A Brief Introduction to Matrix Calculus

Presentor: Chun-Jen Shih

Advisor: Jian-Jiun Ding

Table of Contents

Table of Contents

First Derivative

For a vector-valued function $f(x) = (f_1(x), \dots, f_m(x))' : \mathbb{R}^n \to \mathbb{R}^m$, its first derivative is the $m \times n$ Jacobian

$$\frac{\partial f(x)}{\partial x'} = \begin{pmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \cdots & \partial f_1/\partial x_n \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \cdots & \partial f_2/\partial x_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial f_m/\partial x_1 & \partial f_m/\partial x_2 & \cdots & \partial f_m/\partial x_n \end{pmatrix} \tag{1}$$

- The output $f(x)$ is a column vector, while the input x' is a row vector.

First Derivative

For a matrix-valued function $F(X)$ of a matrix X, one should impose a one-dimensional ordering vec on both outputs and inputs. Then, matrix calculus is vector calculus on $\mathrm{vec}\, F$ and $\mathrm{vec}\, X$:

$$DF(X) = \frac{\partial \,\mathrm{vec}\, F(X)}{\partial (\mathrm{vec}\, X)'} \tag{2}$$

- The vectorization operator vec stacks columns into a single column vector.

First Differential

For a scalar function f of one variable, the first differential at x is defined as the linear map

$$df(x)(u) = f'(x)u, \tag{3}$$

where u is an arbitrary increment.

- The first differential is a **geometric linear map**: change of f at x due to an arbitrary perturbation u.
- Because $dx(u) = u$, we have $df(x)(u) = f'(x)dx(u)$
  $\rightarrow$ short-hand notation: $df(x) = f'(x)dx$.

First Differential

For a vector-valued function $f(x) : \mathbb{R}^n \to \mathbb{R}^m$, its first differential is

$$df(x) = Df(x)dx \tag{4}$$

- The derivative $Df(x)$ is an $m \times n$ matrix, but the differential $df(x)$ is still an $m \times 1$ object.
- While a derivative changes shape depending on the input dimension, a differential has the **same dimension as the function itself**
- For a matrix-valued function $F(X)$ of a matrix X, its first differential is $d\,\mathrm{vec}\,F(X) = DF(X)d\,\mathrm{vec}\,X$

Basic Computation Properties

- $da = 0$
- $d(x') = (dx)'$
- $d(a'x) = a'dx$
- $d(x+y) = dx + dy$
- $d(x'y) = (dx)'y + x'dy$
- Let $z = f(y)$, $y = g(x)$, $z = f(g(x))$
  - Chain Rule: $\frac{\partial z}{\partial x'} = \frac{\partial z}{\partial y'} \frac{\partial y}{\partial x'}$
  - Cauchy Invariance
    If $dz = A(y)dy$, $dy = B(x)dx$, then $dz = A(y)B(x)dx$.

Easy Examples for the First Differential

- $d(x^2) = 2xdx$, $d(\sin x) = (\cos x)dx$
- $d\left(e^{\sin(x^2)}\right) = e^{\sin(x^2)}d(\sin(x^2)) = e^{\sin(x^2)}\cos(x^2)d(x^2) = 2xe^{\sin(x^2)}\cos(x^2)dx$
- $f(x) = x_1^2 x_2$
  - Method 1
    $\frac{\partial f}{\partial x_1} = 2x_1x_2$, $\frac{\partial f}{\partial x_2} = x_1^2$. Thus,
    $df(x) = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 = 2x_1x_2dx_1 + x_1^2dx_2$
  - Method 2
    $df(x) = d(x_1^2)x_2 + x_1^2dx_2 = 2x_1x_2dx_1 + x_1^2dx_2$

Easy Examples for the First Differential

- $f(x) = \begin{pmatrix} x_1^2 - x_2^2 \\ x_1 x_2 x_3 \end{pmatrix}, x = (x_1, x_2, x_3)'$
  - Method 1

    $$\frac{\partial f(x)}{\partial x'} = \begin{pmatrix} 2x_1 & -2x_2 & 0 \\ x_2 x_3 & x_1 x_3 & x_1 x_2 \end{pmatrix} \rightarrow df(x) = \frac{\partial f(x)}{\partial x'} \begin{pmatrix} dx_1 \\ dx_2 \\ dx_3 \end{pmatrix}$$
  - Method 2

    $$df(x) = \begin{pmatrix} d(x_1^2) - d(x_2^2) \\ d(x_1 x_2 x_3) \end{pmatrix}$$

    $$= \begin{pmatrix} 2x_1 dx_1 - 2x_2 dx_2 \\ (x_2 x_3) dx_1 + (x_1 x_3) dx_2 + (x_1 x_2) dx_3 \end{pmatrix}$$

Why Differential Matters?

The aforementioned examples are quite simple. So, we can easily compute the first derivative (Jacobian) and the first differential seems to be just a byproduct.

However, how do we compute the first derivatives of non-trivial examples, such as $f(X) = \mathrm{tr}(X'AX)$, $\log|X'X|$, $\mathrm{tr}(X^k)$ or $F(X) = AX^{-1}B$?

An important methodology is to **first compute the differential, express it as the linear form, and then identify the first derivative**.

Table of Contents

First Identification Theorem

$$df(x) = A(x)dx \iff Df(x) = A(x) \tag{5}$$

$$d \operatorname{vec} F(X) = A(X)d \operatorname{vec} X \iff DF(X) = A(X) \tag{6}$$

- $Df(x)dx = df(x) = A(x)dx \Rightarrow (Df(x) - A(x))dx = 0 \ \forall dx$
  This implies $Df(x) - A(x) = 0$.
- $DF(X)d \operatorname{vec} X = d \operatorname{vec} F(X) = A(X)d \operatorname{vec} X \ \forall d \operatorname{vec} X$
  This implies $DF(X) = A(X)$.

Table of Contents

Basic Computation Properties (Continued)

- $dA = 0$, $d(X') = (dX)'$, $d(\alpha X) = \alpha dX$,
  $d(X+Y) = dX + dY$, $d(XY) = (dX)Y + XdY$
- If X is square, then $d\,\mathrm{tr}\,X = \mathrm{tr}(dX)$.
- For a non-singular square matrix X, $d|X| = |X|\,\mathrm{tr}(X^{-1}dX)$.
  - $d\log|X| = \frac{d|X|}{|X|} = \mathrm{tr}(X^{-1}dX)$ for $|X| > 0$
- For a non-singular X, $dX^{-1} = -X^{-1}(dX)X^{-1}$.
- Cauchy Invariance
  If $Z = F(Y)$, $Y = G(X)$, $Z = F(G(X))$, and if
  $dZ = A(Y)dY$, $dY = B(X)dX$, then $dZ = A(Y)B(X)dX$.

Important Linear Algebra Tools

- $\operatorname{tr}(A) = \operatorname{tr}(A')$
- $\operatorname{tr}(A'B) = \operatorname{tr}(BA')$ for any two matrices A and B of the same order
- $\operatorname{tr}(A'B) = (\operatorname{vec} A)'(\operatorname{vec} B)$.

Quadratic Forms

- $f(x) = x'Ax$
  $df(x) = (dx)'Ax + x'(dA)x + x'Adx = (Ax)'dx + x'Adx$
  $\rightarrow Df(x) = x'(A' + A)$

- $f(X) = \operatorname{tr}(X'AX)$
  $$\begin{aligned} df(X) &= \operatorname{tr}(dX'AX) = \operatorname{tr}((dX)'AX) + \operatorname{tr}(X'AdX) \\ &= \operatorname{tr}(X'(A' + A)dX) \\ &= \operatorname{vec}((A' + A)X)'d\operatorname{vec} X \end{aligned}$$
  Thus, $Df(X) = \operatorname{vec}((A' + A)X)'$

Log Determinant

$f(X) = \log|X'X|$, with X assumed to have full column rank

$$\begin{aligned} df(X) &= \operatorname{tr}((X'X)^{-1} dX'X) \\ &= \operatorname{tr}((X'X)^{-1}((dX)'X + X'dX)) \\ &= 2\operatorname{tr}((X'X)^{-1}X'dX) \\ &= 2\operatorname{vec}(X(X'X)^{-1})' d\operatorname{vec} X \end{aligned}$$

Thus, $Df(X) = 2\operatorname{vec}(X(X'X)^{-1})'$

Power

$f_k(X) = \operatorname{tr}(X^k)$, $k = 1, 2, \cdots$

$$\begin{aligned} df_k(X) &= \operatorname{tr}((dX)X^{k-1} + X(dX)X^{k-2} + \cdots + X^{k-1}dX) \\ &= \operatorname{tr}\left(\sum_{i=0}^{k-1} X^i (dX) X^{k-i-1}\right) \\ &= \operatorname{tr}\left(\sum_{i=0}^{k-1} X^{k-1} dX\right) \\ &= k \operatorname{tr}(X^{k-1} dX) \\ &= k \operatorname{vec}(X'^{k-1})' d \operatorname{vec} X \end{aligned}$$

Thus, $Df_k(X) = k \operatorname{vec}(X'^{k-1})'$

Important Linear Algebra Tools (Continued)

- Given $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$, the Kronecker product $A \otimes B$ is the block matrix $A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}$
- $\operatorname{vec}(ABC) = (C' \otimes A) \operatorname{vec} B$

Inverse

$F(X) = AX^{-1}B$, where X is non-singular.

$$\begin{aligned} dF(X) &= A(dX^{-1})B \\ &= -AX^{-1}(dX)X^{-1}B \end{aligned}$$

To compute the Jacobian, we need to vectorize $F(X)$.

$$\begin{aligned} d\,\mathrm{vec}\,F(X) &= -\,\mathrm{vec}(AX^{-1}(dX)X^{-1}B) \\ &= -((X^{-1}B)' \otimes (AX^{-1}))d\,\mathrm{vec}\,X \end{aligned}$$

Thus, $DF(X) = -(X^{-1}B)' \otimes (AX^{-1})$

Table of Contents

Second Derivative and Second Differential

For a **scalar** function $f : \mathbb{R}^n \to \mathbb{R}$, if $df(x) = \nabla f(x)' dx$, where $\nabla f(x)' = \frac{\partial f(x)}{\partial x'}$, then the second derivative (Hessian matrix) is defined through

$$d\nabla F(x) = Hf(x)dx, \tag{7}$$

where $Hf(x) = \frac{\partial \nabla f(x)}{\partial x'} = \frac{\partial^2 f(x)}{\partial x \partial x'}$. Thus,

$$d^2 f(x) = d(\nabla f(x)')dx = (dx)' Hf(x)' dx = (dx)' Hf(x) dx \tag{8}$$

- The Hessian matrix is symmetric.
- The second differential is a quadratic form in dx.

Second Derivative and Second Differential

- For a **scalar** function $f(X)$ of a matrix X, its second differential and Hessian matrix satisfy

$$d^2 f(X) = (d\,\mathrm{vec}\,X)' H f(X)(d\,\mathrm{vec}\,X) \tag{9}$$

- For a vector-valued function $F : \mathbb{R}^n \to \mathbb{R}^m$, its second differential should be $d^2 F(x) = \begin{pmatrix} dx' H f_1(x) dx \\ \vdots \\ dx' H f_m(x) dx \end{pmatrix}$, where $f_1, \dots, f_m$ are the components of F. So there are m Hessian matrices, one for each component.

- Likewise, if $F(X) \in \mathbb{R}^{p \times q}$, then each entry $F_{ij}(X)$ has its own Hessian matrix with respect to $\mathrm{vec}\,X$.

Table of Contents

Second Identification Theorem

$$d^2 f(x) = (dx)' B(x) dx \iff H f(x) = \frac{B(x) + B(x)'}{2} \tag{10}$$

$$d^2 f(X) = (d \operatorname{vec} X)' B(X) d \operatorname{vec} X \iff H f(X) = \frac{B(X) + B(X)'}{2} \tag{11}$$

- $(dx)' H f(x) dx = d^2 f(x) = (dx)' B(x) dx \Rightarrow$
  $(dx)'(H f(x) - B(x)) dx \; \forall dx$

  Does this imply $H f(x) - B(x) = 0$?

  No! This implies **the symmetric part of** $H f(x) - B(x)$ **is zero**
  $\rightarrow \frac{(H f(x) - B(x)) + (H f(x) - B(x))'}{2} = 0 \Rightarrow H f(x) = \frac{B(x) + B(x)'}{2}$

Table of Contents

Important Linear Algebra Tools (Continued)

- $(A \otimes B)' = A' \otimes B'$
- For $A \in \mathbb{R}^{m \times n}$, vec A and vec A' contain the same entries in different order. The commutation matrix K_{mn} is the unique $mn \times mn$ permutation matrix satisfying $K_{mn}\,\mathrm{vec}\,A = \mathrm{vec}\,A'$
- $K'_{mn} = K^{-1}_{mn} = K_{nm}$
- $K_{pm}(A \otimes B) = (B \otimes A)K_{qn}$, for $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{p \times q}$.

Two Special Trace Forms

- $d^2 f(X) = \operatorname{tr} A(dX)' B dX$ for $X \in \mathbb{R}^{n\times q}$, $A \in \mathbb{R}^{q\times q}$, $B \in \mathbb{R}^{n\times n}$

$$\begin{aligned} \operatorname{tr} A(dX)' B dX &= (d \operatorname{vec} X)' \operatorname{vec}(B(dX)A) \\ &= (d \operatorname{vec} X)'(A' \otimes B) d \operatorname{vec} X \end{aligned}$$

  Thus, $Hf(X) = \frac{1}{2}(A' \otimes B + A \otimes B')$

- $d^2 f(X) = \operatorname{tr} A(dX) B dX$ for $X \in \mathbb{R}^{n\times q}$, $A \in \mathbb{R}^{q\times n}$, $B \in \mathbb{R}^{q\times n}$

$$\begin{aligned} \operatorname{tr} A(dX) B dX &= (d \operatorname{vec} X)' \operatorname{vec}(B'(dX)'A') \\ &= (d \operatorname{vec} X)'(A \otimes B')(d \operatorname{vec} X)' \\ &= (d \operatorname{vec} X)'(A \otimes B') K_{nq} (d \operatorname{vec} X) \end{aligned}$$

  Thus, $Hf(X) = \frac{1}{2} K_{qn}(A' \otimes B + B' \otimes A)$

Quadratic Forms

- $f(x) = x'Ax$ and we know that $df(x) = (Ax)'dx + x'Adx$
  $d^2f(x) = (dx)'A'dx + (dx)'Adx = (dx)'(A' + A)dx$
  Thus, $Hf(x) = A' + A$.

- $f(X) = \mathrm{tr}(X'AX)$ and we know that
  $df(X) = \mathrm{tr}(X'(A' + A)dX)$
  $d^2f(X) = \mathrm{tr}((dX)'(A' + A)dX) = \mathrm{tr}(I_q(dX)'(A' + A)dX)$
  Thus, $Hf(X) = I_q \otimes (A' + A)$ by the first special trace form .

Log Determinant

$f(X) = \log|X'X|$, with X assumed to have full column rank

We know that $df(X) = 2\,\mathrm{tr}((X'X)^{-1}X'dX)$

$d^2f(X) = 2\,\mathrm{tr}((d(X'X)^{-1})X'dX) + 2\,\mathrm{tr}((X'X)^{-1}(dX')dX)$

- The first term is $-2\,\mathrm{tr}((X'X)^{-1}(X'dX + (dX)'X)(X'X)^{-1}X'dX)$. By using the two special trace forms, its Hessian matrix is $-2K_{qn}(C \otimes C') - 2(X'X)^{-1} \otimes (X(X'X)^{-1}X')$, where $C = X(X'X)^{-1}$.
- The Hessian matrix of the second term is $2(X'X)^{-1} \otimes I_n$ by using the first special trace form.

Thus,

$Hf(X) = 2(X'X)^{-1} \otimes (I_n - X(X'X)^{-1}X') - 2K_{qn}(C \otimes C')$

Power

$f_k(X) = \mathrm{tr}(X^k)$, $k = 1, 2, \cdots$
We know that $df_k(X) = k\,\mathrm{tr}(X^{k-1}dX)$
$d^2 f_1(X) = 0$, and for $k \geq 2$,

$$\begin{aligned} d^2 f_k(X) &= k\,\mathrm{tr}((dX^{k-1})dX) \\ &= k\,\mathrm{tr}\left(\sum_{i=0}^{k-2} X^i(dX)X^{k-2-i}dX\right) \end{aligned}$$

From this, the Hessian is $Hf_1(X) = 0$, and for $k \geq 2$,
$Hf_k(X) = \frac{k}{2}\sum_{i=0}^{k-2} K_n(X'^i \otimes X^{k-2-i} + X'^{k-2-i} \otimes X^i)$ by using the second special trace form.

Inverse

$F(X) = AX^{-1}B$, where X is non-singular.

We know that $dF(X) = -AX^{-1}(dX)X^{-1}B$

Differentiating again gives $d^2F(X) = 2AX^{-1}(dX)X^{-1}(dX)X^{-1}B$.

The Hessian matrix **for each component** $F_{st}(X)$ is derived as follows.

$$\begin{aligned} d^2F_{st}(X) &= 2e_s' AX^{-1}(dX)X^{-1}(dX)X^{-1}Be_t \\ &= 2 \operatorname{tr} C_{ts}(dX)X^{-1}(dX), \end{aligned}$$

where $C_{ts} \triangleq X^{-1}Be_t e_s' AX^{-1}$.

Thus, $HF_{st}(X) = \frac{\partial^2 F_{st}}{\partial(\operatorname{vec} X)\partial(\operatorname{vec} X)'} = K_n(C_{ts}' \otimes X^{-1} + X'^{-1} \otimes C_{ts})$

Table of Contents

J. R. Magnus, ”A Gentle Introduction to Matrix Calculus,” *Journal of Econometrics*, vol. 244, no. 1, p. 105862, 2024.