SIMPLE ANALYSIS OF SPARSE, SIGN-CONSISTENT JL

MEENA JAGADEESAN

ABSTRACT. Allen-Zhu, Gelashvili, Micali, and Shavit construct a sparse, sign-consistent Johnson-Lindenstrauss distribution, and prove that this distribution yields an essentially optimal dimension for the correct choice of sparsity. However, their analysis of the upper bound on the dimension and sparsity requires a complicated combinatorial graph-based argument similar to Kane and Nelson's analysis of sparse JL. We present a simple, combinatorics-free analysis of sparse, sign-consistent JL that yields the same dimension and sparsity upper bounds as the original analysis. Our proof also yields dimension/sparsity tradeoffs, which were not previously known.

As with previous proofs in this area, our analysis is based on applying Markov's inequality to the pth moment of an error term that can be expressed as a quadratic form of Rademacher variables. Interestingly, we show that, unlike in previous work in the area, the traditionally used Hanson-Wright bound is *not* strong enough to yield our desired result. Indeed, although the Hanson-Wright bound is known to be optimal for Gaussian degree-2 chaos, it was already shown to be suboptimal for Rademachers. Surprisingly, we are able to show a simple moment bound for quadratic forms of Rademachers that is sufficiently tight to achieve our desired result, which given the ubiquity of moment and tail bounds in theoretical computer science, is likely to be of broader interest.

1. INTRODUCTION

In many modern algorithms, which process high dimensional data, it is beneficial to pre-process the data through a dimensionality reduction scheme that preserves the geometry of the data. Such dimensionality reduction schemes have been applied in streaming algorithms[16] as well as algorithms in numerical linear algebra[22], graph sparsification[19], and many other areas.

The geometry-preserving objective can be expressed mathematically as follows. The goal is to construct a probability distribution $\mathscr{A}$ over $m \times n$ real matrices that satisfies the following condition for any $x \in \mathbb{R}^n$:

$$\mathbb{P}_{A \in \mathscr{A}}[(1-\varepsilon)||x||_2 \leq ||Ax||_2 \leq (1+\varepsilon)||x||_2] > 1-\delta. \tag{1}$$

An upper bound on the dimension m achievable by a probability distribution $\mathscr{A}$ that satisfies (1) is given in the following lemma, which is a central result in the area of dimensionality reduction:

Lemma 1.1 (Johnson-Lindenstrauss[10]**)** For any positive integer n and parameters $0 < \varepsilon, \delta < 1$, there exists a probability distribution $\mathscr{A}$ over $m \times n$ real matrices with $m = \Theta(\varepsilon^{-2}\log(1/\delta))$ that satisfies (1).

The optimality of the dimension m achieved by Lemma 1.1 was recently proven in [11, 9].

For many applications of dimensionality reduction schemes, it can be useful to consider probability distributions over sparse matrices in order to speed up the projection time. Here, sparsity refers to the constraint that there are a small number of nonzero entries in each column. In this context, Kane and Nelson[12] constructed a sparse JL distribution and proved the following:

Theorem 1.2 (Sparse JL[12]**)** For any positive integer n and $0 < \varepsilon, \delta < 1$, there exists a probability distribution $\mathscr{A}$ over $m \times n$ real matrices with $m = \Theta(\varepsilon^{-2}\log(1/\delta))$ and sparsity $s = \Theta(\varepsilon^{-1}\log(1/\delta))$ that satisfies (1).

Notice that this probability distribution, even with its sparsity guarantee, achieves the same dimension as Lemma 1.1. The proof of Theorem 1.2 presented in [12] involved complicated combinatorics; however, Cohen, Jayram, and Nelson[3] recently constructed two simple, combinatorics-free proofs of this result. The first approach, which is most relevant to this paper, used the Hanson-Wright bound on moments of quadratic forms. The second approach used the Chernoff-Rubin bound on exponential moments, and an analysis similar to the second approach can also be recovered by specializing the analysis of [2] for sparse oblivious subspace embeddings to the case of "1-dimensional subspaces." Although this recovered analysis is more complex, it has the advantage of yielding dimension-sparsity tradeoffs that were not produced through any of the previous approaches.

Neuroscience-based constraints give rise to the additional condition of sign-consistency on the matrices in the probability distribution. Sign-consistency refers to the constraint that the nonzero entries of each column are either all positive or all negative. The relevance of dimensionality reduction schemes in neuroscience is described in a survey by Ganguli and Sompolinsky[5]. In convergent pathways in the brain, information stored in a massive number of neurons is compressed into a small number of neurons, and nonetheless the ability to perform the relevant computations is preserved. Modeling this information compression scheme requires a hypothesis regarding what properties of the original information must be accurately transmitted to the receiving neurons. A plausible minimum requirement is that convergent pathways preserve the similarity structure of neuronal representations at the source area[1].

It remains to select the appropriate mathematical measure of similarity. The candidate similarity measure considered in [5] is vector inner product, which conveniently gives rise to a model based on the JL distribution[2]. Suppose there are n "input" neurons at a source area and m "output" neurons at a target area. In this framework, the information at the input neurons is represented as a vector in $\mathbb{R}^n$, the synaptic connections to output neurons are represented as a $m \times n$ matrix (with (i, j)th entry corresponding to the strength of the connection between input neuron j and output neuron i), and the information received by the output neurons is represented as a vector in $\mathbb{R}^m$. The similarity measure between two vectors v, w of neural information being taken to be $\langle v, w \rangle$ motivates modeling a synaptic connectivity matrix as a random $m \times n$ matrix drawn from a probability distribution that satisfies (1). Certain constraints on synaptic connectivity matrices arise from the biological limitations of neurons: the matrices must be *sparse* since a neuron is only connected to a small number (e.g. a few thousand) of postsynaptic neurons and *sign-consistent* since a neuron is usually purely excitatory or purely inhibitory.

This biological setting motivates the mathematical question: What is the optimal dimension and sparsity that can be achieved by a probability distribution over sparse, sign-consistent matrices that satisfies (1)? Related mathematical work includes, in addition to sparse JL[12], a construction of a dense, sign-consistent JL distribution[17, 6]. In [1], Allen-Zhu, Gelashvili, Micali, and Shavit constructed a sparse, sign-consistent JL distribution and proved the following upper bound:

Theorem 1.3 (Sparse, sign-consistent JL[1]**)** For every $\varepsilon > 0$, and $0 < \delta < 1/e$, there exists a probability distribution $\mathscr{A}$ over $m \times n$ real, sign-consistent matrices with $m = \Theta(\varepsilon^{-2} \log^2(1/\delta))$ and sparsity $s = \Theta(\varepsilon^{-1} \log(1/\delta))$ that satisfies (1).

In [1], it was also proven that the additional $\log(1/\delta)$ factor on m is essentially necessary: any distribution over sign-consistent matrices satisfying (1) requires $m = \tilde{\Omega}(\varepsilon^{-2} \log(1/\delta) \min(\log(1/\delta), \log n))$. Thus, the dimension in Theorem 1.3 is essentially optimal. However, in order to achieve this upper bound on m, the proof of Theorem 1.3 presented in [1] involved complicated combinatorics even more delicate than in the analysis of sparse JL in [12].

We present a simpler, combinatorics-free proof of Theorem 1.3. Our proof also yields dimension/sparsity tradeoffs, which were not previously known[3]:

Theorem 1.4 For every $\varepsilon > 0$, $0 < \delta < 1$, and $e \leq B \leq \frac{1}{\delta}$, there exists a probability distribution $\mathscr{A}$ over $m \times n$ real, sign-consistent matrices with $m = \Theta(B\varepsilon^{-2} \log_B^2(1/\delta))$ and sparsity $s = \Theta(\varepsilon^{-1} \log_B(1/\delta))$ that satisfies (1).

[1] This requirement is based on the experimental evidence that semantically similar objects in higher perceptual or association areas in the brain elicit similar neural activity patterns[13] and on the hypothesis that the similarity structure of the neural code is the basis of our ability to categorize objects and generalize responses to new objects[18].

[2] It is not difficult to see that for vectors x and y in the ℓ_2 unit ball, a $(1+\varepsilon)$-approximation of $\|x\|_2$, $\|y\|_2$, and $\|x-y\|_2$ implies an additive error $\Theta(\varepsilon)$ approximation of the inner product $\langle x, y \rangle$.

[3] In Appendix A, we point out the limiting lemma in the combinatorial analysis in [1], which prevents dimension-sparsity tradeoffs from being attainable through this approach. Our dimension-sparsity tradeoffs look similar to the Cohen's dimension-sparsity tradeoffs for sparse JL produced by specializing the analysis of [2] for sparse oblivious subspace embeddings to the case of "1-dimensional subspaces. For sparse JL, it is similarly not known how to obtain these tradeoffs via the combinatorial approach of [12].

Notice Theorem 1.3 is recovered if $B = e$; for larger B values, Theorem 1.4 enables a $\log B$ factor reduction in sparsity at the cost of a $B/\log^2 B$ factor gain in dimension.

As in [1, 12, 3], our analysis is based on applying Markov's inequality to the pth moment of an error term. Like in the first combinatorics-free analysis of sparse JL in [3], we express this error term as a quadratic form of Rademachers (uniform ± 1 random variables), and our analysis then boils down to analyzing the moments of this quadratic form. While the analysis in [3] achieves the optimal dimension for sparse JL using an upper bound on the moments of quadratic forms of subgaussians due to Hanson and Wright[7], we give a counterexample in Section 3.2 that shows that the Hanson-Wright bound is too loose in the sign-consistent setting to result in the optimal dimension. Since the Hanson-Wright bound is tight for quadratic forms of gaussians, we thus require a separate treatment of quadratic forms of Rademachers. We construct a simple bound on moments of quadratic forms of Rademachers that, unlike the Hanson-Wright bound, is sufficiently tight in our setting to prove Theorem 1.4. Our bound borrows some of the ideas from Latała's tight bound on the moments of quadratic forms of Rademachers[15]. Although our bound is much weaker than the bound in [15] in the general case, it has the advantage of providing a greater degree of simplicity by consisting of easier-to-analyze terms as well as permitting a cleaner proof, while still retaining the necessary precision to recover the optimal dimension for sparse, sign-consistent JL.

1.1. **Notation.** The main building blocks for our expressions are the following two types of random variables: Rademacher variables, which are uniform ± 1 random variables, and Bernoulli random variables, which have support $\{0,1\}$. For any random variable X and value $p \geq 1$, we use the notation $\|X\|_p$ to denote the p-norm $(\mathbb{E}[|X|^p])^{1/p}$, where $\mathbb{E}$ denotes the expectation. Similarly, for any random variable X and value $p \geq 1$ and any event E, we use the notation $\|X \mid E\|_p$ to denote the conditional p-norm $(\mathbb{E}[|X|^p \mid E])^{1/p}$, which is equivalent to the p-norm of the random variable $(X \mid E)$ since we only work over a finite probability space. We use the following notation to discuss certain asymptotics: given two scalar quantities Q_1 and Q_2 that are functions of some parameters, we use the notation $Q_1 \simeq Q_2$ to denote that there exist positive universal constants $C_1 \leq C_2$ such that $C_1 Q_2 \leq Q_1 \leq C_2 Q_2$, and we use the notation $Q_1 \lesssim Q_2$ to denote that there exists a positive universal constant C such that $Q_1 \leq C Q_2$.

1.2. **A digression on Rademachers versus Gaussians.** The concept that drives our moment bound can be illustrated in the linear form setting. Suppose $\sigma_1, \sigma_2 \ldots, \sigma_n$ are i.i.d Rademachers, $x = [x_1, \ldots, x_n]$ is a vector in $\mathbb{R}^n$ such that $|x_1| \geq |x_2| \geq \ldots \geq |x_n|$, and $2 \leq p \leq n$. The Khintchine inequality, which is tight for linear forms of Gaussians, yields the ℓ_2-norm bound $\|\sum_{i=1}^n \sigma_i x_i\|_p \lesssim \sqrt{p}\, \|x\|_2$. However, this bound *cannot* be a tight bound on $\|\sum_{i=1}^n \sigma_i x_i\|_p$ for the following reason: As $p \to \infty$, the quantity $\sqrt{p}\,\|x\|_2$ goes to infinity, while for any $p \geq 1$, the quantity $\|\sum_{i=1}^n \sigma_i x_i\|_p$ is bounded by $\|x\|_1$. Surprisingly, a result due to Hitczenko[8] indicates that the tight bound is actually the following combination of the ℓ_2 and ℓ_1 norm bounds:

$$\left\| \sum_{i=1}^n \sigma_i x_i \right\|_p \simeq \sum_{i=1}^p |x_i| + \sqrt{p} \sqrt{\sum_{i>p} x_i^2}.$$

In this bound, the "big" terms (i.e. terms involving x_1, x_2, $\ldots$, x_p) are handled with an ℓ_1-norm bound, while the remaining terms are approximated as Gaussians and bounded with an ℓ_2-norm bound.

A similar complication arises when the Hanson-Wright bound on quadratic forms of subgaussians is applied to Rademachers. Let σ be a d-dimensional vector of independent Rademachers, and let $A = (a_{k,l})$ be a symmetric $d \times d$ matrix with zero diagonal. The Hanson-Wright bound[7], which is tight for Gaussians, states for any $p \geq 1$,

$$\left\|\sigma^T A \sigma\right\|_p \lesssim \sqrt{p} \sqrt{\sum_{k=1}^d \sum_{l=1}^d a_{k,l}^2} + p \left(\sup_{\|y\|_2 = 1} |y^T A y| \right).$$

Similar to the linear form setting, this bound *can't* be a tight bound on $\left\|\sigma^T A\sigma\right\|_p$ for the following reason: As $p \to \infty$, the quantity $\sqrt{p}\sqrt{\sum_{k=1}^d \sum_{l=1}^d a_{k,l}^2}$ goes to ∞, while for any $p \geq 1$, the quantity $\left\|\sigma^T A\sigma\right\|_p$ is bounded by the entrywise ℓ_1-norm $\sum_{k=1}^d \sum_{l=1}^d |a_{k,l}|$.

Our quadratic form bound is based on a degree-2 analog of Hitczenko's observation. We analogously handle the "big" terms with an ℓ_1-norm bound and bound the remaining terms by approximating some of the Rademachers by Gaussians. From this, we obtain a combination of ℓ_2 and ℓ_1 norm bounds, similar to the linear form setting. Our simple bound has the surprising feature that it yields tighter guarantees than the Hanson-Wright bound yields for our error term. While our bound is much weaker than Latała's tight bound on the moments of quadratic forms of Rademachers in [15], it provides a greater degree of simplicity through avoiding an operator-norm-like term that is difficult to analyze in Latała's bound. Moreover, despite avoiding this term, our bound still retains the necessary precision to recover the optimal dimension for sparse, sign-consistent JL.

Although our final analysis follows a style that this is perhaps not as popular within the TCS community, in the end, it is quite simple, relying only on our quadratic form bound coupled with a few standard tricks such as repeated use of triangle inequalities on $||\cdot||_p$ norms as well as conditional $||\cdot||_p$ norms of random variables and standard moment bounds involving the binomial distribution (which we rederive in Appendices C-E for the sake of being self-contained). For this reason, we believe that it is likely to be of interest in other theoretical computer science settings involving moments or tail bounds of Rademacher forms.

1.3. **Outline for the rest of the paper.** In Section 2, we describe the construction and analysis of [1] for sparse, sign-consistent JL. In Section 3, we present Nelson's combinatorics-free approach for sparse JL that uses the Hanson-Wright bound, and we discuss why this approach does not yield the optimal dimension in the sign-consistent setting. In Section 4, we derive our bound on the moments of quadratic forms of Rademachers and use this bound to construct a combinatorics-free proof of Theorem 1.4.

2. Existing Analysis for Sparse, Sign-Consistent JL

In Section 2.1, we describe how to construct the probability distribution of sparse, sign-consistent matrices analyzed in Theorem 1.3. In Section 2.2, we briefly describe the combinatorial proof of Theorem 1.3 presented in [1].

2.1. **Construction of Sparse, Sign-Consistent JL.** The entries of a matrix $A \in \mathscr{A}$ are generated as follows[4]. Let $A_{i,j} = \eta_{i,j}\sigma_j/\sqrt{s}$ where $\{\sigma_i\}_{i\in[n]}$ and $\{\eta_{r,i}\}_{r\in[m],i\in[n]}$ are defined as follows:

- The families $\{\sigma_i\}_{i\in[n]}$ and $\{\eta_{r,i}\}_{r\in[m],i\in[n]}$ are independent from each other.
- The variables $\{\sigma_i\}_{i\in[n]}$ are i.i.d Rademachers.
- The variables $\{\eta_{r,i}\}_{r\in[m],i\in[n]}$ are identically distributed Bernoulli random variables with expectation s/m.
- The $\{\eta_{r,i}\}_{r\in[m],i\in[n]}$ are independent across columns but not independent within each column. For every column $1 \leq i \leq n$, it holds that $\sum_{r=1}^m \eta_{r,i} = s$. For every subset $S \subseteq [m]$ and every column $1 \leq i \leq n$, it holds that $\mathbb{E}\left[\prod_{r\in S}\eta_{r,i}\right] \leq \prod_{r\in S}\mathbb{E}[\eta_{r,i}]$. (One common definition of $\{\eta_{r,i}\}_{r\in[m],i\in[n]}$ that satisfies these conditions is the distribution defined by uniformly choosing exactly s of these variables per column to be a 1.)

For every $x \in \mathbb{R}^n$ such that $\|x\|_2 = 1$, we need to analyze an error term, which for this construction is the following random variable:

$$Z := \|Ax\|_2^2 - 1 = \frac{1}{s}\sum_{i\neq j}\sum_{r=1}^m \eta_{r,i}\eta_{r,j}\sigma_i\sigma_j x_i x_j.$$

[4]See Appendix B.1 for a formal construction of the probability space.

Proving that $\mathscr{A}$ satisfies (1) boils down to proving that $\mathbb{P}_{\eta,\sigma}[|Z| > \varepsilon] < \delta$. The main technique to prove this tail bound is the moment method. Bounding a large moment of Z is useful since it follows from Markov's inequality that

$$\mathbb{P}_{\eta,\sigma}[|Z| > \varepsilon] = \mathbb{P}_{\eta,\sigma}[|Z|^p > \varepsilon^p] < \frac{\mathbb{E}[|Z|^p]}{\varepsilon^p}.$$

The usual approach, used in the analyses in [1, 12, 3] as well as in our analysis, is to take $p = \Theta(\log(1/\delta))$ to be an even integer and analyze the p-norm $\|Z\|_p$ of the error term.

2.2. **Discussion of the combinatorial analysis of** [1]**.** In the analysis in [1], a complicated combinatorial argument was used to prove the following lemma, from which Theorem 1.3 follows:

Lemma 2.1 ([1]) If $s^2 \leq m$ and $p < s$, then $\|Z\|_p \lesssim \frac{p}{s}$.

The argument in [1] to prove Lemma 2.1 was based on expanding $\mathbb{E}[Z^p]$ into a polynomial with $\approx n^{2p}$ terms, establishing a correspondence between the monomials and the multigraphs, and then doing combinatorics to analyze the resulting sum. The approach of mapping monomials to graphs is commonly used in analyzing the eigenvalue spectrum of random matrices [21, 4] and was also used in [12] to analyze sparse JL. The analysis in [1] borrowed some methods from the analysis in [12]; however, the additional correlations between the Rademachers imposed by sign-consistency forced the analysis in [1] to require more delicate manipulations at several stages of the computation.

The expression to be analyzed was $s^p\mathbb{E}[Z^p]$, which was written as:

$$\sum_{i_1,\ldots,i_p,j_1,\ldots,j_p \in [n], i_1 \neq j_1,\ldots,i_p \neq j_p} \left(\prod_{u=1}^{p} x_{i_u} x_{j_u}\right) \left(\mathbb{E}_\sigma \prod_{u=1}^{p} \sigma_{i_u}\sigma_{j_u}\right) \left(\mathbb{E}_\eta \prod_{u=1}^{t} \sum_{r=1}^{m} \eta_{r,i_u}\eta_{r,j_u}\right).$$

After layers of computation, it was shown that

$$s^p\mathbb{E}[Z^p] \leq e^p \sum_{v=2}^{p} \sum_{G \in \mathscr{G}_{v,p}} \left((1/p^p) \prod_{q=1}^{v} \sqrt{d_q}^{d_q}\right) \sum_{r_1,\ldots,r_p \in [m]} \prod_{i=1}^{w} (s/m)^{v_i}$$

where $\mathscr{G}_{v,p}$ is a set of directed multigraphs with v labeled vertices and t labeled edges, where d_q is the total degree of vertex $q \in [v]$ in a graph $\mathscr{G}_{v,p}$, and where w and $v_1,\ldots,v_w$ are defined by G and the edge colorings $r_1,\ldots,r_t$. The problem then boiled down to carefully enumerating the graphs in $\mathscr{G}_{v,p}$ in six stages and analyzing the resulting expression.

3. Discussion of Combinatorics-Free Approaches

The main ingredient of the first combinatorics-free approach for sparse JL in [3] is the Hanson-Wright bound on the moments of quadratic forms of subgaussians. In Section 3.1, we discuss the approach in [3]. In Section 3.2, we discuss why this approach, if applied directly to sparse, sign-consistent JL, fails to yield the optimal dimension.

3.1. **Hanson-Wright approach for sparse JL in** [3]**.** The relevant random variable for sparse JL is

$$Z' = ||Ax||^2 - 1 = \frac{1}{s}\sum_{r=1}^{m}\sum_{i \neq j} \eta_{r,i}\eta_{r,j}\sigma_{r,i}\sigma_{r,j}x_i x_j$$

where the n independent Rademachers $\{\sigma_i\}_{i \in [n]}$ from the sign-consistent case are replaced by the mn independent Rademachers $\{\sigma_{r,i}\}_{i \in [n], r \in [m]}$. The main idea in [3] was to view Z' as a quadratic form $\frac{1}{s}\sigma^T A \sigma$. Here, σ is a mn-dimensional vector of independent Rademachers and $A = (a_{k,l})$ is a symmetric, zero diagonal, block diagonal $mn \times mn$ matrix with m blocks of size $n \times n$, where the (i,j)th entry (for $i \neq j$) of the rth block is $\eta_{r,i}\eta_{r,j}x_i x_j$. The quantity $\left\|\sigma^T A \sigma\right\|_p$ was analyzed using the Hanson-Wright bound. In order to bound

$\left\|\sigma^T A\sigma\right\|_p$, since A is a random matrix whose entries depend on the η values, an expectation had to be taken over η in the expression given by the Hanson-Wright bound. This resulted in the following:

$$\left\|\sigma^T A\sigma\right\|_p \lesssim \left\|\sqrt{p}\sqrt{\sum_{k=1}^{mn}\sum_{l=1}^{mn} a_{k,l}^2} + p\sup_{\|y\|_2=1}|y^T Ay|\right\|_p. \quad (2)$$

The analysis then boiled down to bounding the RHS of (2).

3.2. **Failure of the Hanson-Wright approach for sparse, sign-consistent JL.** The approach for sparse JL in [3] using the Hanson-Wright bound cannot be directly applied to the sign-consistent case to obtain a tight bound on $\|Z\|_p$. The loss arises from the fact that while the Hanson-Wright bound is tight for quadratic forms of gaussians, it is not guaranteed to be tight for quadratic forms of Rademachers. As discussed in Section 1.2, when $p \to \infty$, the Hanson-Wright bound goes to ∞, while $||\sigma^T A\sigma||_p$ can be bounded by the entrywise ℓ_1 norm of the matrix A. We give a counterexample, i.e. a vector x, that shows that the Hanson-Wright bound is too loose to give the optimal dimension for the sign-consistent case (when $\{\eta_{r,i}\}_{r\in[m],i\in[n]}$ are defined by uniformly choosing exactly s of the variables per column to be a 1). We present the details in Appendix F.

4. SIMPLE PROOF OF THEOREM 1.4

The main ingredient in our proof of Theorem 1.4 is the following bound on $\|Z\|_p$:

Lemma 4.1 Let $B = m/s^2$. If $2 \le p \le n$, then

$$\|Z\|_p \lesssim \begin{cases} \frac{p}{s\log B}, & \text{if } B \ge e \\ \frac{p}{sB} & \text{if } B < e. \end{cases}$$

We show in Section 4.4 that Theorem 1.4 follows from Lemma 4.1 via Markov's inequality.

In order to analyze $\|Z\|_p$, we view Z as a quadratic form $\frac{1}{s}\sigma^T A\sigma$, where the vector σ is an n-dimensional vector of independent Rademachers, and $A = (a_{i,j})$ is a symmetric, zero-diagonal $n \times n$ matrix where the (i, j)th entry (for $i \neq j$) is $x_i x_j \sum_{r=1}^m \eta_{r,i}\eta_{r,j}$. Since Z is symmetric in $x_1, \ldots, x_n$, we can assume WLOG that $|x_1| \ge |x_2| \ge \ldots \ge |x_n|$. For convenience, we define, like in [3],

$$Q_{i,j} := \sum_{r=1}^{m} \eta_{r,i}\eta_{r,j} \quad (3)$$

to be the number of collisions between the nonzero entries of the ith column and the nonzero entries of the jth column. Now, the (i, j)th entry of A (for $i \neq j$) can be rewritten as $Q_{i,j}x_i x_j$.

Our method to prove Lemma 4.1 resolves the issues imposed by directly applying the approach in [3]. We derive the following moment bound on quadratic forms of Rademachers[5] that yields tighter guarantees than the Hanson-Wright bound yields for $\|Z\|_p$:

Lemma 4.2 If $A = (a_{i,j})$ is a symmetric square $n \times n$ matrix with zero diagonal, $\{\sigma_i\}_{i\in[n]}$ is a set of independent Rademachers, and $q \ge 1$, then

$$\left\|\sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}\sigma_i\sigma_j\right\|_q \lesssim \left(\sum_{i=1}^{\min(q,n)}\sum_{j=1}^{\min(q,n)} |a_{i,j}|\right) + \sqrt{q}\sqrt{\sum_{i=1}^{n}\left\|\sum_{j>q} a_{i,j}\sigma_j\right\|_q^2}.$$

Observe that our bound avoids the weakness of the Hanson-Wright bound in the limit as $q \to \infty$. As discussed in Section 1.2, $\left\|\sum_{i=1}^n \sum_{j=1}^n a_{i,j}\sigma_i\sigma_j\right\|_q$ can be bounded by the entrywise ℓ_1-norm bound $\sum_{i=1}^n \sum_{j=1}^n |a_{i,j}|$ for any $q \ge 1$. While the Hanson-Wright bound goes to ∞ as $q \to \infty$, the bound in Lemma 4.2 approaches the entrywise ℓ_1 bound in the limit. This is because for $q > n$, the second term in Lemma 4.2 vanishes since

[5]Latała[15] provides a tight bound on the moments of $\sigma^T A\sigma$ (and on the moments of quadratic forms of more general random variables). However, his bound consists of terms that are difficult to analyze in this setting, and his proof is quite complicated, though the bound can be used in a black box to generate a much messier solution. Our simplified bound, though much weaker in the general case, consists of much easier-to-analyze terms and has a cleaner proof, while still being sufficiently tight for this setting.

the summand $\sum_{j>q}$ is empty. As a result, the bound becomes the first-term, which becomes $\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{i,j}|$ as desired.

For $1 \le q < n$, our bound becomes an interpolation of ℓ_1 and ℓ_2 norm bounds that bears resemblance to Hitczenko's Rademacher linear form bound in [8] discussed in Section 1.2. Although our bound is much weaker than the bound in [15], it is much simpler to analyze, especially in this setting. The main complication in the bound in [15] arises from the operator-norm-like term $\sup_{||y||_2=1, ||y||_\infty \le \frac{1}{\sqrt{q}}} |y^T A y|$. Due to the asymmetrical geometry of the ℓ_2 ball truncated by ℓ_∞ planes coupled with the fact that A is a random matrix in our setting, this term becomes especially messy. Our bound in Lemma 4.2 manages to avoid this term altogether. Moreover, our ℓ_1 norm term is straightforward to calculate, and our ℓ_2 norm term can be handled through a bound (Lemma C.1) from [14] that is useful in analyzing the q-norm $\left\|\sum_{j>q} a_{i,j}\sigma_j\right\|_q$ even when the $a_{i,j}$ are themselves random variables.

We defer our proof of Lemma 4.2 to Section 4.1. We now use Lemma 4.2 and the triangle inequality to obtain the following bound on $||Z||_p$:

$$\begin{aligned}
||Z||_p &= \frac{1}{s}\left(\mathbb{E}_\eta \mathbb{E}_\sigma \left[\sum_{i=1}^{n} \sum_{j\le n, j\ne i} Q_{i,j} x_i x_j \sigma_i \sigma_j\right]^p\right)^{1/p} \\
&\le \frac{1}{s}\left(\mathbb{E}_\eta \left[\sum_{i=1}^{p} \sum_{j\le p, j\ne i} |Q_{i,j} x_i x_j| + \sqrt{p}\sqrt{\sum_{i=1}^{n}\left(\mathbb{E}_\sigma\left[\sum_{j>p, j\ne i} Q_{i,j} x_i x_j \sigma_j\right]^p\right)^{2/p}}\,\right]^p\right)^{1/p}. \\
&\le \frac{1}{s}\left(\underbrace{\left\|\sum_{i=1}^{p} \sum_{j\le p, j\ne i} |Q_{i,j} x_i x_j|\right\|_p}_{(*)} + \underbrace{\sqrt{p}\sqrt{\sum_{i=1}^{n}\left\|\sum_{j>p, j\ne i} Q_{i,j} x_i x_j \sigma_j\right\|_p^2}}_{(**)}\right).
\end{aligned}$$

We first discuss some intuition for why using this bound avoids the issues of using the Hanson-Wright bound in producing a tight bound on $||Z||_p$. In the Hanson-Wright bound, all of the Rademachers are essentially approximated by Gaussians. In this bound, we instead make use of an ℓ_1-norm bound for $1 \le i \le p$ and $1 \le j \le p$ (the upper left $p \times p$ minor where the $|x_i|$ and $|x_j|$ values are the largest), which avoids the loss incurred in our setting by approximating the Rademachers in this range by Gaussians. Since the original matrix is symmetric, it only remains to consider $1 \le i \le n$ and $p+1 \le j \le n$. In this range, we approximate the σ_i Rademachers by Gaussians and use an ℓ_2-norm bound. Approximating the σ_j Rademachers by Gaussians as well would yield too loose of a bound for our application, so we preserve the σ_j Rademachers. For the remaining Rademacher linear forms, the interaction between the x_j values (all of which are upper bounded in magnitude by $\frac{1}{\sqrt{p}}$) and the σ_j Rademachers yields the desired bound.

In order to prove Lemma 4.1, it remains to prove Lemma 4.2 as well as to bound $(*)$ and $(**)$. In Section 4.1, we prove Lemma 4.2. The main ingredient in our bounds of $(*)$ and $(**)$ is an analysis of weighted sums of the $Q_{i,j}$ random variables, which form the building blocks for these expressions. In Section 4.2, we use the binomial-like properties of the $Q_{i,j}$s coupled with standard moment bounds involving the binomial distribution to analyze the moments of these weighted sums. In Section 4.3, we use these moment bounds to bound $(*)$ and $(**)$, and then finish our proof of Lemma 4.1. In Section 4.4, we show how Lemma 4.1 implies Theorem 1.4.

4.1. **Proof of Lemma** 4.2**.** We use the following standard lemmas in our proof of Lemma 4.2.

The first lemma allows us to decouple the two sets of Rademachers in our quadratic form so that we can reduce analyzing the moments of the quadratic form to analyzing the moments of a linear form.

Lemma 4.3 (Decoupling, Theorem 6.1.1 of [20]**)** If $A = (a_{i,j})$ is a symmetric, zero-diagonal $n \times n$ matrix and $\{\sigma_i\}_{i\in[n]} \cup \{\sigma_i'\}_{i\in[n]}$ are independent Rademachers, then

$$\left\| \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}\sigma_i\sigma_j \right\|_q \lesssim \left\| \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}\sigma_i'\sigma_j \right\|_q .$$

The next lemma is due to Khintchine and gives an ℓ_2-norm bound on linear forms of Rademachers. Since the Khintchine bound is derived from approximating $\sigma_1, \ldots, \sigma_n$ by i.i.d Gaussians, we only use this bound outside of the delicate upper left $p \times p$ minor of our matrix A.

Lemma 4.4 (Khintchine) If $\sigma_1, \sigma_2, \ldots, \sigma_n$ are independent Rademachers, then for all $q \geq 1$ and $a \in \mathbb{R}^n$,

$$\left\| \sum_{i=1}^{n} \sigma_i a_i \right\|_q \lesssim \sqrt{q}||a||_2.$$

Now, we are ready to prove Lemma 4.2.

Proof of Lemma 4.2. By Lemma 4.3 and the triangle inequality, we know

$$\left\| \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}\sigma_i\sigma_j \right\|_q \lesssim \underbrace{\left\| \sum_{i=1}^{\min(q,n)} \sum_{j=1}^{\min(q,n)} a_{i,j}\sigma_i'\sigma_j \right\|_q}_{\alpha} + \underbrace{\left\| \sum_{i=1}^{n} \sum_{j>q} a_{i,j}\sigma_i'\sigma_j \right\|_q}_{\beta} + \underbrace{\left\| \sum_{i>q} \sum_{j=1}^{q} a_{i,j}\sigma_i'\sigma_j \right\|_q}_{\gamma} .$$

We first bound α. Since a Rademacher σ satisfies $|\sigma| = 1$, it follows that α can be upper bounded by the entrywise ℓ_1-norm bound $\sum_{i=1}^{\min(q,n)} \sum_{j=1}^{\min(q,n)} |a_{i,j}|$ as desired. We now bound β using Lemma 4.4 to obtain

$$\beta \leq \sqrt{q} \left\| \sqrt{\sum_{i=1}^{n} \left(\sum_{j>q} a_{i,j}\sigma_j \right)^2} \right\|_q = \sqrt{q} \sqrt{\left\| \sum_{i=1}^{n} \left(\sum_{j>q} a_{i,j}\sigma_j \right)^2 \right\|_{q/2}} \leq \sqrt{q} \sqrt{\sum_{i=1}^{n} \left\| \sum_{j>q} a_{i,j}\sigma_j \right\|_q^2}.$$

We now bound γ. From an analogous computation, it follows that $\gamma \leq \sqrt{q}\sqrt{\sum_{j=1}^{q} \left\| \sum_{i>q} a_{i,j}\sigma_i \right\|_q^2}$, which implies that $\gamma \leq \sqrt{q}\sqrt{\sum_{j=1}^{n} \left\| \sum_{i>q} a_{i,j}\sigma_i \right\|_q^2} = \sqrt{q}\sqrt{\sum_{i=1}^{n} \left\| \sum_{j>q} a_{i,j}\sigma_j \right\|_q^2}$. □

4.2. **Moments of Weighted Sums of $Q_{i,j}$ Random Variables.** Recall that for $1 \leq i \neq j \leq n$, the $Q_{i,j}$ random variables count the number of collisions between the nonzero entries in the ith column and jth column. We first prove that these random variables satisfy conditional independence properties when i is fixed. We also show that the moments of the random variables obtained through this conditioning are bounded by binomial moments.

Proposition 4.5 Let X be a random variable distributed as $\text{Bin}(s, s/m)$. For any i, given any choice of s nonzero rows $r_1 \neq r_2 \neq \ldots \neq r_s$ in the ith column, the set of $n-1$ random variables[6] $(Q_{i,j} \mid \eta_{r_1,i} = \eta_{r_2,i} = \ldots = \eta_{r_s,i} = 1)$ for $1 \leq j \leq n, j \neq i$ are independent. Moreover, for any $q \geq 1$ and any $i \neq j$:

$$\left\| Q_{i,j} \mid \eta_{r_1,i} = \eta_{r_2,i} = \ldots = \eta_{r_s,i} = 1 \right\|_q \leq \|X\|_q .$$

The independence properties essentially follow from the fact that the nonzero entries in different columns are independent. Moreover, the fact that $Q_{i,j}$ moments are upper bounded by binomial moments follows from the sums of independent Bernoulli random variables being binomial random variables and negative correlation only decreasing moments.

[6]See Appendix B for a formal discussion of viewing these quantities as random variables over a different probability space.

Proof of Proposition 4.5. Let A be a matrix drawn from $\mathscr{A}$, and pick any $1 \le i \le n$. We condition on the event that the s nonzero entries in column i of A occur at rows $r_1, \dots, r_s$. For $1 \le j \le n, j \neq i$ and $1 \le k \le s$, let $Y_{k,j} = \eta_{r_k,j}$, so that $(Q_{i,j} \mid \eta_{r_1,i} = \eta_{r_2,i} = \ldots = \eta_{r_s,i}) = \sum_{k=1}^{s} Y_{k,j}$. Notice that the sets $\{Y_{k,j}\}_{k \in [s]}$ for $1 \le j \le n, j \neq i$ are independent from each other, which means random variables in the set $\{Q_{i,j} \mid \eta_{r_1,i} = \eta_{r_2,i} = \ldots = \eta_{r_s,i} = 1\}_{1 \le j \le n, j \neq i}$ are independent. For $1 \le j \le n, j \neq i$, and $1 \le k \le s$, let $Z_{k,j}$ be distributed as i.i.d Bernoulli random variables with expectation s/m. Notice that for a fixed j, each $Y_{k,j}$ is distributed as $Z_{k,j}$ and the random variables $\{Y_{k,j}\}_{1 \le k \le s}$ are negatively correlated (and nonnegative), which means $\|Q_{i,j} \mid \eta_{r_1,i} = \eta_{r_2,i} = \ldots = \eta_{r_s,i} = 1\|_q = \|\sum_{k=1}^{s} Y_{k,j}\|_q \le \|\sum_{k=1}^{s} Z_{k,j}\|_q = \|X\|_q$. □

Now, we need to analyze the moments of weighted sums of $Q_{i,j}$ random variables. Using the independence properties and the fact that the moments of the $Q_{i,j}$ are upper bounded by binomial moments as given in Proposition 4.5, this boils down to studying the moments of weighted sums of binomial random variables. The main tools that we use in analyzing these moments are bounds on moments of sums of nonnegative random variables and sums of symmetric random variables due to Latała[14]. The relevant bounds and proofs in [14] are not complicated; for the sake of being self-contained, we state and sketch proofs of these bounds in Appendix C.

Our first estimate is an upper bound on the moments of binomial random variables, which also gives bounds on moments of the $Q_{i,j}$ by Proposition 4.5. We defer the proof to Appendix E.

Proposition 4.6 Suppose that X is a random variable distributed as $\mathrm{Bin}(N, \alpha)$ for any $\alpha \in (0,1)$ and any integer $N \ge 1$. If $q \ge 1$ and $B = \frac{q}{\alpha \max(N,q)}$, then

$$\|X\|_q \lesssim \begin{cases} \frac{q}{\log B} & \text{if } B \ge e \\ \frac{q}{B} & \text{if } B < e \end{cases}.$$

Our next estimate is essentially an upper bound on the moments of sums of binomial random variables weighted by Rademachers. We defer the proof to Appendix D.

Proposition 4.7 Suppose that $q \ge 2$ is even and $y = [y_1, \dots, y_M]$ is a vector that satisfies $||y||_2 \le 1$ and $||y||_\infty \le \frac{1}{\sqrt{q}}$. Let X be a random variable distributed as $\mathrm{Bin}(N, \alpha)$ for some $\alpha \in (0,1)$ and some integer $N \ge 1$. Suppose that $Y_1, \dots, Y_M$ are independent random variables that satisfy $||Y_k||_l \le ||X||_l$ for $1 \le k \le M$ and for $l \ge 1$. Suppose that $\sigma_1, \dots, \sigma_M$ are independent Rademachers that are also independent of $\{Y_k\}_{k \in [M]}$. If $B = \frac{q}{\alpha \max(N,q)}$, then

$$\left\| \sum_{k=1}^{M} Y_k y_k \sigma_k \right\|_q \lesssim \begin{cases} \frac{\sqrt{q}}{\log B} & \text{if } B \ge e \\ \frac{\sqrt{q}}{B} & \text{if } B < e \end{cases}.$$

4.3. **Bounding $(*)$ and $(**)$ to prove Lemma** 4.1**.** We bound the quantities $(*)$ and $(**)$ in the following sublemmas, which assume the notation used throughout the paper:

Lemma 4.8 If $m/s^2 = B$, then

$$\left\| \sum_{i=1}^{p} \sum_{j \le p, j \neq i} |Q_{i,j} x_j x_i| \right\|_p \lesssim \begin{cases} \frac{p}{\log B} & \text{if } B \ge e \\ \frac{p}{B} & \text{if } B < e \end{cases}.$$

Lemma 4.9 If $m/s^2 = B$, then

$$\sqrt{p} \sqrt{\sum_{i=1}^{n} \left\| \sum_{j > p, j \neq i} Q_{i,j} x_i x_j \sigma_j \right\|_p^2} \lesssim \begin{cases} \frac{p}{\log B} & \text{if } B \ge e \\ \frac{p}{B} & \text{if } B < e \end{cases}.$$

We now use Proposition 4.5 as well as the moment bound on binomial random variables from Proposition 4.6 to prove Lemma 4.8 and thus bound $(*)$.

Proof of Lemma 4.8. We carefully use the triangle inequality to see[7]:

$$\left\|\sum_{i=1}^{p}\sum_{j\le p, j\ne i}|Q_{i,j}x_jx_i|\right\|_p \le 2\left\|\sum_{i=1}^{p}\sum_{i<j\le p}Q_{i,j}|x_j||x_i|\right\|_p \le 2\left\|\sum_{i=1}^{p}x_i^2\sum_{i<j\le p}Q_{i,j}\right\|_p \le 2\sum_{i=1}^{p}x_i^2\left\|\sum_{i<j\le p}Q_{i,j}\right\|_p .$$

Let $X \sim \mathrm{Bin}(s, s/m)$ and $Y \sim \mathrm{Bin}(sp, s/m)$. By Proposition 4.5, for any i and any $r_1 \ne r_2 \ne \ldots \ne r_s$, the random variables $\left\{Q_{i,j} \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1\right\}_{j\ne i}$ are independent and $\left\|Q_{i,j} \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1\right\|_p \le \|X\|_p$. It follows from taking pth powers of both sides that

$$\left\|\left(\sum_{i<j\le p}Q_{i,j}\right) \mid \eta_{r_1,i} = \eta_{r_2,i} = \ldots = \eta_{r_s,i} = 1\right\|_p = \left\|\sum_{i<j\le p}(Q_{i,j} \mid \eta_{r_1,i} = \eta_{r_2,i} = \ldots = \eta_{r_s,i} = 1)\right\|_p \le \|Y\|_p .$$

Now, Proposition 4.6 gives us a bound on $\|Y\|_p$, and the result follows from the law of total expectation[8]. □

We now use Proposition 4.5 as well as the moment bound on weighted sums of binomial random variables from Proposition 4.7 to prove Lemma 4.9 and thus bound $(**)$.

Proof of Lemma 4.9. Observe that

$$\sqrt{p}\sqrt{\sum_{i=1}^{n}\left\|\sum_{j>p, j\ne i}Q_{i,j}x_ix_j\sigma_j\right\|_p^2} = \sqrt{p}\sqrt{\sum_{i=1}^{n}x_i^2\left\|\sum_{j>p, j\ne i}Q_{i,j}x_j\sigma_j\right\|_p^2} \le \sqrt{p}\max_{1\le i\le n}\left\|\sum_{j>p, j\ne i}Q_{i,j}x_j\sigma_j\right\|_p .$$

Let $X \sim \mathrm{Bin}(s, s/m)$ and $Y \sim \mathrm{Bin}(sp, s/m)$. By Proposition 4.5, for any i and any $r_1 \ne r_2 \ne \ldots \ne r_s$, the random variables $\left\{Q_{i,j} \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1\right\}_{j\ne i}$ are independent and $\left\|Q_{i,j} \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1\right\|_p \le \|X\|_p \le \|Y\|_p$. Moreover, $|x_j| \le \frac{1}{\sqrt{p}}$ for $j > p$. Now, we consider $\left\|\sum_{j>p, j\ne i}Q_{i,j}x_j\sigma_j \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1\right\|_p$ which is equal[6] to $\left\|\sum_{j>p, j\ne i}(Q_{i,j} \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1)(\sigma_j \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1)x_j\right\|_p$. Since each $(\sigma_j \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1)$ is distributed as a Rademacher and since the set of $n-1$ random variables $\left\{\sigma_j \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1\right\}_{j\ne i}$ are independent and also independent of $\left\{Q_{i,j} \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1\right\}_{j\ne i}$, we can apply Proposition 4.7 to $\left\|\sum_{j>p, j\ne i}(Q_{i,j} \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1)(\sigma_j \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1)x_j\right\|_p$ and thus get a bound[9] on the conditional p-norm $\left\|\sum_{j>p, j\ne i}Q_{i,j}x_j\sigma_j \mid \eta_{r_1,i} = \ldots = \eta_{r_s,i} = 1\right\|_p$. Now, the result follows from the law of total expectation[8]. □

We now show the bound on $||Z||_p$ follows from the bounds on $(*)$ and $(**)$ in Lemmas 4.8, 4.9.

Proof of Lemma 4.1. Applying Lemmas 4.8, 4.9 after the following simplification proves the lemma:

$$\|Z\|_p \lesssim \frac{1}{s}\left\|\sum_{i=1}^{p}\sum_{j\le p, j\ne i}|Q_{i,j}x_ix_j|\right\|_p + \frac{\sqrt{p}}{s}\sqrt{\sum_{i=1}^{n}\left\|\sum_{j>p, j\ne i}Q_{i,j}x_ix_j\sigma_j\right\|_p^2}.$$

□

4.4. **Proof of Theorem** 1.4. We show Lemma 4.1 implies Theorem 1.4, completing the proof.

Proof of Theorem 1.4. It suffices to show $\mathbb{P}_{\eta,\sigma}[|Z| > \varepsilon] < \delta$. By Markov's inequality, we know

$$\mathbb{P}_{\eta,\sigma}[|Z| > \varepsilon] = \mathbb{P}_{\eta,\sigma}[|Z|^p > \varepsilon^p] < \varepsilon^{-p}\mathbb{E}[|Z|^p] = \left(\frac{\|Z\|_p}{\varepsilon}\right)^p .$$

[7]Naively applying the triangle inequality yields a suboptimal bound, so we require this more careful treatment.

[8]See Appendix B for a formal discussion of why a uniform bound on the conditional p-norm implies a bound on the p-norm here.

[9]Approximating the σ_j by Gaussians yields a suboptimal bound, so we require the bound given in Proposition 4.7.

Suppose that $B \geq e$. Then by Lemma 4.1, we know

$$\left(\frac{\|Z\|_p}{\varepsilon}\right)^p \leq \left(\frac{Cp}{(\log B)s\varepsilon}\right)^p.$$

Thus, to upper bound this quantity by δ, we can set $s = \Theta(\varepsilon^{-1}p/\log B) = \Theta(\varepsilon^{-1}\log_B(1/\delta))$ and $m = \Theta(Bs^2)$. We impose the additional constraint that $B \leq \frac{1}{\delta}$ to guarantee that $s \geq 1$. This proves the desired result[10]. □

5. Acknowledgments

I would like to thank Prof. Jelani Nelson for proposing and advising this project.

References

[1] Z. Allen-Zhu, R. Gelashvili, S. Micali, and N. Shavit. Sparse sign-consistent Johnson–Lindenstrauss matrices: Compression with neuroscience-based constraints. In *Proceedings of the National Academy of Sciences*, volume 111, pages 16872–16876, 2014.

[2] M. B. Cohen. Nearly tight oblivious subspace embeddings by trace inequalities. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 278–287, 2016.

[3] Michael B. Cohen, T. S. Jayram, and Jelani Nelson. Simple analyses of the sparse Johnson-Lindenstrauss transform. In *1st Symposium on Simplicity in Algorithms, SOSA 2018, January 7-10, 2018, New Orleans, LA, USA*, pages 15:1–15:9, 2018.

[4] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 62:233–241, 1981.

[5] S. Ganguli and H. Sompolinsky. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual Review of Neuroscience*, 35:485–508, 2012.

[6] R.T. Gray and P.A. Robinson. Stability and structural constraints of random brain networks with excitatory and inhibitory neural populations. *Journal of Computational Neuroscience*, 27(1):81–101, 2009.

[7] D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.

[8] P. Hitczenko. Domination inequality for martingale transforms of Rademacher sequence. *Israel Journal of Mathematics*, 84:161–178, 1993.

[9] T.S. Jayram and D. P. Woodruff. Optimal bounds for Johnson-Lindenstrauss transforms and steaming problems with subconstant error. In *ACM Transactions on Algorithms (TALG) - Special Issue on SODA'11*, volume 9, pages 1–26, 2013.

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

[11] D. M. Kane, R. Meka, and J. Nelson. Almost optimal explicit Johnson-Lindenstrauss families. In *Proceedings of the Fourteenth International Workshop and Fifteenth International Conference on Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques*, pages 628–639, 2011.

[12] D. M. Kane and J. Nelson. Sparser Johnson-Lindenstrauss transforms. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 16872–16876. ACM Press, 2012.

[13] R. Kiani, H. Esteky, K. Mirpour, and K. Tanaka. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, 97:4296–4309, 2007.

[14] R. Latała. Estimation of moments of sums of independent real random variables. *Ann. Probab.*, 25(3):1502–1513, 07 1997.

[15] R. Latała. Tail and moment estimates for some types of chaos. *Studia Mathematica*, 135(1):39–53, 1999.

[16] S. Muthukrishnan. Data streams: Algorithms and applications. *Foundations and Trends in Theoretical Computer Science*, 1, 2005.

[17] K. Rajan and L.F. Abbot. Eigenvalue spectra of random matrices for neural networks. *Physical Review Letters*, 97:188104, 2006.

[18] T. Rogers and J. McClelland. *Semantic Cognition: A Parallel Distributed Processing Approach*. MIT Press, 2004.

[19] D. Spielman and N. Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing (SICOMP)*, 40:1913–1926, 2011.

[20] R. Vershynin. *High-Dimensional Probability*. Cambridge University Press, To appear. Available at `https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.html`.

[21] E.P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62:548–564, 1955.

[22] D.P. Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10:1–157, 2014.

[10] If we set $B < e$, if we use Lemma 4.1, we know that in order to obtain an upper bound of δ, we would have to set $s = \Theta(\varepsilon^{-1}p/B) = \Theta(\varepsilon^{-1}\log(1/\delta)/B)$ and $m = \Theta(\varepsilon^{-1}\log^2(1/\delta)/B)$. This yields no better s or m values than those achieved when $B = e$.

Appendix A. Limitations of the Combinatorial Approach

The limiting factor in the combinatorial approach of [1] is the lemma that states that

$$s^t \mathbb{E}[Z^t] \le 2^{O(t)} t^t \left(\frac{s^2}{m}\right)^t .$$

This is Lemma 3 in the conference version of [1], and Lemma 4.3 in the arXiv version of [1]. Here, Z is defined analogously as in section 2.1 of this paper, and the t value is equivalent to the p value in this paper.

The proof of this lemma, given in Appendix A.3 in [1], implicitly relies on the fact that $\frac{s^2}{m} \ge 1$, although this condition is not explicitly stated in the lemma statement. This assumption arises from the last line of the proof, where the sum $\sum_{w=1}^{t} \left(\frac{s^2}{m}\right)^w$ is upper bounded by $t\left(\frac{s^2}{m}\right)^t$. Following the end of the proof of Theorem 1 (the top of p. 9 of the arXiv version), this yields $\mathbb{P}[|Z| > \varepsilon] \le \left(\frac{Cts}{\varepsilon m}\right)^t$. Now, suppose we instead set $m = Bs^2$ (where $B \le 1$ as required by the assumption). Then we obtain $\left(\frac{Cts}{\varepsilon m}\right)^t = \left(\frac{Ct}{\varepsilon Bs}\right)^t$. Thus, we can set s to be $C\varepsilon^{-1}B^{-1}\log(1/\delta)$ and m to be $C^2\varepsilon^{-2}\log^2(1/\delta)B^{-1}$. Since $B \le 1$, this is no better than the original theorem statement and thus yields no dimension-sparsity tradeoff.

Now, suppose we instead let $\frac{s^2}{m} \le 1$. Then we can modify the proof of Lemma 4.3 to obtain the weaker upper bound of $\sum_{w=1}^{t} \left(\frac{s^2}{m}\right)^w$ by $t\frac{s^2}{m}$. Let $B = m/s^2$ where $B \ge 1$. In order to ensure that m is polynomial in $\log(1/\delta)$, we assume that $B \le \delta$. In this case, mimicking the calculation at the end of the proof of Theorem 1, we obtain $\mathbb{P}[|Z| > \varepsilon] \le \frac{1}{B}\left(\frac{Ct}{\varepsilon s}\right)^t = \left(\frac{Ct}{\varepsilon s B^{1/t}}\right)^t$. Thus, we can set $s = C\log(1/\delta)\varepsilon^{-1}e^{-\log B/t}$. Observe that $0 \le \log B \le t$, so $1 \ge e^{-\log B/t} \ge e^{-1}$. Thus, $s = \Theta(\log(1/\delta)\varepsilon^{-1})$ and $m = \Theta(Bs^2)$, which does not yield a dimension-sparsity tradeoff.

Thus, it is not clear how to directly obtain the dimension-sparsity tradeoff from the combinatorial approach of [1]. Some intuition for this limitation is that the moment bounds on Z obtained by the combinatorial approach are not sufficiently tight for varying values of B due to the fact that the bounding techniques are implicitly tailored to the case of $B = \Theta(1)$. The combinatorics-free approach in this paper avoids this issue through making use of a more structured method to bound the moments of Z.

Appendix B. Formal Definition of Probability Space and Random Variables

B.1. **Construction of the Probability Space.** Rather than defining our probability space over $m \times n$ sparse, sign-consistent matrices implicitly by random variables, we explicitly define this (finite) probability space here. We take our probability space to be the product space of n probability spaces $(\Omega_i, \Sigma_i, \mathbb{P}_i)$ where Ω_i is the set of m-dimensional column vectors with entries in $\{-1, 0, 1\}$ with exactly s nonzero entries and all nonzero entries the same sign. The sigma algebra Σ_i for $1 \le i \le n$ is all subsets of Ω_i. Thus, the product $\Omega := \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$ is isomorphic to the sample space of $m \times n$ dimensional sign-consistent matrices with entries in $\{-1, 0, 1\}$ with exactly s nonzero entries in each column, and the product $\Sigma := \Sigma_1 \otimes \Sigma_2 \otimes \ldots \otimes \Sigma_n$ consists of all subsets of Ω. Now, we define the probability measure $\mathbb{P}_i$ for each $1 \le i \le n$, which implicitly defines the product probability measure $\mathbb{P}$.

We decompose $\mathbb{P}_i$ as a product of probability spaces $(\Omega_i^1, \Sigma_i^1, \mathbb{P}_i^1) \times (\Omega_i^2, \Sigma_i^2, \mathbb{P}_i^2)$ which will separate sign from the choice of nonzero entries. Here, Ω_i^1 is the set $\{-1, 1\}$ and Ω_i^2 is the set of binary m-dimensional column vectors. Again, Σ_i^1 and Σ_i^2 are the set of all subsets of Ω_i^1 and Ω_i^2 respectively. For an element $\omega_i^1 \in \Omega_i^1$ and $\omega_i^2 \in \Omega_i^2$, we view the element $\omega_i^1 \times \omega_i^2$ as the product of the scalar ω_i^1 and the vector ω_i^2. Now, we define $\mathbb{P}_i^1(-1) = \mathbb{P}_i^1(1) = 0.5$. We take $\mathbb{P}_i^2$ to be any probability measure that satisfies the following negative correlation property: For any m-dimensional vector and any $1 \le j \le m$, let $f^j(x)$ be an indicator function for the jth coordinate being nonzero. We assume that for any subset $S \subseteq [m]$:

$$\sum_{\omega_i^2 \in \Omega_i^2} \mathbb{P}[\omega_i^2] \left(\prod_{j \in S} f^j(\omega_i^2)\right) \le \prod_{j \in S} \left(\sum_{\omega_i^2 \in \Omega_i^2} \mathbb{P}[\omega_i^2] f^j(\omega_i^2)\right).$$

Let $\sigma_i : \Omega \to \{-1,1\}$ be the random variable for the sign of the ith column. Let $\eta_{r,i} : \Omega \to \{-1,1\}$ be the indicator random variable for whether the (r,i)th entry is nonzero. Observe that the independence and negative correlation properties are recovered exactly.

B.2. **Explanation of Proposition** 4.5**.** As before, for $i \neq j$, we define $Q_{i,j} = \sum_{r=1}^m \eta_{r,i}\eta_{r,j}$. For some $r_1 \neq r_2 \neq \ldots \neq r_s$, let $A_{r_1,\ldots,r_s} \subseteq \Omega$ be the event that the nonzero entries in the ith column occur at $r_1,\ldots,r_s$. Now, observe that $(Q_{i,j} \mid A_{r_1,\ldots,r_s})$ is a random variable defined by the restriction of the function $Q_{i,j} : \Omega \to \mathbb{R}$ to a function $A_{r_1,\ldots,r_s} \to \mathbb{R}$ on the probability space $(A_{r_1,\ldots,r_s}, 2^{A_{r_1,\ldots,r_s}}, \mathbb{P}|_A)$ where for $B \subseteq A_{r_1,\ldots,r_s}$, we define $\mathbb{P}|_A(B)$ to be $\frac{\mathbb{P}(B)}{\mathbb{P}(A_{r_1,\ldots,r_s})}$. This formally defines the random variable $(Q_{i,j} \mid \eta_{r_1,i} = \eta_{r_2,i} = \ldots = \eta_{r_s,i} = 1)$ in Proposition 4.5, and observe that the proof proceeds by computing the distribution of this random variable. In this notation, the proposition shows that the random variables $(Q_{i,j} \mid A_{r_1,\ldots,r_s})$ are independent on this new probability space.

B.3. **Explanation of Lemma** 4.8**.** For Lemma 4.8, the random variable $\left(\sum_{i<j\leq p} Q_{i,j}\right) \mid A_{r_1,\ldots,r_s}$ is defined analogously on the restricted probability space. It follows from definition that this random variable is equivalent to the sum of random variables $\sum_{i<j\leq p}(Q_{i,j} \mid A_{r_1,\ldots,r_s})$. Since we know that $(Q_{i,j} \mid A_{r_1,\ldots,r_s})$ for $i < j \leq p$ are independent and have moments bounded by binomials by Proposition 4.5, this implies that we can apply Proposition 4.6 to $\sum_{i<j\leq p}(Q_{i,j} \mid A_{r_1,\ldots,r_s})$ to obtain a bound on the p-norm $\mathbb{E}\left[\left(\sum_{i<j\leq p} Q_{i,j} \mid A_{r_1,r_2,\ldots,r_s}\right)^p\right]^{1/p}$ which we can convert to a bound on the conditional p-norm $\mathbb{E}\left[\left(\sum_{i<j\leq p} Q_{i,j}\right)^p \mid A_{r_1,r_2,\ldots,r_s}\right]^{1/p}$. Let's suppose this bound is α. Since the events $\{A_{r_1,\ldots,r_s}\}_{r_1\neq r_2\neq\ldots\neq r_s}$ form a disjoint partition of the probability space, we know by the law of total expectation that:

$$\alpha^p \geq \sum_{r_1\neq r_2\neq\ldots\neq r_s} \mathbb{P}[A_{r_1,r_2,\ldots,r_s}]\mathbb{E}\left[\left(\sum_{i<j\leq p} Q_{i,j}\right)^p \mid A_{r_1,r_2,\ldots,r_s}\right] = \mathbb{E}\left[\left(\sum_{i<j\leq p} Q_{i,j}\right)^p\right],$$

which means that $\left\|\sum_{i<j\leq p} Q_{i,j}\right\|_p \leq \alpha$, so a uniform bound on the conditional p-norm implies a bound on the p-norm, which finishes the proof.

B.4. **Explanation of Lemma** 4.9**.** For Lemma 4.9, the random variable $\sum_{j>p,j\neq i} Q_{i,j}\sigma_j \mid A_{r_1,\ldots,r_s}$ is defined analogously on the restricted probability space. It follows from definition that this random variable is equivalent to the sum of random variables $\sum_{j>p,j\neq i}(Q_{i,j}\sigma_j \mid A_{r_1,\ldots,r_s}) = \sum_{j>p,j\neq i}(Q_{i,j} \mid A_{r_1,\ldots,r_s})(\sigma_j \mid A_{r_1,\ldots,r_s})$. By Proposition 4.5, we know that the random variables $(Q_{i,j} \mid A_{r_1,\ldots,r_s})$ are independent for $j > p, j \neq i$ and have moments bounded by binomials. Moreover, from the construction of the probability space, we know that $(\sigma_j \mid A_{r_1,\ldots,r_s})$ are independent for $j > p, j \neq i$ and are independent of all $(Q_{i,j} \mid A_{r_1,\ldots,r_s})$, and we also know that $(\sigma_j \mid A_{r_1,\ldots,r_s})$ is distributed as a Rademacher. This implies that we can apply Proposition 4.7 to obtain a bound on the p-norm $\mathbb{E}\left[\left|\sum_{j>p,j\neq i}(Q_{i,j} \mid A_{r_1,\ldots,r_s})(\sigma_j \mid A_{r_1,\ldots,r_s})\right|^p\right]^{1/p}$ which we can view as a bound on the conditional p-norm $\mathbb{E}\left[\left|\sum_{i<j\leq p} Q_{i,j}x_j\sigma_j\right|^p \mid A_{r_1,r_2,\ldots,r_s}\right]^{1/p}$. Let's suppose that this bound is α. Since the events $\{A_{r_1,\ldots,r_s}\}_{r_1\neq r_2\neq\ldots\neq r_s}$ form a disjoint partition of the probability space, we know by the law of total expectation that:

$$\alpha^p \geq \sum_{r_1\neq r_2\neq\ldots\neq r_s} \mathbb{P}[A_{r_1,r_2,\ldots,r_s}]\mathbb{E}\left[\left|\sum_{i<j\leq p} Q_{i,j}\sigma_j x_j\right|^p \mid A_{r_1,r_2,\ldots,r_s}\right] = \mathbb{E}\left[\left|\sum_{i<j\leq p} Q_{i,j}\sigma_j x_j\right|^p\right],$$

which means that $\left\|\sum_{i<j\leq p} Q_{i,j}\sigma_j x_j\right\|_p \leq \alpha$, so a uniform bound on the conditional p-norm implies a bound on the p-norm, which finishes the proof.

Appendix C. Latała's Moment Bounds and Proof Sketches

We sketch proofs of the upper bounds given in the following two lemmas, due to Latała. Full proofs of these lemmas can be found in [14].

Lemma C.1 ([14]) If $q \geq 2$ and $X, X_1, \dots, X_n$ are independent symmetric random variables, then

$$\left\| \sum_{i=1}^{n} X_i \right\|_q \simeq \inf \left\{ T > 0 \text{ such that } \sum_{i=1}^{n} \log \left(\mathbb{E} \left[\left(1 + \frac{X_i}{T} \right)^q \right] \right) \leq q \right\}.$$

Lemma C.2 ([14][11]) If $1 \leq q \leq n$ and $X, X_1, \dots, X_n$ are i.i.d nonnegative random variables, then

$$\left\| \sum_{i=1}^{n} X_i \right\|_q \simeq \sup_{1 \leq t \leq q} \frac{q}{t} \left(\frac{n}{q} \right)^{1/t} \|X\|_t .$$

For a random variable X, we define

$$\phi_q(X) := \mathbb{E}[|1 + X|^q].$$

We begin with the following proposition that relates ϕ_q to the q-norm, which is useful in proving Lemma $C.2$ and Lemma $C.1$.

Proposition C.3 If independent random variables $X_1, \dots, X_n$ and value $q \geq 1$ satisfy the following inequality for any $T > 0$:

$$\left(\left\| \sum_{i=1}^{n} \frac{X_i}{T} \right\|_q \right)^q \leq \prod_{i=1}^{n} \phi_q \left(\frac{X_i}{T} \right),$$

then

$$\left\| \left(\sum_{i=1}^{n} X_i \right) \right\|_q \lesssim \inf \left\{ T > 0 \text{ such that } \sum_{i=1}^{n} \log \left(\mathbb{E} \left[\left(1 + \frac{X_i}{T} \right)^q \right] \right) \leq q \right\}.$$

Proof. Suppose that $\sum_{i=1}^{n} \log \left(\mathbb{E} \left[\left(1 + \frac{X_i}{T} \right)^q \right] \right) \leq q$. Then,

$$\left\| \left(\sum_{i=1}^{n} X_i \right) \right\|_q = T \left\| \left(\sum_{i=1}^{n} \frac{X_i}{T} \right) \right\|_q \leq T \left(\prod_{i=1}^{n} \phi_q \left(\frac{X_i}{T} \right) \right)^{1/q} = T e^{\frac{1}{q} \log \left(\sum_{i=1}^{n} \phi_q \left(\frac{X_i}{T} \right) \right)} \lesssim T.$$

□

The proofs of the upper bounds of Lemma $C.1$ and Lemma $C.2$ boil down to showing that the condition of Proposition $C.3$ is satisfied. In Section C.1, we sketch a proof of the upper bound of Lemma $C.2$. In Section C.2, we sketch a proof of the upper bound of Lemma $C.1$.

C.1. **Proof Sketch of Lemma $C.1$ (Upper Bound).** We first state the following two propositions. The proof of these propositions are straightforward calculations and can be found in [14].

Proposition C.4 If $q \geq 2$ and $Y_1, Y_2, \dots, Y_n$ are independent symmetric random variables, then

$$\phi_q \left(\sum_{i=1}^{n} Y_i \right) \leq \prod_{i=1}^{n} \phi_q(Y_i).$$

Proposition C.5 If $q \geq 2$ and Y is a symmetric random variable, then

$$||Y||_q^q \leq \phi_q(Y).$$

From Proposition $C.4$ and Proposition $C.5$, coupled with the fact that $X_1/T, X_2/T, \dots, X_n/T$, and $\sum_{i=1}^{n} X_i/T$ are independent symmetric random variables, it follows that the condition of Proposition $C.3$ is satisfied.

[11] This result was actually first due to S.J. Montgomery-Smith through a private communication with Latała. Nonetheless, it is also a corollary of a result in [14].

C.2. **Proof Sketch of Lemma** $C.2$ **(Upper Bound).** We first sketch the proof of the upper bound of the following sublemma, which is analogous to Lemma $C.1$:

Lemma C.6 (Latała[14]**)** If $q \geq 1$ and $X, X_1, \ldots, X_n$ are independent nonnegative random variables then

$$\left\| \sum_{i=1}^{n} X_i \right\|_q \simeq \inf \left\{ T > 0 \text{ such that } \sum_{i=1}^{n} \log \left(\mathbb{E} \left[\left| 1 + \frac{X_i}{T} \right|^q \right] \right) \leq q \right\}.$$

The upper bound of Lemma $C.6$ follows from the following propositions, coupled with the fact that X_1/T, X_2/T, $\ldots$, X_n/T, and $\sum_{i=1}^{n} X_i/T$ are independent nonnegative random variables. The proofs of these propositions are straightforward calculations and can be found in [14].

Proposition C.7 If $q \geq 1$ and $Y_1, Y_2, \ldots, Y_n$ are independent nonnegative random variables, then

$$\phi_q\left(\sum_{i=1}^{n} X_i \right) \leq \prod_{i=1}^{n} \phi_q(X_i).$$

Proposition C.8 If $q \geq 1$ and Y is a nonnegative random variable, then

$$\|Y\|_q^q \leq \phi_q(Y).$$

Now, we describe how the upper bound of Lemma $C.6$ implies the upper bound of Lemma $C.2$. It suffices to show if we take

$$T = 2e \sup_{1 \leq t \leq q} \frac{q}{t} \left(\frac{n}{q} \right)^{1/t} \|X\|_t ,$$

then

$$\sum_{i=1}^{n} \log \left(\mathbb{E} \left[\left| 1 + \frac{X_i}{T} \right|^q \right] \right) \leq q. \tag{4}$$

Since $q \leq n$, it follows that

$$\begin{aligned} \mathbb{E} \left[\left| 1 + \frac{X_i}{T} \right|^q \right] &= \frac{\mathbb{E}[X_i^q]}{T^q} + \sum_{k=0}^{q-1} \binom{q}{k} \frac{\mathbb{E}[X_i^k]}{T^k} \\ &\leq \frac{\mathbb{E}[X_i^q]}{\left(\left(\frac{n}{q} \right)^{1/q} \|X_i\|_q \right)^q} + \sum_{k=0}^{q-1} \left(\frac{qe}{k} \right)^k \frac{\mathbb{E}[X_i^k]}{\left(\frac{2eq\|X_i\|_k}{k} \right)^k} \\ &\leq \frac{q}{n} + 1 \\ &\leq e^{\frac{q}{n}} \end{aligned}$$

from which (4) follows.

Appendix D. Proof of Proposition 4.7

The main ingredient in this proof is Lemma $C.1$ (Latała'a bound on moments of sums of symmetric random variables). The proof of this bound is sketched in Appendix C.

Proof of Proposition 4.7. Since the Y_i are independent random variables, we can apply Lemma $C.1$ to obtain:

$$\left\| \sum_{k=1}^{M} Y_k y_k \sigma_k \right\|_q \lesssim \inf \left\{ T > 0 \mid \sum_{k=1}^{M} \log \left(\mathbb{E} \left[\left| 1 + \frac{Y_k \sigma_k y_k}{T} \right|^q \right] \right) \leq q \right\}.$$

Thus, it suffices to show

$$T \simeq \begin{cases} \frac{\sqrt{q}}{\log B} & \text{if } B \geq e \\ \frac{\sqrt{q}}{B} & \text{if } B < e \end{cases}$$

satisfies $\sum_{k=1}^{M}\log\left(\mathbb{E}\left[\left(1+\frac{Y_k\sigma_k y_k}{t}\right)^q\right]\right)\leq q$. We see

$$\begin{aligned}\sum_{k=1}^{M}\log\left(\mathbb{E}\left[\left(1+\frac{Y_k\sigma_k y_k}{T}\right)^q\right]\right)&=\sum_{k=1}^{M}\log\left(1+\sum_{l=1}^{q}\binom{q}{l}\frac{(\mathbb{E}[(Y_k\sigma_k)^l])y_k^l}{T^l}\right)\\&=\sum_{k=1}^{M}\log\left(1+\sum_{l=1}^{q/2}\binom{q}{2l}\frac{\|Y_k\|_{2l}^{2l}y_k^{2l}}{T^{2l}}\right)\\&\leq\sum_{k=1}^{M}\log\left(1+\sum_{l=1}^{q/2}\left(\frac{qe}{2l}\right)^{2l}\left(\frac{\|Y_k\|_{2l}\,y_k}{T}\right)^{2l}\right)\end{aligned}$$

By the bound on moments of binomial random variables in Proposition 4.6, we know if $B\geq e$ that there exists a universal constant C such that $||Q_{i,j}||_{2l}\leq\frac{2lC}{\log B}$. Thus, we obtain

$$\begin{aligned}\sum_{k=1}^{M}\log\left(\mathbb{E}\left[\left(1+\frac{Y_k\sigma_k y_k}{T}\right)^q\right]\right)&\leq\sum_{k=1}^{M}\log\left(1+\sum_{l=1}^{q/2}\left(\frac{qe}{2l}\right)^{2l}\left(\frac{2lCx_j}{T\log B}\right)^{2l}\right)\\&\leq\sum_{k=1}^{M}\log\left(1+\sum_{l=1}^{q/2}\left(\frac{qeCx_j}{T\log B}\right)^{2l}\right).\end{aligned}$$

Since $|y_k|\leq\frac{1}{q}$, if we set $T=\frac{2eC\sqrt{q}}{\log B}$, then we obtain

$$\sum_{k=1}^{M}\log\left(1+(\sqrt{q}y_k)^2\right)=\sum_{k=1}^{M}\log\left(1+qy_k^2\right)\leq\sum_{i=1}^{n}qy_k^2\leq q$$

as desired. An analogous argument shows that if $B<e$, we can set $T=\frac{2eC\sqrt{q}}{B}$.

□

Appendix E. Proof of Proposition 4.6

The main tool that we use in this proof is Lemma $C.2$ (Latała's bound on moments of sums of i.i.d nonnegative random variables). The proof of Lemma $C.2$ is sketched in Appendix C.

Proof of Proposition 4.6. Notice that it suffices to obtain an upper bound on $\|X\|_q$ for all $N\geq q$. (Since $\|X\|_q$ is an increasing function of N, an upper bound on $\|X\|_q$ at $N=q$ is also an upper bound on $\|X\|_q$ for all $N<q$). For the rest of the proof, we assume $N\geq q$.

Notice X has the same distribution as $\sum_{j=1}^{N}Z_j$ where $Z,Z_1,\ldots,Z_N$ are i.i.d Bernoulli random variables with expectation α. Since $\|Z\|_t=\alpha^{1/t}$, we know by Lemma $C.2$,

$$\begin{aligned}\|X\|_q&\simeq\sup_{1\leq t\leq q}\frac{q}{t}\left(\frac{N}{q}\right)^{1/t}\alpha^{1/t}\\&=\sup_{1\leq t\leq q}\frac{q}{t}\left(\frac{1}{B}\right)^{1/t}\end{aligned}$$

At $t=1$, this quantity is equal to $\frac{q}{B}$, and at $t=q$, this quantity is equal to $\left(\frac{1}{B}\right)^{1/q}=e^{\log(1/B)/q}$. The only $t\in\mathbb{R}$ for which this quantity has derivative 0 is $t=\log B$. Notice that $1\leq\log B\leq q$ if and only if $e\leq B\leq e^q$. Thus

$$\|X\|_q\simeq\begin{cases}\max(\frac{q}{B},\frac{q}{\log B},e^{\log(1/B)/q}) & \text{if } e\leq B\leq e^q\\ \max(\frac{q}{B},e^{\log(1/B)/q}) & \text{if } B<e \text{ or if } B>e^q.\end{cases}.$$

For $B\geq e$, we want to show $\|X\|_q\lesssim q/\log B$. Since $\log B>0$, we see $e^{\log(1/B)/q}=e^{-\log B/q}\leq q/\log B$ and $q/B\leq q/\log B$.

For $B < e$, we want to show $\|X\|_q \lesssim q/B$. Since $\frac{1}{B} > \frac{1}{e}$, we see $e^{\log(1/B)/q} = \left(\frac{1}{B}\right)^{1/q} \leq \frac{e}{B} \lesssim \frac{q}{B}$. □

APPENDIX F. WEAKNESS OF BOUND ON $\|Z\|_p$ FROM (5)

Like in Section 4, we view the random variable Z as a quadratic form $\frac{1}{s}\sigma^T A\sigma$, where σ an n-dimensional vector of independent Rademachers and A is a symmetric, zero-diagonal $n \times n$ matrix where the (i, j)th entry (for $i \neq j$) is $x_i x_j \sum_{r=1}^m \eta_{r,i}\eta_{r,j} = Q_{i,j}x_i x_j$. Applying the Hanson-Wright bound followed by an expectation over the η values yields

$$\left\|\sigma^T A\sigma\right\|_p \lesssim \left\|\sqrt{p}\sqrt{\sum_{i=1}^n \sum_{j\leq n, j\neq i} Q_{i,j}^2 x_i^2 x_j^2} + p \sup_{\|y\|_2=1} \left|\sum_{i=1}^n \sum_{j\leq n, j\neq i} Q_{i,j}x_i x_j y_i y_j\right|\right\|_p =: U_p. \tag{5}$$

We show that the vector $x = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \ldots, 0] \in \mathbb{R}^n$ forces U_p to be too large to yield the optimal m value, thus proving that the Hanson-Wright bound does not provide a sufficiently tight bound on $\|Z\|_p$ to achieve Theorem 1.3. The main ingredient in our proof is the following lemma, which we prove in subsection C.1:

Lemma F.1 For every column $1 \leq i \leq n$, suppose that the random variables $\{\eta_{r,i}\}_{r\in[m],i\in[n]}$ have the distribution defined by uniformly choosing exactly s of the variables per column. If $x = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \ldots, 0\right]$, $p < s$ and $B = m/s^2 \leq \frac{e^p}{p}$, then

$$U_p \simeq \begin{cases} \frac{p^2}{\log Bp} & \text{if } B \geq \frac{e}{p} \\ \frac{p}{B} & \text{if } B < \frac{e}{p}. \end{cases}$$

We can obtain bounds on s and m from Lemma $F.1$ via Markov's inequality. We disregard the case where $B \geq \frac{e^p}{p}$, since this case would yield a value for m that is not polynomial in $\log(1/\delta)$. If $B < e/p$, then it follows that $s = \Theta(\varepsilon^{-1}B^{-1}\log(1/\delta)) = \Omega(\varepsilon^{-1}\log^2(1/\delta))$ and $m = \Theta(\varepsilon^{-2}B^{-1}\log^2(1/\delta)) = \Omega(\varepsilon^{-2}\log^3(1/\delta))$. If $B \geq e/p$, then it follows that $s = \Theta(\varepsilon^{-1}p^2/\log(Bp)) = \Omega(\varepsilon^{-1}\log(1/\delta))$ and $m = \Theta(\varepsilon^{-2}p^4 B/\log^2(Bp)) = \Omega(\varepsilon^{-2}\log^3(1/\delta))$. These bounds on m incur an extra $\log(1/\delta)$ factor, and thus the Hanson-Wright bound is too weak for this setting. Now, it suffices to prove Lemma $F.1$, which we do in the next section.

F.1. Proof of Lemma $F.1$.

In this section, we assume that $x = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \ldots, 0\right]$ and that the random variables $\{\eta_{r,i}\}_{r\in[m],i\in[n]}$ have the distribution defined by uniformly choosing exactly s of the variables per column. We first show the following computation of $||Q_{i,j}||_p$.

Proposition F.2 Assume that the random variables $\{\eta_{r,i}\}_{r\in[m],i\in[n]}$ have the distribution defined by uniformly choosing exactly s of the variables per column. Then, if $p < s$ and $X \sim \mathrm{Bin}(s, s/m)$, we have that $||Q_{i,j}||_p \simeq ||X||_p$.

Proof. We condition on the even that the nonzero locations in column i are at $r_1, r_2, \ldots, r_s$. Notice that the random variable $(Q_{i,j} \mid \eta_{r_1,i} = \eta_{r_2,i} = \ldots = \eta_{r_s,i} = 1)$ is distributed as $Z_{r_1} + Z_{r_2} + \ldots + Z_{r_s}$ where Z_{r_k} is an indicator for the kth entry in the jth column being nonzero. Let Z'_{r_k} for $1 \leq k \leq s$ be i.i.d random variables distributed as $\mathrm{Bern}(s/m)$. Now, observe that

$$\mathbb{E}[(Z_{r_1} + Z_{r_2} + \ldots + Z_{r_s})^p] = \sum_{0\leq t_1,t_2,\ldots,t_s\leq p, t_1+t_2+\ldots+t_s=p} \mathbb{E}[\prod_{i=1}^s Z_{r_i}^{t_i}] = \sum_{0\leq t_1,t_2,\ldots,t_s\leq p, t_1+t_2+\ldots+t_s=p} \mathbb{E}[\prod_{i|t_i>0} Z_{r_i}].$$

Notice that $\mathbb{E}[(Z'_{r_1} + Z'_{r_2} + \ldots + Z'_{r_s})^p] = \sum_{0\leq t_1,t_2,\ldots,t_s\leq p, t_1+t_2+\ldots+t_s=p} \mathbb{E}[\prod_{i|t_i>0} Z'_{r_i}]$. Thus, it suffices to compare $\mathbb{E}[\prod_{i|t_i>0} Z_{r_i}]$ and $\mathbb{E}[\prod_{i|t_i>0} Z'_{r_i}]$. We see that $\mathbb{E}[\prod_{i|t_i>0} Z'_{r_i}] = \left(\frac{s}{m}\right)^{|\{i|t_i>0\}|}$. Since $p < s$, we see that $\mathbb{E}[\prod_{i|t_i>0} Z_{r_i}] = \prod_{j=0}^{|\{i|t_i>0\}|-1} \frac{s-j}{m-j}$. It is not difficult to verify that this ratio is bounded by $2^{O(p)}$ as desired, so

$$\frac{\mathbb{E}[(Q_{i,j} \mid \eta_{r_1,i} = \eta_{r_2,i} = \ldots = \eta_{r_s,i} = 1)^p]}{\mathbb{E}[X^p]} = \frac{\mathbb{E}[(Z_{r_1} + Z_{r_2} + \ldots + Z_{r_s})^p]}{\mathbb{E}[X^p]} \geq 2^{-O(p)}.$$

Now, by the law of total expectation, we know that

$$\frac{\mathbb{E}[Q_{i,j}^p]}{\mathbb{E}[X^p]} \geq 2^{-O(p)}$$

as desired. □

We now prove the following relation between U_p and $\|Q_{1,2}\|_p$:

Lemma F.3 Assume the notation and restrictions above. Then $U_p \simeq p\,\|Q_{1,2}\|_p$.

Proof of Lemma F.3. For ease of notation, we define

$$S_1 := p \sup_{\|y\|_2=1} \left| \sum_{i=1}^{n} \sum_{j \leq n, j \neq i} Q_{i,j} x_i x_j y_i y_j \right|$$

$$S_2 := \sqrt{p} \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} Q_{i,j}^2 x_i^2 x_j^2}.$$

Our goal is to calculate $U_p = \|S_1 + S_2\|_p$. We make use of the following upper and lower bounds on $\|S_1 + S_2\|_p$:

$$\left| \|S_1\|_p - \|S_2\|_p \right| \leq \|S_1 - S_2\|_p \leq \|S_1 + S_2\|_p \leq \|S_1\|_p + \|S_2\|_p . \tag{6}$$

In order to compute $\left| \|S_1\|_p - \|S_2\|_p \right|$ and $\|S_1\|_p + \|S_2\|_p$, we first compute $\|S_1\|_p$ and $\|S_2\|_p$. For our choice of x, notice

$$\|S_1\|_p \simeq p \left\| \sup_{\|y\|_2=1} |Q_{1,2} y_1 y_2| \right\|_p \simeq p\,\|Q_{1,2}\|_p$$

$$\|S_2\|_p \simeq \sqrt{p} \left\| \sqrt{Q_{1,2}^2} \right\|_p = \sqrt{p}\,\|Q_{1,2}\|_p .$$

From these bounds, coupled with (6), it follows that $\|U\|_p \simeq p\,\|Q_{1,2}\|_p$ as desired. □

We now show Lemma $F.1$ follows from Lemma $F.3$ and Proposition $F.2$.

Proof of Lemma F.1. After applying Lemma $F.3$, it suffices to calculate $\|Q_{1,2}\|_p$. It follows from Proposition $F.2$ that $\|Q_{1,2}\|_p \simeq \|X\|_p$ where X is distributed as $\mathrm{Bin}(s, s/m)$. Now, the following calculation $\|X\|_p$ for $p < s$ and $B = m/s^2 \leq \frac{e^p}{p}$ follows from the lower and upper bounds of Lemma $C.2$ (Latała's bound on moments of sums of i.i.d nonnegative random variables):

$$\|X\|_p \simeq \begin{cases} \frac{p}{\log Bp} & \text{if } B \geq \frac{e}{p} \\ \frac{1}{B} & \text{if } B < \frac{e}{p} \end{cases} .$$

From this, Lemma $F.1$ follows. □