Метод Ньютона

$\varphi(t) = 0 \qquad \varphi : \mathbb{R} \to \mathbb{R}$

t^0 – в окр t^*

$\Delta t + t^0 \approx t^*$

$$\varphi(t^0 + \Delta t) = \varphi(t^0) + \varphi'(t^0)\,\Delta t + o(\Delta t)$$

$$\varphi(t^0 + \Delta t) \approx \varphi(t^*) = 0 \Rightarrow \varphi(t^0) + \varphi'(t^0)\,\Delta t \approx 0$$

$$\Downarrow$$

$$\underbrace{\Delta t}_{t^1 - t^0} = -\frac{\varphi(t^0)}{\varphi'(t^0)}$$

$$\boxed{t^{k+1} = t^k - \frac{\varphi(t^k)}{\varphi'(t^k)}}$$

важно, что t^0 взяли из окр t^*

$$\varphi(t) = \frac{t}{\sqrt{1+t^2}} \qquad t^* = 0$$

$$\varphi'(t) = \frac{1}{(1+t^2)^{3/2}}$$

$$t^{k+1} = t^k - \frac{\varphi(t^k)}{\varphi'(t^k)} = t^k - \frac{t^k}{\cancel{\sqrt{1+(t^k)^2}}} \cdot \frac{(1+(t^k)^2)^{2/2}}{1} =$$

$$= -(t^k)^3$$

1) $|t^0| < 1 \qquad |t^1| < |t^0|$ – прибл к решению

2) $t^0 = \pm 1 \qquad t^k = \pm 1$ – колебания (нет сход)

3) $|t^0| > 1 \qquad |t^1| > |t^0|$ – расходимся

Переход №1:

$$\varphi(t) = 0 \quad \Rightarrow \quad \nabla f(x) = 0$$

классич. Ньютона $\Rightarrow$

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)$$

$$f(x) \le \underbrace{f(x^k) + \langle \nabla f(x^k); x - x^k \rangle + \frac{L}{2} \|x - x^k\|^2}_{\text{минимизируем по } x}$$

$$x^{k+1} = x^k - \frac{1}{L} \nabla f(x^k)$$

$$f(x) = \underbrace{f(x^k) + \langle \nabla f(x^k); x - x^k \rangle + \frac{1}{2} \langle x - x^k; \nabla^2 f(x^k)(x - x^k) \rangle}_{\text{мин. по } x} + o\left(\|x - x^k\|^2\right)$$

x из окр. x^k

$$\boxed{x^{k+1} = x^k - \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)}$$

для квадратичной задачи $\frac{1}{2}(x-b)^T A (x-b)$

$$x^1 = x^0 - \underbrace{A^{-1}}_{\text{если можно}} A(x^0 - b) = b \leftarrow \text{решение}$$

$$x^{k+1} = x^k - \gamma_k h^k$$

1) GD $\quad h^k = \nabla f(x^k)$

2) Newton $\quad h^k = \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)$

? убывание

$$\langle h^k; \nabla f(x^k) \rangle > 0$$

$$\left(\nabla f(x^k)\right)^T \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k) > 0$$

$$\nabla^2 f(x^k) \succ 0$$

Th. 1) $\|\nabla^2 f(x) - \nabla^2 f(y)\| \le M \|x-y\|$

2) $\nabla^2 f(x^*) \succeq LI \qquad L>0$

3) x^0 – в окр. x^* : $\|x^0 - x^*\| \le \dfrac{2L}{3M}$

Док-во:

$$x^{k+1} - x^* = x^k - \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k) - x^*$$

$$\nabla f(x^k) - \nabla f(x^*) = \int_0^1 \nabla^2 f\left(x^* + \tau(x^k - x^*)\right)(x^k - x^*)\, d\tau$$

$$= x^k - x^* - \left(\nabla^2 f(x^k)\right)^{-1} \int_0^1 \nabla^2 f\left(x^* + \tau(x^k - x^*)\right)(x^k - x^*)\, d\tau$$

$$= \left(\nabla^2 f(x^k)\right)^{-1} \nabla^2 f(x^k)(x^k - x^*) - \left(\nabla^2 f(x^k)\right)^{-1} \int_0^1 \nabla^2 f\left(x^* + \tau(x^k - x^*)\right)(x^k - x^*)\, d\tau$$

$$= \left(\nabla^2 f(x^k)\right)^{-1} \underbrace{G_k}_{\nabla^2 f(x^k) - \int_0^1 \nabla^2 f\left(x^* + \tau(x^k - x^*)\right) d\tau} (x^k - x^*)$$

$$\|G_k\| = \left\| \int_0^1 \left[\nabla^2 f(x^k) - \nabla^2 f\left(x^* + \tau(x^k - x^*)\right)\right] d\tau \right\|$$

$$\le \int_0^1 \left\| \nabla^2 f(x^k) - \nabla^2 f\left(x^* + \tau(x^k - x^*)\right) \right\| d\tau$$

M-лип гладкие

$$\le \int_0^1 M(1-\tau)\,\|x^k-x^*\|\,d\tau = \frac{M\|x^k-x^*\|}{2}$$

$$\|x^{k+1}-x^*\| = \left\| \left(\nabla^2 f(x^k)\right)^{-1} G_k \,(x^k-x^*)\right\|$$

$$\le \left\|\left(\nabla^2 f(x^k)\right)^{-1}\right\| \, \|G_k\| \, \|x^k-x^*\|$$

$$\le \left\|\left(\nabla^2 f(x^k)\right)^{-1}\right\| \cdot \frac{M\|x^k-x^*\|^2}{2}$$

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

спектр $\nabla^2 f(x^k) - \nabla^2 f(x^*)$ по модулю зажат

$$\nabla^2 f(x^k) - \nabla^2 f(x^*) \succeq -M\|x^k-x^*\| I$$

$$\nabla^2 f(x^k) \succeq \nabla^2 f(x^*) - M\|x^k-x^*\| I \succeq (L - M\|x^k-x^*\|) I$$

$$\left\|\left(\nabla^2 f(x^k)\right)^{-1}\right\| \le \frac{1}{L - M\|x^k-x^*\|}$$

$$\|x^{k+1}-x^*\| \le \frac{1}{L-M\|x^k-x^*\|} \cdot \frac{M\|x^k-x^*\|^2}{2} \le$$

$$\|x^0-x^*\| \le \frac{2L}{3M} \leftarrow \text{окрестность}$$

$$\le \frac{3M\|x^k-x^*\|^2}{2L}$$

(анализ сходимости квадратичной)

Сходимость локальная!

Как исправить?

- $$x^{k+1} = x^k - \gamma_k \underbrace{\left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)}_{h^k}$$

γ_k – как - хотите

$$f(x^{k+1}) - f(x^k) \le -c\,\gamma_k \langle \nabla f(x^k); \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k) \rangle$$

$$c \in (0; \tfrac{1}{2})$$

- $$f(x) \le f(x^k) + \langle \nabla f(x^k); x - x^k \rangle + \frac{1}{2} \langle x - x^k; \nabla^2 f(x^k)(x - x^k) \rangle + \frac{M}{6} \|x^k - x\|^3$$

$\forall x$

M – константа липшица гессиана

$$\boxed{x^{k+1} = \arg\min_{x \in \mathbb{R}^n} \varphi(x, x^k) = f(x^k) + \langle \nabla f(x^k); x - x^k \rangle + \frac{1}{2} \langle x - x^k; \nabla^2 f(x^k)(x - x^k) \rangle + \frac{M}{6} \|x^k - x\|^3}$$

кубический метод Ньютона

Новый метод

$$x^{k+1} = x^k - \gamma_k H^k \nabla f(x^k)$$

$\approx (\nabla^2 f(x^k))^{-1}$, но не так дорого

$$H^{k+1} = H^k + \Delta H^k \qquad H^k ?$$

$$\nabla f(x^k) - \nabla f(x^{k+1}) \approx \nabla^2 f(x^{k+1})(x^k - x^{k+1}) \cancel{+ o(\|x^k - x^{k+1}\|)}$$

$$x^{k+1} - x^k \approx \underbrace{(\nabla^2 f(x^{k+1}))^{-1}}_{H^{k+1}} (\nabla f(x^{k+1}) - \nabla f(x^k))$$

$$\underbrace{x^{k+1} - x^k}_{s^k} = H^{k+1} \underbrace{(\nabla f(x^{k+1}) - \nabla f(x^k))}_{y^k}$$

$$\boxed{s^k = H^{k+1} y^k}$$ – квазиньютоновские уравнения

d строк d^2 – переменных $\Rightarrow H^{k+1}$ – бесконечно много решений

$$H^{k+1} = H^k + \Delta H^k$$

$$\Delta H^k y^k = s^k - H^k y^k$$

1) Broyden

$$\Delta H^k = \mu_k q_k q_k^T \qquad \mu_k \in \mathbb{R} \qquad q_k \in \mathbb{R}^d$$

$$\mu_k q_k (q_k^T y_k) = s^k - H^k y^k$$

$$\underbrace{\mu_k (q_k^T y_k)}_{\in \mathbb{R}} \underbrace{q_k}_{\in \mathbb{R}^d} = \underbrace{s^k - H^k y^k}_{\in \mathbb{R}^d}$$

$$q^k = s^k - H^k y^k \qquad \mu_k = \frac{1}{q_k^T y_k}$$

2) DFP (Davidon - Fletcher - Powell)

$$\Delta H^k = \underline{\mu_1 s^k s^{kT} + \mu_2 H^k y^k (H^k y^k)^T}$$

$$\mu_1 s^k ((s^k)^T y^k) + \mu_2 H^k y^k ((H^k y^k)^T y^k) = s^k - H^k y^k$$

$$\underset{\parallel}{(\mu_1 (s^k)^T y^k)} s^k + \mu_2 ((H^k y^k)^T y^k) H^k y^k = s^k - H^k y^k$$

$$\mu_1 = \frac{1}{(s^k)^T y^k} \qquad \mu_2 = \frac{-1}{(H^k y^k)^T y^k}$$

Вычисления: 1) век × век $= O(d)$

2) матрица × век $= O(d^2)$

$+$

1) память $O(d^2)$

$\Downarrow$

L-BFGS

сверхлинейная – лучше GD

– хуже Newton

по доп выч. и памяти – хуже GD

– лучше Newton