运筹与管理科学丛书 1

非线性优化计算方法

袁亚湘 著

科学出版社
www.sciencep.com

中国科学院科学出版基金资助出版

运筹与管理科学丛书 1

非线性优化计算方法

袁亚湘 著

科学出版社

北京

内 容 简 介

本书系统和深入介绍非线性优化的主要计算方法和相关理论，主要内容包括：一维优化方法、梯度法和共轭梯度法、拟牛顿法、直接方法、二次规划、罚函数法、可行方向法、逐步二次规划法、信赖域法、内点法、滤子方法等.

本书可作为相关专业高年级大学生和研究生的教材，同时也可作为广大非线性优化研究人员以及从事实际应用的工程技术人员的参考书.

图书在版编目(CIP)数据

非线性优化计算方法/袁亚湘著. —北京：科学出版社，2008.2

(运筹与管理科学丛书；1)

ISBN 978-7-03-020883-5

I.非… II.袁… III.非线性-最优化算法 IV.O224

中国版本图书馆CIP数据核字(2008)第005821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：王 浩

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2月第 一 版 开本：B5(720×1000)
2008年2月第一次印刷 印张：17
印数：1—3 000 字数：319 000

定价：45.00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

非线性优化计算方法是计算数学与运筹学的交叉学科. 非线性优化在国防、经济、金融、工程、管理等许多领域有着广泛的应用. 许多其他科学领域的问题也可归结为非线性优化问题, 如大气科学中的同化问题、生命科学中的蛋白质折叠问题、信息科学中的模式识别问题、地球科学中的反演问题等. 这些问题往往都是大规模和高度非线性的, 因而研究高效的求解非线性优化的计算方法具有重要的理论意义和实际价值.

作为一门学科, 非线性优化诞生于 20 世纪 50 年代, 奠基性的工作是 Kuhn 和 Tucker 的最优性定理. 由于它的基础性, 从 70 年代起非线性优化就是国际上数学规划中的主流和最受重视的分支之一. 特别是拟牛顿 (quasi-Newton) 法的兴起, 涌现了如 Powell, Fletcher, Dennis, Wright, Rockafellar 等一批国际著名的非线性优化专家. 80 年代对信赖域 (trust region) 法的深入研究使非线性优化得到进一步发展. 世纪之交, 滤子 (filter) 方法的提出又为非线性优化方法创新开辟了一条新路.

本书根据作者 1993 年在上海科技出版社出版的《非线性规划数值方法》修订而成. 与原书相比, 本书增加了近年来国际上关于非线性优化计算方法的前沿研究结果, 如内点法、滤子方法、弱拟牛顿法、梯度法的步长选取、截断共轭梯度法的理论分析等, 其中有些是作者取得的.

本书保持了简洁、深入和系统的特点, 全面介绍各种主要优化方法, 把重点放在方法的基本思想和分析技巧. 本书可作为大学高年级本科生和研究生的教材, 同时也可作为广大非线性优化研究人员以及从事实际应用的工程技术人员的参考书.

我想借此机会对我的博士论文指导导师、英国皇家学会会员、美国科学院外籍院士、首届 Dantzig 奖获得者、剑桥大学的 Powell 教授表示衷心的感谢, 我在科学研究方面取得的成绩和他的指导与鼓励是分不开的. 他对生活的热爱, 对科学的追求, 对名利的淡泊和对学生的关爱是我永远要学习的.

近些年陪女儿袁瑗的时间太少, 我向她表示歉意, 同时以此书作为她 18 岁生日以及考上北京大学光华管理学院的礼物. 为赶书稿常晚上加班, 在此特向夫人张焱的理解和支持表示感谢.

作者的研究工作长期得到国家自然科学基金委员会的资助, 在此表示感谢. 本书的出版得到中国科学院科学出版基金的大力支持, 作者表示诚挚的谢意.

袁亚湘

2007 年 9 月 4 日于北京

目　录

第1章 导　论

§1.1 问　题

非线性优化 (也称非线性规划) 问题是在有限维实空间上求单值函数的极值, 函数的自变量可能受限制于有限个等式或不等式约束. 非线性优化问题通常可以写成

$$\min_{x\in\Re^n} \quad f(x)\ , \tag{1.1.1}$$

$$\text{s. t.} \quad c_i(x)=0, \quad i=1,\cdots,m_e; \tag{1.1.2}$$

$$c_i(x)\geqslant 0, \quad i=m_e+1,\cdots,m, \tag{1.1.3}$$

$f(x)$ 和 $c_i(x)(i=1,\cdots,m)$ 都是定义在 $\Re^n$ 上的实函数, 其中至少有一个不是线性的. n 是正整数, m 和 m_e 是两个非负整数且满足 $m\geqslant m_e$. s. t. 是英文 subject to (满足于) 的缩写. $f(x)$ 称为目标函数, $c_i(x)(i=1,\cdots,m)$ 称为约束函数. (1.1.2) 和 (1.1.3) 称为约束条件. 满足约束条件的点称为可行点. 所有可行点的集合称为可行域, 记为 X. 显然,

$$X=\{x\mid c_i(x)=0, i=1,\cdots,m_e;\ c_i(x)\geqslant 0, i=m_e+1,\cdots,m\}. \tag{1.1.4}$$

利用这一定义, 我们可把问题 (1.1.1)~(1.1.3) 写成如下简洁的形式:

$$\min_{x\in X} f(x). \tag{1.1.5}$$

当 $m=0$, 问题 (1.1.1)~(1.1.3) 是无约束优化问题, 否则, 是约束优化问题. 在约束优化中, 如果只有等式约束, 即 $m=m_e>0$, 则称为等式约束优化问题. 另一种特殊情形是所有的约束函数都是线性函数, 这时称问题 (1.1.1)~(1.1.3) 是一个线性约束优化问题.

当 $f(x)$ 和 $c_i(x)(i=1,\cdots,m)$ 都是线性时, 问题 (1.1.1)~(1.1.3) 是线性规划. 关于线性规划的理论和方法, 读者可参阅文献 (Dantzig, 1998).

非线性规划这一名词是由 Kuhn 和 Tucker(1951) 最先使用, 但非线性规划问题早已有之. 牛顿在著名的《自然哲学的数学原理》中就研究了极值问题. Sylvester (1857) 提出的在平面上找一个能包含 N 个给定点之最小圆的问题实质上是特殊的优化问题, 它可写成如下形式:

$$\min_{x\in\Re^2} f(x)=\max_{1\leqslant i\leqslant N}\|x-x_i\|_2, \tag{1.1.6}$$

其中, $\|\cdot\|$ 是欧氏范数, $x_i(i=1,\cdots,N)$ 是平面上 N 个给定点. 显然, (1.1.6) 的解 x^* 即是需要寻找的圆之圆心, $f(x^*)$ 就是该圆之半径.

非线性曲线拟合是另一个特殊的优化问题, 即非线性最小二乘问题, 它可以写成

$$\min_{x\in\Re^n} f(x)=\sum_{i=1}^{N}[\phi(x,t_i)-y_i]^2, \tag{1.1.7}$$

其中 $t_i\in\Re^m$, $y_i\in\Re(i=1,\cdots,N)$ 是实验值, x 是待定参数, $\phi(x,t)$ 是一非线性函数, 它是取定的拟合曲线的基函数. 设 x^* 是 (1.1.7) 的解, 则

$$y(t)=\phi(x^*,t) \tag{1.1.8}$$

就是利用最小二乘对数据 $(t_i,y_i)(i=1,\cdots,N)$ 拟合所得到的曲线.

非线性规划问题广泛见于国防、经济、工程、管理等许多重要领域, 在结构设计、化学反应设计、电力分配、石油开采、交通调度、物流等方面都有直接的应用. 不少其他科学领域的问题也可归结于优化问题, 如大气科学中的同化问题、自动化中的模式识别问题、计算机科学中的大规模电路设计问题等. 非线性规划计算方法还和计算数学中的数值逼近、微分方程数值解、变分法、反问题计算方法、非线性代数方程数值方法等分支有交叉和应用.

下面我们给出问题 (1.1.1)~(1.1.3) 解的定义.

定义 1.1.1 设 $x^*\in X$, 如果存在 $\delta>0$, 使得

$$f(x)\geqslant f(x^*),\qquad \forall x\in X\cap B(x^*,\delta)\backslash\{x^*\}, \tag{1.1.9}$$

则称 x^* 是问题 (1.1.1)~(1.1.3) 的局部极小点, 其中 $B(x^*,\delta)$ 是以 x^* 为中心半径为 δ 的球:

$$B(x^*,\delta)=\{x|\|x-x^*\|_2\leqslant\delta\}. \tag{1.1.10}$$

如果 (1.1.9) 对严格不等号 “$>$” 成立, 则称 x^* 是局部严格极小点.

定义 1.1.2 设 $x^*\in X$, 如果

$$f(x)\geqslant f(x^*),\qquad \forall x\in X\backslash\{x^*\}, \tag{1.1.11}$$

则称 x^* 是问题 (1.1.1)~(1.1.3) 的全局极小点. 如果 (1.1.11) 对严格不等号 “$>$” 成立, 则称 x^* 是全局严格极小点.

全局极小值点也称作总体极小值点. 从定义可知, 全局 (严格) 极小值点必定是局部 (严格) 极小点. 问题 (1.1.1)~(1.1.3) 的全局 (或局部) 解就是指它的全局 (或局部) 极小点. 目标函数在全局 (或局部) 极小点的值称为问题 (1.1.1)~(1.1.3) 的全局 (或局部) 极小值.

定义 1.1.3 对任何 $x \in \Re^n$, 称集合

$$\mathcal{A}(x) = E \cup I(x) \tag{1.1.12}$$

是在 x 点处的积极集合, 其中

$$E = \{1, 2, \cdots, m_e\}, \tag{1.1.13}$$

$$I = \{m_e + 1, \cdots, m\}, \tag{1.1.14}$$

$$I(x) = \{i | c_i(x) \leqslant 0, i \in I\}. \tag{1.1.15}$$

称 $c_i(x)(i \in \mathcal{A}(x))$ 是在 x 点的积极约束, 称 $c_i(x)(i \notin \mathcal{A}(x))$ 是在 x 点的非积极约束.

§1.2 最优性条件

在一般情况下, 对于给定的 $x^* \in X$, 集合 $X \cap B(x^*, \delta) \backslash \{x^*\}$ 中有无穷多个点. 所以通过验证 (1.1.9) 来判别 x^* 是否是优化问题 (1.1.1)~(1.1.3) 的解几乎是不可能的. 最优性条件就是在一定条件下和 (1.1.9) 等价的条件, 这些条件通常仅依赖于目标函数和约束函数在 x^* 点的性质.

对于无约束优化问题, 显然有以下定理:

定理 1.2.1 假设 x^* 是无约束优化问题 (1.1.1) 的局部极小点, $f(x)$ 在 x^* 处可微, 则必有

$$\nabla f(x^*) = 0, \tag{1.2.1}$$

如果 $f(x)$ 在 x^* 处二次可微, 则不等式

$$d^{\mathrm{T}} \nabla^2 f(x^*) d \geqslant 0 \tag{1.2.2}$$

对任何 $d \in \Re^n$ 都成立.

此定理是个必要性定理, 它给出在解处函数必须满足的条件. 满足 (1.2.1) 的点称为函数 $f(x)$ 的稳定点. 从上述可知, 只要目标函数可微, 局部极小点必定是稳定点. 下面的定理给出稳定点是局部极小点的充分条件.

定理 1.2.2 假设 $f(x)$ 在 x^* 处二次连续可微, 如果 x^* 是 $f(x)$ 的稳定点且对任何非零向量 $d \in \Re^n$ 都有

$$d^{\mathrm{T}} \nabla^2 f(x^*) d > 0, \tag{1.2.3}$$

则 x^* 是无约束优化问题 (1.1.1) 的局部严格极小点.

以上两个定理的证明非常简单, 只需利用 Taylor 展开即可, 故从略.

现在我们考虑约束优化问题. 显然, 可行域上的点是否为局部极小点取决于目标函数在该点以及在该点附近其他可行点上的值. 所以我们需要给出可行方向的定义.

定义 1.2.3　设 $x^* \in X$, $d \in \Re^n$, 如果存在 $\delta > 0$, 使得

$$x^* + td \in X, \qquad \forall t \in [0, \delta], \tag{1.2.4}$$

则称 d 是 X 在 x^* 处的可行方向. X 在 x^* 处的所有可行方向的集合记为 $\mathrm{FD}(x^*, X)$.

定义 1.2.4　设 $x^* \in X$, $d \in \Re^n$, 如果

$$d^{\mathrm{T}} \nabla c_i(x^*) = 0, i \in E, \quad d^{\mathrm{T}} \nabla c_i(x^*) \geqslant 0, i \in I(x^*), \tag{1.2.5}$$

则称 d 是 X 在 x^* 处的线性化可行方向. X 在 x^* 处的所有线性化可行方向的集合记为 $\mathrm{LFD}(x^*, X)$.

定义 1.2.5　设 $x^* \in X$, $d \in \Re^n$, 如果存在 $d_k \in \Re^n$, $\delta_k > 0$ $(k = 1, 2, \cdots)$, 使得 $d_k \to d$, $\delta_k \to 0$ 且 $x^* + \delta_k d_k \in X$, 则称 d 是 X 在 x^* 处的序列可行方向. X 在 x^* 处的所有序列可行方向的集合记为 $\mathrm{SFD}(x^*, X)$.

根据定义, 下列引理显然成立:

引理 1.2.6　如果所有的约束函数都在 x^* 处可微, 则有

$$\mathrm{FD}(x^*, X) \subseteq \mathrm{SFD}(x^*, X) \subseteq \mathrm{LFD}(x^*, X). \tag{1.2.6}$$

引理 1.2.7　设 x^* 是问题 (1.1.1)~(1.1.3) 的局部极小点, 如果 $f(x)$ 在 x^* 处可微, 则必有

$$d^{\mathrm{T}} \nabla f(x^*) \geqslant 0, \quad \forall d \in \mathrm{SFD}(x^*, X). \tag{1.2.7}$$

证明　(1.2.7) 对 $d = 0$ 显然成立. 对任意非零向量 $d \in \mathrm{SFD}(x^*, X)$, 由定义知存在 $\delta_k \to 0$, $d_k \to d$, 使得 $x^* + \delta_k d_k \in X$. 因为 x^* 是局部极小点, 所以对充分大的 k, 有

$$f(x^* + \delta_k d_k) \geqslant f(x^*). \tag{1.2.8}$$

由于 $f(x)$ 在 x^* 处可微, 从上式可知

$$\delta_k d_k^{\mathrm{T}} \nabla f(x^*) + o(\delta_k) \geqslant 0. \tag{1.2.9}$$

因为 $\delta_k > 0$ 且 $\delta_k \to 0$, 在上式的两端除以 δ_k 后, 令 $k \to \infty$, 即可得 (1.2.7).　□

类似, 我们可以证明

引理 1.2.8　设 $x^* \in X$, 如果 $f(x)$ 在 x^* 处可微且

$$d^{\mathrm{T}} \nabla f(x^*) > 0, \quad \forall d \in \mathrm{SFD}(x^*, X), \tag{1.2.10}$$

则 x^* 是问题 (1.1.1)~(1.1.3) 的局部严格极小点.

证明 如果 x^* 不是局部严格极小点, 则存在无穷点列 $x_k(k=1,2,\cdots)$, $x_k \in X$, $x_k \neq x^*$, $x_k \to x^*$, 使得

$$f(x_k) \leqslant f(x^*), \qquad \forall\, k=1,2,\cdots. \tag{1.2.11}$$

对所有 k, 定义 $\delta_k = \|x_k - x^*\|_2$, $d_k = (x_k - x^*)/\delta_k$. 由于 d_k 有界, 必存在收敛子列. 不失一般性 (选取子列代替整个序列), 假定 d_k 收敛于 d^*. 显然 $d^* \in \mathrm{SFD}(x^*, X)$. 由 (1.2.11) 和 $f(x)$ 的可微性知 $(d^*)^{\mathrm{T}}\nabla f(x^*) \leqslant 0$. 这说明引理成立. □

下面的引理是由 Farkas(1902) 给出的, 故称为 Farkas 引理. 由于它在形式上是线性方程组及线性不等式组和线性表达式这两者中必有一个且只有一个成立, 所以该引理也称为择一性引理.

引理 1.2.9 (Farkas 引理) 设 l 和 l' 是两个非负整数, $a_0, a_i(i=1,\cdots,l)$ 和 $b_i(i=1,\cdots,l')$ 是 $\Re^n$ 中的向量, 则线性方程组及不等式组

$$d^{\mathrm{T}}a_0 < 0, \tag{1.2.12}$$

$$d^{\mathrm{T}}a_i = 0, \quad i=1,\cdots,l, \tag{1.2.13}$$

$$d^{\mathrm{T}}b_i \geqslant 0, \quad i=1,\cdots,l' \tag{1.2.14}$$

无解当且仅当存在实数 $\lambda_i(i=1,\cdots,l)$ 和非负实数 $\mu_i(i=1,\cdots,l')$, 使得

$$a_0 = \sum_{i=1}^{l}\lambda_i a_i + \sum_{i=1}^{l'}\mu_i b_i. \tag{1.2.15}$$

证明 假定 (1.2.15) 成立, 且 $\mu_i \geqslant 0(i=1,\cdots,l')$. 如果 (1.2.13) 和 (1.2.14) 成立, 则

$$d^{\mathrm{T}}a_0 = \sum_{i=1}^{l}\lambda_i d^{\mathrm{T}}a_i + \sum_{i=1}^{l'}\mu_i d^{\mathrm{T}}b_i \geqslant 0. \tag{1.2.16}$$

故知 (1.2.12) 不可能成立, 从而 (1.2.12)~(1.2.14) 无解.

现假定不存在实数 $\lambda_i(i=1,\cdots,l)$ 和非负实数 $\mu_i(i=1,\cdots,l')$, 使得 (1.2.15) 成立. 定义集合

$$S = \left\{a \middle| a = \sum_{i=1}^{l}\lambda_i a_i + \sum_{i=1}^{l'}\mu_i b_i, \lambda_i \in \Re, \mu_i \geqslant 0\right\}. \tag{1.2.17}$$

显然 S 是 $\Re^n$ 中的一个闭凸锥. 由于 $a_0 \notin S$, 根据泛函分析中的凸集分离定理, 必存在 $d \in \Re^n$ 和 $\alpha \in \Re$, 使得

$$d^{\mathrm{T}}a_0 < \alpha < d^{\mathrm{T}}a, \qquad \forall a \in S. \tag{1.2.18}$$

由于 $0 \in S$, 所以

$$d^{\mathrm{T}} a_0 < 0. \tag{1.2.19}$$

对任意 $\lambda > 0$ 和任意 $i = 1, \cdots, l'$, 都有 $\lambda b_i \in S$, 于是

$$\lambda d^{\mathrm{T}} b_i > \alpha. \tag{1.2.20}$$

在上面不等式两边同除以 λ, 然后令 $\lambda \to +\infty$, 即得

$$d^{\mathrm{T}} b_i \geqslant 0, \qquad i = 1, \cdots, l'. \tag{1.2.21}$$

对任意 $\lambda > 0$, 都有 $\lambda a_i \in S$ 和 $-\lambda a_i \in S$, 同上可证 $d^{\mathrm{T}} a_i \geqslant 0$ 和 $d^{\mathrm{T}}(-a_i) \geqslant 0$. 故知

$$d^{\mathrm{T}} a_i = 0, \qquad i = 1, \cdots, l. \tag{1.2.22}$$

所以, (1.2.12)~(1.2.14) 有解. □

利用 Farkas 引理和引理 1.2.7 可证明著名的 Kuhn-Tucker 定理:

定理 1.2.10 (Kuhn-Tucker 定理)　*设 x^* 是问题 (1.1.1)~(1.1.3) 的局部极小点, 如果 $f(x)$, $c_i(x)(i = 1, \cdots, m)$ 在 x^* 处可微而且*

$$\mathrm{SFD}(x^*, X) = \mathrm{LFD}(x^*, X), \tag{1.2.23}$$

则存在 $\lambda_i^*(i = 1, \cdots m)$, 使得

$$\nabla f(x^*) = \sum_{i=1}^{n} \lambda_i^* \nabla c_i(x^*), \tag{1.2.24}$$

$$\lambda_i c_i(x^*) = 0, \quad \lambda_i \geqslant 0, \qquad i = 1, \cdots, m. \tag{1.2.25}$$

证明　由引理 1.2.7 和式 (1.2.23) 即知线性方程及不等式组

$$d^{\mathrm{T}} \nabla f(x^*) < 0, \tag{1.2.26}$$

$$d^{\mathrm{T}} \nabla c_i(x^*) = 0, \quad i \in E, \tag{1.2.27}$$

$$d^{\mathrm{T}} \nabla c_i(x^*) \geqslant 0, \quad i \in I(x^*) \tag{1.2.28}$$

无解. 从引理 1.2.9 可知, 存在 $\lambda_i^*(i \in E)$ 及 $\lambda_i^* \geqslant 0(i \in I(x^*))$, 使得

$$\nabla f(x^*) = \sum_{i \in E} \lambda_i^* \nabla c_i(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla c_i(x^*). \tag{1.2.29}$$

定义 $\lambda_i^* = 0(i \in I \backslash I(x^*))$, 即知 (1.2.24)~(1.2.25) 成立. □

与 (1.2.24) 式有密切联系的一个函数是

$$L(x,\lambda)=f(x)-\lambda^{\mathrm{T}}c(x)=f(x)-\sum_{i=1}^{m}\lambda_i c_i(x), \tag{1.2.30}$$

其中 $\lambda=(\lambda_1,\cdots,\lambda_m)^{\mathrm{T}}\in\Re^m$, $c(x)=(c_1(x),\cdots,c_m(x))^{\mathrm{T}}$. 由于这一函数的思想可追溯到 Lagrange(1760~1761), 故它被称为 Lagrange 函数, 其中 $\lambda_i(i=1,\cdots,m)$ 被称为 Lagrange 乘子. 定理 1.2.10 告诉我们, 约束优化的局部解通常是 Lagrange 函数的稳定点. 由于定理 1.2.10 最早是由 Kuhn 和 Tucker(1951) 给出的, 我们给出如下定义:

定义 1.2.11 *如果 $x^*\in X$ 且存在 $\lambda^*=(\lambda_1^*,\cdots,\lambda_m^*)^{\mathrm{T}}\in\Re^m$ 满足 (1.2.24)~(1.2.25), 则称 x^* 是问题 (1.1.1)~(1.1.3) 的 Kuhn-Tucker 点 (简称 K-T 点).*

因为 Karush(1939) 也类似地考虑了约束优化的最优性条件, 所以也有人将定理 1.2.10 称为 Karush-Kuhn-Tucker 定理, 将 K-T 点称为 K-K-T 点.

定理 1.2.10 中的条件 (1.2.23) 称为约束规范条件. 显然, 约束规范条件 (1.2.23) 在所有约束函数都是线性时自动满足. 对于非线性约束, Mangasarian 和 Fromowitz (1967) 提出了另一个约束规范条件:

$$\nabla c_i(x^*)(i\in E)\ \text{线性无关}, \tag{1.2.31}$$

$$S^*=\{d|d^{\mathrm{T}}\nabla c_i(x^*)=0,i\in E,d^{\mathrm{T}}\nabla c_i(x^*)>0,i\in I(x^*)\}\ \text{非空}, \tag{1.2.32}$$

并且给出了下面的定理:

定理 1.2.12 *在定理 1.2.10 中, 将条件 (1.2.23) 换成 (1.2.31)~(1.2.32), 必要性条件 (1.2.24) 和 (1.2.25) 依然成立.*

证明 显然, 只需证明条件 (1.2.31) 和 (1.2.32) 比 (1.2.23) 更强. 对任意非零 $d\in S^*$, 必存在 $d_i(i=1,\cdots,n-m_e-1)$ 组成 $\mathrm{span}\{\nabla c_1(x^*),\cdots,\nabla c_m(x^*),d\}$ 的法空间的一组正交基. 考虑带参数的非线性代数方程组:

$$c_i(x)=0,\qquad i=1,\cdots,m_e; \tag{1.2.33}$$

$$d_i^{\mathrm{T}}(x-x^*)=0,\qquad i=1,\cdots,n-m_e-1; \tag{1.2.34}$$

$$d^{\mathrm{T}}(x-x^*)-\theta=0. \tag{1.2.35}$$

由于该方程组在 $x=x^*$ 处的 Jacobi 矩阵是非奇异的, 根据隐函数定理, 对充分小的 θ, 必存在解 $x=x(\theta)$, 且满足

$$x'(\theta)|_{\theta=0}=d. \tag{1.2.36}$$

对所有 $i \in I(x^*)$ 有 $d^{\mathrm{T}}\nabla c_i(x^*) > 0$, 故当 $\theta > 0$ 充分小时, 必有

$$c_i(x(\theta)) > 0. \tag{1.2.37}$$

所以 $x(\theta) \in X$. 根据定义 1.2.5 和关系式 (1.2.36) 即知

$$d \in \mathrm{SFD}(x^*, X). \tag{1.2.38}$$

由于 $d \in S^*$ 的任意性, 有

$$S^* \subseteq \mathrm{SFD}(x^*, X). \tag{1.2.39}$$

由定义 1.2.5 还可看出 $\mathrm{SFD}(x^*, X)$ 是闭集. 于是

$$\mathrm{cl}(S^*) \subseteq \mathrm{cl}(\mathrm{SFD}(x^*, X)) = \mathrm{SFD}(x^*, X), \tag{1.2.40}$$

其中 $\mathrm{cl}(S)$ 表示集合 S 的闭包. 由于 S^* 非空, 根据 Rockafellar(1970) 中的定理 6.5 可知

$$\begin{aligned}\mathrm{cl}(S^*) &= \{d \mid d^{\mathrm{T}}\nabla c_i(x^*) = 0, i \in E, d^{\mathrm{T}}\nabla c_i(x^*) \geqslant 0, i \in I(x^*)\} \\ &= \mathrm{LFD}(x^*, X).\end{aligned} \tag{1.2.41}$$

从 (1.2.6) 和上两个式子即知 (1.2.23) 成立. □

一个比 (1.2.31) 和 (1.2.32) 更强的约束规范条件是

$$\nabla c_i(x^*)(i \in E \cup I(x^*)) \text{ 线性无关}. \tag{1.2.42}$$

显然, 当 $I(x^*) = \varnothing$ 时, (1.2.42) 与 (1.2.31) 和 (1.2.32) 等价. 当 $I(x^*) \neq \varnothing$ 时, 只要 (1.2.42) 成立就可证 (1.2.31) 和 (1.2.32) 也成立. 于是我们有

定理 1.2.13 *在定理 1.2.10 中, 将条件 (1.2.23) 换成 (1.2.42), 必要性条件 (1.2.24) 和 (1.2.25) 依然成立.*

因为约束规范条件 (1.2.42) 易于检验, 所以定理 1.2.13 是一个最常见也是最实用的关于约束优化一阶最优性条件的结果.

下面我们讨论二阶最优性条件. 二阶条件与一阶条件有个重要的差别是一阶条件是针对所有的可行方向; 而二阶条件通常只是对可行方向的一个子集. 这是由于在那些使得目标函数一阶方向导数非零的方向上, 目标函数的二阶导数可以自由改变而不影响优化问题的局部最优性. 第二个差别是分析一阶条件是在考察目标函数在可行方向的变化; 而分析二阶条件是利用 Lagrange 函数在可行方向上的变化.

首先我们需要如下定义

定义 1.2.14 设 x^* 是问题 (1.1.1)~(1.1.3) 的 K-T 点, λ^* 是相应的 Lagrange 乘子. 如果存在序列 $d_k(1,2,\cdots)$ 和 $\delta_k(k=1,2,\cdots)$, 使得

$$x^* + \delta_k d_k \in X, \tag{1.2.43}$$

$$\sum_{i=1}^{m} \lambda_i^* c_i(x^* + \delta_k d_k) = 0, \tag{1.2.44}$$

且有 $d_k \to d$ 和 $\delta_k \to 0$, 则称 d 是在 x^* 处的序列零约束方向. 在 x^* 处的所有序列零约束方向的集合记为 $S(x^*, \lambda^*)$.

定义 1.2.15 设 x^* 是 (1.1.1)~(1.1.3) 的 K-T 点. 如果 d 是 X 在 x^* 处的线性化可行方向而且

$$d^{\mathrm{T}} \nabla f(x^*) = 0, \tag{1.2.45}$$

则称 d 是在 x^* 处的线性化稳定可行方向. 在 x^* 处的所有线性化稳定可行方向的集合记为 $G(x^*)$.

显然可以看出,

$$S(x^*, \lambda^*) \subseteq \mathrm{SFD}(x^*, X) \cap N_{\nabla f(x^*)}, \tag{1.2.46}$$

$$G(x^*) = \mathrm{LFD}(x^*, X) \cap N_{\nabla f(x^*)}, \tag{1.2.47}$$

其中

$$N_{\nabla f(x^*)} = \{d \mid d^{\mathrm{T}} \nabla f(x^*) = 0\}. \tag{1.2.48}$$

由于 Lagrange 函数在所有满足 (1.2.44) 的可行点 $x^* + \delta_k d_k$ 上与目标函数是相等的, 与引理 1.2.7 类似, 可证如下二阶必要条件:

定理 1.2.16 设 x^* 是问题 (1.1.1)~(1.1.3) 的局部极小点, λ^* 是相应的 Lagrange 乘子. 如果 $f(x)$ 在 x^* 处二次可微, 则必有

$$d^{\mathrm{T}} W(x^*, \lambda^*) d \geqslant 0, \quad \forall d \in S(x^*, \lambda^*), \tag{1.2.49}$$

其中 $W(x^*, \lambda^*)$ 是 Lagrange 函数的二阶导数:

$$W(x^*, \lambda^*) = \nabla^2 f(x^*) - \sum_{i=1}^{m} (\lambda^*)_i \nabla^2 c_i(x^*). \tag{1.2.50}$$

从定义可知, 如果 $d \in S(x^*, \lambda^*)$, 则必有 $d \in \mathrm{LFD}(x^*, X)$, 而且

$$d^{\mathrm{T}} \nabla c_i(x^*) = 0, \qquad 若\ i \in I(x^*), \lambda_i^* > 0. \tag{1.2.51}$$

与 Fletcher (1987) 一样, 定义强积极约束如下.

定义 1.2.17　设 $x^* \in X$ 是问题 (1.1.1)~(1.1.3) 的 K-T 点, λ^* 是相应的 Lagrange 乘子. 如果 $i \in E = \{1, \cdots, m_e\}$ 或者 $\lambda_i^* > 0$, 则称 $c_i(x)$ 是在 x^* 处 (相对于 λ^*) 是强积极的. 我们称

$$\mathcal{A}_+(x^*, \lambda^*) = E \cup \{i \mid i \in I(x^*), \lambda_i^* > 0\} \tag{1.2.52}$$

是在 x^* 处的强积极集合.

利用强积极集合的定义, 有

$$\begin{aligned} G(x^*) &= \mathrm{LFD}(x^*, X) \cap N_{\nabla f(x^*)} \\ &= \mathrm{LFD}(x^*, X) \cap \{d \mid d^{\mathrm{T}} \nabla c_i(x^*) = 0, \forall i \in \mathcal{A}_+(x^*, \lambda^*)\}. \end{aligned} \tag{1.2.53}$$

正因为这一关系, 我们把线性化稳定可行方向也称为线性化零约束方向. 与引理 1.2.8 类似, 可证明如下二阶充分条件:

定理 1.2.18　设 x^* 是问题 (1.1.1)~(1.1.3) 的 K-T 点, λ^* 是相应的 Lagrange 乘子. 如果 $f(x)$ 在 x^* 处二次可微而且

$$d^{\mathrm{T}} W(x^*, \lambda^*) d > 0, \quad \forall\, 0 \neq d \in G(x^*), \tag{1.2.54}$$

则 x^* 是问题 (1.1.1)~(1.1.3) 的局部严格极小点.

在约束规范条件 (1.2.23) 满足时, 有

$$S(x^*, \lambda^*) = G(x^*). \tag{1.2.55}$$

这时, 定理 1.2.16 所给出的二阶必要条件和定理 1.2.18 给出的二阶充分条件已经非常接近了.

在一个 K-T 点, 如果二阶必要性条件满足而二阶充分性条件不满足, 则需要借助更高阶的优化条件来判断它是否为局部极小点, 例如文献 (Lempio et al., 1982). 但是, 绝大多数求解约束优化的方法都不计算二阶导数, 研究更高阶的最优性条件对实际计算的指导意义不大.

§1.3　方法概述

除了少数极特殊情形, 非线性优化问题 (1.1.1)~(1.1.3) 的计算方法都是迭代法. 即给出一个初始值 $x_1 \in \Re^n$, 然后由算法产生迭代点 $x_k (k = 2, 3, \cdots)$, 希望某个 x_k 是 K-T 点或者 $x_k (k = 1, 2, \cdots)$ 收敛于一个 K-T 点.

只有一个变量的优化问题的计算方法是优化方法中最基本的. 由于它的特殊性, 我们在第 2 章专门对其讨论.