

Побудовано модифікації методів мінімізації функцій багатьох змінних. Введення параметрів та додаткової інформації про властивості функції дозволяє адаптувати процес обчислень. Доведено умови збіжності та отримано оцінки швидкості

Ключові слова: градієнтний метод, мінімізація, метод двоїстих напрямків

Построены модификации методов минимизации функций многих переменных. Введение параметров и дополнительной информации о свойствах функции позволяет адаптировать процесс вычислений. Доказано условия сходимости и получены оценки скорости

Ключевые слова: градиентный метод, минимизация, метод двойственных направлений

Some modifications of functions methods minimization is proposed. The introduction of parameters and additional information about the properties of the function allows adapting the process of computing. The conditions for convergence are proved and estimates of speed are obtained

Keywords: gradient method, minimization method, dual directions method

МЕТОДИ ЧИСЕЛЬНОЇ МІНІМІЗАЦІЇ У МОДЕЛЯХ 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) 370-440, 067-935-53-81

E-mail: y_nikol@yahoo.com

1. Вступ

Широке застосування методів машинного навчання у процесах аналізу даних [1] знов поставило на порядок денний необхідність побудови обчислювальних процедур для мінімізації функцій багатьох змінних. Такі процедури повинні дозволяти отримувати розв'язки з прийнятною швидкістю та бути незалежними від локальних властивостей функцій. Це викликано масовим застосуванням нейронних мереж прямого поширення сигналу, навчання яких здійснюють методом зворотного поширення похибки [2]. В основу цього методу покладено градієнтний метод, який володіє відомими недоліками, пов'язаними із особливостями формування траєкторії спуску для функцій із специфічним розташуванням локальних екстремумів. Використання класичних методів найшвидшого спуску та методів спряжених градієнтів пов'язане із традицією застосування простих обчислювальних процедур, які в той самий час володіють відомими недоліками.

2. Постановка задачі дослідження

В останні роки багато зусиль прикладено у дослідження нових методів мінімізації, які основані на ідеї генетичного алгоритму [3]. Такі методи не

набули широкого застосування, оскільки володіють низькою швидкістю збіжності та вимагають адаптації кроків мутації та схрещування до конкретної функції. Крім того, в роботах із застосування нейронних мереж широко застосовують метод Левенберга – Маркуарда [4]. Цей метод володіє складною обчислювальною процедурою, яка вимагає розв'язування допоміжної нетривіальної задачі знаходження розв'язку системи нелінійних рівнянь, яка рівносильна застосуванню методу Ньютона мінімізації функцій багатьох змінних [5]. Пропонується будувати обчислювальні процедури мінімізації, які можна регулювати за рахунок збільшення інформації про поточну точку траєкторії спуску та локальні властивості функції.

3. Аналіз останніх досліджень

Дослідження чисельних методів мінімізації функцій багатьох змінних має довгу історію. Найбільш очевидним серед чисельних методів мінімізації функцій багатьох змінних є метод градієнтного спуску. Його ідея запропонована ще А. Коші. Вивченню цього методу присвячена робота Л. В. Канторовича [6], в якій отримано оцінку швидкості збіжності при мінімізації квадратичної функції на основі інформації про обумовленість відповідної матриці других похідних.

Значний успіх у конструюванні та дослідженні різних аспектів чисельних методів безумовної мінімізації припадає на 70-ті роки ХХ ст., коли було побудовано велику кількість таких методів та закладено основи їх дослідження на швидкодю та обчислювальну складність. Пізніше з'явилися нові принципи конструювання оптимізаційних процедур, основані на емерджентних принципах, зокрема, генетичні методи, які знайшли своє застосування при розв'язуванні великої кількості прикладних задач. В той же час проблема застосування адекватного чисельного методу оптимізації для вирішення прикладної проблеми залишається відкритою. Це пов'язане із особливістю побудови обчислювальної процедури, яка володіє властивістю «жадібності». Це проявляється в тому, що кожне наступне наближення будується на основі невеликої кількості інформації про властивості функції. Тому можна казати, що пошук наступної точки у просторі аргументів здійснюється в умовах часткової невизначеності. Збільшення ефективності методу у разі вирішення практичної задачі та прискорення збіжності обчислювальних процедур має сенс здійснювати на основі врахування властивостей функції як у черговій точці траєкторії спуску мінімізаційної послідовності, так і в її околі. Цю додаткову інформацію доцільно використати при побудові обчислювальних формул. Ще одним можливим способом покращення властивостей чисельних методів є введення спеціальних параметрів в такі обчислювальні формули. З допомогою таких параметрів можна уточнити напрямок спуску, а самі параметри знаходити з допомогою процедур, які узгоджують їх в процесі вирішення задачі.

Розробка методів мінімізації є відносно самостійною галуззю обчислювальної математики, яка має на своєму рахунку значні досягнення. Ефективне розв'язування задачі мінімізації функції дозволяє істотно просунутися при вирішенні багатьох прикладних проблем аналізу даних, які у різний спосіб зводяться до оптимізаційних задач без обмежень. Основною причиною яка викликає зацікавленість методами мінімізації, є їх широке поширення в задачах аналізу даних, зокрема, при застосуванні нейронних мереж прямого поширення. Тривалий час, відтоді як градієнтний метод був запропонований та до 50-х років ХХ-го ст., мінімізація функцій виконувалась градієнтними методами, в яких послідовні наближення будують на основі лінійної апроксимації функції. Було запропоновано та ґрунтовно досліджено багато модифікацій градієнтних методів [7, 8, 9, 10]. Виявилось, що для всіх варіантів градієнтних методів швидкість збіжності невисока, тому застосування їх для вирішення складних завдань в більшості випадків є недоцільним. Дослідження Б. Поляка [11] показали, що швидкість збіжності градієнтного методу за умови спеціального вибору довжини його кроку, оцінюється збіжністю геометричної прогресії, знаменник якої залежить від властивостей функції, що мінімізують. До функцій, мінімізація яких градієнтними методами вважається неможливою, відносять так звані функції «типу яру». Деякі автори пропонували методи прискорення швидкості збіжності таких функцій, які певною мірою були б позбавлені вказаного недоліку. Загальна ідея побудови таких методів була висловлена І. Гельфандом та І.Цетліним [12], які запропонували після отримання

інформації про поведінку функції на двох послідовних ітераціях, зробити великий крок випадкової довжини уздовж «яру». Повторення у такий спосіб вказаної процедури надає можливість отримати точку мінімуму або вийти з області «яру».

4. Цілі статті

У пропонованій статті побудовано та досліджено модифікації методів мінімізації функцій багатьох змінних, які стосуються двох класів методів – градієнтних та двоїстих напрямків, та їх модифікацій, зокрема, градієнтно-параметричного методу та модифікованого методу двоїстих напрямків, відповідно. Запропоновані формули використовують обчислення лише перших похідних функцій. Основна відмінність запропонованих модифікацій методів полягає у використанні додаткової інформації у кожній точці траєкторії спуску. Такий підхід до побудови напрямку спуску дозволив ввести в обчислювальні формули параметри, з допомогою яких виникає можливість регулювати цей напрямок спуску із врахуванням локальних особливостей функцій. Також це дозволяє забезпечити швидкість збіжності до точки мінімуму.

5. Основний матеріал

Градієнтно-параметричний метод. Розглянемо алгоритм мінімізації опуклої достатньо гладкої функції n змінних $f(x)$, який задано рекурентними формулами

$$\begin{aligned} \bar{x}_k &= x_k - \lambda_k f'(x_k), \quad k=0,1,2,\dots, \\ x_{k+1} &= x_k - \beta_k (\theta_k f'(x_k) + (1-\theta_k) f'(\bar{x}_k)), \quad 0 < \theta \leq 1. \end{aligned} \tag{1}$$

де λ_k, β_k – скалярні множники, причому вибір λ_k здійснюється з умови

$$f(x_k - \lambda_k f'(x_k)) = \min_{\lambda} f(x_k - \lambda f'(x_k)). \tag{2}$$

Для вибору параметра β_k можна використати будь-який алгоритм, який забезпечить виконання співвідношення

$$\begin{aligned} f(x_k - \beta_k p_k) - f(x_k) &\leq -\epsilon \beta_k (f'_k(p_k)), \quad 0 < \epsilon < 0.5, \\ p_k &= \theta_k f'(x_k) + (1 - \theta_k) f'(\bar{x}_k). \end{aligned} \tag{3}$$

Нескладно переконатись, що у разі вибору $\theta_k = 1$ та λ_k із співвідношення (2) буде отримано метод найшвидшого спуску.

Використовуємо алгоритм (1) для мінімізації квадратичної функції

$$f(x) = 0.5(Ax, x), \tag{4}$$

де A – симетрична, додатно визначена матриця, для якої виконані умови

$$m \|y\|^2 \leq (Ay, y) \leq M \|y\|^2, \quad 0 < m < M, \quad y \in E^n.$$

Тоді

$$f'(x_k) = Ax_k,$$

$$p_k = \theta_k Ax_k + (1 - \theta_k) A(x_k - \lambda_k Ax_k) = A(\bar{x}_k + \lambda_k \theta_k Ax_k),$$

$$\lambda_k = \|Ax_k\|^2 (A^2 x_k, Ax_k)^{-1}, \quad k = 0, 1, \dots$$

Теорема 1. Якщо функція $f(x)$ обмежена знизу, її градієнт задовольняє умову Ліпшица $\|f'(x) - f'(y)\| \leq R\|x - y\|$ для будь-яких $x, y \in E^n$, вибір параметрів λ_k та β_k здійснюється з умов (2) та (3) відповідно, то для процесу (1) при $k \rightarrow 0$ виконується $\|f'_k\| \rightarrow 0$ незалежно від вибору початкової точки x_0 .

Теорема 2. Нехай для мінімізації квадратичної функції (4), що має симетричну додатно визначену матрицю A застосовано метод (1)-(2). Тоді для швидкості збіжності мінімізуючої послідовності $\{x_k\}$ виконується оцінка

$$\|x_{k+1}\| \leq \sqrt{\frac{M}{m} \prod_{i=0}^k q_i} \|x_0\|, \quad q_i < 1. \quad (5)$$

У формулі (5) позначено

$$N_k(\theta) = \frac{\theta_k^2}{1 - 2\gamma_{1,k}(1 - \theta_k) + \gamma_{2,k}(1 - \theta_k)^2},$$

$$\gamma_{1,k} = \lambda_k^2 \left(\frac{\|A^2 x_k\|}{\|Ax_k\|} \right)^2, \quad \gamma_{2,k} = \lambda_k^3 \frac{(A^3 x_k, A^2 x_k)}{\|Ax_k\|^2},$$

$$q_i = 1 - \frac{4Mm}{(M+m)^2} N_i(\theta).$$

Модифікований метод двоїстих напрямків. Розглянемо модифікацію методу двоїстих напрямків [5] мінімізації сильно опуклої достатньо гладкої функції n змінних $f(x)$, обчислювальну формулу якого запишемо у вигляді

$$\bar{x}_k = x_k - \bar{\alpha}_k A_{k-1}^{-1} f'(x_k), \quad (6)$$

$$x_{k+1} = x_k - \alpha_k A_k^{-1} f'(x_k), \quad k = 0, 1, 2, \dots, \quad (7)$$

а вибір параметрів α_k та $\bar{\alpha}_k$ здійснюється з умов

$$f(x_k + \alpha_k p_k) = \min_{\pm} f(x_k + \alpha p_k),$$

$$f(x_k + \bar{\alpha}_k p_k) = \min_{\pm} f(x_k + \bar{\alpha} p_k). \quad (8)$$

Основна відмінність пропонованої модифікації від базового методу двоїстих напрямків полягає у способі обчислення матриці A_k^{-1} .

Матриця A_k визначається системою рівнянь

$$A_k p_{k-i} = \epsilon_{k-i}, \quad i = 0, 1, 2, \dots, n-1. \quad (9)$$

Тут p_{k-i} послідовність векторів, для якої виконані умови:

1. Якщо Δ_k - визначник зі стовпцями $p_k \|p_k\|^{-1}, p_{k-1} \|p_{k-1}\|^{-1}, \dots, p_{k-n+1} \|p_{k-n+1}\|^{-1}$, то при будь-якому k виконується нерівність $|\Delta_k| \geq C > \epsilon$, ϵ - довільне мале додатне число.

2. $\|p_k\| \rightarrow 0$ при $k \rightarrow \infty$.

Нехай для p_{k-i} та ϵ_{k-i} виконано співвідношення:

$$\begin{cases} p_{k-i} = x_{k-0.5i} - x_{k-0.5i-1}, \\ \epsilon_{k-i} = f'(x_{k-0.5i-1} + p_{k-i}) - f'(x_{k-0.5i-1}), \end{cases}$$

де $i = 0, 2, \dots, n-2$ для парного n або $i = 0, 2, \dots, n-1$ для непарного n , або

$$\begin{cases} p_{k-i} = \bar{x}_{k-0.5(i-1)} - x_{k-0.5(i-1)}, \\ \epsilon_{k-i} = f(x_{k-0.5(i-1)} + p_{k-i}) - f(x_{k-0.5(i-1)}), \end{cases}$$

де $i = 1, 3, \dots, n-1$ для парного n або $i = 1, 3, \dots, n-2$ для n непарного.

Нехай для матриці других похідних $f''(x)$ функції $f(x)$ виконана умова Ліпшица

$$\|f''(x) - f''(y)\| \leq R\|x - y\|, \quad x, y \in E^n, \quad (10)$$

а також виконується

$$m\|y\|^2 \leq (f''(x)y, y) \leq M\|y\|^2, \quad 0 < m < M. \quad (11)$$

Тоді можна отримати оцінки швидкості збіжності методу, які сформульовані у наступній теоремі.

Теорема 3. Якщо для мінімізації функції $f(x)$ для якої виконано умови (9)-(10) використовується метод (6)-(7) у якому вибір параметра α_k з умови (8), то послідовність $\{x_k\}$ збігається до точки мінімуму із швидкістю, яка визначена оцінкою

$$\|x_{k+1} - x_*\| \leq C \|x_{k-0.5(n-2)} - x_*\| \cdot \|x_k - x_*\| \cdot \|x_{k-0.5n} - x_*\|$$

для n парного,

$$\|x_{k+1} - x_*\| \leq C \|x_{k-0.5(n+1)} - x_*\| \cdot \|x_{k-0.5(n-1)} - x_*\| \cdot \|x_k - x_*\|$$

для n непарного.

Систему (9) можна так записати у матричному вигляді $A_k^{-1} = R_k E_k^{-1}$, де E_k, R_k - матриці, стовпцями яких є відповідно вектори ϵ_{k-i}, p_{k-i} . Для побудови матриці A_k^{-1} необхідно обчислити матрицю E_k^{-1} , рядками якої будуть вектори, які утворюють базис $s_k, s_{k-1}, \dots, s_{k-n+1}$, що двоїстий до базису $\epsilon_k, \epsilon_{k-1}, \dots, \epsilon_{k-n+1}$.

Кожна з матриць $E_k, k = 1, 2, \dots$ відрізняється від сусідніх зліва та справа двома стовпцями. Тому побудову базису $s_k, s_{k-1}, \dots, s_{k-n+1}$ можна здійснювати рекурентними співвідношеннями. Покажемо спосіб їх побудови.

Нехай матриця E_k вже побудована, тобто відомий базис $s_k, s_{k-1}, \dots, s_{k-n+1}$. Побудуємо систему векторів $\bar{s}_{k+2}, \bar{s}_{k+1}, \dots, \bar{s}_{k-n+3}$ у такий спосіб

$$\bar{s}_{k+2} = \Delta^{-1}((s_{k-n+1}, \epsilon_{k+1})s_{k-n+2} - (s_{k-n+2}, \epsilon_{k+1})s_{k-n+1}),$$

$$\bar{s}_{k+1} = \Delta^{-1}((s_{k-n+2}, \epsilon_{k+2})s_{k-n+1} - (s_{k-n+1}, \epsilon_{k+2})s_{k-n+2}),$$

де

$$\Delta = (s_{k-n+2}, \epsilon_{k+2})(s_{k-n+1}, \epsilon_{k+1}) - (s_{k-n+2}, \epsilon_{k+1})(s_{k-n+1}, \epsilon_{k+2}),$$

$$\bar{s}_{k+1-j} = s_{k+1-j} - ((s_{k+1-j}, \epsilon_{k+1})\bar{s}_{k+1} + (s_{k+1-j}, \epsilon_{k+2})\bar{s}_{k+2})$$

для $j = 1, \dots, n-2$,

$$\bar{s}_{k+2-j} = s_{k+2-j} - \left((s_{k+2-j}, \epsilon_{k+1}) \bar{s}_{k+1} + (s_{k+2-j}, \epsilon_{k+2}) \bar{s}_{k+2} \right)$$

для $j = 2, 3, \dots, n-1$.
Якщо врахувати, що

$$A_k^{-1} = \sum_{i=0}^{n-1} p_{k-i} \epsilon_{k-i}^T \tag{12}$$

то формули (6) та (7) набудуть вигляду

$$x_{k+1} = x_k - \alpha_k \sum_{i=0}^{n-1} (s_{k-i}, f'_k) p_{k-i},$$

$$\bar{x}_k = x_k - \bar{\alpha}_k \sum_{i=0}^{n-1} (s_{k-i-2}, f'_k) p_{k-i-2}.$$

На основі виразу (12) можна отримати таку рекурентну формулу $A_{k+1}^{-1} = A_k^{-1} + \Delta A_k^{-1}$ для обчислення A_k^{-1} , де

$$\Delta A_k^{-1} = \Delta^{-1} \left((p_{k+2} - A_k^{-1} \epsilon_{k+2}) (a_{11} s_{k-n+2} - a_{12} s_{k-n+1})^T - (p_{k+1} - A_k^{-1} \epsilon_{k+1}) (a_{21} s_{k-n+2} - a_{22} s_{k-n+1})^T \right)$$

$$\Delta = a_{11} a_{22} - a_{21} a_{12},$$

$$a_{11} = (s_{k-n+1}, \epsilon_{k+1}), \quad a_{12} = (s_{k-n+2}, \epsilon_{k+1}), \quad a_{21} = (s_{k-n+1}, \epsilon_{k+2}),$$

$$a_{22} = (s_{k-n+2}, \epsilon_{k+2}).$$

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 застосування запропонованих методів та оцінювання діапазону значень параметрів, які забезпечать стабільні результати у разі розв'язування прикладних задач із застосуванням нейронних мереж, а також широкому обчислювальному експерименту із застосування запропонованих методів.

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. Хайкин С. Нейронные сети: полный курс/ Хайкин С. – М.: Издательский дом «Вильямс», 2006. – 1104 с.
3. Нікольський Ю.В. Генетичні алгоритми в екстремальних задачах/ Ю.В.Нікольський, Ю.М. Щербина // Вісн. Львів. ун-ту. Сер.: Прикладна математика та інформатика. – 2000. – Вип. 2. – С. 191-208.
4. Режим доступу: – http://ru.wikipedia.org/wiki/Алгоритм_Левенберга_-_Марквардта.
5. Пшеничный Б.Н. Численные методы в экстремальных задачах / Пшеничный Б.Н., Данилин Ю.М. – М.: Наука, 1975. – 320 с.
6. Канторович Л.В. О методе наискорейшего спуска/ Канторович Л.В.// дан СССР. – 1947. – Вип.56. – С. 235-236.
7. Бирман М.Ш. Некоторые оценки метода скорейшего спуска/ Бирман М.Ш // УМН. – 1950. – Вип. 5, №3. – С. 152-155.
8. Алексеев В. И. Градиентный метод поиска экстремума с вычислением коэффициента овражности / Алексеев В. И. // Автом. и вычисл. техника. – 1976. – №3. – С.43-46.
9. Фридман В.М. О сходимости методов типа наискорейшего спуска / Фридман В.М. // УМН. – 1962. – Вип.17, №3. – С.201-204.
10. Сафро В.М. О скорости сходимости некоторых градиентных методов/ Сафро В.М. // ЖВМ и МФ. – 1976. – №2, вип.16. – С. 496-499.
11. Поляк Б.Т. Градиентные методы минимизации функционалов / Поляк Б.Т. // ЖВМ и МФ. – 1963. – Вип.3, №4. – С.643-654.
12. Гельфанд И.М. Принцип нелокального поиска в задачах автоматической оптимизации / Гельфанд И.М., Цетлин М.Л. // ДАН СССР. – 1961. – Вип. 137. – С.295-298.