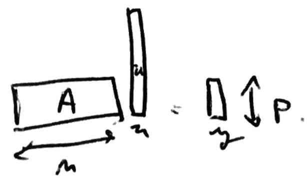
$\boxed{\text{LASSO}}$

Linear systems: $\underset{\mathbb{R}^{P\times n}}{A}\overset{\mathbb{R}^n}{x} = y \in \mathbb{R}^P$ **Undetermined**: $n \gg p$



Pbm: Recover $x_0 \in \mathbb{R}^n$ from $y = Ax_0 + \underset{\text{noise}}{\varepsilon}$ $\to$ ∞ number of sol$^{\underline{o}}$

Ker(A) large.

$\to$ needs a prior.

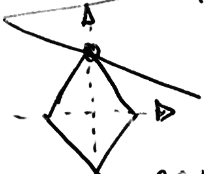
Exampls: ML regress$^{\circ}$: $(a_i, y_i)_i \rightsquigarrow y_i \approx \langle a_i, x\rangle$ $\hookrightarrow$ weight vector.

IP: A measurement device.

Regularized inversion: $\boxed{\min_x \frac{1}{2}\|y - Ax\|^2 + \lambda R(x) \quad (P_\lambda)} \xrightarrow[\lambda\to 0]{} \boxed{\min_{Ax=y} R(x). \quad (P_0)}$

$R = \|x\|^2 = \sum x_i^2$

$\boxed{\text{RIDGE}}$



favors sparse x_0.

$R(x) = \|x\|_1 = \sum_i |x_i|$.

$\boxed{\text{LASSO}}$

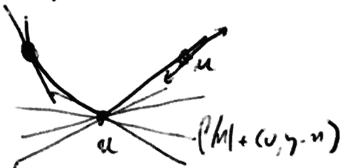
$\boxed{\text{exp}}$

$\boxed{\text{Fundamental Question: Condition on } x_0 \text{ and } A \text{ such that } x_0 = x^* \text{ when } \varepsilon = 0,\ x_0 \approx x_\lambda}$

$x_\lambda = (A^*A + \lambda \mathrm{Id})^{-1}A^*y \xrightarrow{\lambda\to 0} x^+ = A^+y = A^*(\underset{\in\mathbb{R}^{p\times p}}{AA^*})^{-1}y$ inv. if $\mathrm{Im}A = \mathbb{R}^P$

Subdifferential: $\partial f(x) = \{u \in \mathbb{R}^n : \forall y\ f(y) \geq f(x) + \langle u, y - x\rangle\}$

"Slope bellow" $\quad \partial f : \mathbb{R}^n \to 2^{\mathbb{R}^n}$.



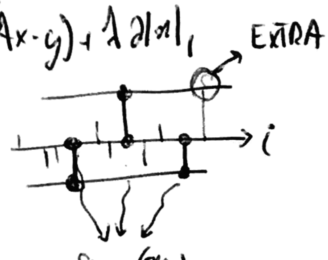
$\boxed{f \text{ diff at } x \iff \partial f(x) = \{\nabla f(x)\}}$ $\boxed{\operatorname{argmin} f = \{x : 0 \in \partial f(x)\}}$

$\partial|\cdot|(x) = \begin{cases} \mathrm{sign}(x) & x \neq 0 \\ [-1,1] & x = 0 \end{cases}$ $\partial\|\cdot\|_1(x) = \prod_{i=1}^n \partial|\cdot|(x_i) = \left\{u : \begin{array}{l}\|u\|_\infty \leq 1 \\ u_I = \mathrm{sign}(x_I)\end{array}\right\}$

$I \triangleq \mathrm{supp}(x)$.

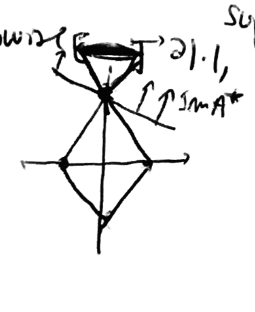
First order condition of lasso: $f(x) = \frac{1}{2}\|y - Ax\|^2 + \lambda\|x\|_1 \longrightarrow \partial f(x) = A^*(Ax - y) + \lambda\,\partial\|x\|_1$

<u>Prop</u>: x_λ sol$^{\underline{c}}$ (P_λ) $\iff \eta_\lambda \triangleq \frac{1}{\lambda}A^*(y - Ax) \in \partial\|x_\lambda\|_1 \iff \begin{cases}\|\eta_\lambda\| \leq 1 \\ \eta_{\lambda I} = \mathrm{sign}(x_\lambda)\end{cases}$



Noiseless: $\min \|x\|_1 + \iota_{\{Ax = y\}}(x)$ $\iota_C = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise}\end{cases}$

$\partial\|\cdot\|_1$ $\partial\iota^{\circ}_{\{Ax=y\}} = \mathrm{Im}A^* = \mathrm{Ker}(A)^\perp$



<u>Prop</u>: x^* with $Ax^* = y$ sol$^{\underline{o}}$ $(P_0) \iff \exists \eta \in \mathrm{Im}A^*,\ \eta \in \partial\|x^*\|_1$

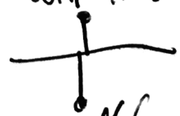
$\iff \eta_I = \mathrm{sign}(x_I)$ and $\|\eta\|_\infty \leq 1$

Compressed Sensing

Question: How to design apriori A such that $x_0 = x^\star$ (or $x_\lambda \approx x_0$) with p as small as possible when $s = |x_0|_0$ is small?

Bad idea: $A = [\,\cdot\ \cdot\ \cdot\,]$ sparse sampling: will "miss" many 1-sparse vectors

$A_{i,:} = \sim$ = low-pass Fourier: will "miss" very simple 2-sparse signals



Good idea: A_{ij} iid random. For simplicity $A_{ij} \sim \mathcal{N}(0, \frac{1}{\sqrt{p}}) \Rightarrow$ Thm becomes probabilistic!

Heuristic thm: if $p \gtrsim \begin{cases} s\log(n) \\ s\log\left(\frac{n}{p}\right) \end{cases}$ then $x^\star = x_0$ with high proba on A

FUCHS' method: Need to "build" $\eta \in \mathrm{Im}(A^*)$ such that $\begin{cases} \eta_I = \mathrm{sign}(x_{0,I}) \to \text{EASY} \\ |\eta|_\infty \leq 1 \to \text{HARD} \end{cases}$

$I \triangleq \mathrm{Supp}(x_0)$, $\eta = A^*\varphi$

FUCHS 2004: Solve $(A^*\varphi)_I = A_I^*\varphi = s_I \triangleq \mathrm{sign}(x_{0,I})$ in least square sense

$A_I = A(:, I) \in \mathbb{R}^{p\times s}$

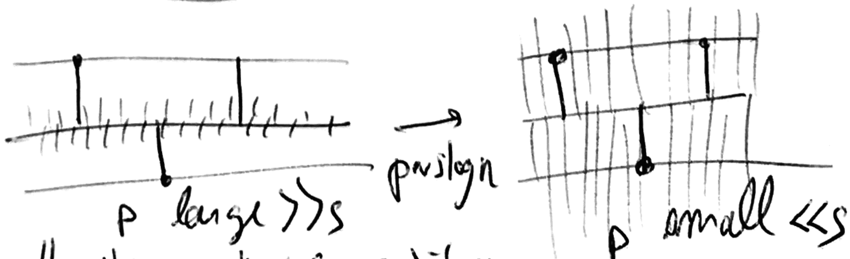
ie $\min_{\varphi\in\mathbb{R}^p} \{\|\varphi\|^2 : A_I^*\varphi = s_I\}$.

sol$^{\underline{\circ}}$: $\varphi_F = (A_I^*)^+ s_I = A_I (A_I^* A_I)^{-1} s_I$

DEF: Fuchs pre-certificate $\eta_F = A^*\varphi_F = A^* A_I \underbrace{(A_I^* A_I)^{-1} s_I}_{\alpha_I}$

Rmq: $C \triangleq A^*A$ cov$^{\circ}$, $\boxed{\eta_F = \sum_{i\in I} \alpha_i C_{\cdot i}}$

Fundamental Q: Is $|\eta_F|_\infty \leq 1$?



Heuristic proof: We have to show $\|\eta_F\| \leq 1$ ie $\|A_J^* A_I (A_I^* A_I)^{-1} s_I\|_\infty \leq 1$

$J \triangleq I^c$ ie $\max_j |\langle a_j, \underbrace{A_I (A_I^* A_I)^{-1} s_I}_{\varphi_F}\rangle| \leq 1$ INDEP!!

$\|\varphi_F\|_2^2 = \langle A_I (A_I^* A_I)^{-1} s_I, A_I (A_I^* A_I)^{-1} s_I\rangle = \langle (A_I^* A_I)^{-1} s_I, s_I\rangle$

Intuition: (TALAGRAND): if $p \sim s\log(s)$, $A_I^* A \approx I \Rightarrow \|\varphi_F\| \sim \sqrt{s}$

For $\boxed{\text{FIXED}}$ φ_F, $\langle a_j, \varphi_F\rangle \sim \mathcal{N}(0, \|\varphi_F\|^2/p)$, and the max of $P-s$ iid gaussians is very close to $\|\varphi_F\|/\sqrt{p} \times \sqrt{2\log(N-s)}$ so $\|\eta\|_\infty \leq 1$ implied by $\frac{s}{p} 2\log(N-s) \leq 1$ ie $\boxed{p \geq 2s\log N}$