研究生教材

不可微最优化

夏尊铨 审校
金洪臻 等译

大连理工大学出版社

不可微最优化

[苏] В・Ф・ДЕМЬЯНОВ
Л・В・ВАСИЛЬЕВ 著

金洪臻 冯恩民 施光燕 译
夏尊铨 审校

大连理工大学出版社

内容提要

本书主要论述不可微函数最优化的基本理论。阐述了点到集映射的理论，研究了某些不可微函数类的性质，并建立了最优性条件。详细研究了凸不可微函数极小化的数值方法。

全书共分 4 章。第 1 章介绍凸函数、凸集及点到集映射的理论；第 2 章论述一类新的不可微函数类以及求最速下降和上升方向的算法；第 3、4 章专门论述求解不可微优化问题的数值方法、凸函数和极大值函数极小化的方法。

本书可供应用数学、计算数学、管理、系统分析与决策、工程控制、工程优化及经济科学领域的研究工作者、研究生、大学生、教师等自学和参考。

本书根据苏联莫斯科物理数学文献出版社 1981 年版：В・Ф・ДЕМЬЯНОВ，Л・В・ВАСИЛЬЕВ：“НЕДИФФЕРЕНЦИРУЕМАЯ ОПТИМЕЗАЦИЯ”译出。

不可微最优化

Bu Kewei Zhi Youhua

大连理工大学出版社出版发行　（邮政编码：116024）

（出版社登记证[辽]第 16 号）　大连理工大学印刷厂印刷

开本：787×1092　1/32　印张：$13\frac{3}{8}$　字数：288 千字

1991 年 2 月　第 1 版　1991 年 2 月第 1 次印刷

印数：0001－1420 册

责任编辑：张亚军　徐润炎　责任校对：正　嫣

封面设计：姜严军

ISBN　7－5611－0363－8/O・59　定价：2.79 元

译　　序

由于科学和经济的迅速发展，在很多应用学科、实际问题及数学自身都包含着求非光滑函数的极值问题。因而，不可微最优化这一领域正在为越来越多的数学工作者以及有关领域中的专家所注目。特别，引起了一些非数学工作者的极大兴趣。

与上述情况适成对照的是，目前我国出版的不可微最优化教材还很少，更缺乏系统介绍不可微最优化基础理论和算法的著作或译作。我们翻译本书希望能为在这方面有兴趣的读者提供一本入门书。

译者非常感谢 B・Ф・ДЕМЬЯНОВ/Vladimir F・Demyanov 教授（列宁格勒大学应用数学系）和 Л. В. ВАСИЛЬЕВ/Leonid V. Vasilev 教授（列宁格勒技术大学数学系）提供他们本书原本以及为翻译工作提供的许多方便，非常感谢笹川徹史 /Tetsushi Sasagawa 教授（日本东京，上智大学机械工程科）为本书翻译工作提供初始英文译稿以及英文译本*。中译本中的第三章 §11 和附录 3 就是取自英译本的。在将本书译

* Nondifferentiable Optimization, Optimization Software, Inc. Publications Division, New York, 1985.

为中文的过程中，译者自始至终参阅了笹川徹史教授的英译本。

本书译稿，第一章由金洪臻译，第二章由潘致强译，第三章由施光燕译(其中 §11 由高岩译)，第四章及附录由冯恩民译。各章译稿由金洪臻整理。全稿由夏尊铨审校。

译者衷心感谢唐焕文研究员、张亚军副教授、滕弘飞副教授和张成学副教授对本书翻译工作的关怀和支持，非常感谢徐润炎教授仔细阅读并修改了译文的原稿，感谢郭健讲师和大连理工大学应用数学系优化课题的研究生们阅读了译稿并提出了许多宝贵建议，感谢吴雷、高左雷、于洪全、胡云姣、陈国庆、单锋和任正斌协助译稿的抄写工作。译者十分感谢大连理工大学出版社为该译本的出版所作的努力。

译稿曾给大连理工大学应用数学系优化课题的研究生试用过两次。虽然我们已经改正了译稿中的一些错误，但仍会有错漏之处未被发现。敬请读者不吝指正。

译　者

1989 年 2 月于大连理工大学

原　　序

“不可微最优化“(НД0/NDO) 是最近才出现在科技领域中的一个术语。它涉及到许多与求不可微函数的极值有关的问题。无论是在解决实际问题的过程中还是数学本身都自然地要遇到求非光滑函数的极小值和极大值等课题。切比雪夫逼近问题就是最好的例证。不失一般性，今后我们以极小化问题为讨论的对象。在非光滑的极小化问题中，极小极大问题和凸问题是被研究得最多(文献[31]，[36]，[57]，[110]，[120])。

特别是近些年来，不可微最优化这一领域引起了人们的极大兴趣(见[30]，[81]，[127])，并出版和发表了很多专集和论文(如[14]，[20]，[87]－[89]，[98]，[130]，[135]，[140]－[142]，[152]，[153]，[160]等都是专门论述非光滑最优化方法和各种理论的)。

为了求解极小化问题就需要：

1. 研究极小化函数的性质，特别是它的可微性及方向导数；

2. 建立全局或局部极小化的必要条件(如可能，就给出充

分条件)；

3. 寻找下降方向(寻求出最速下降方向，或在适当意义下的可行方向)；

4. 构造逐步逼近的算法。

本书中所研究的是有限个变量的非光滑函数的极小化问题。值得指出的是极值的必要条件是一个基本而重要的问题(文献[24]，[45]，[57]，[73]，[74]，[103]，[159]，[163]，[167]，[168]就是专门研究这类问题的)。

对于光滑函数，梯度概念的含义是众所熟知的。但对非光滑函数，已经不存在梯度。对极大值函数和凸函数来讲，相似于梯度的是次微分：与每一点 x_0 相联系着的是某个紧凸集 $\partial f(x_0)$，称其为函数 $f(x)$ 在点 x_0 的次微分。任一元素 $v\in\partial f(x_0)$ 称为函数 $f(x)$ 在点 x_0 处的次梯度。

利用次微分能够：

1. 求函数 $f(x)$ 在 x_0 的方向导数，即

$$\frac{\partial f(x_0)}{\partial g}\equiv\lim_{\alpha\to+0}\alpha^{-1}[f(x_0+\alpha g)-f(x_0)]=\max_{v\in\partial f(x_0)}(v,g)$$

2. 检验极小值的必要条件：点 x^* 是函数 $f(x)$ 在 E_n 上的极小点的必要条件为

$$0\in\partial f(x^*)$$

3. 求出最速下降方向，如果 $O\notin f(x_0)$，则方向

$$g(x_0)=-v(x_0)\,\|v(x_0)\|^{-1}$$

是函数 $f(x)$ 在点 x_0 的最速下降方向，其中 $v(x_0)\in\partial f(x_0)$，$\|v(x_0)\|=\min\limits_{v\in\partial f(x_0)}\|v\|$。

次微分的一个重要作用在于这一概念能在 Lipschitz 函数类上进行推广(例如 F. H. Clarke[133]，[134]，J. Warga[9]，

[168],Б. Н. Пшениный[104],Н. З. Шор[126],[127],А. Голдстейн[139],[140] 等)。

对于凸函数、极大值函数以及某些其它函数类的极小化问题,我们能够利用次微分和次梯度构造出许多逐次逼近算法([36],[51],[91][127],[130],[149]—[151],[155],[156],[170],[171])。

光滑函数 $f(x)$ 在集合 Ω

$$\Omega = \{x \in E_n | h_i(x) \leqslant 0\}, \forall i \in 1 : N\}$$

上的极小化问题,也是一个不可微优化问题,其中 $h_i(x)$ 是在 E_n 上的光滑函数。这是因为集合 Ω 可这样给出:

$$\Omega = \{x \in En | h(x) \leqslant 0\}$$

其中 $h(x) = \max\limits_{i \in 1 : N} h_i(x)$ 就是一个非光滑函数。

本书的目的是系统地阐述不可微函数的最优化理论。第一章中对凸函数、凸集以及点到集的映射理论做出了必要的叙述。对 ε-次微分和 ε-次微分映射的性质给予了极大的重视。凸函数的重要性不仅在于它包含很多的非光滑函数类,还因为凸函数的理论也是研究其它非光滑函数类的工具。

前面曾粗略描述过的次微分概念将凸函数、极大值函数与方向导数有机地联系起来。很多作者(包括上面所提到的)在推广次微分概念时,都没用方向导数。然而,从最优化的角度来看,使用方向导数是更有益的也是很自然的。

第二章中描述了一类新的不可微函数类 —— 拟可微函数类. 对于这类函数,与方向导数密切相关的拟可微概念起了主要的作用。每点都与某凸集对(拟微分)相对应。拟微分是梯度概念(对光滑函数)和次微分(对凸函数)概念的推广。利用拟微分可以简化极值必要条件的描述,也可使求最速下降

和上升方向的问题得以简化。建立了拟微分运算的基本公式，它们是古典微分运算公式的推广。拟可微函数类是个线性空间，关于所有"代数"运算是封闭的，对于逐点取极大和极小的运算也是封闭的(注意，凸函数类以及极大值函数类不是线性空间而是凸锥)。引进拟可微集概念，得到拟微分术语下拟可微函数在拟可微集上取极值的必要条件。于是，把用分析方法来研究的问题的种类扩展了很多，对于一大类拟可微函数，可以把极值必要条件的检验过程和最速下降或上升方向的求法算法化。然而，求解拟可微函数类的极值问题的数值技术还有待进一步研究。

第三、四章专门论述求解不可微最优化问题的数值解法，包括凸函数和极大值函数的极小化的主要方法和思想。逐次逼近法可分为松驰法和非松驰法。如果在每一步的函数值比在上一步的小，方法就称为松驰法。在所论述的方法中，既有松驰法也有非松驰法。我们不去讨论各种方法的优劣，因为从优化的角度来讲，它们都各有所长，需要根据自己的所求对它们加以选择。例如，次梯度下降法实现起来很简单，但收敛得很慢。许多方法的取舍取决于所要实现的目标和方法对问题的适用性。有时可用粗糙的逼近法去求解最优化问题，在一定误差范围内会很快地得到所需精度。在另一些情形下，需要高精度，在计算上多花些时间也是允许的。

所述方法大都是一阶的，因为它们用的是一次逼近(梯度、次梯度、次微分)。可以预见，随着不可微优化理论的进一步发展，高阶方法将会出现。

书中某些材料是作为习题来处理的。本书没有涉及到随机方法(参阅文献[51]，[83]，[107]，[117])和有关对策论的

问题([21],[58],[63],[64],),以及多目标最优化问题([95]),尽管它们也用到不可微最优化。

本书反映了最近几年列宁格勒州立大学(ЛГУ)的应用数学/控制系和计算数学研究所在非光滑最优化方面的研究成果。

部分成果在全苏第七和第八届最优化夏季会议(1977 年于 Жукин,1979 年于 Шушенкий)上报告过。

本书作者对 Н. Н. Моисеев 自始至终的鼓励和支持表示诚恳的感谢。

感谢 А·В·Певный 和 А·М·Рубинов 对手稿的仔细阅读,并提出了许多建设性意见,感谢 Е·Ф·Войтон,М·К·Гавурин,Ю·М·Ермольев,С·С·Кутателадз,В·Н·Малоземов,Б·Н·Пшеничн,В·М·Тихомиров 为本书的完善所做的贡献,以及 Э·В·Демьянов 在本书定稿时给予的巨大帮助。最后,作者对所有对本书提出批评和建议的读者表示衷心的感谢。

本书第一章 §13:1,2 段及第二章 §5 由 Л·Н·Полякова 所写,第一章 §15,§16 由 А·М·Рубинов 所写第一章 §14 及第四章 §4:1,2 段由 В·К·Шомесова 所写,第一章 §9 及第四章 §5,§7 由 Л·В·Васильев 所写。他还写了第一章 §4:6 段,§6,§8:5,6 段,§17,第三章 §1,§2,§3,§4:3 段,§6,第四章 §8 — 10。

本书采用了标准缩写符号。

表达式 $\inf\{f(x) \mid x \in A\}$ 也就是 $\inf\limits_{x\in A} f(x)$.

记法

$$\Omega = \{v \in E_n \mid \exists\ \alpha_0 > 0: x_0 + \alpha v \in A, \forall\ \alpha \in [0,\alpha_0]\}$$

读作:Ω 是点 $v \in E_n$ 的集合,对于每个这样的 v 存在一个 $\alpha_0 >$

0,使得对任一 $\alpha \in [0,\alpha_0]$ 有 $x_0 + \alpha v \in A$.

符号□意味着证明结束。

$p:q$ 表示从 p 到 q 的整数集合。

$\overline{\lim}$ 和 $\underline{\lim}$ 分别表示上,下极限。

集合 A 的元素数量记作 $|A|$。

记号 $\equiv$ 表示"等同于定义"。

带 * 号的节表示以后有用的内容。

В・Ф・Демьянов

Л・В・Васильев

1981 年

目　　录

译序……………………………………………………………（ Ⅰ ）
原序……………………………………………………………（ Ⅲ ）
第一章　凸分析初步和有关问题 ……………………………（　1）
　§1　凸集，凸包和分离定理 ……………………………（　1）
　§2　点到集的映射 ………………………………………（ 10）
　§3　凸锥，可行方向锥和共轭锥 ………………………（ 19）
　§4　凸函数，连续性和方向可微性 ……………………（ 29）
　§5　凸函数的次梯度和次微分 …………………………（ 44）
　§6　集合与锥的距离，极小化条件 ……………………（ 61）
　§7　ε-次微分 ……………………………………………（ 69）
　§8　ε-方向导数和 ε-次微分映射的连续性 ……（ 79）
　§9　凸函数的某些性质和不等式 ………………………（ 92）
　§10　条件 ε-次微分 ……………………………………（103）
　§11　条件方向导数及条件 ε-次微分映射的连续性
　　　………………………………………………………（113）
　§12　利用不等式表示凸集 ………………………………（124）
　§13　正则锥，圆锥映射 …………………………………（133）
　§14　上确界函数的方向可微性 …………………………（138）
　§15　凸函数的可微性 ……………………………………（146）
　§16*　共轭函数 …………………………………………（159）
　§17　某些凸函数类的 ε-次梯度的计算 ………………（172）
第二章　拟可微函数 …………………………………………（179）

§ 1　拟可微函数的定义与例子 …………………… (179)
§ 2　拟可微函数的性质及拟微分运算的基本公式
…………………………………………………… (185)
§ 3　拟可微运算的例子 ………………………… (194)
§ 4　凸-凹函数的拟可微性 ……………………… (203)
§ 5　E_n 空间上的拟可微函数取极值的必要条件
…………………………………………………… (212)
§ 6　拟可微集合 ………………………………… (217)
§ 7　拟可微函数在拟可微集合上取极值的必要条件
…………………………………………………… (227)
§ 8*　点到集合的距离函数 ……………………… (239)
§ 9　隐函数 ……………………………………… (248)
第三章　无约束极小化 ………………………… (252)
§ 1　凸函数在 E_n 上取极小值的必要和充分条件
…………………………………………………… (252)
§ 2　光滑函数的极小化 ………………………… (254)
§ 3　最速下降法 ………………………………… (257)
§ 4　凸函数极小化的次梯度法 ………………… (264)
§ 5　多步次梯度法 ……………………………… (275)
§ 6　松驰次梯度法 ……………………………… (283)
§ 7　松驰 ε-次梯度法 …………………………… (299)
§ 8　Kelley 方法 ………………………………… (308)
§ 9　上确界函数的极小化 ……………………… (317)
§ 10　凸极大值函数的极小化与极值基方法 …… (320)
§ 11　一类拟可微函数极小化的数值方法 ……… (327)
第四章　约束条件下的极小化 ………………… (334)

§ 1 凸函数在凸集上极小化的充要条件 ……… (334)
§ 2 ε—平稳点 ……………………………………… (343)
§ 3 条件次梯度法 ………………………………… (346)
§ 4 凸函数极小化的最速下降法 ……………… (351)
§ 5 具有约束的修正(ε,μ)—次梯度法 ………… (357)
§ 6 定步长次梯度法 ……………………………… (360)
§ 7 具有约束的修正(ε,μ) — 次梯度法………… (364)
§ 8 非光滑的罚函数法 …………………………… (368)
§ 9 在凸集上极小化的 Kelley 方法 …………… (375)
§ 10 具有约束的松驰次梯度法 ………………… (378)
附录 1 **文献注释**………………………………………… (382)
附录 2 **拟微分学文献**…………………………………… (387)
附录 3 **英译本注与有关文献**…………………………… (405)

第一章　凸分析初步和有关问题

§1　凸集，凸包和分离定理

1　下面我们研究 n 维 Euclid 空间 E_n 的向量 $x=(x^{(1)},\cdots,x^{(n)})$. 空间 E_n 是线性空间。引入记号：

$$\mathbb{O}_n=\mathbb{O}=(0,\cdots,0)\in E_n, x_k=(x_k^{(1)},\cdots,x_k^{(n)}),$$

$$(x_1,x_2)=\sum_{i=1}^{n}x_1^{(i)}x_2^{(i)},\ \|x\|=\sqrt{(x,x)}$$

$$x^2=(x,x)$$

有 Cauchy — Буняковский 不等式

$$|(x_1,x_2)|\leq\|x_1\|\cdot\|x_2\|,$$

对于所有向量 $x_1,x_2\in E_n$ 都成立。如果等式 $\sum_{k=1}^{r}\alpha_kx_k=\mathbb{O}$ 当且仅当所有系数 α_k 均为零时才成立，则向量 $x_1,\cdots,x_r$ 称为线性无关。如果 $r\geqslant n+1$，则向量 $x_1,\cdots,x_r$ 线性相关，也就是，可求得数 $\beta_1,\cdots,\beta_r$，使得 $\sum_{k=1}^{r}\beta_k^2>0$（即它们不能同时为零）并且

$$\sum_{k=1}^{r}\beta_kx_k=\mathbb{O}\tag{1.1}$$

如果 $r\geqslant n+2$，则除(1.1)式外还须满足等式

$$\sum_{k=1}^{r}\beta_k=0\tag{1.2}$$

以下证明这一点。引入向量

$\bar{x}_k = (1, x_k^{(1)}, \cdots, x_k^{(n)}) \in E_{n+1}, k \in 1 : r, r \geqslant n+2$. 在 E_{n+1} 中，任意 $n+2$ 个向量总是线性相关的，即，存在数 β_k，$\sum_{k=1}^{r} \beta_k^2 > 0$，使

$$\sum_{k=1}^{r} \beta_k \bar{x}_k = \mathbb{O}_{n+1}. \tag{1.3}$$

由(1.3)可得

$$\sum_{k=1}^{r} \beta_k \bar{x}_k = \mathbb{O}_n, \sum_{k=1}^{r} \beta_k = 0$$

（把第一个坐标撤出，单独等于零，而其余的 n 个分量保留在原式中）。

不含任何元素的集合称为空集并记以 $\varnothing$.

设 $S_\delta(x_0) = \{x \in E_n \mid \| x - x_0 \| \leqslant \delta\} \quad \delta > 0$.
集合 $S_\delta(x_0)$ 称为点 x_0 的 δ-邻域。如果存在这样的 $\delta > 0$，使得 $S_\delta(x_0) \subset G$，则称点 x_0 是集合 G 的内点。由集合 G 的所有内点构成的集合记以 $\mathrm{int}G$（该集合可为空集）。

如果对于 G 中的任意点 x_0，存在 $\delta > 0$，使得 $S_\delta(x_0) \subset G$，则称集合 $G \subset E_n$ 为开集。显然在这种情形下，$G = \mathrm{int}G$.

由点 $x = \lim\limits_{k\to\infty} x_k$（其中 $x_k \in G, \forall\, k \in 1 : \infty$）构成的集合称为集合 $G \subset E_n$ 的闭包，G 的闭包记为 $\bar{G}$。

集合 $G \subset En$ 称为闭集，如果由关系式 $x_k \xrightarrow[k\to\infty]{} x_0, x_k \in G$ $\forall\, k \in 1 : \infty$，可得 $x_0 \in G$. 显然，对闭集来说，$G = \bar{G}$.

设 $G \subset E_n$，对任意 $\delta > 0$，如果 x_0 的任一 δ-邻域 $S_\delta(x_0)$ 既包含 G 中的点又包含不属于 G 的点，则称 x_0 为集合 G 的边界点（x_0 可以是，也可以不是 G 中的点）。集合 G 的边界点构成的集

合记以 G_{fr}。

如果存在这样的正数 $K<+\infty$，使得 $\|x\|\leqslant K$，$\forall x\in G$，则称集合 G 是有界集合。

如果对任意 $K>0$ 都能找到 $x\in G$，使得 $\|x\|>K$，则集合 G 是无界集。

显然，两个有界集的并、交、和及差集仍然是有界集。

设有两个集合，若其中之一是有界的，则它们的交集是有界集。

设 A 和 B 是闭集，则它们的并和交集仍然是闭集。但对和、差、以及代数差，该结论不成立。

例 1 设

$$A=\{x=(x^{(1)},x^{(2)})\in E_2|x^{(2)}\geqslant\frac{1}{x^{(1)}},x^{(1)}>0\}$$

$$B=\{x=(x^{(1)},x^{(2)})\in E_2|x^{(1)}=0,x^{(2)}\leqslant 0\}$$

显然，A 和 B 是闭集，但不是有界集。集合

$$C=A+B=\{x=(x^{(1)},x^{(2)})\in E_2$$

$$|x^{(1)}>0,x^{(2)}\in(-\infty,\infty)\}$$ 不是闭集，因为

$$\overline{C}=\{x=(x^{(1)},x^{(2)})\in E_2|x^{(1)}\geqslant 0,x^{(2)}\in(-\infty,\infty)\}\neq C.$$

然而，如果集合 A 和 B 是闭的并且即使只有一个是有界的，则和（及代数差）也是闭的。

如果一个集合有如下的性质：集合中元素所构成的每个序列都含有一收敛的子序列，且该子序列的极限又属于这个集合，则这集合称为紧的或紧集。显然，E_n 中的集合是紧的当且仅当它是闭的且为有界的。

定义 1 集合 $\Omega\subset E_n$ 称为凸集，如果连接 Ω 中的任意两点 x_1,x_2 所构成的线段都在该集内，即$[x_1,x_2]\subset\Omega$，其中