Теоретические сведения. Численные методы второго порядка ("высокая" траектория)

Метод Ньютона-Рафсона

МЕТОД НЬЮТОНА— РАФСОНА

ПОСТАНОВКА ЗАДАЧИ

Пусть дана функция $f(x)$, ограниченная снизу на множестве R^n и имеющая непрерывные частные производные 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 $f(x^*) = \min\limits_{x \in R^n} f(x)$, $f(x) \in C^2$.

СТРАТЕГИЯ ПОИСКА

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

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

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

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

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

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

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

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

При численном решении задачи определения величины шага степень близости найденного значения t_k к оптимальному значению t_k^*, удовлетворяющему условиям

$$\frac{d\varphi}{dt_k} = 0,\quad \frac{d^2\varphi}{dt_k^2} > 0,$$

зависит от задания интервала $[a, b]$ и точности методов одномерной минимизации [28].

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

а) если неравенство выполнено, то расчет закончен, $x^* = x^k$;

б) если нет, перейти к шагу 5.

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

а) если неравенство выполнено, расчет окончен, $x^* = x^k$;

б) если нет, перейти к шагу 6.

Шаг 6. Вычислить матрицу $H(x^k)$.

Шаг 7. Вычислить 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 $\varphi(t_k) = f\left(x^k + t_k d^k\right) \to \min\limits_{t_k}$.

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

Шаг 12. Проверить выполнение неравенств

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

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 7.2. 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

$$\|x^{k+1} - x^k\| \le \frac{L}{m}\|x^k - x^*\|^2,$$

где m — оценка наименьшего собственного значения матрицы [29].

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

б) производить анализ точки 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 осуществить проверку выполнения достаточных условий минимума $H(x^k) > 0$. Если условие выполнено, то точка x^k может рассматриваться как найденное приближение точки минимума x^*. Проверку выполнения достаточных условий 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. Определение точки x^k, в которой выполняется по крайней мере один критерий окончания расчетов.

1. Зададим x^0, ε_1, ε_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 $\nabla f(x^0)$: $\nabla f(x^0) = (3;\ 2{,}5)^T$.

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

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

6^0. Вычислим $H(x^0)$:

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

7^0. Вычислим $H^{-1}(x^0)$:

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

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

Так как

$$\Delta_1 = \frac{2}{7} > 0,\quad \Delta_2 = \frac{1}{7} > 0,$$

то согласно критерию Сильвестра $H^{-1}(x^0) > 0$. Поэтому найдем $d^0 = -H^{-1}(x^0)\nabla f(x^0)$: $d^0 = \left(-\frac{1}{2}, -1\right)^T$ (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 $\varphi(t_0) = f(x^0 + t_0 d^0) \to \min\limits_{t_0}$. Получаем

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

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

Из условия

$$\frac{d\varphi}{dt_0} = 2\cdot 2\cdot(1 - t_0)\cdot(-1) = 0$$

находим $t_0^* = 1$. При этом

$$\frac{d^2\varphi}{dt_0^2} = 4 > 0,$$

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 $\|x^1 - x^0\| < \varepsilon_2$, $|f(x^1) - f(x^0)| < \varepsilon_2$:

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

Положим $k = 1$ и перейдем к шагу 3.

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

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

II. Анализ точки x^1.

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

Вопросы