Homework

PB22010344 黄境

2024 年 10 月 8 日

Exercise 1. **Bolzano-Weierstrass Theorem**

Solution.

1. Proof by contradiction.

($\Rightarrow$)Suppose that $\exists\epsilon_0$, s.t. $\forall a \in C, a + \epsilon_0 < u$, thus $u - \epsilon_0$ is also an upper bound of C, which has $u - \epsilon_0 < u$, contradicting to the assumption that u is the least upper bound of C.

($\Leftarrow$)Suppose that $u' < u$ is an upper bound of C, let $\epsilon_0 = u' - u$, thus $\forall a \in C, a + \epsilon_0 = a + u' - u = u' - (u - a) \leq u'$, contradicting to the assumption that $\forall\epsilon > 0, \exists a \in C$, s.t. $a > u - \epsilon$.

2. If the term $a_n > a_m, \forall m > n$, we call it a head. If $\{a_n\}$ contains infinite heads, then the sequence of heads is a decreasing subsequence of $\{a_n\}$, or if $\{a_n\}$ only have finite heads, let a_{i_1} be the next term of the last head, then there must exist a_{i_2} s.t. $i_2 > i_1$ and $a_{i_1} \leq a_{i_2}$. Since a_{i_2} is also not a head, we can find a_{i_3} s.t. $i_2 > i_3$ and $a_{i_2} \leq a_{i_3}$. Follow the step and we have a increasing subsquence $\{a_{i_n}\}$. So we have proved that every squence has a monotune subsequence.

Now for bounded sequence $\{a_n\}$, without loss of generality, we assume that it has an increasing subsequence $\{a_{i_n}\}$, then $\{a_{i_n}\}$ is also bounded. By the least upper bound axiom, let $c = sup\{a_{i_n}\}$, then $c \in [a, b]$, and we have $\forall \epsilon > 0, \exists a_{i_N}$, s.t. $c - \epsilon < a_{i_N}$, which implies $c - \epsilon < a_{i_M}, \forall M > N$, for the sequence is increasing. That is $0 \leq c - a_{i_M} < \epsilon, \forall M > N$, which implies $\lim\limits_{n\to\infty} a_{i_n} = c$.

Exercise 2. **Limit and Limit Points**

Solution.

1.($\Rightarrow$) Since $\lim\limits_{n\to\infty} x_n = x$, let $\epsilon = 1$, $\exists N$, s.t. $\forall n > N, \|x_n - x\| \leq 1$. Let $r = 1 + max\{\|x_1 - x\|, \|x_2 - x\|, ..., \|x_N - x\|, 1\}$, then $\{x_n\} \in B_r(x)$, thus $\{x_n\}$ is bounded. Obviously x is a limit point of $\{x_n\}$, and if x' is another limit point of $\{x_n\}$, then $\lim\limits_{n\to\infty} x_n = x' = x$. Thus x is unique.

($\Leftarrow$) Suppose that $\exists \{y_n\}$ is a subsequence of $\{x_n\}$, s.t. $\lim\limits_{n\to\infty} y_n$ does not converge to x. Since x is the unique limit point, y_n must be divergent. By Bolzano-Weierstrass theorem, $\exists A$ and a subsequence $\{y_{k_n}\}$ of y_n, s.t. $\lim\limits_{n\to\infty} y_{k_n} = A$. By the definition of limit, $\exists \epsilon$, s.t. $B^c_\epsilon(A)$ has infinite terms of y_n. These terms are also a bounded sequence, thus by the Bolzano-Weierstrass theorem, $\exists \{y_{s_n}\}$, s.t. $\lim\limits_{n\to\infty} y_{s_n} = B \neq A$, contradicting to the assumption that x is the unique limit point of $\{x_n\}$. Thus we have $\lim\limits_{n\to\infty} x_n = x$.

2.(a) Since $\{2^{-n}\}_{n=1}^\infty \in C$, $\{1 - 2^{-n}\}_{n=1}^\infty \in C$, $\lim\limits_{n\to\infty} 2^{-n} = 0$, $\lim\limits_{n\to\infty} 1 - 2^{-n} = 1$, and $\forall n$, $0 \neq 2^{-n}$, $1 \neq 1 - 2^{-n}$, thus 0 and 1 are limit points of C. $\forall x \in (0, 1)$, let $d = min\{x, 1 - x\}$, then $\{x + 2^{-n}d\}_{n=1}^\infty \in C$, s.t. $x + 2^{-n}d \to x$ and $x + 2^{-n}d \neq x$, thus x is also a limit point of C. By definition, x is a limit point of C $\Leftrightarrow \forall \epsilon > 0, B_\epsilon(x) \bigcap C \setminus \{x\} \neq \emptyset$. Thus $\forall x \notin (0, 1)$, let $d = min\{\|x\|, \|1 - x\|\}$, then $B_{\frac{d}{2}}(x) \bigcap C \setminus \{x\} = \emptyset$. That is, $C' = [0, 1]$, and $\{2\} = C \setminus C'$ is an isolated point of C.

(b) C' is closed $\Leftrightarrow (C')^c$ is open. $\forall a \in (C')^c$, Since a is not a limit point of C, $\exists B_r(a)$, s.t. $B_r(a)\bigcap C=\emptyset$. Thus $\forall b$ in $B_r(a)$, $b\notin C'$, which implies $B_r(a)\subset (C')^c$, thus $(C')^c$ is open, thus C' is closed.

EXERCISE 3. **Norms**

SOLUTION. 1.(a) Obviously l_p is nonnegative and definite. Since $\|\alpha\mathbf{x}\| = (\sum_{i=1}^n(|\alpha x_i|)^p)^{\frac{1}{p}} = \alpha(\sum_{i=1}^n|x_i|^p)^{\frac{1}{p}} = \alpha\|\mathbf{x}\|$, l_p is homogeneous. By the Minkowski inequality, l_p satisfies the triangle inequality.

(b) On the one hand, $\|\mathbf{x}\|_p \le (\sum_{i=1}^n \max_i |x_i|^p)^{\frac{1}{p}} = n^{\frac{1}{p}} \max_i |x_i|$, thus $\lim_{p\to\infty} \|\mathbf{x}\|_p \le \lim_{p\to\infty} n^{\frac{1}{p}} \max_i |x_i| = \max_i |x_i|$. On the other hand, $\|\mathbf{x}\|_p \ge (\max_i |x_i|^p)^{\frac{1}{p}} = \max_i |x_i|$, thus $\lim_{p\to\infty}\|\mathbf{x}\|_p \ge \lim_{p\to\infty}\max_i|x_i| = \max_i |x_i|$. Thus $\lim_{p\to\infty}\|\mathbf{x}\|_p = \max_i |x_i|$

2.(a) Let $\mathbf{x} = \sum_{i=1}^n \lambda_i e_i$, the ith element of e_i is 1, the others are 0. Thus $\|\mathbf{A}\mathbf{x}\|_1 = \sum_{i=1}^m |a_i(\sum_{j=1}^n \lambda_j e_j)| = \sum_{i=1}^m |\sum_{j=1}^n a_{ij}\lambda_j|$, a_i is the ith row of $\mathbf{A}$. Since $|\sum_{j=1}^n a_{ij}\lambda_j| = \sum_{j=1}^n |a_{ij}\lambda_j|$ if $a_{ij}\lambda_j$ are all positive or all negative $\forall i,j$, and $\|\mathbf{x}\| = \sum_{i=1}^n|\lambda_i|$, we can simply assume that $a_{ij}\lambda_j$ are all positive. Thus $\|\mathbf{A}\mathbf{x}\|_1 = \sum_{i=1}^m\sum_{j=1}^n|a_{ij}\lambda_j| = \sum_{i=1}^n|\lambda_i|(\sum_{j=1}^m|a_{ji}|)$, $\|\mathbf{A}\|_1 = \sup_{\lambda_i}\frac{\sum_{i=1}^n(|\lambda_i|(\sum_{j=1}^m|a_{ji}|))}{\sum_{i=1}^n|\lambda_i|}$. On the one hand, $\|\mathbf{A}\|_1 \le \sup_{\lambda_i}\frac{\sum_{i=1}^n(|\lambda_i|(\max_i\sum_{j=1}^m|a_{ji}|))}{\sum_{i=1}^n|\lambda_i|} = \sup_{\lambda_i}\frac{(\sum_{i=1}^n|\lambda_i|)}{\sum_{i=1}^n|\lambda_i|}(\max_i\sum_{j=1}^m|a_{ji}|) = \max_i\sum_{j=1}^m|a_{ji}|$, on the other hand, let $\max_i\sum_{j=1}^m|a_{ji}| = \sum_{j=1}^m|a_{ji_0}|$, then $\|\mathbf{A}\|_1 \ge \sup_{\lambda_i}\frac{\lambda_{i_0}\sum_{j=1}^m|a_{ji_0}|}{\sum_{i=1}^n|\lambda_i|}$. Let $\lambda_j = 0, \forall j\ne i$, then $\sup_{\lambda_i}\frac{\lambda_{i_0}\sum_{j=1}^m|a_{ji_0}|}{\sum_{i=1}^n|\lambda_i|} \ge \frac{\lambda_{i_0}}{\lambda_{i_0}}\sum_{j=1}^m|a_{ji_0}| = \sum_{j=1}^m|a_{ji_0}|$. That is $\|\mathbf{A}\|_1 = \max_i\sum_{j=1}^m|a_{ji}|$.

(b) From (a), we have $\|\mathbf{A}\mathbf{x}\|_\infty = \max_i|\sum_{j=1}^n\lambda_j a_{ij}|$, $\|\mathbf{x}\|_\infty = \max_j|\lambda_j|$, so we can also assume that $|\sum_{j=1}^n\lambda_j a_{ij}| = \sum_{j=1}^n|\lambda_j a_{ij}|$. On the one hand, $\max_i\sum_{j=1}^n|\lambda_j a_{ij}| \le (\max_j|\lambda_j|)(\max_i\sum_{j=1}^n|a_{ij}|)$, which implies $\|\mathbf{A}\|_\infty \le \frac{(\max_j|\lambda_j|)}{\max_j|\lambda_j|}(\max_i\sum_{j=1}^n|a_{ij}|) = \max_i\sum_{j=1}^n|a_{ij}|$. On the other hand, let $\max_i\sum_{j=1}^n|a_{ij}| = \sum_{j=1}^n|a_{i_0j}|$, thus $\|\mathbf{A}\|_\infty \ge \sup_{\lambda_i}\frac{\sum_{j=1}^n|\lambda_j a_{i_0j}|}{\max_j|\lambda_j|}$. Let $\lambda_i = \lambda \ne$

$0, \forall i$, thus $\|\mathbf{A}\|_\infty \geq \frac{|\lambda| \sum_{j=1}^n |a_{i_0 j}|}{|\lambda|} = \max\limits_i \sum_{j=1}^n |a_{ij}|$. Thus, $\|\mathbf{A}\|_\infty = \max\limits_i \sum_{j=1}^n |a_{ij}|$.

Exercise 4. **Open and Closed Sets**

Solution. 1.(a)$\Rightarrow$(b) Since $\forall x \in C'$, $x \in C$, $\forall y \in C^c$, $y \notin C'$, thus $\exists \epsilon$, s.t. $B_\epsilon(y) \bigcap C = \emptyset$, which implies $B_\epsilon(y) \subset C^c$, then $\forall y \in C^c$, y is an inner point of C^c. Thus the complement of C is open.

(b)$\Rightarrow$(c) For $B_\epsilon(x) \bigcap C \neq \emptyset$, $\forall \epsilon$, if $x \in C$ then the proof is obvious, thus we only need to show that points in C^c do not have the property. Since C^c is open, $\forall x \in C^c$, $\exists \epsilon_0$, s.t. $B_{\epsilon_0}(x) \subset C^c$, that is $B_{\epsilon_0}(x) \bigcap C = \emptyset$.

(c)$\Rightarrow$(a) There are two situations. Firstly, if $\{x\} \bigcap C \neq \emptyset$, then obviously $x \in C$. Secondly, since $(B_\epsilon(x) \setminus \{x\}) \bigcap C \neq \emptyset$, $\forall \epsilon$ implies $x \in C'$, then (c) imples $C' \subset C$. That is $\mathbf{cl}\ C = C$.

2.(a) $\forall \epsilon > 0$, $1 + \epsilon \notin [0, 1]$, which implies 1 is not a inner point of $[0, 1]$, thus $[0, 1]$ is not an open set in $\mathbb{R}$. Let $\epsilon = \frac{1}{2}$, then $\forall x \in [0, 1]$, $B_\epsilon(x) \bigcap B \subset [0, 1]$, thus $[0, 1]$ is open in B. On the other hand, for $\{2\} = B \setminus [0, 1]$, $B_\epsilon(2) \bigcap B = \{2\} \subset B$, thus $[0, 1]$ is closed in B.

(b)($\Leftarrow$) Since $C = A \bigcap U$, $\forall x \in C$, $\exists \epsilon$, s.t. $B_\epsilon(x) \subset U$. Thus $B_\epsilon(x) \bigcap A = B_\epsilon(x) \bigcap (A \bigcap U) = B_\epsilon(x) \bigcap C \subset C \subset A$. Thus C is open in A.

($\Rightarrow$) Since $\forall x \in C$, $\exists \epsilon = \epsilon(x)$, s.t. $B_{\epsilon(x)}(x) \bigcap A \subset C$, consider $\bigcup\limits_x B_{\epsilon(x)}(x)$ is open in $\mathbb{R}^n$. On the one hand, $A \bigcap (\bigcup\limits_x B_{\epsilon(x)}(x)) = \bigcup\limits_x (A \bigcap B_{\epsilon(x)}(x))$. Since $B_{\epsilon(x)}(x) \bigcap A \subset C$, $\bigcup\limits_x (A \bigcap B_{\epsilon(x)}(x)) \subset C$. On the other hand, $\forall x \in C$, $x \in B_{\epsilon(x)}(x) \bigcap A$. Thus $C \subset \bigcup\limits_x (A \bigcap B_{\epsilon(x)}(x)) = A \bigcap (\bigcup\limits_x B_{\epsilon(x)}(x))$. Thus $C = A \bigcap U$, $U = (\bigcup\limits_x B_{\epsilon(x)}(x))$ is open in $\mathbb{R}^n$.

Exercise 5. **Extreme Value Theorem and Fixed Point**

Solution. 1. If $f(0)=0$ or $f(1)=1$ then it's obvious, thus we suppose $f(0)>0$ and $f(1)<1$. Let $g(x)=x-f(x)$, then $g(0)<0$, $g(1)>0$, $g\in C[0,1]$. By the zero theorem, $\exists x_0\in[0,1]$, s.t. $g(x_0)=0$, which implies $f(x_0)=x_0$.
2. Let $f(x)=x^2$, then $f:(0,1)\to(0,1)$, $f\in C(0,1)$. Since $f(x)=x \Leftrightarrow x^2-x=0 \Leftrightarrow x=0$ or $x=1$, f has no fixed point in $(0,1)$.
3. Suppose that $\exists x_1,x_2$, s.t. $x_1\neq x_2$, $f(x_1)=f(x_2)$, then $x_1=f^{(n)}(x_1)=f^{(n)}(x_2)=x_2$. Therefore f must be a bijection, thus a monotune function in $[0,1]$. Since $f(0)<f(1)$, f is increasing in $[0,1]$. Suppose that $\exists x_0\in[0,1]$, s.t. $f(x)\neq x$. If $f(x)>x$, since f is increasing we have $f^{(2)}(x)=f(f(x))>f(x)$, $f^{(3)}(x)=f(f^{(2)}(x))>f^{(2)}(x)$, $\dots$, lead to $x=f^{(n)}(x)>f^{(n-1)}(x)>\cdots>f(x)>x$. If $f(x)<x$, similarly we have $x=f^{(n)}(x)<f^{(n-1)}(x)<\cdots<f(x)<x$, both lead to contradiction. Thus $f(x)=x$, $\forall x\in[0,1]$.
4. Suppose $Imf=[a,b]$, then let $\lambda=\frac{1}{b}$, $\lambda f=g:[0,1]\to[\frac{a}{b},1]$. Suppose x is not a fixed point of g, let $a=\sup\limits_{y:g(y)=y,y\leq x}\{y\}$, $b=\sup\limits_{y:g(y)=y,y\geq x}\{y\}$, thus $\forall x\in(a,b)$, $g(x)-x$ does not change its signal. Without loss of generality, suppose that $g(x)>x$ in (a,b). $\forall x_0\in(a,b)$, denote $g(x_i)$ by x_{i+1}, that is, $g(x_0)=x_1, g(x_1)=x_2,\dots$. Thus $g(x_n)=x_{n+1}=x_n+(g(x_n)-x_n)=x_{n-1}+(g(x_{n-1})-x_{n-1})+(g(x_n)-x_n)=\cdots=x_0+\sum_{i=1}^n(g(x_i)-x_i)$. Since $g(x_i)>x_i$, $x_0>0$, along with the continuity of g, we have $|g(x_n)|=g(x_n)<\infty$, $\forall n$. Thus $\sum_{i=1}^\infty g(x_i)-x_i<\infty \Rightarrow \lim\limits_{n\to\infty} g(x_n)-x_n=0 \Rightarrow \lim\limits_{n\to\infty} g(x_n)=\lim\limits_{n\to\infty} x_n$. Since $g(x_n)=g^n(x_0)$, $x_n=g^{n-1}(x_0)$, which implies x_0 satisfies the second case, $\forall x_0\in(a,b)$. Obviously $x\in(a,b)$, Thus we have proved that if x is not a limit point, x satisfies the second case.

Exercise 6. **Linear Space**

Solution. 1. In fact we only need to check that $P_n[x]$ is closed to the addition. $\forall f,\ g \in P_n[x]$, $degf + g \leq degf \leq n$, thus $f + g \in P_n[x]$. Thus $P_n[x]$ is a linear space.

2.(a) Suppose that u and v are two zero vectors of V, then $u = u + v = v$.

(b) Suppose that a and b are two additive inverses of c, then $a + c = 0 = b+c \Rightarrow a+(c+a) = a+(c+b) \Rightarrow (a+c)+a = (a+c)+b \Rightarrow 0+a = 0+b \Rightarrow a = b$. $\forall a$, $a + (-1)a = 1a + (-1)a = (1 + (-1))a = 0$, thus $(-1)a = -a$, $\forall a \in V$

(c) $0v + 0v = (0 + 0)v = 0v \Rightarrow 0v + 0v + (-0v) = 0v + (-0)v = 0 \Rightarrow 0v = 0$, $\forall v \in V$. $\lambda\mathbf{0} + \lambda\mathbf{0} = \lambda(\mathbf{0} + \mathbf{0}) = \lambda\mathbf{0} \Rightarrow \lambda\mathbf{0} + \lambda\mathbf{0} + (-\lambda\mathbf{0}) = \lambda\mathbf{0} + (-\lambda\mathbf{0}) = 0 \Rightarrow \lambda\mathbf{0} = 0$, $\forall \lambda \in \mathbb{F}$.

(d) Suppose $\lambda\mathbf{a} = \mathbf{0}$, if $\lambda \neq 0$, then $\lambda^{-1}(\lambda\mathbf{a}) = \mathbf{0} \Rightarrow (\lambda^{-1}\lambda)\mathbf{a} = \mathbf{0} \Rightarrow 1\mathbf{a} = \mathbf{0} \Rightarrow \mathbf{a} = \mathbf{0}$.

3. A linear space can have infinite vectors and have at least one subspace $\emptyset$, since $\emptyset$ is also a linear space, and only hace one subspace.

Exercise 7. **Basis and Coordinates**

Solution. 1. Since $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ is a basis of V, $\forall v \in V$, v can be uniquely decomposed to the linear combination of $\{\mathbf{a}_i\}$, denote by $v = \sum_{i=1}^n \beta_i\mathbf{a}_i$. Since $\lambda_i \neq 0$, $v = \sum_{i=1}^n \frac{\beta_i}{\lambda_i}\lambda_i\mathbf{a}_i$, $\frac{\beta_i}{\lambda_i}$ are also unique, which implies $\{\lambda_1\mathbf{a}_1, \lambda_2\mathbf{a}_2, \ldots, \lambda_n\mathbf{a}_n\}$ is also a basis of V.

2. $\forall v \in V$, $v = \sum_{i=1}^n \beta_i\mathbf{a}_i = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)(\beta_1, \beta_2, \ldots, \beta_n)^T$. Since $(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n) = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)\mathbf{P}$ and $\mathbf{P}$ is invertible, $v = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n)\mathbf{P}^T(\beta_1, \beta_2, \ldots, \beta_n)^T = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n)(\gamma_1, \gamma_2, \ldots, \gamma_n)^T$. Since β_i are unique, γ_i are also unique. Thus $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ is also a basis of V for an invertible matrix $\mathbf{P}$.

3.(a) From 1, it is $(\frac{x_1}{\lambda_1}, \frac{x_2}{\lambda_2}, \ldots, \frac{x_n}{\lambda_n})$.

(b) Obviously, it is $(1, 1, \ldots, 1)$ under $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$, $(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_n})$ under $\{\lambda_1\mathbf{a}_1, \lambda_2\mathbf{a}_2, \ldots, \lambda_n\mathbf{a}_n\}$.

4.(a) It is $(-x, y)$. Since $\{\mathbf{c}, \mathbf{b}\}$ is a basis, it is unique.

(b) In fact, we only need $x' + z' = x$ and $y' = y$.

(c) $\|(x', y', z')\|_1 = |x'| + |y'| + |z'| = |y| + |x'| + |x - x'| \geq |x| + |y|$. The equality holds if and only if $x'(x - x') < 0$.

Exercise 8. **Derivatives with Matrices**

Solution. 1.(a) Denote $\mathbf{a}$ by $(a_1, a_2, \ldots, a_n)$. Let $f'(\mathbf{x}) = L(\mathbf{x}) = \sum_{i=1}^n a_i x_i$, then $f(\mathbf{x}) - f(\mathbf{x}_0) = L(\mathbf{x} - \mathbf{x}_0)$, thus $\lim_{\mathbf{x}\to\mathbf{x}_0;\mathbf{x}\neq\mathbf{x}_0} \frac{\|f(\mathbf{x})-f(\mathbf{x}_0)-L(\mathbf{x}-\mathbf{x}_0)\|_2}{\|\mathbf{x}-\mathbf{x}_0\|_2} = 0$.

(b) Let $f'(\mathbf{x}) = L(\mathbf{x}) = \sum_{i=1}^n 2x_i$, thus by the L'Hospital's rule, $\lim_{\mathbf{x}\to\mathbf{x}_0;\mathbf{x}\neq\mathbf{x}_0} \frac{\|f(\mathbf{x})-f(\mathbf{x}_0)-L(\mathbf{x}-\mathbf{x}_0)\|_2}{\|\mathbf{x}-\mathbf{x}_0\|_2} = 0$.

(c) $f(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^T \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$. Let $f'(\mathbf{x}) = L(\mathbf{x}) = 2(1, 1, \ldots, 1)(\mathbf{A}\mathbf{x} - \mathbf{y})^T\mathbf{A}$, thus $\lim_{\mathbf{x}\to\mathbf{x}_0;\mathbf{x}\neq\mathbf{x}_0} \frac{\|f(\mathbf{x})-f(\mathbf{x}_0)-L(\mathbf{x}-\mathbf{x}_0)\|_2}{\|\mathbf{x}-\mathbf{x}_0\|_2} = 0$.

2. Let $f : \mathbb{R}^{n\times n} \to \mathbb{R}$ be a function, $\mathbf{x}_0 \in \mathbb{R}^{n\times n}$ be a matrix, and let $L : \mathbb{R}^{n\times n} \to \mathbb{R}$ be a function. We say that f is *differentiable at* $\mathbf{x}_0$ *with derivative* L if we have

$$\lim_{\mathbf{x}\to\mathbf{x}_0;\mathbf{x}\neq\mathbf{x}_0} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)\|_2}{\|\mathbf{x} - \mathbf{x}_0\|_2} = 0.$$

3. $\mathrm{d}tr(\mathbf{A}^T\mathbf{X}) = tr(\mathrm{d}\mathbf{A}^T\mathbf{X}) = tr(\mathbf{A}^T\mathrm{d}\mathbf{X} + \mathbf{X}\mathrm{d}\mathbf{A}^T) = tr(\mathbf{A}^T\mathrm{d}\mathbf{X})) = \sum_{i=1}^m \sum_{j=1}^n a_{ij}\mathrm{d}x_{ji}$, thus $f' = \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_{ji}$.

Exercise 9. **Rank of Matrices**

Solution. 1. We will prove the theorem with the conclusion in (2), and lately we will prove (2) without using the conclusion in (1).

(a) By the conclusion in (2), $rank(\mathbf{A})$ is the dimension of the vector space spanned by its rows (row rank) or its columns (column rank). Since row space of $\mathbf{A}$ is the same as the column space of $\mathbf{A}^T$, we have $dim(Row(\mathbf{A})) = dim(Col(\mathbf{A}^T))$, which implies $rank\mathbf{A} = rank\mathbf{A}^T$. Since the columns of $\mathbf{A}^T\mathbf{A}$ are linear combinations of the columns of $\mathbf{A}$, we have $rank\mathbf{A} = rank\mathbf{A}^T\mathbf{A}$. Similarly $rank\mathbf{A} = rank\mathbf{A}\mathbf{A}^T$.

(b) By the conclusion in (2), $rank(\mathbf{AB})$ is the dimension of the image of the transformation defined by $\mathbf{AB}$. Since $Im(\mathbf{AB}) \subset Im(\mathbf{A})$, $rank(\mathbf{AB}) \leq rank(\mathbf{A})$. Example: $\mathbf{A} = \mathbf{B} = \mathbf{I}_n$.

2.(a) In fact, the dimension of $\mathcal{C}(\mathbf{A})$ is equal to the number of vectors in a basis for $\mathcal{C}(\mathbf{A})$. Denote $\mathbf{A}$ by $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)$. Without loss of generality, let $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_r\}$ be a maximal linearly independent system of $\mathbf{A}$, then $\forall j \in \{r+1, \ldots, n\}$, let $\mathbf{a}_j = \sum_{i=1}^r \lambda_{ji}\mathbf{a}_i$. Since $\mathbf{x} = \sum_{i=1}^n x_i\mathbf{e}_i$, $\mathbf{y} = \sum_{i=1}^n x_i\mathbf{a_i} = \sum_{i=1}^r (x_i + \sum_{j=r+1}^n x_j\lambda_{ji})\mathbf{a}_i$, $\forall \mathbf{y} \in \mathcal{C}(\mathbf{A})$, which implies $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_r\}$ is a basis of $\mathcal{C}(\mathbf{A})$. Thus $dim(\mathcal{C}(\mathbf{A})) = r = rank(\mathbf{A})$.

(b) Denote $n - dim(\mathcal{N}(\mathbf{A}))$ by k, Suppose that $\{\mathbf{x}_{k+1}, \mathbf{x}_{k+2}, \ldots, \mathbf{x}_n\}$ is a basis of $\mathcal{N}$, we can find a basis of $\mathbb{R}^n$, which is $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{x}_{k+2}, \ldots, \mathbf{x}_n\}$. Thus we can get $\{\mathbf{y}_i\}_{i=1}^k$, $y_i = \mathbf{A}\mathbf{x}_i$. $\forall \mathbf{y} \in \mathcal{C}(\mathbf{A})$, $\exists \mathbf{x} \in \mathbb{R}^n$, s.t. $\mathbf{y} = \mathbf{A}\mathbf{x}$. Since $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is a basis of $\mathbb{R}^n$, $\mathbf{x} = \sum_{i=1}^n \lambda_i\mathbf{x}_i$, λ_i are unique. Thus $\mathbf{y} = \mathbf{A}\sum_{i=1}^n \lambda_i\mathbf{x}_i = \sum_{i=1}^n \lambda_i(\mathbf{A}\mathbf{x}_i) = \sum_{i=1}^k \lambda_i\mathbf{y}_i + \sum_{i=k+1}^n \mathbf{0} = \sum_{i=1}^k \lambda_i\mathbf{y}_i$, $\forall \mathbf{y} \in \mathcal{C}$, λ_i are unique. Obviously $\mathbf{y}_i$ are linear independent, thus $\{\mathbf{y}_i\}_{i=1}^k$ is a basis of $\mathcal{C}$, which implies $rank(\mathbf{A}) = dim(\mathcal{C}(\mathbf{A})) = k = n - dim(\mathcal{N}(\mathbf{A}))$. That is $rank(\mathbf{A}) + dim(\mathcal{N}(\mathbf{A})) = n$.

3. Since $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times p}$, $\mathcal{C}(\mathbf{B})$ and $\mathcal{N}(\mathbf{A})$ are subspaces of $\mathbb{R}^n$. Therefore we have $n = dim(\mathbb{R}^n) \geq dim(\mathcal{C}(\mathbf{B}) + \mathcal{N}(\mathbf{A})) = dim(\mathcal{C}(\mathbf{B})) + dim(\mathcal{N}(\mathbf{A})) - dim(\mathcal{C}(\mathbf{B}) \bigcap \mathcal{N}(\mathbf{A})) = rank(\mathbf{B}) + n - rank(\mathbf{A}) - dim(\mathcal{C}(\mathbf{B}) \bigcap \mathcal{N}(\mathbf{A}))$, thus $rank(\mathbf{A}) \geq rank(\mathbf{B}) - dim(\mathcal{C}(\mathbf{B}) \bigcap \mathcal{N}(\mathbf{A})) = rank(\mathbf{AB})$.

EXERCISE 10. **Properties of Eigenvalues and Singular Values**

SOLUTION. 1. $\forall \mathbf{x} \in \mathbb{R}^n$, replace $\mathbf{x}$ with $\frac{\mathbf{x}}{\sqrt{\|\mathbf{x}\|}}$, then $\frac{\frac{\mathbf{x}^T}{\sqrt{\|\mathbf{x}\|}}\mathbf{A}\frac{\mathbf{x}}{\sqrt{\|\mathbf{x}\|}}}{\frac{\mathbf{x}^T}{\sqrt{\|\mathbf{x}\|}}\frac{\mathbf{x}}{\sqrt{\|\mathbf{x}\|}}} = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$. Thus without loss of generality, we can suppose that $\|\mathbf{x}\| = 1$. Therefore $\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \mathbf{x}^T\mathbf{A}\mathbf{x}$. Since $\mathbf{A} \in S^n$, let $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ be the eigenvalues of $\mathbf{A}$, which satisfies $\lambda_i \le \lambda_j$, $\forall i \le j$. We can decompose $\mathbf{A}$ by $\mathbf{P}^T\mathbf{S}\mathbf{P}$, $\mathbf{P}$ is an orthogonal matrix, $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n)^T, \mathbf{S} = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Thus $\mathbf{x}^T\mathbf{A}\mathbf{x} = (\mathbf{P}\mathbf{x})^T\mathbf{S}\mathbf{P}\mathbf{x}$. Since $\mathbf{P}$ is an orthogonal matrix and $\|\mathbf{x}\| = 1$, $(\mathbf{P}\mathbf{x})^T\mathbf{P}\mathbf{x} = \mathbf{x}^T\mathbf{P}^T\mathbf{P}\mathbf{x} = \mathbf{x}^T\mathbf{x} = 1$, which implies $\mathbf{P}\mathbf{x} \in \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\| = 1\}$. Since $\mathbf{P}$ is invertible, $\sup\limits_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq\mathbf{0}} \mathbf{x}^T\mathbf{A}\mathbf{x} = \sup\limits_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq\mathbf{0}} (\mathbf{P}\mathbf{x})^T\mathbf{S}\mathbf{P}\mathbf{x} = \sup\limits_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq\mathbf{0}} \mathbf{x}^T\mathbf{S}\mathbf{x}$, Let $\mathbf{x} = \{y_1, y_2, \ldots, y_n\}$, $\sum_{i=1}^n y_i^2 = 1$, then $\mathbf{x}^T\mathbf{S}\mathbf{x} = \sum_{i=1}^n \lambda_i y_i^2$. Therefore $\sum_{i=1}^n \lambda_i y_i^2 \le \sum_{i=1}^n \lambda_n y_i^2 = \lambda_n \sum_{i=1}^n y_i^2 = \lambda_n = \lambda_{\max}$, equality holds when $\mathbf{x} = \{0, 0, \ldots, 0, 1\}$ and $\sum_{i=1}^n \lambda_i y_i^2 \ge \sum_{i=1}^n \lambda_1 y_i^2 = \lambda_1 \sum_{i=1}^n y_i^2 = \lambda_1 = \lambda_{\min}$, equality holds when $\mathbf{x} = \{1, 0, 0, \ldots, 0\}$. Therefore, $\lambda_{\max}(\mathbf{A}) = \sup\limits_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$, $\lambda_{\min}(\mathbf{A}) = \inf\limits_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$.