Solutions to Homework 2 Problems

Nguyen Trung Hai

September 07, 2019

Problem 1

(a) Show that M is a symmetric matrix

Proof. To show that M is symmetric, we need to show that $B = C$, or equivalently,

$$\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X] \tag{1}$$

From the definition of covariance:

$$\begin{aligned} \mathrm{Cov}[X, Y] &= \mathrm{E}\left[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])\right] \\ &= \mathrm{E}\left[(Y - \mathrm{E}[Y])(X - \mathrm{E}[X])\right] \\ &= \mathrm{Cov}[Y, X] \end{aligned} \tag{2}$$

So M is a symmetric matrix. □

(b) Show that M is a positive semidefinite matrix

Proof. For all vectors $\gamma = (\alpha, \beta)^T \neq (0, 0)^T$, we have

$$\begin{aligned} \gamma^T M \gamma &= (\alpha, \beta) \begin{pmatrix} \mathrm{Var}[X] & \mathrm{Cov}[X, Y] \\ \mathrm{Cov}[Y, X] & \mathrm{Var}[Y] \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\ &= (\alpha, \beta) \begin{pmatrix} \alpha \mathrm{Var}[X] + \beta \mathrm{Cov}[X, Y] \\ \alpha \mathrm{Cov}[Y, X] + \beta \mathrm{Var}[Y] \end{pmatrix} \\ &= \alpha^2 \mathrm{Var}[X] + \alpha\beta \mathrm{Cov}[X, Y] + \alpha\beta \mathrm{Cov}[Y, X] + \beta^2 \mathrm{Var}[Y] \\ &= \alpha^2 \mathrm{Var}[X] + 2\alpha\beta \mathrm{Cov}[X, Y] + \beta^2 \mathrm{Var}[Y] \\ &= \mathrm{Var}[\alpha X + \beta Y] \geq 0 \end{aligned} \tag{3}$$

So M is a positive semidefinite matrix. □

(c) Show that the determinant of M, $\det(M)$ is nonnegative

Proof.

$$\det(M) = \mathrm{Var}[X]\mathrm{Var}[Y] - \mathrm{Cov}[X, Y]^2 \tag{4}$$

According to **Cauchy-Schwarz inequality**, we have

$$\mathrm{Cov}[X,Y]^2 \leq \mathrm{Var}[X]\mathrm{Var}[Y]. \tag{5}$$

The equal sign holds if X and Y are independent random variables. So

$$\det(M) \geq 0 \tag{6}$$

□

(d) Suppose $\det(M) > 0$, derive the formula for the inverse of M, M^{-1}

$$M^{-1} = \frac{1}{\det(M)}\mathrm{cof}(M)^T, \tag{7}$$

where $\mathrm{cof}(M)$ is the matrix of cofactors of M,

$$\mathrm{cof}(M) = \begin{pmatrix} \mathrm{Var}[Y] & -\mathrm{Cov}[Y,X] \\ -\mathrm{Cov}[X,Y] & \mathrm{Var}[X] \end{pmatrix}. \tag{8}$$

So

$$\begin{aligned} M^{-1} &= \frac{1}{\mathrm{Var}[X]\mathrm{Var}[Y] - \mathrm{Cov}[X,Y]^2} \begin{pmatrix} \mathrm{Var}[Y] & -\mathrm{Cov}[X,Y] \\ -\mathrm{Cov}[Y,X] & \mathrm{Var}[X] \end{pmatrix} \\ &= \frac{1}{\mathrm{Cov}[X,Y]^2\left(\frac{1}{\rho_{XY}^2} - 1\right)} \begin{pmatrix} \mathrm{Var}[Y] & -\mathrm{Cov}[X,Y] \\ -\mathrm{Cov}[Y,X] & \mathrm{Var}[X] \end{pmatrix}, \end{aligned} \tag{9}$$

where ρ_{XY} is the correlation coefficient.

Problem 2

Prove the first order case of Taylor's theorem

Proof. Keep x and p fixed and consider the function $g_1 : \mathbb{R} \to \mathbb{R}$

$$g_1(s) = f(x + sp) - f(x). \tag{10}$$

We can see that $g_1(t)$ for some $t \in (0,1)$ is the residual term in the first order Taylor's expansion. We have

$$g_1(0) = 0, \tag{11}$$

and

$$g_1(1) = f(x + p) - f(x). \tag{12}$$

According to the **Mean Value Theorem**, there exists some $t \in (0,1)$ such that

$$\frac{g_1(1) - g_1(0)}{1 - 0} = \left.\frac{dg_1}{ds}\right|_t. \tag{13}$$

Or

$$\left.\frac{dg_1}{ds}\right|_t = f(x+p) - f(x). \tag{14}$$

On the other hand, by using the chain rule for derivative of composite functions, the derivative of g_1 with respect to s evaluated at t is

$$\left.\frac{dg_1}{ds}\right|_t = \nabla f(x+tp)^T p. \tag{15}$$

From Eqs. (14) and (15), we get

$$\nabla f(x+tp)^T p = f(x+p) - f(x), \tag{16}$$

or

$$f(x+p) = f(x) + \nabla f(x+tp)^T p. \tag{17}$$

□

Prove the second order case of Taylor's theorem

Proof. Similar to above, let's consider the function $g_2 : \mathbb{R} \to \mathbb{R}$

$$g_2(s) = f(x+sp) - f(x) - \nabla f(x)^T p. \tag{18}$$

$$g_2(0) = -\nabla f(x)^T p. \tag{19}$$

$$g_2(1) = f(x+p) - f(x) - \nabla f(x)^T p. \tag{20}$$

Since f is twice continuously differentiable, so is g_2. Therefore, we can apply the **Extended Mean Value Theorem** which states that:

Theorem 1 (Extended Mean Value Theorem)**.** *If f and f' are continuous on $[a,b]$ and f' is differentiable on (a, b) then there exists $c \in (a,b)$ such that*

$$f(b) - f(a) = f'(a)(b-a) + \frac{1}{2}f''(c)(b-a)^2 \tag{21}$$

Applying this theorem to g_2, we have

$$g_2(1) - g_2(0) = \left.\frac{dg_2}{ds}\right|_0 (1-0) + \frac{1}{2}\left.\frac{d^2 g_2}{ds^2}\right|_t (1-0)^2, \tag{22}$$

for some $t \in (0,1)$. This is equivalent to

$$f(x+p) - f(x) = \nabla f(x)^T p + \frac{1}{2}p^T \nabla^2 f(x+tp)p. \tag{23}$$

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2}p^T \nabla^2 f(x+tp)p. \tag{24}$$

□

Problem 3

(a) First-order necessary conditions

First, let's prove the **Descent Direction Theorem**.

Theorem 2 (Descent Direction Theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable in an open neighborhood of x^*. If there exists a vector p such that*

$$\nabla f(x^*)^T p < 0 \tag{25}$$

then there exists $\delta > 0$ such that

$$f(x^* + \lambda p) < f(x^*) \tag{26}$$

for all $\lambda \in (0, \delta)$, so that p is a descent direction of f at x^*.

Proof. Because f is differentiable, we can apply the first order case of the Taylor's theorem

$$f(x^* + \lambda p) = f(x^*) + \lambda \nabla f(x^* + tp)^T p, \tag{27}$$

for some $p \in (0, \lambda)$. Because $\nabla f(x^* + tp)^T p$, we have $f(x^* + \lambda p) < f(x^*)$. □

Now let's prove the first-order necessary conditions.

Proof. Suppose that x^* is a local minimizer and that $\nabla f(x^*) \neq 0$. Let $p = -\nabla f(x^*)$, we have

$$\nabla f(x^*)^T p = -||\nabla f(x^*)||^2 \leq 0. \tag{28}$$

By the Theorem of Descent Direction, there exists $\delta > 0$ such that $f(x^* + \lambda p) < f(x^*)$ for all $\lambda \in (0, \delta)$. This contradicts our assumption that x^* is a local minimizer. Hence $\nabla f(x^*) = 0$. □

(b) Second-order necessary conditions

Proof. We have $\nabla f(x^*) = 0$ by the first-order necessary conditions.

Suppose that $\nabla^2 f(x^*)$ is not positive semidefinite, then for some vector p we have $p^T \nabla^2 f(x^*) p < 0$. Since $\nabla^2 f(x)$ is continuous near x^*, there exists δ such that $p^T \nabla^2 f(x^* + \lambda p) p < 0$ for all $\lambda \in [0, \delta]$.

Now expand f around x^* using Taylor's theorem up to the second order

$$f(x^* + \lambda p) = f(x^*) + \nabla f(x^*)^T p + \frac{1}{2} p^T \nabla^2 f(x^* + tp) p, \tag{29}$$

for some $t \in (0, \lambda)$.

But we have $\nabla f(x^*)^T p = 0$ and $p^T \nabla^2 f(x^* + tp) p < 0 \to f(x^* + \lambda p) < f(x^*)$ which means x^* is not a local minimizer. This contradicts our assumption. So $\nabla^2 f(x^*)$ must be positive semidefinite.

□

(c) Second-order sufficient conditions

Proof. Since $\nabla^2 f$ is continuous and positive definite in a neighborhood of x^*, for an arbitrary vector p, there exists $\delta > 0$ such that $\nabla^2 f(x^* + \lambda p)$ remains positive definite for all $\lambda \in [0, \delta]$.

Now let's do Taylor's expansion around x^*

$$\begin{aligned} f(x^* + \lambda p) &= f(x^*) + \lambda \nabla f(x^*)^T p + \frac{\lambda^2}{2} p^T \nabla^2 f(x^* + tp) p \\ &= f(x^*) + \frac{\lambda^2}{2} p^T \nabla^2 f(x^* + tp) p \end{aligned} \tag{30}$$

for some $t \in (0, \lambda)$. $\nabla^2 f(x^* + tp)$ is also positive definite because $(0, \lambda)$ is a subset of $[0, \delta]$. This implies $p^T \nabla^2 f(x^* + tp) p > 0$ or equivalently, $f(x^* + \lambda p) > f(x^*)$. So x^* is a strict local minimizer of f.

□

Problem 4

(a) Eigenvalue and eigenvector of matrix power

Prove that

$$A^s u = \lambda^s u \tag{31}$$

for all integer $s > 0$, where u and λ are eigenvector and eigenvalue, respectively, of A.

We will use proof by induction.

Proof. Equality (31) holds for $s = 1$ by definition of the eigenvector and eigenvalue.

Suppose that equality (31) holds for $s = i$, $A^i u = \lambda^i u$, we need to prove that it also holds for $s = i + 1$.

Indeed

$$\begin{aligned} A^{i+1} u &= A(A^i u) \\ &= A(\lambda^i u) \\ &= \lambda^i A u \\ &= \lambda^i \lambda u \\ &= \lambda^{i+1} u \end{aligned} \tag{32}$$

So equality (31) hold for all $s > 0$. □

(b) Woodbury matrix identity

Prove that

$$(A + XBX^T)^{-1} = A^{-1} - A^{-1} X (B^{-1} + X^T A^{-1} X)^{-1} X^T A^{-1}. \tag{33}$$

Proof. Multiply both side of Eq (33) by $(A + XBX^T)$, we get the identity matrix I on the left side

$$
\begin{aligned}
I &= (A + XBX^T)[A^{-1} - A^{-1}X(B^{-1} + X^TA^{-1}X)^{-1}X^TA^{-1}] \\
&= I - X(B^{-1} + X^TA^{-1}X)^{-1}X^TA^{-1} + XBX^TA^{-1} - XBX^TA^{-1}X(B^{-1} + X^TA^{-1}X)^{-1}X^TA^{-1} \\
&= I + XBX^TA^{-1} - X(B^{-1} + X^TA^{-1}X)^{-1}X^TA^{-1} - XBX^TA^{-1}X(B^{-1} + X^TA^{-1}X)^{-1}X^TA^{-1} \\
&= I + XBX^TA^{-1} - (X + XBX^TA^{-1}X)(B^{-1} + X^TA^{-1}X)^{-1}X^TA^{-1} \\
&= I + XBX^TA^{-1} - XB(B^{-1} + A^TA^{-1}X)(B^{-1} + X^TA^{-1}X)^{-1}X^TA^{-1} \\
&= I + XBX^TA^{-1} - XBX^TA^{-1} \\
&= I
\end{aligned}
\tag{34}
$$

□

Problem 5

Prove that all isolated local minimizers are strict.

Proof. Suppose that x^* is a local minimizer but not strict. Then for any vector p, there exists $\delta > 0$ such that $f(x^* + \lambda p) = f(x^*)$ for all $\lambda \in [0, \delta]$. This means that there are infinitely many local minimizers within the neighborhood of x^* determined by p and δ. This contradicts our assumption that x^* is an isolated local minimizer. Therefore, x^* must be strict local minimizer. □