possible to embed an n-point set of the Euclidean space onto an $O(k\varepsilon^{-2}\log n)$-dimensional Euclidean space, such that no set of size $s \leq k$ changes its volume by more than a factor of $(1+\varepsilon)^{s-1}$.The exponent $s-1$ should be thought as a natural normalization measure. Notice that scaling a set of size s by a factor of $1+\varepsilon$ will change its volume by precisely the above factor. In the current work we improve this result by showing that $O(\max\{\frac{k}{\varepsilon}, \varepsilon^{-2}\log n\})$ dimensions suffice in order to get the same guarantee.

A. Goel et al. (Eds.): APPROX and RANDOM 2008, LNCS 5171, pp. 523–534, 2008.

Theorem 1 (Main theorem). *Let $0 < \varepsilon \leq 1/2$ and let k,n,d be positive integers, such that $d = O(\max\{\frac{k}{\varepsilon}, \varepsilon^{-2}\log n\})$. Then for any n-point subset P of the Euclidean space $\mathbb{R}^n$, there is a mapping $f : P \to \mathbb{R}^d$, such that for all subsets S of P, $1 < |S| < k$,*

$$1-\varepsilon \leq \left(\frac{\mathrm{vol(f(S))}}{\mathrm{vol(S)}}\right)^{\frac{1}{s-1}} \leq 1+\varepsilon.$$

Moreover, the mapping f can be constructed efficiently in randomized polynomial time using a Gaussian random matrix.

The line of work presented here (as well as in [14]) is related to, however quite different from, Feige's work on volume-respecting embeddings [8]. Feige defined a notion of volume for sets in general metric spaces that is very different than ours, and measured the quality of an embedding from such spaces into Euclidean space. For the image of the points of these embeddings, Feige's definition of volume is identical to the one used here and in [14]. Embeddings that do not significantly change volumes of small sets (of size $\leq k$) are presented in [8] and further it is shown how these embeddings lead to important algorithmic applications (see also [19]). The typical size k of sets in [8] is $O(\log n)$ and therefore our work shows that the embedding that are obtained in [8] can be assumed to use no more than $O(\log n)$ dimensions. Compare this to the $O(n)$ as in the original embedding or the $O(\log^2 n)$ bound that can be obtained by [14].

As was shown by Alon [5], the upper bound on the dimensionality of the projection of Johnson Lindenstauss is nearly tight[1], giving a lower bound of $\Omega(\varepsilon^{-2}\log n/\log(\frac{1}{\varepsilon}))$ dimensions. Further, in our setting it is immediate that at least $k-1$ dimensions are needed (otherwise the image of sets of size k will not be full dimensional and will not therefore have a positive volume). These two facts provide a dimension lower bound of $\Omega(\max\{\varepsilon^{-2}\log n/\log(\frac{1}{\varepsilon}), k\})$ which makes our upper bound tight up to a factor of $1/\varepsilon$ throughout the whole range of the parameter k.

Similarly to other dimension reduction results our embedding uses random projections. Several variants have been used in the past, each defining 'random' projection in a slightly different way. Originally, Johnson and Lindenstrauss considered projecting onto a random d-dimensional subspace, while Frankl and Mehara [9] used projection onto d independent random unit vectors. In most later works the standard approach has been to use projection onto d n-dimensional Gaussian, an approach that we adopt here.

In our analysis it is convenient to view the projection as applying a matrix multiplication with a random matrix (of appropriate dimensions) whose entries are i.i.d. Gaussians. A critical component in our analysis is the following "invariance" claim.

Claim. Let S be a subset of the Euclidean space and let $V^{\pi}(S)$ be the volume of the projection of S by a Gaussian random matrix. Then the distribution of $V^{\pi}(S)$ depends linearly on vol(S) but does not depend on other properties of S.

When S is of size 2 the claim is nothing else but an immediate use of the fact that the projections are rotational invariant: indeed, any set of size 2 is the same up to an orthonormal transformation, translation and scaling. For $|S| > 2$, while the claim is still easy to show, it may seem somewhat counterintuitive from a geometric viewpoint. It is

[1] With respect to any embedding, not necessarily a projection.

certainly no longer the case that any two sets with the same volume are the same up to an orthonormal transformation. Specifically, it does not seem clear why a very 'flat' (for example a perturbation of a co-linear set of points) set should behave similarly to a 'round' set (like a symmetric simplex) of the same volume, with respect to the volume of their projections. The question of the distortion of subsets readily reduces to a stochastic question about one particular set S. Essentially, one needs to study the probability that the volume of the projection of this set deviates from its expected value. This makes the effectiveness of the above claim clear: it means that the question can be further reduced to the question of concentration of volume with respect to a *particular set S* of our choice! Since there are roughly n^s sets of size s to consider, we need to bound the probability of a bad event with respect to any arbitrary set by roughly $e^{-\Omega(sd)}$. The previous bound of [14] (implicitly) showed a concentration bound of only $e^{-\Omega(d)}$ which is one way to understand the improvement of the current work.

Other works have extended the Johnson Lindenstauss original work. From the computational perspective, emphasis was placed on derandomizing the embedding [7,18] and on speeding its computation. This last challenge has attracted considerable amount of attention. Achlioptas [1] has shown that projection onto a (randomly selected) set of discrete vectors generated the same approximation guarantee using the same dimensionality. Ailon and Chazelle [3] supplied a method that uses *Sparse* Gaussian matrices for the projection to achieve fast computation (which they call "Fast Johnson Lindenstrauss Transform"). See also [15,4,13] for a related treatment and extensions. On another branch of extentions (closer in flavour to our result) are works that require that the embeddings will preserve richer structure of the geometry. For example, in [2] the authors ask about distance between points that are *moving* according to some algebraically-limited curve; in [17] for affine subspaces, in [11] for sets with bounded doubling dimension, and in [2,6,20] for curves, (smooth) surfaces and manifolds.

2 Notation and Preliminaries

We think of n points in $\mathbb{R}^n$ as an $n \times n$ matrix P, where the rows correspond to the points and the columns to the coordinates. We call the set $\{0, e_1, e_2, \ldots, e_n\}$ i.e., the n-dimensional standard vectors of $\mathbb{R}^n$ with the origin, *regular*. We associate with a set of k points a volume which is the $(k-1)$-dimensional volume of its convex-hull. For $k = 2$ notice that $\mathrm{vol}([\mathrm{x}, \mathrm{y}]) = \mathrm{d}(\mathrm{x}, \mathrm{y})$, and for $k = 3$ is the area of the triangle with vertices the points of the set, etc. Throughout this paper we denote the volume of a set S in the Euclidean space by $\mathrm{vol}(\mathrm{S})$.

We use $\|\cdot\|$ to denote the Euclidean norm. Let X_i $i = 1, \ldots, k$ be k independent, normally distributed random variables with zero mean and variance one, then the random variable $\chi_k^2 = \sum_{i=1}^k X_i^2$ is a Chi-square random variable with k degrees of freedom. If A is an $r \times s$ matrix and B is a $p \times q$ matrix, then the Kronecker product $A \otimes B$ is the $rp \times sq$ block matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \ldots & a_{1s}B \\ \vdots & \ddots & \vdots \\ a_{r1}B & \ldots & a_{rs}B \end{bmatrix}.$$

By $\mathrm{vec(A)} = [\mathrm{a_{11}}, \dots, \mathrm{a_{r1}}, \mathrm{a_{12}}, \dots, \mathrm{a_{r2}}, \dots, \mathrm{a_{1s}}, \dots, \mathrm{a_{rs}}]^{\mathrm{t}}$ we denote the vectorization of the matrix A. We will use P_S to denote a subset S of rows of P. Let X, Y be random variables. We say that X is *stochastically larger* than Y ($X \succeq Y$) if $\Pr[X > x] \geq \Pr[Y > x]$ for all $x \in \mathbb{R}$. Also $X \sim \mathcal{N}(\mu, \sigma^2)$ denotes that X follows the normal distribution with mean μ and variance σ^2, also $\mathcal{N}_n(\mu', \Sigma)$ is the multivariate n dimensional normal distribution with mean vector μ' and covariance matrix Σ. Similarly, we can define the matrix variate Gaussian distribution, $\mathcal{N}_{n,d}(M, \Sigma'_{nd \times nd})$ with mean matrix M and covariance matrix Σ' of dimension $nt \times nt$. Note that the latter definition is equivalent with the multivariate case, considering its vectorization. However, if we restrict the structure of the correlation matrix Σ' we can capture the matrix form of the entries (see the following Definition).

Definition 1 (Gaussian Random Matrix). *The random matrix X of dimensions $n \times d$ is said to have a matrix variate normal distribution with mean matrix M of size $n \times d$ and covariance matrix $\Sigma \otimes \Psi$ (denoted by $X \sim \mathcal{N}_{n,d}(M, \Sigma'_{nd \times nd})$), where Σ, Ψ are positive definite matrices of size $n \times n$ and $d \times d$ respectively, if* $\mathrm{vec}(\mathrm{X^t}) \sim \mathcal{N}_{\mathrm{nd}}(\mathrm{vec}(\mathrm{M^t}), \Sigma \otimes \Psi)$.

A brief explanation of the above definition is the following: The use of the tensor product ($\Sigma \otimes \Psi$) is chosen to indicate that the correlation[2] between its entries has a specific structure. More concretely, every row is correlated with respect to the Σ covariance matrix and every column with respect to Ψ. Hence, the correlation between two entries of X, say X_{ij} and X_{lk}, $E[X_{ij}X_{lk}]$ is equal to $\Sigma_{il} \cdot \Psi_{jk}$.

3 A Regular Set of Points Preserves Its Volume

Assume that the set in Euclidean space we wish to reduce its dimensionality is the regular one. Consider a subset S of the regular set of size $s \leq k$ *with* the origin. Our goal is to show that the volume of the projection of such a set is very concentrated, assuming s is sufficient small. Also denote by $X \sim \mathcal{N}_{n,d}(0, I_{nd})$ the projection matrix[3]. Notice that since our input set is the regular (identity matrix), the image of their projection is simply X (the projection matrix), and recall that the points that correspond to S are represented by X_S. It is well known that the volume of the *projected* points of S is

$$\sqrt{\mathrm{det}(\mathrm{X_S X_S^t})}/s!.$$

Therefore the question of volumes is now reduced to one about the determinant of the Gram matrix of X_S.

We will use the following lemma which gives a simple characterization of this latter random variable.

Lemma 1 ([16]). *Let $X \sim \mathcal{N}_{k,d}(0, I_{kd})$. The k-dimensional volume of the parallelotope determined by $X_{\{i\}}$, $i = 1, \dots, k$ is the product of two independent random variables one of which has a χ-distribution with $d - k + 1$ degrees of freedom and the other is distributed as the $k - 1$ dimensional volume of the parallelotope spanned by $k - 1$ independent Gaussian random vectors, i.e. $\mathcal{N}_{k-1,d}(0, I_{(k-1)d})$. Furthermore,*

[2] Since the entries have zero mean, the correlation between the entries ij and lk is $E[X_{ij}X_{lk}]$.

[3] For ease of presentation, we will not consider the normalization parameter $d^{-1/2}$ at this point.

$$\mathrm{det}(\mathrm{XX^t}) \sim \prod_{\mathrm{i}=1}^{\mathrm{k}} \chi^2_{\mathrm{d-i+1}}.$$

The proof is simple and geometric thus we supply it here for completeness.

Proof. Let $\Delta_d^{(k)} = \sqrt{\mathrm{det}(\mathrm{XX^t})}$ denote the volume of the parallelotope of the k random vectors. Then

$$\Delta_d^{(k)} = a_k \Delta_d^{(k-1)},$$

where $\Delta_d^{(k-1)}$ is the k-dimensional volume of the parallelotope determined by the set of vectors $X_1, X_2, \ldots, X_{k-1}$ and a_k is the distance of X_k from the subspace spanned by $X_1, X_2, \ldots, X_{k-1}$.

Now we will show that a_k is distributed as a Chi random variable with $d-k+1$ degrees of freedom. Using the spherical symmetry of the distribution of the points we can assume w.l.o.g. that the points X_i $i = 1, \ldots, k-1$ span the subspace $W = \{x \in \mathbb{R}^d | x(k) = x(k+1) = \cdots = x(d) = 0\}$, i.e. the set of points that the $d-k+1$ last coordinates are equal to zero. Next we will show that $a_k \sim \chi_{d-k+1}$. Notice that the distance of the point X_k from the subspace that the rest $k-1$ points span is equal to[4] $\mathrm{dist}(X_k, W) = \sqrt{\sum_{i=k}^d X_{ki}^2}$, which is a Chi random variable of $d-k+1$ degrees of freedom, since $X_{i,j} \sim \mathcal{N}(0,1)$. Also note that a_k is independent of $\Delta_d^{(k-1)}$. Using the above statement recursively, we conclude that $\mathrm{det}(\mathrm{XX^t}) \sim \prod_{\mathrm{i}=1}^{\mathrm{k}} \chi^2_{\mathrm{d-i+1}}$ with the Chi-square random variables being independent. □

Due to the normalization (see Theorem 1), it turns out that the random variable we are actually interested in is $(\mathrm{det}(\mathrm{X_S X_S^t}))^{\frac{1}{2s}}$ and so is the geometric mean of a sequence of Chi-square independent random variables with similar numbers of degrees of freedom. This falls under the general framework of law-of-large-numbers, and we should typically expect an amplification of the concentration which grows exponentially with s. This statement is made formal by a concentration result of a (single) Chi-square random variable.

Theorem 2 (Theorem 4, [10]). *Let $u_i := \chi^2_{d-i+1}$ be independent Chi-square random variables for $i = 1, 2, \ldots, s$. If u_i are independent, then the following holds for every $s \geq 1$,*

$$\chi^2_{s(d-s+1)+\frac{(s-1)(s-2)}{2}} \succeq s\left(\prod_{i=1}^{s} u_i\right)^{1/s} \succeq \chi^2_{s(d-s+1)}. \tag{1}$$

We are now ready to prove that the random embedding $f: \mathbb{R}^n \mapsto \mathbb{R}^d$ defined by $p \mapsto \frac{p^t X}{\sqrt{d}}$, $X \sim \mathcal{N}_{n,d}(0, I_{nd})$ for $p \in \mathbb{R}^n$ preserves the volume of regular sets of bounded size with high enough probability.

Lemma 2. *Let $0 < \varepsilon \leq 1/2$ and let f be the random embedding defined as above. Further, let S be a subset of $\mathbb{R}^n$ that contains the origin and s standard vectors, with $s < k < \frac{d\varepsilon}{2}$. Then we have that*

[4] The length of the orthogonal projection of X_k to the subspace W.

$$\Pr\left[1-\varepsilon < \left(\frac{\mathrm{vol(f(S))}}{\mathrm{vol(S)}}\right)^{\frac{1}{s}} < 1+\varepsilon\right] \geq 1-2\exp\left(-s(d-(s-1))\frac{\varepsilon^2}{24}\right). \qquad (2)$$

Proof. We define the random variable $Z = \left(\det(\mathrm{X_S X_S^t})\right)^{1/s}$, $U = \frac{1}{s}\chi^2_{sd-\frac{s^2+s}{2}+1}$ its upper stochastic bound and $L = \frac{1}{s}\chi^2_{sd-s^2+s}$ its lower stochastic bound i.e.,

$$U \succeq Z \succeq L$$

holds from Theorem 2. Also note that this implies upper and lower bounds for the expectation of Z, $d - \frac{s+1}{2} + 1/s \geq E[Z] \geq d-s+1$, with $E[L]-E[U] \geq -\frac{s}{2}$ for $s \geq 1$. Now we will relate the volume of an arbitrary subset of P with the random variable Z. Using that $\mathrm{vol(f(S))} = \frac{\sqrt{\det(\mathrm{X_S X_S^t})}}{\mathrm{d^{s/2} s!}}$ and $\mathrm{vol(S)} = \frac{1}{\mathrm{s!}}$, we get for the upper tail:

$$\begin{aligned}\Pr\left[\left(\frac{\mathrm{vol(X_S)}}{d^{s/2}\cdot \mathrm{vol(I_S)}}\right)^{\frac{1}{s}} > 1+\varepsilon\right] &= \Pr\left[\sqrt{\frac{Z}{d}} > (1+\varepsilon)\right] \\ &\leq \Pr\left[\frac{Z}{E[Z]} > (1+\varepsilon)^2\right] \\ &\leq \Pr\left[\frac{Z}{E[Z]} > 1+2\varepsilon\right]\end{aligned}$$

using that $d \geq E[Z]$. Similarly, the lower tail becomes

$$\begin{aligned}\Pr\left[\left(\frac{\mathrm{vol(X_S)}}{d^{s/2}\cdot \mathrm{vol(I_S)}}\right)^{\frac{1}{s}} < 1-\varepsilon\right] &= \Pr\left[\sqrt{\frac{Z}{d}} < (1-\varepsilon)\right] \\ &= \Pr\left[\frac{Z}{d} < (1-\varepsilon)^2\right] \\ &= \Pr\left[\frac{Z}{E[Z]} < \frac{d}{E[Z]}(1-\varepsilon)^2\right] \\ &\leq \Pr\left[\frac{Z}{E[Z]} < (1+\varepsilon)(1-\varepsilon)^2\right] \\ &\leq \Pr\left[\frac{Z}{E[Z]} < 1-\varepsilon\right]\end{aligned}$$

using that $\frac{d}{E[Z]} \leq 1+\varepsilon$, which is true since $d \geq 2k/\varepsilon$ and $\varepsilon \leq 1$. Now we bound the right tail of Z.

$$\begin{aligned}\Pr\left[Z-E[Z] \geq 2\varepsilon E[Z]\right] &\leq \Pr\left[U-E[Z] \geq 2\varepsilon E[Z]\right] \\ &= \Pr[U-E[U] \geq 2\varepsilon E[U] + (1+2\varepsilon)(E[Z]-E[U])] \\ &\leq \Pr[U-E[U] \geq 2\varepsilon E[U] + (1+2\varepsilon)(E[L]-E[U])]\end{aligned}$$

using $U \succeq Z$ and $E[Z] \geq E[L]$. Now we bound $(1+2\varepsilon)(E[L]-E[U])$ from below. It is not hard to show that $(1+2\varepsilon)(E[L]-E[U]) \geq -\frac{3\varepsilon}{4}E[U]$ since $d \geq 2k/\varepsilon$. Therefore

$$\Pr\left[Z-E[Z] \geq 2\varepsilon E[Z]\right] \leq \Pr[U-E[U] \geq \varepsilon E[U]].$$

Now applying Lemma 4 on U we get the bound

$$\Pr[Z \geq (1+2\varepsilon)E[Z]] \leq \exp\left(-(sd - s(s-1)/2+1)\frac{\varepsilon^2}{6}\right).$$

For the other tail of the random variable Z, we have that

$$\begin{aligned}\Pr[Z - E[Z] < -\varepsilon E[Z]] &\leq \Pr[L - E[Z] < -\varepsilon E[Z]]\\ &\leq \Pr[L - E[L] < -\varepsilon E[L] + (1-\varepsilon)(E[Z] - E[L])]\\ &\leq \Pr[L - E[L] < -\varepsilon E[U] + (E[U] - E[L])]\end{aligned}$$

using that $Z \succeq L$ and $E[Z] \leq E[U]$. Again we bound $(E[U] - E[L])$ from above. It is not hard to show that $(E[U] - E[L]) \leq \frac{3}{8}\varepsilon E[L]$ since $d \geq 2k/\varepsilon$ and $\varepsilon \leq 1/2$, so

$$\Pr[Z - E[Z] < -\varepsilon E[Z]] \leq \Pr[L - E[L] < -\varepsilon/2 E[L]]$$

holds. Therefore applying Lemma 4 on L we get

$$\Pr[Z < (1-\varepsilon)E[Z]] \leq \exp\left(-(sd - s(s-1))\frac{\varepsilon^2}{24}\right).$$

Comparing the upper and lower bound the lemma follows. □

Remark: The bound on k ($k = O(d\varepsilon)$) is tight. While the probabilistic arguments show that the volume of a projection of a subset is concentrated around its mean, we really have to show that it is concentrated around the volume of the set (before the projection). In other words, it is a necessary condition that

$$\frac{\mu_s}{1/s!} = 1 \pm O(\varepsilon) \tag{3}$$

where μ_s is the expected normalized volume of a regular set of size s. As long as we deal with sets of fixed cardinality, we can easily scale Equation 3 making the LHS equal to 1. However, it turns out that $\frac{\mu_s}{1/s!}$ is decreasing in s and furthermore for sufficiently large s it may be smaller than $1 - O(\varepsilon)$. Here is why, $\frac{\mu_s}{1/s!} = E[(\prod_{i=1}^{s} \chi^2_{d-i+1})^{\frac{1}{2s}}] \leq \left(\prod_{i=1}^{s} E[\chi^2_{d-i+1}]\right)^{\frac{1}{2s}} \leq \sqrt[2s]{d(d-1)\ldots(d-s+1)} \leq \sqrt{d-(s-1)/2}$ using independence, Jensen's inequality and arithmetic-geometric mean inequality. On the other hand[5], $\frac{\mu_1}{1/1!} = E[\chi_d] \geq \sqrt{d-1}$. Therefore, no matter what scaling is used we must have that $\sqrt{d-(s-1)/2}/\sqrt{d-1} \geq 1 - O(\varepsilon)$ for all $s \leq k$, from which it follows that $k \leq O(d\varepsilon)$.

4 Extension to the General Case

In this section we will show that if we randomly project a set of s points that are in general position, the (distribution of the) volume of the projection depends linearly *only* on

[5] A simple calculation using $E[\chi_d] = \sqrt{2}\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})}$ and $E[\chi_d] \geq \sqrt{E[\chi_d]E[\chi_{d-1}]}$ gives the result.

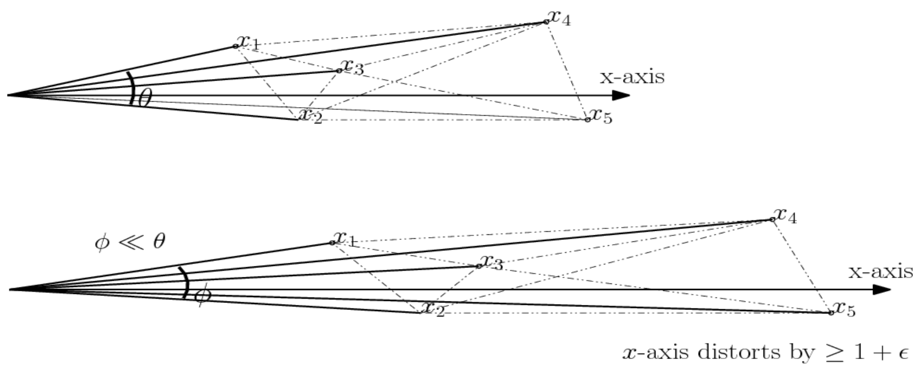

Fig. 1. Example that illustrates the extension of the regular case to the general

the volume of the original set. To gain some intuition, let's consider an example that is essentially as different as possible from the regular case. Consider the one-dimensional set of size s in $\mathbb{R}^n$, $(i, 0, \ldots, 0)$ with $i = 1, \ldots, s$. By adding a small random perturbation (and changing the location of points by distance at most $\delta \ll \varepsilon$) the points will be in general position, and the perturbed set will have positive volume. Consider a random projection π onto d dimensions, normalized so that in expectation distances do not change. Now, look at the event $E := \{\|\pi(e_1)\| > 1+\varepsilon\}$ where e_1 is the first standard vector. We know that $\Pr[E] = \exp(-\Theta(d\varepsilon^2))$. But notice that when E occurs then π expands *all* distances in the set by a factor $1+\varepsilon - O(\delta)$. At this point it may be tempting to conclude that event E implies that the set was roughly scaled by some factor that is at least $1+\varepsilon$. If that were the case then it would mean that the probability of bad projections for this set would be too big, that is $e^{-\Theta(d\varepsilon^2)}$ instead of $e^{-\Theta(sd\varepsilon^2)}$.

However, this is not the case. The reason is that conditioning on the event E does not provide any information about the expansion or contraction of the perpendicular space of the x-axis. Conditioning on E, we observe that the angles between the x-axis and any two points will decrease, since the x-axis expands (see Figure 1). Therefore the intuition that this set is scaled (conditioned on E) is wrong, since it is "squeezed" in the e_1 direction.

Next we will prove a technical lemma that will allows us to extend the volume concentration from the regular set to a set of points in general position.

Lemma 3. *Let S be a $s \times n$ matrix so that every row corresponds to a point in $\mathbb{R}^n$. Assume Y_S of size $s \times d$ be the projected points of S, $|S| = s \leq d$ then*

$$\frac{\det(\mathrm{Y_S Y_S^t})}{\det(\mathrm{SS^t})} \sim \prod_{i=1}^{s} \chi^2_{d-i+1}. \tag{4}$$

Proof. First, observe that if $X \sim \mathcal{N}_{n,d}(0, I_n \otimes I_d)$ then $Y_S = SX \sim \mathcal{N}_{s,d}(0, (SS^t) \otimes I_d)$. To see this argument, note that any linear (fixed) combination of Gaussian random variables is Gaussian from the stability of Gaussian. Now by the linearity of expectation we can easily show that every entry of SX has expected value zero. Also the correlation between two entries $E[(SX)_{ij}(SX)_{lk}] = E[(\sum_{r=1}^{d} S_{ir}X_{rj})\ (\sum_{r=i}^{d} S_{lr}X_{rk})]$ is zero if $j \neq k$, and $S_i^t S_l$ otherwise.

We know that $Y_S \sim \mathcal{N}_{s,d}(0, SS^t \otimes I_d)$. Assuming that S has linearly independent rows (otherwise both determinants are zero), there exists an s-by-s matrix R so that $SS^t = RR^t$ (Cholesky Decomposition).

Now we will evaluate $\det(\mathrm{R}^{-1}\mathrm{Y_S Y_S^t}(\mathrm{R^t})^{-1})$ in two different ways. First note that $R^{-1}, Y_S Y_S^t, (R^t)^{-1}$ are square matrices so

$$\det(\mathrm{R}^{-1}\mathrm{Y_S Y_S^t}(\mathrm{R^t})^{-1}) = \frac{\det(\mathrm{Y_S Y_S^t})}{(\det(\mathrm{R}))^2}. \tag{5}$$

Now note that $R^{-1}Y_S$ is distributed as $\mathcal{N}_{s,d}(0, R^{-1}SS^t(R^t)^{-1} \otimes I_d)$ which is equal to $\mathcal{N}_{s,d}(0, I_s \otimes I_d)$, since $R^{-1}SS^t(R^t)^{-1} = R^{-1}RR^t(R^t)^{-1} = I_s$. Lemma 1 with $R^{-1}Y_S$ implies that

$$\det(\mathrm{R}^{-1}\mathrm{Y_S Y_S^t}(\mathrm{R^t})^{-1}) \sim \prod_{\mathrm{i}=1}^{\mathrm{s}} \chi^2_{\mathrm{d-i+1}}. \tag{6}$$

Using the fact that $(\det(\mathrm{R}))^2 = \det(\mathrm{P_S P_S^t})$ with (5), (6) completes the proof. □

Remark: A different and simpler proof of the above lemma can be achieved by using the more abstract property of the projections, namely the rotationally invariance property. Consider two sets of s vectors, S and T. Assume for now that $W = \mathrm{span(S)} = \mathrm{span(T)}$. Then for *every* transformation ϕ it holds that $\det^2(\mathrm{A}) = \det(\phi(\mathrm{S})\phi(\mathrm{S})^{\mathrm{t}})/\det(\mathrm{SS^t}) = \det(\phi(\mathrm{T})\phi(\mathrm{T})^{\mathrm{t}})/\det(\mathrm{TT^t})$ where A is the $s \times s$ matrix that describes ϕ using any choice of basis for W and $\phi(W)$. To remove the assumption that $\mathrm{span(S)} = \mathrm{span(T)}$, simply consider a rigid tansformation ψ from span(S) to span(T). By rotational invariance of the projection, the distribution of the volume of $\phi(S)$ and that of $\phi(\psi(S))$ is the same, hence we reduce to the case where the span of the sets is the same subspace. Putting it together, this shows that the LHS of (4) distributes the same way for all sets of (linearly independent) vectors of size s, which by Lemma 1, must also be the same as the RHS of (4). We note that we have opted to use the previous proof since Gaussian projections is the tool of choice in our analysis throughout.

To conclude, Lemma 3 implies that the distribution of the volume of any subset of points is *independent* of their geometry up to a multiplicative factor. However, since we are interested in the distortion (fraction) of the volume $\mathrm{vol(Y_S)}/\mathrm{vol(P_S)} = \frac{(\det(\mathrm{Y_S Y_S^t}))^{1/2}/\mathrm{s!}}{(\det(\mathrm{P_S P_S^t}))^{1/2}/\mathrm{s!}} = \sqrt{\frac{\det(\mathrm{Y_S Y_S^t})}{\det(\mathrm{P_S P_S^t})}}$ everything boils down to the orthonormal case.

Notice that so far we proved that any subset of the regular set that *contains* the origin gives us a good enough concentration. Combining this fact with the previous Lemma we will show that the general case also holds. Let a subset $P_S = \{p_0, p_1, \ldots, p_{s-1}\}$ of P. We can translate the set P_S (since volume is translation-invariant) so that p_0 is at the origin, and call the resulting set $P'_S = \{0, p_1 - p_0, \ldots, p_{s-1} - p_0\}$. Now it is not hard to see that combining Lemmata 2,3 on the set P'_S, we get the following general result.

Theorem 3. *Let $0 < \varepsilon \leq 1/2$ and let $f : \mathbb{R}^n \to \mathbb{R}^d$ be the random embedding defined as above. Further, let S be an arbitrary subset of $\mathbb{R}^n$, with $|S| = s < \frac{d\varepsilon}{2}$. Then we have that*

$$\Pr\left[1 - \varepsilon < \left(\frac{\mathrm{vol(f(S))}}{\mathrm{vol(S)}}\right)^{\frac{1}{s-1}} < 1 + \varepsilon\right] \geq 1 - 2\exp\left(-s(d-(s-1))\frac{\varepsilon^2}{24}\right). \tag{7}$$

A closer look at the proof of Lemma 1 and Lemma 3 implies that the distance between any point and a subset of s points follows a Chi distribution with $d-s+1$ degrees of freedom. This fact can be used to simplify the proof for the preservation of affine distances as stated in [14], using the same number of dimensions.

5 Proof of the Main Theorem

We now prove the main theorem.

Proof. (of Theorem 1) Let B_S be the event: "The volume of the subset S of P distorts (under the embedding) its volume by more that $(1+\varepsilon)^{s-1}$". Clearly, the embedding fails if there is any S so that the event B_S occurs. We now bound the failure probability of the embedding from above

$$\Pr[\exists\, S:\ |S|<k,\ B_S] \leq \sum_{S:\ |S|<k} \Pr[B_S] \leq$$

$$2\sum_{s=1}^{k-1}\binom{n}{s}\exp\left(-s(d-(s-1))\frac{\varepsilon^2}{24}\right) \leq 2\sum_{s=1}^{k-1}\frac{n^s}{s^s}\exp\left(-s\left[(d-(s-1))\frac{\varepsilon^2}{24}-1\right]\right)$$

using the union bound, Theorem 3 for any subset of size $s<k$ and bounds on binomial coefficients, i.e. $\binom{n}{s} \leq \left(\frac{ne}{s}\right)^s$. Now if

$$2\sum_{s=1}^{k-1}\frac{n^s}{s^s}\exp\left(-s\left[(d-(s-1))\frac{\varepsilon^2}{24}-1\right]\right) < 1$$

then the probability that a random projection onto d dimesions doesn't distort the volume of any subset of size at most k by a relative error of ε, is positive.

Since $d > 2k/\varepsilon$, setting $d = 30\varepsilon^{-2}(\log n+1)+k-1 = O(\max\{k/\varepsilon, \varepsilon^{-2}\log n\})$ we get that, with positive probability, f has the desired property. □

6 Discussion

We have shown a nearly tight dimension reduction that approximately preserves volumes of sets of size up to k. The main outstanding gap is in the range where $k \geq \log n$ where the dimension required to obtain a k-volume respecting embedding is between k and k/ε. We conjecture that the upper bound we have is tight, and that the lower bound should come from a regular set of points. This conjecture can be phrased as the following linear algebraic statement.

Conjecture. Let A be an $n\times n$ positive semidefinite matrix such that the determinant of every $s\times s$ principal minor ($s\leq k$) is between $(1-\varepsilon)^{s-1}$ and 1. Then the rank of A is at least $\min\{\Omega(k/\varepsilon), n\}$.

We believe that closing gaps in questions of the type discussed above is particularly important as they will reaffirm a reccuring theme: the oblivious method of random Gaussian projections does as well as any other method. More interesting is to show that this is in fact not the case, and that sophisticated methods can go beyond this standard naive approach.

There is still a gap in our understanding with respect to dimension reduction that preserves all distances to affine subspaces spanned by small sets. Interestingly, this questions seems to be asking whether we can go beyond union bound reasoning when we deal with random projections. An example that captures this issue is a regular set where $\varepsilon < 1/k$. Here, it is implied by the proof in [14] that only $O(\varepsilon^{-2}\log n)$ dimensions are needed. However, the probability of failure for a particular event with this dimensionality is $n^{-O(1)}$, in other words not small enough to supply a proof simply by using the union bound. Does our technique extend to other dimension reduction techniques? Particularly, would projections onto ± 1 vectors provide the same dimension guarantees? Could Ailon and Chazelle's Fast JL transform substitute the original (dense) Gaussian matrix? As was mentioned in [14] the answer is yes when dealing with the weaker result that pays the extra factor of k, simply because the JL lemma is used as a "black box" there. We don't know what are the answers with respect to the stronger result of the current work, and we leave this as an open question.

References

1. Achlioptas, D.: Database-friendly random projections. In: PODS 2001: Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pp. 274–281. ACM, New York (2001)
2. Agarwal, P.K., Har-Peled, S., Yu, H.: Embeddings of surfaces, curves, and moving points in euclidean space. In: SCG 2007: Proceedings of the twenty-third annual symposium on Computational geometry, pp. 381–389. ACM, New York (2007)
3. Ailon, N., Chazelle, B.: Approximate nearest neighbors and the fast johnson-lindenstrauss transform. In: STOC 2006: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing, pp. 557–563. ACM, New York (2006)
4. Ailon, N., Liberty, E.: Fast dimension reduction using rademacher series on dual bch codes. In: SODA 2008: Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics, pp. 1–9 (2008)
5. Alon, N.: Problems and results in extremal combinatorics, i. Discrete Math. (273), 31–53 (2003)
6. Clarkson, K.L.: Tighter bounds for random projections of manifolds. In: SCG 2008: Proceedings of the twenty-fourth annual symposium on Computational geometry, pp. 39–48. ACM, New York (2008)
7. Engebretsen, L., Indyk, P., O'Donnell, R.: Derandomized dimensionality reduction with applications. In: SODA 2002: Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics, pp. 705–712 (2002)
8. Feige, U.: Approximating the bandwidth via volume respecting embeddings (extended abstract). In: STOC 1998: Proceedings of the thirtieth annual ACM symposium on Theory of computing, pp. 90–99. ACM, New York (1998)
9. Frankl, P., Maehara, H.: The johnson-lindenstrauss lemma and the sphericity of some graphs. J. Comb. Theory Ser. A 44(3), 355–362 (1987)
10. Gordon, L.: Bounds for the distribution of the generalized variance. The Annals of Statistics 17(4), 1684–1692 (1989)

11. Indyk, P., Naor, A.: Nearest-neighbor-preserving embeddings. ACM Trans. Algorithms 3(3), 31 (2007)
12. Johnson, W.B., Lindenstrauss, J.: Extensions of lipschitz mappings into a hilbert space. In: Amer. Math. Soc. (ed.) Conference in modern analysis and probability, pp. 189–206. Providence, RI (1984)
13. Liberty, E., Ailon, N., Singer, A.: Fast random projections using lean walsh transforms. In: RANDOM (to appear, 2008)
14. Magen, A.: Dimensionality reductions in ℓ_2 that preserve volumes and distance to affine spaces. Discrete & Computational Geometry 38(1), 139–153 (2007)
15. Matousek, J.: On the variants of johnson lindenstrauss lemma (manuscript) (2006)
16. Prekopa, A.: On random determinants i. Studia Scientiarum Mathematicarum Hungarica (2), 125–132 (1967)
17. Sarlos, T.: Improved approximation algorithms for large matrices via random projections. In: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science, Washington, DC, USA, pp. 143–152. IEEE Computer Society, Los Alamitos (2006)
18. Sivakumar, D.: Algorithmic derandomization via complexity theory. In: STOC 2002: Proceedings of the thirty-fourth annual ACM symposium on Theory of computing, pp. 619–626. ACM, New York (2002)
19. Vempala, S.: Random projection: A new approach to vlsi layout. In: FOCS 1998: Proceedings of the 39th Annual Symposium on Foundations of Computer Science, Washington, DC, USA, p. 389. IEEE Computer Society, Los Alamitos (1998)
20. Wakin, M.B., Baraniuk, R.G.: Random projections of signal manifolds. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP 2006, vol. 5, p. V (May 2006)

Appendix

Concentration Bounds for χ^2

Lemma 4 ([1]). *Let* $\chi_t^2 = \sum_{i=1}^t X_i^2$, *where* $X_i \sim \mathcal{N}(0,1)$. *Then for every* ε, *with* $0 < \varepsilon \leq 1/2$, *we have that*

$$\Pr\left[\chi_t^2 \leq (1-\varepsilon)E[\chi_t^2]\right] \leq \exp(-t\frac{\varepsilon^2}{6})$$

and

$$\Pr\left[\chi_t^2 \geq (1+\varepsilon)E[\chi_t^2]\right] \leq \exp(-t\frac{\varepsilon^2}{6}).$$