История метода Ньютона

Найти t^*, что $\varphi(t^*)=0$ $\qquad \varphi: \mathbb{R} \to \mathbb{R}$

стартуем $t^0 \in \mathbb{R}$

хотим? Δt, что $t^0 + \Delta t \approx t^*$

$$\underset{\substack{\Updownarrow \\ \varphi(t^*)=0}}{\varphi(t^0+\Delta t)} \simeq \varphi(t^0) + \varphi'(t^0)\Delta t + o(\Delta t)$$

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

$$\Downarrow$$

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

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

метод Ньютона в классическом варианте →

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

Пример работы

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

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

Итерация метода Ньютона

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

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

Есть ли сходимость?

- $|t^0| > 1$ расходимость $2 \to -8 \to 8^3 \to \dots$
- $|t^0| = 1$ колебания в $1; -1$ $1 \to -1 \to 1 \to -1 \dots$
- $|t^0| < 1$ сходимость $1/2 \to -1/8 \to 1/8^3 \to$ все быстро

Выводы из примера

- 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 оптимизации

Интуиция: раскладываем f в окр. x^k

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

≈ решение

минимум квадратичной задачи:

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

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

Стоит ли игра свеч: $\min\limits_{x\in\mathbb{R}^d} \frac{1}{2}x^T A x \qquad A \succ 0 \qquad A \in \mathbb{S}$

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

для квадратичной $\Rightarrow$ за 1 итерацию, но дорого

(GD — только ∇

Ньютон — ∇^2)

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

Предположения:

- μ-сильная выпуклость $\nabla^2 f(x) \succeq \mu I$
- M-Липшицевость $\nabla^2 f$: $\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \le M\|x-y\|_2$

спектральная

Док-во сходимости:

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

формула Ньютона-Лейбница (на 1 лекции)

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

Подставим (2) в (1)

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

Умножаю „единицу"

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

Вынесем $\sim$, $\sim$ за скобки

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

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

Расстояние до решения

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

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

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

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

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

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

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

M- лип. ∇^2

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

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

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

Сходимость метода Ньютона

квадратичная сходимость (экв. мону-тонно вогнутой) →

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

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

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

локальность сходимости

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

Пример снизу скорости сходимости

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

тогда

$$\|x^1 - x^*\|_2 \le \frac{1}{2^2} \qquad \|x^2 - x^*\|_2 \le \left(\frac{1}{2^2}\right)^2 \ldots$$

Модификации для глобальной сходимости

- 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. 1 лекцию
- $\arg\min\limits_{\gamma \in \mathbb{R}} f(x^k + \gamma p^k) \qquad p^k = -\left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)$

одномерная выпуклая: деление, золотое сечение

- Кубический метод Ньютона

Интуиция:

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

$\Downarrow$ GD с шагом $\frac{1}{L}$

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

То же самое для ньютона:

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

M - липшицев, то это оцен. сверху

Квазиньютоновские методы

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

- в Ньютоне: $H_k = (\nabla^2 f(x^k))^{-1}$ => хотим держать в H_k
- какие-то св-ва гессиана заложить

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

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

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

$$\boxed{s^k = H_{k+1} y^k}$$

квазиньютоновские ур-е

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

сколько решений на H_{k+1}?

H $d^2/2$ неизвестных

d уравнений

∞ решений

=> еще как-то сузить правила поиска H

Частные случаи:

- SR1 / Broyden

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

одно-ранговое изменение

Квазиньютоновское ур-е

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

$$= H_k y^k + \mu_k \left((q^k)^T y^k\right) q^k$$

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

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

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

- BFGS

$$H_{k+1} = \underset{H \in \mathbb{R}^{d \times d}}{\arg\min} \|H - H_k\|^2$$
$$\text{s.t.} \quad s^k = H y^k$$
$$H^T = H$$

норма Фробениуса (взвешенная): $\|A\|_W = \|W^{1/2} A W^{1/2}\|_F$

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

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

По-другому: $B_{k+1} = H_{k+1}^{-1}$

для B квазиньют. ур: $B_{k+1} s^k = y^k$

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

для B_k (по аналогии)

чуть-чуть поменяем SR1 для B_k
до 2х-рангового

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

применяем квазиньютоновское

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

B_{k+1} можно обратить и получить H_{k+1}
(Шермана-Морисона-Вудберри) BFGS

Итоги:

- дешевле итерация $O(d^2)$ по сравнению с $O(d^3)$ у Ньютона
- допущение глобальная сверхлинейная сходимость