Compressed Sensing

Lecture 11

Background reading

These notes are based on Chapters 2,4,6,9 of Foucart, Simon, and Holger Rauhut. *A mathematical introduction to compressive sensing.* Vol. 1. No. 3. Boston: Birkhäuser, 2013.

1 Introduction

Consider the following inverse problem:

> Given $A \in \mathbb{C}^{m\times N}$ and $y \in \mathbb{C}^m$ with $N \gg m$, find $x \in \mathbb{C}^N$ such that $Ax = y$.

In general, this is impossible. However, signals we are typically interested in are not arbitrary, but often 'sparse' or approximately sparse with respect to some representation system (e.g. a wavelet basis).

We begin with a couple of definitions. Note that the second definition is simply the finite dimensional analogue of the nonlinear approximation error discussed in the previous section.

Definition 1. *For $x \in \mathbb{C}^N$, let $\mathrm{Supp}(x) = \{j : x_j \neq 0\}$. A vector $x \in \mathbb{C}^N$ is s-sparse if at most s of its entries are nonzero:*

$$\|x\|_0 := \mathrm{card}(\mathrm{Supp}(x)) \leqslant s.$$

Let Σ_s denote the set of all s-sparse vectors.

Definition 2. *For $p > 0$, the ℓ_p-error of the best s-term approximation to $x \in \mathbb{C}^N$ is*

$$\sigma_s(x)_p := \inf\left\{\|x - z\|_p \,;\, z \in \Sigma_s\right\}.$$

2 Minimal number of measurements

> What conditions should we impose on a measurement matrix $A \in \mathbb{C}^{m\times N}$ so that we can recover every s sparse vector from Ax?

Throughout, given a matrix $A \in \mathbb{C}^{m\times N}$ and an index set $S \subset [N]$, let A_S denote the matrix A with its columns restricted to those indexed by S. Similarly, given a vector $v \in \mathbb{C}^N$, let v_S denote the restriction of the vector v to its coefficients indexed by S.

Theorem 1. *Given $A \in \mathbb{C}^{m\times N}$, the following are equivalent:*

(i) If $Ax = Az$ and $x, z \in \Sigma_s$, then $x = z$.

(ii) The null space $\mathcal{N}(A)$ does not contain any $2s$ sparse vector other than 0, that is $\mathcal{N}(A) \cap \Sigma_{2s} = \{0\}$.

(iii) For all $S \subset [N]$ with $\mathrm{Card}(S) \leqslant 2s$, A_S is injective from $\mathbb{C}^S$ to $\mathbb{C}^N$.

(iv) Every set of $2s$ columns of A is linearly independent.

Proof. It is clear that (iii) and (iv) are equivalent.

To see that (ii) is equivalent to (iii), note that (ii) is true if and only if for all $v \in \Sigma_{2s}$, $Av = A_S v_S = 0$ implies that $v = 0$. This is true if and only if A_S is injective for every subset $S \subset [N]$ of cardinality $2s$.

To see that (ii) implies (i), let $x, z \in \Sigma_s$ be such that $Ax = Az$. Then, $A(x - z) = 0$ and since $x - z \in \Sigma_{2s}$, we must have $x = z$ if we assume (ii).

Finally, to see that (i) implies (ii), let $v \in \mathcal{N}(A)$ be such that $v \in \Sigma_{2s}$. Then, we can write $v = z + x$ for some $z, x \in \Sigma_s$ with $\mathrm{Supp}(x) \cap \mathrm{Supp}(z) = \emptyset$. By (i), since $Ax = -Az$, we have that $x = -z$ and since they have disjoint support, it follows that $x = -z = 0$.

□

Remark 1. From the previous theorem, we see that if we have a measurement matrix $A \in \mathbb{C}^{m\times N}$ for which it is possible to reconstruct every s-sparse vector x from measurements Ax, then (iv) holds, so $\mathrm{rank}(A) \geqslant 2s$. However, since $\mathrm{rank}(A) \leqslant m$, we must have that $m \geqslant 2s$.

If any one of the conditions in the previous lemma hold, then $x \in \Sigma_s$ is the unique solution to

$$\min_{z\in\mathbb{C}^N} \|z\|_0 \text{ subject to } Az = Ax. \tag{1}$$

Theorem 2. *For any integer $N \geqslant 2s$, there exists a measurement matrix $A \in \mathbb{C}^{m\times N}$ with $m = 2s$ such that every s-sparse vector $x \in \mathbb{C}^N$ can be recovered from $y = Ax \in \mathbb{C}^m$ as a solution of* (1).

Proof. Let us fix the following real numbers,

$$t_N > \cdots > t_2 > t_1 > 0.$$

Consider the following matrix $A \in \mathbb{C}^{m\times N}$ with $m = 2s$:

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_N \\ \vdots & & & \vdots \\ t_1^{2s-1} & t_2^{2s-1} & \cdots & t_N^{2s-1} \end{pmatrix}.$$

Given any index set $S \subset [N]$ of cardinality $2s$ with $S = \{j_1 < j_2 < \cdots < j_{2s}\}$,

$$A_S = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_{j_1} & t_{j_2} & \cdots & t_{j_{2s}} \\ \vdots & & & \vdots \\ t_{j_1}^{2s-1} & t_{j_2}^{2s-1} & \cdots & t_{j_{2s}}^{2s-1} \end{pmatrix} \in \mathbb{C}^{2s\times 2s}$$

is a Vandermonde matrix with $\det(A_S) = \prod_{k<l}(t_{j_l} - t_{j_k}) > 0$. Therefore, A_S is invertible and by Theorem 1, the conclusion follows. □

Bad news: In general, solving (1) is not feasible: since the minimizer has sparsity s, a naive approach would be to solve $AP_S u = y$ for all subsets $S \subset \{1, \ldots, N\}$ of size s, this is $\binom{N}{s}$ linear systems. If $N = 1000$ and $s = 10$ and each linear system took 10^{-10} seconds, this approach would take over 300 years. In fact, for general A and y, one can prove that (1) is NP-hard.

Convex relaxation Since $\|z\|_p^p \to \|z\|_0$ as $p \to 0$, it is natural to consider the approximation of (1) by

$$\min_{z\in\mathbb{C}^N} \|z\|_p \text{ subject to } Az = y. \tag{2}$$

For $p \in (0, 1)$, this is again NP-hard in general, for $p > 1$, even 1-sparse vectors cannot be recovered as solutions of (2). However, we will see that sparse recovery can be guaranteed when $p = 1$.

3 Basis Pursuit

In this section, we study the solutions to the following minimization problem (basis pursuit):

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \text{ subject to } Az = y. \tag{3}$$

What are the necessary and sufficient conditions on $A \in \mathbb{C}^{m\times N}$ for which every s sparse vector $x \in \mathbb{C}^N$ can be recovered from $y = Ax$ as the unique solution to (3)? When can we guarantee reconstructions which are stable to sparsity defect or inexact measurements?

Definition 3. *A matrix $A \in \mathbb{C}^{m\times N}$ is said to satisfy the null space property (NSP) relative to $S \subset [N]$ if*

$$\|v_S\|_1 < \|v_{S^c}\|_1, \qquad \forall v \in \mathcal{N}(A) \setminus \{0\}.$$

It is said to satisfy the NSP of order s if this is true for all $S \subset [N]$ with $\mathrm{Card}(S) \leqslant s$.

Theorem 3. *Given $A \in \mathbb{C}^{m\times N}$, every $x \in \mathbb{C}^N$ supported on $S \subset [N]$ is the unique solution to* (3) *with $y = Ax$ if and only if A satisfies the NSP relative to S.*

Proof. For fixed $S \subset [N]$, assume that every vector $x \in \mathbb{C}^N$ supported on S is the unique solution to (3). Then, given any $v \in \mathcal{N}(A) \setminus \{0\}$, v_S is the unique solution to (3) with $y = Av_S$. But, $Av_S = -Av_{S^c}$ and $v_S \neq -v_{S^c}$. This implies that $\|v_S\|_1 < \|v_{S^c}\|_1$.

Conversely, suppose that A satisfies the NSP relative to S. Then, given x with $\mathrm{Supp}(x) \subset S$ and $z \in \mathbb{C}^N$ such that $Ax = Az$ and $z \neq x$, we have that $v := x - z \in \mathcal{N}(A) \setminus \{0\}$. So,

$$\|x\|_1 = \|x - z_S\|_1 + \|z_S\|_1 = \|v_S\|_1 + \|z_S\|_1 < \|v_{S^c}\|_1 + \|z_S\|_1 = \|z_{S^c}\|_1 + \|z_S\|_1 = \|z\|_1.$$

Therefore, x is the unique minimizer to (3). ☐

By applying the above theorem to all subsets of cardinality s, we have the following result.

Corollary 1. *Given $A \subset \mathbb{C}^{m\times N}$, every s-sparse vector $x \in \mathbb{C}^N$ is the unique solution to* (3) *with $y = Ax$ if and only if A satisfies the NSP of order s.*

3.1 Stability to sparsity defect and inexact measurements

Suppose we want to recover x from a measurement vector $y \in \mathbb{C}^m$ such that

$$\|Ax - y\|_2 \leqslant \eta.$$

Then, we will solve instead

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \text{ subject to } \|Az - y\|_2 \leqslant \eta. \tag{4}$$

In the case where x is not perfectly sparse and we are given inexact measurements, it is desirable to recover x in a stable manner, such that the reconstruction error can be controlled by the amount of sparsity defect $\sigma_s(x)_1$, and the noise level η.

Definition 4. *$A \in \mathbb{C}^{m\times N}$ is said to satisfy the robust NSP relative to $S \subset [N]$, with $\rho \in (0,1)$ and $\tau > 0$ if*

$$\|v_S\|_1 \leqslant \rho\|v_{S^c}\|_1 + \tau\|Av\|_2, \qquad \forall v \in \mathbb{C}^N.$$

It is said to satisfy the robust NSP of order s with $\rho \in (0,1)$ and $\tau > 0$ if the above inequality holds for all subsets S of cardinality at most s.

Theorem 4. *Suppose that $A \in \mathbb{C}^{m\times N}$ satisfies the robust NSP of order s with $\rho \in (0,1)$ and $\tau > 0$. Then, for all $x \in \mathbb{C}^N$, given $y = Ax + e$ and $\|e\| \leqslant \eta$, any solution $\hat{x}$ to (4) satisfies*

$$\|x - \hat{x}\|_1 \leqslant \frac{2(1+\rho)}{(1-\rho)}\sigma_s(x)_1 + \frac{4\tau}{1-\rho}\eta.$$

To prove this theorem, we will prove the following (stronger) result.

Theorem 5. *$A \in \mathbb{C}^{m\times N}$ satisfies the robust NSP with $\rho \in (0,1)$ and $\tau > 0$ relative to S if and only if*

$$\|z - x\|_1 \leqslant \frac{1+\rho}{1-\rho}\left(\|z\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1\right) + \frac{2\tau}{1-\rho}\|A(x-z)\|_2. \tag{5}$$

Proof. Assume that A satisfies (5) for all $x, z \in \mathbb{C}^N$. Let $v \in \mathbb{C}^N$. Then, by writing $x = -v_S$ and $z = v_{S^c}$,

$$\|v\|_1 \leqslant \frac{1+\rho}{1-\rho}(\|v_{S^c}\|_1 - \|v_S\|_1) + \frac{2\tau}{1-\rho}\|Av\|_2.$$

By rearranging the above equation, we have that

$$\|v_S\|_1 \leqslant \rho\|v_{S^c}\|_1 + \tau\|Av\|_2.$$

Conversely, assume that A satisfies the robust NSP relative to S. Then, for $x, z \in \mathbb{C}^N$, let $v := z - x$. By the robust NSP,

$$\|v_S\|_1 \leqslant \rho\|v_{S^c}\|_1 + \tau\|Av\|, \tag{6}$$

and by Lemma 1,

$$\|v_{S^c}\|_1 \leqslant \|z\|_1 - \|x\|_1 + \|v_S\|_1 + 2\|x_{S^c}\|_1. \tag{7}$$

By plugging (6) into (7), we have that

$$\|v_{S^c}\|_1 \leqslant \rho\|v_{S^c}\|_1 + \tau\|Av\| + \|z\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1$$

and rearranging yields

$$\|v_{S^c}\|_1 \leqslant \frac{1}{1-\rho}\left(\tau\|Av\| + \|z\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1\right). \tag{8}$$

So, by applying the robust NSP, we have that

$$\|v\|_1 = \|v_{S^c}\|_1 + \|v_S\|_1 \leqslant (1+\rho)\|v_{S^c}\|_1 + \tau\|Av\|_2,$$

and a further application of (8) yields the desired result. □

Lemma 1. *Given $S \subset [N]$, and $x, z \in \mathbb{C}^N$,*

$$\|(x-z)_{S^c}\|_1 \leqslant \|z\|_1 - \|x\|_1 + \|(x-z)_S\|_1 + 2\|x_{S^c}\|_1.$$

Proof. Observe that

$$\|x\|_1 = \|x_{S^c}\|_1 + \|x_S\|_1 \leqslant \|x_{S^c}\|_1 + \|(x-z)_S\|_1 + \|z_S\|_1$$

and

$$\|(x-z)_{S^c}\|_1 \leqslant \|x_{S^c}\|_1 + \|z_{S^c}\|_1.$$

By summing the two inequalities,

$$\|x\|_1 + \|(x-z)_{S^c}\|_1 \leqslant 2\|x_{S^c}\|_1 + \|z\|_1 + \|(x-z)_S\|_1.$$

□

Lecture 12

Deriving ℓ^2 error bounds

Definition 5. *For $q \geqslant 1$, $A \in \mathbb{C}^{m\times N}$ is said to satisfy the ℓ^q-robust NSP of order s with $\rho \in (0,1)$ and $\tau > 0$ if for all $S \subset [N]$ with $\mathrm{Card}(S) \leqslant s$,*

$$\|v_S\|_q \leqslant \frac{\rho}{s^{1-1/q}}\|v_{S^c}\|_1 + \tau\|Av\|_2, \qquad \forall v \in \mathbb{C}^N.$$

Remark 2. For $1 \leqslant p \leqslant q$, observe that by Hölder's inequality,

$$\sum_{j\in S}|v_j|^p \leqslant \left(\sum_{j\in S}|v_j|^q\right)^{p/q} s^{(q-p)/q} = \|v_S\|_q^p s^{1-p/q}.$$

Therefore, $\|v_S\|_p \leqslant s^{1/p-1/q}\|v_S\|_q$. So, if the ℓ^q-robust NSP holds, then

$$\|v_S\|_p \leqslant \frac{\rho}{s^{1-1/p}}\|v_{S^c}\|_1 + s^{1/p-1/q}\tau\|Av\|_2, \qquad \forall v \in \mathbb{C}^N.$$

Theorem 6. *Suppose that $A \in \mathbb{C}^{m\times N}$ satisfies the ℓ^2 robust NSP of order s with $\rho \in (0,1)$ and $\tau > 0$. Then, for all $x \in \mathbb{C}^N$, any solution $\hat{x}$ of (4) with $y = Ax + e$ and $\|e\|_2 \leqslant \eta$ approximates x with ℓ^p error:*

$$\|x - \hat{x}\|_p \leqslant \frac{C}{s^{1-1/p}}\sigma_s(x)_1 + Ds^{1/p-1/2}\eta, \qquad p \in [1,2].$$

Here, C and D are constants which depend only on ρ and τ.

This theorem follows from the stronger result Theorem 7, with $q = 2$ and $z = \hat{x}$. However, before proving that theorem, we first derive a lemma. In saying that a vector is 'compressible', we generally mean that its s-term approximation error decays quickly in s. The following lemma essentially shows that elements belonging to the nonconvex unit ℓ_p balls with $p < 1$ serve as good models for compressible vectors. Moreover, this result hints that the error bound obtained in Theorem 6 is natural: if $p \in [1,2]$ and $\|x\|_q \leqslant 1$ for $q < 1$, then $\sigma_s(x)_p \leqslant s^{1/p-1/q}$. Now, assuming that $\eta = 0$ (no measurements error), the result of Theorem 6 says that

$$\|x - \hat{x}\|_p \leqslant Cs^{1/p-1/q}.$$

So, the error has the same decay in s as the s-term approximiation error in ℓ^p.

Lemma 2. *For any $p > q > 0$, and any $x \in \mathbb{C}^N$, the inequality*

$$\sigma_s(x)_p \leqslant \frac{1}{s^{1/q-1/p}}\|x\|_q.$$

Proof. Let x^* be a rearrangement of $(|x_j|)_{j=1}^N$ in nonincreasing order. Then,

$$\sigma_s(x)_p^p = \sum_{j=s+1}^N (x_j^*)^p \leqslant (x_s^*)^{p-q}\sum_{j=s+1}^N (x_j^*)^q \leqslant \left(s^{-1}\sum_{j=1}^s (x_s^*)^q\right)^{\frac{p-q}{q}} \|x\|_q^q \leqslant s^{-\frac{p-q}{q}}\|x\|_q^{p-q}\|x\|_q^q.$$

Therefore, $\sigma(x)_p \leqslant s^{-1/q+1/p}\|x\|_q$. □

Theorem 7. *Given $1 \leqslant p \leqslant q$, suppose that $A \in \mathbb{C}^{m\times N}$ satisfies the ℓ^q robust NSP of order s with $\rho \in (0,1)$ and $\tau > 0$. Then, for all $x, z \in \mathbb{C}^N$,*

$$\|x - z\|_p \leqslant \frac{C}{s^{1-1/p}}\left(\|z\|_1 - \|x\|_1 + 2\sigma_s(x)_1\right) + Ds^{1/p-1/q}\|A(x-z)\|_2,$$

where

$$C = \frac{(1+\rho)^2}{1-\rho} \quad \text{and} \quad D = \frac{(3+\rho)\tau}{1-\rho}.$$

Proof. The ℓ^q robust NSP (along with Hölder's inequality) implies that for all $v \in \mathbb{C}^N$ and all $S \subset [N]$ with $\mathrm{Card}(S) \leqslant s$,

$$\|v_S\|_1 \leqslant s^{1-1/q}\|v_S\|_q \leqslant \rho\|v_{S^c}\|_1 + s^{1-1/q}\tau\|Av\|_2, \tag{9}$$

and

$$\|v_S\|_p \leqslant s^{1/p-1/q}\|v_S\|_q \leqslant \frac{\rho}{s^{1-1/p}}\|v_{S^c}\|_1 + s^{1/p-1/q}\tau\|Av\|_2. \tag{10}$$

So, from (9), we may apply Theorem 5 to obtain

$$\|z-x\|_1 \leqslant \frac{1+\rho}{1-\rho}\left(\|z\|_1 - \|x\|_1 + 2\sigma_s(x)_1\right) + \frac{2\tau s^{1-1/q}}{1-\rho}\|A(x-z)\|_2. \tag{11}$$

Now, by choosing S to be the largest s entries of $z-x$, by Lemma 2,

$$\|z-x\|_p \overbrace{\leqslant}^{\Delta\text{-inequality}} \|(x-z)_{S^c}\|_p + \|(x-z)_S\|_p \leqslant \frac{1}{s^{1-1/p}}\|x-z\|_1 + \|(x-z)_S\|_p.$$

Therefore, by combining this inequality with (10), we have that

$$\|z-x\|_p \leqslant \frac{1}{s^{1-1/p}}\|x-z\|_1 + \frac{\rho}{s^{1-1/p}}\|v_{S^c}\|_1 + s^{1/p-1/q}\tau\|Av\|_2 \leqslant \frac{1+\rho}{s^{1-1/p}}\|x-z\|_1 + s^{1/p-1/q}\tau\|Av\|_2.$$

Substituting in (11) yields the desired result. □

4 Restricted isometry property

In this section, we introduce the notion of the restricted isometry property (RIP) and show that it is a sufficient condition for stable and robust recovery. It is often easier to work with the RIP than the null space property.

4.1 Definition and basic properties

Definition 6. *The sth restricted isometry constant $\delta_s = \delta_s(A)$ of a matrix $A \in \mathbb{C}^{m\times N}$ is the smallest $\delta > 0$ such that*

$$(1-\delta)\|x\|_2^2 \leqslant \|Ax\|_2^2 \leqslant (1+\delta)\|x\|_2^2, \qquad \forall x \in \Sigma_s$$

Equivalently,

$$\delta_s = \max_{S\subset[N],\mathrm{Card}\,S\leqslant s} \|A_S^*A_S - I_S\|_{2\to 2}.$$

Note that $\delta_1 \leqslant \delta_2 \leqslant \cdots \leqslant \delta_N$.

Proposition 1. *Let $u, v \in \mathbb{C}^N$ and suppose that $\|u\|_0 \leqslant s$ and $\|v\|_0 \leqslant t$. If $\mathrm{Supp}(u) \cap \mathrm{Supp}(v) = \emptyset$, then*

$$|\langle Au,\, Av\rangle| \leqslant \delta_{s+t}\|u\|_2\|v\|_2.$$

Proof. Let $S := \mathrm{Supp}(u) \cup \mathrm{Supp}(v)$. Let $u_S, v_S \in \mathbb{C}^S$ be the restriction of u and v to S. Note that since u and v have disjoint support, $\langle u_S,\, v_S\rangle = \langle u,\, v\rangle = 0$. Hence,

$$|\langle Au,\, Av\rangle| = |\langle A_Su_S,\, A_Sv_S\rangle - \langle u_S,\, v_S\rangle| = |\langle (A_S^*A_S - I)u_S,\, v_S\rangle| \leqslant \|A_S^*A_S - I\|_{2\to 2}\|u_S\|_2\|v_S\|_2.$$

The result follows since $\|A_S^*A_S - I\|_{2\to 2} \leqslant \delta_{s+t}$.

□

In the following, we show that the RIP is necessary for stability to noisy measurements.

Definition 7. *Let $A \in \mathbb{C}^{m\times N}$ and let $\Delta : \mathbb{R}^m \to \mathbb{R}^N$ denote a recovery algorithm. We say that (A, Δ) is C-stable of order s if for all $x \in \Sigma_s$ and any $e \in \mathbb{C}^m$,*

$$\|\Delta(Ax + e) - x\|_2 \leqslant C\|e\|_2.$$

Theorem 8. *If (A, Δ) is C-stable of order s, then $\frac{1}{C}\|x\|_2 \leqslant \|Ax\|_2$ for all $x \in \Sigma_{2s}$.*

Proof. Let $x, z \in \Sigma_s$ and define

$$e_x = \frac{A(z-x)}{2} \quad \text{and} \quad e_z = \frac{A(x-z)}{2}.$$

Observe that

$$Ax + e_x = \frac{A(x+z)}{2} = Az + e_z.$$

Let $\hat{x} = \Delta(Ax + e_x) = \Delta(Az + e_z)$. Then,

$$\begin{aligned} \|x - z\|_2 &= \|x - \hat{x} + \hat{x} - z\|_2 \leqslant \|x - \hat{x}\|_2 + \|\hat{x} - z\|_2 \\ &\leqslant C\|e_x\|_2 + C\|e_z\|_2 = C\left(\frac{\|A(z-x)\|_2}{2} + \frac{\|A(x-z)\|_2}{2}\right) = C\|A(x-z)\|_2. \end{aligned}$$

Since this is true for all $x, z \in \Sigma_s$, the conclusion follows. □

Minimal number of measurements

Theorem 9. *Let $A \in \mathbb{R}^{m\times N}$ such that it has RIP constant $\delta_{2s} \in (0, 1/2]$. Then,*

$$m \geqslant Cs\log\left(\frac{N}{s}\right), \qquad C = \frac{1}{2\log(\sqrt{24}+1)}.$$

To prove this theorem, we first require a preliminary lemma. [1]

Lemma 3. *Let s and N satisfy $s \leqslant N/2$. Then, there exists a set $X \subset \Sigma_s$ such that*

(i) *for any $x \in X$, we have $\|x\|_2 \leqslant \sqrt{s}$.*

(ii) *for any $x, z \in X$ with $x \neq z$, we have $\|x - z\|_2 \geqslant \sqrt{s/2}$.*

(iii) $\log|X| \geqslant \frac{s}{2}\log(\frac{N}{s})$.

This lemma gives a lower bound, in dimension N, on the number of balls of radius $\sqrt{s/2}$, centred at s-sparse vectors that we can pack into the ball of radius $\sqrt{s}$. Since the RIP means that A roughly preserves the distance between s-sparse vectors, we can make a similar statement about balls of dimension m. This allows us to obtain a lower bound on m.

Proof of Theorem 9. Let X be as in Lemma 3. Given any $x, z \in X$, we have that $x, z, x-z \in \Sigma_{2s}$. Therefore, by applying (ii) of Lemma 3 and using the fact that $\delta_{2s} \in (0, 1/2]$,

$$\|Ax - Az\|_2 \geqslant \sqrt{1-\delta_{2s}}\|x - z\|_2 \geqslant \sqrt{1-\delta_{2s}}\sqrt{\frac{s}{2}} \geqslant \sqrt{\frac{s}{4}}.$$

Since $\|x\|_2 \leqslant \sqrt{s}$ and $\delta_{2s} \in (0, 1/2]$,

$$\|Ax\|_2 \leqslant \sqrt{1+\delta_{2s}}\|x\|_2 \leqslant \sqrt{\frac{3s}{2}}. \tag{12}$$

[1] We will use this lemma without proof, the interested reader can refer to Lemma A.1 in *Compressed sensing: theory and applications. Cambridge University Press, Y. Eldar and G. Kutyniok, eds. (2012).*

Let B_x and B_z be balls of radius $\frac{\sqrt{s/4}}{2}$, centred at Ax and Az respectively, then they are disjoint.
By (12), $B_x \cup B_z \subset B_r$ where $r \leqslant \sqrt{3s/2} + \sqrt{s/16}$. Now,

$$\begin{aligned} & \mathrm{Vol}\left[B\left(\sqrt{\frac{3s}{2}}+\sqrt{\frac{s}{16}}\right)\right] \geqslant |X| \,\mathrm{Vol}\left[B\left(\sqrt{\frac{s}{16}}\right)\right] \\ \iff \quad & \left(\sqrt{\frac{3s}{2}}+\sqrt{\frac{s}{16}}\right)^m \geqslant |X| \left(\sqrt{\frac{s}{16}}\right)^m \\ \iff \quad & (\sqrt{24}+1)^m \geqslant |X| \\ \iff \quad & m \geqslant \frac{\log|X|}{\log(\sqrt{24}+1)}. \end{aligned}$$

Finally, the conclusion follows by applying (iii) of Lemma 3. □

Lecture 13

4.2 Analysis of basis pursuit with the RIP

Theorem 10 (RIP implies sparse recovery)**.** *Suppose that the* $2s$ *restricted isometry constant of* $A \in \mathbb{C}^{m\times N}$ *satisfies* $\delta_{2s} < 1/3$. *Then, every* s*-sparse vector* $x \in \mathbb{C}^N$ *is the unique solution of*

$$\min_{z\in\mathbb{C}^N} \|z\|_1 \text{ subject to } Ax = Az.$$

To prove this theorem, we first require a technical lemma.
Lemma 4. *Given* $q, p > 0$, *if* $u \in \mathbb{C}^s$ *and* $v \in \mathbb{C}^t$ *are such that*

$$\max_{i\in[s]} |u_i| \leqslant \min_{j\in[t]} |v_j|,$$

then $\|u\|_q \leqslant \frac{s^{1/q}}{t^{1/p}}\|v\|_p$. *In the case* $p = 1$, $q = 2$ *and* $t = s$, *we have* $\|u\|_2 \leqslant s^{-1/2}\|v\|_1$.

Proof. The result follows by combining the following two inequalities:

- $s^{-1/q}\|u\|_q = \left(s^{-1}\sum_{i=1}^s |u_i|^q\right)^{1/q} \leqslant \max_{i\in[s]} |u_i|$.
- $t^{-1/p}\|v\|_p = \left(t^{-1}\sum_{i=1}^t |v_i|^p\right)^{1/p} \geqslant \min_{i\in[t]} |v_i|$.

□

Proof of Theorem 10. By Corollary 1, it is enough to show A satisfies the NSP of order s

$$\|v_S\|_1 < \frac{1}{2}\|v\|_1, \qquad \forall v \in \mathcal{N}(A)\setminus\{0\}, \quad S \subset [N], \quad \mathrm{Card}(S) = s.$$

Since $\|v_S\|_1 \leqslant \sqrt{s}\|v_S\|_2$ by the Cauchy-Schwarz inequality, it is enough to show that

$$\|v_S\|_2 \leqslant \frac{\rho}{2\sqrt{s}}\|v\|_1, \qquad \forall v \in \mathcal{N}(A)\setminus\{0\}, \quad S \subset [N], \quad \mathrm{Card}(S) = s,$$

where $\rho := 2\delta_{2s}(1-\delta_{2s})^{-1} < 1$.

For $v \in \mathcal{N}(A)$, it is enough to consider the case where $S = S_0$ is the index set corresponding to the largest s entries of v. Partition $S_0^c = S_1 \cup S_2 \cup \cdots$ into sets of cardinality at most s by

$$\begin{aligned} S_1 &= \text{ index set of largest } s \text{ entries in } S_0^c, \\ S_2 &= \text{ index set of largest } s \text{ entries in } S_0^c \setminus S_1, \end{aligned}$$

and so on. Then, for $v \in \mathcal{N}(A)$, $Av_{S_0} = A(-v_{S_1} - v_{S_2} - \cdots)$. By the RIP,

$$\begin{aligned} \|v_{S_0}\|_2^2 &\leqslant \frac{1}{1-\delta_{2s}} \|Av_{S_0}\|_2^2 = \frac{1}{1-\delta_{2s}} \langle Av_{S_0},\, A(-v_{S_1} - v_{S_2} - \cdots) \rangle \\ &= \frac{1}{1-\delta_{2s}} \sum_{k \geqslant 1} \langle Av_{S_0},\, A(-v_{S_k}) \rangle \\ \text{by Proposition 1} \qquad &\leqslant \frac{\delta_{2s}}{1-\delta_{2s}} \sum_{k \geqslant 1} \|v_{S_0}\|_2 \|v_{S_k}\|_2. \end{aligned}$$

Therefore,

$$\|v_{S_0}\|_2 \leqslant \frac{\delta_{2s}}{1-\delta_{2s}} \sum_{k \geqslant 1} \|v_{S_k}\|_2 = \frac{\rho}{2} \sum_{k \geqslant 1} \|v_{S_k}\|_2.$$

By Lemma 4,

$$\|v_{S_k}\|_2 \leqslant \frac{1}{\sqrt{s}} \|v_{S_{k-1}}\|_1.$$

So,

$$\|v_{S_0}\|_2 \leqslant \frac{\rho}{2\sqrt{s}} \sum_{k \geqslant 1} \|v_{S_k}\|_1 \leqslant \frac{\rho}{2\sqrt{s}} \|v\|_1.$$

□

Theorem 11 (RIP implies stability and robustness)**.** *Suppose that the* $2s$ *restricted isometry constant of* $A \in \mathbb{C}^{m \times N}$ *satisfies*

$$\delta_{2s} < \frac{4}{\sqrt{41}} \approx 0.6246, \tag{13}$$

then for all $x \in \mathbb{C}^N$ *and* $y \in \mathbb{C}^m$ *with* $\|Ax - y\|_2 \leqslant \eta$, *any solution* $\hat{x}$ *to* (4) *satisfies*

$$\begin{aligned} \|x - \hat{x}\|_1 &\leqslant C\sigma_s(x)_1 + D\sqrt{s}\eta \\ \|x - \hat{x}\|_2 &\leqslant \frac{C}{\sqrt{s}} \sigma_s(x)_1 + D\eta, \end{aligned}$$

where $C, D > 0$ *are constants which depend only on* δ_{2s}.

By Theorem 6, Theorem 11 is a direct consequence of the following result.

Theorem 12. *If the* $2s$ *restricted isometry constant* $A \in C^{m \times N}$ *satisfies* (13), *then the matrix* A *satisfies the* ℓ^2 *robust NSP of order* s *with constants* $\rho \in (0,1)$ *and* $\tau > 0$ *which depend only on* δ_{2s}.

We first prove a technical lemma, which can be viewed as a counter part to the Cauchy-Schwarz inequality $\|a\|_1 \leqslant \sqrt{s}\|a\|_2$.

Lemma 5 (Square root lifting)**.** *For* $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_s \geqslant 0$,

$$\sqrt{\sum_{j=1}^{s} a_j^2} \leqslant \frac{\sum_{j=1}^{s} a_j}{\sqrt{s}} + \frac{\sqrt{s}}{4}(a_1 - a_s).$$

Proof. Proving this lemma is equivalent to showing that given any $a \in \mathbb{R}^s$,

$$a_1 \geqslant a_2 \geqslant \cdots \geqslant a_s \geqslant 0 \quad \text{and} \quad \frac{\sqrt{s}}{4}a_1 + \frac{\sum_{j=1}^s a_j}{\sqrt{s}} \leqslant 1$$

implies that $\|a\|_2 + \frac{\sqrt{s}}{4}a_s \leqslant 1$. Let $f(a) = \|a\|_2 + \frac{\sqrt{s}}{4}a_s$ and let

$$C = \left\{ a \in \mathbb{R}^s \; ; \; a_1 \geqslant a_2 \geqslant \cdots \geqslant a_s \geqslant 0 \quad \text{and} \quad \frac{\sqrt{s}}{4}a_1 + \frac{\sum_{j=1}^s a_j}{\sqrt{s}} \leqslant 1 \right\}.$$

Then, we simply need to show that $\sup_{a\in C} f(a) \leqslant 1$. Observe that C is a convex set and f is a convex function. So, the supremum is achieved on one of the vertices of C. Moreover, the vertices are exactly the intersection points of the hyperplanes obtained by setting s of the $s+1$ inequality constraints to equalities. We are thus left to consider the following three cases:

- $a_1 = \cdots = a_s = 0$. Then, $f(a) = 0$.
- $a_1 = \cdots = a_k > a_{k+1} = \cdots = a_s = 0$ with $k < s$ and $\frac{\sqrt{s}}{4}a_1 + \frac{\sum_{j=1}^s a_j}{\sqrt{s}} = 1$. In this case, $a_1 = \sqrt{s}/(k+s/4)$ and $f(a) = \sqrt{sk}/(k+s/4) \leqslant 1$.
- $a_1 = \cdots = a_s > 0$ and $\frac{\sqrt{s}}{4}a_1 + \frac{\sum_j a_j}{\sqrt{s}} = 1$. In this case, $a_1 = 4/(5\sqrt{s})$ and $f(a) = \frac{4}{5} + \frac{1}{5} = 1$.

□

Proof of Theorem 12. We need to find $\rho \in (0,1)$ and $\tau > 0$ such that for all $v \in \mathbb{C}^N$, $S \subset [N]$ with $\mathrm{Card}(S) = s$,

$$\|v_S\|_2 \leqslant \frac{\rho}{\sqrt{s}}\|v_{S^c}\|_1 + \tau\|Av\|_2.$$

Given $v \in \mathbb{C}^N$, it is enough to consider the index set $S := S_0$ of the s largest absolute entries of v. As before, partition $S_0^c = S_1 \cup S_2 \cup \cdots$, so that

$$\begin{aligned} S_1 &= \text{ index set of largest } s \text{ entries in } S_0^c, \\ S_2 &= \text{ index set of largest } s \text{ entries in } S_0^c \setminus S_1, \end{aligned}$$

and so on. Since v_{S_0} is s-sparse, we can write

$$\|Av_{S_0}\|_2^2 = (1+t)\|v_{S_0}\|_2^2$$

for some t such that $|t| \leqslant \delta_s$. We aim to show that

$$|\langle Av_{S_0},\, Av_{S_k}\rangle| \leqslant \sqrt{\delta_{2s}^2 - t^2}\,\|v_{S_0}\|_2\|v_{S_k}\|_2. \tag{14}$$

Let

$$u := \frac{v_{S_0}}{\|v_{S_0}\|_2} \quad \text{and} \quad w := \frac{e^{i\theta}v_{S_k}}{\|v_{S_k}\|_2}$$

where θ is such that $|\langle Au,\, Aw\rangle| = \Re\langle Au,\, Aw\rangle$. Then, for $\alpha, \beta \geqslant 0$,

$$\begin{aligned} 2\,|\langle Au,\, Aw\rangle| &= \frac{1}{\alpha+\beta}\left[\|A(\alpha u + w)\|_2^2 - \|A(\beta u - w)\|_2^2 - (\alpha^2 - \beta^2)\|Au\|_2^2\right] \\ &\leqslant \frac{1}{\alpha+\beta}\left[(1+\delta_{2s})\|\alpha u + w\|_2^2 - (1-\delta_{2s})\|\beta u - w\|_2^2 - (\alpha^2-\beta^2)(1+t)\|u\|_2^2\right] \\ \text{since } \langle u,\, w\rangle = 0 \qquad &= \frac{1}{\alpha+\beta}\left[(1+\delta_{2s})(\alpha^2+1) - (1-\delta_{2s})(\beta^2+1) - (\alpha^2-\beta^2)(1+t)\right] \\ &= \frac{1}{\alpha+\beta}\left[(\delta_{2s}-t)\alpha^2 + (t+\delta_{2s})\beta^2 + 2\delta_{2s}.\right] \end{aligned}$$

By substituting in the values

$$\alpha = \frac{\delta_{2s} + t}{\sqrt{\delta_{2s}^2 - t^2}} \quad \text{and} \quad \beta = \frac{\delta_{2s} - t}{\sqrt{\delta_{2s}^2 - t^2}},$$

we have that

$$2\,|\langle Au,\, Aw\rangle| \leqslant \frac{\sqrt{\delta_{2s}^2 - t^2}}{2\delta_{2s}}\,[\delta_{2s} + t + (\delta_{2s} - t) + 2\delta_{2s}.] = 2\sqrt{\delta_{2s}^2 - t^2},$$

and thereby proving (14).

Therefore,

$$\begin{aligned}
\|Av_{S_0}\|_2^2 &= \left\langle Av_{S_0}, A\left(v - \sum_{k\geqslant 1} v_{S_k}\right)\right\rangle \\
&= \langle Av_{S_0},\, Av\rangle - \sum_{k\geqslant 1}\langle Av_{S_0},\, Av_{S_k}\rangle \\
&\leqslant \|Av_{S_0}\|_2\|Av\|_2 + \sum_{k\geqslant 1}\sqrt{\delta_{2s}^2 - t^2}\|v_{S_0}\|_2\|v_{S_k}\|_2 \\
&= \|v_{S_0}\|_2\left(\sqrt{1+t}\|Av\|_2 + \sum_{k\geqslant 1}\sqrt{\delta_{2s}^2 - t^2}\|v_{S_k}\|_2\right).
\end{aligned}$$

For $k \geqslant 1$, let v_k^- and v_k^+ denote respectively the smallest and largest absolute entries of v on S_k. By Lemma 5,

$$\begin{aligned}
\sum_{k\geqslant 1}\|v_{S_k}\|_2 &\leqslant \sum_{k\geqslant 1}\frac{\|v_{S_k}\|_1}{\sqrt{s}} + \frac{\sqrt{s}}{4}(v_k^+ - v_k^-) \leqslant \sum_{k\geqslant 1}\frac{\|v_{S_k}\|_1}{\sqrt{s}} + \frac{\sqrt{s}}{4}v_1^+ \\
&\leqslant \frac{1}{\sqrt{s}}\|v_{S_0^c}\|_1 + \frac{1}{4}\|v_{S_0}\|_2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
(1+t)\|v_{S_0}\|_2^2 = \|Av_{S_0}\|_2^2 &\leqslant \|v_{S_0}\|_2\left(\sqrt{1+t}\|Av\|_2 + \frac{\sqrt{\delta_{2s}^2 - t^2}}{\sqrt{s}}\|v_{S_0^c}\|_1 + \frac{\sqrt{\delta_{2s}^2 - t^2}}{4}\|v_{S_0}\|_2\right) \\
\implies \|v_{S_0}\|_2 &\leqslant \left(\frac{1}{\sqrt{1+t}}\|Av\|_2 + \frac{\sqrt{\delta_{2s}^2 - t^2}}{(1+t)\sqrt{s}}\|v_{S_0^c}\|_1 + \frac{\sqrt{\delta_{2s}^2 - t^2}}{4(1+t)}\|v_{S_0}\|_2\right).
\end{aligned}$$

Since

$$\frac{\sqrt{\delta_{2s}^2 - t^2}}{1+t} \leqslant \frac{\delta_{2s}}{\sqrt{1-\delta_{2s}^2}} \quad \text{and} \quad \frac{1}{\sqrt{1+t}} \leqslant \frac{1}{\sqrt{1-\delta_{2s}}},$$

we have the robust NSP as required:

$$\begin{aligned}
\|v_{S_0}\|_2 &\leqslant \left(1 - \frac{\delta_{2s}^2}{4\sqrt{1-\delta_{2s}}}\right)^{-1}\left(\frac{1}{\sqrt{1-\delta_{2s}}}\|Av\|_2 + \frac{\delta_{2s}}{\sqrt{s(1-\delta_{2s})}}\|v_{S_0^c}\|_1\right) \\
&= \frac{\sqrt{1+\delta_{2s}}}{\sqrt{1-\delta_{2s}^2} - \delta_{2s}/4}\|Av\|_2 + \frac{\delta_{2s}}{\sqrt{1-\delta_{2s}^2} - \delta_{2s}/4}\frac{\|v_{S_0^c}\|_1}{\sqrt{s}}.
\end{aligned}$$

Therefore, the ℓ^2-robust null space property hold if

$$\frac{\delta_{2s}}{\sqrt{1-\delta_{2s}^2} - \delta_{2s}/4} < 1 \iff \delta^2 < \frac{16}{41}.$$

□

Lecture 14

5 Sparse recovery with random matrices

In this section, we will show that random matrices satisfy the RIP.

5.1 Basics from probability theory

We recall some notions from probability theory: Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, where Σ is a σ-algebra on Ω (that is, a collection of subsets of Ω that includes $\emptyset$, is closed under complement, and is closed under countable unions and intersections), and $\mathbb{P}$ is a probability measure on (Ω, Σ).

- for all $B \in \Sigma$,
$$\mathbb{P}(B) = \int_B \mathrm{d}\mathbb{P}(\omega) = \int_\Omega \mathbb{1}_B \mathrm{d}\mathbb{P}(\omega) \in [0,1],$$
- We have the *union bound*: for all sequences $(B_l)_l \subset \Sigma$,
$$\mathbb{P}(\bigcup_{l=1}^N B_l) \leqslant \sum_{l=1}^N \mathbb{P}(B_l).$$
- A random variable (r.v.) is a real valued measurable function on (Ω, Σ), where we say that X is measurable if
$$X^{-1}(A) = \{\omega \in \Omega \,;\, X(\omega) \in A\} \in \Sigma$$
for all Borel measurable subsets $A \subset \mathbb{R}$.
- The distribution function $F = F_X$ of a random variable X is
$$F(t) = \mathbb{P}(X \leqslant t), \qquad t \in \mathbb{R}.$$
- A random variable X has probability density function $\varphi_X : \mathbb{R} \to \mathbb{R}_+$ if
$$\mathbb{P}(X \in [a,b]) = \int_a^b \varphi_X(t)\mathrm{d}t, \qquad \forall a < b \in \mathbb{R}.$$
- The expectation of a random variable X is $\mathbb{E}[X] = \int_\Omega X(\omega)\mathrm{d}\mathbb{P}(\omega)$.
- Given $g : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}[g(X)] = \int_{-\infty}^\infty g(t)\varphi_X(t)\mathrm{d}t$.

Definition 8 (Common random variables).
- *A Rademacher variable (sometimes also called symmetric Bernoulli variable) is a random variable that takes the values $+1$ and -1 with equal probability.*
- *A normally distributed random variable (a.k.a. Gaussian random variable) with mean $\mathbb{E}[X] = \mu$ and variance $\mathbb{E}[(X-\mu)^2] = \sigma^2$, has probability density function*
$$\psi(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right).$$
A Gaussian random variable with mean 0 and variance 1 is called a standard Gaussian random variable.

- *A random variable is called subgaussian if there exists $\beta, \kappa > 0$ such that*

$$\mathbb{P}(|X| \geqslant t) \leqslant \beta \exp(-\kappa t^2), \qquad \forall t > 0.$$

It is called subexponential if there exists constants $\beta, \kappa > 0$ such that

$$\mathbb{P}(|X| \geqslant t) \leqslant \beta \exp(-\kappa t), \qquad \forall t > 0.$$

Remark 3. One can show that if X is a standard Gaussian random variable, then

$$\mathbb{P}(|X| \geqslant u) \leqslant \exp\left(-\frac{u^2}{2}\right) \quad \text{and} \quad \mathbb{E}(\exp(\theta X)) = \exp\left(\frac{\theta^2}{2}\right).$$

So, X is subgaussian with $\kappa = 1/2$ and $\beta = 1$. Note also that Rademacher random variables and in fact, any bounded random variables are also subgaussian.

Some useful results on subgaussian random variables

Proposition 2 (Equivalent characterization)**.** *Let X be a random variable.*

(a) If X is subgaussian with $\beta, \kappa > 0$ and $\mathbb{E}X = 0$, then there exists a constant c (dependent only on β and κ) such that

$$\mathbb{E}[\exp(\theta X)] \leqslant \exp(c\theta^2), \qquad \forall \theta \in \mathbb{R}. \tag{15}$$

(b) Conversely, if (15) *holds, then $\mathbb{E}X = 0$ and X is subgaussian with $\beta = 2$ and $\kappa = 1/(4c)$.*

Theorem 13 (Sum of subgaussian r.v.'s)**.** *Let $\{X_j\}_{j=1}^M$ be a sequence of independent mean zero subgaussian random variables, with parameter c in* (15)*. For $a \in \mathbb{R}^M$, the random variable $Z = \sum_{l=1}^M a_l X_l$ is subgaussian. In particular,*

$$\mathbb{E}[\exp(\theta Z)] \leqslant \exp\left(c\|a\|_2^2\theta^2\right), \qquad \forall \theta \in \mathbb{R}.$$

and

$$\mathbb{P}\left(|Z| \geqslant t\right) \leqslant 2\exp\left(-\frac{t^2}{4c\|a\|_2^2}\right), \qquad \forall t > 0.$$

Proposition 3 (Bernstein inequality for subexponential random variables)**.** *Let $\{X_l\}_{l=1}^M$ be independent mean zero subexponential random variables. i.e. $\mathbb{P}(|X_l| \geqslant t) \leqslant \beta e^{-\kappa t}$ with $\beta, \kappa > 0$ for all $t > 0$. Then,*

$$\mathbb{P}(\left|\sum_{l=0}^M X_l\right| \geqslant t) \leqslant 2\exp\left(-\frac{(\kappa t)^2/2}{2\beta M + \kappa t}\right), \qquad \forall t > 0.$$

5.2 Random matrices and the RIP

Definition 9. *Let $A \in \mathbb{R}^{m \times N}$.*

(a) If the entries of A are independent Rademacher variables then A is called a Bernoulli random matrix.

(b) If the entries of A are independent standard Gaussian random variables, then A is called a Gaussian random matrix.

(c) If the entries of A are independent mean zero subgaussian random variable with variance 1, then A is called a subgaussian random matrix.

Note that Gaussian random matrices and Bernoulli random matrices are also subgaussian random matrices.

Theorem 14. *Let $A \in \mathbb{R}^{m\times N}$ be a subgaussian random matrix. Then there exists $C > 0$ (dependent only on subgaussian parameters β, κ) such that the s-restricted isometry constant of $m^{-1/2}A$ satisfies $\delta_s \leqslant \delta$ with probability at least $1 - \varepsilon$, provided that*

$$m \geqslant C\delta^{-2}\left(s\ln(eN/s) + \ln(2\varepsilon^{-1})\right).$$

Setting $\varepsilon = 2\exp\left(-\delta^2 m/(2C)\right)$ yields

$$m \geqslant 2C\delta^{-2}s\ln(eN/s).$$

Note that the scaling factor of $m^{-1/2}$ makes sense because $\mathbb{E}\|m^{-1/2}Ax\|_2^2 = \|x\|_2^2$.

Definition 10. *Let $Y \in \mathbb{R}^N$ be a random vector.*

(a) If $\mathbb{E}[|\langle Y, x\rangle|^2] = \|x\|^2$ for all $x \in \mathbb{R}^N$, then Y is called isotropic.

(b) If for all $x \in \mathbb{R}^N$ with $\|x\|_2 = 1$, the random variable $\langle Y, x\rangle$ is subgaussian with subgaussian parameter c (independent of x) so that

$$\mathbb{E}[\exp(\theta\langle Y, x\rangle)] \leqslant \exp(c\theta^2), \qquad \theta \in \mathbb{R},$$

then Y is called a subgaussian random vector.

As a consequence of the following lemma, it is sufficient to consider matrices $A \in \mathbb{R}^{m\times N}$ of the form

$$A = \begin{pmatrix} Y_1^T \\ \vdots \\ Y_m^T \end{pmatrix}$$

with rows specified by independent subgaussian and isotropic vectors $Y_1, \ldots, Y_m$.

Lemma 6. *Let Y be a vector with independent mean zero and subgaussian entries with variance 1 and subgaussian parameter c. Then, Y is an isotropic and subgaussian random vector with subgaussian parameter c.*

Proof. Let $Y = (y_j)_{j=1}^M$. Then, for $x = (x_j)_{j=1}^M$,

$$\mathbb{E}[\langle Y, x\rangle^2] = \mathbb{E}(\sum_{j=1}^N x_j y_j)^2 = \sum_{j=1}^N\sum_{l=1}^N x_j x_l \mathbb{E}(y_j y_l).$$

Note that $\mathbb{E}(y_j y_l) = \mathbb{E}(y_j)\mathbb{E}(y_l) = 0$ for $j \neq l$ by independence of the entries of Y and $\mathbb{E}(y_j^2) = 1$. Therefore, condition (a) of Definition 10 is satisfied.

Condition (b) follows because by Theorem 13, a linear combination of independent mean zero subgaussian random variables is also subgaussian with the same subgaussian parameter. □

For the remainder of this section, we shall prove the following theorem:

Theorem 15. *Let $A \in \mathbb{R}^{m\times N}$ be a random matrix with independent, isotropic and subgaussian rows with the same subgaussian parameter c. If*

$$m \geqslant C\delta^{-2}(s\ln(eN/s) + \ln(2\varepsilon^{-1})),$$

then the restricted isometry constant of $m^{-1/2}A$ satisfies $\delta_s \leqslant \delta$ with probability at least $1 - \varepsilon$.

Lemma 7. *Let $A \in \mathbb{R}^{m\times N}$ be a random matrix with independent, isotropic and subgaussian rows with the same subgaussian parameters β, κ. Then for all $x \in \mathbb{R}^N$, and all $t \in (0, 1)$,*

$$\mathbb{P}(\left|m^{-1}\|Ax\|_2^2 - \|x\|_2^2\right| \geqslant t\|x\|_2^2) \leqslant 2\exp(-\tilde{c}t^2 m)$$

where $\tilde{c}$ depends only on β, κ.

Proof. Let $x \in \mathbb{R}^N$. W.l.o.g. assume that $\|x\|_2 = 1$. Denote the rows of A by $Y_1, \ldots, Y_m \in \mathbb{R}^N$ and consider the random variables

$$Z_l = |\langle Y_l,\, x\rangle|^2 - \|x\|_2^2, \qquad l \in [m].$$

Since Y_l is isotropic, we have that $\mathbb{E}(Z_l) = 0$. Furthermore, Z_l is subexponential since $\langle Y_l,\, x\rangle$ is subgaussian:

$$\mathbb{P}(|Z_l| \geqslant r) \leqslant \mathbb{P}(|\langle Y_l,\, x\rangle|^2 \geqslant r - \|x\|_2^2) = \mathbb{P}(|\langle Y_l,\, x\rangle|^2 \geqslant r - 1) \leqslant \begin{cases} 1 & r < 1, \\ \beta \exp(-\kappa(r-1)) & r \geqslant 1. \end{cases}$$

So,

$$\mathbb{P}(|Z_l| \geqslant r) \leqslant \tilde{\beta} \exp(-\tilde{\kappa} r), \qquad \forall r > 0$$

where $\tilde{\kappa} = \kappa$ and $\tilde{\beta} = \max\{\beta e^{\kappa}, e^{\kappa}\}$. Observe that

$$m^{-1}\|Ax\|_2^2 - \|x\|_2^2 = \frac{1}{m}\sum_{l=1}^m \left(|\langle Y_l,\, x\rangle|^2 - \|x\|_2^2\right) = \frac{1}{m}\sum_{l=1}^m Z_l.$$

Since Y_l are independent, Z_l are also independent. So, by the Bernstein inequality for subexponential random variables, for all $t \in (0,1)$,

$$\mathbb{P}\left(\Big|\frac{1}{m}\sum_{l=1}^m Z_l\Big| \geqslant t\right) = \mathbb{P}\left(\Big|\sum_{l=1}^m Z_l\Big| \geqslant tm\right) \leqslant 2\exp\left(-\frac{\tilde{\kappa}^2 m^2 t^2/2}{2\tilde{\beta} m + \tilde{\kappa} m t}\right) \leqslant 2\exp\left(-\frac{\tilde{\kappa}^2}{4\tilde{\beta} + 2\tilde{\kappa}} m t^2\right),$$

where the last inequality follows because $t \in (0,1)$. □

Lemma 8 (Upper bound on covering)**.** *Let $N \in \mathbb{N}$ and $\rho \in (0, 1/2)$. For any $S \subset [N]$, there exists a finite subset U of*

$$B_S = \left\{x \in \mathbb{R}^N \;;\; \mathrm{Supp}(x) \subset S, \|x\|_2 \leqslant 1\right\}$$

such that

$$\mathrm{Card}(U) \leqslant \left(1 + \frac{2}{\rho}\right)^s \quad \textit{and} \quad \min_{u \in U} \|z - u\|_2 \leqslant \rho, \quad \forall z \in B_S.$$

Proof. Given a subset Y of $\{x \in \mathbb{R}^s : \|x\|_2 \leqslant 1\}$, let $\mathcal{N}(Y, r)$ be the smallest integer such that there exists $\{x_j\}_{j=1}^{\mathcal{N}} \subset Y$ is an r-covering for Y: i.e.

$$Y \subset \bigcup_{j=1}^{\mathcal{N}} B(x_j, r).$$

Define the packing number $\mathcal{P}(Y, r)$ to be the largest integer for which there exists a r-packing for Y: i.e. there exists $\{x_j\}_{j=1}^{\mathcal{P}} \subset Y$ for which $|x_j - x_k| > r$ for all $j \neq k$.

First note that $\mathcal{N}(Y, r) \leqslant \mathcal{P}(Y, r)$. Indeed, any maximal packing $\{x_j\}_{j=1}^{\mathcal{P}}$ of Y is also a r-covering of Y because if there exists x not covered by $\bigcup_{j=1}^{\mathcal{P}} B(x_j, r)$, then $|x - x_j| > r$ for all j and hence, $\{x_1, \ldots, x_{\mathcal{P}}, x\}$ is an r-packing of Y which contradicts our assumption that $\{x_j\}_{j=1}^{\mathcal{P}}$ is maximal.

Let $\mathcal{P} = \mathcal{P}(Y, \rho)$ and let $\{x_j\}_{j=1}^{\mathcal{P}}$ be the maximal packing. Then, we have $\mathcal{P}$ nonintersecting balls of radius $\rho/2$ inside $B(0, 1+\rho/2)$. By comparing volumes, we have that

$$\mathrm{Vol}\left(\bigcup_{j=1}^{\mathcal{P}} B(x_j, \rho/2)\right) = \mathcal{P} \cdot \mathrm{Vol}(B(x_j, \rho/2)) \leqslant \mathrm{Vol}(B(0, 1+\rho/2)).$$

Since $\mathrm{Vol}(B(0,t)) = t^s \mathrm{Vol}(B(0,1))$,

$$\mathcal{P} \leqslant \left(1 + \frac{2}{\rho}\right)^s.$$

Therefore, $\mathcal{N}(Y,\rho) \leqslant \left(1+\frac{2}{\rho}\right)^s$. This concludes the proof since B_S can be identified with the unit ball of dimension s.

□

Lecture 15

Theorem 16. *Let $A \in \mathbb{R}^{m\times N}$ be a random matrix such that for all $t \in (0,1)$,*

$$\mathbb{P}(\left|\|Ax\|_2^2 - \|x\|_2^2\right| \geqslant t\|x\|_2^2) \leqslant 2\exp(-\tilde{c}t^2 m). \tag{16}$$

For $S \subset [N]$ with $\mathrm{Card}(S) = s$ and $\delta, \varepsilon \in (0,1)$ if

$$m \geqslant C\delta^{-2}(7s + 2\ln(2\varepsilon^{-1})),$$

where $C = 2/(3\tilde{c})$, then with probability at least $1-\varepsilon$,

$$\|A_S^* A_S - I\|_2 < \delta.$$

Proof. Let U be as in Lemma 8. Let $t \in (0,1)$, whose exact value will be determined later. Then,

$$\begin{aligned}
&\mathbb{P}\left(\left|\|Au\|_2^2 - \|u\|_2^2\right| \geqslant t\|u\|_2^2 \quad \text{for some } u \in U\right) \leqslant \sum_{u\in U} \mathbb{P}\left(\left|\|Au\|_2^2 - \|u\|_2^2\right| \geqslant t\|u\|_2^2\right) \\
&\leqslant 2\,\mathrm{Card}(U)\exp\left(-\tilde{c}t^2 m\right) \leqslant 2\left(1+\frac{2}{\rho}\right)^s \exp\left(-\tilde{c}t^2 m\right).
\end{aligned}$$

So, we have just shown that with probability at least

$$1 - 2\left(1+\frac{2}{\rho}\right)^s \exp\left(-\tilde{c}t^2 m\right),$$

the matrix A satisfies:

$$\left|\|Au\|_2^2 - \|u\|_2^2\right| < t\|u\|_2^2, \qquad \forall u \in U. \tag{17}$$

Let $B = A_S^* A_S - I$, then (17) is equivalent to

$$|\langle Bu,\, u\rangle| < t\|u\|_2^2 < t, \qquad \forall u \in U.$$

Now, let $x \in \mathcal{B}_S := \{z\ ;\ \mathrm{Supp}(z) \subseteq S, \|z\|_2 \leqslant 1\}$. Then, there exists $u \in U$ such that

$$\|u - x\|_2 \leq \rho < \frac{1}{2}.$$

So,

$$\begin{aligned}
|\langle Bx,\, x\rangle| &= |\langle Bu,\, u\rangle + \langle B(x+u),\, x-u\rangle| \\
&\leqslant |\langle Bu,\, u\rangle| + |\langle B(x+u),\, x-u\rangle| \\
&< t + \|B\|_{2\to 2}\|x+u\|_2\|x-u\|_2 \\
&\leqslant t + 2\|B\|_{2\to 2}\rho.
\end{aligned}$$

Taking the maximum over all $x \in \mathcal{B}_S$, it follows that

$$\|B\|_{2\to 2} < t + 2\rho\|B\|_{2\to 2} \implies \|B\|_{2\to 2} < \frac{t}{1-2\rho}.$$

By choosing $t = (1-2\rho)\delta$, we have that $\|B\|_{2\to 2} < \delta$. Therefore,

$$\mathbb{P}(\|A_S^* A_S - I\|_{2\to 2} \geqslant \delta) \leqslant 2\left(1+\frac{2}{\rho}\right)^s \exp\left(-\tilde{c}(1-2\rho)^2\delta^2 m\right).$$

It follows that $\|A_S^* A_S - I\|_{2\to 2} \leqslant \delta$ with probability at least $1-\varepsilon$ provided that

$$m \geqslant \frac{1}{\tilde{c}(1-2\rho)^2}\delta^{-2}\left(\ln(2+2/\rho)s + \ln(2\varepsilon^{-1})\right).$$

The conclusion follows by choosing $\rho = 2/(e^{7/2}-1) \approx 0.0623$ and this implies that

$$\frac{1}{(1-2\rho)^2} \leqslant \frac{4}{3} \quad \text{and} \quad \frac{\ln(1+2/\rho)}{(1-2\rho)^2} \leqslant \frac{14}{3}.$$

□

Theorem 17. *Suppose that A is a random matrix such that* (16) *holds. If, for $\delta, \varepsilon \in (0,1)$,*

$$m \geqslant C\delta^{-2}\left(s(9+2\ln(N/s)) + 2\ln(2\varepsilon^{-1})\right),$$

where $C = 2/(3\tilde{c})$, then with probability at least $1-\varepsilon$, the restricted isometry constant δ_s of A satisfies $\delta_s < \delta$.

Proof. Recall that

$$\delta_s = \sup_{S\subset[N], \mathrm{Card}(S)\leqslant s} \|A_S^* A_S - I\|_{2\to 2}.$$

From the proof of Theorem 16,

$$\mathbb{P}(\|A_S^* A_S - I\|_{2\to 2} \geqslant \delta) \leqslant 2\left(1+\frac{2}{\rho}\right)^s \exp\left(-\tilde{c}(1-2\rho)^2\delta^2 m\right).$$

So,

$$\begin{aligned}\mathbb{P}(\delta_s \geqslant \delta) &\leqslant \sum_{S\subset[N], \mathrm{Card}(S)=s} \mathbb{P}(\|A_S^* A_S - I\|_{2\to 2} \geqslant \delta) \leqslant 2\binom{N}{s}\left(1+\frac{2}{\rho}\right)^s \exp\left(-\tilde{c}(1-2\rho)^2\delta^2 m\right)\\ &\leqslant 2\left(\frac{eN}{s}\right)^s\left(1+\frac{2}{\rho}\right)^s \exp\left(-\tilde{c}(1-2\rho)^2\delta^2 m\right).\end{aligned}$$

where we have applied Stirling's estimate that for all $n \geqslant k > 0$, $\binom{n}{k} \leqslant (en/k)^k$. Making the choice $\rho = 2/(e^{7/2}-1)$ yields the required result.

□

Recovery of vectors which are sparse with respect to some ONB

Corollary 2. *Let $U \in \mathbb{R}^{N\times N}$ be a fixed orthogonal matrix. Suppose that $A \in \mathbb{R}^{m\times N}$ is a random matrix which is drawn in accordance to a probablity distribution for which*

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - \|x\|_2^2\right| > t\|x\|_2^2\right) \leqslant 2\exp(-\tilde{c}t^2 m),$$

for all $t \in (0,1)$, $x \in \mathbb{R}^N$. Then, given $\delta, \varepsilon \in (0,1)$, the restricted isometry constant of AU satisfies $\delta_s < \delta$ with probability at least $1-\varepsilon$ provided that

$$m \geqslant \frac{2}{3\tilde{c}}\delta^{-2}\left(s(9+2\ln(\frac{N}{s})) + 2\ln(2\varepsilon^{-1})\right).$$

Proof. Let $x \in \mathbb{R}^N$ and $x' = Ux$. Then

$$\mathbb{P}\left(\left|\|AUx\|_2^2 - \|x\|_2^2\right| > t\|x\|_2^2\right) = \mathbb{P}\left(\left|\|Ax'\|_2^2 - \|x'\|_2^2\right| > t\|x'\|_2^2\right) \leqslant 2\exp(-\tilde{c}t^2 m),$$

The conclusion follows from Theorem 17.

□

5.3 Relationship to Johnson-Lindenstrauss embeddings

The Johnson-Lindenstrauss lemma (Lemma 9) is not about sparsity, but closely related to the concentration inequality for subgaussian matrices that we proved in Lemma 7.

Given a point set $\{x_1, \ldots, x_M\} \subset \mathbb{R}^N$, it is expensive to process these points when N is large. It is therefore of interest to project them onto a low dimensional space while preserving geometric properties such as mutual distances.

Lemma 9. *Let $x_1, \ldots, x_M \in \mathbb{R}^N$ and let $\eta > 0$. If $m > C\eta^{-2}\ln(M)$, then there exists $B \in \mathbb{R}^{m\times N}$ such that*

$$(1-\eta)\|x_j - x_l\|_2^2 \leqslant \|B(x_j - x_l)\|_2^2 \leqslant (1+\eta)\|x_j - x_l\|_2^2$$

for all $j, l \in [M]$ with $j \neq l$. The constant $C > 0$ is universal.

Proof. Consider $E = \{x_j - x_l \; ; \; 1 \leqslant j < l \leqslant M\}$. Then, $|E| \leqslant M(M-1)/2$. It is enough to show that there exists B such that

$$(1-\eta)\|x\|^2 \leqslant \|Bx\|^2 \leqslant (1+\eta)\|x\|^2, \qquad \forall x \in E. \tag{18}$$

Let $B = \frac{1}{\sqrt{m}}A \in \mathbb{R}^{m\times N}$ be a subgaussian random matrix. Then, for any fixed $x \in E$, by Lemma 7,

$$\mathbb{P}\left(\left|\|Bx\|^2 - \|x\|^2\right| \geqslant \eta\|x\|^2\right) \leqslant 2\exp(-\tilde{c}m\eta^2).$$

Therefore, by applying the union bound, (18) holds with probability at least

$$1 - M^2\exp(-\tilde{c}m\eta^2) \geqslant 1 - \varepsilon,$$

provided that

$$m \geqslant \tilde{c}^{-1}\eta^{-2}\ln(M^2/\varepsilon).$$

Now, the existence of a map B is established as soon as $\varepsilon < 1$. Letting $\varepsilon \to 1$, the claim of this lemma is true with $C = \tilde{c}^{-1}$. □

The following theorem shows that matrices with small restricted isometry constants give rise to Johnson-Lindenstrauss embeddings:

Theorem 18. *Let $E \subset \mathbb{R}^N$ be a finite point set such that $|E| = M$. For $\eta, \varepsilon \in (0,1)$, let $A \in \mathbb{R}^{m\times N}$ with restricted isometry constant satisfying $\delta_{2s} \leqslant \eta/4$ for some $s \geqslant 16\ln(4M/\varepsilon)$. Let $\gamma = (\gamma_1, \ldots, \gamma_N)$ be a Rademacher sequence and let D_γ be the diagonal matrix with γ as its diagonal. Then, with probability at least $1-\varepsilon$,*

$$(1-\eta)\|x\|_2^2 \leqslant \|AD_\gamma x\|_2^2 \leqslant (1+\eta)\|x\|_2^2, \qquad \forall x \in E. \tag{19}$$

Remark 4. We shall not prove this result, but simply make some comments:

1. The theorem is false without the sign randomization of the columns of A: there is not assumption on the set E, and in particular, by choosing $E \subset \mathcal{N}(A)$, we see that (19) is false without the matrix D_γ. The randomization of column signs essentially ensure that the probability that $E \cap \mathcal{N}(AD_\gamma) \neq \emptyset$ is very small.

2. There is an indirect lower bound on the embedding dimension m. Since we require that $\delta_{2s} \leqslant \eta/4$, one would expect that $m \geqslant C\eta^{-2} s \ln^{\alpha}(N)$ for some $\alpha \geqslant 1$. Since $s \geqslant 16\ln(4M/\varepsilon)$, it follows that

$$m \geqslant C\eta^{-2} s \ln^{\alpha}(N)\ln(4M/\varepsilon),$$

so there is an extra factor of $\ln^{\alpha}(N)$ compared with the original Johnson-Lindenstrauss lemma.

6 Nonuniform recovery guarantees

The kind of recovery conditions (NSP, RIP) presented so far lead to universal recovery guarantees "Under certain conditions, we can recover all s-sparse vectors". However, one could consider the conditions for nonuniform recovery guarantees: "For a fixed s-sparse vector x, under certain conditions, we can recover x".

Theorem 19. *Let $A \in \mathbb{C}^{m\times N}$ and let $x \in \mathbb{C}^N$ with $\mathrm{Supp}(x) = S$. Suppose that either of the following conditions are satisfied:*

(a) $|\langle v_S, \mathrm{sign}(x)_S\rangle| < \|v_{S^c}\|_1$ *for all* $v \in \mathcal{N}(A) \setminus \{0\}$.

(b) A_S *is injective and there exists* $\rho = A^* h$ *such that*

$$\rho_S = \mathrm{sign}(x)_S, \qquad |\rho_j| < 1,\ \forall j \in S^c.$$

Then, x is the unique solution to

$$\min \|z\|_1 \text{ subject to } Ax = Az.$$

Proof. Suppose that (a) holds. Let $Ax = Az$ with $z \neq x$. Let $v = x - z \in \mathcal{N}(A) \setminus \{0\}$. Then,

$$\begin{aligned}\|z\|_1 &= \|z_S\|_1 + \|z_{S^c}\|_1 = \|(x-v)_S\|_1 + \|v_{S^c}\|_1 \\ &> |\langle (x-v)_S, \mathrm{sign}(x)_S\rangle| + |\langle v, \mathrm{sign}(x)_S\rangle| \geqslant |\langle x, \mathrm{sign}(x)_S\rangle| = \|x\|_1.\end{aligned}$$

Therefore, x is the unique solution to basis pursuit.

Suppose that (b) holds. We aim to show that (a) holds: note that $Av_S = -Av_{S^c}$ for all $v \in \mathcal{N}(A) \setminus \{0\}$. Then,

$$|\langle v, \mathrm{sign}(x)_S\rangle| = |\langle v_S, A^* h\rangle| = |\langle Av_S, h\rangle| = |\langle Av_{S^c}, h\rangle| \leqslant \|(A^*h)_{S^c}\|_\infty \|v_{S^c}\|_1 < \|v_{S^c}\|_1$$

where the last inequality follows because $v_{S^c} \neq 0$. This is true since $v_{S^c} = 0$, implies that $A_S v_S = 0$ which is a contradiction to A_S being injective. □

Remark 5. In fact, we also have that (a) implies (b). In general, the converse to Theorem 19 is false. However, in the case of real matrices and real vectors, one can show that uniqueness to basis pursuit for a vector x implies that (a) holds. See the exercise sheet for details.