ДОКЛАДЫ АКАДЕМИИ НАУК, 2011, том 437, № 2, с. 168-172

ИНФОРМАТИКА

УДК 518.853

ВАРИАНТЫ МЕТОДА НЕРАВНОМЕРНЫХ ПОКРЫТИЙ ДЛЯ ГЛОБАЛЬНОЙ ОПТИМИЗАЦИИ ЧАСТИЧНО-ЦЕЛОЧИСЛЕННЫХ НЕЛИНЕЙНЫХ ЗАДАЧ

© 2011 г. Академик Ю.Г. Евтушенко, М.А. Посыпкин

Метод неравномерных покрытий для поиска глобального экстремума функций многих переменных был предложен Ю.Г. Евтушенко в 1971 г. [1]. Дальнейшее развитие этот подход нашел в многочисленных работах. Укажем лишь некоторые из них [2]–[5]. Различные варианты метода были программно реализованы и использовались для расчетов на многопроцессорных системах [4, 5].

В данной работе дано более общее, чем в [1, 2], трактование метода неравномерных покрытий. Метод применен к простейшей задаче нелинейного программирования, в которой найден глобальный изолированный минимум. Расчеты проведены с использованием и без использования требования целочисленности. Введение условия целочисленности существенно сократило время расчетов.

Для непрерывной функции $f : \mathbb{R}^n \to \mathbb{R}$ ставится задача отыскания глобального минимума на допустимом множестве $X \subseteq \mathbb{R}^n$:

$$f_* = \text{glob} \min_{x \in X} f(x) = f(x_*), \tag{1}$$

где x_* — любая точка, в которой достигается глобальный минимум, равный f_*. Для этой задачи определим множество решений X_* и множество ϵ-оптимальных решений X_ϵ:

$$\begin{array}{rcl} X_* & = & \{x \in X : f(x) = f_*\}, \\ X_\epsilon & = & \{x \in X : f(x) \le f_* + \epsilon\}, \qquad \epsilon > 0. \end{array} \tag{2}$$

Предполагаем, что множество X_* непусто. Требуется найти хотя бы одну точку из множества X_ϵ.

Для множества $Z \subseteq \mathbb{R}^n$ функции $f(x) : \mathbb{R}^n \to \mathbb{R}$ и числа $\lambda \in \mathbb{R}$ определим лебеговское множество $S(Z, f(x), \lambda) = \{x \in Z : f(x) \ge \lambda\}$ и открытое лебеговское множество $S'(Z, f(x), \lambda) = \{x \in Z : f(x) > \lambda\}$.

Рассмотрим совокупность множеств $\{X_i\}$, $X_i \subseteq \mathbb{R}^n$, $i = 1, 2, \ldots, k$. На множестве X_i определим миноранту $\mu_i(x)$, такую что $f(x) \ge \mu_i(x)$ для всех $x \in X_i$. Пусть заданы допустимая точка $x_r \in X$ и совокупность множеств $\{S_i\}$, удовлетворяющих условиям

$$S_i \subseteq S(X_i, \mu_i(x), f(x_r) - \epsilon), \qquad i = 1, 2, \ldots, k. \tag{3}$$

Будем говорить, что совокупность множеств $\{S_i\}$ покрывает множество X, если

$$X \subseteq \bigcup_{i=1}^{k} S_i. \tag{4}$$

Теорема 1. *Для того чтобы допустимая точка x_r была ϵ-оптимальным решением задачи* (1) *и на множестве X имела место оценка*

$$f(x_r) \geq f_* \geq f(x_r) - \epsilon, \tag{5}$$

достаточно существования совокупностей множеств $\{X_i\}$ и минорант $\{\mu_i(x)\}$, для которых выполнено условие покрытия (4).

Теорема 1 является теоретической базой для реализации различных вычислительных схем метода неравномерных покрытий. При этом множество X может иметь произвольную мощность и структуру. Для непосредственного применения теоремы 1 при поиске глобального минимума требуется определить способ нахождения допустимой точки x_r и проверки условия (4). Рассмотрим случай, когда в задаче (1) допустимое множество задано с помощью функциональных ограничений:

$$X = \{x \in \mathbb{R}^n : g(x) \leq 0_m\}, \tag{6}$$

где $g(x) : \mathbb{R}^n \to \mathbb{R}^m$ — непрерывная вектор-функция.

Будем предполагать, что допустимое множество X ограничено и $X \subseteq P$, где P — n-мерный параллелепипед.

Для $\delta \in \mathbb{R}$ определим δ-допустимое множество следующим образом:

$$X^\delta = \{x \in \mathbb{R}^n : g^j(x) \leq \delta,\ j = 1, 2, \ldots, m\}. \tag{7}$$

Множество X^δ может быть эквивалентным образом определено с использованием функции

$$\phi(x) = \max\left(g^1(x), \ldots, g^m(x)\right) : X^\delta = \{x \in P : \phi(x) \leq \delta\}.$$

Так как $g(x)$ — непрерывная функция, то функция $\phi(x)$ также непрерывна.

Введем обозначения: $\underline{\delta} = \inf\{\delta : X^\delta \neq \emptyset\}$, $\bar{\delta} = \sup\{\delta : X^\delta \subseteq P\}$.

При $\underline{\delta} < \delta < \bar{\delta}$ положим

$$f_*^\delta = \min_{x \in X^\delta} f(x).$$

Функция f_*^δ называется функцией чувствительности. Хорошо известны следующие ее свойства [6, 7]:

1) $f_*^0 = f_*$;

2) множество X^δ не сужается, а функция чувствительности монотонно не возрастает с увеличением δ:

$$X^{\delta_1} \subseteq X^{\delta_2}, \quad f_*^{\delta_1} \geq f_*^{\delta_2} \quad \text{для любых} \quad \delta_1, \delta_2, \qquad \underline{\delta} < \delta_1 \leq \delta_2 < \bar{\delta}; \tag{8}$$

3) для любого δ_0, $\underline{\delta} < \delta_0 < \bar{\delta}$, функция чувствительности непрерывна справа:

$$\lim_{\delta \to \delta_0 + 0} f_*^\delta = f_*^{\delta_0}. \tag{9}$$

Для $\epsilon \in \mathbb{R}_+$ и δ_1, δ_2, $\underline{\delta} < \delta_1 \leq \delta_2 < \bar{\delta}$, определим множество

$$X_\epsilon^{\delta_1, \delta_2} = \left\{x \in X^{\delta_2} : f(x) \leq f_*^{\delta_1} + \epsilon\right\}.$$

В частном случае $\delta_1 = 0$, $\delta_2 = \delta$ это множество будем называть множеством ϵ, δ-оптимальных решений и обозначать через X_ϵ^δ:

$$X_\epsilon^\delta = \{x \in X^\delta : f(x) \leq f_* + \epsilon\}.$$

Рассмотрим совокупность множеств $P_1, P_2, \ldots, P_k$, $P_i \subseteq P$. Определим миноранты $\mu_i(x)$, такие что $f(x) \geq \mu_i(x)$ для всех $x \in P_i$, и миноранты $\nu_i(x)$, такие что $\phi(x) \geq \nu_i(x)$ для всех $x \in P_i$. Пусть $\underline{\delta} < \delta_1 \leq \delta_2 < \bar{\delta}$, $x_r \in X^{\delta_2}$ и множества S_i удовлетворяют соотношению

$$S_i \subseteq S(P_i, \mu_i(x), f(x_r) - \epsilon) \cup S'(P_i, \nu_i(x), \delta_1). \tag{10}$$

Условие покрытия в данном случае имеет вид

$$P = \bigcup_{i=1}^{k} S_i. \tag{11}$$

Справедлива следующая

Теорема 2. *Пусть $\underline{\delta} < \delta_1 \leq \delta_2 < \bar{\delta}$. Пусть также заданы точка $x_r \in X^{\delta_2}$ и совокупность множеств $\{P_i\}$, такие что выполнено условие покрытия* (11)*. Тогда $x_r \in X_\epsilon^{\delta_1,\delta_2}$ и справедлива оценка*

$$f_*^{\delta_1} + \epsilon \geq f(x_r) \geq f_*^{\delta_2}. \tag{12}$$

Выделим два наиболее значимых с точки зрения практики частных случая теоремы 2. В первом случае $\delta_1 < 0$, $\delta_2 = 0$ точка x_r будет приближенным допустимым решением задачи, для которого справедливо неравенство $f_*^{\delta_1} + \epsilon \geq f(x_r) \geq f_*$. Из-за отсутствия непрерывности слева в этом случае приближенное и точное решения могут сколь угодно различаться. Кроме того, для задач, где $\delta = 0$, данный подход неприменим.

Во втором случае $\delta_1 = 0$, $\delta_2 > 0$ точка x_r будет приближенным, но, возможно, недопустимым решением, для которого справедливо неравенство $f_* + \epsilon \geq f(x_r) \geq f_*^{\delta_2}$. Согласно свойству (9) и теореме 2, с уменьшением ϵ и δ_2 значение $f(x_r)$ будет стремиться к f_*.

Рассмотрим одну из возможных схем реализации метода неравномерных покрытий — алгоритм половинных делений (бисекций), предложенный в [2, 3]. Ограничимся рассмотрением частного случая $\delta_1 = 0$, $\delta_2 = \delta > 0$. Другой важный случай $\delta_1 < 0$, $\delta_2 = 0$ рассматривается по аналогии. Как и выше, предполагается, что задан n-мерный параллелепипед P с гранями, параллельными координатным плоскостям, содержащий множество X^δ. В методе бисекций множества P_i также являются параллелепипедами, которые получаются в результате разбиения исходного.

Пусть набор $\{x_i\}$ состоит из точек, принадлежащих P. В каждой точке x_i вычисляется значение функции $f(x_i)$ и определяется рекордная точка x_r среди точек, принадлежащих множеству X^δ, и рекордная точка x'_r из X с использованием следующего правила.

Правило REC-UPDATE. Если $x_i \in X^\delta$ и при этом $f(x_i) < f(x_r)$ либо x_r не определена, то выполнить присваивание $x_r := x_i$. Если $x_i \in X$ и при этом $f(x_i) < f(x'_r)$ либо x'_r не определена, то выполнить присваивание $x'_r := x_i$.

Помимо рекордной точки x_r определим также рекордное значение

$$f_r = \begin{cases} f(x_r), & \text{если } x_r \text{ определена,} \\ \infty & \text{в противном случае.} \end{cases}$$

Пусть функции $f(x)$ и $\phi(x)$ удовлетворяют условию Липшица с константами ℓ_1 и ℓ_2 соответственно. Пусть точка x_i — центр параллелепипеда P_i. Определим опорные миноранты $\mu_i(x)$ и $\nu_i(x)$ для функций $f(x)$ и $\phi(x)$ следующим образом:

$$\mu_i(x) = f(x_i) - \ell_1\|x - x_i\|, \qquad \nu_i(x) = \phi(x_i) - \ell_2\|x - x_i\|. \tag{13}$$

Теорема 3. *Пусть ℓ_1, ℓ_2 — константы Липшица для функций $f(x)$ и $\phi(x)$. Тогда справедливы следующие два утверждения.*

1. *Множество $S_i = \{x \in P_i : \|x - x_i\| < \rho_i\}$, где*

$$\rho_i = \max\left\{\frac{f(x_i) - f_r + \epsilon}{\ell_1}, \frac{\phi(x_i)}{\ell_2}\right\} \tag{14}$$

удовлетворяет включению (10).

2. *Если применяется правило* REC-UPDATE, *то $\rho_i \geq \rho_*$, где $\rho_* = \min\left\{\frac{\epsilon}{\ell_1}, \frac{\delta}{\ell_2}\right\}$.*

Приведем простейший вариант метода, использующий половинные деления параллелепипедов по максимальному ребру.

Алгоритм BISECT

Входные параметры:

P — исходный параллелепипед;

ϵ, δ — заданная точность по целевой функции и ограничениям;

ℓ_1, ℓ_2 — константы Липшица для целевой функции и ограничений.

Выходные параметры:

x_r — рекордная точка;

x'_r — допустимая рекордная точка.

Внутренние переменные:

$\mathcal{L}$ — список покрываемых параллелепипедов.

1. Поместить параллелепипед P в список $\mathcal{L}$: $\mathcal{L} = \{P\}$, положить $f_r = \infty$.
2. Выбрать и исключить из списка текущий параллелепипед $P_i \in \mathcal{L}$.
3. Вычислить значение функции в центре x_i параллелепипеда P_i и обновить рекордные точки x_r и x'_r согласно правилу REC-UPDATE.
4. Взять в качестве центра покрывающего шара точку x_i а радиус покрывающего шара ρ_i вычислить по формуле (14).
5. Если покрывающий шар содержит P_i, то выполнить переход к шагу 2. В противном случае разбить параллелепипед P_i по направлению самого длинного ребра на два одинаковых новых параллелепипеда, которые добавить к списку $\mathcal{L}$.
6. Если список пуст, то завершить работу алгоритма, в противном случае перейти к шагу 2.

Шаги 2–6 алгоритма BISECT циклически повторяются. Теорема 4 дает верхнюю оценку на число итераций этого цикла.

Теорема 4. *Алгоритм* BISECT *завершится за конечное число итераций, не превосходящее величины*

$$4\left(\frac{\rho_*}{\rho}\right)^{\theta(n)} - 1, \tag{15}$$

где

$$\theta(n) = \frac{2}{\log_2\left(1 - \dfrac{3}{4n}\right)}$$

и ρ *— радиус (половина расстояния между двумя самыми удаленными точками) параллелепипеда* P.

Теорема 5. *Если алгоритм* BISECT *завершился за конечное количество шагов, то справедливы следующие два утверждения.*

1. *Если рекордная точка* x_r *не найдена, то допустимое множество* X *задачи* (6) *пусто. В противном случае найденная рекордная точка* x_r *будет* ϵ, δ*-оптимальным решением задачи* (6)*, для которого справедливо неравенство*

$$f_* + \epsilon \geq f_r \geq f_*^\delta. \tag{16}$$

2. *Если найдена допустимая рекордная точка* x'_r*, то допустимое множество решений задачи* (6) *непусто и справедливо неравенство*

$$f(x'_r) \geq f_*. \tag{17}$$

Рассмотренный алгоритм можно с успехом использовать для решения задач частично-целочисленного программирования. Для этого в алгоритме BISECT нужно модифицировать способ разбиения параллелепипеда и правило обновления рекорда. Рассмотрим задачу (1), в которой к ограничениям (6) также добавляются условия целочисленности $x_j \in \mathbb{Z}$, $j \in J$, где J — подмножество индексов целочисленных переменных.

Бисекция производится также с учетом целочисленности. Пусть $Q = [a^1, b^1] \times \times [a^2, b^2] \times \cdots \times [a^n, b^n]$ — параллелепипед, подвергаемый декомпозиции, и

$$s = \arg \min_{i=1,\dots,n} |a^i - b^i|$$

есть индекс самого длинного ребра параллелепипеда. В результате разбиения получаются два параллелепипеда

$$\begin{aligned} Q_1 &= [a^1, b^1] \times \cdots \times [a^s, c^s] \times \cdots \times [a^n, b^n], \\ Q_2 &= [a^1, b^1] \times \cdots \times [d^s, b^s] \times \cdots \times [a^n, b^n], \end{aligned}$$

где

$$c^s = \begin{cases} \left\lfloor \dfrac{a^s + b^s}{2} \right\rfloor, & \text{если } s \in J, \\ \dfrac{a^s + b^s}{2}, & \text{если } s \notin J, \end{cases} \qquad d^s = \begin{cases} \left\lceil \dfrac{a^s + b^s}{2} \right\rceil, & \text{если } s \in J, \\ \dfrac{a^s + b^s}{2}, & \text{если } s \notin J. \end{cases}$$

Перед применением правила REC-UPDATE точка x_i подвергается следующему преобразованию:

$$x_i^j := \begin{cases} \lfloor x_i^j \rceil, & \text{если } j \in J, \\ x_i^j, & \text{если } j \in J. \end{cases}$$

Остальные шаги алгоритма BISECT остаются без изменений.

На практике применяются различные приемы ускорения процесса поиска решения в рассмотренной выше базовой реализации метода бисекций. В частности, для дифференцируемых функций более эффективным является использование следующей миноранты:

$$\mu_i(x) = f(x_i) + \langle f_x(x_i), x - x_i \rangle + \frac{k_i}{2} \|x - x_i\|^2, \tag{18}$$

где k_i — нижняя граница спектра гессиана функции $f(x)$ на параллелепипеде P_i. Также ускорению расчетов способствует сокращение области поиска с помощью исключения части текущего параллелепипеда, не содержащей оптимум, в случаях, когда нельзя удалить его целиком. Результаты, касающиеся конечности числа шагов базового варианта алгоритма BISECT и свойств приближенных решений, остаются справедливыми и в случае усложненных вариантов.

Различные варианты метода неравномерных покрытий успешно использовались для решения задач большой размерности. Например, в [4] приведены результаты расчетов задачи минимизации энергии взаимодействия молекулярного кластера. При этом с помощью метода неравномерных покрытий успешно решалась задача минимизации функции 255 переменных.

Здесь ограничимся только рассмотрением следующего простейшего примера, взятого из [8]:

$$\begin{aligned} f(x) &= x^1 \to \min, \\ g^1(x) &= (x^1-5)^2 + 2(x^2-5)^2 + (x^3-5)^2 - 18 \le 0, \\ g^2(x) &= 100 - (x^1+7-2x^2)^2 - 4(2x^1+x^2-11)^2 - 5(x^3-5)^2 \le 0. \end{aligned} \tag{19}$$

Особенностью данной задачи является тот факт, что глобальный минимум достигается в изолированной допустимой точке $x_* = (1,4,5)$, $f(x_*) = 1$, $g^1(x_*) = 0$, $g^2(x_*) = 0$ (в [8] ошибочно указана точка (1,3,5)). Примененный в [8] метод поиска глобального минимума определил допустимую точку $\bar{x} = (3.747692, 7.171420, 2.362317)$, $f(\bar{x}) = 3.747692$ за 33.703 секунды на компьютере Pentium IV 2.53 GHz.

Для применения метода неравномерных покрытий в качестве множества P был выбран параллелепипед $-10 \le x^i \le 10$, $i = 1,2,3$, заведомо содержащий допустимое множество задачи (19). Использовались миноранты (18). Полагалось $\epsilon = |\delta|$. Эксперименты проводились на персональном компьютере с процессором Intel Core 2 Quad, 2.33 GHz. Использовалось одно ядро процессора. Результаты экспериментов приведены в табл. 1, значение $f(x_r)$ совпадает с первой компонентой вектора x_r, приведенного во втором столбце.

Таблица 1. Результаты экспериментов при различных значениях δ

δ	Рекорд x_r	$\phi(x_r)$	Время расчетов, с	Число итераций
-0.01	$(3.722, 7.276, 7.452)$	-0.00204	0.61	10602
-0.0001	$(3.721, 7.15, 2.331)$	-0.00001	24.57	506351
0.0001	$(0.997, 4.007, 4.999)$	0.00007	7.94	165547
0.01	$(0.965, 4.071, 4.994)$	0.00993	0.16	2671

При $\delta > 0$ получены приближенные решения, близкие к оптимальному решению x_*. Найденные при $\delta < 0$ точки допустимы и близки к точке $\bar{x}$; при этом полученное значение целевой функции меньше, чем найденное в [8], в третьем знаке.

Данная задача была использована для тестирования возможностей нахождения целочисленного решения. Для этого все переменные считали целочисленными. Расчеты проводили в трех различных вариантах при $\epsilon = \delta = 0$. В первом использовалась базовая схема алгоритма BISECT, применялись миноранты (13), константы Липшица оценивались сверху с помощью интервального анализа на текущих параллелепипедах. Во втором варианте использовались миноранты (18). В третьем варианте дополнительно применялась техника исключения части параллелепипеда, не содержащей оптимального решения. Глобальный минимум был найден за 585, 121, 55 итераций в первом, втором и третьем вариантах соответственно. Таким образом, в данной задаче введение условия целочисленности существенно ускоряет расчеты. Поэтому при решении сложных задач представляется целесообразным вводить это условие искусственно для нахождения приближенного решения исходной непрерывной задачи.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (коды проектов 09-01-12098-офи_м, 10-01-90011-БЕЛ_а), Программы П-14 Президиума РАН и программы грантов Президента РФ для поддержки ведущих научных школ (проект НШ-4096.2010.1).

Литература

1. *Евтушенко Ю.Г.* // ЖВМиМФ. 1971. Т. 11. № 6. С. 1390–1403.

2. *Евтушенко Ю.Г., Ратькин В.А.* // Техн. кибернетика. 1987. № 1. С. 119–127.

3. *Ratchek H., Rokne J.* New Computer Methods for Global Optimization. Chichester: Ellis Horwood, 1988.

4. *Евтушенко Ю.Г., Малкова В.У., Станевичюс А.А.* // ЖВМиМФ. 2009. Т. 49. № 2. С. 255–269.

5. *Evtushenko Y., Posypkin M., Sigal I.* // Comput. Sci. Res. and Develop. 2009. V. 23. № 3. P. 211–215.

6. *Evans J.P., Gould F.J.* // Opns. Res. 1970. V. 18. P. 107–118.

7. *Голиков А.И., Коткин Г.Г.* // ЖВМиМФ. 1988. Т. 28. № 10. С. 1461–1474.

8. *Tuy H.* // J. Glob. Optimization. 2010. V. 47. P. 485–501.