09.04.03 Прикладная информатика

Профиль «Машинное обучение и анализ данных»

Дисциплина «Математические основы анализа данных»

Лекция 6

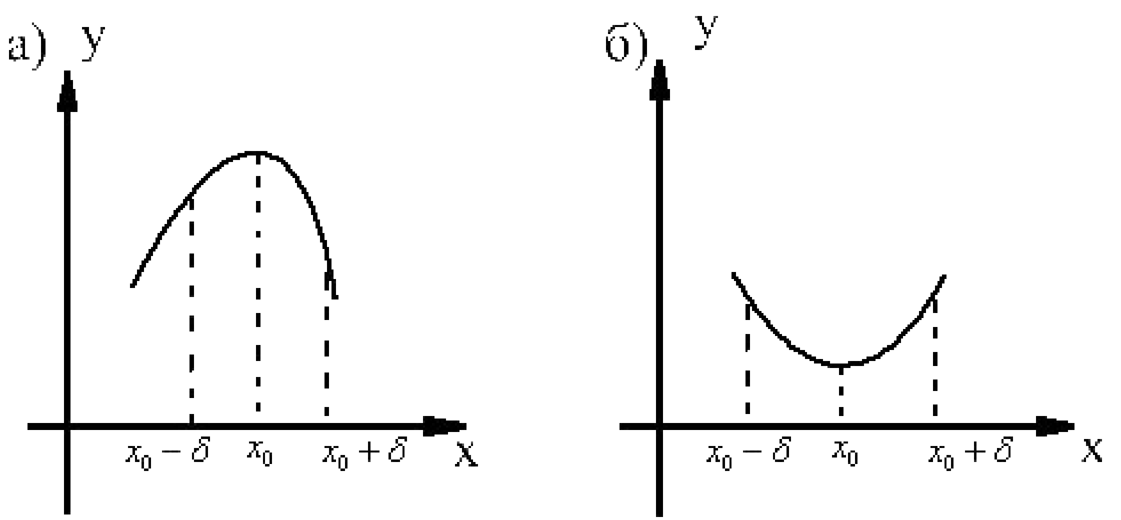
Одномерная оптимизация

План лекции

- Постановка задачи одномерной оптимизации
- Классификация методов оптимизации
- Аналитическое нахождение максимума и минимума функции одной переменной
- Численные методы одномерной оптимизации:
  - *Метод дихотомии (половинного деления)*
  - *Метод золотого сечения*
  - *Метод средней точки*
  - *Метод хорд*
  - *Метод касательных*

- Оптимизация — целенаправленная деятельность, заключающаяся в получении наилучших результатов при соответствующих условиях.
- Оптимизация — в математике, информатике и исследовании операций задача нахождения экстремума (минимума или максимума) целевой функции в некоторой области конечномерного векторного пространства, ограниченной набором линейных и/или нелинейных равенств и/или неравенств.
- Поиски оптимальных решений привели к созданию специальных математических методов и уже в 18 в. были заложены математические основы оптимизации (вариационное исчисление, численные методы и др.).
- Однако до второй половины 20 в. методы оптимизации во многих областях науки и техники применялись очень редко, поскольку практическое использование математических методов оптимизации требовало огромной вычислительной работы, которую без ЭВМ реализовать было крайне трудно.

- Говорят, что функция $f(x)$ имеет в точке x_0 максимум (или минимум), если существует некоторая ε-окрестность $(x_0 — \varepsilon;\ x_0 + \varepsilon)$ в промежутке, где функция определена, что для всех точек этой окрестности выполняется неравенство:
- для максимума: $f(x) \leq f(x_0)$ (рис. а)
- для минимума: $f(x) \geq f(x_0)$ (рис. б)



Постановка задачи оптимизации

- Среди элементов х из множества X, найти такой элемент x^*, который доставляет минимальное (или максимальное) значение $f(x^*)$ заданной функции $f(x)$.
- Минимум и максимум объединяются термином **экстремум**.
- Для того, чтобы корректно поставить задачу оптимизации, необходимо задать:

1. *Допустимое множество* X
2. *Целевую функцию* $f: X \to \mathbf{R}$
3. *Критерий поиска* (max или min).

Классификация методов оптимизации

1) Методы оптимизации классифицируют в соответствии с задачами оптимизации:

- Локальные методы: сходятся к какому-нибудь локальному экстремуму целевой функции.
- Глобальные методы: имеют дело с многоэкстремальными целевыми функциями.

2) Методы численного поиска можно разбить на три большие группы:

- детерминированные;
- случайные (стохастические);
- комбинированные.

Классификация методов оптимизации

3) По критерию размерности допустимого множества, методы оптимизации делят на

- методы *одномерной оптимизации*
- и методы *многомерной оптимизации*.

4) По требованиям к гладкости и наличию у целевой функции частных производных, их также можно разделить на:

- *прямые методы*, требующие только вычислений целевой функции в точках приближений;
- *методы первого порядка*: требуют вычисления первых частных производных функции;
- *методы второго порядка*.

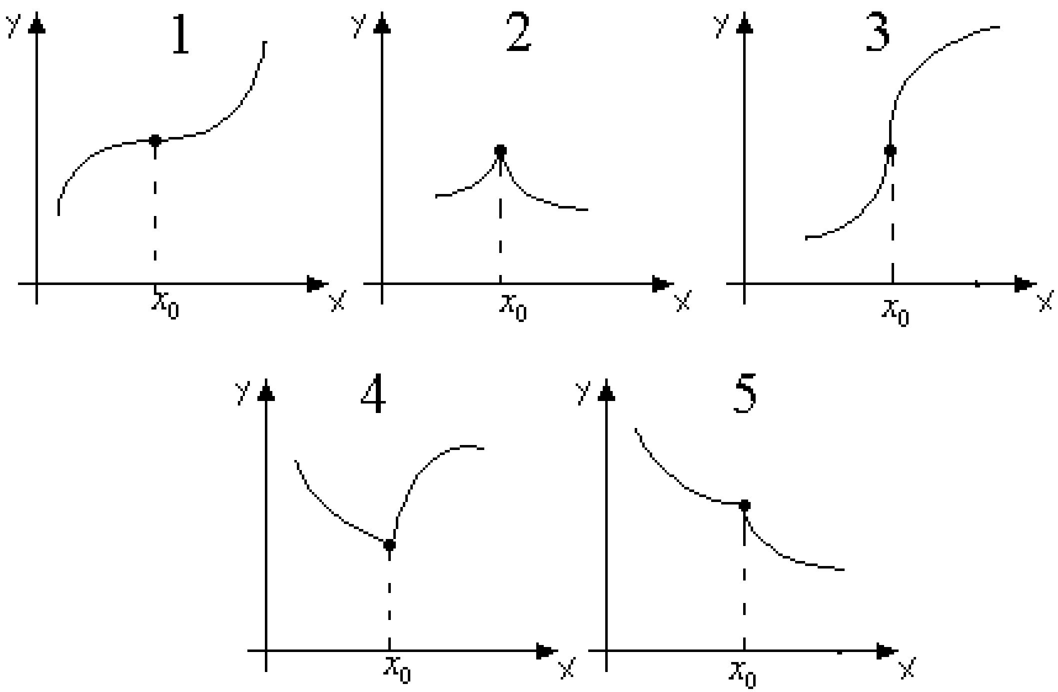
Нахождение максимума и минимума функции одной переменной

- Будем использовать методы классического математического анализа для исследования функции на экстремум.
- **Необходимое условие экстремума (теорема Ферма):**

Пусть функция $f(x)$ определена на некотором промежутке и во внутренней точке *c* этого промежутка принимает наибольшее (наименьшее) значение. Если существует конечная производная $f'(c)$, то необходимо $f'(c) = 0$.

- Если выполняется равенство $f'(x_0) = 0$, то точку x_0 будем называть **стационарной точкой**.

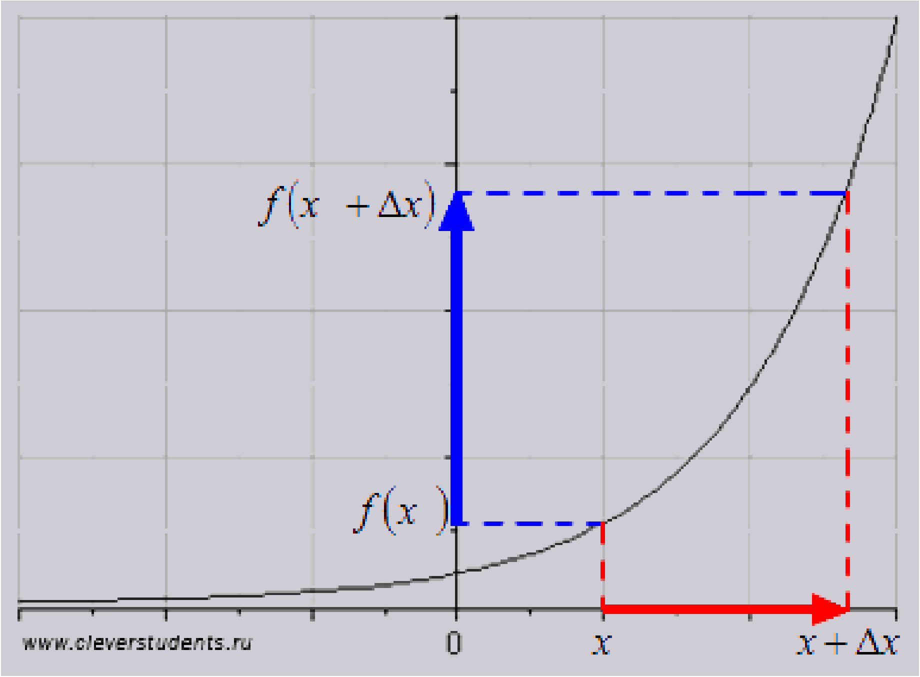
- Стационарные точки и точки, в которых не существует конечной производной, будем называть ***точками, подозрительными на экстремум***.
- Иллюстрация некоторых случаев:



1) Экстремума нет, первая производная равна нулю.
2) Точка максимума, первая производная слева и справа бесконечна.
3) Экстремума нет, первая производная слева и справа бесконечна.
4) Точка минимума, первая производная слева не равна первой производной справа.
5) Экстремума нет, первая производная слева не равна первой производной справа.

Понятие производной функции в точке

- Пусть x – аргумент функции $f(x)$ и Δx - малое число, отличное от нуля.
- Δx (читается «дельта икс») называют *приращением аргумента функции*.
- При переходе от значения аргумента x_0 к $x_0 + \Delta x$ значения функции изменяются соответственно от $f(x_0)$ до $f(x_0 + \Delta x)$.
- Разность $f(x_0 + \Delta x) - f(x_0) = \Delta f(x_0)$ называют приращением функции $f(x)$, соответствующем данному приращению аргумента.



- На рисунке красной линией показано изменение аргумента от значения x до значения формула (отсюда видна суть названия «приращение» аргумента).
- Приращение функции показано синей линией.

- **Производной** функции $f(x)$ в точке x_0 называется предел отношения приращения функции к приращению аргумента при $\Delta x \to 0$. Обозначается

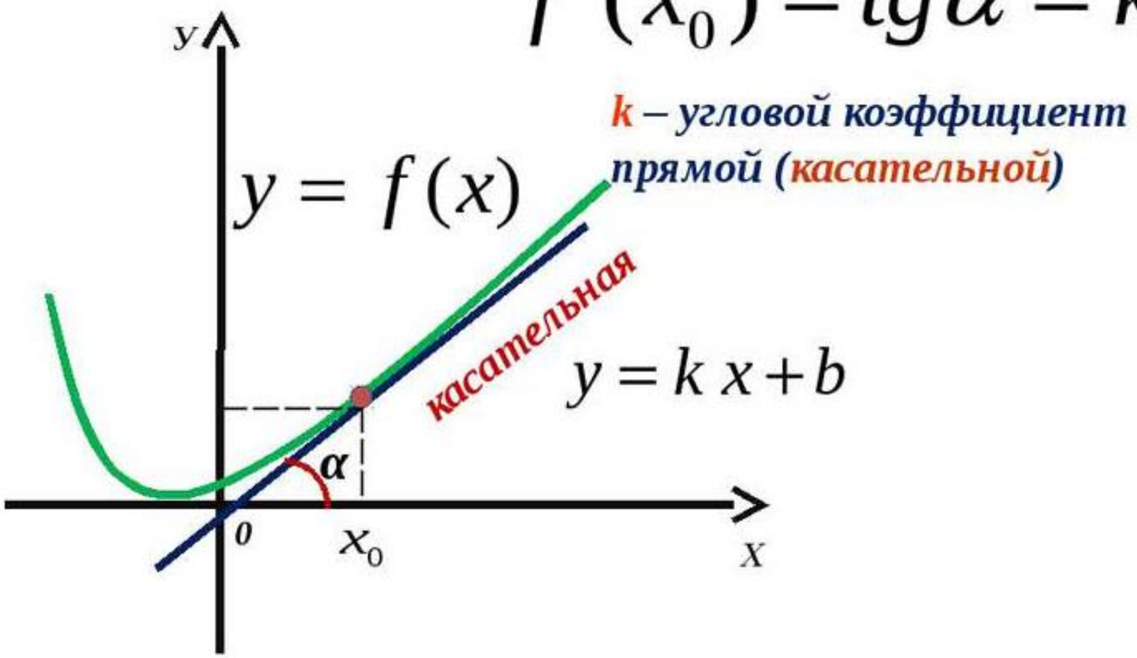
$$\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f(x_0)}{\Delta x} = f'(x_0)$$

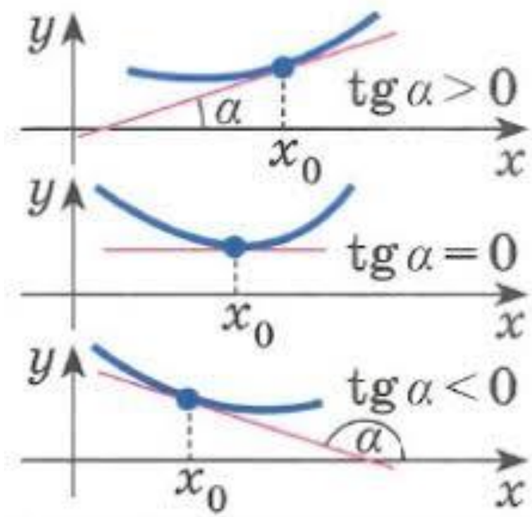
- Когда предел принимает конкретное конечное значение, то говорят о существовании конечной производной в точке.
- Если предел бесконечен, то говорят, что производная бесконечна в данной точке.
- Если же предел не существует, то и производная функции в этой точке не существует.
- Функцию $f(x)$ называют **дифференцируемой** в точке x_0, когда она имеет в ней конечную производную.
- Если функция $f(x)$ дифференцируема в каждой точке некоторого промежутка $(a;\ b)$, то функцию называют дифференцируемой на этом промежутке.
- Таким образом, любой точке x из промежутка $(a;\ b)$ можно поставить в соответствие значение производной функции в этой точке $f'(x)$, то есть, мы имеем возможность определить новую функцию $f'(x)$, которую называют производной функции $f(x)$ на интервале $(a;\ b)$.
- Операция нахождения производной называется **дифференцированием**.

Геометрический смысл производной

- Производная функции $f'(x_0)$ в точке x_0 численно равна угловому коэффициенту касательной к графику функции $f(x)$, проведенной в точке $(x_0; f(x_0))$.

$$f'(x_0) = tg\,\alpha = k$$





Физический смысл производной функции в данной точке

$$V_{ср.} = \frac{\Delta x}{\Delta t}$$

Или, если Δx – перемещение тела, а Δt – промежуток времени, в течении которого выполнялось движение, то

$\frac{\Delta x}{\Delta t}$ *– средняя скорость движения на промежутке времени t.*

При $\Delta t \to 0$ $V_{ср.} \to$ к мгновенной скорости $V(t)$, следовательно, $V(t) = S'(t)$.

$$S'(t) = V(t) \quad \text{или} \quad x'(t) = V(t)$$

Производная от функции в данной точке – это скорость изменения функции. $f'(x) = V(x)$

Таблица производных элементарных функций

1. $c' = 0,\ c = \text{const}$;
2. $\left(x^n\right)' = nx^{n-1}$;
3. $\left(a^x\right)' = a^x \cdot \ln a$;
4. $\left(e^x\right)' = e^x$;
5. $(\log_a x)' = \dfrac{1}{x \ln a}$;
6. $(\ln x)' = \dfrac{1}{x}$;
7. $(\sin x)' = \cos x$;
8. $(\cos x)' = -\sin x$;
9. $\left(\sqrt{x}\right)' = \dfrac{1}{2\sqrt{x}}$;
10. $(\operatorname{tg} x)' = \dfrac{1}{\cos^2 x}$;
11. $(\operatorname{ctg} x)' = -\dfrac{1}{\sin^2 x}$;
12. $(\arcsin x)' = \dfrac{1}{\sqrt{1-x^2}}$;
13. $(\arccos x)' = -\dfrac{1}{\sqrt{1-x^2}}$;
14. $(\operatorname{arctg} x)' = \dfrac{1}{1+x^2}$;
15. $(\operatorname{arcctg} x)' = -\dfrac{1}{1+x^2}$;
16. $(\operatorname{sh} x)' = \operatorname{ch} x$;
17. $(\operatorname{ch} x)' = \operatorname{sh} x$;
18. $(\operatorname{th} x)' = \dfrac{1}{\operatorname{ch}^2 x}$;
19. $(\operatorname{cth} x)' = -\dfrac{1}{\operatorname{sh}^2 x}$;

Правила дифференцирования

$$(u+v)' = u' + v'$$

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

$$(uv)' = u'v + uv'$$

$$\left(\frac{1}{u}\right)' = -\frac{u'}{u^2}$$

$$\left(c \cdot u(x)\right)' = c \cdot u'(x)$$

$$\left(uvw\right)' = u'vw + uv'w + uvw'$$

$$f'(g(x)) = f'(g(x)) \cdot g'(x)$$

Примеры:

- $y = 2\cos x$
- $y = \log_3 x^{\sqrt{2}-1}$
- $y = x^3 + 3^{x+1} - \ln x^{\ln(5+\sqrt{3})}$
- $y = \operatorname{tg} x \cdot \arcsin x$
- $y = (1 + x) \cdot \sin x \cdot \ln x$
- $y = \frac{\sin x}{2x+1}$

Теорема (условие монотонности функции):

- Пусть функция $f(x)$ определена и непрерывна в некотором промежутке и внутри него имеет конечную производную $f'(x)$. Для того, чтобы была на этом промежутке монотонно возрастающей (убывающей) в широком смысле, необходимо и достаточно условие

$$f'(x) \le 0 \; (f'(x) \ge 0)$$

Достаточное условие экстремума

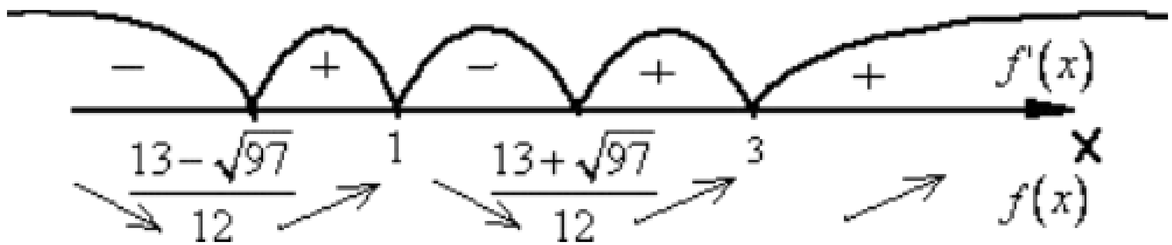
- Пусть функция $f(x)$ определена и дифференцируема в некоторой окрестности точки x_0, кроме, быть может, самой точки x_0 и непрерывна в этой точке. Тогда:
- Если производная $f'(x)$ меняет знак с «—» на «+» при переходе через точку x_0 : для всех x из $(x_0 - \Delta x;\, x_0)$ $f'(x) < 0$ и для всех x из $(x_0;\, x_0 + \Delta x)$ $f'(x) > 0$, то x_0 — точка строго минимума функции $f(x)$.
- Если производная $f'(x)$ меняет знак с «+» на «—» при переходе через точку x_0: для всех x из $(x_0 - \Delta x;\, x_0)$ $f'(x) > 0$ и для всех x из $(x_0;\, x_0 + \Delta x)$ $f'(x) < 0$, то x_0 — точка строго максимума функции $f(x)$.

Пример: Рассмотрим функцию $y = x(x-1)^2(x-3)^3$

$$y' = 6(x-1)(x-3)^2\left(x-\frac{13+\sqrt{97}}{12}\right)\left(x-\frac{13-\sqrt{97}}{12}\right)$$

- Точки, подозрительные на экстремум:

$$x_{1,2} = \frac{13\pm\sqrt{97}}{12},\ x_3 = 1,\ x_4 = 3$$



- В точке x = 1 производная меняет свой знак с плюса на минус, т.е. при функция имеет максимум.
- В точке $x = \frac{13-\sqrt{97}}{12}$ производная меняет свой знак с минуса на плюс, т.е. при функция имеет минимум.
- В точке x = 3 производная своего знака не меняет, т.е. экстремума там нет.

Алгоритм поиска экстремума $f(x)$

1) Найти производную функции $f(x)$.

2) Найти стационарные точки (точки, подозрительные на экстремум), решив уравнение $f'(x) = 0$.Обратить внимание на точки, в которых не существует двусторонней конечной производной.

3) Выяснить, меняет ли производная свой знак в точках, подозрительных на экстремум.

Если она меняет знак с минуса на плюс, то в этой точке функция имеет свой <u>минимум</u>.

Если с плюса на минус, то максимум, а если знак производной не меняется, то экстремума в этой точке нет.

4) Найти значение функции в точках минимума (максимума).

Пример: Исследовать на экстремум функцию

$$y=\sqrt[3]{\left(x^2-1\right)^2}$$

1) Найти производную функции

$$y'=\left(\sqrt[3]{\left(x^2-1\right)^2}\right)'=\left(\left(x^2-1\right)^{\frac{2}{3}}\right)'=\frac{2}{3}\left(x^2-1\right)^{-\frac{1}{3}}2x=\frac{4x}{3\sqrt[3]{x^2-1}}$$

2) Найти стационарные точки (точки, подозрительные на экстремум)

$$\frac{4x}{3\sqrt[3]{x^2-1}}=0$$

$$x\neq\pm1$$

$$x=0$$

Решим неравенство $y' > 0$.

$$\frac{4x}{3\sqrt[3]{x^2-1}} > 0$$

$$x\left(x^2-1\right) > 0$$

$$x(x-1)(x+1) > 0$$

$$-1 < x < 0;\ \ x > 1$$

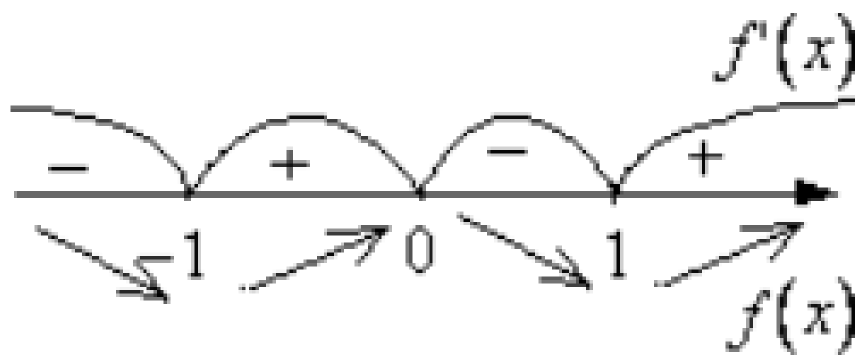
Решим неравенство $y' < 0$.

$$\frac{4x}{3\sqrt[3]{x^2-1}} < 0$$

$$x\left(x^2-1\right) < 0$$

$$x(x-1)(x+1) < 0$$

$$x < -1;\ \ 0 < x < 1$$



Вывод: при x = ±1 функция имеет минимумы,

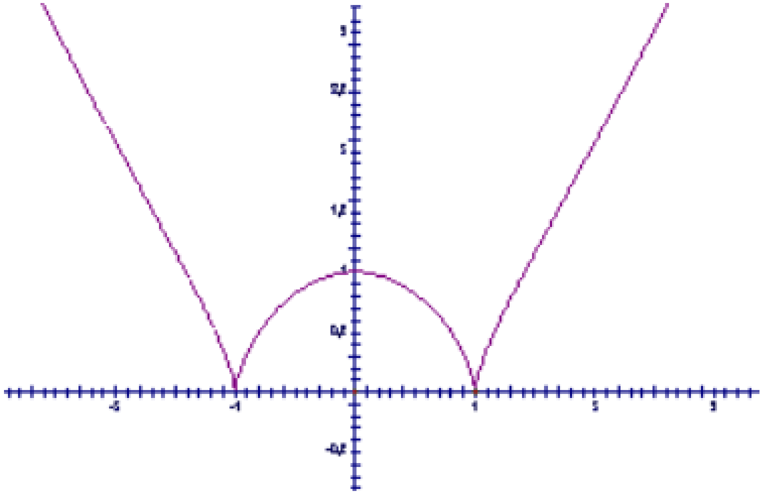
при x=0 — максимум.

4) Найти значение функции в точках минимума (максимума).

$$f(-1)=\sqrt[3]{\left((-1)^2-1\right)^2}=0$$

$$f(1)=\sqrt[3]{\left(1^2-1\right)^2}=0$$

$$f(0)=\sqrt[3]{\left(0^2-1\right)^2}=1$$



Точка максимума функции: (0; 1), точки минимума: (—1; 0), (1; 0)

Численные методы одномерной оптимизации

Решается задача поиска экстремума функции одной переменной с заданной точностью ε

- Непрерывная функция $y = f(x)$ называется **унимодальной** на отрезке $[a, b]$, если:

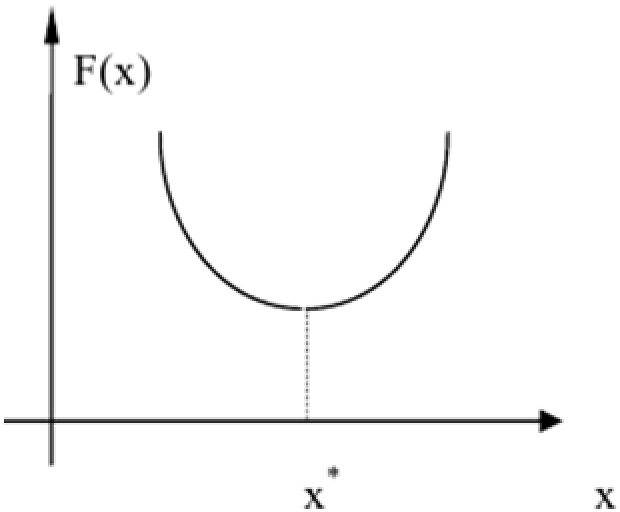
1. точка x_0 локального минимума функции принадлежит отрезку $[a, b]$;

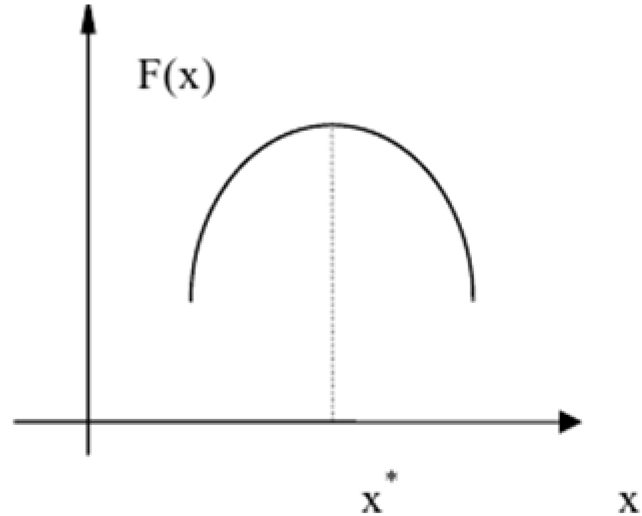
2. для любых двух точек отрезка x_1, x_2, взятых по одну сторону от точки минимума, точке, более близкой к точке минимума соответствует меньшее значение функции; то есть из условий $x_2 < x_1 < x_0$ или $x_0 < x_1 < x_2$ следует условие

$$f(x_1) < f(x_2).$$

- В противном случае функцию называют **полимодальной**.

Примеры унимодальных функций





Пример полимодальной функции

Схема сужения промежутка унимодальности функции

- Пусть требуется решить задачу

$$f(x) \to min,\ x \in X,\ X \subset \boldsymbol{R}$$

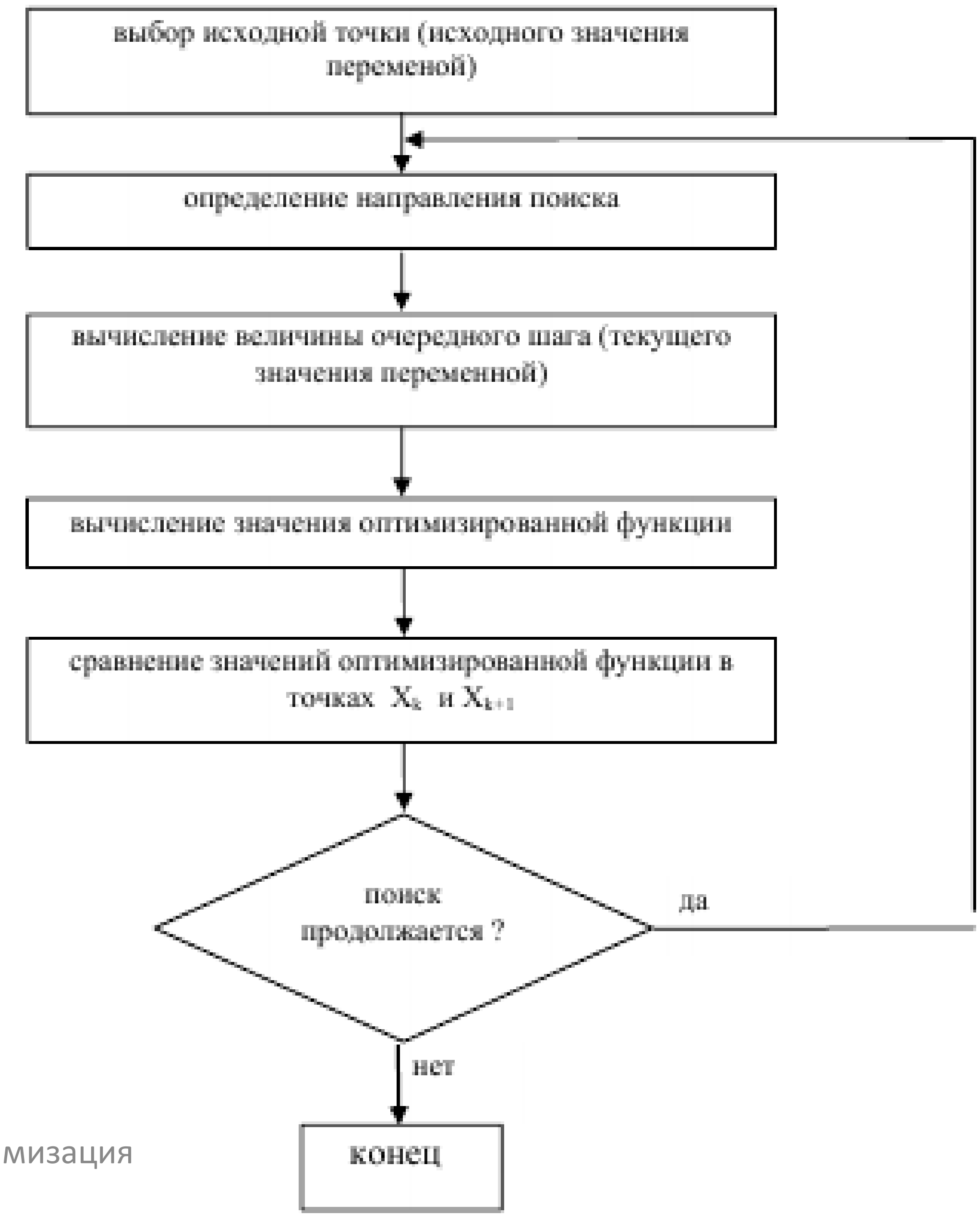
- Применение численных методов для отыскания точек x_0 локального минимума предполагает:

1. Определение промежутков унимодальности функции, то есть нахождение отрезков, каждому из которых принадлежит одна точка локального минимума;

2. Вычисление значения x_0, принадлежащего выбранному промежутку, с заданной точностью.

1. Для непрерывной функции $f(x)$ строят её график на некотором отрезке $[a, b]$ и, если окажется, что на этом отрезке функция выпукла вниз, то $[a, b]$ — отрезок унимодальности функции.
   Отрезок $[a, b]$ берётся, по возможности, малым.
2. При вычислении точки минимума точность достигается последовательным уменьшением отрезка $[a, b]$, содержащего точку x_0, до размеров, не превышающих заданную точность ε.

Алгоритм решения задачи одномерной оптимизации



Сходимость методов оптимизации

- На практике часто используются следующие условия остановки метода:

$$\left\| x^{k+1} - x^k \right\| \le \varepsilon_1,$$

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

$$\left\| f'(x^{k+1}) \right\| \le \varepsilon_3.$$

Методы безусловной оптимизации нулевого порядка

- В этих методах для определения направления спуска не требуется вычислять производные целевой функции.
- Направление минимизации в данном случае полностью определяется последовательными вычислениями значений функции.
- *Метод дихотомии (половинного деления)*
- *Метод золотого сечения*

Метод дихотомии (половинного деления)

1) На каждом шаге процесса поиска делим отрезок $[a, b]$ пополам, $x = (a + b)/2$ – координата середины отрезка $[a, b]$.

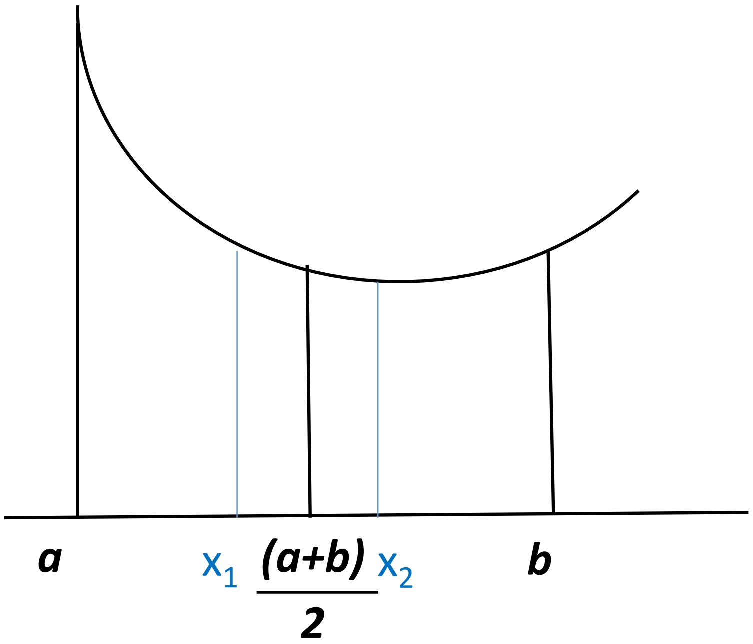
2) Вычисляем значение функции $f(x)$ в окрестности вычисленной точки x, т.е.

$$f_1 = f(x - \varepsilon),$$

$$f_2 = f(x + \varepsilon).$$

3) Сравниваем f_1 и f_2 и отбрасываем одну из половин отрезка $[a, b]$

Метод дихотомии (половинного деления)



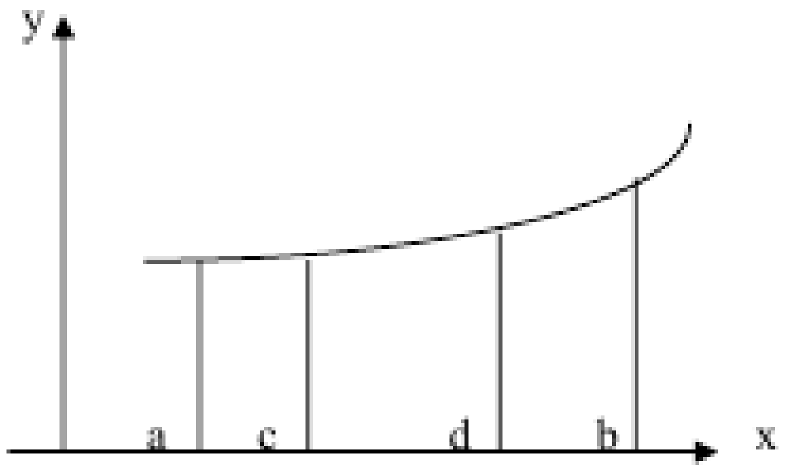
При поиске минимума:

- Если $f_1 < f_2$, то отбрасываем отрезок $[x, b]$, тогда $b = x$.
- Иначе отбрасываем отрезок $[a, x]$, тогда $a = x$

Метод золотого сечения

- В соответствии с методом золотого сечения внутри отрезка $[a, b]$ выделяют две промежуточные точки c и d на расстоянии $s = aL$ от его конечных точек, где $L = b - a$ — длина отрезка, $a = \frac{\sqrt{5}+1}{2}$ – коэффициент золотого сечения.
- Затем вычисляют значения целевой функции в точках c и d и смещают область поиска в соответствующую часть отрезка.

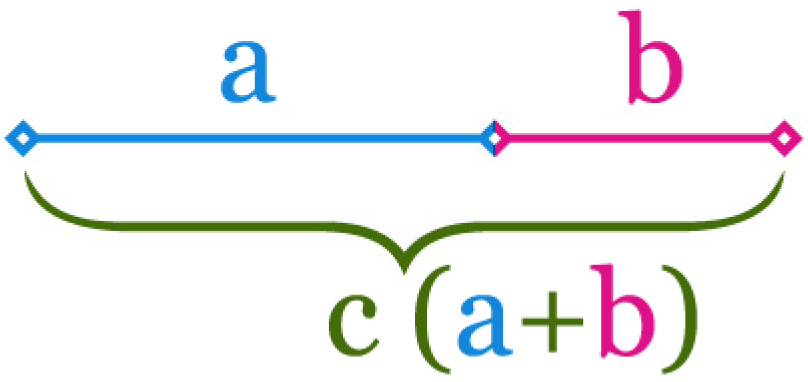
Метод золотого сечения



если $F(d) > F(c)$ $b \to d$,

если $F(d) < F(c)$ $a \to c$.

- Золотое сечение



Методы безусловной оптимизации первого порядка

- К методам первого порядка относятся алгоритмы, в которых в процессе поиска кроме информации о самой функции используется информация о её <u>производной</u> первого порядка.
- *Метод средней точки*
- *Метод хорд*
- *Метод касательных*

Метод средней точки

- В методе средней точки используется необходимое условие экстремума: производная функции в точке экстремума обращается в нуль.
- Значит, чем ближе к нулю значение $f'(x^*)$, тем точнее найдена точка экстремума.

Алгоритм метода средней точки

- ***Начальный этап***. Выбрать начальный интервал $[a_1, b_1]$ и точность поиска ε. Задать $k = 1$ и перейти к основному этапу.
- ***Основной этап***.
- Шаг 1. Вычислить среднюю точку $x_k = (a_k + b_k)/2$ и значение производной $f\,'(x_k)$.
- Шаг 2. Если $|f\,'(x_k)| < \varepsilon$, то расчёт закончен и экстремум находится в точке x_k. Иначе перейти к шагу 3.
- Шаг 3. Если $f\,'(x_k) < 0$, то $a_{k+1} = x_k$ и $b_{k+1} = b_k$, иначе $a_{k+1} = a_k$ и $b_{k+1} = x_k$, увеличить $k = k+1$ и перейти к шагу 1.

Метод хорд

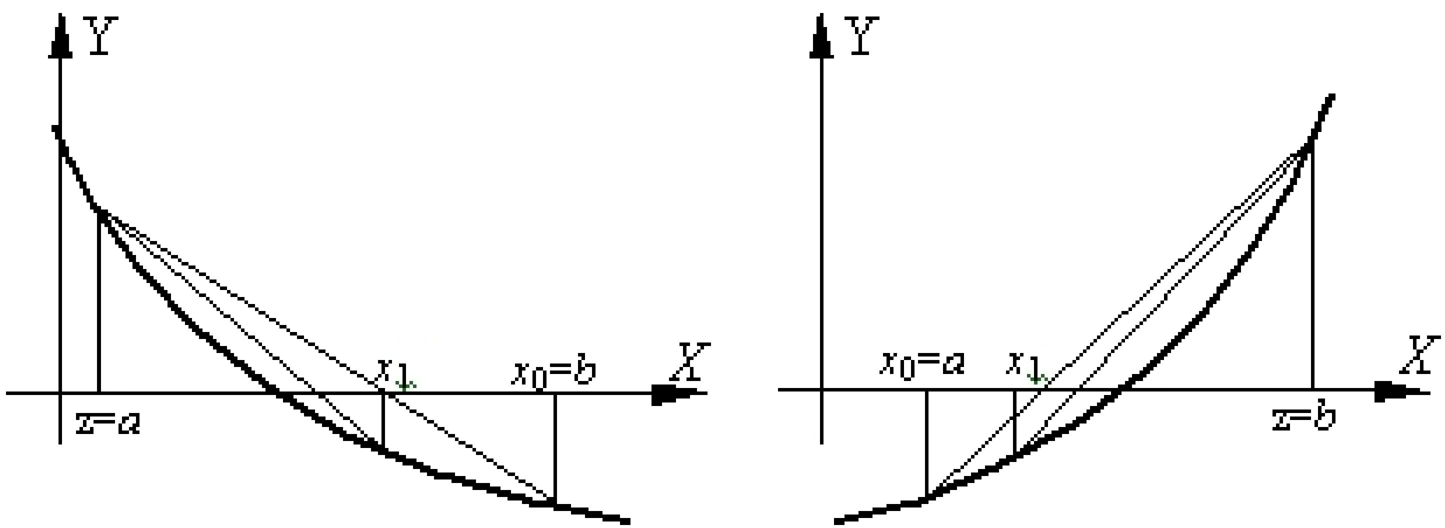
- **Метод хорд** предлагает делить отрезок в точке, отстоящей от краёв отрезка пропорционально абсолютному значению производной функции на краях.
- Метод основан на замене функции $f'(x)$ на каждом шаге поиска хордой, пересечение которой с осью X дает приближение корня уравнения $f'(x) = 0$.

Алгоритм метода хорд

- При этом в процессе поиска семейство хорд может строиться:

а) при фиксированном левом конце хорд, т.е. $z = a$, тогда начальная точка $x_0 = b$;

б) при фиксированном правом конце хорд, т.е. $z = b$, тогда начальная точка $x_0 = a$.



- Процесс поиска продолжается до тех пор, пока не выполнится условие $|x_{n+1} - x_n| < \varepsilon$ или $| f'(x_n) | < \varepsilon$.
- Итерационный процесс схождения к корню производной реализуется рекуррентной формулой:

для случая а):

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

для случая б):

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

Метод Ньютона (касательных)

- Следующий из рассматриваемых методов оптимизации является градиентным методом **второго порядка**.
- В нём при поиске экстремума целевой функции используется её <u>первые и вторые производные</u>.
- Метод применим функции, у которой монотонна её первая производная $f'(x)$.

- Если x_n – точка, полученная на n-ном шаге, то функция $f'(x)$ аппроксимируется своим уравнением касательной:

$$y = f'(x_n) + (x - x_n)f''(x_n)$$

- а точка x_{n+1} выбирается как пересечение этой прямой с осью Ox, т.е.

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

- Неудобство этого метода состоит в необходимости вычисления в каждой точке первой и второй производных.
- Значит, он применим лишь тогда, когда функция $f(x)$ имеет достаточно простую аналитическую форму, чтобы производные могли быть вычислены в явном виде вручную.

Пример

- Исследуем на экстремум функцию

$$f(x) = \frac{x}{1+x^2}$$