09.10.17 момо VI

Квазиньютоновские методы QN

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

$f(x_k + d) \approx m_k(d) = f_k + \nabla f_k^T d + \frac{1}{2} d^T B_k d \to \min\limits_{d}$

$\nabla_d m_k(d) = \nabla f_k + B_k d = 0 \Rightarrow d = -B_k^{-1} \nabla f_k$

$x_{k+1} = x_k - \alpha_k B_k^{-1} \nabla f_k$

условие секущей: $\begin{cases} \nabla m_{k+1}(0) = \nabla f_{k+1} \\ \nabla m_{k+1}(-\alpha d_k) = \nabla f_k \end{cases}$

$\nabla m_{k+1}(d) = \nabla f_{k+1} + B_{k+1} d$

$\nabla m_{k+1}(0) = \nabla f_{k+1} \quad \forall B_{k+1}$

$\nabla m_{k+1}(-\alpha_k d_k) = \nabla f_{k+1} + B_{k+1} \underbrace{(-\alpha_k d_k)}_{x_k - x_{k+1}} = \nabla f_k$

$\Rightarrow B_{k+1} \underbrace{(x_{k+1} - x_k)}_{s_k} = \underbrace{\nabla f_{k+1} - \nabla f_k}_{y_k} \Rightarrow B_{k+1} s_k = y_k \mid s_k^T \times$

$s_k^T B_{k+1} s_k = s_k^T y_k \Rightarrow s_k^T y_k > 0$

утв достаточные условия

1. Если f – сильно выпукла, то $s_k^T y_k > 0$
2. $d_k^T \nabla f_k < 0$ и α_k выбирается из условия Вольфа

$\forall c_2 \in (0, 1) \Rightarrow s_k^T y_k > 0$

Идея QN: $B_{k+1} = B_k + \mathrm{low\_rank\_update}(B_k, s_k, y_k)$

Symmetric Rank 1 (SR1)

$B_{k+1} = B_k + \sigma_k v_k v_k^T$

$B_{k+1} s_k = B_k s_k + \sigma_k v_k (v_k^T s_k) = y_k$

$$\sigma_k (v_k^T s_k) v_k = y_k - B_k s_k \Rightarrow v_k = \delta_k (y_k - B_k s_k)$$

$$\sigma_k \delta_k^2 (y_k - B_k s_k)^T s_k (y_k - B_k s_k) = y_k - B_k s_k$$

$$\sigma_k \delta_k^2 (y_k - B_k s_k)^T s_k = 1$$

$$B_{k+1}^{SR1} = B_k + \sigma_k \delta_k^2 (y_k - B_k s_k)(y_k - B_k s_k)^T =$$

$$= B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$

$$d_k = -B_k^{-1} \nabla f_k$$

$C_k = B_k^{-1}$, тождество Вудберри:

$$C_{k+1}^{SR1} = C_k + \frac{(C_k y_k - s_k)(C_k y_k - s_k)^T}{(C_k y_k - s_k)^T y_k}$$

Схема SR1

x_0, ε

$d_0 = -\nabla f(x_0)$

для $k = 0, 1, 2, \ldots$

$\alpha_k = \arg\min_\alpha f(x_k + \alpha d_k)$ ← неточная

$x_{k+1} = x_k + \alpha_k d_k$

$\nabla f_{k+1} = \nabla f(x_{k+1})$, если $\|\nabla f_{k+1}\|^2 < \varepsilon$, то стоп

$s_k = x_{k+1} - x_k$, $y_k = \nabla f_{k+1} - \nabla f_k$

$C_{k+1}^{SR1} = C_k + (\ldots)$

$d_{k+1} = -C_{k+1}^{SR1} \nabla f_{k+1}$

$f(x) = \frac{1}{2} x^T A x - x^T b \to \min_x$, $A > 0$

<u>утв</u> $\forall x_0, C_0 > 0$, $(C_k y_k - s_k)^T y_k \neq 0$

$\Rightarrow$ в SR1 для $f(x)$ верно $C_k y_j = s_j$ $\forall j < k$

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

$$\nabla f_{k+1} - \nabla f_k = A(x_{k+1} - x_k) \Rightarrow y_k = A s_k$$

$\exists$ $[s_0 | \dots | s_{n-1}]$ л.н.з. : $y_k = A s_k \quad \forall k = 0, \dots, n-1$

$$A = [y_0 | \dots | y_{n-1}][s_0 | \dots | s_{n-1}]^{-1}$$

$\exists\ B_k > 0$

$$B_{k+1}^{SR1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$

$$x^T B_{k+1}^{SR1} x = \underset{\vee\atop 0}{x^T B_k x} + \frac{\overset{0}{\wedge}\, x^T (y_k - B_k s_k)(y_k - B_k s_k)^T x}{\underset{\vee\atop 0}{y_k^T s_k} - \underset{\vee}{s_k^T B_k s_k}} \not> 0$$

$$\tilde{B}_{k+1}^{SR1} = \begin{cases} B_k, & \text{если } (y_k - B_k s_k)^T s_k \le \rho \|y_k - B_k s_k\| \|s_k\| \\ B_k^{SR1}, & \text{иначе} \end{cases} \qquad \rho = 10^{-8}$$

{пропуск обновления}

Выбор C_0

$C_0 = \gamma_0 I$, $\hat{x}_1 = x_0 - \alpha_0 \nabla f(x_0)$

$s_0 = \hat{x}_1 - x_0$, $y_0 = \nabla f(\hat{x}_1) - \nabla f(x_0)$

$\|\gamma_0 y_0 - s_0\|^2 \to \min\limits_{\gamma_0}$, $\gamma^2 y_0^T y_0 - 2\gamma_0 y_0^T s_0 + s_0^T s_0 \to \min\limits_{\gamma_0}$

$$\gamma_0 = \frac{y_0^T s_0}{y_0^T y_0}$$

BFGS

Хотим: $B_{k+1} s_k = y_k$, $B_{k+1} > 0$

$B_{k+1} = J_{k+1} J_{k+1}^T$ параметризация

$$B_{k+1} s_k = y_k \Leftrightarrow J_{k+1} \underbrace{J_{k+1}^T s_k}_{u_k} = y_k \Leftrightarrow \begin{cases} J_{k+1}^T s_k = u_k \\ J_{k+1} u_k = y_k \end{cases}$$

$$B_k = J_k J_k^T : \begin{cases} \frac{1}{2}\|J_{k+1} - J_k\|^2 \to \min\limits_{J_k} \\ J_{k+1}^T s_k = u_k \\ J_{k+1} u_k = y_k \end{cases}$$

$$L(J_{k+1}, \lambda) = \frac{1}{2} tr((J_{k+1} - J_k)^T (J_{k+1} - J_k)) +$$
$$+ \lambda^T (J_{k+1}^T s_k - u_k)$$

$$\nabla_{J_{k+1}} L = J_{k+1} - J_k + s_k \lambda^T = 0 \quad , \quad J_{k+1} = J_k - s_k \lambda^T$$

$$J_{k+1}^T s_k = J_k^T s_k - \lambda s_k^T s_k = u_k . \Rightarrow \lambda = \frac{J_k^T s_k - u_k}{s_k^T s_k}$$

$$J_{k+1} = J_k - \frac{s_k (J_k^T s_k - u_k)^T}{s_k^T s_k}$$

$$J_{k+1} u_k = J_k u_k - \frac{s_k (J_k^T s_k - u_k)^T u_k}{s_k^T s_k} = y_k \quad ?$$

$$\begin{cases} \frac{1}{2} \| J_{k+1} - J_k \|_F^2 \to \min\limits_{J_{k+1}} \\ J_{k+1} u_k = y_k \end{cases} \Rightarrow J_{k+1} = J_k - \frac{(J_k u_k - y_k) u_k^T}{u_k^T u_k}$$

$$J_{k+1}^T s_k = J_k^T s_k - \frac{u_k (J_k u_k - y_k)^T s_k}{u_k^T u_k} = u_k$$

$$\left(1 + \frac{(J_k u_k - y_k)^T s_k}{u_k^T u_k} \right) u_k = J_k^T s_k \Rightarrow u_k = \gamma_k J_k^T s_k$$

$$\gamma_k^2 = \frac{u_k^T s_k}{s_k^T B_k s_k}$$

$$B_{k+1}^{BFGS} = J_{k+1} J_{k+1}^T = \left(J_k - \frac{(J_k u_k - y_k) u_k^T}{u_k^T u_k} \right) \left(J_k - \frac{(J_k u_k - y_k) u_k^T}{u_k^T u_k} \right)^T =$$

$$= B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$$

$$C_k = B_k^{-1} , \quad C_{k+1}^{BFGS} = (I - \rho_k (y_k s_k^T))^T C_k (I - \rho_k y_k s_k^T) +$$
$$+ \rho_k s_k s_k^T \quad , \quad \rho_k = 1 / y_k^T s_k > 0$$

L-BFGS

Идея: вычислить $-C_k \nabla f_k$ с помощью $\{y_i, s_i\}_{i=k-m}^{k}$ через BFGS обновления, начиная с $C_k^0 = \dfrac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}} I$

$$q_k = -\nabla f_k, \quad r_k = C_k q_k = (I - \rho_{k-1}(y_{k-1} s_{k-1}^T)^T) C_{k-1} \cdot$$
$$\cdot (I - \rho_{k-1} y_{k-1} s_{k-1}^T) q_k + \rho_{k-1} s_k (s_{k-1}^T q_k) =$$
$$= \underbrace{(I - \rho_{k-1}(y_{k-1} s_{k-1}^T)^T) C_{k-1} (\underbrace{q_k - \rho_{k-1} (s_{k-1}^T q_k) y_{k-1}}_{q_{k-1}})}_{r_{k-1}} +$$
$$+ \rho_{k-1} (s_{k-1}^T q_k) s_{k-1} =$$
$$= r_{k-1} - \rho_{k-1} \overset{s}{\cancel{y}}_{k-1} (\overset{y^T}{\cancel{s}}_{k-1} r_{k-1}) + \rho_{k-1} (s_{k-1}^T \overset{q_k}{\cancel{q_m}}) s_{k-1}$$

хотим вычислить

$$\begin{tikzcd}
r_k & r_{k-1} \arrow[l] & r_{k-2} \arrow[l] & \cdots \arrow[l] & r_{k-m} \arrow[l] \\
q_k \arrow[u, bend right] \arrow[r] & q_{k-1} \arrow[u, bend right] \arrow[r] & q_{k-2} \arrow[u, bend right] \arrow[r] & \cdots \arrow[r] & q_{k-m} \arrow[u]
\end{tikzcd}$$

знаем

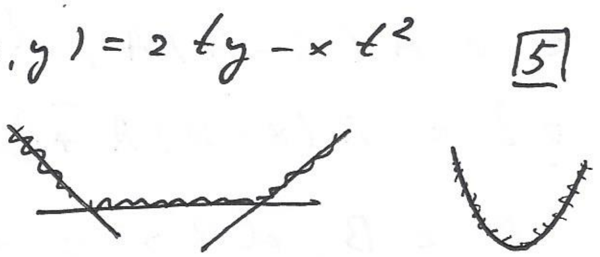
$$r_{k-m} = C_k^0 q_{k-m} = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}} q_{k-m}$$

сложность итерации: $O(mn)$, память: $O(mn)$

семинар

$$f(x,y) = \begin{cases} \dfrac{y^2}{x}, & x > 0 \\ 0, & x = y = 0 \end{cases} \qquad F_t(x,y) = 2ty - xt^2 \qquad \boxed{5}$$

$$\max_{t \in \mathbb{R}} F_t(x,y) = f(x,y)$$



$\forall$ выпуклая ф-ия представима как максимум множества линейных функций

BFGS $\quad x_{k+1}^{*} = x_k - (\nabla^2 F(x_k))^{-1} \nabla F(x_k)$

$\nabla^2 F(x_k) \approx B_k \,, \quad x_{k+1} = x_k - B_k^{-1} \nabla F(x_k) \,, \quad B_k^{-1} = H_k$

$x_{k+1} = x_k - H_k \nabla f(x_k)$

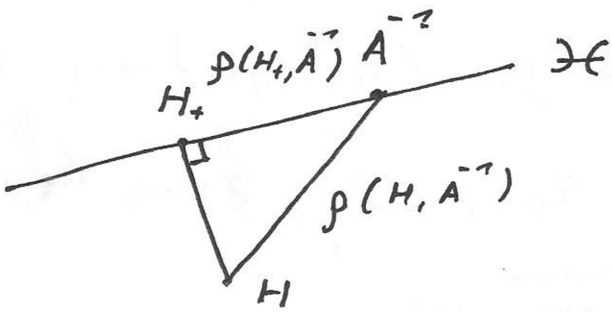
$F(x) = \frac{1}{2} \langle Ax, x \rangle \,, \quad \nabla^2 F(x) = A$

$A \in S^n_{++}$ хотим генерировать $H_k \to A^{-1}, \; k \to \infty$

$H \longrightarrow H_+ \,, \quad s \in \mathbb{R}^n \,, \quad As = y \iff A^{-1} y = s \iff \underline{A^{-1} A s = s}$

$\boxed{H_+ A s = s} \qquad \rho(X, Y) \qquad \mathcal{H} = \{ X \in S^n : XAs = s \}$

$\min\limits_{X \in S^n} \rho(X, H) \;:\; XAs = s$



$\rho(X, Y) = \| A^{\frac{1}{2}} (X - Y) A^{\frac{1}{2}} \|_F$

$\rho(H, A^{-1}) = \| A^{\frac{1}{2}} H A^{\frac{1}{2}} - I_n \|_F$

$F(x) = \langle Ax, x \rangle = \| A^{\frac{1}{2}} x \|^2 \,, \quad x_+ = x - HAx$

$A^{\frac{1}{2}} x_+ = A^{\frac{1}{2}} x - (A^{\frac{1}{2}} H A^{\frac{1}{2}}) A^{\frac{1}{2}} x$

$F(x_+) = \| (I_n - A^{\frac{1}{2}} H A^{\frac{1}{2}}) \|_F^2 \, F(x)$

$$\begin{cases} \min\limits_{X \in S^n} \frac{1}{2} \| A^{\frac{1}{2}} (X - H) A^{\frac{1}{2}} \|^2 \\ \text{st} \quad XAs = s \end{cases}$$

$L(X, \lambda) = \frac{1}{2} \| A^{\frac{1}{2}} (X - H) A^{\frac{1}{2}} \|^2 - \langle \lambda, XAs - s \rangle$

$dL(X, \lambda) = \langle A^{\frac{1}{2}} (X - H) A^{\frac{1}{2}}, A^{\frac{1}{2}} (dX) A^{\frac{1}{2}} \rangle - \langle \lambda, (dX) As \rangle$

$= \langle A (X - H) A, dX \rangle - \langle \lambda s^T A, dX \rangle$

$\nabla L = A (X - H) A - \frac{1}{2} (\lambda s^T A + A s \lambda^T)$

$\| \quad \langle B, dX \rangle = \langle \underbrace{\tfrac{1}{2} (B + B^T)}_{\in S^n}, dX \rangle$

$$A(X-H)A = \frac{1}{2}(\lambda s^T A + A s \lambda^T)$$

$$X = H + \frac{1}{2} A^{-1}(\lambda s^T A + A s \lambda^T) A^{-1}$$

$$X = H + \frac{1}{2} A^{-1} \lambda s^T + \frac{1}{2} s \lambda^T A^{-1} \to X A s = s$$

отнормируем s:

$$\langle As, s \rangle = 1$$

$$HAs + \frac{1}{2} A^{-1}\lambda (s^T A s) + \frac{1}{2} s(\lambda^T s) = s$$

$$HAs + \frac{1}{2}\langle As, s\rangle A^{-1}\lambda + \frac{1}{2}\langle \lambda, s\rangle s = s$$

$$HAs + \frac{1}{2} A^{-1}\lambda + \frac{1}{2}\langle \lambda, s\rangle s = s$$

$$A^{-1}\lambda = 2s - \langle \lambda, s\rangle s - 2HAs$$

$$\langle s, \lambda\rangle = \langle s, 2As - 2AHAs - \langle \lambda, s\rangle As\rangle$$

$$\langle s, \lambda\rangle = 2\underbrace{\langle s, As\rangle}_{1} - 2\langle s, AHAs\rangle - \underbrace{\langle As, s\rangle}_{1}\langle \lambda, s\rangle$$

$$\cancel{2}\langle \lambda, s\rangle = 1 - \langle s, AHAs\rangle$$

$$A^{-1}\lambda = 2s - (1 - \langle s, AHAs\rangle)s - 2HAs =$$

$$= s + \langle s, AHAs\rangle s - 2HAs$$

$$X = H + \frac{1}{2} A^{-1}\lambda s^T + \frac{1}{2} s\lambda^T A^{-1}$$

$$X = H + \frac{1}{2} ss^T + \frac{1}{2}\langle s, AHAs\rangle ss^T - HAss^T +$$

$$+ \frac{1}{2} ss^T + \frac{1}{2}\langle s, AHAs\rangle ss^T - ss^T AH =$$

$$= H + ss^T + \langle s, AHAs\rangle ss^T - HAss^T - ss^T AH =$$

$$= (I_n - ss^T A) H (I_n - Ass^T)$$

$$H_+ = ss^T + (I_n - ss^T A) H (I - Ass^T)$$

$$H_+ = \frac{ss^T}{\langle As, s\rangle^2} + \left(I - \frac{ss^T A}{\langle As, s\rangle^2}\right) H \left(I - \frac{Ass^T}{\langle As, s\rangle^2}\right)$$

$y = As$:

$$H_+ = \frac{ss^T}{\langle y, s\rangle^2} + \left(I_n - \frac{sy^T}{\langle s, y\rangle^2}\right) H \left(I_n - \frac{ys^T}{\langle s, y\rangle^2}\right)$$

$F(x) = \frac{1}{2} \langle Ax, x \rangle$, $y = As = \nabla F(s)$

$s = x^1 - x^2$, $y = Ax^1 - Ax^2 = \nabla F(x^1) - \nabla F(x^2)$

$\min\limits_{x \in Q} F(x)$

1) если F - сильно выпукла, F - непрерывна

$$F(y) \geq F(x) + \langle \nabla F(x), y - x \rangle + \frac{\mu}{2} \|y - x\|^2 \quad \forall x, y \in Q$$

2) Q - замкнуто

тогда $\exists!$ минимум

$\square$ $F(x) \geq F(x^*) + \frac{\mu}{2} \|x - x^*\|^2 \Rightarrow !$

$\exists?$ $L_0 = \{x \in Q : F(x) \leq F(x_0)\}$ ограниченное мн-во

$x \in L_0 \Rightarrow F(x_0) \geq F(x) \geq F(x_0) + \langle \nabla F(x_0), x - x_0 \rangle + \frac{\mu}{2} \|x - x_0\|^2$

$\frac{\mu}{2} \|x - x_0\|^2 \leq |\langle \nabla F(x_0), x - x_0 \rangle| \leq \|\nabla F(x_0)\| \|x - x_0\|$

$\|x - x_0\| \leq \frac{2}{\mu} \|\nabla F(x_0)\|$ мн-во ограничено

минимум $\exists$. $\blacksquare$