УДК 004.942:517.958:519.6

МОДИФІКОВАНІ МЕТОДИ СПРЯЖЕНИХ НАПРЯМКІВ ЧИСЕЛЬНОЇ МІНІМІЗАЦІЇ ФУНКЦІЙ БАГАТЬОХ ЗМІННИХ ДЛЯ ЗАДАЧ АНАЛІЗУ ДАНИХ

Ю.В. Нікольський
Кандидат фізико-математичних наук, доцент
Кафедра "Інформаційні системи та мережі»
Інститут комп'ютерних наук та інформаційних технологій
Національного університету «Львівська політехніка»
вул. С. Бандери, 12, м. Львів, 79013
Контактний тел.: (0322) 37-04-40
E-mail: y_nikol@yahoo.com

Побудовано модифікації методів спряжених напрямків для мінімізації функцій багатьох змінних, які виникають в задачах аналізу даних. Введення параметрів та використання додаткової інформації про властивості функції дозволяє забезпечити умови збіжності та її швидкість

Ключові слова: мінімізація, метод спряжених напрямків

Построены модификации методов сопряженных направлений для минимизации функций многих переменных, которые возникают в задачах анализа данных. Введение параметров и использование дополнительной информации о свойствах функции позволяет обеспечить условия сходимости и ее скорость

Ключевые слова: минимизация, метод сопряженных направлений

Some modifications of conjugate directions methods for functions minimization are constructed. The introduction of parameters and additional information about the properties of the function allows ensuring the conditions for convergence and speed

Keywords: minimization method, conjugate directions method

1. Вступ

Застосування методів машинного навчання у процесах аналізу даних [1] викликає необхідність побудови обчислювальних процедур, пов'язаних із розв'язування задачі мінімізації функцій багатьох змінних. Такі задачі є актуальними, зокрема, у зв'язку із широким застосуванням нейронних мереж прямого поширення сигналу, навчання яких здійснюють методом зворотного поширення похибки. Цей метод в своїй основі має градієнтний метод, який володіє відомими недоліками, пов'язаними із специфічним розташуванням локальних екстремумів та особливостями формування траєкторії спуску. Застосування методів спряжених градієнтів дозволяє уникати цих проблем, хоча й призводить до збільшення кількості обчислень на кожному кроці. Одним з основних недоліків методів спряжених напрямків і методу спряжених градієнтів, зокрема, є те, що вони чутливі до властивостей гесіана функції, що мінімізується. Це вимагає забезпечувати умови збіжності, які визначені такими властивостями. Тому шлях вирішення цієї проблеми полягає у введенні в обчислювальні формули методів спряжених напрямків спеціальних обчислювальних процедур, які дозволять коректувати властивості вектора спуску в залежності від локальних властивостей функції без порушення принципів формування методу. Саме такий підхід пропонується у цій статті.

2. Постановка задачі дослідження

В останні роки багато зусиль у застосуванні методів мінімізації у задачах навчання нейромереж прямого поширення сигналу прикладено до адаптації відомих методів мінімізації та нових методів мінімізації, які основані на ідеї генетичного алгоритму. Крім того, в роботах із застосування нейронних мереж широко застосовують метод Левенберга – Маркуарда. Цей метод вимагає розв'язування допоміжної нетривіальної задачі знаходження розв'язку системи нелінійних рівнянь, яка має таку ж складність, що й метод Ньютона мінімізації функцій багатьох змінних. Застосування генетичного алгоритму не набуло широкого застосування, оскільки цей алгоритм володіє низькою швидкістю збіжності та вимагає адаптації кроків мутації та схрещування до конкретної функції. У цій статті запропоновано та досліджено новий клас методів спряжених напрямків. Методи цього класу дозволяють будувати обчислювальні процедури мінімізації, які можна регулювати за рахунок збільшення інформації про поточну точку траєкторії спуску та локальні властивості функції, а також вводити спеціальні параметри, з допомогою яких забезпечити умови збіжності та її швидкість. Крім того, за рахунок вибору параметрів можна побудувати як усі відомі методи цього класу, так і отримати обчислювальні формули нових методів.

3. Аналіз останніх досліджень

Для розв'язування операторних рівнянь був розроблений метод Ньютона, який для задач мінімізації функцій багатьох змінних записують такою рекурентною формулою

$$x_{k+1} = x_k - \left(f''(x_k)\right)^{-1} f'(x_k) .$$

Метод Ньютона має високу швидкість збіжності, проте йому властиві й очевидні недоліки, які або взагалі не дозволяють використовувати його для вирішення багатьох завдань, або роблять його ефективність низькою.

Алгоритми, які перевершують метод Ньютона за ефективністю, володіють такими властивостями: по-перше, послідовність $\{x_k\}$ збігається для довільного x_0, а, по-друге, трудомісткість кожної їх ітерації нижча за трудомісткість ітерації методу Ньютона. В першу чергу, вона позбавлена необхідності обчислення та обертання матриці Гессе.

Для розширення області збіжності ще Л. В. Канторович запропонував регулювати довжину кроку шляхом введення множника α_k за правилом

$$x_{k+1} = x_k - \alpha_k \left(f''(x_k)\right)^{-1} f'(x_k). \qquad (1)$$

Тут скалярний множник α_k вибирають так, щоб гарантувати спадання функції на кожній ітерації. Вперше теоретичні результати про збіжність процесу (1) отримав М. Н. Яковлев [2], який вивчав метод

$$x_{k+1} = x_k + \alpha_k p_k ,\ p_k = -A_k^{-1} f'(x_k), \qquad (2)$$

де $A_k \to f''(x_k)$ при $k \to \infty$. У цій праці було доведено, що процес вигляду , (2) при виборі α_k з умови

$$f(x_k + \alpha_k p_k) = \min_{\alpha} f(x_k + \alpha p_k) \qquad (3)$$

збігається до точки мінімуму, а для опуклої гладкої функції – з надлінійною швидкістю. Проте спосіб (3) вибору α_k не вважають найвдалішим, оскільки у такому разі потрібно розв'язувати на кожній ітерації одновимірну задачу мінімізації, що є теоретично нескінченною процедурою. При цьому, якщо обчислення $f(x_k)$ виявляється трудомістким, то вибір α_k з умови (3) збільшує трудомісткість усього процесу (2)-(3).

Результати робіт В. Девідона [3], Р. Флетчера і М. Пауелла [4] з'явилися основою розвитку нового класу методів, квазіньютонівських методів із змінною метрикою, в основу яких покладена ідея побудови системи спряжених напрямків. Ці методи, в загальному випадку, мають вигляд

$$x_{k+1} = x_k + \alpha_k p_k ,\ p_k = -H_k^T f'(x_k) , \qquad (4)$$

де матрицю H_k будують за рекурентними формулами, а параметр α_k визначають з умови (3). Низкою авторів побудовано варіанти методів спряжених напрямів. Серед них можна відмітити праці G. McCormic та J. Pearson [5], J. Pearson [6].

Дослідження властивостей, умов і швидкості збіжності даного класу методів проводилося низкою авторів. В результаті досліджень A. Cohen [7], Б. Поляка [8], Г. Майстровського [9], С. Смоляка [10] доведено, що швидкість збіжності методів спряжених градієнтів оцінюється нерівністю

$$\left\| x_{k+n} - x_* \right\| \le C \left\| x_k - x_* \right\|^2 ,\ k = 0, 1, 2 ....$$

У праці H. Huang [11] запропонована загальна методика дослідження властивостей методів типу (4) у разі мінімізації квадратичної функції із вибором множника α_k з умови (3). Ця методика дозволила зобразити всі існуючі алгоритми побудови матриці H_k частковими випадками формули, яку він запропонував. Також доведено, що всі методи спряжених напрямків у разі мінімізації квадратичної функції володіють властивістю збігатись до точки мінімуму за кількість ітерацій, яка не перевершує кількості змінних цієї функції. Послідовності, які побудовано за формулами (4), у цьому випадку співпадають для будь-якого методу спряжених напрямків.

Дослідження збіжності та отримання оцінок швидкості збіжності методів спряжених напрямків проводилося низкою авторів. Серед отриманих результатів, можна відзначити доведену M. Powell [12] та E. Polak [13] надлінійної швидкість збіжності послідовності $\{x_k\}$, $k = 0,1,...$ та квадратичну збіжність послідовності $\{x_{\xi n+1}\}$, $\xi = 0,1,...$, побудованої за методом Davidon-Fletcher-Powell [4]. Також встановлена надлінійна швидкість збіжності методів спряжених напрямків. Останні результати подано у працях Ю. Даніліна [14], а також Б. Пшенічного та Ю. Даніліна [15]. Дослідження та уточнення оцінок Ю. Даніліним [16] дозволили встановити, що швидкість збіжності методів спряжених напрямків не нижча від квадратичної, причому вона залежить від способу побудови матриці H_k та може бути отримана надквадратичною, коли послідовність $\{H_k\}$ в (4) володіє такою властивістю $H_k \to \left(f''(x_k)\right)^{-1}$.

4. Цілі статті

У пропонованій статті побудовано та досліджено модифікації методів спряжених напрямків мінімізації функцій багатьох змінних. Запропоновані формули використовують обчислення лише перших похідних функцій. Вибором параметрів методів можна отримати усі відомі обчислювальні формули методів цього класу, а також забезпечити виконання умов збіжності та високі оцінки її швидкості. Основна відмінність пропонованих модифікацій методів полягає у використанні додаткової інформації для побудови точок траєкторії спуску. Це дозволяє ввести в обчислювальні формули групи параметрів, з допомогою яких можна регулювати напрямок спуску із врахуванням локальних особливостей функцій та забезпечує високу швидкість збіжності до точки мінімуму.

5. Основний матеріал

Спочатку розглянемо принцип побудови модифікацій методів спряжених напрямків для мінімізації квадратичної функції

$$f(x)=0.5(Ax,x)+(b,x)+C\,, \tag{5}$$

де $(Ax,x)>0$ для довільного $x\neq 0, x\in E^n$. Застосуємо метод, послідовність $\{x_k\}$ якого побудовано за формулою

$$x_{k+1}=x_k+\alpha_k p_k\,,\ p_k=-H_k^T f'(x_k)\,,\ \ k=0,1,2,\ldots\,. \tag{6}$$

Відомо, що послідовність векторів спуску p_k, $1\leq k\leq n-1$ повинна задовольняти вимоги

$$(p_k, Ap_j)=0\,,\ 0\leq j\leq k-1 \tag{7}$$

при виборі $p_k=H_k^T f_k$, де H_k – певна квадратна матриця розмірів $n\times n$. Умова (7) приймає вигляд $(f'(x_k), H_k Ap)=0$, $0\leq j\leq k-1$.

Нехай вибір параметра α_k в (6) здійснюється із співвідношення

$$f(x_k+\alpha_k p_k)=\min_{\alpha} f(x_k+\alpha p_k)\,. \tag{8}$$

Такий спосіб вибору a_k приводить до виразу

$$\alpha_k=\frac{(f'(x_k),p_k)}{(Ap_k,p_k)}\,,\ \ k=0,1,\ldots\,, \tag{9}$$

а це, у свою чергу, до

$$(f'(x_k),p_j)=0\,,\ \ 0\leq j\leq k-1\,. \tag{10}$$

З порівняння (9) та (10) можна отримати

$$H_k Ap_j=ap_j\,,\ \ 0\leq j\leq k-1\,, \tag{11}$$

де a – довільна стала, звідки, в припущенні $(f'(x_i),p_i)=0$ і сильної опуклості функції $f(x)$, випливає обмеженість параметра α_k, $0<|\alpha_k|<\infty$, а умови (7) та (10) зобразити у вигляді $(r_k,e_j)=0$, $0\leq j\leq k-1$, $(f'(x_k),r_j)=0$, $0\leq j\leq k-1$.

У цьому випадку (11) можна записати так

$$H_k e_j=ar_j\,,\ \ 0\leq j\leq k-1\,, \tag{12}$$

де $r_j=x_{j+1}-x_j$, $e_j=f'(x_j+r_j)-f'(x_j)=\alpha_j Ap_j$.

Отже, виконання вимоги (7) зводиться до перевірки співвідношення (12)(12). Побудуємо послідовність матриць $\{H_k\}$ з рекурентних співвідношень, де (12) записано так

$$(H_{k-1}+\Delta H_{k-i})e_j=ar_j\,,\ \ 0\leq j\leq k-1\,.$$

Через виконання останньої рівності, будуть правильними такі рівності

$$H_{k-1}e_j=ar_j\,,\ \ 0\leq j\leq k-2\,,$$

$$\Delta H_{k-1}e_j=0\,,$$

$$\Delta H_{k-1}e_{k-1}=ar_{k-1}-H_{k-1}e_{k-1} \tag{13}$$

для $0\leq j\leq k-2$.

Остання рівність виконується, якщо

$$\begin{aligned}\Delta H_{k-1}={}&\frac{\alpha_1\left(a_{11}^{(k-1)}r_{k-1}H_{k-1}^T-a_{12}^{(k-1)}r_{k-1}\tilde{H}_{k-1}^T\right)}{\Delta_1^{(k-1)}}-\\&-\frac{\alpha_2\left(a_{21}^{(k-1)}\tilde{r}_{k-1}H_{k-1}^T-a_{22}^{(k-1)}\tilde{r}_{k-1}\tilde{H}_{k-1}^T\right)}{\Delta_1^{(k-1)}}-\\&-H_{k-1}\frac{\left(b_{11}^{(k-1)}e_{k-1}-\beta\ \ b_{21}^{(k-1)}\tilde{e}_{k-1}\right)\ v_{k-1}^T-\left(b_{12}^{(k-1)}e_{k-1}-\beta b_{22}^{(k-1)}\tilde{e}_{k-1}\right)\ \tilde{v}_{k-1}^T}{\Delta_2^{(k-1)}},\\ \Delta_1^{(k-1)}={}&a_{11}^{(k-1)}a_{22}^{(k-1)}-a_{12}^{(k-1)}a_{21}^{(k-1)},\\ \Delta_2^{(k-1)}={}&b_{11}^{(k-1)}b_{22}^{(k-1)}-b_{12}^{(k-1)}b_{21}^{(k-1)}.\end{aligned} \tag{14}$$

$$\begin{aligned}&a_{11}^{(k-1)}=(\tilde{u}_{k-1},\tilde{e}_{k-1}), &&b_{11}^{(k-1)}=(\tilde{v}_{k-1},\tilde{e}_{k-1}),\\ &a_{12}^{(k-1)}=(u_{k-1},\tilde{e}_{k-1}), &&b_{12}^{(k-1)}=(v_{k-1},\tilde{e}_{k-1}),\\ &a_{21}^{(k-1)}=(\tilde{u}_{k-1},e_{k-1}), &&b_{21}^{(k-1)}=(\tilde{v}_{k-1},e_{k-1}),\\ &a_{22}^{(k-1)}=(u_{k-1},e_{k-1}), &&b_{22}^{(k-1)}=(v_{k-1},e_{k-1}),\end{aligned} \tag{15}$$

причому

$$\begin{aligned}&u_{k-1}=t_{11}r_{k-1}+t_{12}H_{k-1}^T e_{k-1}, &&\tilde{u}_{k-1}=t_{21}r_{k-1}+t_{22}H_{k-1}^T e_{k-1},\\ &v_{k-1}=t_{31}r_{k-1}+t_{32}H_{k-1}^T e_{k-1}, &&\tilde{v}_{k-1}=t_{41}r_{k-1}+t_{42}H_{k-1}^T e_{k-1}.\end{aligned} \tag{16}$$

Тут $\alpha_1, \alpha_2, \beta, t_{11}, t_{12}, t_{21}, t_{22}, t_{31}, t_{32}, t_{41}, t_{42}$ – скалярні множники, вибір яких дозволяє побудувати відповідний метод спряжених напрямків. У співвідношеннях (13)(13)-(16) позначено

$$e_{k-1}=f'(x_k)-f'(x_{k-1})\,,\ \ r_{k-1}=x_k-x_{k-1}\,.$$

Розглянемо алгоритм, який можна отримати з формул (14)-(16).

Тепер нехай $f(x)$ – строго опукла й двічі неперервно диференційована в усіх точках множини $S=\{x\,|\,f(x)\leq c\}$ функція, причому для матриці $f''(x)$ виконані умови

$$m\|z\|^2\leq (f''(x)z,z)\leq M\|z\|^2\,, \tag{17}$$

$$\|f''(x)-f''(y)\|\leq R\,\|x-y\|,\ R>0\,. \tag{18}$$

Послідовні наближення побудуємо за формулою (6), а вибір скалярного множника проводиться з умови (8).

Розглянемо алгоритм, у якому матрицю H_k обчислено за формулою

$$H_{k+1}=H_k-H_k\frac{t_k a_1 e_k r_k^T-a_3\tilde{e}_k r_k^T-a_2 e_k e_k^T H_k+a_4\tilde{e}_k e_k^T H_k}{t_k a_1 a_4-a_2 a_3}, \tag{19}$$

де $a_1=\left(H_k^T e_k,\tilde{e}_k\right)$, $a_3=\left(H_k^T e_k,e_k\right)$,

$a_2=\left(r_k,\tilde{e}_k\right)$, $a_4=\left(r_k,e_k\right)$.

Алгоритми можна виконувати або без відновлення H_k матриці або з її відновленням, наприклад, через d ітерацій $(d>1)$. В останньому випадку, вважаємо $H_{\xi d}=H$, $\xi=0,1,\ldots$ Тут H – певна матриця, для якої виконано умови

$$m_1\|z\|^2\le(Hz,z)\le M_1\|z\|^2,\quad \forall z\in E^n,\quad 0<m_1<M_1. \qquad (20)$$

Лема 1. Нехай $f(x)$ функція, що двічі неперервно диференціюється, для якої виконані умови (17) та (18). Якщо послідовність $\{x_k\}$ побудована за формулами (8), (19), для H_0 виконана умова (20), причому через n ітерацій проводиться відновлення матриці H, то при виборі параметра t_i із співвідношень $t_i\ge\left((m_0-\varepsilon)^{-1}\|r_0\|^4+a_2a_3\right)\cdot(a_1a_4)^{-1}$ виконуються оцінки

$$\left|(r_i,e_j)\right|\le C\|r_j\|^2\|r_i\|,\ \|r_i\|\le C\|f'(x_i)\|,\ 0\le j<i<k-1.$$

Лема 2. Нехай для мінімізації двічі функції, що неперервно диференціюється $f(x)$ застосовано метод (19), а матриця других похідних $f''(x)$ задовольняє умови (17)-(18). Тоді для достатньо великих значень $\|\vartheta_1\|\ge C_2\|r_{n-1}\|$ виконуються $\left\|H_n^{-1}-f''(x_n)\right\|\to 0$ та $\left\|H^{-1}-f''(x)\right\|\le C\|r_0\|$, $0\le k\le n-1$.

Використанням результатів лем можна встановити оцінку швидкості збіжності алгоритму спряжених напрямків, у якому матрицю H_k побудовано за формулою (19).

Теорема. Нехай функція $f(x)$, що двічі неперервно диференціюється, має матрицю других похідних, яка задовольняє умови (17), (18). Тоді, якщо здійснювати мінімізацію даної функції алгоритмом (19), то при виборі скалярного множника α_k з умови (8) для всіх достатньо великих ξ оцінка швидкості збіжності не нижча від $\|x_{k+1}-x_*\|\le C\|x_k-x_*\|\cdot\|x_{k-n+1}-x_*\|$.

6. Висновки

Запропоновані модифікації методів використовують лише перші похідні функцій, що мінімізуються. Побудовані формули дозволяють адаптувати обчислювальний процес до особливостей функції шляхом вибору відповідних значень параметрів, так і використанням додаткової інформації про функцію у кожній точці. Запропоновані модифікації методів можуть дозволити значною мірою послабити проблему уповільнення збіжності градієнтних методів у разі мінімізації певних класів функцій. Широке застосування методу зворотного поширення похибки, який базується на методі градієнтного спуску, в задачах навчання нейромереж, вимагає побудови саме таких методів. Подальші дослідження запропонованих методів можуть бути проведені в напрямку отримання практичних рекомендацій щодо їх застосування у прикладних задачах та оцінювання діапазону значень параметрів, які забезпечать стабільні результати у разі застосування нейронних мереж.

Література

1. Нікольський Ю.В. Невизначеність та надлишковість даних у моделі процесу їх аналізу / Ю.В.Нікольський // Східно-Європейський журнал передових технологій. – Харків, 2009. – № 6/2 (42). – С. 27-31.
2. Яковлев М.Н. О некоторых методах решения нелинейных уравнений / Яковлев М.Н. // Труды математического ин-та АН СССР, 84, 1975. – С. 8-43.
3. Davidon W.C. Variable Metric Method for Minimization / Davidon W.C. // A.E.C. Research and Development Report, ANL-5990, 1959. – P.1-27.
4. Fletcher R. A rapidly convergent descent method for minimization / Fletcher R., Powell M. // Comput. J. – 1963. – Vol.7. – P. 163-168.
5. McCormick G. P. Variable metric methods and unconstrained optimization/ McCormick G. P., Pearson J. D. // Proceedings of the Joint Conference on Optimization/ University of Keele. – 1968. – March. – P. 154-158.
6. Pearson J.D. Variable metric methods of minimization / Pearson J.D. // Comput J. – 1969. – Vol.12, №2. – P. 171–181.
7. Cohen A. Rate of coversgance for root finging and optimization algorithms: Ph. D. Dissertation / University of California Berkeley; Cohen A. // Полак Е. Численные методы оптимизации/ Полак Е. – М. :Мир,1974. – С.43-49.
8. Полак Б.Т. Метод сопряженных градиентов / Полак Б.Т. // Труды второй школы по матем. прогр. и смежным вопросам. – Вып. 1. – М., 1969. – С. 152–202.
9. Майстровский Г.Д. О сходимости метода сопряженных градиентов/ Майстровский Г.Д.// ЖВМ и МФ. – 1971. – Вип.11, №5. – С.1291–1294.
10. Смоляк С.А. Квадратичная сходимость метода сопряженных градиентов/ Смоляк С.А. // Труды 111 зимней школы по математическому программированию/ МИСИ. – М., 1970. – С.134-143.
11. Huang H.Y. Unitied approach to quadratically convergent algoritmus for function minimization / Huang H.Y. // JoTA. – 1970. – Vol.5, №6. – P.405-425.
12. Powell M. J. D. On the convergence of the variable metric algoritm / Powell M. J. D. // Report T.P. / AERE, Harwell, England. – 1970. – №382. – P.148-156.
13. Полак Е. Численные методы оптимизации. Единый подход / Полак Е. – М.: Мир, 1974.
14. Данилин Ю.М. Методы сопряженных направлений для решения задач минимизации / Данилин Ю.М. // Кібернетика. – 1971. – №5. – С.37-43.
15. Пшеничный Б.Н. Численные методы в экстремальных задачах / Пшеничный Б.Н., Данилин Ю.М. – М.: «Наука», 1975. – 319 с.
16. Данилин Ю.М. Скорость сходимости методов сопряженных направлений / Данилин Ю.М. // Кібернетика. – 1977. – №6. – С.24-37.