

Fast and Compact Random Mappings with Uniform Guarantees and Applications

Ying Feng

Piotr Indyk

Abstract

Random orthonormal or Gaussians maps from \mathbb{R}^n to \mathbb{R}^m are a fundamental tool in geometric functional analysis, design of algorithms and machine learning. For example, it is known that, with high probability, a random mapping \mathcal{F} from \mathbb{R}^n to \mathbb{R}^m yields a $(1 + \varepsilon)$ -distortion embedding from ℓ_2^n to ℓ_1^m , i.e., such that $\|x\|_2 \leq \|\mathcal{F}x\|_1 \leq (1 + \varepsilon)\|x\|_2$ for all $x \in \mathbb{R}^n$, as long as $m = \Omega(n/\varepsilon^2)$ [FLM77, Gor85]. However, the algorithmic applications of such mappings have been stymied by the $\Theta(nm)$ time needed to evaluate $\mathcal{F}x$ for a given x . Several alternative constructions of randomized mappings were proposed, with runtimes near-linear in n , but at the price of increasing the dimension m by poly-logarithmic factors.

In this paper we give a new construction of randomized mappings that, in several settings, yields the best-known dimension bound of $m = \mathcal{O}(n/\varepsilon^2)$, while maintaining a near-linear mapping time. Our result applies to the general uniform approximations framework of Cherapanamjeri-Nelson'22. As a result, we obtain improved dimension bounds for applications such as ℓ_2 to ℓ_1 embeddings, adaptive distance estimation data structures, and more.

1 Introduction

Random linear mappings from \mathbb{R}^n to \mathbb{R}^m are a ubiquitous tool in geometric functional analysis, machine learning and design of algorithms for high-dimensional data. They are commonly used to reduce the dimensionality of the data while approximately preserving the distances between the points [JL⁺84, Ach03, DG03, IN07] or for probabilistic constructions of geometric structures. The latter class of applications dates back to Milman's proof of Dvoretzky's theorem [Dvo64, Mil71], which states that every normed space contains a near-Euclidean subspace with non-trivial dimension. Since then, random linear mappings have found many diverse applications, to constructing low-distortion embeddings between ℓ_2 and ℓ_1 norms [FLM77, Gor85, Pis89], compressed sensing matrices [CT06, CRT06b, Don06, FR13] or distance estimation data structures [CN20].

The early applications of random linear mappings used random orthonormal or fully random $n \times m$ matrices \mathcal{F} . Unfortunately, for such distributions, computing a matrix-vector product $\mathcal{F}x$ takes $\Theta(mn)$ time, which is inefficient in many applications. To alleviate this issue, several alternative matrix constructions were proposed, with faster matrix-vector multiplication procedures. A popular example involves partial Fourier matrices, where the rows of the matrix \mathcal{F} consist of randomly selected rows of Fourier or Hadamard matrices, often multiplied by a random variable from some distribution [AC09, CRT06a, Sar06, RV08, KW11, AL13, CN22, DMS24]. Such structured matrices exhibit near-linear-time matrix-vector multiplication procedure thanks to Fast Fourier or Fast Hadamard Transforms. However, proving the desired properties of such matrices becomes more complex, as such the entries in such matrices are no longer independent. Recently [CN22] has greatly simplified this process by presenting a general uniform concentration inequality for such matrices, with applications to distance estimation data structures, random Fourier features and low-distortion embeddings (the latter application was observed in [CSWZ23]). Specifically, they show:

Theorem 1.1 ([CN22]). *Fix a Lipschitz function $f : \mathbb{R} \rightarrow \mathbb{R}$. For any $n \in \mathbb{Z}_{>0}$, and $\varepsilon \in (0, 1/2)$, there exists $m = \mathcal{O}(n \log^5(n/\varepsilon)/\varepsilon^2)$ such that with constant probability, we can sample a linear map $\mathcal{F} \in \mathbb{R}^{m \times n}$ with the following uniform guarantee:*

For all $x \in \mathcal{B}(0, 1)$:

$$\left| \frac{1}{m} \sum_{i=1}^m f((\mathcal{F}x)_i) - \mathbb{E}_{g \sim \mathcal{N}(0, \|x\|_2^2)} [f(g)] \right| \leq \mathcal{O}(\varepsilon).$$

Moreover, $\mathcal{F}x$ can be computed in $\mathcal{O}(n \log n \log^5(n/\varepsilon)/\varepsilon^2)$ time.

However, the significant reduction in the running time is tempered by the increase in the dimension m . Specifically, for a linear mapping with independent random entries, a better dimension bound of $m = \mathcal{O}(n/\varepsilon^2)$ is known, at the price of increasing the running time to $\mathcal{O}(mn) = \mathcal{O}(n^2/\varepsilon^2)$. This raises the question of whether a “best-of-both-worlds” results is possible, matching the best-known dimension bound while preserving near-linear multiplication time.

Our results. In this paper we answer this question in *affirmative*. Specifically, we present a distribution over mappings \mathcal{F} which guarantees the dimension bound of $m = \mathcal{O}(n/\varepsilon^2)$, while being supported by a matrix-vector multiplication procedure with a running time of $\mathcal{O}(n \log \frac{n}{\varepsilon}/\varepsilon + npoly(\log \log n/\varepsilon))$. The formal statement of the result is analogous to Theorem 1.1 and stated in Section 4. We note that for constant ε , the dimension bound of $\mathcal{O}(n)$ is the best possible [BLM89], while the running time of $\mathcal{O}(n \log n)$ matches the best-known bound for the Fourier/Hadamard transform, a ubiquitous tool for generating fast (pseudo)-random maps.

We illustrate the implications of our result in the context of low-distortion embeddings. A mapping $\mathcal{F} : \ell_2^n \rightarrow \ell_1^m$ is said to have $(1 + \varepsilon)$ -*distortion* if for any $x \in \ell_2^n$, we have that

$$\|x\|_2 \leq \|\mathcal{F}x\|_1 \leq (1 + \varepsilon)\|x\|_2.$$

The following table depicts the best known running time and dimension bounds for such mappings.

Dimension m	Time	References
n/ε^2	n^2/ε^2	[FLM77, Gor85]
$n2^{\mathcal{O}(\log \log n)^2}$	$n2^{\mathcal{O}(\log \log n)^2}$	[Ind07]
$n \log^5(n/\varepsilon)/\varepsilon^2$	$n \log n \log^5(n/\varepsilon)/\varepsilon^2$	[CN22, CSWZ23]
n/ε^2	$n \log \frac{n}{\varepsilon}/\varepsilon + npoly(\log \log n/\varepsilon)$	This paper

Table 1: Past work and our result for low-distortion embeddings of ℓ_2^n into ℓ_1^m . For ease of exposition, we drop $\mathcal{O}(\cdot)$ and assume that $\varepsilon > 1/\log n$.

It can be seen that our construction yields the best-to-date dimension bound of $\mathcal{O}(n/\varepsilon^2)$, while being supported by a near-linear-time matrix-vector multiplication procedure.

Applications. Low-distortion embeddings from ℓ_2^n to ℓ_1^m are useful tools in algorithm design, as they reduce the instance of the problem from ℓ_2 to ℓ_1 , where algorithms are often easier to design. For example, our embedding can be combined with the $\mathcal{O}(Nn(\log \log N + \log 1/\varepsilon)/\varepsilon^2)$ -time algorithm for computing the Chamfer distance between two sets of N points in ℓ_1^n [FI25]. This yields an algorithm with a running time of

$$\mathcal{O}(Nn(\log \log(N)/\varepsilon^2 + \log \frac{n}{\varepsilon}/\varepsilon + poly(\log \log n/\varepsilon)))$$

For comparison, the best prior algorithm for Chamfer distance over ℓ_2 had a running time of $\mathcal{O}(Nn \log(N)/\varepsilon^2)$ [BIJ⁺23].

In addition, our result also improves guarantees for two applications studied in [CN22]: efficient kernel approximation and an adaptive data structure for distance estimation. In both cases, our map yields more compact representations (vectors with a smaller dimension) than [CN22], and the representations can be computed quickly, as guaranteed by Theorem 4.1. See Sections 5.2 and 5.3 for details of these applications.

1.1 Our techniques

We explain our techniques using $\ell_2 \rightarrow \ell_1$ embeddings as an example. Intuitively, the goal of a random mapping is to “flatten” the input vector, so that the deviation of its entries from their expectation is constant on average. This ensures that a vector $\mathcal{F}x$ whose ℓ_2 norm is constant has $\Omega(\sqrt{m})$ norm with respect to ℓ_1 . Combining this with the universal upper bound $\|\mathcal{F}x\|_1 \leq \sqrt{m} \|\mathcal{F}x\|_2$ ensures norm preservation up to a constant factor, after a universal scaling of \sqrt{m} .

One way to formalize this flattening is via the so-called *uncertainty principle*, which states that for every vector $x \in \mathbb{R}^n$, the top m/c entries of $\mathcal{F}x$ contain at most a $1/C$ fraction of the total mass $\|\mathcal{F}x\|_2^2$, where $c > 0$ and $C > 0$ are constants. Such a guarantee is easy to achieve using fully random matrices \mathcal{F} with independent Gaussian entries, since in that case each entry of $\mathcal{F}x$ is an independent Gaussian random variable. However, this task becomes much more difficult for partial Hadamard matrices, because then the entries of $\mathcal{F}x$ are no longer independent. The insight of [CN22] is that one can recover the flattening properties of fully random matrices by increasing the dimension m by a $\text{poly}(\log n)$ factor.

In our first step, we observe that instead of increasing the *dimension* by $\text{poly}(\log n)$, we can weaken the flattening property by a similar factor. That is, we can use a mapping L derived from a partial Fourier matrix and retain $m = O(n)$, while guaranteeing that the uncertainty principle holds with $c = \text{poly}(\log n)$ and $C = O(1)$. This property follows from the Restricted Isometry Property of random partial Fourier matrices, which we do not need to reprove (see Section 3 for details).

Then, in order to obtain flattening with the target parameters $c, C = O(1)$, we use the approach of [Ind07]. Specifically, we replicate each entry of Lx d times and permute each copy using a pseudorandom permutation derived from extractor graphs (with left degree $d = O(\log \log n)$). We then group the coordinates mapped together by these permutations into “blocks,” each containing $\text{poly}(\log n)$ coordinates. This step guarantees that the resulting vectors are flattened at the granularity of blocks, though the distribution of coordinate magnitudes within each block can still be uneven. Note that in this paper we use highly efficient *probabilistic* constructions of extractor graphs, since our goal is a probabilistic (as opposed to explicit, as in [Ind07]) construction of efficient mappings. In particular, this means that we only need to replicate each coordinate $O(\log \log n)$ times.

The above construction is then repeated within each block of coordinates, to reduce the block size to $\text{poly}(\log \log n)$. Finally, within each block, we apply standard random Gaussian mappings to reduce the dimension and flatten the vector inside the block. Note that this step is necessary because the two steps above replicate coordinates, increasing the overall dimension by a $\text{poly}(\log \log n)$ factor.

Crucially, the random bits defining the random mappings in each block are independent. This makes it possible to ensure that each vector x is flattened with very high probability, which in turn makes it possible to guarantee that this holds uniformly over all vectors x . To achieve the best dimension bound, we use generic chaining to transform the probabilistic guarantee into the uniform one.

The running time of the whole procedure is dominated by the first application of the Fast Fourier Transform to evaluate the mapping L . The other steps incur only a $\text{poly}(\log \log n)$ overhead per coordinate, for a total running time of $n \cdot \text{poly}(\log \log n)$.

Future directions. Although our construction is not particularly complex (it uses two rounds of Hadamard transforms and random rearrangements, followed by Gaussian projections), it is interesting whether our dimension bound of $O(n/\epsilon^2)$ can be obtained using more straightforward mappings. In this direction, a recent paper [DMS24] showed that a “randomized” circulant matrix mapping yields (with a high probability), a $(1 + \epsilon)$ -distortion mapping from a fixed finite set of points $T \subset \ell_2^n$ into ℓ_1^m , as long as the dimension m is $\Omega(\log(|T|)/\epsilon^2)$. Unfortunately, their theorem only holds assuming $|T| = \exp(O(n/\log^6))$, which precludes extending it to the whole space ℓ_2^n (where the ϵ -net has size $\exp(n)$). Nevertheless, we hope that a more careful analysis of similar mappings could yield a better dimension bound.

2 Preliminaries

Notations. For any integer $n \geq 1$, we use $[n]$ to denote the set of all integers from 1 to n . For a matrix M , we use $\|M\|$ to denote its Euclidean operator norm $\|M\| = \sup_{\|x\|_2=1} \|Mx\|_2$. For an n -by- n matrix M and $S, T \subseteq [n]$, we use $M_{S:}$ to denote the submatrix containing rows of M indexed by set S , and use $M_{:T}$ to denote the submatrix containing columns of M indexed by T . We use $M_{S:T}$ to denote the submatrix whose entries are $M_{i,j} : i \in S, j \in T$ in their original order. For a length- n vector x and $S \subseteq [n]$, we use x_S to denote the segment of this vector indexed by S . For any integer $n > 0$, we use \mathcal{S}^{n-1} to denote the n -dimensional unit Euclidean sphere, i.e. $\mathcal{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$. And we use $\mathcal{B}^n(a, r)$ to denote the n -dimensional Euclidean ball centered at some point $a \in \mathbb{R}^n$ with radius $r \in \mathbb{R}$, i.e. $\mathcal{B}^n(a, r) := \{x \in \mathbb{R}^n : \|x - a\|_2 \leq r\}$. We omit the superscript n when the dimension is clear in the context. For a set $\mathcal{W} \subseteq \mathbb{R}^n$, we use $\text{Diam}(\mathcal{W})$ to denote its Euclidean diameter, i.e. $\sup_{x,y \in \mathcal{W}} \|x - y\|_2$.

In this section, we state some definitions and facts that will appear in our construction.

Definition 2.1 (Hadamard matrix). *For any $l \in \mathbb{N}$ and $N = 2^l$, the order N Hadamard matrix is defined recursively as follows:*

$$H_N := \begin{bmatrix} H_{N/2} & H_{N/2} \\ H_{N/2} & -H_{N/2} \end{bmatrix} \quad \text{with } H_1 := [1].$$

We say $\sqrt{\frac{1}{N}}H_N$ is the normalized Hadamard matrix of order N .

Fact 2.2. *The normalized Hadamard matrix $\sqrt{\frac{1}{N}}H_N$ has orthonormal rows, and $\sqrt{\frac{1}{N}}H_N = \sqrt{\frac{1}{N}}H_N^\top$.*

Fact 2.3. For any $x \in \mathbb{R}^N$, Hx can be computed in time $\mathcal{O}(N \log N)$.

This can be done using the fast Walsh-Hadamard transform, due to the recursive structure of the Hadamard matrix.

We also need the definition of the Restricted Isometry Property (RIP). For our purpose, it will be more convenient to define RIP as a bound on the spectrum of submatrices.

Definition 2.4 (Restricted Isometry Property (RIP)). Given a matrix $U \in \mathbb{R}^{n \times N}$ and parameters $k \leq N$ and $\delta \in (0, 1)$, we say U satisfies (k, δ) -RIP if the following holds:

$$\sup_{|S| \leq k} \|I - (U_{:S})^\top U_{:S}\| \leq \delta,$$

where the supremum is taken over all subsets $S \subseteq [N]$ with at most k elements, $U_{:S}$ denotes the column submatrix of U indexed by S , and I is the identity matrix in $\mathbb{R}^{|S|}$.

Lastly, we state some lemmas that will be useful for our proof.

Lemma 2.5 ([Ind07]). For any $x \in \mathbb{R}^n$ and any partitioning P_1, \dots, P_b of $[n]$, we have

$$\sum_{j=1}^b \|x_{P_j}\|_2 \leq \sqrt{b} \|x\|_2.$$

Definition 2.6. Given a random variable Z , the subgaussian norm of Z , denoted as $\|Z\|_{\psi_2}$, is

$$\|Z\|_{\psi_2} = \inf\{u > 0 : \mathbb{E}[\exp(\frac{Z^2}{u^2})] \leq 2\}.$$

Fact 2.7. For $Z \sim \mathcal{N}(0, \sigma^2)$, $\|Z\|_{\psi_2} = \sqrt{\frac{8}{3}}\sigma$.

Lemma 2.8 ([Ver18], Theorem 2.7.8). For any random variable Z with $\|Z\|_{\psi_2} < \infty$, we have that

$$\|Z - \mathbb{E}[Z]\|_{\psi_2} \leq c_0 \|Z\|_{\psi_2}$$

for some absolute constant $c_0 > 0$.

Lemma 2.9 ([Ver18], Theorem 2.7.3). Let Z_1, \dots, Z_l be l independent, mean-zero random variables such that for each $i \in [l]$, $\|Z_i\|_{\psi_2} \leq \sigma_i$. Then we have

$$\mathbb{P}\left[\left|\sum_{i=1}^l Z_i\right| > t\right] \leq 2 \exp\left(-\Theta\left(\frac{t^2}{\sum_{i=1}^l \sigma_i^2}\right)\right).$$

Given a graph \mathcal{G} and a vertex $i \in \mathcal{G}$, we use $\Gamma_{\mathcal{G}}(i)$ to denote the set of neighbors of i in \mathcal{G} . Moreover, given that $\deg(i) = d$, for any $t \in [d]$, we use $\Gamma_{\mathcal{G}}(i)_t$ to denote the t -th neighbor of i , where we assume that the neighbors of each vertex have a fixed, arbitrary order.

Definition 2.10. A bipartite graph $\mathcal{G} = (\mathcal{A}, \mathcal{B}, \mathcal{E})$, $\mathcal{A} = [a]$, $\mathcal{B} = [b]$, with each left node having degree d , is called an (η, l) -extractor, if it satisfies the following property:

Consider any distribution \mathcal{P} over \mathcal{A} such that for any $i \in \mathcal{A}$, $\Pr_{\mathcal{P}}[i] \leq 1/l$. Let j be a random variable over \mathcal{B} obtained by choosing a vertex i from \mathcal{A} , choosing t uniformly at random from $[d]$, and setting $j = \Gamma_{\mathcal{G}}(i)_t$. Let $\mathcal{G}(\mathcal{P})$ be the resulting distribution of the random variable j , and let \mathcal{I} be the uniform distribution over \mathcal{B} . We require that $\mathcal{G}(\mathcal{P})$ and \mathcal{I} are η -close, i.e., $\|\mathcal{G}(\mathcal{P}) - \mathcal{I}\|_1 \leq \eta$, where both distributions are interpreted as vectors in \mathbb{R}^b .

Lemma 2.11 ([Zuc97]). For any $a, b, l, \eta > 0$, there exists an (η, l) -extractor with $|\mathcal{A}| = a$, $|\mathcal{B}| = n$, and degree $d = \mathcal{O}((\frac{b}{l} + \log \frac{a}{l})/\eta^2)$.

Lemma 2.11 is obtained by choosing the neighbor lists completely at random, and the failure probability of the construction is exponentially small in l . The average right degree of this extractor is $\mathcal{O}(ad/b)$. As argued in Claim 3.2 of [Ind07], we can convert this into an extractor with maximum right degree $\mathcal{O}(ad/b)$. All parameters a, b, η, l, d in the resulting extractor are asymptotically unchanged.

3 Uncertainty Principle

Given any $n \in \mathbb{Z}_{>0}$, the goal of this section is to construct a matrix that satisfies the following guarantee:

Definition 3.1. We say a matrix $L \in \mathbb{R}^{N \times n}$ satisfies the uncertainty principle with parameters $c, C > 1$ if for any vector $x \in \mathbb{R}^n$ and any subset of coordinates $S \subseteq [N]$ with $|S| \leq N/c$,

$$\|(Lx)_S\|_2^2 \leq \|Lx\|_2^2/C = \|x\|_2^2/C \quad (1)$$

We observe that this can be viewed as a dual property of RIP: for a matrix having orthonormal rows and satisfying RIP after appropriate scaling, its transpose satisfies Equation 1.

Lemma 3.2. Let $M \in \mathbb{R}^{n \times N}$ be a matrix with orthonormal rows. If $\sqrt{\frac{N}{n}}M$ satisfies (k, δ) -RIP, then for any vector $x \in \mathbb{R}^n$ and any subset of coordinates $S \subseteq [N]$ with $|S| \leq k$,

$$\|(M^\top x)_S\|_2^2 \leq \|M^\top x\|_2^2/C = \|x\|_2^2/C$$

where $C = N/(1 + \delta)n$.

Proof. For every $|S| \leq k$, all eigenvalues of $\frac{N}{n}(M_{:S})^\top M_{:S}$ lie in $[1 - \delta, 1 + \delta]$, thus the operator norms of $M_{:S}$ and its transpose $(M_{:S})^\top$ are at most $\sqrt{(1 + \delta)n/N}$. We observe that for any $x \in \mathbb{R}^n$, $\|(M^\top x)_S\|_2 = \|(M_{:S})^\top x\|_2$, which is at most $\sqrt{(1 + \delta)n/N}\|x\|_2$. Moreover, M^\top has orthonormal columns, thus $\|M^\top x\|_2 = \|x\|_2$. Combining these together and then taking the squares conclude the proof. \square

We remark that a similar relationship has been observed in [LV06]: roughly, they show that for M satisfying the premise of Lemma 3.2, if the *input* vector x is k -sparse, then $\|Mx\|_2^2 \leq \|x\|_2^2/C$. Since we care about restricting the *output* vector to a k -sparse subset of coordinates, we instead work with the transpose matrix M^\top .

To instantiate Lemma 3.2, we start with a random matrix M' sampled as follows: Let H be a normalized Hadamard matrix of order N , and let Y_1, Y_2, \dots, Y_N be independent Bernoulli random variables, which take value 1 each with probability n/N . These random variables define a subset $\Omega = \{i \in [N] : Y_i = 1\}$, and we take $M' := H_{\Omega:}$ to be the row submatrix indexed by Ω . Such M' is called a random partial Fourier matrix, and it has been shown to satisfy RIP with high probability.

Lemma 3.3 ([RV08]). *There exists an absolute constant $r > 0$ such that given parameters $N \in \mathbb{Z}_{>0}$ and $k \leq N$, if $n \geq r \cdot (k \cdot \log^2 N \log^3 k)$, then once we sample a random subset Ω via $Y_1, \dots, Y_N \sim \text{Bernoulli}(n/N)$ as described above and let $M' := H_{\Omega:}$,*

$$\mathbb{P}\left[\sqrt{\frac{N}{|\Omega|}} M' \text{ satisfies } (k, 1)\text{-RIP}\right] \geq 1 - n^{-10}.$$

Combining Lemma 3.2 and 3.3, we can see that $(M')^\top$ is close to what we want for L in Equation 1, except that $(M')^\top$ have n columns *only in expectation*, while we need L to have exactly n columns. To fix this, we instead sample a random submatrix M of H with a fixed number of rows n . Using Lemma 3.2 and the fact that $\mathbb{E}[|\Omega|] = n$, we can see that M gives the similar guarantee as M' . The details are given in Appendix A.

Combining this with Lemma 3.2, we obtain the following construction for L .

Corollary 3.4. *Given $n \in \mathbb{Z}_{>0}$ and $C > 1$, there exists some $N = \Theta(Cn)$ such that with probability at least $1 - n^{-9}$, we can sample a matrix $L \in \mathbb{R}^{N \times n}$ that satisfies the follows: for any vector $x \in \mathbb{R}^n$ and any subset of coordinates $S \subseteq [N]$ with $|S| \leq \Omega(n/\log^5(Cn))$,*

$$\|(Lx)_S\|_2^2 \leq \|Lx\|_2^2/C = \|x\|_2^2/C.$$

Proof. We fix n and $N = 2Cn$. Then there exists $k = \Omega(n/\log^5(Cn))$ such that the set of parameters N, k, n satisfies the premise of Lemma A.1. Thus $\sqrt{N/n}M \in \mathbb{R}^{n \times N}$ satisfies $(k, 1)$ -RIP with probability $1 - n^{-9}$. We condition on the success case and take $L := M^\top$; it follows immediately from Lemma 3.2 that $\|(Lx)_S\|_2^2 \leq \|Lx\|_2^2/C = \|x\|_2^2/C$ since $\frac{N}{(1+\delta)n} = \frac{2Cn}{(1+1)n} = C$. \square

Finally, we claim that Lx can be computed efficiently.

Lemma 3.5. *For any $x \in \mathbb{R}^n$, Lx can be computed in $\mathcal{O}(Cn \log(Cn))$ time.*

Proof. For any $x \in \mathbb{R}^n$, we define $x' \in \mathbb{R}^N$ to be x at coordinates in Ξ and have zeros anywhere else. Then we can observe that $Lx = M^\top x = (H^\top)_{:\Xi} x = Hx'$, where H is the $N \times N$ normalized Hadamard matrix such that Hx' can be computed in $\mathcal{O}(N \log N) = \mathcal{O}(Cn \log(Cn))$ time. \square

4 Concentration of Averages

Given an error parameter $\varepsilon > 0$, our goal is to construct a linear map $\mathcal{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ for some $m = \mathcal{O}(n/\varepsilon^2)$ such that (1) the mapping $\mathcal{F}(x)$ can be computed efficiently, and (2) for any Lipschitz function f , the average of f applied on every coordinate of $\mathcal{F}(x)$ is ε -close to the expectation $\mathbb{E}_{z \sim \mathcal{N}(0,1)}[f(z)]$. Formally, we will prove the following theorem:

Theorem 4.1. *Fix a Lipschitz function $f : \mathbb{R} \rightarrow \mathbb{R}$. For any $n \in \mathbb{Z}_{>0}$ and $\varepsilon \in (0, 1/2)$, there exists $m = \mathcal{O}(n/\varepsilon^2)$ such that with constant probability, we can sample a linear map $\mathcal{F} \in \mathbb{R}^{m \times n}$ with the following uniform guarantee:*

For all $x \in \mathcal{B}(0, 1)$:

$$\left| \frac{1}{m} \sum_{i=1}^m f((\mathcal{F}x)_i) - \mathbb{E}_{g \sim \mathcal{N}(0, \|x\|_2^2)}[f(g)] \right| \leq \mathcal{O}(\varepsilon).$$

Moreover, $\mathcal{F}x$ can be computed in $\mathcal{O}(n \log \frac{n}{\varepsilon}/\varepsilon + npoly((\log \log n)/\varepsilon))$ time.

Next, we describe our construction.

4.1 Construction.

Let $n \in \mathbb{Z}_{>0}$ and $\varepsilon \in (0, 1/2)$ be arbitrary. And let $C := 1/\varepsilon$. We first sample a matrix $L \in \mathbb{R}^{a \times n}$ as described in Corollary 3.4, where $a = \Theta(Cn)$. Then we sample an (η, l) -extractor \mathcal{G} with parameters:

- $a = \Theta(Cn)$,
- $b = \Theta(n/\log^5(Cn))$,
- $\eta = 1/C$,
- $l = (1 - 1/C)b = \Theta((1 - 1/C)n/\log^5(Cn))$,
- $d = \Theta((\frac{b}{l} + \log \frac{a}{l})/\eta^2) = \Theta(C^2(\log \log n + \log C))$

using Lemma 2.11. Here, a is chosen to be the dimension of Lx . The parameter b is the maximum size of S such that $\|(Lx)_S\|_2^2 \leq \|Lx\|_2^2/C$ in Corollary 3.4. The parameters l, d are chosen depending on a, b, η .

Given $x \in \mathbb{R}^n$, we first compute $Lx \in \mathbb{R}^a$. Then in the second step, we view Lx as the vertex set \mathcal{A} in the extractor graph \mathcal{G} . For each $j \in [b]$, we define a block vector as the coordinates of Lx indexed by the neighbor list $\Gamma_{\mathcal{G}}(j)$. This outputs a vector

$$y := \bigoplus_{j=1}^b (Lx)_{\Gamma_{\mathcal{G}}(j)} \in \mathbb{R}^{ad},$$

where \oplus is the direct sum operator, i.e., concatenation. The dimension is ad because by the definition of the extractor, $\bigoplus_{j=1}^b (Lx)_{\Gamma_{\mathcal{G}}(j)}$ is a permutation of Lx repeated for d times. If we

denote $y_j := (Lx)_{\Gamma_{\mathcal{G}}(j)}$ for each $j \in [b]$, then $y = \oplus_{j=1}^b y_j$, and the maximum dimension of y_j is $\mathcal{O}(ad/b) = \mathcal{O}(C^3 \log^5(Cn) \cdot (\log \log n + \log C))$. To reduce this to $\text{poly}(C) \cdot o(\log n)$, we repeat the above procedure on every y_j .

We let $n' = \mathcal{O}(ad/b)$ denote the maximum dimension of y_1, \dots, y_b , and then pad every y_j to be n' -dimensional by adding zeros at the end. After that, we sample a map $L' \in \mathbb{R}^{\Theta(Cn') \times n'}$ from Corollary 3.4 and an extractor \mathcal{G}' from Lemma 2.11 with all parameters (a', b', η, l', d') depending on n' instead of n . Using L', \mathcal{G}' , we compute and output

$$z := \oplus_{j=1}^b \oplus_{j'=1}^{b'} (L'y_j)_{\Gamma_{\mathcal{G}'}(j')} \in \mathbb{R}^{b \cdot (a'd')}.$$

We denote $z_{j,j'} := (L'y_j)_{\Gamma_{\mathcal{G}'}(j')}$. Each $z_{j,j'}$ has a maximum dimension $\mathcal{O}(a'd'/b')$. Finally, we sample bb' independent Gaussian random matrices $G_{1,1}, \dots, G_{b,b'}$ such that for each $j \in [b]$ and $j' \in [b']$,

- $G_{j,j'} \in \mathbb{R}^{w \times |\Gamma_{\mathcal{G}'}(j')|}$ for $w = \Theta(C^2 n / bb')$, and
- entries of G_j are i.i.d. standard Gaussian random variables scaled by $1/\sqrt{w}$.

The final map \mathcal{F} is defined as follows: for any $x \in \mathbb{R}^n$,

$$\mathcal{F}x := \oplus_{j=1}^b \oplus_{j'=1}^{b'} \sqrt{\frac{bb'}{dd'}} G_{j,j'} z_{j,j'} \in \mathbb{R}^{\Theta(C^2 n)}.$$

This concludes our description of \mathcal{F} . We remark that \mathcal{F} is linear because all block-wise operations can be implemented as multiplying with a block diagonal matrix. Before proving the concentration of averages (Theorem 4.1), we make some observations of \mathcal{F} .

Claim 4.2. *With at least constant probability, $L, \mathcal{G}, L', \mathcal{G}'$ are sampled successfully.*

This is because the failure probability is dominated by that of sampling L' , which is at most $1/\text{poly}(\log n)$. In the successful case of Claim 4.2, our map \mathcal{F} is successfully initialized. We will condition on this case in our analysis.

Lemma 4.3. *Given $x \in \mathbb{R}^n$, $\mathcal{F}x$ can be computed in $\mathcal{O}(Cn \log Cn + n \text{poly}(C \log \log n))$ time.*

Proof. L runs in $\mathcal{O}(Cn \log(Cn))$ time. Then the extractor step with respect to \mathcal{G} takes time proportional to the dimension of the output vector y , which is $\mathcal{O}(ad) = \mathcal{O}(C^3 n (\log \log n + \log C))$. After that, we have block size $n' = \mathcal{O}(ad/b) = \mathcal{O}(C^3 \log^5(Cn) \cdot (\log \log n + \log C))$. So L' takes $\mathcal{O}(b \cdot Cn' \log(Cn'))$ time, and \mathcal{G}' takes $\mathcal{O}(b \cdot a'd')$ time. At this point, the block size is reduced to $\mathcal{O}(a'd'/b')$. Finally, applying bb' -many block-wise Gaussians with w rows takes $\mathcal{O}(bb' \cdot (wa'd'/b')) = \mathcal{O}(bwa'd') = \mathcal{O}(C^2 na'd'/b')$ time.

Therefore, the total runtime is

$$\mathcal{O}(\underbrace{Cn \log(Cn)}_L + \underbrace{C^3 n \xi}_{\mathcal{G}} + \underbrace{C^4 n \xi}_{L'} + \underbrace{C^6 n \xi (\log \log \log n + \log C)}_{\mathcal{G}'} + \underbrace{C^5 n \log^5(C \log n) (\log \log \log n + \log C)}_{G_{j,j'}})$$

where $\xi := (\log \log n + \log C)$. This is $\mathcal{O}(Cn \log Cn + n \text{poly}(C, \log \log n))$, with the exponents of C and $\log \log n$ in the second term are at most 6. \square

4.2 Analysis

It remains to show that \mathcal{F} gives the claimed concentration.

Proof (of Theorem 4.1). The proof proceeds in two parts: in the first part, we bound the norms of block vectors, and in the second part, we use such bounds to show our uniform concentration result.

Bounding block vectors. We first focus on the block vectors $y_j := (Lx)_{\Gamma_G(j)}$ and $z_{j,j'} := (L'y_j)_{\Gamma_{G'}(j')}$. We will show that, due to the uncertainty principle (Corollary 3.4) and the property of the extractor (Definition 2.10), we have that

$$\forall x \in \mathbb{R}^n, \sum_{j=1}^b \sum_{j'=1}^{b'} \left| \frac{b}{d} \frac{b'}{d'} \|z_{j,j'}\|_2^2 - \|x\|_2^2 \right| \leq \mathcal{O}(bb' \|x\|_2^2 / C). \quad (2)$$

This holds deterministically once \mathcal{F} is initialized successfully. This means that the blocks vectors $z_{j,j'}$ have similar norms, in the sense that we can bound the total deviation. We prove this step by step in the following Claims 4.4 and 4.5.

Claim 4.4. $\forall x \in \mathbb{R}^n, \sum_{j=1}^b \left| \frac{b}{d} \|y_j\|_2^2 - \|x\|_2^2 \right| \leq \mathcal{O}(b \|x\|_2^2 / C)$.

Proof. Without loss of generality, we assume that $\|x\|_2 = 1$. We let $S \subseteq [a]$ be the set of indices containing b largest (in magnitude) entries of Lx . Then we define the vectors $v := (Lx)_S$ and $\bar{v} := (Lx)_{[a] \setminus S}$, interpreted as vectors in \mathbb{R}^a . This gives a decomposition $Lx = v + \bar{v}$, and for any block index $j \in [b]$,

$$\|y_j\|_2^2 = \|(Lx)_{\Gamma_G(j)}\|_2^2 = \|v_{\Gamma_G(j)}\|_2^2 + \|\bar{v}_{\Gamma_G(j)}\|_2^2.$$

It suffices to show that $\sum_{j=1}^b \frac{b}{d} \|v_{\Gamma_G(j)}\|_2^2 \leq b/C$ and $\sum_{j=1}^b \left| \frac{b}{d} \|\bar{v}_{\Gamma_G(j)}\|_2^2 - 1 \right| \leq \mathcal{O}(b/C)$. To see the former, recall that by Corollary 3.4, we have $\|v\|_2^2 \leq 1/C$. It follows that $\sum_{j=1}^b \|v_{\Gamma_G(j)}\|_2^2 \leq d \cdot 1/C$ because every coordinate of v is repeated for at most d times in sets $\Gamma_G(1), \dots, \Gamma_G(b)$. Thus $\frac{b}{d} \sum_{j=1}^b \|v_{\Gamma_G(j)}\|_2^2 \leq b/C$.

For the latter, we use the idea from [Ind07] to show that $\sum_{j=1}^b \left| \frac{\|\bar{v}_{\Gamma_G(j)}\|_2^2}{d} - 1/b \right| \leq \mathcal{O}(1/C)$. From Corollary 3.4, we know that $\|\bar{v}\|_2^2 \geq 1 - \frac{1}{C}$. At the same time, for each $i = 1, \dots, a$, we have

$$\bar{v}_i^2 \leq \|Lx\|_2^2 / |S| = \|Lx\|_2^2 / b = 1/b$$

because x is a unit vector and L is an isometry.

We use \bar{v} to construct a probability distribution \mathcal{P} over $[a]$, by defining $p_i = \bar{v}_i^2 / \|\bar{v}\|_2^2 \leq \frac{1}{(1-1/C)b}$. It follows that $p_i \leq 1/l$ for $l = (1-1/C)b$. Therefore, \mathcal{P} satisfies the conditions on using the extractor G . This implies that the distribution $\mathcal{Q} := G(\mathcal{P})$ over $\mathcal{B} = [b]$ is η -close to the uniform distribution over \mathcal{B} . For any $j \in [b]$, the probability q_j with respect to \mathcal{Q} is equal to

$$q_j = 1/d \cdot \sum_{i \in \Gamma_G(j)} p_i = \frac{1}{\|\bar{v}\|_2^2 d} \sum_{i \in \Gamma_G(j)} \bar{v}_i^2 = \frac{\|\bar{v}_{\Gamma_G(j)}\|_2^2}{\|\bar{v}\|_2^2 d}.$$

Since \mathcal{Q} is η -close to the uniform distribution, we get

$$\sum_{j=1}^b \left| \frac{\|\bar{v}_{\Gamma_{\mathcal{G}}(j)}\|_2^2}{\|\bar{v}\|_2^2 d} - 1/b \right| \leq \eta = 1/C.$$

Because $\|\bar{v}\|_2^2 \in [1 - 1/C, 1]$,

$$\sum_{j=1}^b \left| \frac{\|\bar{v}_{\Gamma_{\mathcal{G}}(j)}\|_2^2}{d} - 1/b \right| \leq 1/C + \sum_{j=1}^b |(1 - 1/C)/b - 1/b| = \mathcal{O}(1/C)$$

by triangle's inequality. Thus $\sum_{j=1}^b \left| \frac{b}{d} \|\bar{v}_{\Gamma_{\mathcal{G}}(j)}\|_2^2 - 1 \right| \leq \mathcal{O}(b/C)$. □

Claim 4.5. For any $j \in [b]$, $\sum_{j'=1}^{b'} \left| \frac{b'}{d'} \|z_{j,j'}\|_2^2 - \|y_j\|_2^2 \right| \leq \mathcal{O}(b' \|y_j\|_2^2 / C)$.

Proof Sketch. The proof parallels that of Claim 4.4. If we define $T \subseteq [a']$ to contain b' largest (in magnitude) entries of $L'y_j$ and define $u := (L'y_j)_T$ and $\bar{u} := (L'y_j)_{[a'] \setminus T}$, then we can show that $\sum_{j'=1}^{b'} \|u_{\Gamma_{\mathcal{G}'}(j')}\|_2^2 \leq \|y_j\|_2^2 \cdot d'/C$ and $\sum_{j'=1}^{b'} \left| \frac{\|\bar{u}_{\Gamma_{\mathcal{G}'}(j')}\|_2^2}{d'} - \frac{\|y_j\|_2^2}{b'} \right| \leq \|y_j\|_2^2 \cdot \mathcal{O}(1/C)$. Thus

$$\sum_{j'=1}^{b'} \left| \frac{b'}{d'} (\|u_{\Gamma_{\mathcal{G}'}(j')}\|_2^2 + \|\bar{u}_{\Gamma_{\mathcal{G}'}(j')}\|_2^2) - \|y_j\|_2^2 \right| \leq \mathcal{O}(b' \|y_j\|_2^2 / C).$$

□

Combining Claim 4.4, 4.5, we get that

$$\begin{aligned} \sum_{j=1}^b \sum_{j'=1}^{b'} \left| \frac{b}{d} \frac{b'}{d'} \|z_{j,j'}\|_2^2 - \|x\|_2^2 \right| &\leq \sum_{j=1}^b \sum_{j'=1}^{b'} \left| \frac{b}{d} \frac{b'}{d'} \|z_{j,j'}\|_2^2 - \frac{b}{d} \|y_j\|_2^2 \right| + \sum_{j=1}^b \sum_{j'=1}^{b'} \left| \frac{b}{d} \|y_j\|_2^2 - \|x\|_2^2 \right| \\ &\leq \left(\sum_{j=1}^b \frac{b}{d} \cdot \mathcal{O}(b' \|y_j\|_2^2 / C) \right) + \left(b' \cdot \mathcal{O}(b \|x\|_2^2 / C) \right) \\ &\leq \sum_{j=1}^b \|y_j\|_2^2 \cdot \mathcal{O}(bb'/Cd) + \mathcal{O}(bb' \|x\|_2^2 / C) \end{aligned}$$

Here $\sum_{j=1}^b \|y_j\|_2^2 = \sum_{j=1}^b \|(Lx)_{\Gamma_{\mathcal{G}}(j)}\|_2^2 \leq d \|x\|_2^2$ because $\|Lx\|_2^2 = \|x\|_2^2$ and each coordinate of Lx is repeated at most d times in $\Gamma_{\mathcal{G}}(1), \dots, \Gamma_{\mathcal{G}}(b)$. Therefore, $\sum_{j=1}^b \sum_{j'=1}^{b'} \left| \frac{b}{d} \frac{b'}{d'} \|z_{j,j'}\|_2^2 - \|x\|_2^2 \right| \leq \mathcal{O}(bb' \|x\|_2^2 / C)$.

Uniform Concentration of Averages. Recall that $\mathcal{F}x := \bigoplus_{j=1}^b \bigoplus_{j'=1}^{b'} \sqrt{\frac{bb'}{dd'}} G_{j,j'} z_{j,j'}$, where $G_{j,j'} \in \mathbb{R}^{w \times |\Gamma_{\mathcal{G}'}(j')|}$ are random Gaussian matrices. For $j \in [b]$, $j' \in [b']$, $t \in [w]$, we sample

independent standard Gaussian vectors $g_{j,j'}^t \sim \mathcal{N}(0, 1)^{|\Gamma_{G'}(j')|}$, which should be considered as rows of $G_{j,j'}$. For $x \in \mathcal{B}(0, 1)$, we let T_x denote the quantity

$$T_x := \frac{1}{bb'w} \sum_{j=1}^b \sum_{j'=1}^{b'} \sum_{t=1}^w f(\langle g_{j,j'}^t, \sqrt{\frac{bb'}{dd'}} z_{j,j'} \rangle) - \mu$$

where $\mu = \mu(x) := \mathbb{E}_{g \sim \mathcal{N}(0, \|x\|_2^2)}[f(g)]$. Equivalently, for $m = bb'w = \Theta(C^2n)$, this is saying

$$T_x := \frac{1}{m} \sum_{i=1}^m f((\mathcal{F}x)_i) - \mu.$$

In the following, for succinctness, we switch the indexing from $[b] \times [b'] \times [w]$ to $[m]$. We also let $\bar{T}_x := T_x - \mathbb{E}[T_x]$.

To prove Theorem 4.1, we need to show that for all $x \in \mathcal{B}(0, 1)$, $|T_x| \leq \mathcal{O}(1/C)$. Since $|T_x| \leq |\bar{T}_x| + |\mathbb{E}[T_x]|$, we will bound the two terms separately:

- (Claim 4.7) $|\mathbb{E}[T_x]| \leq \mathcal{O}(1/C)$ holds deterministically for all $x \in \mathcal{B}(0, 1)$.
- (Claim 4.15) Using a generic chaining argument, we can bound $|\bar{T}_x| \leq \mathcal{O}(1/C)$ uniformly for all x . This relies on a sub-gaussian tail on $|\bar{T}_x - \bar{T}_{x'}|$ for $x, x' \in \mathcal{B}(0, 1)$.

Before bounding these two terms, we make a crucial observation on the distributions of coordinates of $\mathcal{F}x$.

Claim 4.6. *Let $x \in \mathbb{R}^n$ be arbitrary and fixed. There exists $\delta_1, \delta_2, \dots, \delta_m \in \mathbb{R}$ (depending on x) such that for each $i \in m$, $(\mathcal{F}x)_i$ independently follows the distribution $\mathcal{N}(0, (1 + \delta_i)\|x\|_2^2)$, and $\sum_{i=1}^m |\delta_i| \leq \mathcal{O}(m/C)$.*

Proof. This follows from Equation 2 from the previous part of the proof. We have that

$$\sum_{j=1}^b \sum_{j'=1}^{b'} \sum_{t=1}^w \left| \frac{b}{d} \frac{b'}{d'} \|z_{j,j'}\|_2^2 - \|x\|_2^2 \right| \leq \mathcal{O}(bb'\|x\|_2^2/C) \cdot w = \mathcal{O}(m\|x\|_2^2/C).$$

For indices j, j', t corresponding to i such that $(\mathcal{F}x)_i = \langle g_{j,j'}^t, \sqrt{\frac{bb'}{dd'}} z_{j,j'} \rangle$, $(\mathcal{F}x)_i \sim \mathcal{N}(0, \frac{b}{d} \frac{b'}{d'} \|z_{j,j'}\|_2^2)$, and the independence follows from the fact that $g_{j,j'}^t$'s are independent. \square

Claim 4.7. *For any $x \in \mathcal{B}(0, 1)$, $|\mathbb{E}[T_x]| \leq \mathcal{O}(1/C)$.*

Proof. We observe that for each $i \in [m]$,

$$\begin{aligned} \left| \mathbb{E}[f((\mathcal{F}x)_i)] - \mu \right| &= \left| \mathbb{E}_{g' \sim \mathcal{N}(0, (1 + \delta_i)\|x\|_2^2)}[f(g')] - \mathbb{E}_{g \sim \mathcal{N}(0, \|x\|_2^2)}[f(g)] \right| \\ &\leq \mathbb{E}_{g \sim \mathcal{N}(0, \|x\|_2^2)} [|f(\sqrt{1 + \delta_i} \cdot g) - f(g)|] \\ &\leq \mathcal{O}(\|x\|_2 \mathbb{E}_{g \sim \mathcal{N}(0, 1)} [\sqrt{1 + \delta_i} \cdot g - g]) \\ &\leq \mathcal{O}(|\sqrt{1 + \delta_i} - 1| \cdot \mathbb{E}_{g \sim \mathcal{N}(0, 1)} [|g|]) \\ &= \mathcal{O}(|\delta_i|) \end{aligned}$$

In the first inequality, we couple the two Gaussians as a scaling of each other. The second inequality follows from f being Lipschitz, and the third inequality follows from $\|x\|_2 \leq 1$. Taking the average across all coordinates, we get

$$|\mathbb{E}[T_x]| = \left| \frac{1}{m} \sum_{i=1}^m \mathbb{E}[f((\mathcal{F}x)_i)] - \mu \right| \leq \frac{1}{m} \sum_{i=1}^m \left| \mathbb{E}[f((\mathcal{F}x)_i)] - \mu \right| \leq \frac{1}{m} \sum_{i=1}^m \mathcal{O}(|\delta_i|) \leq \mathcal{O}(1/C).$$

□

To make our chaining argument, we start with the following standard definitions and results.

Definition 4.8. Let (S, d_S) be a metric space. We define

$$\gamma_2(S, d_S) := \inf_{\{S_r\}_r} \sup_{s \in S} \sum_{r=0}^{\infty} 2^{r/2} d_S(s, S_r),$$

where the infimum is taken over admissible sequences, i.e. $S_0 \subset S_1 \subset \dots S$ such that $|S_0| = 1$ and $|S_r| \leq 2^{2r}$.

Lemma 4.9 ([Tal21], Exercise 2.7.4). If $d'_S \leq B \cdot d_S$ for some $B > 0$, then $\gamma_2(S, d'_S) \leq B\gamma_2(S, d_S)$.

Corollary 4.10. If $d'_S = B \cdot d_S$ for some $B > 0$, then $\gamma_2(S, d'_S) = B\gamma_2(S, d_S)$.

Proof. This follows from Claim 4.9 applied on $d'_S \leq B \cdot d_S$ and $d_S \leq \frac{1}{B} \cdot d'_S$. □

Lemma 4.11 (Majorizing Theorem. [Tal21], Theorem 2.10.1). There exists a constant $c_0 \geq 1$ such that for any metric space $(S, \|\cdot\|_2)$,

$$\frac{1}{c_0} \gamma_2(S, \|\cdot\|_2) \leq \mathbb{E}_{v \sim \mathcal{N}(0,1)^n} [\sup_{x \in S} \langle v, x \rangle] \leq c_0 \gamma_2(S, \|\cdot\|_2).$$

Corollary 4.12. $\gamma_2(\mathcal{B}(0,1), \|\cdot\|_2) = \mathcal{O}(\sqrt{n})$.

Proof. By the Majorizing Measure Theorem,

$$\mathbb{E}_{v \sim \mathcal{N}(0,1)^n} [\sup_{x \in \mathcal{B}(0,1)} \langle v, x \rangle] = \mathcal{O}(\gamma_2(\mathcal{B}(0,1), \|\cdot\|_2)).$$

By the Cauchy-Schwarz Inequality, $\sup_{x \in \mathcal{B}(0,1)} \langle v, x \rangle = \|v\|_2$, thus

$$\mathbb{E}_{v \sim \mathcal{N}(0,1)^n} [\sup_{x \in \mathcal{B}(0,1)} \langle v, x \rangle] = \mathbb{E}[\|\mathcal{N}(0,1)^n\|_2] = \sqrt{n}.$$

□

Theorem 4.13 ([Tal21], Theorem 2.7.11). Let (S, d_S) be a metric space, and let $(X_s)_{s \in S}$ be a process, i.e. a collection of variables indexed by S . If $(X_s)_{s \in S}$ satisfies the following:

- (Centered.) $\forall s \in S, \mathbb{E}[X_s] = 0$, and
- (Increment Condition.) $\forall u > 0$ and $\forall s, t \in S, \mathbb{P}[|X_s - X_t| > u] \leq 2 \exp(-u^2/2d_S(s, t)^2)$,

then we have

$$\mathbb{E}[\sup_{s \in S} X_s] \leq c_0 \gamma_2(S, d_S),$$

where $c_0 > 0$ is an absolute constant.

We will use the combination of Theorem 4.13 and existing bounds on γ_2 (Corollary 4.10 and 4.12) as a black box. Before that, we still need to prove that \bar{T}_x satisfies the increment condition.

Claim 4.14. *For any $x, x' \in \mathcal{B}(0, 1)$ and $u > 0$, $\mathbb{P}[|\bar{T}_x - \bar{T}_{x'}| > u] \leq e^{-\Omega(mu^2/\|x-x'\|_2^2)}$.*

Proof. We observe that for each i , the centered random variable

$$(f((\mathcal{F}x)_i) - f((\mathcal{F}x')_i)) - (\mathbb{E}[f((\mathcal{F}x)_i)] - \mathbb{E}[f((\mathcal{F}x')_i)])$$

is sub-gaussian with ψ_2 norm at most $\mathcal{O}(\sqrt{1+\delta_i}\|x-x'\|_2)$, where $\delta_1, \dots, \delta_m$ are with respect to the increment $x - x'$. To see this, we define

$$t_i := f((\mathcal{F}x)_i) - f((\mathcal{F}x')_i).$$

Because f is Lipschitz and \mathcal{F} is linear, $|t_i| \leq c_0 |(\mathcal{F}(x - x'))_i|$ for some Lipschitz constant c_0 . Then we know that for any $u > 0$, $\exp(t_i^2/u^2) \leq \exp(c_0^2((\mathcal{F}(x - x'))_i)^2/u^2)$. Taking the expectation and the infimum of $u > 0$, this gives $\|t_i\|_{\psi_2} \leq c_0 \|(\mathcal{F}(x - x'))_i\|_{\psi_2}$. Also, we know from Claim 4.6 that $(\mathcal{F}(x - x'))_i \sim \|x - x'\|_2 \cdot \mathcal{N}(0, 1 + \delta_i)$, thus

$$\|t_i\|_{\psi_2} \leq \mathcal{O}(\|(\mathcal{F}(x - x'))_i\|_{\psi_2}) \leq \mathcal{O}(\sqrt{1 + \delta_i}\|x - x'\|_2).$$

Because centering a random variable does not change subgaussianity up to a constant, it follows that $\|t_i - \mathbb{E}[t_i]\|_{\psi_2} \leq \mathcal{O}(\sqrt{1 + \delta_i}\|x - x'\|_2)$.

Since the rows of \mathcal{F} are independent, the random variables $t_i - \mathbb{E}[t_i]$ are independent across $i \in [m]$. Therefore, we can show the concentration of the average using the Hoeffding inequality for subgaussians (Lemma 2.9). This gives

$$\mathbb{P}\left[\left|\sum_{i=1}^m t_i - \mathbb{E}[t_i]\right| > u\right] \leq \exp\left(-\Theta\left(\frac{u^2}{\sum_{i=1}^m (1 + \delta_i)\|x - x'\|_2^2}\right)\right) = \exp\left(-\Theta\left(\frac{u^2}{(m + m/C)\|x - x'\|_2^2}\right)\right),$$

thus

$$\mathbb{P}\left[|\bar{T}_x - \bar{T}_{x'}| > u\right] = \mathbb{P}\left[\left|\frac{1}{m} \sum_{i=1}^m t_i - \mathbb{E}[t_i]\right| > u\right] \leq \exp\left(-\Theta\left(\frac{mu^2}{\|x - x'\|_2^2}\right)\right).$$

□

Claim 4.15. *With constant probability, it holds uniformly for all $x \in \mathcal{B}(0, 1)$ that $|\bar{T}_x| \leq \mathcal{O}(1/C)$.*

Proof. We apply Theorem 4.13 on the metric space $(\mathcal{B}(0, 1), d_{\mathcal{B}})$, where we define $d_{\mathcal{B}}(x, x') := c_0 \|x - x'\|_2 / \sqrt{m}$ for some constant $c_0 > 0$ such that

$$\mathbb{P}[|\bar{T}_x - \bar{T}_{x'}| > u] \leq 2 \exp(-u^2/2d_{\mathcal{B}}(x, x')^2).$$

By Claim 4.14, such a constant c_0 must exist. From the existing results on γ_2 (Corollary 4.10 and 4.12), we know that

$$\gamma_2(\mathcal{B}(0, 1), d_{\mathcal{B}}) = c_0 \cdot \gamma_2(\mathcal{B}(0, 1), \|\cdot\|_2) / \sqrt{m} = \mathcal{O}(\sqrt{n/m}).$$

Then from the fact that $(\bar{T}_x)_{x \in \mathcal{B}(0,1)}$ are centered random variables and satisfy the increment condition, we can conclude that

$$\mathbb{E}[\sup_{x \in \mathcal{B}(0,1)} \bar{T}_x] \leq \mathcal{O}(\sqrt{n/m}).$$

On the other hand, we can get $\mathbb{E}[\sup_{x \in \mathcal{B}(0,1)} (-\bar{T}_x)] \leq \mathcal{O}(\sqrt{n/m})$ by applying Theorem 4.13 on $(-\bar{T}_x)_{x \in \mathcal{B}(0,1)}$. Indeed, this process is still centered and satisfies the increment condition. Combining these two bounds, we get that $\mathbb{E}[\sup_{x \in \mathcal{B}(0,1)} |\bar{T}_x|] \leq \mathcal{O}(\sqrt{n/m})$. Then applying the Markov's inequality, we conclude that with constant probability, it holds uniformly for all $x \in \mathcal{B}(0, 1)$ that $|\bar{T}_x| \leq \mathcal{O}(\sqrt{n/m}) = \mathcal{O}(1/C)$ for $m = \Theta(C^2 n)$. \square

Concluding the above, we get that with constant probability, it holds uniformly for all $x \in \mathcal{B}(0, 1)$ that

$$\left| \frac{1}{m} \sum_{i=1}^m f((\mathcal{F}x)_i) - \mathbb{E}_{g \sim \mathcal{N}(0, \|x\|_2^2)} [f(g)] \right| = |T_x| = \mathcal{O}(1/C) = \mathcal{O}(\varepsilon).$$

\square

5 Applications

We use our inequality (Theorem 4.1) to derive improved algorithms for three applications: ℓ_2 to ℓ_1 embedding, kernel approximation, and adaptive data structure for distance estimation. For the last two applications, we follow the same proof strategy as in [CN22]. For the first application, we just take the Lipschitz function f to be the absolute value function.

5.1 ℓ_2 to ℓ_1 Embedding

Corollary 5.1. *For any $n \in \mathbb{Z}_{>0}$ and $\varepsilon \in (0, 1/2)$, there exists $m = \mathcal{O}(n/\varepsilon^2)$ such that with constant probability, we can sample a linear map $\mathcal{F} \in \mathbb{R}^{m \times n}$ with the following uniform guarantee:*

$$\forall x \in \mathbb{R}^n : (1 - \mathcal{O}(\varepsilon))\|x\|_2 \leq \|\mathcal{F}x\|_1 \leq (1 + \mathcal{O}(\varepsilon))\|x\|_2.$$

Moreover, $\mathcal{F}x$ can be computed in $\mathcal{O}(n \log \frac{n}{\varepsilon}/\varepsilon + npoly(\log \log n/\varepsilon))$ time.

Proof. We first sample $\tilde{\mathcal{F}}$ as described in Theorem 4.1 and condition on it being successful. Then we define $f : \mathbb{R} \rightarrow \mathbb{R}$ to be the absolute function, i.e. $f(x) := |x|$, which is 1-Lipschitz. It follows from Theorem 4.1 that for all $x \in \mathcal{S}^{n-1}$,

$$\left| \frac{1}{m} \sum_{i=1}^m |(\tilde{\mathcal{F}}x)_i| - \mathbb{E}_{g \sim \mathcal{N}(0, 1)} [|g|] \right| = \left| \frac{1}{m} \|\tilde{\mathcal{F}}x\|_1 - \mathbb{E}_{g \sim \mathcal{N}(0, 1)} [|g|] \right| \leq \mathcal{O}(\varepsilon).$$

Because $\mathbb{E}_{g \sim \mathcal{N}(0,1)}[|g|] = \sqrt{2/\pi}$ is a fixed constant and the ℓ_1 norm is positive homogeneous, we get that for any $x \in \mathbb{R}^n$

$$(1 - \mathcal{O}(\varepsilon))\|x\|_2 \leq \frac{1}{m \cdot \sqrt{2/\pi}} \|\tilde{\mathcal{F}}x\|_1 \leq (1 + \mathcal{O}(\varepsilon))\|x\|_2.$$

Therefore, setting $\mathcal{F} := \frac{1}{m\sqrt{2/\pi}}\tilde{\mathcal{F}}$ concludes the proof. \square

5.2 Kernel Approximation

Definition 5.2. For any $x, x' \in \mathbb{R}^n$, we define the RBF kernel $\mathcal{K}(x, x') := \exp(-\|x - x'\|_2^2/2)$.

Corollary 5.3. For any $n \in \mathbb{Z}_{>0}$, $\varepsilon \in (0, 1/2)$, and $\mathcal{W} \subset \mathbb{R}^d$, there exists $m = \mathcal{O}(n \cdot \text{Diam}(\mathcal{W})^2/\varepsilon^2)$ such that if we sample a random function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ as follows:

- Sample $\mathcal{F} : \mathbb{R}^n \rightarrow \mathbb{R}^{m=\mathcal{O}(n \cdot \text{Diam}(\mathcal{W})^2/\varepsilon^2)}$ using Theorem 4.1, with $\varepsilon_{\mathcal{F}}$ set to $\varepsilon/\text{Diam}(\mathcal{W})$;
- then sample $b \sim \text{Unif}[0, 2\pi]^m$.

Define $\psi(x) := \sqrt{\frac{2}{m}} \cos(\mathcal{F}x + b)$, where the function $\cos(\cdot)$ is applied elementwise. Then with constant probability, we get the following guarantee:

$$\forall x, x' \in \mathcal{W} : |\langle \psi(x), \psi(x') \rangle - \mathcal{K}(x, x')| \leq \varepsilon.$$

Moreover, ψx can be computed in $\mathcal{O}(n \log \frac{n}{\varepsilon_{\mathcal{F}}} / \varepsilon_{\mathcal{F}} + n \text{poly}(\log \log n / \varepsilon_{\mathcal{F}}))$ time.

Proof Sketch. This is similar to Section 4 of [CN22]. We will use most parts from [CN22] without proof, and focus on the modified part (which is another chaining argument.)

Lemma 5.4 ([CN22]). For all $x, x' \in \mathbb{R}^n$, $\langle \psi(x), \psi(x') \rangle = \frac{1}{m} \sum_{i=1}^m \cos((\mathcal{F}(x+x'))_i + 2b_i) + \cos((\mathcal{F}(x-x'))_i)$.

[CN22] then shows that the first term is close to 0 and the second term is close to $\mathbb{E}_{Z \sim \mathcal{N}(0, \|x-x'\|_2^2)}[\cos(Z)]$, which is exactly $\exp(-\|x - x'\|_2^2/2) = \mathcal{K}(x, x')$. We state the bound for the second term, which follows from an application of Theorem 4.1 and scaling by $\text{Diam}(\mathcal{W})$ as proven in [CN22].

Lemma 5.5 ([CN22], Lemma 4.2). There exists $m = \mathcal{O}(n \cdot \text{Diam}(\mathcal{W})^2/\varepsilon^2)$ such that with constant probability, it holds for all $x, x' \in \mathcal{W}$ that

$$\left| \frac{1}{m} \sum_{i=1}^m \cos((\mathcal{F}(x-x'))_i) - \mathbb{E}_{Z \sim \mathcal{N}(0, \|x-x'\|_2^2)}[\cos(Z)] \right| \leq \mathcal{O}(\varepsilon).$$

It remains to bound the first term. One way is to again use the proof from [CN22]. However, that relies on a net argument, thus requires $m = \Omega(n \log(\frac{1}{\varepsilon})/\varepsilon^2)$. In the regime where $\text{Diam}(\mathcal{W})^2 \lesssim \log(\frac{1}{\varepsilon})$, this dominates and exceeds the claimed dimension. To remove this $\log(\frac{1}{\varepsilon})$ factor, we invoke Theorem 4.13 again.

Lemma 5.6. *There exists $m = \mathcal{O}(n \cdot \text{Diam}(\mathcal{W})^2/\varepsilon^2)$ such that with constant probability, it holds for all $x, x' \in \mathcal{W}$ that*

$$\left| \frac{1}{m} \sum_{i=1}^m \cos((\mathcal{F}(x+x'))_i + 2b_i) \right| \leq \mathcal{O}(\varepsilon).$$

Proof. Let $\mathcal{H} := \{x+x' : x, x' \in \mathcal{W}\}$. We consider a process $(Y_v)_{v \in \mathcal{H}}$ where

$$Y_v := \frac{1}{m} \sum_{i=1}^m \cos((\mathcal{F}v)_i + 2b_i),$$

where the randomness is over both \mathcal{F} and b , which are independent by construction. Because $b \sim \text{Unif}[0, 2\pi]^m$, we have that $\mathbb{E}_{\mathcal{F}, b}[Y_v] = 0$. It remains to show the increment condition:

Claim 5.7. *For any $\lambda > 0$ and $u, v \in \mathcal{H}$,*

$$\mathbb{P}[|Y_u - Y_v| > \lambda] \leq 2 \exp(-\lambda^2/2d_{\mathcal{H}}(u, v)^2)$$

where $d_{\mathcal{H}}(\cdot) = c_0 \|u - v\|_2 / \sqrt{m}$ for some constant $c_0 > 0$.

Proof. We define

$$t_i := \cos((\mathcal{F}u)_i + 2b_i) - \cos((\mathcal{F}v)_i + 2b_i),$$

which is a random variable depending on both \mathcal{F} and b . Such t_i has zero expectation because $b \sim \text{Unif}[0, 2\pi]^m$. Moreover, if we condition on a fixed realization of b , then $f_i(a) := \cos(a + 2b_i)$ is 1-Lipschitz, thus for any $\lambda > 0$,

$$\mathbb{E}[\exp(\lambda t_i^2) | b] \leq \mathbb{E}[\exp(\lambda(\mathcal{F}(u-v))_i^2)].$$

Taking the expectation in b and then use the law of total expectation, we obtain that unconditionally,

$$\|t_i\|_{\psi_2} \leq \|(\mathcal{F}(u-v))_i\|_{\psi_2} \leq \mathcal{O}(\sqrt{1+\delta_i} \|u-v\|_2)$$

where the last inequality follows from $(\mathcal{F}(u-v))_i \sim \|u-v\|_2 \mathcal{N}(0, 1+\delta_i)$ for $\sum_{i=1}^m |\delta_i| = \mathcal{O}(\varepsilon m)$. Then we can use the Hoeffding inequality (Lemma 2.9) to show that $\mathbb{P}[|\sum_{i=1}^m t_i| > \lambda] \leq \exp(-\Theta(\frac{\lambda^2}{m\|u-v\|_2^2}))$, thus

$$\mathbb{P}[|Y_u - Y_v| > \lambda] = \mathbb{P}\left[\left| \frac{1}{m} \sum_{i=1}^m t_i \right| > \lambda \right] \leq \exp\left(-\Theta\left(\frac{m\lambda^2}{\|u-v\|_2^2}\right)\right).$$

□

We can now apply Theorem 4.13 on the metric space $(\mathcal{H}, d_{\mathcal{H}})$, where we define $d_{\mathcal{H}}(u, v) := c_0 \|u - v\|_2 / \sqrt{m}$ for some appropriate constant $c_0 > 0$ satisfying the increment condition. It follows that

$$\mathbb{E}[\sup_{v \in \mathcal{H}} Y_v] \leq \mathcal{O}(\gamma_2(\mathcal{H}, d_{\mathcal{H}})).$$

Claim 5.8. $\gamma_2(\mathcal{H}, d_{\mathcal{H}}) = \Theta(\text{Diam}(\mathcal{W})\sqrt{n}/\sqrt{m})$.

Proof. Because $\mathcal{H} := \{x + x' : x, x' \in \mathcal{W}\}$, we know that $\text{Diam}(\mathcal{H}) \leq 2\text{Diam}(\mathcal{W})$, i.e. there exists a “center” $v^* \in \mathbb{R}^n$ such that $\mathcal{H} \subseteq \mathcal{B}(v^*, 2\text{Diam}(\mathcal{W}))$. We consider the set $\tilde{\mathcal{H}} := \mathcal{H} - v^*$, then $\tilde{\mathcal{H}} \subseteq \mathcal{B}(0, 2\text{Diam}(\mathcal{W}))$ and

$$\mathbb{E}_{w \sim \mathcal{N}(0,1)^n} [\sup_{v \in \mathcal{H}} \langle v, w \rangle] = \mathbb{E}_{w \sim \mathcal{N}(0,1)^n} [\sup_{v \in \tilde{\mathcal{H}}} \langle v, w \rangle]$$

because $\mathbb{E}_{w \sim \mathcal{N}(0,1)^n} [\langle v^*, w \rangle] = 0$ for any fixed offset v^* . So we bound

$$\mathbb{E}_{w \sim \mathcal{N}(0,1)^n} [\sup_{v \in \tilde{\mathcal{H}}} \langle v, w \rangle] \leq \mathbb{E} [\sup_{v \in \mathcal{B}(0, 2\text{Diam}(\mathcal{W}))} \langle v, w \rangle] = 2\text{Diam}(\mathcal{W}) \mathbb{E} [\sup_{v \in \mathcal{B}(0,1)} \langle v, w \rangle] \leq 2\text{Diam}(\mathcal{W})\sqrt{n}.$$

By the Majorizing Theorem, this gives $\gamma_2(\mathcal{H}, \|\cdot\|_2) = \Theta(\text{Diam}(\mathcal{W})\sqrt{n})$. Using Corollary 4.10, this gives $\gamma_2(\mathcal{H}, d_{\mathcal{H}}) = \Theta(\text{Diam}(\mathcal{W})\sqrt{n}/\sqrt{m})$. \square

Therefore, for our choice of $m = \Theta(n \cdot \text{Diam}(\mathcal{W})^2/\varepsilon^2)$, we get that

$$\mathbb{E} [\sup_{x+x': x, x' \in \mathcal{W}} \frac{1}{m} \sum_{i=1}^m \cos((\mathcal{F}(x+x'))_i + 2b_i)] \leq \Theta(\varepsilon)$$

and the Markov’s inequality tells us that with constant probability, this supremum is within a constant factor of its expectation. Finally, the same argument from the proof of Claim 4.15 shows that the supremum of absolute is asymptotically the same. \square

\square

5.3 Distance Estimation

Definition 5.9. In the Distance Estimation problem for a known metric $d(\cdot, \cdot)$, we are given $X = \{x_i\}_{i=1}^s \subset \mathbb{R}^n$ and $\varepsilon \in (0, 1/2)$, and we are required to construct a data structure \mathcal{D} , which when given input query q , outputs distance estimates $\{d_i\}_{i=1}^s$ satisfying:

$$(1 - \varepsilon)d(q, x_i) \leq d_i \leq (1 + \varepsilon)d(q, x_i).$$

Corollary 5.10. For any $n \in \mathbb{Z}_{>0}$ and $\varepsilon \in (0, 1/2)$, there exists $m = \mathcal{O}(n/\varepsilon^2)$ such that with constant probability, we can correctly initialize a data structure for Distance Estimation in Euclidean space, which supports the following operations:

- Output a correct answer to a possibly adaptively chosen distance estimation query with constant probability;
- Add input $x \in \mathbb{R}^n$ to the database, X .

Furthermore, the processing time for each data vector is $\mathcal{O}(n \log \frac{n}{\varepsilon}/\varepsilon + npoly(\log \log n/\varepsilon))$ and the space to store each processed vector is $\mathcal{O}(n/\varepsilon^2)$ words (assuming that every coordinate fits in a word). The query time of the data structure is $\mathcal{O}(n \log \frac{n}{\varepsilon}/\varepsilon + npoly(\log \log n/\varepsilon) + s \log(s) \log(\frac{1}{\varepsilon})/\varepsilon^2)$.

Proof Sketch. We make only a minimum change to the algorithms described in Section 5.1 of [CN22]: when initializing the data structure, we sample \mathcal{F} using Theorem 4.1; and when responding to queries and making updates, we use \mathcal{F} instead of \tilde{h} as in [CN22]. Since the proof of correctness relies solely on the concentration of averages (Theorem 4.1), this immediately gives a data structure with the desired guarantees. \square

References

- [AC09] Nir Ailon and Bernard Chazelle. The fast johnson–lindenstrauss transform and approximate nearest neighbors. *SIAM Journal on computing*, 39(1):302–322, 2009.
- [Ach03] Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003.
- [AL13] Nir Ailon and Edo Liberty. An almost optimal unrestricted fast johnson-lindenstrauss transform. *ACM Transactions on Algorithms (TALG)*, 9(3):1–12, 2013.
- [BIJ⁺23] Ainesh Bakshi, Piotr Indyk, Rajesh Jayaram, Sandeep Silwal, and Erik Waingarten. Near-linear time algorithm for the chamfer distance. *Advances in Neural Information Processing Systems*, 36:66833–66844, 2023.
- [BLM89] Jean Bourgain, Joram Lindenstrauss, and Vitali Milman. Approximation of zonoids by zonotopes. 1989.
- [CN20] Yeshwanth Cherapanamjeri and Jelani Nelson. On adaptive distance estimation. *Advances in Neural Information Processing Systems*, 33:11178–11190, 2020.
- [CN22] Yeshwanth Cherapanamjeri and Jelani Nelson. Uniform approximations for randomized hadamard transforms with applications. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 659–671, New York, NY, USA, 2022. Association for Computing Machinery.
- [CRT06a] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.
- [CRT06b] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.
- [CSWZ23] Yeshwanth Cherapanamjeri, Sandeep Silwal, David P Woodruff, and Samson Zhou. Optimal algorithms for linear algebra in the current matrix multiplication time. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4026–4049. SIAM, 2023.
- [CT06] Emmanuel J. Candès and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.
- [DG03] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and Lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.

- [DMS24] Sjoerd Dirksen, Shahar Mendelson, and Alexander Stollenwerk. Fast metric embedding into the hamming cube. *SIAM Journal on Computing*, 53(2):315–345, 2024.
- [Don06] David Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.
- [Dvo64] Aryeh Dvoretzky. Some results on convex bodies and banach spaces. *Matematika*, 8(1):73–102, 1964.
- [FI25] Ying Feng and Piotr Indyk. Even faster algorithm for the chamfer distance. In *52nd International Colloquium on Automata, Languages, and Programming (ICALP 2025)*, pages 76–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2025.
- [FLM77] T Figiel, J Lindenstrauss, and VD Milman. The dimension of almost spherical sections of convex bodies. *Acta Mathematica*, 139(1):53–94, 1977.
- [FR13] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.
- [Gor85] Yehoram Gordon. Some inequalities for gaussian processes and applications. *Israel Journal of Mathematics*, 50(4):265–289, 1985.
- [IN07] Piotr Indyk and Assaf Naor. Nearest-neighbor-preserving embeddings. *ACM Trans. Algorithms*, 3(3):31–es, August 2007.
- [Ind07] Piotr Indyk. Uncertainty principles, extractors, and explicit embeddings of ℓ_2 into ℓ_1 . In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, STOC ’07, pages 615–620, New York, NY, USA, 2007. Association for Computing Machinery.
- [JL⁺84] William B Johnson, Joram Lindenstrauss, et al. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984.
- [KW11] Felix Krahmer and Rachel Ward. New and improved johnson-lindenstrauss embeddings via the restricted isometry property. *SIAM Journal on Mathematical Analysis*, 43(3):1269–1281, 2011.
- [LV06] Yurii Lyubarskii and Roman Vershynin. Uncertainty principles and vector quantization. *IEEE Transactions on Information Theory*, 56:3491–3501, 2006.
- [Mil71] Vitali D Milman. A new proof of a. dvoretzky’s theorem on cross-sections of convex bodies. *Funkcional. Anal. i Prilozhen*, 5:28–37, 1971.
- [Pis89] Gilles Pisier. *The Volume of Convex Bodies and Banach Space Geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, 1989.
- [RV08] Mark Rudelson and Roman Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 61(8):1025–1045, 2008.

- [Sar06] Tamas Sarlos. Improved approximation algorithms for large matrices via random projections. In *2006 47th annual IEEE symposium on foundations of computer science (FOCS'06)*, pages 143–152. IEEE, 2006.
- [Tal21] Michel Talagrand. Upper and lower bounds for stochastic processes. 2021.
- [Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.
- [Zuc97] David Zuckerman. Randomness-optimal oblivious sampling. *Random Struct. Algorithms*, 11(4):345–367, December 1997.

A RIP matrices of fixed size

Lemma A.1. *Let N, k, n be parameters satisfying the relationship in Lemma 3.3, and let Ξ be a uniformly random subset of size n sampled from $[N]$ without replacement. If we define $M := H_{\Xi} \in \mathbb{R}^{n \times N}$, then*

$$\mathbb{P}\left[\sqrt{\frac{N}{n}}M \text{ satisfies } (k, 1)\text{-RIP}\right] \geq 1 - n^{-9}.$$

Proof. We let p denote the failure probability, i.e. $p := \mathbb{P}\left[\sqrt{\frac{N}{n}}M \text{ does not satisfy } (k, 1)\text{-RIP}\right]$. Also, we let Ω be a subset sampled via $Y_1, \dots, Y_N \sim \text{Bernoulli}(n/N)$ and $M' := H_{\Omega}$ as described previously. Then $\mathbb{E}[|\Omega|] = n$. By a standard Stirling approximation, we have that

$$\mathbb{P}[|\Omega| = n] = \binom{N}{n} \left(\frac{n}{N}\right)^n \left(1 - \frac{n}{N}\right)^{N-n} = \frac{t}{\sqrt{n}}$$

for some constant $t > 0$. Conditioning on $|\Omega| = n$, the distribution of Ω is the same as the distribution of Ξ . Therefore, it follows from Lemma 3.3 that

$$\begin{aligned} 1 - n^{-10} &\leq \mathbb{P}\left[\sqrt{\frac{N}{|\Omega|}}M' \text{ satisfies } (k, 1)\text{-RIP}\right] \\ &= \sum_{i=0}^N \mathbb{P}\left[\sqrt{\frac{N}{|\Omega|}}H_{\Omega} \text{ satisfies } (k, 1)\text{-RIP} \mid |\Omega| = i\right] \cdot \mathbb{P}[|\Omega| = i] \\ &\leq 1 \cdot \left(1 - \frac{t}{\sqrt{n}}\right) + \mathbb{P}\left[\sqrt{\frac{N}{n}}H_{\Xi} \text{ satisfies } (k, 1)\text{-RIP}\right] \cdot \frac{t}{\sqrt{n}} \\ &= 1 - \frac{t}{\sqrt{n}} + (1 - \mathbb{P}\left[\sqrt{\frac{N}{n}}M \text{ does not satisfy } (k, 1)\text{-RIP}\right]) \cdot \frac{t}{\sqrt{n}} \\ &= 1 - p \cdot \frac{t}{\sqrt{n}} \end{aligned}$$

Thus $p \leq n^{-10} \frac{\sqrt{n}}{t} \leq n^{-9}$ for large enough n . \square