Introductory Lectures on Optimization
Homework (1)

Yang Yichou
Student ID 3230105697

October, 2025

Excercise 1. Norms

In this exercise, we will give some examples of norms and a useful theorem related to norms in finite dimensional vector space

1. l_p **norm**: The l_p norm is defined by

$$\|x\|_p = (\Sigma_{i=1}^n |x_i|^p)^{\frac{1}{p}} \tag{1}$$

where $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ and p≥1.

   a. Please show that l_p norm is a norm.

   b. Please show that the following equality.

$$\lim_{p\to\infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty = \underset{1\le i\le n}{max} |x_i|. \tag{2}$$

   The l_∞ norm is defined as above

2. **Operator norms**: Suppose that $\mathbf{A} \in \mathbb{R}^{m\times n}$, which can be viewed as linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Please show that the following operator norms' equality.

   a. Let $\|A\|_1 = \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq 0} \frac{\|\mathbf{Ax}\|_1}{\|\mathbf{x}\|_1}$. Please show that

$$\|A\|_1 = \underset{1\le j\le n}{max} \Sigma_{i=1}^m |a_{ij}|. \tag{3}$$

   b. Let $\|A\|_\infty = \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq 0} \frac{\|\mathbf{Ax}\|_\infty}{\|\mathbf{x}\|_\infty}$. Please show that

$$\|\mathbf{A}\|_\infty = \underset{1\le i\le m}{max} \Sigma_{j=1}^n |a_{ij}|. \tag{4}$$

Proof of Excercise 1: Write down your solutions step by step here.
1.a. A norm satisfies the following properties:
(1) positive definiteness:$\|x\| \ge 0$,and $\|x\| = 0$ if and only if $x = 0$
(2) linear:$\forall a \in \mathbb{C}, \|ax\| = |a|\|x\|$
(3) triangle inequality:$\|a\| + \|b\| \ge \|a + b\|$
Then we show that l_p satisfies all the properties listed above:
(1) $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} \ge 0$ is obvious

If $\exists x_i \neq 0, i=1,...,n$.Then $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} \geq |x_i| > 0$
$\therefore \|x\| = 0 \iff x = 0$.
(2)$\forall a \in \mathbb{C}, \|ax\|_p = (\sum_{i=1}^n |ax_i|^p)^{\frac{1}{p}} = (|a|^p)^{\frac{1}{p}}(\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} = |a|\|x\|_p$
(3)$\|a\|_p + \|b\|_p = (\sum_{i=1}^n |a_i|^p)^{\frac{1}{p}} + (\sum_{i=1}^n |b_i|^p)^{\frac{1}{p}}, \|a+\||_p = (\sum_{i=1}^n |a_i+b_i|^p)^{\frac{1}{p}}$
In coordinates,$\forall i, |a_i|^p + |b_i|^p \geq |a_i+b_i|^p$
$(\|a\|_p + \|b\|_p)^p \geq \sum_{i=1}^n |a_i|^p + \sum_{i=1}^n |b_i|^p \geq \sum_{i=1}^n |a_i+b_i|^p = (\|a+b\|)^p$
$\therefore \|a\|_p + \|b\|_p = \|a+b\|_p$

1.b. Let $M = \underset{1\leq i\leq n}{max}\, x_i$,$r$ be the number of different $i, s.t. |x_i| = M$
$\lim\limits_{p\to\infty} \|\mathbf{x}\|_p = \lim\limits_{p\to\infty} (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} = \lim\limits_{p\to\infty} M \cdot (\sum_{i=1}^n (\frac{|x_i|}{M})^p)^{\frac{1}{p}} = M \cdot \lim\limits_{p\to\infty} r^{\frac{1}{p}} = M = \|\mathbf{x}\|_\infty$

2.a.$\mathbf{Ax} = \begin{bmatrix} a_{11} & ... & a_{1n} \\ ... & ... & ... \\ a_{m1} & ... & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ ... \\ x_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n a_{1i}x_i \\ ... \\ \sum_{i=1}^n a_{mi}x_i \end{bmatrix}$
$\|A\|_1 = \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq 0} \frac{\|\mathbf{Ax}\|_1}{\|\mathbf{x}\|_1} = \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq 0} \frac{\sum_{j=1}^m \sum_{i=1}^n |a_{ji}x_i|}{\sum_{i=1}^n |x_i|}$
Let I be the only $\underset{1\leq i\leq n}{argmax} \sum_{j=1}^m |a_{ji}|$
$\|A\|_1 = \underset{1\leq i\leq n}{max} \sum_{j=1}^m |a_{ji}| - \frac{\sum_{i\neq I}(\underset{1\leq i\leq n}{max} \sum_{j=1}^m |a_{ji}| - \sum_{j=1}^m |a_{ji}|)|x_i|}{\sum_{i=1}^n |x_i|} \geq \underset{1\leq i\leq n}{max} \sum_{j=1}^m |a_{ji}|$
The equality holds when $\forall i \neq I, x_i = 0$
Then $\|A\|_1 = \underset{1\leq j\leq n}{max} \sum_{i=1}^m |a_{ij}|$.

2.b.$\|A\|_\infty = \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq 0} \frac{\|\mathbf{Ax}\|_\infty}{\|\mathbf{x}\|_\infty} = \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq 0} \frac{\underset{1\leq i\leq m}{max} |\sum_{j=1}^n a_{ij}x_j|}{\underset{1\leq i\leq n}{max} |x_i|}$
Recall the Holder inequation:$p, q \in [1, +\infty), \frac{1}{p} + \frac{1}{q} = 1$,we have $\sum_{i=1}^n |a_ib_i| \leq (\sum_{i=1}^n |a_i|^p)^{\frac{1}{p}} (\sum_{i=1}^n |b_i|^q)^{\frac{1}{q}}$
Let $p = 1, q = \infty$,$\sum_{i=1}^n |a_ib_i| \leq (\sum_{i=1}^n |a_i|) \underset{1\leq i\leq n}{max} |b_i|$
Then $\sum_{j=1}^n |a_{ij}x_j| \leq \sum_{j=1}^n |a_{ij}| \underset{1\leq i\leq n}{max} |x_i|$
$\therefore \frac{\|\mathbf{Ax}\|_\infty}{\|\mathbf{x}\|_\infty} \leq \underset{1\leq i\leq m}{max} \sum_{j=1}^n |a_{ij}|$
The equality holds when $x_j = \overline{sgn(a_{ij})}$,where $sgn(z) = \begin{cases} \frac{z}{|z|}, z \neq 0 \\ 0, z = 0 \end{cases}$,this ensures $\underset{1\leq i\leq n}{max} |x_i| = 1$.
$\therefore \|A\|_\infty = \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq 0} \frac{\|\mathbf{Ax}\|_\infty}{\|\mathbf{x}\|_\infty} = \underset{1\leq i\leq m}{max} \sum_{j=1}^n |a_{ij}|, x_j = \overline{sgn(a_{ij})}$

$\square$

Excercise 2. Basis and Coordinates

Suppose that $\{\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_n}\}$ is a basis of an n-dimensional vector Space V.

1. Show that $\{\lambda_1\mathbf{a_1}, \lambda_2\mathbf{a_2}, ..., \lambda_n\mathbf{a_n}\}$ is also a basis of V for nonzero scalars $\lambda_1, \lambda_2, ..., \lambda_n$.
2. Suppose that the coordinate of a vector $\mathbf{v}$ under the basis $\{\mathbf{a_1}, ..., \mathbf{a_n}\}$ is $\mathbf{x} = (x_1, x_2, ..., x_n)$
   a. What is the coordinate of $\mathbf{v}$ under $\{\lambda_1\mathbf{a_1}, ..., \lambda_n\mathbf{a_n}\}$?
   b. What are the coordinates of $\mathbf{w} = \Sigma_{i=1}^n \mathbf{a_i}$ under $\{\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_n}\}$ and $\{\lambda_1\mathbf{a_1}, \lambda_2\mathbf{a_2}, ..., \lambda_n\mathbf{a_n}\}$? Note that $\lambda_i \neq 0\ \forall i \in \{1, ..., n\}$.

Proof of Excercise 2: Write down your solutions step by step here.
1.To show $\{\lambda_1\mathbf{a_1}, \lambda_2\mathbf{a_2}, ..., \lambda_n\mathbf{a_n}\}$ is also a basis of V,we only need to show
(1)$\forall i \neq j, i, j \in \{1, ..., n\}, < \lambda_i\mathbf{a_i}, \lambda_j\mathbf{a_j} >= 0$;
(2)$\forall x \in V, x = \sum_{i=1}^n k_i\lambda_i\mathbf{a_i}$
Because $\{\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_n}\}$ is a basis of an n-dimensional vector Space V,we have
$\forall i \neq j, i, j \in \{1, ..., n\}, < \mathbf{a_i}, \mathbf{a_j} >= 0$
$\forall x \in V, x = \sum_{i=1}^n l_i\mathbf{a_i}$
(1)$\forall i \neq j, < \lambda_i\mathbf{a_i}, \lambda_j\mathbf{a_j} >= \lambda_i\overline{\lambda_j} < \mathbf{a_i}, \mathbf{a_j} >= 0$
(2)Because $\lambda_i \neq 0, \forall i$,we can let $k_i = \frac{l_i}{\lambda_i}$,then $\forall x \in V, x = \sum_{i=1}^n l_i\mathbf{a_i} = \sum_{i=1}^n k_i\lambda_i\mathbf{a}_i$

2.a.The coordinate of a vector $\mathbf{v}$ under the basis $\{\mathbf{a_1}, ..., \mathbf{a_n}\}$ is $\mathbf{x} = (x_1, x_2, ..., x_n)$
$\iff \mathbf{v} = \sum_{i=1}^n x_i\mathbf{a}_i$
Then $\mathbf{v} = \sum_{i=1}^n \frac{x_i}{\lambda_i}\lambda_i\mathbf{a}_i$
$\therefore$ the coordinate of $\mathbf{v}$ under $\{\lambda_1\mathbf{a_1}, ..., \lambda_n\mathbf{a_n}\}$ is $\mathbf{x}' = (\frac{x_1}{\lambda_1}, \frac{x_2}{\lambda_2}, ..., \frac{x_n}{\lambda_n})$

2.b.$\mathbf{w} = \sum_{i=1}^n \mathbf{a_i} = \sum_{i=1}^n 1 \cdot \mathbf{a_i} \Rightarrow$the coordinates of $\mathbf{w} = \Sigma_{i=1}^n\mathbf{a_i}$ under $\{\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_n}\}$ is $(1, 1, ..., 1)$
Similarly,$\mathbf{w} = \sum_{i=1}^n \mathbf{a_i} = \sum_{i=1}^n \frac{1}{\lambda_i} \cdot \lambda_i\mathbf{a_i} \Rightarrow$the coordinates of $\mathbf{w} = \Sigma_{i=1}^n\mathbf{a_i}$ under $\{\lambda_1\mathbf{a_1}, \lambda_2\mathbf{a_2}, ..., \lambda_n\mathbf{a_n}\}$ is $(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, ..., \frac{1}{\lambda_n})$

□

Excercise 3. Differentiability

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function, $\mathbf{x_0} \in \mathbb{R}^n$ be a point, and let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. We say that f is *differentiable* at $\mathbf{x_0}$ with *derivative* L if we have

$$\lim_{\mathbf{x}\to\mathbf{x_0};\mathbf{x}\neq\mathbf{x_0}} \frac{\|f(x) - f(x_0) - L(x - x_0)\|_2}{\|x - x_0\|_2} = 0 \tag{5}$$

We denote this derivative L by $f'(\mathbf{x_0})$.

1. Let $x, a \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Consider the functions as follows. Please show that they are differentiable and find $f'(\mathbf{x})$
   (a) $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$.
   (b) $f(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$.
   (c) $f(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$.

Proof of Excercise 3: Write down your solutions step by step here.

1.(a)$f'(\mathbf{x}) = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}] = [a_1, a_2, ..., a_n] = \mathbf{a}^T$

$\lim\limits_{\mathbf{x}\to\mathbf{x_0};\mathbf{x}\neq\mathbf{x_0}} \frac{\|f(x)-f(x_0)-L(x-x_0)\|_2}{\|x-x_0\|_2} = \lim\limits_{\mathbf{x}\to\mathbf{x_0};\mathbf{x}\neq\mathbf{x_0}} \frac{\|\mathbf{a}^T\mathbf{x}-\mathbf{a}^T\mathbf{x_0}-\mathbf{a}^T(\mathbf{x}-\mathbf{x_0})\|_2}{\|x-x_0\|_2} = 0$

(b)$f'(\mathbf{x}) = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}] = [2x_1, 2x_2, ..., 2x_n] = 2\mathbf{x}^T$

$\lim\limits_{\mathbf{x}\to\mathbf{x_0};\mathbf{x}\neq\mathbf{x_0}} \frac{\|f(x)-f(x_0)-L(x-x_0)\|_2}{\|x-x_0\|_2} = \lim\limits_{h\to 0} \frac{\|f(x_0+h)-f(x_0)-2\mathbf{x_0}^T h\|_2}{\|h\|_2} =$

$\lim\limits_{h\to 0} \frac{\|\mathbf{x_0}^T\mathbf{x_0}+2\mathbf{x_0}^T h+h^T h-\mathbf{x_0}^T\mathbf{x_0}-2\mathbf{x_0}^T h\|_2}{\|h\|_2} = \lim\limits_{h\to 0} \frac{\|h\|_2^2}{\|h\|_2} = 0$

(c) $x = [x_1, ..., x_n]^T, y = [y_1, ..., y_m]^T, \mathbf{A} = (a_{ij})_{m*n}$

$f(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 = (\mathbf{y} - \mathbf{A}\mathbf{x})^T(\mathbf{y} - \mathbf{A}\mathbf{x}) = \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{A}\mathbf{x} + (\mathbf{A}\mathbf{x})^T\mathbf{A}\mathbf{x}$

$f(\mathbf{x})' = -2\mathbf{y}^T\mathbf{A} + 2(\mathbf{A}\mathbf{x})^T\mathbf{A} = 2(\mathbf{A}\mathbf{x} - \mathbf{y})^T\mathbf{A}$

$\lim\limits_{\mathbf{x}\to\mathbf{x_0};\mathbf{x}\neq\mathbf{x_0}} \frac{\|f(x)-f(x_0)-L(x-x_0)\|_2}{\|x-x_0\|_2} = \lim\limits_{h\to 0} \frac{\|f(x_0+h)-f(x_0)-L(h)\|_2}{\|h\|_2}$

$f(x_0 + h) = \mathbf{y^T y} - \mathbf{2y^T A}(\mathbf{x_0} + \mathbf{h}) + (\mathbf{x_0} + \mathbf{h})^\mathbf{T}\mathbf{A^T A}(\mathbf{x_0} + \mathbf{h}) =$

$\mathbf{y^T y} - \mathbf{2y^T A x_0} - \mathbf{2y^T A h} + \mathbf{x_0^T A^T A x_0} + \mathbf{2x_0^T A^T A h} + \mathbf{h^T A^T A h}$

$f(x_0) = \mathbf{y^T y} - \mathbf{2y^T A x_0} + \mathbf{x_0^T A^T A x_0}$

$f(x_0 + h) - f(x_0) = -\mathbf{2y^T A h} + \mathbf{2x_0^T A^T A h} + \mathbf{h^T A^T A h} = \mathbf{2}(\mathbf{A x_0} - \mathbf{y})^\mathbf{T}\mathbf{A h} + \mathbf{h^T A^T A h}$

$L(h) = \mathbf{2}(\mathbf{A x_0} - \mathbf{y})^\mathbf{T}\mathbf{A h}$

$\therefore \lim\limits_{\mathbf{x}\to\mathbf{x_0};\mathbf{x}\neq\mathbf{x_0}} \frac{\|f(x)-f(x_0)-L(x-x_0)\|_2}{\|x-x_0\|_2} = \lim\limits_{h\to 0} \frac{\|h^T A^T A h\|_2}{\|h\|_2} \leq \lim\limits_{h\to 0} \frac{\|A^T A\|_2 \cdot \|h\|_2^2}{\|h\|_2} = 0$

$\square$

Excercise 4. Properties of Eigenvalues and Singular Values

1. Suppose the maximum eigenvalue, minimum eigenvalue of a given symmetric matrix $A \in S^n$ are denoted by $\lambda_{max}(\mathbf{A})$ and $\lambda_{min}(\mathbf{A})$, respectively. Please show that

$$\lambda_{max}(\mathbf{A}) = \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq 0} \frac{\mathbf{x^T A x}}{\mathbf{x^T x}}, \ \lambda_{min}(\mathbf{A}) = \inf_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq 0} \frac{\mathbf{x^T A x}}{\mathbf{x^T x}}. \tag{6}$$

   (**Hint:**consider the orthogonal decomposition of **A**.)

2. Suppose $\mathbf{B} = (b_{ij}) \in \mathbb{R}^{m\times n}$ with maximum singular value $\sigma_{max}(\mathbf{B})$.

   a. Let $\|\mathbf{B}\|_2 := \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq \mathbf{0}} \frac{\|\mathbf{Bx}\|_2}{\|x\|_2}$. Please show that

$$\sigma_{max}(\mathbf{B}) = \|\mathbf{B}\|_2. \tag{7}$$

   b. Please show that

$$\sigma_{max}(\mathbf{B}) = \sup_{\mathbf{x}\in\mathbb{R}^m, \mathbf{y}\in\mathbb{R}^n, \mathbf{x},\mathbf{y}\neq 0} \frac{\mathbf{x^T B y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2} \tag{8}$$

3. Please show the following two equalities:

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))^T(\mathbf{y} - \mathbf{x})\, dt \tag{9}$$

$$\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}) = \int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x})\, dt \tag{10}$$

   (**Hint**: you may consider the function $g(t) = f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))$ and apply the fundamental theorem of calculus.)

4. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, and the gradient of f is Lipschitz continuous, i.e.,
$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \le L\|\mathbf{x} - \mathbf{y}\|_2, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \tag{11}$$
where $L \ge 0$ is Lipschitz constant. Please show that $\lambda_{max}(\nabla^2 f(\mathbf{x})) \le L$ for any $\mathbf{x} \in \mathbb{R}^n$, where $\lambda_{max}(\nabla^2 f(\mathbf{x}))$ is the largest eigenvalue of $\nabla^2 f(\mathbf{x})$.

Proof of Excercise 4: Write down your solutions step by step here.
1.Because A is symmetric matrix,A can be decomposed to $A = Q\mathbf{\Lambda}Q^T$,where Q is an orthogonal matrix, $Q^TQ = QQ^T = I$,$\mathbf{\Lambda} = diag\,\{\lambda_1, ..., \lambda_n\}$,assume $\lambda_1 \ge ... \ge \lambda_n$.
Then let $x = Qy$,we have

$$\frac{\mathbf{x^TAx}}{\mathbf{x^Tx}} = \frac{\mathbf{y^TQ^TQ\Lambda Q^TQy}}{\mathbf{y^TQ^TQy}} = \frac{\mathbf{y^T\Lambda y}}{\mathbf{y^Ty}} = \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2}$$

Select $\mathbf{x}$ that makes $\mathbf{y} = (1, 0, ..., 0)^T$,we have $\sup\limits_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq 0} \frac{\mathbf{x^TAx}}{\mathbf{x^Tx}} = \lambda_1 = \lambda_{max}(\mathbf{A})$.
Similarly,select $\mathbf{x}$ that makes $\mathbf{y} = (0, ..., 0, 1)^T$,we have $\inf\limits_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq 0} \frac{\mathbf{x^TAx}}{\mathbf{x^Tx}} = \lambda_n = \lambda_{min}(\mathbf{A})$.

2.a. w.l.o.g.Assume $n \le m$
Matrix B has an SVD:$B = U\Sigma V^T$,where U, V are both orthogonal matrix,$U^TU = V^TV = I, \Sigma = diag\,\{\sigma_1, ..., \sigma_n\}$,assume $\sigma_1 \ge ... \ge \sigma_n$.
Then let $x = Vy$,we have

$$\|\mathbf{B}\|_2^2 = \frac{\|\mathbf{Bx}\|_2^2}{\|\mathbf{x}\|_2^2} = \frac{\|\mathbf{U\Sigma V^TVy}\|_2^2}{\|\mathbf{Vy}\|_2^2} = \frac{\|\mathbf{U\Sigma y}\|_2^2}{\|\mathbf{Vy}\|_2^2} = \frac{\|\mathbf{\Sigma y}\|_2^2}{\|\mathbf{y}\|_2^2} = \frac{\sum_{i=1}^n \sigma_i^2 y_i^2}{\sum_{i=1}^n y_i^2}$$

Select $\mathbf{x}$ that makes $y = (1, 0, ..., 0)^T$,we have $\|\mathbf{B}\|_2^2 = \sigma_{max}^2(\mathbf{B})$
$\therefore \|\mathbf{B}\|_2 = \sigma_{max}(\mathbf{B})$

2.b.let $x = Ua, y = Vb$,we have

$$\frac{\mathbf{x^TBy}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2} = \frac{\mathbf{a^TU^TU\Sigma V^TVb}}{\|\mathbf{Ua}\|_\mathbf{2}\|\mathbf{Vb}\|_\mathbf{2}} = \frac{\mathbf{a^T\Sigma b}}{\|\mathbf{a}\|_\mathbf{2}\|\mathbf{b}\|_\mathbf{2}} = \frac{\sum_{i=1}^n \sigma_i a_i b_i}{\|\mathbf{a}\|_\mathbf{2}\|\mathbf{b}\|_\mathbf{2}}$$

Recall the Cauchy-Schwarz inequation:

$$|\sum_{i=1}^n \sigma_i a_i b_i| \le \sqrt{\sum_{i=1}^n \sigma_i^2 a_i^2}\sqrt{\sum_{i=1}^n b_i^2} \le \sigma_1\|\mathbf{a}\|_\mathbf{2}\|\mathbf{b}\|_\mathbf{2}$$

$$\therefore \sup_{\mathbf{x}\in\mathbb{R}^m,\mathbf{y}\in\mathbb{R}^n,\mathbf{x},\mathbf{y}\neq 0} \frac{\mathbf{x^TBy}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2} = \sigma_1 = \sigma_{max}(\mathbf{B})$$

3.The fundamental theorem of calculus is:

$$\int_a^b f(x)dx = F(b) - F(a), F(x)' = f(x)$$

Let $g(t) = f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), g(0) = f(\mathbf{x}), g(1) = f(\mathbf{y})$
$\nabla g(t) = \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))^T(\mathbf{y} - \mathbf{x})$

Then apply the fundamental theorem of calculus:

$$g(1) - g(0) = f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \nabla g(t)dt = \int_0^1 \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))^T(\mathbf{y} - \mathbf{x})dt$$

For $\nabla g(t) = \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))^T(\mathbf{y} - \mathbf{x})$,we have
$\nabla g(1) = \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}), \nabla g(0) = \nabla f(\mathbf{y})^T(\mathbf{y} - \mathbf{x})$
Then apply the fundamental theorem of calculus:

$$\nabla g(1) - \nabla g(0) = (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}))^T(\mathbf{y} - \mathbf{x}) = \int_0^1 \nabla^2 g(t)dt = \int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))^T(\mathbf{y} - \mathbf{x})^2 dt$$

Thus,we have

$$\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}) = \int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x})dt$$

4.To show $\lambda_{max}(\nabla^2 f(\mathbf{x})) \leq L$,according to (6),we only need to show $\sup\limits_{\|\mathbf{v}\|=\mathbf{1}} \mathbf{v^T \nabla^2 f(x) v \leq L}$
Similar to the solution to question number 3,let $g(t) = f(\mathbf{x} + t\mathbf{v})$
Then $\nabla g(t) = \nabla f(\mathbf{x} + t\mathbf{v})^T\mathbf{v}, \nabla^2 g(t) = \mathbf{v}^T\nabla^2 f(\mathbf{x} + t\mathbf{v})\mathbf{v}$
f is twice continuously differentiable,so we have:(remember that $\mathbf{v}$ is a unit vector)

$$\forall t, \|\nabla f(\mathbf{x} + t\mathbf{v}) - \nabla f(\mathbf{x})\|_2 \leq Lt$$

On the direction of v,we have:

$$|\nabla f(\mathbf{x} + t\mathbf{v})^T\mathbf{v} - \nabla f(\mathbf{x})^T\mathbf{v}| \leq Lt$$
$$|\nabla g(t) - \nabla g(0)| \leq Lt$$

Applying the Lagrange's Mean Value Theorem:

$$\exists \xi \in (0, t), |\nabla^2 g(\xi)| = \frac{|\nabla g(t) - \nabla g(0)|}{t} \leq L$$

$$\therefore \forall \mathbf{v}, \mathbf{v}^T\nabla^2 f(\mathbf{x} + t\mathbf{v})\mathbf{v} \leq L$$

Let $t \to 0$,$\lambda_{max}(\nabla^2 f(\mathbf{x})) = \sup\limits_{\|\mathbf{v}\|=\mathbf{1}} \mathbf{v^T \nabla^2 f(x) v \leq L}$

$\square$