

文章编号: 1001-0920(2008)04-0409-06

改进 Lagrange 乘子法及收敛性分析

黄远灿

(北京理工大学 智能机器人研究所, 北京 100081)

摘要: 将与不等式约束相关的乘子重新定义为原乘子的正定函数, 则 Karush-Kuhn-Tucker 必要条件中关于不等式约束乘子的非负约束可以去掉, 并能构造出直接处理不等式约束的 Lagrange 乘子法. 分析了算法的收敛性, 利用 LaSalle 不变集原理揭示其稳定机制, 并讨论如何减弱收敛条件和扩大收敛域.

关键词: 非线性规划; Lagrange 乘子法; 不等式约束; 算法收敛性; LaSalle 不变集原理
中图分类号: O323 **文献标识码:** A

Modified Lagrange multiplier method and its convergence analysis

HUANG Yuan-can

(Intelligent Robotics Institute, Beijing Institute of Technology, Beijing 100081, China. E-mail: yuancanhuang@bit.edu.cn)

Abstract: By redefining multiplier associated with inequality constraint as a positive definite function of the originally defined multiplier, nonnegative constraint imposed on inequality constraints in Karush-Kuhn-Tucker necessary conditions is removed, and a modified Lagrange multiplier method, which may handle inequality constraints directly, is constructed. Then its convergence is analyzed rigorously. By using LaSalle invariance principle, the underlying mechanism that attains the algorithmic convergence is uncovered. Some measures for relaxing the convergence conditions and enlarging the attractive domain are discussed.

Key words: Nonlinear programming; Lagrange multiplier method; Inequality constraints; Convergence; LaSalle invariance principle

1 引言

非线性规划中处理不等式约束的方法分为 3 类: 1) 直接方法, 即直接在可行域内搜索最优解; 2) 通过求解一系列简单的约束优化问题来逼近原问题的解, 如 SQP (successive quadratic programming); 3) 将约束优化问题转化为等价的无约束优化问题, 通过求解一个或一系列无约束优化问题, 获得或逼近问题的最优解, 如罚函数法和 Lagrange 乘子法^[1-8]. 其中 Lagrange 乘子法应用非常广泛. 若使用处理等式约束的方法来处理不等式约束, 在 Lagrange 乘子法中需添加松弛变量, 将不等式约束变成等式约束. 尽管通过简单推导可以去掉松弛变量, 但获得的增广 Lagrange 函数的光滑性比目标函数和约束函数要差^[1].

如果重新定义不等式约束的乘子为原乘子的正定函数, 则 Karush-Kuhn-Tucker 必要条件中关于不等式约束乘子的非负约束可去掉. 利用该技术, 可

以构造出一种直接对不等式约束进行处理的改进 Lagrange 乘子法. 与 Hestenes-Rockafellar-Powell 的 Lagrange 乘子法比较, 新算法有两个优点: 首先, 可直接移植有关等式约束的方法处理不等式约束, 不再需要添加松弛变量将不等式约束转换成等式约束; 其次, 增广 Lagrange 函数具有与目标函数和约束函数相同的光滑性. 因为常用的无约束优化算法 (如共轭梯度法、拟 Newton 法、Newton 法等) 的计算效率依赖于极小化函数的二阶导数的连续性^[3,4], 因此, 改进 Lagrange 乘子法能提高计算效率.

本文给出算法的构造方法和工作机理; 利用隐函数定理分析算法的局部收敛性; 根据 LaSalle 不变集原理揭示算法的稳定机制, 以及讨论如何减弱收敛条件和扩大收敛域.

2 改进 Lagrange 乘子法及其机理

考虑如下只有不等式约束的非线性规划问题:

收稿日期: 2007-01-16; 修回日期: 2007-06-11.

基金项目: 国家 863 计划项目 (2006AA04Z264).

作者简介: 黄远灿 (1969—), 男, 贵州盘县人, 副教授, 博士, 从事机器人控制、神经网络等研究.

$$\begin{aligned} \min f(x), x \in R^n, \\ \text{s.t. } g(x) \leq 0, \end{aligned} \quad (1)$$

式中 $f: R^n \rightarrow R$ 和 $g: R^n \rightarrow R^m$ 分别是二阶连续可微的标量和向量函数.

令 x^* 满足约束条件, $I(x^*)$ 表示对应 $g_i(x^*) = 0$ 的 i 的指标集, 即

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

如果梯度 $\nabla g_i(x^*)$, $i \in I(x^*)$ 线性无关, 则称 x^* 为正则点.

定义 Lagrange 函数为

$$L(x, v) = f(x) + \sum_{i=1}^m v_i g_i(x), \quad (2)$$

式中 $v_i = u_i^2$, u_i 表示原 Lagrange 乘子. 于是, Karush-Kuhn-Tucker 定理中的非负约束可以去掉.

定理 1 (Karush-Kuhn-Tucker 定理)^[2-5] 设 x^* 是问题(1)的局部最优解和正则点, 那么存在唯一向量 v^* , 使得

$$\begin{aligned} \nabla_x L(x^*, v^*) &= 0, \\ v_i^* g_i(x^*) &= 0, \quad i = 1, 2, \dots, m, \\ g_i(x^*) &\leq 0, \quad i = 1, 2, \dots, m. \end{aligned}$$

下面, 介绍一些基本概念. 如果约束 $g_i(x^*) = 0$, 则称为起作用的约束; 反之, 称为不起作用的约束; (x^*, v^*) 称为问题(1)的 KKT 对.

定义增广 Lagrange 函数为

$$\begin{aligned} L_c(x, v) &= \\ f(x) + \sum_{i=1}^m v_i g_i(x) + \frac{c}{2} \sum_{i=1}^m v_i g_i^2(x) &= \\ L(x, v) + \frac{c}{2} \sum_{i=1}^m v_i g_i^2(x), \end{aligned} \quad (3)$$

式中 c 为正罚参数.

通过直接计算, 增广 Lagrange 函数对 x 的梯度和 Hess 阵分别为

$$\begin{aligned} \nabla_x L_c(x, v) &= \\ \nabla_x L(x, v) + c \sum_{i=1}^m v_i g_i(x) \nabla g_i(x), \quad (4) \\ \nabla_{xx}^2 L_c(x, v) &= \\ \nabla_{xx}^2 L(x, v) + c \left[\sum_{i=1}^m v_i g_i(x) \nabla^2 g_i(x) + \right. \\ \left. \sum_{i=1}^m v_i \nabla g_i(x) \nabla g_i^T(x) \right]. \quad (5) \end{aligned}$$

改进的 Lagrange 乘子法如下:

Step1: 给定初始点 x^0 , 初始罚参数 $c_0 > 0$, 罚参数的上界 c_{\max} , 初始乘子 $v^0 > 0$ 和收敛容限 ϵ .

Step2: 极小化 $L_{c_k}(\cdot, v^k)$, 求出极小点 x^k .

Step3: 取 $c_{k+1} > c_k$, 且满足约束 $c_{k+1} \leq c_{\max}$.

Step4: 根据下式修正乘子:

$$v_i^{k+1} = v_i^k [1 + c_{k+1} g_i(x^k)], \quad i = 1, 2, \dots, m.$$

如果 $g_i(x^k) < 0$, 必须适当选择 c_{k+1} , 使得 $0 < 1 + c_{k+1} g_i(x^k) < 1$.

Step5: 如果 $\text{diag}(v^k) g(x^k) \leq -\epsilon$, 停止; 否则, 转到 Step2, 重复计算.

注 1 改进乘子法与原乘子法的区别在于, Step4 中不等式约束的乘子迭代公式不同. 该公式可以用下式替代:

$$v_i^{k+1} = v_i^k e^{c_{k+1} g_i(x^k)}, \quad i = 1, 2, \dots, m. \quad (6)$$

使用式(6)的优点是, 当 $g_i(x^k) < 0$ 时, 无论如何选择 c_{k+1} , 总能保证 $0 < e^{c_{k+1} g_i(x^k)} < 1$. 当指数函数中的指数项非常小时, 有 $e^{c_{k+1} g_i(x^k)} \approx 1 + c_{k+1} g_i(x^k)$. 因此, 对起作用的约束而言, 在 KKT 对的充分小邻域内, 利用指数函数或其一阶近似公式更新乘子, 算法的渐近行为是一样的.

注 2 罚参数 c_k 的选择将会对算法的收敛性产生影响, 一般应选择 c_k 大于某正常数.

根据乘子迭代公式, 容易算出

$$v_i^k = v_i^0 \prod_{l=0}^{k-1} [1 + c_{l+1} g_i(x^l)], \quad i = 1, 2, \dots, m. \quad (7)$$

式(7)表明, 如果约束满足, 则对应乘子或者持续衰减, 或者保持不变; 如果约束不满足, 则对应乘子持续增长. 因此, 如果 x^k 位于问题可行域之外, 即至少存在某个 i , 使得 $g_i(x^k) > 0$, 则对应乘子指数增加; 如果 x^k 在可行域内, 则乘子指数衰减到零或保持不变.

总而言之, 如果 x^k 位于问题可行域之外, 增广 Lagrange 函数中的惩罚项将不断增大, 通过 Step2 的极小化, 驱使 x^k 逐渐接近可行域. 如果算法没有震荡或不产生混沌, 当这样的过程重复有限多次后, 所有约束都将满足.

3 基本结论

假设 1 对任意 c_k 和 v^k , $L_{c_k}(\cdot, v^k)$ 的极小点 x^k 存在.

根据动力学的观点, 改进 Lagrange 乘子法可以看成如下离散时间动态系统:

$$\begin{aligned} x^{k+1} &= A(k, x^k, v^k), \\ v_i^{k+1} &= v_i^k [1 + c_{k+1} g_i(x^k)], \\ i &= 1, 2, \dots, m, \end{aligned} \quad (8)$$

式中 A 是第 k 步的极小点 x^k 和乘子到第 $k+1$ 步的极小点的映射 x^{k+1} (可能是时变多值映射). 算法能展现出复杂的动力学行为, 如混沌、震荡、发散等.

命题 1 令假设 1 成立和算法收敛, 则其聚点 (x^*, v^*) 为问题(1)的 KKT 对.

证明 首先,用反证法说明 x^* 是问题(1)的可行解. 如果 x^* 不是可行解,则至少有一个 $g_i(x^*) > 0$. 由连续性假设,可以找到一个 $k, \forall k > k$, 有 $g_i(x^k) > 0$. 根据乘子迭代公式, $\lim_k L_{c_k}(x^k, v^k)$, 与假设 1 矛盾.

其次,如果 v_i^* 对应于不起作用的乘子,由于 x^* 是可行解,可以推出 $g_i(x^*) < 0$. 于是,存在 k 和小于 1 的正数 ϵ , $\forall k > k$, 有

$$v_i^* = \prod_{l=0}^k [1 + c_{l+1} g_i(x^l)] \prod_{j=k}^l [1 + c_{j+1} g_i(x^j)] v_i^0 = 0. \tag{9}$$

于是,对起作用或不起作用的约束,都有

$$v_i^* g_i(x^*) = 0, \quad i = 1, 2, \dots, m. \tag{10}$$

根据算法的 Step2 可知, $\forall k$, 有

$$\nabla_x L_{c_k}(x^k, v^k) = 0. \tag{11}$$

由 $L_{c_k}(\cdot, \cdot)$ 的连续性,可得

$$\lim_k \nabla_x L_{c_k}(x^k, v^k) = \nabla_x L_c(x^*, v^*) = 0. \tag{12}$$

根据式(4)和(10),可以推出

$$\nabla_x L(x^*, v^*) = 0. \tag{13}$$

综上,命题 1 得证.

4 算法收敛性分析

假设 2 设 x^* 是问题(1)的正则点,且存在向量 v^* 满足

$$\begin{aligned} \nabla_x L(x^*, v^*) &= 0, \\ v_i^* g_i(x^*) &= 0, \quad i = 1, 2, \dots, m, \\ g(x^*) &\leq 0 \end{aligned} \tag{14}$$

和对任何使 $\nabla g_i(x^*)^T y = 0, i \in I(x^*)$ 的非零向量 y , 有

$$y^T \left[\nabla^2 f(x^*) + \sum_{i=1}^m v_i^* \nabla^2 g_i(x^*) \right] y > 0.$$

另外, v_i^* 满足严格互补假设

$$v_i^* > 0, \quad \forall i \in I(x^*).$$

条件 1(二阶充分条件)^[1-6,8] 如果假设 2 满足,则 x^* 为问题(1)的严格局部极小点.

为了证明收敛性定理,先介绍如下引理^[1,11].

引理 1 令 P 为 $n \times n$ 对称阵, Q 为 $n \times n$ 半正定对称阵. 如果对任何满足 $x^T Qx = 0$ 的非零向量 x , 有 $x^T Px > 0$, 那么,存在标量 $c > 0$, 使

$$P + cQ > 0.$$

令假设 2 成立,则

$$\nabla_x L_c(x^*, v^*) = 0, \quad \forall c > 0. \tag{15}$$

根据引理 1, 存在 $\bar{c} > 0$, 使得

$$\nabla_{xx}^2 L_{\bar{c}}(x^*, v^*) =$$

$$\begin{aligned} &\nabla_{xx}^2 L(x^*, v^*) + \\ &\bar{c} \sum_{i=1}^m v_i^* \nabla g_i(x^*) \nabla g_i(x^*)^T > 0. \end{aligned} \tag{16}$$

式(15)和(16)说明, $\nabla c > \bar{c}, x^*$ 是 $L_c(\cdot, v^*)$ 的严格局部极小点.

设所有不等式约束均起作用,与等式情形的证明类似^[1,2], 可得下列结论.

命题 2 若假设 1 和假设 2 成立,且 \bar{c} 是使下式成立的正数:

$$\nabla_{xx}^2 L_{\bar{c}}(x^*, v^*) > 0, \tag{17}$$

那么,存在正数 ϵ 和 B , 使得:

1) 定义集合 $D \subset R^{m+1}$ 为

$$D = \{(v, c) \mid v - v^* < \epsilon, \bar{c} - \epsilon < c < c_{\max}\}. \tag{18}$$

于是, $\forall (v, c) \in D$, 极小化问题

$$\begin{aligned} \min L_c(x, v), \quad x \in R^n, \\ \text{s.t. } x \in S(x^*; \epsilon) \end{aligned} \tag{19}$$

有唯一的极小点 $x(v, c)$, 且在 D 的内部连续可微, 其中 $S(x^*; \epsilon)$ 表示 x^* 的 ϵ -邻域.

2) $\forall (v, c) \in D$, 有

$$\begin{aligned} \|x(v, c) - x^*\| &\leq B \|v - v^*\| / c, \\ \|\tilde{v}(v, c) - v^*\| &\leq B \|v - v^*\| / c. \end{aligned} \tag{20}$$

其中

$$\begin{aligned} \tilde{v}(v, c) &= v_i [1 + c g_i(x(v, c))], \\ &i = 1, 2, \dots, m. \end{aligned}$$

此时, c 的选择与算法中 Step4 的要求一样.

3) $\forall (v, c) \in D$, 矩阵 $\nabla_{xx}^2 L_c[x(v, c), v]$ 正定,

$\nabla g[x(v, c)]$ 列满秩.

4) $\forall (v, c) \in D$, 有

$$\begin{aligned} \tilde{v}(v, c) - v^* &= \\ &\int_0^1 N_c[v^* + (v - v^*)] (v - v^*) \, d. \end{aligned} \tag{21}$$

式中 $m \times m$ 阵 N_c 为

$$\begin{aligned} N_c(u) &= \\ &I_m - c(v) \nabla g[x(u, c)]^T \times \\ &\{ \nabla_{xx}^2 L_c[x(u, c), u] \}^{-1} \nabla g[x(u, c)], \end{aligned}$$

$= \text{diag}(v)$ 表示向量对角化算子.

证明 考虑如下方程:

$$\begin{aligned} \nabla f(x) + \nabla g(x) \tilde{v} &= 0, \\ (v) g(x) + (v - \tilde{v}) / c &= 0. \end{aligned} \tag{22}$$

令 $t = (v - v^*) / c, \epsilon = 1/c$, 则上式可重写为

$$\begin{aligned} \nabla f(x) + \nabla g(x) \tilde{v} &= 0, \\ (v) g(x) + t + (v^* - \tilde{v}) &= 0. \end{aligned} \tag{23}$$

当 $t = 0$ 时, $x = x^*$ 和 $\tilde{v} = v^*$ 为方程(23)的解, 且在该点的 Jacob 矩阵为

$$\begin{bmatrix} \nabla_{xx}^2 L(x^*, v^*) & \nabla g(x^*) \\ (v^*) \nabla g(x^*)^T & -I_m \end{bmatrix}. \quad (24)$$

可以表明, $\forall [1/c_{\max}, 1/\bar{c}]$, 该矩阵非奇异. 若不然, 设存在非零向量 $(y^T, w^T)^T$, 使

$$\begin{bmatrix} \nabla_{xx}^2 L(x^*, v^*) & \nabla g(x^*) \\ (v^*) \nabla g(x^*)^T & -I_m \end{bmatrix} \begin{bmatrix} y \\ w \end{bmatrix} = 0. \quad (25)$$

上式等价于

$$\begin{aligned} \nabla_{xx}^2 L(x^*, v^*) y + \nabla g(x^*) w &= 0, \\ (v^*) \nabla g(x^*)^T y - w &= 0. \end{aligned} \quad (26)$$

从式(26)的第2式解出 w , 代入第1式, 得

$$\begin{aligned} [\nabla_{xx}^2 L(x^*, v^*) + \\ (1/c) \nabla g(x^*) (v^*) \nabla g(x^*)^T] y = 0, \end{aligned}$$

即

$$\nabla_{xx}^2 L_c(x^*, v^*) y = 0.$$

因为 $c \in [1/c_{\max}, 1/\bar{c}]$ 时, $\nabla_{xx}^2 L_c(x^*, v^*)$ 正定, 所以, $\forall [1/c_{\max}, 1/\bar{c}]$, 推出 $y = 0$. 再由式(26)的第2式容易得出 $w = 0$, 与前面的非零假设矛盾.

定义闭集 $K = \{(t, v) \mid [1/c_{\max}, 1/\bar{c}]\}$ 及其邻域 $S(K; \epsilon)$. 根据隐函数定理, 存在 ϵ 以及定义在 $S(K; \epsilon)$ 上的唯一连续可微函数 $\hat{x}(t, v)$ 和 $\hat{v}(t, v)$, $\forall (t, v) \in S(K; \epsilon)$, 满足

$$(\hat{x}(t, v), \hat{v}(t, v)) - (x^*, v^*) < \epsilon$$

和

$$\begin{aligned} \nabla f(\hat{x}(t, v)) + \nabla g(\hat{x}(t, v)) \hat{v}(t, v) &= 0, \\ (v) g(\hat{x}(t, v)) + t + (v^* - \hat{v}(t, v)) &= 0. \end{aligned} \quad (27)$$

将 t 和 v 的表达式代入 $\hat{x}(t, v)$ 和 $\hat{v}(t, v)$, 定义

$$\begin{aligned} x(v, c) &= \hat{x}(v - v^*/c, 1/c), \\ \tilde{v}(v, c) &= \hat{v}(v - v^*/c, 1/c). \end{aligned}$$

再由式(27), $\forall (v, c) \in D$, 得

$$\begin{aligned} \nabla f(x(v, c)) + \nabla g(x(v, c)) \tilde{v}(v, c) &= 0, \\ \tilde{v}(v, c) &= v_i [1 + c g_i(x(v, c))], \\ i &= 1, 2, \dots, m. \end{aligned} \quad (28)$$

由连续性假设, 可以选择合适的 ϵ 和 δ , 保证 $\nabla g_i(x(v, c))$ ($i = 1, 2, \dots, m$) 线性独立和 $\nabla_{xx}^2 L_c(x(v, c), \tilde{v}(v, c))$ 正定.

下面证明命题2中的式(20)和(21). 将式(26)对 t 和 v 微分, 得

$$\begin{aligned} \begin{bmatrix} \nabla_t \hat{x}(t, v)^T & \nabla_v \hat{x}(t, v)^T \\ \nabla_t \hat{v}(t, v)^T & \nabla_v \hat{v}(t, v)^T \end{bmatrix} = \\ A(t, v) \begin{bmatrix} 0 & 0 \\ -I_m & v^* - \hat{v}(t, v) \end{bmatrix}. \end{aligned} \quad (29)$$

式中

$$A(t, v) =$$

$$\begin{bmatrix} \nabla_{xx}^2 L(\hat{x}(t, v), \hat{v}(t, v)) & \nabla g(\hat{x}(t, v)) \\ (v) \nabla g(\hat{x}(t, v))^T & -I_m \end{bmatrix}^{-1}.$$

使用一阶 Taylor 展开, 推出

$$\begin{aligned} \begin{bmatrix} \hat{x}(t, v) - x^* \\ \hat{v}(t, v) - v^* \end{bmatrix} = \\ \int_0^1 A(t, v) \begin{bmatrix} 0 & 0 \\ -I_m & v^* - \hat{v}(t, v) \end{bmatrix} \begin{bmatrix} t \\ t \end{bmatrix} dt. \end{aligned} \quad (30)$$

由于矩阵(24)非奇异, 可选择充分小的 ϵ , 使 $A(t, v)$ 在集合 $\{(t, v) \mid t < \epsilon, [1/c_{\max}, 1/\bar{c}]\}$ 上一致有界, 即 $\|A(t, v)\| < M$. 选取足够大的 \bar{c} 和 c_{\max} , 保证 $M < 1$. 将上式两边取范数, 得

$$\begin{aligned} \begin{bmatrix} \|\hat{x}(t, v) - x^*\| \\ \|\hat{v}(t, v) - v^*\| \end{bmatrix} \leq \\ M \left(t + \max_0^1 \|\hat{v}(t, v) - v^*\| \right), \end{aligned} \quad (31)$$

从而

$$\begin{aligned} \|\hat{v}(t, v) - v^*\| \leq \\ M \left(t + \max_0^1 \|\hat{v}(t, v) - v^*\| \right). \end{aligned}$$

将不等式左边的变量 (t, v) 用 (t, v) 替换, 则

$$\max_0^1 \|\hat{v}(t, v) - v^*\| \leq \frac{M}{1-M} t. \quad (32)$$

所以, 根据式(31)和(32), 有

$$\begin{aligned} \begin{bmatrix} \|\hat{x}(t, v) - x^*\| \\ \|\hat{v}(t, v) - v^*\| \end{bmatrix} \leq \\ \left(1 + \frac{M}{1-M} t\right) \begin{bmatrix} t \\ t \end{bmatrix}. \end{aligned} \quad (33)$$

令 $B = M/(1-M)$, 由 t 的定义, 得

$$\begin{aligned} x(v, c) - x^* &\leq B \|v - v^*\| / c, \\ \tilde{v}(v, c) - v^* &\leq B \|v - v^*\| / c. \end{aligned} \quad (34)$$

应用矩阵求逆公式, 可得

$$A(t, v) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

其中

$$\begin{aligned} A_{11} &= \nabla_{xx}^2 L_c(\hat{x}(t, v), \hat{v}(t, v))^{-1}, \\ A_{12} &= -A_{11} \nabla g(\hat{x}(t, v)), \\ A_{21} &= - (v) \nabla g(\hat{x}(t, v))^T, \\ A_{22} &= -I_m^{-1} + \\ &\quad - (v) \nabla g(\hat{x}(t, v))^T A_{11} \nabla g(\hat{x}(t, v)). \end{aligned}$$

由式(30), 直接推出

$$\begin{aligned} \begin{bmatrix} \hat{x}(t, v) - x^* \\ \hat{v}(t, v) - v^* \end{bmatrix} = \\ \int_0^1 A_{22} [v^* - \hat{v}(t, v)] dt. \end{aligned}$$

将 $t = (v - v^*)/c, \epsilon = 1/c$ 代入, 可知命题中的最后一个结论成立.

注 3 命题 2 中结论 1) 表明, $\forall (v, c) \in D$, 在最优解 x^* 的邻域, 增广 Lagrange 函数 $L_c(\cdot, v)$ 存在唯一的极小点.

注 4 命题 2 中结论 2) 表明, 只要恰当选择参数 c , 则增广 Lagrange 函数的极小点序列和乘子迭代序列线性地收敛于问题的 KKT 对.

注 5 命题没有考虑不起作用约束的影响. 实际上, 通过惩罚的作用, 增广 Lagrange 函数的极小点序列一般可以进入到可行域, 不起作用约束的乘子将指数地衰减为零.

注 6 算法的收敛性分析与采用的无约束优化算法无关, 但要求增广 Lagrange 函数极小点序列 $\{x^k\}$ 在 x^* 的邻域内, 或至少对某个指标后的所有 k 满足要求, 否则, 收敛性分析不成立. 当然, 无约束优化算法可能会产生其他的局部极小点; 然而, 根据大量的计算经验, 如果使用第 k 步获得的极小点 x^k 作为第 $k + 1$ 步优化的初始点, 则产生的极小点序列 $\{x^k\}$ 在 x^* 的邻域内.

注 7 如果相邻两次极小化的初始点充分接近, 那么通过无约束优化算法获得的极小点也应当充分接近. 所以, 离散时间动态系统 (8) 中的多值映射 A 可以认为是连续的.

下面, 用离散时间动态系统稳定性的观点讨论算法的收敛性. 考虑如下离散时间动态系统:

$$y_{n+1} = f(n, y_n), \quad y_{n_0} = y_0. \quad (35)$$

其解用 $y(n, n_0, y_{n_0})$ 表示, 并假设它对初始值 y_0 连续.

原理 1 (LaSalle 不变集原理)^[9,10] $\forall y \in G \subset R^n$, 设:

1) 两个实值函数 $V(n, y)$ 和 $W(x) \geq 0$ 对 y 连续, $V(n, y)$ 有下界, $\forall n \geq n_0$, 有

$$V(n, y_n) = V(n + 1, y_{n+1}) - W(y_n); \quad (36)$$

2) $\forall n \geq n_0, y(n, n_0, y_{n_0}) \in G$.

于是, 或者 $y(n, n_0, y_{n_0})$ 无界, 或者趋于最大不变集 $E = \{x \in \bar{G} \mid W(x) = 0\}$.

将前文的算法修正为

Step1: 给定初始点 x^0 , 罚参数 $c_0 > 0$, 乘子系数 $b_0 > 0$, 罚参数和乘子系数上界 c_{\max} 和 b_{\max} , 初始 Lagrange 乘子 $v^0 > 0$, 收敛容限 ϵ .

Step2: 极小化 $L^k(\cdot, v^k)$, 求出极小点 x^* .

Step3: 根据下列公式修正乘子:

$$v_i^{k+1} = v_i^k e^{b_k g_i(x^k)}, \quad i = 1, 2, \dots, m.$$

Step4: 取 $c_{k+1} = c_k$ 和 $b_{k+1} = b_k$, 并满足约束 $c_{k+1} \leq c_{\max}$ 和 $b_{k+1} \leq b_{\max}$.

Step5: 如果 $\text{diag}(v^k) g(x^k) \leq \epsilon$, 停止; 否

则, 转到 Step2, 重复计算.

利用 LaSalle 不变集原理讨论算法的收敛条件, 将增广 Lagrange 函数选为 LaSalle 不变集原理中的 V 函数. 为了保证算法的收敛性, 需满足 3 个条件: 1) 式 (36) 成立; 2) 迭代序列有界; 3) 增广 Lagrange 函数有下界.

首先, 讨论第 1) 个条件. 考虑

$$L = L_{c_{k+1}}(x^{k+1}, v^{k+1}) - L_{c_k}(x^k, v^k) = L_{c_{k+1}}(x^{k+1}, v^{k+1}) - L_{c_{k+1}}(x^k, v^{k+1}) + L_{c_{k+1}}(x^k, v^{k+1}) - L_{c_k}(x^k, v^k). \quad (37)$$

由于 x^{k+1} 是 $L_{c_{k+1}}(\cdot, v^{k+1})$ 的极小点, 有

$$L_{c_{k+1}}(x^{k+1}, v^{k+1}) - L_{c_{k+1}}(x^k, v^{k+1}) \leq 0. \quad (38)$$

根据增广 Lagrange 函数的定义和乘子迭代公式, 式 (37) 的最后 2 项展开为

$$L_{c_{k+1}}(x^k, v^{k+1}) - L_{c_k}(x^k, v^k) = \sum_{i=1}^m v_i^k g_i(x^k) \{e^{b_k g_i(x^k)} - 1 + [c_{k+1} e^{b_k g_i(x^k)} - c_k] g_i(x^k)\}. \quad (39)$$

显然, 如果 $g_i(x^k) > 0$, 则式 (39) 中相应项为正, 是系统的不稳定因素, 导致惩罚加大. 于是, 算法通过极小化增广 Lagrange 函数, 使极小点序列 $\{x^k\}$ 向 $g_i(x^k) = 0$ 的区域移动. 如果 $g_i(x^k) = 0$, 则相应的项不影响稳定性; 如果 $g_i(x^k) < 0$, 且为较大的负数, 应选择合适的参数 c_k, c_{k+1} 和 b_k 使相应的项为负, 成为系统的稳定因素; 如果 $g_i(x^k)$ 为非常小的负数, 则该项可能变为正, 成为算法的不稳定因素. 此时, 应选择参数使该项的影响忽略不计.

至于第 2) 个条件可以在一个有界集内搜索极小点来满足. 但是, 由此可能导致式 (38) 不满足, 即 x^{k+1} 不是 $L_{c_{k+1}}(\cdot, v^{k+1})$ 的极小点. 解决办法是选择一个充分大的界, 使解轨线不会达到边界; 或通过选择合适的参数 b 和 c , 增大式 (39) 的减小量, 保证增广 Lagrange 函数的总增量小于零.

下面讨论第 3) 个条件. 由 $f(x)$ 和 $g(x)$ 的连续性可知, $f(x^k)$ 和 $g(x^k)$ 有界, 若 c^k 有上界, 则可推出 $L_{c_k}(x^k, v^k)$ 有下界. 因此, 如果第 1) 个条件满足, 且迭代序列有界, 则一定收敛于增广 Lagrange 函数的最大不变集, 也就是问题 (1) 的 KKT 对.

从上面分析可以看出, 算法的收敛性是多种因素综合作用的结果. 如果 x^k 位于可行域外, 则不稳定因素持续增长, 并将最终占优, 迫使序列 $\{x^k\}$ 趋向可行域; 如果 x^k 位于可行域内, 则可通过选择罚参数和乘子系数, 使稳定因素占优, 驱使序列 $\{x^k\}$ 收敛于问题 (1) 的 KKT 对. 算法收敛的关键是极小化增广 Lagrange 函数产生的减小量与式 (39) 产生

的增量和为负,或至少在某一指标后该条件成立. 罚参数的作用是使惩罚作用加强,但罚参数过大,可能导致系统震荡. 因此,罚参数并不是越大越好. 为了增加惩罚作用,可采用增长速度更快的罚函数代替平方罚函数.

值得说明的是,在应用 LaSalle 不变集原理分析算法的收敛性时,只用到假设 1. 所以,根据 LaSalle 不变集原理获得的结论也可用于退化情形以及正则性和 / 或严格互补性条件不满足的情况.

5 数值计算

选择 Matlab 中的两个非线性规划的例子说明算法的有效性.

例 1

$$\begin{aligned} \min \quad & e^{x_1} (4x_1^2 + 2x_2^2 + 4x_1x_2 + 2x_2 + 1), \\ \text{s. t.} \quad & x_1x_2 - x_1 - x_2 + 1.5 = 0, \\ & -x_1x_2 - 10 = 0. \end{aligned}$$

初值选择为

$$\begin{aligned} x^0 &= (-6.3523, 1.3632), \\ v^0 &= (2.4542, 1.6186), \end{aligned}$$

初始参数选择为

$$b_0 = 1, c_0 = 1,$$

每次循环后

$$b_{k+1} = 10b_k, c_{k+1} = 100c_k,$$

经过 5 次循环,最优解

$$x^* = (-9.5474, 1.0474).$$

例 2

$$\begin{aligned} \min \quad & e^{x_1} (4x_1^2 + 2x_2^2 + 4x_1x_2 + 2x_2 + 1), \\ \text{s. t.} \quad & x_1^2 + x_2 - 1 = 0, \\ & -x_1x_2 - 10 = 0. \end{aligned}$$

将等式约束变成两个等价的 inequality 约束. 初值选择为

$$\begin{aligned} x^0 &= (-3.2540, 13.8489), \\ v^0 &= (2.0503, 1.4053, 2.3443), \end{aligned}$$

初始参数选择为

$$b_0 = 1, c_0 = 1,$$

每次循环后

$$b_{k+1} = 10b_k, c_{k+1} = 100c_k,$$

经过 7 次循环,获得最优解

$$x^* = (-0.7527, 0.4335).$$

6 结 论

通过重新定义 Lagrange 乘子,获得一种改进的 Lagrange 乘子法. 其优点在于,不需要用松弛变量将不等式约束转换成等式约束,可直接对不等式约束进行处理和将关于等式约束的结论完全移植过来,并且证明非常相似. 文中根据隐函数定理分析了改进算法的局部稳定性,并利用 LaSalle 不变集原理讨论算法的大范围收敛性质,给出罚参数的选择原则. 从分析中可以看出,算法在稳定因素和不稳定因素的综合作用下获得最优解,二者缺一不可. 但为了使算法收敛,必须保证稳定因素最终占优. 最后,利用两个算例说明了算法的有效性.

参考文献(References)

- [1] Bertsekas D P. Constrained optimization and Lagrange methods[M]. New York: Academic Press, 1982.
- [2] Avriel M. Nonlinear programming analysis and methods [M]. New Jersey: Prentice-Hall, 1976.
- [3] Bertsekas D P. Nonlinear programming [M]. Massachusetts: Athena Scientific, 1999.
- [4] Fletcher R. Practical methods of optimization[M]. New York: John Wiley and Sons, 1987.
- [5] Luenberger D G. Linear and nonlinear programming [M]. Massachusetts: Addison Wesley, 1984.
- [6] Fiacco A V, McCormick G P. Nonlinear programming-sequential unconstrained minimization techniques [M]. New York: SIAM, 1990.
- [7] Rockafellar R T. Lagrange multiplier and optimality[J]. SIAM Review, 1993, 35(2): 183-238.
- [8] Ortega J M, Rheinboldt W C. Iterative solution of nonlinear equations in several variables [M]. New York: Academic Press, 1970.
- [9] Lakshmikantham V, Trigiante D. Theory of difference equations-numerical methods and applications[M]. New York: Marcel Dekker, 2002.
- [10] LaSalle J P. The stability of dynamical systems[M]. New York: Springer, 1976.
- [11] Nocedal J, Wright S J. Numerical optimization [M]. New York: Springer-Verlag, 1999.