§ 7.22. Локальный относительный экстремум

Пусть Ω есть открытое множество n-мерного пространства и $f, \varphi_1, \ldots, \varphi_m$ $(1 \leqslant m < n)$ — определенные на Ω функции.

Обозначим через E множество точек $\mathbf{x}$, для которых выполняются одновременно равенства (связи):

$$\varphi_j(\mathbf{x}) = 0 \qquad (j = 1, \ldots m;\ m < n). \tag{1}$$

По определению, точка $\mathbf{x}_0 \in \Omega$ есть *точка локального относительного максимума (минимума) функции* f при наличии связей (1), если $\mathbf{x}^0 \in E$ и существует $\delta > 0$ такое, что для всех $\mathbf{x} \in E$, удовлетворяющих неравенству $|\mathbf{x} - \mathbf{x}^0| = \sqrt{\sum_1^n (x_k - x_k^0)^2} < \delta$, имеет место $f(\mathbf{x}) \leqslant f(\mathbf{x}^0)$ (в случае максимума) и $f(\mathbf{x}) \geqslant f(\mathbf{x}^0)$ (в случае минимума).

Точка локального относительного максимума или минимума называется *точкой локального относительного экстремума.*

Займемся сначала выяснением вопроса о необходимых условиях, чтобы $\mathbf{x}^0$ была точкой локального относительного экстремума.

Будем предполагать, что в окрестности точки $\mathbf{x}^0$ функции $f, \varphi_1, \ldots, \varphi_n$ имеют непрерывные частные производные. Больше того, будем предполагать, что в точке $\mathbf{x}^0$ ранг матрицы $\left\| \left(\frac{\partial \varphi_j}{\partial x_k} \right)_0 \right\|$ $(j = 1, \ldots, m,\ k = 1, \ldots, n)$ равен m. Таким образом, среди определителей порядка m, порождаемых этой матрицей, имеется не равный нулю. Для определенности будем считать, что это есть определитель

$$\frac{D(\varphi_1, \ldots, \varphi_m)}{D(x_1, \ldots, x_m)_0} = \begin{vmatrix} \left(\frac{\partial \varphi_1}{\partial x_1}\right)_0 & \ldots & \left(\frac{\partial \varphi_1}{\partial x_m}\right)_0 \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \left(\frac{\partial \varphi_m}{\partial x_1}\right)_0 & \ldots & \left(\frac{\partial \varphi_m}{\partial x_m}\right)_0 \end{vmatrix} \neq 0. \tag{2}$$

Мы считаем, что символ $(\quad)_0$ обозначает тот факт, что в функцию, стоящую в скобках, вместо $\mathbf{x}$ подставлено $\mathbf{x}^0$. На основании теоремы о неявных функциях существует прямоугольник

$$\Delta = \Delta' \times \Delta'', \tag{3}$$

$$\Delta' = \{ |x_j - x_j^0| < \delta,\ j = 1, \ldots, m \},$$

$$\Delta'' = \{ |x_i - x_i^0| < \sigma,\ i = m + 1, \ldots, n \}$$

и (единственные) непрерывно дифференцируемые функции $x_j = \mu_j(x_{m+1}, \ldots, x_n)$ $(j = 1, \ldots, m)$, определенные на Δ'', удовлетворяющие равенствам (1):

$$\varphi_j(\mu_1, \ldots, \mu_m, x_{m+1}, \ldots, x_n) = 0 \qquad (j = 1, \ldots, m).$$

При этом $|\mu_j - x_j^0| < \delta$ $(j = 1, \ldots, m)$.

Подставим в f вместо $x_1, \ldots, x_m$ соответствующие функции $\mu_1, \ldots, \mu_m$. Тогда f будет функцией только от $(n-m)$ переменных $x_{m+1}, \ldots, x_n$, независимых между собой:

$$f(\mu_1, \ldots, \mu_m, x_{m+1}, \ldots, x_n) = \Phi(x_{m+1}, \ldots, x_n).$$

Очевидно, что если f достигает локального максимума или минимума (относительного) в $\mathbf{x}^0 = (x_1^0, \ldots, x_n^0)$, то $\Phi(x_{m+1}, \ldots, x_n)$ достигает в точке $(x_{m+1}^0, \ldots, x_n^0)$ локального абсолютного максимума (минимума). Но тогда, как мы знаем, точка $(x_{m+1}^0, \ldots, x_n^0)$ должна быть стационарной для функции Φ, т. е. выполняются равенства

$$\left(\frac{\partial \Phi}{\partial x_l}\right)_0 = 0 \qquad (l = m+1, \ldots, n). \tag{4}$$

Здесь знак $(\quad)_0$ теперь уже означает, что в функции, стоящей в скобках, полагается $x_{m+1} = x_{m+1}^0, \ldots, x_n = x_n^0$. Мы думаем, что такое двоякое обозначение к путанице не приведет.

Равенства (4) эквивалентны одному равенству

$$d\Phi = \sum_{j=m+1}^{n} \left(\frac{\partial \Phi}{\partial x_j}\right)_0 dx_j = 0, \tag{4'}$$

которое должно быть верным для произвольных (независимых между собой) $dx_{m+1}, \ldots, dx_n$. В самом деле, из (4) следует (4′) при любых $dx_{m+1}, \ldots, dx_n$. Обратно, если верно (4′) для любых дифференциалов $dx_{m+1}, \ldots, dx_n$, то, в частности, оно верно, когда один из этих дифференциалов равен 1, а остальные равны нулю, а это приводит к равенствам (4).

По определению точка $x^0 \in E$ называется *стационарной точкой функции f при наличии связей* (1), если для нее выполняются равенства (4), или, что все равно, как мы выяснили, если выполняется одно равенство (4′) для любых независимых между собой dx_l $(l = m+1, \ldots, n)$.

Это определение, очевидно, эквивалентно следующему:

Точка $\mathbf{x}^0 \in E$ называется *стационарной точкой f при наличии связей* (1), если для нее полный дифференциал

$$df = \sum_{k=1}^{n} \left(\frac{\partial f}{\partial x_k}\right)_0 dx_k = 0 \tag{5}$$

для всех dx_k, удовлетворяющих линейным связям

$$\sum_{k=1}^{n}\left(\frac{\partial\varphi_j}{\partial x_k}\right)_0 dx_k = 0 \qquad (j = 1, \ldots, m). \tag{6}$$

В самом деле, в силу инвариантного свойства дифференциала $df = d\Phi$, где входящие в df (зависимые) дифференциалы $dx_1, \ldots$ $\ldots, dx_m$ соответственно равны

$$dx_k = \sum_{j=m+1}^{n}\frac{\partial\mu_k}{\partial x_j}dx_j \qquad (k = 1, \ldots, m).$$

Но последние вместе с независимыми дифференциалами $dx_{m+1}, \ldots$ $\ldots, dx_n$ связаны соотношениями (6).

Приведенные рассуждения попутно дают способ отыскания стационарной точки. Он сводится к решению $n-m$ уравнений (4) относительно $(x_{m+1}, \ldots, x_n)$. Однако предварительно надо было еще решить уравнения (1) относительно $x_1, \ldots, x_m$. Этот способ в сколько-нибудь сложных случаях является неудобным. Более простым является способ, называемый *методом множителей Лагранжа.*

Метод множителей Лагранжа (отыскания стационарных точек) заключается в том, что вводится вспомогательная функция

$$F(x) = f(\mathbf{x}) - \sum_{1}^{m}\lambda_i\varphi_j(\mathbf{x}) \tag{7}$$

от независимых переменных $x = (x_1, \ldots, x_n)$, где λ_j — постоянные числа (множители Лагранжа), подлежащие определению вместе с координатами неизвестной стационарной точки x^0. Сначала задаются произвольные числа $\lambda_1, \ldots, \lambda_m$ и для соответствующей им функции $F(x)$ от независимых переменных (не связанных связями (1)) решается задача на абсолютный экстремум. Точнее, приравниваются к нулю все ее частные производные:

$$\frac{\partial F}{\partial x_k} = \frac{\partial f}{\partial x_k} - \sum_{j=1}^{n}\lambda_j\frac{\partial\varphi_j}{\partial x_k} = 0 \qquad (k = 1, \ldots, n). \tag{8}$$

К полученной системе (8) из n уравнений присоединяется еще система (1) из m уравнений связи. Совокупная система из $n+m$ уравнений решается затем относительно $n+m$ неизвестных $x_1, \ldots$ $\ldots, x_n, \lambda_1, \ldots, \lambda_m$. Оказывается, что каждому решению $x_1^0, \ldots, x_n^0,$ $\lambda_1^0, \ldots, \lambda_m^0$ соответствует стационарная точка $\mathbf{x}^0 = (x_1^0, \ldots, x_n^0)$ задачи и, наоборот, если $\mathbf{x}^0$ есть стационарная точка, то ей соответствует единственная система множителей $\lambda_1^0, \ldots, \lambda_m^0$ такая, что составленные для них уравнения (8) удовлетворяются при $\mathbf{x} = \mathbf{x}^0$.

Введем n-мерные векторы

$$(\operatorname{grad} f)_0 = \left\{\left(\frac{\partial f}{\partial x_1}\right)_0, \ldots, \left(\frac{\partial f}{\partial x_n}\right)_0\right\}, \tag{9}$$

$$(\operatorname{grad} \varphi_j)_0 = \left\{\left(\frac{\partial \varphi_j}{\partial x_1}\right)_0, \ldots, \left(\frac{\partial \varphi_j}{\partial x_n}\right)_0\right\} \qquad (j = 1, \ldots, m). \tag{10}$$

Факт, что $\mathbf{x}^0$ есть стационарная точка, можно выразить так. Из того, что вектор $(dx_1, \ldots dx_n)$ ортогонален ко всем векторам $(\operatorname{grad} \varphi_i)_0$ $(i = 1, \ldots, m)$, т. е. из того, что удовлетворяется система (6), следует, что удовлетворяется равенство (5), т. е. вектор $(dx_1, \ldots, dx_n)$ ортогонален к вектору $(\operatorname{grad} f)_0$. Но тогда, как известно из линейной алгебры (см. ниже лемму), существует m чисел $\lambda_1, \ldots, \lambda_m$ таких, что

$$(\operatorname{grad} f)_0 = \lambda_1 (\operatorname{grad} \varphi_1)_0 + \ldots + \lambda_m (\operatorname{grad} \varphi_m)_0. \tag{11}$$

Иначе говоря, при $\mathbf{x} = \mathbf{x}^0$ выполняются равенства (1) и (8) и мы доказали наше утверждение в одну сторону. Наоборот, если при $\mathbf{x} = \mathbf{x}^0$ при некоторых числах $\lambda_1, \ldots, \lambda_m$ выполняются равенства (8), т. е. векторное равенство $(\operatorname{grad} f)_0 = \sum\limits_{i=1}^{m} \lambda_i (\operatorname{grad} \varphi_i)_0$, то из того, что выполняются равенства (6), т. е. из того, что вектор $(dx_1, \ldots \ldots, dx_n)$ ортогонален к векторам (10), следует, очевидно, что он ортогонален к вектору (9), т. е. что выполняется равенство (5), и мы доказали утверждение в обратную сторону.

Выяснение вопроса о том, будет ли данная стационарная точка $\mathbf{x}^0$ точкой относительного экстремума и какого (максимума или минимума), тоже удобно проводить, рассматривая лагранжеву функцию F. Будем считать, что в точке $\mathbf{x}^0$ якобиан (2) не равен нулю. Тогда в силу связей (1) можно считать, что переменные $x_{m+1}, \ldots, x_n$ в окрестности $\mathbf{x}^0$ независимы (между собой), а переменные $x_1, \ldots, x_m$ от них зависят. Для симметрии можно считать, что все переменные $x_1, \ldots, x_n$ — зависимые (от $x_{m+1}, \ldots, x_n$). На основании теории локального абсолютного экстремума достаточные условия максимума можно получить, исследуя второй дифференциал $d^2 f_0$ в точке $\mathbf{x}^0$, считая $x_{m+1}, \ldots, x_n$ независимыми. Мы знаем, что $f(\mathbf{x}) = f(\mathbf{x}) - \sum\limits_{k=1}^{m} \lambda_k \varphi_k(\mathbf{x}) = F(\mathbf{x})$ для всех $\mathbf{x}$, удовлетворяющих связям (1). Поэтому для этих $\mathbf{x}$ $df = dF = \sum\limits_{k=1}^{n} \frac{\partial F}{\partial x_k} dx_k$ для всех dx_k, удовлетворяющих связям (6), и

$$d^2 f = d^2 F = \sum_{k=1}^{n} \sum_{l=1}^{n} \frac{\partial^2 F}{\partial x_k \partial x_l} dx_k dx_l + \sum_{k=1}^{n} \frac{\partial F}{\partial x_k} d^2 x_k$$

для всех dx_k, $d^2 x_k$, удовлетворяющих связям (6) и вытекающим из них связям после их дифференцирования*). Подставим в эти равенства стационарную точку $\mathbf{x}^0$ и будем считать, что λ_i — ее множители Лагранжа. Тогда $\left(\frac{\partial F}{\partial x_k}\right)_0 = 0$, и мы получим равенство

$$d^2f_0 = \sum_{k=1}^{n}\sum_{l=1}^{n}\left(\frac{\partial^2 F}{\partial x_k \partial x_l}\right)_0 dx_k dx_l,$$

верное для всех dx_k, подчиняющихся связям (6).

Полученный результат можно сформулировать так. Пусть надо вычислить второй дифференциал d^2f_0 от функции f в ее стационарной точке $\mathbf{x}^0$ при наличии связей $\varphi_j = 0$ $(j = 1, \ldots, m)$. Для этого определяется лагранжева функция $F = f - \sum_{j=1}^{n} \lambda_j\varphi_j$ и вычисляется ее второй дифференциал $d^2F_0 = \sum_{k=1}^{n}\sum_{l=1}^{n}\left(\frac{\partial^2 F}{\partial x_k \partial x_l}\right)_0 dx_k dx_l$, формально считая, что $x_1, \ldots, x_n$ — независимы. Тогда имеет место равенство

$$d^2f_0 = d^2F_0, \tag{12}$$

справедливое, каковы бы ни были dx_k $(k = 1, \ldots, n)$, подчиняющиеся линейным связям (6). В этом смысле исследование d^2f_0 может быть сведено к исследованию d^2F_0.

Допустим, что d^2F_0 есть строго положительная форма, т. е. $d^2F_0 > 0$, каковы бы ни были (независимые между собой) dx_k, не равные нулю одновременно. Зададим $dx_{m+1}, \ldots, dx_n$, одновременно не равные нулю; через связи (6) им соответствуют вполне определенные значения $dx_1, \ldots, dx_m$; получим систему $dx_1, \ldots, dx_n$ дифференциалов, одновременно не равных нулю; им соответствует $d^2F_0 > 0$, но тогда и $d^2f_0 > 0$, т. е. квадратичная форма d^2f_0 от переменных $dx_{m+1}, \ldots, dx_n$ (которая явно нигде не писалась) строго положительная. В таком случае f, как функция независимых между собой переменных $x_{m+1}, \ldots, x_n$, достигает в $\mathbf{x}^0$ локального абсолютного минимума или, что все равно, f, как функция от $x_1, \ldots, x_n$, достигает в $\mathbf{x}^0$ локального относительного минимума при наличии связей (1). Подобным образом можно заключить, что если d^2F_0 есть отрицательно определенная форма, то f имеет в $\mathbf{x}^0$ локальный относительный максимум.

Но могут быть более сложные случаи, когда d^2F_0 не есть определенная форма, но она делается определенной, если дифференциалы $dx_1, \ldots, dx_n$ подчиняются связям (6). В этом случае применение равенства (12) также удобно при этом методе.

---

*) Имеются в виду связи $\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2\varphi_k}{\partial x_i \partial x_j} dx_i dx_j + \sum_{i=1}^{n} \frac{\partial\varphi_k}{\partial x_i} d^2x_i = 0$ $(k=1, \ldots, m)$.

Схема решения задачи на относительный экстремум на области Ω сводится к следующему.

Выделяется на Ω подмножество Ω_1 точек $\mathbf{x}$, в которых функции $f, \varphi_1, \ldots, \varphi_m$ имеют непрерывные частные производные, а матрица $\left\| \dfrac{\partial \varphi_j}{\partial x_k} \right\|$ имеет ранг m. На Ω_1 описанным выше способом находятся стационарные точки. Каждая из них затем исследуется на экстремум. Если в ней существуют непрерывные частные производные второго порядка, то может оказаться эффективным метод исследования второго дифференциала функции Лагранжа F. Если теория приводит к сомнительному случаю, то требуется специальное исследование. Конечно, и точки $\mathbf{x} \in \Omega - \Omega_1$ требуют также специального исследования.

Пример 1. Найдем локальные экстремумы функции $f(x, y) = xy$ на окружности (Г):

$$\varphi(x, y) = x^2 + y^2 - 1 = 0. \tag{13}$$

Функции f и φ дважды непрерывно дифференцируемы на всей плоскости. Кроме того, ранг матрицы

$$\left\| \frac{\partial \varphi}{\partial x}, \frac{\partial \varphi}{\partial y} \right\| = \| 2x, 2y \|$$

равен 1 (т. е. количеству связей) на всей плоскости x, y, за исключением точки $(0, 0)$. Но последняя не находится на Г. Следовательно, точки, где возможен локальный экстремум, находятся только среди стационарных точек.

Приравнивая нулю частные производные функции Лагранжа задачи $F(x, y) = xy - \lambda(x^2 + y^2 - 1)$, получим уравнения

$$\frac{\partial F}{\partial x} = y - 2\lambda x = 0; \quad \frac{\partial F}{\partial y} = x - 2\lambda y = 0. \tag{14}$$

Решая их вместе с уравнением (13), получим четыре пары стационарных точек $x = \pm 1/\sqrt{2}$, $y = \pm 1/\sqrt{2}$, соответствующих всевозможным распределениям «+» и «—». Паре $x_1 = y_1 = 1/\sqrt{2}$ соответствуют $\lambda_1 = 1/2$ и лагранжева функция

$$F(x, y) = xy - (x^2 + y^2 - 1)/2.$$

Второй дифференциал от F в точке (x_1, y_1) имеет вид $d^2F = -dx^2 + 2dx\,dy - dy^2 = -(dx - dy)^2$. В силу (13)

$$2x\,dx + 2y\,dy = 0,$$

откуда $dy = -dx$, и окончательно

$$d^2F = -(2dx)^2 = -4dx^2,$$

где dx — независимый дифференциал. Следовательно, в точке (x_1, y_1) имеет место локальный относительный максимум задачи, равный $f(1/\sqrt{2}, 1/\sqrt{2}) = 1/2$. Легко заключить, используя симметрические свойства f, что в точке $(-1/\sqrt{2}, -1/\sqrt{2})$ имеет место другой локальный относительный максимум, равный 1/2.

Так как окружность Г есть ограниченное замкнутое множество и непрерывная на Г функция f должна достигать на Г своего максимума, и так как максимум на Г необходимо есть локальный максимум на Г, то $\max_{\Gamma} F =$
$= f(1/\sqrt{2}, 1/\sqrt{2}) = f(-1/\sqrt{2}, -1/\sqrt{2}) = 1/2$ и, аналогично,

$$\min_{\Gamma} f = f(1/\sqrt{2}, -1/\sqrt{2}) = f(-1/\sqrt{2}, 1/\sqrt{2}) = -1/2.$$

Лемма. *Пусть* $\mathbf{a}$, $\mathbf{b}^1, \ldots, \mathbf{b}^m$ — *векторы n-мерного пространства* ($m < n$). *Для того чтобы имело место представление*

$$\mathbf{a} = \sum_{j=1}^{m} \lambda_j \mathbf{b}^j, \tag{15}$$

где λ_j — *некоторые числа, необходимо и достаточно, чтобы всякий вектор* $\mathbf{c}$, *ортогональный ко всем* $\mathbf{b}^j$:

$$(\mathbf{c}, \mathbf{b}^j) = 0 \quad (j = 1, \ldots, m), \tag{16}$$

автоматически был ортогонален к $\mathbf{a}$:

$$(\mathbf{c}, \mathbf{a}) = 0. \tag{17}$$

Доказательство. Если имеет место (15), то (16) влечет

$$(\mathbf{c}, \mathbf{a}) = \left(\mathbf{c}, \sum_{j=1}^{m} \lambda_j \mathbf{b}^j\right) = \sum_{j=1}^{m} \lambda_j (\mathbf{c}, \mathbf{b}^j) = 0,$$

и мы доказали необходимость условия леммы.

Перейдем к доказательству достаточности. Ортогонализируя систему $\mathbf{b}^1, \ldots, \mathbf{b}^m$, получим ортонормированную систему $\mathbf{a}^1, \ldots, \mathbf{a}^m$, обладающую тем же свойством: из равенств

$$(\mathbf{c}, \mathbf{a}^j) = 0, \quad j = 1, \ldots, m,$$

следует, что $(\mathbf{c}, \mathbf{a}) = 0$.

Разложим вектор $\mathbf{a}$ по векторам $\mathbf{a}^j$:

$$\mathbf{a} = \sum_{j=1}^{m} \lambda_j \mathbf{a}^j + \mathbf{r}, \qquad \lambda_j = (\mathbf{a}, \mathbf{a}^j).$$

Вектор $\mathbf{r}$ ортогонален ко всем $\mathbf{a}^j$, но тогда $(\mathbf{r}, \mathbf{a}) = 0$ и, следовательно,

$$0 = (\mathbf{r}, \mathbf{a}) = \left(\mathbf{r}, \sum_{j=1}^{m} \lambda_j \mathbf{a}^j + \mathbf{r}\right) = (\mathbf{r}, \mathbf{r}).$$

Но тогда

$$\mathbf{r} = 0 \text{ и } \mathbf{a} = \sum_{j=1}^{m} \lambda_j \mathbf{a}^j.$$