

Lecture 5

Math 467

Jan 19 2024

Matrix norm

$A \in \mathbb{R}^{m \times n}$, norm $\|\cdot\|$

1. $\|A\| \geq 0$ ($\|A\| = 0$ only if $A = 0$)
2. $\|\alpha \cdot A\| = |\alpha| \|A\|$ for any $\alpha \in \mathbb{R}$.
3. $\|A + B\| \leq \|A\| + \|B\| \rightarrow$ triangle inequality
4. $\|AB\| \leq \|A\| \cdot \|B\|$

Induced Metric:

$A \in \mathbb{R}^{m \times n}$, $A : \mathbb{R}^n \mapsto \mathbb{R}^m$ (if $x \in \mathbb{R}^n$ then $Ax \in \mathbb{R}^m$)

$$\|A\| = \max_{\|x\|=1} \|Ax\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

This is the induced norm. $\|Ax\| \leq \|A\| \cdot \|x\|$.

Theorem

$\|\cdot\|$ is a matrix norm *and* also satisfies (4) (from the conditions)

Proof

(These are pretty trivial)

1. $\|A\| \geq 0$ as $\|Ax\| \geq 0$ and $\|A\| = 0$ only if $A = 0$
2. $\|\alpha A\| \geq |\alpha| \|A\|$ because $x \in \mathbb{R}^n$, $\|(\alpha A)x\| = \|\alpha Ax\| = |\alpha| \|Ax\|$.

(Look at these)

3. let $x \in \mathbb{R}^n$, then $\|(A + B)x\| = \|(Ax) + (Bx)\|$

$$\begin{aligned}
& \leq \|Ax\| + \|Bx\| \\
& \leq \|A\| \cdot \|x\| + \|B\| \cdot \|x\| \\
& = (\|A\| + \|B\|)\|x\| \\
& \rightarrow \|A + B\| = \max_{\|x\|=1} \|(A + B)x\| \\
& \leq \|A\| + \|B\|
\end{aligned}$$

4. $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p} \rightarrow AB \in \mathbb{R}^{m \times p}$. $x \in \mathbb{R}^p \rightarrow Bx \in \mathbb{R}^n$.

$$\begin{aligned}
\|AB\| &= \max_{\|x\|=1} \|ABx\| \\
&\leq \max_{\|x\|=1} \|A\| \cdot \|Bx\| \leq \max_{\|x\|=1} \|A\| \cdot \|B\| \cdot \|x\| \\
&= \|A\| \cdot \|B\|
\end{aligned}$$

Recall:

$$\|B\| = \max_{x \neq 0} \frac{\|Bx\|}{\|x\|} \rightarrow \|B\| \cdot \|x\| \geq \|Bx\|$$

multiplying the $\|x\|$ over.

Eigenvalues and Norms

Let $A \in \mathbb{R}^{m \times n}$. Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be eigenvalues of $A^T A$. If $\|\cdot\|$ is the induced metric,

$$\sqrt{\lambda_n} \leq \|A\| \leq \sqrt{\lambda_1}$$

Proof)

Notice $A^T A \in \mathbb{R}^{n \times n}$ and is also symmetrical ($(A^T A)^T = A^T (A^T)^T = A^T A$). The quadratic form $f(x) = x^T A^T A x$ for $x \in \mathbb{R}^n$. This quadratic form is semi-definite ($f(x) = x^T A^T A x = \langle x, A^T A x \rangle = \langle Ax, Ax \rangle = \|Ax\|^2 \geq 0$) $\rightarrow A^T A$ has eigenvalues $\lambda_i \in \mathbb{R}$ and $\lambda_i \geq 0$. Let's assume λ_i are all distinct \rightarrow There exists orthonormal basis of eigenvectors v_1, v_2, \dots, v_n .

$$A^T A v_j = \lambda_j v_j, \|v_j\| = 1$$

$$\langle v_j, v_k \rangle = \delta_{jk} = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}$$

We have to show that for $x \in \mathbb{R}^n$

$$\sqrt{\lambda_n} \|x\| \leq \|Ax\| \leq \sqrt{\lambda_1} \|x\|$$

Let $x \in \mathbb{R}^n$, then there are coefficient $c_j \in \mathbb{R}$ so that $x = \sum_{j=1}^n c_j v_j$, then

$$\begin{aligned}
||Ax||^2 &= \langle Ax, Ax \rangle = \langle x, A^T Ax \rangle = \left\langle \sum_{j=1}^n c_j v_j, A^T A \sum_{k=1}^n c_k v_k \right\rangle \\
&= \sum_{j=1}^n \sum_{k=1}^n c_j c_k \langle v_j, A^T A v_k \rangle \\
&= \sum_{j,k=1}^n c_j c_k \langle v_j, \lambda_k v_k \rangle \\
&= \sum_{j,k=1}^n c_j c_k \lambda_k \langle v_j, v_k \rangle \\
&= \sum_{j=1}^n c_j^2 \lambda_j
\end{aligned}$$

$$||Ax||^2 = \sum_{j=1}^n c_j^2 \lambda_j \leq \lambda_1 \sum_{j=1}^n c_j^2 = \lambda_1 ||x||^2$$

$$||Ax||^2 = \sum_{j=1}^n c_j^2 \lambda_j \geq \lambda_n \sum_{j=1}^n c_j^2 = \lambda_n ||x||^2$$

so

$$\begin{aligned}
\lambda_n ||x||^2 &\leq ||Ax||^2 \leq \lambda_1 ||x||^2 \\
\sqrt{\lambda_n} ||x|| &\leq ||Ax|| \leq \sqrt{\lambda_1} ||x|| \\
\sqrt{\lambda_n} &\leq \frac{||Ax||}{||x||} \leq \sqrt{\lambda_1}, x \neq 0
\end{aligned}$$

Chapter 4

functions:

$$f(x) \in \mathbb{R}^m, x \in \mathbb{R}^n$$

$$f(x) = \begin{cases} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \dots \\ f_m(x_1, x_2, \dots, x_n) \end{cases}$$

Special case: $m = 1$, then $f : \mathbb{R}^n \mapsto \mathbb{R}$.

$$f(x_1, x_2, \dots, x_n)$$

it has a graph

$$g = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \\ f(x_1, x_2, \dots, x_n) \end{bmatrix} \in \mathbb{R}^{n+1} \right.$$

so for $n = 2$ we get a 3D graph.

Linear approximation

$$f : \mathbb{R}^n \mapsto \mathbb{R}^m$$

which is equivalent to $f(x) \in \mathbb{R}^m, x \in \mathbb{R}^n$

Derivative matrix

$$Df = \begin{bmatrix} \frac{df_1}{dx_1} & \frac{df_1}{dx_2} & \dots & \frac{df_1}{dx_n} \\ \frac{df_2}{dx_1} & \frac{df_2}{dx_2} & \dots & \frac{df_2}{dx_n} \\ \dots & \dots & \dots & \dots \\ \frac{df_m}{dx_1} & \frac{df_m}{dx_2} & \dots & \frac{df_m}{dx_n} \end{bmatrix}$$

The _ a point $\hat{x} \in \mathbb{R}^n$

$$f(x) = f(\hat{x}) + Df(\hat{x})(x - \hat{x}) + o(\|x - \hat{x}\|)$$

Notation:

“little o ”:

we say $\phi(s) = o(s)$ if $\frac{\phi(s)}{s} \rightarrow 0$ as $s \rightarrow 0$.

“Big O ”:

we say $\phi(s) = O(s)$ if there exist a constant C so that $|\phi(s)| \leq C \cdot |s|$

Homework (Entertainment!)

pg 77

5.3, 5.6, 5.10