Recall

- Norm Concentration
  - Random matrix $A \in \mathbb{R}^{m \times n}$

    (Independent, mean zero, isotropic, sub-gaussian rows)

    Fix $x \in \mathbb{R}^n$, $\quad \mathbb{E}\, \|Ax\| \approx \sqrt{m} \cdot \|x\|$

    $Ax = \begin{pmatrix} \vdots \\ A_i^T x \\ \vdots \end{pmatrix} \in \mathbb{R}^m$

- Standard gaussian $g \sim N(0, I_n)$

  $\mathbb{E}\, \|g\| \sim \sqrt{n}$

- JL Lemma

  Rand Matrix $A \in \mathbb{R}^{m \times n}$. Finite set $T \subset \mathbb{R}^n$

$$\mathbb{E}\left[\sup_{z \in T} \Big| \|Az\| - \sqrt{m}\|z\| \Big| \right] \leq C K^2 \sqrt{\log |T|}\ \mathrm{rad}(T)$$

Matrix Deviation Inequality

THM: Let A be an $m \times n$ matrix whose rows A_i are ind, isotropic, and sub-gaussian random vectors in $\mathbb{R}^n$. Then, for <u>any</u> subset $T \subset \mathbb{R}^n$, we have

$$\mathbb{E} \sup_{x \in T} \Big| \|Ax\| - \sqrt{m}\|x\| \Big| \leq C K^2 \gamma(T).$$

Here, $\gamma(T)$ is the gaussian width of T and $K = \max_i \|A_i\|_{\psi_2}$.

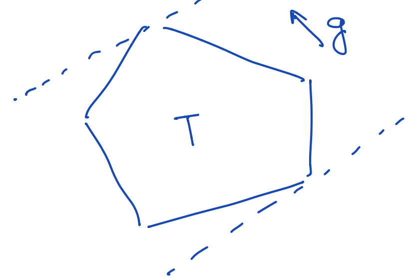
[C.f. Chapter 9, Vershynin's Book

Proof via Talagrand's Comparison Ineq.]

Defn: The gaussian width of a set $T \subset \mathbb{R}^n$ is defined as

$$\gamma(T) := \mathbb{E}_g \sup_{x \in T} |\langle g, x \rangle|.$$

$g \sim N(0, I_n)$.



Examples

① Unit Sphere S^{n-1} (Unit Ball B_2^n)

$$\gamma(S^{n-1}) := \mathbb{E} \sup_{x \in S^{n-1}} |\langle g, x\rangle|$$

$$= \mathbb{E}\, \|g\| = \sqrt{n} \pm c$$

Any $u \in \mathbb{R}^n$

$\sup_{x \in S^{n-1}} |\langle u, x\rangle|$

$= u^T\left(\frac{u}{\|u\|}\right)$ ✓

$= \frac{u^T u}{\|u\|}$

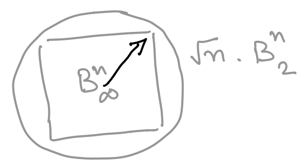
$= \frac{\|u\|^2}{\|u\|} = \|u\|$

② Cube, $B_\infty^n := [-1, 1]^n$ (unit ball wrt ℓ_∞ norm)

$g \sim N(0, I_n)$

$$\gamma(B_\infty^n) = \mathbb{E} \sup_{x \in B_\infty^n} |\langle g, x\rangle|$$

$$= \mathbb{E}\, \|g\|_1 = n\, \mathbb{E}\, |g_1| = n \cdot \sqrt{\frac{2}{\pi}}$$

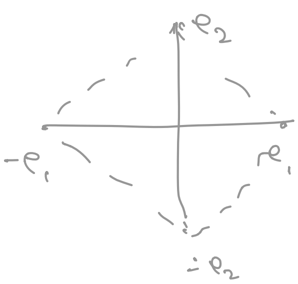


③ ℓ_1 ball $B_1^n := \{x \in \mathbb{R}^n : \|x\|_1 \le 1\}$

$\vec{\mathbb{1}} \in B_\infty^n$

$\|\vec{\mathbb{1}}\|_2 = \sqrt{n}$

$$\gamma(B_1^n) = \mathbb{E} \sup_{x \in B_1^n} |\langle g, x\rangle|$$



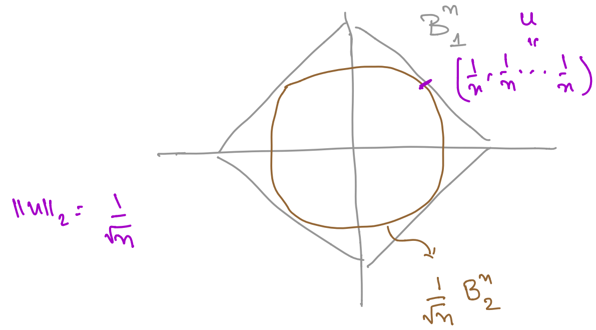
$$= \mathbb{E}\, \|g\|_\infty = \mathbb{E} \max_i |g_i|$$

$$\approx \sqrt{\log n}$$

④ Finite Point Set $T \subset \mathbb{R}^n$

$$\gamma(T) \leq C \cdot \sqrt{\log |T|} \cdot \mathrm{rad}(T)$$

radius $\left[\mathrm{rad}(T) := \max_{x \in T} \|x\|_2 \right]$



- Upto a log factor the gaussian width of these two sets are the same.
- The ℓ_1 ball, B_1^n, has only $2n$ vertices. The bulk of B_1^n lies within the inscribed ball
- Similar obsv holds even from a volumetric perspective.

Properties of Gaussian Width $\gamma(T)$

1. If T is bounded, then $\gamma(T)$ is finite

2. U be any orthogonal matrix, then
$$\gamma(UT) = \gamma(T)$$

[Fix orthogonal U. For $g \sim N(0, I_n)$, $Ug \sim N(0, I_n)$]

Rotational Invariance of Gaussians

3. Gaussian width is invariant under taking convex hulls
$$\gamma(\mathrm{conv}(T)) = \gamma(T)$$

4. Gaussian width respects Minkowski sum of sets & scaling.
For any $T, S \subset \mathbb{R}^n$ and $a \in \mathbb{R}$ we have
$$\gamma(T+S) \leq \gamma(T) + \gamma(S)$$
$$\gamma(aT) = |a| \, \gamma(T)$$

Minkowski sum

$T+S := \{ t+s : t \in T \text{ and } s \in S \}$

$aT := \{ at : t \in T \}$

5. Gaussian width and radius (diameter)

For any $T \subset \mathbb{R}^n$

$$\sqrt{\frac{2}{\pi}}\ \mathrm{rad}(T) \le \gamma(T) \le \sqrt{n}\cdot \mathrm{rad}(T)$$

$\lceil$ $\mathrm{rad}(T) := \sup_{x \in T} \|x\|_2$ $\rfloor$

$\mathrm{diam}(T) := \sup_{x,y \in T} \|x-y\|$

$$\gamma(T) = \mathbb{E} \sup_{x \in T} |\langle g, x\rangle|$$

$$\ge \sup_{x \in T} \left(\mathbb{E}|\langle g, x\rangle| \right)$$

$\lceil$ Jensen's Ineq $\rfloor$

$$= \sup_{x \in T} \left(\sqrt{\frac{2}{\pi}} \cdot \|x\| \right)$$

$$= \sqrt{\frac{2}{\pi}} \cdot \mathrm{rad}(T)$$

Fix x

$\langle g, x\rangle \sim N(0, \|x\|^2)$

$\mathbb{E}|\langle g, x\rangle| = \sqrt{\frac{2}{\pi}} \cdot \|x\|$

$\langle g, x\rangle = \sum_i g_i x_i$

Upper Bound

$$\gamma(T) = \mathbb{E} \sup_{x \in T} |\langle g, x\rangle|$$

$$\le \mathbb{E} \sup_{x \in T} \|g\| \cdot \|x\|$$

$\lceil$ Cauchy Schwarz $\rfloor$

$$= \mathrm{rad}(T) \cdot \mathbb{E}\|g\|$$

$$\approx \mathrm{rad}(T)\sqrt{n}$$

Matrix Deviation Ineq: Tail Bound

THM: Under the conditions of the previous theorem, for any $t \geq 0$, the event

$$\sup_{x \in T} \Big| \|Ax\| - \sqrt{m}\, \|x\| \Big| \leq CK^2 \left(\gamma(T) + t \, \mathrm{rad}(T) \right)$$

holds with prob. at least $1 - 2\exp(-t^2)$

Applications of Matrix Deviation Inequality

- JL lemma

For finite $T \subset S^{n-1}$
$\gamma(T) = \sqrt{\log |T|}$

- Spectra of Random Matrices.

THM: Let A be a $m \times n$ matrix whose rows A_i are ind, mean zero, isotropic, sub-gaussian rand. vectors in $\mathbb{R}^n$. Then, with high prob., we have

$$\sqrt{m} - CK^2\sqrt{n} \leq \sigma_1(A) \leq \sqrt{m} + CK^2\sqrt{n}$$

$\sigma_1(A)$ = largest singular value of A.

$\|A\| = \sigma_1(A)$ Spectral Norm.

$= \max_{x \in S^{n-1}} \|Ax\|$

Largest Eigenvalue if A is symmetric

$A = \begin{array}{cc} & n \\ m & \begin{bmatrix} \; \end{bmatrix} \end{array}$

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

Proof: spectral norm

$\sigma_1(A) = \max_{x \in S^{n-1}} \|Ax\|$

MDI over the sphere.

$$\mathbb{E} \sup_{x \in S^{n-1}} \left| \|Ax\| - \sqrt{m} \cdot 1 \right| \leq CK^2 \cdot \sqrt{n}$$

$\|x\|$ $\qquad$ $\gamma(S^{n-1})$

Markov's ineq. ensures that, with const. prob.,

$$\sqrt{m} - CK^2 \sqrt{n} \leq \sup_{x \in S^{n-1}} \|Ax\| \leq \sqrt{m} + CK^2 \sqrt{n}$$

Exercise

Under the conditions of the previous thm.

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|^2 - m \|x\|^2 \right| \leq CK^4 \gamma(T)^2 + CK^2 \sqrt{m}\, \mathrm{rad}(T)\, \gamma(T)$$

$\ulcorner a^2 - b^2 = (a+b)(a-b) \lrcorner$

Covariance Estimation

Given $X_1, X_2 \ldots X_m \in \mathbb{R}^n$ sampled from an unknown distribution given by rand. vector $X \in \mathbb{R}^n$

$X_1, \ldots, X_m, X$ are iid

$\mathbb{E} X = 0$

Principal Component Analysis (PCA)

- Find the singular vectors of $\mathbb{E} XX^T$ the covariance matrix (eigen vectors)

$$\Sigma := \mathbb{E}\, XX^T$$

$\ulcorner$ If $\mathbb{E}X \neq 0$ $\mathbb{E}XX^T$ is called the second moment matrix $\lrcorner$

- Estimate Σ from data by sample covariance matrix

$$\Sigma_m := \frac{1}{m} \sum_{i=1}^{m} X_i X_i^T$$

Law of Large Numbers

$$\Sigma_m \to \Sigma \quad \text{a.s.} \quad (m \to \infty)$$

Quantitative Question: How many samples, m, are required to get a "good enough" estimate of Σ.

Linear suffices $m \sim n$

Linear is necessary as well – dimension counting

THM: Let X be a sub-gaussian random vector in $\mathbb{R}^n$. $\mathbb{E}X = 0$
Then, for all $m > n$,

$$\mathbb{E}\, \|\Sigma_m - \Sigma\| \le CK^2 \sqrt{\frac{n}{m}} \cdot \|\Sigma\|$$

Proof:

Actual Mean Σ
Empirical Mean Σ_m

For analysis, bring X_i-s to isotropic position

Let Z satisfy

$$X = \Sigma^{1/2} \cdot Z.$$
$$X_i = \Sigma^{1/2} \cdot Z_i$$

Since covariance matrix Σ is PSD, there exists symmetric matrix $\Sigma^{1/2}$ such that $\Sigma = \Sigma^{1/2} \cdot \Sigma^{1/2}$.
$\Sigma^{1/2}$ = Square Root of Σ

$$\mathbb{E}\, ZZ^T = \mathbb{E}\, \Sigma^{-1/2} \cdot XX^T (\Sigma^{-1/2})^T = \Sigma^{-1/2} \cdot \Sigma \cdot \Sigma^{-1/2} = I_n$$

$$\mathbb{E}\, \|\Sigma_m - \Sigma\| = \mathbb{E}\, \left\| \frac{1}{m} \sum_i X_i X_i^T - \Sigma^{1/2} \cdot \Sigma^{1/2} \right\|$$

$$= \mathbb{E}\, \left\| \frac{1}{m} \sum_i \Sigma^{1/2} Z_i Z_i^T \Sigma^{1/2} - \Sigma^{1/2} \cdot \Sigma^{1/2} \right\|$$

$$= \mathbb{E}\, \left\| \Sigma^{1/2} \Big(\underbrace{\frac{1}{m} \sum_i Z_i Z_i^T - I_n}_{R_m} \Big) \Sigma^{1/2} \right\|$$

$$= \mathbb{E}\, \| \Sigma^{1/2} \cdot R_m \Sigma^{1/2} \|$$

$$y^T (\Sigma^{1/2} R_m \Sigma^{1/2}) y = y^T (\Sigma^{1/2})^T R_m (\Sigma^{1/2} y) = (\Sigma^{1/2} y)^T R_m (\Sigma^{1/2} y)$$

$$= \mathbb{E} \max_{y \in S^{n-1}} \left| \langle \Sigma^{1/2} \cdot R_m \Sigma^{1/2} y, y \rangle \right|$$

$$= \mathbb{E} \max_{x \in T} \left| \langle R_m x, x \rangle \right|$$

Ellipse, $T := \Sigma^{1/2} \cdot S^{n-1}$

$x = \Sigma^{1/2} \cdot y$

$$= \mathbb{E} \max_{x \in T} \left| \frac{1}{m} \sum_i \langle Z_i, x \rangle^2 - \|x\|^2 \right|$$

$$= \frac{1}{m} \mathbb{E} \max_{x \in T} \left| \sum_i \langle z_i, x \rangle^2 - m \cdot \|x\|^2 \right|$$

let

$$A := \begin{pmatrix} - z_1 - \\ - z_2 - \\ \vdots \\ - z_m - \end{pmatrix}$$

$$= \frac{1}{m} \mathbb{E} \max_{x \in T} \left| \|Ax\|^2 - m \|x\|^2 \right|$$

$$\underset{(Ex.)}{\leq} \frac{1}{m} \left[\gamma(T)^2 + 2\sqrt{m}\, \mathrm{rad}(T) \cdot \gamma(T) \right]$$

$T := \Sigma^{1/2} \cdot S^{n-1}$

$\mathrm{rad}(T) = \max_{x \in T} \|x\|_2$

$= \max_{y \in S^{n-1}} \|\Sigma^{1/2} \cdot y\|_2$

$= \|\Sigma^{1/2}\| = \|\Sigma\|^{1/2}$

$\gamma(T) := (\mathrm{tr}(\Sigma))^{1/2}$

trace = sum of eigen values

$$\mathbb{E} \|\Sigma_m - \Sigma\| \leq \frac{1}{m} \left(\mathrm{tr}(\Sigma) + 2\sqrt{m} \cdot \sqrt{\|\Sigma\|} \cdot \sqrt{\mathrm{tr}(\Sigma)} \right)$$

$$\leq \frac{1}{m} \left(n \cdot \|\Sigma\| + 2\sqrt{m}\, \sqrt{n} \cdot \|\Sigma\| \right)$$

$$= \left(\frac{n}{m} + 2\sqrt{\frac{n}{m}} \right) \|\Sigma\|$$

$$\leq 3\sqrt{\frac{n}{m}} \cdot \|\Sigma\|$$

Since $m \geq n$

$\sqrt{\frac{n}{m}}$ dominates

$\blacksquare$

While the Covariance Estimate Thm can be derived via simpler arguments (ε-net argument)

The following improvement follows from MDI.

Low Rank Σ

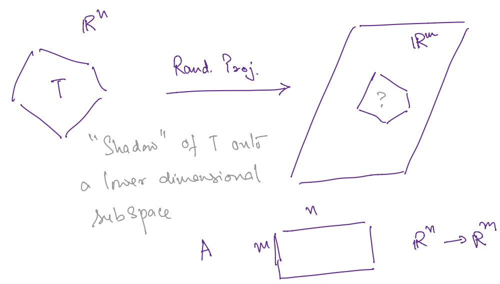
$r << n$

lower dimensional distributions require fewer samples.

$$\mathbb{E}\, \| \Sigma_m - \Sigma \| \lesssim \sqrt{\frac{r}{m}} \cdot \|\Sigma\|$$

Here, r is the effective rank of Σ $\qquad r := \dfrac{\operatorname{tr} \Sigma}{\|\Sigma\|}$

Application ③ : Random Projection of Sets



$T \subset \mathbb{R}^n$

$$\operatorname{diam}(T) := \max_{x, y \in T} \|x - y\|_2$$

If A is a gaussian rand. matrix, then the proj. will be uniform on a rand. m-dim subspace. Grassmanian Manifold.

MDI for $T - T$ and triangle inequality

$$\mathbb{E} \underbrace{\sup_{x \in T-T} \|Ax\|_2}_{\text{Diameter of } AT} \le \underbrace{\sup_{x \in T-T} \sqrt{m} \cdot \|x\|}_{\text{Diameter of } T} + CK^2 \gamma(T-T)$$

$$\mathbb{E} \operatorname{diam}(AT) \le \sqrt{m} \operatorname{diam}(T) + 2CK^2 \gamma(T)$$

Write $P := \frac{1}{\sqrt{m}} A$

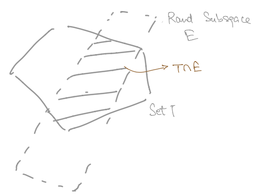
$$\mathbb{E}\ \mathrm{diam}(PT) \leq \sqrt{\frac{m}{n}} \cdot \mathrm{diam}(T) + 2CK^2 \frac{\gamma(T)}{\sqrt{n}}$$

Application (4) Random Section of Sets

Considering diameter of $T \subset \mathbb{R}^n$ intersected with random subspace.

E – Random Subspace of co-dimension m

$\dim(E) = n - m$



$$\frac{\gamma(T)}{\sqrt{n - \dim(E)}}$$

THM (M* bound) (Milman) $\mathbb{E}\ \mathrm{diam}(T \cap E) \leq C \frac{\gamma(T)}{\sqrt{m}}$

Proof:

Random Gaussian Matrix $A \in \mathbb{R}^{m \times n}$ $\quad E := \mathrm{kernel}(A)$

$\dim(E) = n - m$

$\mathrm{Codim}(E) = m$

Consider $T - T$

MDI: $$\mathbb{E} \sup_{x \in T-T} \left| \|Ax\|_2 - \sqrt{m}\|x\|_2 \right| \leq C\gamma(T)$$

$$\geq$$

$$\mathbb{E} \sup_{x \in (T-T)\cap E} \left| \|Ax\|_2 - \sqrt{m}\cdot\|x\| \right| \leq C\gamma(T)$$

Sup over a smaller set

$$\mathbb{E} \sup_{x \in (T-T)\cap E} \left| 0 - \sqrt{m}\|x\| \right| \leq C\gamma(T)$$

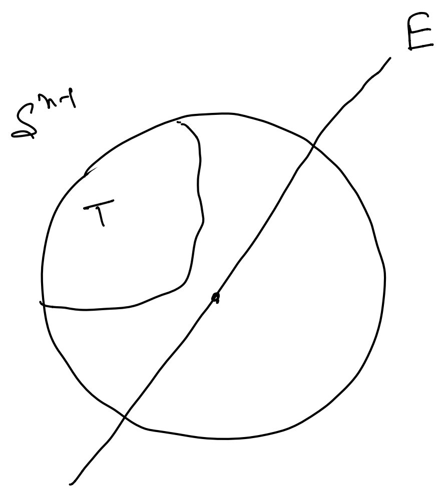
$Ax = 0$ for all $x \in E = \ker(A)$

$$\mathbb{E}\ \operatorname{diam}(T \cap E) \leq c\, \frac{\gamma(T)}{\sqrt{m}}$$

$\square$

Application ⑤ Escape Theorem

Want subspace E to completely avoid $T \subset S^{n-1}$



THM: (Gordon) Let $T \subset S^{n-1}$ be a set of unit vectors and E be a random subspace with $\operatorname{co-dim}(E) = m$.
If $\gamma(T) < c\sqrt{m}$, then, with high probability,

$$T \cap E = \emptyset.$$

Proof: Rand matrix $A \in \mathbb{R}^{m \times n}$

$E = \text{kernel}(A)$

MDI

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\| - \sqrt{m} \cdot 1 \right| \leq C K^2 \gamma(T)$$

With $\sqrt{m} > 2 C K^2 \gamma(T)$

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|_2 - \sqrt{m} \right| < \frac{\sqrt{m}}{2}$$

Event $\mathcal{E}$: $Ax > 0$ for $\underline{\underline{\text{all}}}$ $x \in T$

$\Updownarrow$

$T \cap E = \underline{\emptyset}$

$\ulcorner E = \ker(A) \lrcorner$

Conditioned on $\mathcal{E}^c$,

$$\sup_{x \in T} \left| \underset{\underset{0}{\parallel}}{\|Ax\|} - \sqrt{m} \right| = \sqrt{m}$$

Therefore,

$$\mathbb{P}_r \{ T \cap E = \emptyset \} \geq 1/2 .$$

$\boxed{\checkmark}$

Applications of Matrix Deviation Ineq. (MDI)

0. JL Lemma
1. Spectra of Random Matrices
2. Covariance Estimation
3. Random Projection of Sets
4. Random Section of Sets (M^* Bound)

5. Escape Theorem (Gordon)
6. Compressed Sensing

} Today

7. Community Detection

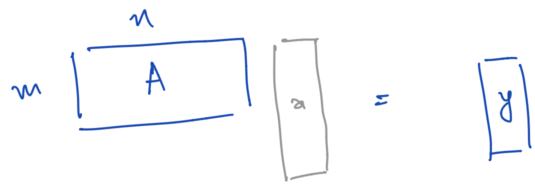
} Next

Compressed Sensing

Efficiently acquiring & reconstructing a signal, by finding solutions to underdetermined linear systems.

Given A & y

Find x



$m \ll n$

Assumption: x is s-sparse

$$\|x\|_0 := |\{ i \; : \; x_i \neq 0 \}|$$

$$s = \|x\|_0 \ll n$$

Supp(x) is unknown.

[In the worst case the problem is NP-Hard. Fix A & then try to find a sparse x.]

The problem is tractable when A is random.

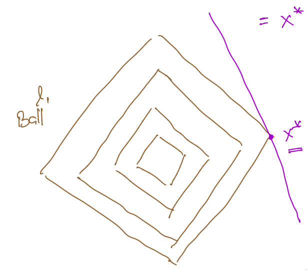
$$\min_{\hat{x}} \; \|\hat{x}\|_0 \quad \text{s.t} \quad A\hat{x} = y.$$

Proposed Method

LP: $\quad \min \|x'\|_1$

relaxation $\quad$ st. $\; Ax' = y$

$F = \{ \tilde{x} : A\tilde{x} = y \} = x^* + \ker(A)$



[If F is random, then likely that it will intersect with the ℓ_1 ball (blown up) at a vertex]

THM: [Compressed Sensing] Let A be a random $m \times n$ matrix and x be an s-sparse vector.
If $m \gtrsim s \log n$, then the solution to the LP, $\hat{x}$, is exact.

Want to show that the error $h := \hat{x} - x$ is equal to zero.

Lemma Write $S := \mathrm{Supp}(x) = \{ i \in [n] : x_i \neq 0 \}$.

$$\| h_{S^c} \|_1 \leq \| h_S \|_1$$

h_S & h_{S^c} : restriction of h on S and S^c, respectively

Pf: We have $\|\hat{x}\|_1 \leq \|x\|_1$. LP opt.

$$\| h + x \|_1 \leq \|x\|_1$$

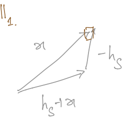
$$\| h_S + h_{S^c} + x \|_1 \leq \|x\|_1$$

$$\| h_S + x \|_1 + \| h_{S^c} \|_1 \leq \|x\|_1$$

Δ-ineq. $\geq$

$$\|x\|_1 - \|h_S\|_1 + \|h_{S^c}\|_1 \leq \|x\|_1$$

Therefore, $\| h_{S^c} \|_1 \leq \| h_S \|_1$.



COR: $\| h \|_1 \leq 2\sqrt{s} \, \| h \|_2$

Pf: $\| h \|_1 = \| h_S \|_1 + \| h_{S^c} \|_1$

$\leq 2 \| h_S \|_1$ (Lem)

$\leq 2\sqrt{s} \, \| h_S \|_2 \leq 2\sqrt{s} \, \| h \|_2$ (Cauchy Schwartz.) $\square$

Proof of Exact Recovery Thm.

Assume $h = \hat{x} - x \neq 0$

① $\frac{h}{\|h\|} \in \ker(A)$

$Ah = A\hat{x} - Ax = y - y = 0$

② Write $T := \{ z \in S^{n-1} : \| z \|_1 \leq 2\sqrt{s} \}$.

$\frac{h}{\|h\|_2} \in T$ [COR.]

$\gamma(T) \leq 2\sqrt{s} \cdot \gamma(B_1^n)$

$= 2\sqrt{s} \sqrt{\log n}$

$h \in \ker(A) \cap T$

However,

$m \gtrsim \gamma^2(T)$, and

in such a case, whp,

$\ker(A) \cap T = \emptyset$.

(Escape Thm.)

A contradiction.

$\square$