凸优化问题（1）
Convex Problems

张伯雷

南京邮电大学 计算机学院

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

凸集

- 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

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

- 重要的凸集
  - 超平面、半空间、多面体
  - 对称半正定矩阵集合
  - 球、椭球
- 凸集的证明
  - 定义
  - 保凸的运算：凸集的交集、凸集的仿射变换

凸函数

- 定义

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

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

对所有的$x, y \in \mathrm{dom} f$，$0 \leq \theta \leq 1$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

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

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

凸函数

- 常见的凸函数
  - 负熵： $x\log x$ on R_{++}
  - 范数函数： $||x||_p = (\sum_{i=1}^{n}|x_i|^p)^{\frac{1}{p}}$ for $p \geq 1$;
  - $f(x) = \max\{\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{n}\}, \mathrm{x} \in R^m$
  - …
- 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
  - $f^*(y) = \sup_{x\in\mathrm{dom}\, f} (y^T x - f(x))$

目录

- 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

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

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

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

- 可行解集（feasible set）
  - $X = \{x \in D$ 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）

$$X_{opt} = \{x | x \in \mathrm{X}_f, f_0(x^*) = p^*\}$$

- ε-次优解集（ ε – suboptimal set）

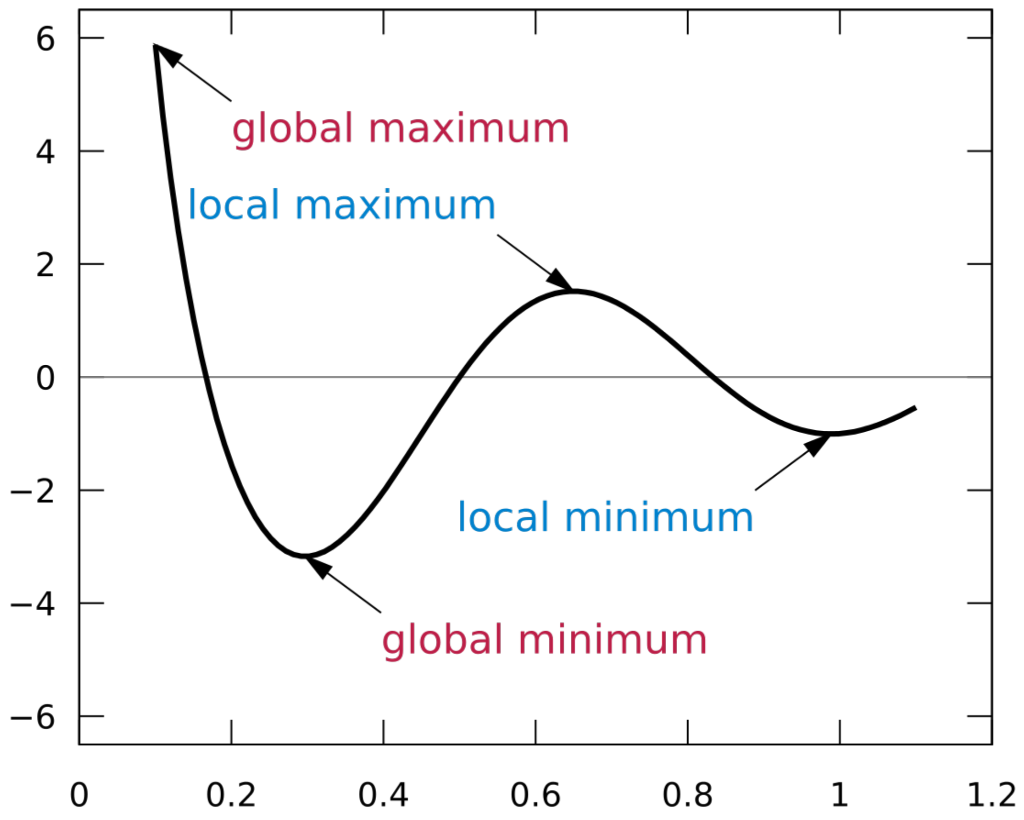
$$X_\varepsilon = \{x | x \in \mathrm{X}_f, f_0(x^*) \le p^* + \varepsilon\}$$

局部最优解（Locally optimal）

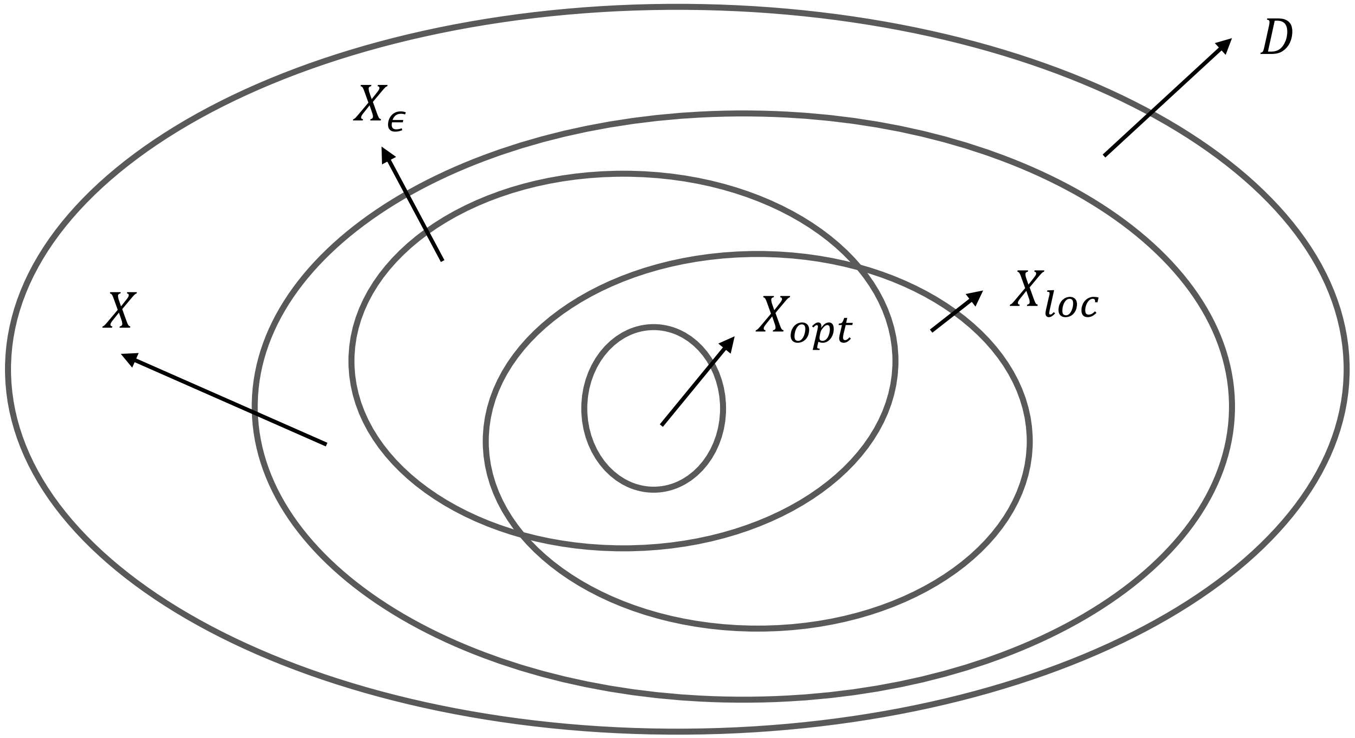
- x为局部最优解，如果 $\exists R>0$使得x为以下问题的最优解：

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

- 所有局部最优解构成的解集为X_{loc}



不同解集的关系



凸优化问题

- （狭义）凸优化问题的标准形式

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

- 其中
  - f_i: $\mathbb{R}^n \to \mathbb{R}, i = 0, \dots, m$：凸函数
  - h_i: $\mathbb{R}^n \to \mathbb{R}, i = 1, \dots, p$：等式约束为仿射函数
- 性质：凸优化问题的可行解集为凸集

例

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

- 是否为标准形式的凸优化问题?
- 等价变换

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

例

- 等价变换

$$
\begin{aligned}
\min.\ & f_0(x)\,. \\
\text{s.t.}\ & s_i \le 0, && i = 1, \dots, m \\
& f_i(x) - s_\mathrm{i} = 0, && i = 1, \dots, m \\
& a_i^T x = b_i, && i = 1, \dots, p
\end{aligned}
$$

- 松弛变量（Slack variable）
  - 引入新的变量s，新的优化问题的优化变量包括x, s

例

$$
\begin{aligned}
\max.\ & f_0(x)\,. \\
\text{s.t.}\ & f_i(x) \le 0, \qquad i = 1, \dots, m \\
& a_i^T x = b_i, \qquad i = 1, \dots, p
\end{aligned}
$$

- 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

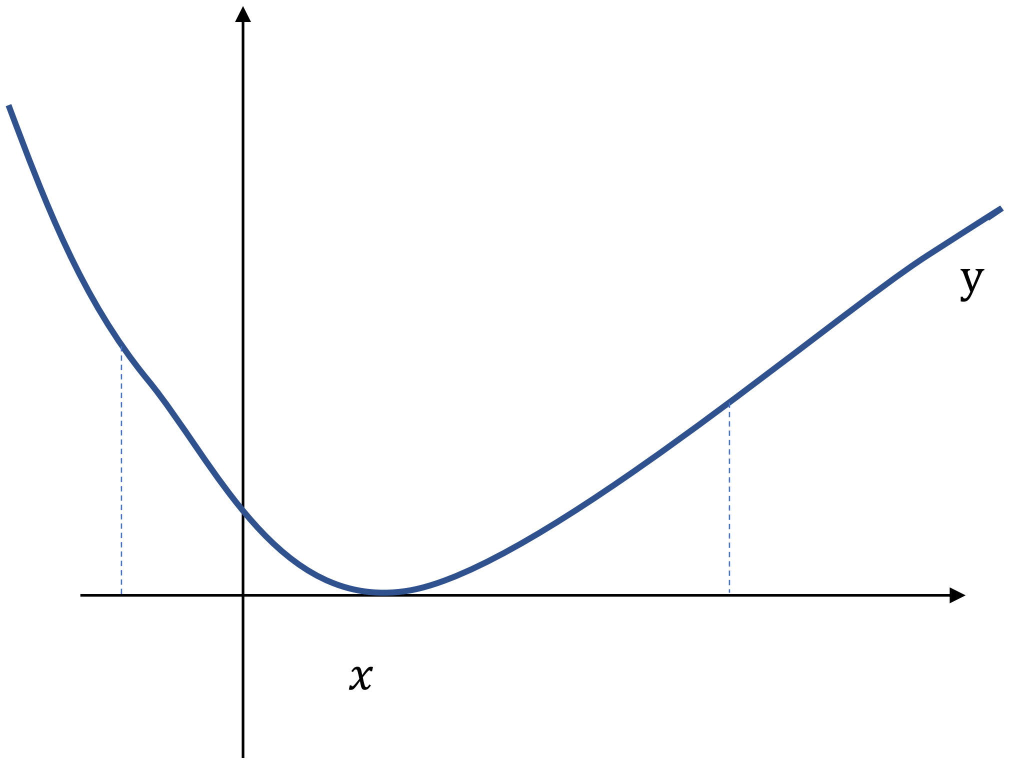
- 局部最优=全部最优
  - 局部最优：存在$R>0$使得x为以下问题的最优解：

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

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

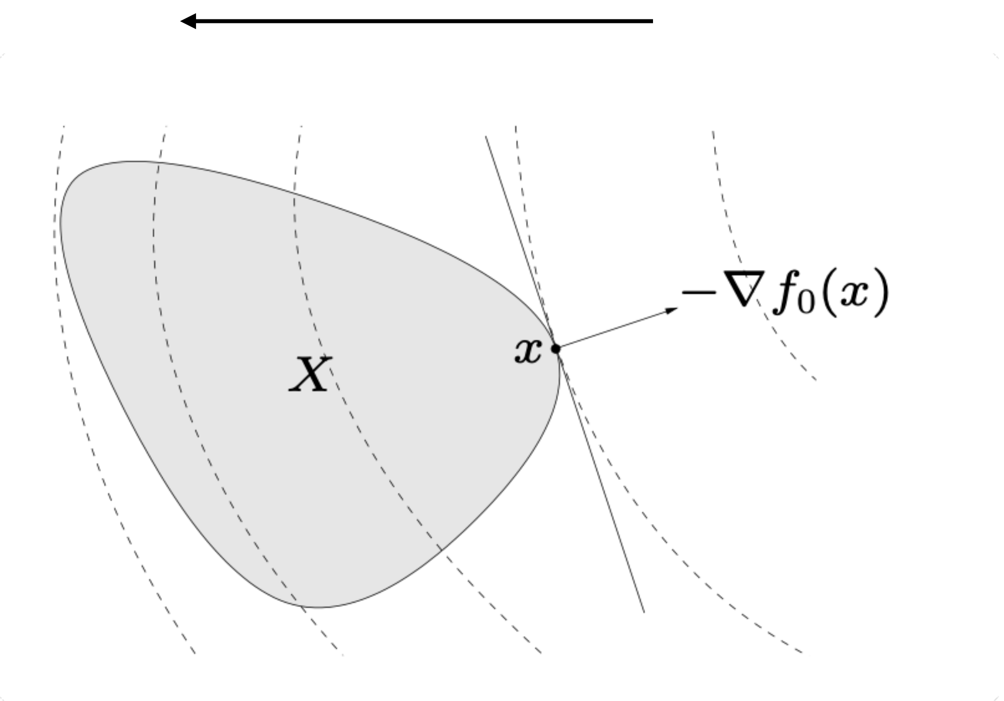
重要性质：局部最优=全部最优

- 证明：



最优条件

- 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 $\Longrightarrow$ $domf$为凸集
    - $f_0(y) \geq f_0(x) + \nabla f_0^T(x)(y-x), \forall x, y \in \mathrm{dom}\, f$
  - 最优条件
    - $x \in X$最优 $\Leftrightarrow$ $\nabla f_0(x)^T(y-x) \geq 0$ 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

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

- 等式约束问题：

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

  - x为最优解的充分必要条件，存在v使得

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

线性规划（Linear Programming）

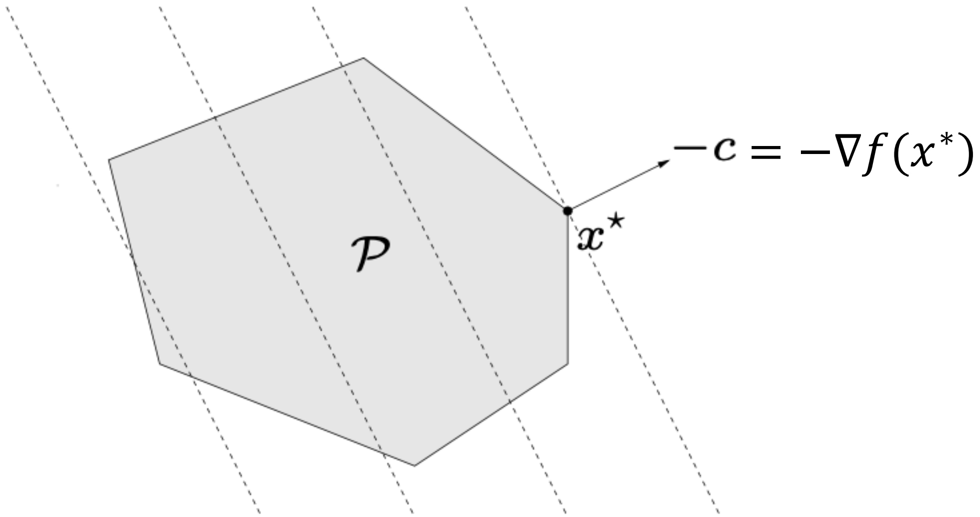
$$\begin{aligned} \min \quad & c^T x + d \\ \text{s.t.} \quad & Gx \le h \\ & Ax = b \end{aligned}$$

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

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

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

- 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



发展历史

- 1930-1940：线性规划最先在第二次世界大战时被提出，用于最大化资源的利用效 率：指按照既定的时刻表去执行任务或者用最佳方式做人员部署。
  - Dantzig（单纯形法）
  - Von Neumann（最优性条件）
  - Kantorovich（最优化问题的形式）

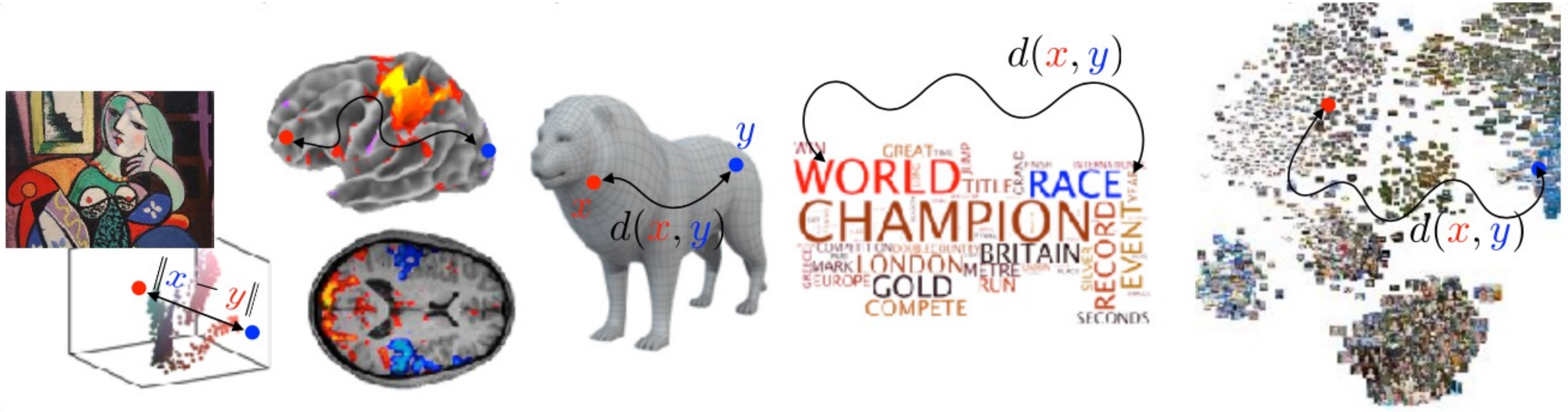
应用举例1-运输问题

- 假设有I个港口$P_1,P_2,\cdots,P_I$，提供某种商品。有J个市场$M_1,M_2,\cdots,M_J$需要这种商品.假设港口P_i有s_i 单 位的这种商品($i = 1,\cdots,I$)，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$\sum_{i=1}^{I} s_i = \sum_{i=1}^{I} r_j$。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.

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

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

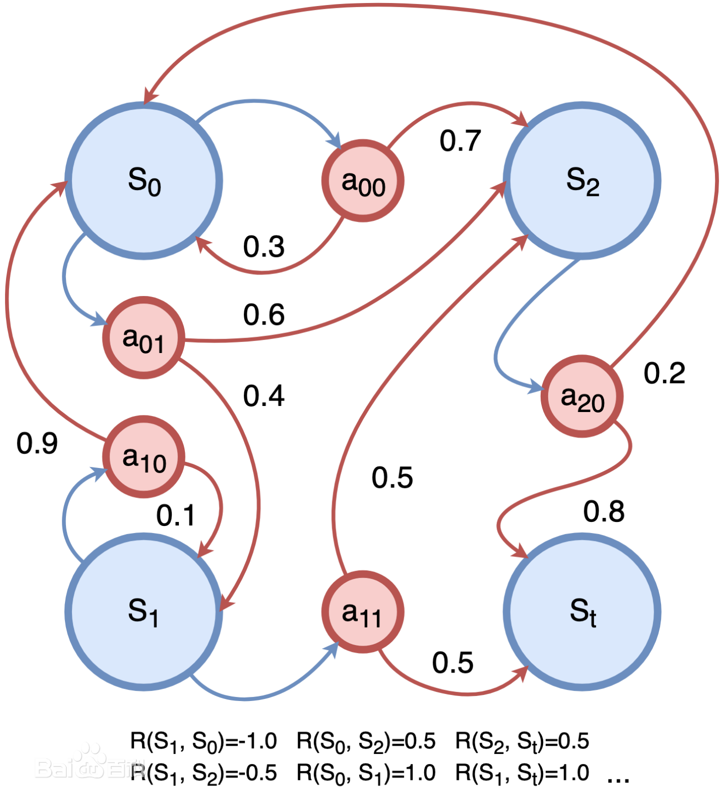


应用举例2-马尔科夫决策过程(Markov Decision Process)

- 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$\mathrm{P}_a(i,j)$
  - $V(i)$ 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

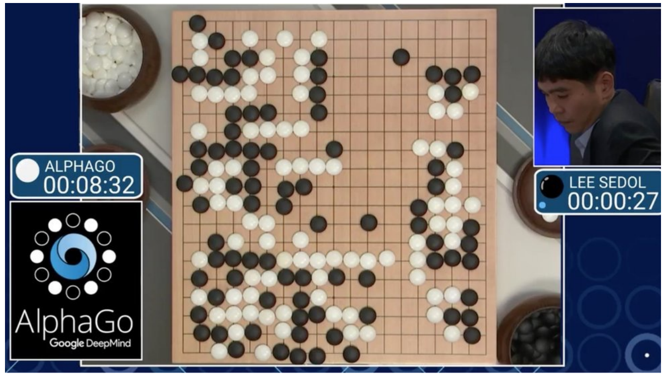
$$\max_{V\in\mathbb{R}^{|\mathcal{S}|}} \quad \sum_i V(i)$$

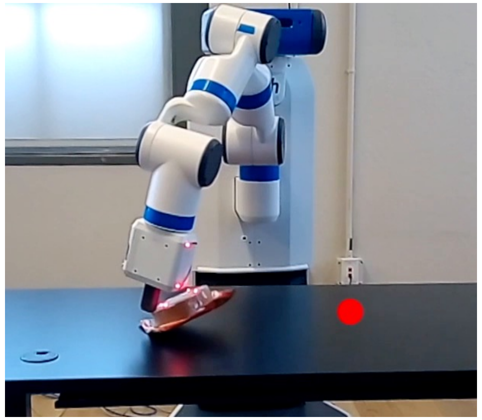
$$\text{s.t.} \quad V(i) \geq \sum_j P_a(i,j)\left(r(i,a)+\gamma V(j)\right), \forall i \in \mathcal{S},\ \forall a \in \mathcal{A},$$



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





谢谢！！