

OR

S.Olivia

September 2023

目录

1	Introduction	5
1.1	数学基础	5
1.1.1	梯度与黑塞矩阵	5
1.1.1.1	梯度	5
1.1.1.2	黑塞矩阵	5
1.1.2	正定矩阵	5
1.2	基本理论	6
1.2.1	数学规划模型的一般形式	6
1.2.2	凸集、凸函数和凸规划	6
1.2.2.1	凸集	6
1.2.2.2	凸函数	6
1.2.2.2.1	水平集	6
1.2.2.2.2	凸函数的性质	7
1.2.2.3	凸规划	8
1.2.3	多面体、极点、极方向	8
2	Linear Programming Problem	9
2.1	Graphical Method	9
2.2	Simplex Method	9
2.2.1	单纯形法的表格计算	10
2.2.2	一般线性规划问题的处理	12
2.2.2.1	大 M 法	12
2.2.2.2	二阶段法	13
2.2.3	线性规划的对偶问题	15
2.2.3.1	对偶问题的形式	15
2.2.3.2	对偶定理	16
2.2.3.3	对偶单纯形法	17
2.2.4	灵敏度分析	17
2.2.4.1	影子价格	18
2.2.4.2	目标函数系数 c 变化*	19
2.2.4.3	右端常数 b 变化*	19
2.2.4.4	约束条件系数 a 变化	20

2.2.4.5	新增变量 x 分析	20
2.2.4.6	新增约束条件	21
3	最优化搜索算法的结构和一维搜索	23
3.1	常用的搜索算法结构	23
3.1.1	收敛性概念	23
3.1.2	收敛准则 (停机条件)	24
3.1.3	收敛速度	24
3.1.4	二次终结性	24
3.1.4.1	共轭方向	24
3.1.5	下降算法模型	24
3.2	一维搜索	25
3.2.1	缩小区间的精确一维搜索	26
3.2.1.1	不确定区间和单峰函数	26
3.2.1.2	进退法求出初始不确定区间	27
3.2.1.3	中点法	27
3.2.1.4	黄金分割法	28
3.2.2	牛顿法和插值法	28
3.2.2.1	牛顿法	28
3.2.2.2	插值法	29
3.2.3	不精确的一维搜索	30
3.2.3.1	Goldstein 法	30
3.2.3.2	Wolfe-Powell 法	32
4	无约束最优化方法	33
4.1	最优性条件	33
4.2	最速下降法	33
4.3	牛顿法及其修正	34
4.3.1	牛顿法	34
4.3.2	减少计算量的直接修正	35
4.3.3	带一维线性搜索的牛顿算法	35
4.3.4	Goldstein-Price 方法	36
4.3.5	L-M 法	36
4.4	共轭梯度法	36
4.4.1	共轭梯度法的方向	36
4.4.2	再开始共轭梯度法	38
4.4.3	算法特点	38
4.5	变尺度法	39
4.5.1	基本思路	39
4.5.1.1	拟牛顿方程	39
4.5.2	变尺度法的特点	40

目录	5
5 约束最优化方法	41
5.1 Kuhn-Tucker 条件	41
5.1.1 等式约束的最优性条件	41
5.1.2 不等式约束问题的 Kuhn-Tucker 条件	42
5.1.3 一般约束问题的 Kuhn-Tucker 条件	43
5.2 既约梯度法	43

Chapter 1

Introduction

1.1 数学基础

1.1.1 梯度与黑塞矩阵

简便记法

1.1.1.1 梯度

$$\nabla f(x_1, x_2, x_3) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3} \right)^T \quad (1.1)$$

1.1.1.2 黑塞矩阵

$$\nabla^2 f(x_1, x_2, x_3) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_3 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_3} & \frac{\partial^2 f}{\partial x_2 \partial x_3} & \frac{\partial^2 f}{\partial x_3^2} \end{pmatrix} \quad (1.2)$$

简而言之，式 (1.2) 即为式 (1.1) 逐行关于参数（即 x_1, x_2, x_3, \dots ）求偏导。

1.1.2 正定矩阵

对一个 n 阶方阵 \mathbf{M} ，如果其对任意非零向量 \mathbf{Z} 都有 $\mathbf{Z}^T \mathbf{M} \mathbf{Z} > 0$ ，其中 \mathbf{Z}^T 为 \mathbf{Z} 的转置，则 \mathbf{M} 为正定矩阵。

定理 1.1 对于 n 阶实对称矩阵 \mathbf{M} ，下列条件是等价的：

1. \mathbf{M} 是正定矩阵。
2. \mathbf{M} 的特征值均为正。
3. \mathbf{M} 的一切顺序主子式均为正。

1.2 基本理论

1.2.1 数学规划模型的一般形式

$$(fS) = \begin{cases} \min & f(x) \\ \text{s.t.} & x \in S \end{cases} \quad (1.3)$$

1.2.2 凸集、凸函数和凸规划

1.2.2.1 凸集

定义 1.1 设 $S \subset \mathbf{R}^n$, 如果 $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in S, \lambda \in [0, 1]$, 均有

$$\lambda \mathbf{x}^{(1)} + (1 - \lambda) \mathbf{x}^{(2)} \in S \quad (1.4)$$

则称 S 为凸集

集合中任意两点连成的线段必属于该集合; 规定空集 \emptyset 为凸集, 单点集 $\{x\}$ 为凸集。

性质:

1. 凸集的交集是凸集
2. 凸集的内点集是凸集
3. 凸集的闭包是凸集
4. 分离和支撑: 凸集边界上任意点存在支撑超平面; 两个相互不交的凸集之间存在分离超平面

有一特殊的凸集: 凸锥

定义 1.2 设非空集合 $C \subset \mathbf{R}^n$, 如果 $\forall \mathbf{x} \in C$ 对 $\forall \lambda > 0$ 有 $\lambda \mathbf{x} \in C$, 则称 C 为以 0 为顶点的锥 (不一定含 0 点)。若 C 又是凸集, 则称 C 为凸锥。

1.2.2.2 凸函数

定义 1.3 设 $S \subset \mathbf{R}^n$, 非空, 凸集, 函数 $f: S \rightarrow \mathbf{R}$, 如果对 $\forall \mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in S, \forall \lambda \in (0, 1)$ 恒有

$$f(\lambda \mathbf{x}^{(1)} + (1 - \lambda) \mathbf{x}^{(2)}) \leq \lambda f(\mathbf{x}^{(1)}) + (1 - \lambda) f(\mathbf{x}^{(2)}) \quad (1.5)$$

则称 f 为 S 上的凸函数。如果式 (1.5) 恒以严格不等式成立, 则称 f 为 S 上的严格凸函数。

几何意义: 任意两个点连线在函数曲线的上方。

1.2.2.2.1 水平集

定义 1.4 设 $S \subset \mathbf{R}^n$, 非空, $f: S \rightarrow \mathbf{R}, a \in \mathbf{R}$, 则称

$$S_\alpha = \{\mathbf{x} | f(\mathbf{x}) \leq \alpha, \mathbf{x} \in S\} \quad (1.6)$$

为 f 的水平集。

水平集的概念相当于在地形图中, 海拔高度不高于某一数值的区域。

注意: 容易证明, 当 f 为凸函数时, $\forall \alpha \in \mathbf{R}, S_\alpha$ 是凸集。但是它的逆不成立。

1.2.2.2.2 凸函数的性质

1. $f(x)$ 为凸集 S 上的凸函数 $\Leftrightarrow S$ 上任意有限点的凸组合的函数值不大于各点函数值的凸组合
2. 设 f_1, f_2 为凸函数, $\lambda_1, \lambda_2 > 0$, 则:
 - $\lambda_1 f_1 + \lambda_2 f_2$ 是凸函数
 - $f(x) = \max\{f_1(x), f_2(x)\}$ 是凸函数
 - $g(x) = \min\{f_1(x), f_2(x)\}$ 不一定是凸函数
3. 若 f 在 S 上凸, 那么 f 在 S 的内点集 ($\text{int}S$) 上连续 (注: f 在 S 上不一定连续)
4. 若 f 在非空凸集 S 上凸, 则对任意方向的方向导数存在

凸函数常用判定条件:

5. 设 S 非空, 凸集, 开集, f 在 S 上可微, 则:
 - f 在 S 上凸 $\Leftrightarrow \forall \bar{x} \in S$, 有 $f(x) \geq f(\bar{x}) + \nabla f^T(\bar{x})(x - \bar{x}), \forall x \in S$ 。
 - f 在 S 上严格凸 $\Leftrightarrow \forall \bar{x} \in S$, 有 $f(x) > f(\bar{x}) + \nabla f^T(\bar{x})(x - \bar{x}), \forall x \in S, x \neq \bar{x}$ 。
6. 设 S 非空, 凸集, 开集, f 在 S 上可微, 则:
 - f 在 S 上凸 $\Leftrightarrow \forall x^{(1)}, x^{(2)} \in S, (\nabla f(x^{(1)}) - \nabla f(x^{(2)}))^T(x^{(1)} - x^{(2)}) \geq 0$
 - f 在 S 上严格凸 $\Leftrightarrow \forall x^{(1)}, x^{(2)} \in S, x^{(1)} \neq x^{(2)}, (\nabla f(x^{(1)}) - \nabla f(x^{(2)}))^T(x^{(1)} - x^{(2)}) > 0$
7. 设 S 是开集, f 在 S 上二次可微, 则:
 - f 在 S 上凸 $\Leftrightarrow \forall x \in S, \nabla^2 f(x)$ 半正定。
 - 如果 $\forall x \in S, \nabla^2 f(x)$ 正定, 则 f 在 S 上严格凸。

但是, 上课讲的时候, 一般是使用**黑塞矩阵是否正定**判断

1. 当 H 为半正定时, f 为凸函数; 若 H 是正半定的, 当且仅当 H 的每一个主子式都大于等于 0
2. 当 H 为正定时, f 为严格凸函数; 若 H 是正定矩阵, 当且仅当 H 的 n 个顺序主子式 (严格) 为正。
3. 当 H 为半负定时, f 为凹函数; 若 H 是负半定的, 当且仅当 H 的每一个奇阶的主子式小于等于 0, 每一个偶阶的主子式大于等于 0
4. 当 H 为负定时, f 为严格凹函数; 若 H 是负定矩阵, 当且仅当 H 的 N 个顺序主子式以如下方式交替出现: $|A_1| < 0, |A_2| > 0, |A_3| < 0, \dots$
5. 当 H 为不定时, f 既非凸也非凹函数。

(! 注意: 只能一个方向判断, 不能由是否凸函数来判断矩阵是否正定)

具体可看: [数学选读 01: 矩阵的主子式与顺序主子式](#)

1.2.2.3 凸规划

定义 1.5 有两种情况

- 当 (fS) 中, S 为凸集, f 是 S 上的凸函数 (求 \min 时), 称 (fS) 为凸规划。
- 对于 (fgh) , 当 f, g_i 为凸函数, h_i 为线性函数时, (fgh) 为凸规划。

定理 1.2 设 $S \subset \mathbf{R}^n$, 非空, 凸, $f: S \rightarrow \mathbf{R}$ 是凸函数。 x^* 为问题 (fS) 的 l.opt., 则 x^* 为 g.opt.; 又如果 f 是严格凸函数, 那么 x^* 是问题 (fS) 的唯一 g.opt.。

1.2.3 多面体、极点、极方向

(看书上 p29 开始的图更好理解)

多面体: 有限个半闭空间的交为多面体

极点: $x \in S$, 不存在 S 中另外两个点 $x^{(1)}$ 和 $x^{(2)}$, 及 $\lambda \in (0, 1)$, 使 $x = \lambda x^{(1)} + (1 - \lambda)x^{(2)}$

根据定义, 闭球体的表面上每一点都是极点; 一般的闭凸锥有唯一极点, 即顶点; 平面没有极点;

极方向:

- 方向: $d \in \mathbf{R}^n, d \neq 0$ 及对于任意 $x \in S, \lambda > 0$, 总有 $x + \lambda d \in S$ (可行方向)。其中, 当 $d^{(1)} = \lambda d^{(2)} (\lambda > 0)$ 时, 称 $d^{(1)}$ 和 $d^{(2)}$ 同方向。
- 极方向: 方向 d 不能表示为两个不同方向的组合 ($d = d^{(1)} + d^{(2)}$)。

定理 1.3 (极点特征定理) 设 A 行满秩, x 是 S 极点的充要条件是:

存在分解 $A = (B, N)$, 其中 B 为 m 阶非奇异矩阵, 使 $x^T = (x_B^T, x_N^T)$, 这里 $x_B = B^{-1}b \geq 0, x_N = 0$

S 中必定存在有限多个极点 ($\leq C_n^m$)

Chapter 2

Linear Programming Problem

2.1 Graphical Method

图解法

具体可看 ppt 或：【运筹学】线性规划图解法（唯一最优解 | 无穷最优解 | 无界解 | 无可行解）

基本解：各个等式约束直线的交点，外加与坐标轴的交点

基本可行解：基本解里面在可行域范围的那些基本解，可行域的顶点

最优解：基本可行解里面使目标函数最大（最小）的基本可行解

2.2 Simplex Method

一些必要概念：（建议先了解解题步骤，再回过头来看概念）

线性规划的标准形式

$$(LP) \begin{cases} \max & \mathbf{c}^T \mathbf{x} \\ s.t. & \mathbf{Ax} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{cases} \quad (2.1)$$

其中， \mathbf{A} 是 $m \times n$ 的矩阵 ($m < n$)

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1m} & a_{1(m+1)} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mm} & a_{m(m+1)} & \cdots & a_{mn} \end{pmatrix}$$

可行解：在公式 (2.1) 中， $\mathbf{x} = (x_1, \dots, x_n)^T$ 即为可行解，可以理解为多面体的极点（顶点？）

基本可行解：可以理解为满足非负约束条件的可行解？（非负约束条件大概就是公式 (2.1) 中 $\mathbf{x} \geq \mathbf{0}$ ）

基： \mathbf{B} 是线性规划问题的一个基

$$\mathbf{B} = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{pmatrix}$$

\mathbf{B} 中的每个列向量 $\mathbf{p}_j = (a_{1j}, a_{2j}, \dots, a_{mj})^T$ 为基向量，基向量中的 a_{mj} 们即为基变量。一般的，我们讲基变量为 m 个线性无关的变量。（一般是标准化后的松弛变量所对应的列们的系数，也是单位矩阵）

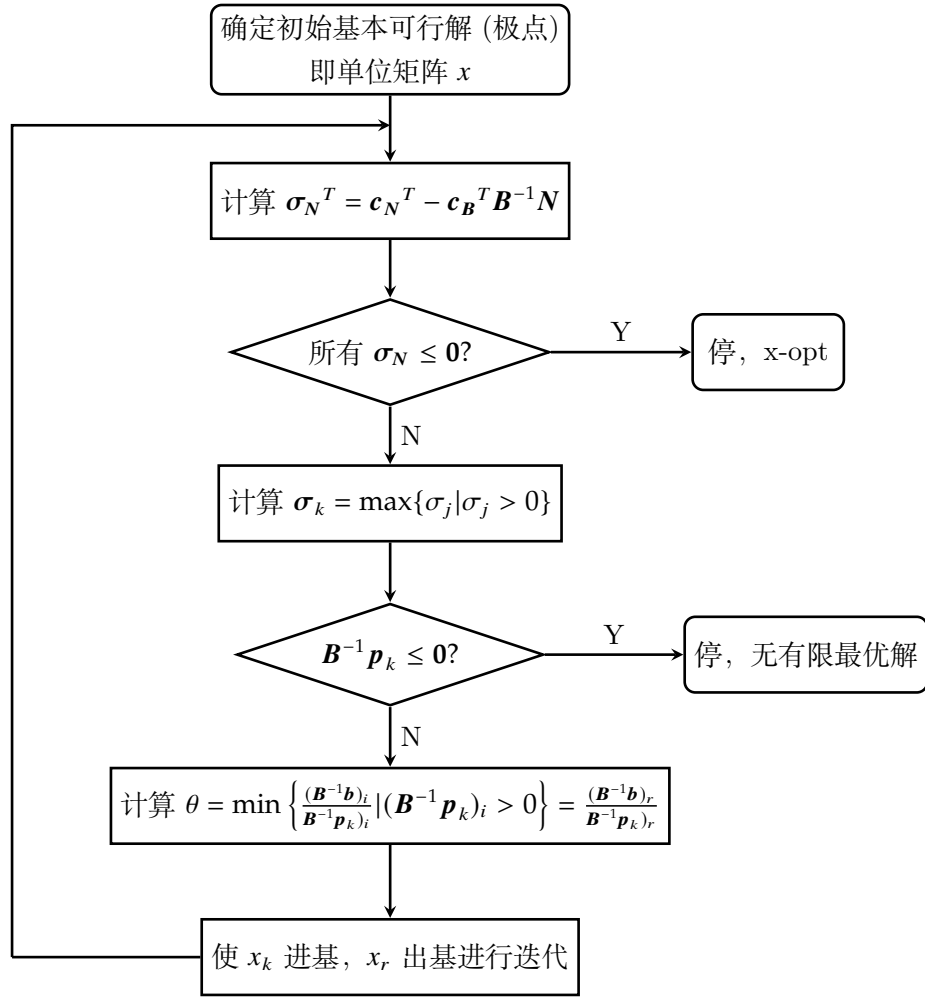


图 2.1: 单纯形法流程图

需要注意的是, σ 在表中的计算没有那么复杂, 可以先略过。

2.2.1 单纯形法的表格计算

考虑规范形式的线性规划问题: $b_i > 0, i = 1, \dots, m$

$$\begin{aligned}
 \max \quad & z = c_1 x_1 + c_2 x_2 + \dots + c_n x_n \\
 \text{s.t.} \quad & \begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m \\ x_1, x_2, \dots, x_n \geq 0 \end{cases}
 \end{aligned}$$

加入松弛变量，化为标准形：

$$\left\{ \begin{array}{ll} \max & z = c_1x_1 + c_2x_2 + \cdots + c_nx_n \\ \text{s.t.} & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + x_{n+1} = b_1 \\ & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + x_{n+2} = b_2 \\ & \vdots \\ & a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + x_{n+m} = b_m \\ & x_1, x_2, \cdots, x_n, x_{n+1}, \cdots, x_{n+m} \geq 0 \end{array} \right. \quad (2.2)$$

STEPS:

1. 根据公式 (2.2) 构造初始单纯形表：

c_B	x_B	b	c_1	c_2	\cdots	c_n	c_{n+1}	c_{n+2}	\cdots	c_{n+m}	θ
			x_1	x_2	\cdots	x_n	x_{n+1}	x_{n+2}	\cdots	x_{n+m}	
c_{n+1}	x_{n+1}	b_1	a_{11}	a_{12}	\cdots	a_{1n}	1	0	\cdots	0	θ_1
c_{n+2}	x_{n+2}	b_2	a_{21}	a_{22}	\cdots	a_{2n}	0	1	\cdots	0	θ_2
\vdots	\vdots	\vdots	\vdots	\vdots		\vdots	\vdots	\vdots		\vdots	\vdots
c_{n+m}	x_{n+m}	b_m	a_{m1}	a_{m2}	\cdots	a_{mn}	0	0	\cdots	1	θ_m
$-z$	$-z'$		σ_1	σ_2	\cdots	σ_n	0	0	\cdots	0	

表 2.1: 单纯形表

表 (2.1) 中；

- 列 c_B 填入目标函数中基变量 x 的系数
- 列 x_B 填入基变量
- 列 b 填入约束方程右端的常数
- 中间 4-11 列填入约束方程中 x 的系数，其中从 x_{n+1} 开始的列组成单位矩阵

2. 求出 $-z' = -\sum_{i=1}^m c_{n+i}b_i$ (初始状态下一一般为 0)

3. 求出**检验数** $\sigma_j = c_j - \sum_{i=1}^m c_{n+i}a_{ij}$ ，可以通俗理解为 $\sigma_1 =$ 第一行中的 c_1 减去 (c_B 的每一项与 c_1 列对应的 a_{11} 相乘的求和)。

需要注意的是，这里的检验数不需要考虑基（即单位矩阵）对应的列，否则会影响下一步。

4. 判断是否所有检验数 $\sigma_j \leq 0$ ，如果全部小于等于 0，则当前的基本可行解是最优解；如果有一检验数大于 0，那么进行下一步的计算。

5. 求入基变量 x_k ：

- 先求出大于 0 的检验数中最大的**检验数** σ_k
- 得到的下标 k ， σ_k 所在的列就是主元列

- 那么所对应的 x_k 就是入基变量

6. 求出基变量 x_r :

- 找出主元列 (x_k 列) 对应元素 $a_{ik} > 0$ 的
- 使每个大于 0 的 a_{ik} 被对应行中的 b_i 除, 得到 $\theta_i = \frac{b_i}{a_{ik}}$
(若 $a_{ik} \leq 0$, 则 $\theta_i = \infty$)
- 找出 $\min\{\theta_i\} = \theta_r$
- 此时 θ_r 所在的行 r 即为主元行, x_r 为出基变量

7. 求出入基变量 x_k 和出基变量 x_r 后, 就可以构建下一张表, 这张表中:

- x_B 上的出基变量被入基变量所取代, 同时需要更改列 c_B 中与之相对应的行的元素
- 在表右侧中间部分, 使用高斯消元法将 x_k 对应的 a 与其他基变量对应的 a 组成一个单位矩阵, 同时列 b 也跟着改变
- 表 2 基本构建完成, 重复第二步, 直至出现流程表中 stop 的情况

2.2.2 一般线性规划问题的处理

在某些情况下, 如基本初始可行解不明显, 即很难在标准形的问题下找到单位矩阵时, 可以考虑使用大 M 法与二阶段法

2.2.2.1 大 M 法

STEPS:

1. 将线性规划问题转化为标准型
2. 观察变量, 若初始基本可行解明显, 直接进行单纯形法; 否则引入人工变量 $x_{n+i} \geq 0 (i = 1, \dots, m)$ 及充分大正数 M , 改写原目标函数, 进行单纯形法
3. 若得到的最优解满足:

$$x_{n+i} = 0 (i = 1, \dots, m)$$

则是原问题的最优解; 否则, 原问题无可行解

例: 使用大 M 法求解下面的问题

$$\begin{aligned} \max \quad & z = 5x_1 + 2x_2 + 3x_3 - x_4 \\ \text{s.t.} \quad & \begin{cases} x_1 + 2x_2 + 3x_3 &= 15 \\ 2x_1 + x_2 + 5x_3 &= 20 \\ x_1 + 2x_2 + 4x_3 + x_4 &= 26 \\ x_1, x_2, x_3, x_4 &\geq 0 \end{cases} \end{aligned}$$

标准化并引入人工变量：

$$\begin{aligned} \max \quad & z = 5x_1 + 2x_2 + 3x_3 - x_4 - Mx_5 - Mx_6 \\ \text{s.t.} \quad & \begin{cases} x_1 + 2x_2 + 3x_3 + x_5 = 15 \\ 2x_1 + x_2 + 5x_3 + x_6 = 20 \\ x_1 + 2x_2 + 4x_3 + x_4 = 26 \\ x_1, x_2, x_3, x_4, x_5, x_6 \geq 0 \end{cases} \end{aligned}$$

最后使用单纯形法计算

c_B	x_B	b	5	2	3	-1	$-M$	$-M$	θ
			x_1	x_2	x_3	x_4	x_5	x_6	
$-M$	x_5	15	1	2	3	0	1	0	5
$-M$	x_6	20	2	1	[5]	0	0	1	4
-1	x_4	26	1	2	4	1	0	0	6.5
$-z$		$35M + 26$	$3M + 6$	$3M + 4$	$8M + 7$	0	0	0	
$-M$	x_5	3	-1/5	[7/5]	0	0	1	-3/5	15/7
3	x_3	4	2/5	1/5	1	0	0	1/5	20
-1	x_4	10	-3/5	6/5	0	1	0	-4/5	25/3
$-z$		$3M - 2$	$-M/5 + 16/5$	$7/5M + 13/5$	0	0	0	$-8/5M - 7/5$	
2	x_2	15/7	-1/7	1	0	0	5/7	-3/7	25/3
3	x_3	25/7	[3/7]	0	1	0	-1/7	2/7	
-1	x_4	52/7	-3/7	0	0	1	-6/7	-2/7	
$-z$		-53/7	25/7	0	0	0	$-M - 13/7$	$-M - 2/7$	
2	x_2	10/3	0	1	1/3	0	2/3	-1/3	
5	x_1	25/3	1	0	7/3	0	-1/3	2/3	
-1	x_4	11	0	0	1	1	-1	-0	
$-z$		-112/3	0	0	-25/3	0	$-M - 2/3$	$-M + 8/3$	

表 2.2: 大 M 法例题

得到基本可行解： $(25/3, 10/3, 0, 11)^T$ 为最优解

得到最优值 $z = 112/3$

2.2.2.2 二阶段法

STEPS:

1. 将线性规划问题转化为标准型
2. 观察变量，若初始基本可行解明显，直接进行单纯形法；否则引入人工变量 $x_{n+i} \geq 0 (i = 1, \dots, m)$ 构造辅助问题 (LP - 1)
3. 第一阶段，求解辅助问题 (LP - 1)，若得到的最优解满足 $x_{n+i} = 0 (i = 1, \dots, m)$ ，则是原问题的基本可行解；否则，原问题无可行解。

4. 第二阶段, 得到原问题的基本可行解后, 直接删除人工变量, 求解原问题

例: 使用二阶段法求解下面的问题

$$\begin{aligned} \max \quad & z = 5x_1 + 2x_2 + 3x_3 - x_4 \\ \text{s.t.} \quad & \begin{cases} x_1 + 2x_2 + 3x_3 & = 15 \\ 2x_1 + x_2 + 5x_3 & = 20 \\ x_1 + 2x_2 + 4x_3 + x_4 & = 26 \\ x_1, x_2, x_3, x_4 \geq 0 \end{cases} \end{aligned}$$

标准化、引入人工变量并构造第一阶段问题 (LP - 1):

$$\begin{aligned} \max \quad & z' = -x_5 - x_6 \\ \text{s.t.} \quad & \begin{cases} x_1 + 2x_2 + 3x_3 + x_5 & = 15 \\ 2x_1 + x_2 + 5x_3 + x_6 & = 20 \\ x_1 + 2x_2 + 4x_3 + x_4 & = 26 \\ x_1, x_2, x_3, x_4, x_5, x_6 \geq 0 \end{cases} \end{aligned}$$

建立第一阶段的单纯形表:

c_B	x_B	b	0	0	0	0	-1	-1	θ
			x_1	x_2	x_3	x_4	x_5	x_6	
-1	x_5	15	1	2	3	0	1	0	5
-1	x_6	20	2	1	[5]	0	0	1	4
0	x_4	26	1	2	4	1	0	0	6.5
$-z$		35	3	3	8	0	0	0	
-1	x_5	3	-1/5	[7/5]	0	0	1	-3/5	15/7
0	x_3	4	2/5	1/5	1	0	0	1/5	20
0	x_4	10	-3/5	6/5	0	1	0	-4/5	25/3
$-z$		3	-1/5	7/5	0	0	0	-8/5	
2	x_2	15/7	-1/7	1	0	0	5/7	-3/7	25/3
3	x_3	25/7	[3/7]	0	1	0	-1/7	2/7	
-1	x_4	52/7	-3/7	0	0	1	-6/7	-2/7	
$-z$		0	0	0	0	0	-1	-1	

表 2.3: 二阶段法例题 1

得到原问题的基本可行解 $(0, 15/7, 25/7, 52/7)^T$

第二阶段删除人工变量, 并把基本可行解填入表中

得到基本可行解: $(25/3, 10/3, 0, 11)^T$ 为最优解

得到最优值 $z = 112/3$

(表见下页)

c_B	x_B	b	0	0	0	0	θ
			x_1	x_2	x_3	x_4	
2	x_2	15/7	-1/7	1	0	0	25/3
3	x_3	25/7	[3/7]	0	1	0	
-1	x_4	52/7	-3/7	0	0	1	
$-z$		-53/7	25/7	0	0	0	
2	x_2	10/3	0	1	1/3	0	
5	x_1	25/3	1	0	7/3	0	
-1	x_4	11	0	0	1	1	
$-z$		-112/3	0	0	-25/3	0	

表 2.4: 二阶段法例题 2

2.2.3 线性规划的对偶问题

给定一个优化问题，我们去理解它的时候，或者设计算法的时候，可以研究它的对偶。

有时原问题不好解，但它的对偶相对容易。这个时候，可以从对偶问题出发，进而寻求原问题的解。

2.2.3.1 对偶问题的形式

1. 对称形式的对偶问题

$$(P) \begin{cases} \max & z = c^T x \\ \text{s.t.} & Ax \leq b \\ & x \geq 0 \end{cases} \quad (D) \begin{cases} \min & f = b^T y \\ \text{s.t.} & A^T y \geq c \\ & y \geq 0 \end{cases}$$

一对对称形式的对偶规划之间具有下面的对应关系

- “max ≤” 和 “min ≥” 相对应
- 从约束系数矩阵看：一个模型中为 A ，则另一个模型中为 A^T ；一个模型是 m 个约束、 n 个变量，则它的对偶模型为 n 个约束、 m 个变量
- 从数据 b 、 c 的位置看：在两个规划模型中， b 和 c 的位置对换
- 两个规划模型中的变量皆非负

2. 非对称形式的对偶问题

(下式并不囊括所有情况)

$$(P) \begin{cases} \max & z = c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{cases} \quad (D) \begin{cases} \min & f = b^T y \\ \text{s.t.} & A^T y \geq c \end{cases}$$

一般称不具有对称形式的一对线性规划为非对称形式的对偶规划。对于非对称形式的规划，可以按照下面的对应关系直接给出其对偶规划。

- 将模型统一为“ $\max \leq$ ”或“ $\min \geq$ ”的形式
- 若原规划的某个约束条件为等式约束，则在对偶规划中与此约束对应的那个变量取值没有非负限制
- 若原规划的某个变量的值没有非负限制，则在对偶问题中与此变量对应的那个约束为等式

原问题与对偶问题的对应关系

原问题（对偶问题）		对偶问题（原问题）	
min		max	
变量	n 个变量	约束	n 个约束
	变量 ≥ 0		约束 \leq
	变量 ≤ 0		约束 \geq
	无正负限制		约束 $=$
约束	m 个约束	变量	m 个变量
	约束 \leq		变量 ≤ 0
	约束 \geq		变量 ≥ 0
	约束 $=$		无正负限制
约束条件右端项		目标函数中的变量系数	
目标函数中的变量系数		约束条件右端项	

表 2.5: 原问题与对偶问题的对应关系

2.2.3.2 对偶定理

设有一对互为对偶的线性规划

$$(P) \begin{cases} \max & z = \mathbf{c}^T \mathbf{x} \\ s.t. & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{cases} \quad (D) \begin{cases} \min & f = \mathbf{b}^T \mathbf{y} \\ s.t. & \mathbf{A}^T \mathbf{y} \geq \mathbf{c} \\ & \mathbf{y} \geq \mathbf{0} \end{cases}$$

定理 2.1 若 \mathbf{x} 和 \mathbf{y} 分别为原规划 (P) 和 (D) 对偶规划，则

$$\mathbf{c}^T \mathbf{x} \leq \mathbf{b}^T \mathbf{y}$$

定理 2.2 设 \mathbf{x} 和 \mathbf{y} 分别为原规划 (P) 和 (D) 的可行解，当 $\mathbf{c}^T \mathbf{x} = \mathbf{b}^T \mathbf{y}$ 时， \mathbf{x}, \mathbf{y} 分别是两个问题的最优解

定理 2.3 若规划 (P) 有可行解，则规划 (P) 有最优解的充分必要条件是规划 (D) 有可行解

定理 2.4 若规划 (D) 有可行解，则规划 (D) 有最优解的充分必要条件是规划 (P) 有可行解

定理 2.5 若原规划 (P) 有最优解，则对偶规划 (D) 也有最优解，反之亦然，且两者的目标函数值相等。

2.2.3.3 对偶单纯形法

对偶单纯形法是求解原规划的一种方法。

原理：

作为原规划的一个解，会有两个性质等待满足：可行性和最优性。而原问题的可行性和最优性恰好对应对偶问题的最优性和可行性。

单纯形法的思路是，先满足可行性，再逐渐逼近最优性；而对偶单纯形法的思路是，先找到最优性，再逐渐逼近可行性。

也就是说，先找到对偶问题的可行解，再找到原问题的可行解（即对偶问题的最优解）。

最优性：看检验数 σ_j 可行性：看右端项 b_i

从原规划的一个基本解出发，此基本解不一定可行，但它对应着一个对偶可行解；就是说可以从一个对偶可行解出发，然后检验原规划的基本解是否可行，即是否有负的分量。如果有负的分量，则进行迭代，求另一个基本解，此基本解对应着另一个对偶可行解（检验数非正）；而得到的基本解的分量皆非负，则该基本解为最优解。

也就是说，对偶单纯形法在迭代过程中始终保持对偶解的可行性（检验数非正），使得原规划的基本解由不可行变为可行，当同时得到对偶规划与原规划的可行解时，得到原规划的最优解。

STEPS:

1. 根据线性规划典式形式，建立初始单纯形表（就是还按照单纯形法填初始单纯形表）。此表对应原规划的一个基本解。表要求：检验数数行各元素一定非正，原规划的基本解可以有小于零的分量。
2. 若基本解的所有分量皆非负，则得到原规划的最优解，停止计算；若基本解中有小于零的分量 b_i ，并且 b_i 所在行各系数 $a_{ij} \geq 0$ ，则原规划无可行解，停止计算；若 $b_i < 0$ ，并且存在 $a_{ir} < 0$ ，则确定 x_r 为出基变量，并计算

$$\theta = \min \left\{ \frac{\sigma_j}{a_{rj}} \mid a_{rj} < 0 \right\} = \frac{\sigma_k}{a_{rk}}$$

确定 x_k 为进基向量。若有多个 $b_i < 0$ ，则选择最小的进行分析计算。

3. 以 b_{rk} 为中心元素，按照与单纯形法类似的方法，在表中进行迭代计算，返回第 2 步。

2.2.4 灵敏度分析

前提：在求灵敏度分析和影子价格中，常常会遇到要使用对偶问题的最优解（或最优基 \mathbf{B} ）求解问题，这里解释一下怎么使用最优单纯形表求得。

1. 首先利用单纯形法得到最优单纯形表
2. 得到 b 列的 n 个数值，最优单纯形表从后往前数 n 列（就是松弛变量对应的列），有矩阵：

$$\begin{pmatrix} a_{1(m-n)} & a_{1(m-n+1)} & \cdots & a_{1m} \\ a_{2(m-n)} & a_{2(m-n+1)} & \cdots & a_{2m} \\ \vdots & & & \vdots \\ a_{n(m-n)} & a_{n(m-n+1)} & \cdots & a_{nm} \\ \hline \sigma_{m-n} & \sigma_{m-n} & \cdots & \sigma_{m-n} \end{pmatrix}$$

其中：

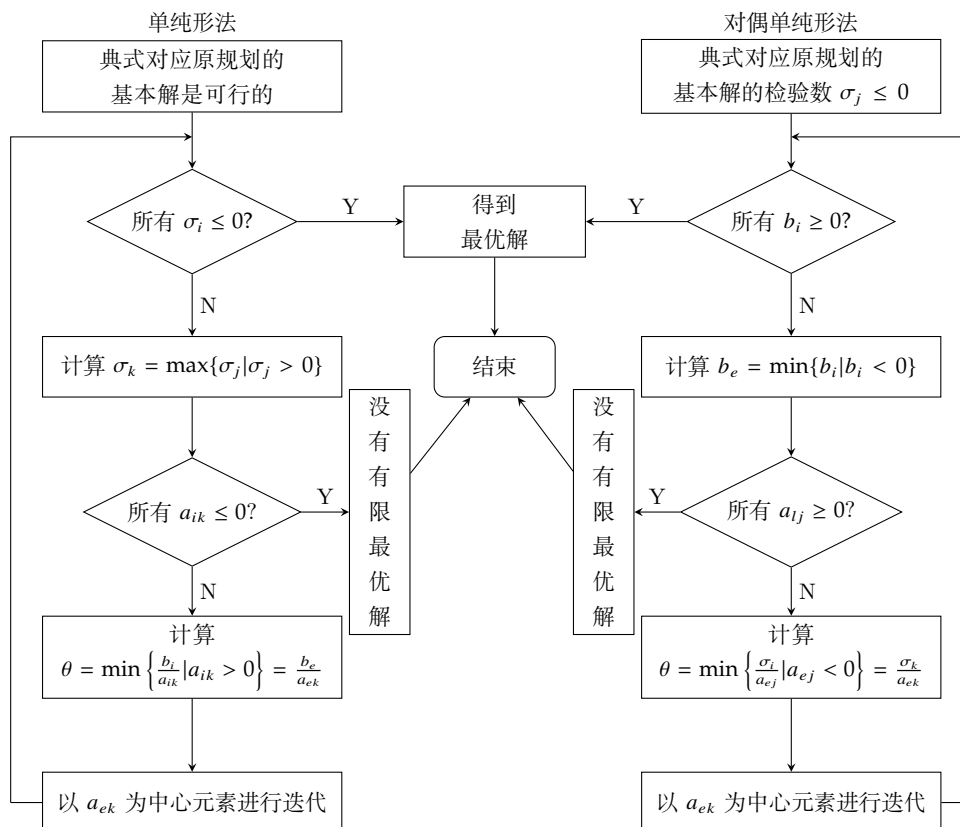


图 2.2: 单纯形法与对偶单纯形法流程图

- 最后一行的检验数 $\sigma^T = -\mathbf{c}_B^T \mathbf{B}^{-1}$, 各检验数取相反数 (即 $\mathbf{c}_B^T \mathbf{B}^{-1}$) 即为对偶问题的最优解。
- 除最后一行外 (即线上方的) 矩阵即为 \mathbf{B}^{-1}

2.2.4.1 影子价格

影子价格是一个向量, 它的分量表示最优目标值随相应资源数量变化的变化率。

若 x^*, y^* 分别为 (LP) 和 (DP) 的最优解, 那么有

$$\mathbf{c}^T \mathbf{x}^* = \mathbf{b}^T \mathbf{y}^*$$

根据 $f = \mathbf{b}^T \mathbf{y}^* = b_1 y_1^*, b_2 y_2^*, \dots, b_m y_m^*$ 可知

$$\frac{\partial f}{\partial b_i} = y_i^*$$

y_i^* 表示 b_i 变化一个单位对目标 f 产生的影响, 称 y_i^* 为 b_i 的影子价格

需要指出, 影子价格不是固定不变的, 当约束条件、产品利润等发生变化时, 有可能使影子价格发生变化。另外, 影子价格的经济含义, 是指资源在一定范围内增加时的情况, 当某种资源的增加超过了这个“一定的范围”时, 总利润的增加量则不是按照影子价格给出的数值线性地增加。

如何求解影子价格:

1. 求出对偶问题的最优解 (可以利用单纯形表, 在求出最优解的情况下, 将松弛变量对应的各检验数取负 $-\sigma_i$ 就能得到对偶问题的最优解)

2. 对偶问题最优解中的数字依次对应的就是原问题的各资源影子价格

经济意义：每增加一单位的某资源，最终收益增加多少单位

判断资源是否有剩余：
$$\begin{cases} y_i^* = 0 & \text{有剩余} \\ y_i^* > 0 & \text{无剩余} \end{cases}$$

2.2.4.2 目标函数系数 c 变化*

若只有一个系数 c_j 变化，其他系数不变。 c_j 的变化只影响检验数 σ_j ，而不影响解的非负性。

$$\sigma_j = c_j - \mathbf{c}_B^T \mathbf{B}^{-1} \mathbf{p}_j, j = 1, 2, \dots, n$$

1. c_k 是非基变量的系数

非基系数变化只影响与 c_k 有关的一个检验数 σ_k 的变化，对其他无影响，故只需要考虑 σ_k 。

设 $c_k \rightarrow \bar{c}_k = c_k + \Delta c_k$ ，有 σ_k 的变化：

$$\bar{\sigma}_k = c_k + \Delta c_k - \mathbf{c}_B^T \mathbf{B}^{-1} \mathbf{p}_j = \sigma_k + \Delta c_k$$

为了保持最优解不变， σ_k 必须满足 $\bar{\sigma}_k = \sigma_k + \Delta c_k \leq 0$ 。也就是说：

$$\Delta c_k \leq -\sigma_k, \bar{c}_k = c_k + \Delta c_k \leq c_k - \sigma_k \quad (2.3)$$

$c_k - \sigma_k$ 是 c_k 变化的上限，若 c_k 不超出上限，最优解不变；否则，将最优单纯形表中的检验数 σ_k 用 $\bar{\sigma}_k$ 取代，取 x_k 为进基变量，继续单纯形的表格计算。

2. c_l 是基变量的系数

设 $c_l \rightarrow c_l + \Delta c_l$ ，引入 $\Delta c = (0, \dots, 0, \Delta c_l, 0, \dots, 0)$ ，有

$$\begin{aligned} \sigma_j \rightarrow \bar{\sigma}_j &= c_j - [\mathbf{c}_B^T + (\Delta c)^T] \mathbf{B}^{-1} \mathbf{p}_j, j \neq l \\ &= \sigma_j - \Delta c_l a'_{lj} \end{aligned}$$

(注意：上式中 a 的 l 与左边基变量的下标对应) 为保证最优解不变， Δc_l 要满足

$$\max \left\{ \frac{\sigma_j}{a'_{lj}} | a'_{lj} > 0 \right\} \leq \Delta c_l \leq \min \left\{ \frac{\sigma_j}{a'_{lj}} | a'_{lj} < 0 \right\}$$

若 Δc_l 超出此范围，应求出新的检验数 $\bar{\sigma}_j$ ，选择其中大于零的检验数对应的变量 x_j 为进基变量，继续迭代。

2.2.4.3 右端常数 b 变化*

b_r 的变化影响解的可行性，但不影响检验数的符号变化。由 $\mathbf{x}_B = \mathbf{B}^{-1} \mathbf{b}$ 可知 b_r 的变化必会引起最优解数值变化。

最优解的变化分为以下两类：

1. 保持 $\mathbf{B}^{-1} \mathbf{b} \geq \mathbf{0}$ ，即最优基 \mathbf{B} 不变（影子价格不变，也就是对偶问题的最优解不变）
只需要将变化后的 b_r 带入 $\mathbf{B}^{-1} \mathbf{b}$ 的表达式重新计算即可
2. $\mathbf{B}^{-1} \mathbf{b}$ 出现负分量，这使最优基 \mathbf{B} 变化
需要通过迭代求解新的最优基和最优解

综合一下，可以利用下面的步骤计算：

设 $b_r \rightarrow \bar{b}_r = b_r + \Delta b_r$, Δb_r , 此时有

$$\mathbf{x}_B \rightarrow \bar{\mathbf{x}}_B = \mathbf{B}^{-1} \begin{pmatrix} b_1 \\ \vdots \\ b_r + \Delta b_r \\ \vdots \\ b_m \end{pmatrix} = \mathbf{B}^{-1} \mathbf{b} + \mathbf{B}^{-1} \begin{pmatrix} 0 \\ \vdots \\ \Delta b_r \\ \vdots \\ 0 \end{pmatrix} = \mathbf{x}_B + \Delta b_r \begin{pmatrix} \beta_{1r} \\ \vdots \\ \beta_{mr} \end{pmatrix} = \begin{pmatrix} b'_1 \\ \vdots \\ b'_m \end{pmatrix} + \Delta b_r \begin{pmatrix} \beta_{1r} \\ \vdots \\ \beta_{mr} \end{pmatrix} \geq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

(上式只需要随便算其中的一个，然后看是否大于等于 0 就行，不需要特别算下面的 Δb_r)

其中，为 $\mathbf{x}_B^{-1} \mathbf{b}$ 原最优解， b'_i 为 $\mathbf{x}_B = \mathbf{B}^{-1} \mathbf{b}$ 的第 i 个分量， β_{ir} 为 \mathbf{B}^{-1} 的第 i 行第 r 列元素要满足

$$\max \left\{ \frac{-b'_i}{\beta_{ir}} | \beta_{ir} > 0 \right\} \leq \Delta b_r \leq \min \left\{ \frac{-b'_i}{\beta_{ir}} | \beta_{ir} < 0 \right\}$$

当 Δb_r 超过此范围时，将使最优解中某个分量小于零，使最优基发生变化。此时可用对偶单纯形法继续迭代新的最优解。

就是说，如果没超范围，直接把 $\bar{\mathbf{x}}_B$ 当最优解得出就行；如果超范围的话，就把 $\bar{\mathbf{x}}_B$ 放到最优单纯形表的 b 列，然后求对偶单纯形。

2.2.4.4 约束条件系数 a 变化

假设只有一个 a_{ij} 变化，其他数据不变，且只讨论 a_{ij} 为非基变量 x_j 的系数的情况。那么此时 a_{ij} 的变化只影响一个检验数 σ_j 。

设 $a_{ij} \rightarrow a_{ij} + \Delta a_{ij}$ ，由检验数的另一种表示形式

$$\sigma_j \rightarrow \bar{\sigma}_j = c_j - \mathbf{y}^T \begin{pmatrix} a_{1j} \\ \vdots \\ a_{ij} + \Delta a_{ij} \\ \vdots \\ a_{mj} \end{pmatrix} = c_j - \mathbf{y}^T \mathbf{p}_j - \mathbf{y}^T \begin{pmatrix} 0 \\ \vdots \\ \Delta a_{ij} \\ \vdots \\ 0 \end{pmatrix} = \sigma_j - y_i^* \Delta a_{ij}$$

其中， \mathbf{y} 为对偶最优解， y_i^* 为 \mathbf{y} 的第 i 个变量

为使最优解不变，要使 $\sigma_j \leq 0$ ，即

$$\begin{aligned} \sigma_j &\leq y_i^* \Delta a_{ij} \\ \Delta a_{ij} &\geq \frac{\sigma_j}{y_i^*}, y_i^* > 0 \\ \Delta a_{ij} &\leq \frac{\sigma_j}{y_i^*}, y_i^* < 0 \end{aligned}$$

2.2.4.5 新增变量 x 分析

增加变量 x_{n+1} ，则有相应的约束条件 \mathbf{p}_{n+1} ，目标函数系数 c_{n+1} ，那么，计算出

$$\sigma_{n+1} = c_{n+1} - \mathbf{c}_B^T \mathbf{B}^{-1} \mathbf{p}_{n+1}$$

填入最优单纯形表，若 $\sigma_{n+1} \leq 0$ ，则最优解不变；否则，进一步用单纯形法求解

2.2.4.6 新增约束条件

增加一个约束之后，应把最优解代入新的约束，若满足，则最优解不变；否则，填入最优单纯形表作为新的一行，引入一个新的非负变量（原约束若是小于等于形式，可引入非负松弛变量；否则，引入非负人工变量），并通过矩阵行变换把对应基变量的元素变为 0，进一步用单纯形法或对偶单纯形法求解。

Chapter 3

最优化搜索算法的结构和一维搜索

从此章开始就是求非线性规划问题，与线性规划问题可以在有限步数内得到解不同，非线性规划问题不一定可以在有限的迭代步数内收敛。

3.1 常用的搜索算法结构

3.1.1 收敛性概念

在规划问题的求解过程中，由于迭代算法是以产生一系列迭代点为目的的，因此算法的收敛性表现在产生的点列 $\{\mathbf{x}^{(k)}\}$ 上。

1. 理想的收敛性概念：设 \mathbf{x}^* 为问题 (fS) 的全局最优解。当 $\mathbf{x}^* \in \{\mathbf{x}^{(k)}\}$ 或 $\mathbf{x}^{(k)} \neq \mathbf{x}^*$ 对于所有 k 成立，并且满足 $\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \mathbf{x}^*$ 时，称算法收敛到最优解 \mathbf{x}^* 。
2. 然而，由于实际问题较为复杂，很难达到理想状态，因此需要引入一些实用的收敛性概念：

首先定义解集 Ω ，是具有某种可接受性质的点集，通常取下面的几种集合：

- (a) $\Omega = \{\mathbf{x}^* | \mathbf{x}^* \text{ is } g.opt\}$
- (b) $\Omega = \{\mathbf{x}^* | \mathbf{x}^* \text{ is } l.opt\}$
- (c) $\Omega = \{\mathbf{x}^* | \mathbf{x}^* \text{ 满足某种最优条件, 或者说 } \nabla f(\mathbf{x}^*) = 0\}$
- (d) $\Omega = \{\mathbf{x}^* | \mathbf{x}^* \in S, f(\mathbf{x}^*) \leq B\}$ ，其中 B 为可接受的目标函数值的一个上界

设算法产生的点列为 $\mathbf{x}^{(k)}$ ，满足下列任一情况时，称算法收敛：

- (a) $\{\mathbf{x}^{(k)}\} \cap \Omega \neq \emptyset$
- (b) $\{\mathbf{x}^{(k)}\}$ 的任意收敛子列的极限点属于 Ω

求解问题时，需要考虑初始点的影响，因此引入下面的概念：

1. 全局收敛性：若算法对任意的初始点 $\mathbf{x}^{(1)}$ 都收敛，则称算法全局收敛。
2. 局部收敛性：若算法只有限制初始点 $\mathbf{x}^{(1)}$ 到解集 Ω 附近（当 Ω 为非连通时，指在 Ω 的某个连通子集附近）时，才有收敛性，则称算法局部收敛。

3.1.2 收敛准则 (停机条件)

1. $\|\mathbf{x}^{(k+m)} - \mathbf{x}^{(k)}\| < \varepsilon$ (最常用的是 $m = 1$)
2. $\frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|}{\|\mathbf{x}^{(k)}\|} < \varepsilon$
3. $|f(\mathbf{x}^{(k+1)}) - f(\mathbf{x}^{(k)})| < \varepsilon$

3.1.3 收敛速度

设 $\{\mathbf{x}^{(k)}\}$ 为算法产生的点列, 收敛于解 \mathbf{x}^* , 且 $\mathbf{x}^{(k)} \neq \mathbf{x}^*$, 有 $\forall k$

1. 线性收敛: $\exists \alpha \in (0, 1)$ $\frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|} \leq \alpha$, k 充分大时 $\mathbf{x}^{(k)}$ 收敛于 \mathbf{x}^*
表示每迭代一次, 比上一次更接近 \mathbf{x}^* ; 收敛的阶数 $p_0 = 1$
2. 超线性收敛: $\lim_{k \rightarrow \infty} \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|} = 0$
迭代速度比线性收敛更快, 收敛的阶数 $p_0 \geq 1$
3. 二阶收敛: $\exists \alpha > 0$, 使得 k 充分大时, 有 $\frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^2} \leq \alpha$
收敛的阶数 $p_0 = 2$, 说明二阶收敛是超线性收敛的一个特例

3.1.4 二次终结性

定义 3.1 一个算法用于解正定二次函数的无约束极小时, 有限步迭代可达最优解, 则称该算法具有二次终结性。

二次终结性 = 共轭方向 + 精确一维搜索

3.1.4.1 共轭方向

定义 3.2 设 $A_{n \times n}$ 对称正定, $d^{(1)}, d^{(2)} \in R^n$, $d^{(1)} \neq 0, d^{(2)} \neq 0$, 满足 $d^{(1)T} A d^{(2)} = 0$, 称 $d^{(1)}, d^{(2)}$ 关于矩阵 A 共轭。

3.1.5 下降算法模型

考虑规范形式的规划问题

$$(fS) \quad \begin{cases} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in S \end{cases}$$

希望能更接近目标函数的最优解, 即对 $\forall k, f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$, 这样的算法称为下降算法。

首先要找到下降可行的方向, 即能够接近最优解的方向。

设 $\bar{\mathbf{x}} \in S$, 对方向 \mathbf{d} , 如果 $\exists \delta > 0$, 使

$$f(\mathbf{x} + \lambda \mathbf{d}) < f(\bar{\mathbf{x}}), \forall \lambda \in (0, \delta) \quad (3.1)$$

那么称 \mathbf{d} 为 $f(\mathbf{x})$ 在点 $\bar{\mathbf{x}}$ 的下降方向; 就是说在某个点上, 沿着这个方向移动可以朝着更优的解靠近。

然而, 对于问题 (fS) , 只找到下降方向, 有时不能产生新的迭代点, 还要求 \mathbf{d} 为可行方向 (换句话说, 沿着可行方向前进, 可以保持在可行解的范围内), 即 $\exists \delta > 0$, 使

$$\mathbf{x} + \lambda \mathbf{d} \in S, \forall \lambda \in (0, \delta) \quad (3.2)$$

如果方向 \mathbf{D} 同时满足式 (3.1) 和 (3.2), 那么称 \mathbf{d} 为问题 (fS) 的下降可行方向。

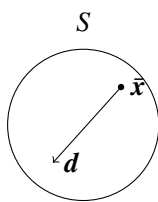


图 3.1: 下降可行方向

下降算法模型如图3.2所示, 接下来先要利用一维搜索找到步长因子 λ_k

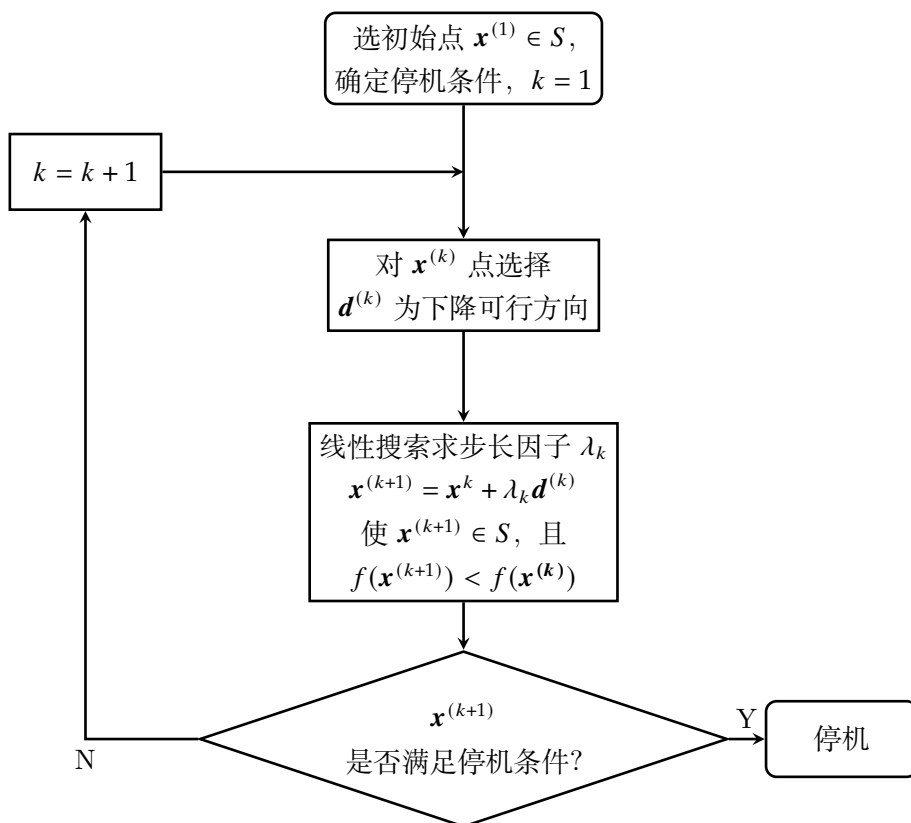


图 3.2: 下降算法模型

3.2 一维搜索

一元函数求最小值和线性搜索均为一维搜索, 用于求步长因子 λ_k 即:

$$\begin{aligned} \min \quad & f(\mathbf{x}^{(k)} + \lambda \mathbf{d}^{(k)}) = \varphi(\lambda) \\ \text{s.t.} \quad & \lambda \in S \end{aligned}$$

这里 S 有三种情况: $(-\infty, +\infty)$, $(0, +\infty)$, $[a, b]$

3.2.1 缩小区间的精确一维搜索

考虑问题 P

$$\begin{aligned} \min \quad & \varphi(\lambda) \\ \text{s.t.} \quad & \lambda \in [\alpha, \beta] \\ & \varphi: \mathbf{R} \rightarrow \mathbf{R} \end{aligned}$$

3.2.1.1 不确定区间和单峰函数

不确定区间：表示 $[\alpha, \beta]$ 含最优解，即 $\varphi(\lambda)$ 的最小值，但是不知道具体位置

定义 3.3 设函数 $\varphi: \mathbf{R} \rightarrow \mathbf{R}$, $[\alpha, \beta]$ 为不确定区间, λ^* 为 φ 在 $[\alpha, \beta]$ 上的最小值点, 如果 $\forall \lambda_1, \lambda_2 \in [\alpha, \beta], \lambda_1 < \lambda_2$ 满足:

1. 当 $\lambda_2 \leq \lambda^*$ 时, $\varphi(\lambda_1) > \varphi(\lambda_2)$
2. 当 $\lambda_1 \geq \lambda^*$ 时, $\varphi(\lambda_1) < \varphi(\lambda_2)$

则称 $\varphi(\lambda)$ 在 $[\alpha, \beta]$ 上为是强单峰的。

如果再加上条件, 当 $\varphi(\lambda_1) \neq \varphi(\lambda^*), \varphi(\lambda_2) \neq \varphi(\lambda^*)$ 时, 上式成立, 则称 $\varphi(\lambda)$ 在 $[\alpha, \beta]$ 上为是单峰的。

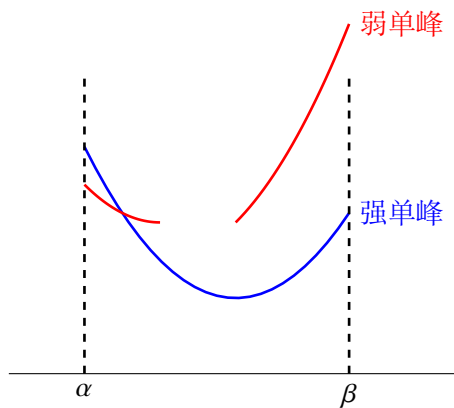


图 3.3: 强单峰与单峰

单峰函数的特点是在最优解 λ^* 的某一侧, 距离 λ^* 远的点的函数值必定不小于距离 λ^* 近的点的函数值。根据该特点有下面的定理:

定理 3.1 设 $\varphi(\lambda): \mathbf{R} \rightarrow \mathbf{R}$, 在区间 $[\alpha, \beta]$ 上单峰。 $\lambda, \mu \in [\alpha, \beta]$, 且 $\lambda < \mu$, 则

1. 若 $\varphi(\lambda) > \varphi(\mu)$, 则 $\varphi(\xi) \geq \varphi(\mu), \forall \xi \in [\alpha, \lambda]$
2. 若 $\varphi(\lambda) < \varphi(\mu)$, 则 $\varphi(\xi) \geq \varphi(\mu), \forall \xi \in [\mu, \beta]$
3. 若 $\varphi(\lambda) = \varphi(\mu)$, 则 $\varphi(\xi) \geq \varphi(\mu), \forall \xi \in [\alpha, \lambda] \cup [\mu, \beta]$

ξ 所在区间去掉后不会失去最优解, 因此可以用缩小区间的方法求解最优解。也就是说, 只需要在 $[\alpha, \beta]$ 上取两个点 λ, μ , 比较函数值 $\varphi(\lambda), \varphi(\mu)$, 可以去掉 ξ 所在区间。

3.2.1.2 进退法求出初始不确定区间

实际计算时，往往不知道最优解的位置，因此需要先找到一个初始的不确定区间，基本思想是找三个点，使两端点的函数值比中间点的函数值大。

先任取点 λ_0 ，选取步长 $\delta > 0$ 取 $\lambda_1 = \lambda_0 + \delta$ ，那么有两种情况

1. 若 $\varphi(\lambda_0) < \varphi(\lambda_1)$ ，向左找，取 $\lambda_2 = \lambda_0 - 2\delta$ ，再考察 $\varphi(\lambda_2)$ ：

(a) 若 $\varphi(\lambda_0) \leq \varphi(\lambda_2)$ ，则停，得到不确定区间 $[\lambda_2, \lambda_1]$

(b) 否则，取 $\lambda_1 = \lambda_0, \lambda_0 = \lambda_2, \delta = 2\delta$ ，重复该过程。

2. 若 $\varphi(\lambda_0) \geq \varphi(\lambda_1)$ ，向右找，取 $\lambda_2 = \lambda_0 + 2\delta$ ，再考察 $\varphi(\lambda_2)$ ：

(a) 若 $\varphi(\lambda_1) \leq \varphi(\lambda_2)$ ，则停，得到不确定区间 $[\lambda_0, \lambda_2]$

(b) 否则，取 $\lambda_0 = \lambda_1, \lambda_1 = \lambda_2, \delta = 2\delta$ ，重复该过程。

如图3.4与3.5所示

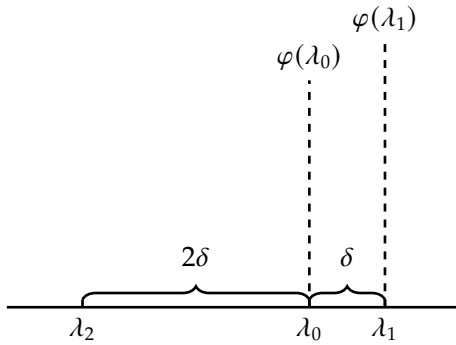


图 3.4: 进退法情况 1

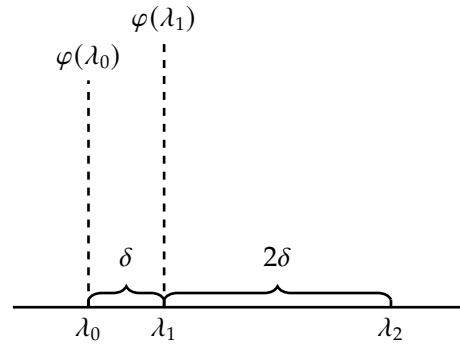


图 3.5: 进退法情况 2

需要注意的是，步长 δ 的选择要适度，太大时含有多个单峰区间，太小时迭代次数过多；另外，当 $\varphi(\lambda)$ 单调时，进退法无结果，因此需要加入迭代次数的限制。

3.2.1.3 中点法

若 $\varphi(\lambda) : \mathbf{R} \rightarrow \mathbf{R}$ 可微，且 $\varphi(\lambda)$ 在 $[\alpha, \beta]$ 上单峰，那么可以用中点法求解最优解。

设 $\varphi(\lambda)$ 在单峰区间 $[\alpha, \beta]$ 可微，且不存在拐点（即 $\varphi'(\lambda)$ 及 $\varphi''(\lambda)$ 同时为零的点），取 $\lambda = \frac{\alpha+\beta}{2}$ ，有中点法一般步骤：

1. 若 $\varphi'(\lambda) = 0$ ，则 λ 为最优解

2. 若 $\varphi'(\lambda) > 0$ ，则 λ 在上升段， $\lambda^* < \lambda$ ，去掉 $(\lambda, \beta]$ ，新区间为 $[\alpha, \lambda]$

3. 若 $\varphi'(\lambda) < 0$ ，则 λ 在下降段， $\lambda^* > \lambda$ ，去掉 $[\alpha, \lambda)$ ，新区间为 $[\lambda, \beta]$

利用中点法进行 k 次迭代后, 存在精度区间 $[\alpha^*, \beta^*] < \varepsilon$, 也就是说有

$$\begin{aligned}\frac{\beta - \alpha}{2^k} &< \varepsilon \\ \frac{\beta - \alpha}{\varepsilon} &< 2^k \\ \ln \frac{\beta - \alpha}{\varepsilon} &< k \ln 2 \\ k &> \frac{\ln \frac{\beta - \alpha}{\varepsilon}}{\ln 2}\end{aligned}$$

需要注意的是, 中点法只能用于单峰函数, 就是说对函数的可微性有要求。

3.2.1.4 黄金分割法

黄金分割法是一种比较经典的缩小区间的方法, 要求考虑条件:

1. 对称性 $\lambda - \alpha = \beta - \mu$
2. 缩减比不变 $\gamma = \frac{\text{保留区间的长度}}{\text{原区间的长度}}$

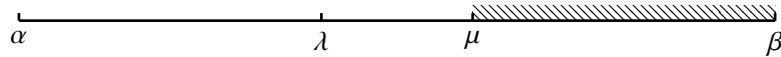


图 3.6: 黄金分割法示意图

不妨设迭代选择的新的两个点为 λ, μ , 通过比较函数值去掉 $[\mu, \beta]$ (或 $[\alpha, \mu]$) 段, 如图3.6, 那么有

$$\begin{aligned}\frac{\mu - \alpha}{\beta - \alpha} &= \frac{\lambda - \alpha}{\mu - \alpha} = \gamma \\ \mu &= \alpha + \gamma(\beta - \alpha), \lambda = \alpha + (1 - \gamma)(\beta - \alpha) \\ \gamma^2 + \gamma - 1 &= 0 \\ \gamma &= \frac{\sqrt{5} - 1}{2} \approx 0.618\end{aligned}$$

下图为黄金分割法的流程图

3.2.2 牛顿法和插值法

3.2.2.1 牛顿法

牛顿法的基本思想是通过迭代逼近一个函数的最小值点。具体来说, 该方法使用函数的二阶导数 (也称为 Hessian 矩阵) 来构建一个二次逼近, 然后通过解这个二次逼近的极小值问题来更新当前估计值。主要用于已经知道 λ_k 求 λ^* 。

设函数 $\varphi(\lambda)$ 二次可微, 考虑 $\varphi(\lambda)$ 在 λ_k 处的泰勒展开

$$\varphi(\lambda) = \varphi(\lambda_k) + \varphi'(\lambda_k)(\lambda - \lambda_k) + \frac{1}{2}\varphi''(\lambda_k)(\lambda - \lambda_k)^2 + o(\lambda - \lambda_k)^2$$

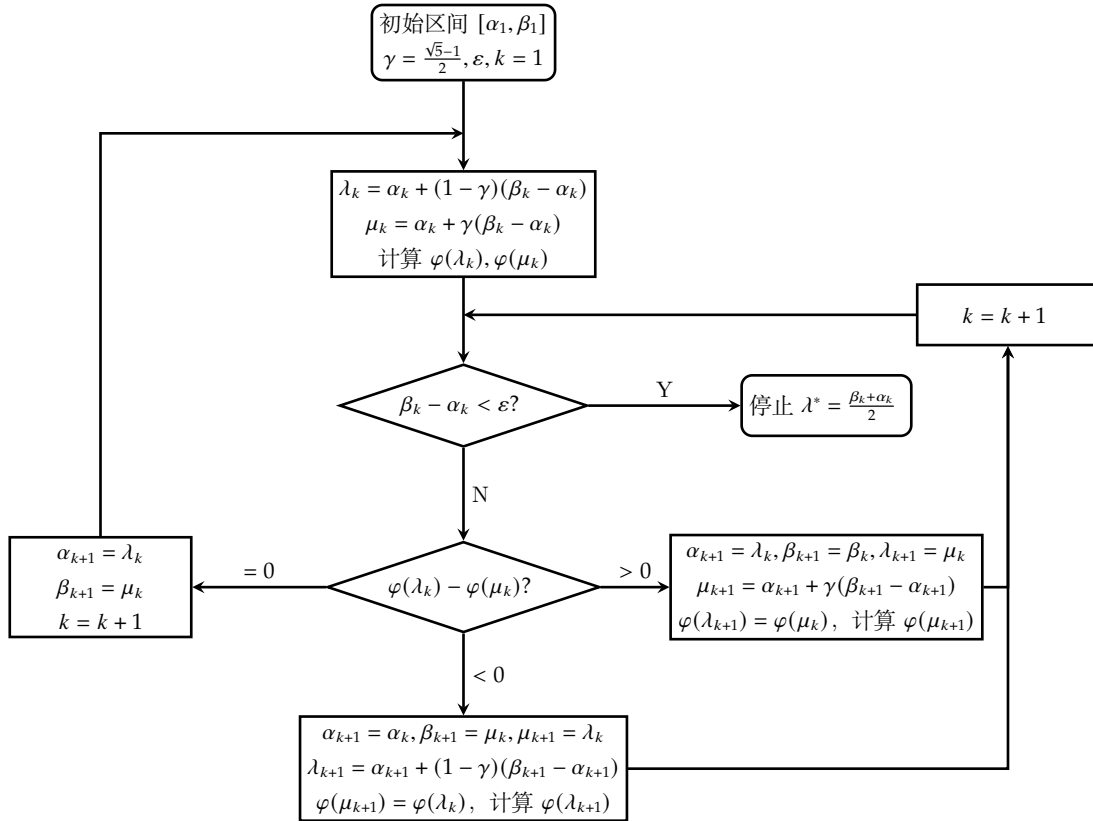


图 3.7: 黄金分割法流程

舍去高阶项，得到二次式

$$q(\lambda) = \varphi(\lambda_k) + \varphi'(\lambda_k)(\lambda - \lambda_k) + \frac{1}{2}\varphi''(\lambda_k)(\lambda - \lambda_k)^2 \quad (3.3)$$

取 $q(\lambda)$ 作为 $\varphi(\lambda)$ 在 λ_k 邻域的近视，当 $\varphi''(\lambda_k) > 0$ 时， $q(\lambda)$ 的驻点即为极小（即为凸函数），作为新的迭代点。对式 3.3 求导，并令其为零

$$q'(\lambda) = \varphi'(\lambda_k) + \varphi''(\lambda_k)(\lambda - \lambda_k) = 0$$

当 $\varphi''(\lambda_k) > 0$ 时，解之得到牛顿迭代公式 ★

$$\lambda_{k+1} = \lambda_k - \frac{\varphi'(\lambda_k)}{\varphi''(\lambda_k)} \quad (3.4)$$

牛顿法的主要特点是二阶收敛，但是局部收敛。而且它的收敛速度很快。

3.2.2.2 插值法

插值法的基本思想是利用插值函数逼近所需求解的目标函数，然后求出插值函数的极小值点，作为新的迭代点。常见的有三点二次插值、二点二次插值和三次插值多项式

（感觉不是很重要，具体见课本）

设 $\varphi(\lambda) : \mathbf{R} \rightarrow \mathbf{R}$

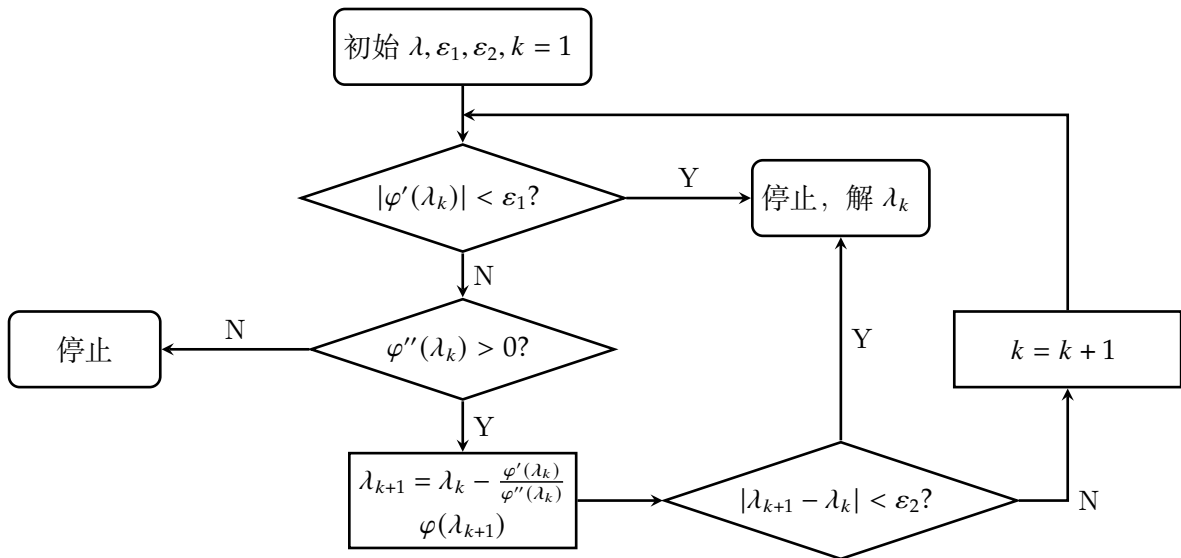


图 3.8: 牛顿法流程

1. 三点二次插值。取 $\lambda_0, \lambda_1, \lambda_2$ ，得到各函数值 $\varphi_0 = \varphi(\lambda_0), \varphi_1 = \varphi(\lambda_1), \varphi_2 = \varphi(\lambda_2)$ ，构造二次插值多项式： $\psi(\lambda) = a\lambda^2 + b\lambda + c$ 有

$$\begin{cases} a\lambda_0^2 + b\lambda_0 + c = \varphi_0 \\ a\lambda_1^2 + b\lambda_1 + c = \varphi_1 \\ a\lambda_2^2 + b\lambda_2 + c = \varphi_2 \end{cases}$$

解得

$$\begin{cases} a = -\frac{(\lambda_0 - \lambda_1)\varphi_2 + (\lambda_1 - \lambda_2)\varphi_0 + (\lambda_2 - \lambda_0)\varphi_1}{(\lambda_0 - \lambda_1)(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_0)} \\ b = \frac{(\lambda_0^2 - \lambda_1^2)\varphi_2 + (\lambda_1^2 - \lambda_2^2)\varphi_0 + (\lambda_2^2 - \lambda_0^2)\varphi_1}{(\lambda_0 - \lambda_1)(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_0)} \end{cases}$$

有极小值点为

$$\bar{\lambda} = -\frac{b}{2a}$$

2. 二点二次插值。
3. 三次插值。

3.2.3 不精确的一维搜索

在实际搜索中，往往不需要特别精确，只要满足一定的条件即可，这样可以减少计算量。

3.2.3.1 Goldstein 法

设 (fS) ，在 $\mathbf{x}^{(k)}$ 取方向 $\mathbf{d}^{(k)}$ ，有 $f^T(\mathbf{x}^{(k)})\mathbf{d}^{(k)} < 0$ ，求解下面的问题

$$(L-R) \begin{cases} \min & f(\mathbf{x}^{(k)} + \lambda \mathbf{d}^{(k)}) \\ s.t. & \lambda \geq 0 \end{cases}$$

那么有 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)}$ ，这里令 $\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x} = \lambda_k \mathbf{d}^{(k)}$ 。

Goldstein 法遵循下面的规则：

$$f(\mathbf{x}^{(k+1)}) - f(\mathbf{x}^{(k)}) \leq \rho \nabla f^T(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} s^{(k)} \quad (3.5)$$

$$f(\mathbf{x}^{(k+1)}) - f(\mathbf{x}^{(k)}) \geq (1 - \rho) \nabla f^T(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} s^{(k)} \quad (3.6)$$

其中, $\rho \in (0, 1/2)$, 实际经验中常取 $\rho = 0.1$ 或更小。

这里式3.5表示 λ_k 取值应使 $f(\mathbf{x}^{(k+1)})$ 的值在直线 $y = f(\mathbf{x}^{(k)}) + \rho \nabla f^T(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \lambda$ 下方; 式 3.6 表示 λ_k 取值应使 $f(\mathbf{x}^{(k+1)})$ 的值在直线 $y = f(\mathbf{x}^{(k)}) + (1 - \rho) \nabla f^T(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \lambda$ 上方。

几何意义如图3.9所示, 其中 $\nabla f^T(\mathbf{x}^{(k)}) s^{(k)} = \nabla f^T(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \lambda_k$ 。

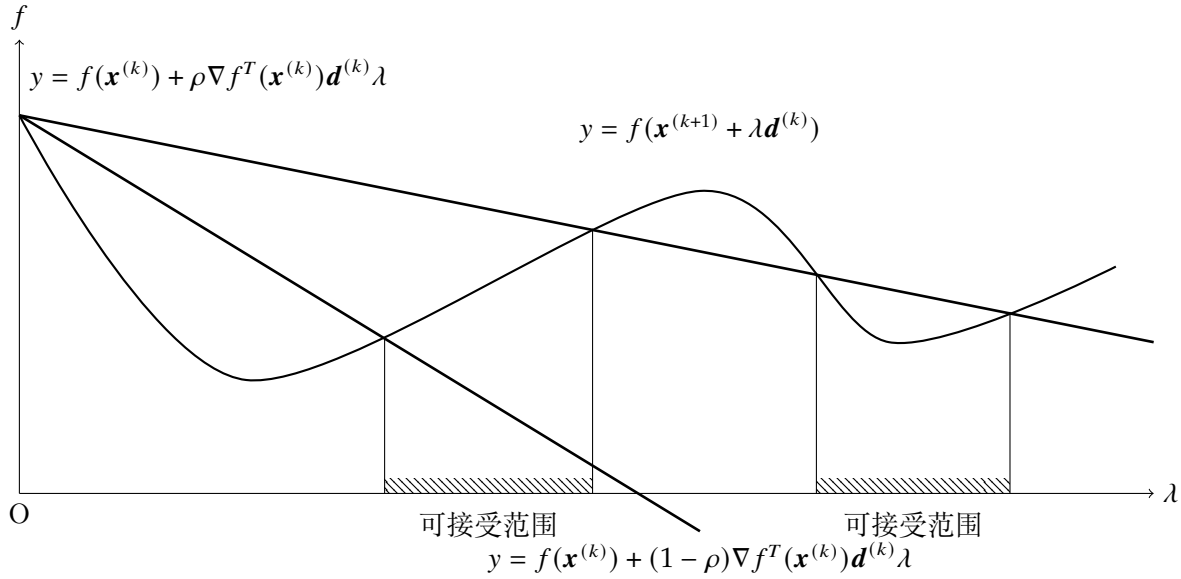


图 3.9: Goldstein 法几何意义

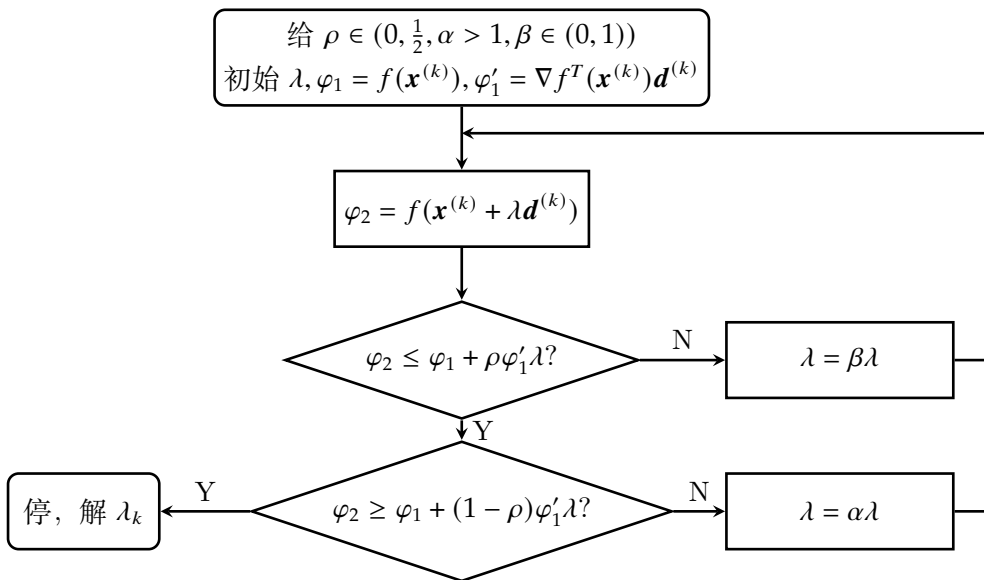


图 3.10: Goldstein 法流程

3.2.3.2 Wolfe-Powell 法

考虑从 $\mathbf{x}^{(k)}$ ，沿方向 $\mathbf{d}^{(k)}$ 搜索新的迭代点 $\mathbf{x}^{(k)} + \lambda \mathbf{d}^{(k)}$

有规则 Wolfe-Powell：

$$f(\mathbf{x}^{(k+1)}) \leq f(\mathbf{x}^{(k)}) + \rho \nabla f^T(\mathbf{x}^{(k)}) \mathbf{s}^{(k)} \quad (3.7)$$

$$\nabla f(\mathbf{x}^{(k+1)}) \mathbf{s}^{(k)} \geq \sigma \nabla f^T(\mathbf{x}^{(k)}) \mathbf{s}^{(k)} \quad (3.8)$$

其中, $\rho \in (0, 1/2), \sigma \in (\rho, 1)$, 实际经验中常取 $\rho = 0.1, \sigma = 0.7$ 或附近。

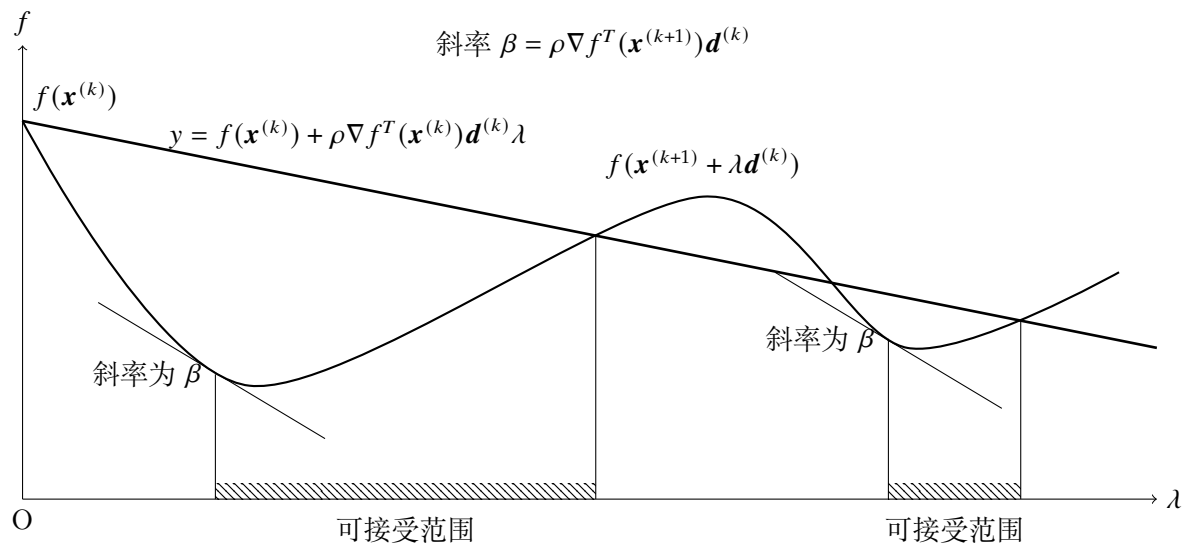


图 3.11: Wolfe-Powell 法几何意义

Chapter 4

无约束最优化方法

无约束最优化问题的，是指问题 (f, S) 中 $S = \mathbf{R}^n$ 的情况。为了简便，记无约束最优化问题为

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

其中, $f: \mathbf{R} \rightarrow \mathbf{R}$

4.1 最优性条件

定理 4.1 (极小点的一阶必要条件) 设 $f(\mathbf{x})$ 可微, \mathbf{x}^* 为 $f(\mathbf{x})$ 的局部极小点, 则 \mathbf{x}^* 为驻点, 即 $\nabla f(\mathbf{x}^*) = 0$

定理 4.2 (极小点的二阶必要条件) 设 $f(\mathbf{x})$ 二次连续可微, \mathbf{x}^* 为 $f(\mathbf{x})$ 的局部极小点, 则 $\nabla f(\mathbf{x}^*) = 0$, 且二阶黑塞矩阵 $\nabla^2 f(\mathbf{x}^*)$ 半正定。

定理 4.3 (极小点的二阶充分条件) 设 $f(\mathbf{x})$ 二次连续可微, $\mathbf{x}^* \in \mathbf{R}^n$, 如果 $\nabla f(\mathbf{x}^*) = 0$, 且二阶黑塞矩阵 $\nabla^2 f(\mathbf{x}^*)$ 正定, 则 \mathbf{x}^* 为 $f(\mathbf{x})$ 的局部极小点。

定理 4.4 当 f 凸时, \mathbf{x}^* 为 $f(\mathbf{x})$ 的局部极小点 $\Leftrightarrow \nabla f(\mathbf{x}) = 0$

注意: $f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla^T f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*), \forall \mathbf{x}$, 故 $f(\mathbf{x}^*) \leq f(\mathbf{x})$

4.2 最速下降法

最速下降法是无约束最优化中最简单也是最古老而基本的方法。

在前面的下降算法中, 方向去负梯度方向 $\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})$ 进行精确一维搜索, 即得到最速下降法。

- 方向: 在迭代点 $\mathbf{x}^{(k)}$ 取方向 $\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})$
- 步长: 精确一维搜索, 即求 λ_k , 使得

$$f(\mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)}) = \min_{\lambda \in \mathbf{R}} f(\mathbf{x}^{(k)} + \lambda \mathbf{d}^{(k)})$$

- 更新: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)}$

定理 4.5 设 $f(\mathbf{x})$ 可微, $\nabla f(\mathbf{x}^*) \neq 0$ 那么, $\mathbf{d}^{(k)} = -\frac{\nabla f(\mathbf{x}^{(k)})}{\|\nabla f(\mathbf{x}^{(k)})\|}$ 是 f 在 $\mathbf{x}^{(k)}$ 处的最速下降方向。

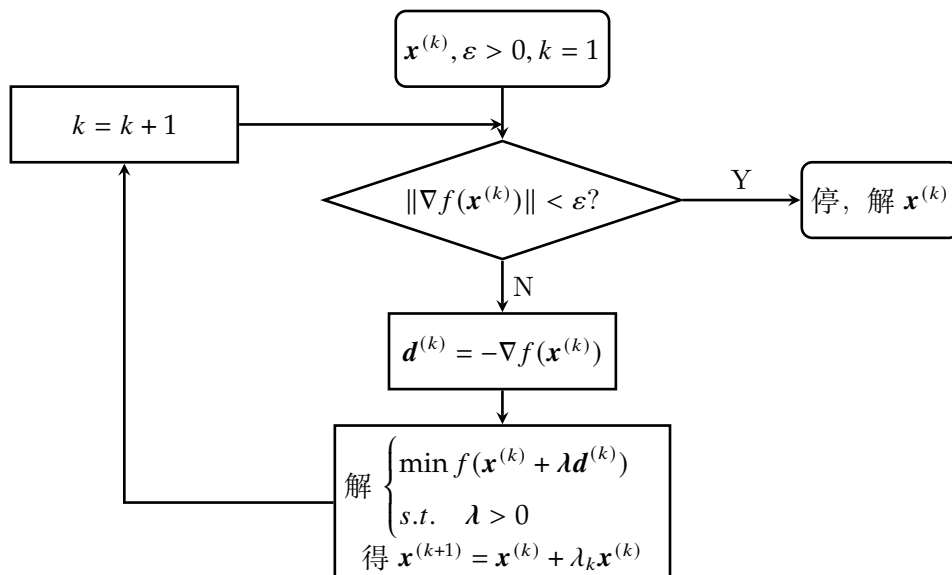


图 4.1: 最速下降法流程

如图4.1所示，即为最速下降法的流程图。

特点：全局收敛，线性收敛，易产生扭摆现象而造成早停（当 $x(k)$ 距最优点较远时，速度快，而接近最优点时，速度下降）

4.3 牛顿法及其修正

4.3.1 牛顿法

牛顿法的基本思想是通过迭代逼近一个函数的最小值点。具体来说，该方法使用函数的二阶导数（也称为 Hessian 矩阵）来构建一个二次逼近，然后通过解这个二次逼近的极小值来作为新的迭代点。

设 $f(x)$ 二次连续可微，考虑 $f(x)$ 在 $x^{(k)}$ 处的二阶 Taylor 近似函数：

$$q^{(k)}(x) = f(x^{(k)}) + \nabla f^T(x^{(k)})(x - x^{(k)}) + \frac{1}{2}(x - x^{(k)})^T \nabla^2 f(x^{(k)})(x - x^{(k)})$$

求驻点，即使 $\nabla q^{(k)}(x) = \nabla f(x^{(k)}) + \nabla^2 f(x^{(k)})(x - x^{(k)}) = 0$ ，那么当 $\nabla^2 f(x^{(k)})$ 正定，有极小点

$$x^{(k+1)} = x^{(k)} - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

在实际计算中， $s^{(k)} = x^{(k+1)} - x^{(k)}$ 得计算常常通过解方程

$$\nabla^2 f(x^{(k)})s^{(k)} = -\nabla f(x^{(k)})$$

得到。

牛顿法流程图如图4.2所示。

特点：二阶收敛，局部收敛。（当 $x^{(k)}$ 充分接近 x^* 时，局部函数可用正定二次函数很好地近似，故收敛很快）

二次终结性：当 $f(x)$ 为正定二次函数时，从任意初始点可一步迭代达到最优解。

主要缺点：

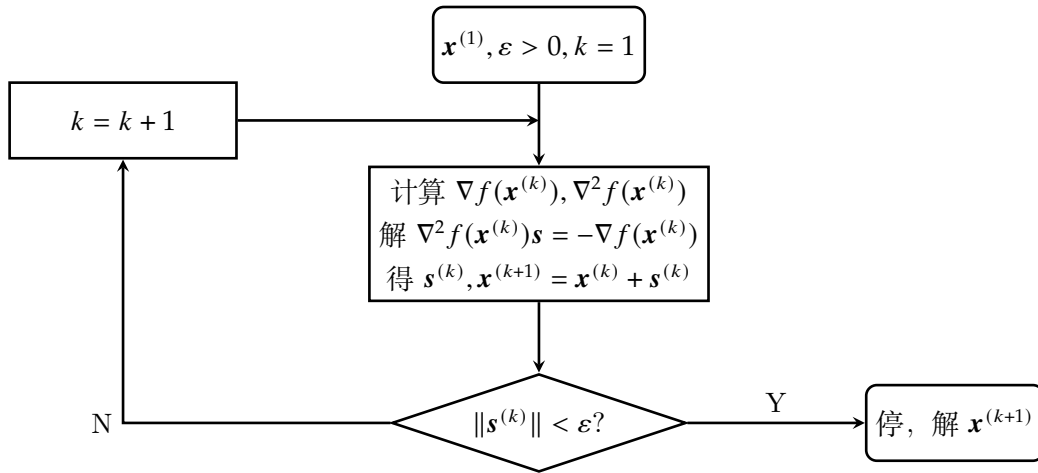


图 4.2: 牛顿法流程

1. 局部收敛，初始点选择不当往往会导致不收敛；
2. 用到二阶 Hesse 矩阵，且要求正定；
3. 需计算 Hesse 矩阵逆或解 n 阶线性方程组，计算量大。

但是牛顿法二阶收敛是非常好的性质，所以牛顿法的修正针对上面的缺点进行。

4.3.2 减少计算量的直接修正

不需要每次迭代都计算 $\nabla^2 f(\mathbf{x}^{(k)})$ ，而是把 m 次迭代作为一组，每一组用同一个 Hesse 矩阵，那么有公式

$$\mathbf{x}^{(km+j+1)} = \mathbf{x}^{(km+j)} - [\nabla^2 f(\mathbf{x}^{(km)})]^{-1} \nabla f(\mathbf{x}^{(km+j)})$$

该算法的收敛速度为 $(m+1)^{1/m}$ 阶，当 m 越大，收敛速度越慢。当 $m = 1$ 时，即为牛顿法，当 $m \rightarrow \infty$ 时即为线性收敛。

4.3.3 带一维线性搜索的牛顿算法

也称阻尼牛顿法，阻尼因子 λ_k

1. 利用 $\nabla^2 f(\mathbf{x}^{(x)}) \mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})$ 求出 $\mathbf{d}^{(k)} = -[\nabla^2 f(\mathbf{x}^{(x)})]^{-1} \nabla f(\mathbf{x}^{(k)})$
2. 解 $\min f(\mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)})$
3. 得到 $\lambda_k, \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)}$

特点：可改善局部收敛性，当 $\mathbf{d}^{(k)}$ 为函数上升方向时，可向负方向搜索，但可能出现 $\pm \mathbf{d}^{(k)}$ 均非下降方向的情况

4.3.4 Goldstein-Price 方法

将牛顿方向同最速下降方向结合起来, 在一维搜索中采用 Goldstein 不精确一维搜索:

假设已有 $\mathbf{x}^{(k)}, \nabla f(\mathbf{x}^{(k)}), \nabla^2 f(\mathbf{x}^{(k)})$, 则有方向

$$\mathbf{d}^{(k)} = \begin{cases} -[\nabla^2 f(\mathbf{x}^{(k)})]^{-1} \nabla f(\mathbf{x}^{(k)}), & \nabla^2 f(\mathbf{x}^{(k)}) \text{ 正定} \\ -\nabla f(\mathbf{x}^{(k)}), & \text{其它} \end{cases} \quad (4.1)$$

再进行非精确一维搜索, $\lambda = 1$ 优先, $\rho \in (0, \frac{1}{2})$

1. $f(\mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)}) \leq f(\mathbf{x}^{(k)}) + \rho \nabla f^T(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \lambda_k$
2. $f(\mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)}) \geq f(\mathbf{x}^{(k)}) + (1 - \rho) \nabla f^T(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \lambda_k$

特点: 在一定条件下, G-P 法全局收敛。但当 $\nabla^2 f(\mathbf{x}^{(k)})$ 非正定情况较多时, 收敛速度降为接近线性。

4.3.5 L-M 法

用 $[\nabla^2 f(\mathbf{x}^{(k)}) + \mu E]$ 取代 $\nabla^2 f(\mathbf{x}^{(k)})$ 进行迭代, 其中 E 为单位矩阵。 $\mu > 0$ 使 $[\nabla^2 f(\mathbf{x}^{(k)}) + \mu E]$ 正定, μ 尽量小。

特点: 全局二阶收敛。

4.4 共轭梯度法

共轭梯度法是介于最速下降法与牛顿法之间的一个方法, 它仅需利用一阶导数信息, 但克服了最速下降法收敛慢的缺点, 又避免了牛顿法需要存储和计算 Hesse 矩阵并求逆的缺点。

共轭梯度法不仅是解大型线性方程组最有用的方法之一, 也是解大型非线性最优化问题最有效的算法之一。

共轭梯度法是一个典型的共轭方向 (定义 3.2) 法, 它的每一个搜索方向是互相共轭的。思路: 搜索方向 $\mathbf{d}^{(k)}$ 仅仅是负梯度方向 $-\mathbf{g}_k$ 与上一次迭代的搜索方向 $\mathbf{d}^{(k-1)}$ 的组合。设

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x} - \mathbf{b}^T \mathbf{x} + c$$

其中 \mathbf{G} 是 $n \times n$ 对称正定矩阵, \mathbf{b} 是 $n \times 1$ 向量, f 的梯度为 $\nabla f(\mathbf{x}) = \mathbf{G} \mathbf{x} + \mathbf{b}$

4.4.1 共轭梯度法的方向

设 $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x} - \mathbf{b}^T \mathbf{x} + c$ 其中 $\mathbf{G}_{n \times n}$ 对称正定, $\mathbf{b} \in \mathbf{R}$, 从最速下降方向开始构造一组共轭方向 (虽然但是步骤并不完整, 仍然需要结合课本):

1. 设初始点 $\mathbf{x}^{(1)}$, 取 $\mathbf{d}^{(1)} = -\nabla f(\mathbf{x}^{(1)})$ (最速下降方向)
2. 设 $k \geq 1$, 已得到 k 个相互共轭的方向 $\mathbf{d}^{(1)}, \mathbf{d}^{(2)}, \dots, \mathbf{d}^{(k)}$, 以及由 $\mathbf{x}^{(1)}$ 开始依次沿上述方向精确一维搜索得到点 $\mathbf{x}^{(2)}, \dots, \mathbf{x}^{(k)}, \mathbf{x}^{(k+1)}$ 。即有下式:

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} + \alpha_i \mathbf{d}^{(i)}, i = 1, 2, \dots, k$$

3. 精确一维搜索保证方向导数为零:

$$\nabla f^T(\mathbf{x}^{(i+1)})\mathbf{d}^{(i)} = 0, i = 1, 2, \dots, k \quad (4.2)$$

4. 在 $\mathbf{x}^{(i+1)}$ 点构造新方向 $\mathbf{d}^{(k+1)}$ 为 $-\nabla f(\mathbf{x}^{(k+1)})$ 与 $\mathbf{d}^{(1)}, \mathbf{d}^{(2)}, \dots, \mathbf{d}^{(k)}$ 的组合:

$$\mathbf{d}^{(k+1)} = -\nabla f(\mathbf{x}^{(k+1)}) + \sum_{j=1}^k \beta_j^{(k)} \mathbf{d}^{(j)}$$

5. 使得 $\mathbf{d}^{(k+1)}$ 与 $\mathbf{d}^{(1)}, \mathbf{d}^{(2)}, \dots, \mathbf{d}^{(k)}$ 都共轭:

$$\mathbf{d}^{(k+1)T} \mathbf{G} \mathbf{d}^{(j)} = 0, j = 1, 2, \dots, k \quad (4.3)$$

6. 下面利用 Gram-Schmidt 过程使其共轭化: $(i, j = 1, 2, \dots, k)$ 记

$$\mathbf{y}^{(j)} = \nabla f(\mathbf{x}^{(j+1)}) - \nabla f(\mathbf{x}^{(j)}) \quad (4.4)$$

$$= \mathbf{G}(\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}) \quad (4.5)$$

$$= \alpha_j \mathbf{G} \mathbf{d}^{(j)} \quad (4.6)$$

7. 由式4.6得到

$$\mathbf{d}^{(i)T} \mathbf{y}^{(j)} = \alpha_j \mathbf{d}^{(i)T} \mathbf{G} \mathbf{d}^{(j)} = 0 \quad (4.7)$$

8. 根据第四到六步的式子, 可得

$$\mathbf{d}^{(k+1)T} \mathbf{y}^{(j)} = \alpha_j \mathbf{d}^{(k+1)T} \mathbf{G} \mathbf{d}^{(j)} = 0, j = 1, 2, \dots, k \quad (4.8)$$

9. 注意, 根据构造式4.3, 有

$$\nabla f(\mathbf{x}^{(j)}) = -\mathbf{d}^{(j)} + \sum_{i=1}^{j-1} \beta_i^{(j-1)} \mathbf{d}^{(i)} \quad (4.9)$$

10. 又根据式4.6, 4.7, 4.2可得, 对于 $\forall j \leq k, i < j$ 有

$$\nabla f^T(\mathbf{x}^{(j+1)})\mathbf{d}^{(i)} = (\nabla f(\mathbf{x}^{(j+1)}) + \sum_{l=i+1}^j \mathbf{y}^{(l)})^T \mathbf{d}^{(i)} = 0 \quad (4.10)$$

11. 根据4.9, 对 $\forall j \leq k, i < j$ 有

$$\nabla f^T(\mathbf{x}^{(j+1)})\nabla f(\mathbf{x}^{(i)}) = 0 \quad (4.11)$$

12. 由4.3, 4.6得到, 当 $j = 1, 2, \dots, k-1$ 时

$$-\nabla f^T(\mathbf{x}^{(k+1)})[\nabla f(\mathbf{x}^{(j+1)}) - \nabla f(\mathbf{x}^{(j)})] + \beta_j^{(k)} \mathbf{d}^{(j)T} \mathbf{y}^{(j)} = 0 \quad (4.12)$$

13. 最终当 $j = k$ 时

$$\beta_k = \frac{\nabla f^T(\mathbf{x}^{(k+1)})[\nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})]}{\mathbf{d}^{(k)T} \mathbf{y}^{(k)}} \quad (4.13)$$

并有等价形式

$$\beta_k = \frac{\nabla f^T(\mathbf{x}^{(k+1)})\nabla f(\mathbf{x}^{(k+1)})}{\nabla f^T(\mathbf{x}^{(k)})\nabla f(\mathbf{x}^{(k)})} \quad \text{FR 法} \quad (4.14)$$

$$\beta_k = \frac{\nabla f^T(\mathbf{x}^{(k+1)})[\nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})]}{\nabla f^T(\mathbf{x}^{(k)})\nabla f(\mathbf{x}^{(k+1)})} \quad \text{PRP 法} \quad (4.15)$$

$$\beta_k = \frac{\nabla f^T(\mathbf{x}^{(k+1)})\nabla f(\mathbf{x}^{(k+1)})}{-\nabla f^T(\mathbf{x}^{(k)})\mathbf{d}^{(k)}} \quad \text{共轭下降法} \quad (4.16)$$

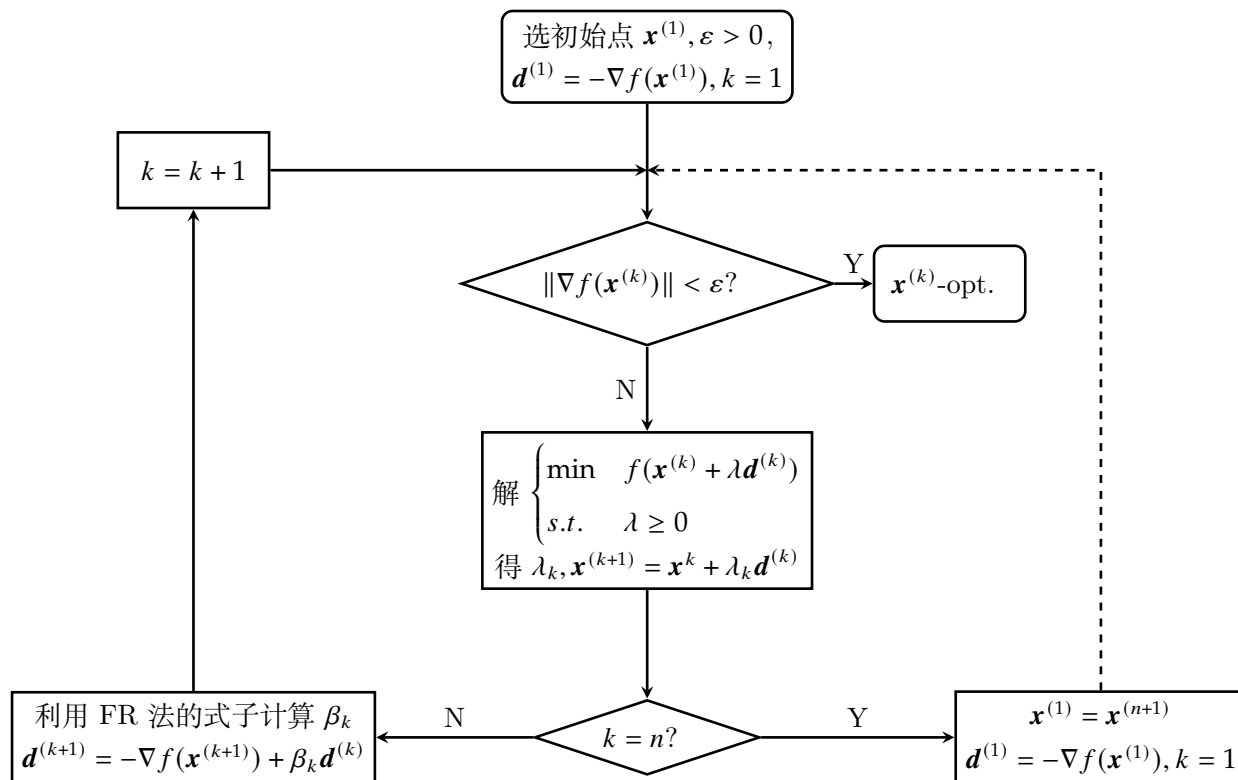


图 4.3: 共轭梯度法流程

图4.3为共轭梯度法的流程图。其中实线为对正定二次函数的FR方法，若加入虚线部分，即为一般函数的再开始的方法。

4.4.2 再开始共轭梯度法

再开始技术：对于一般非二次函数，共轭梯度法常常采用再开始技术，即每 n 步以后周期性地采用最速下降方向作为新的搜索方向从而使方法有较快的收敛速度。

再开始策略：每隔 n 步迭代再开始；还有当 $g_k^T d_k > 0$ ，即 d_k 是上升方向时再开始；对于二次函数，相邻二次迭代的梯度互相直交，因此，如果它们偏离直交性较大，则进行再开始。

精确算法步长计算：

确定搜索方向后，代入目标函数 $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{G}\mathbf{x} - \mathbf{b}^T \mathbf{x} + c$ 得到

$$f(\mathbf{x}_k + \lambda \mathbf{d}_k) = \frac{1}{2}(\mathbf{x}_k + \lambda \mathbf{d}_k)^T \mathbf{G}(\mathbf{x}_k + \lambda \mathbf{d}_k) - \mathbf{b}^T (\mathbf{x}_k + \lambda \mathbf{d}_k) + c$$

关于 λ 求导得

$$f'_\lambda(\mathbf{x}_k + \lambda \mathbf{d}_k) = \mathbf{d}_k^T (\mathbf{G}\mathbf{x}_k + \mathbf{b}) + \lambda \mathbf{d}_k^T \mathbf{G} \mathbf{d}_k$$

令其为 0，解得

$$\lambda = -\frac{\mathbf{d}_k^T (\mathbf{G}\mathbf{x}_k + \mathbf{b})}{\mathbf{d}_k^T \mathbf{G} \mathbf{d}_k} = -\frac{\mathbf{d}_k^T \nabla f(\mathbf{x}_k)}{\mathbf{d}_k^T \mathbf{G} \mathbf{d}_k}$$

4.4.3 算法特点

- 全局收敛，线性收敛，共轭梯度方向即为下降方向

- 每步迭代只需储存若干向量（适用于大规模问题）
- 有二次总结性：对于正定二次函数，最多迭代 n 次即可达 opt.

需要注意的是，不同的 β_k 公式对于正定二次函数的效果是相同的；对于非正定二次函数，经验上 PRP 法效果最好，FR 法次之，共轭下降法最差。

4.5 变尺度法

4.5.1 基本思路

变尺度法从广义上仍可看作是牛顿法的改进。基本想法是用对称正定矩阵近似二阶黑塞矩阵，或黑塞矩阵的逆。是某种意义上的近似，要求满足拟牛顿方程 (4.5.1.1)，因此这类算法又称拟牛顿法。

基本思想为用对称正定矩阵 $\mathbf{H}^{(k)}$ 近似 $\nabla^2 f(\mathbf{x}^{(k)})$ ，而 $\mathbf{H}^{(k)}$ 的产生从给定 $\mathbf{H}^{(1)}$ 开始逐步修正得到。变尺度法的一般模型如图4.4所示。

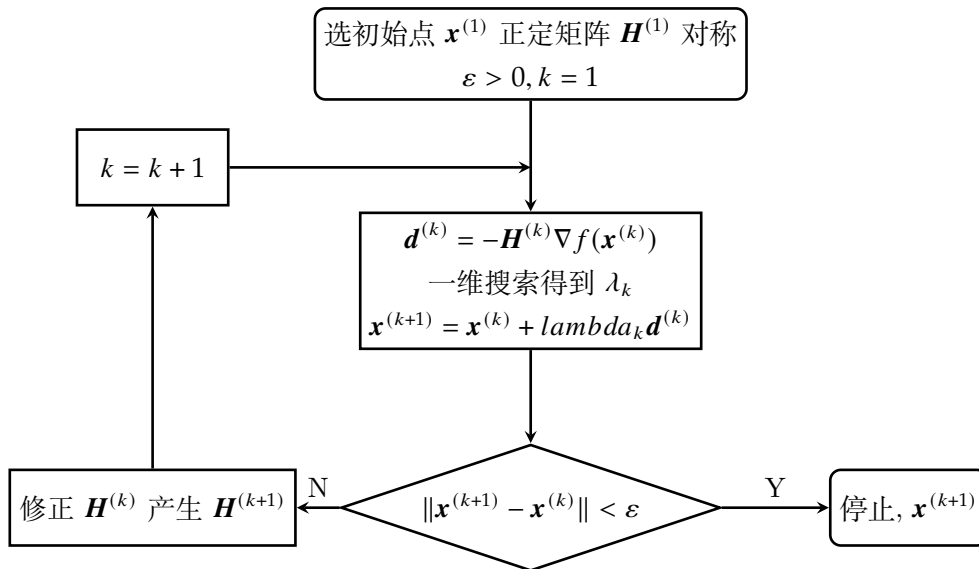


图 4.4: 变尺度法流程

4.5.1.1 拟牛顿方程

设

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x} - \mathbf{b}^T \mathbf{x} + c$$

$$\nabla f = \mathbf{G} \mathbf{x} + \mathbf{b}$$

那么记 $\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$, $\mathbf{y}^{(k)} = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)}) = \mathbf{G} \mathbf{s}^{(k)}$

称 $\mathbf{H} \mathbf{y} = \mathbf{S}$ 或 $\mathbf{y} = \mathbf{G} \mathbf{S}$ 为拟牛顿方程。

显然，当 \mathbf{H} 正定时， $\mathbf{G}^{-1} = \mathbf{H}$

对于 $f(\mathbf{x})$ 的二阶收敛 Taylor 展开，舍去高阶项，有

$$\mathbf{y}^{(k)} \approx \nabla^2 f^T(\mathbf{x}^{(k)}) \mathbf{s}^{(k)} \quad \text{或} \quad \mathbf{s}^{(k)} \approx [\nabla^2 f^T(\mathbf{x}^{(k)})]^{-1} \mathbf{y}^{(k)}$$

用矩阵 $\mathbf{B}^{(k)}$ 或 $\mathbf{H}^{(k)}$ 分别取代 $\nabla^2 f(\mathbf{x}^{(k)})$ 或者 $(\nabla^2 f(\mathbf{x}^{(k)}))^{-1}$ 使拟 Newton 方程成立, 可看做是对 $\nabla^2 f(\mathbf{x}^{(k)})$ 或 $(\nabla^2 f(\mathbf{x}^{(k)}))^{-1}$ 的一种近似。此种近似 \mathbf{H} 或 \mathbf{B} 不唯一。

4.5.2 变尺度法的特点

- 只需用到函数的一阶梯度 (Newton 法用到二阶 Hesse 矩阵);
- 下降算法, 故全局收敛;
- 不需求矩阵逆 (计算量小);
- 一般可达到超线性收敛 (速度快);
- 有二次终结性。

Chapter 5

约束最优化方法

本章问题的一般形式为

$$(fgh) \begin{cases} \min & f(\mathbf{x}) \\ s.t. & g_i(\mathbf{x}) \leq 0, i = 1, 2, \dots, m \\ & h_j(\mathbf{x}) = 0, j = 1, 2, \dots, l \end{cases}$$

或为了便捷写成矩阵形式

$$(fgh) \begin{cases} \min & f(\mathbf{x}) \\ s.t. & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{cases}$$

5.1 Kuhn-Tucker 条件

5.1.1 等式约束的最优性条件

只考虑含有等式约束的问题：

$$(fh) \begin{cases} \min & f(\mathbf{x}) \\ s.t. & \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{cases} \quad (5.1)$$

这是一个条件极值问题，其最优性条件由 Lagrange 函数的一阶必要条件给出。

高等数学中，条件极值为：在 $\Phi(x, y) = 0$ 的条件下，求 $z = f(x, y)$ 的极值，可以对应为

$$\begin{cases} \min & f(x, y) \\ s.t. & \Phi(x, y) = 0 \end{cases}$$

于是可以建立 Lagrange 函数

$$L(\mathbf{x}, \mathbf{y}; \boldsymbol{\lambda}) = f(\mathbf{x}, \mathbf{y}) + \boldsymbol{\lambda} \Phi(\mathbf{x}, \mathbf{y}) \quad (5.2)$$

其中， $\boldsymbol{\lambda}$ 为拉格朗日乘子向量， λ 为拉格朗日乘子。

若 $(\mathbf{x}^*, \mathbf{y}^*)$ 是条件极值, 则存在 λ^* 使得

$$\begin{cases} f_x(\mathbf{x}^*, \mathbf{y}^*) + \lambda^* \Phi_x(\mathbf{x}^*, \mathbf{y}^*) = 0 \\ f_y(\mathbf{x}^*, \mathbf{y}^*) + \lambda^* \Phi_y(\mathbf{x}^*, \mathbf{y}^*) = 0 \\ \Phi(\mathbf{x}^*, \mathbf{y}^*) = 0 \end{cases}$$

推广到多个等式约束, 可得到对于 (f, \mathbf{h}) 的情况: 对于问题 5.1, 可以建立 Lagrange 函数

$$L(\mathbf{x}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{v}^T \mathbf{h}(\mathbf{x}) \quad (5.3)$$

若 \mathbf{x}^* 是问题 5.1 的极小点, 则存在 \mathbf{v}^* 使得

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^l \mathbf{v}_i^* \nabla h_i(\mathbf{x}^*) = \mathbf{0} \quad (5.4)$$

矩阵形式:

$$\nabla f(\mathbf{x}^*) + \frac{\partial \mathbf{h}(\mathbf{x}^*)}{\partial \mathbf{x}} \mathbf{v}^* = \mathbf{0} \quad (5.5)$$

最优性条件即

$$\nabla f(\mathbf{x}^*) = - \sum_{i=1}^l \mathbf{v}_i^* \nabla h_i(\mathbf{x}^*) \quad (5.6)$$

5.1.2 不等式约束问题的 Kuhn-Tucker 条件

考虑含有不等式约束的问题:

$$(f, \mathbf{g}) \begin{cases} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \end{cases} \quad (5.7)$$

设 $\mathbf{x}^* \in S = \{\mathbf{x} | \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$, 记

$$I = \{i | g_i(\mathbf{x}^*) = 0, i = 1, 2, \dots, m\}$$

为 \mathbf{x}^* 点的起作用集或紧约束集或有效集。

如果 \mathbf{x}^* 是 l.opt., 对每一个约束函数来说, 只有当它是起作用约束时, 才产生影响。

对于不等式约束问题有如下最优性定理, 它的结论称为 Kuhn-Tucker 条件。

定理 5.1 (最优性必要条件; K-T 条件) 对于问题 (f, \mathbf{g}) , 设 $\mathbf{x}^* \in S$, I 为 \mathbf{x}^* 点的紧约束集。设函数 $f(\mathbf{x}), g_i(\mathbf{x}), i \in I$ 在 \mathbf{x}^* 点连续可微, $g_i(\mathbf{x}), i \in I$ 在 \mathbf{x}^* 点连续。向量组 $\{\nabla g_i(\mathbf{x}^*), i \in I\}$ 线性无关。如果 \mathbf{x}^* 是 (f, \mathbf{g}) 的 l.opt., 则存在 $\mathbf{u}_i^* \geq 0, i \in I$ 使得

$$\nabla f(\mathbf{x}^*) + \sum_{i \in I} \mathbf{u}_i^* \nabla g_i(\mathbf{x}^*) = \mathbf{0} \quad (5.8)$$

如果进一步有 $g_i(\mathbf{x}), i \notin I$ 在 \mathbf{x}^* 点也可微, 那么下列系统成立

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \mathbf{u}_i^* \nabla g_i(\mathbf{x}^*) = \mathbf{0} \quad (5.9)$$

$$\mathbf{u}^* \geq \mathbf{0} \quad (5.10)$$

$$\mathbf{u}^{*T} \mathbf{g}(\mathbf{x}^*) = 0 \text{ (互补松弛条件)} \quad (5.11)$$

称满足 K-T 条件的 \mathbf{x}^* 为 K-T 点。

5.1.3 一般约束问题的 Kuhn-Tucker 条件

考虑一般约束问题 (fgh) , 结合定理 5.1, 可以平行得到下面的定理。

定理 5.2 考虑问题 (fgh) , $\mathbf{x}^* \in S = \{\mathbf{x} | \mathbf{g}(\mathbf{x}) \leq 0, \mathbf{h}(\mathbf{x}) = 0\}$, I 为作用集。

设 $g_i(\mathbf{x}) (i \in I)$ 在 \mathbf{x}^* 点可微, $g_i(\mathbf{x}) (i \notin I)$ 在 \mathbf{x}^* 连续, $h_j (j = 1, 2, \dots, l)$ 在 \mathbf{x}^* 的某邻域内连续可微。再设约束规格 $(CQ)\{\nabla g_i(\mathbf{x}^*), i \in I, \nabla h_j(\mathbf{x}^*), j = 1, 2, \dots, l\}$ 线性无关。那么存在 $u_i \geq 0, i \in I, v_j, j = 1, 2, \dots, l$ 使得

$$\nabla f(\mathbf{x}^*) + \sum_{i \in I} u_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^l v_j^* \nabla h_j(\mathbf{x}^*) = \mathbf{0} \quad (5.12)$$

如果进一步有 $g_i(\mathbf{x}), i \notin I$ 在 \mathbf{x}^* 点也可微, 那么下列系统成立

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m u_i^* \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^l v_j^* \nabla h_j(\mathbf{x}^*) = \mathbf{0} \quad (5.13)$$

$$\mathbf{u} \geq \mathbf{0} \quad (5.14)$$

$$\mathbf{u}^T \mathbf{g}(\mathbf{x}^*) = 0 \text{ (互补松弛条件)} \quad (5.15)$$

5.2 既约梯度法

一种把约束变化为标准形式的线性约束问题的算法。问题为

$$(fAb) = \begin{cases} \min & f(\mathbf{x}) \\ s.t. & \mathbf{Ax} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{cases} \quad (5.16)$$

其中 \mathbf{A} 为 $m \times n$ 矩阵, $\mathbf{b} \in \mathbf{R}^m$ 。

对问题 5.16 建立非退化假设:

1. \mathbf{A} 的任意 m 列线性无关;
2. 任意可行解 $\mathbf{x} \in S$ 至少有 m 个正分量。

在非退化假设下, $\forall \mathbf{x} \in S$, 存在分解 $\mathbf{A} = (\mathbf{B}, \mathbf{N})$, 使得 $\mathbf{x} = \begin{pmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{pmatrix}$, 其中 $\mathbf{x}_B > \mathbf{0}, \mathbf{x}_N \geq \mathbf{0}$, 称 \mathbf{x}_B 为基变量, \mathbf{x}_N 为非基变量。

相应的, $\nabla f(\mathbf{x}) = \begin{pmatrix} \nabla_B f(\mathbf{x}) \\ \nabla_N f(\mathbf{x}) \end{pmatrix}$, 称 $\mathbf{r}_N^T = \nabla_N f^T(\mathbf{x}) - \nabla_B f^T(\mathbf{x}) \mathbf{B}^{-1} \mathbf{N}$ 为既约梯度。

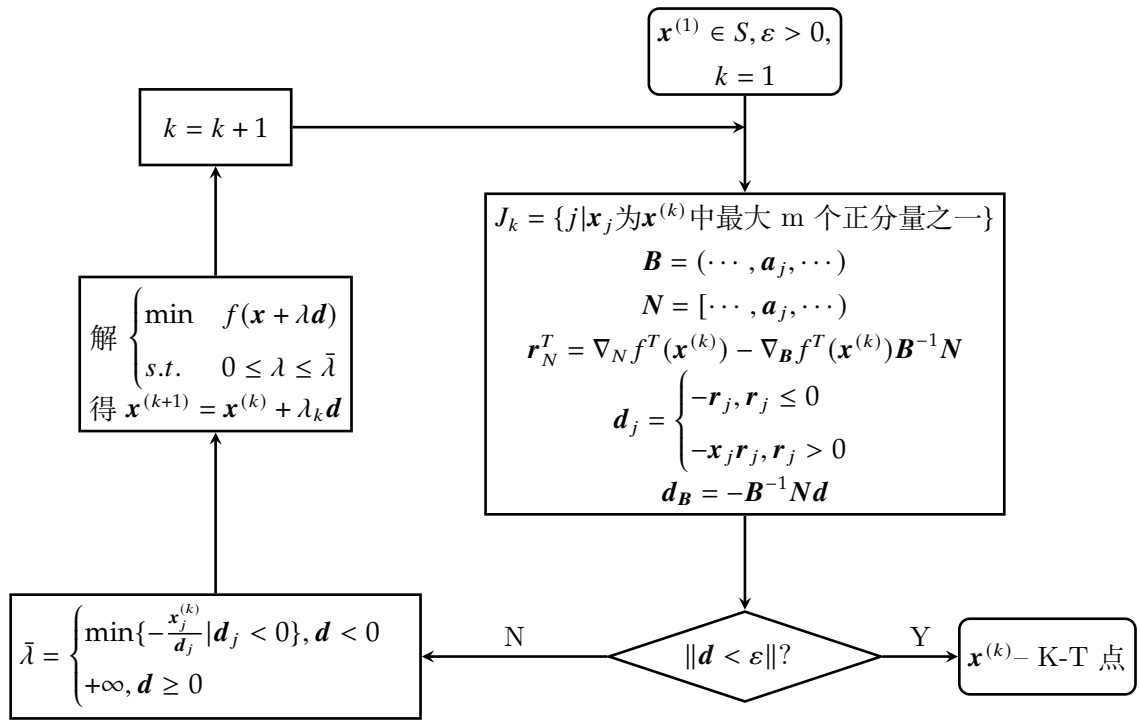


图 5.1: 既约梯度法流程