ОБОБЩЕНИЯ МЕТОДА DIRECT НА ЗАДАЧИ С ФУНКЦИОНАЛЬНЫМИ ОГРАНИЧЕНИЯМИ

С.Ю. Городецкий

Нижегородский госуниверситет им. Н.И. Лобачевского

Метод Direct был разработан в 1993 году для решения задач липшицевой глобальной оптимизации без вычисления оценок константы Липшица. Значения константы могли быть любыми от нуля до бесконечности. Метод не учитывал функциональные ограничения. В данной работе предлагаются два его специальных обобщения на случай наличия таких ограничений. Первое основано на сведении к задаче безусловной глобальной оптимизации с перестраиваемой целевой функцией. Второе – непосредственно распространяет принципы построения Direct на случай липшицевых ограничений при неограниченном диапазоне их констант Липшица. Приведены теоретические основы построения новых методов, обоснована сходимость.

Введение

Метод Direct предложили Jones D.R., Perttunen C.D., Stuckman B.E. в работе [1]. Метод предназначен для поиска на гиперинтервале $D=[a,b]\subset R^N$ глобального минимума липшицевой функции без использования значений константы Липшица L_f:

$$f^* = f(x^*) = \min f(x),\quad x\in D=[a,b]\subset R^N. \qquad (1)$$

Относительно L_f предполагается, что ее значение может быть любым числом из диапазона $[0,\infty)$. Метод Direct использует разбиение области поиска D на гиперинтервалы $D_i=[a_i,b_i]$ по центральной схеме деления на три. Эта схема деления первоначально была предложена Евтушенко Ю.Г. и Ратькиным В.А. для задач с заданным значением константы L_f. В [2] опубликован предварительный вариант этого метода (в форме схемы деления на два). В их методе на каждом шаге выбирался, для последующего деления на три равные части (по большему ребру), наиболее *приоритетный* гиперинтервал D_t. Его выбор определялся тем, что нижняя оценка функции $f(x)$ в пределах D_t по измерению в его геометрическом центре $c_t=(a_t+b_t)/2$, имеющая вид

$$f_t^-(L_f) = f^-(L_f, D_t) = f(c_t) - L_f\cdot\| b_t - a_t\|/2, \qquad (2)$$

оказывалась наименьшей по сравнению с другими гиперинтервалами при заданном значении L_f. При этом из сравнения исключались гиперинтервалы, в которых значения $f_i^-(L_f)$ улучшали достигнутое к данной итерации наименьшее значение $f_{\min}$ функции менее, чем на $\varepsilon>0$. То есть выбор *приоритетного* D_t определялся правилами:

$$\forall D_i:\ f_t^-(L_f)\le f_i^-(L_f), \qquad (3)$$

$$f_t^-(L_f)\le f_{\min}-\varepsilon. \qquad (4)$$

В методе Direct [1], описание которого можно также найти в [3], значение L_f неизвестно, что приводит к следующему изменению принципа (3)-(4) выбора приоритетных гиперинтервалов (их при неизвестном L_f называют *потенциально оптимальными*). D_t

является *потенциально оптимальным*, если существует хотя бы одно значение $L_f \in [0,\infty)$, при котором условие (3) оказывается справедливым, и, кроме того, выполняется измененный аналог (6) условия (4).

Правило выбора *потенциально оптимальных* гиперинтервалов приобретает вид:

$$\exists L_f \in [0,\infty): \ \forall D_i: \ f_t^-(L_f) \le f_i^-(L_f), \tag{5}$$

$$f_t^-(L_f) \le f_{\min} - \varepsilon | f_{\min} |. \tag{6}$$

В работе [1] найдено простое геометрическое представление, позволяющее выделять множество гиперинтервалов D_t, удовлетворяющих условию (5). А именно, если изобразить гиперинтервалы D_i на плоскости точками с координатами (d_i, f_i), где $d_i = \| b_i - a_i \|$, $f_i = f(c_i)$, то множеству точек (d_t, f_t) для гиперинтервалов D_t соответствуют вершины правой нижней части границы выпуклой линейной оболочки множества всех точек (d_i, f_i). Для вычислительно эффективного выделения точек (d_t, f_t) в [1] предложено использовать правило Грэхема [4].

В методе Direct на каждой итерации выполняется разбиение по принципу деления на три (по большим ребрам) всех найденных потенциально оптимальных гиперинтервалов D_t с последующим проведением в новых центрах образовавшихся гиперинтервалов измерений функции. Вычисления значений функций в новых центрах могут выполняться параллельно. Останов вычислений выполняется при превышении предельного количества итераций или предельного количества измерений функции. Метод при тестировании демонстрирует хорошие характеристики.

1. Известные обобщения метода Direct

К настоящему времени разработаны следующие обобщения Direct. В монографии [3] построен более эффективный вариант метода, работающий по диагональной схеме деления на три. В [5] получено нетривиальное обобщение метода Direct на задачи с вычислимой липшицевой производной, константа Липшица которой неизвестна и может изменяться на промежутке $[0,\infty)$; метод построен для задач (1) с размерностью $N = 1$. В [6] метод [5] распространен на многомерные задачи с использованием новой схемы разбиения области.

Обобщений метода Direct на задачи с функциональными ограничениями к настоящему времени построено не было. В некоторых работах предлагалось для метода Direct учитывать ограничения с использованием штрафных функций, однако этот подход нельзя считать корректным, поскольку предлагаемые штрафы не являются точными.

2. Два подхода к обобщению метода Direct на задачи с функциональными ограничениями

Задача с функциональными ограничениями ставится следующим образом:

$$f^* = f(x^*) = \min f(x), \ x \in X, \tag{7}$$

$$X = \{x \in D = [a,b] \subset R^N : g_i(x) \le 0, (i = 1,..,m)\}. \tag{8}$$

Ограничения неравенства сводятся к одному обобщенному ограничению вида

$$g(x) \le 0, \tag{9}$$

$$g(x) = \max\{ g_1(x),..., g_m(x) \}. \tag{10}$$

В двух предлагаемых подходах принимаются разные предположения о свойствах функций задачи.

Первый подход использует сведение исходной задачи (7)-(8) к задаче без функциональных ограничений с перестраиваемой целевой функцией. При этом исходные предположения заключаются в том, что функции f и g являются липшицевыми с равными константами $L_f = L_g = L$, где значение L неизвестно и $L \in [0,\infty)$.

Пусть f_k^* – текущая оценка глобально-минимального значения функции в (7)-(8):

$$f_k^* = \begin{cases} \min \{ f(c_i) : c_i \in D_i \ \& \ g(c_i) \le 0, i = 1,.., M_k \}, \ \exists D_i : g(c_i) \le 0 \\ +\infty, \ \ \forall D_i : \ g(c_i) > 0, \end{cases} \quad (11)$$

где M_k – число гиперинтервалов D_i на k-й итерации, а c_i – их центральные точки.

Введем перестраиваемую целевую функцию $f_k(x)$, зависящую от оценки f_k^*:

$$f_k(x) = \max\{ f(x) - f_k^*; g(x) \}. \quad (12)$$

Ее константа Липшица равна L, значение L неизвестно, и $L \in [0,\infty)$. Вместо задачи (7)-(8) с функциями-ограничениями будем решать задачу без таких ограничений:

$$\min f_k(x), \ x \in D = [a,b] \subset R^N. \quad (13)$$

Заметим, что при $f_k^* = f^*$ множество глобальных минимумов в (12)-(13) совпадает с множеством глобальных минимумов исходной задачи. Отметим также, что предложенный способ учета ограничений использовался в работе [7] для триангуляционного метода с оцениванием константы Липшица производных по направлениям.

К решению (12)-(13) может быть применен модифицированный метод Direct, отличающийся от стандартного только учетом изменяемости функции в процессе поиска ее минимума. Поскольку метод Direct строит неравномерное, но в пределе всюду плотное покрытие области поиска D [3], значение f_k^* будет стремиться к f^* из (7) при $k \to \infty$, что гарантирует сходимость модифицированного метода.

Второй подход не использует сведения задачи с ограничениями к задачам без ограничений и предлагает прямое специальное обобщение принципов построения метода Direct на задачи с ограничениями.

При втором подходе предполагается, что в задаче (7)-(10) функция f является липшицевой с константой L, значение L неизвестно, и $L \in [0,\infty)$. Кроме того, принимается, что функция g из (10) липшицева с константой L_g, значение константы неизвестно, но взаимосвязано с L соотношением $L_g \in [0, \gamma \cdot L]$, где $\gamma > 0$. Поскольку значение L неограниченно, константа Липшица обобщенного ограничения потенциально также может принимать сколь угодно большие значения.

В рассматриваемой задаче каждый гиперинтервал D_i (его диаметр – d_i) с вычисленными в его центре значениями $f_i = f(c_i)$ и $g_i = g(c_i)$ характеризуется уже не парой, а тройкой чисел (d_i, f_i, g_i), что принципиально отличает ситуацию от задачи без функциональных ограничений (1). При выборе «лучших» гиперинтервалов D_t для их деления на три на очередной итерации необходимо сформулировать новый принцип отбора. Отбираемые «лучшие» гиперинтервалы по-прежнему будем назавать *потенциально оптимальными*. Они будут выбираться из категории гиперинтервалов, которые будем называть *недоминируемыми* по аналогии с терминологией, использованной в [3].

Если временно считать, что значения L и L_g известны, то минимально возможные значения функций f и g в пределах гиперинтервала $D_i = [a_i, b_i]$, построенные по результатам вычисления f и g в центральной точке c_i, определяются соотношениями

$$f_i^-(L) = f^-(L, D_i) = f(c_i) - L \cdot d_i / 2; \ g_i^-(L_g) = g^-(L_g, D_i) = g(c_i) - L_g \cdot d_i / 2. \quad (14)$$

Гиперинтервал D_t будем называть *недоминируемым*, если

$$\exists L \in [0,\infty): \ f_t^-(L) = \min\{ f_i^-(L) : \exists L_g \in [0, \gamma \cdot L] : g_i^-(L_g) \le 0, 1 \le i \le M_k \}, \quad (15)$$

$$\exists L_g \in [0, \gamma \cdot L]: \ g_t^-(L_g) \le 0. \quad (16)$$

Каждому недоминируемому D_t отвечает некоторый диапазон $\{L(D_t)\}$ соответствующих ему (в силу (15)-(16)) значений константы L. *Потенциально оптимальными* назовем такие недоминируемые гиперинтервалы D_t, для которых дополнительно

$$\exists L \in \{L(D_t)\}: \ f_t^-(L) \le f_k^* - \varepsilon \left| f_k^* \right|. \quad (17)$$

Введенное определение потенциальной оптимальности соответствует наличию ограничений в (7)-(8), однако является достаточно сложным. Для его практического использования необходимы простые конструктивные правила, последовательное применение которых приводило бы к отбору потенциально оптимальных D_t. Ниже формулируются правила отбора, включающие несколько этапов.

Этап 1. Гиперинтервалы текущего разбиения D_i разделяются на группы – слои. В каждый слой объединяются все D_i, имеющие один и тот же диаметр $d_i = c$ (будем обозначать их как D_i^c). На множестве гиперинтервалов одного слоя вводится понятие *недоминируемости в слое*, аналогичное (15)-(16). Каждому недоминируемому в слое $\{D_i^c\}$ гиперинтервалу (обозначим их через $\hat{D}_{i(s)}^c$, где $s = 1,2,...$ – введенная порядковая нумерация) будет соответствовать диапазон подходящих для $\hat{D}_{i(s)}^c$ в слое значений констант $\{L(\hat{D}_{i(s)}^c)\} = [L_s^c, L_{s+1}^c)$. Для выделения в слое гиперинтервалов $\hat{D}_{i(s)}^c$ элементы слоя удобно представить точками (g_i, f_i) на плоскости (g, f), как показано на рис. 1.

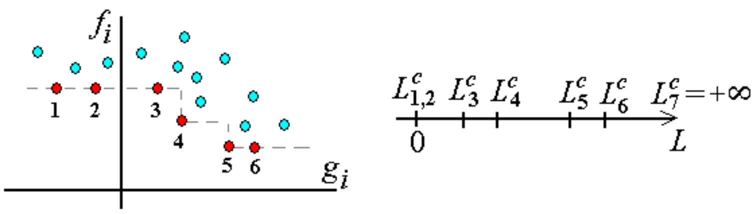

Рис.1. Набор точек соответствует слою $d_i = c$, красные точки отвечают недоминируемым в слое гиперинтервалам $\hat{D}_{i(s)}^c$, $s = 1,2,..,6$

Доказаны следующие леммы.

Лемма 1. Если в слое $\{D_i^c\}$ существуют точки (g_i, f_i) со значениями $g_i \le 0$, то гиперинтервалы D_j^c со значениями $g_j \le 0$, $f_j = f^{*c} = \min\{ f_i : g_i \le 0, D_i \in \{D_i^c\}\}$ являются недоминируемыми в слое (т.е. относятся к типу $\hat{D}_{i(s)}^c$) и для них $L_s^c = 0$.

Если в слое $\{D_i^c\}$ нет гиперинтервалов с $g_j \le 0$, то будем полагать $f_c^* = +\infty$.

Лемма 2. Для того чтобы в слое $\{D_i^c\}$ точка (g_{i^*}, f_{i^*}) с $g_{i^*} > 0$ соответствовала некоторому недоминируемому в слое гиперинтервалу $\hat{D}_{i(s)}^c$ (где $i(s) = i^*$), необходимо и достаточно, чтобы $f_{i^*} \le f^{*c}$ и в слое не существовало другого гиперинтервала D_j, для которого $g_j \le g_{i^*}$, $f_j < f_{i^*}$. Если это выполнено, $L_s^c = 2g_{i^*} / (\gamma \cdot d_{i^*})$, где $i^* = i(s)$.

Перенумеруем индексом s недоминируемые в слое гиперинтервалы $\hat{D}^c_{i(s)}$ в порядке возрастания значений L^c_s. Для гиперинтервала $\hat{D}^c_{i(s)}$ с наибольшим номером s положим $L^c_{s+1} = +\infty$. Правила отбора, выполняемые на этапе 1, иллюстрирует рис.1.

Этап 2. Выполняется упорядочение по возрастанию найденных значений L^c_s для всех s и всех слоев c. Полученный набор значений обозначим как $0 \le L_1 < L_2 < .. < L_{m_k} < L_{m_{k+1}} = +\infty$. Для каждого промежутка $[L_j, L_{j+1})$ выделяются все найденные гиперинтервалы $\hat{D}^c_{i(s)}$ из разных слоев c, для которых $[L_j, L_{j+1}) \cap [L^c_s, L^c_s) \ne \varnothing$. Набор таких гиперинтервалов обозначим $\{\tilde{D}^j_\ell\}$ $(\ell = 1,.., \bar{\ell}(j))$. По построению для значений $L \in [L_j, L_{j+1})$ недоминируемые по совокупности слоев гиперинтервалы могут содержаться только в наборе $\{\tilde{D}^j_\ell\}$.

Этап 3. Для всех $j = 1,..., m_k$ выполняется взаимное сравнение гиперинтервалов в пределах каждого из наборов $\{\tilde{D}^j_\ell\}$ $(\ell = 1,.., \bar{\ell}(j))$. Сравнение проводится на плоскости (d_i, f_i) с выделением *потенциально оптимальных* гиперинтервалов D_t по принципу:

$$D_t \in \{\tilde{D}^j_\ell : \ell = 1,.., \bar{\ell}(j)\},$$

$$\exists L \in [L_j, L_{j+1}):\ \forall \tilde{D}^j_\ell, (\ell = 1,.., \bar{\ell}(j)):\ f^-(L, D_t) \le f^-(L, \tilde{D}^j_\ell), \qquad (18)$$

$$f^-(L, D_t) \le f^*_k - \varepsilon |f^*_k|. \qquad (19)$$

Все выделенные таким образом гиперинтервалы D_t перед началом следующей итерации разделяются на три равные части по их большим ребрам с вычислением функций f и g в новых центрах вновь образованных гиперинтервалов.

Доказано, что для построенного метода $f^*_k \to f^*$ при $k \to \infty$. В докладе будут приведены примеры решения тестовых задач.

Работа выполнена при финансовой поддержке Минобрнауки России, государственное соглашение о предоставлении гранта № 14.B37.21.0878.

Литература

1. Jones D.R., Perttunen C.D., Stuckman B.E. Lipschitzian optimization without the Lipschitz constant // J. Optim. Theory and Appl. 1993. V. 79. № 1. P. 157–181.
2. Евтушенко Ю.Г., Ратькин В.А. Метод половинных делений для глобальной оптимизации функций многих переменных // Изв. АН СССР. Техническая кибернетика. 1987. Т.1. С. 119-127.
3. Сергеев Я.Д., Квасов Д.Е. Диагональные методы глобальной оптимизации. – М.: Физматлит, 2008. – 352 с.
4. Препарата Ф.Ф., Шеймос М.И. Вычислительная геометрия. Введение. – М.: Мир, 1989. – 478 с.
5. Sergeyev Ya.D., Kvasov D.E. A univariate global search working with a set of Lipschitz constants for the first derivative // Optimization Lettes. 3. 2009. – P. 303-318.
6. Kvasov D.E., Sergeyev Ya.D. Lipschitz gradients for global optimization in a one-point-based partitioning scheme // Journal of Computational and Applied Mathematics. 2012. V. 236. P. 4042–4054.
7. Городецкий С.Ю. Триангуляционные методы параболоидов в задачах многоэкстремальной оптимизации с ограничениями для класса функций с липшицевыми производными по направлениям // Вестник Нижегородского университета им. Н.И. Лобачевского. № 1 (1). 2012. – С. 144–155.