Введение в численные методы. Одномерная минимизация. Семинар 7. 11 февраля 2020 г.

Семинарист: Данилова М.

Программа курса

1. Введение в численные методы.
2. Одномерная оптимизация.
3. Градиентный метод.
4. Метод Ньютона.
5. Квазиньютоновские методы.
6. Ускоренные градиентные методы.
7. Стохастический градиентный спуск.
8. Оптимальные методы и нижние оценки.
9. Проксимальные методы.Метод внутренней точки.
10. Введение в pytorch.
11. Введение в оболочку и пакеты cvxpy, варианты солверов.

Литература

1. Б.Т. Поляк. Введение в оптимизацию.
2. Ю.Е. Нестеров. Методы выпуклой оптимизации.
3. А.В. Гасников. Современные численные методы оптимизации.
4. В.Г. Жадан. Методы оптимизации.
5. Matrixcookbook.
6. Sébastien Bubeck. Convex Optimization: Algorithms and Complexity.

Введение в численные методы оптимизации

В этом курсе нас будут интересовать алгоритмы(методы) решения оптимизационных задач вида

$$\min_{x \in \mathrm{X}} \quad f(x)$$

Можно решить задачу аналитически с помощью условий оптимальности или геометрического подхода, но это возможно сделать только в самых простых случаях. Когда это сделать не представляется возможным мы приходим к понятию **численных методов**.

- Точно решить задачу с помощью численных методов невозможно из-за погрешности машинной арифметики.
- Решить задачу - означает найти её приближенное решение с заранее заданной точностью $\varepsilon > 0$.

Численные методы

Общая итеративная схема

1. задаем начальную точку x_0 и требуемую точность $\varepsilon > 0$
2. выбираем направление $d(x_k) = d_k$
3. выбираем длину шага $\alpha(x_k) = \alpha_k$
4. делаем шаг $x_{k+1} = x_k + \alpha_k d(x_k)$ (линейный поиск)
5. проверка условия остановки:
   - либо x_{k+1} - ответ
   - либо $k := k + 1$ и переходим на шаг 2

Критерии остановки

1. сходимость по аргументу: $\|x_k - x^*\| \leq \varepsilon$
2. сходимость по функционалу: $|f_k - f^*| \leq \varepsilon$
3. если $f(x)$ - дифференцируема, то можно проверять: $\|\nabla f(x_k)\| \leq \varepsilon$

Замечания

- в общем случае x^* нам неизвестно, тогда

$$\|x_{k+1} - x_k\| = \|x_{k+1} - x_k + x^* - x^*\| \leq \|x_{k+1} - x^*\| + \|x_k - x^*\| \leq 2\varepsilon$$

- критерии 1) и 2) основаны на использовании абсолютных изменений, лучше использовать их относительные изменения:
  - $\frac{\|x_{k+1}-x_k\|}{\|x_k\|}$
  - $\frac{|f(x_{k+1})-f(x_k)|}{|f(x_k)|}$
- критерии остановки дополняются заданием максимально возможного числа итераций N
- Пусть $\bar{x}$ - ваше решение, а x^* - истинное решение, тогда с помощью методов можно получить $f(\bar{x}) \approx f(x^*)$, но нельзя гарантировать $\|\bar{x} - x^*\| \leq \delta$

Оракул

Предполагается, что численный метод может накапливать специфическую информацию о задаче при помощи некоторого **оракула**.

- **Оракул** – некоторое устройство, которое отвечает на последовательные вопросы численного метода.
- Единственной информацией, получаемой в ходе работы итеративного метода являются ответы оракула.
- Ответы оракула локальные, то есть изменяя задачу далеко от тестовой точки x исходный ответ в точке x не меняется.

Примеры оракулов

- **Оракул нулевого порядка** в запрашиваемой точке x возвращает значение целевой функции $f(x)$.
- **Оракул первого порядка** в запрашиваемой точке возвращает значение функции $f(x)$ и её градиент в данной точке $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)$.

Классификация

1. (a) методы нулевого порядка: оракул возвращает значени функции $f(x)$
   (b) методы первого порядка: $f(x), f'(x)$
   (c) методы второго порядка: $f(x), f'(x), f''(x)$
2. (a) одношаговые методы: $x_{k+1} = \Phi(x_k)$
   (b) многошаговые методы: $x_{k+1} = \Phi(x_k, x_{k-1})$

Как сравнивать методы оптимизации?

1. **Сложность**
   - **Аналитическая сложность** — число обращений к оракулу, необходимое для решения задачи с точностью ε.
   - **Арифметическая сложность** — общее число вычислений (включая работу оракула), необходимых для решения задачи с точностью ε.
2. **Скорость сходимости**

Скорость сходимости

- линейная (геометрическая)

$$\exists\, 0 < C < 1,\ K \geq 0: \quad \|x_{k+1} - x^*\| \leq C\|x_k - x^*\| \quad \forall k \geq K$$

$$\exists\, C > 0,\ 0 < q < 1,\ K \geq 0: \quad \|x_k - x^*\| \leq Cq^k \quad \forall k \geq K$$

- сверхлинейная

$$\|x_{k+1} - x^*\| \leq C_k\|x_k - x^*\|, \quad C_k > 0,\ C_k \to 0,\ k \to \infty$$

- квадратичная

$$\exists\, C > 0,\ k \geq 0: \quad \|x_{k+1} - x^*\| \leq C\|x_k - x^*\|^2 \quad \forall k \geq K$$

$$\exists\, C > 0,\ 0 < q < 1,\ K \geq 0: \quad \|x_k - x^*\| \leq Cq^{2^k} \quad \forall k \geq K$$

Методы спуска

Общая итеративная схема

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k h_k \\ f(x_{k+1}) &< f(x_k) \end{aligned}$$

- $\alpha_k > 0$ - длина шага
- h_k - направление убывания
- x_0 - начальное приближение

Правила выбора длины шага

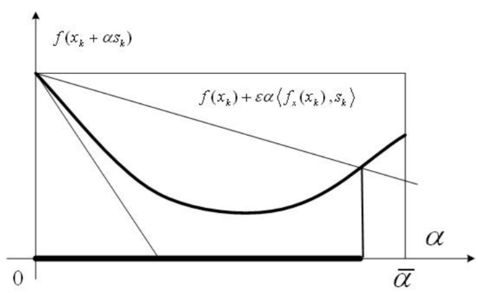
1. **Правило постоянного шага:**
$$\alpha_k = \alpha > 0$$

2. **Правило априорного выбора:**
последовательность $\{\alpha_k\}_{k=0}^{\infty}$ выбирается заранее. Например,
$$\alpha_k = \frac{\alpha}{\sqrt{k+1}}$$

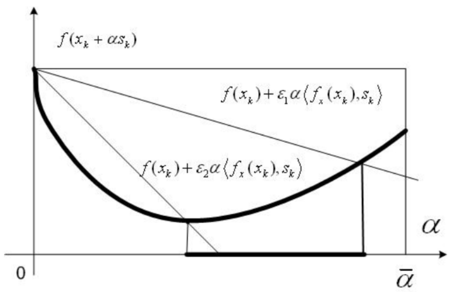
3. **Правило одномерной минимизации:**
$$\alpha_k = \arg\min_{\alpha>0} f(x_k + \alpha h_k)$$

4. **Правило Армихо:** задаются два числа $0 < \varepsilon < 1,\ 0 < \theta < 1$ и выбирают начальное значение длины шага $\bar{\alpha}$. Полагаем $\alpha = \bar{\alpha}$. Выбор α_k проводится согласно следующей двухэтапной процедуре:

   - Шаг 1. Проверяем выполнение условия
   $$f(x_k + \alpha h_k) - f(x_k) \le \varepsilon\alpha \left\langle f'_x(x_k), h_k \right\rangle,$$
   которое называется неравенством Армихо.
   - Шаг 2. Если неравенство Армихо не выполняется, то заменяем α на $\alpha := \theta\alpha$ и идем на шаг 1. В противном случае полагаем $\alpha_k = \alpha$ и заканчиваем процесс.



5. **Правило Голдштейна:** в нем задаются два параметра: $0 < \varepsilon_1 < 1$ и $0 < \varepsilon_2 < 1$, причем $\varepsilon_1 < \varepsilon_2$. Шаг α на k-й итерации подбирается таким образом, чтобы он удовлетворял условиям
$$\varepsilon_1 \le \frac{f(x_k + \alpha h_k) - f(x_k)}{\alpha \left\langle f'_x(x_k), h_k \right\rangle} \le \varepsilon_2.$$
Левое неравенство - это правило Армихо. Правое неравенство вводится для того, чтобы шаг не был достаточно малым.



Замечания

- правила постоянного шага и априорного выбора являются самыми простыми и используются в контексте задач выпуклой оптимизации;
- стратегия одномерной минимазации интересна только с теоретической точки зрения, так как на практике применима только тогда, когда вспомогательная задача решается либо очень быстро, либо аналитически;
- правила Армихо и Гольдштейна используются в большинстве практических алгоритмов;
- правило Армихо гарантирует, что функция в точке x_{k+1} не превосходит её линейной аппроксимации с коэффициентом наклона ε;
- правило Гольдштейна гарантирует то же, что и правило Армихо, а также что функция в точке x_{k+1} убывает не меньше, чем её линейная аппроксимация с коэффициентом наклона ε_2 (то есть расположена между графиками линейных апроксимаций с коэффициентами наклона $\varepsilon_1, \varepsilon_2$.).

Одномерная минимизация

В данном разделе речь пойдет о решении задачи минимизации функции одного аргумента на отрезке:

$$\min_{x \in \mathrm{X}} \quad f(x), \tag{1}$$

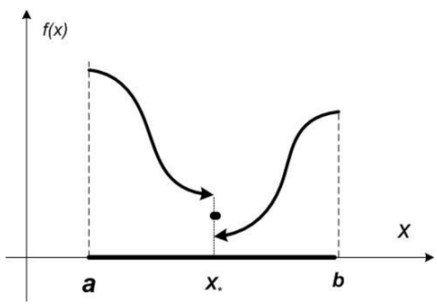
где $x \in \mathbb{R},\ \mathrm{X} = [a, b],\ a < b,\ f(x)$ - непрерывна на $[a, b]$. Данная задача часто встречается в качестве вспомогательной подзадачи во многих численных методах.

Унимодальность

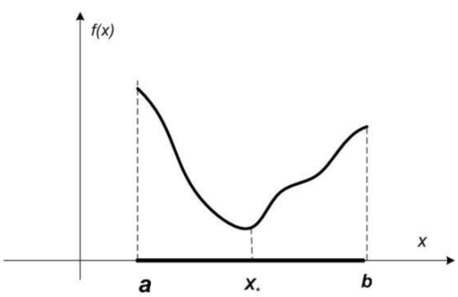
Определение 1. функция $f(x)$ – унимодальная на $[a, b]$, если существует такая точка $x_* \in [a, b]$, что $f(x_1) > f(x_2)$ для любых $a \leq x_1 < x_2 < x_*$ и $f(x_1) < f(x_2)$ для любых $x_* < x_1 < x_2 \leq b$.

Замечание 1.

1. Если на $[a, b]$ унимодальная функция $f(x)$ достигает своего минимума, то она достигает его в точке x_*.
2. Определение для произвольных $f(x)$ не предполагает непрерывности.



3. для непрерывных функций свойство унимодальности означает обязательное наличие у функции $f(x)$ единственного локального минимума на $[a, b]$.



Утверждение 1. Пусть $f(x)$ – унимодальная на отрезке $[a, b]$ функция, достигающего своего минимума в точке $x_* \in [a, b]$. Пусть имеются две точки $x_1 \in [a, b],\ x_2 \in [a, b]$ такие, что $x_1 < x_2$. Тогда

1. если $f(x_1) \le f(x_2)$, то $x_* \in [a, x_2]$,
2. если $f(x_1) \ge f(x_2)$, то $x_* \in [x_1, b]$.

Доказательство. Докажем первый пункт от противного. Пусть $f(x_1) \le f(x_2)$, но $x_* > x_2$. Тогда обязательно $x_1 < x_2 < x_*$ из условий и в силу унимодальности $f(x_1) > f(x_2)$. Получено противоречие. □

Методы последовательной локализации решения

Для решения задач (1) с непрерывными унимодальными функциями являются эффективными **методы последовательной локализации решения**.

Основная идея – построить последовательность вложенных отрезков локализации решения x_* задачи (1):

$$[a, b] \supseteq [a_0, b_0] \supset [a_1, b_2] \supset \dots \supset [a_k, b_k] \supset \dots;$$

$$x_* \in [a_i, b_i] \quad \forall i.$$

В качестве начально отрезка $[a_0, b_0]$ обычно берется сам отрезок $[a, b]$. Длины отрезков $\triangle_k = b_k - a_k$ образуют монотонно убывающую последовательность:

$$\triangle_0 > \triangle_1 > \triangle_2 > \dots > \triangle_k > \dots \to 0.$$

Если построен отрезок $[a_k, b_k]$, то в качестве приближенного решения задачи (1), берется произвольная точка x_k из отрезка $[a_k, b_k]$. Обычно в качестве такой точки целесообразно брать середину отрезка $x_k = \frac{a_k + b_k}{2}$. Это гарантирует, что $|x_k - x_*| \le \frac{\triangle_k}{2}$, а в общем случае $|x_k - x_*| \le \triangle_k$.

Для прерывания процесса задается точность $\varepsilon > 0$, для того, что бы выполнялось следующее неравенство для оценки расстояния между x_* и найденой точки x_k

$$|x_k - x_*| \le \varepsilon.$$

Условие остановки: $\triangle_k \le 2\varepsilon$.

Методы одномерной минимизации

1. метод дихотомии (метод деления отрезка пополам)
2. метод золотого сечения
3. метод Фибоначчи

Рассматривем только унимодальные функции.

Метод дихотомии

Алгоритм

Начальная итерация (k=0)

$$\begin{aligned} \text{полагаем} \quad & a_0 = a \\ & b_0 = b \\ & c_0 = \tfrac{a_0+b_0}{2} \\ \text{вычисляем} \quad & f(c_0) \end{aligned}$$

Общая k-ая итерация

Шаг 1

берем точку $y_k = \frac{a_k+c_k}{2}$ и вычисляем $f(y_k)$

1. если $f(y_k) \leq f(c_k)$, то

$$a_{k+1} = a_k$$
$$b_{k+1} = c_k$$
$$c_{k+1} = y_k$$

переходим на **шаг 3**

2. если $f(y_k) > f(c_k)$, то берем точку $z_k = \frac{c_k+b_k}{2}$ и вычисляем $f(z_k)$

Шаг 2

если $f(c_k) \leq f(z_k)$, то

$$a_{k+1} = y_k$$
$$b_{k+1} = z_k$$
$$c_{k+1} = c_k$$

иначе

$$a_{k+1} = c_k$$
$$b_{k+1} = b_k$$
$$c_{k+1} = z_k$$

Шаг 3

увеличиваем счетчик $k := k + 1$ и переходим на **шаг 1**

Анализ

В силу утверждения 1 мы имеем $x_* \in [a_{k+1}, b_{k+1}]$,

$$\triangle_{k+1} = b_{k+1} - a_{k+1} = \frac{1}{2}(b_k - a_k) = \dots = \frac{1}{2^k}(b - a).$$

На каждой итерации $f(x)$ вычисляется максимум в двух точках, на 0-ой итерации в одной точке.

1. Пусть нам можно выполнить N вычислений $f(x)$ и для удобства пусть N – нечетное, тогда какое число итераций K мы можем выполнить?

$$\frac{N-1}{2} = K$$

$$x_{K+1} = c_{K+1} = \frac{a_{K+1} - b_{K+1}}{2}$$

$$|x_{K+1} - x_*| \le \frac{1}{2}(b_{K+1} - a_{K+1}) = \frac{1}{2^{\frac{N-1}{2}}}\frac{b-a}{2} = \left(\frac{1}{2}\right)^K \frac{b-a}{2}$$

2. Пусть нам нужна точность ε, тогда что мы можем сказать про число итераций K?

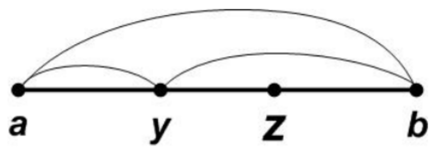
$$|x_{K+1} - x_*| \le \frac{1}{2^K}\frac{b-a}{2} \le \varepsilon$$

$$\left(\frac{1}{2}\right)^{K+1} \le \frac{\varepsilon}{b-a}$$

$$K \ge \log_2 \frac{b-a}{\varepsilon} - 1$$

Метод золотого сечения

Нам заданы отрезок $[a, b]$ и функция $f(x)$. **Золотым сечением** называется такое деление отрезка на две неравные части, что отношение длины всего отрезка к большей части равно отношению большей части к меньшей. Эта пропорция часто встречается в природе и в человеческой деятельности.



1. **Меньшая золотая точка** отрезка $[a, b]$

$$\frac{b-a}{b-y} = \frac{b-y}{y-a}$$

решая данную пропорцию мы получим

$$y = y(a, b) = a + \frac{2}{3+\sqrt{5}}(b-a) \approx a + 0,382(b-a)$$

2. **Большая золотая точка** отрезка $[a, b]$

$$\frac{b-a}{z-a} = \frac{z-a}{b-z}$$

решая данную пропорцию мы получим

$$z = z(a, b) = a + \frac{\sqrt{5}-1}{2}(b-a) \approx a + 0,618(b-a)$$

3. введем величину $\tau = \frac{1+\sqrt{5}}{2}$ и преобразуем

$$y(a, b) = a + \frac{1}{\tau^2}(b-a)$$

$$z(a, b) = a + \frac{1}{\tau}(b-a)$$

Утверждение 2. Пусть y и z - меньшая и большая золотые точки отрезка $[a, b]$. Тогда

1. выполняются следующие равенства

$$z - a = b - y = \frac{\sqrt{5}-1}{2}(b-a) = \frac{1}{\tau}(b-a);$$

2. y - большая золотая точка отрезка $[a, z]$, z - меньшая золотая точка отрезка $[y, b]$.

Алгоритм

Начальная итерация (k=0)

$$\begin{aligned} \text{полагаем} \qquad a_1 &= a \\ b_1 &= b \\ y_1 &= y(a,b) \\ z_1 &= z(a,b) \\ \text{вычисляем} \qquad & f(y_1) \end{aligned}$$

Общая k-ая итерация

Шаг 1

вычисляем то из значений $f(y_k)$, $f(z_k)$, которое еще не вычислено

если $f(y_k) \leq f(z_k)$, то

$$\begin{aligned} a_{k+1} &= a_k \\ b_{k+1} &= z_k \\ z_{k=1} &= y_k \\ y_{k+1} &= y(a_{k+1}, b_{k+1}) \end{aligned}$$

иначе, когда $f(y_k) > f(z_k)$

$$\begin{aligned} a_{k+1} &= y_k \\ b_{k+1} &= b_k \\ z_{k+1} &= z(a_{k+1}, b_{k+1}) \\ y_{k+1} &= z_k \end{aligned}$$

Шаг 2

увеличиваем счетчик $k := k+1$ и переходим на **шаг 1**

Анализ

В силу утверждения 1 и 2 на каждой k-ой итерации мы имеем

1. $x_* \in [a_{k+1}, b_{k+1}]$,
2. $y_{k+1} = y(a_{k+1}, b_{k+1}),\ z_{k+1} = z(a_{k+1}, b_{k+1})$.

Приближение для x_{k+1} – любая точка из отрезка $[a_{k+1}, b_{k+1}]$.

На каждой итерации $f(x)$ вычисляется в одной точке.

$$\triangle_{k+1} = b_{k+1} - a_{k+1} = \frac{1}{\tau}(b_k - a_k) = \left(\frac{1}{\tau}\right)^k (b - a)$$

1. Пусть нам можно выполнить N вычислений $f(x)$, тогда какое число итераций K мы можем выполнить?
$$N - 1 = K$$
$$|x_{K+1} - x_*| \leq \triangle_{K+1} = \left(\frac{1}{\tau}\right)^{N-1} (b - a) \approx 0,618^K (b - a)$$

2. Пусть нам нужна точность ε, тогда что мы можем сказать про число итераций K?
$$|x_{K+1} - x_*| \leq \varepsilon$$
$$\left(\frac{1}{\tau}\right)^K \leq \frac{\varepsilon}{b - a}$$
$$K \geq \frac{\ln \frac{b-a}{\varepsilon}}{\ln \tau}$$

Сравнение методов дихотомии и золотого сечения

- константа геометрической прогрессии метода золотого сечения ($\frac{1}{\tau} = 0,618$) больше, чем у метода дихотомии ($\frac{1}{2} = 0,5$)
- количество вызовов функции у метода золотого сечения меньше, чем у метода дихотомии