11.09.17 момо II

$$f(x) \to \min_x, \quad x \in \mathbb{R}^n, \quad f \in C^2$$

Line search

$x_{k+1} = x_k + \alpha_k d_k$; $x_k, d_k \in \mathbb{R}^n$, $\alpha_k \in \mathbb{R}_+$, $d_k : \nabla f(x_k)^T d_k < 0$,

$\alpha_k = \operatorname{argmin}_{\alpha \geq 0} f(x_k + \alpha d_k) \leftarrow$ неточная 1D оптимизация

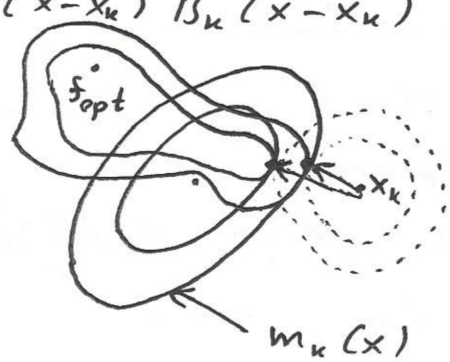
$\{ d_k = -\nabla f(x_k) \}$

Trust Region

$$f(x) \approx m_k(x) = f(x_k) + g_k^T (x - x_k) + \frac{1}{2} (x - x_k)^T B_k (x - x_k)$$

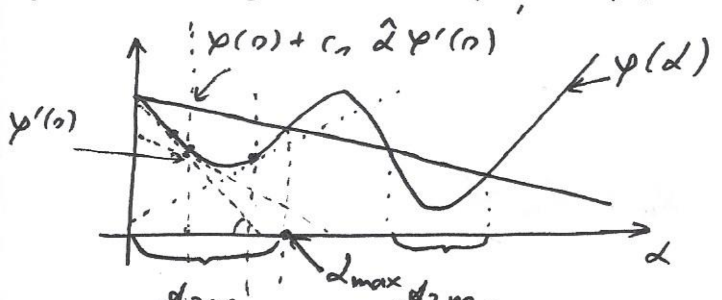
$\{ g_k = \nabla f(x_k), \quad B_k = \nabla f(x_k) \}$

$$\begin{cases} m_k(x) \to \min_x \\ \|x - x_k\|_2^2 \leq \Delta_k \end{cases}$$



Неточная 1D оптимизация

$\varphi(\alpha) = f(x + \alpha d)$, $\alpha \geq 0$, $\varphi'(0) < 0$



условие Armijo:

$$\varphi(\hat{\alpha}) \leq \varphi(0) + c_1 \hat{\alpha} \varphi'(0), \quad c_1 \in (0, 1)$$

Wolfe:
1) $\varphi'(\hat{\alpha}) \geq c_2 \varphi'(0)$, $c_2 \in (0, 1)$ - слабое
2) $|\varphi'(\hat{\alpha})| \leq c_2 |\varphi'(0)|$ - сильное

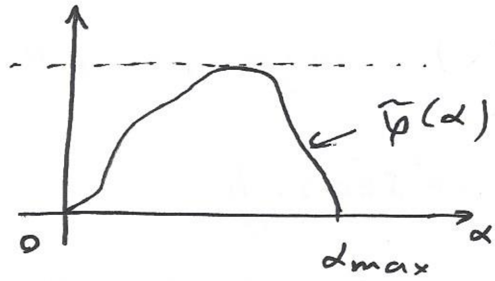
Утв. $\varphi \in C^1$, $\varphi(\alpha) > -\infty$, $\varphi'(0) < 0$, $\alpha > 0$,

$\forall c_1, c_2 : 0 < c_1 \leq c_2 < 1 \Rightarrow \exists \alpha_k$: удовл. Armijo и сильн./слаб. Wolfe

$\square$ $\exists\, \alpha_{max}: \varphi(\alpha_{max}) = \varphi(0) + c_1 \alpha_{max} \varphi'(0)$

$\tilde\varphi(\alpha) = \varphi(0) + c_1 \alpha \varphi'(0) - \varphi(\alpha)$

$\tilde\varphi(0) = \tilde\varphi(\alpha_{max}) = 0$



$\exists\, \alpha_*: \tilde\varphi'(\alpha_*) = 0 \iff$

$\iff c_1 \varphi'(0) - \varphi'(\alpha_*) = 0$, $\varphi'(\alpha_*) = c_1 \varphi'(0) \geq c_2 \varphi'(0)$,

$|\varphi'(\alpha_*)| = c_1 |\varphi'(0)| \leq c_2 |\varphi'(0)|$ $\blacksquare$

На практике $c_1 = 10^{-4}$, $c_2 = 0.1$ или $c_2 = 0.9$

Back tracking

$\alpha = \alpha_{start}$

повторять

  если $\varphi(\alpha) \leq \varphi(0) + c_1 \alpha \varphi'(0)$, то выход

  $\alpha := \alpha \eta$, $\eta < 1$

scipy.optimize.line_search

Градиентный спуск

$f(x) \to \min_{x \in \mathbb{R}^n}$, $f \in C^1$, $x_{k+1} = x_k + \alpha_k d_k$

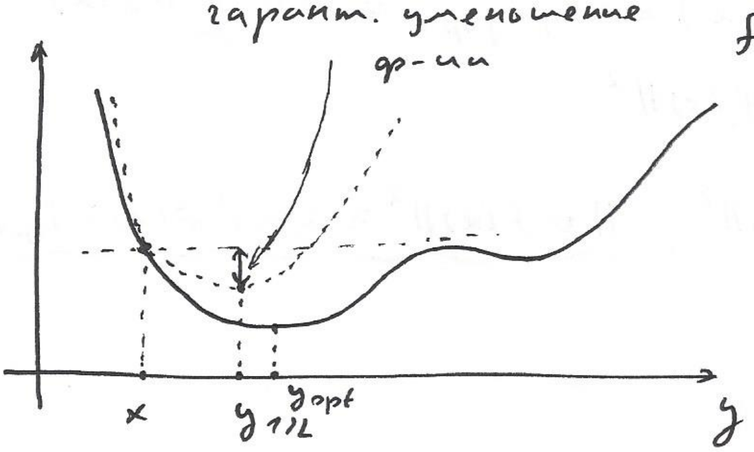
$$\begin{cases} \nabla f(x)^T d \to \min\limits_d \\ \|d\| \leq 1 \end{cases} \Rightarrow d = -\frac{\nabla f(x)}{\|\nabla f(x)\|}$$

GD: $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$

Выбор α_k: 1) $f \in C_L^{1,1}$, $\alpha_k = \frac{1}{L}$

2) Armijo / Wolfe

<u>$f \in C_L^{1,1}$</u>, $y = x - \alpha \nabla f(x)$, $f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|_2^2$



$\underline{f \in C_L^{1,1}}$, $y = x - \alpha \nabla f(x)$,

$f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\|_2^2 = \{ y = x - \alpha \nabla f(x) \} =$

$= f(x) - \alpha \|\nabla f(x)\|_2^2 + \frac{L\alpha^2}{2}\|\nabla f(x)\|^2 = f(x) - \alpha\left(1 - \frac{L\alpha}{2}\right)\|\nabla f(x)\|^2$

$\alpha\left(1 - \frac{L\alpha}{2}\right) \to \max\limits_{\alpha}$, $1 - \frac{L\alpha}{2} + \alpha\left(-\frac{L}{2}\right) = 1 - L\alpha = 0$, $\alpha_{opt} = \frac{1}{L}$

$\overset{\alpha_{opt}}{=} f(x) - \frac{1}{2L}\|\nabla f(x)\|^2$, $\underline{f(y) \le f(x) - \frac{1}{2L}\|\nabla f(x)\|^2}$

$\alpha_k = \frac{1}{L}$: $f(x_{k+1}) \le f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 \le$

$\le f(x_{k-1}) - \frac{1}{2L}\|\nabla f(x_{k-1})\|^2 - \frac{1}{2L}\|\nabla f(x_k)\|^2 \le \ldots \le f(x_0) - \frac{1}{2L}\sum\limits_{i=0}^{k}\|\nabla f(x_i)\|^2$

$g_k = \min\limits_{0 \le i \le k} \|\nabla f(x_i)\|_2^2$: $(k+1) g_k \le \sum\limits_{i=0}^{k}\|\nabla f(x_i)\|^2 \le$

$\le 2L(f(x_0) - f(x_{k+1})) \le 2L(f(x_0) - f_{opt})$

$$g_k \le \frac{2L(f(x_0) - f_{opt})}{k+1}$$

$\underline{f \in C_L^{1,1} \text{ и } \mu \text{ сильно выпукла}}$

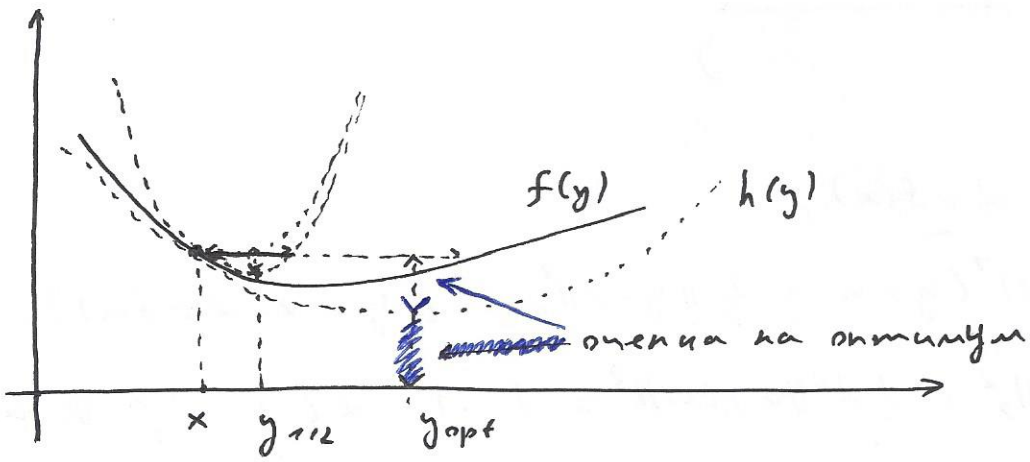
$f(y) \ge f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|y-x\|^2 = h(y)$

$\min\limits_y f(y) \ge \min\limits_y h(y)$

$$\nabla h(y) = \nabla f(x) + \mu(y-x) = 0, \quad y_{opt} = x - \frac{1}{\mu}\nabla f(x)$$

$$h(y_{opt}) = f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2$$

$$f_{opt} \geq f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2, \quad \|\nabla f(x)\|^2 \geq 2\mu(f(x) - f_{opt})$$



$$f(x_{k+1}) - f_{opt} \leq f(x_k) - f_{opt} - \frac{1}{2L}\|\nabla f(x_k)\|^2 \leq$$

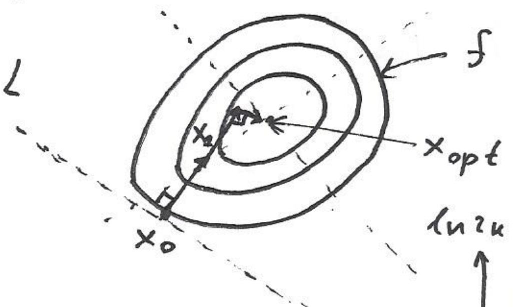
$$\leq f(x_k) - f_{opt} - \frac{2\mu}{2L}(f(x_k) - f_{opt}) =$$

$$= \left(1 - \frac{\mu}{L}\right)(f(x_k) - f_{opt}), \quad c = 1 - \frac{\mu}{L}$$

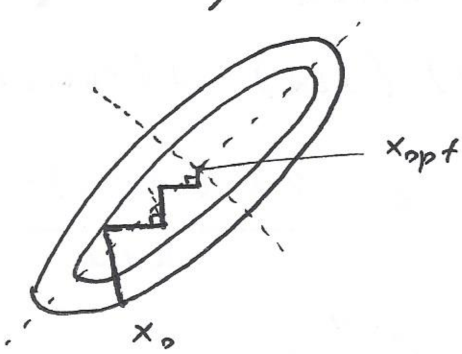
Пример $\quad f(x) = \frac{1}{2}x^T A x - x^T b \to \min\limits_x, \quad \nabla f(x) = Ax - b, \nabla^2 f(x) = A$

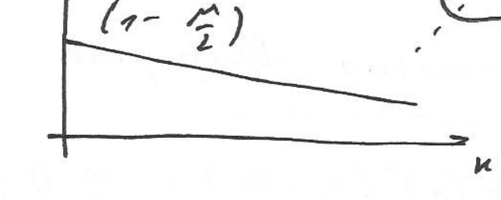
$f \in C_L^{2,1}$, $L = \lambda_{max}(A)$ и μ силно вып.: $\mu = \lambda_{min}(A)$

$\mu \sim L$

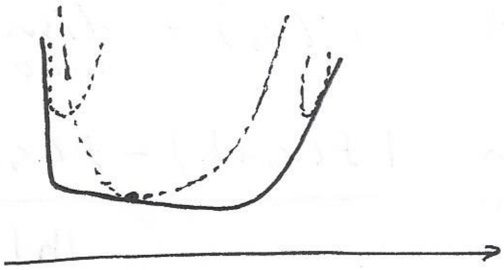


$\mu \ll L$





Подбор L



L_k

повторять

$$y = x - \frac{1}{L_k} \nabla f(x)$$

если $f(y) \le f(x) - \frac{1}{2L_k} \|\nabla f(x)\|^2$, то выход

$L_k \leftarrow L_k \cdot \beta$, $\beta > 1$

$L_k \leftarrow L_k \cdot \eta$, $\eta < 1$

семинар

Матрично-векторное дифференцирование

V, W — векторные пр-ва

$L : V \to W$

О лин. пр-е: $L(x+y) = Lx + Ly$, $L(\alpha x) = \alpha L x$

① $f : V \to W$: $f(x) = w = const$, $w \in W$

$w = f(x+x) = 2w$, лин. $\Leftrightarrow w = 0$

② $f(x) = \langle c, x \rangle$, $c \in V$

лин.

③ $f(x) = \langle c, x \rangle + \alpha$ аффинное преобразование не линейное!

$\alpha \ne 0$

④ $f : \mathbb{R}^{n\times n} \to \mathbb{R}^{n \times n}$: $f(x) = AXB$; $A, B \in \mathbb{R}^{n\times n}$ лин. пр-е

$$\|L\|_{op} = \sup_{\|x\|_v = 1} \|Lx\|_w$$

О ограниченность оператора: $\exists\, C > 0 : \|Lx\| \le C\|x\| \;\; \forall x \in V$

О дифференцируемость

V, W — норм. вект. пр-ва

$f : X \to W$, X — мн-во в V, $x_0 \in X$

$$\begin{cases} f: \mathbb{R} \to \mathbb{R} \qquad f'(x_0) = \lim\limits_{h\to 0} \dfrac{f(x_0+h)-f(x_0)}{h} \\ \Longleftrightarrow \lim\limits_{h \to 0} \dfrac{|f(x_0+h)-f(x_0)-f'(x_0)\cdot h|}{|h|} = 0 \end{cases}$$

$L_0 \in \mathcal{L}(V \to W)$ – мн-во линейных ограниченных пр-й

наз-ся произв-й ф-ии f в т. $x_0 \in X$ - пред. т. X

$$\lim_{h\to 0} \frac{\| f(x_0+h) - f(x_0) - L_0 h \|_W}{\|h\|_V} = 0 \iff$$

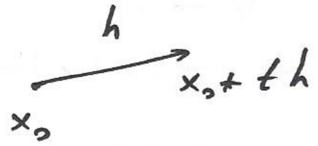
$$\iff f(x_0+h) = f(x_0) + L_0 h + \bar{o}(\|h\|_V)$$

~~Примеры~~

~~$E \subseteq \mathbb{R} : f: E \to \mathbb{R}$~~

<u>Утв</u> Если f диф-ма в т. x_0 с пр. L_0:

$$L_0 h = \lim_{t\to 0, t>0} \frac{f(x_0+th)-f(x_0)}{t} = \quad \forall h$$



$$= \frac{\partial}{\partial t} f(x_0+th)\Big|_{t=0}$$

$\square$ $$\frac{f(x_0+th)-f(x_0)}{t} = \frac{f(x_0) + L h t + \bar{o}(\cancel{\|th\|}) - f(x_0)}{t} = L_0 h + \bar{o}(1)$$ $\blacksquare$

~~Пр-~~ Зм Если пр-я $\exists$, то она $!$

$L_0 \equiv \overset{\text{лин. пр-е}}{Df(x_0)}$, $L_0 h = Df(x_0)[h]$ – вектор из W

пр-я в т. x_0 диф-л в т. x_0 по h

$$Df : X \to \mathcal{L}(V \to W)$$

Примеры ① $E \subseteq \mathbb{R}$, $f: E \to \mathbb{R}$

$$Df(x_0) h = f'(x_0)\cdot h \quad \forall h \in \mathbb{R}$$

$f(x_0+h) = f(x_0) + f'(x_0)\cdot h + \bar{o}(h)$

② пр-я константы

$f: X \to W, \quad f(x) = w \in W \Rightarrow Df(x)[h] = 0 \quad \forall x, h$

$f(x_0+h) - f(x_0) = w - w = 0$ лин. преобр-е.

③ $f: V \to W, \quad f(x) = \langle c, x \rangle$
$\{f: V \to \mathbb{R}\}$

$f(x_0+h) - f(x_0) = \langle c, x_0+h \rangle - \langle c, x_0 \rangle = \langle c, h \rangle$

$\mathcal{D} f(x)[h] = \langle c, h \rangle$

Сл ① $f: \mathbb{R}^n \to \mathbb{R}, \quad f(x) = \langle c, x \rangle$
$\mathcal{D} f(x)[h] = \langle \underset{\in \mathbb{R}^n}{c}, h \rangle$

② $f: \mathbb{R}^{n \times n} \to \mathbb{R}, \quad f(x) = \operatorname{tr} X = \langle I, X \rangle$
$\mathcal{D} f(x)[H] = \langle I, H \rangle \doteq \operatorname{tr} H$

④ $f: V \to \mathbb{R}$

$f: V \to \mathbb{R}, \quad A \in \mathcal{L}(V \to W)$

$f(x_0+h) - f(x_0) = \underline{\langle Ah, x_0 \rangle + \langle Ax_0, h \rangle} + \langle Ah, h \rangle$

$|\langle Ah, h \rangle| \le \|Ah\| \|h\| \le \|A\| \|h\|^2 = \bar{o}(\|h\|)$

$\mathcal{D} f(x)[h] = \langle Ah, x_0 \rangle + \langle Ax_0, h \rangle = \langle (A + A^*)x_0, h \rangle$

Сл ~~$f: V \to \mathbb{R}, \quad A \in \mathcal{L}(V \to W)$~~ $f(x) = \langle Ax, x \rangle$

① $f \in \mathbb{R}^n \to \mathbb{R}, \quad A \in \mathbb{R}^{n \times n}, \quad \mathcal{D} f(x_0)[h] = \langle (A + A^T)x_0, h \rangle =$
$= 2 \langle Ax_0, h \rangle \quad \{A = A^T\}$

② $f: \mathbb{R}^{m \times n} \to \mathbb{R}$

$f(x) = \langle X, X \rangle = \|X\|_F^2, \quad \mathcal{D} f(X)[H] = 2 \langle X, H \rangle$

$\{ \langle IX, X \rangle \}$

⑤ $f(X) = X^{-1}, \quad S = \{ X \in \mathbb{R}^{n \times n} : \det(X) \neq 0 \}$

$f: S \to S, \quad f(X_0 + H) - f(X_0) = (X_0 + H)^{-1} - X_0^{-1} =$

$= (X_0(I + X_0^{-1}H))^{-1} - X_0^{-1} = ((I + X_0^{-1}H)^{-1} - I) X_0^{-1}$

$(I + X_0^{-1}H)^{-1} - I \ominus \qquad \|A^k\| \leq \|A\|^k$

// ряд Неймана: $(I_n - A)^{-1} = \sum_{k=0}^{\infty} A_k \quad , \|A\|_{op} < 1$

$\ominus \; I - \underbrace{X_0^{-1}H}_{\text{лин. по } H} + \sum_{k=2}^{\infty} (-1)^k (X_0^{-1}H)^k - I$

$\sum_{k=2}^{\infty} \|X_0^{-1}\|^k \|H\|^k = \frac{1}{1 - \|X_0^{-1}\|\|H\|} - 1 - \|X_0^{-1}\|\|H\| =$

$= \frac{\|X_0^{-1}\|^2 \|H\|^2}{1 - \|X_0^{-1}\| \|H\|} = \bar{o}(\|H\|)$

$f(x_0 + H) - f(x_0) = -X_0^{-1} H X_0^{-1} + \bar{o}(\|H\|)$

$\mathcal{D} f(x_0)[H] = -X_0^{-2} H X_0^{-2}$

Таблица стандартных производных

$\mathcal{D}(\langle A, X\rangle)[H] = \langle A, H\rangle$

$\mathcal{D}(\langle AX, X\rangle)[H] = \langle (A + A^T)X, H\rangle = \{A = A^T\} = 2\langle AX, H\rangle$

$\mathcal{D}(\det X)[H] = \det X \langle \tilde{X}^T, H\rangle = \det X \; tr\, \tilde{X}^T H$

$\mathcal{D}(X^{-1})[H] = -X^{-1} H X^{-1}$

Правила преобразования

$\mathcal{D}(f_1 + f_2)[h] = \mathcal{D} f_1[H] + \mathcal{D} f_2[H] \quad , \mathcal{D}\alpha f(x)[h] = \alpha \mathcal{D} f(x)[h]$

$\mathcal{D}\langle f_1(x), f_2(x)\rangle[H] = \langle f_1(x), \mathcal{D} f_2(x)[H]\rangle + \langle \mathcal{D} f_1(x)[H], f_2(x)\rangle$

правило Лейбница

$\mathcal{D}(A F(x) B)[H] = A(\mathcal{D} f(x)[H]) B$

$\mathcal{D}(F(x) G(x))[H] = F(x) \mathcal{D} G(x)[H] + \mathcal{D} F(x)[H] G(x)$

Правило композиции

$f: X \to Y \quad , g: Y \to U$

$k(x) = g(f(x)) \quad , k = g \circ f$

$\mathcal{D}(g \circ f)(x)[h] = \mathcal{D} g(f(x))[\mathcal{D} f^{*}(x)[h]]$

// $g'(f(x)) = g'(f(x)) f'(x)$

① $f: \mathbb{R}^n \to \mathbb{R}$

$f(x) = |x|^3 = \langle x, x \rangle^{\frac{3}{2}}$

$\mathcal{D} f(x)[h] = \frac{3}{2} \langle x, x \rangle^{\frac{1}{2}} \mathcal{D}(\langle x, x \rangle)[h] = \frac{3}{2} \|x\| \cdot 2 \langle x, h \rangle =$

$= 3 \|x\| \langle x, h \rangle$

② $f(x) = \ln \det X \quad , f: S^n_{++} \to \mathbb{R}$

$\mathcal{D} \ln \det X [H] = \frac{1}{\det X} [\mathcal{D} \det X [H]] =$

$= \frac{1}{\det X} \det X \langle X^{-1}, H \rangle = \langle X^{-1}, H \rangle$

// $\mathcal{D} \det X [H] = \det X \langle X^{-T}, H \rangle$

③ $f: \mathbb{R}^{n \times n} \to \mathbb{R}$

$f(x) = \operatorname{tr}(AXAX^{-1}) = \langle I, AXAX^{-1} \rangle$

$\mathcal{D} f(X)[H] = \langle I, \mathcal{D} AXAX^{-1}[H] \rangle = \langle I, A \mathcal{D} X[H] A X^{-1} \rangle +$

$+ \langle I, AXA \mathcal{D} X^{-1}[H] \rangle = \langle I, AHAX^{-1} \rangle + \langle I, AXAX^{-1} H X^{-1} \rangle =$

$= \langle A^T X^{-T} A^T, H \rangle - \langle X^{-T} A^T X^T A^T X^{-T}, H \rangle = \langle a(x), H \rangle$

Градиент

$f: V \to \mathbb{R} \quad , \mathcal{D} f(x)[h] = \langle \alpha(x), h \rangle$ {т. Рисса}

$f: \mathbb{R}^n \to \mathbb{R} \quad , L \in \mathcal{L}(\mathbb{R}^n \to \mathbb{R})$

$x = \sum_i \alpha_i e_i \quad , L x = \sum_i \alpha_i L e_i$ ~~...~~ $= \langle \alpha, Le \rangle$

~~$x = \langle \alpha, e \rangle$~~ , ~~$Lx = \langle \alpha, Le \rangle$~~ $Le_i \in \mathbb{R}$, $Le \in \mathbb{R}^n$

	скаляр $\mathbb{R}$	вектор $\mathbb{R}^n$	матрица $\mathbb{R}^{n\times n}$
скаляр $\mathbb{R}$	$Df(x)[h] = f'(x)\cdot h$	—	—
вектор $\mathbb{R}^n$	$Df(x)[h] = \langle \nabla f(x), h\rangle$	$Df(x)[h] = J_x h$	$\times$
матрица $\mathbb{R}^{n\times n}$	$Df(x)[H] = \langle \underbrace{\nabla f(x)}_{\text{матрица}}, H\rangle$	$\times$	$\times$

$f(x) = \|x\|^3$, $\quad Df(x)[h] = 3\|x\|^2 \langle x, h\rangle$

$\nabla f(x) = 3\|x\|^2 x \in \mathbb{R}^n$

$f(x) = \ln\det X$, $\quad Df(x)[H] = \langle X^{-1}, H\rangle$

$\nabla f(x) = X^{-1} \in \mathbb{R}^{n\times n}$

$f(x) = Ax$, $\quad Df(x)[h] = Ah \Rightarrow J_f = A$

$$\langle \nabla f(x), h\rangle = Df(x)[h] = \frac{\partial}{\partial t} f(x+th)\Big|_{t=0}$$

$$h = e_i \Rightarrow [\nabla f(x)]_i = \langle \nabla f(x), e_i\rangle = \frac{\partial}{\partial t} f(x+te_i)\Big|_{t=0} \cancel{= f'_{x_i}} = \frac{\partial}{\partial x_i} f(x)$$

Вторая производная

$$D^2 f(x)[h_1, h_2] = D(Df(x)[h_1])[h_2]$$