最优化方法

张伯雷

南京邮电大学 计算机学院、通达学院

https://bolei-zhang.github.io/course/opt.html

bolei.zhang@njupt.edu.cn

课程简介

- 课程目录
  - 凸优化
    - 判定问题的性质：线性规划、凸规划、非凸规划
    - 最优性理论
    - 算法
  - 近似优化
  - 博弈优化
  - 随机优化
- 预备课程
  - 高等数学，线性代数，概率论

本门课程授课内容

1. 最优化概述（2课时）
2. 凸集与凸函数（4课时）
3. 凸优化问题（2课时）
4. 最优性理论（6课时）
5. 梯度类方法求解凸优化问题（2课时）
6. 组合优化（4课时）
7. 博弈优化（2课时）
8. 随机优化（2课时）

课时计划表与学时分配

学时：32=24(理论课)+8(4次实验)，2学分

- 最优化概述（2课时）
- 凸优化（14课时）
- 组合优化（4课时）
- 博弈优化（2课时）
- 随机优化（2课时）

答疑：每周五计算机楼535，联系QQ：380101771

教材与参考书

教材

1. Stephen Boyd, Lieven Vandenberghe. Convex Optimization[M]. Cambridge University Press, 2004.
2. Vazirani V V. Approximation algorithms. Berlin: springer, 2001.
3. Nisan, Noam , et al. "Algorithmic Game Theory." 2007.
4. Randomized Algorithm

课程评价

- 学时：32=12(理论课)*2+4(实验)*2
- 考试： 开卷 成绩=期末*50%+平时*50%
- 平时成绩：
  - 考勤+提问：10%
  - 作业：20%
  - 上机实验：20%

上机实验

- 4次实验初步安排
  - 实验一：cvxopt使用与优化方法
    - 第5周（周五），第6-7节：计算机学科楼（通达：待定）
  - 实验二：机器学习优化算法
    - 第8周（周五），第6-7节：计算机学科楼（通达：待定）
  - 实验三：近似优化
    - 第11周（周五），第6-7节：计算机学科楼（通达：待定）
  - 实验四：优化建模
    - 第14周（周五），第6-7节：计算机学科楼（通达：待定）
- 实验要求
  - 提前预习、设计代码
  - 独立完成实验（编码、调试、运行成功、检查）
  - 每次实验一份电子档报告，不交纸质
  - 实验成绩根据实验准备、运行、报告等综合打分

最优化方法概论

张伯雷

南京邮电大学 计算机学院、通达学院

https://bolei-zhang.github.io/course/opt.html

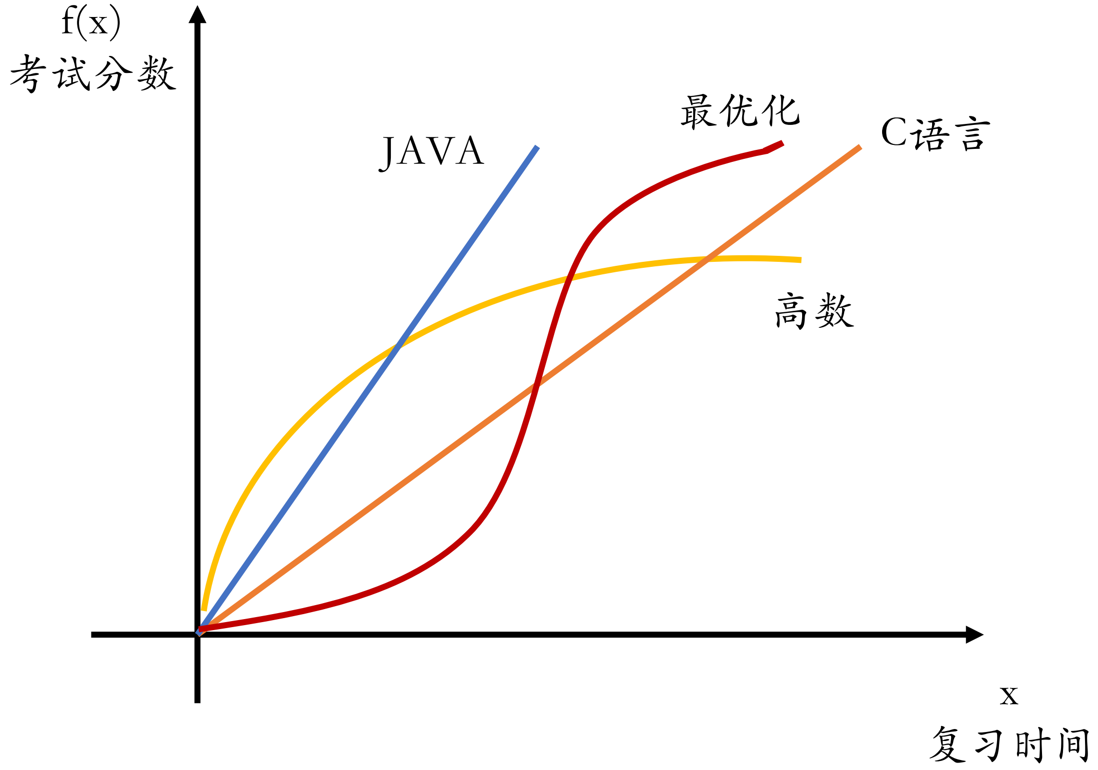
bolei.zhang@njupt.edu.cn

目 录

- **优化问题介绍**
- 优化问题分类
- 典型优化问题
- 求解优化问题

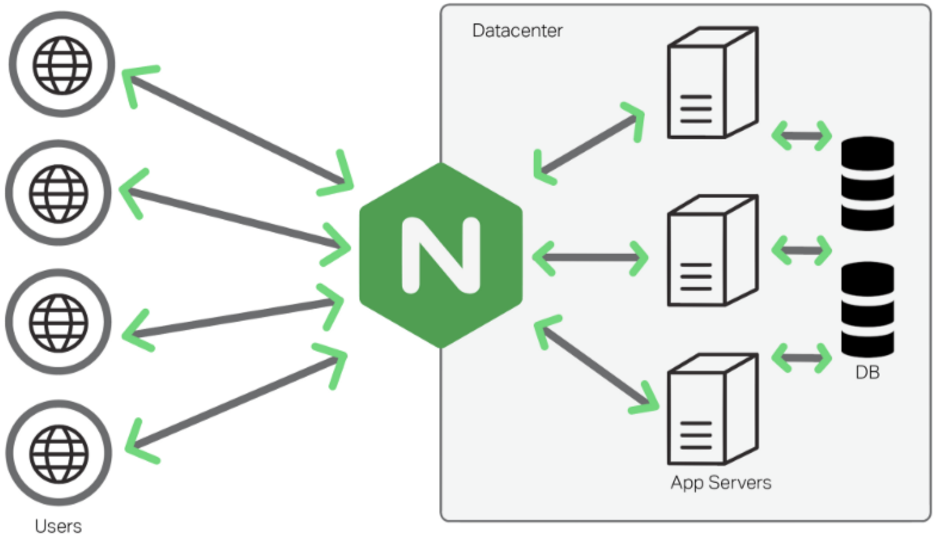
优化问题1

- 还有一个星期就要到期末考试，小明一共要考4门课，但只能复习70个小时，假设$x = (x_1, x_2, \ldots, x_4)$代表小明分别花在4门课上的时间
  - $f_i(x), i = 1,2,3,4$ 为小明在第i门课中的成绩
- 问题建模：
  - 最大化最终的总分数：$\sum_i f_i(x)$
  - 每门课都需要及格，同时花的时间不能超过70：
  - $\sum_i x_i \le 70, x_i \ge 0, f_i(x_i) \ge 60$
- 如何规划复习时间？



优化问题2

- 小李是一个程序员，需要实现一个数据中心服务，当有用户请求服务时，需要尽快地服务用户请求，来降低整体延迟
  - $d_i(x_i), i = 1, 2, \dots, n$
  - x_i为向用户i分配的资源（如CPU的配额），d_i为第i个用户的延迟
- 问题建模：
  - 最小化整体的延迟：$\sum_i d_i(x_i)$
  - 同时不能超过服务器的总资源：$\sum_i x_i \le \mathrm{C}, x_i \ge 0$



- 如何合理分配资源？

优化问题3

- 现在有一百万张图片，一共分成了1000个类别p，我们设计了一个复杂的神经网络，可以输入图片矩阵，输出类别概率q。为了训练这个神经网络，需要优化该网络，最小化预测误差。

- 问题建模：

$$\min -\sum_{x\in X} p(x)\log q(x)$$

- 神经网络可能包括上亿的参数，如何快速优化？

应用领域

- 网络优化
- 大规模电路设计
- 机器学习
- 推荐系统
- 物流运输、路线规划
- 工业设计、工业控制
- 金融经济

优化/数学规划（Optimization/Mathematical Programming）

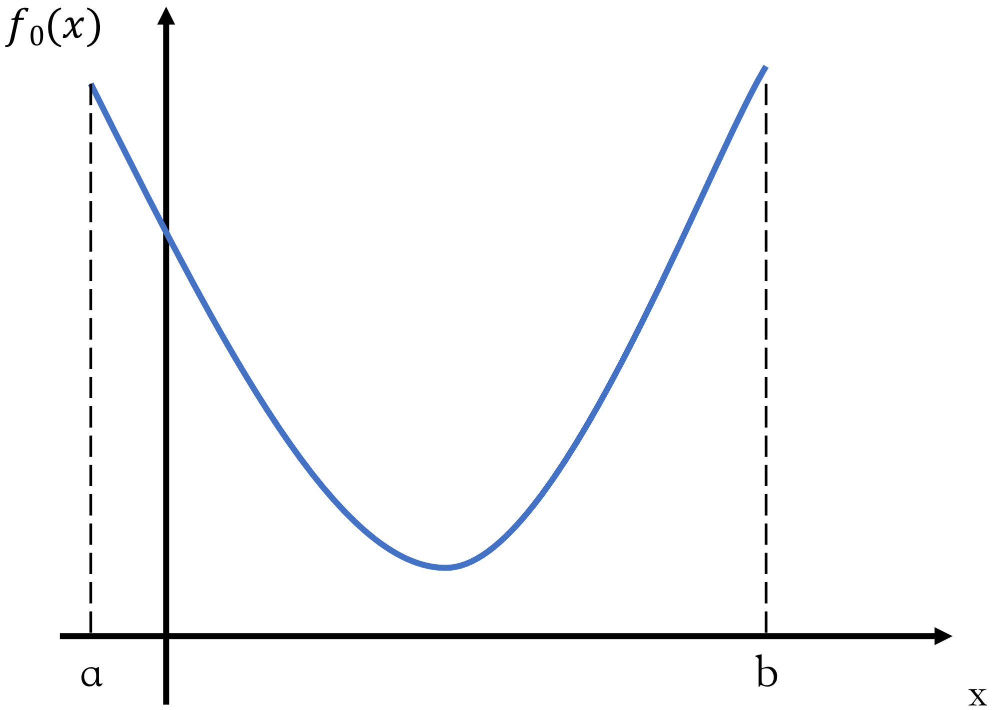
- 优化问题

$$\min \ f_0(x).$$
$$\text{s.t.} \quad f_i(x) \le 0, \qquad i = 1, \dots, m$$

- 其中
  - $x = (x_1, x_2, \dots, x_n)$： 优化变量(variable)
  - f_0: $\mathbb{R}^n \to \mathbb{R}$： 目标函数(Objective function)
  - f_i: $\mathbb{R}^n \to \mathbb{R}, i = 1, \dots, m$： 约束函数(Constraints)
  - 最优解x^*： 所有在定义域且满足约束条件的解中使得目标函数f_0最小的x
  - 最优值p^*： $f_0(x^*)$

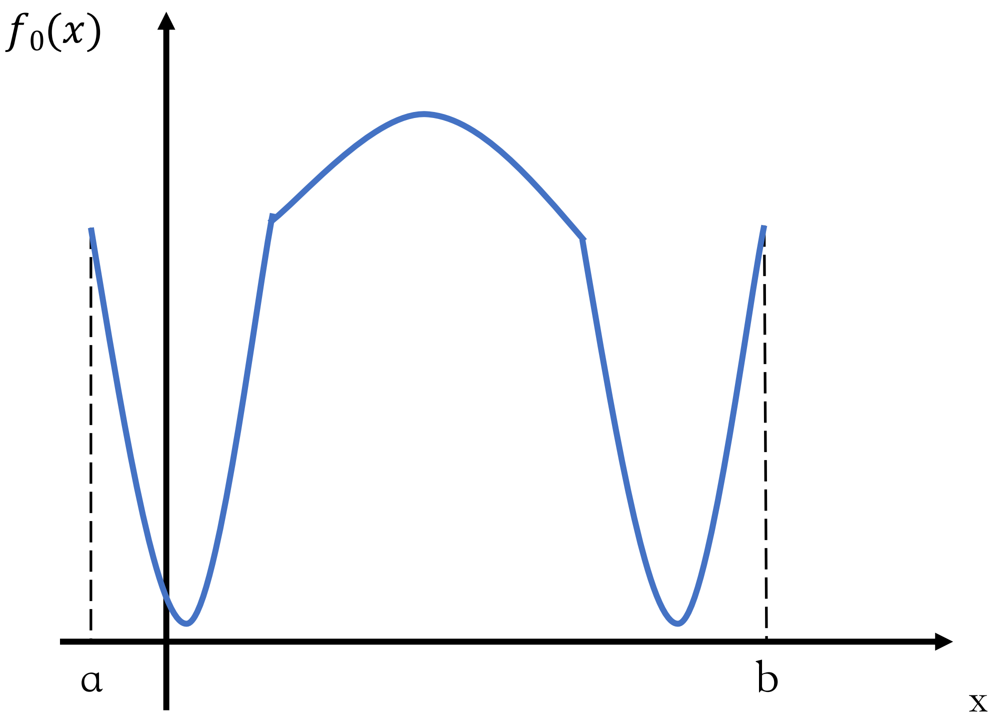
例1

$$f_0(x) = x^2 - 2x + 3$$



- 变量(variable)
- 目标函数(Objective function)
- 约束函数(Constraints)
- 最优解
- 最优值

例2



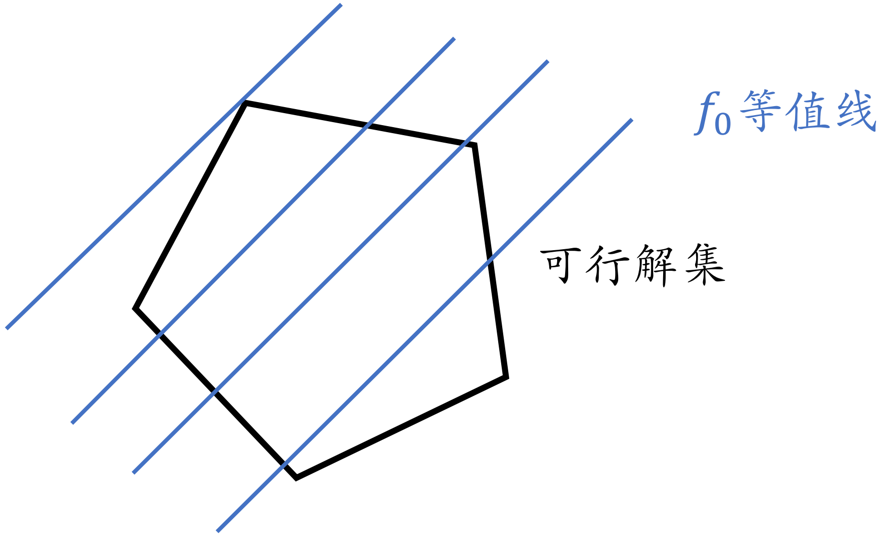
- 变量(variable)
- 目标函数(Objective function)
- 约束函数(Constraints)
- 最优解
- 最优值

目录

- 优化问题介绍
- **优化问题分类**
- 典型优化问题
- 求解优化问题

最优化问题分类：线性/非线性规划

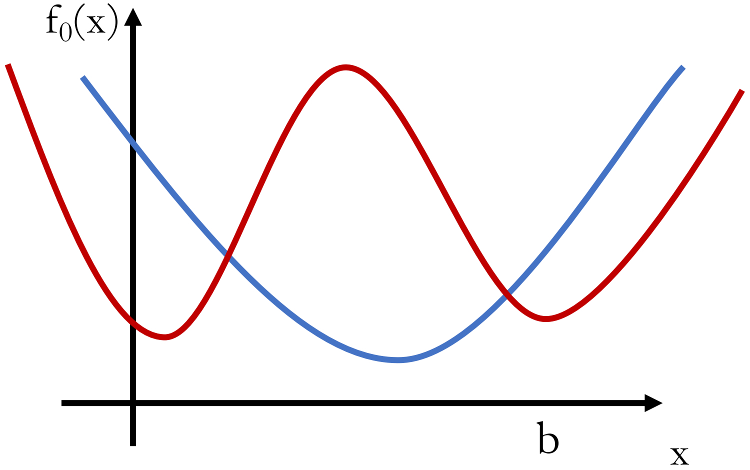
- 线性规划
  - 如果目标函数与约束函数都是线性的，则为线性规划（Linear programming)
  - $f_i(\alpha x+\beta y)=\alpha f_i(x)+\beta f_i(y), i=0,1,\dots,m$
  - 单纯形法(simplex method)



- 非线性规划
  - 如果目标函数与约束函数有一个不是线性的，则为非线性规划（Non-Linear programming)

最优化问题分类：凸/非凸规划

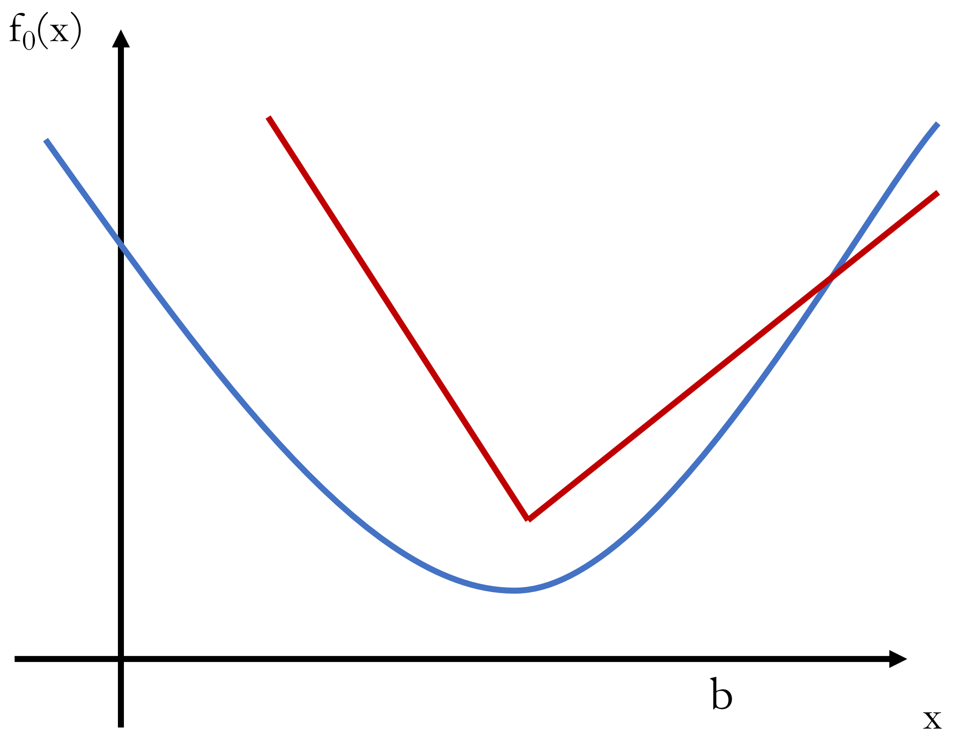
- 凸规划
  - 可行解集是凸集，目标函数是凸函数
  - $f_i(\alpha x+\beta y) \leq \alpha f_i(x)+\beta f_i(y), i=0,1,\dots,m$



- 非凸规划
  - 可行解集不是凸集或目标函数不是凸函数

最优化问题分类：光滑/非光滑规划

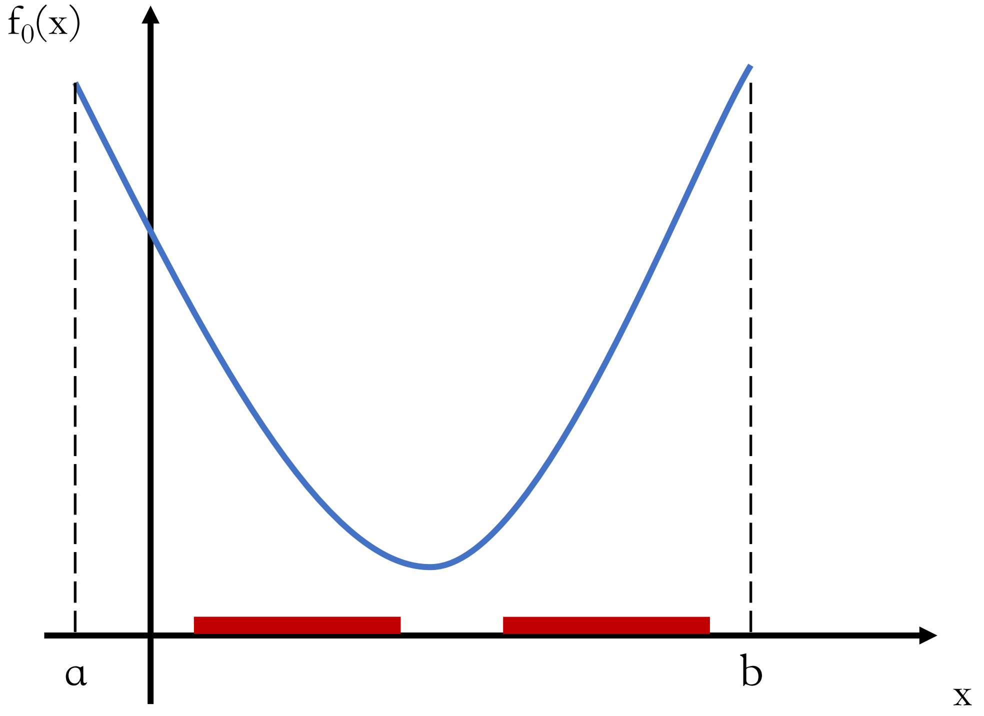
- 光滑规划
  - 如果目标函数$f_0(x)$是光滑的，
  - 光滑：在函数的每一点都可微



- 非光滑规划

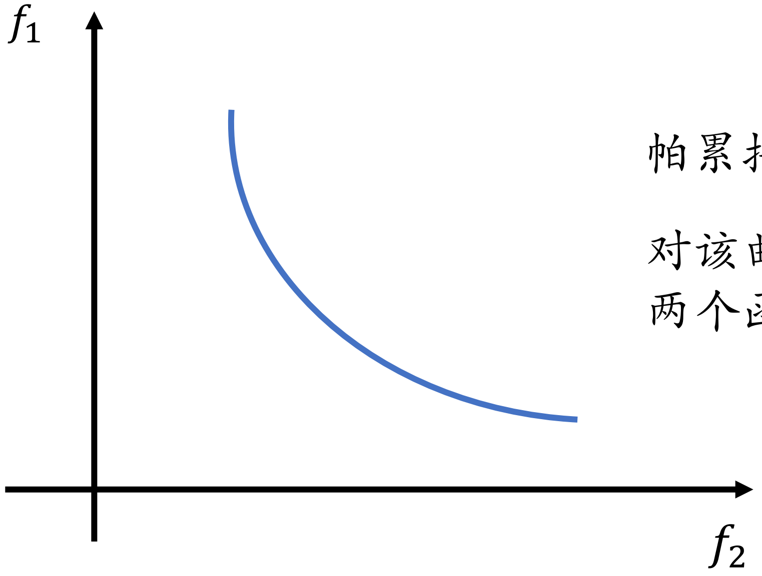
最优化问题分类：连续/离散规划

- 连续规划
  - 可行解集是一个连续的域
- 离散规划
  - 可行解集是离散的



最优化问题分类：单目标/多目标规划

- 多目标规划
  - 同时有多个需要优化的目标
  - $\min f_1(x), f_2(x)$
  - 可以通过加权转化为单目标优化问题



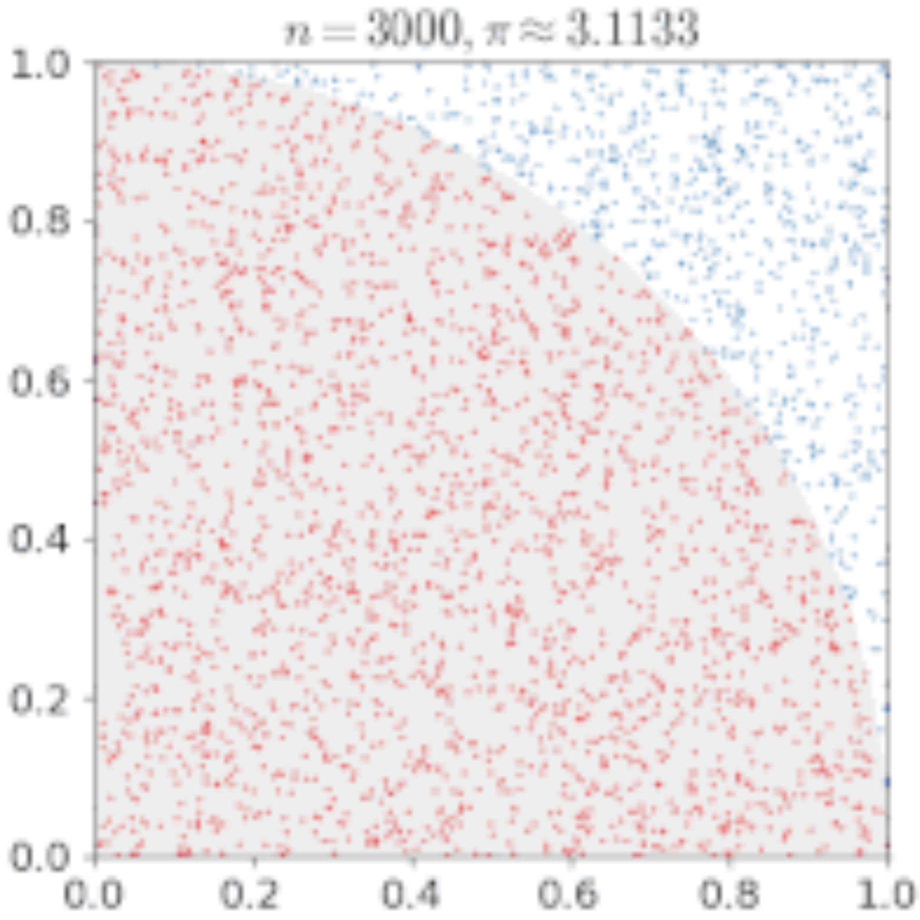
帕累托曲面

对该曲面的任何一点，无法找到其他点，使得其在两个函数上同时可以达到更小

最优化问题分类：博弈优化

Prisoner B / Prisoner A	Prisoner B stays silent (*cooperates*)	Prisoner B testifies (*defects*)
Prisoner A stays silent (*cooperates*)	Each serve 1 year	Prisoner A: 3 years Prisoner B: goes free
Prisoner A testifies (*defects*)	Prisoner A: goes free Prisoner B: 3 years	Each serve 2 years

最优化问题分类：确定/随机优化

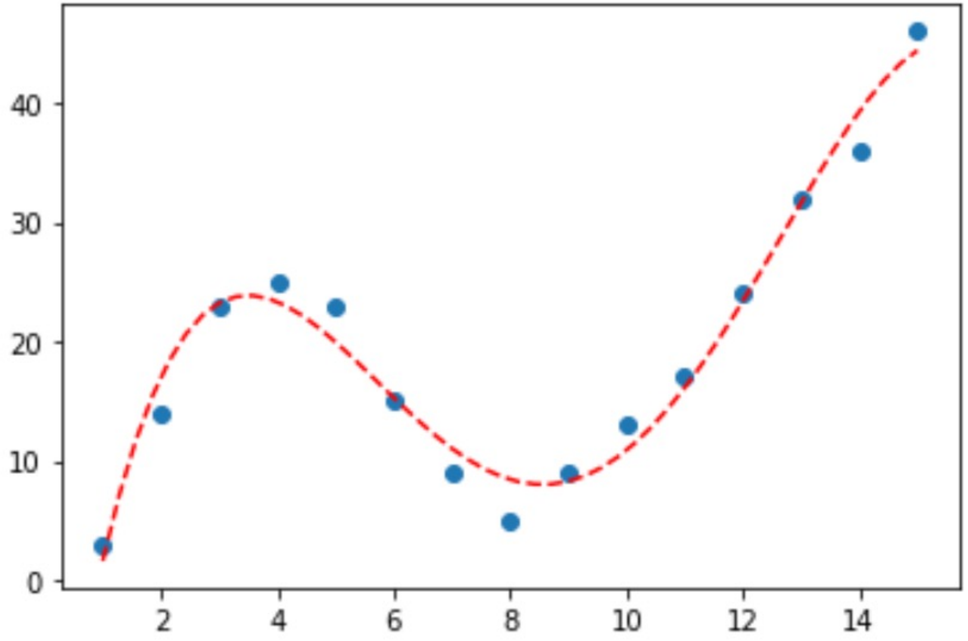


目 录

- 优化问题介绍
- 优化问题分类
- **典型优化问题**
- 求解优化问题

典型优化问题-最小二乘

- 数据拟合问题



- 目标函数

$$\min \sum_{i=1}^{N} \left(y_i - f(x_i)\right)^2$$

典型优化问题-稀疏优化

- 考虑线性方程组求解问题

$$Ax = b$$

其中$x \in R^n, b \in R^m, A \in R^{m \times n}$

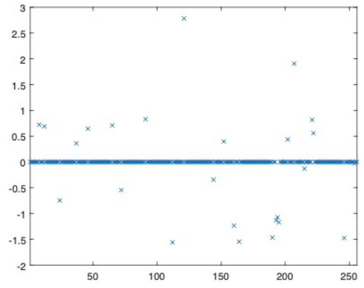
向量b的维度远小于向量x的维度，即$m \ll n$

  - 在信号传输过程中，希望通过接收到长度为m的数字信号b精确地重构原始信号x
  - 方程组是欠定的，因此存在无穷多个解，重构出原始信号看似很难

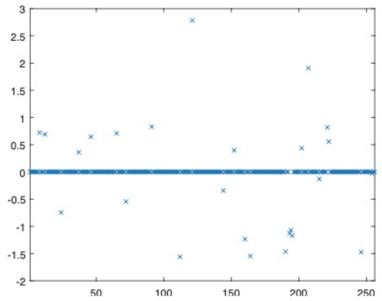
- 稀疏先验：精确解u只有10%的元素非零，即u是如下 l_0 范数问题的最优解:

$$\min_{x \in \mathbb{R}^n} \quad \|x\|_0, \qquad \text{s.t.} \quad Ax = b. \qquad \Rightarrow \qquad \min_{x \in \mathbb{R}^n} \quad \|x\|_1, \qquad \text{s.t.} \quad Ax = b.$$

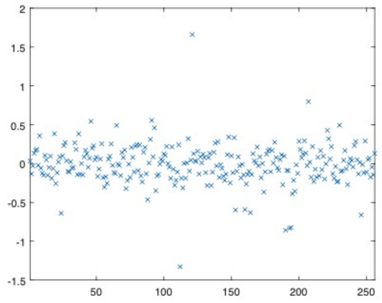
  - l_0范数优化问题是 NP 难问题
  - 但l_1范数优化问题的解可以非常容易地通过现有优化算法得到
  - 若A,b满足一定的条件，向量u也是l_1范数优化问题的唯一最优解



l_1



l_0 (a) 精确解 u



l_2

典型优化问题-低秩矩阵恢复

- 低秩矩阵恢复
  - 某视频网站提供了约 48 万用户对 1 万 7 千多部电影的上亿条评级数据，希望对用户的电影评级进行预测，从而改进用户电影推荐系统，为每个用户更有针对性地推荐影片
  - 用矩阵M代表用户评分，每一行表示不同用户，每一列表示不同电影

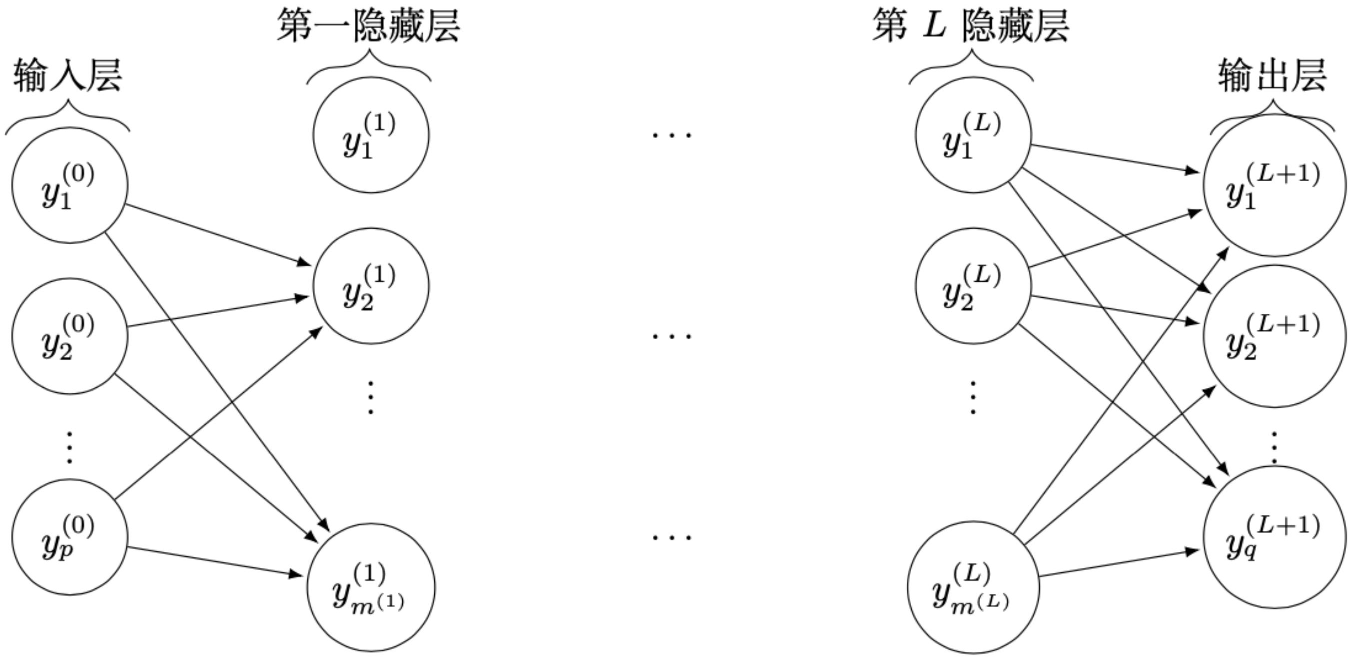
$$
\begin{array}{c} \\ \text{用户 1} \\ \text{用户 2} \\ \text{用户 3} \\ \text{用户 4} \\ \vdots \\ \text{用户m} \end{array}
\begin{array}{c}
\begin{array}{cccccc} \text{电影 1} & \text{电影 2} & \text{电影 3} & \text{电影 4} & \cdots & \text{电影n} \end{array} \\
\begin{bmatrix}
4 & ? & ? & 3 & \cdots & ? \\
? & 2 & 4 & ? & \cdots & ? \\
3 & ? & ? & ? & \cdots & ? \\
2 & ? & 5 & ? & \cdots & ? \\
\vdots & \vdots & \vdots & \vdots & & \vdots \\
? & 3 & ? & 4 & \cdots & ?
\end{bmatrix}
\end{array}
$$

$$
\begin{aligned}
\min_{X\in\mathbb{R}^{m\times n}} \quad & \operatorname{rank}(X), \\
\text{s.t.} \quad & X_{ij} = M_{ij}, \ (i,j)\in\Omega.
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
\min_{X\in\mathbb{R}^{m\times n}} \quad & \|X\|_*, \\
\text{s.t.} \quad & X_{ij} = M_{ij}, \ (i,j)\in\Omega.
\end{aligned}
$$

NP-hard　　　　核范数(nuclear norm)

典型优化问题-深度学习

- 深度学习分类问题



$$y_i^{(l)} = t(z_i^{(l)}), \quad z_i^{(l)} = \sum_{k=1}^{m^{(l-1)}} x_{i,k}^{(l)} y_k^{(l-1)}.$$

$$t(z) = \frac{1}{1+\exp(-z)},$$

$$\min_x \quad \sum_{i=1}^{m} \|h(a_i; x) - b_i\|_2^2 + \lambda r(x),$$

目 录

- 优化问题介绍
- 优化问题分类
- 典型优化问题
- **求解优化问题**

优化问题

对有限资源进行有效分配和控制，并达到某种意义上的最优。

通常需要

- 对需求进行定性和定量分析
- 建立恰当的**数学模型**来描述该问题
- 探索研究模型和算法的**理论性质**，考察算法的**计算性能**等
- 设计合适的**计算方法**来寻找问题的最优解

最优化方法求解

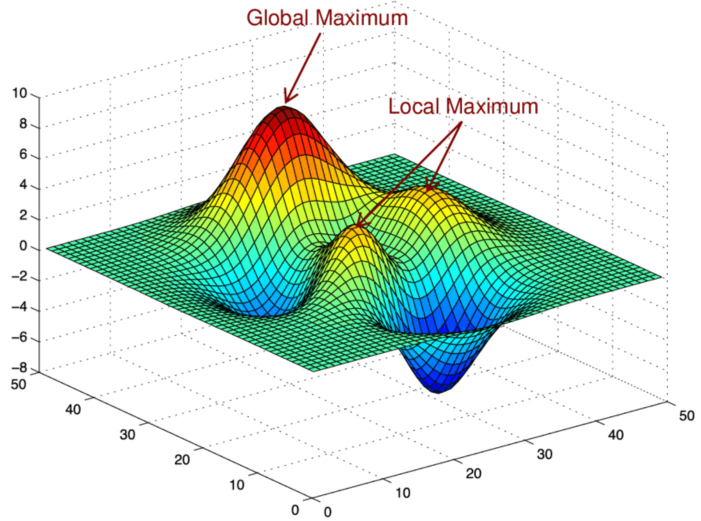
- 一般的最优化问题
  - 非常难以求解
  - 可能需要非常长的计算时间，甚至找不到最优解
- 一类相对简单的问题
  - 最小二乘问题
  - 线性规划问题
  - 凸优化问题
- 更复杂的问题
  - 离散、博弈、随机优化

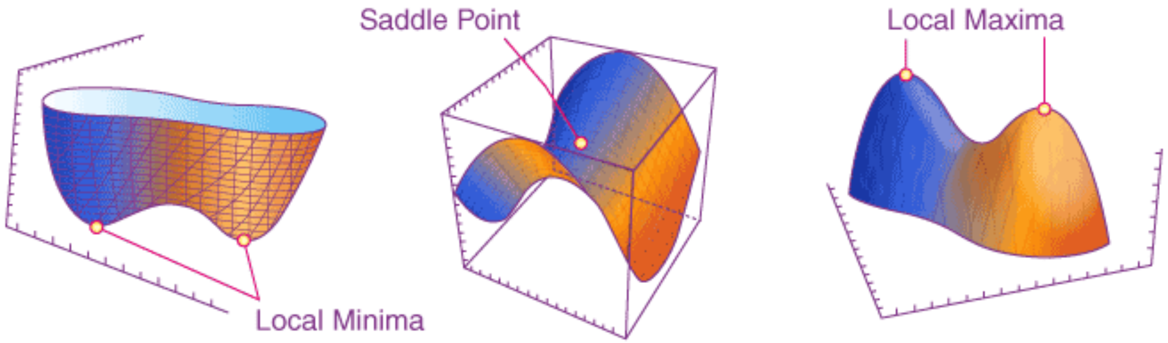
凸优化问题的判定

- 定义域
  - 凸集
- 目标函数
  - 凸函数
- 非凸优化问题与凸优化问题的转化
  - 等价转化
  - 对偶理论

优化算法

- 闭式解（closed form solution）
  - 如果我们能用代数表达式给出其最优解，那么这个解称为闭式解，对应的问题往往比较简单
  - 例如：二次函数在有界区间上
- 实际大多数问题是没有闭式解的，需要通过迭代算法求解
  - 迭代算法的基本思想是:从一个初始点 x^0 出发，按照某种给定的规则进行迭代，得到一个序列 $\{x^k\}$.如果迭代在有限步内终止，那么希望最后 一个点就是优化问题的解
  - 例如：梯度类算法、线搜索法、罚函数法





常用优化技巧

- 泰勒展开

$$f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\frac{f'''(a)}{3!}(x-a)^3+\cdots,$$

- 对偶
  - 每个优化问题都有对应的对偶问题，特别是凸的情形，当原始问题比较难解的时候，其对偶问题可能很容易求解
- 拆分
  - 通过引入变量等方法，讲一个复杂的优化问题拆分成多个相对简单的问题

课程目标

- 目标
  - 识别/建模 哪些问题是凸优化问题
  - 通过代码对相对简单的凸优化问题进行求解
  - 判断最优解，给出算法性能极限等
- 主要内容
  - 凸集、凸函数、凸优化问题
  - 凸优化理论
  - 优化建模
  - 算法
  - 近似算法、博弈算法、组合优化算法

优化问题的历史

- 17世纪：Newton-Raphson：求解$f(x) = 0$
- 18世纪：Lagrange：拉格朗日乘子
- 19世纪：Gauss-Siedel
- 理论（凸分析）：1900-1970
  - 1944 Bellman: 动态规划
  - 1944 Von Neuman，Nash：博弈论，Nash均衡
- 算法
  - 单纯形法（Dantzig 1947）
  - 内点法
  - 次梯度法
  - 1970 多项式内点法应用于线性规划
  - 1990 多项式内点法应用于非线性规划
- 应用
  - 1990以前：运筹研究
  - 1990以后：控制、信号处理、通信、电路设计、金融等等

谢谢！