Семестр 2 (2019), занятие 3

Унимодальные функции

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

- если $a < \alpha$, то $f(x)$ монотонно убывает на отрезке $[a, \alpha]$;
- если $\beta < b$, то $f(x)$ монотонно возрастает на отрезке $[\beta, b]$;
- если $x_1, x_2 \in [\alpha, \beta]$, то $f(x_1) = f(x_2)$.

В дальнейшем, нас будет интересовать нахождение точки $x_{\min} \in [a, b]$, в которой функция $f(x)$ достигает своего минимума. Очевидно, что во всех точка отрезка $[\alpha, \beta]$ функция $f(x)$ принимает минимальное значение. Поэтому, если этот отрезок вырожден, то искомая точка определяется однозначно $x_{\min} = \alpha = \beta$. Иначе, в качестве $x_{\min}$ можно взять любую точку из отрезка $[\alpha, \beta]$.

Локализовать точку $x_{\min}$ помогает следующее свойство унимодальной функции.

Утверждение. Пусть $c, d \in [a, b]$, при этом $a < c < d < b$. Тогда, если $f(c) < f(d)$ (рис. 1), то $x_{\min} \in [a, d]$, иначе $x_{\min} \in [c, b]$.

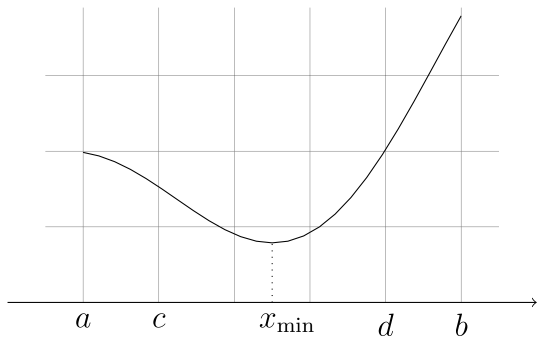

Рис. 1: Локализация точки минимума унимодальной функции.

Постановка задачи

Пусть $f(x)$ – унимодальная функция на отрезке $[a, b]$. Требуется найти приближенное значение точки минимума этой функции $x^* \approx x_{\min} \in [a, b]$ с заданной точностью $\varepsilon > 0$.

В зависимости от метода решения задачи параметр ε может трактоваться по-разному.

1. В качестве параметра ε может выступать абсолютная точность приближения. Приближенное значение x^* принимается, если выполняется неравенство

$$|x^* - x_{\min}| < \varepsilon.$$

2. Многие методы подразумевают последовательное построение приближений

$$x_1, x_2, x_3, \dots x_n, x_{n+1}, \dots \approx x_{\min}.$$

Если $|x_n - x_{n+1}| < \varepsilon$, то x_{n+1} принимается в качестве искомого приближения.

Метод деления отрезка пополам

1. В рамках данного метода строится последовательность вложенных отрезков $[a_n, b_n]$ ($n \geqslant 0$). Середины этих отрезков $x_n = \frac{a_n + b_n}{2}$ рассматриваются в качестве приближений к точке минимума функции.

2. Для каждого построенного отрезка оценивается его длина $l_n = b_n - a_n$. Если $l_n < 2\varepsilon$, то процесс построения завершается. Число x_n является приближением к точке минимума функции с абсолютной погрешностью ε.

3. На начальном шаге $n = 0$ полагаем

$$\begin{aligned} a_0 &= a, \\ b_0 &= b. \end{aligned}$$

4. На каждом шаге $n > 0$ строим две дополнительные точки

$$\begin{aligned} c_{n-1} &= x_{n-1} - \tfrac{\varepsilon}{2}, \\ d_{n-1} &= x_{n-1} + \tfrac{\varepsilon}{2}. \end{aligned}$$

Если $f(c_{n-1}) < f(d_{n-1})$, то положим

$$\begin{aligned} a_n &= a_{n-1}, \\ b_n &= d_{n-1}, \end{aligned}$$

иначе положим

$$\begin{aligned} a_n &= c_{n-1}, \\ b_n &= b_{n-1}. \end{aligned}$$

5. Так как вычисления проводятся приближенно с накоплением вычислительной погрешности, то на каждом шаге $n > 0$ должно проверяться выполнение неравенства

$$a_{n-1} < c_{n-1} < d_{n-1} < b_{n-1}.$$

Если это неравенство не выполняется, то вычислительный процесс прекращается. Принимается решение о том, что решение задачи не может быть найдено.

Следующее утверждение позволяет оценить количество шагов метода деления отрезка пополам, которые необходимо выполнить для нахождения решения задачи.

Утверждение. Неравенство $l_n < 2\varepsilon$ выполняется тогда и только тогда, когда

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

Доказательство. Действительно, при $n > 0$ имеет место равенство

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

Следовательно неравенство $l_n < 2\varepsilon$ может быть переписано в виде

$$2^n > \frac{b-a-\varepsilon}{\varepsilon}.$$

Метод золотого сечения

1. В рамках данного метода строится последовательность вложенных отрезков $[a_n, b_n]$ $(n \geqslant 0)$. Середины этих отрезков $x_n = \frac{a_n+b_n}{2}$ рассматриваются в качестве приближений к точке минимума функции.

2. Для каждого построенного отрезка оценивается его длина $l_n = b_n - a_n$. Если $l_n < 2\varepsilon$, то процесс построения завершается. Число x_n является приближением к точке минимума функции с абсолютной погрешностью ε.

3. Положим

$$\begin{aligned} a_0 &= a, \\ b_0 &= b. \end{aligned}$$

4. Внутри каждого отрезка $[a_n, b_n]$ рассматриваются две различные точки

$$\begin{aligned} c_n &= a_n + r(b_n - a_n), \\ d_n &= b_n - r(b_n - a_n), \end{aligned}$$

где $r = \frac{3-\sqrt{5}}{2}$.

5. Только для отрезка $[a_0, b_0]$ вычисляются обе точки c_0 и d_0 и значения функции в этих точках $f(c_0)$ и $f(d_0)$.

При $n > 0$ вычисляется только одна из точек c_n или d_n. Соответственно вычисляется либо $f(c_n)$, либо $f(d_n)$.

6. Переход от отрезка $[a_{n-1}, b_{n-1}]$ к вложенному отрезку $[a_n, b_n]$ осуществляется по следующим правилам.

Если $f(c_{n-1}) < f(d_{n-1})$, то положим

$$\begin{aligned} a_n &= a_{n-1}, \\ b_n &= d_{n-1}, \\ d_n &= c_{n-1}, \end{aligned}$$

иначе положим

$$\begin{aligned} a_n &= c_{n-1}, \\ b_n &= b_{n-1}, \\ c_n &= d_{n-1}. \end{aligned}$$

7. Так как вычисления проводятся приближенно с накоплением вычислительной погрешности, то как и в методе деления отрезка пополам для каждого $n > 0$ должно проверяться выполнение неравенства

$$a_{n-1} < c_{n-1} < d_{n-1} < b_{n-1}.$$

Если это неравенство не выполняется, то вычислительный процесс прекращается. Принимается решение о том, что решение задачи не может быть найдено.

Утверждение. Неравенство $l_n < 2\varepsilon$ выполняется тогда и только тогда, когда

$$n > \log_{\frac{\sqrt{5}-1}{2}}\left(\frac{2\varepsilon}{b-a}\right).$$

Доказательство. Действительно, при $n > 0$ имеет место равенство

$$l_n = l_{n-1}(1 - r) = (b - a)(1 - r)^n.$$

Следовательно неравенство $l_n < 2\varepsilon$ может быть переписано в виде

$$(1 - r)^n < \frac{2\varepsilon}{b - a}.$$

Замечание. Метод золотого сечения требует выполнения большего количества шагов, чем метод деления отрезка пополам. Однако, вычислительную сложность шага (за исключением первого) метода золотого сечения меньше.

В методе деления отрезка пополам на каждом шаге строятся две внутренние точки и вычисляются значения функции в этих двух точках.

В методе золотого сечения на каждом шаге (кроме первого) заново строится только одна внутренняя точка. В качестве второй может быть выбрана внутренняя точка с предыдущего шага. Соответственно, требуется только один раз вычислить значение функции.

Метод парабол

Точкой экстремума параболы является ее вершина. Предположим, что парабола проходит через три различные точки (x_1, f_1), (x_2, f_2), (x_3, f_3). Тогда ее вершина вычисляется по формуле

$$u = x_2 - \frac{(x_2 - x_1)^2(f_2 - f_3) - (x_2 - x_3)^2(f_2 - f_1)}{2[(x_2 - x_1)(f_2 - f_3) - (x_2 - x_3)(f_2 - f_1)]}.$$

Если $x_1 < x_2 < x_3$, а также $f_2 < f_1$ и $f_2 < f_3$, тогда вершина параболы u гарантированно попадает в интервал (x_1, x_3).

В рамках метода парабол целевая функция приближается параболой, соответственно, в качестве приближения к точке минимума выбирается вершина параболы.

Описание метода.

1. Строится последовательность вложенных отрезков $[a_n, b_n]$ $(n \geqslant 0)$, а также последовательность точек $x_n \in [a_n, b_n]$, которые интерпретируются как приближения к точке минимума.

2. Положим

$$\begin{aligned} a_0 &= a, \\ b_0 &= b, \\ e_0 &= \tfrac{a+b}{2}, \\ x_0 &= \tfrac{a+b}{2}. \end{aligned}$$

3. Построим параболу, проходящую через точки $(a_n, f(a_n))$, $(e_n, f(e_n))$ и $(b_n, f(b_n))$. Обозначим через x_{n+1} вершину этой параболы.

4. Если парабола не может быть построена, или $x_{n+1} \notin [a_n, b_n]$, то вычисления прекращаются. Решение не может быть найдено.

5. Если $|x_{n+1} - x_n| < \varepsilon$, то считаем, что x_{n+1} – искомое приближение к точке минимума функции.

6. Положим

$$\begin{aligned} c_{n+1} &= \min\{e_n, x_{n+1}\}, \\ d_{n+1} &= \max\{e_n, x_{n+1}\}. \end{aligned}$$

7. Если $f(c_{n+1}) < f(d_{n+1})$, то положим

$$\begin{aligned} a_{n+1} &= a_n, \\ b_{n+1} &= d_{n+1}, \\ e_{n+1} &= c_{n+1}, \end{aligned}$$

иначе положим

$$\begin{aligned} a_{n+1} &= c_{n+1}, \\ b_{n+1} &= b_n, \\ e_{n+1} &= d_{n+1}. \end{aligned}$$

Замечание. В методе парабол не может быть гарантирована сходимость длин вложенных отрезков к нулю. Поэтому через эти длины нельзя оценить абсолютную погрешность приближения к точке минимума функции.

Комбинированный метод Брента

В малой окрестности точки минимума функции метод парабол показывает высокую скорость сходимости, превосходящую скорости сходимости методов деления отрезка пополам и золотого сечения. Однако на больших интервалах поиска этот метод может вести себя нестабильно. Метод Брента комбинирует в себе метод парабол и метод золотого сечения, совмещая в себе высокую скорость сходимость одного и стабильность другого метода.

Описание метода.

1. В процессе вычислений используются шесть основных параметров a, b, x, w, v, u и два вспомогательных d_{cur}, d_{prv}.

2. Основные параметры:

- a, b – соответственно, левая и правая граница поиска точки минимума функции;

- x – точка, в которой функция принимает наименьшее из всех вычисленных на текущий момент значений;

- w – точка, в которой функция либо принимает второе снизу из всех вычисленных на текущий момент значений, либо принимает значение $f(x)$;

- v – предыдущее значение w;

- u – текущее приближение к точке минимума функции;

Начальные значения:

$$x, w, v \coloneqq a + r(b-a),$$

где $r = \frac{3-\sqrt{5}}{2}$.

3. Вспомогательные параметры:

- d_{cur} – принимает значение $|u-x|$;

- d_{prv} – предыдущее значение d_{cur}.

Начальные значения:

$$d_{\text{cur}}, d_{\text{prv}} \coloneqq b - a.$$

4. В методе Брента парабола строится по трем точкам x, w, v, которые представляют собой три наилучших приближения к точке минимума функции. Если парабола не может быть построена или вершина параболы отвергается в качестве очередного приближения к точке минимума функции, то выполняется шаг соответствующий методу золотого сечения.

5. Вершина параболы может быть отвергнута в двух случаях. Первый случай, когда $u \notin [a,b]$. Вершина вышла за границы отрезка поиска.

Второй случай, когда $|u-x| > \frac{d_{\text{prv}}}{2}$. Вершина параболы «сильно» отдалилась от точки текущего наименьшего значения функции. Этот критерий представляет собой так называемую эвристику, которую формально обосновать нельзя. Целесообразность эвристики подтверждается положительным опытом ее практического использования.

6. Приведем процедуру, реализующую комбинированный метод Брента.

```
x, w, v := a + r(b − a)
d_cur, d_prv := b − a
пока не превышен лимит шагов
    если max{x − a, b − x} < ε то
        x – найденное решение (остановка)
    g := d_prv/2, d_prv := d_cur
    u := вершина параболы, построенной по x, w, v
    если u = None или u ∉ [a, b] или |u − x| > g то
        если x < (a + b)/2 то
            u := x + r(b − x), d_prv := b − x
        иначе
            u := x − r(x − a), d_prv := x − a
    d_cur := |u − x|
    если f(u) > f(x) то
        если u < x то
            a := u
        иначе
            b := u
        если f(u) ⩽ f(w) или w = x то
            v := w, w := u
        иначе
            если f(u) ⩽ f(v) или v = x или v = w то
                v := u
    иначе
        если u < x то
            b := x
        иначе
            a := x
        v := w, w := x, x := u
```

Метод перебора

1. Разобьем отрезок $[a,b]$ на $n > \frac{b-a}{\varepsilon}$ равных частей. Положим

$$\begin{aligned} h &= \tfrac{b-a}{n}, \\ x_j &= a + jh \quad (j = 0, 1, \ldots, n). \end{aligned}$$

Длина каждого полученного подотрезка $[x_{j-1}, x_j]$ $(j > 0)$ меньше ε.

2. Найдем точку x_m такую, что

$$f(x_m) = \min_{0 \leqslant j \leqslant n} f(x_j),$$

Число x_m является приближением к точке минимума функции с абсолютной погрешностью ε.

Реализации вычислительной процедуры

Вычислительная процедура должна иметь следующие входные и выходные параметры.

Входные параметры.

1. `f` (тип `double (*)(double)`) – указатель на функцию, для которой ищется точка минимума.
2. `a` (тип `double`) – левая граница отрезка поиска точки минимума.
3. `b` (тип `double`) – правая граница отрезка поиска точки минимума.
4. `e` (тип `double`) – требуемая точность.
5. `N` (тип `int`) – максимальное количество шагов. Для метода перебора – максимальное количество отрезков разбиения.
6. `tr` (тип `int`) – ненулевое значение предписывает выводить информационные сообщения о состоянии шага вычислительного процесса.

Выходные параметры.

1. `st` (тип `int`) – статус выполнения процедуры. Возможные значения:
   - 0 – решение было успешно найдено;
   - -1 – некорректные входные параметры (значение `a` должно быть меньше `b`, значение `e` должно быть положительным);
   - -2 – превышен лимит шагов;
   - -3 – в процессе вычислений были получены некорректные промежуточные значения.
2. `x` (тип `double`) – найденное приближение к точке минимума функции.
3. `fx` (тип `double`) – значение функции в точке `x`.
4. `n` (тип `int`) – количество выполненных шагов.
5. `tn` (тип `int`) – теоретическая оценка параметра `n` (для методов деления отрезка пополам и золотого сечения).
6. `pn` (тип `int`) – в случае метода Брента количество шагов выполненных по правилу метода парабол.
7. `gn` (тип `int`) – в случае метода Брента количество шагов выполненных по правилу метода золотого сечения.

При реализации вычислительной процедуры нужно стремиться минимизировать количество вызовов целевой функции. Например, в методе золотого сечения на каждом шаге после первого допускается вызвать целевую функцию только один раз.

Все выходные параметры должны быть оформлены в виде струтуры. Ниже приведен возможный вариант прототипа функции, реализующей метод золотого сечения.

```
typedef double (*fun_t)(double);

typedef struct {
    int    st;
    double x;
    double fx;
    int    n;
    int    tn;
} res_t;

void golden(fun_t  f,
            double a,
            double b,
            double e,
            int    N,
            int    tr,
            res_t *p);
```

Интерфейс программы

При запуске программы все входные параметры передаются через аргументы командной строки. Запуск программы без аргументов командной строки приводит к выводу на экран инструкции по использованию программы.

```
$ ./prog
Usage: ./prog fn a b e N [trace]
fn:
1.  x (x - 2)
...
8.  |sin(x^2)|
...
16. 0.2 x ln(x) + (x - 2.3)^2
```

Основными входными параметрами являются: `fn` – номер тестируемой функции, `a`, `b` – границы отрезка поиска точки минимума, `e` – требуемая точность.

Пример запуска программы.

```
$ ./prog 8 1.5 2.0 1e-12 100
status              : 0
xmin                : 1.772453850905959e+00
f(xmin)             : 1.570621070559757e-12
n                   : 39
tn                  : 39
|xmin - sqrt(Pi)|   : 4.432010314303625e-13
|xmin - sqrt(2Pi)|  : 7.341744237250412e-01
|xmin - Pi|         : 1.369138802683834e+00
|xmin - 2Pi|        : 4.510731456273627e+00
```

Программа должна выводить на экран значения выходных параметров вычислительной процедуры. В случае успешного вычисления точки минимума функции должны также должны выводиться растояния от этой точки до констант $\sqrt{\pi}$, $\sqrt{2\pi}$, π и 2π.

Должен быть предусмотрен необязательный параметр `trace`. Его использование приводит к выводу информации о состоянии каждого шага вычислительного процесса. Должны выводиться номер шага, текущее приближение к точке минимума функции, длина текущего отрезка поиска точки минимума функции и модуль разности значений функции на концах этого отрезка.

Пример запуска программы.

```
$ ./prog bis 8 1.5 2.0 1e-12 100 trace
  n |            xm            |            dx            |            df
----------------------------------------------------------------------------------
  0 | 1.750000000000000e+00 | 5.000000000000000e-01 | 2.127070157999300e-02
  1 | 1.874999999999750e+00 | 2.500000000005000e-01 | 6.777922785587939e-01
  2 | 1.812499999999875e+00 | 1.250000000007501e-01 | 2.863617319573659e-01
  3 | 1.781249999999937e+00 | 6.250000000087508e-02 | 6.406073455382359e-02
...
 36 | 1.772453850902776e+00 | 8.276046514765767e-12 | 1.942737150346067e-11
 37 | 1.772453850904595e+00 | 4.638067707674054e-12 | 6.531464938624700e-12
 38 | 1.772453850905504e+00 | 2.819078304128197e-12 | 8.328961160179163e-14
 39 | 1.772453850905959e+00 | 1.909583602355269e-12 | 3.140798051909663e-12
status                    : 0
xmin                      : 1.772453850905959e+00
f(xmin)                   : 1.570621070559757e-12
n                         : 39
tn                        : 39
|xmin - sqrt(Pi)|         : 4.432010314303625e-13
|xmin - sqrt(2Pi)|        : 7.341744237250412e-01
|xmin - Pi|               : 1.369138802683834e+00
|xmin - 2Pi|              : 4.510731456273627e+00
```

Тесты

Список тестовых функций:

$$
\begin{aligned}
f_1(x) &= x(x-2),\\
f_2(x) &= |f_1(x)|,\\
f_3(x) &= g(f_1(x)),\\
f_4(x) &= |x^3|,\\
f_5(x) &= x(x-2)(x-3),\\
f_6(x) &= |f_5(x)|,\\
f_7(x) &= g(f_5(x)),\\
f_8(x) &= |\sin(x^2)|,\\
f_9(x) &= g(\sin(x^2)),\\
f_{10}(x) &= |e^{0.1x}\sin(x)|,\\
f_{11}(x) &= g(e^{0.1x}\sin(x)),\\
f_{12}(x) &= -5x^5+4x^4-12x^3+11x^2-2x+1,\\
f_{13}(x) &= -\ln^2(x-2)+\ln^2(10-x)-x^{0.2},\\
f_{14}(x) &= -3x\sin(0.75x)+e^{-2x},\\
f_{15}(x) &= e^{3x}+5e^{-2x},\\
f_{16}(x) &= 0.2x\ln(x)+(x-2.3)^2,
\end{aligned}
$$

где

$$
g(x) = \begin{cases} x, & \text{если} \quad x > 0;\\ 0, & \text{иначе.} \end{cases}
$$

Предлагается провести следующие тесты (целевая функция, отрезок поиска точки минимума):

1. функция f_1 на отрезке $[-1.5, 20.1]$;

2. функция f_2 на отрезке $[-20.1, 0.9]$;

3. функция f_3 на отрезке $[-1.5, 20.1]$;

4. функция $f_4(x)$ на отрезке $[-1.5, 20.1]$;

5. функция f_5 на отрезке $[1.1, 20.1]$ (ответ $\frac{5+\sqrt{7}}{3} \approx 2.5485837703548637$);

6. функция f_6 на отрезке $[1.1, 2.5]$;

7. функция f_7 на отрезке $[1.1, 20.1]$;

8a. функция f_8 на отрезке $[1.5, 2.0]$ (ответ $\sqrt{\pi} \approx 1.7724538509055159$);

8b. функция f_8 на отрезке $[2.3, 2.7]$ (ответ $\sqrt{2\pi} \approx 2.5066282746310002$);

9. функция f_9 на отрезке $[1.5, 2.0]$;

10a. функция f_{10} на отрезке $[2.0, 4.5]$ (ответ $\pi \approx 3.1415926535897931$);

10b. функция f_{10} на отрезке $[4.9, 7.5]$ (ответ $2\pi \approx 6.2831853071795862$);

11. функция f_{11} на отрезке $[2.5, 7.5]$;

12. функция f_{12} на отрезке $[-0.5, 0.5]$ (ответ $\approx 0.1098599173884884$);

13. функция f_{13} на отрезке $[6, 9.9]$ (ответ $\approx 9.206243215935171$);

14. функция f_{14} на отрезке $[0, 2\pi]$ (ответ ≈ 2.70647559349924);

15. функция f_{15} на отрезке $[0, 1]$ (ответ $\approx 0.2407945669107351$);

16. функция f_{16} на отрезке $[0.5, 2.5]$ (ответ $\approx 2.124639781251913$).

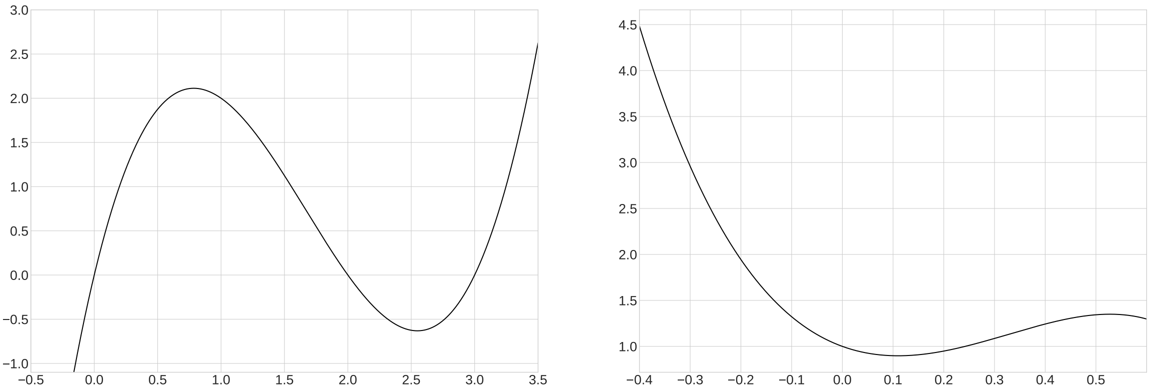

Рис. 2: График функции $f_5(x) = x(x-2)(x-3)$ (слева) и график функции $f_{12}(x) = -5x^5 + 4x^4 - 12x^3 + 11x^2 - 2x + 1$ (справа).

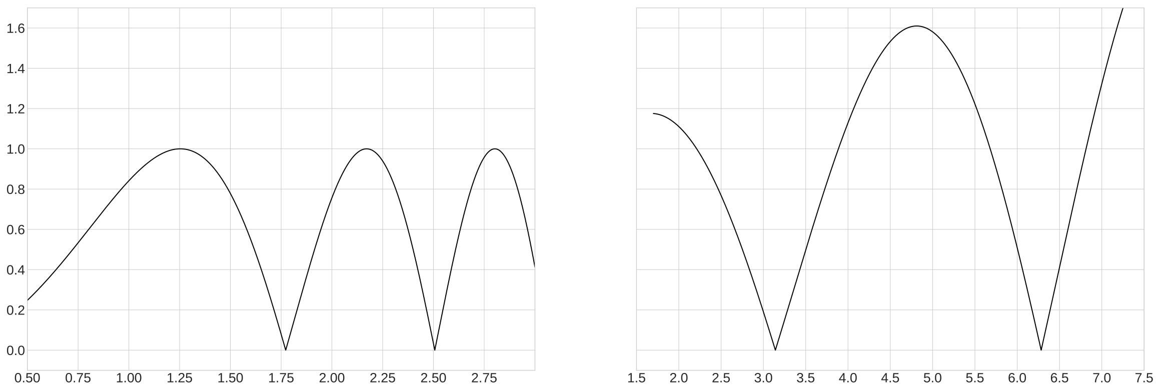

Рис. 3: График функции $f_8(x) = |\sin(x^2)|$ (слева) и график функции $f_{10}(x) = |e^{0.1x}\sin(x)|$ (справа).

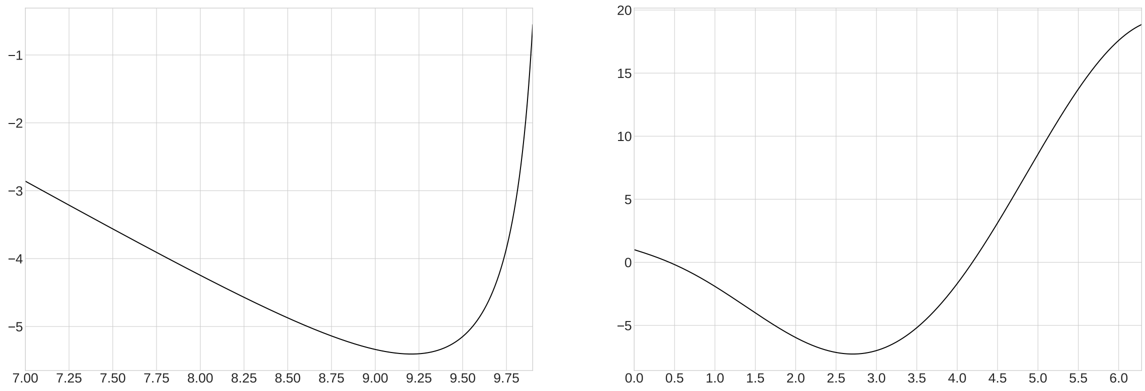

Рис. 4: График функции $f_{13}(x) = -\ln^2(x-2) + \ln^2(10-x) - x^{0.2}$ (слева) и график функции $f_{14}(x) = -3x\sin(0.75x) + e^{-2x}$ (справа).

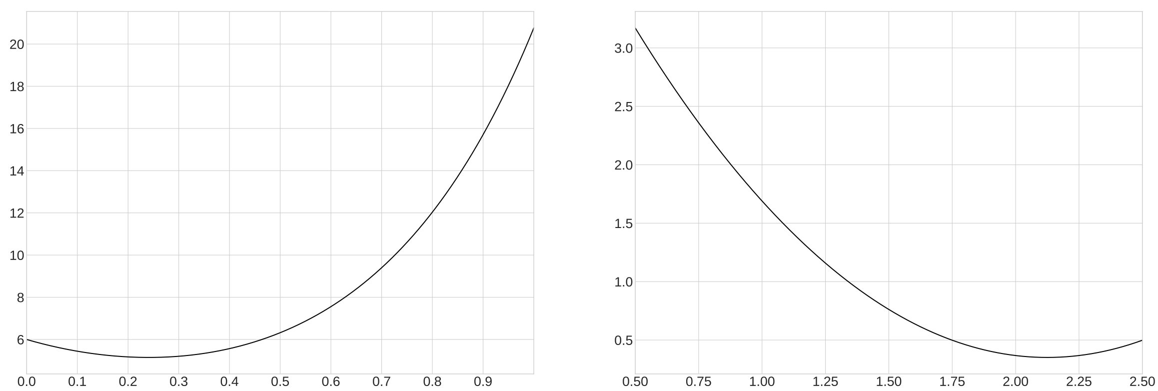

Рис. 5: График функции $f_{15}(x) = e^{3x} + 5e^{-2x}$ (слева) и график функции $f_{16}(x) = 0.2x\ln(x) + (x-2.3)^2$ (справа).