优化模型与软件工具

主讲教师：董庆兴

华中师范大学 信息管理学院
qxdong@mail.ccnu.edu.cn
All rights reserved

2017 年 9 月 5 日

大纲

教材

1. Boyd, Stephen, and Lieven Vandenberghe. Convex optimization. Cambridge University Press. 2004.
2. Stephen Boyd, Lieven Vandenberghe (王书宁，许鋆，黄晓霖译). 凸优化. 清华大学出版社.2013

参考资料

1. 袁亚湘，孙文瑜. 最优化理论与方法. 科学出版社. 1997
2. 李董辉, 童小娇，万中. 数值最优化算法与理论 (第二版). 科学出版社. 2005
3. Bertsekas.Covex Optimization Theory（影印版）. 清华大学出版社.2011
4. Masao Fukushima，非线性最优化基础. 科学出版社，2009

基础要求

高等数学	微积分的基本知识
概率论	基本的概率知识
线性代数	矩阵基本知识、线性空间基本知识

课程情况

教学安排 讲课 + 实验 (coding)

考核标准 考勤 + 课堂表现 + 平时作业 + 课程设计

课程介绍 介绍基本概念 + 算法原理 + 简单实现

What doesn't kill you, will kill you next time

will make you stronger

Office Hour

答疑时间 九号楼 817。提前预约

QQ 群 2017FallConvexOptimization: 620656200

联系方式

最优化问题一

某公交线路各时间段内所需司机人数如下表所示。设司机分别在各时间段开始时上班，并连续工作 8 小时，问该公交线路应怎样安排人员，既能满足需要，又使配备司机的人数最少?

班次	时间	所需人员
1	6:00-10:00	60
2	10:00-14:00	70
3	14:00-18:00	60
4	18:00-22:00	50
5	22:00-2:00	20
6	2:00-6:00	20

设第 i 班次开时开始上班的司机人数为 $x_i, i=1,\cdots,6$ ，从而有
第 i 班的在岗人数 $= x_{i-1} + x_i$

$$\min \quad z = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

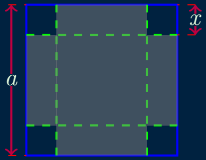
$$\text{s.t.} \begin{cases} x_1 + x_6 \geq 60 \\ x_1 + x_2 \geq 70 \\ x_2 + x_3 \geq 60 \\ x_3 + x_4 \geq 50 \\ x_4 + x_5 \geq 20 \\ x_5 + x_6 \geq 30 \\ x_1, x_2, x_3, x_4, x_5, x_6 \geq 0 \end{cases}$$

最优化问题二

某公司经营两种产品，第一种产品每件售价 30 元，第二种产品每件售价 450 元。根据统计，售出一件第一种产品所需要的服务时间平均为 0.5 小时，第二种产品是（$2+0.5x_2$）小时，其中 x_2 是第二种产品的售出数量。已知该公司在这段时间内的总服务时间为 800 小时，求使其营业额最大的营业计划。
设该公司计划经营第一种产品 x_1 件，第二种产品 x_2 件。则有如下优化模型

$$
\begin{aligned}
\max \quad & f(\mathbf{x}) = 30x_1 + 450x_2 \\
\text{s.t.} \quad & \begin{cases} 0.5x_1 + 2x_2 + 0.25x_2^2 \le 800 \\ x_1, x_2 \ge 0 \end{cases}
\end{aligned}
$$

优化问题三



给定 a，如何截取 x 使得铁皮所围成的容器容积最大?

- 首先列出容积公式
$$v = x(a-2x)^2$$
- 求微分：
$$\frac{\mathrm{d}v}{\mathrm{d}x} = (a-2x)^2 + 2(a-2x)\cdot x\cdot(-2) = 0$$
- 从而求解上式可得 $x = \frac{a}{6}$

最优化问题四

设有 n 个市场，第 j 个市场的位置为 (p_j, q_j)，它对某种货物的需要量为 b_j，现计划建立 m 个仓库，第 i 个仓库的存储量为 $a_i, i = 1, \cdots, m, j = 1, \cdots, n$。试着确定仓库的位置，使得总运输费用最小（运输量与路程乘积之和）。

设第 i 个仓库的位置为 $(x_i, y_i), i = 1, \cdots, m$，第 i 个仓库到第 j 个市场的货物供应量为 $z_{ij}, i = 1, \cdots, m, j = 1, \cdots, n$, 则第 i 个仓库到第 j 个市场的距离为

$$d_{ij} = \sqrt{(x_i - p_j)^2 + (y_i - q_j)^2}$$

$$\min \quad \sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij} d_{ij}$$

$$\text{s.t.} \quad \begin{cases} \sum_{j=1}^{n} z_{ij} \le a_i, & i = 1, \cdots, m \\ \sum_{i=1}^{m} z_{ij} = b_j, & j = 1, \cdots, n \\ z_{ij} \ge 0, & i = 1, \cdots, m, j = 1, \cdots, n \end{cases}$$

问题总结

- 第一个问题为线性问题，后面的问题是非线性问题
- 上述问题均为约束最优化问题，如果第三个问题中不考虑实际情况，有可能是无约束最优化问题 ($a \in (-\infty, +\infty)$)
- 所有上述问题的形式，可以总结为：

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, i = 1 \cdots, m. \end{array}$$

优化问题基本概念

这种形式的问题就被称作最优化问题:

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \le b_i, \ \ i = 1, \cdots, m. \end{aligned} \tag{1}$$

- 向量 $\mathbf{x} = (x_1, \cdots, x_n)$ 被称为问题的优化变量(决策变量)
- 函数 $f_0 : \mathbb{R}^n \to \mathbb{R}$ 为目标函数
- 函数 $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \cdots, m$ 为(不等式)约束函数, 常数 $b_1, \cdots, b_m$ 称为约束边界
- 如果在所有满足约束的向量中向量 $\mathbf{x}^*$ 对应的目标函数值最小: 即对于任意满足约束 $f_1(\mathbf{z}) \le b_1, \cdots, f_m(\mathbf{z}) \le b_m$ 的向量 $\mathbf{z}$ 有 $f_0(\mathbf{z}) \ge f_0(\mathbf{x}^*)$, 那么称 $\mathbf{x}^*$ 为问题 (1) 的最优解

优化问题分类：线性规划 VS 非线性规划

若优化问题（1）的目标函数和约束函数 $f_0, \cdots, f_m$ 都是线性函数，也就是对于任意的 $x, y \in \mathbb{R}^n$ 和 $\alpha, \beta \in \mathbb{R}$ 有

$$f_i(\alpha x + \beta y) = \alpha f_i(x) + \beta f_i(y) \tag{2}$$

则此问题被称作线性规划。若优化问题不是线性的，则称之为非线性规划

优化问题分类：凸优化 VS 非凸优化

若优化问题（1）的目标函数和约束函数 $f_0, \cdots, f_m$ 都是凸函数，也就是对于任意的 $x, y \in \mathbb{R}^n$ 和任意 $\alpha, \beta \in \mathbb{R}$，且满足 $\alpha + \beta = 1, \alpha, \beta \geq 0$

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y) \tag{3}$$

则此问题被称作凸优化问题。若优化问题不是凸的，则称之为非凸优化问题。

比较式（2）和式（3），可以看出凸性是比线性更为一般的性质：线性条件需要严格满足等式，而凸性条件仅需在 α, β 取特定值的情形下满足不等式。因此线性规划问题也是凸优化问题，凸优化是线性规划的拓展。

凸集

凸集

设 K 为 $\mathbb{R}^n$ 中的一个点集，若对任何 $x, y \in K$ 与任何 $\lambda \in [0,1]$ ，都有 $\lambda x + (1-\lambda)y \in K$，则称 K 为凸集

一些保凸运算:

- **交集** 如果 $i \in I$， $C_i \subseteq \mathbb{R}^n$ 均为凸集，则 $\cap_{i\in I} C_i$ 也是凸集
- **向量和** 如果 $C_1, C_2 \subseteq \mathbb{R}^n$ 是凸集，则其向量和 $C = C_1 + C_2 = \{x + y | x \in C_1, y \in C_2\}$ 也是凸集
- **伸缩** 如果 $C \subseteq \mathbb{R}^n$ 为凸集，$\alpha \in \mathbb{R}$，则 αC 为凸集
- **平移** 如果 $C \subseteq \mathbb{R}^n$ 为凸集，$a \in \mathbb{R}$，则 $C + a$ 为凸集

一些特殊的凸集

超平面 $\{\mathbf{x}|\mathbf{a}^{\mathrm{T}}\mathbf{x}=b\}$ 是 Hyperplane，其中 $\mathbf{a}\neq\mathbf{0},\mathbf{a}\in\mathbb{R}^n,b\in\mathbb{R}$

半空间 $\{\mathbf{x}|\mathbf{a}^{\mathrm{T}}\mathbf{x}\leq b\}$ 是 Halfspace, 其中 $\mathbf{a}\neq\mathbf{0},\mathbf{a}\in\mathbb{R}^n,b\in\mathbb{R}$

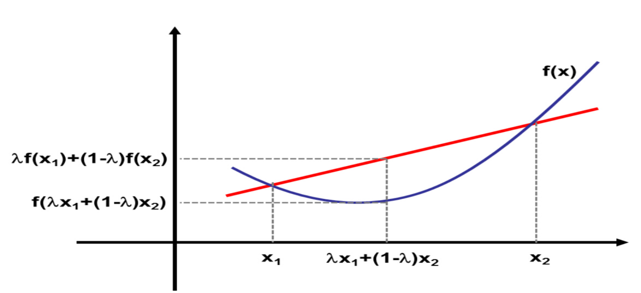
多面体 $\mathcal{P}=\{\mathbf{x}|\mathbf{a}_j^{\mathrm{T}}\mathbf{x}\leq b_j,j=1,\cdots,m\}$ 为 Polyhedral，其中 $\mathbf{a}\neq\mathbf{0},\mathbf{a}_j\in\mathbb{R}^n,b_j\in\mathbb{R},j=1,\cdots,m$

凸函数

凸函数

函数 $f:\mathbb{R}^n \to \mathbb{R}$ 是凸函数，如果 $\mathbf{dom}\ f$ 是凸集，且对任何 $x, y \in \mathbf{dom}\ f$ 与任何 $\theta \in [0,1]$ ，都有

$$f\left(\theta x+(1-\theta) y\right) \leq \theta f(x)+(1-\theta) f(y) \tag{4}$$



凸函数的图像和上境图

图像

函数 $f:\mathbb{R}^n \to \mathbb{R}$，称 $\mathbf{graph}\ \ f=\{(\mathbf{x},f(\mathbf{x}))|\mathbf{x}\in \mathbf{dom}\ \ f\}$ 为 f 的图像

上境图

函数 $f:\mathbb{R}^n \to \mathbb{R}$，称 $\mathbf{epi}\ \ f=\{(\mathbf{x},\mathbf{t}|\mathbf{x}\in \mathbf{dom}\ \ f\ f(\mathbf{x}))\le \mathbf{t}\}$ 为 f 的上境图

- 从几何意义上看，函数 f 为凸函数，等价于 f 的上境图为凸集
- 函数 f 为凹函数，等价于 f 的亚图（上境图的对称定义）为凸集

梯度

梯度与可微

对函数 $f:\mathbb{R}^n \to \mathbb{R}$ 在 $\mathbf{x} \in \mathbf{dom}\ f$ 的某邻域内取有限值，若 f 在 $\mathbf{x}$ 处存在偏导数

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = \lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{e}_i) - f(\mathbf{x})}{t}, i = 1, \cdots, n$$

其中 $\mathbf{e}_i$ 是第 i 个分量为 1，其他分量均为 0 的 n 维单位向量，并且对由 $\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_n}\right)^{\mathrm{T}} \in \mathbb{R}^n$ 定义的向量总有

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{h} + o(\|\mathbf{h}\|) \tag{5}$$

则称 f 在 $\mathbf{x}$ 处可微，其中 o 为高阶无穷小符号，向量 $\nabla f(\mathbf{x})$ 为 f 在 $\mathbf{x}$ 处的梯度

梯度常用公式

- $f(\mathbf{x}) = \mathbf{C}$，则 $\nabla f(\mathbf{x}) = \mathbf{0}$
- $f(\mathbf{x}) = \mathbf{b}^{\mathrm{T}}\mathbf{x}$，则 $\nabla f(\mathbf{x}) = \mathbf{b}$
- $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{x}$，则 $\nabla f(\mathbf{x}) = 2\mathbf{x}$
- $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$，且 $\mathbf{A}$ 为对称矩阵，则 $\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x}$

Hessian 矩阵

Hessian 矩阵

对函数 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $\mathbf{x}\in\mathbf{dom}\ f$ 的某邻域内取有限值，若 f 在 $\mathbf{x}$ 处存在二阶偏导数，并且对由

$$\mathbf{H}(f)=\nabla^2 f(\mathbf{x})=\begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1\partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n\partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n\partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}\in\mathbb{R}^{n\times n}$$

定义的矩阵总有

$$f(\mathbf{x}+\mathbf{h})=f(\mathbf{x})+\nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{h}+\frac{1}{2}\mathbf{h}^{\mathrm{T}}\nabla^2 f(\mathbf{x})\mathbf{h}+o(\|\mathbf{h}\|^2) \tag{6}$$

则称 f 在 $\mathbf{x}$ 处二阶可微，其中 o 为高阶无穷小符号，方阵 $\nabla^2 f(\mathbf{x})$ 为 f 在 $\mathbf{x}$ 处的Hessian 矩阵