EE 546 HW2.

Assigned. Oct. 5 Due Oct. 18

Kangyan Xu

1. Concentration.

Let $X_1, X_2, \cdots, X_m$ be i.i.d Gaussian random variables with $\mathcal{N}(0,1)$ distribution.
Let a_r be i.i.d random Gaussian vectors distributed as $\mathcal{N}(0, I)$ and independent of $X_1, \cdots, X_m$.
Let y be a fixed vector.

Understanding the random variable : $Z = \frac{1}{m} \sum_{r=1}^{m} X_r |X_r| \operatorname{sign}(X_r + a_r^T y)$

(a) What is the expected value of Z ?

(b) How well is the random variable Z concentrated ? That is, can you bound $P\{|Z - E[Z]| \geq t\}$?

(a). We have $E[X^2 \operatorname{sign}(X) \operatorname{sign}(\alpha X + \beta Y)] = \frac{2}{\pi} \tan^{-1}\left(\frac{\alpha}{\beta}\right) + \frac{2}{\pi} \frac{\alpha \beta}{\alpha^2 + \beta^2}$. $(X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1))$

1° $|X_r| = X_r \cdot \operatorname{sign}(X_r)$

2° $a_r^T y = \sum_{i=1}^{n} a_{r_i} y_i = \sum_{i=1}^{n} y_i \cdot a_{r_i}$ $\quad E[a_r^T y] = \sum_{i=1}^{n} y_i \cdot \underbrace{E[a_{r_i}]}_{=0} = 0$ $\quad V[a_r^T y] = \sum_{i=1}^{n} y_i^2 \underbrace{V[a_{ri}]}_{=1} = y^T y$

$a_r^T y \sim \mathcal{N}(0, y^T y) \Rightarrow \frac{a_r^T y}{\sqrt{y^T y}} \sim \mathcal{N}(0,1)$

---

let $\alpha = 1$, $\beta = \sqrt{y^T y}$, $X = X_r$, $Y = \frac{a_r^T y}{\sqrt{y^T y}}$,

$$E[X_r |X_r| \operatorname{sign}(X_r + a_r^T y)] = E\left[X_r^2 \operatorname{sign}(X_r) \operatorname{sign}\left(1 \cdot X_r + \sqrt{y^T y} \cdot \frac{a_r^T y}{\sqrt{y^T y}}\right)\right]$$

$$= \frac{2}{\pi} \tan^{-1}\left(\frac{1}{\sqrt{y^T y}}\right) + \frac{2}{\pi} \cdot \frac{\sqrt{y^T y}}{1 + y^T y} = \text{constant}$$

$$\therefore E[Z] = \frac{1}{m} \cdot m \cdot \text{constant} = \frac{2}{\pi} \cdot \tan^{-1}\left(\frac{1}{\sqrt{y^T y}}\right) + \frac{2}{\pi} \cdot \frac{\sqrt{y^T y}}{1 + y^T y}.$$

(b) $Z_r = X_r |X_r| \operatorname{sign}(X_r + a_r^T y) = X_r^2 \operatorname{sign}(X_r) \operatorname{sign}(X_r + a_r^T y)$

$\because$ X is Gaussian (sub-G) iff X^2 is sub-exponential $\quad P\{|X_r^2| > t\} \leq \exp(1 - t/k_1)$

$P\{|X_r^2| > t\} = P\{|X_r^2 \operatorname{sign}(X_r) \operatorname{sign}(X_r + a_r^T y)| > t\} = P\{|Z_r| > t\}$ $\quad \therefore Z_r$ is sub-exponential

$Z = \frac{1}{m} \sum_{r=1}^{m} Z_r$ $\quad \therefore P\{|Z - E[Z]| \geq t\} \leq \exp(1 - t/k_1)$. $(t \geq 0)$

2. Operator norms.

The l_2 operator norm of $A \in R^{m \times n}$ is defined as $\|A\| := \max\{\|Ax\|_{\ell_2} : \|x\|_{\ell_2} = 1\}$

(i) Prove that the operator norm can alternatively be written in the following two forms:

$\|A\| = \max\{\langle x, Ay\rangle : x \in R^m,\ y \in R^n,\ \|x\|_{\ell_2} = \|y\|_{\ell_2} = 1\}$ & $\|A\| = \sigma_1(A)$ where $\sigma_1(A) \geq \cdots \geq \sigma_n(A)$ are the singular values of A.

(ii) Prove that

$\sum_{s=1}^{r} \sigma_s(A) = \max\{\operatorname{trace}(U^T A V) : U \in R^{m \times r},\ V \in R^{n \times r}, \text{ and } U^T U = V^T V = I_r\}$. where $\sigma_1(A) \geq \cdots \geq \sigma_n(A)$ are ..., and I_r is $r \times r$ identity matrix.

(i)

$\max\{\langle x, Ay\rangle : x \in R^m, y \in R^n, \|x\|_{\ell_2} = \|y\|_{\ell_2} = 1\} = \max\{\|x\|_{\ell_2} \cdot \|Ay\|_{\ell_2} \cdot \overset{\cos\theta}{\vee},\ x \in R^m, y \in R^n, \|x\|_{\ell_2} = \|y\|_{\ell_2} = 1\}$

$= \max\{\|Ay\|_{\ell_2} \cdot \cos\theta,\ y \in R^n,\ \|y\|_{\ell_2} = 1\} \underset{\Delta}{\leq} \max\{\|Ay\|_{\ell_2},\ \|y\|_{\ell_2} = 1\} = \|A\|$

$\because$ max $\quad \therefore \max\{\langle x, Ay\rangle : x \in R^m, y \in R^n, \|x\|_{\ell_2} = \|y\|_{\ell_2} = 1\} = \|A\|$.

SVD: $\quad A = \underset{m \times m}{U} \cdot \underset{m \times n}{\Sigma} \cdot \underset{n \times n}{V^T}$ $\quad , U \cdot U^T = I, V \cdot V^T = I, \quad \|Ux\|_{\ell_2}^2 = x^T \underbrace{U^T U}_{I} x = x^T x = \|x\|_{\ell_2}^2$

$\Rightarrow \|Ux\|_{\ell_2} = \|x\|_{\ell_2}$

$\|A\| = \max\{\|Ax\|_{\ell_2} : \|x\|_{\ell_2} = 1\} = \max\{\|U\Sigma V^T \underline{x}\|_{\ell_2} : \|x\|_{\ell_2} = 1\} = \max\{\|U \underline{\Sigma x}\|_{\ell_2} : \|x\|_{\ell_2} = 1\}$

$= \max\{\|\Sigma x\|_{\ell_2} : \|x\|_{\ell_2} = 1\} = \max\{\sqrt{\sigma_1^2 x_1^2 + \cdots + \sigma_n^2 x_n^2} : \|x\|_{\ell_2} = 1\} = \sigma_1 \quad (x_1 = 1,\ x_2 \sim x_n = 0)$

(ii) $\quad \underset{r \times m}{U^T} \underset{m \times n}{A} \underset{n \times r}{V} = U^T \cdot U' \Sigma V'^T \cdot V$

$U^T U' \cdot (U^T U')^T = I_r$ & $(V'^T \cdot V)^T \cdot (V'^T \cdot V) = I_r$

$\underset{r \times m}{}$ $\qquad \underset{n \times r}{}$

$= \underset{r \times m}{U''} \cdot \underset{m \times n}{\Sigma} \cdot \underset{n \times r}{V''^T}$ $\quad (U'' = U^T \cdot U',\ V''^T = V'^T \cdot V)$

$\operatorname{trace}(U^T A V) = U_{11} \sigma_1 V_{11} + \cdots + U_{rr} \cdot \sigma_r \cdot V_{rr}$, and $U_{ii}^2 \leq 1$, $V_{ii}^2 \leq 1$.

$\therefore \max\{\operatorname{trace}(U^T A V)\} = \sigma_1 + \cdots + \sigma_r = \sum_{s=1}^{r} \sigma_s(A)$ $\qquad$ (when $U_{ii} = V_{ii} = 1$).

3. Prove the following upper and lower bounds $A_{m\times n}$

① $\|A\| \le \sqrt{m} \max_{i\in\{1,2,\cdots m\}} \left(\sum_{j=1}^{n} A_{ij}^2\right)^{1/2}$

② $\|A\| \ge \frac{1}{\sqrt{mn}} \sum_{i=1}^{m} \left|\sum_{j=1}^{n} A_{ij}\right|$

---

① $\|A\|^2 = \max \|Ax\|_{\ell_2}^2 = \max \left[(A_{11}x_1 + \cdots + A_{1n}x_n)^2 + \cdots + (A_{m1}x_1 + \cdots + A_{mn}x_n)^2 \right]$

Cauchy 不等.

$\le \max \left[(A_{11}^2 + \cdots + A_{1n}^2)(x_1^2 + \cdots + x_n^2) + \cdots + (A_{m1}^2 + \cdots + A_{mn}^2)(x_1^2 + \cdots + x_n^2) \right]$

$= \max \left[(A_{11}^2 + \cdots + A_{1n}^2) + \cdots + (A_{m1}^2 + \cdots + A_{mn}^2) \right]$

$\le m \cdot \max (A_{i1}^2 + \cdots + A_{in}^2) = m \cdot \max_{i\in\{1,\cdots m\}} \sum_{j=1}^{n} A_{ij}^2$

$\therefore \|A\| \le \sqrt{m} \cdot \max_{i\in\{1,\cdots m\}} \left(\sum_{j=1}^{n} A_{ij}^2\right)^{1/2}$

---

② $\|A\|^2 = \max \|Ax\|_{\ell_2}^2 = \max \left[(A_{11}x_1 + \cdots + A_{1n}x_n)^2 + \cdots + (A_{m1}x_1 + \cdots + A_{mn}x_n)^2 \right]$

Cauchy 不等.

$\ge \max \left[\left(\frac{A_{11}x_1 + \cdots + A_{1n}x_n + \cdots + A_{m1}x_1 + \cdots + A_{mn}x_n}{\sqrt{m}} \right)^2 \right]$

Pick $x_i = \frac{1}{\sqrt{n}}$ $(A_{ij} > 0)$ or $-\frac{1}{\sqrt{n}}$ $(A_{ij} < 0)$ so $\sum_{i=1}^{n} x_i^2 = 1$

$\ge \left[\left(\frac{|A_{11}| + \cdots + |A_{1n}| + \cdots + |A_{m1}| + \cdots + |A_{mn}|}{\sqrt{mn}} \right)^2 \right] \ge \frac{1}{mn} \left(\sum_{i=1}^{m} \left| \sum_{j=1}^{n} A_{ij} \right| \right)^2$

$\therefore \|A\| \ge \frac{1}{\sqrt{mn}} \sum_{i=1}^{m} \left| \sum_{j=1}^{n} A_{ij} \right|$

---

4. Prove that

$$\|A\|_F = \left(\sum_{i=1}^{\min(m,n)} \sigma_i^2(A) \right)^{1/2}$$

where $\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2}$ is the Frobenius norm of the matrix A. Deduce that

$$\|A\| \leq \|A\|_F \leq \sqrt{\text{rank}(A)} \cdot \|A\|$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2 = A_{11}^2 + \cdots + A_{1n}^2 + \cdots + A_{m1}^2 + \cdots + A_{mn}^2$$

$$= \text{trace}(A \cdot A^T)$$

$$= \text{trace}(\underset{m\times m}{U} \cdot \underset{m\times n}{\Sigma} \cdot \underset{n\times n}{V^T} \cdot V \cdot \Sigma^T \cdot U^T)$$

$$= \text{trace}(\underbrace{U \cdot \Sigma^2} \cdot U^T) = \text{trace}(\underbrace{U^T U} \Sigma^2)$$

$$\cancel{= u_{11}\sigma_1^2 u_{11} + \cdots + u_{kk}\sigma_k^2 u_{kk},\ k = \min(m,n),\ u_{i1}^2 +}$$

$$= \text{trace}(\Sigma^2) = \sigma_1^2 + \cdots + \sigma_k^2, \quad k = \min(m,n)$$

$$\therefore \quad \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2} = \left(\sum_{i=1}^{\min(m,n)} \sigma_i^2(A) \right)^{1/2} = \|A\|_F$$

---

① $\|A\| = \sigma_1(A) = (\sigma_1^2(A))^{1/2} \leq \|A\|_F$

②

$$\Sigma = \begin{array}{c|c} \left[\begin{array}{cc|c} \sigma_1 & & \\ & \ddots \; \sigma_r & 0 \\ \hline & 0 & 0 \end{array}\right] & \begin{array}{l} r \\ m-r \end{array} \\ \begin{array}{cc} r & n-r \end{array} & \end{array}$$

$$\therefore \sum_{i=1}^{\min(m,n)} \sigma_i^2(A) \leq \underbrace{\sigma_1^2 + \cdots + \sigma_1^2}_{\text{rank}(A)} = \text{rank}(A) \cdot \|A\|^2$$

$$\sum_{i=1}^{\min(m,n)} \sigma_i^2(A) = \underbrace{\sigma_1^2 + \cdots + \sigma_r^2}_{\text{rank}(A)} + 0 + \cdots + 0$$

$$\therefore \quad \|A\|_F \leq \sqrt{\text{rank}(A)}\, \|A\|$$

① & ② $\Rightarrow$ $\|A\| \leq \|A\|_F \leq \sqrt{\text{rank}(A)}\, \|A\|.$

End.