Derivative

$f: \mathbb{R}^n \to \mathbb{R}^m \quad x \in \text{int dom} f$

f is differentiable if there exists a matrix $Df(x) \in \mathbb{R}^{m \times n}$ such that

$$\lim_{z \in \text{dom},\ z \to x} \frac{\| f(z) - f(x) - Df(x)(z-x) \|_2}{\| z-x \|_2} = 0$$

$Df(x)$ is the derivative or Jacobian of f at x

First-order approximation

$$f(z) \approx f(x) + Df(x)(z-x)$$

$$Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}$$

$$\begin{bmatrix} df_1(x)/dx \\ df_2(x)/dx \\ \vdots \\ df_m(x)/dx \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} df_1 \\ df_2 \\ \vdots \\ df_m \end{bmatrix}$$

Log-determinant

$f(x) = \log\det X \qquad X \in S^n_{++}$

let $Z \in S^n_{++}$ be close to X, $\quad \Delta X = Z - X$

$$\begin{aligned}
\log\det(Z) &= \log\det(X + \Delta X) \\
&= \log\det\left[X^{1/2} (I + X^{-1/2} \Delta X\, X^{-1/2}) X^{1/2} \right] \\
&= \log\det(X) + \log\det(I + X^{-1/2} \Delta X\, X^{-1/2}) \\
&= \log\det(X) + \sum_{i=1}^{n} \log(1 + \lambda_i)
\end{aligned}$$

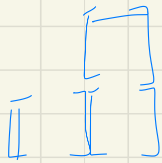
λ_i is the ith eigenvalue of $X^{-1/2} \cdot \Delta X \cdot X^{-1/2}$

$$\log(1+\lambda_i) \approx \log(1) + \log'(1) \cdot \lambda_i$$

$$= 0 + \frac{1}{1} \cdot \lambda_i \approx \lambda_i \qquad \text{for small } \lambda_i$$

$\because \Delta X$ is very small

$\therefore \log(1+\lambda_i) \approx \lambda_i$



$$\log\det(Z) \approx \log\det(X) + \sum_{i=1}^{n} \lambda_i$$

$$= \log\det(X) + tr\left(X^{-1/2}\cdot \Delta X\cdot X^{-1/2}\right) \qquad tr(AB) = tr(BA)$$

$$= \log\det(X) + tr\left(X^{-1}\cdot \Delta X\right)$$

$$= \log\det(X) + tr\left[X^{-1}(Z-X)\right]$$

$$= \log\det(X) + tr\left[X^{-T}(Z-X)\right]$$ Standard form inner product

$$\log\det(Z) \approx \log\det(X) + tr\left[X^{-T}(Z-X)\right]$$

$$f(X) = \log\det(X) \quad X \in S^n_{++} \qquad \nabla f(X) = X^{-T} = X^{-1}$$

Chain Rule

$f: \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in \text{int dom } f$

$g: \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $f(x) \in \text{int dom } g$

$h: \mathbb{R}^n \to \mathbb{R}^p \qquad h(x) = g(f(x))$

$$Dh(x) = Dg(f(x))\cdot Df(x)$$

$$\begin{bmatrix} dh_1/df \\ dh_2/df \\ \vdots \\ dh_p/df \end{bmatrix} \qquad \begin{bmatrix} df_1/dx_1 & df_1/dx_2 & \cdots & df_1/dx_n \\ df_2/dx_1 & \vdots & & \vdots \\ \vdots & & & \\ df_m/dx_1 & & \cdots & df_m/dx_n \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix}$$

$$\begin{bmatrix} dh_1/dx_1 & dh_1/dx_2 & \cdots & \\ \vdots & & & \vdots \\ dh_p/dx_1 & & \cdots & dh_p/dx_n \end{bmatrix} \begin{bmatrix} dh_1 \\ \vdots \\ dh_p \end{bmatrix}$$

Composition with Affine function

$f: \mathbb{R}^n \to \mathbb{R}^m$ is differentiable

$g: \mathbb{R}^p \to \mathbb{R}^m \qquad g(x) = f(Ax+b) \qquad A \in \mathbb{R}^{n\times p}$

$$Dg(x) = Df(Ax+b)\cdot A = A^T \cdot \nabla f(Ax+b)$$

Line restriction

$f: \mathbb{R}^n \to \mathbb{R} \qquad x, v \in \mathbb{R}^n$

$$\tilde{f}(t) = f(x+tv)$$

$$D\tilde{f}'(t) = \tilde{f}'(t) = \nabla f(x+tv)^T v$$

$\tilde{f}'(0)$ is the directional derivative of f at x in the direction v

Eg. $f(x) = \log \sum_i \exp(a_i^T x + b_i)$

$g(y) = \log \sum_i \exp(y_i)$

$$\nabla g(y) = \frac{1}{\sum_i \exp(y_i)} \cdot \begin{bmatrix} \exp(y) \end{bmatrix}$$

$$\nabla f(x) = \frac{1}{\sum_i \exp(y_i)} \cdot A^T \cdot \begin{bmatrix} \exp(y) \end{bmatrix}$$

Eg. $f(x) = \log\det(F_0 + x_1 F_1 \cdots + x_n F_n)$ $\quad F_i \in S^n$

$\mathrm{dom}\, f = \{x \mid F_0 + x_1 F_1 \cdots + x_n F_n \succ 0\}$

$$\frac{\partial f}{\partial x_i} = \mathrm{tr}(F_i^T \nabla \log\det(F)) = \mathrm{tr}(F_i^T F^{-T}) = \mathrm{tr}(F^{-1} \cdot F_i)$$

where $F = F_0 + x_1 F_1 + \cdots + x_n F_n$

Second Derivative

$f: \mathbb{R}^n \to \mathbb{R}$

The second derivative / Hessian matrix at $x \in \operatorname{int} \operatorname{dom} f$

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

Second-order approximation

$$\hat{f}(z) \approx f(x) + \nabla f(x)^T (z-x) + \frac{1}{2}(z-x)^T \nabla^2 f(x)(z-x)$$

$$\lim_{z \in \operatorname{dom} f,\, z \to x} \frac{|\hat{f}(z) - f(z)|}{\|z-x\|_2} = 0$$

$$D \nabla f(x) = \nabla^2 f(x)$$

Eg. $f: S^n \to \mathbb{R}$ $\quad f(X) = \log\det(X)$ $\quad \operatorname{dom} f = S^n_{++}$

$\nabla f(X) = X^{-1}$ $\quad$ for $Z \in S^n_{++}$ near X $\quad \Delta X = Z - X$

$$\begin{aligned}
Z^{-1} &= (X + \Delta X)^{-1} \\
&= \left(X^{1/2}(I + X^{-1/2}\Delta X X^{-1/2}) X^{1/2}\right)^{-1} \\
&= X^{-1/2} \cdot (I + X^{-1/2}\Delta X X^{-1/2})^{-1} \cdot X^{-1/2} \\
&\approx X^{-1/2} \cdot (I - X^{-1/2}\Delta X X^{-1/2}) \cdot X^{-1/2} \\
&= X^{-1} - X^{-1}\Delta X X^{-1}
\end{aligned}$$

$(I + A)^{-1} \approx I - A$ first-order approximation

// TODO :

Chain rule

composition with scalar function

$f: \mathbb{R}^n \to \mathbb{R}$

$g: \mathbb{R} \to \mathbb{R}$ $\qquad h(x) = g(f(x))$

$$\nabla h(x) = g'(f(x)) \cdot \nabla f(x) \qquad [g'] \cdot \begin{bmatrix} \nabla f(x) \end{bmatrix}$$

$$\nabla^2 h(x) = g'(f(x)) \cdot \nabla^2 f(x) + g''(f(x)) \cdot \nabla f(x) \cdot \nabla f(x)^T$$

$$\left[\frac{dg}{df}\right] \cdot \begin{bmatrix} \frac{d^2 f}{dx_1 dx_1} & \cdots & \frac{d^2 f}{dx_1 dx_n} \\ \vdots & & \vdots \\ \frac{d^2 f}{dx_n dx_1} & \cdots & \frac{d^2 f}{dx_n dx_n} \end{bmatrix} + [g''(f(x))] \cdot [\nabla f] \, [\nabla f]$$

$$dx^T \cdot \nabla^2 h(x) \cdot dx = \left[\frac{dg}{df}\right] \cdot [dx] \begin{bmatrix} \frac{\nabla^2 f}{dx_1 dx_1} & \cdots & \frac{\nabla^2 f}{dx_1 dx_n} \\ \vdots & & \vdots \\ \frac{\nabla^2 f}{dx_n dx_1} & \cdots & \frac{\nabla^2 f}{dx_n dx_n} \end{bmatrix} [dx] + [g''(f(x))] \cdot [dx] [\nabla f] \cdot [\nabla f] [dx]$$

$$= \frac{dg}{df} \cdot df + g''(f(x)) \cdot df \cdot df$$

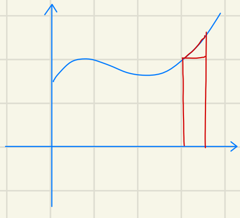
composition with affine function

$f: \mathbb{R}^m \to \mathbb{R}$

$g: \mathbb{R}^n \to \mathbb{R}$ $\qquad g(x) = f(Ax+b) \qquad A \in \mathbb{R}^{m \times n} \qquad b \in \mathbb{R}^m \qquad h(x) = Ax+b$

$$\nabla g(x) = A^T \cdot \nabla f(Ax+b)$$

$$[A][x] \qquad \nabla g(x) = [A^T][\nabla f] \quad \nabla g(x)$$



$$\nabla^2 g(x) = A^T \nabla^2 f(Ax+b) A$$

$$\begin{array}{c} \\ m \\ \end{array} \begin{bmatrix} & A & \end{bmatrix}^{n} \begin{bmatrix} dx \end{bmatrix}^{n} = \begin{bmatrix} dh_1 \\ \vdots \\ dh_m \end{bmatrix} m$$

$$m \begin{bmatrix} \frac{d^2 f}{dh_1 dh_1} & \cdots & \frac{d^2 f}{dh_1 dh_m} \\ \vdots & & \vdots \\ \frac{d^2 f}{dh_m dh_1} & \cdots & \frac{d^2 f}{dh_m dh_m} \end{bmatrix} \begin{bmatrix} d\frac{df}{dh_1} \\ \vdots \\ d\frac{df}{dh_m} \end{bmatrix} m$$

$$n \begin{bmatrix} & A^T & \end{bmatrix} \begin{bmatrix} \\ \end{bmatrix} n$$

$$[dx]$$

$$dx^T [A \; \nabla^2 f(Ax+b) \, A^T] \, dx$$

$$= (A^T dx)^T \; \nabla^2 f(Ax+b) \cdot (A^T dx)$$

$$= dz^T \cdot \nabla^2 f(Ax+b) \cdot dz$$