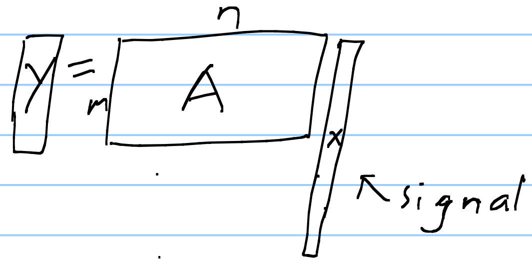
Dudley's $\neq$ to prove RIP

Note Title 2015-10-14

Recall compressed sensing observation model:



$\|x\|_0 = s$, i.e. x has s non-zero entries, or x is <u>s-sparse</u>.

Goal: Given A & y, recover x.

Challenge: A is not invertible.

However, if A is invertible (and well-conditioned) when restricted to sparse vectors, then x can be recovered (efficiently & stably).

<u>Def</u> (<u>Restricted Isometry Property RIP</u>)

We say that a matrix $A \in \mathbb{R}^{m \times n}$ satisfies the $(s, \frac{1}{2})$-<u>RIP</u> if

$$\frac{1}{2}\|x\|_2^2 \leq \|Ax\|_2^2 \leq \frac{3}{2}\|x\|_2^2$$

$$\forall x \in K_s := \{x \in \mathbb{R}^n : \|x\|_0 \leq s\}.$$

- Random A satisfies $(s, \frac{1}{2})$-RIP w.h.p. if $m \geq c\, s \operatorname{polylog}(n)$.
- <u>No</u> known deterministic constructions if $m \leq c s^{2-\epsilon}$.

Thm Let $A \in \mathbb{R}^{m \times n}$ have iid $N(0,1)$ entries, $n \geq 2s$. Then

$$\mathbb{E} \sup_{x \in K_s \cap S^{n-1}} \left| \frac{\|Ax\|_2}{\sqrt{m}} - 1 \right| \leq C \sqrt{\frac{s \log(\frac{n}{s})}{m}}$$

— Implies $(s, \frac{1}{2})$-RIP when $m \geq C s \log(\frac{n}{s})$

Proof (of thm via Dudley's $\neq$)

Let $K' = K_s \cap S^{n-1}$. We will show that

$$\mathbb{E} \sup_{x \in K'} \left| \|Ax\|_2 - \mathbb{E}\|Ax\|_2 \right| \leq C \sqrt{s \log(\tfrac{n}{s})}.$$

Let $X_t = \|At\|_2 - \mathbb{E}\|At\|_2$

Recall: $\|X_t - X_s\|_{\psi_2} \leq C\|s - t\|_2$

Thus, for Dudley's $\neq$, we need only bound $N(K', \|\cdot\|_2, \epsilon)$. Note: K' is the union of $\binom{n}{s}$ spheres S^{s-1}. Thus

$$N(K', \|\cdot\|_2, \epsilon) \leq \binom{n}{s} N(S^{s-1}, \|\cdot\|_2, \epsilon) \leq \left(\frac{ne}{s}\right)^s \cdot \left(\frac{C}{\epsilon}\right)^s \qquad (*)$$

By Dudley's $\neq$,

$$\begin{aligned}
\sup_{t \in K'} |X_t| &\leq C \int_0^2 \sqrt{\log(N(K', \|\cdot\|_2, \epsilon))}\, d\epsilon \\
&= C \int_0^2 \sqrt{s\left(\log\left(\frac{n}{s}\right) + \log\left(\frac{C}{\epsilon}\right)\right)}\, d\epsilon \qquad (\text{By } (*)) \\
&\leq C\sqrt{s} \int_0^2 \left(\sqrt{\log(\tfrac{n}{s})} + \sqrt{\log(\tfrac{C}{\epsilon})}\right) d\epsilon \\
&\leq C \sqrt{s \log(\tfrac{n}{s})} + C' \\
&\leq C \sqrt{s \log(\tfrac{n}{s})}
\end{aligned}$$

QED

In practice, structured-random matrices are vital for compressed sensing.

Structured: Allows fast multiplication, comes up naturally in applications, e.g., DFT for MRI.

Random: Allows provable results such as RIP.

<u>structured-random model</u>:

$A \in \mathbb{R}^{m \times n}$ has rows $a_1^T, a_2^T, \dots a_m^T \overset{iid}{\sim} a^T$, and $a \in \mathbb{R}^n$, satisfies

<u>Incoherence condition</u>: $\|a\|_\infty \leq 2$ a.s. for simplicity

<u>Isotropy condition</u>: $\mathbb{E}\, aa^* = I$.

Example) a_i are random rows of DFT.

<u>Exercise</u>: Why are these conditions that may allow RIP to hold?

<u>Thm</u> (Rudelson-Vershynin '06, Candes-Tao '05)

Let A come from the structured-random model. Then,

$$\mathbb{E} \sup_{x \in K^1} \left| \frac{1}{m} \|Ax\|_2^2 - 1 \right| \leq \frac{1}{10}$$

as long as $n \geq m \geq C_0 s \log^4 n$.

– <u>Exercise</u>: <u>Thm</u> implies $(s, \frac{1}{2})$RIP w.p. $\geq 80\%$

Define the matrix norm $\|\cdot\|_s : Q \mapsto \sup_{x \in K^1} x^T Q x$.

Note that $\sup_{x \in K^1} \left| \frac{1}{m}\|Ax\|_2^2 - 1 \right| = \|\frac{1}{m} A^T A - I\|_s = \|\underbrace{\frac{1}{m} \sum_{i=1}^m a_i a_i^T - I}_{\text{mean-zero}}\|_s$.

To control this quantity, we use the symmetrization trick (to be proven next time). This gives

$$\mathbb{E}\left\|\frac{1}{m}\sum_{i=1}^{m} a_i a_i^T - I\right\|_s \le 2\,\mathbb{E}\left\|\frac{1}{m}\sum_{i=1}^{m} \xi_i a_i a_i^T\right\|_s \qquad (**)$$

where $\xi_1,\dots,\xi_m$ are iid Rademacher.

The right-hand side of $(**)$ is controlled w/ the following lemma (by conditioning on $\{a_i\}$).

Lemma Let $a_1,\dots,a_m$ be fixed vectors in $\mathbb{R}^n$, $\|a_i\|_\infty \le 1$, $m \le n$. Let $\xi_1,\dots,\xi_m$ be iid Rademacher. Then

$$\mathbb{E}\left\|\sum_{i=1}^{m}\xi_i a_i a_i^T\right\|_s \le C\sqrt{s}\,\log^2(n)\sqrt{\left\|\sum_{i=1}^{m} a_i a_i^T\right\|_s}$$

Proof (of thm based on the lemma)

Let $\delta := \mathbb{E}\left\|\frac{1}{m}\sum_{i=1}^{m} a_i a_i^T - I\right\|_s$

By $(**)$

$$\begin{aligned}
\delta &\le 2\,\mathbb{E}\left\|\frac{1}{m}\sum_{i=1}^{m}\xi_i a_i a_i^T\right\|_s \\
&\le \frac{C\sqrt{s}\,\log^2 n}{\sqrt{m}}\,\mathbb{E}\sqrt{\frac{\left\|\sum_{i=1}^{m} a_i a_i^T\right\|_s}{m}} && \text{(By lemma)} \\
&=: Z\cdot \mathbb{E}\sqrt{\left\|\frac{1}{m}\sum_{i=1}^{m} a_i a_i^T\right\|} \\
&\le Z\cdot\sqrt{\mathbb{E}\left\|\frac{1}{m}\sum_{i=1}^{m} a_i a_i^T\right\|_s} && \text{(By Jensen } \neq\text{)} \\
&\le Z\cdot\sqrt{\mathbb{E}\left\|\frac{1}{m}\sum_{i=1}^{m} a_i a_i^T - I\right\|_s + \underbrace{\|I\|_s}_{=1}} && \text{(By } \Delta \neq\text{)} \\
&= Z\sqrt{\delta+1}
\end{aligned}$$

Thus $\delta \le Z\cdot\sqrt{\delta+1} \le 5\cdot Z^2 + \frac{\delta+1}{20}$ $\qquad$ (since $a\cdot b \le \frac{c\,a^2}{2} + \frac{b^2}{2\cdot c}$ for any $c>0$, $a,b\in\mathbb{R}$)

$\Rightarrow \frac{19}{20}\delta \le 5Z^2 + \frac{1}{20} = C'\,\frac{s\log^4 n}{m} + \frac{1}{20}$

Recall $m \ge C_0\, s\log^4(n)$

Thus $\delta \le \frac{c}{C_0}\cdot\frac{20}{19} + \frac{1}{19}$

Take $C_0 \ge \frac{200}{9}c$. $\Rightarrow \delta \le \frac{1}{10}$ $\qquad$ QED

Proof (of the lemma via Dudley's $\neq$)

Random process: $X_t := \sum_{i=1}^m \xi_i \langle a_i, t\rangle^2, \quad t \in K'$.

Sub-Gaussian increments:

$$\|X_t - X_s\|_{\psi_2}^2 = \left\|\sum_{i=1}^m \xi_i \left(\langle a_i, t\rangle^2 - \langle a_i, s\rangle^2\right)\right\|_{\psi_2}^2$$

$$\leq C \sum_{i=1}^m \|\xi_i\|_{\psi_2}^2 \left(\langle a_i, t\rangle^2 - \langle a_i, s\rangle^2\right)^2 \quad \text{(as shown in Lecture 3)}$$

$$\leq C \sum_{i=1}^m \left(\langle a_i, t\rangle - \langle a_i, s\rangle\right)^2 \left(\langle a_i, t\rangle + \langle a_i, s\rangle\right)^2$$

$$\leq C \max_i \langle a_i, s-t\rangle^2 \cdot \sum_{i=1}^m \left(\langle a_i, t\rangle + \langle a_i, s\rangle\right)^2$$

$$\leq C R^2 \|s-t\|_X^2$$

where $R^2 := \sup_{t \in K'} \sum_{i=1}^m \langle a_i, t\rangle^2 = \left\|\sum_{i=1}^m a_i a_i^T\right\|_S^2$

$\|v\|_X := \max_i |\langle a_i, x\rangle|$

$\Rightarrow Y_t := \frac{X_t}{CR}$ has sub-Gaussian increments with associated metric $\|\cdot\|_X$.

We will show that $\mathbb{E} \sup_{t \in K'} Y_t \leq C\sqrt{s}\log^2 n$

thus completing the proof.

We begin with the containment

$K' \subseteq \sqrt{s}\, B_1^n = \sqrt{s}\{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$ $\boxed{\text{Proof: Exercise}}$

$\Rightarrow$ It suffices to bound $\mathbb{E} \sup_{t \in \sqrt{s} B_1} |Y_t|$

We need to control $N(\sqrt{s} B_1^n, \|\cdot\|_X, \epsilon)$ for Dudley $\neq$.

$$\boxed{\underline{\text{Lemma}}\quad \log\left(N\left(B_1^n, \|\cdot\|_X, \epsilon\right)\right) \le \frac{C}{\epsilon^2}\log^2(n)}$$

Proof (using Maurey's empirical method):

Fix $y \in B_1^n$.

Consider iid random vectors

$z_1, \dots, z_\ell$ satisfying

$$z_i = \begin{cases} e_i\,\mathrm{sign}(y_i) & \text{w.p. } |y_i| \\ 0 & \text{w.p. } 1-\|y\|_1 \end{cases}$$

$y_i :=$ i^{th} entry of y, $\quad e_i := (\underbrace{0, 0, \dots 0, 1}_{i}, 0, \dots 0)$

Note: $\mathbb{E}\, z_i = y$

Let $z = \frac{1}{\ell}\sum_{i=1}^{\ell} z_i - y$. We wish to bound $\|z\|_X$ w.h.p.

First consider $\langle z, a_1\rangle = \frac{1}{\ell}\sum_{i=1}^{n}\langle z_i, a_1\rangle - \langle y, a_1\rangle$

Now $\|\langle z_i, a_1\rangle\|_{\psi_2} \le \|\langle z_i, a_1\rangle\|_\infty \le \|a_1\|_\infty = 2$

$$\Rightarrow \|\langle z, a_1\rangle\|_{\psi_2} \le C\cdot\frac{1}{\ell^2}\sum_{i=1}^{m}\|\langle z_i, a_1\rangle - \langle y, a_1\rangle\|_{\psi_2}$$
$$\le C\frac{1}{\ell^2}\cdot\ell\cdot 2$$
$$= C\frac{1}{\ell}$$

Similarly for $\langle z, a_2\rangle, \dots \langle z, a_m\rangle$.

Thus $P(\|z\|_X > \epsilon) \le m P(|\langle z, a_1\rangle| > \epsilon)$
$$\le m e e^{-C\ell\epsilon^2}$$
$$\le \frac{1}{2}$$

by picking ℓ to satisfy $\ell = \left\lceil \frac{C}{\epsilon^2}\log m\right\rceil$

Thus, there is some vector V satisfying $V = \frac{1}{\ell}\sum_{l=1}^{\ell} e_{i_\ell} \cdot z_{i_\ell}$ satisfying $\|V - y\|_X \le \epsilon$, where $z_{i_\ell} \in \{-1, 0, 1\}$. This is true for any $y \in B_1$. There are only $(3n)^\ell$ possible choices for V.

$$\Rightarrow N(B_1, \|\cdot\|_X, \epsilon) \le (3n)^\ell = (3n)^{\frac{C\log m}{\epsilon^2}}$$

Take log of both sides. Use $\log(m) \le \log(n)$. $\boxed{\text{QED}}$

We may now apply Dudley's $\neq$ to bound $\mathbb{E}\sup_{t\in\sqrt{s}B_1} |Y_t| \le \int_0^{\operatorname{diam}(\sqrt{s}B_1)} \sqrt{\log N(\sqrt{s}B_1, \|\cdot\|_X, \epsilon)}\, d\epsilon$

Tidy this expression up:

1. Control $\operatorname{diam}(\sqrt{s}B_1)$: For $t \in \sqrt{s}B_1$, $|\langle a_i, t\rangle| \le \underbrace{\|a_i\|_\infty \|t\|_1}_{\le 2\sqrt{s}}$

$$\Rightarrow \|t\|_X \le 2\sqrt{s} \Rightarrow \operatorname{diam}(\sqrt{s}B_1) \le 4\sqrt{s}.$$

2. Note: $N(\sqrt{s}B_1, \|\cdot\|_X, \epsilon) = N(B_1, \|\cdot\|_X, \frac{\epsilon}{\sqrt{s}})$

3. We will control coarse scales:

$$\mathbb{E}\sup|Y_t| \le \underbrace{\int_{n^{-100}}^{4\sqrt{s}} \sqrt{\log N(B_1, \|\cdot\|_X, \tfrac{\epsilon}{\sqrt{s}})}\, d\epsilon}_{\text{coarse scales}} + \underbrace{\int_0^{n^{-100}} \sqrt{\log N(B_1, \|\cdot\|_X, \tfrac{\epsilon}{\sqrt{s}})}\, d\epsilon}_{\text{fine scales}}$$

$\boxed{\text{HW}}$ Show that $\int_0^{n^{-100}} \sqrt{\log N(B_1, \|\cdot\|_X, \frac{\epsilon}{\sqrt{s}})}\, d\epsilon$ is (much) less than $C\sqrt{s}\log^2(n)$

Coarse scales:

$$\int_{n^{-100}}^{4\sqrt{s}} \sqrt{\log N(B_1, \|\cdot\|_X, \tfrac{\epsilon}{\sqrt{s}})}\, d\epsilon$$

$$\leq \int_{n^{-100}}^{4\sqrt{s}} \sqrt{\frac{Cs}{\epsilon^2} \log^2 n}\; d\epsilon$$ (by lemma above)

$$= C\sqrt{s}\log(n) \int_{n^{-100}}^{4\sqrt{s}} \frac{1}{\epsilon}\, d\epsilon$$

$$= C\sqrt{s}\log n \; \log n \Big|_{n^{-100}}^{4\sqrt{s}}$$

$$= C\sqrt{s}\log n \; \log(n^{100} 4\sqrt{s})$$

$$\leq C\sqrt{s}\log^2(n)$$

$\boxed{\text{QED}}$