# Sparse Recovery

Dhanus M Lal
Advisor: Manjunath Krishnapur

April 2022

**Abstract**

Suppose a linear measurement, $y = Ax$ of an unknown vector $x \in \mathbb{R}^n$ is given, where the measurement matrix $A \in \mathbb{R}^{m\times n}$, with $m \ll n$ is known. The goal is to recover $x$ from the measurement. Without any assumptions on $x$, the system is underdertermined and has a subspace of dimension at least $n-m$ possible solutions. But when it is known that most of the coordinates of $x$ are zero (i.e., $x$ is sparse), it is possible to recover $x$ when $m \sim O(s \log n)$, where $s$ is the number of non-zero coordinates of $x$. This report will discuss the main results that allow recovery of sparse vectors from a linear measurement.

## 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 defined as the smallest $k$ for which $\mathbb{E}\exp(X^2/k^2) \leq 2$. 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.** Suppose $g \sim N(0, I_n)$, then 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 $\gamma(T) = \mathbb{E}\sup_{x\in T}|\langle x, g\rangle|$.

**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 $\{i : x_i \neq 0\}$ (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 Approximate Recovery based on $M^*$ bound

**Definition 2.1.** The class $\mathcal{A}(m, n)$ of random matrices is defined as the collection of all random matrices in $\mathbb{R}^{m\times n}$, whose rows $A_i$ are independent and isotropic sub-gaussian random vectors in $\mathbb{R}^n$. For $A \in \mathcal{A}(m, n)$, $K(A)$ (or simply $K$) is defined as $\max_i \|A_i\|_{\psi_2}$.

In this section, we will take $A$ to be sampled from the class $\mathcal{A}(m, n)$. We will discuss a method for obtaining approximate solution for the sparse recovery problem in this case.

In section 3, we will find conditions under which this solution is exact and later, in section 4, we will solve the same problem when $A$ lies in a deterministic class of matrices.

**Theorem 2.1** (Matrix deviation inequality: Liaw, Mehrabian, Plan, Vershynin. 2017)**.** *Suppose $A \in \mathcal{A}(m, n)$, then for $T \subseteq \mathbb{R}^n$,*

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

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[8].

**Theorem 2.2** ($M^*$ bound)**.** *Suppose $T \subseteq \mathbb{R}^n$ and $A \in \mathcal{A}(m, n)$, then for any $z \in \mathbb{R}^n$, $E = z + \ker(A)$ satisfies*

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

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

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

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 $\sqrt{m}(\mathrm{diam}(T \cap E))$. ∎

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 $\ell_2$ error as

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

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 $\ell_1$ norm. To that end, we observe by Cauchy-Schwarz inequality that the set of unit $s$-sparse vectors is contained in the $\ell_1$ ball scaled by the square root of sparsity. That is, if $B_1^n$ and $B_2^n$ are the unit balls with respect to the $\ell_1$ and $\ell_2$ norms respectively, then

$$S \cap B_2^n \subseteq \sqrt{s} B_1^n. \tag{6}$$

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

$$\mathbb{E} \sup_{\hat{x} \in B_1^n, A\hat{x}=y} \|x - \hat{x}\|_2 \leq CK^2 \sqrt{(s \log n)/m}. \tag{7}$$

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:

$$\begin{aligned} &\hat{x} = \arg\min \|x'\|_1 \\ &\text{subject to: } Ax' = y \end{aligned} \tag{8}$$

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.** *If $\hat{x}$ is the solution of basis pursuit (8), then*

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

*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 (7), 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 Recovery using Escape theorem

In the previous section, we observed that a solution with small expected $\ell_2$ error can be obtained using basis pursuit. In this section, we will find the conditions under which the error becomes zero.

**Theorem 3.1.** *In problem (1), suppose $A \in \mathcal{A}(m, n)$ 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 $A \in \mathcal{A}(m, n)$. 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, which states

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

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 $\ell_1$ 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 $1 - 2\exp(-cm/K^4)$. ∎

## 4 Recovery using RIP

Now we will consider a different class of matrices.

**Definition 4.1.** An $m \times n$ matrix $A$ satisfies Restricted Isometry Property (RIP) with parameters $\alpha, \beta$ and $s$ if for every $s$-sparse vector $w$,

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

The set of $m \times n$ matrices satisfying RIP with parameters $\alpha, \beta$ and $s$ will be denoted by $\mathrm{RIP}(m, n, \alpha, \beta, s)$.

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

**Theorem 4.1** (Candes–Tao)**.** *Suppose in (1), $A$ satisfies RIP with parameters $\alpha, \beta$ and $(1+\lambda)s$, with $\lambda > (\beta/\alpha)^2$. Then basis pursuit (8) exactly recovers the slolution to the sparse recovery problem.*

*Proof.* Let $h_I \in \mathbb{R}^{|I|}$ denote the projection of $h$ onto the subset of indices $I$. Let $I_0$ be the support of $x$, $I_1$ be the $\lambda s$ largest indices of $h_{I_0^c}$, $I_2$ be the next $\lambda s$ 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,

$$\|A_{I_{0,1}} h_{I_{0,1}}\|_2 \leq \|A_{I_{0,1}^c} h_{I_{0,1}^c}\|_2. \tag{9}$$

Now, by RIP,

$$\alpha \|h_{I_{0,1}}\|_2 \leq \|A_{I_{0,1}} h_{I_{0,1}}\|_2. \tag{10}$$

By triangle inequality and RIP,

$$\|A_{I_{0,1}^c} h_{I_{0,1}^c}\|_2 \leq \sum_{j \geq 2} \|A_{I_j} h_{I_j}\|_2 \leq \beta \sum_{j \geq 2} \|h_{I_j}\|_2. \tag{11}$$

Now observe that for any $j$, $\|h_{I_j}\|_2 \leq \sqrt{\lambda s} \frac{\|h_{I_{j-1}}\|_1}{\lambda s}$. Substituting this in (11) gives

$$\|A_{I_{0,1}^c} h_{I_{0,1}^c}\|_2 \leq (\beta/\sqrt{s\lambda}) \|h_{I_0^c}\|_1. \tag{12}$$

From the first part of lemma 3.1 and Cauchy-Schwarz inequality, $\|h_{I_0^c}\|_1 \leq \|h_{I_0}\|_1 \leq \sqrt{s}\|h_{I_0}\|_2$. Substituting this in (12), we get $\|A_{I_{0,1}^c} h_{I_{0,1}^c}\|_2 \leq \frac{\beta}{\sqrt{\lambda}} \|h_{I_0}\|_2$. From (9), (10) and the previous inequality, we see that $\alpha \|h_{I_{0,1}}\|_2 \leq \frac{\beta}{\sqrt{\lambda}} \|h_{I_0}\|_2$. But, from the condition on $\alpha, \beta, \lambda$, 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. Surprisingly, a sample from the collection $\mathcal{A}(m, n)$, that we saw in the previous section satisfies RIP with good parameters, with high probability. In order to prove such a result, we need bounds on the singular values. We will use the following bound on the singular values in the proof.

**Theorem 4.2.** *If $A \in \mathcal{A}(m, n)$, then for any $t \geq 0$,*

$$\sqrt{m} - CK^2(\sqrt{n} + t) \leq s_m(A_I) \leq s_1(A_I) \leq \sqrt{m} + CK^2(\sqrt{n} + t) \tag{13}$$

*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)*

The proof of uses the result that the operator norm $\|\frac{1}{m}A^T A - I_n\| \leq \max(\delta, \delta^2)$ implies all the singular values of $A/\sqrt{m}$ lie between $1 - \delta$ and $1 + \delta$.

An $\epsilon$ net $\mathcal{N}$ is used to bound the supremum over the unit sphere by a supremum over a finite set. After that, Bernstein's inequality is used to find tail bounds for $\langle (\frac{1}{m}A^T A - I_n)x, x\rangle = |\|\frac{Ax}{\sqrt{m}}\|_2^2 - 1| = \sum_i \frac{1}{m}(\langle A_i, x\rangle^2 - 1)$. Then a union bound over $\mathcal{N}$ is taken and finally, an appropriate value of $\delta$ is chosen to arrive at (13).

We can now use the bound (13) to generate random matrices that satisfies RIP with high probability.

**Theorem 4.3** (Candes–Tao, 2005)**.** *Suppose $A \in \mathcal{A}(m, n)$. If $m \geq CK^4 s \log n$, then $A \in RIP(m, n, 0.9, 1.1, s)$ with probability at least $1 - 2\exp(-cm/K^4)$.*

*Sketch of proof:* From remark 2, we need to bound the singular values for each $m \times s$ submatrix of $A$. Using theorem 4.2, bound for singular values of each possible $m \times s$ submatrix $A_I$ can be bounded. Since there are only $\binom{n}{s}$ such matrices, we can use a union bound and use $\binom{n}{s} \leq n^s$. ∎

*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 4.1 with $\lambda = 2 \geq (1.1/0.9)^2$. Thus exact recovery of $s$-sparse signals is possible by the conclusion of theorem 4.1.

## 4.1 Recovering signal from partial Fourier coefficients

In this section we will consider the case when the measurement matrix $A$ in (1) is a Fourier sub-matrix. A signal $f$ is a vector in $\mathbb{R}^n$ whose coordinates $f_n$ will be denoted by $f(n)$. We will also be dealing with signals in $\mathbb{C}^n$, which poses no difficulty since all the analysis done so far in this section for signals in $\mathbb{R}^n$ can be extended to complex signals.

**Definition 4.2.** The *discrete Fourier transform* of a signal $f \in \mathbb{R}^n$ is given by $\hat{f} = \mathcal{F}f$, where $\mathcal{F} \in \mathbb{R}^{n \times n}$ is the Fourier matrix, whose $(a, b)$th entry is $\mathcal{F}(a, b) = e^{(i2\pi ab)/n}/\sqrt{n}$.

The matrix $\mathcal{F}$ is a unitary matrix. The domain of $f$ is called *time* and the domain of $\hat{f}$ is called *frequency*. Since $f = \mathcal{F}^* \hat{f}$, we can write $f(n)$ as a linear combination of $\hat{f}(t)$ for $t = 1, \dots, n$, $\hat{f}(t)$s are also called the *Fourier coefficients* of $f$.

Suppose a subset of Fourier coefficients is known. The objective is to recover the signal $f$, given that $f$ is sparse. In order to obtain a sufficient condition for recovery to be possible we make the following definition.

**Definition 4.3.** Let $\Omega$ be a subset of frequency domain. Let $\hat{f}_\Omega = \mathcal{F}_\Omega f$, where $\mathcal{F}_\Omega$ is the $|\Omega| \times n$ sub matrix of $\mathcal{F}$ formed by the indices in $\Omega$. Then, $\Omega$ is said to have Restricted Isometry Propery with sparsity $s$ and error tolerance $\delta$, if for every $s$ sparse signal $f$,

$$(1 - \delta)\|f\|_2^2/N \leq \|\hat{f}_\Omega\|_2^2/|\Omega| \leq (1 + \delta)\|f\|_2^2/N.$$

On comparing this definition to the RIP property of matrices, it is clear that $\Omega$ satisfies RIP if and only if $\sqrt{N/|\Omega|}\mathcal{F}_\Omega \in \mathrm{RIP}(|\Omega|, n, \sqrt{1-\delta}, \sqrt{1+\delta})$. Thus, by theorem 4.1, the $s$-sparse signal $f$ can be recovered using basis pursuit if $\Omega$ satisfies RIP with say, sparsity $4s$ and error tolerance 0.25.

This result by itself is not very useful, since we do not know which subsets of frequency domain satisfies RIP. But the following theorem says any randomly chosen subset which is sufficiently large satisfies RIP.

**Theorem 4.4** (Candes–Tao, 2006). *When $f$ is $s$-sparse, a randomly chosen subset $\Omega$ of frequency domain of size $Cs\log^6 n$ will obey RIP with probability $1 - O(N^{-C})$.*

**Connection with Donoho-Stark uncertainty principle:** The reason why RIP guarantee exact recovery can be explained qualitatively by Donoho-Stark uncertainty principle. It states that for a signal $f \in \mathbb{R}^n$, the product of the size of supports, $\|f\|_0 \cdot \|\hat{f}\|_0 \geq n$. So, the sparser the signal, the larger is the support of its Fourier transform. The condition for RIP roughly states that the $\ell_2$ norm of $f$ averaged over all the coordinates is aproximately equal to the $\ell_2$ norm of $\hat{f}$ averaged over the coordinates indexed by $\Omega$. Since $\mathcal{F}$ is unitary, $\hat{f}$ and $f$ have the same $\ell_2$ norms. So, the total weight of $f$ is distributed to the coordinates of $\hat{f}$. If the support of $\hat{f}$ is large (i.e., when $f$ is sparse), even for smaller size of $\Omega$, the chances of missing the coordinates with the most weights are low. Hence RIP is satisfied even with smaller number of measurements.

# 5 Low rank matrix recovery

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

$$\text{recover } X \text{ from } y_i = \langle A_i, X\rangle,\ i = 1\ldots, m \text{ when } X \in T \quad (14)$$

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 $\langle A, B\rangle = \operatorname{tr} A^T B$ 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 (14) when $T = R = \{M \in \mathbb{R}^{d\times d} : \operatorname{rank}(M) \leq r\}$. 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 a $d \times d$ matrix A, if we take $v \in \mathbb{R}^d$ to be the $d$ dimensional vector of singular values, $\operatorname{rank}(A) = \|v\|_0$, $\|A\|_* = \|v\|_1$ 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

$$\begin{aligned} &\hat{X} = \arg\min \|X'\|_* \\ &\text{subject to: } \langle A_i, X'\rangle = y_i, \text{ for each } i = 1\ldots n. \end{aligned} \quad (15)$$

From the above remark and (6), 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 (7) in this case gives

$$\mathbb{E} \sup_{\substack{\hat{X}' \in \sqrt{r}B_1 \\ \langle A_i, X'\rangle = y_i}} \|X' - X\|_F \leq (CK^2\sqrt{r}w(B_*))/\sqrt{m}. \quad (16)$$

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

**Proposition 5.1.** *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, $\langle G, M\rangle = \operatorname{tr}(M^T G)$. If $M = A\Sigma B^T$ is the singular value decomposition of $M$, using the cyclic propery of trace, we get $\operatorname{tr}(M^T G) = \operatorname{tr}(\Sigma A^T GB)$. Since $\Sigma$ is a diagonel matrix of singular values, $\operatorname{tr}(\Sigma A^T GB) = \sum_{i=1}^d s_i(M)\langle a_i, Gb_i\rangle$, 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 $\operatorname{tr}(M^T G) \leq \|G\| \sum_i s_i(M) = \|G\|$. ■

To find the expected operator norm, we compare the two gaussian processes $\langle u, Gv\rangle$ and $\langle u, g\rangle + \langle v, h\rangle$, where $u, v \in S^{d-1}$ and $g, h$ independant standard normal random vectors and then apply Sudakov-Fernique inequality[8]. Using this and theorem 2.3, we have the following bound:

**Theorem 5.1.** *If $\hat{X}$ is the solution of problem 15, 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 (16) can be used to get the result. ■

# References


[1] Emmanuel Candes and Terence Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 2006. linked here.

[2] Manjunath Krishnapur. Gaussian processes - lecture notes. linked here.

[3] Manjunath Krishnapur. Topics in fourier analysis - lecture notes. linked here.

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

[5] Michel Talagrand. *Upper and Lower Bounds for Stochastic Processes*, chapter 1. Springer, second edition edition, 2014. [available online].

[6] Terence Tao. Compressed sensing - robust recovery of sparse signals from limited measurements. linked here, February 2008.

[7] Terence Tao. Uniform uncertainity principle and compressed sensing. linked here, December 2008.

[8] 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].