



凸优化问题

Convex Problems

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

- 凸集

- 一个集合C是凸集，则集合中任意两点所组成的线段仍然在该集合中

$$\forall x_1, x_2 \in C, 0 \leq \theta \leq 1. \Rightarrow \theta x_1 + (1 - \theta)x_2 \in C$$

- 重要的凸集

- 超平面、半空间、多面体
 - 对称半正定矩阵集合
 - 球、椭圆

- 凸集的证明

- 定义
 - 保凸的运算：凸集的交集、凸集的仿射变换

• 定义

1. 对于一个函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 如果 $\text{dom} f$ 为凸集, 且

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

对所有的 $x, y \in \text{dom} f$, $0 \leq \theta \leq 1$ 成立, 则 f 为 (下) 凸函数

2. 一阶条件: 对于可微函数 f , 如果 $\text{dom} f$ 为凸集, 则 f 为凸函数当且仅当

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \text{dom } f$$

3. 对于二阶可微函数 f , 如果定义域为凸集, f 为凸函数当且仅当

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \text{dom } f$$

- 常见的凸函数

- 负熵: $x \log x$ on \mathbb{R}_{++}

- 范数函数: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$ for $p \geq 1$;

- $f(x) = \max \{x_1, \dots, x_n\}, x \in \mathbb{R}^m$

- 保凸的运算

- 非负加权和、仿射组合、最大化

- 函数组合

- 函数的共轭

- $f^*(y) = \sup_{x \in \text{dom } f} (y^T x - f(x))$

- 优化问题的标准形式
- 凸优化问题
- 典型优化问题
 - 线性优化
 - 最小二乘问题
 - 复合优化问题
 - 正定规划
 - ...

一般优化问题



- 优化问题

$$\min f_0(x).$$

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

$$h_i(x) = 0, \quad i = 1, \dots, p$$

- 其中

- $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$: 优化变量(optimization variable)
- $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$: 目标/损失/效用函数(Objective/loss/utility function)
- $f_i: \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, m$: 不等式约束(Inequality constraint)
- $h_i: \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, p$: 等式约束(Equality constraint)
- $m=p=0$: 无约束问题

优化问题的最优解

- 优化问题的域 (domain)

$$D = \bigcap_{i=0}^m \text{dom} f_i \cap \bigcap_{i=1}^p \text{dom} h_i$$

- 可行解集 (feasible set) : $X = \{x \in D \text{ 且所有约束条件可以满足}\}$
- 问题的最优值 (optimal value) : $p^* = \inf\{f_0(x) | x \in X_f\}$
- 最优解 (optimal point/solution) : 若 x^* 可行, 且 $f_0(x^*) = p^*$
- 最优解集 (optimal set) : $X_{opt} = \{x | x \in X_f, f_0(x^*) = p^*\}$
- ε -次优解集 (ε -suboptimal set) : $X_\varepsilon = \{x | x \in X_f, f_0(x^*) \leq p^* + \varepsilon\}$

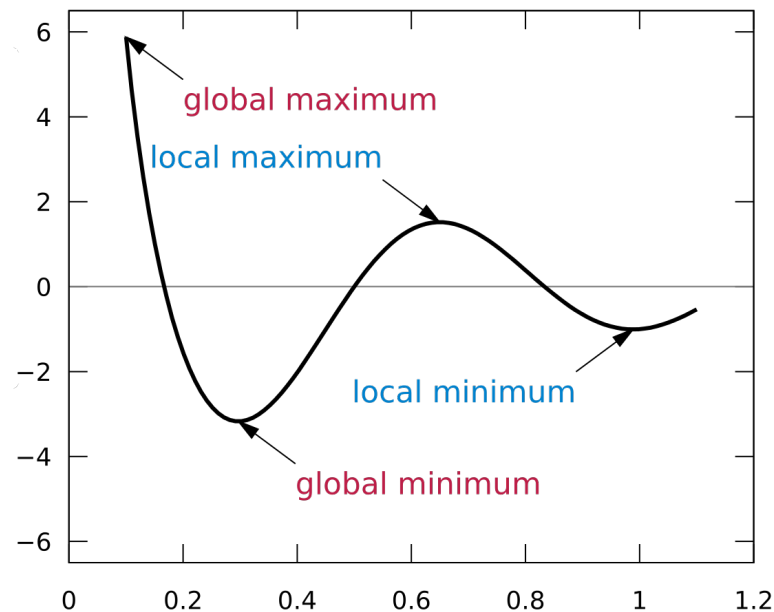
局部最优解 (Locally optimal)



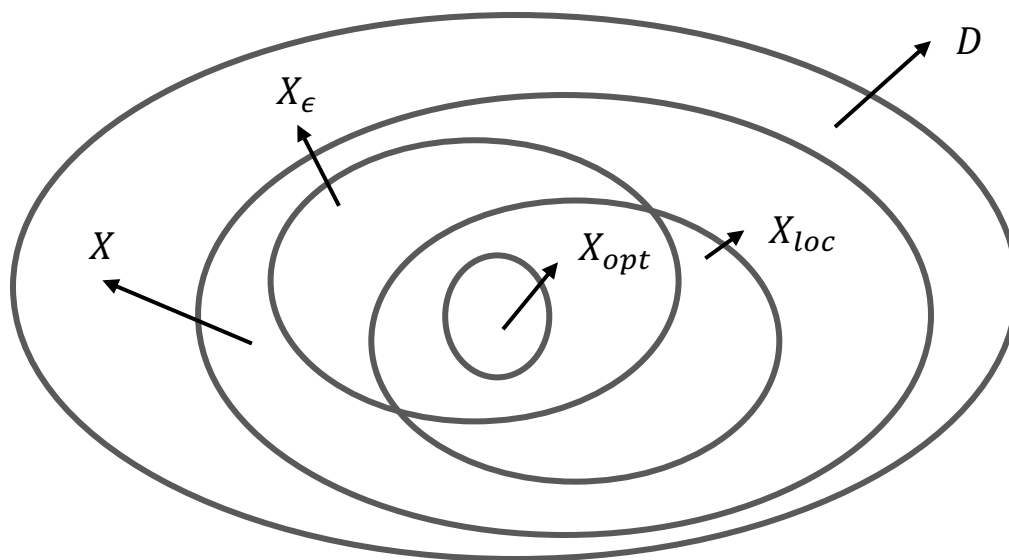
- x 为局部最优解, 如果 $\exists R > 0$ 使得 x 为以下问题的最优解:

$$\begin{array}{ll} \min(\text{over } z) & f_0(z) \\ \text{s.t.} & f_i(z) \leq 0, \quad i = 1, \dots, m \\ & h_i(z) = 0, \quad i = 1, \dots, p \\ & \|z - x\|_2 \leq R \end{array}$$

- 所有局部最优解构成的解集为 X_{loc}



不同解集的关系



- (狭义) 凸优化问题的标准形式

$$\min. f_0(x).$$

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

$$h_i(x) = 0, \quad i = 1, \dots, p$$

- 其中

- $f_i: \mathbb{R}^n \rightarrow \mathbb{R}, i = 0, \dots, m$: 凸函数

- $h_i: \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, p$: 等式约束为仿射函数

- 性质: 凸优化问题的可行解集为凸集

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

- 是否为标准形式的凸优化问题?
- 等价变换

$$\begin{array}{ll}\min & x_1^2 + x_2^2 \\ \text{s.t.} & x_1 \leq 0 \\ & x_1 + x_2 = 0\end{array}$$

$$\max f_0(x).$$

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

$$a_i^T x = b_i, \quad i = 1, \dots, p$$

- 其中 f_0 : 凹函数
- 是否为凸优化问题

重要性质：局部最优=全部最优



- 局部最优=全部最优

- 局部最优：存在 $R > 0$ 使得 x 为以下问题的最优解：

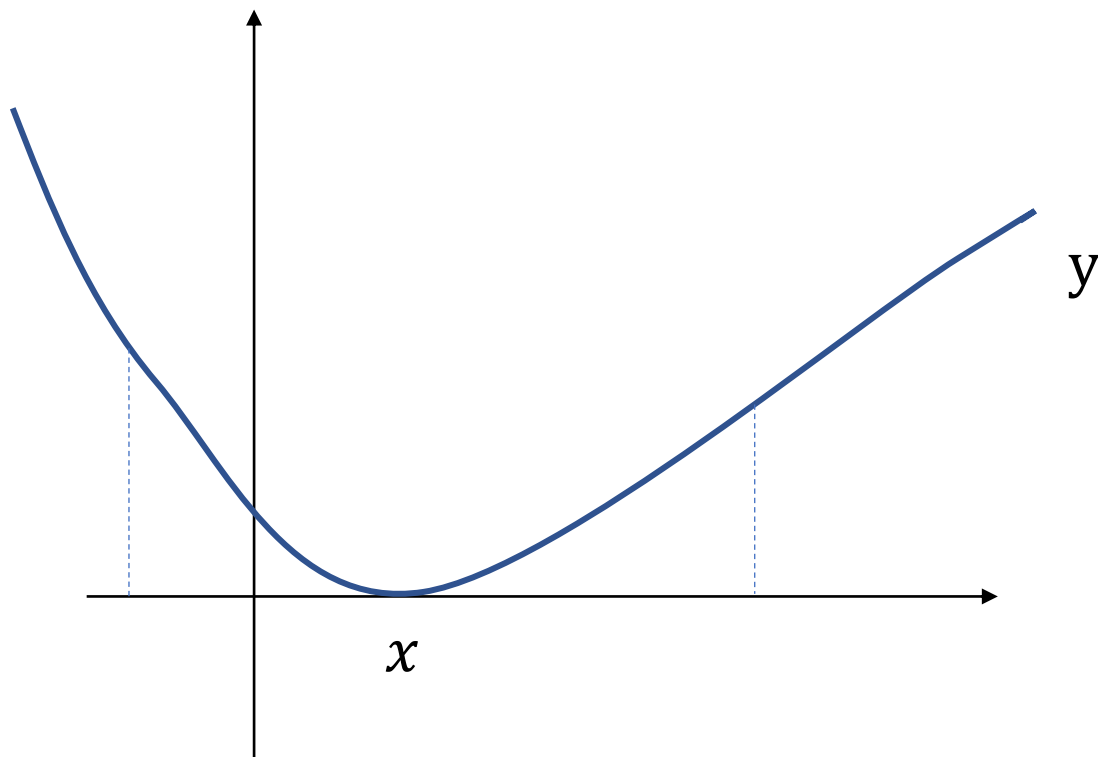
$$\begin{array}{ll} \min(\text{over } z) & f_0(z) \\ \text{s.t.} & f_i(z) \leq 0, \quad i = 1, \dots, m \\ & h_i(z) = 0, \quad i = 1, \dots, p \\ & \|z - x\|_2 \leq R \end{array}$$

- 全局最优： $p^* = \inf\{f_0(x) | x \in D_f\}$, $f_0(x^*) = p^*$

重要性质：局部最优=全部最优



• 证明：



- 可微目标函数下的最优条件

- 凸函数的一阶条件

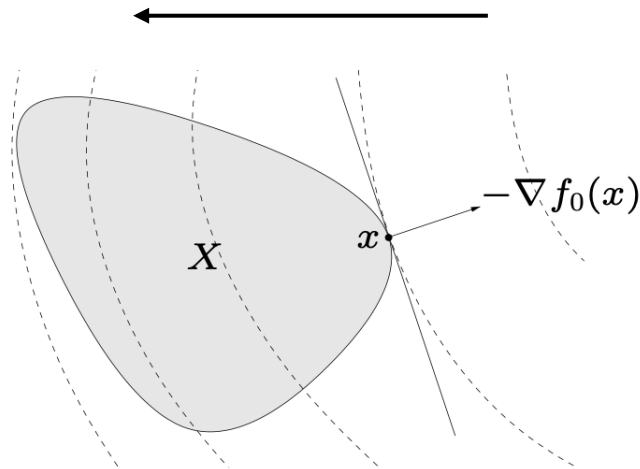
- f_0 可微, 则 f_0 为凸函数 $\Rightarrow \text{dom} f$ 为凸集

- $f_0(y) \geq f_0(x) + \nabla f_0^T(x)(y - x), \forall x, y \in \text{dom} f$

- 最优条件

- $x \in X$ 最优 \Leftrightarrow

- $$\nabla f_0(x)^T(y - x) \geq 0 \text{ for all } y \in X.$$



梯度表示某一函数在该点处的方向导数沿着该方向取得最大值

最优条件（特殊情况）



- 无约束问题：

- x 为最优解的充分必要条件

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



凸优化问题(2)-典型优化问题

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>

线性规划 (Linear Programming)



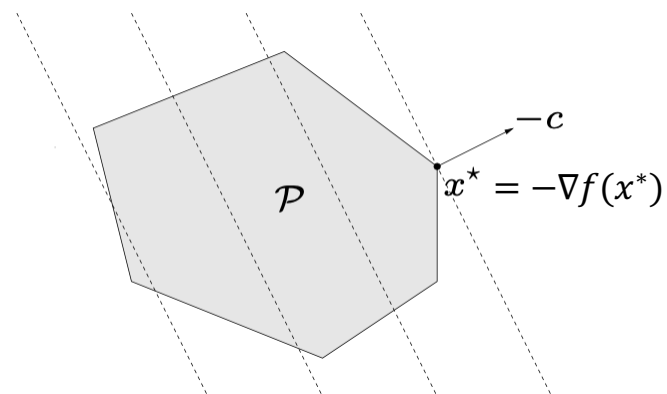
$$\begin{array}{ll}\min & c^T x + d \\ \text{s.t.} & Gx \leq h \\ & Ax = b\end{array}$$

$$c \in \mathbb{R}^n, d \in \mathbb{R}$$

$$G \in \mathbb{R}^{m \times n}, h \in \mathbb{R}^m$$

$$A \in \mathbb{R}^{k \times n}, b \in \mathbb{R}^k$$

- 凸优化问题，目标函数和约束都是仿射函数
- 可行解集是多面体，虚线为等高线
- 在P中任意找一点与 x^* 梯度的内积一定小于0
- 单纯形法 (simplex methods)
 - 任意一个线性规划问题，其最优解一定在边界上



应用举例1-运输问题



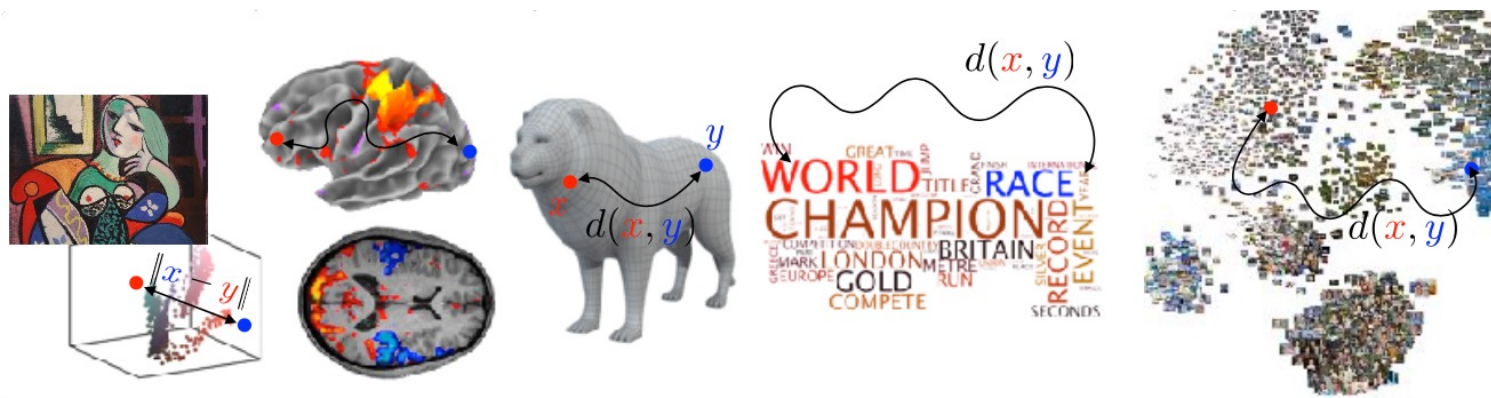
- 假设有 I 个港口 P_1, P_2, \dots, P_I , 提供某种商品。有 J 个市场 M_1, M_2, \dots, M_J 需要这种商品。假设港口 P_i 有 s_i 单位的这种商品 ($i = 1, \dots, I$), 市场 M_j 需要 r_j 单位的这种商品, 且总供应与总需求相等, 即 $\sum_{i=1}^I s_i = \sum_{j=1}^J r_j$ 。令 b_{ij} 为从港口 P_i 运输单位数量商品到市场 M_j 的成本。运输问题是在满足市场需求下使得运输成本最低。

$$\begin{aligned} \min_x \quad & \sum_{i=1}^I \sum_{j=1}^J x_{ij} b_{ij}, \\ \text{s.t.} \quad & \sum_{j=1}^J x_{ij} = s_i, \quad i = 1, 2, \dots, I, \\ & \sum_{i=1}^I x_{ij} = r_j, \quad j = 1, 2, \dots, J, \\ & x_{ij} \geq 0, \quad i = 1, 2, \dots, I, \quad j = 1, 2, \dots, J. \end{aligned}$$

应用举例1-运输问题



- 运输问题还有更一般的情形，即最优运输问题。它关心两个(离散、连续)测度的对应关系。具体地，若测度是离散的，我们想要确定的是离散点之间的对应关系。
- 应用领域包括：计算流体力学，多幅图像之间的颜色转移或图像处理背景下的变形，计算机图形学中的插值方案，以及经济学、通过匹配和均衡问题等、单细胞RNA发育过程中指导分化。
- Domain Adaption: 从源数据分布中学习一个训练良好的模型，并将该模型转换为采用目标数据分布。
- Deep Generative Model: 其目标是将一个固定的分布，例如标准高斯或均匀分布，映射到真实样本的潜在总体分布



应用举例2-马尔科夫决策过程(Markov Decision Process)



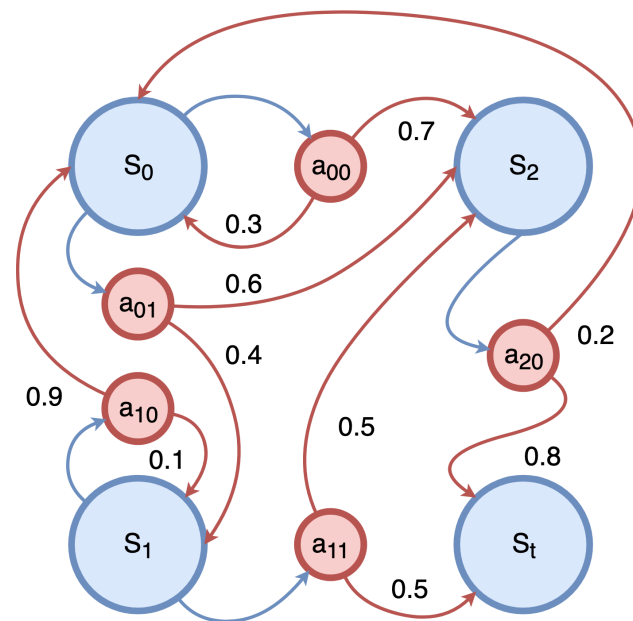
- 智能体对环境进行感知，在每个状态 i ，按照策略 π 实施动作 a ，然后进入下一个状态 j ；获得奖励为 $r(i, a)$ 。其中从状态 i 到状态 j 的概率为

$$P_a(i, j)$$

- $V(i)$ 是向量 V 的第 i 个分量，表示从状态 i 出发得到的累积奖励
- γ 为折扣因子，一般取0.9-1之间

$$\max_{V \in \mathbb{R}^{|S|}} \sum_i V(i)$$

$$\text{s.t. } V(i) \geq \sum_j P_a(i, j) (r(i, a) + \gamma V(j)), \forall i \in S, \forall a \in \mathcal{A},$$

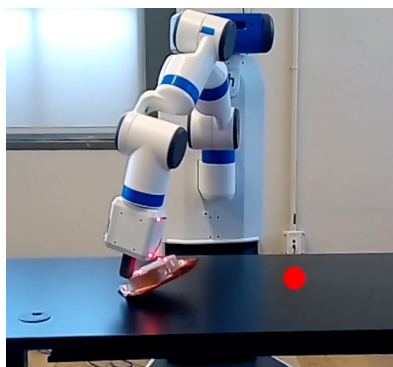
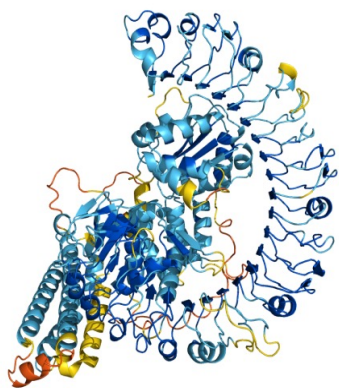
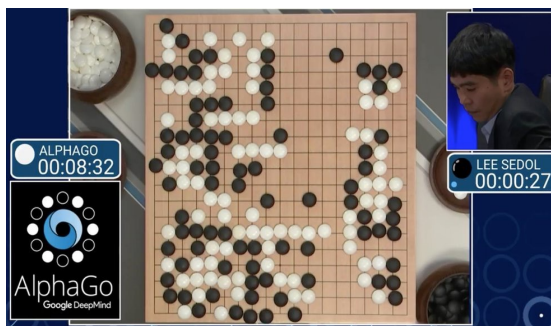


$$\begin{aligned} R(S_1, S_0) &= -1.0 & R(S_0, S_2) &= 0.5 & R(S_2, S_1) &= 0.5 \\ R(S_1, S_2) &= -0.5 & R(S_0, S_1) &= 1.0 & R(S_1, S_1) &= 1.0 \dots \end{aligned}$$

应用举例2-马尔科夫决策过程(Markov Decision Process)



- 马尔科夫决策过程是强化学习的最基本模型。在实际问题中，转移概率矩阵 $P_a(i, j)$ 往往是未知，需要通过探索-利用来求解最优的累积奖励

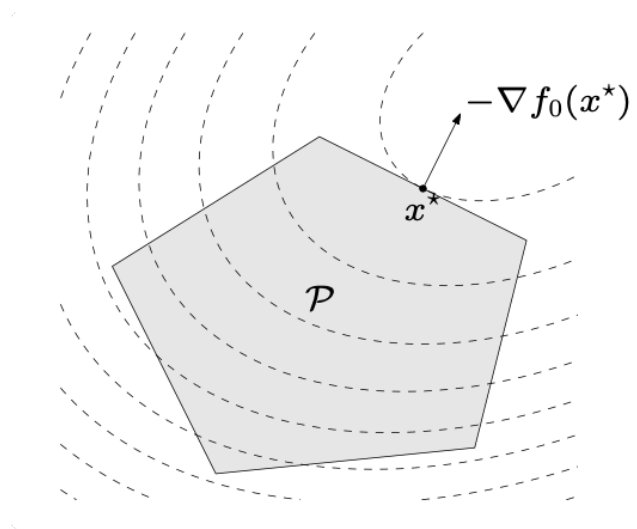


应用举例3-二次规划 (Quadratic Programming)



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

- $P \in S_+^n$, 凸优化问题
- 可行解集是多面体, 虚线为等高线



二次约束二次规划 (QCQP)



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

- $P_i \in S_+^n$, 目标函数与不等式约束都是二次凸函数
- 可行解集为m个椭球和一个仿射集的交集

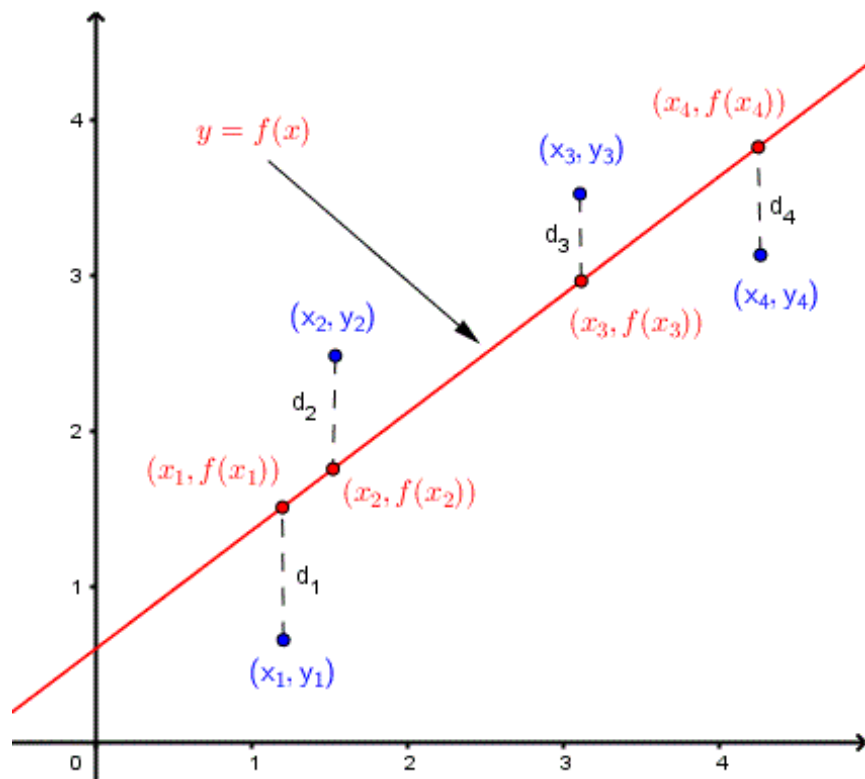
例-最小二乘问题 (Least Squares Problem)



$$\min \|Xw - y\|_2^2$$

$$\min \quad w^T X^T X w - 2y^T X w + y^T y$$

- 无约束QP问题
- 解析解: $(X^T X)^{-1} X^T y$
- 假设有约束 $u_i \leq w_i \leq l_i$, 则是一个具有等式约束的QP问题



例-投资组合问题 (portfolio problem)

$$\begin{aligned} \max \quad & \sum_{i=1}^n x_i p_i . \\ \text{s. t.} \quad & \sum_{i=1}^n x_i \leq B, x_i \geq 0 \end{aligned}$$

其中 x_i 为第 i 种产品的投资额, B 为总资金, p_i 为收益率

- 在实际情况中, p_i 一般为概率分布
 - $\bar{p} = [1.05, 1.05, 1]$
 - $\Sigma = [1.2, 2, 0]$

$$\begin{aligned} \min \quad & \sum_{i=1}^n X^T \Sigma X \\ \text{s. t.} \quad & \bar{p}^T X \geq r_{\min}, I^T X = B, X \geq 0 \end{aligned}$$

应用举例4-复合优化问题

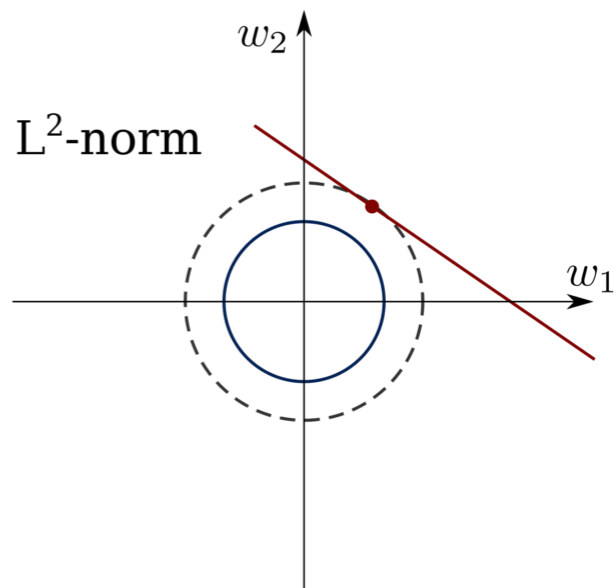
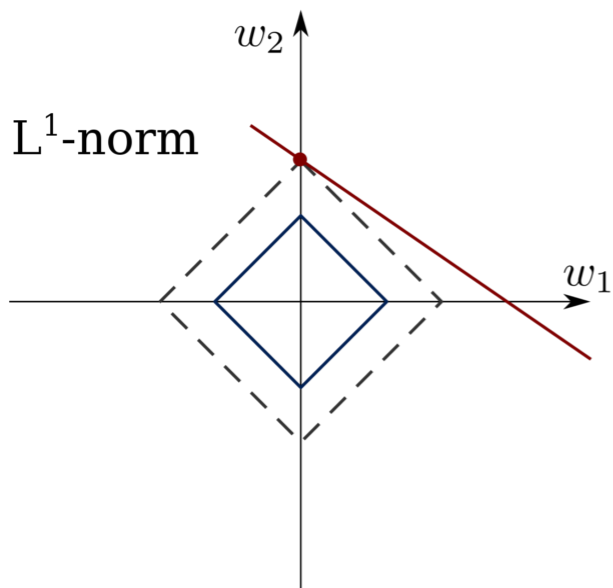
$$\min \psi(x) = f(x) + h(x)$$

其中， $f(x)$ 一般是凸的、光滑的， $h(x)$ 一般是凸的，但不一定光滑

• 应用场景：

- 岭回归： $f(x)$ 为最小二乘， $h(x) = \|x\|_2$
- Lasso： $f(x)$ 为最小二乘， $h(x) = \|x\|_1$
- l1 范数正则化逻辑回归： $f(x) = \sum_{i=1}^m \ln(1 + \exp(-b_i \cdot a_i^T x))$, $h(x) = \|x\|_1$
- l1 范数正则化支持向量机： $f(x) = C \sum_{i=1}^m \max\{1 - b_i(a_i^T x + y), 0\}$, $h(x) = \|x\|_1$

l_1 正则化与 l_2 正则化



应用举例5-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

影	电影 1	电影 2	电影 3	电影 4	...	电影 n
用户 1	4	?	?	3	...	?
用户 2	?	2	4	?	...	?
用户 3	3	?	?	?	...	?
用户 4	2	?	5	?	...	?
\vdots	\vdots	\vdots	\vdots	\vdots		\vdots
用户 m	?	3	?	4	...	?

$$\min_{X \in \mathbb{R}^{m \times n}} \|X\|_*,$$

$$\text{s.t. } X_{ij} = M_{ij}, (i, j) \in \Omega.$$

核范数 (nuclear norm) :

$$\|X\|_* = \sum_{i=1}^r \sigma_i$$

其中 σ_i 为矩阵 X 的所有非零奇异值

应用举例6-整数规划（非凸）



$$\min \quad c^{\top} x$$

$$\text{s.t.} \quad Ax \leq b,$$

$$x_i \geq 0, \quad i = 1, 2, \dots, n,$$

$$x_i \in \mathbb{Z}, \quad i = 1, 2, \dots, n,$$

- 其中 \mathbb{Z} 为所有的整数集合

例-仓库位置选取问题

- 一个管理者需要在 m 个可能的位置上修建一些仓库去满足 n 个消费者的需求
- 根据消费者的需求以及仓库到消费者之间的运输成本，来确定所要修建的仓库的位置，以及如何以最低成本将物资从仓库运输到消费者

$$y_i = \begin{cases} 1, & \text{如果仓库 } i \text{ 被修建,} \\ 0, & \text{如果仓库 } i \text{ 未被修建,} \end{cases}$$

f_i = 仓库 i 的固定运营成本, 比如租赁费,

c_{ij} = 从仓库 i 运输物资到消费者 j 的运输成本.

x_{ij} 为从仓库 i 运输到消费者 j 的物资数量

d_j 为消费者需求

u_i 仓库的容量

$$\begin{aligned} \min_{x,y} \quad & \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} + \sum_{i=1}^m f_i y_i, \\ \text{s.t.} \quad & \sum_{i=1}^m x_{ij} = d_j, \quad j = 1, 2, \dots, n, \\ & \sum_{j=1}^n x_{ij} - y_i u_i \leq 0, \quad i = 1, 2, \dots, m, \\ & x_{ij} \geq 0, \quad i = 1, \dots, m; \quad j = 1, 2, \dots, n, \\ & y_i = 0 \text{ 或者 } 1, \quad i = 1, \dots, m, \end{aligned}$$

- 考虑以下的优化问题

$$\begin{aligned} \min \quad & f_0(x_1, x_2). \\ \text{s.t.} \quad & 2x_1 + x_2 \geq 1 \\ & x_1 + 3x_2 \geq 1 \\ & x_1 \geq 0, x_2 \geq 0 \end{aligned}$$

计算以上优化问题的可行解集，并针对以下的目标函数，计算最优解集与最优值

- (a) $f_0(x_1, x_2) = x_1 + x_2$
- (b) $f_0(x_1, x_2) = -x_1 - x_2$
- (c) $f_0(x_1, x_2) = x_1$
- (d) $f_0(x_1, x_2) = \max\{x_1, x_2\}$
- (e) $f_0(x_1, x_2) = x_1^2 + 9x_2^2$

实验一



- Python, numpy, CVXOPT (<https://cvxopt.org/>)
 - <https://cvxopt.org/userguide/coneprog.html#linear-programming>
 - <https://cvxopt.org/userguide/coneprog.html#quadratic-programming>
- Python基本语法
 - 变量和数据类型
 - 运算符和表达式
 - 字符串
 - 流程控制
- Python组合数据类型
- Python函数
- 数值计算库NumPy

- 1. 食谱问题

- 从 n 种食品中分别选择每种食品的食用量 x_1, \dots, x_n
 - 第 j 种食品的卡路里为 c_j , 包含的第 i 种营养的量为 a_{ij}
 - 为了保持健康, 第 i 种营养的需求至少为 b_i
- 目标: 最小化卡路里的食谱

- 2. 最小二乘问题



谢谢！！

2.8 (鞍点问题) 设函数 $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ 满足如下性质：当固定 $z \in \mathbb{R}^m$ 时， $f(x, z)$ 关于 x 为凸函数；当固定 $x \in \mathbb{R}^n$ 时， $f(x, z)$ 关于 z 是凹函数，则称 f 为**凸 - 凹函数**。

- (a) 设 f 二阶可导，试利用海瑟矩阵 $\nabla^2 f$ 给出 f 为凸 - 凹函数的一个二阶条件；
- (b) 设 f 为凸 - 凹函数且可微，且在点 (\bar{x}, \bar{z}) 处满足 $\nabla f(\bar{x}, \bar{z}) = 0$ ，求证：对任意 x 和 z ，如下鞍点性质成立：

$$f(\bar{x}, z) \leq f(\bar{x}, \bar{z}) \leq f(x, \bar{z}).$$

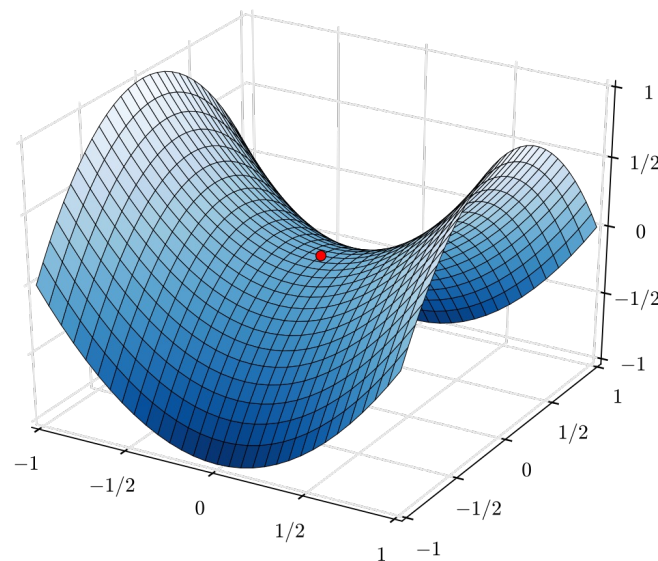
进一步证明 f 满足极小 - 极大性质：

$$\sup_z \inf_x f(x, z) = \inf_x \sup_z f(x, z).$$

- (c) 设 f 可微但不一定是凸 - 凹函数，且在点 (\bar{x}, \bar{z}) 处满足鞍点性质

$$f(\bar{x}, z) \leq f(\bar{x}, \bar{z}) \leq f(x, \bar{z}), \quad \forall x, z$$

求证： $\nabla f(\bar{x}, \bar{z}) = 0$.



(a) 设 f 的海瑟矩阵是

$$\nabla^2 f = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

其中

$$A_{11} \in \mathbb{R}^{n \times n}, A_{12} \in \mathbb{R}^{n \times m}, A_{21} \in \mathbb{R}^{m \times n}, A_{22} \in \mathbb{R}^{m \times m}.$$

则 $\nabla_x^2 f = A_{11}$, $\nabla_z^2 f = A_{22}$. 若固定 z 时 f 关于 x 凸, 则 A_{11} 半正定; 若固定 x 时 f 关于 z 凹, 则 A_{22} 半负定. 这就是由海瑟矩阵给出的二阶条件.

(b) 先证明第一个不等式. 由于 $f(x, z)$ 关于 x 是凸函数, 利用凸函数的性质有

$$f(x, \bar{z}) \geq f(\bar{x}, \bar{z}) + \nabla_x f(\bar{x}, \bar{z})(x - \bar{x}) = f(\bar{x}, \bar{z}),$$

同理可得左半边的不等式.

再证明 $\sup_z \inf_x f(x, z) = f(\bar{x}, \bar{z})$. 首先

$$\inf_x f(x, z) \leq f(\bar{x}, z) \leq f(\bar{x}, \bar{z}),$$

故

$$\sup_z \inf_x f(x, z) \leq f(\bar{x}, \bar{z}),$$

又

$$\sup_z \inf_x f(x, z) \geq \inf_x f(x, \bar{z}) \geq f(\bar{x}, \bar{z}),$$

因此 $\sup_z \inf_x f(x, z) = f(\bar{x}, \bar{z})$. 同理得 $\inf_x \sup_z f(x, z) = f(\bar{x}, \bar{z})$, 因此等式得证.

(c) 只需分别证明 $\nabla_x f(\bar{x}, \bar{z}) = 0$ 和 $\nabla_z f(\bar{x}, \bar{z}) = 0$. 用反证法证明. 先假设 $\nabla_x f(\bar{x}, \bar{z}) \neq 0$, 取向量 $v = (\nabla_x f(\bar{x}, \bar{z})^T, 0)^T$, 对 $f(\bar{x} + tv, \bar{z})$ (其中 $t \neq 0$) 在 (\bar{x}, \bar{z}) 处展开一阶, 得

$$f(\bar{x} + tv, \bar{z}) = f(\bar{x}, \bar{z}) + t \|\nabla_x f(\bar{x}, \bar{z})\|^2 + \mathcal{O}(t^2).$$

若取 $t < 0$ 且绝对值足够小, 就有 $f(\bar{x} + tv, \bar{z}) < f(\bar{x}, \bar{z})$, 与题设矛盾. 因此假设不成立, $\nabla_x f(\bar{x}, \bar{z}) = 0$. 同理可得 $\nabla_z f(\bar{x}, \bar{z}) = 0$. 综上, 命题成立. \square

最优条件（特殊情况）



- 无约束问题：

- x 为最优解的充分必要条件

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

- 等式约束问题：

$$\min f_0(x) \quad \text{s.t.} \quad Ax = b$$

- x 为最优解的充分必要条件，存在 v 使得

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