Вариационное исчисление и оптимальное управление (мехмат МГУ, лекции 2021 г.)

Г. Г. Магарил-Ильяев

Содержание

Введение

Какие причины побуждают решать задачи на максимум и минимум? Интерес к экстремальным задачам, т. е. задачам на максимум и минимум проявился уже на заре развития математики, и основными стимулами были любознательность и стремление к совершенству.

Среди наиболее ранних, точно решенных задач — так называемая *изопериметрическая задача* — задача о форме кривой заданной длины, охватывающей наибольшую площадь[1] и задача о форме поверхности заданной площади, охватывающей наибольший объем. Ответы на эти задачи для мыслителей Древней Греции были символами совершенства человеческого разума. Крупнейшие их представители: Евклид, Архимед и Аполлоний ставили и решали различные геометрические задачи на экстремум. Задача *о параллелограмме наибольшей площади, который можно вписать в треугольник* приводится в “Началах” Евклида (III в. до н. э.); задача *о шаровом сегменте максимального объема при заданной площади шаровой части поверхности этого сегмента* содержится в сочинениях Архимеда (тоже III в. до н. э.); задача *о минимальном расстоянии от точки плоскости до эллипса и о нормалях к эллипсу из произвольной точки плоскости* была поставлена и решена Аполлонием (III–II в. до н. э.) в его знаменитых “Кониках”.

Важная причина, побуждающая интерес к исследованию экстремальных задач, связана с тем, что многие законы природы, как оказалось, подчинены *экстремальным принципам*, т. е. многие природные процессы (неким загадочным образом) являются решениями задач на максимум и минимум. Л. Эйлер по этому поводу высказался так: “В мире не происходит ничего, в чем не был бы виден смысл какого-нибудь максимума или минимума”.

Первый экстремальный принцип в естествознании выдвинул Пьер Ферма (1662). Согласно этому принципу *свет “избирает” такую траекторию, вдоль которой время, затрачиваемое им на преодоление пути от одной точки до другой, минимально.* Законы преломления света, установленные экспериментально, находятся, исходя из этого принципа, как решения экстремальных задач. Отправляясь от принципа Ферма, И. Бернулли в 1696 году решил так называемую *задачу о брахистохроне* — о кривой наибыстрейшего ската, т. е. задачу о форме кривой, соединяющей две точки в вертикальной плоскости, вдоль которой тело под действием силы тяжести без трения проходит путь от одной точки до другой за кратчайшее время (постановка такой задачи возможно была навеяна более ранними размышлениями Галилея на эту тему). Историю развития вариационного исчисления принято отсчитывать именно с этого времени — года решения задачи о брахистохроне.

---

[1] Ответ в ней приводил в своих сочинениях еще Аристотель (IV в. до н. э.)

Нельзя не назвать также и прагматические причины, по которым приходится решать экстремальные задачи. Людям свойственно наилучшим образом использовать ресурсы, находящиеся в их распоряжении, и потому экстремальные задач естественно возникают при управлении различными процессами, в экономике, инженерии.

Первой задачей такого рода была *аэродинамическая задача Ньютона* о форме тела вращения, испытывающем наименьшее сопротивление в некоей "редкой среде", поставленная и решенная Ньютоном в его "Математических началах натуральной философии" (1687 г.). К числу экстремальных задач, возникающих в экономике и теории управления, относятся, например, *транспортная задача* (сформулированная в сороковые годы прошлого века) о наилучшем способе транспортировки продуктов со складов в магазины и *простейшая задача о быстродействии*, формализующая, в частности, задачу о наименьшем времени движения лифта в угольной шахте. Эта задача, также сформулированная в сороковые годы, послужила стимулом для создания *теории оптимального управления*. Считалось, что эта теория родилась в пятидесятые годы прошлого века, но когда она, в основном, сложилась, выяснилось, что первой задачей оптимального управления была именно аэродинамическая задача Ньютона.

Экстремальные задачи и их формализация. Мы видим, что зачастую экстремальные задачи ставятся на языке той области знаний, из которой они происходят. Для того, чтобы эти задачи исследовать математическими средствами, необходимо перевести их на математический язык, т. е. *формализовать*. Точно поставленная 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

$$f(x) \to \min(\max), \quad x \in C \qquad (P)$$

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*.

П. Ферма в 1638 году иллюстрировал свой метод нахождения экстремумов на решении следующей задачи: *найти прямоугольный треугольник наибольшей площади при данной сумме длин его катетов*. Если a — сумма длин катетов, а x — длина одного из них, то формализованная постановка выглядит так $x(a-x)/2 \to \max,\ 0 \le x \le a$.

При решении многих конкретных задач нашей целью будет нахождение глобальных экстремумов, но для этого предварительно приходиться исследовать задачу на наличие локальных экстремумов (т. е. локальных минимумов и максимумов). 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 $f(x) \ge f(\widehat{x})$ ($f(x) \le f(\widehat{x})$) для всех допустимых $x \in U$ (т. е. для всех $x \in C \cap U$).

Цель курса — изложить начала теории экстремума, состоящей из пяти основных глав: база теории (дифференциальное исчисление, обыкновенные дифференциальные уравнения, выпуклый анализ); необходимые условия экстремума; достаточные условия экстремума; теория существования решений экстремальных задач и алгоритмы поиска таких решений. Далее основное внимание будет уделено необходимым условиям экстремума.

1. **Аппарат теории экстремума**

В этом параграфе напоминаются необходимые для дальнейшего факты, связанные с пространством $\mathbb{R}^n$, с понятиями дифференцируемости функций и отображений на $\mathbb{R}^n$, формулируются конечномерные теоремы отделимости и доказывается основной результат этого параграфа — теорема об обратной функции для конуса. Кроме того, здесь много упражнений, которые полезно сделать.

1.1. **Дифференциальные свойства функций и отображений.** 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 $x = (x_1, \dots, x_n)^T$, 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 $a = (a_1, \dots, a_n)$ — вектор-строка из n действительных чисел. Для каждого $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ 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 $a = (l(e_1), \dots, l(e_n))$, где $e_1 = (1, 0, \dots, 0)^T, \dots, e_n = (0, \dots, 0, 1)^T$ — *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 $a = (a_1, \dots, a_n) \in (\mathbb{R}^n)^*$ 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).*

Далее, каждому $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ 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 $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$. 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 $\delta > 0$. Множество $U_{\mathbb{R}^n}(\widehat{x}, \delta) = \{\, x \in \mathbb{R}^n : |x - \widehat{x}| < \delta \,\}$ называется *открытым* шаром с центром в точке $\widehat{x}$ радиуса δ.

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

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

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

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 шар* (*иногда говорят замкнутый шар*) $B_{\mathbb{R}^n}(\widehat{x}, \delta) = \{\, x \in \mathbb{R}^n : |x - \widehat{x}| \leq \delta \,\}$ *с центром в точке* $\widehat{x}$ *радиуса* δ *— 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 $\operatorname{cl} \mathrm{A}$. Ясно, что $A \subset \operatorname{cl} \mathrm{A}$.

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

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 $\|\Lambda\| = \sup_{|x| \le 1} |\Lambda x|$[2], которое называется нормой оператора Λ. Простая проверка показывает, что выполняются следующие условия: (a) $\|\Lambda\| \ge 0$ 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) $\|\Lambda_1 + \Lambda_2\| \le \|\Lambda_1\| + \|\Lambda_2\|$ для любых $\Lambda_1, \Lambda_2 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Ясно также, что $|\Lambda x| \le \|\Lambda\||x|$ 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. 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, определяемый этим представлением однозначно (проверить!), называется *производной функции* f *в точке* $\widehat{x}$ и обозначается $f'(\widehat{x})$.

---

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

Из данного определения легко следует (беря в качестве h векторы $(h_1, 0, \dots, 0)^T, \dots, (0, \dots, 0, h_n)^T$), что a_i есть частная производная функции f по x_i в точке $\widehat{x}$, $1 \le i \le n$. Таким образом, $f'(\widehat{x}) = (\partial f(\widehat{x})/\partial x_1, \dots, \partial f(\widehat{x})/\partial x_n)$.

В классическом анализе анализе обозначают $h = (dx_1, \dots, dx_n)^T$ 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 $df(\widehat{x})$.

Ясно, что функция, дифференцируемая в данной точке, непрерывна в этой точке.

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

Упражнение 1.14. *Постройте функцию, у которой существуют частные производные в точке, но сама функция в этой точке разрывна.*

Упражнение 1.15. *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 $x \in \mathbb{R}^n$.*

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, определяемая этим представлением однозначно, называется *производной отображения F в точке $\widehat{x}$* и обозначается $F'(\widehat{x})$.

Упражнение 1.16. *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, $i = 1, \dots, m$, и отображение $F\colon U \to \mathbb{R}^m$ определено по правилу: $F(x) = (f_1(x), \dots, f_m(x))^T$, 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 $f_1'(\widehat{x}), \dots, f_m'(\widehat{x})$. 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 $F_y(\widehat{x},\widehat{y})$.

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*

$$F'(\widehat{x},\widehat{y})[\xi,\eta] = F_x(\widehat{x},\widehat{y})[\xi] + F_y(\widehat{x},\widehat{y})[\eta]^3$$

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, в силу единственности, следует, что линейный оператор из $\mathbb{R}^k$ в $\mathbb{R}^n$, действующий по правилу: $\xi \mapsto F'(\widehat{x},\widehat{y})[\xi,0]$ совпадает с $F_x(\widehat{x},\widehat{y})$. Аналогично, линейный оператор $\eta \mapsto F'(\widehat{x},\widehat{y})[0,\eta]$ из $\mathbb{R}^m$ в $\mathbb{R}^n$ совпадает с $F_y(\widehat{x},\widehat{y})$. Но тогда $F'(\widehat{x},\widehat{y})[\xi,\eta] = F'(\widehat{x},\widehat{y})([\xi,0]+[0,\eta]) = F_x(\widehat{x},\widehat{y})[\xi] + F_y(\widehat{x},\widehat{y})[\eta]$. □

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

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 $f'(x) \in (\mathbb{R}^n)^*$ и называемое *производной функции* f. Если производная f непрерывна в точке $\widehat{x}$ (на U), то говорят, что функция f *непрерывно дифференцируема в* $\widehat{x}$ (на U).

Упражнение 1.17*. *Доказать, что непрерывная дифференцируемость функции f на U равносильно тому, что все частные производные функции f непрерывны на U.*

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

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 $F'(x)\in\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$ и называемое *производной отображения* F. Если производная F непрерывна в точке $\widehat{x}$ (на U), то говорят, что отображение F *непрерывно дифференцируемо в* $\widehat{x}$ (на U).

Определим теперь понятие строгой дифференцируемости. 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, равносильное дифференцируемости F в точке $\widehat{x}$ (на языке $\varepsilon-\delta$) 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 1.18. *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 $f\colon\mathbb{R}\to\mathbb{R}$, $f(x)=x^2D(x)$, 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.

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 $[x,y]=\{z\in\mathbb{R}^n : z=(1-\alpha)x+\alpha y,\ 0\le z\le 1\}$ называется *отрезком*, соединяющий точки x и y. При $n=1,2$ и 3 — это обычный отрезок.

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

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

Доказательство. Если $F(x) = F(y)$, то утверждение теоремы выполняется очевидным образом. Пусть $F(x) \ne F(y)$. 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 $|F(y) - F(x)|^2 \le |F(y) - F(x)|\|F'(z)\||y - x|$ и тем самым $|F(y) - F(x)| \le \sup_{z\in[x,y]} \|F'(z)\||y - x|$. □

Следствие 1. *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 $|F(x_1) - F(x_2) - F'(\widehat{x})(x_1 - x_2)| \le \sup_{x\in[x_1,x_2]} \|F'(x) - F'(\widehat{x})\||x_1 - x_2| \le \varepsilon|x_1 - x_2|$. □

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.** 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).

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 открытые и замкнутые 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 $\mathrm{co\,A}$.

Упражнение 2.2. *Доказать, что* $\mathrm{co\,A}$ *состоит из выпуклых комбинаций элементов из A, т. е. векторов вида $\alpha_1x_1+\ldots+\alpha_mx_m$, где $x_i\in A$, $\alpha_i\geq 0$, $1\leq i\leq m$, и $\sum_{i=1}^m \alpha_i = 1$.*

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

Пусть A — непустое подмножество $\mathbb{R}^n$. Пересечение всех выпуклых конусов, содержащих A и начало координат называется *конической оболочкой множества A* и обозначается $\mathrm{cone\,A}$.

Упражнение 2.4. *Доказать, что* $\mathrm{cone\,A}$ *состоит из конических комбинаций элементов из A, т. е. векторов вида $\alpha_1x_1+\ldots+\alpha_mx_m$, где $x_i\in A$, $\alpha_i\geq 0$, $1\leq i\leq m$,* и является выпуклым множеством.

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

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

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

Обозначим $\beta = \min\{-\alpha_i/\mu_i : \mu_i < 0\}$. 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 $\alpha_i' \geq 0$, $i = 1, \ldots, k$ и среди чисел α_i' есть нулевые.

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

Следствие 2. *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. □

Перейдем к формулировкам теорем 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 2.5. *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.

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.*

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

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.*

1.3. **Теорема об обратной функции для конуса.** Напомним, что классический метод Ньютона решения нелинейного уравнения $F(x) = y$, где F — функция одного переменного и число y задано, 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. $x_k = x_{k-1} + (F'(x_{k-1}))^{-1}(y - F(x_{k-1})), \quad k \in \mathbb{N}$. При этом, x_0 стараются выбрать так, чтобы $F(x_0)$ было как можно ближе к y.

Если $x \in \mathbb{R}^n$ и $y \in \mathbb{R}^m$, то для решения уравнения $F(x) = y$ мы будем использовать следующую модификацию метода Ньютона: $x_k = x_{k-1} + R(y - F(x_{k-1})),\ k \in \mathbb{N}$, где R — правый обратный к производной отображения F в некоторой точке $\widehat{x}$.

Теорема (об обратной функции для конуса)**.** *Пусть U — окрестность точки $\widehat{x} \in \mathbb{R}^n$, C — выпуклый замкнутый конус в $\mathbb{R}^n$, отображение $F\colon U \to \mathbb{R}^m$ строго дифференцируемо в $\widehat{x}$ и $F'(\widehat{x})(C) = \mathbb{R}^m$. Тогда существуют окрестность W точки $F(\widehat{x})$, отображение $\varphi\colon W \to U \cap (\widehat{x} + C)$ и константа $K > 0$ такие, что $F(\varphi(y)) = y$ и $|\varphi(y) - \widehat{x}| \le K|y - F(\widehat{x})|$ для всех $y \in W$.*

Доказательство. Покажем сначала, что существуют отображение $R\colon \mathbb{R}^m \to C$ и константа $\gamma > 0$ такие, что $F'(0)R(y) = y$ и $|R(y)| \le \gamma|y|$ для всех $y \in \mathbb{R}^m$.

Действительно, пусть $e_1, \dots, e_m$ — базис в $\mathbb{R}^m$. По предположению найдутся такие $f_i \in C$, $1 \le i \le 2m$, что $F'(\widehat{x})f_i = e_i$ и $F'(\widehat{x})f_{m+i} = -e_i$, $1 \le i \le m$. 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 $y_i < 0$. Из выпуклости C следует, что $R(y) \in C$. Далее, $F'(\widehat{x})R(y) = \sum_{i=1}^m |y_i| F'(\widehat{x}) g_i = \sum_{i=1}^m y_i e_i = y$ 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 нужное отображение построено.

В силу строгой дифференцируемости отображения F в точке $\widehat{x}$ найдется $\delta = \delta(\gamma) > 0$ такое, что $U_{\mathbb{R}^n}(\widehat{x}, \delta) \subset U$ и для всех $x, x' \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$ справедливо неравенство

$$|F(x) - F(x') - F'(\widehat{x})(x - x')| \le \frac{1}{2\gamma}|x - x'|. \qquad (i)$$

Положим $W = U_{\mathbb{R}^m}(F(\widehat{x}), \delta/2\gamma)$ и пусть $y \in W$. Рассмотрим последовательность (модифицированный метод Ньютона)

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

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

Используя последовательно (ii), оценку для R, равенство

$$F'(\widehat{x})(x_k-x_{k-1})-y+F(x_{k-1})=0, \qquad (iii)$$

которое следует из (ii) после применения к обеим частям оператора $F'(\widehat{x})$, (i) и затем итерируя процедуру, будем иметь

$$|x_{k+1}-x_k|=|R(y-F(x_k))|\le\gamma|y-F(x_k)|=\gamma|F(x_k)-F(x_{k-1}) \\ -F'(\widehat{x})(x_k-x_{k-1})|\le\frac{1}{2}|x_k-x_{k-1}|\le\ldots\le\frac{1}{2^k}|x_1-\widehat{x}|. \quad (iv)$$

Далее применяя неравенство треугольника, (iv), формулу для суммы геометрической прогрессии, (ii) (при $k=1$), оценку для правого обратного и учитывая выбор y, получим, что

$$|x_{k+1}-\widehat{x}|\le|x_{k+1}-x_k|+\ldots+|x_1-\widehat{x}|\le\left(\frac{1}{2^k}+\frac{1}{2^{k-1}}+\ldots+1\right)|x_1-\widehat{x}| \\ <2\gamma|y-F(\widehat{x})|<\delta, \quad (v)$$

т. е. $x_{k+1}\in U_{\mathbb{R}^n}(\widehat{x},\delta)$. Так как по предположению индукции $x_k\in\widehat{x}+C$, то отсюда и из (ii) следует, что $x_{k+1}\in x_k+C\subset\widehat{x}+C+C\subset\widehat{x}+C$, т. е. $x_{k+1}\in U_{\mathbb{R}^n}(\widehat{x},\delta)\cap(\widehat{x}+C)$ 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

$$|x_{k+l}-x_k|\le|x_{k+l}-x_{k+l-1}|+\ldots+|x_{k+1}-x_k| \\ \le\left(\frac{1}{2^{k+l-1}}+\ldots+\frac{1}{2^k}\right)|x_1-\widehat{x}|<\frac{\gamma}{2^{k-1}}|y-F(\widehat{x})|<\frac{\delta}{2^k}. \quad (vi)$$

Положим $\varphi(y)=\lim_{k\to\infty}x_k$. Тогда из (v) следует, переходя к пределу при $k\to\infty$, что $\varphi(y)\in U_{\mathbb{R}^n}(\widehat{x},\delta)$. Поскольку конус C замкнут, то $\varphi(y)\in U_{\mathbb{R}^n}(\widehat{x},\delta)\cap(\widehat{x}+C)\subset U\cap(\widehat{x}+C)$.

Из (i) вытекает непрерывность F на $U_{\mathbb{R}^n}(\widehat{x},\delta)$. Переходя к пределу в (iii) при $k\to\infty$, получаем, что $F(x(y))=y$, а переход к пределу в (v) приводит к оценке $|x(y)-\widehat{x}|\le K|y-F(\widehat{x})|$, где $K=2\gamma$. □

2. Необходимые условия экстремума в гладких задачах

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 $f : U \to \mathbb{R}$. Рассмотрим задачу

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

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 а $\widehat{x}$ следует, что для достаточно малых t справедливо представление $f(\widehat{x} + tx) = f(\widehat{x}) + t(\langle f'(\widehat{x}), x\rangle + o(t)/t)$. 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. □

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.

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

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

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

Свяжем с задачей (P_2) следующую функцию $\mathcal{L}\colon U \times (\mathbb{R}^{m+1})^* \to \mathbb{R}$, $\mathcal{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x)$, 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 (Правило множителей Лагранжа для задачи (P_2))**.** *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 $f_1'(\widehat{x}), \ldots, f_m'(\widehat{x})$ линейно независимы, то $\lambda_0 \neq 0$.

Доказательство. Необходимые условия в задаче (P_2) заключаются в том, что векторы $f_0'(\widehat{x}), f_1'(\widehat{x}), \ldots, f_m'(\widehat{x})$ 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 $F(x) = (f_0(x), f_1(x), \ldots, f_m(x))^T$. Векторы $f_0'(\widehat{x}), f_1'(\widehat{x}), \ldots, f_m'(\widehat{x})$ образуют строки матрицы $F'(\widehat{x})$. Так как эти векторы линейно независимы, то $F'(\widehat{x})(\mathbb{R}^n) = \mathbb{R}^{m+1}$. Далее, так как функции $f_0, f_1, \ldots, f_m$ строго дифференцируемы в точке $\widehat{x}$, то отображение F строго дифференцируемо в этой точке. Применим теорему об обратной функции в ситуации, когда $C = \mathbb{R}^n$.

Согласно этой теореме найдутся окрестность W точки $F(\widehat{x})$, отображение $\varphi\colon W \to U$ и константа $K > 0$ такие, что $F(\varphi(y)) = y$ и $|\varphi(y) - \widehat{x}| \leq K|y - F(\widehat{x})|$ для всех $y \in W$. Поскольку $F(\widehat{x}) = (f_0(\widehat{x}), 0, \ldots, 0)^T$, то для достаточно малых по модулю

ν точки $y_\nu = (f_0(\widehat{x}) + \nu, 0, \dots, 0)^T$, очевидно, принадлежат W. Обозначим $x_\nu = \varphi(y_\nu)$. Равенство $F(x_\nu) = y_\nu$ 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 $f_0(x_\nu) = f_0(\widehat{x}) + \nu$, $f_i(x_\nu) = 0$, $i = 1, \dots, m$, 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 $\nu < 0$ $(\nu > 0)$. Кроме того, $|\varphi(y_\nu) - \widehat{x}| \le K|y_\nu - F(\widehat{x})|$, или $|x_\nu - \widehat{x}| \le K|\nu|$. Следовательно, 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 утверждение проверяется без труда. $\square$

3. Задачи с ограничениями, задаваемые равенствами неравенствами. Пусть U — открытое подмножество $\mathbb{R}^n$ и $f_i\colon U \to \mathbb{R}$, $i = 0, 1, \dots, m$. Задача

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

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

Теорема (Правило множителей Лагранжа для задачи (P_3))**.** *Пусть точка $\widehat{x} \in U$ является локальным минимумом в задаче (P_3). Если функции f_i, $i = 0, 1, \dots, m$, строго дифференцируемы в $\widehat{x}$, то найдется такой ненулевой набор множителей Лагранжа $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_m)$, 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*) $\lambda_i \ge 0,\ i = 0, 1, \dots, m'$ (*условие неотрицательности*);

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

Доказательство теоремы 4. Заметим сначала, что утверждение (c) можно считать выполненным всегда. В самом деле, отбросим те ограничения среди неравенств, для которых $f_i(\widehat{x}) < 0$. Тогда $\widehat{x}$ будет локальным экстремумом и в новой задаче (проверьте!). Если для этой 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!). Возможно одно из двух: либо $F'(\widehat{x},0)(\mathbb{R}^n \times \mathbb{R}_+^{m'+1}) = \mathbb{R}^{m+1}$, либо нет. Покажем, что первая возможность приводит к противоречию с тем, что $\widehat{x}$ — локальный минимум. В самом деле, в этом случае выполнены условия теоремы 1.3 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 $C = \mathbb{R}^n \times \mathbb{R}_+^{m'+1}$ (C — произведение замкнутых конусов и поэтому есть замкнутый конус).

Пусть W — окрестность точки $F(\widehat{x}, 0)$, отображение $\varphi\colon W \to (U \times \mathbb{R}^{m'+1}) \cap ((\widehat{x}, 0) + (\mathbb{R}^n \times \mathbb{R}_+^{m'+1})) = (U \times \mathbb{R}^{m'+1}) \cap (\mathbb{R}^n \times \mathbb{R}_+^{m'+1}) = U \times \mathbb{R}_+^{m'+1}$ и константа K из теоремы. Заметим, что по предположению $f_i(\widehat{x}) = 0$, $i = 1, \ldots, m'$, и поэтому $F(\widehat{x}, 0) = (f_0(\widehat{x}), 0, \ldots, 0)^T$. Очевидно, что для всех достаточно малых $\varepsilon > 0$ точки $y_\varepsilon = (f_0(\widehat{x}) - \varepsilon, 0 \ldots, 0)^T$ принадлежат W. Тогда, согласно теореме, $F(\varphi(y_\varepsilon)) = y_\varepsilon$. Поскольку $\varphi(y_\varepsilon) \in U \times \mathbb{R}_+^{m'+1}$, то $\varphi(y_\varepsilon) = (x(y_\varepsilon), u(y_\varepsilon))$, где $x(y_\varepsilon) \in U$ и $u(y_\varepsilon) = (u_0(y_\varepsilon), u_1(y_\varepsilon), \ldots, u_{m'}(y_\varepsilon)) \in \mathbb{R}_+^{m'+1}$. Следовательно, равенство $F(\varphi(y_\varepsilon)) = y_\varepsilon$ 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 $u_i(y_\varepsilon) \geq 0$, $i = 1, \ldots, m'$, то $f_i(x(y_\varepsilon)) \leq 0$, $i = 1, \ldots, m'$ и тем самым точки $x(y_\varepsilon)$ допустимы в задаче (P_3). Кроме того, $|x(y_\varepsilon) - \widehat{x}| \leq |(x(y_\varepsilon), u(y_\varepsilon)) - (\widehat{x}, 0)| = |\varphi(y_\varepsilon) - (\widehat{x}, 0)| \leq K|y_\varepsilon - F(\widehat{x}, 0)| = K\varepsilon$ и $f_0(x(y_\varepsilon)) \leq f_0(x(y_\varepsilon)) + u_0(y_\varepsilon) = f_0(\widehat{x}) - \varepsilon < f_0(\widehat{x})$. 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, $F'(\widehat{x}, 0)(\mathbb{R}^n \times \mathbb{R}_+^{m'+1}) \neq \mathbb{R}^{m+1}$. Ясно, что $C_1 = F'(\widehat{x}, 0)(\mathbb{R}^n \times \mathbb{R}_+^{m'+1})$ — выпуклый конус в $\mathbb{R}^{m+1}$. Пусть $y_0 \notin C_1$. По первой теореме отделимости (теорема 1.2) найдется ненулевой вектор $\lambda =$

$(\lambda_0, \lambda_1, \dots, \lambda_m) \in (\mathbb{R}^{m+1})^*$ такой, что $\langle \lambda, y_0 \rangle \le \inf_{y \in C_1} \langle \lambda, y \rangle$. Отсюда легко следует, что $\langle \lambda, y \rangle \ge 0$ для всех $y \in C_1$. Это равносильно тому, что

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

Полагая здесь $u = 0$, получаем, что линейный функционал $\sum_{i=0}^{m} \lambda_i f_i'(\widehat{x})$ неотрицателен на X 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 сразу вытекает утверждение (b). $\square$

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 $[x, y] = \{\, z \in X \mid z = (1-\alpha)x + \alpha y,\ 0 \le \alpha \le 1 \,\}$, 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: $x \mapsto ax^2 + bx + c$, $a \geq 0$; $x \mapsto e^{\alpha x}$, $\alpha \in \mathbb{R}$; $x \mapsto |x|^p$, $p \geq 1$; $x \mapsto -\ln x$, если $x > 0$ и $+\infty$, если $x \leq 0$; $x \mapsto x \log_2 x + (1-x)\log_2(1-x)$, если $0 < x < 1$ и $+\infty$ в остальных случаях.

Упражнение 4.1. *Доказать, что функция $f \colon X \to \overline{\mathbb{R}}$ выпукла тогда и только тогда, когда для любых $x_1, x_2 \in X$ и любых $\alpha_i > 0$, $i = 1, 2$, таких, что $\alpha_1 + \alpha_2 = 1$ справедливо неравенство*

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

которое называется неравенством Иенссена.

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)

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

Упражнение 4.2. *Доказать, что если субдифференциал не пуст, то это выпуклое замкнутое множество.*

Упражнение 4.3. *Привести пример функции, у которой субдифференциал в данной точке пуст.*

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

Предложение 2. *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, $x^* = f'(\widehat{x})$. $\square$

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 $f(\widehat{x}) \leq f(x)$ для всех $x \in U$. Пусть теперь x — произвольная точка из $\mathbb{R}^n$. Для достаточно малых $0 < \alpha \leq 1$ точки $(1-\alpha)\widehat{x} + \alpha x$ принадлежат U и поэтому (по неравенству Иенссена) $f(\widehat{x}) \leq f((1-\alpha)\widehat{x}+\alpha x) \leq (1-\alpha)f(\widehat{x})+\alpha f(x)$, откуда следует, что $f(\widehat{x}) \leq f(x)$.

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 $f(x) - f(\widehat{x}) \geq 0 = 0 \cdot x$ для любого $x \in \mathbb{R}^n$, т. е. $0 \in \partial f(\widehat{x})$. Если $0 \in \partial f(\widehat{x})$, то $f(x) - f(\widehat{x}) \geq \langle 0, x\rangle = 0$, т. е. $f(x) \geq f(\widehat{x})$ 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

$$f_0(x) \to \min, \quad f_i(x) \leq 0, \quad 1 \leq i \leq m, \quad x \in A \tag{P_5}$$

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

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

Если найдется точка $\overline{x} \in A$ *такая, что* $f_i(\overline{x}) < 0$, $1 \leq i \leq m$, *то* $\lambda_0 \neq 0$ (*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 $B = \{\, b = (b_0, b_1, \ldots, b_m)^T \in \mathbb{R}^{m+1} \mid \exists\, x \in A : f_0(x) - f_0(\widehat{x}) \leq b_0,\ f_i(x) \leq b_i,\ i = 1, \ldots, m \,\}$ и $C = \{\, c \in \mathbb{R}^{m+1} \mid c = (c_0, 0, \ldots, 0)^T,\ c_0 < 0 \,\}$. 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 $B \neq \emptyset$. Докажем выпуклость B. Пусть $b = (b_0, \ldots, b_m)^T, b' = (b'_0, \ldots, b'_m)^T \in B$ и $\alpha \in [0, 1]$. Найдутся такие $x, x' \in A$, что $f_0(x) - f_0(\widehat{x}) \leq b_0$, $f_i(x) \leq b_i$, $i = 1, \ldots, m$ и $f_0(x') - f_0(\widehat{x}) \leq b'_0$, $f_i(x') \leq b'_i$, $i = 1, \ldots, m$. Положим $x_\alpha = (1 - \alpha)x + \alpha x'$. Тогда $x_\alpha \in A$. Далее $f_0(x_\alpha) - f_0(\widehat{x}) = f_0((1 - \alpha)x + \alpha x') - f_0(\widehat{x}) \leq (1-\alpha) f_0(x) + \alpha f_0(x') - f_0(\widehat{x}) \leq (1-\alpha) b_0 + \alpha b'_0$ и аналогично $f_i(x_\alpha) \leq (1-\alpha) b_i + \alpha b'_i$, $i = 1, \ldots, m$, т. е. $(1-\alpha) b + \alpha b' \in B$ и тем самым B выпукло.

Если допустить, что $B \cap C \neq \emptyset$, то найдется элемент $\widetilde{x} \in A$ такой, что $f_i(\widetilde{x}) \leq 0$, $i = 1, \ldots, m$, и $f_0(\widetilde{x}) - f_0(\widehat{x}) \leq c_0 < 0$, 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 (теорема 1.2) найдется такой ненулевой вектор $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in (\mathbb{R}^{m+1})^*$, что $\sup_{c \in C} \sum_{i=0}^m \lambda_i c_i \leq \inf_{b \in B} \sum_{i=0}^m \lambda_i b_i$, или $\sup_{c_0 < 0} \lambda_0 c_0 \leq \inf_{b \in B} \sum_{i=0}^m \lambda_i b_i$. Отсюда следует, что

$$\sum_{i=0}^{m} \lambda_i b_i \geq 0 \quad \forall\, b = (b_0, b_1, \ldots, b_m)^T \in B. \qquad (i)$$

Подставляя в (i) векторы $(1, 0, \ldots, 0)^T, \ldots, (0, \ldots, 0, 1)^T$, получаем утверждение (b).

Теперь подставим в (i) векторы $(0, \ldots, 0, f_i(\widehat{x}), 0, \ldots, 0)^T$, $i = 1, \ldots, m$ (они принадлежат B, надо взять $x = \widehat{x}$). Тогда получим, что $\lambda_i f_i(\widehat{x}) \geq 0$. Но $\lambda_i f_i(\widehat{x}) \leq 0$, так как $\lambda_i \geq 0$, а $f_i(\widehat{x}) \leq 0$ и поэтому $\lambda_i f_i(\widehat{x}) = 0$, $i = 1, \ldots, m$, 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 (i), 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 $\mathcal{L}(x, \overline{\lambda}) \geq \mathcal{L}(\widehat{x}, \overline{\lambda})$ и (a) доказано.

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

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

Деля на λ_0, получаем требуемое. □

4. **Необходимые условия экстремума в задачах вариационного исчисления**

Принято считать, что 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, которого он консультировал по научным вопросам) 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. Там, в частности, была рассмотрена задача, которая ныне называется простейшей задачей (классического) вариационного исчисления. В данном параграфе выводятся необходимые условия экстремума в этой задаче, 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), а также в задаче Лагранжа.

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

$$J(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))\, dt \to \text{extr}, \quad x(t_0) = x_0, \quad x(t_1) = x_1 \tag{P_6}$$

называется *простейшей задачей (классического) вариационного исчисления.*

Уточним постановку. 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* (P_6), если $x(t_i) = x_i$, $i = 0, 1$.

Допустимая функция $\widehat{x}(\cdot)$ называется *слабым локальным минимумом (максимумом)* в задаче (P_6), 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 $J(x(\cdot)) \ge J(\widehat{x}(\cdot))$ $(J(x(\cdot)) \le J(\widehat{x}(\cdot)))$. *Слабый локальный экстремум* — это либо слабый локальный минимум, либо слабый локальный максимум.

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

Теорема (Необходимые условия экстремума в задаче (P_6))**.** *Пусть $\widehat{x}(\cdot)$ доставляет слабый локальный экстремум в задаче (P_6). 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 (P_6). 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 $\dot{p}(\cdot) = \widehat{L}_x(\cdot)$. Подставляя это в (i) и интегрируя по частям, приходим к соотношению

$$\int_{t_0}^{t_1} \left(-p(t) + \widehat{L}_{\dot{x}}(t)\right) \dot{x}(t)\, dt = 0,$$

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

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

Выберем $c = c_0$ так, чтобы $\int_{t_0}^{t_1}(-p(t)+\widehat{L}_{\dot{x}}(t)+c_0)\, dt = 0$. Рассмотрим функцию $x(t) = \int_{t_0}^{t}(-p(\tau)+\widehat{L}_{\dot{x}}(\tau)+c_0)\, d\tau$ (ясно, что $x(\cdot) \in C^1([t_0, t_1])$ и $x(t_0) = x(t_1) = 0$) и подставим ее (ii) с $c = c_0$. Тогда получим

$$\int_{t_0}^{t_1} \left(-p(t) + \widehat{L}_{\dot{x}}(t) + c_0\right)^2 dt = 0.$$

Отсюда следует, что $-p(\cdot) + \widehat{L}_{\dot{x}}(\cdot) + c_0 = 0$, откуда 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 в предыдущем случае, $[t_0, t_1]$ — 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 ξ_0 и ξ_1. Задача

$$\mathcal{B}(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))\, dt + l(x(t_0), x(t_1)) \to \mathrm{extr} \qquad (P_6')$$

называется *задачей Больца.*

Любая функция $x(\cdot) \in C^1([t_0, t_1])$ допустима в данной задаче.

Слабый локальный экстремум определяется аналогично предыдущему случаю.

Функции $\widehat{L}_x(\cdot)$ и $\widehat{L}_{\dot{x}}(\cdot)$ определяются как и раньше, и кроме того, полагаем $\widehat{l}_{\xi_i} = l_{\xi_i}(\widehat{x}(t_0), \widehat{x}(t_1))$, $i = 0, 1$.

Теорема (Необходимые условия экстремума в задаче (P_6'))**.** *Пусть $\widehat{x}(\cdot)$ доставляет слабый локальный экстремум в задаче (P_6'). Тогда, если функция L непрерывна вместе со своими частными производными по x и $\dot{x}$, а функция l непрерывна вместе со своими частными производными по ξ_0 и ξ_1, 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:

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

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

$$\left.\frac{d\mathcal{B}(x_\alpha(\cdot))}{d\alpha}\right|_{\alpha=0} = \int_{t_0}^{t_1} \Big(\widehat{L}_x(t)x(t) + \widehat{L}_{\dot{x}}(t)\dot{x}(t)\Big)\, dt + \widehat{l}_{\xi_0} x(t_0) + \widehat{l}_{\xi_1} x(t_1) = 0. \qquad (i)$$

Пусть $p(\cdot)$ — решение задачи Коши $\dot{p} = \widehat{L}_x(t)$, $p(t_1) = -\widehat{l}_{\xi_1}$ (т. е. $p(t) = -\widehat{l}_{\xi_1} - \int_t^{t_1} \widehat{L}_x(\tau)\, d\tau$, $t \in [t_0, t_1]$. 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 $x(\cdot)$ — решение задачи Коши $\dot{x} = -p(t) + \widehat{L}_{\dot{x}}(t)$, $x(t_0) = \widehat{l}_{\xi_0} - p(t_0)$ (т. е. $x(t) = \widehat{l}_{\xi_0} - p(t_0) + \int_{t_0}^t (-p(\tau) + \widehat{L}_{\dot{x}}(\tau))\, d\tau$, $t \in [t_0, t_1]$). 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 $p(\cdot)$. □

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

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

который называется *интегралом импульса.*

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

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

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,

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

4.3. **Задача Лагранжа.** Пусть $[t_0, t_1]$ — конечный отрезок, функция $f\colon \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ и отображение $\varphi\colon \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ (переменных $t \in \mathbb{R}$, $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, $u = (u_1, \ldots, u_r)^T \in \mathbb{R}^r$) непрерывны на $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^r$ и $x_i \in \mathbb{R}^n$, $i = 0, 1$. Задача

$$J(x(\cdot), u(\cdot)) = \int_{t_0}^{t_1} f(t, x(t), u(t))\, dt \to \text{extr}, \quad \dot{x} = \varphi(t, x, u),$$
$$x(t_0) = x_0, \quad x(t_1) = x_1. \quad (P_7)$$

называется *задачей Лагранжа вариационного исчисления.*

Уточним постановку. Пара $(x(\cdot), u(\cdot))$ называется *допустимой в задаче* (P_6), если $(x(\cdot), u(\cdot)) \in Z = C^1([t_0, t_1], \mathbb{R}^n) \times C([t_0, t_1], \mathbb{R}^r)$,[4], $\dot{x}(t) = \varphi(t, x(t), u(t))$ для всех $t \in [t_0, t_1]$ и $x(t_i) - x_i$, $i = 0, 1$.

[4] Z — нормированное пространство с нормой $\|(x(\cdot), u(\cdot))\|_Z = \|x(\cdot)\|_{C^1([t_0,t_1],\mathbb{R}^n)} + \|u(\cdot)\|_{C([t_0,t_1],\mathbb{R}^n)}$.

Допустимая пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ называется *слабым локальным минимумом (максимумом) в задаче* (P_7), если существует такое $\varepsilon > 0$, что для любой допустимой пары $(x(\cdot), u(\cdot))$, для которой $\|(x(\cdot), u(\cdot)) - (\widehat{x}(\cdot), \widehat{u}(\cdot))\|_Z < \varepsilon$ выполнено неравенство $J(x(\cdot), u(\cdot)) \geq J(\widehat{x}(\cdot), \widehat{u}(\cdot))$ $(J(x(\cdot), u(\cdot)) \leq J(\widehat{x}(\cdot), \widehat{u}(\cdot)))$.

Заметим, что простейшая задача классического вариационного исчисления есть частный случай данной задачи, когда $\varphi(t, x, u) = u$ и легко видеть, что если $\widehat{x}(\cdot)$ — слабый экстремум в простейшей задаче, то это равносильно тому, что $(\widehat{x}(\cdot), \widehat{u}(\cdot))$, где $\widehat{u}(\cdot) = \dot{\widehat{x}}(\cdot)$, — слабый экстремум в той же задаче, но записанной в форме задачи (P_7).

Функция Лагранжа задачи (P_7) имеет вид

$$\mathcal{L}(x(\cdot), u(\cdot), \lambda) = \int_{t_0}^{t_1} (\lambda_0 f(t, x(t), u(t)) + \langle p(t), \dot{x}(t) - \varphi(t, x(t), u(t))\rangle)\, dt,$$

где $\lambda = (\lambda_0, p(\cdot))$ — *набор множителей Лагранжа*. Функцию под знаком интеграла обозначим через

$$L = L(t, x, \dot{x}, u, \lambda_0, p) = \lambda_0 f(t, x, u) + \langle p, \dot{x} - \varphi(t, x, u)\rangle.$$

Если фиксирована пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$, то, как и раньше, для сокращения записи используем обозначения: $\widehat{L}_x(t) = L_x(t, \widehat{x}(t), \widehat{u}(t))$ и аналогично для частной производной по u, частных производных отображения f и φ.

Теорема (Необходимые условия минимума в задаче (P_7))**.** *Пусть пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ доставляет слабый локальный экстремум в задаче (P_7). Тогда, если функция f и отображение φ непрерывны вместе со своими частными производными по x и u, то существует такой ненулевой набор множителей Лагранжа $\lambda = (\lambda_0, p(\cdot)) \in \mathbb{R} \times C^1([t_0, t_1], (\mathbb{R}^n)^*)$, что $\widehat{L}_{\dot{x}}(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n)$, для всех $t \in [t_0, t_1]$ выполнено уравнение Эйлера по x :*

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0 \;\Leftrightarrow\; -\dot{p} = p\widehat{\varphi}_x(t) - \lambda_0 \widehat{f}_x(t)$$

и уравнение Эйлера по u :

$$\widehat{L}_u(t) = 0 \;\Leftrightarrow\; p(t)\widehat{\varphi}_u(t) = \lambda_0 \widehat{f}_u(t).$$

Для доказательства этой теоремы, помимо теоремы о разрешимости системы нелинейных уравнений и теоремы отделимости, понадобится еще стандартная теорема о дифференцируемой зависимости решений дифференциальных уравнений от параметров, которую здесь приводим без доказательства и в удобной для нас форме.

Теорема (о дифференцируемой зависимости)**.** *Пусть* $u_i(\cdot) \in C([t_0,t_1],\mathbb{R}^r)$, $i=1,\ldots,k$, $\overline{u}(\cdot)=(u_1(\cdot),\ldots,u_k(\cdot))$ *и* $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ *— допустимая пара в задаче* (P_7). *Тогда найдется такая окрестность нуля* V *в* $\mathbb{R}^k$, *что для любого* $\overline{\alpha}=(\alpha_1,\ldots,\alpha_k)^T \in V$ *существует единственное решение* $x(\cdot,\overline{\alpha},\overline{u}(\cdot))$ *задачи Коши*

$$\dot{x}=\varphi(t,x,\widehat{u}(t)+\sum_{i=1}^{k}\alpha_i u_i(t)), \quad x(t_0)=x_0,$$

определенное на всем отрезке $[t_0,t_1]$.

Кроме того, $x(\cdot,\overline{\alpha},\overline{u}(\cdot)) \to \widehat{x}(\cdot)$ *при* $\overline{\alpha} \to 0$ *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 по* α_i *в нуле на отрезке* $[t_0,t_1]$ *удовлетворяют дифференциальным уравнениям*

$$\dot{y}=\widehat{\varphi}_x(t)y+\widehat{\varphi}_u(t)u_i(t), \quad y(t_0)=0, \quad 1\le i\le k. \tag{i}$$

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

$$\dot{x}=\varphi(t,x,u(t,\alpha,u(\cdot)), \quad x(t_0)=x_0.$$

Если $t \in [t_0,t_1]$, то производную отображения $\alpha \mapsto x(t,\alpha,u(\cdot))$ в нуле обозначим $y(t,u(\cdot))$.

В силу стандартных фактов о дифференцируемости интеграла по параметру, функция $\alpha \mapsto J(x(\cdot,\alpha,u(\cdot)),u(\cdot,\alpha,u(\cdot)))$ непрерывно дифференцируема в некоторой окрестности нуля и ее производная в нуле (которую обозначим $y_0(u(\cdot))$) имеет вид

$$y_0(u(\cdot))=\int_{t_0}^{t_1}(\langle\widehat{f}_x(t),y(t,u(\cdot))\rangle+\langle\widehat{f}_u(t),u(t)\rangle)\,dt. \tag{ii}$$

Покажем, что предположение о том, что $\operatorname{span}\{\,(y_0(u(\cdot)),y(t_1,u(\cdot)))^T \in \mathbb{R}^{n+1} : u(\cdot) \in C([t_0,t_1],\mathbb{R}^r)\,\}=\mathbb{R}^{n+1}$ приводит к противоречию с тем, что $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ — локальный экстремум.

Действительно, пусть данное предположение верно. Тогда существуют такие $u_i(\cdot) \in C([t_0,t_1],\mathbb{R}^r)$, $i=1,\ldots,n+1$, что векторы $(y_0(u_i(\cdot)),y(t_1,u_i(\cdot)))^T$, $i=1,\ldots,n+1$, образуют базис в $\mathbb{R}^{n+1}$. Положим $\overline{u}(\cdot)=(u_1(\cdot),\ldots,u_{n+1}(\cdot))$, $\overline{\alpha}=(\alpha_1,\ldots,\alpha_{n+1})^T$ и $u(\cdot,\overline{\alpha},\overline{u}(\cdot))=\widehat{u}(\cdot)+\sum_{i=1}^{n+1}\alpha_i u_i(\cdot)$. В силу теоремы о дифференцируемой зависимости для достаточно малых $\overline{\alpha}$ существует 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, для малых $\overline{\alpha}$ функция $\overline{\alpha}\mapsto J(x(\cdot,\overline{\alpha},\overline{u}(\cdot)),u(\cdot,\overline{\alpha},\overline{u}(\cdot)))$ определена и непрерывно дифференцируема. Таким образом, существует окрестность V нуля в $\mathbb{R}^{n+1}$, что корректно определено отображение $F\colon V\to\mathbb{R}^{n+1}$, сопоставляющее $\overline{\alpha}\in V$ вектор $(J(x(\cdot,\overline{\alpha},\overline{u}(\cdot)),u(\cdot,\overline{\alpha},\overline{u}(\cdot))),x(t_1,\overline{\alpha},\overline{u}(\cdot)))^T$. Это отображение непрерывно дифференцируемо на V и его производная (матрица Якоби) в нуле невырождена, поскольку ее столбцы суть векторы $(y_0(u_i(\cdot)),y(t_1,u_i(\cdot)))$, $i=1,\ldots,n+1$.

Теперь применим теорему об обратной функции для конуса (§1, п. 1.3), где n и m надо заменить на $n+1$ и считать, что $C=\mathbb{R}^{n+1}$. Отображение F непрерывно дифференцируемо на V и поэтому строго дифференцируемо в нуле. Тогда, рассуждая также как при доказательстве правила множителей Лагранжа, придем к противоречию с тем, что $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ — локальный экстремум.

Действительно, $F(0)=(J(\widehat{x}(\cdot),\widehat{u}(\cdot)),x_1)$. Для достаточно малых по модулю ν вектор $y_\nu=(J(\widehat{x}(\cdot),\widehat{u}(\cdot))+\nu,x_1)$ принадлежит соответствующей окрестности $F(0)$ (окрестность W в теореме лоб обратной функции). Тогда существует вектор $\overline{\alpha}(y_\nu)\in V$ такой, что $F(\overline{\alpha}(y_\nu))=y_\nu$, или $J(x(\cdot,\overline{\alpha}(y_\nu),\overline{u}(\cdot)),u(\cdot,\overline{\alpha}(y_\nu),\overline{u}(\cdot)))=J(\widehat{x}(\cdot),\widehat{u}(\cdot))+\nu$, $x(t_1,\overline{\alpha}(y_\nu),\overline{u}(\cdot))=x_1$ и, кроме того, $|\overline{\alpha}(y_\nu)|\le K|\nu|$. Из этих соотношений следует (в силу того, что $x(\cdot,\overline{\alpha},\overline{u}(\cdot))\to\widehat{x}(\cdot)$ в $C^1([t_0,t_1],\mathbb{R}^n)$ и, очевидно, $u(\cdot,\overline{\alpha},\overline{u}(\cdot))\to\widehat{u}(\cdot)$ в $C([t_0,t_1],\mathbb{R}^r)$ при $\overline{\alpha}\to 0$), что в любой окрестности пары $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ найдется допустимая в задаче (P_7) точка, в которой значение функционала больше (при $\nu>0$) или меньше (при $\nu<0$), чем в $(\widehat{x}(\cdot),\widehat{u}(\cdot))$, в противоречие с тем, что $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ — локальный экстремум.

Таким образом, $L=\operatorname{span}\{\,(y_0(u(\cdot)),y(t_1,u(\cdot)))\in\mathbb{R}^{n+1}:u(\cdot)\in C([t_0,t_1],\mathbb{R}^r)\,\}$ — собственное подпространство в $\mathbb{R}^{n+1}$. Пусть $y_0\notin L$. Тогда по второй теореме отделимости (§1, п. 1.2) найдется ненулевой линейный функционал (вектор) $\overline{\lambda}=(\lambda_0,\lambda_1,\ldots,\lambda_n)\in(\mathbb{R}^{n+1})^*$ такой, что $\langle\overline{\lambda},y_0\rangle<\inf_{y\in L}\langle\overline{\lambda},y\rangle$. Отсюда следует, что $\langle\overline{\lambda},y\rangle=0$ для всех $y\in L$ (продумайте это). Это равносильно тому, что для любого $u(\cdot)\in C([t_0,t_1],\mathbb{R}^r)$ справедливо равенство (с учетом выражения для $y_0(u(\cdot))$ из (ii))

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

где $\lambda=(\lambda_1,\ldots,\lambda_n)$.

Обозначая через $p(\cdot)$ — решение линейного уравнения

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

получаем уравнение Эйлера по x.

Подставим теперь в (iii) вместо функции $\lambda_0\widehat{f}_x(\cdot)$ ее выражение из (iv), а затем вместо функции $\widehat{\varphi}_x(\cdot)y(\cdot, u(\cdot))$ ее выражение из (i) и учитывая, что $p(t_1) = -\lambda$, а $y(t_0, u(\cdot)) = 0$, будем иметь

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

Беря в последнем интеграле в качестве $u(\cdot)$ функцию $(\lambda_0\widehat{f}_u(t) - p(t)\widehat{\varphi}_u(t))^T$, получаем, что $p(t)\widehat{\varphi}_u(t) = \lambda_0\widehat{f}_u(t)$. □

5. Необходимые условия минимума в задаче оптимального управления

Постановка задачи. Пусть $[t_0, t_1]$ — конечный отрезок, U — непустое подмножество $\mathbb{R}^r$, заданы функция $f\colon \mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^r \to \mathbb{R}$ и 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 $x_i \in \mathbb{R}^n$, $i = 0, 1$.

Задача

$$J(x(\cdot), u(\cdot)) = \int_{t_0}^{t_1} f(t, x(t), u(t))\, dt \to \min, \quad \dot{x} = \varphi(t, x, u(t)),$$
$$u(t) \in U, \quad x(t_0) = x_0, \quad x(t_1) = x_1 \quad (P_8)$$

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

Уточним постановку. 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_8), 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_8), 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_8).

Теорема (Принцип максимума Понтрягина)**.** *Если пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ доставляет сильный минимум в задаче (P_8), 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 $u(\cdot)$:

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

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

Перед непосредственным доказательством теоремы приведем некоторые определения и сформулируем одно утверждение.

Пусть $\widehat{u}(\cdot) \in PC([t_0, t_1], \mathbb{R}^r)$, $\tau \in (t_0, t_1)$ — точка непрерывности $\widehat{u}(\cdot)$ и $v \in U$. Для каждого $0 \leq \alpha < \tau - t_0$ такого, что функция $\widehat{u}(\cdot)$ непрерывна на отрезке $[\tau - \alpha, \tau]$ определим функцию из $PC([t_0, t_1], \mathbb{R}^r)$ по формуле

$$u(t, \alpha; \tau, v) = \begin{cases} \widehat{u}(t), & \text{если} \quad t \notin [\tau - \alpha, \tau); \\ v, & \text{если} \quad t \in [\tau - \alpha, \tau), \end{cases}$$

которую называют *игольчатой вариацией* функции $\widehat{u}(\cdot)$, а пару (τ, v) — *иголкой*. При $\alpha = 0$ эта функция совпадает с $\widehat{u}(\cdot)$.

Понятие игольчатой вариации естественным образом распространяется на конечный набор иголок. Пусть $k \in \mathbb{N}$ и $\mathcal{N}_k$ — упорядоченная совокупность пар (иголок) $(\tau_i, v_i) \in (t_0, t_1) \times U$, $1 \le i \le k$, где $t_0 < \tau_1 < \ldots < \tau_k < t_1$ и точки τ_i суть точки непрерывности $\widehat{u}(\cdot)$. Для каждого $\overline{\alpha} = (\alpha_1, \ldots, \alpha_k)^T \in \mathbb{R}^k_+$ такого, что отрезки $[\tau_i - \alpha_i, \tau_i]$, $1 \le i \le k$, не пересекаются и на них $\widehat{u}(\cdot)$ непрерывна, определим функцию из $PC([t_0, t_1], \mathbb{R}^r)$ по формуле

$$u(t, \overline{\alpha}; \mathcal{N}_k) = \begin{cases} \widehat{u}(t), & \text{если} \quad t \notin \cup_{i=1}^k [\tau_i - \alpha_i, \tau_i); \\ v_i, & \text{если} \quad t \in [\tau_i - \alpha_i, \tau_i),\ 1 \le i \le k, \end{cases}$$

которая также называется *игольчатой вариацией* $\widehat{u}(\cdot)$, а набор $\mathcal{N}_k$ называется *пакетом иголок.*

Лемма 1 (о пакете иголок)**.** *Пусть* $(\widehat{x}(\cdot), \widehat{u}(\cdot)) \in PC^1([t_0, t_1], \mathbb{R}^n) \times PC([t_0, t_1], \mathbb{R}^r)$ *и функция* $\widehat{x}(\cdot)$ *является решением задачи Коши*

$$\dot{x} = \varphi(t, x, \widehat{u}(t)), \quad x(t_0) = x_0 \tag{1}$$

на отрезке $[t_0, t_1]$.

Если $k \in \mathbb{N}$ *и* $\mathcal{N}_k = \{(\tau_i, v_i),\ 1 \le i \le k\}$ *— пакет иголок, то найдется такая окрестность нуля* V *в* $\mathbb{R}^k$*, что для любого* $\overline{\alpha} \in V \cap \mathbb{R}^k_+$ *существует единственное решение* $x(\cdot, \overline{\alpha}; \mathcal{N}_k)$ *задачи Коши*

$$\dot{x} = \varphi(t, x, u(t, \overline{\alpha}; \mathcal{N}_k)), \quad x(t_0) = x_0, \tag{2}$$

определенное на всем отрезке $[t_0, t_1]$*, и при этом* $x(\cdot, \overline{\alpha}; \mathcal{N}_k)$ *сходится равномерно к* $\widehat{x}(\cdot)$ *на* $[t_0, t_1]$ *при* $\overline{\alpha} \to 0$.

Для любого $t \in [t_0, t_1]$ *отображение* $\overline{\alpha} = (\alpha_1, \ldots, \alpha_k)^T \mapsto x(t, \overline{\alpha}; \mathcal{N}_k)$ *может быть продолжено до непрерывно дифференцируемого отображения в некоторой окрестности нуля в* $\mathbb{R}^k$*. Частная производная по* α_i*,* $1 \le i \le k$*, в нуле этого отображения, которую обозначим* $x_{\alpha_i}(t) = x_{\alpha_i}(t, 0; \mathcal{N}_k)$*, удовлетворяет на* $[\tau_i, t_1]$ *уравнению*

$$\dot{x}_{\alpha_i}(t) = \widehat{\varphi}_x(t) x_{\alpha_i}(t),\ x_{\alpha_i}(\tau_i) = \varphi(\tau_i, \widehat{x}(\tau_i), v_i) - \varphi(\tau_i, \widehat{x}(\tau_i), \widehat{u}(\tau_i)). \tag{3}$$

Эту лемму доказывать не будем.

Заметим, что из (3) следует, что $x_{\alpha_i}(t, 0; \mathcal{N}_k)$ зависит только от одной иголки (τ_i, v_i), поэтому вместо $x_{\alpha_i}(t, 0; \mathcal{N}_k)$ будем писать $x_{\alpha_i}(t, 0; \tau_i, v_i)$.

Следствие 3. *Функция* $\overline{\alpha} \mapsto J(\overline{\alpha}; \mathcal{N}_k) = J(x(\cdot, \overline{\alpha}, \mathcal{N}_k), u(\cdot, \overline{\alpha}, \mathcal{N}_k))$ *непрерывно дифференцируема на* V *и ее частная производная по*

α_i *выражается формулой*

$$J_{\alpha_i}(0;\mathcal{N}_k) = f(\tau_i, \widehat{x}(\tau_i), v_i) - f(\tau_i, \widehat{x}(\tau_i), \widehat{u}(\tau_i)) + \int_{\tau_i}^{t_1} \langle \widehat{f}_x(t), x_{\alpha_i}(t,0;\tau_i,v_i)\rangle dt, \quad i = 1,\dots,k. \quad (4)$$

Доказательство. Действительно, рассмотрим задачу Коши

$$\begin{cases} \dot{x} = \varphi(t, x, u(t, \overline{\alpha}; \mathcal{N}_k)) \\ \dot{y} = f(t, x, u(t, \overline{\alpha}; \mathcal{N}_k)), \quad (x(t_0), y(t_0)) = (x_0, 0). \end{cases}$$

Согласно лемме о пакете иголок у этой системы существует единственное решение $(\widetilde{x}(\cdot, \overline{\alpha}; \mathcal{N}_k), y(\cdot, \overline{\alpha}; \mathcal{N}_k))$, обладающее свойствами, отмеченными в утверждении леммы. В силу единственности $\widetilde{x}(\cdot, \overline{\alpha}; \mathcal{N}_k) = x(\cdot, \overline{\alpha}; \mathcal{N}_k)$, так что $y(t_1, \overline{\alpha}; \mathcal{N}_k)) = J(x(\cdot, \overline{\alpha}; \mathcal{N}), u(\cdot, \overline{\alpha}; \mathcal{N}_k))$. Так как φ и f не зависят от y, нетрудно видеть, что частная производная по α_i, $1 \le i \le k$, функции $\overline{\alpha} \mapsto y(t, \overline{\alpha}; \mathcal{N}_k)$ в нуле на отрезке $[\tau_i, t_1]$, которую обозначим $y_{\alpha_i}(t)$, есть решение задачи Коши

$$\dot{y}_{\alpha_i}(t) = \langle \widehat{f}_x(t), x_{\alpha_i}(t, 0; \tau_i, v_i)\rangle, \quad y_{\alpha_i}(\tau_i) = f(\tau_i, \widehat{x}(\tau_i), v_i) - f(\tau_i, \widehat{x}(\tau_i), \widehat{u}(\tau_i)).$$

Если записать то уравнение в интегральной форме и положить $t = t_1$, то получим (4), где $y_{\alpha_i}(t_1)$ обозначено через $J_{\alpha_i}(0; \mathcal{N}_k)$. $\square$

Доказательство принципа максимума. Пусть $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — оптимальный процесс в задаче (P_8), $\mathcal{N}_k = \{(\tau_i, v_i),\ 1 \le i \le k\}$ — произвольный пакет иголок и $u(\cdot, \overline{\alpha}; \mathcal{N}_k)$ — соответствующая игольчатая вариация управления $\widehat{u}(\cdot)$. Согласно лемме 1 найдется такая окрестность нуля V в $\mathbb{R}^k$, что для любого $\overline{\alpha} \in V \cap \mathbb{R}^k_+$ существует единственное решение $x(\cdot, \overline{\alpha}; \mathcal{N}_k)$ задачи Коши (2), которое может быть продолжено до непрерывно дифференцируемого отображения на некоторую окрестность V_0 нуля в $\mathbb{R}^k$. Продолженное отображение обозначаем также $x(\cdot, \overline{\alpha}; \mathcal{N}_k)$.

Рассмотрим отображение $F\colon V_0 \to \mathbb{R}^{n+1}$, определенное формулой

$$F(\overline{\alpha}) = F(\overline{\alpha}; \mathcal{N}_k) = (J(\overline{\alpha}; \mathcal{N}_k),\ x(t_1, \overline{\alpha}; \mathcal{N}_k))^T.$$

Из предыдущего следует, что это отображение непрерывно дифференцируемо. Покажем, что равенство $F'(0)(\mathbb{R}^k_+) = \mathbb{R}^{n+1}$ невозможно. Докажем это от противного: предположим, что данное равенство выполняется и придем к противоречию с тем, что $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — сильный минимум в задаче (P_8).

Поскольку отображение F непрерывно дифференцируемо, то оно, в частности, строго дифференцируемо в нуле и тем самым выполнены предположения теоремы об обратной функции (§1, п. 1.3), где $n = k$, $m = n+1$ и $C = \mathbb{R}^k_+$ Пусть W — окрестность точки $F(0) = (J(\widehat{x}(\cdot), \widehat{u}(\cdot)),\, \widehat{x}(t_1))^T$ из этой теоремы. Ясно, что для всех достаточно малых $\varepsilon > 0$ точки $(J(\widehat{x}(\cdot), \widehat{u}(\cdot)) - \varepsilon,\, \widehat{x}(t_1))^T$ принадлежат W. Согласно теореме, найдется такое $\overline{\alpha}_\varepsilon \in V_0 \cap \mathbb{R}^k_+$, что $F(\overline{\alpha}_\varepsilon) = (J(\widehat{x}(\cdot), \widehat{u}(\cdot)) - \varepsilon,\, \widehat{x}(t_1))^T$, или

$$(J(\overline{\alpha}_\varepsilon; \mathcal{N}_k),\ x(t_1, \overline{\alpha}_\varepsilon; \mathcal{N}_k))^T = (J(\widehat{x}(\cdot), \widehat{u}(\cdot)) - \varepsilon,\, \widehat{x}(t_1))^T,$$

т. е. $J(\overline{\alpha}_\varepsilon; \mathcal{N}_k) = J(\widehat{x}(\cdot), \widehat{u}(\cdot)) - \varepsilon$ и $x(t_1, \overline{\alpha}_\varepsilon; \mathcal{N}_k) = \widehat{x}(t_1) = x_1$.

Ясно, что $x(t_0, \overline{\alpha}_\varepsilon; \mathcal{N}_k) = x_0$ (см. (2)). Следовательно, пара $(x(\cdot, \overline{\alpha}_\varepsilon; \mathcal{N}_k),\, u(\cdot, \overline{\alpha}_\varepsilon; \mathcal{N}_k))$ допустима в задаче (P_8) и $J(x(\cdot, \overline{\alpha}_\varepsilon; \mathcal{N}_k),\, u(\cdot, \overline{\alpha}_\varepsilon; \mathcal{N}_k)) = J(\overline{\alpha}_\varepsilon; \mathcal{N}_k) = J(\widehat{x}(\cdot), \widehat{u}(\cdot)) - \varepsilon < J(\widehat{x}(\cdot), \widehat{u}(\cdot))$.

Отсюда, по неравенству $|\overline{\alpha}_\varepsilon| \le K\varepsilon$ (см. теорему об обратной функции (§1, п. 1.3)) и так как $x(\cdot, \overline{\alpha}; \mathcal{N}_k)$ сходится равномерно к $\widehat{x}(\cdot)$ на $[t_0, t_1]$ при $\overline{\alpha} \to 0$ (по лемме 1) получаем, что для любой окрестности $\widehat{x}(\cdot)$ найдется допустимая пара $(x(\cdot, \overline{\alpha}_\varepsilon; \mathcal{N}_k),\, u(\cdot, \overline{\alpha}_\varepsilon; \mathcal{N}_k))$ в задаче (P_8) такая, что $x(\cdot, \overline{\alpha}_\varepsilon; \mathcal{N}_k)$ принадлежит данной окрестности, и на этой паре минимизируемый функционал меньше $J(\widehat{x}(\cdot), \widehat{u}(\cdot))$. Это противоречит тому, $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — оптимальный процесс в задаче (P_8).

Итак, равенство $F'(0)(\mathbb{R}^k_+) = \mathbb{R}^{n+1}$ невозможно. Это означает, другими словами, что коническая оболочка векторов $(J_{\alpha_i}(0; \tau_i, v_i),\, x_{\alpha_i}(t_1, 0; \tau_i, v_i))^T$, $i = 1, \ldots, k$, которые являются столбцами матрицы $F'(0)$, не совпадает с $\mathbb{R}^{n+1}$. Покажем, что отсюда следует, что коническая оболочка всех векторов вида $(J_\alpha(0; \tau, v),\, x_\alpha(t_1, 0; \tau, v))^T$ (по всевозможным иголкам (τ, v)) также не совпадает с $\mathbb{R}^{n+1}$.

Допустим, что эта коническая оболочка, которую обозначим через C, совпадает с $\mathbb{R}^{n+1}$. Тогда согласно следствию 2 из теоремы Каратеодори для конуса (см. §, п. 1.2) любой элемент из $\mathbb{R}^{n+1}$ может быть представлен как коническая комбинация некоторого фиксированного конечного числа элементов из C, что противоречит доказанному выше.

Итак, множество C не совпадает с $\mathbb{R}^{n+1}$. Ясно, что C — выпуклый конус. Пусть точка y_0 ему не принадлежит. Тогда, по теореме отделимости, найдется ненулевой функционал (вектор-строка) $\lambda \in (\mathbb{R}^{n+1})^*$ такой, что $\langle \lambda, y_0 \rangle \le \inf\limits_{y \in C} \langle \lambda, y \rangle$. Покажем, что $\langle \lambda, y \rangle \ge 0$

для любого $y \in C$. Действительно, если $\langle \lambda, y_1 \rangle < 0$ для некоторого $y_1 \in C$, то так как $\alpha y_1 \in C$ для любого $\alpha > 0$, получим, что $\inf\limits_{y \in C} \langle \lambda, y \rangle \le \inf\limits_{\alpha > 0} \alpha \langle \lambda, y_1 \rangle = -\infty$. Но это противоречит неравенству выше.

Итак, представляя λ в виде $\lambda = (\lambda_0, \lambda_1) \in \mathbb{R} \times (\mathbb{R}^n)^*$, неравенство $\langle \lambda, y \rangle \ge 0$ для всех $y \in C$, учитывая (4), запишется в виде

$$\lambda_0 \left(f(\tau, \widehat{x}(\tau), v) - f(\tau, \widehat{x}(\tau), \widehat{u}(\tau)) + \int_\tau^{t_1} \langle \widehat{f}_x(t),\, x_\alpha(t, 0; \tau, v) \rangle \, dt \right) + \\ + \langle \lambda_1,\, x_\alpha(t_1, 0; \tau, v) \rangle \ge 0 \quad (5)$$

для любой пары $(\tau, v) \in (t_0, t_1) \times U$, где τ — точка непрерывности $\widehat{u}(\cdot)$.

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

$$\dot{p} = -p\, \widehat{\varphi}_x(t) + \lambda_0 \widehat{f}_x(t), \quad p(t_1) = -\lambda_1. \tag{6}$$

Подставим теперь в (5) вместо функции $\lambda_0 \widehat{f}_x(\cdot)$ ее выражение из (6), а затем вместо функции $\widehat{\varphi}_x(\cdot) x_{\alpha_i}(\cdot; \tau, v)$ ее выражение из (3) и, учитывая начальные условия в (3) и (6), будем иметь

$$\begin{aligned} 0 \le \lambda_0(f(\tau, \widehat{x}(\tau), v) - f(\tau, \widehat{x}(\tau), \widehat{u}(\tau))) + \int_\tau^{t_1} (\langle \dot{p}(t), x_{\alpha_i}(t, 0; \tau, v) \rangle \\ + \langle p(t), \dot{x}_{\alpha_i}(t, 0; \tau, v) \rangle) \, dt + \langle \lambda_1, x_{\alpha_i}(t_1, 0; \tau, v) \rangle = \lambda_0(f(\tau, \widehat{x}(\tau), v) \\ - f(\tau, \widehat{x}(\tau), \widehat{u}(\tau))) + \langle p(t), x_{\alpha_i}(t, 0; \tau, v) \rangle |_\tau^{t_1} + \langle \lambda_1, x_{\alpha_i}(t_1, 0; \tau, v) \rangle \\ = \lambda_0(f(\tau, \widehat{x}(\tau), v) - f(\tau, \widehat{x}(\tau), \widehat{u}(\tau))) - \langle p(\tau), \varphi(\tau, \widehat{x}(\tau), v) \\ - \varphi(\tau, \widehat{x}(\tau), \widehat{u}(\tau)) \rangle. \end{aligned}$$

Таким образом, для всех $v \in U$ и всех $\tau \in (t_0, t_1)$, где функция $\widehat{u}(\cdot)$ непрерывна, выполняется неравенство $H(\tau, \widehat{x}(\tau), v, \lambda_0, p(\tau)) \le H(\tau, \widehat{x}(\tau), \widehat{u}(\tau), \lambda_0, p(\tau))$, которое вместе с (6) доказывает теорему. □

6. **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. Напомним, что эта задача имеет вид

$$J(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t)) dt \to \text{extr}, \quad x(t_i) = x_i, \quad i = 0, 1, \quad (P_6)$$

где $L\colon \mathbb{R}^3 \to \mathbb{R}$ функция переменных t, x и $\dot{x}$ и $x_i \in \mathbb{R}$, $i = 0, 1$.

Напомним, что функция $x(\cdot) \in C^1([t_0, t_1])$ называется *допустимой* в задаче (P_6), если $x(t_i) = x_i$, $i = 0, 1$, и допустимая функция $\widehat{x}(\cdot)$ называется *слабым локальным минимумом* (*максимумом*) в задаче (P_6), 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 локальный экстремум — это слабый локальный минимум, либо слабый локальный максимум.

В вариационном исчислении, наряду со слабым экстремумом, рассматривают еще и сильный локальный экстремум, где близость функций измеряется в пространстве $C([t_0, t_1])$. Напомним, что $PC^1([t_0, t_1])$ обозначает пространство кусочно-непрерывно-дифференцируемых функций на $[t_0, t_1]$. Допустимая функция $x(\cdot) \in PC^1([t_0, t_1])$ в задаче (P_6) определяется аналогично предыдущему. Скажем, что допустимая функция $\widehat{x}(\cdot)$ доставляет *сильный локальный минимум* (*максимум*) в задаче (P_6), если существует такое $\varepsilon > 0$, что для любой допустимой функции $x(\cdot)$, для которой $\|x(\cdot) - \widehat{x}(\cdot)\|_{C([t_0,t_1])} < \varepsilon$ выполняется неравенство $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$ $(J(x(\cdot)) \leq J(\widehat{x}(\cdot)))$.

Сильный локальный экстремум — это либо сильный локальный минимум, либо сильный локальный максимум.

Заметим, что если функция $\widehat{x}(\cdot)$ доставляет сильный экстремум в задаче (P_6) и при этом $\widehat{x}(\cdot) \in C^1([t_0, t_1])$, то $\widehat{x}(\cdot)$ является и слабым экстремумом в этой задаче. Действительно, пусть $\varepsilon > 0$ такое, что как только $x(\cdot) \in PC^1([t_0, t_1])$ и $\|x(\cdot) - \widehat{x}(\cdot)\|_{C([t_0,t_1])} < \varepsilon$, то, скажем, $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$. Если теперь $x(\cdot) \in C^1([t_0, t_1])$ и $\|x(\cdot) - \widehat{x}(\cdot)\|_{C^1([t_0,t_1])} < \varepsilon$, то так как, в частности, $x(\cdot) \in PC^1([t_0, t_1])$ и $\|x(\cdot) - \widehat{x}(\cdot)\|_{C([t_0,t_1])} \leq \|x(\cdot) - \widehat{x}(\cdot)\|_{C^1([t_0,t_1])} < \varepsilon$, получаем, что $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$.

Таким образом, необходимые условия слабого экстремума для $x(\cdot) \in C^1([t_0, t_1])$ являются необходимыми условиями и сильного экстремума.

Теперь сформулируем условия Лежандра, Якоби и Вейерштрасса. Пусть $\widehat{x}(\cdot)$ — *экстремаль* (или *стационарная точка*) задачи (P_6), т. е. функция, для которой выполнено уравнение Эйлера. Предположим, что интегрант L дважды непрерывно дифференцируем.

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

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

Тогда для достаточно малых по модулю λ

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

и дифференцируя еще раз, получаем, что

$$\varphi''(0) = \int_{t_0}^{t_1} \left(\widehat{L}_{xx}(t)h^2(t) + 2\widehat{L}_{\dot{x}x}(t)h(t)\dot{h}(t) + \widehat{L}_{\dot{x}\dot{x}}(t)\dot{h}^2(t)\right)\,dt,$$

так как $\widehat{L}_{x\dot{x}}(t) = \widehat{L}_{\dot{x}x}(t)$.

Этот функционал (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

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

Это уравнение называется *уравнением Якоби* исходной задачи (P_6).

Введем теперь некоторые определения. Пусть $\widehat{x}(\cdot)$ — *экстремаль* задачи (P_6), т. е. $\widehat{x}(\cdot)$ удовлетворяет уравнению Эйлера для этой задачи.

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

(2) Пусть на $\widehat{x}(\cdot)$ выполнено усиленное условие Лежандра. Точка $\tau \in (t_0, t_1]$ называется *сопряженной точкой* к точке t_0, если существует нетривиальное решение $h(\cdot) \in C^1([t_0, t_1])$ уравнения Якоби такое, что $h(t_0) = h(\tau) = 0$.

(3) Говорят, что на функции $\widehat{x}(\cdot)$ выполнено *условие Якоби*, если в интервале (t_0, t_1) нет точек, сопряженных с t_0 и *усиленное условие Якоби*, если в полуинтервале $(t_0, t_1]$ нет точек, сопряженных с t_0.

Пусть $f : \mathbb{R} \to \mathbb{R}$ — дифференцируемая функция. Функция $\mathcal{E}(x, x') = f(x') - f(x) - f'(x)(x' - x)$ называется *функцией Вейерштрасса* (соответствующей функции f). Геометрически, $\mathcal{E}(x, x')$ — разность между значением функции f и функции $y \mapsto f(x) + f'(x)(y - x)$ (график которой есть касательная плоскость к графику функции f в точке x) в точке x'.

Если f — выпуклая функция, то $\mathcal{E}(x, x') \geq 0$ для всех $x, x' \in \mathbb{R}^n$. Действительно, пусть $x, x' \in \mathbb{R}^n$ и $0 < \alpha < 1$. По неравенству

Иенссена $f((1-\alpha)x+\alpha x')\le(1-\alpha)f(x)+\alpha f(x')$, откуда $\alpha^{-1}(f(x+\alpha(x'-x))-f(x))\le f(x')-f(x)$. Переходя к пределу при $\alpha\to 0$, получаем, что $f(x')-f(x)\ge f'(x)(x'-x)$.

Пусть L — интегрант в задаче (P_6). 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. Ясно, что при каждых t и x — это функция Вейерштрасса, соответствующая функции $L(t,x,\cdot)$.

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

При доказательстве необходимых условий в задаче (P_6) нам понадобится один несложный технический результат, который приводим без доказательства.

Лемма 2 (о скруглении углов)**.** *Пусть в задаче* (P_6) *интегрант* L *непрерывен по совокупности переменных. Тогда*

$$\inf\{\,J(x(\cdot))\mid x(\cdot)\in PC^1([t_0,t_1]),\ x(t_0)=x_0,\ x(t_1)=x_1\,\}=$$
$$=\inf\{\,J(x(\cdot))\mid x(\cdot)\in C^1([t_0,t_1]),\ x(t_0)=x_0,\ x(t_1)=x_1\,\}.$$

Следующая теорема — основной результат данного параграфа.

Теорема (Необходимые условия минимума в простейшей задаче вариационного исчисления)**.** *A) Пусть функция* $\widehat{x}(\cdot)\in C^1([t_0,t_1])$ *доставляет сильный локальный минимум в задаче* (P_6)*. Тогда*

(*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*) *для всех* $t\in[t_0,t_1]$ *и* $u\in\mathbb{R}$ *выполнено условие Вейерштрасса*

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

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

B) Пусть функция $\widehat{x}(\cdot)\in C^1([t_0,t_1])$ *доставляет слабый локальный минимум в задаче* (P_6)*. Тогда выполнено уравнение Эйлера, условие Лежандра и условие Якоби.*

Доказательство. *A*) Запишем задачу (P_6) как задачу оптимального управления:

$$\int_{t_0}^{t_1}L(t,x(t),u(t))\,dt\to\min,\quad \dot x=u,\quad x(t_0)=x_0,\quad x(t_1)=x_1.\quad (\overline{P}_6)$$

Легко видеть, что $\widehat{x}(\cdot)$ доставляет сильный минимум в (P_6) тогда и только тогда, когда пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$, где $\widehat{u}(\cdot) = \dot{\widehat{x}}(\cdot)$ является сильным минимумом в $(\overline{P}_6)$.

Согласно принципу максимума найдутся λ_0 и функция $p(\cdot) \in PC^1([t_0, t_1])$, не равные одновременно нулю, такие, что для всех $t \in [t_0, t_1]$ выполнены условие стационарности по $x(\cdot)$:

$$-\dot{p}(t) + \lambda_0 \widehat{L}_x(t) = 0 \qquad (i)$$

и условие максимума по $u(\cdot)$:

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

Если $\lambda_0 = 0$, то $p(\cdot) = \mathrm{const}$ вследствие (i). Тогда из (ii) следует, что эта константа обязана быть нулевой и тем самым все множители Лагранжа нулевые. 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 $f\colon u \to p(t)u - L(t, \widehat{x}(t), u)$ на $\mathbb{R}$ 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 $p(t) = \widehat{L}_{\dot{x}}(t)$ 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, что тоже $L(t, \widehat{x}(t), u) - L(t, \widehat{x}(t), \dot{\widehat{x}}(t)) - L_{\dot{x}}(t, \widehat{x}(t), \dot{\widehat{x}}(t))(u - \dot{\widehat{x}}(t)) \geq 0$, т. е. выполнено условие Вейерштрасса.

B) Уравнение Эйлера, как необходимое условие слабого экстремума, уже было доказано ранее. Доказательство, в терминах данного параграфа, заключается в том, что если $\widehat{x}(\cdot)$ — локальный минимум, то ноль есть локальный минимум для функции φ, определенной соотношением $(*)$ и тогда необходимо $\varphi'(0) = 0$. Расшифровка этого условия и приводит к уравнению Эйлера.

Теперь докажем условие Лежандра, расшифровывая необходимое условие минимума второго порядка: $\varphi''(0) \geq 0$. Согласно формуле для $\varphi''(0)$, выписанной выше, данное условие равносильно тому, что $Q(h(\cdot)) \geq 0$ для всех $h(\cdot) \in C^1([t_0, t_1])$ таких, что $h(t_0) =$

$h(t_1) = 0$. Это означает, что функция $\widehat{h}(\cdot) = 0$ есть слабый локальный минимум в задаче

$$Q(h(\cdot)) = \int_{t_0}^{t_1} (\widehat{L}_{xx}(t)h^2(t) + 2\widehat{L}_{x\dot{x}}(t)h(t)\dot{h}(t) + \widehat{L}_{\dot{x}\dot{x}}(t)\dot{h}^2(t))dt \to \min, \quad h(t_0) = h(t_1) = 0. \quad (iii)$$

По лемме о скруглении углов функция $\widehat{h}(\cdot) = 0$ доставляет и сильный минимум в этой задаче. Тогда, по уже доказанному, на $\widehat{h}(\cdot)$ должно выполняться условие Лежандра, которое в данном случае имеет тот же вид: $\widehat{L}_{\dot{x}\dot{x}}(t) \geq 0$.

Докажем условие Якоби. Предположим противное, что существует точка $\tau \in (t_0, t_1)$ и нетривиальное решение $\overline{h}(\cdot)$ уравнения Якоби такое, что $\overline{h}(t_0) = \overline{h}(\tau) = 0$. Пусть функция $\widetilde{h}(\cdot)$ такова, что $\widetilde{h}(t) = \overline{h}(t)$, если $t_0 \leq t \leq \tau$ и $\widetilde{h}(t) = 0$, если $\tau \leq t \leq t_1$. Заметим, что $\dot{\overline{h}}(\tau) \neq 0$, так как в противном случае, по теореме единственности, функция $\overline{h}(\cdot)$ была бы тождественным нулем. Далее, интегрируя по частям ($\overline{h}(t_0) = \overline{h}(\tau) = 0$), получим

$$Q(\widetilde{h}(\cdot)) = \int_{t_0}^{\tau} \left(\widehat{L}_{xx}(t)\overline{h}^2(t) + 2\widehat{L}_{\dot{x}x}(t)\overline{h}(t)\dot{\overline{h}}(t) + \widehat{L}_{\dot{x}\dot{x}}(t)\dot{\overline{h}}^2(t)\right) dt$$

$$= \int_{t_0}^{\tau} \left(\widehat{L}_{xx}(t)\overline{h}(t) + \widehat{L}_{x\dot{x}}(t)\dot{\overline{h}}(t)\right) \overline{h}(t)\, dt + \int_{t_0}^{\tau} \left(\widehat{L}_{\dot{x}x}(t)\overline{h}(t) + \widehat{L}_{\dot{x}\dot{x}}(t)\dot{\overline{h}}(t)\right) \dot{\overline{h}}(t)\, dt$$

$$= \int_{t_0}^{\tau} \left(-\frac{d}{dt}\left(\widehat{L}_{\dot{x}x}(t)\overline{h}(t) + \widehat{L}_{\dot{x}\dot{x}}(t)\dot{\overline{h}}(t)\right) + \widehat{L}_{xx}(t)\overline{h}(t) + \widehat{L}_{x\dot{x}}(t)\dot{\overline{h}}(t)\right) \overline{h}(t)\, dt.$$

Поскольку $\overline{h}(\cdot)$ удовлетворяет уравнению Якоби, то отсюда следует, что $Q(\widetilde{h}(\cdot)) = 0$. Это означает, что наряду с $\widehat{h}(\cdot) = 0$, функция $\widetilde{h}(\cdot)$доставляет сильный минимум в задаче (iii). Запишем эту задачу как задачу оптимального управления

$$\int_{t_0}^{t_1} \left(\widehat{L}_{xx}(t)h^2(t) + 2\widehat{L}_{\dot{x}x}(t)h(t)u(t) + \widehat{L}_{\dot{x}\dot{x}}(t)u^2(t)\right) dt \to \min, \quad \dot{h} = u, \quad h(t_0) = h(t_1) = 0.$$

Согласно принципу максимума найдутся такие λ_0 и $p(\cdot) \in PC^1([t_0, t_1])$, не равные одновременно нулю, что выполнены условие стационарности по $h(\cdot)$:

$$-\dot{p}(t) + 2\lambda_0\widehat{L}_{xx}(t)\widetilde{h}(t) + 2\lambda_0\widehat{L}_{x\dot{x}}(t)\dot{\widetilde{h}}(t) = 0 \quad (iv)$$

и условие максимума по $u(\cdot)$:

$$\max_{u\in\mathbb{R}}(p(t)u - 2\lambda_0\widehat{L}_{\dot{x}x}(t)\widetilde{h}(t)u + \lambda_0\widehat{L}_{\dot{x}\dot{x}}(t)u^2) =$$

$$= p(t)\dot{\widetilde{h}}(t) - 2\lambda_0\widehat{L}_{\dot{x}x}(t)\widetilde{h}(t)\dot{\widetilde{h}}(t) + \lambda_0\widehat{L}_{\dot{x}\dot{x}}(t)\dot{\widetilde{h}}^2(t). \quad (v)$$

Как и раньше проверяется, что $\lambda_0 \neq 0$. Пусть $\lambda_0 = 1/2$.

Из (v) следует, что для каждого $t \in [t_0, t_1]$ дифференцируемая на $\mathbb{R}^n$ функция $u \mapsto p(t)u - \widehat{L}_{\dot{x}x}(t)\widetilde{h}(t)u + (1/2)\widehat{L}_{\dot{x}\dot{x}}(t)u^2$ достигает максимума в точке $\dot{\widetilde{h}}(t)$. Следовательно, по теореме Ферма, ее производная в этой точке равна нулю:

$$p(t) = \widehat{L}_{\dot{x}x}(t)\widetilde{h}(t) + \widehat{L}_{\dot{x}\dot{x}}(t)\dot{\widetilde{h}}(t). \quad (vi)$$

По определению $\widetilde{h}(t) = 0$, если $t \geq \tau$ и поэтому из (vi) вытекает, что $p(\tau + 0) = 0$. С другой стороны, в силу (vi), $p(\tau - 0) = \widehat{L}_{\dot{x}x}(\tau)\widetilde{h}(\tau) + \widehat{L}_{\dot{x}\dot{x}}(\tau)\dot{\widetilde{h}}(\tau - 0) = \widehat{L}_{\dot{x}\dot{x}}(\tau)\dot{\bar{h}}(\tau) \neq 0$, так как $\dot{\bar{h}}(\tau) \neq 0$ (как уже было отмечено) и $\widehat{L}_{\dot{x}\dot{x}}(\tau) > 0$, поскольку выполнено усиленное условие Лежандра. Таким образом, функция $p(\cdot)$ разрывна в точке τ, в противоречие с тем, что $p(\cdot) \in PC^1([t_0, t_1])$. Этим условие Якоби доказано. $\square$