Министерство образования и науки российской федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«**ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ**» **(ТУСУР)**

Утверждаю
Зав. отделением каф. ЮНЕСКО
_________________ Ю.М. Осипов
"_____"_________________ 2012 г.

МЕТОДЫ И ТЕОРИЯ ОПТИМИЗАЦИИ

Методические указания к практическим занятиям
по дисциплине «**Методы и теория оптимизации**» для магистрантов 6 курса, обучающихся по направлению 221000.68 "Мехатроника и робототехника" по магистерской программе "Проектирование и исследование мультикоординатных электромехатронных систем движения"

Томск 2012

УДК 621.396.6.671.7

Методы и теория оптимизации: Методические указания к практическим занятиям по дисциплине «Методы и теория оптимизации» для магистрантов 6 курса, обучающихся по направлению 221000.68 "Мехатроника и робототехника" по магистерской программе "Проектирование и исследование мультикоординатных электромехатронных систем движения". – Томск: Изд-во ТУСУР, 2012. – 39 с.

Методические указания рассмотрены и рекомендованы к изданию методическим семинаром отделения кафедры ЮНЕСКО «27 » марта 2012 г.

Составитель к.т.н., доц. _____________С.В. Щербинин

Зав. кафедрой ОКЮ
доктор техн. наук,
доктор экон. наук,
профессор _______________ *Ю.М. Осипов*

Рецензент
Кандидат технических наук,
доцент кафедры МИГ ЮТИ ТПУ
И.Ф. Боровиков

ВВЕДЕНИЕ

Оптимизация – от латинского слова «оптимус» - наилучший – поиск наилучшего, поиск наилучшего проектного изделия.

Процесс оптимизации.

Имеется задача. Для решения задачи нужно формализовать объект и представить его в виде математической модели.

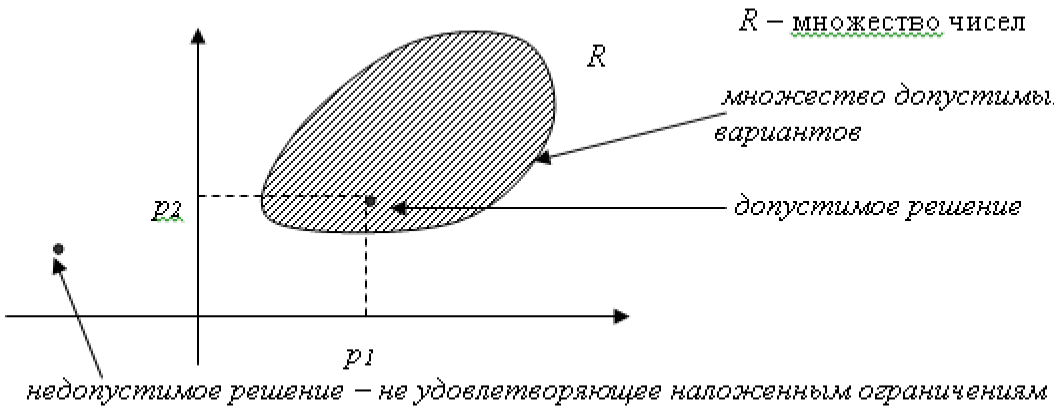
Модели:

- физические;
- геометрические (фотография, рисунок);
- математические.

Математическая модель, та которая определена с помощью математических формализмов. Математическая модель не является точной, а является идеализацией. Модель характеризуется параметрами, которые могут быть и числовыми $p_1, p_2, \ldots, p_n$. Их часть может характеризовать состояние объекта – параметры состояния (p_1, p_2), а другие могут относиться к процессу проектирования – переменные проектирования (p_n).

Определение параметров состояния - задача моделирования. Определение переменных проектирования – задачи проектирования или задачи оптимизации.

Допустим имеются 2 переменные p_1, p_2. Задавая конкретные значения получаем точку.



Плоскость множества возможных вариантов, на нее могут быть наложены ограничения.

Отображение множества $F : G \to R$ - целевая функция позволяет формировать критерий для сравнения различных решений.

2 вида задач оптимизации:

- максимизации;
- минимизации.

Для оптимизационного решения задачи требуется:

1. Сформулировать задачу;
2. Построить математическую модель (определить множество переменных);
3. Определить ограничения на возможные решения;
4. Определить целевую функцию. Далее применим формальные математические методы, позволяющие найти решения.

Занятие № 1. ОБЩАЯ ПОСТАНОВКА ЗАДАЧИ ОПТИМИЗАЦИИ И ОСНОВНЫЕ ПОЛОЖЕНИЯ

Постановка: требуется оптимизировать x (формальная постановка)

$$\begin{cases} Z = f(x) \longrightarrow opt \\ a \le x \le b \end{cases}$$

$f(x)$ - функция одной переменной $a, b, x, f(x) \in R$

$f(x)$ - целевая функция.

Решение: найти x, при котором $f(x)$ принимает оптимальное значение.

2 варианта:

- минимизировать – задача минимизации;
- максимизировать – задача максимизации.

Рассмотрим случай минимизации

$$\begin{cases} Z = f(x) \longrightarrow opt \\ a \le x \le b \end{cases} \qquad a, b, x, f(x) \in R$$

2 способа:

- аналитический
- численный

В аналитическом $f(x)$ задается в виде формулы, в численном $f(x)$ задается в виде черного ящика, на входе подается x, на выходе значение целевой функции в этой точке.

Пусть функция определена в некоторой области S ($x \in S$), в случае одномерной оптимизации S – интервал $S = \{x \mid a \le x \le b\}$:

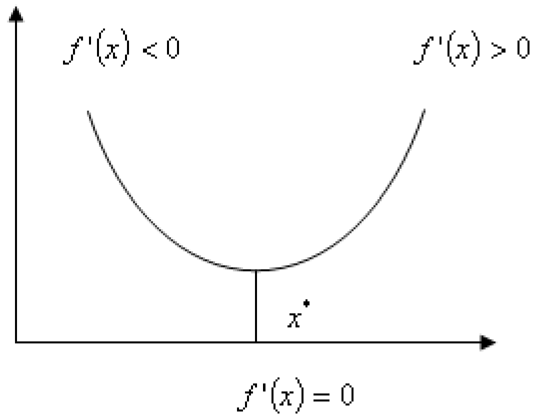
1. точка x^* называется глобальным минимумом, если для $\forall x \in S, f(x^*) \le f(x)$
2. точка x^* называется строгим глобальным минимумом, если для $\forall x \in S, f(x^*) < f(x)$
3. точка x^* называется локальным минимумом, если для $\forall x \in \mathrm{E}(x^*), f(x^*) \le f(x)$
4. точка x^* называется строгим локальным минимумом, если для $\forall x \in \mathrm{E}(x^*), f(x^*) < f(x)$

Следствие: любая точка глобального минимума является локальным минимумом, обратное не верно.

Аналитический способ нахождения локального минимума.

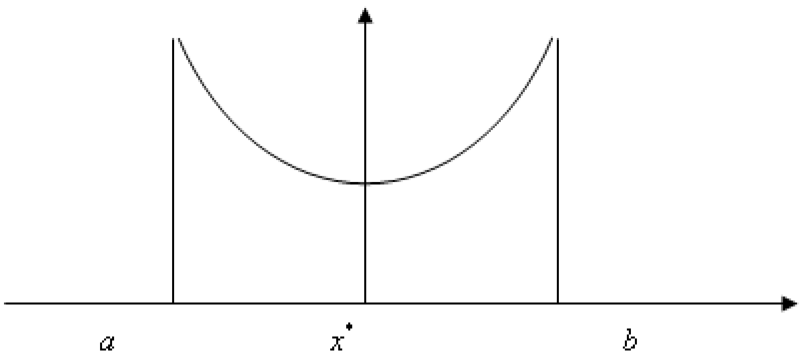
$f(x)$ - дифференцируема

$f'(x) = 0$ - необходимое условие точки локального минимума.



Численные методы.

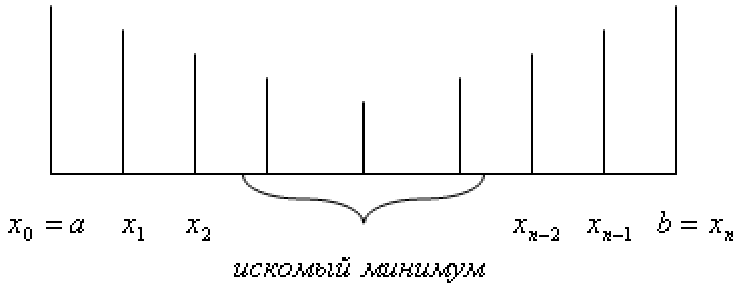
Пусть функция $f(x)$ задана на интервале (a,b), при этом существует такая точка x^*, что на $[a,x^*]$ – монотонно убывает, а на $[x^*,b]$ – монотонно возрастает, то функция унимодальная.



Если из того что $x_1 \le x_2$ следует, что $f(x_1) \le f(x_2)$, то функция называется монотонно возрастающей. Если из того что $x_2 \le x_1$ следует, что $f(x_2) \le f(x_1)$, то функция называется монотонно убывающей.

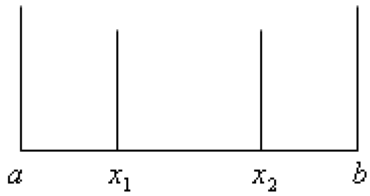
Занятие № 2. ЧИСЛЕННЫЕ МЕТОДЫ ПОИСКА БЕЗУСЛОВНОГО ЭКСТРЕМУМА

Разобьем $[a,b]$ и вычислим значение функции в каждой точке.



В результате остается интервал меньшего размера, к которому применяется тот же метод, и находим еще один интервал, в конце находим интервал с заведомо нужной точкой.

Интервал неопределенности – интервал, в котором заведомо находится точка минимума. Наиболее эффективное разбиение – двумя точками на 3 равных отрезка.



1) $f(x_1) < f(x_2) \Rightarrow [a, x_2]$
2) $f(x_1) > f(x_2) \Rightarrow [x_1, b]$

$\left(\frac{2}{3}\right)^n$ - после выполнения n шагов сокращение исходного интервала

$$\left(\frac{2}{3}\right)^n (b-a) = \delta$$

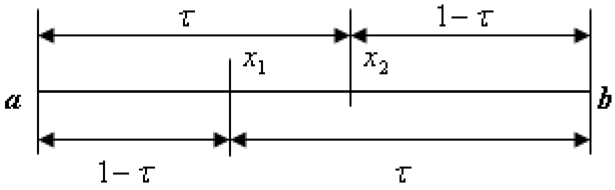
δ - точность с которой надо найти решение задачи.

$$\left(\frac{2}{3}\right)^n = \frac{\delta}{b-a} \qquad n \ln\frac{2}{3} = \ln\frac{\delta}{b-a}$$

N=2n, где n – число шагов, N – число вычислений (мера эффективности данного решения).

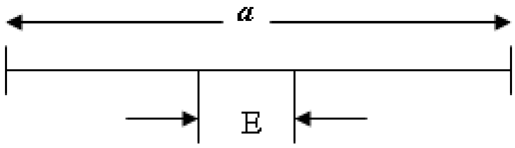
Метод золотого сечения.

Точки должны быть расположены на равном расстоянии.



$$\frac{\tau}{1} = \frac{1-\tau}{\tau};\ \tau^2 = 1-\tau;\ \tau^2 + \tau - 1 = 0;$$

$$\tau_{1,2} = \frac{-1 \pm \sqrt{5}}{2} \Rightarrow \tau = \frac{\sqrt{5}-1}{2} \approx 0{,}618;$$ τ - золотое сечение.



$$\frac{E}{a} < 1$$

λ - величина сокращения на каждом шаге

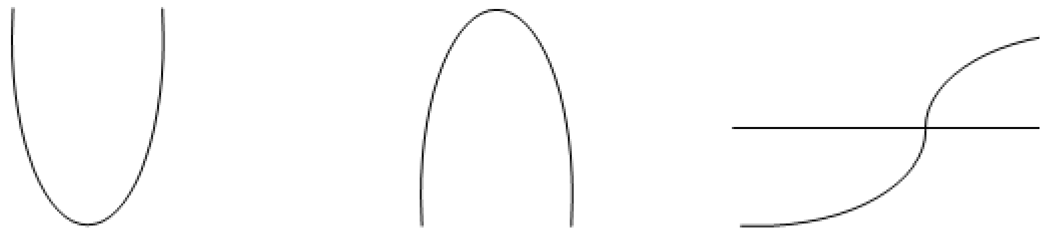
$$\lambda^N = \frac{E}{a};\ N = \ln\frac{E}{a\lambda}$$

число итераций растет как логарифм функции.

Занятие 3. МЕТОДЫ ПЕРВОГО ПОРЯДКА

Одномерная оптимизация с использованием производных.

$f(x) \to \min$. Пусть целевая функция дифференцируема $f'(x) = 0$.

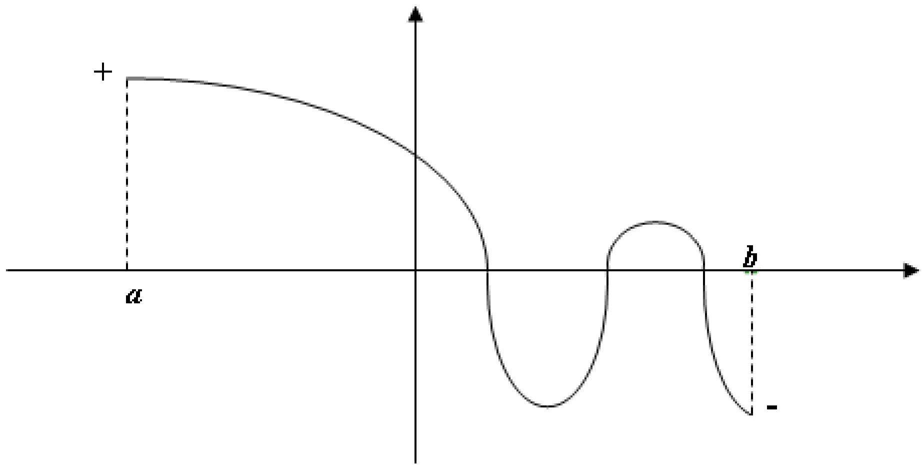


точка локального минимума | точка локального максимума | *точка перегиба*

Методы для нахождения корня уравнения функции 1-ой переменной.

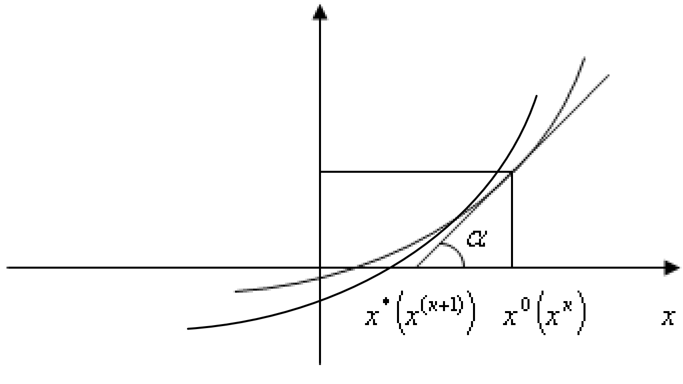
Деление пополам:

Имеется хотя бы 1 корень. Выбираем любую точку и смотрим какой знак она имеет, такой знак нам и искать. Выбираем точку приблизительно в середине интервала, исследуя значения в 3-х можно отбросить половину интервала.



Метод Ньютона (метод касательной):

В случае если известна производная, то выбираем x^0 - начальное приближение.



Допустим, что точка x^0 достаточно близка к корню функции и примерно себя ведет линейно не отклоняется. Проведем касательную и находим точку ближе чем x^0, и повторяем до x^*.

Для метода Ньютона необходимо:

- функция должна иметь производную;
- точка должна быть взята близко к корню;
- функция изменяется близко к линейной функции.

$$\Delta x^{(\kappa)} = x^{(\kappa+1)} - x^{\kappa};$$

$$\underbrace{f(x^{\kappa} + \Delta x^{\kappa})}_{0} = f(x^{\kappa}) + f'(x^{\kappa})\Delta x^{\kappa} + O(\ldots)$$ - уравнение касательной;

$$\begin{cases} \Delta x^{\kappa} = -\dfrac{h(x^{\kappa})}{h'(x^{\kappa})}. \\ x^{\kappa+1} = x^{\kappa} + \Delta x^{\kappa} \end{cases}$$

Если $\left|\Delta x^{\kappa}\right| < \mathrm{E}$, то вычисления можно прекратить и считать что нужный нам корень – условие прекращения поиска. (Е – значение корня с некоторой точностью).

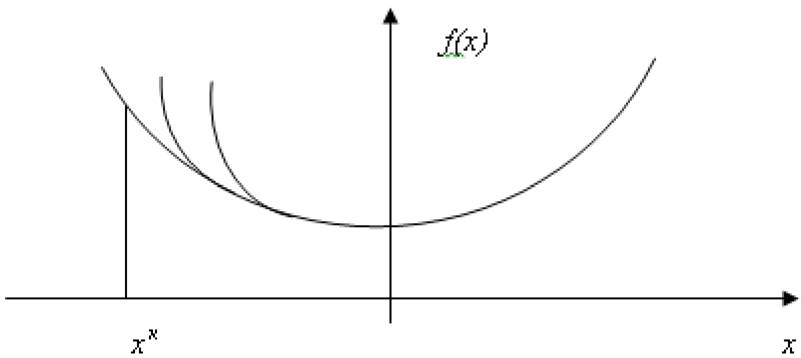
В методе Ньютона каждая его итерация удваивает количество значащих цифр. Если все условия выполнены, то эти методы удваивают (ускоряют) количество значащих цифр:

$$h(x) = f'(x);$$

$$\begin{cases} \Delta x^{\kappa} = -\dfrac{h(x^{\kappa})}{h'(x^{\kappa})} \\ x^{\kappa+1} = x^{\kappa} + \Delta x^{\kappa} \end{cases}$$

Представим что $h(x)$ линейная функция, то метод Ньютона позволяет найти ее корень за *1-у* итерацию. Целевая функция представляет собой квадратичную зависимость следовательно метод Ньютона позволяет найти минимум или максимум квадратичной функции за *1-у* итерацию.

Замена функции на касательную, называется – линейная аппроксимация, и ее применение к целевой функции парабола в точке приближения.

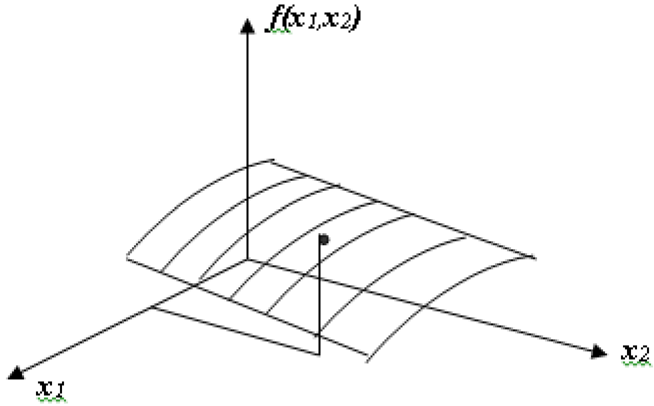


Замена заданной зависимости квадратичной зависимостью, называется – квадратичной аппроксимацией. Метод Ньютона основан на замене заданной зависимости более простой зависимостью.

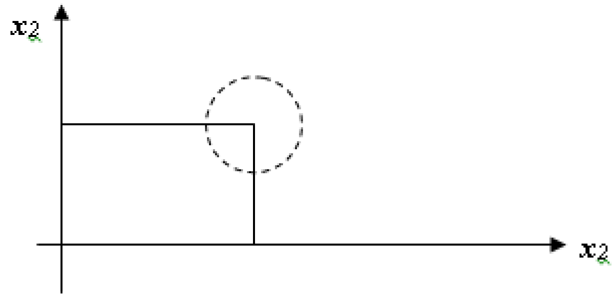
Занятие № 4. МЕТОДЫ ВТОРОГО ПОРЯДКА

Целевая функция зависит от нескольких переменных $f(x_1, x_2, \ldots, x_n) \to \min$. Т.к. нет дополнительных условий накладывающихся на переменные – безусловная оптимизация.

Функции 2-х переменных



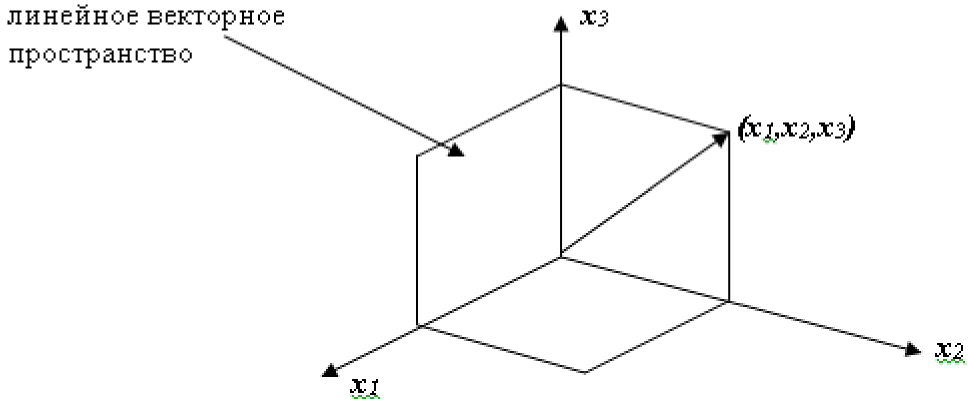
Условия определяющие точку минимума – необходимо проанализировать поведение функции в некоторой точке.



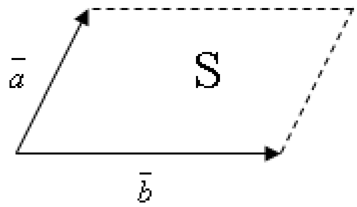
Часто под окрестностью подразумевают шар.

Рассмотрим вспомогательное построение:

$$f(x_1 + \Delta x_2, x_2 + \Delta x_2) = f(x_1, x_2) + \left[\frac{\partial f}{\partial x_1}\Delta x_1 + \frac{\partial f}{\partial x_2}\Delta x_2\right] + O(\ldots)$$



Скалярное произведение векторов $(\bar a,\bar b)=\|\bar a\|\|\bar b\|\cos \bar a\wedge\bar b$, где $\|\cdot\|$ - длина вектора (норма вектора), $\bar a\wedge\bar b$ - угол между векторами.

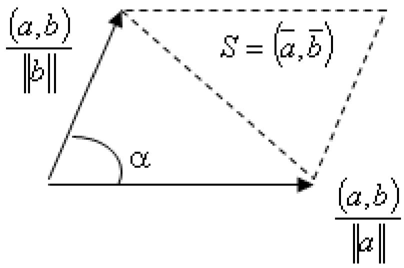


$\|a\|=(a,a)^{1/2}$

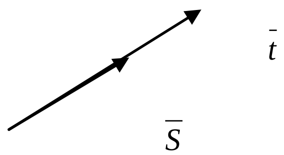
Допустим, что: $a=(a_1,a_2,\ldots,a_n)$, $b=(b_1,b_2,\ldots,b_n)$

Тогда: $(a,b)=\sum_{i=1}^{n}a_ib_i$; $\|a\|=\sqrt{\sum_{i}^{n}a_i^2}$

Допустим, что имеется 2 вектора:



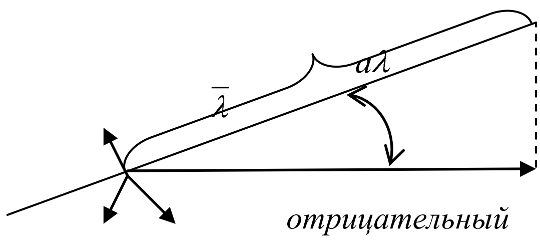
Чтобы задать направление, мы задаем вектор.



Нормируем вектор $\bar\lambda=\frac{\bar S}{\|\bar S\|}$

Нормированный вектор имеет тоже самое направление, но $\|\bar\lambda\|=1$, длина.

Допустим, что задан нормированный вектор $\|\bar\lambda\|=1$.



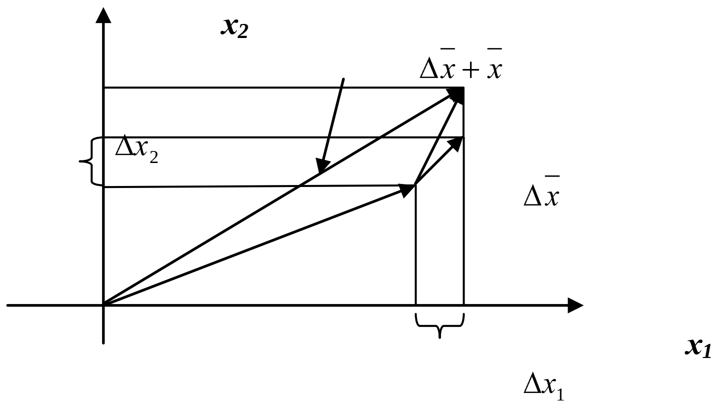
$a\lambda=(\lambda,a)=\|a\|\cos a\wedge\lambda$

Скалярное произведение равно 0, тогда года $\cos a\wedge\lambda$ прямой.

Возвращаемся к функции 2-х переменных:

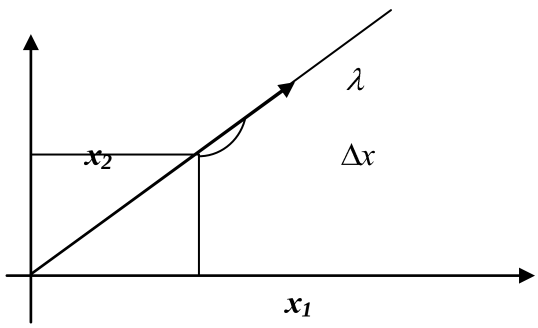
$$f(x_1+\Delta x_2, x_2+\Delta x_2) = f(x_1, x_2) + \underbrace{\frac{\partial f}{\partial x_1}\Delta x_1 + \frac{\partial f}{\partial x_2}\Delta x_2}_{gradf} + O(..)$$

Отбрасываем члены $O(..)$, приращение будет более точным.



$$x = (x_1, x_2)$$
$$\Delta x = (\Delta x_1, \Delta x_2)$$

Вектор $gradf = \left(\frac{\partial f}{\partial x_1}; \frac{\partial f}{\partial x_2}\right) \Rightarrow f(x+\Delta x) = f(x) + (gradf, \Delta x) + O(..)$ - формула Тейлора.



$$\|\bar{\lambda}\| = 1$$

Мы рассматриваем как изменяется точка вдоль данного направления.

Функция становится функцией одной переменной.

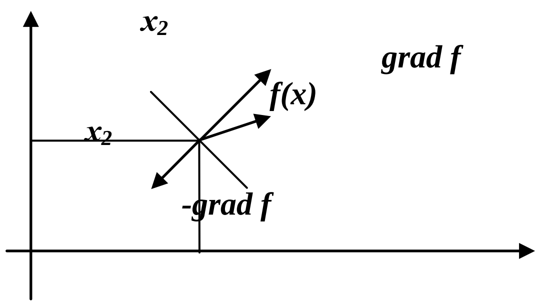
Δx - скалярная величина.

$$f(x+\Delta x) = f(x) + (gradf, \Delta x) + O$$

$\frac{\partial f}{\partial x} = (gradf, x)$ - производная по направлению (вдоль данного направления)

$\max \frac{\partial f}{\partial x}$ - направление ряда равное направлению *grad* (≤ 0).

grad – вектор, в сторону которого функция изменяется более быстро.
Антиградиент – *grad* направленный в другую сторону (-*grad*).



Необходимое условие:

$gradf = 0$ - локальный минимум (или максимум). Точки локального экстремума.

x_1

x_1

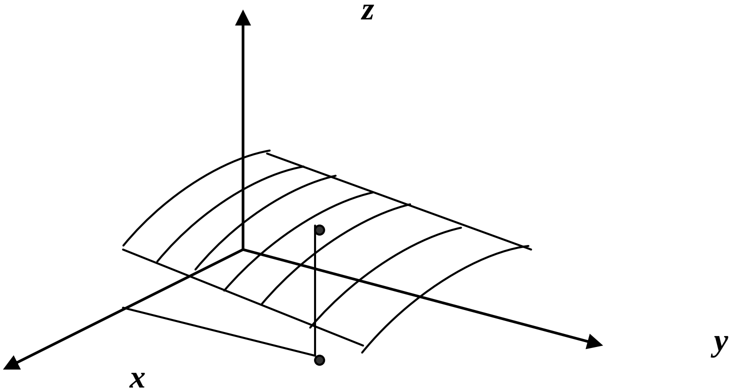
Допустим что мы совершаем малое перемещение Δx. В каком случае (в точке) будет: * больше, чем заданная: тогда, когда угол – острый $\Rightarrow$ $f(x+\Delta x) = f(x) + (gradf, \Delta x) + O(\ldots) = 0$.

* - если под прямым углом, то не изменяется;

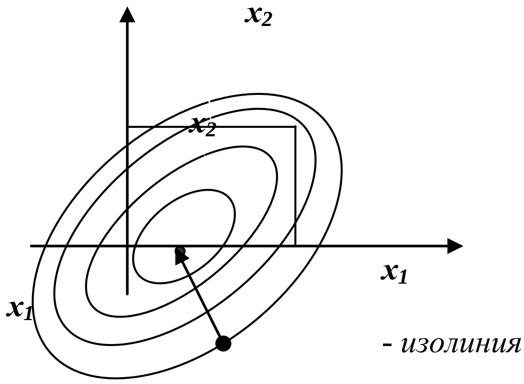
* - если под тупым углом, то приводит к уменьшению функции.

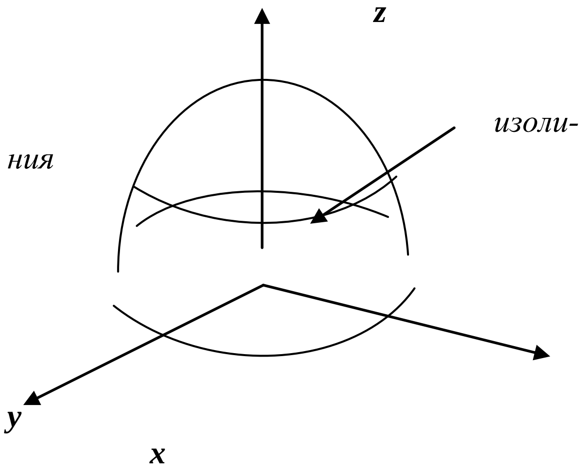
1. $y = f(x_1, x_2)$

строим поверхности



2. Идет построение в плоскости x_1 и x_2. Берут точку – определяющую значение аргумента. Находят точку в которой функция имеет тоже самое значение, в результате получаем линию в которой функция имеет постоянное значение – изолиния (линия уровня).





Вектор *grad* составляет прямой угол с изолинией.

Вернемся к формуле:

$$f(x_1 + \Delta x_2, x_2 + \Delta x_2) = \underbrace{f(x_1, x_2)}_{постоянный\ член} + \frac{\partial f}{\partial x_1}\Delta x_1 + \frac{\partial f}{\partial x_2}\Delta x_2 + \frac{\partial^2 f}{\partial x_1^2}(\Delta x_1)^2 + \frac{\partial^2 f}{\partial x_2^2}(\Delta x_2)^2 + O(\ldots)$$

Квадратичная аппроксимация.

(или квадратичное приращение)

Линейное отображение:

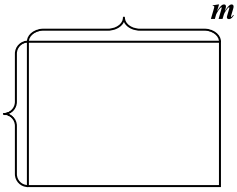
$$L : R^m \to R^n;$$
$$x \in R^m;$$
$$y \in R^n$$

$\bar{y} = y(x)$ - линейное отображение, если:

1. свойство аддитивности - $\varphi(x + y) = \varphi(x) + \varphi(y)$;
2. свойство однородности - $\varphi(kx) = k\varphi(x)$

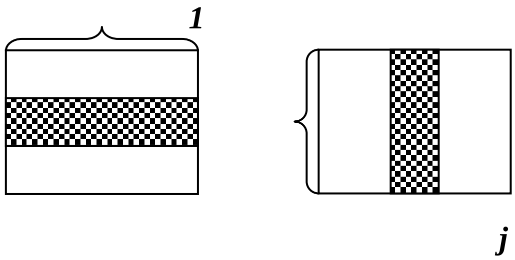
Линейное отображение можно задать матрицей:

$y = Ax$



$y = A(Bx)$; $C = AB$;

$a_j = \sum_k a_{ik} b_{kj}$ - основная формула



$A : R^n \to R^n$ отображение $\quad y = Ax \quad y_i = \sum_j a_{ij} x_j \quad Z = By = (BA)x$

2 задачи:

- решение системы уравнений

  $y = Ax$ – *линейный вектор* $(\ )$

  и обратное отображение – найти x

  A^{-1} – обратное отображение;

  $AA^{-1} = I = A^{-1}A$ следовательно строки матрицы ортогональны столбцам другой матрицы

- нахождение собственных значений

Используя матрицу можно найти более сложную функцию : $Z = (Ax, x)$- квадратичная форма.

$Z = Z(x) = Z(x_1, x_2, \ldots, x_n)$ - функция нескольких переменных $Z = \sum_i \sum_j a_{ij} x_i x_j$.

Рассмотрим подробнее.

Есть матрица: $A = \begin{pmatrix} 2 & 3 \\ -1 & 4 \end{pmatrix}$

$Z = 2x_1x_1 + 3x_1x_2 - 1x_2x_1 + 4x_2x_2$ - квадратичная форма

$Z = 2x_1^2 + 2x_1x_2 + 4x_2^2$

$A' = \begin{pmatrix} 2 & 2 \\ 0 & 4 \end{pmatrix}$

A и A' определяют одну и ту же квадратичную форму следовательно значения этой формы не однозначно. Если по заданной квадратичной форме найдем симметрию, то она будет однозначная.

$\overline{A} = \frac{1}{2}(A + A^T);$

$A^T = \begin{pmatrix} 2 & -1 \\ 3 & 4 \end{pmatrix};\ \frac{1}{2}(A + \overline{A}) = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$

Без ограничения общности можно считать, что матрица определяющая квадратичную форму является симметричной.

Вернемся к квадратичной форме:

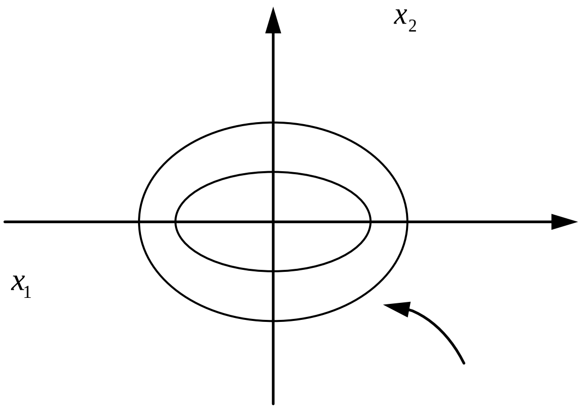
$Z = (Ax, x) = \sum_i \sum_j a_{ij} x_i x_j$

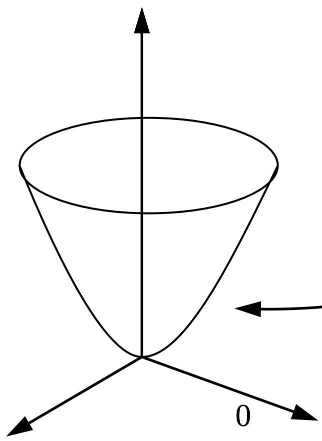
Рассмотрим функцию 2-го порядка:

$$Z = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$$

Допустим, что $a_{12} = 0$, матрица диагональная.

1. $a_{11}, a_{22} > 0$





Эллипти-

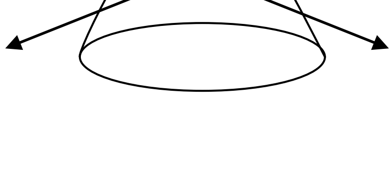
ческий

параболоид

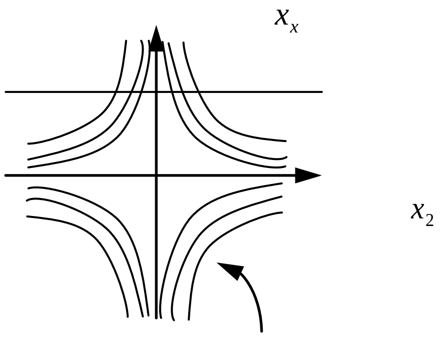
Эл-

липсы

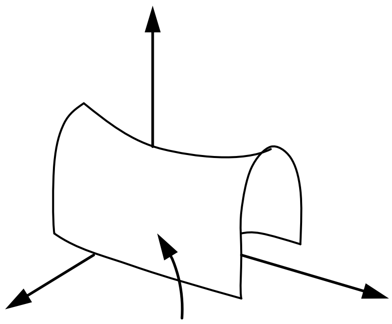
2. $a_{11}, a_{22} < 0$



3. $a_{11} > 0, a_{22} < 0$



Гиперболы



Седло

Допустим, что $a_{12} \neq 0$. Тогда вся картина просто повернется на некоторый угол по оси *Z*.

Рассмотрим *n*-мерный случай.

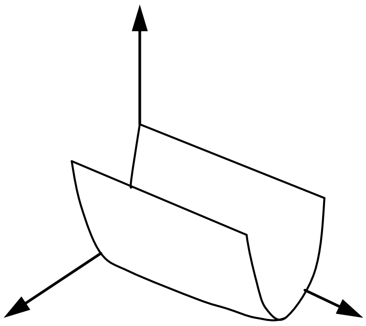
Квадратичная форма называется положительно определенной областью если она не отрицательная.

1. $(Ax, x) \geq 0$, причем обращается в ноль, в том случае если $x = 0$ ($(Ax, x) = 0 \Leftrightarrow x = 0$). Этот случай соответствует эллиптическому параболоиду.
2. $(Ax, x) \leq 0$, $(Ax, x) = 0 \Leftrightarrow x = 0$.
3. Знаконеопределенность.

$(Ax, x) \lessgtr 0$ соответствует *n*-мерному эллиптическому гиперболоиду (*n*-мерное седло)

Рассмотрим 2-мерное пространство:

$$Z = 0x_1^2 + x_2^2$$



Если квадратная матрица называется положительно определенной, то и матрица положительно определенной.

Рассмотрим разложение функции 2-х переменных в ряд Тейлора:

$$f(x_1 + \Delta x_2, x_2 + \Delta x_2) = f(x_1, x_2) + \frac{\partial f}{\partial x_1}\Delta x_1 + \frac{\partial f}{\partial x_2}\Delta x_2 + \frac{\partial^2 f}{\partial x_1^2}(\Delta x_1)^2 + \frac{\partial^2 f}{\partial x_1 \partial x_2}\Delta x_1 \Delta x_2$$
$$+ \frac{\partial^2 f}{\partial x_2^2}(\Delta x_2)^2 + O(\ldots)$$

квадратичная матрица задается матрицей *H*

$$f(x+\Delta x) = f(x) + (grad f, \Delta x) + (H\Delta x, \Delta x) + O(\ldots) = 0$$

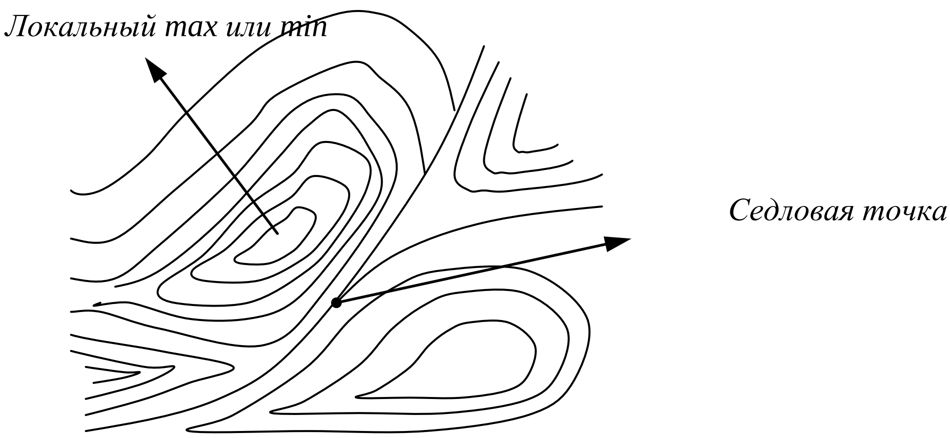
матрица составленная из членов 2-го порядка

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix}$$ - матрица симметрична

Матрица H – матрица Гесса.

$$h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$ - определение матрицы Гесса

Если матрица (матрица Гесса) в точке локального экстремума положительно определена, то это точка – локального минимума, если матрица отрицательно определена, то это точка – локального максимума, а если не определена – седловые точки.



Минимизируем:
$f(x_1, x_2) = 2x_1^2 + \cos x_1 x_2 + 3x_2^2 \to \min$
Найти частные производные:

1. $\frac{\partial f}{\partial x_1} = 4x_1 - x_2 \sin x_1 x_2 = 0$ ($grad =$ 0);
2. $\frac{\partial f}{\partial x_1} = 6x_2 - x_1 \sin x_1 x_2 = 0$

Эта система позволяет найти все точки экстремума:

$$\begin{cases} \frac{\partial f}{\partial x_1} = 4x_1 - x_2 \sin x_1 x_2 \\ \frac{\partial f}{\partial x_1} = 6x_2 - x_1 \sin x_1 x_2 \end{cases}$$

те x_1 и x_2 которые удовлетворяют уравнениям и будет точками экстремума.

Допустим, что $x_1^* = 3{,}5 \quad x_2^* = 2{,}7$. Надо составить функцию второго порядка и подставить $x_1^* \quad x_2^*$ и посмотреть их.

Необходимые условия – помогают охарактеризовать искомую точку:

1. $grad\, f = 0$
2. $H = \dfrac{\partial f}{\partial x_i \partial x_j}$

H ≥ 0 – локальный минимум;
H ≤ 0 – локальный максимум;
H – не определена – седловая точка.

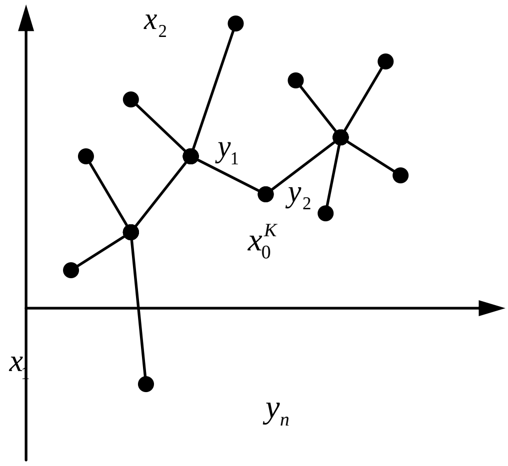
Для поиска используют численные методы.

Постановка:

Требуется $f(x) \to \min$, где x – вектор $x = (x_1, \ldots, x_n)$ - т.к. нет ограничений задача безусловной оптимизации.

Есть черный ящик, который по заданным значениям x позволяет вычислить значение функции.

Занятие № 5. МЕТОДЫ НУЛЕВОГО ПОРЯДКА



Должны задать начальное приближение точки x_0;

$x^к$ - некоторое приближение полученное после $к$ – итераций;

вычислить значение точки в окрестности точки $x^к$;

Из данных точек выбрать точку в которой функция принимает наименьшее значение, выбираем ее и строим вокруг нее окрестность.

Выбираем точку где хуже. В окрестности существующей точки выбираем точку с меньшим значением, опять в ее окрестности есть точки с меньшим значением и т.д.

В таком виде этот метод не эффективен.

Пример:

Шаг по x1 берем больше, а по x2 – сохраняем. Поскольку мы свободны в выборе точек, то можно менять шаг и направление.

Методы:

1. Хука-Дживса;
2. Нелдера-Мида (используется п-1 угольник)

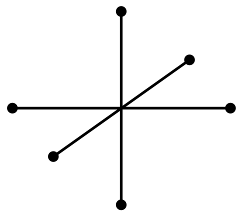
Преимущества метода прямого поиска:

1. простота;
2. не нужны производные.

Недостатки:

1. плохая сходимость;
2. применим для небольшого числа переменных.

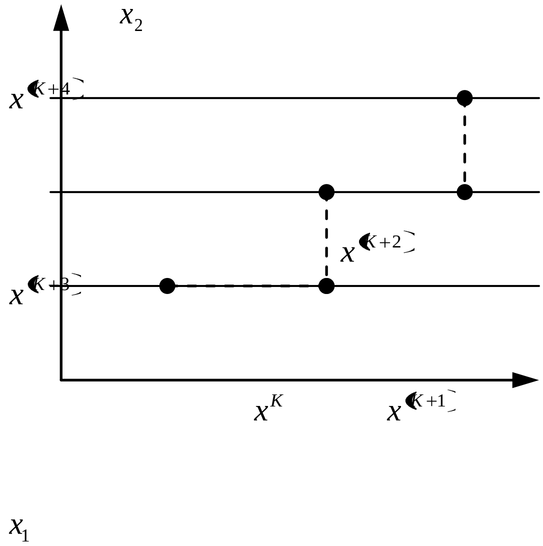
$n \leq 10 \div 20$



2n точек:
в случае 2-х переменных – 4 точки;
в случае 3-х переменных – 6 точек.

Этот метод применим в простых случаях, когда эти недостатки себя не проявляют.

Метод координатного спуска.



Существует приближенная точка минимума. Минимизируя функцию по направлению к x1, на прямой используется любой метод одномерной минимизируют, x2 – фиксируют. Далее выполняют одномерную оптимизацию по x2, фиксируя x1.

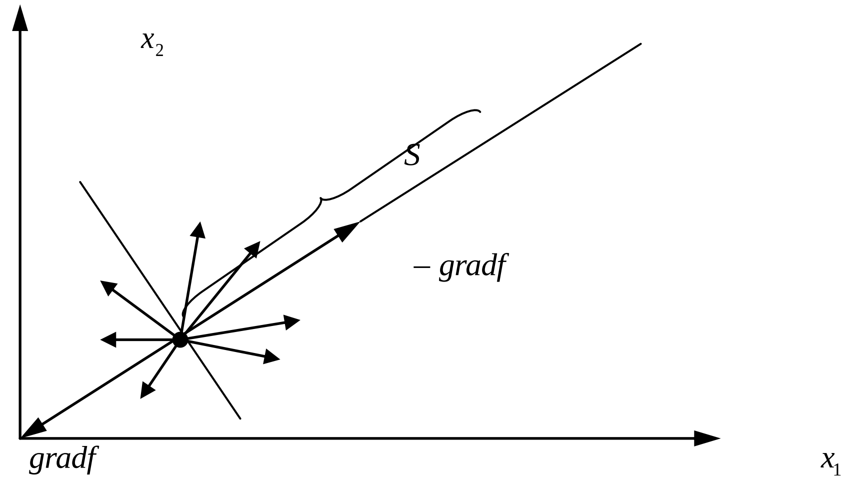
Этот процесс повторяют до тех пор пока следующая точка не окажется близка к точке приближения.

Занятие 6. ЧИСЛЕННЫЕ МЕТОДЫ ПОИСКА УСЛОВНОГО ЭКСТРЕМУМА

Метод наискорейшего спуска.

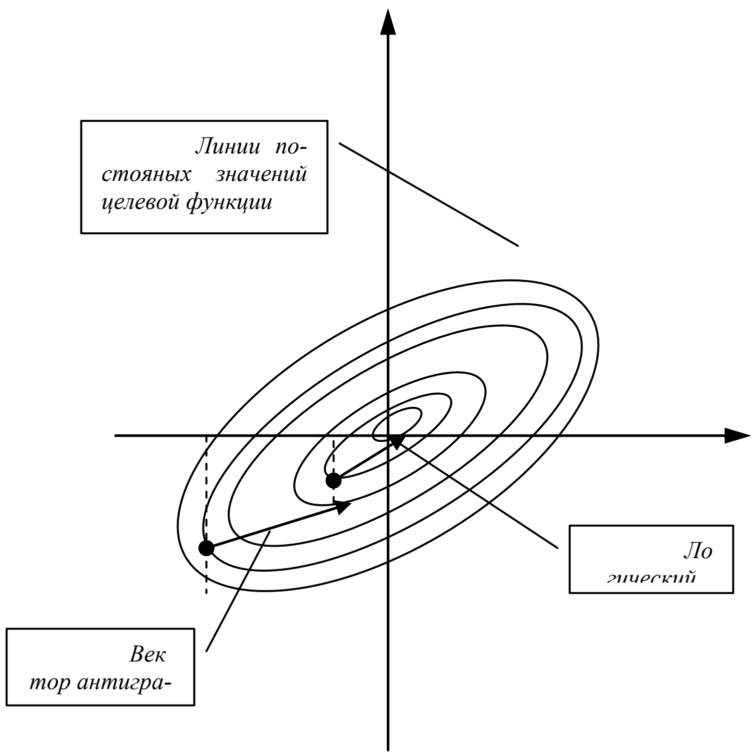
Рассмотрим grad целевой функции.

Движение по направлению вектора под острым углом будет приводить к уменьшению целевой функции, а движение против направления функции к увеличению целевой функции. Разумно за направление движения принять сам вектор – grad f.



Для выбора расстояния нужно применить метод одномерной оптимизации. Прекратить поиск, когда величина grad f станет достаточно малой. Этот метод гарантирует, что найдена либо точка локального минимума, либо седловая точка.

Анализ метода.

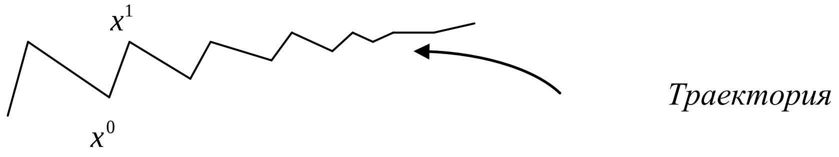


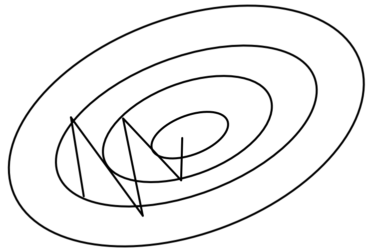
Рассмотрим целевую функцию, которая является квадратичной функцией, точка локального минимума совпадает с точкой начала координат.

Пусть мы выбрали начальное приближение.

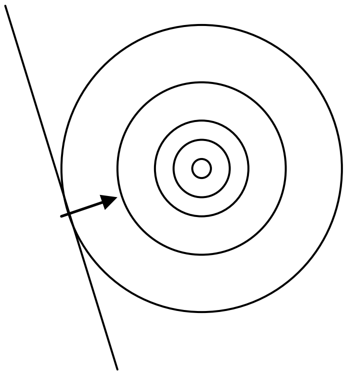
Отыскивая наименьшее значение по направлению траектории (наименьшее значение там где происходит касание grad f линии уровня).

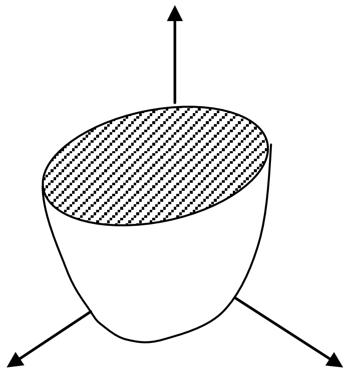
В случае когда масштаб выбирается следующим образом (линии уровня вытянуты).





Если линии уровня $Z = x_1^2 + x_2^2$ - окружности, то приходим сразу в точку локального минимума.





Метод Ньютона.

1. $f(x) = const\ (одночлен)$ - один постоянный член любой точки данной функции является оптимальным – тривиальный случай;
2. линейная функция (двучлен) $f(x) = C_1x_1 + C_2x_2 + \ldots + C_nx_n + C_0 = (C, x) + C_0$

   $gardf = \nabla f = C$ (возможно бесконечное уменьшение и увеличение)

   1 и 2 не подходят для оптимизации**.**

3. трехчлен $f(x) = (C_1x_1 + C_2x_2 + \ldots + C_nx_n) + C_0 + \sum_i \sum_j S_{ij}x_ix_j$

   $f(x) = \frac{1}{2}(Qx, x) + (C, x) + C_0$;

   $(Qx, x) = \sum_j y_jx_j = \sum_i \sum_j q_{ij}x_ix_y$

   без ограничения общности можно положить что матрица q – симметричная $q_{ij} = q_{ji}$

Разложим функцию в ряд Тейлора (должно быть 3 члена). Чтобы найти линейный член квадратичной функции, надо взять *grad*.

$$\frac{\partial}{\partial x_\kappa}\left(\sum_i \sum_j q_{ij}x_ix_j\right) = \sum_i \sum_j q_{ij}\frac{\partial}{\partial x_\kappa}(x_ix_j) = \sum_i \sum_j q_{ij}\left(\frac{\partial}{\partial x_\kappa}x_i + \frac{\partial}{\partial x_\kappa}x_j\right) = \sum_i q_{\kappa i}x_i +$$

$$\sum_j q_{\kappa j}x_j = \sum_j q_{\kappa j}x_j + \sum_j q_{j\kappa}x_j = 2\sum_j q_{\kappa j}x_j = 2Qx$$

$grad\,(Qx, x) = 2Qx$;
$gradf = Qx + C$; $C = 0$

Найдем матрицу Гесса (матрица вторых частных производных)

$$H = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right) = Q$$

элемент матрицы Гесса является элементом функции Q. $h_{ij} = q_{ij}$ (все частные производные высших порядков равны 0).

Функция экстремальна, если $grad$ в данной точке равен 0, $gradf = Qx + C$ следовательно условие экстремальности $Qx = -C$ - система.

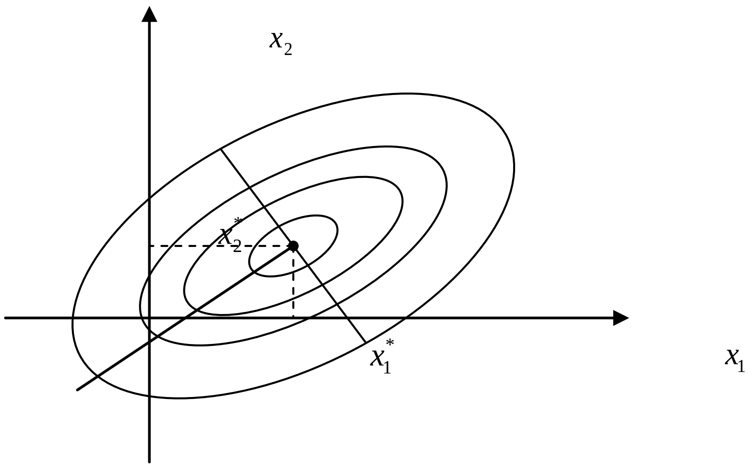
Необходимое условие оптимальности:

Если $Q > 0$ решение данной системы существует и оно единственное (совместная система).

Если $Q < 0$ решение данной системы существует и оно единственное, т.е. если Q знакоопределена, то существует решение и оно единственное.

Если имеем квадратичную функцию и матрица положительно определена, то линии уровня – эллипсы.

Собственные значения определяют оси эллипсов.



Чтобы определить координаты точки локального минимума, нужно решить систему $Qx^2 = -C$.

Пусть *f(x)* – произвольная функция и надо найти точку локального минимума. Разложим функцию в ряд Тейлора в окрестности точки.

$$f\,(x^к + \Delta x^к) = f\,(x^к) + \nabla f \Delta x^к + (H\Delta x^к, \Delta x^к) + O(\ldots)$$

Пусть функция не квадратичная, эллипсы примерно отражают кривизну линий уровня и находятся в окрестности точки $x^к$. В окрестности точки $x^к$ находим приближение и заменяем эту функцию квадратичной функцией, которая получается из разложения в ряд Тейлора. Далее решаем задачу минимизации.

Находим точку минимума и рассматриваем эту точку как следующее приближение и т.д.

Для нахождения точки минимума квадратичной функции (зависит от Δx)необходимо решить систему:

$$H\Delta x^{\kappa} = -\nabla f(x^{\kappa})$$

Окончательно следующее приближение $x^{\kappa+1} = x^{\kappa} + \Delta x^{\kappa}$.

$$x^{\kappa+1} = x^{\kappa} - H^{-1}\nabla f(x^{\kappa})$$ - формула Ньютона

(обобщение формулы минимизации одной переменной)

Выполнение метода останавливается когда $\nabla \approx 0$, т.е. когда Δx^{κ} очень мало. Для получения практической точности достаточно выполнить 4 итерации метода Ньютона.

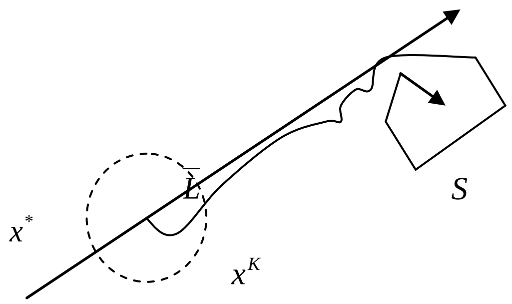
Если f – хороша, то метод Ньютона подходит, если f – квадратичная функция, то метод Ньютона приводит к минимальной точке за 1 шаг, из любой точки.

Недостатки:

1. на каждом шаге итерации надо находить решение системы $H\Delta x^{\kappa} = -\nabla f(x^{\kappa})$;
2. С ростом числа итераций H – разрежается, т.е. большое число членов становится равными 0.

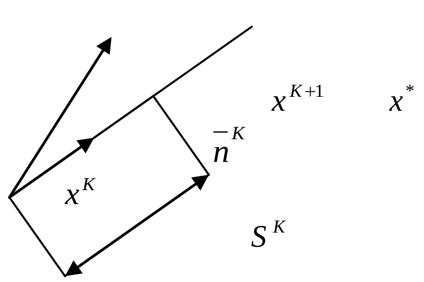
Все формулы безусловной минимизации можно записать в общую схему:

1. выбор направления;
2. выбор шага.



x^{κ} - приближение в точке локального минимума, чтобы приблизиться к искомой точке x^{κ}. Мы должны выбрать направление, в конце получим локальную линию.

Допустим, требуется f(x)→min; x^0 - начальное приближение; x^K - текущее приближение



$\|\bar{n}\| = 1$

а) выбор направления $(gradf, \bar{n}) < 0$;

б) движение вдоль выбранного направления $\overline{X^{K+1}} = \overline{X^K} + S^K * n^K$

Задачи оптимизации с ограничениями – разностями (ЗОР)

$$\begin{cases} f(x_1, x_2, ..., x_n) \to \min \\ g_1(x_1, x_2, ..., x_n) = 0 \\ .... \\ g_m(x_1, x_2, ..., x_n) = 0 \end{cases} \qquad \begin{matrix} f(x) \to \min \\ g(x) = 0 \end{matrix}$$

Пример:

$$\begin{cases} f(x_1, x_2, x_{3,} x_4) \to \min \\ g_1(x_1, x_2, x_{3,} x_4) = 0 \\ g_m(x_1, x_2, x_{3,} x_4) = 0 \end{cases}$$

Функции заданы аналитическим выражением $\Rightarrow$ q_1 можно разрешить относительно одной из переменных $x_1 = h_1(x_2, x_3, x_4)$ $\Rightarrow$ x_1 можно исключить из f и q_2, подставив вместо нее L_1:

$$\begin{cases} \overline{f}(x_2, x_{3,} x_4) \\ \overline{g}(x_2, x_{3,} x_4) = 0 \end{cases} \qquad \Rightarrow \qquad x_2 = h_2(x_3, x_4)$$

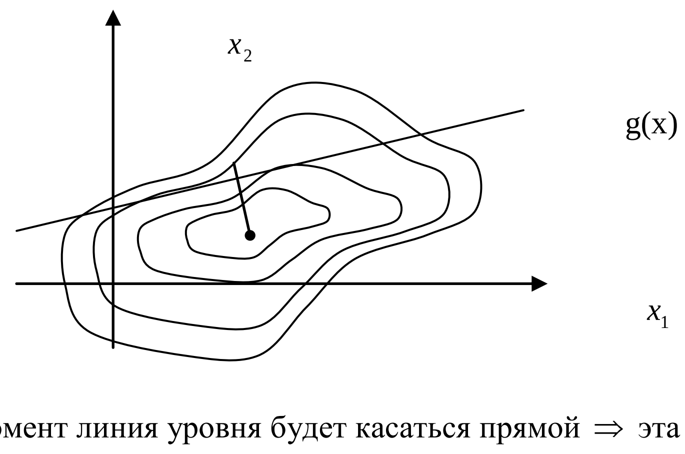
Тогда, $\overline{\overline{f}}(x_3, x_4)$ - задача безусловной оптимизации. Находим $x_3, x_4 \Rightarrow$вычисляем $x_2 \Rightarrow x_1$

Метод исключения.

Численное решение:

$$\begin{cases} f(x_1, x_2) \to \min \\ a_1x_1 + a_2x_2 = b \end{cases}$$

[redacted]$(x_1, x_2) = a_1x_1 + a_2x_2 - b$ $\Rightarrow$ точка min должна лежать на прямой.



В каждый момент линия уровня будет касаться прямой $\Rightarrow$ эта точка и является точкой

условного локального min. Если в окрестности заданной точки, удовлетворяющей

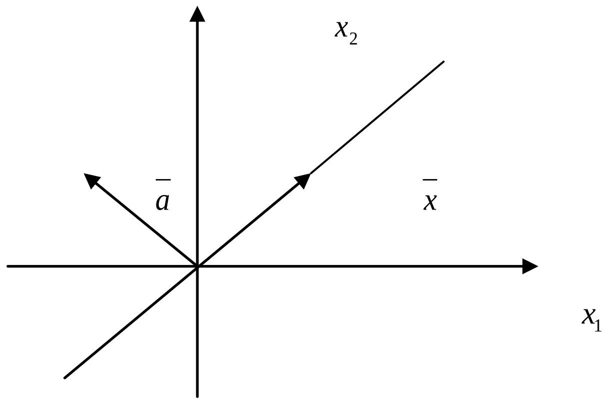
всем значениям равенства, значение функции больше, чем в точке, то эта точка – есть точка условного локального min.

Пример:

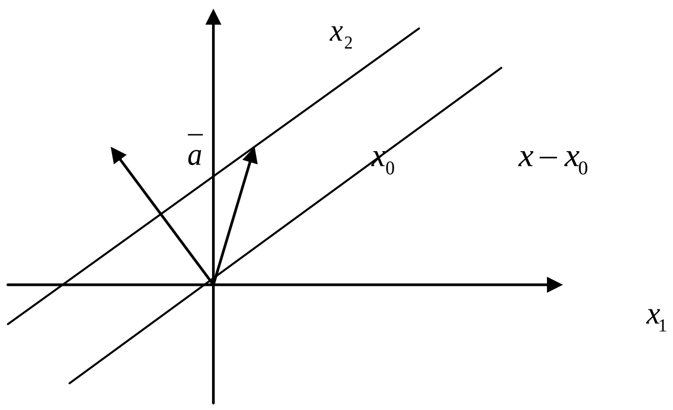
$a_1x_1 + a_2x_2 = 0$

(a,x)=0

$\bar{a} \perp \bar{x}$



Если $(a_1x)=b$



Допустим, $b = (a_1x_0)$

$(a, x) - (a, x_0) = 0$

$(a, x - x_0)$

Прямая будет проходить через некоторую точку удовлетворяющую условию и $\perp \bar{a}$

Для n переменных $f(\bar{x}) \to \min$, Ax=b

Рассмотрим i-ое ограничение:

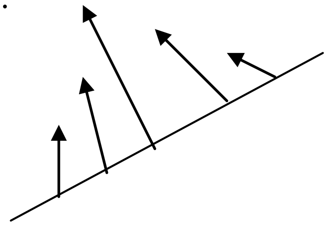
$a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n = 0,\ (a, x) = 0$

$\bar{a}$ - задан $\to$ x - все вектора, лежащие $\perp \bar{a}$. Они и составляют гиперплоскость.

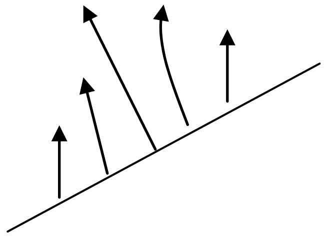
При добавлении еще одного условия, уменьшаются размерности. В конечном итоге получится пространство n-m.

Для двух переменных возможно 2 случая:

1.

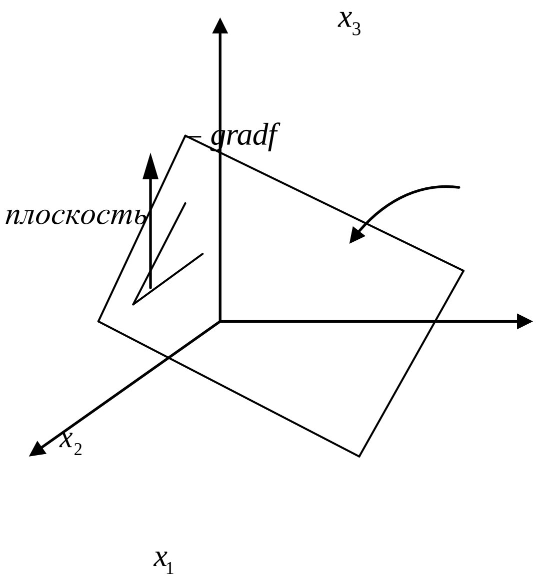


2.



В случае 2 это не точка минимума, а седловая точка.

Рассмотрим точку 3-х переменных:



Ограничение – плоскость, следовательно, все допустимые точки на плоскости.

Если угол *grad* не равен 90 градусам следовательно можно двигаться дальше. На плоскости существует направление, которое будет составлять острый угол с $-grad$, и двигаясь в этом направлении можно уменьшить значение f.

Если -*grad f* перпендикулярен плоскости эта точка может быть точкой минимума.

Пусть существует 2 ограничения:

$$\begin{cases} f(x) \to \min \\ a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n} = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n} = b_2 \end{cases}$$

Рассмотрим опять случай 3-х переменных:

Точка минимума должна принадлежать пересечению плоскостей.

Необходимое условие – вектор антиградиента должен составлять угол 90 градусов с прямой пересечения плоскостей.

Для *n*-мерного случая имеется *n* переменных следовательно рассматривая каждое ограничение, получаем *n-1* гиперплоскость следовательно рассмотрев *m* ограничений получим *n-m* гиперплоскость (*m*<*n*).

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n} = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n} = b_2 \\ \vdots \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn} = b_m \end{cases}$$

$Ax = b$

все ограничения независимы

Если вектор *grad* (*n*-мерный) будет ортогонален *n-m* – пространству.

Допустим имеется *n-1* пространство, *n*-мерный вектор может принадлежать ему или нет. Пусть вектор не принадлежит данному подпространству следовательно его можно разложить на 2 вектора – один который принадлежит подпространству, и второй который ортогонален данному. Ортогональное дополнение – вектора, которые ортогональны данному подпространству.

В 3D – пространстве, если подпространство равно 1 следовательно ортогональное дополнение равно 2.

В *n-m*-мерном подпространстве ортогональное дополнение имеет размерность *m*.

Необходимое условие: Если мы находим точку, где вектор градиента принадлежит ортогональному дополнению к пространству, заданному ограничениями – равенствами, то эта точка может быть точкой локального минимума.

Пусть есть 2 плоскости. Если записать систему ограничений равенств следующим образом:

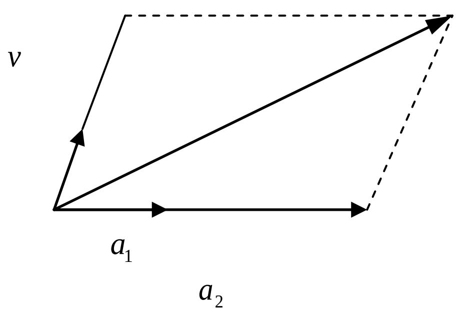
$$\begin{cases} (a_1, x) = b_1 \\ (a_2, x) = b_2 \\ (a_m, x) = b_m \end{cases}$$

где $a_i = a_{i1}, a_{i2}, \ldots, a_{im}$

Т.о. вектора $a_1, a_2, \ldots, a_m$ порождают ортогональное дополнение. Существующие могут быть выбраны в качестве базиса ортогонального дополнения следовательно градиент принадлежит ортогональному дополнению:

$$-gradf = \lambda_1 \overline{a_1} + \lambda_2 \overline{a_2} + \ldots + \lambda_m \overline{a_m}$$

т.е. линейная комбинация базисных векторов.



$$\overline{V} = \lambda_1 \overline{a_1} + \lambda_2 \overline{a_2}$$

$\lambda_1 + \lambda_2 + \ldots + \lambda_m$ - множители Лагранжа.

Рассмотрим матрицу A^T, в ней $a_1, \ldots, a_m$ - столбцы.

$$-gradf = A^T \lambda$$

это условие может быть использовано для численного решения задачи оптимизации с ограничивающими уравнениями.

Пример: $\begin{cases} f(x) \to \min \\ Ax = b \end{cases}$

$$Ax = b$$

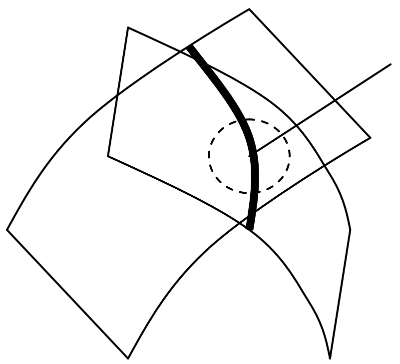
$$A^T \lambda = -gradf$$

Если найдем такие вектора x и λ, для которых эти условия выполняются то точка может быть точкой локального минимума.

Рассмотрим случай когда система ограничений – равенств нелинейная:

$$\begin{cases} f(x_1, \ldots, x_n) \to \min \\ g_1(x_1, \ldots, x_n) = 0 \\ \vdots \\ g_m(x_1, \ldots, x_n) = 0 \end{cases} \qquad \begin{cases} f(x) \to \min \\ g(x) = 0 \end{cases}$$

Если функции дифференцируемы, то в окрестности точки минимума они будут вести себя как линейные.



x^K

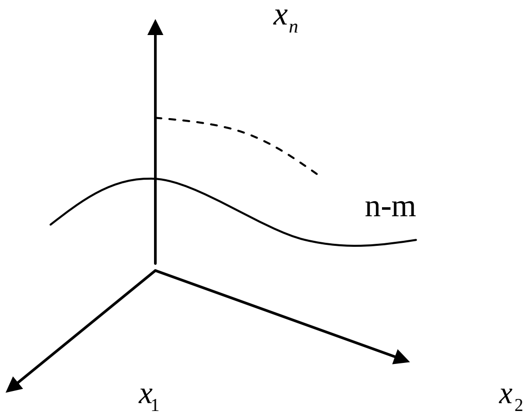
$$\overline{g_i}(x^* + \Delta x) = g_i(x^*) + \frac{\partial g_i}{\partial x_j} \Delta x + O(\ldots)$$

следовательно в окрестности точки локального минимума эта зависимость линейная следовательно получается система вида:

$$A\Delta x = g_i(x^*) - \overline{g_i}(x^* + \Delta x) + O(\ldots),$$

где $A = \dfrac{\partial g_i}{\partial x_j}$

следовательно необходимое условие локального минимума:



$$-gradf = \lambda_1 gradg_1 + \lambda_2 gradg_2 + \ldots + \lambda_m gradg_m$$

$\lambda_1 + \lambda_2 + \ldots + \lambda_m$ - множители Лагранжа.

$$\begin{cases} f(x) \to \min \\ g(x) = 0 \end{cases}$$

$$-gradf=\lambda_1 gradg_1+\lambda_2 gradg_2+\ldots+\lambda_m gradg_m$$

x^* - точка может быть искомой в задаче

$\lambda_1+\lambda_2+\ldots+\lambda_m$ - множители Лагранжа.

$$\begin{matrix} n \\ m \end{matrix}\begin{cases} -gradf=\underbrace{\lambda_1 gradg_1+\lambda_2 gradg_2+\ldots+\lambda_m gradg_m}_{m} \\ g(x)=0 \end{cases}$$

Обозначения для скалярного произведения $(\bar{x},\bar{y})=x^T y$;

$$(\bar{x},\bar{y})=(x_1,\ldots,x_n)\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}=\sum_i x_i y_i \ ;$$

$$L(x,y)=f(x)+\lambda^T g(x)=f(x)+\lambda_1 g_1(x)+\ldots+\lambda_m g_m(x)$$

Необходимое условие точки локального минимума исходное задание с ограничениями представляет собой необходимое условие точки локального экстремума для функции Лагранжа.

$$\frac{\partial L}{\partial x_j}=\ldots=f(x)+\lambda_1 g_1(x)+\ldots+\lambda_m g_m(x)=\frac{\partial f}{\partial x_j}+\lambda_1\frac{\partial g_1}{\partial x_j}+\ldots+\lambda_m\frac{\partial g_m}{\partial x_j}=0$$

$$\frac{\partial L}{\partial x_j}=g_i(x)=0, i=1,\ldots,m$$

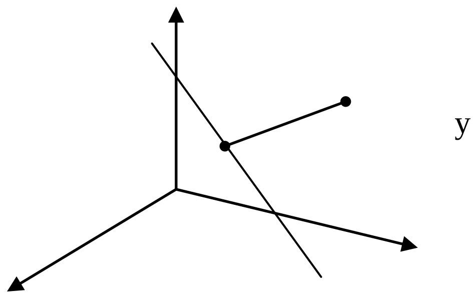
Метод множителей Лагранжа.

Применяется для нахождения точки локального минимума для точек исходной задачи $h(x,\lambda)=f(x)+\lambda^T g(x)$. Экстремальными точками локального минимума являются седловые.

Пример:

Найти расстояние от точки до прямой в 3-х мерном пространстве.

Плоскость : $\begin{matrix} a_{11}x_1+a_{12}x_2+\ldots+a_{1n}=b_1 \\ a_{21}x_1+a_{22}x_2+\ldots+a_{2n}=b_2 \end{matrix}$



Пересечение плоскостей – линия $d=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+(x_3-y_3)^2}\to\min$

$Z=(x_1-y_1)^2+(x_2-y_2)^2+(x_3-y_3)^2\to\min$

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n} - b_1 = 0$$
$$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n} - b_2 = 0$$

$$L(x_1, x_2, x_3, \lambda_1, \lambda_2) = \sum_{j=1}^{3} (x_j - y_j)^2 + \lambda\left(\sum_{j=1}^{3} a_{1j}x_j - b_1\right) + \lambda_2\left(\sum_{j=1}^{3} a_{2j}x_j - b_2\right)$$

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 0 \\ \dfrac{\partial L}{\partial x_2} = 0 \\ \quad\vdots \\ \dfrac{\partial L}{\partial \lambda_2} = 0 \end{cases}$$

5 условий дают систему линейных уравнений

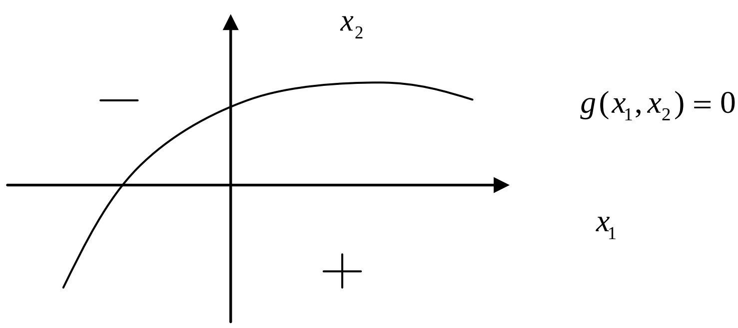
Нелинейное программирование (НЛП).

$$\begin{cases} f(x_1, \ldots, x_n) \to \min \\ g_i(x_1, \ldots, x_n) = 0, i = 1, \ldots, m \\ h_к(x_1, \ldots, x_n) \geq 0 \end{cases}$$

$f, g_i, h_к$ - заданные функции нелинейные

$$\text{НЛП} \begin{cases} f(x) \to \min \\ \bar{g}(x) = 0 \\ \bar{h}(x) \geq 0 \end{cases}$$

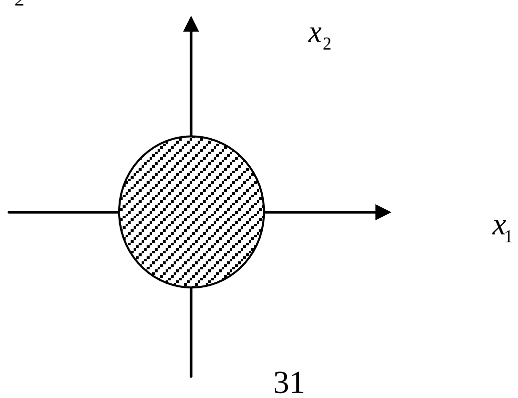
Рассмотрим $g(x_1, x_2) \geq 0$



Пример:

$x_1^2 + x_2^2 \leq 4$

$g(x_1, x_2) = 4 - x_1^2 - x_2^2 \geq 0$

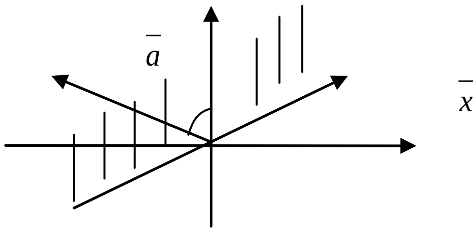


В случае системы неравенств пересечение всех областей. Если *g > 0*, то ограничение неравенства – неактивно (точку можно смещать).

Если точка точно на границе, то говорят, что ограничение активно.

Рассмотрим случай:

$$a^T x = a_1 x_1 + a_2 x_2 \geq 0$$

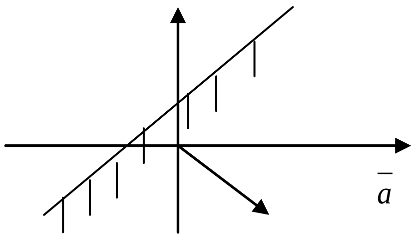


Если задано линейное ограничивающее неравенство, то вектор $\bar{a}$ направлен внутрь допустимой области. Если $a^T x \leq 0$, то вектор $\bar{a}$ будет направлен из допустимой области.

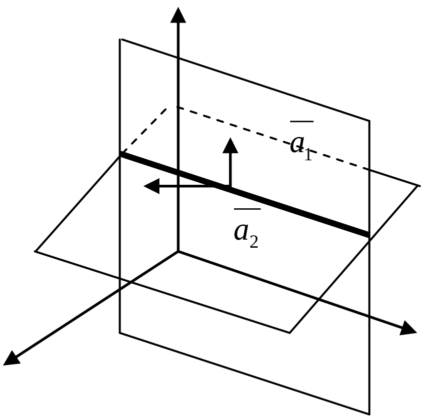
Если $a^T x \geq b$, то граница проходит не через начало координат.

Необходимые условия:

$$f(x_1, x_2) \to \min$$
$$a_1 x_1 + a_2 x_2 \geq b$$



1. Если локальный минимум внутри допустимой области, то $gradf = 0$;
2. Если точка локального минимума точно на границе, то $gradf = \lambda \bar{a}$, точка является точкой локального, если $gradf = \lambda \bar{a}$ и $\lambda \geq 0$



$\bar{a}_1, \bar{a}_2$ - вектора нормали к соответствующей плоскости.

$\lambda_1, \lambda_2 \geq 0$

В общем случае:

$$\begin{cases} f(\bar{x}) \to \min \\ g_1(\bar{x}) = 0 \\ \vdots \\ g_m(\bar{x}) = 0 \\ h_1(\bar{x}) \geq 0 \\ \vdots \\ h_m(\bar{x}) \geq 0 \end{cases}$$

а) $grad f(\bar{x}) = \overline{h_1} grad g_1(\bar{x}) + \ldots + \overline{h_m} grad g_m(\bar{x}) + \lambda_1 grad h_1(\bar{x}) + \ldots + \lambda_p grad h_p(\bar{x})$;

б) $\begin{matrix} \lambda_1 \geq 0 \\ \lambda_2 \geq 0 \\ \vdots \\ \lambda_p \geq 0 \end{matrix}$;

в) Если $h_\kappa = 0$, то $\lambda_\kappa > 0$. Если $h_\kappa \geq 0$, то $\lambda_\kappa = 0$. Т.е. $\lambda_\kappa + h_\kappa = 0, \kappa = 1, \ldots, p$. Условие дополняющей нежесткости.

Все 3 условия в совокупности называются условиями Куна-Таккера (условия оптимальности первого порядка).

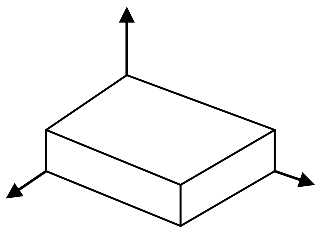
$$\begin{cases} f(\bar{x}) \to \min \\ \bar{g}(\bar{x}) = 0 \\ \bar{h}(\bar{x}) \geq 0 \end{cases}$$

Ограничения неравенства

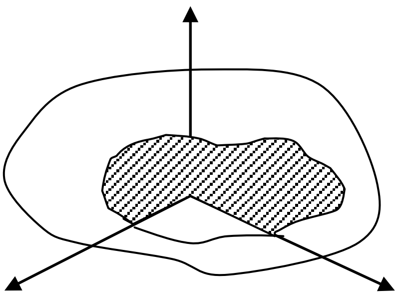
$$t_i = h_i(\bar{x}) \geq 0$$
$$\Downarrow$$
$$t_i - h_i(\bar{x}) = 0, t_i \geq 0$$

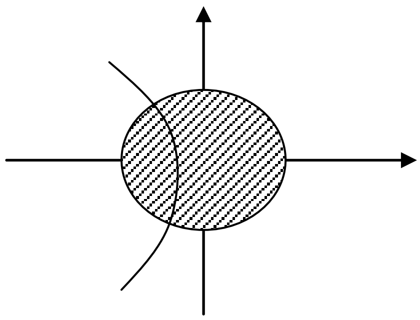
Можно записать и так:

$$\begin{cases} f(\bar{x}) \to \min \\ \bar{g}(\bar{x}) = 0 \\ \bar{x} \geq 0 \end{cases}$$



Поскольку постановка задачи $\bar{h}(\bar{x}) \geq 0$





Основные результаты:

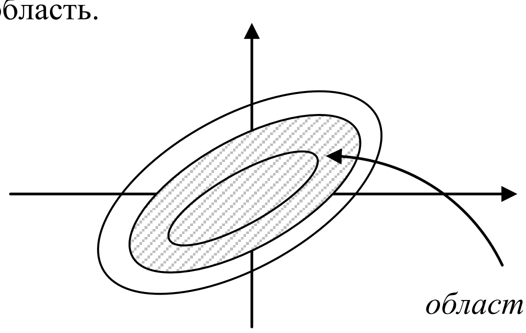
Область n-мерного пространства называется выпуклой если вместе с 2-ми точками, она содержит весь отрезок, соединяющий эти 2 точки.

Пример:

Функция нескольких переменных $f(x)$ называется выпуклой если ее матрица Гесса положительно определена.

$$H = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right); \; x^T H x \geq 0$$

Если мы рассматриваем неравенство $f(x) \leq C$, то данное неравенство определяет выпуклую область.



область будет выпуклой

Th: Пусть дана задача НЛП, если целевая функция этой задачи – выпуклая, и область целевых решений так же выпукла, то локальный оптимум совпадает с ее глобальным оптимумом задачи (задачи выпуклого программирования).

1 случай – когда все ограничительные неравенства являются не активными.

2 случай – когда точка лежит на границе.

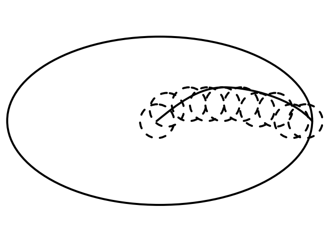
Методы решения НЛП.

Нулевого порядка – поисковые методы (безусловные ориентиры похожи на это). Используется только значение целевой функции (Z).

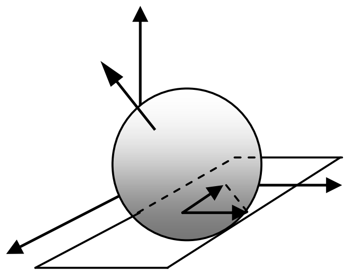
Первого порядка – аналогичны градиентным методам. Условно градиентные методы. Используется и Z и вектор градиента ($grad\ Z$).

Второго порядка. Ньютоновские методы. Они являются специальными вариантами методов Ньютона для оптимизации. Используется Z, $grad\ Z$ и матрица Гесса (H)

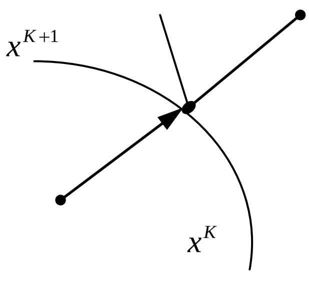
(*)



(**)



(***)



(**)

1 случай – вектор grad направлен по нормали;

2 случай – идет под углом (надо спроецировать поверхность следовательно она будет показывать направление)

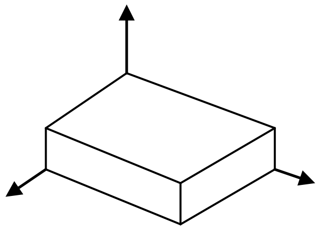
Если мы внутри, двигаемся как в (*), а далее (**). Это более эффективный метод.

(***)
Рассмотреть отрезок, это может дать нам еще один отрезок.

Занятие № 7. ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

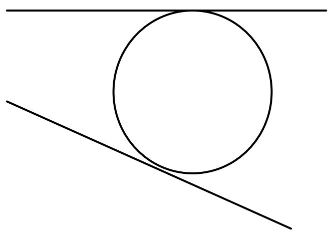
$$\begin{cases} Z = C^T x + C_0 \to \min \\ Ax = b \\ x \geq 0 \end{cases}$$ - *стандартная форма задачи ЛП*

В общем случае, если $Ax \geq b$, то допустимая область представляющая собой многогранник в пространстве.



В случае $Ax = b, x \geq 0$ - многогранник, имеющий неполную размерность.

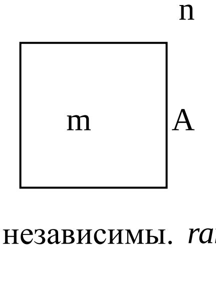
Допустим, имеется некоторое выпуклое множество. Тогда в любой граничной точке этого множества, всегда можно провести опорную гиперплоскость, т.е. такую гиперплоскость которая имеет с множеством только одну общую точку, и все множество находится по одну сторону от гиперплоскости.



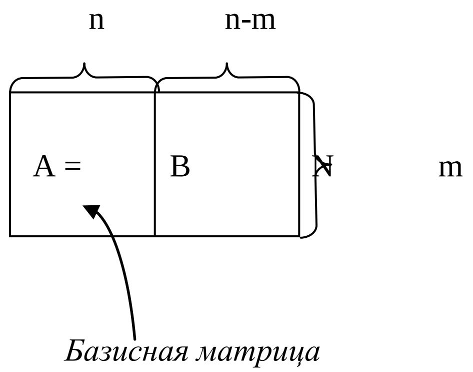
Если $-grad$ является нормалью к гиперплоскости и плоскость не опорная, то можно двигаться под острым углом к $-grad$, тем самым улучшая значение функции. Такое движение невозможно, если антиградиент определяет опорную плоскость. Следовательно в этом случае это точка локального минимума, который является и глобальным.

Геометрически, чтобы найти точку локального минимума, необходимо найти такую вершину глобального множества, что плоскость которая является нормальной к антиградиенту является опорной.

Рассмотрим $Ax = b, x \geq 0$, m – ограничений равенств, n – число переменных.



Первые m столбцов линейно независимы. $rank(B) = m$, $\det(B) \neq 0$.



$$Ax = A_1x_1 + A_2x_2 + \ldots + A_nx_n,\ A_1, A_2, \ldots, A_n \text{ - столбцы матрицы}$$

$$A = \frac{B}{N} \Rightarrow x = \frac{x_B}{x_N} \text{ - базисные переменные}$$

Тогда систему ограничений равенств можно записать

$$\frac{B}{N}\frac{x_B}{x_N} + b;$$

$$Bx_B + Nx_N = b \quad Bx_B = -Nx_N + b;$$

Для B существует обратная матрица $B^{-1}x_B = -B^{-1}Nx_N + B^{-1}b$;

$$x_B = -B^{-1}Nx_N + B^{-1}b$$

Если для данного базиса зафиксируем не базисные переменные в нуле, то получим точку, которая является вершиной многогранника.

Вершины многогранника множества характеризуемые тем, что небазисные переменные равны 0.

Что делать если вершина не точка оптимума.

Рассмотрим целевую функцию:

$$Z = C^T x = \left(\frac{C_B}{C_N}\right)^T \left(\frac{x_B}{x_N}\right) = C_B^T x_B + C_N^T x_N = C_B^T \left(-B^{-1}Nx_N + B^{-1}b\right) + C_N^T x_N =$$

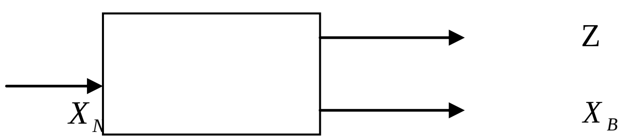
$$-C_B^T B^{-1}Nx_N + C_B^T B^{-1}b + C_N^T x_N = \left(C_N^T - C_B^T B^{-1}N\right)x_N + C_B^T B^{-1}b =$$

$$\Bigl(\underbrace{C_N - N^T B^{-T} C_B}_{d}\Bigr)^T x_N + \underbrace{b^T B^{-T} C_B}_{d_0} = B^{-T} = \left(B^T\right)^{-1}$$

d – показывает суммарное влияние небазисных переменных на целевую функцию

d_0 – множители Лагранжа или относительные оценки небазисных переменных.

$$Z = C^T x = d^T x_N + Z_0$$



Точка будет точкой оптимума, если все $d \geq 0$.

Если имеется один отрицательный коэффициент.

$$Z = d_1 (x_N)_1 + \ldots + d_j (x_N)_j + \ldots + d_n (x_N)_n + Z_0$$

$d_j < 0$ следовательно можно увеличить x_N, тогда целевая функция начнет улучшаться.

$(x_B)_i \geq 0$, если $(x_B)_j = 0$, то дальше x_N увеличивать нельзя x_{Nj} и x_{Bj} меняются местами.

МЕТОДЫ И ТЕОРИЯ ОПТИМИЗАЦИИ

Методические указания к практическим занятиям
по дисциплине «**Методы и теория оптимизации**» для магистрантов 6 курса, обучающихся по направлению 221000.68 "Мехатроника и робототехника" по магистерской программе "Проектирование и исследование мультикоординатных электромехатронных систем движения"

Составитель Щербинин Сергей Васильевич

Подписано к печати
Формат 60х84/16. Бумага офсетная
Печать RISO. Усл.печ.л. Уч.-изд.л.
Тираж 50 экз. Заказ . Бесплатно