Найти $t^* \in \mathbb{R}$: $\varphi(t^*) = 0$

$\varphi : \mathbb{R} \to \mathbb{R}$

$\triangle \quad t^0 \xrightarrow{\text{цель}} \Delta t : \quad t^0 + \Delta t \approx t^*$

$$\varphi(\underbrace{t^0 + \Delta t}_{\overset{??}{t^*}}) = \varphi(t^0) + \varphi'(t^0)\Delta t + o(\Delta t)$$

$\varphi(t^0 + \Delta t) \approx 0$

$$\varphi(t^0) + \varphi'(t^0)\Delta t \approx 0$$

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

$$t^1 = t^0 + \Delta t$$

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

метод Ньютона для поиска корней

Пример:

$$\varphi(t) = \frac{t}{(1+t^2)^{1/2}}$$

$$t^* = 0 \qquad \varphi'(t) = \frac{1}{(1+t^2)^{3/2}}$$

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

$$= t^k - \frac{\dfrac{t^k}{(1+(t^k)^2)^{1/2}}}{\dfrac{1}{(1+(t^k)^2)^{3/2}}}$$

$$= t^k\left(1-\left(1+(t^k)^2\right)\right) = -(t^k)^3$$

- $|t^0|<1$ $\quad \frac{1}{2} \to -\frac{1}{8} \to \frac{1}{8^2} \to -\frac{1}{(8^2)^3}$

  быстрая сходимость к решению

- $|t^0|=1$ $\quad 1 \to -1 \to 1 \to -1$

  нет сходимости

- $|t^0|>1$ $\quad$ расходимость

Метод Ньютона – только локальная сходимость

$$\min_{x\in\mathbb{R}^d} f(x)$$

в гладком случае: $\nabla f(x^*)=0$

Алгоритм 3 Метод Ньютона

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**

2: Вычислить $\nabla f(x^k)$, $\nabla^2 f(x^k)$

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

4: **end for**

Выход: x^K

Интуиция:

$$f(x) \approx f(x^k) + \langle \nabla f(x^k); x-x^k \rangle + \frac{1}{2}\langle x-x^k; \nabla^2 f(x^k)(x-x^k)\rangle \to \min_x$$

$\operatorname{argmin} = x^{k+1}$

$\frac{L}{2}\|x-x^k\|_2^2$ – для град спуска

$$\nabla f(x^k) + \nabla^2 f(x^k)(\underset{x^{k+1}}{x}-x^k) = 0$$

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

Док-во сходимости:

- f – μ-сильно выпукла $\quad \nabla^2 f(x) \succeq \mu I$
- $\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq M \|x-y\|_2$

спектральная норма

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

Формула Н-Л:

$$\nabla f(x^k) - \underset{0}{\underset{\|}{\nabla f(x^*)}} = \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*))(x^k - x^*)d\tau$$

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

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

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

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

$$\| \quad \|_2$$

$$\|x^{k+1} - x^*\|_2 =$$

$$\left\| \left(\nabla^2 f(x^k)\right)^{-1} \left[\int_0^1 \left(\nabla^2 f(x^* + \tau(x^k - x^*)) - \nabla^2 f(x^k)\right) d\tau\right] (x^k - x^*) \right\|_2$$

$$\|Ax\|_2 \leq \|A\|_2 \, \|x\|_2$$

$$\|x^{k+1} - x^*\|_2 \leq$$

$$\left\| \left(\nabla^2 f(x^k)\right)^{-1} \left[\int_0^1 \left(\nabla^2 f(x^* + \tau(x^k - x^*)) - \nabla^2 f(x^k)\right) d\tau\right] \right\|_2 \cdot \|x^k - x^*\|_2$$

$$\|AB\|_2 \leq \|A\|_2 \cdot \|B\|_2$$

$$\leq \|\nabla^2 f(x^k)^{-1}\|_2 \cdot \left\| \int_0^1 \left(\nabla^2 f(x^* + \tau(x^k - x^*)) - \nabla^2 f(x^k)\right) d\tau \right\|_2$$

$$\cdot \|x^k - x^*\|_2$$

1) $$\left\| \int_0^1 \left(\nabla^2 f(x^* + \tau(x^k - x^*)) - \nabla^2 f(x^k)\right) d\tau \right\|_2$$

$$\le \int_0^1 \|\nabla^2 f(x^* + \tau(x^k - x^*)) - \nabla^2 f(x^k)\|_2 \, d\tau$$

M - лип. гессиана

$$\le \int_0^1 M \|x^* + \tau(x^k - x^*) - x^k\|_2 \, d\tau$$

$$= \int_0^1 M(1-\tau) \|x^k - x^*\|_2 \, d\tau = \frac{M}{2}\|x^k - x^*\|_2$$

2) $\|(\nabla^2 f(x^k))^{-1}\|_2$

$\nabla^2 f(x^k) \succeq \mu I$ сильная выпуклость

$(\nabla^2 f(x^k))^{-1} \preceq \frac{1}{\mu} I$

$\|(\nabla^2 f(x^k))^{-1}\|_2 \le \frac{1}{\mu}$

$$\|x^{k+1} - x^*\|_2 \le \frac{1}{\mu} \cdot \frac{M}{2} \|x^k - x^*\|_2 \cdot \|x^k - x^*\|_2$$

Теорема об оценке сходимости метода Ньютона для μ-сильно выпуклых функций с M-Липшецевым гессианом

Пусть задача безусловной оптимизации с μ-сильно выпуклой целевой функцией f с M-Липшецевыми гессианом решается методом Ньютона. Тогда справедлива следующая оценка сходимости за 1 итерацию

$$\|x^{k+1} - x^*\|_2 \le \frac{M}{2\mu}\|x^k - x^*\|_2^2.$$

квадратичная сходимость

- $M = 2$ $\mu = 1$ $\|x^0 - x^*\|_2 = \frac{1}{2}$

$$\frac{1}{2} \to \left(\frac{1}{2}\right)^2 \to \left(\left(\frac{1}{2}\right)^2\right)^2 \to \left(\left(\left(\frac{1}{2}\right)^2\right)^2\right)^2$$

быстро

Локальная сходимость

$$\|x^1 - x^*\|_2 \le \|x^0 - x^*\|_2$$

$$\frac{M}{2\mu}\|x^0 - x^*\|_2 < 1 \Rightarrow \|x^0 - x^*\|_2 \le \frac{2\mu}{M}$$

Модификации Ньютона для глоб. сходимости

- Демпфированный Ньютон

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

$$\gamma = \arg\min_{\gamma} f(x^k + \gamma p_k)$$

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

- $$x^{k+1} = \arg\min_{x \in \mathbb{R}^d} \left\{ f(x^k) + \langle \nabla f(x^k); x - x^k \rangle + \frac{L}{2}\|x - x^k\|_2^2 \right\}$$

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

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^d} \left\{ f(x^k) + \langle \nabla f(x^k); x - x^k \rangle + \frac{1}{2}\langle x - x^k; \nabla^2 f(x^k)(x - x^k)\rangle + \frac{M}{6}\underline{\underline{\|x - x^k\|_2^3}} \right\}$$

M-Лип. гессиана

Кубический метод Ньютона

На что заменить гессиан?

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

$\uparrow$

$(\nabla^2 f(x^k))^{-1}$ для Ньютона

Идея: потребовать от H_k св-ва гессиана

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

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

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

квазиньютонов. уравнение: $\boxed{H_{k+1} y^k = s^k}$ $\boxed{H_{k+1}^T = H_{k+1}}$

беск. много решений

не хватает правил

- SR1/Broyden — одноранговая добавка

$$H_{k+1} = H_k + \underset{\in\mathbb{R}}{\mu_k} \underset{\in\mathbb{R}^d}{q^k} (q^k)^T$$

$$s^k = H_{k+1} y^k = (H_k + \mu_k q^k (q^k)^T) y^k$$

$$= H_k y^k + \underbrace{\mu_k}_{\in\mathbb{R}} q^k \underbrace{(q^k)^T y^k}_{\in\mathbb{R}}$$

$$q^k \parallel s^k - H_k y^k$$

$$\boxed{q^k = s^k - H_k y^k}$$

$$\cancel{q^k} = \mu_k \cancel{q^k} (q^k)^T y^k$$

$$1 = \mu_k (q^k)^T y^k \qquad \boxed{\mu_k = \frac{1}{(q^k)^T y^k}}$$

$O(d^2)$ операций на итерацию

($O(d^3)$ – Ньютон)

- BFGS

$$H_{k+1} = \underset{H \in \mathbb{R}^{d \times d}}{\arg\min} \|H - H_k\|^2$$

$$\text{s.t.} \quad s^k = H y^k$$

$$H^T = H$$

$$\|A\|_W = \|W^{1/2} A W^{1/2}\|_F$$

$$W y^k = s^k$$

$$\boxed{H_{k+1} = (I - \rho_k s^k (y^k)^T) H_k (I - \rho_k y^k (s^k)^T) + \rho_k s^k (s^k)^T \qquad \rho_k = \frac{1}{(y^k)^T s^k}}$$

правильно раскрыть скобки

$O(d^2)$

Квазиньютоновские методы без суперлинейной добавки

Квазиньютоновские методы: BFGS

- До такой формулы можно дойти по-другому. Рассмотрим $B_{k+1} = H_{k+1}^{-1}$. Для B квазиньютоновское уравнение выглядит как
$$B_{k+1}s^k = y^k$$
- Для B_{k+1} можно написать SR1 пересчет матрицы:
$$B_{k+1} = B_k + \frac{(y^k - B_k s^k)(y^k - B_k s^k)^T}{(y^k - B_k s^k)^T s^k}$$
- Смотрим на вид B_{k+1} и делаем из нее двухранговое изменение:
$$B_{k+1} = B_k + \mu_{k,1} y^k (y^k)^T + \mu_{k,2} B_k y^k (B_k y^k)^T$$
- Как и в SR1 можно подогнать $\mu_{k,1}$ и $\mu_{k,2}$:
$$B_{k+1} = B_k + \frac{y^k (y^k)^T}{(y^k)^T s^k} + \frac{B_k s^k (B_k s^k)^T}{(s^k)^T B_k s^k}$$
- Если теперь обратить B_{k+1} (формула Шермана-Маррисона-Вудберри), то получится выражение для H_{k+1}