Matrix Analysis

Part. 4

Chang Jae Ho
jaehochang@konkuk.ac.kr

February 28, 2020

Contents

Contents

Multivariable Differential Calculus I

For the multivariate function $f(x)$,

$$f(x+u) = f(u) + \frac{\partial}{\partial x'} f(x) u + r_1(u,x)$$

where $\frac{r_1(u,x)}{||u||_2} \overset{u \to 0}{\to} 0$. Or,

$$f(x+u) = f(u) + \frac{\partial}{\partial x'} f(x) u + \frac{u' H_f u}{2!} + r_2(u,x)$$

where $\frac{r_2(u,x)}{||u||_2^2} \overset{u \to 0}{\to} 0$ and $H_f := \frac{\partial^2}{\partial x \partial x'} f(x)$.

Contents

Vector and Matrix Functions I

Suppose that $\{f_i\}_{i=1}^m$ is a sequence of functions of a vector $x = (x_1, \cdots, x_n)'$. A vector function can be expressed as

$$f(x) = (f_1(x), \cdots, f_m(x))'.$$

The vector function f is diff'ble at x iff each f_i is diff'ble at x. The Taylor formula is given by

$$f(x+u) = f(x) + \frac{\partial}{\partial x'} f(x) u + r_1(u, x)$$

where $\frac{r_1(u,x)}{||u||_2} \overset{u \to 0}{\to} 0$ and

$$\frac{\partial}{\partial x'} f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f_1(x) & \frac{\partial}{\partial x_2} f_1(x) & \cdots & \frac{\partial}{\partial x_n} f_1(x) \\ \frac{\partial}{\partial x_1} f_2(x) & \frac{\partial}{\partial x_2} f_2(x) & \cdots & \frac{\partial}{\partial x_n} f_2(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} f_m(x) & \frac{\partial}{\partial x_2} f_m(x) & \cdots & \frac{\partial}{\partial x_n} f_m(x) \end{bmatrix} = \left\{ \frac{\partial}{\partial x_j} f_i(x) \right\}_{m \times n}$$

which is referred to as the Jacobian matrix of f at x.

Vector and Matrix Functions II

If we obtain the first differential of f at x in u and write it in the form $df = Bu$, then the $m \times n$ matrix B must be the derivative of f at x. If $y(x) = g\{f(x)\}$ where $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^p$, then the generalization of the chain rule is

$$\frac{\partial}{\partial x_i} y(x) = \sum_{j=1}^{m} \left\{ \frac{\partial}{\partial f_j} g(f) \right\} \left\{ \frac{\partial}{\partial x_i} f_j(x) \right\} = \left\{ \frac{\partial}{\partial f'} g(f) \right\} \left\{ \frac{\partial}{\partial x_i} f_j(x) \right\}$$

for $i = 1, \cdots, n$, or simply

$$\frac{\partial}{\partial x'} y(x) = \left\{ \frac{\partial}{\partial f'} g(f) \right\} \left\{ \frac{\partial}{\partial x'} f(x) \right\}.$$

Vector and Matrix Functions III

The most general case involves the $p \times q$ matrix function $F(X) = \{f_{ij}(X)\}_{p \times q}$ of the $m \times n$ matrix X. If we define $g : \mathbb{R}^{mn} \to \mathbb{R}^{pq}$ so that $g\{vec(X)\} := vec\{F(X)\}$, the Jacobian matrix of F at X is given by the $mn \times pq$ matrix

$$\frac{\partial}{\partial vec(X)'} g\{vec(X)\} = \frac{\partial}{\partial vec(X)'} vec\{F(X)\}.$$

Basic properties of vector and matrix differentials are

1. $da = 0$,
2. $d(\alpha x + \beta y) = \alpha dx + \beta dy$,
3. $d(xy) = (dx)y + x(dy)$,
4. $dx^a = ax^{a-1}dx$,
5. $de^x = e^x dx$,
6. $d \log x = x^{-1}dx$.

Vector and Matrix Functions IV

1. $dA = (0)$,
2. $d(\alpha X + \beta Y) = \alpha dX + \beta dY$,
3. $d(X') = (dX)'$,
4. $d(XY) = (dX)Y + X(dY)$.

Vector and Matrix Functions V

Example

If we define $f(b) = Xb$,

$$\frac{\partial}{\partial b'} f(b) = \frac{\partial}{\partial b'} (X_1 b_1 + \cdots + X_p b_p) = \begin{bmatrix} X_1 & (0) \end{bmatrix} + \cdots + \begin{bmatrix} (0) & X_p \end{bmatrix} = X.$$

Next, if we define $g(b) = b'b$,

$$\frac{\partial}{\partial b'} g(b) = \frac{\partial}{\partial b'} \sum_{j=1}^{p} b_j^2 = 2(b_1 e_1' + \cdots + b_p e_p') = 2b'.$$

Therefore,

$$\frac{\partial}{\partial b'} g\{f(b)\} = \left\{ \frac{\partial}{\partial f'} g(f) \right\} \left\{ \frac{\partial}{\partial b'} f(b) \right\} = 2(f)'X = 2b'X'X.$$

Vector and Matrix Functions VI

Example

Let $z \in S_1 \subseteq \mathbb{R}^m, z \sim \mathcal{N}_p(0, I_m)$ and $x := g(z)$ represent a one-to-one mapping of S_1 to $S_2 \subseteq \mathbb{R}^m$. Then, $z = g^{-1}(x)$ is unique for $x \in S_2$. Denote the Jacobian matrix of z at x as

$$J := \frac{\partial}{\partial x'} z$$

if the partial derivatives in J exist and are continuous functions on the set S_2. Then, the density of x is given by

$$f_2(x) = f_1(z) \left| J \right| .$$

If $x = \mu + Tz$ where $TT' = \Sigma$, $z = T^{-1}(x - \mu)$, so $dz = T^{-1} dx$ and $J = T^{-1}$.

Contents

Maxima and Minima I

Theorem

Let $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ and a be an interior point of S, that is, $\exists \delta > 0$ such that $a + u \in S$ for any $u'u < \delta$. If f has a local maximum at a and f is differential at a, then

$$\frac{\partial}{\partial a'} f(a) = 0'.$$

Here, a is called a **stationary point** *of f.*

Maxima and Minima II

Theorem

Given the assumption above, suppose also that $\exists f^{(2)}$ *at* a*. If* a *is a stationary point of* f *and* H_f *is the Hessian matrix of* f *at* a*, then*

1. a *is a* **local maxima** *if* H_f *is n.d.,*
2. a *is a* **local minima** *if* H_f *is p.d.,*
3. a *is a* **saddle point** *if* H_f *is neither n.d, p.d nor singular,*
4. a *is* **undecided** *if* H_f *is singular.*

Contents

Convex and Concave Functions I

Definition

Let $f : S \subseteq \mathbb{R}^m \to \mathbb{R}$ with a convex set S. Then, f is a convex function on S if

$$f\{cx_1 + (1-c)x_2\} \leq cf(x_1) + (1-c)f(x_2)$$

for all $x_1, x_2 \in S$ and $0 \leq c \leq 1$.

Theorem

Let f be a real-valued convex function on an open convex set $S \subseteq \mathbb{R}^m$. If $\exists f^{(1)}$ and $a \in S$, then

$$f(x) \geq f(a) + \left\{ \frac{\partial}{\partial x'} f(x) \Big|_{x=a} \right\} (x-a)$$

for all $x \in S$.

Convex and Concave Functions II

Theorem

For a convex set $S \subseteq \mathbb{R}^m$ and a random $y \in \mathbb{R}^m$, if $\mathcal{P}(y \in S) = 1$, then $\mathrm{E}(y) \in S$.

Theorem

Let $f : S \subseteq \mathbb{R}^m \to \mathbb{R}$ be a convex function for a convex set S. If y is an $m \times 1$ random vector and $\mathcal{P}(y \in S) = 1$, then

$$\mathrm{E}\{f(y)\} \geq f\{\mathrm{E}(y)\}.$$

Contents

Lagrange Multipliers I

For $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$, let $T = \{x; x \in \mathbb{R}^n, g(x) = 0\} \subset S$ where g is a vector function of $\{g_i\}_{i=1}^m$.
Define Lagrange function

$$L(x, \lambda) = f(x) - \lambda' g(x),$$

where λ is a vector with $\{\lambda_i\}_{i=1}^m$. The stationary point of $L(x, \lambda)$ satisfies:

$$\frac{\partial}{\partial x'} L(x, \lambda) = \frac{\partial}{\partial x'} f(x) - \lambda' \left\{ \frac{\partial}{\partial x'} g(x) \right\} = 0',$$
$$\frac{\partial}{\partial \lambda'} L(x, \lambda) = -g(x)' = 0'.$$

Therefore, we can expect that the local extrema of f would satisfy two equations above.

Lagrange Multipliers II

Theorem

Let $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ and $g : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ where $m < n$. Let a be an interior point of S and the following hold.

- a. $\exists f^{(2)}(a)$ *and* $\exists g^{(2)}(a)$.
- b. $\frac{\partial}{\partial a'} g(a)$ *has full rank* m, *and* $g(a) = 0$.
- c. $\frac{\partial}{\partial a'} L(a, \lambda) = 0'$, *where* $L(x, \lambda) := f(x) - \lambda' g(x)$ *and* $\lambda = (\lambda_1, \cdots, \lambda_m)'$.

If we define $A = H_f(a) - \sum_{i=1}^{m} \lambda_i H_{g_i}(a)$ *and* $B := \frac{\partial}{\partial a'} g(a)$, *then* f *has a local maximum at* $x = a$ *subject to* $g(x) = 0$ *if*

$$x' A x < 0 \text{ for all } x \neq 0 \text{ for which } Bx = 0.$$

Remark The final condition indicates a constrained negative definiteness.

Contents

Idempotent Matrices I

Theorem

Let A be an $m \times m$ idempotent matrix. Then,

- a $\lambda(A) = 0$ *or* 1,
- b A *is diagonalizable,*
- c $rank(A) = tr(A)$.

Theorem

Let A be an $m \times m$ symmetric matrix. Then A is idempotent iff each eigenvalue of A is 0 *or* 1.

Idempotent Matrices II

Theorem

Let A, B be $m \times m$ idempotent matrices. Then,

a. $A + B$ *is idempotent iff* $AB = BA = (0)$.
b. AB *is idempotent if* $AB = BA$.

Theorem

Suppose A is an $m \times m$ symmetric idempotent matrix. Then,

a. $0 \leq a_{ii} \leq 1$ *for* $i = 1, \cdots, m$,
b. *if* $a_{ii} = 0$ *or* 1, *then* $a_{ij} = 0 \; \forall i \neq j$.

Theorem

Suppose that for some positive integer i, the $m \times m$ symmetric matrix A satisfies $A^{i+1} = A^i$. Then A is idempotent.

Contents

Expectations I

Theorem

Let x be an $m \times 1$ random vector with $E|X|^4 < \infty$, so that $\exists E(xx'), E(xx' \otimes xx')$. Put $E(x) = \mu$ and $Var(x) = \Omega$, and take $m \times m$ symmetric matrices A, B. Then,

1. $E(x'Ax) = tr\{AE(xx')\} = tr(A\Omega) + \mu' A\mu$,
2. $Var(x'Ax) = tr\{(A \otimes A)E(xx' \otimes xx')\} - \{tr(A\Omega) + \mu' A\mu\}^2$
3. $Cov(x'Ax, x'Bx) =$ $tr\{(A \otimes B)E(xx' \otimes xx')\} - \{tr(A\Omega) + \mu' A\mu\}\{tr(B\Omega) + \mu' B\mu\}$.

Expectations II

Note that

$$\begin{aligned} E(x'Axx'Bx) &= E\left\{(x'Ax) \otimes (x'Bx)\right\} = E\left\{(x' \otimes x')(A \otimes B)(x \otimes x)\right\} \\ &= E\left[tr\left\{(A \otimes B)(x \otimes x)(x' \otimes x')\right\}\right] \\ &= tr\left\{(A \otimes B)E(xx' \otimes xx')\right\}. \end{aligned}$$