Chapter 7

Nonlinear problems

7.1 Recalling few results of functions of several variables

7.1.1 Continuity

[7.1.1] Fact. Let $f : \mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$. TFAE.

i) The function f is continuous at a.

ii) For each unit vector $u \in \mathbb{R}$, we have $\lim\limits_{t\to 0+} f(a+tu) = f(a)$.

[7.1.2] Fact. Let $f : \mathbb{R}^n \to \mathbb{R}$ and $a \in \mathbb{R}^n$. The following are not equivalent.

i) The function f is continuous at a.

ii) For each unit vector $u \in \mathbb{R}^n$, we have $\lim\limits_{t\to 0+} f(a+tu) = f(a)$.

To see this define

$$f(x,y) = \begin{cases} \frac{y^2}{x}, & \text{if } x \neq 0 \\ 0 & \text{else.} \end{cases}$$

Notice that for each unit vector $u \in \mathbb{R}^2$, we have $\lim\limits_{t\to 0+} f(0+tu) = f(0)$. However, f is not continuous at 0 because if we approach 0 along the curve (imagine a sequence of points on this curve converging $(0,0)$) $y = \sqrt{x}$, the limit is 1, that is, $\lim\limits_{\substack{x\to 0+\\ y=\sqrt{x}}} f(x,y) = 1$.

[7.1.3] Fact. Let $P(x)$ and $Q(x)$ be polynomials in $x = (x_1, \ldots, x_n)$. Then the function $\frac{P(x)}{Q(x)}$ is continuous wherever it is defined on $\mathbb{R}^n$.

[7.1.4] Fact. Let $f : \mathbb{R}^2 \to \mathbb{R}$ such that both $g(y) = f(a,y)$, $h(x) = f(x,b)$ are continuous for each $a, b \in \mathbb{R}$. Even then, $f(x,y)$ my not be continuous.

To see this, consider the function $f(x,y) = \frac{xy}{x^2+y^2}$, $f(0,0) = 0$. If $a \neq 0$ then $f(a,y) = \frac{ay}{a^2+y^2}$ is continuous on $\mathbb{R}$. If $a = 0$ then $f(a,y) = 0$ is continuous on $\mathbb{R}$. Similarly $f(x,b)$ is continuous for any fixed b. The function f is not continuous at $(0,0)$: approaching $(0,0)$ along the line $y = mx$, the limit is $\frac{m}{1+m^2}$, which changes with m.

7.1.2 Differentiability

[7.1.5] Discussion : coefficient matrix of linear terms in $f(a+h) - f(a)$.

a) Consider a polynomial $f(x) = x_1^2 + x_1x_2$. Then

$$f(a+h) - f(a) = 2a_1h_1 + a_2h_1 + a_1h_2 + h_1^2 + h_1h_2.$$

Notice that, the linear terms can also be written as $\begin{bmatrix} 2a_1 + a_2 & a_1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$.

b) So, 'the coefficient matrix of linear terms in $f(a+h) - f(a)$' is $A = \begin{bmatrix} 2a_1 + a_2 & a_1 \end{bmatrix}$.

c) Thus, for $f(x) = x_1^2 - x_1x_2 + x_3^2$ defined on $\mathbb{R}^3$ and $a = (1,2,3)$, the coefficient matrix of the linear terms in h in the expression $f(a+h) - f(a)$ is $\begin{bmatrix} 2a_1 - a_2 & -a_1 & 2a_3 \end{bmatrix}$.

To get this quickly, try to imagine the terms that could give you $(\cdots)h_1$.

[7.1.6] Fact. Let $f(x)$ be a real polynomial in $x = (x_1, \ldots, x_n)$ and $a \in \mathbb{R}^n$. Let A be the coefficient of linear terms in h of $f(a+h) - f(a)$. Then $\lim_{\|h\| \to 0} \frac{|f(a+h) - f(a) - Ah|}{\|h\|} = 0$.

Proof. Note that $f(a+h) - f(a) - Ah$ is a polynomial with terms of degree two or more in h. So $\lim_{\|h\| \to 0} \frac{|f(a+h) - f(a) - Ah|}{\|h\|} = 0$. Note that $\lim_{\|h\| \to 0} \frac{|h_1h_2|}{\|h\|} = \lim_{\|h\| \to 0} \frac{|h_1|}{\|h\|}|h_2| = 0$. ■

[7.1.7] Discussion : differentiability.

a) Let $E \subseteq \mathbb{R}^n$ be open, $f : E \to \mathbb{R}$, and $a \in E$. Suppose that there is a matrix $A_{1\times n}$ such that

$$\lim_{\|h\| \to 0} \frac{|f(a+h) - f(a) - Ah|}{\|h\|} = 0.$$

Then we say f is DIFFERENTIABLE at a and write $f'(a) = A$. We sometimes call A, the TOTAL DERIVATIVE of f at a. The GRADIENT $\nabla f(a)$ of f at a is the vector $f'(a)^t$.

b) Thus for $f(x,y,z) = x^2 + y^2 + xyz$, we have $f'(x,y,z) = \begin{bmatrix} 2x + yz & 2y + xz & xy \end{bmatrix}$.

[7.1.8] Discussion : directional derivatives.

a) On your way to school, the height of the road increases and the road goes as if it is the line $f(x) = x$. What is the slope of the road we are facing?

b) What is the slope we will face on our way back?

c) So, the slope changes according to which direction we are facing.

d) Let $E \subseteq \mathbb{R}^n$ be open, $f : E \to \mathbb{R}$, $a \in E$ and $0 \neq u \in \mathbb{R}^n$. The DIRECTIONAL DERIVATIVE $\boldsymbol{D_u f(a)}$ of f at a in the direction of u is defined as

$$D_u f(a) := \lim_{t\to 0} \frac{f(a+tu) - f(a)}{t},$$

provided the limit exists. It means the 'instantaneous rate of change of f where unit step means u'.

e) Thus, for $f(x,y,z) = x^2 + y^2 + xyz$, $a = (1,2,3)$, $u = (1,0,1)$ and $v = (0,1,1)$, we have

$$D_u f(a) = \lim_{t\to 0} \frac{(1+t)^2 + 2^2 + (1+t)2(3+t) - 1^2 - 2^2 - 1.2.3}{t} = 10$$

and $D_v f(a) = 9$.

f) When $u = e_i$ the directional derivative $D_u f(a)$is called the PARTIAL DERIVATIVE $\boldsymbol{D_i f(a)}$ of f with respect to the ith coordinate.

[7.1.9] **Fact.** Suppose that $D_u f(a) = \beta$ and $\alpha \neq 0$. Then $D_{\alpha u} f(a)$ exists and $D_{\alpha u} f(a) = \alpha D_u f(a)$.

Proof. We have $D_{\alpha u} f(a) = \lim\limits_{t\to 0} \frac{f(a+t\alpha u)-f(a)}{t} = \alpha \lim\limits_{t\to 0} \frac{f(a+t\alpha u)-f(a)}{\alpha t} = \alpha\beta = \alpha D_u f(a)$. ■

[7.1.10] **Fact.** Let $f : \mathbb{R}^n \to \mathbb{R}$ such that $D_u f(a)$ exists for each $u \neq 0$. Even then f need not be continuous at a.

For example, take $f(x,y) = \frac{y^2}{x}$ if $x \neq 0$ and 0, otherwise. Notice that $D_1 f(0) = \lim\limits_{t\to 0} \frac{f(t,0)-f(0,0)}{t} = 0$ and $D_2 f(0) = \lim\limits_{t\to 0} \frac{f(0,t)-f(0,0)}{t} = 0$. Take a unit vector $u = \begin{bmatrix} \frac{1}{\sqrt{m^2+1}} & \frac{m}{\sqrt{m^2+1}} \end{bmatrix}^t$ on the line $y = mx$, we have

$$D_u f(0) = \lim_{t\to 0} \frac{f(tu_1, tu_2) - f(0)}{t} = \frac{m^2}{\sqrt{m^2+1}}.$$

Thus the directional derivatives exist in all directions. We already know that f is not continuous at 0. ┘

[7.1.11] **Fact.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous at a. Then $D_u f(a)$ may not exist even for a single $u \neq 0$. For example, take $f(x) = \|x\|$ on $\mathbb{R}^n$ and check at 0.

[7.1.12] **Fact.** Let $E \subseteq \mathbb{R}^n$ be open, $f : E \to \mathbb{R}$, $a \in E$, and $0 \neq u \in \mathbb{R}^n$. Suppose that f is differentiable at a with $f'(a) = A = \begin{bmatrix} d_1 & \cdots & d_n \end{bmatrix}$. Then f is continuous at a and $D_u f(a)$ exists with $D_u f(a) = \langle \nabla f(a), u \rangle$. In particular, $D_i f(a) = d_i$.

Proof. To prove continuity note that by definition $\lim\limits_{\|h\|\to 0} \frac{|f(a+h)-f(a)-Ah|}{\|h\|} = 0$. So $\lim\limits_{\|h\|\to 0} |f(a + h) - f(a) - Ah| = 0$ and $\lim\limits_{\|h\|\to 0} f(a+h) = f(a)$. So f is continuous.

To prove the next assertion note that as $\lim\limits_{\|h\|\to 0} \frac{|f(a+h)-f(a)-Ah|}{\|h\|} = 0$, in particular, taking $h = tu$, we have $\lim\limits_{t\to 0} \left| \frac{f(a+tu)-f(a)-tAu}{t} \right| = 0$. That is, $\lim\limits_{t\to 0} \left| \frac{f(a+tu)-f(a)}{t} - Au \right| = 0$. That is, $\lim\limits_{t\to 0} \frac{f(a+tu)-f(a)}{t} = Au$. That is, $D_u f(a) = Au = \langle \nabla f(a), u \rangle$. The proof is complete. ■

To verify whether a function is differentiable at a point

1. Find all $D_i f(a)$. If some $D_i f(a)$ does not exist, then conclude that f is not differentiable at a.
2. Form the matrix $A = \begin{bmatrix} D_1 f(a) & \cdots & D_n f(a) \end{bmatrix}$. Substitute A in the limit definition of the derivative and check if the limit is 0.
3. Conclude 'yes', if the limit is 0, otherwise conclude 'no'.

[7.1.13] Example. Take $f(x,y) = \begin{cases} \frac{x^2 y}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$ Check whether it is differentiable at 0.

Answer. We have $D_1 f(0) = 0$ and $D_2 f(0) = 0$. But $\lim\limits_{h \to 0} |\frac{f(h) - \begin{bmatrix} 0 & 0 \end{bmatrix} h}{\|h\|}|$ does not exist. So it is not differentiable at 0. ┘

[7.1.14] Fact : a sufficient condition for differentiability. Let $E \subseteq \mathbb{R}^n$ be open and $f : E \to \mathbb{R}$. If $D_i f$ are continuous <u>on E</u>, then f is differentiable on E.

This condition is not necessary for differentiability. For example, take $f : (-1,1) \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Then $f'(0) = 0$ and $f'(x) = 2x \sin(1/x) - \cos(1/x)$ for $x \neq 0$. But f' is not continuous at 0.

[7.1.15] Exercise(E).

a) Let $f(x,y) = \min\{x,y\}$. Check the directional derivatives and differentiability at 0.

b) Let $f(x,y) = \min\{x^2, y^2\}$. Check the directional derivatives and differentiability at 0, $(1,1)$ and $(1,2)$.

[7.1.16] Discussion : higher order derivatives.

a) Let $E \subseteq \mathbb{R}^n$ be open and $f : E \to \mathbb{R}$. Then f is said to be CONTINUOUSLY DIFFERENTIABLE on E, denoted $f \in \boldsymbol{C^1(E)}$, if $D_i f$ are continuous on E.

b) Let $E \subseteq \mathbb{R}^n$ be open and $f : E \to \mathbb{R}$. Then the SECOND ORDER PARTIAL DERIVATIVE $\boldsymbol{D_{ij} f}$ is defined as $D_{ij} f = D_i(D_j f)$, if it exists.

c) It can happen that $D_{ij} f$ and $D_{ji} f$ both exist and unequal at a point.

For example, consider

$$f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

d) But, it is known from calculus that, if $D_{ij} f, D_{ji} f$ are <u>continuous</u> on E (open), then $D_{ij} f = D_{ji} f$ on E.

e) Let $E \subseteq \mathbb{R}^n$ be open and $f : E \to \mathbb{R}^m$. Then f is said to be TWICE CONTINUOUSLY DIFFERENTIABLE on E, denoted $f \in \mathbf{C^2(E)}$, if all $D_{ij}f$ are continuous on E.

f) All polynomials in $x_1, \ldots, x_n$ infinitely differentiable functions.

g) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2 function. The HESSIAN $\boldsymbol{H_f(a)}$ OR $\boldsymbol{H(a)}$ of f at a is the matrix (it is a real symmetric matrix as the function is C^2)

$$H(a) = \begin{bmatrix} D_{11}f & D_{12}f & \cdots & D_{1n}f \\ D_{21}f & D_{22}f & \cdots & D_{2n}f \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1}f & D_{n2}f & \cdots & D_{nn}f \end{bmatrix}(a).$$

h) Thus the Hessian of $e^{x+y} - xyz$ at $(1,1,0)$ is $\begin{bmatrix} e^2 & e^2 & -1 \\ e^2 & e^2 & -1 \\ -1 & -1 & 0 \end{bmatrix}$.

[7.1.17] **Class workout.** Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a C^2 function and $a \in \mathbb{R}^3$.

a) Is it true that $\sum\limits_{i=1}^{3} D_i f(a) x_i = \langle \nabla f(a), x \rangle$?

b) Is it true that $\sum\limits_{i,j=1}^{3} D_{ij} f(a) x_i x_j = x^t H(a) x$?

7.1.3 Taylor's theorems

[7.1.18] **Fact.** `(Rolle's theorem : single variable.)` Let f be continuous on $[a, b]$ and differentiable on (a, b) with $f(a) = f(b)$. Then $\exists c \in (a, b)$ where $f'(c) = 0$.

[7.1.19] **Fact.** `(Generalized mean value theorem.)` Let f, g be continuous on $[a, b]$ and differentiable on (a, b). Then $\exists c \in (a, b)$ where $f'(c)[g(b) - g(a)] = g'(c)[f(b) - f(a)]$.

Proof. (Self) Consider $H(x) = f(x)[g(b) - g(a)] - g(x)[f(b) - f(a)]$. Apply Rolle's theorem. ∎

[7.1.20] **Fact.** `(Taylor's theorem in one variable.)` Let $E \subseteq \mathbb{R}$ be open, $f : E \to \mathbb{R}$ be in $C^n(E)$ and $[a, a+x] \subseteq E$. Then $\exists t \in (a, a+x)$ such that

$$f(a+x) = f(a) + f'(a)x + \frac{1}{2!}f''(a)x^2 + \cdots + \frac{1}{(n-1)!}f^{(n-1)}(a)x^{n-1} + \frac{1}{n!}f^{(n)}(t)x^n.$$

Proof. (Self) For $r \in [0, x]$, consider

$$F(r) = f(a+r) + f'(a+r)(x-r) + \cdots + \frac{1}{(n-1)!}f^{(n-1)}(a+r)(x-r)^{n-1}, \quad G(r) = (x-r)^n.$$

Then F, G are continuous on $[0, x]$ and differentiable on $(0, x)$. Apply generalized mean value theorem: $\exists c \in (0, x)$ such that $F'(c)[G(x) - G(0)] = G'(c)[F(x) - F(0)]$. So

$$\Big(f'(a+c)+f''(a+c)(x-c)-f'(a+c)+f^{(3)}(a+c)\frac{(x-c)^2}{2!}-f''(a+c)(x-c)+\cdots+f^{(n)}(a+c)\frac{(x-c)^{n-1}}{(n-1)!}$$

$$-f^{(n-1)}(a+c)\frac{(x-c)^{n-2}}{(n-2)!}\Big)[-x^n] = -n(x-c)^{n-1}\Big[f(a+x)-f(a)-f'(a)x-\cdots-f^{(n-1)}(a)\frac{x^{n-1}}{(n-1)!}\Big].$$

So

$$f^{(n)}(a+c)\frac{(x-c)^{n-1}}{(n-1)!}x^n = n(x-c)^{n-1}\Big[f(a+x)-f(a)-f'(a)x-\cdots-f^{(n-1)}(a)\frac{x^{n-1}}{(n-1)!}\Big].$$

That is, $f^{(n)}(a+c)\frac{x^n}{n!} = [f(a+x)-f(a)-f'(a)x-\cdots-f^{(n-1)}(a)\frac{x^{n-1}}{(n-1)!}]$. So, $\exists t := a+c \in (a, a+x)$ such that

$$f(a+x) = f(a) + f'(a)x + \frac{1}{2!}f''(a)x^2 + \cdots + \frac{1}{(n-1)!}f^{(n-1)}(a)x^{n-1} + \frac{1}{n!}f^{(n)}(t)x^n.$$

■

[7.1.21] **Fact** (Taylor's theorem in many variables) Let $E \subseteq \mathbb{R}^n$ be open, $f : E \to \mathbb{R}$ be in $\mathcal{C}^m(E)$ and $[a, a+x] \subseteq E$. Then $\exists t \in (0,1)$ such that

$$\begin{aligned} f(a+x) &= f(a) + \sum_i D_i f(a)\, x_i + \tfrac{1}{2!}\sum_{i,j} D_{ij} f(a)\, x_i x_j + \cdots + \tfrac{1}{(m-1)!}\sum_{i_1,\ldots,i_{m-1}} D_{i_1\cdots i_{m-1}} f(a)\, x_{i_1}\cdots x_{i_{m-1}} \\ &\quad + \tfrac{1}{m!}\sum_{i_1,\ldots,i_m} D_{i_1\cdots i_m} f(a+tx)\, x_{i_1}\cdots x_{i_m}. \end{aligned}$$

Proof. (Self [Rudin-p243]) Define $p(t) = a + tx$, for $0 \le t \le 1$. Define $h(t) = f(p(t))$. For any $t \in (0,1)$, by chain rule,

$$h'(t) = f'(p(t))p'(t) = \sum_{i=1}^n D_i f(p(t)) x_i, \qquad h''(t) = \sum_{i,j} x_i x_j D_{ij} f(p(t)),$$

and so on. By Taylor's theorem in one variable, $\exists t \in (0,1)$ such that $h(1) = \sum_{k=0}^{m-1}\frac{h^{(k)}(0)}{k!} + \frac{h^{(m)}(t)}{(m)!}$. ■

[7.1.22] **Discussion.** (Another look of Taylor's theorem.)

a) Assume that $a = (a_1, \cdots, a_n)^t$. Take $\alpha = (\alpha_1, \cdots, \alpha_n)$ where $\alpha_i \ge 0$ are integers with $|\alpha| := \alpha_1 + \cdots + \alpha_n = k$. That is, α is a solution to the equation $z_1 + \cdots + z_n = k$ in nonnegative integers. There are $\binom{n+k-1}{k}$ many such solutions.

b) Use the notations $\quad \boldsymbol{D}^{\boldsymbol{\alpha}} := D_1^{\alpha_1} D_2^{\alpha_2}\cdots D_n^{\alpha_n}, \quad \boldsymbol{x}^{\boldsymbol{\alpha}} := x_1^{\alpha_1}x_2^{\alpha_2}\cdots x_n^{\alpha_n} \quad \boldsymbol{\alpha}! := \alpha_1!\cdots\alpha_n!.$

d) Then the expression in Taylor's theorem will look like

$$f(a+x) = f(a) + \sum_{|\alpha|=1} D^\alpha f(a)\frac{x^\alpha}{\alpha!} + \cdots + \sum_{|\alpha|=m-1} D^\alpha f(a)\frac{x^\alpha}{\alpha!} + \sum_{|\alpha|=m} D^\alpha f(a+tx)\frac{x^\alpha}{\alpha!}.$$

a) Let $f : \mathbb{R}^n \to \mathbb{R}$ be infinitely differentiable. Then the TAYLOR SERIES $\boldsymbol{T_f(x;a)}$ of f about the point a is defined as

$$f(a) + \sum_{|\alpha|=1} D^\alpha f(a)\frac{(x-a)^\alpha}{\alpha!} + \sum_{|\alpha|=2} D^\alpha f(a)\frac{(x-a)^\alpha}{\alpha!} + \sum_{|\alpha|=3} D^\alpha f(a)\frac{(x-a)^\alpha}{\alpha!} + \cdots$$

b) `(Example.)` Take $f(y,z) = y^2z^4 + yz^3 - 5yz + 6$ and $a = (1,2)$. Find the coefficient of $(y-1)^2(z-2)^2$ in $T_f((y,z);a)$ in two different ways.

Answer. Use $x = (y,z)$. We want $(x-a)^\alpha = (y-1)^2(z-2)^2$, that is, $\alpha = (2,2)$. Hence, the coefficient is

$$\frac{1}{\alpha!}D_{1,1,2,2}f(a) = \frac{1}{2!2!}(2.4.3.2^2) = 24.$$

o Alternately, note that

$$f(y,z) = (y-1+1)^2(z-2+2)^4 - (y-1+1)(z-2+2)^3 - 5(y-1+1)(z-2+2) + 6.$$

The coefficient for $(y-1)^2(z-2)^2$ can only come from the first term. When expanded using binomial expansion, it will look like $(y-1)^2\binom{4}{2}(z-2)^2 2^2$. So the required coefficient is 24.

7.2 Positive definite matrices

[7.2.1] Definition. Let $A \in M_n(\mathbb{C})$.

a) It is called POSITIVE DEFINITE(pd) if $x^*Ax > 0$ holds $\forall x \in \mathbb{C}^n \setminus \{0\}$.

b) It is called POSITIVE SEMIDEFINITE(psd) if $x^*Ax \geq 0$ holds $\forall x \in \mathbb{C}^n$.

[7.2.2] Facts and examples.

a) Let $A \in M_n(\mathbb{C})$ be pd. Then $a_{ii} = e_i^t Ae_i > 0$.

a') Let $A \in M_n(\mathbb{C})$ be psd. Then $a_{ii} = e_i^t Ae_i \geq 0$.

b) Take any matrix $A \in M_n(\mathbb{C})$. Then A^*A and AA^* are psd. If A is nonsingular, then they are pd.

Proof. (Self) For each x, we have $x^*A^*Ax = (Ax)^*Ax = \|Ax\|^2 \geq 0$. If $x \neq 0$ and A is nonsingular, then $Ax \neq 0$ and so $\|Ax\|^2 > 0$. ∎

c) The matrix I is pd. The zero matrix is a psd matrix.

d) A pd matrix is by definition a psd matrix.

e) A singular matrix cannot be pd.

[7.2.3] Fact. `(Psd implies Hermitian)` Let $A \in M_n(\mathbb{C})$ be a psd matrix. Then $A^* = A$.

Proof. (Self) Fix i, j, $i \neq j$. Let $v = e_i + e_j$. As $v^*Av \geq 0$, we have $a_{ii} + a_{ij} + a_{ji} + a_{jj} \geq 0$. But we already know that $a_{ii} + a_{jj} \geq 0$. So, we get $a_{ij} + a_{ji} \in \mathbb{R}$. That is, $Im(a_{ij}) = -Im(a_{ji})$.

Now take $v = e_i + ie_j$. As $v^*Av \geq 0$, we have $a_{ii} + ia_{ij} - ia_{ji} + a_{jj} \geq 0$. But we already know that $a_{ii} + a_{jj} \geq 0$. So, we get $a_{ij} - a_{ji} \in i\mathbb{R}$. That is, $Re(a_{ij}) = Re(a_{ji})$.

So $\overline{a_{ij}} = a_{ji}$. ■

[7.2.4] Lemma. Let $A \in M_n(\mathbb{R})$.

a) Then A is pd iff $A^t = A$ and $x^tAx > 0$ holds $\forall x \in \mathbb{R}^n$, $x \neq 0$.

b) Then A is psd iff $A^t = A$ and $x^tAx \geq 0$ holds $\forall x \in \mathbb{R}^n$.

Proof. (Self) a) Let A be a real pd matrix. As A is pd, it is Hermitian, and in our case it is symmetric. and $x^tAx > 0$ for all real $x \neq 0$.

Conversely, suppose that A is real symmetric and $x^tAx > 0$ for all real $x \neq 0$. Note that, for a real symmetric matrix A, we have $x^tAy = y^tAx$. Hence

$$(x + \mathrm{iy})^*\mathrm{A}(\mathrm{x} + \mathrm{iy}) = \mathrm{x^tAx} + \mathrm{y^tAy} + \mathrm{i}(\mathrm{x^tAy} - \mathrm{y^tAx}) = \mathrm{x^tAx} + \mathrm{y^tAy} > 0,$$

if $x + \mathrm{iy} \neq 0$. The proof for b) is similar. ■

[7.2.5] Theorem. `(Equivalent conditions for pd/psd.)` Let $A \in M_n(\mathbb{C})$. TFAE.

a) A is pd (resp. psd).

b) $A^* = A$ and the eigenvalues of A are positive (resp. nonnegative).

c) $A = B^*B$ for some nonsingular (resp. not necessarily nonsingular) matrix B.

Proof. (Self) a)$\Rightarrow$b). Let A be pd. We already know that $A^* = A$. Let (λ, x) be an eigenpair. Then $x^*Ax = x^*x\lambda \Rightarrow \lambda = \frac{x^*Ax}{x^*x} > 0$, as A is pd.

b)$\Rightarrow$c). Suppose that $A^* = A$ and all eigenvalues of A are positive. Then, by spectral theorem (for Hermitian matrices), there exists a unitary matrix U and a diagonal matrix $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$, $\lambda_i > 0$ such that $A = U^*\Lambda U$. So

$$A = U^*\Lambda^{1/2}\Lambda^{1/2}U = U^*\Lambda^{1/2}UU^*\Lambda^{1/2}U = B^*B.$$

Notice that B is nonsingular as Λ has positive diagonal entries.

c)$\Rightarrow$a). Already done.

The proof for the psd part is similar. ■

We will need the following notations.

[7.2.6] Definition.

a) Let $A \in M_n(\mathbb{C})$ and $S \subseteq [n]$. By $\boldsymbol{A(S,S)}$, we denote the submatrix of A whose entries are indexed by S. This is called the PRINCIPAL SUBMATRIX of A indexed by S.

b) For $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \\ 0 & 2 & 3 & 1 \\ 9 & 8 & 7 & 6 \end{bmatrix}$ and $S = \{1, 4\}$, we have $A(S,S) = \begin{bmatrix} 1 & 4 \\ 9 & 6 \end{bmatrix}$.

c) A LEADING PRINCIPAL SUBMATRIX is one for which the subset S is $\{1, 2, \ldots, k\}$ for some k.

d) For the matrix A in b), and $S = \{1, 2\}$, we have $A(S, S) = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$. It is a leading principal submatrix.

e) The matrix A in b), has only four leading principal submatrices.

f) The determinant of a square submatrix is called a MINOR. The determinant of a leading principal submatrix is called a 'leading principal minor'.

g) Sometimes we use $\boldsymbol{A(S|S)}$ to mean the matrix $A(S^c, S^c)$.

We will need the following two results from advanced linear algebra.

[7.2.7] Theorem. (`Cauchy interlacing theorem.`) Let A and $B = \left[\begin{array}{c|c} A & y \\ \hline y^* & b \end{array}\right]$ be Hermitian. Let $\lambda_1 \leq \cdots \lambda_n$ be the eigenvalues of A and $\mu_1 \leq \cdots \leq \mu_{n+1}$ be the eigenvalues of B. Then

$$\mu_1 \leq \lambda_1 \leq \mu_2 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq \mu_{n+1}.$$

That is, the eigenvalues of A and B interlace.

[7.2.8] Theorem. The eigenvalues of a matrix A are continuous functions of its entries.

Now we can give two useful conditions to test whether a matrix is pd/psd.

[7.2.9] Theorem. (`Test for pd/psd.`) Let $A \in M_n(\mathbb{C})$.

a) Then A is pd iff $A^* = A$ and the leading principal minors of A are positive.

b) Then A is psd iff $A^* = A$ and all principal minors are nonnegative.

Proof. (Self) a) Suppose that A is pd. Then $A^* = A$ and the eigenvalues $\lambda_i > 0$. By Cauchy interlacing theorem, the eigenvalues of the leading principal matrices are > 0. Hence their determinant is positive.

Conversely, suppose that $A^* = A$ and the leading principal minors are positive. By use of interlacing, we see that eigenvalues of A are positive. So A is pd.

b) If A is psd, then showing all principal minors ≥ 0 is similar to that of a). To prove the converse, let $A^* = A$ and assume that all principal minors are ≥ 0. Note that

$$|A + \epsilon e_1 e_1^t| = (a_{11} + \epsilon)|A(1|1)| - a_{12}|A(1|2)| + a_{13}|A(1|3)| - \cdots = |A| + \epsilon|A(1|1)| \geq |A|,$$

so that adding $\epsilon > 0$ to a diagonal entry does not decrease the determinant. Notice that we get $|A + \epsilon e_1 e_1^t| > |A|$, if $|A(1|1)| > 0$ and $\epsilon > 0$.

Next we show that $|A + \epsilon I| > 0$ whenever A has all minors nonnegative. We proceed by induction. Assume that it is true for $n - 1$. Consider the case for n and let B be obtained by adding ϵ to all diagonal entries of A except the first. Notice that $A(1|1)$ is Hermitian and it has all principal minor ≥ 0. By induction hypothesis, we have that $|B(1|1)| > 0$. Hence by earlier argument $|B| > 0$.

It now follows that $A + \epsilon I$ is pd. This is true for each $\epsilon > 0$. As eigenvalues are continuous functions of the entries, it follows that eigenvalues of A are nonnegative. Hence A is psd.

[7.2.10] NoPen.

a) If we say 'a matrix $A \in M_n(\mathbb{C})$ is psd iff $A^* = A$ and all leading principal minors are nonnegative', what would go wrong?

b) Let $A \in M_n(\mathbb{C})$ be pd. Should every (nonempty) principal submatrix of A be pd?

c) Let $A \in M_n(\mathbb{C})$ have positive eigenvalues. Must it be similar to a positive definite matrix?

d) If A is pd and S is a nonsingular matrix, is S^*AS necessarily pd?

e) How do I create a pd matrix $A \in M_{20}(\mathbb{R})$ which does not have a zero entry?

f) Let $f \in C^2(E)$, $E \subseteq \mathbb{R}^n$ be open and $a \in E$. Suppose that $B_\delta(a) \subseteq E$ and for each $x \in B_\delta(a)$, $x \neq a$, we see that $H(x)$ is a pd matrix. Can we conclude that $H(a)$ is a psd matrix?

[7.2.11] Exercise(E+) Let $A \in M_n(\mathbb{C})$ be a positive definite matrix and $\lambda > 0$ be the smallest eigenvalue of A. Show that, for each $x \in \mathbb{C}^n$, we have $x^*Ax \geq \lambda x^*x$.

7.3 Unconstrained optimization

For now, let us consider the general problem of the form

$$(P1) \qquad \begin{array}{ll} \min & f(x) \\ \hline \text{s.t.} & x \in T \subseteq \mathbb{R}^n. \end{array}$$

[7.3.1] Definition. Consider (P1). Let $T \subseteq \mathbb{R}^n$ be nonempty and $a \in T$. A FEASIBLE DIRECTION at a is a vector d (allow $d = 0$) such that we can travel from a in that direction some positive multiple amount and while staying inside T. That is, $\exists \delta > 0$ such that $a + \theta d \in T$, $\forall \theta \in (0, \delta)$.

- By $D(a)$, we denote the set of all feasible directions available at a. (Remember that, 0 is allowed as a feasible direction.) We shall use $\overline{D}(a)$ to denote the closure of $D(a)$.

- Feasible directions at a point a only depend the structure of the feasible region near that point.

[7.3.2] Example. Consider the shaded region in the following figure.