Конспект лекций по матанализу 2014 год, 2 курс
Лекция 1 (01 сентября 2014)

В этом модуле мы будем заниматься функциями многих переменных. В основном, функции все будут гладкими, то есть у них будут существовать производные, чаще всего этих производных будет как минимум одна, реже как минимум 2.

На первой лекции я не расскажу вам ничего особенно нового, однако как обзор она будет полезная.

Изложение содержание этого модуля в основном следует учебнику Зорича.

Все, что до дифференцируемых функций, непрерывность и прочее я напоминать не буду, считаю, что вы все это помните.

Дифференцирование отображений в конечномерных пространствах. Рассмотрим пространства $\mathbb{R}^m$ и $\mathbb{R}^n$ с некоторыми выбранными нормами. Так как это конечномерные пространства, то там все нормы эквивалентны: $c\|x\|_1 \leqslant \|x\|_2 \leqslant C\|x\|_1$.

Отображение $F : \mathbb{R}^m \to \mathbb{R}^n$. **Определение дифференцируемости** отображения в точке x: существует линейное отображение (я иногда буду говорить «*оператор*», это то же самое, что «*отображение*», в моих устах — полная тавтология) $A : \mathbb{R}^m \to \mathbb{R}^n$, такое что $F(x+h)-F(x)-Ah = o(\|h\|)$, то есть

$$\lim_{\|h\|\to 0} \|h\|^{-1} \|F(x+h) - F(x) - Ah\| = 0$$

$$\forall \varepsilon > 0 \ \exists \delta > 0 \ \forall h \in \mathbb{R}^n, \ \|h\| \leqslant \delta \ \text{ справедливо } \ \|h\|^{-1} \|F(x+h) - F(x) - Ah\| \leqslant \epsilon.$$

Подчеркнем, что от нормы производная не зависит.

Линейное отображение A называют по разному (дифференциал, касательно отображение, производная) и обозначают по-разному: $dF(x), DF(x), F'(x)$. Каждое такое обозначение — единый символ линейного оператора, он такой многобуквенный, для того, чтобы подчеркнуть, что это производная, от какой функции, в какой точке мы считаем эту производную. Обозначать все это принято так: $dF(x) \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ — пространство линейных операторов из $\mathbb{R}^m$ в $\mathbb{R}^n$. Значения производной — это векторы $dF(x)h, DF(x)h, F'(x)h$, h — это приращение аргумента, $F(x+h) - F(x)$ — приращение функции.

Сказать про оператор $x \mapsto \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$.

Заметим, что в простейшем случае, производной скалярной функции мы называли число. С точки зрения данного определения, это будет не число, а линейная операция умножения на это число. Ясно, что это формально разные вещи, но по сути — одинаковые.

Теперь я поговорю о важном понятии *касательное пространство*. Мы применяем линейный оператор дифференциал всегда к приращениям аргумента. Никогда не к x, всегда к h. Геометрически это означает, что вместо исходного пространства векторов x мы рассматриваем множество векторов h, торчащих из x. Его тоже можно рассматривать, как линейное пространство той же размерности. В $\mathbb{R}^m$ и в $T_x\mathbb{R}^m$ разные базисы.

Называется касательное пространство, обозначается $T_x\mathbb{R}^m$, $T\mathbb{R}^m(x)$, $T\mathbb{R}^n_x$. Буква T от слова *tangent — касательная, касательный*. Значение дифференциала $dF(x)$ на векторе $h \in T_x\mathbb{R}^m$ есть вектор $dF(x)h \in T_{F(x)}\mathbb{R}^n$.

Не путаем касательное пространство с касательной прямой-плоскостью-гиперплоскостью! Касательная плоскость — это чисто геометрический объект.

Частные производные.

До сих пор у нас были пространства $\mathbb{R}^n$ и $\mathbb{R}^m$ без базисов. Предположим, что мы ввели ортонормированные базисы в обоих этих пространствах, то есть отображение F имеет координатную форму $F(x) = (F^1(x), \dots, F^n(x))$ или $F^1(x_1, \dots, x_m), \dots, F^n(x_1, \dots, x_m)$, где $x = (x_1, \dots, x_m)$.

Утверждение. *Отображение F дифференцируемо в точку x, если и только если дифференцируемы все n скалярнозначных функций F^k.*

Доказательство я приводить не буду, вам его рассказывали, да оно и очевидное.

Теперь мы немножко поговорим о функциях со скалярными значениями. Об их производных, о возникающих при этом понятиях.

Пусть дана функция $f(x_1, \dots, x_m) \mapsto \mathbb{R}$.

Определение. *Частной производной f'_{x_k}* функции f по компоненте x_k называется предел

$$\lim_{h\to 0} h^{-1} \Big(f(x_1, \dots, x_k + h, \dots, x_m) - f(x_1, \dots, x_k, \dots, x_m) \Big),$$

если он существует.

Утверждение. *Если функция дифференцируемая в точке x, то существуют частные производные в точке x. Если функция имеет частные производные в окрестности точки x, и они непрерывны в x, то она дифференцируема в x.*

Это 2 теоремы, одна — о существовании частных производных у дифференцируемой функции, другая — о дифференцируемости функции при наличии непрерывных частных производных, вторая — несколько сложнее.

Соответственно, если частные производные непрерывны в области, то функция дифференцируема в области. Хорошие контрпримеры вы знаете (когда функция непрерывна, когда существуют частные производные в окрестности, но не существует дифференциала).

По определению производная функции — это линейный функционал (линейный функционал на векторном пространстве — это линейная функция на этом пространстве, принимающая скалярные значения). Всякий линейный функционал $a(x) \mapsto \mathbb{R}$ в конечномерном пространстве со скалярным произведением и ортонормированным базисом порождается некоторым вектором: $a(x) = (b, x)$.

Это простое утверждение: Вот есть базис e_k, Рассмотрим b вектор с координатами $a(e_k)$. Теперь для любого вектора $x = \sum x_k e_k$ значение функционала имеет вид

$$a(x) = a(\sum x_k e_k) = \sum a(x_k e_k) = \sum x_k a(e_k) = (x, b).$$

Иными словами, для всякой дифференцируемой в точке x функции f ее дифференциал имеет вид $df(x)h = (b, h)$, равенство справедливо при всех h. Вот этот вектор и называется градиентом функции f (в точке x), обозначается $\mathbf{grad}\, f(x)$. Легко посчитать, что вектор $\mathbf{grad}\, f(x)$ имеет компоненты $\partial f/\partial x_k$, он может быть записан в виде $\nabla f(x)$, где ∇ — это символ, обозначающий «вектор»:

$$\nabla = (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_m}).$$

Это не вектор, не элемент пространства — воспринимаем просто как обозначение, символ. Чисто мнемонически запоминаем формулу $\mathbf{grad}\, f = \nabla f$, как произведение вектора ∇ на скаляр f.

Производная произведения и частного:

$$(fg)' = f'g + fg', \qquad \left(\frac{f}{g}\right)' = \frac{1}{g^2}(gf' - fg')$$

Эти формулы были справедливы для обычных вещественных функций, они так же справедливы и для скалярных функций многих переменных, условие — дифференцируемость функций f и g и ненулевое значение g для частного. Здесь в формулах всюду f' и g' — это векторы, они умножаются и делятся на скаляры f и g.

Другой важный частный случай — это функции из $\mathbb{R}$ в $\mathbb{R}^n$. Для таких функций производная определяется проще всего.

Матрица Якоби. Якобиан. Называют якобианом и матрицу, и определитель (чаще). Матрица может быть и не квадратная, якобиан бывает только у квадратных матриц.

Производная — это линейный оператор из $\mathbb{R}^m$ в $\mathbb{R}^n$. Если в обоих этих пространствах выбраны базис, то каждый линейный оператор — это оператор умножения на $m \times n$ матрицу.

Как мы знаем, элементы этой матрицы задаются частными производными $\partial f/\partial x_k$:

$$\begin{pmatrix} \partial f_1/\partial x_1 & \ldots & \partial f_1/\partial x_m \\ \ldots & \ldots & \ldots \\ \partial f_n/\partial x_1 & \ldots & \partial f_n/\partial x_m \end{pmatrix}$$

Производная по вектору и по направлению.

Производная по вектору:

$$D_v f(x) := \lim_{t\to 0} \frac{f(x+vt) - f(x)}{t}$$

Производная по направлению — производная по вектору $e = (\cos\alpha_1, \ldots, \cos\alpha_m)$ единичной длины:

$$D_e f(x) = (\mathbf{grad}\, f, e) = \sum_{k=1}^{m} \frac{\partial f}{\partial x_k} \cos\alpha_k;$$

α_k — направляющие косинусы, углы α_k — это углы между e и базисными векторами e_k.

Частная производная — производная по направлению координатного вектора.

Характеристическое свойство градиента. *Направление градиента — направление наибольшей скорости функции.*

Дифференцирование композиции отображений.

Основной смысл теоремы: производная композиции равна композиции производных.

Теорема. *Пусть $f : X \to Y$, $g : Y \to Z$, причем отображение f — дифференцируемое в точке x, а g — дифференцируемое в точке $f(x)$, тогда композиция $g(f(x))$ дифференцируемое в точке x и $(g \circ f)' = g' \circ f'$.*

Это общий факт, он справедлив для производных операторов в любых линейных нормированных пространств. Если все расписывать в координатном виде, то получится произведение матриц Якобы, матрицы как раз окажутся нужного размера (они не квадратные), чтобы их можно было умножать.

Дифференцирование обратного отображения.

Эта теорема похожа на теорему о неявной функции, про которую я буду много говорить на следующем занятии. Теорема о неявной функции — это теорема о локальной однозначной разрешимости уравнения $f(x, y) = 0$. Там было примерно так: если $F(x_0, y_0) = 0$ и $F'(y) \neq 0$, то (при выполнении некоторых дополнительных существенных технических условий) локально $y = f(x)$.

Здесь все почти так же: уравнение $y - f(x) = 0$, векторы x и y одинаковой размерности, Если производная по y существует и обратима в некоторой точке, то существует ли обратный оператор? Не обязательно даже для функций одного переменного! Пример у вас был когда-то в листке:

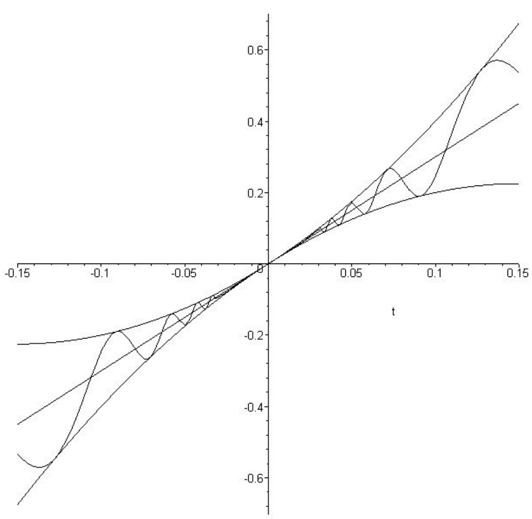

Рис. 1
Это график и функции $y = 3x + 10x^2 \sin(x^{-1})$ вспомогательных функций $y = 3x \pm 10x^2$ и $y = 3x$. В нуле у функции есть производная, она равна 3, однако в окрестности нуля функция колеблется и у нее обратной функции. А обратная функция есть только у монотонной функции.

Теорема. *Пусть f — отображение окрестности $U(x) \subset \mathbb{R}^m$ точки x на окрестность $V(y) \subset \mathbb{R}^m$ точки y. Пусть f непрерывно в окрестности точки x и имеет обратное отображение $f^{-1} : V(y) \to U(x)$, непрерывное в точке $y = f(x)$.*

Если при этом f дифференцируемо в точке x и производная $f' : T\mathbb{R}^m_x \to T\mathbb{R}^m_y$ обратима: существует $(f'(x))^{-1} : T\mathbb{R}^m_y \to T\mathbb{R}^n_y$, то отображение f^{-1} дифференцируемо в точке y и $(f^{-1})'(y) = (f'(x))^{-1}$.

Теорема о неявной функции. *Пусть задано отображение $F : U \subset \mathbb{R}^{m+n} \to \mathbb{R}^n$, U — окрестность точки (x_0, y_0). Пусть*

1) $F(x_0, y_0) = 0$;

2) $F \in C^p(U, \mathbb{R}^n)$, $p \geqslant 1$;

3) Дифференциал $F'_y(x_0, y_0)$ — обратимый линейный оператор;

Тогда существуют промежутки $I^m_x = \{x \in \mathbb{R}^m : \|x - x_0\| \leqslant \alpha\}$ и $I^n_y = \{y \in \mathbb{R}^n : \|y - y_0\| \leqslant \beta\}$ и такое отображение $f \in C^p(I^m_x, I^n_y)$, что что для любой точки $(x, y) \in I^m_x \times I^n_y$ справедливо равенство

$$F(x, y) = 0 \quad \Leftrightarrow y = f(x), \qquad f'(x) = -\big(F'_y(x, f(x))\big)^{-1} F'_x(x, f(x)).$$

Почти эту теорему мы еще раз будем доказывать на одной из двух последующих лекций. Последняя формула означает, что в некоторой окрестности точки (x_0, y_0) неявная функция определена, причем однозначным образом. Я через неделю докажу один из вариантов этой теоремы.

Лекция 2 (8 сентября 2014)

На прошлой лекции мы повторили дифференциалы функций многих переменных. Сегодня я плавно перейду к новым вещам.

Теоремы о конечных приращениях.

Теорема о среднем для скалярнозначных функций. *Пусть отображение f действует из области $D \subset \mathbb{R}^n$ в $\mathbb{R}$. Пусть $h \neq 0$ и $x, x+h \in D$ вместе с отрезком $[x, x+h]$. Пусть f дифференцируемо во всех точках $[x, x+h]$, тогда при некотором $\theta \in [0,1]$ справедливо равенство*

$$f(x+h) - f(x) = df(x+\theta h)h.$$

Равенство можно переписать в виде $f(x+h) - f(x) = (\nabla f(x+\theta h), h)$.

Доказательство. Рассмотрим скалярную функцию $F(t) = f(x+t\,h) : [0,1] \to \mathbb{R}$. Эта функция дифференцируема, для нее справедлива формула Лагранжа конечных приращений: $F(1)-F(0) = F'(\theta)$. Это равенство совпадает с требуемым в теореме. □

Отсюда следует важное **следствие.** *Если в линейно связной области D дифференциал функции равен нулю, то функция постоянная.*

Для выпуклой (или звездчатой) области этот факт очевиден, в любой паре точек функция принимает одинаковые значения, отрезок целиком лежит в области. Для невыпуклых областей ситуация сложнее.

Берем 2 произвольные точки в области. Пусть $x(t), t \in [0,1]$ — путь содержащийся в области, имеющий концами эти 2 точки. Образ компакта при непрерывном отображении — компакт. Покроем каждую точку этого образа–компакта шариком, целиком лежащим в области. Выберем конечное подпокрытие. В каждом шарике по следствию функция постоянна,

Легко видеть, отсюда следует, что функция постоянна везде. Красим точку $x(0)$ в красный цвет. На следующем шаге закрашиваем целиком каждый шарик, имеющий хоть одну красную точку. Во всех красных точках функция имеет одинаковое значение. Очевидно, что в результате конечного числа шагов процесс окраски стабилизируется. Если будут окрашены все шары, то все доказано.

Пусть в результате часть шаров будет закрашена, а часть — нет. Это противоречит замкнутости и связности пути и открытости шаров (берем первую неокрашенную точку, в любой ее окрестности есть окрашенные). □

Теперь перейдем к теореме о среднем для векторнозначных функций.

Во первых, мы знаем, что точно такая же теорема не верна. Берем функцию $f : \mathbb{R} \to \mathbb{R}^2$, определяемую равенствами $t \mapsto f(t) = (\cos t, \sin t)$, видим, что для нее ничего не получится: $|f'(t)| = 1, f(0) = f(\pi)$.

Однако, важная часть теоремы о среднем может быть сформулирована и для общего случая.

Сначала расскажу про **нормы линейных операторов**.

0. Я рассказываю об этом в первую очередь потому, что это абсолютно необходимое понятие «в математике». Оно будет вылезать в разных местах. В частности, здесь.

1. Линейные операторы из $\mathbb{R}^m$ в $\mathbb{R}^n$ образуют линейное пространство размерности mn. Вроде вы это знаете. Обозначается $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$.

2. В это пространстве можно вводить различные нормы, они все эквивалентны. Примеры: если в пространствах $\mathbb{R}^m$ и $\mathbb{R}^n$ введены базисы, то элементы $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ можно отождествить с матрицами, можно строить всякие нормы по элементам матриц. Например, корень из суммы квадратов, сумма модулей, максимальный модуль элемента. При этом утрачивается основная сущность пространства $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ — то, что это пространство линейных операторов.

3. Самая главная конструкция такая. Пусть в пространствах $\mathbb{R}^m$ и $\mathbb{R}^n$ введены нормы, пока что не важно какие. Тогда назовем нормой $A \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ величину

$$\|A\|_{\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)} = \sup_{e \in \mathbb{R}^m, \|e\|_{\mathbb{R}^m} = 1} \|Ae\|_{\mathbb{R}^n}.$$

То, что это норма, — считаем известным, проверить все свойства нормы (только у нуля норма 0, неравенство треугольника и линейность) легко.

Основное определения означает, что для всех $x \in \mathbb{R}^m$ справедливо неравенство $\|Ax\| \leqslant \|A\| \cdot \|x\|$, в котором все нормы имеют естественный смысл: $\|Ax\|_{\mathbb{R}^n} \leqslant \|A\|_{\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)} \cdot \|x\|_{\mathbb{R}^m}$.

4. В частности, дифференциал — это тоже линейный оператор. Значит у него тоже есть норма.

5. Замечу, что эта конструкция определения нормы годится для любых линейных нормированных пространств. В бесконечномерных пространствах бывают линейные ограниченные операторы, а бывают и неограниченные.

6. Впрок, без доказательства. Ограниченность оператора в нормированных пространствах эквивалентна его непрерывности. Непрерывность линейного оператора в одной точке эквивалентна непрерывности в каждой точке и эквивалентна равномерной непрерывности.

7. Пример неограниченного оператора — они чаще всего бывают определены именно на плотных множествах, но не на всем пространстве — оператор дифференцирования в C. Очевидно, что на единичном шаре в C есть дифференцируемые функции со сколь угодно большой производной (например, $\sin nt$ на $0, 2\pi$). Если оператор определен на всем пространстве, аддитивен и однороден, то он довольно часто оказывается ограниченным. Например, если оператор действует на всем гильбертовом пространстве (не буду останавливаться пока на точном определении), аддитивен, однороден и симметричен, то он ограничен.

Теорема о среднем. *Пусть отображение f действует из области $D \subset \mathbb{R}^m$ в $\mathbb{R}^n$. Пусть $h \neq 0$ и $x, x+h \in D$ вместе с отрезком $[x, x+h]$. Пусть f дифференцируемо во всех точках $[x, x+h]$ и пусть существует*

$$S = \sup_{\theta \in [0,1]} \|df(x+\theta h)\|_{\mathcal{L}_{\mathbb{R}^m \to \mathbb{R}^n}}.$$

Тогда $\|f(x+h) - f(x)\|_{\mathbb{R}^n} \leqslant S\,\|h\|_{\mathbb{R}^m}$.

Доказательство. Пусть это не так. Тогда при некотором $\varepsilon > 0$ справедливо неравенство $\|f(x+h)-f(x)\|_{\mathbb{R}^n} > (S+\varepsilon)\,\|h\|_{\mathbb{R}^m}$. Разобьем отрезок $[x, x+h]$ пополам, хотя бы на одном из отрезков $[x, x+h/2]$ или $[x+h/2, x+h]$ справедливо неравенство

$$\|f(x+h)-f(x+h/2)\|_{\mathbb{R}^n} > (S+\varepsilon)\,\|h/2\|_{\mathbb{R}^m}, \qquad \|f(x+h/2)-f(x)\|_{\mathbb{R}^n} > (S+\varepsilon)\,\|h/2\|_{\mathbb{R}^m}.$$

Разобьем такой промежуток на 2 части, снова на одном из частей будет выполнено аналогичное неравенство, продолжим эту процедуру далее, получим последовательность вложенных промежутков $[a_k, b_k] \subset [0,1]$, для которых при всех k справедливы неравенства $\|f(x+a_kh)-f(x+b_kh)\|_{\mathbb{R}^n} > (S+\varepsilon)\,|a_k-b_k|\,\|h\|_{\mathbb{R}^m}$. Эти отрезки имеют общую точку τ,

Более того, для каждого отрезка $[a_k, b_k]$, для которого $\tau \in (a_k, b_k)$ справедливо аналогичное неравенство, либо на отрезке $[\tau, b_k]$, либо на отрезке $[a_k, \tau]$. Теперь хотя бы одно из множеств таких невырожденных отрезков бесконечно, пусть, например, бесконечно множество стягивающих отрезков $[\tau, b_k]$:

$$\|f(x+\tau h)-f(x+b_kh)\|_{\mathbb{R}^n} > (S+\varepsilon)\,|b_k-\tau|\,\|h\|_{\mathbb{R}^m}.$$

А это неравенство противоречит неравенству $\|df(x+\tau h)\|_{\mathcal{L}_{\mathbb{R}^m\to\mathbb{R}^n}} \leqslant S$, так как

$$f(x+\tau h)-f(x+b_kh) = df(x+\tau h)h(b_k-\tau) + o(|b_k-\tau|).$$

□

Следствие. *Пусть отображение f действует из замкнутой выпуклой области $D \subset \mathbb{R}^m$ в $\mathbb{R}^n$. Пусть f дифференцируемо во всех точках D и пусть существует*

$$S = \sup_{x\in D} \|df(x)\|_{\mathcal{L}_{\mathbb{R}^m\to\mathbb{R}^n}}.$$

Тогда $\|f(x)-f(y)\|_{\mathbb{R}^n} \leqslant S\,\|x-y\|_{\mathbb{R}^m}$.

Обсудить теоремы о среднем. Сказать, что справедлива в любых банаховых пространствах (полные линейные нормированный пространства).

Принцип сжимающих отображений

Неподвижные точки, разрешимость уравнений. Примеры: непрерывное отображение отрезка в себя, непрерывное отображение шара или односвязной области в $\mathbb{R}^n$ в себя, отображение монотонным отображением отрезка в себя.

Пусть дано полное метрическое пространство X с метрикой $d(x,y)$.

Определение. Отображение $f : X \to X$ называется *сжимающим*, если оно удовлетворяет условию Липшица: $\forall x_1, x_2 \in X$ справедливо $d(f(x_1), f(x_2)) \leqslant q d(x_1, x_2)$ с константой Липшица $q < 1$.

Теорема. *Каждое сжимающее отображение f имеет единственную неподвижную точку $x_* = f(x_*)$.*

Доказательство. Рассмотрим произвольную точку $x_0 \in X$ и построим последовательность $x_{n+1} = f(x_n)$, $n \geqslant 0$. Эта последовательность фундаментальна.

Для доказательства фундаментальности напишем цепочку неравенств:

$$d(x_\ell, x_{\ell+1}) \leqslant q d(x_{\ell-1}, x_\ell) \leqslant q^2 d(x_{\ell-2}, x_{\ell-1}) \leqslant \ldots \leqslant q^\ell d(x_0, x_1).$$

Теперь для всех n, k

$$d(x_n, x_{n+k}) \leqslant \sum_{\ell=n}^{n+k-1} d(x_\ell, x_{\ell+1}) \leqslant \sum_{\ell=n}^{n+k-1} q^\ell d(x_0, x_1) = q^n(1+q+q^2+\ldots)d(x_0, x_1) \leqslant d(x_0, x_1)\frac{q^n}{1-q}.$$

Отсюда следует фундаментальность последовательности x_n, и, в силу полноты X, её сходимость к некоторому x_*. Теперь в силу непрерывности f имеем $x_* = \lim x_n = \lim x_{n+1} = \lim f(x_n) = f(\lim x_n) = f(x_*)$.

Нет двух предельных точек, очевидно: $d(x^*, x_*) = d(f(x^*), f(x_*)) \leqslant q\, d(x^*, x_*) \;\Rightarrow x^* = x_*$. □

Мы доказали в принципе сжимающих отображений, что последовательность $x_n = f(x_{n-1})$ при любом начальном значении x_0 сходится к единственной неподвижной точке x^*. В доказанной оценке

$$d(x_n, x_{n+k}) \leqslant d(x_0, x_1)\frac{q^n}{1-q}$$

можно перейти к пределу при $k \to \infty$:

$$d(x_n, x^*) \leqslant d(x_0, x_1)\frac{q^n}{1-q}.$$

Эта оценка означает, что сходимость последовательных приближений в условиях теоремы о сжимающих отображениях сходится экспоненциально, как q^n.

Обобщение. *Пусть некоторая степень f^n непрерывного отображения f — сжимающее отображение. Тогда f имеет единственную неподвижную точку $x_* = f(x_*)$.*

Для доказательства сначала рассмотрим неподвижную точку $x = f^n(x) = \lim_{k\to\infty} f^{kn}(x_0)$ отображения f^n. Покажем, что $x = f(x)$.

В силу непрерывности f имеем $f(x) = \lim_{k\to\infty} f^{kn}(f(x_0))$. Теперь

$$d\big(f^{kn+1}(x_0), f^{kn}(x_0)\big) \leqslant q\, d\big(f^{(k-1)n+1}(x_0), f^{(k-1)n}(x_0)\big) \leqslant \ldots \leqslant q^k d(f(x_0), x_0).$$

Следовательно, $d(f^{kn+1}(x_0), f^{kn}(x_0)) \to 0 \;\Rightarrow\; f(x) = x$. □

Параметрические варианты принципа сжимающих отображений.

Пусть отображение $f(x) = f(x,t) : X \to X$ зависит от параметра $t \in D$, где D — компактное метрическое пространство с метрикой ρ. Пусть отображение $f(x,t)$ непрерывно по совокупности переменных.

Пусть при всех значениях t отображения $f(x) = f(x,t) : X \to X$ сжимающие с общей константой Липшица $q < 1$. При каждом $t \in D$ сжимающее отображение $f(x,t)$ имеет единственную неподвижную точку $x_*(t) = f(x_*, t)$.

Неглавная **Теорема.** *Функция $t \mapsto x_*(t)$ непрерывна.*

Доказательство. Для доказательства надо, во-первых, рассмотреть пространство непрерывных функций $C(D)$ со значениями в пространстве X. Это полное метрическое пространство (это не было доказано, однако воспринимайте это как факт, доказательство полноты $C(D) = C(D, X)$ не сильно отличается от доказательства полноты $C(D, \mathbb{R})$, используется компактность D и полнота X). Так как при любых функциях z_1 и z_2 при каждом $t \in D$

$$\|f(z_1(t), t) - f(z_2(t), t)\|_X \leqslant q\|z_1(t) - z_2(t)\|_X,$$

то

$$\|f(z_1(t), t) - f(z_2(t), t)\|_{C(D,X)} \leqslant q\|z_1 - z_2\|_{C(D,X)}.$$

Это значит, что нелинейный оператор $z(t) \mapsto f(z(t), t)$ является сжимающим. □

Теорема об устойчивости неподвижной точки. Пусть отображение $f(x) = f(x, t) : X \to X$ зависит от параметра $t \in D$, где D — метрическое пространство с метрикой ρ (не компактное и даже не обязательно полное). Пусть снова при всех значениях t отображения $f(x) = f(x, t) : X \to X$ сжимающие с общей константой Липшица $q < 1$.

Пусть при каждом x отображение $f(x, t)$ как функция от t непрерывна в точке t_0.

Обозначим через $x^*(t)$ неподвижную точку отображения f.

Теорема. *Функция $t \mapsto x_*(t)$ непрерывна в точке t_0.*

Доказательство. При каждом t решение уравнения $x = f(x, t)$ может быть получено как предел последовательности $x_{n+1} = f(x_n, t)$, причем начальная точка x_0 может выбираться какая угодно. Пусть $x_0 = x^*(t_0) = f(x^*(t_0), t_0)$. Так как

$$d(x^*(t), x^*(t_0)) = d(x^*(t), x_0) \leqslant \frac{1}{1-q} d(x_1, x_0) = \frac{1}{1-q} d(f(x^*(t_0), t), f(x^*(t_0), t_0)) \to 0.$$

□

Лекция 3 (15 сентября 2014)

На прошлой лекции мы рассмотрели несколько важных вещей.

1) Теорема о конечных приращениях.

2) Норма линейного оператора.

3) Принцип сжимающих отображений, в том числе параметрические варианты.

Это все были не очень сложные вещи, но, как мне кажется, очень важные. Я немножко продолжу про принцип сжимающих отображений.

Аналогия с линейными уравнениями. Пусть в полном нормированном пространстве, например, в $\mathbb{R}^n$, дано линейное уравнение $x = Ax + b$. Тогда единственное решение существует, если и только если 1 не является собственным значением линейного оператора A, оно определяется соотношением $x = (Id - A)^{-1}b$.

Пусть $\|A\| < 1$. Тогда для отображения $(Id - A)^{-1}$ справедлива формула

$$(Id - A)^{-1} = \sum_{k=0}^{\infty} A^k.$$

Для доказательства достаточно умножить обе части равенства на невырожденную матрицу $Id - A$. Формула является полным аналогом суммы бесконечной геометрической прогрессии. Соответственно, для решения уравнения верна формула

$$(Id - A)^{-1}b = \sum_{k=0}^{\infty} A^k b.$$

Если взять любое начальное приближение x_0, а потом положить $x_n = Ax_{n-1} + b$, то получатся те самые приближения, что и в нелинейном случае: $x_0,\ Ax_0 + b,\ A^2x_0 + Ab + b,\ A^3x_0 + A^2b + Ab + b$ и так далее. Так как $A^k x_0 \to 0$, то получается естественная аналогия между линейным и нелинейным случаем.

Теорема о неявной функции. Рассмотрим пространства $X = \mathbb{R}^m$ и $Y = \mathbb{R}^n$ и уравнение $f(x, y) = 0$, где отображение f со значениями в Y определено и непрерывно на произведении каких-то подмножеств $D \subset X$ и $G \subset Y$.

Пусть (x_0, y_0) — внутренняя точка области определения и $f(x_0, y_0) = 0$, пусть отображение f дифференцируемо по y при каждом x, пусть «частная» производная $f'_y(x_0, y_0) : Y \to Y$ обратима. Пусть еще производная f'_y непрерывно в точке (x_0, y_0) зависит от x и y, как функция из $D \times G \to \mathcal{L}(Y, Y)$.

Последнее условие означает непрерывность функции $(x, y) \mapsto \mathcal{L}(X, Y)$. Конечно, норма в конечномерном пространстве $\mathcal{L}(X, Y)$ при $X = \mathbb{R}^m$ и $Y = \mathbb{R}^n$ может быть выбрана любой, они все эквивалентны, однако про себя мы помним, что норма в этом пространстве операторная, такая как я рассказывал на прошлой лекции.

Теорема. *Существует непрерывная неявная функция.*

1) Уравнение $f(x, y) = 0$ в Y эквивалентно уравнению $\left(f'_y(x_0, y_0)\right)^{-1} f(x, y) = 0$. Здесь обратимый линейный оператор $\left(f'_y(x_0, y_0)\right)^{-1}$ (он не зависит ни от x, ни от y) применяется к f.

2) Уравнение $\big(f'_y(x_0,y_0)\big)^{-1}f(x,y)=0$ эквивалентно уравнению $y=T(x,y)$, где $T(x,y)=y-\big(f'_y(x_0,y_0)\big)^{-1}f(x,y)$. Очевидно.

3) Производная T'_y в точке (x_0,y_0) равна 0 — это легко посчитать:

$$T'_y(x_0,y_0)=Id-\big(f'_y(x_0,y_0)\big)^{-1}f'_y(x_0,y_0)=0.$$

4) Так как производная f'_y непрерывна в точке (x_0,y_0), то можно выбрать такие окрестности $\|x-x_0\|_X\leqslant\rho$ по x и $\|y-y_0\|_Y\leqslant r$ по y точки (x_0,y_0), что $\|T'_y(x,y)\|_{\mathcal{L}(Y,Y)}\leqslant 1/4$.

5) По вышеизложенному $\|T(x,y_1)-T(x,y_2)\|_Y\leqslant(1/4)\|y_1-y_2\|_Y$, то есть оператор T сжимающий. Значит, при каждом x существует единственный y, по доказанному ранее этот y непрерывно зависит от x.

6) Это было бы хорошо — кто-то видит ошибку? Ошибка в том, что мы не увидели, что какое-то множество (метрическое пространство) отображается в себя.

7) Мы будем выбирать в качестве метрического пространства шар в Y с центром в y_0 и радиуса r. Так как $y_0=T(x_0,y_0)$ и

$$\|y_0-T(x,y)\|_Y=\|T(x_0,y_0)-T(x,y)\|_Y\leqslant\|T(x_0,y_0)-T(x,y_0)\|_Y+\|T(x,y_0)-T(x,y)\|_Y,$$

то при $\|T(x_0,y_0)-T(x,y_0)\|_Y\leqslant r/4$ (а так будет при достаточно малом $\|x-x_0\|_X$)

$$\|y_0-T(x,y)\|_Y\leqslant r/4+(1/4)\|y-y_0\|_Y\leqslant r$$

и шар $\|y-y_0\|_Y\leqslant r$ преобразуется оператором T в себя. $\square$

Замечания.

1. Можно было бы такую теорему о неявной функции доказать, рассмотрев оператор в пространстве непрерывных функций.

2. Для доказательства разрешимости уравнения используется метод неподвижной точки. Есть еще всякие разные методы, я сейчас скажу об одном из них, в Зориче все написано подробно, 1 страница.

Дано уравнение $F(x,y)=0$. Рассмотрим скалярнозначную функцию $\varphi(x,y)=\|F(x,y\|_Y^2$, сумму квадратов компонент. В точке (x_0,y_0) эта функция обращается в ноль, а окрестности точки (x_0,y_0) (так как F'_y не вырожденная матрица), эта функция примерно квадратичная.

Поэтому при каждом x из окрестности у нее есть единственный минимум, в котором значение квадратичной функции равно нулю. Вот так и определяется неявная функция.

Такой подход называется вариационным подходом и используется в разных областях современной математики. В частности, он позволяет доказывать разрешимость всяких хитрых дифференциальных уравнений и краевых задач.

3. Замечание про линейный случай. Пусть отображение F — линейное, $x_0=y_0=0$.

Это значит, что задана система из n уравнений с $n+m$ неизвестными, m штук называются x и n штук называются y. Утверждение теоремы о неявной функции означает, что если определитель по y отличен от нуля, то y можно выразить через x. В линейном случае это факт, который вы проходили по линейной алгебре: есть уравнение $Ax + By = 0$, матрица B — квадратная невырожденная, тогда определено отображение $y = -B^{-1}Ax$, при каждом x соответствующий y решает уравнение.

4. Рассуждение про формулу для производной. Я доказал только что теорему о существовании непрерывной неявной функции, при этом я использовал только существование непрерывной обратимой производной по y. Пусть $F \in C^1$ непрерывно дифференцируемо по x и по y.

Тогда из уравнения $F(x, y) = 0$ следует равенство $F'_x(x, y)(x - x_0) + F'_y(x, y)(y - y_0) +$ малые слагаемые $= 0$. Таким образом, $y = y_0 - \big(F'_y(x, y)\big)^{-1} F'_x(x, y)(x - x_0) +$ малые слагаемые. Это и есть формула для производной y'_x.

Старшие производные.

Что такое вторая производная отображения? Это должна быть производная от производной.

То есть конструкция должна быть такая. Вот есть отображение f, где-то определенное на $\mathbb{R}^m$ и со значениями в каком-то пространстве $\mathbb{R}^n$. Пусть оно в каждой точке дифференцируемое. Тогда в каждой точке определена производная $f'(x)$, это линейный оператор на соответствующих касательных пространствах.

Соответственно, определено отображение $x \mapsto f'(x)$ из пространства $\mathbb{R}^m$ в пространство линейных операторов $\mathcal{L}(\mathbb{R}^m_x, \mathbb{R}^n_x)$. Это конечномерное линейное пространство, его размерность равна mn.

Если там ввести удобную метрику, то можно исследовать непрерывность оператора $x \mapsto f'(x)$, его свойства, можно изучать его дифференцируемость.

Соответственно, вторая производная отображения $x \mapsto f'(x)$ — это будет линейный оператор из $\mathbb{R}^m$ в $\mathbb{R}^{mn}$... и так далее. Мы пока не будем изучать эти конструкции, они в этом модуле нам почти не понадобятся.

Вторую производную $f''(x_0)$ естественно представлять как билинейное отображение $f'' h_1 h_2$, применяем $f''(x_0)$ к вектору из касательного пространства, получаем линейный оператор. применяем его к другому вектору из касательного пространства — получаем вектор из $\mathbb{R}^n$ (из соответствующего другого касательного пространства).

Понадобится только пространство $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ и норма в нём.

Однако, если предположить, что во всех пространствах есть система координат, то можно рассматривать обычные частные производные любых порядков и в любом сочетании.

Я буду обозначать их по-разному, частная производная функции $f(x_1, \ldots, x_m)$ по переменной x_k будет обозначаться

$$f'_{x_k}, \quad \partial_k f, \quad \frac{\partial f}{\partial x_k} \quad \text{и даже (чтобы штрихи не писать!)} \quad f_{x_k}.$$

Их легко определить для $f : \mathbb{R}^m \to \mathbb{R}$ по индукции: если определена производная $\partial_{i_1,\ldots,i_k} f$

функции f порядка k по переменным $i_1, \ldots, i_k$ (индексы могут повторяться), то производную $(k+1)$-го порядка по переменной x_s определим равенством $\partial_{s,i_1,\ldots,i_k} f = \partial_s(\partial_{i_1,\ldots,i_k} f)$.

Старшие производные перестановочны.

Теорема. *Пусть функция $f : \mathbb{R}^m \to \mathbb{R}$ имеет в области $G \subset \mathbb{R}^m$ частные производные второго порядка $\partial_{i_1,i_2} f$ и $\partial_{i_2,i_1} f$. Если обе эти производные непрерывны в некоторой точке $x \in G$, то они совпадают в этой точке.*

Доказательство этой теоремы не слишком громоздкое, и не слишком сложное, вам его уже рассказывали, я не буду его повторять.

Замечу, что условия теоремы все существенные. Нельзя отказаться от непрерывности, нельзя отказаться от дифференцируемости в окрестности.

В частности, обращаю внимание, что в теореме используются не только переменные с индексами i_1 и i_2. Неявно используются и все остальные переменные: все функции и производные обязаны быть непрерывными по всем переменным.

Следствие. *Если функция $f \in C^k(G;\mathbb{R})$ (напомнить обозначение!), то значение частной производной $\partial_{i_1,\ldots,i_k} f(x)$ не зависит от порядка индексов.*

Теперь мы сделаем некоторые выкладки, которые позволят нам сформулировать формулу Тейлора для скалярнозначных функций многих переменных.

1. Пусть дана функция $f(x_1, x_2)$ класса C^k в некоторой области G (это значит, что у функции есть все частные производные до порядка k включительно, причем все они непрерывны).

Пусть $h = (h_1, h_2)$ такое, что отрезок $[x, x+h]$ содержится области G. рассмотрим функцию

$$\varphi(t) = f(x+th) : [0,1] \to \mathbb{R}, \quad f \in C^k.$$

Посчитаем производные этой функции с помощью теорем о производной сложной функции:

$$\varphi'(t) = \partial_1 f(x_1 + th_1, x_2 + th_2)h_1 + \partial_2 f(x_1 + th_1, x_2 + th_2)h_2$$

(здесь обозначено $\partial_i = \partial/\partial x_i$)

$$\varphi''(t) = \partial_{11} f(x+th)(h_1)^2 + 2\partial_{12} f(x+th)h_1h_2 + \partial_{22} f(x+th)(h_2)^2.$$

Если аккуратно еще раз продифференцировать, получится

$$\varphi^{(p)}(t) = (h_1\partial_1 + h_2\partial_2)^p f(x+th) = (\nabla, h)^p f(x+th).$$

2. Выражение $(\nabla, h)^p f$ — это форма записи. Здесь дифференциальные операторы (в этом термин я не вношу никакого глубокого смысла, просто — оператор, содержащий дифференцирование) $(\nabla, h)^p$ осмысленные, так как операторы ∂_ℓ коммутируют для разных ℓ.

3. Теперь пусть функция от m переменных. Тогда снова $\varphi(t) = f(x+th) : [0,1] \to \mathbb{R}, \quad f \in C^k$, и формулу также можно переписать в виде

$$\varphi^{(p)}(t) = \Big(\sum_{i=1}^m h_i\partial_i\Big)^p f(x+th) = (\nabla, h)^p f(x+th).$$

4. Есть еще один способ записывать такие громоздкие формулы. Считается, что если индекс входит в формулу 2 раза, причем один раз в виде нижнего индекса, а другой раз — в виде верхнего индекса, то по этому индексу происходит суммирование. Это часто бывает удобно, если верхний индекс нельзя спутать с показателем степени и если из контекста ясен диапазон суммирования. Формулу для производных функции φ можно переписать тогда

$$\varphi^{(p)}(t) = h^{i_1}, \dots, h^{i_p} \partial_{i_1,\dots,i_p} f(x + th).$$

Здесь индексы у координат вектора h подняты наверх, суммирование идет по p переменным i_p.

Теперь перейдем к формуле Тейлора.

Теорема. *Пусть* $f : U(x) \to \mathbb{R}$, $U \subset \mathbb{R}^m$ *и пусть* $f \in C^k$, *и пусть* $[x, x+h] \in U$. *Тогда*

$$f(x+h) - f(x) = \sum_{\ell=1}^{\ell-1} \frac{1}{\ell!} (\nabla, h)^\ell f(x) + r_{k-1}(x,h), \qquad r_{k-1}(x,h) = \int_0^1 \frac{(1-t)^k}{(k-1)!} (\nabla, h)^k f(x+th)\, dt.$$

Доказательство. Рассматриваем функцию $\varphi(t) = f(x+th) \in C^k$, для нее пишем формулу Тейлора (это функция одной переменной)

$$\varphi(t) = \varphi(0) + \frac{\varphi'(0)}{1!} t + \ldots + \frac{\varphi^{(k-1)}(0)}{(k-1)!} t^{k-1} + \int_0^1 \frac{(1-\tau)^k}{(k-1)!} \varphi^{(k)}(\tau)\, d\tau.$$

Я рассказывал эту формулу сразу после интегрирования по частям. Теперь полагаем $t = 1$ и пользуемся формулами для производных функции φ, которые я раньше предусмотрительно получил. □

Естественно, из интегральной формулы для остаточного члена $r_{k-1}(x,h)$ вытекают иные формы. Например, форма Лагранжа:

$$r_{k-1}(x,h) = \frac{1}{k!} (\nabla, h)^k f(x + \theta h)$$

(используем теорему о среднем) или форма Пеано: $r_{k-1}(x,h) = o(\|h\|^k_{\mathbb{R}^m}$.

Экстремумы функций многих переменных.

Пусть задана функция $f : U(x) \to \mathbb{R}$, $U \subset \mathbb{R}^m$. Мы будем думать про себя, что эта функция гладкая, для определений это не нужно, но мы будем считать, что у нее есть «нужное количество» непрерывных производных. Иногда одна, иногда две.

Теперь мы займемся изучением минимумов и максимумов этой функции, а также седловых точек функций многих переменных.

Определения. 1. *Локальный максимум.* Если у точки x_0 есть окрестность, в которой выполнено неравенство $f(x) \leqslant f(x_0)$, то это локальный максимум. Если в проколотой окрестности выполнено более сильное неравенство $f(x) < f(x_0)$, то это строгий локальный максимум.

2. *Локальный минимум* и строгий локальный минимум. Определяется совершенно аналогично.

3. *Локальный экстремум.* Общее слово для минимумов и максимумов — экстремум.

4. *Критическая точка функции.* Точка x называется критической точкой функции, если $\mathbf{grad}\, f(x) = 0$, то есть $f'_{x_i}(x) = 0$ при всех $i = 1, \ldots, m$.

Критические точки также называют *стационарными* точками.

5. *Седловая точка.* Обезьянье седло. Критическая точка функции двух переменных, не являющаяся экстремумом, является-называется седловой точкой. Седловая точка является аналогом точки перегиба для функций одной переменной.

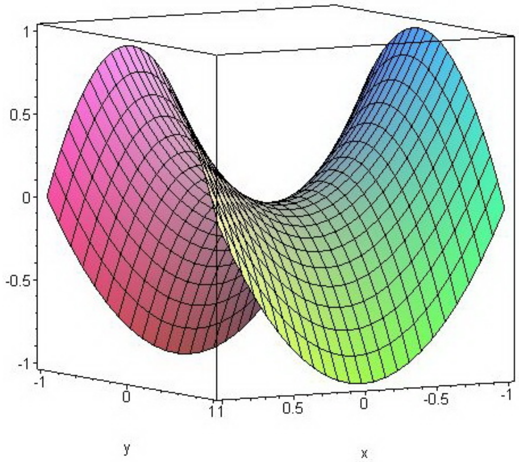

Рис. 2
Это график функции $z = x^2 - y^2$. Начало координат — седловая точка.

6. Глобальный минимум. **Вопрос к залу:** может ли быть гладкая функция на плоскости, у которой есть ровно одна критическая точка, это точка локального минимума, и чтобы этот локальный минимум не был глобальным?

В случае скалярных функций так не бывает, а для плоскости — бывает.

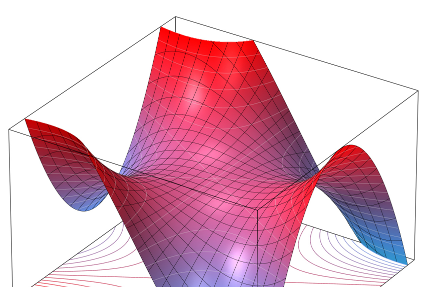

Рис. 3
Это график функции $z = x^3 - 3xy^2$. Начало координат — обезьянье седло.

7. *Критическая точка отображения.* Дам еще определение «впрок». Пусть задано отображение $f : \mathbb{R}^m \to \mathbb{R}^n$, причем пусть в обоих пространствах зафиксированы базисы. Тогда производная f' — это её матрица Якоби. Вообще говор, естественно ожидать, что ранг матрицы размером $m \times n$ равен $\min\{m, n\}$ Точка называется критической, если ранг этой матрицы меньше $\min\{m, n\}$.

Использовать это понятие буду позднее. Оно имеет чисто геометрический смысл.

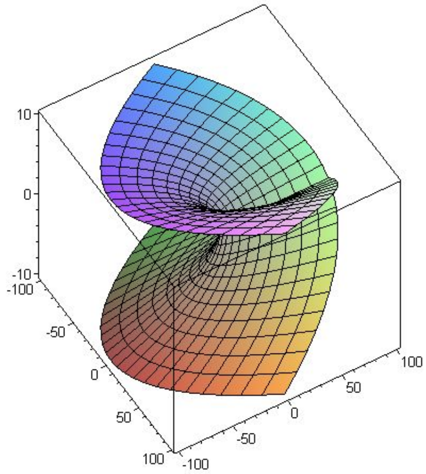

Рис. 4
Это образ квадрата $[-10, 10] \times [-10, 10] \subset \mathbb{R}^2$ при отображении $(u, v) \mapsto (uv, u^2 - v^2, v)$. Начало координат — критическая точка, в ней ранг матрицу Якоби равен 1, во всех остальных точках он равен 2.

Теорема (необходимое условие локального экстремума). *Пусть функция $f : \mathbb{R}^m \to \mathbb{R}$ имеет во внутренней точке x области определения все частные производные (по каждой из переменных). Пусть у f точка x является локальным экстремумом. Тогда $\partial_i f(x) = 0$, $i = 1, \ldots, m$.*

Выберем $i = 1, \ldots, m$ и рассмотрим скалярную функцию $\psi(t) = f(x_1, \ldots, x_i + t, \ldots, x_m)$. Эта функция по предположению имеет при $t = 0$ локальный экстремум, ее производная $\psi'(0) = 0$, поэтому $\partial_i f(x) = 0$. $\square$

Иными словами, каждая точка экстремума дифференцируемой функции — критическая. Обратное, конечно, не верно. У скалярной функции $f(x) = x^3$ ноль — критическая точка, не являющаяся локальным экстремумом. Седловые точки не являются точками локального экстремума.

Из этой теоремы естественным путем следует возможность нахождения локальных экстремумов: пишем систему из m скалярных уравнений $\mathbf{grad}\, f(x) = 0$ с m неизвестными (координатами точки x). Находим решения и дальше как-то надо исследовать полученные точки.

В скалярном случае, если помните, удавалось воспользоваться второй производной: если $f'' > 0$, то это был минимум, если $f'' < 0$, то это был максимум, случай $f'' = 0$ требовал рассмотрения старших производных.

Лекция 4 (22 сентября 2014)

На прошлой лекции мы рассмотрели несколько важных вещей.

1) Я доказал один из вариантов теоремы о неявной функции

2) Дальше мы поговорили о старших производных, что это и как

3) Потом была формула Тейлора для скалярнозначных функций многих переменных

4) Было введено понятие критической точки, приведены необходимые и достаточные условия локального экстремума. Достаточные условия я не доказал, с этого и начну эту лекцию.

Перед тем я решу одну простую задачу с семинара. Она вызвала проблемы у части студентов, ну вот я ее и решу.

Задача такая: есть функция двух переменных $f(x, y) = \cos(x^2 + xy)$. Найти производную $\partial^{60} f / \partial x^{51} / \partial y^9$.

1) Раскладываем в ряд Тейлора $\cos(\cdot)$ в нуле:

$$f(x, y) = 1 - \frac{1}{2}(x^2 + xy)^2 + \frac{1}{24}(x^2 + xy)^4 - \ldots$$

2) Когда мы все это будем дифференцировать, то единственное слагаемое, которое останется в нуле, — это слагаемое $cx^{51}y^9$.

3) Надо найти коэффициент c при этом слагаемом... ясно, что он появится группы слагаемых $-\frac{1}{30!}(x^2 + xy)^{30}$ 60й степени.

4) Наше слагаемое — это $-\frac{1}{30!}C_{30}^9 x^{51}y^9$, так как $x^{51}y^9 = (x^2)^{21} \cdot (xy)^9$.

5) Вот его и надо продифференцировать... Ответ

$$-\frac{\partial^{60} f}{\partial x^{51} \partial y^9} = -\frac{\partial^{60}(\frac{1}{30!}C_{30}^9 x^{51} y^9)}{\partial x^{51} \partial y^9} = -\frac{30!\ 51!\ 9!}{30!\ 9!\ 21!} = -\frac{51!}{21!}.$$

Теорема (необходимое условие локального экстремума). *Если x_0 — локальный экстремум f, то x_0 — критическая точка f $(\partial_i f(x_0) = 0.)$*

Теорема (достаточное условие локального экстремума). *Пусть $f \in C^2$, пусть точка x_0 — критическая. Рассмотрим квадратичную форму*

$$A(h) = \sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) h^i h^j \equiv \partial_{ij} f(x_0) h^i h^j$$

переменных h^ℓ. Если эта квадратичная форма определена строго положительно, то точка x_0 является строгим локальным минимумом. Если эта квадратичная форма определена строго отрицательно, то точка x_0 является строгим локальным максимумом. Если форма принимает значения разных знаков, то точка x_0 не является точкой экстремума.

Квадратичная форма определена строго положительно, это значит, что ее значения всюду положительны, кроме нуля. Форма x^2 на плоскости (x, y) не является строго положительной: она равна нулю при $x = 0$ и любом y. Такие случаи теорема не охватывает.

В случае не строго определенных квадратичных форм сделать вывод (экстремум или нет) нельзя. Примеры легко придумать: $f(x, y) = x^2 - y^4$ и $f(x, y) = x^2 + y^4$, в обоих этих примерах форма $A(h_1, h_2) = h_1^2$ не является строго положительной, в одном случае в нуле локальный минимум, в другом — седло.

Доказательство. Рассмотрим форму A на сфере $h \in S(0, 1) \in \mathbb{R}^m$. Так как A непрерывная функция и так как $S(0, 1)$ — компактное множество, то существуют конечные величины $m = \inf A(h)$ и $M = \sup A(h)$.

Запишем формулу Тейлора с остаточным членом в форме Пеано:

$$f(x_0 + h) - f(x_0) = \frac{\|h^2\|}{2!}\big(A(h) + o(1)\big).$$

Путь квадратичная форма A положительно определена. Это значит, что $m > 0$, следовательно при малых $\|h\| > 0$ правая часть формулы Тейлора положительна, значит, точка x_0 является строгим локальным минимумом.

Аналогичное рассуждение в случае, когда квадратичная форма A отрицательно определена. В этом случае $M < 0$, и точка x_0 является строгим локальным максимумом.

Если форма принимает и положительные, и отрицательные значения, то $m < 0 < M$. В этом случае существуют точки $h_-, h_+ \in S(0, 1)$, в которых $\pm A(h_\pm) > 0$. Тогда при достаточно малых $\theta > 0$ при $h = \theta h_\pm$ знак правой части формулы Тейлора совпадает со знаком формы $A(h_\pm$ то есть принимает и отрицательные, и положительные значения. Поэтому x_0 не является экстремумом.
□

Как определить, является ли форма положительно определенной? Есть критерий Сильвестра, можно найти все собственные значения и увидеть, что они все положительные, но это все вычислительно сложные задачи.

Метод наименьших квадратов. Линейная регрессия. Я быстро расскажу об одном приложении полученных теорем. Когда-то меня этому научили врачи из кардиологического центра, конечно, сразу оказалось что это все знают. Но я после мехмата как-то не знал.

Пусть вы врач, у вас есть данные каких-то измерений, вы меряете 2 величины x и y (или 3, или m), но я буду считать, что их 2. И делаете n одновременных измерений этих величин.

По вашим медицинским соображением вы считаете, что эти величины связаны между собой линейно: $y = ax + b$.

Задача: найти коэффициенты a и b, связывающие ваши величины.

Обычно, в «практической жизни» это делают так. Врач нажимает кнопки на компьютере, вводит свои данные, 2 вектора x_i и y_i, $i = 1, \ldots, n$. А компьютер выдают врачу приблизительные значения коэффициентов a и b, которые связывают величины x и y. Компьютер выдает еще несколько величин, описывающих надежность определения этих коэффициентов, но это выходит за рамки нашего повествования.

Как компьютер считает оптимальные значения a и b? Предлагается такой метод.

Составляем сумму

$$f(a,b)=\sum_{i=1}^{n}(y_i-ax_i-b)^2.$$

Это квадратичная форма, у нее есть минимум (мы это знаем по смыслу), для его определения надо написать 2 уравнения $f'_a=f_b=0$ с неизвестными a и b, эти уравнения имеют вид

$$\sum_{i=1}^{n}2(y_i-ax_i-b)x_i=\sum_{i=1}^{n}2(y_i-ax_i-b)=0.$$

Это система линейных уравнений

$$a\sum x_i^2+b\sum x_i=\sum x_iy_i, \qquad a\sum x_i+nb=\sum y_i.$$

Решаем, находим значения a и b.

Диффеоморфизмы.

Теперь мы временно откладываем экстремумы и начинаем заниматься приложениями теоремы о неявной функции. Это будет ряд довольно громоздких, хотя и философски очень простых утверждений.

Определения.

Начнем с мегаважного определения.

Отображение $f:U\to V$, областей $U,V\subset\mathbb{R}^m$ называется *диффеоморфизмом класса* p, если выполнены 3 условия

1) f — взаимно однозначное отображение (биекция);

2) $f\in C^p(U,V)$,

3) $f^{-1}\in C^p(V,U)$.

Диффеоморфизм класса 0 называется *гомеоморфизм.*

Примеры.

1. Линейное отображение является диффеоморфизмом, если и только если матрица квадратная и определитель отличен от нуля, вроде это очевидно.

2. Отображение $x\mapsto x^3$ любой области в $\mathbb{R}$, содержащей 0, является гомеоморфизмом, но не является диффеоморфизмом класса 1.

3. Проверка условий 1 и 3 сложная!

Теорема о диффеоморфизме (об обратной функции). *Если $f\in C^p(\mathbb{R}^m,\mathbb{R}^m)$, $p>0$ — отображение, причем $y_0=f(x_0)$ и $f'(x_0)$ — обратимое линейное отображение, то существуют* *окрестности точек* x_0 *и* y_0, *в окрестности которых* f *является диффеоморфизмом класса* p, *при этом* $\left(f^{-1}\right)'(y) = \left(f'\right)^{-1}(x)$.

Доказательство простое — это следствие теоремы о неявной функции.

Рассмотрим уравнение $y = f(x) \;\Leftrightarrow\; F(x, y) = f(x) - y = 0$. Мы хотим разрешить это уравнение относительно переменной x в окрестности некоторой точки $(x_0, y_0) \in \mathbb{R}^m \times \mathbb{R}^m$.

По предположению выполнены все условия теоремы о неявной функции, $F'_x(x_0, y_0) = f'(x_0)$ — обратимое отображение.

В силу теоремы о неявной функции отсюда следует существование отображения $x = g(y)$, удовлетворяющего $x = g(y) \;\Leftrightarrow\; y = f(x)$, отсюда же следует требуемая гладкость $g = f^{-1}$ и равенство про производные:

$$g'(y) = -\big(F'_x(x, y)\big)^{-1}\big(F'_y(x, y)\big),\ F'_x(x, y) = f'(x),\ F'_y(x, y) = -Id \quad\Rightarrow\quad g'(y) = \big(f'(x)\big)^{-1}.$$

Осталось сказать, что окрестности выбираются тоже в соответствии с теоремой о неявной функции. □

Эту теорему следует понимать как критерий локального диффеоморфизма: если есть гладкое отображение и его производная в некоторой точке обратима, то это диффеоморфизм.

Естественно, при этом должны совпадать размерности области определения и пространства образа, иначе не возможны условия теоремы (обратимы только квадратные матрицы).

Правильно понимать также эту теорему как аналог соответствующего линейного утверждения: если линейное отображение обратимо, то оно — диффеоморфизм. Локально гладкое отображение ведет себя примерно как его дифференциал.

Диффеоморфизмы используются (в частности) для замены переменных.

Если есть диффеоморфизм $\phi : U \to V$, то говорят, что на U заданы другие координаты. Был вектор координат u, а теперь можно говорить о координатах v, зная, что u вычисляются через v однозначно. Обычно все это делается локально, в окрестности некоторой точки.

Приведу пример использования теоремы о диффеоморфизме для выпрямления кривых.

Путь есть кривая на плоскости, заданная скалярным уравнением $F(x, y) = 0$. Пусть F — гладкая функция, пусть $F(x_0, y_0) = 0$ и пусть точка (x_0, y_0) не является критической: то есть ранг отображения $DF(x_0, y_0)$ равен 1, то есть по крайней мере одна из производных $F'_x(x_0, y_0)$ или $F'_y(x_0, y_0)$ отлична от нуля.

Я буду предполагать, что $F'_y(x_0, y_0) \neq 0$.

Я укажу такой диффеоморфизм окрестности точки (x_0, y_0) на окрестность нуля, что в новых координатах (ξ, η) кривая будет иметь уравнение существенно более простого вида: $\eta = 0$.

На вопрос: «Зачем это надо?» я не буду отвечать, однако, вроде ясно, что в каких-то конструкциях это удобно.

Положим $\xi = x - x_0,\ \eta = F(x, y)$. При таком отображении $\Phi : (x, y) \mapsto (\xi, \eta)$ точка (x_0, y_0) перейдет в начало координат, матрица Якоби этого отображения имеет вид

$$\begin{pmatrix} 1 & 0 \\ F'_x & F'_y \end{pmatrix},$$

якобиан равен $F'_y(x, y)$ и отличен от нуля в точке (x_0, y_0).

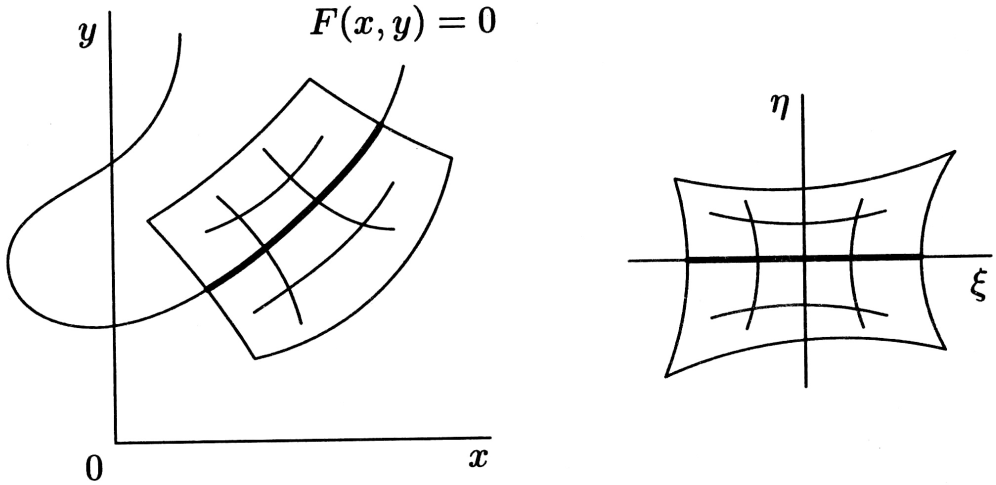

Рис. 5

По доказанной теореме отображение Φ является диффеоморфизмом, у него есть обратное отображение $\Phi^{-1} : (\xi, \eta) \mapsto (x, y)$, кривая $F(x, y) = 0$ в новых координатах (ξ, η) имеет требуемый вид $\eta = 0$ (конечно, только локально, в некоторой окрестности начала координат).

Теперь пусть всё то же самое, уравнение от многих переменных в пространстве $\mathbb{R}$, то есть $F : \mathbb{R}^{m+1} \to \mathbb{R}$, одно скалярное уравнение, переменных $m+1$ — множество решений — поверхность коразмерности 1. Будем считать, что $x \in \mathbb{R}^m,\ y \in \mathbb{R}$, уравнение имеет вид $F(x, y) = 0$.

Пусть $F(x_0, y_0) = 0$ и $F'_y(x_0, y_0) \neq 0$. Тогда отображение $\xi = x - x_0,\ \eta = F(x, y)$ имеет матрицу Якоби

$$\begin{pmatrix} Id_m & 0 \\ F'_x & F'_y \end{pmatrix},$$

Здесь Id_m — тождественное преобразование в $\mathbb{R}^m$, F'_x — вектор-строка, 0 — вектор-столбец с нулевыми координатами. Это — треугольная матрица и её определитель равен F'_y. Снова по теореме о диффеоморфизме отображение $(x, y) \mapsto (\xi, \eta)$ — диффеоморфизм, В координатах (ξ, η) поверхность приобретает локальный вид $\eta = 0$.

Пытливый слушатель спросит, что будет, если $F(x, y) = 0$, $F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, $F(x_0, y_0) = 0$ и матрица $F'_y(x_0, y_0)$ обратима. А то же самое будет. В пространстве размерности 5 задано 3 уравнения. Фиксируем точку двумерной поверхности, предполагаем, что последние 3 переменных выбраны так, что матрица Якоби по последним трем переменным обратима. Снова есть диффеоморфизм окрестности в окрестности начала координат, такой что поверхность имеет вид $(\xi_1, \xi_2, 0, 0, 0)$.

Конечно, такое уже нарисовать не получится. И представить себе такое геометрически я не умею. Однако, аналитически, это хорошо видно. 5 уравнений, 3 хороших переменных, относительно них систему уравнений можно разрешить, 2 параметра остались. Вот и получился такой вид: $(\xi_1, \xi_2, 0, 0, 0)$.

Локальное приведение гладкого отображения к каноническому виду.

Напомню, что ранг матрицы — это размер самого большого невырожденного минора. Если мы умножим матрицу ранга k слева или справа на невырожденную квадратную матрицу, то ранг матрицы не изменится. Это Вы должны были проходить по линейной алгебре и я считаю это известным.

Тем самым корректно говорить о ранге $\operatorname{rang} A$ линейного отображения $A : \mathbb{R}^m \to \mathbb{R}^n$, даже если базисы в пространствах $\mathbb{R}^m$ и $\mathbb{R}^n$ не выбраны и матрица отображения не определена.

Пусть задано нелинейное гладкое отображение $f : \mathbb{R}^m \to \mathbb{R}^n$. Будем называть *рангом* $\operatorname{rang} f(x)$ *отображения* f *в точке* x ранг его дифференциала $f'(x)$.

Естественно, ранг нелинейного отображения может меняться в зависимости от точки x. В нормальной ситуации $\operatorname{rang} f(x)$ локально не может уменьшиться. Ну в самом деле, определитель матрицы — непрерывная функция элементов матрицы, следовательно, если какой-то минор дифференциала отличен от нуля в какой-то точке x, то он и в окрестности этой точки тоже отличен от нуля.

Естественно, возрасти $\operatorname{rang} f(x)$ может. В следующей теореме мы предполагаем, что ранг дифференциала постоянен в окрестности точки x.

Теорема о ранге. *Пусть $U \subset \mathbb{R}^m$ — окрестность точки $x_0 \in \mathbb{R}^m$. Пусть $f : U \to \mathbb{R}^n$, $f \in C^p$, $p \geqslant 1$ и $\operatorname{rang} f(x) = k$ при всех $x \in U$. Тогда существуют такие окрестности $O(x_0)$ и $O(y_0)$ точек x_0 и $y_0 = f(x_0) \in \mathbb{R}^n$ и такие диффеоморфизмы $u = \varphi(x)$, $v = \psi(y)$ класса C^p, что в окрестности $O(u_0) = \varphi(O(x_0))$ точки $u_0 = \varphi(x_0)$ отображение $v = \psi\Big(f\big(\varphi^{-1}(u)\big)\Big)$ имеет вид*

$$(u_1, \dots, u_k, \dots, u_m) = u \mapsto v = (u_1, \dots, u_k, 0, \dots, 0).$$

Для линейных отображений у вас была соответствующая теорема о том, что всегда линейное отображение можно было привести к каноническому виду, теорема о ранге — нелинейный локальный аналог.

Эта теорема утверждает, что всегда можно выбрать новые координаты в области определения и в образе так, чтобы отображение ранга k имело канонический вид.

Иными словами, все такие отображения разбиваются на классы по инварианту k.

Лекция 5 (29 сентября 2014)

На прошлой лекции мы доказали теорему о достаточных условиях экстремума и перешли к изучению диффеоморфизмов. Была сформулирована теорема о диффеоморфизме (если отображение гладкое имеет в точке обратимый дифференциал, то локально — это диффеоморфизм), дающая достаточные условия того, что отображение является диффеоморфизмом. Была сформулирована теорема о ранге, сейчас мы ее будем доказывать.

Теорема о ранге. *Пусть $U \subset \mathbb{R}^m$ — окрестность точки $x_0 \in \mathbb{R}^m$. Пусть $f : U \to \mathbb{R}^n$, $f \in C^p$, $p \geqslant 1$ и $\operatorname{rang} f(x) = k$ при всех $x \in U$. Тогда существуют такие окрестности $O(x_0)$ и $O(y_0)$ точек x_0 и $y_0 = f(x_0) \in \mathbb{R}^n$ и такие диффеоморфизмы $u = \varphi(x)$, $v = \psi(y)$ класса C^p, что в окрестности $O(u_0) = \varphi(O(x_0))$ точки $u_0 = \varphi(x_0)$ отображение $v = \psi\Big(f\big(\varphi^{-1}(u)\big)\Big)$ имеет вид*

$$(u_1, \dots, u_k, \dots, u_m) = u \mapsto v = (u_1, \dots, u_k, 0, \dots, 0).$$

Доказательство.

Обозначения. Без ограничения общности можно считать, что отображение $y = f(x)$ устроено так. Вектор $x \in \mathbb{R}^m$ имеет 2 части: $x^1 = (x_1, \dots, x_k)$ и $x^2 = (x_{k+1}, \dots, x_m)$, вектор $y \in \mathbb{R}^n$ также имеет 2 части $y^1 = (x_1, \dots, x_k)$ и $y^2 = (y_{k+1}, \dots, y_n)$. То, что $m, n \geqslant k$, очевидно. Отображение f теперь будем записывать в виде $y^1 = f^1(x^1, x^2)$, $y^2 = f^2(x^1, x^2)$.

То есть все переменные и функции с индексами сверху — это векторы, с индексом 1 — векторы размерности k, с индексом 2 — векторы размерности $m - k$ или $n - k$, по смыслу.

При этом мы полагаем, что именно дифференциал $(y^1)'_{x^1}$ обратим. Иными словами, именно компоненте y^1 соответствует тот минор с ненулевым определителем.

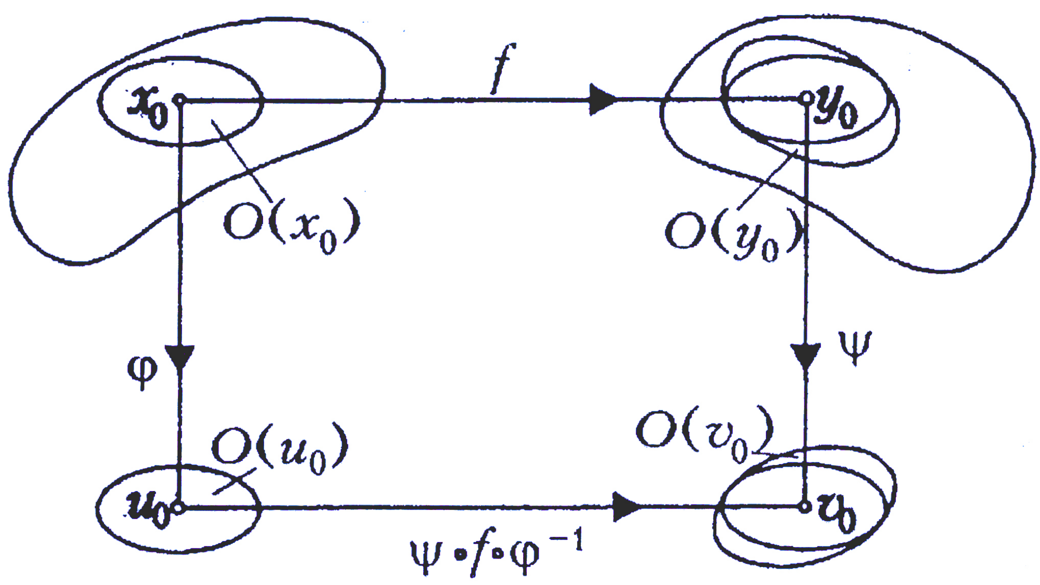

Рис. 6

Рассмотрим отображение $u^1 = f^1(x^1, x^2)$, $u^2 = x^2$ в $\mathbb{R}^m$. Его матрица Якоби имеет вид

$$\begin{pmatrix} (y^1)'_{x^1} & (y^1)'_{x^2} \\ 0 & Id \end{pmatrix}.$$

Это я так записал блочную матрицу размера $m \times m$, у нее блок $(y^1)'_{x^1}$ имеет размер $k \times k$, блок Id имеет размер $(m-k) \times (m-k)$, блок $(y^1)'_{x^2}$ имеет размер $k \times (m-k)$. По предположению, это обратимая матрица.

Значит, по теореме о диффеоморфизме, отображение $u = \varphi(x)$ является C^p-диффеоморфизмом в некоторой окрестности точки x_0, определено обратное отображение $\varphi^{-1} \in C^p$.

По построению, композиция $g = f(\varphi^{-1}(u))$ определена в окрестности окрестности точки u_0 и имеет представление

$$y^1 = f^1(\varphi^{-1}(u)) = u^1, \qquad y^2 = f^2(\varphi^{-1}(u)) = g(u^1, u^2).$$

Посчитаем матрицу Якоби отображения g:

$$\begin{pmatrix} Id & 0 \\ g'_{u^1} & g'_{u^2} \end{pmatrix}.$$

А теперь вспомним, что отображение g не может иметь ранг больше k, так как его дифференциал есть произведение матрицы полного ранга на матрицу ранга k. Значит, $g'_{u^2} = 0$, то есть $g = (g^1, g^2)$, $g^1 = Id$ и компонента g^2 не зависит от переменной u^2, поэтому $g^2 = g(u_1)$.

Значит отображение $g = f(\varphi^{-1}(u))$ можно переписать в виде $y^1 = u^1$, $y^2 = g(u^1)$.

Теперь положим $v = (v^1, v^2) \in \mathbb{R}^n$, $v^1 = \psi^1(y) = y^1$, $v^2 = \psi^2(y) = y^2 - g(y^1)$. Это отображение в $\mathbb{R}^n$, класса C^p, его матрица Якоби имеет вид

$$\begin{pmatrix} Id & 0 \\ g'_{y^1} & Id \end{pmatrix}.$$

Его определитель равен 1, значит это диффеоморфизм.

Диффеоморфизмы φ и ψ — это как раз те, о которых говорилось в формулировке теоремы.

Нужно еще только проследить за окрестностями. Мы 2 раза использовали теорему о диффеоморфизме, там каждый раз выбирались окрестности, вот каждый раз надо брать пересечения всех этих окрестностей. $\square$

Замечания к теореме о ранге.

1. Если ранг отображения f равен n, то точка $y = f(x)$ является внутренней.

Это просто: $f = \psi^{-1} \circ \left(\psi \circ f \circ \varphi^{-1}\right) \circ \varphi$, по краям стоят диффеоморфизмы, переводящие внутренние точки во внутренние, а в середине стоит отображение $u \mapsto v$.

2. Если ранг отображения f равен k и $k < n$, то при правильном разбиении f на компоненты в некоторой окрестности точки x имеют место соотношения $f^2(x) = g^2(f^1(x))$.

Мы это по дороге использовали.

Зависимость функций.

Определение. Система непрерывных функций $f^k,\ k = 1, \ldots, n,\ f^k : \mathbb{R}^m \to \mathbb{R}$ называется *функционально независимой в окрестности точки* $x_0 \in \mathbb{R}^m$, если для любой непрерывной функции $F : \mathbb{R}^n \to \mathbb{R}$, определенной в окрестности точки $y_0 = (f^1(x_0), \ldots, f^1(x_0)) \in \mathbb{R}^n$, соотношение

$$F(f^1(x), \ldots, f^n(x)) \equiv 0$$

возможно только для $F \equiv 0$.

Если система не является функционально независимой в окрестности точки $x_0 \in \mathbb{R}^m$, то она называется *функционально зависимой.*

Линейная зависимость-независимость — близкое понятие. Только функции все линейные.

Пусть дана система гладких функций $f^k,\ k = 1, \ldots, n,\ f^k : \mathbb{R}^m \to \mathbb{R}$. В следующем утверждении мы предполагаем, что ранг дифференциала отображения $f : \mathbb{R}^m \to \mathbb{R}^n,\ f(x) = (f^1(x), \ldots, f^n(x))$ постоянен в окрестности точки x_0 и равен k.

Утверждение.

а) *При $k = n$ система гладких функций f^k является функционально независимой.*

б) *При $k < n$ система функций f^k является функционально зависимой. При этом можно выбрать такие k индексов s_j, что система функций $f^{s_j},\ j = 1, \ldots, k$ функционально независима и остальные функции $f^s,\ s \neq s_j$ можно записать в виде*

$$f^s(x) = g^s(f^{s_1}(x), \ldots, f^{s_k}(x)),$$

где g^s — гладкие функции, определенные в окрестности точки y_0 и зависящие только от k координат с индексами s_j.

Эту теорему легко запомнить так: вспомнить про линейную зависимость, здесь все то же самое: есть n векторов размерности m, из их координат составим матрицу, получим матрицу, если ранг k равен n, то векторы линейно независимы, если $k < n$, то векторы линейно зависимы.

Доказательство этой теоремы легко следует из сделанных замечаний к теореме о ранге.

При $k = n$ при отображении $y = f(x)$ образ окрестности точки x_0 содержит целую окрестность точки y_0. Значит, соотношение

$$F(f^1(x), \ldots, f^n(x)) \equiv 0$$

эквивалентно соотношению $F(y) \equiv 0$ в этой окрестности.

Теперь пусть $k < n$ и ранг k реализуется на первых k функциях f_i. Тогда в силу теоремы о ранге (в силу замечания 2) найдутся $n - k > 0$ функций g_i таких, что $y_i = g_i(y_1, \ldots, y_k)$, каждую из которых можно взять в качестве F. $\square$

Разложение диффеоморфизма в композицию простейших.

Определение. Диффеоморфизм $f : U \subset \mathbb{R}^m \to \mathbb{R}^m$ назовем простейшим, если если его координатное представление имеет при некотором j вид

$$f_i(x) = x_i, \quad i \neq j, \qquad f_j(x) - - - \text{какая-то функция.}$$

Иными словами, простейший диффеоморфизм меняет только одну координату диффеоморфизма.

Теорема. *Всякий диффеоморфизм в окрестности точки может быть представлен в виде композиции простейших диффеоморфизмов.*

Первое соображение. Попробуем представить себе линейные диффеоморфизмы, то есть квадратные невырожденные матрицы. Верно ли это для таких матриц?

Я не помню уже, проходят ли прямо вот такую теорему в обычном курсе линейной алгебры. Но я знаю точно, что всякую квадратную матрицу можно можно привести к треугольному виду элементарными преобразованиями, это метод Гаусса решения систем линейных уравнений. Каждое элементарное преобразование — простейший диффеоморфизм. А треугольную матрицу легко представить в виде композиции простейших: преобразование

$$y_1 = a_{11}x_1, \qquad y_2 = a_{21}x_1 + a_{22}x_2, \qquad y_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3, \qquad \dots$$

очевидно, является произведением простейших: сначала преобразовываем n-ю координату, потом $n-1$-ю координату и так далее.

Второе соображение. Таким образом, преобразование, переставляющие координаты тоже является композицией простейших. Ну это можно и «в лоб» доказать.

Доказательство этой теоремы я буду производить в координатах и по индукции. Я думал, как записать доказательство в некоординатной форме и без индукции, и у меня внятно не вышло. Поэтому я буду доказывать все в координатах, это будет довольно громоздко, но, надеюсь, понятно.

Предположение индукции: любой диффеоморфизм, меняющий не более k координат представим в виде композиции простейших.

База индукции: при $k = 1$ утверждение очевидно — такой диффеоморфизм простейший.

Пусть любой диффеоморфизм, меняющий не более $k-1$ координат представим в виде композиции простейших. Рассмотрим диффеоморфизм, меняющий не более k координат.

Рассмотрим диффеоморфизм, меняющий первые k координат:

$$\begin{aligned} y_1 &= f_1(x_1, \ldots, x_m), \\ &\ldots\ldots\ldots \\ y_k &= f_k(x_1, \ldots, x_m), \\ y_{k+1} &= x_{k+1}, \\ &\ldots\ldots\ldots \\ y_m &= x_m. \end{aligned}$$

Первые координаты мы выбираем не меняя общности в силу сделанных сначала замечаний про линейные диффеоморфизмы.

Поскольку f — диффеоморфизм, то его матрица Якоби не вырожденная, имеет обратную. Зафиксируем некоторую точку x_0, напишем определитель матрицы Якоби

$$\begin{vmatrix} \partial_1 f_1 & \ldots & \partial_k f_1 & \vdots & \partial_{k+1} f_1 & \ldots & \partial_m f_1 \\ \ldots & \ldots & \ldots & \vdots & \ldots & \ldots & \ldots \\ \partial_1 f_k & \ldots & \partial_k f_k & \vdots & \partial_{k+1} f_k & \ldots & \partial_m f_k \\ \ldots & \ldots & \ldots & \vdots & \ldots & \ldots & \ldots \\ & & & \vdots & 1 & \ldots & 0 \\ & 0 & & \vdots & & & \\ & & & \vdots & 0 & \ldots & 1 \end{vmatrix} = \begin{vmatrix} \partial_1 f_1 & \ldots & \partial_k f_1 \\ \ldots & \ldots & \ldots \\ \partial_1 f_k & \ldots & \partial_k f_k \end{vmatrix} \neq 0.$$

Из последнего соотношения можно сделать вывод, что один из миноров размера $k-1$ отличен от нуля. Можно считать, что это главный минор, первые $k-1$ функций, первые $k-1$ переменных. Теперь рассмотрим вспомогательное отображение $g : U \to \mathbb{R}^m$, определяемое равенствами

$$\begin{aligned} u_1 &= f_1(x_1, \ldots, x_m), \\ &\ldots\ldots\ldots \\ u_{k-1} &= f_{k-1}(x_1, \ldots, x_m), \\ u_k &= x_k, \\ &\ldots\ldots\ldots \\ u_m &= x_m. \end{aligned}$$

Якобиан этого отображения имеет вид

$$\begin{vmatrix} \partial_1 f_1 & \dots & \partial_{k-1} f_1 & \vdots & \partial_k f_1 & \dots & \partial_m f_1 \\ \dots & \dots & \dots & \vdots & \dots & \dots & \dots \\ \partial_1 f_{k-1} & \dots & \partial_{k-1} f_{k-1} & \vdots & \partial_k f_{k-1} & \dots & \partial_m f_{k-1} \\ \dots & \dots & \dots & \vdots & \dots & \dots & \dots \\ & & & \vdots & 1 & \dots & 0 \\ & 0 & & \vdots & & & \\ & & & \vdots & 0 & \dots & 1 \end{vmatrix} = \begin{vmatrix} \partial_1 f_1 & \dots & \partial_{k-1} f_1 \\ \dots & \dots & \dots \\ \partial_1 f_k & \dots & \partial_{k-1} f_{k-1} \end{vmatrix} \neq 0$$

(по предположению). Значит, g — диффеоморфизм и у него есть обратный диффеоморфизм g^{-1}.

Пусть $h = f(g^{-1})$. Отображение h переводит u в x, потом f переводит x в y.

Отображение h — это композиция диффеоморфизмов, следовательно h — тоже диффеоморфизм. Его координатная запись имеет вид

$$\begin{aligned} y_1 &= f_1(g^{-1}(u)) = u_1, \\ &\dots\dots\dots \\ y_{k-1} &= f_{k-1}(g^{-1}(u)) = u_{k-1}, \\ y_k &= f_k(g^{-1}(u)), \\ y_{k+1} &= u_{k+1} \\ &\dots\dots\dots \\ y_m &= u_m, \end{aligned}$$

то есть h — это простейший диффеоморфизм.

Теперь всё просто: $f(x) = h(g(x))$, по предположению индукции g раскладывается в композицию простейших, значит и f раскладывается в композицию простейших. $\square$

Теперь можно еще заметить, что по доказательству кажется, что простейших диффеоморфизмов ровно m штук: сколько-то было в разложении h и еще g. Это не так уже в линейном случае. Где-то там в середине шага индукции у нас были перестановки переменных, надо их тоже посчитать.

Теперь мы переходим к следующей теме, немножко в сторону, мы говорили про экстремумы, мы говорили про диффеоморфизмы, сейчас я займусь объединяющей все это темой.

Лемма Морса.

Пусть дана функция $f : U \in \mathbb{R}^m \to \mathbb{R}$, $f \in C^2$. Пусть $x_0 \in U$ — критическая точка этой функции, то есть $f'(x_0) = 0$, или $\nabla f(x_0) = 0$, или $\partial_k f(x_0) = 0$ — это все одно и тоже.

Определение. Критическая точка называется *невырожденной*, если матрица

$$\begin{pmatrix} \partial_{11} f & \dots & \partial_{1m} f \\ \dots & \dots & \dots \\ \partial_{m1} f & \dots & \partial_{mm} f \end{pmatrix}$$

(***гессиан*** функции f) невырожденная, то есть у нее не равен нулю определитель.

Если 0 — критическая точка достаточно гладкой функции, то по формуле Тейлора

$$f(x) - f(0) = \frac{1}{2} \sum \partial_{ij} f(0) x_i x_j + o(\|x\|).$$

Лемма Морса утверждает, что всегда существует такой диффеоморфизм окрестности нуля в окрестность нуля, что в новых координатах $u = (u_1, \dots, u_m)$ функция f будет иметь вид

$$f(u) - f(0) = -u_1^2 - \dots - u_k^2 + u_{k+1}^2 + \dots + u_m^2.$$

Если бы в формуле Тейлора не было малых слагаемых, то $f(x) - f(0)$ была бы квадратичная форма, как известно из линейной алгебры, ее можно было бы привести к указанному виду линейным преобразованием. В общем случае это можно сделать нелинейным диффеоморфизмом, причем только локально.

Для доказательства сначала сформулируем и докажем лемму Адамара.

Лемма Адамара. *Пусть функция $f : U \to \mathbb{R} \in C^p$, $p \geqslant 1$ определена в выпуклой окрестности $U \subset \mathbb{R}^m$ точки $0 \in \mathbb{R}^m$. Пусть $f(0) = 0$. Тогда существует отображение $g(x) = (g_1(x_0, \dots, g_m(x)) \in C^{p-1}$ такие, что справедливо равенство*

$$f(x) = (x, g(x)), \qquad \Leftrightarrow \qquad f(x_1, \dots, x_m) = \sum_{i=1}^{m} x_i g_i(x_1, \dots, x_m),$$

причем $g_i(0) = \partial_i f(0)$.

Доказательство. Требуемая формула вытекает из равенства

$$f(x) = \int_0^1 \frac{df(tx)}{dt}\, dt = \sum_{i=1}^{m} x_i \int_0^1 \partial_i f(tx)\, dt,$$

если положить

$$g_i(x) = \int_0^1 \partial_i f(tx)\, dt, \qquad i = 1, \dots, m.$$

Равенство $g_i(0) = \partial_i f(0)$ следует из формулы и непрерывности g, то, что $g \in C^{p-1}$, — очевидно. $\square$

Лекция 6 (6 октября 2014)

На прошлой лекции мы сделали следующее.

1) Доказали теорему о ранге.

2) Сформулировали следствие — утверждение о локальной функциональной зависимости и независимости функций, обобщающее соответствующую теорему о линейной зависимости-независимос векторов.

3) Доказали теорему о представлении диффеоморфизма в виде композиции простейших.

4) Начали обсуждать лемму Морса с того, что сформулировали и доказали лемму Адамара.

Теперь перейдем к точной формулировке и доказательству леммы Морса.

Лемма Морса. *Пусть $f : G \to \mathbb{R}$, $f \in C^3$, пусть $x_0 \in G$ — внутренняя точка G является невырожденной критической точкой f. Найдется такой диффеоморфизм $g : V \to U$, окрестности V окрестности нуля на окрестность $U \subset G$, что*

$$f(g(u)) = f(x_0) - u_1^2 - \ldots - u_k^2 + u_{k+1}^2 + \ldots + u_m^2.$$

Напоминаю: критическая точка называется *невырожденной*, если матрица $\partial_{ij} f(x_0)$ (*гессиан*) обратима.

Доказательство. Без ограничения общности и линейными заменами переменных можно свести задачу к случаю $x_0 = 0$ и $f(x_0) = 0$.

Применим к функции f лемму Адамара, получим $f = (x, g(x))$. Так как 0 — критическая точка, то $g(0) = 0$ в силу той же леммы. Теперь применим ко всем координатам g_i ту же лемму Адамара:

$$g_i(x) = (x, h_i(x)) = \sum_{j=1}^{m} x_j h_{ij}(x).$$

Таким образом

$$f(x) = \sum_{i,j=1}^{m} x_i x_j h_{ij}(x).$$

Такое представление не единственно, для функций h_{ij} и h_{ji} фиксирована их сумма. Без ограничения общности можно считать, что $h_{ij} = h_{ji}$.

Из непрерывности h_{ij} и единственности тейлоровского разложения следует равенство $h_{ij}(0) = \partial_{ij} f(0)$. По предположению матрица $h_{ij}(0)$ не вырождена.

Теперь функция f записана в матричном виде (что-то вроде билинейной формы), а мы будем приводить её диффеоморфизмами к диагональному виду. Действуем по индукции.

Предположение индукции. Существуют координаты $u = (u_1, \ldots, u_m)$ (иными словами — диффеоморфизм $x = \varphi(u)$, $\varphi(0) = 0$), при которых

$$f(\varphi(u)) = \pm u_1^2 \pm \ldots \pm u_{r-1}^2 + \sum_{i,j=r}^{m} u_i u_j H_{ij}(u).$$

Здесь $H_{ij} = H_{ji}$ — некоторые функции. При $r = 1$ равенство уже доказано с применением леммы Адамара.

Предположим, что при $r > 1$ это равенство верно, докажем его при $r + 1$.

Так как матрица $h_{ij}(0)$ не вырождена и φ — диффеоморфизм, то матрица формы

$$\pm u_1^2 \pm \ldots \pm u_{r-1}^2 + \sum_{i,j=r}^{m} u_i u_j H_{ij}(0)$$

также не вырождена. Линейной заменой переменных u_j, $j = r, \ldots, m$ всегда можно привести форму к диагональному виду, поэтому можно считать, что в предыдущей формуле $H_{rr}(0) \neq 0$. По непрерывности $H_{rr}(x) \neq 0$ в некоторой окрестности нуля.

Положим $\psi(u) = \sqrt{|H_{rr}(u)|}$. Функция ψ непрерывно дифференцируема в некоторой окрестности нуля, сделаем теперь переход к новым координатам

$$\begin{aligned} v_i &= u_i, \qquad i \neq r \\ v_r &= \psi(u)\left(u_r + \sum_{j>r} \frac{u_j H_{jr}(u)}{H_{rr}(u)}\right). \end{aligned}$$

Матрица Якоби (в нуле!) преобразования $v = \phi(u)$ отличается от единичной одной строкой номер r, в этой строке при $u = 0$ на диагонали стоит число $\psi(0)$. Якобиан $\psi(0) \neq 0$, поэтому отображение ϕ — диффеоморфизм.

Теперь запишем основное равенство для $f(\varphi(u))$ в переменных v. Для этого сначала в слагаемых $\sum_{i,j=r}^{m} u_i u_j H_{ij}(u)$ выделим слагаемые, содержащие u_r:

$$B = u_r^2 H_{rr}(u) + 2u_r \sum_{j=r+1}^{m} u_j H_{rj}(u)$$

и запишем их в переменных v:

$$B = \pm v_r^2 - \frac{1}{H_{rr}(v)} \sum_{i,j=r+1}^{m} v_i v_j H_{ij}(v).$$

Знак $\pm$ перед v_r^2 совпадает со знаком $H_{rr}(0)$. Таким образом шаг индукции полностью сделан. □

Замечания к лемме Морса.

1. Из леммы Морса следует, что все невырожденные критические точки являются изолированными. При диффеоморфизмах $u = \varphi(x)$ критические точки u функции $f(u)$ однозначно соответствуют критическим точкам x функции $f(\varphi(x))$ переходят в критические $(\big(f(\varphi(x))\big)' = \varphi'(x) f'(\varphi(x)))$.

У функции $f(x_0) - u_1^2 - \ldots - u_k^2 + u_{k+1}^2 + \ldots + u_m^2$ в окрестности 0 (и не только) есть единственная критическая точка 0, она изолированная, значит и x_0 — невырожденная критическая точка.

2. Как и в линейном случае, справедлив закон инерции: количество знаков $+$ и знаков $-$ в каноническом представлении не зависит от способа приведения к каноническому виду. Иными словами, эти количества являются свойством функции и невырожденной критической точки. Они

полностью определяются гессианом функции. Число знаков минус называется *индекс критической точки.* У локального минимума индекс равен нулю, у локального максимума индекс равен m.

3. Заметьте, что в одномерном случае критическая точка x невырождена, если $f''(x) \neq 0$. Каждая невырожденная критическая точка является точкой экстремума.

В двумерном случае бывают 3 типа невырожденных критических точек: максимум, минимум и обычное седло. Обезьянье седло — вырожденная критическая точка.

Поверхности.

Вам будут рассказывать про то, что такое поверхность, много раз и на разных дисциплинах. Мы представляем себе обычно двумерную поверхность в трехмерном пространстве и как-то переносим такую объемную картинку в многомерные пространства. Кривая — это простейший вид поверхности: поверхность размерности 1.

Двумерную картинку удобно представлять себе

Понятно, что как геометрический объект мы представляем себе всё только в трехмерном пространстве. Решать задачи приходится в многомерном пространстве. Поэтому такую важную роль играют аналитические методы изучения кривых поверхностей. Простейший случай вы изучали — это квадрики.

Вроде очевидно, что кривые и поверхности можно задавать различными формулами. В смысле, что одну и ту же поверхность можно задавать по-разному, различными формулами.

Основных способов 2: параметрический и неявный.

В параметрическом способе мы предполагаем, что геометрическая фигура, которая получается в результате гладкого отображения области $G \in \mathbb{R}^m$ в пространство $\mathbb{R}^n$, имеет при $m < n$ размерность m. Я не обсуждаю, что значит множество размерности m. Это сложное понятие, бывают различные определения, вы слышали слова про дробные размерности, про фракталы. Я в настоящий момент философствую и говорю об интуитивных понятиях.

Типичный пример: график функциии 2х переменных в $\mathbb{R}^3$.

Другой важный способ задания поверхностей — уравнения. Сфера в $\mathbb{R}^3$ задается уравнением $x^2 + y^2 + z^2 = 1$. Кривая — это пересечение 2х поверхностей, то есть множество решений 2х уравнений. Конечно, интуитивно всегда «количество уравнений» + «размернойть поверхности» = «размерность пространства».

Очень важно понимать, что локальное устройство множества и глобальное его устройство - это совершенно разные вещи. Локально каждая точка окружности и каждая точка интервала (неважно, в 2-х или 3-х –мерных пространствах) — одно и то же. Глобально мы все понимаем, что окружность и интервал — вещи разные. Что означает, слово «разные» в этом контексте, очевидно, это вопросы не матанализа, а разных других математических наук.

А граница квадрата и окружность — они разные или одинаковые? На этот вопрос ответы могут быть разные — и правильные. То, что они гомеоморфные (есть гомеоморфизм, переводящий одно в другое), легко увидеть.

Более того, не просто существует гомеоморфизм окружности и квадрата. Существует гомеоморфизм областей, содержащих окружность и квадрат.

Мы не будем заниматься глобальным устройством множеств (кривых, поверхностей). Вместо этого займемся их локальным устройством.

Основное определение. *Множество $S \subset \mathbb{R}^n$ называется k*-мерной *поверхностью гладкости p, если для любой точки $x_0 \in S$ найдется окрестность $U(x_0) \in \mathbb{R}^n$, и диффеоморфизм $\varphi : U(x_0) \mapsto \mathbb{R}^n$ гладкости p окрестности $U(x_0)$ на окрестность $O \subset \mathbb{R}^n$ начала координат, такие что множество $\varphi(S \bigcap U(x_0))$ совпадает с множеством $\{t \in O : t_{k+1} = \ldots = t_n = 0\}$.*

Это определение сразу означает, что локально есть $n - k$ уравнений «диффеоморфизм равен нулю».

Параметрическое определение. *Поверхностью размерности k в $\mathbb{R}^n$ называется множество $S \in \mathbb{R}^n$, каждая точка которого имеет окрестность $S \bigcap \mathbb{R}^n$, гомеоморфную $\mathbb{R}^k$.*

Окрестность нуля может быть любой, например, шар, или промежуток–параллелепипед.

Вместо слова поверхность ещё используют слова многообразие или подмногообразие.

Ясно, что такое определение отличается от предыдущего. В нём нет и не используется никакой гладкости. Никак не используется как именно располагается поверхность в $\mathbb{R}^n$.

Квадрат (как множество из четырёх отрезков) не является 1-мерной поверхностью гладкости 1 в смысле первого определения, но является поверхностью гладкости 0 в смысле второго определения.

Почему квадрат (из-за угловых точек) не является гладкой поверхностью.

1. Если есть диффеоморфизм f гладкости 1 и выше, то его дифференциал обратим. Это просто: $Id = f \circ f^{-1},\ D(Id) = Id,\ D(f \circ f^{-1}) = Df \circ D(f^{-1})$, а это — произведение матриц (по правилу дифференцирования сложной функции).
2. Пусть есть угол квадрата и он диффеоморфен отрезку. То есть есть отображение $(u, v) = f(x, y)$, такое что угол «$\llcorner$» переходит в отрезок $v = 0,\ u \in \Delta$, то есть $\partial_x f(0, 0) = \partial_y f(0, 0) = 0$.
3. Получили противоречие — вырожденный дифференциал.

Другой пример: рогатая сфера Александера — поверхность, гомеоморфная 2-мерной сфере в $\mathbb{R}^3$ ограничивает область не гомеоморфную шару.

Теорема. *Пусть отображение окрестности $\varphi : U \in \mathbb{R}^k \to \mathbb{R}^n,\ n > k$ имеет гладкость $C^p,\ p \geqslant 1$ и пусть $\operatorname{rang} \varphi = k$ в каждой точке U. Тогда образ $\varphi(U(t))$ некоторой окрестности $U(t) \subset U$ каждой точки $t \in U$ — это поверхность гладкости p размерности k в $\mathbb{R}^n$.*

Иными словами, если отображение, задающее поверхность параметрически, гладкое и его ранг максимальный, то локально эта поверхность удовлетворяет первому определению поверхности тоже. В критических точках, где ранг меньше, может быть что угодно.

Глобально поверхность параметрически не получится задать даже с помощью отображения постоянного ранга (пример). Требуется дополнительное условие, например, что φ — гомеоморфизм.

Доказательство. Выберем произвольную точку $x_0 \in \varphi(U)$, пусть $x_0 = \varphi(t_0),\ t_0 \in U$.

Без ограничения общности считаем, что координаты пронумерованы так, что минор $\partial_i \varphi_j,\ i, j = 1, \ldots, k$ отличен от нуля в окрестности точки x_0. Тогда соотношения

$$\begin{aligned} x_1 &= \varphi_1(t_1, \ldots, t_k), \\ &\ldots\ldots\ldots \\ x_k &= \varphi_k(t_1, \ldots, t_k), \\ x_{k+1} &= \varphi_{k+1}(t_1, \ldots, t_k), \\ &\ldots\ldots\ldots \\ x_n &= \varphi_n(t_1, \ldots, t_k) \end{aligned}$$

в силу теоремы о неявной функции можно переписать в виде

$$\begin{aligned} t_1 &= f_1(x_1, \ldots, x_k), \\ &\ldots\ldots\ldots \\ t_k &= f_k(x_1, \ldots, x_k), \\ x_{k+1} &= \varphi_{k+1}(t_1, \ldots, t_k), \\ &\ldots\ldots\ldots \\ x_n &= \varphi_n(t_1, \ldots, t_k) \end{aligned}$$

и, следовательно, в виде

$$\begin{aligned} t_1 &= f_1(x_1, \ldots, x_k), \\ &\ldots\ldots\ldots \\ t_k &= f_k(x_1, \ldots, x_k), \\ x_{k+1} &= f_{k+1}(x_1, \ldots, x_k) = \varphi_{k+1}(f_1(x_1, \ldots, x_k), \ldots, f_k(x_1, \ldots, x_k)), \\ &\ldots\ldots\ldots \\ x_n &= f_n(x_1, \ldots, x_k) = \varphi_n(f_1(x_1, \ldots, x_k), \ldots, f_k(x_1, \ldots, x_k)). \end{aligned}$$

Теперь рассмотрим отображение в $\mathbb{R}^n$ окрестностей точек t_0 и x_0:

$$\begin{aligned} t_1 &= f_1(x_1, \ldots, x_k), \\ &\ldots\ldots\ldots \\ t_k &= f_k(x_1, \ldots, x_k), \\ t_{k+1} &= x_{k+1} - f_{k+1}(x_1, \ldots, x_k), \\ &\ldots\ldots\ldots \\ t_n &= x_n - f_n(x_1, \ldots, x_k). \end{aligned}$$

Это отображение в силу теоремы о диффеоморфизме является диффеоморфизмом окрестностей в $\mathbb{R}^n$, при этом поверхность S переходит в требуемое множество $t_{k+1} = \ldots = t_n = 0$. $\square$

Примеры.

1. Само пространство $\mathbb{R}^n$ является n-мерной поверхностью в себе.

2. Любое k-мерное линейное многообразие, подпространство или афинное, является поверхностью размерности k.

3. Путь есть кривая Γ на плоскости, заданная скалярным уравнением $F(x, y) = 0$. Пусть F — гладкая функция, пусть $F(x_0, y_0) = 0$ и пусть точка (x_0, y_0) не является критической: то есть ранг отображения $DF(x_0, y_0)$ равен 1, то есть по крайней мере одна из производных $F'_x(x_0, y_0)$ или $F'_y(x_0, y_0)$ отлична от нуля.

Мы доказали не так давно, что в некоторой окрестности O точки (x_0, y_0) кривая Γ — гладкое многообразие размерности 1.

Это было утверждение о возможности выпрямления кривой, в точности то, что нам нужно: существует диффеоморфизм окрестности O в окрестность начал координат, так что Γ отображается в ось абсцисс.

4. Теперь пусть всё то же самое, уравнение от многих переменных в пространстве $\mathbb{R}$, то есть $F : \mathbb{R}^{m+1} \to \mathbb{R}$, одно скалярное уравнение, переменных $m+1$ — множество решений — поверхность размерности m. Рассмотрим уравнение $F(x) = 0$. Пусть при всех $x_0 : F(x_0) = 0$ справедливо $\mathbf{grad}\, F(x_0) \neq 0$ (то есть хотя бы одна из производных $\partial_i F \neq 0$).

Мы доказали не так давно, что это есть гладкая поверхность размерности m в $\mathbb{R}^{m+1}$.

5. График гладкой функции $f : \mathbb{R}^m \to \mathbb{R}$ является гладкой поверхностью размерности m в $\mathbb{R}^{m+1}$. Диффеоморфизм зададим формулами

$$t_i = x_i, \quad i = 1, \ldots, m-1; \qquad t_{m+1} = x_{m+1} - f(x_1, \ldots, x_m).$$

6. Окружность S на плоскости есть гладкая поверхность размерности 1 в $\mathbb{R}^2$.

Для того, чтобы это доказать явно укажем диффеоморфизм. Вспомним полярную систему координат, положим $x = (r+1)\cos\varphi,\ y = (r+1)\sin\varphi$. Точка с координатами (x_0, y_0) перейдет в точку с координатами $r = 0,\ \varphi = \varphi_0$, окружность S локально перейдет в прямую $r = 0$.

Лекция 7 (13 октября 2014)

На прошлой лекции мы сделали следующее.

1) Доказали лемму Морса

2) Начали обсуждать понятие поверхности

3) Привели несколько примеров поверхностей

7. Теперь самый главный, «настоящий» пример.

Пусть есть гладкое отображение $F : \mathbb{R}^n \to \mathbb{R}^{n-k}$ постоянного ранга $n-k$. Тогда уравнение $F(x) = 0$ задает в $\mathbb{R}^n$ поверхность размерности k.

В координатах уравнение имеет привычный вид

$$\begin{cases} F_1(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) = 0, \\ \qquad \ldots \qquad \ldots \\ F_{n-k}(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) = 0. \end{cases}$$

Здесь $x^1 = (x_1, \ldots, x_k),\ x^2 = (x_{k+1}, \ldots, x_n)$

Чтобы это доказать, надо написать уравнение $F(x^1, x^2) = 0,\ x^1 \in \mathbb{R}^k, x^2 \in \mathbb{R}^{m-k}$, считать, что дифференциал $\partial_{x_2} F(x_0)$ обратим, а потом сказать, что по теореме о неявной функции исходное уравнение можно переписать в виде $x^2 = \Phi(x^1)$. Поэтому можно написать диффеоморфизм $(y^1, y^2) = \varphi(x^1, x^2) = (x^1, x^2 - \Phi(x^1))$, это диффеоморфизм так как у него $D\Phi = Id$, решения уравнения $F(x) = 0$ переходят в требуемое множество $\{y^2 = 0\}$.

Более того, **всякую гладкую поверхность можно задать в виде уравнения.** В самом деле, по определению существует диффеоморфизм этой поверхности на множество $\{t \in O : t_{k+1} = \ldots = t_m = 0\}$. Вот уравнения и написаны — локальные!

Этот пример показывает, что не случайно во всех или почти всех конструкциях и теоремах задание поверхности в виде уравнений — это основной способ.

8. Рассмотрим уравнение $xy = 0$. В нуле у него градиент равен нулю. И это уравнение — из-за начала координат — не задает поверхности размерности 1 (это оси координат).

Всё дело в том, в начале координат ранг производной равен не 1, а равен 0. Предыдущая конструкция не применима.

Если ноль выкинуть, то останется 4 луча, каждый их которых является поверхностью.

9. Замкнутый луч на плоскости не является поверхностью размерности 1: в крайней точке неприятности. В $\mathbb{R}^3$ — замкнутая полуплоскость не является поверхностью размерности 2. Такое называется *поверхностью с краем.*

10. Параметрический способ задания поверхности был локальный! Окрестность каждой точки из множества параметров была поверхностью, которую можно было задать уравнениями. Глобально множество, заданное параметрически может оказаться плохим.

Достаточно, чтобы φ был гомеоморфизмом.

Пусть V — окрестность t_0, образ которой является гладкой поверхностью в $\mathbb{R}^n$. Нам достаточно найти окрестность $D \subset \mathbb{R}^n$ точки $x_0 = \varphi(t_0) \in \mathbb{R}^n$, которая не пересекается с $\varphi(U) \setminus \varphi(V)$.

Множество $\varphi(V)$ открыто в $\varphi(U)$ (т.к. φ — гомеоморфизм на свой образ), поэтому каждая точка $x \in \varphi(V)$ имеет окрестность $D_x \subset \mathbb{R}^n$, не пересекающуюся с $\varphi(U \setminus V)$, объединение таких окрестностей D_x есть такое открытое множество D в R^n, что $D \cap \varphi(U) = \varphi(V)$. Ясно, что D — искомое.

Примеров поверхностей можно приводить много, я останавливаюсь.

Касательные векторы и касательные пространства.

Пусть есть 2-мерная поверхность $S \subset \mathbb{R}^3$, заданная равенством $z = f(x, y)$, где f — гладкая функция. В точке $M = (x_0, y_0, z_0) \in S,\ z_0 = f(x_0, y_0)$ можно написать формулу Тейлора

$$z - z_0 = f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0) + \text{«малые члены»}.$$

Если малые члены отбросить, то останется равенство

$$z - z_0 = f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0),$$

геометрически оно задает афинную плоскость, проходящую через M — касательную плоскость.

Если задать на этой плоскости структуру линейного пространства, поместив M в начало координат, то мы получим касательное пространство.

Если 2-мерная поверхность задавалась в $\mathbb{R}^3$ уравнением $F(x, y, z) = 0$, то уравнение касательной плоскости в точке (x_0, y_0, z_0) имеет вид

$$F'_x(x_0, y_0, z_0)(x - x_0) + F'_y(x_0, y_0, z_0)(y - y_0) + F'_z(x_0, y_0, z_0)(z - z_0) = 0.$$

В общем случае формулы имеют примерно такой же вид.

Пусть у нас есть k-мерная поверхность $S \subset \mathbb{R}^n$, и в окрестности точки $x_0 \in S$ она задана с помощью гладкого отображения $t \in \mathbb{R}^k \mapsto x = \varphi(t) \in \mathbb{R}^n$, $x_0 = \varphi(0)$ — параметрически. Предположим, что дифференциал $D\varphi(0)$ имеет полный ранг k (тогда и в окрестности $\operatorname{rang} D\varphi(t) = k$).

Тогда k-мерная плоскость в $\mathbb{R}^n$, задаваемая параметрически формулой $x - x_0 = Dx(0)t$, называется касательной плоскостью.

В координатах формула $x - x_0 = Dx(0)t$ имеет вид

$$\left\{\begin{array}{ccccc} x_1 - x_1^0 & = & \frac{\partial x_1}{\partial t_1}(0)t_1 & + \ldots + & \frac{\partial x_1}{\partial t_k}(0)t_k, \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_n - x_n^0 & = & \frac{\partial x_n}{\partial t_1}(0)t_1 & + \ldots + & \frac{\partial x_n}{\partial t_k}(0)t_k. \end{array}\right.$$

Если на касательной плоскости в точку касания поместить начало координат мы получим касательное пространство, естественно, оно имеет размерность k.

Пусть поверхность S в окрестности точки $x_0 \in S \subset \mathbb{R}^n$ задана уравнением $F(x) = 0$, где гладкое отображение $F : \mathbb{R}^n \to \mathbb{R}^{n-k}$ имеет ранг $n-k$. Когда я говорил о примерах поверхностей я приводил эту ситуацию в качестве главного примера.

Напишем уравнение касательной плоскости.

С одной стороны, мы понимаем, что это уравнение должно иметь вид $DF(x_0)(x - x_0) = 0$ — это просто линеаризация, главное приближение уравнения $F(x) = 0$.

Чтобы привести это понимание к определению, для определенности будем считать, что $x = (u, v)$, $x \in \mathbb{R}^k$, $v \in \mathbb{R}^{n-k}$ и перепишем уравнение $F(x) = 0$ в виде $F(u, v) = 0$. Условие о ранге означает обратимость линейного отображения $D_v F(x_0)$. По теореме о неявной функции локально уравнение $F(u, v) = 0$ имеет вид $v = f(u)$.

Теперь по определению получаем параметрическое представление касательной плоскости: $u - u_0 = Id\ t$, $v - v_0 = f'(u_0)t$.

По формуле для производной из теоремы о неявной функции имеем

$$f'(u_0) = -\big(F'_v(u_0, v_0)\big)^{-1} F'_u(u_0, v_0),$$

отсюда следует формула $F'_u(u_0, v_0)(u - u_0) + F'_v(u_0, v_0)(v - v_0) = 0$, после возвращения к переменной x имеем требуемую формулу $DF(x_0)(x - x_0) = 0$, определяющую точки x касательной плоскости.

В координатах уравнения касательной плоскости имеют вид

$$\begin{cases} \frac{\partial F_1}{\partial x_1}(x_0)(x_1 - x_1^0) & + \ldots + & \frac{\partial F_1}{\partial x_n}(x_0)(x_n - x_n^0) & = 0, \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial F_{n-k}}{\partial x_1}(x_0)(x_1 - x_1^0) & + \ldots + & \frac{\partial F_{n-k}}{\partial x_n}(x_0)(x_n - x_n^0) & = 0. \end{cases}$$

Касательный вектор.

Есть гладкое отображение $x : [0, 1] \to \mathbb{R}^n$. Пусть это отображение не вырождено: в каждой точке $x' \neq 0$.

Каждое такое отображение задает в пространстве $\mathbb{R}^n$ кривую Γ. Прямая $x = x_0 + x'(t_0)\tau$, $\tau \in \mathbb{R}$ является касательной к кривой Γ в точке $x_0 = x(t_0)$.

Вектор с началом в точке $x_0 = x(t_0)$ и концом на касательной прямой называется касательным вектором к кривой Γ в точке $x_0 = x(t_0)$.

Теорема. *Пространство, касательное к поверхности S состоит их векторов, касательных в точке x_0 к кривым, лежащим на S и проходящих через точку x_0, и только из таких векторов.*

Вообще говоря, интуитивно (в $\mathbb{R}^3$) эта теорема очевидна.

Для **доказательства** теоремы надо доказать 2 утверждения:

1) Каждый касательный вектор принадлежит касательному подпространству;

2) Каждый элемент касательного пространства является касательным вектором к некоторой кривой на поверхности.

По определению поверхности (существует диффеоморфизм ...) её можно задать в виде уравнения $F(x) = 0$, $\dim x = n$, $\dim F = n - k$. Пусть кривая Γ задана функцией $x(t)$, причем $F(x(t)) \equiv 0$, $x(0) = x_0$.

1) Пусть кривая на S параметрически задана формулой $x(t)$. Продифференцируем тождество $F(x(t)) \equiv 0$, получаем $D\,F(x(t))x'(t) = 0$, при $t = 0$ получаем $D\,F(x_0)x'(0) = 0$, а это и есть уравнение касательного пространства.

2) Теперь выберем произвольный вектор ξ из касательного пространства к S в точке x_0, по определению $D\,F(x_0)\xi = 0$.

Необходимо указать кривую на поверхности, которая имеет вектор ξ вектором скорости.

Считаем, что поверхность задана уравнениями $F(x^1, x^2) = 0,\ x^1 \in \mathbb{R}^k, x^2 \in \mathbb{R}^{n-k},\ F \in \mathbb{R}^{n-k}$, считаем, что дифференциал $\partial_{x_2}F(x_0)$ обратим.

Тогда, зная первые k координат вектора ξ, из уравнения $D\,F(x_0)\xi = 0$ — это система линейных уравнений ранга $n-k$ — мы выразим остальные $n-k$ координат этого вектора.

Это значит, что если для вектора $(\xi^1, \tilde{\xi}^2)$ выполнено равенство $D\,F(x_0)\xi = 0$, то $\tilde{\xi}^2 = \xi^2$.

Рассмотрим прямую $x - x_0^1 = \xi^1 t$ в пространстве $\mathbb{R}^k$. Решим уравнение $F(x^1, x^2) = 0$ относительно второй переменной, получим $x^2 = f(x^1)$.

Теперь построим кривую $x = x_0^1 + \xi^1 t,\ x^2 = f(x_0^1 + \xi^1 t)$. По построению эта кривая лежит на поверхности S (так как $f(x^1, f(x^1) \equiv 0)$. Так же по построению эта кривая проходит через точку (x_0^1, x_0^2).

Теперь продифференцируем тождество $F(x_0^1 + \xi^1 t, f(x_0^1 + \xi^1 t)) = 0$ по t в точке $t = 0$ получается равенство $D_{x^1}\xi^1 + D_{x^2}\tilde{\xi}^2 = 0$.

Как выписать формулу для вектора $\tilde{\xi}^2$ не важно! Мы уже знаем, что $\tilde{\xi}^2 = \xi^2$. $\square$

Условный экстремум.

Постановка задачи: дана функция $f : \mathbb{R}^n \to \mathbb{R}$, дана гладкая поверхность $S \in \mathbb{R}^n$ размерности k, исследуем вопрос, когда в точке $x_0 \in S$ функция принимает экстремальное значение.

Определения локального максимума, локального минимума, строгого локального максимума, строгого локального минимума очевидны. Если в точке x_0 функция f имеет локальный экстремум во всем пространстве, то, естественно, она имеет экстремум на S. Этот случай мы считаем «не интересным».

Однако может быть так что функция имеет экстремум на S, который не является экстремумом в $\mathbb{R}^n$.

Пример. Функция $f(x, y) = x$ на окружности $x^2 + y^2 = 1$ имеет максимум в точке $x = 1, y = 0$ и минимум в точке $x = -1, y = 0$. Конечно, у этой функции нет локальных экстремумов в пространстве.

Первая мысль — считать, что поверхность задана (хотя бы локально) параметрически: $S = \{x : x = \varphi(t), t \in U \subset \mathbb{R}^k\}$, $\operatorname{rang}\varphi = k$, теперь можно переписать функцию f в виде композиции $f \circ \varphi$, и сказать, что локальный экстремум на S функции f в x_0 — это то же самое, что локальный экстремум функции $f \circ \varphi$ во всем пространстве в точке $\varphi^{-1}(x_0)$. Для него мы можем хотя бы проверить необходимое условие: надо проверить, выполнение системы k уравнений $\nabla f \circ \varphi = 0$.

Этот чудесный метод теоретически работает. Но для того, чтобы решить конкретную задачу, он совершенно не годиться. Даже в случаях, когда «всё считается», правильно использовать другой метод, восходящий к Лагранжу.

Философия и геометрия.

Нарисуем картинку на плоскости. Пусть поверхность S имеет размерность 1 (то есть это кривая) и задана уравнением $F(x, y) = 0$, а функцию $f(x, y)$ мы должны исследовать на условный локальный экстремум.

Условный — значит, на кривой, локальный — значит «в окрестности».

Я буду сейчас изображать функцию f не в виде трехмерного графика, а в виде поверхностей уровня. Так как точка (x_0, y_0), в окрестности которой мы хотим найти экстремум, не является критической ($|f'_x(x_0, y_0)| + |f'_y(x_0, y_0)| \neq 0$), то кривые уровня в окрестности этой точки не устроены, как циклические овалы и не являются точками пересечения. Это следует из теоремы о неявной функции: если хотя бы одна из частных производных отлична от нуля, то соответственная переменная может быть локально выражена через другую.

Поэтому мы можем считать, что линии уровня функции f — это такие «почти прямые» кривые $\Gamma_h = \{(x, y) : f(x, y) = h\}$, так сказать, параллельные друг другу и заполняющие окрестность точки (x_0, y_0). С одной стороны от кривой $\Gamma_{f(x_0,y_0)} = \Gamma$ лежат кривые с $h < f(x_0, y_0)$, с другой — кривые с $h > f(x_0, y_0)$, причем неравенства строгие.

Теперь посмотрим, как эти кривые располагаются относительно кривой S, которая проходит через точку (x_0, y_0) и имеет там общую точку с кривой Γ.

В ситуации общего положения эти две кривые (Γ и S) пересекаются под ненулевым углом. В этом случае с одной стороны от точки (x_0, y_0) на кривой S функция f принимает значения больше $f(x_0, y_0)$, с другой — меньше $f(x_0, y_0)$, поэтому в такой ситуации общего положения в точке (x_0, y_0) не может быть экстремума.

Если кривые пересекаются под нулевым углом (касаются), то экстремума тоже нет.

Геометрически единственный случай, когда экстремум возможен и действительно имеет место, это когда кривая S касается кривой Γ, причем они не пересекаются в естественном смысле. Причем если мы по естественным соображениям умеем сказать, как расположены кривые уровня, мы можем сказать, что получится, минимум или максимум.

Вот еще геометрическая картинка.

Представим себе карту местности, гористой, и саму местность. Карта — это уменьшенная вертикальная проекция на горизонтальную плоскость. Пусть по местности идет тропа (кривая линия). Если вы идете по тропе, то вы легко увидите на ней условные экстремумы–минимумы: там будет лужица.

Ровно в этих местах на карте линии уровня будут касаться трека тропы.

Следующая геометрическая картинка в трехмерном пространстве.

Теперь у нас есть функция $f(x, y, z)$, которую мы будем представлять в виде поверхностей уровня. Снова (точка (x_0, y_0, z_0) не критическая) все эти поверхности расположены друг вдоль друга.

Пусть поверхность S — снова одномерная кривая. Тогда из тех же построений легко увидеть, что точка (x_0, y_0, z_0) может быть условным экстремумом, только если в этой точке кривая S касается соответствующей поверхности уровня функции f.

Слова «кривая касается поверхности» означают, естественно, что вектор, касательный кривой в этой точке лежит в касательной плоскости к поверхности уровня.

Теперь мы переходим к общей теореме. Она просматривается из проведенных геометрических конструкций. Пусть в $\mathbb{R}^n$ есть точка x_0, пусть в некоторой окрестности $D \ni x_0$ задана функция $f : D \to \mathbb{R} \in C^1$, причем точка x_0 не является критической для f.

Пусть гладкая поверхность S содержит точку x_0.

Теорема (необходимый признак условного экстремума). *Для того, чтобы точка x_0 была бы точкой условного локального экстремума для функции f, необходимо выполнение условия $TS_{x_0} \subset TN_{x_0}$, где TS_{x_0} — пространство, касательное к S, а TN_{x_0} — пространство, касательное к поверхности уровня $N = \{x \in D : f(x) = f(x_0)\}$.*

Для того, чтобы запомнить, что чему должно принадлежать, надо вспомнить, что размерность S (и размерность TS_{x_0}) может быть любая от 1 до $n-1$, а размерность поверхности уровня N (и размерность TN_{x_0}) обязательно равна $n-1$.

Доказательство. Возьмем произвольный вектор $\xi \in TS_{x_0}$ и такую гладкую кривую $x(t)$ на поверхности S, что $x'(0) = \xi$. это можно всегда сделать, мы это доказали. Если x_0 — точка локального условного экстремума, то точка 0 — точка локального экстремума гладкой функции $f(x(t))$. Поэтому, необходимо, $(f(x))'_t(0) = 0$, то есть $f'(x_0)x'(0) = (\nabla f(x_0), \xi) = 0$. Так как точка x_0 — не критическая, то есть $\nabla f(x_0) \neq 0$, то это равенство как раз и означает, что $\xi \in TN_{x_0}$. $\square$

Лекция 8 (20 октября 2014)

На прошлой лекции мы сделали следующее.

1) Рассмотрели несколько примеров поверхностей.

2) Написали уравнения для касательных векторов и для касательных плоскостей.

3) Доказали теорему о том, что касательная плоскость состоит из касательных векторов.

4) Сформулировали и доказали геометрическое необходимое условие условного экстремума.

Метод множителей Лагранжа.

Теорема с прошлой лекции красиво формулируется и легко доказывается, однако для прямого применения в конкретных задачах она не кажется приемлемой. Зато их нее вытекает замечательный метод практического нахождения точек, подозрительных на условный экстремум — метод множителей Лагранжа.

Пусть поверхность S задана системой из m уравнений

$$\begin{cases} F_1(x_1,\dots,x_n) & = 0, \\ \quad\dots\dots \\ F_m(x_1,\dots,x_n) & = 0. \end{cases}$$

Пусть ранг этой системы в окрестности точки x_0 постоянный и равен m, то эта система уравнений задает в окрестности точки x_0 гладкую поверхность размерности $k = n - m$.

Тогда пространство TS_{x_0} задается системой линейных уравнений

$$\begin{cases} \partial_1 F_1(x_0)\xi_1 & +\dots+ & \partial_n F_1(x_0)\xi_n & = 0, \\ \dots\dots\dots\dots \\ \partial_1 F_m(x_0)\xi_1 & +\dots+ & \partial_n F_m(x_0)\xi_n & = 0. \end{cases}$$

Пространство TN_{x_0} задается уравнением

$$\partial_1 f(x_0)\xi_1 + \dots + \partial_n f(x_0)\xi_n = 0.$$

Условия теоремы означают, что это уравнение является следствием системы уравнений. Таким образом, линейная комбинация уравнений системы является вот этим уравнением.

Иначе говоря, при некоторых λ_i справедливо равенство

$$\mathbf{grad}\, f(x_0) = \sum_{i=1}^{m} \lambda_m\, \mathbf{grad}\, F_i(x_0),$$

Это n скалярных равенств. Существование таких λ_i, что выполнено такое равенство в $\mathbb{R}^n$, и есть необходимое условие локального условного экстремума.

Обычно делают так. Пишут *функцию Лагранжа*: $L(x,\lambda) = f(x) - \sum\limits_{i=1}^{m} \lambda_i F_i(x)$ от $m+n$ переменных, n переменных x и m переменных λ. Теперь выписываем условия обычного безусловного экстремума для функции L по всем переменным — это как раз будут равенства для градиента и для поверхности:

$$\frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial \lambda_j} = 0, \qquad i = 1, \ldots, n; \; j = 1, \ldots, m.$$

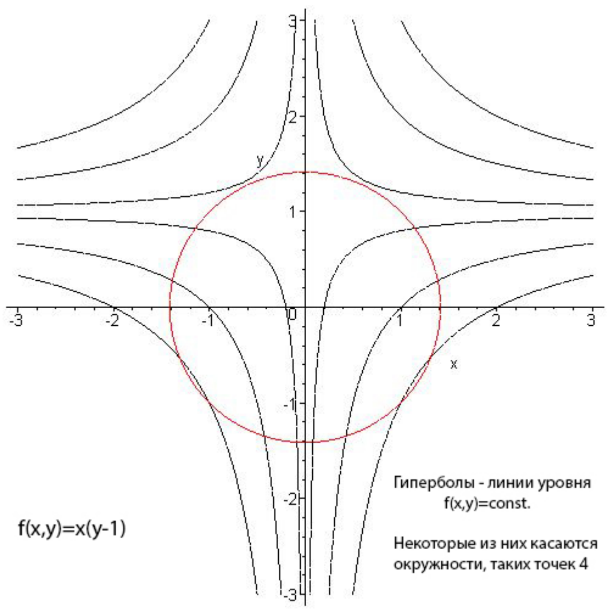

Рис. 7 Пример.
Найти точки условного экстремума функции

$$f(x, y) = x(y - 1)$$

на окружности $x^2 + y^2 = 2$. Пишем функцию Лагранжа

$$L(x, y, \lambda) = x(y - 1) - \lambda(x^2 + y^2 - 2)$$

и ищем её критические точки. Уравнения

$$x^2 + y^2 - 2 = y - 1 + 2\lambda x = x + 2\lambda y = 0.$$

Решаем, получаем 4 решения, 2 условных минимума, 2 условных максимума.

Замечания.

1. Можно метод множителей Лагранжа применять и к случаю $k < n - m$. То есть, не страшно, если будет лишнее условие.

2. Пишут функцию Лагранжа и так: $L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i F_i(x)$.

Достаточные условия условного экстремума.

Достаточные условия экстремума даже в случае безусловном были не слишком простые: нужно было проверять знакоопреденность квадратичной формы. Естественно, в случае условного экстремума они не могут быть «проще».

Пусть в $\mathbb{R}^n$ есть точка x_0, пусть в некоторой окрестности $D \ni x_0$ задана функция $f : D \to \mathbb{R} \in C^2$, причем точка x_0 не является критической для f.

Пусть гладкая поверхность S содержит точку x_0. Пусть в этой точке при некоторых λ_j выполнено необходимое условие локального условного экстремума: функция Лагранжа $L(x) = L(x, \lambda)$ имеет критическую точку.

Теорема (достаточные условия условного экстремума). *Для того, чтобы точка x_0 была точкой условного экстремума функции f достаточно, чтобы квадратичная форма $\sum \partial_{ij} L(x_0) \xi_i \xi_j$*

была знакоопределенной на TS_{x_0}. Для того, чтобы точка x_0 не была точкой условного экстремума функции f достаточно, чтобы квадратичная форма $\sum \partial_{ij} L(x_0)\xi_i\xi_j$ не была знакоопределенной (нестрого) на TS_{x_0}.

Как и ранее, случай нестрогой определённости-неопределённости этой формы оставляет вопрос об экстремуме открытым.

Если мы покажем, что функция L имеет условный экстремум в точке x_0, то и функция f имеет там условный экстремум: $f(x) = L(x)$ при $x \in S$.

Повторю: приведенное достаточное условие должно быть выполнено только для векторов из касательного подпространства, вовсе не для любых векторов из большого пространства. В пространстве $\mathbb{R}^n$ (большом) есть маленькое касательное подпространство. Форма определена на n-мерных векторах, принадлежащих k-мерному касательному подпространству.

Доказательство. Так как выполнено необходимое условие экстремума и так как на S частные производные по λ_j равны нулю, то тейлоровское разложение функции $L(x)$ в x_0 имеет вид

$$L(x) = L(x_0) + \frac{1}{2!}\sum_{i,j=1}^{n} \frac{\partial^2 L(x_0)}{\partial x_i \partial x_j}(x_i - x_{i,0})(x_j - x_{j,0}) + o(\|x - x_0\|^2),$$

линейные слагаемые обращаются в ноль на S.

В силу определения поверхности существует гладкое биективное отображение $x(t)$ окрестности нуля в $\mathbb{R}^k$ в $\mathbb{R}^n$ на множество $U \bigcap S$, где $U \in \mathbb{R}^n$ — окрестность точки x_0.

Из соотношения $x(t) = x(0) + x'(0)t + o(\|t\|)$ (здесь $t \in \mathbb{R}^k$, $x'(0) \in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^n)$, $x \in \mathbb{R}^n$, $x(0) = x_0$) следует соотношение

$$\|x(t) - x(0)\|_{\mathbb{R}^n} \sim \|t\|_{\mathbb{R}^k}.$$

Таким образом,

$$L(x(t)) - L(x(0)) = \frac{1}{2!}\sum_{i,j=1}^{n} \frac{\partial^2 L(x_0)}{\partial x_i \partial x_j} \sum_{r,s=1}^{k} \partial_s x_i(0)\partial_r x_j(0) t_r t_s + o(\|t\|^2).$$

Сейчас покажем, что квадратичная форма в правой части — это другая форма записи формы

$$\sum_{i,j=1}^{n} \frac{\partial^2 L(x_0)}{\partial x_i(0)\partial x_j(0)}\xi_i\xi_j$$

при $\xi \in S$. Тогда все будет доказано: знак выражения $L(x(t)) - L(x(0))$ совпадает со знаком квадратичной формы.

Совпадение форм следует из равенств $\xi_i = \sum_{s=1}^{k} \partial_s x_i(0) t_s$. Эти равенства следуют из параметрической записи $\xi = x'(0)t$ для векторов касательного пространства TS_{x_0}. $\square$

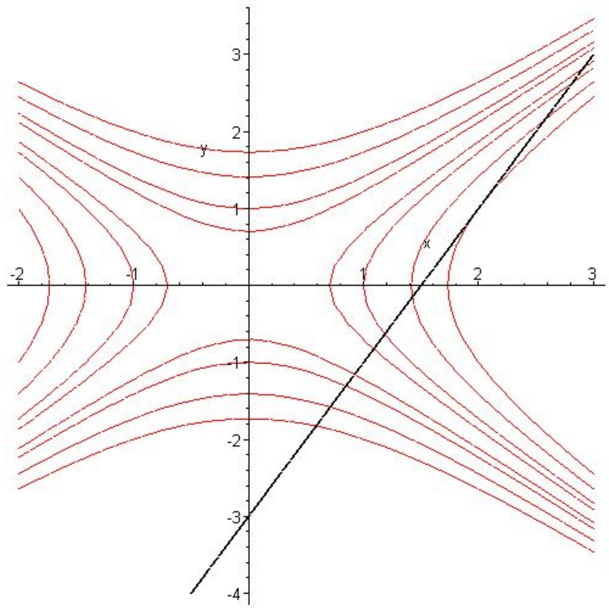

Рис. 8 Игрушечный пример.
$f(x,y) = x^2 - y^2$, $2x - y - 3 = 0$.
Функция Лагранжа
$L(x,\lambda) = x^2 - y^2 - \lambda(2x - y - 3)$.
Уравнения подозрительных точек:
$2x - y - 3 = 2x - 2\lambda = -2y + \lambda = 0$.
Решение: $x = 2,\ y = 1,\ \lambda = 2$.
Форма имеет вид: $\xi_1^2 - \xi_2^2$, уравнение касательной: $2\xi_1 = \xi_2$, поэтому на касательном пространстве форма имеет вид $-3(\xi_2)^2$, то есть это точка максимума.
Такой пример легко решить и без Лагранжа.
Обратите внимание, что форма на всем пространстве не является знакоопределенной, а на касательной становится знакоопределенной.

Замечания.

1. Задача на условный экстремум: $f(x,y) = x^2 + y^2$, $\ x^2/4 + y^2/9 \leqslant 1$.

Это простая задача: сначала надо поискать экстремумы во внутренних точках эллипса. Для этого нужно приравнять к нулю градиент f, естественно, мы найдем минимум в точке $(0,0)$.

Теперь нужно пытаться найти «интересные» точки на границе эллипса.

Таких точек будет 2: $(0,3)$ и $(2,0)$. В этих точках линии уровня f — окружности с центром в нуле касаются границы эллипса. При этом в точке $(2,0)$ касание внутреннее, там точка, подозрительная на минимум, и это в самом деле будет минимум на границе, но внутри эллипса будут меньшие значения. А в точке $(0,3)$ касание внешнее, эта точка подозрительна на максимум, и является максимумом.

Ответ: есть точка $(0,0)$ — минимум, точка $(0,3)$ — максимум. Одна точка внутри области. друга на границе.

2. Задача об условном экстремуме рассматривается не только на гладких многообразиях, да и не только гладкие функции возникают в практических приложениях. Сказать про негладкие теоремы. Что бывает выпуклый анализ, негладкий анализ.

2. Рассказать про принцип максимума, про вариационное исчисление, про уравнения Эйлера–Лагранжа, про бесконечномерные задачи.

3. Сказать что-то про задачу линейного программирования.

4. Задача о брахистохроне — кривой скорейшего спуска. Задача о её нахождении была поставлена в 1696 году Иоганном Бернулли.

Среди плоских кривых, соединяющих две данные точки A и B, лежащих в одной вертикальной плоскости (B ниже A), найти ту, двигаясь по которой под действием только силы тяжести, сонаправленной отрицательной полуоси OY, материальная точка достигнет B из A за кратчайшее время.

Метод решения, полученного 26 января 1697 года Ньютоном, лег в основу важнейшей области естествознания — вариационного исчисления.

Решением задачи о брахистохроне является дуга циклоиды с горизонтальным основанием, точка возврата которой находится в точке A, или иными словами, имеющая вертикальную касательную в точке A.

5. В задаче об условном экстремуме мы решали только локальные задачи. Мы искали точки, подозрительные на экстремум и их исследовали.

Бывает другой круг задач.

Есть поверхность — например сфера, или тор, или крендель. На нем есть гладкая функция, у которой все критические точки невырождены, называется *функция Морса*. Оказывается, тип поверхности определяет минимальное количество критических точек.

На сфере есть как минимум точка минимума и точка максимума. А на торе есть как минимум 4 критических точки.