Математический анализ
Лекция 9

Емельянов Д.П., Никитин А.А.

МГУ им. М.В. Ломоносова, факультет ВМК
Кафедра общей математики

Онлайн-курс «Математике в Data Science»
7 апреля, 2021г.

Подход, использованный для функций одной переменной оказывается не применим, так как в линейном пространстве, вообще говоря, не определено векторное деление.

Определение

Пусть $m, n \in \mathbb{N}$. Отображение $\hat{A} : [1, m]_{\mathbb{N}} \times [1, n]_{\mathbb{N}} \to \mathbb{R}$, где $[1, m]_{\mathbb{N}} = \{1, 2, \ldots, m\}$, называется вещественной матрицей с m строками и n столбцами. Числа $\hat{A}_{ij} \equiv \hat{A}(i, j)$ называются элементами матрицы $(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$.
Множество всех вещественных матриц размера m на n обозначается $\mathbb{R}^{m\times n}$.

$$\hat{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}.$$

Определение

Пусть $\hat{A} \in \mathbb{R}^{m \times n}$, $\vec{x} \in \mathbb{R}^n$. Тогда произведением $\hat{A} \cdot \vec{x}$ называется вектор $\vec{y} \in \mathbb{R}^m$ такой, что

$$y_i = \sum_{k=1}^{n} a_{i,k} \cdot x_k, \quad i = 1, 2, \ldots, m.$$

Если представить $\hat{A}$ как матрицу из векторов-строк $\vec{a}_i$, то произведение можно записать как:

$$\begin{bmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vdots \\ \vec{a}_m \end{bmatrix} \cdot \vec{x} \equiv \begin{bmatrix} (a_{1,1}, \ldots, a_{1,n}) \\ (a_{2,1}, \ldots, a_{2,n}) \\ \vdots \\ (a_{m,1}, \ldots, a_{m,n}) \end{bmatrix} \cdot \vec{x} = (\langle \vec{a}_1, \vec{x} \rangle, \langle \vec{a}_2, \vec{x} \rangle, \ldots, \langle \vec{a}_m, \vec{x} \rangle).$$

Указанное произведение матрицы на вектор является частным случаем матричного умножения.

Определение

Функция $\vec{f} : \mathbb{R}^m \to \mathbb{R}^n$ называется дифференцируемой (по Фреше), если существует $\hat{A} \in \mathbb{R}^{n\times m}$ такая, что

$$\Delta\vec{f} \equiv \vec{f}(\vec{x} + \Delta\vec{x}) - \vec{f}(\vec{x}) = \hat{A} \cdot \Delta\vec{x} + \bar{o}(|\Delta\vec{x}|).$$

Матрица $\hat{A} = \hat{A}(\vec{x}) = \dfrac{\mathcal{D}\vec{f}}{\mathcal{D}\vec{x}}$ называется производной функции $\vec{f}$ по Фреше в точке $\vec{x}$.

Если $f : \mathbb{R}^m \to \mathbb{R}$, то определение дифференцируемости упрощается: существует вектор $\vec{a} \in \mathbb{R}^m$ такой, что

$$\Delta f \equiv f(\vec{x} + \Delta\vec{x}) - f(\vec{x}) = \langle\vec{a}, \Delta\vec{x}\rangle + \bar{o}(|\Delta\vec{x}|).$$

Вектор $\vec{a} = \vec{a}(\vec{x})$ называется градиентом функции f в точке $\vec{x}$. Обозначение: $\operatorname{grad} f(\vec{x})$, $\nabla f(\vec{x})$.

Производная функций многих переменных

Геометрическая интерпретация

$$\Delta f \approx \langle \operatorname{grad} f(\vec{x}), \Delta\vec{x} \rangle,$$

$$\langle \operatorname{grad} f(\vec{x}), \Delta\vec{x} \rangle = |\operatorname{grad} f(\vec{x})| \cdot |\Delta\vec{x}| \cdot \cos \angle(\operatorname{grad} f(\vec{x}), \Delta\vec{x}),$$

следовательно, градиент – локальное направление наискорейшего возрастания функции f.

Теорема

Пусть функция f дифференцируема в точке x по Фреше. Тогда f непрерывна в точке x.

Доказательство следует из непрерывности операции умножения матрицы на вектор.

Определение

Пусть функция $\vec{f}: \mathbb{R}^m \to \mathbb{R}^n$, пусть $\vec{e} \in \mathbb{R}^m$ – вектор единичной длины (направление). Производной функции $\vec{f}$ по направлению $\vec{e}$ (производной по Гато) в точке $\vec{x}$ называется величина

$$\frac{\partial \vec{f}}{\partial \vec{e}} \equiv \lim_{t \to 0} \frac{\vec{f}(\vec{x} + t \cdot \vec{e}) - \vec{f}(\vec{x})}{t}.$$

Замечание

В одномерном случае определения производной по Фреше и Гато совпадают с обыкновенным определением производной.

Теорема

Пусть функция $\vec{f}:\mathbb{R}^m\to\mathbb{R}^n$ дифференцируема в точке $\vec{x}$ (по Фреше). Тогда $\vec{f}$ имеет в точке $\vec{x}$ производную по любому направлению $\vec{e}$ и имеет место тождество

$$\frac{\partial\vec{f}}{\partial\vec{e}}=\hat{A}\cdot\vec{e}.$$

Для $f:\mathbb{R}^m\to\mathbb{R}$:

$$\frac{\partial f}{\partial\vec{e}}=\langle\operatorname{grad} f,\vec{e}\rangle\ .$$

Обратное, вообще говоря, не верно.

Доказательство.

$$\frac{\partial f}{\partial\vec{e}}=\lim_{t\to0}\frac{\vec{f}(\vec{x}+t\cdot\vec{e})-\vec{f}(\vec{x})}{t}=\lim_{t\to0}\frac{\hat{A}\cdot(t\vec{e})+\bar{o}(|t\vec{e}|)}{t}=$$

$$=\lim_{t\to0}\left(\hat{A}\cdot\vec{e}+\bar{o}(|\vec{e}|)\right)=\hat{A}\cdot\vec{e}.$$

□

Производная функций многих переменных

Определение

Пусть $\vec{e} = \vec{e}_i = (0, ..., 0, 1, 0, ..., 0)$, где 1 находится только в i-й позиции. Тогда величина

$$\frac{\partial \vec{f}}{\partial x_i} \equiv \frac{\partial \vec{f}}{\partial \vec{e}_i} = \lim_{\Delta x_i \to 0} \frac{\vec{f}(x_1, ..., x_{i-1}, x_i + \Delta x_i, x_{i+1}, ..., x_m) - \vec{f}(x_1, ..., x_m)}{\Delta x_i}$$

называется частной производной функции f по переменной x_i.

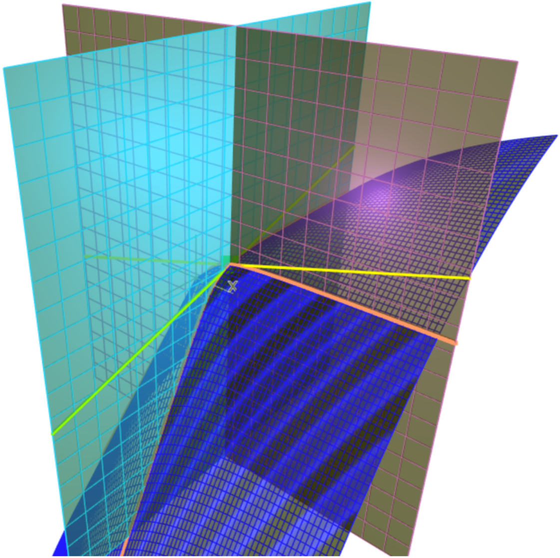
Обозначение: $\dfrac{\partial \vec{f}}{\partial x_i}$, $\vec{f}'_{x_i}$.

Следствие

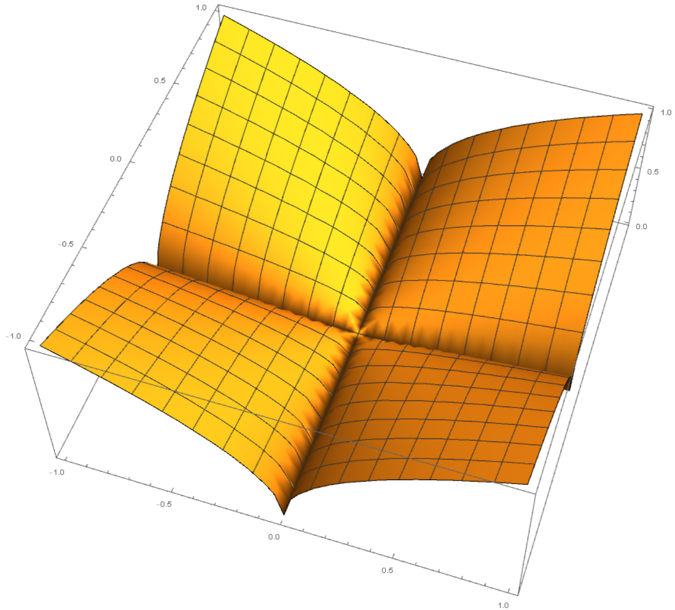
Пусть $\vec{x} = (x_1, ..., x_m)$, $\vec{f}(\vec{x}) = (f_1(\vec{x}), ..., f_n(\vec{x}))$. Так как $\langle \vec{x}, \vec{e}_i \rangle = x_i$, то

$$\hat{A}_{ij} = \frac{\partial f_i}{\partial x_j}, \quad \operatorname{grad} f = \left(\frac{\partial f}{\partial x_1}, \dots \frac{\partial f}{\partial x_m} \right).$$

Матрица $\hat{A}$ называется матрицей Якоби (или якобианом) функции (преобразования) f.



$$f(x, y) = x + y - \sqrt{x^2 + y^2}.$$



$$z(x, y) = \sqrt{|xy|}$$

Пример 1

Пусть $f(x, y, z) = e^{x^2 \cdot y^2 + z^2}$,
$u'_x = 2xy^2 \cdot e^{x^2 \cdot y^2 + z^2}$,
$u'_y = 2yx^2 \cdot e^{x^2 \cdot y^2 + z^2}$,
$u'_z = 2z \cdot e^{x^2 \cdot y^2 + z^2}$.

Пример 2

Пусть $f(\mathbf{x}) = (x^1 + \ldots + x^n)^2$,
$u'_{x_k} = 2(x^1 + \ldots + x^n), \quad k = 1, 2, ..., n.$

Определение

Дифференциалом функции $f : \mathbb{R}^m \to \mathbb{R}$ в точке $\vec{x}$ называется величина

$$df = \langle \operatorname{grad} f, d\vec{x} \rangle = \frac{\partial f}{\partial x_1} \cdot dx_1 + ... + \frac{\partial f}{\partial x_m} \cdot dx_m.$$

Геометрический смысл дифференциала

Пусть $\vec{a} = (a_1, ..., a_m)$, $\vec{x} = (x_1, ..., x_m)$, $f(\vec{a}) = f_0$.
Положим $dx_i = (x_i - a_i)$, $df = (y - f_0)$. Тогда уравнение

$$df = \langle \operatorname{grad} f, d\vec{x} \rangle = \frac{\partial f}{\partial x_1} \cdot dx_1 + ... + \frac{\partial f}{\partial x_m} \cdot dx_m$$

является уравнением касательной плоскости к графику функции f в точке a.

Производная функций многих переменных

Из дифференцируемости следует существование всех частных производных. При каких условиях верно обратное?

Теорема

Пусть у функции $\vec{f}: \mathbb{R}^m \to \mathbb{R}^n$ в точке x и некоторой её окрестности существуют все частные производные и эти частные производные являются непрерывными в точке x.
Тогда f дифференцируема в точке x.

Доказательство.

Докажем для случая двух переменных.

$$f(x+\Delta x, y+\Delta y) = f(x, y+\Delta y) + \frac{\partial f}{\partial x}(x, y+\Delta y)\cdot\Delta x + \bar{o}(\Delta x) =$$

$$= f(x,y) + \frac{\partial f}{\partial y}(x,y)\cdot\Delta y + \bar{o}(\Delta y) + \left(\frac{\partial f}{\partial x}(x,y) + \bar{o}(1)\right)\cdot\Delta x + \bar{o}(\Delta x) =$$

$$= f(x,y) + \frac{\partial f}{\partial x}(x,y)\cdot\Delta x + \frac{\partial f}{\partial y}(x,y)\cdot\Delta y + \bar{o}(\sqrt{\Delta x^2 + \Delta y^2}).$$

$\square$

Теорема (производная композиции)

Пусть функция $\vec{y}: \mathbb{R}^m \to \mathbb{R}^n$ дифференцируема в точке $\vec{x}$, а функция $\vec{z}: \mathbb{R}^n \to \mathbb{R}^k$ дифференцируема в точке $\vec{y} = \vec{y}(\vec{x})$. Тогда функция $z(y(x))$ является дифференцируемой в точке $\vec{x}$ и её производная имеет вид

$$\frac{\mathcal{D}\vec{z}}{\mathcal{D}\vec{x}}(\vec{x}) = \frac{\mathcal{D}\vec{z}}{\mathcal{D}\vec{y}}(\vec{y}(\vec{x})) \cdot \frac{\mathcal{D}\vec{y}}{\mathcal{D}\vec{x}}(\vec{x}),$$

где $\cdot$ – матричное умножение. $\frac{\mathcal{D}\vec{z}}{\mathcal{D}\vec{x}} \in \mathbb{R}^{k \times m}$.
В частном случае $z: \mathbb{R}^n \to \mathbb{R}$ $(k = 1)$ мы получаем формулу:

$$\frac{\partial z}{\partial x_k} = \frac{\partial z}{\partial y_1} \cdot \frac{\partial y_1}{\partial x_k} + \ldots + \frac{\partial z}{\partial y_n} \cdot \frac{\partial y_n}{\partial x_k}.$$

Доказательство.

$$\Delta\vec{z} = \frac{\mathcal{D}\vec{z}}{\mathcal{D}\vec{y}}(\vec{y}(\vec{x})) \cdot \Delta\vec{y} + \bar{o}(\Delta\vec{y}) =$$

$$= \frac{\mathcal{D}\vec{z}}{\mathcal{D}\vec{y}}(\vec{y}(\vec{x})) \cdot \left(\frac{\mathcal{D}\vec{y}}{\mathcal{D}\vec{x}}(\vec{x}) \cdot \Delta\vec{x} + \bar{o}(\Delta\vec{x})\right) + \bar{o}\left(\frac{\mathcal{D}\vec{y}}{\mathcal{D}\vec{x}}(\vec{x}) \cdot \Delta\vec{x} + \bar{o}(\Delta\vec{x})\right) =$$

$$= \frac{\mathcal{D}\vec{z}}{\mathcal{D}\vec{y}}(\vec{y}(\vec{x})) \cdot \frac{\mathcal{D}\vec{y}}{\mathcal{D}\vec{x}}(\vec{x}) \cdot \Delta\vec{x} + \bar{o}(\Delta\vec{x}).$$

□

Определение

Пусть функция f дифференцируема в точке x и все её частные производные являются дифференцируемыми функциями в точке x. Тогда функция f называется дважды дифференцируемой в точке x.
Функции
$$\frac{\partial^2 f}{\partial x_i \partial x_j} \equiv \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)$$
называются частными производными второго порядка функции f.
Если $i \neq j$, то частная производная называется смешанной.

Аналогично вводится понятие N раз дифференцируемой функции многих переменных.

Влияет ли порядок дифференцирования на результат?
В общем случае - да.

Пример

$$f(x,y) = \begin{cases} \dfrac{xy(x^2-y^2)}{x^2+y^2}, & x^2+y^2>0, \\ 0, & x^2+y^2=0. \end{cases}$$

Теорема

Пусть функция $f : \mathbb{R}^m \to \mathbb{R}^n$ дважды дифференцируема в точке x. Тогда

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Теорема

Пусть функция $f : \mathbb{R}^m \to \mathbb{R}^n$ N раз дифференцируема в точке x. Тогда смешанные производные порядка N не зависят от порядка дифференцирования.

Доказательство.

Докажем утверждение для функции двух переменных $f(x, y)$.

$$\Phi(\Delta) = f(x+\Delta, y+\Delta) - f(x+\Delta, y) - f(x, y+\Delta) + f(x, y) =$$

$$= \left(f'_y(x+\Delta, y+\theta_1\Delta) - f'_y(x, y+\theta_1\Delta)\right)\Delta =$$

$$= \left(f'_y(x+\Delta, y+\theta_1\Delta) - f'_y(x, y)\right)\Delta + \left(f'_y(x, y) - f'_y(x, y+\theta_1\Delta)\right)\Delta =$$

$$= f''_{yx}\Delta^2 + f''_{yy}\theta_1\Delta^2 - f''_{yy}\theta_1\Delta^2 + \bar{o}(\Delta^2) = f''_{yx}\Delta^2 + \bar{o}(\Delta^2).$$

Аналогично

$$\Phi(\Delta) = f''_{xy}\Delta^2 + \bar{o}(\Delta^2),$$

$$f''_{yx}\Delta^2 + \bar{o}(\Delta^2) = f''_{xy}\Delta^2 + \bar{o}(\Delta^2),$$

$$f''_{yx} + \bar{o}(1) = f''_{xy} + \bar{o}(1),$$

устремим Δ к нулю и получим

$$f''_{yx} = f''_{xy}.$$

Теорема доказана. $\square$

Формула Тейлора

Теорема

Пусть функция $f : \mathbb{R}^m \to \mathbb{R}$ является $n+1$ раз дифференцируемой в некоторой окрестности точки x^0. Тогда для любого x из данной окрестности имеет место формула Тейлора с остаточным членом в форме Лагранжа:

$$f(x) = f(x^0) + \frac{1}{1!}\sum_{k=1}^{m} \frac{\partial f}{\partial x_k}(x^0)\cdot(x_k - x_k^0) + \ldots +$$

$$+\frac{1}{n!}\sum_{\substack{k_1+k_2+\ldots+k_m\leqslant n\\ 0\leqslant k_i\leqslant n}} \frac{\partial^n f}{\partial x_1^{k_1}\ldots\partial x_m^{k_m}}(x^0)\cdot(x_1 - x_1^0)^{k_1}\cdot\ldots\cdot(x_m - x_m^0)^{k_m} + R_{n+1}(x),$$

$$R_{n+1}(x) = \frac{1}{(n+1)!}\sum_{\substack{k_1+k_2+\ldots+k_m\leqslant n+1\\ 0\leqslant k_i\leqslant n+1}} \frac{\partial^n f}{\partial x_1^{k_1}\ldots\partial x_m^{k_m}}(\xi)\cdot(x_1 - x_1^0)^{k_1}\cdot\ldots\cdot(x_m - x_m^0)^{k_m},$$

где $\xi = (\xi_1, \ldots, \xi_m)$ – некоторая точка из окрестности x^0.

Формула Тейлора

Теорема

Пусть функция $f:\mathbb{R}^m\to\mathbb{R}$ является $n-1$ раз дифференцируемой в некоторой окрестности точки x^0 и n раз в самой точке x^0. Тогда для любого x из данной окрестности имеет место формула Тейлора с остаточным членом в форме Пеано:

$$f(x)=f(x^0)+\frac{1}{1!}\sum_{k=1}^{m}\frac{\partial f}{\partial x_k}(x^0)\cdot(x_k-x_k^0)+\ldots+$$

$$+\frac{1}{n!}\sum_{\substack{k_1+k_2+\ldots+k_m\leqslant n\\ 0\leqslant k_i\leqslant n}}\frac{\partial^n f}{\partial x_1^{k_1}\ldots\partial x_m^{k_m}}(x^0)\cdot(x_1-x_1^0)^{k_1}\cdot\ldots\cdot(x_m-x_m^0)^{k_m}+R_{n+1}(x),$$

$$R_{n+1}(x)=\bar{o}\left(|x-x^0|^n\right).$$

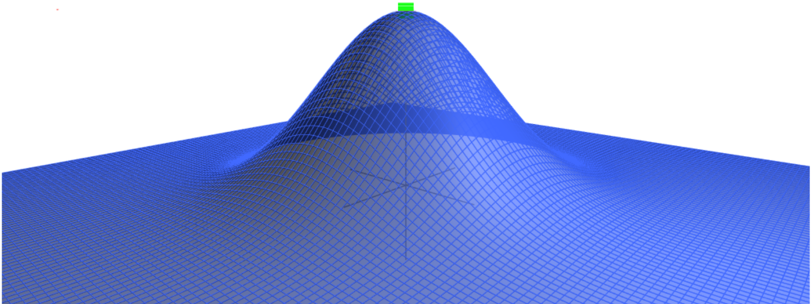
Определение

Далее рассматриваем $f : \mathbb{R}^m \to \mathbb{R}$. Пусть для некоторой точки x^0 существует $U_\delta(x^0)$ такая, что

$$\forall x \in U_\delta(x^0) \implies f(x) \leqslant f(x^0) \quad (f(x) \geqslant f(x^0)),$$

тогда точка x^0 называется точкой локального максимума (минимума) функции f.

Говорят, что f имеет в точке x^0 локальный экстремум, если она имеет в этой точке локальный максимум или локальный минимум.



Экстремумы функций многих переменных

Теорема

Пусть x^0 – точка локального экстремума для функции f, внутренняя для её области определения, Пусть f дифференцируема в x^0. Тогда $\operatorname{grad} f(x^0) = 0$.

Доказательство.

Применим формулу Тейлора с остаточным членом в форме Пеано:

$$f(x) - f(x^0) = \left\langle \operatorname{grad} f(x^0), (x - x^0) \right\rangle + \bar{o}\left(|x - x^0|\right) \leqslant 0.$$

Пусть $\operatorname{grad} f(x^0) \neq 0$. Тогда положим

$$x = x^0 + t \cdot e, \quad e = \frac{\operatorname{grad} f(x^0)}{|\operatorname{grad} f(x^0)|}.$$

Имеем

$$\left\langle \operatorname{grad} f(x^0), t \cdot e \right\rangle + \bar{o}\left(|t|\right) \leqslant 0, \quad t \neq 0,$$

$$t \cdot \left(|\operatorname{grad} f(x^0)| + \bar{o}(1)\right) \leqslant 0, \quad t \neq 0.$$

Выражение в скобках положительно при достаточно малых t, следовательно левая часть меняет знак – противоречие. Следовательно, $\operatorname{grad} f(x^0) = 0$. Теорема доказана. $\square$

Определение

Матрица $\hat{A} \in \mathbb{R}^{m \times m}$ называется симметричной, если $\hat{A}_{ij} = \hat{A}_{ji}$.

Определение

Матрица $\hat{A} \in \mathbb{R}^{m \times m}$ называется положительно определённой, если

$$\forall \vec{x} \in \mathbb{R}^m, \vec{x} \neq 0 \implies \left\langle \hat{A}\vec{x}, \vec{x} \right\rangle > 0.$$

Обозначение $\hat{A} > 0$.

Определение

Матрица $\hat{A} \in \mathbb{R}^{m \times m}$ называется отрицательно определённой, если

$$\forall \vec{x} \in \mathbb{R}^m, \vec{x} \neq 0 \implies \left\langle \hat{A}\vec{x}, \vec{x} \right\rangle < 0.$$

Обозначение $\hat{A} < 0$.

Определение

Пусть функция f дважды дифференцируема в точке x. Тогда матрицей Гессе (гессианом) функции f называется матрица

$$\hat{H}(x) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_m \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_m \partial x_m} \end{bmatrix}.$$

Мы доказали, что при данных условиях матрица Гессе является симметричной.

Теорема

Пусть функция f дважды дифференцируема в точке x и дифференцируема в некоторой её окрестности, $\operatorname{grad} f(x) = 0$. Тогда если матрица Гессе определена положительно, то x – точка локального минимума.
Если матрица Гессе определена отрицательно, то x – точка локального максимума.

Доказательство.

Рассмотрим $\hat{H}(x) > 0$ (случай $\hat{H}(x) < 0$ рассматривается аналогично). Пусть $\Delta x \neq 0$. Разложим функцию f в ряд Тейлора с остаточным членом в форме Пеано до 2-й производной:

$$\Delta f = f(x + \Delta x) - f(x) = \langle \operatorname{grad} f(x), \Delta x \rangle + \left\langle \hat{H}(x) \cdot \Delta x, \Delta x \right\rangle + \bar{o}\left(|\Delta x|^2\right),$$

$$\Delta f = |\Delta x|^2 \left(\left\langle \hat{H}(x) \cdot e, e \right\rangle + \bar{o}(1) \right),$$

где $e = \Delta x / |\Delta x|$.

На единичной сфере функция $F(e) = \left\langle \hat{H}(x) \cdot e, e \right\rangle$ непрерывна и положительна. По теореме Вейерштрасса существует $\delta > 0$ такое, что $F(e) \geqslant \delta$. При достаточно малых Δx получим

$$\Delta f \geqslant |\Delta x|^2 \left(\frac{\delta}{2} \right) > 0,$$

следовательно x – точка локального минимума. Теорема доказана. □

Определение

Матрица $\hat{A} \in \mathbb{R}^{m\times m}$ называется знакопеременной, если

$$\exists \vec{x}_1, \vec{x}_2 \in \mathbb{R}^m : \left\langle \hat{A}\vec{x}_1, \vec{x}_1 \right\rangle > 0, \left\langle \hat{A}\vec{x}_2, \vec{x}_2 \right\rangle < 0.$$

Теорема

Пусть функция f дважды дифференцируема в точке x и дифференцируема в некоторой её окрестности, $\operatorname{grad} f(x) \neq 0$ или $\hat{H}(x)$ – знакопеременна.
Тогда x не является точкой локального экстремума.

Доказательство.

Случай $\operatorname{grad} f(x) \neq 0$ был рассмотрен ранее в необходимом условии экстремума.
Выберем $\Delta x_1 = x_1/|x_1|$, $\Delta x_2 = x_2/|x_2|$, где x_1 и x_2 векторы из определения знакопеременности. Тогда

$$\Delta f_1 = f(x + t\Delta x_1) - f(x) = \left\langle \hat{H}(x) \cdot t\Delta x_1, t\Delta x_1 \right\rangle + \bar{o}\left(|t|^2\right),$$

$$\Delta f_1 = |t|^2 \left(\left\langle \hat{H}(x) \cdot \Delta x_1, \Delta x_1 \right\rangle + \bar{o}(1) \right),$$

аналогично

$$\Delta f_2 = f(x + t\Delta x_2) - f(x) = \left\langle \hat{H}(x) \cdot t\Delta x_2, t\Delta x_2 \right\rangle + \bar{o}\left(|t|^2\right),$$

$$\Delta f_2 = |t|^2 \left(\left\langle \hat{H}(x) \cdot \Delta x_2, \Delta x_2 \right\rangle + \bar{o}(1) \right),$$

Выберем t настолько малым, чтобы было выполнено

$$|\bar{o}(1)| \leqslant \min \left(\left| \left\langle \hat{H}(x) \cdot \Delta x_1, \Delta x_1 \right\rangle \right|, \left| \left\langle \hat{H}(x) \cdot \Delta x_2, \Delta x_2 \right\rangle \right| \right).$$

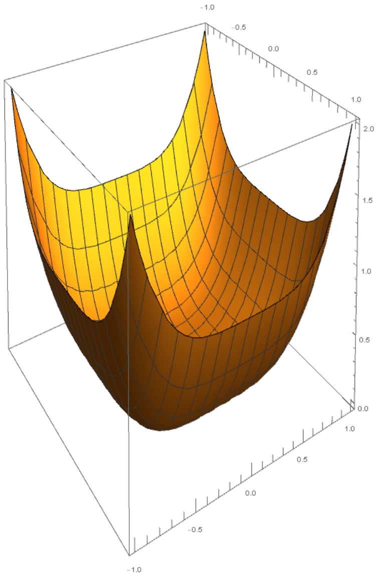
Тогда $\Delta f_1 > 0$, а $\Delta f_2 < 0$, что есть условие отсутствия экстремума. $\square$

Замечание

Если $\hat{H}(x) \geqslant 0$ или $\hat{H}(x) \leqslant 0$, то функция требует дополнительных исследований.

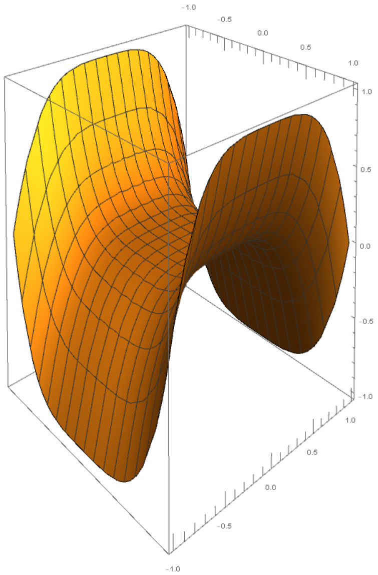
Примеры

$f_1(x, y) = x^4 + y^4$, $f_2(x, y) = x^4 - y^4$.
Имеем $\hat{H}(0, 0) = 0$, но f_1 имеет локальный экстремум в точке $(0, 0)$, а у f_2 в этой точке экстремума нет.



$f(x, y) = x^4 + y^4$

Экстремумы функций многих переменных



$$f(x, y) = x^4 - y^4$$

Определитель матрицы

$$\det\left[x_{11}\right] = x_{11},$$

$$\det\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = x_{11}x_{22} - x_{12}x_{21}.$$

$$\det\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} =$$

$$= x_{11}x_{22}x_{33} + x_{12}x_{23}x_{31} + x_{21}x_{32}x_{13} - x_{31}x_{22}x_{13} - x_{21}x_{12}x_{33} - x_{11}x_{32}x_{23} =$$

$$= x_{11}\cdot\det\begin{bmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{bmatrix} - x_{12}\cdot\det\begin{bmatrix} x_{21} & x_{23} \\ x_{31} & x_{33} \end{bmatrix} + x_{13}\cdot\det\begin{bmatrix} x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix}.$$

Определение

Пусть $\hat{A} \in \mathbb{R}^{m \times m}$. Угловым минором порядка n, $1 \leqslant n \leqslant m$ называется определитель

$$\Delta_n = \det \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix}.$$

Иначе говоря – определитель квадратной подматрицы, расположенной в левом верхнем углу исходной.

Теорема (критерий Сильвестра)

Матрица $\hat{A}$ определена положительно тогда и только тогда, когда

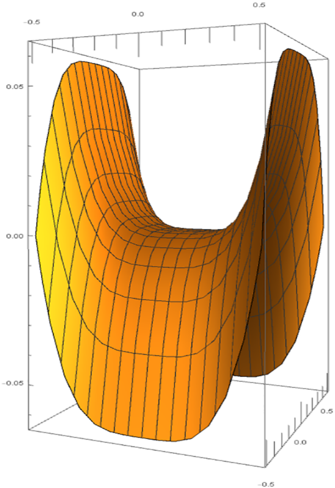
$$\Delta_n > 0, \quad n = 1, 2, ..., m.$$

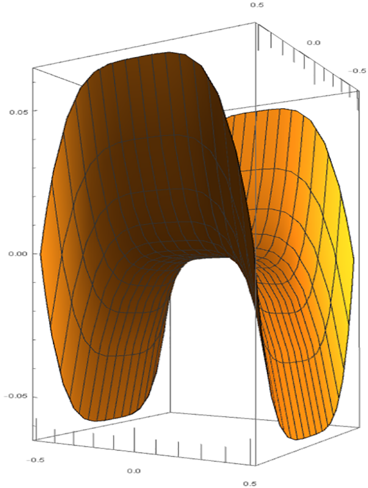
Матрица $\hat{A}$ определена отрицательно тогда и только тогда, когда

$$(-1)^n \cdot \Delta_n > 0, \quad n = 1, 2, ..., m.$$

Пример

Рассмотрим функцию $f(x, y) = x^4 - y^4$. При условии $x = 0$ точка $(0, 0)$ – локальный максимум, при $y = 0$ точка $(0, 0)$ – локальный минимум. Без дополнительных условий в точке $(0, 0)$ локального экстремума нет.





Определение

Пусть на открытом множестве $G \subset \mathbb{R}^m$, заданы функции $f : G \to \mathbb{R}$, $\vec{F} : G \to \mathbb{R}^n$. Уравнение

$$\vec{F}(\vec{x}) = 0$$

называется уравнением связи (условиями связи).

Обозначим за E множество всех решений уравнения $\vec{F}(\vec{x}) = 0$.

Определение

Точка $\vec{x}_0 \in E$ называется точкой условного минимума функции f при наличии уравнений связи $\vec{F}(\vec{x}) = 0$, если

$$\exists \delta > 0 \ : \ \forall \vec{x} \in U_\delta(\vec{x}_0) \cap E \Longrightarrow f(\vec{x}) > f(\vec{x}_0).$$

Определение

Точка $\vec{x}_0 \in E$ называется точкой условного максимума функции f при наличии уравнений связи $\vec{F}(\vec{x}) = 0$, если

$$\exists \delta > 0 \ : \ \forall \vec{x} \in U_\delta(\vec{x}_0) \cap E \Longrightarrow f(\vec{x}) < f(\vec{x}_0).$$

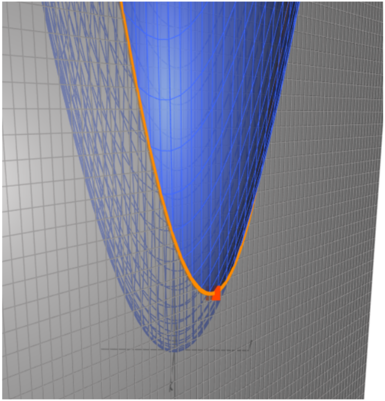
Условный экстремум

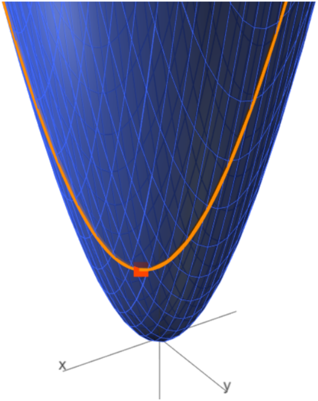
Пример

Исследуем на экстремум функцию $f(x,y)=x^2+y^2$ при дополнительном условии $x+y-1=0$.

$$f(x,1-x)=2x^2-2x+1,\quad f'(x,1-x)=4x-2=0\Leftrightarrow x=\tfrac{1}{2},\ y=1-\tfrac{1}{2}=\tfrac{1}{2}.$$

В точке $\left(\frac{1}{2},\frac{1}{2}\right)$ функция $f(x,y)=x^2+y^2$ достигает минимума относительно уравнения связи $x+y-1=0$.





Определение

Множество X называется выпуклым, если

$$\forall x, y \in X, \forall \alpha \in (0,1) \implies \alpha x + (1-\alpha)y \in X.$$

Иначе говоря, если две точки x и y лежат в X, то и отрезок $[x, y]$ лежит в X.

Определение

Функция f, заданная на выпуклом множестве X называется выпуклой, если

$$\forall x, y \in X, \forall \alpha \in (0,1) \implies f(\alpha x+(1-\alpha)y) \leqslant \alpha f(x)+(1-\alpha)f(y),$$

сильно выпуклой, если

$$f(\alpha x+(1-\alpha)y) \leqslant \alpha f(x)+(1-\alpha)f(y)-\frac{\mu}{2}\alpha(1-\alpha)\cdot|x-y|^2, \quad \mu > 0.$$

Градиентный спуск

Постановка задачи

Пусть задан функционал $F : X \to \mathbb{R}$. Найти экстремаль x_0 такую, что

$$F(x_0) = \min_{x \in X} F(x).$$

Описание метода

Метод градиентного спуска – итерационный метод оптимизации (поиска минимума), при котором очередной шаг задаётся формулой

$$x_{n+1} = x_n - \alpha_n \cdot \operatorname{grad} F(x_n),$$

где α_n – некоторые положительные коэффициенты. Различный выбор α_n порождает различные «версии» градиентного спуска.

Градиентный спуск

Выбор α_n

1. $\alpha_n = \alpha = const$.
2. $\alpha_n = \dfrac{\alpha}{2^k(n)}$ – метод дробления, где $k(n) \nearrow$ и выбирается из условия $F(x_{n+1}) \leqslant F(x_n)$.
3. $\alpha_n = \dfrac{\alpha}{2^k(n)} \cdot \dfrac{1}{|\mathrm{grad}\, F(x_n)|}$ – метод дробления при шаге, не зависящем от длины градиента.
4. α_n выбирается так, чтобы минимизировать функцию одного переменного $\varphi(\alpha) = F(x_n - \alpha \cdot \mathrm{grad}\, F(x_n))$ – метод скорейшего спуска.

Правило остановки метода

1. $|F(x_{n+1}) - F(x_n)|$ достаточно мал на протяжении нескольких последних шагов.
2. $|x_{n+1} - x_n|$ достаточно мал на протяжении нескольких последних шагов.
3. $|\mathrm{grad}\, F(x_n)|$ достаточно мал на протяжении нескольких последних шагов.

Градиентный спуск

Теорема

Пусть F – сильно выпуклый с постоянной $\mu > 0$, дважды дифференцируемый функционал на выпуклом и замкнутом X, его гессиан ограничен:

$$\left|\left\langle \hat{H}(x)\xi, \xi \right\rangle\right| \leqslant L|\xi|^2.$$

Тогда метод градиентного спуска с $\alpha = const$ сходится к экстремали при любом начальном приближении если выполнено

$$\forall x \in X \implies x - \alpha \cdot \operatorname{grad} F(x) \in X,$$

$$0 < \alpha < \frac{2\mu}{L^2}.$$

Доказательство. Рассмотрим функцию шага
$$S(x) = x - \alpha \cdot \mathrm{grad}\, F(x).$$
Для любых $x, y \in X$
$$|S(x) - S(y)|^2 = |x - \alpha \cdot \mathrm{grad}\, F(x) - y + \alpha \cdot \mathrm{grad}\, F(y)|^2 =$$
$$= |x-y|^2 + \alpha^2 |\mathrm{grad}\, F(x) - \mathrm{grad}\, F(y)|^2 - 2\alpha \langle x - y, \mathrm{grad}\, F(x) - \mathrm{grad}\, F(y)\rangle.$$
Отметим, что $\dfrac{\mathcal{D}(\mathrm{grad}\, F)}{\mathcal{D}x}(x) = \hat{H}(x)$. Тогда в силу формулы Тейлора
$$|\mathrm{grad}\, F(x) - \mathrm{grad}\, F(y)| = |\hat{H}(\xi) \cdot (x - y)| \leqslant L \cdot |x - y|.$$
Далее,
$$\frac{\mu}{2}\alpha(1-\alpha) \cdot |x-y|^2 \leqslant \alpha F(x) + (1-\alpha)F(y) - F(\alpha x + (1-\alpha)y) =$$
$$= F(y) - F(y + \alpha(x-y)) + \alpha(F(x) - F(y)) =$$
$$= -\langle \mathrm{grad}\, F(y), \alpha(x-y)\rangle + \alpha(F(x) - F(y)) + \bar{o}(\alpha).$$
Поделим обе части на α и устремим α к нулю.
$$\frac{\mu}{2} \cdot |x-y|^2 \leqslant -\langle \mathrm{grad}\, F(y), (x-y)\rangle + (F(x) - F(y))$$
или
$$F(x) \geqslant F(y) + \langle \mathrm{grad}\, F(y), (x-y)\rangle + \frac{\mu}{2} \cdot |x-y|^2.$$

Градиентный спуск

$$F(x) \geqslant F(y) + \langle \operatorname{grad} F(y), (x-y) \rangle + \frac{\mu}{2} \cdot |x-y|^2.$$

Аналогично,

$$F(y) \geqslant F(x) + \langle \operatorname{grad} F(x), (y-x) \rangle + \frac{\mu}{2} \cdot |x-y|^2.$$

Складывая неравенства, получаем

$$F(x) + F(y) \geqslant F(x) + F(y) + \langle \operatorname{grad} F(y) - \operatorname{grad} F(x), (x-y) \rangle + \mu \cdot |x-y|^2,$$

$$\langle \operatorname{grad} F(x) - \operatorname{grad} F(y), x-y \rangle \geqslant \mu \cdot |x-y|^2.$$

Возвращаясь к S и применяя полученные неравенства, получим:

$$|S(x) - S(y)|^2 \leqslant |x-y|^2 + \alpha^2 L^2 |x-y|^2 - 2\alpha\mu |x-y|^2 =$$

$$= |x-y|^2 \left(1 - 2\alpha\mu + \alpha^2 L^2\right).$$

Обозначим $\left(1 - 2\alpha L + \alpha^2 L^2\right) = \beta^2$. При выбранном α число $\beta < 1$, $\beta > 0$. Тогда

$$|S(x) - S(y)| \leqslant \beta |x-y|,$$

то есть отображение S является сжимающим.

Отображение S является сжимающим и непрерывным на X, так как F дважды дифференцируема.
$S : X \to X$, следовательно, в силу теоремы о неподвижной точке существует единственный $x \in X$ такой, что $x = S(x)$, кроме того, итерационная последовательность метода $x_{n+1} = S(x_n)$ сходится к x.
Покажем, что x – экстремаль.

$$x = S(x) \equiv x - \alpha \cdot \operatorname{grad} F(x) \Longleftrightarrow \operatorname{grad} F(x) = 0,$$

то есть x – единственный потенциальный экстремум функции F. Так как F имеет минимум по теореме Вейерштрасса, то x им и является. Теорема доказана.

Рекомендуемые задачи для решения

«Листочки» – 40.1, 40.3, 41.10, 46.3(авджзн).