
ГЛАВА 1

Теория Морса на поверхностях

Теория Морса изучает, как связаны функции, определённые на многообразии, со структурой многообразия. В первой главе эта взаимосвязь будет показана для функций, определённых на поверхностях. Поверхности проще представить и изобразить, а основные понятия и принципы теории Морса проявляются в полной мере и в двумерном случае.

1.1. Критические точки функций

Рассмотрим функцию одной переменной $y = f(x)$, где x и y — вещественные числа. Напомним, что найти промежутки возрастания и убывания функции можно следующим образом: нужно вычислить производную функции, найти точки x_0 , для которых $f'(x_0) = 0$ и определить знак производной на получившихся интервалах. Точка x_0 , удовлетворяющая условию

$$f'(x_0) = 0,$$

называется *критической точкой* функции f . Примерами критических точек являются точки максимума и минимума, а также точки перегиба, как у функции $y = x^3$ при $x = 0$.

В зависимости от значения второй производной $f''(x_0)$, критические точки разбиваются на два класса. Критическая точка x_0 называется *невыврожденной*, если $f''(x_0) \neq 0$, и *вырожденной*, если $f''(x_0) = 0$.

ПРИМЕР 1.1. Для квадратичной функции $y = x^2$, $y'' = 2$; поэтому критическая точка $x = 0$ является невырожденной.

При $n \geq 3$, критическая точка $x = 0$ функции $y = x^n$ будет вырожденной, поскольку $y'' = n(n-1)x^{n-2}$ при $x = 0$ принимает нулевое значение.

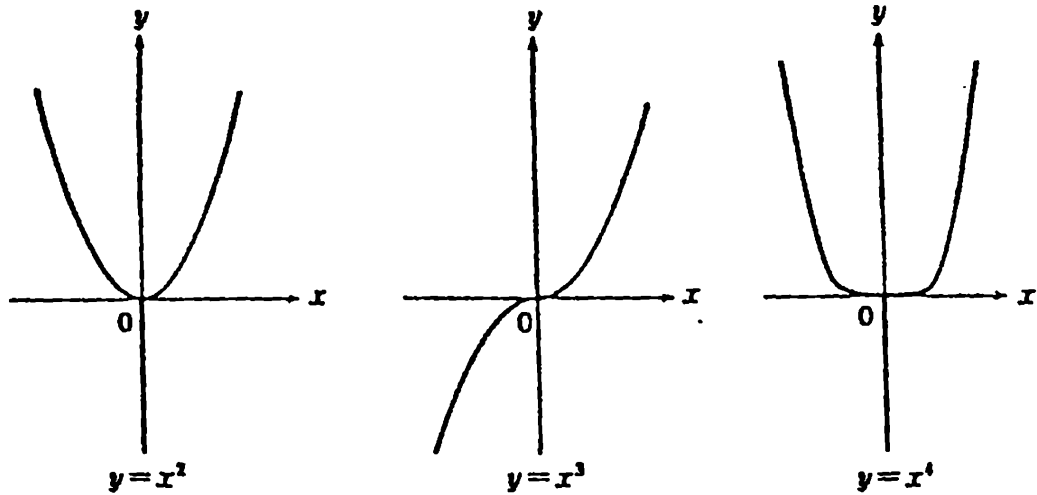


Рис. 1.1: Графики функций $y = x^2$, $y = x^3$ и $y = x^4$.

Посмотрим, как графики разных функций касаются оси абсцисс (см. рисунок 1.1). Визуально, график подходит к критической точке под меньшим углом наклона, когда критическая точка невырождена.

Другое различие между вырожденными и невырожденными критическими точками проявится, если немного «пошевелить» функцию. Рассмотрим квадратичную функцию $y = x^2$ и кубическую функцию $y = x^3$. Для этих функций $x = 0$ является критической точкой: в первом случае — невырожденной, во втором — вырожденной. Немного изменим эти функции, добавив линейную функцию $y = ax + b$.

Найдём критические точки возмущённой квадратичной функции:

$$(1.1) \quad y = x^2 + ax + b.$$

Её производная равна

$$y' = 2x + a.$$

Точка $x = -a/2$, в которой производная принимает нулевое значение, является единственной критической точки функции (1.1). Поскольку вторая производная $y'' = 2$, данная критическая точка также будет невырожденной. Нестрого говоря, при «шевелении» функции невырожденная критическая точка также «шевелится» — смещается на расстояние $a/2$ — и остаётся невырожденной; новых критических точек не появляется.

Теперь рассмотрим случай вырожденной критической точки. Производная возмущённой кубической функции

$$(1.2) \quad y = x^3 + ax + b$$

равна

$$y' = 3x^2 + a.$$

и принимает нулевое значение при

$$(1.3) \quad x = \pm \sqrt{-\frac{a}{3}}.$$

Решения (1.3) являются комплексными при $a > 0$, так что у возмущённой кубической функции при $a > 0$ теряется критическая точка.

Если $a < 0$, то значения x (1.3) вещественны и, следовательно, у функции есть две критические точки. Поскольку вторая производная $y'' = 6x$ принимает ненулевые значения при $x = \pm \sqrt{-a/3}$, эти критические точки невырождены.

Получаем (см. рисунок 1.2), что вырожденная критическая точка $x = 0$ функции $y = x^3$ при «шевелении» исчезает ($a > 0$), либо разделяется на две невырожденные критические точки ($a < 0$).

Можно сказать, что в некотором смысле невырожденные точки являются «устойчивыми», а вырожденные — «неустойчивыми» относительно линейного возмущения.

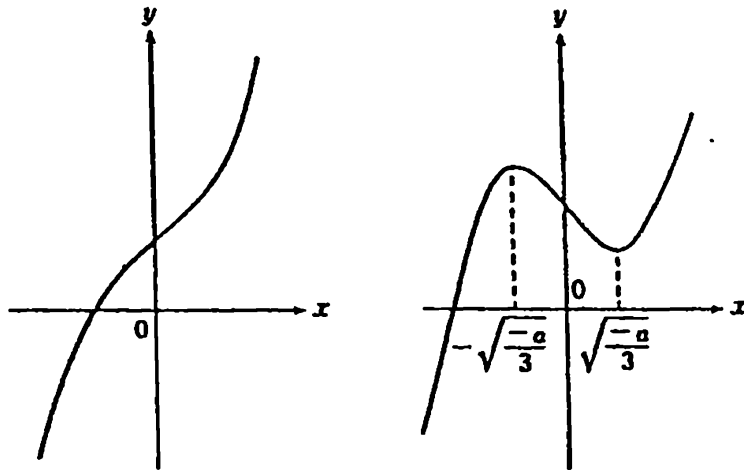


Рис. 1.2: Графики функции $y = x^3 + ax + b$ при $a > 0$ (слева) и $a < 0$ (справа).

1.2. Гессиан

Теперь рассмотрим вещественнозначную функцию

$$(1.4) \quad z = f(x, y)$$

двух вещественных переменных x и y . Будем считать пары аргументов (x, y) точками на плоскости. Тогда f — функция, каждой точке плоскости сопоставляющая вещественное число, и мы можем изобразить её график в трёхмерном пространстве (см. рисунок 1.3).

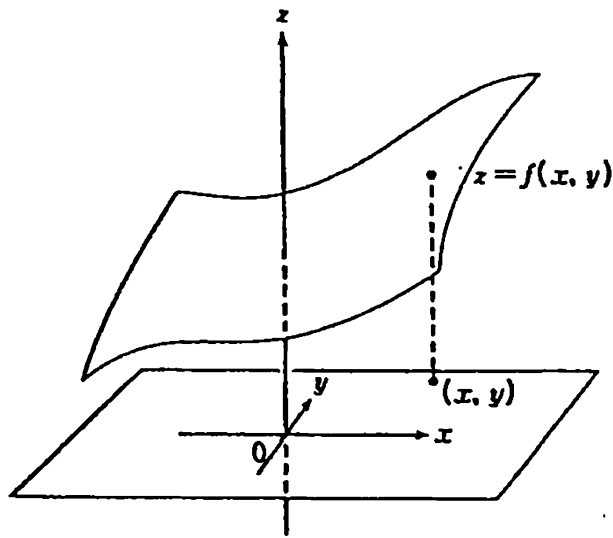


Рис. 1.3: Пример графика функции двух переменных.

ОПРЕДЕЛЕНИЕ 1.2. (Критические точки функции двух переменных) Точка $p_0 = (x_0, y_0)$ называется *критической точкой* функции $z = f(x, y)$, если выполнено условие:

$$(1.5) \quad \frac{\partial f}{\partial x}(p_0) = 0, \quad \frac{\partial f}{\partial y}(p_0) = 0.$$

В этом определении мы предполагаем, что $f(x, y)$ принадлежит классу C^∞ , т. е. бесконечно дифференцируема. В этой книге мы будем рассматривать только функции класса C^∞ .

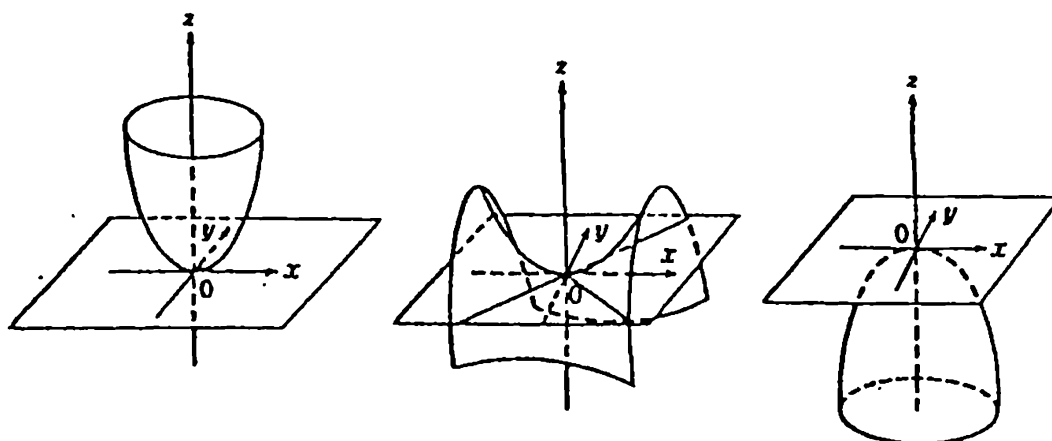


Рис. 1.4: Слева направо: графики функций $z = x^2 + y^2$, $z = x^2 - y^2$, $z = -x^2 - y^2$.

ПРИМЕР 1.3. Точка $O = (0, 0)$ является критической точкой следующих трёх функций:

$$(1.6) \quad z = x^2 + y^2, \quad z = x^2 - y^2, \quad z = -x^2 - y^2.$$

(см. рисунок 1.4).

Зададимся вопросом: как определить вырожденные и невырожденные критические точки в двумерном случае?

После некоторого раздумья, у читателя может возникнуть искушение сказать, что критическая точка является невырожденной, если

$$(1.7) \quad \frac{\partial^2 f}{\partial x^2}(p_0) \neq 0, \quad \frac{\partial^2 f}{\partial y^2}(p_0) \neq 0.$$

Однако, такое определение плохо тем, что условия (1.7) для той же функции f и точки p_0 могут перестать выполняться после замены координат (см. пример 1.6). Мы же хотим обобщить понятие невырожденной (и вырожденной) критической точки так, чтобы оно не зависело от выбора системы координат. Следующее определение удовлетворяет этому условию.

ОПРЕДЕЛЕНИЕ 1.4. (i) Пусть $p_0 = (x_0, y_0)$ — критическая точка функции $z = f(x, y)$. Матрица

$$(1.8) \quad \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(p_0) & \frac{\partial^2 f}{\partial x \partial y}(p_0) \\ \frac{\partial^2 f}{\partial y \partial x}(p_0) & \frac{\partial^2 f}{\partial y^2}(p_0) \end{pmatrix}$$

вторых производных, вычисленных в точке p_0 , называется *матрицей Гессе* функции $z = f(x, y)$ в критической точке p_0 , и обозначается $\mathcal{H}_f(p_0)$. Определитель матрицы Гессе называется *гессианом*.

(ii) Критическая точка p_0 функции $z = f(x, y)$ называется *невырожденной*, если гессиан функции f в точке p_0 принимает нулевое значение; то есть, p_0 — невырожденная критическая точка, если выполнено условие:

$$(1.9) \quad \det \mathcal{H}_f(p_0) = \frac{\partial^2 f}{\partial x^2}(p_0) \frac{\partial^2 f}{\partial y^2}(p_0) - \left(\frac{\partial^2 f}{\partial x \partial y}(p_0) \right)^2 = 0.$$

В противном случае, если $\det \mathcal{H}_f(p_0) = 0$, точка p_0 называется *вырожденной критической точкой*.

Заметим, что матрица $\mathcal{H}_f(p_0)$ симметрична, поскольку $\frac{\partial^2 f}{\partial x \partial y}(p_0) = \frac{\partial^2 f}{\partial y \partial x}(p_0)$.

ПРИМЕР 1.5. Вычислим матрицу Гессе для функций из примера 1.3 в точке $(0, 0)$.

(i) Матрица Гессе в нуле функции $z = x^2 + y^2$ равна

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

(ii) Матрица Гессе в нуле функции $z = x^2 - y^2$ равна

$$\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

(iii) Матрица Гессе в нуле функции $z = -x^2 - y^2$ равна

$$\begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Определители всех перечисленных матриц не равны нулю, поэтому точка $(0, 0)$ будет невырожденной критической точкой всех трёх функций.

ПРИМЕР 1.6. Рассмотрим функцию $z = xy$. Точка $(0, 0)$ является критической. Матрица Гессе в нуле равна

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

и её определитель не равен нулю, поэтому $(0, 0)$ — невырожденная критическая точка. Заметим, что функция $z = xy$ получается из функции $z = x^2 - y^2$ заменой координат.

ПРИМЕР 1.7. Точка $(0, 0)$ является критической для функции $z = x^2 + y^3$, однако определитель матрицы Гессе этой функции в нуле

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

равен нулю. Следовательно, $(0, 0)$ — вырожденная критическая точка функции $z = x^2 + y^3$.

Как изменяется матрица Гессе при замене координат?

ЛЕММА 1.8. Пусть p_0 — критическая точка функции $z = f(x, y)$. Обозначим за $\mathcal{H}_f(p_0)$ матрицу Гессе, вычисленную в координатах (x, y) , и за $\tilde{\mathcal{H}}_f(p_0)$ матрицу Гессе, вычисленную в других координатах (X, Y) . Тогда верно соотношение:

$$(1.10) \quad \tilde{\mathcal{H}}_f(p_0) = J^T(p_0) \mathcal{H}_f(p_0) J(p_0),$$

где $J(p_0)$ — матрица Якоби замены координат, определённая как

$$(1.11) \quad \begin{pmatrix} \frac{\partial x}{\partial X}(p_0) & \frac{\partial x}{\partial Y}(p_0) \\ \frac{\partial y}{\partial X}(p_0) & \frac{\partial y}{\partial Y}(p_0) \end{pmatrix},$$

и $J^T(p_0)$ — транспонированная матрица Якоби.

ДОКАЗАТЕЛЬСТВО. Доказательство предлагается провести читателю самостоятельно. Идея заключается в замене

$$\begin{aligned} \frac{\partial f}{\partial X} &= \frac{\partial f}{\partial x} \frac{\partial x}{\partial X} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial X}, \\ \frac{\partial f}{\partial Y} &= \frac{\partial f}{\partial x} \frac{\partial x}{\partial Y} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial Y} \end{aligned}$$

и последующем выражении $\frac{\partial^2 f}{\partial X^2}$, $\frac{\partial^2 f}{\partial X \partial Y}$ и $\frac{\partial^2 f}{\partial Y^2}$ через $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial x \partial y}$ и $\frac{\partial^2 f}{\partial y^2}$. Затем нужно вычислить эти производные в точке p_0 , не забывая, что $\frac{\partial f}{\partial x}(p_0) = \frac{\partial f}{\partial y}(p_0) = 0$, и сравнить результат с произведением матриц из правой части. \square

ПРИМЕР 1.9. Перепишем функцию $z = xy$ (Пример 1.6) в новых координатах

$$(1.12) \quad \begin{cases} x = X + Y, \\ y = X - Y. \end{cases}$$

Получим

$$xy = (X - Y)(X + Y) = X^2 - Y^2,$$

т. е. вторую функцию из Примера 1.3. Матрицы Гессе в нуле относительно координат (x, y) и (X, Y) равны, соответственно,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Матрица Якоби преобразования координат (1.12) равна

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Убедимся, что утверждение Леммы 1.8 выполняется:

$$\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

СЛЕДСТВИЕ 1.10. (из Леммы 1.8) Свойства невырожденности и вырожденности критической точки не зависят от выбора координат.

ДОКАЗАТЕЛЬСТВО. Перейдём к определителям в равенстве из утверждения $\tilde{\mathcal{H}}_f(p_0) = J^T(p_0)\mathcal{H}_f(p_0)J(p_0)$ Леммы 1.8:

$$(1.13) \quad \det \tilde{\mathcal{H}}_f(p_0) = \det J^T(p_0) \det \mathcal{H}_f(p_0) \det J(p_0).$$

Заметим, что якобиан преобразования координат

$$(1.14) \quad \det J(p_0) \neq 0,$$

поэтому $\det \tilde{\mathcal{H}}_f(p_0) = 0$ тогда и только тогда, когда $\det \mathcal{H}_f(p_0) = 0$, откуда следует утверждение Следствия 1.10. \square

1.3. Лемма Морса

В этом разделе мы докажем следующий факт.

ТЕОРЕМА 1.11. (Лемма Морса). Пусть p_0 — невырожденная критическая точка функции f двух переменных. Тогда существуют локальные координаты (X, Y) , в которых функция f принимает одну из трёх стандартных форм:

$$\begin{aligned} (i) \quad & f = X^2 + Y^2 + c, \\ (ii) \quad & f = X^2 - Y^2 + c, \\ (iii) \quad & f = -X^2 - Y^2 + c, \end{aligned}$$

где $c = f(p_0)$, а p_0 — нуль новой системы координат ($p_0 = (0, 0)$).

Эта теорема утверждает, что функция в окрестности невырожденной критической точки ведёт себя предельно просто: подходящая замена координат приводит в случаю одной из трёх простых функций, рассмотренных в Примере 1.3.

Поскольку ни у одной из этих функций нет критических точек, близких к нулю, верно следующее утверждение.

СЛЕДСТВИЕ 1.12. Невырожденные критические точки функции двух переменных изолированы.

Это утверждение верно и в общем случае, для функции m переменных.

Теперь докажем Теорему 1.11.

ДОКАЗАТЕЛЬСТВО. Рассмотрим произвольную систему координат в окрестности точки p_0 . Пусть p_0 — точка $(0, 0)$ в этих координатах. Также для простоты изложения будем считать¹, что $f(p_0) = 0$. Далее мы покажем, что, не ограничивая общности, мы можем положить

$$(1.15) \quad \frac{\partial^2 f}{\partial x^2}(p_0) \neq 0.$$

Если в нашей системе координат $\frac{\partial^2 f}{\partial x^2}(p_0) \neq 0$, то доказывать нечего. Если $\frac{\partial^2 f}{\partial x^2}(p_0) = 0$, но $\frac{\partial^2 f}{\partial y^2}(p_0) \neq 0$, то мы можем просто поменять местами оси x и y , чтобы условие (1.15) выполнялось. Рассмотрим случай, когда $\frac{\partial^2 f}{\partial x^2}(p_0) = 0$ и $\frac{\partial^2 f}{\partial y^2}(p_0) = 0$.

В этом случае матрица Гессе в точке $p_0 = (0, 0)$ выглядит так:

$$(1.16) \quad \begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix}.$$

Поскольку p_0 — невырожденная критическая точка, $a \neq 0$. Введём новые локальные координаты (X, Y) :

$$(1.17) \quad x = X + Y, \quad y = X - Y.$$

Матрица Якоби перехода от координат (X, Y) к (x, y) :

¹В общем случае можно рассмотреть функцию $g(x) = f(x) - f(p_0)$ и применить к ней все дальнейшие рассуждения

$$(1.18) \quad J = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Пользуясь Леммой 1.8, вычислим матрицу Гессе $\tilde{\mathcal{H}}_f$ в координатах (X, Y) :

$$(1.19) \quad \tilde{\mathcal{H}}_f = J^T \mathcal{H}_f J = \begin{pmatrix} 2a & 0 \\ 0 & -2a \end{pmatrix}.$$

Мы видим, что в новых координатах

$$(1.20) \quad \frac{\partial^2 f}{\partial X^2}(p_0) = 2a \neq 0, \quad \frac{\partial^2 f}{\partial Y^2}(p_0) = -2a \neq 0.$$

Для удобства будем обозначать новые координаты (X, Y) как (x, y) . Заметим, что теперь условие 1.15 выполнено, и мы можем продолжить доказательство в предположении 1.15.

В анализе функций многих переменных известен следующий факт. Пусть функция $z = f(x, y)$ задана в окрестности нуля и $f(0, 0) = 0$. Тогда существуют функции $g(x, y)$ и $h(x, y)$ такие, что

$$(1.21) \quad f(x, y) = xg(x, y) + yh(x, y)$$

в некоторой окрестности нуля, и

$$(1.22) \quad \frac{\partial f}{\partial x}(0, 0) = g(0, 0), \quad \frac{\partial f}{\partial y}(0, 0) = h(0, 0).$$

Этот факт достаточно известен, но для удобства читателя мы его докажем ниже.

Для простоты скажем, что $z = f(x, y)$ определена на всей плоскости Oxy . Фиксируем произвольную точку (x, y) . Рассмотрим функцию $f(xt, yt)$ с параметром t . Если мы продифференцируем эту функцию по параметру t , а затем проинтегрируем, мы получим исходную функцию. В частности, рассмотрим определённый интеграл от 0 до 1 и воспользуемся условием $f(0, 0) = 0$:

$$(1.23) \quad f(x, y) = \int_0^1 \frac{df(tx, ty)}{dt} dt = \int_0^1 \left\{ x \frac{df}{dx}(tx, ty) + y \frac{df}{dy}(tx, ty) \right\} dt = \\ = xg(x, y) + yh(x, y).$$

Здесь мы ввели обозначения

$$(1.24) \quad g(x, y) = \int_0^1 \frac{df}{dx}(tx, ty) dt, \quad h(x, y) = \int_0^1 \frac{df}{dy}(tx, ty) dt.$$

Таким образом, мы нашли функции g и h , удовлетворяющие равенству (1.21). Более того, подстановкой $(x, y) = (0, 0)$ в (1.24) убеждаемся, что равенства (1.22) также верны.

Вернёмся к доказательству теоремы. Поскольку $p_0 = (0, 0)$ — критическая точка функции f , мы имеем:

$$(1.25) \quad g(0, 0) = \frac{\partial f}{\partial x}(0, 0) = 0, \quad h(0, 0) = \frac{\partial f}{\partial y}(0, 0) = 0.$$

Теперь применим доказанное выше утверждение к функциям $g(x, y)$ и $h(x, y)$. Существуют функции h_{11}, h_{12}, h_{21} и h_{22} , для которых верны соотношения:

$$(1.26) \quad g(x, y) = xh_{11}(x, y) + yh_{12}(x, y)$$

и

$$(1.27) \quad h(x, y) = xh_{21}(x, y) + yh_{22}(x, y).$$

Подставим равенства (1.26) и (1.27) в (1.21):

$$(1.28) \quad f(x, y) = x^2h_{11} + xy(h_{12} + h_{21}) + y^2h_{22}.$$

Введём новые обозначения: $H_{11} = h_{11}$, $H_{12} = (h_{12} + h_{21})/2$, $H_{22} = h_{22}$. Приведём (1.28) к более простому виду:

$$(1.29) \quad f(x, y) = x^2H_{11} + xyH_{12} + y^2H_{22}.$$

Отсюда простыми вычислениями получаем:

$$(1.30) \quad \begin{cases} \frac{\partial^2 f}{\partial x^2}(0, 0) = 2H_{11}(0, 0), \\ \frac{\partial^2 f}{\partial x \partial y}(0, 0) = \frac{\partial^2 f}{\partial y \partial x}(0, 0) = 2H_{12}(0, 0), \\ \frac{\partial^2 f}{\partial y^2}(0, 0) = 2H_{22}(0, 0). \end{cases}$$

Теперь вспомним, что левая часть верхнего уравнения не равна нулю. Значит, $H_{11}(0, 0) \neq 0$, а из непрерывности H_{11} следует, что

$$(1.31) \quad H_{11}(x, y) \neq 0 \text{ в некоторой окрестности точки } (0, 0).$$

Введём новую координату X в окрестности нуля:

$$(1.32) \quad X = \sqrt{|H_{11}|} \left(x + \frac{H_{12}}{H_{11}} y \right).$$

Координату y оставим прежней. Поскольку якобиан перехода от (x, y) к (X, y) в точке $(0, 0)$ ненулевой, локальные координаты (X, y) корректно определены в некоторой окрестности нуля. Возведём X в квадрат:

$$(1.33) \quad \begin{aligned} X^2 &= |H_{11}| \left(x^2 + 2 \frac{H_{12}}{H_{11}} xy + \frac{H_{12}^2}{H_{11}^2} y^2 \right) \\ &= \begin{cases} H_{11}x^2 + 2H_{12}xy + \frac{H_{12}^2}{H_{11}} y^2 & (H_{11} > 0), \\ -H_{11}x^2 - 2H_{12}xy - \frac{H_{12}^2}{H_{11}} y^2 & (H_{11} < 0). \end{cases} \end{aligned}$$

Сравнивая (1.33) и (1.29), получаем, что при $H_{11} > 0$

$$(1.34) \quad f = X^2 + \left(H_{22} - \frac{H_{12}^2}{H_{11}} \right) y^2,$$

а при $H_{11} < 0$

$$(1.35) \quad f = -X^2 + \left(H_{22} - \frac{H_{12}^2}{H_{11}} \right) y^2.$$

Из равенств (1.30) следует, что

$$(1.36) \quad H_{11}(0, 0)H_{22}(0, 0) - H_{12}^2(0, 0) = \frac{1}{4} \det \mathcal{H}_f \neq 0,$$

где $\mathcal{H}_f \neq 0$ из условия невырожденности критической точки p_0 . Теперь выберем новую координату Y в окрестности точки $p_0 = (0, 0)$:

$$(1.37) \quad Y = \sqrt{\left| \frac{H_{11}H_{22} - H_{12}^2}{H_{11}} \right|} y,$$

и перепишем равенства (1.34) и (1.35). Тогда f представляется в локальных координатах (X, Y) одним из следующих способов:

$$(1.38) \quad f = \begin{cases} X^2 + Y^2 & (H_{11} > 0, K > 0), \\ X^2 - Y^2 & (H_{11} > 0, K < 0), \\ -X^2 + Y^2 & (H_{11} < 0, K < 0), \\ -X^2 - Y^2 & (H_{11} < 0, K > 0), \end{cases}$$

где для краткости мы обозначили $K = H_{11}H_{22} - H_{12}^2$. Если мы поменяем координаты X и Y поворотом на 90° , то заметим, что $f = X^2 - Y^2$ и $f = -X^2 + Y^2$ — это одна и та же стандартная форма. Это замечание завершает доказательство Теоремы 1.11. \square

ОПРЕДЕЛЕНИЕ 1.13. (Индекс невырожденной критической точки). Пусть p_0 — невырожденная критическая точка функции f двух переменных. В подходящей системе координат (x, y) в некоторой окрестности точки p_0 функция f представляется в одной из стандартных форм, перечисленных в Теореме 1.11. *Индексом* невырожденной критической точки p_0 назовём число 0, 1 или 2, если функция f принимает вид $f = x^2 + y^2 + c$, $f = x^2 - y^2 + c$ или $f = -x^2 - y^2 + c$, соответственно. Иначе говоря, индекс точки p_0 равен количеству знаков «минус» в стандартной форме.

Как можно заметить по соответствующим графикам (см. рисунок 1.4) функций $f = x^2 + y^2$, $f = x^2 - y^2$ и $f = -x^2 - y^2$, если у точки p_0 индекс 0, то p_0 — точка минимума. Если у p_0 индекс 1, то у функции в любой окрестности точки p_0 будут как значения, большие $f(p_0)$, так и значения, меньшие $f(p_0)$. Если у p_0 индекс 2, то p_0 — точка максимума. Таким образом, индекс критической точки p_0 определяется поведением функции вблизи p_0 .

Докажем, что индекс критической точки определён корректно. В Примере 1.5 мы вычислили матрицы Гессе функций $f = x^2 + y^2$, $f = x^2 - y^2$ и $f = -x^2 - y^2$: они равны

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \text{ и } \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix},$$

соответственно.

Мы можем рассматривать эти матрицы как полученные диагонализацией матрицы Гессе $\mathcal{H}_f(p_0)$ функции f . Действительно, пусть J — матрица Якоби преобразования координат, переводящих f в одну из стандартных форм. Тогда $J^T \mathcal{H}_f(p_0) J$ — диагональная матрица (одна из перечисленных выше), т. е. матрица, у которой ненулевыми являются только элементы на диагонали. Согласно закону Сильвестра, количество знаков «минус» у диагонализированной

симметричной матрицы не зависит от способа её диагонализации. Следовательно, функция f приводится ровно к одной из выше перечисленных форм, и индекс критической точки определён корректно.

Во второй главе мы обобщим Теорему 1.11 и связанные с ней определения на случай произвольной (конечной) размерности.

1.4. Функции Морса на поверхностях

В предыдущих разделах мы исследовали локальное поведение функции вблизи её критических точек. Здесь мы перейдём к исследованию глобальных свойств, оказывающих влияние на форму пространства в целом. В этом разделе мы ограничимся рассмотрением двумерных пространств, т. е. *поверхностей*.

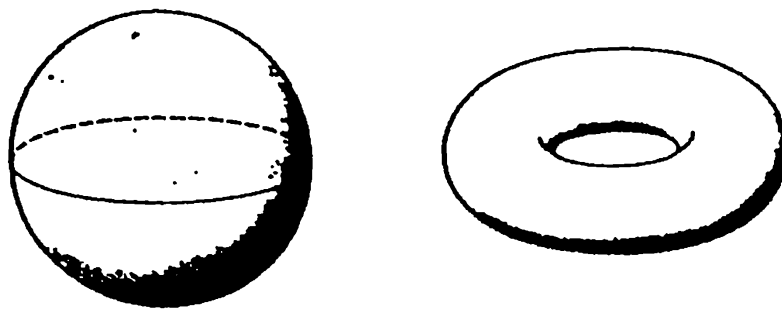


Рис. 1.5: Сфера и тор

Примеры *замкнутых поверхностей* показаны на рисунках 1.5 и 1.6: сфера и тор на рисунке 1.5 и замкнутые поверхности рода два и три на рисунке 1.6. Нестрого говоря, *род* поверхности — это количество отверстий в ней. Род тора равен 1, а род сферы равен 0. Аналогично можно рассмотреть поверхность рода g для любого натурального числа g .

Будем обозначать сферу как S^2 . Показатель 2 указывает на размерность сферы. Тор будем обозначать как T^2 . Также мы будем введём общее обозначение Σ_g для замкнутой поверхности рода g , в этих обозначениях Σ_0 и Σ_1 есть не что иное, как сфера S^2 и тор T^2 , соответственно.

Пусть M — поверхность. Назовём отображение

$$f : M \rightarrow \mathbb{R},$$

сопоставляющее каждой точке p поверхности M вещественное число, функцией на M .

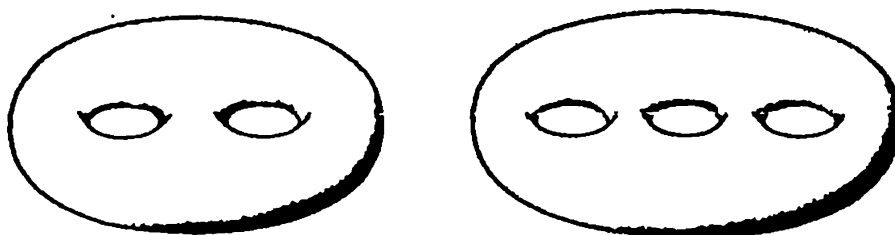


Рис. 1.6: Замкнутые поверхности рода 2 и 3

Заметим, что локальные координаты на рассматриваемых поверхностях будут криволинейными (см. рисунок 1.7).

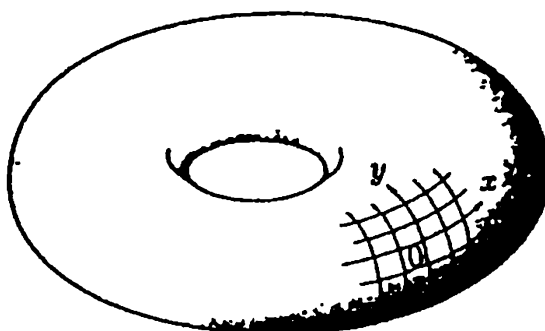


Рис. 1.7: Локальная система координат на торе

Функция $f : M \rightarrow \mathbb{R}$ является гладкой (принадлежит классу C^∞), если она принадлежит классу C^∞ в любых гладких локальных координатах любой точки поверхности M .

С этого момента мы будем рассматривать только гладкие (C^∞) поверхности и гладкие (C^∞) функции на этих поверхностях.

Понятие критической точки, рассмотренное в предыдущем разделе, естественным образом обобщается на функцию $f : M \rightarrow \mathbb{R}$, определённую на поверхности M , при помощи локальных координат. А именно, точка p_0 поверхности M называется *критической точкой* функции $f : M \rightarrow \mathbb{R}$, если

$$(1.39) \quad \frac{\partial f}{\partial x}(p_0) = 0, \quad \frac{\partial f}{\partial y}(p_0) = 0$$

относительно локальных координат в некоторой окрестности точки p_0 . Как мы видели в предыдущих разделах, невырожденные критические точки устойчивы и обладают некоторыми удобными свойствами, в отличие от вырожденных

критических точек. Поэтому нам будет в дальнейшем рассматривать функции, обладающие только невырожденными критическими точками.

ОПРЕДЕЛЕНИЕ 1.14. (Функция Морса). Пусть все критические точки функции $f : M \rightarrow \mathbb{R}$ невырождены. Тогда скажем, что f — функция Морса.

Рассмотрим несколько примеров таких функций.

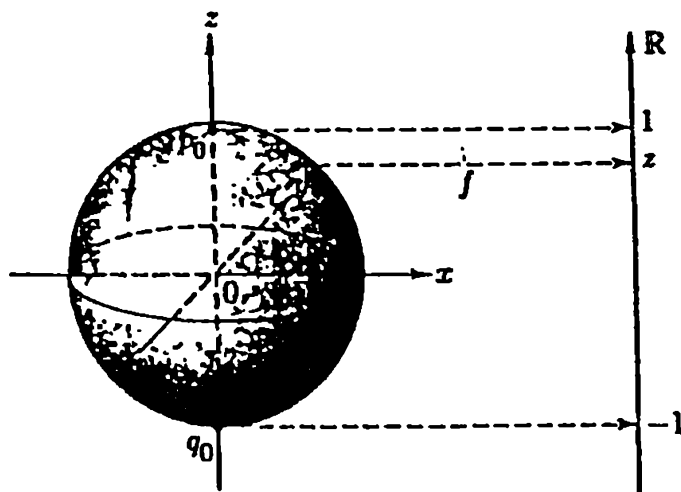


Рис. 1.8: Функция высоты на сфере

ПРИМЕР 1.15. (Функция высоты на сфере). Рассмотрим единичную сферу S^2 в декартовых координатах (x, y, z) евклидова пространства \mathbb{R}^3 ; так, S^2 определена уравнением

$$(1.40) \quad x^2 + y^2 + z^2 = 1.$$

Пусть $f : S^2 \rightarrow \mathbb{R}$ — функция на S^2 , сопоставляющая каждой точке $p = (x, y, z)$ сферы S^2 её координату z . Такую функцию можно назвать «функцией высоты». Тогда f является функцией Морса (рисунок 1.8).

Действительно, у f есть две критические точки: «северный полюс» $p_0 = (0, 0, 1)$ и «южный полюс» $q_0 = (0, 0, -1)$. Нетрудно заметить, что других критических точек на сфере нет. Чтобы проверить, что критические точки p_0 и q_0 невырожденные, достаточно вычислить гессиан функции f относительно локальных координат (x, y) — он будет ненулевым.

В данном примере мы рассмотрели функцию Морса на сфере с ровно двумя критическими точками, обе из которых невырождены. Оказывается, верно обратное.

ТЕОРЕМА 1.16. Пусть M — замкнутая поверхность, $f : M \rightarrow \mathbb{R}$ — функция Морса с ровно двумя невырожденными критическими точками (других критических точек нет). Тогда M диффеоморфна сфере S^2 .

Эта теорема — простой пример применения теории Морса. Перед тем, как доказывать теорему, определим понятие «диффеоморфизм».

Начнём с понятия «гомеоморфизма». Пусть дано взаимно однозначное отображение

$$h : X \rightarrow Y$$

между двумя (топологическими) пространствами X и Y . Мы можем определить обратное отображение

$$h^{-1} : Y \rightarrow X.$$

Если оба отображения $h : X \rightarrow Y$ и $h^{-1} : Y \rightarrow X$ непрерывны, то отображение h называется *гомеоморфизмом*. Два (топологических) пространства называются *гомеоморфными*, если существует гомеоморфизм между ними. В топологии пространства часто рассматриваются «с точностью до гомеоморфизма», т. е. гомеоморфные пространства считаются одинаковыми, «имеющими одну и ту же форму».

ОПРЕДЕЛЕНИЕ 1.17. Гомеоморфизм

$$h : M \rightarrow N$$

между поверхностями M и N называется *диффеоморфизмом*, если отображения $h : M \rightarrow N$ и $h^{-1} : N \rightarrow M$ принадлежат классу C^∞ . Поверхности M и N называются *диффеоморфными*, если между ними существует диффеоморфизм.

Диффеоморфные поверхности рассматриваются как имеющие одинаковую форму и гладкую структуру. В дифференциальной топологии, где объектами изучения являются гладкие геометрические фигуры, пространства рассматриваются «с точностью до диффеоморфизма».

Теперь докажем Теорему 1.16 (см. рисунок 1.9).

ДОКАЗАТЕЛЬСТВО. Заметим, что замкнутая поверхность является компактом (на самом деле, замкнутая поверхность определяется как «двумерное компактное многообразие без края»). Подробнее это понятие будет рассмотрено во 2-й главе. В ходе доказательства мы будем пользоваться теоремой Вейерштрасса о непрерывной функции на компакте:

ТЕОРЕМА 1.18. (Теорема Вейерштрасса о максимальном значении). Пусть $f : X \rightarrow \mathbb{R}$ — непрерывная функция на компакте X . Тогда f принимает максимальное значение в некоторой точке $p_0 \in X$ и минимальное значение в некоторой точке $q_0 \in X$.

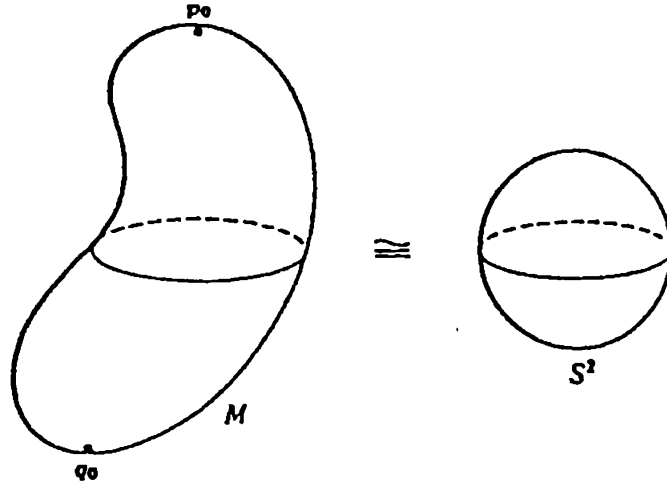


Рис. 1.9: Поверхность M диффеоморфна сфере S^2

Этот факт достаточно известен, поэтому здесь мы опустим его доказательство.

По теореме Вейерштрасса, функция Морса $f : M \rightarrow \mathbb{R}$ принимает максимальное значение в некоторой точке $p_0 \in M$ и минимальное значение в некоторой (другой) точке $q_0 \in M$. Точки p_0 и q_0 — критические точки функции f . Более того, поскольку f — функция Морса, p_0 и q_0 — невырожденные критические точки.

По лемме Морса (Теорема 1.11, функцию f можно представить в стандартной форме в подходящих локальных координатах (x, y) в окрестности точки p_0 и координатах (X, Y) в окрестности точки q_0 .

Поскольку p_0 — максимум, индекс p_0 равен 2; поскольку q_0 — минимум, индекс q_0 равен 0. Следовательно, имеем два представления в окрестности критических точек:

$$(1.41) \quad f = \begin{cases} -x^2 - y^2 + A & (\text{в окрестности } p_0), \\ X^2 + Y^2 + a & (\text{в окрестности } q_0), \end{cases}$$

где $A = f(p_0)$, $a = f(q_0)$.

Пусть ε — достаточно малое положительное число. Обозначим как $D(p_0)$ окрестность точки p_0 , для точек p из которой

$$(1.42) \quad A - \varepsilon \leq f(p) \leq A.$$

«Перевернутая чаша» параболоида слева на рисунке 1.10 соответствует множеству $D(p_0)$ которое в координатах (x, y) характеризуется уравнением

$$(1.43) \quad x^2 + y^2 \leq \varepsilon,$$



Рис. 1.10: Слева: график f в окрестности p_0 ; справа: график f в окрестности q_0 .

как видно из равенства (1.41). Другими словами, $D(p_0)$ диффеоморфно 2-диску (двумерному диску), определённому уравнением (1.43). Аналогично, множество $D(q_0)$ точек окрестности q_0 , удовлетворяющих неравенству

$$(1.44) \quad a \leq f(p) \leq a + \varepsilon,$$

соответствует чаше параболоида справа на рисунке 1.10. В координатах (X, Y) множество $D(q_0)$ принимает вид

$$(1.45) \quad X^2 + Y^2 \leq \varepsilon,$$

и также диффеоморфно 2-диску.

Удалим внутренность дисков $D(p_0)$ и $D(q_0)$ из поверхности M , оставшуюся поверхность с краем обозначим как M_0 . Граница M_0 состоит из граничных окружностей $C(p_0)$ и $C(q_0)$ дисков $D(p_0)$ и $D(q_0)$, соответственно.

В общем случае, граница поверхности с краем M_0 обозначается как

$$(1.46) \quad \partial M_0.$$

В нашем случае граница состоит из двух окружностей $C(p_0)$ и $C(q_0)$, поэтому мы можем записать:

$$(1.47) \quad \partial M_0 = C(p_0) \cup C(q_0).$$

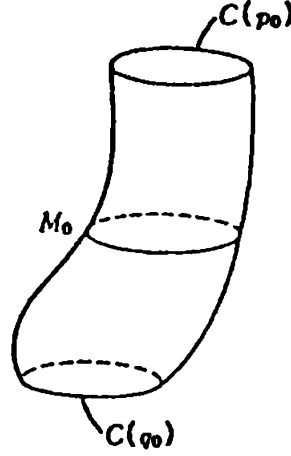
Будем обозначать как

$$(1.48) \quad \text{int}(M_0)$$

внутренность M_0 — то, что останется после удаления границы из M_0 . По определению, $\text{int}(M_0) = M_0 \setminus \partial M_0$.

Из определения $D(p_0)$ и $D(q_0)$ очевидно, что ограничение функции f на M_0 , $f : M_0 \rightarrow \mathbb{R}$, принимает постоянные значения $A - \varepsilon$ и $a + \varepsilon$ на граничных окружностях $C(p_0)$ и $C(q_0)$, соответственно.

Напомним, что по условию Теоремы 1.16 рассматриваемая функция Морса имеет ровно две критические точки. Поэтому после удаления (внутренности)

Рис. 1.11: Поверхность M_0 .

$D(p_0)$ и $D(q_0)$ у функции $f : M_0 \rightarrow \mathbb{R}$ критических точек не останется. Сформулируем факт, которым мы сейчас воспользуемся, а докажем позднее (TODO) в более общем виде.

ЛЕММА 1.19. Пусть $f : M_0 \rightarrow \mathbb{R}$ — функция класса C^∞ , которая является константой на граничных окружностях $C(p_0)$ и $C(q_0)$. Также пусть у f нет критических точек на M_0 . Тогда M_0 диффеоморфна прямому произведению окружности на единичный отрезок, т. е. $M_0 \cong C(q_0) \times [0, 1]$.

Поскольку $C(p_0)$ и $C(q_0)$ диффеоморфны единичной окружности (обозначаемой S^1), из Леммы 1.19 следует, что M_0 диффеоморфна прямому произведению

$$S^1 \times [0, 1].$$

Вообще, поверхность, диффеоморфная прямому произведению $S^1 \times [0, 1]$ называется *кольцом* (в топологическом, а не алгебраическом смысле). Так, если удалить из диска Δ внутренность меньшего концентрического диска Δ_0 , полученная поверхность A будет являться кольцом.

По Лемме 1.19, M_0 — кольцо. Пусть

$$N_0 = M_0 \cup D(q_0);$$

так, N_0 — это объединение M_0 с чашей гиперболоида. Точнее, N_0 получается приклейкой диска $D(q_0)$ к M_0 по граничной окружности $C(q_0)$, и поэтому поверхность N_0 тоже диффеоморфна диску.

Теперь приклеим диск $D(p_0)$ по граничной окружности к N_0 (у них общая граница) так, чтобы снова получить поверхность M . Таким образом, мы получили поверхность M склейкой двух дисков по границе, откуда следует, что M диффеоморфна сфере S^2 . \square

Строго говоря, нам нужна следующая лемма (Упражнение TODO), чтобы показать, что замкнутая поверхность, образованная склейкой двух дисков по границе диффеоморфна сфере S^2 .

ЛЕММА 1.20. Пусть отображение

$$(1.49) \quad k : \partial D_0 \rightarrow \partial D_1$$

является диффеоморфизмом между границами дисков D_0 и D_1 , соответственно. Тогда мы можем продолжить k до диффеоморфизма дисков

$$(1.50) \quad K : D_0 \rightarrow D_1.$$

Доказательство леммы смотрите в Упражнении (TODO). Подчеркнём, что утверждение леммы не является очевидным. На самом деле, оно выполняется для дисков размерности, не превышающей 6, а для больших размерностей, вообще, неверно.

В заключение раздела докажем следующий базовый факт.

ЛЕММА 1.21. Функция Морса $f : M \rightarrow \mathbb{R}$, определённая на замкнутой поверхности M , имеет конечное количество критических точек.

ДОКАЗАТЕЛЬСТВО. Предположим, что у функции Морса $f : M \rightarrow \mathbb{R}$ бесконечно много критических точек

$$p_1, p_2, p_3, \dots$$

Поскольку M — компакт, существует сходящаяся подпоследовательность p_{n_1}, p_{n_2}, \dots данной последовательности. Пусть p_0 — предел этой подпоследовательности. Не ограничивая общности, будем считать, что вся последовательность $\{p_{n_i}\}_{i=1}^\infty$ содержится внутри окрестности U точки p_0 .

Поскольку функция f гладкая, её производные $\frac{\partial f}{\partial x}(p)$ и $\frac{\partial f}{\partial y}(p)$ непрерывно зависят от p . Производные $\frac{\partial f}{\partial x}(p)$ и $\frac{\partial f}{\partial y}(p)$ принимают нулевое значение в точках последовательности p_{n_1}, p_{n_2}, \dots . Следовательно, в предельной точке p_0 производные также равны нулю, значит, p_0 — критическая точка функции f . Но в любой окрестности критической точки p_0 содержится бесконечное число других критических точек, что противоречит утверждению (Следствие 1.12) об изолированности критических точек. Это противоречие завершает доказательство Леммы 1.5. \square

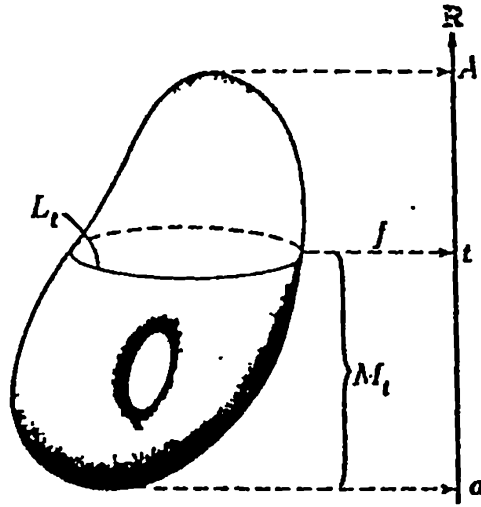


Рис. 1.12: Множество подуровня M_t и кривая уровня L_t .

1.5. Разложение на ручки

В Теореме 1.16 мы показали, что в одном частном случае функция Морса может определять форму поверхности, на которой она определена. Предмет теории Морса (в особенности, теории Морса для конечномерных пространств) заключается в изучении этого явления. В этом разделе мы опишем важный инструмент для изучения формы поверхности — «разложение на ручки».

Пусть

$$f : M \rightarrow \mathbb{R}$$

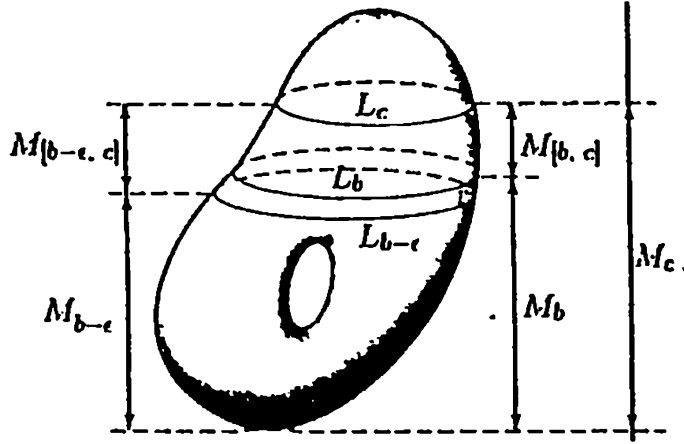
— функция морса, определённая на замкнутой поверхности.

Далее мы будем считать замкнутую поверхность M *связной*. Для поверхности M (а в общем случае, для многообразия M), это предположение означает то же, что и *линейная связность*, т. е. любые две точки p и q поверхности M соединены кривой, лежащей на M .

Для функции Морса $f : M \rightarrow \mathbb{R}$ мы будем обозначать как M_t *множество подуровня* M состоящее из точек, в которых функция f принимает значения, меньшие или равные вещественного числа t :

$$(1.51) \quad M_t = \{p \in M \mid f(p) \leq t\}.$$

Обозначим как L_t *кривую уровня* (или *множество уровня*) функции f , т. е. множество точек, в которых функция f принимает значение t .

Рис. 1.13: Между L_b и L_c нет критических точек.

Пусть f принимает на M максимальное значение A и минимальное значение a . Поскольку не существует точек p , для которых $f(p) < a$,

$$M_t = \emptyset$$

при $t < a$. Также заметим, что $f(p) \leq A$ в любой точке $p \in M$, поэтому $f(p) \leq t$ при $t \geq A$; значит, при $t \geq A$

$$M_t = M.$$

Следовательно, при изменении параметра t в интервале от a до A , множество подуровня M_t изменяется от пустого множества до всей поверхности M . Фундаментальная идея теории Морса состоит в описании изменения формы множества подуровня M_t в зависимости от t .

Можно представить ситуацию так, что $f : M \rightarrow \mathbb{R}$ — функция высоты, и мы погружаем M в воду. Параметр t соответствует уровню воды, а M_t — часть M , ушедшая под воду, при уровне воды t . При увеличении уровня t форма «подводной» части изменяется. Теория Морса изучает изменение формы множества M_t , расположенного «под водой».

ОПРЕДЕЛЕНИЕ 1.22. (Критические значения). Вещественное число c_0 называется критическим значением функции f , если f принимает значение c_0 в некоторой критической точке p_0 : $f(p_0) = c_0$.

ЛЕММА 1.23. Пусть $b < c$ — вещественные числа такие, что у f нет критических значений на отрезке $[b, c]$. Тогда множества M_b и M_c диффеоморфны (см. рисунок 1.13).

ДОКАЗАТЕЛЬСТВО. Поскольку максимальное и минимальное значения, A и a , соответственно, являются критическими значениями f , отрезок $[b, c]$ лежит внутри интервала (a, A) , т. е.

$$a < b < c < A.$$

Обозначим как $M_{[b,c]}$ часть поверхности M , расположенную между кривыми L_b и L_c :

$$(1.52) \quad M_{[b,c]} = \{p \in \mathbb{M} \mid b \leq f(p) \leq c\}.$$

Очевидно,

$$(1.53) \quad M_b \cup M_{[b,c]} = M_c.$$

По условию леммы, $M_{[b,c]}$ не содержит критических точек f .

По Лемме , $f : M \rightarrow \mathbb{R}$ содержит конечное число критических точек. Следовательно, мы можем предположить, что у f нет критических точек на $M_{[b-\varepsilon,c]}$ (образно говоря, мы приклеили короткую «юбку» снизу к $M_{[b,c]}$) для достаточно малого $\varepsilon > 0$.

Согласно Теореме (TODO), которую мы докажем в следующей главе, множество $M_{[b-\varepsilon,c]}$ диффеоморфно произведению $L_{b-\varepsilon} \times [0, 1]$. Также заметим, что $M_{[b-\varepsilon,b]} \subset M_{[b-\varepsilon,c]}$ и поскольку у f нет критических точек на $M_{[b-\varepsilon,b]}$, по той же теореме, множество $M_{[b-\varepsilon,b]}$ также диффеоморфно $L_{b-\varepsilon} \times [0, 1]$. Тогда существует диффеоморфизм

$$h : M_{[b-\varepsilon,b]} \rightarrow M_{[b-\varepsilon,c]},$$

причём мы можем считать ограничение h на кривую $L_{b-\varepsilon}$ тождественным отображением. Теперь «склеим»² h и тождественное отображение

$$\text{id} : M_{b-\varepsilon} \rightarrow M_{b-\varepsilon}$$

по граничной кривой уровня $L_{b-\varepsilon}$. В результате получим диффеоморфизм

$$(1.54) \quad H = \text{id} \cup h : M_{b-\varepsilon} \cup M_{[b-\varepsilon,b]} \rightarrow M_{b-\varepsilon} \cup M_{[b-\varepsilon,c]}$$

Заметив, что $M_{b-\varepsilon} \cup M_{[b-\varepsilon,b]} = M_b$ и $M_{b-\varepsilon} \cup M_{[b-\varepsilon,c]} = M_c$, мы получаем диффеоморфизм

$$H : M_b \rightarrow M_c.$$

Таким образом, Лемма 1.23 доказана.

²Процесс склейки диффеоморфизмов имеет некоторые тонкости, которые указаны в Теореме TODO и разъяснениях к ней в следующей главе.

Основная идея доказательства состоит в следующем: если растянуть тонкий «манжет» $M_{[b-\varepsilon, b]}$ поверхности подуровня M_b , можно получить поверхность M_c . \square

Согласно Лемме 1.23, форма M_t остаётся неизменной до тех пор, пока параметр t не проходит через критическое значение функции f . Следовательно, изменение формы M_t происходит при прохождении параметром t критического значения, и это изменение нуждается в исследовании.

Если c_0 — критическое значение f , то существует хотя бы одна критическая точка p_0 такая, что

$$f(p_0) = c_0.$$

Для простоты предположим, что существует *ровно* одна критическая точка p_0 с критическим значением c_0 . Также предположим, что существует достаточно малое $\varepsilon > 0$, что на участке поверхности $M_{[c_0-\varepsilon, c_0+\varepsilon]}$ нет других критических точек.

Посмотрим, как происходит переход от $M_{c_0-\varepsilon}$ к $M_{c_0+\varepsilon}$ при прохождении критического значения c_0 .

(а) Первый случай: индекс p_0 равен 0

Запишем функцию f в локальных координатах (x, y) в стандартной форме (см. Теорему 1.11):

$$(1.55) \quad f = x^2 + y^2 + c_0.$$

Если c_0 — глобальный минимум функции f , то $M_{c_0-\varepsilon} = \emptyset$. Поверхность $M_{c_0+\varepsilon}$, определённая как

$$(1.56) \quad M_{c_0+\varepsilon} = \{p \in M \mid f(p) \leq c_0 + \varepsilon\} = \{(x, y) \mid x^2 + y^2 \leq \varepsilon\},$$

диффеоморфна диску D^2 . Изменение поверхности M_t можно описать так: M_t — пустое множество при $t \leq c_0$, а при прохождении точки $t = c_0$ поверхность M_t становится диффеоморфной диску. Если же c_0 не является глобальным минимумом, то множество $M_{c_0-\varepsilon}$ непусто. В этом случае при прохождении t значения c_0 появляется новый диск, и $M_{c_0+\varepsilon}$ становится диффеоморфной несвязному объединению $M_{c_0-\varepsilon}$ и 2-диска D^2 (см. рисунок 1.14):

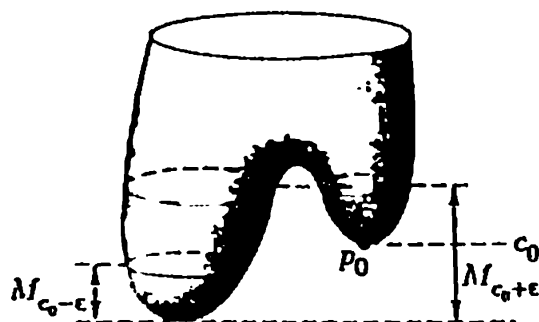
$$(1.57) \quad M_{c_0+\varepsilon} \cong M_{c_0-\varepsilon} \sqcup D^2,$$

где символ \sqcup означает *несвязное объединение* множеств, т. е. объединение непересекающихся множеств.

(б) Второй случай: индекс p_0 равен 1

Запишем функцию f в локальных координатах (x, y) в стандартной форме (см. Теорему 1.11):

$$(1.58) \quad f = -x^2 + y^2 + c_0,$$

Рис. 1.14: Случай (а): индекс p_0 равен нулю.

где мы поменяли местами x и y для соответствия обозначениям в общем случае, который будет рассмотрен в следующей главе. График функции f в окрестности p_0 напоминает горный перевал (см. рисунок 1.15). Кривая, соответствующая $y = 0$, уходит вниз от p_0 , а ортогональная ей кривая $x = 0$, наоборот, восходит из p_0 вверх. Таким образом, утверждение, что критическая точка имеет индекс 1 эквивалентно тому, что нисходящее направление из критической точки одномерно.

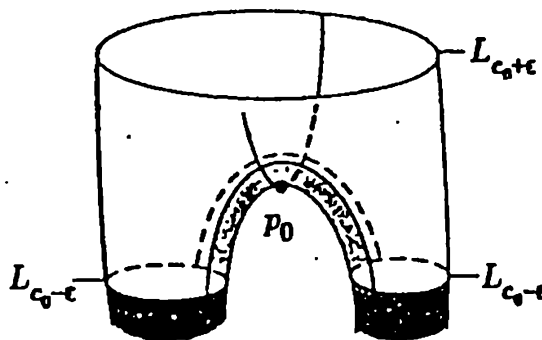


Рис. 1.15: График функции в окрестности критической точки индекса 1.

Рассмотрим нисходящую кривую; добавив к ней точки, лежащие достаточно близко, получим узкую полосу, «путь», проходящий по «горному перевалу» через p_0 . Можно представлять эту полосу, как узкий мост, соединяющий разные участки границы $L_{c_0-\epsilon}$ поверхности $M_{c_0-\epsilon}$. На рисунке 1.16 показан вид сверху графика с рисунка 1.15.

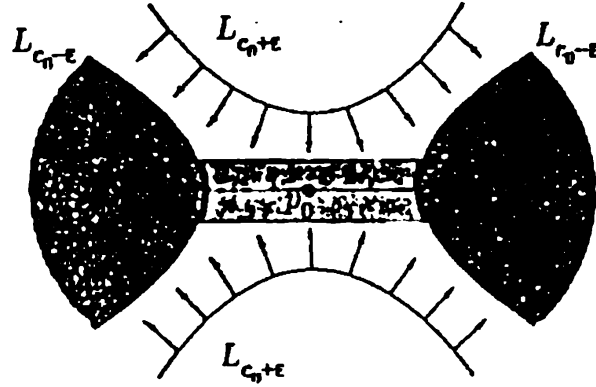


Рис. 1.16: Вид сверху графика с рисунка 1.15.

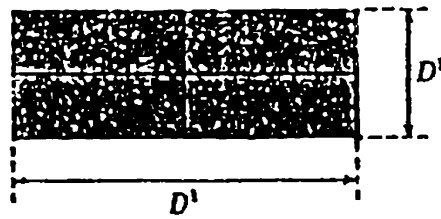
Закрашенная область на рисунке 1.16 соответствует $M_{p_0-\epsilon}$ и задаётся уравнением

$$-x^2 + y^2 \leq -\epsilon,$$

а затенённая область на рисунке 1.16 соответствует мосту, соединяющему $M_{c_0-\epsilon}$. Мост «почти прямоугольный» (диффеоморфен прямоугольнику), хоть и немного искривлён.

Отрезок $[0, 1]$ мы будем называть одномерным диском, или просто 1-диском, и будем обозначать как D^1 . Концы отрезка D^1 являются его границей и обозначаются как ∂D^1 . Таким образом, ∂D^1 состоит из двух точек.

Серый прямоугольный «мостик» на рисунке 1.16 диффеоморфен произведению $D^1 \times D^1$, а его пересечение с $M_{c_0-\epsilon}$ (показано жирными линиями на рисунке 1.16) соответствует паре противоположных рёбер $\partial D^1 \times D^1$ (см. рисунок 1.17).

Рис. 1.17: 1-ручка $D^1 \times D^1$ и пара отрезков, которыми она приклеена к $M_{c_0-\epsilon}$.

Такой прямоугольник $D^1 \times D^1$ называется *1-ручкой*, приклеенной к $M_{c_0-\epsilon}$. Термин «1-ручка» иллюстрирует тот факт, что она соответствует критической

точке индекса 1. Понимание этого термина лучше раскроется в дальнейшем, когда мы будем рассматривать трёхмерные ручки в третьей главе.

Поверхность $M_{c_0+\varepsilon}$ после того, как параметр t прошёл критическое значение c_0 , в локальных координатах описывается неравенством

$$-x^2 + y^2 + c_0 \leq \varepsilon.$$

Сравним поверхности $M_{c_0+\varepsilon}$ и $M_{c_0-\varepsilon}$ с приклеенной 1-ручкой $D^1 \times D^1$ (рис. 1.16). Заметим, что мы можем стянуть $M_{c_0+\varepsilon}$ вдоль стрелок в поверхность $M_{c_0-\varepsilon} \cup D^1 \times D^1$. Следовательно,

$$(1.59) \quad M_{c_0+\varepsilon} \cong M_{c_0-\varepsilon} \cup D^1 \times D^1.$$

Таким образом, для критической точки p_0 индекса 1, изменение M_t при прохождении параметра t от $c_0 - \varepsilon$ до $c_0 + \varepsilon$ может быть описано как приклеивание 1-ручки к $M_{c_0-\varepsilon}$.

Строго говоря, поверхность $M_{c_0-\varepsilon} \cup D^1 \times D^1$ в правой части (1.59) в изображённом виде не является гладкой, поскольку граница $M_{c_0-\varepsilon} \cup D^1 \times D^1$ изгибается под прямым углом в четырёх углах, по которым приклеена 1-ручка $D^1 \times D^1$. Однако, это несложно исправить, используя известную технику «сглаживания» (см., например, TODO). Поэтому мы можем считать, что в диффеоморфизме (1.59) справа находится гладкая поверхность. В некоторых случаях мы будем использовать символ \cong в (1.59) для обозначения гомеоморфизма, и в тех случаях, когда там нужно будет воспользоваться гладкостью поверхности, мы просто будем рассматривать поверхность $M_{c_0+\varepsilon}$. В любом случае, в дальнейших рассуждениях мы будем считать, что поверхность $M_{c_0-\varepsilon} \cup D^1 \times D^1$ гладкая.

(в) Третий случай: индекс p_0 равен 2

Запишем функцию f в локальных координатах (x, y) в стандартной форме (см. Теорему 1.11):

$$(1.60) \quad f = -x^2 - y^2 + c_0.$$

В локальных координатах (x, y) мы можем записать

$$(1.61) \quad M_{c_0-\varepsilon} = \{p \in M \mid f(p) \leq c_0 - \varepsilon\} = \{(x, y) \mid x^2 + y^2 \geq \varepsilon\},$$

откуда $M_{c_0-\varepsilon}$ является дополнением к диску радиуса $\sqrt{\varepsilon}$ (со включённой границей). На рисунке 1.18 этот диск изображён как перевёрнутая чаша параболоида. Поверхность $M_{c_0+\varepsilon}$ получается прилейкой этой чаши к $M_{c_0-\varepsilon}$ по её границе:

$$(1.62) \quad M_{c_0+\varepsilon} \cong M_{c_0-\varepsilon} \cup D^2.$$

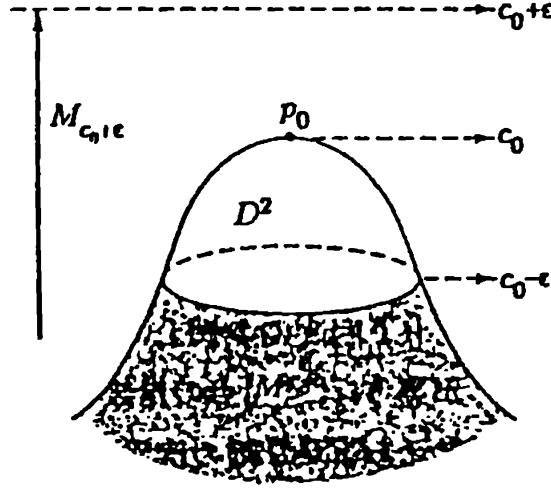


Рис. 1.18: Критическая точка индекса 2.

«Шапочка» D^2 называется *2-ручкой*. Соотношение (1.59) означает, что при прохождении параметра t от $c_0 - \epsilon$ к $c_0 + \epsilon$, поверхность $M_{c_0+\epsilon}$ получается из $M_{c_0-\epsilon}$ приклеиванием 2-ручки D^2 . Заметим, что при приклеивании «шапочки» пропадает одна из граничных окружностей поверхности $M_{c_0-\epsilon}$, таким образом, количество компонент связности границы уменьшается на 1.

(г) Разложение на ручки

Вернёмся к функции Морса $f : M \rightarrow \mathbb{R}$.

По Лемме 1.5, количество критических точек у функции f конечно. Обозначим эти точки как

$$p_1, p_2, \dots, p_n.$$

В следующей главе мы покажем, что можем «пошевелить» функцию Морса f так, чтобы

$$f(p_i) \neq f(p_j) \text{ при } i \neq j.$$

Поэтому будем считать, что f обладает этим свойством. Затем перенумеруем критические точки так, чтобы

$$(1.63) \quad f(p_1) < f(p_2) < \dots < f(p_n).$$

Обозначим $c_i = f(p_i)$. Тогда

$$(1.64) \quad c_1 < c_2 < \dots < c_n.$$

Очевидно, c_1 — минимальное значение f , а c_n — максимальное значение.

Будем наблюдать, как изменяется поверхность M_t при возрастании t , начиная с некоторого значения, меньшего c_1 . При $t < c_1$, $M_t = \emptyset$. При прохождении значения c_1 появляется диск (чаша параболоида), и $M_t = D^2$. Так произойдёт, поскольку индекс p_0 равен нулю, т. к. c_1 — минимум f . Этот 2-диск соответствует критической точке индекса 0 и называется *0-ручкой*. 2-ручка тоже является 2-диском, но 0-ручка соответствует чаше параболоида, а 2-ручка — перевёрнутой чаше, или «шапочке». (Возможна другая интерпретация индекса критической точки — это количество нисходящих направлений. Поскольку 2-ручке-шапочке соответствует два направления спуска, то индекс критической точки равен 2. Для 1-ручки единственным нисходящим направлением будет центральная линия моста, см. Рисунок 1.16, поэтому индекс соответствующей критической точки равен 1. 0-ручка — это чаша параболоида, у которой нет нисходящих направлений, поэтому индекс критической точки равен 0.)

Продолжая описанный процесс, каждый раз, когда t проходит критическое значение c_i , мы приклеиваем 0-, 1- или 2-ручку к $M_{c_i-\varepsilon}$ в зависимости от индекса соответствующей критической точки p_i («приклеивание» 0-ручки означает добавление ручки к $M_{c_i-\varepsilon}$ в качестве новой связной компоненты).

Последнее значение c_n является максимумом f , поэтому оно соответствует 2-ручке. Границей поверхности $M_{c_0-\varepsilon}$ является окружность, по которой приклеивается эта 2-ручка, завершающая замкнутую поверхность M . Таким образом, мы показали справедливость следующей теоремы.

ТЕОРЕМА 1.24. (Разложение на ручки замкнутой поверхности). Если на замкнутой поверхности M задана функция Морса $f : M \rightarrow \mathbb{R}$, то M может быть представлена как объединение конечного числа 0-, 1- и 2-ручек.

В следующей главе мы докажем существование функции Морса на любой замкнутой поверхности M , из чего следует, что любая замкнутая поверхность может быть разложена на ручки.

Краткое содержание главы

- 1.1 Критические точки функции бывают вырожденными и невырожденными.
- 1.2 Функция, у которой все критические точки невырождены, называется функцией Морса.
- 1.3 Если функция Морса f определена на замкнутой поверхности M , то мы можем исследовать M , рассматривая множества подуровня функции f .
- 1.4 Каждый раз, когда значение f проходит через критическое значение, к поверхности подуровня приклеивается ручка. Индекс ручки совпадает с индексом соответствующей критической точки.

-
- 1.5 Таким образом, замкнутая поверхность M может быть разложена в объединение конечного числа ручек.