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

$\gamma(T) = \sqrt{\log |T|}$

- 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

$A =$ 

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.

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

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

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,

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

[$a^2 - b^2 = (a+b)(a-b)$]

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

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

[If $\mathbb{E}X \neq 0$, $\mathbb{E}XX^T$ is called the second moment matrix]

- 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

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

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 $\mathbb{R}^n$. Then, for all $m > n$,

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

Proof:

Actual Mean Σ

Empirical Mean Σ_m

For analysis, bring X_i's to isotropic position

⌈ 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 ⌋

Let Z satisfy

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

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

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

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

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

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

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

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

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

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

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

let

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

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

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

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

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

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

trace = sum of eigen values

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

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

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

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

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

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

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

lower dimensional distributions require fewer samples.

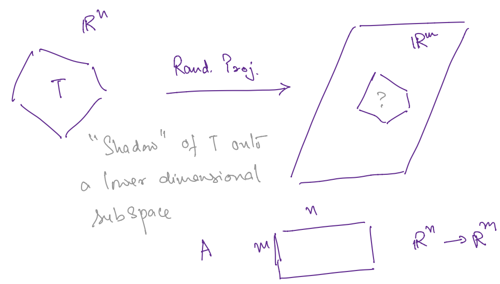
$$\mathbb{E}\,\|\Sigma_m - \Sigma\| \lesssim \sqrt{\frac{r}{m}} \cdot \|\Sigma\|$$

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

Application ③ : Random Projection of Sets

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

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



MDI for $T-T$ and triangle inequality

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

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

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

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

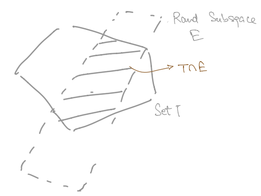
$$\mathbb{E}\ \mathrm{diam}(PT) \le \sqrt{\frac{m}{n}} \cdot \mathrm{diam}(T) + 2CK^2 \frac{\gamma(T)}{\sqrt{n}}$$

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

$\dim(E) = n - m$



THM (M* bound) (Milman)

$$\mathbb{E}\ \mathrm{diam}(T \cap E) \le \frac{C\ \gamma(T)}{\sqrt{m}}$$

Proof:

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

Consider $T - T$

MDI:

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

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

Sup over a smaller set

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

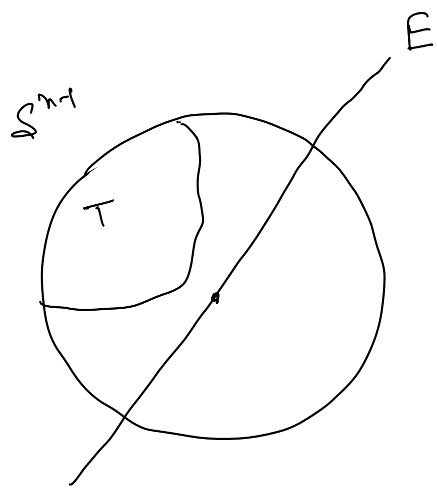
$Ax = 0$ for all $x \in E = \mathrm{Ker}(A)$

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

$\square$

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 $\text{co-dim}(E) = m$.
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 all $x \in T$

$\Updownarrow$

$T \cap E = \emptyset$

$\left[E = \ker(A) \right]$

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

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

Therefore,

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

$\boxtimes$