Course 3

I] OMP $\Leftrightarrow$ Exact Recovery Condition

$\hookrightarrow$ Notebook

II] Stability and Robustness

III] Restricted Isometry Property

I] High-Dimensional linear model:

$$y = Ax + \eta$$

$\|A_i\|_2^2 = 1$

$A \, \mathrm{Diag}(\lambda) \, \mathrm{Diag}(\tfrac{1}{\lambda}) x$

normalised

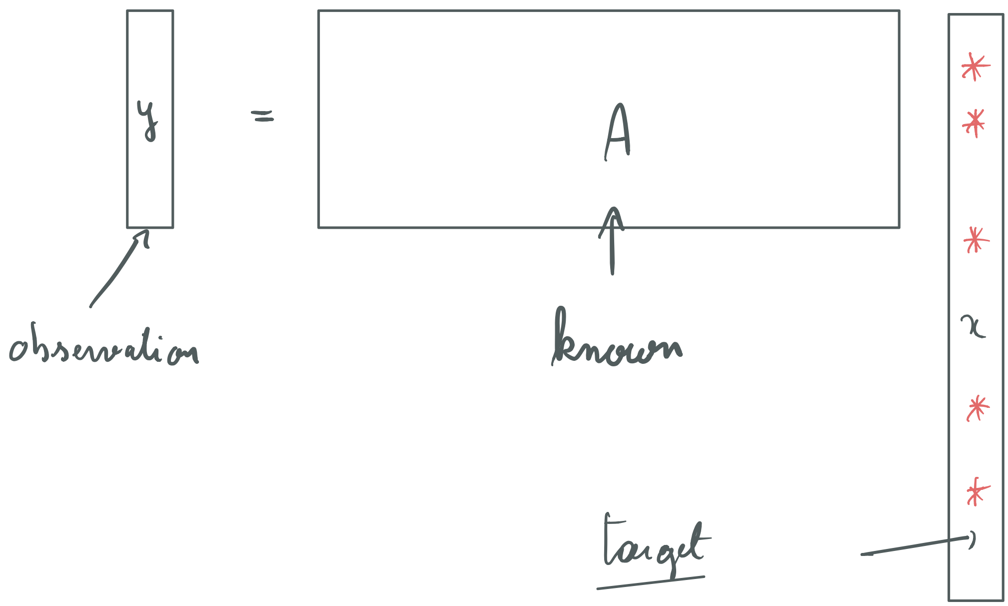
where $A \in \mathbb{R}^{m \times N}$

$m < N$

ill-posed

Sparsity: $x \in \Sigma_k$

where $\Sigma_k = \{ x \in \mathbb{R}^N : \|x\|_0 = k \}$



OMP

Input $A \in \mathbb{R}^{m \times N}$, $y \in \mathbb{R}^m$

Initialization $S^0 = \emptyset \qquad x^0 = 0$

Iteration $j_{m+1} = \underset{j \in [N]}{\arg\max} \left\{ \left| (A^T(y - Ax^m))_j \right| \right\}$

search

[Rk $\quad \left| \underbrace{A^T(y - Ax^n)}_{\text{residuals } \in \mathbb{R}^N}{}_{j_{n+1}} \right| = \| A^T(y - Ax^n) \|_\infty$]

$$S^{n+1} = S^n \cup \{ j_{n+1} \}$$

support update

$$x^{n+1} = \arg\min_{\substack{z \in \mathbb{R}^N \\ \mathrm{Supp}(z) = S^{n+1}}} \left\{ \| y - Az \|_2^2 \right\}$$

solution update

__Lemma__ Given $S \subset [N]$

__If__ $\quad v = \arg\min_{\substack{z \in \mathbb{R}^N \\ \mathrm{Supp}(z) = S}} \left\{ \| y - Az \|_2^2 \right\}$

__Then__ $\cdot \quad \underbrace{A_S^T(y - A_S v)}_{\text{residuals on } S} = 0 \qquad (Av = A_S v)$

• $v = (A_S^T A_S)^{-1} A_S^T y$

(Ordinary Least Squares)

A normalized

<u>Proposition</u> : $\forall x \in \Sigma_k$ is recovered in k steps by OMP

$\iff$

• A_S is injective and

$\forall r \in \mathrm{Span}(A_S) \setminus \{0\}$

$\| A_S^T r \|_\infty > \| A_{S^c}^T r \|_\infty$

S complement.

$r \in \{ A z : \mathrm{Supp}(z) \subseteq S \} \setminus \{0\}$

$\forall S \subset [N]$ s.t $\# S = k$

$\Leftrightarrow$ • $\forall S \subset [N]$ s.t. $\# S = k$

$$\cdot A_S^+ := \underline{(A_S^T A_S)^{-1}} A_S^T$$ exists (A_S invertible)

$\forall u \in \mathbb{R}^k \setminus \{0\}$,

$$\boxed{\| A_S^T A_S u \|_\infty > \| A_{S^c}^T A_S u \|_\infty} \quad (*)$$

Put $v = A_S^T A_S u$

so that $u = (A_S^T A_S)^{-1} v$

$(*) \Leftrightarrow \| v \|_\infty > \| A_{S^c}^T A_S u \|_\infty$

$\Leftrightarrow \| v \|_\infty > \| A_{S^c}^T A_S (A_S^T A_S)^{-1} v \|_\infty$

$$\Leftrightarrow \| A_{S^c}^T \underbrace{A_S (A_S^T A_S)^{-1}}_{(A_S^+)^T} \|_{\infty \to \infty} < 1$$

$\Leftrightarrow \| (A_S^+) A_{S^c} \|_{1 \to 1} < 1$

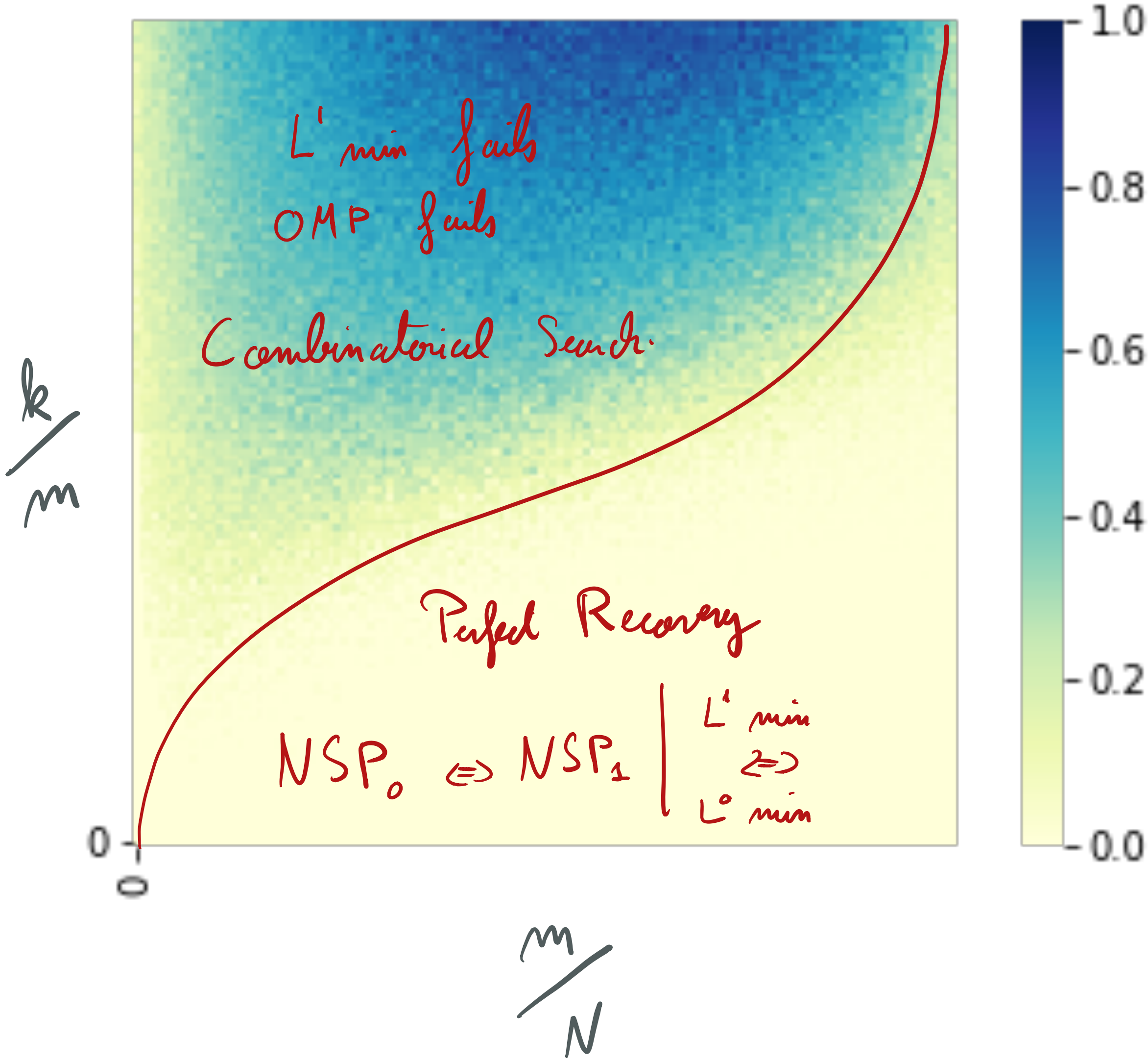
$$\boxed{\underline{ERC}: \quad \left\| \left(A_S^{\dagger} \right) A_{S^c} \right\|_{1\to 1} < 1 \qquad \forall S \subset [N],\ \# S = k}$$

$$\underline{NSC}: \quad \left\| \left(A_S^T A_S \right)^{-1} A_S^T A_{S^c} \right\|_{1\to 1} < 1$$

$$\Updownarrow$$

$$\max_{l \in S^c} \left\| \underbrace{\left(A_S^T A_S \right)^{-1} A_S^T A_l}_{\in \mathbb{R}^k} \right\|_1 < 1$$

Coordinates of Projection (A_l) onto Span (A_S) w.r.t A_S.



True Random A matrices

"$A_{ij} \overset{iid}{\sim} \mathcal{N}(0, 1/m_{rows})$"

- ## Coherence Property

Def: A with normalized columns

$$\mu = \mu(A) = \max_{1 \leq i \neq j \leq N} |\langle A_i, A_j \rangle|$$

Rk: A normalized

$$A^T A - Id_N = \left(\delta_{i \neq j} \langle A_i, A_j \rangle \right)_{i,j}$$

Def: ℓ_1-coherence function

$$\mu_1(k) = \max_{i \in [N]} \max \left\{ \sum_{j \in S} |\langle A_i, A_j \rangle| , \ S \subset [N], \ \# S = k, \ i \notin S \right\}$$

$k \in [N-1]$

Note that: $\forall k \in [N-1]$,

$$\boxed{\mu \leq \mu_1(k) \leq k\mu}$$

<u>Theorem</u> (Coherence – RIP)

- A normalized
- $\forall x \in \Sigma_{\underset{2k}{k}}$,

$$\boxed{(1-\mu_1(k-1))\,\|x\|_2^2 \leq \|Ax\|_2^2 \leq \cdots \quad \cdots (1+\mu_1(k-1))\,\|x\|_2^2}$$

- <u>If</u> $\underset{2k-1}{\mu_1(k-1) < 1}$ <u>Then</u> $RIP_{\underset{2k}{k}}$ holds

<u>Proof</u>: • Let $S \subset [N]$ be s.t $\# S = k$

• Let $x \in \Sigma_k$ be s.t $\mathrm{Supp}\, x = S$

$$\|Ax\|_2^2 = x_S^T A_S^T A_S x_S$$

(indeed $Ax = A_S x_S$)

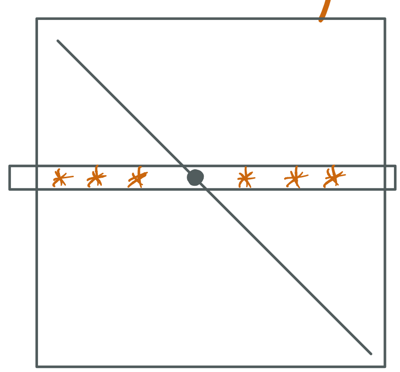
$$\lambda_{max} = \max_{\substack{x \in \mathbb{R}^N \\ \mathrm{Supp}\, x \subset S \\ \|x\|_2 = 1}} \left\{ \langle A_S^T A_S x_S, x_S \rangle \right\}$$

$$\lambda_{min} = \min_{\substack{x \in \mathbb{R}^N \\ \mathrm{Supp}\, x \subset S \\ \|x\|_2 = 1}} \left\{ \langle A_S^T A_S x_S, x_S \rangle \right\}$$

Note that:

- $(A_S^T A_S)_{ii} = 1$

- $r_j = \sum_{l \in S,\, l \neq j} \underbrace{|(A_S^T A_S)_{j,l}|}_{*}$

j  $= A_S^T A_S$

$$r_j = \sum_{l \in S,\, l \neq j} |\langle A_l, A_j \rangle| \leq \mu_1 (k-1)$$

$$\forall j \in S$$

By Gershgorin's disk theorem

$$\{\text{eigenvalues}\} \subset [1 - \mu_1(k-1), 1 + \mu_1(k-1)]$$

Theorem A is normalized

If $\boxed{\underline{\mu_1(k) + \mu_1(k-1) < 1}}$ Coherence Condition $\quad CC_k$

then ERC_k holds

Rk: $\cdot \ \mu_1(2k-1) \leq \mu_1(k) + \mu_1(k-1)$

$CC_k \Rightarrow \mu_1(2k-1) < 1 \Rightarrow RIP_{2k}$

$\cdot \ CC_k \Rightarrow ERC_k$

Proof: $CC_k \Rightarrow ERC_k$

We need to prove that:

- A_S injective $\Big)$ implied by $\mu_1(2k-1) < 1 \Rightarrow RIP_{2k}$

- $\| A_S^T r \|_\infty > \| A_{S^c}^T r \|_\infty$

  where $r = A_S z$ with $z \neq 0$

- Let $z \neq 0$, set $r = A_S z$, and chose:

$$\ell \in S \quad \text{s.t.} \quad |z_\ell| = \|z\|_\infty$$

- Note that for $j \in S^c$

$$\begin{aligned} |\langle A_j, r \rangle| &= \Big| \sum_{i \in S} z_i \langle A_i, A_j \rangle \Big| \\ &\leq \sum_{i \in S} |z_i| \, |\langle A_i, A_j \rangle| \\ &\leq |z_\ell| \, \mu_1(k) \\ &\lesseqgtr \end{aligned}$$

· And for $i \in S$,

$$|\langle A_{\ell}, r \rangle| = \Big| \sum_{j \in S} z_j \langle A_{\ell}, A_j \rangle \Big|$$

$$\geq |z_\ell| \, |\langle A_{\ell}, A_{\ell} \rangle| - \sum_{\substack{j \neq \ell \\ j \in S}} |z_j| \, |\langle A_j, A_\ell \rangle|$$

$$\geq |z_\ell| - |z_\ell| \, \mu_1 (k-1)$$

$$= |z_\ell| \, \big(\underbrace{1 - \mu_1 (k-1)}_{> \mu_1 (k)} \big)$$

$$\|A_S^T r\|_\infty \geq |\langle A_\ell, r \rangle| \geq |z_\ell| \, (1 - \mu_1 (k-1))$$

$$> |z_\ell| \, \mu_1 (k)$$

$$\geq \|A_{S^c}^T r\|_\infty \qquad \square$$

$$CC_k \Rightarrow RIP_k$$

$$CC_k \Rightarrow ERC_k$$

Proposition $ERC_k \Rightarrow NSP_1$

Proof: $v \in \ker A \setminus \{0\}$

$$A_S v_S = - A_{S^c} v_{S^c}$$

$$\|v_S\|_1 = \| A_S^+ A_S v_S \|_1 = \| A_S^+ A_{S^c} v_{S^c} \|_1$$

(A_S injective, $A_S^+ A_S = (A_S^\top A_S)^{-1} A_S^\top A_S = Id$)

$$\leq \underbrace{\| A_S^+ A_{S^c} \|_{1 \to 1}}_{< 1} \| v_{S^c} \|_1 < \| v_{S^c} \|_1$$

$\square$

II) Stability and robustness

<u>Stability</u> : x is not sparse

<u>Robustness</u> : η is not zero

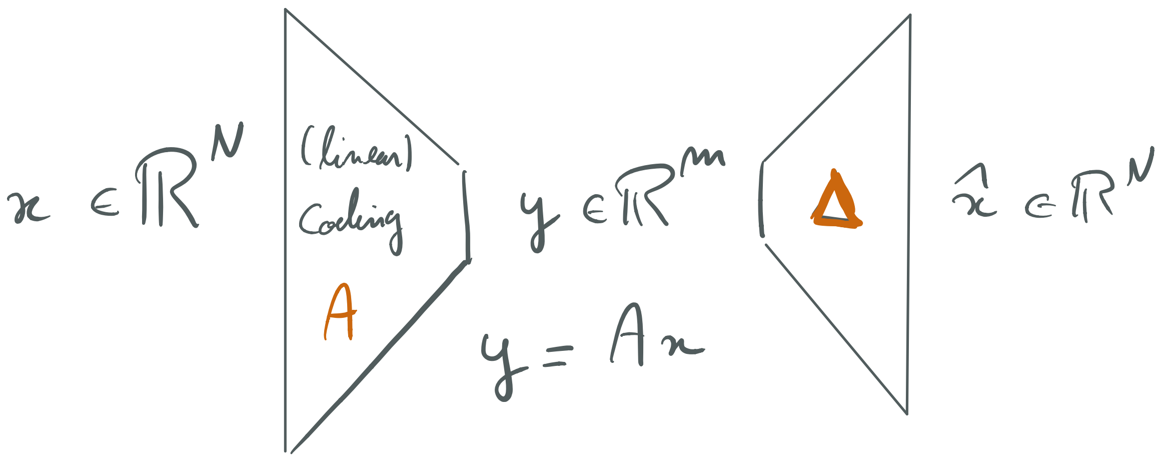
Instance Optimality

$$\boxed{\sigma_k(x)_p := \inf \left\{ \|x-z\|_p : z \in \Sigma_k \right\}}$$

- $k \in [N]$ sparsity
- $p \in [1, \infty]$ l_p-norm

<u>Rk</u>: $x \in \Sigma_k$, $\sigma_k(x)_p = 0$

Coding - Decoding Scheme



$\Delta : \mathbb{R}^m \longrightarrow \mathbb{R}^N$ <u>decoder</u>

<u>Def</u> ℓ_p- instance optimality <u>of order k</u>

We say that Δ is

ℓ_p- instance optimal of order k

iff $\exists C > 0, \forall x \in \mathbb{R}^N,$

$$\|x - \underbrace{\Delta(Ax)}\|_p \leq C\, \sigma_k(x)_p$$

Theorem Let $A \in \mathbb{R}^{m \times N}$ be given.

If $\exists \Delta$ s.t (A, Δ) is l_1-instance optimal of order k with constant $C > 0$

then

$$\boxed{\begin{array}{l} \forall v \in \ker A, \\ \|v\|_1 \leq C\, \sigma_{2k}(v) \qquad (**) \end{array}}$$

($C = 2$ is the NSP_1)

<u>Conversely</u> <u>if</u> $(**)$ holds

<u>then</u> $\exists \Delta$ s.t (A, Δ) is l_1-instance optimal of order k and constant $\underline{2C}$.

<u>Proof</u>: Let $v \in \ker A$
Let S be an index of the k largest entries of v

Inst. Opt $\Rightarrow$ $-v_S = \Delta(\underbrace{A(-v_S)}_{A v_{S^c}}) = \Delta(A v_{S^c})$

$$\| v \|_1 = \| v_S + v_{S^c} \|_1 = \| v_{S^c} - \Delta(A v_{S^c}) \|_1$$

$$\leq C \underbrace{\sigma_k (v_{S^c})_1}_{\text{sum on the } N-2k \text{ least abs. val coeff}} = C \sigma_{2k}(v)_1$$

Conversely, Assume (**) condition.

Define Δ as:

$$\boxed{\underbrace{\Delta(y)}_{z^*} = \arg\min_{\substack{z:\\ Az=y}} \left\{ \sigma_k(z)_1 \right\}}$$

$y = Ax$

x is feasible

Let $x \in \mathbb{R}^N$, (**) with $v = x - \Delta(Ax) \in \ker A$

$$\Big(Av = Ax - A(\underbrace{\Delta(Ax)}_{z^*}) = 0 \Big)$$

$$\underbrace{\qquad\qquad}_{Ax}$$

$$\|x - \Delta(Ax)\|_2 \leq C\, \sigma_{2k}(x - \Delta(Ax))_1$$

$$\leq C \left[\sigma_k(x)_1 + \underbrace{\sigma_k(\Delta(Ax))_1}_{\leq \sigma_k(x)_1} \right]$$

(x feasible)

$$\leq 2C\, \sigma_k(x)_1 \qquad \square$$

Theorem If (A, Δ) are ℓ_1-instance optimal of order k and constant $C>0$

then
$$\boxed{m \geq c\, k \ln(eN/k)}$$

where c depends only on C.

Def Stable NSP with constant $0<\rho<1$

iff $\forall v \in \ker A$, $\|v_S\|_1 \leq \rho \|v_{\bar{S}}\|_1$
$\forall S \subset [N]$ s.t. $\#S = k$,

<u>Theorem</u> <u>Assume</u> A satisfies <u>Stable NSP</u> of order k and constant $0 < \rho < 1$

<u>Then</u> $\forall x \in \mathbb{R}^N$, any solution $\hat{x}$ of <u>Basis Pursuit</u>:

$$\hat{x} \in \arg\min_{Az = Ax} \|z\|_1$$

is such that

$$\boxed{\|x - \hat{x}\|_1 \leq \frac{2(1+\rho)}{1-\rho} \sigma_k(x)_1}$$