Assignment #1

Shao Hua, Huang 0750727
ECM5901 - Optimization Theory and Application

December 27, 2019

Exercise 1. Verify that the ℓ_1-norm on $\mathbb{R}^n$ defined by $\|x\|_1 = \Sigma_{i=1}^n |x_i|$ for $x \in \mathbb{R}^n$ is a norm on $\mathbb{R}^n$. (10%)

Verify. Check the four properties of a *norm* function.

(1) nonnegative: $\|x\|_1 = \sum_{i=1}^{n} |x_i| \geq 0 \quad \forall x \in \mathbb{R}^n$

(2) definite: $\|x\|_1 = \sum_{i=1}^{n} |x_i| = 0 \Leftrightarrow x_1 = x_2 = \cdots = x_n = 0 \Leftrightarrow x = 0$

(3) homogeneous:

$$\|tx\|_1 = \sum_{i=1}^{n} |tx_i| = \sum_{i=1}^{n} |t||x_i| = |t| \sum_{i=1}^{n} |x_i| = |t|\|x\|_1 \quad \forall x \in \mathbb{R}^n, t \in \mathbb{R}$$

(4) triangle inequality:

$$\|x+y\|_1 = \sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} (|x_i| + |y_i|) = \|x\|_1 + \|y\|_1 \quad \forall x, y \in \mathbb{R}^n$$

Therefore, ℓ_1-norm is a *norm.* □

Exercise 2. The operator norm on $\mathbb{R}^{m \times n}$ induced by two norms $\|\cdot\|_a$ on $\mathbb{R}^m$ and $\|\cdot\|_b$ on $\mathbb{R}^n$ is defined by

$$\|X\|_{a,b} = \sup\{\|Xu\|_a \mid \|u\|_b \leq 1\}$$

for $X \in \mathbb{R}^{m \times n}$. Verify that

$$\|X\|_{1,1} = \max_{j=1,\ldots,n} \sum_{i=1}^{n} |X_{ij}|.$$

(10%)

Verify. Separate the verify into two parts.

(a) $\|X\|_{1,1} \leq \max_{j=1,\ldots,n} \sum_{i=1}^{n} |X_{ij}|$

Let $X = [X_1 \quad X_2 \quad \ldots \quad X_n]$ where $X_1, X_2, \ldots, X_n$ are column vectors of X

$$\begin{aligned}
\|X\|_{1,1} &= \sup\{\|Xu\|_1 \mid \|u\|_1 \leq 1\} \\
&= \sup\left\{\|\sum_{i=1}^{n} X_i u_i\|_1 \mid \|u\|_1 \leq 1\right\} \\
&\leq \sup\left\{\sum_{i=1}^{n} \|X_i u_i\|_1 \mid \|u\|_1 \leq 1\right\} \quad \text{(triangle inequality of norm)} \\
&= \sup\left\{\sum_{i=1}^{n} \|X_i\|_1 |u_i| \mid \|u\|_1 \leq 1\right\} \quad \text{(homogeneous of norm)} \\
&\leq \sup\left\{\max\{\|X_1\|_1, \|X_2\|_1, \ldots, \|X_n\|_1\} \sum_{i=1}^{n} |u_i| \mid \|u\|_1 \leq 1\right\} \\
&= \sup\{\max\{\|X_1\|_1, \|X_2\|_1, \ldots, \|X_n\|_1\}\|u\|_1 \mid \|u\|_1 \leq 1\} \\
&= \max\{\|X_1\|_1, \|X_2\|_1, \ldots, \|X_n\|_1\} \\
&= \max\left\{\sum_{i=1}^{n} |X_{i1}|, \sum_{i=1}^{n} |X_{i2}|, \ldots, \sum_{i=1}^{n} |X_{in}|\right\} = \max_{j=1,\ldots,n} \sum_{i=1}^{n} |X_{ij}|
\end{aligned}$$

(b) $\|X\|_{1,1} \geq \max_{j=1,\ldots,n} \sum_{i=1}^{n} |X_{ij}|$

$$\|X\|_{1,1} = \sup\{\|Xu\|_1 \mid \|u\|_1 \leq 1\} \geq \|Xu\|_1$$

We choose $u = e_k$ where $k = \arg\max_j \left\{\sum_{i=1}^{n} |X_{i1}|, \ldots, \sum_{i=1}^{n} |X_{ij}|, \ldots, \sum_{i=1}^{n} |X_{in}|\right\}$

Therefore, $\|u\|_1 = 1$ is under the constraint, and

$$\begin{aligned}
\|X\|_{1,1} &\geq \|Xu\|_1 \\
&= |X_{1k}| + |X_{2k}| + \cdots + |X_{nk}| \\
&= \sum_{i=1}^{n} |X_{ik}| \\
&= \max\left\{\sum_{i=1}^{n} |X_{i1}|, \sum_{i=1}^{n} |X_{i2}|, \ldots, \sum_{i=1}^{n} |X_{in}|\right\} = \max_{j=1,\ldots,n} \sum_{i=1}^{n} |X_{ij}|
\end{aligned}$$

(c) From (a) and (b), we know that $\|X\|_{1,1} = \max_{j=1,\ldots,n} \sum_{i=1}^{n} |X_{ij}|$. $\square$

Exercise 3. Show that the dual norm of ℓ_1-norm is the ℓ_∞-norm. (10%)

Proof. Separate the proof into two parts. $(z, x \in \mathbb{R}^n)$

(a) $\|z\|_* \le \|z\|_\infty$

$$\begin{aligned}
\|z\|_* &= \sup\{z^T x \mid \|x\|_1 \le 1\} \\
&= \sup\left\{\sum_{i=1}^{n}(z_i x_i) \mid \|x\|_1 \le 1\right\} \\
&\le \sup\left\{\sum_{i=1}^{n}(|z_i||x_i|) \mid \|x\|_1 \le 1\right\} \\
&= \sup\left\{\max\{z_1, z_2, \ldots, z_n\}\sum_{i=1}^{n}|x_i| \mid \|x\|_1 \le 1\right\} \\
&= \max\{z_1, z_2, \ldots, z_n\} = \|z\|_\infty
\end{aligned}$$

(b) $\|z\|_* \ge \|z\|_\infty$

$$\|z\|_* = \sup\{z^T x \mid \|x\|_1 \le 1\} \ge z^T x$$

We choose $x = \mathbf{sign}(z_k)e_k$ where $k = \arg\max_i\{z_1, \ldots, z_i, \ldots, z_n\}$
Therefore, $\|x\|_1 = 1$ is under the constraint, and

$$\begin{aligned}
\|z\|_* &\ge z^T x \\
&= z^T\,\mathbf{sign}(z_k)e_k = |z_k| \\
&= \max\{z_1, z_2, \ldots, z_n\} = \|z\|_\infty
\end{aligned}$$

(c) From (a) and (b), we know that $\|z\|_* = \|z\|_\infty$. □

Exercise 4. The trace of a square matrix $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ is defined as $\mathbf{tr}(A) = \sum_{i=1}^{n} a_{ii}$. Show that

(a) $\mathbf{tr}(AB) = \mathbf{tr}(BA)$ for $A, B \in \mathbb{R}^{n\times n}$. (5%)

(b) $\mathbf{tr}(tA + B) = t\,\mathbf{tr}(A) + \mathbf{tr}(B)$ for $A, B \in \mathbb{R}^{n\times n}$ and $t \in \mathbb{R}$. (5%)

Proof.

(a) $\mathbf{tr}(AB) = \sum_{i=1}^{n}\left(\sum_{k=1}^{n} a_{ik}b_{ki}\right) = \sum_{k=1}^{n}\left(\sum_{i=1}^{n} b_{ki}a_{ik}\right) = \mathbf{tr}(BA)$

(b) $\mathbf{tr}(tA+B) = \sum_{i=1}^{n}(ta_{ii}+b_{ii}) = t\sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n} b_{ii} = t\,\mathbf{tr}(A) + \mathbf{tr}(B)$ $\square$

Exercise 5. Let $\langle , \rangle$ be the inner product on $\mathbb{R}^n$. Prove the Cauchy-Schwarz inequality that for $x, y \in \mathbb{R}^n$, $|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle$. (10%)

Proof. We can easily know that $\langle x, x\rangle = |x|^2 = \|x\|_2^2$ where $x \in \mathbb{R}^n$.
The proof starts from equation $\|x - \frac{\langle x,y\rangle}{\|y\|_2^2} y\|_2^2 \geq 0$, and then we have

$$\begin{aligned}
\|x - \frac{\langle x,y\rangle}{\|y\|_2^2} y\|_2^2 &= \langle x - \frac{\langle x,y\rangle}{\|y\|_2^2} y, x - \frac{\langle x,y\rangle}{\|y\|_2^2} y\rangle \\
&= \langle x, x\rangle - 2\frac{\langle x,y\rangle}{\|y\|_2^2}\langle x, y\rangle + \frac{\langle x,y\rangle\langle x,y\rangle}{\|y\|_2^4}\langle y, y\rangle \\
&= \langle x, x\rangle - 2\frac{|\langle x,y\rangle|^2}{\|y\|_2^2} + \frac{|\langle x,y\rangle|^2}{\|y\|_2^2} \\
&= \langle x, x\rangle - \frac{|\langle x,y\rangle|^2}{\|y\|_2^2} \\
&= \langle x, x\rangle - \frac{|\langle x,y\rangle|^2}{\langle y, y\rangle} \geq 0
\end{aligned}$$

Therefore, we can show that $|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle$ $\square$

Exercise 6. Verify that Frobenious inner product on $\mathbb{R}^{m\times n}$ defined by

$$\langle X, Y\rangle_F = \mathbf{tr}(X^T Y)$$

for $X, Y \in \mathbb{R}^{m\times n}$ is an inner product. (10%)

Verify. Check the three properties of *inner product.*

(1) conjugate symmetry: Since $X, Y \in \mathbb{R}^{m\times n}$, we have $\overline{\langle X, Y\rangle}_F = \langle X, Y\rangle_F$.

$$\langle X, Y\rangle_F = \mathbf{tr}(X^T Y) = \mathbf{tr}((X^T Y)^T) = \mathbf{tr}(Y^T X) = \langle Y, X\rangle_F = \overline{\langle Y, X\rangle}_F$$

(2) linearity: $X, Y, Z \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}$

(i) $\langle X+Z, Y\rangle_F = \mathbf{tr}((X+Z)^T Y) = \mathbf{tr}((X^T + Z^T)Y) = \mathbf{tr}(X^T Y + Z^T Y)$
$= \mathbf{tr}(X^T Y) + \mathbf{tr}(Z^T Y) = \langle X, Y\rangle_F + \langle Z, Y\rangle_F$

(ii) $\langle cX, Y\rangle_F = \mathbf{tr}((cX)^T Y) = \mathbf{tr}(cX^T Y) = c\ \mathbf{tr}(X^T Y) = c\langle X, Y\rangle_F$

(3) positive-definite: $\langle X, X\rangle_F = \mathbf{tr}(X^T X) = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}^2 > 0$ where $X \in \mathbb{R}^{m\times n} \setminus \{0\}$

Therefore, *Frobenious inner product* is an inner product $\square$

Exercise 7. Read Appendix A.4. Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Consider $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(x) = \|Ax - b\|_2^2$ for $x \in \mathbb{R}^n$. Show that

(a) $\nabla f(x) = 2A^T(Ax - b)$. (10%)

(b) $\nabla^2 f(x) = 2A^T A$. (10%)

Lemma 1. $Df(x) = A$ where $f(x) = Ax + b$ and $A \in \mathbb{R}^{m\times n} \quad x, b \in \mathbb{R}^n$
Let $A = [A_1^T \quad A_2^T \quad \dots \quad A_n^T]$ where $A_1, A_2, \dots, A_n$ are row vectors of A.
We have $Df(x)_{ij} = \dfrac{\partial f_i(x)}{\partial x_j} = \dfrac{\partial(A_i x + b_i)}{\partial x_j} = A_{ij} \Rightarrow Df(x) = A$

Lemma 2. $Df(x) = 2x^T$ where $f(x) = x^T x$ and $x \in \mathbb{R}^n$
We have $Df(x)_i = \dfrac{\partial f(x)}{\partial x_i} = \dfrac{\partial(\sum_{j=1}^n x_j^2)}{\partial x_i} = 2x_i \Rightarrow Df(x) = 2x^T$

Proof. Back to exercise, we have $f(x) = \|Ax - b\|_2^2 = (Ax - b)^T(Ax - b)$.
Employ chain rule to prove these equations.

(a) $\nabla f(x) = Df(x)^T = \left(\dfrac{\partial f(x)}{\partial(Ax - b)} \dfrac{\partial(Ax - b)}{\partial x}\right)^T = (2(Ax - b)^T A))^T = 2A^T(Ax - b)$

(b) $\nabla^2 f(x) = D\nabla f(x) = \dfrac{\partial(2A^T(Ax - b))}{\partial x} = \dfrac{\partial(2A^T Ax - 2A^T b)}{\partial x} = 2A^T A$ $\square$

Exercise 8. For a symmetric matrix $A \in \mathbb{R}^{n\times n}$, it has an eigenvalue (spectral) decomposition

$$A = Q\Lambda Q^T,$$

where $Q \in \mathbb{R}^{n\times n}$ is an orthogonal matrix such that $Q^T Q = QQ^T = I$ and $\Lambda = \mathbf{diag}(\lambda_1, \dots, \lambda_n)$ is a diagonal matrix with entries that are eigenvalues of A.

(a) Show that $\mathbf{tr}(A) = \displaystyle\sum_{i=1}^n \lambda_i$. (5%)

(b) Show that $x^T Ax \geq 0$ for any $x \in \mathbb{R}^n$ if and only if $\lambda_i \geq 0$ for $i = 1, \dots, n$. (A is called positive semidefinite if all the eigenvalues λ_i are nonnegative.) (10%)

(c) Let $\lambda_{\max} = \max\{\lambda_1, \dots, \lambda_n\}$ and $\lambda_{\min} = \min\{\lambda_1, \dots, \lambda_n\}$. Show that

$$\lambda_{\min} x^T x \leq x^T Ax \leq \lambda_{\max} x^T x.$$

(10%)

Proof.

(a) eigenvalues are roots of the characteristic equation, *i.e.*, roots of

$$\begin{aligned}
\det(A - \lambda I) &= \begin{vmatrix} (a_{11} - \lambda) & a_{12} & \cdots & a_{1n} \\ a_{21} & (a_{22} - \lambda) & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & (a_{nn} - \lambda) \end{vmatrix} \\
&= (-1)^n \lambda^n + (-1)^{n-1} \left(\sum_{i=1}^{n} a_{ii} \right) \lambda^{n-1} + \cdots + \det(A) \\
&= (-1)^n \lambda^n + (-1)^{n-1} \mathbf{tr}(A) \lambda^{n-1} + \cdots + \det(A) \qquad (1) \\
&= 0
\end{aligned}$$

We can rewrite $\det(A - \lambda I) = 0$ by its roots $\lambda_1, \ldots, \lambda_n$ with the information of term $(-1)^n \lambda^n$.

$$\begin{aligned}
\det(A - \lambda I) &= (\lambda_1 - \lambda)(\lambda_2 - \lambda) \ldots (\lambda_n - \lambda) \\
&= (-1)^n \lambda^n + (-1)^{n-1} \left(\sum_{i=1}^{n} \lambda_i \right) \lambda^{n-1} + \cdots + \prod_{i=1}^{n} \lambda_i \qquad (2)
\end{aligned}$$

Compare (1) with (2), we know that $\mathbf{tr}(A) = \sum_{i=1}^{n} \lambda_i$.

(b) Separate the proof into two parts.

(i) Proposition of $\rightarrow$
Choose x to be eigenvectors of A, *i.e.*, $v_i, i = 1, 2, \ldots, n$
Then $Av_i = \lambda_i v_i$ holds where λ_i represents corresponding eigenvalue

$$x^T A x = v_i^T A v_i = v_i^T \lambda_i v_i = \lambda_i \|v_i\|_2^2 \geq 0 \Rightarrow \lambda_i \geq 0$$

$x^T A x \geq 0$ for any $x \in \mathbb{R}^n \Rightarrow \lambda_i \geq 0$ for $i = 1, \ldots, n$.

(ii) Proposition of $\leftarrow$
Let $y = Q^T x$, *i.e.*, $x = Qy$

$$x^T A x = (Qy)^T A (Qy) = y^T Q^T Q \Lambda Q^T Q y = y^T \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2$$

We know that all $\lambda_i \geq 0$; therefore, $x^T A x \geq 0$
$x^T A x \geq 0$ for any $x \in \mathbb{R}^n \Leftarrow \lambda_i \geq 0$ for $i = 1, \ldots, n$.

(iii) From (i) and (ii), we get
$x^T A x \geq 0$ for any $x \in \mathbb{R}^n \Leftrightarrow \lambda_i \geq 0$ for $i = 1, \ldots, n$.

(c) Separate the proof into two parts. Similar to (b)(ii), let $y = Q^T x$, *i.e.*, $x = Qy$.

(i) $\lambda_{\min} x^T x \le x^T A x$

$$x^T A x = (Qy)^T A(Qy) = y^T Q^T Q \Lambda Q^T Q y = y^T \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2$$

$$\lambda_{\min} x^T x = \lambda_{\min} (Qy)^T (Qy) = \lambda_{\min} y^T Q^T Q y = \lambda_{\min} y^T y = \lambda_{\min} \sum_{i=1}^{n} y_i^2$$

It is easy to show that $\lambda_{\min} \sum_{i=1}^{n} y_i^2 \le \sum_{i=1}^{n} \lambda_i y_i^2 \Rightarrow \lambda_{\min} x^T x \le x^T A x$.

(ii) $x^T A x \le \lambda_{\max} x^T x$

$$\lambda_{\max} x^T x = \lambda_{\max} (Qy)^T (Qy) = \lambda_{\max} y^T Q^T Q y = \lambda_{\max} y^T y = \lambda_{\max} \sum_{i=1}^{n} y_i^2$$

Similar to (i), $\sum_{i=1}^{n} \lambda_i y_i^2 \le \lambda_{\max} \sum_{i=1}^{n} y_i^2 \Rightarrow x^T A x \le \lambda_{\max} x^T x$.

(iii) From (i) and (ii), we get $\lambda_{\min} x^T x \le x^T A x \le \lambda_{\max} x^T x$. $\square$