The Fast Johnson-Lindenstrauss Transform Is Even Faster

Ora Nova Fandina [*1] **Mikael Møller Høgsgaard** [*1] **Kasper Green Larsen** [*1]

Abstract

The Johnson-Lindenstaruss lemma (Johnson & Lindenstrauss, 1984) is a cornerstone result in dimensionality reduction, stating it is possible to embed a set of n points in d-dimensional Euclidean space into optimal $k = O(\varepsilon^{-2} \ln n)$ dimensions, while preserving all pairwise distances to within a factor $(1 \pm \varepsilon)$.

The seminal Fast Johnson-Lindenstrauss (Fast JL) transform by Ailon and Chazelle (SICOMP'09) supports computing the embedding of a data point in $O(d \ln d + k \ln^2 n)$ time, where the $d \ln d$ term comes from multiplication with a $d \times d$ Hadamard matrix and the $k \ln^2 n$ term comes from multiplication with a sparse $k \times d$ matrix. Despite the Fast JL transform being more than a decade old, it is one of the fastest dimensionality reduction techniques for many tradeoffs between ε, d and n.

In this work, we give a surprising new analysis of the Fast JL transform, showing that the $k \ln^2 n$ term in the embedding time can be improved to $(k \ln^2 n)/\alpha$ for an $\alpha = \Omega(\min\{\varepsilon^{-1} \ln(1/\varepsilon), \ln n\})$. The improvement follows by using an even sparser matrix. We complement our improved analysis with a lower bound showing that our new analysis is in fact tight.

1. Introduction

Dimensionality reduction is a central technique for speeding up algorithms and reducing the memory footprint of large data sets. The basic idea is to map a set $X \subset \mathbb{R}^d$ of n high-dimensional points to a lower dimensional representation, while approximately preserving similarities between the points. The most fundamental result in dimensionality reduction, is the Johnson-Lindenstrauss transform (Johnson & Lindenstrauss, 1984), which for any precision $0 < \varepsilon < 1$, gives a mapping $f : X \to \mathbb{R}^k$ with $k = O(\varepsilon^{-2} \ln n)$ such that

$$\forall x, y \in X : \|f(x) - f(y)\|_2 \in (1 \pm \varepsilon)\|x - y\|_2. \quad (1)$$

That is, the pairwise Euclidean distance between the embeddings of any two points $x, y \in X$ is within a factor $(1 \pm \varepsilon)$ of the original distance. The target dimensionality of $k = O(\varepsilon^{-2} \ln n)$ is known to be optimal (Larsen & Nelson, 2017; Alon & Klartag, 2017). For algorithmic applications where one can tolerate a small loss of precision, one can apply a Johnson-Lindenstrauss transform as a preprocessing step to reduce the dimensionality of the input. Since the running time of most algorithms depend on the dimensionality of the input, this typically speeds up the analysis while also reducing memory consumption.

A simple construction of a mapping f satisfying Equation (1) is to let $f(x) = k^{-1/2}Ax$, where A is a random $k \times d$ matrix, having each entry i.i.d. $\mathcal{N}(0, 1)$ distributed (Indyk & Motwani, 1998). This results in an embedding time of $O(kd)$ to compute the matrix-vector product Ax. For some applications, this embedding time may dominate the running time of the algorithms applied to the embedded data, hence dimensionality reducing maps with a faster embedding time has been the focus of much research. The line of research on faster dimensionality reducing maps splits roughly into two categories: 1) maps based on sparse matrices, and 2), maps based on structured matrices with fast matrix-vector multiplication algorithms.

Sparse JL. A sparse JL transform is obtained by replacing the dense matrix A above with a matrix having only t non-zero entries per column. Computing the product Ax now takes only $O(td)$ time instead of $O(kd)$. Perhaps even more importantly, if the input vectors $x \in X$ are themselves sparse vectors, then the embedding time is further reduced to $O(t\|x\|_0)$, where $\|x\|_0$ denotes the number of non-zero entries in x. This is particularly useful when applying JL on e.g. bag-of-words, n-gram or tf-idf representations of text documents (Manning & Schütze, 1999), which are often very sparse. The fastest (sparsest) known construction, due to Kane and Nelson (Kane & Nelson,

[*] Equal contribution [1] Department of Computer Science, Aarhus University, Aarhus, Denmark. Correspondence to: Ora Nova Fandina <fandina@gmail.com>, Mikael Møller Høgsgaard <hogsgaard@cs.au.dk>, Kasper Green Larsen <larsen@cs.au.dk>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

2014), achieves $t = O(\varepsilon^{-1} \ln n)$, which nearly matches a sparsity lower bound by Nelson and Nguyen (Nelson & Nguyen, 2013), stating that any Sparse JL must have $t = \Omega(\varepsilon^{-1} \ln n / \ln(1/\varepsilon))$. Sparse JL thus improves over classic JL by an ε^{-1} factor.

While the lower bound by Nelson and Nguyen rules out significant further improvements, the Feature Hashing technique by Weinberger et al. (Weinberger et al., 2009) study the extreme case of $t = 1$. Since this is below the sparsity lower bound, they have to assume that the ratio $\nu = \|z\|_\infty / \|z\|_2$ is small for all pairwise difference vectors $z = y - x$ for $x, y \in X$ to ensure Equation (1) holds. Determining the exact ratio ν for which Equation (1) holds was subsequently done by Freksen et al. (Freksen et al., 2018) and generalized to t-sparse embeddings for all $t \geq 1$ by Jagadeesan (Jagadeesan, 2019).

Fast JL. Ailon and Chazelle (Ailon & Chazelle, 2009) initiated the study of JL transforms that exploit dense matrices with fast matrix-vector multiplication algorithms. Concretely, they defined the Fast JL transform where the embedding of a vector x is computed as $PHDx$, such that D is a diagonal matrix with random signs on the diagonal, H is a $d \times d$ standardized Hadamard matrix and P is a sparse $k \times d$ matrix. Computing Dx takes only $O(d)$ time, and multiplication with the Hadamard matrix can be done in $O(d \ln d)$ time. The key observation that permits a very sparse matrix P, is that with high probability, the vector $y = HDx$ has a small ratio $\nu = \|y\|_\infty / \|y\|_2$, i.e. no single entry contributes most of the "mass". As was the case for Feature Hashing, such a bound allows for an even sparser random projection matrix P than what a Sparse JL transform could achieve. Ailon and Chazelle proved that a matrix P in which each entry is non-zero only with probability $q = O((\ln^2 n)/d)$ suffices for Equation (1). Thus the expected number of non-zeroes in P is $kdq = O(k \ln^2 n)$ (also with high probability) and the product Py can be computed in $O(k \ln^2 n)$ time. This yields a total embedding time of $O(d \ln d + k \ln^2 n)$.

Numerous follow-up works have attempted to improve over the Fast JL construction of Ailon and Chazelle, in particular attempting to shave off the $k \ln^2 n$ additive term to obtain a clean $O(d \ln d)$ time embedding. These approaches naturally divide into a couple of categories. First, a number of constructions sacrifice the optimal target dimensionality of $k = O(\varepsilon^{-2} \ln n)$ for faster embedding time. This includes e.g. five solutions with $O(d \ln d)$ embedding time, but different sub-optimal $k = O(\varepsilon^{-2} \ln n \ln^4 d)$ (Krahmer & Ward, 2011), $k = O(\varepsilon^{-2} \ln^3 n)$ (Do et al., 2009), $k = O(\varepsilon^{-1} \ln^{3/2} n \ln^{3/2} d + \varepsilon^{-2} \ln n \ln^4 d)$ (Krahmer & Ward, 2011), $k = O(\varepsilon^{-2} \ln^2 n)$ (Hinrichs & Vybíral, 2011; Vybiral, 2010; Freksen & Larsen, 2020) and $k = O(\varepsilon^{-2} \ln n \ln^2(\ln n) \ln^3 d)$ (Jain et al., 2020), respectively. The second category is solutions where one assumes that k is significantly smaller than d. Here there are two solutions that both achieve $O(d \ln k)$ embedding time under the assumption that $k = o(d^{1/2})$ (Ailon & Liberty, 2008; Bamberger & Krahmer, 2021). Among solutions that insist on optimal $k = O(\varepsilon^{-2} \ln n)$ and that make no assumption about the relationship between k and d (other than the obvious $k \leq d$), only the recent analysis (Jain et al., 2020) of the Kac JL transform (Kac, 1958) improves over the classic Fast JL solution by Ailon and Chazelle for some tradeoffs between ε, d and n. The Kac JL transform works by repeatedly picking two coordinates and doing a random unitary rotation on the two coordinates. After a sufficient number of steps, one projects on to the first $k = O(\varepsilon^{-2} \ln n)$ coordinates and scales the coordinates appropriately. Since each rotation takes $O(1)$ time, the running time is proportional to the number of steps needed. Jain et al. (Jain et al., 2020) showed that

$$O(d \ln d + \min\{d \ln n, k \ln n \ln^2(\ln n) \ln^3 d\}) \quad (2)$$

rotations suffice. Compared to the $O(d \ln d + k \ln^2 n)$ embedding time of Fast JL, Kac JL is an improvement unless $\ln^3 d > \ln n / \ln^2(\ln n)$. Despite these numerous approaches to Fast JL, we still lack a clean $O(d \ln d)$ or $O(d \ln k)$ time solution.

Our Contributions. While Fast JL has been the focus of a considerable amount of research, we give a surprising new analysis of the classic Fast JL transform by Ailon and Chazelle. Our analysis shows that the sparsity parameter q in the matrix P can be lowered by a factor $\Omega(\min\{\varepsilon^{-1} \ln(1/\varepsilon), \ln n\})$, thereby yielding a similar improvement in embedding time. Concretely, we show that Fast JL can embed a vector x in time:

$$O\left(d \ln d + \min\left\{\frac{d \ln n}{\varepsilon}, k \ln n \cdot \max\left\{1, \frac{\varepsilon \ln n}{\ln(1/\varepsilon)}\right\}\right\}\right). \quad (3)$$

While this rather complicated expression might seem like an artifact of our proof, we complement our improved upper bound by showing the existence of a vector requiring precisely this embedding time using the $PHDx$ Fast JL construction. In later sections, we also give an intuitive description of where the different terms originate from.

Before giving more details on our results, let us thoroughly compare the bound to previous work. Compared to the classic $O(d \ln d + k \ln^2 n)$ Fast JL bound, we observe that Equation (3) is always bounded by $O(d \ln d + k \ln n \max\{1, \varepsilon \ln n / \ln(1/\varepsilon)\})$, i.e. the term $O(k \ln^2 n)$ is improved by a factor $\Omega(\min\{\varepsilon^{-1} \ln(1/\varepsilon), \ln n\})$. Also, if we consider the case of $\varepsilon = O(\ln(\ln n)/\ln n)$, then 1 takes the maximum value in the max-expression and the bound simplifies to $O(d \ln d + k \ln n)$. Comparing this clean bound to the Kac JL bound in Equation (2), this is a strict improvement (for $\varepsilon < \ln(\ln n)/\ln n$).

We further show that one can replace the Normal variables in the projection matrix P with i.i.d. Rademacher variables and obtain the same embedding time as in Equation (3), thereby providing more efficient way to implement the Fast JL construction with the same theoretical guarantees. This is an improvement by a factor $\Omega(\min\{\varepsilon^{-1}\ln(1/\varepsilon), \ln n\})$ over the bound proved by Matoušek (Matoušek, 2008) for the same construction, who showed that the embedding time is bounded by $O(d \ln d + k \ln^2 n)$.

In the next section, we give a detailed description of the Fast JL transform and formally state our new results.

2. The Fast JL Transform

In the spirit of (Ailon & Chazelle, 2009) we now introduce the notation for the Fast JL transform. Here we let d denote the input dimension and k the output dimension. We assume d is a power of two, which can always be ensured by padding with 0's. The Fast JL transform is the composition of three matrices $P \in \mathbb{R}^{k\times d}$ and $H, D \in \mathbb{R}^{d\times d}$. Here D is a random diagonal matrix with independent Rademacher variables ($D_{i,i}$ is 1 or -1 with equal probability) on its diagonal, H is the normalized $d \times d$ Hadamard matrix (denoted H_d in the following):

$$H_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}, H_d = \frac{1}{\sqrt{2}}\begin{pmatrix}H_{d/2} & H_{d/2}\\ H_{d/2} & -H_{d/2}\end{pmatrix}$$

and P is a random matrix with the (i,j)'th entry being $\sqrt{1/q}\ b_{i,j}N_{i,j}$ where $b_{i,j}$ is a Bernoulli random variable with success probability/sparsity parameter q and $N_{i,j}$ is a standard normal random variable, where all the $b_{i,j}$'s, $N_{i,j}$'s and $D_{i,i}$'s are independent of each other. The final embedding of a vector x is then computed as $k^{-1/2}PHDx$.

Analysis Sketch. As is standard in the analysis of JL transforms, we observe that $k^{-1/2}PHD$ is a linear transformation. Hence for $k^{-1/2}PHD$ to satisfy Equation (1) for a set of points X, it suffices that $k^{-1/2}PHD$ preserves the norm of every vector $z = x - y$ with $x, y \in X$ to within a factor $(1 \pm \varepsilon)$. Also by linearity, we guarantee this by arguing that $k^{-1/2}PHD$ preserves the norm of a fixed unit vector x to within $(1 \pm \varepsilon)$ with probability $1 - \delta$ when $k = O(\varepsilon^{-2}\lg(1/\delta))$. Setting $\delta = 1/n^3$ and applying a union bound over all normalized difference vectors $z/\|z\|$ with $z = x - y$ for $x, y \in X$ ensures Equation (1) holds with probability $1 - 1/n$. For shorthand, we from here on use $\|\cdot\|$ to denote the norm $\|\cdot\|_2$.

To build some intuition for the key ideas used to show that the PHD construction approximately preserves the norm of a unit vector with high probability, we first observe that H and D are both unitary matrices, hence HDx preserves the norm of any vector x. Moreover, if we examine a single coordinate $(HDx)_i$, then it is distributed as $d^{-1/2}\sum_j \sigma_j x_j$ for independent Rademachers $\sigma_j = \mathrm{sign}(H_{i,j})D_{j,j}$. Standard tail bounds show that $(HDx)_i$ is bounded by $\sqrt{\ln(d/\delta)/d}$ in absolute value with probability $1 - \delta/d$ when x has unit norm. A union bound over all d coordinates gives that they are all bounded by $\sqrt{\ln(d/\delta)/d}$ with probability $1 - \delta$. Now that HDx has only small coordinates (recall x has unit norm), it suffices to use a very sparse matrix P, precisely as in the analysis of Feature Hashing. Recall that we will set $\delta \leq 1/n^3$ and thus the d term in $\ln(d/\delta)$ is irrelevant for $d \leq n$. For simplicity, we will thus assume $d \leq n$, which is also consistent with previous work (it was assumed both for Fast JL (Ailon & Chazelle, 2009) and Kac JL (Jain et al., 2020)).

Upper Bounds. In their work, Ailon and Chazelle showed that it suffices to set

$$q = O(\ln^2(n)/d) \tag{4}$$

to guarantee Equation (1) for a set X of n points (with probability $1 - 1/n$ by setting $\delta = 1/n^3$). Their proof follows the template above, union bounding over preserving the norm of all normalized pairwise difference vectors. This results in an expected $kdq = O(k \ln^2 n)$ number of non-zero entries in P. Our main upper bound result is an improved analysis, showing that an even sparser P suffices:

Theorem 1. *Let X be a set of n vectors in $\mathbb{R}^d$ and let $k = \Theta(\varepsilon^{-2}\ln n)$. Let further $0 < \varepsilon \leq C$ where C is some universal constant. Then for*

$$q = O\left(\min\left\{\varepsilon, \frac{\ln n}{d}\cdot\max\left\{1, \frac{\varepsilon \ln n}{\ln(1/\varepsilon)}\right\}\right\}\right),$$

it holds that $k^{-1/2}PHD$ satisfies Equation (1) with probability at least $1 - 1/n$.

The above is also true for matrices P where instead of $N_{i,j}$ variables one samples i.i.d. Rademacher variables.

Compared to Equation (4), we notice that even if we ignore the first term in the min-expression, our guarantee on q is $q = O(\max\{\ln(n)/d, \varepsilon\ln^2(n)/(d\ln(1/\varepsilon))\})$, i.e. always at least a factor $\Omega(\min\{\ln n, \varepsilon^{-1}\ln(1/\varepsilon)\})$ better. Also, for the case of $\varepsilon = O(\ln(\ln n)/\ln n)$, the 1-term in the max dominates, and the expression for q simplifies to a clean $q = O(\ln(n)/d)$. Plugging in the value of q from Theorem 1 (and recalling $k = \Theta(\varepsilon^{-2}\ln n)$), we get that the number of non-zeroes of P is

$$kdq = O\left(\min\left\{\varepsilon^{-1}d\ln n, k\ln n\cdot\max\left\{1, \frac{\varepsilon\ln n}{\ln(1/\varepsilon)}\right\}\right\}\right)$$

in expectation. Moreover, since this number is larger than $\ln n$, it follows from a Chernoff bound that the number of non-zeroes is strongly concentrated around its mean.

Lower Bound. A natural question to ask now is whether the above q is optimal, or an even more refined analysis can lead to further improvements. We show that our analysis is tight for the case of Normal entries in the matrix P. In particular, we give an example of a unit vector x such that for the mapping $k^{-1/2}PHDx$ to preserve the norm of x to within $(1 \pm \varepsilon)$ with probability $1-\delta$, we cannot make P sparser than in Theorem 1:

Theorem 2. *For $\delta > 0, \varepsilon \leq r$ where r is a universal constant and $k = \varepsilon^{-2}\ln(1/\delta)$, there is a unit vector $x \in \mathbb{R}^d$ for which we must have*

$$q = \Omega\left(\min\left\{\varepsilon, \frac{\ln(1/\delta)}{d} \cdot \max\left\{1, \frac{\varepsilon\ln(1/\delta)}{\ln(1/\varepsilon)}\right\}\right\}\right)$$

for $\frac{1}{\sqrt{k}}\|PHDx\| \in (1 \pm \varepsilon)$ to hold with probability at least $1-\delta$, where P is the projection matrix with Normal entries.

For the reader concerned with assuming $k = \varepsilon^{-2}\ln(1/\delta)$, we remark that Theorem 2 can also be shown with $k = \tilde{c}\varepsilon^{-2}\ln(1/\delta)$ for $\tilde{c} \geq 1$, and another universal constant r'.

Comparing Theorem 2 to Theorem 1, we observe that the bounds on q match exactly when setting $\delta = n^{-\Theta(1)}$. This means that the analysis of Fast JL cannot be improved if one attempts to show that any fixed vector has its norm preserved except with probability $n^{-\Theta(1)}$ and doing a union bound over all pairwise difference vectors. It is however still conceivable that a more refined analysis could somehow argue that there are only very few worst case vectors in any set X. However, such an improved analysis remains to be seen for any JL transform (when focusing only on the type of guarantee in Equation (1), whereas net-based arguments have been used e.g. for subspace embeddings (Clarkson & Woodruff, 2013)). In this light, Theorem 2 can be seen either as a hard barrier for Fast JL, or as hinting at a way towards further improvements.

In the next section, we formally prove Theorem 1 and also discuss how our analysis differs from the previous analysis by Ailon and Chazelle and conclude by giving more intuition on where the different terms in the expression for q come from.

3. Upper Bound

In this section we give the proof of Theorem 1 for Normal variables in P, while the sketch for the result for Rademacher variables in P can be found in Appendix C. We start by giving the high level ideas of our proof. As in previous works, our analysis follows by arguing that for any fixed unit vector x, it holds with probability at least $1 - 1/n^3$ that $\|k^{-1/2}PHDx\| \in (1 \pm \varepsilon)$.

First, we observe that HD is a unitary matrix and thus $\|HDx\| = \|x\| = 1$ for a unit vector x. Moreover, any single coordinate $(HDx)_i$ equals $d^{-1/2}\sum_{j=1}^{d}\sigma_j x_j$, where the $\sigma_j = D_{j,j}\operatorname{sign}(H_{i,j})$'s are independent Rademacher random variables. Thus in line with the analysis by Ailon and Chazelle, we get that any coordinate $(HDx)_i$ is bounded by $O(\sqrt{\ln(n)/d})$ in absolute value with probability $1 - 1/n^4$. A union bound over all $d \leq n$ coordinates (this assumption is also made in previous work) gives that all coordinates of HDx are bounded by $O(\sqrt{\ln(n)/d})$ with probability $1 - 1/n^3$.

What remains now is to argue that $k^{-1/2}\|Pu\| \in (1 \pm \varepsilon)$ with high probability when $u = HDx$ is a unit vector with all coordinates bounded by $O(\sqrt{\ln(n)/d})$.

To simplify the analysis, we will argue that $k^{-1}\|Pu\|^2 \in (1 \pm \varepsilon)$ with probability $1 - 1/n^3$. This is stronger since $\sqrt{1 \pm \varepsilon} \subset (1 \pm \varepsilon)$. To understand the distribution of $\|Pu\|^2$ for a fixed u, notice that the i'th coordinate of Pu is given by $\sum_{j=1}^{d} q^{-1/2}u_j b_{i,j} N_{i,j}$. Let us assume that the Bernoulli random variables $b_{i,j}$ have been fixed. In this case, $(Pu)_i$ is a sum of weighted and independent $\mathcal{N}(0,1)$ random variables, hence $(Pu)_i$ is itself $\mathcal{N}(0, q^{-1}\sum_{j=1}^{d} b_{i,j}u_j^2)$ distributed. Now define $Z_i = \sum_{j=1}^{d} b_{i,j}u_j^2$ and let $N_1, \ldots, N_k$ be i.i.d. $\mathcal{N}(0,1)$ variables. Then for fixed values of the Bernoullis, $\|Pu\|^2$ is distributed as $\sum_{i=1}^{k} q^{-1}Z_i N_i^2$. Our proof now has two steps: 1.) Give a bound on the Z_i's that holds with high probability over the random choice of the Bernoullis $b_{i,j}$, and 2.) Use the bound on the Z_i's to argue that $\sum_{i=1}^{k} q^{-1}Z_i N_i^2$ behaves in a desirable manner.

In order to understand what type of bounds we need on the Z_i's, we start by examining step 2. For this step, we need a tail bound on $\sum_{i=1}^{k} q^{-1}Z_i N_i^2$. When the Z_i's are fixed, this is a weighted sum of sub-exponential random variables. To analyse it, we use Proposition 5.16 from (Vershynin, 2012), which gives upper bounds on the tails of centered sub-exponential random variables:

Lemma 3 ((Vershynin, 2012))**.** *Let $Y_1, \ldots, Y_k$ be independent centered sub-exponential random variables. i.e., there a constant $C > 0$ such that $\mathbb{E}[\exp(CY_i)] \leq e$. Then for any $a_1, \ldots, a_k \in \mathbb{R}$ and $R = a_1Y_1 + \cdots + a_kY_k$ we have*

$$\mathbb{P}\left[|R| \geq x\right] \leq 2\exp\left(-\tfrac{cx^2}{\|a\|_2^2}\right), \quad \forall 0 \leq x \leq \tfrac{\|a\|_2^2}{\|a\|_\infty}$$
$$\mathbb{P}[|R| \geq x] \leq 2\exp\left(-\tfrac{cx}{\|a\|_\infty}\right), \quad \forall x \geq \tfrac{\|a\|_2^2}{\|a\|_\infty}.$$

where $c > 0$ is an absolute constant.

Note that for a random variable $N \sim \mathcal{N}(0,1)$, its centered square (i.e. $N^2 - 1$) is a sub-exponential random variable. Therefore, we can apply Lemma 3 to $\sum_{i=1}^{k} q^{-1}Z_i N_i^2$ by rewriting as $\sum_{i=1}^{k} q^{-1}Z_i(N_i^2 - 1) + \sum_{i=1}^{k} q^{-1}Z_i$.

Examining Lemma 3, we see that we need two bounds on the Z_i's, one on $\sum_i Z_i^2$ and one on $\max_i |Z_i|$. Thus, we focus on giving bounds on these two quantities. For this, we will use that $u = HDx$ has all coordinates bounded in absolute value by $O(\sqrt{\ln(n)/d})$ as observed earlier. We then argue that the hardest such vector u, is one in which precisely m coordinates all take the value $m^{-1/2} = O(\sqrt{\ln(n)/d})$ and the remaining coordinates of u are all 0. This is also the hard vector analysed by Ailon and Chazelle. In their analysis, they simply bound $\sum_{i=1}^{k} Z_i^2$ by $k(\max_i |Z_i|)^2$ and this is where we improve over their work. Giving a tight analysis of $\sum_i Z_i^2$ is far from trivial and takes up the majority of Appendix A.

For now, we state the concentration inequalities we need and we provide the proofs in Appendix A.

Lemma 4. *For $i = 1, \dots, k$ let $Z_i = \sum_{j=1}^{d} u_j^2 b_{i,j}$ where the $b_{i,j}$'s are independent Bernoulli random variables with success probability q and the u_j^2's are positive real numbers upper bounded by $1/m$ and summing to 1. For $\alpha \leq 1/4$ it holds*

$$\mathbb{P}\left[\max_{i=1,\dots,k} Z_i > \frac{q}{2\alpha}\right] \leq k \exp\left(-\frac{mq\ln(1/\alpha)}{32\alpha}\right).$$

And to bound $\sum_i Z_i^2$, we show the following:

Lemma 5. *Let $Z_1, \dots, Z_k$ be i.i.d. random variables distributed as the Z_i's in Lemma 4. Then for any $t \geq 64 \cdot 24e^3q^2k$ and $q \geq 8/(em)$, we have:*

$$\mathbb{P}\left[\sum_{i=1}^{k} Z_i^2 > t\right] < 14\exp\left(-\frac{m\sqrt{t}\ln(\sqrt{t/2^3}/(eq))}{200\cdot 44\cdot 2^{\frac{5}{2}}}\right).$$

Before continuing, let us briefly argue that Lemma 5 is tighter than using the approach of Ailon and Chazelle where $\sum_i Z_i^2$ is merely bounded as $k(\max_i Z_i)^2$. For large enough t, Lemma 5 roughly gives that $\mathbb{P}[\sum_i Z_i^2 > t] < \exp(-m\sqrt{t}\ln(\sqrt{t}/q))$. If we instead bounded $\sum_i Z_i^2$ by $k(\max_i Z_i)^2$, then for any t, their approach would need $\max_i Z_i \leq \sqrt{t/k}$. Choosing α such that $\sqrt{t/k} = q/(2\alpha)$ and examining Lemma 4, we would roughly get $\mathbb{P}[\sum_i Z_i^2 > t] < k\exp(-(m(\sqrt{t/k})\ln((\sqrt{t/k})/q)))$. We would thus lose almost a factor $\sqrt{k}$ in the exponent. This is basically where our improvement comes from.

Since Lemma 5 does not capture all tradeoffs between ε, d and n that we need, we also prove the following lemma:

Lemma 6. *Let $Z_1, \dots, Z_k$ be i.i.d. random variables distributed as the Z_i's in Lemma 4, with $m = c_2d/\ln n$ and $k = c_1\varepsilon^{-2}\ln n$ and $q = c_1\varepsilon$, where $c_1 \geq 1/c_2$. For $\varepsilon \leq c_1^{-1}/(e4)$ and $t \geq 2c_1^3e^8\ln n$, we have that*

$$\mathbb{P}\left[\sum_{i=1}^{k} Z_i^2 > t\right] \leq 3n^{-4c_1}.$$

With the central lemmas laid out, we now give the full proof details of Theorem 1. We prove here the case of Normal entries in P, and the case of Rademacher variables is proved in Appendix C.

Proof of Theorem 1, Normal entries.

Proof. Let $m = c_2d/\ln n$ for a constant $c_2 > 0$, and let $k = c_1\varepsilon^{-2}\ln n$ be such that $c_1 \geq 1/c_2$. Let the success probabilities of the binomial random variables $b_{i,j}$ in P be

$$q = \max\left\{c_1/m, c_1\varepsilon\min\left\{1, \ln(n)/(m\ln(1/\varepsilon))\right\}\right\}.$$

Assume for now that $u \in \mathbb{R}^d$ is such that $u_i^2 \leq 1/m$ for all $i = 1, \dots, d$ and $\|u\|^2 = 1$. By construction of P and the 2-stability of the standard normal distribution

$$\|Pu\|^2 = \sum_{i=1}^{k}\left(\sum_{j=1}^{d}\sqrt{1/q}u_jb_{i,j}N_{i,j}\right)^2 \overset{d}{=} \sum_{i=1}^{k}\frac{1}{q}Z_iN_i^2,$$

where $Z_i = \sum_{j=1}^{d} u_j^2 b_{i,j}$ and N_i's are independent standard normal random variables. We first prove a bound on $\sum_{i=1}^{k} Z_i$. For this, notice that $\sum_{i=1}^{k} Z_i$ is a sum of independent random variables, where each Z_i is a sum of independent random variables with values between $[0, 1/m]$, and $\mathbb{E}[Z_i] = q$. In Appendix A, LemmaA1, we prove

$$\mathbb{P}\left[\sum_{i=1}^{k} Z_i \notin (1 \pm \varepsilon/4)qk\right] \leq 2\exp\left(-\frac{qmk\varepsilon^2}{48}\right),$$

which is bounded by $2n^{-c_1^2/48}$ when $q \geq c_1/m$ and $k = c_1\varepsilon^{-2}\ln n$. Therefore, $\sum_{i=1}^{k} Z_i \in (1 \pm \varepsilon/4)qk$ with probability at least $1 - 2n^{-c_1^2/48}$.

We continue with case analysis based on the value of q. Our goal is to show that $\|Pu\|^2 = \sum_{i=1}^{k} Z_iN_i^2/q \in (1 \pm \varepsilon/4)k$ with high probability (conditioned on u having bounded coordinates as remarked earlier).

CASES $q = c_1/m$ AND $q = c_1\varepsilon\ln(n)/(m\ln(1/\varepsilon))$.

In both these cases we have $q \geq c_1\varepsilon\ln(n)/(m\ln(1/\varepsilon))$ (due to the $\max$ in the definition of q and $c_1/m \leq c_1\varepsilon$). By Lemma 4, taking $\alpha = \varepsilon$ we have that $\|Z\|_\infty \leq q/(2\varepsilon)$ with probability at least $1 - k\exp(-(mq\ln(1/\varepsilon))/(32\varepsilon)) \geq 1 - n^{-c_1/32+1}$, because $mq\ln(1/\varepsilon)/\varepsilon \geq c_1\ln n$ and $k \leq n$.

Since $q \geq c_1/m$, by Lemma 5 with $t := 64 \cdot 24e^3q^2k$, and since $q \geq c_1\varepsilon\ln(n)/(m\ln(1/\varepsilon))$ we conclude that $\|Z\|^2 \leq$

$64 \cdot 24 \cdot e^3 q^2 k$ with probability at least

$$1 - 14 \exp \left(- \frac{m\sqrt{64 \cdot 24 \cdot e^3 q^2 k} \ln(\frac{\sqrt{(64 \cdot 24 \cdot e^3 q^2 k)/(2^3)}}{eq})}{200 \cdot 44 \cdot 2^{\frac{5}{2}}} \right)$$
$$\geq 1 - 14 \exp \left(- \frac{(c_1 \ln(n))^{3/2} \ln \left(22\sqrt{k}\right)}{300 \ln (1/\varepsilon)} \right)$$
$$\geq 1 - 14 n^{-c_1^{3/2}/300}.$$

Hence, in these cases we have that $\sum_{i=1}^k Z_i \in (1 \pm \varepsilon/4)qk$, $\|Z\|_\infty \leq \frac{q}{2\varepsilon}$ and $\|Z\|^2 \leq 64 \cdot 24 e^3 q^2 k$ with probability at least $1 - 17n^{-c_1/300+1}$. We call such outcomes of the variables Z_i *desirable*.

Therefore, for desirable outcomes of the Z_i's, we have from Lemma 3 that if $x := (\varepsilon/4) \sum_{i=1}^k Z_i \geq \|Z\|^2/\|Z\|_\infty$, then (with probability over the N_i's)

$$\mathbb{P}\left[\frac{1}{q} \sum_{i=1}^k Z_i N_i^2 \notin (1 \pm \varepsilon/4) \frac{1}{q} \sum_{i=1}^k Z_i \right]$$
$$\leq 2 \exp \left(- \frac{c(\varepsilon/4)\sum_{i=1}^k Z_i}{\|Z\|_\infty} \right)$$
$$\leq 2 \exp \left(- \frac{c\varepsilon qk/8}{q/(2\varepsilon)} \right) = 2n^{-cc_1/4},$$

where we used that $\sum_{i=1}^k Z_i \geq (1 - \varepsilon/4)qk$ for desirable outcomes, and that $k = c_1 \varepsilon^{-2} \ln n$. In addition, if $(\varepsilon/4) \sum_{i=1}^k Z_i \leq \|Z\|^2/\|Z\|_\infty$, then by Lemma 3 (and using $\varepsilon < 1$):

$$\mathbb{P}\left[\sum_{i=1}^k \frac{1}{q} N_i^2 Z_i \notin (1 \pm \varepsilon/4) \sum_{i=1}^k \frac{1}{q} Z_i \right]$$
$$\leq 2 \exp \left(- \frac{c((\varepsilon/4)\sum_{i=1}^k Z_i)^2}{\|Z\|^2} \right)$$
$$\leq 2 \exp \left(- \frac{c\varepsilon^2 q^2 k^2}{16 \cdot 64 \cdot 96 e^3 q^2 k} \right) = 2n^{-cc_1/(16 \cdot 64 \cdot 96 e^3)}.$$

Therefore, for desirable outcomes of the Z_i's, for $r_1 := c/(16 \cdot 64 \cdot 96e^3)$, it holds (with probability over the N_i's):

$$1 - 2n^{-r_1 c_1} \leq \mathbb{P}\left[\sum_{i=1}^k \frac{1}{q} N_i^2 Z_i \in (1 \pm \varepsilon/4) \sum_{i=1}^k \frac{1}{q} Z_i \right]$$
$$\leq \mathbb{P}\left[\sum_{i=1}^k \frac{1}{q} N_i^2 Z_i \in (1 \pm \varepsilon)k \right].$$

Since Z_i's and N_i's are independent, it follows that with probability at least $(1 - 2n^{-r_1 c_1}) \cdot (1 - 17n^{-c_1/300+1}) \geq 1 - 34n^{-\min\{r_1, 1/300\}c_1+1}$ it holds that $\sum_{i=1}^k N_i^2 Z_i/q \in (1 \pm \varepsilon)k$, as required.

CASE $q = c_1\varepsilon$.

In this case (we assume that $\varepsilon < c_1^{-1}/(4e)$) from Lemma 6, taking $t := 2c_1^3 e^8 \ln n$ it follows that $\|Z\|^2 \leq 2c_1^3 e^8 \ln n$ with probability at least $1 - 3n^{-4c_1}$. Therefore, with probability at least $1 - 5n^{-c_1/48}$ it holds that $\sum_{i=1}^k Z_i \in (1 \pm \varepsilon/4)qk$ and $\|Z\|^2 \leq 2c_1^3 e^8 \ln n$, which we call as *desirable* outcomes of Z_i's.

Similarly to the analysis in the previous case, using Lemma 3 for desirable outcomes of Z_i's it holds that if $(\varepsilon/4) \sum_{i=1}^k Z_i \geq \|Z\|^2/\|Z\|_\infty$, then with probability over the N_i's, and using the trivial bound that the Z_i's are at most 1, it follows that

$$\mathbb{P}\left[\sum_{i=1}^k \frac{1}{q} N_i^2 Z_i \notin (1 \pm \varepsilon/4) \sum_{i=1}^k \frac{1}{q} Z_i \right]$$
$$\leq 2 \exp \left(- \frac{c(\varepsilon/4)\sum_{i=1}^k Z_i}{\|Z\|_\infty} \right)$$
$$\leq 2 \exp \left(- \frac{c\varepsilon qk}{8} \right) = 2n^{-cc_1^2/8},$$

where the last inequality follows from $\sum_{i=1}^k Z_i \geq (1 - \varepsilon/4)qk \geq qk/2$ and the equality follows from $\varepsilon qk = c_1^2 \ln n$. And if $(\varepsilon/4) \sum_{i=1}^k Z_i \leq \|Z\|^2/\|Z\|_\infty$, Lemma 3 yields:

$$\mathbb{P}\left[\sum_{i=1}^k \frac{1}{q} N_i^2 Z_i \notin (1 \pm \varepsilon/4) \sum_{i=1}^k \frac{1}{q} Z_i \right]$$
$$\leq 2 \exp \left(- \frac{c((\varepsilon/4)\sum_{i=1}^k Z_i)^2}{\|Z\|^2} \right)$$
$$\leq 2 \exp \left(- \frac{c\varepsilon^2 q^2 k^2}{128 c_1^3 e^8 \ln n} \right) \leq 2n^{-cc_1/(128e^8)},$$

where the last inequality follows from $\varepsilon^2 q^2 k^2 / \ln n = c_1^4 \varepsilon^4 \ln^2(n)/(\varepsilon^4 \ln n) \geq c_1^4 \ln\ n$.

Letting $r_2 = c/(128e^8)$ we conclude that for desirable outcomes of the Z_i's, with probability

$$1 - 2n^{r_2 c_1} \leq \mathbb{P}\left[\sum_{i=1}^k \frac{1}{q} N_i^2 Z_i \in (1 \pm \varepsilon/4) \sum_{i=1}^k \frac{1}{q} Z_i) \right]$$
$$\leq \mathbb{P}\left[\sum_{i=1}^k \frac{1}{q} N_i^2 Z_i \in (1 \pm \varepsilon)k \right].$$

Using the independence of the Z_i's and N_i's, we get that $\sum_{i=1}^k N_i^2 Z_i/q \in (1 \pm \varepsilon)k$ holds with probability at least $(1 - 2n^{-r_2 c_1})(1 - 5n^{-c_1/48}) \geq 1 - 10n^{-\min\{r_2, 1/48\}c_1}$.

CONCLUDING THE PROOF

By a similar argument to (Ailon & Chazelle, 2009) in Equation (4), with probability at least $1 - 1/(2n^3)$ it holds that

$u_i^2 = (HDx)_i^2 \leq \ln(n)/(c_2 d) = 1/m$ for all $i = 1, \ldots, d$ simultaneously, when c_2 is small enough (assuming $d \leq n$ such that $\ln d = O(\ln n)$), thus we have $u_i^2 \leq 1/m$ as required.

Therefore, for all cases it suffices to set c_1 as a sufficiently large constant so with probability at least $1 - 1/(2n^3)$, $\|Pu\|^2 = \sum_{i=1}^{k} \frac{1}{q} N_i^2 Z_i \in (1 \pm \varepsilon)k$. Since D and P are independent, with probability at least $(1 - 1/(2n^3))(1 - 1/(2n^3)) \geq 1 - 1/n^3$, it holds $k^{-1}\|PHDx\|^2 \in (1 \pm \varepsilon)$.

To complete the proof of the theorem, we union bound over all vectors $z/\|z\|$ where $z = x - y$ with $x, y \in X$, to conclude that with probability at least $1 - 1/n$, $k^{-1/2}PHD$ satisfies (1).

It remains to claim that the choice of

$$q = \max\left\{c_1/m, c_1\varepsilon \min\left\{1, \ln(n)/(m\ln(1/\varepsilon))\right\}\right\},$$

is equivalent to that claimed in Theorem 1. Recalling that $m = O(d/\ln n)$, implies that our choice of q is $O(\max\{(\ln n)/d, \varepsilon \min\{1, \ln^2(n)/(d\ln(1/\varepsilon))\}\})$. Since $(\ln n)/d \leq (\ln n)/k = O(\varepsilon^2) = O(\varepsilon)$, we can never have $(\ln n)/d = \omega(\varepsilon)$ and hence we can move the $\max$ into the $\min$

$$q = O\left(\min\left\{\varepsilon, \frac{\ln n}{d} \cdot \max\left\{1, \frac{\varepsilon \ln n}{\ln(1/\varepsilon)}\right\}\right\}\right).$$

This completes the proof of Theorem 1. □

DISCUSSION OF EXPRESSION

Let us conclude by giving some more intuition on where the different terms in the expression for q originate from. Recall from above that the hardest vector for $k^{-1/2}P$ is a unit vector u with $m = O(d/\ln n)$ non-zero entries, each of magnitude $m^{-1/2}$. Also recall that each entry of P is the product of a Bernoulli $b_{i,j}$ with success probability q and a normal distributed random variable with variance $1/q$.

The term $\ln(n)/d$ in the expression for q intuitively comes from the following: There is a total of km Bernoulli random variables $b_{i,j}$ that are each multiplied with the same non-zero value u_j^2. This gives an expected kmq of them that are non-zero. Intuitively, since they are all multiplied with the same coefficient, we need the number of non-zero Bernouillis to be within εkmq of the expectation. A binomial distribution with km trials and success probability q deviates from its expectation by $\Omega(\sqrt{kmq \ln n})$ with probability $n^{-1/2}$ and thus we require $\sqrt{kmq \ln n} < \varepsilon kmq$. This implies that we must set $q > \ln(n)/(\varepsilon^2 mk) = \Omega(1/m) = \Omega(\ln(n)/d)$.

The terms $\varepsilon \ln^2 n/(d\ln(1/\varepsilon))$ and ε in the expression for q come from the event that the square of the first coordinate, $(k^{-1/2}Pu)_1^2$ is larger than ε (which causes a distortion if the rest of the coordinates are concentrated). Conditioned on the Bernoullis $b_{1,j}$, the square of the first coordinate is the square of a normal distributed random variable. Hence it is a factor $\Omega(\ln n)$ larger than its variance with probability $n^{-1/2}$. There are now two cases: 1. $m < c\ln_{1/q} n$ for a small constant $c > 0$, and 2., $m \geq c\ln_{1/q} n$.

In the first case, $m < c\ln_{1/q} n$, it happens with probability at least $n^{-1/2}$ that all Bernoullis $b_{1,j}$ that are multiplied with a non-zero coefficient take the value 1. In that case, the first coordinate of $k^{-1/2}Pu$ is normal distributed with mean zero and variance $1/(qk)$ (since $\sum_j u_j^2 = 1$). We thus need $\ln n/(qk) < \varepsilon$. Using that $k = \Theta(\varepsilon^{-2}\ln n)$, this means we have to set $q = \Omega(\varepsilon)$.

In the second case, $m \geq c\ln_{1/q} n$, we expect to see qm non-zero Bernoullis $b_{1,j}$ that are each multiplied with $1/m$ for the first coordinate of $k^{-1/2}Pu$. However, by a "reverse" Chernoff bound, with probability at least $n^{-1/2}$, we see at least $c\ln_{1/q} n$ non-zero Bernoullis. In that case, the first coordinate of $k^{-1/2}Pu$ is normal distributed with mean zero and variance $\Theta((\ln_{1/q} n)/(mqk)) = \Theta(\varepsilon^2 \ln_{1/q}(n)/(dq))$. Since the square of the first coordinate was a factor $\ln n$ larger than its variance with probability $n^{-1/2}$, we hence need $\varepsilon^2 \ln n \ln_{1/q}(n)/(dq) = O(\varepsilon)$. If we for simplicity approximate q by ε in $\ln_{1/q} n$, this gives precisely $q = \Omega(\varepsilon \ln^2 n/(d\ln(1/\varepsilon)))$.

4. Lower Bound

In this section we give a sketch of the proof of Theorem 2. The complete proof is in Appendix B. In particular, we give an example of a unit vector $x \in \mathbb{R}^d$, such that one must have

$$q = \Omega\left(\min\left\{\varepsilon, \frac{\ln(1/\delta)}{d} \cdot \max\left\{1, \frac{\varepsilon \ln(1/\delta)}{\ln(1/\varepsilon)}\right\}\right\}\right),$$

to guarantee $\mathbb{P}[\|k^{-1/2}PHDx\| \in (1 \pm \varepsilon)] \geq 1 - \delta$, where P is populated with Normal variables.

The proof consists of two steps. In the first step, we show that we must have $q = \Omega(\ln(1/\delta)/d)$. In the second step, we use the result from step one to conclude that q must also be $\Omega(\varepsilon \min\{1, \ln^2(1/\delta)/(d\ln(1/\varepsilon))\})$. Combining the two, we have:

$$q = \Omega\left(\max\{\ln(1/\delta)/d, \varepsilon \min\{1, \ln^2(1/\delta)/(d\ln(1/\varepsilon))\}\}\right).$$

Noticing that it is always $\ln(1/\delta)/d = O(\ln(1/\delta)/k) = O(\varepsilon^2) = O(\varepsilon)$, we can move the max inside the min and obtain the bound claimed above.

In both steps, we use the same hard instance vector x. It has the property that with probability at least δ^c for a small constant $c > 0$, $u = HDx$ has $m = \Theta(d/\ln(1/\delta))$ non-zero entries, each of magnitude $1/\sqrt{m}$. Conditioning on such a transformed vector $u = HDx$ puts a lot of structure on u, which simplifies the analysis of the product Pu. Indeed, if we consider a coordinate $(Pu)_i$, then this coordinate is $\mathcal{N}(0, \sum_j b_{i,j}u_j^2/q)$ distributed if we condition on the Bernoullis $b_{i,j}$. But u_j^2 is $1/m$ for precisely m values of j and 0 for all others. Thus $\sum_j b_{i,j}u_j^2/q$ is distributed as $1/(qm)$ times a binomial distribution with m trials and success probability q. One part of the analysis is thus to study this distribution. Secondly, if we consider $\|Pu\|^2$, then this is a linear combination of k independent χ^2 random variables, with the i'th being scaled by $\sum_j b_{i,j}u_j^2/q$. Hence we also need to understand the tail of such a distribution.

For the first step, i.e. showing $q = \Omega(\ln(1/\delta)/d)$, we argue that the sum of the coefficients $\sum_j b_{i,j}u_j^2/q$ deviates a lot from its expectation with reasonable probability. More precisely, notice that $\mathbb{E}[\sum_j b_{i,j}u_j^2/q] = (mq)/(mq) = 1$ and thus $\mathbb{E}[\sum_i \sum_j b_{i,j}u_j^2/q] = k$. But the sum of these coefficients is itself distributed as $1/(mq)$ times a binomial distribution with mk trials and success probability q. The number of successes in such a binomial distribution deviates by additive $\Omega(\sqrt{\ln(1/\delta)(mkq)})$ from its expectation mkq with probability at least δ^c for a small constant $c > 0$. Intuitively, we need this deviation to be less than εmkq to preserve the norm of x (and thus u) to within $(1 \pm \varepsilon)$. This implies $\sqrt{\ln(1/\delta)(mkq)} = O(\varepsilon mkq) \Rightarrow q = \Omega(\ln(1/\delta)/(\varepsilon^2 mk)) = \Omega(1/m) = \Omega(\ln(1/\delta)/d)$.

In the second step, we now use the fact that we know that q is sufficiently large, such that coordinates $2, \dots, k$ of Pu are reasonably well concentrated around their mean. What establishes the second lower bound on q, namely $q = \Omega(\varepsilon \min\{1, \ln^2(1/\delta)/(d\ln(1/\varepsilon))\})$, is the possibility that the first coordinate $(Pu)_1$ may be so large that it alone distorts the norm $\|k^{-1/2}Pu\|^2$. In more detail, we show that with good probability, we have $\sum_{i=2}^k k^{-1}(Pu)_i^2 \in (1 \pm \varepsilon)(k-1)/k$, i.e. on the last $k-1$ coordinates, the embedding $k^{-1/2}PHDx$ preserves the norm of x as it should (we work with $k^{-1}\|Pu\|^2$ instead of $k^{-1/2}\|Pu\|$ to simplify the analysis - and since the later is a weaker statement by $\sqrt{1 \pm \varepsilon} \subset (1 \pm \varepsilon)$ it suffices to work with $k^{-1}\|Pu\|^2$). In this case, we show that unless q is large enough, the single coordinate $k^{-1}(Pu)_1$ contributes more than ε to $k^{-1}\|Pu\|^2$ with probability more than δ.

In what follows, we show the existence of the vector x for which $u = HDx$ often has $m = O(d/\ln(1/\delta))$ coordinates of magnitude $1/\sqrt{m}$, and the two steps of the proof are given in Appendix B.

The hard instance. For $\varepsilon, \delta > 0$ set l be the integer such that $l \le \lg_2\left(\lg_2(1/\sqrt{2\delta})\right) \le l+1$ and define $x_i = \sqrt{\frac{1}{2^l}}$ for all $i \le 2^l$, and $x_i = 0$ for rest coordinates.

Notice that $Dx = x$ with probability $2^{-2^l} \ge \sqrt{2\delta}$. Multiplying the Hadamard matrix H_d with the vector x results in $Hx = [H_{2^l}\mathbf{1}, \dots, H_{2^l}\mathbf{1}]^T/\sqrt{d}$, where H_{2^l} is the unnormalized Hadamard matrix of size $2^l \times 2^l$, and $\mathbf{1}$ is all-ones vector in $\mathbb{R}^{2^l}$. Since the rows of the Hadamard matrix are orthogonal, and its first row is all-ones, it follows that

$$(Hx)_i = \begin{cases} \sqrt{\frac{2^l}{d}} \text{ , for } i \equiv 0 \mod (2^l) \\ 0 \text{ , otherwise} \end{cases}$$

Therefore, $u := Hx$ has $d/2^l$ non-zero entries, all of value $\sqrt{2^l/d}$. This is the vector u we will analyze throughout the remainder of the lower bound proof.

Using the definition of u we have that

$$\begin{aligned} \|Pu\|^2 &\overset{d}{=} \sum_{i=1}^{k}\left(\sum_{j=1}^{\frac{d}{2^l}} \sqrt{\frac{2^l}{dq}} b_{i,j} N_{i,j}\right)^2 \\ &= \frac{2^l}{dq}\sum_{i=1}^{k}\left(\sum_{j=1}^{\frac{d}{2^l}} b_{i,j} N_{i,j}\right)^2 , \end{aligned}$$

where the $b_{i,j}$'s are Bernoulli random variables with success probability q and the $N_{i,j}$'s are $\mathcal{N}(0,1)$ distributed, all independent of each other. Conditioned on the outcome of the $b_{i,j}$'s it follows from 2-stability of normal distribution

$$\sum_{i=1}^{k}\left(\sum_{j=1}^{\frac{d}{2^l}} b_{i,j} N_{i,j}\right)^2 \overset{d}{=} \sum_{i=1}^{k} b_i N_i^2 ,$$

where the b_i's are $\sum_{j=1}^{d/2^l} b_{i,j}$ and the N_i's are independent standard normal random variables.

5. Conclusion

In this paper we studied the embedding time of the classic Fast JL transform (Ailon & Chazelle, 2009). We showed that this famous algorithm is in fact faster than was proven more than two decades ago in the original paper. In particular, we carefully analyzed the sparsity parameter q in the sparse matrix P of the PHD embedding construction.

We showed that q can be decreased by a factor of $\Omega(\min\{\varepsilon^{-1}\ln(1/\varepsilon), \ln n\})$, resulting in the similar improvement in the $O(k \ln^2 n)$ term of the original embedding time $O(d \lg d + k \ln^2 n)$. Moreover, for the case of $\varepsilon = O(\ln(\ln n)/\ln n)$ our bound simplifies to $O(d \ln d + k \ln n)$ which is a strict improvement of the Kac JL bound in Equation (2) in this regime. It is an interesting open question to investigate the optimality of the Kac JL transform in the regime of larger ϵ's.

We complimented our analysis with the lower bound, effectively showing that the the upper bound on q cannot be further lowered down, if one is about to use the PHD construction together with the union bound for obtaining distortion guarantees on pairwise distances. However, this does not rule out possible further improvements in the Fast JL embedding time using a different argument from the standard union bound. We leave this as an intriguing open question for future research.

As fast dimensionality reduction is widely used across many applied communities, we believe its tight analysis will be beneficial for practitioners.

Acknowledgements

Supported by Independent Research Fund Denmark (DFF) Sapere Aude Research Leader grant No 9064-00068B.

References

Ailon, N. and Chazelle, B. The fast johnson–lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.*, 39:302–322, 2009.

Ailon, N. and Liberty, E. Fast dimension reduction using rademacher series on dual BCH codes. In Teng, S. (ed.), *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2008, San Francisco, California, USA, January 20-22, 2008*, pp. 1–9. SIAM, 2008. URL http://dl.acm.org/citation.cfm?id=1347082.1347083.

Alon, N. and Klartag, B. Optimal compression of approximate inner products and dimension reduction. In Umans, C. (ed.), *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pp. 639–650. IEEE Computer Society, 2017. doi: 10.1109/FOCS.2017.65. URL https://doi.org/10.1109/FOCS.2017.65.

Bamberger, S. and Krahmer, F. Optimal fast johnson–lindenstrauss embeddings for large data sets. *Sampling Theory, Signal Processing, and Data Analysis*, 19(1):3, 2021. doi: 10.1007/s43670-021-00003-5. URL https://doi.org/10.1007/s43670-021-00003-5.

Clarkson, K. L. and Woodruff, D. P. Low rank approximation and regression in input sparsity time. In Boneh, D., Roughgarden, T., and Feigenbaum, J. (eds.), *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pp. 81–90. ACM, 2013. doi: 10.1145/2488608.2488620. URL https://doi.org/10.1145/2488608.2488620.

Do, T. T., Gan, L., Chen, Y., Nguyen, N., and Tran, T. D. Fast and efficient dimensionality reduction using structurally random matrices. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1821–1824, 2009. doi: 10.1109/ICASSP.2009.4959960.

Freksen, C., Kamma, L., and Larsen, K. G. Fully understanding the hashing trick. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 5394–5404, Red Hook, NY, USA, 2018. Curran Associates Inc.

Freksen, C. B. and Larsen, K. G. On using toeplitz and circulant matrices for johnson-lindenstrauss transforms. *Algorithmica*, 82(2):338–354, 2020. URL https://doi.org/10.1007/s00453-019-00644-y.

Hinrichs, A. and Vybíral, J. Johnson-lindenstrauss lemma for circulant matrices**. *Random Structures & Algorithms*, 39(3):391–398, 2011. doi: https://doi.org/10.1002/rsa.20360. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20360.

Indyk, P. and Motwani, R. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, pp. 604–613, New York, NY, USA, 1998. Association for Computing Machinery. ISBN 0897919629. doi: 10.1145/276698.276876. URL https://doi.org/10.1145/276698.276876.

Jagadeesan, M. Understanding sparse JL for feature hashing. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019*, pp. 15177–15187, 2019.

Jain, V., Pillai, N. S., and Smith, A. Kac meets johnson and lindenstrauss: a memory-optimal, fast johnson-lindenstrauss transform. *CoRR*, abs/2003.10069, 2020. URL https://arxiv.org/abs/2003.10069. To appear in Annals of Applied Probability.

Johnson, W. B. and Lindenstrauss, J. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26:28, 1984.

Kac, M. Foundations of kinetic theory. In *Proceedings of The third Berkeley symposium on mathematical statistics and probability*, pp. 171–197. University of California Press Berkeley and Los Angeles, California, 1958.

Kane, D. M. and Nelson, J. Sparser johnson-lindenstrauss transforms. *J. ACM*, 61(1):4:1–4:23, 2014. doi: 10.1145/2559902. URL https://doi.org/10.1145/2559902.

Krahmer, F. and Ward, R. New and improved johnson-lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.

doi: 10.1137/100810447. URL https://doi.org/10.1137/100810447.

Larsen, K. G. and Nelson, J. Optimality of the johnson-lindenstrauss lemma. In Umans, C. (ed.), *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pp. 633–638. IEEE Computer Society, 2017. doi: 10.1109/FOCS.2017.64. URL https://doi.org/10.1109/FOCS.2017.64.

Manning, C. D. and Schütze, H. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts, 1999. URL http://nlp.stanford.edu/fsnlp/.

Matoušek, J. On variants of the johnson–lindenstrauss lemma. *Random Structures & Algorithms*, 33(2): 142–156, 2008. doi: https://doi.org/10.1002/rsa.20218. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20218.

Mousavi, N. How tight is the chernoff bound? https://ece.uwaterloo.ca/~nmousavi/Papers/Chernoff-Tightness.pdf, 2010.

Nelson, J. and Nguyen, H. L. Sparsity lower bounds for dimensionality reducing maps. In Boneh, D., Roughgarden, T., and Feigenbaum, J. (eds.), *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pp. 101–110. ACM, 2013. doi: 10.1145/2488608.2488622. URL https://doi.org/10.1145/2488608.2488622.

Pólya, G. Remarks on computing the probability integral in one and two dimensions. *Statistical Laboratory of the University of California*, 1949.

Rigollet, P. and Hütter, J.-C. High dimensional statistics. *Lecture notes for course 18S997*, 813(814):46, 2015.

Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. In Eldar, Y. C. and Kutyniok, G. (eds.), *Compressed Sensing*, pp. 210–268. Cambridge University Press, 2012. doi: 10.1017/cbo9780511794308.006. URL https://doi.org/10.1017/cbo9780511794308.006.

Vybiral, J. A variant of the johnson-lindenstrauss lemma for circulant matrices. *Journal of Functional Analysis*, 260:1096–1105, 02 2010. doi: 10.1016/j.jfa.2010.11.014.

Weinberger, K. Q., Dasgupta, A., Langford, J., Smola, A. J., and Attenberg, J. Feature hashing for large scale multitask learning. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML 2009, Montreal, Quebec, Canada, June 14-18, 2009*, pp. 1113–1120, 2009.

Zhang, A. R. and Zhou, Y. On the non-asymptotic and sharp lower tail bounds of random variables. *Stat*, 9(1): e314, 2020. doi: https://doi.org/10.1002/sta4.314. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sta4.314. e314 sta4.314.

A. Concentration Inequalities Lemmas

The following lemma and its proof is similar to Lemma 2.3 in (Ailon & Chazelle, 2009), we reprove it here for completeness.

Lemma A1. *For $i = 1, \dots, k$ let $Z_i = \sum_{j=1}^{d} u_j^2 b_{i,j}$ where the $b_{i,j}$'s are independent Bernoulli random variables with success probability q and the u_j^2's are positive real numbers bounded by $1/m$ and summing to 1. It holds that*

$$\mathbb{P}\left[\sum_{i=1}^{k} mZ_i \notin (1 \pm \varepsilon)mqk\right] \leq 2\exp(-mqk\varepsilon^2/3).$$

Proof. For any $c > 0$ we have

$$\mathbb{E}[\exp(cZ_1)] = \sum_{b' \in \{0,1\}^d} \exp\left(c \sum_{j=1}^{d} u_j^2 b'_{1,j}\right) \mathbb{P}\left[b'\right],$$

where $\sum_{j=1}^{d} u_j^2 b'_{i,j}$ is an convex function in $(u_1^2, \dots, u_d^2)$, implying that $\exp(\sum_{j=1}^{d} u_j^2 b'_{i,j})$ is also convex (as a composition of the convex function $\sum_{j=1}^{d} u_j^2 b'_{i,j}$ and the increasing convex function $\exp(\cdot)$). Further, $\mathbb{E}[\exp(cZ_1)]$ is a convex function in $(u_1^2, \dots, u_d^2)$ as a linear combination with positive scalars of convex functions. Since the point $(u_1^2, \dots, u_d^2)$ lies in the set $\{x \in \mathbb{R}^d | x_i \in [0, 1/m], \sum_{i=1}^{d} x_i = 1\}$ (which is a convex polytope), the function $\mathbb{E}[\exp(cZ_1)]$ obtains its maximum on a vertex of the polytope. The choice of vertex does not change the distribution of the random variable, therefore without loss of generality we may assume that $u_1^2, \dots, u_m^2 = 1/m$ and $u_{m+1}^2, \dots, u_d^2 = 0$. Let μ denote mqk and let $\lambda > 0$, then

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^{k} mZ_i \geq (1+\varepsilon)mqk\right] = \mathbb{P}\left[\sum_{i=1}^{k} mZ_i \geq (1+\varepsilon)\mu\right] &\leq \mathbb{E}\left[\exp(\lambda \sum_{i=1}^{k} mZ_i)\right] \exp(-(1+\varepsilon)\mu\lambda) \\ &= \mathbb{E}\left[\exp(\lambda m Z_1)\right]^k \exp(-(1+\varepsilon)\mu\lambda)\end{aligned}$$

where the last inequality follows by Z_i's being i.i.d.

Since

$$\begin{aligned}\mathbb{E}\left[\exp(\lambda m Z_1)\right] \leq \mathbb{E}\left[\exp\left(\lambda m \sum_{i=1}^{m} (1/m) b_{1,j}\right)\right] = \mathbb{E}\left[\exp(\lambda b_{1,1})\right]^m &= \left[\exp(\lambda)q + (1-q)\right]^m \\ &= \left[\exp\left(\lg\left[1 + (\exp(\lambda) - 1)q\right]\right)\right]^m \\ &= \exp((\exp(\lambda) - 1)mq)\end{aligned}$$

we get that

$$\mathbb{P}\left[\sum_{i=1}^{k} mZ_i \geq (1+\varepsilon)mqk\right] \leq \exp((\exp(\lambda) - 1)mqk)\exp(-(1+\varepsilon)\mu\lambda) = \exp((\exp(\lambda) - 1) - (1-\varepsilon)\lambda)\mu).$$

Setting $\lambda = \lg(1+\varepsilon)$, we get

$$\mathbb{P}\left[\sum_{i=1}^{k} mZ_i \geq (1+\varepsilon)mqk\right] \leq \exp((\varepsilon - (1+\varepsilon)\lg(1+\varepsilon))\,\mu),$$

and using $\varepsilon - (1+\varepsilon)\lg(1+\varepsilon) \leq -\epsilon^2/3$ for $0 < \varepsilon < 1$ we obtain

$$\mathbb{P}\left[\sum_{i=1}^{k} mZ_i \geq (1+\varepsilon)mqk\right] \leq \exp(-\varepsilon^2\mu/3). \tag{5}$$

Considering the case $\mathbb{P}\left[\sum_{i=1}^{k} mZ_i \leq (1-\varepsilon)mqk\right]$, let $\lambda < 0$ then

$$\mathbb{P}\left[\sum_{i=1}^{k} mZ_i \leq (1-\varepsilon)mqk\right] = \mathbb{P}\left[\sum_{i=1}^{k} mZ_i \leq (1-\varepsilon)\mu\right] \leq \mathbb{E}\left[\exp(\lambda\sum_{i=1}^{k} mZ_i)\right]\exp(-(1-\varepsilon)\mu\lambda)$$
$$= \mathbb{E}\left[\exp(\lambda m Z_1)\right]^k \exp(-(1-\varepsilon)\mu\lambda).$$

Similar estimations result in

$$\mathbb{P}\left[\sum_{i=1}^{k} mZ_i \leq (1-\varepsilon)mqk\right] = \mathbb{P}\left[\sum_{i=1}^{k} mZ_i \leq (1-\varepsilon)\mu\right] \leq \exp((\exp(\lambda)-1)mqk)\exp(-(1-\varepsilon)\mu\lambda)$$
$$\leq \exp((\exp(\lambda)-1-\lambda(1-\varepsilon))\mu).$$

Setting $\lambda = \lg(1-\varepsilon) < 0$, we obtain

$$\mathbb{P}\left[\sum_{i=1}^{k} mZ_i \leq (1-\varepsilon)mqk\right] \leq \exp((-\varepsilon-(1-\varepsilon)\lg(1-\varepsilon))\mu),$$

and since $-\varepsilon-(1-\varepsilon)\lg(1-\varepsilon) \leq -\varepsilon^2/2$ for $0 < \varepsilon < 1$ we get that

$$\mathbb{P}\left[\sum_{i=1}^{k} mZ_i \leq (1-\varepsilon)mqk\right] \leq \exp(-\varepsilon^2\mu/2). \tag{6}$$

Finally, by Equation (5) and Equation (6) we conclude that

$$\mathbb{P}\left[\sum_{i=1}^{k} mZ_i \notin (1\pm\varepsilon)mqk\right] \leq 2\exp(-mqk\varepsilon^2/3),$$

for $0 < \varepsilon < 1$, as claimed.

□

Next, we restate and prove Lemma 4.

Restatement of Lemma 4.

Lemma 4. *For $i = 1,\ldots,k$ let $Z_i = \sum_{j=1}^{d} u_j^2 b_{i,j}$ where the $b_{i,j}$'s are independent Bernoulli random variables with success probability q and the u_j^2's are positive real numbers upper bounded by $1/m$ and summing to 1. For $\alpha \leq 1/4$ it holds*

$$\mathbb{P}\left[\max_{i=1,\ldots,k} Z_i > \frac{q}{2\alpha}\right] \leq k\exp\left(-\frac{mq\ln(1/\alpha)}{32\alpha}\right).$$

Proof. First notice that by a union bound and Markov's inequality we have that for $c > 0$

$$\mathbb{P}\left[\max_{i=1,\ldots,k} Z_i > t\right] \leq k\mathbb{P}\left[Z_1 > t\right] \leq k\mathbb{E}\left[\exp(cZ_1)\right]\exp(-ct). \tag{7}$$

Since

$$\mathbb{E}[\exp(cZ_1)] = \sum_{b'\in\{0,1\}^d} \exp\left(\sum_{j=1}^{d} u_j^2 b'_{i,j}\right)\mathbb{P}\left[b'\right],$$

where $\sum_{j=1}^{d} u_j^2 b'_{i,j}$ is an convex function in $(u_1^2, \dots, u_d^2)$, implying that $\exp(\sum_{j=1}^{d} u_j^2 b'_{i,j})$ is convex as the composition of the convex function with the increasing convex function $\exp(\cdot)$. Since a linear combination with positive scalars of convex functions is again a convex function, we conclude that $\mathbb{E}[\exp(cZ_1)] = \sum_{b' \in \{0,1\}^d} \exp(\sum_{j=1}^{d} u_j^2 b'_{i,j}) \mathbb{P}[b']$ is a convex function in $(u_1^2, \dots, u_d^2)$. Since $(u_1^2, \dots, u_d^2) \in \{x \in \mathbb{R}^d | x_i \in [0, 1/m] \forall i \in 1, \dots, d$ and $\sum_{i=1}^{d} x_i = 1\}$ (which is a convex polytope), the function $\mathbb{E}[\exp(cZ_1)]$ obtains its maximum on a vertex of the polytope. The choice of vertex does not change the distribution of the random variable, so we can without loss of generality assume that $u_1^2, \dots, u_m^2 = 1/m$ and $u_{m+1}^2, \dots, u_d^2 = 0$.

Using that the maximum of $\mathbb{E}[\exp(cZ_1)]$ is attained in such a vertex, we obtain that

$$\begin{aligned}\mathbb{E}\left[\exp(cZ_1)\right] &\leq \mathbb{E}\left[\exp(\frac{c}{m}\sum_{i=1}^{m} b_{1,i})\right] = \left(\exp\left(\frac{c}{m}\right)q + (1-q)\right)^m \\ &\leq \exp\left(m\left(\exp\left(\frac{c}{m}\right)q - q\right)\right) = \exp\left(mq\left(\exp\left(\frac{c}{m}\right) - 1\right)\right),\end{aligned} \tag{8}$$

where the first equality follows from the bernoulli trailes $b_{1,i}$ being independent and identically distributed. The second inequality uses that $0 \leq (1+x) \leq \exp(x)$ for $x \in \mathbb{R}^+$. Now setting $c = m\ln(t/q)$ (for $t > q$) and using Equation (7) and Equation (8)

$$\mathbb{P}\left[\max_{i=1,\dots,k} Z_i > t\right] \leq k\mathbb{E}\left[\exp\left(cZ_1\right)\right]\exp\left(-ct\right) \leq k\exp\left(mq\left(\frac{t}{q} - 1\right) - mt\ln\frac{t}{q}\right).$$

Now setting $t = q/(2\alpha) > q$ we get that

$$\begin{aligned}\mathbb{P}\left[\max_{i=1,\dots,k} Z_i > t\right] &\leq k\exp\left(mq\left(\frac{1}{2\alpha} - 1 - \frac{1}{2\alpha}\ln\frac{1}{2\alpha}\right)\right) = \\ k\exp\left(\frac{mq}{2\alpha}\left(1 - 2\alpha - \ln\frac{1}{2\alpha}\right)\right) &\leq k\exp\left(-\frac{mq\ln(1/\alpha)}{32\alpha}\right),\end{aligned}$$

where we in the second inequality we used that $\alpha \leq 1/4$ so $(1 - 2\alpha - \ln(1/(2\alpha))) \leq -\ln(1/\alpha)/16$. □

Next we give the proof of Lemma 5. For this, we need the following technical lemma about linear combinations of independent Bernoulli random variables.

Lemma 7. *Let $Z = \sum_{j=1}^{d} u_j^2 b_j$ where b_j are independent Bernoulli random variables with success probability q and u_j^2 are positive real numbers bounded by $1/m$ and summing to 1. We then have for $t > q$:*

$$\mathbb{P}\left[Z > t\right] < \left(\frac{t}{eq}\right)^{-mt}.$$

Proof. The proof follows the proof steps in Lemma 4. For any $c \geq 0$, we have

$$\mathbb{E}\left[\exp\left(cZ\right)\right] \leq \exp\left(mq\left(\exp\left(\frac{c}{m}\right) - 1\right)\right).$$

Thus by Markov's, we have for $c > 0$

$$\mathbb{P}\left[Z > t\right] = \mathbb{P}\left[\exp\left(cZ\right) > \exp\left(ct\right)\right] \leq \exp\left(mq\left(\exp\left(\frac{c}{m}\right) - 1\right)\right)\exp\left(-ct\right) \leq \exp\left(mq\exp\left(\frac{c}{m}\right) - ct\right).$$

Setting $c = m\ln(t/q)$ gives

$$\mathbb{P}\left[Z > t\right] < \exp\left(\frac{mqt}{q} - mt\ln\frac{t}{q}\right) = \exp\left(mt - mt\ln\frac{t}{q}\right) = \exp\left(-mt\ln\frac{t}{eq}\right) = \left(\frac{t}{eq}\right)^{-mt}.$$

□

Restatement of Lemma 5.

Lemma 5. Let $Z_1, \dots, Z_k$ be i.i.d. random variables distributed as the Z_i's in Lemma 4. Then for any $t \geq 64 \cdot 24e^3q^2k$ and $q \geq 8/(em)$, we have:

$$\mathbb{P}\left[\sum_{i=1}^{k} Z_i^2 > t\right] < 14 \exp\left(-\frac{m\sqrt{t}\ln(\sqrt{t/2^3}/(eq))}{200 \cdot 44 \cdot 2^{\frac{5}{2}}}\right).$$

Proof. For simplicity we assume in the following that $\lg_2 k$ is an integer. For $j = 0, \dots, \lg_2(k)/2$, let E_j denote the event that there are at least $2^j/(j+1)^2$ indices i such that $Z_i^2 \geq t/(2^{j+3})$ and let E'_j denote the event that there are at least $k/(2^j(j+1)^2)$ indices i with $Z_i^2 \geq t2^{j-3}/k$. We claim that if $\sum_{i=1}^k Z_i^2 > t$, then one of the events E_j or E'_j must occur for some j. Before we prove this, we briefly motivate why we need the two separate events E_j and E'_j. If we had only defined the events E_j, but let j range all the way to $\lg_2 k$, then either $j = 0$ or $j = \lg_2 k$ term would dominate. The issue with this, is that the $(j+1)^2$ term is sub-optimal (i.e. non-constant) for $j = \lg_2 k$. One could simply try to remove the $1/(j+1)^2$ term, but this would not work as $\sum_j 2^j \cdot t/2^{j+3}$ is $\omega(t)$. Including $1/(j+1)^2$ is precisely used to guarantee that $\sum_j 2^j/(j+1)^2 \cdot t/2^{j+3} = O(t)$. For that reason, we define the events E'_j that will handle the case of many indices with small values.

Assume for the sake of contradiction that none of the events occur, then

$$\begin{aligned}
\sum_{i=1}^{k} Z_i^2 &\leq \sum_{i=1}^{k}\sum_{j=0}^{\infty} 1_{\{Z_i^2 \geq \frac{t}{2^{j+3}}\}} \frac{t}{2^{j+3}} = \sum_{j=0}^{\infty} \frac{t}{2^{j+3}} \sum_{i=1}^{k} 1_{\{Z_i^2 \geq \frac{t}{2^{j+3}}\}} \\
&= \sum_{j=0}^{\lg_2 k} \frac{t}{2^{j+3}} \sum_{i=1}^{k} 1_{\{Z_i^2 \geq \frac{t}{2^{j+3}}\}} + \sum_{j=\lg_2 k+1}^{\infty} \frac{t}{2^{j+3}} \sum_{i=1}^{k} 1_{\{Z_i^2 \geq \frac{t}{2^{j+3}}\}} \\
&\leq \sum_{j=0}^{\lg_2(k)/2} \frac{t}{2^{j+3}} \sum_{i=1}^{k} 1_{\{Z_i^2 \geq \frac{t}{2^{j+3}}\}} + \sum_{j=0}^{\lg_2(k)/2} \frac{t}{2^{\lg_2 k - j + 3}} \sum_{i=1}^{k} 1_{\{Z_i^2 \geq t/2^{\lg_2 k - j + 3}\}} + \sum_{j=\lg_2 k+1}^{\infty} \frac{tk}{2^{j+3}} \\
&\leq \sum_{j=0}^{\lg_2(k)/2} \frac{t2^j}{2^{j+3}(j+1)^2}) + \sum_{j=0}^{\lg_2(k)/2} \frac{t2^{j-3}}{k} \sum_{i=1}^{k} 1_{\{Z_i^2 \geq \frac{t2^{j-3}}{k}\}} + \frac{t}{8} \\
&\leq \frac{t}{8} \sum_{j=0}^{\lg_2(k)/2} \frac{1}{(j+1)^2} + \sum_{j=0}^{\lg_2(k)/2} \frac{kt2^{j-3}}{k(2^j(j+1)^2)} + \frac{t}{8} \\
&\leq \frac{t}{4} \sum_{j=0}^{\infty} \frac{1}{(j+1)^2} + \frac{t}{8} = \frac{t\pi^2}{4 \cdot 6} + \frac{t}{8} < t.
\end{aligned}$$

Therefore, we have that $\mathbb{P}[\sum_{i=1}^k Z_i^2 > t] \leq \sum_{j=0}^{\lg_2(k)/2} \mathbb{P}[E_j] + \mathbb{P}[E'_j]$. To bound $\mathbb{P}[E_j]$, let S be any subset of $2^j/(j+1)^2$ indices in $[k]$ and define the event $E_{j,S}$ which happens when all $i \in S$ satisfy $Z_i^2 \geq t/(2^{j+3})$. Notice since $t \geq 64 \cdot 24e^3q^2k$ and $j \leq \lg_2(k)/2$ we have $t/2^{j+3} \geq 64 \cdot 24e^3q^2k/(8k^{1/2}) \geq 64 \cdot 3e^3q^2k^{1/2}$ implying that the ratio of $\sqrt{t/2^{j+3}}$ with q is larger than 1, Lemma 7 is applicable with $Z \geq \sqrt{t/2^{j+3}}$. Using a union bound over the events $E_{j,S}$ for any such set S, and that the Z_i's on such sets are independent and identically distributed, combined with Lemma 7 yields that,

$$\mathbb{P}\left[E_j\right] \leq \sum_{S} \mathbb{P}\left[E_{j,S}\right] \leq \binom{k}{2^j/(j+1)^2} \left(\sqrt{t/2^{j+3}}/\left(eq\right)\right)^{-m\sqrt{t/2^{j+3}}2^j/(j+1)^2},$$

and bounding $\binom{k}{2^j/(j+1)^2}$ by $k^{2^j/(j+1)^2}$, we obtain

$$\mathbb{P}\left[E_j\right] \leq \exp\left(-\frac{2^j\left(m\sqrt{t/2^{j+3}}\ln\left(\sqrt{t/2^{j+3}}/\left(eq\right)\right) - \ln k\right)}{\left(j+1\right)^2}\right).$$

For $t \geq 8e^2kq^2$ and $j \leq \lg_2(k)/2$, we have $\sqrt{t/2^{j+3}}/(eq) \geq \sqrt{8e^2kq^2/(8\sqrt{k}e^2q^2))} \geq k^{1/4}$ and thus it follows that $\ln(\sqrt{t/2^{j+3}}/(eq)) \geq \ln(k)/4$. Using $q \geq 8/(em)$ we also have $m\sqrt{t/2^{j+3}} \geq m\sqrt{8e^2kq^2/(8\sqrt{k})} \geq meqk^{1/4} \geq 8$. By this we then obtain

$$\left(m\sqrt{t/2^{j+3}}\ln\left(\sqrt{t/2^{j+3}}/\left(eq\right)\right)-\ln k\right)\geq m\sqrt{t/2^{j+3}}\ln\left(\sqrt{t/2^{j+3}}/\left(eq\right)\right)/2.$$

Thus letting $f(j)=2^{\frac{1}{2}j-5/2}m\sqrt{t}\ln(\sqrt{t/2^{j+3}}/(eq))/(j+1)^2$ we get that

$$\begin{aligned}\mathbb{P}\left[E_j\right]&\leq\exp\left(-\left(\frac{2^{j-1}m\sqrt{t/2^{j+3}}\ln(\sqrt{t/2^{j+3}}/(eq))}{(j+1)^2}\right)\right)\\&=\exp\left(-\frac{2^{\frac{1}{2}j-5/2}m\sqrt{t}\ln\left(\sqrt{t/2^{j+3}}/\left(eq\right)\right)}{\left(j+1\right)^2}\right)=\exp\left(-f\left(j\right)\right).\end{aligned}$$

Now using that $\ln(\sqrt{t/2^{j+3}}/(eq))\geq\ln(\sqrt{64\cdot 24e^3q^2k/(8\sqrt{k})}/(eq))\geq\ln\left(32\cdot 3e\right)/2=\ln\left(96e\right)/2$ for any $j\in 0,\ldots,\lg_2(k)/2$ and $t\geq 64\cdot 24e^3q^2k$ we get that the ratio between $f(j)$ and $f(j+1)$ for $j\in 0,\ldots,\lg_2(k)/2-1$ is lower bounded by

$$\frac{f\left(j+1\right)}{f\left(j\right)}=\frac{2^{1/2}\left(1-\ln\left(\sqrt{2}\right)/\ln\left(\sqrt{t/2^{j+3}}/\left(eq\right)\right)\right)\left(j+1\right)^2}{\left(j+2\right)^2}\geq\frac{2^{1/2}\left(1-\ln\left(2\right)/\ln\left(96e\right)\right)\left(j+1\right)^2}{\left(j+2\right)^2}.$$

By iteratively applying the above inequality for the ratio of consecutive terms of f we get that for $j'\in 1,\ldots,\lg_2(k)/2$ that

$$f\left(j'\right)\geq\frac{\left(2^{1/2}\left(1-\ln\left(2\right)/\ln\left(96e\right)\right)\right)^{j'}f\left(0\right)}{\left(j'+1\right)^2}\geq\frac{j'f\left(0\right)}{200},$$

where we in the last inequality have used that $\left(\left(1-2\ln\left(2\right)/\ln\left(96e\right)\right)2^{1/2}\right)^{j'}/\left(j'+1\right)^2\geq j'/200$ for $j'\geq 0$.

Now using the above inequality for f we get by a geometric series argument that,

$$\begin{aligned}\sum_{j=0}^{\lg_2(k)/2}\mathbb{P}\left[E_j\right]&\leq\exp\left(-f\left(0\right)\right)+\sum_{j=1}^{\lg_2(k)/2}\exp\left(-\frac{jf\left(0\right)}{200}\right)\\&\leq\exp\left(-f\left(0\right)\right)+\frac{\exp\left(-\frac{f(0)}{200}\right)}{1-\exp\left(-\frac{f(0)}{200}\right)}\leq 3\exp\left(-2^{-5/2}\cdot m\sqrt{t}\ln\left(\sqrt{t/2^3}/\left(eq\right)\right)/200\right),\end{aligned}$$

where we in the last inequality have used that $f(0)=2^{-5/2}\cdot m\sqrt{t}\ln(\sqrt{t/2^3}/(eq))\geq 250$, to say that $1/(1-\exp(-f(0)/200))\leq 2$.

Next we bound $\mathbb{P}[E'_j]$. Again by a union bound over all sets of $k/(2^j(j+1)^2)$ indices and Lemma 7, we get:

$$\mathbb{P}[E'_j]\leq\binom{k}{k/\left(2^j(j+1)^2\right)}\left(\sqrt{t2^{j-3}/k}/\left(eq\right)\right)^{-m\sqrt{t2^{j-3}/k}\cdot k/\left(2^j(j+1)^2\right)}.$$

Bounding $\binom{k}{k/(2^j(j+1)^2)}$ from above by $(e2^j(j+1)^2)^{k/(2^j(j+1)^2)}$ we get that

$$\mathbb{P}\left[E'_j\right]\leq\exp\left(-\frac{k}{2^j(j+1)^2}\cdot\left(m\sqrt{t2^{j-3}/k}\ln\left(\sqrt{t2^{j-3}/k}/\left(eq\right)\right)-\ln\left(e2^j\left(j+1\right)^2\right)\right)\right).$$

For $t\geq 24e^3kq^2$, we have $\sqrt{t2^{j-3}/k}/(eq)\geq\sqrt{3e2^j}$. Since $(j+1)^2\leq 3\cdot 2^j$ for all $j\geq 0$, $\sqrt{3e2^j}$ is at least $\sqrt{e2^{j/2}(j+1)}\geq(e2^j(j+1)^2)^{1/4}$ and thus $\ln(\sqrt{t2^{j-3}/k}/(eq))\geq\ln(e2^j(j+1)^2)/4$. For $q\geq 8/(em)$, we also have $m\sqrt{t2^{j-3}/k}\geq m\sqrt{3e^3q^2}\geq 8$ and hence:

$$m\sqrt{t2^{j-3}/k}\ln\left(\sqrt{t2^{j-3}/k}/\left(eq\right)\right)-\ln\left(e2^j\left(j+1\right)^2\right)\geq m\sqrt{t2^{j-3}/k}\ln\left(\sqrt{t2^{j-3}/k}/\left(eq\right)\right)/2.$$

Now let $g(j) = m\sqrt{tk}\ln(\sqrt{t2^{j-3}/k}/(eq))/((j+1)^2 2^{1/2j+5/2})$ then we have

$$\mathbb{P}\left[E'_j\right] \leq \exp\left(-\frac{km\sqrt{t2^{j-3}/k}\ln\left(\sqrt{t2^{j-3}/k}/\left(eq\right)\right)}{\left(j+1\right)^2 2^{j+1}}\right) = \exp\left(-g\left(j\right)\right).$$

Now for any $j \in 0,\ldots,\lg_2(k)/2$ and $t \geq 64\cdot 24e^3q^2k$ it holds that $\ln(\sqrt{t2^{j-3}/k}/(eq))$ is at least $\ln(\sqrt{32\cdot 24e^3q^2k/(8k)}/(eq)) \geq \ln\left(192e\right)/2$. This implies that the ratio between $g(j+1)$ and $g(j)$ for $j \in 0,\ldots,\lg_2(k)/2-1$ is

$$\frac{g\left(j+1\right)}{g\left(j\right)} = \frac{2^{-1/2}\left(1+\ln\left(\sqrt{2}\right)/\ln\left(\sqrt{t2^{j-3}/k}/(eq)\right)\right)(j+1)^2}{(j+2)^2} \leq \frac{2^{-1/2}\left(1+\ln\left(2\right)/\ln\left(192e\right)\right)(j+1)^2}{(j+2)^2}.$$

Now iteratively using the above relation on the ratio between $g(j+1)$ and $g(j)$ and that $g(\lg_2(k)/2) = k^{1/4}m\sqrt{t}\ln\left(t/(8e^2q^2\sqrt{k})\right)/(2^{7/2}(\ln(k)/2+1)^2)$ we get for $j' \in 0,\ldots,\lg_2(k)/2-1$ that

$$\begin{aligned}
g(j') &\geq \frac{\left(\lg_2\left(k\right)/2+1\right)^2 g\left(\lg_2\left(k\right)/2\right)}{\left(2^{-1/2}\left(1+\ln\left(2\right)/\ln\left(192e\right)\right)\right)^{\left(\lg_2(k)/2-j'\right)}\left(j'+1\right)^2} \\
&\geq \frac{k^{1/4}m\sqrt{t}\ln\left(t/(8e^2q^2\sqrt{k})\right)}{\left(2^{-1/2}\left(1+\ln\left(2\right)/\ln\left(192e\right)\right)\right)^{\left(\lg_2(k)/2-j'\right)}22k^{1/8}2^{7/2}} \\
&\geq \frac{k^{1/8}m\sqrt{t}\ln\left(t/(8e^2q^2\sqrt{k})\right)}{\left(2^{-1/2}\left(1+\ln\left(2\right)/\ln\left(192e\right)\right)\right)^{\left(\lg_2(k)/2-j'\right)}22\cdot 2^{7/2}} \\
&\geq \frac{\left(\lg_2(k)/2-j'\right)k^{1/8}m\sqrt{t}\ln\left(t/(8e^2q^2\sqrt{k})\right)}{200\cdot 22\cdot 2^{7/2}},
\end{aligned} \tag{9}$$

where we in the second inequality have used that for $j' \geq 0$ we have $(j'+1)^2 \leq 22\cdot 2^{j'/4} \leq 22\cdot k^{1/8}$ and where we in the last inequality have used that for $j' = 0,\ldots,\lg_2(k)/2-1$ we have

$$\left(2^{-1/2}\left(1+\ln(2)/(\ln(192e))\right)\right)^{-\left(\lg_2(k)/2-j'\right)} \geq (\lg_2(k)/2-j')/200.$$

Now using that Equation (9), also holds for $j' = \lg_2(k)/2$, and a geometric series argument we get that,

$$\begin{aligned}
&\sum_{j=0}^{\lg_2(k)/2}\mathbb{P}\left[E'_j\right] \\
&\leq \exp\left(-\frac{k^{1/8}m\sqrt{t}\ln\left(t/\left(8e^2q^2\sqrt{k}\right)\right)}{22\cdot 2^{7/2}}\right) + \sum_{j'=0}^{\lg_2(k)/2-1}\exp\left(-\frac{\left(\lg_2(k)/2-j'\right)k^{1/8}m\sqrt{t}\ln\left(t/\left(8e^2q^2\sqrt{k}\right)\right)}{200\cdot 22\cdot 2^{7/2}}\right) \\
&\leq \exp\left(-\frac{k^{1/8}m\sqrt{t}\ln\left(t/\left(8e^2q^2\sqrt{k}\right)\right)}{22\cdot 2^{7/2}}\right) + \frac{\exp\left(-k^{1/8}m\sqrt{t}\ln\left(t/\left(8e^2q^2\sqrt{k}\right)\right)/(200\cdot 22\cdot 2^{7/2})\right)}{1-\exp\left(-k^{1/8}m\sqrt{t}\ln\left(t/\left(8e^2q^2\sqrt{k}\right)\right)/(200\cdot 22\cdot 2^{7/2})\right)} \\
&\leq 11\exp\left(-\frac{k^{1/8}m\sqrt{t}\ln\left(t/\left(8e^2q^2\sqrt{k}\right)\right)}{200\cdot 22\cdot 2^{7/2}}\right),
\end{aligned}$$

where we in the last inequality have used that $k^{1/8}m\sqrt{t}\ln(t/(8e^2q^2\sqrt{k}))/(200\cdot 22\cdot 2^{7/2}) \geq 1/10$.

By the above upper bounds on $\sum_{j=0}^{\lg_2(k)/2} \mathbb{P}[E'_j]$ and $\sum_{j=0}^{\lg_2(k)/2} \mathbb{P}[E_j]$ we can conclude that

$$\begin{aligned}
&\mathbb{P}\left[\sum_{i=1}^{k} Z_i^2 \geq t\right] \\
&\leq 14\exp\left(-\min\left\{k^{1/8} m\sqrt{t}\ln\left(t/\left(8e^2q^2\sqrt{k}\right)\right)/\left(200\cdot 22\cdot 2^{7/2}\right), m\sqrt{t}\ln\left(\sqrt{t/2^3}/(eq)\right)/\left(200\cdot 2^{5/2}\right)\right\}\right) \\
&\leq 14\exp\left(-m\sqrt{t}/\left(200\cdot 2^{7/2}\right)\min\left\{k^{1/8}\ln\left(t/\left(8e^2q^2\sqrt{k}\right)\right)/22, \ln\left(t/\left(8e^2q^2\right)\right)\right\}\right) \\
&\leq 14\exp\left(-\frac{m\sqrt{t}\ln\left(\sqrt{t/2^3}/(eq)\right)}{200\cdot 44\cdot 2^{5/2}}\right),
\end{aligned}$$

where we have used that the second term in the $\min$ is always the smallest, when it is scaled by $1/44$, this follows from the assumption about $t \geq 64\cdot 24e^3kq^2$ implying that for any such given t there exist $\tilde{c} \geq 1$ such that $t = \tilde{c}8e^2kq^2$ and we get that the first term in the $\min$ is equal to $k^{1/8}\ln\left(\tilde{c}\sqrt{k}\right)/22 = k^{1/8}(\ln(\tilde{c}) + \ln(k)/2)/22$ and the second term in the $\min$ is equal to $\ln(\tilde{c}) + \ln(k)$, where by the claim follows.

□

We now restate and present the proof of Lemma 6.

Restatement of Lemma 6.

Lemma 6. *Let $Z_1, \ldots, Z_k$ be i.i.d. random variables distributed as the Z_i's in Lemma 4, with $m = c_2d/\ln n$ and $k = c_1\varepsilon^{-2}\ln n$ and $q = c_1\varepsilon$, where $c_1 \geq 1/c_2$. For $\varepsilon \leq c_1^{-1}/(e4)$ and $t \geq 2c_1^3e^8\ln n$, we have that*

$$\mathbb{P}\left[\sum_{i=1}^{k} Z_i^2 > t\right] \leq 3n^{-4c_1}.$$

Proof. In the following we assume for simplicity that $\lg_2(k)$ and $\lg_2(t)$ are integers. We proceed in a somewhat similar fashion as in the proof of Lemma 5. For $j = \lg_2 t, \ldots, \lg_2 k$ let E_j be the event that there are at least $2^{j-1}/(j - \lg_2(t) + 1)^2$ indices such that $Z_i^2 \geq t/2^{j+1}$. Assume that none of the events E_j occurs, we then have that

$$\begin{aligned}
&\sum_{i=1}^{k} Z_i^2 \\
&\leq \sum_{i=1}^{k}\sum_{j=\lg_2(t)}^{\infty} 1_{\{Z_i^2 \geq \frac{t}{2^{j+1}}\}}\frac{t}{2^{j+1}} \\
&= \sum_{j=\lg_2(t)}^{\infty}\frac{t}{2^{j+1}}\sum_{i=1}^{k} 1_{\{Z_i^2 \geq \frac{t}{2^{j+1}}\}} \\
&= \sum_{j=\lg_2(t)}^{\lg_2 k}\frac{t}{2^{j+1}}\sum_{i=1}^{k} 1_{\{Z_i^2 \geq \frac{t}{2^{j+1}}\}} + \sum_{j=\lg_2 k+1}^{\infty}\frac{t}{2^{j+1}}\sum_{i=1}^{k} 1_{\{Z_i^2 \geq \frac{t}{2^{j+1}}\}} \\
&\leq \sum_{j=\lg_2(t)}^{\lg_2 k}\frac{t2^{j-1}}{2^{j+1}\left(\lg_2(t) - j + 1\right)^2} + \sum_{j=\lg_2 k+1}^{\infty}\frac{tk}{2^{j+1}} \\
&\leq \frac{t}{4}\sum_{j=1}^{\infty}\frac{1}{j^2} + \frac{t}{4}\sum_{j=0}^{\infty}\frac{1}{2^j} \\
&\leq \frac{t\pi^2}{24} + \frac{t}{2} < t,
\end{aligned}$$

where the first inequality follows by $Z_i^2 \leq 1$, so the sum of the terms $1_{\{Z_i^2 \geq t/2^{j+1}\}} t/2^{j+1}$ starting at $j = \lg_2(t)$ is always greater than Z_i^2. Thus we conclude that one of the events E_j happens when $\sum_{i=1}^k Z_i^2 \geq t$. Now by an union bound over the events E_j we have

$$\mathbb{P}\left[\sum_{i=1}^{k} Z_i^2 \geq t\right] \leq \sum_{j=\lg_2(t)}^{\lg_2(k)} \mathbb{P}\left[E_j\right].$$

When E_j happens we know that there is a set S of $2^{j-1}/(j-\lg_2(t)+1)^2$ indices such that for $i \in S$ we have $Z_i^2 \geq t/2^{j+1}$. Thus the probability of each E_j can be bounded by using a union bound over all such possible sets of indices (k choose $2^{j-1}/(j-\lg_2(t)+1)^2$). Now using that the Z_i's are independent and identically distributed, the probability of each of the sets S splits into a product of probabilities $\mathbb{P}\left[Z_i^2 \geq t/2^{j+1}\right]$, where Lemma 7 can be used to bound each of these probabilities. We note that Lemma 7 with $Z \geq \sqrt{t/2^{j+1}}$ is applicable since $\sqrt{t/2^{j+1}}/q \geq \sqrt{2c_1^3 e^8 \ln(n)/(2k)}/(c_1\varepsilon) = \sqrt{2c_1^3 e^8/(2c_1^3)} \geq e^4$, where we have used the assumption that $t \geq 2c_1^3 e^8 \ln(n)$. We now get that:

$$\begin{aligned}\mathbb{P}\left[E_j\right] &\leq \binom{k}{2^{j-1}/\left(j-\lg_2(t)+1\right)^2}\left(\sqrt{t/2^{j+1}}/\left(eq\right)\right)^{\sqrt{t/2^{j+1}}m2^{j-1}/(j-\lg_2(t)+1)^2} \\ &\leq \exp\left(-\frac{2^{j-1}\left(\sqrt{t/2^{j+1}}m\ln\left(\sqrt{t/2^{j+1}}/(eq)\right)-\ln\left(ek(j-\lg_2(t)+1)^2/2^{j-1}\right)\right)}{(j-\lg_2(t)+1)^2}\right),\end{aligned}$$

where the last inequality follows by $\binom{k}{2^{j-1}/(j-\lg_2(t)+1)^2} \leq \left(ek(j-\lg_2(t)+1)^2/2^{j-1}\right)^{2^{j-1}/(j-\lg_2(t)+1)^2}$.

To evaluate the term $\sqrt{t/2^{j+1}}m\ln\left(\sqrt{t/2^{j+1}}/(eq)\right)-\ln\left(ek(j-\lg_2(t)+1)^2/2^{j-1}\right)$ we notice the following four relations for $j = \lg_2(t), \ldots, \lg_2(k)$

$$\sqrt{t/2^{j+1}}m \geq \sqrt{2c_1^3 e^8 \ln\left(n\right)/\left(2k\right)}c_2 d/\ln\left(n\right) \geq \sqrt{2c_1^3 e^8 \varepsilon^2/\left(2c_1\right)}c_2 k/\ln\left(n\right) \geq \sqrt{2c_1^3 e^8 c_1/2}c_2\varepsilon^{-1} \geq e^4\varepsilon^{-1},$$

$$\left(\sqrt{t/2^{j+1}}/(eq)\right) \geq \sqrt{2c_1^3 e^8 \ln\left(n\right)/\left(2k\right)}/\left(ec_1\varepsilon\right) = \sqrt{2c_1^3 e^8/\left(2e^2c_1^3\right)} \geq e^3,$$

$$\frac{ek}{2^{j-1}} \leq e2k/t \leq e2c_1/\left(2c_1^3 e^8\varepsilon^2\right) \leq 1/\left(e^7\varepsilon^2\right),$$

$$j-\lg_2(t)+1 \leq \lg_2(k/t)+1 \leq \lg_2\left(c_1/\left(2c_1^3e^8\varepsilon^2\right)\right)+1 = \lg_2\left(2c_1/\left(2c_1^3e^8\varepsilon^2\right)\right) \leq \lg_2\left(1/\left(e^8\varepsilon^2\right)\right),$$

where we have used that $c_1 \geq 1/c_2$ $t \geq 2c_1^3 e^8 \ln(n)$, $k = c_1\varepsilon^{-2}\ln(2)$ and $d \geq k$. By the above relations we conclude that for sufficiently small ε, we have that

$$\sqrt{t/2^{j+1}}m\ln\left(\sqrt{t/2^{j+1}}/(eq)\right)-\ln\left(ek(j-\lg_2(t)+1)^2/2^{j-1}\right) \geq \sqrt{t/2^{j+1}}m\ln\left(\sqrt{t/2^{j+1}}/(eq)\right)/2.$$

Hence for such ε and $f(j) = 2^{j/2-5/2}\sqrt{t}m\ln\left(\sqrt{t/2^{j+1}}/(eq)\right)/(j-\lg_2(t)+1)^2$ we have that

$$\mathbb{P}\left[E_j\right] \leq \exp\left(-\frac{2^{j-1}\sqrt{t/2^{j+1}}m\ln\left(\sqrt{t/2^{j+1}}/\left(eq\right)\right)/2}{\left(j-\lg_2(t)+1\right)^2}\right) = \exp\left(-f\left(j\right)\right).$$

Now using the assumptions that $t \geq 2c_1^3 e^8 \ln(n)$ and $q = c_1\varepsilon$ we get that $\sqrt{t/2^{j+1}}/(eq) \geq \sqrt{2c_1^3e^8/2c_1^3}/e \geq e^3$ such that for $j = \lg_2 t, \ldots, \lg_2(k)-1$

$$\frac{f\left(j+1\right)}{f\left(j\right)} \geq \frac{\left(j-\lg_2\left(t\right)+1\right)^2\left(1-\ln\left(2\right)/6\right)\sqrt{2}}{\left(j+1-\lg_2\left(t\right)+1\right)^2},$$

using this iteratively we get that for $j' \in 1, \dots, \lg_2(k) - \lg_2(t)$

$$f(\lg_2(t) + j') \geq \frac{\left((1 - \ln(2)/6)\sqrt{2}\right)^{j'} f(\lg_2 t)}{(j'+1)^2} \geq \frac{j' f(\lg_2 t)}{150},$$

where the last inequality follows by $\left((1 - \ln(2)/6)\sqrt{2}\right)^{j'} / (j'+1)^2 \geq j'/150$ for $j' > 1$.

Now using a geometric series argument we get that

$$\begin{aligned}
&\mathbb{P}\left[\sum_{i=1}^{k} Z_i^2 \geq t\right] \\
&\leq \sum_{j=\lg_2(t)}^{\lg_2(k)} \mathbb{P}\left[E_j\right] \\
&\leq \sum_{j=\lg_2(t)}^{\lg_2(k)} \exp\left(-f(j)\right) \\
&\leq \exp\left(-f(\lg_2 t)/150\right) + \sum_{j=1}^{\infty} \exp\left(-j f(\lg_2 t)/150\right) \\
&\leq 2\frac{\exp(-f(\lg_2 t)/150)}{1 - \exp(-f(\lg_2 t)/150)} \\
&\leq 2\frac{\exp\left(-tm\ln(1/(\sqrt{2}eq))/(600\sqrt{2})\right)}{1 - \exp\left(-tm\ln(1/(\sqrt{2}eq))/(600\sqrt{2})\right)}.
\end{aligned}$$

Now using that $t \geq 2c_1^3 e^8 \ln(n)$ and $\varepsilon \leq c_1^{-1}/(4e)$ so $\ln(1/(\sqrt{2}eq)) \geq \ln(2)$ we end up with the following inequality $t\ln(1/(\sqrt{2}eq))/(600\sqrt{2}) \geq c_1^3 e^8 \ln(2)/(300\sqrt{2})\ln(n) \geq 4c_1^3$ and since $m \geq 1$ we conclude that

$$\mathbb{P}\left[\sum_{i=1}^{k} Z_i^2 \geq t\right] \leq 2\frac{\exp\left(-tm\ln(1/(\sqrt{2}eq))/(600\sqrt{2})\right)}{1 - \exp\left(-tm\ln(1/(\sqrt{2}eq))/(600\sqrt{2})\right)} \leq 2\frac{n^{-4c_1^3}}{1 - n^{-4c_1^3}} \leq 3n^{-4c_1},$$

where we in the last inequality have assumed that $n \geq 2$ and used that $c_1 \geq 1$, which completes the proof.

□

B. Proof of Theorem 2

As described in the sketch of the proof in Section 4, we proceed with the following two cases (and two steps).

First case: $q = \Omega\left(\frac{\ln\frac{1}{\delta}}{d}\right)$**.** For this case we need lower bounds on the tail probabilities for weighted sums of independent χ^2-distributions, thus we now restate Theorem 7 from (Zhang & Zhou, 2020) in a slightly weaker form.

Lemma 8 ((Zhang & Zhou, 2020))**.** *Let $g_1, \dots, g_d$ be independent $N(0,1)$ random variables and $u_1, \dots, u_d$ be non-negative real numbers, then for constants $0 < c_3$ and $C_3 \geq 1$ we have that $\forall x \geq 0$*

$$\mathbb{P}\left[\sum_{i=1}^{d} u_i(g_i^2 - 1) \geq x\right] \geq c_3 \exp\left(-C_3 x^2/\|u\|_2^2\right).$$

We will also need the following reverse Chernoff bound from (Mousavi, 2010) which we restate in a multiplicative version instead of an additive:

Lemma 9 ((Mousavi, 2010))**.** *Let X be a binomial random variable with r trials and success probability $q \leq 1/4$. Then for any $0 \leq \alpha q \leq 1/4$ it holds that*

$$\Pr\left[X \geq (1+\alpha)qr\right] \geq \frac{1}{4}\exp\left(-2\alpha^2 qr\right).$$

With the above lemma stated we proceed with the first step in the proof.

First step in proof of Theorem 2. We condition on the randomness in HD resulting in the fixed vector u as argued earlier. In this case, we start by showing that $\sum_i b_i$ is large with reasonable probability. Observe that $\sum_i b_i$ is binomial distributed with $r = kd/2^l$ trials and success probability q. Hence for $\alpha = \sqrt{\ln(1/(4^4\delta))/(8qr)}$, it follows from Lemma 9 that either $\alpha q > 1/4$ or $q > 1/4$ or $\mathbb{P}[\sum_i b_i \geq qr + \sqrt{\ln(1/(4^4\delta))qr/8}] \geq \delta^{1/4}$.

If $q \geq 1/4$ we are done. Likewise, if $\alpha q \geq 1/4$ then $q \geq 1/(4\alpha)$ implying that $q \geq \sqrt{qr/(2\ln(1/(4^4\delta)))} \geq \Omega(\varepsilon^{-2})$ by assumptions on $r = kd/2^l$, $k = \lg(1/\delta)/\varepsilon^2$ and $d/2^l \geq 1$ and we are done again.

Thus, what remains is the case of $\mathbb{P}[\sum_i b_i \geq qr + \sqrt{\ln(1/(4^4\delta))qr/8}] \geq \delta^{1/4}$. Let us condition on $\sum_i b_i \geq qr + \sqrt{\ln(1/(4^4\delta))qr/8}$. Then by Lemma 8 with $x = 0$ we get $\mathbb{P}[\sum_i b_i(N_i^2 - 1) \geq 0] \geq c_3$. This implies $\sum_i b_i N_i^2 \geq \sum_i b_i \geq qr + \sqrt{\ln(1/(4^4\delta))qr/8}$ with probability at least $c_3\delta^{1/4}$. But $(2^l/(dq))(qr + \sqrt{\ln(1/(4^4\delta))qr/8}) = k + \sqrt{\ln(1/(4^4\delta))2^{2l}r/(8d^2q)} = k + \Omega(\sqrt{\ln(1/\delta)2^l k/(qd)})$. Thus with probability at least $c_3\delta^{1/4}$, we have $(2^l/(dq))\sum_i b_i N_i^2 \geq k + \Omega(\sqrt{\ln(1/\delta)2^l k/(qd)})$. And since $\|Pu\|^2 \overset{d}{=} (2^l/(dq))\sum_i b_i N_i^2$ we also have that $\|Pu\|^2 \geq k + \Omega(\sqrt{\ln(1/\delta)2^l k/(qd)})$ with probability $c_3\delta^{1/4}$. Further since we noticed that the probability of $HDx = u$ is at least $\sqrt{2\delta}$ it now follows what with probability at least $c_3\delta^{3/4}$ we have that

$$\frac{1}{k}\left\|PHDx\right\|^2 > 1 + \Omega(\sqrt{\ln(1/\delta)2^l/(kqd)}).$$

Thus for $\delta \leq c_3^4$ it follows that we must have

$$\Omega\left(\sqrt{\ln(1/\delta)2^l/(kqd)}\right) \leq \varepsilon$$

for $\frac{1}{k}\left\|PHDx\right\|^2$ to satisfy Equation (1) (being a length preserving projection) with probability δ, which implies $q \geq \Omega(\ln(1/\delta)2^l/(\varepsilon^2 kd)) = \Omega(\ln(1/\delta)/d)$ where we have used that 2^l is $\Theta(\ln(1/\delta))$ by the choose of l, which completes the proof of the first step.

Second case $q = \Omega\left(\varepsilon\min\left\{1, \ln^2\left(1/\delta\right)/\left(d\ln\left(1/\varepsilon\right)\right)\right\}\right)$. In this section we show the second step of the lower bound. We use the result from the first step which results in $q = \Omega(\ln(1/\delta)/d)$. The basic idea is to show that there is a reasonably large probability that the first coordinate $(Pu)_1$ is so large that it distorts the embedding of x by too much, even when all other coordinates behave well.

In what follows we state the lemmas we will need in the proof of the second step. By the the first step, we already have our claimed lower bound in Theorem 2 whenever $\Theta\left(\max\left\{\ln(1/\delta)/d, \varepsilon\min\left\{1, \ln^2(1/\delta)/(d\ln(1/\varepsilon))\right\}\right\}\right) = \Theta\left(\ln(1/\delta)/d\right)$, so we now consider the cases where ε, δ, d are such that

$$\begin{aligned}&\Theta\left(\max\left\{\ln(1/\delta)/d, \varepsilon\min\left\{1, \ln^2(1/\delta)/(d\ln(1/\varepsilon))\right\}\right\}\right)\\ &\quad= \Theta\left(\varepsilon\min\left\{1, \ln^2(1/\delta)/(d\ln(1/\varepsilon))\right\}\right),\end{aligned}$$

and then show that for

$$c_4\ln(1/\delta)/d \leq q \leq c_5\varepsilon\min\left\{1, \ln^2\left(1/\delta\right)/\left(d\ln\left(1/\varepsilon\right)\right)\right\}, \tag{10}$$

where c_4 is the constant from the lower bound $q \geq c_4\ln(1/\delta)/d$ and c_5 is a constant to be fixed later (but will be chosen less than 1), we have that the projection fails with at least δ probability.

We construct our hard instance as in step one, except that we will have to slightly adjust the value of l (to deal with constants). We thus set l to be the integer such that $l \leq \lg_2\left(\lg_2((1/\delta)^{\min\{1/50, c_4/\lg_2(e)\}})\right) \leq l + 1$ and define

$$x_i := \begin{cases}\frac{1}{\sqrt{2^l}}\text{ , for } i \leq 2^l\\ 0\text{ , otherwise}\end{cases}$$

It thus follows that with probability $2^{-2^l} \geq \delta^{\min\{1/50, c_4/\lg_2(e)\}}$, the first 2^l signs in D are 1 , thus $Dx = x$ with at least probability $\delta^{\min\{1/50, c_4/\lg_2(e)\}}$. We further notice that for the above x we have that

$$u_i := (Hx)_i = \begin{cases} \sqrt{\frac{2^l}{d}} \text{ , for } i \equiv 0 \mod (2^l) \\ 0 \text{ , otherwise} \end{cases}$$

Notice that u has $d/2^l$ entries being $\sqrt{2^l/d}$, and rest of the entries are 0. Let m denote the number of non-zero entries. If $\ln(1/\delta)/(qm) \leq c_6$ then by the choice of l and $m = d/2^l$ it holds that $q \geq \ln^2(1/\delta)\min\left\{1/50, c_4/\lg_2(e)\right\}/(c_6 d)$, and since $\Theta\left(\max\left\{\ln(1/\delta)/d, \varepsilon\min\{1, \ln^2(1/\delta)/(d\ln(1/\varepsilon))\}\right\}\right) = O(\ln^2(1/\delta)/d)$, we are done. Hence, we may assume in the following that

$$\ln(1/\delta)/(qm) \geq c_6, \tag{11}$$

where c_6 is at least 8, and will be chosen larger later.

For $i \in [1, k]$ let Z_i denote a normalized sum of m independent Bernoulli random variables: $Z_i = (1/m)\sum_{j=1}^{m} b_{i,j}$ and N_i denote a standard normal random variable, where all the Z_i's and the N_i's are independent of each other. Then, for the u as described above we have by linear combinations of independent normal distributions that:

$$\|Pu\|^2 \overset{d}{=} \sum_{i=1}^{k}\left(\sum_{j=1}^{\frac{d}{2^l}}\sqrt{\frac{2^l}{dq}}b_{i,j}N_{i,j}\right)^2 \overset{d}{=} \sum_{i=1}^{k}\frac{1}{q}Z_i N_i^2.$$

Next we present the lemmas we will use in the second step in the proof of Theorem 2. The proof of the lemmas is in Appendix B.1.

The following lemma states that with good probability the first coordinate of the projection vector $Z_1 N_1^2/q$ is large.

Lemma 10. *For $0 < \varepsilon, \delta \leq 1/4$, c_5 sufficiently small (Equation (10)), and c_6 sufficiently large (Equation (11)) we have with probability at least $\delta^{1/50+1/2+1/\pi}$ that*

$$\frac{1}{q}Z_1 N_1^2 \geq \frac{5\ln(1/\delta)}{\varepsilon}.$$

We also would need to show that the sum of the coordinates, except $Z_1 N_1^2/q$, have a good concentration around its mean:

Lemma 11. *For $0 < \varepsilon \leq 1/4$ and $0 < \delta \leq 1/8$ we have with probability at least $\delta^{1/8}$ that*

$$\sum_{i=2}^{k}\frac{1}{q}Z_i N_i^2 \geq (1 - 3\varepsilon)(k - 1).$$

We are now ready to put the above lemmas together and complete the proof of Theorem 2.

Second step in proof of Theorem 2.

Proof. Let $0 < \varepsilon \leq 1/4$ and $0 < \delta \leq 1/8$. We choose c_5 and c_6 according to Lemma 10. This implies that with probability at least $\delta^{1/50+1/2+1/\pi}$, it holds that $Z_1 N_1^2/q \geq 5\ln(1/\delta)\varepsilon^{-1}$. In addition, by Lemma 11, $\sum_{i=2}^{k} Z_i N_i^2/q \geq (1-3\varepsilon)(k-1)$ with probability at least $\delta^{1/8}$.

Therefore, by independence of the Z_i's and the N_i's, with probability $\delta^{1/50+1/2+1/\pi+1/8}$ for the vector u:

$$\begin{aligned}
\|Pu\|^2 &\overset{d}{=} \sum_{i=1}^{k} \frac{1}{q} Z_i N_i^2 \\
&= \frac{1}{q} Z_1 N_1^2 + \sum_{i=2}^{k} \frac{1}{q} Z_i N_i^2 \\
&\geq 5\ln(1/\delta)\varepsilon^{-1} + (1-3\varepsilon)(k-1) \\
&= 5\varepsilon k + k - 3\varepsilon k - 1 + 3\varepsilon \\
&= (1+\varepsilon)k + \varepsilon k - 1 + 3\varepsilon \\
&> (1+\varepsilon)k,
\end{aligned}$$

where the last inequality follows by the assumptions on $\varepsilon \leq 1/4$ implying that $\varepsilon k = \ln(1/\delta)\varepsilon^{-1} > 4 \geq 1 - 3\varepsilon$.

Since we have chosen l such that $l \leq \lg_2\left(\lg_2((1/\delta)^{\min\{1/50, c_4/\lg_2(e)\}})\right) \leq l+1$, we have that with probability at least $\delta^{1/50}$ $u = HDx$, independently of the outcomes of the $b_{i,j}$'s and the $N_{i,j}$'s. Therefore, by the law of conditional probability, with probability at least $\delta^{1/50+1/2+1/\pi+1/8+1/50} \geq \delta$ it holds that $\|PHDx\|^2 > (1+\varepsilon)k$. Thus we have shown that for δ, ε less than sufficiently small constants, we must have $q \geq c_5 \varepsilon \min\{1, \ln(1/\delta)/(d\ln(1/\varepsilon))\}$ for the mapping PHD to be a length preserving random projection with probability $1-\delta$. □

B.1. Inequalities for the lower bound

In this section we proof Lemma 10 and Lemma 11. Lemma 10 states that the first coordinate $Z_1 N_1^2/q$ is $\Omega(\varepsilon k)$ with good probability and Lemma 11 says that $\sum_{i=2}^{k} Z_i N_i^2/q$ is $\Omega(k)$ with good probability.

We consider the cases where ε, δ, d are such that

$$c_4 \ln(1/\delta)/d \leq q \leq c_5 \varepsilon \min\{1, \ln^2(1/\delta)/(d\ln(1/\varepsilon)\} \tag{12}$$

where c_4 is the constant from Theorem 2 and c_5 is a constant to be fixed later and will be chosen to be < 1.

We have $m = d/2^l$ where $l \leq \lg_2\left(\lg_2((1/\delta)^{\min\{1/50, c_4/\lg_2(e)\}})\right) \leq l+1$ implying that

$$m \leq 2d/(\min\{1/50, c_4/\lg_2(e)\}\lg_2(1/\delta)) \leq 2d/(\min\{1/50, c_4/\lg_2(e)\}\ln(1/\delta)),$$

and

$$m \geq d/(\min\{1/50, c_4/\lg_2(e)\}\lg_2(1/\delta)) \geq d/(\min\{1/50, c_4/\lg_2(e)\}\lg_2(e)\ln(1/\delta)).$$

We notice that for q's as in Equation (12) and the above m we have that

$$\min\{1/50, c_4/\lg_2(e)\}\lg_2(e)\ln(1/\delta)/c_4 \geq \ln(1/\delta)/qm \geq \min\{1/50, c_4/\lg_2(e)\}\ln(1/\varepsilon)/(2c_5\varepsilon), \tag{13}$$

especially that $1/(qm) \leq 1$.

We have that

$$\ln(1/\delta)/(qm) \geq c_6, \tag{14}$$

where c_6 is at least 8, and will be chosen larger later.

We consider the random variables $Z_1 N_1^2/q$ and $\sum_{i=2}^{k} Z_i N_i^2/q$, where the Z_i's denotes normalized sums of independent Bernoulli random variables $Z_i = (1/m)\sum_{j=1}^{m} b_j$ and the N_i's denotes standard normal random variable, where all the Z_i's and the N_i's are independent of each other.

We now present a technical lemma we will use in our proofs:

Lemma 12. *For $a, x \in \mathbb{R}$ such that $0 \leq x \leq 1$ and $0 \leq ax \leq 1$ we have that*

$$(1-x)^a \leq (1 - ax/2).$$

Proof. Cases $x = 0, 1$ and $ax = 0$ can be realised by insertion, and the case $ax = 1$ corresponds to $(1-x)^{1/x} \le 1/2$ which holds. Now for the remainding cases we first note by Taylor expansion of $\ln(1-x) = -\sum_{i=1}^{\infty} x^i/i$ that $(1-x)^a = \exp(-a\sum_{i=1}^{\infty} x^i/i)$ and $(1-ax/2) = \exp(-\sum_{i=1}^{\infty}(ax/2)^i/i)$. So it suffices to show that $\sum_{i=1}^{\infty}(ax/2)^i/i \le a\sum_{i=1}^{\infty} x^i/i$. Now using that $ax \le 1$ and that a geometric series with common ratio of $1/2$ equals 2 we get that $\sum_{i=1}^{\infty}(ax/2)^i/i = (ax/2)\sum_{i=1}^{\infty}\frac{(ax/2)^{i-1}}{i} \le (ax/2)2 = ax$. We also have that $ax \le a\sum_{i=1}^{\infty} x^i/i$. Hence we conclude that $\sum_{i=1}^{\infty}(ax/2)^i/i \le a\sum_{i=1}^{\infty} x^i/i$ which proofs the claim. □

In what follows we prove Lemma 13, Remark 14 and Lemma 15 which combined yield that with good probability we have a lower bound of $\Theta(\varepsilon^{-1})$ on the scaled binomial Z_1/q.

Lemma 13. *Let $0 < \varepsilon, \delta \le 1/4$. Let further $c_7 \le 1$ and $L = c_7 \ln(1/\delta)/\ln(\ln(1/\delta)/(qm))$ if $m/L \ge 1$, $qm/L \le 1$ and c_5 (Equation (12)) is chosen so small that $\min\{1/50, c_4/\lg_2(e)\}/(2c_5)$ is greater than 2. We then have with probability at least δ^{c_7} that:*

$$\frac{Z_1}{q} = \frac{1}{q}\sum_{i=1}^{m}\frac{1}{m}b_{1,i} \ge \frac{c_8 c_7}{\varepsilon\sqrt{c_5}},$$

with $c_8 = \ln(2)\sqrt{\min\{1/50, c_4/\lg_2(e)\}}/(4\sqrt{2})$.

Proof. The idea of the proof is to divide the m Bernoulli trails in $Z_1 = \sum_{i=1}^{m}\frac{1}{m}b_{1,i}$ into L disjoint buckets of size m/L (we choose c_7 such that the bucket size is an integer), and then calculate the probability that all the buckets have at least one success, and get thereby obtain the above lower bound on Z_1/q.

Using that the buckets are disjoint so the events of buckets having a success in it is independent of each other the probability of having at least one success in every disjoint bucket is $(1-(1-q)^{m/L})^L$. Now using Lemma 12 with $x = q$ and $a = m/L$ we get that $\left(1-(1-q)^{m/L}\right)^L \ge (1-(1-(qm)/(2L)))^L = ((qm)/(2L))^L$. Now plugging L into this expression we get that

$$\begin{aligned}\left(\frac{qm}{2L}\right)^L &= \left(\frac{\ln\left(\ln\left(1/\delta\right)/\left(qm\right)\right)qm}{2c_7\ln(1/\delta)}\right)^{c_7\ln(1/\delta)/\ln(\ln(1/\delta)/(qm))}\\ &= \left(\frac{\ln\left(\ln\left(1/\delta\right)/\left(qm\right)\right)}{2c_7}\right)^{c_7\ln(1/\delta)/\ln(\ln(1/\delta)/(qm))}\delta^{c_7} \ge \delta^{c_7},\end{aligned}$$

where the last inequality follows from the assumption that $\ln(1/\delta)/(qm) \ge 8$ (Equation (14)) so the first term in the second to last expression is lower bounded by 1. Hence with probability at least δ^{c_7} we have that all the disjoint L buckets have at least one success and hence on this event $Z_1/q \ge L/(qm)$. Plugging L into the expression, using that $x/\ln x$ is increasing for $x \ge 3$ and that $\ln(1/\delta)/(qm)$ is lower bounded by $\min\{1/50, c_4/\lg_2(e)\}\ln(1/\varepsilon)/(2c_5\varepsilon)$ (Equation (13)) which is at least 3 by assumptions on c_5 and $\varepsilon \le 1/4$, it follows that

$$\frac{1}{q}Z_1 \ge \frac{c_7\ln\left(1/\delta\right)}{qm\ln\left(\ln\left(1/\delta\right)/\left(qm\right)\right)} \ge \frac{c_7\min\{1/50, c_4/\lg_2(e)\}\ln(1/\varepsilon)}{2c_5\varepsilon\ln\left(\min\{1/50, c_4/\lg_2(e)\}\ln(1/\varepsilon)/(2c_5\varepsilon)\right)}. \tag{15}$$

Since $\min\{1/50, c_4/\lg_2(e)\}/(2c_5) \ge 2$ by assumption it holds that $\ln\left(\min\{1/50, c_4/\lg_2(e)\}\ln(1/\varepsilon)/(2c_5\varepsilon)\right)$ is less than or equal to $\ln\left((\min\{1/50, c_4/\lg_2(e)\}/(2c_5\varepsilon))^2\right)$, thus

$$\frac{\ln(1/\varepsilon)}{\ln\left(\min\{1/50, c_4/\lg_2(e)\}\ln(1/\varepsilon)/(2c_5\varepsilon)\right)} \ge \frac{\ln(1/\varepsilon)}{2\ln\left(\min\{1/50, c_4/\lg_2(e)\}/(2c_5\varepsilon)\right)}.$$

Now using that $x/(x+a)$ with $a, x > 0$ is increasing in x, with $a = \ln\left(4/(c_5\min\{1/50, c_4/\lg_2(e)\})\right)$, $x = \ln(1/\varepsilon)$ and $\ln(1/\varepsilon) \ge \ln 2$ it follows that

$$\frac{\ln(2)}{2(\ln\left(\min\{1/50, c_4/\lg_2(e)\}/(2c_5)\right)+\ln(2))} \ge \frac{\ln(2)}{4\ln\left(\min\{1/50, c_4/\lg_2(e)\}/(2c_5)\right)}.$$

Plugging this into Equation (15) it follows that

$$\frac{1}{q}Z_1 \geq \frac{c_7 \min\left\{1/50, c_4/\lg_2(e)\right\}\ln(2)}{8c_5\varepsilon \ln\left(\min\left\{1/50, c_4/\lg_2(e)\right\}/(2c_5)\right)}.$$

Now using that $x/\ln(x) \geq \sqrt{x}$ for $x \geq 1$ with $x = \min\left\{1/50, c_4/\lg_2(e)\right\}/(2c_5)$, which is greater than 2 by assumptions, we get that

$$\frac{\min\left\{1/50, c_4/\lg_2(e)\right\}}{2c_5 \ln\left(\min\left\{1/50, c_4/\lg_2(e)\right\}/(2c_5)\right)} \geq \sqrt{\min\left\{1/50, c_4/\lg_2(e)\right\}/(2c_5)}.$$

Thus we get

$$\frac{1}{q}Z_1 \geq \frac{c_7\ln(2)\sqrt{\min\left\{1/50, c_4/\lg_2(e)\right\}}}{4\sqrt{2c_5}\varepsilon} = \frac{c_8 c_7}{\varepsilon\sqrt{c_5}},$$

with $c_8 = \ln(2)\sqrt{\min\left\{1/50, c_4/\lg_2(e)\right\}}/(4\sqrt{2})$. □

We now notice that the assumption of $qm/L \leq 1$ in Lemma 13 for a fixed c_7 maybe be removed.

Remark 14. *We may assume that $qm/L \leq 1$ in Lemma 13 for a fixed c_7 holds by choosing c_6 sufficiently large.*

Proof. To see this we notice that the assumption $qm/L \leq 1$ is equivalent to

$$\frac{qm\ln\left(\ln(1/\delta)/(qm)\right)}{c_7\ln(1/\delta)} \leq 1.$$

So if we can upper bound the left hand side by 1, we are done. To upper bound the left hand side we use that $\ln(x)/x$ is decreasing for $x \geq 3$ so using this fact with $x = \ln(1/\delta)/(qm)$ and $\ln(1/\delta)/(qm)$ being lower bounded by c_6 (Equation (14)) we get that

$$\frac{qm\ln\left(\ln(1/\delta)/(qm)\right)}{c_7\ln(1/\delta)} \leq \frac{\ln c_6}{c_7 c_6},$$

which is less than 1 for sufficiently large c_6 hence the assumption of $qm/L \leq 1$ for a fixed c_7 may be removed.

□

Lemma 15. *Let the setting be as in Lemma 13 other than $m/L \leq 1$ then we have with probability δ^{c_7} that*

$$\frac{1}{q}Z_1 \geq \frac{1}{q} \geq \frac{1}{c_5\varepsilon}.$$

Proof. Now since $1/q \geq Z_1/q$ happens if and only if $Z_1 = (1/m)\sum_{j=1}^m b_{1,j} = 1$, hence all the Bernoulli trails in the binomial being one, the above happens with probability q^m. This probability is less than or equal to $(qm/L)^L$ since $m/L \leq 1$ now the calculations in Lemma 13 for $(qm/(2L))^L$ yields that $q^m \geq \delta^{c_7}$. The later lower bound on $1/q$ follows from $q \leq c_5\varepsilon$ (Equation (12)) □

We now show that with good probability we have that N_1^2 is $\Theta(\ln(1/\delta))$.

Lemma 16. *For $x \geq 0$ it holds that $\mathbb{P}\left[N^2 \geq x\right] \geq 1 - \sqrt{1 - \exp(-2x/\pi)}$.*

Proof. First notice that the event $\{N^2 \leq x\}$ is equivalent to the event $\{-\sqrt{x} \leq N \leq \sqrt{x}\}$. Using this and that the density functions of N is $(2\pi)^{-1/2}\exp(-x^2/2)$ we get that

$$\mathbb{P}\left[N^2 \leq x\right] = \int_{-\sqrt{x}}^{\sqrt{x}} (2\pi)^{-1/2}\exp(-x^2/2)dx.$$

Now using the equation on the top of page 64 and equation (1.5) on the same page in (Pólya, 1949) we get that the above is at most $\sqrt{1 - \exp(-2x/\pi)}$. This conclude the proof since $\mathbb{P}\left[N^2 \geq x\right] = 1 - \mathbb{P}\left[N^2 \leq x\right]$. □

We will now combine Lemma 13, Remark 14, Lemma 15 and Lemma 16 to show Lemma 10, recall that Lemma 10 is.

Restatement of Lemma 10.

Lemma 10. *For* $0 < \varepsilon, \delta \leq 1/4$, c_5 *sufficiently small (Equation* (10)*), and* c_6 *sufficiently large (Equation* (11)*) we have with probability at least* $\delta^{1/50+1/2+1/\pi}$ *that*

$$\frac{1}{q} Z_1 N_1^2 \geq \frac{5\ln(1/\delta)}{\varepsilon}.$$

Proof. Let $c_7 = 1/50$ and now fix c_6 large enough such that $qm/L \leq 1$ as described in Remark 14 and such that c_6 is greater than 8. Then we have with probability $\delta^{1/50}$ by either Lemma 13 (and accordingly small c_5) or Lemma 15 that

$$\frac{1}{q} Z_1 \geq \min\left(\frac{1}{c_5 \varepsilon}, \frac{c_8}{50\varepsilon\sqrt{c_5}}\right).$$

We now also choose c_5 so small that the above is greater than $2 \cdot 5\varepsilon^{-1}$.

Now using $\sqrt{1-x} \leq 1 - x/2$ for $x \leq 1$ and that $\delta \leq 1/4$ it follows by Lemma 16 that with probability $1 - \sqrt{1 - \exp(-\ln(1/\delta)/\pi)} \geq \delta^{1/\pi}/2 \geq \delta^{1/2+1/\pi}$, we have $N_1^2 \geq \ln(1/\delta)/2$.

Now since that Z_1 and N_1^2 are independent we conclude that with probability $\delta^{1/50+1/2+1/\pi}$ we have that

$$\frac{1}{q} Z_1 N_1^2 \geq \frac{2 \cdot 5\ln(1/\delta)}{2\varepsilon} = \frac{5\ln(1/\delta)}{\varepsilon},$$

which concludes the proof of Lemma 10 □

We now restate and prove Lemma 11.

Restatement of Lemma 11.

Lemma 11. *For* $0 < \varepsilon \leq 1/4$ *and* $0 < \delta \leq 1/8$ *we have with probability at least* $\delta^{1/8}$ *that*

$$\sum_{i=2}^{k} \frac{1}{q} Z_i N_i^2 \geq (1 - 3\varepsilon)(k-1).$$

Proof. Let $X = (1/q)\sum_{i=2}^{k} Z_i N_i^2 \overset{d}{=} (1/(mq))\sum_{i=2}^{k} b_i N_i^2$, where the b_i's are binomial random variables with m trails and success probability q, the N_i's are standard normal random variables and the b_i's and the N_i's are all independent of each other. We now notice since the $b_i N_i^2$'s are independent and identically distributed the variance of their sum i equal to $k-1$ times the variance of $b_2 N_2^2$:

$$\mathrm{Var}\left(X\right) = \frac{1}{\left(mq\right)^2}\sum_{i=2}^{k} \mathrm{Var}\left(b_i N_i^2\right) = \frac{k-1}{\left(mq\right)^2}\mathrm{Var}\left(b_2 N_2^2\right).$$

Now using the independence of b_2 and N_2 and that the forth moment of a standard normal distribution is 3, and that the first and second moment of a binomial random variable is respectively mq and $(mq)^2 + mq(1-q)$ we get that

$$\begin{aligned}\mathrm{Var}\left(b_2 N_2^2\right) &= \mathbb{E}\left[\left(b_2 N_2^2\right)^2\right] - E\left[\left(b_2 N_2^2\right)\right]^2 = \mathbb{E}\left[b_2^2\right]\mathbb{E}\left[N_2^4\right] - \left(\mathbb{E}\left[b_2\right]\mathbb{E}\left[N_2^2\right]\right)^2 \\ &= 3\left(\left(mq\right)^2 + mq(1-q)\right) - \left(mq\right)^2 = \left(mq\right)^2\left(2 + (1-q)/\left(mq\right)\right).\end{aligned}$$

Now plugging $\mathrm{Var}(b_2 N_2^2)$ back into the expression of $\mathrm{Var}\left(X\right)$, yields that

$$\mathrm{Var}\left(X\right) = (k-1)\left(2 + (1-q)/\left(mq\right)\right).$$

Now using that $\mathbb{E}\left[X\right] = (k-1)$, the above calculation of the variance of X and Chebyshev-Cantelli's inequality $\mathbb{P}\left[Y - \mathrm{E}[Y] \leq -t\right] \leq \mathrm{Var}(Y)/\left(\mathrm{Var}(Y) + t^2\right)$ which holds for $t > 0$, yields that

$$\mathbb{P}\left[\sum_{i=2}^{k}\frac{1}{q}Z_iN_i^2 \le (1-3\varepsilon)(k-1)\right] \le \frac{(k-1)\left(2+(1-q)/(mq)\right)}{(k-1)\left(2+(1-q)/(mq)\right)+\left(3\varepsilon(k-1)\right)^2}$$
$$\le \frac{\left(2+(1-q)/(mq)\right)}{\left(2+(1-q)/(mq)\right)+(3\varepsilon)^2(k-1)}.$$

Since $y \to y/(y+a)$ is increasing in y for $a, y > 0$, it now follows using this with $a = (3\varepsilon)^3(k-1)$ and $y = 2 + (1-q)/(mq) \le 2 + 1 = 3$, where we have used that $1/(mq) \le 1$ by the comment under Equation (13), we get that

$$\mathbb{P}\left[\sum_{i=2}^{k}\frac{1}{q}Z_iN_i^2 \le (1-3\varepsilon)(k-1)\right] \le \frac{3}{3+(3\varepsilon)^2(k-1)}$$

Lastly using that $k = \ln(1/\delta)/\varepsilon^2$, $\varepsilon \le 1/4$ and $\delta \le 1/8$ we get $\varepsilon^2(k-1) = \ln(1/\delta) - \varepsilon^2 \ge 2$, and we conclude that

$$\mathbb{P}\left[\sum_{i=2}^{k}\frac{1}{q}Z_iN_i^2 \le (1-3\varepsilon)(k-1)\right] \le \frac{3}{3+18} \le 1-(1/8)^{1/8} \le 1-\delta^{1/8},$$

which completes the proof. □

C. Rademacher Entries in P

We sketch the proof of Theorem 1 for the case of P being populated with i.i.d. Rademacher entries. Namely, $P_{i,j} = b_{i,j}r_{i,j}/\sqrt{q}$, where the $r_{i,j}$'s are independent Rademacher variables, and $b_{i,j}$'s are Bernoulli with success probability q.

We will use the following upper bound (which is essentially a special case of the bound given in Lemma 3):

Lemma 17. *Let $a_{i,j} \in \mathbb{R}$ for $i = 1, \ldots, k$ and $j = 1, \ldots, d$ and $a_i = (a_{i,1}, \ldots, a_{i,d}) \in \mathbb{R}^d$. Let further $Y_i = \sum_{j=1}^{d} a_{i,j}r_{i,j}$, where the $r_{i,j}$ denote independent $\{-1, 1\}$-variables with mean 0. We then have*

$$\mathbb{P}\left[\,|\sum_{i=1}^{k}Y_i^2 - \|a_i\|^2\,| \ge x\right] \le 2\exp\left(-\frac{cx^2}{\sum_{i=1}^{k}\|a_i\|^4}\right),\ \ \textit{for } 0 < x \le \frac{16\sum_{i=1}^{k}\|a_i\|^4}{\max_{i=1,\ldots,k}\|a_i\|^2}$$
$$\mathbb{P}\left[\,|\sum_{i=1}^{k}Y_i^2 - \|a_i\|^2\,| \ge x\right] \le 2\exp\left(-\frac{cx}{\max_{i=1,\ldots,k}\|a_i\|^2}\right),\ \ \textit{for } x \ge \frac{16\sum_{i=1}^{k}\|a_i\|^4}{\max_{i=1,\ldots,k}\|a_i\|^2}$$

We note that the above lemma in general holds for any subgaussian random variables (note that Rademachers are indeed subgaussian), implying the statement of Theorem 1 holds for any subgaussian variables in entries of P. We give the proof of Lemma 17 for completeness at the end of this section.

Proof of Theorem 1, Rademacher entries. As in the case of Normal variables let $m = c_2 d/\ln n$ for a constant c_2. Further, let the embedding dimension $k = c_1\varepsilon^{-2}\ln n$, where c_1 is such that $c_1 \ge 1/c_2$, and let q the success probabilities of the binomial random variables $b_{i,j}$ in P be

$$q = \max\left\{c_1/m, c_1\varepsilon\min\left\{1, \ln(n)/(m\ln(1/\varepsilon))\right\}\right\}.$$

We assume that u is a vector in $\mathbb{R}^d$ such that $u_i^2 \le 1/m$ for all $i = 1, \ldots, d$ and $\|u\|^2 = 1$. Thus, the random variable of interest is

$$\|Pu\|^2 = \frac{1}{q}\sum_{i=1}^{k}\left(\sum_{j=1}^{d}u_jb_{i,j}r_{i,j}\right)^2.$$

Using Lemma 17 with $a_{i,j}$'s equal to $b_{i,j}u_j$ we have to consider $\|a_i\|^2 = \sum_{j=1}^{d}u_j^2b_{i,j}$'s which is what we denoted $Z_i = \sum_{j=1}^{d}u_j^2b_{i,j}$ in the upper bound proof with normal random variables. Going through the upper bound proof ones again and using Lemma 17 instead of Lemma 3 will result in exactly the same proof with the only difference in the constants. □

It remains to prove Lemma 17.

Proof of Lemma 17. First, note the we can assume without loss of generality that $\|a_i\|^2 > 0$ for all $1 \leq i \leq d$, since $\|a_i\|^2 = 0$ implies $Y_i = 0$ and the inequalities are trivially true if all $\|a_i\|^2 = 0$. Next, since $\exp(-y) + \exp(y) \leq 2\exp\left(x^2/2\right)$ for $y \in \mathbb{R}$, it follows that for $s \in \mathbb{R}$ we have

$$\mathbb{E}\left[\exp(sa_{i,j}r_{i,j})\right] \leq \left(\exp(-sa_{i,j}) + \exp(sa_{i,j})\right)/2 \leq \exp\left((sa_{i,j})^2/2\right)$$

for all i, j. Thus it follows by independence of the $r_{i,j}$ that for all i and $s \in \mathbb{R}$ we have that

$$\mathbb{E}\left[\exp(sY_i)\right] \leq \prod_{i=1}^{d} \exp\left((sa_{i,j})^2/2\right) = \exp\left(s^2 \|a_i\|^2/2\right).$$

Thus the Y_i's are independent sub-Gaussian random variable with variance proxy parameter $\|a_i\|^2$ as in the Definition 1.2 in (Rigollet & Hütter, 2015). By Lemma 1.12 in (Rigollet & Hütter, 2015) it follows that for $Y_i^2 - \mathbb{E}\left[Y_i^2\right]$, which by the independence of the $r_{i,j}$'s is equal to $Y_i^2 - \|a_i\|^2$, for $|s| \leq 1/(16\|a_i\|^2)$

$$\mathbb{E}\left[\exp\left(s\left(Y_i^2 - \|a_i\|^2\right)\right)\right] \leq \exp\left(16^2 s^2 \|a_i\|^4/2\right) \tag{16}$$

Thus if we consider $0 \leq s \leq 1/(16\max_{1,\ldots,k}\|a_i\|^2)$ we get that

$$\mathbb{E}\left[\exp\left(s\left(\sum_{i=1}^{k} Y_i^2 - \|a_i\|^2\right)\right)\right] \leq \exp\left(16^2 s^2 \sum_{i=1}^{k}\|a_i\|^4/2\right).$$

Now using Markov's inequality it follows that

$$\mathbb{P}\left[\sum_{i=1}^{k} Y_i^2 - \|a_i\|^2 \geq x\right] \leq \mathbb{E}\left[\exp\left(s\left(\sum_{i=1}^{k} Y_i^2 - \|a_i\|^2\right)\right)\right]\exp(-sx) \leq \exp\left(16^2 s^2 \sum_{i=1}^{k}\|a_i\|^4/2 - sx\right).$$

If $x/\left(16^2\sum_{i=1}^{k}\|a_i\|^4\right) \leq 1/(16\max_{1,\ldots,k}\|a_i\|^2)$ we set $s = x/\left(16^2\sum_{i=1}^{k}\|a_i\|^4\right)$ and get that

$$\mathbb{P}\left[\sum_{i=1}^{k} Y_i^2 - \|a_i\|^2 \geq x\right] \leq \exp\left(-x^2/\left(2\cdot 16^2\sum_{i=1}^{k}\|a_i\|^4\right)\right) \tag{17}$$

If $x/\left(16^2\sum_{i=1}^{k}\|a_i\|^4\right) \geq 1/(16\max_{1,\ldots,k}\|a_i\|^2) \geq s$ we get that $x \geq \left(16^2 s\sum_{i=1}^{k}\|a_i\|^4\right)$ implying that

$$\mathbb{P}\left[\sum_{i=1}^{k} Y_i^2 - \|a_i\|^2 \geq x\right] \leq \exp\left(16^2 s^2\sum_{i=1}^{k}\|a_i\|^4/2 - sx\right) \leq \exp\left(sx/2 - sx\right) = \exp\left(-sx/2\right).$$

Now choosing $s = 1/(16\max_{i=1,\ldots,k}\|a_i\|^2)$ we get

$$\mathbb{P}\left[\sum_{i=1}^{k} Y_i^2 - \|a_i\|^2 \geq x\right] \leq \exp\left(-x/\left(2\max_{i=1,\ldots,k}\|a_i\|^2\right)\right) \leq \exp\left(-x/\left(2\cdot 16^2\max_{i=1,\ldots,k}\|a_i\|^2\right)\right). \tag{18}$$

Due to Equation (16) holding true for $|s| \leq 1/(16\max_{1=,\ldots,k}\|a_i\|^2)$, we can carry the above arguments for $-(\sum_{i=1}^{k} Y_i^2 - \|a_i\|^2)$, and get the inequalities in Equations (17) and (18) for $\mathbb{P}\left[\sum_{i=1}^{k} Y_i^2 - \|a_i\|^2 \leq -x\right]$. Now setting $c = 1/(2\cdot 16^2)$ and union bounding over $\sum_{i=1}^{k} Y_i^2 - \|a_i\|^2 \leq -x$ and $\sum_{i=1}^{k} Y_i^2 - \|a_i\|^2 \geq x$ the claim follows. □