同济数学系列丛书

TONGJISHUXUE XILIECONGSHU

运筹数学方法基础

朱经浩 殷俊锋 编著

同济大学出版社
TONGJI UNIVERSITY PRESS

同济数学系列丛书

TONGJISHUXUEXILIECONGSHU

运筹数学方法基础

朱经浩　殷俊锋　编著

同济大學 出版社

TONGJI UNIVERSITY PRESS

内 容 提 要

本书是作者根据多年讲授应用数学专业的“运筹学”课程讲义编写而成。全书的主要内容有线性规划方法基础、非线性规划的 K-T 最优性条件、二次规划、无约束最优化及约束最优化问题的罚函数法等。

本书思路新颖，文字浅显易懂，适用面广，可作为综合大学、师范院校的应用数学专业以及管理学科相关专业的教材或参考书。

图书在版编目(CIP)数据

运筹数学方法基础 / 朱经浩，殷俊锋编著. -- 上海：同济大学出版社，2014.12

ISBN 978-7-5608-5687-2

Ⅰ. ①运…　Ⅱ. ①朱…②殷…　Ⅲ. ①运筹学—数学方法　Ⅳ. ①O22

中国版本图书馆 CIP 数据核字(2014)第 266873 号

同济数学系列丛书

运筹数学方法基础

朱经浩　殷俊锋　编著

责任编辑　李小敏　**责任校对**　徐春莲　**封面设计**　潘向蓁

出版发行　同济大学出版社　www.tongjipress.com.cn

(地址：上海市四平路 1239 号　邮编：200092　电话：021－65985622)

经　销　全国各地新华书店

印　刷　大丰科星印刷有限责任公司

开　本　787mm×960mm　1/16

印　张　11.5

印　数　1—2 100

字　数　230 000

版　次　2014 年 12 月第 1 版　2014 年 12 月第 1 次印刷

书　号　ISBN 978-7-5608-5687-2

定　价　29.00 元

前　言

本书是作者根据多年讲授应用数学专业“运筹学”课程讲义编写而成。主要包含线性规划方法基础、非线性规划的 K-T 最优性条件、二次规划、无约束最优化及约束最优化问题的罚函数法等内容。在选材上，本书既注重运筹数学理论和方法的传授，又强调计算技能的训练，并引导证明能力的培养，做到既方便学生使用，又方便教师的教学。在编写上，本书思路新颖，考虑到运筹学的教学恰逢学生学完数学分析和高等代数，在编写中紧密联系分析和代数的思想方法，例如，在线性规划写作上尝试以基矩阵为红线，把单纯形法和对偶方法以及线性规划的应用串联成一个有机整体，易于对纷扰繁杂的线性规划内容的思想有一个基本的理解和较长久的记忆。另外，在编写中也加入了部分作者的最新科研成果，例如，在二次规划部分给出了球约束下的二次规划的完整的求解理论和方法。

在编写过程中，承蒙同济大学濮定国教授和上海大学邬冬华教授对初稿提出许多宝贵意见和建议，在此表示由衷的感谢。本书得到同济大学出版社和同济大学数学系的大力支持，同济大学出版社责任编辑李小敏老师为本书的出版付出了辛勤的劳动，我的研究生朱礼冬也为本书的形成做了大量电脑应用方面的工作。他们的努力使本书得以顺利出版，谨向他们致以诚挚的谢意。

作者

2014 年 11 月于同济大学

目 录

第 1 章　引　　论

运筹学是应用数学和管理科学中一个十分活跃的分支，在运筹数学的发展中不断萌发卓越的新理论，在运筹管理科学的应用实践中不断涌现新方法和新成果. 所谓运筹，就是在实现一个行动前及其过程中的谋划，从众多方案中选出最好或最可行的方案. 众所周知，谋定而动，方可立于不败之地. 简而言之，即要求最优化，其思想和运作可追溯到古代的生产和科学实践. 在工业化的近代和科学大发展的现代，运筹学的应用更是随处可见. 但是运筹方法千变万化不离其宗，从各种优化思想的实现到运筹方法的更新，都离不开数学的理论框架和计算工具. 本课程正是涉及运筹数学方法的一些理论基础和算法实践.

本章讲最优化的方法结构和作为运筹学重要工具的数学凸分析的初步理论.

1.1　最优化问题的数学形式

运筹学所研究的最优化问题的一般数学形式是

$$P\text{—}\ \min f(x) \qquad (1.1)$$
$$\text{s. t. } x \in X \subset \mathbf{R}^n,$$

其中，$x\in\mathbf{R}^n$ 称为决策变量或可行点，$f(x)$称为目标函数，$X\subset\mathbf{R}^n$ 称为约束集合或可行集.

例 1.1.1　设 $\boldsymbol{A}\in\mathbf{R}^{n\times n}$，$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}$，$\boldsymbol{A}>\boldsymbol{O}$，$\boldsymbol{b}\in\mathbf{R}^n$. 求解

$$\min f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}, \qquad (1.2)$$
$$\text{s. t. } \boldsymbol{x}\in\mathbf{R}^n.$$

解　这个问题的目标函数是正定二次函数，这个问题称为一个无约束的二次规划，可行集为 $X=\mathbf{R}^n$. 由初等微积分，先考虑梯度为零的点，求解$\nabla f(\boldsymbol{x})=\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}=\boldsymbol{0}$，得到唯一解 $\boldsymbol{x}^*=\boldsymbol{A}^{-1}\boldsymbol{b}$，再由 Hessen 矩阵

$$\nabla^2 f(\boldsymbol{x}) = \boldsymbol{A} > \boldsymbol{O},$$

可判断 $\boldsymbol{x}^* = \boldsymbol{A}^{-1}\boldsymbol{b}$ 为无约束的二次规划式(1.2)的最优解. 而最优值为

$$f(\boldsymbol{x}^*) = \frac{1}{2}(\boldsymbol{A}^{-1}\boldsymbol{b})^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{-1}\boldsymbol{b}) - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b} = \frac{-1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}.$$

注: 这里,可以回顾微积分中关于 $\nabla f(\boldsymbol{x})$, $\nabla^2 f(\boldsymbol{x})$ 的求法. 由于 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}$,对每分量 x_i,有

$$\begin{aligned}\frac{\partial f}{\partial x_i} &= \frac{1}{2}\frac{\partial(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial x_i} - b_i = \frac{1}{2}\Big[\Big(\sum_{j=1}^{n} a_{ij}x_j\Big) + \Big(\sum_{j=1}^{n} a_{ji}x_j\Big)\Big] - b_i \\ &= \frac{1}{2}\Big[2\Big(\sum_{j=1}^{n} a_{ij}x_j\Big)\Big] - b_i = \Big(\sum_{j=1}^{n} a_{ij}x_j\Big) - b_i.\end{aligned} \tag{1.3}$$

所以 $\nabla f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$. 而由式(1.3)得到

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial\Big(\Big(\sum_{k=1}^{n} a_{ik}x_k\Big) - b_k\Big)}{\partial x_i} = a_{ij},$$

所以, $\nabla^2 f(\boldsymbol{x}) = \boldsymbol{A}$.

例 1.1.2 设 $\boldsymbol{G} \in \mathbf{R}^{n\times n}$, $\boldsymbol{G}=\boldsymbol{G}^{\mathrm{T}}$, $\boldsymbol{G}>\boldsymbol{O}$, $\boldsymbol{g} \in \mathbf{R}^n$, $\boldsymbol{A} \in \mathbf{R}^{m\times n}$, $\boldsymbol{H} \in \mathbf{R}^{l\times n}$, $\boldsymbol{b} \in \mathbf{R}^m$, $\boldsymbol{d} \in \mathbf{R}^l$. 考虑优化问题

$$\min f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{x} - \boldsymbol{g}^{\mathrm{T}}\boldsymbol{x},$$

$$\text{s.t. } \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b},\ \boldsymbol{H}\boldsymbol{x} = \boldsymbol{d}.$$

这个问题的目标函数也是正定二次函数,但是约束条件由线性函数构成,这个问题称为一个有约束的正定二次规划,可行集为

$$X = \{\boldsymbol{x} \in \mathbf{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b},\ \boldsymbol{H}\boldsymbol{x} = \boldsymbol{d}\}.$$

例 1.1.3 设 $\boldsymbol{A} \in \mathbf{R}^{m\times n}$, $\boldsymbol{b} \in \mathbf{R}^m$, $\boldsymbol{c} \in \mathbf{R}^n$. 考虑优化问题

$$\begin{aligned}&\min f(\boldsymbol{x}) = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \\ &\text{s.t. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geqslant \boldsymbol{0}.\end{aligned}$$

这个问题的目标函数和约束条件都由线性函数构成,这个问题称为一个线性规划.

1.2 运筹数学方法的基本框架

假设目标函数 $f(x)$ 是连续可微的,如果可行集 $X=\mathbf{R}^n$,最优化问题(1.1)称为

一个无约束优化问题，这类问题的理论和方法框架是由初等微积分奠定的. 1.1 节的例 1.1.1 便是一个典型的无约束优化问题. 这里需要引入以下极值点的基本定义和回顾初等微积分中的一些众所周知的结论.

定义 1.2.1 最优化问题(1.1)的局部极小点 $\bar{x}$ 是指满足如此分析表达式的点：$\bar{x}\in X$，存在 $\delta>0$，对于 $\forall x\in X\cap O_\delta(\bar{x})$，$f(x)\geqslant f(\bar{x})$.（这里 $O_\delta(\bar{x})$ 表示以 δ 为半径，$\bar{x}$ 为中心的开球）. 最优化问题(1.1)的全局极小点(或最优点)x^* 是指满足如次分析表达式的点：$x^*\in X$，对于 $\forall x\in X$，$f(x)\geqslant f(x^*)$.

运筹数学源自微积分和高等代数，可直接叙述一些初等结论如次. 一阶必要条件：若最优化问题(1.1)的局部极小点 $\bar{x}\in \mathrm{int}\ X$，则 $\nabla f(\bar{x})=0$. 二阶必要条件：若最优化问题(1.1)的局部极小点 $\bar{x}\in \mathrm{int}\ X$，且 $f(x)$ 是二阶连续可微的，则 $\nabla f(\bar{x})=0$，$\nabla^2 f(\bar{x})\geqslant 0$. 二阶充分条件：设 $f(x)$ 是二阶连续可微的，若 $\bar{x}\in \mathrm{int}\ X$，$\nabla f(\bar{x})=0$，$\nabla^2 f(\bar{x})>0$，则 $\bar{x}$ 是严格局部极小点. 1.1 节的例 1.1.1 便是一个典型的满足这些结论的无约束最优化问题.

而对于一般的最优化问题(1.1)，例如 1.1 节的例 1.1.2 和例 1.1.3，则别指望可以简单应用以上这些理想的最优性条件而得到全局极小点或局部极小点. 而且在更深入的理论研究和实际问题中，有时甚至没有可微性条件，此时寻求最优点一般就只有遵循最原始的搜索思想. 直观而言，即从一个不是最优的可行点出发去搜寻下一个更好的可行点. 这里需要做三件事：①判断出发点是否为最优点，若是则无需再搜索；②检验目标值是否严格下降；③建立由一个可行点出发得到下一个可行点的算法. 本节介绍沿着直线进行搜索的思想，从而引入运筹数学方法的基本框架. 这里需要给出以下线搜索(也称为一维搜索)的基本知识点.

1.2.1 可行方向和下降方向

设 $\hat{x}\in X$，$\hat{x}$ 处的可行方向 $p\neq 0$ 是指满足如下分析表达式的向量：

$$\exists\alpha>0,\ \text{s.t.}\ \{x\,|\,x=\hat{x}+tp,\ 0\leqslant t\leqslant\alpha\}\subset X.$$

设 $\hat{x}\in X$，$\hat{x}$ 处的下降方向 $p\neq 0$ 是指满足如下分析表达式的向量：

$$\exists\alpha>0,\ \forall t\in(0,\alpha],\ f(\hat{x}+tp)<f(\hat{x}).$$

为了后面叙述的方便，不妨称 α 为一个下降步长因子.

设 $\hat{x}\in X$，$\hat{x}$ 处的下降可行方向 $p\neq 0$ 是指满足如下分析表达式的向量：

$$\exists\alpha>0,\{x\,|\,x=\hat{x}+tp,\ 0\leqslant t\leqslant\alpha\}\subset X,\text{且}\ \forall t\in(0,\ \alpha],$$
$$f(\hat{x}+tp)<f(\hat{x}).$$

例 1.2.1 考虑优化问题

$$\begin{aligned} \min\ & f(x) = x_1^2 + x_2, \\ \text{s.t.}\ & -x_1 - x_2 + 1 \geqslant 0, \\ & x_1^2 + x_2^2 - 9 \leqslant 0. \end{aligned} \tag{1.4}$$

验证 $\hat{\boldsymbol{x}}=(0,\ -3)^{\mathrm{T}}$ 是一个可行点,试证:$\hat{\boldsymbol{x}}$ 处的任一可行方向 $\boldsymbol{p}\neq\boldsymbol{0}$ 都不是下降方向.

证明 显然 $\hat{\boldsymbol{x}}=(0,\ -3)^{\mathrm{T}}$ 同时满足两个约束不等式,所以 $\hat{\boldsymbol{x}}=(0,\ -3)^{\mathrm{T}}$ 是一个可行点.

目标函数 $f(\boldsymbol{x})=x_1^2+x_2$ 的等位线 $x_1^2+x_2=c$ 是直角坐标系中的一族开口向下的抛物线:$x_2=-x_1^2+c$,其顶点坐标是 $(0,\ c)^{\mathrm{T}}$. 目标函数在可行集上的最小化也就是这族抛物线的位于可行集上的顶点坐标的第二个分量的最小化. 易见,在可行集上,要使得参数 c 取最小值 $\hat{c}$,抛物线的顶点是 $\hat{\boldsymbol{x}}=(0,\ -3)^{\mathrm{T}}$, $\hat{c}=-3$,且 $\hat{\boldsymbol{x}}$ 位于可行集的边界上. 几何上易见抛物线 $x_2=-x_1^2-3$ 与可行集仅有唯一的交点 $\hat{\boldsymbol{x}}$. 这样就发现 $\hat{\boldsymbol{x}}=(0,\ -3)^{\mathrm{T}}$ 是目标函数 $f(x)=x_1^2+x_2$ 在可行集上的最小点. 由平面解析几何可知,可行集是一个有界闭区域,因而 $\hat{\boldsymbol{x}}$ 处的任一可行方向 $\boldsymbol{p}$ 指向可行区域内部,而通过可行区域内部的抛物线 $x_2=-x_1^2+c$ 必满足 $c>-3$,从而目标函数在可行方向 $\boldsymbol{p}$ 位于可行集内的一段上的取值都大于 $\hat{c}=-3$,则 $\boldsymbol{p}$ 就不是下降方向.

注:以上的观察方法称为求解最优化问题的等位线法.

1.2.2 下降迭代法思想

简而言之,下降迭代即由某点 $\boldsymbol{x}\in X$ 出发寻找下一点 $\boldsymbol{x}'\in X$ 使得 $f(\boldsymbol{x})>f(\boldsymbol{x}')$. 这个过程一般得到一列可行点:

$$\boldsymbol{x}_0,\ \boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_n,\ \cdots,$$

使得

$$f(\boldsymbol{x}_0)>f(\boldsymbol{x}_1)>f(\boldsymbol{x}_2)>\cdots>f(\boldsymbol{x}_n)>\cdots.$$

数学工作者的任务是试图证明,有一点 $\boldsymbol{x}^*\in X$,使得 $\lim\limits_{n\to\infty}\|\boldsymbol{x}_k-\boldsymbol{x}^*\|=0$. 这是建立某个下降迭代算法的基本要求(一般试图说明 $\boldsymbol{x}^*$ 是 $\{x_k\}$ 的一个聚点).

1.2.3 下降算法的基本结构

(i) 确定初始点:$\boldsymbol{x}_0$.

(ii) $k \geqslant 0$,确定下降可行方向:$\boldsymbol{d}_k$.

(iii) 确定步长因子:$\alpha_k > 0$.

(iv) 计算 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$,验证 $\boldsymbol{x}_{k+1}$ 是否满足某种终止条件. 决定 $\boldsymbol{x}_{k+1}$ 为近似最优点,或 $k = k+1$,返回(ii).

1.2.4 收敛速度和算法的终止准则

收敛速度是一个下降迭代序列趋近于极小点的快慢的指标,也是评判一个算法优劣的标准. 若有 $\alpha > 0$, $q \geqslant 0$,使得 $\lim\limits_{n \to \infty} \dfrac{\| \boldsymbol{x}_{k+1} - \boldsymbol{x}^* \|}{\| \boldsymbol{x}_k - \boldsymbol{x}^* \|^{\alpha}} = q$,有以下收敛速度标准的分类:

(i) $\alpha = 1$, $0 < q < 1 \Rightarrow \{\boldsymbol{x}_k\}$ 有线性收敛速度;

(ii) $1 < \alpha < 2$, $q > 0$ 或 $\alpha = 1$, $q = 0 \Rightarrow \{\boldsymbol{x}_k\}$ 有超线性收敛速度;

(iii) $\alpha = 2$, $q > 0$, $\{\boldsymbol{x}_k\}$ 有二阶收敛速度.

显然,在以上收敛速度标准的分类中,一般而言,满足二阶收敛速度的算法是最好的.

终止准则决定算法的完成自然也非常重要. 以下三项之一都分别可作为算法的终止准则:

(i) $0 \leqslant f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}) \leqslant \varepsilon$.

(ii) $\| \boldsymbol{x}_k - \boldsymbol{x}_{k+1} \| \leqslant \varepsilon$.

(iii) $\| \nabla f(\boldsymbol{x}_k) \| \leqslant \varepsilon$.

这里(i),(ii)两条准则在数学分析中很重要,但是在运筹数学中却不很实用,经常出现如次情形:$\| \boldsymbol{x}_k - \boldsymbol{x}_{k+1} \|$ 很小时,$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1})$ 却很大;而 $f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1})$ 很小时,$\| \boldsymbol{x}_k - \boldsymbol{x}_{k+1} \|$ 反而变大. 本书讲的运筹数学也可称为光滑运筹学,即所涉函数都是可微的,因而比较实用的是准则(iii). 但是大家都知道,梯度为零的点可能只是局部极小点,甚至是极大点,所以要判断是否为最优点还需要结合其他手段进行分析.

例 1.2.2 对于最优化问题 $\min f(x) = |x|$, s. t. $x \in \mathbf{R}^1$,显然全局最优点是 $x^* = 0$. 建立下列迭代序列,

$$x_{k+1} = \begin{cases} \dfrac{x_k}{2}, & x_k \leqslant 1, \\ \dfrac{(x_k - 1)}{2} + 1, & x_k > 1. \end{cases} \tag{1.5}$$

对以下三种初始可行点 x_0 的选取方式,观察 $\{x_k\}$ 是否为下降迭代序列:

(i) $0<x_0\leqslant 1$,则 $x_k>0$, $\forall k$,且有 $x_k>x_{k+1}\to 0=x^*$,从而有 $f(x_k)>f(x_{k+1})\to 0$.

(ii) $x_0>1$,则 $x_k>0$, $\forall k$,且有 $x_{k+1}=\dfrac{x_k+1}{2}=\dfrac{x_0+2^{k+1}-1}{2^{k+1}}\to 1\neq x^*$ $(k\to\infty)$,从而有 $f(x_k)\to 1$. 但是,由于 $f(x_k)=|x_k|=x_k>1\Rightarrow 2x_k>x_k+1$,有

$$f(x_k)=|x_k|=x_k>\frac{x_k+1}{2}=x_{k+1}=|x_{k+1}|=f(x_{k+1}).$$

(iii) $x_0\leqslant 0$,则 $x_k\leqslant 0$, $\forall k$,且有 $x_k\leqslant x_{k+1}\to 0=x^*$,从而有 $f(x_k)>f(x_{k+1})\to 0$.

注:可见对任何初始点,此迭代公式构建了下降序列. 但当且仅当初始点不大于1时,迭代序列趋于最优点,即构成一个下降迭代算法. 当初始点不大于1且不为零时,算法具有线性收敛速度,这是因为

$$\lim_{n\to\infty}\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}=\lim_{n\to\infty}\frac{\left|\dfrac{x_k}{2}\right|}{|x_k|}=\frac{1}{2}.$$

由于本题的目标函数不是处处可微的,所以就不能用第(iii)个终止准则. 这里可用终止准则(i)或(ii). 容易看到,当初始点不大于1时,$\|x_k-x_{k+1}\|$和$f(x_k)-f(x_{k+1})$同时都很小.

1.3 一维搜索及其两个常用算法

对于最优化问题(1.1),所谓线搜索(也称为一维搜索)的概念由三个要素和一个一维优化问题组成:一个出发点(可行点)x_0,一个下降可行方向 d 和一个正数(步长因子)α_0,满足 $f(x_0+\alpha_0 d)=\min\limits_{\alpha\geqslant 0}f(x_0+\alpha d)$. 这个一元函数 $\varphi(\alpha):=f(x_0+\alpha d)$在正实轴$(0,+\infty)$内的优化过程称为一维搜索. 一维搜索也就是求解最优化问题

$$\begin{aligned}&\min \varphi(\alpha),\\&\text{s.t. }\alpha\geqslant 0.\end{aligned}\tag{1.6}$$

由于 d 为下降可行方向,最优化问题(1.6)的最优点在$(0,+\infty)$内. 由于下降可行方向往往是一个局部概念,一般在$(0,+\infty)$内的一个合适的单峰区间$[a,b](\subset(0,\delta))$上进行优化工作,即求解最优化问题

$$\min \varphi(\alpha), \quad \text{s.t. } \alpha \in [a, b]. \tag{1.7}$$

这也称为进行近似的一维搜索(或不精确线搜索). 若得到 $\alpha_0 \in [a, b]$,使得 $\varphi(\alpha_0) = \min\limits_{\alpha \in [a, b]} \varphi(\alpha)$,令 $x_1 = x_0 + \alpha_0 d$,则因 d 为下降可行方向,有

$$f(x_0) > \varphi(\alpha) \geqslant \varphi(\alpha_0) = f(x_0 + \alpha_0 d) = f(x_1). \tag{1.8}$$

这样通过不精确一维搜索就从可行点 x_0 出发沿着下降可行方向 d 而得到下一个可行点 x_1,满足 $f(x_0) > f(x_1)$,完成了一次下降迭代.

函数 $\varphi(\alpha)$ 的单峰区间 $[a, b]$ 意指,存在 $\alpha^* \in (a, b)$,使得在 $[a, \alpha^*]$ 上 $\varphi(\alpha)$ 严格单调递减,而在 $[\alpha^*, b]$ 上严格单调递增. 由此可知,在下降可行方向 d 上选取一个步长因子 α_0,可首先在 $(0, +\infty)$ 内获取一个单峰区间 $[a, b]$,由于可微函数 $\varphi(\alpha)$ 在一个单峰区间内存在最小点,接下来的工作就是寻求 $\varphi(\alpha)$ 在 $[a, b]$ 上的最小点. 这个单峰区间 $[a, b]$ 也称为一个搜索区间.

下面介绍在 $(0, +\infty)$ 内获取搜索区间 $[a, b]$ 的一个方法,称为进退法. 其基本思路是设想自变元从某个选定点向左或向右按一定步长进行移动,希望出现使得函数值呈现"高—低—高"的三个位置点. 所谓进退就是,若朝一个方向移动不成功,就退回来,再朝相反的方向寻找理想的位置点. 具体而言,即对于给定初始点 α_0 和初始步长 $h_0 > 0$,若 $\varphi(\alpha_0) > \varphi(\alpha_0 + h_0)$,则下一步从 $\alpha_0 + h_0$ 出发,加大步长或仍记为 h_0,继续向前搜索,直到 $\varphi(\alpha)$ 的值上升为止. 而若 $\varphi(\alpha_0) < \varphi(\alpha_0 + h_0)$,则下一步从 $\alpha_0 + h_0$ 出发朝相反方向以步长 $2h_0$,行至位置点 $\alpha_0 - h_0$,若继续有 $\varphi(\alpha_0 - h_0) < \varphi(\alpha_0)$,则前往位置点 $\alpha_0 - 2h_0$,直到 $\varphi(\alpha)$ 的值上升为止.

1.3.1 进退法

步 1:取定出发点 α_0 和步长 $h_0 > 0$,计算 $\varphi(\alpha_0)$,转到步 2.

步 2:$\alpha_1 = \alpha_0 + h_0$ 计算 $\varphi(\alpha_1)$,若 $\varphi(\alpha_1) < \varphi(\alpha_0)$,转到步 3,否则,转到步 5.

步 3:$h_0 = 2h_0$, $\alpha_2 = \alpha_1 + h_0$,计算 $\varphi(\alpha_2)$. 若 $\varphi(\alpha_2) > \varphi(\alpha_1)$,则得到 $[\alpha_0, \alpha_2]$ 为初始区间,停;若 $\varphi(\alpha_2) < \varphi(\alpha_1)$,转到步 4.

步 4:$\alpha_0 = \alpha_1$, $\alpha_1 = \alpha_2$, $\varphi(\alpha_0) = \varphi(\alpha_1)$, $\varphi(\alpha_1) = \varphi(\alpha_2)$, 转到步 3.

步 5:$h_0 = 2h_0$, $\alpha_2 = \alpha_1 - h_0$,计算 $\varphi(\alpha_2)$. 若 $\varphi(\alpha_0) \leqslant \varphi(\alpha_2)$,则得到 $[\alpha_2, \alpha_1]$ 为初始区间,停;若 $\varphi(\alpha_0) > \varphi(\alpha_2)$,则转到步 6.

步 6:$\alpha_1 = \alpha_0$, $\alpha_0 = \alpha_2$, $\varphi(\alpha_1) = \varphi(\alpha_0)$, $\varphi(\alpha_0) = \varphi(\alpha_2)$, 转到步 5.

例 1.3.1 求一维搜索 $\min\limits_{\alpha > 0} \varphi(\alpha) = \alpha^4 - \alpha + 1$ 的搜索区间,取初始点 $\alpha_0 = 0$,初始步长 $h_0 = 0.5$.

解 由于当 $\alpha\in(0,4^{\frac{-1}{3}})$，$\varphi(\alpha)=\alpha^4-\alpha+1$ 单调下降，而当 $\alpha\geqslant 4^{\frac{-1}{3}}$，$\varphi(\alpha)=\alpha^4-\alpha+1$ 单调上升，所以，由初等微分学可知，单峰区间的存在性是明显的. 若直接应用上述进退法算，借助 MATLAB，对于初始点 $\alpha_0=0$，初始步长 $h_0=0.5$，可得到搜索区间$[0,1]$.

应用进退法可得到函数值呈现"高—低—高"的三个位置点，而以下的三点二次插值法（抛物线法）正是借助这三个位置点在搜索区间上寻求 $\varphi(\alpha)$ 的最小点的数值方法. 为清楚起见，对于 $\varphi(\alpha)$ 的自变量，以下用字母 x 代替 α.

1.3.2 三点二次插值法（抛物线法）

(i) 给出 $x_1<x_0<x_2\in\mathbf{R}^1$，计算 $\varphi_1=\varphi(x_1)$，$\varphi_0=\varphi(x_0)$，$\varphi_2=\varphi(x_2)$.

(ii) 若 $\varphi_1>\varphi_0$，$\varphi_0<\varphi_2$，由三点 (x_1,φ_1)，(x_0,φ_0)，(x_2,φ_2) 作二次插值函数：

$$
\begin{aligned}
q(x)&=\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}\varphi_1+\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}\varphi_0+\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}\varphi_2\\
&=ax^2+bx+c,
\end{aligned}
$$

其中，$a=\dfrac{-[(x_0-x_2)\varphi_1+(x_2-x_1)\varphi_0+(x_1-x_0)\varphi_2]}{(x_1-x_0)(x_0-x_2)(x_2-x_1)}$. 可见若 $|a|$ 很小，则 $q(x)$ 近似一直线段. 并得到 $q'(x)=2ax+b$.

(iii) 解 $q'(x)=0$，得到唯一极小点：

$$
\bar{x}=\frac{1}{2}\,\frac{(x_2^2-x_0^2)\varphi_1+(x_1^2-x_2^2)\varphi_0+(x_0^2-x_1^2)\varphi_2}{(x_2-x_0)\varphi_1+(x_1-x_2)\varphi_0+(x_0-x_1)\varphi_2},
$$

则 $x_1<x_0\leqslant\bar{x}<x_2\in\mathbf{R}^1$ 或 $x_1<\bar{x}\leqslant x_0<x_2\in\mathbf{R}^1$. 若

$$
|(x_0-x_2)\varphi_1+(x_2-x_1)\varphi_0+(x_2-x_1)\varphi_2|<\varepsilon,
$$

则取 $\alpha^*=x_0$（因 $\varphi(x)$ 近似一直线）.

(iv) 若 $|\bar{x}-x_0|<\varepsilon_2$ 取 $\varphi(x)$ 在 $[x_1,x_2]$ 上的最小点 $x^*=x_0$，否则，比较 $\varphi(\bar{x})$ 与 φ_0 的大小，区分以下三种情形：

(1) $\varphi_0<\varphi(\bar{x})$，(i) $x_0<\bar{x}$，取 $x_1=x_1$，$x_0=x_0$，$x_2=\bar{x}$；(ii) $x_0>\bar{x}$，取 $x_1=\bar{x}$，$x_0=x_0$，$x_2=x_2$；返回(i).

(2) $\varphi_0>\varphi(\bar{x})$，(i) $x_0>\bar{x}$，取 $x_1=x_1$，$x_0=\bar{x}$，$x_2=x_0$；(ii) $x_0<\bar{x}$，取 $x_1=x_0$，$x_0=\bar{x}$，$x_2=x_2$；返回(i).

(3) $\varphi_0=\varphi(\bar{x})$，(i) $x_0<\bar{x}$，取 $x_1=x_0$，$x_2=\bar{x}$，$x_0=\dfrac{1}{2}(x_1+x_2)$；(ii) $x_0>\bar{x}$，取 $x_1=\bar{x}$，$x_2=x_0$，$x_0=\dfrac{1}{2}(x_1+x_2)$；返回(i).

由三点二次插值法产生的迭代序列具有很好的收敛性，这里仅介绍以下定理：

定理 1.3.1 设 $\varphi(\alpha)$ 存在四阶连续导数，α^* 满足 $\varphi'(\alpha^*)=0$，$\varphi''(\alpha^*)\neq 0$，则抛物线法产生的序列 $\{\alpha_k\}$ 收敛到 α^*，且有收敛速度为 1.32(超线性收敛).

1.4 数学凸分析的初步理论

欧氏空间 $\mathbf{R}^n$ 上的凸分析是运筹数学的基础，本节介绍一些数学凸集和凸函数的基本知识.

1.4.1 凸集

欧氏空间 $\mathbf{R}^n$ 中的凸集是这样的集合，其中任意两个点所连成的线段含于这个集合中.

定义 1.4.1 如果 $\forall\, \boldsymbol{x},\ \boldsymbol{y}\in D\subset\mathbf{R}^n$，$\forall\,\alpha\in[0,\ 1]$，有 $(1-\alpha)\boldsymbol{x}+\alpha\boldsymbol{y}\in D$，则称 D 是一个凸集.

数学上经常使用函数 $l(\alpha)=\alpha\boldsymbol{x}+(1-\alpha)\boldsymbol{y}$，$\alpha\in[0,\ 1]$表示连接 $\boldsymbol{x}$，$\boldsymbol{y}$ 的线段.

例 1.4.1 设 $\boldsymbol{a}\in\mathbf{R}^n$，$b\in\mathbf{R}$，集合 $H=\{\boldsymbol{x}\in\mathbf{R}^n\,|\,\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}=b\}$，$H^+=\{\boldsymbol{x}\in\mathbf{R}^n\,|\,\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}\geqslant b\}$，$B=\{\boldsymbol{x}\,|\,\|\boldsymbol{x}\|\leqslant r\}$ 依次表示欧氏空间 $\mathbf{R}^n$ 中的超平面，半空间和闭球，它们都是凸集.

与其他许多数学分支不同，运筹数学认定以下记号：

$$\boldsymbol{x}=(x_1,\ x_2,\ \cdots,\ x_n)^{\mathrm{T}}\in\mathbf{R}^n,\ \boldsymbol{x}\geqslant\boldsymbol{0}\Leftrightarrow x_i\geqslant 0,\ i=1,\ 2,\ \cdots,\ n.$$

例 1.4.2 $\{\boldsymbol{x}\in\mathbf{R}^n\,|\,\boldsymbol{x}\geqslant\boldsymbol{0}\}$是凸集集合.

证明 任意给定 $\boldsymbol{x}$，$\boldsymbol{y}\geqslant\boldsymbol{0}$，显然有 $l(\alpha)=\alpha\boldsymbol{x}+(1-\alpha)\boldsymbol{y}\geqslant\boldsymbol{0}$，$\alpha\in[0,\ 1]$，由此得证.

以下罗列凸集的若干性质：

(1) 若干凸集的交集是凸集.

(2) 凸集的数乘是凸集.

(3) 若干凸集的和集是凸集. 若干凸集的并集未必是凸集.

这些性质都可直接由凸集的定义结合集合论知识导出，除了关于第(3)条需指出，欧氏空间 $\mathbf{R}^n$ 中两个集合的和集的元素由任意选自各集合的元素的相加而形成.

以下是关于凸集定义在数学上的一些拓展.

凸组合的定义：给定 $\boldsymbol{x}_i\in\mathbf{R}^n$，$\lambda_i\geqslant 0$，$i=1,\ 2,\ \cdots,\ m$，$\sum\limits_{i=1}^{m}\lambda_i=1$，称 $\boldsymbol{y}=$

$\sum_{i=1}^{m}\lambda_i \boldsymbol{x}_i$ 为一个凸组合. 对给定 $\boldsymbol{x}_i \in \mathbf{R}^n$, $i=1, 2, \cdots, m$, 定义以下凸组合形成的集合:

$$E=\{\boldsymbol{y}=\sum_{i=1}^{m}\lambda_i \boldsymbol{x}_i \mid \boldsymbol{\lambda}=(\lambda_1, \cdots, \lambda_m)^{\mathrm{T}}\in \mathbf{R}^m, \lambda \geqslant 0, \sum_{i=1}^{m}\lambda_i=1\}.$$

直接由凸集的定义可证明:凸组合形成的集合 E 是凸集.

例 1.4.3 设 D 为凸集,则

(1) 对于任给的 $x_i \in D$, $i=1, 2, \cdots, m$ 和任给的 $\boldsymbol{\lambda}=(\lambda_1, \cdots, \lambda_m)^{\mathrm{T}}\in \mathbf{R}^m$, $\boldsymbol{\lambda}\geqslant \mathbf{0}$, $\sum_{i=1}^{m}\lambda_i=1$,有 $\boldsymbol{y}=\sum_{i=1}^{m}\lambda_i \boldsymbol{x}_i \in D$;

(2) 给定 $\boldsymbol{x}_i\in \mathbf{R}^n$, $i=1, 2, \cdots, m$, E 是包含$\{x_i, i=1, 2, \cdots, m\}$的最小凸闭集(凸包).

证明 (1) 当 $m=2$,结论由凸集定义得到. 进行数学归纳,设结论对于 $m=k$ 是正确的.

要证明:当 $m=k+1$ 时, 对于任给的 $\boldsymbol{x}_i \in D$, $i=1, 2, \cdots, m$, $\boldsymbol{\lambda}=(\lambda_1, \cdots, \lambda_m)^{\mathrm{T}}\in \mathbf{R}^m$, $\boldsymbol{\lambda}\geqslant \mathbf{0}$, $\sum_{i=1}^{m}\lambda_i=1$,有 $\boldsymbol{y}=\sum_{i=1}^{m}\lambda_i \boldsymbol{x}_i \in D$.

事实上,因 $\boldsymbol{\lambda}\geqslant \mathbf{0}$, $\sum_{i=1}^{k+1}\lambda_i=1$,若 $\lambda_{k+1}=0$,则回到 $m=k$ 情形,而若 $\lambda_{k+1}=1$,则 $\boldsymbol{y}=\boldsymbol{x}_{k+1}\in D$. 所以考虑 $0<\lambda_{k+1}<1$ 的情形,于是 $1-\lambda_{k+1}>0$,令 $\mu_i=\dfrac{\lambda_i}{1-\lambda_{k+1}}$, $i=1, 2, \cdots, k$,因 $\sum_{i=1}^{k}\mu_i=\sum_{i=1}^{k}\dfrac{\lambda_i}{1-\lambda_{k+1}}=\dfrac{\sum_{i=1}^{k}\lambda_i}{1-\lambda_{k+1}}=1$,由数学归纳法假设可知, $y=\sum_{i=1}^{k}\mu_i x_i \in D$. 注意到 $0<\lambda_{k+1}<1$, 由凸集定义得到,

$$\sum_{i=1}^{k+1}\lambda_i \boldsymbol{x}_i=(1-\lambda_{k+1})\sum_{i=1}^{k}\mu_i \boldsymbol{x}_i+\lambda_{k+1}\boldsymbol{x}_{k+1}=(1-\lambda_{k+1})\boldsymbol{y}+\lambda_{k+1}\boldsymbol{x}_{k+1}\in D.$$

(2) 设 D 为包含 $\boldsymbol{x}_i\in \mathbf{R}^n$, $i=1, 2, \cdots, m$ 的凸集,由已证得的结论(1)可知, $E\subset D$. 又由数学分析可知,E 为闭集. 而由凸集定义又易知 E 是凸集. 所以 E 是包含$\{\boldsymbol{x}_i, i=1, 2, \cdots, m\}$的最小凸闭集,也称为包含$\{\boldsymbol{x}_i, i=1, 2, \cdots, m\}$的凸包.

例 1.4.4 设 $\boldsymbol{A}\in \mathbf{R}^{m\times n}$, $\boldsymbol{b}\in \mathbf{R}^m$, $\boldsymbol{c}\in \mathbf{R}^n$. 考虑以下线性规划

$$\begin{aligned} &\min f(\boldsymbol{x})=\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \\ &\text{s. t. } \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}, \boldsymbol{x}\geqslant \mathbf{0}. \end{aligned} \tag{1.9}$$

设 $x^{(1)}$，$x^{(2)}$ 是上述线性规划的两个相异的可行点，满足 $c^T x^{(1)} > c^T x^{(2)}$，试证：$p = x^{(2)} - x^{(1)}$ 是 $x^{(1)}$ 处的下降可行方向.

证明 因 $x^{(1)} \neq x^{(2)}$，可见 $p = x^{(2)} - x^{(1)} \neq \mathbf{0}$. 容易证明可行集 $D = \{Ax = b, x \geqslant \mathbf{0}\}$ 是凸集. 因 $x^{(1)}, x^{(2)} \in D$，所以对于 $0 \leqslant \alpha \leqslant 1$，有 $x^{(1)} + \alpha p = (1-\alpha)x^{(1)} + \alpha x^{(2)} \in D$，可见 $p = x^{(2)} - x^{(1)}$ 是 $x^{(1)}$ 处的可行方向. 又因 $c^T x^{(1)} > c^T x^{(2)}$，对于任意的 $0 < \alpha \leqslant 1$，有

$$
\begin{aligned}
c^T x^{(1)} &= (1-\alpha)c^T x^{(1)} + \alpha c^T x^{(1)} > (1-\alpha)c^T x^{(1)} + \alpha c^T x^{(2)} \\
&= c^T((1-\alpha)x^{(1)} + \alpha x^2) = c^T(x^{(1)} + \alpha p).
\end{aligned}
$$

所以，$p = x^{(2)} - x^{(1)}$ 是 $x^{(1)}$ 处的下降方向. 总之，$p = x^{(2)} - x^{(1)}$ 是 $x^{(1)}$ 处的下降可行方向.

1.4.2 凸函数

所谓凸函数(严格凸函数)可定义如下：设 $D(\subset \mathbf{R}^n)$ 为凸集，函数 $f(x)$ 定义在 D 上，$\forall x, y \in D$，满足

$$
f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y), \quad \forall \alpha \in [0, 1], \tag{1.10}
$$

则称 $f(x)$ 是定义在 D 上的凸函数，若 $x \neq y$ 时式(1.10)中的不等号是严格的，则称 $f(x)$ 是定义在 D 上的严格凸函数.

容易明白：设 D 为凸集，则 $f(x)$ 为定义在 D 上的凸函数的充要条件是，对于 $\forall x, y \in D$，$\Phi(t) = f(tx + (1-t)y)$ 为定义在 $[0, 1]$ 上的凸函数. 又有以下须知或结论：

(1) 若 $f(x)$ 为凸函数，则 $-f(x)$ 称为凹函数.

(2) 凸函数的线性组合为凸函数.

(3) 线性函数为凸函数，也是凹函数.

(4) 凸函数 $f(x)$ 的水平集 $S(f, \beta) = \{x \mid x \in D, f(x) \leqslant \beta\}$ 为凸集.

证明 对于 $\forall x, y \in S(f, \beta) \subset D$，$\forall \alpha \in [0, 1]$，有 $f(x) \leqslant \beta$，$f(y) \leqslant \beta$，且 $\alpha x + (1-\alpha)y \in D$，因 $f(x)$ 为凸函数，所以，

$$
f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y) \leqslant \alpha\beta + (1-\alpha)\beta = \beta,
$$

由此得到 $\alpha x + (1-\alpha)y \in S(f, \beta)$，从而可知凸函数 $f(x)$ 的水平集为凸集.

(5) 设 $f(x)$ 在凸区域 D 内一阶连续可微，则 $f(x)$ 为凸函数的充分必要条件为 $f(y) \geqslant f(x) + \nabla f(x)^T(y - x)$，$\forall x, y \in D$.

证明 必要性：对任意 $x, y \in D$，$0 < \alpha < 1$，有