Some Math for Compressed Sensing

Jingxing Wang

December 29, 2024

Abstract. *In their work published in 2007 [4], Vershynin and Rudelson proposed a novel approach to bound the number of samples needed for sparse recovery with Fourier measurements. We give an introduction to the field of sparse recovery, then present the following series of key tricks that allow Vershynin and Rudelson to reduce the problem to one that involves Gaussian processes: we first provide an alternative definition of the restricted isometry constant, then derive a bound for its expectation, then rephrase the key lemma in the previous derivation as a chaining problem, and finally state relevant theorems in chaining without proofs to conclude the lemma.*

1 Introduction

1.1 Consider a Setup Where...

A signal $f \in \mathbb{C}^n$ is transformed through the **Discrete Fourier Transform**. This can be thought of as a matrix multiplication where the transform matrix is defined by

$$\Psi_{\omega,t} := \frac{1}{\sqrt{n}} \exp(-i2\pi\omega t/n),\ \omega, t \in \{0, ..., n-1\}. \tag{1}$$

The transform matrix $\Psi \in \mathbb{C}^{n\times n}$'s $(\omega+1)^{th}$, $(t+1)^{th}$ entry is given by the above expression. The transformed signal is given by $\hat{f} = \Psi f$. Each time we make a measurement, a row of Ψ is chosen at random and the $(\omega+1)^{th}$ entry of $\hat{f}$ is revealed to us. We wish to reconstruct the unknown f via sampling from $\hat{f}$ with the following procedure:

- Step 1: k points, denoted as set Ω, are chosen uniformly at random in the set $\{0, ..., n-1\}$.
- Step 2: measurements are observed. That is, for $\omega \in \Omega$, observe $\hat{f}(\omega) = \Psi_\omega f \in \mathbb{C}^n$.
- Step 3: solve the following convex linear program

$$\begin{aligned} &\text{minimize} \quad \|g\|_1 \\ &\text{subject to} \quad \Psi_\omega g = \Psi_\omega f, \quad \omega \in \Omega \end{aligned}$$

- Step 4: claim that the optimal solution to the above problem, $g^* \in \mathbb{C}^n$ to be f.

Of course, at the current level of generality our procedure would fail, so we must also add several conditions.

- Condition 1: f is r-sparse, or $\|f\|_0 = r$, or exactly r entries of f are non-zero.
- Condition 2: Ω or Ψ_Ω satisfies the **Restricted Isometry Property (RIP)**.

The famous paper by Tao and Candes [1] has already revealed to us that under these two conditions, solving the optimization problem in step 3 would grant us exact recovery of f. This is a reduction from a difficult problem (exact recovery is equivalent to solving the problem in step 3 with $\|g\|_0$, which is non-convex) to a much easier problem (since $\|\cdot\|_1$ is convex).

1.2 This Problem Matters Because...

In reality, we're examining a specific instance of a much boarder class of problem: **Compressed Sensing**, or **Sparse Recovery**. To move apart a little from mathematical abstractions, the signal f we considered in the previous section can be thought of as "a function that conveys information about a phenomenon. Any quantity that can vary over space or time can be used as a signal to share messages between observers [5]." Some examples include audio, video, speech, image, sonar, and radar.

What role does sparsity play here? "Sparse signal models provide a mathematical framework for capturing the fact that in many cases these high-dimensional signals contain relatively little information compared to their ambient dimension [2]."

Typically, it is not that the signal itself is sparse but that it is naturally sparse in the measurement basis. In our example we considered a signal that is sparse itself, but if we were to consider another that's sparse in the Fourier basis, then we would be assuming sparsity on $\hat{f}$.

Measurements can also be of various forms. In our example, the Fourier measurement matrix is a deterministic matrix, and when we sample we are choosing its rows at random. We can also consider some other matrices where the entries are sampled from an i.i.d. Gaussian distributions (this would be Gaussian measurements), or if the rows are not chosen uniformly at random. Each method of measurement has its own advantage: for instance, sometimes the data collection process naturally samples from the transformed signal.

1.3 Today, We Will Explore...

Back to our problem setup, we haven't answered every question. How many samples do we need to take if we want to successfully recover f? We know by condition 2 that as long as Ψ_Ω satisfies the RIP condition, we can apply Tao's result. However, because Ω is random, Ψ_Ω is a random matrix, and of course it can only satisfy the RIP with certain probability.

We're interested in the trade-off between this success probability and the number of samples k we take. In reality, for a fixed success rate, the number of samples we take also depends on the signal dimension n and the signal's sparsity r. The following theorem formalizes (more than) what we wish to prove.

Theorem 1. *Let $t > 1$ be a trade-off variable*

- *$k(t, n, r)$ scales as $k = s \log s \log^2 r$, where $s(t, n, r) = O(tr \log n)$.*
- *The success probability p scales as $1 - \exp(-O(t))$.*

Reminder that Ω depends on $k(t, n, r)$. Then, for any t, with probability p, Ψ_Ω satisfies the RIP.

Remark 2. *This theorem is a major improvement over the previous Tao's result in the following way: fix $p = 0.99$ for example, and this would correspond to a t. Plug this t into $k(t, n, r)$, and we would get that $k(n, r) = O(r \log^4 n)$. The previous result by Tao gives a scaling factor $k'(n, r) = O(r \log^6 n)$. In the context when r is small, this improves sampling efficiency by orders of magnitude.*

Of course, we will not be able to prove everything related to Theorem 1. However, we will tackle the key argument in this proof. On a high level, in order to prove theorem 1, we must show Ψ_Ω satisfy the RIP with high probability. This is equivalent to proving that some random variable δ_r (δ_r depends on Ω) is small enough. The proof of theorem 1 first establishes that δ_r has a very small expectation, then uses traditional techniques in concentrated inequalities to tail bound δ_r. The first part of this argument is what we will focus on. Specifically, we will establish from scratch the reduction from the problem of bounding expectation to a problem of the expected supremum of Gaussian process. This is done through the following steps:

- First, we would like to not only formalize the RIP condition but also derive a more convenient form. This is done in section 2.1. As a result, we know that Ψ_Ω satisfies the RIP condition if the random variable $\delta_r > \frac{1}{2}$, where

$$\delta_r = \sup_{|T|\leq r} \left\| id_{\mathbb{C}^T} - \frac{1}{k}\sum_{i\in\Omega} y_i^T \otimes y_i^T \right\|.$$

- Second, section 3 takes a key lemma and proves an upper bound for $\mathbb{E}\delta_r$ through formalizing the randomness of Ω, implementing an argument of **symmetrization** using Rademacher random variables, applies the given lemma, and finally bounds $\mathbb{E}\delta_r$ by a function of itself using triangle inequality:

$$\mathbb{E}\delta_r \leq O(\sqrt{\frac{1}{t}}).$$

- Third, section 4.1 reduces lemma 7 to a problem involving Gaussian processes through the **comparison principle**.
- Finally, section 4.2 states some results in Dudley's inequality and Metric Entropy. It is impossible to build this part of the theory from scratch, but we provide the results necessary for the completion of our argument.

1.4 Preliminaries and Notations

This text is intended to be self-contained. In this section we briefly state all the common tools we will use and our notations.

Symbol	Description
x^T	If $x \in \mathbb{C}^n$ is a row or column vector, $T \subseteq \{1, ..., n\}$ is an index set, then x^T represents the entries of x at the indices in T. Note that $\lvert x^T\rvert = \lvert T\rvert$.
Ψ_T	If $\Psi \in \mathbb{C}^{m\times n}$ is a matrix and T is the same as before, then $\Psi_T \in \mathbb{C}^{\lvert T\rvert\times n}$ represents the rows of Ψ that are indexed by elements in T.
$\Psi_{T_1}^{T_2}$	Similarly, if T_1, T_2 are both index sets, then $\Psi_{T_1}^{T_2} \in \mathbb{C}^{\lvert T_1\rvert\times\lvert T_2\rvert}$ is the matrix Ψ having rows indexed by T_1 and columns indexed by T_2.
$id_{\mathbb{C}^T}$	The identity matrix in the space $\mathbb{C}^T$.
$\lVert A\rVert_{op}$	Operator norm. Suppose $A \in \mathbb{C}^{\lvert T_1\rvert\times\lvert T_2\rvert}$, then this norm is defined between $l_2^{T_1}$ and $l_2^{T_2}$ to be $\sup_{\lVert x\rVert_2=1} \lVert Ax\rVert_2$.
$x \otimes x$	Tensor Product between $x \in \mathbb{C}^n$ and itself. Given a vector $v \in \mathbb{C}^n$, we have $(x \otimes x)(v) = \langle x, v\rangle\, x$.

In addition to this, the other notations are standard. For example, $|T|$ means the size of T. A^* is the conjugate transpose of A. One should know $\|\cdot\|_i$ for $i \in \{0, 1, 2\}$. One should also be familiar with Jensen's inequality and Cauchy Schwarz inequality.

2 The Two Definitions of the RIP Condition

2.1 The original definitions of RIP

We provide, first, the formal definition of the RIP and its associated constant δ_r.

Definition 3. *Conditioned on Ω, the restricted isometry constant δ_r for $\Psi_\Omega \in \mathbb{C}^{k\times n}$ is defined to be the smallest positive number such that the following inequality holds:*

$$(1-\delta_r)\left\|x\right\|_2^2 \leq \left\|\Psi_\Omega^T x\right\|_2^2 \leq (1+\delta_r)\left\|x\right\|_2^2.$$

This relation must hold for any $x \in \mathbb{C}^{|T|}$ and any $T \subseteq \{1, ..., n\} : |T| \leq r$.

Definition 4. *We say Ψ_Ω satisfies the RIP if the constants δ_{3r} and δ_{4r} satisfy*

$$\delta_{3r} + 3\delta_{4r} \leq 2.$$

Now, we make the following observation:

Observation 5. *δ_r is non-decreasing in r, or $\delta_r \leq \delta_{2r} \leq ...$*

This is easy to see from definition 3 because a T with size at most r is also a T with size at most $2r$. Thus, it is sufficient to prove $4\delta_{4r} \leq 2$, or $\delta_{4r} \leq \frac{1}{2}$, for RIP to be satisfied. We can moreover replace r by $4r$ and change the constant in the definition of k.

2.2 A second definition of the RIP constant

Theorem 6. *An equal definition for δ_r is*

$$\delta_r = \sup_{|T|\leq r} \left\| id_{\mathbb{C}^T} - \frac{1}{k}\sum_{i\in\Omega} y_i^T \otimes y_i^T \right\|_{op}.$$

where y_i is the re-normalized i-th row of Ψ: $y_i^T = \sqrt{|\Omega|}\Psi_i^T$. $y_i^T \in \mathbb{C}^{|T|}$.

Proof. Starting with definition 3, we can arrange the expression to be

$$|\left\|\Psi_\Omega^T x\right\|_2^2 - \left\|x\right\|_2^2| \leq \delta_r \left\|x\right\|_2^2.$$

The LHS is simplified to be (we use linearity and conjugate symmetry of the inner product)

$$\left\|\Psi_\Omega^T x\right\|_2^2 - \left\|x\right\|_2^2 = \left\langle \Psi_\Omega^T x, \Psi_\Omega^T x\right\rangle - \langle x, x\rangle = \left\langle ((\Psi_\Omega^T)^*\Psi_\Omega^T - id_{\mathbb{C}^T})x, x\right\rangle.$$

Thus,

$$\delta_r \geq \frac{\left\langle ((\Psi_\Omega^T)^*\Psi_\Omega^T - id_{\mathbb{C}^T})x, x\right\rangle}{\left\|x\right\|_2^2} \text{ for all } x \text{ and } T.$$

Since we want to find the smallest positive number, this is equivalent to

$$\delta_r = \sup_{|T|\leq r, x\in\mathbb{C}^T} \frac{\left\langle ((\Psi_\Omega^T)^*\Psi_\Omega^T - id_{\mathbb{C}^T})x, x\right\rangle}{\left\|x\right\|_2^2} = \sup_{|T|\leq r, \left\|x\right\|_2^2=1} \left\|((\Psi_\Omega^T)^*\Psi_\Omega^T - id_{\mathbb{C}^T})x\right\|_2.$$

The last step is because $((\Psi_\Omega^T)^*\Psi_\Omega^T - id_{\mathbb{C}^T})$ is Hermitian. Now, by definition of operator norm which is provided before, we have

$$\delta_r = \sup_{|T|\leq r} \left\|((\Psi_\Omega^T)^*\Psi_\Omega^T - id_{\mathbb{C}^T})\right\|_{op}.$$

Lastly, consider a vector $v \in \mathbb{C}^{|T|}$. We have ($|\Omega| = k$)

$$(\Psi_\Omega^T)^*\Psi_\Omega^T v = (\Psi_\Omega^T)^*[\left\langle \Psi_1^T, v\right\rangle, ...]^\top = \sum_{i\in\Omega}\left\langle \Psi_i^T, v\right\rangle \Psi_i^T = (\sum_{i\in\Omega}\Psi_i^T \otimes \Psi_i^T)(v) = (\frac{1}{k}\sum_{i\in\Omega} y_i^T \otimes y_i^T)(v).$$

This completes the proof.

□

3 Symmetrization

3.1 Toolbox

Recall that we want to show if we take $k = (Ctr \log n) \log(Ctr \log n) \log^2 r$ uniform measurements, then $\delta_r > \frac{1}{2}$ with high probability. In the previous section we have quipped ourselves with a more convenient form of δ_r. In this section, we will make a further reduction using what's called a symmetrization technique. Specifically, we will first claim a lemma (we'll discuss it more in the next section), and then show how this lemma implies our desired result by symmetrization.

Lemma 7. *For any $m < n$, let $x_1, ..., x_m \in \mathbb{C}^n$ be vectors with uniformly bounded entries: $\|X_i\|_\infty \le K$ for all i, then*

$$\mathbb{E} \sup_{|T| \le r} \left\| \sum_{i=1}^{m} \epsilon_i x_i^T \otimes x_i^T \right\|_{op} \le k_1 \sup_{|T| \le r} \left\| \sum_{i=1}^{m} x_i^T \otimes x_i^T \right\|_{op}^{1/2}.$$

Here,

- *k_1 is a function of m, r, n, K: $k_1 \le C_1 K \sqrt{r} \log r \sqrt{\log n} \sqrt{\log m}$.*
- *ϵ_i are independent Rademacher random variables.*

We also need to formalize what uniform measurements mean:

Definition 8. *We define Ω in the following way: associate each number in $\{0, ..., n-1\}$ with a Bernoulli $p = k/n$. We include the row with ω equals to this number if the corresponding Bernoulli realization is 1. This would give us an Ω with size expected to equal k, or $\mathbb{E}|\Omega| = k$.*

3.2 The Argument

We start by considering the expectation of δ_r, which is

$$\mathbb{E}\delta_r = \mathbb{E} \sup_{|T| \le r} \left\| id_{\mathbb{C}^T} - \frac{1}{k} \sum_{i \in \Omega} y_i^T \otimes y_i^T \right\|_{op}.$$

The first observation we can make is that the two quantities inside of the operator norm equal in expectation (We're using the property that Ψ is orthonormal):

$$id_{\mathbb{C}^T} = \frac{1}{n} \sum_{i=1}^{n} y_i^T \otimes y_i^T = \mathbb{E} \frac{1}{k} \sum_{i \in \Omega} y_i^T \otimes y_i^T.$$

Now, we use the fact that $F(\cdot) := \sup_{|T| \le r} \|\cdot\|_{op}$ is a convex function because operator norm is convex and supremum of convex functions is convex. Moreover, denote $(X_i)_{i \in \Omega} := (\frac{1}{k} \cdot y_i^T \otimes y_i^T)_{i \in \Omega}$ and consider $(X_i')_{i \in \Omega}$ an independent copy of $(X_i)_{i \in \Omega}$,

$$\mathbb{E}\delta_r = \mathbb{E}F((\mathbb{E} \sum_{i \in \Omega} X_i) - \sum_{i \in \Omega} X_i) = \mathbb{E}F(\sum_{i \in \Omega} (X_i - \mathbb{E}X_i)) \le \mathbb{E}F(\sum_{i \in \Omega} (X_i - X_i')).$$

In the last step we used the fact that $\mathbb{E}X_i = \mathbb{E}X_i'$ and then applied Jensen's inequality. We observe that since $(X_i - X_i')_{i \in \Omega}$ is a sequence of independent symmetric random variables, it has the same distribution as $(\epsilon_i (X_i - X_i'))_{i \in \Omega}$. Thus, by convexity of F,

$$\mathbb{E}F(\sum_{i \in \Omega} (X_i - X_i')) = \mathbb{E}F(\sum_{i \in \Omega} \epsilon_i (X_i - X_i')) \le \mathbb{E}(\frac{1}{2} F(2 \sum_{i \in \Omega} \epsilon_i X_i) + \frac{1}{2} F(2 \sum_{i \in \Omega} \epsilon_i X_i')) \le \mathbb{E}F(2 \sum_{i \in \Omega} \epsilon_i X_i).$$

We've thus shown with the symmetrization argument that

$$\mathbb{E}\delta_r \le 2\mathbb{E}\sup_{|T|\le r}\left\|\frac{1}{k}\sum_{i\in\Omega}\epsilon_i y_i^T\otimes y_i^T\right\|_{op}.$$

A key observation is that the RHS an expectation taken over both Ω and $(\epsilon_i)_{i=1}^{|\Omega|}$. Conditioning on Ω would put us in the setting of using lemma 7 because y_i becomes deterministic:

$$\begin{aligned}
2\mathbb{E}\sup_{|T|\le r}\left\|\frac{1}{k}\sum_{i\in\Omega}\epsilon_i y_i^T\otimes y_i^T\right\|_{op} &= 2\mathbb{E}_\Omega\mathbb{E}_\epsilon\sup_{|T|\le r}\left\|\frac{1}{k}\sum_{i=1}^{m}\epsilon_i y_{a_i}^T\otimes y_{a_i}^T\right\|_{op} \mid \Omega=\{a_1,...,a_m\}\\
&\le \frac{1}{k}2\mathbb{E}_\Omega k_1\sup_{|T|\le r}\left\|\sum_{i\in\Omega}y_i^T\otimes y_i^T\right\|_{op}^{1/2}\\
&\le \frac{2C_1K\sqrt{r}\log r\sqrt{\log n}}{k}\mathbb{E}_\Omega\sqrt{\log|\Omega|}\sup_{|T|\le r}\left\|\sum_{i\in\Omega}y_i^T\otimes y_i^T\right\|_{op}^{1/2}\\
&\le \frac{2C_1K\sqrt{r}\log r\sqrt{\log n}}{k}(\mathbb{E}_\Omega\log|\Omega|)^{1/2}(\mathbb{E}_\Omega\sup_{|T|\le r}\left\|\sum_{i\in\Omega}y_i^T\otimes y_i^T\right\|_{op})^{1/2}.
\end{aligned}$$

In the third step, because we applied lemma 7 with $m=|\Omega|$ being a random quantity, the $\sqrt{\log|\Omega|}$ needs to stay in the expectation. In the next step we applied Cauchy-Schwarz's inequality, namely $\mathbb{E}XY\le(\mathbb{E}X^2)^{1/2}(\mathbb{E}Y^2)^{1/2}$. Now, we have to deal with the two terms in expectation respectively. For the first one, Jensen's inequality gives us $\mathbb{E}\log|\Omega|\le\log\mathbb{E}|\Omega|=\log k$. For the second one, we use the triangular inequality for norms, namely $\|u-v\|\ge|\,\|u\|-\|v\|\,|$:

$$\mathbb{E}\sup_{|T|\le r}\left\|id_{\mathbb{C}^T}-\frac{1}{k}\sum_{i\in\Omega}y_i^T\otimes y_i^T\right\|_{op}\ge\mathbb{E}\sup_{|T|\le r}\left\|\frac{1}{k}\sum_{i\in\Omega}y_i^T\otimes y_i^T\right\|_{op}-\mathbb{E}\sup_{|T|\le r}\|id_{\mathbb{C}^T}\|_{op}.$$

This implies

$$\mathbb{E}\sup_{|T|\le r}\left\|\sum_{i\in\Omega}y_i^T\otimes y_i^T\right\|_{op}\le\mathbb{E}\sup_{|T|\le r}\|id_{\mathbb{C}^T}\|_{op}-\mathbb{E}\delta_r\le k(1+\mathbb{E}\delta_r).$$

Plugging this back into our previous result yields

$$\mathbb{E}\delta_r\le\frac{2C_1K\sqrt{r}\log r\sqrt{\log n}}{\sqrt{k}}\sqrt{\log k}\sqrt{\mathbb{E}\delta_r+1}\le C_2\frac{\sqrt{\log tr\log n+\log\log tr\log n+\log\log r}}{\sqrt{t\log(tr\log n)}}.$$

Here, we plugged in the value of k assumed in the very beginning. Since $\mathbb{E}\delta_r$ is small, the term $1+\mathbb{E}\delta_r$ is approximated to be 1. Observe that the $\log\log$ terms all get dominated. We thus have

$$\mathbb{E}\delta_r\le C_3\frac{1}{\sqrt{t}}.$$

4 From Rademacher to Gaussian

4.1 Reduction to Gaussian Process

In this section, we reduce the lemma 7 provided in the previous section to a Gaussian process formulation. Formally, this is called the **Comparison Principle**.

We start with, $F(\cdot) = \sup_{|T|\leq r}$ and $(X_i)_{i=1}^m = (x_i^T \otimes x_i^T)_{i=1}^m$. Consider, again ϵ_i to be independent Rademacher random variable. Consider also g_i as an independent sequence of standard Gaussians.

Symmetry gives us that $\epsilon_i|g_i|$ has the same distribution as g_i. The quantity on the LHS of lemma 7 is

$$\mathbb{E}F(\left\|\sum_{i=1}^m g_i X_i\right\|_{op}) = \mathbb{E}F(\left\|\sum_{i=1}^m \epsilon_i |g_i| X_i\right\|_{op}) \geq \mathbb{E}F(\left\|\sum_{i=1}^m \epsilon_i \mathbb{E}|g_i| X_i\right\|_{op}) \geq \mathbb{E}F(\sqrt{\frac{2}{\pi}}\left\|\sum_{i=1}^m \epsilon_i X_i\right\|_{op}).$$

Where the second step comes from Jensen's inequality and partial integration. Note that we already claimed F is convex in the previous part. The last step comes from the expected absolute value of a standard Gaussian: namely, $\mathbb{E}|g_i| = \sqrt{2/\pi}$. We can take this constant out of F. Now, we have a bound using Gaussian process:

$$\mathbb{E}\sup_{|T|\leq r}\left\|\sum_{i=1}^m \epsilon_i x_i^T \otimes x_i^T\right\|_{op} \leq C_4 \mathbb{E}\sup_{|T|\leq r}\left\|\sum_{i=1}^m g_i x_i^T \otimes x_i^T\right\|_{op}.$$

We apply the definition of operator norm:

$$\left\|\sum_{i=1}^m g_i x_i^T \otimes x_i^T\right\|_{op} = \sup_{v:\|v\|_2\leq 1}\left\|(\sum_{i=1}^m g_i x_i^T \otimes x_i^T)(v)\right\|_2 = \sup_{v:\|v\|_2\leq 1}|\sum_{i=1}^m g_i \left\langle x_i^T, v\right\rangle^2|.$$

Plug this back into the expectation, and let B_2^T denote the unit ball of l_2 on $\mathbb{C}^T$. We have

$$\mathbb{E}\sup_{|T|\leq r}\left\|\sum_{i=1}^m g_i x_i^T \otimes x_i^T\right\|_{op} = \mathbb{E}\sup_{|T|\leq r, v\in B_2^T}|\sum_{i=1}^m g_i \left\langle x_i, v\right\rangle^2|.$$

We've finally reached the end of the series of arguments!

4.2 Some Results from Chaining and Metric Entropy

The expected supremum of a Gaussian process is a well studied problem [3]. Because this is not the direct goal of this project, we will not explore the theory from scratch. However, in this section we still provide some results that would complete the proof of our lemma 7.

Definition 9. *Gaussian Process: $(X_t)_{t\in T}$ is a Gaussian process if for every finite subset of the index set T, every linear combination of the X-s defined by them is Gaussian.*

We see that what we arrived at in the end of the last section is a Gaussian Process on the index set $\cup_{|T|\leq r} B_2^T$. Take an arbitrary number of elements from this set, we now have a jointly Gaussian vector because each element is a linear combination of the g_i-s.

Specifically, with our index set $\cup_{|T|\leq r} B_2^T$, we can apply the following theorem:

Theorem 10. *Dudley's Inequality: let $X = (X_t)_{t\in T}$ be a Gaussian process, then*

$$\mathbb{E}\sup_{t\in T} X_t \leq C_4 \int_0^\infty \log N(T, d_X; \epsilon)^{1/2}\, d\epsilon.$$

where C_4 is some constant.

Definition 11. *Entropy Number: fix $\epsilon > 0$, we define the entropy number $N(T, d; \epsilon)$ is the minimum number of open balls of radius ϵ necessary to cover T.*

In our case, the metric we use ends up being

$$d_X(x,y) = 2 \sup_{|T|\leq r} \left\| \sum_{i=1}^{m} x_i^T \otimes x_i^T \right\|^{1/2} \max_{i\leq k} |\langle x_i, x-y\rangle|.$$

Notice that the first term in the above expression appears in the end result of lemma 7. We basically separate it out and bound the entropy number in order to obtain the desired k_1 constant as given in the lemma. This would complete of proof of the lemma, and thus complete the proof that $\mathbb{E}\delta_r$ is small.

References

[1] Emmanuel Candes and Terence Tao. *Decoding by Linear Programming*. 2005. arXiv: math/0502327 [math.MG]. URL: https://arxiv.org/abs/math/0502327.

[2] Mark A. Davenport et al. "Introduction to compressed sensing". In: *Compressed Sensing: Theory and Applications*. Ed. by Yonina C. Eldar and GittaEditors Kutyniok. Cambridge University Press, 2012, pp. 1–64.

[3] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces isoperimetry and processes Michel Ledoux; Michel Talagrand*. Springer, 1991.

[4] Mark Rudelson and Roman Vershynin. "On sparse reconstruction from Fourier and Gaussian measurements". In: *Communications on Pure and Applied Mathematics* 61.8 (2008), pp. 1025–1045. DOI: https://doi.org/10.1002/cpa.20227. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.20227. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.20227.

[5] Wikipedia contributors. *Signal – Wikipedia, The Free Encyclopedia*. [Accessed: 15-Dec-2024]. 2024. URL: https://en.wikipedia.org/wiki/Signal.