约束最优化方法

何旭初
孙麟平　编著

南京大学出版社

约束最优化方法

何旭初
孙麟平　编著

南京大学出版社

内 容 简 介

本书系统地阐述了约束最优化问题的基本理论和数值方法。全书共分六章。前三章是理论基础，内容包括约束极值问题的最优性条件，对偶问题，算法理论；后三章是求解约束最优化问题的数值方法，分类讨论了线性约束最优化问题、二次规划问题、非线性约束最优化问题的各种经典算法以及近年来提出的一些新算法。为了方便读者，所有算法都给出了详细步骤和综合评述。

本书可作为高等学校有关专业“最优化”课程的教材；也可供从事系统分析、运筹学、数值分析、管理科学的工程技术人员，高年级学生、研究生，以及教师参考或自学。

约 束 最 优 化 方 法

何旭初 孙麟平 编著

南京大学出版社出版

（南京大学校内）

江苏省新华书店发行 国营丹徒印刷厂印刷

开本 787×1092 1/32 印张 6.875 字数 154千

1986年 3 月第 1 版 1986年 5 月第 1 次印刷

印数 1—2,500

统一书号：13336·001 定价：1.30元

责任编辑 秦 涛

序

自1974年起，我校计算数学专业开始设置最优化方法课，当时的内容仅限于无约束最优化方法。所编的教材已于1977年由科学出版社出版。近年来，最优化方法得到了迅速的发展，它的应用也日趋广泛。为了适应这种形势，从1981年起又增设了约束最优化方法这门课程。所用的教材经过几年的教学实践，作了多次修改与补充，于今年春天初步定稿。

本书比较系统地阐述了约束最优化问题的基本理论，分类介绍了求解约束最优化问题的一系列有效算法。全书共分六章，前三章是理论基础，分别阐述了约束极值存在的条件和性质、约束极值问题的对偶以及算法理论；后三章是求解约束最优化问题的数值方法，讨论了线性约束极值问题、二次规划问题以及非线性约束极值问题的常用算法和近年提出的新算法。为了适应不同类型的读者，我们对有些章节加注了“*”。略去“*”号的章节后，本书依然自成一体。仅对某类方法感兴趣的读者，可以在读完第一章后直接阅读有关章节。书末的“文献索引”栏分为两部分，A类文献直接与本书内容有关；B类文献按专题编排，为有志于进一步研究约束最优化方法的读者提供了扩大视野的窗口。此外，本书还设置了“算法索引”和“名词索引”以便读者查阅。

限于作者水平，缺点和错误在所难免，敬请读者批评、指正。

作　者

1985年4月于南京大学数学系

目 录

第1章 约束极值问题的最优性条件

第 2 章　对偶理论

第 3 章　算法理论初步

第 4 章　线性约束问题

第 5 章　二次规划问题

第 6 章　非线性约束问题

第1章

约束极值问题的最优性条件

在实际工作中，为了可靠地反映工程技术和管理系统中的有关问题，往往导致非线性数学模型的建立，即构成所谓非线性规划问题。其一般形式是：

$$(\mathrm{P}) \qquad \min_{x\in R^n} \quad f(x)$$
$$\mathrm{s.t} \quad x\in S$$

当$S=R^n$时，(P)称为无约束最优化问题；当$S\subset R^n$时，(P)称为约束最优化问题，简称约束极值问题。

在约束极值问题中，S称为约束区域。满足约束区域的向量x称为问题(P)的可行解。若x^*是问题(P)的可行解，并且对任一可行解x皆有$f(x)\geqslant f(x^*)$成立，则称x^*为问题(P)的最优解；相应的函数值$f(x^*)$称为最优目标值。

本章主要讨论约束极值问题最优解x^*的有关性质，为建立求解问题(P)的有效算法奠定基础。

1.1 等式约束问题的最优性条件

考虑仅含等式约束的极值问题

$$(\mathrm{NEP}) \qquad \min \quad f(x)$$
$$\mathrm{s.t} \quad h_i(x)=0, \quad i=1,\cdots,l,$$

这里$x\in R^n$，$l<n$。

直观的想法是：若能从l个约束方程中解出l个变量x_{k_1}，

x_{k2}，…，x_{kl}代入目标函数中，则$f(\boldsymbol{x})$降维为$n-l$元函数，原约束问题也随之转化为无约束问题。

当然，从l个方程中解出l个变量是有条件的，这个条件在理论上称为隐函数定理。我们就是从这里出发，着手导出等式约束极值问题的最优性条件。

1.1.1 一阶必要条件

定理 1.1

在（NFP）问题中，设$\boldsymbol{x}^*$是可行解；$f(\boldsymbol{x})$和诸$h_i(\boldsymbol{x})$是$S\subset R^n$上的实值函数，且在邻域$N\varepsilon(\boldsymbol{x}^*)\subset S$上连续可微；对$N\varepsilon(\boldsymbol{x}^*)$中满足诸$h_i(\boldsymbol{x})=0$的一切点$\boldsymbol{x}$而言，$\boldsymbol{x}^*$是$f(\boldsymbol{x})$的一个局部极小点；诸$h_i(\boldsymbol{x}^*)$的Jacobi矩阵之秩为$l$；则存在实数

$\lambda_i^*, i=1,\cdots,l$使得

$$\nabla f(\boldsymbol{x}^*)=\sum_{i=1}^{l}\lambda_i^*\nabla h_i(\boldsymbol{x}^*) \tag{1.1}$$

【证明】 为不失一般性，假定诸$h_i(\boldsymbol{x})$的Jacobi矩阵J的前l行所构成的l阶子阵是非异的。

于是，线性方程组

$$\sum_{i=1}^{l}\frac{\partial h_i(\boldsymbol{x}^*)}{\partial x_j}\lambda_i=\frac{\partial f(\boldsymbol{x}^*)}{\partial x_j},\qquad j=1,\cdots,l \tag{1.2}$$

有唯一解λ_i^*。

令$\hat{\boldsymbol{x}}=(x_{l+1},\cdots,x_n)^T$，由隐函数定理，在$\boldsymbol{x}^*$处有

$$x_j^*=\phi_j(\hat{x}^*),\quad j=1,\cdots,l \tag{1.3}$$

$$f(\boldsymbol{x}^*)=f[\phi_1(\hat{\boldsymbol{x}}^*),\ \cdots,\ \phi_l(\hat{\boldsymbol{x}}^*),\ \boldsymbol{x}_{l+1}^*,\cdots,\boldsymbol{x}_n^*] \tag{1.4}$$

$$h_i[\phi_1(\hat{\boldsymbol{x}}^*),\cdots,\ \phi_l(\hat{\boldsymbol{x}}^*),\ \boldsymbol{x}_{l+1}^*,\cdots\boldsymbol{x}_n^*]=0 \tag{1.5}$$

这时，由无约束极小化的最优性条件可知，$f(\boldsymbol{x})$关于x_{l+1}，$\cdots,x_n$ 的偏导数在$\boldsymbol{x}^*$处必为零。

即
$$\sum_{k=1}^{l}\frac{\partial f(\boldsymbol{x}^*)}{\partial x_k}\cdot\frac{\partial\varphi_k(\hat{\boldsymbol{x}}^*)}{\partial x_j}+\frac{\partial f(\boldsymbol{x}^*)}{\partial x_j}=0$$

$$j=l+1,\ \cdots,\ n \tag{1.6}$$

另一方面，由(1.5)对于每个$j=l+1$，$\cdots$，n 有

$$\sum_{k=1}^{l}\frac{\partial h_i(\boldsymbol{x}^*)}{\partial x_k}\cdot\frac{\partial\phi_k(\hat{\boldsymbol{x}}^*)}{\partial x_j}+\frac{\partial h_i(\boldsymbol{x}^*)}{\partial x_j}=0$$

$$i=1,\cdots,l \tag{1.7}$$

以λ_i^*乘(1.7)中每一个方程并相加，得到

$$\sum_{i=1}^{l}\ \sum_{k=1}^{l}\lambda_i^*\ \frac{\partial h_i(\boldsymbol{x}^*)}{\partial x_k}\cdot\frac{\partial\phi_k(\hat{\boldsymbol{x}}^*)}{\partial x_j}+\sum_{i=1}^{l}\lambda_i^*$$

$$\times\frac{\partial h_i(\boldsymbol{x}^*)}{\partial x_j}=0,\qquad i=l+1,\ \cdots,\ n \tag{1.8}$$

从(1.6)减去(1.8)，整理后对于$i=l+1$，$\cdots$，n应有

$$\sum_{k=1}^{l}\left[\frac{\partial f(\boldsymbol{x}^*)}{\partial x_k}-\sum_{i=1}^{l}\lambda_i^*\ \frac{\partial h_i(\boldsymbol{x}^*)}{\partial x_k}\right]\frac{\partial\varphi_k(\hat{\boldsymbol{x}}^*)}{\partial x_j}$$

$$+\left[\frac{\partial f(\boldsymbol{x}^*)}{\partial x_j}-\sum_{i=1}^{l}\lambda_i^*\ \frac{\partial h_i(\boldsymbol{x}^*)}{\partial x_j}\right]=0 \tag{1.9}$$

由(1.2)可知，对于$i=l+1, \cdots, n$ 成立

$$\frac{\partial f(\boldsymbol{x}^*)}{\partial x_j}-\sum_{i=1}^{l} \lambda_i^* \frac{\partial h_i(\boldsymbol{x}^*)}{\partial x_j}=0 \qquad (1.10)$$

合并(1.2)和(1.10)即得(1.1). ∎

(1.1)式也叫做等式约束问题的**一阶必要条件**. 定理1.1启发我们，根据目标函数和约束函数两者梯度之间的关系去建立计算等式约束最优化问题的数值方法.

构造函数

$$L(\boldsymbol{x},\boldsymbol{\lambda})=f(\boldsymbol{x})-\boldsymbol{\lambda}^T h(\boldsymbol{x}) \qquad (1.11)$$

其中$\boldsymbol{\lambda}$是l维向量，$h(\boldsymbol{x})=(h_1(\boldsymbol{x}),h_2(\boldsymbol{x}),\cdots,h_l(\boldsymbol{x}))^T$.

设$(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)$是$n+l$元函数$L(\boldsymbol{x},\boldsymbol{\lambda})$的极小点、则由无约束问题$\min L(\boldsymbol{x},\boldsymbol{\lambda})$的最优性条件

$$\nabla L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)=0$$

得

$$\nabla_x L=\nabla f(\boldsymbol{x}^*)-\nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^*=0 \qquad (1.12)$$

$$\nabla_\lambda L=h(\boldsymbol{x}^*)=0 \qquad (1.13)$$

由定理1.1可知，$\boldsymbol{x}^*$是(NEP)问题最优解的必要条件为：它是函数$L(\boldsymbol{x},\boldsymbol{\lambda})$的最优解.通常称$L(\boldsymbol{x},\boldsymbol{\lambda})$，为Lagrange函数，$\boldsymbol{\lambda}$为Lagrange乘子向量.从而，等式约束最优化问题可以通过构造相应的Lagrange函数转化为无约束问题。

1.1.2 二阶充分条件

定理 1.2

在(NEP)问题中，设$f(\boldsymbol{x})$和$h_i(\boldsymbol{x})$，$i=1,\cdots,l$是R^n上的二次连续可微实函数，若存在向量$\boldsymbol{x}^*\in R^n$和$\boldsymbol{\lambda}^*\in R^l$，使得Lagrange函数的梯度

$$\nabla L(x^*, \lambda^*) = 0$$

并且，对每个满足

$$z^T \nabla h_i(x^*) = 0, \quad i = 1, \cdots, l$$

的非零向量$z \in R^n$，均有

$$z^T \nabla_x^2 L(x^*, \lambda^*) z > 0$$

成立，则x^*为(NEP)问题的严格局部最优解。

【证明】 用反证法。

假定x^*不是严格局部最优解，则必存在一个邻域$N_\delta(x^*)$和一个收敛于x^*的序列$\{z_k\}$，使得$z_k \in N_\delta(x^*)$，$z_k \neq x^*$，且对于每个$z_k \in \{z_k\}$有

$$h_i(z_k) = 0, \quad i = 1, \cdots, l \tag{1.14}$$

$$f(x^*) \geq f(z_k) \tag{1.15}$$

令$z_k = x^* + \theta_k y_k$，其中$\theta_k > 0$，$\| y_k \| = 1$，序列$\{\theta_k, y_k\}$有子序列收敛于$(0, \bar{y})$，而$\| \bar{y} \| = 1$。根据中值定理，对该子序列的每一个k，我们有

$$h_i(z_k) - h_i(x^*) = \theta_k (y_k)^T \nabla h_i(x^* + \eta_{ik} \theta_k y_k) = 0 \tag{1.16}$$

其中$0 < \eta_{ik} < 1$。

以θ_k除(1.16)，并令$k \to \infty$得

$$\bar{y}^T \nabla h_i(x^*) = 0, \quad i = 1, \cdots, l \tag{1.17}$$

由Taylor展式，可知

$$L(z_k, \lambda^*) = L(x^*, \lambda^*) + \theta_k y_k^T \nabla_x L(x^*, \lambda^*) + \frac{1}{2} \theta_k^2 \times y_k^T \nabla_x^2 L(x^* + \eta_k \theta_k y_k, \lambda^*) y_k \tag{1.18}$$

其中$0 < \eta_k < 1$。

根据定理的假设条件和(1.14)，我们有

$$f(z_k)-f(x^*)=\frac{1}{2}\theta_k^2 y_k^{\mathrm{T}}\nabla_x^2 L(x^*+\eta_k\theta_k y_k,\ \lambda^*)y_k$$

两边除以$\frac{1}{2}\theta_k^2$，并令$k\to\infty$ 得

$$\bar{y}\nabla_x^2 L(x^*,\ \lambda^*)y\leqslant 0 \qquad (1.19)$$

由于$\bar{y}\neq 0$且满足(1.17),故导出矛盾. ∎

定理1.2称为等式约束问题的二阶充分条件，它可以用来判定所求解是否是最优解.

1.1.3 经典Lagrange方法

在一般约束问题中，由于包含不等式约束，所以不能直接应用Lagrange方法。通常是引入松驰变量，使不等式约束变成等式约束，然后再用Lagrange方法求解. 我们通过例子来说明.

例 1.1

$$\begin{aligned}&\min \quad f(x_1,x_2)\\&\mathrm{s.t} \quad e(x_1,x_2)=b\\&\qquad\qquad x_1\geqslant a\end{aligned}$$

【解】引入松驰变量θ，令$\theta^2=x_1-a$,则约束变为

$$\begin{cases}e(x_1,x_2)-b=0\\ \theta^2-x_1+a=0\end{cases}$$

构造Lagrange函数

$$L(x_1,\ x_2,\theta,\lambda_1,\lambda_2)=f(x_1,x_2)+\lambda_1[e(x_1,x_2)-b]+\lambda_2[\theta^2-x_1+a]$$

由$\nabla L=0$，得

$$
\begin{cases}
\dfrac{\partial L}{\partial x_1}=\dfrac{\partial f}{\partial x_1}+\lambda_1\dfrac{\partial e}{\partial x_1}-\lambda_2=0\\[2mm]
\dfrac{\partial L}{\partial x_2}=\dfrac{\partial f}{\partial x_2}+\lambda_1\dfrac{\partial e}{\partial x_2}=0\\[2mm]
\dfrac{\partial L}{\partial \theta}=2\lambda_2\theta=0\\[2mm]
\dfrac{\partial L}{\partial \lambda_1}=e(x_1,\ x_2)-b=0\\[2mm]
\dfrac{\partial L}{\partial \lambda_2}=\theta^2-x_1+a=0
\end{cases}
$$

解此方程组即得问题的最优解。 ∎

鉴于这种求解约束最优化问题，在数值实现上是相当困难的。为此，还须寻找其他途径，以实现将约束问题转化为无约束问题求解的基本思想。

1.2 不等式约束问题的最优性条件

研究不等式约束问题

$$
\text{(NIP)}\qquad \min\ f(\boldsymbol{x}) \qquad \text{s.t}\quad e_i(x)\geqslant 0,\quad i=1,\ \cdots,\ m,
$$

讨论它的最优性条件要比等式约束问题复杂一些，为此需要引进新的概念和理论。

1.2.1 几何最优性条件

定义 1.1

设S是R^n中的非空子集，若$\overline{\boldsymbol{x}}\in S$，且对于某个非零向量$\boldsymbol{d}\in R^n$存在$\delta>0$，当$\alpha\in(0,\delta)$时使得$\overline{\boldsymbol{x}}+\alpha\boldsymbol{d}\in S$，则

称$\boldsymbol{d}$为区域S中点$\overline{\boldsymbol{x}}$处的可行方向向量。所有这样的$\boldsymbol{d}$的集合$D_0(\overline{\boldsymbol{x}})$称为区域$S$中点$\overline{\boldsymbol{x}}$处的可行方向集。

定义 1.2

设$f(\boldsymbol{x})$: $R^n \to R^1$在点$\overline{\boldsymbol{x}}$处可微，若存在向量$\boldsymbol{d} \in R^n$使得$\boldsymbol{d}^T \nabla f(\overline{\boldsymbol{x}}) < 0$，则称$\boldsymbol{d}$为$f(\boldsymbol{x})$在$\overline{\boldsymbol{x}}$处的下降方向。所有这样的$\boldsymbol{d}$的集合$F_0(\overline{\boldsymbol{x}})$称为$f(\boldsymbol{x})$在$\overline{\boldsymbol{x}}$处的下降方向锥。

上述二个定义告诉我们：从$\overline{\boldsymbol{x}}$起沿可行方向作一个很小的移动，可以得到另一个新的可行点；而从$\overline{\boldsymbol{x}}$出发沿下降方向作一个很小的移动，可以使目标函数值$f(\boldsymbol{x})$减少。综合这二点，不难想象：对于约束最优化问题而言，如果$\boldsymbol{x}^*$是(P)问题的局部最优解，那末$\boldsymbol{x}^*$处的下降方向一定不是它的可行方向。这个想法可以用下面的定理严格地描述。

定理 1.3

在问题(P)中，设$\boldsymbol{x}^* \in S$是其局部最优解，$f(\boldsymbol{x})$在点$\boldsymbol{x}^*$处可微，则点$\boldsymbol{x}^*$处的可行方向集和下降方向锥之交是空集，即$D_0(\boldsymbol{x}^*) \cap F_0(\boldsymbol{x}^*) = \phi$。

【证明】 用反证法

若$D_0(\boldsymbol{x}^*) \cap F_0(\boldsymbol{x}^*) \neq \phi$，则至少有一个向量

$$\boldsymbol{d} \in D_0(\boldsymbol{x}^*) \cap F_0(\boldsymbol{x}^*)$$

由定义1.1，必存在$\delta_1 > 0$，当$\alpha_1 \in (0, \delta_1)$时，成立

$$\boldsymbol{x}^* + \alpha_1 \boldsymbol{d} \in S$$

由定义1.2，必存在$\delta_2 > 0$，当$\alpha_2 \in (0, \delta_2)$时，成立

$$f(\boldsymbol{x}^* + \alpha_2 \boldsymbol{d}) < f(\boldsymbol{x}^*)$$

取$\delta = \min\{\delta_1, \delta_2\}$，当$\alpha \in (0, \delta)$时，$\boldsymbol{x}^* + \alpha \boldsymbol{d}$是可行点，并且$f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) < f(\boldsymbol{x}^*)$。这意味着$\boldsymbol{x}^*$不是(P)的局

部最优解，和假设条件矛盾。∎

将定理1.3用于(NIP)问题，即得到不等式约束极值问题的几何最优性条件，但从解析角度来看，还是很不确切的。因此需要寻求易于解析表达的几何最优性条件.

定义 1.3

在(NIP)问题中，若可行点$\bar{x}$使某个不等式约束$e_i(x)\geqslant 0$变为等式，即$e_i(\bar{x})=0$，则称约束$e_i(x)\geqslant 0$是关于可行点$\bar{x}$的有效约束. 反之，若有$e_i(\bar{x})>0$，则称$e_i(x)\geqslant 0$是关于可行点$\bar{x}$的非有效约束。可行点$\bar{x}$的一切有效约束的全体称为关于可行点$\bar{x}$的有效约束集，记为$I\triangleq\{i\mid e_i(\bar{x})=0\}$。

所谓在某点处约束是有效的，实际上就是在该点的邻域内约束起到了限制其可行性范围的作用；而非有效约束对该点在其邻域内的可行性就没有影响. 乍看起来，研究约束最优化问题时只要考虑局部最优解处的有效约束就行了；可惜这是不可能的，因为实际上人们往往无法事先判定哪些约束对于问题的局部最优解是有效约束。尽管如此，概念本身在非线性规划问题中依然居有十分重要的地位。

定理 1.4

在(NIP)问题中，设$\bar{x}$是可行点；$I\triangleq\{i\mid e_i(\bar{x})=0\}$，当$i\in I$时$e_i(x)$在$\bar{x}$处可微，当$i\bar{\in}I$时$e_i(x)$在$\bar{x}$处连续；若对于$n$维向量$d$成立$\nabla e_i(\bar{x})^{\mathrm{T}}d>0$，$i\in I$；则$d$为$\bar{x}$处处的可行方向.

【证明】 当$i\in I$时，按照可微性假设，应该有

$$e_i(\bar{x}+\alpha d)=e_i(\bar{x})+\alpha\nabla e_i(\bar{x})^{\mathrm{T}}d+\alpha\varepsilon(\alpha) \qquad (1.20)$$