第三章 典型优化问题

修贤超

https://xianchaoxiu.github.io

目录

凸优化问题

- 标准形式的凸优化问题

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \cdots, m \\ & a_i^\top x = b_i, \quad i = 1, \cdots, p \end{aligned}$$

  - $f_0, f_1, \ldots, f_m$ 为凸函数
  - $a_i^\top x = b_i$ 为线性等式约束

- 经常写成

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \cdots, m \\ & Ax = b \end{aligned}$$

- 凸问题的可行集为凸集

应用举例

- 考虑

$$
\begin{aligned}
\min \quad & f_0(x) = x_1^2 + x_2^2 \\
\text{s.t.} \quad & f_1(x) = x_1/(1+x_2^2) \le 0 \\
& h_1(x) = (x_1 + x_2)^2 = 0
\end{aligned}
$$

  - f_0 为凸函数, 可行集 $\{(x_1, x_2) \mid x_1 = -x_2 \le 0\}$ 为凸集
  - f_1 非凸, h_1 不是线性函数 $\quad\Rightarrow\quad$ 不是凸问题
  - 等价于 (但不完全相等) 凸优化问题

$$
\begin{aligned}
\min \quad & x_1^2 + x_2^2 \\
\text{s.t.} \quad & x_1 \le 0 \\
& x_1 + x_2 = 0
\end{aligned}
$$

- **凸优化问题的任意局部极小点都是全局最优**

证明 假设 x 是局部极小，y 全局最优且 $f_0(y) < f_0(x)$. x 局部最优意味着存在 $R > 0$ 使得

$$z\text{可行},\quad \|z - x\|_2 \le R \quad \Longrightarrow \quad f_0(z) \ge f_0(x)$$

考虑 $z = \theta y + (1-\theta)x$ 且 $\theta = R/(2\|y - x\|_2)$

- $\|y - x\|_2 > R$, 因此 $0 < \theta < 1/2$
- z 是两个可行点的凸组合，因此也可行
- $\|z - x\|_2 = R/2$，并且

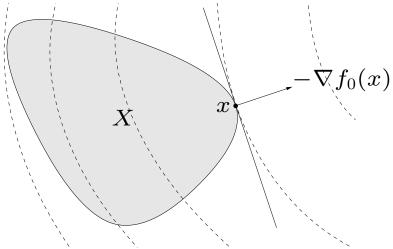
$$f_0(z) \le \theta f_0(x) + (1-\theta) f_0(y) < f_0(x)$$

这与 x 是局部极小的假设矛盾

可微凸优化问题的最优性条件

- 设 x 是凸优化问题 $\min_{x \in X} \; f_0(x)$ 最优解当且仅当 x 可行且满足

$$\nabla f_0(x)^\top (y - x) \geq 0, \quad \forall y \in X$$



- 如果 $\nabla f_0(x)$ 非零，它定义了可行集 X 在 x 处的支撑超平面

具体含义

- 无约束优化 x 是最优解当且仅当

$$x \in \mathrm{dom}\ f_0, \quad \nabla f_0(x) = 0$$

- 等式约束优化问题

$$\min \quad f_0(x) \qquad Ax = b$$

x 是最优解当且仅当存在 v 使得

$$x \in \mathrm{dom}\ f_0, \quad Ax = b, \quad \nabla f_0(x) + A^\top v = 0$$

- 非负约束优化问题

$$\min \quad f_0(x) \qquad x \succeq 0$$

x 是最优解当且仅当

$$x \in \mathrm{dom}\ f_0, \quad x \succeq 0, \quad \begin{cases} \nabla f_0(x)_i \geq 0 & x_i = 0 \\ \nabla f_0(x)_i = 0 & x_i > 0 \end{cases}$$

线性规划基本形式

- 线性规划问题的一般形式

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \quad & c^\top x \\
\text{s.t.} \quad & Ax = b \\
& Gx \le e
\end{aligned}
$$

- 线性规划问题的标准形式

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \quad & c^\top x \\
\text{s.t.} \quad & Ax = b \\
& x \ge 0
\end{aligned}
$$

- 线性规划问题的不等式形式

$$
\begin{aligned}
\max_{y\in\mathbb{R}^n} \quad & b^\top y \\
\text{s.t.} \quad & A^\top y \le c
\end{aligned}
$$

应用举例: 基追踪问题

- 基追踪问题是压缩感知中的一个基本问题，可以写为

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \|x\|_1 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

- 对每个 $|x_i|$ 引入一个新的变量 z_i，可以转化为

$$\begin{aligned} \min_{z \in \mathbb{R}^n} \quad & \sum_{i=1}^n z_i \\ \text{s.t.} \quad & Ax = b \\ & -z_i \le x_i \le z_i, \quad i = 1, \cdots, n \end{aligned}$$

应用举例: 数据拟合

- 最小 ℓ_∞ 范数模型

$$\min_{x\in\mathbb{R}^n} \quad \|Ax-b\|_\infty$$

- 令 $t=\|Ax-b\|_\infty$，则得到等价问题

$$\begin{aligned} \min_{x\in\mathbb{R}^n,\, t\in\mathbb{R}} \quad & t \\ \text{s.t.} \quad & \|Ax-b\|_\infty \le t \end{aligned}$$

- 利用 ℓ_∞ 范数的定义，可以进一步写为

$$\begin{aligned} \min_{x\in\mathbb{R}^n,\, t\in\mathbb{R}} \quad & t \\ \text{s.t.} \quad & -t\mathbf{1} \le Ax-b \le t\mathbf{1} \end{aligned}$$

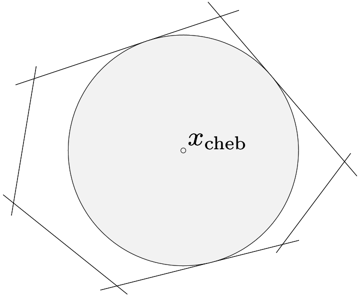
应用举例：多面体的切比雪夫中心

- 多面体

$$\mathcal{P} = \{x \mid a_i^\top x \le b_i, i = 1, \cdots, m\}$$

的切比雪夫中心，即为其最大半径内接球的球心

$$\mathcal{B} = \{x_c + u \mid \ \|u\|_2 \le r\}$$



- 可以转化为

$$\begin{aligned} \max_{x_c, r} \quad & r \\ \text{s.t.} \quad & a_i^\top x_c + r\|a_i\|_2 \le b_i, \quad i = 1, \dots, m \end{aligned}$$

目录

最小二乘问题

- 最小二乘问题的一般形式如下

$$\min_{x\in\mathbb{R}^n} \quad \sum_{i=1}^{m} r_i^2(x)$$

- 如果所有的 $r_i : \mathbb{R}^n \to \mathbb{R}$ 都是线性函数，则称线性最小二乘问题，否则称为非线性最小二乘问题
- 最小二乘问题是线性回归和非线性回归的基础
- 如果噪声服从高斯分布，最小二乘问题的解对应于原问题的最大似然解.

应用举例: 线性最小二乘问题

- 线性最小二乘问题是回归分析中的一个基本模型，它可以表示为

$$\min_{x\in\mathbb{R}^n} \quad \sum_{i=1}^{m}(a_i^\top x - b_i)^2$$

即 $r_i(x) = a_i^\top x - b_i,\ i = 1, 2, \cdots, m$

- 记 $A = [a_1, a_2, \cdots, a_m]^\top$，那么线性最小二乘问题可以等价地写成

$$\min_{x\in\mathbb{R}^n} \quad f(x) = \frac{1}{2}\|Ax - b\|_2^2$$

- $x \in \mathbb{R}^n$ 为其全局极小解当且仅当 x 满足方程

$$\nabla f(x) = A^\top(Ax - b) = 0$$

应用举例: 数据插值

- 给定数据集 $\{a_i \in \mathbb{R}^p, b_i \in \mathbb{R}^q, i = 1, 2, \cdots, m\}$，插值是求一个映射 f，使得

$$b_i = f(a_i), \quad i = 1, 2, \cdots, m$$

- 利用线性函数 $f(a) = Xa + y$ 逼近，可以建立如下最小二乘问题

$$\min_{X \in \mathbb{R}^{q \times p}} \quad \sum_{i=1}^{m} \|Xa_i + y - b_i\|^2$$

- 假设 $\{\phi_i(a)\}_{i=1}^{n} (n \leq m)$ 为插值空间的一组基，数据插值可以写成

$$b_j = f(a_j) = \sum_{i=1}^{n} x_i \phi_i(a_j), \quad j = 1, 2, \cdots, m$$

应用举例: 数据插值

- 假设有一些简单的非线性向量函数 $\phi_i(\theta):\mathbb{R}^q\to\mathbb{R}^q$，并构造如下复合函数

$$f(\theta)=\phi_n(X_n\phi_{n-1}(X_{n-1}\cdots\phi_1(X_1\theta+y_1)\cdots+y_{n-1})+y_n)$$

- 常用的简单非线性函数有 ReLU，即

$$\phi_i(\theta)=(\mathrm{ReLU}(\theta_1),\mathrm{ReLU}(\theta_2),\cdots,\mathrm{ReLU}(\theta_q))^\top,\quad i=1,2,\cdots,n$$

且

$$\mathrm{ReLU}(t)=\begin{cases}t, & t\geq 0\\ 0, & \text{其他}\end{cases}$$

- 在深度学习中，这样的做法往往会带来更多未知的非线性，因而可能在更大的函数空间中得到未知函数的一个更好的逼近

应用举例：带微分方程约束优化问题

- 当约束中含微分方程时，称相应的优化问题为带微分方程约束的优化问题
- 考虑瓦斯油催化裂解生成气体和其他副产物的反应系数的问题反应过程

$$\begin{cases} \dot{y}_1 = -(\theta_1 + \theta_3) y_1^2 \\ \dot{y}_2 = \theta_1 y_1^2 - \theta_2 y_2 \end{cases}$$

- 转化为最小二乘问题

$$\begin{aligned} \min_{\theta \in \mathbb{R}^3} \quad & \sum_{j=1}^{n} \|y(\tau_j; \theta) - z_j\|^2 \\ \text{s.t.} \quad & y(\tau; \theta) \text{满足上述方程组} \end{aligned}$$

这里 z_j 是在时刻 τ_j 的 y 的测量值，n 为测量的时刻数量

目录

- 3.1 线性规划
- 3.2 最小二乘问题
- 3.3 复合优化问题
- 3.4 随机优化问题
- 3.5 半定规划
- 3.6 矩阵优化
- 3.7 优化模型语言

复合优化问题

- 复合优化问题一般可以表示为

$$\min_{x\in\mathbb{R}^n} \quad \psi(x) = f(x) + h(x)$$

  - $f(x)$ 是光滑函数, 如数据拟合项
  - $h(x)$ 可能是非光滑的, 如 ℓ_1 范数正则项，约束集合的示性函数
- 常用算法有次梯度法，近似点梯度法，Nesterov 加速法、交替方向乘子法等

应用举例: 信号处理

- ℓ_1 范数正则化回归分析问题

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}\|Ax-b\|_2^2+\mu\|x\|_1$$

- 矩阵分离问题

$$\begin{aligned}\min_{X,S\in\mathbb{R}^{m\times n}} \quad & \|X\|_*+\mu\|S\|_1 \\ \text{s.t.} \quad & X+S=M\end{aligned}$$

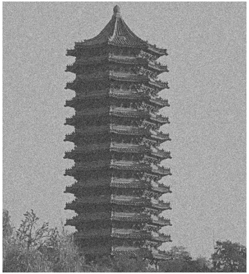
- 字典学习问题

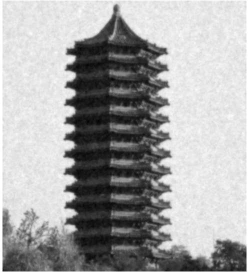
$$\begin{aligned}\min_{X,D\in\mathbb{R}^{m\times n}} \quad & \frac{1}{2n}\|DX-A\|_F^2+\lambda\|X\|_1 \\ \text{s.t.} \quad & \|D\|_F\leqslant 1\end{aligned}$$

应用举例: 图像去噪

- 图像去噪问题是指从一个带噪声的图像中恢复出不带噪声的原图
- 由全变差模型，去噪问题可表示为

$$\min_{x \in \mathbb{R}^{n \times n}} \quad \frac{1}{2}\|x - y\|_F^2 + \lambda \|x\|_{TV}$$





应用举例: 盲反卷积

- 盲反卷积是从一个模糊的图像恢复出原来清晰的图像, 也称为去模糊
- 盲反卷积问题可以表示成

$$y = a * x + \varepsilon$$

- 假设噪声为高斯噪声，则转化为

$$\min_{a,x} \quad \frac{1}{2}\|y - a * x\|_2^2$$

- 假设原始图像信号在小波变换下是稀疏的，进一步得到

$$\min_{a,x} \quad \frac{1}{2}\|y - a * x\|_2^2 + \|\lambda \odot (Wx)\|_1,$$

其中 W 是小波框架，$\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_m)^\top$ 用来控制稀疏度

目录

随机优化问题

- 随机优化问题可以表示为

$$\min_{x\in\mathcal{X}} \quad \mathbb{E}_\xi[F(x,\xi)] + h(x)$$

  - $F(x,\xi)$ 表示样本 ξ 上的损失或者奖励
  - $h(x)$ 用来保证解的某种性质
- 假设有 N 个样本 $\xi_1,\xi_2,\cdots,\xi_N$，令 $f_i(x)=F(x,\xi_i)$，得到经验风险极小化问题 (采样平均极小化问题)

$$\min_{x\in\mathcal{X}} \quad f(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x) + h(x)$$

- 样本数 N 比较多，可行域所在空间维数 n 比较大，导致计算困难

应用举例: 随机主成分分析

- 如果样本点 ξ 服从某个零均值分布 $\mathcal{D}$，那么随机主成分分析可以写成

$$\max_{X \in \mathbb{R}^{p \times d}} \quad \mathrm{Tr}(X^\top A A^\top X) \quad \text{s.t.} \quad X^\top X = I$$

$$\Downarrow$$

$$\max_{X \in \mathbb{R}^{p \times d}} \quad \mathrm{Tr}(X^\top \mathbb{E}_{\xi \sim \mathcal{D}}[\xi \xi^\top] X) \quad \text{s.t.} \quad X^\top X = I$$

- 高逼近解过程中需要的样本数量以及所消耗的时间
- 在有限内存情况下的逼近计算分析

应用举例: 分布式鲁棒优化

- 深度学习的目的是从已有的未知分布的数据中学出一个好的预测器，为了提高预测器的泛化能力，考虑

$$\min_{h} \quad \mathbb{E}_z[F(h,z)]$$

$$\Downarrow$$

$$\min_{h} \quad \max_{\hat{z}\in\Gamma} \mathbb{E}_{\hat{z}}[F(h,\hat{z})]$$

- 集合 Γ 中随机变量的分布与真实数据的分布在一定意义下非常接近
- Wasserstein 距离可以改变原来经验分布的支撑集，常用于选取 Γ

Q&A

Thank you!

感谢您的聆听和反馈