Введение. Выпуклые множества и выпуклые функции

Александр Катруца

Московский физико-технический институт,
Факультет Управления и Прикладной Математики

3 октября 2017 г.

Зачем этот курс?

- Формализация задачи выбора элемента из множества
- Обоснование правильности принятия решения
- Разнообразные приложения:
  - машинное обучение: классификация, кластеризация, регрессия
  - молекулярное моделирование
  - анализ рисков
  - выбор активов (portfolio optimization)
  - оптимальное управление
  - обработка сигналов
  - оценка параметров в статистике
  - и другие[1]

[1] http://www.cvxpy.org/en/latest/examples/index.html

О чём этот курс?

- Основы выпуклого анализа
- Теория двойственности
- Условия оптимальности
- Методы безусловной минимизации первого и второго порядка
- Методы условной оптимизации
- Оптимальные методы
- ...

План на семестр

- Семинар-лекция раз в неделю
- Промежуточная контрольная в середине семестра
- Итоговая контрольная в конце семестра
- Домашние задания в течение семестра

Предварительные навыки

- Линейная алгебра
- Математический анализ
- Программирование: Python (NumPy, SciPy, CVXPY) или MATLAB
- Элементы вычислительной математики

Методология

Основные этапы использования методов оптимизации при решении реальных задач:

1. Определение целевой функции
2. Определение допустимого множества решений
3. Постановка и анализ оптимизационной задачи
4. Выбор наилучшего алгоритма для решения поставленной задачи
5. Реализация алгоритма и проверка его корректности

Постановка задачи

$$\min_{\mathbf{x} \in X} f_0(\mathbf{x})$$
$$\text{s.t. } f_i(\mathbf{x}) = 0, \; i = 1, \ldots, p$$
$$f_j(\mathbf{x}) \leq 0, \; j = n+1, \ldots, m,$$

- $\mathbf{x} \in \mathbb{R}^n$ — искомый вектор
- $f_0(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ — целевая функция
- $f_k(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ — функции ограничений

Пример: выбор объектов для вложения денег и определение в какой объект сколько вкладывать

- $\mathbf{x}$ — размер инвестиций в каждый актив
- f_0 — суммарный риск или вариация прибыли
- f_k — бюджетные ограничения, min/max вложения в актив, минимально допустимая прибыль

Как решать?

В общем случае:

- NP-полные
- рандомизированные алгоритмы: время vs стабильность

НО определённые классы задач могут быть решены быстро!

- Линейное программирование
- Метод наименьших квадратов
- Малоранговое приближение порядка k
- Выпуклая оптимизация

Линейное программирование

$$
\begin{aligned}
&\min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{c}^\mathsf{T} \mathbf{x} \\
&\text{s.t. } \mathbf{a}_i^\mathsf{T} \mathbf{x} \leq c_i, \;\; i = 1, \dots, m
\end{aligned}
\tag{1}
$$

- нет аналитического решения
- существуют эффективные алгоритмы
- разработанная технология
- симплекс-метод для решения задачи (1) входит в Top-10 алгоритмов XX века[2]

[2] https://www.siam.org/pdf/news/637.pdf

Метод наименьших квадратов

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2, \qquad (2)$$

где $\mathbf{A} \in \mathbb{R}^{m\times n}$ и $\mathbf{b} \in \mathbb{R}^m$.

- имеет аналитическое решение: $\mathbf{x}^* = (\mathbf{A}^\intercal\mathbf{A})^{-1}\mathbf{A}^\intercal\mathbf{b}$
- существуют эффективные алгоритмы
- разработанная технология
- имеет статистическую интерпретацию

Малоранговое приближение ранга k

$$\min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \|\mathbf{A} - \mathbf{X}\|_F \quad \text{s.t. } \operatorname{rank}(\mathbf{X}) \leq k \tag{3}$$

Theorem (Eckart–Young, 1993)

Пусть $\mathbf{A} = \mathbf{U\Sigma V}^\intercal$ *— сингулярное разложение матрицы* $\mathbf{A}$, *где* $\mathbf{U} = [\mathbf{U}_k, \mathbf{U}_{r-k}] \in \mathbb{R}^{m \times r}$, $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_k, \ldots, \sigma_r)$, $\mathbf{V} = [\mathbf{V}_k, \mathbf{V}_{r-k}] \in \mathbb{R}^{n \times r}$ *и* $r = rank(\mathbf{A})$. *Тогда решение задачи* (3) *можно записать в виде:*

$$\mathbf{X} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^\intercal,$$

где $\hat{\mathbf{U}} \in \mathbb{R}^{m \times k}$, $\hat{\mathbf{\Sigma}} = \mathrm{diag}(\sigma_1, \ldots, \sigma_k)$, $\hat{\mathbf{V}} \in \mathbb{R}^{n \times k}$.

Алгоритм вычисления сингулярного разложения и быстрый, и устойчивый.

Выпуклая оптимизация

$$\min_{\mathbf{x}\in\mathbb{R}^n} f_0(\mathbf{x})$$
$$\text{s.t. } f_i(\mathbf{x}) \le b_i, \; i = 1, \ldots, m \qquad (4)$$

- f_0, f_i — выпуклые функции:

  $$f(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2) \le \alpha f(\mathbf{x}_1) + \beta f(\mathbf{x}_2),$$

  где $\alpha, \beta \ge 0$ и $\alpha + \beta = 1$.
- нет аналитического решения
- существуют эффективные алгоритмы
- часто сложно «увидеть» задачу выпуклой оптимизации
- существуют приёмы для преобразования задачи к виду (4)

Какая задача проще?

Поиск независимого под-алфавит максимальной мощности

$$\max_{\mathbf{x}\in\mathbb{R}^n} \sum_{i=1}^{n} x_i$$

$$\text{s.t. } x_i^2 - x_i = 0 \quad i = 1, \dots, n$$

$$x_i x_j = 0 \qquad \forall (i,j) \in \Gamma,$$

где Γ — множество пар

Truss design

$$\min -2 \sum_{i=1}^{k} \sum_{j=1}^{m} c_{ij} x_{ij} + x_{00}$$

$$\text{s.t. } \sum_{i=1}^{k} x_i = 1$$

$$\lambda_{\min}(\mathbf{A}) \geq 0,$$

где $\mathbf{A} =$

$$\begin{bmatrix} x_1 & \dots & \dots & \sum\limits_{j=1}^{m} b_{pj} x_{1j} \\ \vdots & \ddots & & \vdots \\ & & x_k & \sum\limits_{j=1}^{m} b_{pj} x_{kj} \\ \sum\limits_{j=1}^{m} b_{pj} x_{1j} & \dots & \sum\limits_{j=1}^{m} b_{pj} x_{kj} & x_{00} \end{bmatrix}$$

Почему выпуклость так важна?

R. Tyrrell Rockafellar (1935 —)

The great watershed in optimization is not between linearity and non-linearity, but convexity and non-convexity.

- Локальный оптимум является глобальным
- Необходимое условие оптимальности является достаточным

Вопросы:

- Любую ли задачу выпуклой оптимизации можно эффективно решить?
- Можно ли эффективно решить невыпуклые задачи оптимизации?

Выпуклое множество

Выпуклое множество

Множество C называется выпуклым, если
$$\forall x_1,\ x_2 \in C, \theta \in [0,1] \to \theta x_1 + (1-\theta)x_2 \in C.$$
$\emptyset$ и $\{x_0\}$ также считаются выпуклыми.

Примеры: $\mathbb{R}^n$, аффинное множество, луч, отрезок.

Выпуклая комбинация точек

Пусть $x_1, \ldots, x_k \in G$, тогда точка $\theta_1 x_1 + \ldots + \theta_k x_k$ при $\sum\limits_{i=1}^{k} \theta_i = 1,\ \theta_i \geq 0$ называется выпуклой комбинацией точек $x_1, \ldots, x_k$.

Выпуклая оболочка точек

Множество $\left\{ \sum\limits_{i=1}^{k} \theta_i x_i \mid x_i \in G, \sum\limits_{i=1}^{k} \theta_i = 1, \theta_i \geq 0 \right\}$ называется выпуклой оболочкой множества G и обозначается $\textbf{conv}(G)$.

Операции, сохраняющие выпуклость

- Пересечение любого (конечного или бесконечного) числа выпуклых множеств — выпуклое множество
- Образ аффинного отображения выпуклого множества — выпуклое множество
- Линейная комбинация выпуклых множеств — выпуклое множество
- Декартово произведение выпуклых множеств — выпуклое множество

Примеры

Проверьте на аффинность и выпуклость следующие множества:

1. Полупространство: $\{\mathbf{x} | \mathbf{a}^\top \mathbf{x} \leq c\}$
2. Многоугольник: $\{\mathbf{x} | \mathbf{A}\mathbf{x} \preceq \mathbf{b},\ \mathbf{C}\mathbf{x} = 0\}$
3. Шар по норме в $\mathbb{R}^n$: $B(r, x_c) = \{x \mid \|x - x_c\| \leq r\}$
4. Эллипсоид:
   $\mathcal{E}(x_c, \mathbf{P}, r) = \{x \mid (x - x_c)^\top \mathbf{P}^{-1}(x - x_c) \leq r^2\}$
5. Множество симметричных положительно-определённых матриц:
   $\mathbf{S}_+^n = \{\mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathbf{X}^\top = \mathbf{X},\ \mathbf{X} \succeq 0\}$
6. $\{\mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathrm{Tr}(\mathbf{X}) = const\}$
7. Гиперболическое множество: $\{\mathbf{x} \in \mathbb{R}_+^n \mid \prod\limits_{i=1}^{n} x_i \geq 1\}$

Определения функций

Выпуклая функция

Функция $f : X \subset \mathbb{R}^n \to \mathbb{R}$ называется выпуклой (строго выпуклой), если X — выпуклое множество и для $\forall \mathbf{x}_1, \mathbf{x}_2 \in X$ и $\alpha \in [0, 1]$ ($\alpha \in (0, 1)$) выполнено:

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \ (<) \ \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2)$$

Вогнутая функция

Функция f вогнутая (строго вогнутая), если $-f$ выпуклая (строго выпуклая).

Сильно выпуклая функция

Функция $f : X \subset \mathbb{R}^n \to \mathbb{R}$ называется сильно выпуклой с константой $m > 0$, если X — выпуклое множество и для $\forall \mathbf{x}_1, \mathbf{x}_2 \in X$ и $\alpha \in [0, 1]$ выполнено:

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) - \frac{m}{2}\alpha(1-\alpha)\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2$$

Определения множеств

Надграфик (эпиграф)

Надграфиком функции f называется множество
$\mathrm{epi}f = \{(\mathbf{x}, y) : \mathbf{x} \in \mathbb{R}^n,\ y \in \mathbb{R},\ y \geq f(\mathbf{x})\} \subset \mathbb{R}^{n+1}$

Множество подуровней (множество Лебега)

Множество подуровня функции f называется следующее множество

$$C_\gamma = \{\mathbf{x} | f(\mathbf{x}) \leq \gamma\}.$$

Замкнутая функция

Функция f называется замкнутой, если её надграфик замкнутое множество.

Квазивыпуклая функция

Функция f называется квазивыпуклой, если её область определения и множество подуровней выпуклое множество.

Критерии выпуклости

Дифференциальный критерий первого порядка

Функция f выпукла тогда и только тогда, когда определена на выпуклом множестве X и $\forall \mathbf{x}, \mathbf{y} \in X \subset \mathbb{R}^n$ выполнено:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f^\mathsf{T}(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

Дифференциальный критерий второго порядка

Непрерывная и дважды дифференцируемая функция f выпукла тогда и только тогда, когда определена на выпуклом множестве X и $\forall \mathbf{x}, \mathbf{y} \in \mathrm{ri}(X) \subset \mathbb{R}^n$ выполнено:

$$\nabla^2 f(\mathbf{x}) \succeq 0.$$

Связь с надграфиком

Функция выпукла тогда и только тогда, когда её надграфик выпуклое множество.

Ограничение на прямую

Функция $f : X \to \mathbb{R}$ выпукла тогда и только тогда, когда X — выпуклое множество и выпукла функция $g(t) = f(\mathbf{x} + t\mathbf{v})$ на множестве $\{t | \mathbf{x} + t\mathbf{v} \in X\}$ для всех $\mathbf{x} \in X$ и $\mathbf{v} \in \mathbb{R}^n$.

Критерии сильной выпуклости

Дифференциальный критерий первого порядка

Функция f сильно выпукла с константой m тогда и только тогда, когда определена на выпуклом множестве X и $\forall \mathbf{x}, \mathbf{y} \in X \subset \mathbb{R}^n$ выполнена:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f^\top(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{m}{2}\|\mathbf{y} - \mathbf{x}\|^2$$

Дифференциальный критерий второго порядка

Непрерывная и дважды дифференцируемая функция f сильно выпукла с константой m тогда и только тогда, когда она определена на выпуклом множестве X и $\forall \mathbf{x} \in \mathbf{relint}(X) \subset \mathbb{R}^n$ выполнено:

$$\nabla^2 f(\mathbf{x}) \succeq m\mathbf{I}.$$

Примеры

1. Квадратичная функция: $f(x) = \frac{1}{2}\mathbf{x}^\top \mathbf{P}\mathbf{x} + \mathbf{q}^\top \mathbf{x} + r$, $\mathbf{x} \in \mathbb{R}^n$
2. Нормы в $\mathbb{R}^n$
3. $f(\mathbf{x}) = \log\left(e^{x_1} + \ldots + e^{x_n}\right)$, $\mathbf{x} \in \mathbb{R}^n$
4. Логарифм детерминанта: $f(\mathbf{X}) = -\log\det \mathbf{X}$, $\mathbf{X} \in \mathbf{S}^n_{++}$
5. Множество выпуклых функций — выпуклый конус
6. Поэлементный максимум: $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$, $\mathrm{dom}\ f = \mathrm{dom}\ f_1 \cap \mathrm{dom}\ f_2$
7. Расширение на бесконечное множество функций: если для $\mathbf{y} \in \mathcal{A}$ функция $f(\mathbf{x}, \mathbf{y})$ выпуклая функция по $\mathbf{x}$, тогда $\sup\limits_{\mathbf{y}\in\mathcal{A}} f(\mathbf{x}, \mathbf{y})$ выпукла по $\mathbf{x}$
8. Максимальное собственное значение: $f(\mathbf{X}) = \lambda_{\max}(\mathbf{X}) = \sup\{\mathbf{y}^\top \mathbf{X}\mathbf{y} \mid \|\mathbf{y}\|_2 = 1\}$

Неравенство Йенсена

Неравенство Йенсена

Для выпуклой функции f выполнено следующее неравенство:

$$f\left(\sum_{i=1}^{n} \alpha_i \mathbf{x}_i\right) \leq \sum_{i=1}^{n} \alpha_i f(\mathbf{x}_i),$$

где $\alpha_i \geq 0$ и $\sum\limits_{i=1}^{n} \alpha_i = 1$.

или в бесконечномерном случае: $p(x) \geq 0$ и $\int\limits_X p(x) = 1$

$$f\left(\int\limits_X p(x) x dx\right) \leq \int\limits_X f(x) p(x) dx$$

при условии, что интегралы существуют.

Примеры

1. Неравенство Гёльдера
2. Неравенство о среднем арифметическом и среднем геометрическом
3. $f(\mathbf{E}(x)) \leq \mathbf{E}(f(x))$
4. Выпуклость множества $\{\mathbf{x} | \prod_{i=1}^{n} x_i \geq 1\}$

Резюме

- Организация работы
- Предмет курса по оптимизации
- Общие факты об оптимизации
- Выпуклые множества
- Выпуклые функции