КОМП'ЮТЕРНІ НАУКИ
ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

УДК 004.021:004.67:519.7
DOI https://doi.org/10.32851/tnv-tech.2021.3.1

ЗНАХОДЖЕННЯ ОПТИМАЛЬНИХ ЗНАЧЕНЬ ФУНКЦІЙ ІЗ ЗАСТОСУВАННЯМ МЕТОДУ СПРЯЖЕНИХ ГРАДІЄНТІВ

Димова Г.О. *– кандидат технічних наук,*
доцент кафедри менеджменту та інформаційних технологій
Херсонського державного аграрно-економічного університету
ORCID ID: 0000-0002-5294-1756

У статті для розв'язання некоректно поставлених задач методом регуляризації розглядається метод спряжених градієнтів. Градієнтом називається вектор, величина якого визначає швидкість змінення функції, а напрямок збігається з напрямком найбільшого зростання цієї функції. Вектор, що вказує напрямок найбільшого зменшення функції, називається антиградієнтом функції. Метод спряжених градієнтів застосовується для розв'язання задач безумовної мінімізації, для відшукання екстремалі згладжуючого функціоналу. Цей метод є ітераційним методом. Загальною властивістю більшості ітераційних алгоритмів є швидке спадання швидкості мінімізації в разі наближення до точки мінімуму функціонала. Тому важливою характеристикою ітераційних алгоритмів є той фактичний мінімальний рівень значень функціонала нев'язки, до якого вдається довести процес мінімізації за реальний час.

У роботі описаний метод найскорішого спуску як метод, що передує методу спряжених градієнтів і поєднує в собі два поняття: «градієнт цільової функції» та «сполучений напрямок векторів». Також приведений метод сполучених напрямків та два методи пошуку вагового коефіцієнта.

У статті аналізуються градієнтні методи пошуку оптимальних значень квадратичних функцій та функцій загального виду. Метод спряжених градієнтів є методом першого порядку, але швидкість його збіжності квадратична, чим цей метод вигідно відрізняється від звичайних градієнтних методів. Недоліком градієнтного пошуку є те, що під час його використання можна виявити тільки локальний екстремум функції. Для відшукання інших локальних екстремумів необхідно проводити пошук з інших початкових точок. Побудований алгоритм мінімізації функціонала за допомогою методу спряжених градієнтів.

Ключові слова: *мінімізація, найскоріший спуск, сполучені напрямки, спряжені градієнти, ітераційний алгоритм.*

Dymova H.O. Finding the optimal values of functions using the conjugate gradient method

The article considers the conjugate gradient method to solve ill-posed problems by the regularization method. A gradient is a vector, the value of which determines the rate of change of a function, and the direction coincides with the direction of the growth of this function itself. The vector indicating the direction of the greatest decrease in the function is called the antigradient of the function. The conjugate gradient method is used to solve unconstrained minimization problems, to find the extremal of the smoothing functional. This method is an iterative method.

A common property of most iterative algorithms is a rapid drop in the minimization rate when approaching the minimum point of the functional. Therefore, an important characteristic of iterative algorithms is the actual minimum level of the residual functional values, to which it is possible to bring the real-time minimization process.

The paper describes the steepest descent method as a method preceding the conjugate gradient method and combines two concepts: the objective function gradient and the conjugate direction of vectors. The method of conjugate directions and two methods of finding the weight coefficient are also given.

The article analyzes gradient methods for finding the optimal values of quadratic functions and functions of a general form. The conjugate gradient method is a first order method, but its rate of convergence is quadratic, which makes this method comparing favorably with conventional gradient methods. The disadvantage of gradient search is that only the local extremum of the function can be found when using it. To find other local extrema, it is necessary to search from other starting points. An algorithm for minimizing the functional using the conjugate gradient method is constructed.

Key words: *minimization, fastest descent, conjugate directions, conjugate gradients, iterative algorithm.*

Метод спряжених градієнтів застосовується для розв’язання задач безумовної мінімізації. Розглядається цей метод для відшукання екстремалі згладжуючого функціоналу під час розв’язання некоректно поставлених задач методом регуляризації. Метод спряжених градієнтів являється ітераційним методом. Загальною властивістю більшості ітераційних алгоритмів є швидке спадання швидкості мінімізації в разі наближення до точки мінімуму функціонала. Тому важною характеристикою ітераційних алгоритмів являється той фактичний мінімальний рівень значень функціонала нев’язки, до якого вдається довести процес мінімізації за реальний час [1].

Метод спряжених градієнтів має тривалу історію розвитку, початковий етап якого був детально описаний Голубом та О’Лірі в оглядовій статті [2]. Перша робота в даній області належить Хестенсу і Штіфель і датується 1952 роком [3]. У цій роботі метод спряжених градієнтів застосовувався для розв’язання систем лінійних рівнянь. У 1964-му році Флетчер і Рівз [4] вперше використовували метод спряжених градієнтів для мінімізації гладких функцій.

Розвиток методу триває і зараз: розробляються нові варіанти методу, досліджуються питання збіжності. Перевагами методу спряжених градієнтів є його простота й низькі витрати пам’яті, що робить його особливо ефективним під час розв’язання задач великої розмірності.

Метою роботи є дослідження методу спряжених градієнтів для розв’язання некоректно поставлених задач методом регуляризації, а також побудова алгоритму мінімізації функціонала за допомогою цього методу.

Для розгляду методу спряжених градієнтів уведемо деякі позначення, що будуть використовуватися. Скалярний добуток двох векторів $\mathbf{x}^T\mathbf{y}$ представляє суму скалярів $\sum_{i=0}^{n}\mathbf{x}_i\mathbf{y}_i$, причому $\mathbf{x}^T\mathbf{y}=\mathbf{y}^T\mathbf{x}$. Якщо $\mathbf{x}$ та $\mathbf{y}$ ортогональні, то $\mathbf{x}^T\mathbf{y}=0$. Вирази, що перетворюються в матриці 1×1, такі як $\mathbf{x}^T\mathbf{y}$ та $\mathbf{x}^T\mathbf{A}\mathbf{x}$, розглядаються як скалярні величини.

Спочатку метод спряжених градієнтів був розроблений для розв’язання систем лінійних алгебраїчних рівнянь виду [5]:

$$\begin{aligned}&a_{0,0}x_0+a_{0,1}x_1+\ldots+a_{0,n-1}x_{n-1}=b_0\\&a_{1,0}x_0+a_{1,1}x_1+\ldots+a_{1,n-1}x_{n-1}=b_1\\&\ldots\\&a_{n-1,0}x_0+a_{n-1,1}x_1+\ldots+a_{n-1,n-1}x_{n-1}=b_{n-1}\end{aligned}\tag{1}$$

У матричній формі (1) виглядає як:

$$\mathbf{Ax} = \mathbf{b}\ , \tag{2}$$

де $\mathbf{x}$ – невідомий вектор;
$\mathbf{b}$ – відомий вектор;
$\mathbf{A}$ – відома квадратна симетрична позитивно визначена матриця.

Розв'язання цієї системи є еквівалентним знаходженню мінімуму відповідної квадратичної форми:

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x} + \mathbf{c}\ . \tag{3}$$

Наявність такого зв'язку між матрицею лінійного перетворення **A** і скалярною функцією $f(\mathbf{x})$ дає можливість продемонструвати деякі функції лінійної алгебри рисунками (рис. 1).

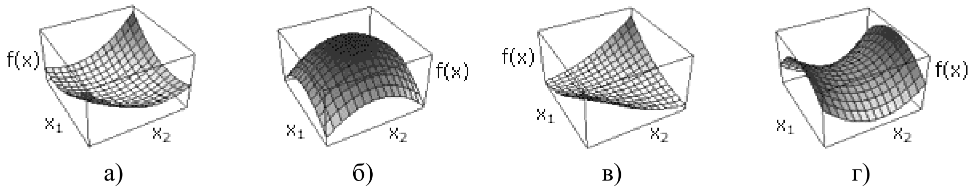

Рис. 1. Квадратичні форми для позитивно визначеної матриці (а), негативно визначеної матриці (б), позитивно невизначеної матриці (в), невизначеної матриці (г) [6]

Матриця **A** є позитивно визначеною, якщо для будь-якого ненульового вектора **x** справедливий вираз:

$$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0\ . \tag{4}$$

Для знаходження позитивно визначеної матриці **A** необхідно знайти мінімум її квадратичної функції. Причому за допомогою методу спряжених градієнтів мінімум квадратичної функції можна знайти за n чи менше кроків, де n – розмірність невідомого вектора **x**. Виходячи з того, що будь-яка гладка функція в околицях точки свого мінімуму гарно апроксимується квадратичною функцією, цей же метод можна використати для мінімізації і неквадратичних функції. При цьому метод перестає бути кінцевим, а стає ітеративним [6].

Для початку розглянемо метод найскорішого спуску як найпростіший метод пошуку екстремуму функції. Представимо алгоритм цього методу:

Крок 1. У початковій точці $x(0)$ обчислюється градієнт. Рух здійснюється в напрямку антиградієнта до тих пір, доки зменшується цільова функція.

Крок 2. У точці, де функція перестає зменшуватися, знову обчислюється градієнт, і спуск продовжується в новому напрямку.

Крок 3. Процес повторюється до тих пір, доки точка не досягне мінімуму.

На рисунку 2 показана траєкторія руху в точку мінімуму методом найскорішого спуску.

У випадку методу найскорішого спуску кожний новий напрямок руху ортогональний попередньому.

Існує іншій спосіб вибору нового напрямку руху – метод сполучених напрямків, до якого відноситься метод спряжених градієнтів. Метод спряжених градієнтів є подальшим розвитком методу найшвидшого спуску, який поєднує в собі два поняття: градієнт цільової функції і сполучений напрямок векторів. На рисунку 3 зображені траєкторії руху в точку мінімуму в разі використання методу найскорішого спуску (1) та методу спряжених градієнтів (2).

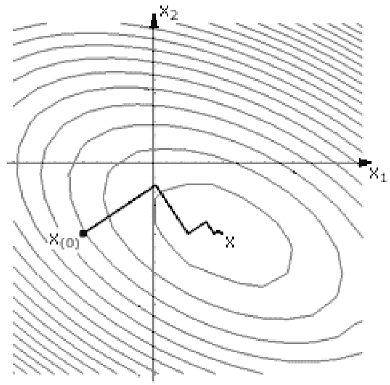

Рис. 2. Траєкторія руху в точку мінімуму методом найскорішого спуску [6]

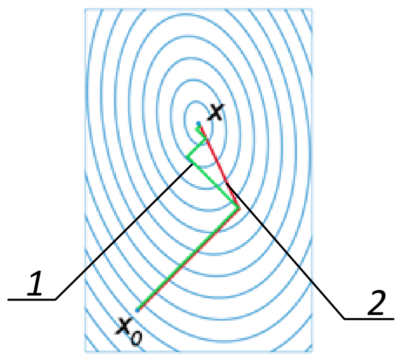

Рис. 3. Траєкторія руху в точку мінімуму [7]:
1 – методом найскорішого спуску; 2 – методом спряжених градієнтів

Спряженість – це узагальнене поняття ортогональності. Коли матриця **A** – одинична матриця, у відповідності з рівнянням (5) вектори **x** та **y** являються ортогональними. Процес ортогоналізації Грамма-Шмідта є одним із можливих способів обчислення сполучених напрямків з використанням методів лінійної алгебри. Однак цей спосіб не застосовується для більшості задач, тому що для його використання необхідно знати матрицю **A**. Існують інші ітеративні способи обчислення сполучених напрямків. Процес знаходження мінімуму функції є ітераційною процедурою, яка записується у векторній формі таким чином:

$$\mathbf{d}_{(i+1)} = \mathbf{d}_{(i+1)} + \beta_{(i+1)}\mathbf{d}_i\,, \qquad (6)$$

де $\beta_{(i+1)}$ – ваговий коефіцієнт, який використовується для визначення сполученого напрямку.

Тобто новий сполучений напрямок (6) отримуємо складанням антиградієнта в точці повороту і попереднього напрямку руху, помноженого на ваговий коефіцієнт $\beta_{(i+1)}$ (7). Напрямки, які отримані з (6), виявляються сполученими, якщо функція, що мінімізується, задана у формі (3), тобто для квадратичних функцій метод спряжених градієнтів відшукує мінімум за n кроків, де n – це розмірність простору пошуку.

Ваговий коефіцієнт можна визначати декількома методами. Одним із таких методів є визначення за формулою Флетчера-Рівса (Fletcher-Reeves) [8]:

$$\beta_{(i+1)} = \frac{\mathbf{r}_{(i+1)}^T \mathbf{r}_{(i+1)}}{\mathbf{r}_i^T \mathbf{r}_i}. \tag{7}$$

Для функцій загального виду алгоритм перестає бути кінцевим і стає ітеративним, при цьому метод Флетчера-Рівса передбачає оновлення алгоритмічної процедури кожні n+1 кроки.

Другим методом розглядають визначення вагового коефіцієнта за формулою Полака-Райбера (Polak-Ribiere):

$$\beta_{(i+1)} = \frac{\mathbf{r}_{(i+1)}^T (\mathbf{r}_{(i+1)} - \mathbf{r}_i)}{\mathbf{r}_i^T \mathbf{r}_i}. \tag{8}$$

Метод Флетчера-Рівса збігається, якщо початкова точка достатньо близька до необхідного мінімуму, тоді як метод Полака-Райбера може в рідких випадках зациклюватися нескінченно, але останній метод часто збігається швидше першого методу. Збіжність методу Полака-Райбера може бути гарантована вибором $\beta = \max\{\beta; 0\}$. Це еквівалентно перезапуску алгоритму за умовою $\beta \le 0$ [6]. Перезапуск алгоритмічної процедури необхідний для очищення останнього напрямку пошуку і старту алгоритму спочатку в напрямку найскорішого спуску.

Приведемо алгоритм методу спряжених градієнтів для мінімізації функцій загального виду.

Крок 1. Обчислюється антиградієнт у довільній точці x_0: $\mathbf{d}_0 = \mathbf{r}_0 = -f'(x_0)$.

Крок 2. Здійснюється спуск в обчисленому напрямку, доки функція зменшується, тобто пошук a_i, який мінімізує $f(x_i + a_i \mathbf{d}_i)$.

Крок 3. Перехід у точку, що знайдена в попередньому пункті $x_{(i+1)} = x_i + a_i \mathbf{d}_i$.

Крок 4. Обчислення антиградієнту в новій точці $\mathbf{r}_{(i+1)} = -f'(x_{(i+1)})$.

Крок 5. Обчислення вагового коефіцієнту $\beta_{(i+1)}$ для перезапуску алгоритму (для методу Флетчера-Рівса привласнити 0 через кожні n–1 кроків, для методу Полака-Райбера – $\beta_{(i+1)} = \max\{\beta_{(i+1)}; 0\}$.

Крок 6. Обчислення нового сполученого напрямку $\mathbf{d}_{(i+1)} = \mathbf{d}_{(i+1)} + \beta_{(i+1)} \mathbf{d}_i$.

Крок 7. Перехід на крок 2.

На кроці 2 алгоритму здійснюється одномірна мінімізація функції. Для цього використовують метод Фібоначчі, метод золотого перерізу або метод бісекцій. Найбільш швидку збіжність забезпечує метод Ньютона-Рафсона, але для цього необхідно мати можливість обчислення матриці Гессе [6]. Змінна, по якій здійснюється оптимізація, обчислюється на кожному кроці ітерації за формулою:

$$a = \frac{f'^T(\mathbf{x})\mathbf{d}}{\mathbf{d}^T f''(\mathbf{x})\mathbf{d}},$$

де $f''(\mathbf{x})$ – матриця Гессе, що має вигляд:

$$f''(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}.$$

Істотним недоліком методу Ньютона-Рафсона є залежність збіжності для невипуклих функцій від початкового наближення x_0. Якщо x_0 знаходиться досить далеко від точки мінімуму, то метод може розбігатися, тобто в разі проведення ітерації кожна наступна точка буде більш віддаленою від точки мінімуму, чим попередня.

Метод спряжених градієнтів є методом першого порядку, але швидкість його збіжності квадратична, чим цей метод вигідно відрізняється від звичайних градієнтних методів. Метод градієнта разом з його численними модифікаціями є розповсюдженим та ефективним методом пошуку оптимуму досліджених об'єктів. Недоліком градієнтного пошуку є те, що в разі його використання можна виявити тільки локальний екстремум функції. Для відшукання інших локальних екстремумів необхідно проводити пошук з інших початкових точок. Для квадратичної функції методи найскорішого та координатного спусків збігаються лише в межі. Метод спряжених градієнтів оптимізує квадратичну функцію за кінцеве число ітерацій. Метод сполучених напрямків за оптимізації функцій загального виду збігається набагато скоріше методу найскорішого спуску, при цьому не потребує складних обчислень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ:

1. Регуляризирующие алгоритмы и априорная информация / А.Н. Тихонов и др. Москва : Наука, 1983. 200 с.
2. Golub G.H., O'Leary D.P. Some History of the Conjugate Gradient Methods and Lanczos Algorithms: 1948–1976. *SIAM Rev.* 1989. Vol. 31. P. 50–102.
3. Hestenes M.R., Stiefel E. Method of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*. 1952. Vol. 49. P. 409–436.
4. Fletcher R., Reeves C.M. Function Minimization by Conjugate Gradients. *Computer Journal*. 1964. Vol. 7. P. 149–154.
5. Решение СЛУ методом сопряженных градиентов. 2012. URL: http://www.hpcc.unn.ru/?dir=847 (дата звернення: 20.05.21).
6. Некипелов Н. Метод сопряженных градиентов – математический аппарат. URL: https://basegroup.ru/community/articles/conjugate (дата звернення: 18.05.21).
7. Метод спряженого градієнта. 2021. URL: https://uk.wikipedia.org/wiki/Метод_спряженого_градієнта#Місцево_оптимальний_метод_найбільш_стрімкого_спуску (дата звернення: 18.05.21).
8. Безусловная оптимизация. Метод сопряженных градиентов. 2017. URL: https://simenergy.ru/math-analysis/solution-methods/90-conjugate-gradient (дата звернення: 11.08.21).

REFERENCES:

1. Tikhonov A.N. Goncharovskiy A.V., Stepanov V.V., Yagola A.G. (1983). Regulyariziruyushchiye algoritmy i apriornaya informatsiya [Regularizing algorithms and a priori information]. M.: Nauka, 200 [in Russian].

2. Golub G. H., O’Leary D. P. (1989). Some History of the Conjugate Gradient Methods and Lanczos Algorithms: 1948–1976. *SIAM* Rev. Vol. 31. 50–102 [in English].

3. Hestenes M. R., Stiefel E. (1952). Method of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*. Vol. 49. 409–436 [in English].

4. Fletcher R., Reeves C. M. (1964). Function Minimization by Conjugate Gradients. *Computer Journal*. Vol. 7. P. 149–154 [in English].

5. Resheniye SLU metodom sopryazhennykh gradiyentov [Solution SLE conjugate gradient method]. (2012). URL: http://www.hpcc.unn.ru/?dir=847 (date of the application 20.05.21) [in Russian].

6. Nekipelov N. Metod sopryazhennykh gradiyentov – matematicheskiy apparat [Method of conjugate gradients – mathematical apparatus]. URL: https://basegroup.ru/community/articles/conjugate (date of the application 18.05.21) [in Russian].

7. Metod spryazhenoho hradiyenta [Conjugate gradient method]. (2021). URL: https://uk.wikipedia.org/wiki/Метод_спряженого_градієнта#Місцево_оптимальний_метод_найбільш_стрімкого_спуску (date of the application 18.05.21) [in Ukranian].

8. Bezuslovnaya optimizatsiya. Metod sopryazhennykh gradiyentov [Unconditional optimization. Conjugate gradient method]. (2017). URL: https://simenergy.ru/math-analysis/solution-methods/90-conjugate-gradient (date of the application 11.08.21) [in Russian].