КОМПЬЮТЕРНАЯ МАТЕМАТИКА

Теория и методы оптимизации

УДК 519.615

В.Ю. СЕМЕНОВ

МЕТОД ГЛОБАЛЬНОЇ МІНІМІЗАЦІЇ ФУНКЦІЙ, ЗАСНОВАНИЙ НА РОЗВ'ЯЗАННІ СИСТЕМ НЕЛІНІЙНИХ РІВНЯНЬ

Запропоновано метод глобальної мінімізації функції від декількох змінних на заданому інтервалі. Метод засновано на розв'язанні системи нелінійних рівнянь, утвореної частковими похідними цільової функції. Застосування методу продемонстровано на чисельних прикладах.

© В.Ю. Семенов, 2018

Вступ. Розглянемо задачу глобальної мінімізації функції декількох змінних:

$$\min f(x_1,...,x_n),$$
$$a_1 \le x_1 \le b_1,...,a_n \le x_n \le b_n. \qquad (1)$$

Припустимо, що $f(x_1,...,x_n) \in C^{(3)}(D)$ (позначення $C^{(l)}(D)$ використовується для класу l разів диференційовних функцій на інтервалі $D=[a_1;b_1]\times...\times[a_n;b_n]$).

У відповідності з традиційною термінологією [1], вектор $\hat{x}$ зветься локальним мінімумом, якщо $f(\hat{x}) \le f(x)$ для усіх x з деякого околу вектора $\hat{x}$. Вектор $\hat{x}$ зветься глобальним мінімумом, якщо $f(\hat{x}) \le f(x)$ для усіх $x \in D$. Ми також називаємо локальний мінімум внутрішнім, якщо він не належить до границі інтервалу D.

Існує багато методів мінімізації різних класів функцій (див., наприклад, [1]). Серед них можна згадати методи імітації відпалу (simulated annealing), генетичні алгоритми, методи «гілок і меж» (branch and bound), градієнтні методи та інші. Найчастіше, ці методи дозволяють знайти один або декілька локальних мінімумів функції f. Проте в багатьох практичних задачах ставиться задача глобальної мінімізації цільової функції.

У статті наведено метод глобальної мінімізації функції, що базується на розв'язанні системи нелінійних рівнянь, яка утворена частковими похідними (градієнтом) цільової функції:

$$\begin{cases} \partial f / \partial x_1(x_1,...,x_n) & = & 0; \\ \partial f / \partial x_2(x_1,...,x_n) & = & 0; \\ \quad ... & & ... \\ \partial f / \partial x_n(x_1,...,x_n) & = & 0; \end{cases} \qquad (2)$$

на інтервалі $D = [a_1;b_1] \times ... \times [a_n;b_n]$.

Запропонований метод заснований на принципі «гілок і меж» (branch-and-bound) [1] і виконує рекурсивний поділ інтервалу D з метою вибору підінтервалів, кожен з яких містить один з глобальних мінімумів цільової функції. Даний метод – це розвиток методу локалізації коренів систем нелінійних рівнянь [2]. Практичне застосування методу продемонстровано на декількох прикладах.

Допоміжні твердження. Як відомо, внутрішні локальні мінімуми є розв'язками системи (2), що мають позитивно визначений гессіан

$$F'' = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & ... & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ ... & ... & ... \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & ... & \frac{\partial^2 f}{\partial x_1 \partial x_n} \end{pmatrix}.$$

Згідно з критерієм Сільвестра, це означає що головні мінори $\psi_1, \psi_2, ..., \psi_n$ матриці F'' мають бути невід'ємними.

Припустимо, що для точок інтервалу $x = (x_1,..., x_n) \in D$ – правильні наступні обмеження:

$$\left| \frac{\partial f}{\partial x_i} \right| \le M_i, \quad i = 1,...,n; \qquad (3)$$

$$\left| \frac{\partial^2 f}{\partial x_i \partial x_k} \right| \le M_{ik}, \quad i,k = 1,...,n; \qquad (4)$$

$$\left| \frac{\partial \varphi_i}{\partial x_k} \right| \le M'_{ik}, \ \ i,k = 1,...,n. \qquad (5)$$

Далі наведемо твердження щодо поведінки функції f та її похідних на n-мірному інтервалі $D_0 = [c_1, d_1] \times [c_2, d_2] \times ... \times [c_n, d_n], D_0 \subset D$ з центром $x^{(0)} = [0.5(c_1 + d_1), 0.5(c_2 + d_2),..., 0.5(c_n + d_n)]$. Ці твердження засновані на розкладі функції f в ряд Тейлора та слідують з леми 1 [2].

Теорема 1. Нехай $f(x_1,...,x_n) \in C^{(1)}(D_0)$ та виконуються обмеження (3). Тоді значення функції f на інтервалі D_0 лежать у наступних межах:

$$f(x^{(0)}) - \frac{1}{2}\sum_{i=1}^{n} M_i (d_i - c_i) \le f(x) \le f(x^{(0)}) + \frac{1}{2}\sum_{i=1}^{n} M_i (d_i - c_i). \qquad (6)$$

Ця теорема дозволяє оцінити верхню та нижню границі функції f на інтервалі D_0.

Теорема 2. Нехай $f(x_1,...,x_n) \in C^{(2)}(D_0)$ та виконуються обмеження (4). Тоді $\forall x \in D_0$

$$\partial f / \partial x_i (x^{(0)}) - \frac{1}{2}\sum_{k=1}^{n} M_{ki}(d_i - c_i) \le \partial f / \partial x_i (x) \le \partial f / \partial x_i (x^{(0)}) + \frac{1}{2}\sum_{k=1}^{n} M_{ki}(d_i - c_i). \qquad (7)$$

Також, якщо існує номер i, такий що

$$\left| \partial f / \partial x_i (x^{(0)}) \right| > \frac{1}{2}\sum_{k=1}^{n} M_{ki}(d_i - c_i), \qquad (8)$$

то похідна $\partial f / \partial x_i$ не обнуляється на інтервалі D_0 і, як наслідок, функція f не може містити на D_0 більше одного внутрішнього локального мінімуму.

Теорема 3. Нехай $f(x_1,...,x_n) \in C^{(3)}(D_0)$ і виконуються границі (5). Тоді $\forall x \in D_0$

$$\psi_k(x^{(0)}) - \frac{1}{2}\sum_{i=1}^{n} M'_{ki}(d_i - c_i) \le \psi_k(x) \le \psi_k(x^{(0)}) + \frac{1}{2}\sum_{i=1}^{n} M'_{ki}(d_i - c_i).$$

Також, якщо існує номер i, такий що

$$\psi_i(x^{(0)}) < -\frac{1}{2}\sum_{k=1}^{n} M'_{ki}(d_i - c_i), \qquad (9)$$

то $\psi_i(x) < 0, \forall x \in D_0$ і, як наслідок, інтервал D_0 не може містити внутрішній локальний мінімум.

Окрім того, якщо $\forall i, \ 1 \le i \le n$ виконується умова

$$\psi_i(x^{(0)}) > \frac{1}{2}\sum_{k=1}^{n} M'_{ki}(d_i - c_i), \qquad (10)$$

то $\psi_i(x) > 0, \forall x \in D_0, \ 1 \le i \le n.$ Внаслідок цього функція f є опуклою на інтервалі D_0 і тому не може містити більше одного локального мінімуму.

Зауваження. Таким чином, якщо на деякому n-мірному інтервалі D_0 умова (10) виконується для всіх головних мінорів, інтервал D_0 може містити один локальний мінімум або жодного (але не більше одного). Точна наявність мінімуму може бути встановлена шляхом перевірки збіжності методу Ньютона в інтервалі D_0 (шляхом мульті-старту методу Ньютона з декількох внутрішніх точок або перевіркою умови Канторовича [3]).

Алгоритм глобальної мінімізації. Як відомо, глобальний мінімум функції досягається на границі області його пошуку або у точках локального мінімуму всередині області пошуку мінімуму. Тому структура запропонованого методу складається з двох частин: 1) пошук глобальних мінімумів на границі інтервалу D; 2) пошук глобальних мінімумів всередині інтервалу D. Перша частина зводиться до мінімізації меншої розмірності і, таким чином, може бути розв'язана рекурсивним шляхом (в одновимірному випадку це означає порівняння значень функції у граничних точках інтервалу). Як результат першої частини, ми отримуємо точку глобального мінімуму функції f на границі інтервалу D, а також відповідне мінімальне значення f_{min} (яке буде використовуватись як початкове наближення для другої частини пошуку мінімуму).

Розглянемо структуру алгоритму глобальної мінімізації функції f на інтервалі D.

Призначення алгоритму. Знаходження глобального мінімуму (або мінімумів) функції f на інтервалі $D = [a_1; b_1] \times ... \times [a_n; b_n]$.

Вхідні дані. Формули для обчислення функції f, її часткових похідних $\partial f / \partial x_k, k = 1,...,n$ і мінорів $\psi_k, k = 1,...,n$; границі M_i, M_{ki} і M'_{ki} що входять у обмеження (3), (4), (5), ε – мінімально припустимий розмір аналізуємого інтервалу (параметр ε дозволяє уникнути безкінечної рекурсії у випадку похибок заокруглення та спільних коренів $\partial f / \partial x_k$ і $\psi_k, k = 1,...,n$).

Вихідні дані. f_{min} – оцінка мінімального значення функції f на інтервалі D і $x^{(k)}, k = 1,...,K$ – відповідні точки глобального мінімуму; множина підінтервалів R з розміром (діагоналлю) меншим за ε (для яких ситуація є невизначеною).

Алгоритм.

Алгоритм мінімізації складається з наступних кроків, що виконуються для поточного інтервалу D_0. На початку алгоритму виконується ініціалізація $D_0 = D$.

1. Якщо розмір інтервалу D_0 не перевищує ε, D_0 додається до множини невизначених інтервалів R.

2. Оцінюється діапазон значень $[M_{\min}, M_{\max}]$ функції f на інтервалі D_0 за допомогою нерівностей (6). Якщо $M_{\max} < f_{\min}$, тоді виконується присвоєння $f_{\min} := M_{\max}$. Якщо $M_{\min} > f_{\min}$, то інтервал D_0 виключається з розгляду. Якщо жодна з цих двох умов не виконується, переходимо до наступного кроку.

3. Виконується перевірка умов (8). Якщо хоча б одна з часткових похідних не обнуляється, інтервал D_0 виключається з розгляду. В інших випадках, переходимо до наступного кроку.

4. Виконується перевірка умов (9), (10). Якщо умова (9) виконується хоча б для одного з головних мінорів, інтервал D_0 виключається з розгляду. Якщо умова (10) виконується для усіх головних мінорів, приймається рішення про наявність локального мінімуму функції f на інтервалі D_0. Значення глобального мінімуму уточнюється за допомогою методу Ньютона (його використання є коректним завдяки несінгулярності Гессіану, тобто $\psi_n > 0$). Якщо обчислене мінімальне значення $M_{\min}$ є меншим за $f_{\min}$, тоді виконується присвоєння $f_{\min} := M_{\min}$. В іншому випадку, інтервал D_0 виключається з розгляду. Якщо зазначені умови не виконуються, інтервал D_0 розділяється на рівні частини, для кожної з яких виконуються кроки 1 – 4.

У відповідності з теоремою 4 [2], якщо часткові похідні функції f та мінори ψ_k не мають спільних коренів, алгоритм гарантовано припинить роботу за кінчену кількість кроків, у результаті чого будуть знайдені всі глобальні мінімуми. Зауважимо, що для прискорення роботи алгоритму доцільним є використання власних обмежень (3), (4), (5) для кожного інтервалу D_0. Крім того, перевірки, що містяться в теоремах 1 – 3 можуть бути покращені за рахунок використання похідних вищих порядків.

Приклади.

Приклад 1. Розглянемо глобальну мінімізацію функції

$$f(x) = x^3/3 - \cos(x) - 1.5x^2 \log(x) + 0.75x^2 - 1.5x$$

на інтервалі [0.2; 6]. Застосування алгоритму [2] для локалізації всіх нулів похідної $f'(x) = x^2 + \sin(x) - 3x\log(x) - 1.5x$ дає три значення: $x_1 = 0.3147$, $x_2 = 1.4305$, $x_3 = 5.5232$. Безпосередня перевірка показує, що x_1, x_3 – локальні мінімуми, а x_2 – локальний максимум. Порівнюючи значення функції $f(x) = x^3/3 - \cos(x) - 1.5x^2\log(x) + 0.75x^2 - 1.5x$ у точках x_1, x_3 та граничних точках $\{0.2, 6\}$, робимо висновок що глобальним мінімумом функції $f(x)$ на інтервалі [0.2; 6] є $x_3 = 5.5232$. При цьому, алгоритм [2] потребує використання шести ділень початкового інтервалу.

Тепер розглянемо глобальну мінімізацію функції $f(x)$ за допомогою запропонованого алгоритму. Відповідне дерево роботи алгоритму представлене на рисунку. За рахунок усього трьох рекурсивних ділень початкового інтервалу, алгоритм дозволив знайти глобальний мінімум $x_1 = 5.5232$ без обчислення інших локальних мінімумів.

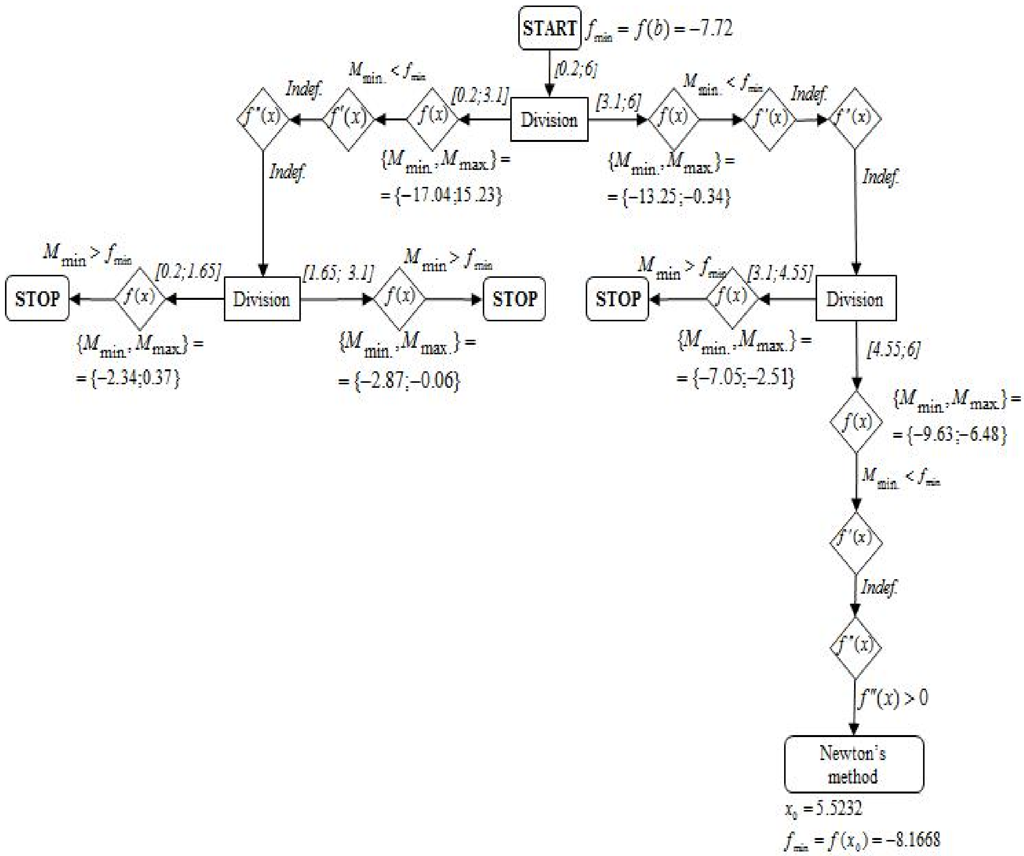

РИСУНОК. Застосування запропонованого алгоритму до глобальної мінімізації функції $f(x) = x^3/3 - \cos(x) - 1.5x^2\log(x) + 0.75x^2 - 1.5x$ на інтервалі $[0.2; 6]$

Приклад 2. Розглянемо глобальну мінімізацію полінома $f(x) = x^4 - 14x^3 + 61x^2 - 84x$ на інтервалі $[0, 8]$. Його похідна $f'(x)$ має нулі в точках $\{1; 3.5; 6\}$. При цьому, точки $\{1; 6\}$ є локальними мінімумами з однаковими значеннями цільової функції: -36. Запропонований у даній роботі алгоритм дозволив знайти точки глобального мінімуму $\{1; 6\}$ за рахунок п'яти рекурсивних ділень початкового інтервалу. Таким чином, у цьому прикладі запропонований метод дозволяє знайти усі точки глобального мінімуму.

Приклад 3. Розглянемо глобальну мінімізацію функції двох змінних з роботи [4]:

$$f(x_1, x_2) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4,$$

на прямокутнику $[-5; 5]\times[-5, 5]$. Обчислення локальних мінімумів зводиться до розв'язання системи рівнянь з двома змінними:

$$\begin{cases} 8x_1 - 8.4x_1^3 + 2x_1^5 + x_2 & = & 0, \\ x_1 - 8x_2 + 16x_2^3 & = & 0. \end{cases}$$

Застосування алгоритму [2] дозволяє знайти 15 коренів системи за рахунок 108 рекурсивних ділень початкового інтервалу. Серед цих точок є два локальних максимуми та шість локальних мінімумів: $\{(-1.61, -0.57),\ (1.61, -0.57),$ $(-1.7, 0.8),\ (1.7, -0.8),\ (0.09, -0.71),\ (-0.09, 0.71)\}$.

Проте застосування алгоритму, запропонованого в даній роботі дозволяє знайти дві точки глобального мінімуму $\{(0.09, -0.71), (-0.09, 0.71)\}$ за рахунок меншої кількості ділень (79) початкового прямокутника.

Якщо глобальний мінімум знаходиться на границі інтервалу D, робота алгоритму може бути прискорена, що показує наступний приклад.

Приклад 4. Розглянемо мінімізацію функції $g(x_1, x_2) = -f(x_1, x_2)$ на інтервалі $[-5, 5]\times[-5, 5]$, де $f(x_1, x_2)$ – функція з прикладу 3. За рахунок шести рекурсивних ділень сторін початкового прямокутника, ми знаходимо початкове наближення до точки мінімуму: $\{(-5, -5)\}$. На другому етапі алгоритму за рахунок лише одного розбиття початкового інтервалу, ми знаходимо, що функція $g(x_1, x_2)$ не містить локальних мінімумів із значеннями, меншими за $g(-5, -5) = -6420.8333$. Таким чином, точка $(-5, -5)$ – шукана точка глобального мінімуму функції $g(x_1, x_2)$.

Приклад 5. Розглянемо глобальну мінімізацію полінома від двох змінних з роботи [4]:

$$f(x_1, x_2) = (1 + (1 + x_1 + x_2)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)) \times$$
$$\times(30 + (2x_1 - 3x_2^2)(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2))$$

на прямокутнику $[-2, 2]\times[-2, 2]$.

Застосування алгоритму [2] дозволяє знайти дев'ять нулів градієнта цільової функції за рахунок 452 рекурсивних ділень початкового інтервалу. Серед цих точок є один локальний максимум та чотири локальні мінімуми: $\{(-0.60, -0.40),\ (0.00, -1.00),\ (0.80, 0.20),\ (1.84, 0.22)\}$.

Проте більш ефективним є застосування алгоритму, запропонованого в даній роботі. Це призводить до 293 ділень початкового прямокутника, що дозволяє визначити точку глобального мінімуму: $(0.00, -1.00)$.

Зауважимо, що застосування запропонованого методу є доцільним у тому випадку, якщо можуть бути обчислені головні мінори матриці Гессе та відповідні їм обмеження (5). Тема подальшої роботи – це конструювання більш ефективних критеріїв виявлення наявності локальних мінімумів у заданому інтервалі, таких як тест Кравчіка [5].

Висновок. У роботі запропоновано метод глобальної мінімізації функції $f(x_1,...,x_n)$ на інтервалі $D=[a_1;b_1]\times...\times[a_n;b_n]$. Метод засновано на розв'язанні системи нелінійних рівнянь, утвореної частковими похідними цільової функції. Головний принцип методу полягає у покритті початкового інтервалу підінтервалами, в яких часткові похідні цільової функції або головні мінори матриці Гессе мають постійний знак. Наведено алгоритм формування такого розбиття початкового інтервалу. Практичне застосування методу продемонстровано на декількох прикладах.

В.Ю. Семенов

МЕТОД ГЛОБАЛЬНОЙ МИНИМИЗАЦИИ ФУНКЦИЙ, ОСНОВАННЫЙ НА РЕШЕНИИ СИСТЕМ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

Предложен метод глобальной минимизации функций от нескольких переменных на заданном интервале. Метод основан на решении системы нелинейных уравнений, образованной частными производными целевой функции. Применение метода продемонстрировано на численных примерах.

V. Semenov

GLOBAL MINIMIZATION METHOD BASED ON THE SOLUTION OF SYSTEMS OF NONLINEAR EQUATIONS

The global minimization method for the functions of several variables on the given interval is proposed. The method is based on the solution of systems of nonlinear equations formed by partial derivatives of the objective function. The application of the method is illustrated on several examples.

Список літератури

1. Neumaier A. Complete Search in Continuous Global Optimization and Constraint Satisfaction, *Acta Numerica*. 2004. P. 383 – 408.
2. Семенов В.Ю. Метод нахождения всех действительных некратных корней системы нелинейных уравнений. *Журнал Вычислительной математики и математической физики*. 2007. № 9. С. 1486 – 1493.
3. Деннис Дж., Шнабель Р. Численные методы безусловной оптимизации и решения нелинейных уравнений. М.: Мир, 1988. 440 с.
4. Wang X., Chang T.-S. A multivariate global optimization using linear bounding functions. *J. Global Optim.* 1998, Vol. 12. P. 383 – 404.
5. *Семенов В.Ю.* Метод нахождения всех корней системы нелинейных алгебраических уравнений, основанный на операторе Кравчика. *Кибернетика и системный анализ.* 2015. № 5. С. 169 – 175.

Одержано 12.05.2018

Про автора:

Семенов Василь Юрійович,
кандидат фізико-математичних наук,
завідувач науково-дослідного відділу алгоритмів ТОВ “Дельта СПЕ”.
E-mail: vasyl.delta@gmail.com