Методи оптимізації. Лекція 29.04.2022

Розглянемо приклад застосування градієнтних методів.

Приклад. Розглянемо задачу мінімізації квадратичної функції

$$f(x) = x_1^2 + 4x_1x_2 + 6x_2^2 - 6x_1 - 20x_2,\ x = (x_1, x_2) \in E^2.$$

Оптимальний розв'язок цієї задачі є точка $x^* = (-1, 2)$, $f(x_*) = -17$. Лінії рівня цієї функції наведено на рис. 1.

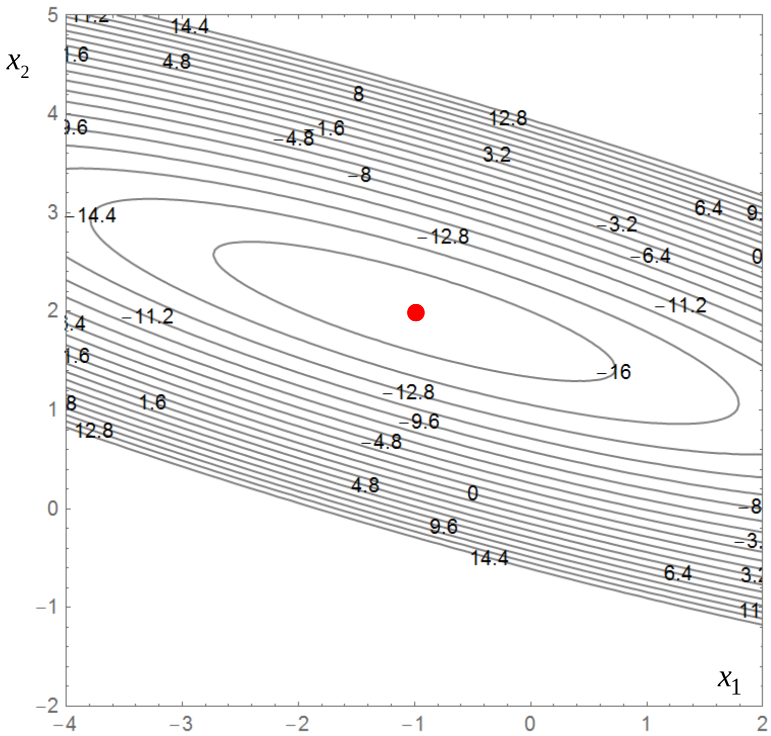


Рис. 1

Градієнт цієї функції має вигляд:

$$f'(x) = \begin{pmatrix} 2x_1 + 4x_2 - 6 \\ 4x_1 + 12x_2 - 20 \end{pmatrix}.$$

Знайдемо два перші наближення до точки мінімуму функції $f(x)$ за ***методом найшвидшого градієнтного спуску***.

Як початкову можна брати будь-яку точку простору E^2.

Нехай $x^{(0)} = (0,0)^T$ – початкова точка, $f\left(x^{(0)}\right) = 0$.

Для визначення крокового множника α_k побудуємо функцію $g(\alpha) = f\left(x^{(k)} - \alpha\ f'\left(x^{(k)}\right)\right)$, мінімум якої при $\alpha \geq 0$ необхідно визначити. Ця функція є функцією однієї змінної, отже, для знаходження мінімуму можна використати класичний метод або, при програмній реалізації, один з методів одновимірної оптимізації.

Перша ітерація. Обчислимо

$$x^{(1)} = x^{(0)} - \alpha_0\ f'\left(x^{(0)}\right),$$

де α_0 – шуканий кроковий множник, $f'\left(x^{(0)}\right)=\begin{pmatrix}-6\\-20\end{pmatrix}$. Тоді отримаємо

$$x^{(1)}=\begin{pmatrix}0\\0\end{pmatrix}-\alpha_0\begin{pmatrix}-6\\-20\end{pmatrix}=\begin{pmatrix}6\alpha_0\\20\alpha_0\end{pmatrix}.$$

Для визначення α_0 побудуємо функцію $g(\alpha_0)=f\left(x^{(0)}-\alpha_0\ f'\left(x^{(0)}\right)\right)$ і знайдемо мінімум цієї функції, тобто

$$g(\alpha_0)=36\alpha_0^2+4\cdot 6\alpha_0\cdot 20\alpha_0+6\cdot 400\alpha_0^2-6\cdot 6\alpha_0-20\cdot 20\alpha_0,$$

$$2916\alpha_0^2-436\alpha_0\to min,\ \alpha_0\geq 0.$$

Знайдемо розв’язок одновимірної задачі оптимізації класичним методом.

$$g'(\alpha_0)=0:\qquad 2\cdot 729\alpha_0-109=0,\ \Rightarrow\alpha_0=\frac{109}{1458}\approx 0.0748.$$

Отже, $x^{(1)}=\begin{pmatrix}0.449\\1.495\end{pmatrix}$, $f\left(x^{(1)}\right)\approx -16.297$. Умова монотонності виконується, так як $f\left(x^{(1)}\right)<f\left(x^{(0)}\right)$.

Перевіримо критерій завершення ітераційного процесу, тобто умову $\left\|x^{(k+1)}-x^{(k)}\right\|<\varepsilon$, де ε – задана точність.

Нехай $\varepsilon=10^{-2}$.

$$\left\|x^{(1)}-x^{(0)}\right\|=\sqrt{\left(x_1^{(1)}-x_1^{(0)}\right)^2+\left(x_2^{(1)}-x_2^{(0)}\right)^2}=\sqrt{(0.449-0)^2+(1.495-0)^2}=1.561>\varepsilon.$$

Друга ітерація. Обчислимо наступну точку $x^{(2)}$:

$$x^{(2)}=\begin{pmatrix}0.449\\1.495\end{pmatrix}-\alpha_1\begin{pmatrix}0.878\\-0.264\end{pmatrix}=\begin{pmatrix}0.449-0.878\cdot\alpha_1\\1.495+0.264\cdot\alpha_1\end{pmatrix}.$$

Отже,

$$g(\alpha_1)=f\left(x^{(1)}-\alpha_1\ f'\left(x^{(1)}\right)\right)=-16.297-0.84\cdot\alpha_1+0.262\cdot\alpha_1^2\to min,\ \alpha_1\geq 0,$$

звідки отримаємо $g'(\alpha_1)=0.524\cdot\alpha_1-0.84=0$ і $\alpha_1=1.6$.

Тоді $x^{(2)}=\begin{pmatrix}-0.956\\1.917\end{pmatrix}$. $f\left(x^{(2)}\right)\approx -16.972$. Умова монотонності виконується, так як $f\left(x^{(2)}\right)<f\left(x^{(1)}\right)$.

Перевіримо критерій завершення ітераційного процесу. Наприклад, умову $\left\|x^{(k+1)}-x^{(k)}\right\|<\varepsilon$, де ε – задана точність. Нехай $\varepsilon=10^{-2}$.

$$\left\|x^{(2)}-x^{(1)}\right\|=\sqrt{\left(x_1^{(2)}-x_1^{(1)}\right)^2+\left(x_2^{(2)}-x_2^{(1)}\right)^2}=1.467>\varepsilon.$$

Наближене значення мінімуму за методом найшвидшого градієнтного спуску після двох ітерацій є $x^* = x^{(2)} = \begin{pmatrix} -0.956 \\ 1.917 \end{pmatrix}$ та $f\left(x^*\right) \approx -16.972$.

Знайдемо наближення до точки мінімуму функції $f\left(x\right)$ за ***градієнтним методом с дробленням кроку***. Виконаємо дві ітерації методу.

Нехай початкова точка $x^{(0)} = \left(0,0\right)^T$, $f\left(x^{(0)}\right) = 0$. Задаємо константи $\beta = 0.1$ та $\lambda = 0.5$.

Перша ітерація. $\alpha_0 = \beta = 0.1$. Обчислимо

$$x^{(1)} = x^{(0)} - \alpha_0 \ f'\left(x^{(0)}\right),$$

При $f'\left(x^{(0)}\right) = \begin{pmatrix} -6 \\ -20 \end{pmatrix}$ отримаємо

$$x^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 0.1 \begin{pmatrix} -6 \\ -20 \end{pmatrix} = \begin{pmatrix} 0.6 \\ 2 \end{pmatrix}.$$

Для коефіцієнта $\alpha_0 = 0.1$ перевіряється умова монотонності

$$f\left(x^{(k)} - \alpha \ f'\left(x^{(k)}\right)\right) < f\left(x^{(k)}\right).$$

Маємо

$$f\left(x^{(1)}\right) < f\left(x^{(0)}\right)?$$

$-14.44 < 0$ умова монотонності виконується.

Перевіримо критерій завершення ітераційного процесу. Наприклад, умову $\left\| x^{(k+1)} - x^{(k)} \right\| < \varepsilon$, де ε – задана точність. Нехай $\varepsilon = 10^{-2}$.

$$\left\| x^{(1)} - x^{(0)} \right\| = \sqrt{\left(0.6 - 0\right)^2 + \left(2 - 0\right)^2} = 2.088 > \varepsilon.$$

Друга ітерація. $\alpha_1 = \beta = 0.1$. Обчислимо

$$x^{(2)} = x^{(1)} - \alpha_1 \ f'\left(x^{(1)}\right),$$

При $f'\left(x^{(1)}\right) = \begin{pmatrix} 3.2 \\ 6.4 \end{pmatrix}$ отримаємо

$$x^{(2)} = \begin{pmatrix} 0.6 \\ 2 \end{pmatrix} - 0.1 \begin{pmatrix} 3.2 \\ 6.4 \end{pmatrix} = \begin{pmatrix} 0.28 \\ 1.36 \end{pmatrix}.$$

Для коефіцієнта $\alpha_1 = 0.1$ перевіряється умова монотонності $f\left(x^{(2)}\right) < f\left(x^{(1)}\right)$? Маємо $-16.181 < -14.44$, умова монотонності виконується.

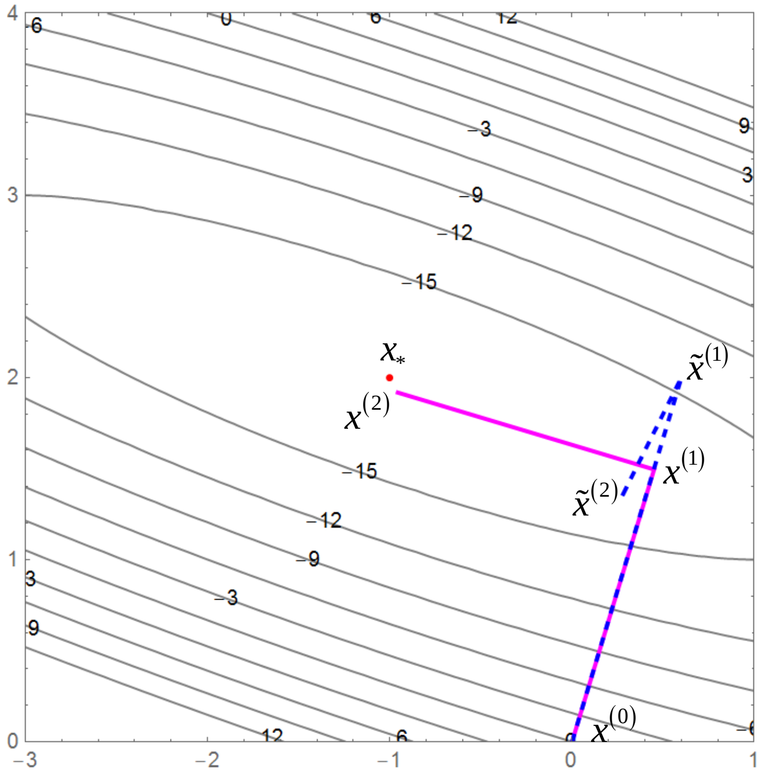
Перевіримо критерій завершення ітераційного процесу.

$$\left\| x^{(2)} - x^{(1)} \right\| = \sqrt{\left(0.28 - 0.6\right)^2 + \left(1.36 - 2\right)^2} = 0.7155 > \varepsilon.$$

Наближене значення мінімуму за градієнтним методом с дробленням кроку після двох ітерацій $x^{*}=x^{(2)}=\begin{pmatrix}0.28\\1.36\end{pmatrix}$ та $f\left(x^{*}\right)\approx-16.181$.

Рис. 2 ілюструє геометричний сенс градієнтних методів:

- для методу **найшвидшого спуску** два послідовно знайдених напрямки спуску $\left(-f'\left(x^{(0)}\right)\right)$ та $\left(-f'\left(x^{(1)}\right)\right)$ є взаємно ортогональними;
- градієнтний метод з дробленням кроку генерує послідовність точок $\left\{x^{(k)}\right\}_0^{\infty}$, які утворюють зигзагоподібну траєкторію.



– метод найшвидшого градієнтного спуску, точки $x^{(0)}, x^{(1)}, x^{(2)}$;

– градієнтний метод с дробленням кроку, точки $x^{(0)}, \tilde{x}^{(1)}, \tilde{x}^{(2)}$.

Рис. 2

Метод Ньютона та його модифікації

Постановка задачі.

$$f\left(x\right)\to min,\quad x\in E^{n}, \tag{1}$$

Умова застосовності:

$$f\left(x\right)\in C^{2}\left(E^{n}\right). \tag{2}$$

Розглянемо квадратичну апроксимацію цільової функції. Запишемо розкладання цільової функції $f\left(x\right)$ в ряд Тейлора в околі точки $x^{(k)}$:

$$f\left(x\right)-f\left(x^{(k)}\right)=\left(f'\left(x^{(k)}\right),x-x^{(k)}\right)+\frac{1}{2}\left(f''\left(x^{(k)}\right)\left(x-x^{(k)}\right),x-x^{(k)}\right)+$$
$$+o\left(\left\|x-x^{(k)}\right\|^{3}\right).$$

Позначимо квадратичну апроксимацію як $f_k(x)$:

$$f_k(x)=\left(f'\left(x^{(k)}\right),x-x^{(k)}\right)+\frac{1}{2}\left(f''\left(x^{(k)}\right)\left(x-x^{(k)}\right),x-x^{(k)}\right).$$

Знайдемо мінімум функції $f_k(x)$:

$$f_k(x)=\left(f'\left(x^{(k)}\right),x-x^{(k)}\right)+\frac{1}{2}\left(f''\left(x^{(k)}\right)\left(x-x^{(k)}\right),x-x^{(k)}\right)\to min,\ x\in E^n.$$

Застосуємо класичний метод.

$$f_k'(x)=0\Rightarrow f'\left(x^{(k)}\right)+f''\left(x^{(k)}\right)\left(x-x^{(k)}\right)=0.$$

Вважаємо, що $f''\left(x^{(k)}\right)\neq 0$, тоді $x=x^{(k)}-\left[f''\left(x^{(k)}\right)\right]^{-1}f'\left(x^{(k)}\right)$.

За наступну точку $x^{(k+1)}$ можна взяти

$$x^{(k+1)}=x^{(k)}-\left[f''\left(x^{(k)}\right)\right]^{-1}f'\left(x^{(k)}\right),\ k=0,1,\ldots \tag{3}$$

Формула (3) – ітераційна формула методу Ньютона.

Так як $f_k''(x)=f''(x)$, то у випадку, якщо функція $f(x)$ є опуклою, наближена точка мінімуму, яку отримано за методом (3), і буде розв'язком задачі.

Якщо цільова функція $f(x)$ не є опуклою, то точка, яку отримано з використанням формули (3), буде стаціонарною точкою. І тоді, ми повинні її додатково досліджувати, щоб перевірити чи є вона точкою мінімуму.

Метод Ньютона, в якому наближення знаходиться за формулою (3), називається ***класичним методом Ньютона***. Маємо, що

$$p^{(k)}=-\left[f''\left(x^{(k)}\right)\right]^{-1}f'\left(x^{(k)}\right),\ \alpha_k=1.$$

Він має істотний недолік: для довільної функції при невдало обраному початковому наближенні $x^{(0)}$ метод може розбігатися. Інший недолік: трудомісткість (на кожному кроці треба обчислювати і обертати матрицю Гессе). Перевага: висока швидкість збіжності.

Теорема про збіжність класичного методу Ньютона.

Нехай:

1. функція $f(x)\in C^2\left(E^n\right)$ і є сильно опуклою функцією:

$$m\|z\|^2\le\left(f''(x)z,z\right)\le M\|z\|^2,\quad \forall x,z\in E^n,\quad 0<m\le M,$$

2. матриця других похідних задовольняє умові Ліпшиця

$$\left\|f''(x)-f''(z)\right\|\le L\|x-z\|,\quad \forall x,z\in E^n;$$

3. початкове наближення $x^{(0)}$ таке, що

$\left\|f'\left(x^{(0)}\right)\right\|\le\dfrac{8M^2}{L}$ або, якщо ввести $q\in(0,1)$, то $\left\|f'\left(x^{(0)}\right)\right\|=\dfrac{8M^2}{L}q$.

Тоді послідовність $\left\{x^{(k)}\right\}$, яку обчислено за формулою (3), збігається до точки x_* із ***квадратичною*** швидкістю:

$$\left\| x^{(k+1)} - x_* \right\| \le \frac{4M^2}{L} q^{2^k}. \tag{4}$$

Неважко бачити, що метод Ньютона мінімізації функції $f(x)$, який визначається формулою (3), збігається з відомим з курсу методів обчислень методом Ньютона розв'язання системи нелінійних рівнянь $f'(x) = 0$.

Модифікації методу Ньютона

Розглянемо ***модифікований метод Ньютона*** (метод Ньютона з регулюванням кроку):

$$x^{(k+1)} = x^{(k)} - \alpha_k \left[f''\left(x^{(k)}\right) \right]^{-1} f'\left(x^{(k)}\right),\ k = 0,1,\dots, \tag{5}$$

де α_k – кроковий множник.

Параметр α_k вибирається як і в градієнтних методах: методом дроблення кроку, адаптивний вибір α_k, вибір з умови мінімуму функції в наступній точці.

1) $f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$.

2) $f\left(x^{(k)} + \alpha p^{(k)}\right) - f\left(x^{(k)}\right) \le \varepsilon\alpha\left(f'\left(x^{(k)}\right), p^{(k)}\right)$,

де $p^{(k)} = -\left[f''\left(x^{(k)}\right) \right]^{-1} f'\left(x^{(k)}\right)$.

3) $f\left(x^{(k)} - \alpha_k \left[f''\left(x^{(k)}\right) \right]^{-1} f'\left(x^{(k)}\right) \right) = \min\limits_{\alpha>0} f\left(x^{(k)} + \alpha p^{(k)}\right)$.

Модифікований метод Ньютона ***збігається*** при ***будь-якому початковому наближенні*** $x^{(0)} \in E^n$.

Модифікований метод Ньютона в залежності від вимог до функції, що мінімізується, збігається або зі надлінійною, або з квадратичною швидкістю.

Є й інші модифікації методу Ньютона. Іноді матрицю $\left[f''\left(x^{(k)}\right) \right]^{-1}$ обчислюють не на кожному кроці, а через наперед задане число кроків,

$$x^{(k+1)} = x^{(k)} - \alpha_k \left[f''\left(x^{\left(s\left[\frac{k}{s}\right]\right)} \right) \right]^{-1} f'\left(x^{(k)}\right), \alpha_k > 0, k = 0,1,2,\dots,$$

де $\left[\frac{k}{s}\right]$ – ціла частина числа $\frac{k}{s}$, тобто найбільше ціле число, яке менше або дорівнює $\frac{k}{s}$.

Або матрицю $\left[f''\left(x^{(k)}\right)\right]^{-1}$ замінюють на матрицю $\left[f''\left(x^{(0)}\right)\right]^{-1}$, тоді

$$x^{(k+1)} = x^{(k)} - \alpha_k \left[f''\left(x^{(0)}\right)\right]^{-1} f'\left(x^{(k)}\right),\ k = 0,1,\ldots, \tag{6}$$

Інший підхід зменшення трудомісткості – створення квазіньютоновських методів, в яких матриця других похідних замінюється її скінчено-різницевою апроксимацією.

Приклад. Методом Ньютона знайти точку мінімуму функції

$$f(x) = 9x_1^2 + x_2^2 - 18x_1 + 6x_2 + 18$$

з точністю $\varepsilon = 10^{-3}$.

Спочатку знайдемо точку мінімумі функції $f(x)$ класичним методом. Градієнт цієї функції має вигляд:

$$f'(x) = \begin{pmatrix} 18x_1 - 18 \\ 2x_2 + 6 \end{pmatrix}.$$

За необхідною умовою мінімуму маємо:

$$\begin{cases} 18x_1 - 18 = 0, \\ 2x_2 - 6 = 0. \end{cases} \Rightarrow \begin{cases} x_1 = 1, \\ x_2 = -3. \end{cases}$$

Будуємо матрицю других похідних (гессіан). Матриця других похідних для квадратичної функції є постійною і для заданої функції має вигляд:

$$f''(x) = \begin{pmatrix} 18 & 0 \\ 0 & 2 \end{pmatrix}.$$

Кутові мінори $M_1 = 18 > 0$, $M_2 = \begin{vmatrix} 18 & 0 \\ 0 & 2 \end{vmatrix} = 36 > 0$. Отже, гессіан є додатно визначеним, отже точка $x^* = (1,-3)$ є точкою мінімуму і $f\left(x^*\right) = 0$.

Тепер застосуємо метод Ньютона.

Матриця $\left[f''(x)\right]^{-1}$, обернена до матриці Гессе $f''(x)$, має вигляд

$$\left[f''(x)\right]^{-1} = \frac{1}{36}\begin{pmatrix} 2 & 0 \\ 0 & 18 \end{pmatrix} = \begin{pmatrix} 1/18 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

За початкове наближення візьмемо точку $x^{(0)} = (0,0)^T$, $f\left(x^{(0)}\right) = 18$. Для першого наближення за ітераційною формулою (3) маємо:

$$x^{(1)} = x^{(0)} - \left[f''\left(x^{(0)}\right)\right]^{-1} f'\left(x^{(0)}\right).$$

Підставляючи, отримаємо

$$x^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1/18 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} -18 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}.$$

Перевіримо критерій закінчення ітераційного процесу. Наприклад, умову $\left\|x^{(k+1)}-x^{(k)}\right\|<\varepsilon$, де ε – задана точність.

$$\left\|x^{(1)}-x^{(0)}\right\|=\sqrt{\left(x_1^{(1)}-x_1^{(0)}\right)^2+\left(x_2^{(1)}-x_2^{(0)}\right)^2}=\sqrt{10}>\varepsilon .$$

Критерій закінчення не виконано. І хоч ми бачимо, що точку мінімуму знайдено, але із-за невиконання умови зупинки робимо другу ітерацію.

$$x^{(2)}=x^{(1)}-\left[f''\left(x^{(1)}\right)\right]^{-1}f'\left(x^{(1)}\right).$$

Підставляючи, отримаємо

$$x^{(2)}=\begin{pmatrix}1\\-3\end{pmatrix}-\begin{pmatrix}1/18&0\\0&1/2\end{pmatrix}\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}1\\-3\end{pmatrix}.$$

$$\left\|x^{(2)}-x^{(1)}\right\|=0<\varepsilon .$$

Отже, оптимальний розв'язок $x^*=(1,-3)$.

Для будь-якої квадратичної функції з додатно визначеною матрицею других похідних метод Ньютона дає точний розв'язок незалежно від початкового наближення за одну ітерацію.

Приклад. Знайти точку мінімуму функції

$$f\left(x_1,x_2\right)=\frac{1}{2}\left(x_1^2-x_2\right)^2+\frac{1}{2}\left(1-x_1\right)^2.$$

Виконати дві ітерації класичного методу Ньютона з початкової точки $x^{(0)}=(2,2)$. Обчислити значення $f\left(x^{(0)}\right)$, $f\left(x^{(1)}\right)$, $f\left(x^{(2)}\right)$.

Градієнт цієї функції має вигляд:

$$f'(x)=\begin{pmatrix}2x_1\left(x_1^2-x_2\right)-\left(1-x_1\right)\\-\left(x_1^2-x_2\right)\end{pmatrix}=\begin{pmatrix}2x_1^3-2x_1x_2+x_1-1\\x_2-x_1^2\end{pmatrix}.$$

Будуємо матрицю других похідних (гессіан).

$$f''(x)=\begin{pmatrix}6x_1^2-2x_2+1&-2x_1\\-2x_1&1\end{pmatrix}.$$

За початкове наближення за умовою задана точка $x^{(0)}=(2,2)^T$, $f\left(x^{(0)}\right)=2.5$. Для першого наближення за ітераційною формулою (3) маємо:

$$x^{(1)}=x^{(0)}-\left[f''\left(x^{(0)}\right)\right]^{-1}f'\left(x^{(0)}\right).$$

Знайдемо

$$f'\left(x^{(0)}\right)=\begin{pmatrix}9\\-2\end{pmatrix},\ f''\left(x^{(0)}\right)=\begin{pmatrix}21&-4\\-4&1\end{pmatrix},\ \left[f''\left(x^{(0)}\right)\right]^{-1}=\frac{1}{5}\begin{pmatrix}1&4\\4&21\end{pmatrix}.$$

Підставляючи, отримаємо

$$x^{(1)}=\begin{pmatrix}2\\2\end{pmatrix}-\frac{1}{5}\begin{pmatrix}1&4\\4&21\end{pmatrix}\begin{pmatrix}9\\-2\end{pmatrix}=\begin{pmatrix}9/5\\16/5\end{pmatrix}=\begin{pmatrix}1.8\\3.2\end{pmatrix},\ f\left(x^{(1)}\right)=0.3208.$$

Робимо другу ітерацію: $x^{(2)}=x^{(1)}-\left[f''\left(x^{(1)}\right)\right]^{-1}f'\left(x^{(1)}\right)$.

Знайдемо

$$f'\left(x^{(1)}\right)=\begin{pmatrix}118/125\\-1/25\end{pmatrix},\ f''\left(x^{(1)}\right)=\begin{pmatrix}351/25&-18/25\\-18/25&1\end{pmatrix},$$

$$\left[f''\left(x^{(1)}\right)\right]^{-1}=\frac{25}{27}\begin{pmatrix}1&-18/5\\-18/5&351/25\end{pmatrix}=\begin{pmatrix}25/27&-10/3\\-10/3&13\end{pmatrix}.$$

Підставляючи, отримаємо

$$x^{(2)}=\begin{pmatrix}9/5\\16/5\end{pmatrix}-\begin{pmatrix}25/27&-10/3\\-10/3&13\end{pmatrix}\begin{pmatrix}118/125\\-1/25\end{pmatrix}=\begin{pmatrix}107/135\\103/15\end{pmatrix}=\begin{pmatrix}0.7926\\6.87\end{pmatrix}.$$

$$f\left(x^{(2)}\right)\approx 19.48.$$

При початковому наближенні $x^{(0)}=(2,2)^T$ класичний метод Ньютона розбігається.