МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
им. М.В. ЛОМОНОСОВА

Механико – математический факультет
Кафедра вычислительной математики

И. О. Арушанян

Практикум на ЭВМ

Безусловная минимизация функций многих переменных

Третье издание,
дополненное

Москва, 2012

ОГЛАВЛЕНИЕ

БЕЗУСЛОВНАЯ МИНИМИЗАЦИЯ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ

1. Постановки задач минимизации

Пусть скалярная функция $f(x)$ определена на множестве $x \in X$, где множество X принадлежит некоторому метрическому пространству. Говорят, что на элементе (точке) $\bar{x} \in X$ функция $f(x)$ имеет *локальный минимум*, если существует такая конечная ε-окрестность точки $\bar{x}$, что для всех $x \in X$, удовлетворяющих $\|x - \bar{x}\| < \varepsilon$, выполняется неравенство

$$f(\bar{x}) \leq f(x). \tag{1}$$

Такая точка $\bar{x}$ называется точкой *локального минимума*. Если указанное неравенство выполняется как строгое при $\bar{x} \neq x$, то говорят, что $\bar{x}$ — точка *строгого* локального минимума. Подобных локальных минимумов у функции $f(x)$ может быть много. Если выполняется

$$f(\bar{x}) = \inf_{x} f(x), \tag{2}$$

то говорят, что $f(\bar{x})$ является *глобальным* (абсолютным) минимумом $f(x)$ на заданном множестве X, т.е. $f(x) > f(\bar{x})$ для всех $x \in X$. Всякая точка глобального минимума является и точкой локального минимума, но не наоборот.

Поиск хотя бы одной точки минимума $\bar{x}$ и минимума $f(\bar{x})$ называется *минимизацией* функции $f(x)$. Нахождение точки максимума сводится к задаче минимизации при помощи замены $f(x)$ на $-f(x)$.

В дальнейшем будем предполагать, что множество X компактно (т.е. из каждого бесконечного и ограниченного его подмножества можно выделить сходящуюся последовательность) и замкнуто (т.е. предел любой сходящейся последовательности его элементов принадлежит этому множеству). В частности, если множество X само является пространством, то это пространство должно быть банаховым. Будем также предполагать, что функция $f(x)$ непрерывна или, по крайней мере, кусочно-непрерывна.

Если перечисленные требования не выполняются, то поиск минимума затруднителен. Например, если $f(x)$ не является кусочно-непрерывной функцией, то единственный способ состоит в переборе всех точек x, на которых определена $f(x)$.

Заметим, что чем более жестким требованиям удовлетворяет $f(x)$ (например, требованию существования непрерывных производных различного порядка), тем легче строить численные алгоритмы.

Если множество является числовой осью, то задача минимизации состоит в поиске минимума функции одного вещественного переменного (*одномерная минимизация*).

Если же есть n-мерное векторное пространство, то говорят о поиске минимума функции n переменных (*многомерная минимизация*).

В случае когда — пространство функций $x(t)$, то задачу (1) называют задачей на *минимум функционала*. Для решения этих задач используются методы вариационного исчисления.

Глобальный минимум может быть определен только тогда, когда вычислены все локальные минимумы: наименьший из них и есть глобальный. Поэтому в основном рассматривают задачу поиска локальных минимумов.

Из курса математического анализа известно, что в точке минимума удовлетворяется уравнение

$$\frac{\partial f}{\partial x} = 0. \tag{3}$$

Для задачи одномерной минимизации $\dfrac{\partial f}{\partial x}$ является обычной производной $\dfrac{df}{dx}$. Тогда уравнение (3) становится одним нелинейным (в общем случае) уравнением с одним неизвестным, которое может быть решено каким-либо из численных методов вычисления нулей нелинейных уравнений.

В случае многомерной минимизации уравнение (3) представляет собой систему нелинейных уравнений

$$\frac{\partial f}{\partial x_i} = 0, \quad 1 \le i \le n,$$

которая решается специальными методами. (Заметим, что при минимизации функционалов уравнение (3) оказывается дифференциальным или интегро-дифференциальным.)

Однако на практике указанные уравнения являются сложными и для них известные итерационные методы решения нелинейных уравнений сходятся медленно или вообще не сходятся. Поэтому разработаны методы решения задачи (1) без приведения ее к виду (3).

Если множество является пространством, то говорят о *безусловной минимизации* функции $f(x)$.

Если же множество принадлежит какому-либо пространству, то задачу (1) называют задачей на *минимум в ограниченной области.* Когда множество выделяется из пространства системой ограничений типа равенств и/или неравенств, то говорят об *условной минимизации* и задачу (1) называют задачей на *условный экстремум* (или задачей *математического программирования*).

Задачи математического программирования по виду функции $f(x)$ разбиваются на следующие классы:

— функция $f(x)$ линейная и ограничения линейные: задача *линейного программирования*;

— функция $f(x)$ нелинейная и/или ограничения нелинейные (или ограничения нелинейные, а $f(x)$ — линейная функция): задача *нелинейного программирования.*

В свою очередь, если ограничения линейны, то задача нелинейного программирования может быть разбита на следующие подклассы:

— $f(x)$ дробно-рациональная функция: задача *дробно-рационального программирования*;

— $f(x)$ выпуклая квадратичная функция: задача *квадратичного программирования.*

Все перечисленные задачи называют еще *задачами оптимизации.*

Отдельный класс оптимизационных задач представляют задачи *оптимального управления.* Если в задачах оптимального управления процесс оптимизации можно представить в виде ряда последовательных этапов (шагов), то такие задачи называют *многошаговыми* задачами оптимизации (управления). Для их решения используются методы *динамического программирования*, которые применимы к непрерывной модели многошагового процесса оптимизации, когда управления и векторы состояния могут непрерывно изменяться. Однако для многих экономических и производственных задач характерной является дискретная модель, когда величины, описывающие процесс, могут принимать только дискретный ряд значений. В таких задачах применяются дискретные методы динамического программирования.

Оптимизационная задача называется *детерминированной* в том случае, если погрешностями вычисления или экспериментального определения значений функции $f(x)$ можно пренебречь. В противном случае оптимизационная задача называется *стохастической*. Для этого класса задач разработаны специальные методы.

Данное учебное пособие посвящено рассмотрению некоторых методов безусловной минимизации функций многих переменных, предлагаемых для практической реализации на ЭВМ при выполнении заданий студенческого практикума на механико-математическом факультете Московского университета.

2. Методы безусловной минимизации функций многих переменных

2.1. Вводные понятия

Пусть заданы множество X, принадлежащее некоторому метрическому пространству, и скалярная функция $f(x)$, определенная на этом множестве X. Задача на минимум функции $f(x)$ записывается в виде

$$f(x) \to \min, \quad x \in X. \tag{4}$$

В этой записи функцию $f(x)$ называют *целевой функцией*, X — *допустимым множеством*, любой элемент $x \in X$ — *допустимой точкой* задачи (4).

Поиск максимума функции $f(x)$ на X эквивалентен задаче вычисления минимума функции $-f(x)$ и записывается в виде

$$-f(x) \to \min, \quad x \in X. \tag{5}$$

Точки минимума и максимума называют *точками экстремума*, а задачи (4) и (5) называются *экстремальными задачами.* Вопрос о существовании решений этих задач базируется на теореме Вейерштрасса:

Пусть X — компакт в евклидовом n-мерном пространстве R^n (т.е. X — замкнутое ограниченное множество), а $f(x)$ — непрерывная функция на X. Тогда существует точка глобального минимума $f(x)$ на X.

Теорема Вейерштрасса имеет важное следствие: если функция $f(x)$ непрерывна на R^n и $\lim\limits_{\|x\|_2 \to \infty} f(x) \to +\infty$, то $f(x)$ достигает своего глобального минимума на любом замкнутом подмножестве в R^n.

Мы будем иметь дело с конечномерными задачами, когда допустимое множество X совпадает с R^n, т.е. когда задача (4) является задачей безусловной минимизации функций многих переменных.

Дадим ряд определений.

Градиентом функции $f(x)$ называется вектор первых частных производных

$$\operatorname{grad} f = \operatorname{grad} f(x) = f'(x) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

Антиградиентом функции $f(x)$ называется вектор первых частных производных, взятых со знаком минус, т.е. $-\operatorname{grad} f$.

Матрицей Гессе функции $f(x)$ называется матрица вторых частных производных

$$f''(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j=1,\ldots,n}.$$

Ниже будем предполагать, что смешанные производные функции $f(x)$ второго порядка непрерывны; следовательно, имеем

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right) = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) = \frac{\partial^2 f}{\partial x_j \partial x_i},$$

а это означает, что матрица Гессе является симметричной.

Функция $f(x)$ называется *дифференцируемой* в точке x^*, если она имеет в этой точке полный дифференциал, т.е. для полного приращения $f(x)$ в точке x^* имеет место равенство

$$\Delta f = f(x^* + \Delta x) - f(x^*) = \big(f'(x), \Delta x\big) + o\big(\|\Delta x\|_2\big).$$

Здесь и далее под $(\cdot,\cdot)$ подразумевается скалярное произведение векторов. Заметим, что если все частные производные непрерывны, то функция дифференцируема.

Разложение в ряд Тейлора функции $f(x)$ в точке x^* имеет вид

$$\begin{aligned} f(x^* + \Delta x) \quad &= \quad f(x^*) + \big(f'(x), \Delta x\big) + \frac{1}{2}\left(f''(x)\Delta x, \Delta x\right) + \\ &\quad + o\left(\|\Delta x\|_2^2\right). \end{aligned}$$

В приведенной записи удержаны три члена разложения. Полезны следующие частные случаи этого разложения.

а) Формула Лагранжа:

$$f(x) = f(x^*) + \big(f'(x^* + \alpha h), h\big), \quad h = x - x^*, \quad 0 < \alpha < 1.$$

б) Формула Ньютона–Лейбница:

$$f(x) = f(x^*) + \int\limits_0^1 \big(f'(x^* + \alpha h), h\big)\, d\alpha, \quad h = x - x^*, \quad 0 < \alpha < 1.$$

в) Формула Тейлора с остаточным членом в форме Лагранжа:

$$\begin{aligned} &f(x) = f(x^*) + \big(f'(x^*), h\big) + \frac{1}{2}\left(f''(x^* + \alpha h)h, h\right), \\ &h = x - x^*, \quad 0 < \alpha < 1. \end{aligned}$$

Частную производную функции $f(x)$ по x_i в точке x^* можно представить в виде

$$\frac{\partial f}{\partial x_i}(x^*) = \lim_{\alpha \to 0} \frac{f(x^* + \alpha e_i) - f(x^*)}{\alpha},$$

где e_i — вектор-столбец, у которого i-я координата равна единице, а остальные равны нулю.

Функция $f(x)$ называется *дифференцируемой* в точке x^*, если градиент $f'(x^*)$ существует и при всех достаточно малых $h \in R^n$ справедливо представление

$$f(x^* + h) = f(x^*) + \big(f'(x^*), h\big) + o\big(\|h\|_2\big).$$

Функция $f(x)$ называется *дважды дифференцируемой* в точке x^*, если матрица Гессе $f''(x^*)$ существует и симметрична и при всех достаточно малых $h \in R^n$ справедливо представление

$$f(x^* + h) = f(x^*) + \big(f'(x^*), h\big) + \frac{1}{2}\big(f''(x^*)h, h\big) + o\left(\|h\|_2^2\right).$$

Величина

$$f'(x^*, h) = \lim_{\alpha \to 0+} \frac{f(x^* + \alpha h) - f(x^*)}{\alpha}, \quad \|h\|_2 = 1$$

называется производной функции $f(x)$ в точке x^* по *направлению* вектора h. Функция $f(x)$ называется дифференцируемой в точке x^* по направлению вектора h, если величина $f'(x^*, h)$ существует и конечна. Если функция $f(x)$ дифференцируема в точке x^*, то она дифференцируема в точке x^* по направлению любого вектора h, причем выполняется равенство

$$f'(x^*, h) = \big(f'(x^*), h\big).$$

Условие, которому *необходимо* должна удовлетворять точка локального минимума (необходимое условие локальной оптимальности), дается следующей теоремой.

Теорема 1. *Пусть функция $f(x)$ дифференцируема в точке $\bar{x} \in R^n$. Если $\bar{x}$ — точка локального минимума, то*

$$\operatorname{grad} f(\bar{x}) = f'(\bar{x}) = \mathbf{0}.$$

Доказательство. Если $\bar{x}$ — точка локального минимума, то по определению существует такая ε-окрестность этой точки (ε-шар), что

$$f(\bar{x}) \le f(\bar{x} + \alpha h),$$

где h — любой вектор из R^n и $\|(\bar{x}+\alpha h)-\bar{x}\|_2 \le \varepsilon$, т.е. выполняется неравенство $\|\alpha h\|_2 \le \varepsilon$. Поскольку $f(x)$ дифференцируема, то

$$0 \le f(\bar{x}+\alpha h) - f(\bar{x}) = \big(f'(\bar{x}), \alpha h\big) + o\big(\|\alpha h\|_2\big).$$

Разделим обе части неравенства на α:

$$\big(f'(\bar{x}), h\big) + \frac{o\big(\|\alpha h\|_2\big)}{\alpha} \ge 0$$

и перейдем к пределу при $\alpha \to 0$:

$$\big(f'(\bar{x}), h\big) \ge 0.$$

Это неравенство верно при любых h, в том числе и для вектора $h = -f'(\bar{x})$, для которого имеем

$$-\big(f'(\bar{x}), f'(\bar{x})\big) = -\left\|f'(\bar{x})\right\|_2^2 \ge 0.$$

Следовательно, $\|f'(\bar{x})\|_2 = 0$, т.е. $f'(\bar{x}) = 0$. Теорема доказана.

Определение. Точка $\bar{x}$, для которой $f'(\bar{x}) = 0$, называется *стационарной* точкой функции $f(x)$.

Стационарная точка не обязательно является точкой минимума, поскольку $f'(\bar{x}) = 0$ — только необходимое, но не достаточное условие оптимальности. Приведем пример, когда стационарная точка не является точкой минимума. Рассмотрим функцию

$$f(x) = x_1^3 + x_2^3 - 3x_1x_2, \quad x \in R^n.$$

Градиент этой функции имеет вид

$$f'(x) = \left(3x_1^2 - 3x_2, 3x_2^2 - 3x_1\right).$$

Выпишем решения уравнения $f'(x) = 0$:

$$\bar{x}^{(1)} = \left(\bar{x}_1^{(1)}, \bar{x}_2^{(1)}\right) = (0,0), \quad \bar{x}^{(2)} = \left(\bar{x}_1^{(2)}, \bar{x}_2^{(2)}\right) = (1,1).$$

Точка $\bar{x}^{(1)}$ является стационарной, но не является точкой минимума, т.е. нет такого ε-шара с центром в $\bar{x}^{(1)}$, для которого при всех $x : \left\|x - \bar{x}^{(1)}\right\| < \varepsilon$ выполнено неравенство $f(x) \ge f\big(\bar{x}^{(1)}\big)$.

Действительно, для любой точки $x = \bar{x}^{(1)} + \varepsilon$ (где $\varepsilon = (\varepsilon, \varepsilon)$ и $0 < \varepsilon < 3/2$) имеем

$$\begin{aligned} f(x) &= \left(\bar{x}_1^{(1)} + \varepsilon\right)^3 + \left(\bar{x}_2^{(1)} + \varepsilon\right)^3 - 3\left(\bar{x}_1^{(1)} + \varepsilon\right)\left(\bar{x}_2^{(1)} + \varepsilon\right) = \\ &= \varepsilon^3 + \varepsilon^3 - 3\varepsilon\varepsilon = 2\varepsilon^3 - 3\varepsilon^2 = \varepsilon^2(2\varepsilon - 3) < 0 = f\big(\bar{x}^{(1)}\big). \end{aligned}$$

Отметим, что каждая точка минимума является стационарной.

Теорему 1 называют необходимым условием оптимальности *первого порядка*. Для выявления посторонних стационарных точек может использоваться необходимое условие оптимальности *второго порядка*:

Теорема 2. *Пусть функция $f(x)$ дважды дифференцируема в точке $\bar{x} \in R^n$. Если $\bar{x}$ — точка локального минимума, то матрица Гессе $f''(\bar{x})$ неотрицательно определена, т.е.*

$$\big(f''(\bar{x})\, h, h\big) \geq 0$$

при всех $h \in R^n$.

Доказательство. Поскольку $\bar{x}$ — точка локального минимума, то

$$f(\bar{x}) \leq f(\bar{x} + \alpha h)$$

для достаточно малых α. По определению дважды дифференцируемой функции имеем

$$\begin{aligned} 0 \leq f(\bar{x} + \alpha h) - f(\bar{x}) &= \big(f'(\bar{x}), \alpha h\big) + \frac{1}{2}\big(f''(\bar{x})\alpha h, \alpha h\big) + \\ &\quad + o\left(\|\alpha h\|_2^2\right). \end{aligned}$$

Поскольку $f'(\bar{x}) = 0$, то

$$0 \leq f(\bar{x} + \alpha h) - f(\bar{x}) = \frac{1}{2}\,\alpha^2\big(f''(\bar{x})h, h\big) + o\left(\alpha^2\right)$$

при всех достаточно малых α. Поделим обе части последнего неравенства на α^2 и перейдем к пределу при $\alpha \to 0$:

$$\frac{1}{2}\big(f''(\bar{x})h, h\big) \geq 0.$$

Следовательно, приходим к заключению: если $\bar{x}$ — точка локального минимума, то матрица $f''(\bar{x})$ неотрицательно определена. Теорема доказана.

Теперь сформулируем *достаточное* условие локальной оптимальности.

Теорема 3. *Пусть функция $f(x)$ дважды дифференцируема в точке $\bar{x} \in R^n$ и пусть $f'(\bar{x}) = 0$, а матрица $f''(\bar{x})$ положительно определена, т.е.*

$$\big(f''(\bar{x})\, h, h\big) > 0$$

при всех $h \in R^n$, $h \neq 0$. Тогда $\bar{x}$ — точка строгого локального минимума.

Доказательство. Наши рассуждения будем проводить от противного. Пусть в R^n существует такая последовательность $\{x^k\}$, что

$$x^k \neq \bar{x}, \quad x^k \to \bar{x}, \quad f(x^k) \leq f(\bar{x}).$$

Представим x^k в виде

$$x^k = \bar{x} + \alpha_k h^k, \quad \alpha_k = \left\| x^k - \bar{x} \right\|_2, \quad h^k = \frac{x^k - \bar{x}}{\alpha_k}.$$

Поскольку $||h^k||_2 = 1$ (т.е. множество векторов h^k ограничено), то из последовательности h^k можно выделить сходящуюся подпоследовательность. Для определенности будем считать, что это сама последовательность h^k, т.е. $h^k \to h \neq 0$. Из определения дважды дифференцируемой функции имеем

$$\begin{aligned}
0 \quad &\geq \quad f(x^k) - f(\bar{x}) = f(\bar{x} + \alpha_k h^k) - f(\bar{x}) = \\
&= \quad \big(f'(\bar{x}), \alpha_k h^k\big) + \frac{1}{2}\left(f''(\bar{x})\alpha_k h^k, \alpha_k h^k\right) + o\left(\left\|\alpha_k h^k\right\|_2^2\right) = \\
&= \quad \frac{1}{2}\,\alpha_k^2 \big(f''(\bar{x}) h^k, h^k\big) + o\left(\alpha_k^2\right).
\end{aligned}$$

Поделим обе части этого неравенства на α_k^2 и перейдем к пределу при $\alpha_k \to 0$:

$$0 \geq \big(f''(\bar{x}) h^k, h^k\big).$$

Полученное неравенство противоречит условию теоремы. Следовательно, $\bar{x}$ — точка строгого локального минимума. Теорема доказана.

Вернемся к анализу стационарных точек рассмотренного выше примера. Предварительно напомним формулировку *критерия Сильвестра*: симметричная матрица положительно (неотрицательно) определена тогда и только тогда, когда все ее ведущие миноры положительны (неотрицательны).

Матрица Гессе для функции $f(x) = x_1^3 + x_2^3 - 3x_1x_2$ имеет вид

$$f''(x) = \begin{pmatrix} 6x_1 & -3 \\ -3 & 6x_2 \end{pmatrix}.$$

Для ранее найденных стационарных точек $\bar{x}^{(1)} = (0, 0)$ и $\bar{x}^{(2)} = (1, 1)$ имеем

$$f''(\bar{x}^{(1)}) = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix}, \quad f''(\bar{x}^{(2)}) = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}.$$

По критерию Сильвестра матрица $f''(\bar{x}^{(1)})$ не является неотрицательно определенной, т.е. необходимое условие оптимальности второго порядка не выполняется. Таким образом, еще раз показано, что точка $\bar{x}^{(1)}$ не является точкой минимума.

Что же касается матрицы $f''(\bar{x}^{(2)})$, то по критерию Сильвестра эта матрица положительно определена. Это означает, что $\bar{x}^{(2)}$ — точка минимума по достаточному условию оптимальности.

2.2. Общие сведения о численных методах безусловной минимизации

Методы безусловной минимизации, использующие информацию только о значениях минимизируемой функции, называются методами *нулевого порядка*. Если при этом используются значения первых и вторых производных минимизируемой функции, то такие методы называют методами *первого* и *второго* порядков соответственно.

Алгоритм минимизации называют *последовательным*, если каждое следующее приближение к точке минимума строится через предыдущие приближения.

Для записи методов минимизации используются соотношения вида

$$x^{k+1} = x^k + \alpha_k h^k, \quad \alpha_k \in R, \quad k = 0, 1, 2, \dots \, .$$

Каждый конкретный алгоритм минимизации определяется заданием *начальной* точки x^0 (начального приближения к точке минимума), правилами выбора векторов h^k и чисел α_k, а также *критериями окончания счета.* Вектор h^k задает *направление* $(k+1)$-го шага алгоритма, а коэффициент α_k — *длину* этого шага.

Название метода минимизации определяется способом выбора векторов h^k, в то время как модификации метода связаны с различными способами выбора α^k. Термины *шаг метода* и *итерация метода* эквивалентны.

Если метод гарантирует получение точки минимума за *конечное* число шагов, то его называют *конечношаговым.* Такие методы удается построить для специальных типов задач (например, для задач линейного и квадратичного программирования). Если же достижение решения гарантируется лишь в пределе, то соответствующий метод называется *бесконечношаговым.*

Говорят, что метод

$$x^{k+1} = x^k + \alpha_k h^k$$

сходится, если $x^k \to \bar{x}$ при $k \to \infty$, где $\bar{x}$ — точка минимума функции $f(x)$. Если $f(x^k) \to f(\bar{x})$, то говорят, что метод сходится *по функции*, а последовательность $\left\{x^k\right\}$ называют *минимизирующей.* Отметим, что минимизирующая последовательность может не сходиться к точке минимума.

Говорят, что вектор h задает *направление убывания* функции $f(x)$ в точке x, если $f(x+\alpha h) < f(x)$ при всех достаточно малых $\alpha > 0$. Сам вектор h называют *направлением убывания.* Если при всех достаточно малых $\alpha > 0$ выполняется $f(x+\alpha h) > f(x)$, то вектор h называют *направлением возрастания.*

Сформулируем достаточный и необходимый признак направления убывания.

Теорема 4. *Пусть функция $f(x)$ дифференцируема в точке $x \in R^n$. Если вектор h удовлетворяет условию*

$$\big(f'(x), h\big) < 0,$$

то h — направление убывания функции $f(x)$ в точке x. Если h — направление убывания функции $f(x)$ в точке x, то выполняется неравенство

$$\big(f'(x), h\big) \le 0.$$

Доказательство. Пусть $\big(f'(x), h\big) < 0$. По определению дифференцируемой функции можно записать, что

$$\begin{aligned} f(x + \alpha h) - f(x) &= \big(f'(x), \alpha h\big) + o\big(\|\alpha h\|_2\big) = \\ &= \alpha \left(\big(f'(x), h\big) + \frac{o(\alpha)}{\alpha} \right). \end{aligned} \tag{6}$$

Поскольку $\big(f'(x), h\big) < 0$ по предположению теоремы, то начиная с некоторого достаточно малого значения α имеем неравенство $\big(f'(x), h\big) + o(\alpha)/\alpha < 0$, т.е. $f(x + \alpha h) - f(x) < 0$. Следовательно, h — направление убывания.

Вторую часть утверждения теоремы докажем от противного. Пусть h задает направление убывания в точке x, однако $\big(f'(x), h\big) > 0$. Тогда из (6) следует, что в действительности h является направлением возрастания. Полученное противоречие показывает, что должно быть выполнено неравенство $\big(f'(x), h\big) \le 0$, если h — направление убывания. Теорема доказана.

Метод $x^{k+1} = x^k + \alpha_k h^k$ называют *методом спуска*, если вектор h^k задает направление убывания функции $f(x)$ в точке x^k, а число α_k положительно и таково, что $f(x^{k+1}) < f(x^k)$.

Простейшим примером метода спуска является *градиентный метод*, в котором $h^k = -f'(x^k)$. Действительно, предположим, что $f'(x^k) \ne 0$. Тогда вектор $-f'(x^k)$ есть направление убывания в силу достаточного признака, поскольку

$$\big(f'(x^k), -f'(x^k)\big) = -\left\|f'(x^k)\right\|_2^2 < 0.$$

Напомним, что вектор $h^k = -f'(x^k)$ называют *антиградиентом.*

Теперь рассмотрим два подхода к выбору шага α_k по направлению убывания минимизируемой функции.

Первый из них называют *дроблением шага.* Пусть h^k — направление убывания. Выберем некоторые постоянные $\beta > 0$ и $0 < \lambda < 1$. Полагаем вначале $\alpha = \beta$ и проверим условие

$$f(x^k + \alpha h^k) < f(x^k). \tag{7}$$

Если это условие не выполняется, то осуществляем дробление шага $\alpha = \lambda\beta$ и вновь проверяем условие (7). Процесс дробление шага продолжаем до тех пор, пока условие (7) не окажется выполненным. Первое α, при котором это условие выполнено, принимается за α_k. Описанный процесс не может быть бесконечным, поскольку h^k — направление убывания.

Если при $\alpha = \beta$ условие (7) выполнено, то полезно увеличить шаг: $\alpha = \mu\beta$, $\mu > 1$. Если будет выполнено

$$f(x^k + \alpha h^k) < f(x^k + \beta h^k),$$

то текущее значение α опять умножается на μ и так до тех пор, пока значение функции не перестанет уменьшаться. Последнее α, при котором произошло уменьшение, берется в качестве α_k.

На практике часто выбирают $\lambda = 1/2$ и $\mu = 2$. Величину β относят к *параметрам управления* процессом минимизации и подбирают в зависимости от характера поведения минимизируемой функции вблизи x^k. Полезно также ограничить сверху увеличение шага.

Согласно второму подходу выбор длины шага по направлению убывания осуществляется из условия минимизации функции вдоль этого направления:

$$f(x^k + \alpha_k h^k) = \min_{\alpha} f(x^k + \alpha h^k) = \min_{\alpha} f(\alpha).$$

Для методов спуска минимум берется по $\alpha > 0$. Такой способ выбора α_k является наилучшим, поскольку при нем не только выполняется условие (7), но и обеспечивается достижение наименьшего значения $f(x)$ вдоль заданного направления убывания. Недостаток данного подхода состоит в том, что на каждом шаге требуется решение одномерной задачи минимизации, что приводит к дополнительному увеличению объема вычислений.

2.3. Скорость сходимости. Критерии окончания счета

Эффективность применяемого метода минимизации характеризуют при помощи понятия *скорости сходимости.*

Говорят, что метод сходится к точке минимума $\bar{x}$ *линейно* (с линейной скоростью, или со скоростью геометрической прогрессии), если существуют такие постоянные $q \in (0,1)$ и k_0, что

$$\left\|x^{k+1} - \bar{x}\right\| \le q \left\|x^k - \bar{x}\right\| \qquad k \ge k_0.$$

Скорость сходимости становится *сверхлинейной*, если

$$\left\|x^{k+1} - \bar{x}\right\| \le q_{k+1} \left\|x^k - \bar{x}\right\|, \quad q_k \to 0+ \qquad k \to \infty.$$

Говорят, что имеет место *квадратичная* скорость сходимости, если существуют такие постоянные $c \ge 0$ и k_0, что

$$\left\|x^{k+1} - \bar{x}\right\| \le c \left\|x^k - \bar{x}\right\|^2 \qquad k \ge k_0.$$

Иногда указанные неравенства заменяют на неравенства

$$\begin{aligned}
&\left\|x^{k+1} - \bar{x}\right\| \le c_1 q^{k+1}, \quad q \in (0,1), \quad k \ge k_0,\\
&\left\|x^{k+1} - \bar{x}\right\| \le c_2 q_{k+1} q_k \dots q_1, \quad q_k \to 0+,\\
&\left\|x^{k+1} - \bar{x}\right\| \le c_3 q^{2^{k+1}}, \quad q \in (0,1), \quad k \ge k_0.
\end{aligned}$$

Большинство теорем о сходимости методов минимизации доказываются в предположении *выпуклости* целевой функции, а скорость сходимости устанавливается в предположении ее *сильной* выпуклости. Для невыпуклых задач методы обычно позволяют отыскивать только локальные решения (точнее говоря, стационарные точки). Требования, которые накладываются в теоремах сходимости на минимизируемую функцию, называют *областью применимости* метода. Часть из них формулируют требования к начальному приближению.

На практике часто используют следующие критерии окончания счета:

$$\begin{aligned}
&\left\|x^{k+1}-x^k\right\| \le \varepsilon,\\
&\left\|f(x^{k+1})-f(x^k)\right\| \le \varepsilon,\\
&\left\|f'(x^{k+1})\right\| \le \varepsilon,
\end{aligned}$$

где ε — заданная абсолютная точность, с которой ищется точка минимума, а в качестве нормы может быть выбрана любая векторная норма. Как правило, требуют одновременного выполнения указанных критериев.

В тех случаях, когда желательно достижение относительной точности δ, используются такие критерии:

$$\begin{aligned}
&\left\|x^{k+1}-x^k\right\| \le \delta\left(1+\left\|x^{k+1}\right\|\right),\\
&\left\|f(x^{k+1})-f(x^k)\right\| \le \delta\left(1+\left\|f(x^{k+1})\right\|\right),\\
&\left\|f'(x^{k+1})\right\| \le \delta\left(1+\left\|f'(x^{k+1})\right\|\right).
\end{aligned}$$

Иногда применяют *комбинированные* критерии, объединяющие контроль по абсолютной и относительной погрешностям. В пользу такого подхода можно высказать следующие соображения. Рассмотрим неравенство

$$\left\|x^{k+1}-x^k\right\| \le \varepsilon+\delta\left\|x^k\right\|, \tag{8}$$

где x^{k+1} и x^k — два последовательных приближения к точке минимума.

Если задана только допустимая абсолютная погрешность ε (т.е. $\delta = 0$), то тем самым фиксируется разряд приближенных значений координат точки минимума, соответствующий требуемым самым младшим верным цифрам этих значений. Однако если задать абсолютную погрешность без учета величины порядка искомого минимума и длины разрядной сетки используемой вычислительной машины, то контроль точности вычислений по абсолютной погрешности может оказаться невозможным. Например, если вычисления проводятся с семью десятичными разрядами и искомый минимум (для одномерного случая) равен 55555.55, то задание абсолютной погрешности, равной 10^{-4} окажется бессмысленным и приведет к зацикливанию итерационного процесса. Поэтому если мы хотим, чтобы четвертый разряд приближенного значения минимума соответствовал самой младшей верной цифре, то в данном примере мы должны положить абсолютную погрешность равной 10. Такое задание абсолютной погрешности в отрыве от величины порядка искомого минимума и количества разрядов, с которыми проводятся вычисления, может показаться нелепым, поскольку обычно абсолютная погрешность используется для задания количества верных цифр после точки, отделяющей целую часть от дробной.

Таким образом, чтобы разумно задать абсолютную погрешность вычислений, нужно предварительно знать величину порядка нормы решения и учитывать величину нормы начального приближения.

Если задана только допустимая относительная погрешность δ (т.е. $\varepsilon = 0$), то тем самым фиксируется общее требуемое количество верных цифр в приближенных значениях координат точки минимума. Однако если искомый минимум мал и значение $\left\|x^k\right\|$ становится слишком близким к нулю, то даже при разумном задании δ неравенство (8) может никогда не достигаться или же при вычислении $\delta\left\|x^k\right\|$ может произойти образование машинного нуля (потеря значимости).

Поясним на примере одномерной минимизации, почему это неравенство может никогда не достигаться, даже если в машинном представлении произведение $\delta|x^k|$ не равно нулю и итерационный процесс гарантированно сходится. Для этого напомним фундаментальное свойство систем представления чисел с плавающей точкой: расстояние между числом x и соседним по отношению к нему числом не меньше $\mathrm{masheps}\cdot|x|/\beta$ и не больше $\mathrm{masheps}\cdot|x|$, если только само число x или соседнее число не равны нулю. Здесь β — основание системы счисления машины, а машинно-зависимый параметр masheps (называемый машинным эпсилон) характеризует относительную точность машинной арифметики.

Таким образом, если $\delta|x^k|$ окажется меньше $\mathrm{masheps}\cdot|x|/\beta$, то неравенство (8) при $\varepsilon = 0$ никогда не будет достигаться, а основанный на нем итерационный процесс никогда не завершится. Если мы хотим, чтобы $|x^{k+1}|$ и $|x^k|$ стали максимально близкими друг к другу, т.е. стали соседними числами, то критерий точности должен быть таким:

$$\left|x^{k+1} - x^{k}\right| \leq \text{masheps} \cdot \max\left(\left|x^{k+1}\right|, \left|x^{k}\right|\right).$$

Конечно, данный критерий неприменим для небольшой окрестности нуля, в которой происходит образование машинного нуля при вычислении правой части этого неравенства. Отметим, что расстояние от нуля до правого (левого) соседнего числа не связано с машинным эпсилон и представляет собой самостоятельный машинно-зависимый параметр.

Часто применяют следующие две модификации рассмотренного комбинированного критерия:

$$\left\|x^{k+1} - x^{k}\right\| \leq \max\left\{\varepsilon, \delta\left\|x^{k}\right\|\right\}$$

или

$$\left\|x^{k+1} - x^{k}\right\| \leq \begin{cases} \varepsilon, & \text{если } \left\|x^{k}\right\| \leq 1; \\ \delta\left\|x^{k}\right\|, & \text{если } \left\|x^{k}\right\| > 1. \end{cases}$$

Таким образом, применение критерия типа (8) или его модификаций позволяет избегать тех тупиковых ситуаций, которые могут возникнуть, если задавать только абсолютную или только относительную погрешности, и дает возможность задавать требуемое количество верных знаков в приближенном решении, не заботясь о величине его порядка.

2.4. Выпуклые множества и выпуклые функции

Пусть R^n — n-мерное евклидово пространство вещественных векторов $x = (x_1, x_2, \ldots, x_n)^T$. Множество $X \in R^n$ называется *выпуклым*, если вместе с любыми двумя точками $x^{(1)}$ и $x^{(2)}$ оно содержит и отрезок, соединяющий эти точки; это означает, что

$$\lambda x^{(1)} + (1-\lambda)\, x^{(2)} \in X, \quad \lambda \in [0, 1].$$

На числовой прямой R^1 выпуклыми множествами являются всевозможные промежутки (сама прямая, отрезки, интервалы, полупрямые).

Функция $f(x)$, определенная на некотором выпуклом множестве $X \in R^n$, называется выпуклой на X, если выполнено неравенство

$$f\big(\lambda x^{(1)} + (1-\lambda)\, x^{(2)}\big) \leq \lambda f(x^{(1)}) + (1-\lambda)\, f(x^{(2)})$$

при всех $x^{(1)}, x^{(2)} \in X$, $\lambda \in [0,1]$. Если это неравенство строгое, то $f(x)$ называют строго выпуклой функцией на X. Функция $f(x)$ называется *вогнутой*, если $-f(x)$ выпукла. Геометрически выпуклость означает, что любая хорда графика $f(x)$ располагается выше кривой $f(x)$.

Задача минимизации (оптимизации) называется выпуклой, если X — выпуклое множество, а $f(x)$ — выпуклая на X функция.

Теорема 5. *Если задача минимизации выпукла, то любое ее локальное решение является также глобальным.*

Доказательство. Пусть $\bar{x}$ — точка локального минимума, т.е. при некотором $\varepsilon > 0$ имеем $f(\bar{x}) \le f(x)$ для всех $x \in X \bigcap U_\varepsilon(\bar{x})$, где $U_\varepsilon(\bar{x}) = \{x \in R^n \mid \ \|x - \bar{x}\| \le \varepsilon\}$ — шар радиуса ε с центром в точке $\bar{x}$.

Для любого $x \in X$, $x \ne \bar{x}$, положим $\lambda = \min\big\{\varepsilon/\big(\|x-\bar{x}\|\big), 1\big\}$. Тогда $\lambda x + (1-\lambda)\,\bar{x} \in X \bigcap U_\varepsilon(\bar{x})$. Действительно, имеет место неравенство

$$\|\lambda x + (1-\lambda)\,\bar{x} - \bar{x}\| = \lambda\, \|x - \bar{x}\| \le \varepsilon.$$

Следовательно, в силу выпуклости $f(x)$ имеем

$$f(\bar{x}) \le f\big(\lambda x + (1-\lambda)\,\bar{x}\big) \le \lambda f(x) + (1-\lambda)\, f(\bar{x}).$$

Отсюда заключаем, что $f(\bar{x}) \le f(x)$. Теорема доказана.

Для выпуклых задач необходимые условия оптимальности являются также и достаточными.

Теорема 6. *Пусть функция $f(x)$ — выпукла на X и дифференцируема в точке $\bar{x} \in X$. Если $f'(\bar{x}) = 0$, то $\bar{x}$ — точка минимума $f(x)$ на X.*

Доказательство. В силу выпуклости $f(x)$ имеем

$$f\big(\lambda x + (1-\lambda)\,\bar{x}\big) \le \lambda f(x) + (1-\lambda)\, f(\bar{x}), \quad \lambda \in [0,1].$$

Отсюда

$$f(x) - f(\bar{x}) \ge \frac{f\big(\bar{x} + \lambda\,(x-\bar{x})\big) - f(\bar{x})}{\lambda}.$$

Разложим $f\big(\bar{x} + \lambda\,(x-\bar{x})\big)$ в ряд Тейлора:

$$f(x) - f(\bar{x}) \ge \frac{\big(f'(\bar{x}), \lambda\,(x-\bar{x})\big) + o\,(\lambda\,\|x-\bar{x}\|_2)}{\lambda} = \frac{o(\lambda)}{\lambda}.$$

После предельного перехода при $\lambda \to 0$ получим $f(x) - f(\bar{x}) \geq 0$. Отсюда $f(x) \leq f(\bar{x})$. Теорема доказана.

Из этой теоремы следует, что для выпуклых задач оптимизации отыскание стационарной точки означает отыскание точки глобального минимума.

Для выявления выпуклости функции можно воспользоваться следующим критерием: если функция $f(x)$ дважды дифференцируема на выпуклом множестве $X \subset R^n$ и матрица ее вторых производных $f''(x)$ положительно определена при всех $x \in X$, то $f(x)$ является выпуклой функцией на множестве X.

Если к матрице $f''(x)$ применить критерий Сильвестра, то критерий выпуклости формулируется так: если все ведущие миноры матрицы $f''(x)$ положительны при всех $x \in X$, то функция $f(x)$ выпукла на множестве X.

Укажем еще одно полезное свойство выпуклых задач.

Теорема 7. *Пусть рассматривается выпуклая задача оптимизации. Тогда множество ее решений* $X^* = \{\bar{x}\}$ *выпукло. Если при этом функция* $f(x)$ *строго выпукла на* X*, то решение задачи единственно, т.е. множество* X^* *состоит из одной точки.*

Доказательство. Пусть x_1 и x_2 принадлежат X^* и $\lambda \in [0,1]$. Тогда $f(x_1) = f(x_2) = f(\bar{x})$. В силу выпуклости функции $f(x)$ имеем:

$$f\big(\lambda x_1 + (1-\lambda)\, x_2\big) \leq \lambda f(x_1) + (1-\lambda)\, f(x_2) = f(\bar{x}).$$

Поскольку $f(\bar{x})$ — минимальное значение $f(x)$ на X, то это неравенство может выполняться только как равенство. Следовательно, $\lambda x_1 + (1-\lambda)\, x_2$ — точка минимума. Значит, по определению, множество X^* выпукло.

Пусть теперь функция $f(x)$ строго выпукла. Если предположить, что в X^* существуют две различные точки x_1 и x_2, то при $\lambda \in [0,1]$ приведенное выше неравенство должно быть строгим, что невозможно, поскольку $f(\bar{x})$ — минимальное значение $f(x)$ на X. Теорема доказана.

2.5. Квадратичные функции

Во многих задачах оптимизации рассматриваются *квадратичные* функции, т.е. функции вида

$$f(x) = \sum_{i,j=1}^{n} c_{ij} x_i x_j + \sum_{j=1}^{n} b_j x_j.$$

Положим $a_{ij} = c_{ij} + c_{ji}$. Тогда матрица $A = (a_{ij})$ будет симметричной. С ее помощью квадратичную функцию можно представить в виде

$$f(x) = \frac{1}{2}\,(Ax, x) + (b, x),$$

где $x = (x_1, \ldots, x_n)^T$ и $b = (b_1, \ldots, b_n)^T$.

Градиент и матрица Гессе квадратичной функции представляются следующим образом:

$$\operatorname{grad} f(x) = f'(x) = Ax + b, \quad f''(x) = A.$$

Чтобы квадратичная функция была выпуклой на R^n, достаточно, чтобы матрица A была положительно определена.

В случае минимизации выпуклой квадратичной функции выбор шага α_k на $(k+1)$-й итерации по направлению убывания может быть осуществлен из следующих соображений. Запишем

$$\begin{aligned}
p(\alpha) &= f\big(x^k + \alpha h^k\big) = \\
&= \frac{1}{2}\Big(A\big(x^k + \alpha h^k\big), x^k + \alpha h^k\Big) + \Big(b, x^k + \alpha h^k\Big) = \\
&= \frac{1}{2}\Big(\big(Ax^k, x^k\big) + \alpha\,\big(Ax^k, h^k\big) + \alpha\,\big(Ah^k, x^k\big) + \\
&\quad + \alpha^2\big(Ah^k, h^k\big)\Big) + \big(b, x^k\big) + \alpha\,\big(b, h^k\big) = \\
&= \frac{1}{2}\,\big(Ah^k, h^k\big)\,\alpha^2 + \big(Ax^k + b, h^k\big)\,\alpha + \left(\frac{1}{2}\,Ax^k + b, x^k\right).
\end{aligned}$$

Здесь мы воспользовались равенством $\big(Ah^k, x^k\big) = \big(Ax^k, h^k\big)$, поскольку A — симметричная матрица.

Итак, мы выписали квадратный трехчлен $p(\alpha)$. Его минимум достигается при том значении α, которое может быть получено из уравнения $p'(\alpha) = 0$:

$$\big(Ah^k, h^k\big)\,\alpha + \big(Ax^k + b, h^k\big) = 0.$$

Отсюда получаем, что

$$\alpha_k = -\frac{\left(Ax^k + b, h^k\right)}{\left(Ah^k, h^k\right)}.$$

Полученное значение α_k неотрицательно, поскольку числитель не положителен по признаку убывания, а знаменатель строго больше нуля в силу положительной определенности матрицы A.

Если квадратичная функция выпукла, то точку минимума можно также найти из уравнения $f'(x) = Ax + b = 0$, т.е. решая систему линейных алгебраических уравнений с симметричной положительно определенной матрицей.

2.6. Градиентные методы

Рассмотрим методы безусловной минимизации, основанные на идее замены минимизируемой функции $f(x)$ в окрестности очередного приближения x^k первым членом (линейной частью) ее разложения в ряд Тейлора. Такие методы называют *градиентными*, поскольку при вычислении x^{k+1} используются производные функции $f(x)$ первого порядка.

Градиентные методы относятся к классу методов спуска, в которых два последовательных приближения к точке минимума связаны соотношением

$$x^{k+1} = x^k + \alpha_k h^k,$$

где h^k — направление убывания функции $f(x)$ в точке x^k и α_k — длина шага по направлению убывания h^k. Вектор h^k берется равным антиградиенту функции $f(x)$ в точке x^k, т.е. $h^k = -f'(x^k)$:

$$x^{k+1} = x^k - \alpha_k f'(x^k), \quad \alpha_k > 0, \quad k = 0, 1, 2, \dots \,. \tag{9}$$

В пользу такого выбора направления убывания могут быть высказаны следующие соображения. В предположении, что функция $f(x)$ дифференцируема на R^n, рассмотрим линейную часть приращения $f(x) - f(x^k)$:

$$\begin{aligned} f(x) &= f\big(x^k + (x - x^k)\big) = \\ &= f(x^k) + \big(f'(x^k), x - x^k\big) + o\left(\left\|x - x^k\right\|_2\right). \end{aligned} \tag{10}$$

Все возможные направления перемещений от точки x^k с конечным шагом α образуют шар X радиуса α с центром в точке x^k: $X = \{x : \|x - x^k\|_2 \le \alpha\}$. Наша цель — найти такое направление убывания, при котором на границе этого шара выполнялись условия, чтобы $f(x) < f(x^k)$ и чтобы разность $f(x^k) - f(x)$ при этом была наибольшей (т.е. чтобы при фиксированной длине шага по искомому направлению достигалось наименьшее значение $f(x)$).

Из (10) можно заключить, что эта разность будет наибольшей, если мы минимизируем по x на шаре X линейную часть приращения $f(x) - f(x^k)$, равную $\big(f'(x^k), x - x^k\big)$. Воспользовавшись неравенством Коши–Буняковского, запишем

$$\big(f'(x^k), x - x^k\big) \ge -\left\|f'(x^k)\right\|_2 \left\|x - x^k\right\|_2 \ge -\alpha \left\|f'(x^k)\right\|_2 .$$

Легко видеть, что нижняя грань последнего неравенства достигается при

$$x = x^{k+1} = x^k - \frac{\alpha\, f'(x^k)}{\|f'(x^k)\|_2} \in X.$$

Таким образом, приходим к выводу, что при фиксированной длине шага α минимум линейной части разложения функции $f(x)$ в ряд Тейлора в окрестности точки x^k достигается, если направление вектора $h = x^{k+1} - x^k$ совпадает с направлением антиградиента $-f'(x^k)$. Это означает, что направление антиградиента является самым выгодным из всех направлений убывания.

Для квадратичной функции градиентный метод (9) принимает вид

$$x^{k+1} = x^k - \alpha_k\big(Ax^k + b\big).$$

В численных расчетах шаг α_k по направлению убывания может быть получен методом дробления шага, рассмотренном в п. 2.2. Если же α_k выбирается при помощи одномерной минимизации функции $f(x^k + \alpha h^k)$ вдоль антиградиента, то такая модификация градиентного метода называется методом *наискорейшего спуска*, при котором достигается максимальное уменьшение функции $f(x)$ вдоль направления ее антиградиента. Для квадратичных функций соответствующее значение α_k приведено в п. 2.5.

Градиентный метод сходится к точке минимума линейно, т.е. со скоростью геометрической прогрессии. Если на текущем шаге итераций наименьшее и наибольшее собственные значения матрицы Гессе мало отличаются друг от друга, то знаменатель прогрессии уменьшается, а скорость сходимости увеличивается. Если же эти собственные значения значительно отличаются, то направление антиградиента может сильно отклоняться от направления в точку минимума; из-за этого движение к минимуму приобретает зигзагообразный характер и сходимость замедляется.

Чувствительность градиентного метода минимизации к погрешностям вычислений повышается в окрестности точки минимума, когда норма градиента мала. Поэтому градиентный метод и его модификации лучше использовать в начальной стадии поиска минимума, чем на его заключительном этапе.

2.7. Метод Ньютона многомерной минимизации

Если в окрестности очередного приближения x^k мы разложим минимизируемую функцию $f(x)$ в ряд Тейлора и возьмем квадратичную часть этого разложения, то получим метод второго порядка (метод Ньютона), который использует информацию о вторых производных функции $f(x)$. Этот метод применяется для безусловной минимизации выпуклых дважды дифференцируемых функций и при определенных условиях обеспечивает более быструю, нежели градиентный метод и его модификации, скорость сходимости.

Пусть функция $f(x)$ выпукла и дважды дифференцируема на R^n, причем матрица $f''(x)$ не выраждена на R^n. Исходя из определения дважды дифференцируемой функции, можно выписать следующее разложение для $f(x)$ в окрестности точки x^k:

$$\begin{aligned} f(x) - f(x^k) &= \left(f'(x^k), x - x^k\right) + \\ &+ \frac{1}{2}\left(f''(x^k)(x - x^k), x - x^k\right) + \\ &+ o\left(\left\|x - x^k\right\|_2^2\right). \end{aligned}$$

Обозначим квадратичную часть приращения $f(x)-f(x^k)$ через

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

Найдем точку x^{k+1}, в которой достигается минимум функции $f_k(x)$. По предположению функция $f(x)$ выпукла; значит, матрица $f''(x)$ положительно определена. Поскольку $f_k''(x) = f''(x^k)$, то $f_k''(x)$ — также положительно определенная матрица. Следовательно, функция $f_k(x)$ выпукла в силу необходимого и достаточного условия выпуклости. Отсюда заключаем, что по теоремам 5 и 6 необходимое и достаточное условие ее минимума имеет вид

$$f_k'(x) = f'(x^k) + f''(x^k)\bigl(x - x^k\bigr) = 0.$$

Теперь решим полученную систему линейных уравнений, получим точку минимума функции $f_k(x)$ и возьмем ее в качестве очередного приближения x^{k+1} к точке минимума исходной функции $f(x)$:

$$x^{k+1} = x^k - \bigl[f''(x^k)\bigr]^{-1} f'(x^k). \tag{11}$$

Здесь $\bigl[f''(x^k)\bigr]^{-1}$ — матрица, обратная к матрице вторых производных $f''(x^k)$. Выписанное соотношение называют методом Ньютона.

При достаточно хорошем приближении метод (11) имеет квадратичную скорость сходимости. Поэтому его удобно применять на завершающем этапе минимизации при уточнении приближения к точке минимума, найденного каким-либо другим, менее трудоемким способом. Если начальное приближение выбрано неудачно, то сходимость отсутствует. Указанный недостаток устраняется, если применить следующую модификацию метода Ньютона, называемую *модифицированным* методом Ньютона (или методом Ньютона с регулировкой шага):

$$x^{k+1} = x^k - \alpha_k\bigl[f''(x^k)\bigr]^{-1} f'(x^k), \quad \alpha_k > 0. \tag{12}$$

При $\alpha_k = 1$ итерационный метод (12) совпадает с классическим методом (11). Легко видеть, что эти методы относятся к классу методов спуска $x^{k+1} = x^k + \alpha_k h^k$, где вектор направления убывания h^k находится из решения линейной системы $f''(x^k)h^k = -f'(x^k)$. Отсюда следует, что в практических расчетах на каждой итерации нет необходимости обращать матрицу $f''(x^k)$: достаточно решить указанную линейную систему. Выбор шага α_k по направлению убывания можно осуществлять либо методом дробления шага, рассмотренном в п. 2.2., либо при помощи одномерной минимизации функции $f(x^k + \alpha h^k)$ вдоль направления убывания.

Может быть показано, что модифицированный метод Ньютона (12) сходится при любом начальном приближении $x^0 \in R^n$, причем скорость сходимости будет сверхлинейной или квадратичной в зависимости от свойств функции $f(x)$. Таким образом, с помощью регулировки шага по направлению убывания преодолевается недостаток метода (11), связанный с необходимостью выбора хорошего начального приближения.

Если по каким-либо причинам сложно вычислять матрицу $f''(x^k)$, то можно строить ее аппроксимации при помощи формул численного дифференцирования. Построенные при таком подходе методы называют *квазиньютоновскими*. Остановимся на этом вопросе подробнее.

Поскольку матрица $f''(x^k)$ содержит частные производные второго порядка, то достаточно рассмотреть случай функции двух переменных $f(x, y)$. Для аппроксимации производных $\dfrac{\partial^2 f}{\partial x^2}$ и $\dfrac{\partial^2 f}{\partial y^2}$ воспользуемся известными соотношениями

$$\frac{\partial^2 f}{\partial x^2} = \frac{f(x-h, y) - 2f(x, y) + f(x+h, y)}{h^2} + O\left(h^2\right),$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{f(x, y-h) - 2f(x, y) + f(x, y+h)}{h^2} + O\left(h^2\right).$$

Здесь h — малый параметр, определяющий погрешность выписанных формул численного дифференцирования.

Теперь выведем разностное соотношение для аппроксимации смешанной производной $\dfrac{\partial^2 f}{\partial x \partial y}$. Для произвольной достаточно гладкой функции $g(x,y)$ введем в рассмотрение разностные операторы

$$g_x = \frac{g(x+h,y)-g(x,y)}{h}, \quad g_{\bar{x}} = \frac{g(x,y)-g(x-h,y)}{h},$$
$$g_y = \frac{g(x,y+h)-g(x,y)}{h}, \quad g_{\bar{y}} = \frac{g(x,y)-g(x,y-h)}{h}.$$

Имеем
$$g_x = \frac{\partial g}{\partial x} + \frac{h}{2}\frac{\partial^2 g}{\partial x^2} + O\left(h^2\right),$$
$$f_{\bar{y}} = \frac{\partial f}{\partial y} - \frac{h}{2}\frac{\partial^2 f}{\partial y^2} + O\left(h^2\right).$$

Используя эти разложения для $g = f_{\bar{y}}$, получим

$$\left(f_{\bar{y}}\right)_x = \frac{\partial^2 f}{\partial x \partial y} + \frac{h}{2}\frac{\partial^3 f}{\partial^2 x \partial y} - \frac{h}{2}\frac{\partial^3 f}{\partial x \partial^2 y} + O\left(h^2\right).$$

Аналогично можно получить, что

$$\left(f_y\right)_{\bar{x}} = \frac{\partial^2 f}{\partial x \partial y} - \frac{h}{2}\frac{\partial^3 f}{\partial^2 x \partial y} + \frac{h}{2}\frac{\partial^3 f}{\partial x \partial^2 y} + O\left(h^2\right).$$

Сложив два последних соотношения, получаем, что

$$\frac{1}{2}\left(\left(f_{\bar{y}}\right)_x + \left(f_y\right)_{\bar{x}}\right) = \frac{\partial^2 f}{\partial x \partial y} + O\left(h^2\right),$$

где

$$\left(f_{\bar{y}}\right)_x = \frac{f(x+h,y)-f(x,y)-f(x+h,y-h)+f(x,y-h)}{h^2},$$
$$\left(f_y\right)_{\bar{x}} = \frac{f(x,y+h)-f(x,y)-f(x-h,y+h)+f(x-h,y)}{h^2}.$$

Эти разностные соотношения получаются последовательным применением приведенных выше разностных операторов.

Для квадратичной функции

$$f(x) = \frac{1}{2}(Ax,x) + (b,x), \quad f'(x) = Ax + b, \quad f''(x) = A$$

метод (11) примет вид

$$x^{k+1} = x^k - A^{-1}\big(Ax^k + b\big),$$

т.е. при любом начальном приближении точное решение достигается за одну итерацию. Если применяется метод (12) с регулировкой шага при помощи одномерной минимизации вдоль направления убывания, то α_k для квадратичной функции выражается явно (см. п. 2.5).

3. Методы одномерной минимизации

Рассмотрим некоторые методы поиска минимума функций одного вещественного переменного, которые могут быть использованы для выбора шага вдоль направления убывания в задачах многомерной минимизации.

3.1. Унимодальные функции

Пусть $f(x)$ — функция одного вещественного переменного x из некоторого множества $X \subset R^1$. Отметим, что в общем случае минимум функции $f(x)$ на множестве X может и не существовать. Тогда говорят о точной *нижней грани* функции $f(x)$, что является обобщением понятия минимума.

Пусть функция $f(x)$ ограничена снизу на множестве X (это означает, что $f(x) \geq A > -\infty$ для всех $x \in X$). Число f^* называется нижней гранью функции $f(x)$ на множестве X: $f^* = \inf\limits_{x \in X} f(x)$, если $f(x) \geq f^*$ при всех $x \in X$ и для любого $\varepsilon > 0$ существует такая точка x_ε, что $f(x_\varepsilon) < f^* + \varepsilon$. Если $f(x)$ не ограничена снизу, то полагают $f^* = -\infty$.

Рассмотрим пример, когда функция $f(x)$ не имеет минимума на множестве X.

Возьмем $f(x) = 1/x$ на $X = [1, \infty)$ и предположим, что $\bar{x}$ — одна из возможных точек минимума функции $f(x)$ на X. Для произвольной точки x, такой, что $x > \bar{x}$, $x \in X$, имеем $f(\bar{x}) = 1/\bar{x} > 1/x = f(x)$, т.е. $\bar{x}$ не является точкой минимума. Получили противоречие. Покажем теперь, что $f^* = \inf\limits_{X} f(x) = 0$. Действительно, для произвольного $x \in X$ справедливо неравенство

$f(x) = 1/x > 0$. Возьмем произвольное число $\varepsilon > 0$ и произвольную точку $x_\varepsilon > \max(1/\varepsilon, 1)$. Тогда $x_\varepsilon \in X$ и $f(x_\varepsilon) < \varepsilon = 0 + \varepsilon$. Следовательно, $f^* = 0$.

Если функция $f(x)$ имеет на X несколько локальных минимумов, то их поиск затруднен. Поэтому многие методы минимизации применимы тогда, когда на X существует один локальный минимум, который является на X одновременно и глобальным. Если локальных минимумов несколько, то эти методы дают сходимость к одному из них.

Выделим класс унимодальных функций, у которых на X любой локальный минимум является глобальным.

Определение. Функция $f(x)$ называется *унимодальной* на отрезке $[a, b]$, если она непрерывна и существуют такие числа α и β, $a \le \alpha \le \beta \le b$, что

1) если $a < \alpha$, то $f(x)$ монотонно убывает на $[a, \alpha]$;

2) если $\beta < b$, то $f(x)$ монотонно возрастает на $[\beta, b]$;

3) при $x \in [\alpha, \beta]$ имеет место равенство

$$f(x) = f^* = \min_{[a,b]} f(x).$$

Отметим, что возможно вырождение в точку одного или двух отрезков $[a, \alpha]$, $[\alpha, \beta]$, $[\beta, b]$. Для проверки унимодальности могут быть применены следующие критерии:

1) функция $f(x)$ дифференцируема на отрезке $[a, b]$ и первая производная $f'(x)$ не убывает на этом отрезке;

2) функция $f(x)$ дважды дифференцируема на отрезке $[a, b]$ и $f''(x) \ge 0$.

3.2. Прямые методы одномерной минимизации

Отдельную группу численных методов поиска минимума функций одного вещественного переменного составляют *прямые* методы, основанные на вычислении только значений минимизируемой функции в некоторых точках множества X и не использующих значений ее производных.

3.2.1. Метод перебора

Метод перебора является простейшим из прямых методов минимизации.

Пусть $f(x)$ — унимодальная функция на отрезке $[a, b]$. Требуется найти какую-либо из ее точек минимума на этом отрезке с абсолютной точностью $\varepsilon > 0$. Разобьем отрезок $[a, b]$ на n равных частей равномерной сеткой

$$x_j = a + jh \quad (j = 0, 1, \ldots, n)$$

с шагом h, где $h = \dfrac{b-a}{n}$ и n — целое число, такое, что $n \geq \dfrac{b-a}{\varepsilon}$. Таким образом, длина каждого подотрезка не превосходит ε. По вычисленным значениям $f(x_j)$ найдем точку x_m, для которой

$$f(x_m) = \min_{0 \leq j \leq n} f(x_j).$$

После этого полагаем $\bar{x} \approx x_m$ и $f(\bar{x}) \approx f(x_m)$. При этом максимальная погрешность ε_n определения точки минимума $\bar{x}$ равна $\varepsilon_n = \dfrac{b-a}{n} \leq \varepsilon$.

Описанный метод, предусматривающий предварительное задание достаточно мелкой равномерной сетки $\{x_j\}$, называют также пассивной стратегией поиска точки минимума.

Такую стратегию применяют в тех случаях, когда удобно единовременно получить $n+1$ значение минимизируемой функции $f(x)$ (например, в результате эксперимента или с использованием многопроцессорных вычислительных систем), а получение значений этой функции в других точках затруднено или же вовсе невозможно.

3.2.2. Метод деления отрезка пополам

Более эффективны методы, использующие уже имеющиеся значения функции $f(x)$ для определения очередного приближения к точке минимума. Такие методы называются *последовательными.*

Простейшим последовательным методом является метод деления отрезка пополам. Сущность этого метода состоит в том, что строится последовательность вложенных подотрезков, каждый из которых содержит в себе хотя бы одну из точек минимума.

Пусть требуется вычислить точку минимума функции $f(x)$ с абсолютной точностью $\varepsilon > 0$ на отрезке $[a, b]$. Построим следующие рекуррентные соотношения:

$$
\begin{aligned}
&n = 1; \quad a_{n-1} = a; \quad b_{n-1} = b;\\
&x_1^{(n-1)} = \frac{a_{n-1} + b_{n-1}}{2} - \frac{\varepsilon}{2}; \;\; x_2^{(n-1)} = \frac{a_{n-1} + b_{n-1}}{2} + \frac{\varepsilon}{2};\\
&a_n = a_{n-1}, \;\; b_n = x_2^{(n-1)}, \qquad f\bigl(x_1^{(n-1)}\bigr) \le f\bigl(x_2^{(n-1)}\bigr);\\
&a_n = x_1^{(n-1)}, \;\; b_n = b_{n-1}, \qquad f\bigl(x_1^{(n-1)}\bigr) > f\bigl(x_2^{(n-1)}\bigr).
\end{aligned}
\tag{13}
$$

Теперь рассмотрим критерий окончания счета и исследуем скорость сходимости этого процесса.

Длина $l_n = b_n - a_n$ текущего подотрезка $[a_n, b_n]$ удовлетворяет соотношению

$$l_n = \frac{l_{n-1}}{2} + \frac{\varepsilon}{2}, \quad n = 1, 2, \dots \, .$$

Выразим l_n через длину исходного отрезка $[a, b]$:

$$
\begin{aligned}
l_n \;&=\; \frac{l_{n-1}}{2} + \frac{\varepsilon}{2} = \frac{l_{n-2}}{2^2} + \frac{\varepsilon}{2^2} + \frac{\varepsilon}{2} = \frac{l_{n-3}}{2^3} + \frac{\varepsilon}{2^3} + \frac{\varepsilon}{2^2} + \frac{\varepsilon}{2} = \dots =\\
&=\; \frac{l_0}{2^n} + \frac{\varepsilon}{2^n} + \frac{\varepsilon}{2^{n-1}} + \dots + \frac{\varepsilon}{2^2} + \frac{\varepsilon}{2} =\\
&=\; \frac{b-a}{2^n} + \frac{\varepsilon}{2}\left(1 + \frac{1}{2} + \dots + \frac{1}{2^{n-1}}\right) =\\
&=\; \frac{b-a}{2^n} + \varepsilon\left(1 - \frac{1}{2^n}\right) = \frac{b-a-\varepsilon}{2^n} + \varepsilon.
\end{aligned}
$$

Поскольку минимум функции $f(x)$ заведомо находится на отрезке $[a_n, b_n]$, то за искомую точку минимума можно взять $\bar{x} \approx x_n = (b_n + a_n)/2$. Тогда абсолютная погрешность отклонения от точки минимума не превосходит

$$\varepsilon_n = \frac{l_n}{2} = \frac{b-a-\varepsilon}{2^{n+1}} + \frac{\varepsilon}{2}. \tag{14}$$

Процесс (13) заканчивается, когда выполняется неравенство

$$\frac{b-a-\varepsilon}{2^{n+1}} < \frac{\varepsilon}{2},$$

так как верхняя граница ε_n абсолютной погрешности становится меньше ε.

Поскольку длина каждого последующего подотрезка примерно вдвое меньше длины предыдущего, то процесс (13) сходится со скоростью геометрической прогрессии со знаменателем, почти равным $1/2$.

Если функция $f(x)$ не унимодальна, то сходимость будет к одному из локальных минимумов, не обязательно наименьшему (глобальному). Видно, что изложенный метод деления отрезка пополам применим и к недифференцируемым функциям.

Так как для завершения процесса (13) должно быть выполнено неравенство

$$|\bar{x} - x_n| \le \varepsilon_n = \frac{b-a-\varepsilon}{2^{n+1}} + \frac{\varepsilon}{2} < \varepsilon,$$

то отсюда может быть вычислено значение n, для которого заданная точность ε будет достигнута:

$$\frac{b-a-\varepsilon}{\varepsilon} < 2^n; \quad n > \log_2\left(\frac{b-a-\varepsilon}{\varepsilon}\right).$$

Этим критерием окончания счета пользуются тогда, когда требуется найти точку минимума с относительной точностью ε.

При построении процесса (13) точки $x_1^{(n-1)}$ и $x_2^{(n-1)}$ можно выбирать другим способом:

$$x_1^{(n-1)} = \frac{a_{n-1}+b_{n-1}}{2} - \frac{\delta}{2}, \quad x_2^{(n-1)} = \frac{a_{n-1}+b_{n-1}}{2} + \frac{\delta}{2},$$

где $0 < \delta < \varepsilon$ — постоянная, являющаяся параметром метода (параметром управления процесса). Величина δ может выбираться так, чтобы увеличить скорость сходимости рассматриваемого метода и приблизить ее к скорости сходимости метода бисекций для нахождения нулей функций. Чем меньше δ, тем ближе точки $x_1^{(n-1)}$ и $x_2^{(n-1)}$ к середине подотрезка $[a_{n-1}, b_{n-1}]$. Однако этот выбор ограничен количеством верных десятичных знаков m, с которыми задаются значения аргумента x для вычисления $f(x)$: $10^{-m} < \delta$.

Если в процесс (13) ввести параметр δ, то соотношение (14) для абсолютной погрешности, с которой определена точка минимума, примет вид

$$\varepsilon_n = \frac{b-a-\delta}{2^{n+1}} + \frac{\delta}{2} = \frac{b-a-\delta}{2^{k/2+1}} + \frac{\delta}{2},$$

где k — количество вычислений функции $f(x)$, выполненных за n итераций. Значение k, необходимое для достижения заданной точности ε, может быть определено из неравенства

$$\frac{k}{2} = n > \log_2\left(\frac{b-a-\varepsilon}{2\varepsilon-\delta}\right).$$

Если вычисление $f(x)$ трудоемко, то количество итераций иногда ограничивают заранее заданным числом k вычислений значений функции $f(x)$ в точках $x_1^{(n-1)}$ и $x_2^{(n-1)}$. Тогда точность, с которой получен минимум $\bar{x}$, будет примерно равна $(b-a)/2^{k/2+1}$. В этой связи разумно поставить вопрос: можно ли за одно и то же количество вычислений $f(x)$ найти точку $\bar{x}$ более точно? Таким методом является, например, метод золотого сечения.

3.2.3. Метод золотого сечения

Метод золотого сечения более рационально использует вычисленные значения функции $f(x)$ и позволяет переходить к очередному подотрезку, содержащему точку минимума $\bar{x}$, после вычисления одного, а не двух значений минимизируемой функции.

Напомним, что золотым сечением отрезка называется такое его деление на две неравные части, что отношение длины всего отрезка к длине большего подотрезка равно отношению длины большего подотрезка к длине меньшего.

Метод золотого сечения также применим к недифференцируемым функциям. Будем считать, что функция $f(x)$ унимодальна на отрезке $[a, b]$.

При построении этого метода мы будем поступать так же, как и в методе бисекций поиска нулей функций: вычисляем $f(x)$

внутри отрезка $[a, b]$ и отбрасываем подотрезок, который не содержит точки минимума, после чего повторяем процесс на выбранном подотрезке до тех пор, пока длина текущего подотрезка не станет меньше заданной абсолютной точности.

Легко видеть, что одного значения функции $f(x)$ внутри отрезка $[a, b]$ не достаточно для того, чтобы отбросить отрезок, не содержащий точку минимума. Однако двух ее значений достаточно для определения направления поиска. Следовательно, нам надо построить такой итерационный процесс, при котором вычисления значений $f(x)$ осуществляются из таких условий:

1) на каждом шаге итераций вычисленное значение $f(x)$ на текущем подотрезке может быть использовано повторно;

2) для получения следующего подотрезка, имеющего меньшую длину, чем предыдущий, нужно только одно новое значение $f(x)$;

3) отбрасываемый подотрезок должен быть как можно большей длины для ускорения сходимости.

Выберем произвольным образом точку c на отрезке $[a, b]$, а затем покажем, как это следует сделать так, чтобы удовлетворялись перечисленные условия. Для определенности будем считать, что $b - c > c - a$. Обозначим через r отношение

$$r = \frac{c - a}{b - a}.$$

Тогда можно записать:

$$\frac{b - c}{b - a} = 1 - \frac{c - a}{b - a} = 1 - r.$$

Здесь r и $1 - r$ — относительные длины подотрезков $[a, c]$ и $[c, b]$ соответственно. Теперь возьмем точку d на большем подотрезке $[c, b]$ и обозначим через w отношение

$$w = \frac{d - c}{b - a},$$

определяющее относительную длину подотрезка $[c, d]$.

Если отвлечься от конкретных значений функции $f(x)$ в точках a, c, d и b, то в общем случае следующим для рассмотрения подотрезком должен быть либо подотрезок $[a, d]$, либо подотрезок $[c, b]$, причем для общего алгоритма (который строится не под конкретную функцию) выбор того или иного из этих подотрезков равновероятен. Следовательно, чтобы минимизировать будущие трудозатраты, длины подотрезков $[a, d]$ и $[c, b]$ должны быть равны. Это означает, что в каком бы направлении мы ни пошли дальше, объемы последующих вычислений с равной вероятностью будут одинаковыми. В этом смысле такой выбор точек c и d будет оптимальным.

Итак, точку d выберем на отрезке $[c, b]$ так, чтобы $d - a = b - c$. Представив это равенство как $d - c + c - a = b - c$, поделим его на $b - a$:

$$\frac{d-c}{b-a} + \frac{c-a}{b-a} = \frac{b-c}{b-a}.$$

Следовательно, мы получили соотношение, связывающее r и w:

$$w = 1 - 2r. \tag{15}$$

Однако конкретное значение r, т.е. как следует однозначно выбрать точку c, остается неопределенным. Для определения r взглянем на выбор точки d под другим углом зрения.

Предположим, что мы находимся в начале процесса, имея не отрезок $[a, b]$, а отрезок $[c, b]$. Тем самым, выбор точки d на $[c, b]$ эквивалентен выбору точки c на $[a, b]$. Посмотрим, можно ли взять точку d так, чтобы сохранилось соотношение (15), и так, чтобы точка d делила отрезок $[c, b]$ в том же отношении, что и точка c отрезок $[a, b]$. Для этого должно выполняться равенство

$$\frac{d-c}{b-c} = r \qquad \frac{(d-c)/(b-a)}{(b-c)/(b-a)} = r.$$

Переходя к относительным длинам подотрезков, получим второе соотношение на r и w:

$$\frac{w}{1-r} = r, \quad w = r(1-r). \tag{16}$$

Решая систему (15), (16) относительно r, приходим к квадратному уравнению

$$r^2 - 3r + 1 = 0,$$

корни которого равны

$$r_1 = \frac{3+\sqrt{5}}{2}, \quad r_2 = \frac{3-\sqrt{5}}{2}.$$

Поскольку $r_1 > 1$, то этот корень отбрасываем. Таким образом, мы получили искомое значение r, равное

$$r = \frac{3-\sqrt{5}}{2} \approx 0.38197.$$

Другими словами, оптимальное разбиение текущего подотрезка, на котором ищется минимум, состоит в выборе такой точки, расстояние от которой до ближайшего к ней конца подотрезка, отнесенное к его длине (назовем это относительным расстоянием), равно $r \approx 0.38197$, а относительное расстояние до дальнего конца подотрезка равно $1 - r \approx 0.61803$.

Такой выбор точки разбиения и есть золотое сечение отрезка. Действительно, по определению золотого сечения имеем:

$$\frac{b-a}{b-c} = \frac{b-c}{c-a}, \quad (b-a)(c-a) = (b-c)^2.$$

Поделив последнее равенство на $(b-a)^2$, получим следующее квадратное уравнение относительно r:

$$r = (1-r)^2,$$

корни которого также равны

$$r_1 = \frac{3+\sqrt{5}}{2}, \quad r_2 = \frac{3-\sqrt{5}}{2}.$$

Основываясь на рассмотренном выше способе выбора точек c и d на исходном отрезке $[a, b]$, мы можем построить следующий алгоритм.

Вычисляем $r = \left(3-\sqrt{5}\right)/2$ и полагаем

$$c = a + r(b-a), \quad d = b - r(b-a). \tag{17}$$

Далее процесс разветвляется:

а) если $f(c) < f(d)$, то в качестве текущего подотрезка выбираем $[a, d]$ (т.е. полагаем $b = d$), вычисляем новую точку c по формуле (17), а в качестве новой точки d берем старую точку c;

б) если $f(c) \geq f(d)$, то в качестве текущего подотрезка выбираем $[c, d]$ (т.е. полагаем $a = c$), вычисляем новую точку d по формуле (17), а в качестве новой точки c берем старую точку d.

Процесс продолжается до тех пор, пока длина текущего подотрезка не станет меньше заданной абсолютной погрешности ε.

Приведем формализованную запись изложенного алгоритма:

```
    r=(3-sqrt(5))/2
    c=a+r(b-a)
    d=b-r(b-a)
10  if(b-a<eps) go to 30
      if(f(c)<f(d)) go to 20
        a=c
        c=d
        d=b-r(b-a)
        go to 10
20    b=d
      d=c
      c=a+r(b-a)
      go to 10
30  min=(b-a)/2
```

Поскольку по предположению функция $f(x)$ унимодальна на $[a,b]$, то этот алгоритм можно усовершенствовать следующим образом. Если $f(a) < f(c) < f(b)$, то мы сразу можем отбросить подотрезок $[c,b]$ и перейти к рассмотрению подотрезка $[a,c]$. Так же поступаем, когда $f(a) > f(c) > f(b)$: отбрасываем подотрезок $[a,c]$ и переходим к подотрезку $[c,b]$.

Напомним: в самом начале мы для определенности предположили, что $b - c > c - a$. Однако аналогичный итерационный процесс можно построить и в предположении, что $b - c < c - a$. Единственное отличие состоит в том, что для вычисления r получается другое квадратное уравнение $r^2 + r - 1 = 0$, корни которого равны

$$r_1 = \frac{\sqrt{5} - 1}{2}, \quad r_2 = \frac{-\sqrt{5} - 1}{2}.$$

Отбрасывая отрицательный корень, получаем, что в этом случае $r = \left(\sqrt{5} - 1\right)/2 \approx 0.61803$. Такое значение r соответствует второй точке золотого сечения отрезка.

Теперь рассмотрим вопрос о скорости сходимости метода золотого сечения. На каждом шаге длина очередного подотрезка сокращается в $(1-r)$ раз (т.е. примерно в 0.61803 раза). Это означает, что метод сходится со скоростью геометрической прогрессии со знаменателем $(1-r)$. Поскольку точка минимума $\bar{x}$ расположена на текущем подотрезке, то имеет место оценка

$$\begin{aligned} |\bar{x}-x_n| \le \varepsilon_n &= (1-r)^n(b-a) = \left(\frac{\sqrt{5}-1}{2}\right)^n (b-a) \approx \\ &\approx 0.61803^n(b-a), \end{aligned}$$

где x_n — текущее приближение к $\bar{x}$. Вычисления заканчиваются, когда $\varepsilon_n < \varepsilon$, где ε — заданная абсолютная точность.

Если ε — заданная относительная точность, то число шагов n метода золотого сечения, обеспечивающего эту точность, должно удовлетворять неравенству

$$n \ge \frac{\ln\big(\varepsilon/(b-a)\big)}{\ln\big((\sqrt{5}-1)/2\big)} \approx -2.1 \ln\frac{\varepsilon}{b-a}.$$

По сравнению с методом деления отрезка пополам скорость сходимости метода золотого сечения медленнее, однако последний требует на каждом шаге итерационного процесса только одного вычисления значения функции $f(x)$, а а не двух. Это позволяет сделать суждение о том, что метод золотого сечения для достижения одной и той же точности требует меньших вычислительных затрат, чем метод деления отрезка пополам. Отмеченное преимущество становится ощутимым уже при небольших количествах вычислений минимизируемой функции. Действительно, для метода золотого сечения при k вычислениях функции $f(x)$ абсолютная погрешность $\varepsilon_n^{(1)}$ приближения к $\bar{x}$ примерно равна

$$\varepsilon_n^{(1)} \approx \left(\frac{\sqrt{5}-1}{2}\right)^k (b-a),$$

тогда как для метода деления отрезка пополам достигнутую абсолютную погрешность $\varepsilon_n^{(2)}$ можно записать в виде

$$\varepsilon_n^{(2)} \approx \frac{b-a}{2^{k/2+1}} < \frac{b-a}{2^{k/2}}.$$

Для $\varepsilon_n^{(1)}$ и $\varepsilon_n^{(2)}$ можно выписать следующее соотношение:

$$\frac{\varepsilon_n^{(1)}}{\varepsilon_n^{(2)}} \approx \frac{\big((\sqrt{5}-1)/2\big)^k (b-a)}{(b-a)/2^{k/2}} = \left(\frac{2\sqrt{2}}{\sqrt{5}+1}\right)^k \approx (0.87)^k.$$

Отсюда видно, что метод золотого сечения менее трудоемок, чем метод деления отрезка пополам.

Если функция $f(x)$ не унимодальна и на $[a,b]$ имеется несколько локальных минимумов, описанный выше процесс сойдется к одному из них (не обязательно наименьшему).

3.3. Методы с использованием производных минимизируемой функции

3.3.1. Метод касательных

Пусть $f(x)$ — выпуклая дважды дифференцируемая функция на отрезке $[a,b]$; предположим, что $f'(a)f'(b) < 0$, т.е. точка минимума $\bar{x}$ находится внутри этого отрезка.

Заметим, что если $f'(a) = 0$ или $f'(b) = 0$, то соответственно точки a или b являются точками минимума; если же $f'(a) > 0$ и $f'(b) > 0$, то $\bar{x} = a$, а если $f'(a) < 0$ и $f'(b) < 0$, то $\bar{x} = b$.

Возьмем $n = 1$ и положим $a_n = a$, $b_n = b$. Выпишем уравнения касательных к $f(x)$ в точках a_n и b_n:

$$y = f'(a_n)x + f(a_n) - f'(a_n)a_n,$$
$$y = f'(b_n)x + f(b_n) - f'(b_n)b_n.$$

Обозначим через c_n точку пересечения этих касательных:

$$f'(a_n)c_n + f(a_n) - f'(a_n)a_n = f'(b_n)c_n + f(b_n) - f'(b_n)b_n.$$

Следовательно,

$$c_n = \frac{f(a_n) - f(b_n) + f'(b_n)b_n - f'(a_n)a_n}{f'(b_n) - f'(a_n)}.$$

Если $\big|f'(c_n)\big| < \varepsilon$, где ε — заданная абсолютная точность, то в качестве точки минимума берем c_n. В противном случае выполняем следующие действия:

1) если $f'(c_n) < 0$, то полагаем $a_{n+1} = c_n$ и $b_{n+1} = b_n$ (сдвиг вправо);

2) если $f'(c_n) > 0$, то полагаем $a_{n+1} = a_n$ и $b_{n+1} = c_n$ (сдвиг влево).

Затем полагаем $n = n + 1$ и повторяем описанный процесс, получивший название метода касательных. После достижения заданной точности последнюю точку c_n берем в качестве приближения к точке минимума $\bar{x}$, причем имеет место оценка

$$\bigl|c_n - \bar{x}\bigr| \le \bigl|b_n - a_n\bigr|.$$

Теперь оценим скорость сходимости метода касательных, предварительно отметив, что $f''(x) > 0$ всюду на $[a, b]$ в силу необходимого и достаточного признака выпуклости дважды дифференцируемой функции. Перепишем выражение для c_n в виде

$$\begin{aligned} c_n &= \frac{f(a_n) - f(b_n) + f'(b_n)b_n}{f'(b_n) - f'(a_n)} + \\ &\quad + \frac{\bigl(f'(b_n)a_n - f'(b_n)a_n\bigr) - f'(a_n)a_n}{f'(b_n) - f'(a_n)} = \\ &= \frac{f(a_n) - f(b_n) + f'(b_n)(b_n - a_n)}{f'(b_n) - f'(a_n)} + a_n. \end{aligned}$$

Отсюда

$$c_n - a_n = \frac{f(a_n) - f(b_n) + f'(b_n)(b_n - a_n)}{f'(b_n) - f'(a_n)}.$$

Разложим $f(a_n)$ и $f'(a_n)$ в ряд Тейлора в окрестности точки b_n с шагом $b_n - a_n$:

$$\begin{aligned} f(a_n) &= f\bigl(b_n - (b_n - a_n)\bigr) = f(b_n) - (b_n - a_n)f'(b_n) + \\ &\quad + \frac{(b_n - a_n)^2}{2} f''(\xi_n), \quad a_n \le \xi_n \le b_n, \\ f'(a_n) &= f'\bigl(b_n - (b_n - a_n)\bigr) = \\ &= f'(b_n) - (b_n - a_n)f''(\mu_n), \quad a_n \le \mu_n \le b_n. \end{aligned}$$

Подставим полученные разложения в выражение для $c_n - a_n$:

$$c_n - a_n = \frac{\dfrac{(b_n - a_n)^2}{2} f''(\xi_n)}{(b_n - a_n)f''(\mu_n)} = \frac{1}{2}\,(b_n - a_n)\,\frac{f''(\xi_n)}{f''(\mu_n)}.$$

Аналогично получим:

$$b_n - c_n = \frac{1}{2}\,(b_n - a_n)\,\frac{f''(\eta_n)}{f''(\mu_n)}, \quad a_n \le \eta_n \le b_n.$$

По построению метода касательных в качестве следующего подотрезка $[a_{n+1}, b_{n+1}]$ выбирается либо подотрезок $[a_n, c_n]$, либо подотрезок $[c_n, b_n]$. Следовательно, из последних двух равенств заключаем, что

$$b_{n+1} - a_{n+1} \le \max\{c_n - a_n, b_n - c_n\} \le \frac{q_n}{2}\,(b_n - a_n),$$

где

$$q_n = \max\left\{\frac{f''(\xi_n)}{f''(\mu_n)}, \frac{f''(\eta_n)}{f''(\mu_n)}\right\}.$$

Поскольку последовательности ξ_n, η_n, μ_n вместе с последовательностями a_n и b_n стремятся с увеличением n к точке минимума $\bar{x}$, то в силу непрерывности функции $f''(x)$ и условия $f''(x) > 0$ имеем $\lim\limits_{n\to\infty} q_n = 1$. Следовательно, для любого $\delta > 0$ найдется номер $N = N(\delta)$, такой, что $q_n \le 1+\delta$ при всех $n \ge N$. Тогда при $n \ge N$ получим неравенство

$$b_{n+1} - a_{n+1} \le \frac{1+\delta}{2}\,(b_n - a_n).$$

Так как точка минимума $\bar{x}$ лежит на $[a_{n+1}, b_{n+1}]$, то в силу построения этого подотрезка можно записать:

$$\begin{aligned}
\left|c_n - \bar{x}\right| &\le b_{n+1} - a_{n+1} \le \\
&\le \frac{1+\delta}{2}\,(b_n - a_n) \le \frac{(1+\delta)^2}{2^2}\,(b_{n-1} - a_{n-1}) \le \ldots \le \\
&\le \left(\frac{1+\delta}{2}\right)^{n-N} (b_N - a_N), \quad n \ge N.
\end{aligned}$$

Отсюда заключаем, что скорость сходимости метода касательных не меньше скорости сходимости геометрической прогрессии со знаменателем $q = \dfrac{1+\delta}{2} \approx \dfrac{1}{2}$.

Заметим, что полученная выше оценка скорости сходимости носит теоретический характер и не является конструктивной, поскольку она не позволяет оценить число итераций для достижения заданной точности через величины, вычисляемые в процессе реализации метода.

3.3.2. Метод Ньютона одномерной минимизации

Пусть $f(x)$ — выпуклая дважды дифференцируемая функция, заданная на отрезке $[a, b]$. Тогда можно построить более быстрый метод, основанный на решении уравнения $f'(x) = 0$. Напомним, что корень $\bar{x}$ этого уравнения является точкой минимума, если $f''(x) > 0$, и точкой максимума, если $f''(x) < 0$.

Будем решать уравнение $f'(x) = 0$ *методом Ньютона*

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}. \tag{18}$$

При этом сохраняются все свойства метода Ньютона, применяемого для нахождения нулей нелинейных функций. В частности, имеет место квадратичная скорость сходимости, если искомый корень уравнения $f'(x) = 0$ простой; если же корень имеет кратность p, то сходимость становится линейной (метод сходится со скоростью геометрической прогрессии со знаменателем $(p-1)/p$).

Итерационный процесс (18) можно построить другим способом. Разложим $f(x)$ в точке x_n в ряд Тейлора и удержим в этом разложении три члена:

$$\begin{aligned} f(x) &= f\big(x_n + (x - x_n)\big) \approx \\ &\approx f(x_n) + (x - x_n) f'(x_n) + \frac{1}{2}(x - x_n)^2 f''(x_n) = p_2(x). \end{aligned}$$

Выписанное разложение эквивалентно приближению функции $f(x)$ в окрестности точки x_n параболой. Минимум этой параболы достигается в точке, определяемой формулой (18). Действительно, уравнение для вычисления минимума параболы $p_2(x)$ имеет вид $p_2'(x) = 0$, т.е.

$$(x - x_n) f''(x_n) + f'(x_n) = 0, \quad x = x_n - \frac{f'(x_n)}{f''(x_n)}.$$

Найденную точку минимума для параболы $p_2(x)$ берем за следующее приближение x_{n+1} к минимуму исходной функции и повторяем процесс. Поэтому метод (18) называют еще *методом парабол.*

Если выражения для $f'(x)$ и $f''(x)$ громоздки, то их можно заменить, например, конечно-разностными аппроксимациями с порядком $O\left(h^2\right)$:

$$f'(x_n) = \frac{f(x_n+h) - f(x_n-h)}{2h} + O\left(h^2\right),$$
$$f''(x_n) = \frac{f(x_n+h) - 2f(x_n) + f(x_n-h)}{h^2} + O\left(h^2\right).$$

Тогда вместо (18) получаем следующий итерационный процесс:

$$x_{n+1} = x_n - \frac{h}{2} \cdot \frac{f(x_n+h) - f(x_n-h)}{f(x_n+h) - 2f(x_n) + f(x_n-h)}. \qquad (19)$$

Можно показать, что вблизи точки минимума характер сходимости процесса (19) близок к квадратичному. Он эквивалентен замене кривой $f(x)$ в окрестности точки x_n интерполяционной параболой, построенной по точкам $x_n - h$, x_n, $x_n + h$.

Как и в случае многомерной минимизации, метод Ньютона часто используется на завершающем этапе, когда найдено грубое приближение к точке минимума при помощи какого-либо менее трудоемкого метода и требуется найти минимум с большой точностью.

3.4. О влиянии погрешностей вычислений

Погрешности (ошибки) вычислений, возникающие при выполнении расчетов на ЭВМ, обусловлены тремя причинами. Во-первых, ошибки могут содержаться в исходных данных (*неустранимая погрешность*). Во-вторых, ошибки возникают в результате замены бесконечного процесса конечным (например, вследствие отбрасывания остаточного члена разложения). В-третьих, ошибки возникают из-за конечной точности, с которой числа могут быть представлены в машине (*ошибки округления*).

Каждый из этих типов ошибок является неизбежным в вычислениях. В этой связи весьма важной становится проблема исследования распространения (накопления) ошибок в процессе вычислений. Поясним сказанное на примере анализа накопления вычислительной погрешности для приведенного ниже другого алгоритма одномерной минимизации методом золотого сечения.

Пусть $a_1 = a$ и $b_1 = b$. На отрезке $[a_1, b_1]$ возьмем точки золотого сечения u_1 и u_2 и вычислим $f(u_1)$ и $f(u_1)$. Далее поступаем следующим образом:

— если $f(u_1) \le f(u_2)$, то $a_2 = a_1$, $b_2 = u_2$, $\bar{u}_2 = u_1$;

— если $f(u_1) > f(u_2)$, то $a_2 = u_1$, $b_2 = b_1$, $\bar{u}_2 = u_2$.

При этом $\bar{u}_2$ является одной из точек золотого сечения отрезка $[a_2, b_2]$.

Предположим, что уже определены точки $u_1, u_2, \ldots, u_{n-1}$, вычислены значения $f(u_1), f(u_2), \ldots, f(u_{n-1})$ и определен отрезок $[a_{n-1}, b_{n-1}]$, содержащий точку минимума. Считаем, что нам известна точка $\bar{u}_{n-1}$, производящая золотое сечение отрезка $[a_{n-1}, b_{n-1}]$, причем выполнено $f(\bar{u}_{n-1}) = \min f(u_i)$, $i = 1, \ldots, n-1$, $n \ge 2$. Тогда в качестве следующей точки возьмем $u_n = a_{n-1} + b_{n-1} - \bar{u}_{n-1}$, также производящую золотое сечение отрезка $[a_{n-1}, b_{n-1}]$.

Рассмотренный алгоритм лишь незначительно отличается от построенного в п. 3.2.3, но практически неприменим для численной реализации из-за роста вычислительной погрешности. Действительно, число $\sqrt{5}$ вычисляется приближенно; следовательно, выражение $u_1 = a_1 + \left(3 - \sqrt{5}\right)(b_1 - a_1)/2$ находится с некоторой погрешностью. Оценим ее влияние на последующие шаги алгоритма.

Величина $\Delta_n = b_n - a_n$ является решением разностного уравнения

$$\Delta_{n-2} = \Delta_{n-1} + \Delta_n \tag{20}$$

с начальными условиями

$$\Delta_1 = b_1 - a_1, \quad \Delta_2 = b_1 - u_1. \tag{21}$$

Общее решение уравнения (20) может быть записано в виде

$$\Delta_n = A\, q_1^n + B\, q_2^n, \quad n = 1, 2, \ldots\ , \tag{22}$$

где $q_1 = \left(\sqrt{5} - 1\right)/2$, $q_2 = -\left(\sqrt{5} + 1\right)/2$.

Подставим выражение (22) в начальные условия (21). Получим систему для определения коэффициентов A и B:

$$\begin{aligned} A\,q_1 + B\,q_2 &= \Delta_1, \\ A\,q_1^2 + B\,q_2^2 &= \Delta_2. \end{aligned} \tag{23}$$

Из этой системы находим

$$A = \frac{2(b-a)}{\sqrt{5}-1}, \quad B = 0.$$

Таким образом, в случае отсутствия ошибок округлений имеем

$$\Delta_n = q_1^{n-1}(b-a).$$

Однако точка u_1 вычислена с погрешностью; следовательно, в системе (23) вместо точного значения Δ_2 следует взять некоторое приближенное значение $\tilde{\Delta}_2 = \Delta_2 + \delta$. Тогда коэффициенты A и B определяются из системы (23) с какими-то погрешностями:

$$\tilde{A} = A + \delta_1, \quad \tilde{B} = B + \delta_2.$$

В результате вместо величины Δ_n мы получим

$$\tilde{\Delta}_n = \tilde{A}\,q_1^n + \tilde{B}\,q_2^n.$$

Поскольку $0 < q_1 < 1$ и $|q_2| > 1$, то погрешность

$$\left|\Delta_n - \tilde{\Delta}_n\right| = |\delta_1 q_1^n + \delta_2 q_2^n|$$

с возрастанием n будет расти и уже при небольших n точки $\bar{u}_n$ и $u_{n+1} = a_n + b_n - \bar{u}_n$ могут оказаться лежащими вне текущего отрезка, на котором локализована точка минимума.

Разобранный пример показывает, что ошибки округления, которые сами по себе кажутся незначительными, могут оказать существенное влияние на конечный результат, если для его получения выполняется большое количество арифметических операций. Поэтому следует стараться минимизировать погрешности в каждой операции или в последовательности операций, уменьшая тем самым их распространение и воздействие на конечный результат.

Рассмотрим, как можно это сделать в случае выполнения часто встречающейся в задачах минимизации простой операции вычисления среднего арифметического $(a+b)/2$ двух чисел a и b. Возможны два способа выполнения этой операции:

$$c = \frac{a+b}{2} \tag{24}$$

и

$$c = a + \frac{b-a}{2}. \tag{25}$$

Очевидно, что формула (24) требует на одну операцию сложения меньше, чем формула (25), но с точки зрения точности не всегда лучше. Действительно, пусть вычисления проводятся в десятичной арифметике с тремя цифрами и с правильным округлением для $a = 0.596$ и $b = 0.600$. Тогда можно записать, что

$$c = (0.596 + 0.600)/2 = 1.200/2 = 0.600,$$

хотя правильное значение c равно 0.598. Если же мы будем проводить вычисления по формуле (25), то получим следующий результат

$$c = 0.596 + (0.600 - 0.596)/2 = 0.596 + 0.004/2 = 0.598.$$

Заметим, что в данном примере, для которого формула (25) оказалась предпочтительней, числа a и b имеют одинаковые знаки.

Теперь рассмотрим другой пример для десятичной четырехзначной арифметики, в которой вместо правильного округления применяется отбрасывание лишних разрядов. Пусть $a = -3.483$ и $b = 8.765$. Тогда по формуле (24) будем иметь

$$c = (-3.483 + 8.765)/2 = 5.282/2 = 2.641,$$

что представляет собой точный результат. Вычисление же по формуле (25) дают

$$\begin{aligned} c &= -3.483 + (8.765 + 3.483)/2 = -3.483 + 12.24/2 = \\ &= -3.483 + 6.120 = 2.637. \end{aligned}$$

Даже если бы в этом примере применялось правильное округление, то результат по формуле (25) все равно отличался бы от точного: $c = 2.642$. Следовательно, в данном примере, в котором числа a и b имеют разные знаки, формула (24) оказалась предпочтительней.

Отсюда можно сделать вывод, что для достижения наивысшей точности следует применять либо формулу (24), либо формулу (25) в зависимости от знаков чисел a и b. Поэтому наилучший способ вычисления среднего арифметического чисел a и b можно представить следующим образом:

$$\operatorname{sign}(a) \neq \operatorname{sign}(b), \quad c = (a+b)/2 \qquad c = a + (b-a)/2.$$

4. Этапы выполнения заданий. Содержание отчета

Каждому студенту предлагается решить конкретную задачу на безусловную минимизацию функций многих переменных, определяемую одним из вариантов. Перечень вариантов в свою очередь определяется следующим образом.

Запишем рассматриваемые методы многомерной минимизации в виде

$$x^{k+1} = x^k + \alpha_k h^k,$$

где x^k — текущее приближение к точке минимума $\bar{x}$, h^k — направление убывания минимизируемой функции $f(x)$ в точке x^k, α_k — шаг по направлению убывания и x^{k+1} — следующее приближение к точке минимума.

Варианты методов многомерной минимизации определяются выбором способа построения вектора h^k:

1) $h^k = -f'(x^k)$ — градиентный метод;

2) $h^k = -\left[f''(x^k)\right]^{-1} f'(x^k)$ — метод Ньютона в следующих модификациях:

— когда элементы матрицы $f''(x^k)$ вычисляются явно,

— когда для вычисления элементов матрицы $f''(x^k)$ используются формулы численного дифференцирования.

Способы выбора шага α_k по направлению убывания определяются следующими вариантами:

1) дробление шага (см. п. 2.2);

2) метод перебора (см. п. 3.2.1);

3) метод деления отрезка пополам (см. п. 3.2.2);

4) метод золотого сечения (см. п. 3.2.3);

5) метод касательных (см. п. 3.3.1);

6) метод Ньютона одномерной минимизации (см. п. 3.3.2);

7) метод Ньютона одномерной минимизации с использованием формул численного дифференцирования (см. п. 3.3.2).

Выполнение задания состоит из следующих этапов.

Этап 1. Сначала применяется градиентный метод с одним из перечисленных способов выбора шага α_k при заданном начальном приближении к точке минимума. Результатом этого этапа является уточнение приближения к минимуму с погрешностью $\sqrt{\varepsilon}$, где ε — требуемая точность решения задачи.

Этап 2. Затем применяется метод Ньютона многомерной минимизации в одной из его модификаций с одним из перечисленных способов выбора шага α_k при начальном приближении к точке минимума, совпадающим с уточненным приближением, полученным после завершения первого этапа. Результатом второго этапа является вычисление приближения к минимуму с погрешностью ε.

Критерием окончания счета на каждом этапе является одновременное выполнение следующих неравенств:

$$\begin{aligned}
&\left\|x^{k+1} - x^k\right\| \leq \delta,\\
&\left|f(x^{k+1}) - f(x^k)\right| \leq \delta,\\
&\left\|f'(x^{k+1})\right\| \leq \delta,
\end{aligned}$$

где $\delta = \sqrt{\varepsilon}$ для первого этапа и $\delta = \varepsilon$ для второго этапа.

По окончании выполнения задания практикума оформляется отчет о проделанной работе по следующей схеме.

1) Постановка задачи.

2) Описание реализованных алгоритмов.

3) Описание отладочных тестов и анализ их результатов.

4) Графическое представление результатов расчетов.

Приведем два простых примера анализа экстремумов функций. Эти примеры могут быть использованы для отладки программ.

Пример 1. Рассмотрим функцию

$$f(x) = x_1^2 + x_2^2, \quad x \in R^n.$$

Градиент этой функции имеет вид

$$f'(x) = (2x_1, 2x_2).$$

Выпишем решение уравнения $f'(x) = 0$: $x^{(1)} = (0, 0)$. Матрица Гессе для рассматриваемой функции имеем вид

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

По критерию Сильвестра матрица $f''(x^{(1)})$ положительно определена. В силу достаточного условия локальной оптимальности точка $x^{(1)}$ является точкой локального минимума. На основании следствия из теоремы Вейерштрасса заключаем, что минимум $f(x)$ на R^n достигается. Следовательно, $x^{(1)}$ — точка глобального минимума.

Пример 2. Рассмотрим функцию

$$f(x) = x_1^4 + x_2^4 - (x_1 + x_2)^2, \quad x \in R^n.$$

Градиент этой функции имеет вид

$$f'(x) = (4x_1^3 - 2(x_1 + x_2), 4x_2^3 - 2(x_1 + x_2)).$$

Выпишем решение уравнения $f'(x) = 0$:

$$x^{(1)} = (0, 0), \quad x^{(2)} = (-1, -1), \quad x^{(3)} = (1, 1).$$

Матрица Гессе для рассматриваемой функции имеем вид

$$f''(x) = \begin{pmatrix} 12x_1^2 - 2 & -2 \\ -2 & 12x_1^2 - 2 \end{pmatrix}.$$

Для найденных стационарных точек имеем

$$f''(x^{(1)}) = \begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix}, \quad f''(x^{(2)}) = f''(x^{(3)}) = \begin{pmatrix} 10 & -2 \\ -2 & 10 \end{pmatrix}.$$

По критерию Сильвестра матрица $f''(x^{(1)})$ неотрицательно определена, т.е. необходимое условие оптимальности второго порядка выполняется. Однако легко видеть, что в любой достаточно малой окрестности точки $x^{(1)}$ наша функция принимает отрицательные значения. Следовательно, эта точка не является точкой минимума.

Теперь рассмотрим матрицу Гессе для двух других стационарных точек. По критерию Сильвестра эта матрица положительно определена. Это означает, что в силу достаточного условия локальной оптимальности точки $x^{(2)}$ и $x^{(3)}$ являются точками локального минимума. На основании следствия из теоремы Вейерштрасса заключаем, что минимум $f(x)$ на R^n существует. Следовательно, эти точки доставляют и глобальный минимум.

В заключение приведем примеры некоторых функций, которые могут быть выбраны для отладки программ и проверки правильности выполнения заданий практикума.

1. $f(x_1, x_2) = x_1^2 - x_2^2$.
2. $f(x_1, x_2) = x_1^2 + x_2^4$.
3. $f(x_1, x_2) = (x_1 - 1)^2 + (x_2 + 1)^2$.
4. $f(x_1, x_2) = x_1^3 + x_2^3 - 3x_1x_2$.
5. $f(x_1, x_2) = x_1^2 - x_1x_2 + x_2^2 - 2x_1 + x_2$.
6. $f(x_1, x_2) = x_1^2 - x_2^2 - 4x_1 + 6x_2$.
7. $f(x_1, x_2) = 2x_1^2 + x_1x_2 + x_2^2$.
8. $f(x_1, x_2) = (1 - x_1)^2 + 10(x_2 - x_1^2)^2$.
9. $f(x_1, x_2) = (x_2 - x_1^2)^2 + (1 - x_1)^2$.
10. $f(x_1, x_2) = 3x_1x_2 - x_1x_2^2 - x_1^2x_2$.
11. $f(x_1, x_2) = 3x_1^2 + 4x_1x_2 + x_2^2 - 8x_1 - 12x_2$.
12. $f(x_1, x_2, x_3) = x_1^2 + 5x_2^2 + 3x_3^2 + 4x_1x_2 - 2x_2x_3 - 2x_1x_3$.

13. $f(x_1, x_2) = x_1^3 - x_1x_2 + x_2^2 - 2x_1 + 3x_2 - 4$.

14. $f(x_1, x_2, x_3) = -x_1^2 - x_2^2 - x_3^2 - x_1 + x_1x_2 + 2x_3$.

15. $f(x_1, x_2, x_3) = x_1^3 + x_2^2 + x_3^2 + x_2x_3 - 3x_1 + 6x_2 + 2$.

Литература

1. *Бахвалов Н.С.* Численные методы. М.: Наука, 1975.
2. *Бахвалов Н.С., Жидков Н.П., Кобельков Г.М.* Численные методы. М.: Наука, 1987.
3. *Сухарев А.Г., Тимохов А.В., Федоров В.В.* Курс методов оптимизации. М.: Наука, 1986.
4. *Васильев Ф.П.* Численные методы решения экстремальных задач. М.: Наука, 1987.
5. *Алексеев В.М., Галеев Э.М., Тихомиров В.М.* Сборник задач по оптимизации. М.: Наука, 1984.
6. *Летова Т.А., Пантелеев А.В.* Экстремум функций в примерах и задачах. М.: Изд-во МАИ, 1998.
7. Сборник задач по математике. Методы оптимизации / Под ред. Ефимова А.В. М.: Наука, 1990.
8. *Федоров В.В.* Численные методы максимина. М.: Наука, 1979.