运筹与管理科学丛书 5

非光滑优化

高岩 著

科学出版社
www.sciencep.com

运筹与管理科学丛书 5

非光滑优化

高 岩 著

科 学 出 版 社
北 京

内 容 简 介

本书旨在系统介绍非光滑优化理论与方法, 全书共分为九章. 第 1 章和第 2 章分别介绍凸集和凸函数的概念和有关性质; 第 3 章引入凸函数的次微分, 给出凸函数的极值条件与中值定理, 并介绍次微分的性质和特殊凸函数的次微分表达式; 第 4 章介绍局部 Lipschitz 函数的广义梯度, 给出极大值函数广义 Jacobi 的计算; 第 5 章阐述拟可微函数及拟微分的定义和性质; 第 6 章针对凸规划、Lipschitz 优化、拟可微优化给出最优性条件; 第 7 章提出非光滑优化算法, 包括下降方法、凸规划的次梯度法、凸规划的割平面法; 第 8 章研究非光滑方程组及非线性互补问题; 第 9 章介绍非光滑理论在控制论中的应用.

本书可作为应用数学、运筹学与控制论及经济管理有关专业的高年级本科生或研究生教材, 也可供相关专业的科研工作者参考.

图书在版编目(CIP)数据

非光滑优化/高岩著. —北京: 科学出版社, 2008

(运筹与管理科学丛书; 5)

ISBN 978-7-03-021502-4

I. 非… II. 高… III. 光滑化(数学) IV. O189

中国版本图书馆 CIP 数据核字 (2008) 第 041433 号

责任编辑: 赵彦超 / 责任校对: 陈玉凤

责任印制: 赵德静 / 封面设计: 陈 敬

科 学 出 版 社出版

北京东黄城根北街 16 号

邮政编码: 100717

http://www.sciencep.com

新 蕾 印 刷 厂 印刷

科学出版社发行 各地新华书店经销

*

2008 年 5 月第 一 版 开本: B5(720 × 1000)

2008 年 5 月第一次印刷 印张: 13 1/4

印数: 1—3 000 字数: 245 000

定价: 39.00 元

(如有印装质量问题, 我社负责调换〈环伟〉)

《运筹与管理科学丛书》序

运筹学是运用数学方法来刻画、分析以及求解决策问题的科学. 运筹学的例子在我国古已有之, 春秋战国时期著名军事家孙膑为田忌赛马所设计的排序就是一个很好的代表. 运筹学的重要性同样在很早就被人们所认识, 汉高祖刘邦在称赞张良时就说道: "运筹帷幄之中, 决胜千里之外."

运筹学作为一门学科兴起于第二次世界大战期间, 源于对军事行动的研究. 运筹学的英文名字 Operational Research 诞生于 1937 年. 运筹学发展迅速, 目前已有众多的分支, 如线性规划、非线性规划、整数规划、网络规划、图论、组合优化、非光滑优化、锥优化、多目标规划、动态规划、随机规划、决策分析、排队论、对策论、物流、风险管理等.

我国的运筹学研究始于 20 世纪 50 年代, 经过半个世纪的发展, 运筹学队伍已具相当大的规模. 运筹学的理论和方法在国防、经济、金融、工程、管理等许多重要领域有着广泛应用, 运筹学成果的应用也常常能带来巨大的经济和社会效益. 由于在我国经济快速增长的过程中涌现出了大量迫切需要解决的运筹学问题, 因而进一步提高我国运筹学的研究水平、促进运筹学成果的应用和转化、加快运筹学领域优秀青年人才的培养是我们当今面临的十分重要、光荣、同时也是十分艰巨的任务. 我相信,《运筹与管理科学丛书》能在这些方面有所作为.

《运筹与管理科学丛书》可作为运筹学、管理科学、应用数学、系统科学、计算机科学等有关专业的高校师生、科研人员、工程技术人员的参考书, 同时也可作为相关专业的高年级本科生和研究生的教材或教学参考书. 希望该丛书能越办越好, 为我国运筹学和管理科学的发展做出贡献.

袁亚湘

2007 年 9 月

前 言

非光滑优化又称不可微优化, 是最优化理论与方法中的一个重要分支. 所谓非光滑优化, 是指目标函数或约束函数中至少有一个不是连续可微 (光滑) 的非线性规划问题. 由于不具有连续可微性质, 传统的基于微分 (梯度) 概念的优化理论和方法已不再适用于非光滑优化问题. 对经典的微分概念进行推广, 建立各种广义微分概念, 基于广义的微分理论建立相应的最优性理论和算法, 正是非光滑优化研究之所在.

非光滑优化具有广泛的应用背景. 下面给出几个非光滑优化的例子.

例 1 设 $x_1, \cdots, x_m \in \mathbf{R}^n$ 为 m 个实验数据, 要建立一个线性模型, 即求一个超平面 $H = \{ x \in \mathbf{R}^n \mid a^{\mathrm{T}} x = b \}$, 其中 $a \in \mathbf{R}^n$, $b \in \mathbf{R}$, 使得 $x_1, \cdots, x_m$ 尽可能接近 H, 这就引出一个不可微优化问题

$$\min_{(a,b)\in\mathbf{R}^{n+1}} \sum_{k=1}^{m} \left| \sum_{i=1}^{n} a_i x_k^i - b \right|, \tag{1}$$

其中, $a \in \mathbf{R}^n$, $b \in \mathbf{R}$ 为变量, a_i, x_k^i 分别为 a 和 x_k 的第 i 个分量. 易见, 问题 (1) 的目标函数带有绝对值, 是非光滑函数. 但是, 人们以往为了回避非光滑问题的困难, 通常考虑下述问题

$$\min_{(a,b)\in\mathbf{R}^{n+1}} \sum_{k=1}^{m} \left(\sum_{i=1}^{n} a_i x_k^i - b \right)^2, \tag{2}$$

即人们通常所熟知的最小二乘问题. 需要指出的是, 问题 (1) 和问题 (2) 的解一般来讲是不一致的. 对大多数情形而言, 问题 (1) 的解较问题 (2) 的解更加符合实际需要.

例 2 考虑非线性互补问题

$$f(x) \geqslant 0, \quad h(x) \geqslant 0, \quad f(x)^{\mathrm{T}} h(x) = 0, \tag{3}$$

其中

$$f(x) = (f_1(x), \cdots, f_n(x))^{\mathrm{T}}, \quad h(x) = (h_1(x), \cdots, h_n(x))^{\mathrm{T}},$$

$f_i(x)$, $h_i(x)$, $i = 1, \cdots, n$ 均为 $\mathbf{R}^n$ 上的连续可微函数, 求解问题 (3) 可等价地转化为求解如下非光滑方程组

$$\min\{ f_i(x), h_i(x) \} = 0, \quad i = 1, \cdots, n. \tag{4}$$

求解方程组 (4) 也等价于求解如下的优化问题:

$$\min_{x\in\mathbf{R}^n}\sum_{i=1}^{n}(\min\{f_i(x),h_i(x)\})^2, \tag{5}$$

显然, (5) 是一个非光滑优化问题.

还有一种非光滑优化来自于优化问题本身, 即求解非线性规划的罚函数方法. 考虑约束优化问题

$$\begin{aligned}&\min\ f(x),\\&\text{s.t.}\quad g(x)\leqslant 0,\end{aligned} \tag{6}$$

其中, $f(x)$, $g(x)$ 为 $\mathbf{R}^n$ 上的连续可微函数. 利用罚函数法, 求解约束优化问题 (6) 转化为求解下述无约束优化问题

$$\min_{x\in\mathbf{R}^n} f(x)+M\max\{0,g(x)\}, \tag{7}$$

其中 M 为较大的正数. 由于最后一项 $\max\{0,g(x)\}$ 的原因, 问题 (7) 的目标函数是非光滑的.

尽管非光滑优化有广泛的应用, 然而到目前为止, 我们还没有有效的方法处理一般形式的非光滑优化问题, 只能针对一些特殊形式的非光滑优化问题分别进行研究. 在各种类型的非光滑优化中, 凸规划和 Lipschitz 规划是目前影响最大, 也是最被广泛接受的一类非光滑优化问题. 当然, 凸规划是 Lipschitz 规划的特殊形式.

非光滑优化包括许多内容, 目前国外出版的著作都有自身的一套体系. 本书从非光滑优化基本内容入手, 较详细地介绍了凸函数的次微分、局部 Lipschitz 函数的广义梯度、拟可微函数的拟微分及它们的最优性理论, 最后介绍了在控制理论中的应用.

本书的完成得到了国家自然科学基金 (10671126) 和上海市重点学科建设项目 (T0502) 的资助. 作者在写作过程中结合了多年关于非光滑优化的学习和研究工作, 并参阅了国内外相关文献. 由于作者水平有限, 加之时间仓促, 书中一定有许多不妥和错误之处, 敬请各位专家、同行批评指正.

作　者

2008 年 2 月于上海理工大学

目 录

第 1 章　凸　　集

凸集是非光滑分析与优化中最基本的概念之一, 本书以后各章都与本章的内容有关.

1.1　凸集的基本概念

本节引入凸集的概念, 并给出它的一些基本性质.

1.1.1　凸集与凸组合

定义 1.1.1　设 $S \subset \mathbf{R}^n$, 如果对任意 $x_1, x_2 \in S, 0 \leqslant \lambda \leqslant 1$, 有 $\lambda x_1 + (1-\lambda)x_2 \in S$, 则称 S 为凸集.

由定义可以看出, 所谓凸集就是这样的集合, 它的任意两点的连线都在集合中, 可以说凸集具有明显的几何意义.

例 1.1.1　超平面 $H = \{x \in \mathbf{R}^n | p^{\mathrm{T}} x = \alpha\}$ 是凸集, 其中 p 为 n 维向量, α 为实数. 对任意 $x_1, x_2 \in H, 0 \leqslant \lambda \leqslant 1$, 有

$$p^{\mathrm{T}}(\lambda x_1 + (1-\lambda)x_2) = \lambda p^{\mathrm{T}} x_1 + (1-\lambda)p^{\mathrm{T}} x_2 = \alpha,$$

因此 $\lambda x_1 + (1-\lambda)x_2 \in H$, 根据定义, H 是凸集.

例 1.1.2　设 $x_0 \in \mathbf{R}^n$, $\delta > 0$, 容易验证, 以 x_0 为圆心 δ 为半径的球体:

$$B(x, \delta) = \{x \in \mathbf{R}^n |\, ||x - x_0|| \leqslant \delta\}$$

为 $\mathbf{R}^n$ 中的凸集.

另外, 根据凸集的定义很容易验证, $\mathbf{R}^n$ 中空集 $\varnothing$、全空间、所有子空间都是凸集.

定理 1.1.1　设 I 是任意指标集, $S_i \subset \mathbf{R}^n$, $i \in I$ 是凸集, 则 S_i 的交 $S = \bigcap\limits_{i \in I} S_i$ 是 $\mathbf{R}^n$ 中凸集.

证明　若 S 为空集或单点集, 结论显然成立. 假设 $x_1, x_2 \in S$, 则 $x_1, x_2 \in S_i$, $i \in I$. 由于 S_i 是凸集, 则对于 $0 \leqslant \lambda \leqslant 1$, 有

$$\lambda x_1 + (1-\lambda)x_2 \in S_i, \quad i \in I,$$

故

$$\lambda x_1 + (1-\lambda)x_2 \in \bigcap_{i\in I} S_i,$$

所以 S 是凸集. 定理得证.

定义 1.1.2 设 $x_1,\cdots,x_m \in \mathbf{R}^n, \lambda_i \geqslant 0,\ i=1,\cdots,m,\ \sum\limits_{i=1}^m \lambda_i = 1$, 则点 $x = \sum\limits_{i=1}^m \lambda_i x_i$ 称为 $x_1,\cdots,x_m$ 的一个凸组合.

凸组合是凸分析中一个重要概念, 它与凸集有密切联系. 定义 1.1.1 意味着凸集就是 "其中任意两点的凸组合仍属于它本身的集合". 而实际上, 我们也可以通过任意有限点的凸组合来定义凸集, 下面的定理就刻画了这样一个事实.

定理 1.1.2 $S \subset \mathbf{R}^n$ 是凸集的充要条件是 S 中所有元素的凸组合还在 S 中.

证明 设 S 是凸集, $x_1,\cdots,x_m \in S$, 我们将证明 $x_1,\cdots,x_m$ 的凸组合属于 S. 对 m 用数学归纳法. 当 $m=1$ 时, 结论显然成立; 当 $m=2$ 时, 由凸集的定义, 结论也成立. 设结论在 $m \leqslant k$ 时成立, 要证明对于 $\lambda_i \geqslant 0, i=1,\cdots,k+1$ 满足

$$\sum_{i=1}^{k+1} \lambda_i = 1,$$

如果

$$x_i \in S, \quad i=1,\cdots,k+1,$$

则

$$x = \sum_{i=1}^{k+1} \lambda_i x_i \in S.$$

不失一般性, 假设 $\lambda_i > 0, i=1,\cdots,1+k$, 这时

$$1-\lambda_{k+1} = \sum_{i=1}^{k} \lambda_i > 0.$$

根据归纳法假设

$$y = \frac{1}{1-\lambda_{k+1}} x_1 + \cdots + \frac{\lambda_k}{1-\lambda_{k+1}} x_k \in S,$$

这是因为

$$\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} = 1,$$

上式为 $x_1,\cdots,x_k$ 的凸组合. 再由凸集的定义, 有

$$x = (1-\lambda_{k+1})y + \lambda_{k+1}x_{k+1} \in S.$$

另一方面, 设集合 S 中元素的所有凸组合都在 S 自身中, 则 S 的任意两个元素的凸组合也在 S 中, 于是 S 是凸集. 定理得证.

定义 1.1.3 $\mathbf{R}^n$ 中集合 S 的凸包是由 S 中的一切凸组合形成的集合, 记为 $\mathrm{co}S$, 换言之, $x\in \mathrm{co}S$ 当且仅当 x 可表示为 $x=\sum\limits_{i=1}^{k}\lambda_i x_i$, 其中 $x_i\in S,\lambda_i\geqslant 0, i=1,\cdots,k,\sum\limits_{i=1}^{k}\lambda_i=1$, k 为一正整数.

很容易验证, S 的凸包是包含 S 的最小凸集. 事实上, 不难验证它是包含 S 的所有凸集的交集. 凸包也是对一个给定非凸集合进行凸化的手段.

$\mathbf{R}^n$ 中有限点集 $\{\alpha_1,\cdots,\alpha_m\}$, 其中 $\alpha_i\in\mathbf{R}^n, i=1,\cdots,m$ 的凸包由形如 $\lambda_1\alpha_1+\cdots+\lambda_m\alpha_m$ 的向量构成, 其中 $\lambda_i\geqslant 0,\ i=1,\cdots,m$, 且有 $\sum\limits_{i=1}^{m}\lambda_i=1$, 亦可表示为

$$\left\{\sum_{j=1}^{m}\lambda_i a_i|\lambda_i\geqslant 0,\ i=1,\cdots,m\right\},$$

它是 $\mathbf{R}^n$ 空间中的一个凸多面体.

由定义 1.1.3 知, 凸包 $\mathrm{co}S$ 是由 S 的所有有限多个点的凸组合构成的集合, 但定义 1.1.3 没有对构成这个凸组合所需的点数给出任何限制, 实际上, 对于 n 维空间中的集合 S, 只需至多 $n+1$ 个点的凸组合就可以表示 $\mathrm{co}S$ 中的点, 下面的 Caraheodory 定理就揭示了这样的事实.

定理 1.1.3 (Caraheodory 定理) 设 $S\subset\mathbf{R}^n$, 则 S 的凸包 $\mathrm{co}S$ 中的任意一点可以表示成 S 中至多 $n+1$ 个点的凸组合, 即对任意 $x\in\mathrm{co}S$, 存在常数 $r\leqslant n+1$ 以及 $x_i\in S,\lambda_i\geqslant 0, i=1,\cdots,r$ 满足 $\sum\limits_{i=1}^{r}\lambda_i=1$, 使得

$$x=\sum_{i=1}^{r}\lambda_i x_i. \tag{1.1.1}$$

证明 根据定理 1.1.2, 只要证明 $r\leqslant n+1$ 即可. 以下证明, 如果 $r>n+1$, 则式 (1.1.1) 右边的非零项可以减少. 不妨假设

$$\lambda_i>0,\quad i=1,\cdots,r,\quad r>n+1.$$

取 $n+1$ 维向量 $(x_i,1)$, $i=1,\cdots,r$, 因为向量的个数 $r>n+1$, 因此线性相关, 故存在不全为零的常数 α_i, $i=1,\cdots,r$, 使得

$$\sum_{i=1}^{r}\alpha_i(x_i,1)=0,$$

于是有

$$\sum_{i=1}^{r}\alpha_i x_i = 0, \quad \sum_{i=1}^{r}\alpha_i = 0. \tag{1.1.2}$$

由

$$\sum_{i=1}^{r}\alpha_i = 0$$

可知, α_i, $i=1,\cdots,r$ 中一定存在正数, 记

$$\varepsilon_0 = \min\left\{\frac{\lambda_i}{\alpha_i}|\alpha_i > 0,\ i=1,\cdots,r\right\},$$

于是存在一个 i_0, 使得

$$\varepsilon_0 = \frac{\lambda_{i_0}}{\alpha_{i_0}},$$

进而有

$$\bar{\lambda}_i = \lambda_i - \varepsilon_0\alpha_i \geqslant 0, \quad i=1,\cdots,r, \tag{1.1.3}$$

特别地, $\bar{\lambda}_{i_0}=0$. 由式 (1.1.2) 得

$$\sum_{i=1}^{r}\bar{\lambda}_i x_i = \sum_{i=1}^{r}\lambda_i x_i - \varepsilon_0\sum_{i=1}^{r}\alpha_i x_i = x,$$

$$\sum_{i=1}^{r}\bar{\lambda}_i = \sum_{i=1}^{r}\lambda_i - \varepsilon_0\sum_{i=1}^{r}\alpha_i = 1, \tag{1.1.4}$$

这说明, x 还可以表示为式 (1.1.1) 的形式, 但却减少了一项 (在式 (1.1.4) 中 $\bar{\lambda}_{i_0}=0$). 定理得证.

1.1.2　代数运算

在非光滑分析中, 通常的集合加法和数乘运算按如下法则, 通常也称为 Minkowski 加法和数乘.

定义 1.1.4　设 $S_1, S_2 \subset \mathbf{R}^n$, $\lambda \in \mathbf{R}^n$, 则 $\lambda S_1 = \{\lambda x | x \in S_1\}$ 称为集合 S_1 和 λ 的数乘;

$$S_1 + S_2 = \{x_1 + x_2 | x_1 \in S_1,\ x_2 \in S_2\}$$

称为 S_1 和 S_2 的和.

下述结论显然成立.

命题 1.1.1　设 $S_1, S_2 \subset \mathbf{R}^n$ 为凸集, $\lambda \in \mathbf{R}$, 则 λS_1 和 $S_1 + S_2$ 也为凸集.

定理 1.1.4　设 S 为 $\mathbf{R}^n$ 中凸集, $\lambda_1 \geqslant 0, \lambda_2 \geqslant 0$, 则有

$$(\lambda_1+\lambda_2)S = \lambda_1 S + \lambda_2 S. \tag{1.1.5}$$

证明 当 $\lambda_1=\lambda_2=0$ 时, 式 (1.1.5) 显然成立. 设 $\lambda_1+\lambda_2>0$, 容易验证, 即使集合 S 不是凸, 下述性质也成立:

$$(\lambda_1+\lambda_2)S\subset\lambda_1 S+\lambda_2 S. \tag{1.1.6}$$

对于凸集 S, 容易验证

$$\lambda S+(1-\lambda)S\subset S,$$

其中 $0\leqslant\lambda\leqslant 1$, 于是有

$$\frac{\lambda_1}{\lambda_1+\lambda_2}S+\frac{\lambda_2}{\lambda_1+\lambda_2}S\subset S,$$

上式两边乘以 $\lambda_1+\lambda_2$ 得

$$\lambda_1 S+\lambda_2 S\subset(\lambda_1+\lambda_2)S, \tag{1.1.7}$$

联立式 (1.1.6) 和式 (1.1.7) 得式 (1.1.5). 定理得证.

1.1.3 凸锥

在凸集中, 一个比较重要的特殊情形是凸锥. 凸锥是非光滑分析和优化中研究的重要对象之一.

定义 1.1.5 设 S 为 $\mathbf{R}^n$ 中集合, $x_0\in S$, 且对任意 $\lambda>0$, 有 $x_0+\lambda(x-x_0)\in S$, 则称 S 是以 x_0 为顶点的锥. 特别当 S 为凸集时, S 称为凸锥.

在最优化研究中, 人们最关心的是以 0 为顶点的锥, 以后除特殊声明, 所提到的锥均指以 0 为顶点的锥. 易见, $S\subset\mathbf{R}^n$ 是锥的充要条件是对任意 $\lambda>0$, 有 $\lambda S=S$.

例 1.1.3 下述两个集合是 $\mathbf{R}^n$ 中的凸锥:

$$S_1=\{(x_1,\cdots,x_n)^{\mathrm{T}}|x_i\geqslant 0,\ i=1,\cdots,n\},$$

$$S_2=\{(x_1,\cdots,x_n)^{\mathrm{T}}|x_i>0,\ i=1,\cdots,n\}.$$

定理 1.1.5 $S\subset\mathbf{R}^n$ 是凸锥的充分必要条件是它对加法和正数乘法封闭.

证明 设 S 是凸锥, 由于是锥, 它对正数乘法是封闭的, 又 S 是凸集, 则对任意 $x_1,x_2\in S$, 有

$$z=\frac{1}{2}x_1+\frac{1}{2}x_1\in S,$$

所以 $x_1+x_2=2z\in S$, 即 S 关于加法封闭.

设 S 对加法和正数乘法封闭. 由于对正数乘法封闭, 所以 S 是锥, 故对任意

$$x_1,x_2\in S,\quad 0<\lambda<1,$$

有

$$\lambda x_1 \in S, \quad (1-\lambda)x_2 \in S.$$

而由 S 对加法封闭性, 得 $\lambda x_1 + (1-\lambda)x_2 \in S$, 所以 S 是凸集, 进而是凸锥. 定理得证.

由定理 1.1.5 易得下面的推论.

推论 1.1.1　$S \subset \mathbf{R}^n$ 是凸锥的充分必要条件是 S 包含它的元素的全部正线性组合, 即对任意 $x_i \in S$, $\lambda > 0$, $i = 1, \cdots, m$, 有

$$\lambda_1 x_1 + \cdots + \lambda_m x_m \in S.$$

定义 1.1.6　设 $S \subset \mathbf{R}^n$ 是凸锥, 它的极锥定义如下:

$$S^\circ = \{d \in \mathbf{R}^n | d^{\mathrm{T}} x \leqslant 0, \ \forall x \in S\}.$$

从极锥的定义不难看出, S° 是闭凸锥. 不难验证: 如果 S 是一个子空间, S° 是它的正交补; 如果 $S_1 \subset S_2$, 则 $S_1^\circ \supset S_2^\circ$; 锥与它的极锥交非空, 即 $S \bigcap S^\circ = \{0\}$.

例 1.1.4　设集合 S 为有限点形成的凸锥, 即

$$S = \left\{ \sum_{j=1}^{m} \lambda_i a_i | \lambda_i \geqslant 0, \ i = 1, \cdots, m \right\},$$

其中 $a_i \in \mathbf{R}^n, i = 1, \cdots, m$, 则 S 的极锥为

$$S^\circ = \{y \in \mathbf{R}^n | y^{\mathrm{T}} x_i \leqslant 0, i = 1, \cdots, m\}.$$

1.2　凸集上的投影

点到子空间的投影算子有很多好的性质, 例如线性、对称性、半正定性、幂等性、非膨胀性等. 本节讨论点到凸集的投影, 它在平衡问题以及变分不等式问题中有许多重要应用.

给定非空闭集 $S \subset \mathbf{R}^n$ 和固定的 $x \in \mathbf{R}^n$, 考虑下述问题:

$$\inf_{y \in S} \frac{1}{2} ||y - x||^2, \tag{1.2.1}$$

它是寻找 x 到集合 S 中的最近点, 即投影. 固定 $x \in \mathbf{R}^n$, 定义下述函数:

$$f_x(y) = \frac{1}{2} ||y - x||^2, \quad y \in \mathbf{R}^n, \tag{1.2.2}$$

给定 $s \in S$, 考虑水平集

$$L_s = \{y \in \mathbf{R} | f_x(y) \leqslant f_x(s)\}.$$

显然, 问题 (1.2.1) 与问题 $\inf\limits_{y \in S \cap L_s} f_x(y)$ 等价. 由于函数 $f_x(y)$ 连续且非负, 则 $S \bigcap L_s$ 为紧集, 于是点 x 到集合 S 的最近点, 即投影是存在的. 这样式 (1.1.2) 中的下确界 inf 可用极小值 min 来代替.

事实上, 投影的存在性并不需要集合的凸性来保证, 但是凸性却可以保证它的唯一性. 假设 y_1 和 y_2 都是问题 (1.2.1) 的解, 令 $x_i = y_i - x, i = 1, 2$, 记 $y_0 = \frac{1}{2}(y_1 + y_2)$, 利用关系式

$$\frac{1}{2}||y_1 + y_2||^2 = ||y_1||^2 + ||y_2||^2 - \frac{1}{2}||y_2 - y_1||^2,$$

得

$$f_x(y_0) = \frac{1}{2}(f_x(y_1) + f_x(y_2)) - \frac{1}{8}||y_2 - y_1||^2.$$

由 y_1 和 y_2 均为 (1.2.1) 的解和 $y_0 \in S$, 则有 $y_1 = y_2$, 这就说明了凸集投影的唯一性.

记 $p_S(x)$ 为点 x 到集合 S 上的投影, 我们有下述定理.

定理 1.2.1 设 $S \subset \mathbf{R}^n$ 为凸集, $x \in \mathbf{R}^n$, 则 $y_x \in S$ 为 x 到 S 的投影 $p_S(x)$ 的充要条件是

$$(x - y_x)^{\mathrm{T}}(y - y_x) \leqslant 0, \quad \forall y \in S. \tag{1.2.3}$$

证明 必要性. 设 S 为凸集. 注意到 y_x 是问题 (1.2.1) 的解, 取任意的 $y \in S$, 根据 S 的凸性有

$$y_x + \lambda(y - y_x) \in S, \quad 0 < \lambda < 1,$$

由式 (1.2.2) 可得

$$\begin{aligned}\frac{1}{2}||y_x - x||^2 &= f_x(y_x) \\ &\leqslant f_x(y_x + \alpha(y - y_x)) \\ &= \frac{1}{2}||y_x - x + \lambda(y - y_x)||^2.\end{aligned}$$

展开平方项并整理得

$$0 \leqslant \lambda(y_x - x)^{\mathrm{T}}(y - y_x) + \frac{1}{2}\lambda^2||y - y_x||^2,$$

两边除以 λ, 并令 $\lambda \to 0^+$, 即得式 (1.2.3).

充分性. 假设 $y_x \in S$ 满足式 (1.2.3), 如果 $y_x = x$, 则 y_x 显然是 (1.2.1) 的解. 考虑 $y_x \neq x$, 对任意的 $y \in S$, 利用 Cauchy-Schwarz 不等式, 直接推导得

$$\begin{aligned}
0 &\geqslant (x-y_x)^{\mathrm{T}}(y-y_x) \\
&= (x-y_x)^{\mathrm{T}}(y-x+x-y_x) \\
&= ||x-y_x||^2 + (x-y_x)^{\mathrm{T}}(y-x) \\
&\geqslant ||x-y_x||^2 - ||x-y||\,||x-y_x||.
\end{aligned}$$

注意到 $||x-y_x||>0$, 两边除以 $||x-y_x||$ 知 y_x 是式 (1.2.3) 的解. 定理得证.

进一步, 我们有下述定理.

定理 1.2.2　设 $S \subset \mathbf{R}^n$ 为凸集, 对任意 $x_1, x_2 \in \mathbf{R}^n$, 下式成立:

$$||p_S(x_1)-p_S(x_2)||^2 \leqslant (p_S(x_1)-p_S(x_2))^{\mathrm{T}}(x_1-x_2). \tag{1.2.4}$$

证明　在式 (1.2.3) 中, 取

$$x = x_1, \quad y = p_S(x_2) \in S,$$

则有

$$(p_S(x_2)-p_S(x_1))^{\mathrm{T}}(x_1-p_S(x_1)) \leqslant 0.$$

类似地有

$$(p_S(x_1)-p_S(x_2))^{\mathrm{T}}(x_2-p_S(x_2)) \leqslant 0.$$

上述两式相加得

$$(p_S(x_1)-p_S(x_2))^{\mathrm{T}}(x_2-x_1+p_S(x_1)-p_S(x_2)) \leqslant 0,$$

故式 (1.2.3) 成立. 定理得证.

由定理 1.2.2, 我们立刻得到下述两个有趣的结论:

$$0 \leqslant (p_S(x_1)-p_S(x_2))^{\mathrm{T}}(x_1-x_2), \quad \forall x_1, x_2 \in \mathbf{R}^n,$$

即投影算子 p_S 的单调性. 结合 Cauchy-Schwarz 不等式得到

$$||p_S(x_1)-p_S(x_2)|| \leqslant ||x_1-x_2||, \tag{1.2.5}$$

即投影算子 p_S 的非膨胀性. 特别是如果 $0 \in S$, 则有 $||p_S(x)|| \leqslant ||x||$.

定理 1.2.3　设 $S \subset \mathbf{R}^n$ 是闭凸锥, 则 y_x 是点 x 到 S 投影的充要条件是

$$y_x \in S, \quad x-y_x \in S^{\circ}, \quad (x-y_x)^{\mathrm{T}}y_x = 0. \tag{1.2.6}$$

证明　必要性. 设 y_x 是 x 到 S 的投影, 根据定理 1.2.1, 有

$$(x-y_x)^{\mathrm{T}}(y-y_x) \leqslant 0, \quad \forall y \in S. \tag{1.2.7}$$

将 $y=\alpha y_x$, $\alpha \geqslant 0$, 代入式 (1.2.7) 得

$$(\alpha-1)(x-y_x)^{\mathrm{T}}y_x \leqslant 0,$$

由于 $\alpha-1$ 可取正值或负值, 于是有

$$(x-y_x)^{\mathrm{T}}y_x=0,$$

这样式 (1.2.7) 变为

$$y^{\mathrm{T}}(x-y_x)\leqslant 0, \quad \forall y\in S,$$

即

$$x-y_x\in S^{\circ}.$$

充分性. 设 y_x 满足式 (1.2.6). 对任意 $y\in S$, 由式 (1.2.2) 的记号, 有

$$\begin{aligned} f_x(y)=\frac{1}{2}||x-y_x+y_x-y||^2 \geqslant f_x(y)\\ =f_x(y_x)+(x-y_x)^{\mathrm{T}}(y_x-y). \end{aligned}$$

又由式 (1.2.6) 可得

$$(x-y_x)^{\mathrm{T}}(y_x-y)=-(x-y_x)^{\mathrm{T}}\geqslant 0,$$

于是

$$f_x(y)\geqslant f_x(y_x),$$

故 y_x 是式 (1.2.1) 的解. 定理得证.

1.3 凸集的分离定理

平面中两个不相交凸集的一个明显几何性质, 就是存在一条直线将它们分开, 使得一个集合在直线一侧, 另外一个集合在直线的另一侧. 这一事实在一般的 n 维空间中就是所谓的凸集分离定理. 凸集的分离定理在最优化理论, 特别是在最优性条件的建立中起着重要的作用. 本节介绍几个凸集分离定理.

1.3.1 分离定理

首先介绍单点集与凸集的分离定理.

定理 1.3.1 (分离定理) 设 $S\subset \mathbf{R}^n$ 是非空闭凸集, $x\in\mathbf{R}^n$ 且 $x\notin S$, 则存在 $p\in\mathbf{R}^n$, 使得

$$\sup_{y\in S} p^{\mathrm{T}}y<p^{\mathrm{T}}x. \tag{1.3.1}$$