Sparse recovery

Dhanus M Lal
Advisor: Manjunath Krishnapur

April 2020

Abstract

The matrix A of dimension $m \times n$, where $m \ll n$ is known. And a linear measurement of x, $y = Ax$ is known, where A is a known matrix. The goal is to find x from the measurement. Without any assumptions on x, the system is underdetermined and has a subspace of dimension at least $n - m$ solutions. But when it is known that most of the coordinates of x are zero (i.e., x is sparse), useful estimates for the solution x can be obtained and under some circumstances, the solution can be exactly recovered. This report will discuss the main results that allow recovery of sparse vectors from a linear measurement Ax.

1 Preliminaries

Definition 1.1. A random variable X is said to be *sub-gaussian* if $\mathbb{P}(|X| \geq t) \leq 2\exp(-t^2/K^2)$ for every $t \geq 0$. The *sub-gaussian norm*, $\|X\|_{\psi_2}$, of X is then proportional to K. A random vector X in $\mathbb{R}^n$ is sub-gaussian if $\langle X, x\rangle$ is subgaussian for every $x \in \mathbb{R}^n$. Its sub-gaussian norm, $\|X\|_{\psi_2} := \sup_{x\in\mathbb{R}^n} \|\langle X, x\rangle\|_{\psi_2}$.

Definition 1.2. The *gaussian width* $w(T)$ of a subset $T \subseteq \mathbb{R}^n$ is defined as $\mathbb{E}\sup_{x\in T}\langle x, g\rangle$ and its *gaussian complexity* as $\gamma(T) = \mathbb{E}\sup_{x\in T}|\langle x, g\rangle|$, where $g \sim N(0, I_n)$.

Definition 1.3. A random vector $X \in \mathbb{R}^n$ is said to be *isotropic* if $\mathbb{E}[XX^T] = I_n$, where I_n is the identity matrix in $\mathbb{R}^n$.

Definition 1.4 (Sparsity)**.** The *support* of a vector $x \in \mathbb{R}^n$ is the set $\{i \in \{1, \dots, n\} : x_i \neq 0\}$ (the set of non-zero indices). The *sparsity* of a vector x in $\mathbb{R}^n$ is defined as the cardinality of its support and is denoted by $\|x\|_0$. A vector with $\|x\|_0 \leq s$ is said to be $s-$sparse.

The problem we are trying to solve is:

$$\text{recover } x \text{ from } y = Ax, \text{ when } x \in T \tag{1}$$

where $y \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ and $m \ll n$, when it is known that $x \in T$. Without the condition that $x \in T$, the set of all possible values for x forms a subspace of dimension at least $n - m$. Adding this restriction can reduce the number of possibilities for x significantly. For sparse recovery, we are interested in the case when $T = S$, where $S = \{x \in \mathbb{R}^n : \|x\|_0 \leq s\}$.

Henceforth, x will denote the correct solution of problem (1), y will represent the linear measurement Ax of x and S will denote the set of s-sparse vectors. C, C', c, c' and c with sub-scripts will always denote universal constants.

2 Recovery based on M^* bound

Theorem 2.1 (Matrix deviation inequality)**.** *Suppose A is an $m \times n$ real matrix, whose rows A_i are independent, isotropic, sub-gaussian random vectors in $\mathbb{R}^n$ then for $T \subseteq \mathbb{R}^n$,*

$$\mathbb{E}\sup_{x\in T}|\|Ax\|_2 - \sqrt{m}\|x\|_2| \leq CK^2\gamma(T). \tag{2}$$

Here $K = \max_i \|A_i\|_{\psi_2}$.

This theorem is proved by first showing that the increments of the random process in the argument is bounded by a gaussian random process and then using Talagrand's comparison inequality.

Theorem 2.2 (M^* bound)**.** *Suppose $T \subseteq \mathbb{R}^n$ and A is an $m \times n$ real matrix whose rows A_i are independent, isotropic, sub-gaussian random vectors in $\mathbb{R}^n$, then for any $z \in \mathbb{R}^n$, $E = z + \ker(A)$ satisfies*

$$\mathbb{E}diam(T \cap E) \leq \frac{CK^2 w(T)}{\sqrt{m}}. \tag{3}$$

Here, K is the same as that in the previous theorem.

Proof. Use matrix deviation inequality on the set $T - T$ to get

$$\mathbb{E}\sup_{a,b\in T}|\|Aa - Ab\|_2 - \sqrt{m}\|a - b\|_2| \leq CK^2\gamma(T - T)$$

The right hand side is equal to $2CK^2w(T)$ because $\gamma(T - T) = 2w(T)$. If we restrict a, b to be in E, then $Aa - Ab = z - z = 0$ and the left side of the inequality above becomes $\sqrt{m}(\text{diam}(T \cap E))$. ∎

A candidate solution for the problem (1) is any vector $\hat{x} \in T$ satisfying $A\hat{x} = y$. If we take A in (1) to be as in the hypothesis of the above theorem, Then theorem 2.2 gives a bound on expected ℓ_2 error as

$$\mathbb{E}\sup_{\substack{x'\in T\\ Ax'=y}}\|x - x'\|_2 \leq \frac{CK^2}{\sqrt{m}}w(T) \tag{4}$$

From the bound (4), an approximate solution for the recovery problem (1) in the case when $T = S$ can be posed as an optimization problem as follows:

$$\begin{aligned} &\tilde{x} = \arg\min \|x'\|_0 \\ &\text{subject to: } x' \in S, Ax' = y \end{aligned} \tag{5}$$

Then the bound in (4) is valid for $\tilde{x}$.

This problem is computationally hard. So, we try to approximate by using the closest norm to $\|\cdot\|_0$, which is the ℓ_1 norm. To that end, we show that the set of unit s-sparse vectors is contained in a scaled ℓ_1 ball:

Proposition 2.1. *Let $B_1^n, B_2^n \in \mathbb{R}^n$ denote the unit balls with respect to the ℓ_1 and ℓ_2 norms respectively, then $S \cap B_2^n \subseteq \sqrt{s} B_1^n$*

Proof. For any $x \in S \cap B_2^n$, using Cauchy-Schwarz inequality, $\|x\|_1 \leq \sqrt{\|x\|_0} \cdot \|x\|_2 \leq \sqrt{s}$. ■

Now, applying (4) with $T = S \cap B_2^n$, we get

$$\mathbb{E} \sup_{\hat{x} \in B_1^n, A\hat{x} = y} \|x - \hat{x}\|_2 \leq CK^2 \sqrt{\frac{s \log n}{m}}. \tag{6}$$

Here we have used the linearity propery of Gaussian width, $w(\sqrt{s}B_1^n) = \sqrt{s} w(B_1^n)$ and that $w(B_1^n) = \mathbb{E} \max_{i \leq n} g_i$, where g_i are standard normal random variables. And finally we used the result that the expected maximum of n gaussians can be bounded by $C\sqrt{\log N}$ for a universal constant C.

The condition that $\|x\|_2 \leq 1$ can be removed and the problem can be stated as an optimization problem like (5) as follows:

$$\begin{aligned} &\hat{x} = \arg\min \|x'\|_1 \\ &\text{subject to: } Ax' = y \end{aligned} \tag{7}$$

This is a convex optimization problem called basis-pursuit. It is computationally feasible and can be solved using standard convex optimization methods like linear programming.

Theorem 2.3. *$\hat{x}$, the solution of basis pursuit (7) satisfies*

$$\mathbb{E}\|x - \hat{x}\|_2 \leq CK^2 \sqrt{\frac{s \log n}{m}} \|x\|_2$$

Proof. Since $\hat{x}$ is the minimizer, $\|\hat{x}\|_1 \leq \|x\|_1 \leq \sqrt{s}\|x\|_2$. Thus $\hat{x}/\|x\|_2, x/\|x\|_2 \in \sqrt{s}B_1^n$. Using (6), we get $\mathbb{E}\|(\hat{x} - x)/\|x\|_2\|_2 \leq CK^2\sqrt{(s \log n)/m}$. Taking $\|x\|_2$ to the other side gives the result. ■

3 Exact recovery

In the previous section, we observed that a solution with small expected ℓ_2 error can be obtained using basis pursuit. In this section, we will find the conditions under which the error becomes zero.

3.1 Recovery using Escape theorem

Theorem 3.1. *In problem (1), suppose the rows of A are independent, isotropic sub-gaussian random vectors and $T = S$. If $m \geq CK^4 s \log n$, then with probability at least $1 - 2\exp(-cm/K^4)$, basis pursuit solves the sparse recovery problem (that is, $\hat{x} = x$).*

Remark 1. The number of measurements m depends linearly on sparsity s and logarithmically on the dimension of the underlying space. This suggests that even for larger dimensional spaces, sparse recovery is possible with few meausrements.

The main result that will be used to prove the theorem is escape theorem, which states that the kernel of a matrix with sub-gaussian rows misses sufficiently small subset of the unit sphere with high probability.

Theorem 3.2. *Consider $T \subseteq S^{n-1}$. Let A be an $m \times n$ matrix whose rows are independent, isotropic sub gaussian random vectors in $\mathbb{R}^n$. If $m \geq CK^4 w(T)^2$, then $T \cap \ker(A) = \emptyset$ with probability at least $1 - 2\exp(-cm/K^4)$.*

Proof. The proof uses tail bound version of matrix deviation inequality

$$\sup_{a \in T} |\|Aa\|_2 - \sqrt{m}\|a\|_2| \leq CK^2(w(T) + u)$$

with probability at least $1 - 2\exp(-u^2)$. Suppose the above event happens and $T \cap \ker(A) \neq \emptyset$. If $a \in T \cap \ker(A)$, then putting $u = \sqrt{m}/(2CK^2)$, the above inequality becomes $\sqrt{m} \leq CK^2 w(T) + \sqrt{m}/2$. That is, $m \leq C'K^4 w(T)^2$, which is contrary to the hypothesis. So $T \cap E = \emptyset$ happens with probability at least $1 - 2\exp(-(cm)/K^4)$ ■

Before proving theorem 3.1, we will prove the following lemma:

Lemma 3.1. *Let $h = \hat{x} - x$*

1. *If $S = supp(x)$, then $\|h_S\|_1 \geq \|h_{S^c}\|_1$*
2. *$\|h\|_1 \leq 2\sqrt{s}\|h\|_2$*

Proof. Since $\hat{x}$ is the ℓ_1 norm minimizer and $x_{S^c} = 0$, we have $0 \geq \|\hat{x}\|_1 - \|x\|_1 = \|\hat{x}_S\|_1 + \|\hat{x}_{S^c}\|_1 - \|x_S\|_1 = (\|\hat{x}_S\|_1 - \|x_S\|_1) + \|h_{S^c}\|_1 \geq -\|\hat{x}_S - x_S\|_1 + \|h_{S^c}\|_= -\|h_S\|_1 + \|h_{S^c}\|_1$. This proves the first part.
For the second part, $\|h\|_1 = \|h_S\|_1 + \|h_{S^c}\|_1 \leq 2\|h_S\|_1 \leq 2\sqrt{s}\|h\|_2$

■

Proof of theorem 3.1. Suppose $h \neq 0$, then $h/\|h\|_2 \in S^{n-1}$. Also, from part 2 of the lemma, $h/\|h\|_2 \in 2\sqrt{s}B_1^n$. So, if $T = S^{n-1} \cap (2\sqrt{s}B_1^n)$, then $h/\|h\|_2 \in T \cap (\ker A)$, since $Ax = A\hat{x}$. Now, $m \geq CK^4 s \log n \geq C'K^4 w(T)^2$ (because $w(T) \leq w(2\sqrt{s}B_1^n)$), therefore, by theorem 3.2 $T \cap (\ker A) = \emptyset$ with high probability. Thus $\hat{x} = x$ with probability at least $1 - 2\exp(-cm/K^4)$. ■

3.2 Recovery using RIP

Definition 3.1. An $m \times n$ matrix A satisfies Restricted Isometry Property (RIP) with parameters α, β and s if for every s-sparse vector w,

$$\alpha\|w\|_2 \leq \|Aw\|_2 \leq \beta\|w\|_2$$

Remark 2. It is clear from the definition that A satisfies RIP with parameters α, β and s if and only if all the singular values of all $m \times s$ sub-matrices of A lie between α and β.

Theorem 3.3. *Suppose in* (1)*, A satisfies RIP with parameters α, β and $(1+\lambda)s$, with $\lambda > (\beta/\alpha)^2$. Then basis pursuit* (7) *exactly recovers the slolution to the sparse recovery problem.*

Proof. Let I_0 be the support of x, I_1 be the λs largest indices of $h_{I_0^c}$, I_2 be the next λs largest indices and so on. Let $I_{0,1} = I_0 \cup I_1$. For any set of indices $I \subseteq \{1, \dots n\}$, let A_I be the $m \times |I|$ submatrix of A with all but the columns in I deleted. Let $h = \hat{x} - x$.
By triangle inequality, $\|Ah\|_2 = \|A_{I_{0,1}}h_{I_{0,1}} + A_{I_{0,1}^c}h_{I_{0,1}^c}\|_2 \geq \|A_{I_{0,1}}h_{I_{0,1}}\|_2 - \|A_{I_{0,1}^c}h_{I_{0,1}^c}\|_2$. Left hand side is zero, since $Ah = A\hat{x} - Ax = 0$. So,

$$\|A_{I_{0,1}}h_{I_{0,1}}\|_2 \leq \|A_{I_{0,1}^c}h_{I_{0,1}^c}\|_2 \tag{8}$$

Now, by RIP,

$$\alpha\|h_{I_{0,1}}\|_2 \leq \|A_{I_{0,1}}h_{I_{0,1}}\|_2 \tag{9}$$

By triangle inequality and RIP,

$$\|A_{I_{0,1}^c}h_{I_{0,1}^c}\|_2 \leq \sum_{j\geq 2}\|A_{I_j}h_{I_j}\|_2 \leq \beta\sum_{j\geq 2}\|h_{I_j}\|_2 \tag{10}$$

Now observe that for any j, $\|h_{I_j}\|_2 \leq \sqrt{s\lambda}|\max h_{I_j}| \leq \sqrt{s\lambda}|\min h_{I_{j-1}}|$ (here max and min are largest and smallest coordinates respectively). Now, the smallest coordinate of $h_{I_{j-1}}$ is smaller than the average of the other coordinates so, $\|h_{I_j}\|_2 \leq \sqrt{\lambda s}\frac{\|h_{I_{j-1}}\|_1}{\lambda s}$. Substituting this in (10) gives

$$\|A_{I_{0,1}^c}h_{I_{0,1}^c}\|_2 \leq (\beta/\sqrt{s\lambda})\|h_{I_0^c}\|_1 \tag{11}$$

From the first part of lemma 3.1, $\|h_{I_0^c}\|_1 \leq \|h_{I_0}\|_1$ By Cauchy-Schwarz inequality, $\|h_{I_0}\|_1 \leq \sqrt{s}\|h_{I_0}\|_2$ Substituting this in (11), we get

$$\|A_{I_{0,1}^c}h_{I_{0,1}^c}\|_2 \leq \frac{\beta}{\sqrt{\lambda}}\|h_{I_0}\|_2$$

From (8), (9) and the above inequality, we get $\alpha\|h_{I_{0,1}}\|_2 \leq \frac{\beta}{\sqrt{\lambda}}\|h_{I_0}\|_2$ From the condition on α, β, λ, This can happen only if $\|h_{I_{0,1}}\|_2 = 0$. Since $0 = \|h_{I_1}\|_2 \geq \|h_{I_2}\|_2 \geq \dots$, we must have $h_{I_0^c} = h_{I_0} = 0$. ■

Even though RIP gives a sufficient condition for exact recovery, constructing matrices that satisfy RIP is difficult. However, one can generate random matrices that satisfy RIP with high probability. We need bounds on the singular values of random matrices to prove such a result. We will use the following bound on the singular values in the proof.

Theorem 3.4. *If A is an $m \times n$ matrix with whose rows are mean zero, isotropic, sub-gaussian random vectors in $\mathbb{R}^n$, then for any $t \geq 0$,*

$$\sqrt{m} - CK^2(\sqrt{n}+t) \leq s_m(A_I) \leq s_1(A_I) \leq \sqrt{m} + CK^2(\sqrt{n}+t) \tag{12}$$

with probability at least $1 - 2\exp(-t^2)$. (here $s_i(A)$ represents the ith singular value of A and K is as before)

Sketch of proof: The proof of uses the result that the operator norm $\|\frac{1}{m}A^TA - I_n\| \leq \max(\delta, \delta^2)$ implies all the singular values of $A/\sqrt{m}$ lie between $1-\delta$ and $1+\delta$.
First, an ϵ net $\mathcal{N}$ is used to bound the supremum over the unit circle by a supremum over a finite set:

$$\sup_{x\in S^{n-1}}\langle(\frac{1}{m}A^TA - I_n)x, x\rangle \leq C(\epsilon)\sup_{x\in\mathcal{N}}\langle(\frac{1}{m}A^TA - I_n)x, x\rangle$$

After that, Bernstein's inequality is used to find tail bounds for $\langle(\frac{1}{m}A^TA - I_n)x, x\rangle = |\|\frac{Ax}{\sqrt{m}}\|_2^2 - 1| = \sum_i \frac{1}{m}(\langle A_i, x\rangle^2 - 1)$. Then a union bound over $\mathcal{N}$ is taken and finally, an appropriate value of δ is chosen to arrive at (12). ■

We can now use the bound (12) to generate random matrices that satisfies RIP with high probability.

Theorem 3.5. *Suppose A is an $m \times n$ matrix with mean zero, isotropic, subgaussian rows. If $m \geq CK^4 s\log n$, then A satisfies RIP with parameters $\alpha = 0.9\sqrt{m}, \beta = 1.1\sqrt{m}$ and s with probability at least $1 - 2\exp(-cm/K^4)$.*

Proof. From remark 2, we need to bound the singular values for each $m \times s$ submatrix of A.
Fix a subset I, of size s, of $\{1, \dots, n\}$. Consider the $m \times s$ submatrix A_I. Applying the bounds on singular values of random matrices with mean zero, isotropic, subgaussian rows to A_I, we get:

$$\sqrt{m} - CK^2(\sqrt{s}+t) \leq s_s(A_I) \leq s_1(A_I) \leq \sqrt{m} + CK^2(\sqrt{s}+t)$$

for any $t \geq 0$, with probability at least $1 - 2\exp(-t^2)$. Putting $t = \sqrt{m}/(DCK^2)$ and setting appropriate values for the constants, we can get the singular values to be bounded by $0.9\sqrt{m}$ and $1.1\sqrt{m}$ with probability at least $1 - 2\exp(-2mc_1/K^4)$ (Here $2c_1 = 1/(D^2C^2)$). A union bound can be used to find the probability that this happens for each subset I of size s. This gives the probability that A satisfies RIP to be at least $1 - 2\exp(\frac{-2mc_1}{K^4})\binom{n}{s}$. Now, using $\binom{n}{s} \leq n^s$, the bound on the probability becomes $1 - 2\exp(\frac{-2mc_1}{K^4} + s\log n)$. From the hypothesis, the exponent is at most $-\frac{mc_1}{K^4}$, proving the result. ■

Remark 3. The above theorem can be used to prove theorem 3.1: By replacing s by $3s$ in the theorem, A satisfies RIP with parameters $0.9\sqrt{m}, 1.1\sqrt{m}$ and $3s$ with probability at lest $1 - 2\exp(-cm/K^4)$. A then satisfies the hypothesis of theorem 3.3 with $\lambda = 2 \geq (1.1/0.9)^2$. Thus exact recovery of s-sparse signals is possible by the conclusion of theorem 3.3.

4 Low rank matrix recovery

In this section, the problem we are trying to tackle is

$$\text{recover } X \text{ from } y_i = \langle A_i, X\rangle,\ i = 1\dots,m \text{ when } X \in T \tag{13}$$

where $y_i \in \mathbb{R}^m$, X, A_i are $d \times d$ matrices and T is a subset of real $d \times d$ matrices. The inner product is defined as $\langle A, B\rangle = \mathrm{tr} A^T B$ for $d \times d$ matrices.

Till now, our notion of sparsity was $\|\cdot\|_0$. For matrices, there is another notion of sparsity, which is the rank. So we are interested in the solution of problem (13) when $T = R = \{M \in \mathbb{R}^{d\times d} : \mathrm{rank}(M) \leq r\}$. But, like $\|\cdot\|_0$, rank is also not a norm. Hence S is not a convex set. As before, to convexify the set, we replace rank with the closest norm, which is the nuclear norm $\|\cdot\|_*$, defined as the sum of singular values.

Remark 4. For the $d \times d$ matrix A, if we take $v \in \mathbb{R}^d$ to be the d dimensional vector of singular values, $\mathrm{rank}(A) = \|v\|_0$, $\|A\|_* = \|v\|_1$ and the spectral norm $\|A\| = \|v\|_\infty$. The Frobenius norm of A, which is the square root of sum of squares of all entries of A can also be written in terms of singular values as $\|A\|_F = \|v\|_2$.

The equivalent to basis pursuit here is

$$\begin{aligned}&\hat{X} = \arg\min\|X'\|_* \\ &\text{subject to: } \langle A_i, X'\rangle = y_i,\ \text{for each } i = 1\dots n\end{aligned} \tag{14}$$

From the above remark and proposition 2.1, we have $R \cap B_F \subseteq \sqrt{r}B_*$ (B_F and B_* are unit balls in Frobenius and nuclear norms). Now, if we assume $T = B_F \cap R$, the calculations that lead to (6) in this case gives

$$\mathbb{E} \sup_{\substack{\hat{X}' \in \sqrt{r}B_1 \\ \langle A_i, X'\rangle = y_i}} \|X' - X\|_F \leq \frac{CK^2\sqrt{r}w(B_*)}{\sqrt{m}} \tag{15}$$

Note that $K = \max_i \|A_i\|_{\psi_2}$, where A_i must be seen as a $d \times d$ vector.

Proposition 4.1. *The gaussian width of the unit ball under nuclear norm, $w(B_*) \leq \mathbb{E}\|G\|$, where $\|\cdot\|$ is the operarator norm and G is a matrix with each entry being independent $N(0, 1)$ random variables.*

Proof. $w(B_*) = \mathbb{E} \sup\langle G, M\rangle$, where $\|M\|_* = 1$. Now, $\langle G, M\rangle = \mathrm{tr}(M^T G)$. If $M = A\Sigma B^T$ is the singular value decomposition of M, using the cyclic propery of trace, we get $\mathrm{tr}(M^T G) = \mathrm{tr}(\Sigma A^T GB)$. Since Σ is a diagonel matrix of singular values, $\mathrm{tr}(\Sigma A^T GB) = \sum_{i=1}^d s_i(M)\langle a_i, Gb_i\rangle$, where a_i and b_i are the columns of A and B respectively. Now, a_i and b_i are unit vectors, so by using Cauchy-Schwarz inequality, we get $\mathrm{tr}(M^T G) \leq \|G\| \sum_i s_i(M) = \|G\|$. ■

To find the expected operator norm of G, we observe that $(G_{uv})_{u,v\in S^{d-1}}$ is a mean zero gaussian process and that the increments of the random process G_{uv}, $\mathbb{E}(G_{u_1v_1} - G_{u_2v_2})^2$ are bounded by the increments of another mean zero gaussian process $(H_{uv})_{u,v\in S^{d-1}} = \langle g, u\rangle + \langle h, v\rangle$, where g, v are independent standard normal random vectors in $\mathbb{R}^d$. Now, Sudakov-Fernique comparison inequality can be used to obtain $\sup_{u,v\in S^{d-1}} G_{uv} \leq \sup_{u,v\in S^{d-1}} H_{uv} = \mathbb{E}\|g\|_2 + \mathbb{E}\|h\|_2 \leq 2\mathbb{E}\sqrt{\|h\|_2^2} = 2\sqrt{d}$. Using this and theorem 2.3, we have the following bound:

Theorem 4.1. *If $\hat{X}$ is the solution of problem 14, then*

$$\mathbb{E}\|\hat{X} - X\|_F \leq CK^2\sqrt{\frac{rd}{m}}\|X\|_F$$

Proof. $\hat{X}/\|X\|_F$ and $X/\|X\|_F$ are in $\sqrt{r}B_*$. Now (15) can be used to get the result. ■

References

[1] Manjunath Krishnapur. Gaussian processes.

[2] Holger Rauhut Simon Foucart. *A Mathematical Introduction to Compressive Sensing*. Applied and Numeric Harmonic Analysis Series. Springer, 2013.

[3] Michel Talagrand. *Upper and Lower Bounds for Stochastic Processes*, chapter 1. Springer, second edition edition, 2014. [Accessible online: http://michel.talagrand.net/ULBSPRINGER.pdf].

[4] Terrence Tao. Compressed sensing - robust recovery of sparse signals from limited measurements. https://terrytao.files.wordpress.com/2008/03/anziam.pdf, February 2008.

[5] Terrence Tao. Uniform uncertainity principle and compressed sensing. https://terrytao.files.wordpress.com/2008/12/uup.pdf, December 2008.

[6] Roman Vershynin. *High-Dimensional Probability - An Introduction with Applications in Data Science*. Cambridge Series in Statistic and Probabilistic Mathematics. Cambridge University Press, 2020. [Draft version : https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.pdf].