

## ИССЛЕДОВАНИЕ НА ЭКСТРЕМУМ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ

В начале рассмотрим случай поиска локального минимума функции *двух* переменных, поскольку уже для  $n = 2$  все существенные отличия от одномерной задачи можно легко продемонстрировать.

Пусть дана дважды непрерывно дифференцируемая в некоторой области  $\Omega \subseteq E^2$  с ОНБ функция  $f(x, y)$ . Выясним, при каких условиях эта функция имеет в точке  $\begin{pmatrix} x^* \\ y^* \end{pmatrix} \in \Omega$  локальный минимум. Дадим

Определение 1. Будем говорить, что функция  $f(x, y)$  имеет в точке  $\begin{pmatrix} x^* \\ y^* \end{pmatrix}$  *строгий локальный минимум*, если существует  $U_\varepsilon^\circ$  - проколота окружность, такая, что для *любой* точки  $\begin{pmatrix} x \\ y \end{pmatrix} \in U_\varepsilon^\circ$  выполнено неравенство

$$f(x, y) > f(x^*, y^*). \quad (1)$$

Понятно, что проверить выполнение условия определения 1 для *всех* точек окрестности  $U_\varepsilon^\circ$  невозможно. Поэтому необходимо получить условия существования экстремума, проверка которых практически реализуема.

В некоторых случаях, удастся преобразовать запись функции  $f(x, y)$  к виду, в котором выполнение неравенства (1) очевидно или легко проверяется. Например, функцию  $f(x, y) = x^4 - x^2y + y^2$  можно записать так:

$$f(x, y) = x^4 - x^2y + \frac{y^2}{4} + \frac{3y^2}{4} = \left(x^2 - \frac{y}{2}\right)^2 + \left(\frac{y\sqrt{3}}{2}\right)^2.$$

Откуда следует, что точка  $\begin{pmatrix} x^* \\ y^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$  есть точка локального минимума. Впрочем, подобная ситуация есть исключение, а не правило.

Более удобным способом использования определения 1 является оценка знака разности  $f(x, y) - f(x^*, y^*)$  для точек окрестности  $U_\varepsilon$  при помощи формулы Тейлора.

Формулу Тейлора для функции  $f(x, y)$  в точке  $\begin{pmatrix} x^* \\ y^* \end{pmatrix}$ , как известно, можно записать в виде

$$f(x, y) = f(x^*, y^*) + df + \frac{1}{2}d^2f + o(\rho^2), \quad (2)$$

где  $\rho = \sqrt{dx^2 + dy^2}$ ,  $dx = x - x^*$  и  $dy = y - y^*$ , а дифференциалы  $df$  и  $d^2f$  соответственно равны

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy, \quad d^2f = \frac{\partial^2 f}{\partial x^2} dx^2 + 2 \frac{\partial^2 f}{\partial x \partial y} dx dy + \frac{\partial^2 f}{\partial y^2} dy^2.$$

В двух последних формулах частные производные вычислены в точке  $\begin{pmatrix} x^* \\ y^* \end{pmatrix}$ .

Из (2) получаем, что

$$f(x, y) - f(x^*, y^*) = df + \frac{1}{2}d^2f + o(\rho^2), \quad (3)$$

Здесь отметим, что, согласно теореме Тейлора (об остаточном члене в форме Пеано) первое слагаемое в правой части (3) в силу дважды непрерывной дифференцируемости функции  $f(x, y)$ , имеет порядок малости  $\rho = \sqrt{dx^2 + dy^2}$ . Второе же слагаемое в правой части имеет порядок малости  $\rho^2 = dx^2 + dy^2$ .

Это означает, что в достаточно малой окрестности точки  $\left\| \begin{matrix} x^* \\ y^* \end{matrix} \right\|$  знак приращения  $f(x, y) - f(x^*, y^*)$  определяется знаком величины  $df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy$ , который может быть *любым* в силу линейной зависимости  $df$  от  $dx$  и  $dy$ .

Иначе говоря, если  $\text{grad } f(x, y) \neq 0$  и для некоторого вектора  $\left\| \begin{matrix} dx \\ dy \end{matrix} \right\|$  мы имеем  $f(x, y) - f(x^*, y^*) > 0$ , то для вектора  $\left( -\left\| \begin{matrix} dx \\ dy \end{matrix} \right\| \right)$  мы обязательно будем иметь, что  $f(x, y) - f(x^*, y^*) < 0$ , так как производные  $\frac{\partial f}{\partial x}$  и  $\frac{\partial f}{\partial y}$  от значений  $dx$  и  $dy$  не зависят.

Значит, у непрерывно дифференцируемой функции  $f(x, y)$  в точках, где  $\text{grad } f(x, y) \neq 0$ , экстремума (т.е. минимума или максимума) быть не может. Заметим, что точки, в которых  $\text{grad } f(x, y) = 0$ , принято называть *стационарными точками* для функции  $f(x, y)$ .

В итоге мы приходим к следующему *необходимому* условию существования экстремума:

Если функция  $f(x, y)$  имеет экстремум в точке  $\left\| \begin{matrix} x^* \\ y^* \end{matrix} \right\|$ , то

**либо**  $\text{grad } f(x^*, y^*) = 0$ ,

**либо** *градиент в этой точке не существует.*

*Это необходимое условие не является достаточным.* Пример:  $f(x, y) = xy$ . Здесь в начале координат градиент есть нулевой вектор, а экстремума нет.

Пусть теперь мы рассматриваем только точки, в которых  $\text{grad } f(x, y) = 0$ . В этом случае знак разности  $f(x, y) - f(x^*, y^*)$  будет совпадать со знаком второго слагаемого в правой части (3), т.е.

$$d^2 f = \frac{\partial^2 f}{\partial x^2} dx^2 + 2 \frac{\partial^2 f}{\partial x \partial y} dx dy + \frac{\partial^2 f}{\partial y^2} dy^2. \quad (4)$$

В формуле (4) значения производных вычислены в фиксированной точке  $\begin{pmatrix} x^* \\ y^* \end{pmatrix}$  и не зависят от значений  $dx$  и  $dy$ . Для простоты записей будем обозначать эти значения так:  $A = \frac{\partial^2 f}{\partial x^2}$ ,  $B = \frac{\partial^2 f}{\partial x \partial y}$ ,  $C = \frac{\partial^2 f}{\partial y^2}$ . Напомним, что матрица  $\begin{pmatrix} A & B \\ B & C \end{pmatrix}$  называется *матрицей Гессе*.

Тогда можно утверждать, что знак разности  $f(x, y) - f(x^*, y^*)$  совпадает со знаком квадратичной формы

$$A dx^2 + 2B dx dy + C dy^2. \quad (5)$$

Если квадратичная форма  $A dx^2 + 2B dx dy + C dy^2$  *положительно определенная*, то в точке  $\begin{pmatrix} x^* \\ y^* \end{pmatrix}$  функция  $f(x, y)$  будет иметь *строгий локальный минимум*, а, если эта форма *отрицательно определена*, то – *строгий локальный максимум*. Наконец, если форма не имеет знаковой определенности (как строгой, так и нестрогой), то экстремума гарантировано нет. Оставшиеся возможные случаи будут требовать дополнительного исследования.

Итак, мы пришли к достаточным условиям вида:

**Если у дважды непрерывно дифференцируемой функции  $f(x, y)$  в стационарной точке квадратичная форма (5) положительно определена, то  $f(x, y)$  имеет в этой точке *строгий локальный минимум*.**

**Если у дважды непрерывно дифференцируемой функции  $f(x, y)$  в стационарной точке квадратичная форма (5) отрицательно определена, то  $f(x, y)$  имеет в этой точке *строгий локальный максимум*.**

**Если у дважды непрерывно дифференцируемой функции  $f(x, y)$  в стационарной точке квадратичная форма (5) не имеет знаковой определенности, то  $f(x, y)$  не имеет в этой точке *строгого локального экстремума*.**

Заметим, что, например, *первое достаточное условие не является необходимым*.  
Пример:  $f(x, y) = x^4 + y^4$ . Здесь в начале координат есть *строгий минимум*, а *строгой положительной определенности нет*.

Напомним теперь, доказываемые в курсе линейной алгебры, методы исследования квадратичной формы (5) на наличие или отсутствия знаковой неопределенности.

1) *Метод Лагранжа*. Он сводится к построению диагонального (или канонического) базиса методом выделения полных квадратов, т.е. базиса, в котором коэффициент  $B$  у квадратичной формы (5) равен нулю.

2) *Критерий Сильвестра*. Этот критерий утверждает, что для положительной определенности формы (5) необходимо и достаточно, чтобы одновременно выполнялись неравенства  $A > 0$  и  $\det \begin{vmatrix} A & B \\ B & C \end{vmatrix} > 0$ .

Для отрицательной определенности формы (5) необходимо и достаточно, чтобы одновременно выполнялись неравенства  $A < 0$  и  $\det \begin{vmatrix} A & B \\ B & C \end{vmatrix} > 0$ .

3) *Диагонализация матрицы квадратичной формы в евклидовом пространстве*. Этот метод основан на теореме о том, что в базисе из собственных векторов самосопряженного преобразования, имеющего в ОНБ матрицу вида  $\begin{vmatrix} A & B \\ B & C \end{vmatrix}$ , матрица такого преобразования диагональная, причем на ее главной диагонали стоят собственные значения этого преобразования.



Пример 1. В  $E^2$  исследовать на экстремум функцию  $f(x, y) = x^3 + y^3 - 3xy$ .

Решение: 1) Найдем вначале для бесконечно дифференцируемой функции  $f(x, y)$  все ее стационарные точки, т.е. точки, подозрительные на экстремум. Эти точки находятся из *необходимого* условия

$$\text{grad } f(x, y) = 0 \quad \text{или} \quad \begin{cases} \frac{\partial f}{\partial x} = 3x^2 - 3y = 0, \\ \frac{\partial f}{\partial y} = 3y^2 - 3x = 0 \end{cases} \Rightarrow \begin{bmatrix} \parallel 1 \parallel \\ \parallel 1 \parallel \\ \parallel 0 \parallel \\ \parallel 0 \parallel \end{bmatrix}.$$

2) Проверим теперь выполнение *достаточных* условий в стационарных

точках. Строим матрицу Гессе 
$$\begin{vmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{vmatrix} = \begin{vmatrix} 6x & -3 \\ -3 & 6y \end{vmatrix}.$$

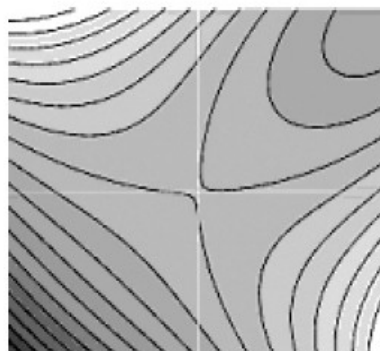
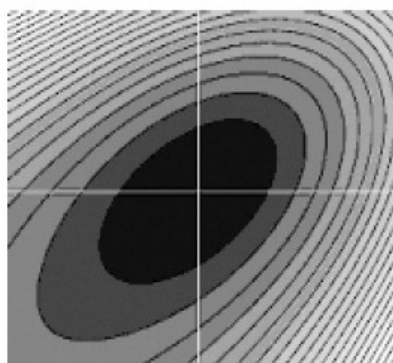
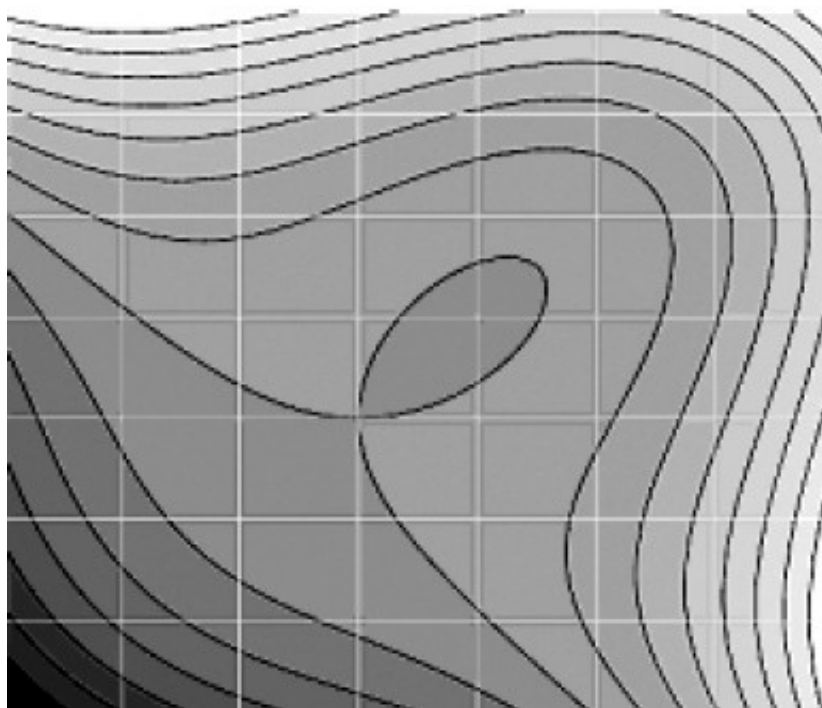
В первой стационарной точке матрица Гессе будет  $\begin{vmatrix} 6 & -3 \\ -3 & 6 \end{vmatrix}$ . Для нее выполняется критерий Сильвестра положительной определенности квадратичной формы (5). Значит в точке  $\begin{vmatrix} 1 \\ 1 \end{vmatrix}$  у функции строгий локальный минимум.

Во второй стационарной точке матрица Гессе равна  $\begin{vmatrix} 0 & -3 \\ -3 & 0 \end{vmatrix}$ . Для нее выполняется достаточное условие отсутствия знаковой определенности квадратичной формы (5).

В точке  $\begin{vmatrix} 0 \\ 0 \end{vmatrix}$  у функции экстремума нет, поскольку в малой окрестности начала координат  $f(x, y) = -3xdy + dx^3 + dy^3$

$$\text{и при } dx = dy \text{ имеем } \Delta f = -3dx^2 + 2dx^3 < 0,$$

$$\text{а при } dx = -dy \text{ имеем } \Delta f = 3dx^2 - 2dx^3 > 0.$$

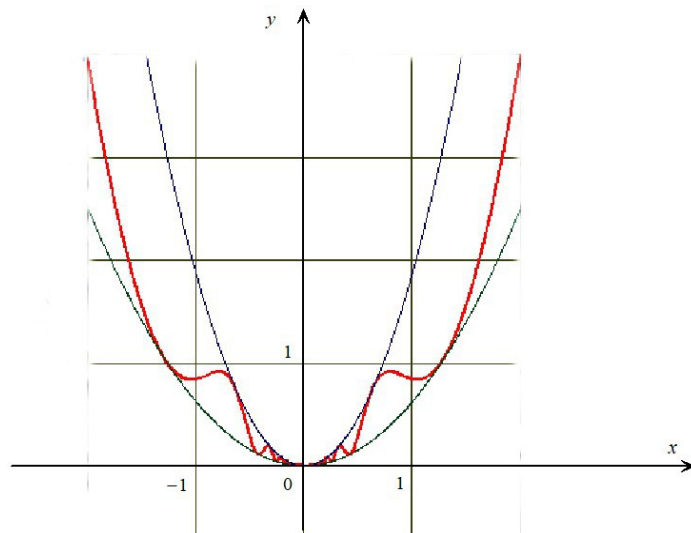


Пример 2. Пусть непрерывная функция  $F(x)$  имеет в точке  $x^*$  локальный экстремум. Верно ли утверждение: в этом случае найдется такая окрестность точки  $x^*$ , для которой  $F(x)$  монотонна как в правой, так и в левой полуокрестности этой точки? Ответ обосновать.

Решение: Нет, неверно. Пример: непрерывная функция

$$F(x) = \begin{cases} 0, & \text{если } x = 0, \\ x^2 \left( 2 + \cos \frac{1}{x} \right), & \text{если } x \neq 0 \end{cases}$$

имеет в точке  $x^* = 0$  минимум, но не является монотонной в любой полуокрестности этой точки.



### Общая схема поиска локального экстремума

Вполне очевидно, что метод поиска в  $E^n$  локальных экстремумов функции  $F(\xi_1, \xi_2, \dots, \xi_n)$ , основанный на исследовании ее критических элементов (или же на решении уравнения  $\text{grad } F(x) = 0$ ), практически пригоден лишь для крайне ограниченного числа случаев.

Реально реализуемой альтернативой является использование так называемых *итеративных численных методов*.

Эти методы могут быть представлены в виде следующей схемы поиска, например, максимума.

- 1°. Для некоторого начального элемента  $x^0$  находятся  
ненулевое *улучшающее направление максимизации*  $w^0 \in E^n$  и  
положительное число  $\sigma_0 < +\infty$  – *величина шага* по данному направлению  
– такие, что на элементе  $x^1 = x^0 + \sigma_0 \cdot w^0$   
верно неравенство  $F(x^1) > F(x^0)$ .
- 2°. Если задача в пункте 1° решена успешно, то элемент  $x^0$  заменяется на  $x^1$  и  
процедура пункта 1° повторяется для некоторого нового  $w^1 \in E^n$ .

Если же оказалось, что множество улучшающих направлений состоит только из  
нулевого элемента или же  $\sigma_0 = 0$ , то

либо элемент  $x^0$  принимается за  $x^*$  – максимальный,  
либо, при  $\sigma_0 = +\infty$ , констатируется факт отсутствия  
максимума у функции  $F(x)$ .

Таким образом, поиск экстремального элемента сводится к итерационной процедуре  
вида

$$x^{k+1} = x^k + \sigma_k \cdot w^k \quad k = 0, 1, 2, \dots,$$

*условия сходимости* которой к искомому экстремальному элементу  $x^*$  (то есть,  
 $\rho(x^k, x^*) \rightarrow 0$  при  $k \rightarrow \infty$ ) зависят, как от способа выбора  $w^k$  и  $\sigma_k$ , так и от свойств  
максимизируемой функции.

**Примеры методов поиска локального гладкого экстремума**

Предположим, что функция  $F(x)$  непрерывно дифференцируема в  $E^n$  и выполнены следующие условия:

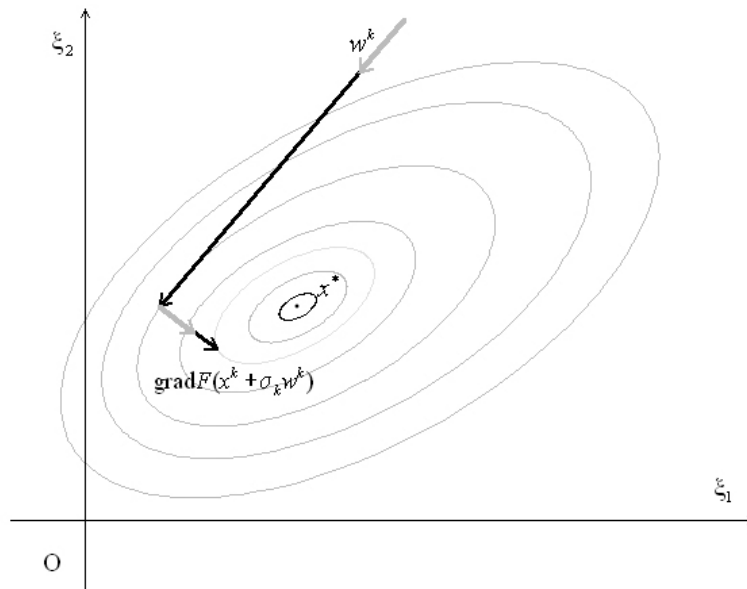
- 1°. Начальный элемент  $x^0$  принадлежит окрестности максимального элемента  $x^*$ , в которой функция  $F(x)$  *строго выпукла вверх*;
- 2°. Элемент *направления максимизации*  $w^k$  удовлетворяет ограничению  $(\text{grad } F(x^k), w^k) > 0$ .
- 3°. Величина шага по выбранному направлению максимизации  $\sigma_k$  находится путем решения одномерной задачи

$$\sigma_k = \max_{\sigma \geq 0} F(x^k + \sigma \cdot w^k).$$

Тогда последовательность элементов  $\{x^k\}$  будет сходиться к  $x^*$ , если величина шага  $\sigma_k$  по направлению  $w^k$  находится из уравнения

$$(\text{grad } F(x^k + \sigma_k w^k), w^k) = 0.$$

Это правило выбора величины шага по направлению  $w^k$  гарантирует наибольшее увеличение значения функции по направлению  $w^k$ , что иллюстрирует следующий рисунок.





В случаях, когда оптимизируемая функция достаточно гладкая (например, имеет непрерывные частные производные первого и второго порядка), для выбора  $w_k$  можно использовать более эффективные схемы, такие как:

- 1) *метод наискорейшего подъема (линейной аппроксимации)*, основанный на оценке скорости возрастания функции  $F(x)$  на элементе  $x^k$  по направлению  $w^k$ , которая в силу неравенства Коши–Буняковского максимальна при  $w^k = \text{grad } F(x^k)$ , поскольку  $\forall w$  с  $|w| = 1$

$$\left| \frac{\partial F}{\partial w} \right| = |(\text{grad } F, w)| \leq |\text{grad } F| |w| = |\text{grad } F|.$$

Этот метод, однако, оказывается малоэффективным в случаях "плохой обусловленности" целевой функции.

- 2) метод квадратичной аппроксимации (метод Ньютона), при котором улучшающее направление  $w_k$  удовлетворяет условию

$$\widehat{\text{Hess}} F(x_k) w^k = -\text{grad} F(x_k),$$

что в матричном и координатном представлениях есть система линейных уравнений

$$\left\| \widehat{\text{Hess}} F(x_k) \right\| w^k = -\text{grad} F(x_k)$$

или 
$$\sum_{j=1}^n \frac{\partial^2 F}{\partial \xi_i \partial \xi_j} \omega_j^k = -\frac{\partial F}{\partial \xi_i}, i = [1, n],$$

где  $\|w^k\| = \begin{pmatrix} \omega_1^k \\ \omega_2^k \\ \dots \\ \omega_n^k \end{pmatrix}$  – координатное представление элемента  $w^k$ .

Убедимся, что для квадратичной формы  $F(x)$  стационарный (то есть такой, что  $\text{grad } F(x) = 0$ ) элемент находится по методу Ньютона за одну итерацию, причем с  $\sigma = 1$ .

Действительно, в этом случае стационарный элемент существует и единственный. Разложение  $F(x)$  по формуле Тейлора в окрестности любого  $x_0$  при произвольном элементе вариации  $dx = w$  (то есть отклонении от  $x_0$ ) записывается *без остаточного члена*

$$F(x_0 + w) = F(x_0) + \sum_{j=1}^n \frac{\partial F}{\partial \xi_j} \omega_j + \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \frac{\partial^2 F}{\partial \xi_j \partial \xi_i} \omega_j \omega_i.$$

Тогда условие стационарности  $F(x)$  будет иметь вид

$$\frac{\partial F}{\partial \omega_i} = \frac{\partial F}{\partial \xi_i} + \sum_{j=1}^n \frac{\partial^2 F}{\partial \xi_j \partial \xi_i} \omega_j = 0 \quad ; \quad i = [1, n],$$

что и доказывает проверяемое утверждение.

Тип стационарного элемента (максимум, минимум или "седло") в этом случае зависит от знаковой определенности  $\left\| \hat{\text{Hess}} F(x) \right\|$ .

В заключение приведем пример использования схемы *градиентного подъема* с непрерывным изменением направления (*метод Коши*).

Можно заметить, что поиск локального максимального элемента для непрерывно дифференцируемой функции  $F(x)$  методом наискорейшего подъема сводится к решению следующей задачи Коши:

$$\frac{dx}{d\tau} = \text{grad } F(x); \quad x(0) = x^0.$$

Например, в  $E^2$  для  $F(\xi_1, \xi_2) = -2\xi_1^2 - \xi_2^2$  с  $\|x^0\| = \left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\|$  имеем

$$\begin{cases} \dot{\xi}_1 = -4\xi_1, \\ \dot{\xi}_2 = -2\xi_2, \end{cases} \quad \text{что дает} \quad \begin{cases} \xi_1(\tau) = e^{-4\tau}, \\ \xi_2(\tau) = e^{-2\tau} \end{cases}$$

с фазовой траекторией, задаваемой в силу условий  $\xi_1(0) = 1; \xi_2(0) = 1$  уравнением  $\xi_1 - \xi_2^2 = 0$ , которая при  $\tau = 0$  выходит из начального элемента  $x_0$  и при  $\tau \rightarrow +\infty$  асимптотически стремится к максимальному элементу  $\|x^*\| = \left\| \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\|$ .