Историческая справка

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

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

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

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

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

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

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

метод Ньютона в классической форме

Пример

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

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

Итерации м. 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?

- $|t^0|>1$ $\quad t^0=2 \to -8 \to 8^3 \to -(8^3)^3 \not\to 0$ расход.
- $|t^0|=1$ $\quad t^0=1 \to -1 \to 1 \to -1 \not\to 0$ колеб в $-1;1$
- $|t^0|<1$ $\quad t^0=1/2 \to -1/8 \to \frac{1}{8^3} \to 0$ сходимость

  очень быстро

Вывод:

⊕ быстрая сходимость

⊖ локальная сходимость (только в окр.

Обратно к оптимизации

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

Тоже ищем 0, но $\nabla f(x^*)=0$

Метод Ньютона для $\nabla f(x^*)=0$

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

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

метод Ньютона для задачи безусл. оптим.

Другая интуиция

Раскладываем в ряд в окр x^k

$$f(x) \approx \underbrace{f(x^k) + \langle \nabla f(x^k); x-x^k\rangle + \frac{1}{2}\langle x-x^k; \nabla^2 f(x^k)(x-x^k)\rangle}_{\text{миним. кв. апр вместо } f}$$

$\approx$ решение для f

$\nabla = 0$

Минимизируем кв. аппроксимацию:

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

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

Пример

для кв. задачи $\min\limits_{x\in\mathbb{R}^d} \frac{1}{2}x^T A x \qquad A \succeq 0 \qquad A \in S$

$$x^1 = x^0 - \underbrace{A^{-1}}_{\nabla^2}\underbrace{Ax^0}_{\nabla} = 0 \leftarrow \text{решение}$$

для кв. задачи за 1 итерацию, но дорогая

Сходимость

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

спектральная

Док-во:

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

ф-ла Нютон-Лайбниц

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

Подставяме (2) в (1)

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

Умножаваме „1“:

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

Вадим за „скоби“

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

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

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

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

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

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

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

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

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

M – липшицевость гессиана

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

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

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

Сходимость м. Ньютона

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

квадратичная сходимость

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

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

отвечает за лок. сходимость

Пример сходимости:

$M = 2 \quad \mu = 1 \quad \|x^0 - x^*\| = 1/2$

$\|x^k - x^*\|_2 : \quad 1/2 \to 1/4 \to (1/4)^2 \to \left(\left(\frac{1}{4}\right)^2\right)^2$ квадрат.

Выводы по м. Ньютона:

⊕ квадратичная сходимость

⊖ локальная сходимость

⊖ дороговизна итерации

Модификации для глобальной сходимости

1) Демпфированный (добавить шаг)

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

как подобрать шаг?

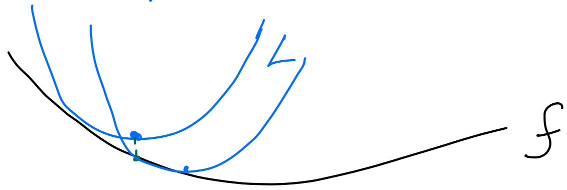
- см. лекцию 2
- $\underbrace{\arg\min_{\gamma \in \mathbb{R}} f(x^k + \gamma p^k)}_{\text{если } f \text{ выпукла, то з. по } \gamma \text{ тоже выпукла}}$ $\qquad p^k = -\left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)$

2) Кубический метод Ньютона

Интуиция

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

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



Та же самая идея в сл. нв строг:

$$x^{k+1} = \arg\min_{x\in\mathbb{R}^d} \Big(f(x^k) + \langle \nabla f(x^k); x - x^k \rangle + \frac12 \langle x - x^k; \nabla^2 f(x^k)(x - x^k) \rangle + \frac{M}{6}\|x - x^k\|_2^3 \Big)$$

M - лип. гессиана

далеко не самая простая задача

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: использ. не $(\nabla^2)^{-1}$, а некотор. приближен. со св-ми гессиана

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

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

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

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

сколько реш. на H_{k+1}

∞

$\Downarrow$

правило H_{k+1} через H_k

Частные случаи

• SR1 / Broyden (одноранг. приближение):

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

$O(d^2)$ операций

Квазиньютон. ур-е:

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

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

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

$q^k \parallel s^k - H_k y^k \Rightarrow \boxed{q^k = s^k - H_k y^k}$ $O(d^2)$

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

• BFGS

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

любая

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

$$H^T = H$$

BFGS исп. взвеш. Фроб. норма:

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

$$H_{k+1} = \underbrace{\left(I - \rho_k s^k (y^k)^T\right) H_k \left(I - \rho_k y^k (s^k)^T\right)}_{O(d^2)} + \rho_k s^k (s^k)^T$$

где $\rho^k = \dfrac{1}{(y^k)^T s^k}$ избежали mat x mat $O(d^3)$

Идея получить по-другому:

$$B_{k+1} = H_{k+1}^{-1}$$

для B_{k+1} тоже квазиньютоновское: $B_{k+1} s^k = y^k$

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 возникли $y^k (y^k)^T$ $\quad B_k s^k (B_k s^k)^T$

$\cancel{B_k s^k (y^k)^T}$ ← идея

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

найти

применили квазиньютон. ур-е

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

Шермана-Морриса-Вудбери (SMW)

(ф-ла обр. матриц вида $A + BCD$

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

Выводы по квазиньютоновским методам

⊕ дешёвая итерация $O(d^2)$ и не надо вычисл. гессиан

⊕ глоб. сходимость суперлинейная (Ньютон квадрат.)

(лучше Nesterov)