1-1　问题模型与基本概念

问题实例1：线性规划

问题描述 某公司生产甲乙两种商品，每件甲商品要耗A种原料3kg, B种原料3kg, 产值为180元; 每件乙商品要耗A种原料4.5kg，B种原料1.5kg, 产值为150元. 现公司共有A种原料900kg, B种原料600kg, 试确定甲乙两种商品共生产多少件, 使得公司的总产值最大.

建模分析:

变量： 生产甲商品x_1件, 乙商品x_2件, 故公司总产值为

$$180x_1 + 150x_2(\text{目标函数})$$

问题实例1：线性规划

约束： 考虑到A,B两种原材料总量限制，故x_1, x_2需满足如下不等式约束条件:

$$\begin{cases} 3x_1 + 4.5x_2 \leq 900 \\ 3x_1 + 1.5x_2 \leq 600 \end{cases}$$

非负约束： $x_1 \geq 0, x_2 \geq 0.$

线性规划模型：

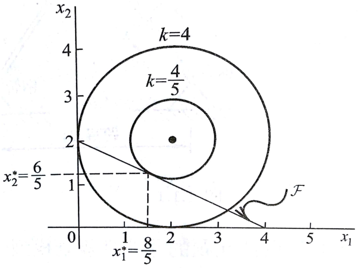
$$\begin{aligned} \max \quad & z = 180x_1 + 150x_2 \\ \text{s.t.} \quad & 3x_1 + 4.5x_2 \leq 900 \\ & 3x_1 + 1.5x_2 \leq 600 \\ & x_1 \geq 0, x_2 \geq 0 \end{aligned}$$

问题实例2：非线性规划

问题模型:

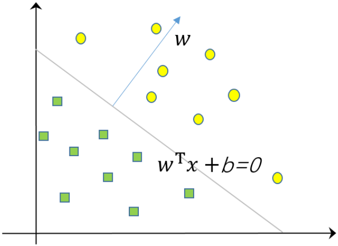
$$\begin{aligned} \min_{x_1,x_2} \quad & (x_1-2)^2+(x_2-2)^2 \\ \text{s.t.} \quad & x_1+2x_2=4 \\ & x_1\ge 0, x_2\ge 0 \end{aligned}$$

求解分析： 可行域$\Omega=\{(x_1,x_2)|x_1+2x_2=4, x_1\ge 0, x_2\ge 0\}$



实例问题3：感知机二分类问题

感知机（1957年）是较早的二分类的线性分类模型，也是后来神经网络和支持向量机的基础。感知机的目标为确定一个超平面$\boldsymbol{w}^T\boldsymbol{x}+b=0$，使得将训练集中的正类点和负类点分开。



感知机的分类模型如下:

$$f(\boldsymbol{x})=\text{sign}(\boldsymbol{w}^T\boldsymbol{x}+b)=\begin{cases}1 & \boldsymbol{w}^T\boldsymbol{x}+b\geq 0\\ -1 & \text{otherwise}\end{cases}$$

实例问题3：感知机二分类问题

为确定分类模型中参数$\boldsymbol{w}, b$, 考虑极小化误分类样本点到超平面的距离，即

$$-\frac{1}{\|\boldsymbol{w}\|}\sum_{i=1}^{N} y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)$$

其中$(\boldsymbol{x}_i, y_i), i = 1, \ldots, N$, 为样本点与相应分类结果对。这里，$y_i = 1$或者$y_i = -1$。

注意成倍增加$(\boldsymbol{w}, b)$不改变超平面, 故感知机只需极小化如下损失函数:

$$\text{minimize} \quad L(\boldsymbol{w}, b) = -\sum_{i=1}^{N} y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)$$

问题4：投资规划问题

假定一个投资方案里有n个投资项目，项目i的回报率为：

$$r_i, \quad i = 1, 2, \cdots, n$$

单个投资项目i的通常假定满足正态分布的随机变量，其均值和方差表分别表示如下：

$$\mu_i = E[r_i](\text{收益}) \qquad \sigma_i^2 = E[(r_i - \mu_i)^2](\text{风险})$$

假定投资在项目i上的资金比为：$x_i, i = 1, \ldots, n$，则投资方案的利润回报与期望回报分别为：

$$R = \sum_{i=1}^{n} x_i r_i. \quad E[R] = E[\sum_{i=1}^{n} x_i r_i] = \sum_{i=1}^{n} x_i E[r_i] = \boldsymbol{x}^T \boldsymbol{\mu}$$

问题4：投资规划问题

整个投资的方差由下式给出：

$$E[(R-E(R))^2] = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_i \sigma_j \rho_{ij} = \boldsymbol{x}^T G \boldsymbol{x},$$

其中项目回报协方差为：$\rho_{ij} = \frac{E[(r_i-\mu_i)(r_j-\mu_j)]}{\sigma_i\sigma_j}$, $i,j = 1,2,\cdots,n$. 矩阵$G_{n\times n}$定义为：$G_{ij} = \rho_{ij}\sigma_i\sigma_j$. 可以证明$G$是半正定的。

最优组合投资Markowitz模型：

$$\begin{aligned} \max \quad & \boldsymbol{x}^T \boldsymbol{\mu} - \kappa \boldsymbol{x}^T G \boldsymbol{x} \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = 1, \boldsymbol{x} \geq 0, \end{aligned}$$

这里风险允许参数$\kappa \in [0,\infty)$，依赖于投资者的偏好。

最优化问题数学模型

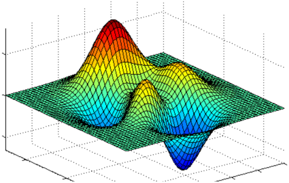
一般优化问题数学模型可表述为：

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & c_i(\boldsymbol{x}) = 0, i \in \mathcal{E}; c_j(\boldsymbol{x}) \geq 0, j \in \mathcal{I} \end{aligned}$$

其中：

- 决策变量 - $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T \in R^n$;
- 目标函数 - $f: \ R^n \to R$;
- 约束函数 - $c_i: \ R^n \to R, i \in \mathcal{E} \cup \mathcal{I}$.

  $\mathcal{E} = \{1, ..., m_e\}; \mathcal{I} = \{m_e + 1, \ldots, m\}$



- 可行域：$\Omega = \{\boldsymbol{x} | c_i(\boldsymbol{x}) = 0, i \in \mathcal{E}; c_j(\boldsymbol{x}) \geq 0, j \in \mathcal{I}\}$

缩写min 对应minimize （极小化）s.t. 对应subject to（受约束）

问题转换

- **极大极小类型转换：**

$$\max f(\boldsymbol{x}) \leftrightarrow \min -f(\boldsymbol{x})$$

- **约束类型转换：**
  - 不等式符号转换：

$$c_i(\boldsymbol{x}) \leq 0 \leftrightarrow -c_i(\boldsymbol{x}) \geq 0$$

  - 一般形式转换标准形式：

$$a(\boldsymbol{x}) \leq b(\boldsymbol{x}) + c \leftrightarrow h(\boldsymbol{x}) = b(\boldsymbol{x}) - a(\boldsymbol{x}) + c \geq 0$$

优化问题分类

- 连续性与光滑性：
  - 连续与离散(组合优化)、光滑与非光滑
- 约束条件：
  - 无约束优化与约束优化
- 目标与约束的线性性：
  - 线性规划与非线性规划
- 变量随机性：
  - 确定性优化与不确定性优化(随机优化)
- 优化目标数量：
  - 单目标优化与多目标优化

约束优化问题

- 等式约束优化问题(equality constraint)：

$$\begin{cases} \min & f(\boldsymbol{x}), \\ \text{s.t.} & c_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \cdots, m_e.(m = m_e) \end{cases}$$

- 不等约束优化问题(inequality constraint)：

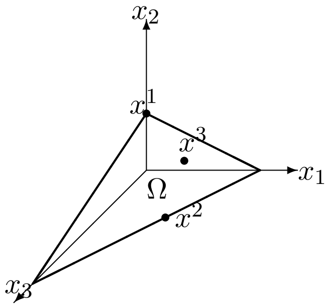
$$\begin{cases} \min & f(\boldsymbol{x}), \\ \text{s.t.} & c_i(\boldsymbol{x}) \geq 0, \quad i = 1, 2, \cdots, m.(m_e = 0) \end{cases}$$

- 混合约束优化问题 既有等式约束，又有不等式约束。

注意： 约束优化问题经常转化为无约束优化问题来求解。

约束优化可行域

定义1.1.1： 若点$\boldsymbol{x} \in R^n$满足优化问题模型中所有约束条件, 则称$\boldsymbol{x}$为**可行点** (Feasible Point)。可行点的全体称为**可行域**(Feasible Region)。



例子： 可行域$\Omega = \{\boldsymbol{x} \in R^3 | c_1(\boldsymbol{x}) = 2x_1 + 3x_2 + x_3 - 6 = 0, x_i \geq 0, i = 1, 2, 3\}$的边界由粗线表示。

特殊优化问题：线性规划

线性规划(Linear Programming)数学模型：

$$\begin{aligned} \min \quad & \boldsymbol{c}^T\boldsymbol{x}, \\ \text{s.t.} \quad & A_1\boldsymbol{x} = \boldsymbol{b}_1, A_2\boldsymbol{x} \geq \boldsymbol{b}_2, \end{aligned}$$

其中：

- $\boldsymbol{c} = (c_1, c_2, \cdots, c_n)^T$, $\boldsymbol{b}_1 = (b_1, b_2, \cdots, b_m)^T$, $\boldsymbol{b}_2 = (b_{m+1}, \cdots, b_p)^T$,
- 约束矩阵：

$$A_1 = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, A_2 = \begin{bmatrix} a_{m+1,1} & a_{m+1,2} & \cdots & a_{m+1,n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{bmatrix}.$$

二次规划(Quadratic Programming)数学模型:

$$
\begin{array}{ll}
\text{min} & q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} + d, \\
\text{s.t.} & A_1\boldsymbol{x} = \boldsymbol{b}_1; A_2\boldsymbol{x} \geq \boldsymbol{b}_2
\end{array}
$$

其中：$A_1, A_2, \boldsymbol{b}_1, \boldsymbol{b}_2$的表示与线性规划模型类似，$\boldsymbol{c} = (c_1, c_2, \cdots, c_n)^T$，$d$为标量，$G$为$n \times n$阶对称矩阵：

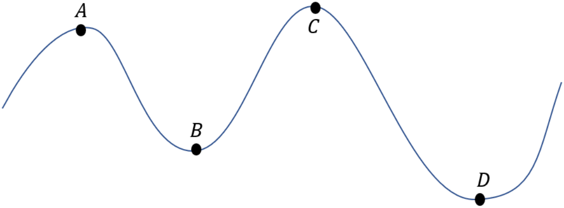
$$
G = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1n} \\ G_{21} & G_{22} & \cdots & G_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ G_{n1} & G_{n2} & \cdots & G_{nn} \end{bmatrix},
$$

满足$G_{ij} = G_{ji}, \ \forall\ i \neq j$.

全局最优和局部最优

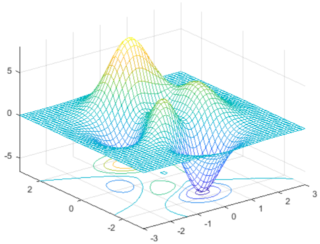
定义1.1.2： 对于任意$\boldsymbol{x} \in \Omega$，若可行点$\boldsymbol{x}^* \in \Omega$满足$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$，则称$\boldsymbol{x}^*$为最优化问题的**全局最优解**(或**全局极小点**)。若对于任意$\boldsymbol{x} \in \Omega$且$\boldsymbol{x} \neq \boldsymbol{x}^*$，有$f(\boldsymbol{x}^*) < f(\boldsymbol{x})$，则称$\boldsymbol{x}^*$为**严格全局最优解**(或**严格全局极小点**)。

定义1.1.3： 若可行点$\boldsymbol{x}^*$的邻域为$\mathcal{N}(\boldsymbol{x}^*) = \{\boldsymbol{x} | \|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \delta\}$。若任意$\boldsymbol{x} \in \mathcal{N}(\boldsymbol{x}^*) \cap \Omega$，满足$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$,则称$\boldsymbol{x}^*$为最优化问题的**局部最优解**(或**局部极小点**)。若对于任意$\boldsymbol{x} \in \mathcal{N}(\boldsymbol{x}^*) \cap \Omega$且$\boldsymbol{x} \neq \boldsymbol{x}^*$，有$f(\boldsymbol{x}^*) < f(\boldsymbol{x})$，则称$\boldsymbol{x}^*$为**严格局部最优解**(**严格局部极小点**)。



函数图形与等高线

欲求函数的极小值，通过函数图形可直观地了解函数的特性。对于二维问题，所谓等高线就是函数值取常数值的所有点的集合，即$f(\boldsymbol{x}) = c$。常数值c取不同值对应一族曲线。



若目标函数为连续的单值函数，则具有以下性质：(i) 不同等值线不相交; (2) 除极值点外，等值线不会中断; (3) 等值线稠密(稀疏)的地方，函数值变化较快(慢); (4) 极值点附近，等值线近似地呈现同心椭圆族。

梯度与海森矩阵

在研究最优性条件与优化算法时，通常假定目标函数或者约束函数的导数存在，比如：

- **一阶梯度向量**：
$$\boldsymbol{g}(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \ldots, \frac{\partial f(\boldsymbol{x})}{\partial x_n}\right]$$
- **二阶海森矩阵**：
$$G(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}$$

在最优解$\boldsymbol{x}^*$处，通常使用以下记号表示$\boldsymbol{x}^*$的函数值/梯度/海森矩阵：
$$f^* = f(\boldsymbol{x}^*), \boldsymbol{g}^* = \boldsymbol{g}(\boldsymbol{x}^*), G^* = G(\boldsymbol{x}^*).$$

梯度与海森矩阵

梯度的性质：

- 若$\nabla f(\boldsymbol{x}) \neq 0$, 则$\nabla f(\boldsymbol{x})$必与过$\boldsymbol{x}$点的等值线垂直;
- 沿梯度方向函数具有最大的变化率。

海森矩阵的性质：

- 海森矩阵可描述函数的局部曲率;
- 海森矩阵具有对称性;
- 海森矩阵的正定性可用于判断函数的极值;
- 函数二阶近似中需用到海森矩阵。

一阶方向导数

定义1.1.4： 对$\forall \boldsymbol{d} \in R^n \backslash \{\mathbf{0}\}$(非零方向向量)，若极限:

$$\lim_{\alpha \to 0^+} \frac{f(\boldsymbol{x} + \alpha \boldsymbol{d}) - f(\boldsymbol{x})}{\alpha \|\boldsymbol{d}\|}$$

存在，则称该极限为函数f在$\boldsymbol{x}$处沿着方向$\boldsymbol{d}$的方向导数，记为：$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{d}}$。方向导数是$f(\boldsymbol{x})$在点$\boldsymbol{x}$处沿着$\boldsymbol{d}$的函数值变化率。

利用一阶方向导数($f \in C^1$)可以定义二阶方向导数($f \in C^2$)，二者具体计算公式如下：

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{d}} = \frac{1}{\|\boldsymbol{d}\|} g(\boldsymbol{x})^T \boldsymbol{d}$$

$$\frac{\partial f^2(\boldsymbol{x})}{\partial \boldsymbol{d}^2} = \frac{1}{\|\boldsymbol{d}\|^2} \boldsymbol{d}^T G(\boldsymbol{x})^T \boldsymbol{d}$$