

Almost Optimal Explicit Johnson-Lindenstrauss Transformations

Daniel M. Kane[†] Raghu Meka[‡] Jelani Nelson[§]

Abstract

The Johnson-Lindenstrauss lemma is a fundamental result in probability with several applications in the design and analysis of algorithms. Constructions of linear embeddings satisfying the Johnson-Lindenstrauss property necessarily involve randomness and much attention has been given to obtain explicit constructions minimizing the number of random bits used. In this work we give explicit constructions with an almost optimal use of randomness: For $0 < \varepsilon, \delta < 1/2$, we obtain explicit generators $G : \{0, 1\}^r \rightarrow \mathbb{R}^{s \times d}$ for $s = O(\log(1/\delta)/\varepsilon^2)$ such that for all $w \in \mathbb{R}^d$, $\|w\| = 1$,

$$\Pr_{y \in_u \{0,1\}^r} [|\|G(y)w\|^2 - 1| > \varepsilon] \leq \delta,$$

with seed-length $r = O\left(\log d + \log(1/\delta) \cdot \log\left(\frac{\log(1/\delta)}{\varepsilon}\right)\right)$. In particular, for $\delta = 1/\text{poly}(d)$ and fixed $\varepsilon > 0$, we obtain seed-length $O((\log d)(\log \log d))$. Previous constructions required $\Omega(\log^2 d)$ random bits to obtain polynomially small error.

We also give a new elementary proof of the optimality of the JL lemma showing a lower bound of $\Omega(\log(1/\delta)/\varepsilon^2)$ on the embedding dimension. Previously, Jayram and Woodruff [JW11] used information and communication complexity techniques to show a similar bound.

1 Introduction

The celebrated Johnson-Lindenstrauss lemma (JLL) [JL84] is by now a standard technique for handling high dimensional data. Among its many known variants (see [AV99, DG03, IM98, Mat08]), we use the following version originally proven in [Ach03, AV06]¹.

Theorem 1. *For all $w \in \mathbb{R}^d$, $\|w\| = 1$, $0 < \varepsilon < 1/2$, $s \geq 1$,*

$$\Pr_{S \in_u \{1,-1\}^{s \times d}} [|\|(1/\sqrt{s})Sw\|^2 - 1| \geq \varepsilon] \leq C \cdot e^{-C'\varepsilon^2k}.$$

We say a family of random matrices has the JL property (or is a JL family) if the above condition holds. In typical applications of JLL, δ is taken to be $1/\text{poly}(d)$ and the goal is to embed a given set of $\text{poly}(d)$ points in d dimensions to $O(\log d)$ dimensions with distortion at most ε for a fixed constant ε . This is the setting we concern ourselves with.

Linear embeddings of Euclidean space as above necessarily require randomness as else one can take the vector w to be in the kernel of the fixed transformation. To formalize this we use the following definition.

¹Harvard University, Department of Mathematics. dankane@math.harvard.edu.

²University of Texas at Austin, Department of Computer Science. raghu@cs.utexas.edu.

³MIT Computer Science and Artificial Intelligence Laboratory. minilek@mit.edu.

¹Throughout, C, C' denote universal positive constants. For a multiset S , $x \in_u S$ denotes a uniformly random element of S . For a vector w , $\|w\|$ denotes the Euclidean norm of w .

Definition 2. For $\varepsilon, \delta > 0$, a generator $G : \{0, 1\}^r \rightarrow \mathbb{R}^{s \times d}$ is a $(d, s, \delta, \varepsilon)$ -JL generator if for every $w \in \mathbb{R}^d$, $\|w\| = 1$,

$$\Pr_{y \in_u \{0,1\}^r} [| \|G(y)w\|^2 - 1 | \geq \varepsilon] \leq \delta.$$

1.1 Derandomizing JLL

A simple probabilistic argument shows that there exists a $(d, O(\log(1/\delta)/\varepsilon^2), \delta, \varepsilon)$ -JL generator with seed-length $r = O(\log d + \log(1/\delta))$. On the other hand, despite much attention the best known explicit generators have seed-length at least $\Omega(\log(1/\delta) \log d)$ [CW09]. Besides being a natural problem in geometry as well as derandomization, an explicit JL generator with minimal randomness would likely help derandomize other geometric algorithms and metric embedding constructions. Further, having an explicit construction is of fundamental importance for streaming algorithms as storing the entire matrix (as opposed to the randomness required to generate the matrix) is often too expensive in the streaming context.

Our main result is an explicit generator that takes roughly $O(\log d \log \log d)$ random bits for constant ε and $\delta = 1/\text{poly}(d)$, and outputs a matrix $A \in \mathbb{R}^{s \times d}$ satisfying the JL property.

Theorem 3 (Main). *There exists a constant C such that for every $0 < \varepsilon, \delta < 1/2$, there exists an explicit $(d, C \log(1/\delta)/\varepsilon^2, \delta, \varepsilon)$ -JL generator $G : \{0, 1\}^r \rightarrow \mathbb{R}^{s \times d}$ with seed-length*

$$r = O \left(\log d + \log(1/\delta) \cdot \log \left(\frac{\log(1/\delta)}{\varepsilon} \right) \right).$$

Our constructions are elementary in nature using only standard tools in derandomization such as k -wise independence and oblivious samplers [Zuc97]. We give two different constructions. Our first construction is simpler and gives a generic template for derandomizing most known JL families. The second construction has the advantage of allowing fast matrix-vector multiplications: the matrix-vector product $G(y)w$ can be computed efficiently in time $O(d \log d + \text{poly}(\log(1/\delta)/\varepsilon))^2$.

Further, as one of the motivations for derandomizing JLL is its potential applications in streaming, it is important that the entries of the generated matrices be computable in small space. We observe that for any $i \in [s], j \in [d], y \in \{0, 1\}^r$, the entry $G(y)_{ij}$ can be computed in space $O(\log d \cdot \text{poly}(\log \log d))$ and time $O(d^{1+o(1)})$ (for fixed $\varepsilon, \delta > 1/\text{poly}(d)$). (See the proof of [Theorem 18](#) for the exact bound)

1.2 Optimality of JLL

We also give a new proof of the optimality of the JL lemma showing a lowerbound of $s_{\text{opt}} = \Omega(\log(1/\delta)/\varepsilon^2)$ for the target dimension. Previously, Jayram and Woodruff [JW11] used information and communication complexity techniques to show a similar bound in the case $s_{\text{opt}} < d^{1-\gamma}$ for some fixed constant $\gamma > 0$. In contrast, our argument is more direct in nature and is based on linear algebra and elementary properties of the uniform distribution on the sphere, and it only requires the assumption $s_{\text{opt}} < d/2$. Note the JLL is only interesting for $s_{\text{opt}} < d$.

²The computational efficiency does not follow directly from the dimensions of $G(y)$, as the matrices in our constructions are obtained by composing several matrices some of which are of dimension $\tilde{O}(\sqrt{d}) \times d$.

Theorem 4. *There exists a universal constant $c > 0$, such that for any distribution \mathcal{A} over linear transformations $A : \mathbb{R}^d \rightarrow \mathbb{R}^s$ with $s < d/2$, there exists a vector $w \in \mathbb{R}^d$, $\|w\| = 1$, such that*

$$\Pr_{S \sim \mathcal{A}} [| \|Sw\|^2 - 1 | > \varepsilon] \geq \exp(-c(s\varepsilon^2 + 1)).$$

1.3 Related Work

The ℓ_2 streaming sketch of Alon et al. [AMS99] implies an explicit JL family with seed-length $O(\log d)$ for embedding \mathbb{R}^d into \mathbb{R}^s with distortion ε and error δ , where $s = O(1/(\varepsilon^2\delta))$. Karnin et al. [KRS11] construct an explicit family for embedding \mathbb{R}^d into \mathbb{R}^s with $s = \text{poly}(1/(\varepsilon\delta))$, but with seed-length only $(1 + o(1)) \log d + O(\log^2 s)$ (i.e., nearly linear in $\log d$).

The works of Diakonikolas et al. [DKN10] and Meka and Zuckerman [MZ10] construct pseudorandom generators for d -variate degree 2 threshold functions achieving a seed-length of $\log d \cdot \text{poly}(1/\delta)$ for fooling with error at most δ . As derandomizing the JL lemma is a special case of fooling degree 2 threshold functions, these works give a JL family with seed-length $\log d \cdot \text{poly}(1/\delta)$.

The best known explicit JL family is the construction of Clarkson and Woodruff [CW09] who show that a random scaled Bernoulli matrix with $O(\log(1/\delta))$ -wise independent entries satisfies the JL lemma. We make use of their result in our construction.

We also note that there are efficient non-black box derandomizations of JLL, [EIO02], [Siv02]. These works take as input n points in \mathbb{R}^d , and deterministically compute an embedding (that depends on the input set) into $\mathbb{R}^{O(\log n)/\varepsilon^2}$ which preserves all pairwise distances between the given set of n points.

1.4 Outline of Constructions

For intuition, suppose that $\delta > 1/d^c$ is polynomially small and ε is a constant. Our constructions are based on a simple iterative scheme: We reduce the dimension from d to $\tilde{O}(\sqrt{d})$ ³ and iterate for $O(\log \log d)$ steps.

Generic Construction. Our first construction gives a generic template for reducing the randomness required in standard JL families and is based on the following simple observation. Starting with any JL family, such as the random Bernoulli construction of [Theorem 1](#), there is a trade-off that we can make between the amount of independence required to generate the matrix and the final embedding dimension. For instance, if we only desire to embed to a dimension of $O(\sqrt{d})$ (as opposed to $O(\log d)$), it suffices for the entries of the random Bernoulli matrix to be $O(1)$ -wise independent. We exploit this idea by iteratively decreasing the dimension from d to $\tilde{O}(\sqrt{d})$ and so on by using a random Bernoulli matrix with an increasing amount of independence at each iteration.

Fast JL Construction. Fix a vector $w \in \mathbb{R}^d$ with $\|w\| = 1$ and suppose $\delta = 1/\text{poly}(d)$. We first use an idea of Ailon and Chazelle [AC09] who give a family of unitary transformations \mathcal{R} from \mathbb{R}^d to \mathbb{R}^d such that for every $w \in \mathbb{R}^d$ and $V \in_u \mathcal{R}$, the vector Vw is *regular*, in the sense that $\|Vw\|_\infty = O(\sqrt{(\log d)/d})$, with high probability. We derandomize their construction using limited independence to get a family of rotations \mathcal{R} such that for $V \in_u \mathcal{R}$, $\|Vw\|_\infty = O(d^{-(1/2-\alpha)})$ with high probability, for a sufficiently small constant $\alpha > 0$.

³We say $f = \tilde{O}(g)$ if $f = O(g \cdot \text{polylog}(g))$.

We next observe that for a vector $w \in \mathbb{R}^d$, with $\|w\|_\infty = O(d^{-(1/2-\alpha)}\|w\|_2)$ projecting onto a random set of $O(d^{2\alpha} \log(1/\delta)/\varepsilon^2)$ coordinates preserves the ℓ_2 norm with distortion at most ε with high probability. We then note that the random set of coordinates can be chosen using oblivious samplers as in [Zuc97]. The idea of using samplers is due to Karnin et al. [KRS11] who use samplers for a similar purpose.

Finally, iterating the above scheme $O(\log \log d)$ times we obtain an embedding of \mathbb{R}^d to $\mathbb{R}^{\text{poly}(\log d)}$ using $O(\log d \log \log d)$ random bits. We then apply the result of Clarkson and Woodruff [CW09] and perform the final embedding into $O(\log(1/\delta)/\varepsilon^2)$ dimensions by using a random scaled Bernoulli matrix with $O(\log(1/\delta))$ -wise independent entries.

As all of the matrices involved in the construction are either Hadamard matrices or projection operators, the final embedding can actually be computed in $O(d \log d + \text{poly}(\log(1/\delta)/\varepsilon))$ time.

Outline of Lowerbound. To show a lowerbound on the embedding dimension s , we use Yao's min-max principle to first transform the problem to that of finding a hard distribution on \mathbb{R}^d , such that no single transformation can embed a random vector drawn from the distribution well with very high probability. We then show that the uniform distribution over the d -dimensional sphere is one such hard distribution. The proof of the last fact involves elementary linear algebra and some direct calculations.

2 Preliminaries

We first state the classical Khintchine-Kahane inequalities (cf. [LT91]) which gives tight moment bounds for linear forms.

Lemma 5 (Khintchine-Kahane). *For every $w \in \mathbb{R}^n$, $x \in_u \{1, -1\}^n$, $k > 0$,*

$$\mathbb{E}[|\langle w, x \rangle|^k] \leq k^{k/2} \mathbb{E}[|\langle w, x \rangle|^2]^{k/2} = k^{k/2} \|w\|^k.$$

We use randomness efficient oblivious samplers due to Zuckerman [Zuc97] (See Theorem 3.17 and the remark following the theorem in [Zuc97]).

Theorem 6 (Zuckerman [Zuc97]). *There exists a constant C such that for every $\varepsilon, \delta > 0$ there exists an explicit collection of subsets of $[d]$, $\mathcal{S}(d, \varepsilon, \delta)$, with each $S \in \mathcal{S}$ of cardinality $|S| = s(\varepsilon, \delta, d) = ((\log d + \log(1/\delta))/\varepsilon)^C$, such that for every function $f : [d] \rightarrow [0, 1]$,*

$$\Pr_{S \in_u \mathcal{S}} \left[\left| \frac{1}{s} \sum_{i \in S} f(i) - \mathbb{E}_{i \in_u [d]} f(i) \right| > \varepsilon \right] \leq \delta,$$

and there exists a NC algorithm that generates random elements of \mathcal{S} using $O(\log d + \log(1/\delta))$ random bits.

Corollary 7. *There exists a constant C such that for every $\varepsilon, \delta, B > 0$ there exists an explicit collection of subsets of $[d]$, $\mathcal{S}(d, B, \varepsilon, \delta)$, with each $S \in \mathcal{S}$ of cardinality $|S| = s(d, B, \varepsilon, \delta) = ((\log d + \log(1/\delta))B/\varepsilon)^C$, such that for every function $f : [d] \rightarrow [0, B]$,*

$$\Pr_{S \in_u \mathcal{S}} \left[\left| \frac{1}{s} \sum_{i \in S} f(i) - \mathbb{E}_{i \in_u [d]} f(i) \right| > \varepsilon \right] \leq \delta,$$

and there exists a NC algorithm that generates random elements of \mathcal{S} using $O(\log d + \log(1/\delta))$ random bits.

Proof. Follows by taking $\mathcal{S} = \mathcal{S}(d, \varepsilon/B, \delta)$ as in [Theorem 6](#) and using the condition of the theorem for $\bar{f} : [d] \rightarrow [0, 1]$ defined by $\bar{f}(i) = f(i)/B$. \blacksquare

Let $H_d \in \{-1/\sqrt{d}, 1/\sqrt{d}\}^{d \times d}$ be the normalized Hadamard matrix such that $H_d^T H_d = I_d$ (we drop the suffix d when dimension is clear from context). While the Hadamard matrix is known to exist for powers of 2, for clarity, we ignore this technicality and assume that it exists for all d . Finally, let S^{d-1} denote the Euclidean sphere $\{w : w \in \mathbb{R}^d, \|w\| = 1\}$.

The following definitions will be useful in giving an abstract description of our construction.

Definition 8. A distribution \mathcal{D} over $\mathbb{R}^{s \times d}$ is said to be an $(d, s, \delta, \varepsilon)$ -JL distribution if for any $w \in S^{d-1}$, $\Pr_{S \sim \mathcal{D}} [|\|Sw\|^2 - 1| > \varepsilon] < \delta$.

Definition 9. A distribution \mathcal{D} over $\mathbb{R}^{s \times d}$ is said to have the $(d, s, t, \delta, \varepsilon)$ -JL moment property if for any $w \in S^{d-1}$, $\mathbf{E}_{S \sim \mathcal{D}} [|\|Sw\|^2 - 1|^t] < \varepsilon^t \cdot \delta$.

Definition 10. A distribution \mathcal{D} is called a strong (d, s) -JL distribution if it is an $(d, s, \exp(-\Omega(\varepsilon^2 s)), \varepsilon)$ -JL distribution for all $\varepsilon > 0$. \mathcal{D} is said to have the strong (d, s) -JL moment property if it has the $(d, s, \ell, O(\sqrt{\ell/(\varepsilon^2 s)})^\ell, \varepsilon)$ -JL moment property for all $0 < \varepsilon < 1/2$ and integer $\ell \geq 2$.

Note that by [Theorem 1](#), random Bernoulli form a (d, s) -JL distribution. Sometimes we omit the d, s terms in the notation above if these quantities are clear from context, or if it is not important to specify them.

Throughout, we let logarithms be base-2 unless otherwise specified. We often assume various quantities, like $1/\varepsilon$ or $1/\delta$, are powers of 2; this is without loss of generality.

3 Strong JL Distributions and the Strong JL Moment Property

It is not hard to show that having the strong JL moment property and being a strong JL distribution are equivalent. We use the following fact, which is standard, but we provide a proof in the appendix for completeness.

Fact 11. Let Y, Z be nonnegative random variables such that $\Pr[Z \geq t] = O(\Pr[Y \geq t])$ for any $t \geq 0$. Then for $\ell \geq 1$ if $\mathbf{E}[Y^\ell] < \infty$, we have $\mathbf{E}[Z^\ell] = O(\mathbf{E}[Y^\ell])$.

Theorem 12. A distribution \mathcal{D} is a strong (d, s) -JL distribution if and only if it has the strong (d, s) -JL moment property.

Proof. First assume \mathcal{D} has the strong JL moment property. Then, for arbitrary $w \in S^{d-1}, \varepsilon > 0$,

$$\Pr_{S \sim \mathcal{D}} [|\|Sw\|^2 - 1| > \varepsilon] < \varepsilon^{-\ell} \cdot \mathbf{E}[|\|Sw\|^2 - 1|^\ell] < O(\sqrt{\ell/(\varepsilon^2 s)})^\ell.$$

The claim follows by setting $\ell = O(\varepsilon^2 s)$.

Now assume \mathcal{D} is a strong JL distribution. Set $Z = |\|Sw\|^2 - 1|$. Since \mathcal{D} is a strong JL distribution, the right tail of Z is big-Oh of that of the absolute value of the nonnegative random variable Y which is a Gaussian with mean 0 and variance $O(1/s)$. Now, apply [Fact 11](#). \blacksquare

Remark 13. [Theorem 12](#) implies that any strong JL distribution can be derandomized using $2 \log(1/\delta)$ -wise independence giving an alternate proof of the derandomized JL result of Clarkson and Woodruff ([Theorem 2.2](#) in) [[CW09](#)]. This is because, by Markov's inequality with ℓ even,

$$\Pr_{S \sim \mathcal{D}} [|\|Sw\|^2 - 1| > \varepsilon] < \varepsilon^{-\ell} \cdot \mathbf{E}_{S \sim \mathcal{D}} [(\|Sw\|^2 - 1)^\ell] \leq 2^{O(\ell)} \cdot (\varepsilon^{-1} \cdot \sqrt{\ell/s})^\ell. \quad (3.1)$$

Iterative dimensionality reduction:

// Output a matrix S distributed according to a $(d, s, \delta, \varepsilon)$ -JL distribution.

1. Define $m = \log((\log 1/\delta)/(2 \log \log 1/\delta))$, $\varepsilon' = \varepsilon/(e(m+2))$, $\delta' = \delta/(m+2)$.
2. Define $s_i = C(\varepsilon')^{-2} \delta'^{-1/2^i}$, $\ell_i = \Theta(2^i)$ an even integer for $i \geq 0$. Define $s_{-1} = d$.
3. Let S_i be a random matrix drawn from a distribution with the $(s_{i-1}, s_i, \ell_i, \delta', \varepsilon')$ -JL moment property for $i = 0, \dots, m$.
4. Let S_{final} be drawn from a $(s_m, O(\varepsilon^{-2} \log(1/\delta)), \delta', \varepsilon')$ -JL distribution.
5. $S \leftarrow S_{\text{final}} \cdot S_m \cdots S_0$.

Figure 1: A general derandomization scheme for distributions with JL moment properties.

Setting $\ell = \log(1/\delta)$ and $s = C\ell/\varepsilon^2$ for $C > 0$ sufficiently large makes the above probability at most δ . Now, note the ℓ th moment is determined by 2ℓ -wise independence of the entries of S .

4 A Generic JL Derandomization Template

Theorem 12 and Remark 13 provide the key insight for our construction. If we use $\ell = 2 \log(1/\delta)$ -wise independent Bernoulli entries suggested in Remark 13, the seed length would be $O(\ell \log d) = O(\log(1/\delta) \log d)$ for $s = \Theta(\varepsilon^{-2} \log(1/\delta))$. However, note that in Eq. (3.1), a trade-off can be made between the amount of independence needed and the final embedding dimension without changing the error probability. In particular, it suffices to use 4-wise independence if we embed into $s = \Omega(\varepsilon^{-2} \delta^{-1})$ dimensions. In general, if $s = C\varepsilon^{-2}q$ for $\log^2(1/\delta) \leq q \leq 1/\delta$, it suffices to set $\ell = O(\log_q(1/\delta))$ to make the right hand side of Eq. (3.1) at most δ . By gradually reducing the dimension over the course of several iterations, using higher independence in each iteration, we obtain shorter seed length.

Our main construction is described in Figure 1. We first embed into $O(\varepsilon^{-2} \delta^{-1})$ dimension using 4-wise independence. We then iteratively project from $O(\varepsilon^{-2} \delta^{1/2^i})$ dimensions into $O(\varepsilon^{-2} \delta^{1/2^{i+1}})$ dimensions until we have finally embedded into $O(\varepsilon^{-2} \log^2(1/\delta))$ dimensions. In our final step, we embed into the optimal target dimension using $2 \log(1/\delta)$ -wise independence. Note the Bernoulli distribution is not special here; we could use any family of strong JL distributions.

Theorem 14. *The output matrix S in Figure 1 is distributed according to a $(d, s, \delta, \varepsilon)$ -JL distribution for $s = O(\log(1/\delta)/\varepsilon^2)$.*

Proof. For a fixed vector w , let $w^i = S_i \cdots S_0 w$, and let w^{-1} denote w . Then by our choice of s_i and a Markov bound on the ℓ_i th moment,

$$\Pr [\|w^i\|^2 - \|w^{i-1}\|^2 > \varepsilon' \|w^{i-1}\|^2] < \varepsilon'^{-\ell_i} \cdot \mathbf{E}[(\|w^i\|^2 / \|w^{i-1}\|^2 - 1)^{\ell_i}] < \delta'$$

for $0 \leq i \leq m$. We also have $\Pr [\|S_{\text{final}} w^m\|^2 - \|S w^m\|^2 > \varepsilon' \|w^m\|^2] < \delta'$. By a union bound, $\|S_{\text{final}} w^m\|^2 \leq (1 + \varepsilon')^{m+2} \leq e^{(m+2)\varepsilon'} \leq 1 + \varepsilon$ with probability $1 - (m+2)\delta' = 1 - \delta$. ■

As a corollary, we obtain our main theorem, [Theorem 3](#).

Proof. [Proof of [Theorem 3](#)] We let the distributions in Steps 3 and 4 of [Figure 1](#) be strong JL distributions. Then the ℓ_i th moment of the error from \mathcal{D}_i is $\varepsilon'^{-\ell_i} \cdot O(\sqrt{\ell_i/(\varepsilon'^2 s_i)})^{\ell_i} = \delta'$, thus satisfying Step 3. We know Step 4 is satisfied by [Remark 13](#).

Now, notice our proof of [Theorem 14](#) only required a union bound over the $m + 2$ applications of the S_i and S_{final} . Thus, we can always use the same random seed across all levels. The seed length required to generate S_0 is $O(\log d)$. For S_i for $i > 0$ the seed length is $O(\ell_i \log(\varepsilon'^{-2} \delta'^{1/2^i})) = O(2^i \log(1/\varepsilon') + \log(1/\delta'))$, which is never larger than $O((\log(1/\delta')/\log \log(1/\delta')) \log(1/\varepsilon') + \log(1/\delta'))$, which is $O((\log(1/\delta)/\log \log(1/\delta)) \log(1/\varepsilon) + \log(1/\delta))$. The seed length required for S_{final} is $O(\log(1/\delta') \log(\log(1/\delta')/\varepsilon')) = O(\log(1/\delta) \log(\log(1/\delta)/\varepsilon))$. Thus, the total seed length is dominated by generating S_0 and S_{final} , giving the claim. \blacksquare

5 Explicit JL Families via Samplers

We now give an alternate construction of an explicit JL family. The construction is similar in spirit to that of the previous section and has the additional property that matrix-vector products for matrices output by the generator can be computed in time roughly $O(d \log d + s^3)$, as it is based on the Fast Johnson-Lindenstrauss Transform (FJLT) of [\[AC06\]](#). For clarity, we concentrate on the case of $\delta = \Theta(1/d^c)$ polynomially small. The case of general δ can be handled similarly with some minor technical issues⁴ that we skip in this extended abstract. Further, we assume that $\log(1/\delta)/\varepsilon^2 < d$ as else JLL is not interesting..

As outlined in the introduction, we first give a family of rotations to regularize vectors in \mathbb{R}^d . For a vector $x \in \mathbb{R}^d$, let $D(x) \in \mathbb{R}^{\dim \times d}$ be the diagonal matrix with $D(x)_{ii} = x_i$.

Lemma 15. *Let $x \in \{1, -1\}^n$ be drawn from a k -wise independent distribution. Then, for every $w \in \mathbb{R}^d$ with $\|w\| = 1$,*

$$\Pr[\|HD(x)w\|_\infty > n^{-(1/2-\alpha)}] \leq \frac{k^{k/2}}{n^{\alpha k - 1}}.$$

Proof. Let $v = HD(x)w$. Then, for $i \in [d]$, $v_i = \sum_j H_{ij} x_j w_j$ and $\mathbb{E}[v_i^2] = \sum_j H_{ij}^2 w_j^2 = 1/d$. By Markov's inequality and the Khintchine-Kahane inequality ([Lemma 5](#)),

$$\Pr[|v_i| > d^{-(1/2-\alpha)}] \leq \mathbb{E}[v_i^k] \cdot d^{(1/2-\alpha)k} \leq k^{k/2} d^{(1/2-\alpha)k} / d^{k/2} = k^{k/2} d^{-\alpha k}.$$

The claim now follows from a union bound over $i \in [d]$. \blacksquare

We now give a family of transformations for reducing d dimensions to $\tilde{O}(d^{1/2})$ dimensions using oblivious samplers. For $S \subseteq [d]$, let $\mathcal{P}_S : \mathbb{R}^d \rightarrow \mathbb{R}^{|S|}$ be the projection onto the coordinates in S . In the following let C be the universal constant from [Corollary 7](#).

Lemma 16. *Let $\mathcal{S} \equiv \mathcal{S}(d, d^{1/2C}, \varepsilon, \delta)$, $s = O(d^{1/2} \log^C(1/\delta)/\varepsilon^C)$ be as in [Corollary 7](#) and let \mathcal{D} be a k -wise independent distribution over $\{1, -1\}^d$. For $S \in_u \mathcal{S}$, $x \leftarrow \mathcal{D}$, define the random linear transformation $A_{S,x} : \mathbb{R}^d \rightarrow \mathbb{R}^s$ by $A_{S,x} = \sqrt{d/s} \cdot \mathcal{P}_S \cdot HD(x)$. Then, for every $w \in \mathbb{R}^d$ with $\|w\| = 1$,*

$$\Pr[|\|A_{S,x}(w)\|^2 - 1| \geq \varepsilon] \leq \delta + k^{k/2} / d^{k/4C - 1}.$$

⁴In case of very small δ , we need to ensure that we never increase the dimension - which can be done trivially by using the identity transformation. In case of large δ , we first embed the input vector into $O(1/\delta\varepsilon^2)$ dimensions using 4-wise independence as in [Section 4](#).

Proof. Let $v = HD(x)w$. Then, $\|v\| = 1$ and by [Lemma 15](#) applied for $\alpha = 1/4C$,

$$\Pr[\|v\|_\infty > d^{-(1/2-1/4C)}] \leq k^{k/2}/d^{k/4C-1}.$$

Now suppose that $\|v\|_\infty \leq d^{-(1/2-1/4C)}$. Define $f : [d] \rightarrow \mathbb{R}$ by $f(i) = d \cdot v_i^2 \leq d^{1/2C} = B$. Then,

$$\|A_{S,x}(w)\|^2 = (d/s)\|\mathcal{P}_S(v)\|^2 = \frac{1}{s} \sum_{i \in S} dv_i^2 = \frac{1}{s} \sum_{i \in S} f(i),$$

and $\mathbb{E}_{i \in_u [d]} f(i) = (1/d) \sum_i d \cdot v_i^2 = 1$. Therefore, by [Corollary 7](#),

$$\Pr[|\|A_{S,x}(w)\|^2 - 1| \geq \varepsilon] = \Pr_{S \in_u \mathcal{S}} \left[\left| \frac{1}{s} \sum_{i \in S} f(i) - \mathbb{E}_{i \in_u [d]} f(i) \right| \geq \varepsilon \right] \leq \delta.$$

The claim now follows. ■

We now recursively apply the above lemma. Fix $\varepsilon, \delta > 0$. Let $\mathcal{A}(d, k) : \mathbb{R}^d \rightarrow \mathbb{R}^{s(d)}$ be the collection of transformations $\{A_{S,x} : S \in_u \mathcal{S}, x \leftarrow D\}$ as in the above lemma for $s(d) = s(d, d^{1/2C}, \varepsilon, \delta) = c_1 d^{1/2} (\log d/\varepsilon)^C$, for a constant c_1 . Note that we can sample from $\mathcal{A}(d, k)$ using $r(d, k) = k \log d + O(\log d + \log(1/\delta)) = O(k \log d)$ random bits.

Let $d_0 = d$, and let $d_{i+1} = s(d_i)$. Let $k_0 = 8C(c+1)$ (recall that $\delta = 1/d^c$) and $k_{i+1} = 2^i k_0$. The parameters d_i, k_i are chosen so that $1/d_i^{k_i}$ is always polynomially small. Fix $t > 0$ to be chosen later so that $k_i < d_i^{1/4C}$ for $i < t$.

Lemma 17. *For $A_0 \in_u \mathcal{A}(d_0, k_0), A_1 \in_u \mathcal{A}(d_1, k_1), \dots, A_{t-1} \in_u \mathcal{A}(d_{t-1}, k_{t-1})$ chosen independently, and $w \in \mathbb{R}^d, \|w\| = 1$,*

$$\Pr[(1 - \varepsilon)^t \leq \|A_{t-1} \cdots A_1 A_0(w)\|^2 \leq (1 + \varepsilon)^t] \geq 1 - t\delta - \sum_{i=0}^{t-1} \frac{k_i^{k_i/2}}{d_i^{k_i/4C-1}}.$$

Proof. The proof is by induction on $i = 1, \dots, t$. For $i = 1$, the claim is same as [Lemma 16](#). Suppose the statement is true for $i - 1$ and let $v = A_{i-1} \cdots A_0(w)$. Then, $v \in \mathbb{R}^{d_i}$ and the lemma follows by [Lemma 16](#) applied to $\mathcal{A}(d_i, k_i)$, and v . ■

What follows is a series of elementary calculations to bound the seed-length and error from the above lemma. Observe that

$$d^{(1/2)^i} \leq d_i = d^{(1/2)^i} \cdot \left(\frac{c_1 \log^C(d)}{\varepsilon^C} \right)^{1+(1/2)+\dots+(1/2)^{i-1}} \leq d^{(1/2)^i} \left(\frac{c_1 \log^C d}{\varepsilon^C} \right)^2. \quad (5.1)$$

Let $t = O(\log \log d)$ be such that $2^t = \log d/4C \log \log d$. Then, $d_t \leq \log^{4C} d \cdot (c_1 \log^C d/\varepsilon^C)^2 = O(\log^{6C} d/\varepsilon^{2C})$, and for $i < t$,

$$k_i < k_t = 8C(c+1)2^t = 2(c+1) \log d / \log \log d < \log d = d^{(1/2)^t/4C} < d_t^{1/4C} < d_i^{1/4C}, \quad (5.2)$$

where we assumed that $\log \log d > 2c + 2$. Therefore, the error in [Lemma 17](#) can be bounded by

$$t\delta + \sum_{i=0}^{t-1} \frac{k_i^{k_i/2}}{d_i^{k_i/4C-1}} \leq t\delta + d \sum_{i=0}^{t-1} d_i^{-k_i/8C} \quad (\text{Equation 5.2})$$

$$\leq t\delta + d \sum_{i=0}^{t-1} (d^{1/2^i})^{-8C(c+1) \cdot 2^i/8C} \quad (\text{Equation 5.1})$$

$$\leq t\delta + t/d^c \leq 2t\delta \quad (\text{as } \delta > 1/d^c).$$

Note that,

$$k_i \log d_i \leq 8C(c+1) \cdot 2^i (\log d/2^i + 2C \log \log d + 2C \log(1/\varepsilon)) = O(\log d + \log d \log(1/\varepsilon) / \log \log d).$$

Therefore, the randomness needed after $t = O(\log \log d)$ iterations is

$$\sum_{i=0}^{t-1} O(k_i \log d_i) = O(\log d \log \log d + (\log d) \log(1/\varepsilon)).$$

Combining the above arguments (applied to $\delta' = \delta / \log \log d$ and $\varepsilon' = \varepsilon / \log \log d$ and simplifying the resulting expression for seed-length) we obtain our fast derandomized JL family.

Theorem 18 (Fast Explicit JL Family). *There exists a $(\varepsilon, \delta, d, O(\log(1/\delta)/\varepsilon^2))$ -JL generator with seed-length $r = O(\log d + \log(1/\delta)(\log(\log(1/\delta)/\varepsilon)))$ such that for every vector $w \in \mathbb{R}^d$, $y \in \{0, 1\}^r$, $G(y)w$ can be evaluated in time $O(d \log d + \text{poly}(\log(1/\delta)/\varepsilon))$.*

Proof. We suppose that $\delta = \Theta(1/d^c)$ - the analysis for the general case is similar.

From the above arguments there is an explicit generator that takes $O(\log(d/\delta) \cdot \log(\log(d/\delta)/\varepsilon))$ random bits and outputs a linear transformation $A : \mathbb{R}^d \rightarrow \mathbb{R}^m$ for $m = \text{poly}(\log(d/\delta), 1/\varepsilon)$, satisfying the JL property with error at most δ and distortion at most ε . The theorem now follows by composing the transformations of the above theorem with a Bernoulli matrix having $2 \log(1/\delta)$ -wise independence. The additional randomness required is $O(\log(1/\delta) \log m) = O(\log(1/\delta)(\log \log(d/\delta) + \log(1/\varepsilon)))$.

We next bound the time for computing matrix-vector products for the matrices we output. Note that for $i < t$, the matrices A_i of [Lemma 17](#) are of the form $\mathcal{P}_S \cdot H_{d_i} D(x)$ for a k -wise independent string $x \in \{1, -1\}^{d_i}$. Thus, for any vector $w_i \in \mathbb{R}^{d_i}$, $A_i w_i$ can be computed in time $O(d_i \log d_i)$ using the discrete Fourier transform. Therefore, for any $w = w_0 \in \mathbb{R}^{n_0}$, the product $A_{t-1} \cdots A_1 A_0 w_0$ can be computed in time

$$\begin{aligned} \sum_{i=0}^{t-1} O(d_i \log d_i) &\leq O(d \log d) + \log d \cdot \sum_{i=1}^{t-1} O\left(d^{1/2^i} (\log(1/\delta)/\varepsilon^2)^2\right) && \text{(Equation 5.1)} \\ &= O(d \log d + \sqrt{d} \log d \log^2(1/\delta)/\varepsilon^4). \end{aligned}$$

The above bound dominates the time required to perform the final embedding.

A similar calculation shows that for indices $i \in s, j \in [d]$, the entry $G(y)_{ij}$ of the generated matrix can be computed in space $O(\sum_i \log d_i) = O(\log d + \log(1/\varepsilon) \cdot \log \log d)$ by expanding the product of matrices and enumerating over all intermediary indices⁵. The time required to perform the calculation is $O(s \cdot d_t \cdot d_{t-1} \cdots d_1) = d \cdot (\log d/\varepsilon)^{O(\log \log d)}$. \blacksquare

Remark 19. *It is also possible to use the FJLT in the framework of [Figure 1](#) to get the same seed length with fast update time as above. Details are in [Section B](#) of the appendix.*

⁵We also need to account for the time and space needed by the samplers and for generating a k -wise independent strings. However, these are dominated by the task of enumerating over all indices; for instance, the samplers of [\[Zuc97\]](#) are in NC.

6 Optimality of JL Lemma

We next prove [Theorem 4](#), the optimality of the number of rows in the JL Lemma. Note that for a distribution \mathcal{D} , if $\Pr_{S \sim \mathcal{D}}[\|Sw\|^2 - 1] < \delta$ for any $w \in S^{d-1}$, then it must be the case that $\Pr_{S \sim \mathcal{D}}[\Pr_{w \in S^{d-1}}[\|Sw\|^2 - 1]] < \delta$. The following theorem shows that no $S \in \mathbb{R}^{k \times d}$ can have $\Pr_{w \in S^{d-1}}[\|Sw\|^2 - 1] < \delta$ unless k is at least as large as in the statement of the JL lemma.

Theorem 20. *If $S : \mathbb{R}^d \rightarrow \mathbb{R}^k$ is a linear transformation with $d > 2k$ and $\varepsilon > 0$ sufficiently small, then for w a randomly chosen vector in S^{d-1} , $\Pr[\|Sw\|^2 - 1] > \varepsilon] \geq \exp(-O(k\varepsilon^2 + 1))$.*

Proof. First we note that we can assume that S is surjective since if it is not, we may replace \mathbb{R}^k by the image of S . Let $V = \ker(A)$ and let $U = V^\perp$. Then $\dim(U) = k$, $\dim(V) = d - k$. Now, any $w \in \mathbb{R}^d$ can be written uniquely as $w_V + w_u$ where w_V and w_u are the components in V and U respectively. We may then write $w_V = r_V \Omega_V$, $w_u = r_u \Omega_u$, where r_V, r_u are positive real numbers and Ω_V and Ω_u are unit vectors in V and U respectively. Let $s_V = r_V^2$ and $s_u = r_u^2$. We may now parameterize the unit sphere by $(s_V, \Omega_V, s_u, \Omega_u) \in [0, 1] \times S^{d-k-1} \times [0, 1] \times S^{k-1}$, so that $s_V + s_u = 1$. It is clear that the uniform measure on the sphere is given in these coordinates by $f(s_u) ds_u d\Omega_V d\Omega_u$ for some function f . To compute f we note that $f(s_u)$ should be proportional to the limit as $\delta_1, \delta_2 \rightarrow 0^+$ of $(\delta_1 \delta_2)^{-1}$ times the volume of points w so that $\|w\|^2 \in [1, 1 + \delta_1]$ and $\|w_u\|^2 \in [s_u, s_u + \delta_2]$. Equivalently, $\|w_u\|^2 \in [s_u, s_u + \delta_2]$, and $\|w_V\|^2 \in [1 - \|w_u\|^2, 1 - \|w_u\|^2 + \delta_1]$. For fixed w_u , the latter volume is within $O(\delta_1 \delta_2)$ of the volume of w_V so that $\|w_V\|^2 \in [s_V, s_V + \delta_1]$. Now the measure on V is $r_V^{d-k-1} dr_V d\Omega_V$. Therefore it also is $\frac{1}{2} s_V^{(d-k-2)/2} ds_V d\Omega_V$. Therefore this volume over V is proportional to $s_V^{(d-k-2)/2} (\delta_1 + O(\delta_1 \delta_2 + \delta_1^2))$. Similarly the volume of w_u so that $\|w_u\|^2 \in [s_u, s_u + \delta_2]$ is proportional to $s_u^{(k-2)/2} (\delta_2 + O(\delta_2^2))$. Hence f is proportional to $s_V^{(d-k-2)/2} s_u^{(k-2)/2}$.

We are now prepared to prove the theorem. The basic idea is to first condition on Ω_V, Ω_u . We let $C = \|S\Omega_u\|^2$. Then if w is parameterized by $(s_V, \Omega_V, s_u, \Omega_u)$, $\|Sw\|^2 = C s_u$. Choosing w randomly, we know that $s = s_u$ satisfies the distribution $\frac{s^{(k-2)/2} (1-s)^{(d-k-2)/2}}{\beta((k-2)/2, (d-k-2)/2)} ds = f(s) ds$ on $[0, 1]$. We need to show that for any $c = \frac{1}{C}$, the probability that s is not in $[(1 - \varepsilon)c, (1 + \varepsilon)c]$ is $\exp(-O(\varepsilon^2 k))$. Note that $f(s)$ attains its maximum value at $s_0 = \frac{k-2}{d-4} < \frac{1}{2}$. Notice that $\log(f(s_0(1+x)))$ is some constant plus $\frac{k-2}{2} \log(s_0(1+x)) + \frac{d-k-2}{2} \log(1-s_0-x s_0)$. If $|x| < 1/2$, then this is some constant plus $-O(kx^2)$. So for such w , $f(s_0(1+x)) = f(s_0) \exp(-O(kx^2))$. Furthermore, for all x , $f(s_0(1+x)) = f(s_0) \exp(-\Omega(kx^2))$. This says that f is bounded above by a normal distribution and checking the normalization we find that $f(s_0) = \Omega(s_0^{-1} k^{1/2})$.

We now show that both $\Pr(s < (1 - \varepsilon)s_0)$ and $\Pr(s > (1 + \varepsilon)s_0)$ are reasonably large. We can lower bound either as

$$\begin{aligned} s_0 \int_\varepsilon^{1/2} f(s_0) \exp(-O(kx^2)) dx &\geq \Omega(k^{1/2}) \int_\varepsilon^{\varepsilon + k^{-1/2}} \exp(-O(kx^2)) dx \\ &\geq \Omega(\exp(-O(k(\varepsilon + k^{-1/2})^2))) \\ &\geq \exp(-O(k\varepsilon^2 + 1)). \end{aligned}$$

Hence since one of these intervals is disjoint from $[(1 - \varepsilon)c, (1 + \varepsilon)c]$, the probability that s is not in $[(1 - \varepsilon)c, (1 + \varepsilon)c]$ is at least $\exp(-O(k\varepsilon^2 + 1))$. \blacksquare

References

- [AC06] 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.
- [AC09] Nir Ailon and Bernard Chazelle. The fast Johnson-Lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.*, 39(1):302–322, 2009.
- [Ach03] Dimitris Achlioptas. Database-friendly random projections: Johnson-lindenstrauss with binary coins. *J. Comput. Syst. Sci.*, 66(4):671–687, 2003.
- [AMS99] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137 – 147, 1999.
- [AV99] Rosa I. Arriaga and Santosh Vempala. An algorithmic theory of learning: Robust concepts and random projection. In *FOCS*, pages 616–623, 1999.
- [AV06] Rosa I. Arriaga and Santosh Vempala. An algorithmic theory of learning: Robust concepts and random projection. *Machine Learning*, 63(2):161–182, 2006.
- [CCFC04] Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. *Theor. Comput. Sci.*, 312(1):3–15, 2004.
- [CW09] Kenneth L. Clarkson and David P. Woodruff. Numerical linear algebra in the streaming model. In *STOC*, pages 205–214, 2009.
- [DG03] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Struct. Algorithms*, 22(1):60–65, 2003.
- [DKN10] Ilias Diakonikolas, Daniel M. Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 11–20, 2010.
- [EIO02] Lars Engebretsen, Piotr Indyk, and Ryan O’Donnell. Derandomized dimensionality reduction with applications. In *SODA*, pages 705–712, 2002.
- [IM98] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *STOC*, pages 604–613, 1998.
- [JL84] W B Johnson and J Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemp Math*, 26:189–206, 1984.
- [JW11] T. S. Jayram and David P. Woodruff. Optimal bounds for Johnson-Lindenstrauss transforms and streaming problems with low error. In *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, to appear, 2011.
- [KRS11] Zohar Karnin, Yuval Rabani, and Amir Shpilka. Explicit dimension reduction and its applications. In *Proceedings of the 23rd Annual IEEE Conference on Computational Complexity (CCC)*, to appear, 2011.

- [LT91] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces: isoperimetry and processes*. Springer, 1991.
- [Mat08] Jiri Matousek. On variants of the Johnson-Lindenstrauss lemma. *Random Struct. Algorithms*, 33(2):142–156, 2008.
- [MZ10] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. In *STOC*, pages 427–436, 2010.
- [Siv02] D. Sivakumar. Algorithmic derandomization via complexity theory. In *STOC*, pages 619–626, 2002.
- [TZ04] Mikkel Thorup and Yin Zhang. Tabulation based 4-universal hashing with applications to second moment estimation. In *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 615–624, 2004.
- [Zuc97] David Zuckerman. Randomness-optimal oblivious sampling. *Random Struct. Algorithms*, 11(4):345–367, 1997.

Appendix

A Missing Proofs

Proof. [Proof of Fact 11]

$$\mathbf{E}[Z^\ell] = \int_0^\infty z^\ell \varphi(z) dz = \left[z^\ell \cdot (1 - \Phi(z)) \right]_0^\infty + \ell \cdot \int_0^\infty z^{\ell-1} \cdot (1 - \Phi(z)) dz = \ell \cdot \int_0^\infty z^{\ell-1} \cdot \mathbf{Pr}[Z \geq z] dz,$$

where the second equality holds as long as $z^\ell \cdot (1 - \Phi(z)) \rightarrow 0$ as $z \rightarrow \infty$, which holds since $1 - \Phi(z) = \mathbf{Pr}[Z \geq z] = O(\mathbf{Pr}[Y \geq z])$, and Y has bounded ℓ th moment. The first equality is via integration by parts. Here, φ is the pdf for Z , and Φ is the cdf. The above argument also gives

$$\mathbf{E}[Y^\ell] = \ell \cdot \int_0^\infty z^{\ell-1} \cdot \mathbf{Pr}[Y \geq z] dz,$$

from which the claim follows. ■

B Another approach to Fast Iterative JL reduction

Here we discuss how to plug the Fast Johnson-Lindenstrauss Transform (FJLT) of Ailon and Chazelle [AC09] into the template of Figure 1 to obtain the same seed length as in Theorem 3, but with a JL distribution that in the end has fast embedding time. That is, rather than use samplers as in Section 5, we purely use bounded independence.

We first describe a (slightly modified) version of the construction of [AC09]. $D \in \mathbb{R}^{d \times d}$ represents a diagonal matrix with random signs on the diagonal (label the i th random sign $\sigma(i)$), and H is the $d \times d$ Hadamard matrix, i.e. $H_{i,j} = (-1)^{\langle i,j \rangle} / \sqrt{d}$, where we interpret $i, j \in \{0, \dots, d-1\}$ are interpreted as length- $(\log d)$ binary vectors. Q is a $s \times d$ sampling matrix (with replacement). That is, each row of Q is e_i for some random i . Our final embedding matrix is $S = \sqrt{d/s} \cdot QHD$.

The following follows from Lemma 15.

Lemma 21.

$$\Pr \left[\|HDw\|_\infty > \sqrt{\frac{\log(d/\delta)}{d}} \right] < \delta$$

Proof. Apply [Lemma 15](#) with $\alpha = \log(\sqrt{\log(d/\delta)}/2)(\log d)$ and $k = \log(d/\delta)$. ■

Theorem 22 (Chernoff Bound). *Let X_1, \dots, X_n be independent with means μ_i , variances σ_i^2 , and such that $|X_i| < K$ almost surely. Let $X = \sum_i X_i$, $\mu = \sum_i \mu_i$, $\sigma^2 = \sum_i \sigma_i^2$. Then,*

$$\Pr [|X - \mu| > \lambda] < e^{-\Omega(\min\{\lambda^2/\sigma^2, \lambda/K\})}.$$

Let $y = HDw$. Now, let X_i be the square of the i th sample of HDw taken by Q , scaled by d/s . Then $\mu_i = 1/s$, $\sigma_i^2 \leq (d \cdot \|y\|_\infty^2)/s^2$, $K = (d \cdot \|y\|_\infty^2)/s$. The ℓ_2^2 of our embedded vector is then $\sum_i X_i$, and we want to bound the probability that it deviates by $\mu = 1$ by more than ε . By the Chernoff bound, this probability is at most

$$e^{-\Omega(\min\{\varepsilon^2 s/(d \cdot \|y\|_\infty^2), \varepsilon s/(d \cdot \|y\|_\infty^2)\})} = e^{-\Omega(\varepsilon^2 s/(d \cdot \|y\|_\infty^2))}.$$

But by [Lemma 21](#), we know $\|y\|_\infty^2$ is larger than $\log(d) + \varepsilon\sqrt{s}$ with probability less than $e^{-\Omega(\varepsilon\sqrt{s})}$. Thus, conditioning on this event, we have that

$$\Pr_S[| \|Sw\|^2 - 1 | > \varepsilon] < e^{-\Omega(\varepsilon\sqrt{s})} + e^{-\Omega((\varepsilon^2 s)/(\varepsilon\sqrt{s} + \log d))} = O(e^{-\Omega(\varepsilon\sqrt{s})} + e^{-\Omega(\varepsilon^2(s/\log d))}). \quad (\text{B.1})$$

The above equation now implies the following theorem.

Theorem 23. *The modified FJLT given in this section satisfies the $(\varepsilon, O(\sqrt{(\ell \log d)/(\varepsilon^2 s)})^\ell + O(\ell/\sqrt{\varepsilon^2 s})^\ell, \ell, d, s)$ -JL moment property for any $\ell \geq 2$.*

Proof. The tail bound in [Eq. \(B.1\)](#) is bounded by the max of a Gaussian with variance $\Theta((\log d)/s)$ and an exponential random variable with variance $\Theta(1/s)$. Thus, by [Fact 11](#), the ℓ th moment of the error term is bounded by the max of the absolute value of such a Gaussian and exponential random variable, up to a $2^{O(\ell)}$ factor. ■

[Theorem 23](#) allows us to plug the modified FJLT into the template of [Figure 1](#) to achieve a derandomization. We pick S_0 to be a random Bernoulli matrix with $s_0 = C\varepsilon'^{-2}\delta'^{-1}$. In fact, it is known that with this choice of s_0 , one can even take a matrix which has only one-nonzero entry per column, chosen by a 4-wise independent hash function, with a random sign in that location, again chosen from a 4-wise independent family [[CCFC04](#), [TZ04](#)]. Then for $i > 0$, we set $s_i = C\varepsilon'^{-2}\delta'^{1/2^i} \log(s_{i-1})$ and again set $m = O(\log \log(1/\delta'))$ so that $s_m = C\varepsilon'^{-2} \log^2(1/\delta') \log(s_{m-1})$. The matrix S_{final} is then chosen to have $2 \log(1/\delta)$ -wise independent Bernoulli entries, scaled by $1/\sqrt{s}$. Thus, we have the following theorem.

Theorem 24. *For any $d > 0$ and $0 < \varepsilon, \delta < 1/2$, there is a $(d, O(\varepsilon^{-2} \log(1/\delta)), \delta, \varepsilon)$ -JL distribution which can be sampled using $O(\log d + \log(1/\varepsilon) \log(1/\delta) + \log(1/\delta) \log \log(1/\delta))$ uniform random bits, and where any vector can be embedded in $\tilde{O}(d + \varepsilon^{-4} \log^3(1/\delta))$ time.*

Proof. The matrix S_0 can be applied in $O(d)$ time, since each column has only one non-zero entry. Each S_i can be applied in $O(s_i \log s_i)$ time via the fast Hadamard transform. Then, S_{final} can be applied in $O(s_m \cdot s) = \tilde{O}(\varepsilon^{-4} \log^3(1/\delta))$ time. ■