Метод Ньютона. Квазиньютоновские методы. Семинар 3. 26 февраля 2020 г.

Семинарист: Данилова М.

Метод Ньютона

Эвристические соображения

Основная идея: квадратичная аппроксимация $f(x)$ в точке x_k функцией $f_k(x)$:

$$f_k(x) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \left\langle \frac{1}{2} \nabla^2 f(x_k)(x - x_k), x - x_k \right\rangle.$$

Новое приближение: точка минимума $f_k(x)$:

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} f_k(x)$$

$$\nabla f_k(x) = 0$$

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

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

Algorithm 1 Классический метод Ньютона

1: Пусть $x_0 \in \mathbb{R}^n, \quad h_k = -\left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k), \quad \alpha_k = 1.$
2: Вычислим $x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k) \quad k = 0, 1, ..$

Сходимость

Теорема 1 (**Поляк**). Пусть функция $f(x)$ - дважды дифференцируема, $\nabla^2 f(x)$ удовлетворяет условию Липшица с константой L, $f(x)$ сильно выпукла с константой μ и начальное приближение удовлетворяет условию

$$q = \left(\frac{\mathrm{L}}{2\mu^2}\right) \|\nabla f(x_0)\| < 1.$$

Тогда метод Ньютона сходится к точке глобального минимума x^* с квадратичной скоростью:

$$\|x_k - x^*\| \leq \frac{2\mu}{\mathrm{L}} q^{(2^k)}$$

Теорема 2 (Нестеров). Пусть функция $f(x)$:

1. дважды непрерывно дифференцируема и её гессиан удовлетворяет условию Липшица с константой L;
2. существует точка локального минимума с положительно определённым гессианом

$$\nabla^2 f(x^*) \succeq \mu I_n, \ \mu > 0;$$

3. начальная точка x_0 достаточно близка к точке минимума x^*

$$\|x_0 - x^*\|_2 \leq \frac{2\mu}{3\mathrm{L}}.$$

Тогда метод Ньютона сходится квадратично:

$$\|x_{k+1} - x^*\|_2 \leq \frac{\mathrm{L}\|x_k - x^*\|_2^2}{2(\mu - \mathrm{L}\|x_k - x^*\|_2)}.$$

Метод Ньютона для уравнений

Метод может применяться для решения произвольных нелинейных уравнений:

$$g(x) = 0, \quad g : R^n \to R^n.$$

На к-й итерации решается линеаризированное уравнение:

$$g(x_k) + g'(x_k)(x - x_k) = 0$$
$$x_{k+1} = x_k - g'(x_k)^{-1} g(x_k)$$

Замечания

1. метод второго порядка
2. квадратичная скорость сходимости вблизи решения, нельзя говорить о сходимости при любом x_0
3. при отсутствии условия Липшица $\to$ геометрическая скорость сходимости
4. сложность $O(n^3)$ - обращение гессиана
5. требования к $f(x)$: дважды дифференцируема, гессиан липшицев и положительно определен (для min)

Модификации метода Ньютона

Классический метод Ньютона

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

Резюме:

- локальная сходимость
- жесткие требования к $f(x)$
- большой объем вычислений
- быстрая сходимость

Демпфированный метод Ньютона (с переменным шагом)

Цель: придать глобальную сходимость

$$x_{k+1} = x_k - \alpha_k [\nabla^2 f(x_k)]^{-1} \nabla f(x_k), \quad \alpha_k \text{ - длина шага}$$

Одним из недостатков метода Ньютона с постоянным шагом является его локальная сходимость.

Применяемый даже для минимизации выпуклых функций, он не всегда может найти решение задачи.

Пример 2.

$$f(x) = x \operatorname{arctg} x - \frac{1}{2} \ln(1 + x^2), \quad x \in \mathbb{R}$$

$$f'(x) = \operatorname{arctg} x \quad f''(x) = (1 + x^2)^{-1} > 0$$

Данная функция является строго выпуклой и достигает своего минимума на $\mathbb{R}$ в нуле.

Если воспользоваться методом Ньютона с постоянным шагом, то он сойдется только при условии $|x_0| < 1,392$.

Для того, чтобы расширить область сходимости метода Ньютона, применяют его вариант с переменным шагом:

$$x_{k+1} = x_k - \alpha_k [\nabla^2 f(x_k)]^{-1} \nabla f(x_k).$$

Выбор длины шага α_k?

- одномерный поиск

$$\alpha_k = \arg\min_{\alpha > 0} f\left(x_k - \alpha [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)\right)$$

- дробление начиная с $\alpha = 1,\ \alpha_k = \gamma\alpha_k,\ 0 < \gamma < 1$ до выполнения каких-либо условий, например

$$f(x_{k+1}) \leq f(x_k) - \alpha q \left([\nabla^2 f(x_k)]^{-1}\nabla f(x_k), \nabla f(x_k)\right)$$
$$\|\nabla f(x_{k+1})\|^2 \leq (1-\alpha q)\|\nabla f(x_k)\|^2$$
$$0 < q < 1$$

Теорема 3 (Жадан). Пусть $f(x)$ - сильно выпуклая дважды непрерывно дифференцируемая функция, удовлетворяющая условию

$$\mu\|s\|^2 \leq \langle s, \nabla^2 f(x)s\rangle \leq \mathrm{L}\|s\|^2.$$

Демпфированный метод Ньютона с правилом выбора шага Армихо (или одномерной минимизации, наискорейший спуск) для сильно выпуклой функций сходится из любой точки, со сверхлинейной скоростью, в случае липшицевой второй производной - с квадратичной.

Регуляризация Левенберга - Марквардта

Цель: избавиться от требования $\nabla^2 f(x) \succ 0$

Если матрица $\nabla^2 f$ не положительно определена, её можно регуляризовать с помощью единичной матрицы.

То есть в шаге метода вместо гессиана использовать:

$$\nabla^2 f(x) + \gamma \mathrm{I} > 0$$
$$x_{k+1} = x_k - [\nabla^2 f(x_k) + \gamma \mathrm{I}]^{-1}\nabla f(x_k)$$

Квазиньютоновские методы

Цель: облегчить вычисления

Основная идея: не вычислять каждый раз обратную матрицу вторых производных, а приближатся к ней постепенно, итеративно накапливая инфрмацию.

Итерационный процесс общего вида:

$$x_{k+1} = x_k - \alpha_k \mathrm{H}_k \nabla f(x_k)$$

где H_k - симметричная положительно определенная матрица отличная от $\nabla^2 f(x_k)^{-1}$

H_k пересчитывается рекуррентным способом на основе информации, полученной на k-ой итерации при условии

$$\mathrm{H}_k - [\nabla^2 f(x_k)]^{-1} \to 0$$

в пределе $\to$ классический метод Ньютона

Algorithm 2 Квазиньютоновский метод

1: Выбираем $x_0 \in \mathbb{R}^n$. Положим $\mathrm{H}_0 = \mathrm{I}_n$. Вычислим $f(x_0)$, $\nabla f(x_0)$.
2: k-я итерация $(k \geq 0)$
- положим $h_k = \mathrm{H}_k \nabla f(x_k)$
- найдем $x_{k+1} = x_k - \alpha_k h_k$, используя правила выбора длины шага α_k
- вычислим $f(x_{k+1})$, $\nabla f(x_{k+1})$
- обновим H_k : $\mathrm{H}_k \to \mathrm{H}_{k+1}$

Квазиньютоновское правило:

Выбираем H_{k+1} так, чтобы выполнялось равенство

$$\mathrm{H}_{k+1} \left(\nabla f(x_{k+1}) - \nabla f(x_k) \right) = x_{k+1} - x_k$$

Это правило вытекает из следующего свойства квадратичной функции:

$$f(x) = \frac{1}{2} (\mathrm{A}x, x) + (b, x) + c$$

$$\nabla f(x) = \mathrm{A}x + b$$

$$\forall x, y \in \mathbb{R}^n \quad \nabla f(x) - \nabla f(y) = \mathrm{A}(x - y)$$

Существует много способов удовлетворить квазинютоновское правило.

Ниже приводится несколько наиболее популярных.

Обозначим

$$\Delta \mathrm{H}_k = \mathrm{H}_{k+1} - \mathrm{H}_k, \quad \gamma_k = \nabla f(x_{k+1}) - \nabla f(x_k), \quad \delta_k = x_{k+1} - x_k$$

Правило одноранговой коррекции:

$$\Delta \mathrm{H}_k = \frac{(\delta_k - \mathrm{H}_k\gamma_k)(\delta_k - \mathrm{H}_k\gamma_k)^\top}{\langle \delta_k - \mathrm{H}_k\gamma_k, \gamma_k\rangle}.$$

Правило Давидона - Флетчера - Пауэла (ДФП)

$$\Delta \mathrm{H}_k = \frac{\delta_k\delta_k^\top}{\langle \gamma_k, \delta_k\rangle} - \frac{\mathrm{H}_k\gamma_k\gamma_k^\top \mathrm{H}_k}{\langle \mathrm{H}_k\gamma_k, \gamma_k\rangle}.$$

Правило Бройдена-Флетчера-Гольдфарба-Шенно (БФГШ):

$$\Delta \mathrm{H}_k = \frac{\mathrm{H}_k\gamma_k\delta_k^\top + \delta_k\gamma_k^\top \mathrm{H}_k}{\langle \mathrm{H}_k\gamma_k, \gamma_k\rangle} - \beta_k \frac{\mathrm{H}_k\gamma_k\gamma_k^\top \mathrm{H}_k}{\langle \mathrm{H}_k\gamma_k, \gamma_k\rangle}.$$

$$\beta_k = 1 + \frac{\langle \gamma_k, \delta_k\rangle}{\mathrm{H}_k\gamma_k, \gamma_k}$$

БФГШ обычно упоминается в литературе как наиболее устойчивое к вычислительным погрешностям

Доверительные области

В соответствии с этим подходом вокруг точки x_k надо зафиксировать окрестность, в которой аппроксимация второго порядка обязана быть достаточно хорошей. Эта окрестность $\Delta(x_k)$ называется доверительной областью. Можно, например, взять $\Delta(x_k) = \{x : \|x - x_k\| \leq \varepsilon,\ \varepsilon > 0\}$. Тогда следующая точка x_{k+1} будет выбираться как решение задачи

$$\min_{x \in \Delta(x_k)} \left[\langle \nabla f(x_k), x - x_k\rangle + \frac{1}{2}\langle x - x_k, \nabla^2 f(x_k)(x - x_k)\rangle \right]$$

Если $\Delta(x_k) = \mathbb{R}^n$, то это в точностью классический ньютоновский шаг.