Оптимальное управление

Старший преподаватель
каф. “Кибернетика”
Локтюшев Александр Викторович

1. Постановка задачи оптимизации

Понятие “оптимизация”

Людям свойственно стремление к лучшему, и если им приходится выбирать из нескольких возможностей, то желание найти среди них лучшую представляется вполне естественным.

- Для постановки задачи принятия решения необходимо выполнение двух условий:
- 1) должно быть много вариантов решений;
- 2) лучший вариант должен быть выбран по определенному принципу.

Понятие “оптимизация”

Критериальный выбор заключается в принятии некоторого критерия и сравнении возможных вариантов, соответствующих критерию.

Если такой выбор предусматривает проведение количественного анализа ситуации путем сравнения различных вариантов, то говорят о решении задачи оптимизации (по латыни optimus — наилучший).

Понятие “оптимизация”

- Под ***оптимизацией*** понимают процесс выбора наилучшего варианта из всех возможных

Для того чтобы найти оптимальную из возможностей, приходится решать задачи на отыскание максимума или минимума

Оба эти понятия объединяются термином «***экстремум***» (от латинского extremum, означающего «крайнее»). Поэтому задачи на отыскание максимума или минимума называют **экстремальными задачами**.

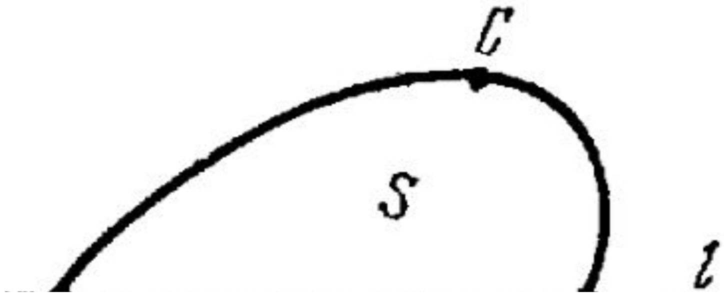
История развития оптимизации

825 г. до н.э. – задача царицы Дидоны

- Финикийская царевна Дидона и с ней небольшой отряд жителей города Тира, спасаясь от преследований тирана, покинули родной город и отправились на кораблях на запад вдоль берегов Средиземного моря. Выбрав на африканском побережье удобное место (нынешний Тунисский залив), Дидона и ее спутники решили основать здесь поселение.
- Дидоне, удалось "уговорить предводителя местных жителей Ярба, и он неосторожно согласился уступить Дидоне клочок земли, «который можно окружить бычьей шкурой».
- Как поступила царица Дидона, чтобы территория охваченной земли оказалась наибольшей?

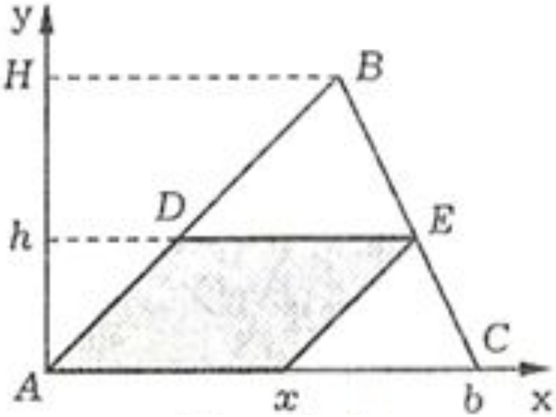
Параметрические задачи

Разрезав шкуру на тонкие полоски, Дидона связала их в один длинный ремень и, окружив им значительную территорию, заложила на ней город **Карфаген** (финикийское Картадашт— новый город) или **Бирса** (пунийское (так римляне называли жителей Карфагена) — **шкура**).



- требуется указать оптимальную форму участка земли, который при заданной длине периметра L, имеет наибольшую площадь S.

Задача Евклида (IV в. до н.э.)



В заданный треугольник *ABC* с высотой *H* и основанием *b* вписать параллелограмм наибольшей площади S, стороны которого параллельны двум сторонам треугольника

Критерием оптимальности - достижение площадью параллелограмма наибольшего значения, а *ограничения* связаны с условиями параллельности сторон и размещения параллелограмма в пределах заданного треугольника.

$$\begin{cases} S = hx \to \max; \\ \dfrac{H-h}{H} = \dfrac{x}{b}, \qquad h \geq 0, \qquad x \geq 0. \end{cases}$$

$x^* = b/2$

$h^* = (1 - x/b)H = H/2$

Этапы постановки задачи оптимального управления

1. Вербальное (словесное) описание задачи. Определение основных целей, достигаемых при решении задачи управления.

При постановке задачи оптимизации необходимо:

- Наличие **объекта оптимизации -** устройства, процессы и ситуации, применительно к которым предстоит решать задачу оптимизации, объединим общим названием **объект оптимизации**. В качестве объекта оптимизации может фигурировать **объект (технологический процесс)** или его **математическая модель**.
- Наличие ресурсов оптимизации - возможность изменения значений некоторых параметров объекта оптимизации (**управляющими воздействиями -** варьируемые, поисковые переменные).

Этапы постановки задачи оптимального управления

2) **Построение математической модели**. Вводятся обозначения для всех переменных (желательно с указанием их размерности).

3) **Формулировка критерия оптимальности** (целевой функции), соответствующего цели поставленной задачи управления.

   2) Частные случаи – функция вещественных переменных или функционал

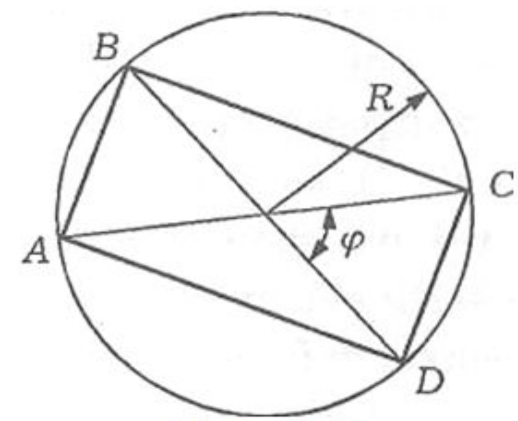
4) Выделение **ограничений -** множества допустимых значений переменных, а также условия, наложенные на совокупность переменных. Множество допустимых решений будем обозначать D.

Постановка задачи оптимизации

1) **Вербальное (словесное) описание задачи**

Найти стороны прямоугольника, вписанного в окружность радиуса *R* и имеющего наибольшую площадь *S*

2) Построение математической модели



Обозначим длины сторон прямоугольника a и b – варьируемые параметры

3) Формулировка критерия оптимальности

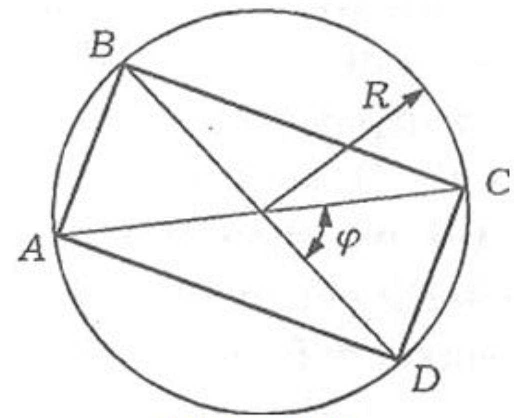
Площадь $S = a \cdot b \to \max$

4) Выделение **ограничений**

$a>0,\ b>0,\ a^2+b^2 = (2R)^2$

Решение задачи оптимизации

Найти стороны прямоугольника, вписанного в окружность радиуса R и имеющего наибольшую площадь S



$$\begin{cases} S = ab \to \max; \\ a^2 + b^2 = 4R^2, \quad a \geq 0, \quad b \geq 0. \end{cases}$$

$(a - b)^2 = a^2 + b^2 - 2ab = 4R^2 - 2S$

$2S = 4R^2 - (a - b)^2$

$S \to \max$ при $(a - b)^2 \to \min$, т.е. a=b

2) Площадь S прямоугольника равна половине произведения его диагоналей на синус угла φ между диагоналями, т.е. $S = ½*(2R)^2\sin\varphi$.

$S \to \max$ при $\sin \phi = 1$, т.е. ABCD - квадрат

Переменные, характеризующие объект управления:

Входные переменные:

a) возмущающие (внешние) воздействия (x_k, k=1,2,…,m);

b) управляющие воздействия

(u_j, j=1,2,…,r);

Выходные переменные (y_i, i=1,2,…,n).

Состояния объекта управления

- Статическое состояние
- Динамическое состояние

Статическое состояние

Признаком статического состояния объекта управления является постоянство во времени переменных, характеризующих состояние объекта управления, т.е. $dx_i/dt = 0$,

где x_i – переменные, характеризующие состояние объекта управления.

Физически статическое состояние есть состояние, при котором имеет место

Приход (*энергии, вещества*) =

= Расход (*энергии, вещества*)

Связь переменных при статическом состоянии объекта

Связь переменных при статическом состоянии объекта управления:

$$y_i = \varphi_i(x_1,\ldots,x_m,u_1,\ldots,u_r),\;\; i = 1,\ldots,n,$$

или $\boldsymbol{y} = \varphi(\boldsymbol{x},\boldsymbol{u})$,

где $\boldsymbol{y} = (y_1,\ldots,y_n),\; \boldsymbol{x} = (x_1,\ldots,x_m),\; \boldsymbol{u} = (u_1,\ldots,u_r)$.

Целевая функция есть математический оператор, который набору чисел на входе ставит в соответствие число на выходе.

Виды критериев оптимальности

Пример. Во время второй мировой войны несколько сот английских торговых судов на Средиземном море были вооружены зенитными орудиями для защиты от вражеских бомбардировщиков. Поскольку это мероприятие было достаточно дорогим (требовалось иметь на каждом судне боевую команду), через несколько месяцев решили оценить его эффективность

Какой из параметров оптимизации более подходит для этой цели?

- Число сбитых самолетов.
- Потери в судах, оснащенных орудиями, по сравнению с судами без орудий.

Требования к критерию оптимальности

- Параметр оптимизации – это признак, по которому мы хотим оптимизировать процесс. Он должен быть *количественным*, *задаваться числом*. Мы должны уметь его измерять при любой возможной комбинации выбранных уровней факторов.

- Множество значений, которые может принимать параметр оптимизации, будем называть *областью его определения*.

Классификация задач оптимизации

1) По количеству варьируемых переменных

В зависимости от числа управляемых параметров различают методы *одномерной* (*варьируемый параметр единственный*) и *многомерной* (*размер вектора X не менее двух) оптимизации*.

Реальные задачи в АСУ многомерны.

Методы одномерной оптимизации играют вспомогательную роль на отдельных этапах многомерного поиска.

Классификация задач оптимизации

2) По области определения

- **Неограниченными: x∈ R** –
  Безусловная задача оптимизации
- **Ограниченными: x∈ D(x) –**

Условные задачи оптимизации, или *задачи с ограничениями*, — это такие, при формулировке которых задаются некоторые условия (ограничения) на множестве;

Теория и методы решения задач оптимизации при наличии ограничений составляют предмет исследования одного из разделов прикладной математики — ***математического программирования***.

- *Ограничения* представляют собой различные технические, экономические, экологические условия, учитываемые при решении задачи. Ограничения представляют собой зависимости между переменными x^1, x^2, ... x^n, задаваемые в форме неравенств или равенств

$$f_1(x_1, x_2, \dots x_n) \le b_1;$$
$$f_2(x_1, x_2, \dots x_n) = b_2;$$
$$\dots\dots\dots\dots\dots\dots\dots$$
$$f_m(x_1, x_2, \dots x_n) > b_m.$$

- Общее количество ограничений равно m. Правые части ограничений, представляющие собой постоянные коэффициенты b_j (j=1, 2, … m), называются свободными членами.

- Как и в выражении целевой функции, зависимости между переменными в системе ограничений могут быть линейными и нелинейными.

- ***Граничные условия*** устанавливают диапазон изменения искомых переменных $d_i \le x_i \le D_i$, i=1, 2, … n, где d_i и D_i - соответственно нижняя и верхняя границы диапазона изменения переменной x_i.
- Наиболее часто в технических задачах все искомые переменные, как правило, неотрицательны. В этом случае граничные условия имеют следующий вид: $x_i \ge 0$, где i = 1, 2, ... n.

При наличии ограничений и граничных условий ищется уже не абсолютный, а **относительный экстремум целевой функции**. На рис. показана некоторая функция одного переменного $Z(x)$. Указан диапазон изменения переменной x (нижняя граница d и верхняя граница D). Видно, что абсолютный минимум функции соответствует точке 1, а относительный минимум – точке 2, принадлежащей заданному диапазону изменения переменной x.

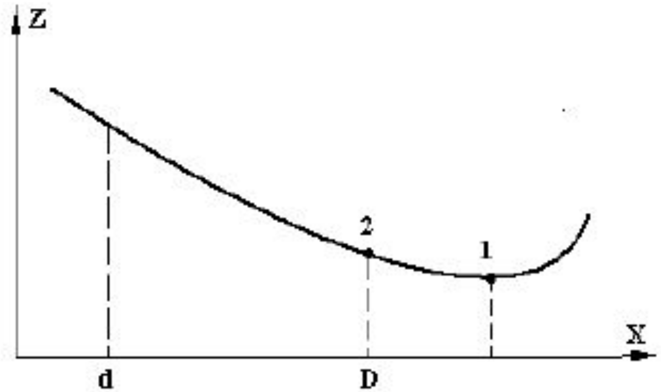


Рис. 1.2. Абсолютный (точка 1) и относительный (точка 2) минимумы функции

3) По виду зависимости

Зависимость между переменными в целевой функции может быть ***линейной*** или ***нелинейной***. Линейной называется такая зависимость, в которую переменные x_i входят только в первой степени и с этими переменными выполняются только действия сложения, вычитания и умножения на постоянный коэффициент.

Такие задачи оптимизации называются **задачами линейного программирования**.

Во всех других случаях зависимость будет нелинейной, а задачи оптимизации – **задачами нелинейного программирования**.

Пример задачи линейного программирования

- Обозначим через t_j - время, в течении которого изделия выпускаются по j-ой технологии. Выпуск изделий за это время составит $c_j t_j$ штук, а израсходовано будет единиц $a_{ij} t_j$ i- ого ресурса .
- Общее потребление каждого j-ого ресурса всеми технологическими процессами не должно превышать величины b_i .
- Требуется найти $\max F = \sum_{j=1}^{n} c_j t_j$

  при ограничениях

.

$$\sum_{j=1}^{n} a_{ij} t_j \le b_j (i = \overline{1, m})$$

Классификация задач оптимизации

4) По количеству экстремумов

- Нелинейная целевая функция в заданном диапазоне изменения переменных может иметь один экстремум (***одноэкстремальная***) или несколько экстремумов (*многоэкстремальная*).

Функции с одним экстремумом называются ***унимодальными***.

В многоэкстремальных функциях вводят понятие ***локального*** и ***глобального*** оптимумов.

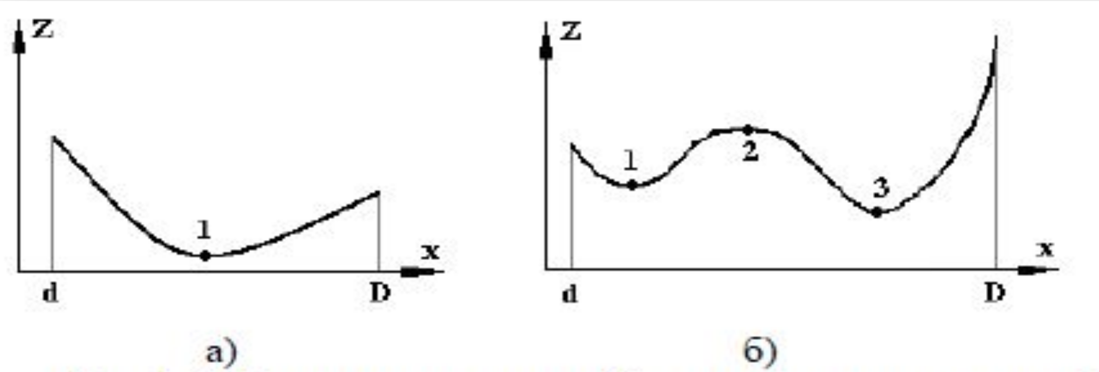


Рис. 1.1. Одноэкстремальная (а) и многоэкстремальная (б) функции

5) По характеру изменения переменных

- Если переменные могут принимать любые значения, такие переменные называются *непрерывными*.
- Если переменные могут принимать только значения целых чисел, такие переменные называются *целочисленной*, а задача оптимизации – *задачей целочисленного программирования*.
- Если переменные могут принимать только определенные значения, такие переменные называются *дискретными* , а задача оптимизации – *задачей дискретного программирования*.

6) По количеству критериев оптимальности

- Задачи, в которых оптимизация проводится не по одному, а по нескольким критериям, относятся к классу задач ***многокритериальной оптимизации***.
- Решение таких задач заключается в нахождении компромисса между принятыми критериями оптимальности.

- Обычно проводится ***операция свертки*** критериев с целью получения значения компромиссного решения.
- Если все показатели измеряются по одной шкале, то возможно использовать аддитивный

$$F = \sum_{i-1}^{n} \beta_i f_i(\overline{x})$$

или мультипликативный критерий

$$F = \prod_{i=1}^{p} \left[f_i(\overline{x}) \right]^{\beta_i}$$

Если частные показатели качества измеряются по разным шкалам, то формируется обобщенный критерий, приведенный к единой шкале измерений

$$\min F(\overline{x}) = \sum_{i=1}^{p} \beta_i \frac{c_i - f_i(\overline{x})}{|c_i|}$$

где

$$c_i = \max f_i(\overline{x}), \overline{x} \in X$$

Виды объектов управления:
объект с сосредоточенными параметрами;
объект с распределенными параметрами.

Объект с сосредоточенными параметрами – есть объект, в каждой точке которого в рассматриваемый момент времени характеризующие его состояние переменные принимают одни и те же значения.

Объект с распределенными параметрами – есть объект, в направлении координатных осей которого в рассматриваемый момент времени характеризующие его состояние переменные имеют различные значения (распределены в направлении координатных осей).

Выбор критериев оптимальности для задач оптимизации объектов, находящихся в статическом и динамическом состояниях

Критерии оптимальности для объектов, находящихся в статическом состоянии:

a) для объектов с сосредоточенными параметрами – целевая функция в виде $Q = Q(\boldsymbol{x}, \boldsymbol{u})$;
b) для объектов с распределенными параметрами – функционал в виде $Q(l) = \int_{l_0}^{l} \partial Q(l) / \partial l \, dl$, где $\partial Q(l) / \partial l = F(\boldsymbol{x}(l), \boldsymbol{u}(l))$,

l – пространственная координата.

Динамическое состояние

Признаком динамического состояния объекта управления является изменение во времени переменных, характеризующих состояние объекта управления, т.е. $dx_i/dt \neq 0$.

Физически динамическое состояние есть состояние, при котором имеет место

Приход (*энергии, вещества*) - Расход (*энергии, вещества*) =

= Накопление или истечение (*энергии, вещества*).

Связь переменных при динамическом состоянии объекта

Связь переменных при динамическом состоянии объекта управления:

$$dy_i(t)/dt = f_i(y_1(t),\dots,y_n(t),u_1(t),\dots,u_r(t)),$$

$$i = 1,\dots,n,$$

или $$d\boldsymbol{y}(t)/dt = f(\boldsymbol{y}(t),\boldsymbol{u}(t))$$

где $$d\,\mathbf{y}(t)/dt = (d\,y_1(t)/dt,\dots,d\,y_n(t)/dt), \quad \mathbf{f} = (f_1,\dots,f_n),$$

$$\mathbf{y}(t) = (y_1(t),\dots,y_n(t)), \quad \mathbf{u} = (u_1(t),\dots,u_r(t))$$

Здесь $y_i(t)$ - переменная, характеризующая состояние объекта управления.

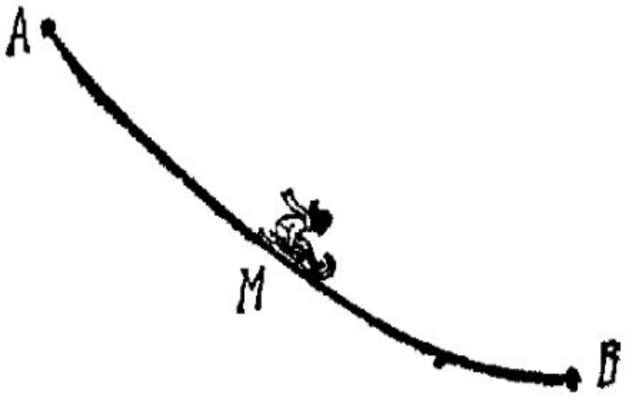
Вариационное исчисление

- Задача о брахистохроне.

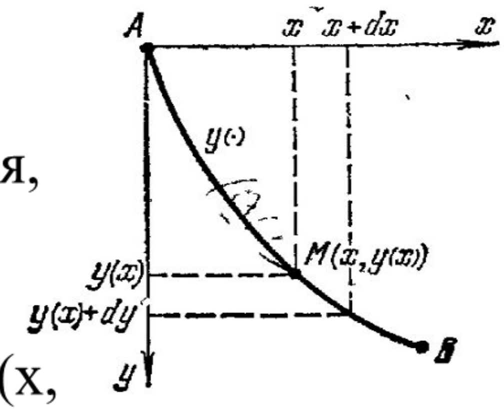
В 1696 г. появилась заметка И. Бернулли с интригующим заглавием: «Новая задача, к решению которой приглашаются математики».

В ней была поставлена следующая задача: В вертикальной плоскости даны две точки A и B.

Определить путь AMB, спускаясь по которому под действием собственной тяжести тело M, начав двигаться из точки A, дойдет до точки B в кратчайшее время.



Введем в плоскости систему координат (x, y) так, чтобы ось x была горизонтальна, а ось y направлена вниз. В соответствии с законом Галилея, скорость тела M в точке с координатами (x, y (x)) если тело без трения спускается по кривой F(*) не зависит от формы кривой F(*) между точками A и (x, y(x)), а зависит лишь от ординаты y(x) и равна $\sqrt{gy(x)}$, где g—ускорение силы тяжести.



При этом требуется найти наименьшее время, которое потребуется на. преодоление пути от A к B, т. е. надо минимизировать интеграл

$$T = \int_{\breve{AB}} \frac{ds}{v} = \int_{\breve{AB}} \frac{ds}{\sqrt{2gy(x)}}$$

В задаче же о брахистохроне множество всех кривых, соединяющих две точки, бесконечномерно. Здесь требуется найти экстремум функции бесконечного числа переменных.

Критерии оптимальности для задач оптимизации объектов, находящихся в динамическом состоянии

Критерии оптимальности для объектов, находящихся в динамическом состоянии есть функционал в виде:

$$Q(t) = \int_{t_0}^{t} dQ(t)/dt\, dt$$, где $$dQ(t)/dt = F(\boldsymbol{x}(t), \boldsymbol{u}(t))$$

t – время.

Функционал есть математический оператор, который функции на входе ставит в соответствие число на выходе.

2. Одномерная оптимизация

2.1 Необходимые и достаточные условия экстремума функции одной переменной

Теорема Вейерштрасса.

Всякая функция f(x), непрерывная на отрезке [a,b], принимает на этом отрезке наименьшее и наибольшее значения, т. е. на отрезке [a,b] существуют такие точки x1 и x2, что для любого x $\in$ [a,b] имеют место неравенства

$$f(x_1) \leq f(x) \leq f(x_2)$$

.

ПОСТАНОВКА ЗАДАЧИ

Одномерная оптимизация заключается в нахождении точки x^*, в которой целевая функция $f(x^*)$ принимает максимальное или минимальное значение. Часто в постановках задачи может быть задан отрезок $[a, b]$, в котором находится оптимальное значение.

Функция $f(x)$ имеет локальный минимум в точке x^*, если при $\varepsilon > 0$ существует окрестность $[x^* - \varepsilon, x^* + \varepsilon]$ такая, что для всех значений x в этой окрестности $f(x)$ больше $f(x^*)$. Функция $f(x)$ имеет глобальный минимум в точке x^*, если для всех x справедливо неравенство $f(x) > f(x^*)$.

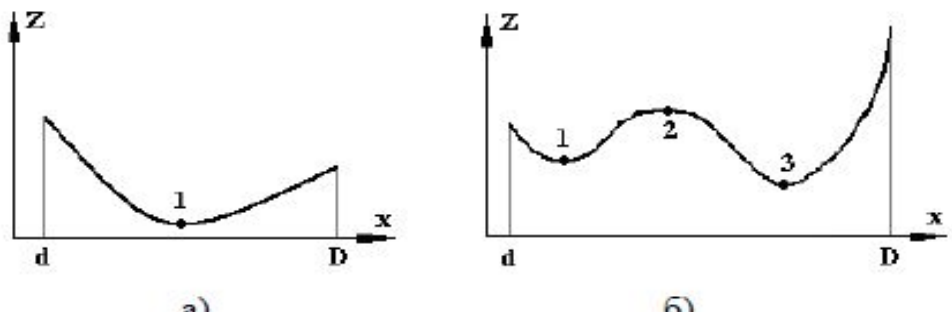


Рис. 1.1. Одноэкстремальная (а) и многоэкстремальная (б) функции

АНАЛИТИЧЕСКИЙ АНАЛИЗ ФУНКЦИИ

Необходимые и достаточные условия экстремума

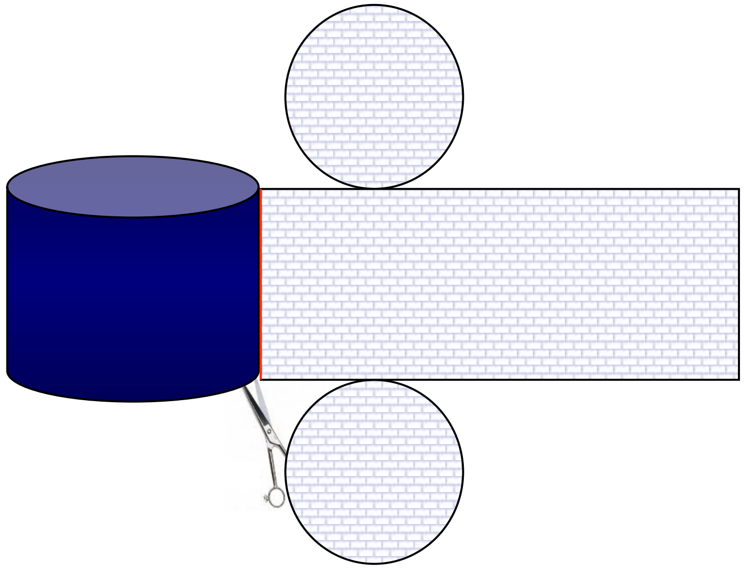
Классический подход к задаче нахождения экстремумов функции состоит в поиске условий, которым они должны удовлетворять. ***Необходимым условием*** экстремума в точке x^* является равенство нулю первой производной (теорема Ферма), т.е. требуется решить уравнение

$$f'(x) = 0. \tag{1}$$

Данному уравнению удовлетворяют как локальные и глобальные экстремумы, так и точки перегиба функции, поэтому приведенное условие является только необходимым, но недостаточным.

Задачи оптим. проектирования

Пусть требуется спроектировать бак горючего в виде прямого цилиндра заданного объема *V*, на изготовление которого будет затрачено наименьшее количество листовой стали.



$$V = \pi r^2 h \qquad S = 2\pi r^2 + 2\pi r h$$

$$h = \frac{V}{\pi r^2} \qquad S(r) = 2\pi r^2 + \frac{2V}{r} \qquad S'(r) = 0, \quad r_1 = \sqrt[3]{\frac{V}{2\pi}}$$

$$S'(r) = 4\Pi r - \frac{2V}{r^2} = \frac{2}{r^2}(2\Pi r^3 - V)$$

$$\sqrt[3]{\frac{V}{2\pi}}$$

S’ - +

S min

Необходимые и достаточные условия минимума функции одной переменной

Необходимое условие экстремума $f(x)$:

$$f'(x)\big|_{x=x^*} = 0$$

Достаточное условие минимума (гессиан размера 1x1):

$$f''(x)\big|_{x=x^*} > 0$$

Однако, может быть что $f''(x)\big|_{x=x^*} = 0$.
Если все $f^{(i)} = 0,\ i = 1..n-1$, а $f^{(n)} \neq 0$, то:

- Если n - нечетное, то минимума нет (а есть перегиб)
- Если n - четное, то при $f^{(n)}\big|_{x=x^*} < 0$ - минимум, при $f^{(n)}\big|_{x=x^*} > 0$ - максимум

Пример 1. Аналитический поиск безусловного экстремума. Функция одной переменной

$$f(x) = x^3 + 10x^2 + 2x + 38$$

$$f'(x) = 3x^2 + 20x + 2$$

Достаточное условие

$$f'(x) = 0$$

$$3x^2 + 20x + 2 = 0$$

$$x_{1,2} = \frac{-20 \pm \sqrt{400 - 4 \times 2 \times 3}}{6} = (-6.57, -0.1)$$

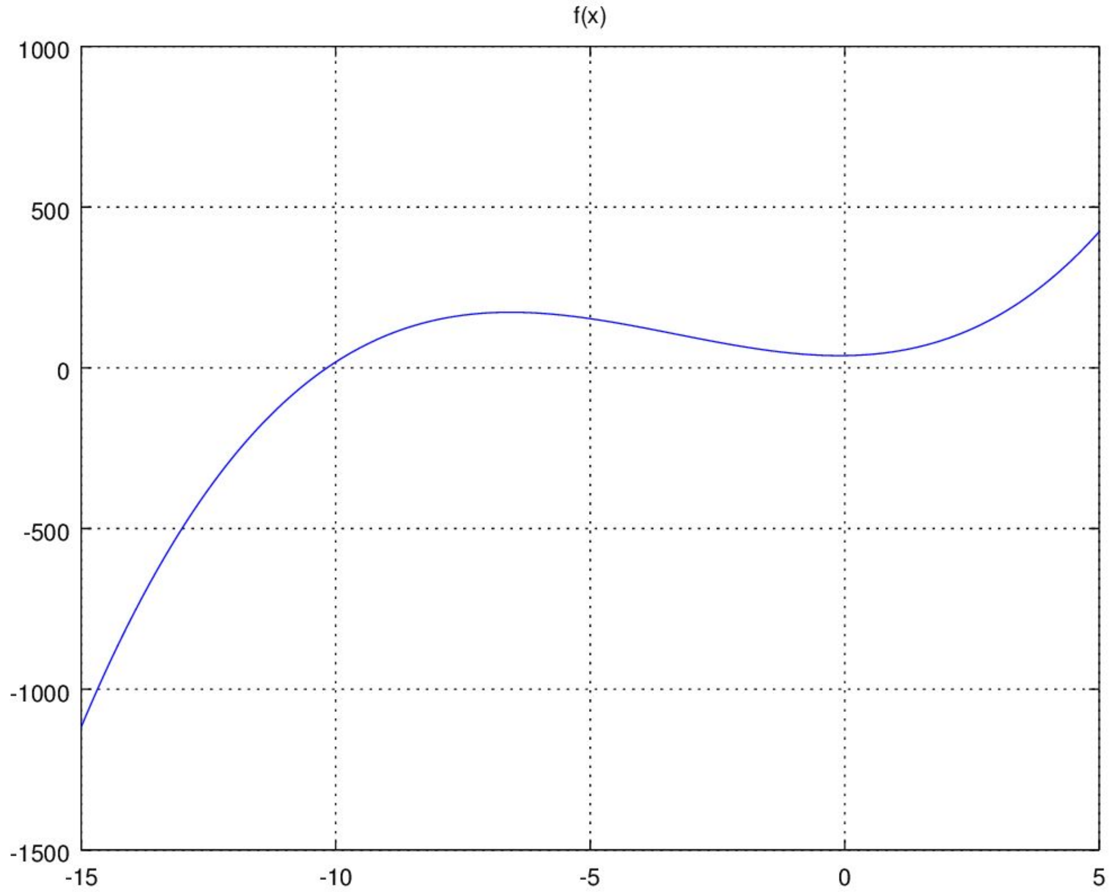
Две точки подозрительны на экстремум.
Определим значения вторых производных

$$f''(x) = 6x + 20$$

в точках, подозрительных на экстремум принимает значение -19.39, 19.39 соответственно.

Значит ($n = 2$, четно - экстремум есть) - точка $x_1 = -6.57$ - точка максимума, точка $x_2 = -0.1$ - точка минимума.

К примеру 1



Пример 2. Аналитический поиск условного экстремума. Функция одной переменной

$$f(x) = x^3 + 10x^2 + 2x + 38$$

$$x \in [-5, 5]$$

Точки, подозрительные на экстремум:

- $x_2 = -0.1$ (локальный минимум)
- левый край интервала $x_a = -5$
- правый край интервала $x_b = 5$

Вычисляем значение функции.

$$f(x_a) = 153$$

$$f(x_b) = 423$$

$$f(x_2) = 37.9$$

Минимум: x_2 Максимум: x_b

Примеры исследования функции одной переменной на экстремум

Пример 1. Q = $(1-u)^3$.

Необходимое условие экстремума $dQ(u)/du|_{u*} = 0$

$3(1-u^*)^2(-1) = 0$,

$(1-u^*)^2 = 0$, $u^* = 1$.

Достаточное условие экстремума

$d^2Q(u)/du^2|_{u*}$ = $-6(1-u^*) = -6(1-1) = 0$, .

$k = 3$ – нечетное. $d^3Q(u)/du^3|_{u*} = -6 \neq 0$

Ответ: при $u^* = 1$ исследуемая функция не имеет экстремума, это точка перегиба

Примеры исследования функции одной переменной на экстремум

Пример 2. $Q = (1-u)^4$.

Необходимое условие экстремума: $dQ(u)/du\big|_{u*} = 0$

$4(1-u^*)^3(-1) = 0$, $u^* = 1$.

Достаточное условие экстремума

$d^2Q(u)/du^2\big|_{u*} = -12(1-u^*)^2 = -12(1-1)^2 = 0,$

$d^3Q(u)/du^3\big|_{u*} = 24(1-u^*) = 24(1-1) = 0,$

$d^4Q(u)/du^4\big|_{u*} = -24 \neq 0.$

$k = 4$ – четное, $d^4Q(u)/du^4\big|_{u*} < 0$

Ответ: при $u^* = 1$ исследуемая функция имеет минимум.

2. Одномерная оптимизация

2.2 Численные методы прямого поиска экстремума функции одной переменной

Методы одномерной оптимизации

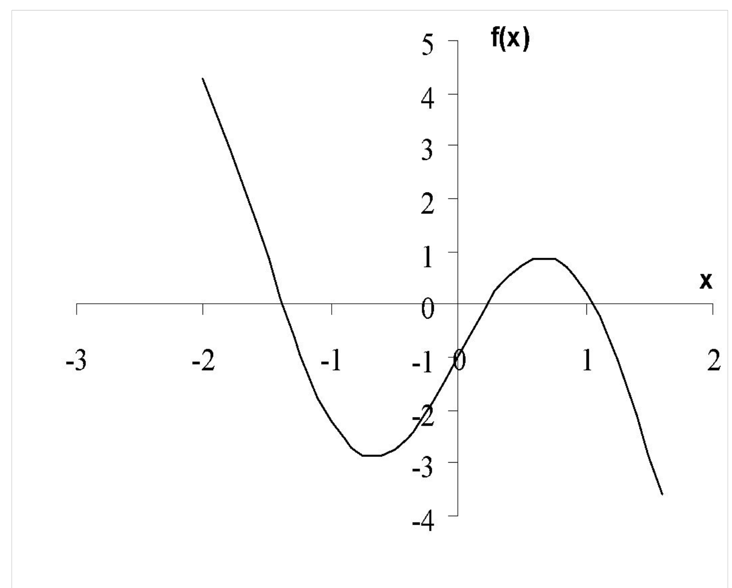
Дана некоторая функция f(x) от одной переменной x, надо определить такое значение x*, при котором функция f(x) принимает экстремальное значение.

Нахождение этих точек с заданной точностью можно разбить на два этапа.

1) Сначала экстремальные точки отделяют, т.е. определяются отрезки, которые содержат по одной экстремальной точке
2) Затем уточняют до требуемой точности ε.

Отделение можно осуществить, как графически, так и табулированием. Все методы уточнения точек экстремумов будем рассматривать относительно уточнения минимума на заданном отрезке.

пример: f(x) = 3*sin(2*x)-1.5*x-1



Унимодальные функции

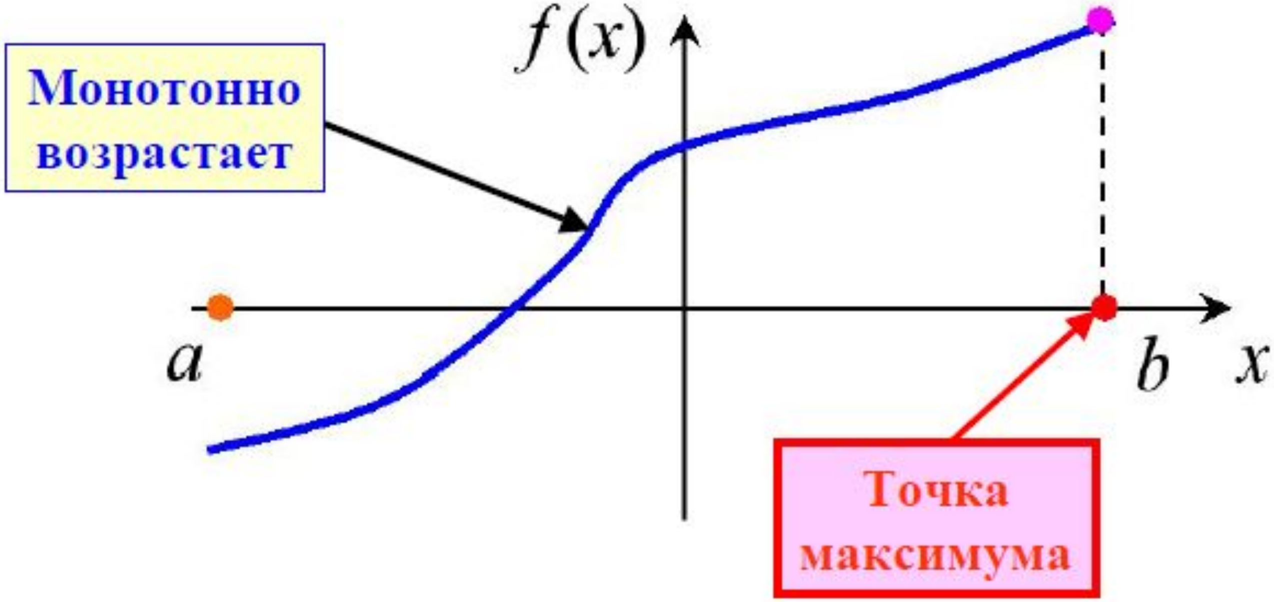
Дадим определение унимодальной функции при поиске минимума.

Определение. Непрерывная функция $f(x)$ называется унимодальной на отрезке $[a, b]$ если:

- точка x^* локального минимума функции принадлежит отрезку $[a, b]$;
- для любых двух точек отрезка x_1 и x_2 , взятых по одну сторону от точки минимума, точке x_1 более близкой к точке минимума соответствует меньшее значение функции, т.е. при $x^* < x_1 < x_2$ либо при $x_2 < x_1 < x^*$ справедливо неравенство $f(x_1) < f(x_2)$.

Унимодальная на отрезке [a, b] функция с максимумом

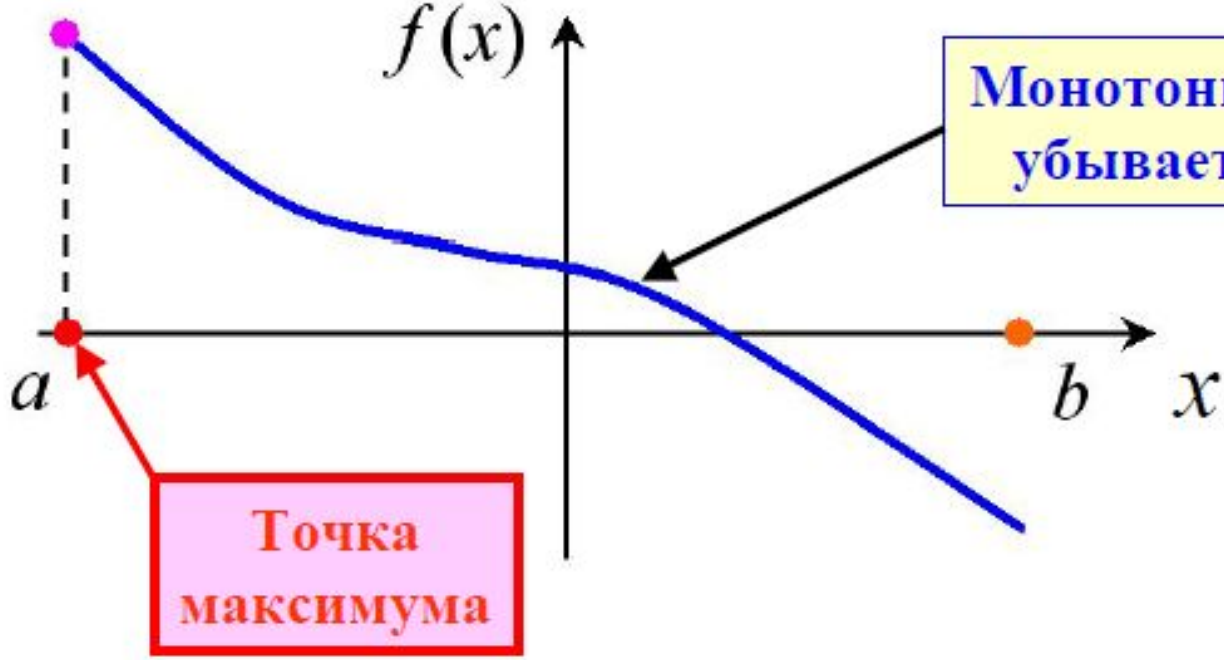
Случай 1



Непрерывная монотонно возрастающая на отрезке [a, b] функция имеет на нем *единственную* точку максимума на *конце* отрезка

Унимодальная на отрезке [a, b] функция с максимумом

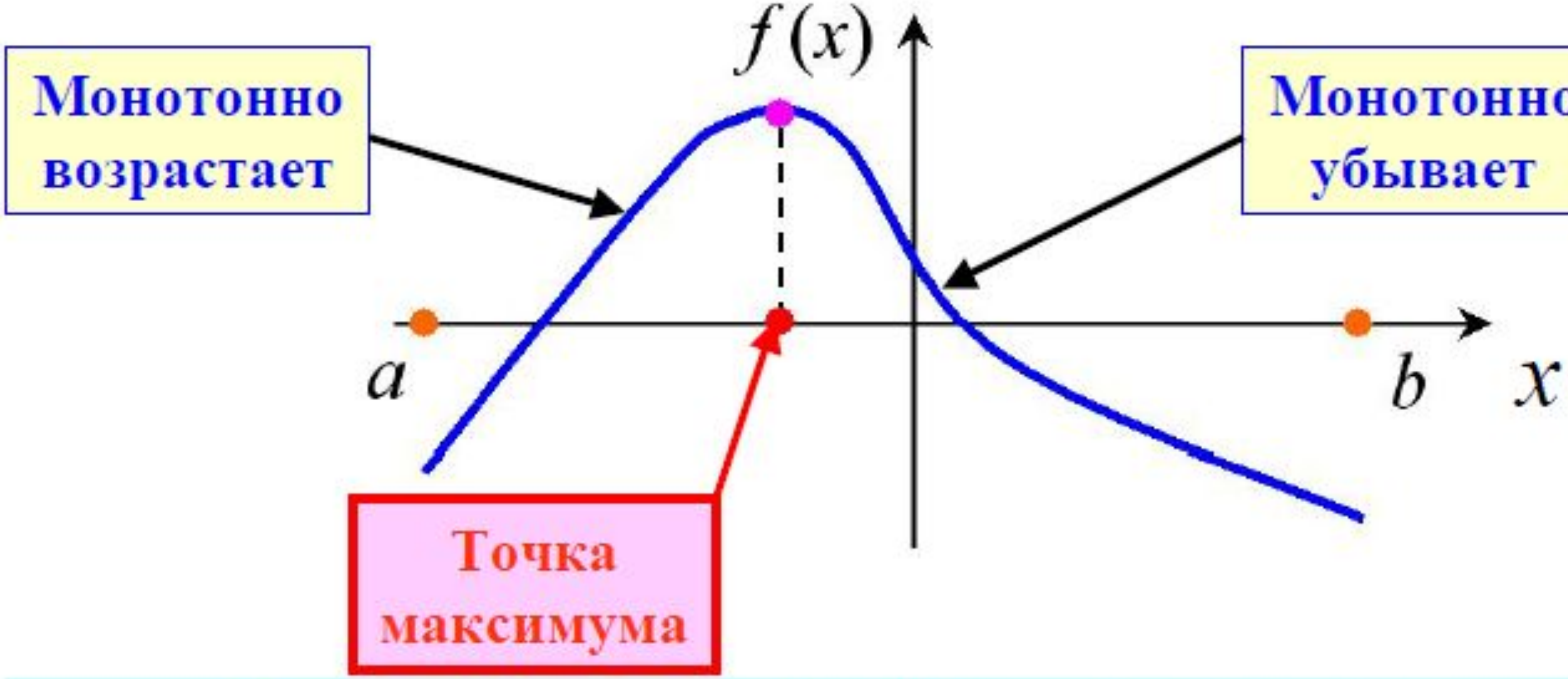
Случай 2



Непрерывная монотонно убывающая на отрезке [a, b] функция имеет на нем *единственную* точку максимума на *конце* отрезка

Унимодальная на отрезке [*a*, *b*] функция с максимумом

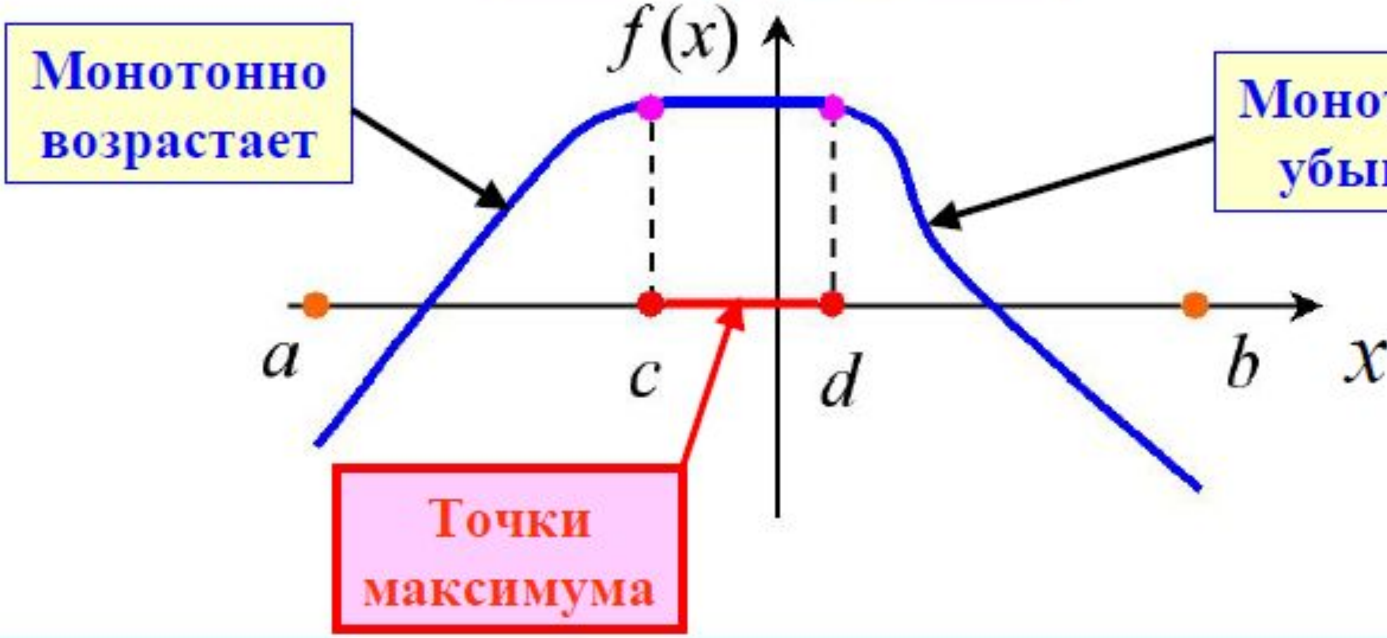
Случай 3



Непрерывная **на отрезке [*a*, *b*] функция имеет на нем *единственную* точку максимума *внутри* этого отрезка.**
Слева **от нее функция *монотонно возрастает*,**
справа – монотонно убывает

Унимодальная на отрезке $[a, b]$ функция с максимумом

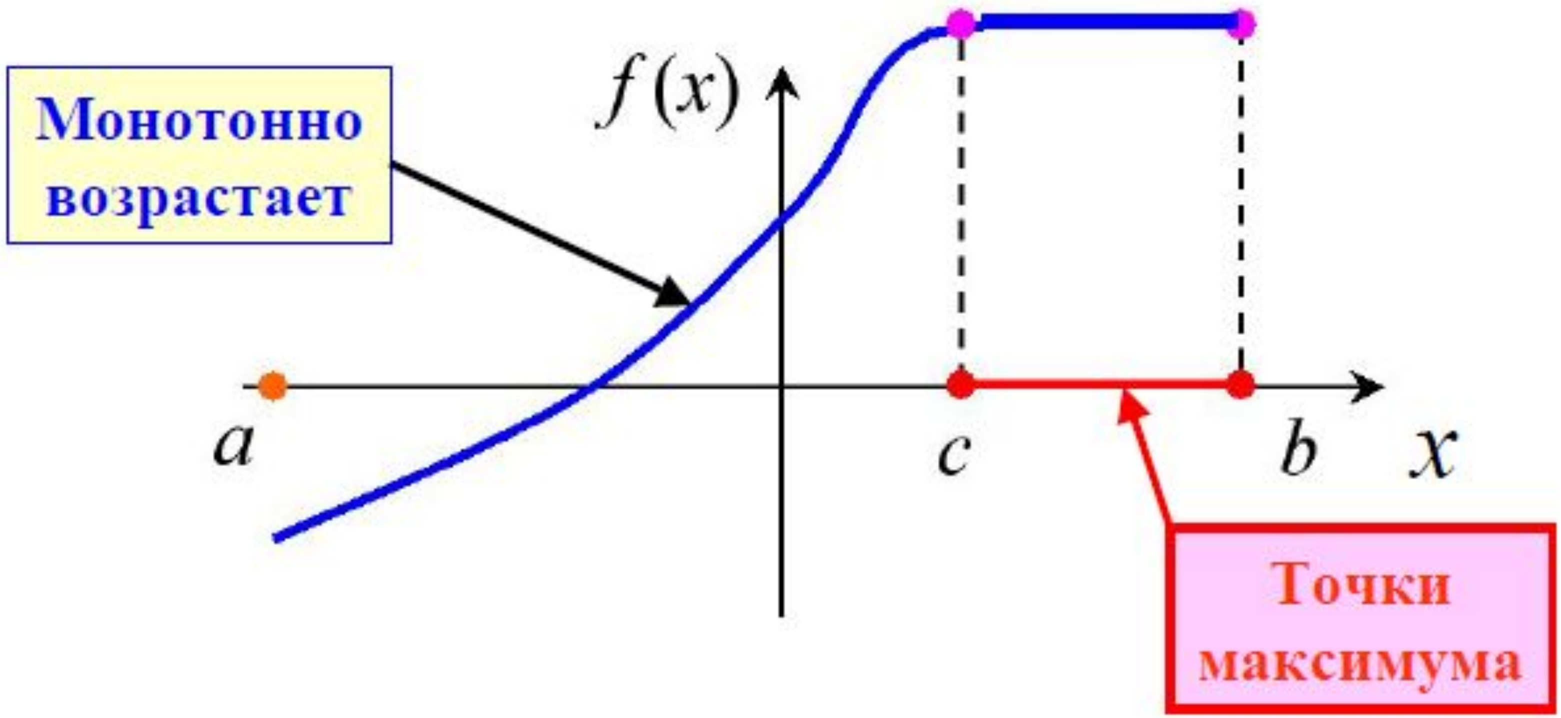
Случай 4



Непрерывная **на отрезке $[a, b]$ функция имеет на нем *бесконечно много* точек максимума, образующих *один* отрезок $[c, d]$ *внутри* отрезка $[a, b]$.**
На отрезке $[a, c]$ функция *монотонно возрастает*.
На отрезке $[d, b]$ функция *монотонно убывает*.

Унимодальная на отрезке [a, b] функция с максимумом

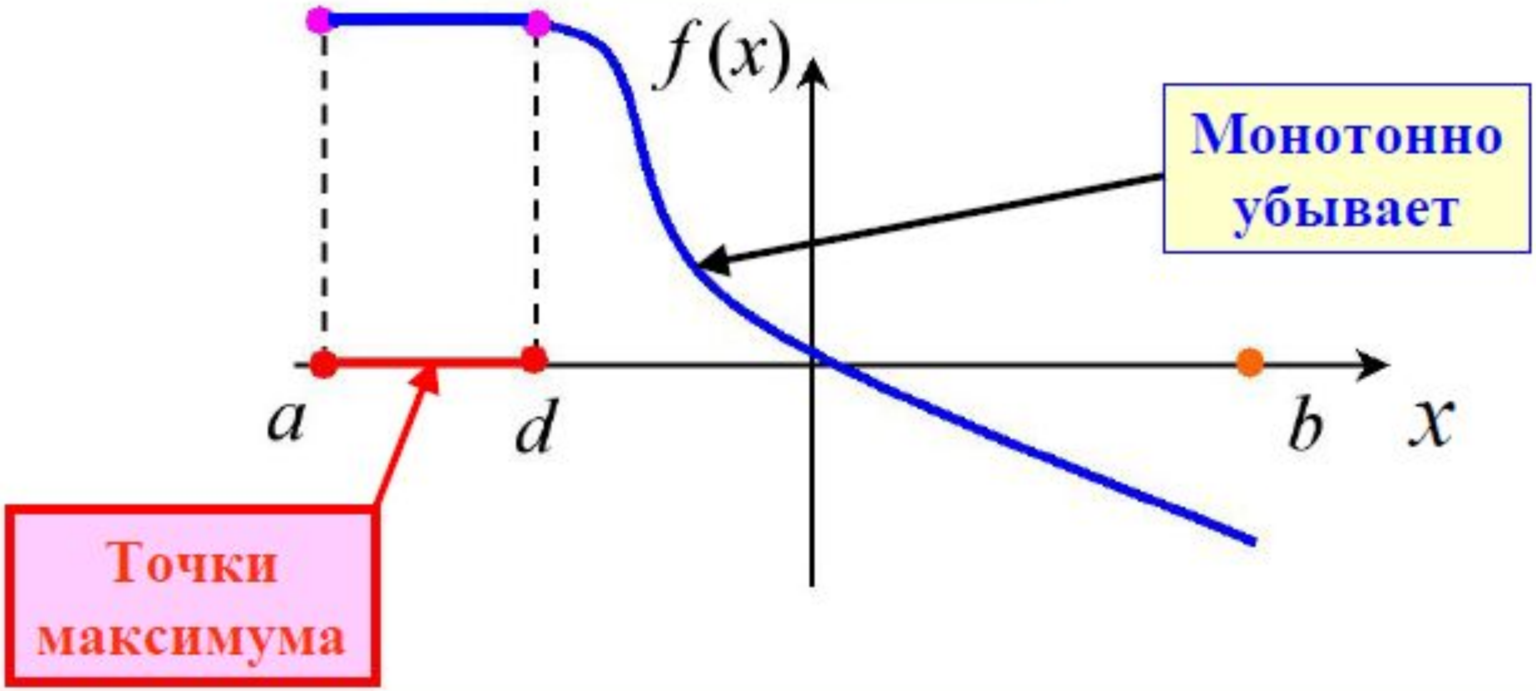
Случай 5



Непрерывная **на отрезке [a, b] функция имеет на нем** ***бесконечно много*** **точек максимума, образующих** ***один*** **отрезок [c, b].**
На отрезке [a, c] функция ***монотонно возрастает.***

Унимодальная на отрезке [a, b] функция с максимумом

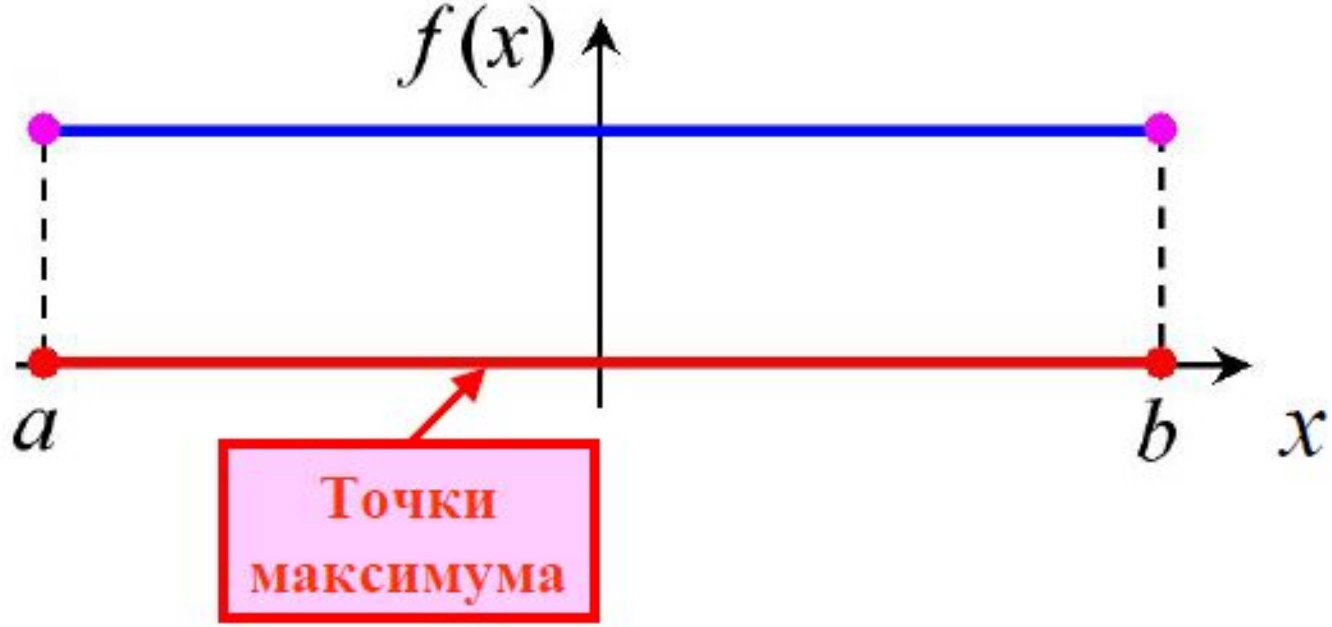
Случай 6



Непрерывная **на отрезке [*a*, *b*] функция имеет на нем** ***бесконечно много*** **точек максимума, образующих** ***один*** **отрезок [*a*, *d*].**
На отрезке [*d*, *b*] функция ***монотонно убывает*****.**

Унимодальная на отрезке [a, b] функция с максимумом

Случай 7



Непрерывная на отрезке [a, b] функция имеет на нем *постоянное* значение.
Есть *бесконечно много* точек максимума, образующих *один* отрезок [a, b].

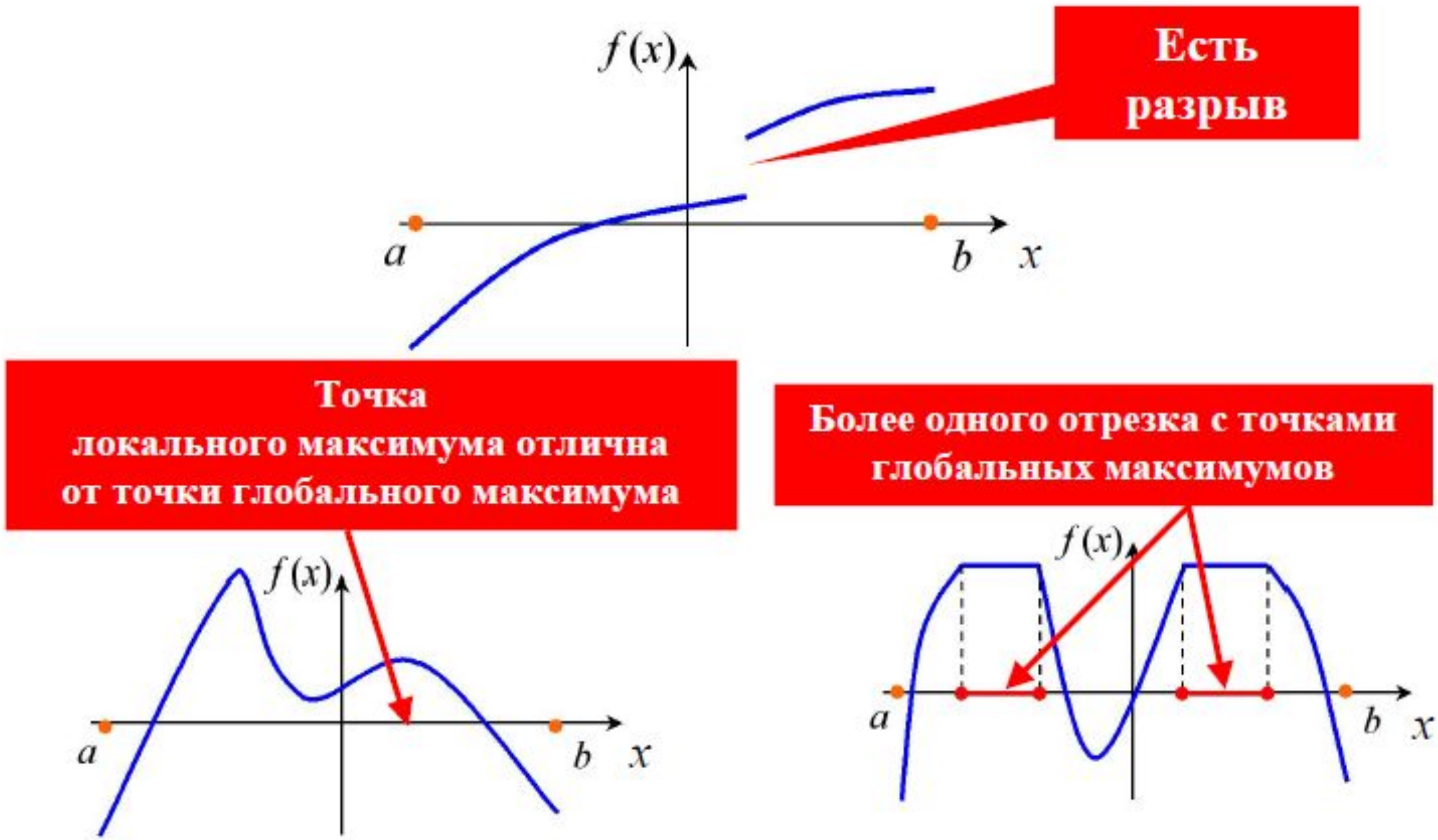
Унимодальная на отрезке функция с <u>максимумом</u>

1. Непрерывна на этом отрезке.

2. *Не имеет* на этом отрезке точек *локальных* максимумов, *отличных от точек глобальных* максимумов.

3. Имеет на этом отрезке *единственную* точку глобального максимума или *бесконечно много* точек глобальных максимумов; эти точки образуют <u>*один*</u> отрезок.

Примеры неунимодальных на отрезке $[a, b]$ функций



$f(x) = x^2 + 2x$

Определение начального отрезка унимодальности

Для определения отрезка унимодальности используем начальные вычисления метода обратного переменного шага, описанные в пункте А.3 для заданной начальной точки $x^{(0)} = 10$ и начальном шаге $\Delta = - 5$.

$f(x_0) = f(10) = 120$;

$f(x_0 + \Delta_0) = f(10 - 5) = 35$; $f(x_0 + \Delta_0) < f(x_0)$;

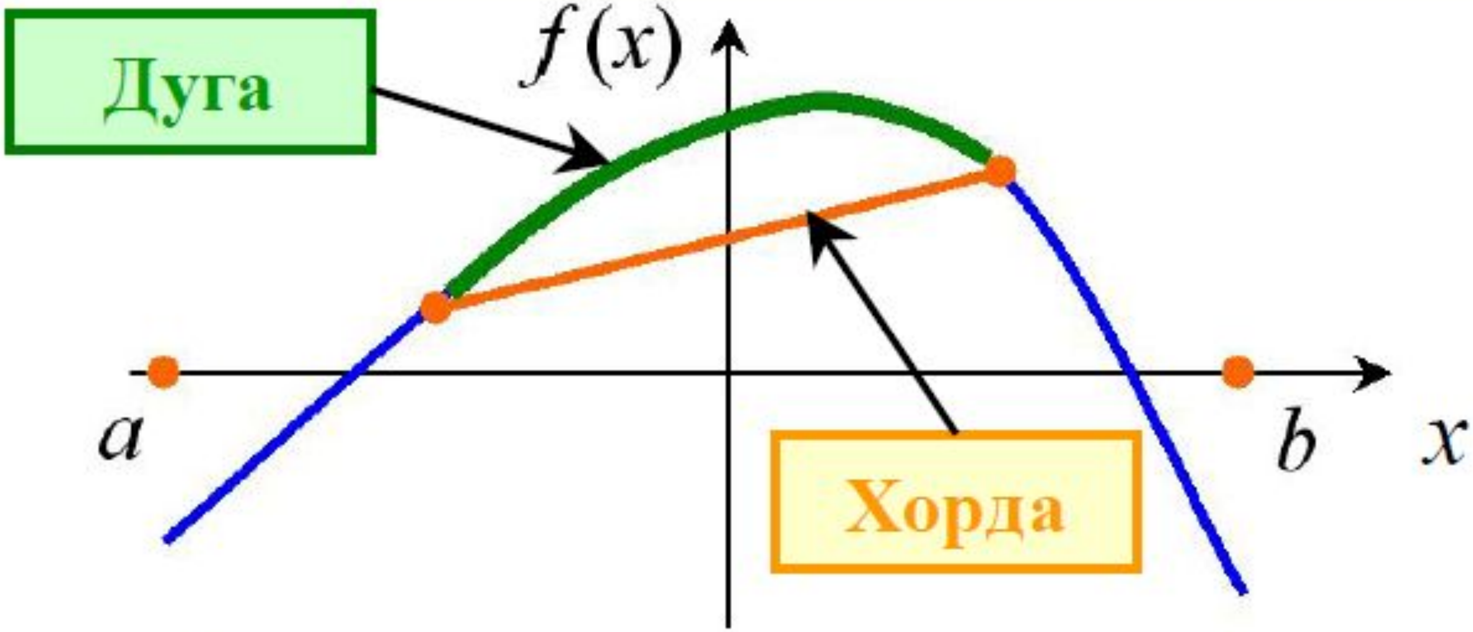
$f(x_0 + 2\Delta_0) = f(10 - 10) = 0$; $f(0) < f(5)$;

$f(x_0 + 3\Delta_0) = f(10 - 15) = 15$; $f(-5) > f(0)$ и т.о. точка минимума $5 > x^* > -5$.

Вывод: минимум функции находится на отрезке [-5, 5] , который можно принять в качестве начального отрезка унимодальности.

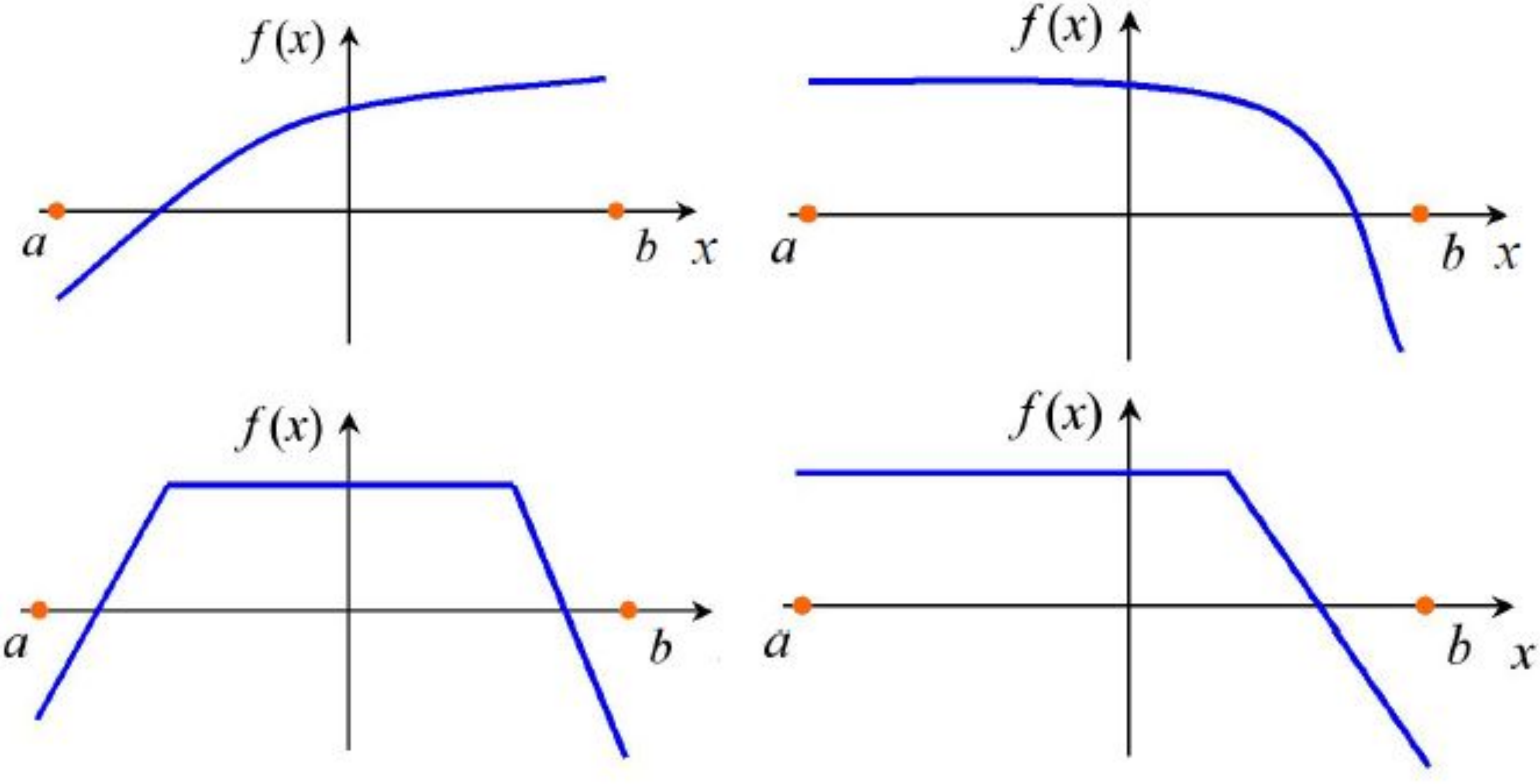
Частный случай унимодальной функции – непрерывная выпуклая функция

Непрерывная выпуклая <u>вверх</u>
на отрезке $[a, b]$ функция

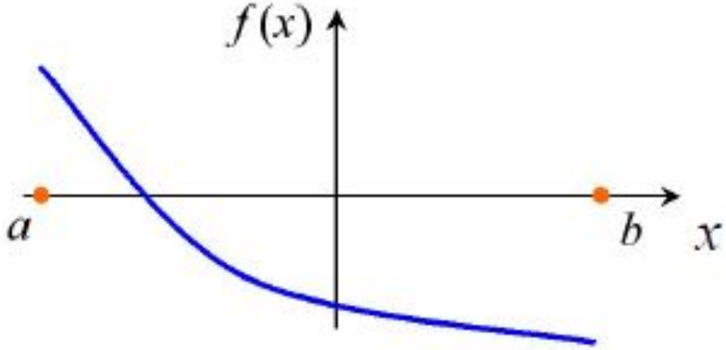


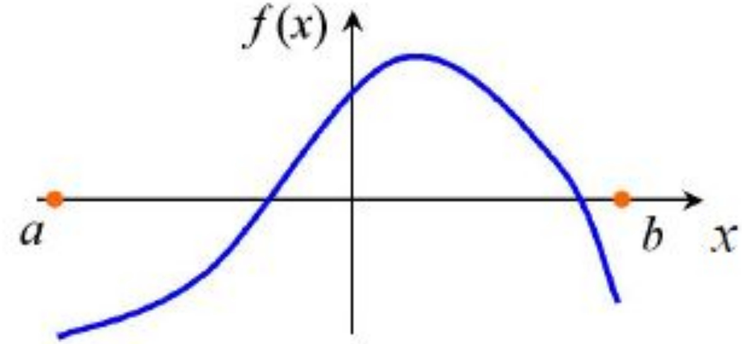
<u>Любая</u> дуга
не ниже
стягивающей ее хорды

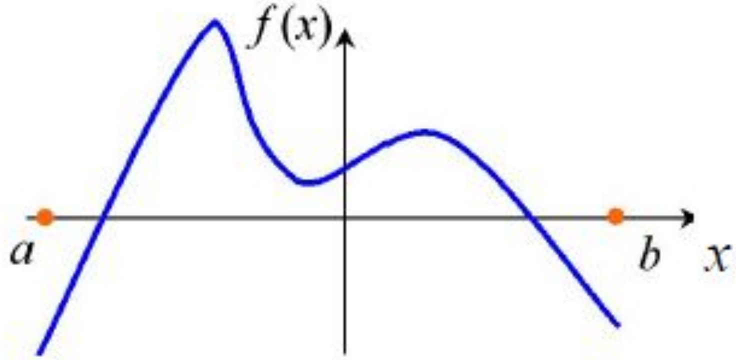
Примеры
непрерывных выпуклых вверх
на отрезке [a, b] функций



Примеры
<u>не</u>выпуклых вверх
на отрезке [*a*, *b*] функций







Связь выпуклости вверх со второй производной

Дважды дифференцируемая на отрезке $[a, b]$ функция $f(x)$ выпукла вверх на этом отрезке, если и только если

$$\boxed{f''(x) \le 0}$$

для всех

$$\boxed{x \in [a, b]}$$

Достаточное условие унимодальности

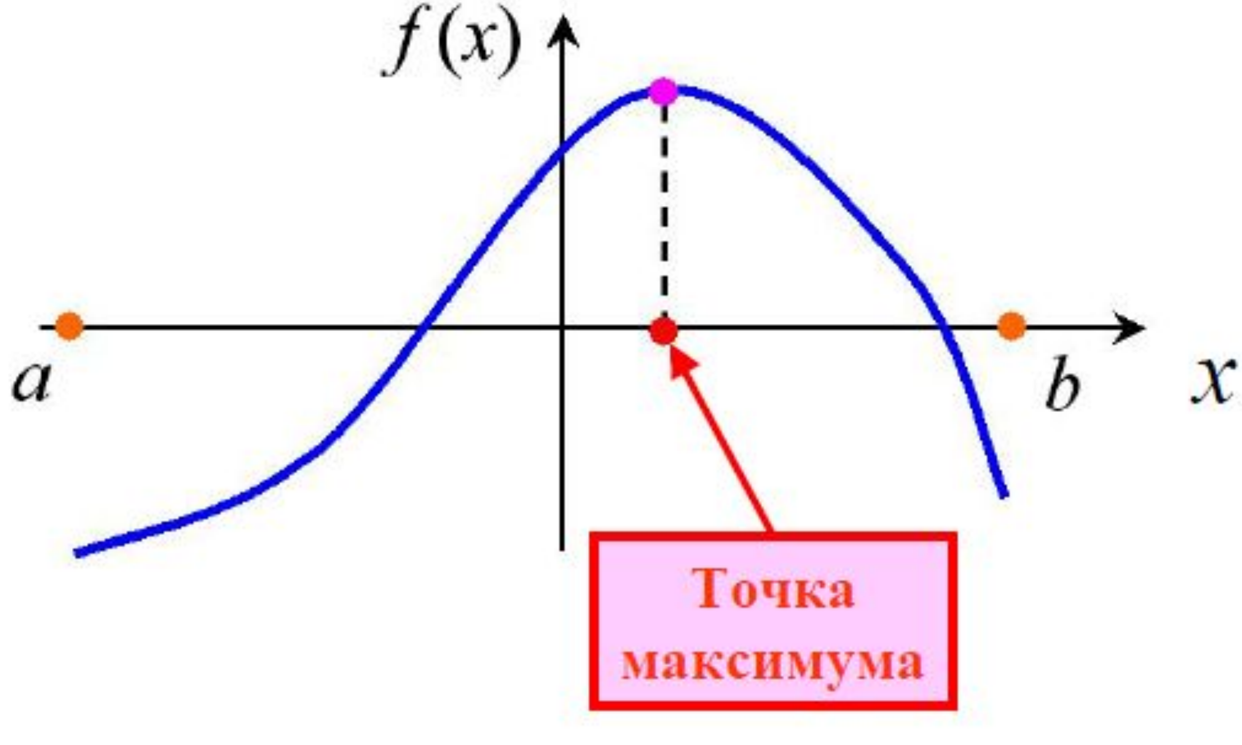
Достаточное условие унимодальности функции $f(x)$ на отрезке $[a, b]$ содержится в следующей теореме.

Теорема. Если функция $f(x)$ дважды дифференцируема на отрезке $[a, b]$ и $f''(x^*) > 0$ в любой точке этого отрезка, то $f(x)$ – унимодальная функция на $[a, b]$.

Заметим, что условие $f''(x^*) > 0$ определяет множество точек, на котором функция является выпуклой (вниз). Условие же $f''(x^*) < 0$ определяет вогнутую функцию, которая на отрезке $[a, b]$ имеет максимум и также является унимодальной.

Может ли
унимодальная
на отрезке
функция с максимумом
не быть выпуклой вверх
на этом отрезке?

Унимодальная
на отрезке [a, b] функция
с максимумом
<u>не</u>выпуклая вверх
на этом отрезке



Стратегия поиска экстремума

Так как в прикладных задачах вычисление каждого значения функции может быть достаточно трудоемким, то целесообразно выбрать такую стратегию поиска, чтобы значение f^* с заданной точностью было найдено наиболее экономным путем.

Будем считать, что стратегия поиска определена, если:

1) определен алгоритм выбора точек x_k, $k = 1, 2, \ldots, N$;

2) определено условие прекращения поиска, т.е. условие, при выполнении которого значение f_* считают найденным с заданной точностью.

Можно выделить две группы *методов прямого поиска*, соответствующие двум принципиально различным ситуациям:

1) все N точек x_k, $k = 1, 2, \ldots, N$, в которых будут вычислены значения функции, выбирают заранее (до вычисления функции в этих точках) – ***пассивный поиск***;

2) точки x_k выбирают последовательно (для выбора последующей точки используют значения функции, вычисленные в предыдущих точках) – ***последовательный поиск***

МЕТОДЫ ОДНОМЕРНОЙ ОПТИМИЗАЦИИ

МЕТОДЫ ПОСЛЕДОВАТЕЛЬНОГО ПОИСКА

- **МЕТОДЫ НУЛЕВОГО ПОРЯДКА**
  - метод половинного деления
  - метод «золотого»сечения
  - метод Фибоначчи
- **МЕТОДЫ ПОЛИНОМИАЛЬНОЙ АППРОКСИМАЦИИ**
- **МЕТОДЫ С ИСПОЛЬЗОВАНИЕМ ПРОИЗВОДНЫХ**

МЕТОДЫ ПАССИВНОГО ПОИСКА

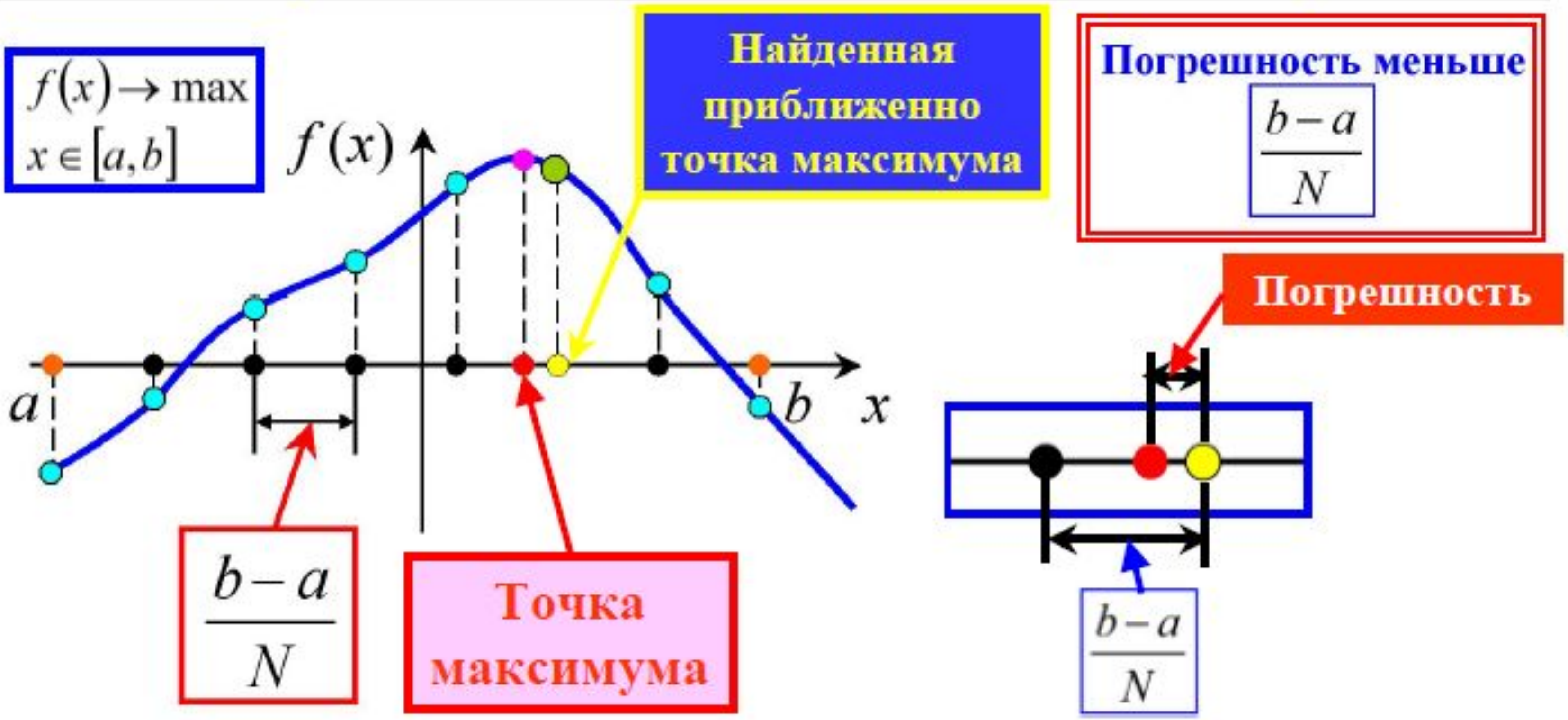
Оптимальный пассивный поиск состоит в выборе точек, равномерно расположенных на отрезке $[a, b]$, координаты которых

$$x_k = \frac{(b-a)\,k}{N+1}, \quad k = 1, 2, \ldots, N.$$

При этом длина интервала, содержащего минимум $l_N = \dfrac{2(b-a)}{N+1}$ дает оценку скорости сходимости пассивного поиска с ростом числа N точек, так как скорость сходимости любого метода прямого поиска можно характеризовать скоростью уменьшения интервала неопределенности с возрастанием N. Вычисляем значения целевой функции в каждой точке $f_k = f(x_k)$, найдем наименьшее значение f_j и, тогда точка оптимума $x^* \in \left(x_{j-1}, x_{j+1}\right)$ и можно считать, что $x^* \cong x_j \pm \varepsilon$ с точностью ε.

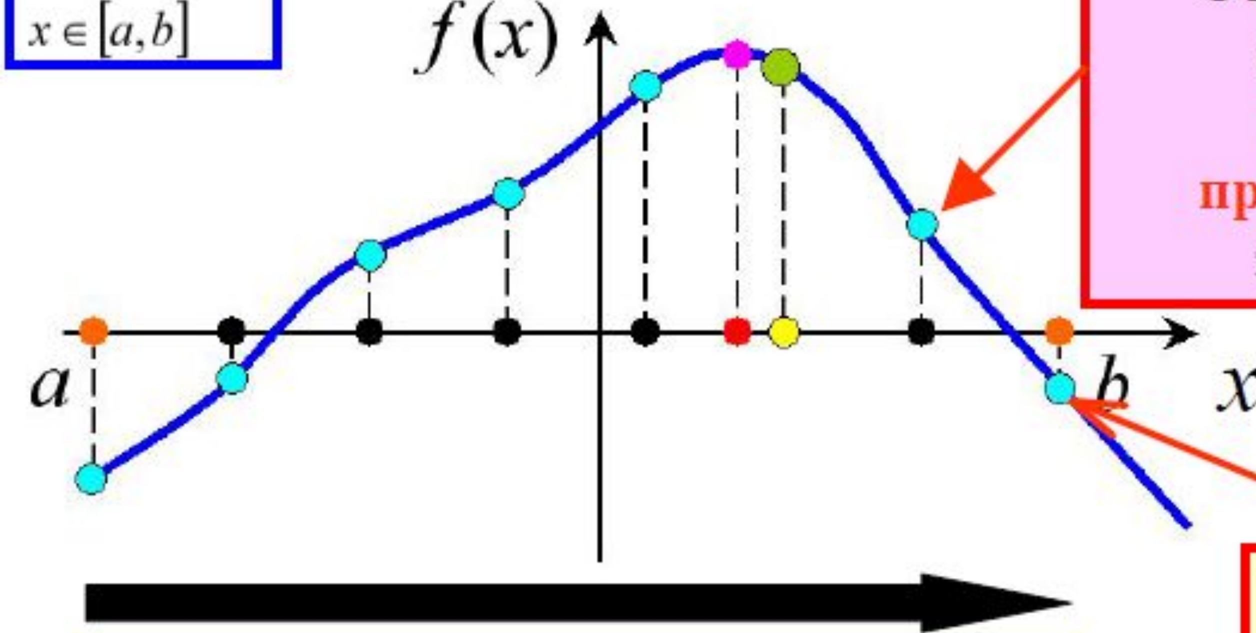
Метод перебора

Отрезок [*a*, *b*] делится на *N* равных частей.
В точках *a*, *b* и в точках деления
вычисляются значения целевой функции.
Среди них находится максимальное значение.
Ему приближенно соответствует точка максимума.



Унимодальная функция. Метод перебора

$$f(x) \to \max$$
$$x \in [a,b]$$



Это значение функции меньше предыдущего значения

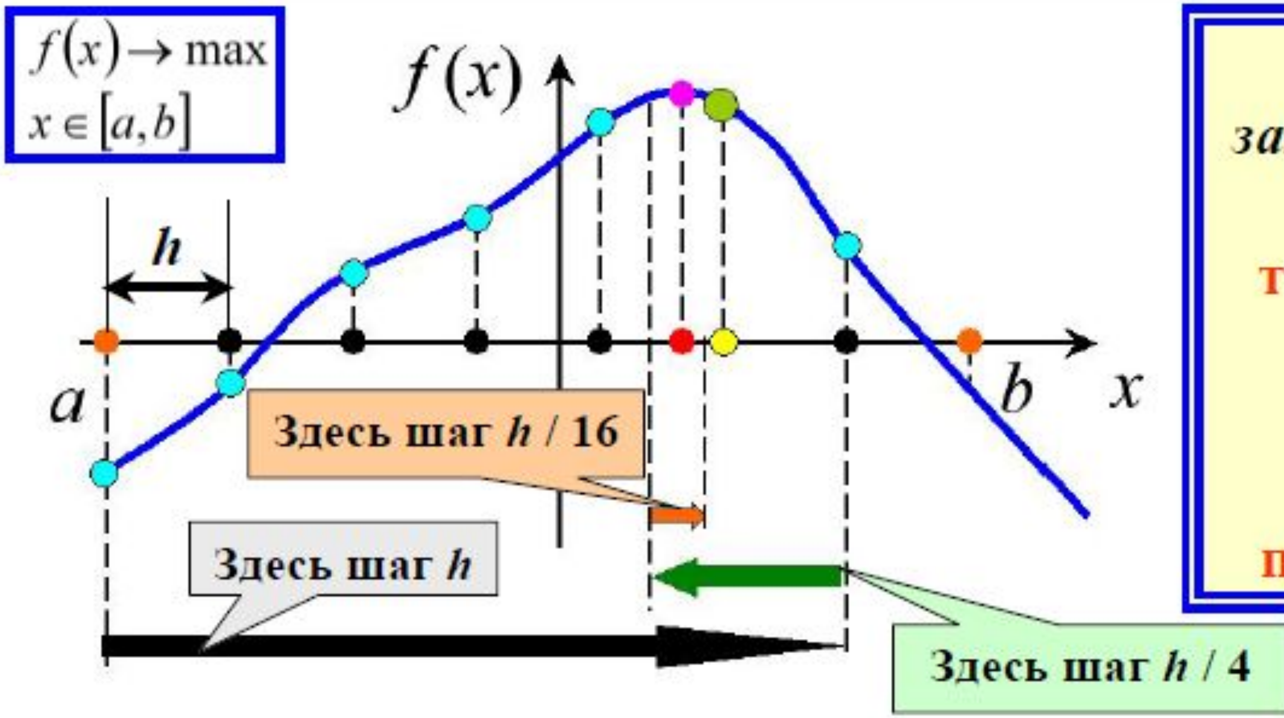
Не нужно вычислять это значение функции

Значения *x* изменялись от *a* до *b*

Когда очередное вычисленное значение функции оказалось меньше предыдущего значения, надо «изменить направление движения»

Метод **поразрядного поиска**

Идем от a к b с шагом h.
Когда значение функции меньше предыдущего, идем к a с шагом h / 4.
Когда значение функции меньше предыдущего, идем к b с шагом h / 16, и т.д.



Поиск *заканчивается*, когда текущий шаг станет меньше заданной погрешности

Методы последовательного сокращения отрезка унимодальности

Основой многих одномерных численных методов является сокращение отрезка унимодальности, а именно: построение последовательности отрезков $[a_k, b_k]$, стягивающихся к точке x^* – минимуму функции на исходном отрезке. Методы оптимизации отличаются друг от друга лишь различным выбором точек на начальном отрезке унимодальности.

Общая последовательность реализации методов:

- выбор точек на начальном отрезке унимодальности;
- вычисление значений функции в этих точках и сравнение этих значений;
- определение нового отрезка;
- проверка критерия останова.

Общая схема сужения промежутка унимодальности

В методах, рассматриваемых далее, для дальнейшего сужения промежутка унимодальности используют следующую идею.

Возьмем две точки x_1 и x_2 , принадлежащие начальному отрезку $[a_0\ ,\ b_0]$ такие, что $x_1 < x_2$. В каждом из трех следующих очевидных случаев можно указать отрезок меньших размеров $[a_1,\ b_1]$, содержащий точку минимума x^* и принадлежащий первоначальному отрезку (рисунок 1):

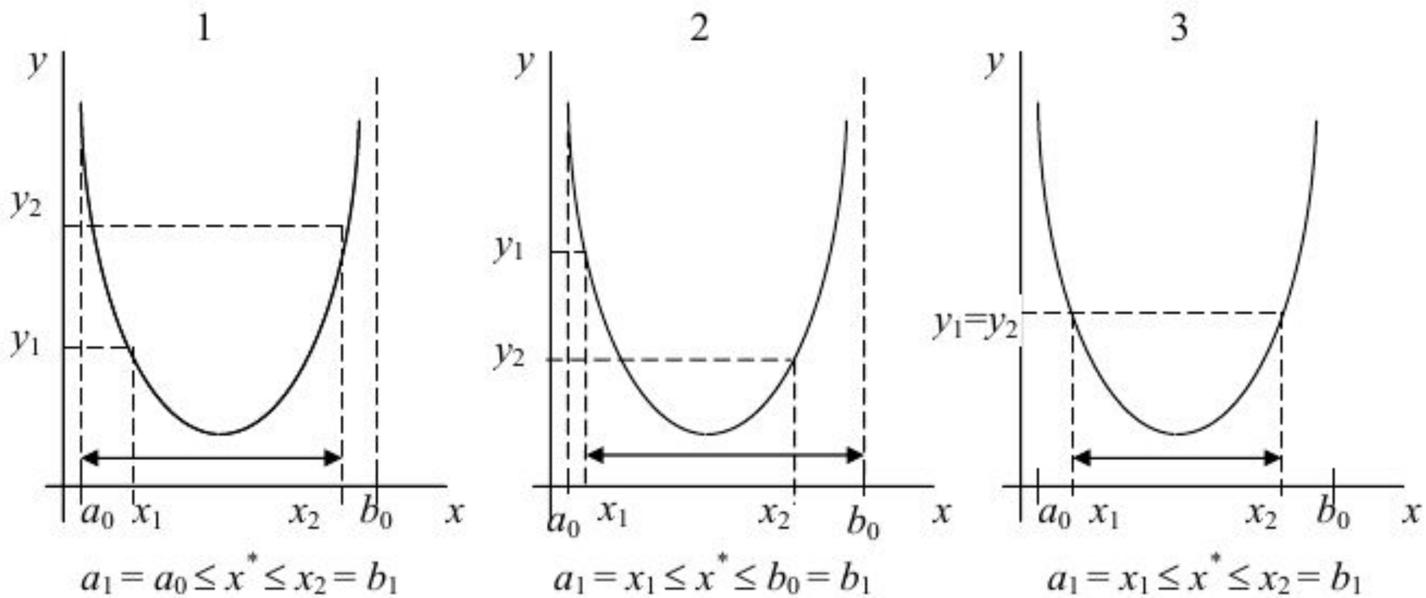


Рисунок 1 – Возможные ситуации при сужении отрезка

1 Если $y_1 = f(x_1) < y_2 = f(x_2)$, то положим $a_1 = a_0$ и $b_1 = x_2$ и получим меньший отрезок унимодальности $[a_1,\ b_1]$.
2 Если $y_1 = f(x_1) > y_2 = f(x_2)$, то естественно принять $a_1 = x_1$ и $b_1 = b_0$.
3 Если $y_1 = f(x_1) = y_2 = f(x_2)$, то $a_1 = x_1$ и $b_1 = x_2$.

Метод половинного деления

Метод половинного деления, называемый также ***методом дихотомии***, является процедурой последовательного поиска. Пусть определен отрезок $[a_0, b_0]$, которому принадлежит точка локального минимума x^*, и функция $f(x)$ является унимодальной на этом отрезке. Далее для сужения промежутка унимодальности используем две точки x_1 и x_2, расположенные симметрично на расстоянии $\delta > 0$ от середины отрезка:

$$x_1 = \frac{a_0 + b_0}{2} - \delta \ ;$$

$$x_2 = \frac{a_0 + b_0}{2} + \delta \ .$$

Константа δ должна быть меньше допустимой конечной длины отрезка, $\Delta_k = b_k - a_k > 0$.

Рассчитываем значение функции в этих точках $y_1=f(x_1)$ и $y_2=f(x_2)$ и в зависимости от их соотношения новые границы отрезка унимодальности $[a_1, b_1]$ будут следующие:

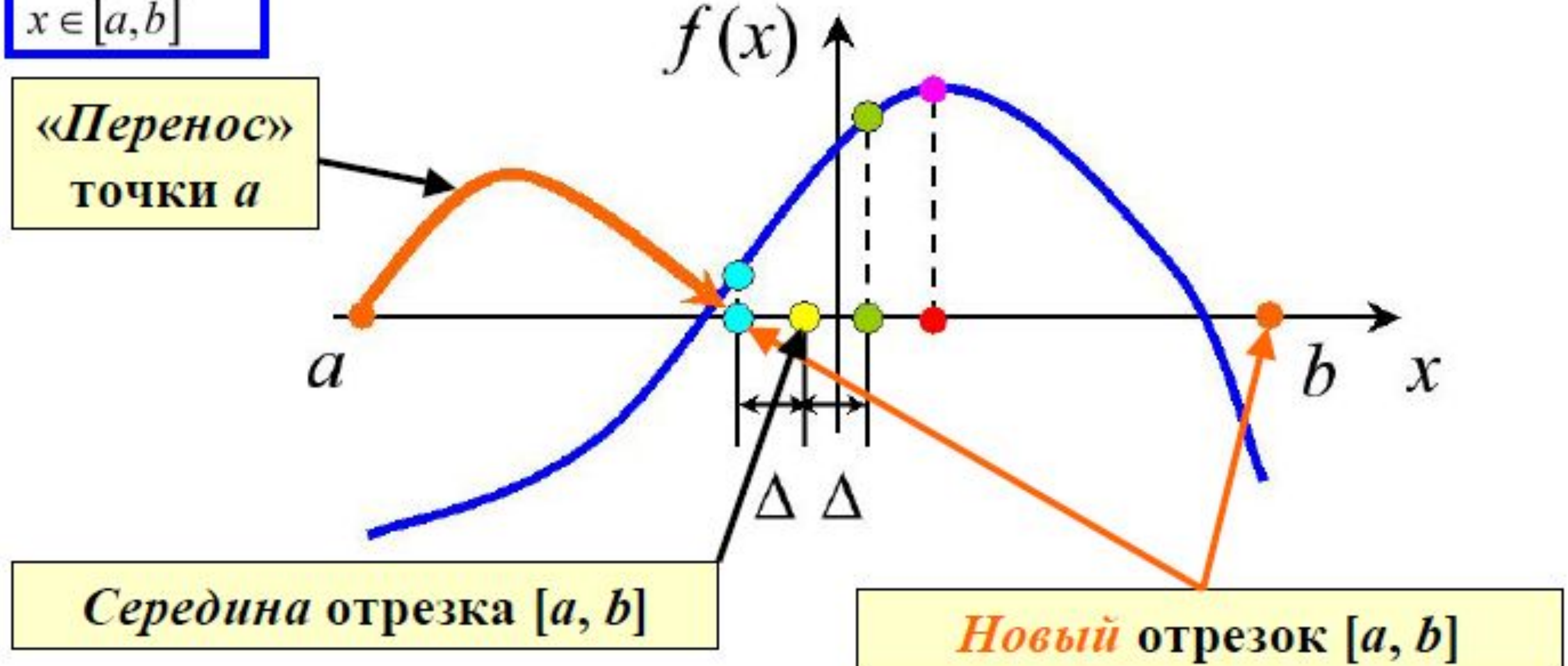
- $y_1 < y_2$, $a_1=a_0$ и $b_1=x_2$;
- $y_1 > y_2$, $a_1=x_1$ и $b_1=b_0$;
- $y_1 = y_2$, $a_1=x_1$ и $b_1= x_2$.

Название *метода половинного деления* мотивировано тем, что если величина ε достаточно мала, то длина отрезка унимодальности $(b - a)$ уменьшается почти вдвое.

Метод дихотомий

Вычисляем значения целевой функции в двух точках: слева и справа от середины отрезка [*a*, *b*]. Если значение в правой точке больше, чем в левой, то «переносим» точку *a* в левую точку.

$$f(x) \to \max$$
$$x \in [a, b]$$



Когда надо закончить поиск в методе дихотомии?

Поиск *прекращается*, когда

$$b - a \le 2\varepsilon$$

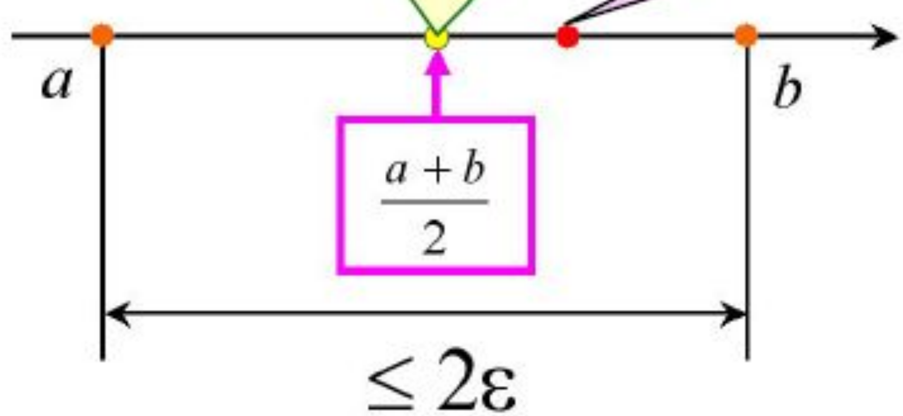
ε — заданная погрешность вычисления точки максимума

Выбираем

$$\Delta < 2\varepsilon$$

Вычисленная *приближенно* точка максимума – середина *последнего* отрезка [*a*, *b*]

Истинная точка максимума



Отсюда можно вычислить число итераций для достижения необходимой точности ε, потребуется $n \geq \dfrac{\ln\left((b_0 - a_0)/\varepsilon\right)}{\ln 2}$ итераций. На каждой итерации минимизируемая функция вычисляется дважды.

3.2. Метод деления пополам. На каждой итерации исключается половина интервала.

АЛГОРИТМ (рис. 5)

Шаг 0. Зададим точность ε>0

Шаг 1. Найти $x^* = \dfrac{a+b}{2}$ и $l = b - a$. Вычислить $f(x^*)$.

Шаг 2. Найти $x_1 = a + l/4$ и $x_2 = b - l/4$. Вычислить $f(x_1), f(x_2)$

Шаг 3. Если $f(x_1) < f(x^*)$, то исключается интервал (x^*, b), при этом

$b = x^*, \quad x^* = x_1$; перейти к п. 5, иначе перейти к п. 4.

Шаг 4. Если $f(x_2) < f(x^*)$, то исключается интервал (a, x^*), при этом $a = x^*, \quad x^* = x_2$; перейти к п. 5. Иначе исключить интервалы $(a, x_1), \quad (x_2, b)$, то есть $a = x_1, b = x_2$; перейти к п. 4.

Шаг 5. Вычислить $l = b - a$. Если $l \le \varepsilon$, то закончить поиск. Иначе перейти к п. 2.

Метод золотого сечения

Термин "золотое сечение" ввел Леонардо да Винчи. Точка x_1 является золотым сечением отрезка $[a, b]$, если отношение длины b-a всего отрезка к длине b-x_1 большей части равно отношению длины большей части к длине x_1–a меньшей части (рисунок 2), т.е. x_1 – золотое сечение, если справедливо соотношение $\frac{b-a}{b-x_1}=\frac{b-x_1}{x_1-a}$. Аналогично, точка x_2 симметричная точке x_1 относительно середины отрезка $[a, b]$, является вторым золотым сечением этого отрезка.

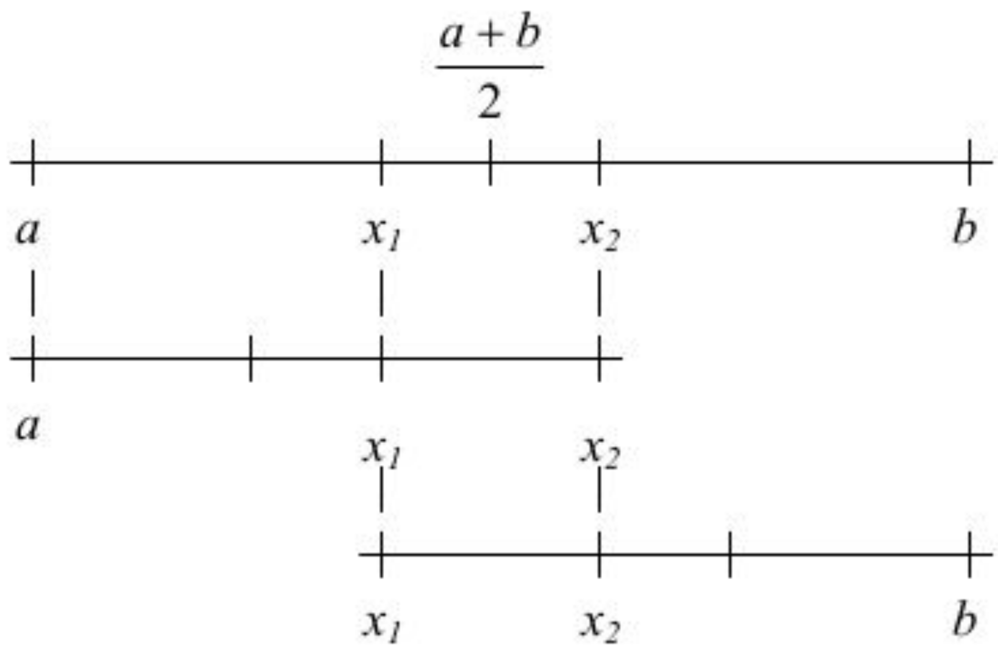


Рисунок 2 – Метод золотого сечения

Отметим свойство золотого сечения: точка x_1 одновременно является золотым сечением отрезка $[a, x_2]$, а другая точка x_2 - золотым сечением отрезка $[x_1, b]$.

Суть метода золотого сечения заключается в следующем. Сначала на исходном отрезке $[a_0, b_0]$ находятся точки x_1 и x_2 по следующим формулам:

$$x_1= a_0 +(1-k)\cdot(b_0-a_0) ;$$

$$x_2= a_0 +k\cdot(b_0-a_0) ;$$

где $k = \dfrac{\sqrt{5}-1}{2} = 0{,}618$ – коэффициент сжатия.

Затем вычисляются значения функции в точках x_1 и x_2 , т.е. $y_1=f(x_1)$ и $y_2=f(x_2)$. При этом возможны два случая:

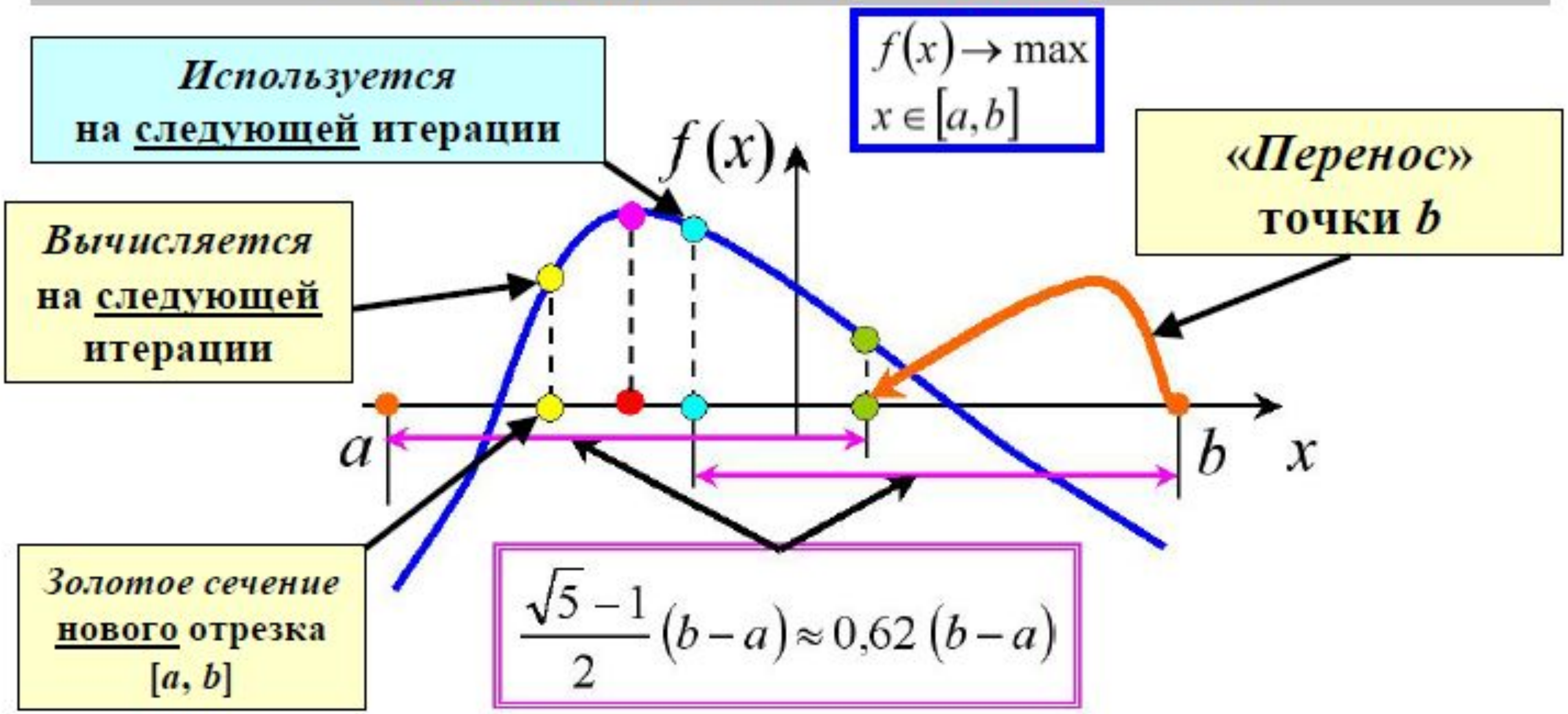
1 $y_1 < y_2$, в этом случае новый отрезок будет равен: $a_1=a_0$ и $b_1=x_2$. В этом отрезке вновь выбираются две точки: $x_1^{(1)} = a_1+ (1- k)\cdot(b_1-a_1)$ и $x_2^{(1)}=x_1$.

2 $y_1 < y_2$, тогда новый отрезок будет составлять: $a_1=x_1$ и $b_1=b_0$. В новом отрезке также выбираются две точки: $x_1^{(1)} = x_2$ и $x_2^{(1)}= a_1 + k\cdot(b_1 - a_1)$.

И в первом и во втором случаях рассчитывается лишь одна новая точка (вторая известна). В новой точке рассчитывается значение функции и вновь производится сравнение в двух точках, и в зависимости от этого выбирается новый отрезок. Процедура повторяется до тех пор, пока не будет выполняться условие $(b_k - a_k) \leq \varepsilon$, где ε – точность поиска.

Метод золотого сечения

Вычисляем значения целевой функции в двух точках:
левой и правой,
полученных *золотыми сечениями* отрезка [*a*, *b*].
Если значение в левой точке не меньше, чем в правой,
то «переносим» точку *b* в правую точку.



Метод Фибоначчи

В методе Фибоначчи требуется, чтобы общее число n вычислений функции было выбрано заранее, так как точки, в которых производится вычисление, определяются по формулам:

$$x_1^{(k)} = a_k + \frac{F_{n-k-2}}{F_{n-k}} \cdot (b_k - a_k), \qquad k=0, ..., n\text{-}2\,;$$

$$x_2^{(k)} = a_k + \frac{F_{n-k-1}}{F_{n-k}} \cdot (b_k - a_k), \qquad k=0, ..., n\text{-}2\,;$$

где $F_i = F_{i-1} + F_{i-2}$, $i=1, 2, \ldots$ $F_0 = F_1 = 1$ - называется последовательностью чисел Фибоначчи (1, 1, 2, 3, 5, 8, 13, 21, 34, 55, ... , т.е. каждый член последовательности рассчитывается как сумма двух предыдущих членов).

Алгоритм поиска.

Выбирается начальный отрезок $[a_0, b_0]$ и число вычислений n таким образом, чтобы $F_n > \dfrac{b_0 - a_0}{\Delta}$, где $\Delta > 0$ – конечная длина отрезка.

Затем рассчитываются координаты двух точек:

$$x_1^{(0)} = a_0 + \frac{F_{n-2}}{F_n} \cdot (b_0 - a_0),$$

$$x_2^{(0)} = a_0 + \frac{F_{n-1}}{F_n} \cdot (b_0 - a_0)$$

и значение функции в этих точках $y_1=f(x_1^{(0)})$ и $y_2=f(x_2^{(0)})$.

В случае $y_1 < y_2$

$$a_1=a_0 \text{ и } b_1=x_2^{(0)}, \qquad x_1^{(1)} = a_1 + \frac{F_{n-3}}{F_{n-1}} \cdot (b_1 - a_1), \qquad x_2^{(1)}= x_1^{(0)}.$$

В случае $y_1 > y_2$

$$a_1= x_1^{(0)} \text{ и } b_1= b_0, \qquad x_1^{(1)}= x_2^{(0)}, \qquad x_2^{(1)} = a_1 + \frac{F_{n-2}}{F_{n-1}} \cdot (b_1 - a_1).$$

Таким образом данная процедура повторяется $(n – 2)$ раза.

При $k = (n – 2)$ точки x_k и y_k совпадают и соответствуют середине отрезка, поэтому чтобы обеспечить дальнейшее сокращение отрезка, точка последнего вычисления функции перемещается вправо на величину константы различимости $\delta > 0$, которая выбирается заранее существенно меньше заданной точности.

Метод квадратичной аппроксимации

Основан на аппроксимации функции полиномом второго порядка в некоторой окрестности и расчета на его основе координаты точки оптимума.

Пусть известны значения функции в трех точках x_0 , x_1 , x_2 , составляющие соответственно y_0 , y_1 , y_2 . Тогда функцию $f(x)$ можно аппроксимировать полиномом

$$g(x)=a_0 + a_1(x- x_0) + a_2(x- x_0)\,(x- x_1)$$

с коэффициентами

$$a_0 = y_0 \quad ;$$

$$a_1 = \frac{y_1 - y_0}{x_1 - x_0} \quad ;$$

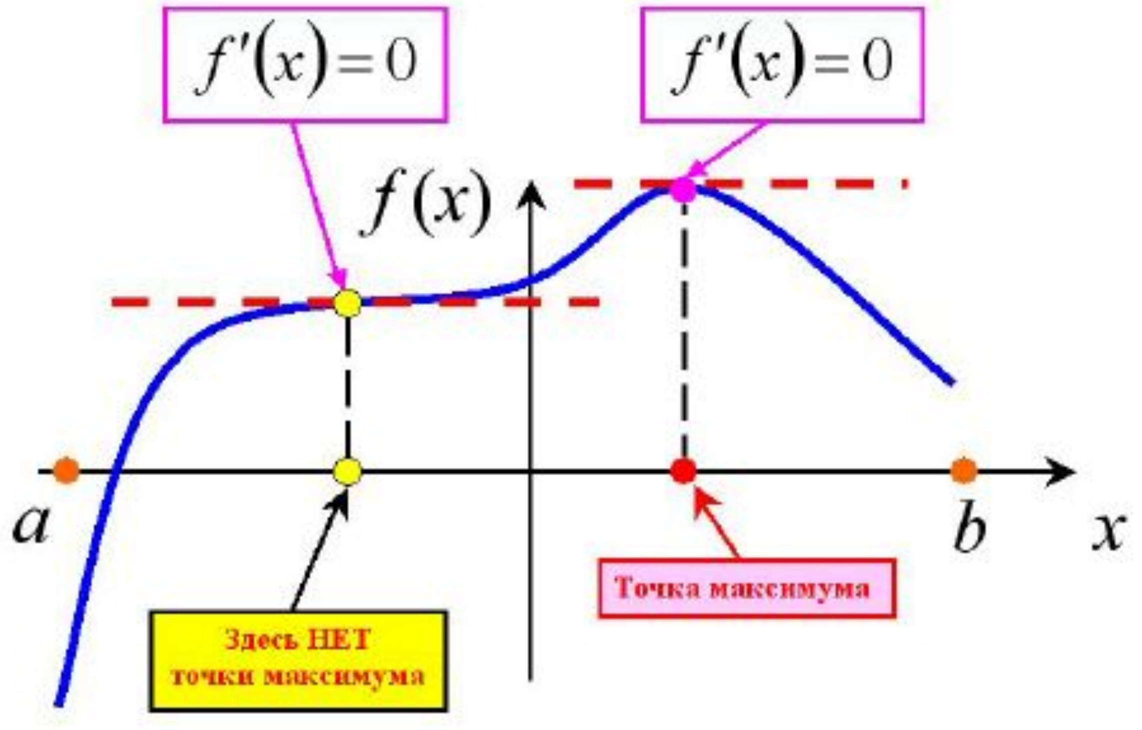
$$a_2 = \frac{1}{x_2 - x_1}\left(\frac{y_2 - y_0}{x_2 - x_0} - \frac{y_1 - y_0}{x_1 - x_0}\right) \quad .$$

Оптимальное значение оценивается по формуле

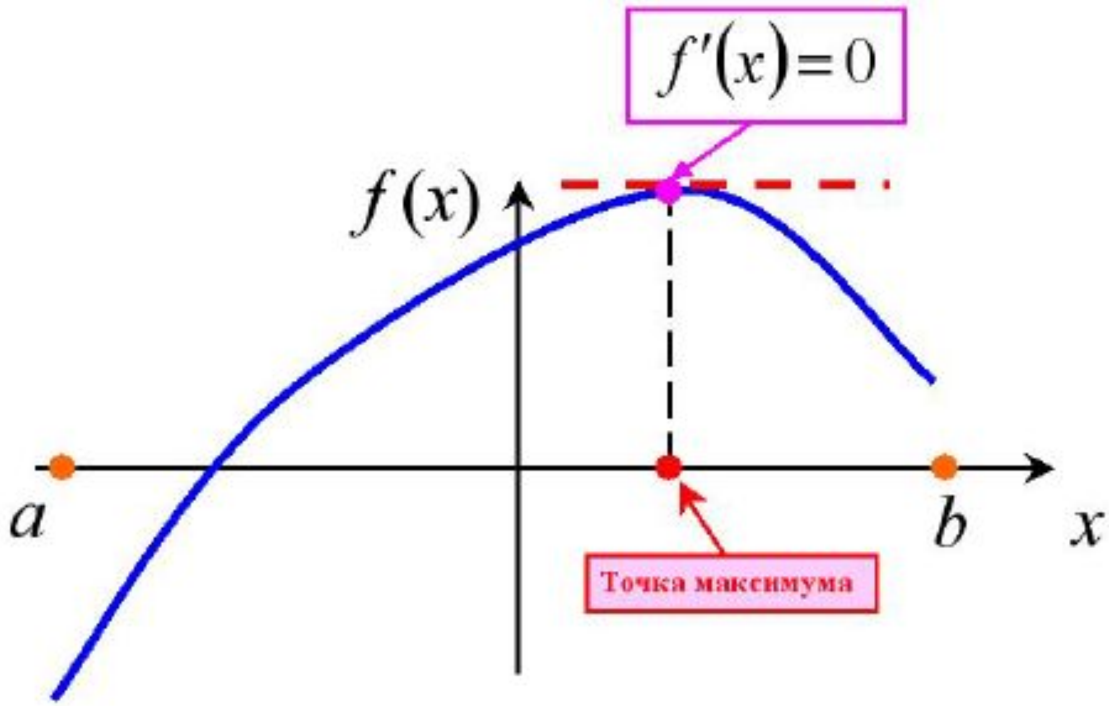
$$x^* \approx \frac{x_1 + x_0}{2} - \frac{a_1}{2a_2} \quad .$$

Методы одномерной нелинейной оптимизации с использованием производных

Для <u>произвольной</u> дифференцируемой на отрезке функции равенство нулю первой производной *внутри* отрезка – это необходимое, но <u>не</u> достаточное условие экстремума

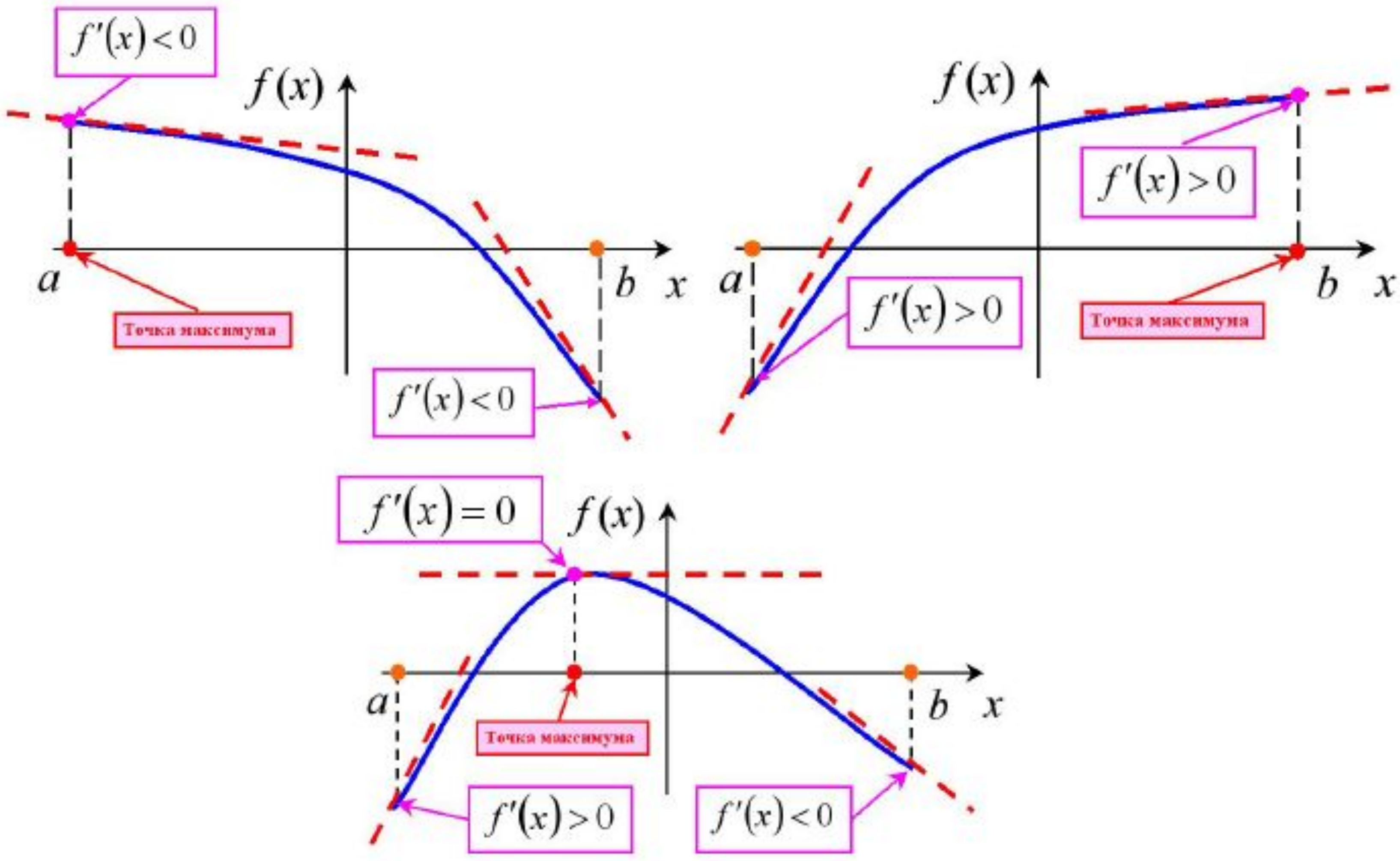


Для дифференцируемой выпуклой на отрезке функции равенство нулю первой производной – это достаточное условие экстремума



Три случая

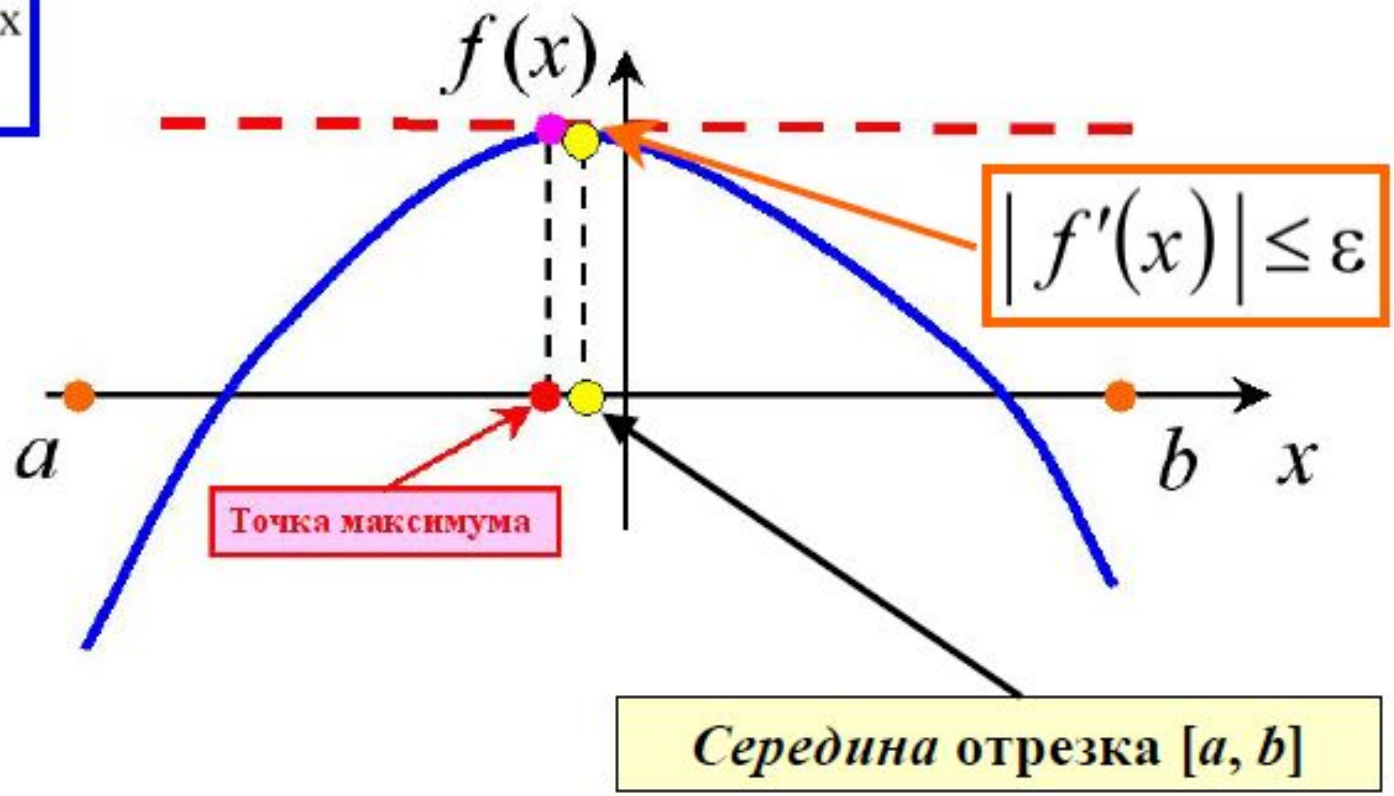
для дифференцируемой выпуклой вверх на отрезке [*a*, *b*] функции



Метод средней точки

Вычисляем первую производную целевой функции
в середине отрезка [a, b].
Если *модуль* первой производной в середине отрезка
не более заданной погрешности,
то эта середина – *приближенно* точка максимума.

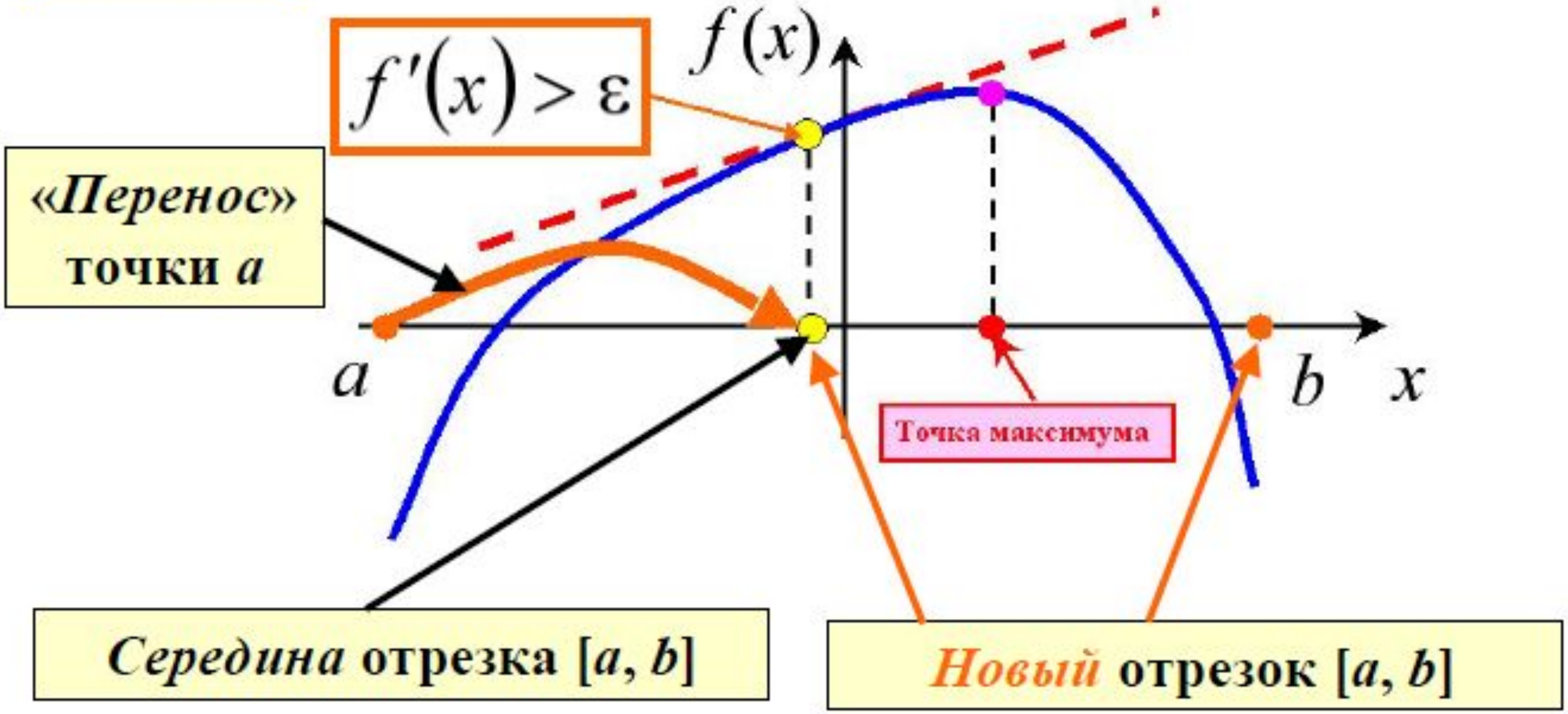
$$f(x) \to \max$$
$$x \in [a, b]$$



Метод **средней точки**

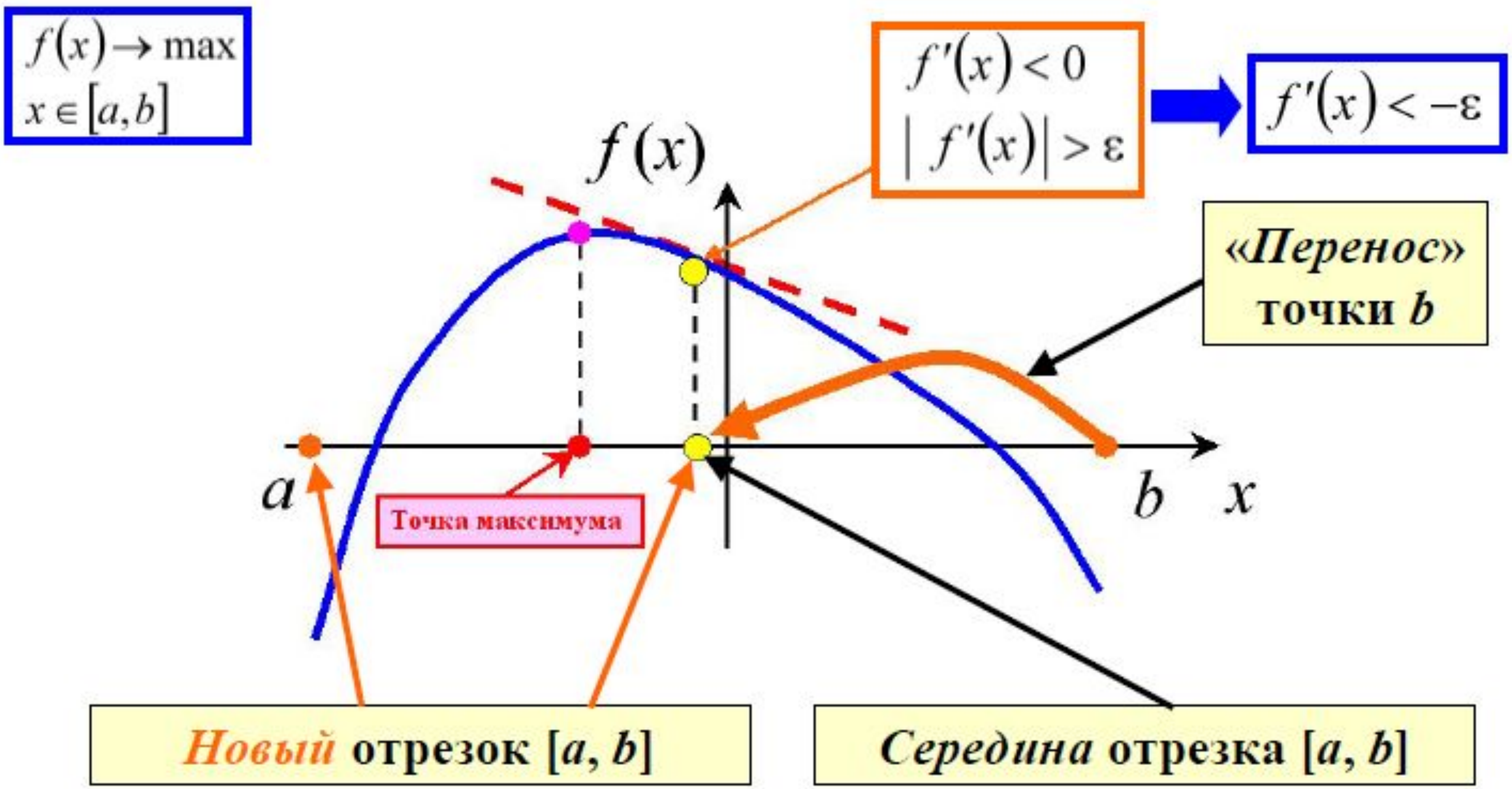
Если первая производная в середине отрезка больше заданной погрешности, то «переносим» точку *a* в середину отрезка.

$$f(x) \to \max$$
$$x \in [a, b]$$



Метод **средней точки**

Если первая производная в середине отрезка отрицательная и по модулю больше заданной погрешности, то «переносим» точку *b* в середину отрезка.



В методе хорд

надо решить

методом хорд

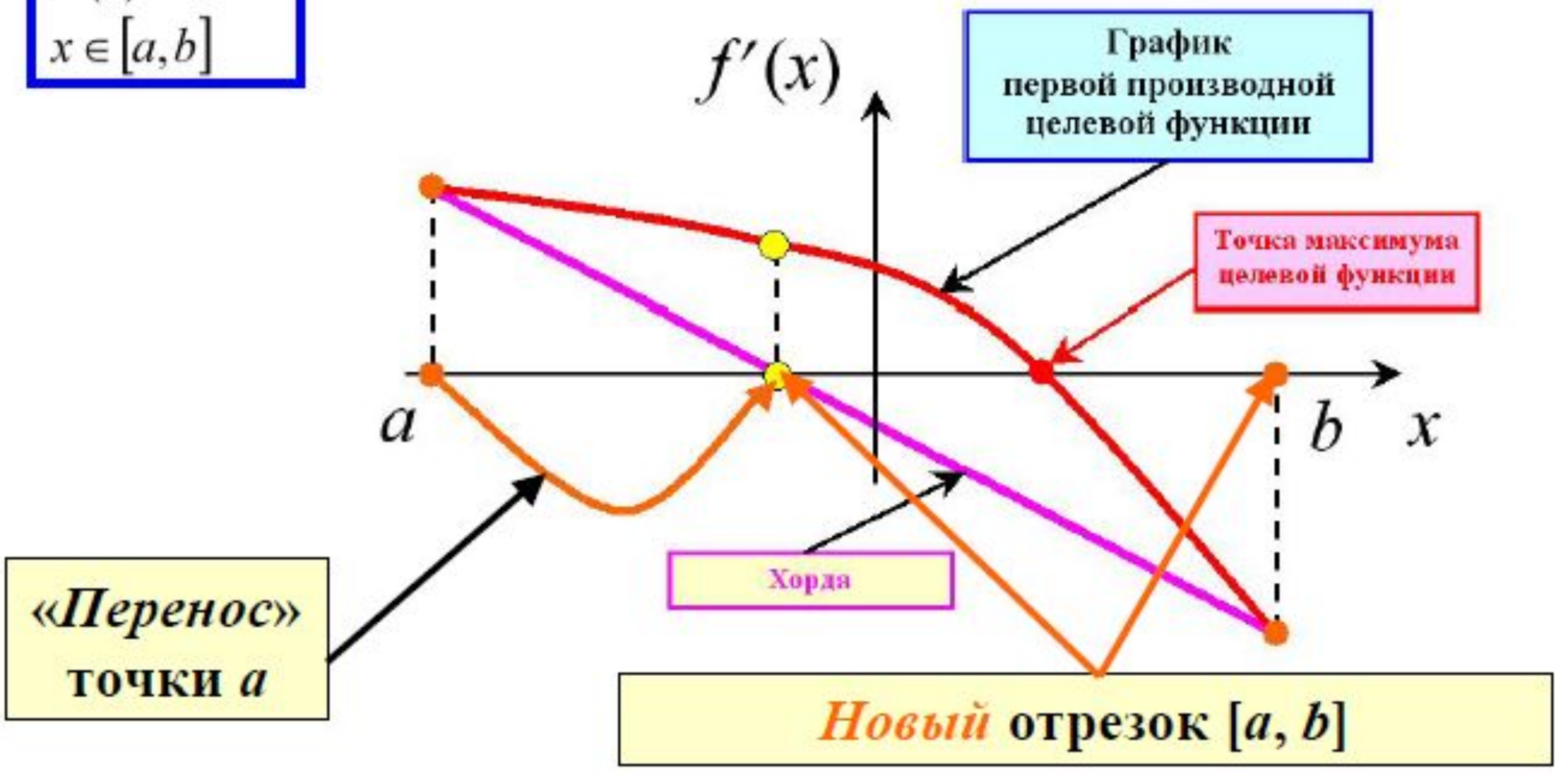
уравнение

$$\boxed{f'(x) = 0}$$

Метод хорд

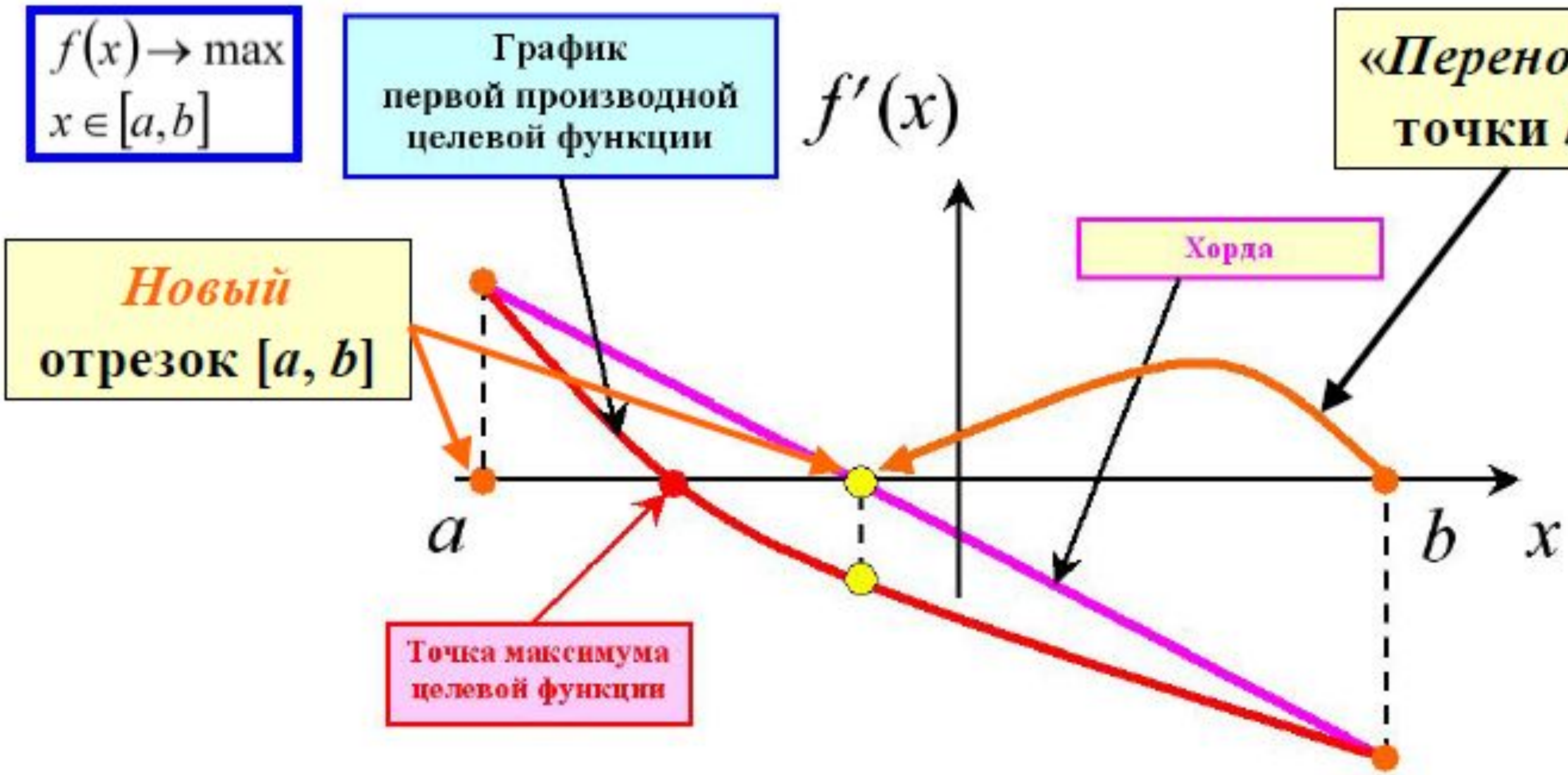
Знаки первой производной целевой функции на концах нового отрезка [*a*, *b*] должны быть различны.

$$f(x) \to \max$$
$$x \in [a, b]$$



Метод хорд

Знаки первой производной целевой функции на концах <u>нового</u> отрезка [a, b] должны быть различны.



Когда надо **закончить** поиск в методе хорд?

Поиск прекращается, когда

$$\left| x_n - x_{n-1} \right| \le \varepsilon$$

x_n, x_{n-1} — текущая и предыдущая точки пересечения хорд с осью x

ε — заданная погрешность

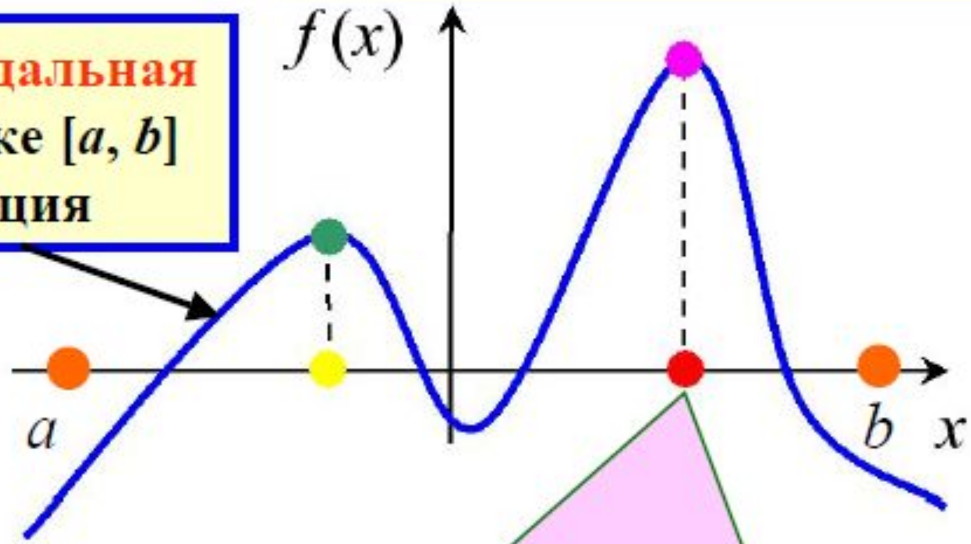
n – номер итерации

Одномерная нелинейная оптимизация *с ограничениями*

$$f(x) \to \max$$
$$x \in [a, b]$$

На отрезке [*a*, *b*]
различных **значений**
локальных максимумов
более одного

<u>Не</u>унимодальная
на отрезке [*a*, *b*]
функция



Точка глобального максимума

Гарантированно
найти
точку глобального экстремума
неунимодальной функции
можно,
если она на отрезке
удовлетворяет
условию Липшица

Условие Липшица

Функция $f(x)$

удовлетворяет на отрезке $[a, b]$

условию Липшица,

если ***существует***

такое число $L > 0$

(**константа Липшица**),

что

$$\boxed{|f(x_1) - f(x_2)| \leq L\,|x_1 - x_2|}$$

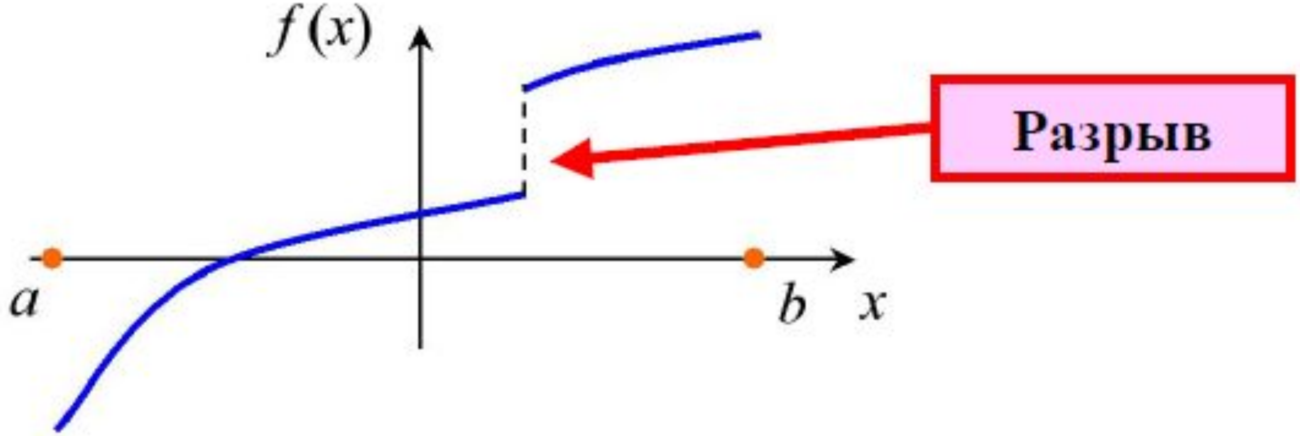
для всех x_1 и x_2, принадлежащих $[a, b]$.

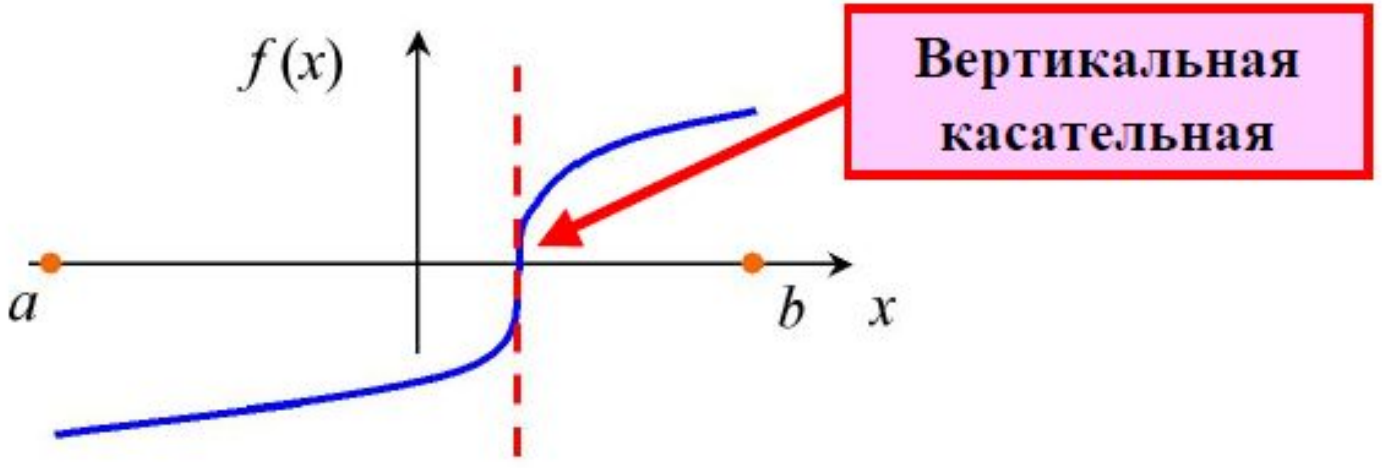
Если функция удовлетворяет на каком-то отрезке **условию Липшица**, то:

1) функция **непрерывна** на этом отрезке;

2) **нет вертикальных касательных** к графику функции на этом отрезке.

Примеры функций, не удовлетворяющих на отрезке $[a, b]$ условию Липшица

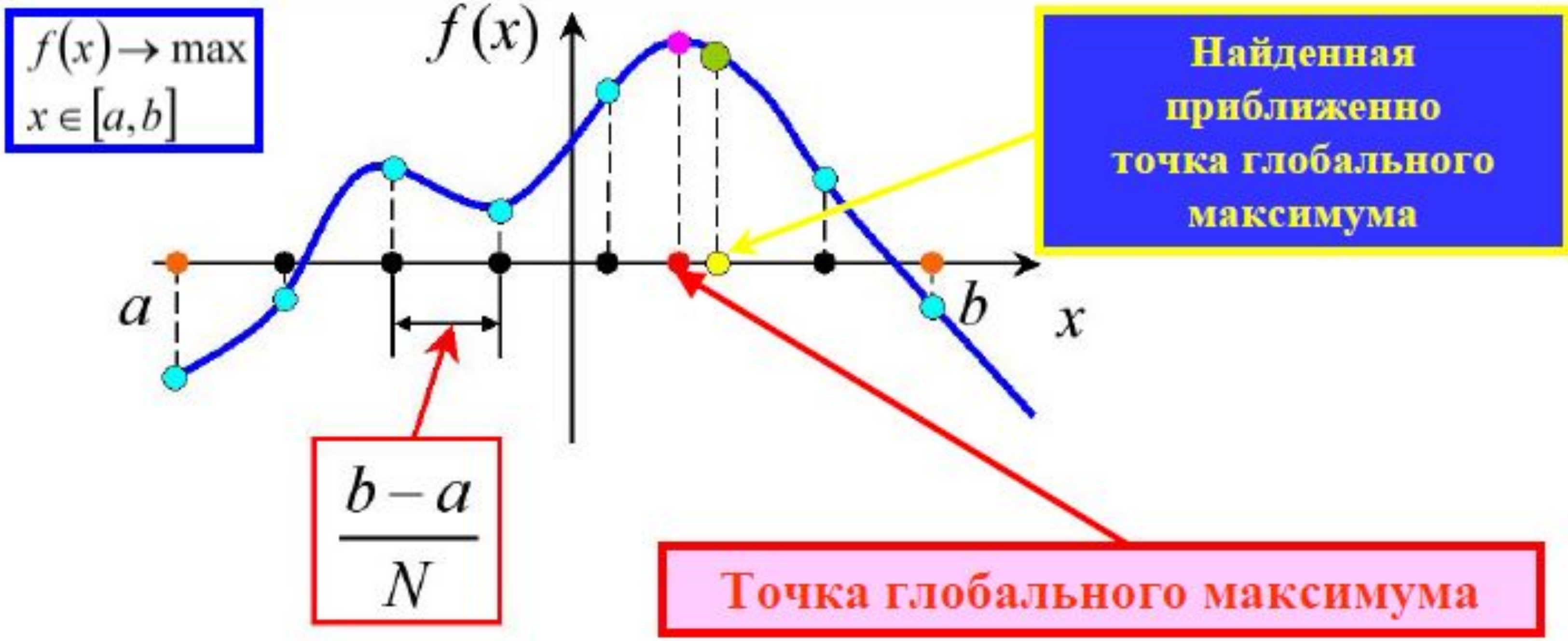




Метод **перебора** для **неунимодальной** функции, удовлетворяющей условию Липшица

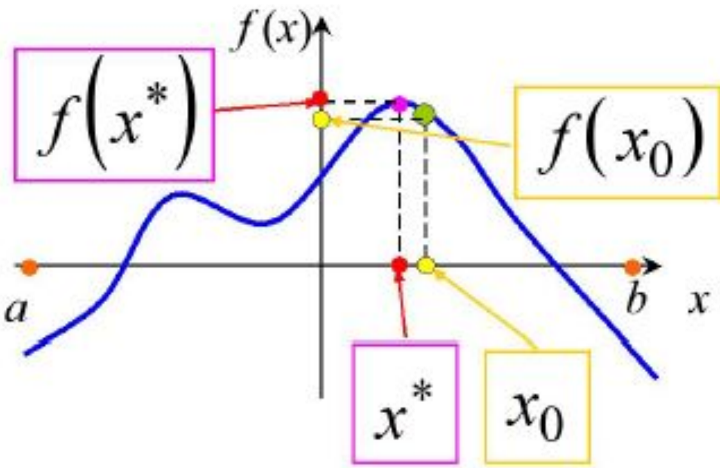
Метод перебора

Отрезок [a, b] делится на N равных частей.
В точках a, b и в точках деления
вычисляются значения целевой функции.
Среди них находится максимальное значение.
Ему приближенно соответствует
точка глобального максимума.



Погрешность метода перебора для непрерывно дифференцируемой на отрезке [*a*, *b*] функции, удовлетворяющей на нем условию Липшица

$$\boxed{f\left(x^*\right) - f(x_0) \le L\,\frac{b-a}{2N}}$$



$$L = \max_{x \in [a,\, b]} \left| f'(x) \right|$$

***Минимальная* константа Липшица**

Другие методы одномерной нелинейной оптимизации – в книгах:

Автор:
Гончаров В.А.
Название:
«Методы оптимизации».
Москва,
издательство «Юрайт»,
2010 год

Авторы:
Пантелеев А.В., Летова Т.А.
Название:
«Методы оптимизации в примерах и задачах».
Москва,
издательство «Высшая школа»,
2005 год