НАО «Восточно-Казахстанский технический университет имени Д. Серикбаева»

Дисциплина «Методы оптимизации»

ЛЕКЦИЯ 6

Методы одномерной оптимизации: основные понятия и классификация

Образовательная программа 6B06103 «Математическое и компьютерное моделирование»

Объём: 2 академических часа. **Место в программе:** лекция 6.

Цель: научиться правильно ставить задачу минимизации функции одной переменной, обосновывать сокращение области поиска и выбирать класс численных методов с учётом доступной информации.

После изучения темы студент должен уметь: различать локальный и глобальный минимум; проверять достаточные признаки унимодальности; объяснять роль начального отрезка; выполнять один шаг исключения интервала; оценивать погрешность по его длине; связывать одномерную минимизацию с выбором шага в многомерной задаче.

1. Зачем нужны методы одномерной оптимизации

Во многих прикладных задачах решение зависит от одного управляемого параметра: времени обработки, скорости движения, объёма выпуска или коэффициента настройки. Требуется подобрать его значение, при котором затраты, ошибка или расход энергии минимальны. Даже если исходная модель содержит несколько переменных, одномерная задача возникает при движении вдоль выбранного направления.

Предыдущие лекции были посвящены линейному программированию. Теперь начинается раздел численных методов безусловной оптимизации. В этой лекции вводятся общие понятия. Подробные алгоритмы дихотомии, золотого сечения, Фибоначчи и касательных рассматриваются в следующих темах рабочей программы. Здесь важно понять, какую информацию использует метод и на каком основании он исключает часть допустимых значений [1].

Пример постановки. Пусть удельные затраты зависят от положительного параметра x:

$$f(x) = x^2 + \frac{2}{x}, \qquad x \in [0{,}5;\, 3].$$

Первое слагаемое возрастает при увеличении x, второе убывает. Нельзя минимизировать их по отдельности и затем объединить ответы: требуется минимизировать сумму. Для этой простой модели решение находится аналитически. В более сложной модели вместо формулы может использоваться результат численного расчёта или эксперимента.

2. Математическая постановка и существование решения

Задача одномерной минимизации записывается в виде

$$f(x) \to \min, \qquad x \in D \subseteq \mathbb{R}.$$

Множество D задаёт допустимые значения аргумента. Если решение достигается в точке x^*, то

$$f(x^*) \le f(x) \quad \text{для всех } x \in D.$$

Число x^* называется точкой глобального минимума, а $f^* = f(x^*)$ — минимальным значением. В отчёте необходимо указывать оба результата. Максимизация сводится к минимизации заменой целевой функции на $-f(x)$; допустимое множество при этом сохраняется.

Локальный минимум требует выполнения такого неравенства только для допустимых точек из некоторой окрестности x^*. Поэтому локальное решение может не быть лучшим на всём множестве D. У функции $(x^2-1)^2$ на $[-2;2]$ две точки глобального минимума: -1 и 1. Единственность решения нельзя предполагать без исследования функции.

Теорема Вейерштрасса. Непрерывная функция на непустом замкнутом ограниченном отрезке $[a;b]$ достигает своего наименьшего и наибольшего значений. Теорема гарантирует существование решения, но сама по себе не даёт ни его единственности, ни алгоритма поиска.

Если какое-либо условие нарушено, решение всё ещё может существовать, однако общей гарантии уже нет. Например, у $f(x)=x$ на $(0;1)$ нижняя грань равна нулю, но допустимой точки с таким значением нет. У $f(x)=-x$ на всей числовой прямой нет конечного минимального значения: функция не ограничена снизу.

Практический вывод. До запуска программы нужно записать область определения и ограничения. Нельзя подставлять ноль в $1/x$, отрицательные числа в вещественный логарифм или рассматривать выход за границы физической модели как обычный шаг алгоритма.

3. Стационарные точки и границы отрезка

Пусть функция дифференцируема в окрестности внутренней точки локального минимума. Тогда необходимое условие имеет вид

$$f'(x^*) = 0.$$

Обратное утверждение неверно. Для $f(x)=x^3$ производная в нуле равна нулю, но слева значения отрицательны, а справа положительны: экстремума в нуле нет. Уравнение $f'(x)=0$ даёт кандидатов, которые требуют дальнейшего анализа.

Если $f'(x^*)=0$ и $f''(x^*)>0$, то при обычных условиях непрерывности второй производной точка является строгим локальным минимумом. При $f''(x^*)=0$ этот признак не решает задачу: у x^4 в нуле минимум есть, а у x^3 его нет.

На границе отрезка производная не обязана обращаться в ноль. Функция $f(x)=x$ на $[0;1]$ имеет минимум при $x=0$, хотя $f'(0)=1$. Поэтому классический поиск глобального минимума на отрезке предусматривает сравнение значений во внутренних стационарных точках, в допустимых точках недифференцируемости и на обоих концах.

Для исходного примера

$$f'(x) = 2x - \frac{2}{x^2}, \qquad f''(x) = 2 + \frac{4}{x^3} > 0 \quad (x>0).$$

Уравнение $f'(x)=0$ эквивалентно $x^3=1$. Получаем $x^*=1$, $f^*=3$. Проверка концов даёт $f(0{,}5)=4{,}25$ и $f(3)=29/3$. Строгая выпуклость на положительной полуоси подтверждает единственность найденного минимума.

4. Унимодальность и выпуклость

4.1. Определение, используемое в данной лекции

Функцию будем называть **строго унимодальной относительно минимума** на $[a; b]$, если существует единственная точка x^*, до которой функция строго убывает, а после которой строго возрастает. Если x^* совпадает с концом отрезка, соответствующая ветвь отсутствует. Ниже правила исключения интервала формулируются именно для этого определения. В литературе встречаются и нестрогие определения унимодальности, для которых некоторые выводы требуют уточнения.

Геометрически график имеет одну «впадину». Гладкость не обязательна: $f(x) = |x|$ на $[-2; 2]$ строго унимодальна и не имеет производной в точке минимума. Поэтому методы, использующие только значения функции, применимы к более широкому классу задач, чем методы на основе производных.

Если дифференцируемая функция имеет отрицательную производную слева от x^* и положительную справа, она строго унимодальна. Это достаточный признак. Проверка знака производной в нескольких точках графика не заменяет доказательства на всём соответствующем промежутке.

4.2. Связь с выпуклостью

Функция выпукла на отрезке, если её значение в любой внутренней точке хорды не превосходит соответствующей линейной интерполяции:

$$f((1-t)x + ty) \le (1-t)f(x) + tf(y), \qquad 0 \le t \le 1.$$

Для дважды дифференцируемой функции условие $f''(x) \ge 0$ на интервале является признаком выпуклости; условие $f''(x) > 0$ — достаточным признаком строгой выпуклости. Строго выпуклая функция, достигающая минимума на отрезке, имеет единственный минимум и строго унимодальна относительно него.

Обратное утверждение неверно. Функция $f(x) = \sqrt{|x|}$ строго унимодальна на $[-2; 2]$, но не выпукла: на положительной полуоси квадратный корень вогнут. Например, $f(0{,}5) > [f(0) + f(1)]/2$. Таким образом, требование унимодальности для исключения интервала слабее требования строгой выпуклости.

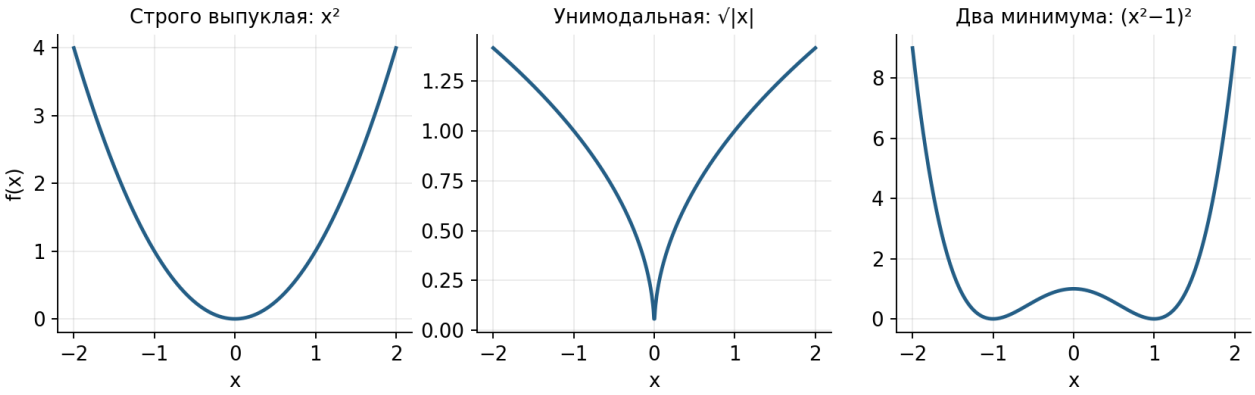


Рисунок 1 — Выпуклость, унимодальность и наличие нескольких минимумов.

У функции $(x^2-1)^2$ между двумя минимумами находится локальный максимум. Если применять к ней правило для одной впадины без проверки предпосылок, можно исключить область с глобальным решением. Построенный по редкой сетке график полезен для разведки, но сам по себе не является доказательством унимодальности.

5. Отрезок неопределённости и начальная локализация

Отрезок неопределённости $[a_k; b_k]$ — отрезок, про который обоснованно известно, что он содержит искомую точку минимума. Метод последовательно строит вложенные отрезки:

$$[a_{k+1}; b_{k+1}] \subseteq [a_k; b_k], \qquad x^* \in [a_k; b_k].$$

Начальный отрезок может быть задан условиями задачи. Если ограничения заданы явно, поиск не должен выходить за них. Если функция определена на неограниченной области, сначала требуется локализовать подходящую впадину. Один из подходов — последовательное увеличение пробного шага в направлении уменьшения функции.

Набор точек $a < m < b$ с неравенствами $f(m) < f(a)$ и $f(m) < f(b)$ у непрерывной функции гарантирует наличие хотя бы одного внутреннего минимума на $[a; b]$. Он не доказывает, что этот минимум единственен, и не исключает более низкие значения вне найденного отрезка. Если дополнительно известно, что функция унимодальна на области поиска, такая локализация становится основанием для дальнейшего сужения.

Если значения продолжают уменьшаться при расширении шага, нельзя бесконечно увеличивать его. Нужны ограничение числа вычислений, контроль области определения и сообщение о том, что локализация не получена. Например, для $f(x) = -x$ движение вправо всегда уменьшает функцию и не приводит к конечному минимуму.

6. Правило исключения части отрезка

Пусть f строго унимодальна на $[a; b]$, а $a < y < z < b$. Достаточно сравнить два значения $f(y)$ и $f(z)$ [1, § 5.1.1].

Результат сравнения	Новый отрезок, содержащий минимум	Обоснование
$f(y) < f(z)$	$[a; z]$	Минимум не может лежать правее z: до него функция должна убывать.
$f(y) > f(z)$	$[y; b]$	Минимум не может лежать левее y: после него функция должна возрастать.
$f(y) = f(z)$	$[y; z]$	Две точки с равными значениями не могут лежать на одной строго монотонной ветви.

Доказательство первого случая. Предположим противное: $x^* > z$. Тогда обе точки $y < z$ лежат на убывающей ветви. Следовательно, $f(y) > f(z)$, что противоречит исходному сравнению. Значит, $x^* \le z$. Второй случай доказывается аналогично; в третьем минимум расположен между пробными точками.

Правило отвечает на вопрос, какую часть отрезка можно исключить, но ещё не задаёт полноценный алгоритм. Алгоритмы отличаются способом выбора y и z, повторным использованием уже вычисленных значений и условием остановки.



Рисунок 2 — Иллюстрация первого правила для $f(x) = (x - 1{,}7)^2 + 2$.

6.1. Числовой пример трёх последовательных шагов

Рассмотрим строго выпуклую функцию $f(x) = (x - 1{,}7)^2 + 2$ на $[0; 5]$. В учебном примере точка 1,7 известна из формулы, но при выполнении шагов используются только значения в пробных точках.

Шаг	Отрезок до сравнения	$y; z$	$f(y); f(z)$	Отрезок после сравнения
1	$[0; 5]$	1; 3	2,49; 3,69	$[0; 3]$
2	$[0; 3]$	1; 2	2,49; 2,09	$[1; 3]$
3	$[1; 3]$	1,4; 1,8	2,09; 2,01	$[1{,}4; 3]$

После третьего шага середина отрезка равна 2,2, а гарантированная погрешность по аргументу не превышает 0,8. Фактическая ошибка равна 0,5. Пример показывает правильность исключения, но выбор пробных точек здесь специально не оптимизирован. Более систематические способы выбора рассматриваются далее.

Если каждое сравнение заново вычисляет два значения, выполнено шесть обращений к функции. Сохранение результата $f(1)$ сокращает это число до пяти. Совпадение численных значений $f(2)$ и $f(1{,}4)$ само по себе не означает, что второе значение можно было узнать заранее.

7. Погрешность и критерии остановки

Если отрезок содержит минимум, то для его середины получается оценка

$$\hat{x}_k = \frac{a_k + b_k}{2}, \qquad |\hat{x}_k - x^*| \le \frac{b_k - a_k}{2}.$$

Поэтому для гарантированной абсолютной погрешности ε_x достаточно остановиться при $b_k - a_k \le 2\varepsilon_x$. Эта оценка относится к середине отрезка. Если программа возвращает другую пробную точку, её погрешность необходимо оценивать отдельно.

Если длина отрезка уменьшается с коэффициентом не более q, где $0 < q < 1$, то после k сокращений она не превышает $q^k(b_0 - a_0)$. При начальной длине больше $2\varepsilon_x$ достаточное число сокращений удовлетворяет условию

$$k \geq \left\lceil \frac{\ln((b_0 - a_0)/(2\varepsilon_x))}{\ln(1/q)} \right\rceil.$$

При начальной длине 8, коэффициенте $q = 1/2$ и точности 0,01 нужно не менее девяти сокращений: $8/2^9 = 0{,}015625 \leq 0{,}02$. Это оценка для метода с указанным коэффициентом; её нельзя автоматически переносить на любой способ выбора пробных точек.

Погрешность аргумента и погрешность функции различны. Малое изменение f не обязательно означает близость к точке минимума. Для $f(x) = 10^{-8}(x - 2)^2$ значение при $x = 12$ отличается от минимального только на 10^{-6}, хотя ошибка аргумента равна 10.

При внутреннем минимуме $f'(x^*) = 0$ и оценке $|f''(x)| \leq M$ на отрезке между x^* и $\hat{x}$ формула Тейлора даёт

$$0 \leq f(\hat{x}) - f(x^*) \leq \frac{M}{2}|\hat{x} - x^*|^2.$$

Если дополнительно $f''(x) \geq \mu > 0$, то

$$f(\hat{x}) - f(x^*) \geq \frac{\mu}{2}|\hat{x} - x^*|^2.$$

Эти оценки связывают две погрешности через свойства функции. Для минимума на границе первая квадратичная оценка без условия нулевой производной в общем случае неверна.

В вычислениях с плавающей точкой также контролируют число итераций и различимость пробных точек. Нельзя считать два значения равными только потому, что они совпали при печати до двух знаков. При шумных вычислениях надёжность сравнения требует отдельной оценки ошибки; обычное правило для точных значений такой гарантии не даёт.

8. Классификация методов

Группа	Используемая информация и примеры	Условия и особенности
Нулевого порядка	Только значения f: перебор по сетке, дихотомия, золотое сечение, Фибоначчи	Производные не нужны. Для стандартного исключения интервала требуется унимодальность.
На основе интерполяции	Приближение функции по нескольким значениям: параболическая интерполяция	Обычно относятся к нулевому порядку; модель может дать неудачную точку, поэтому полезно сохранять надёжный отрезок.
Первого порядка	Значения f и/или f': поиск нуля производной, методы с касательными	Нужна соответствующая гладкость; поиск корня производной сам по себе не доказывает глобальный минимум.
Второго порядка	Производные f' и f'': метод Ньютона	Может быстро сходиться вблизи решения, но требует контроля кривизны и начальной точки.

Дополнительно различают **пассивные** и **последовательные** схемы. В пассивной схеме все пробные точки задаются заранее, например узлы равномерной сетки. В последовательной схеме следующая точка зависит от ранее полученных результатов. Такое использование информации позволяет сосредоточить вычисления в перспективной части области.

Равномерная сетка. Пусть шаг равен h и выбран узел с наименьшим значением строго унимодальной функции. Минимум лежит между его соседями; при выборе крайнего узла используется соседний крайний отрезок. Погрешность лучшего узла по аргументу не превосходит h. Оценку $h/2$ без дополнительных предпосылок применять нельзя: минимальное значение среди узлов не всегда достигается в ближайшем к x^* узле при несимметричной форме впадины.

В методе дихотомии сравнивают две точки вблизи середины. В золотом сечении точки располагают так, чтобы после сокращения переиспользовать одно вычисленное значение. Метод Фибоначчи учитывает заранее заданный бюджет поиска. Эти различия влияют на число вычислений функции, а не только на число итераций.

Для оценки работы алгоритма следует отдельно указывать точность, число итераций, число вычислений f, а при необходимости — производных. Одна итерация разных методов может иметь существенно разную стоимость.

9. Одномерный поиск в многомерной задаче

В многомерной оптимизации новая точка часто строится по формуле

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{d}^k, \qquad \alpha_k \geq 0.$$

После выбора направления $\mathbf{d}^k$ задача выбора длины шага становится одномерной:

$$\varphi(\alpha) = f(\mathbf{x}^k + \alpha \mathbf{d}^k) \to \min.$$

Если направление является направлением спуска, то $\nabla f(\mathbf{x}^k)^T \mathbf{d}^k < 0$ и достаточно малый положительный шаг уменьшает функцию. Это локальное утверждение: произвольно большой шаг может привести к увеличению значения.

Для квадратичной функции с положительно определённой матрицей Гессе H и ненулевым направлением спуска $\mathbf{d}$ точный шаг имеет вид

$$\alpha^* = -\frac{\nabla f(\mathbf{x})^T \mathbf{d}}{\mathbf{d}^T H \mathbf{d}}.$$

При направлении $\mathbf{d} = -\nabla f(\mathbf{x})$ получается формула наискорейшего спуска, которая подробно разбирается в лабораторной работе 6. Для общей нелинейной функции такой готовой формулы обычно нет, и нужны методы одномерного поиска.

10. Самопроверка и задания

1. Чем точка минимума отличается от минимального значения функции?
2. Какие условия гарантируют существование минимума на отрезке? Гарантируют ли они единственность?
3. Почему равенства $f'(x) = 0$ недостаточно для вывода о минимуме?
4. Может ли минимум на отрезке достигаться в точке с ненулевой производной?
5. Приведите пример строго унимодальной, но невыпуклой функции.

6. Докажите правило исключения левой части отрезка при $f(y) > f(z)$.
7. Почему правило равных значений требует внимания к определению унимодальности?
8. Как длина отрезка неопределённости связана с погрешностью его середины?
9. Почему нельзя сравнивать методы только по числу итераций?
10. Как одномерная минимизация используется при движении вдоль направления?

Задание 1. Для $f(x) = (x-3)^2 + 1$ на $[0; 8]$ сравните значения в точках 2 и 5. Найдите новый отрезок, его середину и гарантированную погрешность середины.

Задание 2. Начальная длина отрезка равна 12. После каждого шага она сокращается вдвое. Сколько сокращений достаточно для погрешности середины не более 0,005?

Задание 3. Исследуйте $f(x) = x + 1/x$ на $[0{,}25; 4]$: найдите производные, обоснуйте единственность минимума и проверьте концы отрезка.

Задание 4. Объясните, почему применение одного унимодального поиска к функции $(x^2 - 1)^2$ на $[-2; 2]$ не имеет стандартной гарантии. Предложите допустимое предварительное разбиение области.

Краткие ответы. 1: $f(2) = 2 < f(5) = 5$, новый отрезок $[0; 5]$, середина 2,5, оценка ошибки 2,5. 2: одиннадцать сокращений. 3: $f' = 1 - 1/x^2$, $f'' = 2/x^3 > 0$, минимум при $x = 1$, значение 2; на обоих концах 4,25. 4: на всём отрезке две впадины; можно исследовать $[-2; 0]$ и $[0; 2]$ отдельно и сравнить результаты.

11. Литература

[1] Пантелеев А. В., Летова Т. А. Методы оптимизации в примерах и задачах: учебное пособие. 4-е изд., испр. СПб.: Лань, 2015. 512 с. Для данной лекции: § 5.1.1, с. 95–97; связь с выбором шага — § 6.2, с. 165–166.

Объяснения, числовые примеры, рисунки и задания подготовлены для данной лекции. Материал задаёт основу для следующих тем рабочей программы; подробные реализации дихотомии, золотого сечения и метода Фибоначчи изучаются отдельно.