Введение в теорию методов оптимизации

Александр Катруца

Сочи 2017

Введение

- Что такое методы оптимизации?
- Какой математический аппарат используется?
- Как развивалась теория методов оптимизации?
- Какие главные результаты?

Обозначения и типы задач

$$\min f(x)$$
$$\text{s.t. } x \in X \subseteq \mathbb{R}^n$$

- Глобальный vs. локальный минимум
- Условная vs.безусловная задача
- Непрерывная vs. дискретная задача
- Детерминированная vs. стохастическая задача

Как сравнивать методы оптимизации?

- Теоретическая сложность — об этом ниже
- Масштабируемость
- Время работы
- Простота понимания и реализации

Общая схема

Математически:

- Поиск направления h_k, чаще всего это направление убывания
- Выбор шага α_k, такого что $f(x_k + \alpha_k h_k) < f(x_k)$
- Проверка критерия остановки

Алгоритмически:

```
while (True):
    h = FindDirection(...)
    alpha = FindStepSize(...)
    x = x + alpha * h
    if StopCriterion(...):
        break
```

Методы первого порядка

Идея: помимо значения функции в точке, использовать значение первой производной.

- Градиентный спуск
- Метод сопряжённых градиентов
- Квазиньютоновские методы

Методы второго порядка

Идея: помимо значения функции в точке и значения первой производной использовать значения второй производной в точке.

- Метод Ньютона

Теория двойственности: идея

Общий принцип

Полезно иметь несколько взглядов на одну сущность

Например:

- множество: перечисление элементов vs. задание ограничений на элементы
- функция: табличное vs. аналитическое определение
- задачи: прямое решение vs. переформулировка или решение вспомогательной задачи
- связь между эллиптическими кривыми и теорией чисел: великая теорема Ферма, криптография, etc

Двойственность в оптимизации

Построение задачи, связанной с данной, которая

- всегда решается легко не зависимо от сложности решения исходной
- даёт нижнюю оценку на оптимальное значение целевой функции исходной задачи

Теория двойственности: конкретика

Задача

$$\min_{x \in \mathcal{D}} f(x) = p^*$$
$$\text{s.t. } g_i(x) = 0,\ i = 1, \ldots, m$$
$$h_j(x) \le 0,\ j = 1, \ldots, p$$

Лагранжиан

$$L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

Двойственные переменные

Вектора $\boldsymbol{\mu}$ и $\boldsymbol{\lambda}$ называются двойственными переменными.

Двойственная функция

Функция $g(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \inf\limits_{x \in \mathcal{D}} L(x, \boldsymbol{\lambda}, \boldsymbol{\mu})$ называется двойственной функцией Лагранжа.

Свойства двойственной функции

Вогнутость

Двойственная функция является вогнутой как инфимум аффинных функций по $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ вне зависимости от того, является ли исходная задача выпуклой.

Нижняя граница

Для любого $\boldsymbol{\lambda}$ и для $\boldsymbol{\mu} \geq 0$ выполнено $g(\boldsymbol{\mu}, \boldsymbol{\lambda}) \leq p^*$.

Двойственная задача

$$\max g(\boldsymbol{\mu}, \boldsymbol{\lambda}) = d^*$$

$$\text{s.t. } \boldsymbol{\mu} \geq 0$$

Зачем?

- Двойственная задача выпукла независимо от того, выпукла ли прямая
- Нижняя оценка **может** достигаться

Главный вопрос

~~Быть или не быть?~~

Какой метод оптимизации лучше для данной задачи?

Возможные ответы:

- экспериментально сравнить несколько методов
- определить к какому классу задач относится данная и использовать наилучший метод для этого класса задач

Оптимальные методы

Заключение

- Зачем нужны методы оптимизации?
- Итеративная схема
- Классификация методов и их анализ
- Теория двойственности
- Элементы теории сложности