Concentration Inequality

(a) $\text{Prob}[L \le \beta\mu] \le \beta^{k/2}\left(1 - \frac{(1-\beta)k}{d-k}\right)^{d-k/2}$, $\beta < 1$

Bernstein, Chernoff, inequalities

* Markov, * independence

Markov – Chebyshev Ineq.

** $X \ge 0$. $\text{Prob}[X \ge \mu] \le \frac{\mathbb{E}[X]}{\mu}$, $\mu > 0$

eg. " $\text{Prob}(X^2 \ge t] \le \frac{\mathbb{E}[X^2]}{t} = \frac{\text{Var}(X)}{t}$ $\mathbb{E}[X] = 0$ "

Proof of Lemma (a)

$L = \|z\|^2$ $z = \frac{(x_1 \cdots x_k, 0 \cdots 0)}{\|x\|}$

$X = (x_1 \cdots x_d)$, $x_i \sim N(0,1)$

$$\text{Prob}(L \le \beta\mu) = \text{Prob}\left\{\frac{x_1^2 + \cdots x_k^2}{x_1^2 + \cdots + x_d^2} \le \beta\mu\right\}$$

$$= \text{Prob}\left(\sum_{i=1}^{k} x_i^2 \le \beta\mu \sum_{i=1}^{d} x_i^2\right)$$

$$= \text{Prob}\left\{t\left(\beta\mu \sum_{i=1}^{d} x_i^2 - \sum_{i=1}^{k} x_i^2\right) \ge 0.\right\} \qquad t > 0$$

$$= \text{Prob}\left\{\exp\left[\underbrace{(t\beta\mu - 1)\sum_{i=1}^{k} x_i^2 + t\beta\mu \sum_{i=k+1}^{d} x_i^2}_{\ge 0}\right] \ge 1\right\}$$

Markov Ineq

$$\le \mathbb{E} \exp\left[(t\beta\mu - 1)\sum_{i=1}^{k} x_i^2 + t\beta\mu \sum_{i=k+1}^{d} x_i^2\right]$$

Independence

$$= \prod_{i=1}^{k}\left(\mathbb{E} \exp (t\beta\mu - 1) x_i^2\right) \prod_{i=k+1}^{d}\left(\mathbb{E} \exp (t\beta\mu x_i^2)\right)$$

$X \sim N(0,1)$

$$= \left\{\mathbb{E} \exp[(t\beta\mu - 1)X^2]\right\}^k \left[\mathbb{E} \exp(t\beta\mu X^2)\right]^{d-k}$$

$X \sim N(0,1)$ $\quad \mathbb{E}[e^{sX^2}] = \frac{1}{\sqrt{1-2s}}$ $\quad s \in (-\infty, \frac{1}{2})$

"moment bound. sub-Gaussian"

$$\text{above} = (1-2t(\beta\mu-1))^{-k/2}(1-2t\beta\mu)^{-(d-k)/2}$$

$$=: g^{-1}(t)$$

$$g(t) = (1-2t(\beta\mu-1))^{k/2}(1-2t\beta\mu)^{\frac{d-k}{2}}$$

$$\min_t g^{-1}(t) = \max_t g(t) \qquad t \in (0, \frac{1}{2\beta\mu})$$

$$0 = g'(t) \rightarrow t_0 = \frac{1-\beta}{2\beta(d-\beta k)}$$

$$g(t_0) = \left(\frac{d-k}{d-k\beta}\right)^{\frac{(d-k)}{2}} \left(\frac{1}{\beta}\right)^{k/2}$$

$$g^{-1}(t_0) = \beta^{k/2}\left(1 - \frac{(1-\beta)k}{d-k}\right)^{\frac{d-k}{2}}. \qquad \square$$

注：(b) 证明相似.

Exponential Probablistic Ineq. (Markov + Indep.)

Compressed Sensing

$x_0 \in \mathbb{R}^p$.

$b = Ax_0$. $b \in \mathbb{R}^m$ m-measurement

$m > p$. indep.

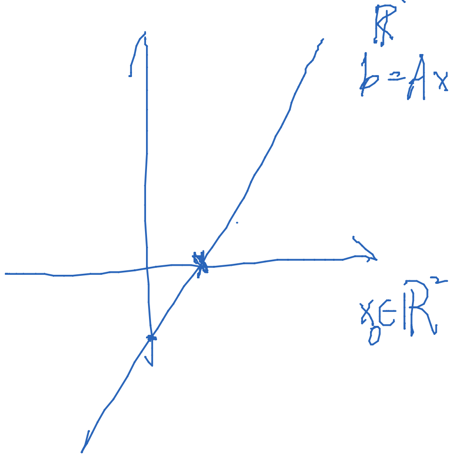
* $m < p$. ? underdetermint eq. No!

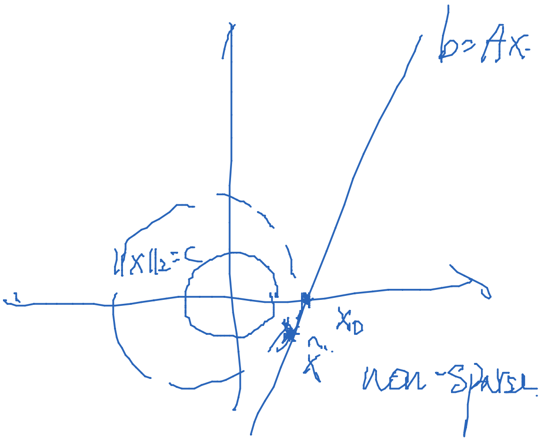
Assumption: x_0 is sparse

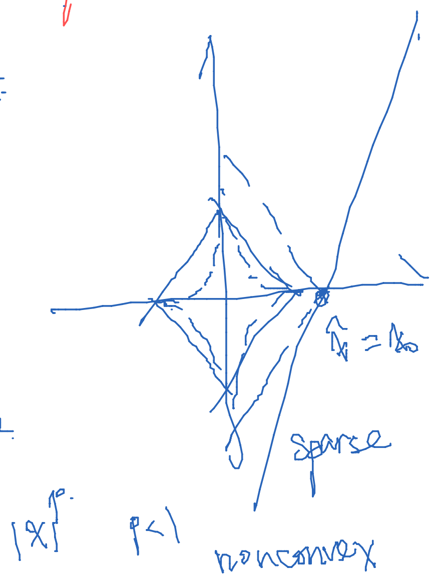
$s = \|x_0\|_0 = \#\{i : x_0(i) \neq 0\} < \min(p, m)$

(P_0) $\min \|x\|_0$ sparsest sdn NP-hard.
s.t. $b = Ax$.

(P_1) $\min \|x\|_1$ $\hat{x} \neq x_0$. Convex Relaxation
s.t. $b = Ax$. p.







Basis Pursuit. (Donoho-Chen' 1996) "LASSO"

P_1 $\quad \min \|x\|_1 := \sum_{i=1}^{p} |x_i|$

$\quad$ s.t. $Ax = b$.

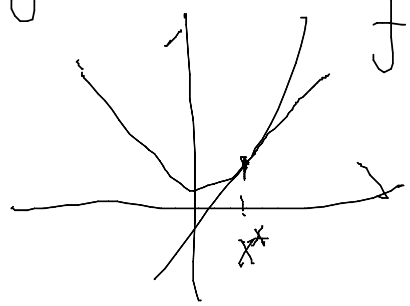
$\hat{x} = x_0$? ERC ?

$$\max_{\lambda} \min_{x} L(x;\lambda) = \|x\|_1 + \lambda^T(Ax-b) \qquad \lambda \in \mathbb{R}^m$$

$$0 \in \partial L(\hat{x},\hat{\lambda}) = \text{sign}(x_i) + A_i^T\lambda = 0 \quad \text{nonsmooth.} \quad x_i \neq 0.$$

"Smooth $0 = \partial L(\hat{x},\hat{\lambda})$" Subgradient $\quad A_i^T\lambda \in [-1,1] \Leftrightarrow |A_i^T\lambda| \le 1 \quad x_i = 0$

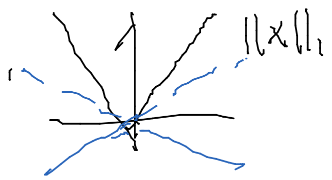
Subgradient



$$f(x) \geq f(x^*) + \overset{g_{x^*}}{\partial f(x^*)} \cdot (x - x^*)$$

$g_{x^*} \in \partial f(x^*)$ subgradient of f at x^*

- $f(x)$ differentiable at x^*. $\partial f(x^*) = \{f'(x^*)\}$
- $f(x)$ non- ...



$$\partial \|x\|_1 \Big|_{x=0} = [-1, 1].$$

Witness Method λ

1) $A_i^T\lambda = \text{sign}(x_0)$ $\quad i \in \text{supp}(x_0) = ?$

2) $|A_i^T\lambda| < 1$ $\quad i \notin \ldots$

$\hat{x} = x_0$. 依赖于 x_0 sign pattern.

Universal Recovery Condition. $\forall x_0 \quad \|x_0\|_0 \le s$.

Incoherence Condition

$$\max_{i\neq j} \frac{|\langle A_i, A_j\rangle|}{\|A_i\|\,\|A_j\|} =: \mu(A).$$

Donoho-Huo '2001 $\dfrac{1+\dfrac{1}{\mu(A)}}{2} > s \Leftrightarrow \mu(A) < \dfrac{1}{2s-1}$

Elad-Bruckstein '2001. $\mu(A) < \dfrac{\sqrt{2}-\frac{1}{2}}{s}$

Candes-Tao '2006 Restricted Isometry Property (RIP)

$\exists$ constant δ_k s.t.

$$\underbrace{1-\delta_k \le \frac{\|Ax\|^2}{\|x\|^2} \le 1+\delta_k}, \quad \forall x : k\text{-sparse } \|x\|_0 \le k$$

A is RIP (δ_k)

" $A^{m\times p}$. $\to$ RIP ? NP-hard "

Incoherence $\Rightarrow$ RIP more general

Orthogonal $\to$ RIP

Random Projection $\to$ RIP matrix

Thm If $\Phi \in \mathbb{R}^{m\times p}$. is a random matrix satisfying

$$\text{Prob}\left(\|\Phi x\|_2^2 - \|x\|_2^2 \ge \epsilon \|x\|_2^2 \right) \le 2e^{-n\,c(\epsilon)}$$

Then: $\forall \delta > 0$. $\forall$ k-sparse x. RIP(δ) holds with prob. $1 - c(\delta)e^{-c(\epsilon)n}$

Johnson-Lindenstrauss Lemma

RM $\longrightarrow$ RIP

Universal Exact Recovery Condi.

Thm $\forall$ k-sparse x_0. s.t. $Ax_0 = b$ A is $RIP(\delta_k)$

(1) A is $RIP(\delta_k)$. $\delta_{2k} < 1$. $\Rightarrow$ P_0 $\exists!$ sol'n x_0

(2) $\delta_{2k} < \sqrt{2} - 1 \approx 0.414 \Rightarrow P_1$ $\exists!$ sol'n $\hat{x} = x_0$