# 非线性优化算法

韦增欣　陆　莎　著

科学出版社

# 非线性优化算法

韦增欣　陆　莎　著



科学出版社
北　京

# 内 容 简 介

本书主要对几类常用的非线性优化算法：共轭梯度法、拟牛顿法、邻近点法、信赖域方法以及求解约束优化问题的梯度投影法、有限记忆 BFGS 方法、Topkis-Veinott 方法等逐一作了介绍，尤其着重对这几类算法的改进和扩展应用，包含对共轭梯度法参数的讨论、修正的共轭梯度法、修正的拟牛顿公式及对应的修改的拟牛顿算法、非单调的 BFGS 类算法、非光滑凸优化的一类邻近点模式算法、邻近束方法、带非单调线搜索的 Barzilai-Borwein 梯度法、自适应三次正则化信赖域算法、结合有限记忆 BFGS 的有效集投影信赖域方法、初始点任意的梯度投影法、变形 Topkis-Veinott 方法、子空间有限记忆 BFGS 方法等，以及随机规划 SQP 算法和随机极限载荷分析模型. 对应算法均给出了收敛性质的分析，部分算法给出一些算例和数值试验结果.

本书可供具备最优化理论基础的高等院校数学系高年级本科生，运筹学与控制论、应用数学、计算数学等相关专业的研究生，以及对最优化理论和算法感兴趣的教师和科技工作者阅读.


图书在版编目(CIP)数据

非线性优化算法/韦增欣，陆莎著. —北京：科学出版社，2015.11

ISBN 978-7-03-046237-4

Ⅰ. ①非… Ⅱ. ①韦… ②陆… Ⅲ. ①非线性—最优化算法—算法分析 Ⅳ. ①O224

中国版本图书馆 CIP 数据核字（2015）第 264612 号

责任编辑：李 欣／责任校对：邹慧卿
责任印制：张 倩／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

三河市骏杰印刷有限公司 印刷
科学出版社发行 各地新华书店经销
*
2016 年 1 月第 一 版 开本：720 × 1000 1/16
2016 年 1 月第一次印刷 印张：19
字数：383 000

**定价：118.00 元**

(如有印装质量问题，我社负责调换)

# 前　言

最优化是一门应用性很强的学科, 它讨论决策问题最优解的存在性、稳定性、求解方法以及这些方法的收敛性态和数值表现. 随着计算机技术的高速发展, 最优化理论与方法的应用越来越广泛, 涉及经济、工程、生产、交通、国防等重要领域. 最优化的思想方法与技术是各类技术人员、管理人员所必备的思维方式和基础知识.

本书主要是对几类常用的求解非线性规划问题的方法进行介绍和分析, 大部分内容是作者及合作者多年的研究成果. 对于非线性规划中的无约束优化问题, 重点介绍了在共轭梯度法、拟牛顿法、邻近点法和信赖领域法等的主要结果; 对于非线性规划中的约束优化问题, 则介绍了梯度投影法、有限记忆 BFGS、Topkis-Veinott 法等.

本书分为 6 章, 下面简要叙述每一章的内容.

第 1 章为本书的引论部分, 主要介绍在后面 5 章将要用到的数学基础知识, 包括最优化算法的迭代框架和数值实验等.

第 2 章为共轭梯度法, 在简要介绍了一些传统的共轭梯度法及其收敛性后给出了一族不依赖线搜索而自动具有充分下降性的共轭梯度算法; 利用 Moreau-Yosida 正则函数给出新的线搜索技术进而获得一类修正的 Armijo 型线搜索下的 PRP 共轭梯度法; 由于共轭梯度法中参数非负性的重要性, 作者给出了一个修正的 PRP 方法, 此处所给的参数自动保持非负性, 而且此法中参数的选取技巧很容易推广至其他共轭梯度法.

第 3 章为拟牛顿方法, 通过在第 $k$ 次迭代用逼近函数代替原函数得出新的拟牛顿方程, 由此导出了各类拟牛顿公式. 重点研究了 BFGS 型的拟牛顿算法, 包括两个非单调 BFGS 型算法, 对这些算法的全局收敛性和超线性收敛性都给予了详细的讨论.

第 4 章为邻近点方法, 主要是通过 Moreau-Yosida 正则函数来研究正常下半连续广义值凸函数的最优解的求解问题. 给出了一个很一般的算法框架并讨论其收敛性质; 利用一族二次函数逼近目标函数给出了一类近似方法; 最后求解非凸类优化的非单调线搜索 Barzilai-Borwein 梯度法.

第 5 章为信赖域方法, 首先利用 Moreau-Yosida 正则化技术给出了一个自适应三次正则化信赖域算法, 由于采用的是近似技术, 超线性收敛证明具有一定的难度; 利用有限记忆 BFGS 技术, 给出了求解带框型约束的非光滑方程的有效集投影信

赖域方法; 结合 Moreau-Yosida 正则化函数, 得到带有限记忆 BFGS 的非光滑凸优化信赖域算法; 虽然上述两算法求解的问题不具有光滑性, 它们仍然获得了局部超线性收敛性, 初步的数值结果表明最后一个算法是很有效的.

第 6 章为约束优化问题的一些方法, 核心的内容是首次给出了初始点任意且全局收敛的梯度投影法, 而且只要迭代点落入可行域, 后来的所有迭代均在可行域中进行; 给出变形的 Topkis-Veinott 方法有一些良好的理论性质; 给出的近似 SQP 算法可以用来求解一些复杂问题, 这些问题的目标函数值以及梯度值都很难精确求得; 最后给出了求解大规模有界约束问题的子空间有限记忆 BFGS 方法.

借此机会, 韦增欣对引领他进入优化领域的薛声家教授、赖炎连教授、韩继业教授、越民义教授表示感谢; 对博士生导师祁力群教授表示感谢, 感谢他的培养关心和多年来的鼓励和帮助; 感谢郭柏灵院士、袁亚湘院士、孙德峰教授、戴彧虹教授、修乃华教授、杨新民教授、李董辉教授、李声杰教授、万中教授、凌晨教授、孙文瑜教授、吴至友教授、莫降涛教授、袁功林教授等多年来的支持和帮助.

本书得到国家自然科学基金 (11161003) 的资助, 特此表示致谢.

由于作者水平有限, 本书的不妥之处在所难免, 欢迎读者批评和指正.

作　者

# 目　录

# 第 1 章　引　　论

## 1.1　引　　言

最优化方法是一个重要的数学分支, 源自于军事、管理、经济和工程技术领域的各个方面, 其内容的深度和广度也随着各个不同阶段的科学技术进步而发展. 它所研究的问题是讨论如何对众多方案进行判别与衡量, 并以合适的方法从这些方案中选出最优方案. 例如在确定投资项目时希望选择期望收益最大或者风险最小的项目; 城建规划中, 希望能合理安排工厂、机关、学校等单位的布局; 在输送管道和运输路线的设计中, 希望在满足设计要求的条件下能尽可能的短. 类似这样的问题, 不胜枚举. 要应用最优化技术确定最优方案, 需对具体实际问题进行建模, 再根据相应模型的形式与特征选择出适当的方法进行求解. 最优化为解决这些问题, 提供了理论基础与求解方法, 从而在一切可能的方案中选择一个最好的方案以达到最优的目标. 它是一门应用广泛、实用性强的学科.

非线性最优化问题的一般表达形式为

$$
\text{(NLP)}\begin{cases}\min & f(x);\\ \text{s.t.} & c_i(x)=0,\quad i\in E,\\ & c_i(x)\leqslant 0,\quad i\in I,\\ & x\in R^n.\end{cases}\tag{1.1.1}
$$

其中 $x$ 是决策变量, $f(x)$ 为目标函数, $E$ 和 $I$ 分别表示等式约束指标集和不等式约束指标集, $c_i(x)$ 是约束函数. 记集合

$$
D=\{x|c_i(x)=0,i\in E;c_i(x)\leqslant 0,i\in I,x\in R^n\},
$$

则称 $D$ 为 (NLP) 问题的可行域, 将 $x\in D$ 的点称为可行点. 特别地, 如果 $D=R^n$, 则最优化问题 (1.1.1) 称为无约束优化问题

$$
\min_{x\in R^n} f(x);\tag{1.1.2}
$$

否则, 称最优化问题 (1.1.1) 为约束优化问题.

最优化问题可应用于多个方面, 如下面的一个例子.

**例 1.1.1** 投资组合问题

假设有一笔资金为 $a$ 亿元, 准备将其投资于 $n$ 种证券. 已知第 $i$ 种证券的期望收益率为 $r_i$, 证券收益率之间的协方差矩阵为 $A$. 假设投资到各种证券所构成的资金向量为 $x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$, 且期望收益不低于事先指定的数 $r_0$, 如何根据 Markowitz 的投资组合理论, 建立最优化模型.

**解** 这项投资的风险是 $V(x)=x^{\mathrm{T}}Ax$, 期望收益率是 $R(x)=r_1x_1+r_2x_2+\cdots+r_nx_n$. 我们有如下最优化数学模型, 即投资组合模型:

$$\begin{cases}\min & V(x)=x^{\mathrm{T}}Ax,\\ \text{s.t.} & r_1x_1+r_2x_2+\cdots+r_nx_n\geqslant ar_0,\\ & \displaystyle\sum_{i=1}^{n}x_i=a,\quad x_i\geqslant 0,\quad i=1,\cdots,n.\end{cases}$$

不仅经济金融, 在工程技术、现代化管理、军事领域中还有许多问题可归结为最优化问题来解决, 限于篇幅不再介绍, 有兴趣的读者可以参考相关文献. 本书的相关章节也有一些应用举例, 如随机规划在工程中的应用等.

本书主要研究求解最优化问题的理论和方法, 其中第 2 至第 5 章研究无约束最优化问题, 第 6 章研究约束最优化问题. 书中的大部分是作者及合作者近年来的研究结果.

## 1.2 数学基础

凸性和最优性条件是最优化理论分析中较为重要的一部分内容, 本节将扼要地介绍凸集和凸函数的一些概念, 凸集的分离和支撑、无约束问题和约束优化问题最优性条件的一些结果, 对其证明过程感兴趣的读者可参阅这方面的专著, 如 Rockafellar 的 *Convex Analysis*[154] 等.

### 1.2.1 凸集和凸函数

**定义 1.2.1** (凸集) 设集合 $X\subset R^n$, 若 $\forall x,y\in X$ 与 $\forall\lambda\in[0,1]$, 有 $\lambda x+(1-\lambda)y\in X$, 则称集合 $X$ 是凸集. 称 $\lambda x+(1-\lambda)y$ 为 $x,y$ 的凸组合. 当 $\lambda\in(0,1)$ 时, 称 $\lambda x+(1-\lambda)y$ 为 $x,y$ 的严格凸组合.

**定理 1.2.2** $R^n$ 中任意两个凸集的交集仍然是凸集.

应注意的是, 两个凸集的并集未必是凸集.

**定理 1.2.3** $X$ 为凸集的充分必要条件是: $\forall x_1,x_2,\cdots,x_m\in X$ ($m\geqslant 2$ 为正整数) 及任意非负实数 $\lambda_1,\lambda_2,\cdots,\lambda_m$, 当 $\lambda_1+\lambda_2+\cdots+\lambda_m=1$ 时, 有 $\lambda_1x_1+\lambda_2x_2+\cdots+\lambda_mx_m\in X$.

**定义 1.2.4** (凸函数) 设集合 $X$ 是非空凸集, $f$ 是定义在 $X$ 上的函数, 若 $\forall x,y\in X$ 与 $\lambda\in(0,1)$, 均有 $f(\lambda x+(1-\lambda)y)\leqslant\lambda f(x)+(1-\lambda)f(y)$, 则称 $f$ 为 $X$ 上的凸函数; 若 $\forall x,y\in X$ 与 $\lambda\in(0,1)$, 均有 $f(\lambda x+(1-\lambda)y)<\lambda f(x)+(1-\lambda)f(y)$, 则称 $f$ 为 $X$ 上的严格凸函数.

利用凸函数的定义及相关性质可以判断函数的凸性, 但有时候计算过于繁琐且不便, 因此需要进一步研究函数的判别问题. 在介绍凸函数的判别定理之前先给出以下定义.

**定义 1.2.5** 设函数 $f(x)$ 存在一阶偏导数, $x\in R^n$, 则称向量

$$\nabla f(x)=\left(\frac{\partial f(x)}{\partial x_1},\frac{\partial f(x)}{\partial x_2},\cdots,\frac{\partial f(x)}{\partial x_n}\right)^{\mathrm{T}}$$

为 $f(x)$ 在点 $x$ 处的梯度或一阶导数, 记 $g(x)=\nabla f(x)$.

**定义 1.2.6** 设函数 $f(x)$ 存在二阶偏导数, $x\in R^n$, 则称矩阵

$$\nabla^2 f(x)=\begin{bmatrix}\dfrac{\partial^2 f(x)}{\partial x_1^2} & \dfrac{\partial^2 f(x)}{\partial x_1\partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1\partial x_n}\\ \dfrac{\partial^2 f(x)}{\partial x_2\partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_2\partial x_n}\\ \vdots & \vdots & & \vdots\\ \dfrac{\partial^2 f(x)}{\partial x_n\partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_n\partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2}\end{bmatrix}$$

为 $f(x)$ 在点 $x$ 处的 Hesse 矩阵或二阶导数, 记 $G(x)=\nabla^2 f(x)$.

当 $f(x)$ 为二次函数时, 函数的梯度及 Hesse 阵很容易求得. 二次函数可以写成以下形式

$$f(x)=\frac{1}{2}x^{\mathrm{T}}Ax+b^{\mathrm{T}}x+c.$$

容易验证, 函数 $f(x)$ 在 $x$ 处的梯度为

$$\nabla f(x)=Ax+b,$$

Hesse 阵为

$$\nabla^2 f(x)=A.$$

显然, 二次函数的 Hesse 阵与点的位置无关.

**定理 1.2.7** $f(x)$ 为凸函数的充分必要条件是: $\forall x,y\in R^n,\alpha\in R$, 一元函数 $\varphi(\alpha)=f(x+\alpha y)$ 是关于 $\alpha$ 的凸函数.

**定理 1.2.8** 设 $X \subset R^n$ 为非空开凸集, $f(x)$ 是定义在 $X$ 上的可微函数, 则

(1) $f(x)$ 是 $X$ 上凸函数的充分必要条件是

$$f(x) \geqslant f(x) + \nabla f(x)^{\mathrm{T}}(y-x), \quad \forall x, y \in X. \tag{1.2.1}$$

(2) $f(x)$ 是 $X$ 上严格凸函数的充分必要条件是

$$f(x) > f(x) + \nabla f(x)^{\mathrm{T}}(y-x), \quad \forall x, y \in X, x \neq y. \tag{1.2.2}$$

下面叙述判别凸函数的二阶条件.

**定理 1.2.9** 设 $X \subset R^n$ 为非空开凸集, $f(x)$ 是定义在 $X$ 上的二阶可微函数, 则

(1) $f$ 是 $X$ 上凸函数的充分必要条件是 $f(x)$ 的 Hesse 阵 $\nabla^2 f(x)$ 在 $X$ 上半正定.

(2) 如果 $f(x)$ 的 Hesse 阵 $\nabla^2 f(x)$ 在 $X$ 上正定, 则 $f(x)$ 是 $X$ 上的严格凸函数; 反之, 如果 $f(x)$ 是 $X$ 上的严格凸函数, 则 $\nabla^2 f(x)$ 在 $X$ 上正定.

利用以上几个定理容易判断一个可微函数是否为凸函数, 特别对于二次函数, 可以通过函数的 Hesse 阵来判别函数的凸性. 由于二次函数的 Hesse 阵为常数矩阵, 与所考虑的点无关, 故很容易对其凸性进行判断.

另外, 凸函数亦与水平集关系密切. 下面的定理指出凸函数的水平集是凸集.

**定义 1.2.10** 设 $f(x)$ 是定义在 $X \subset R^n$ 上的函数, $\alpha \in R$, 集合

$$L_\alpha = \{x | x \in X, f(x) \leqslant \alpha\}$$

称为函数 $f$ 的 $\alpha$ 水平集.

**定理 1.2.11** 设 $X \subset R^n$ 是非空凸集, $f$ 是定义在 $X$ 上的凸函数, $\alpha$ 是一个实数, 则水平集 $L_\alpha = \{x | x \in X, f(x) \leqslant \alpha\}$ 是凸集.

**定理 1.2.12** 设水平集 $L(x_0) = \{x \in X | f(x) \leqslant f(x_0)\}$, $f(x)$ 在 $X \subset R^n$ 上二次连续可微, 且存在常数 $m > 0$, 使得

$$u^{\mathrm{T}} \nabla^2 f(x) u \geqslant m\|u\|^2, \quad \forall x \in L(x_0),\ u \in R^n, \tag{1.2.3}$$

则水平集 $L(x_0)$ 是有界闭凸集.

### 1.2.2 最优性条件

优化问题解的类型有局部极小点和全局极小点两种.

**定义 1.2.13** 设 $x^* \in R^n$ (或在约束优化问题中 $x^* \in$ 可行域$D$), 若存在 $\delta > 0$, 使得对于所有满足 $\|x - x^*\| < \delta$ 的 $x \in R^n$(或 $x^* \in$ 可行域$D$), 都有 $f(x) \geqslant f(x^*)$

成立, 则称 $x^*$ 为 $f(x)$ 的局部极小点; 若对所有满足 $x \in R^n$(或 $x^* \in$ 可行域$D$), $x \neq x^*$ 和 $\|x - x^*\| < \delta$ 的 $x$, 都有 $f(x) > f(x^*)$, 则称 $x^*$ 为 $f(x)$ 的严格局部极小点.

**定义 1.2.14**　若对于任意的 $x \in R^n$(或在约束优化问题中 $x^* \in$ 可行域$D$), 都有 $f(x) \geqslant f(x^*)$, 则称 $x^*$ 为 $f(x)$ 的全局极小点; 若对于任意的 $x \in R^n$ (或在约束优化问题中 $x^* \in$ 可行域$D$), 都有 $f(x) > f(x^*)$, 则称 $x^*$ 为 $f(x)$ 的严格全局极小点.

显然, 全局极小点也是一个局部极小点, 而局部极小点却未必是全局极小点.

虽然目前国内外已有许多求全局极小点的算法, 但一般来说这仍是一项相对艰巨的任务. 尤其对较复杂的目标函数或大型数据规模问题, 求出全局极小点往往不可行. 在现实运用中, 求出局部极小点有时也能满足许多问题的要求. 因此, 本书所指的优化问题求解, 通常是指求局部极小点, 仅当问题具有某种凸性时, 局部极小点才是全局极小点.

**定理 1.2.15**　设 $X \subset R^n$ 为凸集, $f: X \to R$, 假定 $\overline{x} \in X$ 是 $f(x)$ 在 $X$ 上的一个局部极小点,

(1) 如果 $f$ 是凸的, 则 $\overline{x}$ 是 $f$ 在 $X$ 上的全局极小点;

(2) 如果 $f$ 是严格凸的, 则 $\overline{x}$ 是 $f$ 在 $X$ 上的唯一的全局极小点.

下面分别给出无约束优化问题及有约束优化问题的最优性条件. 相关定理的证明在许多优化书籍中都能找到, 故在此仅列出结论而略去其证明.

1. 无约束优化问题的最优性条件

考虑无约束优化问题:

$$\min_{x \in R^n} f(x), \tag{1.2.4}$$

对极小点的判断, 有如下一些结论:

**定理 1.2.16** (无约束优化一阶必要条件)　设 $f(x)$ 一阶连续可微. 如果 $x^*$ 是无约束优化问题 (1.2.4) 的一个局部极小点, 则

$$\nabla f(x^*) = 0. \tag{1.2.5}$$

**定理 1.2.17** (无约束优化二阶必要条件)　设 $f(x)$ 二阶连续可微. 如果 $x^*$ 是无约束优化问题 (1.2.4) 的一个局部极小点, 则

$$\nabla f(x^*) = 0, \quad \nabla^2 f(x^*) \text{ 为半正定.}$$

**定理 1.2.18** (无约束优化二阶充分条件)　设 $f(x)$ 二阶连续可微. 如果 $\nabla f(x^*) = 0$, $\nabla^2 f(x^*)$ 为正定, 则 $x^*$ 是无约束优化问题的一个严格局部极小点.

**定理 1.2.19** (无约束凸优化充要条件)　如果 $f: X \to R^n$ 是凸函数, 且 $f(x)$ 一阶连续可微, 则 $x^*$ 是全局极小点的充分必要条件是 $\nabla f(x^*) = 0$.

2. 有约束优化问题的最优性条件

考虑一般约束优化问题:

$$\begin{cases} \min & f(x); \\ \text{s.t.} & c_i(x) = 0, \quad i \in E, \\ & c_i(x) \leqslant 0, \quad i \in I, \\ & x \in R^n. \end{cases} \tag{1.2.6}$$

记指标集 $E = \{1, 2, \cdots, l\}$, $I = \{1, 2, \cdots, m\}$, $I(x) = \{i | c_i(x) = 0\}$, 可行域为 $D$.

约束优化问题 (1.2.6) 的拉格朗日函数为

$$L(x, \mu, \lambda) = f(x) - \sum_{i=1}^{l} \mu_i c_i(x) - \sum_{i=1}^{m} \lambda_i c_i(x), \tag{1.2.7}$$

它关于变量 $x$ 的梯度和 Hesse 矩阵分别为

$$\nabla_x L(x, \mu, \lambda) = \nabla f(x) - \sum_{i=1}^{l} \mu_i \nabla c_i(x) - \sum_{i=1}^{m} \lambda_i \nabla c_i(x),$$

$$\nabla_{xx}^2 L(x, \mu, \lambda) = \nabla^2 f(x) - \sum_{i=1}^{l} \mu_i \nabla^2 c_i(x) - \sum_{i=1}^{m} \lambda_i \nabla^2 c_i(x).$$

对约束优化问题极小点的判断, 有如下定理:

**定理 1.2.20** (KKT 一阶必要条件)　设 $x^*$ 是约束优化问题 (1.2.6) 的一个局部极小点, 在 $x^*$ 处的有效约束集为

$$\mathcal{A}(x^*) = E \bigcup I(x^*),$$

并且 $f(x)$ 和约束函数 $c_i(x)$, $i \in E \bigcup I$ 在 $x^*$ 上均一阶连续可微. 若向量组 $\nabla c_i(x^*)$, $i \in \mathcal{A}(x^*)$ 线性无关, 则存在向量 $(\mu^*, \lambda^*) \in R^l \times R^m$, 其中 $\mu^* = (\mu_1^*, \cdots, \mu_l^*)$, $\lambda^* = (\lambda_1^*, \cdots, \lambda_m^*)$, 使得

$$\begin{cases} \nabla f(x^*) - \sum_{i=1}^{l} \mu_i^* \nabla c_i(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0, \\ c_i(x^*) = 0, \quad i \in E, \\ c_i(x^*) \geqslant 0, \ \lambda_i^* \geqslant 0, \ \lambda_i^* c_i(x^*) = 0, \quad i \in I. \end{cases} \tag{1.2.8}$$

(1.2.8) 称为约束优化问题 (1.2.6) 的 KKT 条件, 满足这一条件的点 $x^*$ 称为 KKT 点, $(x^*; (\mu^*, \lambda^*))$ 称为 KKT 对, 其中 $(\mu^*, \lambda^*)$ 称为约束优化问题的拉格朗日乘子;KKT 条件中的 $\lambda_i^* c_i(x^*) = 0$, $i \in I$ 称为互补松弛条件, 它意味着 $\lambda_i^*$ 和 $c_i(x^*)$ 必须至少有一个为 0, 若两者中有一个为 0 而另一个严格大于 0, 则称之为满足严格互补松弛条件.

**定理 1.2.21** (约束优化二阶充分条件)　对约束优化问题 (1.2.6), 设 $f(x)$ 和约束函数 $c_i(x), i \in E \bigcup I$ 均二阶连续可微, 并且 $x^*$ 为问题的 KKT 点, $(\mu^*, \lambda^*)$ 是对应的拉格朗日乘子. 若对任意的 $d \neq 0, d \in R^n$, $\nabla c_i(x^*)^{\mathrm{T}} d = 0 (i \in \mathcal{A}(x^*))$, 均有

$$d^{\mathrm{T}} \nabla_{xx}^{L}(x^*, \mu^*, \lambda^*) d > 0,$$

则 $x^*$ 是约束优化问题 (1.2.6) 的一个严格局部极小点.

一般而言, 约束优化问题的 KKT 点不一定是局部极小点, 但如果是凸优化问题且满足一定的约束规格, 则 KKT 点、局部极小点、全局极小点三者是等价的.

**定义 1.2.22** (凸优化问题)　当 $f(x)$ 是凸函数, $c_i(x)\,(i \in E)$ 是线性函数, $c_i(x)\,(i \in I)$ 是凸函数时, 约束优化问题 (1.2.6) 称为凸优化问题.

**定理 1.2.23**　设 $x^*$ 是带约束凸优化问题的 KKT 点, 则 $x^*$ 是该问题的全局极小点.

## 1.3　优化算法结构

最优化方法通常采用迭代的方法求出其最优解, 其基本思想是: 从已知点 $x_k$ 出发, 按照某种规则求出后继点 $x_{k+1}$, 继而用 $k+1$ 代替 $k$ 并重复以上过程, 这样就得到一个点列 $\{x_k\}$, 并且使得当 $\{x_k\}$ 是有穷点列时, 其最后一个点是该最优化问题的最优解; 当 $\{x_k\}$ 是无穷点列时, 它有极限点, 且其极限点是最优化问题的最优解. 当然, 一般算法需要首先给出初始点 $x_0$, 以使迭代可以开始进行. 一个好的算法应该具备的典型特征是: 迭代点 $x_k$ 能稳定地接近局部极小点 $x^*$ 的邻域, 然后迅速收敛于 $x^*$. 当给定的某种收敛准则满足时, 迭代即可终止.

下面给出最优化算法的基本结构:

**算法 1.3.1** (最优化算法的基本结构)

步 1. 给定最优解的一个初始估计 $x_0$, 令 $k := 0$;

步 2. 若 $x_k$ 满足对最优解估计的某种终止条件, 则停止迭代;

步 3. 按照某种规则得到改善的下一个迭代点 $x_{k+1}$;

步 4. 令 $k := k+1$, 转步 2.

在上述基本结构中涉及了初始点的选取、迭代的终止准则、收敛速度、搜索方向和下一个迭代点的产生等问题. 下面对这些分别加以简单的讨论.

### 1.3.1　初始点的选取

初始点的选取往往对算法的收敛性能产生影响. 如果一个算法产生的序列 $\{x_k\}$ 满足

$$\lim_{k \to \infty} \|x_k - x^*\| = 0, \tag{1.3.1}$$

称其为收敛的. 一个算法如果对于任意给定的初始点都能够收敛, 则说明该方法是全局收敛的. 如果只有当初始点接近或者充分接近最优解时才有收敛性, 则称这样的算法为局部收敛的. 因此, 对于全局收敛的算法, 没有任何对于初始点的选取的限制. 而对于局部收敛的算法, 一方面要求初始点应尽可能的接近最优解; 另一方面, 由于最优解是未知的, 选取一个好的初始点也是一个困难的问题. 对于大量的实际优化问题一般可以从以往的实践经验确定出合适的最优解的初始估计.

### 1.3.2　终止准则

在最优化方法中, 一般会选用一个评价函数来评价迭代点的好坏. 对于无约束最优化问题, 由于没有约束条件, 通常便以目标函数 $f(x)$ 作为评价函数. 以无约束极小化问题 $\min f(x)$ 为例, 如果有 $f(x_{k+1}) < f(x_k)$, 就是说点 $x_{k+1}$ 要好于点 $x_k$, 即要求产生的迭代序列 $\{x_k\}$ 使评价函数值单调下降. 而对于约束最优化问题, 则要复杂一些. 如果迭代点都是可行点, 当然可以直接用目标函数作评价函数, 这适用于迭代点都是可行点的方法. 但是如果迭代点不是可行点, 判定一个点的好坏既要考虑目标函数值的大小, 还要考虑这个点的可行程度. 因此, 这类方法采用的评价函数中一般既包含目标函数又包含约束函数. 对此有许多不同类型的评价函数, 有关评价函数将在第 6 章加以介绍.

迭代的终止条件在不同的最优化方法中也是不同的. 一个理想的算法的终止准则为

$$\|f(x_k) - f(x^*)\| \leqslant \varepsilon.$$

但是由于 $x^*$ 是未知的, 这样的准则并不具有任何的实用价值. 注意到

$$\begin{aligned}\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} &= \frac{\|x_{k+1} - x_k + x_k - x^*\|}{\|x_k - x^*\|} \\ &\geqslant \left|\frac{\|x_{k+1} - x_k\| - \|x_k - x^*\|}{\|x_k - x^*\|}\right| \\ &= \left|\frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} - 1\right| \geqslant 0,\end{aligned}$$

在序列 $\{x_k\}$ 超线性收敛于 $x^*$ 时, 可以得到

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} = 1.$$

上式表明, $\|x_{k+1} - x_k\|$ 可以用来代替 $\|x_k - x^*\|$ 给出终止判断, 并且这个估计随着 $k$ 的增加而改善. 因此, 对于具有超线性收敛速度的方法,

$$\|x_{k+1} - x_k\| \leqslant \varepsilon \tag{1.3.2}$$

是比较合适的收敛准则. 同样也可以用评价函数值序列来确定终止准则, 由于 $x^*$ 未知, $\|f(x_k)-f(x^*)\|\leqslant\varepsilon$ 不可能直接作为收敛准则, 但当目标函数二次连续可微时可以推得 $|f(x_{k+1})-f(x_k)|=O(\|x_{k+1}-x_k\|^2)$, 因此对于快速收敛的算法,

$$\|f(x_{k+1})-f(x_k)\|\leqslant\varepsilon \tag{1.3.3}$$

也是一个相对有效的收敛准则. 但是, 在有些情况下, 终止准则 (1.3.2) 和 (1.3.3) 是不适当的. 因为有时虽然 $\|x_{k+1}-x_k\|$ 是小的, 但 $f(x_{k+1})-f(x_k)$ 是大的, 极小点 $x^*$ 远离 $x_k$; 有时虽然 $f(x_k)-f(x_{k+1})$ 是小的, 但 $\|x_{k+1}-x_k\|$ 又是大的, 极小点 $x^*$ 也远离 $x_k$. 因此, 在有些方法中为了确保所得的是最优解理想的估计, 往往采用两个或几个收敛准则同时使用的方法. 例如, 当 $\|x_k\|>\varepsilon_1$ 和 $\|f(x_k)\|>\varepsilon_2$ 时, 采用

$$\frac{\|x_{k+1}-x_k\|}{\|x_k\|}\leqslant\varepsilon_1,\quad \frac{\|f(x_k)-f(x_{k+1})\|}{\|f(x_k)\|}\leqslant\varepsilon_2, \tag{1.3.4}$$

否则采用

$$\|x_{k+1}-x_k\|\leqslant\varepsilon_1,\quad \|f(x_k)-f(x_{k+1})\|\leqslant\varepsilon_2. \tag{1.3.5}$$

对于有一阶导数信息, 且收敛不太快的算法, 例如共轭梯度法, 可以采用如下终止准则:

$$\|g_k\|\leqslant\varepsilon_3, \tag{1.3.6}$$

其中, $g_k=g(x_k)=\nabla f(x_k)$. 此外, 也可以把 (1.3.6) 和其他准则组合起来使用.

### 1.3.3 收敛速度

收敛速度也是衡量最优化方法的重要方面. 对于一个不能在有限步内找到最优解的优化算法, 我们不仅要求它收敛, 还要求它要有较快的收敛速度, 这是因为一个收敛很慢的方法一般都需要较长的时间才能找到满足精度要求的最优解的近似, 在实际应用中不是一个有效的方法.

若算法满足 (1.3.1), 且存在实数 $\alpha>0$ 及一个与迭代次数 $k$ 无关的常数 $q>0$, 使得

$$\lim_{k\to\infty}\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|^\alpha}=q,$$

则称算法产生的迭代点列 $\{x_k\}$ 具有 $Q$-$\alpha$ 阶收敛速度, 特别地,

(1) 当 $\alpha=1,q>0$ 时, 称迭代点列 $\{x_k\}$ 具有 $Q$-线性收敛速度;

(2) 当 $1<\alpha<2,q>0$ 或 $\alpha=1,q=0$ 时, 称迭代点列 $\{x_k\}$ 具有 $Q$-超线性收敛速度;

(3) 当 $\alpha=2$ 时, 称迭代点列 $\{x_k\}$ 具有 $Q$-二阶收敛速度.

一般认为, 具有超线性收敛速度和二阶收敛速度的方法是比较快速的. 当然我们也应该意识到, 一个算法收敛性和收敛速度的理论结果并不保证算法在实际执行