计算方法丛书

约束最优化计算方法

赵凤治 蔚继英 著

科学出版社

计算方法丛书

约束最优化计算方法

赵凤治 尉继英 著

科学出版社

1991

内 容 简 介

本书系统地论述了约束最优化中常用的计算方法和新算法，以及这些方法的计算框图和在计算机上实现的计算方案. 主要内容包括：二次规划算法、直接法、系列无约束最优化方法、容许方向法、简约梯度法、约束变尺度法等. 本书取材着眼于方法的实用性和全面性.

本书读者对象：从事计算机应用、工程设计、管理科学研究的工作者，以及大专院校有关专业的师生.

计算方法丛书

约束最优化计算方法

赵凤治　尉继英　著

责任编辑　林　鹏

科学出版社出版

北京东黄城根北街16号

邮政编码：100707

中国科学院印刷厂印刷

新华书店北京发行所发行　　各地新华书店经售

*

1991年8月第一版　　开本：850×1168 1/32

1991年8月第一次印刷　　印张：10 7/8

印数：0001—1 800　　字数：285 000

ISBN 7-03-002309-9/O・434

定价：12.20元

《计算方法丛书》编委会

序

在科学研究、工农业生产、金融及商业活动、军事对抗等领域中，参与者往往有许多决策可供选择．采用不同的决策常常导致不同的结局．人们自然希望自己能采用最好的决策．研究最优决策的学科称为最优化．

最优化的数学形式是极值问题．极值问题很自然地分为无约束极值及约束极值两类问题．对于约束极值(或称为约束最优化)问题,尽管早已有了 Lagrange 函数法,但现实生活中大量出现的问题远远地超出了古典极值方法研究的范围．继 Kuhn-Tucker (1951) 的工作之后，约束最优化问题引起了学者们的广泛重视，得到了极大的发展．

最优化是一个面向实际应用的学科，单纯地认识和刻画最优化决策满足不了客观要求，必须具体地给出最优决策的数值描述才能让人们见诸行动．很多实际问题都很复杂，而构造一个最优决策必须借助于计算机才能实现．因此讨论计算机上采用的约束最优化数值方法就成了最优化学科的基础课题之一．

我们自 1960 年以来,一直从事最优化数值方法的基础研究和应用研究工作．通过多年的学习及研究工作中的体会，对于最优化的数值方法有了一些理性和感性认识．在此基础上形成了这本书．作者曾接受一些单位的邀请，以本书稿为教材给研究生开过课,在教学过程中的许多反映证实了其内容是能够为读者接受的．

约束最优化问题内容十分丰富．诸如建模的理论和方法、最优化的数学理论、对偶理论、稳定性分析等许多方面都有丰硕的成就．本书作为约束最优化数值方法的专著，我们把它的内容限制在以介绍数值方法为主,对于方法的理论分析只做了有限的叙述．但是,那种不管解的性态而去构造算法的做法,我们认为是不可取

的，所以在第二章用了较多的篇幅来介绍解的性质.

本书对计算方法的取材首先着眼于实用性. 对于那些虽然在理论上是完整的、严密的，但在实用中不够方便的算法我们是弃而不取的. 第二，力求于新，我们整理了一些宜于列入本书的方法，尽管它们还未公开发表. 第三考虑了方法的全面性，既注意方法的实效性，也注意方法的代表性. 书中有经长期考验被认为是有效的方法，也有一些很有实用价值的新方法.

本书完整地介绍了近 20 种计算方法，还粗略地介绍了其它一些方法. 但是，在实用中会发现，这些方法在使用时还要做适当的改动才会收到好的效果. 我们认为，学习方法必须把握方法的实质. 而修改方法、构造方法时要以解的性质做指导.

我们虽然努力追求书本身的完备性，但是，我们还是认为读者对最优化有了一定的了解. 在介绍方法时，我们没有给出帮助理解的例子，仅在每章后面留了一些习题.

为了查阅方便，在附录中介绍了解线性规划的单纯形法.

书中收入了作者本人的部分研究结果及对方法的理解，但受水平限制，有些内容很粗浅.

在本书撰写过程中得到了许多朋友的关心和支持. 参阅了很多同志的专著、论文、讲义、手稿. 对此，作者表示衷心的感谢!

作者希望广大读者提出宝贵的批评意见，以便使本书有所充实和提高.

目　录

第一章　引　　言

§1　问题的数学描述

本书自始至终都是讨论如下的数学问题：

$$\min f(\boldsymbol{x}) \tag{1.1}$$

满足

$$\begin{aligned} &g_i(\boldsymbol{x}) = 0, \quad i = 1,2,\cdots,p, \\ &g_i(\boldsymbol{x}) \leqslant 0, \quad i = p+1,p+2,\cdots,m, \end{aligned} \tag{1.2}$$

其中 $\boldsymbol{x} \in \mathbf{R}^n$，今后不再说明．整个问题称为约束最优化问题，或者称为约束极值问题、数学规划问题、规划问题等．$f(\boldsymbol{x})$ 称为问题的目标函数，(1.2) 称为问题的约束条件．

对于约束最优化问题有很多内容值得讨论，这里只讨论约束最优化问题的数值解法及有关的一部分理论问题．

如果没有约束条件，则上述问题称为无约束极值问题．有关无约束极值问题的解法，可参看本丛书的《无约束最优化计算方法》[1]和其他有关资料．

当出现的函数 $f(\boldsymbol{x})$，$g_i(\boldsymbol{x})$ $(i = 1,2,\cdots,m)$都是线性函数时，上述问题称为线性规划问题．有关线性规划问题的解法，可参看本丛书的《线性规划计算方法》[2]和其他有关资料．

本书讨论当 $m \neq 0$ 时，(1.1)，(1.2) 中有非线性函数出现的非线性约束最优化问题的计算方法．

记

$$\begin{aligned} R = \{\boldsymbol{x} \mid g_i(\boldsymbol{x}) = 0, i = 1,2,\cdots,p;\ g_i(\boldsymbol{x}) \geqslant 0, \\ i = p+1,p+2,\cdots,m\}, \end{aligned}$$

问题又可表示为

$$\min\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in R\},$$

$\boldsymbol{R}$ 称为容许集合.

若 $\boldsymbol{x}^*$ 满足

$$\boldsymbol{x}^* \in R, f(\boldsymbol{x}^*) \leqslant f(\boldsymbol{x}), \forall \boldsymbol{x} \in R, \tag{1.3}$$

则称 $\boldsymbol{x}^*$ 为 $f(\boldsymbol{x})$ 于 R 上的整体极小值点，或者称 $\boldsymbol{x}^*$ 为非线性最优化问题的解. $f(\boldsymbol{x}^*)$ 称为非线性最优化问题的最优值.

记

$$U_\varepsilon(\boldsymbol{x}^0) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^0\| \leqslant \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} > 0\},$$

若存在 $\varepsilon > 0$,使 $\boldsymbol{x}^0$ 满足

$$\boldsymbol{x}^0 \in R, f(\boldsymbol{x}^0) \leqslant f(\boldsymbol{x}), \forall \boldsymbol{x} \in R \cap U_\varepsilon(\boldsymbol{x}^0),$$

则称 $\boldsymbol{x}^0$ 为 $f(\boldsymbol{x})$ 于 R 上的局部极小值点.

显然一个非线性最优化问题可能出现的情形是:

i) $R = \varnothing$,问题无容许解;

ii) $R \neq \varnothing$, $\inf\limits_{\boldsymbol{x} \in R}\{f(\boldsymbol{x})\}$ 有界,存在 $\boldsymbol{x}^*$ 使

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x} \in R}\{f(\boldsymbol{x})\} = \inf_{\boldsymbol{x} \in R}\{f(\boldsymbol{x})\};$$

iii) $R \neq \varnothing$, $\inf\limits_{\boldsymbol{x} \in R}\{f(\boldsymbol{x})\} = f^0$ 有界，但不存在 $\boldsymbol{x}^0 \in \boldsymbol{R}$ 使 $f(\boldsymbol{x}^0) = f^0$. 此时问题无精确解;

iv) $R \neq \varnothing$, $\inf\limits_{\boldsymbol{x} \in R}\{f(\boldsymbol{x})\} = -\infty$,问题为无界解情形.

显然对于一个约束条件 $g_i(\boldsymbol{x}) \geqslant 0$ 一个点 $\boldsymbol{x}^0$ 可能出现的情形是:

i) $g_i(\boldsymbol{x}^0) < 0$, $\boldsymbol{x}^0$ 为不满足约束条件 $g_i(\boldsymbol{x}) \geqslant 0$ 的点;

ii) $g_i(\boldsymbol{x}^0) = 0$, $\boldsymbol{x}^0$ 为约束条件 $g_i(\boldsymbol{x}) \geqslant 0$ 的边界点;

iii) $g_i(\boldsymbol{x}^0) > 0$, $\boldsymbol{x}^0$ 为约束条件 $g_i(\boldsymbol{x}) \geqslant 0$ 的内点.

对于一个约束最优化问题来说，若 $\boldsymbol{x}^0$ 不满足 R 的一个约束条件,则 $\boldsymbol{x}^0 \notin R$. 若 $\boldsymbol{x}^0$ 在 R 的至少一个约束条件边界上,则 $\boldsymbol{x}^0$ 为 R 的边界点.若 $\boldsymbol{x}^0$ 为 R 的所有约束条件的内点,则 $\boldsymbol{x}^0$ 为 R 的内点.

§2 凸 规 划

凸规划是一类重要的约束最优化问题，尤其在理论上有它特

殊的作用．因此我们引进凸规划这一概念．

一个集合 C，若对任一 $\boldsymbol{x}_1 \in C$，及任一 $\boldsymbol{x}_2 \in C$ 及对任一 $\lambda \in [0,1]$ 都有

$$\boldsymbol{x} = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in C,$$

则称 C 为凸集合．

一个函数定义在凸集 C 上，若对任意 $\boldsymbol{x}_1 \in C$，$\boldsymbol{x}_2 \in C$ 及任一 $\lambda \in [0,1]$，有

$$f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \leqslant \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2), \quad (1.4)$$

则称 $f(\boldsymbol{x})$ 为凸函数．

当 $-f(\boldsymbol{x})$ 为凸函数时，则称 $f(\boldsymbol{x})$ 为凹函数．

在(1.4)式中，若当 $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ 时，且 $\lambda \neq 0, \lambda \neq 1$，不等式一定成立为"<"，则称 $f(\boldsymbol{x})$ 为严格凸函数．

同样可定义严格凹函数 $f(\boldsymbol{x})$．

可微凸函数有下面的定理

定理 1.1 C 为非空开凸集，$f(\boldsymbol{x})$ 在 C 上可微，则 $f(\boldsymbol{x})$ 为 C 上凸函数的充分必要条件是：对于任意的 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in C$，总有

$$f(\boldsymbol{x}_1) \geqslant f(\boldsymbol{x}_2) + \nabla f^T(\boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2).$$

证明 首先证明必要性．由凸函数的定义可以得出

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

对任意 $0 \leqslant \lambda \leqslant 1$，即

$$f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2) \geqslant \frac{1}{\lambda}[f(\boldsymbol{x}_2 + \lambda(\boldsymbol{x}_1 - \boldsymbol{x}_2)) - f(\boldsymbol{x}_2)].$$

取 λ 趋于 0 就得出

$$f(\boldsymbol{x}_1) \geqslant f(\boldsymbol{x}_2) + \nabla f^T(\boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2).$$

现在证明充分性．假定所给不等式成立，取

$$\boldsymbol{z} = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2,$$

则有

$$f(\boldsymbol{x}_1) \geqslant f(\boldsymbol{z}) + \nabla f^T(\boldsymbol{z})(\boldsymbol{x}_1 - \boldsymbol{z}),$$
$$f(\boldsymbol{x}_2) \geqslant f(\boldsymbol{z}) + \nabla f^T(\boldsymbol{z})(\boldsymbol{x}_2 - \boldsymbol{z}).$$

由此便可得出

$$\lambda(f(\boldsymbol{x}_1)-f(\boldsymbol{z}))+(1-\lambda)(f(\boldsymbol{x}_2)-f(\boldsymbol{z}))$$
$$\geqslant \nabla f^T(\boldsymbol{z})[\lambda(\boldsymbol{x}_1-\boldsymbol{z})+(1-\lambda)(\boldsymbol{x}_2-\boldsymbol{z})]=0,$$

即

$$\lambda f(\boldsymbol{x}_1)+(1-\lambda)f(\boldsymbol{x}_2)\geqslant f(\boldsymbol{z})=f(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2).$$

对于约束最优化问题

$$\min f(\boldsymbol{x})$$

满足

$$g_i(\boldsymbol{x})=0, i=1,2,\cdots,p,$$
$$g_i(\boldsymbol{x})\leqslant 0, i=p+1,p+2,\cdots,m.$$

若 $f(\boldsymbol{x})$，$g_i(\boldsymbol{x})$ $(i=p+1,\ p+2,\ \cdots,\ m)$ 是凸函数，$g_i(\boldsymbol{x})$ $(i=1,2,\cdots,p)$ 是线性函数，则问题称为凸规划.

定理 1.2 对于凸规划有

i) 容许集合 R 是凸集合；

ii) 每一个局部极小值点都是整体极小值点，即规划问题的最优解；

iii) 最优解的集合为凸集合；若目标函数为严格凸函数，则最优解唯一.

证明 i) 设有 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$ 使 $g_i(\boldsymbol{x}_1)=0$，$(i=1,2,\cdots,p)$，$g_i(\boldsymbol{x}_2)=0,(i=1,2,\cdots,p)$；$g_i(\boldsymbol{x}_1)\leqslant 0,(i=p+1,p+2,\cdots,m)$，$g_i(\boldsymbol{x}_2)\leqslant 0,(i=p+1,p+2,\cdots,m)$.对于任一 $\lambda\in[0,1]$，有

$$g_i(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2)=\lambda g_i(\boldsymbol{x}_1)+(1-\lambda)g_i(\boldsymbol{x}_2)=0 \quad (i=1,2,\cdots,p),$$

$$g_i(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2)\leqslant\lambda g_i(\boldsymbol{x}_1)+(1-\lambda)g_i(\boldsymbol{x}_2)=0 \quad (i=p+1,p+2,\cdots,m),$$

这便证明了容许集合 R 是凸集合.

ii) 记 $\boldsymbol{x}^0$ 为任一局部极小值点，$\boldsymbol{x}^*$ 为整体极小值点，且 $\boldsymbol{x}^0\neq\boldsymbol{x}^*$，$f(\boldsymbol{x}^*)<f(\boldsymbol{x}^0)$. 由于 R 是凸集合，故对任何 $\lambda\in[0,1]$ 有

$$\lambda x^* + (1-\lambda)x^0 \in R.$$

又由 $f(x)$ 为凸函数，故有

$$f(\lambda x^* + (1-\lambda)x^0) \leqslant \lambda f(x^*) + (1-\lambda)f(x^0)$$
$$< \lambda f(x^0) + (1-\lambda)f(x^0) = f(x^0).$$

取 λ 趋近于 0，与 x^0 为局部极小值点矛盾。

iii) 设 x_1, x_2 均为最优解，且 $x_1 \neq x_2$。由 i) 知，对任何 $\lambda \in [0,1]$ 有 $\lambda x_1 + (1-\lambda)x_2 \in R$。因为 $f(x)$ 为严格凸函数，故有

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2)$$
$$= f(x_1) = f(x_2).$$

这与 x_1, x_2 为最优解矛盾。

定理 1.3 设 x^* 为凸规划

P: $$\min\{f(x) \mid g_i(x) \leqslant 0, i = 1,2,\cdots,m\}$$

的最优解，记

$$\bar{I} = \{i \mid g_i(x^*) = 0\},$$

则 x^* 也是规划

$P1$: $$\min\{f(x) \mid g_i(x) \leqslant 0, i \in \bar{I}\}$$

的最优解。

证明 x^* 明显地是 $P1$ 的容许解。若 x^* 不是 $P1$ 的最优解，记 x^0 为 $P1$ 的最优解，则有 $f(x^0) < f(x^*)$。记

$$I_0 = \{i \mid g_i(x^0) > 0\} \setminus \bar{I}.$$

我们取

$$\lambda_0 = \max_{i \in I_0}\left\{\frac{g_i(x^0)}{g_i(x^0) - g_i(x^*)}\right\}. \tag{1.5}$$

明显地有 $0 < \lambda_0 < 1$。容易验证

$$g_i(\lambda_0 x^* + (1-\lambda_0)x_0) \leqslant 0, \ i = 1,2,\cdots,m.$$

由 $f(x)$ 的凸性，有

$$f(\lambda_0 x^* + (1-\lambda_0)x_0)$$
$$\leqslant \lambda_0 f(x^*) + (1-\lambda_0)f(x^0) < f(x^*).$$

这与 x^* 是 P 的最优解矛盾。

定理 1.4 令 $\boldsymbol{x}^*$ 为凸规划问题

$$P:\quad \min\{f(\boldsymbol{x})\,|\,g_i(\boldsymbol{x})=0,\ i=1,\ 2,\cdots,\ p;g_i(\boldsymbol{x})\leqslant 0,\ i=p+1,\ p+2,\cdots,m\}$$

的容许解,问题中出现的所有函数于 $\boldsymbol{x}^*$ 点均可微. 若存在常数 $u_i(i=1,2,\cdots,p),u_i\geqslant 0(i=p+1,p+2,\cdots,m)$,使

$$\nabla f(\boldsymbol{x}^*)+\sum_{i=1}^{m} u_i\nabla g_i(\boldsymbol{x}^*)=0 \tag{1.6}$$

及

$$u_i=0,\ 当\ g_i(\boldsymbol{x}^*)<0时, \tag{1.7}$$

则 $\boldsymbol{x}^*$ 是规划问题 P 的最优解.

证明 对于线性函数 $g_i(\boldsymbol{x})(i=1,2,\cdots,p)$,一定有

$$g_i(\boldsymbol{x})=g_i(\boldsymbol{x}^*)+(\boldsymbol{x}-\boldsymbol{x}^*)^T\nabla g_i(\boldsymbol{x}^*).$$

对于凸函数 $g_i(\boldsymbol{x})(i=p+1,p+2,\cdots,m)$,一定有

$$g_i(\boldsymbol{x})\geqslant g_i(\boldsymbol{x}^*)+(\boldsymbol{x}-\boldsymbol{x}^*)^T\nabla g_i(\boldsymbol{x}^*).$$

同样对 $f(\boldsymbol{x})$, 也有

$$f(\boldsymbol{x})\geqslant f(\boldsymbol{x}^*)+(\boldsymbol{x}-\boldsymbol{x}^*)^T\nabla f(\boldsymbol{x}^*).$$

于是

$$f(\boldsymbol{x})+\sum_{i=1}^{m} u_ig_i(\boldsymbol{x})\geqslant f(\boldsymbol{x}^*)+\sum_{i=1}^{m} u_ig_i(\boldsymbol{x}^*)+[\nabla f(\boldsymbol{x}^*)+\sum_{i=1}^{m} u_i\nabla g_i(\boldsymbol{x}^*)]^T(\boldsymbol{x}-\boldsymbol{x}^*).$$

由 (1.6),上式可化为

$$f(\boldsymbol{x})+\sum_{i=1}^{m} u_ig_i(\boldsymbol{x})\geqslant f(\boldsymbol{x}^*)+\sum_{i=1}^{m} u_ig_i(\boldsymbol{x}^*),$$

注意 (1.7),有

$$f(\boldsymbol{x})+\sum_{i=1}^{m} u_ig_i(\boldsymbol{x})\geqslant f(\boldsymbol{x}^*).$$

对于任意容许解 $\boldsymbol{x}$, 一定有

$$f(\boldsymbol{x})\geqslant f(\boldsymbol{x})+\sum_{i=1}^{m} u_ig_i(\boldsymbol{x}),$$

故有

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}^*).$$

所以 $\boldsymbol{x}^*$ 是规划问题 P 的最优解.

凸规划还有一些有趣的性质,凸性也还有一些有趣的扩展,这里我们就不一一介绍了. 读者可参看 [3],[4],[5].

§3 Farkas 引 理

Farkas 引理在约束最优化中有很重要的作用,为了后面讨论问题的方便,我们在这里给予介绍.

如果用线性不等式组的对偶理论,证明 Farkas 引理是并不复杂的,对此可参看 [2] 及 [6] 的有关部分.我们这里采用 [5] 中的方法.

我们首先讨论分离平面定理.

假定 S_1 及 S_2 为 n 维空间中的两个集合,如果有一个超平面

$$H = \{x \mid p^T x = \alpha\}$$

存在,并且使

$$p^T x \geqslant \alpha,\ \text{对所有}\ x \in S_1;$$

$$p^T y \leqslant \alpha,\ \text{对所有}\ y \in S_2,$$

成立,则称 H 为 S_1 及 S_2 的分离平面. 特别,如果有

$$p^T x > \alpha,\ \text{对所有}\ x \in S_1;$$

$$p^T y < \alpha,\ \text{对所有}\ y \in S_2,$$

则称 H 严格分离集合 S_1 及 S_2.

引理 1.5 设 R 为不包含原点的非空闭凸集,则存在一个超平面严格分离 R 及原点.

证明 令

$$U_\delta(\mathbf{0}) = \{x \mid \|x\| \leqslant \delta\}, \delta > 0.$$

选择 δ 使 $R \cap U_\delta(\mathbf{0}) \neq \varnothing$. $R \cap U_\delta(\mathbf{0})$ 是一紧集,连续函数 $\|\boldsymbol{x}\|$ 于其上的极小值点为 $\boldsymbol{x}^0$. 明显地有 $\|\boldsymbol{x}^0\| > 0$. 对于任意 $\boldsymbol{x} \in R$,均有 $\|\boldsymbol{x}\| \geqslant \|\boldsymbol{x}^0\|$. 因为 R 为凸集,自然也有

$$\|\lambda x + (1-\lambda)x^0\| \geqslant \|x^0\|, \forall x \in R, 0 \leqslant \lambda \leqslant 1.$$

由此得出

$$\|\lambda x + (1-\lambda)x^0\|^2 \geqslant \|x^0\|^2,$$

$$(\lambda x + (1-\lambda)x^0)^T(\lambda x + (1-\lambda)x^0) \geqslant x^{0T}x^0,$$

$$\lambda^2(x - x^0)^T(x - x^0) + 2\lambda x^{0T}(x - x^0) \geqslant 0. \qquad (1.8)$$

假定 $x^{0T}(x - x^0) < 0$,我们只要选

$$\lambda < \frac{-2x^{0T}(x - x^0)}{(x - x^0)^T(x - x^0)},$$

便知和 (1.8) 矛盾. 所以一定有

$$x^{0T}(x - x^0) \geqslant 0,$$

$$x^{0T}x \geqslant x^{0T}x^0.$$

令 $\alpha = x^{0T}x^0/2$,定义超平面

$$H = \{x \mid x^{0T}x = \alpha\}.$$

对此超平面 H,任取 $x \in R$, 有

$$x^{0T}x \geqslant x^{0T}x > \alpha;$$

而对原点 $\mathbf{0}$ 有

$$x^{0T}\mathbf{0} = 0 < \alpha.$$

所以超平面H把原点和凸集 R 分离开.

引理 1.6 R_1, R_2 为不相交的非空闭凸集，其中 R_2 是紧的，则存在超平面H严格分离 R_1, R_2.

证明 由于 R_2 是紧集,所以 $R = R_1 - R_2$ 是闭凸集，并且 $\mathbf{0} \notin R$. 据引理 1.4 有 H_R 严格分离 R 及原点. 此处取

$$H_R = \{x \mid x^{0T}x = \alpha = x^{0T}x^0/2\}$$

x^0 为R中离原点最近的点. 令

$$x = x_1 - x_2,\ x_1 \in R_1,\ x_2 \in R_2,$$

$$x^0 = x_1^0 - x_2^0,\ x_1^0 \in R_1, x_2^0 \in R_2,$$

我们得到

$$(x_1^0 - x_2^0)^T(x_1 - x_2) > \alpha > 0.$$

所以

$$\inf_{x_1 \in R_1}(x_1^0 - x_2^0)^T x_1 \geqslant \sup_{x_2 \in R_2}(x_1^0 - x_2^0)^T x_2 + \alpha$$

$$> \sup_{x_2 \in R_2} (\boldsymbol{x}_1^0 - \boldsymbol{x}_2^0)^T \boldsymbol{x}_2.$$

任选 β 使

$$\inf_{x_1 \in R_1} (\boldsymbol{x}_1^0 - \boldsymbol{x}_2^0)\boldsymbol{x}_1 > \beta > \sup_{x_2 \in R_2} (\boldsymbol{x}_1^0 - \boldsymbol{x}_2^0)^T \boldsymbol{x}_2,$$

则定义

$$H = \{\boldsymbol{x} | \boldsymbol{x}^{0T}\boldsymbol{x} = \beta\}$$

即为严格分离 R_1, R_2 的超平面.

引理 1.7 (Farkas 引理) A 为 $m \times n$ 实矩阵, $\boldsymbol{b}$ 为 n 维实向量. 则对所有满足 $A\boldsymbol{y} \geqslant \boldsymbol{0}$ 的 $\boldsymbol{y}$ 均有 $\boldsymbol{b}^T\boldsymbol{y} \geqslant 0$ 的充分必要条件为: 存在 m 维向量 $\boldsymbol{x} \geqslant \boldsymbol{0}$ 使 $A^T\boldsymbol{x} = \boldsymbol{b}$.

证明 本命题等价于 $A\boldsymbol{y} \geqslant 0$, $\boldsymbol{b}^T\boldsymbol{y} < 0$ 有解的充要条件是 $A^T\boldsymbol{x} = \boldsymbol{b}$, $\boldsymbol{x} \geqslant \boldsymbol{0}$ 无解. 后者说明下列非空闭凸集 R_1, R_2 不相交:

$$R_1 = \{\boldsymbol{z} | \boldsymbol{z} = A^T\boldsymbol{x}, \boldsymbol{x} \geqslant 0\},$$
$$R_2 = \{\boldsymbol{b}\}.$$

明显 R_2 是紧集. 由引理 1.6, 存在一个非零向量 $\boldsymbol{x}^0$ 及一个实数 α 使 $\boldsymbol{x}^{0T}\boldsymbol{b} < \alpha$, 对每一个 $\boldsymbol{x} \in R_1$ 有 $\boldsymbol{x}^{0T}\boldsymbol{x} > \alpha$, 或者等价地对每一个 $\boldsymbol{w} \geqslant \boldsymbol{0}$ 有 $\boldsymbol{x}^{0T}\boldsymbol{A}^T\boldsymbol{w} > \alpha$. 令 $\boldsymbol{w} = \boldsymbol{0}$, 得出 $\alpha < 0$. 对 $j = 1, 2, \cdots, m$, 令 $\boldsymbol{w} = (0, \cdots, w_j, 0, \cdots, 0)^T$, 其中 w_j 是一个很大的正数, 得出 $\boldsymbol{x}^{0T}A^T \geqslant \boldsymbol{0}$, 因此 $A\boldsymbol{x}^0 \geqslant \boldsymbol{0}$, 所以 $\boldsymbol{x}^0$ 是 $A\boldsymbol{y} \geqslant 0$, $\boldsymbol{b}^T\boldsymbol{y} < 0$ 的解. 反之, 若 $A^T\boldsymbol{x} = \boldsymbol{b}$, $\boldsymbol{x} \geqslant \boldsymbol{0}$, 又 $A\boldsymbol{y} \geqslant 0$, 则 $\boldsymbol{b}^T\boldsymbol{y} = \boldsymbol{x}^T A\boldsymbol{y} \geqslant 0$.

引理 1.8 设 R 为不包含原点的非空凸集合, 存在超平面 H 分离原点和集合 R.

证明 记

$$P(\boldsymbol{x}) = \{\boldsymbol{y} | \boldsymbol{y}^T\boldsymbol{y} = 1, \boldsymbol{y}^T\boldsymbol{x} \geqslant 0\}.$$

$P(\boldsymbol{x})$ 为非空闭集合. 对任一 k 及 $\boldsymbol{x}_i \in R (i = 1, 2, \cdots, k)$ 记

$$Q = \{\boldsymbol{x} | \boldsymbol{x} = \sum_{i=1}^{k} \alpha_i \boldsymbol{x}_i, \sum_{i=1}^{k} \alpha_i = 1, \boldsymbol{x}_i \in R, \alpha_i \geqslant 0,$$
$$i = 1, 2, \cdots, k\},$$