Elementary proof of the spherical section property for random matrices

Stephen A. Vavasis

April 9, 2009

Abstract

We provide elementary proofs that Bernoulli and Gaussian random matrices satisfy the so-called approximate spherical section property. The best possible of this type was established by Kashin and by Garnaev and Gluskin. In the case of Gaussian matrices, our bound is weaker than theirs (by a factor of $\sqrt{\log n}$) but uses only elementary arguments. This analysis provides elementary proofs of the main results of compressive sensing.

1 Introduction

In recent papers by Candès and Tao; Donoho; and others, a paradigm called *compressive sensing* has been described. The main results of compressive sensing take the following form. Almost every $m \times n$ matrix A, $m \ll n$, chosen from certain simple distributions (e.g., i.i.d. entries drawn from $N(0,1)$) has the following properties. If $\bar{\mathbf{x}} \in \mathbf{R}^n$ is a sufficiently sparse vector, then $\bar{\mathbf{x}}$ is uniquely recovered by solving the convex optimization problem of minimizing $\|\xi\|_1$ subject to $A\xi = \mathbf{b}$, where $\mathbf{b} = A\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ is well approximated by a sparse vector, then the solution ξ to the above problem is also a good approximation to $\bar{\mathbf{x}}$.

Compressive sensing requires that A have some properties; Candès and Tao [3] identify the Uniform Uncertainty Principle (UUP) as the key property and in later work [2] the RIP (restricted isometry property). Donoho proposes CS1, CS2, CS3 as the properties the matrix should have. These authors then proceed to prove that almost all matrices from certain common distributions have the properties they define. However, these proofs are rather complicated because they involve recent results about extremal eigenvalues of random matrices. Baraniuk et al. [1] simplified some of these proofs, but even these simplified proofs require some advanced results about concentration of measure.

It has emerged recently in the work of Kashin and Temlyakov [8] and Zhang [10] that in fact most of the results about compressive sensing can be derived from another property of the random matrix called the approximate spherical section property. The spherical section property has the advantage, when compared to the UUP, RIP or CS1-3, is that it does not involve estimates of extremal eigenvalues and hence potentially admits simpler analysis. A further advantage, noted by Zhang, is that the spherical section property is invariant under left multiplication by a nonsingular matrix (unlike the UUP, RIP or CS1-3). This is a desirable property since many of the compressive sensing theorems depend only on null(A) and not on A itself. A notable exception are theorems that involve measurement error, i.e., theorems that concern the case that one must recover $\bar{\mathbf{x}}$ given $A\bar{\mathbf{x}}+\mathbf{r}$, where $\mathbf{r}$ is a small error in the measurement [2]. In this case, the result is dependent on the matrix so the spherical section property alone is not enough to obtain the result.

2 Spherical section property

The main definition of these notes is as follows.

Definition 1. *A subspace $V \subset \mathbf{R}^n$ of dimension $n-m$ is said to have the* spherical section property *with* distortion Δ *provided that for all $\mathbf{v} \in V - \{\mathbf{0}\}$,*

$$\frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_2} \geq \sqrt{m/\Delta}. \tag{1}$$

Thus, a subspace with the spherical section property has no vectors with unusually small 1-norm.

The main result of this paper is the following theorem. Let $N(0,1)$ denote the normal distribution with variance 1. Let $\mathcal{B}$ denote the Bernoulli distribution in which -1 has probability 1/2 and 1 has probability 1/2.

Theorem 1. *If V is defined as the null space of a random $m \times n$ matrix A, such that the entries of A are i.i.d. drawn from either $N(0,1)$ or $\mathcal{B}$, then, with probability $1-\exp(-C_1 m)$, V has the spherical section property with distortion $\Delta = C_0 \log(n) \cdot (1 + \log(n/m))$, where $C_0, C_1 > 0$ are constants.*

Note: the constants depend on whether Bernoulli or Gaussian is under consideration, but otherwise they do not depend on any other aspect of the problem.

In contrast, the optimal bound for Δ in the case of Gaussian matrices is $\Delta = C_0(1 + \log(n/m))$, as proved by Kashin [7] and Garnaev and Gluskin [4]. The result is stated in this form explicitly and attributed to these authors by Gluskin and Milman [5]. (The probability of success is $1-\exp(-C_1(n-m))$ rather than $1-\exp(-C_1 m)$ according to [5].)

The proof of this theorem occupies most of the rest of the paper. In the remainder of this section, we develop some consequences of the spherical section property. In Section 3, we state two properties of the distribution, which we denote **P1** and **P2**, that are the key to proving Theorem 1. The proofs that these properties hold for the Gaussian and Bernoulli distributions are presented after the properties are stated. Finally, in Sections 5–7, the proof that Theorem 1 can be deduced from **P1** and **P2** is presented.

The following lemma is well known, so we omit its proof.

Lemma 2. *For any $\mathbf{x} \in \mathbf{R}^n$,*

$$\|\mathbf{x}\|_2 \leq \left(\|\mathbf{x}\|_1 \cdot \|\mathbf{x}\|_\infty\right)^{1/2}.$$

The first main lemma characterizes vectors with a small 1-norm.

Lemma 3. *Suppose* $\mathbf{x} \in \mathbf{R}^n$ *satisfies* $\|\mathbf{x}\|_2 = 1$ *and* $\|\mathbf{x}\|_1 \le C$. *(Clearly* $C \ge 1$ *else this is impossible.) Suppose* m *is an integer satisfying* $1 \le m < n$. *Then if* $T \subset \{1, \ldots n\}$ *contains the indices of the* m *entries of* $\mathbf{x}$ *with largest magnitude, then* $\|\mathbf{x}(\bar{T})\|_2 \le C/(2\sqrt{m})$, *where* $\bar{T}$ *denotes* $\{1, \ldots, n\} - T$.

Proof. Assume without loss of generality that all entries of $\mathbf{x}$ are nonnegative. Let γ be the value of the largest entry of $\mathbf{x}(\bar{T})$, and let β denote $\|\mathbf{x}(\bar{T})\|_2$. Then each x_i for $i \in T$ satisfies $x_i \ge \gamma$, hence $\|\mathbf{x}(T)\|_1 \ge m\gamma$. The 1-norm of $\mathbf{x}(\bar{T})$ is at least $\|\mathbf{x}(\bar{T})\|_2^2 / \|\mathbf{x}(\bar{T})\|_\infty$ by the preceding lemma, i.e., at least β^2/γ. Thus, $\|\mathbf{x}\|_1 = \|\mathbf{x}(T)\|_1 + \|\mathbf{x}(\bar{T})\|_1 \ge m\gamma + \beta^2/\gamma$ so $m\gamma + \beta^2/\gamma \le C$. Applying the AGM inequality on the left, this means $2\sqrt{m}\beta \le C$ which proves the lemma. □

We now introduce some numbering notation used for the remainder of the paper. Given positive integers m', n such that $m' \le n$, let the sequence $\tau_1, \tau_2, \ldots, \tau_r$ be defined according to $\tau_1 = m'$, $\tau_2 = 2m'$, $\tau_3 = 4m'$, up to $\tau_{r-1} = 2^{r-2}m'$, with r chosen so that $2^{r-2}m' < n$ but $2^{r-1}m' \ge n$. We assume for the rest of this paper that $r \ge 2$, i.e., $m' < n$. Finally, let τ_r be chosen so that $\tau_1 + \cdots + \tau_r = n$. Note that $r \le 1 + \log(n/m')/\log(2)$.

Lemma 4. *Let*

$$V_{C,n} = \{\mathbf{x} \in \mathbf{R}^n : \|\mathbf{x}\|_2 = 1 \text{ and } \|\mathbf{x}\|_1 \le C\}. \tag{2}$$

Let m' *be a positive integer such that* $m' < n$, *and let* $\{1, \ldots, n\}$ *be partitioned into* $T_1 \cup \cdots \cup T_r$ *such that* $|T_i| = \tau_i$ *where* τ_i *was defined above. Then* $V_{C,n}$ *is contained in the union (over all possible such choices of* $T_1, \ldots, T_r$*) of the sets*

$$\begin{aligned} V_{C,n,T_1,\ldots,T_r} \;=\; & \{\mathbf{x} \in \mathbf{R}^n : \|\mathbf{x}(T_1)\|_2 \le 1, \|\mathbf{x}(T_2)\|_2 \le C/(2\sqrt{m'}), \|\mathbf{x}(T_3)\|_2 \le C/(2\sqrt{2m'}), \ldots \\ & \text{and } \|\mathbf{x}(T_r)\|_2 \le C/(2\sqrt{2^{r-2}m'})\}. \end{aligned}$$

Proof. Suppose $\mathbf{x} \in V_{C,n}$. Let T_1 index the $\tau_1 (= m')$ entries of $\mathbf{x}$ with largest absolute values, let T_2 index the τ_2 entries with the next largest absolute values, etc. Choose an $i \in \{1, \ldots, r-1\}$, and take $T = T_1 \cup \cdots \cup T_i$. Note that $|T| = m'(1 + 2 + 4 + \cdots + 2^{i-1}) = (2^i - 1)m'$. Apply the preceding lemma to conclude that $\|\mathbf{x}(\bar{T})\|_2 \le C/(2\sqrt{2^{i-1}m'})$, and in particular, $\|\mathbf{x}(T_{i+1})\|_2 \le C/(2\sqrt{2^{i-1}m'})$. □

Let $\mathbf{0}_m$ denote the all-zero vector lying in $\mathbf{R}^m$, and let $B(\mathbf{x}, r)$ denote the closed Euclidean ball of radius r centered at $\mathbf{x}$. The following result is from Pisier [9].

Lemma 5. *The ball* $B(\mathbf{0}_n, 1)$ *can be covered with at most* $(1 + 2/\epsilon)^n$ *balls of radius* ϵ.

Proof. Construct a sequence of points $\mathbf{z}_1, \mathbf{z}_2, \ldots$, all lying in $B(\mathbf{0}, 1)$, according to the following rule. Choose $\mathbf{z}_0$ arbitrarily in $B(\mathbf{0}, 1)$. Let $\mathbf{z}_i$, $i \ge 1$, be any point in $B(\mathbf{0}, 1)$ such that $\min_{j=1,\ldots,i-1} \|\mathbf{z}_i - \mathbf{z}_j\|_2 > \epsilon$. Stop when no such $\mathbf{z}_i$ can be found; let N be the value of i on the last successful insertion. Then $B(\mathbf{0}, 1) \subset B(\mathbf{z}_1, \epsilon) \cup \cdots \cup B(\mathbf{z}_N, \epsilon)$; this is because if not, there is a $\mathbf{z} \in B(\mathbf{0}, 1)$ such that $\|\mathbf{z} - \mathbf{z}_i\|_2 > \epsilon$ for $i = 1, \ldots, N$, in which case the insertion procedure would not have halted. Next, observe that $B(\mathbf{z}_1, \epsilon/2)$, $\ldots$, $B(\mathbf{z}_N, \epsilon/2)$ are mutually disjoint because $\|\mathbf{z}_i - \mathbf{z}_j\|_2 > \epsilon$ for all pairs (i, j) of distinct indices. The volume of $B(\mathbf{z}_i, \epsilon/2)$ is $v_n(\epsilon/2)^n$, where v_n is the volume of the unit n-ball. All of these balls lie in $B(\mathbf{0}, 1+\epsilon/2)$. The volume of $B(\mathbf{0}, 1+\epsilon/2)$ is $v_n(1+\epsilon/2)^n$, hence

$$N \leq \frac{(1+\epsilon/2)^n}{(\epsilon/2)^n}$$

proving the lemma. □

The following lemma is an immediate consequence of the two preceding lemmas.

Lemma 6. *Let $V_{C,n}$ be as in Lemma 4. Suppose $\epsilon > 0$. There is a set $\mathbf{x}_1, \ldots, \mathbf{x}_N$ of points in $\mathbf{R}^{m'}$ such that $N \leq (1+2/\epsilon)^{m'}$ and such that*

$$B(\mathbf{0}_{m'}, 1) \subset B(\mathbf{x}_1, \epsilon) \cup \cdots \cup B(\mathbf{x}_N, \epsilon).$$

Furthermore, $V_{C,n}$ is a contained by the union of all sets of the form

$$\begin{aligned} V_{T_1,\ldots,T_n,i} &= \{\mathbf{x} \in \mathbf{R}^n : \|\mathbf{x}(T_1) - \mathbf{x}_i\|_2 \leq \epsilon, \|\mathbf{x}(T_2)\|_2 \leq C/(2\sqrt{m'}), \|\mathbf{x}(T_3)\|_2 \leq C/(2\sqrt{2m'}), \ldots \\ &\quad \text{and } \|\mathbf{x}(T_r)\|_2 \leq C/(2\sqrt{2^{r-2}m'})\}, \end{aligned}$$

where $T_1, \ldots, T_n$ have the same range as in Lemma 4 and i ranges over $\{1, \ldots, N\}$.

Note that without loss of generality, we may presume that $\mathbf{x}_1, \ldots, \mathbf{x}_N$ in the preceding lemma all satisfy $\|\mathbf{x}_i\|_2 \leq 1$.

The final lemma of this section, also from Pisier [9], bounds a matrix-vector product in terms of the ϵ-net.

Lemma 7. *Suppose $B(\mathbf{0}, 1)$ is covered by N balls of radius ϵ centered at the points of W, where $W = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. Let C be an arbitrary matrix, and suppose $\mathbf{x} \in B(\mathbf{0}, 1)$. Then*

$$\|C\mathbf{x}\|_2 \leq \frac{1}{1-\epsilon} \max_{i=1,\ldots,N} \|C\mathbf{x}_i\|_2.$$

Proof. We claim that $\mathbf{x}$ may be expressed as an infinite sum of the form $\mathbf{x} = \mu_0\mathbf{y}_0 + \mu_1\mathbf{y}_1 + \cdots$, where each $\mathbf{y}_j \in W$ and where $|\mu_j| \leq \epsilon^j$. Let $\mathbf{y}_0 \in W$ be the point of W closest to $\mathbf{x}$. Since the balls of radius ϵ centered at points of W cover $B(\mathbf{0}, 1)$, $\|\mathbf{x} - \mathbf{y}_0\|_2 \leq \epsilon$. Take $\mu_0 = 1$. Next, let $\mathbf{x}' = (\mathbf{x} - \mathbf{y}_0)/\|\mathbf{x} - \mathbf{y}_0\|_2$ so that $\mathbf{x}' \in B(\mathbf{0}, 1)$. Find a $\mathbf{y}_1 \in W$ such that $\|\mathbf{x}' - \mathbf{y}_1\|_2 \leq \epsilon$. Then $\|\mathbf{x}' - \mathbf{y}_1\|_2 = \|(\mathbf{x} - \mu_0\mathbf{y}_0)/\|\mathbf{x} - \mu_0\mathbf{y}_0\|_2 - \mathbf{y}_1\|_2 \leq \epsilon$ so $\|\mathbf{x} - \mu_0\mathbf{y}_0 - \mu_1\mathbf{y}_1\|_2 \leq \epsilon^2$, where $\mu_1 = \|\mathbf{x} - \mu_0\mathbf{y}_0\|_2$. Continuing in this manner proves the claim.

Then for all $\mathbf{x} \in B(\mathbf{0}, 1)$,

$$\begin{aligned} \|C\mathbf{x}\|_2 &= \left\| C \sum_{i=0}^{\infty} \mu_i \mathbf{y}_i \right\|_2 \\ &\leq \sum_{i=0}^{\infty} \mu_i \|C\mathbf{y}_i\|_2 \\ &\leq \max_{i=1,\ldots,N} \|C\mathbf{x}_i\|_2 \cdot \sum_{i=0}^{\infty} \mu_i \\ &\leq \frac{\max_{i=1,\ldots,N} \|C\mathbf{x}_i\|_2}{1-\epsilon}. \end{aligned}$$

□

3 Required properties of the distribution

In this section we state two properties of the distribution that will imply the spherical section property. Then we prove in this section that the Gaussian distribution has these properties. In the next section, we show that the Bernoulli distribution also has the properties. Then in Sections 5–7 we show that the two properties imply Theorem 1.

Let $\mathcal{D}$ be a probability distribution over $\mathbf{R}$ symmetric with respect to the origin. In this paper, we will consider either $\mathcal{D} = N(0,1)$ or $\mathcal{D} = \mathcal{B}$.

The two properties that imply the spherical section property are as follows.

Property P1. There is a universal constant $c_0 > 0$ with the following property. Let $a_1, \ldots, a_n$ be i.i.d. according to distribution $\mathcal{D}$. Let $\mathbf{x}$ be an arbitrary deterministic unit vector and t an arbitrary positive number. Then $\text{Prob}(\mathbf{a}^T\mathbf{x} \geq t) \leq \exp(-c_0 t^2)$.

Property P2. There are universal constants $c_1 > 0$ and $\gamma_0 \in (0,1)$ with the following property. Let $a_1, \ldots, a_n$ be i.i.d. according to distribution $\mathcal{D}$. Let $\mathbf{x}$ be an arbitrary deterministic unit vector. Then $\text{Prob}(|\mathbf{a}^T\mathbf{x}| < c_1) \leq \gamma_0$.

Lemma 8. *The Gaussian distribution has* **P1***.*

Proof. We can find an $n \times n$ orthogonal matrix Q such that $Q\mathbf{x} = [1; 0; 0 \cdots; 0]$. If we replace $\mathbf{a}$ by $Q^T\mathbf{a}$ its distribution is unchanged because the Gaussian distribution is invariant under orthogonal transformation. Therefore, without loss of generality, $\mathbf{x} = [1; 0; 0 \cdots; 0]$. In this case, $\mathbf{a}^T\mathbf{x} = a_1$, which is normally distributed. Therefore,

$$\text{Prob}(\mathbf{a}^T\mathbf{x} \geq t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty \exp(-s^2/2)\, ds.$$

An upper bound on the integral on the right is $\exp(-ct^2)$. □

Lemma 9. *The Gaussian distribution has* **P2***.*

Proof. As in the previous lemma, we may assume without loss of generality that $\mathbf{x} = [1; 0; 0; \cdots; 0]$. In this case, it is required to show that there exists a c_1 and γ_0 such that $\text{Prob}(|a_1| < c_1) \leq \gamma_0$. In fact, we can choose any $c_1 > 0$ to make this true since the $\text{Prob}(|a_1| < c_1) < 1$ for any $c_1 > 0$. □

4 The Bernoulli distribution has P1 and P2

In this section we establish **P1** and **P2** for the Bernoulli distribution. The arguments are more complicated because we cannot use orthogonal invariance.

In what follows, we will use S. Bernstein's lemma (see [6]) several times.

Lemma 10. *(S. Bernstein.) Suppose $z_1, \ldots, z_n$ are i.i.d. random variables drawn from a distribution Ω, and let $t_1, \ldots, t_n$ be deterministic scalars. Then for any $h > 0$,*

$$\text{Prob}(z_1 + \cdots + z_n \geq t_1 + \cdots + t_n) \leq \prod_{i=1}^n E\left(\exp(h(z_i - t_i))\right).$$

Proof. Let $\phi(z_1, \ldots, z_n)$ be the random variable

$$\phi(z_1,\ldots,z_n) = \begin{cases} 1 & \text{if } z_1 + \cdots + z_n - t_1 - \cdots - t_n \geq 0, \\ 0 & \text{if } z_1 + \cdots + z_n - t_1 - \cdots - t_n < 0. \end{cases}$$

Then $\text{Prob}(z_1 + \cdots + z_n \geq t_1 + \cdots + t_n) = E(\phi(z_1, \ldots, z_n))$ by definition of ϕ. Also, we observe that $\phi(z_1, \ldots, z_n) \leq \exp(h(z_1 + \cdots + z_n - t_1 - \cdots - t_n))$ for any $h > 0$. Thus,

$$\begin{aligned} \text{Prob}(z_1 + \cdots + z_n \geq t_1 + \cdots + t_n) &= E(\phi(z_1, \ldots, z_n)) \\ &\leq E(\exp(h(z_1 + \cdots + z_n - t_1 - \cdots - t_n))) \\ &= E\left(\prod_{i=1}^n \exp(hz_i - ht_i)\right) \\ &= \prod_{i=1}^n E(\exp(hz_i - ht_i)). \end{aligned}$$

The last line follows from independence of the z_i's. □

We start with the following simple lemma.

Lemma 11. *There exists a universal constant $c > 0$ such that $\cosh(x) \leq \exp(cx^2)$ for all $x \in \mathbf{R}$.*

Proof. If $f(x) = \cosh(x)$ and $g(x) = \exp(cx^2)$, then we observe that f and g are both even, so it suffices to establish the result for positive x. Next, we observe that $f(0) = g(0) = 1$ and $f'(0) = g'(0) = 0$, so it suffices to compare their second derivatives: $f''(x) = \cosh(x) = O(e^x)$ for large x, while $g''(x) = (4c^2x^2 + 2c)\exp(cx^2) > \Omega(e^{x^2})$. Thus, both second derivatives are positive, and g'' asymptotically dominates f''. Therefore, choosing c sufficiently large ensures that g'' dominates f'' not just asymptotically but for all x. Numerically we have found that $c = 0.5$ seems to suffice. □

Lemma 12. *The Bernoulli distribution has Property* **P1***.*

Proof. Suppose $a_1, \ldots, a_n$ i.i.d. random variables such that $a_i \sim \mathcal{B}$. Let $\mathbf{x}$ be a (deterministic) unit n-vector.

We write $t = tx_1^2 + \cdots + tx_n^2$ (since $\mathbf{x}$ is a unit vector). Then by Bernstein's lemma,

$$\text{Prob}(\mathbf{a}^T\mathbf{x} \geq t) \leq \prod_{i=1}^n E(\exp(ha_ix_i - htx_i^2)),$$

for an $h > 0$ to be determined. Now let us consider one factor in the product above.

$$\begin{aligned} E(\exp(ha_ix_i - htx_i^2)) &= \frac{1}{2}(\exp(-hx_i - htx_i^2) + \exp(hx_i - htx_i^2)) \\ &= \frac{\exp(-htx_i^2)}{2}(\exp(-hx_i) + \exp(hx_i)) \\ &\leq \exp(-htx_i^2)\exp(ch^2x_i^2) \\ &= \exp((ch^2 - ht)x_i^2). \end{aligned}$$

The third line follows from the preceding lemma for a $c > 0$. We now choose h to minimize the above quantity; the optimal choice is easily seen to be $h = t/(2c)$, so $E(\exp(ha_i x_i - htx_i^2)) \leq \exp(-t^2 x_i^2/(4c))$.

Combining,

$$\begin{aligned}
\text{Prob}(\mathbf{a}^T\mathbf{x} \geq t) &\leq \prod_{i=1}^{n} \exp(-t^2 x_i^2/(4c)) \\
&= \exp(-t^2(x_1^2 + \cdots + x_n^2)/(4c)) \\
&= \exp(-t^2/(4c)).
\end{aligned}$$

□

Lemma 13. *The Bernoulli distribution has Property* **P2**.

Proof. Observe that

$$\begin{aligned}
E((\mathbf{a}^T\mathbf{x})^2) &= E\left(a_1^2 x_1^2 + \cdots + a_n^2 x_n^2 + 2\sum_{i \neq j} a_i a_j x_i x_j\right) \\
&= \sum_{i=1}^{n} E(a_i^2 x_i^2) + 2\sum_{i \neq j} E(a_i a_j x_i x_j) \\
&= \sum_{i=1}^{n} x_i^2 + 0 \\
&= 1.
\end{aligned}$$

The second line follows by linearity of the expected value. The first term of the third line follows because $a_i^2 \equiv 1$ for $a_i \sim \mathcal{B}$. The second term follows because a_i, a_j are independent so $E(a_i a_j x_i x_j) = x_i x_j E(a_i) E(a_j) = 0$ since the mean of a_i is 0. The fourth line follows because $\mathbf{x}$ is a unit vector.

On the other hand, we can write $E((\mathbf{a}^T\mathbf{x})^2)$ as a sum of three terms depending on the value of $|\mathbf{a}^T\mathbf{x}|$. Let c_0, L be scalars to be determined such that $0 < c_0 < L$ and L is integral. Then we have:

$$E((\mathbf{a}_i^T\mathbf{x})^2) \leq T_1 + T_2 + T_3$$

where

$$\begin{aligned}
T_1 &= \text{Prob}(|\mathbf{a}^T\mathbf{x}| < c_0)c_0^2, \\
T_2 &= \text{Prob}(|\mathbf{a}^T\mathbf{x}| \in [c_0, L))L^2, \qquad \text{and} \\
T_3 &= \sum_{l=L}^{\infty} \text{Prob}(|\mathbf{a}^T\mathbf{x}| \in [l, l+1))(l+1)^2.
\end{aligned}$$

We can overestimate the probability in the equation for T_3 as $\text{Prob}(|\mathbf{a}^T\mathbf{x}| \geq l)$ and apply Lemma 12:

$$T_3 \leq \sum_{l=L}^{\infty} e^{-l^2/(4c)}(l+1)^2.$$

This is a very rapidly converging series since the squared exponentially decays quickly, so in particular there exists a universal value of L to ensure that $T_3 \leq 1/2$. Now let $\text{Prob}(|\mathbf{a}^T\mathbf{x}| <$

c_0) be denoted as γ_0; the purpose of the lemma is to estimate γ_0. Then $T_1 = \gamma_0 c_0^2$ and $T_2 \le (1-\gamma_0)L^2$ hence

$$E((\mathbf{a}^T\mathbf{x})^2) \le \gamma_0 c_0^2 + (1-\gamma_0)L^2 + 1/2.$$

We have already shown that the left-hand side is equal to 1. Substituting this and rearranging,

$$\gamma_0 \le \frac{L^2 - 1/2}{L^2 - c_0^2}.$$

Thus, choosing c_0 sufficiently small (in particular, less than $\min(L, 1/2)$) ensures that γ_0 has a universal value strictly less than 1, proving the lemma. □

5 Maximum singular value

In this section we extend property **P1** to estimates of the maximum singular value of a random matrix and its submatrices.

Lemma 14. *Suppose A is an $m \times n$ random matrix with i.i.d. entries such that $A_{ij} \sim \mathcal{D}$, where $\mathcal{D}$ has Property* **P1**. *Let $\mathbf{x}$ be a fixed (deterministic) unit m-vector. Then for any $q \ge 0$,*

$$\mathrm{Prob}(\|A^T\mathbf{x}\|_2 \ge \sqrt{qn}) \le \exp((-c_3 q + c_2)n),$$

where c_2, c_3 are constants depending only on $\mathcal{D}$ and $c_3 > 0$.

Proof. Let $z_i = \mathbf{a}_i^T\mathbf{x}$. Then the z_i are independent random variables symmetrically distributed around 0, and by Property **P1**, $\mathrm{Prob}(z_i \ge t) \le \exp(-c_0 t^2))$ for all $t \ge 0$. Let us write $qn = q + \cdots + q$ (n times) to apply Bernstein's lemma:

$$\mathrm{Prob}(\|A^T\mathbf{x}\|_2 \ge \sqrt{qn}) \le \prod_{i=1}^{n} E(\exp(hz_i^2 - hq)).$$

Let us now analyze one term in above expectation. Let $f(z)$ be the PDF of z_i. In the analysis that follows, we interpret the integral as a discrete summation in the case of $\mathcal{B}$ (a discrete distribution for z_i).

$$\begin{aligned}
E(\exp(hz_i^2 - hq)) &= \int_{-\infty}^{\infty} \exp(hz^2 - hq) f(z)\,dz \\
&= 2\int_{0}^{\infty} \exp(hz^2 - hq) f(z)\,dz \\
&= 2\sum_{l=0}^{\infty} \int_{l}^{l+1} \exp(hz^2 - hq) f(z)\,dz \\
&\le 2\sum_{l=0}^{\infty} \int_{l}^{l+1} \exp(h(l+1)^2 - hq) f(z)\,dz \\
&= 2\sum_{l=0}^{\infty} \mathrm{Prob}(z \in [l, l+1)) \exp(h(l+1)^2 - hq) \\
&\le 2\sum_{l=0}^{\infty} \mathrm{Prob}(z \ge l) \exp(h(l+1)^2 - hq)
\end{aligned}$$

$$\leq \; 2\sum_{l=0}^{\infty}\exp(-c_0 l^2)\exp(h(l+1)^2 - hq).$$

Let us now select $h = c_0/2$ and continue the above derivation.

$$\begin{aligned} E(\exp(hz_i^2 - hq)) &\leq 2\sum_{l=0}^{\infty}\exp(c_0(l+1)^2/2 - c_0 l^2 - c_0 q/2) \\ &\leq 2\exp(-c_3 q)\sum_{l=0}^{\infty}\exp(c_0(l+1)^2//2 - c_0 l^2) \end{aligned}$$

Here, we have defined $c_3 = c_0/2$. The summation mutiplied by 2 is bounded by a constant, say $\exp(c_2)$, since for large l the argument of the exponential behaves like $-c_0 l^2/2$, and so the terms of the summation decrease superexponentially fast. Thus, $E(\exp(hz_i^2 - hq)) \leq \exp(-c_3 q + c_2)$. Thus, combining with the previous paragraph, $\mathrm{Prob}(\|A^T\mathbf{x}\|_2 \geq \sqrt{qn}) \leq \exp((-c_3 q + c_2)n)$. □

We can use this immediately to obtain an upper bound on the maximum singular value of a random Gaussian matrix. Recall the definition that $\sigma_{\max}(A) = \sup\{\|A\mathbf{x}\|_2 : \|\mathbf{x}\|_2 \leq 1\}$. Also, recall the fact that $\sigma_{\max}(A) = \sigma_{\max}(A^T)$.

Lemma 15. *Let A be an $m \times n$ random matrix with i.i.d. entries satisfying* **P1***. Then for any $q > 0$,*

$$\mathrm{Prob}(\sigma_{\max}(A) \geq 10\sqrt{qn}/9) \leq \exp(m\log(21) + (-c_3 q + c_2)n).$$

Proof. By the previous lemma,

$$\mathrm{Prob}(\|A^T\mathbf{x}\|_2 \geq \sqrt{qn}) \leq \exp((-c_3 q + c_2)n).$$

Let us now cover $B(\mathbf{0}_m, 1)$ with balls of radius ϵ for an $\epsilon > 0$ to be determined. Let there be N balls; by Lemma 5, $N \leq (2/\epsilon + 1)^m$. Let the centers of the balls be $\mathbf{x}_1, \ldots, \mathbf{x}_N$. By the union bound,

$$\mathrm{Prob}(\|A^T\mathbf{x}_i\|_2 \geq (qn)^{1/2} \text{ for some } i = 1, \ldots, N) \leq N\exp((-c_3 q + c_2)n). \tag{3}$$

Then applying Lemma 7,

$$\mathrm{Prob}\left(\|A^T\mathbf{x}\|_2 \geq \frac{(qn)^{1/2}}{1-\epsilon} \text{ for some } \mathbf{x} \in B(\mathbf{0}, 1)\right) \leq (2/\epsilon + 1)^m \exp((-c_3 q + c_2)n).$$

Let us now fix ϵ at a constant value, e.g., $\epsilon = 0.1$. Then

$$\mathrm{Prob}(\|A^T\mathbf{x}\|_2 \geq (10/9)(qn)^{1/2}) \leq \exp(m\log(21) + (-c_3 q + c_2)n).$$

□

Lemma 16. *Let $A \in \mathbf{R}^{m\times n}$ have i.i.d. entries following a distribution with Property* **P1**. *Let $\mathcal{T}_{m',n}$ denote the collection of subsets $T \subset \{1,\ldots,n\}$ of cardinality at least m'. Then*

$$\text{Prob}\Big(\sigma_{\max}(A(:,T)) \geq (20/9)\sqrt{q|T|} \tag{4}$$

$$\textit{for some } T \in \mathcal{T}_{m',n}\Big) \leq \binom{n}{m'} \cdot \exp(m\log(21) + (-c_3 q + c_2)m'). \tag{5}$$

Proof. First, consider the case that $|T| = m'$. There are $\binom{n}{m'}$ ways to choose this T; by the union bound and Lemma 15, (5) is valid for this special case, and the constant is 10/9 rather than 20/9.

Now assume that there is a uniform bound of $(10/9)\sqrt{qm'}$ on $\sigma_{\max}(A(:,T))$ for all T of cardinality m'; as just argued, this holds with the claimed probability. Note that the same bound $(10/9)\sqrt{qm'}$ also holds for the case when $|T| < m'$ as well, since any such T is contained in a larger one of cardinality m', and $\sigma_{\max}$ is monotonic with respect to submatrix containment.

Consider a T such that $m' \leq |T| \leq n$. Write this T as a disjoint union $T_1 \cup T_2 \cup \cdots \cup T_s$, where $s = \lceil |T|/m' \rceil$ and each T_i except possibly T_s has cardinality exactly m'; $|T_s| \leq m'$. It is a direct consequence of the operator norm definition of $\sigma_{\max}(\cdot)$ that

$$\sigma_{\max}([A_1, A_2, \ldots, A_k]) \leq \sqrt{k}\max_i(\sigma_{\max}(A_i)).$$

Therefore, $\sigma_{\max}(A(:,T)) \leq \sqrt{s}(10/9)\sqrt{qm'}$. By choice of s, $s \leq 2|T|/m'$ so $\sigma_{\max}(A(:,T)) \leq (20/9)\sqrt{q|T|}$, proving the lemma. □

6 Probability of a low norm

In this section we derive an upper bound on the probability that $A\mathbf{x}$ will have a small norm using Property **P2**. The result is as follows.

Lemma 17. *There exist constants $c_6 > 0$ and $\gamma_1 \in (0,1)$ such that if $A \in \mathbf{R}^{m\times n}$ has random i.i.d. entries in $\mathcal{D}$, where $\mathcal{D}$ satisfies* **P1** *and* **P2**, *and $\mathbf{x}$ is any deterministic unit vector, then*

$$\text{Prob}(\|A\mathbf{x}\|_2 \leq \sqrt{c_6 m}) \leq \gamma_1^m.$$

Proof. Let $z_i = -(\mathbf{a}_i^T\mathbf{x})^2$, where $\mathbf{a}_i^T$ denotes the ith row of A. For now let us leave c_6 and γ_1 undetermined. Note that the z_i's are independent, so we apply Bernstein's analysis again:

$$\begin{aligned}\text{Prob}(\|A\mathbf{x}\|_2 \leq \sqrt{c_6 m}) &= \text{Prob}(-z_1 - \cdots - z_m \leq c_6 m)\\ &= \text{Prob}(z_1 + \cdots + z_m \geq -c_6 m)\\ &\leq \prod_{i=1}^m E(\exp(hz_i + hc_6)). \qquad (6)\end{aligned}$$

Let us analyze one term in this summation. We bound the expectation as a sum of three terms:

$$E(\exp(-h(\mathbf{a}_i^T\mathbf{x})^2 + hc_6)) \le T_1 + T_2 + T_3$$

where

$$\begin{aligned} T_1 &= \text{Prob}(|\mathbf{a}_i^T\mathbf{x}| < c_1) \cdot \exp(hc_6), \\ T_2 &= \text{Prob}(|\mathbf{a}_i^T\mathbf{x}| \in [c_1, M)) \cdot \exp(-hc_1^2 + hc_6), \qquad \text{and} \\ T_3 &= \text{Prob}(|\mathbf{a}_i^T\mathbf{x}| \in [M, \infty)) \cdot \exp(-hM^2 + hc_6). \end{aligned}$$

Here, c_1 is the constant from Property **P2**, and M is another constant to be determined. Let us first recall that Property **P2** states that $\text{Prob}(|\mathbf{a}_i^T\mathbf{x}| < c_1) \le \gamma_0$, where $\gamma_0 \in (0,1)$ is also constant. We now have to select constants h, c_6, γ_1 and M to make the bound work. Let us first select a sufficiently large $h > 0$ so that $\exp(-hc_1^2) \le (1 - 2\gamma_0 + \gamma_0^2)/6$; this is possible since the right-hand side is positive, and $-hc_1^2$ is negative. Next, let us select $c_6 > 0$ sufficiently small so that $\exp(hc_6) \le 2 - \gamma_0$. Again, this is possible because the right-hand side is greater than 1. Clearly, we have more strongly that $\exp(hc_6) < 2$. With these bounds, we can now estimate:

$$T_1 \le \gamma_0 \cdot (2 - \gamma_0) = 2\gamma_0 - \gamma_0^2.$$

Note that this quantity is strictly less than 1 since $1 - 2\gamma_0 + \gamma_0^2$ is positive.

We can now estimate T_2:

$$\begin{aligned} T_2 &= \text{Prob}(|\mathbf{a}_i^T\mathbf{x}| \in [c_1, M)) \cdot \exp(-hc_1^2 + hc_6) \\ &\le \exp(-hc_1^2 + hc_6) \\ &= \exp(hc_6)\exp(-hc_1^2) \\ &\le 2 \cdot (1 - 2\gamma_0 + \gamma_0^2)/6. \end{aligned}$$

Finally, observe that the probability appearing in T_3 tends to 0 superexponentially fast for large M by Property **P1**, so let us select M large enough so that this probability is at most $(1 - 2\gamma_0 + \gamma_0^2)/6$. Then

$$\begin{aligned} T_3 &= \text{Prob}(|\mathbf{a}_i^T\mathbf{x}| \in [M, \infty)) \cdot \exp(-hM^2 + hc_6) \\ &\le \text{Prob}(|\mathbf{a}_i^T\mathbf{x}| \in [M, \infty)) \cdot \exp(hc_6) \\ &\le (1 - 2\gamma_0 + \gamma_0^2)/6 \cdot 2. \end{aligned}$$

Assembling the three terms,

$$\begin{aligned} E(\exp(-h(\mathbf{a}_i^T\mathbf{x})^2 + hc_6)) &\le T_1 + T_2 + T_3 \\ &\le 2\gamma_0 - \gamma_0^2 + (1 - 2\gamma_0 + \gamma_0^2)/3 + (1 - 2\gamma_0 + \gamma_0^2)/3 \\ &\le 2/3 + 2\gamma_0/3 - \gamma_0^2/3. \end{aligned}$$

If we define γ_1 to be the quantity on the final line, then we see that $\gamma_1 < 1$ since $1 - \gamma_1 = 1/3 - 2\gamma_0/3 + \gamma_0^2/3 = (1 - \gamma_0)^2/3 > 0$.

Finally, substituting this bound into (6) gives the result. □

7 Proof of the main theorem

Now finally we prove the main theorem. The proof uses only **P1** and **P2** as well as their consequences, hence it applies to both the Gaussian and Bernoulli cases. Let A be an $m \times n$ matrix with i.i.d. entries from a distribution satisfying **P1** and **P2**. Let m' be a parameter depending on m, n to be determined below. Let us first exclude the possibility that $\sigma_{\max}(A(:,T))$ is large for some $|T|$ of cardinality m' or larger. By Lemma 16, $\sigma_{\max}(A(:,T))$ can be assumed to have a uniform bound of $(20/9)\sqrt{q|T|}$, where q will be determined later (depending on m, n). According to the lemma, this fails with probability at most

$$\exp(m\log(21) - c_3 qm' + c_2 m')\begin{pmatrix} n \\ m' \end{pmatrix},$$

which we overestimate as

$$\exp(m\log(21) - c_3 qm' + c_2 m' + m'\log n), \tag{7}$$

which we will check below is exponentially small after the parameters q, m' are selected.

Thus, for the remainder of the proof, we assume $\sigma_{\max}(A(:,T)) \le (20/9)(q|T|)^{1/2}$. Let C be chosen to depend on m and n; the precise value of C is specified below. Let $V_{C,n}$ be defined as in (2), that is, the set of $\mathbf{x} \in \mathbf{R}^n$ such that $\|\mathbf{x}\|_2 = 1$ and $\|\mathbf{x}\|_1 \le C$. We wish to show that with high probability over the choice of A that satisfy the condition in the previous paragraph, there is no element of $V_{C,n}$ in the null space of A.

Let $\epsilon > 0$ be another parameter depending on m, n. Recalling Lemma 6, let us find N points $\mathbf{x}_1, \ldots, \mathbf{x}_N$ whose ϵ-balls cover $B(\mathbf{0}_{m'}, 1)$. Select one such point $\mathbf{x}_i$, and also select a set T_1 such that $|T_1| = m'$. Let U_{i,T_1} denote the subset of $\mathbf{R}^n$ defined by the lemma once i and T_1 are selected, i.e., U_{i,T_1} is the union over all possible choices of $T_2, \ldots, T_r$ of the set $V_{T_1,\ldots,T_r,i}$ defined by the lemma. Let us consider the probability that A has a unit-length null vector in U_{i,T_1}. Let $\mathbf{x}$ be such a point. Then $\|\mathbf{x}(T_1) - \mathbf{x}_i\|_2 \le \epsilon$, and there are $T_2, \ldots, T_r$ such that $\|\mathbf{x}(T_i)\|_2 \le C/(2\sqrt{2^{i-2}m'})$ for $i = 2, \ldots, r$ as in the lemma.

Since $A\mathbf{x} = \mathbf{0}$, this means that $A(:,T_1)\mathbf{x}(T_1) + \cdots + A(:,T_r)\mathbf{x}(T_r) = \mathbf{0}$. If $\mathbf{v} = \mathbf{x}(T_1) - \mathbf{x}_i$, then $\|\mathbf{v}\|_2 \le \epsilon$ by assumption and $A(:,T_1)\mathbf{x}_i = -A(:,T_1)\mathbf{v} - A(:,T_2)\mathbf{x}(T_2) - \cdots - A(:,T_r)\mathbf{x}(T_r)$. Therefore,

$$\begin{aligned}
\|A(:,T_1)\mathbf{x}_i\|_2 &= \|-A(:,T_1)\mathbf{v} - A(:,T_2)\mathbf{x}(T_2) - \cdots - A(:,T_r)\mathbf{x}(T_r)\|_2 \\
&\le \|A(:,T_1)\mathbf{v}\|_2 + \|A(:,T_2)\mathbf{x}(T_2)\|_2 + \cdots + \|A(:,T_r)\mathbf{x}(T_r)\|_2 \\
&\le \sigma_{\max}(A(:,T_1))\|\mathbf{v}\|_2 + \sigma_{\max}(A(:,T_2))\|\mathbf{x}(T_2)\|_2 + \cdots + \sigma_{\max}(A(:,T_r))\|\mathbf{x}(T_r)\|_2 \\
&\le (20/9)\left((q\tau_1)^{1/2}\epsilon + (q\tau_2)^{1/2}C/(2\sqrt{m'}) + \cdots + (q\tau_r')^{1/2}C/(2\sqrt{2^{r-2}m'})\right) \\
&\le (20/9)q^{1/2}((m')^{1/2}\epsilon + (r-1)\sqrt{2}C/2).
\end{aligned} \tag{8}$$

In the fourth line, we used the fact that $\sigma_{\max}(A(:,T_j)) \le (20/9)(q|T_j|)^{1/2} = (20/9)(q\tau_j)^{1/2}$ for $j = 2, \ldots, r$. In that same line, we use the notation τ_r' to denote $2^{r-1}m'$, which is an upper bound on τ_r. Note that all the terms $(q\tau_j)^{1/2}C/(2\sqrt{2^{j-2}m'})$ in the fourth line the fourth line are equal to $q^{1/2}\sqrt{2}C/2$.

Let us now fix

$$C = \frac{2\epsilon\sqrt{m'}}{\sqrt{2}(r-1)} \tag{9}$$

to make the two terms of (8) equal so that

$$\|A(:,T_1)\mathbf{x}_i\|_2 \le (40/9)(qm')^{1/2}\epsilon. \tag{10}$$

Thus, $A(:,T_1)\mathbf{x}_i$ is a small vector. We want to claim based on earlier lemmas that this is unlikely. In order to make this claim, we first need a lower bound on $\|\mathbf{x}_i\|_2$. This is established as follows:

$$\begin{aligned}
\|\mathbf{x}_i\|_2 &\ge \|\mathbf{x}(T_1)\|_2 - \epsilon \\
&\ge \|\mathbf{x}\|_2 - \sum_{j=2}^{r} \|\mathbf{x}(T_j)\|_2 - \epsilon \\
&= 1 - \sum_{j=2}^{r} \|\mathbf{x}(T_j)\|_2 - \epsilon \\
&\ge 1 - \epsilon - \sum_{j=2}^{r} C/(2\sqrt{2^{j-2}m'}) \\
&\ge 1 - \epsilon - \frac{C/(2\sqrt{m'})}{1-1/\sqrt{2}} \\
&= 1 - \epsilon - \frac{\sqrt{2}\epsilon}{(r-1)(1-1/\sqrt{2})} \\
&\ge 1 - \epsilon - \frac{\sqrt{2}\epsilon}{(r-1)(1-1/\sqrt{2})} \\
&\ge 1 - 8\epsilon.
\end{aligned}$$

Here, we used (9) and the assumption $r \ge 2$. Thus, to ensure a lower bound of 1/2 on $\|\mathbf{x}_1\|$, we need to assert that $\epsilon < 1/16$. After we have selected all the parameters, we can check this bound and hence assume $\|\mathbf{x}_1\|_2 \ge 1/2$.

Thus, there is a unit vector $\hat{\mathbf{x}}_i = \mathbf{x}_i/\|\mathbf{x}_i\|_2$ such that

$$\|A(:,T_1)\hat{\mathbf{x}}_i\|_2 \le (80/9)\sqrt{qm'}\epsilon. \tag{11}$$

By Lemma 17, there is a c_6 and γ_1 such that

$$\mathrm{Prob}(\|A(:,T_1)\hat{\mathbf{x}}_i\|_2 \le \sqrt{c_6 m}) \le \gamma_1^m. \tag{12}$$

These constants depend only on whether we are analyzing the Gaussian or Bernoulli distributions but are otherwise universal. So we now select parameters $m' = c_9 m/\log n$, $q = c_{10}\log n$, and $\epsilon = \min(1/16, (9/80)\sqrt{c_6/(c_9 c_{10})})$, where $c_9, c_{10} > 0$ are universal constants still to be determined. With these choices, the right-hand side of (11) is bounded above by $\sqrt{c_6 m}$, and hence (12) may be applied.

Thus, γ_1^m is an upper bound on the probability that $A(:,T_1)\mathbf{x}_1$ is small, which we showed to be a necessary condition for A to have a unit-length null vector in $U_{T_1,i}$. Thus, it is also an upper bound on the probability of the existence of such a null vector. Now use the union bound to obtain an upper bound on the probability that A has a unit-length null vector anywhere in the cover of $V_{C,n}$, which consists of $\binom{n}{m'}(1+2/\epsilon)^{m'}$ sets isometric to $U_{T_1,i}$:

$$\begin{aligned} \text{Prob}(A\mathbf{x} = \mathbf{0} \text{ for some } \mathbf{x} \in V_{C,n}) &\leq \binom{n}{m'}(1+2/\epsilon)^{m'}\gamma_1^m \\ &\leq n^{m'} 2.1^{m'} \epsilon^{-m'} \gamma_1^m &(13) \\ &= \exp(m'(\log n + \log 2.1 - \log \epsilon) + m \log \gamma_1), &(14) \end{aligned}$$

where we have imposed the assumption that $\epsilon \leq 1/16$ so $1 + 2/\epsilon \leq 2.1/\epsilon$.

Next, substituting our choices for m', ϵ into (14) yields a probability of a null vector in $V_{C,n}$ at most

$$\exp(c_9 m + (\log 2.1 + \log(1/\epsilon))m/\log n + m \log \gamma_1).$$

Regardless of the choices of c_9, c_{10} (which affect ϵ), the middle term is asympotically dominated by the first and third for large m, so we can focus only on the those terms. We see that this bound has the form $\exp(-\text{const} \cdot m)$ provided that $c_9 \in (0, -\log \gamma_1)$ (recall $\gamma_1 < 1$ so its logarithm is negative). Thus, we have exponential decay provided c_9 is sufficiently small.

Now we finally determine the universal constant c_{10} to satisfy all the bounds. We substitute the choices for q and m' into (7), which is the probability that our assumption about the largest singular value fails. After substituting, this probability is written

$$\exp(m \log(21) - c_3 c_9 c_{10} m + c_2 c_9 m/(\log n) + c_9 m) \quad (15)$$

The third term may be ignored since it is asymptotically dominated. This bound is $\exp(-\text{const} \cdot m)$ provided that c_{10} is sufficiently large (in terms of $c_3, c_9, \log 21$) so that the second term dominates the others.

Finally, since we selected $C = 2\sqrt{m'}\epsilon/(\sqrt{2}(r-1))$ in (9) and recall that $r = 1 + \text{const} \cdot \log(n/m)$, $m' = c_9 m/\log n$, $\epsilon = \text{const}$, we conclude that

$$C = \frac{\text{const}\sqrt{m}}{\sqrt{(\log n)(1 + \log(n/m))}}.$$

Thus, for a random Bernoulli or Gaussian matrix, with probability $1 - \exp(-\text{const} \cdot m)$, A has no null vector in $V_{C,n}$ where $V_{C,n}$ was defined by (2). This is the same as saying that with this probability, A has the property for all nonzero $\mathbf{x}$ in the null space of A,

$$\|\mathbf{x}\|_1/\|\mathbf{x}\|_2 \geq \frac{\text{const}\sqrt{m}}{\sqrt{(\log n)(1 + \log(n/m))}}$$

which proves Theorem 1.

References

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28:253–263, 2008.

[2] E. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Information Theory*, 51(12):4203–4215, 2005.

[3] E. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Information Theory*, 52(12):5406–5425, 2006.

[4] A. Y. Garnaev and E. D. Gluskin. On the widths of the Euclidean ball. *Soviet Math. Dokl.*, 30:200–203, 1984.

[5] E. Gluskin and V. Milman. Note on the geometric-arithmetic mean inequality. In *Geometric aspects of Functional analysis*, pages 131–135. Springer, Berlin, 2003.

[6] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. American Statistical Assoc.*, 58:13–30, 1962.

[7] B. Kashin. Diameters of certain finite-dimensional sets in classes of smooth function. *Izv. Akad. Nauk SSSR, Ser. Mat.*, 41(2):334–351, 1977.

[8] B. Kashin and V. Temlyakov. A remark on compressed sensing. *Mathematical notes*, 82:748–755, 2007.

[9] Gilles Pisier. *The volume of convex bodies and Banach space geometry.* Cambridge University Press, Cambridge, 1989.

[10] Y. Zhang. On theory of compressive sensing via ℓ_1-minimization: Simple derivations and extensions. Technical Report TR08-11, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 2008.