优化理论与算法

第五章

凸问题

郭加熠|助理教授

目录

优化问题的标准形式

优化问题的标准形式：

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

- $x \in \mathbb{R}^n$ 是优化变量
- $f_0 : \mathbb{R}^n \to \mathbb{R}$ 是目标函数
- $f_i : \mathbb{R}^n \to \mathbb{R}$ 是不等式约束函数
- $h_i : \mathbb{R}^n \to \mathbb{R}$ 是等式约束函数

- $p^\star = \inf\{f_0(x) \mid f_i(x) \le 0,\ h_i(x) = 0\}$
- 若问题不可行: $p^\star = \infty$
- 若问题在可行域内无下界： $p^\star = -\infty$

最优解与局部最优解

- 若 $x \in \operatorname{dom} f_0$ 且满足约束，则 x 为**可行解**
- 若可行解 x 满足 $f_0(x) = p^\star$，则为**最优解**，最优解集合记为 X_{opt}
- 若存在 $R > 0$，使 x 在以下问题中为最优解，则称其为**局部最优解**：

$$
\begin{array}{ll}
\displaystyle\min_{z} & f_0(z) \\
\text{s.t.} & f_i(z) \le 0,\ i = 1, \ldots, m \\
& h_i(z) = 0,\ i = 1, \ldots, p \\
& \|z - x\|_2 \le R
\end{array}
$$

- $f_0(x) = 1/x,\ \operatorname{dom} f_0 = \mathbb{R}_{++}$：
  $p^\star = 0$，无最优解
- $f_0(x) = -\log x,\ \operatorname{dom} f_0 = \mathbb{R}_{++}$：
  $p^\star = -\infty$
- $f_0(x) = x \log x,\ \operatorname{dom} f_0 = \mathbb{R}_{++}$：
  $p^\star = -1/e,\ x = 1/e$ 是最优解
- $f_0(x) = x^3 - 3x$：$p^\star = -\infty$，
  但 $x = 1$ 是局部最优解

隐式约束（Implicit Constraints）

标准形式的优化问题包含隐式约束：

$$x \in \mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom} f_i \ \cap \bigcap_{i=1}^{p} \operatorname{dom} h_i$$

- $\mathcal{D}$ 被称为问题的定义域（domain）
- $f_i(x) \le 0$，$h_i(x) = 0$ 是显式约束
- 若 $m = p = 0$，即无显式约束，则问题为无约束问题（unconstrained）

例子：

$$\min f_0(x) = -\sum_{i=1}^{k} \log(b_i - a_i^\top x)$$

这是一个无约束问题，但其隐式约束为 $a_i^\top x < b_i$

可行性问题（Feasibility Problem）

$$\begin{aligned} \text{Find} \quad & x \\ \text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

可行性问题是目标函数为常数 $f_0(x) = 0$ 的一般优化问题特例：

$$\begin{aligned} \min \quad & 0 \\ \text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

- 若约束可行，则 $p^\star = 0$，任意可行 x 为最优
- 若约束不可行，则 $p^\star = \infty$

目录

凸优化问题（Convex Optimization Problem）

标准形式：

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

- $f_0, f_1, \ldots, f_m$ 是凸函数，等式约束是仿射的
- 若 f_0 为拟凸函数，$f_1, \ldots, f_m$ 为凸函数，则问题为拟凸优化问题

常写作：

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

重要性质： 凸优化问题的可行解集是凸集

非凸约束示例

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

- f_0 是凸函数；可行域 $\{(x_1, x_2) \mid x_1 = -x_2 \le 0\}$ 是凸集
- 不是凸优化问题：f_1 非凸，h_1 非仿射
- 它等价（但不完全相同）于下列凸优化问题：

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

局部最优与全局最优

任意凸优化问题的局部最优点也是全局最优点。

证明： 假设 x 是局部最优，但存在可行 y 使得 $f_0(y) < f_0(x)$。

局部最优意味着存在 $R > 0$，使得

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

令 $z = \theta y + (1 - \theta)$，不妨设 $\theta = R/(2\|y - x\|_2)$，那么可以验证：

$$\|y - x\|_2 > R \quad \Rightarrow \quad 0 < \theta < 1/2 \in (0, 1)\text{，且 } \|z - x\|_2 = R/2\text{。}$$

由目标函数凸性可得：$f_0(z) \le \theta f_0(y) + (1 - \theta) f_0(x) < f_0(x)$

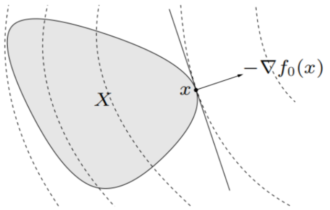
这与 x 是局部最优点矛盾，故 x 必为全局最优点。

可微函数的最优性判据

若目标函数 f_0 可微，点 x 是最优的当且仅当：

- x 是可行解；
- 对所有可行点 y，有：

$$\nabla f_0(x)^T(y-x) \geq 0$$



图示说明：

- 若 $\nabla f_0(x) \neq 0$，则其定义了一个在点 x 处支撑可行域 X 的超平面（supporting hyperplane）

常见结构下的最优性条件

1. 无约束问题： $\min f_0(x)$，则最优的充要条件为：

$$\nabla f_0(x) = 0$$

2. 等式约束问题： $\min f_0(x) \quad \text{s.t.} \quad Ax = b$。最优性充要条件为存在 ν 使得：

$$x \in \operatorname{dom} f_0, \quad Ax = b, \quad \nabla f_0(x) + A^T \nu = 0$$

3. 非负正交域上的最优化： $\min f_0(x) \quad \text{s.t.} \quad x \geq 0$。最优性充要条件为：

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

等价凸优化问题：消除等式约束

若两个问题的解可相互转换，我们称它们等价。

目标：消除等式约束 $Ax = b$，保持凸性不变。

原问题:

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

等价于如下问题（变量变为 z）:

$$
\begin{aligned}
\min \quad & f_0(Fz + x_0) \\
\text{s.t.} \quad & f_i(Fz + x_0) \le 0, \quad i = 1, \ldots, m
\end{aligned}
$$

其中 F 和 x_0 满足：$Ax = b \iff x = Fz + x_0, \quad \text{for some} \quad z$

等价凸问题：引入等式

引入等式约束：

$$\begin{aligned} \min \quad & f_0(A_0x + b_0) \\ \text{s.t.} \quad & f_i(A_ix + b_i) \leq 0, \quad i = 1, \ldots, m \end{aligned}$$

等价于引入变量 $y_i = A_ix + b_i$：

$$\begin{aligned} \min_{x, y_i} \quad & f_0(y_0) \\ \text{s.t.} \quad & f_i(y_i) \leq 0, \quad i = 1, \ldots, m \\ & y_i = A_ix + b_i, \quad i = 0, \ldots, m \end{aligned}$$

等价凸问题：引入松弛变量

引入松弛变量：

$$\min \quad f_0(x) \quad \text{s.t.} \quad a_i^T x \leq b_i, \quad i = 1, \ldots, m$$

等价于：

$$\begin{aligned} \min_{x,s} \quad & f_0(x) \\ \text{s.t.} \quad & a_i^T x + s_i = b_i, \quad i = 1, \ldots, m \\ & s_i \geq 0, \quad i = 1, \ldots, m \end{aligned}$$

等价凸优化形式：改写成上图形式

凸问题标准形式问题：

$$\min f_0(x) \quad \text{s.t.} \;\; f_i(x) \le 0, \;\; Ax = b$$

等价于引入变量 t 的形式：

$$\begin{aligned} \min_{x,t} \quad & t \\ \text{s.t.} \quad & f_0(x) - t \le 0 \\ & f_i(x) \le 0 \\ & Ax = b \end{aligned}$$

等价凸优化形式：对部分变量先进行最小化

原问题

$$\min_{x_1, x_2} f_0(x_1, x_2) \quad \text{s.t.} \ \ f_i(x_1) \le 0$$

等价于令 $\tilde{f}_0(x_1) = \min_{x_2} f_0(x_1, x_2)$

$$\min_{x_1} \tilde{f}_0(x_1) \quad \text{s.t.} \ \ f_i(x_1) \le 0,$$

含义：先优化掉 x_2，之后仅对 x_1 优化。

- **降维**：减少优化变量数量；
- **结构优化**：保留部分变量结构，有利于分布式算法（如 ADMM）实现。

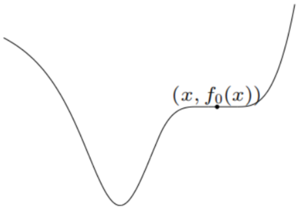
目录

拟凸优化问题（Quasiconvex Optimization）

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

其中：

- $f_0 : \mathbb{R}^n \to \mathbb{R}$ 是拟凸函数；
- $f_1, \dots, f_m$ 是凸函数；
- $Ax = b$ 是仿射约束。



图示：拟凸函数可能出现“鞍点式”的局部最优，见右图。

拟凸函数的下水平集的凸表示

若 $f_0(x)$ 是拟凸函数，则存在一族函数 $\phi_t(x)$ 满足：

- 对每个固定的 t，$\phi_t(x)$ 是凸函数；同时，$f_0(x) \leq t \iff \phi_t(x) \leq 0$

示例： 设 $f_0(x) = \dfrac{p(x)}{q(x)}$，其中：

- $p(x)$ 是凸函数，$p(x) \geq 0$
- $q(x)$ 是凹函数，$q(x) > 0$

令：

$$\phi_t(x) = p(x) - tq(x)$$

则当 $t \geq 0$ 时，ϕ_t 是凸函数，且有：

$$f_0(x) \leq t \iff \phi_t(x) \leq 0$$

拟凸优化：通过凸可行性问题 + 二分法求解

将拟凸优化转化为可行性问题： 固定参数 t，考虑以下凸可行性问题

$$\phi_t(x) \le 0, \quad f_i(x) \le 0, \quad Ax = b \tag{1}$$

- 若 (1) 可行，则 $t \ge p^\star$
- 若 (1) 不可行，则 $t < p^\star$

二分法（bisection method）：

- 输入：$l \le p^\star$，$u \ge p^\star$，$\epsilon > 0$
- 重复直到 $u - l \le \epsilon$：
  1. $t := (l + u)/2$
  2. 解可行性问题 (1)
  3. 若可行，设 $u := t$；否则 $l := t$

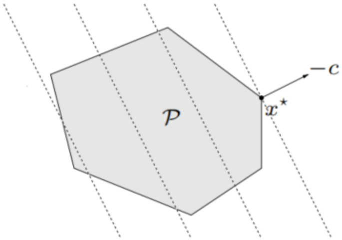
收敛次数：

$$\left\lceil \log_2\left(\frac{u-l}{\epsilon}\right)\right\rceil$$

目录

线性规划（Linear Program, LP）

$$
\begin{aligned}
\min \quad & c^T x + d \\
\text{s.t.} \quad & Gx \le h \\
& Ax = b
\end{aligned}
$$



- 具有仿射目标函数与仿射约束函数的凸优化问题
- 可行解集合是一个多面体（polyhedron）

示例（Examples）

饮食问题（diet problem）：选择 $x_1, \ldots, x_n$ 种食物的数量

- 每单位食物 j 的价格为 c_j，含有 a_{ij} 单位的营养素 i
- 健康饮食要求营养素 i 的摄入量至少为 b_i

为找到最便宜的健康饮食方案，求解如下问题：

$$
\begin{aligned}
\min \quad & c^T x \\
\text{s.t.} \quad & Ax \leq b \\
& x \geq 0
\end{aligned}
$$

Chebyshev 中心（Chebyshev center of a polyhedron）

Chebyshev 中心定义于：

$$\mathcal{P} = \{x \mid a_i^T x \leq b_i,\ i = 1, \ldots, m\}$$

是其最大内切球的中心，记为：$\mathcal{B} = \{x_c + u \mid \|u\|_2 \leq r\}$

- $a_i^T x \leq b_i$ 对所有 $x \in \mathcal{B}$ 成立，当且仅当：

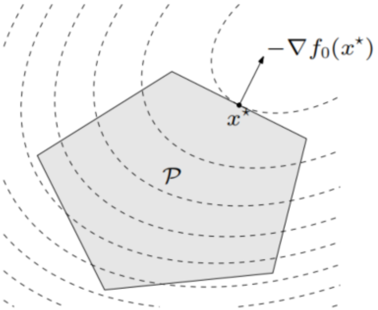
$$\sup\{a_i^T(x_c + u) \mid \|u\|_2 \leq r\} = a_i^T x_c + r\|a_i\|_2 \leq b_i$$

- 因此，可以通过求解以下线性规划来确定 x_c, r：

$$\begin{array}{ll} \max & r \\ \text{s.t.} & a_i^T x_c + r\|a_i\|_2 \leq b_i, \quad i = 1, \ldots, m \end{array}$$

二次规划（Quadratic Program, QP）

$$
\begin{aligned}
\min \quad & \frac{1}{2}x^T P x + q^T x + r \\
\text{s.t.} \quad & Gx \leq h \\
& Ax = b
\end{aligned}
$$



- 要求 $P \in \mathbb{S}^n_+$，所以目标函数是凸二次函数
- 目标是在多面体可行域上最小化一个凸二次函数

示例：最小二乘与随机成本线性规划

最小二乘（least-squares） $\min \quad \|Ax - b\|_2^2$

- 解析解 $x^\star = A^\dagger b$，其中 $A^\dagger$ 是广义逆
- 可添加线性约束，例如 $l \le x \le u$

具有随机成本的线性规划（linear program with random cost）

$$
\begin{aligned}
\min \quad & \bar{c}^T x + \gamma x^T \Sigma x = \mathbb{E}[c^T x] + \gamma \mathbb{V}\mathrm{ar}(c^T x) \\
\text{s.t.} \quad & Gx \le h, \quad Ax = b
\end{aligned}
$$

- 向量 c 是具有均值 $\bar{c}$、协方差为 Σ 的随机向量
- 因此 $c^T x$ 是具有均值 $\bar{c}^T x$、方差 $x^T \Sigma x$ 的随机变量
- $\gamma > 0$ 是风险规避参数，调节期望成本与方差（风险）之间的权衡

带二次约束的二次规划（QCQP）

$$
\begin{aligned}
\min \quad & \frac{1}{2}x^T P_0 x + q_0^T x + r_0 \\
\text{s.t.} \quad & \frac{1}{2}x^T P_i x + q_i^T x + r_i \le 0, \quad i = 1, \dots, m \\
& Ax = b
\end{aligned}
$$

- 要求 $P_i \in \mathbb{S}_+^n$，所以目标函数与约束均为凸二次函数
- 要求 $P_1, \dots, P_m \in \mathbb{S}_{++}^n$，所以可行域是 m 个椭球与一个仿射集的交集

二阶锥规划（Second-order Cone Programming）

$$
\begin{aligned}
\min \quad & f^T x \\
\text{s.t.} \quad & \|A_i x + b_i\|_2 \le c_i^T x + d_i, \quad i = 1, \dots, m \\
& Fx = g
\end{aligned}
$$

其中 $A_i \in \mathbb{R}^{n_i \times n}$，$F \in \mathbb{R}^{p \times n}$

- 不等式为二阶锥（second-order cone, SOC）约束，也可以表示为：

  $(A_i x + b_i,\ c_i^T x + d_i) \in n_i + 1$ 维的二阶锥

- 若 $n_i = 0$，则退化为线性规划；若 $c_i = 0$，则退化为 QCQP
- 比 QCQP 和 LP 更一般

鲁棒线性规划（Robust Linear Programming）

优化问题中的参数常常具有不确定性，例如在线性规划中：

$$
\begin{aligned}
\min \quad & c^T x \\
\text{s.t.} \quad & a_i^T x \leq b_i, \quad i = 1, \ldots, m
\end{aligned}
$$

其中 c，a_i，b_i 都可能具有不确定性。

我们以 a_i 为例，常见的两种建模方式：

- **确定性模型**：约束对所有可能的 $a_i \in \mathcal{E}_i$ 都必须成立

$$\min \quad c^T x \quad \text{s.t.} \quad a_i^T x \leq b_i \quad \forall a_i \in \mathcal{E}_i, \quad i = 1, \ldots, m$$

- **随机模型**：a_i 是随机变量，约束以概率 η 成立

$$\min \quad c^T x \quad \text{s.t.} \quad \mathrm{prob}(a_i^T x \leq b_i) \geq \eta, \quad i = 1, \ldots, m$$

确定性鲁棒优化中的 SOCP 表达

- 对于不确定集 $\mathcal{E}_i$，选用椭球形式：

$$\mathcal{E}_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \leq 1\}, \quad \bar{a}_i \in \mathbb{R}^n, \; P_i \in \mathbb{R}^{n \times n}$$

  中心为 $\bar{a}_i$，半轴由 P_i 的奇异值与奇异向量决定

- 鲁棒线性规划问题：

$$\min \; c^T x \quad \text{s.t.} \; a_i^T x \leq b_i, \quad \forall a_i \in \mathcal{E}_i, \quad i = 1, \ldots, m$$

  可等价转化为二阶锥规划：

$$\min \; c^T x \quad \text{s.t.} \; \bar{a}_i^T x + \|P_i^T x\|_2 \leq b_i, \quad i = 1, \ldots, m$$

  （推导基于：$\sup_{\|u\|_2 \leq 1} (\bar{a}_i + P_i u)^T x = \bar{a}_i^T x + \|P_i^T x\|_2$）

随机鲁棒优化中的 SOCP 表达

- 假设 $a_i \sim \mathcal{N}(\bar{a}_i, \Sigma_i)$，均值为 $\bar{a}_i$，协方差为 Σ_i
- 则 $a_i^T x$ 是正态分布随机变量，均值为 $\bar{a}_i^T x$，方差为 $x^T \Sigma_i x$
- 因此概率约束可写为：$\mathrm{prob}(a_i^T x \le b_i) = \Phi\left(\frac{b_i - \bar{a}_i^T x}{\|\Sigma_i^{1/2} x\|_2}\right)$
  其中 $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} e^{-t^2/2}\, dt$ 为标准正态分布的累积分布函数
- 鲁棒线性规划问题：

$$\min \ c^T x \quad \text{s.t.} \ \mathrm{prob}(a_i^T x \le b_i) \ge \eta, \quad i = 1, \ldots, m$$

当 $\eta \ge 1/2$ 时，可转化为 SOCP：

$$\min \ c^T x \quad \text{s.t.} \ \bar{a}_i^T x + \Phi^{-1}(\eta) \|\Sigma_i^{1/2} x\|_2 \le b_i, \quad i = 1, \ldots, m$$

半正定规划（Semidefinite Program, SDP）

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & x_1 F_1 + x_2 F_2 + \cdots + x_n F_n + G \preceq 0 \\ & Ax = b \end{aligned}$$

其中给定参数为：c, A, b 以及 $F_i, G \in \mathbb{S}^k$；决策变量：$x \in \mathbb{R}^n$

- 不等式约束被称为线性矩阵不等式
- 多个不等式，如：$x_1 \hat{F}_1 + \cdots + x_n \hat{F}_n + \hat{G} \preceq 0, \quad x_1 \tilde{F}_1 + \cdots + x_n \tilde{F}_n + \tilde{G} \preceq 0$

  等价于单个：$x_1 \begin{bmatrix} \hat{F}_1 & 0 \\ 0 & \tilde{F}_1 \end{bmatrix} + x_2 \begin{bmatrix} \hat{F}_2 & 0 \\ 0 & \tilde{F}_2 \end{bmatrix} + \cdots + x_n \begin{bmatrix} \hat{F}_n & 0 \\ 0 & \tilde{F}_n \end{bmatrix} + \begin{bmatrix} \hat{G} & 0 \\ 0 & \tilde{G} \end{bmatrix} \preceq 0$

LP 与 SOCP 的 SDP 表达

LP 与等价 SDP

- LP: $\min c^T x \quad \text{s.t. } Ax \le b$
- SDP: $\min c^T x \quad \text{s.t. } \operatorname{diag}(Ax - b) \preceq 0$

SOCP 与等价 SDP

- SOCP: $\min f^T x \quad \text{s.t. } \|A_i x + b_i\|_2 \le c_i^T x + d_i, \quad i = 1, \ldots, m$
- SDP:

$$\min f^T x \quad \text{s.t. } \begin{bmatrix} (c_i^T x + d_i) I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{bmatrix} \succeq 0, \quad i = 1, \ldots, m$$

特征值最小化问题（Eigenvalue Minimization）

$$
\begin{aligned}
\min \quad & \lambda_{\max}(A(x)) \\
\text{s.t.} \quad & A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n, \quad A_i \in \mathbb{S}^k
\end{aligned}
$$

等价的 SDP 形式：

$$
\min \ t \quad \text{s.t.} \ A(x) \preceq tI
$$

- 决策变量为 $x \in \mathbb{R}^n, t \in \mathbb{R}$
- 推理基于：

$$
\lambda_{\max}(A) \leq t \iff A \preceq tI
$$