Line search. Matrix calculus.

Seminar

Optimization for ML. Faculty of Computer Science. HSE University

Theory recap. Differential

- Differential $df(x)[\cdot] : U \to V$ in point $x \in U$ for $f(\cdot) : U \to V$:

$$f(x+h) - f(x) = \underbrace{df(x)[h]}_{\text{differential}} + \bar{o}(||h||)$$

$U \to V$	$\mathbb{R}$	$\mathbb{R}^n$	$\mathbb{R}^{n \times m}$
$\mathbb{R}$	$f'(x)dx$	$\nabla f(x)dx$	$\nabla f(x)dx$
$\mathbb{R}^n$	$\nabla f(x)^T dx$	$J(x)dx$	—
$\mathbb{R}^{n \times m}$	$tr(\nabla f(X)^T dX)$	—	—

Theory recap. Differential

- Differential $df(x)[\cdot] : U \to V$ in point $x \in U$ for $f(\cdot) : U \to V$:

$$f(x+h) - f(x) = \underbrace{df(x)[h]}_{\text{differential}} + \bar{o}(||h||)$$

- Canonical form of the differential:

$U \to V$	$\mathbb{R}$	$\mathbb{R}^n$	$\mathbb{R}^{n \times m}$
$\mathbb{R}$	$f'(x)dx$	$\nabla f(x) dx$	$\nabla f(x) dx$
$\mathbb{R}^n$	$\nabla f(x)^T dx$	$J(x) dx$	—
$\mathbb{R}^{n \times m}$	$tr(\nabla f(X)^T dX)$	—	—

Theory recap. Differentiation Rules

- Useful differentiation rules and standard derivatives:

Differentiation Rules	Standard Derivatives
$dA = 0$	$d(\langle A, X\rangle) = \langle A, dX\rangle$
$d(\alpha X) = \alpha(dX)$	$d(\langle Ax, x\rangle) = \langle (A + A^T)x, dx\rangle$
$d(AXB) = A(dX)B$	$d(Det(X)) = Det(X)\langle X^{-T}, dX\rangle$
$d(X + Y) = dX + dY$	$d(X^{-1}) = -X^{-1}(dX)X^{-1}$
$d(X^T) = (dX)^T$	
$d(XY) = (dX)Y + X(dY)$	
$d(\langle X, Y\rangle) = \langle dX, Y\rangle + \langle X, dY\rangle$	
$d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$	

Matrix Calculus. Problem 1

Example

Find $\nabla f(x)$, if $f(x) = \frac{1}{2}x^T Ax + b^T x + c$.

Matrix Calculus. Problem 2

Example

Find $\nabla f(X)$, if $f(X) = tr(AX^{-1}B)$

- $h(x) = f(g(x)) \Rightarrow dh(x_0)[dx] = df(g(x_0))[dg(x_0)[dx]]$

Matrix Calculus. Problem 3

Example

Find the gradient $\nabla f(x)$ and hessian $\nabla^2 f(x)$, if $f(x) = \frac{1}{3}\|x\|_2^3$

- $d^2 f(x)[h_1, h_2] = d\left(df(x)[\underbrace{h_1}_{\text{fixed when take outer } d(\cdot)}]\right)[h_2]$

Matrix Calculus. Problem 3

Example

Find the gradient $\nabla f(x)$ and hessian $\nabla^2 f(x)$, if $f(x) = \frac{1}{3}\|x\|_2^3$

- $d^2 f(x)[h_1, h_2] = d\left(df(x)[\underbrace{h_1}_{\text{fixed when take outer } d(\cdot)}]\right)[h_2]$
- Canonic form for $f : \mathbb{R}^n \to \mathbb{R}$: $d^2 f(x)[h_1, h_2] = h_1^T \underbrace{\nabla^2 f(x)}_{\text{hessian}} h_2$