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

поиск корня

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

$\exists\ t^0 \in \mathbb{R}$ хотим Δt : $t^0+\Delta t \approx t^*$

разложим в ряд

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

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

$$\Downarrow$$

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

$$\Downarrow$$

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

$$\Downarrow$$

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

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

Фишка: суть t^0 в „хорош." окр t^*

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

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

Метод Ньютона:

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

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

- $|t^0| < 1$ — сходится $\quad t^0 = \frac{1}{2} \to -\frac{1}{8} \to \left(\frac{1}{8}\right)^3 \to \ldots$ быстро!
- $|t^0| = 1$ — колебания $1; -1$
- $|t^0| > 1$ — расходится

Локальная сходимость, но быстрая

Возвращаемся к

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

хотим $\nabla f(x^*) = 0$

$$\frac{\varphi(t^k)}{\varphi'(t^k)} \Rightarrow \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)$$

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

По-другому: разложение

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

в окр x^k $\qquad (x-b)^T A (x-b)$

$\min f(x) \Rightarrow \min$ аппроксимации

∇ аппроксимации $= 0$

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

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

проблем

- вычисл. гессиана
- обращение матрицы

<u>для квадрат. задачи</u> $\frac{1}{2} x^T A x \to \min\limits_{x \in \mathbb{R}^d}$ $\quad A \in S^+$

$$x^0 - \underbrace{A^{-1}}_{\nabla^2}\big(\underbrace{A x^0}_{\nabla}\big) = x^0 - x^0 = 0$$

↑ решение

за 1 дорогую итерацию!

<u>Сходимость</u> $\quad A \succeq B \quad (A-B)$ – неотрицат. полуопред.

- $\nabla^2 f(x) \succeq \mu I \quad \forall x \in \mathbb{R}^d$ ← μ-сильная выпуклость
- $\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \le M \|x - y\|_2 \quad \forall x, y \in \mathbb{R}$ ← M-липш. гессиан

<u>NB</u> $\underset{\in \mathbb{R}^d}{\|x\|}$ ← векторная норма

$$\underset{\in \mathbb{R}^{d \times d}}{\|A\|} = \sup_{x \in \mathbb{R}^d} \frac{\|Ax\|}{\|x\|} = \sup_{\|\tilde{x}\| = 1} \|A \tilde{x}\|$$

Итерация:

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

Формула Ньютона-Лейбница

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

$\nabla f(x^*) = 0$, подставим $\nabla f(x^k)$ в шаг

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

умножим „1"

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

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

выносим за скобка

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

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

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

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

$$\frac{\|Ax\|_2}{\|x\|_2} \leq \sup_x \frac{\|Ax\|_2}{\|x\|_2} = \|A\|_2 \qquad \|Ax\|_2 \leq \|A\|_2 \cdot \|x\|_2$$

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

$$\left\| \left(\nabla^2 f(x^k)\right)^{-1} \right\|_2 \leq \frac{1}{\mu} \quad \Leftarrow \quad \|\nabla^2 f(x^k)\|_2 \geq \mu$$

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

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

$$\|\Sigma\| \leq \Sigma \|\ \|$$

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

M- липшиц. гессиана

$$\le \frac{1}{\mu}\int_0^1 M \|x^k - x^*\| \tau \, d\tau \cdot \|x^k - x^*\|_2$$

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

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

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

Физика оценки:

$$\|x^1 - x^*\| < \|x^0 - x^*\| \leftarrow \text{хотим}$$

$\exists$ $M = 2$ $\mu = 1$, тогда

$$\|x^1 - x^*\| \le \|x^0 - x^*\|_2^2$$

нужно, чтобы $\|x^0 - x^*\|_2 < 1$

$$\|x^0 - x^*\|_2 = \frac{1}{2} \quad \to \frac{1}{2} \to \frac{1}{4} \to \frac{1}{16} \to \frac{1}{(16)^2}$$

⊕ квадратичная скорость сходимости

⊖ локальная (из окрестности)

⊖ дорогая итерация ($\nabla^2 f$ и обращение + хранение)

Борьба с „локальностью" сходимости:

• шаг

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

шаг

демпфированный метод Ньютона

как подбирать γ? $\quad \arg\min_{\gamma} f(x^k + \gamma p^k)$

золотое сечение $\qquad \in \mathbb{R}$

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

• по стопам град. спуска:

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

$$\nabla f(x^k) + L(\underbrace{x}_{x^{k+1}} - x^k) = 0$$

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

для оценки сверху с произв. 2х порядков

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

M-лип. гессиан

неприятно

кубический метод Ньютона

Борьба с „дороговизной" итерации

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

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

Свойства $\nabla^2 f$ для H:

- симметричность $H_{k+1}^T = H_{k+1}$
- разложение

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

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

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

Хотим искать H_{k+1}, исходя из этих св-в.

сущ. ли H_{k+1}?

d уравнений, $\sim d^2$ неизвестных $\}$ ∞ много решений на H

Нужны ещё правила подсчета H_{k+1}:

- SR1 – одно-ранговое приближение

$$H_{k+1} = H_k + \underbrace{\mu_k}_{\in\mathbb{R}} \underbrace{q^k}_{\in\mathbb{R}^d} (q^k)^T \rightarrow O(d^2)$$

найти μ и q

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

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

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

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

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

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

SR1

- BFGS

близко к H_k

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

но с условиями

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

можно любую норму

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

взвеш. Фробениуса $\quad W y^k = s^k$

$H_{k+1} =$ см. презентацию

+ стоимость $O(d^2)$ 1 итерации

+ глобальная сходимость, суперлинейная (не квадрат.)

+ только градиенты

+ есть версии эффективные по памяти

− не подходит для больших невыпуклых задач