Определение максимума, минимума и седлового элемента

Определение	Говорят, что функция $F(x)$ имеет *локальный максимум* (*минимум*) на элементе $x^* \in E^n$, если существует $U_\delta(x^*)$ – окрестность этого элемента такая, что $$F(x) \le F(x^*),\ \forall x \in U_\delta(x^*)$$ (для случая *минимума* $$F(x) \ge F(x^*),\ \forall x \in U_\delta(x^*)$$).

В этом случае принята терминология: x^* – *экстремальный элемент* для $F(x)$, а $F(x^*)$ – *экстремальное значение* (*экстремум*) функции $F(x)$.

Задачу поиска экстремального элемента

$$x^* = \arg\max_{x \in D} F(x)$$

на D – открытом множестве в области определения функции $F(x)$, принято называть *классической экстремальной задачей*, или *задачей безусловной оптимизации*.

Определение *Седловым элементом* функции $F(x)$ называется элемент x^* такой, что при переносе начала координат в элемент x^* найдутся два подпространства Ω_1 и Ω_2 $\Omega_1 \oplus \Omega_2 = E^n$ такие, что

$$x^* = \arg\max_{x \in \Omega_1} F(x) \quad \text{и} \quad x^* = \arg\min_{x \in \Omega_2} F(x).$$

Необходимые условия экстремума

Введем для задачи поиска максимального значения функции понятие вариаций, "*улучшающих значение функции* $F(x)$" на элементе

.

Определение

Пусть для некоторого элемента $\delta x \in E^n$ найдется число $\varepsilon > 0$ такое, что $x^0 + \tau \delta x \in D$ и

$$F(x^0 + \tau \delta x) > F(x^0)\,,\ \forall \tau : 0 < \tau < \varepsilon\,.$$

Тогда $P(x^0)$ – совокупность всех таких элементов δx назовем *множеством улучшающих значение функции вариаций на элементе* x^0.

Необходимое условие существования экстремума функционала $F(x)$ на элементе $x^* \in E^n$ формулирует следующая теорема.

Теорема **Для того чтобы функция $F(x)$ имела экстремум на элементе x^*, необходимо, чтобы на этом элементе градиент функции либо *не существовал*, либо *равнялся нулевому элементу* в E^n.**

Доказательство.

Действительно, пусть $\operatorname{grad} F$ существует на максимальном элементе x^* и $\operatorname{grad} F \neq o$, тогда из соотношения

$$F(x^* + dx) - F(x^*) = (\operatorname{grad} F, dx) + \mathrm{o}(|\, dx \,|)$$

следует, что если для некоторой ненулевой, достаточно малой по норме не улучшающей вариации dx выполнено условие

$$F(x^* + dx) < F(x^*) \,,$$

то будет верно и неравенство

$$F(x^* - dx) > F(x^*) \,.$$

То есть, $-dx$ – улучшающая вариация и $F(x^*)$ не является максимальным значением для функционала $F(x)$.

Случай минимума рассматривается аналогично.

Теорема доказана.

Элементы в E^n, для которых $\operatorname{grad} F = o$ или не существует, называются *критическими* элементами функционала $F(x)$. Для дифференцируемых функционалов элементы в E^n, для которых $\operatorname{grad} F = o$, принято называть *стационарными*.

Необходимое условие экстремума не является достаточным. Это следует из факта отсутствия экстремума на нулевом элементе у функционала $F(x) = \xi_1 \xi_2$, $x \in E^2$, хотя на этом элементе его градиент

$$\| \operatorname{grad} F \| = \begin{Vmatrix} \dfrac{\partial F}{\partial \xi_1} & \dfrac{\partial F}{\partial \xi_2} \end{Vmatrix}^{\mathrm{T}} = \begin{Vmatrix} \xi_2 \\ \xi_1 \end{Vmatrix}$$

нулевой, то есть $\operatorname{grad} F(x) \big|_{x=o} = o$.

Достаточные условия экстремума

Достаточные условия существования (или отсутствия) строгого экстремума функции $F(x)$ на элементе $x^* \in E^n$ дает

Теорема **Пусть функция $F(x)$ дважды непрерывно дифференцируема в некоторой окрестности стационарного элемента x^*. Тогда у функция $F(x)$ на x^*:**

1º. имеется *строгий минимум*, если на этом элементе

$$d^2F = (dx, \widehat{\mathrm{Hess}}\, F\, dx) > 0\ ,\ \forall dx \neq o\ ;$$

2º. имеется *строгий максимум*, если на этом элементе

$$d^2F = (dx, \widehat{\mathrm{Hess}}\, F\, dx) < 0\ ,\ \forall dx \neq o\ ;$$

3. отсутствует строгий экстремум, если на этом элементе $d^2F = (dx, \widehat{\mathrm{Hess}}\, F\, dx)$ не является знакопостоянным.

Доказательство.

Действительно, из формулы Тейлора при условии стационарности x_0 следует, что

$$F(x_0 + dx) - F(x_0) = \frac{1}{2} \| dx \|^{\mathrm{T}} \| \mathrm{Hess}\, F \| \| dx \| + \mathrm{o}(| dx |^2),$$

но тогда для достаточно малых по норме dx знак приращения значения функции будет зависеть от знаковой определенности *квадратичного формы*

$$d^2 F = (dx, \hat{\mathrm{Hess}}\, F\, dx) > 0 ,$$

порождаемой в E^n симметрической матрицей Гессе.

Теорема доказана.

Как известно из курса линейной алгебры, исследование квадратичной формы на знаковую определенность можно выполнять

либо методом приведения ее координатного представления к *диагональному виду*, построив специальный базис,

либо непосредственно в исходном базисе – при помощи *критерия Сильвестра*.

Использование критерия Сильвестра позволяет утверждать, что для дважды непрерывно дифференцируемой функцииа $F(x)$ на стационарном элементе:

1°. Имеет строгий минимум, если на этом элементе у матрицы Гессе $\left\| \dfrac{\partial^2 F}{\partial \xi_j \partial \xi_i} \right\|$ все главные диагональные миноры *положительные*;

2°. Имеет строгий максимум, если на этом элементе у матрицы Гессе $\left\| \dfrac{\partial^2 F}{\partial \xi_j \partial \xi_i} \right\|$ все четные главные диагональные миноры *положительные*, а все нечетные – *отрицательные.*

3°. Не имеет строго экстремума в остальных случаях.

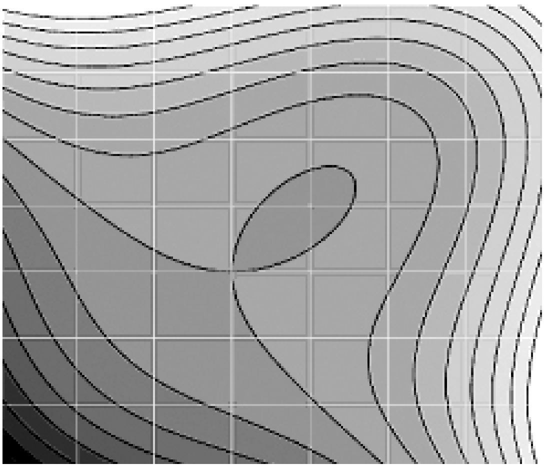
В качестве примера рассмотрим в E^2 функцию

$$F(\xi_1, \xi_2) = \xi_1^3 + \xi_2^3 - 3\xi_1\xi_2,$$

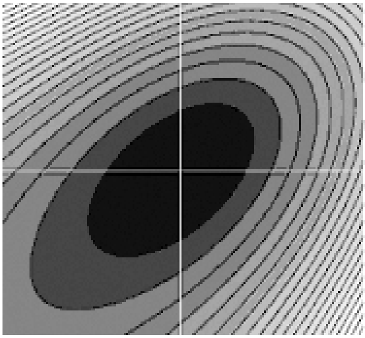
градиент которой имеет координатное представление

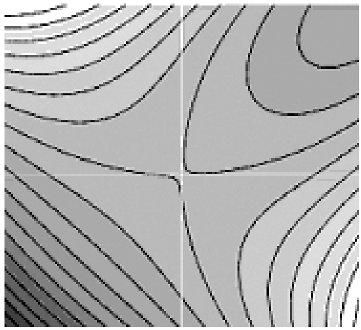
$$\| \operatorname{grad} F \| = \begin{Vmatrix} 3\xi_1^2 - 3\xi_2 \\ 3\xi_2^2 - 3\xi_1 \end{Vmatrix},$$

а гессиан представляется матрицей $\left\| \overset{\Lambda}{\operatorname{Hess}} F \right\| = \begin{Vmatrix} 6\xi_1 & -3 \\ -3 & 6\xi_2 \end{Vmatrix}.$



Система изолиний для $F(\xi_1, \xi_2) = \xi_1^3 + \xi_2^3 - 3\xi_1\xi_2$





Окрестность x_1 Окрестность x_2

Легко видеть, что рассматриваемая функция определена $\forall x \in E^2$ и имеет только два стационарных элемента:

$$\| x_1 \| = \left\| \begin{matrix} 1 \\ 1 \end{matrix} \right\| \quad \text{и} \quad \| x_2 \| = \left\| \begin{matrix} 0 \\ 0 \end{matrix} \right\|,$$

причем (согласно критерию Сильвестра) на x_1 $F(\xi_1, \xi_2)$ она имеет *локальный минимум*, в то время как, x_2 – *седловой элемент*, не являющийся экстремальным.

Условия теоремы не являются, вообще говоря, *необходимыми условиями экстремума*. Например, в E^2 для функционала $\xi_1^4 + \xi_2^4$ в начале координат имеется строгий минимум.

Отметим, однако, что даже существование окрестности стационарного элемента с определенной локальной выпуклостью может не являться необходимым условием строгого экстремума, что иллюстрируется следующим примером. В E^1 у функционала

$$F(\xi_1) = \begin{cases} \xi_1^2\left(2 + \sin\dfrac{1}{\xi_1}\right), & \xi_1 \neq 0, \\ 0\,, & \xi_1 = 0 \end{cases}$$

имеется строгий минимум при $\xi_1 = 0$, хотя в любой, сколь угодно малой окрестности начала координат имеются области локальной выпуклости как вниз, так и вверх.

Общая схема поиска локального экстремума

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

Реально реализуемой альтернативой является использование так называемых *итеративных численных методов*. Все эти методы также основаны на необходимых или достаточных условиях экстремума функции в E^n и 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 элемент x^0 принимается за x^* – максимальный, либо, при $\sigma_0 = +\infty$, констатируется факт отсутствия максимума у $F(x)$.

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, $\rho(x^k, x^*) \to 0$ при $k \to \infty$) зависят, как от способа выбора w^k и σ_k, так и от свойств максимизируемой функции.

Методы 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 $(\operatorname{grad} F(x^k), w^k) > 0$.

3°. Величина шага по выбранному направлению максимизации σ_k находится путем решения одномерной задачи

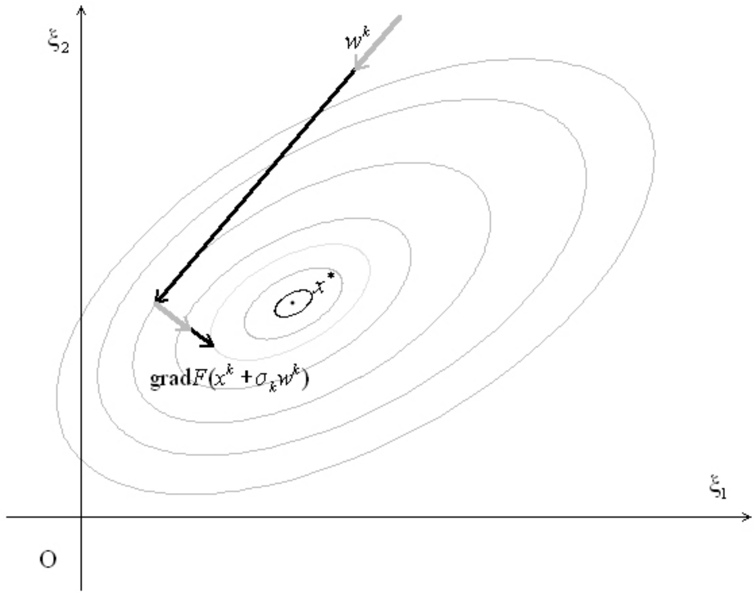
$$\sigma_k = \max_{\sigma \geq 0} F(x^k + \sigma \cdot w^k).$$

Тогда последовательность элементов $\{x^k\}$ будет сходиться к x^*.

Оптимальная величина шага σ_k по направлению w^k в этом случае может быть найдена из уравнения

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

Это правило гарантирует наибольшее увеличение значение функции по направлению w^k, что иллюстрирует следующий рисунок.



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

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

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

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

$$\overset{\wedge}{\mathrm{Hess}}\, F(x_k) w^k = -\,\mathrm{grad}\, F(x_k)\,,$$

что в матричном и координатном представлениях есть система линейных уравнений

$$\left\| \overset{\wedge}{\mathrm{Hess}}\, F(x_k) \right\| \left\| w^k \right\| = -\left\| \mathrm{grad}\, F(x_k) \right\|$$

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

где $\left\| w^k \right\| = \begin{Vmatrix} \omega_1^k \\ \omega_2^k \\ \cdots \\ \omega_n^k \end{Vmatrix}$ – координатное представление элемента w^k .

Убедимся, что для квадратичной формы $F(x)$ стационарный (то есть такой, что $\operatorname{grad} F(x) = o$) элемент находится по методу Ньютона за одну итерацию 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 $\left\| \operatorname{H\hat{e}ss} F(x) \right\|$.

В заключение приведем пример использования схемы *градиентного подъема* с непрерывным изменением величины "шага по направлению" (*метод Коши*).

Формально поиск локального максимального элемента для функции $F(x)$ в этом случае может быть сведен к решению следующей задачи Коши: $$\frac{dx}{d\tau} = \operatorname{grad} F(x); \quad x(0) = x^0 .$$

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

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

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 $\tau \to +\infty$ асимптотически стремится к максимальному элементу $\| x^* \| = \left\| \begin{matrix} 0 \\ 0 \end{matrix} \right\|$.