1 Аппарат теории экстремума

Данный раздел содержит те факты, которые относятся к базе теории экстремума: дифференциальные свойства функций и отображений, выпуклые множества и теоремы отделимости.

1.1 Дифференциальные свойства функций и отображений

В этом параграфе напоминаются необходимые для дальнейшего факты, связанные с пространством $\mathbb{R}^n$, с понятиями дифференцируемости функций и отображений на $\mathbb{R}^n$. Здесь также много упражнений, которые полезно сделать.

Пусть n — натуральное число. Пространство $\mathbb{R}^n$ — это совокупность всех упорядоченных наборов $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ из n действительных чисел (если $n = 1$, то это просто множество действительных чисел, и мы пишем $\mathbb{R}$ вместо $\mathbb{R}^1$), называемых *векторами* или *вектор-столбцами*, а числа x_i, $1 \le i \le n$ — *координатами вектора* x. Ради экономии места, элементы $\mathbb{R}^n$ будем записывать так $x = (x_1, \ldots, x_n)^T$, где ^{T} обозначает *транспонирование* (перевод строки в столбец и наоборот). В $\mathbb{R}^n$ естественным образом вводится операция (покоординатного) сложения векторов и операция (покоординатного) умножения вектора на число, превращающие это множество в вещественное векторное пространство.

Пусть $a = (a_1, \ldots, a_n)$ — вектор-строка из n действительных чисел. Для каждого $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ обозначим $\langle a, x \rangle = \sum_{i=1}^n a_i x_i$. Это матричное произведение вектор-строки a на вектор-столбец x, которое иногда называют *внутренним произведением.* Ясно, что отображение $x \mapsto \langle a, x \rangle$ есть линейная функция (или говорят, линейный функционал) на $\mathbb{R}^n$. Легко видеть, что и любой линейный функционал l на $\mathbb{R}^n$ задается подобным образом с $a = (l(e_1), \ldots, l(e_n))$, где $e_1 = (1, 0, \ldots, 0)^T, \ldots, e_n = (0, \ldots, 0, 1)^T$ — *стандартный базис* в $\mathbb{R}^n$. Таким образом, если обозначить через $(\mathbb{R}^n)^*$ множество, элементы которого суть те же наборы из n действительных чисел, но расположенные в строку (с аналогичными операциями сложения и умножения на числа), то его можно отождествить с пространством всех линейных функционалов на $\mathbb{R}^n$, которое называют *сопряженным* или *двойственным* пространством к $\mathbb{R}^n$.

Упражнение 1.1. *Доказать, что отображение, сопоставляющее $a = (a_1, \ldots, a_n) \in (\mathbb{R}^n)^*$ линейный функционал $x \mapsto \langle a, x \rangle$ на $\mathbb{R}^n$ является изоморфизмом $(\mathbb{R}^n)^*$ и двойственного к $\mathbb{R}^n$ (с естественными операциями поточечного сложения функций и умножения их на числа).*

Каждому $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ можно сопоставить линейный функционал $a \mapsto \langle a, x \rangle$ на $(\mathbb{R}^n)^*$ и тогда совершенно аналогично устанавливается, что сопряженное (двойственное) пространство к $(\mathbb{R}^n)^*$ можно отождествить с $\mathbb{R}^n$ (говорят еще, что второе сопряженное к $\mathbb{R}^n$ совпадает с ним самим, т. е. $((\mathbb{R}^n)^*)^* = (\mathbb{R}^n)^{**} = \mathbb{R}^n$).

В последующем, как правило, элементы $(\mathbb{R}^n)^*$ будем обозначать x^*, y^* и т. д.

Далее мы говорим о свойствах пространства $\mathbb{R}^n$, но все без исключения переносится на пространство $(\mathbb{R}^n)^*$.

Пусть $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$. Величина $|x| = \sqrt{x_1^2 + \ldots + x_n^2}$ (или короче, $|x| = \sqrt{\langle x^T, x \rangle}$) называется *длиной* или *модулем* или *евклидовой нормой* вектора

x. Длина вектора удовлетворяет следующим свойствам: $|x| \geq 0$ и $|x| = 0$ тогда и только тогда, когда $x = 0$, $|\alpha x| = |\alpha||x|$ для любых $x \in \mathbb{R}$ и $\alpha \in \mathbb{R}$, $|x+y| \leq |x| + |y|$ для любых $x, y \in \mathbb{R}$. Первые два свойства очевидны, третье (называемое неравенством треугольника) устанавливается с помощью неравенства Коши–Буняковского: $|\langle x^T, y\rangle| \leq |x||y|$, проверка которого достаточно проста. Величина $d(x,y) = |x-y|$ называется *расстоянием* между векторами x и y.

Пусть $\widehat{x} \in \mathbb{R}^n$ и $r > 0$. Множество

$$U_{\mathbb{R}^n}(\widehat{x}, \delta) = \{\, x \in \mathbb{R}^n : |x - \widehat{x}| < r \,\}$$

называется *открытым* шаром с центром в точке $\widehat{x}$ радиуса r.

Множество

$$B_{\mathbb{R}^n}(\widehat{x}, \delta) = \{\, x \in \mathbb{R}^n : |x - \widehat{x}| \leq r \,\}$$

называется *замкнутым* шаром с центром в точке $\widehat{x}$ радиуса r.

Пусть $A \subset \mathbb{R}^n$ и $x \in A$. Говорят, что x — *внутренняя точка* A, если x входит в A вместе с некоторым шаром с центром в этой точке.

Множество внутренних точек A называется *внутренностью* A и обозначается $\operatorname{int} A$.

Множество $G \subset \mathbb{R}^n$ называется *открытым*, если каждая его точка внутренняя, т. е. $\operatorname{int} G = G$.

Упражнение 1.2. *Доказать, что открытый шар — открытое множество.*

Упражнение 1.3. *Доказать, что объединение любого и пересечение конечного числа открытых множеств — открытое множество.*

Окрестностью точки называется любое открытое множество, содержащее данную точку.

Множество $F \subset \mathbb{R}^n$ называется *замкнутым*, если его дополнение $\mathbb{R}^n \setminus F$ — открытое множество.

Упражнение 1.4. *Доказать, что замкнутый шар — замкнутое множество.*

Упражнение 1.5. *Доказать, что пересечение любого числа и объединение конечного числа замкнутых множеств — замкнутое множество.*

Упражнение 1.6. *Доказать, что подпространства в $\mathbb{R}^n$ — замкнутые множества.*

Пусть $A \subset \mathbb{R}^n$. Точка $x \in \mathbb{R}^n$ называется *предельной точкой множества* A, если каждая ее окрестность имеет непустое пересечение с A. Совокупность всех предельных точек множества A называется *замыканием* A и обозначается $\mathrm{cl\,A}$. Ясно, что $A \subset \mathrm{cl\,A}$.

Упражнение 1.7. *Доказать, что* $\mathrm{cl\,A}$ *— замкнутое множество и что множество A замкнуто тогда и только тогда, когда оно совпадает со своим замыканием, т. е.* $A = \mathrm{cl\,A}$.

Множество $A \subset \mathbb{R}^n$ называется *компактным* (или *компактом*), если из любой последовательности его элементов можно выделить подпоследовательность, сходящуюся к элементу из A.

Упражнение 1.8. *Доказать, что множество в $\mathbb{R}^n$ компактно тогда и только тогда, когда оно ограничено (т. е. содержится в некотором шаре) и замкнуто.*

Упражнение 1.9. *Доказать, что если K — компакт, а F — замкнутое множество, то их алгебраическая сумма $K + F = \{\, z = x + y : x \in K,\ y \in F \,\}$ — замкнутое множество.*

Пусть $A \subset \mathbb{R}^n$ и $x_0 \in A$. Функция $f : A \to \mathbb{R}$ называется *непрерывной в точке* $x_0 \in A$, если для любого $\varepsilon > 0$ найдется такое $\delta = \delta(\varepsilon) > 0$, что $|f(x) - f(x_0)| < \varepsilon$ для всех $x \in A \cap U_{\mathbb{R}^n}(x_0, \delta)$, или равносильно: для любой последовательности векторов $\{x_k\}_{k\in\mathbb{N}} \subset A$, сходящейся к x_0, последовательность $\{f(x_k)\}_{k\in\mathbb{N}}$ сходится к $f(x_0)$.

Если функция непрерывна в каждой точке множества A, то говорят, что она *непрерывна на* A.

Упражнение 1.10. *Доказать, что функции $x \to |x|$ и $x \to \langle x^*, x\rangle + \alpha$, где $x^* \in (\mathbb{R}^n)^*$ и $\alpha \in \mathbb{R}$, непрерывны на $\mathbb{R}^n$.*

Теорема (Вейерштрасса)**.** *Функция непрерывная на компакте в $\mathbb{R}^n$ достигает на нем своего максимального и минимального значений.*

Доказательство. Пусть A — компактное подмножество $\mathbb{R}^n$ и $f : A \to \mathbb{R}$ — непрерывная функция. Обозначим $\gamma = \sup\{\, f(x) : x \in A \,\}$. По определению верхней грани (конечной или бесконечной) существует последовательность $\{x_k\}$ элементов из A такая, что $\{f(x_k)\}$ сходится к γ. Из $\{x_k\}$ можно выделить подпоследовательность $\{x_{k_l}\}$, которая сходится к $\overline{x} \in A$. Тогда последовательность $\{f(x_{k_l})\}$ сходится к $f(\overline{x}) = \gamma$, т. е. γ конечно и тем самым $\overline{x}$ — точка, где f принимает свое максимальное значение. Аналогичные рассуждения относительно минимума функции. $\square$

В конце этого пункта напомним еще понятие непрерывного отображения. Пусть A — подмножество $\mathbb{R}^n$ и задано отображение $F\colon A \to \mathbb{R}^m$.

Отображение $F\colon A \to \mathbb{R}^m$ называется *непрерывным в точке* $\widehat{x} \in A$, если для любого $\varepsilon > 0$ найдется такое $\delta = \delta(\varepsilon) > 0$, что $|F(x) - F(\widehat{x})| < \varepsilon$ для всех $x \in A \cap U_{\mathbb{R}^n}(\widehat{x}, \delta)$, или равносильно: для любой последовательности $\{x_k\}$ точек из A, сходящейся к $\widehat{x}$ последовательность $\{F(x_k)\}$ сходится к $F(\widehat{x})$.

Говорят, что отображение F *непрерывно на* A, если оно непрерывно в каждой точке A.

Упражнение 1.11. *Доказать, что отображение $F\colon \mathbb{R}^n \to \mathbb{R}^m$ непрерывно тогда и только тогда, когда прообраз любого открытого множества открыт.*

Упражнение 1.12. *Доказать, что если G — открытое множество, то для любых $y \in \mathbb{R}^n$ и $\alpha \in \mathbb{R}$, $\alpha \neq 0$, множество*

$$y + \alpha G = \{\, z \in \mathbb{R}^n : z = y + \alpha x,\ x \in G \,\}$$

также открыто.

Пусть $\Lambda\colon \mathbb{R}^n \to \mathbb{R}^m$ — линейный оператор. Мы будем отождествлять этот оператор с его матрицей в стандартных базисах $e_1 = (1, 0, \ldots, 0)^T, \ldots, e_n = (0, \ldots, 0, 1)^T$ и $e'_1 = (1, 0, \ldots, 0)^T, \ldots, e'_m = (0, \ldots, 0, 1)^T$ в $\mathbb{R}^n$ и $\mathbb{R}^m$ соответственно, т. е. если $\Lambda e_j = \sum_{i=1}^m a_{ij} e'_i$, $j = 1, \ldots, n$, то матрицей оператора Λ называется матрица (мы ее обозначаем той же буквой) $\Lambda = (a_{ij})_{1\le i\le n,\, 1\le j\le m}$ размера $m \times n$. В этом случае Λx — произведение матрицы Λ на вектор x.

Обозначим через $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ множество всех линейных операторов из $\mathbb{R}^n$ в $\mathbb{R}^m$. Это векторное пространство с обычными (поэлементными) операциями сложения матриц и умножения их на числа. Каждому $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ сопоставим число

$$\|\Lambda\| = \sup_{|x|\le 1} |\Lambda x|, \text{ [1]}$$

[1] На самом деле здесь можно поставить max, так как по теореме Вейерштрасса непрерывная функция $x \to |\Lambda x|$ достигает своего максимума на шаре $B_{\mathbb{R}^n}(0, 1)$.

которое называется нормой оператора Λ.

Простая проверка показывает, что выполняются следующие условия: (a) $\|\Lambda\| \geq 0$ для любого $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ и $\|\Lambda\| = 0$ тогда и только тогда, когда $\Lambda = 0$; (b) $\|\alpha\Lambda\| = |\alpha|\|\Lambda\|$ для любых $\alpha \in \mathbb{R}$ и $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$; (c) $\|\Lambda_1 + \Lambda_2\| \leq \|\Lambda_1\| + \|\Lambda_2\|$ для любых $\Lambda_1, \Lambda_2 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Ясно также, что $|\Lambda x| \leq \|\Lambda\||x|$ для любого $x \in \mathbb{R}^n$.

Данное определение нормы в $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ позволяет (также как и в $\mathbb{R}^n$) ввести этом пространстве понятия открытого и замкнутого множества, окрестности точки и т. д.

Упражнение 1.13. *Доказать, что оператор $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ есть непрерывное отображение из $\mathbb{R}^n$ в $\mathbb{R}^m$.*

Теперь мы можем перейти к понятию дифференцируемости функций и отображений.

Определение. *Пусть U — окрестность точки $\widehat{x} \in \mathbb{R}^n$. Говорят, что функция $f : U \to \mathbb{R}$ дифференцируема в точке $\widehat{x}$, если существует линейный функционал на $\mathbb{R}^n$, т. е. вектор $a = (a_1, \ldots, a_n) \in (\mathbb{R}^n)^*$ такой, что для всех $h \in \mathbb{R}^n$, для которых $\widehat{x} + h \in U$ справедливо представление*

$$f(\widehat{x} + h) = f(\widehat{x}) + \langle a, h \rangle + r(h),$$

где $|r(h)|/|h| \to 0$ при $h \to 0$, т. е. $r(h) = o(h)$. Вектор a называется производной функции f в точке $\widehat{x}$ и обозначается $f'(\widehat{x})$.

Упражнение 1.14. *Доказать, что производная определена однозначно.*

Из данного определения легко следует (беря в качестве h векторы $(h_1, 0, \ldots, 0)^T, \ldots, (0, \ldots, 0, h_n)^T$), что a_i есть частная производная функции f по x_i в точке $\widehat{x}$, $1 \leq i \leq n$. Таким образом,

$$f'(\widehat{x}) = \left(\frac{\partial f}{\partial x_1}(\widehat{x}), \ldots, \frac{\partial f}{\partial x_n}(\widehat{x}) \right).$$

В классическом анализе анализе обозначают $h = (dx_1, \ldots, dx_n)^T$ и тогда

$$\langle f'(\widehat{x}),\, h \rangle = \frac{\partial f}{\partial x_1}(\widehat{x})dx_1 + \ldots + \frac{\partial f}{\partial x_n}(\widehat{x})dx_n.$$

Это выражение называется дифференциалом f и обозначается $df(\widehat{x})$.

Легко проверить, что функция, дифференцируемая в данной точке, непрерывна в этой точке.

Существование частных производных функции в данной точке не гарантирует ее дифференцируемость и даже ее непрерывность в данной точке.

Упражнение 1.15. *Постройте функцию, у которой существуют частные производные в точке, но сама функция в этой точке разрывна.*

Упражнение 1.16. *Пусть A — симметричная матрица размера $n \times n$, $b \in (\mathbb{R}^n)^*$, $c \in \mathbb{R}$ и функция $f\colon \mathbb{R}^n \to \mathbb{R}$ определена по правилу:*

$$f(x) = \langle x^T,\, Ax \rangle + \langle b,\, x \rangle + c.$$

Найдите ее производную для каждого $x \in \mathbb{R}^n$.

Определение. *Пусть U — окрестность точки $\widehat{x} \in \mathbb{R}^n$. Говорят, что отображение $F\colon U \to \mathbb{R}^m$ дифференцируемо в точке $\widehat{x}$, если существует линейный оператор из $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, т. е. матрица Λ размера $m \times n$, такая, что для всех $h \in \mathbb{R}^n$, для которых $\widehat{x} + h \in U$, справедливо представление*

$$F(\widehat{x} + h) = F(\widehat{x}) + \Lambda h + r(h),$$

где $|\rho(h)|/|h| \to 0$ при $h \to 0$, т. е. $\rho(h) = o(h)$. Матрица Λ называется производной отображения F в точке $\widehat{x}$ и обозначается $F'(\widehat{x})$.

Упражнение 1.17. *Доказать, что производная определена однозначно.*
Упражнение 1.18. *Пусть $\widehat{x} \in \mathbb{R}^n$. Чему равна производная отображения $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ в данной точке?*

Если на U определены функции f_i, $i = 1, \ldots, m$, и отображение $F\colon U \to \mathbb{R}^m$ определено по правилу: $F(x) = (f_1(x), \ldots, f_m(x))^T$, то легко проверить, что F дифференцируемо в точке $\widehat{x}$ тогда и только тогда, когда функции f_i, $1 \le i \le m$, дифференцируемы в $\widehat{x}$. При этом строки матрицы $F'(\widehat{x})$ суть векторы $f_1'(\widehat{x}), \ldots, f_m'(\widehat{x})$. Производную $F'(\widehat{x})$ называют *матрицей Якоби* отображения F в точке $\widehat{x}$.

Если отображение F дифференцируемо в каждой точке U, то говорят, что F *дифференцируемо на U*.

Пусть W — окрестность точки $(\widehat{x}, \widehat{y}) \in \mathbb{R}^k \times \mathbb{R}^m$ и $F\colon W \to \mathbb{R}^n$. Если отображение $x \mapsto F(x, \widehat{y})$, определенное на открытом множестве $\{x \in \mathbb{R}^k : (x, y) \in W\}$, дифференцируемо в точке $\widehat{x}$, то соответствующую производную называют *частной производной* отображения F по x в точке $(\widehat{x}, \widehat{y})$ и обозначают $F_x(\widehat{x}, \widehat{y})$. Аналогично определяется частная производная F по y в точке $(\widehat{x}, \widehat{y})$, которую обозначаем $F_y(\widehat{x}, \widehat{y})$.

Легко проверить, что частные производные определены однозначно.

Предложение 1. *Если отображение $F\colon W \to \mathbb{R}^n$ дифференцируемо в точке $(\widehat{x}, \widehat{y})$, то оно имеет в этой точке частные производные по x и y, и для всех $\xi \in \mathbb{R}^k$ и $\eta \in \mathbb{R}^m$ справедливо представление*

$$F'(\widehat{x}, \widehat{y})[\xi, \eta] = F_x(\widehat{x}, \widehat{y})[\xi] + F_y(\widehat{x}, \widehat{y})[\eta]^2$$

Доказательство. Так как F дифференцируемо в точке $(\widehat{x}, \widehat{y})$, то, в частности, справедливо представление

$$F(\widehat{x} + h, \widehat{y}) = F(\widehat{x}, \widehat{y}) + F'(\widehat{x}, \widehat{y})(h, 0) + r(h),$$

где $|r(h)|/|h| \to 0$ при $h \to 0$. Отсюда следует, что существует частная производная отображения F по x в точке $(\widehat{x}, \widehat{y})$, действующая по правилу: $\xi \mapsto F'(\widehat{x}, \widehat{y})[\xi, 0]$. Аналогично устанавливается существование частной производной F по y в точке $(\widehat{x}, \widehat{y})$, действующей по правилу: $\eta \mapsto F'(\widehat{x}, \widehat{y})[0, \eta]$. Следовательно,

$$F'(\widehat{x}, \widehat{y})[\xi, \eta] = F'(\widehat{x}, \widehat{y})([\xi, 0] + [0, \eta]) = F_x(\widehat{x}, \widehat{y})[\xi] + F_y(\widehat{x}, \widehat{y})[\eta].$$

$\square$

[2] $F'(\widehat{x}, \widehat{y})[\xi, \eta]$ обозначает действие (значение) оператора $F'(\widehat{x}, \widehat{y})$ на элементе (ξ, η). Аналогично для частных производных.

Пусть функция $f\colon U \to \mathbb{R}$ дифференцируема на U. Тогда определено отображение $f'\colon U \to (\mathbb{R}^n)^*$, сопоставляющее $x \in U$ вектор $f'(x) \in (\mathbb{R}^n)^*$. Если это отображение непрерывно в точке $\widehat{x}$ (на U), то говорят, что функция f *непрерывно дифференцируема в* $\widehat{x}$ (на U).

Упражнение 1.19*. *Доказать, что непрерывная дифференцируемость функции f на U равносильно тому, что все частные производные функции f непрерывны на U.*

Аналогично, если отображение $F\colon U \to \mathbb{R}^m$ дифференцируемо на U, то определено отображение $F'\colon U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, сопоставляющее $x \in U$ матрицу $F'(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Если это отображение непрерывно в точке $\widehat{x}$ (на U), то говорят, что отображение F *непрерывно дифференцируемо в* $\widehat{x}$ (на U).

В теории экстремальных задач необходимые условия экстремума формулируются на языке производных (впервые такой факт был обнаружен П. Ферма в 1638 году, названный впоследствии теоремой Ферма и утверждающий, что если функция в точке $\widehat{x}$ достигает своего локального экстремума и дифференцируема в $\widehat{x}$, то ее производная в этой точке равна нулю). Однако в более общих ситуациях, для получения необходимых условий, одной дифференцируемости, как правило, недостаточно, а непрерывная дифференцируемость оказывается излишней. Следующее определение занимает промежуточное положение между дифференцируемостью и непрерывной дифференцируемостью и естественным образом возникает в доказательствах необходимых и достаточных условиях экстремума.

Прежде чем привести определение строгой дифференцируемости, сформулируем определение дифференцируемости на языке ε-δ (заменяя $\widehat{x} + h$ на x). Отображение $F\colon U \to \mathbb{R}^m$ дифференцируемо в точке $\widehat{x}$, если существует оператор $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ такой, что для любого $\varepsilon > 0$ найдется $\delta = \delta(\varepsilon) > 0$, обладающее тем свойством, что если $x \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$, то

$$|F(x) - F(\widehat{x}) - \Lambda(x - \widehat{x})| \le \varepsilon |x - \widehat{x}|.$$

Отсюда “напрашивается” естественное обобщение этого определения.

Говорят, что отображение $F\colon U \to \mathbb{R}^m$ *строго дифференцируемо в точке* $\widehat{x}$, если существует такой линейный оператор из $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, т. е. матрица Λ размера $m \times n$, что для любого $\varepsilon > 0$ найдется $\delta = \delta(\varepsilon) > 0$, обладающее тем свойством, что для всех $x, x' \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$ выполняется неравенство

$$|F(x) - F(x') - \Lambda(x - x')| \le \varepsilon |x - x'|.$$

Если $x' = \widehat{x}$, то получаем определение дифференцируемости F в точке $\widehat{x}$ и значит, $\Lambda = F'(\widehat{x})$. Таким образом, строго дифференцируемое отображение в $\widehat{x}$ дифференцируемо в этой точке.

Если $m = 1$, т. е. F — функция, то получаем определение строгой дифференцируемости функции в точке.

Если отображение F порождено набором функций, то простая проверка показывает, что строгая дифференцируемость F в точке $\widehat{x}$ равносильна строгой дифференцируемости в $\widehat{x}$ каждой из этих функций.

Упражнение 1.20. *Доказать, что если отображение строго дифференцируемо в точке $\widehat{x}$, то для каждого $\varepsilon > 0$ оно непрерывно в соответствующей окрестности $U_{\mathbb{R}^n}(\widehat{x}, \delta)$.*

Множество функций, строго дифференцируемых в данной точке уже множества функций, которые просто дифференцируемы в этой точке. Действительно, рассмотрим функцию $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2 D(x)$, где $D(\cdot)$ — функция Дирихле (равная единице, если x рационально и нулю, если иррационально). Легко проверить, что f дифференцируема в нуле и, очевидно, разрывна во всех остальных точках и значит, не может быть строго дифференцируемой в нуле.

Пусть U — окрестность точки $\widehat{x}$ и $F_i\colon U \to \mathbb{R}^{m_1}$, $i = 1, 2$. Рассмотрим отображение

$$F = (F_1, F_2)\colon U \to \mathbb{R}^{m_1} \times \mathbb{R}^{m_2},$$

действующее по правилу: $F(x) = (F_1(x), F_2(x))$ для всех $x \in U$.

Предложение 2. *Отображение F дифференцируемо (строго дифференцируемо) в точке $\widehat{x}$ тогда и только тогда, когда отображения F_1 и F_2 дифференцируемы (строго дифференцируемы) в точке $\widehat{x}$ и при этом $F'(\widehat{x}) = (F_1'(\widehat{x}), F_2'(\widehat{x}))$.*

Доказательство. Пусть отображения F_1 и F_2 дифференцируемы в точке $\widehat{x}$. Тогда по определению

$$F(\widehat{x} + h) = (F_1(\widehat{x} + h), F_2(\widehat{x} + h)) = (F_1(\widehat{x}), F_2(\widehat{x})) + (F_1'(\widehat{x}), F_2'(\widehat{x}))h + (r_1(h), r_2(h)),$$

где $|r_i(h)|/|h| \to 0$ при $h \to 0$, $i = 1, 2$. Следовательно, отображение F дифференцируемо в точке $\widehat{x}$ и $F'(\widehat{x}) = (F_1'(\widehat{x}), F_2'(\widehat{x}))$.

Для проверки того, что из дифференцируемости F следует дифференцируемость F_1 и F_2, надо заметить, что $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^{m_1} \times \mathbb{R}^{m_2})$ тогда и только тогда, когда $\Lambda = (\Lambda_1, \Lambda_2)$, где $\Lambda_i \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^{m_i})$, $i = 1, 2$.

Рассуждения, связанные со строгой дифференцируемостью, аналогичны. □

Теорема (о суперпозиции дифференцируемых отображений)**.** *Пусть U — окрестность точки $\widehat{x} \in \mathbb{R}^n$, $F_1\colon U \to \mathbb{R}^m$, V — окрестность точки $\widehat{y} = F_1(\widehat{x})$, $F_2\colon V \to \mathbb{R}^k$ и $F = F_2 \circ F_1\colon U \to \mathbb{R}^k$. Тогда, если F_1 дифференцируемо (строго дифференцируемо, непрерывно дифференцируемо) в точке $\widehat{x}$, F_2 дифференцируемо (строго дифференцируемо, непрерывно дифференцируемо) в точке $\widehat{y}$, то отображение F дифференцируемо (строго дифференцируемо, непрерывно дифференцируемо) в точке $\widehat{x}$ и $F'(\widehat{x}) = F_2'(\widehat{y}) \circ F_1'(\widehat{x})$.* [3]

Эту теорему доказывать не будем (см. ....).

Если отображение непрерывно дифференцируемо в точке, то оно строго дифференцируемо в этой точке. Это следствие теоремы о среднем, которую здесь докажем.

Если $x, y \in \mathbb{R}^n$, то множество

$$[x, y] = \{z \in \mathbb{R}^n : z = (1 - \alpha)x + \alpha y,\ 0 \le \alpha \le 1\}$$

называется *отрезком*, соединяющий точки x и y. При $n = 1, 2$ и 3 — это обычный отрезок.

Теорема (о среднем)**.** *Пусть U — открытое множество в $\mathbb{R}^n$, $[x, y] \subset U$ и отображение $F\colon U \to \mathbb{R}^m$ дифференцируемо в каждой точке отрезка $[x, y]$. Тогда*

$$|F(y) - F(x)| \le \sup_{z \in [x,y]} \|F'(z)\| |y - x|.$$

[3] На языке матриц — это произведение матриц.

Доказательство. Если $F(x) = F(y)$, то утверждение теоремы выполняется очевидным образом. Пусть $F(x) \neq F(y)$. Рассмотрим функцию $\varphi\colon [0,1] \to \mathbb{R}$, определенную равенством

$$\varphi(t) = \langle (F(y) - F(x))^T,\, F((1-t)x + ty \rangle.$$

Эта функция, по теореме о суперпозиции непрерывных и дифференцируемых функций, непрерывна на отрезке $[0,1]$ и дифференцируема на $(0,1)$. Следовательно, по теореме Лагранжа существует такое $0 < \theta < 1$, что $\varphi(1) - \varphi(0) = \varphi'(\theta)$, или

$$\langle (F(y) - F(x))^T,\, F(y) \rangle - \langle (F(y) - F(x))^T,\, F(x) \rangle = \langle (F(y) - F(x))^T,\, F'(z)(y - x) \rangle,$$

где $z = (1-\theta)x + \theta y$. Отсюда вытекает, что

$$|F(y) - F(x)|^2 \leq |F(y) - F(x)| \|F'(z)\| |y - x|.$$

Деля обе части этого неравенства на $|F(y) - F(x)|$, приходим к нужному утверждению. □

Следствие. *Пусть U — окрестность точки $\widehat{x} \in \mathbb{R}^n$. Если отображение $F\colon U \to \mathbb{R}^m$ непрерывно дифференцируемо в $\widehat{x}$, то оно строго дифференцируемо в $\widehat{x}$.*

Доказательство. Пусть $\varepsilon > 0$. Так как F' непрерывно в $\widehat{x}$, то найдется $\delta > 0$ такое, что из условия $x \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$ следует неравенство $\|F'(x) - F'(\widehat{x})\| < \varepsilon$. Если $x_i \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$, $i = 1, 2$, и $x \in [x_1, x_2]$, то как легко проверить $x \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$. Отображение $x \mapsto F(x) - F'(\widehat{x})x$ дифференцируемо на U и тогда по теореме о среднем для этого отображения

$$|F(x_1) - F(x_2) - F'(\widehat{x})(x_1 - x_2)| \leq \sup_{x \in [x_1, x_2]} \|F'(x) - F'(\widehat{x})\| |x_1 - x_2| \leq \varepsilon |x_1 - x_2|.$$

□

Пусть отображение $f'\colon U \to (\mathbb{R}^n)^*$ дифференцируемо на U. Тогда определено отображение $f''\colon U \to \mathcal{L}(\mathbb{R}^n, (\mathbb{R}^n)^*)$, сопоставляющее $x \in U$ линейный оператор (матрицу) $f''(x) \in \mathcal{L}(\mathbb{R}^n, (\mathbb{R}^n)^*)$, которую называем *второй производной* f *в точке* x, и говорят, что функция f *дважды дифференцируема в* $\widehat{x}$ (*на* U, *если вторая производная существует в каждой точке* U).

Легко проверить, что $f''(x) = (\partial^2 f(x)/\partial x_j \partial x_i)$, $i, j = 1, \ldots, n$. Эту матрицу называют *матрицей Гесса* или *гессианом функции* f *в точке* x.

Если отображение $f''\colon U \to \mathcal{L}(\mathbb{R}^n, (\mathbb{R}^n)^*)$ непрерывно в $\widehat{x}$ (на U), то говорят, что функция f *дважды непрерывно дифференцируема в* $\widehat{x}$ (на U). Последнее равносильно тому, что *все вторые частные производные функции* f *непрерывны на* U. Заметим еще, что непрерывность вторых частных производных в точке x влечет равенство $\partial^2 f(x)/\partial x_i \partial x_j = \partial^2 f(x)/\partial x_j \partial x_i$, $i, j = 1, \ldots, n$, откуда следует, что гессиан дважды непрерывно дифференцируемой функции f в данной точке есть симметричная матрица.

Отметим здесь еще вариант формулы Тейлора: *если функция $f\colon U \to \mathbb{R}$ дважды дифференцируема в $\widehat{x} \in U$, то для любого $h \in \mathbb{R}^n$, для которого $\widehat{x} + h \in U$, справедливо равенство*

$$f(\widehat{x} + h) = f(\widehat{x}) + \langle f'(\widehat{x}),\, h \rangle + \frac{1}{2} h^T f''(\widehat{x}) h + r(h),$$

где $|r(h)|/|h|^2 \to 0$ *при* $h \to 0$, *т. е.* $r(h) = o(|h|^2)$.

Исторический комментарий. Понятие производной функции одного переменного появилось вместе с рождением анализа. Оно принадлежит Ньютону и Лейбницу (и было опубликовано в первой работе Лейбница по анализу в 1684 г., хотя Ньютон владел этим понятием раньше). Для функций многих переменных понятие производной появилось в лекциях Вейрштрасса. Строгая дифференцируемость была введена Личем в 1961 г.

1.2 Выпуклые множества и теоремы отделимости

Выпуклые множества

Пусть $x, y \in \mathbb{R}^n$. Напомним, что множество

$$[x, y] = \{z \in \mathbb{R}^n \mid z = (1-\alpha)x + \alpha y,\ 0 \le \alpha \le 1\}$$

называется *отрезком* (соединяющим точки x и y). Определим еще интервал

$$(x, y) = \{z \in \mathbb{R}^n \mid z = (1-\alpha)x + \alpha y,\ 0 < \alpha < 1\}.$$

Непустое множество $A \subset \mathbb{R}^n$ называется *выпуклым*, если вместе с любыми двумя своими точками x и y оно содержит отрезок $[x, y]$.

Пустое множество выпукло по определению.

Следующие свойства выпуклых множеств непосредственно следуют из определений.

Если $\{A_i\}_{i\in\mathcal{J}}$ — произвольное семейство выпуклых подмножеств $\mathbb{R}^n$, то $\bigcap_{i\in\mathcal{J}} A_i$ — выпуклое множество.

Если $\{A_i\}_{i=1}^n$ — конечное семейство выпуклых подмножеств $\mathbb{R}^n$, то $A_1 + \dots + A_n = \{x \in \mathbb{R}^n : x = x_1 + \dots + x_n,\ x_i \in A_i,\ i = 1, \dots, n\}$ — выпуклое множество.

Если A — выпуклое подмножество $\mathbb{R}^n$ и $\lambda \in \mathbb{R}$, то множество $\lambda A = \{\lambda x : x \in A\}$ выпукло.

Упражнение 2.1. *Доказать, что открытый и замкнутый шары в* $\mathbb{R}^n$ *— выпуклые множества.*

Пусть A — непустое подмножество $\mathbb{R}^n$. Наименьшее выпуклое множество, содержащее A (т. е. пересечение всех выпуклых множеств, содержащих A) называется *выпуклой оболочкой множества* A и обозначается $\mathrm{co}\,\mathrm{A}$.

Предложение 3. *Множество* $\mathrm{co}\,\mathrm{A}$ *состоит из выпуклых комбинаций элементов из* A, *т. е. векторов вида* $\alpha_1 x_1 + \ldots + \alpha_m x_m$, *где* $x_i \in A$, $\alpha_i \ge 0$, $1 \le i \le m$, *и* $\sum_{i=1}^m \alpha_i = 1$.

Доказательство. Покажем, что множество A' всех выпуклых комбинаций элементов из A выпукло. Действительно, пусть $x, y \in A'$, т. е.

$$x = \sum_{i=1}^m \alpha_i x_i, \quad x_i \in A, \quad \alpha_i \ge 0, \quad i = 1, \dots m, \quad \sum_{i=1}^m \alpha_i = 1$$

и

$$y = \sum_{j=1}^s \beta_j y_i, \quad y_j \in A, \quad \beta_j \ge 0, \quad j = 1, \dots s, \quad \sum_{j=1}^s \beta_j = 1.$$

Для любого $\alpha \in [0,1]$ имеем

$$(1-\alpha)x + \alpha y = \sum_{i=1}^{m}(1-\alpha)\alpha_i x_i + \sum_{j=1}^{s}\alpha\beta_j y_i.$$

Так как

$$\sum_{i=1}^{m}(1-\alpha)\alpha_i + \sum_{j=1}^{s}\alpha\beta_j = 1,$$

то $(1-\alpha)x + \alpha y \in A'$ и значит, множество A' выпукло.

Покажем теперь, что если B — выпуклое множество, то оно содержит все выпуклые комбинации своих элементов. Доказываем индукцией по числу элементов m в такой комбинации. При $m = 1$ утверждение очевидно. Пусть оно верно для m. Докажем, что тогда оно верно и для $m+1$. Пусть $x = \sum_{i=1}^{m+1}\alpha_i x_i$ — выпуклая комбинация элементов из B. Запишем ее в виде

$$x = \sum_{i=1}^{m+1}\alpha_i x_i = \sum_{i=1}^{m}\alpha_i x_i + \alpha_{m+1}x_{m+1}. \tag{i}$$

Если α_{m+1} равняется нулю или единице, то доказывать нечего. Пусть $0 < \alpha_{m+1} < 1$. Тогда равенство (i) можно записать так:

$$x = \sum_{i=1}^{m+1}\alpha_i x_i = (1-\alpha_{m+1})\sum_{i=1}^{m}\frac{\alpha_i}{1-\alpha_{m+1}}\,x_i + \alpha_{m+1}x_{m+1}.$$

Сумма коэффициентов при x_i, $i = 1,\dots,m$, в сумме справа равна единице, и поэтому эта сумма, по предположению индукции, принадлежит B, а так как B выпукло, то $x \in B$.

Из доказанного следует, что A' принадлежит любому выпуклому множеству, содержащему A, т. е. A' принадлежит пересечению таких множеств. Но так как A' само выпукло, то оно совпадает с этим пересечением и значит, $A' = \operatorname{co}\mathrm{A}$. $\square$

Предложение 4. *Пусть A — непустое выпуклое подмножество $\mathbb{R}^n$.*

(a) *Если $x \in \operatorname{int} A$ и $y \in \operatorname{cl} A$, то $(x,y) \subset \operatorname{int} A$;*

(b) $\operatorname{int} A$ *и* $\operatorname{cl}\mathrm{A}$ — *выпуклые множества.*

Доказательство. (a) Пусть $z \in (x,y)$, тогда $z = (1-\alpha)x + \alpha y$ для некоторого $0 < \alpha < 1$. Так как $y \in \operatorname{cl} A$, то для любого $\delta > 0$ найдется $x = x(\delta) \in A$ такое, что $x \in A \cap U_{\mathbb{R}^n}(y,\delta)$, т. е. $x \in A$ и $x \in y + U_{\mathbb{R}^n}(0,\delta)$. Отсюда, поскольку $U_{\mathbb{R}^n}(0,\delta) = -U_{\mathbb{R}^n}(0,\delta)$, следует включение $y \in x + U_{\mathbb{R}^n}(0,\delta)$ и значит, $\operatorname{cl} A \subset A + U_{\mathbb{R}^n}(0,\delta)$.

Далее

$$\begin{aligned}(1-\alpha)x + \alpha y + U_{\mathbb{R}^n}(0,\delta) &\subset (1-\alpha)x + \alpha(A + U_{\mathbb{R}^n}(0,\delta)) + U_{\mathbb{R}^n}(0,\delta)\\ &= (1-\alpha)\left(x + \frac{1+\alpha}{1-\alpha}U_{\mathbb{R}^n}(0,\delta)\right) + \alpha A.\end{aligned}$$

Так как $x \in \operatorname{int} A$, то для достаточно малого δ выражение справа в больших скобках будет принадлежать A, а тогда, в силу выпуклости A, все выражение справа принадлежит A. Таким образом, $(1-\alpha)x + \alpha y + U_{\mathbb{R}^n}(0,\delta) \subset A$ и тем самым $z = (1-\alpha)x + \alpha y \in \operatorname{int} A$.

(b) Выпуклость $\operatorname{int} A$ сразу следует из (a). Докажем выпуклость $\operatorname{cl} \mathrm{A}$. Пусть $x_i \in \operatorname{cl} \mathrm{A}$, $i = 1, 2$ и $x = (1-\alpha)x_1 + \alpha x_2$, $0 \le \alpha \le 1$. Для любого $\delta > 0$ найдутся $\overline{x}_i \in (x_i + U_X(0,\delta)) \cap A$, $i = 1, 2$. Положим $\overline{x} = (1-\alpha)\overline{x}_1 + \alpha\overline{x}_2$. Тогда $\overline{x} \in A$ и

$$\overline{x} \in (1-\alpha)(x_1 + U_X(0,\delta)) + \alpha(x_2 + U_X(0,\delta)) = x + U_X(0,\delta),$$

т. е. каждая окрестность точки $x = (1-\alpha)x_1 + \alpha x_2$ имеет непустое пересечение с A и значит, $x \in \operatorname{cl} \mathrm{A}$. □

Непустое множество $C \subset \mathbb{R}^n$ называется *конусом*, если из включения $x \in C$ следует, что $\alpha x \in C$ для любого $\alpha > 0$.

Упражнение 2.2. *Доказать, что если конус C выпуклый, то $C + C \subset C$.*

Пусть A — непустое подмножество $\mathbb{R}^n$. Наименьший выпуклый конус, содержащий A и начало координат, называется *конической оболочкой множества A* и обозначается $\operatorname{cone} \mathrm{A}$.

Упражнение 2.3. *Доказать, что* $\operatorname{cone} \mathrm{A}$ *состоит из конических комбинаций элементов из A, т. е. векторов вида $\alpha_1 x_1 + \ldots + \alpha_m x_m$, где $\alpha_i \ge 0$, $x_i \in A$, $i = 1, \ldots, m$.*

Теорема (Каратеодори для конуса)**.** *Каждый элемент конической оболочки множества $A \subset \mathbb{R}^n$ может быть представлен как коническая комбинация не более чем n элементов из A.*

Доказательство. Пусть $x \in \operatorname{cone} A$. Тогда найдутся $k \in \mathbb{N}$, $x_i \in A$, $\alpha_i > 0$, $i = 1, \ldots, k$ такие, что $x = \sum_{i=1}^k \alpha_i x_i$.

Если $k \le n$, то все доказано. Пусть $k > n$. Тогда векторы x_i, $i = 1, \ldots, k$, линейно зависимы и поэтому найдутся числа $\lambda_1, \ldots, \lambda_k$, не все равные нулю, такие, что $\sum_{i=1}^k \lambda_i x_i = 0$. Пусть, для определенности, $\lambda_k \ne 0$. Тогда $x_k = \sum_{i=1}^{k-1} \mu_i x_i$, где $\mu_i = -\lambda_i/\lambda_k$, $i = 1, \ldots, k-1$.

Положим $\mu_k = -1$. Тогда предыдущее равенство означает, что $\sum_{i=1}^k \mu_i x_i = 0$.

Обозначим $\beta = \min\{-\alpha_i/\mu_i : \mu_i < 0,\ i = 1, \ldots, k\}$. Тогда

$$x = \sum_{i=1}^k \alpha_i x_i + \beta \sum_{i=1}^k \mu_i x_i = \sum_{i=1}^k (\alpha_i + \beta\mu_i) x_i = \sum_{i=1}^k \alpha_i' x_i,$$

причем, как легко проверить, $\alpha_i' \ge 0$, $i = 1, \ldots, k$ и среди чисел α_i' есть нулевые (так как минимум в определении β достигается хотя бы на одном элементе).

Таким образом, $x = \sum_{j=1}^{k'} \alpha_j' x_j$, где $\alpha_j' > 0$ и $k' < k$. Если $k' \le n$, то все доказано. Если же $k' > n$, то, повторяя процедуру, придем за конечное число шагов к утверждению теоремы. □

Следствие. *Если множество $A \subset \mathbb{R}^n$ таково,что* $\operatorname{cone} A = \mathbb{R}^n$*, то найдется не более чем $2n^2$ элементов из A таких, что любой элемент из $\mathbb{R}^n$ есть их коническая комбинация.*

Доказательство. Пусть $e_1, \ldots, e_n$ — базис в $\mathbb{R}^n$. Согласно теореме Каратеодори каждый из векторов $\pm e_i$, $i = 1, \ldots, n$, есть коническая комбинация не более чем n векторов из A. Множество всех этих векторов обозначим через B. Ясно, что число их не превосходит $2n^2$. Пусть $x \in \mathbb{R}^n$. Тогда $x = \sum_{i=1}^n \alpha_i e_i$. Если $\alpha_i < 0$, то запишем слагаемое $\alpha_i e_i$ в виде $-\alpha_i(-e_i)$. Подставляя вместо e_i или $-e_i$ их конические разложения, получаем, что любой элемент из $\mathbb{R}^n$ может быть представлен как коническая комбинация элементов из B. $\square$

Теоремы отделимости

Пусть $x^* \in (\mathbb{R}^n)^*$, $x^* \neq 0$ и $\gamma \in \mathbb{R}$. Множество

$$H = H(x^*, \gamma) = \{\, x \in \mathbb{R}^n : \langle x^*,\, x\rangle = \gamma \,\}$$

называется *гиперплоскостью.* Гиперплоскость порождает два полупространства

$$H_+(x^*, \gamma) = \{\, x \in \mathbb{R}^n : \langle x^*,\, x\rangle \leq \gamma \,\}$$

и

$$H_-(x^*, \gamma) = \{\, x \in \mathbb{R}^n : \langle x^*,\, x\rangle \geq \gamma \,\}.$$

Упражнение 2.4. *Доказать, что гиперплоскости и полупространства — выпуклые замкнутые множества.*

Пусть A и B — непустые подмножества $\mathbb{R}^n$. Говорят, что эти множества *отделимы*, если существует такая гиперплоскость, что A и B принадлежат различным полупространствам, порожденным этой гиперплоскостью.

Пусть, например, $A \subset H(x^*_+, \gamma)$, а $B \subset H(x^*_-, \gamma)$. Тогда легко видеть, что данное геометрическое определение отделимости равносильно, очевидно, следующему алгебраическому: множества A и B отделимы, если существует такой ненулевой элемент $x^* \in (\mathbb{R}^n)^*$, что

$$\sup_{a\in A} \langle x^*,\, a\rangle \leq \inf_{b\in B} \langle x^*,\, b\rangle.$$

Если неравенство строгое, то говорят, что множества A и B *строго отделимы.*

Мы сначала докажем теорему о строгой отделимости точки от множества (которую обычно называют *второй теоремой отделимости*), а затем выведем из нее теорему об отделимости множеств (которую называют *первой теоремой отделимости*).

Теорема (Вторая теорема отделимости)**.** *Пусть A — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$ и $b \notin A$. Тогда множество A и точка b строго отделимы.*

Доказательство. Пусть $\overline{x} \in A$. Обозначим $r = |\overline{x} - b|$ и $A_1 = A \cap B_{\mathbb{R}^n}(b, r)$. Рассмотрим функцию $f\colon \mathbb{R}^n \to \mathbb{R}$, определенную равенством $f(x) = |x - b|$ (расстояние от точки x до точки b). Эта функция непрерывна (упр. 1.10) и поэтому на ограниченном замкнутом множестве (компакте) A_1 она достигает, по теореме Вейерштрасса, своей нижней грани в некоторой точке $\widehat{x} \in A_1$. Положим $x^* = (\widehat{x} - b)^T$ и рассмотрим гиперплоскость

$$H = \{\, x \in \mathbb{R}^n : \langle x^*,\, x\rangle = \langle x^*,\, \widehat{x}\rangle \,\}.$$

Покажем, что вектор b и множество A находятся в разных полупространствах, порожденных этой гиперплоскостью. Действительно,

$$\langle x^*, b\rangle = \langle x^*, b - \widehat{x} + \widehat{x}\rangle = \langle x^*, -x^{*T} + \widehat{x}\rangle = -|x^*|^2 + \langle x^*, \widehat{x}\rangle < \langle x^*, \widehat{x}\rangle.$$

Теперь покажем, что A находится в другом полупространстве, т. е. $\langle x^*, x\rangle \geq \langle x^*, \widehat{x}\rangle$ для любого $x \in A$. Предположим противное, что существует такой элемент $x_0 \in A$, для которого $\langle x^*, x_0\rangle < \langle x^*, \widehat{x}\rangle$, или $\langle x^*, x_0 - \widehat{x}\rangle < 0$. Поскольку A выпукло, то $(1-t)\widehat{x} + tx_0 \in A$ при $0 \leq t \leq 1$ и при малых t (учитывая, что $\langle x^*, x_0 - \widehat{x}\rangle < 0$) будем иметь

$$|(1-t)\widehat{x} + tx_0 - b|^2 = |\widehat{x} - b + t(x_0 - \widehat{x})|^2 = |x^{*T} + t(x_0 - \widehat{x})|^2$$
$$= |x^*|^2 + 2t\langle x^*, x_0 - \widehat{x}\rangle + t^2|x_0 - \widehat{x}|^2 < |x^*|^2 = |\widehat{x} - b|^2 \leq |\overline{x} - b|^2 = r^2.$$

Следовательно, точки $(1-t)\widehat{x}+tx_0$ принадлежат множеству A_1, но это противоречит тому, что $\widehat{x}$ — минимум функции f на этом множестве. Итак, $\langle x^*, b\rangle < \langle x^*, \widehat{x}\rangle \leq \langle x^*, x\rangle$ для всех $x \in A$, или

$$\langle x^*, b\rangle < \inf_{x\in A}\langle x^*, x\rangle,$$

т. е. A и b строго отделимы. □

Пусть L — подпространство в $\mathbb{R}^n$. Множество

$$L^{\perp} = \{\, x^* \in (\mathbb{R}^n)^* : \langle x^*, x\rangle = 0, \quad \forall\, x \in L \,\}$$

называется *аннулятором* L. Легко видеть, что это подпространство в $(\mathbb{R}^n)^*$.

Следующее утверждение есть полезное следствие доказанной теоремы.

Лемма (о нетривиальности аннулятора)**.** *Если L — собственное подпространство в $\mathbb{R}^n$, то $L^{\perp} \neq \{0\}$.*

Доказательство. Подпространство L замкнуто, поскольку конечномерно и тем самым есть выпуклое замкнутое множество. Пусть $x_0 \in \mathbb{R}^n \setminus L$. Тогда по второй теореме отделимости найдется элемент $x^* \in (\mathbb{R}^n)^*$, $x^* \neq 0$, такой, что

$$\langle x^*, x_0\rangle < \inf_{x\in L}\langle x^*, x\rangle. \qquad (i)$$

Отсюда следует, что $x^* \in L^{\perp}$. Действительно, пусть $\langle x^*, \overline{x}\rangle \neq 0$ для некоторого $\overline{x} \in L$. Так как $\alpha\overline{x} \in L$ для любого $\alpha \in \mathbb{R}$, то правая часть часть в (i) будет равна $-\infty$, что, очевидно, невозможно. □

Теорема (Первая теорема отделимости)**.** *Пусть A и B — непустые выпуклые подмножества $\mathbb{R}^n$ и $A \cap B = \emptyset$. Тогда A и B отделимы.*

Доказательство. Положим $C = A - B$. Ясно, что множество C выпукло и $0 \notin C$. Если $0 \notin \mathrm{cl\,C}$, то, согласно предложению 4, множество $\mathrm{cl\,C}$ также выпукло, и тогда по второй теореме отделимости найдется ненулевой вектор $x^* \in (\mathbb{R}^n)^*$ такой, что $\langle x^*, x\rangle < 0$ для всех $x \in \mathrm{cl\,C}$ и тем более для $x = a-b \in A-B = C$, т. е. $\langle x^*, a\rangle < \langle x^*, b\rangle$ для всех $a \in A$ и $b \in B$ и значит, множества A и B отделимы.

Пусть $0 \in \mathrm{cl\,C}$. Предположим, что по крайней мере одно из множеств A и B имеет непустую внутренность. Пусть $\mathrm{int}\,A \neq \emptyset$. Тогда, так как множество $\mathrm{int}\,A - B = \cup_{b\in B}(\mathrm{int}A - b)$ открыто (как объединение открытых множеств) и принадлежит C, то $\mathrm{int}\,C \neq \emptyset$. Если $x \in \mathrm{int}\,C$, то для любого $\lambda > 0$ вектор $-\lambda x$ не принадлежит $\mathrm{cl\,C}$, ибо в противном случае, в силу предложения 4,

$$0 = \frac{\lambda}{1+\lambda}\,x + \frac{1}{1+\lambda}(-\lambda x)$$

принадлежал бы $\mathrm{int}\,C$ и тем самым $\mathrm{cl\,C}$. Следовательно, существует последовательность точек $\{x_k\}$, не принадлежащих $\mathrm{cl\,C}$, таких, что $x_k \to 0$ при $k \to \infty$.

По второй теореме отделимости найдутся такие ненулевые векторы $x_k^* \in (\mathbb{R}^n)^*$, что

$$\langle x_k^*,\, x\rangle < \langle x_k^*,\, x_k\rangle \tag{i}$$

для всех $x \in \mathrm{cl\,C}$ и $k \in \mathbb{N}$. Деля, если необходимо (i) на $|x_k^*|$, считаем, что $|x_k^*| = 1$, $k \in \mathbb{N}$. Тогда из последовательности $\{x_k^*\}$ можно выделить подпоследовательность, сходящуюся к некоторому вектору x^* и при этом $|x^*| = 1$. Переходя в неравенстве (i) для этой подпоследовательности к пределу, получаем, что $\langle x^*,\, x\rangle \le 0$ для всех $x \in \mathrm{cl\,C}$. Отсюда, как и в предыдущем случае, следует, что множеств A и B отделимы.

Случай, когда $0 \in \mathrm{cl\,C}$, но $\mathrm{int}\,C = \emptyset$ не доказываем. □

1.3 Теорема о разрешимости нелинейного уравнения и ее следствия

В основе доказательства необходимых условий экстремума лежат, как правило, либо теорема об обратной функции, либо теорема Люстерника о касательном пространстве. Обе теоремы являются простыми следствиями теоремы о разрешимости нелинейного уравнения, доказательство которого основано на методе Ньютона. В этой связи представляется естественным сказать несколько слов об истории этого метода.

И. Ньютон неоднократно излагал способ нахождения корня нелинейного уравнения $F(x) = 0$ для функции одного переменного. В частности, в письме к секретарю Королевского лондонского общества Генриху Ольденбургу (1676 г.) он демонстрирует его на примере решения уравнения $x^3 - 2x - 5 = 0$. Идея этого метода очень проста и естественна. Если точка x_0 близка к решению нелинейного уравнения, то заменяем в окрестности этой точки нелинейное уравнение его линейным приближением. Находим решение x_1 линейного уравнения, которое, вообще говоря, должно быть ближе к решению нелинейного уравнения, чем x_0. Беря x_1 за исходную точку, повторяем процедуру. Получаем последовате5льность $\{x_k\}$, которая, по-видимому, должна сходиться к решению нелинейного уравнения. Это и есть метод Ньютона.

Для нахождения решения уравнения $F(x) = y$ он заключается в построении последовательности $\{x_k\}$, где x_k находится как решение линейного уравнения $F'(x_{k-1})(x - x_{k-1}) = y - F(x_{k-1})$, т. е.

$$x_k = x_{k-1} + (F'(x_{k-1}))^{-1}(y - F(x_{k-1})), \quad k \in \mathbb{N}.$$

Мы же будем использовать *модифицированный метод Ньютона*

$$x_k = x_{k-1} + R(y - F(x_{k-1})), \quad k \in \mathbb{N},$$

где R — правый обратный к производной отображения F в некоторой точке $\widehat{x}$. Кроме того, мы будем предполагать, что решение уравнения принадлежит некоторому множеству. Это удобно для различных приложение, что будет видно из дальнейшего.

Теорема (о разрешимости нелинейного уравнения). *Пусть C — выпуклый замкнутый конус в $\mathbb{R}^n$, U — окрестность точки $\widehat{x} \in \mathbb{R}^n$, отображение $F\colon U \to \mathbb{R}^m$ строго дифференцируемо в $\widehat{x}$ и $F'(\widehat{x})(C) = \mathbb{R}^m$. Тогда найдутся окрестности $V_1 \subset U$ и V_2 точек $\widehat{x}$ и $F(\widehat{x})$ соответственно и константа $c > 0$ такие, что для любых $x \in V_1 \cap (\widehat{x} + C)$ и $y \in V_2$ существует элемент $\varphi(x, y) \in U \cap (\widehat{x} + C)$, для которого справедливы соотношения*

$$F(\varphi(x, y)) = y, \qquad |\varphi(x, y) - x| \le c|y - F(x)|,$$

и при этом отображение $(x, y) \mapsto \varphi(x, y)$ непрерывно на $(V_1 \cap (\widehat{x} + C)) \times V_2$.

Доказательство. Обозначим для краткости $F'(\widehat{x}) = \Lambda$ и покажем сначала, что существуют отображение $R\colon \mathbb{R}^m \to C$ и константа $\gamma > 0$ такие, что

$$\Lambda R(y) = y, \qquad |R(y)| \le \gamma|y| \tag{i}$$

для всех $y \in \mathbb{R}^m$.

Действительно, пусть $e_1, \ldots, e_m$ — базис в $\mathbb{R}^m$. По предположению найдутся такие $f_i \in C$, $1 \le i \le 2m$, что $\Lambda f_i = e_i$ и $\Lambda f_{m+i} = -e_i$, $1 \le i \le m$. Для каждого $y = \sum_{i=1}^m y_i e_i \in \mathbb{R}^m$ положим

$$R(y) = \sum_{i=1}^m |y_i| g_i,$$

где $g_i = f_i$, если $y_i \ge 0$ и $g_i = f_{m+i}$, если $y_i < 0$.

Так как C — выпуклый конус и содержит ноль (в силу замкнутости), то легко убедиться, что C содержит все конические комбинации своих элементов и тем самым $R(y) \in C$. Далее,

$$\Lambda R(y) = \sum_{i=1}^m |y_i| \Lambda g_i = \sum_{i=1}^m y_i e_i = y$$

и

$$|R(y)| \le \sum_{i=1}^m |y_i||g_i| \le \max_{1 \le i \le m} |y_i| \sum_{i=1}^{2m} |f_i| \le \gamma|y|,$$

где $\gamma = \sum_{i=1}^{2m} |f_i|$, и соотношения (i) доказаны.

Перейдем непосредственно к доказательству теоремы. Пусть $\varepsilon > 0$ таково, что $\varepsilon\gamma < 1$. Согласно определению строгой дифференцируемости F в точке $\widehat{x}$ найдется $\delta = \delta(\varepsilon) > 0$ такое, что для всех $x, x' \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \subset U$ справедливо неравенство

$$|F(x) - F(x') - \Lambda(x - x')| \le \varepsilon|x - x'| = \frac{\theta}{\gamma}|x - x'|, \tag{ii}$$

где $\theta = \varepsilon\gamma$.

Отсюда, как отмечалось выше, следует непрерывность F на $U_{\mathbb{R}^n}(\widehat{x}, \delta)$. Окрестность $V_1 \subset U_{\mathbb{R}^n}(\widehat{x}, \delta/3)$ точки $\widehat{x}$ выберем так, что $|F(x) - F(\widehat{x})| < (1 - \theta)\delta/3\gamma$, если $x \in V_1$, а $V_2 = U_{\mathbb{R}^m}(F(\widehat{x}), (1 - \theta)\delta/3\gamma)$.

Пусть $x \in V_1 \cap (\widehat{x} + C)$ и $y \in V_2$. Рассмотрим последовательность (модифицированный метод Ньютона)

$$x_k = x_{k-1} + R(y - F(x_{k-1})), \quad k \in \mathbb{N}, \quad x_0 = x. \tag{iii}$$

Покажем, что эта последовательность принадлежит $U_{\mathbb{R}^n}(\widehat{x}, \delta) \cap (\widehat{x} + C)$ и фундаментальна. Первое доказываем по индукции. Очевидно, что $x_0 = x \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \cap (\widehat{x} + C)$. Пусть $x_s \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \cap (\widehat{x} + C)$, $1 \leq s \leq k$. Докажем, что $x_{k+1} \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \cap (\widehat{x} + C)$.

Используя последовательно (iii), неравенство в (i), равенство

$$\Lambda(x_k - x_{k-1}) - y + F(x_{k-1}) = 0, \tag{iv}$$

которое следует из (iii) после применения к обеим частям оператора Λ, (ii) и затем итерируя процедуру, будем иметь

$$|x_{k+1} - x_k| = |R(y - F(x_k))| \leq \gamma |y - F(x_k)| = \gamma |F(x_k) - F(x_{k-1}) - \Lambda(x_k - x_{k-1})|$$
$$\leq \theta |x_k - x_{k-1}| \leq \ldots \leq \theta^k |x_1 - x|. \tag{v}$$

Далее применяя неравенство треугольника, (v), формулу для суммы геометрической прогрессии, (iii) (при $k = 1$), неравенство в (i), выбор x и y, получим, что

$$|x_{k+1} - \widehat{x}| \leq |x_{k+1} - x_k| + \ldots + |x_1 - \widehat{x}| \leq |x_{k+1} - x_k| + \ldots + |x_1 - x| +$$
$$+ |x - \widehat{x}| \leq \left(\theta^k + \theta^{k-1} + \ldots + 1\right) |x_1 - x| + |x - \widehat{x}| < \frac{\gamma}{1-\theta} |y - F(x)| + |x - \widehat{x}|$$
$$\leq \frac{\gamma}{1-\theta} |y - F(\widehat{x})| + \frac{\gamma}{1-\theta} |F(x) - F(\widehat{x})| + |x - \widehat{x}| < \frac{\delta}{3} + \frac{\delta}{3} + \frac{\delta}{3} = \delta, \tag{vi}$$

т. е. $x_{k+1} \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$.

По предположению индукции $x_k \in \widehat{x} + C$, а тогда, всилу выпуклости C, из (iii) вытекает $x_{k+1} \in x_k + C \subset \widehat{x} + C + C \subset \widehat{x} + C$ и значит, вся последовательность $\{x_k\}$ принадлежит $U_{\mathbb{R}^n}(\widehat{x}, \delta) \cap (\widehat{x} + C)$.

Последовательность $\{x_k\}$ фундаментальна. Действительно, для любых $k, l \in \mathbb{N}$, рассуждая точно также как и при доказательстве предыдущего неравенства, будем иметь

$$|x_{k+l} - x_k| \leq |x_{k+l} - x_{k+l-1}| + \ldots + |x_{k+1} - x_k| \leq$$
$$\leq \left(\theta^{k+l-1} + \ldots + \theta^k\right) |x_1 - x| < \frac{\theta^k \gamma}{1-\theta} |y - F(x)|, \tag{vii}$$

откуда, очевидно, следует фундаментальность последовательности $\{x_k\}$.

Положим $\varphi(x, y) = \lim_{k \to \infty} x_k$. Из (vi) следует, что $\varphi(\xi, y) \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \subset U$, а из замкнутости C следует, что $\varphi(x, y) \in \widehat{x} + C$ и тем самым $\varphi(x.y) \in U \cap (\widehat{x} + C)$.

Переходя к пределу в (iv), получаем, что $F(\varphi(x, y)) = y$.

Полагая $k = 0$ в (vii) и переходя к пределу при $l \to \infty$, обозначая $c = \gamma/(1-\theta)$, приходим к нужному неравенству в формулировке теоремы.

Осталось доказать, что отображение φ непрерывно. Действительно, легко видеть, что отображение R непрерывно. Тогда из (iii) следует (по индукции), что элементы x_k, как функции на $(V_1 \cap (\widehat{x} + C)) \times V_2$, непрерывны. Переходя к пределу при $l \to \infty$ в (vii), получаем, что отображение φ есть равномерный предел непрерывных функций и значит, само непрерывно на $(V_1 \cap (\widehat{x} + C)) \times V_2$. $\square$

Следствие (Теорема об обратной функции для конуса)**.** *Пусть выполнены предположения теоремы о разрешимости нелинейного уравнения. Тогда найдутся окрестность W точки $F(\widehat{x})$, непрерывное отображение $\psi\colon W \to U \cap (\widehat{x} + C)$ и константа $K > 0$ такие, что*

$$F(\psi(y)) = y, \qquad |\psi(y) - \widehat{x}| \le K|y - F(\widehat{x})|$$

для всех $y \in W$.

Заметим, что если $C = \mathbb{R}^n$, то это классическая теорема об обратной функции.

Доказательство. Полагая в теореме о разрешимости нелинейного уравнения $x = \widehat{x}$, $W = V_2$ и $\psi(\cdot) = \varphi(\widehat{x}, \cdot)$, получаем утверждение данной теоремы с $K = c$. □

Пусть M — непустое подмножество $\mathbb{R}^n$. Элемент $h \in \mathbb{R}^n$ называется *касательным вектором к M в точке $\widehat{x} \in M$*, если существуют $\varepsilon > 0$ и отображение $r\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ такие, что $\widehat{x} + th + r(t) \in M$ для всех $t \in (-\varepsilon, \varepsilon)$, где $r(t) = o(t)$ при $t \to 0$. Множество всех касательных векторов к M в точке $\widehat{x} \in M$ обозначим $T_{\widehat{x}}M$.

Следствие (Теорема Люстерника)**.** *Пусть выполнены условия теоремы о разрешимости нелинейного уравнения с $C = \mathbb{R}^n$ и пусть $M = \{\, x \in \mathbb{R}^n \mid F(x) = F(\widehat{x}) \,\}$. Тогда*

$$T_{\widehat{x}}M = \operatorname{Ker} F'(\widehat{x}).$$

Доказательство. Пусть $h \in T_{\widehat{x}}M$. Тогда существуют $\varepsilon > 0$ и отображение $r\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ такие, что $F(\widehat{x} + th + r(t)) = F(\widehat{x})$ для всех $t \in (-\varepsilon, \varepsilon)$ и $r(t) = o(t)$ при $t \to 0$. Отсда, вследствие дифференцируемости F в точке $\widehat{x}$ имеем

$$0 = F(\widehat{x} + th + r(t)) - F(\widehat{x}) = tF'(\widehat{x})h + F'(\widehat{x})r(t) + o(th + r(t)) = tF'(\widehat{x})h + o(t).$$

Деля это равенство на t и переходя к пределу при $t \to 0$, получаем, что $h \in \operatorname{Ker} F'(\widehat{x})$.

Обратно, пусть $h \in \operatorname{Ker} F'(\widehat{x})$. Из утверждения теоремы о разрешимости нелинейного уравнения, где $C = \mathbb{R}^n$ следует существование такого $\varepsilon > 0$, что для всех $t \in (-\varepsilon, \varepsilon)$ справедливо равенство $F(\varphi(\widehat{x} + th, F(\widehat{x}))) = F(\widehat{x})$ и

$$|\varphi(\widehat{x} + th, F(\widehat{x})) - \widehat{x} - th| \le c|F(\widehat{x}) - F(\widehat{x} + th)| = c|F(\widehat{x}) - F(\widehat{x}) - tF'(\widehat{x})h - o(t)| = o(t),$$

т. е. $\varphi(\widehat{x} + th, F(\widehat{x})) = \widehat{x} + th + o(t)$ или $F(\widehat{x} + th + o(t)) = F(\widehat{x})$ и значит, h — касательный вектор к M в точке $\widehat{x}$. □

Исторический комментарий. Если классический метод Ньютона, о котором было сказано в начале параграфа, применить к нахождению решения уравнения $x^2 = 2$, то придем к последовательности

$$x_k = x_{k-1} + \frac{1}{2x_{k-1}}(2 - x_{k-1}^2) = \frac{1}{2}\left(x_{k-1} + \frac{2}{x_{k-1}}\right),$$

известной еще Герону (I век до н. э.). Модифицированный метод Ньютона стал активно применяться в двадцатом веке. В зарубежной литературе его иногда называют методом ... . В бесконечномерном случае этот метод изучался Л. В. Канторовичем.

2 Необходимые условия экстремума в гладких задачах

Экстремальная задача включает в себя функционал f (со значениями, вообще говоря, в расширенной прямой $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$) вместе со своей областью определения X и множество ограничений $C \subset X$. Формализованная экстремальная задача записывается так

$$f(x) \to \min(\max), \quad x \in C \tag{P}$$

и заключается в нахождении таких точек $x \in C$, в которых функционал f достигает своего минимума (максимума) на C. Такие точки называются *глобальными* или *абсолютными минимумами* (*максимумами*) в задаче (P) или ее *решениями*. Если нас интересуют и точки минимума и точки максимума, то вместо min(max) пишем extr и говорим о задаче на экстремум функционала f.

Отметим еще, что если $\widehat{x}$ — решение задачи (P) на минимум (максимум), то ясно, что $\widehat{x}$ — решение аналогичной задачи на максимум (минимум) с функционалом $-f$ вместо f.

Точки из множества ограничений C называются *допустимыми* в задаче (P). Если $C = X$, то задача (P) называется задачей *без ограничений*.

В гладких задачах, как правило, изучают локальные экстремумы (т. е. локальные минимумы и максимумы). Если в X определено понятие "окрестности точки", то точка $\widehat{x} \in C$ называется *локальным минимумом* (*максимумом*) в задаче (P), если существует такая ее окрестность U, что $f(x) \geq f(\widehat{x})$ ($f(x) \leq f(\widehat{x})$) для всех допустимых $x \in U$ (т. е. для всех $x \in C \cap U$).

2.1 Гладкие задачи без ограничений

Когда величина является максимальной или минимальной,
в этот момент она не течет ни вперед, ни назад
И. Ньютон

Пусть U — открытое подмножество $\mathbb{R}^n$ и $f\colon U \to \mathbb{R}$. Рассмотрим задачу

$$f(x) \to \mathrm{extr}, \quad x \in U. \tag{P_1}$$

Теорема (Ферма для гладких задач без ограничений)**.** *Если точка $\widehat{x} \in U$ является локальным экстремумом в задаче (P_1) и функция f дифференцируема в $\widehat{x}$, то*

$$f'(\widehat{x}) = 0.$$

Доказательство. Предположим, что (линейный функционал) $f'(\widehat{x})$ отличен от нуля. Тогда найдется элемент $x \in \mathbb{R}^n$, $x \neq 0$, такой, что $\langle f'(\widehat{x}), x\rangle \neq 0$. Из дифференцируемости f в $\widehat{x}$ следует, что для достаточно малых t справедливо представление

$$f(\widehat{x} + tx) = f(\widehat{x}) + t\left(\langle f'(\widehat{x}), x\rangle + \frac{o(t)}{t}\right).$$

Отсюда вытекает, что $f(\widehat{x}+tx) > f(\widehat{x})$ для всех малых $t \neq 0$ одного знака с $\langle f'(\widehat{x}), x\rangle$ и $f(\widehat{x}+tx) < f(\widehat{x})$, если знаки противоположны. Получили противоречие с тем, что $\widehat{x}$ — локальный экстремум. $\square$

Исторический комментарий. Впервые мысль о том, что в точке максимума или минимума функции “приращение несущественно”, высказал И.Кеплер в 1615 году. В словах И. Ньютона, вынесенных в эпиграф, говорится об аналогичном наблюдении. Аналитически этот факт выразил П. Ферма в 1638 году в письме к Робервалю и Мерсену, предназначенном для Р. Декарта (хотя понятие производной тогда еще не было, но с ее появлением рассуждения Ферма стало возможно интерпретировать, как равенство нулю производной функции в точке ее локального экстремума), и с этого года обычно отсчитывают начало теории экстремума.

Теорема (Необходимые условия экстремума второго порядка для задачи (P_1)). *Пусть $\widehat{x}$ — локальный минимум (максимум) функции f в задаче (P_1). Если функция f дважды дифференцируема в точке $\widehat{x} \in U$, то*

$$f'(\widehat{x}) = 0$$

и для любого $h \in \mathbb{R}^n$ выполняется неравенство

$$h^T f''(\widehat{x})h \geq 0 \ \ (\leq 0).$$

Доказательство. Если функция дважды дифференцируема в точке $\widehat{x}$, то она дифференцируема в этой точке, и поэтому первое утверждение теоремы следует из предыдущей теоремы.

Пусть, для определенности, $\widehat{x}$ — локальный минимум функции f. Учитывая первое утверждение, будем иметь по формуле Тейлора для любого $h \in \mathbb{R}^n$ и достаточно малых $t \in \mathbb{R}$ имеем

$$0 \leq f(\widehat{x} + th) - f(\widehat{x}) = \frac{t^2}{2} h^T f''(\widehat{x})h + o(t^2).$$

Деля на t^2 и переходя к пределу при $t \to 0$, получаем требуемое неравенство.

Для локального максимума рассуждения аналогичны. □

Теорема (Достаточные условия экстремума второго порядка для задачи (P_1)). *Пусть в задаче (P_1) функция f дважды дифференцируема в $\widehat{x} \in U$. Тогда, если $f'(\widehat{x}) = 0$ и для любого $h \in \mathbb{R}^n$, $h \neq 0$, выполняется неравенство*

$$h^T f''(\widehat{x})h > 0 \ \ (< 0),$$

то $\widehat{x}$ — локальный минимум (максимум) в задаче (P_1).

Доказательство. Пусть $h^T f''(\widehat{x})h > 0$ для любого ненулевого $h \in \mathbb{R}^n$. Функция $h \to h^T f''(\widehat{x})h$ непрерывна на $\mathbb{R}^n$. Обозначим через m ее минимальное значение на единичной сфере

$$\mathbb{S}^{n-1} = \{\, x \in \mathbb{R}^n : |x| = 1 \,\}.$$

Ясно, что $m > 0$. Снова по формуле Тейлора имеем для любого $h \neq 0$ такого, что $\widehat{x} + h \in U$ (учитывая, что $h/|h|$ и $h^T/|h|$ принадлежат $\mathbb{S}^{n-1}$)

$$f(\widehat{x} + h) = f(\widehat{x}) + \frac{1}{2} h^T f''(\widehat{x})h + o(|h|^2) = f(\widehat{x})$$
$$+ \frac{|h|^2}{2} \left(\frac{h^T}{|h|} f''(\widehat{x}) \frac{h}{|h|} + \frac{o(|h|^2)}{|h|^2} \right) \geq f(\widehat{x}) + \frac{|h|^2}{2} \left(m + \frac{o(|h|^2)}{|h|^2} \right).$$

Выражение в скобках в правой части неравенства положительно для достаточно малых h. Следовательно, $f(\widehat{x}+h) > f(\widehat{x})$ для таких h и значит, $\widehat{x}$ — локальный минимум в задаче (P_1).

Если $h^T f''(\widehat{x})h < 0$, то аналогичные рассуждения показывают, что $\widehat{x}$ — локальный максимум в задаче (P_1). $\square$

Отметим, что неравенство в теореме означают, что квадратичная форма $h \to h^T f''(\widehat{x})h = \sum_{i,j=1}^n a_{ij}h_ih_j$, где $a_{ij} = \partial^2 f(\widehat{x})/\partial x_i \partial x_j$, $i,j = 1,\ldots,n$, положительно (отрицательно) определена. Согласно критерию Сильвестра это равносильно тому, что главные миноры матрицы $f''(\widehat{x})$ (т. е. определители матриц $A_k = (a_{ij})$, $i,j = 1,\ldots,k$, $k = 1,\ldots,n$) положительны (чередуют знаки, причем $A_1 = a_{11} < 0$).

2. Задачи с ограничениями, задаваемые равенствами

Можно высказать следующий общий принцип. Если ищется максимум или минимум некоторой функции многих переменных при условии, что между этими переменными имеется связь, задаваемая одним или несколькими уравнениями, нужно прибавить к минимизируемой функции функции, задающие уравнения связи, умноженные на неопределенные множители, и искать затем максимум или минимум построенной суммы, как если бы переменные были независимы. Полученные уравнения, присоединенные к уравнениям связи, послужат для определения всех неизвестных

Ж. Л. Лагранж

Пусть U — открытое подмножество $\mathbb{R}^n$ и $f_i\colon U \to \mathbb{R}$, $i = 0,1,\ldots,m$. Задача

$$f_0(x) \to \text{extr}, \quad f_i(x) = 0, \quad i = 1,\ldots,m \qquad (P_2)$$

называется *задачей с ограничениями, задаваемые равенствами* или *задачей с ограничениями типа равенств.*

Свяжем с задачей (P_2) следующую функцию $\mathcal{L}\colon U \times (\mathbb{R}^{m+1})^* \to \mathbb{R}$, определяемую формулой

$$\mathcal{L}(x,\lambda) = \sum_{i=0}^m \lambda_i f_i(x),$$

которая называется *функцией Лагранжа задачи* (P_2), числа $\lambda_0, \lambda_1, \ldots, \lambda_m$ — *множителями Лагранжа*, а вектор $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$ — *набором множителей Лагранжа.*

Теорема (Необходимые условия первого порядка для задачи (P_2). Правило множителей Лагранжа)**.** *Пусть $\widehat{x} \in U$ является локальным экстремумом в задаче (P_2). Тогда, если функции f_i, $i = 0,1,\ldots,m$, строго дифференцируемы в $\widehat{x} \in U$, то найдется такой ненулевой набор множителей Лагранжа $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$, что*

$$\mathcal{L}_x(\widehat{x},\lambda) = 0 \;\Leftrightarrow\; \sum_{i=0}^m \lambda_i f_i'(\widehat{x}) = 0.$$

Если векторы $f_1'(\widehat{x}), \dots, f_m'(\widehat{x})$ линейно независимы, то $\lambda_0 \neq 0$.

Доказательство. Необходимые условия в задаче (P_2) заключаются в том, что векторы $f_0'(\widehat{x}), f_1'(\widehat{x}), \dots, f_m'(\widehat{x})$ линейно зависимы. Предположим, что они линейно независимы и придем к противоречию с тем, что $\widehat{x}$ — локальный экстремум.

Определим отображение $F\colon U \to \mathbb{R}^{m+1}$ по правилу

$$F(x) = (f_0(x), f_1(x), \dots, f_m(x))^T.$$

Векторы $f_0'(\widehat{x}), f_1'(\widehat{x}), \dots, f_m'(\widehat{x})$ образуют строки матрицы $F'(\widehat{x})$. Так как эти векторы линейно независимы, то $F'(\widehat{x})(\mathbb{R}^n) = \mathbb{R}^{m+1}$. Далее, поскольку функции $f_0, f_1, \dots, f_m$ строго дифференцируемы в точке $\widehat{x}$, то отображение F строго дифференцируемо в этой точке. Следовательно, выполнены предположения теоремы об обратной функции в ситуации, когда $C = \mathbb{R}^n$. Пусть окрестность W точки $F(\widehat{x})$, отображение $\psi\colon W \to U$ константа K из этой теоремы.

Так как $F(\widehat{x}) = (f_0(\widehat{x}), 0, \dots, 0)^T$, то для достаточно малых по модулю ν точки $y_\nu = (f_0(\widehat{x}) + \nu, 0, \dots, 0)^T$ принадлежат W, и мы имеем

$$F(\psi(y_\nu)) = y_\nu, \qquad |\psi(y_\nu) - \widehat{x}| \leq K|y_\nu - F(\widehat{x})| = K|\nu|. \tag{i}$$

Обозначим $x_\nu = \psi(y_\nu)$. Равенство в (i) означает, что

$$(f_0(x_\nu), f_1(x_\nu), \dots, f_m(x_\nu))^T = (f_0(\widehat{x}) + \nu, 0, \dots, 0)^T,$$

или

$$f_0(x_\nu) = f_0(\widehat{x}) + \nu, \quad f_i(x_\nu) = 0, \quad i = 1, \dots, m, \tag{ii}$$

т. е. элементы x_ν допустимы в задаче (P_2) и при этом, $f_0(x_\nu) < f_0(\widehat{x})$ ($f_0(x_\nu) > f_0(\widehat{x})$), если $\nu < 0$ ($\nu > 0$). Отсюда, из (ii) и неравенства в (i) следует, что в любой окрестности точки $\widehat{x}$ находятся допустимые в задаче (P_2) точки x_ν, в которых значения f_0 и больше и меньше, чем в точке $\widehat{x}$. Это противоречит тому, что $\widehat{x}$ — локальный экстремум. Первое утверждение теоремы доказано.

Второе проверяется без труда. □

Если $\lambda_0 \neq 0$, то поскольку множители Лагранжа в необходимых условиях определены с точностью до ненулевого множителя, можно считать, что $\lambda_0 = 1$ (как и полагал Лагранж, согласно его словам выше) и тогда необходимые условия экстремума и “уравнения связи” $f_i(x) = 0$, $i = 1, \dots, m$, дают $n + m$ соотношений для определения $n + m$ неизвестных: $x_1, \dots, x_n, \lambda_1, \dots, \lambda_m$.

Отметим, что Правило множителей Лагранжа для задачи (P_2) доказано в предположении строгой дифференцируемости всех функций в точке $\widehat{x}$, из чего следует, что все эти функции непрерывны в окрестности этой точки. Можно ли отказаться от этого и попробовать доказать теорему в предположении только дифференцируемости функций в $\widehat{x}$? Следующий пример показывает, что этого сделать нельзя.

Рассмотрим задачу

$$f_0(x) \to \mathrm{extr}, \quad f_1(x) = 0,$$

где функции $f_i\colon \mathbb{R}^2 \to \mathbb{R}$, $i = 0, 1$, таковы, что $f_0(x) = x_1$, а

$$f_1(x) = \begin{cases} x_1^2 x_2 + x_2, & x_2 \neq 0; \\ x_1^2(D(x_1) + 1), & x_2 = 0, \end{cases}$$

где $D(\cdot)$ — функция Дирихле.

Единственная допустимая точка — это нуль, и поэтому она является решением данной задачи. Ясно, что $f_0'(0) = (1, 0)$ и несложно проверить, что $f_1'(0) = (0, 1)$. Функция f_1 имеет разрывы в любой окрестности нуля. Правило множителей Лагранжа для этой задачи не выполняется, поскольку векторы $f_0'(0)$ и $f_1'(0)$ линейно независимы.

Теорема (Необходимые условия экстремума второго порядка в задаче (P_2))**.** *Пусть $\widehat{x} \in U$ является локальным минимумом (максимумом) в задаче (P_2). Если функции f_i, $i = 0, 1, \ldots, m$, дважды дифференцируемы в точке $\widehat{x}$ и векторы $f_1'(\widehat{x}), \ldots, f_m'(\widehat{x})$ линейно независимы, то существует набор множителей Лагранжа $\lambda = (1, \lambda_1, \ldots, \lambda_m)$ такой, что*

$$\mathcal{L}_x(\widehat{x}, \lambda) = 0 \ \Leftrightarrow \ \sum_{i=0}^{m} \lambda_i f_i'(\widehat{x}) = 0$$

и для любого $h \in \mathbb{R}^n$, для которого $\langle f_i'(\widehat{x}), h \rangle = 0$, $i = 1, \ldots, m$, справедливо неравенство

$$h^T \mathcal{L}_{xx}(\widehat{x}, \lambda) h \geq 0 \ \ (\leq 0).$$

Доказательство. Так как функции f_i, $i = 0, 1, \ldots, m$, дважды дифференцируемы в точке $\widehat{x}$, то они непрерывно дифференцируемы в $\widehat{x}$ и значит, строго дифференцируемы в этой точке. Следовательно, первое утверждение теоремы сразу вытекает (с $\lambda_0 = 1$) из предыдущей теоремы (векторы $f_i'(\widehat{x})$, $i = 1, \ldots, m$, линейно независимы и поэтому $\lambda_0 \neq 0$. Деля все множители Лагранжа на λ_0, можно считать, что $\lambda_0 = 1$).

Для доказательства второго утверждения воспользуемся теоремой Люстерника о касательном пространстве. Определим отображение $F\colon U \to \mathbb{R}^m$ по правилу:

$$F(x) = (\, f_1(x), \ldots, f_m(x)\,)^T, \quad x \in U.$$

Строгая дифференцируемость функций f_i, $1 \leq i \leq m$, в точке $\widehat{x}$ равносильна строгой дифференцируемости F в этой точке. Кроме того, $F'(\widehat{x})(\mathbb{R}^n) = \mathbb{R}^m$ в силу линейной независимости векторов $f_1'(\widehat{x}), \ldots, f_m'(\widehat{x})$.

Пусть $h \in \mathbb{R}^n$ таково, что $\langle f_i'(\widehat{x}), h \rangle = 0$, $i = 1, \ldots, m$, или. равносильно, $h \in \operatorname{Ker} F'(\widehat{x})$. Тогда из теоремы Люстерника следует, что h — касательный вектор ко множеству $M = \{\, x \in \mathbb{R}^n : F(x) = F(\widehat{x}) = 0 \,\}$ в точке $\widehat{x}$, т. е. существует $\varepsilon > 0$ и отображение $r\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ такие, что

$$F(\widehat{x} + th + r(t)) = 0, \ \ \forall\, t \in (-\varepsilon, \varepsilon),$$

где $r(t) = o(t)$ при $t \to 0$. Это равносильно тому, что для таких t

$$f_i(\widehat{x} + th + r(t)) = 0, \quad i = 1, \ldots, m. \qquad (i)$$

Таким образом, элементы $\widehat{x} + th + r(t)$ допустимы для задачи (P_2).

Пусть $\widehat{x}$ — локальный минимум, тогда

$$f_0(\widehat{x} + th + r(t)) \geq f_0(\widehat{x})$$

для достаточно малых по модулю t.

Теперь по формуле Тейлора, учитывая (i), первое утверждение теоремы и формулу Тейлора, имеем для малых t

$$0 \le f_0(\widehat{x}+th+r(t))-f_0(\widehat{x}) = f_0(\widehat{x}+th+r(t))+\sum_{i=1}^{m}\lambda_i f_i(\widehat{x}+th+r(t))-f_0(\widehat{x})-\sum_{i=1}^{m}\lambda_i f_i(\widehat{x})$$
$$= \mathcal{L}(\widehat{x}+th+r(t),\lambda)-\mathcal{L}(\widehat{x},\lambda) = \frac{t^2}{2}h^T L_{xx}(\widehat{x},\lambda)h + o(t^2),$$

откуда, деля на t^2 и переходя к пределу при $t \to 0$, получаем второе утверждение теоремы, когла $\widehat{x}$ — локальный минимум.

В случае локального максимума рассуждения аналогичны. □

Теорема (Достаточные условия экстремума второго порядка в задаче (P_2))**.** *Пусть в задаче (P_2) функции f_i, $i = 0, 1, \ldots, m$, дважды дифференцируемы в допустимой точке $\widehat{x} \in U$, векторы $f_1'(\widehat{x}), \ldots, f_m'(\widehat{x})$ линейно независимы. Тогда, если найдется такой набор множителей Лагранжа $\lambda = (1, \lambda_1, \ldots, \lambda_m)$, что*

$$\mathcal{L}_x(\widehat{x},\lambda) = 0$$

и для любого ненулевого $h \in \mathbb{R}^n$, для которого $\langle f_i'(\widehat{x}),\, h\rangle = 0$, $i = 1, \ldots, m$, справедливо неравенство

$$h^T\mathcal{L}_{xx}(\widehat{x},\lambda)h > 0 \ \ (< 0),$$

то $\widehat{x}$ — локальный минимум (максимум) в задаче (P_2).

Доказательство. Будем доказывать от противного. Пусть $\widehat{x}$ не является локальным экстремумом в задаче (P_2). Для определенности считаем, что $\widehat{x}$ — не локальный минимум. Покажем, что в этом случае найдется ненулевое $\overline{h} \in \mathbb{R}^n$ такое, что $\langle f_i'(\widehat{x}),\, \overline{h}\rangle = 0$, $i = 1, \ldots, m$, и для любого набора $\lambda = (1, \lambda_1, \ldots, \lambda_m)$, для которого $\mathcal{L}_x(\widehat{x},\lambda) = 0$, выполняется неравенство

$$\overline{h}^T\mathcal{L}_{xx}(\widehat{x},\lambda)\overline{h} \le 0. \qquad (i)$$

Действительно, так как $\widehat{x}$ — не локальный минимум, то существует последовательность $\{x_k\}_{k\in\mathbb{N}}$ допустимых элементов в (P_2) такая, что $x_k \ne \widehat{x}$, $x_k \to \widehat{x}$ при $k \to \infty$ и $f_0(x_k) < f_0(\widehat{x})$. Обозначая $h_k = x_k - \widehat{x}$, будем иметь по формуле Тейлора для каждого $0 \le i \le m$ и каждого $k \in \mathbb{N}$

$$f_i(x_k) = f_i(\widehat{x}) + \langle f_i'(\widehat{x}),\, h_k\rangle + \frac{|h_k|^2}{2}\frac{h_k^T}{|h_k|}f_i''(\widehat{x})\frac{h_k}{|h_k|} + o(|h_k|^2). \qquad (ii)$$

Пусть $\lambda = (1, \lambda_1, \ldots, \lambda_m)$ и $\mathcal{L}_x(\widehat{x},\lambda) = 0$. Умножая i-ое равенство в (ii) на λ_i, $i = 0, 1, \ldots, m$, и затем их складывая, учитывая, что $f_i(x_k) = f_i(\widehat{x})$, $i = 1, \ldots, m$, будем иметь

$$f_0(x_k) - f_0(\widehat{x}) = \langle \mathcal{L}_x(\widehat{x},\lambda),\, h_k\rangle + \frac{|h_k|^2}{2}\frac{h_k^T}{|h_k|}\mathcal{L}_{xx}(\widehat{x},\lambda)\frac{h_k}{|h_k|} + o(|h_k|^2). \qquad (iii)$$

Выражение слева отрицательно, а первое слагаемое справа равно нулю по условию. Деля (iii) на $|h_k|^2/2$, приходим к неравенству

$$\frac{h_k^T}{|h_k|}\mathcal{L}_{xx}(\widehat{x},\lambda)\frac{h_k}{|h_k|} + \frac{o(|h_k|^2)}{|h_k|^2} < 0. \qquad (iv)$$

Последовательность $\{|h_k|^{-1}h_k\}$ принадлежит единичной сфере в $\mathbb{R}^n$ и поэтому из нее можно выделить подпоследовательность, сходящуюся к некоторому $\overline{h} \in \mathbb{R}^n$, $|\overline{h}| = 1$. Будем считать, что сама последовательность сходится к $\overline{h}$. Переходя к пределу в (iv), получаем неравенство (i).

Осталось проверить, что $\langle f_i'(\widehat{x}), \overline{h}\rangle = 0$, $1 \le i \le m$. Для всех $k \in \mathbb{N}$ имеем

$$f_i(x_k) = f_i(\widehat{x}) + \langle f_i'(\widehat{x}),\, h_k\rangle + o(h_k) \ \Leftrightarrow\ \langle f_i'(\widehat{x}),\, h_k\rangle + o(h_k) = 0, \quad i = 1, \ldots, m.$$

Деля каждое равенство на $|h_k|$ и переходя к пределу, получаем, что $\langle f_i'(\widehat{x}), \overline{h}\rangle = 0$, $i = 1, \ldots, m$. □

3. Задачи с ограничениями, задаваемые равенствами и неравенствами

Пусть U — открытое подмножество $\mathbb{R}^n$ и $f_i\colon U \to \mathbb{R}$, $i = 0, 1, \ldots, m$. Задача

$$f_0(x) \to \min, \quad f_i(x) \le 0, \quad 1 \le i \le m', \quad f_i(x) = 0, \quad m' + 1 \le i \le m \qquad (P_3)$$

называется *задачей с ограничениями, задаваемые равенствами и неравенствами* или *задачей с ограничениями типа равенств и неравенств*.

Ясно, что задача (P_2) является частным случаем данной задачи, но она была рассмотрена отдельно ввиду ее основополагающей роли в теории экстремума.

Свяжем с задачей (P_3) функцию Лагранжа

$$\mathcal{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x),$$

где $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$, которая, как мы видим, имеет тот же вид, что и в задаче (P_2).

Теорема (Необходимые условия минимума для задачи (P_3). Правило множителей Лагранжа)**.** *Пусть точка $\widehat{x} \in U$ является локальным минимумом в задаче (P_3). Если функции f_i, $i = 0, 1, \ldots, m$, строго дифференцируемы в $\widehat{x}$, то найдется такой ненулевой набор множителей Лагранжа $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$, что*

(a) $\mathcal{L}_x(\widehat{x}, \lambda) = 0 \ \Leftrightarrow\ \sum\limits_{i=0}^{m} \lambda_i f_i'(\widehat{x}) = 0$ *(условие стационарности);*

(b) $\lambda_i \ge 0,\ i = 0, 1, \ldots, m'$ *(условие неотрицательности);*

(c) $\lambda_i f_i(\widehat{x}) = 0,\ i = 1, \ldots, m'$ *(условие дополняющей нежесткости).*

Если векторы $f_{m'+1}'(\widehat{x}), \ldots, f_m'(\widehat{x})$ линейно независимы и существует вектор $h \in \mathbb{R}^n$ такой, что $\langle f_i'(\widehat{x}),\, h\rangle = 0$, $i = m' + 1, \ldots, m$, и $\langle f_i'(\widehat{x}),\, h\rangle < 0$ для тех индексов $1 \le i \le m'$, для которых $f_i(\widehat{x}) = 0$, то $\lambda_0 \ne 0$.

Доказательство. Заметим сначала, что утверждение (c) можно считать выполненным всегда. В самом деле, отбросим те ограничения среди неравенств, для которых

$f_i(\widehat{x}) < 0$. Пусть, для определенности, $f_i(\widehat{x}) < 0$, $i = 1, \ldots, k$, $k \le m'$. Тогда $\widehat{x}$ будет локальным минимумом и в новой задаче. Действительно, допустим, что это не так, и пусть V — окрестность точки $\widehat{x}$ такая, что $f_i(x) < 0$, $i = 1, \ldots, k$, для всех $x \in V$ (это возможно, так как функции строго дифференцируемы и тем самым непрерывны). Поскольку, по предположению, точка $\widehat{x}$ не является локальным минимумом для новой задачи, то в любой ее окрестности $V_0 \subset V$ найдется допустимая точка для новой задаче, значение функционала f_0 в которой меньше, чем в $\widehat{x}$. Но эта точка допустима и для исходной задачи, что противоречит тому $\widehat{x}$ ее локальный минимум.

Если теперь для новой задачи доказаны утверждения (a) и (b), то (c) выполняется автоматически. Дополнив найденный набор множителей Лагранжа нулевыми компонентами, соответствующие тем номерам, где $f_i(\widehat{x}) < 0$, получим утверждения (a), (b) и (c) для исходной задачи.

Рассмотрим отображение $F\colon U \times \mathbb{R}^{m'+1} \to \mathbb{R}^{m+1}$, определенное по правилу:

$$F(x,u) = (f_0(x) + u_0, f_1(x) + u_1, \ldots, f_{m'}(x) + u_{m'}, f_{m'+1}(x), \ldots, f_m(x))^T,$$

где $u = (u_0, u_1, \ldots, u_{m'})^T$. Оно строго дифференцируемо в точке $(\widehat{x}, 0)$ (проверьте!). Покажем, что если

$$F'(\widehat{x}, 0)(\mathbb{R}^n \times \mathbb{R}_+^{m'+1}) = \mathbb{R}^{m+1}, \qquad (i)$$

то $\widehat{x}$ не может быть локальным минимумом.

В самом деле, если справедливо (i), то выполнены условия теоремы об обратной функции для конуса, в которой вместо $\mathbb{R}^n$, $\mathbb{R}^m$, U и $\widehat{x}$ надо рассмотреть $\mathbb{R}^n \times \mathbb{R}^{m'+1}$, $\mathbb{R}^{m+1}$, $U \times \mathbb{R}^{m'+1}$ и $(\widehat{x}, 0)$, а $C = \mathbb{R}^n \times \mathbb{R}_+^{m'+1}$ (выпуклость и замкнутость C очевидны).

Пусть W — окрестность точки $F(\widehat{x}, 0) = (f_0(\widehat{x}), 0, \ldots, 0)^T$, отображение ψ, действующее из множества W во множество

$$(U \times \mathbb{R}^{m'+1}) \cap ((\widehat{x}, 0) + (\mathbb{R}^n \times \mathbb{R}_+^{m'+1})) = (U \times \mathbb{R}^{m'+1}) \cap (\mathbb{R}^n \times \mathbb{R}_+^{m'+1}) = U \times \mathbb{R}_+^{m'+1}$$

и константа K из этой теоремы.

Очевидно, что для всех достаточно малых $\varepsilon > 0$ точки $y_\varepsilon = (f_0(\widehat{x}) - \varepsilon, 0 \ldots, 0)^T$ принадлежат W и мы имеем

$$F(\psi(y_\varepsilon)) = y_\varepsilon, \qquad |\psi(y_\varepsilon) - (\widehat{x}, 0)| \le K|y_\varepsilon - F(\widehat{x}, 0)| = K\varepsilon. \qquad (ii)$$

Поскольку $\psi(y_\varepsilon) \in U \times \mathbb{R}_+^{m'+1}$, то $\psi(y_\varepsilon) = (x(y_\varepsilon), u(y_\varepsilon))$, где $x(y_\varepsilon) \in U$ и $u(y_\varepsilon) = (u_0(y_\varepsilon), u_1(y_\varepsilon), \ldots, u_{m'}(y_\varepsilon))^T \in \mathbb{R}_+^{m'+1}$. Тогда равенство в (ii) означает, что

$$(f_0(x(y_\varepsilon)) + u_0(y_\varepsilon),\ f_1(x(y_\varepsilon)) + u_1(y_\varepsilon), \ldots, f_{m'}(x(y_\varepsilon)) + u_{m'}(y_\varepsilon),$$
$$f_{m'+1}(x(y_\varepsilon)), \ldots, f_m(x(y_\varepsilon))^T = (f_0(\widehat{x}) - \varepsilon, 0 \ldots, 0)^T,$$

или

$$f_0(x(y_\varepsilon)) + u_0(y_\varepsilon) = f_0(\widehat{x}) - \varepsilon,\ \ f_1(x(y_\varepsilon)) + u_1(y_\varepsilon) = 0, \ldots, f_{m'}(x(y_\varepsilon)) + u_{m'}(y_\varepsilon) = 0,$$
$$f_{m'+1}(x(y_\varepsilon)) = 0, \ldots, f_m(x(y_\varepsilon) = 0.$$

Так как $u_i(y_\varepsilon) \ge 0$, $i = 1, \ldots, m'$, то $f_i(x(y_\varepsilon)) \le 0$, $i = 1, \ldots, m'$ и тем самым точки $x(y_\varepsilon)$ допустимы в задаче (P_3). Кроме того,

$$f_0(x(y_\varepsilon)) \le f_0(x(y_\varepsilon)) + u_0(y_\varepsilon) = f_0(\widehat{x}) - \varepsilon < f_0(\widehat{x}).$$

Из (ii) следует, что

$$|x(y_\varepsilon) - \widehat{x}| \le |(x(y_\varepsilon), u(y_\varepsilon)) - (\widehat{x}, 0)| = |\psi(y_\varepsilon) - (\widehat{x}, 0)| \le K\varepsilon.$$

Таким образом, в любой окрестности $\widehat{x}$ существуют допустимые в задаче (P_3) точки, в которых значение f_0 меньше, чем в $\widehat{x}$. Это противоречит тому, что $\widehat{x}$ — локальный минимум.

Итак,

$$F'(\widehat{x}, 0)(\mathbb{R}^n \times \mathbb{R}_+^{m'+1}) \ne \mathbb{R}^{m+1}.$$

Легко проверяется, что при линейном отображении образ выпуклого множества есть выпуклое множество и образ конуса есть конус и поэтому $C_1 = F'(\widehat{x}, 0)(\mathbb{R}^n \times \mathbb{R}_+^{m'+1})$ — выпуклый конус в $\mathbb{R}^{m+1}$.

Пусть $y_0 \notin C_1$. По первой теореме отделимости найдется ненулевой вектор $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_m) \in (\mathbb{R}^{m+1})^*$ такой, что

$$\langle \lambda,\, y_0 \rangle \le \inf_{y \in C_1} \langle \lambda,\, y \rangle. \qquad (iii)$$

Отсюда следует, что $\langle \lambda, y \rangle \ge 0$ для всех $y \in C_1$ (так как если $\langle \lambda,\, \overline{y} \rangle < 0$ для некоторого $\overline{y} \in C_1$, то $\alpha\overline{y} \in C_1$ для любого $\alpha > 0$, и это приводит к противоречию с неравенством (iii)). Поскольку производная отображения F в точке $(\widehat{x}, 0)$ действует, как нетрудно убедиться, по правилу:

$$F'(\widehat{x}, 0)[x, y] = (\langle f_0'(\widehat{x}),\, x \rangle + u_0, \dots, \langle f_{m'}'(\widehat{x}),\, x \rangle + u_{m'}, \langle f_{m'+1}'(\widehat{x}),\, x \rangle, \dots, \langle f_m'(\widehat{x}),\, x \rangle)^T$$

для любых $(x, u) \in \mathbb{R}^n \times \mathbb{R}^{m'+1}$, где $u = (u_0, u_1, \dots, u_{m'})$, то условие: $\langle \lambda, y \rangle \ge 0$ для всех $y \in C_1$ равносильно соотношению

$$\sum_{i=0}^{m'} \lambda_i(\langle f_i'(\widehat{x}),\, x \rangle + u_i) + \sum_{i=m'+1}^{m} \lambda_i \langle f_i'(\widehat{x}),\, x \rangle \ge 0, \quad \forall\, (x, u) \in \mathbb{R}^n \times \mathbb{R}_+^{m'+1}.$$

Полагая здесь $u = 0$, получаем, что линейный функционал $\sum_{i=0}^{m} \lambda_i f_i'(\widehat{x})$ неотрицателен на $\mathbb{R}^n$ и значит, он нулевой, т. е. доказано утверждение (a) теоремы. Полагая $x = 0$, получаем, что $\sum_{i=0}^{m'} \lambda_i u_i \ge 0$ для всех $u \in \mathbb{R}_+^{m'+1}$, откуда следует (b).

Докажем последнее утверждение теоремы. Предположим, что $\lambda_0 = 0$. Тогда из (a) последует, что среди чисел $\lambda_1, \dots, \lambda_{m'}$ есть ненулевые (так как в противном случае это противоречило бы линейной независимости векторов $f_{m'+1}'(\widehat{x}), \dots, f_m'(\widehat{x})$). Поскольку $\langle f_i'(\widehat{x}),\, h \rangle = 0$, $i = m'+1, \dots, m$, то из (a) вытекает, что $0 = \sum_{i=1}^{m'} \lambda_i \langle f_i'(\widehat{x}),\, h \rangle$. Если здесь какое-то $\lambda_{i_0} > 0$, то $f_{i_0}(\widehat{x}) = 0$ согласно (c). Следовательно, $\langle f_{i_0}'(\widehat{x}),\, h \rangle < 0$ по свойству h, и мы приходим к противоречию: $0 = \sum_{i=1}^{m'} \lambda_i \langle f_i'(\widehat{x}),\, h \rangle < 0$. □

3 Необходимые и достаточные условия минимума в выпуклых задачах

В этом параграфе будут получены необходимые и достаточные условия минимума в выпуклых задачах, т. е. в задачах минимизации выпуклых функций на выпуклых множествах.

3.1 Выпуклые задачи без ограничений

Понятие выпуклого множества было определено выше для пространства $\mathbb{R}^n$. Но для векторного пространства оно такое же. Пусть X — вещественное векторное пространство. Непустое множество $A \subset X$ называется *выпуклым,* если вместе с любыми двумя своими точками x и y оно содержит отрезок

$$[x, y] = \{\, z \in X : z = (1 - \alpha)x + \alpha y,\ 0 \le \alpha \le 1 \,\},$$

соединяющий точки x и y.

Пустое множество выпукло по определению.

Пусть $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ — расширенная прямая, т. е. прямая, пополненная символам $+\infty$, продолжающим естественное отношение порядка: $a \le +\infty$, $a \in \mathbb{R}$. Кроме того, предполагается, что $a + (+\infty) = +\infty$ для всех $a \in \mathbb{R}$, $a(+\infty) = +\infty$, если $a > 0$ и $+\infty + (+\infty) = +\infty$.

С каждой функцией $f\colon X \to \overline{\mathbb{R}}$ свяжем два множества

$$\operatorname{dom} f = \{\, x \in X : f(x) < +\infty \,\}$$

и

$$\operatorname{epi} f = \{\, (x, \alpha) \in X \times \mathbb{R} : \alpha \ge f(x),\ x \in \operatorname{dom} f \,\},$$

которые называются соответственно *эффективным множеством* и *надграфиком* (или *эпиграфом*) *функции* f.

Функция $f\colon X \to \overline{\mathbb{R}}$ называется *выпуклой*, если ее надграфик выпуклое множество в $X \times \mathbb{R}$.

Вот примеры выпуклых функций на прямой: $x \mapsto ax^2 + bx + c$, $a \ge 0$; $x \mapsto e^{\alpha x}$, $\alpha \in \mathbb{R}$; $x \mapsto |x|^p$, $p \ge 1$; $x \mapsto -\ln x$, если $x > 0$ и $+\infty$, если $x \le 0$; $x \mapsto x \log_2 x + (1 - x)\log_2(1 - x)$, если $0 < x < 1$ и $+\infty$ в остальных случаях.

Предложение 5. *Функция $f\colon X \to \overline{\mathbb{R}}$ выпукла тогда и только тогда, когда для любых $x_1, x_2 \in X$ и любых $\alpha_i \ge 0$, $i = 1, 2$, таких, что $\alpha_1 + \alpha_2 = 1$ справедливо неравенство*

$$f(\alpha_1 x_1 + \alpha_2 x_2) \le \alpha_1 f(x_1) + \alpha_2 f(x_2),$$

которое называется неравенством Иенссена.

Доказательство. Пусть $\operatorname{epi} f$ — выпуклое множество. Если $f(x_1)$ или $f(x_2)$ равны $+\infty$, то неравенство Иенссена выполняется очевидным образом. Пусть $f(x_i) < +\infty$, $i = 1, 2$. Ясно, что в этом случае $(x_i, f(x_i)) \in \operatorname{epi} f$. Для $\alpha_i \ge 0$, $i = 1, 2$, и $\alpha_1 + \alpha_2 = 1$, имеем

$$\alpha_1(x_1, f(x_1)) + \alpha_2(x_2, f(x_2)) = (\alpha_1 x_1 + \alpha_2 x_2,\ \alpha_1 f(x_1) + \alpha_2 f(x_2)) \in \operatorname{epi} f$$

и значит, выполнено неравенство Иенссена.

Обратно, пусть $(x_i, \beta_i) \in \operatorname{epi} f$, $i = 1, 2$. Следовательно, $\beta_i \ge f(x_i)$, $i = 1, 2$. Если $\alpha_i > 0$, $i = 1, 2$, и $\alpha_1 + \alpha_2 = 1$, то по неравенству Иенссена

$$f(\alpha_1 x_1 + \alpha_2 x_2) \le \alpha_1 f(x_1) + \alpha_2 f(x_2) \le \alpha_1 \beta_1 + \alpha_2 \beta_2.$$

Отсюда вытекает, что

$$\alpha_1(x_1, \beta_1) + \alpha_2(x_2, \beta_2) = (\alpha_1 x_1 + \alpha_1 x_2,\ \alpha_1 \beta_1 + \alpha_1 \beta_2) \in \operatorname{epi} f$$

и тем самым $\operatorname{epi} f$ — выпуклое множество. $\square$

Введем понятие субдифференциала для функций на $\mathbb{R}^n$, которое является аналогом производной для выпуклых функций. Пусть выпуклая функция $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ конечна в точке $\widehat{x}$.

Субдифференциалом функции f в точке $\widehat{x}$ называется множество (возможно пустое)

$$\partial f(\widehat{x}) = \{\, x^* \in (\mathbb{R}^n)^* : f(x) - f(\widehat{x}) \geq \langle x^*,\, x - \widehat{x}\rangle,\ \ \forall\, x \in \mathbb{R}^n \,\}.$$

Если субдифференциал не пуст, то это выпуклое замкнутое множество в $(\mathbb{R}^n)^*$. Действительно, субдифференциал можно записать так

$$\partial f(\widehat{x}) = \bigcap_{x \in \mathbb{R}^n} \{\, x^* \in (\mathbb{R}^n)^* : \langle x^*,\, x\rangle \leq \langle x^*,\, \widehat{x}\rangle + f(x) - f(\widehat{x}) \,\},$$

т. е. субдифференциал есть пересечение замкнутых полупространств, которые, очевидно, выпуклы.

Упражнение 4.1. *Привести пример функции, у которой субдифференциал в данной точке пуст.*

Следующее предложение показывает, что субдифференциал — достаточно естественное обобщение понятия производной на выпуклые функции.

Предложение 6. *Пусть f — выпуклая функция на $\mathbb{R}^n$, дифференцируемая в точке $\widehat{x}$. Тогда $\partial f(\widehat{x}) = \{f'(\widehat{x})\}$.*

Доказательство. Пусть $x \in \mathbb{R}^n$. Для любого $0 < \alpha < 1$ имеем по неравенству Иенссена: $f((1-\alpha)\widehat{x} + \alpha x) \leq (1-\alpha)f(\widehat{x}) + \alpha f(x)$, откуда

$$\frac{f(\widehat{x} + \alpha(x - \widehat{x})) - f(\widehat{x})}{\alpha} \leq f(x) - f(\widehat{x}).$$

В силу дифференцируемости f в точке $\widehat{x}$

$$\frac{f(\widehat{x} + \alpha(x - \widehat{x})) - f(\widehat{x})}{\alpha} = \frac{f(\widehat{x}) + \alpha\langle f'(\widehat{x}),\, x - \widehat{x}\rangle + o(\alpha) - f(\widehat{x})}{\alpha}.$$

Переходя здесь к пределу при $\alpha \to 0$ и учитывая предыдущее неравенство, получаем, что

$$\langle f'(\widehat{x}),\, x - \widehat{x}\rangle \leq f(x) - f(\widehat{x}),$$

т. е. $f'(\widehat{x}) \in \partial f(\widehat{x})$.

Обратно, если $x^* \in \partial f(\widehat{x})$, то для любого $x \in X$ и любого $t > 0$ имеем $f(\widehat{x} + tx) - f(\widehat{x}) \geq t\langle x^*,\, x\rangle$. Следовательно,

$$\frac{f(\widehat{x} + tx) - f(\widehat{x})}{t} \geq \langle x^*,\, x\rangle,$$

и аналогично предыдущему, отсюда следует, что $\langle f'(\widehat{x}),\, x\rangle \geq \langle x^*,\, x\rangle$ для любого x и значит, $x^* = f'(\widehat{x})$. ☐

Пусть $f_0\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ — выпуклая функция. Рассмотрим задачу

$$f(x) \to \min, \quad x \in \mathbb{R}^n, \tag{P_4}$$

которая называется *выпуклой задачей без ограничений.*

Отметим, что в этой задаче нет смысла говорить о локальных минимумах, поскольку любой локальный минимум является и глобальным. Действительно, пусть $\widehat{x}$ — локальный минимум, т. е. существует такая окрестность U точки $\widehat{x}$, что $f(\widehat{x}) \le f(x)$ для всех $x \in U$. Пусть теперь x — произвольная точка из $\mathbb{R}^n$. Для достаточно малых $0 < \alpha \le 1$ точки $(1-\alpha)\widehat{x} + \alpha x$ принадлежат U и поэтому (по неравенству Иенссена)

$$f(\widehat{x}) \le f((1-\alpha)\widehat{x} + \alpha x) \le (1-\alpha)f(\widehat{x}) + \alpha f(x),$$

откуда следует, что $f(\widehat{x}) \le f(x)$.

Теорема (Ферма для выпуклых функций)**.** *Точка $\widehat{x}$ является минимумом в задаче (P_4) тогда и только тогда, когда $0 \in \partial f(\widehat{x})$.*

Доказательство. Если $\widehat{x}$ — минимум в (P_3), то $f(x) - f(\widehat{x}) \ge 0 = \langle 0, x\rangle$ для любого $x \in \mathbb{R}^n$, т. е. $0 \in \partial f(\widehat{x})$. Если $0 \in \partial f(\widehat{x})$, то $f(x) - f(\widehat{x}) \ge \langle 0, x\rangle = 0$, т. е. $f(x) \ge f(\widehat{x})$ для любого $x \in \mathbb{R}^n$. $\square$

3.2 Выпуклые задачи с ограничениями

Пусть X — вещественное векторное пространство, $f_i\colon X \to \mathbb{R}$, $i = 0, 1, \ldots, m$ — выпуклые функции и A — выпуклое подмножество X. Задача

$$f_0(x) \to \min, \quad f_i(x) \le 0, \quad 1 \le i \le m, \quad x \in A \qquad (P_5)$$

называется *выпуклой задачей* или *задачей выпуклого программирования*.

С задачей (P_5) связывается функция Лагранжа $\mathcal{L}\colon X \times (\mathbb{R}^{m+1})^* \to \mathbb{R}$,

$$\mathcal{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x),$$

которая имеет тот же вид, что и в задаче (P_2).

Теорема (Каруша–Куна–Таккера)**.** 1) *Необходимость. Если $\widehat{x}$ — минимум в задаче (P_5), то найдется такой ненулевой набор множителей Лагранжа $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$, что выполняются*

(*a*) $\min\limits_{x \in A} \mathcal{L}(x, \lambda) = \mathcal{L}(\widehat{x}, \lambda)$ (*условие минимума*);

(*b*) $\lambda_i \ge 0,\ i = 0, 1, \ldots, m$ (*условие неотрицательности*);

(*c*) $\lambda_i f_i(\widehat{x}) = 0, \quad i = 1, \ldots, m$ (*условие дополняющей нежесткости*).

Если найдется точка $\overline{x} \in A$ такая, что $f_i(\overline{x}) < 0$, $1 \le i \le m$, то $\lambda_0 \ne 0$ (условие Слейтера).

2) *Достаточность. Если существует допустимая в (P_5) точка $\widehat{x}$ и набор множителей Лагранжа $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$, удовлетворяющие условиям (a), (b) и (c) и при этом $\lambda_0 > 0$, то $\widehat{x}$ — решение задачи (P_5).*

Доказательство. 1) Пусть $\widehat{x}$ — решение задачи (P_5). Рассмотрим множества

$$B = \{\, b = (b_0, b_1, \dots, b_m)^T \in \mathbb{R}^{m+1} \mid \exists\, x \in A : f_0(x) - f_0(\widehat{x}) \le b_0, \\ f_i(x) \le b_i,\ i = 1, \dots, m \,\}$$

и

$$C = \{\, c \in \mathbb{R}^{m+1} \mid c = (c_0, 0, \dots, 0)^T,\ c_0 < 0 \,\}.$$

Покажем, что эти множества непусты, выпуклы и $B \cap C = \emptyset$.

Действительно, непустота и выпуклость C очевидны. Если в B взять $x = \widehat{x}$, то ясно, что $\mathbb{R}^{m+1}_+ \subset B$ и тем самым $B \ne \emptyset$.

Докажем выпуклость B. Пусть $b = (b_0, \dots, b_m)^T$, $b' = (b'_0, \dots, b'_m)^T \in B$ и $\alpha \in [0, 1]$. Найдутся такие $x, x' \in A$, что

$$f_0(x) - f_0(\widehat{x}) \le b_0, \quad f_i(x) \le b_i, \quad i = 1, \dots, m$$

и

$$f_0(x') - f_0(\widehat{x}) \le b'_0, \quad f_i(x') \le b'_i, \quad i = 1, \dots, m.$$

Положим $x_\alpha = (1 - \alpha)x + \alpha x'$. Тогда $x_\alpha \in A$. Далее, по неравенству Иенссена,

$$f_0(x_\alpha) - f_0(\widehat{x}) = f_0((1-\alpha)x + \alpha x') - f_0(\widehat{x}) \le (1-\alpha) f_0(x) + \alpha f_0(x') - f_0(\widehat{x}) \\ = (1-\alpha)(f_0(x) - f_0(\widehat{x})) + \alpha(f_0(x') - f_0(\widehat{x})) \le (1-\alpha) b_0 + \alpha b'_0$$

и аналогично

$$f_i(x_\alpha) \le (1-\alpha) b_i + \alpha b'_i, \quad i = 1, \dots, m,$$

т. е. $(1-\alpha)b + \alpha b' \in B$ и тем самым B выпукло.

Если допустить, что $B \cap C \ne \emptyset$, то найдется элемент $\widetilde{x} \in A$ такой, что $f_i(\widetilde{x}) \le 0$, $i = 1, \dots, m$, и $f_0(\widetilde{x}) - f_0(\widehat{x}) \le c_0 < 0$, т. е. вектор $\widetilde{x}$ допустим в задаче (P_5) и $f_0(\widetilde{x}) < f_0(\widehat{x})$ в противоречие с тем, что $\widehat{x}$ — минимум.

Согласно первой теореме отделимости (см ...) найдется такой ненулевой вектор $\overline{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_m) \in (\mathbb{R}^{m+1})^*$, что

$$\sup_{c \in C} \sum_{i=0}^{m} \lambda_i c_i = \sup_{c_0 < 0} \lambda_0 c_0 \le \inf_{b \in B} \sum_{i=0}^{m} \lambda_i b_i. \tag{i}$$

Если $\lambda_0 < 0$, то верхняя грань равна $+\infty$, что противоречит (i). Если же $\lambda_0 \le 0$, то эта верхняя грань равна нулю и, таким образом,

$$\sum_{i=0}^{m} \lambda_i b_i \ge 0, \quad \forall\, b = (b_0, b_1, \dots, b_m)^T \in B. \tag{ii}$$

Подставляя в (ii) векторы $(1, 0, \dots, 0)^T, \dots, (0, \dots, 0, 1)^T$, принадлежащие, очевидно, B $(x = \widehat{x})$, получаем утверждение (b).

Теперь подставим в (ii) векторы $(0, \dots, 0, f_i(\widehat{x}), 0, \dots, 0)^T$, $i = 1, \dots, m$ (они принадлежат B, надо взять $x = \widehat{x}$). Тогда получим, что $\lambda_i f_i(\widehat{x}) \ge 0$. Но $\lambda_i f_i(\widehat{x}) \le 0$, так как $\lambda_i \ge 0$, а $f_i(\widehat{x}) \le 0$ и поэтому $\lambda_i f_i(\widehat{x}) = 0$, $i = 1, \dots, m$, что доказывает утверждение (c).

Докажем условие минимума (a). Пусть $x \in A$. Ясно, что $(f_0(x) - f_0(\widehat{x}), f_1(x), \ldots, f_m(x))^T \in B$. Подставляя этот вектор в (ii), приходим к неравенству $\sum_{i=0}^m \lambda_i f_i(x) \geq \lambda_0 f_0(\widehat{x})$. Добавляя справа нулевые слагаемые $\lambda_i f_i(\widehat{x})$, $i = 1, \ldots, m$, получаем, что $\mathcal{L}(x, \lambda) \geq \mathcal{L}(\widehat{x}, \lambda)$ и (a) доказано.

Докажем второе утверждение теоремы. Если $\lambda_0 = 0$, то ненулевой множитель Лагранжа находится среди остальных и поэтому (с учетом (c)) $\mathcal{L}(\overline{x}, \lambda) = \sum_{i=1}^m \lambda_i f_i(\overline{x}) < 0 = \sum_{i=1}^m \lambda_i f_i(\widehat{x}) = \mathcal{L}(\widehat{x}, \lambda)$, что противоречит (a).

2) Пусть x — допустимый вектор в задаче (P_5). Тогда, используя это обстоятельство вместе с (b), затем (a) и (c), будем иметь

$$\lambda_0 f_0(x) \geq \lambda_0 f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) = \mathcal{L}(x, \lambda) \geq \mathcal{L}(\widehat{x}, \lambda) = \lambda_0 f_0(x) + \sum_{i=1}^m \lambda_i f_i(\widehat{x}) = \lambda_0 f_0(\widehat{x}).$$

Деля на λ_0, получаем требуемое. □

4 Необходимые условия экстремума в задачах вариационного исчисления

Вариационное исчисление родилось с задачи о брахистохроне, предложенной в 1696 г. И. Бернулли для решения своим современникам. Задача была решена самим Бернулли, его братом Яковом, Ньютоном, Лейбницем и Лопиталем. Решения были разные, и вскоре еще было решено несколько сходных задач.

В начале 18 века И. Бернулли предложил Л. Эйлеру (тогда молодому человеку, учившегося вместе с сыновьями Бернулли) найти общие методы решения подобных задач. Начиная с 1732 г. Л. Эйлер начал активно этим заниматься и через 12 лет завершил свой фундаментальный труд “Modus invineniondi lineas curvas maximive proprietate gemdenties sive soluto problematis isoperimetrice latissimo sensu accepti” (“Метод нахождения кривых линий, обладающих свойствами максимума или минимума, или решение изопериметрической задачи, взятой в самом широком смысле”), Лозанна, 1744 г.

Эйлер доказывал необходимые условия экстремума в задачах (по сути своей бесконечномерных), переходя к их конечномерным аппроксимациям (так называемый метод ломаных). В 1759 г. он получил письмо от совсем молодого математика из Турина — Ж.-Л. Лагранжа, который доказывал необходимые условия экстремума, варьируя экстремальную функцию, сводя задачу к конечномерной. Идея Лагранжа привела Эйлера в восторг, о чем он и написал Лагранжу. Класс экстремальных задач, где такой подход применим, Эйлер назвал “вариационным исчислением”.

В упомянутом труде Эйлера была рассмотрена, в частности, задача, которая ныне называется простейшей задачей (классического) вариационного исчисления. Ниже выводятся необходимые условия экстремума в этой задаче, а также в задаче Больца, которая была поставлена лишь в 20 веке, но она близка к простейшей задаче и важна для понимания структуры необходимых условий экстремума в задачах вариационного исчисления с ограничениями.

4.1 Простейшая задача вариационного исчисления

Пусть $[t_0, t_1]$ — отрезок числовой прямой, $L\colon \mathbb{R}^3 \to \mathbb{R}$ — непрерывная функция трех переменных (обозначаемых t, x и $\dot{x}$) и $x_i \in \mathbb{R}$, $i = 0, 1$. Задача

$$J(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))\, dt \to \text{extr}, \quad x(t_0) = x_0, \quad x(t_1) = x_1 \qquad (Pr)$$

называется *простейшей задачей (классического) вариационного исчисления.*

Уточним постановки задачи. Обозначим через $C([t_0, t_1])$ и $C^1([t_0, t_1])$ множества всех непрерывных и непрерывно дифференцируемых функций $x(\cdot)$ на $[t_0,\, t_1]$. Это нормированные пространства соответственно с нормами

$$\|x(\cdot)\|_{C([t_0,\, t_1])} = \max_{t \in [t_0,\, t_1]} |x(t)|$$

и

$$\|x(\cdot)\|_{C^1([t_0,\, t_1])} = \max(\|x(\cdot)\|_{C([t_0,\, t_1])}, \|\dot{x}(\cdot)\|_{C([t_0,\, t_1])}).$$

Функция $x(\cdot) \in C^1([t_0,\, t_1])$ называется *допустимой в задаче* (Pr), если $x(t_i) = x_i$, $i = 0, 1$.

Допустимая функция $\widehat{x}(\cdot)$ называется *слабым локальным минимумом (максимумом)* в задаче (Pr) (далее слово "локальный" будем опускать), если существует такое $\varepsilon > 0$, что для любой допустимой функции $x(\cdot)$, для которой $\|x(\cdot) - \widehat{x}(\cdot)\|_{C^1([t_0,\, t_1])} < \varepsilon$, выполняется неравенство $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$ $(J(x(\cdot)) \leq J(\widehat{x}(\cdot)))$. *Слабый экстремум* — это либо слабый минимум, либо слабый максимум.

Далее, если фиксирована функция $\widehat{x}(\cdot)$, то для сокращения записи используем обозначения: $\widehat{L}_x(t) = L_x(t, \widehat{x}(t), \dot{\widehat{x}}(t))$ (частная производная L по x) и аналогично для частной производной L по $\dot{x}$.

Теорема (Необходимые условия экстремума первого порядка в задаче (Pr). Уравнение Эйлера)**.** *Пусть $\widehat{x}(\cdot)$ доставляет слабый экстремум в задаче (Pr). Тогда, если функция L непрерывна вместе со своими частными производными по x и $\dot{x}$, то $\widehat{L}_{\dot{x}}(\cdot) \in C^1([t_0, t_1])$ и для всех $t \in [t_0, t_1]$ выполнено уравнение Эйлера*

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0.$$

Доказательство. Пусть $x(\cdot) \in C^1([t_0, t_1])$ и $x(t_0) = x(t_1) = 0$. Положим $x_\alpha(\cdot) = \widehat{x}(\cdot) + \alpha x(\cdot)$, где $\alpha \in \mathbb{R}$. Очевидно, что $x_\alpha(\cdot) \in C^1([t_0, t_1])$ и $x_\alpha(t_i) = x_i$, $i = 0, 1$, т. е. функции $x_\alpha(\cdot)$ допустимы в задаче (Pr). Далее, ясно, что функция $\alpha \mapsto J(x_\alpha(\cdot))$ достигает в нуле локального экстремума. В силу условий теоремы, теоремы о производной суперпозиции функций и стандартных утверждений относительно дифференцируемости интеграла по параметру, эта функция дифференцируема в нуле и тогда по теореме Ферма ее производная в нуле равна нулю:

$$\left.\frac{dJ(x_\alpha(\cdot))}{d\alpha}\right|_{\alpha=0} = \int_{t_0}^{t_1} \left(\widehat{L}_x(t)x(t) + \widehat{L}_{\dot{x}}(t)\dot{x}(t)\right) dt = 0. \qquad (i)$$

Пусть $p(\cdot)$ такое, что $\dot p(\cdot) = \widehat L_x(\cdot)$ (например $p(t) = \int_{t_0}^t \widehat L_x(\tau)\,d\tau$). Подставляя $p(\cdot)$ в (i) и интегрируя по частям, приходим к соотношению

$$\int_{t_0}^{t_1} \Big(-p(t) + \widehat L_{\dot x}(t)\Big)\,\dot x(t)\,dt = 0,$$

справедливому для всех функций $x(\cdot) \in C^1([t_0, t_1])$, для которых $x(t_0) = x(t_1) = 0$. Но тогда для любой константы c справедливо и такое равенство

$$\int_{t_0}^{t_1} \Big(-p(t) + \widehat L_{\dot x}(t) + c\Big)\,\dot x(t)\,dt = 0. \qquad (ii)$$

Выберем $c = c_0$ так, чтобы

$$\int_{t_0}^{t_1} (-p(t) + \widehat L_{\dot x}(t) + c_0)\,dt = 0 \qquad (iii)$$

(возможность этого очевидна) и рассмотрим функцию

$$x(t) = \int_{t_0}^{t} (-p(\tau) + \widehat L_{\dot x}(\tau) + c_0)\,d\tau, \quad t \in [t_0, t_1].$$

Ясно, что $x(\cdot) \in C^1([t_0, t_1])$, $x(t_0) = 0$ и $x(t_1) = 0$ (в силу (iii)). Подставим $x(\cdot)$ в (ii) с $c = c_0$, получим

$$\int_{t_0}^{t_1} \Big(-p(t) + \widehat L_{\dot x}(t) + c_0\Big)^2\,dt = 0.$$

Отсюда следует равенство $-p(\cdot) + \widehat L_{\dot x}(\cdot) + c_0 = 0$, из которого вытекает, что $\widehat L_{\dot x}(\cdot) \in C^1([t_0, t_1])$. Дифференцируя равенство $p(\cdot) = \widehat L_{\dot x}(\cdot) + c_0$ и учитывая, что $\dot p(\cdot) = \widehat L_x(\cdot)$, получаем уравнение Эйлера. $\square$

4.2 Задача Больца

Пусть, как и выше, $[t_0, t_1]$ — отрезок числовой прямой, $L\colon \mathbb{R}^3 \to \mathbb{R}$ — непрерывная функция трех переменных: t, x и $\dot x$ и $l\colon \mathbb{R}^2 \to \mathbb{R}$ — непрерывная функция переменных ζ_0 и ζ_1. Задача

$$\mathcal{B}(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot x(t))\,dt + l(x(t_0), x(t_1)) \to \mathrm{extr} \qquad (Bl)$$

называется *задачей Больца*.

Эта задача также рассматривается на пространстве $C^1([t_0, t_1])$ и любая функция $x(\cdot) \in C^1([t_0, t_1])$, очевидно, допустима в данной задаче. Локальный экстремум также называется *слабым экстремумом*.

Функции $\widehat L_x(\cdot)$ и $\widehat L_{\dot x}(\cdot)$ определяются как и раньше, и кроме того, полагаем $\widehat l_{\zeta_i} = l_{\zeta_i}(\widehat x(t_0), \widehat x(t_1))$, $i = 0, 1$ (частные производные функции l по ζ_0 и ζ_1 в точке $(\widehat x(t_0), \widehat x(t_1))$).

Теорема (Необходимые условия экстремума первого порядка в задаче (Bl)). *Пусть $\widehat{x}(\cdot)$ доставляет слабый экстремум в задаче (Bl). Тогда, если функция L непрерывна вместе со своими частными производными по x и $\dot{x}$, а функция l непрерывно дифференцируема, то $\widehat{L}_{\dot{x}}(\cdot) \in C^1([t_0,t_1])$, для всех $t \in [t_0,t_1]$ выполнено уравнение Эйлера:*

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0$$

и условие трансверсальности:

$$\widehat{L}_{\dot{x}}(t_i) = (-1)^i \widehat{l}_{\zeta_i}, \quad i = 0, 1.$$

Доказательство. Пусть $x(\cdot) \in C^1([t_0,t_1])$ и $x_\alpha(\cdot) = \widehat{x}(\cdot) + \alpha x(\cdot)$. Рассуждая точно также как и в предыдущей теореме, приходим к соотношению

$$\left.\frac{d\mathcal{B}(x_\alpha(\cdot))}{d\alpha}\right|_{\alpha=0} = \int_{t_0}^{t_1} \Big(\widehat{L}_x(t)x(t) + \widehat{L}_{\dot{x}}(t)\dot{x}(t)\Big)\, dt + \widehat{l}_{\zeta_0}x(t_0) + \widehat{l}_{\zeta_1}x(t_1) = 0, \qquad (i)$$

справедливому для любого $x(\cdot) \in C^1([t_0,t_1])$.

Пусть $p(\cdot)$ таково, что $\dot{p} = \widehat{L}_x(t)$, $p(t_1) = -\widehat{l}_{\xi_1}$, т. е.

$$p(t) = -\widehat{l}_{\xi_1} + \int_t^{t_1} \widehat{L}_x(\tau)\, d\tau, \quad t \in [t_0,t_1].$$

Подставляя это в (i) и интегрируя по частям, получаем, что

$$\int_{t_0}^{t_1} \Big(-p(t) + \widehat{L}_{\dot{x}}(t)\Big)\, \dot{x}(t)\, dt + (\widehat{l}_{\xi_0} - p(t_0))x(t_0) = 0. \qquad (ii)$$

Пусть теперь $x(\cdot)$ такое, что $\dot{x} = -p(t) + \widehat{L}_{\dot{x}}(t)$, $x(t_0) = \widehat{l}_{\xi_0} - p(t_0)$, т. е.

$$x(t) = \widehat{l}_{\xi_0} - p(t_0) + \int_{t_0}^{t} (-p(\tau) + \widehat{L}_{\dot{x}}(\tau))\, d\tau, \quad t \in [t_0,t_1].$$

Ясно, что $x(\cdot) \in C^1([t_0,t_1])$. Подставляя это $x(\cdot)$ в (ii), приходим к равенству

$$\int_{t_0}^{t_1} \Big(-p(t) + \widehat{L}_{\dot{x}}(t)\Big)^2 dt + (\widehat{l}_{\xi_0} - p(t_0))^2 = 0,$$

откуда следует равенство $p(\cdot) = \widehat{L}_{\dot{x}}(\cdot)$, равносильное, в силу определения $p(\cdot)$, уравнению Эйлера, а также соотношение $p(t_0) = \widehat{l}_{\xi_0}$, или $\widehat{L}_{\dot{x}}(t_0) = \widehat{l}_{\xi_0}$. Условие $\widehat{L}_{\dot{x}}(t_1) = -\widehat{l}_{\xi_1}$ входит в определение $p(\cdot)$. □

Исторический комментарий. Уравнение Эйлера для простейшей задачи вариационного исчисления содержится в его уже упоминавшемся основополагающем труде Эйлера, Lausanne, 1744. Роль этого труда в развитии теории экстремума огромна. О. Больца (1857–1942) учился в Геттингене и Берлине (где в 1879 г. слушал лекции Вейерштрасса). В 1908 г. он издал учебник по вариационному исчислению, в котором изложил теорию Вейерштрасса. Этот учебник оказал большое влияние на преподавание вариационного исчисление во всем мире.

Интегралы уравнения Эйлера

Напомним, что первым интегралом дифференциального уравнения называется функция, которая постоянна на решениях данного уравнения.

Если лагранжиан L не зависит от переменной x, то уравнение Эйлера имеет очевидный первый интеграл

$$p(t) = L_{\dot{x}}(t, \dot{\widehat{x}}(t)) = \text{const},$$

который называется *интегралом импульса.*

Если лагранжиан L не зависит от переменной t, то уравнение Эйлера имеет первый интеграл

$$H(t) = L_{\dot{x}}(\widehat{x}(t), \dot{\widehat{x}}(t))\dot{\widehat{x}}(t) - L(\widehat{x}(t), \dot{\widehat{x}}(t)) = \text{const}.$$

Он называется *интегралом энергии.*

Для доказательства вычислим производную функции $H(\cdot)$ (учитывая, что $\widehat{x}(\cdot)$ удовлетворяет уравнению Эйлера):

$$\frac{dH}{dt} = L_{\dot{x}}(\widehat{x}(t), \dot{\widehat{x}}(t))\ddot{\widehat{x}}(t) + \frac{d}{dt}L_{\dot{x}}(\widehat{x}(t), \dot{\widehat{x}}(t))\dot{\widehat{x}}(t) - L_x(\widehat{x}(t), \dot{\widehat{x}}(t))\dot{\widehat{x}}(t) -$$
$$- L_{\dot{x}}(\widehat{x}(t), \dot{\widehat{x}}(t))\ddot{\widehat{x}}(t) = \left(\frac{d}{dt}L_{\dot{x}}(\widehat{x}(t), \dot{\widehat{x}}(t)) - L_x(\widehat{x}(t), \dot{\widehat{x}}(t))\right)\dot{\widehat{x}}(t) = 0$$

При доказательстве мы предположили существование $\ddot{\widehat{x}}(\cdot)$.

Мы рассмотрели "одномерные" варианты простейшей задачи и задачи Больца. Совершенно аналогично рассматриваются их векторные аналоги, когда $x = (x_1, \ldots, x_n)^T$. В этом случае роль пространств $C([t_0, t_1])$ и $C^1([t_0, t_1])$ играют пространства $C([t_0, t_1], \mathbb{R}^n)$ и $C^1([t_0, t_1], \mathbb{R}^n)$ — соответственно непрерывных и непрерывно дифференцируемых вектор-функций со значениями в $\mathbb{R}^n$. Они определяются аналогично одномерным вариантам, где $|x(t)| = \sqrt{x_1^2(t) + \ldots + x_n^2(t)}$. Необходимые условия экстремума здесь имеют тот же вид и их доказательства остаются прежними. Но формулы, разумеется, надо понимать векторно. Например,

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0 \;\Leftrightarrow\; -\frac{d}{dt}\widehat{L}_{\dot{x}_i}(t) + \widehat{L}_{x_i}(t) = 0, \quad i = 1, \ldots, n.$$

5 Задача Лагранжа вариационного исчисления

Пусть $[t_0, t_1]$ — конечный отрезок, заданы непрерывная функция $f\colon \mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^r \to \mathbb{R}$, непрерывное отображение $\varphi\colon \mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^r \to \mathbb{R}^n$ (переменных $t \in \mathbb{R}$, $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, $u = (u_1, \ldots, u_r)^T \in \mathbb{R}^r$) и векторы $x^i \in \mathbb{R}^n$, $i = 0, 1$. Рассмотрим следующую задачу Лагранжа

$$J(x(\cdot), u(\cdot)) = \int_{t_0}^{t_1} f(t, x(t), u(t))\, dt \to \text{extr}, \;\; \dot{x} = \varphi(t, x, u), \;\; x(t_0) = x^0, \;\; x(t_1) = x^1. \tag{P_6}$$

Эта задача рассматривается на пространстве $Z = C^1([t_0, t_1],\, \mathbb{R}^n) \times C([t_0, t_1],\, \mathbb{R}^r)$ относительно переменной $(x(\cdot), u(\cdot)) \in Z$. Понятия допустимой пары и слабого минимума (максимума) определяются естественным образом.

Функция
$$L(t,x,\dot x,u,\lambda_0,p)=\lambda_0 f(t,x,u)+\langle p,\,\dot x-\varphi(t,x,u)\rangle.$$
где $p\in(\mathbb{R}^n)^*$ и $\lambda_0\in\mathbb{R}$, называется лагранжианом задачи (P_6) .

Если фиксированы функции $\widehat{x}(\cdot)\in C^1([t_0,t_1],\mathbb{R}^n)$, $\widehat{u}(\cdot)\in C([t_0,t_1],\mathbb{R}^r)$, $p(\cdot)\colon[t_0,t_1]\to(\mathbb{R}^n)^*$ и λ_0, то для краткости записи, пишем
$$\widehat{L}(t)=L(t,\widehat{x}(t),\dot{\widehat{x}}(t),\widehat{u}(t),\lambda_0,p(t))$$
и аналогично для частных производных по x, $\dot x$ и u этой функции.

Предполагается, что функция f и отображение φ непрерывны вместе со своими частными производными по x и u.

Теорема (Необходимые условия экстремума в задаче (P_6))**.** *Если пара $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ доставляет слабый экстремума в задаче (P_6), то найдется ненулевая пара $(\lambda_0,\,p(\cdot))\in\mathbb{R}\times C^1([t_0,t_1],\,(\mathbb{R}^n)^*)$ такая, что $\widehat{L}_{\dot x}(\cdot)\in C^1([t_0,t_1],\,(\mathbb{R}^n)^*)$, выполняется условие стационарности по $x(\cdot)$:*
$$-\frac{d}{dt}\widehat{L}_{\dot x}(t)+\widehat{L}_x(t)=0\;\Leftrightarrow\;\dot p(t)=-p(t)\widehat{\varphi}_x(t)+\lambda_0\widehat{f}_x(t),\quad\forall\,t\in[t_0,t_1],$$
и условие стационарности по $u(\cdot)$:
$$\widehat{L}_u(t)=0\;\;\Leftrightarrow\;\;p(t)\widehat{\varphi}_u(t)=\lambda_0\widehat{f}_u(t),\quad\forall\,t\in[t_0,t_1].$$

Для доказательства этой теоремы понадобится стандартная теорема о дифференцируемой зависимости решения дифференциального уравнений от параметров, которую здесь приводим без доказательства и в удобной для нас форме.

Теорема (о дифференцируемой зависимости)**.** *Пусть $\widehat{u}(\cdot)\in C([t_0,t_1],\mathbb{R}^r)$ и $\widehat{x}(\cdot)$ — решение задачи Коши*
$$\dot x=\varphi(t,x,\widehat{u}(t)),\quad x(t_0)=x^0.$$
Тогда для любого $k\in\mathbb{N}$ и любого набора $\overline{u}(\cdot)=(u_1(\cdot),\dots,u_k(\cdot))$, где $u_i(\cdot)\in C([t_0,t_1],\mathbb{R}^r)$, $i=1,\dots,k$, найдется окрестность нуля V в $\mathbb{R}^k$ такая, что для каждого $\overline{\alpha}=(\alpha_1,\dots,\alpha_k)^T\in V$ существует единственное решение $x(\cdot,\overline{\alpha},\overline{u}(\cdot))$ задачи Коши
$$\dot x=\varphi\big(t,x,\widehat{u}(t)+\sum_{i=1}^k\alpha_iu_i(t)\big),\quad x(t_0)=x^0,$$
определенное на всем отрезке $[t_0,t_1]$.

Кроме того, $x(\cdot,\overline{\alpha},\overline{u}(\cdot))\to\widehat{x}(\cdot)$ при $\overline{\alpha}\to0$ в метрике $C^1([t_0,t_1],\mathbb{R}^n)$ и для каждого $t\in[t_0,t_1]$ отображение $\overline{\alpha}\mapsto x(t,\overline{\alpha},\overline{u}(\cdot))$ непрерывно дифференцируемо на V и его частные производные $x_{\alpha_i}(\cdot)=x_{\alpha_i}(\cdot,\overline{u}(\cdot))$ по α_i в нуле, как функции переменной t, удовлетворяют на отрезке $[t_0,t_1]$ дифференциальным уравнениям
$$\dot x_{\alpha_i}(t)=\widehat{\varphi}_x(t)x_{\alpha_i}(t)+\widehat{\varphi}_u(t)u_i(t),\quad x_{\alpha_i}(t_0)=0,\quad i=1,\dots,k.\tag{1}$$

Заметим, что $x_{\alpha_i}(\cdot,\overline{u}(\cdot))$ зависит только от $u_i(\cdot)$ и поэтому будем писать $x_{\alpha_i}(\cdot,u_i(\cdot))$, $i=1,\dots,k$.

Доказательство теоремы. Пусть $u(\cdot) \in C([t_0, t_1], \mathbb{R}^r)$. Согласно теореме о дифференцируемой зависимости, для всех достаточно малых по модулю α существует единственное решение $x(\cdot, \alpha, u(\cdot))$ задачи Коши

$$\dot{x} = \varphi(t, x, \widehat{u}(\cdot) + \alpha u(\cdot)), \quad x(t_0) = x^0$$

на отрезке $[t_0, t_1]$. Соответственно производную отображения $\alpha \mapsto x(t, \alpha, u(\cdot))$ в нуле обозначаем $x_\alpha(\cdot, u(\cdot))$.

В силу теоремы о дифференцируемой зависимости, функция

$$J(\alpha, u(\cdot)) = J(x(\cdot, \alpha, u(\cdot)), u(\cdot, \alpha, u(\cdot))) = \int_{t_0}^{t_1} f(t, x(t, \alpha, u(\cdot)), u(t, \alpha, u(\cdot)))\, dt,$$

как функция α, определена в некоторой окрестности нуля и, в силу стандартных фактов о дифференцируемости интеграла по параметру, непрерывно дифференцируема в этой окрестности, а ее производная в нуле, которую обозначим через $J_\alpha(u(\cdot))$, имеет вид

$$J_\alpha(u(\cdot)) = \int_{t_0}^{t_1} (\langle \widehat{f}_x(t),\, x_\alpha(t, u(\cdot))\rangle + \langle \widehat{f}_u(t),\, u(t)\rangle)\, dt. \tag{i}$$

Положим

$$M = \operatorname{span}\{\, (J_\alpha(u(\cdot)),\, x_\alpha(t_1, u(\cdot))) \in \mathbb{R}^{n+1} : u(\cdot) \in C([t_0, t_1],\, \mathbb{R}^r)\,\},$$

где span обозначает линейную оболочку множества в фигурных скобках, т. е. совокупность всех линейных комбинаций элементов этого множества.

Покажем, что если $M = \mathbb{R}^{n+1}$, то $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ не может быть слабым экстремумом в задаче (P_6).

Так как $M = \mathbb{R}^{n+1}$, то найдутся функции $u_i(\cdot) \in C([t_0, t_1], \mathbb{R}^r)$, $i = 1, \ldots, n+1$ такие, что векторы $(J_\alpha(u_i(\cdot)),\, x_\alpha(t_1, u_i(\cdot)))$, $i = 1, \ldots, n+1$, образуют базис в $\mathbb{R}^{n+1}$. Положим $\overline{u}(\cdot) = (u_1(\cdot), \ldots, u_{n+1}(\cdot))$ и

$$u(\cdot, \overline{\alpha}, \overline{u}(\cdot)) = \widehat{u}(\cdot) + \sum_{i=1}^{n+1} \alpha_i u_i(\cdot),$$

где $\overline{\alpha} = (\alpha_1, \ldots, \alpha_{n+1})^T \in \mathbb{R}^{n+1}$. В силу теоремы о дифференцируемой зависимости для достаточно малых по норме $\overline{\alpha}$ существует на $[t_0, t_1]$ единственное решение $x(\cdot, \overline{\alpha}, \overline{u}(\cdot))$ задачи Коши

$$\dot{x} = \varphi(t, x, u(t, \overline{\alpha}, \overline{u}(\cdot))), \quad x(t_0) = x^0.$$

Снова, в силу стандартных фактов анализа, для малых по норме $\overline{\alpha}$ определена функция

$$\overline{\alpha} \mapsto J(x(\cdot, \overline{\alpha},\, \overline{u}(\cdot)), u(\cdot, \overline{\alpha}, \overline{u}(\cdot))),$$

которая непрерывно дифференцируема в некоторой окрестности V_0 нуля в $\mathbb{R}^{n+1}$ и тем самым определено отображение $F\colon V_0 \to \mathbb{R}^{n+1}$, сопоставляющее $\overline{\alpha} \in V_0$ вектор

$$(J(x(\cdot, \overline{\alpha},\, \overline{u}(\cdot)), u(\cdot, \overline{\alpha}, \overline{u}(\cdot))),\, x(t_1, \overline{\alpha}, \overline{u}(\cdot))) \in \mathbb{R}^{n+1}.$$

Это отображение непрерывно дифференцируемо на V_0 (поскольку непрерывно дифференцируемы его компоненты) и нетрудно проверить, что его производная (матрица Якоби) в нуле $F'(0)$ такова, что ее столбцы суть векторы

$$\left(\int_{t_0}^{t_1} (\langle \widehat{f}_x(t),\, x_{\alpha_i}(t, u_i(\cdot))\rangle + \langle \widehat{f}_u(t),\, u_i(t)\rangle)\, dt,\ x_{\alpha_i}(t_1, u_i(\cdot))\right),\quad i = 1, \ldots, n+1.$$

Ясно, что производная $x_{\alpha_i}(t, u_i(\cdot))$ удовлетворяет тому же дифференциальному уравнению, что и $x_\alpha(t, u_i(\cdot))$ и поэтому они совпадают. Таким образом, столбцы матрицы Якоби — это векторы $(J_\alpha(u_i(\cdot)),\, x_\alpha(t_1, u_i(\cdot)))$, $i = 1, \ldots, n+1$, которые, как сказано выше, образуют базис $\mathbb{R}^{n+1}$ и тем самым линейно независимы. Следовательно, матрица Якоби невырождена и значит, $F'(0)(\mathbb{R}^{n+1}) = \mathbb{R}^{n+1}$.

Теперь применим теорему об обратной функции для конуса, где роль пространств $\mathbb{R}^n$ и $\mathbb{R}^m$ играет пространство $\mathbb{R}^{n+1}$, $C = \mathbb{R}^{n+1}$, $\widehat{x} = 0$. Отображение F непрерывно дифференцируемо на V_0 и поэтому строго дифференцируемо в нуле, $F'(0)(\mathbb{R}^{n+1}) = \mathbb{R}^{n+1}$. Согласно этой теореме найдутся окрестность W точки $F(0)$, отображение $\psi\colon W \to V_0$ и константа $K > 0$ такие, что

$$F(\psi(y)) = y, \qquad |\psi(y)| \le K|y - F(0)|$$

для всех $y \in W$.

Ясно, что $F(0) = (J(\widehat{x}(\cdot), \widehat{u}(\cdot)),\, x^1)$. Для достаточно малых по модулю ν вектор $y_\nu = (J(\widehat{x}(\cdot), \widehat{u}(\cdot)) + \nu,\, x^1)$ принадлежит окрестности W. Тогда $F(\psi(y_\nu)) = y_\nu$ или, обозначая $\psi(y_\nu) = \overline{\alpha}(y_\nu)$,

$$(J(x(\cdot, \overline{\alpha}(y_\nu), \overline{u}(\cdot)), u(\cdot, \overline{\alpha}(y_\nu), \overline{u}(\cdot))),\, x(t_1, \overline{\alpha}(y_\nu), \overline{u}(\cdot))) = (J(\widehat{x}(\cdot), \widehat{u}(\cdot)) + \nu,\, x^1),$$

то есть

$$J(x(\cdot, \overline{\alpha}(y_\nu), \overline{u}(\cdot)),\, u(\cdot, \overline{\alpha}(y_\nu), \overline{u}(\cdot))) = J(\widehat{x}(\cdot), \widehat{u}(\cdot)) + \nu, \quad x(t_1, \overline{\alpha}(y_\nu), \overline{u}(\cdot)) = x^1 \qquad (ii)$$

и, кроме того, $|\overline{\alpha}(y_\nu)| \le K|\nu|$.

По построению, пары $(x(\cdot, \overline{\alpha}(y_\nu), \overline{u}(\cdot)),\, u(\cdot, \overline{\alpha}(y_\nu), \overline{u}(\cdot)))$ удовлетворяют дифференциальному уравнению в задаче (P_6), $x(t_0, \overline{\alpha}(y_\nu), \overline{u}(\cdot)) = x^0$ и, согласно (ii), $x(t_1, \overline{\alpha}(y_\nu), \overline{u}(\cdot)) = x^1$. Таким образом, данные пары допустимы в задаче (P_6), и на каждой такой паре минимизируемый функционал меньше (больше), чем на паре $(\widehat{x}(\cdot), \widehat{u}(\cdot))$, если $\nu < 0$ ($\nu > 0$).

Далее, так как $\overline{\alpha}(y_\nu) \to 0$ при $\nu \to 0$, то согласно теореме о дифференцируемой зависимости, $x(\cdot, \overline{\alpha}(y_\nu), \overline{u}(\cdot)) \to \widehat{x}(\cdot)$ в $C^1([t_0, t_1],\, \mathbb{R}^n)$ при $\nu \to 0$ и, очевидно, $u(\cdot, \overline{\alpha}(y_\nu), \overline{u}(\cdot)) \to \widehat{u}(\cdot)$ в $C([t_0, t_1],\, \mathbb{R}^r)$ при $\nu \to 0$. Это означает, что в любой окрестности пары $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ найдется допустимая пара в задаче (P_6), в которой значение функционала и меньше и больше, чем в $(\widehat{x}(\cdot), \widehat{u}(\cdot))$, в противоречие с тем, что $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — слабый экстремум.

Таким образом, M — собственное подпространство в $\mathbb{R}^{n+1}$.

По лемме о нетривиальности аннулятора существует ненулевой линейный функционал (вектор) $\overline{\lambda} = (\lambda_0, \lambda) \in \mathbb{R} \times (\mathbb{R}^n)^*$, $\lambda = (\lambda_1, \ldots, \lambda_n)$, такой, что $\langle \overline{\lambda},\, y\rangle = 0$ для любого $y \in M$. В частности, это верно для элементов вида (см. (i))

$$(J_\alpha(u(\cdot)),\, x_\alpha(t_1, u(\cdot))) = \left(\int_{t_0}^{t_1} (\langle \widehat{f}_x(t),\, x_\alpha(t, u(\cdot))\rangle + \langle \widehat{f}_u(t),\, u(t)\rangle)\, dt,\, x_\alpha(t_1, u(\cdot))\right)$$

и значит, для любого $u(\cdot) \in C([t_0, t_1], \mathbb{R}^r)$ справедливо равенство

$$\lambda_0 \int_{t_0}^{t_1} (\langle \widehat{f}_x(t),\, x_\alpha(t, u(\cdot))\rangle + \langle \widehat{f}_u(t),\, u(t)\rangle)\, dt + \langle \lambda,\, x_\alpha(t_1, u(\cdot))\rangle = 0. \qquad (iii)$$

Обозначим через $p(\cdot)$ решение линейного уравнения

$$\dot p = -p\, \widehat{\varphi}_x(t) + \lambda_0 \widehat{f}_x(t), \quad p(t_1) = -\lambda. \qquad (iv)$$

Такое решение существует, единственно и принадлежит $C^1([t_0, t_1], \mathbb{R}^n)$. Подставляя в L указанные λ_0 и $p(\cdot)$, получим, что $\widehat{L}_{\dot x}(\cdot) = p(\cdot)$ и поэтому $\widehat{L}_{\dot x}(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n)$. Отсюда и (iv) следует справедливость условие стационарности по $x(\cdot)$.

Подставим теперь в (iii) вместо функции $\lambda_0 \widehat{f}_x(\cdot)$ ее выражение из (iv), а затем вместо функции $\widehat{\varphi}_x(\cdot) x_\alpha(\cdot, u(\cdot))$ ее выражение из (1) и учитывая, что $\langle p(t)\widehat{\varphi}_x(t),\, x_\alpha(t, u(\cdot))\rangle = \langle p(t),\, \widehat{\varphi}_x(t) x_\alpha(t, u(\cdot))\rangle$, $p(t_1) = -\lambda$ и $x_\alpha(t_0, u(\cdot)) = 0$, будем иметь для любого $u(\cdot) \in C([t_0, t_1], \mathbb{R}^r)$

$$\begin{aligned}\int_{t_0}^{t_1} (\langle \dot p(t),\, x_\alpha(t, u(\cdot))\rangle + \langle p(t),\, \dot x_\alpha(t, u(\cdot))\rangle - \langle p(t),\, \widehat{\varphi}_u(t) u(t)\rangle + \langle \lambda_0 \widehat{f}_u(t),\, u(t)\rangle)\, dt \\ + \langle \lambda,\, x_\alpha(t_1, u(\cdot))\rangle = \langle p(t),\, x_\alpha(t, u(\cdot))\rangle\big|_{t_0}^{t_1} + \int_{t_0}^{t_1} \langle \lambda_0 \widehat{f}_u(t) - p(t)\widehat{\varphi}_u(t),\, u(t)\rangle\, dt \\ + \langle \lambda,\, x_\alpha(t_1, u(\cdot))\rangle = \int_{t_0}^{t_1} \langle \lambda_0 \widehat{f}_u(t) - p(t)\widehat{\varphi}_u(t),\, u(t)\rangle\, dt = 0.\end{aligned}$$

Беря в последнем интеграле в качестве $u(\cdot)$ функцию $t \mapsto (\lambda_0 \widehat{f}_u(t) - p(t)\widehat{\varphi}_u(t))^T$, получаем, что $p(\cdot)\widehat{\varphi}_u(\cdot) = \lambda_0 \widehat{f}_u(\cdot)$ и тем самым доказано условие стационарности по $u(\cdot)$.

Пара $(\lambda_0,\, p(\cdot))$ — ненулевая. Действительно, если она нулевая, то из (iv) последует, что вектор $(\lambda_0, \lambda_1, \ldots, \lambda_n) \in (\mathbb{R}^{n+1})^*$ нулевой, что не так. $\square$

В качестве следствия доказанной теоремы можно получить необходимые условия экстремума в двух известных задачах классического вариационного исчисления —- в задаче со старшими производными и изопериметрической задаче. Доказательства не приводим.

Задача со старшими производными. Уравнение Эйлера–Пуассона

Пусть $[t_0, t_1]$ — отрезок числовой прямой, $L\colon \mathbb{R}^{n+2} \to \mathbb{R}$ — непрерывная функция (переменные которой обозначаем $t, x, \dot x, \ldots, x^{(n)}$) и $x_j^k \in \mathbb{R}$, $k = 0, 1, \ldots, n-1$, $j = 0, 1$. Задача

$$J(x(\cdot)) = \int_{t_0}^{t_1} f(t, x(t), \dot x(t), \ldots, x^{(n)}(t))\, dt \to \mathrm{extr}, \quad x^{(k)}(t_j) = x_j^k,$$
$$k = 0, 1, \ldots, n-1, \quad j = 0, 1, \quad (SP)$$

называется *задачей со старшими производными.*

Задачу (SP) рассматриваем на пространстве $C^n([t_0,t_1])$, функции которого n раз непрерывно дифференцируемы. Это банахово пространство с нормой

$$\|x(\cdot)\|_{C^n([t_0,t_1])} = \|x(\cdot)\|_{C([t_0,t_1])} + \|\dot{x}(\cdot)\|_{C([t_0,t_1])} + \ldots + \|x^{(n)}(\cdot)\|_{C([t_0,t_1])}.$$

Локальный экстремум в этом пространстве также называем *слабым экстремумом.*

Теорема (Необходимые условия экстремума в задаче (SP). Уравнение Эйлера–Пуассона)**.** *Пусть $\widehat{x}(\cdot)$ доставляет слабый экстремум в задаче (SP). Тогда, если функция f такова, что $f_{x^{(k)}}(\cdot) \in C^k([t_0,t_1])$, $k = 0,1,\ldots,n$, то для всех $t \in [t_0,t_1]$ выполнено уравнение Эйлера–Пуассона*

$$\sum_{k=0}^{n}(-1)^k \frac{d^k}{dt^k}\widehat{f}_{x^{(k)}}(t) = 0.$$

Изопериметрическая задача

Пусть $[t_0,t_1]$ — отрезок числовой прямой, функции $f_i\colon \mathbb{R}^{2n+1} \to \mathbb{R}$, $i = 0,1,\ldots,m$ (переменных $t \in \mathbb{R}$, $x \in \mathbb{R}^n$ и $\dot{x} \in \mathbb{R}^n$) непрерывны, $\alpha_i \in \mathbb{R}$, $i = 1,\ldots,m$, и $x_i \in \mathbb{R}^n$, $i = 0,1$. Задача

$$J(x(\cdot)) = \int_{t_0}^{t_1} f_0(t,x(t),\dot{x}(t))\,dt \to \text{extr}, \quad \int_{t_0}^{t_1} f_i(t,x(t),\dot{x}(t))\,dt = \alpha_i,$$
$$i = 1,\ldots,m, \quad x(t_0) = x^0, \quad x(t_1) = x^1 \quad (IZ)$$

называется *изопериметрической задачей.*

Эта задача также рассматривается на пространстве $C^1([t_0,t_1], \mathbb{R}^n))$, и локальный экстремум, по прежнему, называется *слабым укстремумом.*

Сопоставим задаче (IZ) следующую функцию

$$L(t,x,\dot{x},\lambda) = \sum_{i=0}^{m}\lambda_i f_i(t,x,\dot{x}),$$

где $\lambda = (\lambda_0,\lambda_1,\ldots,\lambda_m) \in (\mathbb{R}^{m+1})^*$.

Теорема (Необходимые условия минимума в задаче (IZ). Уравнение Эйлера)**.** *Пусть $\widehat{x}(\cdot)$ доставляет слабый минимум в задаче (IZ). Тогда, если функции f_i, $i = 0,1,\ldots,m$, непрерывны вместе со своими частными производными по x и $\dot{x}$, то найдется такой ненулевой набор множителей Лагранжа $\lambda = (\lambda_0,\lambda_1,\ldots,\lambda_m)$, что $\widehat{L}_{\dot{x}}(\cdot) \in C^1([t_0,t_1], (\mathbb{R}^n)^*)$ и для всех $t \in [t_0,t_1]$ выполнено уравнение Эйлера*

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0.$$

6 Необходимые условия минимума в задаче оптимального управления. Принцип максимума Понтрягина

Постановка задачи

Пусть $[t_0,t_1]$ — конечный отрезок, U — непустое подмножество $\mathbb{R}^r$, $f\colon \mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^r \to \mathbb{R}$ — непрерывная функция и $\varphi\colon \mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^r \to \mathbb{R}^n$ — непрерывное отображение переменных $t \in \mathbb{R}$, $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$, $u = (u_1, \dots, u_r)^T \in \mathbb{R}^r$ и $x_i \in \mathbb{R}^n$, $i = 0, 1$.

Задача

$$J(x(\cdot), u(\cdot)) = \int_{t_0}^{t_1} f(t, x(t), u(t))\, dt \to \min, \quad \dot{x} = \varphi(t, x, u), \quad u(t) \in U,$$
$$x(t_0) = x^0, \quad x(t_1) = x^1 \quad (P_7)$$

называется *задачей оптимального управления* (в понтрягинской форме). Функцию $u(\cdot)$ называют *управлением*, а функцию $x(\cdot)$ *фазовой* переменной или траекторией.

Обозначим через $PC([t_0, t_1], \mathbb{R}^r)$ совокупность кусочно-непрерывных вектор-функций на $[t_0, t_1]$ со значениями в $\mathbb{R}^r$, а через $PC^1([t_0, t_1], \mathbb{R}^n)$ совокупность кусочно-непрерывно дифференцируемых вектор-функций (т. е. непрерывных функций, у которых производная кусочно непрерывна) на $[t_0, t_1]$ со значениями в $\mathbb{R}^n$.

Пара $(x(\cdot), u(\cdot)) \in PC^1([t_0, t_1], \mathbb{R}^n) \times PC([t_0, t_1], \mathbb{R}^r)$ называется *допустимой в задаче* (P_6), если включение $u(t) \in U$ и равенство $\dot{x}(t) = \varphi(t, x(t), u(t))$ выполняются для всех $t \in [t_0, t_1]$, где функция $u(\cdot)$ непрерывна и $x(t_i) = x^i$, $i = 0, 1$.

Допустимая пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ называется *сильным минимумом* или *оптимальным процессом* в задаче (P_7), если существует такое $\varepsilon > 0$, что для любой допустимой пары $(x(\cdot), u(\cdot))$, для которой $\|x(\cdot) - \widehat{x}(\cdot)\|_{C([t_0,t_1], \mathbb{R}^n)} < \varepsilon$ выполнено неравенство $J(x(\cdot), u(\cdot)) \geq J(\widehat{x}(\cdot), \widehat{u}(\cdot))$.

Всюду далее предполагается, что *функция f и отображение φ непрерывны по совокупности переменных вместе со своими частными производными по x.*

Если фиксирована пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$, то, как и раньше, пишем $\widehat{f}(t) = f(t, \widehat{x}(t), \widehat{u}(t))$ и $\widehat{\varphi}(t) = \varphi(t, \widehat{x}(t), \widehat{u}(t))$ и аналогично для частных производных по x этих отображений.

Функцию

$$H(t, x, u, \lambda_0, p) = \langle p, \varphi(t, x, u) \rangle - \lambda_0 f(t, x, u),$$

где $\lambda_0 \in \mathbb{R}$ и $p \in (\mathbb{R}^n)^*$, называют *функцией Понтрягина* задачи (P_6).

Теорема (Необходимые условия минимума в задаче (P_7). Принцип максимума Понтрягина)**.** *Если пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ доставляет сильный минимум в задаче (P_7), то найдется $\lambda_0 \in \mathbb{R}$ и функция $p(\cdot) \in PC^1([t_0, t_1], (\mathbb{R}^n)^*)$, не равные одновременно нулю, такие, что в точках непрерывности $\widehat{u}(\cdot)$ выполнено условие стационарности по $x(\cdot)$:*

$$\dot{p}(t) = -p(t)\widehat{\varphi}_x(t) + \lambda_0 \widehat{f}_x(t)$$

и условие максимума по u:

$$\max_{u \in U} H(t, \widehat{x}(t), u, \lambda_0, p(t)) = H(t, \widehat{x}(t), \widehat{u}(t), \lambda_0, p(t)).$$

Принципом максимума Понтрягина обычно называют оба необходимых условия (и стационарность по $x(\cdot)$ и само условие максимума).

7 Необходимые условия слабого и сильного экстремума в простейшей задаче вариационного исчисления

В этом параграфе, опираясь на принцип максимума Понтрягина, будут получены классические необходимые условия экстремума Лежандра, Якоби и Вейерштрасса в простейшей задаче вариационного исчисления.

Все утверждения данного параграфа справедливы для векторного варианта простейшей задачи вариационного исчисления. Но, чтобы не усложнять выкладки, всюду далее будем иметь дело с классическим (скалярным) вариантом этой задачи, которая, напомним, имеет вид

$$J(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))dt \to \min, \quad x(t_i) = x_i, \quad i = 0, 1, \tag{Pr}$$

и, для определенности, рассматриваем задачу на минимум.

Понятие слабого минимума для этой задачи было дано выше.

В вариационном исчислении, наряду со слабым экстремумом, рассматривают еще и сильный экстремум, где близость функций измеряется в пространстве $C([t_0, t_1])$. Точнее говоря, обозначим через $PC^1([t_0, t_1])$ пространство кусочно-непрерывно-дифференцируемых функций на $[t_0, t_1]$. Допустимая функция $x(\cdot) \in PC^1([t_0, t_1])$ в задаче (Pr) определяется аналогично предыдущему.

Скажем, что допустимая функция $\widehat{x}(\cdot)$ доставляет *сильный минимум* в задаче (Pr), если существует $\varepsilon > 0$ такое, что для любой допустимой функции $x(\cdot)$, для которой $\|x(\cdot) - \widehat{x}(\cdot)\|_{C([t_0,t_1])} < \varepsilon$, выполняется неравенство $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$.

Теперь сформулируем условия Лежандра, Якоби и Вейерштрасса. Пусть $\widehat{x}(\cdot)$ — *экстремаль* (или *стационарная точка*) задачи (Pr), т. е. функция, для которой выполнено уравнение Эйлера. Предположим, что интегрант L — дважды непрерывно дифференцируемая функция.

Говорят, что на функции $\widehat{x}(\cdot)$ выполнено *условие Лежандра*, если $\widehat{L}_{\dot{x}\dot{x}}(t) = L_{\dot{x}\dot{x}}(t, \widehat{x}(t), \dot{\widehat{x}}(t)) \geq 0$ для всех $t \in [t_0, t_1]$ и *усиленное условие Лежандра*, если $\widehat{L}_{\dot{x}\dot{x}}(t) > 0$ для всех $t \in [t_0, t_1]$.

Пусть, далее, $h(\cdot) \in C^1([t_0, t_1])$, $h(t_0) = h(t_1) = 0$. Рассмотрим функцию (одного переменного)

$$\varphi(\lambda) = \int_{t_0}^{t_1} L(t, \widehat{x}(t) + \lambda h(t), \dot{\widehat{x}}(t) + \lambda \dot{h}(t))\, dt.$$

Тогда

$$\varphi'(\lambda) = \int_{t_0}^{t_1} (L_x(t, \widehat{x}(t) + \lambda h(t), \dot{\widehat{x}}(t) + \lambda \dot{h}(t))h(t) + L_{\dot{x}}(t, \widehat{x}(t) + \lambda h(t), \dot{\widehat{x}}(t) + \lambda \dot{h}(t))\dot{h}(t))\, dt$$

и дифференцируя еще раз, получаем, что

$$\begin{aligned}\varphi''(0) = \int_{t_0}^{t_1} \Big(\widehat{L}_{xx}(t)h^2(t) + \widehat{L}_{x\dot{x}}(t)h(t)\dot{h}(t) + \widehat{L}_{\dot{x}x}(t)\dot{h}(t)h(t) + \widehat{L}_{\dot{x}\dot{x}}(t)\dot{h}^2(t)\Big)dt \\ = \int_{t_0}^{t_1} \Big(\widehat{L}_{xx}(t)h^2(t) + 2\widehat{L}_{x\dot{x}}(t)h(t)\dot{h}(t) + \widehat{L}_{\dot{x}\dot{x}}(t)\dot{h}^2(t)\Big)dt.\end{aligned}$$

Функционал $\varphi''(0)$ (как функцию от $h(\cdot)$) обозначим через $Q(h(\cdot))$ и рассмотрим задачу

$$Q(h(\cdot)) \to \min, \quad h(t_0) = h(t_1) = 0.$$

Уравнение Эйлера для данной задачи имеет вид

$$-\frac{d}{dt}\Big(\widehat{L}_{\dot{x}x}(t)h(t) + \widehat{L}_{\dot{x}\dot{x}}(t)\dot{h}(t)\Big) + \widehat{L}_{xx}(t)h(t) + \widehat{L}_{x\dot{x}}(t)\dot{h}(t) = 0$$

и называется *уравнением Якоби* исходной задачи (Pr).

Пусть на $\widehat{x}(\cdot)$ выполнено усиленное условие Лежандра. Точка $\tau \in (t_0, t_1]$ называется *сопряженной точкой к* t_0, если существует нетривиальное решение $h(\cdot)$ уравнения Якоби, для которого $h(t_0) = h(\tau) = 0$.

Говорят, что на $\widehat{x}(\cdot)$ выполнено *условие Якоби*, если в интервале (t_0, t_1) нет точек сопряженных к t_0 и *усиленное условие Якоби*, если полуинтервал $(t_0, t_1]$ не содержит точек сопряженных к t_0.

Если L — дифференцируемая функция, то функция

$$\mathcal{E}(t, x, \dot{x}, u) = L(t, x, u) - L(t, x, \dot{x}) - L_{\dot{x}}(t, x, \dot{x})(u - \dot{x})$$

называется *функцией Вейерштрасса функционала* J.

Говорят, что на экстремали $\widehat{x}(\cdot)$ выполнено *условие Вейрштрасса*, если $\mathcal{E}(t, \widehat{x}(t), \dot{\widehat{x}}(t), u) \geq 0$ для всех $u \in \mathbb{R}$ и $t \in [t_0, t_1]$.

Теорема (Необходимые условия сильного минимума в задаче (Pr))**.** *Пусть функция* $\widehat{x}(\cdot) \in C^1([t_0, t_1])$ *доставляет сильный минимум в задаче* (Pr)*. Тогда, если интегрант* L *непрерывно дифференцируем, то*

(a) *для всех* $t \in [t_0, t_1]$ *выполнено уравнение Эйлера*

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0;$$

(b) *выполнено условие Вейерштрасса, т. е. для всех* $t \in [t_0, t_1]$ *и* $u \in \mathbb{R}$ *справедливо неравенсво*

$$\mathcal{E}(t, \widehat{x}(t), \dot{\widehat{x}}(t), u) \geq 0;$$

(c) *если существует* $\widehat{L}_{\dot{x}\dot{x}}(\cdot)$*, то выполнено условие Лежандра, т. е.* $\widehat{L}_{\dot{x}\dot{x}}(t) \geq 0$ *для всех* $t \in [t_0, t_1]$.

Доказательство. Запишем задачу (Pr) как задачу оптимального управления:

$$\int_{t_0}^{t_1} L(t, x(t), u(t))\, dt \to \min, \quad \dot{x} = u, \quad x(t_0) = x_0, \ \ x(t_1) = x_1. \qquad (Pr')$$

Легко видеть, что $\widehat{x}(\cdot)$ доставляет сильный минимум в задаче (Pr) тогда и только тогда, когда пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$, где $\widehat{u}(\cdot) = \dot{\widehat{x}}(\cdot)$ является сильным минимумом в задаче (Pr').

Согласно принципу максимума найдется ненулевая пара $(\lambda_0, p(\cdot)) \in \mathbb{R} \times PC^1([t_0, t_1])$ такая, что выполнено условие стационарности по $x(\cdot)$:

$$\dot{p}(t) = \lambda_0 \widehat{L}_x(t) \qquad (i)$$

и условие максимума по u:

$$\max_{u \in \mathbb{R}} (p(t)u - \lambda_0 L(t, \widehat{x}(t), u)) = p(t)\dot{\widehat{x}}(t) - \lambda_0 L(t, \widehat{x}(t), \dot{\widehat{x}}(t)), \quad \forall\, t \in [t_0, t_1]. \qquad (ii)$$

Если $\lambda_0 = 0$, то из (ii) последует, что функция $p(\cdot)$ равна нулю и тем самым пара $(\lambda_0, p(\cdot))$ нулевая. Итак, $\lambda_0 \neq 0$ и можно считать, что $\lambda_0 = 1$.

Условие (ii) означает, что для всех $t \in [t_0, t_1]$ функция

$$f(u) = p(t)u - L(t, \widehat{x}(t), u), \quad u \in \mathbb{R}$$

достигает максимума в точке $\dot{\widehat{x}}(t)$ и, следовательно, по теореме Ферма производная этой функции в данной точке равна нулю, т. е. $p(t) = \widehat{L}_{\dot{x}}(t)$. Вместе с (i) это дает уравнение Эйлера.

Необходимое условия минимума второго порядка функции f заключаются в том, что $\widehat{L}_{\dot{x}\dot{x}}(t) \geq 0$, т. е. выполнено условие Лежандра.

Из соотношения (ii) и доказанного равенства $p(t) = \widehat{L}_{\dot{x}}(t) = L_{\dot{x}}(t, \widehat{x}(t), \dot{\widehat{x}}(t))$ следует, что

$$L(t, \widehat{x}(t), u) - L_{\dot{x}}(t, \widehat{x}(t), \dot{\widehat{x}}(t))u \geq L(t, \widehat{x}(t), \dot{\widehat{x}}(t)) - L_{\dot{x}}(t, \widehat{x}(t), \dot{\widehat{x}}(t))\dot{\widehat{x}}(t)$$

для всех $u \in \mathbb{R}$ и $t \in [t_0, t_1]$ или

$$L(t, \widehat{x}(t), u) - L(t, \widehat{x}(t), \dot{\widehat{x}}(t)) - L_{\dot{x}}(t, \widehat{x}(t), \dot{\widehat{x}}(t))(u - \dot{\widehat{x}}(t)) \geq 0,$$

т. е. выполнено условие Вейерштрасса. $\square$

Теорема (Необходимые условия слабого минимума в задаче (Pr))**.** *Пусть* $\widehat{x}(\cdot) \in C^1([t_0, t_1])$ *доставляет слабый минимум в задаче* (Pr)*. Тогда, если интегрант* L *дважды непрерывно дифференцируемая функция, то для всех* $t \in [t_0, t_1]$ *выполнено уравнение Эйлера, условие Лежандра и условие Якоби.*