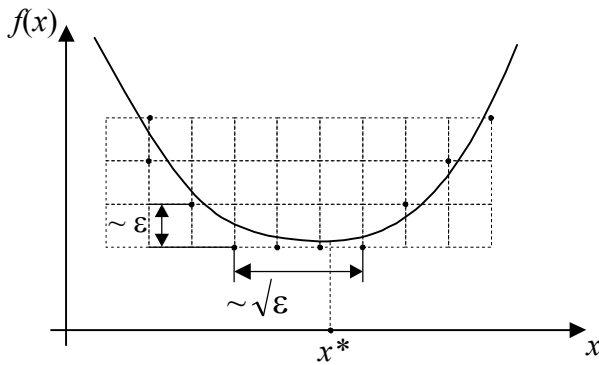


Розглянемо питання максимально досяжної точності при розв'язанні задач оптимізації. Нехай x^* — точка мінімуму функції $f(x)$. Розкладемо в околиці цієї точки функцію $f(x)$ в ряд Тейлора і врахуємо, що в точці мінімуму $f'(x^*) = 0$. Тоді

$$f(x^* + \Delta x) = f(x^*) + \Delta x f'(x^*) + \Delta x^2 \frac{f''(x^*)}{2} + \dots \approx f(x^*) + c \Delta x^2,$$

де $c = \frac{f''(x^*)}{2}$.



Звідси випливає, що для того щоб відхилення функції $f(x)$ від значення в стаціонарній точці $f(x^*)$ були помітними, необхідно, щоб:

$$\Delta x \sim \sqrt{\varepsilon}, \quad (3)$$

де ε — відносна похибка округлення дійсних чисел. Таким чином, якщо нелінійне рівняння $f(x)=0$ можна розв'язати з відносною похибкою ε , то задачу $\min\{f(x)\}$ тільки з відносною похибкою $\sqrt{\varepsilon}$.

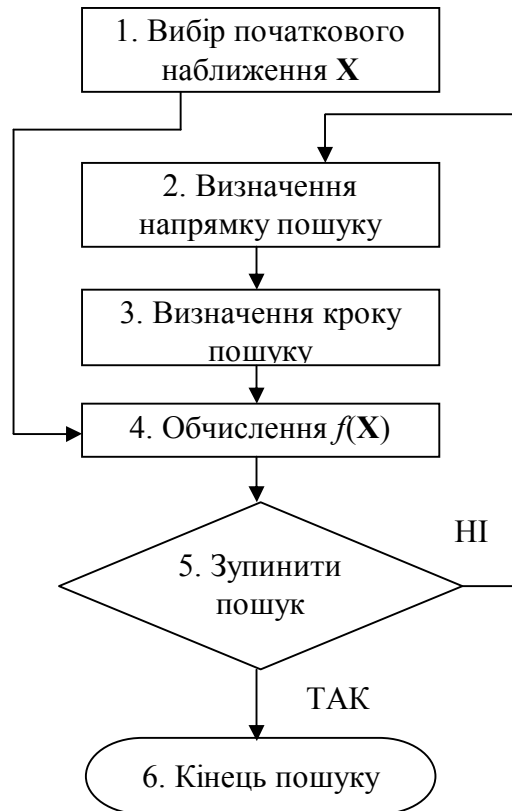
Найчисельнішу групу методів розв'язання задач оптимізації складають локальні методи безумовної оптимізації. Деяке уявлення про широко вживані методи цієї групи дає рисунок, наведений нижче.



В залежності від порядку використовуваних похідних цільової функції методи безумовної оптимізації поділяють на методи нульового, першого і другого порядків.

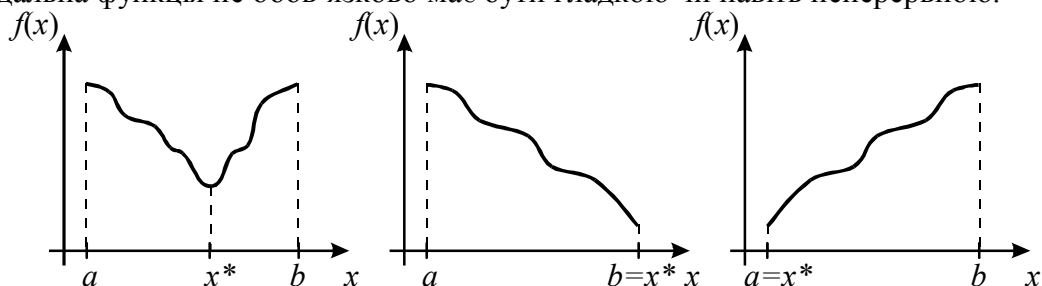
Практично всі методи оптимізації прагнуть побудувати таку послідовність значень X_0, X_1, X_2, \dots , при якій $f(X_0) > f(X_1) > f(X_2) > \dots$. В цьому випадку метод забезпечує збіжність і можна сподіватися, що мінімум функції буде знайдено.

Схема алгоритму пошуку для загального випадку показана на рисунку. Суть методу оптимізації визначається етапами 2 і 3 алгоритму, на яких вибирається напрямок подальшого пошуку і обчислюються координати чергової точки X_{i+1} на траєкторії пошуку.



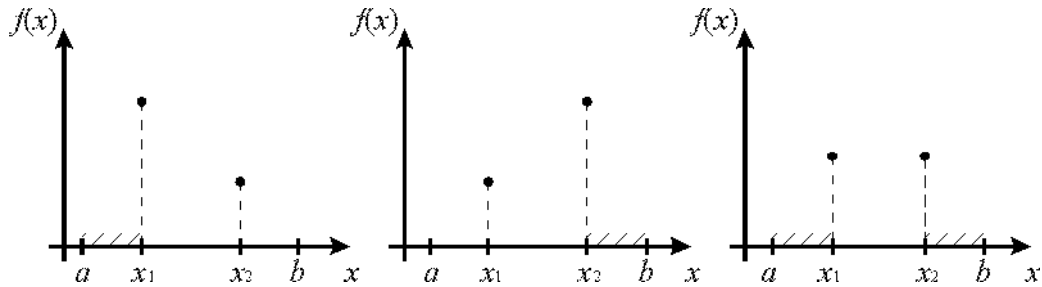
Методи одномірного пошуку, в основному, будують виходячи із припущення про унімодальність цільової функції $f(x)$ на відрізку $[a, b]$. Дійсну функцію $f(x)$ називають унімодальною, якщо існує єдина точка мінімуму x^* , і що $f(x)$ строго спадає для $x \leq x^*$ і строго зростає для $x \geq x^*$.

Приклади унімодальних функцій показані на рисунку нижче. Зауважимо, що унімодальна функція не обов'язково має бути гладкою чи навіть неперервною.



Методи одномірного пошуку можна розділити на методи послідовного пошуку (метод дихотомії, Фібоначчі і золотого перетину) і методи, які використовують апроксимацію функції (методи квадратичної і кубічної інтерполяції).

Розглянемо методи послідовного пошуку. Нехай відомо, що $f(x)$ унімодальна на $[a, b]$. Тоді за будь-якими двома значеннями $f(x_1), f(x_2)$ можна вказати інтервал, в якому знаходиться точка x^* , яка мінімізує $f(x)$, причому цей інтервал має довжину меншу за початкову. Нехай $x_1 < x_2$. Можливі наступні три варіанти:



В першому випадку слід відкинути інтервал $[a, x_1]$, в іншому випадку $[x_2, b]$, в третьому $[a, x_1]$, $[x_2, b]$, оскільки в цих інтервалах не може знаходитися x^* , інакше порушується припущення про унімодальність $f(x)$. Задача полягає в тому, щоб знайти множину абсцис x_1, x_2, \dots, x_n , в яких обчислюється функція, таку, що мінімум функції $f(x)$ лежить в деякому інтервалі $x^* \in [x_{i-1}, x_i]$. Такий інтервал називають інтервалом невизначеності.

Стратегія вибору значень x_{i-1} і x_i для проведення досліджень із врахуванням попередніх результатів визначає суть різних методів послідовного пошуку.

В методі дихотомії (половинного поділу) на кожному кроці аргументи x_1 і x_2 вибираються на відстані $\delta/2$ справа і зліва від середини інтервалу:

$$x_1 = \frac{a+b}{2} - \frac{\delta}{2}, \quad x_2 = \frac{a+b}{2} + \frac{\delta}{2}.$$

Обчислюючи $f(x_1)$ і $f(x_2)$ знаходять новий інтервал невизначеності:

- якщо $f(x_1) < f(x_2)$, то $b=x_2$;
- якщо $f(x_1) = f(x_2)$, то $a=x_1, b=x_2$;
- якщо $f(x_1) > f(x_2)$, то $a=x_1$.

Потім знову обчислюють x_1 і x_2 і продовжують пошук. Пошук зупиняють, коли довжина інтервалу невизначеності $|b-a|$ стає меншою, ніж задана похибка визначення x^* .

Золотий перетин, відкритий Евклідом, полягає в розбитті інтервалу $[a, b]$ точкою x_1 на дві частини таким чином, щоб відношення довжини всього інтервалу до великої частини дорівнювало відношенню великої частини до меншої:

$$\frac{b-a}{b-x_1} = \frac{b-x_1}{x_1-a}.$$

Легко перевірити, що золотий перетин виконують дві точки:

$$x_1 = a + (1 - \delta)(b - a), \\ x_2 = a + \delta(b - a),$$

де $\delta = \frac{\sqrt{5}-1}{2} \approx 0.618$.

Точка x_1 виконує золотий перетин відрізка $[a, x_2]$, а точка x_1 інтервалу $[x_1, b]$. Тому на інтервалі, який залишився треба визначити тільки одну точку, яка виконує золотий перетин. На кожному кроці довжина першого інтервалу невизначеності дорівнює ≈ 0.618 довжини старого інтервалу.

Алгоритм методу золотого перетину складається із таких кроків:

1. Обчислити

- $$x_1 := a + (1 - \delta)(b - a)$$
- $$x_2 := a + \delta(b - a).$$
2. Обчислити $f(x_1), f(x_2)$.
 3. Якщо $f(x_1) < f(x_2)$, то покласти $b := x_2, x_2 := x_1, f(x_2) := f(x_1), x_1 := a + (1 - \delta)(b - a)$ і обчислити $f(x_1)$
інакше покласти $a := x_1, x_1 := x_2, f(x_1) := f(x_2), x_2 := a + \delta(b - a)$ і обчислити $f(x_2)$
 4. Якщо $|b - a| > \varepsilon \left| \frac{b + a}{2} \right|$, то перейти до кроку 3
інакше $x^* \approx \frac{b + a}{2}$.

Із методів, які використовують апроксимацію функції, розглянемо алгоритм методу квадратичної інтерполяції:

1. Обчислити $f(x)$ в початковій точці x_0 .
2. Вибрати крок h . Якщо $f(x_0 + h) > f(x_0)$, то $h := -h$.
3. Обчислити $x_{i+1} = x_i + h$ и $f(x_{i+1})$.
4. Якщо $f(x_{i+1}) \leq f(x_i)$, то $h := 2h$ і перейти до етапу 3.
Якщо $f(x_{i+1}) > f(x_i)$, то $x_m := x_{i+1}, x_{m-1} := x_i, h := h/2$ і перейти до етапу 3 в останній раз.
5. Із чотирьох рівновіддалених значень $x_{m+1}, x_m, x_{m-1}, x_{m-2}$ виключити або x_{m+1} , або x_{m-2} , в залежності від того, яка із цих точок знаходиться далше від точки x , в якій $f(x)$ має найменше значення. Нехай x_a, x_b, x_c — три точки, які залишилися, де x_c — центральна точка, а $x_a = x_c - h, x_b = x_c + h$.
6. Провести квадратичну інтерполяцію для визначення координат точки:

$$x^* = x_c + \frac{h(f(x_a) - f(x_b))}{2(f(x_a) - 2f(x_c) + f(x_b))}.$$

До методів першого порядку відносять градієнтні методи. Відомо, що градієнт ортогональний до поверхні рівня цільової функції в точці його визначення і цей напрямок співпадає з напрямком найшвидшого зростання цільової функції.

Всі градієнтні методи використовують вказані особливості поведінки градієнта, а стратегія їх пошуку будується на рекурентному виразі вигляду:

$$\mathbf{X}_{i+1} = \mathbf{X}_i + h\mathbf{S}_i,$$

де h — величина кроку, а \mathbf{S}_i — одиничний вектор напрямку пошуку на i -му кроці.

Спосіб вибору кроку, напрямку пошуку або того і іншого одночасно визначає суть методу.

При пошуку мінімуму цільової функції треба рухатися в напрямку, протилежному градієнту $f(\mathbf{X})$. Тому напрямок:

$$\mathbf{S}_i = -\frac{\nabla f(\mathbf{X}_i)}{\|\nabla f(\mathbf{X}_i)\|_2}.$$

дозволяє отримати рекурентну формулу для метода градієнтного спуску:

$$\mathbf{X}_{i+1} = \mathbf{X}_i - h \frac{\nabla f(\mathbf{X}_i)}{\|\nabla f(\mathbf{X}_i)\|_2}. \quad (4)$$

Особливістю методу найшвидшого спуску являється рух із оптимальним кроком, розрахованим за допомогою одномірної мінімізації цільової функції по h вздовж антиградієнтного напрямку. Дійсно, якщо в якій-небудь точці \mathbf{X}_i напрямок пошуку визначено, то значення цільової функції в наступній точці \mathbf{X}_{i+1} виявляється функцією тільки кроку спуску:

$$f(\mathbf{X}_{i+1}) = f\left(\mathbf{X}_i - h \frac{\nabla f(\mathbf{X}_i)}{\|\nabla f(\mathbf{X}_i)\|_2}\right) = g(h).$$

Тому крок h можна вибрати так, щоб $f(\mathbf{X}_{i+1})$ максимально зменшила своє значення:

$$f(\mathbf{X}_{i+1}) = \min_h f\left(\mathbf{X}_i - h \frac{\nabla f(\mathbf{X}_i)}{\|\nabla f(\mathbf{X}_i)\|_2}\right) = \min_h g(h).$$

При цьому вибір оптимального кроку зводиться до розв'язання одновимірної оптимізації функції $f(h)$. Алгоритм методу найшвидшого спуску містить такі елементи:

1. Обчислення всіх часткових похідних цільової функції.
2. Знаходження одним із методів одновимірного пошуку оптимального кроку вздовж антиградієнтного напрямку. Величина кроку h визначається із умови мінімуму

функції $f\left(\mathbf{X}_i - h \frac{\nabla f(\mathbf{X}_i)}{\|\nabla f(\mathbf{X}_i)\|_2}\right)$ по h .

3. Обчислення координат нової точки \mathbf{X}_{i+1} по формулі (4).

4. Якщо умова припинення пошуку не виконується, то перехід до кроку 1.

Траєкторія пошуку цим методом показана на рисунку. Із рисунку видно, що рух вздовж одного напрямку припиняється, коли лінія напрямку пошуку стає дотичною до якої-небудь лінії рівного рівня. Кожен новий напрямок руху до екстремуму ортогональний попередньому, якщо цільова функція квадратична.

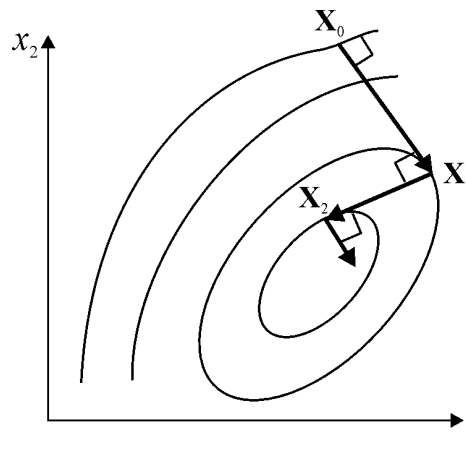


Рис. Ілюстрація методу найшвидшого спуску

Розглянуті методи сходяться до локального мінімуму зі швидкістю геометричної прогресії, тобто лінійно.

Розглянемо метод спряжених градієнтів. Вектори \mathbf{A} і \mathbf{B} називають спряженими по відношенню до матриці \mathbf{Q} (або \mathbf{Q} -спряженими), якщо скалярний добуток векторів \mathbf{A} і \mathbf{QB} дорівнює нулю, тобто $\mathbf{A}^T \mathbf{QB} = 0$. Спряженість векторів є узагальненим поняттям ортогональності, так якщо $\mathbf{Q} = \mathbf{E}$, то \mathbf{Q} -спряженість векторів означає їх ортогональність.

Припустимо, що пошук починається в точці \mathbf{X}_0 з початковим напрямком \mathbf{S}_0 , вибраним довільно. Тоді нова точка:

$$\mathbf{X}_1 = \mathbf{X}_0 + h\mathbf{S}_0, \quad (5)$$

де величину кроку h слід визначити із умови мінімуму функції $f(\mathbf{X}_0+h\mathbf{S}_0)$ по h . Для цього розкладемо функцію $f(\mathbf{X})$ в ряд Тейлора в околиці точки $\mathbf{X}=\mathbf{X}_0$ і обмежимося розглядом тільки трьох членів, що є точним для квадратичної цільової функції:

$$f(\mathbf{X}) \approx f(\mathbf{X}_0) + \nabla f(\mathbf{X}_0)^T (\mathbf{X} - \mathbf{X}_0) + \frac{1}{2} (\mathbf{X} - \mathbf{X}_0)^T \mathbf{H}(\mathbf{X}_0) (\mathbf{X} - \mathbf{X}_0),$$

$$\text{де } \mathbf{H}(\mathbf{X}_0) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_2^2} & \dots & \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_1 \partial x_n} & \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_2 \partial x_n} & \dots & \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_n^2} \end{bmatrix} \text{ — матриця Гессе.}$$

Тоді:

$$f(\mathbf{X}_0 + h\mathbf{S}_0) = f(\mathbf{X}_0) + \nabla f(\mathbf{X}_0)^T h\mathbf{S}_0 + \frac{1}{2} h\mathbf{S}_0^T \mathbf{H}(\mathbf{X}_0) h\mathbf{S}_0.$$

Із умови мінімуму останньої функції по h маємо:

$$\frac{df(\mathbf{X}_0 + h\mathbf{S}_0)}{dh} = \nabla f(\mathbf{X}_0)^T \mathbf{S}_0 + h\mathbf{S}_0^T \mathbf{H}(\mathbf{X}_0) \mathbf{S}_0 = 0.$$

Звідки

$$h_0 = -\frac{\nabla f(\mathbf{X}_0)^T \mathbf{S}_0}{\mathbf{S}_0^T \mathbf{H}(\mathbf{X}_0) \mathbf{S}_0}. \quad (6)$$

Можна показати, що після того як по (5) і (6) обчислена точка \mathbf{X}_1 , для продовження пошуку повинен бути вибраний новий напрямок \mathbf{S}_1 , так, щоб

$$\mathbf{S}_1^T \mathbf{H}(\mathbf{X}_0) \mathbf{S}_0 = 0,$$

тобто новий напрямок \mathbf{S}_1 повинен бути спряженим до старого напрямку \mathbf{S}_0 . Збіжність такого методу строго обґрунтовано для квадратичних функцій вигляду:

$$f(\mathbf{X}) = \mathbf{C}^T \mathbf{X} + \frac{1}{2} \mathbf{X}^T \mathbf{H} \mathbf{X},$$

де \mathbf{C} — вектор-стовпчик сталих коефіцієнтів, \mathbf{H} — матриця сталих коефіцієнтів. Мінімізація такої функції при додатно-визначеній матриці \mathbf{H} може бути виконана за n (або менше) кроків. В загальному випадку цільові функції не є квадратичними. Тому з використанням такого методу розв'язок буде знайдено за більше число кроків, ніж n . Недолік такого методу — необхідність трудомісткого розрахунку матриці Гессе. Цей недолік усунений в модифікованому Р. Флетчером і Ц.М. Рівсом варіанті, що називається методом спряженого градієнта.

В методі спряжених градієнтів будується послідовність напрямків пошуку \mathbf{S}_i , які є лінійними комбінаціями антиградієнта $-\nabla f(\mathbf{X}_i)$ і попередніх напрямків пошуку $(\mathbf{S}_0, \dots, \mathbf{S}_{i-1})$. Так, якщо $\mathbf{S}_0 = -\nabla f(\mathbf{X}_0)$, то $\mathbf{X}_1 = h_0 \mathbf{S}_0$ потрібно знайти новий $\mathbf{S}_1 = -\nabla f(\mathbf{X}_1) + \beta_1 \mathbf{S}_0$ напрямок, підбравши коефіцієнт β_1 такий, щоб \mathbf{S}_1 і \mathbf{S}_0 були спряженими. β_1 буде задовольняти умову спряженості векторів \mathbf{S}_0 і \mathbf{S}_1 якщо

$$\beta_1 = \frac{\nabla f(\mathbf{X}_1)^T \nabla f(\mathbf{X}_1)}{\nabla f(\mathbf{X}_0)^T \nabla f(\mathbf{X}_0)}.$$

В загальному випадку алгоритм методу спряжених градієнтів складається із таких етапів:

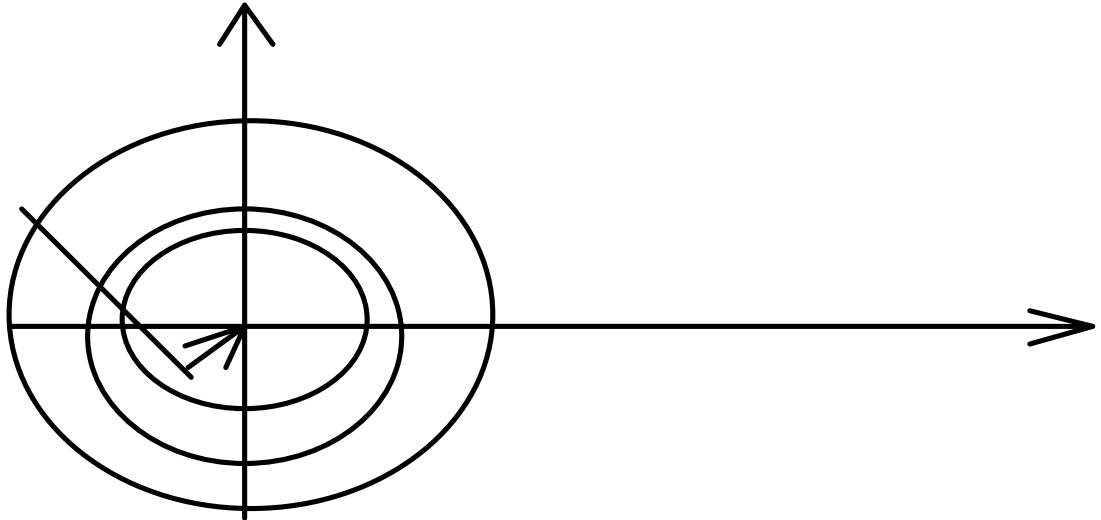
1. Обчислення вектора $\mathbf{S}_0 = -\nabla f(\mathbf{X}_0)$.
2. Знаходження мінімуму $f(\mathbf{X})$ одним із методів одномірного пошуку в напрямку \mathbf{S}_i . Звідки знаходимо \mathbf{X}_{i+1} , $f(\mathbf{X}_{i+1})$, $\nabla f(\mathbf{X}_{i+1})$.

3. Визначення нового спряженого напрямку \mathbf{S}_{i+1} із співвідношення:

$$\mathbf{S}_{i+1} = -\nabla f(\mathbf{X}_{i+1}) + \mathbf{S}_i \frac{\nabla f(\mathbf{X}_{i+1})^T \nabla f(\mathbf{X}_{i+1})}{\nabla f(\mathbf{X}_i)^T \nabla f(\mathbf{X}_i)}.$$

Після обчислення $(n+1)$ -го кроку (при $i=n$) обчислення циклічно повторюються.

4. Якщо $\|\mathbf{S}_i\| < \varepsilon$ пошук припиняється.



Методи спряжених градієнтів мають n -крокову квадратичну швидкість збіжності, тобто:

$$\|\mathbf{X}_{i+n} - \mathbf{X}^*\| \leq c \|\mathbf{X}_i - \mathbf{X}^*\|^2.$$

До методів другого порядку відноситься метод Ньютона. Нехай \mathbf{X}_i — i -е наближення до точки, яка відповідає мінімуму функції $f(\mathbf{X})$. В околиці цієї точки розкладемо функцію $f(\mathbf{X})$ в ряд Тейлора і обмежимося розглядом тільки трьох членів. Тоді,

$$f(\mathbf{X}) = f(\mathbf{X}_i) + \nabla f(\mathbf{X}_i)^T (\mathbf{X} - \mathbf{X}_i) + \frac{1}{2} (\mathbf{X} - \mathbf{X}_i)^T \mathbf{H}(\mathbf{X}_i) (\mathbf{X} - \mathbf{X}_i), \quad (7)$$

де $\mathbf{H}(\mathbf{X}_i)$ — матриця Гессе в точці $\mathbf{X} = \mathbf{X}_i$.

Нове наближення \mathbf{X}_{i+1} знайдемо із необхідної умови існування мінімуму функції (7):

$$\nabla f(\mathbf{X}_{i+1}) = \nabla f(\mathbf{X}_i) + \mathbf{H}(\mathbf{X}_i) (\mathbf{X}_{i+1} - \mathbf{X}_i). \quad (8)$$

Звідки

$$\mathbf{X}_{i+1} = \mathbf{X}_i - \mathbf{H}^{-1}(\mathbf{X}_i) \nabla f(\mathbf{X}_i).$$

Формулу (8) можна отримати застосуванням методу Ньютона до розв'язання системи нелінійних рівнянь $\nabla f(\mathbf{X}) = 0$. Тому такий метод називають методом Ньютона розв'язання задачі безумовної мінімізації.

При практичному використанні формули (8) не вдаються до безпосереднього обчислення оберненої матриці Гессе, а на кожному ітераційному кроці задачу розв'язують у два етапи. Спочатку розв'язують СЛАР відносно $\Delta \mathbf{X}_i$:

$$\mathbf{H}(\mathbf{X}_i)\Delta\mathbf{X}_i = -\nabla f(\mathbf{X}_i), \quad (9)$$

а потім знаходять нове наближення:

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \Delta\mathbf{X}_i. \quad (10)$$

Якщо початкове наближення \mathbf{X}_0 задане достатньо близько до точки \mathbf{X}^* , то метод (9), (10) має квадратичну швидкість збіжності, тобто:

$$\|\mathbf{X}_{i+1} - \mathbf{X}^*\| \leq c \|\mathbf{X}_i - \mathbf{X}^*\|^2,$$

де c — стала величина.

При розв'язанні СЛАР (9) слід мати на увазі дві важливі властивості матриці Гессе. По-перше, матриця Гессе є симетричною. По-друге, із (7) випливає, що якщо \mathbf{X}_i є точкою мінімуму, то $\nabla f(\mathbf{X}_i) = 0$ і, отже, для того, щоб функція $f(\mathbf{X})$ зростала при відхиленні від точки \mathbf{X}_i , необхідно щоб $(\mathbf{X} - \mathbf{X}_i)^T \mathbf{H}(\mathbf{X}_i)(\mathbf{X} - \mathbf{X}_i) > 0$. Отже, поблизу мінімуму матриця Гессе додатно визначена. Тому для розв'язання СЛАР (9) застосовується факторизація Холеського.

Метод Ньютона безумовної мінімізації має ті ж недоліки, що і метод Ньютона розв'язання нелінійних рівнянь: метод не має глобальної збіжності і вимагає аналітично заданих перших і других похідних цільової функції. Однак метод Ньютона безумовної оптимізації має додатковий недолік. Навіть будучи локальним, він не завжди призводить до точки мінімуму, оскільки в ньому немає нічого, щоб утримувало б його від просування в сторону максимуму або сідлової точки функції $f(\mathbf{X})$, в яких $\nabla f(\mathbf{X})$ також дорівнює нулю. Останнє можливе якщо матриця Гессе перестане бути додатно визначеною. Тому в практично використовуваних методах Ньютона, коли матриця Гессе не є додатно визначеною (з'являються від'ємні власні значення) її замінюють матрицею:

$$\tilde{\mathbf{H}} = \mathbf{H}(\mathbf{X}_i) + \mu\mathbf{E},$$

в якій μ вибирається так, щоб матриця $\tilde{\mathbf{H}}$ була додатно визначеною і добре обумовленою. При цьому найменше серед можливих μ повинне бути дещо більше, ніж модуль найменшого від'ємного власного значення матриці Гессе $\mathbf{H}(\mathbf{X}_i)$.

З метою збільшення області збіжності часто застосовують модифікований метод Ньютона, в якому нове наближення \mathbf{X}_{i+1} визначають за формулою:

$$\mathbf{X}_{i+1} = \mathbf{X}_i - \alpha \tilde{\mathbf{H}}^{-1} \nabla f(\mathbf{X}_i),$$

де $0 < \alpha \leq 1$. Величину α вибирають такою, щоб забезпечити зменшення цільової функції, тобто виконання умови:

$$f(\mathbf{X}_{i+1}) < f(\mathbf{X}_i).$$

Якщо аналітичний вираз для першої і другої похідних цільової функції не може бути заданий, то для їх визначення використовують скінчено-різницеву апроксимацію:

$$\frac{\partial f_m(\mathbf{X}_k)}{\partial x_i} \approx \frac{f_m(\mathbf{X}_k + \mathbf{I}\Delta x_i) - f_m(\mathbf{X}_k)}{\Delta x_i},$$

$$\frac{\partial^2 f_m(\mathbf{X}_k)}{\partial x_i^2} \approx \frac{f_m(\mathbf{X}_k - \mathbf{I}\widetilde{\Delta x}_i) - 2f_m(\mathbf{X}_k) + f_m(\mathbf{X}_k + \mathbf{I}\widetilde{\Delta x}_i)}{\widetilde{\Delta x}_i^2},$$

$$\frac{\partial^2 f_m(\mathbf{X}_k)}{\partial x_i \partial x_j} \approx \frac{f_m(\mathbf{X}_k + \mathbf{I}\widetilde{\Delta x}_i + \mathbf{J}\widetilde{\Delta x}_j) + f_m(\mathbf{X}_k - \mathbf{I}\widetilde{\Delta x}_i - \mathbf{J}\widetilde{\Delta x}_j)}{4\widetilde{\Delta x}_i \widetilde{\Delta x}_j} - \frac{f_m(\mathbf{X}_k - \mathbf{I}\widetilde{\Delta x}_i + \mathbf{J}\widetilde{\Delta x}_j) + f_m(\mathbf{X}_k + \mathbf{I}\widetilde{\Delta x}_i - \mathbf{J}\widetilde{\Delta x}_j)}{4\widetilde{\Delta x}_i \widetilde{\Delta x}_j},$$

де \mathbf{I} -вектор, i -й елемент якого дорівнює 1, а всі решта елементів дорівнюють 0, \mathbf{J} -вектор, j -й елемент якого дорівнює 1, а всі решта елементів дорівнюють 0. Для правильного вибору кроків апроксимації Δx_i , $\widetilde{\Delta x}_i$, $\widetilde{\Delta x}_j$ слід враховувати, що їх зменшення призводить до зменшення похибки апроксимації, але збільшує обчислювальну похибку. Тому на практиці їх вибирають виходячи із розумного компромісу як:

$$\Delta x_i = \begin{cases} \sqrt{\varepsilon_{\text{маши}}} x_i, & \sqrt{\varepsilon_{\text{маши}}} x_i \neq 0 \\ \sqrt{\varepsilon_{\text{маши}}} \text{ type } x_i, & \sqrt{\varepsilon_{\text{маши}}} x_i = 0 \end{cases}, \quad (5)$$

$$\widetilde{\Delta x}_i = \begin{cases} \sqrt[4]{\varepsilon_{\text{маши}}} x_i, & \sqrt[4]{\varepsilon_{\text{маши}}} x_i \neq 0 \\ \sqrt[4]{\varepsilon_{\text{маши}}} \text{ type } x_i, & \sqrt[4]{\varepsilon_{\text{маши}}} x_i = 0 \end{cases}$$

де $\varepsilon_{\text{маши}}$ - відносна похибка округлення дійсних чисел (у мові C для типу float $\varepsilon_{\text{маши}} \approx 1,2 \cdot 10^{-7}$), type x_i - типове значення змінної x_i .

Основним недоліком методу Ньютонa є громіздкі обчислення матриці Гессе. В методах змінної метрики (матрицю Гессе можна інтерпретувати як метрику в просторі градієнтів) використовують матрицю, що обчислюється простіше. Один із способів побудови методів змінної метрики ґрунтується на застосуванні методу січних розв'язання нелінійних рівнянь стосовно системи $\nabla f(\mathbf{X}) = 0$.

При цьому матриця Гессе є матрицею Якобі, побудованою для компонентів градієнта цільової функції. При використанні метода січних на кожному ітераційному кроці нова матриця Гессе $\tilde{\mathbf{H}}_{i+1}$ перераховується на основі попередньої $\tilde{\mathbf{H}}_i$.

Нехай \mathbf{H}_0 — обчислена матриця Гессе в початковій точці пошуку \mathbf{X}_0 , або яка інша додатно-визначена симетрична матриця, наприклад одинична матриця. Тоді вектор напрямку $\Delta \mathbf{X}_i$ пошуку на i -му кроці і наступна точка траєкторії можуть бути обчислені як:

$$\tilde{\mathbf{H}}_i \Delta \mathbf{X}_i = -\nabla f(\mathbf{X}_i),$$

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \alpha \Delta \mathbf{X}_i,$$

де α визначається із умови мінімуму функції $f(\mathbf{X}_i + \alpha \Delta \mathbf{X}_i)$ вздовж $\Delta \mathbf{X}_i$ одним із методів одновимірного пошуку.

Наступні наближення матриці $\tilde{\mathbf{H}}$ можна знайти із виразів:

$$\tilde{\mathbf{H}}_{i+1} = \tilde{\mathbf{H}}_i + \frac{\alpha \Delta \mathbf{X}_i \Delta \mathbf{X}_i^T}{\Delta \mathbf{X}_i^T \mathbf{P}_i} - \frac{\tilde{\mathbf{H}}_i \mathbf{P}_i \mathbf{P}_i^T \tilde{\mathbf{H}}_i}{\mathbf{P}_i^T \tilde{\mathbf{H}}_i \mathbf{P}_i},$$

де $\mathbf{P}_i = \nabla f(\mathbf{X}_{i+1}) - \nabla f(\mathbf{X}_i)$.

Робоче завдання

1. Побудувати алгоритми рішення задачі n -мірної оптимізації методами найшвидшого градієнтного спуску, спряжених градієнтів, Ньютона й змінної метрики.
2. Скласти робочу програму для розв'язання задачі оптимізації кожним із методів з відносною похибкою $\varepsilon = 5 \cdot 10^{-4}$. Кожен метод повинен бути оформлений у вигляді окремої універсальної функції. Формальні параметри функції повинні включати: ім'я функції, що реалізує обчислення цільової функції $f(X)$; кількість параметрів оптимізації n ; відносну похибку рішення; початкове наближення; розв'язок. У програмі передбачити можливість підрахунку числа викликів функцій.
3. Набрати й налагодити програму на комп'ютері, використовуючи цільову функцію з відомою точкою мінімуму.
4. Для задачі, вибраної відповідно до варіанту, побудувати цільову функцію й розв'язати задачу оптимізації розробленими методами. Для кожного методу одержати й записати значення й аргументи цільової функції на кожному ітераційному кроці алгоритмів.
5. Порівняти різні методи за швидкістю збіжності, надійності, необхідним машинним ресурсам (обсяг оперативної пам'яті, кількість арифметичних операцій, час виконання).

Зміст звіту

1. Назва роботи.
2. Мета.
3. Електрична схема, відповідно до варіанту, і отримана цільова функція.
4. Робоче завдання.
5. Математичні формулювання використаних алгоритмів розв'язання задачі оптимізації кожним методом.
6. Лістинг робочої програми.
7. Результати розрахунків.
8. Висновки.

Задачі

1.

Варіанти завдань

Номер бригади	Номер задачі	Номер варіанту задачі
1	2	1
2	3	1
3	4	1
4	5	1
5	6	1
6	7	1
7	1	1
8	2	2
9	3	2
10	4	2
11	5	2
12	6	2
13	7	2
14	1	2
15	3	3
16	4	3