

Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

«Тихоокеанский государственный университет»

МЕТОДЫ ОДНОМЕРНОЙ ОПТИМИЗАЦИИ

Методические указания и задания
к выполнению лабораторных работ по дисциплине «Методы
оптимизации»

Хабаровск
Издательство ТОГУ
2010

УДК 517.51(75.8)

ББК 22.161

Методы одномерной оптимизации : методические указания и задания к выполнению лабораторных работ по дисциплине «Методы оптимизации»/ сост. Т. М. Попова. – Хабаровск : Изд-во Тихоокеан. гос. ун-та, 2011. – 26 с.

Методические указания составлены на кафедре прикладной математики и посвящены одному из важнейших направлений подготовки выпускника университета — математической теории оптимизации. Рассмотрены теоретические, вычислительные и прикладные аспекты методов конечномерной оптимизации. Много внимания уделено описанию алгоритмов численного решения задач безусловной минимизации функций нескольких переменных. Приведены задания для лабораторной работы по одномерной оптимизации. Данное пособие будет способствовать выработке у студентов практических навыков применения методов оптимизации. Для студентов специальности Прикладная математика и других, изучаю методы оптимизации.

Печатается в соответствии с решениями кафедры «Прикладная математика» и методического совета факультета математического моделирования и процессов управления.

© Тихоокеанский государственный
университет, 2011

В своей жизни человек часто сталкивается с ситуацией, когда ему из некоторой совокупности возможных вариантов своего поведения или принятия решения в какой-либо области деятельности необходимо выбрать

один. Выбор осуществляется путем сравнения различных вариантов при помощи некоторой количественной их оценки. В этом случае говорят о необходимости решения задачи оптимизации. *Оптимизация* (от латинского слова «*optimus*» – наилучший) – поиск наилучшего варианта, при наличии множества альтернативных. По содержанию задачи оптимизации весьма разнообразны. Они могут быть связаны с проектированием технических устройств и технологических процессов, с распределением ограниченных ресурсов и планированием работы предприятий, наконец, с решением проблем, возникающих в повседневной жизни человека. Всевозможные устройства, процессы и ситуации, применительно к которым предстоит решать задачу оптимизации, называют *объектом оптимизации*.

1. Постановка задачи оптимизации

Имеется задача, для ее решения нужно формализовать объект и представить его в виде математической модели. *Математическая модель* – модель, которая определена с помощью математических формализмов. Математическая модель не является точной, а является идеализацией.

Для применения теории оптимизации к решению конкретных задачи нужно выполнить определённую последовательность действий, которая называется постановкой задачи оптимизации. Она включает этапы.

Этап I. Установление границ подлежащей оптимизации системы. *Система* – некая изолированная часть внешнего мира. Границы системы задают пределы, отделяющие её от внешнего мира. При этом предполагается, что взаимосвязи с внешним миром зафиксированы. Первоначальный выбор границ системы может оказаться слишком жёстким. Для получения адекватного решения нужно включить в систему дополнительные подсистемы, однако это ведёт к увеличению размерности задачи. Следует стремиться к представлению системы в виде изолированных подсистем, которые можно рассматривать независимо от других.

Этап II. Выбор количественного критерия, позволяющего выявить наилучший вариант, называемого характеристическим критерием. Критерии могут быть, в зависимости от конкретной задачи, экономического или технологического характера (минимальная стоимость, максимальный крутящий момент). Независимо от того, какой критерий принят в качестве характеристического, он должен принимать максимальное (или минимальное) значение для наилучшего варианта. Критериев может быть много, тогда задача становится многокритериальной. Существуют методы решения многокритериальных задач, но можно привести многокритериальную задачу к однокритериальной. Для этого один из критериев выбирается в качестве первичного, а остальные становятся вторичными. Первичный критерий

используется как характеристический, а вторичные формируют ограничения задачи.

Этап III: определение внутрисистемных переменных, через которые выражается характеристический критерий. Выбор переменных осуществляется с учётом следующих рекомендаций. Нужно разделить переменные, которые меняются в широком диапазоне и переменные, которые фиксированы или меняются слабо. Первые определяются, как независимые переменные, а вторые – как параметры задачи. Параметры задачи разделяют на фиксированные и те, которые испытывают флуктуации под воздействием внешней среды.

Нужно выбрать только те переменные, которые оказывают наибольшее влияние на характеристический критерий.

Этап IV. Построение модели, которая описывает взаимосвязь внутрисистемных переменных. Модель системы описывает взаимосвязь между переменными и отражает степень влияния этих переменных на характеристический критерий. Модель включает в себя основные уравнения материальных и энергетических балансов; уравнения, описывающие физические процессы в системе; неравенства, определяющие области допустимых значений переменных.

Таким образом, задача в виде, пригодном для решения методом оптимизации состоит в минимизации (максимизации) вещественнозначной функции $f(x)$ N -мерного аргумента x , компоненты которого удовлетворяют системе ограничений в виде уравнений $H_k(x) = 0$, $k = 1, 2, \dots, m$ или неравенств $g_j(x) \geq 0$, $j = m + 1, \dots, s$. Такая задача называется *задачей условной оптимизации*. Если задача не содержит ограничения и рассматривается на всем пространстве, то это задача *безусловной оптимизации*.

Задачи оптимизации классифицируются в соответствии с видом функций $f(x)$, $H_k(x)$, $g_j(x)$, и размерностью вектора x .

Задачи без ограничений с $N = 1$ называются задачами *одномерной оптимизации*, с $N \geq 2$ – *многомерной оптимизации*.

Если в задаче функции $H_k(x)$, $g_j(x)$ линейны, то это задача с *линейными ограничениями*. При этом целевая функция $f(x)$ может быть как линейной, так и нелинейной. Задача условной оптимизации, в которой все функции линейны, называется *задачей линейного программирования*. Задачи с нелинейной целевой функцией называются задачами *нелинейного программирования*. При этом если $f(x)$ квадратичная функция, то задачей квадратичного программирования. Если $f(x)$ отношение линейных функций, то рассматривается задача дробно-линейного программирования.

В соответствии с классификацией задач оптимизации классифицируются и методы оптимизации

2. Методы одномерной оптимизации

В некоторых случаях ограничения в задаче оптимизации позволяют через один из параметров выразить остальные и исключить их из целевой функции. В результате данных действий, задача будет сведена к поиску наибольшего или наименьшего значения скалярной действительной функции $f(x)$, выражающей критерий оптимальности. Выбирая тот или иной знак перед этой функцией, всегда можно ограничиться лишь поиском ее наименьшего значения в области определения $D(f)$, заданной с учетом ограничений на параметр оптимизации x . Для краткости будем говорить об одномерной минимизации, имея в виду нахождение наименьшего значения функции $f(x)$ на множестве $D(f)$ и точек, в которых это значение достигается. Изучение методов одномерной минимизации важно не только для решения задачи $\min_{x \in [a,b]} \{z = f(x)\}$, имеющей самостоятельное значение. Эти методы являются

также существенной составной частью методов многомерной минимизации, при помощи которых находят наименьшее значение действительных функций многих переменных, так как целый ряд методов нелинейного программирования включает в себя в качестве составной части решение задач одномерной оптимизации.

Методы одномерной оптимизации разделяются на подклассы по следующим принципам:

- использование в процессе поиска экстремума информации о самой функции, так как в ряде задач целевая функция задана таким образом, что точных значений производных найти нельзя (только оценить).
- использование в процессе поиска экстремума информации о самой функции или ее производных.
- по виду целевой функции (методы решения одно- и многоэкстремальных задач).

Введем некоторые определения.

Монотонность функции. Функция $f(x)$ является монотонной на интервале, если для любых x_1 и x_2 из этого интервала, таких, что $x_1 < x_2$ выполняется неравенство $f(x_1) < f(x_2)$, если функция монотонно возрастающая или $f(x_1) > f(x_2)$, если функция монотонно убывающая.

Унимодальность. Функция $f(x)$ является унимодальной на отрезке, если она монотонна по обе стороны от единственной на отрезке точки x_0 , то есть

функция $f(x)$ в полуинтервале $[a, x_0)$ убывает, а в полуинтервале $(x_0, b]$ возрастает. Примеры графиков унимодальных функций приведены на рис. 1. Точка x_0 может быть внутренней точкой отрезка $[a, b]$ или совпадать с одним из его концов. Унимодальная функция не обязательно непрерывна на отрезке $[a, b]$.

Определение глобального минимума Функция $f(x)$, определённая на множестве D достигает глобального минимума в точке $x^* \in D$, если $f(x^*) < f(x)$ для всех $x \in D$.

Определение локального минимума. Функция $f(x)$, определённая на множестве D имеет локальный минимум в точке $x^* \in D$, если существует такая ε -окрестность точки x^* , что для всех x из этой ε -окрестности $f(x^*) < f(x)$.

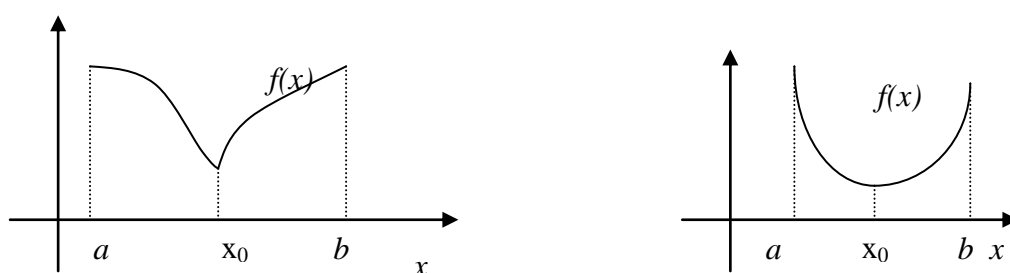


Рис. 1

Если функция $f(x)$ не унимодальна, то наименьший из локальных минимумов будет глобальным минимумом (аналогично – наибольший из локальных максимумов будет глобальным максимумом). Для нахождения аналитического оптимума необходимо найти стационарные или критические точки функции, использовать какое-либо из достаточных условий экстремума (изменение знака первой производной или определенного знака четной производной) и сравнить значения функции в точках локальных экстремумов и граничных значениях.

Однако в прикладных задачах нередко ситуации, когда трудно вычислить производные функции (например, если функция не задана в аналитическом виде). Более того, не исключено, что значения функции известны или могут быть вычислены только в отдельных точках. В таких ситуациях использование необходимого и достаточного условий локального минимума невозможно и следует применять другие методы решения задачи

оптимизации. Методы минимизации функции одного переменного, в которых используют значения функции в точках рассматриваемого промежутка и не используют значения ее производных, называют *методами прямого поиска*. Можно выделить две группы методов прямого поиска, соответствующие двум принципиально различным ситуациям:

- 1) все N точек x_k , $k = 1, \dots, N$, в которых будут вычислены значения функции, выбирают заранее (до вычисления функции в этих точках);
- 2) точки x_k выбирают последовательно (для выбора последующей точки используют значения функции, вычисленные в предыдущих точках).

В первом случае поиск наименьшего значения называют *пассивным*, а во втором — *последовательным*.

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

Будем считать, что стратегия поиска определена, если:

- определен алгоритм выбора точек x_k , $k = 1, \dots, N$;
- определено условие прекращения поиска, т.е. условие, при выполнении которого значение $f(x^*)$ считают найденным с заданной точностью.

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

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

При этом длина интервала, содержащего минимум $l_N = \frac{2(b-a)}{N+1}$ 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 $x^* \in (x_{j-1}, x_{j+1})$ и можно считать, что $x^* \cong x_j \pm \varepsilon$ с точностью ε .

3. Методы последовательного поиска (методы интервалов)

Предположение об унимодальности функции позволяет разработать адаптивные (или последовательные) алгоритмы поиска оптимального значения). При этом используется понятие *интервала неопределенности*, как интервала содержащего точку минимума, хотя ее точное положение неизвестно. Важной характеристикой данных методов является количество вычислений значений оптимизируемой функции для получения конечной величины интервала неопределенности, не превышающей заданной

величины. Методы ориентированы на нахождение точки оптимума внутри заданного интервала и основаны на свойстве унимодальности функции и не используют информацию о производной функции.

В алгоритмах этих методов вычисляются значения функции в промежуточных точках λ_k и μ_k ($\lambda_k < \mu_k$) интервала неопределенности:

если $f(\lambda_k) > f(\mu_k)$, то в качестве границ нового интервала рассматривается интервал $[\lambda_k, b_k]$ (рис. 2)

если $f(\lambda_k) < f(\mu_k)$ это будет интервал $[a_k, \mu_k]$ (рис. 3) если $f(\lambda_k) = f(\mu_k)$, то оставим интервал $[\lambda_k, \mu_k]$

Все алгоритмы различаются только способом определения промежуточных точек λ_k и μ_k .

3.1. Метод дихотомии. Идея метода состоит в вычислении на каждой очередной итерации двух значений целевой функции в точках, отстоящих на величину α в обе стороны от середины интервала неопределенности. Величина α в этом методе называется *константой различимости*, такова, что, с одной стороны, величина 2α была близка к желаемому конечному значению интервала неопределенности, с другой, значения оптимизируемой функции на краях интервала 2α были различимы.

Введем обозначения: ε – конечная длина интервала, $[a_1, b_1]$ – начальный интервал неопределенности, k – номер итерации.

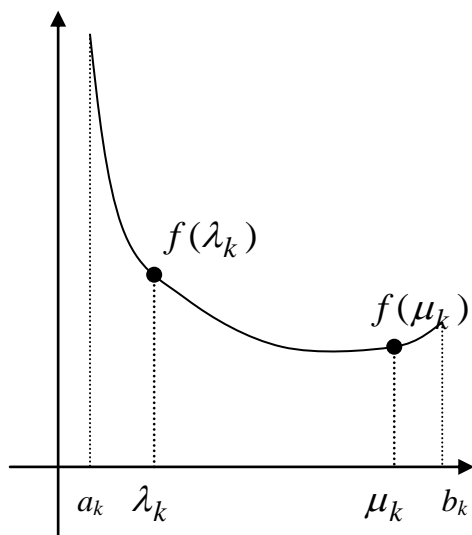


Рис. 2

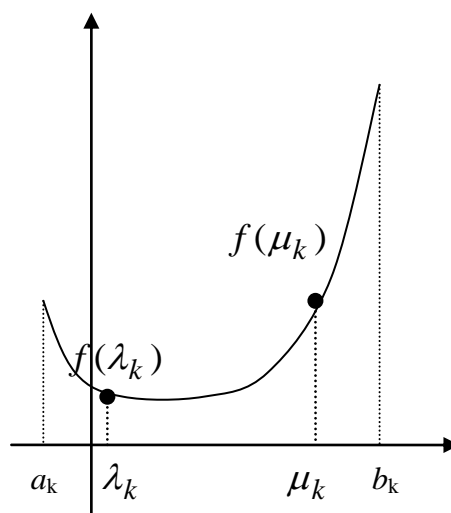


Рис. 3

АЛГОРИТМ (рис. 4)

Шаг 0. Задать $\alpha > 0$, $\varepsilon > 0$, $[a_1, b_1]$, $k=1$

Шаг 1. Если $b_k - a_k \leq \varepsilon$, то $x^* \in [a_k, b_k]$ и $x^* = \frac{a_k + b_k}{2}$ конец.

$$\text{Иначе } \lambda_k = \frac{a_k + b_k}{2} - \alpha, \quad \mu_k = \frac{a_k + b_k}{2} + \alpha$$

Шаг 2 Вычислить $f(\lambda_k), f(\mu_k)$ если $f(\lambda_k) > f(\mu_k)$,

то $a_{k+1} = \lambda_k, b_{k+1} = b_k$, иначе $a_{k+1} = a_k$ и $b_{k+1} = \mu_k$.

Шаг 3 $k := k + 1$ переходим на шаг 1.

Длина интервала неопределенности после k -ой итерации

$$|b_{k+1} - a_{k+1}| = \frac{1}{2^k} (b_1 - a_1) + 2\alpha \left(1 - \frac{1}{2^k} \right).$$

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

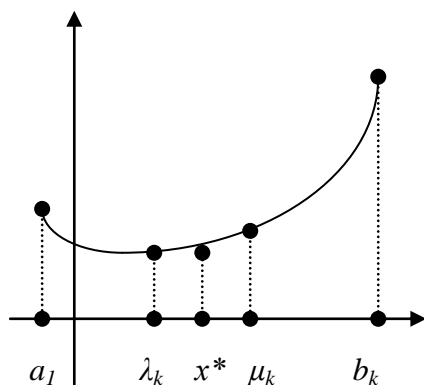


Рис. 4

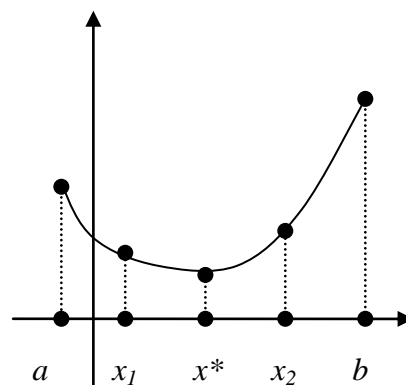


Рис. 5

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

АЛГОРИТМ (рис. 5)

Шаг 0. Зададим точность $\varepsilon > 0$

Шаг 1. Найти $x^* = \frac{a+b}{2}$ 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

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

Шаг 4. Если $f(x_2) < f(x^*)$, то исключается интервал (a, x^*) , при этом

$a = x^*$, $x^* = x_2$; перейти к п. 5. Иначе исключить интервалы (a, x_1) , (x_2, b) , то есть $a = x_1$, $b = x_2$; перейти к п. 4.

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

Средняя точка последовательности получаемых интервалов всегда совпадает с одной из пробных точек x_1, x_2 , или x^* , найденных на предыдущих итерациях. На каждой итерации требуется не более 2-х вычислений значений функции. После N вычислений длина интервала равна $l = (1/2)^{1/N}$ длины исходного интервала

3.3. Метод золотого сечения. Идея метода состоит в использовании на каждой итерации для сокращения интервала неопределенности одной из внутренних точек предыдущей итерации. Должны быть выполнены условия:

- Пробные точки на каждой итерации находятся на одинаковых расстояниях от концов интервала неопределенности $\lambda_k - a_k = b_k - \mu_k$
- Для новой итерации точки λ_{k+1} , μ_{k+1} выбираются так, чтобы λ_{k+1} совпало с μ_k либо μ_{k+1} совпало с λ_k .
- Сжатие интервала неопределенности осуществляется на каждой итерации с одним и тем же коэффициентом сжатия τ , удовлетворяющее уравнению

$$\tau^2 + \tau - 1 = 0. \text{ Золотое сечение можно вычислить как } \tau = \frac{\sqrt{5} - 1}{2} = 0,61803...$$

- При выполнении этих условий λ_k и μ_k вычисляются по формулам

$$\begin{aligned} \lambda_k &= a_k + (1 - \tau)(b_k - a_k), \\ \mu_k &= a_k + \tau(b_k - a_k), \end{aligned} \quad (1)$$

здесь $\tau = \frac{b_{k+1} - a_{k+1}}{b_k - a_k}$, определяется из условий выше.

АЛГОРИТМ

Шаг 0. Задать $\varepsilon > 0$, $[a_1, b_1]$, $k=1$. Вычислить λ_1, μ_1 по формулам (1),

$$f(\lambda_1), f(\mu_1),$$

Шаг 1. если $b_k - a_k \leq \varepsilon$, то $x^* \in [a_k, b_k]$ и $x^* = \frac{a_k + b_k}{2}$ конец.

Иначе: если $f(\lambda_k) > f(\mu_k)$, то перейти на шаг 2, если $f(\lambda_k) \leq f(\mu_k)$, то перейти на шаг 3.

Шаг 2. Положить $a_{k+1} = \lambda_k$, $b_{k+1} = b_k$, $\lambda_{k+1} = \mu_k$,

$\mu_{k+1} = a_{k+1} + \tau(b_{k+1} - a_{k+1})$, вычислить $f(\mu_{k+1})$, перейти на шаг 4.

Шаг 3 Положить $a_{k+1} = a_k$, $b_{k+1} = \mu_k$, $\mu_{k+1} = \lambda_k$,

$\lambda_{k+1} = a_{k+1} - (1 - \tau)(b_{k+1} - a_{k+1})$, вычислить $f(\mu_{k+1})$, перейти на шаг 4.

Шаг 4. $k := k + 1$ переходим на шаг 1.

Если исходный интервал имеет единичную длину, длина интервала после N вычислений равна τ^{N-1} , иначе $\tau^{n-1} \leq \frac{\varepsilon}{b_1 - a_1}$. Для достижения точности ε

потребуется $n \geq \frac{\ln \frac{b_1 - a_1}{\varepsilon}}{\ln \frac{\sqrt{5} - 1}{2}}$ итераций.

3.4. Метод Фибоначчи. Метод аналогичен методу золотого сечения. Отличие состоит в том, что коэффициент сжатия интервала неопределенности меняется от итерации к итерации согласно последовательности Фибоначчи. Последовательность чисел определяется следующим образом:

$$F_0 = F_1 = 1, \quad F_{k+1} = F_k + F_{k-1}, \quad k = 1, 2, \dots$$

Пробные точки λ_k и μ_k вычисляются по формулам

$$\begin{aligned} \lambda_k &= a_k + \frac{F_{n-k-1}}{F_{n-k+1}}(b_k - a_k), \\ \mu_k &= a_k + \frac{F_{n-k}}{F_{n-k+1}}(b_k - a_k). \end{aligned} \tag{2}$$

При этом число итераций выбирается до начала вычислений и обусловлено требуемой точностью $\varepsilon = \frac{b_1 - a_1}{F_n}$.

АЛГОРИТМ

Шаг 0. Задать ε , $\alpha > 0$, $[a_1, b_1]$, $k=1$. Вычислить λ_1, μ_1 по формулам (2),

$$f(\lambda_1), f(\mu_1).$$

Шаг 1. Если $f(\lambda_k) > f(\mu_k)$, то перейти на шаг 2, если $f(\lambda_k) \leq f(\mu_k)$, то перейти на шаг 3.

Шаг 2. Положить $a_{k+1} = \lambda_k$, $b_{k+1} = b_k$, $\lambda_{k+1} = \mu_k$,

$$\mu_{k+1} = a_{k+1} + \frac{F_{n-k-1}}{F_{n-k}}(b_{k+1} - a_{k+1}).$$

Если $k = n - 2$, то перейти на шаг 5, иначе вычислить $f(\mu_{k+1})$,

перейти на шаг 4.

Шаг 3. Положить $a_{k+1} = a_k$, $b_{k+1} = \mu_k$, $\mu_{k+1} = \lambda_k$,

$$\lambda_{k+1} = a_{k+1} + \frac{F_{n-k-2}}{F_{n-k}} (b_{k+1} - a_{k+1}),$$

если $k = n - 2$, то перейти на шаг 5, иначе вычислить $f(\lambda_{k+1})$,

перейти на шаг 4.

Шаг 4. $k = k + 1$ перейти на шаг 1.

Шаг 5. Положить $\lambda_n = \lambda_{n-1}$, $\mu_n = \lambda_n + \alpha$, вычислить $f(\lambda_n)$ и $f(\mu_n)$.

Если $f(\lambda_n) > f(\mu_n)$, то $a_n = \lambda_n$, $b_n = b_{n-1}$, иначе $a_n = a_{n-1}$, $b_n = \lambda_n$.

Оптимальное решение $x^* \in [a_n, b_n]$ и $x^* = \frac{a_n + b_n}{2}$.

4. Методы аппроксимации. Метод Пауэлла

В методах прямого поиска мы не имели никакой информации о минимизируемой функции за исключением ее значений в выбранных нами точках и предположения, что она непрерывна и является унимодальной функцией на рассматриваемом отрезке. Если функцию в некоторой окрестности точки ее минимума можно достаточно точно заменить (аппроксимировать) многочленом, то для ее минимизации целесообразно использовать так называемые *методы полиномиальной аппроксимации*. Их общая особенность состоит в вычислении коэффициентов многочлена по известным значениям функции в отдельных точках и последующем нахождении минимума этого многочлена с использованием необходимых и достаточных условий экстремума. Основная идея метода: возможность аппроксимации гладкой функции полиномом достаточно высокого порядка и использование этого полинома для оценивания точки оптимума.

Качество этой оценки может быть повышено двумя способами:

- увеличением степени полинома;
- уменьшением интервала аппроксимации.

Второй способ предпочтительнее, так как построение полинома порядка более 3 – достаточно сложная задача, а сужение интервала для унимодальной функции – достаточно простая.

Использование квадратичной аппроксимации для нахождения оптимума.

Чтобы функция имела минимум внутри отрезка она должна быть, по крайней мере квадратичной. Для построения квадратичной функции достаточно трех точек: $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, $M_3(x_3, y_3)$, . Можно задать аппроксимацию функции полиномом вида: $P_2(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2)$ и

И выбрать a_0, a_1 и a_2 так, чтобы $P_2(x_1) = y_1, P_2(x_2) = y_2, P_2(x_3) = y_3$.

Отсюда следует, что

$$a_0 = y_1, \quad a_1 = \frac{y_2 - y_1}{x_2 - x_1}, \quad a_2 = \frac{1}{x_3 - x_2} \cdot \left(\frac{y_3 - y_1}{x_3 - x_1} - \frac{y_2 - y_1}{x_2 - x_1} \right) \quad (3)$$

Найдём стационарную точку x^* полинома $P_2(x)$:

$$x^* = \frac{x_2 - x_1}{2} - \frac{a_1}{2a_2} \quad (4)$$

Так как функция унимодальна на рассматриваемом интервале и полином $P_2(x)$ тоже унимодальная функция, то x^* является приемлемой оценкой истинного оптимума. *Метод Пауэлла* основан на последовательном применении процедуры оценивания с использованием квадратичной аппроксимации.

АЛГОРИТМ

Шаг 1. Задать x_1 и шаг $h > 0$, точность $\varepsilon > 0$.

Шаг 2. Найти $x_2 = x_1 + h$, вычислить $f(x_1)$ и $f(x_2)$

Шаг 3 Если $f(x_1) > f(x_2)$, то $x_3 = x_1 + 2h$, иначе $x_3 = x_1 - h$.

Шаг 4. Вычислить $f(x_3)$; определить $F_{\min} = \min\{f(x_1), f(x_2), f(x_3)\}$,
определить соответствующее x_{\min} .

Шаг 5. Найти x^* по формулам (3), (4).

Шаг 6 Если $|x^* - x_{\min}| < \varepsilon$, то поиск окончен x^* является оценкой оптимума,
иначе перейти к шагу 7.

Шаг 7. Вычислить $f(x^*)$, если $f(x^*) < F_{\min}$, то $x_1 = x^*$, иначе $x_1 = x_{\min}$.

Перейти к п. 2.

Метод квадратичной аппроксимации удобно применять после локализации точки минимума методом золотого сечения или методом Фибоначчи. Это объясняется тем, что для дважды дифференцируемой функции многочлен второго порядка достаточно хорошо аппроксимирует функцию в окрестности точки минимума.

5. Методы с использованием информации о производной функции

В методах прямого поиска при вычислении значений минимизируемой функции $f(x)$ неизбежно возникают погрешности, к которым чувствительны алгоритмы прямого поиска, основанные на сравнении значений функции в отдельных точках. Если унимодальная функция $f(x)$ непрерывно дифференцируема на отрезке минимизации, то точку x^* наименьшего

значения функции можно вычислять как корень уравнения $f'(x) = 0$ с помощью тех или иных методов численного решения нелинейных уравнений. В этом случае на точность решения задачи решающее влияние оказывает погрешность вычисления производной функции. Рассмотрим некоторые методы одномерной минимизации, основанные на использовании производной минимизируемой функции.

5.1. Метод средней точки. Будем искать минимум функции $f(x)$ непрерывно дифференцируемой и строго унимодальной на отрезке $[a, b]$. В этом случае единственной точкой $x^* \in [a, b]$ минимума будет стационарная точка, в которой $f'(x^*) = 0$. Отметим, что непрерывно дифференцируемая унимодальная на отрезке функция может иметь на нем более одной стационарной точки. На отрезке определяются две точки a_k, b_k , в которых производные имеют разные знаки, $f'(a_k)f'(b_k) < 0$. Искомый оптимум находится между ними. Делим интервал пополам: $x_k = \frac{a_k + b_k}{2}$, если $f'(x_k) > 0 (< 0)$, то из двух интервалов оставляем тот, на концах которого производная имеет разные знаки.

АЛГОРИТМ

Шаг 1. Определим точность $\varepsilon > 0$, $a_1 = a, b_1 = b$, вычисляем

$$f'(a_1) < 0, f'(b_1) > 0$$

$$\text{Вычисляем } x_1 = \frac{a_1 + b_1}{2}, k=1.$$

Шаг 2 Вычисляем $f'(x_k)$, если $f'(x_k) = 0$, или $f'(x_k) < \varepsilon$,

$$\text{то } x_k = \frac{a_k + b_k}{2} \text{ оптимум, конец.}$$

Шаг 3 если $f'(x_k) < 0$, то $a_{k+1} = x_k, b_{k+1} = b_k$, иначе $a_{k+1} = a_k, b_{k+1} = x_k$

Шаг 3 $x_k = \frac{a_k + b_k}{2}$, переход на 2.

Метод средней точки напоминает метод дихотомии, но сходится к искомому значению x^* быстрее. Поскольку для метода средней точки после вычисления n значений производной минимизируемой на отрезке $[0, 1]$ функции $f(x)$ для длины интервала неопределенности получаем $l_n = \frac{1}{2^n}$.

Таким образом, для одинакового уменьшения значения l_n в методе средней

точки нужно вычислить вдвое меньше значений производной функции по сравнению с числом значений самой функции в методе дихотомии.

5.2. Метод Ньютона (метод касательной). Если строго унимодальная на отрезке $[a, b]$ функция $f(x)$ дважды непрерывно дифференцируема на этом отрезке, то точку $x^* \in [a, b]$ минимума этой функции можно найти путем решения уравнения $f'(x) = 0$ методом Ньютона, иногда называемым методом касательных. Выбираем x_0 - начальное приближение, называемое обычно начальной точкой. Линеаризуем функцию $f'(x)$ в окрестности начальной точки, приближенно заменив дугу графика этой функции касательной в точке $(x_0, f'(x_0))$.

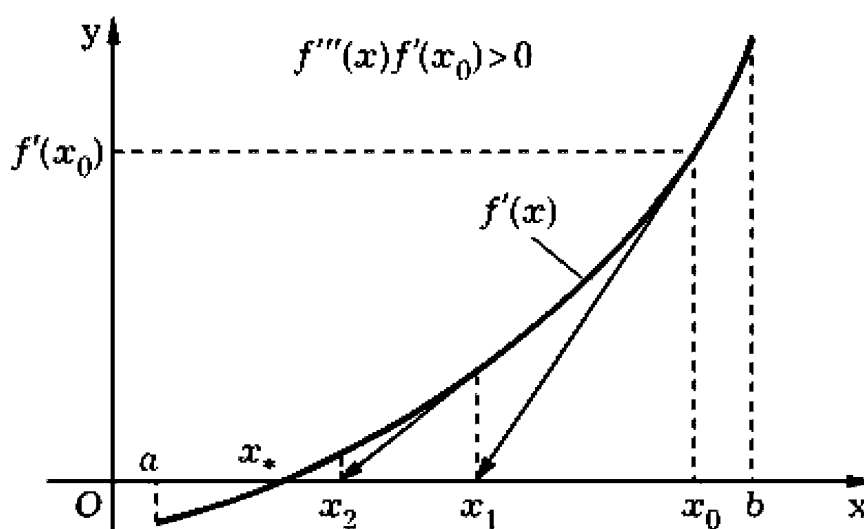


Рис. 6

Уравнение касательной имеет вид $f'(x) \approx f'(x_0) + f''(x_0) \cdot (x - x_0)$. Выберем в качестве следующего приближения к x^* точку x_1 пересечения касательной с осью абсцисс (рис. 6). Получаем первый элемент $x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)}$ итерационной последовательности $\{x_k\}$. На $(k+1)$ -м шаге по найденной на предыдущем шаге точке x_k можно найти точку

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \quad (5)$$

В общем случае сходимость метода Ньютона существенно зависит от выбора начальной точки x_0 . Для надежной работы этого метода необходимо, чтобы вторая производная $f''(x)$ в некоторой окрестности искомой точки x^* сохраняла знак, а начальная точка x_0 выбиралась из такой окрестности. В

противном случае второе слагаемое в правой части (5) может стать неограниченным. Поскольку для дважды непрерывно дифференцируемой функции в точке минимума $f''(x^*) > 0$, то должно быть и $f''(x_0) > 0$. Поэтому говорят, что метод Ньютона обладает *локальной сходимостью* в том смысле, что надо выбрать хорошее начальное приближение, попадающее в такую окрестность точки x^* , где $f''(x) > 0$. Однако проверка выполнения этого условия не всегда возможна. Достаточным условием монотонной сходимости метода Ньютона будут постоянство в интервале между точками x_0 и x^* знака производной $f''(x)$ и совпадение его со знаком $f'(x)$. Оказывается, что в этом случае метод Ньютона обладает квадратичной скоростью сходимости в некоторой окрестности точки x^* , причем

$$|x^* - x_k| \leq \frac{|x^* - x_{k-1}|^2}{C}.$$

5.3. Метод секущих похож на метод Ньютона, но строится не касательная к графику производной, а секущая. Геометрическая интерпретация этого метода (рис. 7) состоит в том, что в качестве очередного приближения x_{k+1} выбирают точку пересечения с осью абсцисс секущей к графику функции $f'(x)$, то есть

$$x_{k+1} = x_k - \frac{f'(x_k)(x_k - x_{k-1})}{(f'(x_k) - f'(x_{k-1}))} \quad (6)$$

Выбор начальной точки x_0 проводят следующим образом. Если на отрезке $[a, b]$ функция $f(x)$ имеет знакопостоянную третью производную $f'''(x)$, то в качестве x_0 выбирают тот конец отрезка $[a, b]$, на котором совпадают знаки $f'(x)$ и $f'''(x)$, тогда $x_1 = \frac{af'(b) - bf'(a)}{f'(b) - f'(a)}$. Точка x_1 – точка пересечения с осью

абсцисс хорды, стягивающей дугу графика функции $f'(x)$ на отрезке $[a, b]$ (рис 7). Таким образом, первый шаг метода секущих выполняют согласно *методу хорд*, а последующие шаги — в соответствии (6).

Этот метод имеет *сверхлинейную скорость* сходимости, причем $|x^* - x_k| \leq C|x^* - x_{k-1}|^\tau$, где $C - \text{const}$, а $\tau = \frac{1 + \sqrt{5}}{2} = 1,618...$ — отношение золотого сечения.

Модификации метода Ньютона обладают только локальной сходимостью, т.е. сходятся, если начальная точка выбрана в некоторой окрестности точки минимума функции. Если же начальная точка расположена слишком далеко от точки минимума, то подобный метод может расходиться или «зацикливаться». В отличие от метода средней точки метод

секущих использует информацию не только о знаке производной, но и о значениях в пробных точках.

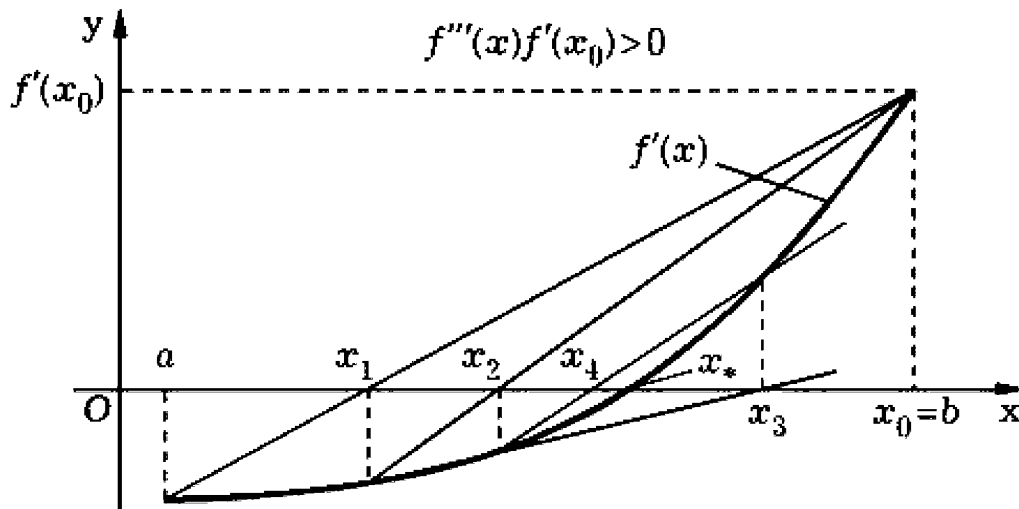


Рис. 7

5.4. Метод кубической аппроксимации. В методах полиномиальной аппроксимации при построении многочлена, аппроксимирующего минимизируемую функцию в окрестности искомой точки x^* минимума, помимо значений функции в отдельных точках могут быть использованы и значения ее производных.

Пусть для непрерывно дифференцируемой функции $f(x)$, строго выпуклой на отрезке $[x_1, x_2]$ известны значения $y_1 = f(x_1)$, $y_2 = f(x_2)$ функции и значения ее производной $y_{11} = f'(x_1)$, $y_{12} = f'(x_2)$. Если $y_{11} \cdot y_{12} < 0$, то рассматриваемая функция строго унимодальна на этом отрезке.

Рассмотрим метод поиска точки x^* при условии $y_{11} \cdot y_{12} < 0$, называемый методом кубической аппроксимации, поскольку в этом случае на отрезке $[x_1, x_2]$ можно построить единственный многочлен третьей степени, располагая значениями функции и производных на концах этого отрезка. Этот многочлен, называемый *кубическим интерполяционным многочленом Эрмита*, можно преобразовать к виду

$$P_3(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2) + a_3(x - x_1)^2(x - x_2), \quad (7)$$

$$a_0 = y_1, \quad a_1 = \frac{y_2 - y_1}{x_2 - x_1}, \quad a_2 = \frac{y_2 - y_1}{(x_2 - x_1)^2} - \frac{y_{11}}{x_2 - x_1}, \quad (8)$$

$$a_3 = \frac{y_{12} - y_{11}}{(x_2 - x_1)^2} - 2 \frac{y_2 - y_1}{(x_2 - x_1)^3} \quad (9)$$

Из необходимого условия $P_3(x)=0$ экстремума этого многочлена с учетом коэффициентов (8), (9) получаем квадратное уравнение

$$3a_3(x-x_1)^2 - 2 \frac{y_{11} + y_{12} - 3a_1}{x_2 - x_1} (x-x_1) + y_{11} = 0,$$

решение которого представим в виде

$$x_0 = x_1 + \mu(x_2 - x_1), \quad (10)$$

$$\text{где } \mu = \frac{\omega + z - y_{11}}{2\omega - y_{11} + y_{12}}, \quad z = y_{11} + y_{12} - 3 \frac{y_2 - y_1}{x_2 - x_1}, \quad \omega = \sqrt{z^2 - y_{11} \cdot y_{12}}.$$

Легко показать, что $\mu \in (0, 1)$ и $x_0 \in (x_1, x_2)$.

Если $f'(x_0)=0$, то $x_0 = x^*$ — искомая точка минимума функции $f(x)$ на отрезке $[x_1, x_2]$. Если же $f'(x_0) \neq 0$ то оставляем меньший отрезок $[x_0, x_2]$ или $[x_1, x_0]$, такой чтобы $f'(x_0)y_{1j} < 0, j=1,2$ и продолжать описанным выше способом поиск точки минимума на новом отрезке. После каждого приближения правильность вычислений подтверждается уменьшением минимального значения многочлена по сравнению с его минимальным значением на предыдущем шаге. Вычисления можно прекратить, когда длина интервала неопределенности, в котором гарантированно находится искомая точка x^* , станет меньше заданной наибольшей допустимой величины ε .

Алгоритм

Шаг 0. Задать $\varepsilon > 0$, найти точки x_1, x_2 так чтобы $f'(x_1) \cdot f'(x_2) < 0$.

Шаг 1. Вычислить $y_1 = f(x_1), y_2 = f(x_2)$ и $y_{11} = f'(x_1), y_{12} = f'(x_2)$.

Определить коэффициенты многочлена $P_3(x)$ по формулам (8), (9).

Шаг 2. Найти x_0 по формулам (10). Вычислить $f'(x_0)$, если $|f'(x_0)| < \varepsilon$,

то $x_0 = x^*$ — искомая точка минимума, конец, иначе шаг 3.

Шаг 3. Если $f'(x_0) < 0$, то $x_1 = x_0$, если $f'(x_0) > 0$, то $x_2 = x_0$. Переход на шаг 1

6. Сравнение методов

Для быстрого получения предварительных результатов (начальной точки для применения других методов), а также, если требуется надёжная работа алгоритма при неизвестной заранее целевой функции, лучше использовать методы исключения интервалов.

Если требуется точное решение, необходимо воспользоваться градиентными методами (особенно кубической аппроксимацией).

С другой стороны, если требуются высокая точность, но функция не задана аналитически, лучше пользоваться методами точечного оценивания, так как при использовании градиентных методов накапливается погрешность при конечно-разностной аппроксимации производных.

Если сравнить методы с точки зрения поставленной задачи и вида функции, то при минимуме информации о функции следует использовать метод исключения интервалов.

Если функция квадратичная или близка к таковой, то следует использовать метод Пауэлла

Если функция дважды дифференцируемая, непрерывная и задана аналитически, то следует использовать градиентные методы.

Методы точечного оценивания при прочих равных условиях (интервалы, гладкая функция) быстрее методов исключения интервалов

Задание для лабораторной работы

Лабораторная работа № 1 Методы одномерного поиска

Цель работы

Ознакомиться с методами одномерного поиска. Сравнить различные алгоритмы по эффективности на тестовых примерах.

Порядок выполнения работы

1. Найти аналитическое решение задачи $\min_{x \in [a, b]} f(x)$.
2. Реализовать методы с различной точностью.
Варианты 1-4 методы пассивного поиска, золотого сечения, метод Ньютона.
Варианты 5-8 методы дихотомии, Фибоначчи, секущих.
Варианты 9-12 методы деления пополам, Пауэлла, Ньютона.
Варианты 13-15 методы золотого сечения, Пауэлла, использование кубической аппроксимации.
3. Исследовать их сходимость и провести сравнение по числу вычислений функции для достижения заданной точности.

Варианты тестовых заданий

1. $f(x) = \sin(x)$, $x \in [-\pi, \pi/2]$	2. $f(x) = \cos(x)$, $x \in [0, \pi]$.
3. $f(x) = (x-2)^2$, $x \in [0, 3]$	4. $f(x) = (x-15)^2 + 5$, $x \in [12, 20]$
5. $f(x) = (x+5)^4$, $x \in [-6, 2]$	6. $f(x) = xe^x$, $x \in [-2, 0]$,

7. $f(x) = x^2 + 2x - 4, x \in [-2, 1]$	8. $f(x) = x^3 - x, x \in [0, 1]$
9. $f(x) = x^5 - x^2, x \in [0, 1]$	10. $f(x) = -x/e^x, x \in [0, 3]$
11. $f(x) = x^4 - x, x \in [0, 1]$	12. $f(x) = x^4/\ln x, x \in [1.1, 1.5]$
13. $f(x) = xe^{-x}, x \in [-2, 6]$	14. $f(x) = xe^{-2x}, x \in [-2, 6]$

Содержание отчета

- титульный лист;
- цель работы; задание;
- таблицы с результатами исследований по каждому методу, где должны быть отражены границы и длины интервалов на каждой итерации,
- соотношение длины интервала на $k-1$ итерации к длине интервала на k итерации;
- график зависимости количества вычислений целевой функции от логарифма задаваемой точности ε ;
- выводы по всем пунктам задания.

Библиографический список

1. *Аттетков А.В.* Методы оптимизации: учеб. для вузов / *А.В. Аттетков*, СВ. Галкин, В.С. Зарубин ; под ред. В.С. Зарубина, А.П. Крищенко. - 2-е изд., стереотип. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2003. -440 с. (Сер. Математика в техническом университете; Вып. XIV).
2. *Алексеев В.М.* Сборник задач по оптимизации. / *В.М. Алексеев*, Э.М. Галлеев, В.М. Тихомиров. – М. : Наука, 1984. 287 с.
3. *Ашманов С.А.*, Теория оптимизации в задачах и упражнениях. / *С.А. Ашманов*, А.В. Тимохов. – М. : Наука, 1991. 448 с.
4. *Лесин В.В.*, Основы методов оптимизации: учебник / *Лесин В.В.*, Лисовец Ю.П. М.: Изд-во МАИ, 1995. 341 с.
5. *Глебов Н. И.* Методы оптимизации: учебное пособие / *Глебов Н. И.*, Кочетов Ю. А, Плясунов А. В. Из-во Новосибирского ун-та, Новосибирск, 2000. 105 с.

МЕТОДЫ ОДНОМЕРНОЙ ОПТИМИЗАЦИИ

Методические указания и задания
к выполнению лабораторных работ по дисциплине «Методы оптимизации»

Попова Татьяна Михайловна

Главный редактор Л. А. Суевалова
Редактор Л. С. Бакаева

Подписано в печать Формат 60x84 1/16.
Бумага писчая. Гарнитура «Таймс». Печать цифровая.
Усл. печ. л. Тираж 200 экз. Заказ

Издательство Тихоокеанского государственного университета.
680035, Хабаровск, ул. Тихоокеанская, 136.

Отдел оперативной полиграфии издательства
Тихоокеанского государственного университета.
680035, Хабаровск, ул. Тихоокеанская, 136.