Matrix Calculus Note

Yi Jiang

August 11, 2021

1 Introductory Example

$$\frac{\partial f}{\partial x} = a \tag{1}$$

for multivariate, we have:

$$f(x) = \sum_i a_i x_i = a^T x$$

$$\frac{\partial f}{\partial x_k} = \frac{\partial(\sum_i a_i x_i)}{\partial x_k} = a_k$$

Thenn we organize n partial derivatives in the following way:

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a \tag{2}$$

2 Derivation

2.1 Organization of Elements

For a scalar valued function $f(x)$, the result of $\frac{\partial f}{\partial x}$ has the same size with x. That

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix} \tag{3}$$

By this definition, we have:

$$\frac{\partial f}{\partial x} = (\frac{\partial f}{\partial x})^T = a^T$$

2.2 Deal with Inner Product

For $f(x) = a^T x$, we have$\frac{\partial f}{\partial x} = a$.

2.3 Properties of Trace

Defintion 2. Trace $Tr[A] = \sum_i A_{ii}$ $\frac{\partial Tr[A]}{\partial A} = I$

Theorem 1 *Matrix traces has the following properties.*

1. $Tr[A^T B] = \sum_i \sum_j A_{ij} Bij$

If therel's a multivariate scalar function $f(x) = \mathrm{Tr}[A^T x]$, we have $\frac{\partial f}{\partial x} = A$.

2.4 Deal with Generalized Inner Product

2.5 Define Matrix Differential

$\mathrm{dTr}[A] = \mathrm{Tr}[\mathrm{d}A]$

$\mathrm{d}f = \mathrm{Tr}\left[(\frac{\partial f}{\partial x})^T dx\right]$

2.6 Matrix Differential Properties

eg. 4. Given function $f(x) = x^T Ax$, where A is square and x is a column vertor, we can compute:

eg. 5. $d(X^{-1}) = -X^{-1} dX X^{-1}$.

2.7 Schame of Handlding Scalar Function

2.8 Determinant