Лабораторная работа № 1

Исследование методов безусловной одномерной оптимизации

<u>Целью работы</u> являются изучение и моделирование в среде MATLAB различных методов одномерной оптимизации и сравнение эффективности их применения для конкретных целевых функций.

1. Краткие теоретические сведения

1.1. Постановка задачи одномерной безусловной оптимизации

Поиск экстремума функции одной переменной имеет определенный интерес, так как относительно часто встречается в инженерной практике. Правильная организация одномерного поиска определяет успех решения всей задачи. Кроме того, одномерная оптимизация является составной частью многих методов многомерной оптимизации.

В силу эквивалентности двух типов оптимизационных задач (максимизации и минимизации) далее рассматривается задача одномерной минимизации. Задача поиска минимума целевой функции формулируется в виде:

$$x^* = arg\,\min f(x),\ x \in X,$$

где X – множество допустимых решений, среди которых ищется точка x^*, дающая минимум $f(x)$ целевой функции.

Другая распространенная запись задачи минимизации:

$$f(x) \to \min_{x \in X}.$$

Когда $X=R$, где R – множество вещественных чисел, то говорят об одномерной безусловной задаче минимизации, т.е. целевая функция $f(x)$ имеет только один аргумент и множество X есть вся вещественная ось чисел.

В методах одномерной оптимизации вместо $X=R$ рассматривается отрезок $X=[a, b]$, содержащий искомое решение x^*. Такой отрезок называется отрезком локализации (неопределенности). Относительно целевой функции $f(x)$ часто предполагается, что она унимодальная. Если ограничиваться рассмотрением лишь непрерывных функций $f(x)$, то свойство унимодальности функции означает наличие у нее единственного локального минимума и этот минимум достигается в точке $x=x^*$. Кроме того, в ряде методов предполагается, что целевая функция $f(x)$ является выпуклой на множестве X. Непрерывная строго выпуклая функция является унимодальной. Однако не всякая унимодальная функция является выпуклой или непрерывной. Наконец, точность поиска точки минимума δ задается заранее.

1.2. Метод перебора на равномерной сетке

Идея метода состоит в переборе некоторого множества точек отрезка $X=[a, b]$ и вычислений соответствующих значений целевой функции в этих точках. Точка с минимальным значением целевой функции объявляется решением задачи одномерной оптимизации.

Отрезок $X=[a, b]$ разбивается на n равных частей, $n \geq (b-a)/\delta$, и вычисляются точки деления $x_i = a + i(b-a)/n$, $i=1,2,3\ldots,n$. Затем вычисляются значения целевой функции $f(x)$ в каждой точке деления.

Пусть окажется, что в некоторой точке x_m выполняется условие

$$f(x_m) = \min_{i=\overline{0,n}} \{f(x_i)\}.$$

Тогда очевидно, что $x^* \in [x_{m-1}, x_{m+1}]$, то есть $[x_{m-1}, x_{m+1}]$ – отрезок локализации. Центр этого отрезка принимается за аппроксимацию x^*: $x^* \approx x_m$. Абсолютная ошибка в определении точки минимума не превышает половины длины отрезка локализации и оказывается меньше заданной точности:

$$\left| x^* - x_m \right| \le \frac{(b-a)}{n} \le \delta.$$

Одним из важных показателей эффективности методов одномерной оптимизации принято считать число обращений к целевой функции. Очевидно, что для метода перебора на равномерной сетке число обращений к целевой функции будет равно $N=n+1$.

1.3. Метод деления отрезка пополам (метод дихотомии)

Более эффективными являются методы, в которых выбор очередной точки x_i производится на основании информации о функции $f(x)$, полученной на предыдущих итерациях; такие методы называют последовательными.

Процедура поиска включает следующие шаги.

1. Вычисляются координата средней точки $x_0 = \dfrac{b+a}{2}$ отрезка $[a, b]$, его длина $l = b - a$ и значение целевой функции в средней точке $f(x_0)$.

2. Определяются координаты двух пробных точек $x_1 = a + l/4$, $x_2 = b - l/4$, и вычисляются значения целевой функции в этих точках $f(x_1)$, $f(x_2)$.
3. Сравниваются значения $f(x_1)$ и $f(x_0)$.

Возможны ситуации:

а) если $f(x_1)<f(x_0)$, то исключается полуинтервал $(x_0, b]$. Тогда отрезок локализации будет $[a, x_0]$. Средней точкой нового отрезка локализации становится точка x_1. Полагаем $b=x_0$, $x_0= x_1$. Переход к п. 5;

б) если $f(x_1)\geq f(x_0)$, то переход к п. 4.

4. Сравниваются значения $f(x_2)$ и $f(x_0)$.

Возможны ситуации:

а) если $f(x_2)<f(x_0)$, то новый отрезок локализации $[x_0, b]$. Средней точкой отрезка локализации становится точка x_2. Полагаем $a=x_0$, $x_0=x_2$. Переход к п. 5;

б) если $f(x_2)\geq f(x_0)$, то исключаются полуинтервалы $[a, x_1)$ и $(x_2, b]$. Тогда новый отрезок локализации будет $[x_1, x_2]$. Средней точкой отрезка локализации остается точка x_0. Полагаем $a=x_1$, $b= x_2$. Переход к п. 5.

5. Вычисляются новая длина отрезка $l = b - a$ и ошибка вычисления точки минимума $\Delta = l/2$. Если $\Delta \leq \delta$, то поиск заканчивается. В противном случае – переход к п. 2.

Таким образом, как следует из описания алгоритма, на каждом шаге исключается ровно половина отрезка локализации.

Средняя точка отрезка локализации x_0 объявляется точкой минимума $f(x)$. Число обращений к целевой функции $N=2n+1$.

1.4. Метод золотого сечения

Деление отрезка $[a, b]$ на две неравные части $[a, x]$ и $[x, b]$ так, что отношение длины всего отрезка к большей части равно отношению длины большей части к длине меньшей части, называется *золотым сечением* этого отрезка. Квадратное уравнение

$$\frac{b-a}{b-x} = \frac{b-x}{x-a}$$

содержит два решения: $x_1 = a + (1-\lambda)(b-a)$, $x_2 = a + \lambda(b-a)$, где

$\lambda = \dfrac{\sqrt{5}-1}{2} \approx 0{,}618$. Точки x_1, x_2 обладают следующими свойствами:

а) размещаются на одинаковых расстояниях от середины отрезка;

б) точка x_1 является второй точкой золотого сечения отрезка $[a, x_2]$, а точка x_2 – первой точкой золотого сечения отрезка $[x_1, b]$;

в) выполняется условие $x_1 + x_2 = a+b$, откуда следует:

$$x_1 = a+b-x_2,$$

или (*)

$$x_2 = a+b-x_1 .$$

Алгоритм содержит следующие шаги.

1. Определяется отрезок $[a, b]$, задается ошибка δ, вычисляются координаты двух точек золотого сечения x_1, x_2 и значения целевой функции в этих точках $f(x_1)$, $f(x_2)$.
2. Сравниваются значения целевых функций $f(x_1)$ и $f(x_2)$.

Возможны ситуации:

а) если $f(x_1) < f(x_2)$, то исключается полуинтервал $(x_2, b]$, новый отрезок локализации $[a, x_2]$. Тогда полагаем $b = x_2$, $x_2 = x_1$, $f(x_2) = f(x_1)$. По одному из соотношений (*) вычисляются точка золотого сечения x_1 и новое значение целевой функции в этой точке $f(x_1)$. Переход к п. 3;

б) если $f(x_1) \geq f(x_2)$, то исключается полуинтервал $[a, x_1)$, новый отрезок локализации $[x_1, b]$. Тогда полагаем $a=x_1$, $x_1=x_2$, $f(x_1)= f(x_2)$. По одному из соотношений (*) вычисляются новое значение точки золотого сечения x_2 и новое значение целевой функции в этой точке $f(x_2)$. Переход к п. 3.

3. Вычисляется ошибка $\Delta = (b-a)/2$. Если $\Delta < \delta$, то $x^* \approx \dfrac{a+b}{2}$ и останов поиска. Если $\Delta > \delta$, то переход к п. 2.

Длина отрезка локализации на итерации с номером n равна $L_n = \lambda^n (b-a)$. Число обращений к целевой функции $N=n+1$.

1.5. Метод Фибоначчи

Этот метод применяется, когда число итераций n заранее задано. Метод Фибоначчи, также как и метод золотого сечения, относится к симметричным методам, т.е. точки, в которых выполняются два эксперимента (т.е. два обращения к целевым функциям на каждой итерации), на основе которых происходит уменьшение отрезка локализации, расположены симметрично относительно середины отрезка. Однако выбор точки x_1 происходит по другим соотношениям. Для этого используются числа Фибоначчи: $F_0, F_1, F_2, F_3, \ldots$, где $F_i = F_{i-2} + F_{i-1}$ $(i = 2,3,\ldots)$ и $F_0 = F_1 = 1$.

Точка x_1 определяется из соотношения:

$$\frac{длина[a, x_1]}{длина[a, b]} = \frac{F_{n-2}}{F_n},$$

т.е. $x_1 = a + (b-a)\frac{F_{n-2}}{F_n}$. Точка x_1 делит отрезок $[a,b]$ на две неравные части. Отношение малого отрезка к большему равно F_{n-2} / F_{n-1}. Точка x_2 определяется как точка, симметричная по отношению к точке x_1 относительно середины отрезка $[a,b]$. Поэтому $x_2 = b - (b-a)\frac{F_{n-2}}{F_n} = a + (b-a)\frac{F_{n-1}}{F_n}$. При этом будет выполняться условие $x_1 < x_2$.

В результате экспериментов в точках x_1 и x_2 у нас получится отрезок локализации $[a, x_2]$, содержащий точку x_1, или отрезок локализации $[x_1, b]$, содержащий точку x_2. Остающаяся точка делит новый отрезок неопределённости на две неравные части в отношении:

$\frac{меньшая\_часть}{большая\_часть} = F_{n-3} / F_{n-2}$. То есть в методе Фибоначчи остающаяся точка делит отрезок на две неравные части в пропорциях, определяемых числами Фибоначчи. Тогда на *k*-м шаге это отношение равно: $\frac{меньшая\_часть}{большая\_часть} = \frac{F_{n-k-2}}{F_{n-k}}$, а длины отрезков равны: $меньшая\_часть = \frac{F_{n-k-1}}{F_n}(b-a)$ и $большая\_часть = \frac{F_{n-k}}{F_n}(b-a)$.

Для того чтобы, в свою очередь, уменьшить получившийся отрезок локализации, надо определить симметричную точку относительно середины отрезка и произвести эксперимент в ней. Этот процесс продолжается, пока не будет проведено *n* итераций.

Алгоритм включает следующие шаги.

1. Задаются значения параметров a, b, n. Вычисляются числа Фибоначчи $F_0, F_1, ..., F_n$. Определяется:

$$x_1 = a + (b-a)F_{n-2} / F_n,\ x_2 = a + (b-a)F_{n-1} / F_n,$$
$$y_1 = f(x_1), y_2 = f(x_2).$$

2. а) Если $y_1 \le y_2$, то полагают $b = x_2, x_2 = x_1, y_2 = y_1$ и вычисляют $x_1 = a + b - x_2, y_1 = f(x_1)$.

б) Если $y_1 > y_2$, то полагают $a = x_1, x_1 = x_2, y_1 = y_2$ и вычисляют $x_2 = a + b - x_1, y_2 = f(x_2)$.

Повторить шаг 2 $n-2$ раза.

3. Если $y_1 < y_2$, то полагают $\tilde{x} = x_1$ и $\tilde{y} = y_1$. Если $y_1 \ge y_2$, то полагают $\tilde{x} = x_2$ и $\tilde{y} = y_2$. Закончить поиск.

Длина отрезка локализации в методе Фибоначчи $L_n \approx (b-a)/F_n$.

1.6. Методы оптимизации с использованием производной

Все рассмотренные ранее алгоритмы представляют алгоритмы нулевого порядка, т.е. используют информацию только о целевой функции. Если дополнить условие унимодальности функции на отрезке [*a*, *b*] условием ее дифференцируемости, то получим методы первого, второго и т.д. порядков. Поскольку функция *f*(*x*) унимодальная, то выполняется условие $f'(x^*)=0$, тогда решение задачи минимизации первого порядка сводится к численному решению уравнения $f'(x)=0$. Метод средней точки (метод Больцано) использует схему дихотомии (деления отрезка пополам) и содержит следующие шаги.

Пусть заданы отрезок локализации [a, b] и некоторое число ε>0.

1. Вычисляются значения производной целевой функции в крайних точках отрезка локализации $\dot{f}(a)$ и $\dot{f}(b)$. При этом должно выполняться условие $\dot{f}(a)<0$ и $\dot{f}(b)>0$.
2. Вычисляются средняя точка $x_0=\dfrac{a+b}{2}$ и значение производной функции в этой точке $\dot{f}(x_0)$.
3. Если выполняется условие $\left|\dot{f}(x_0)\right|\le\varepsilon$, то $x^*\approx x_0$ и поиск заканчивается. В противном случае – переход к п. 4.
4. Если $\dot{f}(x_0)<0$, то исключается полуинтервал [a, x_0), новый отрезок локализации [x_0, b], полагаем a=x_0. Переход к п. 2. Если окажется, что $\dot{f}(x_0)>0$, то исключается полуинтервал (x_0, b], новый отрезок локализации [a, x_0], полагаем b=x_0. Переход к п. 2.

1.7. Методы поиска, основанные на аппроксимации целевой функции

Суть этих методов заключается в том, что по полученной в ходе вычислений информации строится аппроксимирующая функция и её минимум принимается за точку очередного вычисления. Такие методы дают хорошие результаты при минимизации достаточно гладких унимодальных функций.

1.7.1. Метод касательных

Пусть функция $f(x)$ выпукла и дифференцируема на отрезке $[a,b]$. Идея метода состоит в следующем. Пусть $[a,b]$ - отрезок локлизации и $f(a), f'(a), f(b), f'(b)$ - результаты вычислений в точках a и b. По этой информации строится аппроксимирующая функция, представляющую собой кусочно-линейную функцию, состоящую из касательной $L_a(x) = f(a) + f'(a)(x-a)$ к кривой целевой функции $f(x)$ в точке a и касательной $L_b(x) = f(b) + f'(b)(x-b)$ к $f(x)$ в точке b (рис.1).

Полученная аппроксимирующая функция есть ломаная, состоящая из прямой $L_a(x)$ на $[a,c]$ и $L_b(x)$ на $[c,b]$, где c – точка пересечения касательных. Легко заметить, что при $f(a) < 0$ и $f(b) > 0$ минимум аппроксимирующей функции достигается в точке c. Значение точки пересечения c можно определить по формуле:

$$c = \frac{(bf'(b) - af'(a)) - (f(b) - f(a))}{f'(b) - f'(a)}.$$

В точке c производятся вычисления $f(c)$ и $f'(c)$. Если $f'(c)=0$, то решением задачи будет $x^*=c$. Если $f'(c)>0$, то в качестве следующего отрезка неопределённости будет $[a,c]$. Если $f'(c)<0$, то – отрезок $[c,b]$. Процесс повторяется до тех пор, пока $f'(c)=0$ или отрезок локализации не достигнет заданной точности.

Алгоритм содержит следующие шаги.

1. Заданы a,b,ε. Вычислить:

$y_1=f(a),\quad y_2=f(b),\quad z_1=f'(a),\quad z_2=f'(b)$.

2. Если $b-a\le 2\varepsilon$, то полагаем $\tilde{x}=(a+b)/2,\quad \tilde{y}=f(\tilde{x})$. Поиск окончен. Если $b-a>2\varepsilon$, то вычислить:

$c=\dfrac{(bz_2-az_1)-(y_2-y_1)}{z_2-z_1},\quad y=f(c),\quad z=f'(c)$. Если z=0, то полагаем $\tilde{x}=c,\quad \tilde{y}=y$. Поиск окончен. Если $z<0$, то $a=c,\quad y_1=y,\quad z_1=z$. Если $z>0$, то $b=c,\quad y_2=y,\quad z_2=z$. Повторить шаг 2.

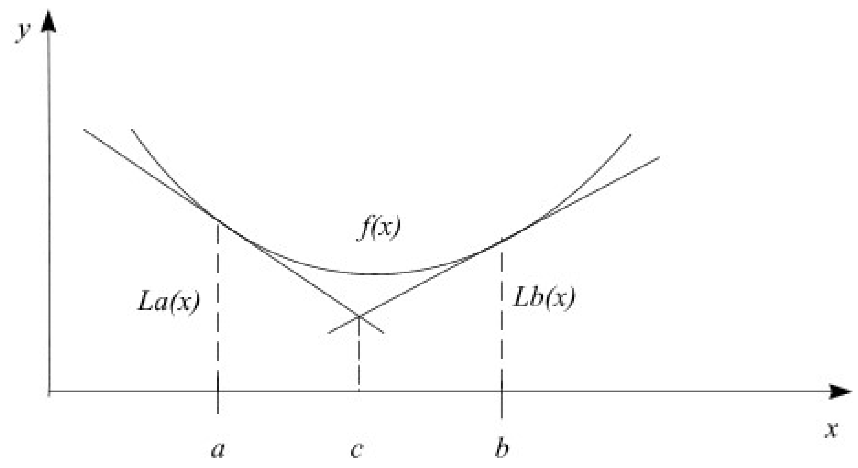

Рис. 1

1.7.2. Метод парабол

Рассмотрим алгоритм квадратичной интерполяции, или метод парабол (т.е. в качестве аппроксимирующей функции используется парабола). Для однозначного задания параболы необходимы три точки. Пусть имеются три точки, для которых выполняется: $a<c<b$, $f(c)\leq f(a)$, $f(c)\leq f(b)$. Так как $[a, b]$ – отрезок локализации и $f(x)$ – унимодальная функция, то найти такую точку c нетрудно. Парабола, проходящая через три точки $(a,f(a))$, $(c,f(c))$, $(b,f(b))$, имеет вид

$$P(x)=\left(\frac{f(b)-f(c)}{(b-c)}+\frac{f(a)-f(c)}{(a-c)}\right)\frac{(x-c)(x-b)}{b-a}+\frac{f(b)-f(c)}{b-c}(x-c)+f(c)\cdot$$

Поскольку $f(x)$ – унимодальная функция, то хотя бы одно из неравенств $f(c)\leq f(a)$, $f(c)\leq f(b)$ строгое и, следовательно, коэффициент при старшем члене $P(x)$ положителен. Тогда $P(x)$ достигает минимума в точке:

$$t=c+\frac{1}{2}\frac{(b-c)^2(f(a)-f(c))-(c-a)^2(f(b)-f(c))}{(b-c)(f(a)-f(c))+(c-a)(f(b)-f(c))},$$

причем $(a+c)/2\leq t\leq(c+b)/2$. Эта точка и выбирается в качестве точки очередного вычисления значения функции.

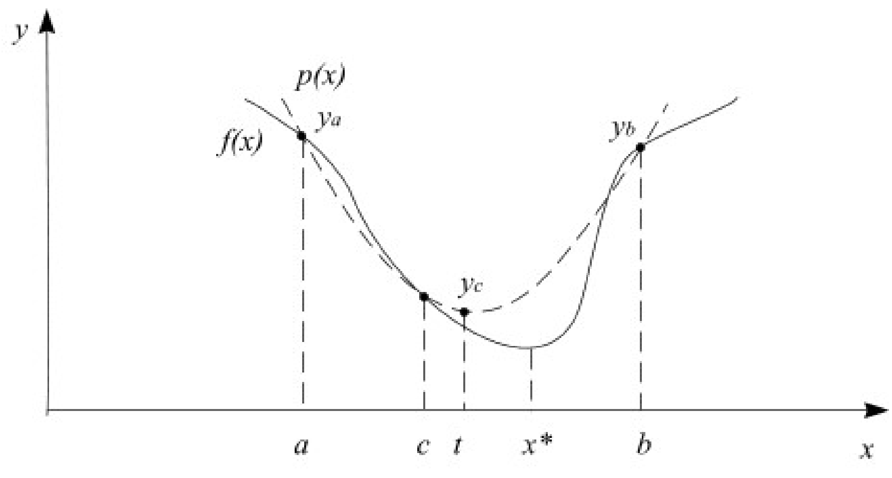

Рис. 2

Если оказалось, что $t=c$, то в качестве точки очередного вычисления выбирается точка $(a+c)/2$. Итак, следующее вычисление проводится в точке:

$$x = \begin{cases} t, если\_t \neq c, \\ (a+c)\big/2, если\_t = c. \end{cases}$$

Определим новый отрезок локализации с лежащей внутри него точкой, для которой выполняются условия, аналогичные условиям, которым удовлетворяла точка c. В силу унимодальности функции $f(x)$ и в зависимости от выполнения или невыполнения условий $x<c$, $f(x)<f(c)$, $f(x)=f(c)$ это будут отрезки с точкой внутри: $[a,c]$ и x; $[x,b]$ и c; $[x,c]$ и $(x+c)/2$; $[a,x]$ и c; $[c,b]$ и x; $[c,x]$ и $(x+c)/2$ (рис. 2). Затем строится парабола, определяется ее минимум, и далее – до тех пор, пока длина отрезка локализации не удовлетворит заданной точности.

Алгоритм содержит следующие шаги.

1. Задаются параметры a, c, b и ε. Вычислить $y_a=f(a)$, $y_c=f(c)$, $y_b=f(b)$.

2. Вычислить $x = \begin{cases} t, если\_t \neq c, \\ (a+c)\big/2, если\_t = c. \end{cases}$, $y=f(x)$,

где $$t = c + \frac{1}{2} \cdot \frac{(b-c)^2 (y_a - y_c) - (c-a)^2 (y_b - y_c)}{(b-c)(y_a - y_c) + (c-a)(y_b - y_c)}.$$

3. а) При $x<c$.

Если $y<y_c$, то $b=c$, $c=x$, $y_b=y_c$, $y_c=y$.

Если $y>y_c$, то $a=x$, $y_a=y$.

Если $y=y_c$, то $a=x$, $b=c$, $c=(x+c)/2$, $y_a=y$, $y_b=y_c$, $y_c=f(c)$.

б) При $x>c$.

Если $y<y_c$, то $a=c$, $c=x$, $y_a=y_c$, $y_c=y$.

Если $y>y_c$, то $b=x$, $y_b=y$.

Если $y=y_c$, то $a=c$, $b=x$, $c=(x+c)/2$, $y_a=y_c$, $y_b=y$, $y_c=f(c)$.

4. Если $b-a\leq\varepsilon$, то закончить поиск, положив $\widetilde{x}=x, \widetilde{y}=y$, иначе перейти к п. 2.

2. Порядок выполнения лабораторной работы

2.1. Изучить предлагаемые методы одномерной безусловной оптимизации, используя дополнительную литературу и конспект лекций, если необходимо.

2.2. В соответствии с вариантом задания, определенным преподавателем, в среде MATLAB составить программы, реализующие методы поиска, и найти точку минимума функции $f(x)$ на отрезке $[a,b]$.

а) Если в варианте задания указано число экспериментов n, то сравнить заданные в варианте методы по получаемой длине отрезка локализации.

б) Если указана точность искомого решения δ, то сравнить методы по числу обращений к целевой функции, понадобившихся для достижения заданной точности.

2.3. Оформить отчет о выполнении задания с приведением условия задачи, алгоритмов и программ указанных методов поиска, таблицы результатов сравнения рассмотренных методов, заключения по результатам сравнения методов.

3. Варианты заданий

3.1. Методы одномерной безусловной оптимизации:

а) метод перебора на равномерной сетке;

б) метод дихотомии (деления интервала пополам);

в) метод золотого сечения;

г) метод Фибоначчи;

д) метод Больцано;

е) метод касательных;

ж) метод парабол.

3.2. Варианты задач

№	***Целевая функция***	***Отрезок [a,b]***	***Точность δ или число итераций (экспериментов) n***
1	$x^2+6\cdot e^{0,15x}$	[-1,0]	n =22
2	$x^2+4\cdot e^{-0,25x}$	[0,1]	n = 23
3	$x^4+0,4\cdot arctg5x$	[-1,0]	n =20
4	$x^4-1,5arctgx$	[0,1]	n =21
5	$x^2+8\cdot e^{0,55x}$	[-2,0]	$\delta=10^{-3}$
6	$-4x+e^{\lvert x-0,2\rvert}$	[0,2]	$\delta=1.5*10^{-3}$
7	$1,4x+e^{\lvert x-2\rvert}$	[0,2]	$\delta=5*10^{-3}$
8	x^2+e^x	[-1,0]	n =18
9	$\lvert x\rvert+e^{10x}$	[-1,0]	$\delta=10^{-3}$
10	$10\cos x+e^x$	[0,3]	$\delta=5*10^{-4}$