Euclidean Embeddings that Preserve Volumes

A thesis

submitted to the department of Computer Science

and the committee on graduate studies

of University of Toronto

in partial fulfillment of the requirements

for the degree of

Master of Science

Anastasios Zouzias

February 2009

I certify that I have read this thesis and that, in my opinion, it is fully adequate in scope and quality as a thesis for the degree of Master of Science.

(Avner Magen) Principal Adviser

I certify that I have read this thesis and that, in my opinion, it is fully adequate in scope and quality as a thesis for the degree of Master of Science.

(Allan Borodin)

Approved for the University Committee on Graduate Studies.

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 a factor of $1+\varepsilon$ at most. 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 $O(\max\{k/\varepsilon, \varepsilon^{-2}\log n\})$ 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 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.

Instead of fixing the distortion and finding the required dimension, we also consider

the setting where the dimension is fixed and try to find the best possible distortion. More formally, the question is what is the smallest possible, in the worst-case, relative change of the volume of subsets up to size k, when P is projected onto a fixed d-dimensional Euclidean space. This problem was first studied by Matoušek for sets of size 2, namely distances; he proved that there is a projection of P onto a subspace of fixed dimension d, with $3 \leq d \leq O(\log n)$, such that the distances change by a factor of $O(n^{2/d}\sqrt{\log n/d})$ at most. Here, we extend Matoušek's result and prove that there exists a mapping where not only pairs, but all subsets of at most $d/2$ points maintain their volume approximately with roughly the same (volume) distortion $O(n^{2/d}\log^{3/2} n)$. Note that our results are tight up to polylogarithmic factors, since $\Omega(n^{2/d})$ distortion is necessary even for sets of size two.

Acknowledgements

First and foremost I want to thank my adviser Avner Magen for his guidance, inspiration and patience during my last two years at university of Toronto. I also gratefully thank Allan Borodin for his comments and suggestions on an earlier draft of this thesis.

I would also like to thank the members of the theory group for their helpful discussions about my research. Many thanks go to Tasos Sidiropoulos for posing the question studied in Chapter 4.

Anastasios Zouzias

Contents

Chapter 1

Prior and Related Work

A classical result of Johnson and Lindenstrauss [JL84] shows that a set of n points in 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 to achieve a computational speedup, 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 geometric parameters other than distances between points, for example, the centre of gravity of a set 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 [Mag07], in which it was shown that it is possible to embed an n-point set of Euclidean space into 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 of as a natural normalization measure. Note that scaling a set of size s by a factor of $1+\varepsilon$ will change its volume by precisely this factor. In the current work, we improve this result by showing that $O(\max\{k/\varepsilon, \varepsilon^{-2}\log n\})$ dimensions suffice to get the same guarantee. Here, the improvement is achieved by analyzing the probability distribution of volumes of subsets under a random mapping directly, instead of preserving distances between points and affine subspaces by adding $O(n^k)$ additional points which was followed in [Mag07].

What if instead of fixing the required distortion and looking for a sufficiently large dimension, we fix the dimension and bound the distortion that may guarantee? This type of question was handled by Matoušek [Mat90] in the content of distances. Here we extend Matoušek's result by proving that there exists a mapping where not only pairs, but all subsets of at most $d/2$ points maintain their volume approximately within a relative factor of $O(n^{2/d}\log^{3/2} n)$. Also, there exists n-point metric spaces such that any embedding of them onto $\mathbb{R}^d$ has distortion $\Omega(n^{1/\lfloor (d+1)/2 \rfloor})$ [Mat90], and thus the above worst-case upper bound cannot be much improved; in particular, for every even dimension, it is tight up to the poly-logarithmic factor.

Other works have extended Johnson-Lindenstauss original work. From the computational perspective, emphasis was placed on derandomizing the embedding [EIO02, Siv02] and on speeding-up its computation. This last challenge has attracted considerable amount of attention. Achlioptas [Ach01] has shown that projection onto a (randomly selected) set of discrete vectors gives the same approximation guarantee, while using the same number

of dimensions. Ailon and Chazelle [AC06] supplied a method that uses *Sparse* Gaussian matrices for the projection to achieve fast computation (which they call "Fast Johnson Lindenstrauss Transform"). See also [Mat08, AL08, LAS08] for a related treatment and extensions. On a different branch of extensions, closer in flavour to our result, are works that require that the embeddings will preserve more structure of the geometry of the points. For example, in [AHPY07] the authors ask about distance between points that are *moving* according to some algebraically-limited curve; in [Sar06] about affine subspaces, in [IN07] about sets with bounded doubling dimension, and in [AHPY07, Cla08, WB06] about curves, (smooth) surfaces and manifolds.

Chapter 2

Preliminaries and Notation

In this chapter we introduce the notation that will be used throughout this thesis and prove some basic results about the Gaussian and Chi-square Distribution that will be used heavily in this thesis. The latter are of independent interest.

We present two basic notions that will be used heavily in this thesis. First we start by defining the (distance) distortion of a mapping, which roughly speaking measures how well a mapping preserves the metric structure of the point-set. Let $P \subset \mathbb{R}^n$. We say that a mapping $f : P \to \mathbb{R}^d$ has *distortion* D, if

$$\forall x, y \in P, \qquad \|x-y\|_2 \leq \|f(x)-f(y)\|_2 \leq D\|x-y\|_2.$$

Next, we define the (volume) distortion of a mapping between Euclidean spaces. As discussed above, there is a more elaborate notion of volume for an abstract metric space, but it is not useful here since, in our case, the host and target spaces are Euclidean and thus the notion of volume is well-defined. For a subset $S \subset \mathbb{R}^n$, we denote by $\mathrm{Vol}(S)$,

the $(|S|-1)$-dimensional volume of the convex hull of the set S in the standard Lebesgue measure sense. Clearly, if S is affinely dependent, it has zero volume. Also note that if $|S|=2$, then the volume is simply the distance between the two points.

Now we are ready to define the volume distortion of a mapping $f : P \to \mathbb{R}^d$. We say that a mapping f has *volume distortion* (k,D) if

$$\forall\ |S| \le k,\ S \subset P \qquad \mathrm{Vol}(S) \le \mathrm{Vol}(f(S)) \le D^{|S|-1} \cdot \mathrm{Vol}(S).$$

By the above remark about S of size 2, we have that in particular, pair-wise distances do not change by much.

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 in the standard Lebesgue measure sense. For $k=2$, notice that $\mathrm{Vol}(\{x,y\}) = d(x,y)$, for $k=3$ is the area of the triangle with vertices the points of the set, etc. Throughout this thesis, we denote the volume of a set S in the Euclidean space by $\mathrm{Vol}(S)$. We use dist(f) to denote the distortion of the mapping f.

We use $\|\cdot\|$ to denote the Euclidean norm. 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 $\text{vec}(A) = [a_{11}, \ldots, a_{r1}, a_{12}, \ldots, a_{r2}, \ldots, a_{1s}, \ldots, a_{rs}]^\top$ we denote the vectorization of the matrix A. We will use P_S to denote a subset S of rows of a matrix P. Let X, Y be arbitrary random variables. $X \sim \mathcal{N}(\mu, \sigma^2)$ denotes that X follows the normal distribution with mean μ and variance σ^2, and 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 $nd \times nd$. 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 $vec(X^\top) \sim \mathcal{N}_{nd}(vec(M^\top), \Sigma \otimes \Psi)$.*

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[1] between its entries has a specific structure; every row is correlated with respect to the Σ covariance matrix and every column with respect to Ψ. Hence, the correlation $\mathbb{E}[X_{ij}X_{lk}]$ of the entries X_{ij}, X_{lk} is equal to $\Sigma_{il} \cdot \Psi_{jk}$.

We now define a useful notion from probability theory.

Definition 2 (Stochastic domination)**.** *Let X and Y be two random variables, not necessarily on the same probability space. The random variable X is stochastically smaller than*

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

the random variable Y when, for every $x \in \mathbb{R}$, the inequality

$$\mathbf{Pr}(X \leq x) \geq \mathbf{Pr}(Y \leq x) \tag{2.1}$$

holds. We denote this by $X \preceq Y$.

This notion is useful when we have an "ordering" between two random variables, X and Y. Assume that for Y we know its distribution function and can compute its moments, and assume X has a complicated distribution function which is hard to deal with. Using this notion, we can overcome this obstacle and easily give tail bounds for the "hard" random variable X. See for example Theorem 2, which relates the product of the geometric mean of Chi-square random variables with a single Chi-square.

2.1 Stability of the Gaussian Distribution

Stable distributions is an interesting class of distributions its own right. Here, since the stability property of the Gaussian distribution is heavily used, we give a proof that the Gaussian distribution is 2-stable. Roughly speaking, for the Gaussian case, stability is the property that the Gaussian distribution is closed under linear combinations.

Lemma 1 (2-Stability of Gaussian)**.** *Let $X_i \sim \mathcal{N}(0,1)$ independent random variables, $a_i \in \mathbb{R}$ for $i = 1, \ldots, n$. If $Y = \sum_{i=1}^{n} a_i X_i$, then $Y \sim \mathcal{N}(0, \|a\|_2^2)$.*

Proof. The proof is done using the characteristic function of the Gaussian. First we

know [Fel71, Chapter XV, p. 476, Table 1] that the characteristic function (Fourier Transform of the distribution function) of a Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is

$$\varphi_X(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}.$$

Also we know that the distribution of the sum of independent random variables corresponds to the convolution of their distributions; therefore (by Fourier transform) the characteristic function of the sum is the product of their characteristic functions.

First note that $a_i X_i \sim \mathcal{N}(0, a_i^2)$. Therefore, by the above discussion and noting that X_i are independent we get that

$$\begin{aligned} \varphi_Y(t) &= \prod_{i=1}^{n} \varphi_{X_i}(t) = \prod_{i=1}^{n} e^{-\frac{a_i^2 t^2}{2}} \\ &= e^{-\sum_{i=1}^{n} a_i^2 t^2/2} = e^{-\frac{t^2}{2}\sum_{i=1}^{n} a_i^2} \\ &= e^{-\frac{\|a\|_2^2 t^2}{2}}. \end{aligned}$$

Since $\varphi_Y(t) = e^{-\frac{\|a\|_2^2 t^2}{2}}$, it follows that $Y \sim \mathcal{N}(0, \|a\|^2)$. □

2.2 Chi-square Distribution

In this section, we recall some results related to the Chi-square distribution and also give bounds on the Gamma function. Let $X_i,\ i = 1, \ldots, d$ be d independent, normally distributed random variables, then the random variable $\chi_d^2 = \sum_{i=1}^{d} X_i^2$ is a Chi-square random variable with d degrees of freedom. It is well known [Fel71, Chapter II, p. 47] that the Chi-square distribution is a special case of the Gamma distribution and its cumulative distribution

function is given by

$$\mathbf{Pr}[\chi^2_d \leq t] = \frac{\gamma(d/2, t/2)}{\Gamma(d/2)}, \tag{2.2}$$

where $\Gamma(x)$ is the Gamma function, $\gamma(a,x) = \int_0^x t^{a-1}e^{-t}\,dt$, $\Gamma(a,x) = \int_x^\infty t^{a-1}e^{-t}\,dt$ is the lower and upper *incomplete gamma function*, respectively. Next we present some bounds on Gamma and incomplete Gamma function that we need to prove our result in Chapter 4. We start by presenting the following bound on the Gamma function, see for instance [CD05, Lemmata 2.5, 2.6, 2.7] and [WW63, p.253] .

Lemma 2 (Stirling Bound on Gamma Function)**.** *If* $\Gamma(a) = \int_0^\infty e^{-t}t^{a-1}\,dt$*, where* $a > 0$*, then*

$$\sqrt{2\pi}a^{a+1/2}e^{-a} < \Gamma(a+1) < \sqrt{2\pi}a^{a+1/2}e^{-a+\frac{1}{12a}}, \tag{2.3}$$

and

$$\Gamma(a+\frac{1}{2}) < \Gamma(a)\sqrt{a}. \tag{2.4}$$

We next bound $\gamma(a,x)$ from above. Note that $\gamma(a,x) = \int_0^x t^{a-1}e^{-t}\,dt \leq \gamma(a,x) = \int_0^x t^{a-1}\,dt$, hence

$$\gamma(a,x) \leq x^a/a. \tag{2.5}$$

Now for the upper incomplete gamma, we have the following stronger bound.

Lemma 3. *If* $\Gamma(a,x) = \int_x^\infty e^{-t}t^{a-1}\,dt$ *where* $x > 2(a+1)$*, then*

$$\Gamma(a,x) < 2\exp(-x)x^{a+1}. \tag{2.6}$$

Proof. In [CD05, Lemma 2.6] set $\alpha = 1$ and $d = 2$. □

It is straight-forward to compute the first two moments of the Chi-square distribution,

$$\mathbb{E}[\chi_r^2] = r, \tag{2.7}$$

$$\mathrm{Var}[\chi_r^2] = 2r. \tag{2.8}$$

Next we show that the Chi-square distribution is concentrated around its' expected value; roughly speaking, a Chi-square random variable with r degrees of freedom has concentration $e^{-\Omega(r)}$ around its expected value. We formalize this statement with the following lemma, which is a folklore Chernoff type bound (see for example [Ach01, DG03]).

Lemma 4 (Concentration of Chi-square)**.** *Let* $\chi_r^2 = \sum_{i=1}^r X_i^2$*, where* $X_i \sim \mathcal{N}(0,1)$*. Then for every* ε*, with* $0 < \varepsilon \leq 1/2$*, we have that*

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

and

$$\mathbf{Pr}\left[\chi_r^2 \geq (1+\varepsilon)E[\chi_r^2]\right] \leq \exp(-r\frac{\varepsilon^2}{6})$$

Proof. Define $Y \sim \chi_r^2$, $\mathbb{E}[Y] = r$. We will compute the probability that Y deviates from its

expected value, i.e.,

$$\begin{aligned}
\mathbf{Pr}[Y \geq (1+\varepsilon)r] &= \mathbf{Pr}[e^{sY} \geq e^{s(1+\varepsilon)r}] \quad \text{(holds for all } s \geq 0) \\
&= \mathbf{Pr}[e^{sY} e^{-s(1+\varepsilon)r} \geq 1] \\
\text{(Markov Ineq.)} &\leq \mathbb{E}[e^{sY} e^{-sa}] \\
&= e^{-s(1+\varepsilon)r} \mathbb{E}[e^{s\sum_i^r X_i^2}] \\
&= e^{-s(1+\varepsilon)r} \mathbb{E}[\prod_{i=1}^{r} e^{sX_i^2}] \\
\text{(by independence)} &= e^{-s(1+\varepsilon)r} \prod_{i=1}^{r} \mathbb{E}[e^{sX_i^2}] \\
&= e^{-s(1+\varepsilon)r} \left(\mathbb{E}[e^{sX_i^2}]\right)^r \\
&= \left(e^{-s(1+\varepsilon)} \mathbb{E}[e^{sX_i^2}]\right)^r.
\end{aligned}$$

It remains to compute $\mathbb{E}[e^{sX_i^2}]$. We know that $X_i \sim N(0,1)$. For $s \geq 0$,

$$\begin{aligned}
\mathbb{E}[e^{sX_i^2}] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{st^2} e^{-t^2/2} dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-t^2(1-2s)/2} dt \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-2s}} \int_{\mathbb{R}} e^{-z^2/2} dz \quad \text{(Substituting } z = t\sqrt{1-2s} \text{ assuming } s < 1/2) \\
&= \frac{1}{\sqrt{1-2s}}.
\end{aligned}$$

We thus have

$$\mathbf{Pr}\left[Y \geq (1+\varepsilon)r\right] \leq \left(\frac{e^{-s(1+\varepsilon)}}{\sqrt{1-2s}}\right)^r.$$

Minimizing the above expression over s, gives $s = \frac{\varepsilon}{2(1+\varepsilon)}$. Substituting it into the above

expression, we obtain

$$\begin{aligned}
\mathbf{Pr}[Y \geq (1+\varepsilon)r] &\leq \left(\frac{e^{\varepsilon}}{1+\varepsilon}\right)^{-r/2} \\
&= \exp\left(-\frac{r}{2}(\varepsilon - \log(1+\varepsilon))\right) \quad \text{(Using Taylor expansion of } \log(1+\varepsilon)) \\
&= \exp\left(-\frac{r}{2}\left(\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3}\right)\right).
\end{aligned}$$

which gives our bound for $\varepsilon \leq 1/2$. A similar argument shows that $\mathbf{Pr}[Y \leq (1-\varepsilon)r] \leq \exp\left(-r\frac{\varepsilon^2}{6}\right)$. □

Chapter 3

Johnson-Lindenstrauss Lemma for Volumes

In this chapter, we improve the result of Magen [Mag07] for volumes, i.e., we prove that there exists an embedding with volume distortion $(k, O(\max\{k/\varepsilon, \varepsilon^{-2}\log n\}))$ for any n-point subset of Euclidean space. On the one hand, our analysis generalizes the analysis of Johnson and Lindenstrauss, which can be thought of as the special case of $k = 2$. On the other hand, our result shows that more geometric properties of the input points are approximately preserved: volumes of subsets of size up to $O(\log n/\varepsilon)$.

Theorem 1 (Main theorem)**.** *Let* $0 < \varepsilon \leq 1/3$ *and let* k, n, d *be positive integers, such that* $d = O(\max\{k/\varepsilon, \varepsilon^{-2}\log n\})$. *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,* $1 < |S| \leq k$*,*

$$1 \leq \left(\frac{Vol(f(S))}{Vol(S)}\right)^{\frac{1}{|S|-1}} \leq 1+\varepsilon.$$

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

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 [FM87] 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.

Notice that in this approach, the projection is simply the multiplication by a matrix (of appropriate dimensions) whose entries are i.i.d. Gaussian. 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 $\mathrm{Vol}(S)$, but does not depend on other properties of S. In other words, the fraction $\mathrm{Vol}^{\pi}(S)/\mathrm{Vol}(S)$ is a fixed distribution that does not depend on 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 two 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 counter-intuitive from a geometric point of view. It is 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 what's the concentration of the volume of 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 [Mag07] implicitly showed a concentration bound of only $e^{-\Omega(d)}$, which is one way to understand the improvement of the current work.

3.1 A Regular Set of Points Preserves its Volume

Assume that the set in Euclidean space for which 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 its projection is sufficiently concentrated, assuming s is small enough. Also denote by $X \sim \mathcal{N}_{n,d}(0, I_{nd})$ the projection matrix[1]. Note that our input set is the regular (identity matrix), therefore the image of its projection is simply X (the projection matrix); also recall that the points that correspond to S are represented by X_S. It is well known (see for example [FB90, pp. $220-235$]) that the volume of the *projected* points of S is

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

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

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 5 ([Pre67])**.** *Let $X \sim \mathcal{N}_{k,d}(0, I_{kd})$. The k-dimensional volume of the parallelotope determined by $X_{\{i\}}$, $i = 1, \ldots, k$ 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,*

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

Proof. Let $\Delta_d^{(k)} = \sqrt{\det(XX^\top)}$ 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 $W = \{x \in \mathbb{R}^d \mid x(k) = x(k+1) = \cdots = x(d) = 0\}$, i.e. the set of points that the $d-k+1$ last coordinates are equal to zero. Next we will show that $a_k \sim \chi_{d-k+1}$. Note that the distance from the point X_k to the

subspace that the rest $k-1$ points span[2] is equal to $\text{dist}(X_k, W) = \sqrt{\sum_{i=k}^{d} X_{ki}^2}$, 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 $\det(XX^\top) \sim \prod_{i=1}^{k} \chi_{d-i+1}^2$ 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 $(\det(X_S X_S^\top))^{\frac{1}{2s}}$ and so it 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, [Gor89])**.** *Let $u_i := \chi_{d-i+1}^2$ be independent Chi-square random variables for $i = 1, 2, \ldots, s$. If u_i are independent, then the following holds for every $s \geq 1$,*

$$\chi^2_{s(d-s+1)+\frac{(s-1)(s-2)}{2}} \succeq s \left(\prod_{i=1}^{s} u_i \right)^{1/s} \succeq \chi^2_{s(d-s+1)}. \tag{3.1}$$

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

Lemma 6. *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 $s \leq k < \frac{d\varepsilon}{2}$.*

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

Then we have that

$$\mathbf{Pr}\left[1-\varepsilon < \left(\frac{Vol(f(S))}{Vol(S)}\right)^{\frac{1}{s}} < 1+\varepsilon\right] \geq 1-2\exp\left(-s(d-(s-1))\frac{\varepsilon^2}{24}\right). \tag{3.2}$$

Proof. We define the random variable $Z = \left(\det(X_S X_S^\top)\right)^{1/s}$, $U = \frac{1}{s}\chi^2_{sd-\frac{s^2+s}{2}+1}$ 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, $d - \frac{s+1}{2} + 1/s \geq \mathbb{E}[Z] \geq d - s + 1$, with $\mathbb{E}[L] - \mathbb{E}[U] \geq -\frac{s}{2}$ for $s \geq 1$. Now we relate the volume of an arbitrary subset of P with the random variable Z. Using that $\mathrm{Vol}(f(S)) = \frac{\sqrt{\det(X_S X_S^\top)}}{d^{s/2}s!}$ and $\mathrm{Vol}(S) = \frac{1}{s!}$, we get for the upper tail

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

using that $d \geq \mathbb{E}[Z]$. Similarly for the lower tail

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

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

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

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

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

Now applying Lemma 4 on U, we get the bound

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

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

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

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

$$\mathbf{Pr}\left[Z - \mathbb{E}[Z] < -\varepsilon\mathbb{E}[Z]\right] \quad \leq \quad \mathbf{Pr}\left[L - \mathbb{E}[L] < -\varepsilon/2\mathbb{E}[L]\right]$$

holds. Therefore applying Lemma 4 on L we get

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

Comparing the upper and lower bound, the lemma follows. □

Remark: The bound on $k = O(d\varepsilon)$ 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

$$\frac{\mu_s}{1/s!} = 1 \pm O(\varepsilon), \tag{3.3}$$

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 3.3 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, $\frac{\mu_s}{1/s!} = \mathbb{E}[(\prod_{i=1}^{s} \chi^2_{d-i+1})^{\frac{1}{2s}}] \leq \left(\prod_{i=1}^{s} \mathbb{E}[\chi^2_{d-i+1}]\right)^{\frac{1}{2s}} \leq \sqrt[2s]{d(d-1)\dots(d-s+1)} \leq \sqrt{d-(s-1)/2}$ using independence, Jensen's inequality and arithmetic-geometric mean inequality. On the other hand[3], $\frac{\mu_1}{1/1!} = \mathbb{E}[\chi_d] \geq \sqrt{d-1}$. 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 $k \leq O(d\varepsilon)$.

3.2 Extension to the General Case

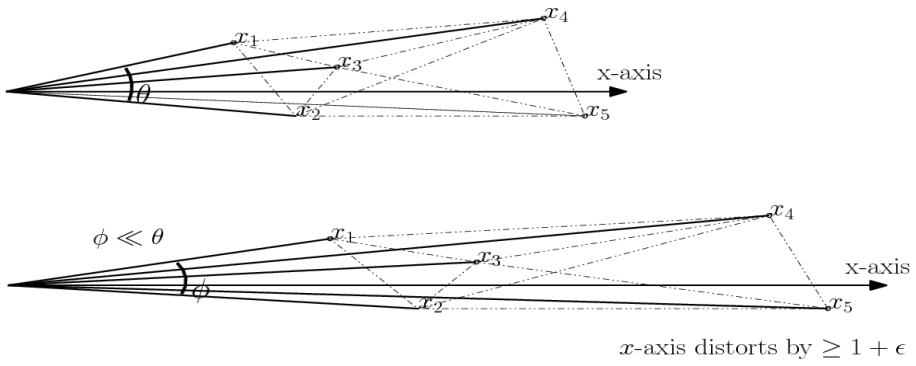

Figure 3.1: Example that illustrates the extension of the regular case to the general.

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

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 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 $\mathbf{Pr}[E] = \exp(-\Theta(d\varepsilon^2))$. 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 3.1). 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 7. *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

$$\frac{det(Y_S Y_S^\top)}{det(SS^\top)} \sim \prod_{i=1}^{s} \chi^2_{d-i+1}. \tag{3.4}$$

Proof. First, observe that if $X \sim \mathcal{N}_{n,d}(0, I_n \otimes I_d)$ then $Y_S = SX \sim \mathcal{N}_{s,d}(0, (SS^\top) \otimes I_d)$. 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 $\mathbb{E}[(SX)_{ij}(SX)_{lk}] = \mathbb{E}[(\sum_{r=1}^{d} S_{ir}X_{rj})\ (\sum_{r=i}^{d} S_{lr}X_{rk})]$ is zero if $j \neq k$, and $S_i^\top S_l$ otherwise.

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

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

$$\det(R^{-1}Y_S Y_S^\top (R^\top)^{-1}) = \frac{\det(Y_S Y_S^\top)}{(\det(R))^2}. \tag{3.5}$$

Now note that $R^{-1}Y_S$ is distributed as $\mathcal{N}_{s,d}(0, R^{-1}SS^\top(R^\top)^{-1} \otimes I_d)$ which is equal to $\mathcal{N}_{s,d}(0, I_s \otimes I_d)$, since $R^{-1}SS^\top(R^\top)^{-1} = R^{-1}RR^\top(R^\top)^{-1} = I_s$. Lemma 5 with $R^{-1}Y_S$ implies that

$$\det(R^{-1}Y_S Y_S^\top (R^\top)^{-1}) \sim \prod_{i=1}^{s} \chi^2_{d-i+1}. \tag{3.6}$$

Using the fact that $(\det(R))^2 = \det(SS^\top)$ with (3.5), (3.6) 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 $W = \text{span}(S) = \text{span}(T)$. Then for *every* transformation φ it holds that $\det^2(A) = \det(\varphi(S)\varphi(S)^\top)/\det(SS^\top) = \det(\varphi(T)\varphi(T)^\top)/\det(TT^\top)$ where A is the $s \times s$ matrix that describes φ using any choice of basis for W and $\varphi(W)$. To remove the assumption that $\text{span}(S) = \text{span}(T)$, simply consider a rigid transformation ψ from $\text{span}(S)$ to $\text{span}(T)$. By rotational invariance of the projection, the distribution of the volume of $\varphi(S)$ and that of $\varphi(\psi(S))$ 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 (3.4) have the same distibution for all sets of (linearly independent) vectors of size s, which by Lemma 5, must also be the same as the RHS of (3.4). 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 7 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 $\text{Vol}(Y_S)/\text{Vol}(P_S) = \frac{(\det(Y_S Y_S^\top))^{1/2}/s!}{(\det(P_S P_S^\top))^{1/2}/s!} = \sqrt{\frac{\det(Y_S Y_S^\top)}{\det(P_S P_S^\top)}}$ everything boils down to the orthonormal case.

Note that so far we proved that any subset of the regular set that *contains* the origin gives us sufficient 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 6,7 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*

$$\mathbf{Pr}\left[1-\varepsilon < \left(\frac{Vol(f(S))}{Vol(S)}\right)^{\frac{1}{s-1}} < 1+\varepsilon\right] \geq 1 - 2\exp\left(-s(d-(s-1))\frac{\varepsilon^2}{24}\right). \tag{3.7}$$

A closer look at the proof of Lemma 5 and Lemma 7 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 [Mag07], using the same number of dimensions.

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

$$\begin{aligned} \mathbf{Pr}[\exists\, S:\ |S| \leq k,\ B_S] &\leq \sum_{S:\ |S|\leq k} \mathbf{Pr}[B_S] \leq \\ 2\sum_{s=2}^{k}\binom{n}{s}\exp\left(-s(d-(s-1))\frac{\varepsilon^2}{24}\right) &\leq 2\sum_{s=2}^{k}\frac{n^s}{s^s}\exp\left(-s\left[(d-(s-1))\frac{\varepsilon^2}{24}-1\right]\right) \end{aligned}$$

using union bound, Theorem 3 for any subset of size $s \leq k$ and the inequality $\binom{n}{s} \leq \left(\frac{ne}{s}\right)^s$.

Now if

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

then the probability that a random projection onto d dimensions does not 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. Rescaling f by $\frac{1}{1-\varepsilon}$ completes the proof. □

Chapter 4

Low-Dimensional Volume Embeddings

In this chapter we begin by re-proving an $O(n^{2/d}\sqrt{\log n/d})$ upper bound on the distortion of embedding any n-point subset of Euclidean space into $\mathbb{R}^d$, with $3 \leq d \leq O(\log n)$. This was first proved in [Mat90] by projecting the points into $\mathbb{R}^d$ using a random d-dimensional subspace. Here the analysis follows the proof of [Mat90] and differs slightly in that we generate our random (linear) mapping via a Gaussian random matrix. This will be useful for the new result presented in this chapter: there exists an embedding that preserves the volumes of all subsets of size up to $\lfloor d/2 \rfloor$ ($k \leq \lfloor d/2 \rfloor$), with volume distortion $O(n^{2/d}\log^{3/2} n)$. Note that for $k = 2$ (distances), we get the same exponent $2/d$ as in [Mat90]; hence our results can be thought as a generalization of the usual distance embeddings. Also there exists n-point metric spaces that any embedding onto $\mathbb{R}^d$ has distortion $\Omega(n^{1/\lfloor (d+1)/2 \rfloor})$ [Mat90], and thus the above worst-case upper bound cannot be much improved; in particular, for every even dimension it is tight up to the poly-logarithmic factor.

4.1 Statement of Results

In this section, we prove that there exists an embedding into $\mathbb{R}^d$ with distortion at most $O(n^{2/d}\sqrt{\log n/d})$. We reproduce the proof of [Mat90], but instead of using a random d-dimension subspace of $\mathbb{R}^n$ we utilize a random Gaussian matrix, i.e., a matrix whose entries are i.i.d. Gaussian $\mathcal{N}(0,1)$.

Theorem 4. *[Mat90, Theorem 2.2] Let P be a n-point subset of $\mathbb{R}^n$ and let $3 \leq d \leq O(\log n)$. Then there is a mapping $f : P \mapsto \mathbb{R}^d$ with distortion $dist(f) = O(n^{2/d}\sqrt{\log n/d})$, i.e., there exists an absolute constant $c > 0$ such that*

$$\forall x, y \in P, \qquad \|x-y\| \leq \|f(x)-f(y)\| \leq cn^{2/d}\sqrt{\log n/d}\|x-y\|.$$

The extension of the above theorem is the following.

Theorem 5. *Let P be a n-point subset of $\mathbb{R}^n$ and let $3 \leq d \leq O(\log n)$. Then there is a mapping $f : P \mapsto \mathbb{R}^d$ such that*

$$\forall S \subset P, |S| \leq \lfloor d/2 \rfloor \quad 1 \leq \left(\frac{Vol(f(S))}{Vol(S)}\right)^{\frac{1}{|S|-1}} \leq c_1 n^{2/d}\log^{3/2} n$$

where $c_1 > 0$ is an absolute constant.

Proof of Theorem 4

Consider a random matrix R of size $d \times n$ with entries $r_{ij} \sim \mathcal{N}(0,1)$. Let $y = R \cdot x = f(x)$ for any vector $x \in \mathbb{R}^n$. The matrix R encodes our linear embedding f.

An important observation is that for every unit vector x and $i \in \{1, 2, \ldots, d\}$, the random variable y_i is normally distributed (see Section 2.1). Thus the random variable $\|y\|^2 = \sum_{i=1}^{d} y_i^2$ is distributed as a Chi-square random variable with d degrees of freedom. Our goal is to show that $\|y\|^2$ is sufficiently concentrated. More specifically, we have to show for the aforementioned random mapping f that the norm of its image (with domain an n-point set P) doesn't fall outside an interval $[a,b]$, i.e., $a \leq \|y\| \leq b$ with very high probability. This aims to upper bound the probabilities $Pr[\chi_d^2 \leq a^2]$ and $Pr[\chi_d^2 \geq b^2]$ (it is more convenient to work with the squares of the distances; note that the Chi-square is the squared ℓ_2-norm of a Gaussian random vector).

If X is a n-point subset of $\mathbb{R}^n$, the points of X determine at most $\binom{n}{2}$ distinct direction vectors. Let $a, b > 0$. Applying union bound over all pairs of vectors gives that if

$$\binom{n}{2} \left(\mathbf{Pr}\left[\chi_d^2 \leq a^2\right] + \mathbf{Pr}\left[\chi_d^2 \geq b^2\right] \right) < 1, \tag{4.1}$$

then there exists a projection f onto d dimensions, which expands every distance in X by at most a times and contracts at least b times (modulo a normalization factor), so dist(f) $\leq$ b/a.

First we bound $\Gamma(d/2)$ from below, which will be used for bounding both a, b. By

Lemma 2, we have that

$$\begin{aligned}\Gamma(d/2) &> \sqrt{2\pi}(d/2-1)^{d/2-1+1/2}\exp(-(d/2-1)) \\ &> \sqrt{2\pi}(\frac{d-2}{2})^{d/2-1/2}\exp(-d/2) \\ &> \sqrt{2\pi}\frac{\exp(-d/2)(d-2)^{(d-1)/2}}{2^{d/2-1/2}} \\ &> \frac{\exp(-d/2)(d-2)^{(d-1)/2}}{2^{d/2}}.\end{aligned}$$

Approximate a We will find a such that $\binom{n}{2}\mathbf{Pr}[\chi^2_d \leq a^2] < 1/2$. Using Equation 2.5 and the previous analysis we require that

$$\frac{n^2}{2}\frac{a^d}{e^{-d/2}(d-2)^{(d-1)/2}} < 1/2,$$

which holds if $a = c_1\frac{\sqrt{d}}{n^{2/d}}$, $c_1 > 0$ an absolute constant.

Approximate b Similarly as before, we will find b such that $\binom{n}{2}\mathbf{Pr}[\chi^2_d \geq b^2] < 1/2$. It suffices to bound $\mathbf{Pr}[\chi^2_d \geq b^2]$. Using Lemma 3, note that it must hold that $b^2 > 2d-2$, we have that

$$\begin{aligned}\mathbf{Pr}[\chi^2_d \geq b^2] &\leq \frac{e^{-b^2/2}(b^2/2)^{d/2-1}}{\Gamma(d/2)} \\ &\leq \frac{2e^{-b^2/2}b^{d-2}}{2^{d/2}e^{d/2}(d-2)^{(d-1)/2}} \\ &\leq \frac{b^{d-2}e^{-b^2/2-d/2}}{(d-2)^{(d-1)/2}}.\end{aligned}$$

Note that it suffices to show that $\ln\left(n^2 \frac{b^{d-2}e^{-b^2/2-d/2}}{(d-2)^{(d-1)/2}}\right)$ is negative.

$$\ln\left(n^2 \frac{b^{d-2}e^{-b^2/2-d/2}}{(d-2)^{(d-1)/2}}\right) \quad < \quad 2\ln n + (d-2)\ln b - b^2/2 - d/2 - \frac{d-1}{2}\ln(d-2).$$

Note that if $d > d'$ then $\mathbf{Pr}[\chi^2_{d'} \geq b^2] \leq \mathbf{Pr}[\chi^2_d \geq b^2]$, so we can assume that $d = \ln n$. Define $g(b,n) = 2\ln n + (d-2)\ln b - b^2/2 - d/2 - \frac{d-1}{2}\ln(d-2)$. We want to show that $g(b,\ n) < 0$ for large enough n. By choosing $b = 5\sqrt{\ln n}$, and recall that $d = \ln n$, we see that $\lim_{n\to\infty} g(5\sqrt{\ln n},\ n) = -\infty$ as desired.

Hence, we can choose a, b functions of n such that

$$\frac{b}{a} = \frac{5\sqrt{\log n}}{c_1 \frac{\sqrt{d}}{n^{2/d}}} = cn^{2/d}\sqrt{\log n/d},$$

where $c_1, c > 0$ are absolute constants. This concludes the proof of Theorem 4, since $\mathrm{dist(f)} \leq \mathrm{b/a} = O(\mathrm{n}^{2/\mathrm{d}}\sqrt{\log \mathrm{n/d}})$.

4.2 Volume Preservation

In this section, we consider the problem of preserving the volume of subsets of points when are projected onto a (fixed) d-dimensional Euclidean space. Our goal is to prove Theorem 5. As in the analysis in Chapter 1, we can assume w.l.o.g. that the input points are the regular set (see also the discussion in Section 3.2).

So let's see what happens when the input points form the regular n-dimensional simplex. It is not hard to see in this case that the projected points are independent random

Gaussian vectors and using Lemma 5 we have the following claim.

Claim 1. *Let $f : \{e_1, e_2, \ldots, e_n\} \mapsto \mathbb{R}^d$ be a Gaussian random embedding. Let S be a regular subset of $\{e_1, e_2, \ldots, e_n\}$ with $|S| = s < d$, then*

$$\left(\frac{Vol(f(S))}{Vol(S)}\right)^{\frac{2}{|S|-1}} \sim \left(\prod_{i=1}^{s-1} \chi^2_{d-i+1}\right)^{\frac{1}{s-1}}.$$

Recall that our goal is to find a mapping $f : P \to \mathbb{R}^d$ such that

$$\forall S \subset P, |S| \leq k \quad 1 \leq \left(\frac{\mathrm{Vol}(f(S))}{\mathrm{Vol}(S)}\right)^{\frac{1}{|S|-1}} \leq D, \tag{4.2}$$

where D is the volume distortion of the mapping.

Similarly to the previous section (distances), we will take a random mapping using a Gaussian random matrix and we will show that it satisfies the constraints of (4.2) with high probability.

First we start with the expansion, which as we will see gives the dominant term on the worst case bound of the volume distortion.

Expansion of Volumes

Lemma 8. *Fix any subset $S \subset P$ of size $|S| = s+1$ with $1 \leq s < k$. Then*

$$\mathbf{Pr}\left[\left(\frac{Vol(f(S))}{Vol(S)}\right)^{\frac{1}{|S|-1}} \leq a\right] \leq \sqrt{\frac{2}{\pi e}} \frac{(esa^2)^{t/2}}{t(t-2)^{(t-1)/2}},$$

where $t = s(d-s+1)$.

Proof. By translation of the points (volume is translation invariant) and Claim 1, we know that the above probability is equal to $\mathbf{Pr}\left[\left(\prod_{i=1}^{s}\chi^2_{d-i+1}\right)^{1/s} \leq a^2\right]$.

Note that from Theorem 2, we can bound the above probability of the product of Chi-square random variables with a single Chi-square. More specifically, we have the following inequality

$$\mathbf{Pr}\left[\left(\prod_{i=1}^{s}\chi^2_{d-i+1}\right)^{1/s} \leq a^2\right] \leq \mathbf{Pr}\left[\chi^2_{s(d-s+1)} \leq s\cdot a^2\right]$$

for $1 \leq s < k$.

Let define $t = s(d-s+1)$. Now, we have to deal with a single Chi-square random variable and thus we can bound it from above, similarly with the distance case, using Lemmas (2), (2.5). It follows that

$$\begin{aligned}\mathbf{Pr}\left[\chi^2_t \leq s\cdot a^2\right] &= \frac{\gamma(t/2, sa^2/2)}{\Gamma(t/2)} \\ &\leq \frac{(sa^2/2)^{t/2}}{\Gamma(t/2)t/2} \\ &\leq \sqrt{\frac{2}{\pi e}}\frac{(esa^2)^{t/2}}{t(t-2)^{(t-1)/2}}.\end{aligned}$$

□

Contraction of Volumes

Lemma 9. *Fix any subset $S \subset P$ of size $|S| = s+1$ with $1 \leq s < k$. Then*

$$\mathbf{Pr}\left[\left(\frac{Vol(f(S))}{Vol(S)}\right)^{\frac{1}{|S|-1}} \geq b\right] \leq \frac{e^{-\frac{sb^2-l+2}{2}}(sb^2)^{l/2+1}}{2\sqrt{\pi}(l-2)^{(l-1)/2}},$$

where $l = s(d-s+1) + \frac{(s-1)(s-2)}{2}$.

Proof. As in the previous Lemma the above probability is equal to $\mathbf{Pr}[\left(\prod_{i=1}^{s}\chi^2_{d-i+1}\right)^{1/s} \geq b^2]$, and again using Theorem 2 we can bound it using a single Chi-square random variable,

$$\mathbf{Pr}\left[\left(\prod_{i=1}^{s}\chi^2_{d-i+1}\right)^{1/s} \geq b^2\right] \leq \mathbf{Pr}\left[\chi^2_{s(d-s+1)+\frac{(s-1)(s-2)}{2}} \geq s\cdot b^2\right]$$

Let again define $l = s(d-s+1) + \frac{(s-1)(s-2)}{2}$. Hence

$$\begin{aligned}\mathbf{Pr}\left[\chi^2_l \geq s\cdot b^2\right] &= \frac{\Gamma(l/2, sb^2/2)}{\Gamma(l/2)} \\ &\leq \frac{2e^{-sb^2/2}(b^2 s)^{l/2+1}}{2^{l/2+1}\Gamma(l/2)} \\ &\leq \frac{e^{-\frac{sb^2-l+2}{2}}(sb^2)^{l/2+1}}{2\sqrt{\pi}(l-2)^{(l-1)/2}}.\end{aligned}$$

□

Union Bound for Expansion of Volumes Our goal is to find a such that with probability at least $1/2$, our embedding does not expands volumes of subsets of size up to k by

a factor a.

We will now apply union bound for all the sets of fixed size i, $1 \leq i \leq k$. We want to find a such that

$$\binom{n}{i+1}\sqrt{\frac{2}{\pi e}}\frac{(eia^2)^{t_i/2}}{t_i(t_i-2)^{(t_i-1)/2}} < \frac{1}{2k},$$

where $t_i = i(d-i+1)$. Note that if we sum over all different size of subsets ($i = 1,\ldots,k$) we get that the failure probability is at most $1/2$.

It suffices to show that $\ln\left(2k\binom{n}{i+1}\sqrt{\frac{2}{\pi e}}\frac{(eia^2)^{t_i/2}}{t_i(t_i-2)^{(t_i-1)/2}}\right)$ is negative for every $1 \leq i \leq k$.

$$\ln\left(2k\binom{n}{i+1}\sqrt{\frac{2}{\pi e}}\frac{(eia^2)^{t_i/2}}{t_i(t_i-2)^{(t_i-1)/2}}\right) < \ln 2 + \ln k + (i+1)\ln n + t_i \ln a \quad + \\ (t_i/2-i)\ln i + (t_i/2+i) - \ln t_i - (\frac{t_i-1}{2})\ln(t_i-2).$$

Let's group the terms of the right hand size and bound them individually.

- It is not hard to see that the following inequality

$$(t_i/2-i)\ln i - (\frac{t_i-1}{2})\ln(t_i-2) < 0$$

  holds for $i = 1,\ldots,k$ and for $d \geq 3$, since $t_i = i(d-i+1)$.

- Also we have that

$$\ln k - \ln t_i \leq 0,$$

since $k < t_i$ for all $i = 1, \ldots, k$.

- Finally, we have to show that the following quantity is negative, i.e.,

$$\ln 2 + (i+1)\ln n + t_i \ln a + (t_i/2 + i) < 0.$$

Define $a = c_e n^{-\gamma}$, for some positive γ that will be specified shortly and c_e a suitable constant. Recall that we want the above inequality to hold for every $1 \leq i \leq k$. We can choose the constant $0 < c_e < e^{-1}$ and take care of the $t_i/2 + i$ term. Let focus now on the dominate term $(i+1)\ln n$. It follows that the above quantity is negative if

$$\gamma \geq \frac{i+1}{i(d-i+1)}, \quad \text{for all } i = 1, \ldots, k.$$

Let's study closer the function $h(x) = \frac{x+1}{x(d-x+1)}$. We will show that $h(x)$ is convex on the domain $[1, d/2]$ and also is increasing in the domain $[d/4, d]$. The first and second derivatives of h with respect to x are :

$$\begin{aligned} h'(x) &= \frac{x^2 + 2x - d - 1}{x^2(d-x+1)^2} \\ h''(x) &= \frac{2(x^3 + 3x^2 - 3dx - 3x + d^2 + 2d + 1)}{x^3(d-x+1)}. \end{aligned}$$

It is not hard to see that $h''(x) > 0$ for $x \in [1, d]$ and $h'(x) > 0$ for $x \in [\frac{d}{4}, d]$ (details omitted), hence h is convex and increasing in $[d/4, d]$. Also note that $h(1) = h(d/2) = 2/d$. By convexity, we get that $h(x) \leq 2/d$ for all $x \in [1, d/2]$.

The above analysis gives a bound on the parameter k, i.e., the maximum size of subsets that we can consider. Thus, we get that k should be less than or equal to $\lfloor d/2 \rfloor$.

To sum up, we have proved that if $a = c_e n^{-2/d}$, with probability at most $1/2$ (over the randomness of our embedding) our embedding contracts the normalized volumes of subsets of size at most $d/2$ by more than a multiplicative factor of a, i.e.,

$$\mathbf{Pr}\left[\forall S \subset P, |S| \leq \lfloor d/2 \rfloor, \left(\frac{\mathrm{Vol}(f(S))}{\mathrm{Vol}(S)}\right)^{\frac{1}{|S|-1}} \leq a\right] < \frac{1}{2}.$$

Union Bound for Contraction of Volumes Our goal is to find b such that with probability at least $1/2$, our embedding does not contract volumes by more than a factor of b.

We apply union bound for all the sets of fixed size i, $1 \leq i \leq k$. We want to find b such that

$$\binom{n}{i+1} \frac{e^{-\frac{ib^2-l_i+2}{2}} (ib^2)^{l_i/2+1}}{2\sqrt{\pi}(l_i-2)^{(l_i-1)/2}} < \frac{1}{2k}.$$

where $l_i = i(d-i+1) + \frac{(i-1)(i-2)}{2}$. Note that if we sum over all different size of subsets we get the desired property with probability at least $1/2$.

It suffices to show that $\ln\left(2k\binom{n}{i+1} \frac{e^{-\frac{ib^2-l_i+2}{2}} (ib^2)^{l_i/2+1}}{2\sqrt{\pi}(l_i-2)^{(l_i-1)/2}}\right)$ is negative for every $1 \leq i \leq k$ and $d \in [3, \log n]$. Similarly with the distance case in the previous section, using the stochastic domination, we can assume w.l.o.g. that $d = \log n$, see previous section.

Now, since there are at most $\binom{n}{i+1} \leq \left(\frac{ne}{i+1}\right)^{i+1}$ subsets of size $i+1$, we need to show that

the following quantity is negative,

$$\ln\left(\frac{kn^{i+1}e^{-\frac{ib^2-l_i-2i}{2}}(ib^2)^{l_i/2+1}(i+1)^{(i+1)}}{\sqrt{\pi}(l_i-2)^{(l_i-1)/2}}\right) \leq \ln\left(\frac{kn^{i+1}e^{-\frac{ib^2-l_i-2i}{2}}i^{l_i/2-i+1}b^{l_i+2}}{(l_i-2)^{(l_i-1)/2}}\right) =$$

$$\ln k+(i+1)\ln n-\frac{ib^2-l_i-2i}{2}+(l_i/2-i+1)\ln i+(l_i-2)\ln b-(\frac{l_i-1}{2})\ln(l_i-2)<$$

$$(i+1)\ln n+(l_i/2-i+1)\ln i+(l_i-2)\ln b+l_i/2+2i+\ln k-\left(\frac{ib^2}{2}+\frac{l_i-1}{2}\ln(l_i-2)\right).$$

Note that in the last quantity the positive terms are $O(i\ln i\ln n)$. It is not hard to see that by choosing $b=c_2\log^{3/2}n$, $c_2>0$ a sufficient large constant, the above quantity will go to $-\infty$ as n grows for all $1\leq i\leq k$.

To sum up, we have proved that with probability at most $1/2$ (over the randomness of our embedding), we have that our embedding expands the normalized volumes by more than a multiplicative factor of b, i.e.,

$$\mathbf{Pr}\left[\forall S\subset P,|S|\leq k,\left(\frac{\mathrm{Vol}(f(S))}{\mathrm{Vol}(S)}\right)^{\frac{1}{|S|-1}}\geq b\right]<\frac{1}{2}.$$

Therefore we can find a,b with $a<b$ such that the relative change on the volumes is

$$\frac{b}{a}=\frac{c_2\log^{3/2}n}{c_e n^{-2/d}}=O(n^{2/d}\log^{3/2}n).$$

This concludes the proof of Theorem 5.

Chapter 5

Applications

Motivated by the minimum bandwidth problem, Feige [Fei98] introduced a combinatorial notion of *volume* of a metric, and described embeddings that approximately preserve volumes of subsets of a metric. Feige applied a volume-preserving embedding in an algorithm for minimizing bandwidth and obtained a polylogarithmic approximation. The previous best approximation was a trivial $\Theta(n)$. Subsequently, volume-preserving embeddings were used in [Vem98] to give approximation algorithms for VLSI layout and related problems. It turns out that Feige's embedding can be viewed as a sligh generalization of Bourgain's embedding. An immediate result of our work is that Feige's embedding can be assumed to use no more than $O(\log n)$ dimensions, compared to the $O(n)$ as in the original embedding or the $O(\log^2 n)$ bound that can be obtained by [Mag07].

Next we give an application of our volume preserving embedding of Chapter 3 to sampling-based algorithms that efficiently compute a low-rank approximation of a matrix.

The problem of low-rank approximation of a matrix is defined as follows: Given an $m \times n$ matrix A and an integer k, find a matrix B of rank at most k that minimizes

$\|A-B\|_F^2 = \sum_{i,j}(A_{ij}-B_{ij})$. This problem has received much attention in the past decade. Recall that the classical optimal solution to this problem is the matrix A_k consisting of the first k terms of the Singular Value Decomposition (SVD) of A:

$$A = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^\top$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ are the singular values and $\{u_i\}_1^n, \{v_i\}_1^n$ are the left and right singular vectors, respectively. Computing the SVD and hence the best low-rank approximation takes $O(\min\{mn^2, m^2n\})$ time [GvL96]. Although polynomial, is still too high for most applications.

In their seminal work, Frieze et al. [FKV04], and later Drineas et al. [DK03], introduced matrix sampling for fast low-rank approximation. These results introduced the sampling approach where rows of A are picked with probabilities proportional to their squared lengths and return an approximate solution with additive error. Achlioptas and McSherry [AM01] using a different approach gave similar results to the same problem, see also [AHK06] for a simplified proof.

These algorithms can also be viewed in the *streaming* model of computation [HRR99]. In this model, we do not have random access to data; the data comes as a stream of data items in arbitrary order and we are allowed one or a few sequential passes over the data. Algorithms for the streaming model have been designed for many problems including computing frequency moments, histograms, and others, which have mainly focused on what can be done in one pass. There has been some recent work on what can be done in multiple passes. The "pass-efficient" model of computation was introduced in [HRR99],

see also [DKM06].

Subsequently, Deshpande et al. [DRVW06, DV06] generalized the sampling appoach of Frieze et al. to sample subsets of rows proportional to the volume of their span. Using this approach, they construct a relative approximation algorithm for the low-rank matrix approximation problem, which then utilize to give the first polynomial time approximation scheme (PTAS) for the Projective Clustering problem. These results were later applied to solve a generalization of the same problem [DV07].

However, both of the relative error approximation algorithms do not fit in the pass-efficient model of computation, i.e., do not perform a constant number of sequential passes over the data. The need for perfoming multiple passes over the data is based on the fact that the algorithms approximate the distribution of the volume of subsets in an adaptive way, which force the algorithm to perform multiple passes over the data.

Now we will sketch an enhancement of the above relative error algorithms utilizing our volume preserving embeddings of Chapter 3, we will make them pass-efficient; more precisely, they will perform only three passes over the data. In the first pass, using Theorem 1 with ε a small constant we can store[1] the output of the embedding that preserves volumes of subsets. Next we execute the relative error algorithm on the ouput of the embedding and keep the indices of the rows that the algorithm selects. Then, in two more passes, we can compute and return the solution. This resolves an open question of [DV06] and also gives the first pass-efficient algorithm for the generalized Projective Clustering problem [DV07].

[1] To prevent confusion, we can do that since our model has additional sublinear RAM.

Chapter 6

Conclusions and Future Work

As was shown by Alon [Alo03], the upper bound on the dimensionality of the projection of Johnson-Lindenstauss is nearly tight[1], giving a lower bound of $\Omega(\varepsilon^{-2}\log n/\log(1/\varepsilon))$ dimensions. Further, in our setting, it is obvious that at least $k-1$ dimensions are needed; otherwise, the image of a set of size k will not be full-dimensional and will not therefore have a positive volume. These two facts provide a lower bound of $\Omega(\max\{\,\varepsilon^{-2}\log n/\log(1/\varepsilon), k\})$ dimensions, which makes our upper bound tight up to a factor of $1/\varepsilon$ throughout the whole range of the parameter k.

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 $k \geq \log n$ where the dimension required to obtain a k-volume respecting embedding is between k and k/ε. We conjecture that the k/ε 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.

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

Conjecture: Let A be an $n \times n$ positive semi-definite 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 recurring theme: the oblivious method of Gaussian random 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 [Mag07] 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.

We also re-proved an upper bound on the distortion of embedding any n-point subset of Euclidean space into a fixed d-dimensional Euclidean space, due to Matoušek, using a Gaussian random matrix. Then, we generalize this result and prove that there exists a mapping where not only pairs, but all subsets of at most $k \leq d/2$ points maintain their volume approximately with volume distortion $O(n^{2/d} \log^{3/2} n)$. Our results are tight up to polylogarithmic factors, since $\Omega(n^{2/d})$ distortion is necessary even for the distance case.

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 [Mag07], 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 them as open questions.

Bibliography

[AC06] N. Ailon and B. Chazelle. Approximate nearest neighbors and the fast johnson-lindenstrauss transform. In *STOC '06: Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, pages 557–563, New York, NY, USA, 2006. ACM. (Cited on page 3)

[Ach01] D. Achlioptas. Database-friendly random projections. In *PODS '01: Proceedings of the Twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 274–281, New York, NY, USA, 2001. ACM. (Cited on pages 2 and 10)

[AHK06] S. Arora, E. Hazan, and S. Kale. A fast random sampling algorithm for sparsifying matrices. In *APPROX-RANDOM*, pages 272–279, 2006. (Cited on page 40)

[AHPY07] P. K. Agarwal, S. Har-Peled, and H. Yu. Embeddings of surfaces, curves, and moving points in euclidean space. In *SCG '07: Proceedings of the Twenty-Third Annual Symposium on Computational Geometry*, pages 381–389, New

York, NY, USA, 2007. ACM. (Cited on page 3)

[AL08] N. Ailon and E. Liberty. Fast dimension reduction using rademacher series on dual *bch* codes. In *SODA '08: Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1–9, Philadelphia, PA, USA, 2008. Society for Industrial and Applied Mathematics. (Cited on page 3)

[Alo03] N. Alon. Problems and results in extremal combinatorics, i. *Discrete Math.*, 1(273):31–53, 2003. (Cited on page 42)

[AM01] D. Achlioptas and F. McSherry. Fast computation of low rank matrix approximations. In *STOC '01: Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 611–618, New York, NY, USA, 2001. ACM Press. (Cited on page 40)

[CD05] Z. Chen and J. J. Dongarra. Condition numbers of gaussian random matrices. *SIAM Journal on Matrix Analysis and Applications*, 27(3):603–620, 2005. (Cited on page 9)

[Cla08] K. L. Clarkson. Tighter bounds for random projections of manifolds. In *SCG '08: Proceedings of the twenty-fourth annual symposium on Computational geometry*, pages 39–48, New York, NY, USA, 2008. ACM. (Cited on page 3)

[DG03] S. Dasgupta and A. Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Struct. Algorithms*, 22(1):60–65, 2003. (Cited on page 10)

[DK03] P. Drineas and R. Kannan. Pass efficient algorithms for approximating large matrices. In *Proc. 14th ACM-SIAM Symp. on Discrete Algorithms (SODA'03)*, pages 223–232. SIAM, 2003. (Cited on page 40)

[DKM06] P. Drineas, R. Kannan, and M. W. Mahoney. Fast monte carlo algorithms for matrices i: Approximating matrix multiplication. *SIAM J. Comput.*, 36(1):132–157, 2006. (Cited on page 41)

[DRVW06] A. Deshpande, L. Rademacher, S. Vempala, and G. Wang. Matrix approximation and projective clustering via volume sampling. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 1117–1126, New York, NY, USA, 2006. ACM Press. (Cited on page 41)

[DV06] A. Deshpande and S. Vempala. Adaptive sampling and fast low-rank matrix approximation. In *APPROX-RANDOM*, pages 292–303, 2006. (Cited on page 41)

[DV07] A. Deshpande and K. R. Varadarajan. Sampling-based dimension reduction for subspace approximation. In *STOC*, pages 641–650, 2007. (Cited on page 41)

[EIO02] L. Engebretsen, P. Indyk, and R. O'Donnell. Derandomized dimensionality reduction with applications. In *SODA '02: Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 705–712, Philadelphia, PA, USA, 2002. Society for Industrial and Applied Mathematics. (Cited on page 2)

[FB90] J. B. Fraleigh and R. A. Beauregard. *Linear Algebra*. Addison-Wesley Press, 2 edition, 1990. (Cited on page 15)

[Fei98] U. Feige. Approximating the bandwidth via volume respecting embeddings (extended abstract). In *STOC '98: Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pages 90–99, New York, NY, USA, 1998. ACM. (Cited on page 39)

[Fel71] W. Feller. *An Introduction to Probability Theory and its Applications*, volume II. Wiley, New York, 1971. (Cited on page 8)

[FKV04] A. Frieze, R. Kannan, and S. Vempala. Fast monte-carlo algorithms for finding low-rank approximations. *J. ACM*, 51(6):1025–1041, 2004. (Cited on page 40)

[FM87] P. Frankl and H. Maehara. The johnson-lindenstrauss lemma and the sphericity of some graphs. *J. Comb. Theory Ser. A*, 44(3):355–362, 1987. (Cited on page 14)

[Gor89] L. Gordon. Bounds for the distribution of the generalized variance. *The Annals of Statistics*, 17(4):1684–1692, 1989. (Cited on page 17)

[GvL96] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins, third edition, 1996. (Cited on page 40)

[HRR99] M. Henzinger, P. Raghavan, and S. Rajagopalan. Computing on data streams. *External Memory Algorithms – DIMACS Series in Discrete Mathematics and Computer Science*, 50:107–118, 1999. (Cited on page 40)

[IN07] P. Indyk and A. Naor. Nearest-neighbor-preserving embeddings. *ACM Trans. Algorithms*, 3(3):31, 2007. (Cited on page 3)

[JL84] W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. In Amer. Math. Soc., editor, *In* Conference in Modern Analysis and Probability, pages 189–206, Providence, RI, 1984. (Cited on page 1)

[LAS08] E. Liberty, N. Ailon, and A. Singer. Dense fast random projections and lean walsh transforms. In *APPROX-RANDOM*, volume 5171 of *Lecture Notes in Computer Science*, pages 512–522. Springer, 2008. (Cited on page 3)

[Mag07] A. Magen. Dimensionality reductions in ℓ_2 that preserve volumes and distance to affine spaces. *Discrete & Computational Geometry*, 38(1):139–153, 2007. (Cited on pages 1, 2, 13, 15, 25, 39, 43 and 44)

[Mat90] J. Matoušek. Bi-lipschitz embeddings into low dimensional euclidean spaces. In *Comment. Math. Univ. Carolinae*, volume 31, pages 589–600, 1990. (Cited on pages 2, 27 and 28)

[Mat08] J. Matoušek. On variants of the johnson-lindenstrauss lemma. *Random Struct. Algorithms*, 33(2):142–156, 2008. (Cited on page 3)

[Pre67] A. Prekopa. On random determinants i. *Studia Scientiarum Mathematicarum Hungarica*, 1(2):125–132, July 1967. (Cited on page 16)

[Sar06] T. Sarlos. Improved approximation algorithms for large matrices via random projections. In *FOCS '06: Proceedings of the 47th Annual IEEE Symposium*

on Foundations of Computer Science, pages 143–152, Washington, DC, USA, 2006. IEEE Computer Society. (Cited on page 3)

[Siv02] D. Sivakumar. Algorithmic derandomization via complexity theory. In *STOC '02: Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 619–626, New York, NY, USA, 2002. ACM. (Cited on page 2)

[Vem98] S. Vempala. Random projection: A new approach to vlsi layout. In *FOCS '98: Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, page 389, Washington, DC, USA, 1998. IEEE Computer Society. (Cited on page 39)

[WB06] M.B. Wakin and R.G. Baraniuk. Random projections of signal manifolds. *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, 5:V–V, May 2006. (Cited on page 3)

[WW63] E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis*. Cambridge University Press, 4 edition, 1963. (Cited on page 9)