- **История метода Ньютона**

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

Найти $t^* : \varphi(t^*) = 0$

Идея:

стартовая точка $t^0 \in \mathbb{R}$

В идеале найти $\Delta t : \quad t^0 + \Delta t \approx t^*$

Разложение в ряд:

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

$$\overset{\wr\wr}{\varphi(t^*) = 0} \Rightarrow \quad \varphi(t^0) + \varphi'(t^0)\Delta t \approx 0$$

$$\Downarrow$$

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

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

$$\boxed{t^{k+1} = t^k - \frac{\varphi(t^k)}{\varphi'(t^k)}} \leftarrow \text{метод Ньютона в классическом варианте}$$

- **Пример работы**

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

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

Итерация метода Ньютона:

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

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

Есть ли сходимость?

- $|t^0| > 1$ $\quad t^0 = 2 \to -8 \to 8^3 \to \dots$ расходится
- $|t^0| = 1$ $\quad t^0 = 1 \to -1 \to 1 \to -1$ колеблется в $-1; 1$
- $|t^0| < 1$ $\quad t^0 = 1/2 \to -\frac{1}{8} \to \frac{1}{8^3} \to \dots$ сходимость

очень быстро

Выводы из истор. справки:

⊖ локальная сходимость

⊕ быстрая сходимость

Обратно к оптимизации:

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

Тоже ищем 0, но $\nabla f(x^*) = 0$

Метод Ньютона для $\nabla f(x^*) = 0$

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

метод Ньютона для безусловной задачи оптимизации

Другая интуиция

Раскладываем f в окр. x^k

$$f(x) \approx \underbrace{f(x^k) + \langle \nabla f(x^k); x - x^k \rangle + \frac{1}{2} \langle x - x^k; \nabla^2 f(x^k)(x - x^k) \rangle}_{\text{минимизируем квадр. апроксимацию}}$$

≈ решение

Минимум квадрат. апроксимации:

$$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0$$

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

Для квадрат. з.: $\min \frac{1}{2} x^T A x \quad A \succ 0 \quad A \in S$

$$x^1 = x^0 - A^{-1} A x^0 = 0 \leftarrow \text{решение}$$

для квадр. задачи — за 1 итерацию, не дорогую

- Сходимость

Предположения:

1) f — μ- сильно выпуклая $\quad \nabla^2 f(x) \succeq \mu I$
2) M- Липшицев гессиан $\quad \|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \le M\|x-y\|_2$

спектральная

Док-во сходимости:

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

формулу Ньютона-Лейбница (1 лекция)

$$\underbrace{\nabla f(x^k)}_{(2)} - \underbrace{\cancel{\nabla f(x^*)}}_{0} = \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*))(x^k - x^*)\, d\tau$$

Подставляем (2) в (1):

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

Умножим „1":

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

Вынесем за „скобки"

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

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

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

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

$$\leq \underbrace{\| (\nabla^2 f(x^k))^{-1} \|_2}_{} \, \|G_k\|_2 \, \|x^k - x^*\|_2$$

$$\nabla^2 f(x) \succeq \mu I \Rightarrow (\nabla^2 f(x))^{-1} \preceq \frac{1}{\mu} I \Rightarrow \|(\nabla^2 f(x^k))^{-1}\|_2 \leq \frac{1}{\mu}$$

$$\leq \frac{1}{\mu} \|G_k\|_2 \, \|x^k - x^*\|_2 \leq$$

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

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

M – липшицевость гессиана

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

$$= \frac{M}{2} \|x^k - x^*\|_2$$

$$\leq \frac{M}{2\mu} \|x^k - x^*\|_2^2$$

Сходимость метода Ньютона

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

квадратичная сходимость

Сходимость есть? $\|x^1-x^*\|_2 < \|x^0-x^*\|_2$

выполнено, если $\|x_0-x^*\|_2 < \frac{2\mu}{M}$

локальная сходимость

Пример сходимости:

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

$\|x^1-x^*\|_2 = \left(\frac{1}{2}\right)^2 \to \left(\left(\frac{1}{2}\right)^2\right)^2 \to \left(\left(\left(\frac{1}{2}\right)^2\right)^2\right)^2$ квадратичная

Итоги по методу Ньютона:

⊕ квадратичная сходимость

⊖ локальная сходимость (от опт. решения)

⊖ дорогостоящие итерации

- Модификации для глобальной сходимости

1) Демпфирование (добавить шаг)

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

как подобрать шаг?

– см. 2 лекцию

– $\underset{\gamma \in \mathbb{R}}{\operatorname{argmin}} f(x^k+\gamma p^k)$ $\qquad p^k = -\left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)$

выпуклая по γ: дихотомия, золотое сечение

2) Кубический метод Ньютона

Интуиция:

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

$\Downarrow$ GD с шагом $\frac{1}{L}$

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

Та же самая идея в случае квадр. аппроксимации

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

далеко не самая простая задача

M-липшицев гессиан

- **Квазиньютоновские методы**

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

– в Ньютоне: $H_k = \left(\nabla^2 f(x^k)\right)^{-1}$ ← хотим дешевле

– идея: не использовать $\nabla^2 f$, но $H_k \to H_{k+1}$ за счет свойств гессиана

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

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

$$\boxed{s^k = H_{k+1} y^k}$$ ← квазиньютоновское ур-е (св-во гессиана)

$$\boxed{H_{k+1}^T = H_{k+1}}$$

сколько решений на H_{k+1}? ∞

$\Downarrow$

правило H_{k+1}
(H_{k+1} из H_k)

Частные случаи:

- SR1 / Broyden (одноранговое приближение):

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

Квазиньютоновское ур-е

$$s^k = H_{k+1} y^k = H_k y^k + \mu_k q^k \underbrace{(q^k)^T y^k}_{\in \mathbb{R}}$$

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

$$s^k - H_k y^k = \mu_k (q^k)^T y^k \cdot q^k$$

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

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

- BFGS

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

любая

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

BFGS исп. взвешен. фробениусову норму:

$$\|A\|_W = \|W^{1/2} A W^{1/2}\|_F \qquad W y^k = s^k$$

$$\boxed{H_{k+1} = \underbrace{\left(I - \rho_k s^k (y^k)^T\right) H_k \left(I - \rho_k y^k (s^k)^T\right)}_{\text{реализовывать не в виде } mat \times mat \ O(d^3)} + \rho_k s^k (s^k)^T \qquad \rho_k = \frac{1}{(y^k)^T s^k}}$$

Идея получить по-другому:

Пусть $B_{k+1} = H_{k+1}^{-1}$

для B_k запишем квазиньютон.: $B_{k+1} s^k = y^k$

SR1 для B_k: $B_{k+1} = B_k + \frac{(y^k - B_k s^k)(y^k - B_k s^k)^T}{(y^k - B_k s^k)^T s^k}$

отсюда выделим $y^k (y^k)^T$ $\quad B_k s^k (B_k s^k)^T$ $\quad \cancel{B_k s^k (y^k)^T}$

идея

$$B_{k+1} = B_k + \mu_1 y^k (y^k)^T + \mu_2 (B_k s^k)(B_k s^k)^T$$

применим квазиньютон. ур-е

$$B_{k+1} = B_k + \frac{(y^k)(y^k)^T}{(y^k)^T s^k} + \frac{(B_k s^k)(B_k s^k)^T}{(s^k)^T B_k s^k}$$

Шерман - Моррисон - Вудбери (SMW)
(формула обращения матрицы вида $A + BCD$)

$$H_{k+1} \xRightarrow{SMW} B_{k+1}$$

Итоги по квазиньютоновским методам:

⊕ дешевая итерация $O(d^2)$ и не надо вычислять гессиан

⊕ глобальная суперлинейная сходимость
(хуже Ньютона
лучше Nesterov, GD)