Optimal Bounds for Johnson-Lindenstrauss Transformations

Michael Burr   BURR2@CLEMSON.EDU
Shuhong Gao   SGAO@CLEMSON.EDU
Department of Mathematical Sciences
Clemson University, Clemson SC 29634, USA

Fiona Knoll   FKNOLL@G.CLEMSON.EDU
Department of Mathematical Sciences
University of Cincinnati, Cincinnati, OH 45221, USA

Editor: Kenji Fukumizu

Abstract

In 1984, Johnson and Lindenstrauss proved that any finite set of data in a high-dimensional space can be projected to a lower-dimensional space while preserving the pairwise Euclidean distances between points up to a bounded relative error. If the desired dimension of the image is too small, however, Kane, Meka, and Nelson (2011) and Jayram and Woodruff (2013) proved that such a projection does not exist. In this paper, we provide a precise asymptotic threshold for the dimension of the image, above which, there exists a projection preserving the Euclidean distance, but, below which, there does not exist such a projection.

Keywords: Johnson-Lindenstrauss transformation, Dimension reduction, Phase transition, Asymptotic threshold

1. Introduction

In 1984, Johnson and Lindenstrauss (1984), in establishing a bound on the Lipschitz constant for the Lipschitz extension problem, proved that any finite set of data in a high-dimensional space can be projected into a lower-dimensional space while preserving the pairwise Euclidean distance within any desired relative error. In particular, for any finite set of vectors $x_1, \ldots, x_N \in \mathbb{R}^d$ and for any error factor $0 < \epsilon < \frac{1}{2}$, there exists an absolute constant c such that for all $k \geq c\epsilon^{-2} \log N$, there exists a linear map $A : \mathbb{R}^d \to \mathbb{R}^k$ such that for all pairs $1 \leq i, j \leq N$,

$$(1 - \epsilon)\|x_i - x_j\|_2 \leq \|Ax_i - Ax_j\|_2 \leq (1 + \epsilon)\|x_i - x_j\|_2,$$

where $\|\cdot\|_2$ denotes the Euclidean norm. These inequalities are implied by the following theorem (by setting $\delta = \frac{1}{N^2}$ and using the union bound):

Theorem 1 (Johnson and Lindenstrauss (1984)) *For any real numbers $0 < \epsilon, \delta < \frac{1}{2}$, there exists an absolute constant $c > 0$ such that for any integer $k \geq c\epsilon^{-2} \log \frac{1}{\delta}$, there exists a probability distribution $\mathcal{D}$ on $k \times d$ real matrices such that for any fixed $x \in \mathbb{R}^d$,*

$$\mathrm{Prob}_{A \sim \mathcal{D}} \left[(1 - \epsilon)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \epsilon)\|x\|_2^2 \right] > 1 - \delta, \tag{1}$$

where $A \sim \mathcal{D}$ indicates that the matrix A is a random matrix with distribution $\mathcal{D}$.

Note that, in order to project a large number of vectors, δ must be sufficiently small. For instance, suppose we wish to project a set of $N = 2^{20}$ vectors to a smaller dimensional space. To apply the union bound to Inequality (1), we use any $\delta < 2^{-39}$. In this case, Inequality (1) implies that the probability of preserving all pairwise distances between N points (up to a relative error of ϵ) is at least $1 - \delta N^2/2 > 0$. Since the probability is nonzero, such a projection exists.

A probability distribution $\mathcal{D}$ satisfying Inequality (1) is called an (ϵ, δ)-JL distribution, or simply a JL distribution. Since these transformations are linear, without loss of generality, we assume for the rest of the paper that $\|x\|_2 = 1$. When a JL distribution is specified via an explicit construction, we may call a random projection $x \mapsto Ax$ generated in this way a JL transformation.

Since the introduction of JL distributions, there has been considerable work on explicit constructions of JL distributions, see, e.g., Johnson and Lindenstrauss (1984); Frankl and Maehara (1988); Indyk and Motwani (1998); Achlioptas (2003); Ailon and Chazelle (2006); Matousek (2008); Dasgupta et al. (2010); Kane and Nelson (2014); and the references therein. A simple and easily described JL distribution is that of Achlioptas (2003). In this construction, the entries of A are distributed as follows:

$$
a_{ij} = \begin{cases} k^{-1/2}, & \text{with probability } 1/2, \\ -k^{-1/2}, & \text{with probability } 1/2. \end{cases}
$$

The recent constructions in Ailon and Chazelle (2006); Matousek (2008); Dasgupta et al. (2010); Kane and Nelson (2014) have focused on the complexity of computing a projection for the purpose of applications. We note that the ability to project a vector to a smaller dimensional space, independent of the original dimension, while preserving the Euclidean norm up to a prescribed relative error, is highly desirable. In particular, dimension reduction has applications in many fields, including machine learning (Arriaga and Vempala (2006); Weinberger et al. (2009)), low rank approximation (Clarkson and Woodruff (2013); Nguyen et al. (2009); Ubaru et al. (2017)), approximate nearest neighbors (Ailon and Chazelle (2006); Indyk and Motwani (1998)), data storage (Candès (2008); Cormode and Indyk (2016)), and document similarity (Bingham and Mannila (2001); Lin and Gunopulos (2003)).

For both practical and theoretical purposes, it is important to know the smallest possible dimension k of a potential image space for any given ϵ and δ. Note that, for any $d_1 < d$, each (ϵ, δ)-JL distribution $\mathcal{D}$ on $\mathbb{R}^{k \times d}$ induces an (ϵ, δ)-JL distribution $\mathcal{D}_1$ on $\mathbb{R}^{k \times d_1}$ in a natural way: the matrices of $\mathcal{D}_1$ are obtained from $\mathcal{D}$ by deleting the last $d - d_1$ columns of a random matrix, together with the induced probability distribution. This construction is a JL distribution since $\mathbb{R}^{d_1}$ can be naturally embedded into $\mathbb{R}^d$ by extending a vector in $\mathbb{R}^{d_1}$ by $d - d_1$ zeros. Hence, if there exists an (ϵ, δ)-JL distribution on $\mathbb{R}^{k \times d}$, then there is an (ϵ, δ)-JL distribution on $\mathbb{R}^{k \times d_1}$ for all $1 \le d_1 \le d$. Similarly, if an (ϵ, δ)-JL distribution does not exist on $\mathbb{R}^{k \times d}$, then, for any $k_1 < k$, there cannot be an (ϵ, δ)-JL distribution on $\mathbb{R}^{k_1 \times d}$. In particular, since $\mathbb{R}^{k_1}$ can be naturally embedded into $\mathbb{R}^k$ by extending a vector in $\mathbb{R}^{k_1}$ by $k - k_1$ zeros, if an (ϵ, δ)-JL distribution existed for $\mathbb{R}^{k_1 \times d}$, it could be extended to an (ϵ, δ)-JL distribution for $\mathbb{R}^{k \times d}$.

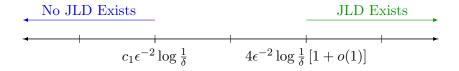Figure 1: For fixed ϵ and δ, there exists a JL distribution for $k \geq 4\epsilon^{-2} \log(1/\delta) [1 + o(1)]$. For $k < c_1 \epsilon^{-2} \log(1/\delta)$, for some absolute constant $c_1 > 0$, there is no JL distribution. In this paper, we close this gap in the limit.

For any ϵ and δ, we define

$$k_0(\epsilon, \delta) = \min\{k : \text{there exists an } (\epsilon, \delta)\text{-JL distribution on } \mathbb{R}^{k \times d} \text{ for every } d \geq 1\}.$$

By our definition, $k_0 = k_0(\epsilon, \delta)$ is independent of d, and, by Theorem 1, we have that $k_0 \leq c\epsilon^{-2} \log(1/\delta)$ for some absolute constant $c > 0$. Frankl and Maehara (1988) show that $c \leq 9$. Achlioptas (2003) further improves this bound by providing a JL distribution with

$$k > 2 \log(2/\delta) \left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3} \right)^{-1},$$

resulting in the following upper bound:

$$k_0 \leq 2 \log(2/\delta) \left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3} \right)^{-1} = 4\epsilon^{-2} \log(1/\delta) [1 + o(1)],$$

where $o(1)$ approaches zero as both ϵ and δ approach zero.

A lower bound on k_0 was not given until 2003 when Alon (2003) proved that

$$k_0 \geq c\epsilon^{-2} \log(1/\delta) \Big/ \log(1/\epsilon)$$

for some absolute constant $c > 0$. Improving Alon's work, Jayram and Woodruff (2013) and Kane et al. (2011) showed, through different methods, that, for some absolute constant $c_1 > 0$, there is no (ϵ, δ)-JL distribution for $k \leq c_1 \epsilon^{-2} \log \frac{1}{\delta}$. Hence, there is a lower bound of the form $k_0 \geq c_1 \epsilon^{-2} \log \frac{1}{\delta}$. This situation is summarized in Figure 1.

The goal of the current paper is to close the gap between the upper and lower bounds in the limit. In particular, we prove an optimal lower bound that asymptotically matches the known upper bound when ϵ and δ approach 0, see Theorem 2. This means that $4\epsilon^{-2} \log(1/\delta)$ is an asymptotic threshold for k_0 where a phase change phenomenon occurs. The main result of this paper is captured in the following theorem:

Theorem 2 *For ϵ and δ sufficiently small, $k_0 \approx 4\epsilon^{-2} \log(1/\delta)$. More precisely,*

$$\lim_{\epsilon, \delta \to 0} \frac{k_0(\epsilon, \delta)}{4\epsilon^{-2} \log(1/\delta)} = 1.$$

3

The rest of the paper is organized as follows: To prove Theorem 2, we follow the approach of Kane et al. (2011). To make their constant c_1 explicit, however, we must use a more careful argument. In Section 2, we provide explicit conditions under which we prove the main result, Theorem 2. We delay the proofs of these conditions until Section 3 in order to make the main result more accessible, since only the statements of these results are needed and not their more technical proofs. In Section 3, we prove probabilistic bounds on $s = x_1^2 + \cdots + x_k^2$ where $x = (x_1, \cdots, x_k, \cdots, x_d)$ is a random variable uniformly distributed on S^{d-1}. These bounds explicitly determine the hidden constants in the bounds on $\mathrm{Prob}\,[s < s_0(1 - \epsilon)]$ and $\mathrm{Prob}\,[s > s_0(1 + \epsilon)]$ from (Kane et al., 2011, Theorem 20). These bounds can be viewed as explicit bounds for concentration theorems for laws of large numbers from probability theory. In the Appendix, we provide an alternate proof of a result in Kane et al. (2011), showing that if $d\Omega_{d-1}$ is the surface area for the $(d-1)$-dimensional sphere S^{d-1}, then for any $1 \le k \le d$,

$$d\Omega_{d-1} = \frac{1}{2} f(s) ds \, d\Omega_{k-1} d\Omega_{d-k-1},$$

where $s \in [0,1]$ and $f(s) = s^{\frac{k-2}{2}}(1 - s)^{\frac{d-k-2}{2}}$.

2. Asymptotic Threshold Bound

In this section, we prove the asymptotic threshold bound for JL transformations. In particular, we provide specific conditions that result in the asymptotic threshold bound of $4\epsilon^{-2}\log(1/\delta)$. In Section 3, we prove that these conditions hold, but the details of these proofs are more technical and are unnecessary to understand the main results of this paper.

2.1. The Uniform Distribution on S^{d-1}

There is a unique probability distribution, called the uniform distribution, on the $(d-1)$-dimensional sphere S^{d-1} that is invariant under the orthonormal group. We express the uniform distribution on S^{d-1} in terms of the surface area differential form $d\Omega_{d-1}$, which means that, for any measurable subset $V \subset S^{d-1}$, the $(d-1)$-dimensional surface area of V is equal to the integral with respect to $d\Omega_{d-1}$, i.e., $\mathrm{Vol}_{d-1}(V) = \int_V d\Omega_{d-1}$.[1] Thus, the uniform distribution on S^{d-1} corresponds to the measure $d\Omega_{d-1}/\mathrm{Vol}_{d-1}\left(S^{d-1}\right)$.

As we are interested in reducing a d-dimensional vector to a k-dimensional vector for $1 \le k < d$, we derive a relationship between the uniform distribution on S^{d-1} and the uniform distributions on S^{k-1} and S^{d-k-1}. Following the approach of Kane et al. (2011), for $1 \le k < d$, we define an injective map

$$\Psi : S^{d-1} \to [0,1] \times S^{k-1} \times S^{d-k-1}$$

as follows: For any $x = (x_1, x_2, \ldots, x_d)^t \in S^{d-1}$, we define s in $\Psi(x) = (s, u, v)$ as $s = x_1^2 + \cdots + x_k^2$. In the case where $0 < s < 1$, we define

$$u = (x_1, \ldots, x_k)^t / \sqrt{s} \quad \text{and} \quad v = (x_{k+1}, \ldots, x_d)^t / \sqrt{1 - s}.$$

---

1. In this paper, we suppress the pullback maps on equalities for differential forms since there is a unique (almost) bijective map under consideration in each case. We leave the details to the interested reader.

When $s = 0$, i.e., $x_1 = \cdots = x_k = 0$, we define $u = (1, 0, \ldots, 0)^t$ (or any point in S^{k-1}) and $v = (x_{k+1}, \ldots, x_d)^t$. Similarly, for $s = 1$, we define $u = (x_1, \ldots, x_k)^t$ and $v = (1, 0, \ldots, 0)^t$ (or any point in S^{d-k-1}). It is straight-forward to check that Ψ is injective. In addition, the complement of the image of Ψ is a subset of $\{0, 1\} \times S^{k-1} \times S^{d-k-1}$, which has, in turn, $(d-1)$-dimensional surface area equal to 0. Therefore, when convenient, we assume that $s \in (0, 1)$.

For $s \in [0, 1]$, we define
$$f(s) = s^{\frac{k-2}{2}}(1-s)^{\frac{d-k-2}{2}}.$$

In the Appendix, we provide alternative proof to the computation in Kane et al. (2011) which shows that, via the map Ψ,

$$d\Omega_{d-1} = \frac{1}{2}f(s)ds\, d\Omega_{k-1}d\Omega_{d-k-1}.$$

Equivalently, in term of probability distributions,

$$\frac{d\Omega_{d-1}}{\mathrm{Vol}_{d-1}\left(S^{d-1}\right)} = Bf(s)ds\frac{d\Omega_{k-1}}{\mathrm{Vol}_{k-1}\left(S^{k-1}\right)}\frac{d\Omega_{d-k-1}}{\mathrm{Vol}_{d-k-1}\left(S^{d-k-1}\right)}, \tag{2}$$

where
$$B = \frac{1}{2}\frac{\mathrm{Vol}_{k-1}\left(S^{k-1}\right)\mathrm{Vol}_{d-k-1}\left(S^{d-k-1}\right)}{\mathrm{Vol}_{k-1}\left(S^{k-1}\right)} = \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{k}{2})\Gamma(\frac{d-k}{2})}.$$

is an appropriate scaling constant, depending on the gamma function, for more details, see the Appendix. Moreover, in this situation, we observe that $Bf(s)$ is a probability distribution on $[0, 1]$. This implies that the uniform distribution on S^{d-1} is a direct product of the distributions on the factors. In other words, a uniformly distributed random variable X_{d-1} on S^{d-1} can be decomposed into three random variables $\Psi(X_{d-1}) = (S, X_{k-1}, X_{d-k-1})$ with the following properties:

(i) S is a random variable on $[0, 1]$ with density function $Bf(s)$,

(ii) X_{k-1} and X_{d-k-1} are uniformly distributed on S^{k-1} and S^{d-k-1}, and

(iii) The random variables S, X_{k-1}, and X_{d-k-1} are *independent*.

The independence of these three random variables is a key property in our proof as it allows us to study the three spaces separately.

2.2. Upper Bound: Explicit JL Distribution

We recall that Achlioptas (2003) proved that

$$k_0(\epsilon, \delta) \le 2\log(2/\delta)\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)^{-1} = 4\epsilon^{-2}\log(1/\delta)\left[1 + o(1)\right].$$

In this section, we give an alternate proof of this result using the approach and bounds from this paper.

We recall the following construction by Dasgupta and Gupta (2003): A distribution $\mathcal{D}$ on $k \times d$ matrices is formed by picking a $d \times d$ orthonormal matrix $V = (v_1, \ldots, v_d)^t$ uniformly

at random with respect to the Haar measure on orthonormal matrices and then letting $A = \frac{1}{\sqrt{s_0}}(v_1, \ldots, v_k)^t$, where $s_0 = k/d$. The following proposition shows that $k_0(\epsilon, \delta) \leq 4\epsilon^{-2} \log(1/\delta) [1 + o(1)]$, which, in turn, implies that the limit appearing in Theorem 2 (if it exists) is at most 1:

Proposition 3 *Let $0 < \epsilon, \delta < \frac{1}{2}$ and $s_0 = k/d$. Suppose that there is some constant C so that*

$$\max\{\mathrm{Prob}_{x \sim S^{d-1}}[s < s_0(1-\epsilon)], \mathrm{Prob}_{x \sim S^{d-1}}[s > s_0(1+\epsilon)]\} \leq Ce^{-\frac{k-2}{4}\epsilon^2\left(1-\frac{2}{3}\epsilon\right)},$$

where $s = x_1^2 + \cdots + x_k^2$ is defined as in $\Psi(x) = (s, u, v)$. Then, there exists an $o(1)$ function, which approaches zero as both ϵ and δ approach zero so that if $k > 4\epsilon^{-2} \log\left(\frac{1}{\delta}\right) [1 + o(1)]$, then the distribution on $k \times d$ random matrices defined as above is an (ϵ, δ)-JL distribution, that is, for any $w \in S^{d-1}$,

$$\mathrm{Prob}_{A \sim \mathcal{D}}\left[\left|\|Aw\|_2^2 - 1\right| < \epsilon\right] \geq 1 - \delta.$$

Proof Let V be the random orthogonal matrix as defined above, and let $x = (x_1, \ldots, x_d)^t = Vw$. Then $Aw = \sqrt{s_0^{-1}}(x_1, \ldots, x_k)^t$, and

$$\|Aw\|_2^2 = \frac{1}{s_0}(x_1^2 + \cdots + x_k^2).$$

Since V is orthonormal and $\|w\|_2 = 1$, we have that $\|x\|_2 = 1$, and, hence, $x \in S^{d-1}$. We observe that since V is a random orthogonal matrix, for fixed $w \in S^{d-1}$, $x = Vw$ is a random variable, uniformly distributed on S^{d-1}. Hence,

$$\mathrm{Prob}_{A \sim \mathcal{D}}\left[\left|\|Aw\|_2^2 - 1\right| > \epsilon\right] = \mathrm{Prob}_{x \sim S^{d-1}}\left[\left|\frac{1}{s_0}\sum_{i=1}^{k} x_i^2 - 1\right| > \epsilon\right],$$

where $x \sim S^{d-1}$ means that x is a random variable uniformly distributed on S^{d-1}. Let $s = \sum_{i=1}^{k} x_i^2$. Then, $s \in [0, 1]$ and the probability above becomes

$$\mathrm{Prob}_{x \sim S^{d-1}}[s < s_0(1-\epsilon)] + \mathrm{Prob}_{x \sim S^{d-1}}[s > s_0(1+\epsilon)] \leq 2Ce^{-\frac{k-2}{4}\epsilon^2\left(1-\frac{2}{3}\epsilon\right)}, \qquad (3)$$

by assumption. We observe that when

$$k > 4\epsilon^{-2} \log\left(\frac{1}{\delta}\right)\left[1 + \frac{2\epsilon}{3 - 2\epsilon} + \frac{\log(2C)}{\log\left(\frac{1}{\delta}\right)}\frac{1}{1 - 2\epsilon/3} + \frac{2\epsilon^2}{4\log\left(\frac{1}{\delta}\right)}\right] = 4\epsilon^{-2} \log\left(\frac{1}{\delta}\right)[1 + o(1)],$$
$$(4)$$

the right-hand-side of Inequality (3) is less than δ. In this case, the $o(1)$ term needed in the theorem statement appears in Inequality (4). Therefore, when $k > 4\epsilon^{-2} \log\left(\frac{1}{\delta}\right)[1 + o(1)]$, the distribution $\mathcal{D}$ is an (ϵ, δ)-JL distribution. ∎

We observe that this result also follows from work in Dasgupta and Gupta (2003). In particular, (Dasgupta and Gupta, 2003, Lemma 2.2) can be used to derive bounds similar to the assumed bounds in this statement.

2.3. Lower Bound for Arbitrary Distributions

In this section, we prove an optimal lower bound on the limit in Theorem 2 that matches the upper bound from the previous section. The proof of this lower bound is the main challenge in this paper. We begin with the following key lemma:

Lemma 4 *Let $x = (x_1, \ldots, x_d)^t$ be a random variable, uniformly distributed on S^{d-1}, $\Psi(x) = (s, u, v)$, and $s_0 = k/d$. Suppose that for a fixed $\epsilon > 0$,*

$$\min\{\mathrm{Prob}\left[s > s_0(1 + \epsilon)\right], \mathrm{Prob}\left[s < s_0(1 - \epsilon)\right]\} \geq L,$$

where s is a random variable with probability distribution $Bf(s)$ on $[0, 1]$. For any function $c(u, v) > 0$ depending only on $u \in S^{k-1}$ and $v \in S^{d-k-1}$ (i.e., independent of s), we have

$$\mathrm{Prob}_{x \sim S^{d-1}}\left[|sc - 1| > \epsilon\right] \geq L.$$

Proof By the equality of differential forms in Equation (2),

$$\mathrm{Prob}\left[|sc - 1| > \epsilon\right] = \int_{|sc-1|>\epsilon} Bf(s)ds \frac{d\Omega_{k-1}}{\mathrm{Vol}_{k-1}\left(S^{k-1}\right)} \frac{d\Omega_{d-k-1}}{\mathrm{Vol}_{d-k-1}\left(S^{d-k-1}\right)}$$

$$= \int_{S^{k-1}\times S^{d-k-1}} \left(\int_{|sc-1|>\epsilon} Bf(s)ds\right) \frac{d\Omega_{k-1}}{\mathrm{Vol}_{k-1}\left(S^{k-1}\right)} \frac{d\Omega_{d-k-1}}{\mathrm{Vol}_{d-k-1}\left(S^{d-k-1}\right)}.$$

Our goal is to find a lower bound on the integral $\int_{|sc-1|>\epsilon} Bf(s)ds$. Due to the independence of u, v, and s, $c(u, v)$ is a fixed positive constant within this integral. We observe that $|sc - 1| > \epsilon$ consists of two intervals, $s < (1 - \epsilon)/c$ and $s > (1 + \epsilon)/c$, and we consider two cases depending on the value of c.

We begin by recalling that

$$\mathrm{Prob}\left[s > s_0(1 + \epsilon)\right] = \int_{s>s_0(1+\epsilon)} Bf(s)ds \quad \text{and} \quad \mathrm{Prob}\left[s < s_0(1 - \epsilon)\right] = \int_{s<s_0(1-\epsilon)} Bf(s)ds.$$

If $c \geq s_0$, then $(1 + \epsilon)/c \leq (1 + \epsilon)/s_0$, and, hence

$$\int_{|sc-1|>\epsilon} Bf(s)ds \geq \int_{s>(1+\epsilon)/c} Bf(s)ds \geq \int_{s>(1+\epsilon)/s_0} Bf(s)ds \geq L.$$

On the other hand, if $c < s_0$, then $(1 - \epsilon)/s_0 < (1 - \epsilon)/c$, then

$$\int_{|sc-1|>\epsilon} Bf(s)ds \geq \int_{s<(1-\epsilon)/c} Bf(s)ds \geq \int_{s<(1-\epsilon)/s_0} Bf(s)ds \geq L.$$

Therefore, the integral $\int_{|sc-1|>\epsilon} Bf(s)ds$ is bounded from below by L, and

$$\mathrm{Prob}\left[|sc - 1| > \epsilon\right] \geq \int_{S^{k-1}\times S^{d-k-1}} L \frac{d\Omega_{k-1}}{\mathrm{Vol}_{k-1}\left(S^{k-1}\right)} \frac{d\Omega_{d-k-1}}{\mathrm{Vol}_{d-k-1}\left(S^{d-k-1}\right)} = L. \qquad \blacksquare$$

We now show that when $k \le \eta\epsilon^{-2}\log(1/\delta)$ with $\eta < 4$, and ϵ and δ are sufficiently small, there does not exist an (ϵ, δ)-JL distribution on $\mathbb{R}^{k \times d}$. This fact, combined with the results in Section 2.2, shows that the limit appearing in Theorem 2 exists and equals 1. In order to show this, we consider the following related problem: By definition, for a probability distribution $\mathcal{D}$ on $\mathbb{R}^{k \times d}$ to be an (ϵ, δ)-JL distribution, the following inequality must hold for every $w \in S^{d-1}$:

$$\text{Prob}_{A \sim \mathcal{D}}\left[\left|\|Aw\|_2^2 - 1\right| > \epsilon\right] < \delta.$$

Hence,

$$\text{Prob}_{A \sim \mathcal{D},\ w \sim S^{d-1}}\left[\left|\|Aw\|_2^2 - 1\right| > \epsilon\right] < \delta, \tag{5}$$

where $w \in S^{d-1}$ is a random variable distributed uniformly on S^{d-1}. Following the approach of Kane et al. (2011), our goal is to prove that, for every $A \in \mathbb{R}^{k \times d}$,

$$\text{Prob}_{w \sim S^{d-1}}\left[\left|\|Aw\|_2^2 - 1\right| > \epsilon\right] > \delta. \tag{6}$$

When Inequality (6) holds for all A, then Inequality (5) cannot hold for any distribution $\mathcal{D}$ on $\mathbb{R}^{k \times d}$. Therefore, an (ϵ, δ)-JL distribution does not exist. We make this precise in the following theorem:

Theorem 5 *Suppose that $\eta < 4$ and let $k(\epsilon, \delta) = \left\lfloor \eta\epsilon^{-2}\log\left(\frac{1}{\delta}\right) \right\rfloor$. Let $s_0 = k/d$, and suppose that, for every ϵ, δ, and s_0 sufficiently small (to make s_0 sufficiently small, d must be sufficiently large),*

$$\min\{\text{Prob}\left[s > s_0(1 + \epsilon)\right], \text{Prob}\left[s < s_0(1 - \epsilon)\right]\} \ge C\delta^{\frac{\eta}{4}\gamma},$$

where $C > 0$ is an absolute constant, and γ approaches 1 as ϵ, δ, and s_0 approach 0. Then, by decreasing ϵ, δ, and s_0 as needed, for every matrix $A \in \mathbb{R}^{k(\epsilon,\delta) \times d}$,

$$\text{Prob}_{w \sim S^{d-1}}\left[\left|\|Aw\|_2^2 - 1\right| > \epsilon\right] > \delta.$$

Proof We assume that A has rank $k = k(\epsilon, \delta)$ since, if not, we may reduce k (and decrease η correspondingly) to the rank of A. Let $A = U\Sigma V^t$ be the singular value decomposition of A where U is a $k \times k$ orthonormal matrix, $V = (v_1, \ldots, v_d)$ is a $d \times d$ orthonormal matrix, and Σ is a $k \times d$ diagonal matrix with $\lambda_i > 0$ its entry at $\Sigma_{i,i}$ for $1 \le i \le k$. Let

$$x = (x_1, \ldots, x_d)^t = V^t w.$$

Since V is orthonormal, we have $x \in S^{d-1}$. We observe that since w is a uniformly distributed random variable on S^{d-1}, $V^t w$ is also a uniformly distributed random variable on S^{d-1}. Therefore, since U is orthonormal, we have

$$\|Aw\|_2^2 = \|U\Sigma x\|_2^2 = \|\Sigma x\|_2^2 = \sum_{i=1}^{k} \lambda_i^2 x_i^2.$$

We recall the definition of the function $\Psi(x) = (s, u, v)$ where $s = x_1^2 + \cdots + x_k^2$. Moreover, we restrict our attention to the case where $s \in (0, 1)$ since the complement has zero measure. Let

$$c = \sum_{i=1}^{k} \lambda_i^2 x_i^2 / s = \|\Sigma_{k \times k} u\|_2^2,$$

8

where $\Sigma_{k \times k}$ denotes the $k \times k$ principal submatrix of Σ, then

$$\mathrm{Prob}_{w \sim S^{d-1}} \left[\left| \|Aw\|_2^2 - 1 \right| > \epsilon \right] = \mathrm{Prob}_{x \sim S^{d-1}} \left[|sc - 1| > \epsilon \right].$$

Due to the independence of u, v, and s, it follows that c depends only on u. Therefore, by Lemma 4, it follows that

$$\mathrm{Prob}_{w \sim S^{d-1}} \left[\left| \|Aw\|_2^2 - 1 \right| > \epsilon \right] \geq C \delta^{\frac{\eta}{4}\gamma}.$$

It follows that for ϵ, δ, and s_0 sufficiently small, $C\delta^{\frac{\eta}{4}\gamma} > \delta$. ■

Therefore, by selecting ϵ, δ, and s_0 sufficiently small, it follows from Theorem 5 that there is no (ϵ, δ)-JL distribution when $k < \eta \epsilon^{-2} \log\left(\frac{1}{\delta}\right)$ for $\eta < 4$. Therefore, $k_0(\epsilon, \delta) \geq \eta \epsilon^{-2} \log\left(\frac{1}{\delta}\right)$. We collect the results of Proposition 3 and Theorem 5 in the following corollary:

Corollary 6 *Assume the hypotheses on* $\mathrm{Prob}\left[s > s_0(1 + \epsilon)\right]$ *and* $\mathrm{Prob}\left[s < s_0(1 - \epsilon)\right]$ *from Proposition 3 and Theorem 5 hold.*

(a) *There exists an $o(1)$ function that approaches 0 as ϵ and δ approach zero such that if $k > 4\epsilon^{-2} \log\left(\frac{1}{\delta}\right)[1 + o(1)]$, then there exists a JL distribution.*

(b) *If $k(\epsilon, \delta) = \left\lfloor \eta \epsilon^{-2} \log\left(\frac{1}{\delta}\right) \right\rfloor$, then, by decreasing ϵ and δ, and increasing d, there is no (ϵ, δ)-JL distribution for any $k' \leq k(\epsilon, \delta)$.*

This proves the main result in the paper. In the following section, we provide the more technical results that verify the assumptions in Proposition 3 and Theorem 5.

3. Explicit Concentration Bounds

Throughout this section, we assume that s is a random variable with probability distribution $Bf(s)$. We define $s_0 = \frac{k}{d}$, and we further assume that $0 \leq \epsilon, \delta \leq 1/2$, $k - 4 \geq \epsilon^{-2}$ and $s_0 < 0.4$. We derive lower and upper bounds for the following probabilities:

$$\mathrm{Prob}\left[s > s_0(1 + \epsilon)\right] \quad \text{and} \quad \mathrm{Prob}\left[s < s_0(1 - \epsilon)\right],$$

using the probability density $Bf(s)$ for s and $f(s) = s^{(k-2)/2}(1 - s)^{(d-k-2)/2}$. These probabilities appear in (Kane et al., 2011, Theorem 20) without the explicit constants that are derived in this paper. These bounds are instances of explicit concentration theorems (or explicit laws of large numbers) from probability theory. Our goal is to formulate these bounds as precisely as possible so that the lower and upper bounds are asymptotically the same when ϵ and δ approach 0.

3.1. Bounds for B

We recall that $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$, and $\Gamma(1 + z) = z\Gamma(z)$. Hence

$$\Gamma\left(\frac{d}{2}\right) = \left(\frac{d}{2} - 1\right)! \text{ if } d \text{ is even, and } \Gamma\left(\frac{d}{2}\right) = \left(\frac{d}{2} - 1\right)\left(\frac{d}{2} - 2\right) \cdots \frac{3}{2}\frac{1}{2}\sqrt{\pi} \text{ if } d \text{ is odd.}$$

In this section, we derive lower and upper bounds for B, see Equation (38), by using the following form of Stirling's approximation of $n!$ due to Robbins (1955):

$$\sqrt{2\pi}n^{n+1/2}e^{-n}e^{\frac{1}{12n+1}} < \Gamma(n+1) = n! < \sqrt{2\pi}n^{n+1/2}e^{-n}e^{\frac{1}{12n}}.$$

Since we are interested in the asymptotic behavior, we focus on the case where d is even. This choice does not affect the asymptotic results of our paper, but the calculations are more straight-forward in this case. We leave the details for the case where d is odd to the interested reader.

Lemma 7 *Suppose k and d are both even. Then we have the following inequality:*[2]

$$\frac{e^{-2}}{2\sqrt{\pi}}\frac{(d-2)^{(d-1)/2}}{(k-2)^{(k-1)/2}(d-k-2)^{(d-k-1)/2}} \le B \le \frac{e^{-1}}{2\sqrt{\pi}}\frac{(d-2)^{(d-1)/2}}{(k-2)^{(k-1)/2}(d-k-2)^{(d-k-1)/2}}.$$

Proof Using the bound on $n!$ from Robbins (1955), we obtain

$$C_0\frac{(d-2)^{(d-1)/2}}{(k-2)^{(k-1)/2}(d-k-2)^{(d-k-1)/2}} \le B \le C_1\frac{(d-2)^{(d-1)/2}}{(k-2)^{(k-1)/2}(d-k-2)^{(d-k-1)/2}},$$

where

$$C_0 = \frac{1}{2\sqrt{\pi}}e^{-1}e^{\frac{1}{6(d-2)+1}}e^{\frac{-1}{6(k-2)}}e^{\frac{-1}{6(d-k-2)}} \ge \frac{e^{-2}}{2\sqrt{\pi}}, \quad \text{and}$$

$$C_1 = \frac{1}{2\sqrt{\pi}}e^{-1}e^{\frac{1}{6(d-2)}}e^{\frac{-1}{6(k-2)+1}}e^{\frac{-1}{6(d-k-2)+1}} \le \frac{e^{-1}}{2\sqrt{\pi}}. \qquad \blacksquare$$

Corollary 8 *With $s_0 = k/d$, we have*

$$\frac{e^{-2}}{2\sqrt{\pi}}\sqrt{k} \le Bs_0f(s_0) \le \frac{9e^{-1}}{\sqrt{2\pi}}\sqrt{k}.$$

Proof By evaluating f at s_0 and replacing B by its lower bound found in Lemma 7, we obtain the lower bound

$$Bs_0f(s_0) \ge \frac{e^{-2}}{2\sqrt{\pi}}\sqrt{k}\left(\frac{d-2}{d-k}\right)^{\frac{1}{2}}\left(\frac{k(d-2)}{d(k-2)}\right)^{\frac{k-1}{2}}\left(\frac{(d-k)(d-2)}{d(d-k-2)}\right)^{\frac{d-k-1}{2}} \ge \frac{e^{-2}}{2\sqrt{\pi}}\sqrt{k}.$$

Similarly, by using the upper bound in Lemma 7, we obtain the upper bound

$$Bs_0f(s_0) \le \frac{e^{-1}}{2\sqrt{\pi}}\sqrt{k}\left(\frac{d-2}{d}\right)^{\frac{d-1}{2}}\left(\frac{k}{k-2}\right)^{\frac{k-1}{2}}\left(\frac{d-k}{d-k-2}\right)^{\frac{d-k-1}{2}}\frac{\sqrt{d}}{\sqrt{d-k}} \le \frac{9e^{-1}}{\sqrt{2\pi}}\sqrt{k},$$

where the last inequality follows from $\frac{d}{d-k} \le 2$ since $s_0 < 0.4$, and $\left(\frac{x}{x-2}\right)^{\frac{x-1}{2}} \le 3$ for $x \ge 3$.
$\blacksquare$

3.2. Bounds on $\mathrm{Prob}\left[s > s_0(1 + \epsilon)\right]$

We begin by mentioning the following inequalities which are used in our arguments below:

$$\log(1 + x) \geq x - \frac{x^2}{2} \quad \text{for } 0 < x < 1, \tag{7}$$

$$\log(1 - x) \geq -x - x^2 \quad \text{for } 0 < x < 0.68, \tag{8}$$

$$\log(1 + x) \leq x \quad \text{for } x > -1, \quad \text{and} \tag{9}$$

$$\log(1 + x) \leq x - \frac{x^2}{2} + \frac{x^3}{3} \quad \text{for } x > -1. \tag{10}$$

These bounds can be verified by employing basic calculus techniques (e.g., derivatives and Taylor expansions) as well as sufficiently accurate approximations. Using these inequalities, we derive the following bounds:

Lemma 9

$$\mathrm{Prob}\left[s > s_0(1 + \epsilon)\right] \geq \frac{e^{-2}}{4\sqrt{\pi}} e^{-\frac{1}{4}(\sqrt{k}\epsilon + 1)^2 \frac{1 + s_0}{1 - s_0}}.$$

Moreover, when $k < \eta\epsilon^{-2}\log\frac{1}{\delta}$,

$$\mathrm{Prob}\left[s > s_0(1 + \epsilon)\right] \geq \frac{e^{-2}}{4\pi} \delta^{\frac{\eta}{4}\gamma_1},$$

where

$$\gamma_1 = \left(1 + (\eta\log(1/\delta))^{-1/2}\right)^2 \left(\frac{1 + s_0}{1 - s_0}\right).$$

Additionally, γ_1 approaches 1 as ϵ, δ, and s_0 approach 0.

Proof Note that

$$\mathrm{Prob}\left[s > s_0(1 + \epsilon)\right] = B \int_{s > s_0(1 + \epsilon)} f(s)ds = Bs_0 \int_{\epsilon}^{\frac{1}{s_0} - 1} f(s_0(1 + x))dx, \tag{11}$$

via the substitution $s = s_0(1 + x)$.

Let $g(s) = s^{k/2}(1 - s)^{(d-k)/2}$, then $f(s_0(1+x))$ can be expressed in terms of $g(s)$, namely,

$$f(s_0(1 + x)) = \frac{g(s_0(1 + x))}{s_0(1 + x)\left(1 - s_0(1 + x)\right)}. \tag{12}$$

To find a lower bound on $\mathrm{Prob}\left[s > s_0(1 + \epsilon)\right]$, we compute a bound on $g(s_0(1 + x))$ from below. Taking the logarithm of $g(s_0(1 + x))$, we find

$$\log\left(g(s_0(1 + x))\right) = \log g(s_0) + \frac{d}{2}\left(s_0\log(1 + x) + (1 - s_0)\log\left(1 - \frac{s_0}{1 - s_0}x\right)\right). \tag{13}$$

Restricting x to the interval $0 \leq x < 1$, it then follows that $0 < \frac{s_0}{1 - s_0}x < 0.68$ from the assumption that $s_0 < 0.4$. We now bound the second term in Equation (13) using Inequalities (7) and (8), as follows:

$$s_0\log(1 + x) + (1 - s_0)\log\left(1 - \frac{s_0}{1 - s_0}x\right) \geq -\left(\frac{s_0(1 + s_0)}{2(1 - s_0)}\right)x^2. \tag{14}$$

Hence, by substituting Inequality (14) into Equation (13) and exponentiating, we obtain the following lower bound for $g(s_0(1 + x))$:

$$g(s_0(1 + x)) \geq g(s_0)e^{-\frac{k}{4}x^2\frac{1+s_0}{1-s_0}}. \tag{15}$$

We also observe that since $s_0 < 0.4$ and $0 \leq x < 1$, the denominator of Equation (12) is bounded from below as follows:

$$\frac{1}{s_0(1 + x)(1 - s_0(1 + x))} \geq \frac{1}{2s_0(1 - s_0)}. \tag{16}$$

Therefore, by substituting Inequalities (15) and (16) into Equation (12) when $0 \leq x < 1$, we have

$$f(s_0(1 + x)) \geq \frac{g(s_0)}{2s_0(1 - s_0)}e^{-\frac{k}{4}x^2\frac{1+s_0}{1-s_0}} = \frac{1}{2}f(s_0)e^{-\frac{k}{4}x^2\frac{1+s_0}{1-s_0}}. \tag{17}$$

Since $s_0 < 0.4$, we observe that $\frac{1}{s_0} - 1 = \frac{d-k}{k} > 1.5$. Since we assumed that $\epsilon < \frac{1}{2}$ and $k \geq 4 + \epsilon^{-2} > 4$, it follows that $\epsilon + k^{-1/2} < 1$ and so $\epsilon + k^{-1/2} < 1 < \frac{1}{s_0} - 1$. Therefore, we further restrict x to the interval $(\epsilon, \epsilon + k^{-1/2})$ and observe that Inequality (17) applies in this range. Therefore, since f is a positive function,

$$Bs_0\int_\epsilon^{\frac{1}{s_0}-1} f(s_0(1 + x))dx \geq Bs_0\int_\epsilon^{\epsilon+k^{-1/2}} f(s_0(1 + x))dx \geq \frac{1}{2}Bs_0f(s_0)\int_\epsilon^{\epsilon+k^{-1/2}} e^{-\frac{k}{4}x^2\frac{1+s_0}{1-s_0}}dx. \tag{18}$$

Replacing $Bs_0f(s_0)$ with its lower bound given in Corollary 8 and observing that the integrand is decreasing over an interval of width $k^{-1/2}$, Inequality (18) is bounded from below by

$$\frac{1}{2}Bs_0f(s_0)\int_\epsilon^{\epsilon+k^{-1/2}} e^{-\frac{k}{4}x^2\frac{1+s_0}{1-s_0}}dx \geq \frac{e^{-2}}{4\sqrt{\pi}}e^{-\frac{1}{4}(\sqrt{k}\epsilon+1)^2\frac{1+s_0}{1-s_0}},$$

which completes the first inequality. The second inequality follows by replacing k by the given upper bound and simplifying. ∎

Lemma 10

$$\text{Prob}\left[s > s_0(1 + \epsilon)\right] \leq \frac{27e^{-1}}{\sqrt{2\pi}}e^{-\frac{k-2}{4}\epsilon^2(1-\frac{2}{3}\epsilon)}.$$

Proof To derive an upper bound, we start with the expression for $\text{Prob}\left[s > (1 + \epsilon)s_0\right]$ from Equation (11). We first find an upper bound on $f(s_0(1+x))$. Then, we bound $f(s_0(1+x))$ using the inequality $1 - x \leq e^{-x}$ for all x as follows:

$$f(s_0(1+x)) = f(s_0)(1+x)^{\frac{k-2}{2}}\left(1 - \frac{s_0}{1 - s_0}x\right)^{\frac{d-k-2}{2}} \leq f(s_0)(1+x)^{\frac{k-2}{2}}\left(e^{\frac{-s_0}{1-s_0}x}\right)^{\frac{d-k-2}{2}}, \tag{19}$$

Moreover, since $s_0 < 0.4$,

$$\frac{s_0}{1 - s_0}\frac{d - k - 2}{2} = \frac{k}{d - k}\frac{d - k - 2}{2} > \frac{k - 2}{2}. \tag{20}$$

12

By applying Inequality (20) to Inequality (19), we derive the upper bound

$$f(s_0)(1+x)^{\frac{k-2}{2}}e^{-\frac{k-2}{2}x}.$$

Therefore, by extending the region of integration in Inequality (11), we find the following upper bound on the probability:

$$Bs_0\int_{\epsilon}^{\frac{1}{s_0}-1}f(s_0(1+x))dx \leq Bs_0f(s_0)\int_{\epsilon}^{\infty}(1+x)^{\frac{k-2}{2}}e^{-\frac{k-2}{2}x}dx. \tag{21}$$

By integrating by parts, we observe that for any ℓ and m with $1 \leq \ell \leq m$,

$$\int_{\epsilon}^{\infty}(1+x)^{\ell}e^{-mx}dx \leq \frac{1}{m}(1+\epsilon)^{\ell}e^{-m\epsilon} + \int_{\epsilon}^{\infty}(1+x)^{\ell-1}e^{-mx}dx. \tag{22}$$

Applying Inequality (22) $\frac{k-2}{2}$ times to the integral in Inequality (21) and bounding the resulting geometric series from above gives

$$\int_{\epsilon}^{\infty}(1+x)^{\frac{k-2}{2}}e^{-\frac{k-2}{2}x}dx \leq \frac{2e^{-\frac{k-2}{2}\epsilon}}{k-2}\sum_{i=0}^{\frac{k-2}{2}}(1+\epsilon)^{i} \leq \frac{2(1+\epsilon)}{\epsilon(k-2)}(1+\epsilon)^{\frac{k-2}{2}}e^{-\frac{k-2}{2}\epsilon}.$$

By applying Inequality (10) to $(1+\epsilon)^{\frac{k-2}{2}} = e^{\frac{k-2}{2}\log(1+\epsilon)}$, we obtain the upper bound

$$\frac{2(1+\epsilon)}{\epsilon(k-2)}(1+\epsilon)^{\frac{k-2}{2}}e^{-\frac{k-2}{2}\epsilon} \leq \frac{2(1+\epsilon)}{\epsilon(k-2)}e^{-\frac{k-2}{4}\epsilon^2(1-\frac{2}{3}\epsilon)}.$$

Since $k - 4 \geq \epsilon^{-2}$, it follows that $\frac{(k-2)^2}{k} \geq k - 4 \geq \epsilon^{-2}$, and, hence, that $\epsilon(k-2) \geq \sqrt{k}$. Therefore, we can further simplify our bound to

$$\frac{2(1+\epsilon)}{\epsilon(k-2)}e^{-\frac{k-2}{4}\epsilon^2(1-\frac{2}{3}\epsilon)} \leq \frac{2(1+\epsilon)}{\sqrt{k}}e^{-\frac{k-2}{4}\epsilon^2(1-\frac{2}{3}\epsilon)}. \tag{23}$$

By combining Inequalities (21) and (23), we find an upper bound on the probability as follows:

$$Bs_0\int_{\epsilon}^{\frac{1}{s_0}-1}f(s_0(1+x))dx \leq Bs_0f(s_0)\frac{2(1+\epsilon)}{\sqrt{k}}e^{-\frac{k-2}{4}\epsilon^2(1-\frac{2}{3}\epsilon)}.$$

Applying the upper bound on $Bs_0f(s_0)$ from Corollary 8 and the assumption that $\epsilon \leq \frac{1}{2}$ completes the inequality. ∎

3.3. Bounds on Prob $[s < s_0(1-\epsilon)]$

We begin this section by including the following two additional inequalities on $\log(1-x)$:

$$\log(1-x) \geq -x - \frac{x^2}{2} - x^3 \quad \text{for } 0 < x < 0.815, \tag{24}$$

$$\log(1-x) \leq -x - \frac{x^2}{2} \quad \text{for } 0 \leq x < 1. \tag{25}$$

These bounds can be justified using a similar approach as for Inequalities (7-10). Using these inequalities, we derive the following bounds:

13

Lemma 11

$$\text{Prob}\left[s < s_0(1 - \epsilon)\right] \geq \frac{e^{-2}}{2\sqrt{\pi}} e^{-\frac{1}{4}\left(\frac{(\sqrt{k}\epsilon+1)^2}{1-s_0}+2(\sqrt[3]{k}\epsilon+k^{-1/6})^3\right)}.$$

Moreover, when $k < \eta\epsilon^{-2}\log\frac{1}{\delta}$,

$$\text{Prob}\left[s < s_0(1 - \epsilon)\right] \geq \frac{e^{-2}}{2\sqrt{\pi}}\delta^{\frac{\eta}{4}\gamma_2},$$

where

$$\gamma_2 = \frac{1}{1 - s_0}\left(1 + \frac{1}{\sqrt{\eta\log\frac{1}{\delta}}}\right)^2 + 2\left(\epsilon^{1/3} + \frac{1}{\sqrt[3]{\eta\log\frac{1}{\delta}}}\right)^3.$$

Additionally, γ_2 approaches 1 as ϵ, δ, and s_0 approach 0.

Proof The proof of this lemma is very similar to the proof of Lemma 9, so we focus on the new details. The probability can be rewritten, using the substitution $s = s_0(1 - x)$, as

$$\text{Prob}\left[s < (1 - \epsilon)s_0\right] = B\int_{s<s_0(1-\epsilon)} f(s)ds = Bs_0\int_{\epsilon}^{1} f(s_0(1 - x))dx. \tag{26}$$

Using $g(s)$ as in Lemma 9, it follows that

$$f(s_0(1 - x)) = \frac{g(s_0(1 - x))}{s_0(1 - x)(1 - s_0(1 - x))} \tag{27}$$

and

$$\log g(s_0(1 - x)) = \log g(s_0) + \frac{d}{2}\left(s_0\log(1 - x) + (1 - s_0)\log\left(1 + \frac{s_0}{1 - s_0}x\right)\right). \tag{28}$$

Since $0 < x \leq 1$, it follows that $0 < \frac{s_0}{1-s_0}x < 0.68$ from the assumption that $s_0 < 0.4$. Therefore, we can bound the second term in Equation (28) using Inequalities (7) and (24), as follows:

$$s_0\log(1 - x) + (1 - s_0)\log\left(1 + \frac{s_0}{1 - s_0}x\right) \geq \frac{-s_0}{2(1 - s_0)}x^2 - s_0x^3. \tag{29}$$

Substituting Inequality (29) into Expression (28) and exponentiating, we get

$$g(s_0(1 - x)) \geq g(s_0)e^{-\frac{k}{4}\left(\frac{x^2}{1-s_0}+2x^3\right)}. \tag{30}$$

Since $s_0 < 0.4$, it follows that $\frac{s_0}{1-s_0} < 1$, and, hence, that $(1-x)\left(1 + \frac{s_0}{1-s_0}x\right) < 1$. Therefore, the denominator in Equation (27) can be bounded from above by

$$s_0(1 - x)(1 - s_0(1 - x)) = s_0(1 - s_0)(1 - x)\left(1 + \frac{s_0x}{1 - s_0}\right) \leq s_0(1 - s_0). \tag{31}$$

Therefore, by substituting Inequalities (30) and (31) into Expression (27), when $\epsilon < x < 1$, we have

$$f(s_0(1-x)) \geq f(s_0)e^{-\frac{k}{4}\left(\frac{x^2}{1-s_0}+2x^3\right)}. \tag{32}$$

Since $\epsilon < \frac{1}{2}$ and $k \geq 4+\epsilon^{-2} > 4$, it follows that $\epsilon + k^{-1/2} < 1$. Therefore, we restrict x to the interval $(\epsilon, \epsilon + k^{-1/2})$, and observe that Inequality (32) applies in this range. Therefore,

$$Bs_0 \int_\epsilon^1 f(s_0(1-x))dx \geq Bs_0f(s_0)\int_\epsilon^{\epsilon+k^{-1/2}} e^{-\frac{k}{4}\left(\frac{x^2}{1-s_0}+2x^3\right)}dx. \tag{33}$$

Replacing $Bs_0f(s_0)$ with its lower bound given in Corollary 8 and observing that the integrand is decreasing on an interval of width $k^{-1/2}$, Inequality (33) is bounded from below by

$$Bs_0f(s_0)\int_\epsilon^{\epsilon+k^{-1/2}} e^{-\frac{k}{4}\left(\frac{x^2}{1-s_0}+2x^3\right)}dx \geq \frac{e^{-2}}{2\sqrt{\pi}}e^{-\frac{1}{4}\left(\frac{(\sqrt{k}\epsilon+1)^2}{1-s_0}+2(\sqrt[3]{k}\epsilon+k^{-1/6})^3\right)},$$

which completes the first inequality. The second inequality follows by replacing $k \geq 1$ by the given upper bound and simplifying. ∎

Lemma 12

$$\mathrm{Prob}\left[s < s_0(1-\epsilon)\right] \leq \frac{18\sqrt{2}e^{1/2}}{\sqrt{\pi}}e^{-\left(\frac{k}{4}\right)\epsilon^2} \leq \frac{18\sqrt{2}e^{1/2}}{\sqrt{\pi}}e^{-\left(\frac{k-2}{4}\right)\epsilon^2\left(1-\frac{2}{3}\epsilon\right)}.$$

Proof The proof of this lemma is very similar to the proof of Lemma 10, so we focus on the new details. To prove an upper bound, we start with the bound on $\mathrm{Prob}\left[s < s_0(1-\epsilon)\right]$ from Equation (26). We first observe that

$$f(s_0(1-x)) = f(s_0)(1-x)^{\frac{k-2}{2}}\left(1+\frac{s_0}{1-s_0}x\right)^{\frac{d-k-2}{2}}. \tag{34}$$

We now bound the logarithm of the second and third factors in Equation (34) using Inequalities (9) and (25) as follows:

$$\log\left((1-x)^{\frac{k-2}{2}}\left(1+\frac{s_0}{1-s_0}x\right)^{\frac{d-k-2}{2}}\right) \leq \frac{k-2}{2}\left(-x-\frac{x^2}{2}\right)+\frac{d-k-2}{2}\left(\frac{s_0}{1-s_0}x\right) \tag{35}$$

Since $\frac{s_0}{1-s_0} = \frac{k}{d-k}$ and $\epsilon < x < 1$, Inequality (35) further simplifies to

$$\frac{k-2}{2}\left(-x-\frac{x^2}{2}\right)+\frac{d-k-2}{2}\left(\frac{k}{d-k}x\right) \leq x-\left(\frac{k-2}{4}\right)x^2 \leq \frac{3}{2}-\left(\frac{k}{4}\right)x^2.$$

Hence, for $\epsilon \leq x < 1$, we have

$$f(s_0(1-x)) \leq f(s_0)e^{3/2}e^{-\left(\frac{k}{4}\right)x^2} \leq f(s_0)e^{3/2}e^{-\left(\frac{k}{4}\right)\epsilon x}. \tag{36}$$

Substituting Inequality (36) into the integral of Equation (26), we find

$$Bs_0 \int_\epsilon^1 f(s_0(1-x))dx \leq Bs_0 f(s_0)e^{3/2} \int_\epsilon^1 e^{-\left(\frac{k}{4}\right)\epsilon x}dx$$

$$\leq Bs_0 f(s_0)e^{3/2} \int_\epsilon^\infty e^{-\left(\frac{k}{4}\right)\epsilon x}dx \leq Bs_0 f(s_0)\frac{4e^{3/2}}{k\epsilon}e^{-\frac{k}{4}\epsilon^2}. \qquad (37)$$

Since $k \geq \epsilon^{-2}$, Inequality (37) can be further simplified to

$$Bs_0 f(s_0)\frac{4e^{3/2}}{\sqrt{k}}e^{-\frac{k}{4}\epsilon^2}$$

Applying the upper bound on $Bs_0 f(s_0)$ from Corollary 8 completes the first inequality of the proof. The final inequality follows from the fact that $k \geq (k-2)\left(1 - \frac{2}{3}\epsilon\right)$. ∎

This completes the proof all of the conditions in Section 2, and, therefore, completes the proof of our main theorem, Theorem 2.

Acknowledgments

Appendix A. Uniform Distributions on Unit Spheres in High Dimensions

In this section, we study uniform distributions and surface areas (or hypervolumes) on high-dimensional spheres. More precisely, for any $d \geq 1$, let S^{d-1} denote the unit sphere of dimension $d-1$ and $d\Omega_{d-1}$ denote the surface area measure for S^{d-1}. We show that, for any $1 \leq k \leq d$,

$$d\Omega_{d-1} = \frac{1}{2}f(s)ds\, d\Omega_{k-1}d\Omega_{d-k-1},$$

where $s \in [0,1]$ and $f(s) = s^{\frac{k-2}{2}}(1-s)^{\frac{d-k-2}{2}}$. This is a more precise version of a result in Kane et al. (2011), replacing an unspecified constant by $1/2$. We include the proof of this result here in order to make this result more accessible to a larger population of researchers. This formula is of independent interest since it shows that the uniform distribution on S^{d-1} is a product of uniform distributions on S^{k-1} and S^{d-k-1} with a distribution on $[0,1]$.

Theorem 13 *Under the almost bijective map $\Psi : S^{d-1} \to [0,1] \times S^{k-1} \times S^{d-k-1}$, we have equality of the surface area differential forms on S^{d-1}, S^{k-1}, and S^{d-k-1}, i.e.,*

$$d\Omega_{d-1} = \frac{1}{2}f(s)ds\, d\Omega_{k-1}d\Omega_{d-k-1},$$

where $f(s) = s^{(k-2)/2}(1-s)^{(d-k-2)/2}$. Equivalently, in terms of probability distribution measures,

$$\frac{d\Omega_{d-1}}{Vol_{d-1}\left(S^{d-1}\right)} = Bf(s)ds\frac{d\Omega_{k-1}}{Vol_{k-1}\left(S^{k-1}\right)}\frac{d\Omega_{d-k-1}}{Vol_{d-k-1}\left(S^{d-k-1}\right)}.$$

Hence, the uniform distribution on S^{d-1} can be identified with the product distribution on $[0,1] \times S^{k-1} \times S^{d-k-1}$ where the distribution of s on $[0,1]$ has density function $Bf(s)$ and

$$B = \frac{1}{2} \frac{Vol_{k-1}\left(S^{k-1}\right) Vol_{d-k-1}\left(S^{d-k-1}\right)}{Vol_{k-1}\left(S^{k-1}\right)} = \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{k}{2})\Gamma(\frac{d-k}{2})}. \tag{38}$$

This theorem is based on the following lemma, which is well-known to experts, but is included here for completeness.

Lemma 14 *Let $x = (x_1, \ldots, x_d)$ with $x_d > 0$ be a point on the upper hemisphere of S^{d-1}. Then the surface area measure of the unit sphere S^{d-1} at x is*

$$d\Omega_{d-1} = \frac{1}{x_d} dx_1 \ldots dx_{d-1}.$$

Before we begin the proof, we recall the approach for S^2 in 3-dimensional space. We consider the upper hemisphere of S^2 as the graph of a function over D^2, where D^{d-1} denotes $(d-1)$-dimensional disk, namely

$$D^{d-1} = \left\{ \hat{x} \in \mathbb{R}^{d-1} : \sum_{i=1}^{d-1} \hat{x}_i^2 \leq 1 \right\}.$$

We then integrate over the disk D^2 to calculate the surface area of S^2. In particular, the integrand is the limit of the ratios of the area of a square in D^2 to the area of the corresponding parallelogram above the square in the tangent space of S^2 as the square shrinks a point. In the case of the sphere, the parallelogram's area is calculated using the cross product, but we must replace the use of the cross product in higher dimensions.

Proof In d-dimensional space, we consider the upper hemisphere of S^{d-1} as the graph of a function over the $(d-1)$-dimensional disk D^{d-1}. We construct a pair of $(d-1)$-dimensional parallelepipeds as follows: P^{d-1} is in the tangent space of D^{d-1} and Q^{d-1} is in the tangent space of S^{d-1}. Then, we take the limit of their $(d-1)$-dimensional volumes as P^{d-1} approaches a point. Due to complications in taking the $(d-1)$-dimensional volume in d-dimensional space, we extend both P^{d-1} and Q^{d-1} to associated, full-dimensional parallelepipeds.

Let $(\hat{x}_1, \ldots, \hat{x}_{d-1}) \in D^{d-1}$ and define $\phi : D^{d-1} \to \mathbb{R}_{\geq 0}$ as

$$\phi(\hat{x}_1, \ldots, \hat{x}_{d-1}) = \sqrt{1 - \sum_{i=1}^{d-1} \hat{x}_i^2}.$$

We observe that the graph of this function is the upper hemisphere of S^{d-1}. We now extend this map to $D^{d-1} \times \mathbb{R}$ as $\Phi_d : D^{d-1} \times \mathbb{R} \to \mathbb{R}^d$ defined by

$$(\hat{x}_1, \ldots, \hat{x}_{d-1}, \hat{x}_d) \mapsto ((1+\hat{x}_d)\hat{x}_1, (1+\hat{x}_d)\hat{x}_2, \ldots, (1+\hat{x}_d)\hat{x}_{d-1}, (1+\hat{x}_d)\phi(\hat{x}_1, \ldots, \hat{x}_{d-1})).$$

We observe that $\Phi_d|_{D^{d-1} \times \{0\}}$ maps the disk $D^{d-1} \times \{0\}$ surjectively onto the graph of ϕ, i.e., the upper hemisphere of S^{d-1}, see Figure 2.
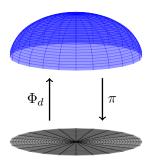
Figure 2: Φ_d maps the disk $D^{d-1} \times \{0\}$ surjectively onto the upper hemisphere of the $(d-1)$-dimensional unit sphere S^{d-1}. We observe that $(\Phi_d)^{-1} = \pi$ is the projection map onto the first $d-1$ coordinates.

We recall that, at $(\hat{x}_1, \ldots, \hat{x}_{d-1}) \in D^{d-1}$, the tangent space is $\mathbb{R}^{d-1}$ and we define the parallelepiped P^{d-1} in the tangent space by the vectors $\Delta \hat{x}_i e_i$ of length $\Delta \hat{x}_i$ in the direction of the ith standard basis vector e_i of $\mathbb{R}^{d-1}$. Similarly, the tangent space at $(\hat{x}_1, \ldots, \hat{x}_{d-1}, 0) \in D^{d-1} \times \mathbb{R}$ is $\mathbb{R}^d$, and we define the parallelepiped P^d in this tangent space by the vectors $\Delta \hat{x}_i e_i$ for $1 \leq i \leq d-1$ and he_d for the final direction. Then, as the vector he_d is perpendicular to the tangent vectors of the disk D^{d-1}, the d-dimensional volume of P^d can be computed in terms of the $(d-1)$-dimensional volume of P^{d-1} and the height h, i.e.,

$$\text{Vol}_d\left(P^d\right) = h\text{Vol}_{d-1}\left(P^{d-1}\right).$$

Next, we let Q^{d-1} and Q^d be the images of P^{d-1} and P^d under the Jacobian of Φ_d, i.e., $\text{Jac}\,\Phi_d$, respectively. Note that the Jacobian of Φ_d, when restricted to $D^{d-1} \times \{0\}$, is

$$\left(\text{Jac}\,\Phi_d\right)|_{D^{d-1} \times \{0\}} = \left[\begin{array}{ccc|c} & & & \hat{x}_1 \\ & I & & \vdots \\ & & & \hat{x}_{d-1} \\ \hline -\frac{\hat{x}_1}{\phi(\hat{x}_1, \ldots, \hat{x}_{d-1})} & \cdots & -\frac{\hat{x}_{d-1}}{\phi(\hat{x}_1, \ldots, \hat{x}_{d-1})} & \phi(\hat{x}_1, \ldots, \hat{x}_{d-1}) \end{array}\right].$$

Since the Jacobian acts on tangent vectors, Q^{d-1} is defined by the vectors

$$\Delta \hat{x}_i \left(f_i - \frac{\hat{x}_i}{\phi(\hat{x}_1, \ldots, \hat{x}_{d-1})} f_d\right),$$

for $1 \leq i \leq d-1$, where the f_i is the ith standard basis vector of $\mathbb{R}^d$. Moreover, Q^d is defined by these vectors as well as the image of he_d, i.e.,

$$h\left(\hat{x}_1, \ldots, \hat{x}_{d-1}, \phi(\hat{x}_1, \ldots, \hat{x}_{d-1})\right).$$

We observe that the vectors $\Delta \hat{x}_i \left(f_i - \frac{\hat{x}_i}{\phi(\hat{x}_1, \ldots, \hat{x}_{d-1})} f_d \right)$ for $1 \le i \le d-1$ are tangent vectors to the sphere S^{d-1}, and that $h(\hat{x}_1, \ldots, \hat{x}_{d-1}, \phi(\hat{x}_1, \ldots, \hat{x}_{d-1}))$ is the outward pointing surface normal with length h. Since the tangent vectors are perpendicular to the outward pointing normal, the volumes of Q^{d-1} and Q^d have a similar relationship as the volumes of P^{d-1} and P^d, i.e.,

$$\mathrm{Vol}_d \left(Q^d \right) = h \mathrm{Vol}_{d-1} \left(Q^{d-1} \right).$$

Therefore, the ratio between the d-dimensional volumes of Q^d and P^d is the same as the ratio of the $(d-1)$-dimensional volumes of Q^{d-1} and P^{d-1}.

Since Q^d is the image of P^d under the linear map $\mathrm{Jac}\ \Phi_d|_{D^{d-1} \times \{0\}}$, it follows that

$$\mathrm{Vol}_d \left(Q^d \right) = \left| \det \left(\mathrm{Jac}\ \Phi_d \right) \right|_{(\hat{x}_1, \ldots, \hat{x}_{d-1}, 0)} \left| \mathrm{Vol}_d \left(P^d \right). \right.$$

Therefore, the ratio of the volumes of Q^{d-1} and P^{d-1} is $\left| \det \left(\mathrm{Jac}\ \Phi_d \right) \right|_{(\hat{x}_1, \ldots, \hat{x}_{d-1}, 0)} \left|. \right.$ It is straight-forward to compute the determinant of $\mathrm{Jac}(\Phi_d)$ at the point $(\hat{x}_1, \ldots, \hat{x}_{d-1}, 0)$ via a few row reductions to eliminate the first $d-1$ entries in the last row and turn the matrix into an upper triangular matrix whose lower right corner is $\frac{1}{\phi(\hat{x}_1, \ldots, \hat{x}_{d-1})}$. Hence, the determinant of $\mathrm{Jac}(\Phi_d)$ is $\frac{1}{\phi(\hat{x}_1, \ldots, \hat{x}_{d-1})}$, which is the desired scaling factor.

Therefore, $\frac{1}{\phi(\hat{x}_1, \ldots, \hat{x}_{d-1})}$ is the local factor in the stretching of the surface area in the map from D^{d-1} to S^{d-1}. We recall that the coordinates $x_1, \ldots, x_d$ are the coordinates on the upper hemisphere and $\hat{x}_1, \ldots, \hat{x}_{d-1}$ are the coordinates on D^{d-1}. Since, under the map $\Phi_d|_{D^{d-1} \times \{0\}}$, $x_i = \hat{x}_i$ for $1 \le i \le d-1$, it follows that

$$d\hat{x}_1 \ldots d\hat{x}_{d-1} = dx_1 \ldots dx_{d-1} \qquad \text{and} \qquad \phi(\hat{x}_1, \ldots, \hat{x}_{d-1}) = x_d.$$

From here, the result follows directly. ∎

Proof of Theorem 13 Let $(w_1, \ldots, w_d)$, $(x_1, \ldots, x_k)$, and $(y_1, \ldots, y_{d-k})$ be coordinates of points on the $(d-1)$-dimensional unit sphere, the $(k-1)$-dimensional unit sphere, and the $(d-k-1)$-dimensional unit sphere, respectively. Let $(\hat{w}_1, \ldots, \hat{w}_{d-1})$, $(\hat{x}_1, \ldots, \hat{x}_{k-1})$, and $(\hat{y}_1, \ldots, \hat{y}_{d-k-1})$ be the coordinates of points on the disks D^{d-1}, D^{k-1} and D^{d-k-1}, respectively. Let

$$\varphi : [0,1] \times D^{k-1} \times D^{d-k-1} \to D^{d-1}$$

be defined by

$$s \times (\hat{x}_1, \ldots, \hat{x}_{k-1}) \times (\hat{y}_1, \ldots, \hat{y}_{d-k-1}) \mapsto$$

$$\left(\sqrt{s}\hat{x}_1, \ldots, \sqrt{s}\hat{x}_{k-1}, \sqrt{s}\sqrt{1 - \sum_{i=1}^{k-1} \hat{x}_i^2}, \sqrt{1-s}\hat{y}_1, \ldots, \sqrt{1-s}\hat{y}_{d-k-1} \right).$$

We observe that φ maps the disks D^{k-1} and D^{d-k-1} onto the half of the disk D^{d-1} whose k^{th} coordinate is nonnegative. As the measure of the image is the measure of the preimage scaled by the determinant of the Jacobian of φ, the surface area measure of the disk D^{d-1} is

$$d\hat{w}_1 \ldots d\hat{w}_{d-1} = |\det \mathrm{Jac}(\varphi)| ds\, d\hat{x}_1 \ldots d\hat{x}_{k-1} d\hat{y}_1 \ldots d\hat{y}_{d-k-1}. \tag{39}$$

19

The Jacobian of φ for $s \in (0,1)$ is

$$
\text{Jac } \varphi =
\left[
\begin{array}{ccccc|ccc}
\frac{\hat{x}_1}{2\sqrt{s}} & \sqrt{s} & \cdots & 0 & & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\
\frac{\hat{x}_{k-1}}{2\sqrt{s}} & 0 & \cdots & \sqrt{s} & & 0 & \cdots & 0 \\
\hline
\frac{\sqrt{1-\sum_{i=1}^{k-1}\hat{x}_i^2}}{2\sqrt{s}} & \frac{-\sqrt{s}\hat{x}_1}{\sqrt{1-\sum_{i=1}^{k-1}\hat{x}_i^2}} & \cdots & \frac{-\sqrt{s}\hat{x}_{k-1}}{\sqrt{1-\sum_{i=1}^{k-1}\hat{x}_i^2}} & & 0 & \cdots & 0 \\
\hline
\frac{-\hat{y}_1}{2\sqrt{1-s}} & 0 & \cdots & 0 & & \sqrt{1-s} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\
\frac{-\hat{y}_{d-k-1}}{2\sqrt{1-s}} & 0 & \cdots & 0 & & 0 & \cdots & \sqrt{1-s}
\end{array}
\right].
$$

Eliminating all but the kth entry of the first column by adding multiples of the other columns to the first column, we obtain

$$
\det \text{Jac}(\varphi) = \frac{1}{2\hat{x}_k} s^{\frac{k-2}{2}} (1-s)^{\frac{d-k-1}{2}}.
$$

Substituting this value into Expression (39), we have the surface area measure of D^{d-1} in terms of the disks D^{k-1} and D^{d-k-1}. That is,

$$
d\hat{w}_1 \ldots d\hat{w}_{d-1} = \frac{1}{2\hat{x}_k} s^{(k-2)/2}(1-s)^{(d-k-1)/2} ds\, d\hat{x}_1 \ldots d\hat{x}_{k-1} d\hat{y}_1 \ldots d\hat{y}_{d-k-1}. \tag{40}
$$

We observe that the coordinates of the disk D^{t-1} correspond to the first $t-1$ entries of coordinates of the unit sphere S^{t-1}. Therefore, we may extend φ to the map Ψ, as defined above, where

$$
\Psi^{-1} = \Phi_d \circ \varphi \circ (id_s \times (\Phi_k)^{-1} \times (\Phi_{d-k})^{-1}).
$$

Employing the results of Lemma 14 in various dimensions, we rewrite the surface measure of a unit sphere in terms of the surface measure of the corresponding disks:

$$
d\hat{x}_1 \ldots d\hat{x}_{k-1} = dx_1 \ldots dx_{k-1} = x_k d\Omega_{k-1}
$$
$$
d\hat{y}_1 \ldots d\hat{y}_{d-k-1} = dy_1 \ldots dy_{d-k-1} = y_{d-k} d\Omega_{d-k-1}
$$
$$
d\hat{w}_1 \ldots d\hat{w}_{d-1} = dw_1 \ldots dw_{d-1} = w_d d\Omega_{d-1}.
$$

By applying the Ψ, we can substitute these three equalities into Equation (40) to obtain

$$
d\Omega_{d-1} = \frac{1}{w_d} dw_1 \ldots dw_{d-1} = \frac{y_{d-k}}{2w_d} s^{(k-2)/2}(1-s)^{(d-k-1)/2} ds\, d\Omega_{k-1} d\Omega_{d-k-1}
$$
$$
= \frac{1}{2} s^{(k-2)/2}(1-s)^{(d-k-2)/2} ds\, d\Omega_{k-1} d\Omega_{d-k-1} = \frac{1}{2} f(s) ds\, d\Omega_{k-1} d\Omega_{d-k-1},
$$

where the third equality follows from the fact that $w_d = \sqrt{1-s}\, y_{d-k}$ by the map Ψ. Since the cases where $s = 0$ or $s = 1$ have measure 0, the result follows. ∎

References

Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003.

Nir Ailon and Bernard Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In *Proceedings of the 38th ACM Symposium on Theory of Computing (STOC)*, pages 557–563, 2006.

Noga Alon. Problems and results in extremal combinatorics. *Discrete Mathematics*, 273: 31–53, 2003.

Rosa I. Arriaga and Santosh Vempala. An algorithmic theory of learning: Robust concepts and random projection. *Machine Learning*, 63:161–182, 2006.

Ella Bingham and Heikki Mannila. Random projection in dimensionality reduction: applications to image and text data. In *Knowledge Discovery and Data Mining*, pages 245–250, 2001.

Emmanuel J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique. Académie des Sciences. Paris*, 346:589–592, 2008.

Kenneth L. Clarkson and David P. Woodruff. Low rank approximation and regression in input sparsity time. In *Proceedings of the 45th ACM Symposium on Theory of Computing (STOC)*, pages 81–90, 2013.

Graham Cormode and Piotr Indyk. Stable distributions in streaming computation. In Garofalakis M., Gehrke J., and Rastoi R., editors, *Data Stream Management*. Springer, Berlin, Heidelberg, 2016.

Anirban Dasgupta, Ravi Kumar, and Tamás Sarlós. A sparse Johnson-Lindenstrauss transform. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*, pages 341–350, 2010.

Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structure and Algorithms*, 22(1):60–65, 2003.

Peter Frankl and Hiroshi Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *Journal of Combinatorial Theory Series B*, 44(3):355–362, 1988.

Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th ACM Symposium on Theory of Computing (STOC)*, pages 604–613, 1998.

T. S. Jayram and David P. Woodruff. Optimal bounds for Johnson-Lindenstrauss transforms and streaming problems with subconstant error. *ACM Transactions on Algorithms*, 9(26), 2013.

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

Daniel M. Kane and Jelani Nelson. Sparser Johnson-Lindenstrauss transforms. *Journal of the ACM*, 61(1), 2014.

Daniel M. Kane, Raghu Meka, and Jelani Nelson. Almost optimal explicit Johnson-Lindenstrauss families. In *Proceedings of the 15th International Workshop on Randomization and Computation (RANDOM)*, pages 628–639, 2011.

Jessica Lin and Dimitrios Gunopulos. Dimensionality reduction by random projection and latent semantic indexing. In *Proceedings of the Text Mining Workshop, at the 3rd SIAM International Conference on Data Mining*, May 2003.

Jiri Matousek. On variants of the Johnson-Lindenstrauss lemma. *Random Structure and Algorithms*, 33(2):142–156, 2008.

Nam H. Nguyen, Thong T. Do, and Trac D. Tran. A fast and efficient algorithm for low-rank approximation of a matrix. In *Proceedings of the 41st ACM Symposium on Theory of Computing (STOC)*, 2009.

Herbert Robbins. A remark on Stirling formula. *American Mathematical Monthly*, 62:26–29, 1955.

Shashanka Ubaru, Arya Mazumdar, and Yousef Saad. Low rank approximation and decomposition of large matrices using error correcting codes. *IEEE Transactions on Information Theory*, 63(9):5544–5558, 2017.

Kilian Q. Weinberger, Anirban Dasgupta, John Langford, Alexander J. Smola, and Josh Attenberg. Feature hashing for large scale multitask learning. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*, volume 1113-1120, 2009.