18.09.17 МОМО $\underline{\text{III}}$ ~~IV~~

$f(x) \to \min\limits_x \quad , x \in \mathbb{R}^n$

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

α_k : Armijo/Wolfe

1) $f(x_{k+1}) \le f(x_k) + c_1 \alpha_k \nabla f(x_k)^T d_k$

2) $\nabla f(x_{k+1})^T d_k \ge c_2 \nabla f(x_k)^T d$

сходимость? $f(x_{k+1}) \le f(x_k) - const \|\nabla f(x_k)\|^2$

$$\cos\theta_k = \frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\| \, \|d_k\|}$$

$$(c_2 - 1)\nabla f(x_k)^T d_k \le (\nabla f(x_{k+1}) - \nabla f(x_k))^T d_k \overset{КБ}{\le} \|\nabla f(x_{k+1}) - \nabla f(x_k)\| \, \|d_k\| \le$$

$$\le L \|x_{k+1} - x_k\| \, \|d_k\| = L \alpha_k \|d_k\|^2 \quad \Rightarrow$$

$$\alpha_k \geqslant \frac{(c_2 - 1)\nabla f(x_k)^T d_k}{L \|d_k\|^2} = \frac{(c_2-1)\cos\theta_k \|\nabla f(x_k)\| \cancel{\|d_k\|}}{L \|d_k\|^{\cancel{2}}}$$

① $$f(x_{k+1}) \le f(x_k) + \frac{c_1(c_2-1)\cos^2\theta_k \|\nabla f(x_k)\|^2}{L}$$

$$f(x_{k+1}) \le f(x_k) - \frac{c_1(1-c_2)\cos^2\theta_k}{L} \|\nabla f(x_k)\|^2$$

$[\, d_k = -\nabla f(x_k) \,] \Rightarrow \cos\theta_k = -1 , \ \cos^2\theta_k = 1 \,]$

$$f(x_{k+1}) \le f(x_k) - \frac{c_1(1-c_2)}{L}\|\nabla f(x_k)\|^2$$

линейная сх-ть для сильно вып. сублинейная $\frac{1}{k}$ для остальных

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

СЛАУ $Ax = b$, $A = A^T \ne 0 \Rightarrow x = A^{-1} b$

$A \in \mathbb{R}^{m \times n}$, $m > n \Rightarrow x = (A^T A)^{-1} A^T b$

~~ниж~~треугольная матрица L

① $A = A^T > 0 \Rightarrow A = L L^T$, $L = \left(\begin{smallmatrix} * & 0 \\ * & * \end{smallmatrix} \right)$, $Chol = \frac{2}{3} n^3$

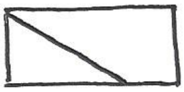
$A x = b$, $L \underbrace{L^T x}_{y} = b \Rightarrow \begin{cases} y = L^{-1} b \leftarrow O(n^2) \\ x = L^{-T} y \end{cases}$

② $A = A^T \neq 0$, $A \in \mathbb{R}^{n\times n}$

$P^T A P = L \mathcal{D} L^T$, $L = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ & & 1 \end{pmatrix}$, $\mathcal{D} = \begin{bmatrix} \square & & & 0 \\ & \boxplus & & \\ & & \boxplus & \\ 0 & & & \ddots \end{bmatrix}$

$P = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$, $P^T = P^{-1}$ метод обратного исключения неизвестных

$Ax = b \iff P L \underbrace{\mathcal{D} \underbrace{L^T P^T x}_{y}}_{z} = b$ $\quad \begin{cases} z = L^{-1} P^T b \\ y = \mathcal{D}^{-1} z \\ x = P L^{-T} y \end{cases}$ $\quad O(n^2)$

③ $A \in \mathbb{R}^{m\times n}$, $m \neq n$ 

$PA = LU$, $L = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ & & 1 \end{pmatrix}$, $U = \begin{pmatrix} \diagdown & \\ 0 & \end{pmatrix}$, $\frac{4}{3} n^3$

алгоритм Гаусса! P определяет ведущие элементы

$L \in \mathbb{R}^{m \times \min(m,n)}$, $U \in \mathbb{R}^{\min(m,n) \times n}$

④ $A \in \mathbb{R}^{m \times n}$, $m \neq n$

$AP = QR$, $Q^T = Q^{-1} \in \mathbb{R}^{m\times m}$, $R = \begin{pmatrix} \diagdown & \\ 0 & \end{pmatrix} \in \mathbb{R}^{m\times n}$

Q содержит ортогональный базис ядра A, $Q = \begin{bmatrix} q_1 & \dots & q_m \end{bmatrix}$

пример: $Ax = b$, $x = (A^T A^{-1}) A^T b \iff$

$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ $\quad R_1 \in \mathbb{R}^{n\times n}$, $Q_1 \in \mathbb{R}^{m\times n}$, $Q_2 \in \mathbb{R}^{m\times(m-n)}$

$\iff (R_1^T \underbrace{Q_1^T Q_1}_{I} R_1)^{-1} R_1^T Q_1^T b = R_1^{-1} R_1^{-T} R_1^T Q_1^T b = R_1^{-1} Q_1^T b$

Метод Ньютона $f(x) \to \min_x$

$f(x_k + d) \approx m_k(d) = f(x_k) + g_k^T d + \frac{1}{2} d^T B_k d \to \min_d$

1) $g_k = \nabla f(x_k)$, $B_k = I \Rightarrow d = -\nabla f(x_k)$

$\nabla_d m_k(d) = g_k + B_k d = 0$, $d = -B_k^{-1} g_k$

$\Downarrow$

метод градиентного спуска! $d = -\nabla f(x_k)$

2) $g_k = \nabla f(x_k)$, $B_k = \nabla^2 f(x_k)$ Newton

тогда оптимальное $d = -B_k^{-1} d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$

необходимо $\nabla^2 f(x_k) > 0$, $\underline{x_{k+1} = x_k - \alpha_k (\nabla^2 f(x_k))^{-1} \nabla f(x_k)}$

$$\nabla f(x_k)^T d_k = -\nabla f(x_k)^T (\nabla^2 f(x_k))^{-1} \nabla f(x_k) < 0 \iff \nabla^2 f(x_k) > 0$$

Глобальная сходимость

$x_{k+1} = x_k + \alpha_k d_k$, $f(x_{k+1}) \le f(x_k) - \dfrac{c_2(1-c_2)\cos^2\theta_k}{L} \|\nabla f(x_k)\|^2$

необходимо $\cos^2\theta_k > 0$; $\cos\theta_k = \dfrac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\| \|d_k\|} =$

$$= \frac{-\nabla f(x_k)^T (\nabla^2 f(x_k))^{-1} \nabla f(x_k)}{\|\nabla f(x_k)\| \|d_k\|} =$$

$$= \frac{-\nabla f(x_k)^T (\nabla^2 f(x_k))^{-1} \nabla f(x_k)}{\|\nabla f(x_k)\| (\nabla f(x_k)^T (\nabla^2 f(x_k))^{-2} \nabla f(x_k))^{\frac{1}{2}}} \le$$

$$\le -\frac{\lambda_{min}((\nabla^2 f(x_k))^{-1}) \|\nabla f(x_k)\|^2}{\|\nabla f(x_k)\|^2 \lambda_{max}((\nabla^2 f(x_k))^{-1})} \le -\frac{\lambda_{min}(\nabla^2 f(x_k))^{-1}}{\lambda_{max}(\nabla^2 f(x_k))^{-1}}$$

метод сх-ся для $\forall$ начального приближения

Локальная сходимость

<u>Утв</u> $f \in C_M^{2,2}$, $\nabla^2 f(x_{opt}) \succeq \mu I$, $\mu > 0$, $\alpha_k = 1$ $x_0 : \|x_0 - x_{opt}\| \le r$

$\Rightarrow$ Newton $\|x_{k+1} - x_{opt}\| \le C \|x_k - x_{opt}\|^2$, квадратичная сх.

$\square$ $x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$

$\|x_{k+1} - x_{opt}\| = \|x_k - x_{opt} - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)\| \le$

$\le \|(\nabla^2 f(x_k))^{-1}\| \|\nabla^2 f(x_k)(x_k - x_{opt}) - (\nabla f(x_k) - \underbrace{\nabla f(x_{opt})}_{0})\| \le$

$\le \|\nabla^2 f(x_k)^{-1}\| \|\nabla^2 f(x_k)(x_k - x_{opt}) - \int_0^1 \nabla^2 f(x_{opt} + \tau(x_k - x_{opt})) \cdot$

$\cdot (x_k - x_{opt}) d\tau\| = \|\nabla^2 f(x_k)^{-1}\| \|\int_0^1 (\nabla^2 f(x_k) - \nabla^2 f(x_{opt} +$

$+ \tau(x_k - x_{opt})))(x_k - x_{opt}) d\tau\| \le \|\nabla^2 f(x_k)^{-1}\| \cdot$

$\cdot \int_0^1 M(1-\tau)\|x_k - x_{opt}\|^2 d\tau = \|\nabla^2 f(x_k)^{-1}\| \frac{M}{2} \|x_k - x_{opt}\|^2$

$\forall\ x_k : \|x_k - x_{opt}\| < r$

$\|x_{k+1} - x_{opt}\| \le r \|\nabla^2 f(x_{opt})^{-1}\| \frac{M}{2} \|x_k - x_{opt}\|^2$

$r_{k+1} \le C r_k^2 \quad , \quad r_0 : \quad C r_0 < 1 \quad !$

$\|\nabla^2 f(y) - \nabla^2 f(x)\| \le M \|y - x\|$

$f(x) = \frac{1}{2} x^T A x - x^T b \quad , \nabla^2 f = A \Rightarrow f \in C_M^{2,2}$ для $M = 0$

Методы модификации Гессиана

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

$B_k = \nabla^2 f(x_k) + E_k \quad : \quad B_k > 0 \quad \forall k$

① Модификация с.зн.

$\nabla^2 f(x_k) = Q^T \Lambda Q \quad , \quad Q^T = Q^{-1} \quad , \quad \Lambda = diag(\lambda_1 \dots \lambda_n)$

$B_k = Q^T \tilde{\Lambda} Q \quad , \quad \tilde{\lambda}_i = \begin{cases} \lambda_i & \text{, если } \lambda_i \ge \delta > 0 \\ \delta & \text{, иначе} \end{cases}$

② Модификация диагонали

$B_k = \nabla^2 f(x_k) + \lambda_k I > 0 \iff \lambda_k > \cancel{\lambda_{min}(\nabla^2 f(x_k))}$ $\max\{0, -\lambda_{min}(\nabla^2 f(x_k))\}$ $\quad L_k L_k^T$

λ_0, повторять

$B_k = \nabla^2 f(x_k) + \lambda_k I$

$B_k = L_k L_k^T$, если успех, то выход

$\lambda_k = \lambda_k \gamma, \quad \gamma > 1$

$\lambda_k = \lambda_k \rho, \quad \rho < 1$

$d_k = -(\nabla^2 f(x_k) + \lambda_k I)^{-1} \nabla f(x_k)$

$\lambda_k \gg 1$ GD

$\lambda_k \ll 1$ Newton

③ Модификация LDL разложения

$D = Q^T \Lambda Q \quad , \quad \tilde{D} = Q^T \tilde{\Lambda} Q \quad , \quad P^T \nabla^2 f(x_k) P = L D L^T$

$B_k = P L \tilde{D} L^T P^T$

$\varkappa = \frac{L}{\mu}$

L — константа липшицевости градиента

μ — константа сильной выпуклости

$L = \max_x \lambda_{max}(\nabla^2 f(x))$

$\mu = \min_x \lambda_{min}(\nabla^2 f(x))$