Matrix Cookbook, wiki: matrix calculus

(1) $f: \mathbb{R}^n \to \mathbb{R}$ $\quad x \in \mathbb{R}^n$ $\quad x_i$ – column vector.

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

a) $f = f(x_1, \ldots, x_n)$

b) k-th : $\nabla f_k = \dfrac{\partial f}{\partial x_k}$

c) $\nabla f \in \mathbb{R}^n$

(2) $f(x) = c^T x$

$f(x) = \sum_{i=1}^{n} c_i x_i$

k-th : $\nabla f_k = \dfrac{\partial f}{\partial x_k} = c_k \Rightarrow \nabla f = c$

(3) $f(x) = \|x\|_2^2 = \sum_{i=1}^{n} x_i^2$

k-th: $\nabla f_k = \dfrac{\partial f}{\partial x_k} = 2x_k \Rightarrow \nabla f = 2x$

(4) $f(x) = \|x\|_B^2 = x^T B x = \sum_{i,j} b_{ij} x_i x_j$

$\nabla f(x)$ – ? $\quad B = B^T$ and $B \succ 0$: $B \in S^n_{++}$

k-th : $\nabla f_k = \dfrac{\partial f}{\partial x_k} = \sum_i b_{ik} x_i + \sum_j b_{kj} x_j =$

$= 2 \sum_j b_{kj} x_j \longrightarrow \nabla f = 2Bx \; (B = B^T)$

(5) $f(x) = \underbrace{\frac{1}{2} x^T A x - b^T x}$, $A \in S^n_{++}$ $\to \min_x$

$\nabla f(x) = Ax - b$

$x^* = \arg\min f(x) \to \nabla f(x^*) = 0$

$Ax^* = b$

(6) $f(x) = \underbrace{\|Ax - b\|_2^2} = x^T \underbrace{A^T A} x + b^T b -$

$- 2 b^T A x$

$A^T A = (A^T A)^T$ $\quad \nabla f(x) = 2A^T A x - 2A^T b$

$A^T A \succcurlyeq 0$ $\quad \underbrace{A^T A x = A^T b}$

if $A \in \mathbb{R}^{n \times n}$ and nonsingular $\Rightarrow Ax = b$

$f(x) = \|Ax - b\|_2^2 + \alpha \|x\|_2^2$, $\alpha > 0$

$\nabla f = 2A^T A x - 2A^T b + 2\alpha x =$

$= 2(A^T A + \alpha I)x - 2A^T b$

(7) Chain rule:

$f(x) = \sin(c^T x)$ $\quad \nabla f = \cos(c^T x) \cdot c$

(8) $f(w) = \sum_i \log(1 + e^{-y_i w^T x_i})$

$x_i \in \mathbb{R}^n$ $i = 1, \dots, m$

$y_i \in \mathbb{R}$ $i = 1, \dots, m$

$$\nabla f(w) = \sum_i \frac{1 \cdot e^{-y_i w^T x_i}}{1 + e^{-y_i w^T x_i}} (-y_i x_i)$$

Matrix case $\quad A \in \mathbb{R}^{n \times n}$

$f : \mathbb{R}^{n \times n} \to \mathbb{R} \quad tr(A) = \sum_{i=1}^{n} a_{ii}$

① $tr(X) = \sum_i x_{ii}$

$$\nabla f_{jk} = \frac{\partial f}{\partial x_{jk}} = \begin{cases} 1, & j = k \\ 0, & \text{otherwise} \end{cases} \Rightarrow \nabla f = I_n$$

② $tr(AX) = \sum_{i,j} a_{ij} x_{ji}$

$$\frac{\partial(\ldots)}{\partial x_{pq}} = a_{qp} = a^T_{pq} \Rightarrow \nabla f = A^T$$

③ $tr(AXBX) = \sum_{i,j,k,l} a_{ij} x_{jk} b_{kl} x_{li}$

$$\nabla_4 f_{pq} = \underbrace{\sum a_{ip} b_{ql} x_{li}}_{j=p \ k=q} + \underbrace{\sum a_{qj} x_{jk} b_{kp}}_{l=p \ i=q} =$$

$$= \sum a^T_{pi} x^T_{il} b^T_{lq} + \sum b^T_{pk} x^T_{kj} a^T_{jq} \to$$

$$\to \nabla f = A^T X^T B^T + B^T X^T A^T$$

④ $f(A) = x^T A x = \sum_{i,j} a_{ij} x_i x_j$

$$\frac{\partial f}{\partial a_{pq}} = x_p x_q \longrightarrow \nabla f = x x^T \in \mathbb{R}^{n\times n}$$ – rank-one matrix

$x^T x$ – scalar

⑤ $f(X) = \det(X)$

$$\nabla f = (X^{-1})^T \det(X)$$

$X = [x_{ij}]$

pq-th elem x_{pq} $\det(X) = \sum_j (-1)^{p+j} x_{pj} M_{pj}$

$$\frac{\partial(\det(X))}{\partial x_{pq}} = \underbrace{(-1)^{p+q} M_{pq}} = C_{pq}$$ – element of the cofactor matrix

Adjugate matrix $A = \mathrm{adj}(X) = C^T$

$$X \cdot \mathrm{adj}(X) = \det(X) I. \Rightarrow \frac{\partial(\det(X))}{\partial X} = (\mathrm{adj}(X))^T = $$
$$= (X^{-1})^T \det(X)$$

Applications:

1) maximum likelihood estimation of covariance matrix in multivariat Gaussian distribution
2) $\log \det(X)$ – barrier function in the interior point methods for solving constrained convex optimization problem. More details see in book „Convex optimization" by S. Boyd & L. Vandenberghe.