# On variants of the Johnson–Lindenstrauss lemma

Jiří Matoušek

Department of Applied Mathematics and
Institute of Theoretical Computer Science (ITI)
Charles University, Malostranské nám. 25
118 00 Praha 1, Czech Republic, and
Institute of Theoretical Computer Science
ETH Zurich, 8092 Zurich, Switzerland

**Abstract**

The Johnson–Lindenstrauss lemma asserts that an $n$-point set in any Euclidean space can be mapped to a Euclidean space of dimension $k = O(\varepsilon^{-2} \log n)$ so that all distances are preserved up to a multiplicative factor between $1 - \varepsilon$ and $1 + \varepsilon$. Known proofs obtain such a mapping as a linear map $\mathbf{R}^n \to \mathbf{R}^k$ with a suitable random matrix. We give a simple and self-contained proof of a version of the Johnson–Lindenstrauss lemma that subsumes a basic versions by Indyk and Motwani and a version more suitable for efficient computations due to Achlioptas. (Another proof of this result, slightly different but in a similar spirit, was given independently by Indyk and Naor.) An even more general result was established by Klartag and Mendelson using considerably heavier machinery.

Recently Ailon and Chazelle showed, roughly speaking, that a good mapping can also be obtained by composing a suitable Fourier transform with a linear mapping that has a *sparse* random matrix $M$; a mapping of this form can be evaluated very fast. In their result the nonzero entries of $M$ are normally distributed. We show that the nonzero entries can be chosen as random $\pm 1$, which further speeds up the computation. We also discuss the case of embeddings into $\mathbf{R}^k$ with the $\ell_1$ norm.

# 1 Introduction

The Johnson–Lindenstrauss lemma is the following surprising fact:

**Theorem 1.1** *Let $\varepsilon \in (0, \frac{1}{2})$ be a real number, and let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of $n$ points in $\mathbf{R}^n$. Let $k$ be an integer with $k \geq C\varepsilon^{-2} \log n$, where $C$ is a sufficiently large absolute constant. Then there exists a mapping $f\colon \mathbf{R}^n \to \mathbf{R}^k$ such that*

$$(1-\varepsilon)\|p_i - p_j\| \leq \|f(p_i) - f(p_j)\| \leq (1+\varepsilon)\|p_i - p_j\|$$

*for all $i, j = 1, 2, \ldots, n$, where $\|.\|$ denotes the Euclidean norm.*

That is, every set of $n$ points in a Euclidean space of any dimension can be "flattened" to dimension only $O(\varepsilon^{-2} \log n)$ in such a way that all distances between the points are preserved up to a multiplicative factor between $1-\varepsilon$ and $1+\varepsilon$. (It is known that the dependence of $k$ on both $\varepsilon$ and $n$ is optimal up to the value of $C$, as was shown by Alon [3]; also see [14].)

The Johnson–Lindenstrauss lemma is an interesting mathematical fact, and it has also become one of the basic tools in modern algorithm design. Indeed, $f(P)$ can be regarded as an approximate but compact representation of $P$. For example, while storing $P$ exactly requires storing $n^2$ coordinates, $f(P)$ can be stored in space only $O(n \log n)$ (for a fixed $\varepsilon$). Moreover, the complexity of many geometric algorithms depends significantly on the dimension, and so dimension reduction from $n$ to $O(\log n)$ is a great advantage. See Indyk [7] for an overview of algorithmic applications.

All known proofs of the Johnson–Lindenstrauss lemma proceed according to the following scheme: For given $n$ and an appropriate $k$, one defines a suitable probability distribution $\mathcal{F}$ on the set of all linear maps $\mathbf{R}^n \to \mathbf{R}^k$. Then one proves the following statement:

**Statement 1.2** *If $T\colon \mathbf{R}^n \to \mathbf{R}^k$ is a random linear mapping drawn from the distribution $\mathcal{F}$, then for every vector $x \in \mathbf{R}^n$ we have*

$$\mathrm{Prob}\Big[(1-\varepsilon)\|x\| \leq \|T(x)\| \leq (1+\varepsilon)\|x\|\Big] \geq 1 - \frac{1}{n^2}.$$

Having established this statement for the considered distribution $\mathcal{F}$, Theorem 1.1 follows easily: We choose $T$ at random according to $\mathcal{F}$. Then for every $i < j$, using linearity of $T$ and Statement 1.2 with $x = p_i - p_j$, we get that $T$ fails to satisfy $(1-\varepsilon)\|p_i-p_j\| \le \|T(p_i)-T(p_j)\| \le (1+\varepsilon)\|p_i-p_j\|$ with probability at most $\frac{1}{n^2}$. Consequently, the probability that any of the $\binom{n}{2}$ pairwise distances is distorted by $T$ by more than $1\pm\varepsilon$ is at most $\binom{n}{2}/n^2 < \frac{1}{2}$. Therefore, a random $T$ works with probability at least $\frac{1}{2}$.

Theorem 1.1 was discovered by Johnson and Lindenstrauss [10]. They needed it as a lemma in a result on extendability of Lipschitz maps. In their proof, the random linear map $T$ is chosen as the orthogonal projection on a random $k$-dimensional subspace of $\mathbf{R}^n$ (with an appropriate scaling factor, which turns out to be $\sqrt{n/k}$). This choice can still be considered the most intuitive geometrically, and it is also not difficult technically given an appropriate tool, namely, measure concentration on the sphere. Indeed, the proof boils down to showing that if $x$ is a random point on the unit sphere $S^{n-1}$ in $\mathbf{R}^n$, then the length of its orthogonal projection on the first $k$ coordinates, or in other words, the quantity $\sqrt{x_1^2+x_2^2+\cdots+x_k^2}$, is sharply concentrated around $\sqrt{k/n}$. This is a simple consequence of measure concentration on $S^{n-1}$; see, e.g., [14] for a detailed presentation of such a proof.

For algorithmic applications of the Johnson–Lindenstrauss lemma it is important to be able to generate and evaluate the random linear map $T$ fast. Moreover, one would like to obtain a good estimate of the constant $C$ in the dimension bound. Finally, measure concentration on the sphere cannot be considered a completely elementary tool (suitable for presenting the result in a class for students of computer science, say). These were the main motivations for an ongoing research on variants of the Johnson–Lindenstrauss lemma in the context of combinatorics and computer science [6], [8], [5], [1], [2].

In the following discussion, let $A$ be the $k\times n$ matrix of $T$; that is, $T(x)=Ax$. Choosing $T$ as a random orthogonal projection as in the original proof of Johnson and Lindenstrauss means that the rows of $A$ are chosen as a random $k$-tuple of orthonormal vectors in $\mathbf{R}^n$ (multiplied by $\sqrt{n/k}$ for proper scaling; for simplicity, we will not mention scaling factors for the various matrices $A$ in the rest of the discussion in this section).

Indyk and Motwani [8] noted that the condition of orthogonality can be dropped, and in their proof they choose the entries of $A$ as independent random variables with the standard normal distribution $N(0,1)$. Such an $A$ is much easier to generate. By simple properties of the normal distribution it follows that in this case, for every fixed unit vector $x\in\mathbf{R}^n$, the quantity $\|T(x)\|^2$ has the chi-square distribution with $k$ degrees of freedom, and one can use known tail estimates for this distribution to prove Statement 1.2.

Dasgupta and Gupta [5] use a similar construction of $T$ (but with each row rescaled to a unit vector), and they give a self-contained proof based mainly on calculation with moment generating functions.

Achlioptas [1] presented a still slightly different view of the proof, and more significantly, he showed that $T$ can be generated in a computationally simpler way. Namely, he proved that the entries of $A$ can be chosen as independent $\pm 1$ random variables (each attaining values $+1$ and $-1$ with probability $\frac{1}{2}$). Another variant of his result has the entries of $A$ attaining value 0 with probability $\frac{2}{3}$ and values $+1$ and $-1$ with probability $\frac{1}{6}$ each. This latter setting allows for computing the image $T(x)$ about 3 times faster than the former, since $A$ is sparse—only about one third of the entries are nonzero.

Recently Ailon and Chazelle [2] came up with an ingenious extension of this idea (speeding up the evaluation of $T$ by using a sparse random matrix $A$). A significant obstacle they had to overcome is that once $A$ becomes significantly sparse, with the fraction of nonzero entries tending to 0, the length of the image $\|T(x)\|$ is *not* sufficiently concentrated for some vectors, for example, for $x = (1, 0, 0, \ldots, 0)$. They proved that the concentration is sufficient even for $A$ sparse provided that the vector $x$ is "well-spread," which can be quantified as follows: Assuming $\|x\| = 1$,we require that $\|x\|_\infty = \max_j |x_j|$ be close to $\frac{1}{\sqrt{n}}$. This means that the "mass" of $x$ has to be distributed over many components.

In order to deal with vectors $x$ that are not well-spread, Ailon and Chazelle defined the matrix $A$ as the product $MHD$, where

- $M$ is a sparse $k \times n$ random matrix. Its entries are independent random variables, and each of them attains value 0 with probability $1-q$, and a value drawn from the normal distribution with zero mean and variance $\frac{1}{q}$ with probability $q$. Here $q \in (0, 1)$ is a "sparsity parameter," which can be chosen as $\frac{1}{n}$ times a factor polylogarithmic in $n$.

- $H$ is an $n \times n$ Walsh matrix (assuming $n$ to be a power of 2). The important thing is that $H$ acts as a (scaled) isometry and that, given $x$, the product $Hx$ can be evaluated by $O(n \log n)$ arithmetic operations by a Fast Fourier Transform algorithm.

- $D$ is a diagonal matrix with independent random $\pm 1$ entries.

Let $x$ be a fixed unit vector and let $y = \frac{1}{\sqrt{n}} HDx$. Then $\|y\| = 1$ and it is not difficult to show that $\|y\|_\infty = O(\sqrt{(\log n)/n}\,)$ with high probability. Therefore, with high probability, the sparse matrix $M$ is applied only to well-spread unit vectors $y$, and the length of the image is concentrated as needed.

**This paper.** We first consider the version of the Johnson–Lindenstrauss lemma (Theorem 3.1 below) that allows the entries of $A$ to be arbitrary independent random variables with zero mean, unit variance, and a subgaussian tail. We give a complete and self-contained proof in Section 3, for expository purposes and also as a preparation for the subsequent development. Our calculation is similar in spirit to that of Achlioptas, but slightly simpler and divided into several independent conceptual steps, which should hopefully make it easier to grasp.

After a preliminary version of the present paper was written, I learned about a preprint of Indyk and Naor [9]. It contains, as Remark 3.1, a very concise proof (noted by Assaf Naor long ago but not published before) of Theorem 3.1. That proof is similar in many respects to ours, but it uses an additional beautiful trick (introducing a new "artificial" Gaussian random variable at a right place and taking expectation with respect to it), which makes it, on the one hand, slightly shorter and, on the other hand, less pedestrian. It should also be mentioned that Klartag and Mendelson [11] proved a new and even more general version of the Johnson–Lindenstrauss lemma, with the same conditions on the projection matrix (as in Theorem 3.1) but with the $\log n$ factor in the dimension bound replaced by a structural parameter of the considered $n$-point set, which is always bounded by $O(\log n)$ but can be smaller for some sets. Their proof use the celebrated Majorizing Measure Theorem from the theory of stochastic processes and other sophisticated tools.

We continue our development by essentially re-proving the main technical part of the Ailon–Chazelle result (more precisely, the result concerning embeddings in $\mathbf{R}^k$ with Euclidean norm; they also consider the case of $\ell_1$ norm, which we will discuss separately). We need to add only two simple steps to our previous proof of the Achlioptas-style result. On the technical side, we do for the Ailon–Chazelle result what Achlioptas did for the Indyk–Motwani version: We replace normal distribution in the definition of the sparse matrix $M$ with a (suitably scaled) $\pm 1$ distribution. This makes $M$ computationally simpler to generate, and for input points with integer coordinates, all calculations can be done in integer arithmetic. All of our proofs up to this point are self-contained and use nothing beyond elementary calculus and probability theory.

In Section 5 we discuss a variant of the Johnson–Lindenstrauss lemma where the embedding $T$ goes in the space $\mathbf{R}^k$ with the $\ell_1$ norm $\|.\|_1$, rather than with the Euclidean norm. This case was also investigated by Ailon and Chazelle in the context of embeddings with a sparse matrix (they actually need this setting for an algorithmic application). They obtained sparsity whose dependence on $n$ is better than for the Euclidean case, while the dependence on $\varepsilon$ is worse.

We indicate how this result can be re-proved by our approach, and we again provide an analogous result with sparse matrices whose nonzero entries are $\pm 1$. However, this time the sparsity achieved for this $\pm 1$ case is worse, by a factor of $\varepsilon$, than that in the case of matrices with nonzero entries distributed normally. In this part our treatment is somewhat less detailed and we also use nontrivial results of probability theory.

We conclude the present section with several estimates for the exponential function, which will be used throughout the paper.

(E0) $1 + x \leq e^x$ for all $x \in \mathbf{R}$;

(E1) $e^x \leq 1 + 2x$ for all $x \in [0, 1]$;

(E2) $e^x \leq 1 + x + x^2$ for all $x \leq 1$;

(E3) $\frac{1}{2}(e^x + e^{-x}) \leq e^{x^2/2}$ for all $x \in \mathbf{R}$.

All of these can be verified by elementary calculus. Moreover, (E0) is standard and (E3) is often used in proofs of Chernoff-type inequalities.

## 2 Subgaussian tails and a Chernoff-type inequality

There is an extensive literature concerning concentration of random variables around their expectation, and because of phenomena related to the Central Limit Theorem, tail bounds similar to the tail of the standard normal distribution play a prominent role. We introduce the following convenient terminology:

**Definition 2.1** *Let $X$ be a real random variable with $\mathbf{E}[X] = 0$. We say $X$ has a* subgaussian upper tail *if there exists a constant $a > 0$ such that for all $\lambda > 0$,*

$$\mathrm{Prob}[X > \lambda] \leq e^{-a\lambda^2}.$$

*We say that $X$ has a* subgaussian upper tail up to $\lambda_0$ *if the previous bound holds for all $\lambda \leq \lambda_0$. We say that $X$ has a* subgaussian tail *if both $X$ and $-X$ have subgaussian upper tails.*

If $X_1, X_2, \ldots, X_n$ is a sequence of random variables, by saying that they have a *uniform subgaussian tail* we mean that all of them have subgaussian tails with the same constant $a$.

One of the first examples of a random variable with a subgaussian tail encountered in a course on the probabilistic method is provided by the Chernoff inequality (which should probably be more properly attributed to Bernstein). If $X = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are independent random variables with each $X_i$ attaining values $+1$ and $-1$, each with probability $\frac{1}{2}$, then $\mathrm{Var}[X] = n$ and the normalized random variable $Y = \frac{X}{\sqrt{n}}$ has a subgaussian tail.

There are several directions of generalizing this fact, some of them quite deep. Most of the generalizations encountered in the literature require that the individual variables contributing to $Y$ be *bounded*—say attain values in the interval $[-1, 1]$. The starting point of the present paper is the observation (which may not be new, but it doesn't commonly occur in textbooks and surveys) that the boundedness assumption may be replaced by requiring a subgaussian tail of each $X_i$. The usual proof of Chernoff-type inequalities then goes through with only a small enhancement.

For our subsequent use, it will be convenient to allow for $Y$ to be a linear combination of the $X_i$ with arbitrary coefficients (normalized so that the variance of $Y$ is 1):

**Lemma 2.2** *Let $X_1, \ldots, X_n$ be independent random variables, satisfying $\mathbf{E}[X_i] = 0$, $\mathrm{Var}[X_i] = 1$, and having a uniform subgaussian tail. Let $\alpha_1, \ldots, \alpha_n$ be real coefficients satisfying $\alpha_1^2 + \cdots + \alpha_n^2 = 1$. Then the sum*

$$Y = \alpha_1 X_1 + \cdots + \alpha_n X_n$$

*has $\mathbf{E}[Y] = 0$, $\mathrm{Var}[Y] = 1$, and a subgaussian tail.*

Let us remark that the special case of the lemma with all the $X_i$ having the standard normal distribution is an immediate consequence of the

2-stability of the normal distribution (then $Y$ has the standard normal distribution as well).

The following lemma is commonly used as a step in proofs of Chernoff-type inequalities.

**Lemma 2.3 (Moment generating function and subgaussian tail)** *Let $X$ be a random variable with $\mathbf{E}[X] = 0$. If $\mathbf{E}\left[e^{uX}\right] \le e^{Cu^2}$ for some constant $C$ and for all $u > 0$, then $X$ has a subgaussian upper tail. If $\mathbf{E}\left[e^{uX}\right] \le e^{Cu^2}$ holds for all $u \in (0, u_0]$, then $X$ has a subgaussian upper tail up to $2Cu_0$.*

**Proof.** For all $u \in (0, u_0]$ and all $t \ge 0$ we have

$$\begin{aligned}
\mathrm{Prob}[X \ge t] &= \mathrm{Prob}\left[e^{uX} \ge e^{ut}\right] \\
&\le e^{-ut}\mathbf{E}\left[e^{uX}\right] \qquad \text{(by the Markov inequality)} \\
&\le e^{-ut+Cu^2}.
\end{aligned}$$

For $t \le 2Cu_0$ we can set $u = t/2C$, use the above estimate, and obtain $\mathrm{Prob}[X \ge t] \le e^{-t^2/4C}$. $\square$

Maybe less well known is the following partial converse:

**Lemma 2.4** *If $\mathbf{E}[X] = 0$, $\mathrm{Var}[X] = \mathbf{E}[X^2] = 1$ (this is the extra condition compared to Lemma 2.3), and $X$ has a subgaussian upper tail, then $\mathbf{E}\left[e^{uX}\right] \le e^{Cu^2}$ for all $u > 0$, where the constant $C$ depends only on the constant $a$ in the subgaussian tail.*[1]

**Proof.** Let $F$ be the distribution function of $X$; that is, $F(t) = \mathrm{Prob}[X < t]$. We have $\mathbf{E}\left[e^{uX}\right] = \int_{-\infty}^{\infty} e^{ut}\,\mathrm{d}F(t)$. We split the integration interval into two subintervals, corresponding to $ut \le 1$ and $ut \ge 1$, and in the first one we use (E2). So we calculate

$$\begin{aligned}
\int_{-\infty}^{1/u} e^{ut}\,\mathrm{d}F(t) \le \int_{-\infty}^{1/u} 1 + ut + u^2t^2\,\mathrm{d}F(t) &\le \int_{-\infty}^{\infty} 1 + ut + u^2t^2\,\mathrm{d}F(t) \\
&= 1 + u\mathbf{E}[X] + u^2\mathbf{E}[X^2] = 1 + u^2.
\end{aligned}$$

[1] A subgaussian upper tail plus $\mathbf{E}[X] = 0$ do *not* imply $\mathbf{E}\left[e^{uX}\right] \le e^{Cu^2}$ for $C$ depending only on the constant in the subgaussian tail ($\mathrm{Var}[X]$ can be arbitrarily large). Of course, if we have a (two-sided) subgaussian tail, then we get bounded variance and thus the conclusion of the lemma.

For the second interval we can estimate the integral by sum, for instance:

$$\int_{1/u}^{\infty} e^{ut}\,\mathrm{d}F(t) \le \sum_{k=1}^{\infty} e^{k+1}\mathrm{Prob}\left[X \ge \frac{k}{u}\right] \le \sum_{k=1}^{\infty} e^{2k}e^{-ak^2/u^2} = \sum_{k=1}^{\infty} e^{k(2-ak/u^2)}.$$

For $u \le \sqrt{a}/2$ we have $2 - ak/u^2 \le -a/2u^2$ and the sum is bounded by the geometric series with both first term and quotient $e^{-a/2u^2} \le e^{-1} < \frac{1}{2}$. So the sum is at most $2e^{-a/2u^2} = O(u^2)$ (to see the last estimate, we can start with $e^x \ge 1 + x > x$, take reciprocal values to obtain $e^{-x} \le \frac{1}{x}$ for $x > 0$, and finally substitute $x = a/2u^2$). Hence $\mathbf{E}\left[e^{uX}\right] \le 1 + O(u^2) \le e^{O(u^2)}$.

For $u > \sqrt{a}/2$, the largest terms in the considered sum are those with $k$ near $u^2/a$, and the sum is $O(e^{u^2/2a})$. So here we also arrive at $\mathbf{E}\left[e^{uX}\right] \le e^{O(u^2)}$ as desired. □

**Proof of Lemma 2.2.** We have $\mathbf{E}[Y] = 0$ by linearity of expectation, and $\mathrm{Var}[Y] = \sum_{i=1}^{n}\alpha_i^2\mathrm{Var}[X_i] = \sum_{i=1}^{n}\alpha_i^2 = 1$ since the variance is additive for independent random variables.

We want to check that $Y$ has a subgaussian tail. Since $\mathbf{E}\left[e^{uX_i}\right] \le e^{Cu^2}$ by Lemma 2.4, we have

$$\mathbf{E}\left[e^{uY}\right] = \prod_{i=1}^{n}\mathbf{E}\left[e^{u\alpha_i X_i}\right] \le e^{Cu^2(\alpha_1^2+\cdots+\alpha_n^2)} = e^{Cu^2},$$

and $Y$ has a subgaussian tail by Lemma 2.3 (and by symmetry). □

# 3 A Johnson-Lindenstrauss lemma with independent subgaussian projection coefficients

Here we present an elementary proof of the following version of Statement 1.2 (with the probability $1/n^2$ replaced with a new parameter $\delta$):

**Theorem 3.1** *Let $n$ be an integer, $\varepsilon \in (0, \frac{1}{2}]$, and $\delta \in (0,1)$, and let us set $k = C\varepsilon^{-2}\log\frac{\delta}{2}$, where $C$ is a suitable constant. Let us define a random linear map $T\colon \mathbf{R}^n \to \mathbf{R}^k$ by*

$$T(x)_i = \frac{1}{\sqrt{k}}\sum_{j=1}^{n} R_{ij}x_j, \qquad i = 1, 2, \ldots, k,$$

*where the $R_{ij}$ are independent random variables with* $\mathbf{E}[R_{ij}] = 0$,$\mathrm{Var}[R_{ij}] = 1$*, and a uniform subgaussian tail (the constant $C$ above depends on the constant $a$ in the subgaussian tail). Then for every $x \in \mathbf{R}^n$ we have*

$$\mathrm{Prob}\Big[(1-\varepsilon)\|x\| \leq \|T(x)\| \leq (1+\varepsilon)\|x\|\Big] \geq 1-\delta.$$

This contains both the Indyk–Motwani result and the results of Achlioptas (possibly with worse constants). As was mentioned in the introduction, another short and elementary proof can be found in [9]

The theorem is easily proved from the following proposition, which in turn can be derived using Lemma 2.2 and calculations very similar to those in the proof of Lemma 2.4.

**Proposition 3.2** *Let $k \geq 1$ be an integer. Let $Y_1, \ldots, Y_k$ be independent random variables with* $\mathbf{E}[Y_i] = 0$, $\mathrm{Var}[Y_i] = 1$*, and a uniform subgaussian tail. Then $Z = \frac{1}{\sqrt{k}}(Y_1^2 + Y_2^2 + \cdots + Y_k^2 - k)$ has a subgaussian tail up to $\sqrt{k}$.*

Assuming this proposition, the proof of Theorem 3.1 is standard:

**Proof of Theorem 3.1.** Let $x \in \mathbf{R}^n$ be a fixed unit vector and let us set $Y_i = \sum_{j=1}^n R_{ij}x_j$. Then by Lemma 2.2, $\mathbf{E}[Y_i] = 0$, $\mathrm{Var}[Y_i] = 1$, and the $Y_i$ have a uniform subgaussian tail. So Proposition 3.2 applies and shows that $Z = \frac{1}{\sqrt{k}}(Y_1^2 + \cdots + Y_k^2 - k)$ has a subgaussian tail up to $\sqrt{k}$. We note that for $x$ fixed and unit, the quantity $\|T(x)\|^2 - 1$ is distributed as $\frac{1}{\sqrt{k}}Z$. Thus, still with $x$ fixed and unit, $\mathrm{Prob}[\|T(x)\| \geq 1+\varepsilon] \leq \mathrm{Prob}\left[\|T(x)\|^2 \geq 1+2\varepsilon\right] = \mathrm{Prob}\left[Z \geq 2\varepsilon\sqrt{k}\right]$. Since we assume $\varepsilon \leq \frac{1}{2}$, we are in the allowed range and the last probability is at most $e^{-a(2\varepsilon\sqrt{k})^2} = e^{-4a\varepsilon^2 C\varepsilon^{-2}\log(2/\delta)} \leq \frac{1}{2}\delta$ for $C \geq \frac{1}{2a}$. The calculation showing $\mathrm{Prob}[\|T(x)\| \leq 1-\varepsilon] \leq \frac{1}{2}\delta$ is almost the same. $\square$

The real work in proving Proposition 3.2 is in the proof of the next lemma:

**Lemma 3.3** *If $Y$ is as the $Y_i$ in Proposition 3.2, then there are constants $C$ and $u_0$ such that for all $u \in [0, u_0]$ we have* $\mathbf{E}\left[e^{u(Y^2-1)}\right] \leq e^{Cu^2}$ *and* $\mathbf{E}\left[e^{u(1-Y^2)}\right] \leq e^{Cu^2}$.

**Proof.** We begin with the first inequality. First we note that $\mathbf{E}[Y^4]$ is finite (a constant); this follows from the subgaussian tail of $Y$ by direct calculation, or in a simpler way, from Lemma 2.4 and from $t^4 = O(e^t + e^{-t})$ for all $t$.

Let $F$ be the distribution function of $Y^2$; that is, $F(t) = \text{Prob}\left[Y^2 < t\right]$. We again split the integral defining $\mathbf{E}\left[e^{uY^2}\right]$ into two intervals, corresponding to $uY^2 \le 1$ and $uY^2 \ge 1$. That is,

$$\mathbf{E}\left[e^{uY^2}\right] = \int_0^{1/u} e^{ut}\,\mathrm{d}F(t) + \int_{1/u}^{\infty} e^{ut}\,\mathrm{d}F(t).$$

The first integral is estimated using (E2) by

$$\int_0^{1/u} 1 + ut + u^2t^2\,\mathrm{d}F(t) \le \int_0^{\infty} 1 + ut + u^2t^2\,\mathrm{d}F(t)$$

$$= 1 + u\mathbf{E}[Y^2] + u^2\mathbf{E}[Y^4] = 1 + u + O(u^2)$$

The second integral can be estimated by a sum:

$$\sum_{k=1}^{\infty} e^{k+1}\text{Prob}\left[Y^2 \ge k/u\right] \le 2\sum_{k=1}^{\infty} e^{2k}e^{-ak/u}.$$

We may assume that $u \le u_0 = a/4$; then $k(2-a/u) \le -ka/2u$, and the sum is of order $e^{-\Omega(1/u)}$. Similar to the proof of Lemma 2.4 we can bound this by $O(u^2)$, and for $\mathbf{E}[e^{uY^2}]$ we thus get the estimate $1 + u + O(u^2) \le e^{u+O(u^2)}$.

Then we calculate $\mathbf{E}\left[e^{u(Y^2-1)}\right] = \mathbf{E}\left[e^{uY^2}\right]e^{-u} \le e^{O(u^2)}$ as required.

The calculation for estimating $\mathbf{E}\left[e^{-uY^2}\right]$ is simpler, since (E2) gives $e^{-ut} \le 1 - ut + u^2t^2$ for all $t > 0$ and $u > 0$:

$$\mathbf{E}\left[e^{-uY^2}\right] = \int_0^{\infty} e^{-ut}\mathrm{d}F(t) \le \int_0^{\infty} 1 - ut + u^2t^2\mathrm{d}F(t)$$

$$= 1 - u\mathbf{E}[Y^2] + u^2\mathbf{E}[Y^4] \le 1 - u + O(u^2) \le e^{-u+O(u^2)}.$$

This yields $\mathbf{E}\left[e^{u(1-Y^2)}\right] \le e^{O(u^2)}$. $\square$

**Proof of Proposition 3.2.** For $Z = \frac{1}{\sqrt{k}}(Y_1^2 + \cdots + Y_k^2 - k)$ and $0 < u \le u_0\sqrt{k}$, with $u_0$ as in Lemma 3.3, we calculate

$$\mathbf{E}\left[e^{uZ}\right] = \mathbf{E}\left[e^{(u/\sqrt{k})(Y_1^2+\cdots+Y_k^2-k)}\right] =$$

$$= \mathbf{E}\left[e^{(u/\sqrt{k})(Y^2-1)}\right]^k \le (e^{Cu^2/k})^k = e^{Cu^2}.$$

Lemma 2.3 implies that $Z$ has a subgaussian upper tail up to $2C\sqrt{k} \ge \sqrt{k}$ (assuming, as we may, that $2C \ge 1$). The calculation for the lower tail is identical. □

# 4 Sparse projection matrices

Here we prove an analog of Statement 1.2 for a random mapping $T$ with a "sparse" matrix. As was explained in the introduction, this is similar to the results of Ailon and Chazelle, but the particular probability distribution for the matrix entries is computationally more convenient than theirs. As we have also mentioned, the statement cannot hold for all vectors $x$, but we must assume that $x$ has a sufficiently small $\ell_\infty$ norm (or something in that spirit).

**Theorem 4.1** *Let $n$, $\varepsilon \in (0, \frac{1}{2})$, $\delta \in (0,1)$, and $\alpha \in [\frac{1}{\sqrt{n}}, 1]$ be parameters. Let us set*

$$q = C_0\alpha^2 \log(n/\varepsilon\delta)$$

*for a sufficiently large constant $C_0$, and let us assume that $q \le 1$. Let $S$ be a random variable with*

$$S = \begin{cases} +q^{-1/2} & \text{with probability } \frac{1}{2}q, \\ -q^{-1/2} & \text{with probability } \frac{1}{2}q, \\ 0 & \text{with probability } 1-q. \end{cases}$$

*Let us set $k = C\varepsilon^{-2}\log\frac{4}{\delta}$ for a sufficiently large constant $C$ (assuming $k$ integral), and let us define a random linear mapping $T\colon \mathbf{R}^n \to \mathbf{R}^k$ by*

$$T(x)_i = \frac{1}{\sqrt{k}}\sum_{j=1}^n S_{ij}x_j, \qquad i = 1, 2, \ldots, k$$

*with the $S_{ij}$ independent and distributed as $S$. Then for every unit vector $x \in \mathbf{R}^n$ satisfying $\|x\|_\infty \le \alpha$ we have*

$$\mathrm{Prob}\Big[(1-\varepsilon)\|x\| \le \|T(x)\| \le (1+\varepsilon)\|x\|\Big] \ge 1-\delta.$$

First we check that the components $T(x)_i$ have subgaussian tails at least up to a suitable threshold.

**Lemma 4.2** *Let $\alpha^2 \le q \le 1$, let $x \in \mathbf{R}^n$ be a unit vector with $\|x\|_\infty \le \alpha$, and let $Y = \sum_{j=1}^n S_j x_j$, where the $S_j$ are independent and distributed as $S$ in Theorem 4.1. Then $Y$ has a subgaussian tail up to $\sqrt{2q}/\alpha$.*

**Proof.** We have

$$\begin{aligned}
\mathbf{E}\left[e^{uY}\right] &= \prod_{j=1}^n \mathbf{E}\left[e^{ux_jS}\right] \\
&= \prod_{j=1}^n \Big(\frac{q}{2}\, e^{ux_j/\sqrt{q}} + \frac{q}{2}\, e^{-ux_j/\sqrt{q}} + 1 - q\Big) \\
&\le \prod_{j=1}^n \Big(q e^{u^2x_j^2/2q} + 1 - q\Big)
\end{aligned}$$

using (E3). For $u \le u_0 = \sqrt{2q}/\alpha$ we have $u^2x_j^2/2q \le 1$ for all $j$, and so we can estimate, using (E1), $qe^{u^2x_j^2/2q} + 1 - q \le q(1+u^2x_j^2/q) + 1 - q = 1+u^2x_j^2 \le e^{u^2x_j^2}$. Then $\mathbf{E}\left[e^{uY}\right] \le e^{u^2\|x\|^2} = e^{u^2}$, and the proof of Lemma 4.2 is finished by applying Lemma 2.3. □

**Proof of Theorem 4.1.** Let $x$ be fixed and let $Y = \sum_{j=1}^n S_j x_j$ be the random variable as in Lemma 4.2. By the same reasoning as in the proof of Theorem 3.1, we get that $\mathbf{E}[Y^2] = 1$ (using $\mathrm{Var}[S] = 1$) and that $\|T(x)\|^2 - 1$ is distributed as $\frac{1}{\sqrt{k}}\, Z$, where $Z = \frac{1}{\sqrt{k}}(Y_1^2 + \cdots + Y_k^2 - k)$. Thus it suffices to prove

$$\mathrm{Prob}\left[|Z| \ge 2\varepsilon\sqrt{k}\right] \le \delta.$$

This time we cannot use Proposition 3.2 for $Y$ directly, since the subgaussian tail of $Y$ is guaranteed only up to the threshold $\lambda_0 = \frac{\sqrt{q}}{\alpha}$. (It can actually be shown that $Y$ may indeed fail to have a subgaussian tail everywhere.)

A way of bringing Proposition 3.2 into play is to define a new random variable $\tilde{Y}$ by

$$\tilde{Y} = \begin{cases} Y & \text{if } |Y| \le \lambda_0, \\ 0 & \text{otherwise.} \end{cases}$$

The $k$ independent copies of $\tilde{Y}$ are denoted by $\tilde{Y}_1, \ldots, \tilde{Y}_k$, and $\tilde{Z} = \frac{1}{\sqrt{k}}(\tilde{Y}_1^2 + \cdots + \tilde{Y}_k^2 - k)$. Since $Y$ has a subgaussian tail up to $\lambda_0$, we have $\mathrm{Prob}\left[\tilde{Y} \ne Y\right] \le 2e^{-a\lambda_0^2}$. Then we can write

$$\mathrm{Prob}\left[|Z| \ge 2\varepsilon\sqrt{k}\right] \le \mathrm{Prob}\left[|\tilde{Z}| \ge 2\varepsilon\sqrt{k}\right] + \mathrm{Prob}\left[\tilde{Y}_i \ne Y_i \text{ for some } i\right]$$

$$\le \mathrm{Prob}\left[|\tilde{Z}| \ge 2\varepsilon\sqrt{k}\right] + ke^{-a\lambda_0^2}.$$

With our choice of parameters, the second term is at most $\frac{\delta}{2}$, and so it suffices to deal with the first term.

Now $\tilde{Y}$ has a subgaussian tail, and it might seem that Proposition 3.2 can be used immediately to estimate $\mathrm{Prob}\left[|\tilde{Z}| \ge 2\varepsilon\sqrt{k}\right]$. However, by passing from $Y$ to $\tilde{Y}$ we possibly decreased the variance, so the assumptions of Proposition 3.2 are not met by $\tilde{Y}$. Indeed, if we knew nothing else about $Y$ but that it has zero mean, unit variance, and a subgaussian tail up to $\lambda_0$, it could even happen that $\mathrm{Var}[\tilde{Y}] = 0$ (for example, we could set $M = e^{a\lambda_0^2} > \lambda_0^2$ and have $Y$ attain value $\sqrt{M}$ with probability $\frac{1}{2M}$, value $-\sqrt{M}$ with probability $\frac{1}{2M}$, and value 0 otherwise). So we need to use more information on $Y$. In the considered case, we know that $\max|Y| \le n\alpha/\sqrt{q}$. Therefore,

$$1 = \mathbf{E}[Y^2] \le \mathbf{E}[\tilde{Y}^2] + \max(Y^2) \cdot \mathrm{Prob}[|Y| \ge \lambda_0] \le$$

$$\le \mathbf{E}[\tilde{Y}^2] + \frac{n^2\alpha^2}{q} e^{-a\lambda_0^2} \le \mathbf{E}[\tilde{Y}^2] + \varepsilon.$$

Let us write $\mathrm{Var}[\tilde{Y}] = 1 - \nu \ge 1 - \varepsilon$. The random variable $\tilde{\tilde{Y}} = \frac{1}{\sqrt{1-\nu}}\tilde{Y}$ already has $\mathbf{E}[\tilde{\tilde{Y}}] = 0$, $\mathrm{Var}[\tilde{\tilde{Y}}] = 1$, and a subgaussian tail, so we can apply Proposition 3.2 to it, and obtain that for all positive $\lambda \le \sqrt{k}$ we have

$$\mathrm{Prob}\left[\tilde{\tilde{Y}}_1^2 + \cdots + \tilde{\tilde{Y}}_k^2 \ge k + \lambda\sqrt{k}\right] \le e^{-a\lambda^2}$$

and

$$\mathrm{Prob}\left[\tilde{\tilde{Y}}_1^2 + \cdots + \tilde{\tilde{Y}}_k^2 \le k - \lambda\sqrt{k}\right] \le e^{-a\lambda^2}.$$

Then

$$\mathrm{Prob}\left[\tilde{Z} \ge 2\varepsilon\sqrt{k}\right] = \mathrm{Prob}\left[\tilde{Y}_1^2 + \cdots + \tilde{Y}_k \ge (1+2\varepsilon)k\right] =$$

$$= \mathrm{Prob}\left[\tilde{\tilde{Y}}_1^2 + \cdots + \tilde{\tilde{Y}}_k^2 \ge \frac{1+2\varepsilon}{1-\nu}k\right] \le \mathrm{Prob}\left[\tilde{\tilde{Y}}_1^2 + \cdots + \tilde{\tilde{Y}}_k^2 \ge k + 2\varepsilon k\right] \le$$

$$\le e^{-4a\varepsilon^2 k} \le \frac{\delta}{4},$$

and

$$\mathrm{Prob}\left[\tilde{Z} \le -2\varepsilon\sqrt{k}\right] = \mathrm{Prob}\left[\tilde{Y}_1^2 + \cdots + \tilde{Y}_k^2 \le (1-2\varepsilon)k\right] =$$

$$= \mathrm{Prob}\left[\tilde{\tilde{Y}}_1^2 + \cdots + \tilde{\tilde{Y}}_k^2 \le \frac{1-2\varepsilon}{1-\nu}k\right] \le \mathrm{Prob}\left[\tilde{\tilde{Y}}_1^2 + \cdots + \tilde{\tilde{Y}}_k^2 \le (1-\varepsilon)k\right] \le$$

$$\le e^{-a\varepsilon^2 k} \le \frac{\delta}{4}.$$

Theorem 4.1 is proved. □

**Remark.** The sparsity parameter $q$ in Theorem 4.1 is essentially optimal in the following sense: If $\alpha$ is not too large, say $\alpha \le n^{-0.1}$, and if we set $q = \alpha^2\varphi$, where $1 \le \varphi = \varphi(n) = o(\log n)$, then the random variable $Z = \frac{1}{\sqrt{k}}(Y_1^2 + \cdots + Y_k^2 - k)$ doesn't necessarily have a subgaussian tail up to $\sqrt{k}$, e.g., for $k = \log n$ (which is a value of interest for the Johnson–Lindenstrauss lemma with $\varepsilon = \frac{1}{2}$). To see this, we consider an $x$ with $m = \alpha^{-2}$ components equal to $\alpha$ and the rest zero. In this case $Y_1$ is the sum of $m$ independent random variables, attaining values $\pm\varphi^{-1/2}$ with probability $\frac{\varphi}{2m}$ and value 0 otherwise. The probability of the event $E$ that $r$ of these variables give $+\varphi^{-1/2}$ and the rest 0 is $\binom{m}{r}(\frac{\varphi}{2m})^r(1-\frac{\varphi}{2m})^{m-r}$. For $r$ not too large this is roughly $(\frac{\mathrm{const}\cdot\varphi}{r})^r e^{-\mathrm{const}\cdot\varphi}$. Setting $r = \sqrt{2k\varphi}$, we calculate that the last expression is at least $n^{-o(1)}$. If the event $E$ occurs, it causes deviation of $Y_1^2$ from its expectation by $r^2/\varphi = 2k$, and hence a deviation of $Z$ by at least $\sqrt{k}$. However, if $Z$ had a subgaussian upper tail up to $\sqrt{k}$, such a deviation should have probability at most $e^{-\Omega(k)} = n^{-\Omega(1)}$.

## 5 The $\ell_1$ case

As was mentioned in the introduction, for some algorithmic purposes it is also interesting to investigate almost-isometric embeddings of an $n$-point set in a Euclidean space into the normed space $(\mathbf{R}^k, \|.\|_1)$. It turns out that the basic version of the Johnson–Lindenstrauss lemma can be proved with the same dependence of $k$ on $n$ and $\varepsilon$. (To prevent confusion, we stress that we are talking about $\ell_1$ norm in the *target* space. No analog of the Johnson–Lindenstrauss lemma holds for point sets in $\mathbf{R}^n$ with the $\ell_1$ norm; such sets generally cannot be almost-isometrically "flattened" to a logarithmic dimension [4], [13].)

When considering embeddings in $\mathbf{R}^k$ with the Euclidean metric, we were dealing with a random quantity of the form $\sum_{i=1}^k Y_i^2$, where $Y_i = \sum_{j=1}^n R_{ij}x_j$, $\|x\| = 1$ is unit, and the $R_{ij}$ were suitable independent random variables with zero expectation and unit variance. For $\|.\|_1$ as the target norm we need to deal with the quantity $\sum_{i=1}^k |Y_i|$ instead.

To prove a Johnson–Lindenstrauss type result, we need to verify that the considered quantity has the right expectation, and that it is sufficiently concentrated. In the $\ell_1$ case there is no problem with the concentration: actually, weaker conditions on the $Y_i$ suffice than for the Euclidean case. However, the expectation poses a new challenge. Indeed, in the Euclidean case, $\mathbf{E}[Y_i^2]$ is just the variance of $Y_i$, which equals $\sum_{j=1}^n \mathrm{Var}\,[R_{ij}]\,x_j^2$ and is thus *exactly the same* for all (unit) $x$. In contrast, $\mathbf{E}[|Y_i|]$ generally does depend on $x$. For example, if the random coefficients $R_{ij}$ are independent uniform $\pm 1$, then Khintchine's inequality from the geometry of Banach spaces tells us that $\mathbf{E}[|Y_i|]$ is between two absolute constants, but this is all one can say in general. Hence we generally do not obtain an Achlioptas-style result, analogous to Theorem 1.1, for $\|.\|_1$ as the target norm. (More precisely, we can get an embedding with distortion bounded by a constant in this way, but not with distortion arbitrarily close to 1.)

There is one distribution of the random coefficients $R_{ij}$ where this obstacle doesn't arise. If the $R_{ij}$ are independent with the standard normal distribution $N(0, 1)$, then the $Y_i$ have exactly the standard normal distribution, and hence $\mathbf{E}[|Y_i|]$ doesn't depend on $x$. (We have $\mathbf{E}[|Y_i|] = \sqrt{2/\pi}$, as is well known and not difficult to calculate.) This is a consequence of a remarkable property of the normal distribution, known as 2-*stability*: If $X$ and $Y$ are independent and normally distributed, then $X + Y$ is normally distributed too (and $\mathrm{Var}\,[X + Y] = \mathrm{Var}\,[X] + \mathrm{Var}\,[Y]$ by independence).

Fortunately, in the case considered by Ailon and Chazelle, we deal only with "well-spread" vectors $x$, and it turns out that in this case $\mathbf{E}[|Y_i|]$, although not completely independent of $x$, varies sufficiently little with $x$, even if the sparse matrix has the nonzero entries $\pm 1$. This is a consequence of a quantitative version of the Central Limit Theorem, as we will see below (Lemma 5.3). However, the sparsity we can afford for $\pm 1$ coefficients is worse (by a factor of $\varepsilon$) than the one obtained by Ailon and Chazelle, who consider nonzero entries distributed normally.

Below we give an alternative proof of the $\ell_1$ result of Ailon and Chazelle using the tools built above, and we also establish the version with $\pm 1$ coefficients.

**Theorem 5.1** *Let $n$, $\varepsilon \in (0, \frac{1}{2})$, $\delta \in (0,1)$, and $\alpha \in [\frac{1}{\sqrt{n}}, 1]$ be parameters, and let $k = C\varepsilon^{-2} \log \frac{4}{\delta}$ ($C$ a sufficiently large constant). Let $\beta_0 = \sqrt{2/\pi}$.*

(i) (Ailon and Chazelle [2]) *Let $q = 2\alpha^2/\varepsilon \le 1$, and let $T: \mathbf{R}^n \to \mathbf{R}^k$ be given by $T(x)_i = \frac{1}{\beta_0 k} \sum_{j=1}^n \sigma_{ij} x_j$, with the $\sigma_{ij}$ independent and attaining value 0 with probability $1-q$, and a value drawn from the normal distribution with zero mean and variance $\frac{1}{q}$ with probability $q$. Then*

$$\mathrm{Prob}\Big[(1-\varepsilon)\|x\| \le \|T(x)\|_1 \le (1+\varepsilon)\|x\|\Big] \ge 1-\delta.$$

*for every unit $x \in \mathbf{R}^n$ with $\|x\|_\infty \le \alpha$.*

(ii) *Let $q = C_1\alpha^2/\varepsilon^2 \le 1$ for a suitable constant $C_1$, and let $T: \mathbf{R}^n \to \mathbf{R}^k$ be given by $T(x)_i = \frac{1}{\beta_0 k} \sum_{j=1}^n S_{ij} x_j$, where the $S_{ij}$ are as in Theorem 4.1 (0 with probability $1-q$, and $+q^{-1/2}$ or $-q^{-1/2}$ with probability $\frac{1}{2}q$ each). Then*

$$\mathrm{Prob}\Big[(1-\varepsilon)\|x\| \le \|T(x)\|_1 \le (1+\varepsilon)\|x\|\Big] \ge 1-\delta.$$

*for every unit $x \in \mathbf{R}^n$ with $\|x\|_\infty \le \alpha$.*

We begin the proof with a concentration result, which is an $\ell_1$ version of Proposition 3.2.

**Proposition 5.2** *Let $k \ge 1$ be an integer. Let $Y_1, \ldots, Y_k$ be independent random variables with $\mathbf{E}[Y_i] = 0$, $\mathrm{Var}[Y_i] = 1$, $\mathbf{E}[|Y_i|] = \beta$ (where $\beta \in (0,1)$ is a constant) and a uniform* exponential *tail, meaning that for some constant $b > 0$ and all $t \ge 0$ we have $\mathrm{Prob}[Y_i \ge t] \le e^{-bt}$ and $\mathrm{Prob}[Y_i \le -t] \le e^{-bt}$. Then $Z_1 = \frac{1}{\sqrt{k}}(|Y_1| + |Y_2| + \cdots + |Y_k| - \beta k)$ has a subgaussian tail up to $\sqrt{k}$.*

**Proof.** First we imitate the proof of Lemma 3.3. We let $F$ be the distribution function of $|Y|$, where $Y$ satisfies the conditions imposed on the $Y_i$, and we estimate

$$\mathbf{E}\left[e^{u|Y|}\right] \le \int_0^{1/u} 1 + ut + u^2t^2 \,\mathrm{d}F(t) + \int_{1/u}^{\infty} e^{ut}\,\mathrm{d}F(t)$$

$$\le 1 + u\mathbf{E}[|Y|] + u^2\mathbf{E}[Y^2] + \sum_{k=1}^{\infty} e^{k+1}\mathrm{Prob}\left[|Y| \ge \frac{k}{u}\right]$$

$$\le 1 + \beta u + u^2 + \sum_{k=1}^{\infty} e^{2k} \cdot 2e^{-bk/u}.$$

The sum is the one we have encountered in the proof of Lemma 3.3, and for $u \le u_0 = b/4$ it is bounded by $O(u^2)$. So we obtain $\mathbf{E}[e^{u|Y|}] \le e^{\beta u + O(u^2)}$ for $u \le u_0$. Then, as in the proof of Proposition 3.2, for $0 < u \le u_0\sqrt{k}$ we have

$$\mathbf{E}\left[e^{uZ_1}\right] = \prod_{i=1}^{k} \mathbf{E}\left[e^{(u/\sqrt{k})(|Y_i|-\beta)}\right] \le (e^{O(u^2/k)})^k = e^{O(u^2)}.$$

The calculation for $\mathbf{E}\left[e^{-uZ_1}\right]$ is analogous but simpler, and we omit it. Hence $Z_1$ has a subgaussian tail up to $\sqrt{k}$, and Proposition 5.2 is proved. □

Next, we establish a result needed for dealing with the expectation of the $Y_i$.

**Lemma 5.3** *Let $x \in \mathbf{R}^n$ with $\|x\| = 1$ and $\|x\|_\infty \le \alpha$.*

(i) (Ailon and Chazelle [2]) *Let $Y = \sigma_1 x_1 + \cdots + \sigma_n x_n$, where the $\sigma_i$ are as the $\sigma_{ij}$ in Theorem 5.1(i) (with $q = 2\alpha^2/\varepsilon$). Then $\beta_0 - \frac{\varepsilon}{2} \le \mathbf{E}[|Y|] \le \beta_0$.*

(ii) *Let $Y = S_1 x_1 + \cdots + S_n x_n$, where the $S_j$ are as the $S_{ij}$ in Theorem 5.1(ii) (with $q = C_1\alpha^2/\varepsilon^2$). Then $\beta_0 - \frac{\varepsilon}{2} \le \mathbf{E}[|Y|] \le \beta_0 + \frac{\varepsilon}{2}$.*

**Proof.** For part (i), we essentially reproduce the neat argument in [2], for the reader's convenience and delight.

Let $I_1, \ldots, I_n$ be independent random variables, each attaining value 1 with probability $q$ and value 0 with probability $1 - q$. Then we can write

$Y = q^{-1/2}\sum_{j=1}^n I_j\gamma_j x_j$, with the $\gamma_j$ independent $N(0,1)$. We set $Z = \sum_{j=1}^n I_j x_j^2$. Conditioning on $Z = z$, the random variable $Y$ has the normal distribution with mean 0 and variance $\frac{z}{q}$. Hence $\mathbf{E}[|Y|\,|\,Z=z] = \beta_0\sqrt{z/q}$, and $\mathbf{E}[|Y|] = \beta_0 q^{-1/2}\mathbf{E}[\sqrt{Z}]$.

By convexity we have $\mathbf{E}[\sqrt{Z}] \le \sqrt{\mathbf{E}[Z]} = \sqrt{q}$. For bounding $\mathbf{E}[\sqrt{Z}\,]$ from below we use the inequality $\sqrt{1+t} \ge 1 + \frac{t}{2} - t^2$, valid for all $t \ge -1$, with $t = \frac{Z}{q} - 1$. Then $\mathbf{E}[\sqrt{Z}\,] = \sqrt{q}\cdot\mathbf{E}[\sqrt{1+t}\,] \ge \sqrt{q}(1 + \frac{1}{2}\mathbf{E}[Z/q-1] - \mathbf{E}[(Z/q-1)^2]) = \sqrt{q}(1 + 0 - q^{-2}\mathrm{Var}\,[Z])$. Now $\mathrm{Var}\,[Z] = \sum_{j=1}^n x_j^4 \mathrm{Var}\,[I_j] \le \sum_{j=1}^n x_j^4 q \le q\alpha^2\sum_{j=1}^n x_j^2 = q\alpha^2$, and hence $\mathbf{E}[\sqrt{Z}] \ge \sqrt{q}(1-\alpha^2/q) \le \sqrt{q}(1-\frac{\varepsilon}{2})$. This proves part (i).

For part (ii), we follow a derivation of a similar fact in König, Schütt, and Tomczak-Jaegermann [12], remark on page 20 (who in turn followed an advice of Schechtman). They use the following quantitative version of the Central Limit Theorem, related to the Berry–Esséen theorem: *Let $X_1, X_2, \ldots, X_n$ be independent symmetric random variables (symmetric meaning that $-X_i$ has the same distribution as $X_i$) with $\sum_{i=1}^n \mathbf{E}[X_i^2] = 1$, let $F(t) = \mathrm{Prob}[X_1 + X_2 + \cdots + X_n < t]$ be the distribution function of $X_1 + X_2 + \cdots + X_n$, and let $\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^t e^{-x^2/2}\,\mathrm{d}x$ be the distribution function of the standard normal distribution. Then*

$$|F(t) - \Phi(t)| \le \frac{C}{1+|t|^3}\cdot\sum_{i=1}^n \mathbf{E}[|X_i|^3]$$

*for all $t \in \mathbf{R}$, with $C$ a constant.*

We apply this with $X_i = S_i x_i$, where we have $\sum_{i=1}^n \mathbf{E}[|X_i|^3] = \sum_{i=1}^n q\cdot q^{-3/2}|x_j|^3 \le q^{-1/2}\alpha\sum_{i=1}^n x_i^2 = C_1^{-1/2}\varepsilon$. Then

$$\left|\mathbf{E}[|Y|] - \beta_0\right| = \left|\int_{-\infty}^{\infty}|t|\,\mathrm{d}F(t) - \int_{-\infty}^{\infty}|t|\,\mathrm{d}\Phi(t)\right|$$

$$\le \int_{-\infty}^{\infty}|F(t)-\Phi(t)|\,\mathrm{d}t \le \frac{C\varepsilon}{\sqrt{C_1}}\int_{-\infty}^{\infty}\frac{\mathrm{d}t}{1+|t|^3} \le \frac{\varepsilon}{2}$$

if $C_1$ is sufficiently large. (The first inequality in this chain can be derived by considering the integral $\int_{-K}^{K}$, integrating by parts, and then taking the limit for $K \to \infty$.) Lemma 5.3 is proved. □

**Proof of Theorem 5.1.** In both (i) and (ii), we consider $x$ unit and fixed and we set $\beta = \mathbf{E}[|Y_i|]$. Lemma 5.3 shows that $\beta \in [\beta_0 - \frac{\varepsilon}{2}, \beta_0 + \frac{\varepsilon}{2}]$. In order to apply Proposition 5.2, we need to check that the $Y_i$ have an exponential tail. (Unlike in the Euclidean case in the previous section, here we need not "trim" the $Y_i$.)

For case (ii), we have calculated in Lemma 4.2 that $\mathbf{E}\left[e^{uY_i}\right] \le e^{u^2}$ for $u \le u_0$. The proof of Lemma 2.3 gives $\mathrm{Prob}[Y_i \ge t] \le e^{-u_0 t + u_0^2} = O(e^{-u_0 t})$. This is almost an exponential (upper) tail, up to the multiplicative constant $e^{u_0^2}$. However, one can easily check that the multiplicative constant doesn't really matter in the proof of Proposition 5.2. (Alternatively, we could derive an exponential tail with multiplicative constant 1 by dealing with small $t$ separately using $\mathrm{Var}[Y_i] = 1$, the symmetry of $Y_i$, and the Chebyshev inequality.)

In order to obtain an exponential tail for the $Y_i$ in case (i), we can more or less repeat the calculation in the proof of Lemma 4.2, but we use the standard fact that for an $N(0,1)$ random variable $X$ we have $\mathbf{E}\left[e^{uX}\right] = e^{-u^2/2}$ for all $u \ge 0$. After that we proceed as above. This concludes the proof of Theorem 5.1. $\square$


## Acknowledgments

I would like to thank Assaf Naor for advising me the reference [12] as an appropriate source for a Khintchine-style inequality needed in the $\ell_1$ case and also for telling me about his paper with Indyk [9], Piotr Indyk for pointing out the reference [11], Uli Wagner for stimulating discussions about the Ailon–Chazelle paper, and Anastasios Zouzias and anonymous referees for pointing out mistakes and for useful comments.


# References


[1] D. Achlioptas. Database-friendly random projections: Johnson–Lindenstrauss with binary coins. *J. Comput. Syst. Sci.*, 66(4):671–687, 2003.

[2] N. Ailon and B. Chazelle. Approximate nearest neighbors and the fast Johnson–Lindenstrauss transform. In *Proc. 38th Annual ACM Symposium on Theory of Computing*, 2006.

[3] N. Alon. Problems and results in extremal combinatorics, I. *Discrete Math.*, 273:31–53, 2003.

[4] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in $\ell_1$. In *Proc. 44th IEEE Symposium on Foundations of Computer Science*, pages 514–523, 2003.

[5] S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Struct. Algorithms*, 22:60–65, 2003.

[6] P. Frankl and H. Maehara. The Johnson-Lindenstrauss lemmas and the sphericity of some graphs. *J. Combin. Theory Ser. B*, 44:355–362, 1988.

[7] P. Indyk. Algorithmic applications of low-distortion embeddings. In *Proc. 42nd IEEE Symposium on Foundations of Computer Science*, 2001.

[8] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proc. 30th Annual ACM Symposium on Theory of Computing*, pages 604–613, 1998.

[9] P. Indyk and A. Naor. Nearest neighbor preserving embeddings. *ACM Trans. Algorithms*, 3(3), 2007. Article no. 31.

[10] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemp. Math.*, 26:189–206, 1984.

[11] B. Klartag and S. Mendelson. Empirical processes and random projections. *J. Functional Analysis*, 225:229–245, 2005.

[12] H. Köning, C. Schütt, and N. Tomczak-Jaegermann. Projection constants of symmetric spaces and variants of Khintchine's inequality. *J. Reine Angew. Math.*, 511:1–42, 1999.

[13] J. R. Lee and A. Naor. Embedding the diamond graph in $L_p$ and dimension reduction in $L_1$. *Geom. Funct. Anal.*, 14(4):745–747, 2004.

[14] J. Matoušek. *Lectures on Discrete Geometry.* Springer, New York, 2002.