最优化方法及其 Matlab 程序设计

马昌凤

2010 年 3 月 22 日

目　录

第十章　可行方向法

本章介绍的可行方向法是一类直接处理约束优化问题的方法, 其基本思想是要求每一步迭代产生的搜索方向不仅对目标函数是下降方向, 而且对约束函数来说是可行方向, 即迭代点总是满足所有的约束条件. 各种不同的可行方向法的主要区别在于选取可行方向 d_k 的策略不同, 我们主要介绍 Zoutendijk 可行方向法、投影梯度法和简约梯度法三类可行方向法.

10.1 Zoutendijk 可行方向法

Zoutendijk 可行方向法是用一个线性规划来确定搜索方向— 下降可行方向的方法, 它最早是由 Zoutendijk 于 1960 年提出来的. 我们分线性约束和非线性约束两种情形来讨论其算法原理.

10.1.1 线性约束下的可行方向法

1. 基本原理

考虑下面的非线性优化问题

$$\begin{cases} \min\ f(x), \\ \text{s.t.}\ \ Ax \geq b, \\ \qquad Ex = e, \end{cases} \tag{10.1}$$

其中 $f(x)$ 连续可微, A 是 $m \times n$ 矩阵, E 是 $l \times n$ 矩阵, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $e \in \mathbb{R}^l$. 即 (10.1) 中有 m 个线性不等式约束和 l 个线性等式约束.

下面的引理指出了问题 (10.1) 的下降可行方向 d 应满足的条件.

引理 10.1 设 $\bar{x}$ 是问题 (10.1) 的一个可行点, 且在 $\bar{x}$ 处有 $A_1\bar{x}=b_1$, $A_2\bar{x}>b_2$, 其中

$$A=\begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad b=\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

则 $d\in\mathbb{R}^n$ 是点 $\bar{x}$ 处的下降可行方向的充要条件是

$$A_1 d\geq 0,\ \ Ed=0,\ \ \nabla f(\bar{x})^T d<0.$$

证 不难发现 d 是 $f(x)$ 在 $\bar{x}$ 处的下降方向的充要条件是 $\nabla f(\bar{x})^T d<0$. 另外, 注意到条件 $A_1\bar{x}=b_1$ 表明约束条件 $A_1x\geq b_1$ 是点 $\bar{x}$ 出的有效约束, 而条件 $A_2\bar{x}>b_2$ 表明约束条件 $A_2x\geq b_2$ 是点 $\bar{x}$ 处的非有效约束. 因此, 在可行点 $\bar{x}$ 处将约束矩阵 A 分裂为相应的 A_1 和 A_2 两部分.

充分性. 设 $A_1 d \geq 0$, $Ed = 0$. 因 $\bar{x}$ 是可行点, 且 $A_1\bar{x} = b_1$, $E\bar{x} = e$. 故对任意的 $\alpha > 0$, 都有

$$\begin{aligned} A_1(\bar{x} + \alpha d) &= A_1\bar{x} + \alpha(A_1 d) \geq A_1\bar{x} = b_1, \\ E(\bar{x} + \alpha d) &= E\bar{x} + \alpha(Ed) = E\bar{x} = e. \end{aligned}$$

又由 $A_2\bar{x} > b_2$, 故必存在一个 $\bar{\alpha} > 0$, 使得对于任意的 $\alpha \in (0, \bar{\alpha}]$, 都有

$$A_2(\bar{x} + \alpha d) = A_2\bar{x} + \alpha A_2 d \geq b_2.$$

以上三式表明, 存在 $\bar{\alpha}$, 使得对于任意的 $\alpha \in (0, \bar{\alpha}]$, 有

$$A(\bar{x} + \alpha d) \geq b, \;\; E(\bar{x} + \alpha d) = e,$$

即 $\bar{x} + \alpha d$ 是可行点, 从而 d 是 $\bar{x}$ 处的可行方向.

必要性. 设 $\bar{x}$ 是可行点, d 是 $\bar{x}$ 处的一个可行方向. 由可行方向的定义, 存在 $\bar{\alpha}$, 使得对于任意的 $\alpha \in (0, \bar{\alpha}]$, 有

$$A(\bar{x} + \alpha d) \geq b, \;\; E(\bar{x} + \alpha d) = e,$$

或

$$A_1(\bar{x}+\alpha d) \geq b_1,\ \ A_2(\bar{x}+\alpha d) \geq b_2,\ \ E(\bar{x}+\alpha d) = e.$$

于是由

$$A_1(\bar{x}+\alpha d) = A_1\bar{x}+\alpha(A_1 d) \geq b_1,\ \ A_1\bar{x} = b_1,\ \ \alpha > 0$$

可推出 $A_1 d \geq 0$. 又由

$$E(\bar{x}+\alpha d) = E\bar{x}+\alpha(A_1 d) \geq e,\ \ E\bar{x} = e,\ \ \alpha > 0$$

可推出 $Ed = 0$. 证毕. □

上面的引理启发我们, 要寻找问题 (10.1) 的可行点 $\bar{x}$ 处的一个下降可行方向 d, 可以通过求解下述线性规划问题得到:

$$\begin{cases} \min\ \nabla f(\bar{x})^T d, \\ \text{s.t.}\ \ A_1 d \geq 0, \\ \qquad Ed = 0, \\ \qquad -1 \leq d_i \leq 1,\ i = 1, \cdots, n, \end{cases} \tag{10.2}$$

其中 $d = (d_1, \cdots, d_n)^T$. 增加约束条件 $-1 \leq d_i \leq 1$, $i = 1, \cdots, n$ 是为了防止 $\|d\| \to \infty$.

注意到, $d = 0$ 显然是子问题 (10.2) 的一个可行解, 故目标函数 $\nabla f(\bar{x})^T d$ 的最优值必然小于或等于 0. 若目标函数的最优值 $\bar{z} = \nabla f(\bar{x})^T \bar{d} < 0$, 则由引理 10.1 可知, $\bar{d}$ 即为 $\bar{x}$ 处的下降可行方向. 否则, 若标函数的最优值 $\bar{z} = \nabla f(\bar{x})^T \bar{d} = 0$, 则可以证明 $\bar{x}$ 是问题 (10.1) 的 KT 点.

定理 10.1 设 $\bar{x}$ 是问题 (10.1) 的一个可行点, 且在 $\bar{x}$ 处有 $A_1 \bar{x} = b_1$,

$A_2\bar{x} > b_2$, 其中

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

则 $\bar{x}$ 是问题 (10.1) 的 KT 点的充要条件是子问题 (10.2) 的最优值为 0.

由于上述定理的证明需要用到 Farkas 引理 (引理 **??**), 为了使用方便, 我们给出 Farkas 引理的一个等价描述方式.

引理 10.2 (**Farkas 引理**) *设 A 为 $m \times n$ 矩阵, c 为 n 维向量. 则 $A^T y = c$, $y \geq 0$ 有解的充分必要条件是 $Ax \geq 0$, $c^T x > 0$ 无解, 其中 x, y 分别是为 n, m 维向量.*

定理 10.1 的证明: 注意到, $\bar{x}$ 是 KT 点充要条件是, 存在 $\lambda \geq 0$ 和 μ, 使得

$$\nabla f(\bar{x}) - A_1^T \lambda - E^T \mu = 0. \tag{10.3}$$

令 $\mu = \nu_1 - \nu_2\,(\nu_1, \nu_2 \geq 0)$, 把 (10.3) 写成

$$\left(-A_1^T, -E^T, E^T\right)\begin{pmatrix}\lambda \\ \nu_1 \\ \nu_2\end{pmatrix} = -\nabla f(\bar{x}), \qquad \begin{pmatrix}\lambda \\ \nu_1 \\ \nu_2\end{pmatrix} \geq 0. \tag{10.4}$$

根据 Farkas 引理, (10.4) 有解的充要条件是

$$\begin{pmatrix}-A_1 \\ -E \\ E\end{pmatrix} d \leq 0, \quad -\nabla f(\bar{x})^T d > 0 \tag{10.5}$$

无解, 即

$$A_1 d \geq 0, \quad Ed = 0, \quad \nabla f(\bar{x})^T d < 0$$

无解. 故 $\bar{x}$ 是问题 (10.1) 的 KT 点的充要条件是子问题 (10.2) 的最优值为 0. $\square$

由上述定理可知, 求解子问题 (10.2) 的结果, 或者得到下降可行方向, 或者得到原问题的一个 KT 点.

2. 计算步骤

下面讨论可行方向法的具体计算步骤. 首先分析如何确定搜索步长 α_k. 设问题的可行域为 $\mathcal{F}$. 第 k 次迭代的出发点 $x_k \in \mathcal{F}$ 是可行点, d_k 是其下降可行方向, 则后继点 x_{k+1} 为

$$x_{k+1} = x_k + \alpha_k d_k. \tag{10.6}$$

为了使得 $x_{k+1} \in \mathcal{F}$, 且使 $f(x_{k+1})$ 的值尽可能小, 可以通过求解下面的一维搜索问题来解决:

$$\begin{aligned} &\min_{0\le\alpha\le\bar{\alpha}} f(x_k + \alpha d_k), \\ &\bar{\alpha} = \max\{\alpha | x_k + \alpha d_k \in \mathcal{F}\}. \end{aligned} \tag{10.7}$$

在求解 (10.7) 式时, 考虑到线性约束情形时的 (10.1) 式, 先求解

$$\begin{cases} \min \ \ f(x_k + \alpha d_k), \\ \text{s.t.} \ \ A(x_k + \alpha d_k) \geq b, \\ \qquad E(x_k + \alpha d_k) = e. \end{cases} \tag{10.8}$$

而 (10.8) 可作进一步的简化: 因为 d_k 是可行方向, 有 $Ed_k = 0$; 而 x_k 是可行点, 有 $Ex_k = e$. 因此 (10.8) 中的等式约束条件自然成立, 可不必再考虑它. 此外, 在 x_k 处, 将不等式约束分为有效约束和非有效约束, 设

$$A_1 x_k = b_1, \quad A_2 x_k > b_2, \tag{10.9}$$

其中

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

则 (10.8) 中的不等式约束条件可分裂成:

$$A_1x_k + \alpha A_1 d_k \geq b_1, \tag{10.10}$$
$$A_2x_k + \alpha A_2 d_k \geq b_2. \tag{10.11}$$

又因 d_k 是可行方向, 由引理 10.1 知 $A_1 d_k \geq 0$. 注意到 $A_1 x_k = b_1$ 及 $\alpha \geq 0$, 因此条件 (10.10) 也自然成立. 于是 (10.8) 中的约束条件只剩下 (10.11), 故 (10.8) 可简化为:

$$\begin{cases} \min \quad f(x_k + \alpha d_k), \\ \text{s.t.} \quad A_2(x_k + \alpha d_k) > b_2, \\ \qquad\quad \alpha \geq 0. \end{cases} \tag{10.12}$$

以下讨论 (10.12) 中求 α 上限的公式. 将 (10.12) 中的第一个约束条件改写成:

$$\alpha A_2 d_k \geq b_2 - A_2 x_k.$$

若记

$$\bar{b} = b_2 - A_2 x_k, \quad \bar{d} = A_2 d_k, \tag{10.13}$$

则有

$$\alpha \bar{d} \geq \bar{b}, \quad \alpha \geq 0.$$

注意到 (10.9) 式, 我们有 $\bar{b} < 0$. 由此可得 α 的上界计算公式:

$$\bar{\alpha} = \begin{cases} \min \left\{ \dfrac{\bar{b}_i}{\bar{d}_i} = \dfrac{(b_2 - A_2 x_k)_i}{(A_2 d_k)_i} \,\middle|\, \bar{d}_i < 0 \right\}, & \bar{d} \ngeq 0, \\ +\infty, & \bar{d} \geq 0, \end{cases} \tag{10.14}$$

其中 $\bar{b}_i$, $\bar{d}_i$ 分别是向量 $\bar{b}$, $\bar{d}$ 的第 i 个分量. 因此, 求解 (10.12) 等价于求解

$$\begin{cases} \min \quad f(x_k + \alpha d_k), \\ \text{s.t.} \quad 0 \leq \alpha \leq \bar{\alpha}, \end{cases} \tag{10.15}$$

其中 $\bar{\alpha}$ 由 (10.14) 式计算.

至此, 我们可以写出求解问题 (10.1) 的可行方向法的详细计算步骤:

算法 10.1 (线性约束的可行方向法)

步 0 给定初始可行点 x_0, 终止误差 $0 < \varepsilon_1 \ll 1$, $0 < \varepsilon_2 \ll 1$. 令 $k := 0$.

步 1 在 x_k 处, 将不等式约束分为有效约束和非有效约束:

$$A_1 x_k = b_1, \quad A_2 x_k > b_2,$$

其中

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

步 2 若 x_k 是可行域的一个内点 (此时问题 (10.1) 中没有等式约束, 即 $E = 0$ 且 $A_1 = 0$), 且 $\|\nabla f(x^k)\| < \varepsilon_1$, 停算, 得到近似极小点 x_k; 否则, 若 x_k 是可行域的一个内点但 $\|\nabla f(x^k)\| \geq \varepsilon_1$,则取搜索方向 $d_k = -\nabla f(x_k)$, 转步 5(即用目标函数的负梯度方向作为搜索方向再求步长, 此时类似于无约束优化问题). 若 x_k 不是可行域的一个内点, 则转步 3.

步 3 求解线性规划问题

$$
\begin{cases}
\min \ z = \nabla f(x_k)^T d, \\
\text{s.t.} \ \ A_1 d \geq 0, \\
\qquad\ \ Ed = 0, \\
\qquad\ \ -1 \leq d_i \leq 1, \ i = 1, \cdots, n,
\end{cases} \tag{10.16}
$$

其中 $d = (d_1, \cdots, d_n)^T$. 设求得最优解和最优值分别为 d_k 和 z_k.

步 4 若 $|z_k| < \varepsilon_2$, 停算, 输出 x_k 作为近似极小点. 否则, 以 d_k 作为搜索方向, 转步 5.

步 5 首先由 (10.13) 和 (10.14) 计算 $\bar{\alpha}$, 然后作一维线搜索:

$$
\begin{cases}
\min \ \ f(x_k + \alpha d_k), \\
\text{s.t.} \ \ 0 \leq \alpha \leq \bar{\alpha},
\end{cases}
$$

求得最优解 α_k.

步 6 置 $x_{k+1} := x_k + \alpha_k d_k$, $k := k+1$, 转步 1.

10.1.2 非线性约束下的可行方向法

1. 基本原理

考虑下面带有非线性不等式约束的优化问题:

$$\begin{cases} \min \ \ f(x), \\ \text{s.t.} \ \ g_i(x) \geq 0, \ i = 1, \cdots, m, \end{cases} \tag{10.17}$$

其中 $x \in \mathbb{R}^n$, $f(x)$, $g_i(x)\,(i = 1, \cdots, m)$ 都是连续可微的函数.

下面的引理指出了问题 (10.17) 的一个下降可行方向 所应满足的条件.

引理 10.3 设 $\bar{x}$ 是问题 (10.17) 的一个可行点, 指标集 $I(\bar{x}) = \{i \,|\, g_i(\bar{x}) = 0\}$, $f(x)$, $g_i(x)\,(i \in I(\bar{x}))$ 在 $\bar{x}$ 处可微, $g_i(x)\,(i \notin I(\bar{x}))$ 在 $\bar{x}$ 处连续. 若

$$\nabla f(\bar{x})^T d < 0, \quad \nabla g_i(\bar{x})^T d > 0 \,(i \in I(\bar{x})),$$

那么 d 是问题 (10.17) 在 $\bar{x}$ 处的下降可行方向.

证 由引理 ?? 中的下降可行方向的代数条件 (??) 可知, d 必是问题 (10.17) 在 $\bar{x}$ 处的一个下降可行方向. 证毕. □

由上述引理可知, 问题 (10.17) 在可行点 $\bar{x}$ 处的下降可行方向 d 应满足:

$$\begin{cases} \nabla f(\bar{x})^T d < 0, \\ \nabla g_i(x)^T d > 0,\ i \in I(\bar{x}). \end{cases} \tag{10.18}$$

而在 (10.18) 中引进辅助变量 z 后, 等价于下面的线性不等式组求 d 和 z:

$$\begin{cases} \nabla f(\bar{x})^T d \le z, \\ -\nabla g_i(x)^T d \le z,\ i \in I(\bar{x}), \\ z \le 0. \end{cases} \tag{10.19}$$

注意到, 满足 (10.19) 的下降可行方向 d 及数 z 一般有很多个, 我们自然希望求出能使目标函数下降最多的方向 d. 故而可将 (10.19) 转化为以 z 为

目标函数的线性规划问题

$$
\begin{cases}
\min \ z, \\
\text{s.t. } \nabla f(\bar{x})^T d \le z, \\
\qquad -\nabla g_i(x)^T d \le z, \ i \in I(\bar{x}), \\
\qquad -1 \le d_i \le 1, \ i = 1, \cdots, n,
\end{cases}
\tag{10.20}
$$

其中 $d = (d_1, \cdots, d_n)^T$.

设 (10.20) 的最优解为 $\bar{d}$, 最优值为 $\bar{z}$. 那么, 若 $\bar{z} < 0$, 则 $\bar{d}$ 是问题 (10.17) 在 $\bar{x}$ 处的下降可行方向; 否则, 若 $\bar{z} = 0$, 则下面的定理将证明: 相应的 $\bar{x}$ 必为问题 (10.17) 的 Fritz John 点.

定理 10.2 设 $\bar{x}$ 是问题 (10.17) 的可行点, $I(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}$. 则 $\bar{x}$ 是问题 (10.17) 的 Fritz John 点的充要条件是子问题 (10.20) 的最优值为 0.

证 对于子问题 (10.20), 其最优值为 0 的充要条件是不等式组

$$\begin{cases} \nabla f(\bar{x})^T d < 0, \\ \nabla g_i(\bar{x})^T d > 0, \ \ i \in I(\bar{x}), \end{cases}$$

即

$$\begin{cases} \nabla f(\bar{x})^T d < 0, \\ -\nabla g_i(\bar{x})^T d < 0, \ \ i \in I(\bar{x}), \end{cases} \tag{10.21}$$

无解. 根据 Gordan 引理 (引理 **??**), 不等式组 (10.21) 无解的充要条件是存在不全为 0 的数 $\lambda_0 \geq 0$ 和 λ_i $(i \in I(\bar{x}))$, 使得

$$\lambda_0 \nabla f(\bar{x}) - \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) = 0,$$

即 $\bar{x}$ 是问题 (10.17) 的 Fritz John 点. 证毕. □

2. 计算步骤

与线性约束情形一样, 为了确定搜索步长 α_k, 仍然需要求解一个一维搜索问题:

$$\begin{aligned} &\min\ f(x_k + \alpha d_k), \\ &\text{s.t. } 0 \le \alpha \le \bar{\alpha}, \end{aligned} \tag{10.22}$$

其中

$$\bar{\alpha} = \sup\{\alpha \,|\, g_i(x_k + \alpha d_k) \ge 0,\ i = 1, \cdots, m\}. \tag{10.23}$$

现在, 我们写出求解问题 (10.17) 的可行方向法的详细计算步骤:

算法 10.2 (**非线性约束的可行方向法**)

步 0　给定初始可行点 x_0, 终止误差 $0 < \varepsilon_1 \ll 1$, $0 < \varepsilon_2 \ll 1$. 令 $k := 0$.

步 1　确定 x_k 处的有效约束指标集 $I(x_k)$:

$$I(x_k) = \{i \,|\, g_i(x_k) = 0\}.$$

若 $I(x_k) = \emptyset$ 且 $\|\nabla f(x_k)\| < \varepsilon_1$, 停算, 得到近似极小点 x_k; 否则若

$I(x_k)=\emptyset$ 但 $\|\nabla f(x_k)\|\geq\varepsilon_1$, 则取搜索方向 $d_k=-\nabla f(x_k)$, 转步 4. 反之, 若 $I(x_k)\neq\emptyset$, 转步 2.

步 2 求解线性规划问题 (10.20), 得最优解 d_k, 最优值 z_k.

步 3 若 $|z_k|<\varepsilon_2$, 停算, 输出 x_k 作为近似极小点. 否则, 以 d_k 作为搜索方向, 转步 4.

步 4 首先由 (10.23) 计算 $\bar{\alpha}$, 然后作一维线搜索:

$$\begin{cases}\min & f(x_k+\alpha d_k),\\ \text{s.t.} & 0\leq\alpha\leq\bar{\alpha},\end{cases}$$

求得最优解 α_k.

步 5 置 $x_{k+1}:=x_k+\alpha_k d_k$, $k:=k+1$, 转步 1.

注 (1) 步 1 中的 $I(x_k)=\emptyset$, 表明 x_k 是可行域 $\mathcal{F}$ 的内点, 因此任意方向都是可行方向. 此时, 若不满足终止条件, 类似于无约束优化问题, 可用最速下降法寻求下一个迭代点. 但毕竟不是真正的无约束问题, 步长要受到可行域边界的限制.

(2) 步 3 中若 $z_k \approx 0$, 说明在 x_k 处找不到下降可行方向, 可以认为 x_k 是原问题的一个 Fritz-John 点.

(3) 算法 10.2 若推广到包含非线性等式约束的优化问题, 迭代过程会出现一些困难. 因为对于等式约束和当前可行迭代点 x_k, 一般难于找到一个可行方向. 这与罚函数类算法刚好相反, 罚函数类算法容易处理等式约束.

例 10.1 用 Zoutendijk 可行方向法求解下列问题:

$$
\begin{aligned}
\min\ & x_1^2 + x_2^2 - 2x_1 - 4x_2 + 3, \\
\text{s.t.}\ & -2x_1 + x_2 \geq -1, \\
& -x_1 - x_2 \geq -2, \\
& x_1 \geq 0,\ x_2 \geq 0.
\end{aligned}
$$

解 取初始可行点 $x_0 = (0, 0)^T$.

第 1 次迭代. $\nabla f(x_0) = (-2, -4)^T$. 有效约束和非有效约束的系数矩阵和右端向量分别为:

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -2 & 1 \\ -1 & -1 \end{bmatrix}, \quad b_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad b_2 = \begin{bmatrix} -1 \\ -2 \end{bmatrix}.$$

先求在 x_0 处的下降可行方向 d_0. 解线性规划问题

$$\begin{aligned} &\min\ \nabla f(x_0)^T d, \\ &\text{s.t. } A_1 d \geq 0, \\ &\qquad |d_i| < 1,\ i = 1, 2, \end{aligned}$$

即

$$
\begin{aligned}
&\min\ -2d_1-4d_2,\\
&\text{s.t. } d_1\ge 0,\ d_2\ge 0,\\
&\qquad -1\le d_1\le 1,\\
&\qquad -1\le d_2\le 1.
\end{aligned}
$$

由单纯形方法求得最优解为 $d_0=(1,1)^T$.

再求步长因子 α_0. 分别计算 $\bar{b}$, $\bar{d}$ 和 $\bar{\alpha}$:

$$\bar{d} = A_2 d_0 = \begin{bmatrix} -2 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix},$$

$$\bar{b} = b_2 - A_2 x_0 = \begin{bmatrix} -1 \\ -2 \end{bmatrix} - \begin{bmatrix} -2 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix},$$

$$\bar{\alpha} = \min\left\{\frac{-1}{-1}, \frac{-2}{-2}\right\} = 1.$$

于是解下面的一维搜索问题求 α_0:

$$\min\ f(x_0 + \alpha d_0) = 2\alpha^2 - 6\alpha + 3,$$

$$\text{s.t. } 0 \le \alpha \le 1.$$

求得 $\alpha_0 = 1$. 令 $x_1 = x_0 + \alpha_0 d_0 = (1, 1)^T$. 至此第 1 次迭代完成.

第 2 次迭代. $\nabla f(x_1) = (0, -2)^T$. 有效约束和非有效约束的系数矩阵和右端向量分别为:

$$A_1 = \begin{bmatrix} -2 & 1 \\ -1 & -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad b_1 = \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

解线性规划问题

$$\begin{aligned} &\min \ -2d_2, \\ &\text{s.t. } 2d_1 + d_2 \geq 0, \\ &\qquad -d_1 - d_2 \geq 0, \\ &\qquad -1 \leq d_1 \leq 1, \\ &\qquad -1 \leq d_2 \leq 1. \end{aligned}$$

由单纯形方法求得最优解为 $d_1 = (-1, 1)^T$.

再求步长因子 α_1. 分别计算 $\bar{b}$, $\bar{d}$ 和 $\bar{\alpha}$:

$$\bar{d} = A_2 d_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

$$\bar{b} = b_2 - A_2 x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix},$$

得 $\bar{\alpha} = 1$. 于是解下面的一维搜索问题求 α_1:

$$\begin{aligned} &\min\ f(x_1 + \alpha d_1) = 2\alpha^2 - 2\alpha - 1, \\ &\text{s.t.}\ 0 \le \alpha \le 1. \end{aligned}$$

求得 $\alpha_1 = 0.5$. 令 $x_2 = x_1 + \alpha_1 d_1 = (1,1)^T + 0.5(-1,1)^T = (0.5, 1.5)^T$. 至此第 2 次迭代完成.

第 3 次迭代. $\nabla f(x_2) = (-1,-1)^T$. 有效约束和非有效约束的系数矩阵和右端向量分别为:

$$A_1 = \begin{bmatrix} -1 & -1 \end{bmatrix},\quad A_2 = \begin{bmatrix} -2 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},\quad b_1 = \begin{bmatrix} -2 \end{bmatrix},\quad b_2 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}.$$

解线性规划问题

$$\begin{aligned} \min\ & -d_1 - d_2, \\ \text{s.t.}\ & -d_1 - d_2 \geq 0, \\ & -1 \leq d_1 \leq 1, \\ & -1 \leq d_2 \leq 1. \end{aligned}$$

由单纯性方法求得最优解为 $d_1 = (0,0)^T$. 由定理 10.1 知, $x_2 = (0.5,1.5)^T$ 是 KT 点. 由于此例是凸规划, 因此, x_2 也是最优解, 相应地, 目标函数的最优值为 $f_{\min} = f(x_2) = 1.5$.

10.2 梯度投影法

梯度投影法是 Rosen 于 1961 年针对线性约束的优化问题首先提出来的一种优化算法, 次年 Rosen 又将他的这一算法推广到处理非线性约束的情形. 后来这一方法又得到了进一步的发展, 成为求解非线性规划问题的一类重要的方法.

10.2.1 梯度投影法的理论基础

我们考虑线性约束的优化问题

$$\begin{cases} \min f(x), \\ \text{s.t.} \;\; Ax \geq b, \\ \qquad\quad Ex = e, \end{cases} \tag{10.24}$$

其中 f 是连续可微的 n 元实函数, $A \in \mathbb{R}^{m\times n}$, $E \in \mathbb{R}^{l\times n}$, $b \in \mathbb{R}^m$, $e \in \mathbb{R}^l$, $x \in \mathbb{R}^n$. 其可行域为 $\mathcal{D} = \{x \in \mathbb{R}^n \,|\, Ax \geq b,\ Ex = e\}$.

梯度投影法的基本思想是: 当迭代点 x_k 是可行域 $\mathcal{D}$ 的内点时, 取 $d = -\nabla f(x_k)$ 作为搜索方向; 否则, 当 x_k 是可行域 $\mathcal{D}$ 的边界点时, 取 $-\nabla f(x_k)$ 在这些边界面交集上的投影作为搜索方向. 这也是“梯度投影法”名称的由来.

在具体介绍梯度投影法之前, 我们先引入投影矩阵的概念及其有关性质.

定义 10.1 称矩阵 $P \in \mathbb{R}^{n\times n}$ 为投影矩阵, 是指 P 满足

$$P = P^T, \quad P^2 = P.$$

由上述定义可知, 一个对称幂等矩阵就是投影矩阵. 投影矩阵具有如下一些基本性质, 其证明可参看文献 [?].

引理 10.4 设矩阵 $P \in \mathbb{R}^{n\times n}$.

(1) 若 P 为投影矩阵, 则 P 是半正定的. (2) P 是投影矩阵当且仅当 $I - P$ 也是投影矩阵, 其中 I 是 n 阶单位阵. (3) 设 P 是投影矩阵,

$Q = I - P$, 则

$$L = \{y = Px | x \in \mathbb{R}^n\}, \quad L^{\perp} = \{z = Qx | x \in \mathbb{R}^n\}$$

是互相正交的线性子空间, 并且对于任意的 $x \in \mathbb{R}^n$ 可唯一地表示为

$$x = y + z,\ y \in L,\ z \in L^{\perp}.$$

定理 10.3 设 $\bar{x}$ 是问题 (10.24) 的一个可行点, 且满足 $A_1\bar{x} = b_1$, $A_2x > b_2$, 其中

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

又设

$$M = \begin{bmatrix} A_1 \\ E \end{bmatrix}$$

是满秩矩阵, $P = I - M^T(MM^T)^{-1}M$, $P\nabla f(\bar{x}) \neq 0$. 若取 $d = -P\nabla f(\bar{x})$, 则 d 是问题 (10.24) 的一个下降可行方向.

证 不难验证, $P = I - M^T(MM^T)^{-1}M$ 是投影矩阵, 故由 $P\nabla f(\bar{x}) \neq 0$ 可得

$$\nabla f(\bar{x})^T d = -\nabla f(\bar{x})^T P \nabla f(\bar{x})^T = -\|P\nabla f(\bar{x})^T\|^2 < 0,$$

即 d 为下降方向. 又因

$$\begin{aligned} Md &= -MP\nabla f(\bar{x})^T \\ &= -M(I - M^T(MM^T)^{-1}M)\nabla f(\bar{x})^T \\ &= (-M + M)\nabla f(\bar{x})^T = 0, \end{aligned}$$

即

$$Md = \begin{bmatrix} A_1 \\ E \end{bmatrix} d = \begin{bmatrix} A_1 d \\ Ed \end{bmatrix} = 0,$$

从而 $A_1 d = 0$, $Ed = 0$. 根据引理 10.1, d 是在 $\bar{x}$ 的可行方向. 证毕. □

上述定理在 $P\nabla f(\bar{x}) \neq 0$ 的假设下, 给出了用投影来求下降可行方向的一种方法. 但是, 当 $P\nabla f(\bar{x}) = 0$ 时, 情况又该如何呢？下面的定理指

出, 此时有两种可能: 要么 $\bar{x}$ 是已是 KT 点, 要么构造新的投影矩阵, 以便求得下降可行方向.

定理 10.4 设 $\bar{x}$ 是问题 (10.24) 的一个可行点, 且满足 $A_1\bar{x} = b_1$, $A_2x > b_2$, 其中

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

又设

$$M = \begin{bmatrix} A_1 \\ E \end{bmatrix}$$

是行满秩矩阵，令

$$P = I - M^T(MM^T)^{-1}M,$$

$$\omega = (MM^T)^{-1}M\nabla f(\bar{x}) = \begin{bmatrix} \lambda \\ \mu \end{bmatrix},$$

其中 λ 和 μ 分别对应于 A_1 和 E. 若 $P\nabla f(\bar{x}) = 0$, 则

(1) 如果 $\lambda \geq 0$, 那么 $\bar{x}$ 是 KT 点.

(2) 如果 $\lambda \ngeq 0$, 不妨设 $\lambda_j < 0$, 那么先从 A_1 中去掉 λ_j 所对应的行，得到新矩阵 $\tilde{A}_1$, 然后令

$$\tilde{M} = \begin{bmatrix} \tilde{A}_1 \\ E \end{bmatrix}, \quad \tilde{P} = I - \tilde{M}^T(\tilde{M}\tilde{M}^T)^{-1}\tilde{M}, \quad d = -\tilde{P}\nabla f(\bar{x}),$$

那么，d 是下降可行方向.

证 (1) 设 $\lambda \geq 0$. 注意到 $P\nabla f(\bar{x}) = 0$, 我们有

$$\begin{aligned} 0 &= P\nabla f(\bar{x}) = [I - M^T(MM^T)^{-1}M]\nabla f(\bar{x}) \\ &= \nabla f(\bar{x}) - M^T(MM^T)^{-1}M\nabla f(\bar{x}) \\ &= \nabla f(\bar{x}) - M^T\omega = \nabla f(\bar{x}) - [A_1^T, E^T]\begin{bmatrix} \lambda \\ \mu \end{bmatrix} \\ &= \nabla f(\bar{x}) - A_1^T\lambda - E^T\mu. \end{aligned} \tag{10.25}$$

(10.25) 恰为 KT 条件, 因此 $\bar{x}$ 是 KT 点.

(2) 设 $\lambda_j < 0$. 先证明 $\tilde{P}\nabla f(\bar{x}) \neq 0$. 用反证法. 若 $\tilde{P}\nabla f(\bar{x}) = 0$, 则由 $\tilde{P}$ 定义可得

$$0 = \tilde{P}\nabla f(\bar{x}) = [I - \tilde{M}^T(\tilde{M}\tilde{M}^T)^{-1}\tilde{M}]\nabla f(\bar{x}) = \nabla f(\bar{x}) - \tilde{M}^T\tilde{\omega}, \tag{10.26}$$

其中 $\tilde{\omega} = (\tilde{M}\tilde{M}^T)^{-1}\tilde{M}\nabla f(\bar{x})$. 设 A_1 中对应于 λ_j 的行为 r_j (第 j 行). 由于

$$A_1^T\lambda + E^T\mu = \tilde{A}_1^T\tilde{\lambda} + \lambda_j r_j^T + E^T\mu = \tilde{M}^T\bar{\omega} + \lambda_j r_j^T. \tag{10.27}$$

将 (10.27) 代入 (10.25) 得

$$0 = \nabla f(\bar{x}) - \tilde{M}^T \bar{\omega} - \lambda_j r_j^T. \tag{10.28}$$

(10.26) 减去 (10.28) 得

$$\tilde{M}^T(\bar{\omega} - \tilde{\omega}) + \lambda_j r_j = 0. \tag{10.29}$$

上式左端是矩阵 M 的行向量的一个线性组合, 且至少有一个系数 $\lambda_j \neq 0$. 由此可得 M 的行向量线性相关, 这与 M 是行满秩矛盾. 因此必有 $\tilde{P}\nabla f(\bar{x}) \neq 0$.

由于 $\tilde{P}$ 亦为投影矩阵, 且 $\tilde{P}\nabla f(\bar{x}) \neq 0$, 故

$$\nabla f(\bar{x})^T d = -\nabla f(\bar{x})^T \tilde{P} \nabla f(\bar{x})^T = -\|\tilde{P}\nabla f(\bar{x})\|^2 < 0,$$

即 d 是下降方向. 以下只需证明 d 是可行方向即可. 事实上, 因

$$\begin{aligned}\tilde{M}d &= -\tilde{M}\tilde{P}\nabla f(\bar{x}) \\ &= -\tilde{M}[I - \tilde{M}^T(\tilde{M}\tilde{M}^T)^{-1}\tilde{M}]\nabla f(\bar{x}) \\ &= -(\tilde{M} - \tilde{M})\nabla f(\bar{x}) = 0,\end{aligned}$$

即

$$\tilde{A}_1 d = 0, \quad Ed = 0. \tag{10.30}$$

将 (10.28) 两边左乘 $r_j \tilde{P}$ 得

$$r_j \tilde{P} \nabla f(\bar{x}) - r_j \tilde{P} \tilde{M}^T \bar{\omega} - \lambda_j r_j \tilde{P} r_j^T = 0.$$

注意到 $\tilde{P}\tilde{M}^T = 0$ 及 $d = -\tilde{P}\nabla f(\bar{x})$, 上式即

$$r_j d + \lambda_j r_j \tilde{P} r_j^T = 0. \tag{10.31}$$

因 $\tilde{P}$ 半正定 $(r_j \tilde{P} r_j^T \geq 0)$ 及 $\lambda_j < 0$, 故有

$$r_j d = -\lambda_j r_j \tilde{P} r_j^T \geq 0. \tag{10.32}$$

由 (10.30) 和 (10.32) 即得

$$A_1 d \geq 0, \quad Ed = 0.$$

最后, 根据引理 10.1, d 是在 $\bar{x}$ 的可行方向. 证毕. $\square$

10.2.2　梯度投影法的计算步骤

基于上述分析与讨论, 我们给出 Rosen 梯度投影法的详细计算步骤如下:

算法 10.3 (Rosen **梯度投影法**)

步 0　给定初始可行点 x_0, 令 $k := 0$.

步 1　在 x_k 处确定有效约束 $A_1x_k = b_1$ 和非有效约束 $A_2x_k > b_2$, 其中

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

步 2　令

$$M = \begin{bmatrix} A_1 \\ E \end{bmatrix}.$$

若 M 是空的, 则令 $P = I$ (单位矩阵). 否则, 令 $P = I - M^T(MM^T)^{-1}M$.

步 3　计算 $d_k = -P\nabla f(x_k)$. 若 $\|d_k\| \neq 0$, 转步 5; 否则, 转步 4.

步 4　计算

$$\omega = (MM^T)^{-1}M\nabla f(x_k) = \begin{bmatrix} \lambda \\ \mu \end{bmatrix}.$$

若 $\lambda \geq 0$, 停算, 输出 x_k 为 KT 点. 否则, 选取 λ 的某个负分量, 比如 $\lambda_j < 0$, 修正矩阵 A_1, 即去掉 A_1 中对应于 λ_j 的行, 转步 2.

步 5　求解一维搜索问题, 确定步长 α_k :

$$\begin{aligned} &\min\ f(x_k + \alpha d_k), \\ &\text{s.t.}\ \ 0 \leq \alpha \leq \bar{\alpha}, \end{aligned}$$

其中 $\bar{\alpha}$ 由下式确定

$$\bar{\alpha} = \begin{cases} \min\left\{\dfrac{(b_2 - A_2x_k)_i}{(A_2d_k)_i}\middle|(A_2d_k)_i < 0\right\}, & A_2d_k \ngeq 0, \\ +\infty, & A_2d_k \geq 0. \end{cases}$$

步 6　令 $x_{k+1} := x_k + \alpha_k d_k$, $k := k + 1$, 转步 1.

例 10.2 用 Rosen 梯度投影法求解下面的优化问题

$$
\begin{cases}
\min f(x) = x_1^2 + x_2^2 + 6x_1 + 9, \\
\text{s.t. } 2x_1 + x_2 \geq 4, \\
\qquad x_1, x_2 \geq 0.
\end{cases}
$$

取初始可行点为 $x_0 = (2,0)^T$.

解 首先目标函数的梯度

$$
\nabla f(x) = \begin{bmatrix} 2x_1 + 6 \\ 2x_2 \end{bmatrix}.
$$

第 1 次迭代. 在 x_0 处的梯度为

$$
\nabla f(x_0) = \begin{bmatrix} 10 \\ 0 \end{bmatrix}.
$$

在 x_0 处有效约束的指标集为 $I=\{1,3\}$, 即 $2x_1+x_2\geq 4$ 和 $x_2\geq 0$ 是在 $x_0=(2,0)^T$ 处的有效约束. 将约束系数矩阵 A 和右端向量 b 分解为

$$A_1=\begin{bmatrix}2&1\\0&1\end{bmatrix},\quad b_1=\begin{bmatrix}4\\0\end{bmatrix},\quad A_2=\begin{bmatrix}1&0\end{bmatrix},\quad b_2=\begin{bmatrix}0\end{bmatrix}.$$

因 $E=\emptyset$, 故 $M=A_1$. 计算投影矩阵

$$\begin{aligned}P&=I-A_1^T(A_1A_1^T)^{-1}A_1\\&=\begin{bmatrix}1&0\\0&1\end{bmatrix}-\begin{bmatrix}2&0\\1&1\end{bmatrix}\left(\begin{bmatrix}2&1\\0&1\end{bmatrix}\begin{bmatrix}2&0\\1&1\end{bmatrix}\right)^{-1}\begin{bmatrix}2&1\\0&1\end{bmatrix}\\&=\begin{bmatrix}1&0\\0&1\end{bmatrix}-\begin{bmatrix}2&0\\1&1\end{bmatrix}\left(\frac{1}{4}\begin{bmatrix}1&-1\\-1&5\end{bmatrix}\right)\begin{bmatrix}2&1\\0&1\end{bmatrix}=\begin{bmatrix}0&0\\0&0\end{bmatrix}.\end{aligned}$$

令

$$d_0 = -P\nabla f(x_0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

计算

$$\begin{aligned} \lambda &= (A_1 A_1^T)^{-1} A_1 \nabla f(x_0) \\ &= \left(\frac{1}{4} \begin{bmatrix} 1 & -1 \\ -1 & 5 \end{bmatrix} \right) \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 10 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -5 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}. \end{aligned}$$

修正 A_1. 去掉 A_1 中对应 $\lambda_2 = -5$ 的行, 得到

$$\tilde{A}_1 = [2,\ 1].$$

再求投影矩阵

$$\begin{aligned}\tilde{P} &= I - \tilde{A}_1^T(\tilde{A}_1\tilde{A}_1^T)^{-1}\tilde{A}_1 \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix} \left(\begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 2 & 1 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}.\end{aligned}$$

令

$$\tilde{d}_0 = -\tilde{P}\nabla f(x_0) = -\frac{1}{5} \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 10 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}.$$

求步长 α_0:

$$\begin{aligned} &\min\ f(x_0 + \alpha\tilde{d}_0), \\ &\text{s.t.}\ \ 0 \le \alpha \le \bar{\alpha}. \end{aligned} \tag{10.33}$$

由于

$$\begin{aligned}\hat{b} &= b_2 - A_2 x_0 = 0 - \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = -2. \\ \hat{d} &= A_2 \tilde{d}_0 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 4 \end{bmatrix} = -2.\end{aligned}$$

故

$$\bar{\alpha} = \frac{-2}{-2} = 1.$$

这样, 问题 (10.33) 即为

$$\begin{aligned}&\min\ 20\alpha^2 - 20\alpha + 25, \\ &\text{s.t.}\ \ 0 \le \alpha \le 1.\end{aligned}$$

解之得 $\alpha_0 = \dfrac{1}{2} = 0.5$. 从而

$$x_1 = x_0 + \alpha_0 d_0 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} -2 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

第 2 次迭代.

$$\nabla f(x_1) = \begin{bmatrix} 2 \times 1 + 6 \\ 2 \times 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}.$$

$$A_1 = \begin{bmatrix} 2 & 1 \end{bmatrix}, \quad b_1 = \begin{bmatrix} 4 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

因 $E = \emptyset$, 故 $M = A_1$. 计算投影矩阵

$$P = I - A_1^T(A_1A_1^T)^{-1}A_1 = \frac{1}{5}\begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}.$$

令

$$d_2 = -P\nabla f(x_1) = -\frac{1}{5}\begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

计算

$$\lambda = (A_1A_1^T)^{-1}A_1\nabla f(x_1) = \frac{1}{5}\begin{bmatrix} 2 & 1 \end{bmatrix}\begin{bmatrix} 8 \\ 4 \end{bmatrix} = 4 > 0,$$

故 $x_1 = (1, 2)^T$ 是 KT 点. □

10.3 简约梯度法

10.3.1 Wolfe 简约梯度法

Wolfe 于 1963 年针对线性等式约束的非线性优化问题提出一种新的可行方向法, 称之为简约梯度法. 我们来介绍这种方法.

考虑具有约性约束的非线性优化问题

$$\begin{cases} \min\ f(x), \\ \text{s.t. } Ax = b, \\ \qquad x \geq 0, \end{cases} \tag{10.34}$$

其中 $A \in \mathbb{R}^{m\times n}$, 秩为 m, $b \in \mathbb{R}^m$, $f : \mathbb{R}^n \to \mathbb{R}$ 连续可微. 设矩阵 A 的任意 m 个列都线性无关，并且约束条件的每个基本可行点都有 m 个正分量. 那么在此假设下, 每个可行解至少有 m 个正分量, 至多有 $n-m$ 个零分量. 简约梯度法的基本思想把求解线性规划的单纯形法推广到解线性约束的非线性优化问题 (10.34). 先利用等式约束条件消去一些变量，然后

利用降维所形成的简约梯度来构造下降方法, 接着作线性搜索求步长, 重复此过程逐步逼近极小点. 下面依次介绍如何确立简约梯度、如何构造下降方向和计算线搜索的步长上界等.

先介绍简约梯度的确立. 将 A 和 x 进行分解. 不失一般性, 可令

$$A = [B,\ N], \quad x = \begin{bmatrix} x_B \\ x_N \end{bmatrix},$$

其中 B 是 $m \times m$ 可逆矩阵, x_B, x_N 分别是由基变量和非基变量构成的向量. 那么线性约束 $Ax = b$ 就可以表示为

$$Bx_B + Nx_N = b,$$

而 $x \geq 0$ 则变成

$$x_B = B^{-1}b - B^{-1}Nx_N \geq 0, \quad x_N \geq 0.$$

现假设 x 是非退化的可行解, 即 $x_B > 0$. 由于 x_B 可以用 x_N 来表示, 因此 $f(x)$ 可以化成关于 x_N 的函数, 即

$$f(x) = f(x_B, x_N) = f(B^{-1}b - B^{-1}Nx_N, x_N) := F(x_N).$$

称 $n-m$ 维向量 x_N 的函数 $F(x_N)$ 的梯度为 $f(x)$ 的简约梯度, 记为 $r(x_N)$, 即

$$\begin{aligned} r(x_N) &= \nabla_{x_N} F(x_N) = \nabla_{x_N} f(B^{-1}b - B^{-1}Nx_N,\ x_N) \\ &= \nabla_N f(x_B, x_N) - (B^{-1}N)^T \nabla_B f(x_B, x_N), \end{aligned} \tag{10.35}$$

上式中 $\nabla_N = \nabla_{x_N}$, $\nabla_B = \nabla_{x_B}$.

再确定搜索方向, 即怎样确定在点 x_k 处的下降可行方向 d_k, 使得后继点 $x_{k+1} = x_k + \alpha_k d_k$ 是可行点且目标函数值下降. 令

$$d_k = \begin{bmatrix} d_k^B \\ d_k^N \end{bmatrix}.$$

欲使 d_k 为下降可行方向, 需其满足

$$\begin{cases} \nabla f(x_k)^T d_k < 0, \\ A d_k = 0, \\ (d_k)_j \geq 0, \quad \text{若 } (x_k)_j = 0. \end{cases}$$

由等式 $Ad_k = 0$ 得

$$Bd_k^B + Nd_k^N = 0,$$

即

$$d_k^B = -B^{-1}Nd_k^N, \tag{10.36}$$

这表明 d_k^B 是由 d_k^N 确定的. 再由下降性条件 $\nabla f(x_k)^T d_k < 0$ 可得

$$\begin{aligned} \nabla f(x_k)^T d_k &= \nabla_B f(x_k)^T d_k^B + \nabla_N f(x_k)^T d_k^N \\ &= -\nabla_B f(x_k)^T B^{-1} N d_k^N + \nabla_N f(x_k)^T d_k^N \\ &= r(x_k^N)^T d_k^N < 0. \end{aligned} \tag{10.37}$$

由非负性条件可知

$$\text{当 } (x_k^N)_j = 0 \text{ 时，} (d_k^N)_j \geq 0. \tag{10.38}$$

不难发现, 满足 (10.36)∼(10.38) 的 d_k 有许多种选取方法, 其中一种简单的取法为

$$(d_k^N)_j = \begin{cases} -(x_k^N)_j\, r_j(x_k^N), & \text{如果 } r_j(x_k^N) \geq 0 \\ -r_j(x_k^N), & \text{否则}. \end{cases} \tag{10.39}$$

$$d_k = \begin{bmatrix} -B^{-1}Nd_k^N \\ d_k^N \end{bmatrix} = \begin{bmatrix} -B^{-1}N \\ I_{n-m} \end{bmatrix} d_k^N \tag{10.40}$$

.

余下的问题就是确定步长 α_k. 为保持

$$x_{k+1} \geq 0,$$

即

$$(x_{k+1})_j = (x_k)_j + \alpha(d_k)_j \geq 0, \;\; j = 1\cdots, n, \tag{10.41}$$

需确定 α 的取值范围.

注意到, 当 $(d_k)_j \geq 0$ 时, 对任意的 $\alpha \geq 0$, (10.41) 恒成立. 而当 $(d_k)_j < 0$ 时, 应取

$$\alpha \leq \frac{(x_k)_j}{-(d_k)_j}.$$

因此, 令

$$\bar{\alpha} = \begin{cases} +\infty, & d_k \geq 0 \\ \min\left\{ -\dfrac{(x_k)_j}{(d_k)_j} \middle| (d_k)_j < 0 \right\}, & \text{否则}. \end{cases} \tag{10.42}$$

可以证明, 按照上述方式构造的方向 d_k, 若 $d_k \neq 0$, 则它必为下降可行方向, 否则, 相应的 x_k 必为 KT 点.

定理 10.5 设 $A = (B, N)$ 是 $m \times n$ 矩阵, B 是 m 阶非奇异矩阵, $x_k = \begin{bmatrix} x_k^B \\ x_k^N \end{bmatrix}$ 是问题 (10.34) 的可行点, 其中 $x_k^B > 0$ 是相应于 B 的 m 维

向量. 又假定函数 f 在 x_k 处连续可微, d_k 是由 (10.39) 和 (10.40) 定义的方向向量. 那么, 若 $d_k \neq 0$, 则 d_k 是下降可行方向, 且 $d_k = 0$ 的充要条件是 x_k 是 KT 点.

证 由 d_k 的定义, 有

$$Ad_k = Bd_k^B + Nd_k^N = B(-B^{-1}Nd_k^N) + Nd_k^N = 0.$$

又由 (10.38), 当 $(x_N)_j = 0$ 时, $(d_k)_j \geq 0$. 注意到 $x_k^B > 0$, 因此根据引理 10.1, d_k 是可行方向. 此外, 我们有

$$\begin{aligned}\nabla f(x_k)^T d_k &= \nabla_B f(x_k)^T d_k^B + \nabla_N f(x_k)^T d_k^N \\ &= \nabla_B f(x_k)^T(-B^{-1}Nd_k^N) + \nabla_N f(x_k)^T d_k^N \\ &= r(x_k^N)^T d_k.\end{aligned}$$

注意到, 当 $d_k^N \neq 0$ 时, 根据 (10.39) 知 $r(x_k^N)^T d_k < 0$, 因此 d_k 是下降可行方向. 现在证明 $d_k = 0$ 当且仅当 x_k 是 KT 点. 事实上, 我们知道, x_k 是问

题 (10.34) 的 KT 点的充要条件是, 存在乘子 $\lambda = (\lambda_B^T, \lambda_N^T)^T \geq 0$ 及 μ, 使得

$$\nabla f(x_k) - A^T\mu - \lambda = 0,$$
$$\lambda \geq 0,\ \ x_k \geq 0,\ \ \lambda^T x = 0,$$

即

$$\begin{bmatrix} \nabla_B f(x_k) \\ \nabla_N f(x_k) \end{bmatrix} - \begin{bmatrix} B^T \\ N^T \end{bmatrix} \mu - \begin{bmatrix} \lambda_B \\ \lambda_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{10.43}$$

$$\lambda_B^T x_k^B = 0, \quad \lambda_N^T x_k^N = 0. \tag{10.44}$$

我们先设 x_k 是 KT 点, 则上述条件成立. 由于 $x_k^B > 0$ 且 $\lambda_B \geq 0$, 则由 (10.44) 的第一式推出 $\lambda_B = 0$. 从而由 (10.43) 的第一个方程可得

$$\mu = (B^T)^{-1}\nabla_B f(x_k). \tag{10.45}$$

将上式代入 (10.43) 的第二个方程可求得

$$\lambda_N = \nabla_N f(x_k) - (B^{-1}N)^T \nabla_B f(x_k) = r(x_k^N) \geq 0. \tag{10.46}$$

由 (10.46) 及 (10.44) 的第二式可得

$$r(x_k^N)^T x_k^N = 0. \tag{10.47}$$

注意到 $x_k^N \geq 0$, 故由上式可推出

$$r_j(x_k^N)(x_k^N)_j = 0, \quad j = 1, \cdots, n. \tag{10.48}$$

因此, 根据 (10.46), (10.48) 和 (10.36), (10.39) 可推得 $d_k = 0$.

反之, 设 $d_k = 0$, 则 $r(x_k^N)$ 均非负. 令

$$\lambda_N = r(x_k^N) = \nabla_N f(x_k) - (B^{-1}N)^T \nabla_B f(x_k) \geq 0.$$

故由 (10.39) 可知必有 $\lambda_N^T x_k^N = 0$ 成立. 再令

$$\lambda_B = 0, \quad \mu = (B^T)^{-1} \nabla_B f(x_k),$$

则有 $\lambda_B^T x_k^B = 0$ 及 (10.43) 式成立. 故 x_k 是 KT 点. 证毕. $\square$

下面, 我们写出简约梯度法的详细计算步骤如下:

算法 10.4 (Wolfe 简约梯度法)

步 0 初始化. 选取初始可行点 $x_0 \in \mathbb{R}^n$. 令 $k := 0$.

步 1 计算搜索方向. 将 x_k 分解成

$$x_k = \begin{bmatrix} x_k^B \\ x_k^N \end{bmatrix},$$

其中, x_k^B 为基变量, 由 x_k 的 m 个最大分量组成, 这些分量的下标集记作 J_k. 相应地, 将 A 分解成 $A = (B, N)$. 按下式计算 d_k :

$$\begin{aligned} r(x_k^N) &= \nabla_N f(x_k^B, x_k^N) - (B^{-1}N)^T \nabla_B f(x_k^B, x_k^N), \\ (d_k^N)_j &= \begin{cases} -(x_k^N)_j \, r_j(x_k^N), & \text{如果 } r_j(x_k^N) \geq 0, \\ -r_j(x_k^N), & \text{否则}, \end{cases} \\ d_k &= \begin{bmatrix} d_k^B \\ d_k^N \end{bmatrix} = \begin{bmatrix} -B^{-1}N \\ I_{n-m} \end{bmatrix} d_k^N. \end{aligned}$$

步 2　检验终止准则. 若 $d_k = 0$, 则 x_k 为 KT 点, 停算. 否则, 转步 3.

步 3　计算步长上界 $\bar{\alpha}$:

$$\bar{\alpha} = \begin{cases} +\infty, & d_k \geq 0 \\ \min\left\{ -\dfrac{(x_k)_j}{(d_k)_j} \middle| (d_k)_j < 0 \right\}, & \text{否则}. \end{cases}$$

步 4　进行一维搜索. 求解下面的一维极小化问题得步长 α_k :

$$\begin{cases} \min f(x_k + \alpha d_k), \\ \text{s.t. } 0 \leq \alpha \leq \bar{\alpha}. \end{cases}$$

令 $x_{k+1} := x_k + \alpha_k d_k$.

步 5　修正基变量. 若 $x_{k+1}^B > 0$, 则基变量不变. 否则, 若有 j 使得 $(x_{k+1}^B)_j = 0$, 则将 $(x_{k+1}^B)_j$ 换出基, 而以 $(x_{k+1}^N)_j$ 中最大分量换入基, 构成新的基向量 x_{k+1}^B 和 x_{k+1}^N.

步 6　令 $k := k + 1$, 转步 1.

例 10.3 用 Wolfe 简约梯度法重新求解例 10.1, 即

$$
\begin{aligned}
\min\ & x_1^2 + x_2^2 - 2x_1 - 4x_2 + 3, \\
\text{s.t.}\ & -2x_1 + x_2 \geq -1, \\
& -x_1 - x_2 \geq -2, \\
& x_1 \geq 0,\ x_2 \geq 0.
\end{aligned}
$$

解 首先, 引入松弛变量 $x_3, x_4 \geq 0$, 将原问题转化为等价的"标准形式":

$$
\begin{aligned}
\min\ & x_1^2 + x_2^2 - 2x_1 - 4x_2 + 3, \\
\text{s.t.}\ & \begin{cases} 2x_1 - x_2 + x_3 = 1, \\ x_1 + x_2 + x_4 = 2, \\ x_1 \geq 0,\ x_2 \geq 0, \\ x_3 \geq 0,\ x_4 \geq 0. \end{cases}
\end{aligned}
$$

首先写出约束矩阵和梯度

$$A = \begin{bmatrix} 2 & -1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}, \quad \nabla f(x) = \begin{bmatrix} 2x_1 - 2 \\ 2x_2 - 4 \\ 0 \\ 0 \end{bmatrix}.$$

取初始可行点 $x_0 = (0, 0, 1, 2)^T$.

第 1 次迭代. $k = 0$. $J_0 = \{3, 4\}$. $\nabla f(x_0) = (-2, -4, 0, 0)^T$. 确立 B, N 等得

$$x_0^B = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad x_0^N = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad N = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}.$$

计算

$$B^{-1}N = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}, \quad \nabla_N f(x_0) = \begin{bmatrix} -2 \\ -4 \end{bmatrix}, \quad \nabla_B f(x_0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

计算简约梯度 $r(x_0^N)$ 得

$$\begin{aligned} r(x_0^N) &= \nabla_N f(x_0) - (B^{-1}N)^T \nabla_B f(x_0) \\ &= \begin{bmatrix} -2 \\ -4 \end{bmatrix} - \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}^T \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \end{bmatrix}. \end{aligned}$$

当 $r_j(x_0^N) \leq 0$ 时, 取 $(d_0^N)_j = -r_j(x_0^N)$, 故

$$d_0^N = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \quad d_0^B = -B^{-1}Nd_0^N = -\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ -6 \end{bmatrix}.$$

从而 $d_0 = (2, 4, 0, -6)^T$.

求步长上界

$$\bar{\alpha} = -\frac{2}{-6} = \frac{1}{3}.$$

从 x_0 出发, 沿 d_0 搜索:

$$x_0 + \alpha d_0 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \end{bmatrix} + \alpha \begin{bmatrix} 2 \\ 4 \\ 0 \\ -6 \end{bmatrix} = \begin{bmatrix} 2\alpha \\ 4\alpha \\ 1 \\ 2-6\alpha \end{bmatrix},$$

$$f(x_0 + \alpha d_0) = 20\alpha^2 - 20\alpha + 3.$$

求解一维极小问题

$$\min 20\alpha^2 - 20\alpha + 3,$$

$$\text{s.t. } 0 \le \alpha \le \frac{1}{3},$$

得 $\alpha_0 = \frac{1}{3}$, 从而 $x_1 = x_0 + \alpha_0 d_0 = \left(\frac{2}{3}, \frac{4}{3}, 1, 0\right)^T$.

第 2 次迭代. $J_1=\{2,3\}$. $\nabla f(x_1)=\left(-\frac{2}{3},-\frac{4}{3},0,0\right)^T$.

$$x_1^B=\begin{bmatrix}\frac{4}{3}\\1\end{bmatrix},\quad x_1^N=\begin{bmatrix}\frac{2}{3}\\0\end{bmatrix},\quad B=\begin{bmatrix}-1&1\\1&0\end{bmatrix},\quad N=\begin{bmatrix}2&0\\1&1\end{bmatrix}.$$

计算

$$B^{-1}N=\begin{bmatrix}1&1\\3&1\end{bmatrix},\quad \nabla_N f(x_1)=\begin{bmatrix}-\frac{2}{3}\\0\end{bmatrix},\quad \nabla_B f(x_1)=\begin{bmatrix}-\frac{4}{3}\\0\end{bmatrix}.$$

计算简约梯度 $r(x_1^N)$ 得

$$\begin{aligned}r(x_1^N)&=\nabla_N f(x_1)-(B^{-1}N)^T\nabla_B f(x_1)\\&=\begin{bmatrix}-\frac{2}{3}\\0\end{bmatrix}-\begin{bmatrix}1&1\\3&1\end{bmatrix}^T\begin{bmatrix}-\frac{4}{3}\\0\end{bmatrix}=\begin{bmatrix}\frac{2}{3}\\\frac{4}{3}\end{bmatrix}.\end{aligned}$$

当 $r_j(x_1^N) > 0$ 时, 取 $(d_1^N)_j = -(x_1^N)_j r_j(x_1^N)$, 故

$$d_1^N = \begin{bmatrix} -\frac{4}{9} \\ 0 \end{bmatrix}, \quad d_1^B = -(B^{-1}N)d_1^N = -\begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} -\frac{4}{9} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{4}{9} \\ \frac{4}{3} \end{bmatrix}.$$

从而

$$d_1 = \left(-\frac{4}{9}, \frac{4}{9}, \frac{4}{3}, 0\right)^T.$$

求步长上界

$$\bar{\alpha} = -\frac{2/3}{-4/9} = \frac{3}{2}.$$

从 x_1 出发, 沿 d_1 搜索:

$$x_1 + \alpha d_1 = \begin{bmatrix} 2/3 \\ 4/3 \\ 1 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} -4/9 \\ 4/9 \\ 4/3 \\ 0 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 6-4\alpha \\ 12+4\alpha \\ 9+8\alpha \\ 0 \end{bmatrix},$$

$$f(x_1 + \alpha d_1) = \frac{32}{81}\alpha^2 - \frac{24}{81}\alpha - \frac{117}{81}.$$

求解一维极小问题

$$\min \frac{32}{81}\alpha^2 - \frac{24}{81}\alpha - \frac{117}{81},$$
$$\text{s.t. } 0 \le \alpha \le \frac{3}{2},$$

得 $\alpha_1 = \frac{3}{8}$, 从而 $x_2 = x_1 + \alpha_1 d_1 = \left(\frac{1}{2}, \frac{3}{2}, \frac{4}{3}, 0\right)^T$.

第 3 次迭代. $J_2=\{2,3\}$. $\nabla f(x_2)=(-1,-1,0,0)^T$.

$$x_2^B=\begin{bmatrix}\frac{3}{2}\\ \frac{4}{3}\end{bmatrix},\quad x_2^N=\begin{bmatrix}\frac{1}{2}\\ 0\end{bmatrix},\quad B=\begin{bmatrix}-1 & 1\\ 1 & 0\end{bmatrix},\quad N=\begin{bmatrix}2 & 0\\ 1 & 1\end{bmatrix}.$$

计算

$$B^{-1}N=\begin{bmatrix}1 & 1\\ 3 & 1\end{bmatrix},\quad \nabla_N f(x_1)=\begin{bmatrix}-1\\ 0\end{bmatrix},\quad \nabla_B f(x_1)=\begin{bmatrix}-1\\ 0\end{bmatrix}.$$

计算简约梯度 $r(x_1^N)$ 得

$$\begin{aligned} r(x_1^N) &= \nabla_N f(x_1)-(B^{-1}N)^T\nabla_B f(x_1)\\ &= \begin{bmatrix}-1\\ 0\end{bmatrix}-\begin{bmatrix}1 & 1\\ 3 & 1\end{bmatrix}^T\begin{bmatrix}-1\\ 0\end{bmatrix}=\begin{bmatrix}0\\ 1\end{bmatrix}. \end{aligned}$$

当 $r_j(x_1^N) > 0$ 时, 取 $(d_1^N)_j = -(x_1^N)_j r_j(x_1^N)$, 而 $r_j(x_1^N) = 0$ 时, 取 $(d_1^N)_j = -r_j(x_1^N)$. 故

$$d_2^N = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad d_2^B = -(B^{-1}N)d_2^N = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

从而 $d_2 = (0,0,0,0)^T$. 根据定理 10.5, x_2 即为 KT 点, 故 $x^* = \left(\frac{1}{2}, \frac{3}{2}\right)^T$ 是原问题的全局极小点. □

10.3.2　广义简约梯度法

Abadie 和 Carpentier 于 1969 年将 Wolfe 简约梯度法推广到一般非线性约束的情形, 提出了所谓的广义简约梯度法. 设一般非线性约束优化问

题为

$$\begin{cases} \min f(x), \\ \text{s.t. } h_i(x) = 0, \; i \in E = \{1, \cdots, l\}, \\ \qquad g_i(x) \geq 0, \; i \in I = \{1, \cdots, m\}, \end{cases}$$

其中 f, $h_i\,(i \in E)$, $g_i\,(i \in I)$ 是连续可微的函数.

假设 x_k 是第 k 次可行迭代点, 记 $I_k = E \cup \{i \,|\, g_i(x_k) = 0\}$,并设

$$c(x_k) = \big(h_1(x_k), \cdots, h_l(x_k); g_i(x_k)\,(i \in I_k \backslash E)\big)^T.$$

则在第 k 步可考虑子问题:

$$\begin{cases} \min f(x), \\ \text{s.t. } c(x_k) = 0. \end{cases} \tag{10.49}$$

现在我们讨论如何确立简约梯度 $r(x_k^N)$. 不失一般性, 设 $I_k = \{1, 2, \cdots, s\}$, $s \geq l$, 且约束函数的 Jacobi 矩阵

$$\left[\nabla h_1(x_k), \cdots, \nabla h_l(x_k),\ \nabla g_i(x_k)\,(i \in I_k \backslash E)\right]^T$$

行满秩 (不妨设其前 s 列构成的方阵非奇异). 那么可以解出前 s 个变量, 即可以用其余的 $n-s$ 个变量来表示这 s 个变量. 通常称这 s 个变量组成的子向量为基向量, 记为 x_k^B, 其余 $n-s$ 个变量组成的子向量为非基向量, 记为 x_k^N. 为方便计, 去掉下标 k, 并记 $s \times s$ 矩阵

$$\nabla_B c(x) = \left[\nabla_B h_1(x), \cdots, \nabla_B h_l(x), \nabla_B g_1(x), \cdots, \nabla_B g_{s-l}(x)\right]^T,$$

其中

$$c(x) = \big(h_1(x), \cdots, h_l(x), g_1(x), \cdots, g_{s-l}(x)\big)^T := \big(c_1(x), \cdots, c_s(x)\big)^T$$

及

$$\nabla_B c_i(x) = \begin{bmatrix} \dfrac{\partial c_i(x)}{\partial x_1} \\ \vdots \\ \dfrac{\partial c_i(x)}{\partial x_s} \end{bmatrix}, \ i = 1, 2, \cdots, s.$$

由假设可知 $\nabla_B c(x)$ 非奇异. 再记矩阵

$$\nabla_N c(x) = \Big[\nabla_N c_1(x),\ \nabla_N c_2(x),\ \cdots,\ \nabla_N c_s(x)\Big]^T \in \mathbb{R}^{s\times(n-s)},$$

其中

$$\nabla_N c_i(x) = \begin{bmatrix} \dfrac{\partial c_i(x)}{\partial x_{s+1}} \\ \vdots \\ \dfrac{\partial c_i(x)}{\partial x_n} \end{bmatrix}, \ i = 1, 2, \cdots, s.$$

在 x 的某邻域内, 由非线性方程组 $c(x) = c(x_B, x_N) = 0$ 可以确定 x_B 为 x_N 的函数, 即 $x_B = \varphi(x_N)$.

对等式 $c(x) = c(x_B, x_N) = 0$ 两边关于 x_N 求梯度得 $J_{BN}(x_N)\nabla_B c(x) + \nabla_N c(x) = 0$, 其中

$$J_{BN}(x_N) = \left[\frac{\partial(x_1, \cdots, x_s)}{\partial(x_{s+1}, \cdots, x_n)}\right]^T = \begin{bmatrix} \dfrac{\partial x_1}{\partial x_{s+1}} & \dfrac{\partial x_2}{\partial x_{s+1}} & \cdots & \dfrac{\partial x_s}{\partial x_{s+1}} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial x_1}{\partial x_n} & \dfrac{\partial x_2}{\partial x_n} & \cdots & \dfrac{\partial x_s}{\partial x_n} \end{bmatrix}.$$

从而, 我们有

$$J_{BN}(x_N) = -\nabla_N c(x)\left[\nabla_B c(x)\right]^{-1}. \tag{10.50}$$

注意到 $f(x) \equiv f(\varphi(x_N), x_N)$, 对其求关于 x_N 的梯度 (即简约梯度) 得

$$r(x_N) = \nabla_N f(x) + J_{BN}(x_N)\nabla_B f(x).$$

将 (10.50) 代入上式即得

$$r(x_N) = \nabla_N f(x) - \nabla_N c(x)\left[\nabla_B c(x)\right]^{-1}\nabla_B f(x). \tag{10.51}$$

现在设下降可行方向为 $d = (d_B^T, d_N^T)^T$, 则由下降可行条件知 d 应满足

$$\nabla f(x)^T d < 0, \qquad \nabla c(x)^T d = 0.$$

由 $\nabla c(x)^T d = 0$ 可得

$$\nabla_B c(x)^T d_B + \nabla_N c(x)^T d_N = 0.$$

于是有

$$d_B = -[\nabla_B c(x)^T]^{-1} \nabla_N c(x)^T d_N = J_{BN}(x_N)^T d_N. \tag{10.52}$$

又由 $\nabla f(x)^T d < 0$, 得

$$\nabla_B f(x)^T d_B + \nabla_N f(x)^T d_N < 0.$$

将 d_B 的表达式代入上式得

$$\nabla_N f(x)^T d_N - \nabla_B f(x)^T [\nabla_B c(x)^T]^{-1} \nabla_N c(x)^T d_N < 0,$$

即

$$r(x_N)^T d_N < 0.$$

因此, d_N 的一种简单的选取方法是 $d_N = -r(x_N)$. 至此, 我们在上述的推导过程中以 $x = x_k$ 代入得 x_k 处的简约梯度为 $r(x_k^N)$, 下降可行方向为

$$d_k = \begin{bmatrix} d_k^B \\ d_k^N \end{bmatrix} = \begin{bmatrix} -J_{BN}(x_k^N)^T r(x_k^N) \\ -r(x_k^N) \end{bmatrix} = \begin{bmatrix} -J_{BN}(x_k^N)^T \\ -I_{n-s} \end{bmatrix} r(x_k^N) \tag{10.53}$$

最后确定步长 α_k. 可以通过求解下述一维极小问题

$$\begin{cases} \min f(x_k + \alpha d_k), \\ \text{s.t. } c_i(x_k + \alpha d_k) = 0, \ i \in E, \\ \qquad c_i(x_k + \alpha d_k) \geq 0, \ i \in I \end{cases} \tag{10.54}$$

获得搜索步长 α_k, 然后令 $x_{k+1} := x_k + \alpha_k d_k$ 即得到后继可行迭代点 x_{k+1}.

最后讨论拉格朗日乘子的估计. 由定理 ?? 知, 在极小点 x^* 处成立

$$\nabla f(x^*) = \sum_{i \in E} \mu_i^* \nabla h_i(x^*) + \sum_{i \in I^*} \lambda_i^* \nabla g_i(x^*), \tag{10.55}$$

其中 $\lambda_i^* \geq 0,\ i \in I^* = \{i \,|\, g_i(x^*) = 0\}$. 记

$$\nabla c(x^*) = \big[\nabla h_1(x^*), \cdots, \nabla h_l(x^*),\ \nabla g_i(x^*),\ (i \in I^*)\big],$$
$$\nu^* = \big((\mu^*)^T,\ (\lambda^*)^T\big)^T = \big(\mu_1, \cdots, \mu_l, \lambda_i\,(i \in I^*)\big)^T.$$

那么 (10.55) 可以写成 $\nabla f(x^*) = \nabla c(x^*)\nu^*$. 由广义逆知识可得其极小最小二乘解 $\nu^* = [\nabla c(x^*)]^+ \nabla f(x^*)$. 因此, 计算相应的乘子估计

$$\nu_k = (\mu_k^T, \lambda_k^T)^T = [\nabla c(x_k)]^+ \nabla f(x_k). \tag{10.56}$$

下面给出广义简约梯度法的详细计算步骤.

算法 10.5 (**广义简约梯度法**)

步 0 选取初始值. 给定初始可行点 $x_0 \in \mathbb{R}^N$, $0 \leq \varepsilon \ll 1$. 令 $k := 0$.

步 1 检验终止条件. 确定基变量 x_k^B 和非基变量 x_k^N. 由 (10.51) 计算简约梯度 $r(x_k^N)$. 若 $\|r(x_k^N)\| \leq \varepsilon$, 则 x_k 为近似极小点, 停算.

步 2 确定搜索方向. 由 (10.53) 计算下降可行方向 d_k.

步 3 进行线搜索. 解子问题 (10.54) 得步长因子 α_k. 令 $x_{k+1} := x_k + \alpha_k d_k$.

步 4　修正有效集. 先求 x_{k+1} 处的有效集, 设为 $\bar{I}_{k+1}$. 由 (10.56) 计算 λ_k. 若 $\lambda_{k+1} \geq 0$, 则 $I_{k+1} = \bar{I}_{k+1}$. 否则, I_{k+1} 是 $\bar{I}_{k+1}$ 中删去 λ_{k+1} 最小分量所对应的约束指标集.

步 5　令 $k := k + 1$, 转步 1.

注 (1) 在算法 10.5 的步 2 中, 当 $\|r(x_k^N)\| \leq \varepsilon$ 时, 实际还需要判别对应于不等式约束的拉格朗日乘子的非负性, 若不满足还需进行改进. (2) 广义简约梯度法通过消去某些变量在降维空间中运算, 能够较快确定最优解, 可用来求解大型问题, 因而它是目前求解非线性优化问题的最有效的方法之一.