Numerical Linear Algebra

Lecture Notes 6:Applications

Professor: Zhihua Zhang

Theorem 0.1. *If A and B are $n \times n$ matrices, then*

- $\sigma(A+B) \prec_w \sigma(A) + \sigma(B)$
- $\sum_{i=1}^{k} \sigma_i(A) - \sum_{i=1}^{k} \sigma_{n-i+1}(B)$, $k = 1, ..., n$
- *If $A = [A_1, A_2]$, $A_1 \in \mathbb{R}^{n \times p}$, $A_2 \in \mathbb{R}^{n \times (n-p)}$, and $B = [A_1, 0]$, then $\sigma(A) \succ_w \sigma(B)$.*

Definition 0.1. *A real-valued function $\|\cdot\|$ on $\mathbb{R}^{m \times n}$ is called a unitarily invariant norm if the following conditions are satisfied*

1. $\|X\| > 0$, $X \neq 0$
2. $\|\alpha X\| = |\alpha| \|X\|$
3. $\|X + Y\| \leq \|X\| + \|Y\|$
4. $\|XV\| = \|UX\| = \|X\|$. *Here U and V are any unitary matrix.*

   Conduct SVD decomposition $X = U\Sigma V^T$, then we have

- $\|X\| = \|U\Sigma V^T\| = \|\Sigma\| = \|\sigma(X)\|$
- $\|\sigma(X)\|_2 = \|X\|_F$
- $\|\sigma(X)\|_\infty = |||X|||_2$.

Theorem 0.2. *Let A and B be $m \times n$ matrices, then for all $n \times n$ matrices X,*

$$\|A(A^\dagger B) - B\| \leq \|AX - B\|$$

for every unitarily invariant norm $\|\cdot\|$.

Proof. Here is an easy proof from symmetric gauge function.

$$\begin{aligned} A^\dagger B &= \underset{X \in \mathbb{R}^{n \times n}}{\arg\min} \|AX - B\| \\ &\Leftrightarrow \phi(\sigma(AA^\dagger B - B)) \leq \phi(\sigma(AX - B)) \\ &\Leftrightarrow \sigma(AA^\dagger B - B) \prec_w \sigma(AX - B) \end{aligned}$$

Define $L = AX - B$, $P = AX - AA^\dagger B$, $Q = AA^\dagger B - B$. It is easy to verify $L = P + Q$ and $P^T Q = 0$. Thus we have

$$L^T L = P^T P + Q^T Q,$$

which indicates that $\lambda_i(L^T L) \leq \lambda_i(Q^T Q) \Leftrightarrow \sigma_i(L) \leq \sigma_i(Q)$. $\square$

Theorem 0.3 (Eckart and Young 1936, Minksy 1960)**.** *Let A be an $m \times n$ matrix of rank r. Let $A = U_A \Sigma_S V_A^T$ be the full SVD and let $k \leq r$. Define $A_k = U_{A,k} \Sigma_{A,k} V_{A,k}^T$. Then $\|A - A_k\| \leq \|A - X\|$ for all $m \times n$ matrices X of rank k and every unitarily invariant norm.*

1 Subgradient of Matrix

Let us consider an optimization problem,

$$\min f(X) \triangleq \|A - X\|_F^2 + \lambda \|X\|_*.$$

First, let us review the gradient of scalar value and vector,

- $f : \mathbb{R} \to \mathbb{R}$, $\partial f(x) = \{z \in \mathbb{R} : f(y) \geq f(x) + (y - x)z \quad \forall y \in \mathbb{R}\}$
- $f : \mathbb{R}^n \to \mathbb{R}$, $\partial f(x) = \{z \in \mathbb{R}^n : f(y) \geq f(x) + (y - x)^T z \quad \forall y \in \mathbb{R}^n\}$

Definition 1.1. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^{m\times n}$ real-value matrix, and $A \in \mathbb{R}^{m\times n}$. Then subgradient of $\|A\|$ is defined by*

$$\partial\|A\| = \{G \in \mathbb{R}^{m\times n} : \|B\| \geq \|A\| + tr((B - A)^T G) \quad \forall B \in \mathbb{R}^{m\times n}\}$$

.

Lemma 1.1. *$G \in \partial\|A\|$ is equivalent to that*

1. $\|A\| = tr(G^T A)$
2. $\|G\|_D \leq 1$

Proof. (Not complete) If $G \in \partial\|A\|$, then $\forall B \in \mathbb{R}^{m\times n}$, $\|B\| \geq \|A\| + tr((B - A)^T G)$. Let $B = 2A$, then we have

$$2\|A\| \geq \|A\| + tr(A^T G) \implies \|A\| \geq tr(A^T G).$$

Let $B = \frac{1}{2}A$, then we have $\|A\| \leq tr(A^T G)$ similarily. Then we get $\|A\| = tr(G^T A)$. This implies that

$$tr(B^T G) \leq \|B\| \quad \forall B.$$

Then we have

$$\|G\|_D = \max_{\|Z\|=1} tr(G^T Z) \leq \|Z\| = 1.$$

Next we prove another direction. $\|\cdot\|$ is a unitarily invariant norm, then there exists a symmetric gauge function $\phi(\sigma(A)) = \|A\|$.
$Z \in \partial\phi(X)$ is equivalent to

1. $\phi(x) = x^T z$
2. $\phi^*(z) \leq 1$, where $\phi * (z) = \max_{\phi(y)=1} z^T y$.

□

Theorem 1.1. *Let $\|\cdot\|$ be unitarily invariant norm on $\mathbb{R}^{m\times n}$, $D = diag(\vec{d})$ and $\vec{d} = dg(D)$. Then $\partial\|A\| = conv\{UDV^T : A = U\Sigma V^T, \vec{d} \in \partial\phi(\vec{\sigma})\}$.*