最优化理论与应用

成像科学与图像分析研究中心 刘晓云
主楼C2-505

教材与参考书

- 教材

最优化理论与算法 陈宝林 清华大学出版社

- 参考书

1. 最优化计算原理与算法程序设计
   栗塔山 国防科学出版社
2. 最优化理论与方法 袁亚湘 科学出版社
3. 精通MATLAB最优化计算 龚纯 王正林
   电子工业出版社
4. 数值最优化Numerical Optimization (美)劳斯特(Nocedal.J.)等著

第一章 引言

- 最优化理论和方法是近几十多年来发展十分迅速的一个数学分支。
- 在数学上，最优化是一种求极值的方法。
- 最优化已经广泛的渗透到工程、经济、电子技术等领域。

- 在实际生活当中，人们做任何事情，不管是分析问题，还是进行决策，都要用一种标准衡量一下是否达到了最优。
- 在各种科学问题、工程问题、生产管理、社会经济问题中，人们总是希望在有限的资源条件下，用尽可能小的代价，获得最大的收获。

最优化问题的研究已经有很多年的历史。

以前解决最优化问题的数学方法只限于古典求导方法和变分法（求无约束极值问题）,Lagrange乘子法解决等式约束下的条件极值问题。

计算机技术的出现，使得人们研究出了许多最优化方法和算法用以解决以前难以解决的问题。

第一节 基本概念

- 最优化方法　研究在一定限制条件下, 选择某种方案, 以达到最优目标的一门学科, 它是一个重要的数学分支。
- 最优方案　达到最优目标的方案。
- 最优化理论　最优化方法的理论。

“把事情做的更好，就是优化”----最优化首先应该学习的是理念，然后才是模型，方法，理论

模型与分类

几乎所有的优化问题都可以概括为这样的数学模型：

给定一个集合 D (称为可行集),和该集合上定义的实值函数$f(x)$,需计算函数在可行集上的极值。

$$\min_{x\in D} f(x) \quad 或 \quad \max_{x\in D} f(x)$$

模型分类：按可行集性质分类

- 可行集中的元素是有限的-----“组合优化”或“网络规划”（如：最短路问题，最小最大流）
- 可行集是有限维空间中的一个连续子集---线性或非线性规划

- 可行集中的元素是依赖于时间的决策序列——“动态规划”
- 可行集是无穷维空间中的连续子集(如：可行集中的元素是一条由常微分方程组描述的曲线，而目标函数为定积分)——“最优控制”

本课程讨论线性规划与非线性规划问题

本课程主要内容

第一章 引言

第二章 线性规划

第三章 对偶原理与对偶单纯形法

第四章 优化算法的结构与一维搜索

第五章 无约束非线性规划

第六章 约束优化最优性条件

第七章 惩罚函数法

第八章 可行方向法

经典极值问题

包括：

①无约束极值问题

②有约束极值问题

1、无约束极值问题的数学模型

$$\min_{x} f(x)$$

2、约束条件下极值问题的数学模型

$$\begin{aligned} &\min_{x} f(x) \\ &s.t. \quad g_i(x) \geq 0, \quad i = 1, 2, \ldots, m \\ &\qquad\ h_j(x) = 0, \quad j = 1, 2, \ldots, l \end{aligned}$$

其中，极大值问题可以转化为极小值问题来进行求解。如求：

$$\max_{x} f(x)$$

可以转化为：$\min_{x}(-f(x))$

解决优化问题的一般方法步骤

① 前期分析：分析问题，找出要解决的目标，约束条件，并确立最优化的目标。

② 定义变量，建立最优化问题的数学模型，列出目标函数和约束条件。

③ 针对建立的模型，选择合适的求解方法或数学软件。

④ 编写程序，利用计算机求解。

⑤ 对结果进行分析，讨论诸如：结果的合理性、正确性，算法的收敛性，模型的适用性和通用性，算法效率与误差等。

例:某豆腐店用黄豆制作两种不同口感的豆腐出售。制作口感较鲜嫩的豆腐每千克需要0.3千克一级黄豆及0.5千克二级黄豆，售价10元；制作口感较厚实的豆腐每千克需要0.4千克一级黄豆及0.2千克二级黄豆，售价5元。现小店购入9千克一级黄豆和8千克二级黄豆。

问：应如何安排制作计划才能获得最大收益。

一、问题前期分析

该问题是在不超出制作两种不同口感豆腐所需黄豆总量条件下合理安排制作计划，使得售出各种豆腐能获得最大收益。

二、模型假设

1．假设制作的豆腐能全部售出。

2．假设豆腐售价无波动。

变量假设：

设计划制作口感鲜嫩和厚实的豆腐各x_1千克和x_2千克，可获得收益f元。

目标函数：获得的总收益最大。

总收益可表示为：$f=10x_1+5x_2$

受一级黄豆数量限制：$0.3x_1+0.4x_2\le 9$

受二级黄豆数量限制：$0.5x_1+0.2x_2\le 8$

综上分析，得到该问题的数学模型

$$\max f = 10x_1 + 5x_2$$

$$\text{s.t}\quad \begin{cases} 0.3x_1 + 0.4x_2 \le 9 \\ 0.5x_1 + 0.2x_2 \le 8 \\ x_1, x_2 \ge 0 \end{cases}$$

例（运输问题）设某种物资有m个产地 $A_1,A_2,\ldots A_m$,各产地的产量是$a_1,a_2,\ldots,a_m$;有 n个销地 $B_1,B_2,\ldots,B_n$.各销地的销量是$b_1,b_2,\ldots,b_n$.

假定从产地$A_i(i=1,2,\ldots,m)$到销地$B_j(j=1,2,\ldots,n)$运输单位物品的运价是c_{ij}问怎样调运这些物品才能使总运费最小？

如果运输问题的总产量等于总销量，即有

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$$

则称该运输问题为产销平衡问题;

反之，称产销不平衡问题。

令x_{ij}表示由产地A_i运往销地B_j的物品数量，

则产销平衡问题的数学模型为：

$$\min f=\sum_{i=1}^{n}\sum_{j=1}^{m}c_{ij}x_{ij}$$

$$s.t\begin{cases}\sum\limits_{j=1}^{n}x_{ij}=a_i & i=1,2,\cdots,m\\ \sum\limits_{i=1}^{m}x_{ij}=b_j & j=1,2,\cdots n\\ x_{ij}\ge 0 & \begin{array}{l} i=1,2,\cdots,m\\ j=1,2,\cdots n\end{array}\end{cases}$$

第二节 梯度、Hesse阵与 Taylor展式

1）局部极值和全局极值

2）梯度 $$\nabla f=\begin{bmatrix}\frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n}\end{bmatrix}^T$$

3）方向导数：设h为一方向向量，即长度为1的单位向量，可微函数f(x）在 x 处沿方向h的方向导数为:

$$\frac{\partial f}{\partial h}=\lim_{\theta\to 0^+}\frac{f(x+\theta h)-f(x)}{\theta}$$

$$=\nabla f(x)^T h=\left\|\nabla f(x)^T\right\|\cos(\nabla f(x),h)$$

方向导数性质

- 若 $\frac{\partial f(x)}{\partial h} = \nabla f(x)^T \cdot h > 0$,则$\nabla f(\boldsymbol{x})$ 与$\boldsymbol{h}$ 的夹角为锐角，此时沿$\boldsymbol{h}$的方向，$f(\boldsymbol{x})$ 是上升
- 若 $\frac{\partial f(x)}{\partial h} = \nabla f(x)^T \cdot h < 0$,则$\nabla f(\boldsymbol{x})$ 与$\boldsymbol{h}$ 的夹角为钝角，此时沿$\boldsymbol{h}$的方向，$f(\boldsymbol{x})$ 是下降
- 当 $h = \frac{\nabla f(x)}{\|\nabla f(x)\|}$ 时，$\frac{\partial f(x)}{\partial h}$取得最大值，即：正梯度方向$\nabla f(\boldsymbol{x})$ 是$f(\boldsymbol{x})$ 在 x 处的最速上升方向，而负梯度方向$-\nabla f(\boldsymbol{x})$ 是$f(\boldsymbol{x})$ 在 x 处的最速下降方向。可见，梯度在最优化理论与算法中占有特殊重要的地位。

常见函数梯度

$$\nabla(b^T x) = b$$

$$\nabla(x^T x) = 2x$$

$$A^T = A \Rightarrow \nabla(x^T A x) = 2Ax$$

$$A^T = A, f(x) = \frac{1}{2} x^T A x + b^T x + c \Rightarrow \nabla f(x) = Ax + b$$

（4）Hesse矩阵

$$\nabla^2 f(X)=\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ & \cdots & \cdots & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

例1 $f(X)=\frac{1}{3}x_1^3+\frac{1}{3}x_2^3-2x_2^2-4x_1$

求 $\nabla f,\nabla^2 f$

极值问题

定义1: 若 $\min\limits_{x\in D} f(x) = f(x^*)$, $x^* \in D$ ，也就是

$\exists x^* \in D, \forall x \in D,$ 有 $f(x) \geq f(x^*)$

称 x^* 是全局极小值点。

如果有 $f(x)\ \mathrm{f}\ f(x^*)$, $\forall x \in D, x \neq x^*$

称 x^* 是严格全局极小点，