История метода Ньютона

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

найти $t^* : \varphi(t^*) = 0$

Идея:

стартовая точка $t^0 \in \mathbb{R}$

в идеальном мире найти $\Delta t : t^0 + \Delta t \approx t^*$

Разложение в ряд

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

$\varphi(t^0+\Delta t) \approx \varphi(t^*) = 0$, $o(\Delta t) \approx 0$

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

$$\Downarrow$$

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

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

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

← метод Ньютона в классическом варианте

Пример работы:

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

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

Итерация м. Ньютона:

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

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

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

Есть ли сходимость?

- $|t^0| > 1$ $\quad t^0 = 2 \to -8 \to 8^3 \to -(8^3)^3$ расход.
- $|t^0| = 1$ $\quad t^0 = 1 \to -1 \to 1 \to -1$ колеб в ± 1
- $|t^0| < 1$ $\quad t^0 = \frac{1}{2} \to -\frac{1}{8} \to \left(\frac{1}{8}\right)^3 \to -\left(\left(\frac{1}{8}\right)^3\right)^3$ сходимость

очень быстро

Вывод из истории:

⊕ быстрая сходимость

⊖ локальная сходимость

Обратимся к оптимизации:

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

Тогда ищем 0, но $\nabla f(x^*) = 0$

Метод Ньютона для $\nabla f(x^*) = 0$:

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

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

метод Ньютона для задачи безусл. оптимизации

Другая интуиция:

Раскладываем f в окр x^k

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

≈ решение $f(x)$

$\nabla = 0$

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

$$\Downarrow$$

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

Для квадр задач

$$\min_{x \in \mathbb{R}^d} \frac{1}{2} x^T A x \qquad A \succ 0 \qquad A \in S$$

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

для кв задачи — за 1 итерацию, но дорого

Сходимость:

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

спектральная

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

квадратичная сходимость

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

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

лок. сходимость

Пример

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

$\|x^1-x^*\|_2 = \left(\frac{1}{2}\right)^2 \longrightarrow \left(\left(\frac{1}{2}\right)^2\right)^2 \rightarrow \left(\left(\left(\frac{1}{2}\right)^2\right)^2\right)^2$ квадр. сход.

Итог по м. Ньютона:

⊕ кв. сход.

⊖ лок. сход. (из окр. решения)

⊖ дороговизна 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 подбирать шаг?

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

если f вып., то вып. по γ: дихотомия, золотое сечение

2) Кубический метод Ньютона:

Интуиция

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

$$\Downarrow$$

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

Идея:

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

М-лип. гессиан

не самая простая задача

NB $3 \to 2$

Квазиньютоновские методы

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

- в Ньютоне: $H_k = (\nabla^2 f(x^k))^{-1}$ ← дорого
- идея: не использ. $\nabla^2 f$, но $H_k \to H_{k+1}$ с учетом св-в гессиана
  1) симметричен
  2) разложение в ряд:

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

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

$$\boxed{s^k = H_{k+1} y^k}$$ ← квазиньютон ур-е

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

сколько решений? ∞

$\Downarrow$

правило H_{k+1} из H_k

Частный случай!

- SR1 / Broyden (одноранговая добавка):

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

Квазиньютон. ур-е

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

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

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

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

$\mu_k (q^k)^T y^k = 1$

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

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

• BFGS (д/з)

<u>Итоги по кв.-ньютон. методы</u>

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

⊕ глоб. суперлинейная сходимость
(хуже Ньютона
лучше Нестерова)