·信息与计算科学专业系列教材·

万仲平 费浦生 编

优化理论与方法

全国优秀出版社
武汉大学出版社

信息与计算科学专业系列教材

万仲平 费浦生 编

优化理论与方法

全国优秀出版社
武汉大学出版社

图书在版编目(CIP)数据

优化理论与方法/万仲平，费浦生编. —武汉：武汉大学出版社，2004.6

ISBN 7-307-04102-2

Ⅰ.优…　Ⅱ.①万…　②费…　Ⅲ.最优化算法　Ⅳ.O242.23

中国版本图书馆 CIP 数据核字(2003)第 118028 号

责任编辑：顾素萍　　责任校对：鄢春梅　　版式设计：支　笛

出版发行：**武汉大学出版社**　(430072　武昌　珞珈山)

(电子邮件：wdp4@whu.edu.cn　网址：www.wdp.whu.edu.cn)

印刷：武汉市新华印刷有限责任公司

开本：850×1168　1/32　印张：6.625　字数：163 千字

版次：2004 年 6 月第 1 版　　2004 年 6 月第 1 次印刷

ISBN 7-307-04102-2/O · 285　　定价：10.00 元

出版说明

1998年，教育部颁布了经调整后的高等学校新的专业目录，从1999年秋季开始，各院校开始按新的专业设置进行招生。信息与计算科学专业是在这次调整中设置的，是以信息处理和科学与工程计算为背景的，由信息科学、计算科学、运筹与控制科学等交叉渗透而形成的一个新的理科专业。目前，社会对这方面的人才需求越来越多，开办这个专业的院校也越来越多。因此，系统地出版一套高质量的相关教材是一项当务之急。

由于信息与计算科学专业是一个新设的专业，有关该专业的人才培养模式、培养目标、教学计划、课程体系、教材建设等一系列专业建设问题，各院校目前正在积极的研究和探索之中。为了配合全国各类高校信息与计算科学专业的教学改革和课程建设，推进高校信息与计算科学专业教材的出版工作，在有关专家的倡议和有关部门的大力支持下，我们组织成立了信息与计算科学专业系列教材编委会，并制定了教材出版规划。

编委会一致认为，规划教材应该能够反映当前教学改革的需要，要有特色和一定的前瞻性。规划的教材由个人申报或由有关专家推荐，经编委会认真评审，最后由出版社审定出版。对这批教材的编写力求做到教学改革力度大、有创新精神、有特色风格、深入浅出、可读性好、实用性强，以满足全国各类高校21世纪信息与计算科学专业及相关专业的教学需要。

限于我们的水平和经验，这批教材在编审、出版工作中还可

能存在不少的缺点和不足，希望使用本系列教材的教师、同学和其他广大读者提出批评和建议。

信息与计算科学专业系列教材编委会

前　言

最优化方法是计算数学与运筹学的一门重要交叉学科，也是一门古老而年轻的学科。最优化方法最早可以追溯到 Lagrange 关于一个函数在一组等式约束条件下的极值问题。近半个多世纪以来，最优化理论与算法已得到了充分的研究，并取得了非常重要的研究成果。随着高速度大容量电子计算机的普及，最优化方法已广泛应用于工程、国防、管理与经济等许多重要领域。

非线性规划（约束与无约束）问题内容十分丰富，求解非线性规划问题的优化算法更加繁多。本书对约束和无约束非线性最优化的一些常用而有效算法及其相关理论作了系统的介绍，特别是对这些算法所产生的背景及其主要思想作了简洁的描述。内容包括最优化基本理论；无约束优化中的常用线性搜索方法、共轭梯度法以及拟 Newton 算法；约束优化问题中的可行方向法、有效集方法、二次规划问题算法、罚函数方法、乘子方法和约束变尺度方法等。此外，对内点算法和信赖域方法也作了简单介绍。

本书可作为信息与计算科学、应用数学、工程领域各专业、管理与经济等专业的高年级本科生和研究生的教材或教学参考书，也可作为从事优化技术应用的工程技术人员的参考用书。

本书在编写的过程中得到了许多同事的关心和支持，参阅了很多专著和文稿。对此，作者表示衷心的感谢。

限于作者水平，缺点和错误在所难免，敬请广大读者批评、指正，不吝赐教。

编　者

2004 年 3 月

目　录

第 1 章　预备知识

本章首先介绍非线性规划问题的数学模型分类及其有关概念，然后简单介绍多元函数的有关分析基础和凸规划问题的基本知识以及非线性规划最优解所满足的最优性条件，最后简要介绍求解方法产生的迭代序列及其收敛速度的概念. 这一章的内容是以后各章的基础.

1.1　最优化问题的模型与分类

最优化问题(P)的数学模型的一般形式为

$$\left.\begin{aligned}\min\quad & f(\boldsymbol{x}),\\ \text{s.t.}\quad & g_i(\boldsymbol{x})\leqslant 0,\quad i\in I\triangleq\{1,2,\cdots,m\},\\ & h_j(\boldsymbol{x})=0,\quad j\in E\triangleq\{1,2,\cdots,l\},\end{aligned}\right\}\tag{1.1}$$

其中，$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbf{R}^n$ 称为**优化向量**或**决策变量**；f: $\mathbf{R}^n\to\mathbf{R}$ 称为**目标函数**；g_i: $\mathbf{R}^n\to\mathbf{R}$ $(i\in I)$，h_j: $\mathbf{R}^n\to\mathbf{R}$ $(j\in E)$ 称为**约束函数**，相应地，$g_i(\boldsymbol{x})\leqslant 0$ $(i\in I)$与 $h_j(\boldsymbol{x})=0$ $(j\in E)$ 分别称为**不等式约束**与**等式约束**，等式约束与不等式约束统称为**约束条件**. 令

$$D=\{\boldsymbol{x}\in\mathbf{R}^n\mid g_i(\boldsymbol{x})\leqslant 0,\ i\in I,\ h_j(\boldsymbol{x})=0,\ j\in E\}.$$

称 D 为问题(P)的**约束集合**(**约束区域**)或**可行域**. 若 $\boldsymbol{x}\in D$，则称 $\boldsymbol{x}$ 为**可行点**或**可行解**. 设 $\boldsymbol{x}^*\in D$.

1) 若对任意的 $\boldsymbol{x}\in D$，都有 $f(\boldsymbol{x})\geqslant f(\boldsymbol{x}^*)$，则称 $\boldsymbol{x}^*$ 为优

化问题(P)的**全局最优解(极小点)**.

2) 若存在 x^* 的一个 $\delta\ (>0)$ 邻域

$$N(x^*,\delta)=\{x\mid \|x-x^*\|\leqslant\delta\},$$

当 $x\in D\cap N(x^*,\delta)$ 时, 有 $f(x)\geqslant f(x^*)$, 则称 x^* 为问题(P)的**局部最优解(极小点)**. 若对任意的 $x\in D\cap N(x^*,\delta)$, 有 $f(x)>f(x^*)$, 则称 x^* 是**严格局部最优解**.

1.1.1 问题的分类

1. 无约束优化问题

当 $E=I=\varnothing$ (空集)时, 称问题(P)为**无约束优化问题**:

$$\min f(x),\quad x\in\mathbf{R}^n. \tag{1.2}$$

无约束优化问题是在空间 $\mathbf{R}^n$ 上寻求使目标函数 $f(x)$ 达到极小或最小的点(解) x^*.

2. 约束优化问题

若指标集 I 与 E 中至少有一个非空, 则称问题(P)为**约束优化问题**. 约束优化问题是在约束集 D 上寻求使目标函数 $f(x)$ 达到极小或最小的点(解) x^*.

若 $I=\varnothing$, $E\neq\varnothing$, 即问题(P)只含等式约束, 此时称问题(P)为**等式约束优化问题**; 若 $I\neq\varnothing$, $E=\varnothing$, 即问题(P)只含不等式约束, 此时称问题(P)为**不等式约束优化问题**. 否则称问题(P)为**混合约束优化问题**.

如果目标函数和约束函数都是线性函数, 则称问题(P)为**线性规划**, 线性规划常记为 **LP**. 如果目标函数和约束函数中至少有一个是非线性函数, 则称问题(P)为**非线性规划**, 非线性规划常记为(**NP**)或简记为 **NP**. 在非线性规划问题中, 如果约束函数均是线性函数, 则称问题(P)是**线性约束非线性规划问题**, 这类规划问题常记为 **NLP**. 特别地, 目标函数 $f(x)$ 为二次函数的线性约束问题称为**二次规划问题**, 常记为 **QP**.

由于 $\max f(x)$ 等价于 $-\min\{-f(x)\}$, 因此, 我们只讨论

在 D 上的极小化问题. 此外，本书不讨论线性规划问题，只讨论非线性规划问题的求解方法及其有关理论.

1.1.2　解法的分类

1. 解析方法

解析方法主要是利用函数的分析性质去构造迭代公式，使得到的函数值序列和解序列分别收敛到问题的极小值和极小解(极小值点). 本书介绍的大部分方法属于解析方法.

2. 直接方法

直接方法对函数的分析性质(如可微性等)没有要求，而只是根据一定的数学原理，通过直接比较函数值的大小来确定问题的极小值(点).

3. 智能方法

智能方法是通过智能技术来达到问题的求解目的. 其方法主要包括：模拟退火算法、神经网络算法、遗传算法以及禁忌搜索算法等，本书不予介绍，有兴趣的读者可参阅相关文献资料.

1.2　多元函数基础知识

1.2.1　多元函数的梯度、Hesse 矩阵及向量函数的 Jaccobi 矩阵

1. 梯度

定义 1.1　若 n 元函数 $f(\boldsymbol{x})$ 对自变量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 的各分量 x_i 的偏导数 $\dfrac{\partial f(\boldsymbol{x})}{\partial x_i}$ $(i=1,2,\cdots,n)$ 都存在，则称函数 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处一阶可导，并称向量

$$\nabla f(\boldsymbol{x})=\left(\frac{\partial f(\boldsymbol{x})}{\partial x_1},\frac{\partial f(\boldsymbol{x})}{\partial x_2},\cdots,\frac{\partial f(\boldsymbol{x})}{\partial x_n}\right)^{\mathrm{T}} \tag{1.3}$$

为函数 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处的**梯度**或**一阶导数**. 为书写简便，有时记

$$g(\boldsymbol{x}) \triangleq \nabla f(\boldsymbol{x}), \quad \boldsymbol{g}_k \triangleq \nabla f(\boldsymbol{x}^{(k)}),$$

其中，$\nabla f(\boldsymbol{x}^{(k)})$为$\nabla f(\boldsymbol{x})$在点 $\boldsymbol{x}^{(k)}$处的值.

梯度的几何意义：如果函数 $f(\boldsymbol{x})$在 $\bar{\boldsymbol{x}}$ 处的梯度$\nabla f(\bar{\boldsymbol{x}}) \neq \boldsymbol{0}$，则$\nabla f(\bar{\boldsymbol{x}})$为 $f(\boldsymbol{x})$的等值线(面)在 $\bar{\boldsymbol{x}}$ 处的法向量，它垂直于等值线(面)在点 $\bar{\boldsymbol{x}}$ 处的切平面，且沿此方向 $f(\boldsymbol{x})$的函数值增大.

2. Hesse 矩阵

定义 1.2 若 n 元函数$f(\boldsymbol{x})$具有二阶偏导数，即

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} \quad (i, j = 1, 2, \cdots, n) \tag{1.4}$$

都存在，则称矩阵

$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_n} \end{pmatrix} \tag{1.5}$$

为 $f(\boldsymbol{x})$在 $\boldsymbol{x}$ 处的 **Hesse 矩阵(海色矩阵)**. 有时将$\nabla^2 f(\boldsymbol{x})$记为 $G(\boldsymbol{x})$或 $H(\boldsymbol{x})$，将$\nabla^2 f(\boldsymbol{x}^{(k)})$记为 $\boldsymbol{G}_k$. 显然，当 $f(\boldsymbol{x})$的所有二阶偏导数都连续时，其 Hesse 矩阵$\nabla^2 f(\boldsymbol{x})$是一个对称矩阵(为简便起见，今后所言正定矩阵时，它们均为对称正定矩阵).

例 设 $f(\boldsymbol{x}) = \dfrac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + c$，其中 $\boldsymbol{Q} \in \mathbf{R}^{n \times n}$是 n 阶对称矩阵(以后简称为对称矩阵)，$\boldsymbol{b} \in \mathbf{R}^n$，$c \in \mathbf{R}$ (一维实数空间). 易验证：$\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}$，$\nabla^2 f(\boldsymbol{x}) = \boldsymbol{Q}$.

3. 向量函数的 Jaccobi 矩阵

定义 1.3 设向量函数 $F(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_m(\boldsymbol{x}))^{\mathrm{T}}$，其中 f_i：$\mathbf{R}^n \to \mathbf{R}$. 若$\dfrac{\partial f_i(\boldsymbol{x})}{\partial x_j}$ $(i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n)$都存

在，则称 $F(\boldsymbol{x})$在点 $\boldsymbol{x}$ 处一阶可导，$F(\boldsymbol{x})$在点 $\boldsymbol{x}$ 处的导数 $F'(\boldsymbol{x}) \in \mathbf{R}^{m\times n}$为

$$F'(\boldsymbol{x})=\begin{pmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_1(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_n} \\ \frac{\partial f_2(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_2(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\boldsymbol{x})}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_m(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n} \end{pmatrix}, \tag{1.6}$$

它是向量函数 $F(\boldsymbol{x})$在 $\boldsymbol{x}$ 处的 **Jaccobi 矩阵**，其转置叫做 $F(\boldsymbol{x})$在 $\boldsymbol{x}$ 处的**梯度**，即 $F'(\boldsymbol{x})=\nabla F(\boldsymbol{x})^{\mathrm{T}}$.

1.2.2 多元函数的 Taylor 展开式

n 元函数的 Taylor 展开式在非线性规划问题的研究中具有非常重要的作用. 在最优化理论中，常用带有 Peano 型或 Lagrange 型余项公式的一阶和二阶 Taylor 展(开)式.

1. 一阶 Taylor 展式

设 $f(\boldsymbol{x})$在点 $\bar{\boldsymbol{x}}$ 处具有一阶连续偏导，则 $f(\boldsymbol{x})$在点 $\bar{\boldsymbol{x}}$ 处的一阶 Taylor 展开式为

$$f(\boldsymbol{x})=f(\bar{\boldsymbol{x}})+\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})+o(\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|), \tag{1.7}$$

其中 $o(\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|)$为变量 $\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|$ 的高阶无穷小量($\boldsymbol{x}\to\bar{\boldsymbol{x}}$)，或者

$$f(\boldsymbol{x})=f(\bar{\boldsymbol{x}})+\nabla f(\boldsymbol{\xi})^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}}), \tag{1.8}$$

其中 $\boldsymbol{\xi}=\bar{\boldsymbol{x}}+\theta(\boldsymbol{x}-\bar{\boldsymbol{x}})$ $(0<\theta<1)$.

2. 二阶 Taylor 展式

设 $f(\boldsymbol{x})$在点 $\bar{\boldsymbol{x}}$ 处二阶连续可微(或具有二阶连续偏导数)，则 $f(\boldsymbol{x})$在点 $\bar{\boldsymbol{x}}$ 处的二阶 Taylor 展开式为

$$f(\boldsymbol{x})=f(\bar{\boldsymbol{x}})+\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})+\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}})^{\mathrm{T}}\nabla^2 f(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}}) + o(\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2), \tag{1.9}$$

或者

$$f(\boldsymbol{x}) = f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x} - \bar{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathrm{T}}\nabla^2 f(\boldsymbol{\xi})(\boldsymbol{x} - \bar{\boldsymbol{x}}), \tag{1.10}$$

其中 $\boldsymbol{\xi} = \bar{\boldsymbol{x}} + \theta(\boldsymbol{x} - \bar{\boldsymbol{x}})$ $(0<\theta<1)$.

1.3 凸规划问题

1.3.1 凸集与凸函数

1. 凸集及其基本性质

定义 1.4 设集合 $D \subset \mathbf{R}^n$. 若 $\forall\, \boldsymbol{x}, \boldsymbol{y} \in D$, $\forall\, \lambda \in [0,1]$ 有 $\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in D$, 则称集合 D 为**凸集**.

由定义可直接验证：设 $D_1, D_2 \subset \mathbf{R}^n$ 为凸集，$\alpha \in \mathbf{R}$, 则

1) $D_1 \cap D_2 = \{\boldsymbol{x} \mid \boldsymbol{x} \in D_1,\ \boldsymbol{x} \in D_2\}$ 是凸集；

2) $D_1 + D_2 = \{\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in D_1,\ \boldsymbol{y} \in D_2\}$ 是凸集；

3) $D_1 - D_2 = \{\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{x} \in D_1,\ \boldsymbol{y} \in D_2\}$ 是凸集；

4) $\alpha D_1 = \{\alpha \boldsymbol{x} \mid \boldsymbol{x} \in D_1\}$ 是凸集.

利用数学归纳法可证：设 D 是凸集，则 $\forall\, \boldsymbol{x}^{(i)} \in D$ 和数 α_i, 都有 $\sum\limits_{i=1}^{m} \alpha_i \boldsymbol{x}^{(i)} \in D$, 其中 $\alpha_i \geqslant 0$, $\sum\limits_{i=1}^{m} \alpha_i = 1$.

2. 凸函数的定义及其基本性质

定义 1.5 设 f: $D \subset \mathbf{R}^n \to \mathbf{R}$, 且 D 为非空凸集. 若 $\forall\, \boldsymbol{x}, \boldsymbol{y} \in D$, $\forall\, \lambda \in [0,1]$ 都有

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leqslant \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}), \tag{1.11}$$

则称 $f(\boldsymbol{x})$ 是定义在凸集 D 上的**凸函数**.

在式(1.11)中，若 $\boldsymbol{x} \neq \boldsymbol{y}$ 时严格不等式成立，则称 $f(\boldsymbol{x})$ 是凸集 D 上的**严格凸函数**.

若 $D = \mathbf{R}^n$, 我们就不强调凸集 D, 而称 $f(\boldsymbol{x})$ 是**凸函数**或**严**

格凸函数.

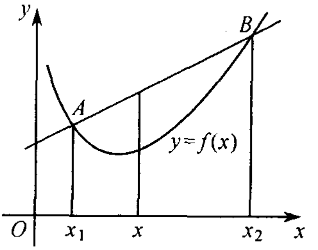

图 1.1 凸函数的几何解释

凸函数的几何意义：以 $n=1$ 来说明，当 $f(x)$是凸函数时，其图形 $y=f(x)$上任意两点 $A(x_1,f(x_1)),B(x_2,f(x_2))$之间的弧必在连接这两点的弦 $\overline{AB}$ 的下方(如图 1.1).

注 由定义 1.5，我们不难用归纳法证明：若 $f(\boldsymbol{x})$是凸集 D 上的凸函数，则 $\forall\, \boldsymbol{x}^{(i)}\in D$ 和数 $\alpha_i\geqslant 0$ $(i=1,2,\cdots,m)$，$\sum\limits_{i=1}^{m}\alpha_i=1$，有

$$f\Big(\sum_{i=1}^{m}\alpha_i\boldsymbol{x}^{(i)}\Big)\leqslant\sum_{i=1}^{m}\alpha_i f(\boldsymbol{x}^{(i)}).$$

定义 1.6 设 $D\subset\mathbf{R}^n$ 为非空凸集，f: $D\to\mathbf{R}$. 称集合

$$E_p(f)=\{(\boldsymbol{x},y)\,|\,f(\boldsymbol{x})\leqslant y,\ \boldsymbol{x}\in D\}\triangleq \mathrm{epi}\,f$$

为 $f(\boldsymbol{x})$的**上图(像)**. 称集合

$$S_\alpha=\{\boldsymbol{x}\,|\,\boldsymbol{x}\in D,\ f(\boldsymbol{x})\leqslant\alpha\}$$

为 $f(\boldsymbol{x})$的 **α-水平集**.

定理 1.1 设 D 为凸集. 则 $f(\boldsymbol{x})$为 D 上的凸函数当且仅当 $\mathrm{epi}\,f$ 为 $\mathbf{R}^{n+1}$中的凸子集.

证 设 $f(\boldsymbol{x})$为 D 上的凸函数，则 $\forall\,(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2)\in \mathrm{epi}\,f$，即有 $f(\boldsymbol{x}_1)\leqslant y_1$，$f(\boldsymbol{x}_2)\leqslant y_2$. 对 $0\leqslant\lambda\leqslant 1$ 有

$$\begin{aligned}f(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2)&\leqslant\lambda f(\boldsymbol{x}_1)+(1-\lambda)f(\boldsymbol{x}_2)\\&\leqslant\lambda y_1+(1-\lambda)y_2,\end{aligned}$$

即有

$$(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2,\lambda y_1+(1-\lambda)y_2)\in\mathrm{epi}\,f,$$

从而 $\mathrm{epi}\,f$ 为凸集.

反之，设 epi f 为凸集，则 $\forall\, \boldsymbol{x}_1, \boldsymbol{x}_2 \in D$，$0 \leqslant \lambda \leqslant 1$，易见

$$(\boldsymbol{x}_1, f(\boldsymbol{x}_1)) \in \operatorname{epi} f, \quad (\boldsymbol{x}_2, f(\boldsymbol{x}_2)) \in \operatorname{epi} f.$$

而

$$\begin{aligned}&(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2, \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2))\\ &\quad = \lambda(\boldsymbol{x}_1, f(\boldsymbol{x}_1)) + (1-\lambda)(\boldsymbol{x}_2, f(\boldsymbol{x}_2)) \in \operatorname{epi} f,\end{aligned}$$

所以

$$f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \leqslant \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2),$$

故 $f(\boldsymbol{x})$为 D 上的凸函数. □

定理 1.2 设 D 为凸集且$f(\boldsymbol{x})$是 D 上的凸函数，则 S_α 为凸集.

证 设 $f(\boldsymbol{x})$为 D 上的凸函数. $\forall\, \alpha \in \mathbf{R}$，$\forall\, \boldsymbol{x}_1, \boldsymbol{x}_2 \in S_\alpha$ 及 $\lambda \in (0,1)$，有

$$\begin{aligned}f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) &\leqslant \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2)\\ &\leqslant \lambda\alpha + (1-\lambda)\alpha = \alpha.\end{aligned}$$

因此 $\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in S_\alpha$，即 S_α 为凸集. □

注 若 $f(\boldsymbol{x})$为闭凸集 D 上的连续凸函数，则 S_α 为闭凸集. 事实上，设$\{\boldsymbol{x}^{(k)}\} \subset S_\alpha$，且 $\bar{\boldsymbol{x}}$ 为$\{\boldsymbol{x}^{(k)}\}$的一个极限点，显见 $\bar{\boldsymbol{x}} \in D$. 由 $f(\boldsymbol{x}^{(k)}) \leqslant \alpha$ 和$f(\boldsymbol{x})$的连续性，则有

$$f(\bar{\boldsymbol{x}}) = \lim_{k \to \infty} f(\boldsymbol{x}^{(k)}) \leqslant \alpha,$$

所以 $\bar{\boldsymbol{x}} \in S_\alpha$.

定理 1.3 设 D 为 $\mathbf{R}^n$ 中非空凸集，f: $D \to \mathbf{R}$ 是凸函数，则 f 在D 的内部(记为 int D)是连续函数.

证 设 $\bar{\boldsymbol{x}} \in \operatorname{int} D$，为证 f 在$\bar{\boldsymbol{x}}$ 处连续，需证：$\forall\, \varepsilon > 0$，$\exists\, \delta > 0$，使得当$\| \boldsymbol{x} - \bar{\boldsymbol{x}} \| \leqslant \delta$ 时，有$| f(\boldsymbol{x}) - f(\bar{\boldsymbol{x}}) | \leqslant \varepsilon$.

由于 $\bar{\boldsymbol{x}} \in \operatorname{int} D$，故存在 $\delta' > 0$，使得当$\| \boldsymbol{x} - \bar{\boldsymbol{x}} \| \leqslant \delta'$时，有 $\boldsymbol{x} \in D$. 令