

Лекции ИУ7. Методы Вычислений. Семестр 2

Власов П. А.*

21 марта 2016 г.

Содержание

1	Одномерная оптимизация	2
1.1	Основные понятия одномерной оптимизации	2
1.1.1	Минимум функции	2
1.1.2	Унимодальные функции	2
1.1.3	Выпуклые функции	3
1.1.4	Липшицевы функции	4
1.2	Методы одномерной оптимизации	6
1.2.1	Классический метод	6
1.2.2	Методы перебора и поразрядного поиска	7
1.2.3	Методы исключения отрезков	9
1.2.4	Метод парабол	15
1.2.5	Метод бисекции и хорд	17
1.2.6	Метод Ньютона	21
1.2.7	Метод перебора	24
1.2.8	Метод ломаных	26

Основные понятия

Типовая задача оптимизации имеет следующий вид

$$\begin{cases} f(x) \rightarrow \min \\ x \in G \end{cases} \quad (1)$$

Замечание:

1. Если требуется задачу максимизации, то обычно вместо функции $f(x)$ рассматривают функцию $g(x) = -f(x)$ и решают задачу минимизации для G .
2. В прошлом семестре мы рассматривали задачу (1) для:
 - (а) случая, когда G конечно или счетно
 - (б) случая, когда f линейна, а G — выпуклый многоугольник в пространстве \mathbb{R}^n .
(В этом случае задачу (1) называют *задачей исследования операций*)
3. В этом семестре будем рассматривать задачу (1) для

*Законспектировано Абакумкиным А. В.

- (а) произвольной (не обязательно скалярной) функции f и
 (б) для произвольного множества $G \subseteq \mathbb{R}^n$.

Используется следующая терминология:

Функция f	Множество G	Название задачи
$f : G \rightarrow \mathbb{R}$	$[a; b] \subset \mathbb{R}$	Задача одномерной оптимизации
$f : G \rightarrow \mathbb{R}$	$G = \mathbb{R}^n, n \geq 2$	Задача многомерной безусловной оптимизации
$f : G \rightarrow \mathbb{R}$	$G \subset \mathbb{R}^n, n \geq 2$	Задача многомерной условной оптимизации
$f : G \rightarrow \mathbb{R}^m, m \geq 2$	$G \subseteq \mathbb{R}^n$	Задача многокритериальной оптимизации

1. Одномерная оптимизация

$$\begin{cases} f(x) \rightarrow \min \\ x \in [a; b] \end{cases} \quad (2)$$

1.1. Основные понятия одномерной оптимизации

1.1.1. Минимум функции

Пусть $f : G \rightarrow \mathbb{R}, G \subseteq \mathbb{R}$

Определение: Точка $x^* \in G$ называется *точкой глобального минимума* функции f на множестве $\forall x \in G \quad f(x^*) \leq f(x)$.

При этом число f^* называется *минимумом* (глобальным) функции f на G и обозначается $f^* = \min_{x \in G} f(x)$.

Замечание: Обозначим *множество всех точек глобальных минимумов* f на G , как

$$G^* = \left\{ x^* \in G : f(x^*) = \min_{x \in G} f(x) \right\}$$

Определение: Точка $\tilde{x} \in G$ называется *точкой локального минимума* функции на множестве G , если

$$\exists \varepsilon > 0 \quad \forall x \in u_\varepsilon(\tilde{x}) \cap G \quad f(\tilde{x}) \leq f(x),$$

где $u_\varepsilon(\tilde{x}) = \{x : |\tilde{x} - x| < \varepsilon\}$.

Замечание:

1. Точка глобального минимума является точкой локального минимума. Обратное неверное.
2. Задача (2) имеет решение тогда и только тогда, когда $G^* \neq \emptyset$
3. Согласно теореме Вейерштрасса, всякая функция, непрерывная на замкнутом ограниченном множестве, достигает на этом множестве своих \inf и \sup (которые являются в этом случае минимум и максимумом этой функции на этом множестве).
 Таким образом задача (2) всегда имеет решение в случае непрерывной функции f .

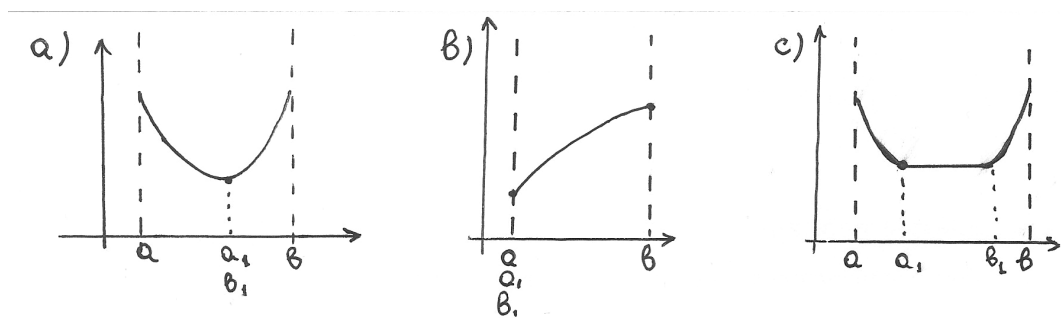
1.1.2. Унимодальные функции

Пусть $f : [a; b] \rightarrow \mathbb{R}$

Определение: f называется *унимодальной* на отрезке $[a; b]$, если $\exists a_1, b_1 \in \mathbb{R}$:

1. $a \leq a_1 \leq b_1 \leq b$
2. Если $a < a_1$, то f монотонно убывает на $[a; a_1]$
3. Если $b_1 < b$, то f монотонно возрастает на $[b_1; b]$.

$$4. \forall \tilde{x} \in [a_1; b_1] \quad f(\tilde{x}) = \min_{x \in G} f(x)$$



Свойства унимодальных функций

1° Каждая точка локального минимума унимодальной функции является одновременно точкой её глобального минимума.

2° Если f унимодально на $[a; b]$, то f унимодально и на любом отрезке $[a_1, b_1] \subset [a; b]$.

3° Пусть:

1. f унимодальна на отрезке $[a; b]$
2. $a \leq x_1 < x_2 \leq b$
3. x^* — точка минимума функции f .

Тогда

1. Если $f(x_1) \leq f(x_2)$, то $x^* \in [a; x_2]$
2. Если $f(x_1) > f(x_2)$, то $x^* \in [x_1; b]$

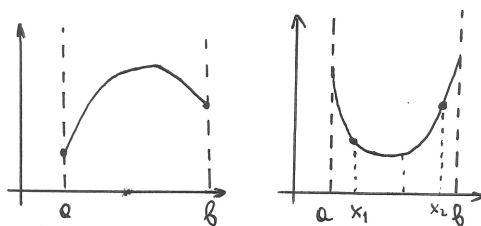
1.1.3. Выпуклые функции

Пусть $f : [a; b] \rightarrow \mathbb{R}$

Определение: Функция f называется *выпуклой*, если

$$\forall x_1, x_2 \in [a; b] \quad \forall \alpha \in [0; 1]$$

$$f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2) \quad (3)$$



Замечание:

1. Неравенство (3) означает, что для любой хорды графика функции $f(x)$, которая соединяет точки $(x_1, f(x_1))$ и $(x_2, f(x_2))$, график функции $f(x)$ на отрезке, соединяющий x_1 и x_2 , лежит не выше этой хорды.

2. В классическом математическом анализе такие функции называются выпуклыми вниз. Функции, которые в классическом математическом анализе являются выпуклыми вверх, мы не будем считать выпуклыми (так как они не удовлетворяют нашему определению). Эта «дискриминация» связана с тем, что в дальнейшем будем рассматривать только задачу минимизации.

Свойства выпуклых функций

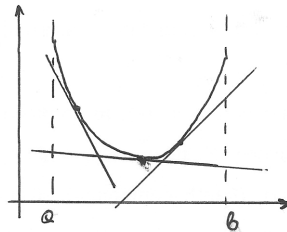
Через $C^{(k)}[a; b]$ будем обозначать функции, которые непрерывны на отрезке $[a; b]$ и имеют на $[a; b]$ непрерывные производные до порядка k включительно.

1° Пусть $f \in C^{(1)}[a; b]$

Тогда f выпукла тогда и только тогда, когда $f'(x)$ не убывает на $[a; b]$

2° Пусть $f \in C^{(2)}[a; b]$, тогда f выпукла на $[a; b] \Leftrightarrow f''(x) \geq 0, \quad x \in [a; b]$

3° Пусть $f \in C^{(3)}[a; b]$, тогда f выпукла $\Leftrightarrow \forall x_0 \in [a; b]$ касательная к графику функции $f(x)$ в точке x_0 лежит не выше графика $f(x)$.



4° Пусть

1. $f \in C^{(1)}[a; b]$
2. f выпукла на $[a; b]$
3. $f'(x^*) = 0, \quad x^* \in [a; b]$

Тогда x^* — точка глобального минимума $f(x), x \in [a; b]$.

5° $C[a; b] = C^0[a; b]$

Пусть

1. $f \in C[a; b]$
2. f выпукло на $[a; b]$

Тогда f унимодальна на $[a; b]$

Замечание:

1. Многие методы минимизации разработанны для унимодальных функций. При этом эти методы хорошо сходятся, если f выпукла.
2. На практике проверку выпуклости целевой функции осуществляют не с помощью использования определения, а с использованием свойств 1-3 или физических соображений.

1.1.4. Липшицевы функции

Пусть $f: [a; b] \rightarrow \mathbb{R}$

Определение: Говорят, что f удовлетворяет на отрезке $[a; b]$ условию Липшица (является липшицевой), если

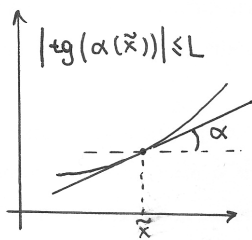
$$\exists L \geq 0 \quad \forall x_1, x_2 \in [a; b]$$

$$|f(x_1) - f(x_2)| \leq L \cdot |x_1 - x_2|$$

При этом L называется константой Липшица для f на $[a; b]$.

Замечание: Для дифференцируемой на $[a; b]$ функции условие Липшица означает, что для любой точки $\tilde{x} \in [a; b]$ угловой коэффициент касательной к графику $f(x)$ в этой точке по абсолютной величине не превосходит L .

$$\forall \tilde{x} \quad |\operatorname{tg} \alpha(\tilde{x})| \leq L$$



Свойства липшицевых функции

1° Если f удовлетворяет условию Липшица с константой L , то f удовлетворяет условию и с любой константой $L_1 > L$.

2° Если f липшицева на $[a; b]$, то f является липшицевой и на любом отрезке $[a_1, b_1] \subseteq [a, b]$.

3° Если $f \in C^{(1)}[a; b]$, то

1. f липшицева на $[a; b]$
2. константа Липшица для f на $[a; b]$ может быть выбрана

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

4° Пусть

1. $x_0 < x_1 < \dots < x_n$
2. f является липшицевой на $[x_{i-1}, x_i]$ с константой $L_i, i = \overline{1; n}$.

Тогда f является липшицевой на $[x_0; x_n]$ с константой

$$L = \max_{i=\overline{1; n}} L_i.$$

5° Если f липшицева на $[a; b]$, то она непрерывна на $[a; b]$.

Пример:

1. $f(x) = \sin x$ является липшицевой на любом отрезке $[a; b]$, так как она непрерывно дифференцируема на $[a; b]$
2. $f(x) = \sqrt{x}$ не является липшицевой на $[0; a], a > 0$. Если бы f была липшицевой, то угловые коэффициенты касательных к графику были бы ограничены некоторой константой. Для \sqrt{x} на $[0; a]$ это не так.

1.2. Методы одномерной оптимизации

1.2.1. Классический метод

$$\begin{cases} f(x) \rightarrow \min \\ x \in [a; b] \end{cases}$$

Из курса математического анализа известно:

1. Если

- (a) $f(x)$ дифференцируемая в точке x^* ,
- (b) $f(x)$ имеет локальный экстремум в точке x^* ,

то $f'(x) = 0$

2. Если

- (a) $f(x)$ дифференцируемая в окрестности x^* ,
- (b) $f'(x^*) = 0$,

то

- (a) Если $f(x)$ при переходе через x^* меняет знак с «−» на «+», то x^* — точка локального минимума
- (b) Если $f(x)$ при переходе через x^* меняет знак с «+» на «−», то x^* — точка локального

3. Если

- (a) $f(x)$ n раз дифференцируемая в точке x^* ,
- (b) $f'(x) = f^{(n-1)}(x^*) = 0$,
- (c) $f^{(n)}(x^*) \neq 0$,

то

- (a) если n нечетно, то $f(x)$ не имеет локального экстремума в точке x^* ,
- (b) если n четно, а $f^{(n)}(x^*) > 0$, то x^* — точка локального минимума,
- (c) если n четно и $f^{(n)}(x^*) < 0$, то x^* — точка локального максимума.

Классический метод

1. Вычисляем $f'(x)$, $x \in (a; b)$, решаем уравнение

$$f'(x) = 0 \tag{4}$$

Пусть x_1, \dots, x_n — его решения

2. Для каждой точки x_k , $k = \overline{1, n}$ проверяем условие 2 или 3 и отбираем точки $\tilde{x}_1, \dots, \tilde{x}_p$, которые отвечают условию локального минимума.

3. Полагаем

$$f^* = \min \{f(\tilde{x}_1), \dots, f(\tilde{x}_p), f(a), f(b)\}$$

Замечание: На практике для применения этого метода затруднительно по следующим причинам

- 1. Для практически интересных(?) функций аналитическое решение (4) часто затруднительно
- 2. Функция может быть известна из наблюдений, что ведет к тому, что невозможно получить аналитическое представление для $f'(x)$
- 3. Проверка достижимости условий затруднительна

Эти трудности привели к появлению численных методов.

Их делят

1. Прямые методы

- методы перебора и поразрядного поиска
- методы исключения отрезков
- метод парабол

2. Методы использующие производные целевой функции

- метод бисекций
- метод хорд и Ньютона

1 и 2 используются для унимодальных функций

3. Для минимизации многомодальных функций:

- метод перебора
- метод ломаных

Замечание: *Прямыми* называются методы, которые используют только значения целевой функции и не используют значения её производных.

1.2.2. Методы перебора и поразрядного поиска

Всегда предполагаем, что функция является унимодальной

I метод перебора

1. Разобьем $[a, b]$ системой точек $x_i = a + i\Delta$, $i = \overline{0, n}$, где $\Delta = \frac{b-a}{n}$
2. Затем вычислим $f(x_i)$, где $i = \overline{0, n}$
3. Выбираем точки x_m , $m \in \{0, \dots, n\}$ так, чтобы $f(x_m) = \min_{i=\overline{0, n}} f(x_i)$. Положим $x^* = x_m$, $f^* = f(x_m)$

Замечание:

1. Погрешность нахождения x^* с использованием этого метода

$$\varepsilon_n \leq \frac{b-a}{n}$$

2. Если принять $n \gg 1$, то $\frac{1}{n} \approx \frac{1}{n+1}$ поэтому точность $\varepsilon(N)$, которую обеспечивает этот метод для N кратного вычисления(?) целевой функции

$$\varepsilon(N) \approx \frac{b-a}{N}$$

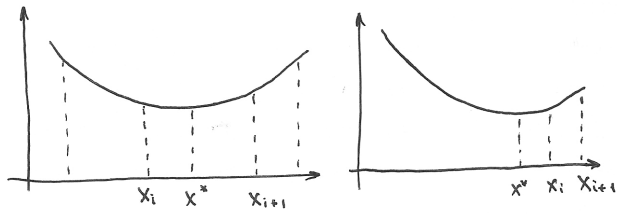
II метод поразрядного поиска

Этот метод является усовершенствованием метода перебора с целью уменьшения количества значений целевой функции f , которое необходимо найти для достижения заданной точности.

Замечание:

1. Если в методе перебора $f(x_{i+1}) \geq f(x_i)$, то $x^* \in [a, x_{i+1}]$ и следовательно $f(x_{i+2}), f(x_{i+3}), \dots$ можно не вычислять.

Пример:



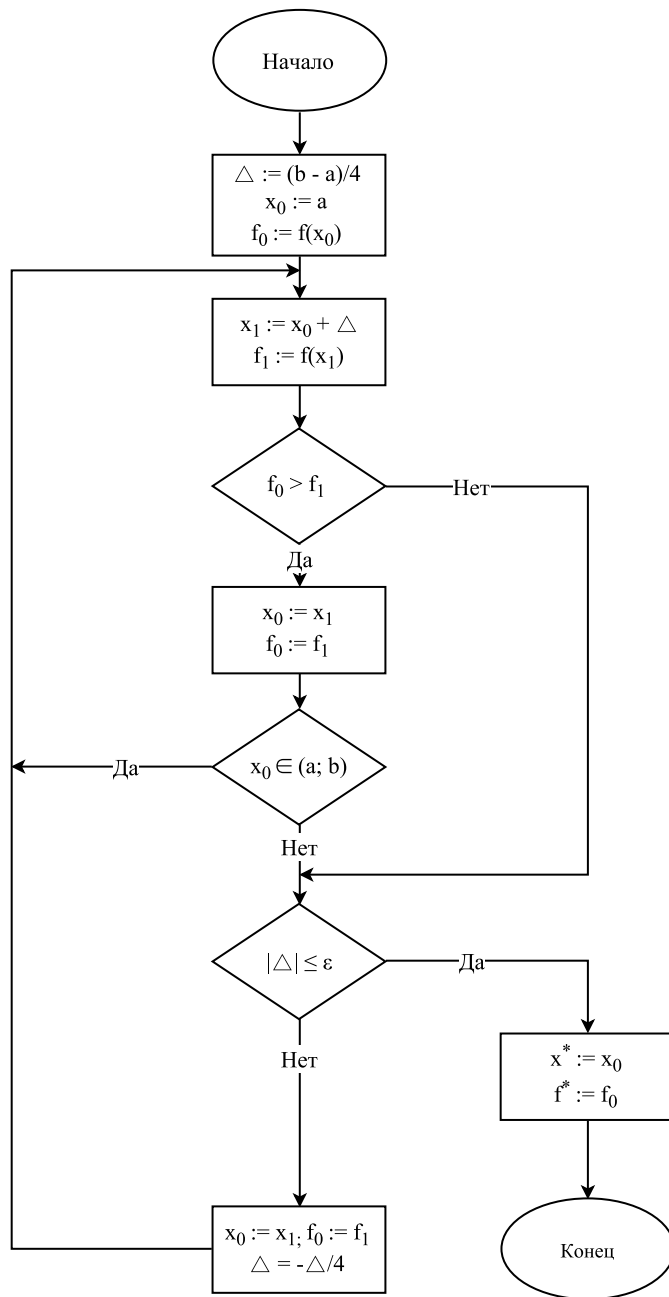
2. Целесообразно сперва найти приближенное (грубо) значение x^* , а затем уточнить это значение, используя более точный шаг.

Пусть ε — требуемая точность нахождения x^* (глобальный минимум). При реализации, обычно, сперва фиксируют $\Delta > \varepsilon$, вычисляют $f_i = f(x_i)$, $x_i = a + i\Delta$, до тех пор, пока не будет выполнено условие $f_{i+1} \geq f_i$.

При выполнении этого условия шаг Δ уменьшается (как правило в четыре раза, а процесс поиска запускается в обратную сторону).

Пусть ε — искомая точность.

Метод поразрядного поиска



1.2.3. Методы исключения отрезков

Один из подходов к построению основан на использовании следующих свойств.

Если $x_1 < x_2$, то

1. Если $f(x_1) \leq f(x_2)$, то $x^* \in [a, x_2]$

2. Если $f(x_1) > f(x_2)$, то $x^* \in [x_1, b]$

При построении соответствующих методов выбираем две произвольные точки x_1, x_2 :

$$a < x_1 < x_2 < b$$

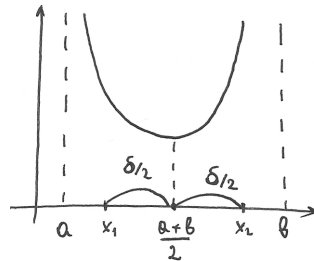
Далее проверяем условия 1-2 и по результатам этой проверки отбрасываем часть отрезка $[a, b]$.

Вычисления продолжаются до тех пор, пока длина текущего отрезка не станет меньше ε .

Пробные точки x_1, x_2 выбирают обычно симметричными от середины отрезка. Это делается для того, чтобы отношение длины нового отрезка к длине предыдущего не зависело от того, какая часть (правая или левая) отбрасывается.

Способ выбора пробных точек x_1 и x_2 определяет конкретный метод поиска минимума.

I Метод дихотомии



Выбираем достаточно малое $\delta > 0$ и положим $x_1 = \frac{a+b}{2} - \frac{\delta}{2}$, $x_2 = \frac{a+b}{2} + \frac{\delta}{2}$.

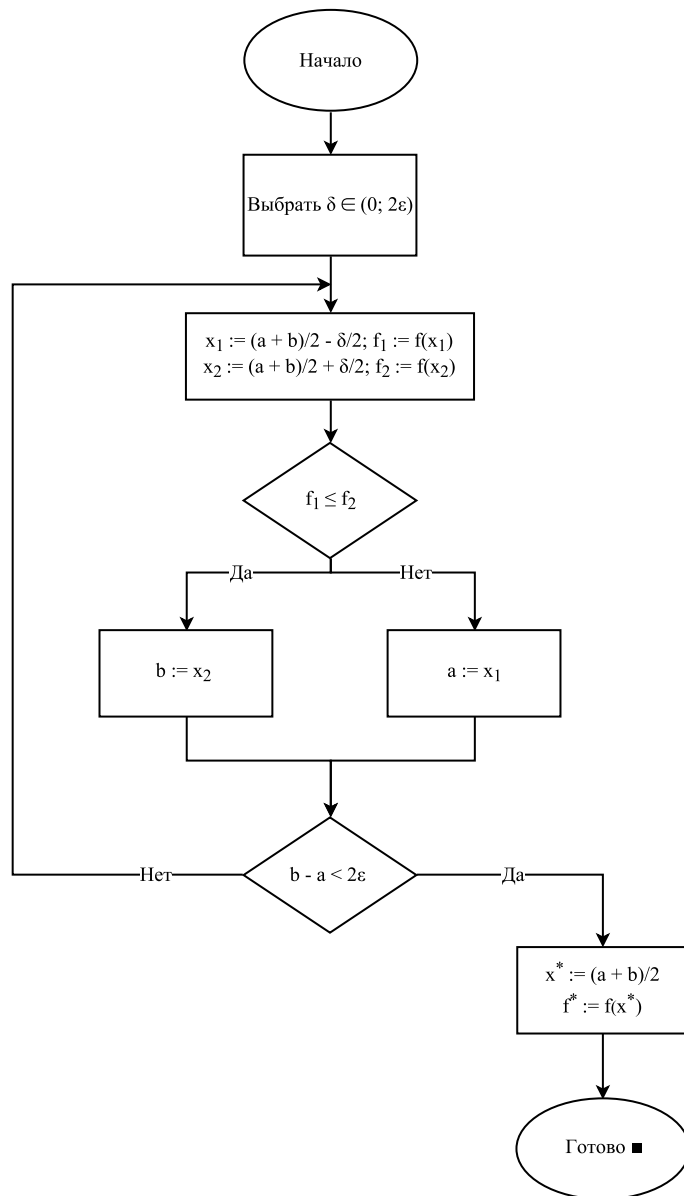
В этом случае отношение длины нового отрезка к длине предыдущего:

$$\tau = \frac{b - x_1}{b - a} = \frac{x_2 - a}{b - a} \approx \frac{1}{2}$$

Вычисления прекращаются, когда для очередного отрезка его длина

$$b - a < 2\varepsilon \quad (5)$$

Использование ослабленного неравенства (5) связано с тем, что в алгоритме принимается $x^* = \frac{a+b}{2}$



Замечание:

1. О выборе δ :

- (а) Чем меньше δ , тем метод лучше сходится
- (б) При слишком малых значениях δ значения $f(x_1) \approx f(x_2)$, если эти значения содержат ошибки измерений или вычислений, то возможно выполнение «не того» неравенства.

2. Число n итераций метода дихотомии необходимое для достижения заданной точности ε , определяется условием

$$n \geq \log_2 \frac{b-a-\delta}{2\varepsilon-\delta}$$

Доказательство

Пусть $\Delta_0 = b - a$ — длина искомого отрезка. Тогда длина искомого отрезка после первой итерации метода:

$$\Delta_1 = \frac{\Delta_0}{2} + \frac{\delta}{2}$$

Длина отрезка после второй итерации:

$$\Delta_2 = \frac{\Delta_1}{2} + \frac{\delta}{2} = \frac{\Delta_0}{4} + \delta \left(\frac{1}{2} + \frac{1}{4} \right)$$

Длина отрезка после третьей итерации:

$$\Delta_3 = \frac{\Delta_2}{2} + \frac{\delta}{2} = \frac{\Delta_0}{8} + \delta \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} \right)$$

...

После n итераций:

$$\Delta_n = \frac{\Delta_0}{2^n} + \delta \left(\frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^n} \right) = \frac{\Delta_0}{2^n} + \delta \left(1 - \frac{1}{2^n} \right)$$

Условие окончания: $\Delta_n \leq 2\varepsilon$

Тогда

$$\frac{b-a-\delta}{2^n} + \delta \leq 2\varepsilon$$

$$\frac{b-a-\delta}{2^n} \leq 2\varepsilon - \delta$$

$$2^n \geq \frac{b-a-\delta}{2\varepsilon-\delta}$$

$$n \geq \log_2 \frac{b-a-\delta}{2\varepsilon-\delta}$$

3. Так как δ обычно выбирают достаточно малым, то точность ε_n , которая обеспечивается после выполнения n итераций алгоритма,

$$n \approx \log_2 \frac{b-a-\delta}{2\varepsilon_n} \Rightarrow \varepsilon_n \approx \frac{b-a}{2^{n+1}}$$

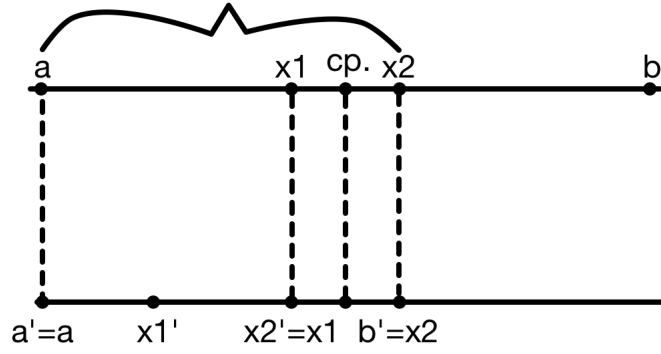
Поскольку для выполнения n итераций алгоритма требуется $N = 2n$ вычислений значений целевой функции f , то точность $\varepsilon(N)$, которая будет гарантированно после N вычислений значений функции

$$\varepsilon(N) = \varepsilon_{\frac{N}{2}} = \frac{b-a}{2^{N/2+1}}$$

II Метод золотого сечения

Для уменьшения количества значений целевой функции, которые приходится вычислять в ходе реализации алгоритма постараемся выбрать пробные точки x_1 и x_2 внутри отрезка $[a; b]$ так, чтобы при переходе к очередному отрезку одна из этих точек стала новой пробной точкой.

При этом будем считать, что отношение длины нового отрезка к длине текущего отрезка не зависит от номера итерации и равно τ . Так же будем считать, что x_1 и x_2 располагаются симметрично относительно середины отрезка $[a; b]$.



$$\tau = \frac{b' - a'}{b - a}$$

1.

$$x_2 = a + \tau(b - a)$$

$$x_1 = b - \tau(b - a)$$

2. Отношение длины отрезка $[a', x_2']$ к длине отрезка $[a', b']$ должны быть равны τ :

$$\text{Дл}([a', x_2']) = \text{Дл}([a, x_1]) = x_1 - a = b - a - \tau(b - a)$$

$$\text{Дл}([a', b']) = \text{Дл}([a, x_2]) = x_2 - a = \tau(b - a)$$

Таким образом

$$\frac{b - a - \tau(b - a)}{\tau(b - a)} = \tau \Rightarrow \frac{1 - \tau}{\tau} = \tau \Rightarrow \tau^2 + \tau - 1 = 0$$

$$\tau = \frac{-1 + \sqrt{1 + 4}}{2} = \frac{\sqrt{5} + 1}{2} \approx 0.6183$$

отрицательное решение не рассматриваем.

Таким образом

$$\tau = 0.6183$$

$$x_1 = b - \tau(b - a)$$

$$x_2 = a + \tau(b - a)$$

Замечание:

1. Каждая из точек x_1 и x_2 , используемых в рассматриваемом методе, делит отрезок $[a; b]$ на неравные части так, что

$$\frac{\text{длина отрезка } [a; b]}{\text{длина большей части отрезка } [a; b]} = \frac{\text{длина большей части отрезка } [a; b]}{\text{длина меньшей части отрезка } [a; b]}$$

Про такие точки говорят, что они реализуют золотое сечение отрезка $[a; b]$.

2. На каждой итерации длина отрезка уменьшается в $\tau = \frac{\sqrt{5}-1}{2}$ раз. Поэтому после n итераций длина соответствующего отрезка равна

$$\frac{1}{2}\tau^n(b - a)$$

так как в конце берем $x^* = \frac{a+b}{2}$

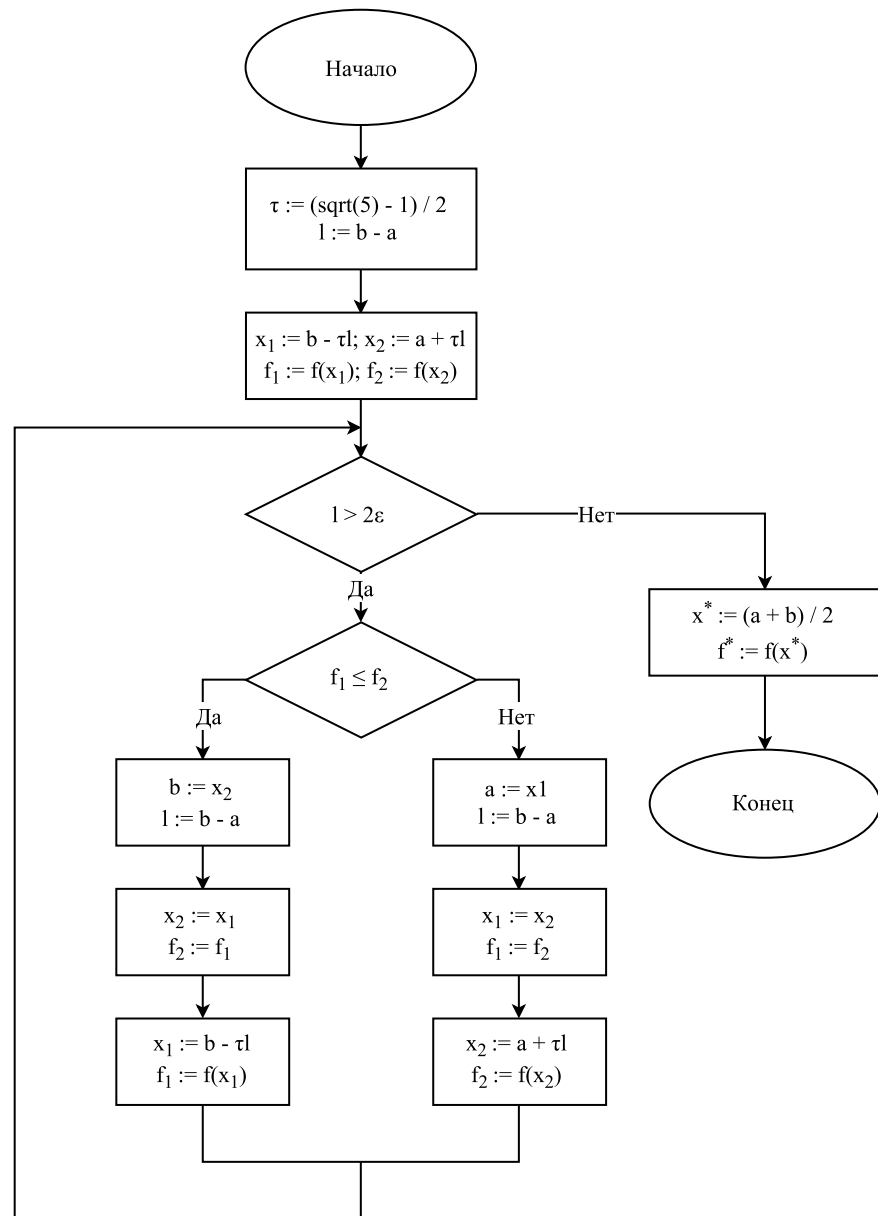
3. Число n итераций, необходимых для достижения заданной точности ε , составляет

$$\frac{1}{2}\tau^n(b - a) \leq \varepsilon \Rightarrow \tau^n \leq \frac{2\varepsilon}{b - a} \Rightarrow$$

$$n \geq \log_{\tau} \frac{2\varepsilon}{b - a} = \frac{\ln \frac{2\varepsilon}{b - a}}{\ln \tau} \approx -2.1 \cdot \ln \frac{2\varepsilon}{b - a} = 2.1 \ln \frac{b - a}{2\varepsilon}$$

4. Для выполнения первой итерации необходимо вычисление двух значений целевой функции f . Для выполнения второй, третьей, ... итераций необходимо вычисление одного значения функции. Поэтому для выполнения n итераций необходимо вычислить $N + 1$ значений функции. Поэтому

$$\varepsilon(N) = \varepsilon_n \Big|_{n=N-1} = \frac{1}{2} \tau^{N-1} (b-a) \approx \tau^{N-2} (b-a)$$



1.2.4. Метод парабол

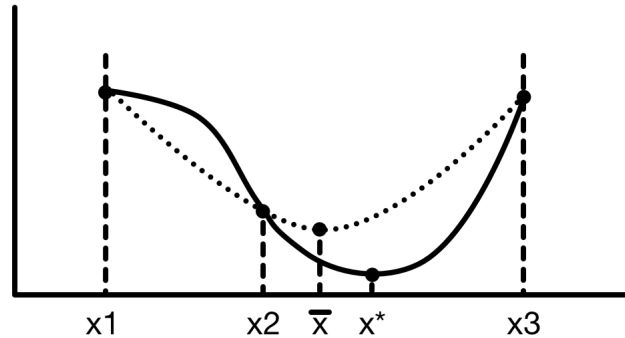
Метод парабол является представителем группы методов, основанных на аппроксимации целевой функции некоторой более простой функцией (как правило полиномом), минимум которой можно легко найти. Точка минимума этой аппроксимирующей функции и принимается за очередное приближение точки минимума целевой функции.

Пусть

1. f унимодальна на $[a; b]$
2. f достигает минимума во внутренней точке отрезка $[a; b]$

Выберем три точки $x_1, x_2, x_3 \in [a; b]$, так чтобы (*):

1. $x_1 < x_2 < x_3$
2. $f(x_1) \geq f(x_2) \leq f(x_3)$ принимает по крайней мере одно неравенство строгое



Тогда в силу унимодальности функции f точка минимума $x^* \in [x_1, x_3]$.

Аппроксимируем целевую функцию параболой, проходящей через точки (x_1, f_1) , (x_2, f_2) , (x_3, f_3) , где $f_i = f(x_i)$, $i = \overline{1; 3}$.

В силу условий (*) ветви параболы направлены вверх. Это значит, что точка \bar{x} минимума этой параболы также принадлежит отрезку $[x_1, x_3]$.

Точка \bar{x} принимается за очередное приближение точки x^* .

Пусть $q(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2)$ — уравнение параболы.

Можно показать, что условия $q(x_i) = f_i$, $i = \overline{1; 3}$, приводят к (**):

$$\begin{aligned} a_0 &= f_1 \\ a_1 &= \frac{f_2 - f_1}{x_2 - x_1} \\ a_2 &= \frac{1}{x_3 - x_2} \left[\frac{f_3 - f_1}{x_3 - x_1} - \frac{f_2 - f_1}{x_2 - x_1} \right] \\ \bar{x} &= \frac{1}{2} \left[x_1 + x_2 - \frac{a_1}{a_2} \right] \end{aligned}$$

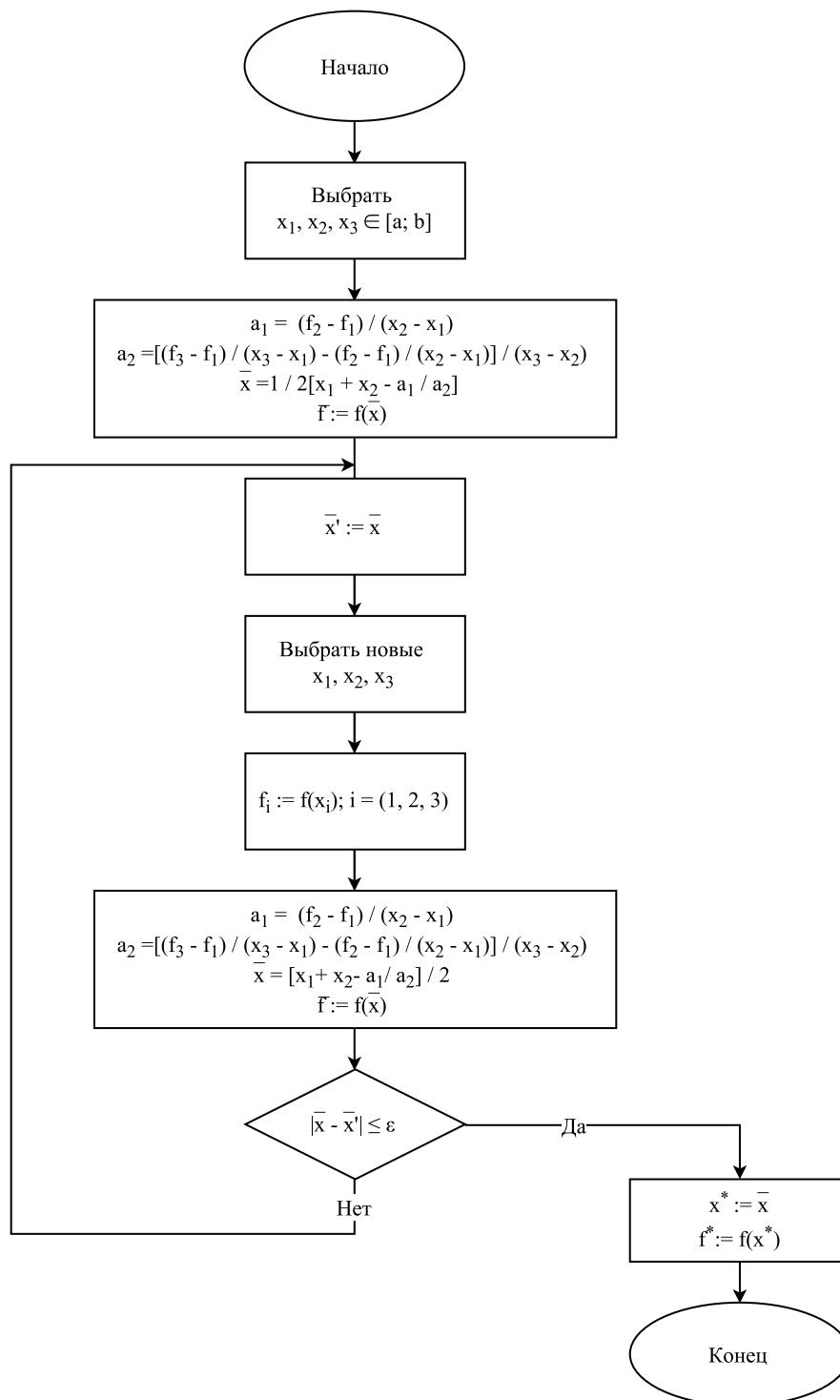
Метод парабол

Замечание:

1. В качестве критерия окончания вычислений используется условие $|\bar{x} - \bar{x}'| < \varepsilon$, означающее близость друг к другу двух последовательных приближений точки x^* . Вообще говоря, выполнение этого условия не гарантирует близость этих точек к x^* . Однако на практике такое условие удовлетворительно работает. Дополнительно точность текущего приближения можно оценивать (если получится) с использованием длины отрезка $[x_1, x_3]$.

2. О выборе точек x_1, x_2, x_3

- (a) На первой итерации для выбора точек x_1, x_2, x_3 обычно достаточно использование нескольких пробных точек. Если это не получается за разумное время, можно выполнить несколько итераций метода золотого сечения до тех пор, пока пробные точки этого метода и одна из граничных точек текущего отрезка не будут удовлетворять условиям (*).
 - (b) На второй и последующих итерациях на отрезке $[x_1, x_3]$ рассматриваются две пробные точки x_2 и \bar{x} , для которых используется метод исключения отрезков. В новом отрезке $[x'_1, x'_3]$ в качестве x'_2 выбирается та точка из x_2 и \bar{x} , которая оказалась внутри.
3. На каждой итерации метода парабол, кроме первой, вычисляется только одно значение целевой функции: \bar{f} .



1.2.5. Метод бисекции и хорд

Согласно сформулированным в п. 1 свойствам для дифференцируемой выпуклой (а значит и уни-модальной) функции f условие:

$$f'(x) = 0$$

является не только необходимым, но и достаточным условием точки минимума.

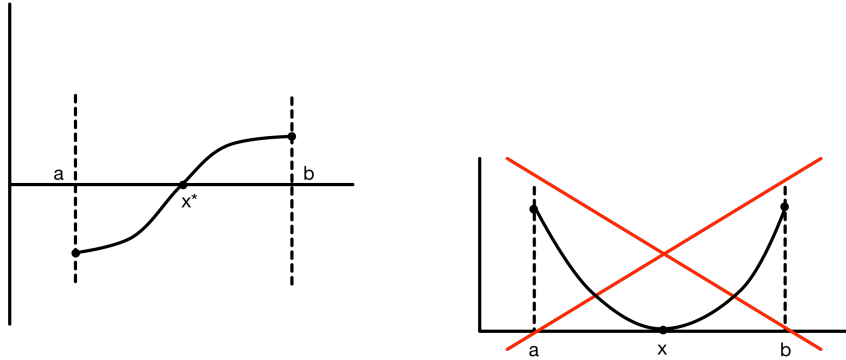
Метод бисекции поиска минимума функции $f(x)$

Является методом решения уравнения $f'(x) = 0$.

Замечание: Метод бисекции решения уравнения $g(x) = 0$

Пусть

1. $g(x)$ имеет единственный корень на $[a; b]$
2. $g(a)g(b) < 0$



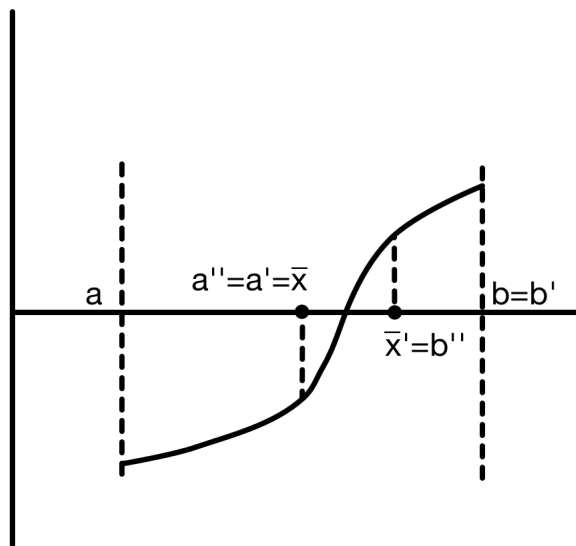
В качестве очередного приближения корня x^* в методе бисекции принимают значение

$$\bar{x} = \frac{a+b}{2}$$

Далее:

если $g(\bar{x})g(a) < 0 \Rightarrow b := \bar{x}$;

если $g(\bar{x})g(a) > 0 \Rightarrow a := \bar{x}$.



Вычисления останавливают, когда

$$|b - a| < 2\varepsilon$$

и полагают

$$x^* = \frac{x + b}{2}$$

Конец замечания

Метод бисекции решения задачи

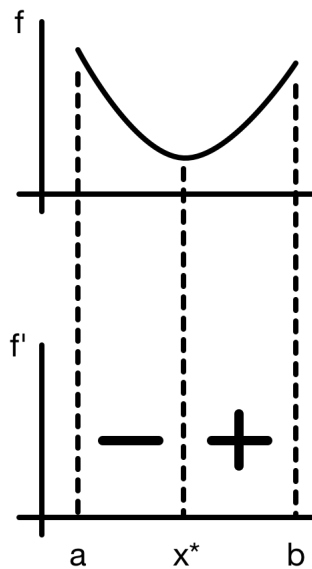
$$\begin{cases} f(x) \rightarrow \min \\ x \in [a; b] \end{cases}$$

< !!! БЛОК СХЕМА >

Замечание:

1. Так как мы используем метод бисекции для минимизации именно выпуклой функции, то

$$\begin{cases} f'(a) < 0 \\ f'(b) > 0 \end{cases}$$



2. На каждой итерации отрезок уменьшается вдвое, следовательно после n итераций будет достигнута точность

$$\varepsilon_n = \frac{b - a}{2^{n+1}}$$

Для достижения заданной точности ε необходимо сделать определенное число шагов, которое можно посчитать заранее:

$$n \geq \log_2 \frac{(b - a)}{\varepsilon} - 1$$

Метод хорд

Метод хорд решения задачи минимизации

$$\begin{cases} f(x) \rightarrow \min \\ x \in [a; b] \end{cases}$$

является методом хорд решения уравнения $f'(x) = 0$

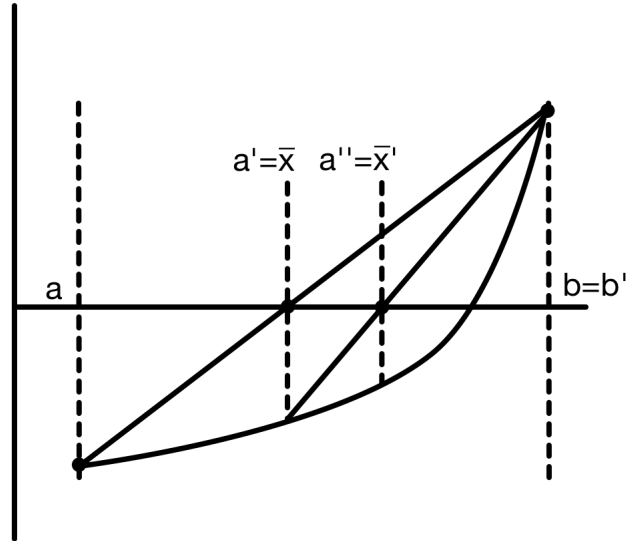
Замечание: Метод хорд решения уравнения $g(x) = 0$.

В методе хорд предполагается, что

1. $g(x)$ имеет единственный корень на $[a; b]$,

$$2. \quad g(a)g(b) < 0.$$

В качестве очередного приближения \bar{x} корня x^* используется точка пересечения с осью Ox хорды, соединяющей точки $(a, g(a))$, $(b, g(b))$.



В качестве условия окончания вычислений используется

либо $|\bar{x} - \bar{x}'| \leq \varepsilon$, где \bar{x}' — приближение x^* с предыдущей итерации

либо $|g(\bar{x})| < \varepsilon$

Получим расчетное соотношение метода хорд.

$$\begin{aligned} &(a, g(a)) \\ &(b, g(b)) \end{aligned}$$

Уравнение хорды:

$$\frac{x - a}{b - a} = \frac{y - g(a)}{g(b) - g(a)}$$

Пересечение с Ox :

$$\begin{aligned} &y = 0 \\ x &= \frac{b - a}{g(b) - g(a)} \cdot (-g(a)) + a \end{aligned}$$

Метод хорд

Решение задачи

$$\begin{cases} f(x) \rightarrow \min \\ x \in [a; b] \end{cases}$$

< !!! БЛОК СХЕМА >

Замечание: На каждой итерации, кроме первой, необходимо вычислять только одно значение функции $f'(\bar{x})$.

1.2.6. Метод Ньютона

Пусть

1. $f \in C^2[a; b]$
2. $f''(x) > 0$

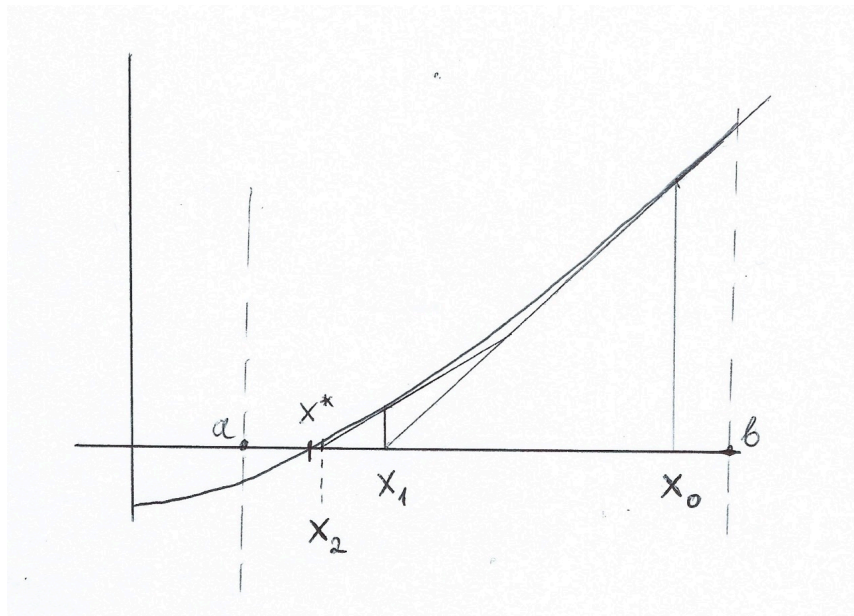
Из условия 2 вытекает, что f выпукла на $[a; b]$. Совместно с условием 1 это значит, что f унимодальна.

Метод Ньютона поиска минимума функции $f(x)$ является метод касательных (Ньютона) решения уравнения $f'(x) = 0$.

Замечание: Метод касательных решения $g(x) = 0$.

Пусть $g'(x)$ имеет постоянный знак на $[a; b]$,

В качестве очередного приближения неизвестного корня x^* используется точка \bar{x} пересечения касательной к графику функции $g(x)$ в точке \bar{x}' , где x' — текущее приближение известного корня.

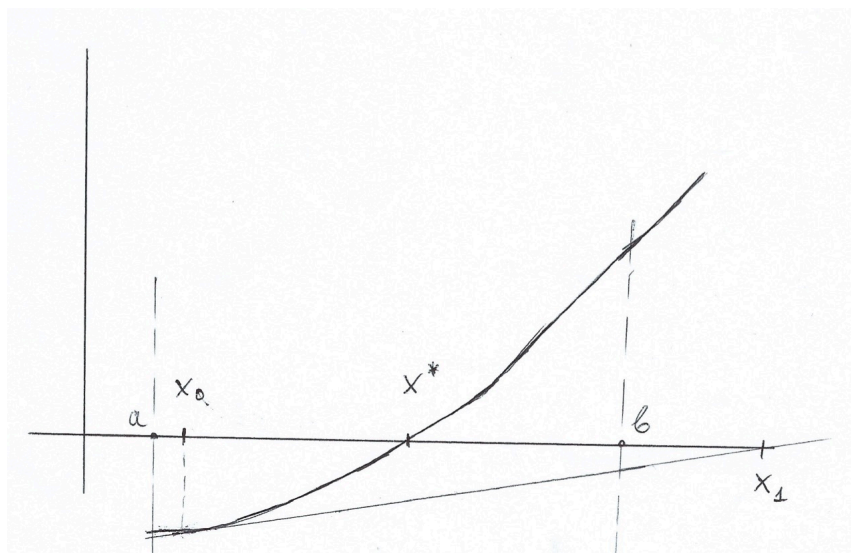


Условием окончания итераций служит:

либо $|\bar{x}' - \bar{x}| \leq \varepsilon,$

либо $|g(\bar{x})| \leq \varepsilon.$

Замечание: Метод Ньютона обладает высокой точностью и скоростью сходимости, только в том случае, когда начальное приближение x_0 достаточно близко к x^* . В случае неудачного выбора x_0 метод может расходиться. Как правило, чем больше значения функции $g'(x)$ в окрестности x^* , тем лучше сходится метод.



Расчетные соотношения метода Ньютона:

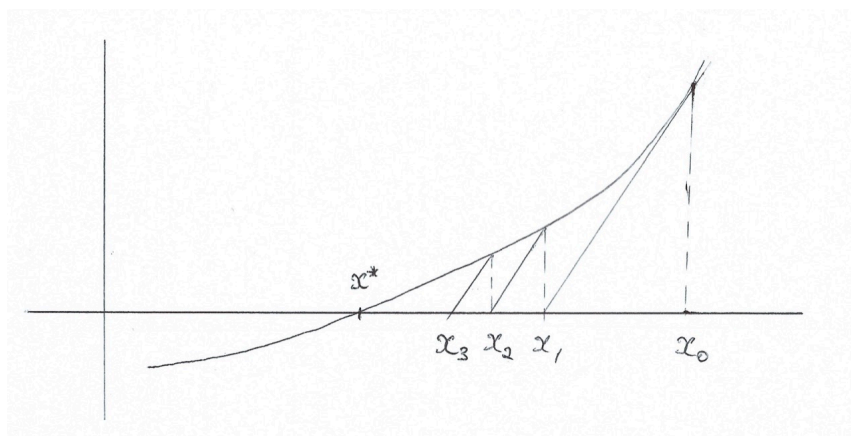
1. Уравнение касательной в точке \bar{x}' :

$$y = g'(\bar{x}') \cdot (x - \bar{x}') + g(\bar{x}')$$

2. Пересечение с Ox

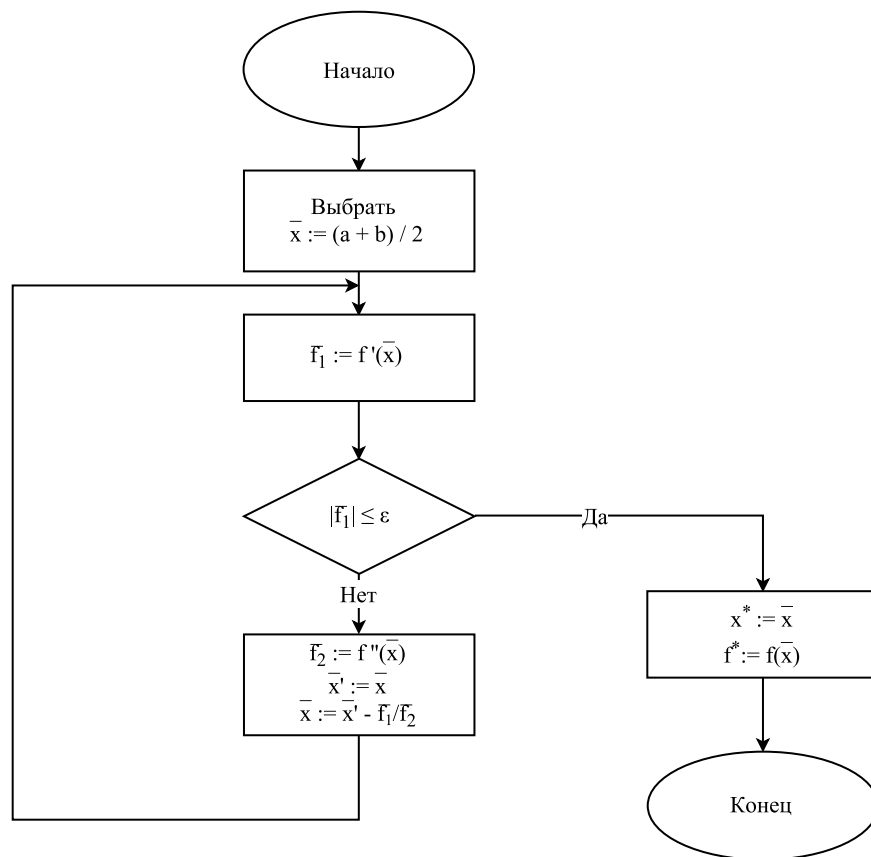
Замечание: Иногда, когда вычисление $g'(x)$ очень трудоемко, используют модификацию метода Ньютона, которая называется «Методом одной касательной».

В качестве очередного приближения \bar{x} неизвестного корня x^* используют точку пересечения Ox прямой, проходящей через точку $(\bar{x}', g(\bar{x}'))$, где \bar{x}' — текущее приближение, параллельно касательной к графику $g(x)$ в точке x_0 , x_0 — начальное приближение.



Метод Ньютона решения задачи

$$\begin{cases} f(x) \rightarrow \min \\ x \in [a; b] \end{cases}$$



Замечание:

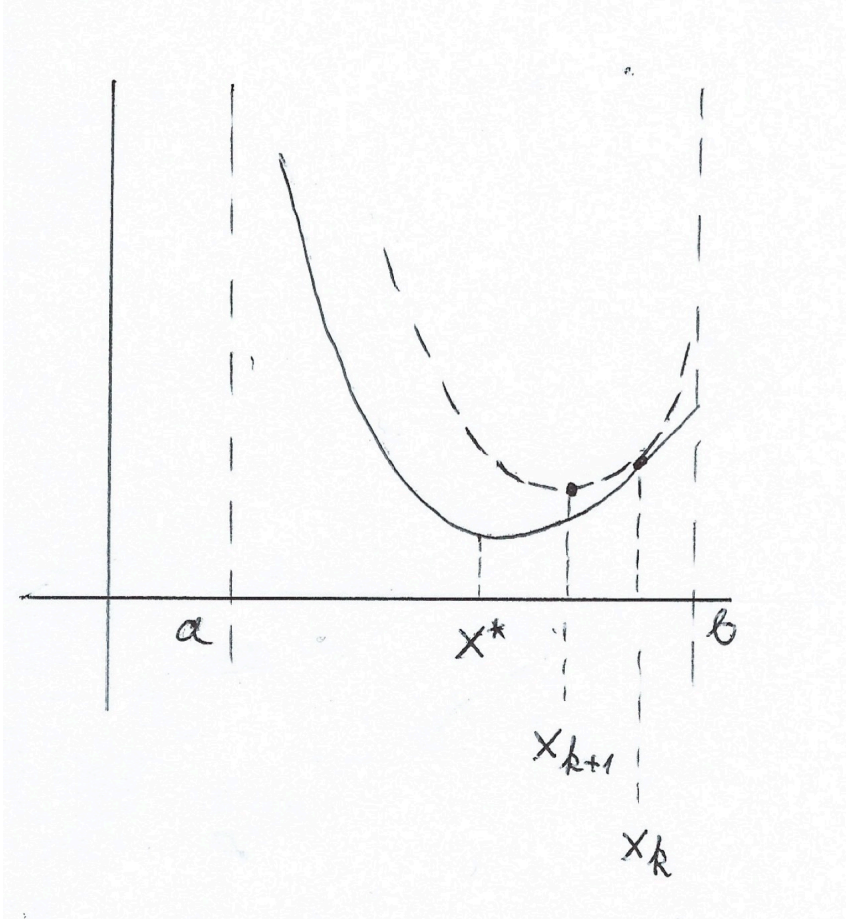
1. Формулу $x_{k+1} := x_k - \frac{f'(x_k)}{f''(x_k)}$, с помощью которой вычисляется очередное приближение точки x^* , так же можно получить из следующих соображений. Рассмотрим квадратный трехчлен.

$$q(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2}f''(x_k)(x - x_k)^2$$

Точка минимума трехчлена $q(x)$:

$$\bar{x} = x_k + \left(-\frac{b}{2a} \right) = x_k + \frac{-f'(x_k)}{2 \cdot \frac{1}{2} \cdot f''(x_k)} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

Так как функция f выпукла и дважды дифференцируемая то её график в окрестности точки x_k «похожа» на параболу и, следовательно, парабола хорошо аппроксимирует f . Поэтому точка x_{k+1} минимума функции $q(x)$ близка к точке x^* функции $f(x)$, если x_k выбрана удачно.



- Можно показать, что погрешность k -ой итерации метода Ньютона

$$|x_{k+1} - x^*| \leq C \cdot q^{2^k},$$

где $C > 0$, $q \in (0; 1)$, но только в случае, если начальные x_0 выбраны удачно. Константы C , q зависят от f и от выбора x_0 .

- Если метод Ньютона расходится, то можно выполнить несколько некоторых итераций какого-нибудь другого метода, например метода золотого сечения.

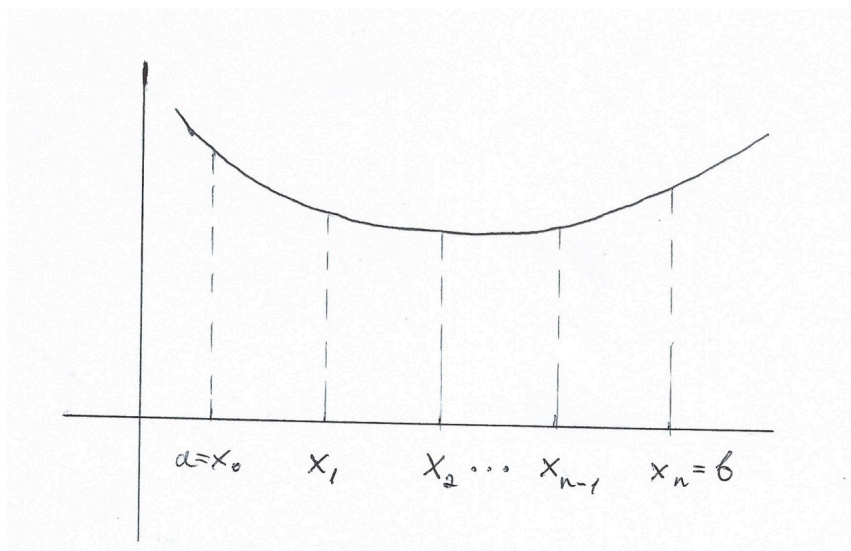
1.2.7. Метод перебора

Метод был описан выше. Отрезок $[a; b]$ разбивается на n равных частей $n + 1$ точками:

$$\begin{aligned} x_i &= a + i\Delta; & i &= \overline{0; n} \\ \Delta &= \frac{b - a}{n} \end{aligned}$$

В качестве точки x^* принимается точка x_m , для которой

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



Ранее была доказана сходимость этого метода для унимодальных на $[a; b]$ функций.

Оказывается этот метод сходится и для многомодальных функций, если дополнительно потребовать липшицевость целевой функции.

Теорема: Пусть

1. Функция f удовлетворяет на отрезке $[a; b]$ условию Липшица с константой L
2. $x^* := x_m$ — точка минимума, найденная методом перебора
3. $f^* := f(x^*)$

Тогда

$$\delta_n \leq L \cdot \frac{b-a}{2n},$$

где n — число отрезков разбиения $[a; b]$, $\delta_n = |f^* - f_{\text{точн}}^*|$.

□

1. Так как f непрерывна на отрезке $[a; b]$, то f достигает на $[a; b]$ своей точки минимума (?), тогда $\exists \min_{x \in [a; b]} f(x) = f_{\text{точн}}^*$
2. Пусть x^* — точка глобального минимума $f(x)$ на $[a; b]$

Очевидно, что среди точек x_i , $i = \overline{0; n}$, найдется точка x_k такая, что

$$|x_k - x_{\text{точн}}^*| \leq \frac{\Delta}{2} = \frac{b-a}{2n}$$

Тогда

$$0 \leq f(x_m) - f_{\text{точн}}^* \leq f(x_k) - f_{\text{точн}}^* \leq L \cdot |x_k - x_{\text{точн}}^*| \leq L \cdot \frac{b-a}{2n}$$

Ввиду того, что $f(x_k) \geq f(x_m)$

■

Замечание:

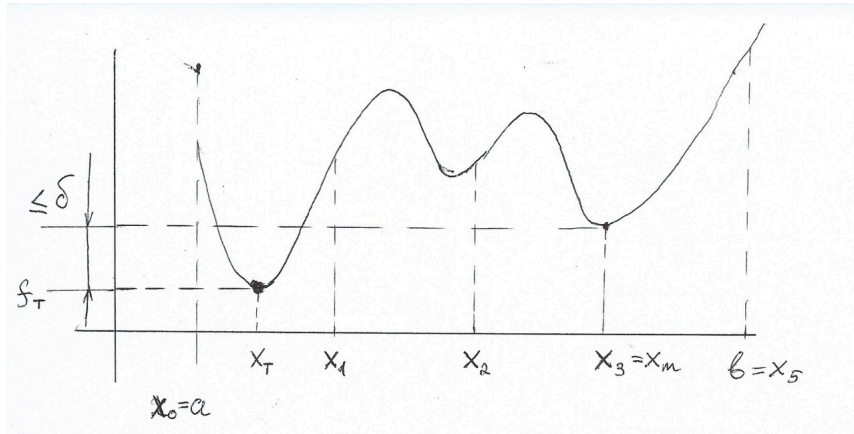
1. Из доказанной теоремы вытекает, что вычисления n значений целевой функции гарантирует точность

$$\delta(N) \leq L \cdot \frac{(b-a)}{2(N-1)}$$

2. Для обеспечения заданной точности δ необходимо вычислить

$$N \geq \frac{L(b-a)}{2\delta} + 1$$

3. Может получиться так, что значение f^* найдено с заданной точностью δ , но точка минимума $x^* = x_m$ далека от $x_{\text{точн}}^*$



1.2.8. Метод ломаных

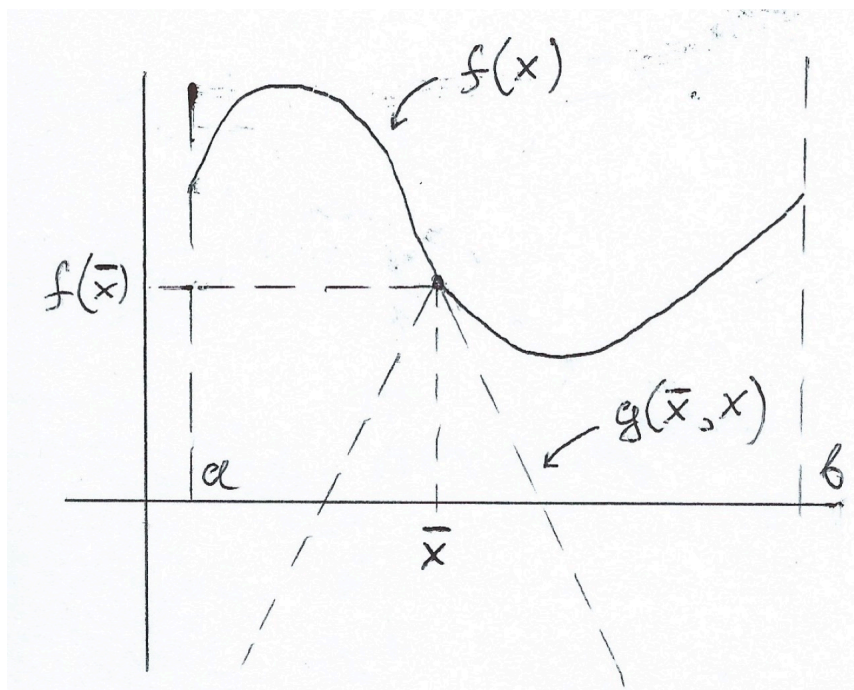
Метод ломаных также является прямым методом минимизации многомодальных функций.

Пусть

1. f удовлетворяет на $[a; b]$ условию Липшеца с константой L
2. $\bar{x} \in [a; b]$
3. Введем в рассмотрение вспомогательную функцию

$$\begin{aligned} g(\bar{x}, x) &= f(\bar{x}) - L|x - \bar{x}| \\ \bar{x} &= \text{const} \\ x &= \text{var} \end{aligned}$$

Замечание: График $g(\bar{x}, x)$ располагается ниже графика $f(x)$ так как L — константа Липшеца для f .



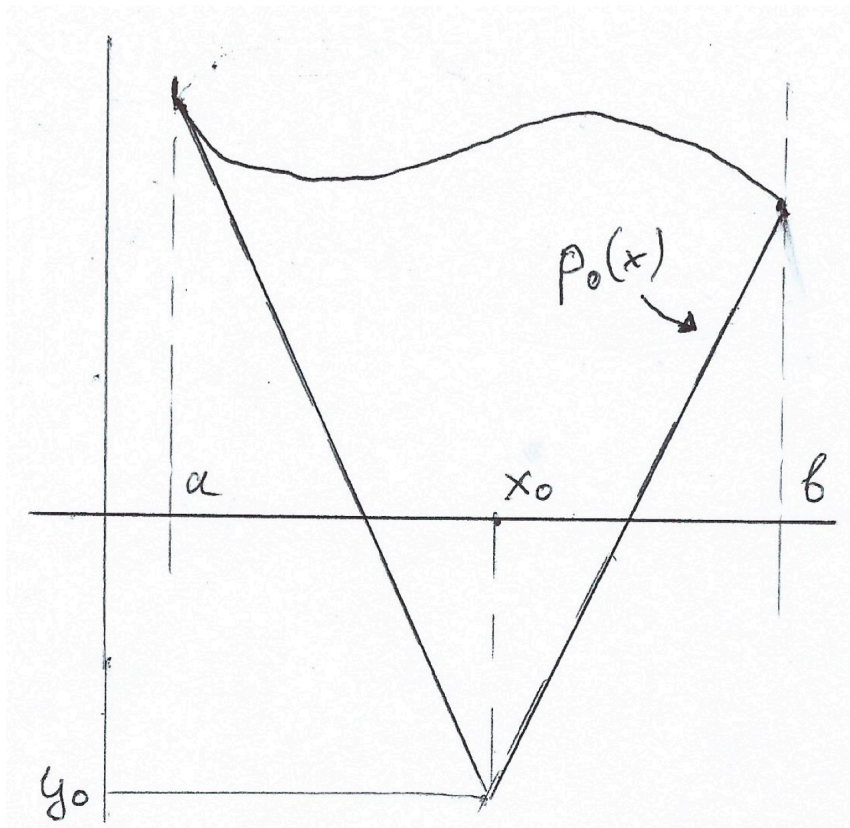
Идея метода ломаных заключается в аппроксимации целевой функции f кусочно-линейной функцией, звенья которой имеют угловые коэффициенты $\pm L$. В качестве минимального значения функции f принимается минимальное значение функции, графиком которой является эта ломанная.

Обозначим: $p_k(x)$, $k = 0, 1, 2, \dots$ кусочно-линейная функция, построенная на k -ой итерации.

#0 Рассмотрим прямые

$$y = f(a) - L(x - a)$$

$$y = f(b) - L(x - b)$$



Эти прямые пересекаются в точке с координатами

$$x_0 = \frac{1}{2L} [f(a) - f(b) + L(a + b)]$$

$$y_0 = \frac{1}{2} [f(a) - f(b) + L(a - b)]$$

Примем

$$p_0(x) = \begin{cases} f(a) - L(x - a), & a \leq x \leq x_0 \\ f(b) + L(x - b), & x_0 \leq x \leq b \end{cases}$$

#1

Построим $p_1(x)$.

Функция $p_0(x)$ имеет единственную точку глобального минимума $x_0^* = x_0$.

Положим

$$p_1(x) = \max \{p_0(x), g(x_0^*, x)\}$$

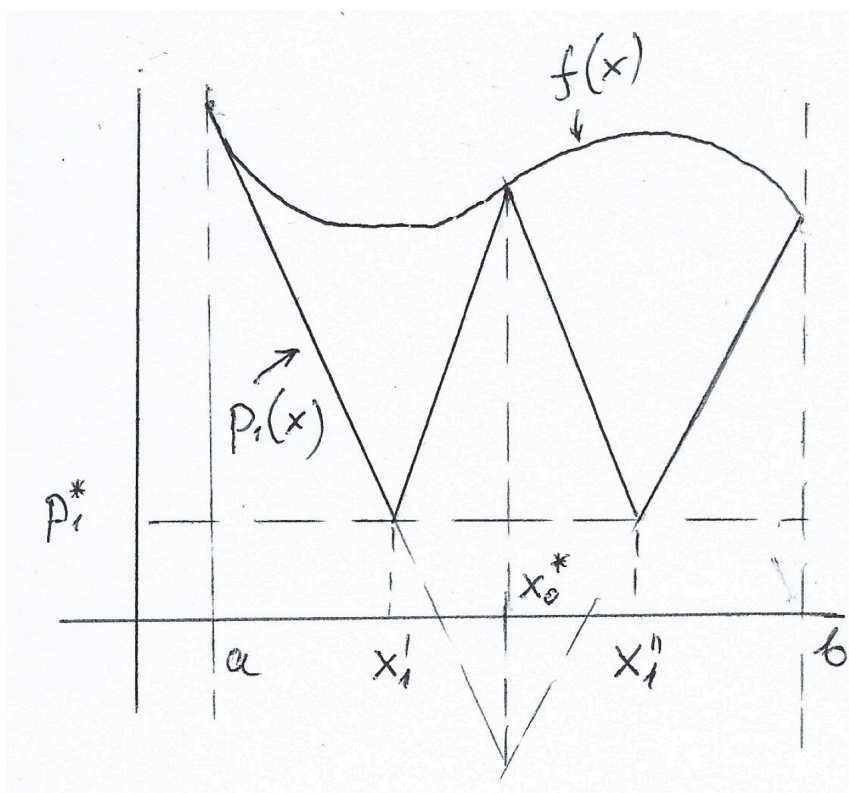
$p_1(x)$ в отличие от $p_0(x)$ вместо одной точки x_0^* глобального минимума имеет две точки x_1' и x_1'' локального минимума.

Прим этом

$$x_1' = x_0^* - \Delta_1$$

$$x_1'' = x_0^* + \Delta_1$$

$$\Delta_1 = \frac{1}{2L} [f(x_0^*) - p_0^*]$$



#2

Опишем построение функции $p_2(x)$.

Выберем произвольную точку x_1^* , в которой $p_1(x)$ имеет глобальный минимум (в нашем случае таких точек две: x_1' и x_1'' , выберем $x_1^* = x_1'$).

Положим

$$p_2(x) = \max \{p_1(x), g(x_1^*, x)\}$$

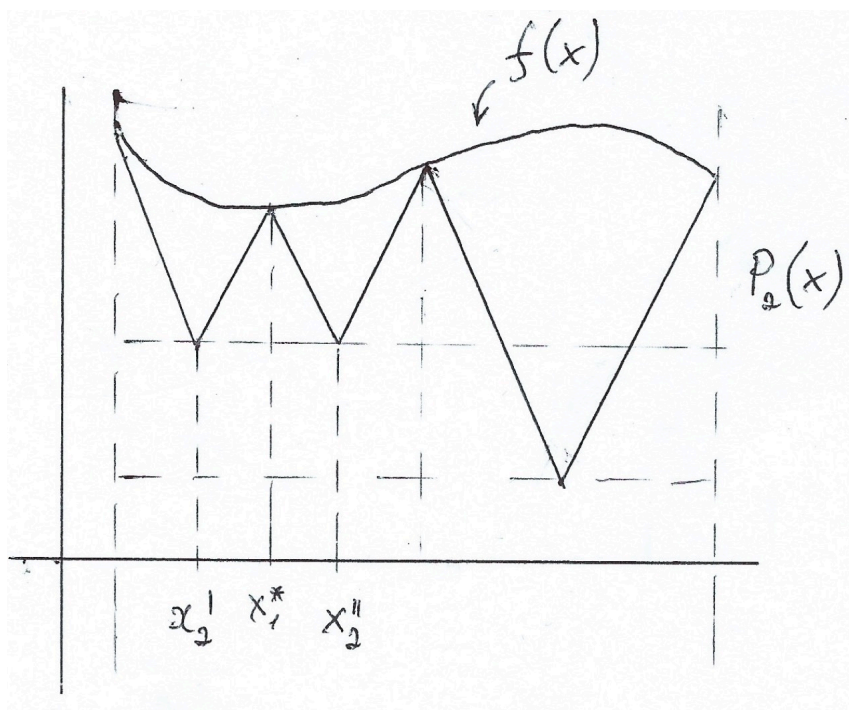
По сравнению с $p_1(x)$ функция $p_2(x)$ вместо точки x_1^* глобального минимума имеет две точки x_2' и x_2'' локального минимума

$$\begin{aligned} x_2' &= x_1 - \Delta_2 \\ x_2'' &= x_1 + \Delta_2 \\ \Delta_2 &= \frac{1}{2} [f(x_1^*) - p_1^*] \end{aligned}$$

Причем

$$p_2(x_2') = p_2(x_2'') = \frac{1}{2} [f(x_1^*) + p_1^*]$$

...



#k

Пусть построена функция $p_{k-1}(x)$. Опишем построение $p_k(x)$.

Пусть x_{k-1}^* — точка глобального минимума функции $p_{k-1}(x)$, $p_{k-1}^* = p_{k-1}(x_{k-1}^*)$.

Положим

$$p_k(x) = \max \{ p_{k-1}(x), g(x_{k-1}^*, x) \}.$$

Функция $p_k(x)$ по сравнению с $p_{k-1}(x)$ будет иметь две точки x_k' и x_k'' локального минимума вместо одной точки x_{k-1}^* глобального минимума.

При этом

$$\begin{aligned} x_k' &= x_{k-1}^* - \Delta_k \\ x_k'' &= x_{k-1}^* + \Delta_k \\ \Delta_k &= \frac{1}{2L} [f(x_{k-1}^*) - p_{k-1}^*] \end{aligned}$$

Причем

$$p_k(x_k') = p_k(x_k'') = \frac{1}{2} [f(x_{k-1}^*) + p_{k-1}^*]$$

Свойства функций $p_k(x)$

1. $p_k(x)$ — непрерывная кусочно-линейная функция, каждое звено которой имеет угловой коэффициент $+L$ или $-L$.
2. $p_k(x)$ имеет ровно $k+1$ точку локального минимума.
3. $p_{k-1}(x) \leq p_k(x) \leq f(x)$, $x \in [a; b]$
4. Ломанные $p_k(x)$ при $k \rightarrow \infty$ приближаются снизу к графику функции $f(x)$ в окрестностях точек её глобального минимума

Основное достоинство метода ломаных заключается в том, что минимум кусочно-линейной функции искать существенно проще, чем минимум $f(x)$. При этом на каждом шаге (кроме нулевого) метода ломаных требуется вычисление лишь одного значения целевой функции.

Теорема: Пусть

1. $f(x)$ удовлетворяет условию Липшеца на $[a; b]$ с константной L .
2. $p_k(x)$, $k = 0, 1, 2, \dots$ — последовательность ломаных, построенных по указанному методу.

3. x_k^* — точка глобального минимума функции $p_k(x)$

Тогда

1. $\lim_{k \rightarrow \infty} p_k(x_k^*) = f(x^*) = \min_{x \in [a; b]} f(x)$
2. Если точка x^* глобальный минимум функции $f(x)$ единственна на $[a; b]$, то $x_k^* \xrightarrow{k \rightarrow \infty} x^*$
3. Оценка погрешности

$$\delta_k = 2L\Delta_k,$$

где $\delta_k = f(x_k^*) - f^*$

Замечание:

1. Последняя оценка погрешности используется в качестве критерия для остановки вычислений.
2. Если $f(x)$ имеет несколько глобальных минимумов на $[a; b]$, то есть $|G^*| \geq 2$, то пункт 2 теоремы можно сформулировать таким образом

$$\lim_{k \rightarrow \infty} \rho_k = 0,$$

где ρ_k — расстояние от x_k^* до ближайшей ей точки из G^* .

< БЛОК СХЕМА !!! >

Замечание: об определении константы L

1. Если получится, то можно оценить производную $|f'(x)| \leq A$ и принять $L = A$.
2. Найти угловые коэффициенты k_1, \dots, k_l некоторого количества хорд графика функции $f(x)$. Эти значения являются нижними оценками для постоянной Липшица, тогда можно принять

$$L = \max_{i=1; \overline{l}} \{k_i\} + \{\text{некоторая величина}\}$$

Если {некоторую величину} выбрать слишком большой, то метод будет долго сходиться.

Если {некоторую величину} выбрать слишком малой, так, что полученное значение L не будет постоянной Липшица для f , то метод может разойтись.