Math 164 Homework 3

Nick Monozon

Problem 4.1

Proof. ($\Rightarrow$) Suppose that $\mathcal{S} = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$ is a linear variety. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$ and $\alpha \in \mathbb{R}$. Then we have that

$$\boldsymbol{A}\big(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}\big) = \alpha\boldsymbol{A}\boldsymbol{x} + (1-\alpha)\boldsymbol{A}\boldsymbol{y} = \alpha\boldsymbol{b} + (1-\alpha)\boldsymbol{b} = \boldsymbol{b}.$$

So $\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in \mathcal{S}$.

($\Leftarrow$) If $\mathcal{S} = \emptyset$, then we're done. So suppose that $\boldsymbol{x}_0 \in \mathcal{S}$. Define the set $\mathcal{S}' = \mathcal{S} - \boldsymbol{x}_0 := \{\boldsymbol{x} - \boldsymbol{x}_0 : \boldsymbol{x} \in \mathcal{S}\}$. Then for every $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}'$ and $\alpha \in \mathbb{R}$, we have that $\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in \mathcal{S}'$. We will now show that $\mathcal{S}'$ is a subspace. This time, let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}'$, $\alpha \in \mathbb{R}$ where $\boldsymbol{x} = \bar{\boldsymbol{x}} - \boldsymbol{x}_0$ and $\boldsymbol{y} = \bar{\boldsymbol{y}} - \boldsymbol{x}_0$ for $\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}} \in \mathcal{S}'$. We thus have that

$$\begin{aligned}
\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} &= \alpha(\bar{\boldsymbol{x}} - \boldsymbol{x}_0) + (1-\alpha)(\bar{\boldsymbol{y}} - \boldsymbol{x}_0) \\
&= \underbrace{\alpha\bar{\boldsymbol{x}} + (1-\alpha)\bar{\boldsymbol{y}}}_{\in\mathcal{S}} - \boldsymbol{x}_0 \\
&\in \mathcal{S}'.
\end{aligned}$$

Now observe that $\boldsymbol{0} \in \mathcal{S}'$ because $\boldsymbol{x}_0 - \boldsymbol{x}_0 = \boldsymbol{0} \in \mathcal{S}'$. Furthermore, if $\boldsymbol{x} \in \mathcal{S}'$ and $a \in \mathbb{R}$, we have that

$$a\boldsymbol{x} = a\boldsymbol{x} + (1-a)\boldsymbol{0} \in \mathcal{S}'.$$

Now, given $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}'$, we have from the previous result that

$$\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y} \in \mathcal{S}' \Rightarrow \quad 2\Big(\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}\Big) \in \mathcal{S}'.$$

As such, we have shown that $\mathcal{S}'$ is a subspace by demonstrating that it is closed under scalar multiplication, closed under addition, and contains the zero vector. From the results of Exercise 3.13, there exists a matrix $\boldsymbol{A}$ such that $\mathcal{S}' = \mathcal{N}(\boldsymbol{A}) = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$. From this, we show that

$$\begin{aligned}
\mathcal{S} = \mathcal{S}' + \boldsymbol{x}_0 &= \{\boldsymbol{x} + \boldsymbol{x}_0 : \boldsymbol{x} \in \mathcal{S}'\} \\
&= \{\boldsymbol{x} + \boldsymbol{x}_0 : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\} \\
&= \{\boldsymbol{x} + \boldsymbol{x}_0 : \boldsymbol{A}(\boldsymbol{x} + \boldsymbol{x}_0) = \boldsymbol{b}\} \\
&= \{\boldsymbol{y} : \boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}\}.
\end{aligned}$$

where we let $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{x}_0$. Therefore, $\mathcal{S}$ is a linear variety. This concludes the proof. $\square$

Problem 4.2

Proof. Let $\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| \le r\}$ for $r > 0$ and $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^\top\boldsymbol{x}}$ (the standard Euclidian norm). Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$ and $\alpha \in [0,1]$ and suppose that $\boldsymbol{z} = \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}$. We must show that $\|\boldsymbol{z}\| \le r$. We have that

$$\begin{aligned}
\|\boldsymbol{z}\|^2 &\le \big(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}\big)^\top\big(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}\big) \\
&= \big(\alpha\boldsymbol{x}^\top + (1-\alpha)\boldsymbol{y}^\top\big)\big(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}\big) \\
&= \alpha^2\boldsymbol{x}^\top\boldsymbol{x} + 2\alpha(1-\alpha)\boldsymbol{x}^\top\boldsymbol{y} + (1-\alpha)^2\boldsymbol{y}^\top\boldsymbol{y} \\
&= \alpha^2\|\boldsymbol{x}\|^2 + 2\alpha(1-\alpha)\boldsymbol{x}^\top\boldsymbol{y} + (1-\alpha)^2\|\boldsymbol{y}\|^2.
\end{aligned}$$

Since $\|x\|^2, \|y\|^2 \le r^2$ and $x^\top y \le \|x\|\|y\| \le r^2$ by the Cauchy-Schwarz Inequality, it follows that

$$\begin{aligned} \|z\|^2 &\le \alpha^2 r^2 + 2\alpha(1-\alpha) r^2 + (1-\alpha)^2 r^2 \\ &= r^2(\alpha^2 + 2\alpha - 2\alpha^2 + 1 - 2\alpha + \alpha^2) \\ &= r^2(\alpha^2 + \alpha^2 - 2\alpha^2 + 2\alpha - 2\alpha + 1) \\ &= r^2. \end{aligned}$$

Taking the square root of both sides, we find that $\|z\| \le r$. Therefore, $z \in \mathcal{C}$ and hence $\mathcal{C}$ is a convex set. □

Problem 4.4

Proof. Let $\mathcal{C} = \{x \in \mathbb{R}^n : x \ge \mathbf{0}\}$. Let $x, y \in \mathcal{C}$ and $\alpha \in [0,1]$ and suppose that $z = \alpha x + (1-\alpha) y$. To show that this set is convex, we must show that $z \in \mathcal{C}$. For $x, y \in \mathcal{C}$, we can write $x = (x_1, x_2, \ldots, x_n)^\top$ and $y = (y_1, y_2, \ldots, y_n)^\top$ where $x_i, y_i \ge 0$ for each $i = 1, 2, \ldots, n$. We can also define $z = (z_1, z_2, \ldots, z_n)^\top$ where for each $i = 1, 2, \ldots, n$, the ith component is given by

$$z_i = \alpha x_i + (1-\alpha) y_i.$$

Since $0 \le \alpha \le 1$, we have that $\alpha, 1-\alpha \ge 0$. From this, we can clearly see that for every i

$$z_i = \underbrace{\alpha x_i}_{\ge 0} + \underbrace{(1-\alpha) y_i}_{\ge 0} \ge 0.$$

Therefore, $z \ge \mathbf{0}$, so $z \in \mathcal{C}$ and hence $\mathcal{C}$ is a convex set. □

Problem 5.1

Proof. We have that

$$\|A^k\| \le \|A^{k-1}\|\|A\| \le \|A^{k-2}\|\|A\|^2 \le \cdots \le \|A\|^k.$$

Thus, if $\|A\| < 1$, we have that $\lim_{k\to\infty} \|A^k\| = O$, which hence implies that $\lim_{k\to\infty} A^k = O$. As such, $\|A\| < 1$ is a sufficient condition for $\lim_{k\to\infty} A^k = O$. □

Problem 5.2

Proof. Let $A \in \mathbb{R}^{n\times n}$. We will first show that this statement is true when A has all real eigenvalues. Let λ be the eigenvalue associated with the normalized eigenvector v^* such that λ has the largest absolute value. By definition, we have that $Av^* = \lambda v^*$ and $\|v^*\| = 1$. It follows that

$$\|A\| \ge \|Av^*\| = \|\lambda v^*\| = |\lambda|\|v^*\| = |\lambda|.$$

This completes the proof. However, the eigenvalues of A may be complex, so we must consider this case as well. Consider the matrix

$$M = \frac{A}{\|A\| + \epsilon}$$

where $\epsilon > 0$. By definition, we have that

$$M = \frac{A}{\|A\| + \epsilon} < 1.$$

Using Exercise 5.1, $\lim_{k\to\infty} M^k = O$, so by Lemma 5.1, we have that $|\lambda_i(M)| < 1$ for each $i = 1, 2, \ldots, n$. Furthermore, for each $i = 1, 2, \ldots, n$, it follows that

$$\lambda_i(M) = \frac{\lambda_i(A)}{\|A\| + \epsilon}$$

and from this that

$$|\lambda_i(M)| = \frac{|\lambda_i(A)|}{\|A\| + \epsilon} < 1 \Rightarrow \quad |\lambda_i(A)| < \|A\| + \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we therefore have that $\lambda_i(A)| \le \|A\|$. Lastly, since this inequality holds for each $i = 1, 2, \ldots, n$, we have that

$$\|A\| \ge \max_{1\le i\le n} |\lambda_i(A)|.$$

This completes the proof.

□

Problem 5.3

(a) We are given the function

$$f(\boldsymbol{x}) = (\boldsymbol{a}^\top \boldsymbol{x})(\boldsymbol{b}^\top \boldsymbol{x})$$

where $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{x}$ are n-dimensional vectors. We find the gradient to be

$$\begin{aligned}\nabla f(\boldsymbol{x}) &= \boldsymbol{b}\boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{a}\boldsymbol{b}^\top \boldsymbol{x} \\ &= (\boldsymbol{a}\boldsymbol{b}^\top + \boldsymbol{b}\boldsymbol{a}^\top)\boldsymbol{x}.\end{aligned}$$

(b) Given the gradient from part (a), which we computed to be

$$\nabla f(\boldsymbol{x}) = (\boldsymbol{a}\boldsymbol{b}^\top + \boldsymbol{b}\boldsymbol{a}^\top)\boldsymbol{x},$$

we find the Hessian matrix to be

$$\boldsymbol{F}(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x}) = \boldsymbol{a}\boldsymbol{b}^\top + \boldsymbol{b}\boldsymbol{a}^\top.$$

Problem 5.6

We find that

$$Df(\boldsymbol{x}) = \left[3x_1^2 x_2 x_3^2 + x_2, x_1^3 x_3^2 + x_1, 2x_1^3 x_2 x_3 + 1\right]$$

and

$$\frac{d\boldsymbol{x}(t)}{dt} = \begin{bmatrix} e^t + 3t^2 \\ 2t \\ 1 \end{bmatrix}.$$

Using the Chain Rule, we have that

$$\begin{aligned}\frac{d}{dt} f(\boldsymbol{x}(t)) &= Df(\boldsymbol{x}(t))\frac{d\boldsymbol{x}(t)}{dt} \\ &= \left[3x_1(t)^2 x_2(t) x_3(t)^2 + x_2(t), x_1(t)^3 x_3(t)^2 + x_1(t), 2x_1(t)^3 x_2(t) x_3(t) + 1\right] \begin{bmatrix} e^t + 3t^2 \\ 2t \\ 1 \end{bmatrix} \\ &= 12t(e^t + 3t^2)^3 + 2te^t + 6t^2 + 2t + 1.\end{aligned}$$

Problem 5.10

(a) We have that

$$f(\boldsymbol{x}) = x_1 e^{-x_2} + x_2 + 1,$$

so we compute

$$Df(\boldsymbol{x}) = \left[e^{-x_2}, -x_1 e^{-x_2} + 1\right]$$

and

$$Df^2(\boldsymbol{x}) = \begin{bmatrix} 0 & -e^{-x_2} \\ -e^{-x_2} & x_1 e^{-x_2} \end{bmatrix}.$$

So the Taylor expansion about $\boldsymbol{x}_0 = [1, 0]^\top$ is

$$\begin{aligned}f(\boldsymbol{x}) &= f(\boldsymbol{x}_0) + \frac{1}{1!} Df(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \frac{1}{2!}(\boldsymbol{x} - \boldsymbol{x}_0)^\top D^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \cdots \\ &= 2 + [1, 0][x_1 - 1, x_2]^\top + \frac{1}{2}[x_1 - 1, x_2] \begin{bmatrix} 0 & -1 \\ -1 & 1 \end{bmatrix} [x_1 - 1, x_2]^\top + \cdots \\ &= 2 + x_1 - 1 - x_1 x_2 + x_2 + \frac{1}{2} x_2^2 + \cdots \\ &= 1 + x_1 + x_2 - x_1 x_2 + \frac{1}{2} x_2^2 + \cdots.\end{aligned}$$

(b) We are given that

$$f(\boldsymbol{x}) = x_1^4 + 2x_1^2 x_2^2 + x_2^4.$$

Hence, we find that

$$Df(\boldsymbol{x}) = \left[4x_1^3 + 4x_1 x_2^2, 4x_2^3 + 4x_1^2 x_2\right]$$

and

$$D^2 f(\boldsymbol{x}) = \begin{bmatrix} 12x_1^2 + 4x_2^2 & 8x_1 x_2 \\ 8x_1 x_2 & 12x_2^2 + 4x_1^2 \end{bmatrix}.$$

So the Taylor expansion about $\boldsymbol{x}_0 = [1,1]^\top$ is

$$\begin{aligned} f(\boldsymbol{x}) &= f(\boldsymbol{x}_0) + \frac{1}{1!} Df(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \frac{1}{2!}(\boldsymbol{x} - \boldsymbol{x}_0)^\top D^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \cdots \\ &= 4 + [8,8]\left[x_1 - 1, x_2 - 1\right]^\top + \frac{1}{2}\left[x_1 - 1, x_2 - 1\right] \begin{bmatrix} 16 & 8 \\ 8 & 16 \end{bmatrix} \left[x_1 - 1, x_2 - 1\right]^\top \\ &= 8x_1^2 + 8x_2^2 - 16x_1 - 16x_2 + 8x_1 x_2 + 12 + \cdots. \end{aligned}$$