

极大值复合函数 Clarke 广义梯度 计算的一个新方法*

高 岩

(中国矿业大学北京校区数学教研室, 北京 100083)

摘 要 对于一类非光滑函数—光滑函数的有限次极大值复合函数, 给出了计算它在一点处 Clarke 广义梯度中一个元素的新方法. 与以往方法比较, 本文的方法不需判别线性不等式组的相容性, 因而易于实现.

关键词 非光滑优化, Clarke 广义梯度, 极大值复合函数

1 引言和预备知识

考虑下述一般形式的极大值复合函数

$$f(x) = g(x, \max_{j \in J_1} f_{1j}(x), \dots, \max_{j \in J_m} f_{mj}(x)), \quad (1)$$

其中 $x \in R^n$, $g, f_{ij}, j \in J_i, i = 1, \dots, m$ 是 R^{m+n} 和 R^n 上连续可微函数, $J_i, i = 1, \dots, m$ 为有限指标集. [1] 给出了计算 $f(x)$ 在一点的 Clarke 广义梯度 [2] 中一个元素的一种方法, 该方法做为极小化 $f(x)$ 算法 (例如 Bundle 方法 [3,4]) 的子算法, 可得到极小化 $f(x)$ 的可执行算法. 然而, 从实用角度看, [1] 所给的方法计算量太大, 不易实现. 本文将对 [1] 中方法进行改进, 得到一个及为简便的计算 $f(x)$ 在一点的一个 Clarke 广义梯度的方法.

首先回顾 [1] 中有关内容. 记

$$J_i(x) = \{j \in J_i \mid f_{ij}(x) = \max_{k \in J_i} f_{ik}(x)\}, \quad i = 1, \dots, m. \quad (2)$$

按下述原则确定指标集 $\bar{J}_i(x), i = 1, \dots, m$:

$$\bar{J}_i(x) \subset J_i(x), \quad i = 1, \dots, m; \quad (3a)$$

$$\nabla f_{ij}(x) \neq \nabla f_{ik}(x), \quad \forall j, k \in \bar{J}_i(x), j \neq k, i = 1, \dots, m; \quad (3b)$$

$$\forall t_i \in J_i(x), \exists k_i \in \bar{J}_i(x) \text{ 使得 } \nabla f_{it_i}(x) = \nabla f_{ik_i}(x), \quad i = 1, \dots, m. \quad (3c)$$

事实上, 指标集 $\bar{J}_i(x)$ 是指标集 $J_i(x)$ 的子集, 它可由在指标 $j, k \in J_i(x)$ 满足 $\nabla f_{ij}(x) = \nabla f_{ik}(x), i \in \{1, \dots, m\}$ 中仅保留一个指标而得到. 对每一组指标 $j_1 \in \bar{J}_1(x), \dots, j_m \in$

本文 1998 年 2 月 9 日收到. 1999 年 8 月 10 日收到修改稿.

* 辽宁省博士启动基金资助项目.

$\bar{J}_m(x)$, 引入下述线性不等式系统

$$\begin{aligned} L_{j_1 \dots j_m}(x) \quad & (\nabla f_{1k_1}(x) - \nabla f_{1j_1}(x))^T y < 0, \quad \forall k_1 \in \bar{J}_1(x) \setminus \{j_1\} \\ & (\nabla f_{2k_2}(x) - \nabla f_{2j_2}(x))^T y < 0, \quad \forall k_2 \in \bar{J}_2(x) \setminus \{j_2\} \\ & \vdots \\ & (\nabla f_{mk_m}(x) - \nabla f_{mj_m}(x))^T y < 0, \quad \forall k_m \in \bar{J}_m(x) \setminus \{j_m\}, \end{aligned}$$

其中 $y \in R^n$. 易见 $L_{j_1 \dots j_m}(x)$ 是具有 $\prod_{i=1}^m (\text{card } \bar{J}_i(x) - 1)$ 个不等式 n 个变量的严格线性不等式组.

定理1^[1] 如果指标组 $j_1 \in \bar{J}_1(x), \dots, j_m \in \bar{J}_m(x)$ 使得系统 $L_{j_1 \dots j_m}(x)$ 相容, 则 $\xi = \nabla g(x, f_{1j_1}(x), \dots, f_{mj_m}(x)) \in \partial f(x)$, 其中 $\partial f(x)$ 为 $f(x)$ 的 Clarke 广义梯度.

对每一组指标 $j_1 \in \bar{J}_1(x), \dots, j_m \in \bar{J}_m(x)$ 判别相应系统 $L_{j_1 \dots j_m}(x)$ 的相容性, 可至少得到函数 $f(x)$ 在 x 点的一个 Clarke 广义梯度. 按此方法计算 $f(x)$ 在 x 点的一个 Clarke 广义梯度, 需要 $\prod_{i=1}^m \text{card } \bar{J}_i(x)$ 次判别线性不等式组的相容性, 或等价地求解线性规划问题, 因而计算量很大.

2 算法及有关定理

给定点 $x \in R^n$ 和一组 $y_i \in R^n, i = 1, \dots, m$, 定义下述指标集

$$K_1(x, y_1) = \{j \in \bar{J}_1(x) \mid \nabla f_{1j}(x)^T y_1 > \nabla f_{1k}(x)^T y_1, \quad \forall k \in \bar{J}_1(x) \setminus \{j\}\}, \quad (4a)$$

$$K_i(x, y_1) = \{j \in \bar{J}_i(x) \mid \nabla f_{ij}(x)^T y_1 = \max_{k \in \bar{J}_i(x)} \nabla f_{ik}(x)^T y_1\}, \quad 2 \leq i \leq m, \quad (4b)$$

$$\begin{aligned} K_i(x, y_1, y_2) &= \{j \in K_i(x, y_1) \mid \nabla f_{ij}(x)^T y_2 \\ &= \max_{k \in K_i(x, y_1)} \nabla f_{ik}(x)^T y_2\}, \quad 3 \leq i \leq m, \end{aligned} \quad (4c)$$

⋮

$$\begin{aligned} K_i(x, y_1, \dots, y_{i-1}) &= \{j \in K_i(x, y_1, \dots, y_{i-2}) \mid \nabla f_{ij}(x)^T y_{i-1} \\ &= \max_{k \in K_i(x, y_1, \dots, y_{i-2})} \nabla f_{ik}(x)^T y_{i-1}\}, \quad 3 \leq i \leq m, \end{aligned} \quad (4d)$$

$$K_i(x, y_1, \dots, y_i) = \{j \in K_i(x, y_1, \dots, y_{i-1}) \mid \nabla f_{ij}(x)^T y_i > \nabla f_{ik}(x)^T y_i, \quad \forall k \in K_i(x, y_1, \dots, y_{i-1}) \setminus \{j\}\}, \quad 2 \leq i \leq m. \quad (4e)$$

易见, 对任意点 x 和一组 $y_i \in R^n, i = 1, \dots, m$, 指标集 $K_i(x, y_1, \dots, y_j), 1 \leq j < i$ 是非空的, 而指标集 $K_i(x, y_1, \dots, y_i), i = 1, \dots, m$ 可能是空集, 但是有下述性质

$$K_i(x, y_1, \dots, y_j) \subset K_i(x, y_1, \dots, y_{j-1}), \quad 2 \leq j \leq i, \quad (5)$$

$$(\nabla f_{ik_t}(x) - \nabla f_{ij_t}(x))^T y_t \leq 0, \quad \forall k_t \in \bar{J}_i(x), j_t \in K_i(x, y_1, \dots, y_i), t \leq i. \quad (6)$$

引理1^[5] 设 A 是 $m \times n$ 实矩阵, 则线性不等式组 $Ay < 0, y \in R^n$ 的解集为 R^n 中开凸锥.

引理 2^[5] 设 $a_i \in R^n$, $i \in I$, 其中 I 为有限指标集, 且 $a_i \neq a_j$, $\forall i, j \in I, i \neq j$. 若指标 $i_0 \in I$ 满足 $\|a_{i_0}\| = \max_{i \in I} \|a_i\|$, 则有

$$a_i^T a_{i_0} < a_{i_0}^T a_{i_0}, \quad \forall i \in I \setminus \{i_0\}.$$

下面给出本文的主要定理.

定理 2 若指标集 $K_i(x, y_1, \dots, y_i)$, $i = 1, \dots, m$ 非空, 设 $j_1 \in K_1(x, y_1), \dots, j_m \in K_m(x, y_1, \dots, y_m)$, 则系统 $L_{j_1 \dots j_m}(x)$ 相容, 进一步 $\nabla g(x, f_{j_1}(x), \dots, f_{j_m}(x)) \in \partial f(x)$.

证 只需证明存在常数 $\varepsilon_i > 0$, $i = 2, \dots, m$ 使得 $y = y_1 + \varepsilon_2 y_2 + \dots + \varepsilon_m y_m$ 为系统 $L_{j_1 \dots j_m}(x)$ 的解. 对 m 利用归纳法. 当 $m = 1$ 时, 由 $K_1(x, y_1)$ 的定义知

$$(\nabla f_{1k_1}(x) - \nabla f_{1j_1}(x))^T y_1 < 0, \quad \forall k_1 \in \bar{J}_1(x) \setminus \{j_1\}, \quad (7)$$

即 $y = y_1$ 是系统 $L_{j_1}(x)$ 的解. 假设存在 $\varepsilon_i > 0$, $i = 1, \dots, m-1$ 使得 $y = y_1 + \varepsilon_2 y_2 + \dots + \varepsilon_{m-1} y_{m-1}$ 为系统 $L_{j_1 \dots j_{m-1}}(x)$ 的解, 即

$$\begin{aligned} (\nabla f_{ik_i}(x) - \nabla f_{ij_i}(x))^T (y_1 + \varepsilon_2 y_2 + \dots + \varepsilon_{m-1} y_{m-1}) < 0, \\ \forall k_i \in \bar{J}_i(x) \setminus \{j_i\}, \quad i = 1, \dots, m-1. \end{aligned} \quad (8)$$

由引理 1, $L_{j_1 \dots j_{m-1}}(x)$ 的解集为 R^n 中开凸锥, 于是存在 $\varepsilon^{(1)} > 0$, 当 $0 < \varepsilon_m < \varepsilon^{(1)}$ 时, 有

$$\begin{aligned} (\nabla f_{ik_i}(x) - \nabla f_{ij_i}(x))^T (y_1 + \varepsilon_2 y_2 + \dots + \varepsilon_m y_m) < 0, \\ \forall k_i \in \bar{J}_i(x) \setminus \{j_i\}, \quad i = 1, \dots, m-1. \end{aligned} \quad (9)$$

再根据 $K_m(x, y_1, \dots, y_m)$ 的定义, 得

$$(\nabla f_{mk_m}(x) - \nabla f_{mj_m}(x))^T y_m < 0, \quad \forall k_m \in K_m(x, y_1, \dots, y_{m-1}) \setminus \{j_m\}. \quad (10)$$

结合 (6) 式和 (10) 式得

$$\begin{aligned} (\nabla f_{mk_m}(x) - \nabla f_{mj_m}(x))^T (y_1 + \varepsilon_2 y_2 + \dots + \varepsilon_m y_m) < 0, \\ \forall k_m \in K_m(x, y_1, \dots, y_{m-1}) \setminus \{j_m\}. \end{aligned} \quad (11)$$

注意到下述关系式

$$K_m(x, y_1, \dots, y_m) \subset \dots \subset K_m(x, y_1, y_2) \subset K_m(x, y_1) \subset \bar{J}_m(x). \quad (12)$$

于是, 对每个指标 $k_m \in (\bar{J}_m(x) \setminus K_m(x, y_1, \dots, y_{m-1})) \setminus \{j_m\}$ 可分为这样两种情形: 一种情形是存在指标 t 满足 $2 \leq t \leq m-1$ (此处 t 依赖于 k_m) 使得 $k_m \in K_m(x, y_1, \dots, y_{t-1}) \setminus \{j_m\}$, $k_m \notin K_m(x, y_1, \dots, y_t) \setminus \{j_m\}$, 这时有

$$(\nabla f_{mk_m}(x) - \nabla f_{mj_m}(x))^T y_t < 0.$$

另一种情形是 $k_m \notin K_m(x, y_1) \setminus \{j_m\}$, 这时有

$$(\nabla f_{mk_m}(x) - \nabla f_{mj_m}(x))^T y_1 < 0.$$

综上所述, 对任意 $k_m \in (\bar{J}_m(x) \setminus K_m(x, y_1, \dots, y_{m-1})) \setminus \{j_m\}$, 存在 $t \leq m-1$ 使得

$$(\nabla f_{mk_m}(x) - \nabla f_{mj_m}(x))^T y_t < 0. \quad (13)$$

于是, 存在 $\varepsilon^{(2)} > 0$, 当 $0 < \varepsilon_m < \varepsilon^{(2)}$ 时, 有

$$\begin{aligned} & (\nabla f_{mk_m}(x) - \nabla f_{mj_m}(x))^T (y_1 + \varepsilon_2 y_2 + \dots + \varepsilon_m y_m) < 0, \\ & \forall k_m \in (\bar{J}_m(x) \setminus K_m(x, y_1, \dots, y_{m-1})) \setminus \{j_m\}. \end{aligned} \quad (14)$$

选取 ε_m 满足 $0 < \varepsilon_m < \min\{\varepsilon^{(1)}, \varepsilon^{(2)}\}$, 结合式 (11) 和式 (14), 得

$$(\nabla f_{mk_m}(x) - \nabla f_{mj_m}(x))^T (y_1 + \varepsilon_2 y_2 + \dots + \varepsilon_m y_m) < 0, \quad \forall k_m \in \bar{J}_m(x) \setminus \{j_m\}. \quad (15)$$

进一步, 结合式 (9) 和式 (15), 得 $y = y_1 + \varepsilon_2 y_2 + \dots + \varepsilon_m y_m$ 为 $L_{j_1 \dots j_m}(x)$ 的解. 根据归纳法, $L_{j_1 \dots j_m}(x)$ 是相容的. 再由定理 1, $\xi = \nabla g(x, f_{1j_1}(x), \dots, f_{mj_m}(x)) \in \partial f(x)$. 证毕.

定理 2 说明计算 $f(x)$ 在 x 点的一个 Clarke 广义梯度可转化为寻求一组向量 $y_i \in R^n$, $i = 1, \dots, m$ 使得指标集 $K_i(x, y_1, \dots, y_i)$, $i = 1, \dots, m$ 非空. 这一工作可由下述步骤实现.

(1) 选取指标 $j_1 \in \bar{J}_1(x)$ 满足 $\|\nabla f_{1j_1}(x)\| = \max\{\|\nabla f_{1j}(x)\| \mid j \in \bar{J}_1(x)\}$, 令 $y_1 = \nabla f_{1j_1}(x)$.

(2) 选取指标 $j_i \in K_i(x, y_1, \dots, y_{i-1})$, $i = 2, \dots, m$ 满足

$$\|\nabla f_{ij_i}(x)\| = \max\{\|\nabla f_{ij}(x)\| \mid j \in K_i(x, y_1, \dots, y_{i-1})\}, \quad i = 2, \dots, m,$$

令 $y_i = \nabla f_{ij_i}(x)$.

由引理 2 可知, 按上述原则确定的 y_i 和 j_i , $i = 1, \dots, m$ 满足 $j_i \in K_i(x, y_1, \dots, y_i)$, $i = 1, \dots, m$. 于是由定理 1, $\nabla g(x, f_{1j_1}(x), \dots, f_{mj_m}(x)) \in \partial f(x)$. 下边给出算法的具体步骤.

算法

步 0 给定点 $x \in R^n$.

步 1 计算函数值 $f_{ij}(x)$, $j \in J_i$, $i = 1, \dots, m$, 确定指标集 $J_i(x)$, $i = 1, \dots, m$.

步 2 计算梯度 $\nabla f_{ij}(x)$, $j \in J_i(x)$, $i = 1, \dots, m$, 确定指标集 $\bar{J}_i(x)$, $i = 1, \dots, m$.

步 3 令 $K_i = \bar{J}_i(x)$, $i = 1, \dots, m$ 和 $n = 1$.

步 4 确定指标集 $K'_n = \{j \in K_n \mid \|\nabla f_{nj}(x)\| = \max_{k \in K_n} \|\nabla f_{nk}(x)\|\}$, 选取任意一个指标 $j_n \in K'_n$.

步 5 如果 $n = m$, 计算 Clarke 广义梯度 $\xi = \nabla g(x, f_{1j_1}(x), \dots, f_{mj_m}(x))$, 停止; 否则, 转步 6.

步 6 确定指标集

$$K'_i = \{j \in K_i \mid \|\nabla f_{ij}(x)\|^T \nabla f_{nj_n}(x) = \max_{k \in K_i} \|\nabla f_{ik}(x)\|^T \nabla f_{nj_n}(x)\}, \quad i = n+1, \dots, m.$$

步 7 令 $K_i = K'_i$, $i = n+1, \dots, m$ 和 $n = n+1$, 转步 4.

注 不考虑函数值 $f_{ij}(x)$, $j \in J_i$, $i = 1, \dots, m$ 和梯度 $\nabla f_{ij}(x)$, $j \in J_i$, $i = 1, \dots, m$, $\nabla g(x, f_{1j_1}(x), \dots, f_{mj_m}(x))$ 的计算量, 执行上述算法最多进行 $n \sum_{i=1}^m i \text{card} \bar{J}_i(x)$ 次乘法运算.

例 设

$$\begin{aligned} f(x) &= g\left(\max_{j=1,2} f_{1j}(x), \max_{j=1,2,3} f_{2j}(x), \max_{j=1,2,3} f_{3j}(x)\right) \\ &= \max_{j=1,2} f_{1j}(x) - \max_{j=1,2,3} f_{2j}(x) \max_{j=1,2,3} f_{3j}(x), \end{aligned}$$

其中

$$\begin{aligned} x &= (x_1, x_2, x_3)^T, \\ f_{11}(x) &= x_1 + x_2, & f_{12}(x) &= x_1 - x_2, \\ f_{21}(x) &= \sin x_1, & f_{22}(x) &= \sin x_2, & f_{23}(x) &= \sin x_3, \\ f_{31}(x) &= e^{x_1} - 1, & f_{32}(x) &= x_1^2 + \sin x_2, & f_{33}(x) &= x_2 + \cos x_3 - 1. \end{aligned}$$

通过计算 $f_{ij}(x)$ 在 $x = 0$ 处函数值可确定指标集 $J_1(0) = \{1, 2\}$, $J_2(0) = \{1, 2, 3\}$, $J_3(0) = \{1, 2, 3\}$. 计算梯度

$$\begin{aligned} \nabla f_{11}(0) &= (1, 1, 0)^T, & \nabla f_{12}(0) &= (1, -1, 0)^T, \\ \nabla f_{21}(0) &= (1, 0, 0)^T, & \nabla f_{22}(0) &= (0, 1, 0)^T, & \nabla f_{23}(0) &= (0, 0, 1)^T, \\ \nabla f_{31}(0) &= (1, 0, 0)^T, & \nabla f_{32}(0) &= (0, 1, 0)^T, & \nabla f_{33}(0) &= (0, 1, 0)^T. \end{aligned}$$

确定指标集 $\bar{J}_1(0) = J_1(0) = \{1, 2\}$, $\bar{J}_2(0) = J_2(0) = \{1, 2, 3\}$, $\bar{J}_3(0) = \{1, 2\}$. 当然, 指标集 $\bar{J}_3(0)$ 也可选为 $\{1, 3\}$. 使用 [1] 中方法计算 $f(x)$ 在 $x = 0$ 点的一个 Clarke 广义梯度, 需要判别 12 个线性不等式系统:

$$\begin{aligned} L_{111}(0), L_{112}(0), L_{121}(0), L_{122}(0), L_{131}(0), L_{132}(0), \\ L_{211}(0), L_{212}(0), L_{221}(0), L_{222}(0), L_{231}(0), L_{232}(0) \end{aligned}$$

的相容性, 其中每个线性不等式系统含有 4 个不等式和 3 个变量, 例如, $L_{111}(0)$ 具有下述形式:

$$\begin{aligned} L_{111}(0) \quad & (\nabla f_{12}(0) - \nabla f_{11}(0))^T y < 0 \\ & (\nabla f_{22}(0) - \nabla f_{21}(0))^T y < 0 \\ & (\nabla f_{23}(0) - \nabla f_{21}(0))^T y < 0 \\ & (\nabla f_{32}(0) - \nabla f_{31}(0))^T y < 0 \end{aligned}$$

其中, $y \in R^3$. 下面用本文的方法计算 $f(x)$ 在 $x = 0$ 点的一个 Clarke 广义梯度. 由于 $\|\nabla f_{11}(0)\| = \|\nabla f_{12}(0)\| = \max\{\|\nabla f_{1j}(0)\| \mid j \in \bar{J}_1(0)\}$, 选取 $j_1 = 1$, 这时 $y_1 = \nabla f_{11}(0) = (1, 1, 0)^T$. 确定指标集

$$\begin{aligned} K_2(0, y_1) &= \{j \in \bar{J}_2(0) \mid \nabla f_{2j}(0)^T y_1 = \max_{k \in \bar{J}_2(0)} \nabla f_{2k}(0)^T y_1\} = \{1, 2\}, \\ K_3(0, y_1) &= \{j \in \bar{J}_3(0) \mid \nabla f_{3j}(0)^T y_1 = \max_{k \in \bar{J}_3(0)} \nabla f_{3k}(0)^T y_1\} = \{1, 2\}. \end{aligned}$$

由于 $\|\nabla f_{21}(0)\| = \|\nabla f_{22}(0)\| = \max\{\|\nabla f_{2j}(0)\| \mid j \in K_2(0, y_1)\}$, 选取 $j_2 = 2$, 这时

$y_2 = \nabla f_{22}(0) = (0, 1, 0)^T$. 确定指标集

$$K_3(0, y_1, y_2) = \{j \in K_3(0, y_1) \mid \nabla f_{3j}(0)^T y_2 = \max_{k \in K_3(0, y_1)} \nabla f_{3k}(0)^T y_2\} = \{2\}.$$

$K_3(0, y_1, y_2)$ 是单点集, 选取 $j_3 = 2$. 至此一组指标 $j_1 = 1, j_2 = 2, j_3 = 2$ 已确定. 于是

$$\begin{aligned} \xi &= \nabla g(f_{11}(x), f_{22}(x), f_{32}(x)) \Big|_{x=0} = \nabla (f_{11}(x) - f_{22}(x)f_{32}(x)) \Big|_{x=0} \\ &= \nabla (x_1 + x_2 - \sin x_2(x_1^2 + \sin x_2)) \Big|_{x=0} = (1, 1, 0)^T \end{aligned}$$

为 $f(x)$ 在 $x = 0$ 点的一个 Clarke 广义梯度.

参 考 文 献

- 1 Gao Yan, Xia Zunquan. Finding a Clarke Subgradient for Smooth Composition of Max-type Functions. *Archives of Control Sciences*, 1994, 3(XXXIX): 181-191
- 2 Clarke F H. Optimization and Nonsmooth Analysis. New York: John Wiley & Sons, 1983
- 3 Lemarechal C, Zow J. A Condensed Introduction to Bundle Methods in Nonsmooth Optimization. Algorithms for Continuous Optimization: The State of the Art, Dordrecht: Kluwer Academic Publishers, 1994, 357-382
- 4 Makela M M, Neittaanmaki P. Nonsmooth Optimization. Singapore: World Scientific, 1994
- 5 Hiriart-Urruty J B, Lemarechal C. Convex Analysis and Minimization Algorithms. Berlin: Springer-Verlag, 1993

A NEW APPROACH TO COMPUTING A CLARKE GENERALIZED GRADIENT FOR A SMOOTH COMPOSITION OF MAX-TYPE FUNCTIONS

GAO YAN

(Division of Mathematics, China University of Mining and Technology, Beijing 100083)

Abstract This paper refers to computing Clarke generalized gradient for a smooth composition of max-type functions. A new approach to computing an element of Clarke generalized gradient for this class of functions at a point, is proposed. In contrast to previous publications, the present approach does not employ the procedure of determining consistencies of system of strictly linear inequalities. Hence, it is easy to be implemented.

Key words Nonsmooth optimization, Clarke generalized gradient, max-type functions