Вариационное исчисление и оптимальное управление (курс лекций 2013 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.

Нельзя не назвать также и прагматические причины, по которым приходится решать экстремальные задачи. Людям свойственно наилучшим образом использовать ресурсы, находящиеся в их

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

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

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

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

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

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

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 $f(x) \geq f(\widehat{x})$ $(f(x) \leq f(\widehat{x}))$ для всех допустимых $x \in U$ (т. е. для всех $x \in C \cap U$).

Цель курса — изложить начала теории экстремума, которая обычно делится на следующие главы: необходимые условия экстремума; достаточные условия экстремума; теория существования решений экстремальных задач и алгоритмы поиска таких решений. Далее будут затронуты вопросы, связанные, в основном, с условиями экстремума.

Первые параграфы относятся к базе теории экстремума. В них доказываются результаты, играющие ключевую роль в построении этой теории — теорема о разрешимости системы нелинейных уравнений конечного числа переменных и конечномерные теоремы отделимости для выпуклых множеств.

1. Дифференциальные свойства функций и отображений

В этом пункте напоминаются необходимые для дальнейшего факты, связанные с пространством $\mathbb{R}^n$, с понятиями дифференцируемости и строгой дифференцируемости функций и отображений на $\mathbb{R}^n$.

1.1. **Пространство $\mathbb{R}^n$ и двойственное к нему. Открытые и замкнутые множества. Компактность и теорема Вейерштрасса.** 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, $1 \leq i \leq n$ — *координатами вектора* x. Ради экономии места, элементы $\mathbb{R}^n$ будем записывать так $x = (x_1, \ldots, x_n)^T$, где ^{T} обозначает *транспонирование* (перевод строки в столбец и наоборот). В $\mathbb{R}^n$ естественным образом вводится операция (покоординатного) сложения векторов и операция (покоординатного) умножения вектора на число, превращающие это множество в вещественное векторное пространство.

Пусть $a = (a_1, \ldots, a_n)$ — вектор-строка из n действительных чисел. Для каждого $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ обозначим $a \cdot x = \sum_{i=1}^n a_i x_i$. Это матричное произведение вектор-строки a на вектор-столбец x, которое иногда называют *внутренним произведением*.

Ясно, что отображение $x \mapsto a \cdot x$ 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 $x \mapsto a \cdot x$ 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 $a \mapsto a{\cdot}x$ 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, $|x| = \sqrt{x^T \cdot x}$) 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: $|x^T \cdot y| \leq |x||y|$, 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 $U_{\mathbb{R}^n}(\widehat{x}, \delta) = \{\, x \in \mathbb{R}^n \mid |x - \widehat{x}| < \delta \,\}$ 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 $\mathrm{int\,A}$. Множество $G \subset \mathbb{R}^n$ называется *открытым*, если каждая его точка внутренняя, т. е. $\mathrm{int\,G} = \mathrm{G}$.

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 множества называется любое открытое множество, содержащее данное множество. В частности, *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) $B_{\mathbb{R}^n}(\widehat{x}, \delta) = \{\, x \in \mathbb{R}^n \mid |x - \widehat{x}| \le \delta \,\}$ 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 $\mathrm{cl\,A}$. Ясно, что $A \subset \mathrm{cl\,A}$.

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

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

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

Упражнение 1.9. *Доказать, что если K — компакт, а F — замкнутое множество, то их алгебраическая сумма $K + F = \{\, z = x + y \mid x \in K,\ y \in F \,\}$ — 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 $x \to x^* \cdot x + \alpha$, 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 $\gamma = \sup\{\, f(x) \mid x \in A\,\}$. 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 $F\colon A \to \mathbb{R}^m$. Это равносильно тому, что на A задано m функций. Действительно, координаты $F(x)$ суть m функций на A. С другой стороны, если на A заданы функции $f_1, \ldots, f_m$, то они определяют отображение $F\colon A \to \mathbb{R}^m$ по правилу $F(x) = (f_1(x), \ldots, f_m(x))^T$, $x \in A$. В этом смысле многие вопросы, связанные с отображениями можно свести к соответствующим вопросам для функций. Однако часто удобно (и полезно, имея в ввиду обобщения на бесконечномерный случай) работать именно с отображениями.

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

Линейный оператор $\Lambda\colon \mathbb{R}^n \to (\mathbb{R}^m)^*$ мы также отождествляем с его матрицей в стандартных базисах $\mathbb{R}^n$ и $(\mathbb{R}^m)^*$ (но $e'_1, \ldots, e'_m$ — уже вектор-строки). Тем самым, если $\Lambda e_i = \sum_{j=1}^m b_{ij} e'_j$, $i = 1, \ldots, n$, то матрицей этого оператора называется матрица $\Lambda =$

$(b_{ij})_{1\le i\le n,1\le j\le m}$ размера $n\times m$ и в этом случае Λx — произведение вектор строки x^T на матрицу Λ.

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 $\mathbb{R}^m$.*

1.2. **Дифференцируемость и непрерывная дифференцируемость.** 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 $a=(a_1,\dots,a_n)\in(\mathbb{R}^n)^*$ такой, что для всех $h\in\mathbb{R}^n$, для которых $\widehat{x}+h\in U$ справедливо представление

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

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

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

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

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

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

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

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: $f(x) = x^T \cdot Ax + b \cdot x + c$. 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, $i = 1, \ldots, m$, и отображение $F\colon U \to \mathbb{R}^m$ определено по правилу: $F(x) = (f_1(x), \ldots, f_m(x))^T$, то легко проверить, что F дифференцируемо в точке $\widehat{x}$ тогда и только тогда, когда функции f_i, $1 \le i \le m$, дифференцируемы в $\widehat{x}$. При этом строки матрицы $F'(\widehat{x})$ суть векторы $f_1'(\widehat{x}), \ldots, f_k'(\widehat{x})$. 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.

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.*

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 $\widehat{x}$, то говорят, что отображение F *непрерывно дифференцируемо в* $\widehat{x}$ (на U).

Из упр. 1.17* следует, что если отображение F задается функциями $f_1, \ldots, f_m$, то непрерывная дифференцируемость F на U равносильна тому, что все частные производные всех функций f_i, $1 \le i \le m$, непрерывны на U.

Если производная $f'\colon U \to (\mathbb{R}^n)^*$ дифференцируема в точке $\widehat{x}$ (на U), то говорят, что функция f *дважды дифференцируема в* $\widehat{x}$ *(на* U*)* и производную отображения f' в точке $x \in U$ называют *второй производной* f *в точке* x и обозначают $f''(x)$. Легко проверить, что $f''(x) = (\partial^2 f(x)/\partial x_j \partial x_i)$, $i, j = 1, \ldots, n$. Эту матрицу называют *матрицей Гесса* или *гессианом функции* f *в точке* x. Можно доказать, что эта матрица симметрична.

Пусть функция f дважды дифференцируема на U. Тогда определено отображение $f'' : U \to \mathcal{L}(\mathbb{R}^n, (\mathbb{R}^n)^*)$. Если оно непрерывно в $\widehat{x}$ (на U), то говорят, что функция f *дважды непрерывно дифференцируема в* $\widehat{x}$ *(на* U*)*. Последнее равносильно тому, что *все вторые частные производные функции* f *непрерывны на* U.

Теорема (формула Тейлора)**.** *Пусть* U *— окрестность точки* $\widehat{x} \in \mathbb{R}^n$ *и функция* $f : U \to \mathbb{R}$ *дважды дифференцируема в* $\widehat{x}$*. Тогда для любого* $h \in \mathbb{R}^n$ *такого, что* $\widehat{x} + h \in U$ *справедливо равенство*

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

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

1.3. **Строгая дифференцируемость.** 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 $\widehat{x}$ 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 $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2 D(x)$, где $D(\cdot)$ — функция Дирихле (равная единице, если x рационально и нулю, если иррационально). Легко проверить, что f дифференцируема в нуле и, очевидно, разрывна во всех остальных точках и значит, не может быть строго дифференцируемой в нуле.

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

Теорема (о среднем). *Пусть G — открытое множество в $\mathbb{R}^n$, отображение $F\colon G \to \mathbb{R}^m$ дифференцируемо на G и $[x, y] \subset G$. 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 $\varphi(t) = (F(y) - F(x))^T \cdot F((1-t)x + ty)$. Легко проверяется, что она 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 $(F(y) - F(x))^T \cdot F(y) - (F(y) - F(x))^T \cdot F(x) = (F(y) - F(x))^T \cdot F'(z)(y - x)$, 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. *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.

2. Модифицированный метод Ньютона и разрешимость системы нелинейных уравнений

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 $F(x) = y$ (или, равносильно, системы нелинейных уравнений $f_1(x_1, \ldots, x_n) = y_1, \ldots, f_m(x_1, \ldots, x_n) = y_m$) 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 $\widehat{x}$. При этом выбор начальной точки x_0 допускает некоторый произвол, и это приводит к разным решениям уравнения $F(x) = y$. Точное утверждение содержится в следующей теореме.

Теорема (о разрешимости системы уравнений)**.** *Пусть U — окрестность точки $\widehat{x} \in \mathbb{R}^n$, отображение $F\colon U \to \mathbb{R}^m$ строго дифференцируемо в точке $\widehat{x}$ и $F'(\widehat{x})(\mathbb{R}^n) = \mathbb{R}^m$. Тогда существуют окрестности $V \subset U$ и W точек $\widehat{x}$ и $F(\widehat{x})$, непрерывное отображение $\varphi\colon V \times W \to U$ и константа $K > 0$ такие, что $F(\varphi(\xi, y)) = y$ и $|\varphi(\xi, y) - \xi| \le K|y - F(\xi)|$ для всех $\xi \in V$ и $y \in W$.*

Доказательство. Обозначим для краткости $\Lambda = F'(\widehat{x})$ и покажем, что из условия $\Lambda(\mathbb{R}^n) = \mathbb{R}^m$ следует существование линейного отображения $R\colon \mathbb{R}^m \to \mathbb{R}^n$ и константы $\gamma > 0$ таких, что $\Lambda R(y) = y$ и $|R(y)| \le \gamma|y|$ для всех $y \in \mathbb{R}^m$ (т. е. R является правым обратным к Λ).

Это утверждение есть следствие стандартных фактов о разрешимости систем линейных уравнений. С другой стороны, легко сразу предъявить одно из таких отображений: $R(y) = \Lambda^T(\Lambda\Lambda^T)^{-1}y$, где Λ^T — матрица, транспонированная к Λ ($\Lambda\Lambda^T$ — обратимая матрица, как нетрудно проверить).

Перейдем непосредственно к доказательству теоремы. Пусть $\varepsilon > 0$ таково, что $\varepsilon\gamma < 1$. Согласно определению строгой дифференцируемости F в точке $\widehat{x}$ найдется $\delta = \delta(\varepsilon) > 0$ такое, что для всех

$x, x' \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$ справедливо неравенство

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

где $\theta = \varepsilon\gamma$.

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

Пусть $\xi \in V$ и $y \in W$. Рассмотрим последовательность (модифицированный метод Ньютона)

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

Докажем, что эта последовательность принадлежит $U_{\mathbb{R}^n}(\widehat{x}, \delta)$ и фундаментальна. Первое доказываем по индукции. Очевидно, что $x_0 = \xi \in U_{\mathbb{R}^n}(\widehat{x}, \delta/3) \subset U_{\mathbb{R}^n}(\widehat{x}, \delta)$. Пусть $x_s \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$, $1 \le s \le k$. Докажем, что $x_{k+1} \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$.

Используя последовательно (ii), оценку для правого обратного, равенство

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

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

$$|x_{k+1} - x_k| = |R(y - F(x_k))| \le \gamma|y - F(x_k)| = \gamma|F(x_k) - F(x_{k-1}) \\ - \Lambda(x_k - x_{k-1})| \le \theta|x_k - x_{k-1}| \le \ldots \le \theta^k|x_1 - \xi|. \qquad (iv)$$

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

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

т. е. $x_{k+1} \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$ и значит, вся последовательность $\{x_k\}$ принадлежит $U_{\mathbb{R}^n}(\widehat{x}, \delta)$.

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

$$|x_{k+l} - x_k| \le |x_{k+l} - x_{k+l-1}| + \ldots + |x_{k+1} - x_k| \le \\ \le \left(\theta^{k+l-1} + \ldots + \theta^k\right)|x_1 - \xi| < \frac{\theta^k\gamma}{1-\theta}|y - F(\xi)|, \qquad (vi)$$

что и доказывает фундаментальность последовательности $\{x_k\}$.

Положим $\varphi(\xi, y) = \lim_{k\to\infty} x_k$. Из (v) следует, что $\varphi(\xi, y) \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \subset U$. Переходя к пределу в (iii), получаем равенство $F(\varphi(\xi, y)) = y$. Полагая в (vi) $k = 0$ и переходя к пределу при $l \to \infty$, получаем, что $|\varphi(\xi, y) - \xi| \leq (\gamma/(1-\theta))|y - F(\xi)|$.

Функции x_k, $k \in \mathbb{N}$, как функции ξ и y, непрерывны на $V \times W$. Тогда переходя к пределу при $l \to \infty$ в (vi), получаем, что отображение φ есть равномерный предел непрерывных функций и значит, само непрерывно на $V \times W$. $\square$

При доказательстве необходимых условий экстремума удобно пользоваться так называемой теоремой Люстерника, которая есть простое следствие доказанного утверждения. Предварительно дадим понятие касательного пространства.

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

Следствие (Теорема Люстерника)**.** *Пусть выполнены условия теоремы и* $M = \{\, x \in \mathbb{R}^n \mid F(x) = F(\widehat{x}) \,\}$. *Тогда* $T_{\widehat{x}}M = \operatorname{Ker} F'(\widehat{x})$.

Доказательство. Пусть $h \in T_{\widehat{x}}M$ и r из определения касательного вектора. Тогда вследствие дифференцируемости F в точке $\widehat{x}$ имеем $0 = F(\widehat{x} + th + r(t)) - F(\widehat{x}) = tF'(\widehat{x})h + o(t)$, откуда следует, что $h \in \operatorname{Ker} F'(\widehat{x})$. Обратно, пусть $h \in \operatorname{Ker} F'(\widehat{x})$. Из утверждения теоремы следует существование такого $\varepsilon > 0$, что для всех $t \in (-\varepsilon, \varepsilon)$ справедливо равенство $F(\varphi(\widehat{x}+th, F(\widehat{x}))) = F(\widehat{x})$ и $|\varphi(\widehat{x}+th, F(\widehat{x})) - \widehat{x} - th| \leq K|F(\widehat{x}+th) - F(\widehat{x})| = K|F(\widehat{x}) + tF'(\widehat{x})h + o(t) - F(\widehat{x})| = |o(t)|$, т. е. $\varphi(\widehat{x} + th, F(\widehat{x})) = \widehat{x} + th + o(t)$ и значит, h — касательный вектор. $\square$

Исторический комментарий. Если классический метод Ньютона, о котором было сказано в начале параграфа, применить к нахождению решения уравнения $x^2 = 2$, то придем к последовательности $x_k = x_{k-1} + \frac{1}{2x_{k-1}}(2 - x_{k-1}^2) = \frac{1}{2}(x_{k-1} + \frac{2}{x_{k-1}})$, известной еще Герону (I век до н. э.). Модифицированный метод Ньютона стал активно применяться в двадцатом веке. В бесконечномерном случае этот метод изучался Л. В. Канторовичем.

3. Выпуклые множества и теоремы отделимости

Пусть $x, y \in \mathbb{R}^n$. Напомним, что множество $[x, y] = \{z \in \mathbb{R}^n \mid z = (1-\alpha)x + \alpha y,\ 0 \leq \alpha \leq 1\}$ называется *отрезком* (соединяющим точки x и y). Множество $(x, y) = \{z \in \mathbb{R}^n \mid z = (1-\alpha)x + \alpha y,\ 0 < \alpha < 1\}$ называется *интервалом.*

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 $A_1+\cdots+A_n := \{x\in\mathbb{R}^n \mid x = x_1+\cdots+x_n,\ x_i\in A_i,\ i=1,\ldots,n\}$ — выпуклое множество.

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

Упражнение 3.1. *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 $\operatorname{co}\mathrm{A}$.

Упражнение 3.2. *Доказать, что* $\operatorname{co}\mathrm{A}$ *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* $\sum_{i=1}^m\alpha_i = 1$.

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

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

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

Доказательство. (a) Пусть $z\in(x,y)$, тогда $z=(1-\alpha)x+\alpha y$ для некоторого $0<\alpha<1$. Так как $x\in\operatorname{int}\mathrm{A}$, то существует такое $\delta>0$, что $U_{\mathbb{R}^n}(x,\delta)\subset A$. Множество $(1-\alpha)U_{\mathbb{R}^n}(x,\delta)+\alpha y$ является окрестностью z (см. упр. 1.12) и принадлежит A, т. е. $z\in\operatorname{int}\mathrm{A}$.

(b) Выпуклость $\operatorname{int}\mathrm{A}$ сразу следует из (a). Докажем выпуклость $\operatorname{cl}\mathrm{A}$. Пусть $x_i\in\operatorname{cl}\mathrm{A}$, $i=1,2$ и $x=(1-\alpha)x_1+\alpha x_2$, $0\leq\alpha\leq 1$. Для любого $\delta>0$ найдутся $\overline{x}_i\in(x_i+U_{\mathbb{R}^n}(0,\delta))\cap A$, $i=1,2$. Положим $\overline{x}=(1-\alpha)\overline{x}_1+\alpha\overline{x}_2$. Тогда $\overline{x}\in A$ и $\overline{x}\in(1-\alpha)(x_1+U_{\mathbb{R}^n}(0,\delta))+\alpha(x_2+U_{\mathbb{R}^n}(0,\delta))=x+U_{\mathbb{R}^n}(0,\delta)$, т. е. каждая окрестность точки $x=(1-\alpha)x_1+\alpha x_2$ имеет непустое пересечение с A и значит, $x\in\operatorname{cl}\mathrm{A}$. $\square$

Пусть $x^*\in(\mathbb{R}^n)^*$, $x^*\neq 0$ и $\gamma\in\mathbb{R}$. Множество $H=H(x^*,\gamma)=\{\,x\in\mathbb{R}^n \mid x^*\cdot x=\gamma\,\}$ называется *гиперплоскостью.* Гиперплоскость порождает два полупространства $H_+(x^*,\gamma)=\{\,x\in\mathbb{R}^n \mid x^*\cdot x\leq\gamma\,\}$ и $H_-(x^*,\gamma)=\{\,x\in\mathbb{R}^n \mid x^*\cdot x\geq\gamma\,\}$.

Упражнение 3.3. *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

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

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

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

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

Доказательство. Пусть $\overline{x} \in A$. Обозначим $r = |\overline{x} - b|$ и $A_1 = A \cap B_{\mathbb{R}^n}(b, r)$. Рассмотрим функцию $f\colon \mathbb{R}^n \to \mathbb{R}$, определенную равенством $f(x) = |x - b|$ (расстояние от точки x до точки b). Эта функция непрерывна (упр. 1.10) и поэтому на ограниченном замкнутом множестве (компакте) A_1 она достигает, по теореме Вейерштрасса, своей нижней грани в некоторой точке $\widehat{x} \in A_1$. Положим $x^* = (\widehat{x} - b)^T$ и рассмотрим гиперплоскость $H = \{\, x \in \mathbb{R}^n \mid x^* \cdot x = x^* \cdot \widehat{x} \,\}$. Покажем, что вектор b и множество A находятся в разных полупространствах. Действительно, $x^* \cdot b = x^* \cdot (b - \widehat{x} + \widehat{x}) = x^* \cdot (-x^{*T} + \widehat{x}) = -|x^*|^2 + x^* \cdot \widehat{x} < x^* \cdot \widehat{x}$. Теперь покажем, что A находится в другом полупространстве, т. е. $x^* \cdot x \geq x^* \cdot \widehat{x}$ для любого $x \in A$. Предположим противное, что существует такой элемент $x_0 \in A$, что $x^* \cdot x_0 < x^* \cdot \widehat{x}$, или $x^* \cdot (x_0 - \widehat{x}) < 0$. Поскольку A выпукло, то $(1-t)\widehat{x} + t x_0 \in A$ при $0 \leq t \leq 1$ и при малых t (учитывая, что $x^* \cdot (x_0 - \widehat{x}) < 0$) будем иметь

$$\begin{aligned} |(1-t)\widehat{x} + t x_0 - b|^2 = |\widehat{x} - b + t(x_0 - \widehat{x})|^2 = |x^{*T} + t(x_0 - \widehat{x})|^2 = \\ = |x^*|^2 + 2t\, x^* \cdot (x_0 - \widehat{x}) + t^2 |x_0 - \widehat{x}|^2 < |x^*|^2 = |\widehat{x} - b|^2. \end{aligned}$$

Величина справа не превосходит r^2, следовательно, точки $(1-t)\widehat{x} + t x_0$ принадлежат множеству A_1, но это противоречит тому, что $\widehat{x}$ — минимум функции f на этом множестве. Итак, $x^* \cdot b < x^* \cdot \widehat{x} \leq x^* \cdot x$ для всех $x \in A$, или $x^* \cdot b < \inf_{x \in A} x^* \cdot x$, т. е. A и b строго отделимы. □

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

Доказательство. Положим $C = A - \operatorname{int} \mathrm{B}$. Множество C выпукло (как разность двух выпуклых множеств) и $0 \notin C$. Если $0 \notin \operatorname{cl} \mathrm{C}$, то так как множество $\operatorname{cl} C$ также выпукло, то по предыдущей теореме найдется ненулевой вектор $x^* \in (\mathbb{R}^n)^*$ такой, что $x^* \cdot x < x^* \cdot 0 = 0$ для всех $x \in \operatorname{cl} C$ и тем более для $x = a - b \in A - \operatorname{int} B = C$, т. е. $x^* \cdot a < x^* \cdot b$ для всех $a \in A$ и $b \in \operatorname{int} B$. Отсюда легко следует, что $x^* \cdot a \le x^* \cdot b$ для всех $a \in A$ и $b \in B$ (продумайте это) и значит, множества A и B отделимы.

Пусть $0 \in \operatorname{cl} C$. Так как $C = \cup_{a \in A}(a - \operatorname{int} B)$, то $\operatorname{int} C \neq \emptyset$. Если $x \in \operatorname{int} C$, то для любого $\lambda > 0$ вектор $-\lambda x$ не принадлежит $\operatorname{cl} C$, ибо в противном случае (согласно Предложению) $0 = \lambda(1+\lambda)^{-1}x + (1+\lambda)^{-1}(-\lambda x)$ принадлежал бы $\operatorname{int} C$ и тем самым C. Следовательно, существует последовательность точек $\{x_k\}$ (например, $x_k = -\lambda_k x$, где $\lambda_k > 0$ и $\lambda_k \to 0$ при $k \to \infty$), не принадлежащих $\operatorname{cl} C$ таких, что $x_k \to 0$ при $k \to \infty$. По доказанной второй теореме отделимости найдутся ненулевые векторы $x_k^* \in (\mathbb{R}^n)^*$, что

$$x_k^* \cdot x < x_k^* \cdot x_k \qquad (i)$$

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

Следует заметить, что, на самом деле, теорема справедлива и без предположения непустоты внутренности какого-либо из множеств. Для доказательства надо дополнительно провести еще одно рассуждение, показывающее, что выпуклое множество с пустой внутренностью содержится в некоторой гиперплоскости. В дальнейшей надо это не понадобится и поэтому ограничиваемся здесь доказанным утверждением.

Отметим еще, что теоремы отделимости аналогично формулируются для выпуклых множеств в $(\mathbb{R}^n)^*$. Двойственным к этому пространству будет пространство $\mathbb{R}^n$, как уже отмечалось ранее. Например, если $x \in \mathbb{R}^n$, $x \neq 0$, и $\gamma \in \mathbb{R}$, то гиперплоскость, порожденная элементом x (линейным функционалом на $(\mathbb{R}^n)^*$) и числом γ имеет вид $H(x,\gamma) = \{\, x^* \in (\mathbb{R}^n)^* \mid x^* \cdot x = \gamma \,\}$.

Исторический комментарий. Первый результат, связанный с отделимостью, а именно, теорема отделимости точки от выпуклого множества в конечномерном случае была доказана Минковским (1910 г.). Впоследствии, когда были доказаны теоремы об отделимости множеств в бесконечномерных пространствах, первой теоремой отделимости стали называть теорему об отделимости выпуклых множеств, а второй — теорему об отделимости точки от выпуклого замкнутого множества.

4. **Необходимые и достаточные условия экстремума для гладких задач без ограничений**

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

В этом пункте будет доказан изначальный результат теории экстремума — теорема Ферма (необходимое условие экстремума для гладких задач без ограничений), а также для таких задач будут получены необходимые и достаточные условия экстремума второго порядка (т. е. когда участвуют вторые производные).

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

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

С точки зрения общей задачи (P) здесь $U = C = X$, т. е. (P_1) — задача без ограничений.

Теорема (Ферма, необходимые условия экстремума первого порядка 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. Допустим, что (линейный функционал) $f'(\widehat{x})$ отличен от нуля. Тогда найдется элемент $x \in \mathbb{R}^n$, $x \neq 0$, такой, что $f'(\widehat{x}) \cdot x \neq 0$. Дифференцируемость f в $\widehat{x}$ означает, что $f(\widehat{x} + tx) = f(\widehat{x}) + t(f'(\widehat{x}) \cdot x + o(t)/t)$. Отсюда вытекает, что $f(\widehat{x} + tx) > f(\widehat{x})$ для всех малых $t \neq 0$ одного знака с $f'(\widehat{x}) \cdot x$ и $f(\widehat{x} + tx) < f(\widehat{x})$, если знаки противоположны. Получили противоречие с тем, что $\widehat{x}$ — локальный экстремум. $\square$

Исторический комментарий. Впервые мысль о том, что в точке максимума или минимума функции “приращение несущественно”, высказал И.Кеплер в 1615 году. В словах И. Ньютона, вынесенных в эпиграф, говорится об аналогичном наблюдении. Аналитически этот факт выразил П. Ферма в 1638 году в письме к Р. Декарту. (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). С 1638 года обычно отсчитывают начало теории экстремума.

Теорема (необходимые и достаточные условия экстремума второго порядка для гладких задач без ограничений). *Пусть в задаче (P_1) функция f дважды дифференцируема в $\widehat{x} \in U$. Тогда*

1) *если* $\widehat{x}$ — *локальный минимум (максимум) функции* f, *то для любого* $h \in \mathbb{R}^n$ *выполняется неравенство*
$$h^T f''(\widehat{x})h \geq 0 \ \ (\leq 0);$$
2) *если* $f'(\widehat{x}) = 0$ *и для любого* $h \in \mathbb{R}^n$, $h \neq 0$, *выполняется неравенство*
$$h^T f''(\widehat{x})h > 0 \ \ (< 0),$$
то $\widehat{x}$ — *локальный минимум (максимум) функции* f.

Доказательство. 1) Пусть, для определенности, $\widehat{x}$ — локальный минимум функции f. По теорема Ферма $f'(\widehat{x}) = 0$ и тогда по формуле Тейлора для любого $h \in \mathbb{R}^n$ и достаточно малых $t \in \mathbb{R}$ имеем $0 \leq f(\widehat{x} + th) - f(\widehat{x}) = (t^2/2)h^T f''(\widehat{x})h + o(t^2)$. Отсюда (деля на t^2 и устремляя t к нулю) получаем требуемое неравенство. Для локального максимума рассуждения аналогичны.

2) Пусть $h^T f''(\widehat{x})h > 0$ для любого ненулевого $h \in \mathbb{R}^n$. Функция $h \to h^T f''(\widehat{x})h$ непрерывна на $\mathbb{R}^n$. Обозначим через m ее минимальное значение на единичной сфере $S^{n-1} = \{x \in \mathbb{R}^n \mid |x| = 1\}$. Ясно, что $m > 0$. Снова по формуле Тейлора имеем для любого $h \neq 0$ такого, что $\widehat{x} + h \in U$ (учитывая, что $h/|h|$ и $h^T/|h|$ принадлежат S^{n-1})
$$f(\widehat{x} + h) = f(\widehat{x}) + \frac{1}{2}h^T f''(\widehat{x})h + o(|h|^2) = f(\widehat{x}) +$$
$$\frac{|h|^2}{2}\left(\frac{h^T}{|h|} f''(\widehat{x})\frac{h}{|h|} + \frac{o(|h|^2)}{|h|^2}\right) \geq f(\widehat{x}) + \frac{|h|^2}{2}\left(m + \frac{o(|h|^2)}{|h|^2}\right).$$
Выражение в скобках в правой части неравенства положительно для достаточно малых h. Следовательно, $f(\widehat{x}+h) > f(\widehat{x})$ для таких h и значит, $\widehat{x}$ — локальный минимум функции f. Для локально максимума рассуждения аналогичны. □

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

5. Необходимые и достаточные условия экстремума для гладких задач с ограничениями

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

Рекомендации Лагранжа относятся к случаю, когда ограничения задаются уравнениями (связи). Мы сразу рассмотрим случай, когда ограничения задаются и уравнениями и неравенствами.

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

$$f_0(x) \to \min, \quad f_i(x) \le 0, \quad 1 \le i \le m', \qquad f_i(x) = 0, \quad m'+1 \le i \le m, \quad (P_2)$$

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

С точки зрения общей задачи (P) здесь $X = U$ и множество ограничений $C = \{x \in U \mid f_i(x) \le 0,\ 1 \le i \le m',\ f_i(x) = 0,\ m'+1 \le i \le m\}$.

Свяжем с задачей (P_2) функцию Лагранжа $\mathcal{L}\colon U \times (\mathbb{R}^{m+1})^* \to \mathbb{R}$, $\mathcal{L}(x, \overline{\lambda}) = \sum_{i=0}^m \lambda_i f_i(x)$.

Теорема (Правило множителей Лагранжа для задачи (P_2))**.** *Пусть $\widehat{x} \in U$ — локальный минимум в задаче (P_2). Если функции f_i, $i = 0, 1, \dots, m'$, дифференцируемы, а функции f_i, $i = m'+1, \dots, m$, строго дифференцируемы в $\widehat{x}$, то найдется такой ненулевой набор множителей Лагранжа $\overline{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_m)$, что выполняются*

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

Теперь заметим, что если векторы $f'_{m'+1}(\widehat{x}), \ldots, f'_m(\widehat{x})$ линейно зависимы, т. е. найдутся такие числа $\lambda_{m'+1}, \ldots, \lambda_m$, не все равные нулю, что $\sum_{i=m'+1}^m \lambda_i f'_i(\widehat{x}) = 0$, то теорема доказана, так как остальные множители Лагранжа можно взять нулевыми.

Итак, пусть выполнено (c) и векторы $f'_{m'+1}(\widehat{x}), \ldots, f'_m(\widehat{x})$ линейно независимы. Обозначим $A = \operatorname{co}\{\, f'_0(\widehat{x}), f'_1(\widehat{x}), \ldots, f'_{m'}(\widehat{x})\,\}$ и $L = \operatorname{span}\{f'_{m'+1}(\widehat{x}), \ldots, f'_m(\widehat{x})\,\}$ — линейная оболочка векторов $f'_{m'+1}(\widehat{x}), \ldots, f'_m(\widehat{x})$. Тогда утверждение теоремы (т. е. условия (a) и (b) с $\overline{\lambda} \neq 0$) равносильны тому, что $0 \in A + L$. Действительно, если $0 \in A + L$, то $\sum_{i=0}^m \lambda_i f'_i(\widehat{x}) = 0$, где $\lambda_i \geq 0$, $i = 0, 1, \ldots, m'$ и $\sum_{i=0}^{m'+1} \lambda_i = 1$, т. е. $\overline{\lambda} \neq 0$. Обратно, пусть выполнены соотношения (a) и (b) с $\overline{\lambda} \neq 0$. Из (a) и линейной независимости векторов $f'_{m'+1}(\widehat{x}), \ldots, f'_m(\widehat{x})$ вытекает, что среди чисел $\lambda_0, \ldots, \lambda_{m'}$ есть ненулевые. Деля равенство (a) на $\sum_{i=0}^{m'+1} \lambda_i$, получаем, что $0 \in A + L$.

Утверждение теоремы докажем от противного — предположим, что $0 \notin A + L$ и придем к противоречию с тем, что $\widehat{x}$ — локальный минимум. Множество A компактно, а L замкнуто и поэтому $A + L$ — выпуклое и замкнутое (см. упр. 1.9) подмножество в $(\mathbb{R}^n)^*$. По второй теореме отделимости найдется ненулевой элемент $\overline{x} \in \mathbb{R}^n$ такой, что $\sup_{x^* \in A+L} x^* \cdot \overline{x} < 0 \cdot \overline{x} = 0$, или

$$\sup_{\substack{\sum_{i=0}^{m'} \lambda_i = 1 \\ \lambda_i \geq 0,\ 0 \leq i \leq m'}} \sum_{i=0}^{m'} \lambda_i f'_i(\widehat{x}) \cdot \overline{x} + \sup_{\substack{\lambda_i \in \mathbb{R} \\ m'+1 \leq i \leq m}} \sum_{i=m'+1}^{m} \lambda_i f'_i(\widehat{x}) \cdot \overline{x} < 0. \qquad (i)$$

Оба слагаемых слева, очевидно, конечны. Конечность второго слагаемого означает, что $f'_i(\widehat{x}) \cdot \overline{x} = 0$, $i = m' + 1, \ldots, m$, а тогда из первого слагаемого следует, что $f'_i(\widehat{x}) \cdot \overline{x} < 0$, $i = 0, 1, \ldots, m'$.

Теперь воспользуемся теоремой Люстерника. Определим отображение $F\colon U \to \mathbb{R}^{m-m'}$ по формуле $F(x) = (\, f_{m'+1}(x), \ldots, f_m(x)\,)^T$. Это отображение строго дифференцируемо в $\widehat{x}$ в силу строгой дифференцируемости его компонент и $F'(\widehat{x})(\mathbb{R}^n) = \mathbb{R}^{m-m'}$, поскольку векторы $f'_{m'+1}(\widehat{x}), \ldots, f'_m(\widehat{x})$ линейно независимы. Равенства $f'_i(\widehat{x}) \cdot \overline{x} = 0$, $i = m' + 1, \ldots, m$, означают, что $\overline{x} \in \operatorname{Ker} F'(\widehat{x})$. Следовательно, по теореме Люстерника, $\overline{x}$ — касательный вектор ко множеству $\{\, x \in \mathbb{R}^n \mid F(x) = F(\widehat{x}) = 0\,\}$, т. е. существуют $\varepsilon > 0$ и отображение $r\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ такие, что $F(\widehat{x} + t\overline{x} + r(t)) = 0$ (или $f_i(\widehat{x} + t\overline{x} + r(t)) = 0$, $i = m' + 1, \ldots, m$) для $t \in (-\varepsilon, \varepsilon)$, где $r(t) = o(t)$ при $t \to 0$.

Далее, так как $f'_i(\widehat{x}) \cdot \overline{x} < 0$, $i = 0, 1, \ldots, m'$, то для достаточно малых $t > 0$ имеем $f_i(\widehat{x} + t\overline{x} + r(t)) = f_i(\widehat{x}) + t f'_i(\widehat{x}) \cdot \overline{x} + o(t) < 0$, если $i = 1, \ldots, m'$ и $f_0(\widehat{x} + t\overline{x} + r(t)) = f_0(\widehat{x}) + t f'_0(\widehat{x}) \cdot \overline{x} + o(t) < f_0(\widehat{x})$, т. е.

в любой окрестности $\widehat{x}$ есть точка вида $\widehat{x}+t\overline{x}+r(t)$, $t>0$, которая допустима в задаче (P_2) и на ней значение функционала f_0 меньше, чем на $\widehat{x}$ в противоречие с тем, что $\widehat{x}$ — локальный минимум. $\square$

Классическое правило множителей Лагранжа относится к задачам с ограничениями типа равенств и при этом обычно требуют непрерывную дифференцируемость всех функций. Ясно, что этот результат сразу следует из доказанной теоремы, так как непрерывная дифференцируемость влечет строгую дифференцируемость. Но, отдавая дань традиции, сформулируем отдельно эту теорему и воспользовавшись более сильными условиями гладкости, дадим ее доказательство, которое более просто, чем приведенное выше.

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.*

Функция Лагранжа задачи (P_2') имеет вид: $\mathcal{L}(x,\overline{\lambda})=\sum_{i=0}^m \lambda_i f_i(x)$, где $\overline{\lambda}=(\lambda_0,\lambda_1,\ldots,\lambda_m)$.

Теорема (Классическое правило множителей Лагранжа для задачи (P_2'))**.** *Пусть $\widehat{x}\in U$ — локальный экстремум в задаче (P_2') и функции f_i, $i=0,1,\ldots,m$, непрерывно дифференцируемы на U. Тогда найдется такой ненулевой набор множителей Лагранжа $\overline{\lambda}=(\lambda_0,\lambda_1,\ldots,\lambda_m)$, что*

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

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 противное и придем к противоречию с тем, что $\widehat{x}$ — локальный экстремум. Действительно, линейная независимость векторов $f_0'(\widehat{x}),f_1'(\widehat{x}),\ldots,f_m'(\widehat{x})$ означает, что производная в точке $\widehat{x}$ отображения $F\colon U\to\mathbb{R}^{m+1}$, определенного формулой $F(x) = (f_0(x),f_1(x),\ldots,f_m(x))^T$, такова, что $F'(\widehat{x})(\mathbb{R}^n)=\mathbb{R}^{m+1}$. Кроме того, так как функции $f_0,f_1,\ldots,f_m$ непрерывно дифференцируемы на U (в частности, в точке $\widehat{x}$), то они строго дифференцируемо в $\widehat{x}$ и значит, отображение F строго дифференцируемо в этой точке. Следовательно, применима теорема о разрешимости системы уравнений, согласно которой найдется окрестности V и W точек $\widehat{x}$ и $F(\widehat{x})$, непрерывное отображение $\varphi\colon V\times W\to U$ такие, что (при $\xi=\widehat{x}$) $F(\varphi(\widehat{x},y))=y$ для всех $y\in W$. Так как $F(\widehat{x})=(f_0(\widehat{x}),0,\ldots,0)^T$, то для достаточно малых по модулю ν точки $y_\nu=(f_0(\widehat{x})+\nu,0,\ldots,0)^T$, очевидно, принадлежат W. Обозначим $x_\nu=\varphi(\widehat{x},y_\nu)$. Поскольку

$F(x_\nu) = y_\nu$, или $(f_0(x_\nu), f_1(x_\nu), \dots, f_m(x_\nu))^T = (f_0(\widehat{x}) + \nu, 0, \dots, 0)^T$ и отображение φ непрерывно ($x_\nu \to \widehat{x}$ при $\nu \to 0$), то в любой окрестности $\widehat{x}$ находятся допустимые в задаче (P_2') точки x_ν и при этом, $f_0(x_\nu) < f_0(\widehat{x})$ ($f_0(x_\nu) > f_0(\widehat{x})$), если $\nu < 0$ ($\nu > 0$) в противоречие с тем, что $\widehat{x}$ — локальный экстремум.

Последнее утверждение теоремы проверяется без труда. $\square$

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

Теорема (Необходимые и достаточные условия экстремума второго порядка для задачи (P_2'))**.** *Пусть в задаче (P_2') функции f_i, $i = 0, 1, \dots, m$, дважды дифференцируемы в точке $\widehat{x}$ и векторы $f_1'(\widehat{x}), \dots, f_m'(\widehat{x})$ линейно независимы. Тогда*

1) *если $\widehat{x}$ — локальный минимум (максимум) в (P_2'), то существует набор множителей Лагранжа $\overline{\lambda} = (1, \lambda_1, \dots, \lambda_m)$ такой, что для любого $h \in \mathbb{R}^n$, для которого $f_i'(\widehat{x}) \cdot h = 0$, $i = 1, \dots, m$, справедливо неравенство*
$$h^T \mathcal{L}_{xx}(\widehat{x}, \overline{\lambda}) h \geq 0 \ \ (\leq 0);$$
2) *если $\widehat{x}$ — допустимая точка в (P_2'), существует такой набор множителей Лагранжа $\overline{\lambda} = (1, \lambda_1, \dots, \lambda_m)$, что $\mathcal{L}_x(\widehat{x}, \overline{\lambda}) = 0$ и для любого ненулевого $h \in \mathbb{R}^n$, для которого $f_i'(\widehat{x}) \cdot h = 0$, $i = 1, \dots, m$, справедливо неравенство*
$$h^T \mathcal{L}_{xx}(\widehat{x}, \overline{\lambda}) h > 0 \ \ (< 0),$$
то $\widehat{x}$ — локальный минимум (максимум) в задаче (P_2').

Доказательство. 1) Воспользуемся теоремой Люстерника. Для этого определим отображение $F\colon U \to \mathbb{R}^m$ по правилу $F(x) = (\, f_1(x), \dots, f_m(x)\,)^T$. Так как функции f_i, $i = 0, 1, \dots, m$, дважды дифференцируемы в $\widehat{x}$, то их производные непрерывны в этой точке и значит, строго дифференцируемы в $\widehat{x}$ (см. п. 1.3), а тогда отображение F строго дифференцируемо в $\widehat{x}$. Далее, $F'(\widehat{x})(\mathbb{R}^n) = \mathbb{R}^m$, поскольку векторы $f_1'(\widehat{x}), \dots, f_m'(\widehat{x})$ линейно независимы. Если $h \in \mathbb{R}^n$ таково, что $f_i'(\widehat{x}) \cdot h = 0$, $i = 1, \dots, m$, то это означает, что $h \in \operatorname{Ker} F'(\widehat{x})$. Следовательно, по теореме Люстерника, h — касательный вектор ко множеству $\{\, x \in \mathbb{R}^n \mid F(x) = F(\widehat{x}) = 0 \,\}$, т. е. существуют $\varepsilon > 0$ и отображение $r\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ такие, что $F(\widehat{x} + th + r(t)) = 0$ (или $f_i(\widehat{x} + th + r(t)) = 0$, $i = 1, \dots, m$) для $t \in (-\varepsilon, \varepsilon)$, где $r(t) = o(t)$ при $t \to 0$.

Таким образом, для малых t точки $\widehat{x} + th + r(t)$ допустимы в задаче (P_2'). Так как $\widehat{x}$ — локальный экстремум, то согласно правилу множителей Лагранжа найдется ненулевой набор $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m)$ такой, что $\mathcal{L}_x(\widehat{x}, \overline{\lambda}) = 0$, причем $\lambda_0 \neq 0$ и можно считать, что $\lambda_0 = 1$. Учитывая это, будем иметь по формуле Тейлора, считая для определенности, что $\widehat{x}$ — локальный минимум:

$$0 \le f_0(\widehat{x} + th + r(t)) - f_0(\widehat{x}) = \mathcal{L}(\widehat{x} + th + r(t), \overline{\lambda}) - \mathcal{L}(\widehat{x}, \overline{\lambda})$$
$$= \mathcal{L}_x(\widehat{x}, \overline{\lambda}) \cdot (th + r(t))$$
$$+ \frac{1}{2}(th + r(t))^T \mathcal{L}_{xx}(\widehat{x}, \overline{\lambda})(th + r(t)) + o(t^2) = \frac{t^2}{2} h^T L_{xx}(\widehat{x}, \overline{\lambda}) h + o(t^2),$$

откуда следует нужное утверждение.

Для случая локального максимума рассуждения аналогичны.

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

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

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

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

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

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

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

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

Последовательность $\{|h_k|^{-1} h_k\}$ принадлежит единичной сфере в $\mathbb{R}^n$ и поэтому из нее можно выделить подпоследовательность, сходящуюся к некоторому $\overline{h} \in \mathbb{R}^n$, $|\overline{h}| = 1$. Будем, для простоты, считать, что сама последовательность сходится к $\overline{h}$. Переходя к пределу в (iv), получаем неравенство (i). Осталось проверить, что $f_i'(\widehat{x}) \cdot \overline{h} = 0$, $1 \le i \le m$. Для всех $k \in \mathbb{N}$ имеем

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

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

6. **Выпуклые задачи**

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

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

С каждой функцией $f\colon X\to\overline{\mathbb{R}}$ свяжем два множества $\operatorname{dom}f=\{\,x\in X\mid f(x)<+\infty\,\}$ и $\operatorname{epi}f=\{\,(x,\alpha)\in X\times\mathbb{R}\mid \alpha\ge f(x),\ x\in\operatorname{dom}f\,\}$, которые называются соответственно *эффективным множеством* и *надграфиком* (или *эпиграфом*) *функции* f.

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

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

Функцию f называют *собственной*, если $\operatorname{dom}f\ne\emptyset$ и $f(x)>-\infty$ для всех $x\in X$.

Упражнение 6.1. *Доказать, что собственная функция f на X 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*

$$f(\alpha_1x_1+\alpha_2x_2)\le\alpha_1f(x_1)+\alpha_2f(x_2),$$

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)

$$\partial f(\widehat{x})=\{\,x^*\in(\mathbb{R}^n)^*\mid f(x)-f(\widehat{x})\ge x^*\cdot(x-\widehat{x}),\ \ \forall\,x\in\mathbb{R}^n\,\}.$$

Упражнение 6.2. *Доказать, что если субдифференциал не пуст, то это выпуклое замкнутое множество.*
Упражнение 6.3. *Привести пример функции, у которой субдифференциал в данной точке пуст.*

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

Предложение. *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: $f((1-\alpha)\widehat{x}+\alpha x) \le (1-\alpha)f(\widehat{x})+\alpha f(x)$, откуда

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

Переходя к пределу при $\alpha \to 0$, получаем, что $f'(\widehat{x}) \cdot (x-\widehat{x}) \le f(x)-f(\widehat{x})$, т. е. $f'(\widehat{x}) \in \partial f(\widehat{x})$.

Обратно, если $x^* \in \partial f(\widehat{x})$, то для любого $x \in \mathbb{R}^n$ и любого $t > 0$ имеем $f(\widehat{x}+tx)-f(\widehat{x}) \ge tx^* \cdot x$. Следовательно,

$$\frac{f(\widehat{x}+tx)-f(\widehat{x})}{t} \ge x^* \cdot x,$$

т. е. $f'(\widehat{x}) \cdot x \ge x^* \cdot x$ для любого x и значит, $x^* = f'(\widehat{x})$. ☐

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

$$f(x) \to \min, \quad x \in \mathbb{R}^n, \tag{P_3}$$

которая называется *выпуклой задачей без ограничений.*

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

Теорема (Ферма для выпуклых функций)**.** *Точка $\widehat{x}$ является минимумом в задаче (P_3) тогда и только тогда, когда $0 \in \partial f(\widehat{x})$.*

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

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

$$f_0(x) \to \min, \quad f_i(x) \le 0, \quad 1 \le i \le m, \quad x \in A \tag{P_4}$$

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

С задачей (P_4) связывается функция Лагранжа $\mathcal{L}\colon X \times (\mathbb{R}^{m+1})^* \to \mathbb{R}$, $\mathcal{L}(x, \overline{\lambda}) = \sum_{i=0}^{m} \lambda_i f_i(x)$, которая имеет тот же вид, что и в задаче (P_2).

Теорема (Каруша–Куна–Таккера)**.** *Если $\widehat{x}$ — минимум в задаче (P_4), то найдется такой ненулевой набор множителей Лагранжа $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m)$, что выполняются*

(*a*) $\min_{x \in A} \mathcal{L}(x, \overline{\lambda}) = \mathcal{L}(\widehat{x}, \overline{\lambda})$ *(условие минимума);*
(*b*) $\lambda_i \ge 0,\ i = 0, 1, \ldots, m$ *(условие неотрицательности);*
(*c*) $\lambda_i f_i(\widehat{x}) = 0,\ i = 1, \ldots, m$ *(условие дополняющей нежесткости).*

Если существует допустимая в (P_4) точка $\widehat{x}$ и набор множителей Лагранжа $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m)$, удовлетворяющие условиям (a), (b) и (c) и при этом $\lambda_0 > 0$, то $\widehat{x}$ — решение задачи (P_4).

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

Доказательство. Пусть $\widehat{x}$ — решение задачи (P_4). Рассмотрим множества $B = \{\, b = (b_0, b_1, \ldots, b_m)^T \in \mathbb{R}^{m+1} \mid \exists\, x \in A : f_0(x) - f_0(\widehat{x}) \le b_0,\ f_i(x) \le b_i,\ i = 1, \ldots, m \,\}$ и $C = \{\, c \in \mathbb{R}^{m+1} \mid c = (c_0, 0, \ldots, 0)^T,\ c_0 < 0 \,\}$. Покажем, что эти множества выпуклы, внутренность B непуста и $B \cap C = \emptyset$.

Действительно, выпуклость C очевидна. Если в B взять $x = \widehat{x}$, то ясно, что $\mathbb{R}^{m+1}_+ \subset B$ и тем самым $\operatorname{int} \mathrm{B} \ne \emptyset$. Докажем выпуклость B. Пусть $b = (b_0, \ldots, b_m)^T, b' = (b'_0, \ldots, b'_m)^T \in B$ и $\alpha \in [0, 1]$. Надо показать, что $(1 - \alpha)b + \alpha b \in B$. Так как b и b' принадлежат B, то найдутся такие $x, x' \in A$, что $f_0(x) - f_0(\widehat{x}) \le b_0$, $f_i(x) \le b_i$, $i = 1, \ldots, m$ и $f_0(x') - f_0(\widehat{x}) \le b'_0$, $f_i(x') \le b'_i$, $i = 1, \ldots, m$. Положим $x_\alpha = (1 - \alpha)x + \alpha x'$. Тогда $x_\alpha \in A$. Далее $f_0(x_\alpha) - f_0(\widehat{x}) = f_0((1 - \alpha)x + \alpha x') - f_0(\widehat{x}) \le (1 - \alpha) f_0(x) + \alpha f_0(x') - f_0(\widehat{x}) \le (1 - \alpha) b_0 + \alpha b'_0 - f_0(\widehat{x})$ и аналогично $f_i(x_\alpha) \le (1 - \alpha) b_i + \alpha b'_i$, $i = 1, \ldots, m$, т. е. $(1 - \alpha)b + \alpha b' \in B$ и тем самым B выпукло.

Если допустить, что $B \cap C \ne \emptyset$, то найдется элемент $\widetilde{x} \in A$ такой, что $f_i(\widetilde{x}) \le 0$, $i = 1, \ldots, m$, и $f_0(\widetilde{x}) - f_0(\widehat{x}) \le c_0 < 0$, т. е. вектор $\widetilde{x}$ допустим в задаче (P_4) и $f_0(\widetilde{x}) < f_0(\widehat{x})$ в противоречие с тем, что $\widehat{x}$ — минимум.

Согласно первой теореме отделимости найдется такой ненулевой вектор $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in (\mathbb{R}^{m+1})^*$, что $\sup_{c \in C} \sum_{i=0}^{m} \lambda_i c_i \le$

$\inf_{b\in B}\sum_{i=0}^m \lambda_i b_i$, или $\sup_{c_0<0}\lambda_0 c_0 \le \inf_{b\in B}\sum_{i=0}^m \lambda_i b_i$. Отсюда следует, что

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

Подставляя в (i) векторы $(1,0,\dots,0)^T,\dots,(0,\dots,0,1)^T$, получаем, что $\lambda_i\ge 0$, $i=0,1,\dots,m$, и утверждение (b) доказано.

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

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

Докажем второе утверждение теоремы. Пусть x — допустимый вектор в задаче (P_4). Тогда, используя это обстоятельство вместе с (b), затем (a) и (c), будем иметь

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

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

Деля на λ_0, получаем требуемое.

Докажем последнее 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). $\square$

6.3. **Минимум максимума гладких функций.** Здесь мы рассмотрим задачу о минимуме максимума конечного числа гладких функций, в которой естественным образом переплетены гладкая и выпуклая структуры. Предварительно приведем два определения.

Пусть U — окрестность точки $\widehat{x}\in\mathbb{R}^n$ и $f\colon U\to\mathbb{R}$. Скажем, что f выпукла в $\widehat{x}$, если найдутся такие $x^*\in(\mathbb{R}^n)^*$ и окрестность $U_0\subset U$ точки $\widehat{x}$, что $f(x)-f(\widehat{x})\ge x^*\cdot(x-\widehat{x})$ для всех $x\in U_0$. Другими словами, в окрестности U_0 график функции f лежит выше гиперплоскости $y=x^*\cdot(x-\widehat{x})+f(\widehat{x})$.

Пусть U — открытое подмножество $\mathbb{R}^n$, $f_i\colon U\to\mathbb{R}$, $i=1,\dots,m$. Рассмотрим задачу

$$f(x)=\max_{1\le i\le m}\{f_i(x)\}\to\min,\quad x\in U. \tag{P_4'}$$

Теорема. *Пусть в задаче* (P_4') *функции* f_i, $1 \le i \le m$, *дифференцируемы в* $\widehat{x} \in U$ *и* $f_1(\widehat{x}) = \ldots = f_m(\widehat{x})$. *Тогда условие* $0 \in \operatorname{co}\{\, f_1'(\widehat{x}), \ldots, f_m'(\widehat{x}) \,\}$ *необходимо, а если функции* f_i, $1 \le i \le m$, *выпуклы в* $\widehat{x}$, *то и достаточно для того, чтобы* $\widehat{x}$ *было локальным минимумом функции* f.

Доказательство. Необходимость. Пусть $\widehat{x}$ — локальный минимум f. Обозначим $A = \operatorname{co}\{\, f_1'(\widehat{x}), \ldots, f_m'(\widehat{x}) \,\}$. Легко видеть, что A — выпуклое замкнутое множество. Допустим, что $0 \notin A$. Тогда по второй теореме отделимости найдется такой вектор $\overline{x} \in \mathbb{R}^m$, что $\sup_{y \in A} y \cdot \overline{x} < 0$ или что тоже $\sup\{\, \sum_{i=1}^m \alpha_i f_i'(\widehat{x}) \cdot \overline{x} \mid \alpha_i \ge 0,\ 1 \le i \le m,\ \sum_{i=1}^m \alpha_i = 1 \,\} = \max_{1 \le i \le m}\{\, f_i'(\widehat{x}) \cdot \overline{x} \,\} < 0$. Отсюда следует, что для каждого $1 \le i \le m$ и всех достаточно малых $t > 0$ выполняется неравенство $f_i(\widehat{x} + t\overline{x}) = f_i(\widehat{x}) + t(f_i'(\widehat{x}) \cdot \overline{x} + o(t)/t) < f_i(\widehat{x})$, из которого вытекает, что $f(\widehat{x} + t\overline{x}) < f(\widehat{x})$ в противоречие с тем, что $\widehat{x}$ — локальный минимум f.

Достаточность. Пусть $0 \in A$. Тогда $\sup_{y \in A} y \cdot x \ge 0$ для всех $x \in \mathbb{R}^m$, что, как показано выше, равносильно неравенству $\max_{1 \le i \le m}\{\, f_i'(\widehat{x}) \cdot x \,\} \ge 0$. Для каждого $1 \le i \le m$ по условию существует такой элемент $x_i^* \in (\mathbb{R}^n)^*$, что $f_i(x) - f_i(\widehat{x}) \ge x_i^* \cdot (x - \widehat{x})$ для x близких к $\widehat{x}$. Так как функция f_i, дифференцируема в $\widehat{x}$, то $x_i^* = f_i'(\widehat{x})$ (см. доказательство предложения). Тогда, для x достаточно близких к $\widehat{x}$, учитывая, что $f_1(\widehat{x}) = \ldots = f_m(\widehat{x})$, будем иметь $f(x) - f(\widehat{x}) = \max_{1 \le i \le m}\{\, f_i(x) - f_i(\widehat{x}) \,\} \ge \max_{1 \le i \le m}\{\, f_i'(\widehat{x}) \cdot (x - \widehat{x}) \,\} \ge 0$, т. е. $\widehat{x}$ — локальный минимум f. $\square$

Исторический комментарий. Основы теории специального класса выпуклых задач — задач линейного программирования — были заложены советским математиком Л. В. Канторовичем в его работе 1939 г. и впоследствии были переоткрыты американскими математиками. Важную роль в формировании этого направления сыграл Дж. фон Нейман, который в 1935 г. совместно экономистом О. Маргенштерном написал книгу “Теория игр и экономическое развитие”, положившую начало интенсивным исследованиям в области математической экономики. В середине сороковых годов прошлого века в США была осознана роль выпуклых задач в вопросах, которые ставились “военно-промышленным комплексом”, в частности, проблемами управления войсками. Это стало, с одной стороны, стимулом для развития вычислительных средств, а с другой, привлекло к математической теории экстремума многих исследователей — Дж. Б. Данцига, Г. Куна, А. У. Таккера и др. Их исследования несколько позже были объединены общей теорией выпуклости, включающей в себя теорию выпуклых множеств (основы которой заложены Г. Минковским) и выпуклых функций (основы которой разработаны В. Фенхелем).

7. Задачи 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) и в задаче Лагранжа, в качестве следствий из которой выводятся необходимые условия экстремума в задаче со старшими производными и в изопериметрической задаче.

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

$$J(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot x(t))\, dt \to \mathrm{extr}, \quad x(t_0) = x_0, \quad x(t_1) = x_1 \qquad (P_5)$$

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_5), если $\Gamma(x(\cdot)) = \{(t, x(t), \dot x(t)) \mid t \in [t_0, t_1]\} \subset G$ и $x(t_i) = x_i$, $i = 0, 1$.

Допустимая функция $\widehat{x}(\cdot)$ называется *слабым локальным минимумом (максимумом)* в задаче (P_5), 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 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_5))**.** *Пусть $\widehat{x}(\cdot)$ доставляет слабый локальный экстремум в задаче (P_5). Тогда, если функция L непрерывна вместе со своими частными производными по x и $\dot{x}$ в окрестности $\Gamma(\widehat{x}(\cdot))$, 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 $\alpha \in \mathbb{R}$. Для достаточно малых α функция $x_\alpha(\cdot)$ принадлежит окрестности $\Gamma(\widehat{x}(\cdot))$ и, очевидно, что $x_\alpha(\cdot) \in C^1([t_0, t_1])$ и $x_\alpha(t_i) = x_i$, $i = 0, 1$, т. е. для данных α функции $x_\alpha(\cdot)$ допустимы в задаче (P_5). 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:

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

Пусть $p(\cdot)$ такое, что $\dot{p}(\cdot) = \widehat{L}_x(\cdot)$. Подставляя это в (i) и интегрируя по частям, приходим к соотношению

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

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

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

Выберем $c = c_0$ так, чтобы $\int_{t_0}^{t_1}(-p(t)+\widehat{L}_{\dot{x}}(t)+c_0)\, dt = 0$. 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

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

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. $\square$

7.2. **Задача Больца.** Пусть, как и в предыдущем случае, $[t_0,t_1]$ — отрезок числовой прямой, G — открытое подмножество $\mathbb{R}^3$, $L\colon G\to\mathbb{R}$ — непрерывная функция трех переменных: t, x и $\dot x$, W — открытое подмножество $\mathbb{R}^2$ и $l\colon W\to\mathbb{R}$ — непрерывная функция переменных ξ_0 и ξ_1. Задача

$$\mathcal{B}(x(\cdot))=\int_{t_0}^{t_1}L(t,x(t),\dot x(t))\,dt+l(x(t_0),x(t_1))\to\mathrm{extr}\qquad (P_5')$$

называется *задачей Больца.*

Функция $x(\cdot)\in C^1([t_0,t_1])$ называется *допустимой в задаче* (P_5'), если $\Gamma(x(\cdot))=\{(t,x(t),\dot x(t))\mid t\in[t_0,t_1]\}\subset G$ и $(x(t_0),x(t_1))\in W$.

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_5'))**.** *Пусть $\widehat{x}(\cdot)$ доставляет слабый локальный экстремум в задаче (P_5'). Тогда, если функция L непрерывна вместе со своими частными производными по x и $\dot x$ в окрестности $\Gamma(\widehat{x}(\cdot))$, а функция l непрерывна вместе со своими частными производными по ξ_0 и ξ_1 в окрестности точки $(\widehat{x}(t_0),\widehat{x}(t_1))$, 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

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

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 $p(\cdot)$. $\square$

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,

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

7.3. **Интегралы уравнения Эйлера.** Напомним, что первым интегралом дифференциального уравнения называется функция, которая постоянна на решениях данного уравнения.

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

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

который называется *интегралом импульса.*

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

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

Он называется *интегралом энергии.*

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

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

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

Замечание. Эйлер доказывал необходимые условия экстремума в задачах вариационного исчисления (по сути своей бесконечномерных), переходя к их конечномерным аппроксимациям (так называемый метод ломаных). В 1759 г. он получил письмо от совсем молодого математика из Турина — Ж.-Л. Лагранжа, который выводил необходимые условия экстремума, вводя семейство функций типа $\widehat{x}(\cdot) + \alpha x(\cdot)$, как в доказательстве выше. Эти функции были названы "вариациями", а сам класс экстремальных задач, где такой подход применим, Эйлер назвал "вариационным исчислением". Своему молодому коллеге Эйлер ответил так:"Твое аналитическое решение изопериметрической проблемы содержит, насколько я вижу, все, что только можно пожелать в этой области, и я чрезвычайно рад, что эта теория, с которой после моих первых попыток, я занимался едва ли не один, доведена до величайшего совершенства. Важность вопроса побудила меня к тому, что я с помощью твоего освещения сам вывел аналитическое решение. Я, однако, решил скрывать это, пока ты не опубликуешь свои результаты, так как я никоим образом не хочу отнимать заслуженной тобой славы".

Исторический комментарий. Уравнение Эйлера для простейшей задачи вариационного исчисления, как уже было сказано, было получено Эйлером в его основополагающем труде Methodus invenitn, роль которого в развитии теории экстремума огромна. О. Больца (1857–1942) учился в Геттингене и Берлине (где в 1879 г. слушал лекции Вейерштрасса). В 1908 г. он издал учебник по вариационному исчислению, в котором изложил теорию Вейерштрасса. Этот учебник оказал большое влияние на преподавание вариационного исчисление во всем мире и, в частности, в нашей стране.

7.4. **Задача Лагранжа. Общая постановка.** Пусть $[t_0, t_1]$ — конечный отрезок, G — открытое подмножество $\mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^r$, W — открытое подмножество $\mathbb{R}^n\times\mathbb{R}^n$, функции $L_i\colon G\to\mathbb{R}$, $i = 0, 1, \ldots, m$, и отображение $\varphi\colon G \to \mathbb{R}^n$ (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 функции $l_i\colon W \to \mathbb{R}$, $i = 0, 1, \ldots, m$ (переменных ξ_0 и ξ_1) непрерывны на своей области определения. Задача

$$f_0(x(\cdot), u(\cdot)) = \int_{t_0}^{t_1} L_0(t, x(t), u(t))\,dt + l_0(x(t_0), x(t_1)) \to \min,$$

$$\dot{x} = \varphi(t, x, u),$$

$$f_i(x(\cdot), u(\cdot)) = \int_{t_0}^{t_1} L_i(t, x(t), u(t))\,dt + l_i(x(t_0), x(t_1)) \le 0, \quad 1 \le i \le m',$$

$$f_i(x(\cdot), u(\cdot)) = \int_{t_0}^{t_1} L_i(t, x(t), u(t))\,dt + l_i(x(t_0), x(t_1)) = 0,$$

$$m' + 1 \le i \le m \quad (P_6)$$

называется *задачей Лагранжа вариационного исчисления* (в понтрягинской форме).

Уточним постановку. Пара $(x(\cdot), u(\cdot))$ называется *допустимой в задаче* (P_6), если $(x(\cdot), u(\cdot)) \in Z = C^1([t_0, t_1], \mathbb{R}^n) \times C([t_0, t_1], \mathbb{R}^r)$,[3] $\Gamma(x(\cdot), u(\cdot)) = \{\, (t, x(t), u(t)) \mid t \in [t_0, t_1] \,\} \subset G$, $(x(t_0), x(t_1)) \in W$, $\dot{x}(t) = \varphi(t, x(t), u(t))$ для всех $t \in [t_0, t_1]$, $f_i(x(\cdot), u(\cdot)) \le 0$, $1 \le i \le m'$ и $f_i(x(\cdot), u(\cdot)) = 0$, $m' + 1 \le i \le m$.

Допустимая пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ называется *слабым локальным минимумом в задаче* (P_6), 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 $f_0(x(\cdot), u(\cdot)) \ge f_0(\widehat{x}(\cdot), \widehat{u}(\cdot))$.

Функция Лагранжа задачи (P_6) имеет вид:

$$\mathcal{L}(x(\cdot), u(\cdot), \overline{\lambda}) = \int_{t_0}^{t_1} L(t, x(t), u(t))\, dt + l(x(t_0), x(t_1)),$$

где $L(t, x, u) = \sum_{i=0}^m \lambda_i L_i(t, x, u) + p(t) \cdot (\dot{x} - \varphi(t, x, u))$, $l(\xi_0, \xi_1) = \sum_{i=0}^m \lambda_i l_i(\xi_0, \xi_1)$ и вектор $\overline{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_m, p(\cdot)) \in (\mathbb{R}^{m+1})^* \times C^1([t_0, t_1], (\mathbb{R}^n)^*)$ — набор множителей Лагранжа.

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 φ, l и т. д.

Теорема (Необходимые условия минимума в задаче (P_6))**.** *Пусть* $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ *доставляет слабый локальный минимум в задаче* (P_6). *Тогда, если функции* L_i, $0 \le i \le m$, *и отображение* φ *непрерывны вместе со своими частными производными по* x *и* u *в окрестности множества* $\Gamma(\widehat{x}(\cdot), \widehat{u}(\cdot))$, *а функции* l_i, $0 \le i \le m$, *непрерывно дифференцируемы в окрестности точки* $(\widehat{x}(t_0), \widehat{x}(t_1))$, *то найдется ненулевой набор множителей Лагранжа* $\overline{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_m, p(\cdot)) \in (\mathbb{R}^{m+1})^* \times C^1([t_0, t_1], (\mathbb{R}^n)^*)$ *такой, что выполняются*

(a) *условия стационарности (уравнения Эйлера–Лагранжа):*

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0 \quad \Leftrightarrow \quad -\dot{p} = p\widehat{\varphi}_x(t) - \sum_{i=0}^m \lambda_i \widehat{L}_{ix}(t),$$

$$\widehat{L}_u(t) = 0 \quad \Leftrightarrow \quad p(t)\widehat{\varphi}_u(t) = \sum_{i=0}^m \lambda_i \widehat{L}_{iu}(t);$$

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

(c) $\lambda_i f_i(\widehat{x}(\cdot), \widehat{u}(\cdot)) = 0$, $i = 1, \dots, m'$ *(условия дополняющей нежесткости);*

(d) $\widehat{L}_{\dot{x}}(t_i) = (-1)^i \widehat{l}_{\xi_i}$, $i = 0, 1$ *(условия трансверсальности).*

---

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

7.5. **Задача Лагранжа. Упрощенный вариант.** Необходимые условия минимума докажем для следующего частного случая общей постановки:

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

$$x(t_0) = x_0, \quad x(t_1) = x_1. \quad (P_6')$$

При этом следует заметить, что доказательство в общем случае в существенном не отличается от доказательства необходимых условий для задачи (P_6').

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_6').

Функция Лагранжа задачи (P_6') имеет вид

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

где $\overline{\lambda} = (\lambda_0, p(\cdot))$ — *набор множителей Лагранжа*. Функцию под знаком интеграла обозначим через $L = L(t,x,\dot{x},u,\lambda_0,p) = \lambda_0 f(t,x,u) + p\cdot(\dot{x} - \varphi(t,x,u))$. Мы докажем следующий результат.

Теорема (Необходимые условия минимума в задаче (P_6'))**.** *Пусть пара $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ доставляет слабый локальный минимум в задаче (P_6'). Тогда, если функция f и отображение φ непрерывны вместе со своими частными производными по x и u в окрестности $\Gamma(\widehat{x}(\cdot),\widehat{u}(\cdot))$, то существует такой ненулевой набор множителей Лагранжа $\overline{\lambda} = (\lambda_0,p(\cdot)) \in \mathbb{R}\times C^1([t_0,t_1],(\mathbb{R}^n)^*)$, 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.

Теорема (о зависимости решения дифференциального уравнения от параметров)**.** *Пусть G — открытое подмножество $\mathbb{R}\times\mathbb{R}^n$, W*

— открытое подмножество $\mathbb{R}^k$*, функция* $f\colon G\times W\to\mathbb{R}^n$ *(переменных* $t\in\mathbb{R}$*,* $x\in\mathbb{R}^n$ *и* $\alpha\in\mathbb{R}^k$*) и ее частные производные по* x *и* α *непрерывны на* $G\times W$*,* $\widehat{\alpha}\in W$ *и* $x_0\in\mathbb{R}^n$*. Пусть далее функция* $\widehat{x}(\cdot)\in C^1([t_0,t_1],\mathbb{R}^n)$*, для которой* $\Gamma(\widehat{x}(\cdot))=\{\,(t,\widehat{x}(t))\mid t\in[t_0,t_1]\,\}\subset G$*, является решением задачи Коши*

$$\dot{x}=f(t,x,\widehat{\alpha}),\quad x(t_0)=x_0.$$

Тогда существует окрестность $V\subset W$ *точки* $\widehat{\alpha}$ *такая, что для каждого* $\alpha\in V$ *определено на отрезке* $[t_0,t_1]$ *единственное решение* $x(\cdot,\alpha)$ *задачи Коши*

$$\dot{x}=f(t,x,\alpha),\quad x(t_0)=x_0$$

и при этом, $x(\cdot,\alpha)\to\widehat{x}(\cdot)$ *в метрике* $C^1([t_0,t_1],\mathbb{R}^n)$ *при* $\alpha\to\widehat{\alpha}$*; для каждого* $t\in[t_0,t_1]$ *отображение* $\alpha\to x(t,\alpha)$ *из* V *в* $\mathbb{R}^n$ *непрерывно дифференцируемо и его производная в точке* $\widehat{\alpha}$ *на* $[t_0,t_1]$ *удовлетворяет дифференциальному уравнению*

$$\dot{y}=f_x(t,\widehat{x}(t),\widehat{\alpha})y+f_\alpha(t,\widehat{x}(t),\widehat{\alpha}),\quad y(t_0)=0.$$

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

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

Если $t\in[t_0,t_1]$, то производную отображения $\alpha\mapsto x_\alpha(t,u(\cdot))$ в нуле обозначим $y(t,u(\cdot))$. Она удовлетворяет, согласно теореме, дифференциальному уравнению

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

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

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

Покажем, что если $M=\operatorname{span}\{\,(y_0(u(\cdot)),y(t_1,u(\cdot)))\in\mathbb{R}^{n+1}\mid u(\cdot)\in C([t_0,t_1],\mathbb{R}^r)\,\}$ (линейная оболочка множества в фигурных скобках) есть все пространство $\mathbb{R}^{n+1}$, то это приводит к противоречию с тем, что $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ — локальный минимум. Действительно, если это так, то существуют такие $u_i(\cdot)\in C([t_0,t_1],\mathbb{R}^r)$, $i=1,\ldots,n+1$, что векторы $(y_0(u_i(\cdot)),y(t_1,u_i(\cdot)))$, $i=1,\ldots,n+1$, 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 $u_{\overline{\alpha}}(\cdot,\overline{u}(\cdot)) = \widehat{u}(\cdot) + \sum_{i=1}^{n+1} \alpha_i u_i(\cdot)$. В силу теоремы о дифференцируемой зависимости для достаточно малых $\overline{\alpha}$ существует решение $x_{\overline{\alpha}}(\cdot,\overline{u}(\cdot))$ задачи Коши

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

и, снова, в силу стандартных фактов, для малых $\overline{\alpha}$ функция $\overline{\alpha} \mapsto J(x_{\overline{\alpha}}(\cdot,\overline{u}(\cdot)), u_{\overline{\alpha}}(\cdot,\overline{u}(\cdot)))$ определена и непрерывно дифференцируема. Таким образом, существует окрестность U нуля в $\mathbb{R}^{n+1}$, что корректно определено отображение $F\colon U \to \mathbb{R}^{n+1}$, сопоставляющее $\overline{\alpha} \in U$ вектор $(J(x_{\overline{\alpha}}(\cdot,\overline{u}(\cdot)), u_{\overline{\alpha}}(\cdot,\overline{u}(\cdot))), x_{\overline{\alpha}}(t_1,\overline{u}(\cdot)))$. Это отображение непрерывно дифференцируемо на U 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 применяя обычную теорему об обратной функции или теорему о разрешимости системы нелинейных уравнений (условия, которой выполнены, поскольку отображение F непрерывно дифференцируемо на U и поэтому строго дифференцируемо в нуле) придем (рассуждая также как при доказательстве правила множителей Лагранжа) к противоречию с тем, что $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — локальный минимум. Действительно, $F(0) = (J(\widehat{x}(\cdot), \widehat{u}(\cdot)), x_1)$. Для достаточно малых $\varepsilon > 0$ вектор $y_\varepsilon = (J(\widehat{x}(\cdot), \widehat{u}(\cdot)) - \varepsilon, x_1)$ принадлежит соответствующей окрестности $F(0)$ (окрестность W в теореме о разрешимости). Тогда существует вектор $\overline{\alpha}_\varepsilon = \varphi(0, y_\varepsilon) \in U$ такой, что $F(\overline{\alpha}_\varepsilon) = y_\varepsilon$, или $J(x_{\overline{\alpha}_\varepsilon}(\cdot,\overline{u}(\cdot)), u_{\overline{\alpha}_\varepsilon}(\cdot,\overline{u}(\cdot))) = J(\widehat{x}(\cdot), \widehat{u}(\cdot)) - \varepsilon$, $x_{\overline{\alpha}_\varepsilon}(t_1,\overline{u}(\cdot)) = x_1$. Так как φ — непрерывное отображение, то $(x_{\overline{\alpha}_\varepsilon}(\cdot,\overline{u}(\cdot)), u_{\overline{\alpha}_\varepsilon}(\cdot,\overline{u}(\cdot))) \to (\widehat{x}(\cdot), \widehat{u}(\cdot))$ при $\varepsilon \to 0$ и поэтому в любой окрестности пары $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ найдется допустимая в задаче (P_6') точка, в которой значение функционала меньше, чем в $(\widehat{x}(\cdot), \widehat{u}(\cdot))$, в противоречие с тем, что $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — локальный минимум.

Таким образом, M — собственное подпространство в $\mathbb{R}^{n+1}$. Пусть $y_0 \notin M$. Тогда по второй теореме отделимости найдется ненулевой линейный функционал (вектор) $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_n) \in (\mathbb{R}^{n+1})^*$ такой, что $\overline{\lambda} \cdot y_0 < \inf_{y \in M} \overline{\lambda} \cdot y$. Отсюда следует, что $\overline{\lambda} \cdot y = 0$ для всех $y \in M$ (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))

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

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

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

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

В качестве следствия доказанной теоремы получим необходимые условия экстремума в задаче со старшими производными и изопериметрической задаче.

7.6. **Задача со старшими производными. Уравнение Эйлера–Пуассона.** Пусть $[t_0,t_1]$ — отрезок числовой прямой, G — открытое подмножество $\mathbb{R}^{n+2}$, $L\colon G\to\mathbb{R}$ — непрерывная функция (переменные которой обозначаем $t,x,\dot x,\dots,x^{(n)}$) и $x_i^{(k)}\in\mathbb{R}$, $k=0,1,\dots,n-1$, $i=0,1$. Задача

$$J(x(\cdot))=\int_{t_0}^{t_1}L(t,x(t),\dot x(t),\dots,x^{(n)}(t))\,dt\to\min,$$
$$x^{(k)}(t_i)=x_i^{(k)},\quad k=0,1,\dots,n-1,\quad i=0,1,\quad (P_6'')$$

называется *задачей со старшими производными.*

Обозначим через $C^n([t_0,t_1])$ пространство всех n раз непрерывно дифференцируемых функций $x(\cdot)$ на $[t_0,t_1]$ с нормой $\|x(\cdot)\|_{C^n([t_0,t_1])}=\max(\|x(\cdot)\|_{C([t_0,t_1])},\|\dot x(\cdot)\|_{C([t_0,t_1])},\dots,\|x^{(n)}(\cdot)\|_{C([t_0,t_1])})$.

Функцию $x(\cdot)\in C^n([t_0,t_1])$ называется *допустимой в задаче* (P_6''), если $\Gamma(x(\cdot))=\{(t,x(t),\dot x(t),\dots,x^{(n)}(t))\mid t\in[t_0,t_1]\}\subset G$ и $x^{(k)}(t_i)=x_i^{(k)}$, $k=0,1,\dots,n-1$, $i=0,1$.

Допустимая функция $\widehat{x}(\cdot)$ называется *локальным минимумом в задаче* (P_6'')*, если существует такое* $\varepsilon>0$*, что для любой допустимой функции* $x(\cdot)$*, для которой* $\|x(\cdot)-\widehat{x}(\cdot)\|_{C^n([t_0,t_1])}<\varepsilon$ *выполняется неравенство* $J(x(\cdot))\ge J(\widehat{x}(\cdot))$*.*

Следствие (Необходимые условия минимума в задаче (P_6''). Уравнение Эйлера–Пуассона)**.** *Пусть* $\widehat{x}(\cdot)$ *— допустимая функция в задаче* (P_6'') *и функция* L *такова, что* $L_{x^{(k)}}(\cdot)\in C^k([t_0,t_1])$, $k=$

$0, 1, \ldots, n$, в окрестности $\Gamma(\widehat{x}(\cdot))$. Тогда, если $\widehat{x}(\cdot)$ доставляет локальный минимум в (P_6''), то для всех $t \in [t_0, t_1]$ выполнено уравнение Эйлера–Пуассона

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

Доказательство. Обозначая $x = x_1,\ \dot{x}_1 = x_2, \ldots, \dot{x}_{n-1} = x_n,\ \dot{x}_n = u$, задачу (P_6'') можно записать как задачу Лагранжа

$$\int_{t_0}^{t_1} L(t, x_1(t), \ldots, x_n(t), u(t))\, dt \to \min, \quad \dot{x}_1 = x_2, \ldots, \dot{x}_{n-1} = x_n,$$
$$\dot{x}_n = u,\ \ x_k(t_i) = x_i^{(k-1)},\ \ k = 1, \ldots, n,\ \ i = 0, 1. \quad (i)$$

Простая проверка показывает, что если $\widehat{x}(\cdot)$ — локальный минимум в задаче (P_6''), то вектор-функция $(\widehat{x}_1(\cdot), \ldots, \widehat{x}_n(\cdot), \widehat{u}(\cdot))^T$ — локальный минимум в данной задаче. Согласно общей теореме найдутся такие множители Лагранжа λ_0 и $p(\cdot) = (p_1(\cdot), \ldots, p_n(\cdot)) \in C^1([t_0, t_1], (\mathbb{R}^n)^*)$, не равные одновременно нулю, что для функции Лагранжа

$$\mathcal{L} = \int_{t_0}^{t_1} \Bigg(\lambda_0 L(t, x_1(t), \ldots, x_n(t), u(t)) + \sum_{k=1}^{n-1} p_k(t)(\dot{x}_k(t) - x_{k+1}(t))$$
$$+ p_n(t)(\dot{x}_n(t) - u(t))\Bigg)\, dt$$

задачи (i) выполнены соотношения: $-\dot{p}_1(t) + \lambda_0 \widehat{L}_{x_1}(t) = 0,\ -\dot{p}_2(t) + \lambda_0 \widehat{L}_{x_2}(t) - p_1(t) = 0, \ldots,\ -\dot{p}_n(t) + \lambda_0 \widehat{L}_{x_n}(t) - p_{n-1}(t) = 0$ и $\lambda_0 \widehat{L}_u(t) - p_n(t) = 0$.

Если $\lambda_0 = 0$, то сразу видно, что и $p(\cdot) = 0$. Считая $\lambda_0 = 1$, из выписанных соотношений выводим (учитывая условия теоремы и переходя к прежним обозначениям: $x_1 = x, x_2 = \dot{x}, \ldots, x_n = x^{(n-1)}, u = x^{(n)}$), что

$$\widehat{L}_x(t) = \dot{p}_1(t) = -\ddot{p}_2(t) + \frac{d}{dt}\widehat{L}_{\dot{x}}(t) = \ldots = (-1)^{n-1} p_n^{(n)}(t)$$
$$+ \sum_{k=1}^{n-1}(-1)^{k-1}\frac{d^k}{dt^k}\widehat{L}_{x^{(k)}}(t) = (-1)^{n-1}\frac{d^n}{dt^n}\widehat{L}_{x^{(n)}}(t)$$
$$+ \sum_{k=1}^{n-1}(-1)^{k-1}\frac{d^k}{dt^k}\widehat{L}_{x^{(k)}}(t).$$

Это, очевидно, равносильно уравнению Эйлера–Пуассона. □

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

$$J(x(\cdot))=\int_{t_0}^{t_1} f_0(t,x(t),\dot x(t))\,dt\to\min,\quad \int_{t_0}^{t_1} f_i(t,x(t),\dot x(t))\,dt=\alpha_i,$$
$$1\le i\le m,\quad x(t_0)=x_0,\quad x(t_1)=x_1 \quad (P_6''')$$

называется *изопериметрической задачей*. Ясно, что это частный случай задачи (P_6') (дифференциальная связь: $\dot x=u$).

Функция $x(\cdot)$ называется *допустимой в задаче* (P_6'''), если $x(\cdot)\in C^1([t_0,t_1],\mathbb{R}^n)$, $\Gamma(x(\cdot))=\{(t,x(t),\dot x(t))\mid t\in[t_0,t_1]\}\subset G$, $\int_{t_0}^{t_1} f_i(t,x(t),\dot x(t))\,dt=\alpha_i$, $1\le i\le m$, и $x(t_i)=x_i$, $i=0,1$.

Допустимая функция $\widehat x(\cdot)$ называется *локальным минимумом* в задаче (P_6'''), если существует такое $\varepsilon>0$, что для любой допустимой функции $x(\cdot)$, для которой $\|x(\cdot)-\widehat x(\cdot)\|_{C^1([t_0,t_1],\mathbb{R}^n)}<\varepsilon$ выполняется неравенство $J(x(\cdot))\ge J(\widehat x(\cdot))$.

Функция Лагранжа задачи (P_6''') (в соответствии с функцией Лагранжа задачи (P_6')) имеет вид

$$\mathcal{L}(x(\cdot),\overline\lambda)=\lambda_0\int_{t_0}^{t_1} f_0(t,x(t),\dot x(t))\,dt + \sum_{i=1}^m\lambda_i\Big(\int_{t_0}^{t_1} f_i(t,x(t),\dot x(t))\,dt-\alpha_i\Big),$$

где $\overline\lambda=(\lambda_0,\lambda_1,\ldots,\lambda_m)$.

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

Следствие (Необходимые условия минимума в задаче (P_6'''))**.** *Пусть $\widehat x(\cdot)$ доставляет слабый локальный минимум в задаче (P_6'''). Тогда, если функции f_i, $i=0,1,\ldots,m$, непрерывны вместе со своими частными производными по x и $\dot x$ в окрестности $\Gamma(\widehat x(\cdot))$, то найдется такой ненулевой набор множителей Лагранжа $\overline\lambda=(\lambda_0,\lambda_1,\ldots,\lambda_m)$, что $\widehat L_{\dot x}(\cdot)\in C^1([t_0,t_1],\mathbb{R}^n)$ и для всех $t\in[t_0,t_1]$ выполнено уравнение Эйлера*

$$-\frac{d}{dt}\widehat L_{\dot x}(t)+\widehat L_x(t)=0.$$

Доказательство. Рассмотрим следующую задачу Лагранжа

$$\int_{t_0}^{t_1} f_0(t,x(t),u(t))\,dt\to\min,\quad \dot x=u,\quad \dot y_1=f_1(t,x,u),\ldots,$$
$$\dot y_m=f_m(t,x,u),\quad x(t_0)=x_0,\quad x(t_1)=x_1,\quad y_i(t_0)=0,$$
$$y_i(t_1)=\alpha_i,\quad i=1,\ldots,m.\quad (i)$$

Несложная проверка показывает, что если $\widehat{x}(\cdot)$ — локальный экстремум в задаче (P_6'''), то $(\widehat{x}(\cdot), \widehat{y}(\cdot), \widehat{u}(\cdot))$, где $\widehat{u}(\cdot) = \dot{\widehat{x}}(\cdot)$ и $\widehat{y}(\cdot) = (\widehat{y}_1(\cdot), \ldots, \widehat{y}_m(\cdot))$ однозначно определяется по $\widehat{x}(\cdot)$ и $\widehat{u}(\cdot)$, — локальный экстремум в данной задаче.

Согласно теореме о необходимых условиях минимума в задаче (P_6') найдутся такие множители Лагранжа λ_0 и $p(\cdot) = (p_1(\cdot), \ldots, p_{n+m}(\cdot)) \in C^1([t_0, t_1], (\mathbb{R}^{n+m})^*)$, не все равные нулю, что для функции Лагранжа

$$\int_{t_0}^{t_1} \left(\lambda_0 f_0(t, x(t), u(t)) + \sum_{i=1}^{n} p_i(\dot{x}_i - u_i) + \sum_{j=1}^{m} p_{n+j}(\dot{y}_j - f_j(t, x, u)) \right) dt$$

выполнены соотношения: $-\dot{p}_i(t) + \lambda_0 \widehat{f}_{0x_i}(t) - \sum_{j=1}^m p_{n+j} \widehat{f}_{jx_i}(t) = 0$, $i = 1, \ldots, n$; $\dot{p}_{n+1}(t) = \ldots = \dot{p}_{n+m}(t) = 0$ и $\lambda_0 \widehat{f}_{0u_i}(t) - p_i(t) - \sum_{j=1}^m p_{n+j} \widehat{f}_{ju_i}(t) = 0$, $i = 1, \ldots, n$. Из второго соотношения следует, что $p_{n+j}(\cdot) = \mathrm{const}$, $j = 1, \ldots, m$. Положим $\lambda_j = -p_{n+j}$, $j = 1, \ldots, m$. Из последнего соотношения вытекает (заменяя u_i на $\dot{x}_i$, $i = 1, \ldots, n$), что $\sum_{j=0}^m \lambda_j \widehat{f}_{j\dot{x}_i}(\cdot) \in C^1([t_0, t_1])$, $i = 1, \ldots, n$, а это равносильно тому, что $\widehat{L}_{\dot{x}}(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n)$. Дифференцируя последнее соотношение по t и сравнивая его с первым, получаем уравнение Эйлера. Заметим еще, что не все множители $\lambda_0, \lambda_1, \ldots, \lambda_m$ нулевые, так как из выписанных соотношений следовало бы, что и $p_i(\cdot) = 0$, $i = 1, \ldots, n$. $\square$

Исторический комментарий. Как уже говорилось, вариационное исчисление зародилось в 18 веке в трудах Эйлера и Лагранжа. Эйлер доказал необходимые условия экстремума и в изопериметрической задачи, и в задаче со старшими производными. Лагранж изучал более общую задачу...

8. Задачи оптимального управления

Вариационное исчисление, как уже говорилось, интенсивно развивалось в 18 веке (в основном усилиями Эйлера, Лагранжа и Лежандра). В 19 веке в его развитии приняли участие такие математики как Пуассон, Вейерштрасс, Гильберт и Пуанкаре. К началу 20 века предмет, в существенном, оказался исчерпанным, хотя в одной, несколько изолированной, “чикагской школе” велась достаточно плодотворная работа по развитию вариационного исчисления. Итоги этой деятельности были подведены Блиссом в его монографии 1942 г. Построение теории экстремума (как она представлялась в те времена), казалось, завершено. Но впоследствии появились выпуклые задачи, а затем, в начале 50-х годов прошлого века родилось оптимальное управление — новое направление в теории экстремума, охватывающее вариационное исчисление.

Необходимые условия экстремума в задачах оптимального управления были получены в школе Л. С. Понтрягина. Основной результат называется принципом максимума Понтрягина. В этом разделе рассматривается задача оптимального управления и доказываются для нее необходимые условия минимума.

Пусть $[t_0, t_1]$ — конечный отрезок, U — непустое подмножество $\mathbb{R}^r$, G — открытое подмножество $\mathbb{R} \times \mathbb{R}^n$, функция $f\colon G \times U \to \mathbb{R}$ и отображение $\varphi\colon G \times U \to \mathbb{R}^n$ (переменных $t \in \mathbb{R}$, $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ и $u = (u_1, \dots, u_r)^T \in \mathbb{R}^r$) непрерывны на $G \times U$ и $x_0 \in \mathbb{R}^n$. Задача

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

называется *задачей оптимального управления*. Переменную $x(\cdot)$ часто называют фазовой переменной, а $u(\cdot)$ — управлением.

Уточним постановку. Пусть $PC^1([t_0, t_1], \mathbb{R}^n)$ — совокупность всех кусочно-непрерывно дифференцируемых, а $PC([t_0, t_1], \mathbb{R}^r)$ — кусочно-непрерывных функций на $[t_0, t_1]$ со значениями соответственно в $\mathbb{R}^n$ и $\mathbb{R}^r$. 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_7), если $\Gamma(x(\cdot)) = \{\, (t, x(t)) \mid t \in [t_0, t_1] \,\} \subset G$, 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 $x(t_0) = x_0$, $x(t_1) = x_1$.

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

Функция Лагранжа для задачи (P_7) имеет тот же вид, что и для задачи (P_6').

Функцию $H(t, x, u, \lambda_0, p) = p \cdot \varphi(t, x, u) - \lambda_0 f(t, x, u)$ называют *функцией Понтрягина задачи* (P_7).

Ниже пользуемся теми же соглашениями об обозначениях, о которых сказано перед формулировкой теоремы о необходимых условиях экстремума в задаче (P_6).

Теорема (Необходимые условия минимума в задаче (P_7). Принцип максимума Понтрягина)**.** *Пусть $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ доставляет сильный минимум в задаче (P_7). Тогда, если функция f и отображение φ непрерывны вместе со своими частными производными по x в $G \times U$, то найдется ненулевой набор множителей Лагранжа $\overline{\lambda} = (\lambda_0, p(\cdot)) \in \mathbb{R} \times PC^1([t_0, t_1], (\mathbb{R}^n)^*)$ такой, что выполнено условие стационарности по $x(\cdot)$:*

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

и в точках непрерывности $\widehat{u}(\cdot)$ условие минимума по $u(\cdot)$:

$$\min_{u\in U} L(t,\widehat{x}(t),\dot{\widehat{x}}(t),u,\lambda_0,p(t)) = L(t,\widehat{x}(t),\dot{\widehat{x}}(t),\widehat{u}(t),\lambda_0,p(t)),$$

или (что то же) условие максимума по $u(\cdot)$:

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

Из-за последнего соотношения необходимые условия в задаче оптимального управления и называют “Принципом максимума Понтрягина”.

Сформулированная задача оптимального управления не самая общая, но достаточно представительная. Мы докажем ее более простой вариант, когда правый конец свободен, т. е. получим необходимые условия минимума в задаче

$$J(x(\cdot),u(\cdot)) = \int_{t_0}^{t_1} f(t,x(t),u(t))\,dt \to \min, \quad \dot{x} = \varphi(t,x,u),$$
$$u(t)\in U, \quad x(t_0)=x_0. \quad (P_7')$$

Эти условия минимума те же, что и в задаче (P_7), но добавляется еще условие трансверсальности: $p(t_1)=0$.

Перед доказательством сформулируем две леммы.

Лемма (об игольчатой вариации). *Пусть $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ — допустимая пара в задаче (P_7) (или (P_7')), $\tau\in(t_0,t_1)$ — точка непрерывности $\widehat{u}(\cdot)$, $\alpha>0$ столь мало, что функция $\widehat{u}(\cdot)$ непрерывна на $[\tau-\alpha,\tau]$ и $v\in U$. Положим*

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

Тогда найдется такое $\alpha_0>0$, что для любого $0\le\alpha<\alpha_0$ существует единственное решение $x_\alpha(\cdot;\tau,v)$ задачи Коши

$$\dot{x} = \varphi(t,x,u_\alpha(t;\tau,v)), \quad x(t_0)=x_0,$$

определенное на всем отрезке $[t_0,t_1]$. При этом, $x_\alpha(\cdot;\tau,v)\to\widehat{x}(\cdot)$ при $\alpha\to 0$ равномерно на $[t_0,t_1]$, производная отображения $\alpha\mapsto x_\alpha(t,\tau,v)$ при каждом $t\in[\tau,t_1]$ непрерывно дифференцируема для достаточно малых $\alpha\ge 0$ и в нуле на $[\tau,t_1]$ удовлетворяет уравнению

$$\dot{y} = \widehat{\varphi}_x(t)y, \quad y(\tau) = \varphi(\tau,\widehat{x}(\tau),v) - \varphi(\tau,\widehat{x}(\tau),\widehat{u}(\tau)). \qquad (i)$$

Эту лемму доказывать не будем.

Функцию $u_\alpha(\cdot;v)$ называют *игольчатой вариацией* $\widehat{u}(\cdot)$, а пару (τ,v) — *иголкой.*

Лемма (о производной функционала). *Пусть $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ — допустимая пара в задаче (P_7) (или (P_7')). Тогда функция $\alpha\mapsto J(\alpha)=$*

$J(x_\alpha(\cdot,\tau,v),u_\alpha(\cdot,\tau,v))$ для достаточно малых $\alpha \geq 0$ непрерывно дифференцируема и ее производная в нуле имеет вид

$$\Delta_{\tau v} f + \int_\tau^{t_1} \widehat{f}_x(t)\cdot y(t;\tau,v)\,dt, \qquad (ii)$$

где $\Delta_{\tau v} f = f(\tau,\widehat{x}(\tau),v) - f(\tau,\widehat{x}(\tau),\widehat{u}(\tau))$ и $y(\cdot;\tau,v)$ — решение задачи (i).

Доказательство. Докажем только формулу (ii). Имеем

$$J'(0+0) = \lim_{\alpha\downarrow 0}\frac{J(\alpha)-J(0)}{\alpha} = \lim_{\alpha\downarrow 0}\frac{1}{\alpha}\int_{\tau-\alpha}^{\tau}(f(t,x_\alpha(t;\tau,v),v) - f(t,\widehat{x}(t),\widehat{u}(t)))\,dt + \lim_{\alpha\downarrow 0}\frac{1}{\alpha}\int_\tau^{t_1}(f(t,x_\alpha(t;\tau,v),\widehat{u}(t)) - f(t,\widehat{x}(t),\widehat{u}(t)))\,dt.$$

Второе слагаемое справа — производная в нуле интеграла по параметру, к которому применима стандартная теорема о дифференцируемости под знаком интеграла, ибо функция $\alpha \mapsto x_\alpha(t;\tau,v)$ непрерывно дифференцируема для каждого $t\in[\tau,t_1]$, а отрезок $[\tau,t_1]$ можно разбить на конечное число интервалов, на каждом из которых функция $\widehat{u}(\cdot)$ непрерывна. К первому интегралу применим теорему о среднем для интегралов и тогда в итоге получим

$$J'(0+0) = \lim_{\alpha\downarrow 0}(f(\xi,x_\alpha(\xi;\tau,v),v) - f(\xi,\widehat{x}(\xi),\widehat{u}(\xi))) + \int_\tau^{t_1}\widehat{f}_x(t)\cdot y(t;\tau,v)\,dt,$$

где $\xi\in[\tau-\alpha,\tau]$. Когда $\alpha\to 0$, то, очевидно, $\xi\to\tau$, $x_\alpha(\xi;\tau,v)\to\widehat{x}(\tau)$ согласно лемме об игольчатой вариации, а $\widehat{u}(\xi)\to\widehat{u}(\tau)$, так как $\widehat{u}(\cdot)$ непрерывна в точке τ. Формула (ii) доказана. □

Доказательство принципа максимума. Пусть $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ — сильный минимум в задаче (P_7'). Обозначая через $p(\cdot)$ — решение линейного уравнения

$$-\dot{p} = p\widehat{\varphi}_x(t) - \widehat{f}_x(t), \quad p(t_1)=0, \qquad (iii)$$

получаем условие стационарности по $x(\cdot)$ и условие трансверсальности.

Так как пара $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ доставляет минимум, то необходимо $J'(0+0)\geq 0$, или согласно (ii)

$$\Delta_{\tau v} f + \int_\tau^{t_1}\widehat{f}_x(t)\cdot y(t,\tau,v)\,dt \geq 0.$$

Подставим сюда вместо функции $\widehat{f}_x(\cdot)$ ее выражение из (iii), а затем вместо функции $\widehat{\varphi}_x(\cdot)y(\cdot;\tau,v)$ ее выражение из (i) и учитывая, что

$p(t_1) = 0$, а $y(\tau;\tau,v) = \varphi(\tau,\widehat{x}(\tau),v) - \varphi(\tau,\widehat{x}(\tau),\widehat{u}(\tau))$, будем иметь

$$\begin{aligned}0 \le \Delta_{\tau v} f &+ \int_\tau^{t_1} (\dot p(t)\cdot y(t;\tau,v) + p(t)\cdot \dot y(t;\tau,v)\,dt \\ &= \Delta_{\tau v} f + p(t)\cdot y(t;\tau,v)|_\tau^{t_1} = f(\tau,\widehat{x}(\tau),v) - f(\tau,\widehat{x}(\tau),\widehat{u}(\tau)) \\ &\qquad - p(\tau)(\varphi(\tau,\widehat{x}(\tau),v) - \varphi(\tau,\widehat{x}(\tau),\widehat{u}(\tau))),\end{aligned}$$

или

$$\begin{aligned}&p(\tau)\varphi(\tau,\widehat{x}(\tau),v) - f(\tau,\widehat{x}(\tau),v) \\ &\qquad\le p(\tau)\varphi(\tau,\widehat{x}(\tau),\widehat{u}(\tau)) - f(\tau,\widehat{x}(\tau),\widehat{u}(\tau))\end{aligned}$$

Таким образом, для любой точки τ, где функция $\widehat{u}(\cdot)$ непрерывна, максимум выражения слева по всем $v \in U$ достигается в точке $\widehat{u}(\tau)$. Это и есть условие максимума из теоремы. □

9. **Необходимые условия слабого и сильного экстремума в простейшей задаче вариационного исчисления**

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

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

$$J(x(\cdot)) = \int_{t_0}^{t_1} L(t,x(t),\dot x(t))dt \to \min, \quad x(t_i) = x_i, \quad i = 0,1, \quad (P_5)$$

и, для определенности, рассматриваем задачу на минимум.

Здесь $[t_0,t_1]$ — отрезок числовой прямой, непрерывная функция L переменных $t \in \mathbb{R}$, $x \in \mathbb{R}$ и $\dot x \in \mathbb{R}$ определена на открытом подмножестве $G \subset \mathbb{R}\times\mathbb{R}\times\mathbb{R}$ и $x_i \in \mathbb{R}$, $i = 0,1$.

Напомним также, что функция $x(\cdot) \in C^1([t_0,t_1])$ называется *допустимой в задаче* (P_5), если $\Gamma(x(\cdot)) = \{(t,x(t),\dot x(t)) \mid t \in [t_0,t_1]\} \subset G$ и $x(t_i) = x_i$, $i = 0,1$, и допустимая функция $\widehat{x}(\cdot)$ называется *слабым локальным минимумом* в задаче (P_5), 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 $J(x(\cdot)) \ge J(\widehat{x}(\cdot))$.

В вариационном исчислении, наряду со слабым экстремумом, рассматривают еще и сильный локальный экстремум, где близость функций измеряется в пространстве $C([t_0,t_1])$. Точнее говоря, обозначим через $PC^1([t_0,t_1])$ пространство кусочно-непрерывно-дифференцируемых функций на $[t_0,t_1]$. Допустимая функция $x(\cdot) \in PC^1([t_0,t_1])$ в задаче (P_5) определяется аналогично предыдущему.

Скажем, что допустимая функция $\widehat{x}(\cdot)$ доставляет *сильный локальный минимум* в задаче (P_5), 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 $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$.

Заметим, что если функция $\widehat{x}(\cdot)$ доставляет сильный минимум в задаче (P_5) и при этом $\widehat{x}(\cdot) \in C^1([t_0,t_1])$, то $\widehat{x}(\cdot)$ является и слабым минимумом 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 $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$. 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 минимума для $x(\cdot) \in C^1([t_0,t_1])$ являются необходимыми условиями и сильного минимума.

Теперь сформулируем условия Лежандра, Якоби и Вейерштрасса. Пусть $\widehat{x}(\cdot)$ — *экстремаль* (или *стационарная точка*) задачи (P_5), т. е. функция, для которой выполнено уравнение Эйлера. Предположим, что интегрант L дважды непрерывно дифференцируем в некоторой окрестности $\Gamma(\widehat{x}(\cdot))$.

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

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

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

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

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

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

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

Уравнение Эйлера для данной задачи имеет вид

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

и называется *уравнением Якоби* исходной задачи (P_5).

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

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

Пусть $f\colon\mathbb{R}\to\mathbb{R}$ — 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 к графику функции f в точке x) в точке x'.

Если f — выпуклая функция, то $\mathcal{E}(x,x')\geq 0$ для всех $x,x'\in\mathbb{R}$. Действительно, пусть $x,x'\in\mathbb{R}$ и $0<\alpha<1$. По неравенству Иенссена $f((1-\alpha)x+\alpha x')\leq(1-\alpha)f(x)+\alpha f(x')$, откуда $\alpha^{-1}(f(x+\alpha(x'-x))-f(x))\leq f(x')-f(x)$. Переходя к пределу при $\alpha\to 0$, получаем, что $f(x')-f(x)\geq f'(x)(x'-x)$.

Пусть L — интегрант в задаче (P_5). Если L — дифференцируемая функция на некотором открытом множестве $G\times\mathbb{R}$, где G — открытое подмножество $\mathbb{R}\times\mathbb{R}$, то функция $\mathcal{E}(t,x,\dot{x},u)=L(t,x,u)-L(t,x,\dot{x})-L_{\dot{x}}(t,x,\dot{x})(u-\dot{x})$, определенная на $G\times\mathbb{R}$, 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 $\mathcal{E}(t,\widehat{x}(t),\dot{\widehat{x}}(t),u)\geq 0$ для всех $u\in\mathbb{R}$ и $t\in[t_0,t_1]$.

Теорема (Необходимые условия сильного минимума в задаче (P_5))**.** *Пусть функция* $\widehat{x}(\cdot)\in C^1([t_0,t_1])$ *доставляет сильный локальный минимум в задаче* (P_5). *Тогда, если интегрант* L *непрерывно дифференцируем в окрестности графика* $\Gamma(\widehat{x}(\cdot))=\{(t,\widehat{x}(t),\dot{\widehat{x}}(t))\mid t\in[t_0,t_1]\}$, *то*

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

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

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

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

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

Доказательство. Запишем задачу (P_5) как задачу оптимального управления:

$$\int_{t_0}^{t_1} L(t, x(t), u(t))\, dt \to \min, \quad \dot{x} = u, \quad x(t_0) = x_0, \;\; x(t_1) = x_1. \;\; (P_5')$$

Легко видеть, что $\widehat{x}(\cdot)$ доставляет сильный минимум в (P_5) тогда и только тогда, когда пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$, где $\widehat{u}(\cdot) = \dot{\widehat{x}}(\cdot)$ является сильным минимумом в (P_5').

Согласно принципу максимума найдутся такой ненулевой набор множителей Лагранжа $\overline{\lambda} = (\lambda_0, p(\cdot)) \in \mathbb{R} \times PC^1([t_0, t_1])$, что для функции Лагранжа $\mathcal{L}(x(\cdot), u(\cdot), \overline{\lambda}) = \int_{t_0}^{t_1} (\lambda_0 L(t, x(t), u(t)) + p(t)(\dot{x}(t) - u(t)))\, dt$ для всех $t \in [t_0, t_1]$ выполнено условие стационарности по $x(\cdot)$:

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

и условие минимума по $u(\cdot)$:

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

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 $f\colon u \to L(t, \widehat{x}(t), u) - p(t)u$ на $\mathbb{R}$ достигает минимума 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. $\square$

При доказательстве необходимых условий слабого минимума в задаче (P_5) нам понадобится один несложный технический результат, который приводим без доказательства (см. В. М. Алексеев, В. М. Тихомиров, С. В. Фомин. Оптимальное управление. М.: Наука, 1979, стр. 69).

Лемма (о скруглении углов)**.** *Пусть в задаче* (P_9) *интегрант* L *непрерывен по совокупности переменных. Тогда*

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

Теорема (Необходимые условия слабого минимума в задаче (P_5))**.** *Пусть* $\widehat{x}(\cdot) \in C^1([t_0,t_1])$ *доставляет слабый локальный минимум в задаче* (P_5). *Тогда, если интегрант* L *дважды непрерывно дифференцируем в окрестности* $\Gamma(\widehat{x}(\cdot))$, *то для всех* $t \in [t_0,t_1]$ *выполнено уравнение Эйлера, условие Лежандра и условие Якоби.*

Доказательство. Уравнение Эйлера, как необходимое условие слабого экстремума, уже было доказано раньше. 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 $\varphi''(0) \geq 0$ 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

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

По лемме о скруглении углов $\widehat{h}(\cdot)=0$ 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

$$Q(\widetilde{h}(\cdot)) = \int_{t_0}^{\tau} \Big(\widehat{L}_{xx}(t)\overline{h}^2(t) + 2\widehat{L}_{\dot{x}x}(t)\overline{h}(t)\dot{\overline{h}}(t) + \widehat{L}_{\dot{x}\dot{x}}(t)\dot{\overline{h}}^2(t)\Big)\,dt$$
$$= \int_{t_0}^{\tau} \Big(\widehat{L}_{xx}(t)\overline{h}(t) + \widehat{L}_{x\dot{x}}(t)\dot{\overline{h}}(t)\Big)\,\overline{h}(t)\,dt + \int_{t_0}^{\tau} \Big(\widehat{L}_{\dot{x}x}(t)\overline{h}(t)$$
$$+ \widehat{L}_{\dot{x}\dot{x}}(t)\dot{\overline{h}}(t)\Big)\dot{\overline{h}}(t)\,dt = \int_{t_0}^{\tau} \Big(- \frac{d}{dt}\Big(\widehat{L}_{\dot{x}x}(t)\overline{h}(t) + \widehat{L}_{\dot{x}\dot{x}}(t)\dot{\overline{h}}(t)\Big)$$
$$+ \widehat{L}_{xx}(t)\overline{h}(t) + \widehat{L}_{x\dot{x}}(t)\dot{\overline{h}}(t)\Big)\overline{h}(t)\,dt.$$

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 $\widetilde{h}(\cdot)$ также доставляет сильный минимум в задаче (i). Запишем эту задачу как задачу оптимального управления

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

Согласно принципу максимума найдутся такие множители Лагранжа λ_0 и $p(\cdot) \in PC^1([t_0, t_1])$, не равные одновременно нулю, что для функции Лагранжа

$$\int_{t_0}^{t_1} \Big(\lambda_0\Big(\widehat{L}_{xx}(t)h^2(t) + 2\widehat{L}_{\dot{x}x}(t)h(t)u(t) + \widehat{L}_{\dot{x}\dot{x}}(t)u^2(t)\Big) +$$
$$+ p(t)(\dot{h}(t) - u(t))\Big)\,dt$$

выполнено условие стационарности по $h(\cdot)$:

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

и условие минимума по $u(\cdot)$:

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

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

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

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

По определению $\widetilde{h}(t) = 0$, если $t \geq \tau$ и поэтому из (iii) вытекает, что $p(\tau + 0) = 0$. Но функция $p(\cdot)$ непрерывна и поэтому (снова из (iii)) получаем $0 = p(\tau - 0) = \widehat{L}_{\dot{x}\dot{x}}(\tau)\dot{\widetilde{h}}(\tau - 0) = \widehat{L}_{\dot{x}\dot{x}}(\tau)\dot{\bar{h}}(\tau) \neq 0$, так как $\dot{\bar{h}}(\tau) \neq 0$ (как уже было отмечено) и $\widehat{L}_{\dot{x}\dot{x}}(\tau) > 0$ в силу того, что выполнено усиленное условие Лежандра. Пришли к противоречию и тем самым условие Якоби доказано. □

Вопросы к экзамену

1. Дифференцируемость, непрерывная дифференцируемость и строгая дифференцируемость конечномерных отображений.

2. Конечномерные теоремы отделимости.

3. Модифицированный метод Ньютона и разрешимость конечномерной системы нелинейных уравнений. Теорема Люстерника.

4. Теорема Ферма для конечномерной гладкой задачи без ограничений. Необходимые и достаточные условия экстремума второго порядка.

5. Правило множителей Лагранжа для конечномерной гладкой задачи с ограничениями, задаваемыми равенствами.

6. Правило множителей Лагранжа для конечномерной гладкой задачи с ограничениями, задаваемыми равенствами и неравенствами.

7. Необходимые и достаточные условия экстремума второго порядка для конечномерной гладкой задачи с ограничениями, задаваемыми равенствами.

8. Теорема Ферма для выпуклых задач без ограничений и правило множителей Лагранжа для конечномерных выпуклых задач с ограничениями (теорема Каруша–Куна–Таккера).

9. Необходимые и достаточные условия минимума максимума конечного числа гладких функций.

10. Уравнение Эйлера для простейшей задачи классического вариационного исчисления и его первые интегралы.

11. Необходимые условия слабого экстремума для задачи Больца.

12. Уравнения Эйлера–Лагранжа для задачи Лагранжа вариационного исчисления.

13. Необходимые условия экстремума для задачи со старшими производными (как следствие необходимых условий экстремума в задаче Лагранжа).

14. Необходимые условия экстремума для изопериметрической задачи (как следствие необходимых условий экстремума в задаче Лагранжа).

15. Принцип максимума Понтрягина для понтрягинской задачи оптимального управления

16. Необходимые условия сильного минимума для простейшей задачи вариационного исчисления.

17. Необходимые условия слабого минимума для простейшей задачи вариационного исчисления.