Глава 7. МЕТОДЫ ВТОРОГО ПОРЯДКА

7.1. МЕТОД НЬЮТОНА

Постановка задачи

Пусть дана функция $f(x)$, ограниченная снизу на множестве R^n и имеющая непрерывные частные производные первого и второго порядков во всех его точках.

Требуется найти локальный минимум функции $f(x)$ на множестве допустимых решений $X = R^n$, т.е. найти такую точку $x^* \in R^n$, что

$$f(x^*) = \min_{x \in R^n} f(x).$$

Стратегия поиска

Стратегия метода Ньютона (Newton) состоит в построении последовательности точек $\left\{x^k\right\}, k = 0,1,\ldots$, таких, что $f\left(x^{k+1}\right) < f\left(x^k\right)$, $k = 0,1,\ldots$. Точки последовательности вычисляются по правилу

$$x^{k+1} = x^k + d^k, \quad k = 0,1,\ldots , \tag{7.1}$$

где x^0 – задается пользователем, а направление спуска d^k определяется для каждого значения k по формуле

$$d^k = -H^{-1}\left(x^k\right) \nabla f\left(x^k\right). \tag{7.2}$$

Выбор d^k по формуле (7.2) гарантирует выполнение требования $f\left(x^{k+1}\right) < f\left(x^k\right)$ при условии, что $H\left(x^k\right) > 0$. Формула (7.2) получена из следующих соображений:

1) функция $f(x)$ аппроксимируется в каждой точке последовательности $\left\{x^k\right\}$ квадратичной функцией $F_k = f\left(x^k\right) + \left(\nabla f\left(x^k\right), d^k\right) + \frac{1}{2}\left(d^k, H\left(x^k\right) d^k\right)$;

2) направление d^k определяется из необходимого условия экстремума первого порядка: $\frac{dF_k}{dd^k} = 0$. Таким образом, при выполнении требования $H\left(x^k\right) > 0$ последовательность является последовательностью точек минимумов квадратичных функций F_k, $k = 0,1,\ldots$(рис.7.1). Чтобы обеспечить выполнение требования $f\left(x^{k+1}\right) < f\left(x^k\right)$, $k = 0,1,\ldots$, даже в тех случаях, когда для каких-либо значений матрица Гессе $H\left(x^k\right)$ не окажется положительно определенной, рекомендуется для соответствующих значений k вычислить точку x^{k+1} по методу градиентного спуска $x^{k+1} = x^k - t_k \nabla f\left(x^k\right)$ с выбором величины шага t_k из условия $f\left(x^k - t_k \nabla f\left(x^k\right)\right) < f\left(x^k\right)$.

Построение последовательности $\left\{x^k\right\}$ заканчивается в точке x^k, для которой $\left\|\nabla f\left(x^k\right)\right\| < \varepsilon_1$, где ε_1 – заданное малое положительное число, или при $k \geq M$, где M – предельное число итераций, или при двукратном одновременном выполнении двух неравенств $\left\|x^{k+1} - x^k\right\| < \varepsilon_2$, $\left| f\left(x^{k+1}\right) - f\left(x^k\right)\right| < \varepsilon_2$, где ε_2 – малое положительное число.

Вопрос о том, может ли точка x^k рассматриваться как найденное приближение искомой точки минимума, решается путем проведения дополнительного исследования, которое описано ниже.

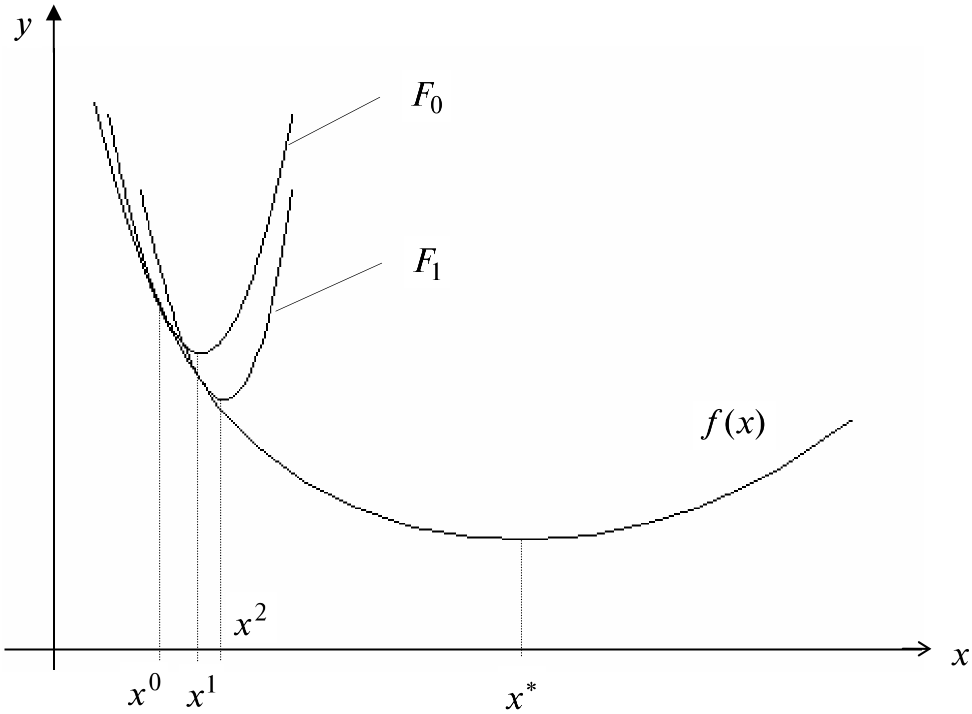

Рис. 7.1

Алгоритм

Шаг 1. Задать x^0, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, M – предельное число итераций. Найти градиент $\nabla f(x)$ и матрицу Гессе $H(x)$.

Шаг 2. Положить $k = 0$.

Шаг 3. Вычислить $\nabla f\left(x^k\right)$.

Шаг 4. Проверить выполнение критерия окончания $\left\|\nabla f\left(x^k\right)\right\| \leq \varepsilon_1$:

а) если неравенство выполнено, то расчет окончен и $x^* = x^k$;
б) в противном случае перейти к шагу 5.

Шаг 5. Проверить выполнение неравенства $k \geq M$:

а) если неравенство выполнено, расчет окончен и $x^* = x^k$;
б) если нет, перейти к шагу 6.

Шаг 6. Вычислить элементы матрицы $H\left(x^k\right)$.

Шаг 7. Найти обратную матрицу $H^{-1}\left(x^k\right)$.

Шаг 8. Проверить выполнение условия $H^{-1}\left(x^k\right) > 0$:

а) если $H^{-1}\left(x^k\right) > 0$, то перейти к шагу 9;

б) если нет, то перейти к шагу 10, положив $d^k = -\nabla f\left(x^k\right)$.

Шаг 9. Определить $d^k = -H^{-1}\left(x^k\right)\nabla f\left(x^k\right)$.

Шаг 10. Найти точку $x^{k+1} = x^k + t_k\, d^k$,
положив $t_k = 1$, если $d^k = -H^{-1}\left(x^k\right)\nabla f\left(x^k\right)$,

или выбрав t_k из условия $f\left(x^{k+1}\right) < f\left(x^k\right)$, если $d^k = -\nabla f\left(x^k\right)$.

Шаг 11. Проверить выполнение условий

$$\left\|x^{k+1} - x^k\right\| < \varepsilon_2, \qquad \left|f\left(x^{k+1}\right) - f\left(x^k\right)\right| < \varepsilon_2:$$

а) если оба условия выполнены при текущем значении k и $k = k-1$, то расчет окончен, $x^* = x^{k+1}$;

б) в противном случае положить $k = k+1$ и перейти к шагу 3.

Сходимость

Утверждение 7.1. *Пусть* $f(x)$ *дважды непрерывно дифференцируемая сильно выпуклая функция с константой* $l > 0$ *на* R^n *и удовлетворяющая условию*

$$\| H(x) - H(y) \| \le L \, \| x - y \| \qquad \forall x, y \in R^n,$$

где $L > 0$*, а начальная точка такова, что* $\left\|\nabla f\left(x^0\right)\right\| \le 8\dfrac{l^2}{L}$*, т.е.*

$$\left\| \nabla f(x^0) \right\| = \frac{8l^2 q}{L}, \tag{7.3}$$

где $q \in (0,1)$*. Тогда последовательность* $\left\{x^k\right\}$ *сходится к точке минимума с квадратичной скоростью* $\left\|x^k - x^*\right\| \le \dfrac{4lq^{2^k}}{L}$ [39].

З а м е ч а н и я 7.1.

1. Сходимость метода Ньютона доказана лишь для сильно выпуклых функций и для достаточно хорошего начального приближения, определяемого условием (7.3), практическое использование которого крайне затруднено, так как постоянные *l* и *L*, как правило, неизвестны или требуют трудоемкого исследования для их определения. Поэтому при практическом использовании метода Ньютона следует:

а) проанализировать матрицу $H\left(x^k\right)$ на выполнение условия $H\left(x^k\right) > 0$ $\forall k = 0,1,\ldots$ и заменить формулу $x^{k+1} = x^k - H^{-1}\left(x^k\right)\nabla f\left(x^k\right)$ на формулу $x^{k+1} = x^k - t_k\,\nabla f\left(x^k\right)$ в случае его невыполнения;

б) проанализировать точку x^k с целью выяснения, является ли она найденным приближением искомой точки x^*.

2. При решении задачи поиска безусловного максимума формула (7.2) не изменяется, так как в этом случае $H\left(x^k\right) < 0$.

Процедура решения задачи

1. Используя алгоритм Ньютона, найти точку x^k, в которой выполняется по крайней мере один критерий окончания расчета.

2. Так как $f(x) \in C^2$, то осуществить проверку выполнения достаточного условия минимума $H\left(x^k\right) > 0$. Если условие выполнено, то точка x^k может рассматриваться как найденное приближение точки минимума x^*. Проверку выполнения достаточного условия минимума можно заменить проверкой функции $f(x)$ на выпуклость.

Пример 7.1. Найти локальный минимум функции

$$f(x) = 2x_1^2 + x_1 x_2 + x_2^2 .$$

☐ I. Определим точку x^k, в которой выполняется по крайней мере один критерий окончания расчетов.

1. Зададим $x^0, \varepsilon_1, \varepsilon_2, M$: $x^0 = (0{,}5;1)^T$; $\varepsilon_1 = 0{,}1$; $\varepsilon_2 = 0{,}15$; $M = 10$. Найдем градиент функции $\nabla f(x) = (4x_1 + x_2; x_1 + 2x_2)^T$ и матрицу Гессе $H(x) = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$.

2. Положим $k = 0$.

3^0. Вычислим $\nabla f\left(x^0\right)$: $\nabla f\left(x^0\right) = (3; 2{,}5)^T$.

4^0. Проверим выполнение условия $\left\|\nabla f\left(x^0\right)\right\| \le \varepsilon_1$: $\left\|\nabla f\left(x^0\right)\right\| = 3{,}9 > 0{,}1$. Перейдем к шагу 5.

5^0. Проверим выполнение условия $k \ge M$: $k = 0 < 10$. Перейдем к шагу 6.

6^0. Вычислим $H\left(x^0\right)$: $H\left(x^0\right) = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$.

7^0. Вычислим $H^{-1}\left(x^0\right)$: $\quad H^{-1}\left(x^0\right) = \begin{pmatrix} \frac{2}{7} & -\frac{1}{7} \\ -\frac{1}{7} & \frac{4}{7} \end{pmatrix}$.

8^0. Проверим выполнение условия $H^{-1}\left(x^0\right) > 0$. Так как $\Delta_1 = \frac{2}{7} > 0$, $\Delta_2 = \frac{1}{7} > 0$, то согласно критерию Сильвестра $H^{-1}\left(x^0\right) > 0$ (см. гл. 2).

9^0. Определим $d_0 = -H^{-1}\left(x^0\right)\nabla f\left(x^0\right) = -\begin{pmatrix} \frac{2}{7} & -\frac{1}{7} \\ -\frac{1}{7} & \frac{4}{7} \end{pmatrix}\begin{pmatrix} 3 \\ 2{,}5 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}$.

10^0. Вычислим $x^1 = \left(\frac{1}{2},1\right)^T + \left(-\frac{1}{2},-1\right)^T = (0,0)^T$.

11^0. Проверим выполнение условий $\left\|x^1 - x^0\right\| < \varepsilon_2$, $\left|f\left(x^1\right) - f\left(x^0\right)\right| < \varepsilon_2$:

$$\left\|x^1 - x^0\right\| = 1{,}12 > 0{,}15; \qquad \left|f\left(x^1\right) - f\left(x^0\right)\right| = 2 > 0{,}15 \;.$$

Положим $k = 1$ и перейдем к шагу 3.

3^1. Вычислим $\nabla f\left(x^1\right)$: $\nabla f\left(x^1\right) = (0,0)^T$.

4^1. Проверим выполнение условия $\left\|\nabla f\left(x^1\right)\right\| < \varepsilon_1$: $\left\|\nabla f\left(x^1\right)\right\| = 0 < 0{,}1$. Расчет окончен. Заметим, что в точке x^1 выполняется необходимое условие первого порядка, поэтому она является стационарной точкой.

II. Проведем анализ точки x^1.

Функция $f(x) = 2x_1{}^2 + x_1 x_2 + x_2{}^2$ является строго выпуклой, так как ее матрица
x_2
вторых производных $H = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} > 0$ в силу того, что $\Delta_1 = 4 > 0$, $\Delta_2 = 7 > 0$. Найденная точка $x^1 = (0,0)^T$ есть точка локального и одновременно глобального минимума целевой функции $f(x)$.

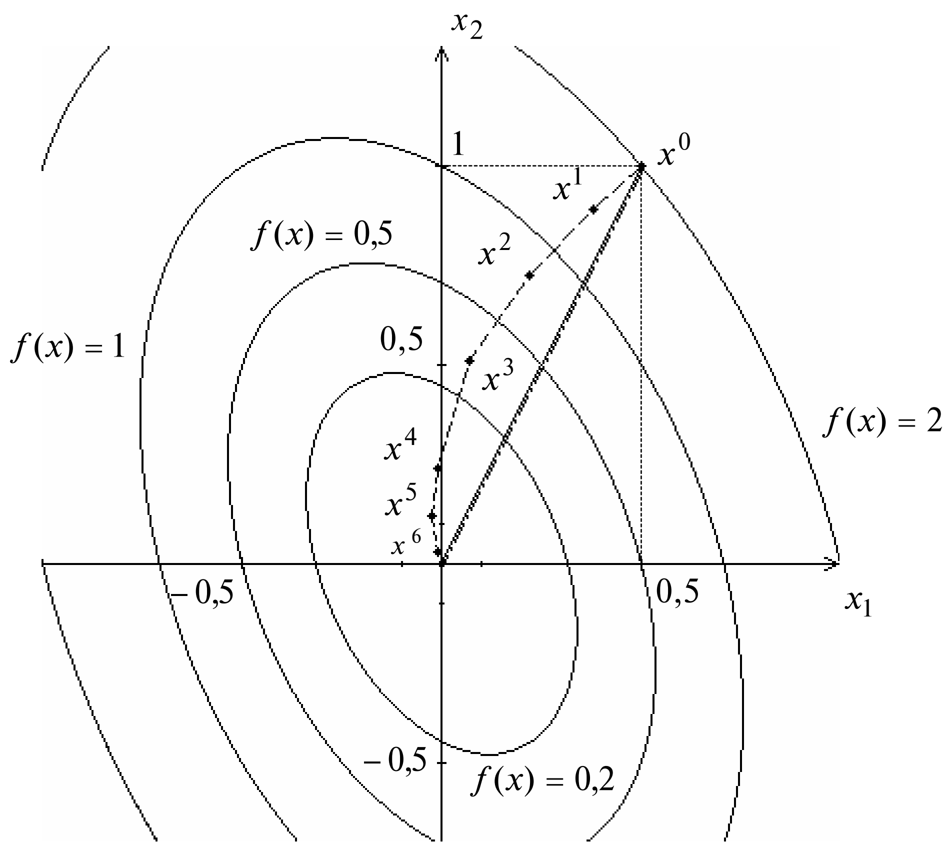

Рис. 7.2

На рис. 7.2 траектория спуска изображена сплошной линией. ■

7.2. МЕТОД НЬЮТОНА–РАФСОНА

Постановка задачи

Пусть дана функция $f(x)$, ограниченная снизу на множестве R^n и имеющая непрерывные частные производные первого и второго порядков во всех его точках.

Требуется найти локальный минимум функции $f(x)$ на множестве допустимых решений $X = R^n$, т.е. найти такую точку $x^* \in R^n$, что

$$f(x^*) = \min_{x \in R^n} f(x).$$

Стратегия поиска

Стратегия метода Ньютона–Рафсона (Newton–Raphson) состоит в построении последовательности точек $\left\{x^k\right\}, k = 0,1,\ldots$, таких, что $f\left(x^{k+1}\right) < f\left(x^k\right)$, $k = 0,1,\ldots$. Точки последовательности вычисляются по правилу

$$x^{k+1} = x^k - t_k \, H^{-1}\left(x^k\right) \nabla f\left(x^k\right), \; k = 0,1,\ldots , \tag{7.4}$$

где x^0 задается пользователем, а величина шага t_k определяется из условия

$$\varphi\left(t_k\right) = f\left(x^k - t_k \, H^{-1}\left(x^k\right) \nabla f\left(x^k\right)\right) \to \min_{t_k}. \tag{7.5}$$

Задача (7.5) может решаться либо аналитически с использованием необходимого условия минимума $\dfrac{d\varphi}{dt_k} = 0$ и последующей проверкой достаточного условия $\dfrac{d^2\varphi}{d\,{t_k}^2} > 0$, либо численно как задача

$$\varphi\left(t_k\right) \to \min_{t_k \in [a,b]}, \tag{7.6}$$

где интервал $[a,b]$ задается пользователем.

Если функция $\varphi\left(t_k\right)$ достаточно сложна, то возможна ее замена полиномом $P\left(t_k\right)$ второй или третьей степени и тогда шаг t_k может быть определен из условия $\dfrac{dP}{dt_k} = 0$ при выполнении условия $\dfrac{d^2P}{d{t_k}^2} > 0$.

При численном решении задачи определения величины шага степень близости найденного значения t_k к оптимальному значению t_k^*, удовлетворяющему условиям $\dfrac{d\varphi}{dt_k} = 0$, $\dfrac{d^2\varphi}{dt_k^2} > 0$, зависит от задания интервала $[a,b]$ и точности методов одномерной минимизации [27].

Построение последовательности $\{x^k\}$ заканчивается в точке x^k, для которой $\left\| \nabla f(x^k) \right\| < \varepsilon_1$, где ε_1 – заданное число, или при $k \ge M$, где M – предельное число итераций), или при двукратном одновременном выполнении двух неравенств $\left\| x^{k+1} - x^k \right\| < \varepsilon_2$, $\left| f(x^{k+1}) - f(x^k) \right| < \varepsilon_2$, где ε_2 – малое положительное число.

Вопрос о том, может ли точка x^k рассматриваться как найденное приближение искомой точки минимума, решается путем проведения дополнительного исследования, которое описано ниже.

Алгоритм

Шаг 1. Задать x^0, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, M – предельное число итераций. Найти градиент $\nabla f(x)$ и матрицу Гессе $H(x)$.

Шаг 2. Положить $k = 0$.

Шаг 3. Вычислить $\nabla f(x^k)$.

Шаг 4. Проверить выполнение условия $\left\| \nabla f(x^k) \right\| \le \varepsilon_1$:

а) если неравенство выполнено, то расчет закончен и $x^* = x^k$;
б) если нет, перейти к шагу 5.

Шаг 5. Проверить выполнение условия $k \ge M$:

а) если неравенство выполнено, расчет окончен и $x^* = x^k$;
б) если нет, перейти к шагу 6.

Шаг 6. Вычислить элементы матрицы $H(x^k)$.

Шаг 7. Найти обратную матрицу $H^{-1}(x^k)$.

Шаг 8. Проверить выполнение условия $H^{-1}(x^k) > 0$:

а) если условие выполняется, то найти $d^k = -H^{-1}(x^k)\nabla f(x^k)$;

б) если нет, то положить $d^k = -\nabla f(x^k)$.

Шаг 9. Определить $x^{k+1} = x^k + t_k d^k$.

Шаг 10. Найти шаг t_k^* из условия $\varphi(t_k) = f(x^k + t_k d^k) \to \min\limits_{t_k}$.

Шаг 11. Вычислить $x^{k+1} = x^k + t_k^* d^k$.

Шаг 12. Проверить выполнение неравенств

$$\left\| x^{k+1} - x^k \right\| < \varepsilon_2, \quad \left| f(x^{k+1}) - f(x^k) \right| < \varepsilon_2:$$

а) если оба условия выполнены при текущем значении k и $k = k - 1$, то расчет окончен и $x^* = x^{k+1}$;
б) в противном случае положить $k = k + 1$ и перейти к шагу 3.

Сходимость

Утверждение 7.2. *Пусть функция $f(x)$ дважды непрерывно дифференцируема и сильно выпукла на R^n, а ее матрица Гессе $H(x)$ удовлетворяет условию Липшица*

$$\| H(x) - H(y) \| \le L \| x - y \| \quad \forall x, y \in R^n .$$

Тогда последовательность $\{x^k\}$ сходится независимо от выбора начальной точки x^0 к точке минимума x^* с квадратичной скоростью $\left\| x^{k+1} - x^k \right\| \le \frac{L}{m} \left\| x^k - x^* \right\|^2$, где m – оценка наименьшего собственного значения матрицы [28].

З а м е ч а н и е 7.2. Сходимость к точке минимума метода Ньютона−Рафсона гарантируется независимо от выбора начального приближения лишь для сильно выпуклых функций. Поэтому при практическом использовании метода Ньютона−Рафсона следует:

а) проанализировать матрицу Гессе $H\left(x^k\right)$ на выполнение условия $H\left(x^k\right) > 0$, $k = 0,1,\ldots$, и заменить формулу $x^{k+1} = x^k - t_k H^{-1}\left(x^k\right) \nabla f\left(x^k\right)$ на формулу метода градиентного спуска $x^{k+1} = x^k - t_k \nabla f\left(x^k\right)$ в случае его невыполнения;

б) проанализировать точку x^k с целью выяснения, является ли она найденным приближением искомой точки x^*.

Процедура решения задачи

1. Используя алгоритм Ньютона−Рафсона, построить точку x^k, в которой выполняется по крайней мере один критерий окончания расчетов.

2. Так как $f(x) \in C^2$, то проверить выполнение достаточного условия минимума $H\left(x^k\right) > 0$. Если условие выполнено, то точка x^k может рассматриваться как найденное приближение точки минимума x^*. Проверку выполнения достаточного условия минимума можно заменить проверкой функции $f(x)$ на выпуклость.

Пример 7.2. Найти локальный минимум функции $f(x) = 2x_1^2 + x_1 x_2 + x_2^2$.

☐ I. Определим точку x^k, в которой выполняется по крайней мере один критерий окончания расчетов.

1. Зададим $x^0, \varepsilon_1, \varepsilon_2, M$: $x^0 = (0{,}5;1)^T$, $\varepsilon_1 = 0{,}1$; $\varepsilon_2 = 0{,}15$; $M = 10$. Найдем градиент функции $\nabla f(x) = (4x_1 + x_2; x_1 + 2x_2)^T$ и матрицу Гессе $H(x) = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$.

2. Положим $k = 0$.

3^0. Вычислим $\nabla f\left(x^0\right)$: $\nabla f\left(x^0\right) = (3; 2{,}5)^T$.

4^0. Проверим выполнение условия $\left\| \nabla f\left(x^0\right) \right\| \le \varepsilon_1$: $\left\| \nabla f\left(x^0\right) \right\| = 3{,}9 > 0{,}1$.

5^0. Проверим выполнение условия $k \ge M$: $k = 0 < 10$. Перейдем к шагу 6.

6^0. Вычислим $H\left(x^0\right)$: $H\left(x^0\right) = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$.

7^0. Вычислим $H^{-1}\left(x^0\right)$: $H^{-1}\left(x^0\right) = \begin{pmatrix} \frac{2}{7} & -\frac{1}{7} \\ -\frac{1}{7} & \frac{4}{7} \end{pmatrix}$.

8^0. Проверим выполнение условия $H^{-1}\left(x^0\right) > 0$.

Так как $\Delta_1 = \frac{2}{7} > 0,\ \Delta_2 = \frac{1}{7} > 0$, то согласно критерию Сильвестра $H^{-1}\left(x^0\right) > 0$. Поэтому найдем $d^0 = -H^{-1}\left(x^0\right)\nabla f\left(x^0\right)$: $d^0 = \left(-\frac{1}{2}, -1\right)^T$ (см. шаг 9^0 примера 7.1).

9^0. Определим: $x^1 = x^0 + t_0 d^0 = \left(\frac{1}{2}, 1\right)^T + t_0\left(-\frac{1}{2}, -1\right)^T = \left(\frac{1}{2} - \frac{1}{2}t_0, 1 - t_0\right)^T$.

10^0. Определим t_0^* из условия $\varphi(t_0) = f(x^0 + t_0 d^0) \to \min\limits_{t_0}$. Получим

$$f(x^0 + t_0 d^0) = f\left(\left(\frac{1}{2} - \frac{1}{2}t_0, 1 - t_0\right)^T\right) =$$

$$= 2 \cdot \left(\frac{1}{2} - \frac{1}{2}t_0\right)^2 + \left(\frac{1}{2} - \frac{1}{2}t_0\right) \cdot (1 - t_0) + (1 - t_0)^2 = 2 \cdot (1 - t_0)^2 = \varphi(t_0).$$

Из условия $\frac{d\varphi}{dt_0} = 2 \cdot 2 \cdot (1 - t_0) \cdot (-1) = 0$ находим $t_0^* = 1$. При этом $\frac{d^2\varphi}{d\,t_0^2} = 4 > 0$, т.е. найденная величина шага обеспечивает минимум функции $\varphi(t_0)$.

11^0. Вычислим $x^1 = x^0 + t_0^* d^0$: $x^1 = \left(\frac{1}{2} - \frac{1}{2} \cdot 1,\ 1 - 1\right)^T = (0,\ 0)^T$.

12^0. Проверим выполнение условий $\left\| x^1 - x^0 \right\| < \varepsilon_2$, $\left| f\left(x^1\right) - f\left(x^0\right) \right| < \varepsilon_2$:

$$\left\| x^1 - x^0 \right\| = 1{,}12 > 0{,}15; \quad \left| f\left(x^1\right) - f\left(x^0\right) \right| = 2 > 0{,}15.$$

Положим $k = 1$ и перейдем к шагу 3.

3^1. Вычислим $\nabla f\left(x^1\right)$: $\nabla f\left(x^1\right) = (0;0)^T$.

4^1. Проверим выполнение условия $\left\| \nabla f\left(x^1\right) \right\| < \varepsilon_1$: $\left\| \nabla f\left(x^1\right) \right\| = 0 < 0{,}1$. Расчет окончен: $x^* = x^1$.

II. Проведем анализ точки x^1.

Точка $x^* = (0;0)^T$ – точка локального и одновременно глобального минимума $f(x)$ (см. пример 7.1). На рис. 7.2 траектория спуска изображена штрихпунктирной линией. ■

7.3. МЕТОД МАРКВАРДТА

Постановка задачи

Пусть дана функция $f(x)$, ограниченная снизу на множестве R^n и имеющая непрерывные частные производные первого и второго порядков во всех его точках.

Требуется найти локальный минимум функции $f(x)$ на множестве допустимых решений $X = R^n$, т.е. найти такую точку $x^* \in R^n$, что

$$f(x^*) = \min_{x \in R^n} f(x).$$

Стратегия поиска

Стратегия метода Марквардта (Marquardt) состоит в построении последовательности точек $\left\{x^k\right\}$, $k = 0,1,\ldots$, таких, что $f\left(x^{k+1}\right) < f\left(x^k\right)$, $k = 0,1,\ldots$.

Точки последовательности $\left\{x^k\right\}$ вычисляются по правилу

$$x^{k+1} = x^k - \left[H\left(x^k\right) + \mu^k E\right]^{-1} \nabla f\left(x^k\right),\ k = 0,1,\ldots, \tag{7.7}$$

где точка x^0 задается пользователем, E – единичная матрица, μ^k – последовательность положительных чисел, таких, что матрица $\left[H\left(x^k\right) + \mu^k E\right]^{-1}$ положительно определена. Как правило, число μ^0 назначается как минимум на порядок больше, чем самый большой элемент матрицы $H\left(x^0\right)$, а в ряде стандартных программ полагается $\mu^0 = 10^4$ [36] . Если $f\left(x^k - \left(H\left(x^k\right) + \mu^k E\right)^{-1} \nabla f\left(x^k\right)\right) < f\left(x^k\right)$, то $\mu^{k+1} = \frac{\mu^k}{2}$. В противном случае $\mu^{k+1} = 2\mu^k$. Легко видеть, что алгоритм Марквардта в зависимости от величины μ^k на каждом шаге по своим свойствам приближается либо к алгоритму Ньютона, либо к алгоритму градиентного спуска.

Построение последовательности $\left\{x^k\right\}$ заканчивается, когда либо $\left\|\nabla f\left(x^k\right)\right\| < \varepsilon_1$, либо число итераций $k \ge M$, где ε_1 – малое положительное число, а M – предельное число итераций.

Вопрос о том, может ли точка x^k рассматривается как найденное приближение искомой точки минимума, решается путем проведения дополнительного исследования, которое описано ниже.

Алгоритм

Шаг 1. Задать $x^0, \varepsilon_1 > 0$, M – предельное число итераций. Найти градиент $\nabla f(x)$ и матрицу Гессе $H(x)$.

Шаг 2. Положить $k = 0,\ \mu^k = \mu^0$.

Шаг 3. Вычислить $\nabla f\left(x^k\right)$.

Шаг 4. Проверить выполнение условия $\left\| \nabla f\left(x^k\right) \right\| \le \varepsilon_1$:

а) если неравенство выполнено, то расчет окончен и $x^* = x^k$;
б) если нет, перейти к шагу 5.

Шаг 5. Проверить выполнение условия $k \ge M$:

а) если неравенство выполнено, расчет окончен и $x^* = x^k$;
б) если нет, перейти к шагу 6.

Шаг 6. Вычислить $H\left(x^k\right)$.

Шаг 7. Вычислить $H\left(x^k\right) + \mu^k E$.

Шаг 8. Найти обратную матрицу $\left[H\left(x^k\right) + \mu^k E\right]^{-1}$.

Шаг 9. Вычислить $d^k = -\left[H\left(x^k\right) + \mu^k E\right]^{-1} \nabla f\left(x^k\right)$.

Шаг 10. Вычислить $x^{k+1} = x^k - \left[H\left(x^k\right) + \mu^k E\right]^{-1} \nabla f\left(x^k\right)$.

Шаг 11. Проверить выполнение условия $f\left(x^{k+1}\right) < f\left(x^k\right)$:

а) если неравенство выполняется, то перейти к шагу 12;
б) если нет, перейти к шагу 13.

Шаг 12. Положить $k = k + 1,\ \mu^{k+1} = \dfrac{\mu^k}{2}$ и перейти к шагу 3.

Шаг 13. Положить $\mu^k = 2\mu^k$ и перейти к шагу 7.

Процедура решения задачи

1. Используя алгоритм Марквардта, построить точку x^k, в которой выполняется по крайней мере один критерий окончания расчетов.

2. Так как $f(x) \in C^2$, то проверить выполнение достаточного условия минимума $H\left(x^k\right) > 0$. Если условие выполнено, то точка x^k может рассматриваться как найденное приближение точки минимума x^*. Проверку выполнения достаточного условия минимума можно заменить проверкой функции $f(x)$ на выпуклость.

З а м е ч а н и я 7.3.

1. Метод Марквардта за счет выбора μ^k обеспечивает построение последовательности $\left\{x^k\right\}$, такой, что $f\left(x^{k+1}\right) < f\left(x^k\right),\ k = 0,1,\ldots$ [28].

2. В окрестности точки минимума x^* метод Марквардта обладает скоростью сходимости, близкой к квадратичной [28].

Пример 7.3. Найти локальный минимум функции $f(x) = 2x_1^{\,2} + x_1 x_2 + x_2^{\,2}$.

☐ I. Определим точку x^k, в которой выполняется по крайней мере один критерий окончания расчетов.

1. Зададим $x^0 = (0{,}5;1)^T$; $\varepsilon_1 = 0{,}1$; $M = 10$. Найдем градиент функции $\nabla f(x) = (4x_1 + x_2; x_1 + 2x_2)^T$ и матрицу Гессе $H(x) = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$.

2. Положим $k = 0$, $\mu^0 = 20$.

3^0. Вычислим $\nabla f(x^0)$: $\nabla f(x^0) = (3; 2{,}5)^T$.

4^0. Проверим выполнение условия $\|\nabla f(x^0)\| \le \varepsilon_1$: $\|\nabla f(x^0)\| = 3{,}9 > 0{,}1$. Перейдем к шагу 5.

5^0. Проверим выполнение условия $k \ge M$: $k = 0 < 10$. Перейдем к шагу 6.

6^0. Вычислим $H(x^0)$: $H(x^0) = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$.

7^0. Вычислим $H(x^0) + \mu^0 E$: $H(x^0) + \mu^0 E = \begin{pmatrix} 24 & 1 \\ 1 & 22 \end{pmatrix}$.

8^0. Вычислим $\left[H(x^0) + \mu^0 E\right]^{-1}$: $\left[H(x^0) + \mu^0 E\right]^{-1} = \begin{bmatrix} 0{,}0417 & -0{,}0019 \\ -0{,}0019 & 0{,}0455 \end{bmatrix}$.

9^0. Вычислим $d^0 = -\left[H(x^0) + \mu^0 E\right]^{-1} \nabla f(x^0)$: $d^0 = (-0{,}119; -0{,}108)^T$.

10^0. Вычислим $x^1 = x^0 - \left[H(x^0) + \mu^0 E\right]^{-1} \nabla f(x^0)$: $x^1 = (0{,}381; 0{,}892)^T$.

11^0. Проверим выполнение условия $f(x^1) < f(x^0)$: $f(x^1) = 1{,}438 < 2 = f(x^0)$.

12^0. Положим $k = 1$, $\mu^1 = \dfrac{\mu^0}{2} = 10$ и перейдем к шагу 3.

3^1. Вычислим $\nabla f(x^1)$: $\nabla f(x^1) = (2{,}41; 2{,}16)^T$.

4^1. Проверим выполнение условия $\|\nabla f(x^1)\| < \varepsilon_1$: $\|\nabla f(x^1)\| = 3{,}18 > 0{,}1$. Перейдем к шагу 5.

5^1. Проверим выполнение условия $k \ge M$: $k = 1 < 10$. Перейдем к шагу 6.

6^1. Вычислим $H(x^1)$: $H(x^1) = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$.

7^1. Вычислим $H(x^1) + \mu^1 E$: $H(x^1) + \mu^1 E = \begin{pmatrix} 14 & 1 \\ 1 & 12 \end{pmatrix}$.

8^1. Вычислим $\left[H(x^1) + \mu^1 E\right]^{-1}$:

$$\left[H(x^1) + \mu^1 E\right]^{-1} = \begin{bmatrix} 0{,}072 & -0{,}0059 \\ -0{,}0059 & 0{,}084 \end{bmatrix}.$$

9^1. Вычислим $d^1 = -\left[H\left(x^1\right) + \mu^1 E\right]^{-1} \nabla f\left(x^1\right)$: $d^1 = \left(-0{,}160; -0{,}168\right)^T$.

10^1. Вычислим $x^2 = x^1 - \left[H\left(x^1\right) + \mu^1 E\right]^{-1} \nabla f\left(x^1\right)$:

$$x^2 = \left(0{,}381; 0{,}892\right)^T - \left(0{,}160; 0{,}168\right)^T = \left(0{,}221; 0{,}724\right)^T.$$

11^1. Проверим выполнение условия $f\left(x^2\right) < f\left(x^1\right)$:

$$f\left(x^2\right) = 0{,}791 < 1{,}438 = f\left(x^1\right).$$

12^1. Положим $k = 2$, $\mu^2 = \dfrac{\mu^1}{2} = 5$ и перейдем к шагу 3.

3^2. Вычислим $\nabla f\left(x^2\right)$: $\nabla f\left(x^2\right) = \left(1{,}60; 1{,}67\right)^T$.

4^2. Проверим выполнение условия $\left\|\nabla f\left(x^2\right)\right\| < \varepsilon_1$: $\left\|\nabla f\left(x^2\right)\right\| = 2{,}31 > 0{,}1$. Перейдем к шагу 5.

5^2. Проверим выполнение условия $k \geq M$: $k = 2 < 10$. Перейдем к шагу 6.

6^2. Вычислим $H\left(x^2\right)$: $H\left(x^2\right) = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$.

7^2. Вычислим $H\left(x^2\right) + \mu^2 E$:

$$H\left(x^2\right) + \mu^2 E = \begin{pmatrix} 9 & 1 \\ 1 & 7 \end{pmatrix}.$$

8^2. Вычислим $\left[H\left(x^2\right) + \mu^2 E\right]^{-1}$:

$$\left[H\left(x^2\right) + \mu^2 E\right]^{-1} = \begin{pmatrix} 0{,}113 & -0{,}016 \\ -0{,}016 & 0{,}145 \end{pmatrix}.$$

9^2. Вычислим $d^2 = -\left[H\left(x^2\right) + \mu^2 E\right]^{-1} \nabla f\left(x^2\right)$: $d^2 = \left(-0{,}155; -0{,}217\right)^T$.

10^2. Вычислим $x^3 = x^2 - \left[H\left(x^2\right) + \mu^2 E\right]^{-1} \nabla f\left(x^2\right)$:

$$x^3 = \left(0{,}221; 0{,}724\right)^T - \left(0{,}155; 0{,}217\right)^T = \left(0{,}07; 0{,}51\right)^T.$$

11^2. Проверим выполнение условия $f\left(x^3\right) < f\left(x^2\right)$:

$$f\left(x^3\right) = 0{,}3 < 0{,}791 = f\left(x^2\right).$$

12^2. Положим $k = 3$, $\mu^3 = \dfrac{\mu^2}{2} = 2{,}5$ и перейдем к шагу 3.

3^3. Вычислим $\nabla f\left(x^3\right)$: $\nabla f\left(x^3\right) = \left(0{,}79; 1{,}09\right)^T$.

4^3. Проверим выполнение условия $\left\|\nabla f\left(x^3\right)\right\| < \varepsilon_1$: $\left\|\nabla f\left(x^3\right)\right\| = 1{,}34 > 0{,}1$. Перейдем к шагу 5.

5^3. Проверим выполнение условия $k \geq M$: $k = 3 < 10$. Перейдем к шагу 6.

6^3. Вычислим $H\left(x^3\right)$: $H\left(x^3\right)=\begin{pmatrix}4 & 1\\ 1 & 2\end{pmatrix}$.

7^3. Вычислим $H\left(x^3\right)+\mu^3 E$: $H\left(x^3\right)+\mu^3 E=\begin{pmatrix}6{,}5 & 1\\ 1 & 4{,}5\end{pmatrix}$.

8^3. Вычислим $\left[H\left(x^3\right)+\mu^3 E\right]^{-1}$:

$$\left[H\left(x^3\right)+\mu^3 E\right]^{-1}=\begin{pmatrix}0{,}159 & -0{,}035\\ -0{,}035 & 0{,}23\end{pmatrix}.$$

9^3. Вычислим $d^3=-\left[H\left(x^3\right)+\mu^3 E\right]^{-1}\nabla f\left(x^3\right)$: $d^3=\left(-0{,}078;-0{,}22\right)^T$.

10^3. Вычислим $x^4=x^3-\left[H\left(x^3\right)+\mu^3 E\right]^{-1}\nabla f\left(x^3\right)$:

$$x^4=\left(0{,}07;0{,}51\right)^T-\left(0{,}078;0{,}22\right)^T=\left(-0{,}008;0{,}29\right)^T.$$

11^3. Проверим выполнение условия $f\left(x^4\right)<f\left(x^3\right)$:

$$f\left(x^4\right)=0{,}082<0{,}3=f\left(x^3\right).$$

12^3. Положим $k=4$, $\mu^4=\dfrac{\mu^3}{2}=1{,}25$ и перейдем к шагу 3.

3^4. Вычислим $\nabla f\left(x^4\right)$: $\nabla f\left(x^4\right)=\left(0{,}26;0{,}57\right)^T$.

4^4. Проверим выполнение условия $\left\|\nabla f\left(x^4\right)\right\|<\varepsilon_1$: $\left\|\nabla f\left(x^4\right)\right\|=0{,}62>0{,}1$. Перейдем к шагу 5.

5^4. Проверим выполнение условия $k\ge M$: $k=4<10$. Перейдем к шагу 6.

6^4. Вычислим $H\left(x^4\right)$: $H\left(x^4\right)=\begin{pmatrix}4 & 1\\ 1 & 2\end{pmatrix}$.

7^4. Вычислим $H\left(x^4\right)+\mu^4 E$: $H\left(x^4\right)+\mu^4 E=\begin{pmatrix}5{,}25 & 1\\ 1 & 3{,}25\end{pmatrix}$.

8^4. Вычислим $\left[H\left(x^4\right)+\mu^4 E\right]^{-1}$:

$$\left[H\left(x^4\right)+\mu^4 E\right]^{-1}=\begin{pmatrix}0{,}203 & -0{,}0623\\ -0{,}0623 & 0{,}327\end{pmatrix}.$$

9^4. Вычислим $d^4=-\left[H\left(x^4\right)+\mu^4 E\right]^{-1}\nabla f\left(x^4\right)$: $d^4=\left(-0{,}017;-0{,}17\right)^T$.

10^4. Вычислим $x^5=x^4-\left[H\left(x^4\right)+\mu^4 E\right]^{-1}\nabla f\left(x^4\right)$:

$$x^5=\left(-0{,}008;0{,}29\right)^T-\left(0{,}017;0{,}17\right)^T=\left(-0{,}025;0{,}12\right)^T.$$

11^4. Проверим выполнение условия $f\left(x^5\right)<f\left(x^4\right)$:

$$f\left(x^5\right)=0{,}012<0{,}082=f\left(x^4\right).$$

12^4. Положим $k = 5$, $\mu^5 = \dfrac{\mu^4}{2} = 0{,}625$ и перейдем к шагу 3.

3^5. Вычислим $\nabla f\left(x^5\right)$: $\nabla f\left(x^5\right) = (0{,}02; 0{,}22)^T$.

4^5. Проверим выполнение условия $\left\|\nabla f\left(x^5\right)\right\| < \varepsilon_1$: $\left\|\nabla f\left(x^5\right)\right\| = 0{,}22 > 0{,}1$. Перейдем к шагу 5.

5^5. Проверим выполнение условия $k \ge M$: $k = 5 < 10$. Перейдем к шагу 6.

6^5. Вычислим $H\left(x^5\right)$: $H\left(x^5\right) = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$.

7^5. Вычислим $H\left(x^5\right) + \mu^5 E$: $H\left(x^5\right) + \mu^5 E = \begin{pmatrix} 4{,}622 & 1 \\ 1 & 2{,}622 \end{pmatrix}$.

8^5. Вычислим $\left[H\left(x^5\right) + \mu^5 E\right]^{-1}$:

$$\left[H\left(x^5\right) + \mu^5 E\right]^{-1} = \begin{pmatrix} 0{,}236 & -0{,}09 \\ -0{,}09 & 0{,}416 \end{pmatrix}.$$

9^5. Вычислим $d^5 = -\left[H\left(x^5\right) + \mu^5 E\right]^{-1} \nabla f\left(x^5\right)$: $d^5 = (0{,}015; -0{,}090)^T$.

10^5. Вычислим $x^6 = x^5 - \left[H\left(x^5\right) + \mu^5 E\right]^{-1} \nabla f\left(x^5\right)$:

$$x^6 = (-0{,}025; 0{,}12)^T + (0{,}015; -0{,}09)^T = (-0{,}01; 0{,}03)^T.$$

11^5. Проверим выполнение условия $f\left(x^6\right) < f\left(x^5\right)$:

$$f\left(x^6\right) = 0{,}0006 < 0{,}012 = f\left(x^5\right).$$

12^5. Положим $k = 6$, $\mu^6 = \dfrac{\mu^5}{2} = 0{,}311$ и перейдем к шагу 3.

3^6. Вычислим $\nabla f\left(x^6\right)$: $\nabla f\left(x^6\right) = (-0{,}01; 0{,}05)^T$.

4^6. Проверим выполнение условия $\left\|\nabla f\left(x^6\right)\right\| < \varepsilon_1$: $\left\|\nabla f\left(x^6\right)\right\| = 0{,}051 < 0{,}1$. Расчет окончен.

II. Проведем анализ точки x^6.

Точка $x^6 = (-0{,}01; 0{,}03)$ (см. рис. 7.2) является найденным приближением точки минимума x^*, так как функция $f(x) = 2x_1^2 + x_1 x_2 + x_2^2$ является строго выпуклой (см. пример 7.1). На рис. 7.2 полученная траектория спуска изображена пунктирной линией. ■

Задачи для самостоятельного решения

1. Будет ли удачной начальная точка $x^0 = (1; 2; 1; 1)^T$ для решения задачи $f(x) = \left(x_1^2 - x_2\right)^2 + (x_3 - x_4)^2 \to \min$ методом Ньютона?

Ответ: нет, так как матрица $H\left(x^0\right)$ не является положительно определенной.

2. После десяти итераций по методу Марквардта при решении задачи $f(x) = \left(x_1 - x_2^2\right)^2 + (1 - x_1)^2 \to \min$ программа остановилась в точке $x = (1;\ 1)^T$. Поясните причину остановки.

Ответ: точка $x = (1;\ 1)^T$ - точка минимума.

3. Методом Ньютона–Рафсона найдите точку безусловного минимума функции $f(x) = x_1^2 + x_1 x_2 + 2x_2^2$.

Ответ: точное решение $x^* = (0; 0)^T$.

4. В задаче $f(x) = 100x_1^2 + x_2^2 \to \min$, $x^0 = (0; 10)^T$ определите координаты точки x^1 с помощью метода Ньютона.

Ответ: $x^1 = (0; 0)^T$.

5. Будет ли удачной минимизация функции $f(x) = x_1^3 + x_1 x_2 + x_2^2 x_1^2 - 3x_1$ методом Ньютона из точки $x^0 = (2; 2)^T$?

Ответ: нет, так как $H\left(x^0\right)$ не является положительно определенной.

6. Решается задача

$$f(x) = \frac{1}{(x_1 + 1)^2 + x_2^2} \to \max.$$

Укажите, из какой начальной точки ее решение методом Ньютона потребует не более одной итерации.

Ответ: из любой начальной точки, если решать эту задачу как задачу поиска минимума функции $\dfrac{1}{f(x)} = (x_1 + 1)^2 + x_2^2$.

7. Методами Ньютона, Ньютона–Рафсона и Марквардта найти безусловный экстремум функций:

а) $f(x) = n^2 x_1^2 - x_1 x_2 + \dfrac{(n+1)^2}{2} {x_2}^2 - n x_1$;

б) $f(x) = (n+10)\, x_1^2 + 2n x_1 x_2 + (n+30)\, {x_2}^2 - 4(n+15)\, x_1 - 12n x_2$.