Вариационное исчисление и оптимальное управление (курс лекций 2010 г.)

Г. Г. Магарил-Ильяев

Содержание

Введение

Какие причины побуждают решать задачи на максимум и минимум? Несколько причин побуждают ставить и решать экстремальные задачи, т. е. задачи на максимум и минимум. Интерес к ним проявился уже на заре развития математики и основными стимулами были любознательность и стремление к совершенству. Но эти черты были присущи человеку во все времена, и они дают поводы для поиска наилучших решений.

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”.

Вторая причина связана с не совсем объяснимым свойством природы: многие ее законы, 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 задач на минимум. Отправляясь от принципа Ферма, И. Бернулли в 1696 году решил так называемую *задачу о брахистохроне* — о кривой наибыстрейшего ската, т. е. о форме кривой, соединяющей две точки в вертикальной плоскости, вдоль которой тело под действием силы тяжести без трения проходит путь от одной точки до другой за кратчайшее время (постановка такой задачи возможно была навеяна более ранними размышлениями Галилея на эту тему). Историю развития вариационного исчисления принято отсчитывать именно с этого времени — года решения задачи о брахистохроне.

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

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

Первой задачей такого рода была *аэродинамическая задача Ньютона* о форме тела вращения, испытывающем наименьшее сопротивление в некоей "редкой среде", поставленная и решенная Ньютоном в его "Математических началах натуральной философии" (1687 г.). К числу экстремальных задач, возникающих в экономике и теории управления, относятся, например, *транспортная задача* (сформулированная в сороковые годы прошлого века) о наилучшем способе транспортировки продуктов со складов в магазины и *простейшая задача о быстродействии*, формализующая, в частности, задачу о наименьшем времени движения лифта в угольной шахте. Эта задача, также сформулированная в сороковые годы, послужила стимулом для создания *теории оптимального управления*. Считалось, что эта теория родилось 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_0 (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

$$f_0(x) \to \min(\max), \quad x \in C \qquad (P_1)$$

и заключается в нахождении таких точек $x \in C$, в которых функционал f_0 достигает своего минимума (максимума) на C. Такие точки называются *глобальными* или *абсолютными минимумами (максимумами)* в задаче (P_1) 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_0.

Отметим еще, что если $\widehat{x}$ — решение задачи (P_1) на минимум (максимум), то ясно, что $\widehat{x}$ — решение аналогичной задачи на максимум (минимум) с функционалом $-f_0$ вместо f_0.

Точки из множества ограничений C называются *допустимыми* в задаче (P_1). Если $C = X$, то задача (P_1) 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_1), если существует такая ее окрестность U, что $f_0(x) \ge f_0(\widehat{x})$ $(f_0(x) \le f_0(\widehat{x}))$ для всех допустимых $x \in U$ (т. е. для всех $x \in C \cap U$).

Цель курса — изложить начала теории экстремума. Эта теория состоит из пяти основных глав: база теории (дифференциальное исчисление, обыкновенные дифференциальные уравнения, выпуклый анализ); необходимые условия экстремума; достаточные условия экстремума; теория существования решений экстремальных задач и алгоритмы поиска таких решений. Далее будут затронуты, так или иначе, все из этих направлений, но наибольшее внимание будет уделено условиям экстремума.

Первые два параграфа относятся к базе теории экстремума. В них доказываются два результата, играющих ключевую роль в построении этой теории — конечномерная теорема отделимости для выпуклых множеств и теорема о разрешимости системы нелинейных уравнений конечного числа переменных.

1. Дифференциальные свойства функций и отображений на $\mathbb{R}^n$. Конечномерные теоремы отделимости

В этом пункте напоминаются необходимые для дальнейшего факты, связанные с пространством $\mathbb{R}^n$, с понятиями дифференцируемости и строгой дифференцируемости функций и отображений на $\mathbb{R}^n$, а также доказываются теоремы об отделимости выпуклых множеств.

Пространство $\mathbb{R}^n$ и двойственное к нему, открытые и замкнутые множества, компактность, теорема Вейрштрасса. Пусть n — натуральное число. Пространство $\mathbb{R}^n$ — это совокупность всех упорядоченных наборов $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ из n действительных чисел (если $n = 1$, то это просто множество действительных чисел, и мы пишем $\mathbb{R}$ вместо $\mathbb{R}^1$), называемых *векторами* или *вектор-столбцами*, а числа x_i, $1 \le i \le n$ — *координатами вектора* x. Ради экономии места, элементы $\mathbb{R}^n$ будем записывать так $x = (x_1, \ldots, x_n)^T$, где ^{T} обозначает *транспонирование* (перевод строки в столбец). 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 $a = (l(e_1), \ldots, l(e_n))$, где $e_1 = (1, 0, \ldots, 0)^T, \ldots, e_n = (0, \ldots, 0, 1)^T$ — *стандартный базис* в $\mathbb{R}^n$. Таким образом, если обозначить через $(\mathbb{R}^n)^*$ множество, элементы которого суть те же наборы из n действительных чисел, но расположенные в строку (с аналогичными операциями сложения и умножения на числа), то его можно отождествить с пространством всех линейных функционалов на $\mathbb{R}^n$, которое называют *сопряженным* или *двойственным* пространством к $\mathbb{R}^n$.

Далее, каждому $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ 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 $x, y \in \mathbb{R}^n$. Первые два свойства очевидны, третье (называемое |it 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 $\delta > 0$. Множества $U_{\mathbb{R}^n}(\widehat{x}, \delta) = \{\, x \in \mathbb{R}^n \mid |x - \widehat{x}| < \delta \,\}$ и $B_{\mathbb{R}^n}(\widehat{x}, \delta) = \{\, x \in \mathbb{R}^n \mid |x - \widehat{x}| \leq \delta \,\}$ называются соответственно *открытым* и *замкнутым* 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 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. $\mathrm{int\,G} = \mathrm{G}$. *Окрестностью точки* $x \in \mathbb{R}^n$ 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.

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 $A \subset \mathrm{cl\,A}$.

Говорят, что последовательность $\{x_k\}_{k \in \mathbb{N}}$ элементов из $\mathbb{R}^n$ сходится к $x \in \mathbb{R}^n$ (и пишут $x = \lim_{k \to \infty} x_k$ или $x_k \to x$ при $k \to \infty$), если $|x_k - x| \to 0$ при $k \to \infty$.

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

Множество в $\mathbb{R}^n$ компактно тогда и только тогда, когда оно ограничено (т. е. содержится в некотором шаре) и замкнуто (продумайте это).

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.

В курсе анализа доказывается

Теорема (Вейерштрасса)**.** *Функция непрерывная на компакте достигает на нем своего максимального и минимального значений.*

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 $f_1, \dots, f_m$, то они определяют отображение $F\colon A \to \mathbb{R}^m$ по правилу $F(x) = (f_1(x), \dots, f_m(x))^T$, $x \in A$. 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.

Пусть $\Lambda\colon \mathbb{R}^n \to \mathbb{R}^m$ — линейный оператор. Мы будем отождествлять этот оператор с его матрицей в стандартных базисах в $\mathbb{R}^n$ и $\mathbb{R}^m$ соответственно, так что Λx — произведение матрицы Λ на вектор x.

Обозначим через $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ множество всех линейных операторов из $\mathbb{R}^n$ в $\mathbb{R}^m$. Это векторное пространство с обычными (поэлементными) операциями сложения матриц и умножения их на числа. Каждому $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ сопоставим число $\|\Lambda\| = \sup_{|x| \le 1} |\Lambda x|$[2], которое называется нормой оператора Λ. Простая проверка показывает, что выполняются следующие условия: (a) $\|\Lambda\| \ge 0$ для любого $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ и $\|\Lambda\| = 0$ тогда и только тогда, когда $\Lambda = 0$; (b) $\|\alpha\Lambda\| = |\alpha|\|\Lambda\|$ для любых $\alpha \in \mathbb{R}$ и $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$; (c) $\|\Lambda_1 + \Lambda_2\| \le \|\Lambda_1\| + \|\Lambda_2\|$ для любых $\Lambda_1, \Lambda_2 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Ясно также, что $|\Lambda x| \le \|\Lambda\||x|$ для любого $x \in \mathbb{R}^n$.

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

---

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

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

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

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

Упражнение. 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: $f(x) = x^T Ax + b \cdot x + c$. Найдите ее производную для каждого $x \in \mathbb{R}^n$.

Если функция f дифференцируема в $\widehat{x} = (\widehat{x}_1, \ldots, x_n)^T$, то, очевидно, существуют производные в нуле функций $t \mapsto f(\widehat{x}_1 + t, \widehat{x}_2, \ldots, \widehat{x}_n), \ldots, t \mapsto f(\widehat{x}_1, \ldots, \widehat{x}_{n-1}, \widehat{x}_n + t)$, которые называются частными производными функции f по $x_1, \ldots, x_n$ в точке $\widehat{x}$ и обозначаются соответственно $\partial f(\widehat{x})/\partial x_1, \ldots, \partial f(\widehat{x})/\partial x_n$. Легко видеть, что $a_i = \partial f(\widehat{x})/\partial x_i$, $i = 1, \ldots, n$, и таким образом, $f'(\widehat{x}) = (\partial f(\widehat{x})/\partial x_1, \ldots, \partial f(\widehat{x})/\partial x_n)$.

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

Существование частных производных функции в данной точке не гарантирует ее дифференцируемости и даже непрерывности в этой точке. Соответствующих примеров множество. Скажем, функция на $\mathbb{R}^2$, равная единице на множестве $\{\, (x_1, x_2) \mid x_1 x_2 = 0 \,\}$ и нулю в остальных точках, разрывна в нуле, но частные производные в этой точке существуют.

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

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

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 $F'(\widehat{x})$.

Если на U заданы 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, $1 \leq i \leq m$, 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.

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

Непрерывная дифференцируемость функции f на U равносильна тому, что все частные производные функции f непрерывны на U (продумайте это).

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

Если отображение f' (производная функции f на U) дифференцируемо 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.

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.* $r(h) = o(|h|^2)$.

Строгая дифференцируемость функций и отображений. Отображение $F\colon U \to \mathbb{R}^m$ называется *строго дифференцируемым 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.

Из определения строгой дифференцируемости F в точке $\widehat{x}$ легко следует, что для каждого $\varepsilon > 0$ это отображение 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. Рассмотрим, например, 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 окрестности точки $\widehat{x}$, то оно строго дифференцируемо в этой точке. Это простое следствие теоремы о среднем, примененной к отображению $x \mapsto F(x) - F'(\widehat{x})x$.

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

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.*

Контрольный вопрос. Напишите уравнение гиперплоскости в $\mathbb{R}^2$, отделяющей точку $(1,1)$ от круга $\{\,(x_1, x_2) \mid x_1^2 + x_2^2 < 2\,\}$.

Мы докажем только теорему о строгой отделимости точки от множества (которую обычно называют *второй теоремой отделимости*). Теорему об отделимости множеств (которую называют *первой теоремой отделимости*) приведем без доказательства.

Напомним понятие выпуклого множества. Непустое подмножество A векторного пространства X называется *выпуклым*, если 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 эти точки.

Теорема 1.2 (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 (продумайте это) 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 $H = \{x \in \mathbb{R}^n \mid x^* \cdot x = x^* \cdot \widehat{x}\}$. Тогда $x^* \cdot b = x^* \cdot (b - \widehat{x} + \widehat{x}) = x^* \cdot (-x^{*T} + \widehat{x}) = -|x^*|^2 + x^* \cdot \widehat{x} < x^* \cdot \widehat{x}$. Теперь покажем, что A находится в другом полупространстве, т. е. $x^* \cdot x \ge x^* \cdot \widehat{x}$ для любого $x \in A$. Предположим противное, что существует такой элемент $x_0 \in A$, что $x^* \cdot x_0 < x^* \cdot \widehat{x}$. Поскольку A выпукло, то $(1-t)\widehat{x} + t x_0 \in A$ при $0 \le t \le 1$ 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, $x^* \cdot b < x^* \cdot \widehat{x} \le x^* \cdot x$ для всех $x \in A$, или $x^* \cdot b < \inf_{x \in A} x^* \cdot x$, т. е. A и b строго отделимы. $\square$

Теорема 1.1 (первая теорема отделимости). *Пусть A и B — непустые выпуклые подмножества $\mathbb{R}^n$ и $A \cap B = \emptyset$. Тогда A и B отделимы.*

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. Теорема отделимости точки от выпуклого множества доказана Минковским (1910 г.).

2. Модифицированный метод Ньютона и разрешимость конечномерной 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. $x_k = x_{k-1} + (F'(x_{k-1}))^{-1}(y - F(x_{k-1}))$, $k \in \mathbb{N}$. 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 $f_1(x_1, \dots, x_n) = y_1, \dots, f_m(x_1, \dots, x_n) = y_m$) 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 $y \in \mathbb{R}^m$. Таким образом, 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: $R(y) = \Lambda^T(\Lambda\Lambda^T)^{-1}$, где Λ^T — матрица, транспонированная к Λ ($\Lambda\Lambda^T$ — обратимая матрица, как нетрудно проверить). Но, имея в виду некоторые дальнейшие обобщения, приведем еще и следующее доказательство. Пусть $e_1, \dots, e_m$ — базис в $\mathbb{R}^m$. По условию найдутся такие $f_i \in \mathbb{R}^n$, что $\Lambda f_i = e_i$, $1 \le i \le m$. Для каждого $y = \sum_{i=1}^m y_i e_i \in \mathbb{R}^m$ положим $R(y) = \sum_{i=1}^m y_i f_i$. Ясно, что $\Lambda R(y) = \sum_{i=1}^m y_i \Lambda f_i = \sum_{i=1}^m y_i e_i = y$ и $|R(y)| \le \sum_{i=1}^m |y_i||f_i| \le \max_{1\le i\le m} |y_i| \sum_{i=1}^m |f_i| \le \gamma|y|$, где $\gamma = \sum_{i=1}^m |f_i|$.

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

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

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

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

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

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

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

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

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, что $F(\varphi(\xi, y)) = y$. Полагая в (v) $k = 0$ и переходя к пределу при $l \to \infty$, получаем, что $|\varphi(\xi, y) - \xi| \le (\gamma/(1-\theta))|y - F(\xi)|$.

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 (v), получаем, что отображение φ есть равномерный предел непрерывных функций и значит, само непрерывно на $V \times W$. $\square$

При доказательстве всех необходимых условий экстремума мы будем пользоваться непосредственно данной теоремой (либо незначительными ее модификациями). Но часто вместо нее используют теорему 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 $|\varphi(\widehat{x}+th, F(\widehat{x}))-\widehat{x}-th| \le K|F(\widehat{x}+th)-F(\widehat{x})| = K|F(\widehat{x})+tF'(\widehat{x})h+o(t)-F(\widehat{x})| = |o(t)|$, 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 зарубежной литературе его иногда называют методом ... . В бесконечномерном случае этот метод изучался Л. В. Канторовичем.

3. Теорема Ферма и условия экстремума второго порядка для конечномерных 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
$$f(x) \to \text{extr}, \quad x \in U. \qquad (P_2)$$
С точки зрения общей задачи (P_1) здесь $U = C = X$, т. е. (P_2) — задача без ограничений.

Теорема (Ферма, необходимые условия экстремума в гладких задач без ограничений)**.** *Если точка $\widehat{x} \in U$ является локальным экстремумом в задаче (P_2) и функция f дифференцируема в $\widehat{x}$, то*
$$f'(\widehat{x}) = 0.$$

Доказательство. Предположим, что (линейный функционал) $f'(\widehat{x})$ отличен от нуля. Тогда найдется элемент $x \in \mathbb{R}^n$, $x \neq 0$, такой, что $f'(\widehat{x}) \cdot x \neq 0$. 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. □

Исторический комментарий. Впервые мысль о том, что в точке максимума или минимума функции “приращение несущественно”, высказал И.Кеплер в 1615 году. В словах И. Ньютона, вынесенных в эпиграф, говорится об аналогичном наблюдении. Аналитически этот факт выразил П. Ферма в 1638 году в письме к Робервалю и Мерсену, предназначенном для Р. Декарта. (Хотя понятие производной тогда еще не было, но с ее появлением рассуждения Ферма стало возможно интерпретировать, как равенство нулю производной функции в точке ее локального экстремума). 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_2) 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 $h \to h^T f''(\widehat{x})h = \sum_{i,j=1}^n a_{ij}h_ih_j$, где $a_{ij} = \partial^2 f(\widehat{x})/\partial x_i \partial x_j$, $i, j = 1, \ldots, n$, положительно (отрицательно) определена. Согласно критерию Сильвестра это равносильно тому, что главные миноры матрицы $f''(\widehat{x})$ (т. е. определители матриц $A_k = (a_{ij})$, $i, j = 1, \ldots, k$, $k = 1, \ldots, n$) положительны (чередуют знаки, причем $A_1 = a_{11} < 0$).

3. Правило множителей Лагранжа для конечномерных гладких задач с ограничениями, задаваемые равенствами. Условия второго порядка

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

Как и теорема Ферма, правило множителей Лагранжа для гладких задач с ограничениями, задаваемыми равенствами — изначальный результат теории экстремума. Из дальнейшего будет видно, что замысел Лагранжа "снятия ограничений" (с помощью функции Лагранжа) оказался поразительно плодотворным. Необходимые условия экстремума в самых разных задачах (от истоковых до исследуемых в настоящее время) имеют форму правила множителей Лагранжа.

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

$$f_0(x) \to \mathrm{extr}, \quad f_i(x) = 0, \quad i = 1, \dots, m \qquad (P_3)$$

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

С точки зрения общей задачи (P_1) здесь $X = U$ и множество ограничений $C = \{x \in U \mid f_i(x) = 0,\ 1 \le i \le m\}$.

Свяжем с задачей (P_3) следующую функцию $\mathcal{L}\colon U \times (\mathbb{R}^{m+1})^* \to \mathbb{R}$, $\mathcal{L}(x, \overline{\lambda}) = \sum_{i=0}^m \lambda_i f_i(x)$, которая называется *функцией Лагранжа задачи* (P_3), числа $\lambda_0, \lambda_1, \dots, \lambda_m$ — *множителями Лагранжа*, а вектор $\overline{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_m)$ — *набором множителей Лагранжа*.

Теорема 3 (Правило множителей Лагранжа для задачи (P_3)). *Если точка $\widehat{x} \in U$ является локальным экстремумом в задаче (P_3) и функции f_i, $i = 0, 1, \dots, m$, строго дифференцируемы в*

$\widehat{x} \in U$, то 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_3) 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, F строго дифференцируемо в $\widehat{x}$ в силу строгой дифференцируемости в этой точке функций $f_0, f_1, \ldots, f_m$. Следовательно, применима теорема о разрешимости системы уравнений, согласно которой найдется окрестности V и W точек $\widehat{x}$ и $F(\widehat{x})$, 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 $(f_0(x_\nu), f_1(x_\nu), \ldots, f_m(x_\nu))^T = (f_0(\widehat{x}) + \nu, 0, \ldots, 0)^T$ и отображение φ непрерывно ($x_\nu \to \widehat{x}$ при $\nu \to 0$), то в любой окрестности $\widehat{x}$ находятся допустимые в задаче (P_3) 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. □

Сделаем несколько замечаний по поводу доказанной теоремы.

1. 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 Лагранжа, согласно его словам выше) и тогда необходимые условия экстремума и "уравнения связи" $f_i(x) = 0$, $i = 1, \ldots, m$, дают $n + m$ соотношений для определения $n + m$ неизвестных: $x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m$.

2. Стандартные требования о гладкости функций в правиле множителей Лагранжа — это их непрерывная дифференцируемость в окрестности точки $\widehat{x}$. В силу доказанного выше, классическое правило множителей Лагранжа следует из данной теоремы.

3. В утверждение теоремы входят только производные функций в точке $\widehat{x}$. Насколько существенны требования строгой дифференцируемости? От непрерывности функций $f_1, \ldots, f_m$ в окрестности точки точки $\widehat{x}$ (которая следует из строгой дифференцируемости) отказаться нельзя (соответствующий пример см. ...). От функции f_0 достаточно требовать, только дифференцируемость в $\widehat{x}$ (см. замечания в следующем параграфе).

Теорема (Необходимые и достаточные условия экстремума второго порядка для задачи (P_3))**.** *Пусть в задаче (P_3) функции f_i, $i = 0, 1, \ldots, m$, строго дифференцируемы в точке $\widehat{x}$, имеют в этой* *точке вторую производную и векторы* $f_1'(\widehat{x}),\ldots,f_m'(\widehat{x})$ *линейно независимы. Тогда*

1) *если* $\widehat{x}$ *— локальный минимум (максимум) в* (P_3)*, то существует набор множителей Лагранжа* $\overline{\lambda}=(1,\lambda_1,\ldots,\lambda_m)$ *такой, что для любого* $h\in\mathbb{R}^n$*, для которого* $f_i'(\widehat{x})\cdot h=0$*,* $i=1,\ldots,m$*, справедливо неравенство*

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

2) *если* $\widehat{x}$ *— допустимая точка в* (P_3)*, существует такой набор множителей Лагранжа* $\overline{\lambda}=(1,\lambda_1,\ldots,\lambda_m)$*, что* $\mathcal{L}_x(\widehat{x},\overline{\lambda})=0$ *и для любого ненулевого* $h\in\mathbb{R}^n$*, для которого* $f_i'(\widehat{x})\cdot h=0$*,* $i=1,\ldots,m$*, справедливо неравенство*

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

то $\widehat{x}$ *— локальный минимум (максимум) в задаче* (P_3)*.*

Доказательство. 1) Воспользуемся теоремой о разрешимости системы уравнений. Пусть отображение $F\colon U\to\mathbb{R}^m$ определено по правилу: $F(x)=(\,f_1(x),\ldots,f_m(x)\,)^T$. Строгая дифференцируемость функций f_i, $1\leq i\leq m$ в точке $\widehat{x}$ равносильна строгой дифференцируемости F в этой точке. Кроме того, $F'(\widehat{x})(\mathbb{R}^n)=\mathbb{R}^m$ в силу линейной независимости векторов $f_1'(\widehat{x}),\ldots,f_m'(\widehat{x})$. Пусть φ — соответствующее отображение из теоремы о разрешимости и $h\in\mathbb{R}^n$ таково, что $f_i'(\widehat{x})\cdot h=0$, $i=1,\ldots,m$ (или, равносильно, $F'(\widehat{x})h=0$). Так как $F(\widehat{x})=0$, то для всех достаточно малых t будем иметь $F(\varphi(\widehat{x}+th,0))=0$ и $|\varphi(\widehat{x}+th,0)-\widehat{x}-th|\leq K|F(\widehat{x}+th)|=K|F(\widehat{x})+tF'(\widehat{x})h+o(t)|=|o(t)|$, т. е. $\varphi(\widehat{x}+th,0)=\widehat{x}+th+r(t)$, где $r(t)=o(t)$.

Таким образом, $f_i(\widehat{x}+th+r(t))=0$, $i=1,\ldots,m$, и значит, для таких t точки $\widehat{x}+th+r(t)$ допустимы в задаче (P_3). 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, что $\widehat{x}$ — локальный минимум:

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

откуда следует нужное утверждение.

Для случая локального максимума рассуждения аналогичны.

2) Доказываем от противного. Пусть $\widehat{x}$ не является локальным экстремумом в задаче (P_3). Для определенности считаем, что $\widehat{x}$ — не локальный минимум. Покажем, что в этом случае найдется такое ненулевое $\overline{h}\in K(\widehat{x})$, 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

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

Действительно, так как $\widehat{x}$ — не локальный минимум, то существует последовательность $\{x_k\}_{k\in\mathbb{N}}$ допустимых элементов в (P_3) 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 $0\leq i\leq m$ 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 $f_i(x_k) = f_i(\widehat{x})$, $i = 1, \ldots, m$, будем иметь

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

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

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

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

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

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

4. Правило множителей Лагранжа для конечномерных гладких задач с ограничениями, задаваемые равенствами и неравенствами (теорема Каруша–Джона)

В этом параграфе будут получены необходимые условия экстремума в гладких задачах с равенствами и неравенствами. Удивительно, что такие задачи попали в поле зрения математиков только в тридцатые годы прошлого века. По сути дела, подобные задачи представляют собой некий симбиоз гладких задач с равенствами и выпуклых задач (неравенства “порождают” выпуклость).

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

$$f_0(x) \to \min, \quad f_i(x) \le 0, \quad 1 \le i \le m', \quad f_i(x) = 0, \quad m'+1 \le i \le m, \quad (P_4)$$

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

Ясно, что задача (P_3) является частным случаем данной задачи, но она была рассмотрена отдельно ввиду ее основополагающей роли в теории экстремума.

С точки зрения общей задачи (P_1) 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_4) функцию Лагранжа $\mathcal{L}\colon U \times (\mathbb{R}^{m+1})^* \to \mathbb{R}$, $\mathcal{L}(x, \overline{\lambda}) = \sum_{i=0}^{m} \lambda_i f_i(x)$, которая, как мы видим, имеет тот же вид, что и в задаче (P_3).

Теорема 4 (Каруша–Джона, правило множителей Лагранжа для задачи (P_4)). *Если точка $\widehat{x} \in U$ является локальным минимумом в задаче (P_4), функции f_i, $i = 0, 1, \ldots, m'$, дифференцируемы, а функции f_i, $i = m'+1, \ldots, m$, строго дифференцируемы в $\widehat{x}$, то найдется такой ненулевой набор множителей Лагранжа $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m)$, 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*) $\lambda_i \ge 0,\ i = 0, 1, \ldots, m'$ (*условие неотрицательности*);
(*c*) $\lambda_i f_i(\widehat{x}) = 0,\ i = 1, \ldots, m'$ (*условие дополняющей нежесткости*).

4. Необходимые и достаточные условия минимума второго порядка в гладких задач с равенствами и неравенствами теорема Левитина–Милютина–Осмоловского)

Пусть функции f_i, $i = 0, 1, \dots, m$, в задаче (P_4) дифференцируемы в точке $\widehat{x}$. Свяжем с этой задачей два множества. Первое — это нормированное множество всех множителей Лагранжа, удовлетворяющих необходимым условиям минимума:

$$\Omega(\widehat{x}) = \{\, \overline{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_m) \in (\mathbb{R}^{m+1})^* \mid \lambda_i \ge 0, \ \ i = 0, 1, \dots, m'; \\ \sum_{i=0}^{m} \lambda_i f_i'(\widehat{x}) = 0, \quad \sum_{i=0}^{m'} \lambda_i + \sum_{i=m'+1}^{m} |\lambda_i| = 1 \,\}$$

и так называемый *конус критических вариаций*:

$$K(\widehat{x}) = \{\, h \in \mathbb{R}^n \mid f_i'(\widehat{x}) \cdot h \le 0, \ \ i = 0, 1, \dots, m'; \quad f_i'(\widehat{x}) \cdot h = 0, \\ i = m'+1, \dots, m \,\}.$$

Теорема (Левитина–Милютина–Осмоловского, необходимые и достаточные условия минимума второго порядка для задачи (P_4))**.** *Пусть в задаче (P_4) функции f_i, $i = 0, 1, \dots, m$, дважды непрерывно дифференцируемы на U. Тогда*

1) *если $\widehat{x}$ — локальный минимум в (P_4), то $\Omega(\widehat{x}) \ne \emptyset$ и для любого $h \in K(\widehat{x})$ найдется вектор $\overline{\lambda} = \overline{\lambda}(h) \in \Omega(\widehat{x})$ такой, что*
$$h^T \mathcal{L}_{xx}(\widehat{x}, \overline{\lambda}) h \ge 0;$$
2) *если $\widehat{x}$ — допустимая точка в (P_4), $\Omega(\widehat{x}) \ne \emptyset$ и для любого $h \in K(\widehat{x})$, $h \ne 0$, найдется вектор $\overline{\lambda} = \overline{\lambda}(h) \in \Omega(\widehat{x})$ такой, что*
$$h^T \mathcal{L}_{xx}(\widehat{x}, \overline{\lambda}) h > 0,$$
то $\widehat{x}$ — локальный минимум в задаче (P_3).

Заметим, что множество $\Omega(\widehat{x})$ компактно в $(\mathbb{R}^{m+1})^*$ и для любого $h \in \mathbb{R}^n$ функция $\overline{\lambda} \to h^T \mathcal{L}_{xx}(\widehat{x}, \overline{\lambda}) h$ непрерывна. Тогда по теореме Вейерштрасса она достигает на $\Omega(\widehat{x})$ своего минимального и максимального значений. Учитывая это, утверждения теоремы можно сформулировать так: *если $\widehat{x}$ — локальный минимум, то $\Omega(\widehat{x}) \ne \emptyset$ и для любого $h \in K(\widehat{x})$ справедливо неравенство*

$$\max_{\lambda \in \Omega(\widehat{x})} h^T \mathcal{L}_{xx}(\widehat{x}, \overline{\lambda}) h \ge 0,$$

а если $\widehat{x}$ — допустимая точка, $\Omega(\widehat{x}) \ne \emptyset$ и для любого $h \in K(\widehat{x})$, $h \ne 0$, справедливо неравенство

$$\max_{\lambda \in \Omega(\widehat{x})} h^T \mathcal{L}_{xx}(\widehat{x}, \overline{\lambda}) h > 0,$$

то $\widehat{x}$ — локальный минимум.

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

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

Основы теории специального класса выпуклых задач — задач линейного программирования — были заложены Л. В. Канторовичем в его работе 1939 г. и впоследствии были переоткрыты американскими математиками. Особую роль в формировании этого направления сыграл Дж. фон Нейман. В 1935 г. совместно в авторитетным экономистом О. Маргенштерном им была написана книга "Теория игр и экономическое развитие", положившая 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), выпуклых функций (основы которой разработаны В. Фенхелем) и выпуклые экстремальные задачи, которым посвящен данный параграф.

5.1. Выпуклая задача без ограничений. Пусть X — вещественное векторное пространство и $x, y \in X$. Напомним, что 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: $x \mapsto ax^2 + bx + c$, $a \geq 0$; $x \mapsto e^{\alpha x}$, $\alpha \in \mathbb{R}$; $x \mapsto |x|^p$, $p \geq 1$; $x \mapsto -\ln x$, если $x > 0$ и $+\infty$, если $x \leq 0$; $x \mapsto x \log_2 x + (1-x)\log_2(1-x)$, если $0 < x < 1$ и $+\infty$ в остальных случаях.

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

Собственная функция f на X выпукла тогда и только тогда, когда для любых $x_1, x_2 \in X$ и любых $\alpha_i \geq 0$, $i = 1, 2$, таких, что $\alpha_1 + \alpha_2 = 1$ справедливо неравенство $f(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2)$, которое называется *неравенством Иенссена* (продумайте это).

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

Из определений следует, что субдифференциал есть выпуклое замкнутое множество.
Контрольный вопрос. Чему равен субдифференциал функции $x \mapsto |x|$, $x \in \mathbb{R}^n$, в нуле (начните со случая $n = 1$).

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: $f((1-\alpha)\widehat{x} + \alpha x) \leq (1-\alpha) f(\widehat{x}) + \alpha f(x)$, откуда

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

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

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

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

т. е. $f'(\widehat{x}) \cdot x \geq x^* \cdot x$ 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

$$f_0(x) \to \min, \quad x \in \mathbb{R}^n,$$

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 $f_0(\widehat{x}) \leq f_0(x)$ для всех $x \in U$. Пусть теперь x — произвольная точка из $\mathbb{R}^n$. Для достаточно малых $0 < \alpha \leq 1$ точки $(1-\alpha)\widehat{x} + \alpha x$ принадлежат U и поэтому (по неравенству Иенссена) $f_0(\widehat{x}) \leq f_0((1-\alpha)\widehat{x} + \alpha x) \leq (1-\alpha) f_0(\widehat{x}) + \alpha f_0(x)$, откуда следует, что $f_0(\widehat{x}) \leq f_0(x)$.

Теорема (Ферма для выпуклых функций)**.** *Точка $\widehat{x}$ является минимумом в выпуклой задаче без ограничений тогда и только тогда, когда $0 \in \partial f_0(\widehat{x})$.*

Доказательство. Если $\widehat{x}$ — минимум, то $f_0(x) - f_0(\widehat{x}) \geq 0 = 0 \cdot x$ для любого $x \in \mathbb{R}^n$, т. е. $0 \in \partial f_0(\widehat{x})$. Если $0 \in \partial f_0(\widehat{x})$, то $f_0(x) - f_0(\widehat{x}) \geq 0 \cdot x = 0$, т. е. $f_0(x) \geq f_0(\widehat{x})$ для любого $x \in \mathbb{R}^n$. □

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

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

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

С задачей (P_5) связывается функция Лагранжа $\mathcal{L} \colon X \times (\mathbb{R}^{m+1})^* \to \mathbb{R}$, $\mathcal{L}(x, \overline{\lambda}) = \sum_{i=0}^{m} \lambda_i f_i(x)$, которая имеет тот же вид, что и в задачах (P_3) и (P_4).

Теорема 5 (Каруша–Куна–Таккера). *Если $\widehat{x}$ — минимум в задаче (P_5), 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)* $\lambda_i \geq 0$, $i = 0, 1, \ldots, m$ *(условие неотрицательности);*
- *(c)* $\lambda_i f_i(\widehat{x}) = 0$, $i = 1, \ldots, m$ *(условие дополняющей нежесткости).*

Если существует допустимая в (P_5) 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_5).

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

Доказательство. Пусть $\widehat{x}$ — решение задачи (P_5). Рассмотрим множества $M = \{\, \mu = (\mu_0, \mu_1, \ldots, \mu_m)^T \in \mathbb{R}^{m+1} \mid \exists x \in A : f_0(x) - f_0(\widehat{x}) < \mu_0,\ f_i(x) \leq \mu_i,\ i = 1, \ldots, m \,\}$. Покажем, что это множество не пусто и выпукло.

Действительно, непустота следует из того, что вектор $\mu \in \mathbb{R}^{m+1}$ с положительными компонентами принадлежит M ($x = \widehat{x}$). Докажем выпуклость M. Пусть $\mu = (\mu_0, \ldots, \mu_m)^T, \mu' = (\mu'_0, \ldots, \mu'_m)^T \in M$ и $\alpha \in [0, 1]$. Найдутся такие $x, x' \in A$, что $f_0(x) - f_0(\widehat{x}) < \mu_0$, $f_i(x) \leq \mu_i$, $i = 1, \ldots, m$ и $f_0(x') - f_0(\widehat{x}) < \mu'_0$, $f_i(x') \leq \mu'_i$, $i = 1, \ldots, m$. Положим $x_\alpha = (1 - \alpha)x + \alpha x'$. Тогда $x_\alpha \in A$. Далее $f_0(x_\alpha) - f_0(\widehat{x}) = f_0((1 - \alpha)x + \alpha x') - f_0(\widehat{x}) \leq (1 - \alpha) f_0(x) + \alpha f_0(x') - f_0(\widehat{x}) = (1 - \alpha)(f_0(x) - f_0(\widehat{x})) + \alpha(f_0(x') - f_0(\widehat{x})) < (1 - \alpha)\mu_0 + \alpha\mu'_0$ и аналогично $f_i(x_\alpha) \leq (1 - \alpha)\mu_i + \alpha\mu'_i$, $i = 1, \ldots, m$, т. е. $(1 - \alpha)\mu + \alpha\mu' \in M$ и тем самым M выпукло.

Покажем, что $0 \notin M$. Действительно, если $0 \in M$, то найдется элемент $\widetilde{x} \in A$ такой, что $f_i(\widetilde{x}) \leq 0$, $i = 1, \ldots, m$, и $f_0(\widetilde{x}) - f_0(\widehat{x}) < 0$, т. е. вектор $\widetilde{x}$ допустим в задаче (P_5) и $f_0(\widetilde{x}) < f_0(\widehat{x})$ в противоречие с тем, что $\widehat{x}$ — минимум.

Согласно первой теореме отделимости (теорема 1.1) найдется такой ненулевой вектор $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in (\mathbb{R}^{m+1})^*$, что

$$\sum_{i=0}^{m} \lambda_i \mu_i \geq 0 \quad \forall\ \mu = (\mu_0, \mu_1, \ldots, \mu_m)^T \in M. \tag{i}$$

Пусть $\varepsilon > 0$. Подставляя в (i) вектор $(\varepsilon, \ldots, \varepsilon, 1, \varepsilon, \ldots, \varepsilon)^T$, где 1 стоит на i_0-ом месте, получаем, что $\lambda_{i_0} \geq -\varepsilon \sum_{i \neq i_0} \lambda_i$, откуда, в силу произвольности ε, следует, что $\lambda_{i_0} \geq 0$ и значит, доказано утверждение (b).

Подставим в (i) вектор $(\delta, 0, \ldots, 0, f_i(\widehat{x}), 0, \ldots, 0)^T$, где $\delta > 0$ (этот вектор принадлежат M, надо взять $x = \widehat{x}$). Тогда получим, что $\lambda_i f_i(\widehat{x}) \geq -\lambda_0 \delta$, откуда $\lambda_i f_i(\widehat{x}) \geq 0$ в силу произвольности δ. Но $\lambda_i f_i(\widehat{x}) \leq 0$, так как $\lambda_i \geq 0$, а $f_i(\widehat{x}) \leq 0$ и поэтому $\lambda_i f_i(\widehat{x}) = 0$, $i = 1, \ldots, m$, что доказывает утверждение (c).

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

Докажем второе утверждение теоремы. Пусть x — допустимый вектор в задаче (P_5). Тогда, используя это обстоятельство вместе с (b), затем (a) и (c), будем иметь $\lambda_0 f_0(x) \geq \lambda_0 f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) = \mathcal{L}(x, \overline{\lambda}) \geq \mathcal{L}(\widehat{x}, \overline{\lambda}) = \lambda_0 f_0(\widehat{x}) + \sum_{i=1}^m \lambda_i f_i(\widehat{x}) = \lambda_0 f_0(\widehat{x})$. 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). $\square$

5.3. 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 $f(x) - f(\widehat{x}) \geq x^* \cdot (x - \widehat{x})$ 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 A — непустое подмножество векторного пространства X. Наименьшее выпуклое множество, содержащее A называется выпуклой оболочкой A и обозначается $\operatorname{co} A$.

Упражнение. Доказать, что $\operatorname{co} A$ — множество всех выпуклых комбинаций элементов из A, т. е. векторов вида $x = \alpha_1 x_1 + \ldots + \alpha_m x_m$, где $x_i \in A$, $\alpha_i \geq 0$, $i = 1, \ldots, m$, и $\sum_{i=1}^m \alpha_i = 1$.

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

$$f(x) = \max_{1 \leq i \leq m} \{ f_i(x) \} \to \max, \quad x \in U. \tag{P_5'}$$

Теорема. *Пусть в задаче (P_5') функции f_i, $1 \leq i \leq m$, 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, $1 \leq i \leq m$, 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 $A = \operatorname{co}\{\, f_1'(\widehat{x}), \dots, f_m'(\widehat{x}) \,\}$. 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 $\sup\{\, \sum_{i=1}^m \alpha_i f_i'(\widehat{x}) \cdot \overline{x} \mid \alpha_i \geq 0,\ 1 \leq i \leq m,\ \sum_{i=1}^m \alpha_i = 1 \,\} = \max_{1 \leq i \leq m}\{\, f_i'(\widehat{x}) \cdot \overline{x} \,\} < 0$. Отсюда следует, что для каждого $1 \leq i \leq m$ 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 $\sup_{y \in A} y \cdot x \geq 0$ для всех $x \in \mathbb{R}^m$, что, как показано выше, равносильно неравенству $\max_{1 \leq i \leq m}\{\, f_i'(\widehat{x}) \cdot x \,\} \geq 0$. Для каждого $1 \leq i \leq m$ по условию существует такой элемент $x_i^* \in (\mathbb{R}^n)^*$, что $f_i(x) - f_i(\widehat{x}) \geq x_i^* \cdot (x - \widehat{x})$ 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 $f(x) - f(\widehat{x}) = \max_{1 \leq i \leq m}\{\, f_i(x) - f_i(\widehat{x}) \,\} \geq \max_{1 \leq i \leq m}\{\, f_i'(\widehat{x}) \cdot (x - \widehat{x}) \,\} \geq 0$, т. е. $\widehat{x}$ — локальный минимум f. $\square$

6. Уравнение Эйлера в простейшей задаче вариационного исчисления и его интегралы. Необходимые условия экстремума в задаче Больца

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 об этом еще будет идти речь). 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, а также в задаче Больца, которая была поставлена лишь в 20 веке, но она близка к простейшей задаче и важна для понимания структуры необходимых условий экстремума в задачах вариационного исчисления с ограничениями.

Простейшая задача вариационного исчисления и задача Больца. 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

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

называется *простейшей задачей (классического) вариационного исчисления.*

Пусть, далее, W — открытое подмножество $\mathbb{R}^2$ и $l\colon W\to\mathbb{R}$ — непрерывная функция переменных ξ_0 и ξ_1. Задача

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

называется *задачей Больца.*

Уточним постановки. Обозначим через $C([t_0,t_1])$ и $C^1([t_0,t_1])$ множества всех непрерывных и непрерывно дифференцируемых функций $x(\cdot)$ на $[t_0, t_1]$. Это нормированные пространства соответственно с нормами $\|x(\cdot)\|_{C([t_0,t_1])} = \max_{t\in[t_0,t_1]} |x(t)|$ и $\|x(\cdot)\|_{C^1([t_0,t_1])} = \max(\|x(\cdot)\|_{C([t_0,t_1])}, \|\dot{x}(\cdot)\|_{C([t_0,t_1])})$.

Функция $x(\cdot) \in C^1([t_0, t_1])$ называется *допустимой в задаче* (P_6), если $\Gamma(x(\cdot)) = \{(t, x(t), \dot{x}(t)) \mid t \in [t_0, t_1]\} \subset G$ и $x(t_i) = x_i$, $i = 0, 1$.

Допустимая функция $\widehat{x}(\cdot)$ называется *слабым локальным минимумом (максимумом)* в задаче (P_6), если существует такое $\varepsilon > 0$, что для любой допустимой функции $x(\cdot)$, для которой $\|x(\cdot) - \widehat{x}(\cdot)\|_{C^1([t_0,t_1])} < \varepsilon$ выполняется неравенство $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$ $(J(x(\cdot)) \leq J(\widehat{x}(\cdot)))$. *Слабый локальный экстремум* — это либо слабый локальный минимум, либо слабый локальный максимум.

Для (P_6') аналогично определяются понятия допустимой функции и слабого локального экстремума.

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 $\dot{x}$, а также $\widehat{l}_{\xi_i} = l_{\xi_i}(\widehat{x}(t_0), \widehat{x}(t_1))$, $i = 0, 1$.

Теорема 6 (Необходимые условия экстремума в задачах (P_6) и (P_6')). (*a*) *Пусть $\widehat{x}(\cdot)$ — допустимая функция в задаче (P_6), функция L непрерывна вместе со своими частными производными по x и $\dot{x}$ в окрестности $\Gamma(\widehat{x}(\cdot))$. Тогда, если $\widehat{x}(\cdot)$ доставляет слабый локальный экстремум в (P_6), 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*

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

(*b*) *Пусть $\widehat{x}(\cdot)$ — допустимая функция в задаче (P_6'), функция L непрерывна вместе со своими частными производными по x и $\dot{x}$ в окрестности $\Gamma(\widehat{x}(\cdot)) = \{\,(t, \widehat{x}(t), \dot{\widehat{x}}(t)) \mid t \in [t_0, t_1]\,\}$, а функция l непрерывна вместе со своими частными производными по ξ_0 и ξ_1 в окрестности точки $(\widehat{x}(t_0), \widehat{x}(t_1))$. Тогда, если $\widehat{x}(\cdot)$ доставляет слабый локальный экстремум в (P_6'), 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. (*a*) 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 $\Gamma(\widehat{x}(\cdot))$ и, очевидно, что $x_\alpha(\cdot) \in C^1([t_0, t_1])$ и $x_\alpha(t_i) = x_i$, $i = 0, 1$. 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:

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

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

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

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

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

Выберем c так, чтобы $\int_{t_0}^{t_1}(-p(t) + \widehat{L}_{\dot{x}}(t) + c)\, dt = 0$. Тогда подставляя в (iii) функцию $x(t) = \int_{t_0}^{t}(-p(\tau) + \widehat{L}_{\dot{x}}(\tau) + c)\, d\tau$ (ясно, что $x(\cdot) \in C^1([t_0, t_1])$ и $x(t_0) = x(t_1) = 0$), получаем, что интеграл от квадрата непрерывной функции равен нулю и значит, $-p(\cdot) + \widehat{L}_{\dot{x}}(\cdot) + c = 0$. Отсюда следует, что $\widehat{L}_{\dot{x}}(\cdot) \in C^1([t_0, t_1])$. Дифференцируя равенство $p(\cdot) = \widehat{L}_{\dot{x}}(\cdot) + c$ и учитывая, что $\dot{p}(\cdot) = \widehat{L}_x(\cdot)$, получаем уравнение Эйлера.

(b) 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 выше, приходим к соотношению

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

Пусть $p(\cdot)$ — решение задачи Коши $\dot{p} = \widehat{L}_x(t)$, $p(t_1) = -\widehat{l}_{\xi_1}$ (т. е. $p(t) = -\widehat{l}_{\xi_1} + \int_t^{t_1} \widehat{L}_x(\tau)\, d\tau$, $t \in [t_0, t_1]$[3]). Подставляя это в (iv) и интегрируя по частям, получаем, что

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

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 (v), приходим к равенству

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

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

[3] Мы говорим о задаче Коши в этой простейшей ситуации, чтобы потом было видно, что эти же рассуждения применимы и при выводе необходимых условий в общей задаче — задаче Лагранжа вариационного исчисления.

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

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

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

который называется *интегралом импульса.*

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

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

Он называется *интегралом энергии.*

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

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

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

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

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

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. Уравнения Эйлера–Лагранжа для задачи Лагранжа. Следствия для изопериметрической задачи и задачи со старшими производными

1. 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

$$f_0(x(\cdot),u(\cdot))=\int_{t_0}^{t_1}L_0(t,x(t),u(t))\,dt+l_0(x(t_0),x(t_1))\to\min,$$
$$\dot{x}=\varphi(t,x,u),$$
$$f_i(x(\cdot),u(\cdot))=\int_{t_0}^{t_1}L_i(t,x(t),u(t))\,dt+l_i(x(t_0),x(t_1))\le 0,\ \ 1\le i\le m',$$
$$f_i(x(\cdot),u(\cdot))=\int_{t_0}^{t_1}L_i(t,x(t),u(t))\,dt+l_i(x(t_0),x(t_1))=0,$$
$$m'+1\le i\le m \quad (P_7)$$

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

Уточним постановку. Пара $(x(\cdot),u(\cdot))$ называется *допустимой в задаче* (P_7), если $(x(\cdot),u(\cdot))\in Z=C^1([t_0,t_1],\mathbb{R}^n)\times C([t_0,t_1],\mathbb{R}^r)$,[4] $\Gamma(x(\cdot),u(\cdot))=\{\,(t,x(t),u(t))\mid t\in[t_0,t_1]\,\}\subset G$, $(x(t_0),x(t_1))\in W$, $\dot{x}(t)=\varphi(t,x(t),u(t))$ для всех $t\in[t_0,t_1]$, $f_i(x(\cdot),u(\cdot))=0$, $1\le i\le m'$ 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_7), если существует такое $\varepsilon>0$, что для любой допустимой пары $(x(\cdot),u(\cdot))$, для которой $\|(x(\cdot),u(\cdot))-(\widehat{x}(\cdot),\widehat{u}(\cdot))\|_Z<\varepsilon$ выполнено неравенство $J(x(\cdot),u(\cdot))\ge J(\widehat{x}(\cdot),\widehat{u}(\cdot))$.

Функция Лагранжа задачи (P_7) 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 $L(t,x,u)=\sum_{i=0}^m\lambda_iL_i(t,x,u)+p(t)\cdot(\dot{x}-\varphi(t,x,u))$, $l(\xi_0,\xi_1)=\sum_{i=0}^m\lambda_il_i(\xi_0,\xi_1)$ и вектор $\overline{\lambda}=(\lambda_0,\lambda_1,\ldots,\lambda_m,p(\cdot))\in(\mathbb{R}^{m+1})^*\times C^1([t_0,t_1],(\mathbb{R}^n)^*)$ — набор множителей Лагранжа.

---

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

Отметим, что функция Лагранжа составлена, следуя (формально) рекомендациям Лагранжа: дифференциальную связь воспринимаем как континуум равенств, каждое из которых надо умножить на неопределенный множитель $p(t)$, а затем "проссумировать", т. е. проинтегрировать.

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 7 (Необходимые условия экстремума в задаче (P_7)). *Пусть $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — допустимый элемент в задаче (P_7), функции L_i, $0 \le i \le m$, и отображение φ непрерывны вместе со своими частными производными по x и u в окрестности множества $\Gamma(x(\cdot), u(\cdot)) = \{\, (t, \widehat{x}(t), \widehat{u}(t)) \mid t \in [t_0, t_1] \,\}$, а функции l_i, $0 \le i \le m$, непрерывно дифференцируемы в окрестности точки $(\widehat{x}(t_0), \widehat{x}(t_1))$. Тогда, если $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — слабый локальный минимум в (P_7), то найдется ненулевой набор множителей Лагранжа $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m, p(\cdot)) \in (\mathbb{R}^{m+1})^* \times C^1([t_0, t_1], (\mathbb{R}^n)^*)$ 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) $\lambda_i \ge 0$, $i = 0, 1, \ldots m'$ *(условия неотрицательности);*
(c) $\lambda_i f_i(\widehat{z}(\cdot)) = 0$, $i = 1, \ldots, m'$ *(условия дополняющей нежесткости);*
(d) $\widehat{L}_{\dot{x}}(t_i) = (-1)^i \widehat{l}_{\xi_i}$, $i = 0, 1$ *(условия трансверсальности).*

2. Задача Лагранжа. Упрощенный вариант. Теорему 7 докажем для следующего частного случая общей постановки:

$$J(x(\cdot), u(\cdot)) = \int_{t_0}^{t_1} f(t, x(t), u(t))\, dt \to \min, \quad \dot{x} = \varphi(t, x, u),$$
$$x(t_0) = x_0, \quad x(t_1) = x_1. \quad (P_7')$$

При этом следует заметить, что доказательство в общем случае в существенном не отличается от доказательства необходимых условий для задачи (P_7').

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 задаче (P_6), то это равносильно тому, что $(\widehat{x}(\cdot), \widehat{u}(\cdot))$, где $\widehat{u}(\cdot) = \dot{\widehat{x}}(\cdot)$, — слабый экстремум в той же задаче, но записанной в форме задачи (P_7').

Функция Лагранжа задачи (P_7') имеет вид

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

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 7’ (Необходимые условия экстремума в задаче (P_7')). *Пусть пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ допустима в задаче (P_7'), функция f и отображение φ непрерывно вместе со своими частными производными по x и u в окрестности $\Gamma(\widehat{x}(\cdot), \widehat{u}(\cdot))$. Тогда, если пара $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ доставляет слабый локальный минимум в (P_7'), 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 единственное решение $x(\cdot, \alpha)$ данной задачи Коши на отрезке $[t_0, t_1]$ 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 $7'$. 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: $\dot x = \varphi(t,x,u_\alpha(t,u(\cdot))$, $x(t_0)=x_0$. 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: $\dot x = \varphi(t,x,u_{\overline{\alpha}}(t,\overline{u}(\cdot)))$, $x(t_0)=x_0$ 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_7') 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 $y_0\notin L$. Тогда по второй теореме отделимости (теорема 1.2) 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

$$0=\int_{t_0}^{t_1}(\dot{p}(t)\cdot y(t,u(\cdot))+p(t)\cdot\dot{y}(t,u(\cdot))-p(t)\cdot\widehat{\varphi}_u(t)u(t)+\lambda_0\widehat{f}_u(t)\cdot u(t))\,dt$$

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

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

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.

2. 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 $t, x, \dot{x}, \ldots, x^{(n)}$) и $x_i^{(k)} \in \mathbb{R}$, $k = 0, 1, \ldots, n-1$, $i = 0, 1$. Задача

$$J(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t), \ldots, x^{(n)}(t))\, dt \to \min,$$
$$x^{(k)}(t_i) = x_i^{(k)}, \quad k = 0, 1, \ldots, n-1, \quad i = 0, 1, \quad (P_7'')$$

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

Функцию $x(\cdot) \in C^n([t_0, t_1])$ называется *допустимой в задаче* (P_7''), если $\Gamma(x(\cdot)) = \{(t, x(t), \dot{x}(t), \ldots, x^{(n)}(t)) \mid t \in [t_0, t_1]\} \subset G$ и $x^{(k)}(t_i) = x_i^{(k)}$, $k = 0, 1, \ldots, n-1$, $i = 0, 1$.

Допустимая функция $\widehat{x}(\cdot)$ называется *локальным минимумом в задаче* (P_7'')*, 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* $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$.

Следствие 1 *(Необходимые условия минимума в задаче* (P_7'')*. Уравнение Эйлера–Пуассона). Пусть* $\widehat{x}(\cdot)$ *— допустимая функция в задаче* (P_7'') *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_7'')*, 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_7'') можно записать как задачу Лагранжа

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

Простая проверка показывает, что если $\widehat{x}(\cdot)$ — локальный минимум в задаче (P_7''), 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

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

задачи (i) выполнены соотношения: $-\dot{p}_1(t) + \lambda_0 \widehat{L}_{x_1}(t) = 0$, $-\dot{p}_2(t) + \lambda_0 \widehat{L}_{x_2}(t) - p_1(t) = 0, \dots,$ $-\dot{p}_n(t) + \lambda_0 \widehat{L}_{x_n}(t) - p_{n-1}(t) = 0$ 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: $x_1 = x, x_2 = \dot{x}, \dots, x_n = x^{(n-1)}, u = x^{(n)}$), 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. □

Отметим, что в функцию Лагранжа не включены граничные условия, так как соответствующие условия трансверсальности (говорящие о том, что функция $p(\cdot)$ в точках t_0 и t_1 какие-то значения) не информативны.

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

$$J(x(\cdot)) = \int_{t_0}^{t_1} f_0(t, x(t), \dot{x}(t))\, dt \to \min, \quad \int_{t_0}^{t_1} f_i(t, x(t), \dot{x}(t))\, dt = \alpha_i,$$
$$1 \le i \le m, \quad x(t_0) = x_0, \quad x(t_1) = x_1 \quad (P_7''')$$

называется *изопериметрической задачей*. Ясно, что это частный случай задачи (P_7) (дифференциальная связь: $\dot{x} = u$) и можно было бы сослаться на теорему 7. Но мы доказали только теорему $7'$ и поэтому будем опираться на нее.

Функция $x(\cdot)$ называется *допустимой в задаче* (P_7'''), 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_7'''), 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_7''') (в соответствии с функцией Лагранжа задачи (P_7)) 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 $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m)$. Здесь мы снова не включили граничные условия вследствие их неинформативности.

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

Следствие 2 (Необходимые условия экстремума в задаче (P_7''')). *Пусть $\widehat{x}(\cdot)$ — допустимая функция в задаче (P_7'''), функции f_i, $i = 0, 1, \ldots, m$, непрерывны вместе со своими частными производными по x и $\dot{x}$ в окрестности $\Gamma(\widehat{x}(\cdot))$. Тогда, если $\widehat{x}(\cdot)$ — локальный минимум в (P_7'''), 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

$$\int_{t_0}^{t_1} f_0(t, x(t), u(t))\, dt \to \min, \quad \dot{x} = u, \quad \dot{y}_1 = f_1(t, x, u), \ldots, \dot{y}_m = f_m(t, x, u),$$
$$x(t_0) = x_0, \quad x(t_1) = x_1, \quad y_i(t_0) = 0, \quad y_i(t_1) = \alpha_i, \quad i = 1, \ldots, m. \quad (i)$$

Несложная проверка показывает, что если $\widehat{x}(\cdot)$ — локальный экстремум в задаче (P_7'''), 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 $7'$ 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 $p_i(\cdot) = 0$, $i = 1, \ldots, n$. □

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, ... механика ....

8. Принцип максимума Понтрягина для понтрягинской 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, "чекагской школе" (в Чекагском университете работали Блисс, Больца, МакШеин и др.) велась достаточно плодотворная работа по развитию вариационного исчисления. Итоги этой деятельности были подведены Блиссом в его монографии 1942 г. Казалось, что построение теории экстремума (как она представлялась в те времена) завершено. Но как уже упоминалось, 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 необходимые условия минимума в ней.

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 $t\in\mathbb{R}$, $x=(x_1,\ldots,x_n)^T\in\mathbb{R}^n$ и $u=(u_1,\ldots,u_r)^T\in\mathbb{R}^r$) непрерывны на $G\times U$ и $x_i\in\mathbb{R}^n$, $i=0,1$. Задача

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

называется *задачей оптимального управления* (в понтрягинской форме). 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_8), 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 $x(t_i)=x_i$, $i=0,1$.

Допустимая пара $(\widehat x(\cdot),\widehat u(\cdot))$ называется *сильным (локальным) минимумом в задаче* (P_8), если существует такое $\varepsilon>0$, что для любой допустимой пары $(x(\cdot), u(\cdot))$, для которой $\|x(\cdot) - \widehat{x}(\cdot)\|_{C([t_0,t_1],\mathbb{R}^n)} < \varepsilon$ выполнено неравенство $J(x(\cdot), u(\cdot)) \geq J(\widehat{x}(\cdot), \widehat{u}(\cdot))$.

Функция Лагранжа для задачи (P_8) имеет тот же вид, что и для задачи (P_7').

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_8).

Ниже пользуемся теми же соглашениями об обозначениях, о которых сказано перед формулировкой теоремы о необходимых условиях экстремума в задаче (P_7).

Теорема 8 (Принцип максимума Понтрягина). *Пусть $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — допустимая пара в задаче (P_8), функция f и отображение φ непрерывны вместе со своими частными производными по x в окрестности $\Gamma(\widehat{x}(\cdot), \widehat{u}(\cdot))$. Тогда, если $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — сильный минимум в (P_8), 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 $x(\cdot)$:*

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

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

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

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

Замечание. Обратим внимание на сходство условий минимума в теореме 5 (выпуклые задачи) и в данной теореме. Причина этого — “скрытая выпуклость”, которая сопутствует интегрированию. Подробнее об этом будет идти речь в дальнейшем.

Доказательство этой теоремы проходит по той же схеме, что и доказательство теоремы из предыдущего параграфа. Вместо теоремы о дифференцируемой зависимости решений дифференциальных уравнений от параметров понадобится ее аналог — так называемая теорема о пакете иголок, которая есть следствие стандартных теорем о дифференцируемой зависимости решений дифференциальных уравнений от начальных данных. Кроме того, нужна некая модификация теоремы о разрешимости системы нелинейных уравнений для отображений, которые определены только на конусе неотрицательных векторах. Приведем здесь эти результаты без доказательства.

Пусть U — окрестность точки $\widehat{x} \in \mathbb{R}^n$ и $F\colon U \cap (\widehat{x} + \mathbb{R}^n_+) \to \mathbb{R}^m$. Понятия дифференцируемости и строгой дифференцируемости отображения F, приведенные в §2, фактически, дословно переносятся на эту ситуацию. Производную F в точке $\widehat{x}$ (определяемую однозначно) будем обозначать $F'(\widehat{x}+0)$. Например, если $n = m = 1$, то $F'(\widehat{x}+0)$ — производная в $\widehat{x}$ справа, а если F задается m функциями, то $F'(\widehat{x}+0)$ — матрица Якоби частных производных этих функций, вычисленных в $\widehat{x}$ справа.

Лемма (о пакете иголок)**.** *Пусть $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — допустимая пара в задаче (P_8), $\mathcal{N}$ — совокупность пар $(\tau_i, v_i) \in [t_0, t_1] \times U$, $1 \le i \le k$, где $t_0 < \tau_1 < \ldots < \tau_k < t_1$ и точки τ_i принадлежат T — множеству точек непрерывности $\widehat{u}(\cdot)$, $\overline{\alpha} = (\alpha_1, \ldots, \alpha_k)^T \in \mathbb{R}^k_+$, причем α_i столь малы, что отрезки $[\tau_i - \alpha_i, \tau_i]$, $1 \le i \le k$, не пересекаются и принадлежат T. Положим*

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

Тогда найдется такая окрестность нуля V в $\mathbb{R}^k$, что для любого $\overline{\alpha} \in V \cap \mathbb{R}^k_+$ существует единственное решение $x(\cdot, \overline{\alpha}, \mathcal{N})$ задачи Коши: $\dot{x} = \varphi(t, x, u(t, \overline{\alpha}, \mathcal{N}))$, $x(t_0) = x_0$, определенное на всем отрезке $[t_0, t_1]$. При этом, $x(\cdot, \overline{\alpha}, \mathcal{N}) \to \widehat{x}(\cdot)$ при $\overline{\alpha} \to 0$ равномерно на $[t_0, t_1]$, частная производная по α_i, $1 \le i \le k$, отображения $\overline{\alpha} \mapsto x(t, \overline{\alpha}, \mathcal{N})$ при $t \in [\tau_i, t_1]$ непрерывно дифференцируема для достаточно малых $\overline{\alpha} \in \mathbb{R}^k_+$ и в нуле на $[\tau_i, t_1]$ удовлетворяет уравнению

$$\dot{y} = \widehat{\varphi}_x(t) y, \quad y(\tau_i) = \varphi(\tau_i, \widehat{x}(\tau_i), v_i) - \varphi(\tau_i, \widehat{x}(\tau_i), \widehat{u}(\tau_i)). \tag{i}$$

Функцию $u(\cdot, \overline{\alpha}, \mathcal{N})$ называют *игольчатой вариацией* $\widehat{u}(\cdot)$, а сам набор $\mathcal{N}$ — *пакетом иголок.* Если пакет состоит из одной иголки (соответствующей $(\tau, v) \in T \times U$), то игольчатую вариацию $\widehat{u}(\cdot)$ записываем так $u(\cdot, \alpha, (\tau, v))$.

Теперь докажем теорему о разрешимости системы нелинейных уравнений к отображениям, заданным на конусе неотрицательных векторов.

Теорема (о разрешимости системы нелинейных уравнений для конуса)**.** *Пусть V — окрестность точки $\widehat{x}$ в $\mathbb{R}^n$, отображение $F\colon V \cap (\widehat{x} + \mathbb{R}^n_+) \to \mathbb{R}^m$ строго дифференцируемо в $\widehat{x}$ и $F'(\widehat{x}+0)(\mathbb{R}^n_+) = \mathbb{R}^m$. Тогда существуют окрестность W точки $F(\widehat{x})$ и константа $K > 0$ такие, что для каждого $y \in W$ найдется элемент $x(y) \in V \cap (\widehat{x} + \mathbb{R}^n_+)$ такой, что $F(x(y)) = y$ и $|x(y) - \widehat{x}| \le K|y - F(\widehat{x})|$.*

Отметим, что $\operatorname{cone} A$ обозначает выпуклую коническую оболочку множества $A \subset \mathbb{R}^n$, т. е. наименьший выпуклый конус, содержащий A. Как нетрудно проверить, $\operatorname{cone} A$ состоит из элементов вида $\alpha_1 x_1 + \ldots + \alpha_m x_m$, где $x_i \in A$ и $\alpha_i > 0$, $i = 1, \ldots, m$.

Доказательство теоремы 8. Пусть $(\tau, v) \in T \times U$. Согласно лемме о пакете иголок для всех достаточно малых $\alpha \ge 0$ существует единственное решение $x(\cdot, \alpha, (\tau, v))$ задачи Коши: $\dot{x} = \varphi(t, x, u(t, \alpha, (\tau, v))$, $x(t_0) = x_0$. Если $t \in [\tau, t_1]$, то производную отображения $\alpha \mapsto x(t, \alpha, (\tau, v))$ в нуле обозначим $y(t, \tau, v)$.

Покажем, что функция $\alpha \mapsto J(x(\cdot, \alpha, (\tau, v)), u(\cdot, \alpha, (\tau, v)))$ для достаточно малых $\alpha \ge 0$ непрерывно дифференцируема и ее производная в нуле, которую обозначим $y_0(\tau, v)$, имеет вид

$$y_0(\tau, v) = \Delta_{\tau v} f + \int_{\tau}^{t_1} \widehat{f}_x(t) \cdot y(t, \tau, v)\, dt, \qquad (ii)$$

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).

Действительно, рассмотрим задачу Коши: $\dot{z} = \psi(t, z, u(t, \alpha, (\tau, v)), z(t_0) = z_0$, где $z = (x, y)^T \in \mathbb{R}^n \times \mathbb{R}$, $\psi(t, z, u) = (\varphi(t, x, u), f(t, x, u))^T$ и $z(t_0) = (x(t_0), 0)^T$. Решение $\widetilde{z}(\cdot, \alpha, (\tau, v)) = (\widetilde{x}(\cdot, \alpha, (\tau, v)), \widetilde{y}(\cdot, \alpha, (\tau, v))^T$ этой задачи существует согласно лемме о пакете иголок и если $t \in [\tau, t_1]$, то отображение $\alpha \mapsto \widetilde{z}(t, \alpha, (\tau, v))$ непрерывно дифференцируемо для всех достаточно малых $\alpha \geq 0$. Далее, в силу единственности, $\widetilde{x}(\cdot, \alpha, (\tau, v)) = x(\cdot, \alpha, (\tau, v))$ и поскольку $y(t_1, \alpha, (\tau, v))) = \int_{t_0}^{t_1} f(t, x(t, \alpha, (\tau, v)), u(t, \alpha, (\tau, v)))dt = J(x(\cdot, \alpha, (\tau, v)), u(\cdot, \alpha, (\tau, v)))$, то требуемая непрерывная дифференцируемость доказана.

Докажем (ii). Обозначим $J(\alpha) = J(x(\cdot, \alpha, (\tau, v)), u(\cdot, \alpha, (\tau, v)))$. Тогда

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

Второе слагаемое справа — производная в нуле интеграла по параметру, к которому применима стандартная теорема о дифференцируемости под знаком интеграла, ибо функция $\alpha \mapsto x(t, \alpha, (\tau, v))$ 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

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

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

Покажем теперь, что предположение о том, что $\operatorname{cone} \{\, (y_0(\tau, v), y(t_1, \tau, v)) \in \mathbb{R}^{n+1} \mid (\tau, v) \in T \times U \,\} = \mathbb{R}^{n+1}$ приводит к противоречию с тем, что $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — локальный минимум. Действительно, пусть данное предположение верно. Тогда существуют такие $(\tau_i, v_i) \in T \times U$, $\tau_1 \leq \ldots \leq \tau_k$, что выпуклая коническая оболочка векторов $(y_0(\tau_i, v_i), y(t_1, \tau_i, v_i))$, $i = 1, \ldots, k$, совпадает с $\mathbb{R}^{n+1}$. Легко видеть, что функции $\tau \mapsto y(t_1, u_i(\cdot))$ и $\tau \mapsto y_0(u_i(\cdot))$ непрерывны в некоторой окрестности точки τ_i, $i = 1, \ldots, k$, и поэтому можно считать, что $\tau_1 < \ldots < \tau_k$. Положим $\mathcal{N} = \{\, (\tau_i, v_i),\ 1 \leq i \leq k \,\}$ и определим (как в лемме) $u(\cdot, \overline{\alpha}, \mathcal{N})$ для достаточно малых $\overline{\alpha} \in \mathbb{R}^k_+$. По теореме о пакете иголок для малых $\overline{\alpha} \in \mathbb{R}^k_+$ существует решение $x(\cdot, \overline{\alpha}, \mathcal{N})$ задачи Коши: $\dot{x} = \varphi(t, x, u(t, \overline{\alpha}, \mathcal{N}))$, $x(t_0) = x_0$. Далее, как и для одной иголки, доказывается что функция $\overline{\alpha} \mapsto J(x(\cdot, \overline{\alpha}, \mathcal{N}), u(\cdot, \overline{\alpha}, \mathcal{N}))$ непрерывно дифференцируема для малых $\overline{\alpha} \in \mathbb{R}^k_+$.

Таким образом, существует такая окрестность нуля V в $\mathbb{R}^k$, что корректно определено отображение $F\colon V \cap \mathbb{R}^k_+ \to \mathbb{R}^{n+1}$, сопоставляющее $\overline{\alpha} \in V \cap \mathbb{R}^k_+$ вектор $F(\overline{\alpha}) = (J(x(\cdot, \overline{\alpha}, \mathcal{N}), u(\cdot, \overline{\alpha}, \mathcal{N})), x(t_1, \overline{\alpha}, \mathcal{N}))$. Это отображение непрерывно дифференцируемо на $V \cap \mathbb{R}^k_+$ и значит, строго дифференцируемо в нуле. Поскольку любой элемент из $\mathbb{R}^{n+1}$ представляется как линейная комбинация с неотрицательными множителями векторов $(y_0(\tau_i, v_i), y(t_1, \tau_i, v_i))$, $i = 1, \ldots, k$ (образующих столбцы матрицы $F'(0+0)$)), то $F'(0+0)(\mathbb{R}^k_+) = \mathbb{R}^{n+1}$.

Применим теперь теорему о разрешимости системы нелинейных уравнений для конуса и придем (рассуждая точно также как в задаче 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 по $\nu>0$ вектор $y_\nu = (J(\widehat{x}(\cdot),\widehat{u}(\cdot))-\nu,x_1)$ принадлежит соответствующей окрестности $F(0)$ (окрестность W в теореме о разрешимости). Тогда существует вектор $\overline{\alpha}(y_\nu)\in V\cap\mathbb{R}^k_+$ такой, что $F(\overline{\alpha}(y_\nu))=y_\nu$, или $J(x(\cdot,\overline{\alpha}(y_\nu),\mathcal{N}),u(\cdot,\overline{\alpha}(y_\nu),\mathcal{N})=J(\widehat{x}(\cdot),\widehat{u}(\cdot))-\nu$, $x(t_1,\overline{\alpha}(y_\nu),\mathcal{N})=x_1$ и, кроме того, $|\overline{\alpha}(y_\nu)|\le K\nu$. Из этих соотношений следует (в силу того, что $x(\cdot,\overline{\alpha},\mathcal{N})\to\widehat{x}(\cdot)$ в $C([t_0,t_1],\mathbb{R}^n)$ при $\overline{\alpha}\to 0$), что в любой окрестности $\widehat{x}(\cdot)$ найдется функция $x(\cdot,\overline{\alpha}(y_\nu),\mathcal{N})$ такая, что пара $(x(\cdot,\overline{\alpha}(y_\nu),\mathcal{N}),u(\cdot,\overline{\alpha}(y_\nu),\mathcal{N}))$ допустима в задаче (P_8), а значение функционала на ней меньше, чем на $(\widehat{x}(\cdot),\widehat{u}(\cdot))$, в противоречие с тем, что $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ — локальный минимум.

Итак, конус $C=\operatorname{cone}\{\,(y_0(\tau,v),y(t_1,\tau,v))\in\mathbb{R}^{n+1}\mid(\tau,v)\in T\times U\,\}$ не совпадает с $\mathbb{R}^{n+1}$. Пусть $(z_0,z)\in(\mathbb{R}\times\mathbb{R}^n)\setminus C$. Тогда по первой теореме отделимости (теореме 1.1) найдется ненулевой вектор $(\lambda_0,\lambda)\in(\mathbb{R}^{n+1})^*$ такой, что $\lambda_0z_0+\lambda\cdot z\le\inf_{(y_0(\tau,v),y(t_1,\tau,v))\in C}(\lambda_0y_0(\tau,v)+\lambda\cdot y(t_1,\tau,v))$. Отсюда легко следует, что $\lambda_0y_0(\tau,v)+\lambda\cdot y(t_1,\tau,v)\ge 0$ для всех $(y_0(\tau,v),y(t_1,\tau,v))\in C$. Это равносильно тому (с учетом выражения для $y_0(\tau,v)$ из (ii)), что для любой пары $(\tau,v)\in T\times U$ справедливо неравенство

$$\lambda_0\left(\Delta_{\tau v}f+\int_\tau^{t_1}\widehat{f}_x(t)\cdot y(t,\tau,v)\,dt\right)+\lambda\cdot y(t_1,\tau,v)\ge 0. \qquad (iii)$$

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 условие стационарности по x.

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

$$\begin{aligned}0\le\lambda_0\Delta_{\tau v}f+\int_\tau^{t_1}(\dot{p}(t)\cdot y(t,\tau,v)+p(t)\cdot\dot{y}(t,\tau,v)\,dt+\lambda\cdot y(t_1,\tau,v)=\\=\lambda_0\Delta_{\tau v}f+p(t)\cdot y(t,\tau,v)|_\tau^{t_1}+\lambda\cdot y(t_1,\tau,v)=\lambda_0(f(\tau,\widehat{x}(\tau),v)-f(\tau,\widehat{x}(\tau),\widehat{u}(\tau)))-\\-p(\tau)(\varphi(\tau,\widehat{x}(\tau),v)-\varphi(\tau,\widehat{x}(\tau),\widehat{u}(\tau))),\end{aligned}$$

или $p(\tau)\varphi(\tau,\widehat{x}(\tau),v)-\lambda_0f(\tau,\widehat{x}(\tau),v)\le p(\tau)\varphi(\tau,\widehat{x}(\tau),\widehat{u}(\tau))-\lambda_0f(\tau,\widehat{x}(\tau),\widehat{u}(\tau))$. Поскольку это верно для любой пары $(\tau,v)\in T\times U$, то условие максимума доказано. $\square$

Исторический комментарий. В пятидесятые годы, откликаясь на запросы практики, связанные с вопросами управления различными процессами, Л. С. Понтрягин организовал семинар, основными участниками которого были В. Г. Болтянский, Р. В. Гамкрелидзе и Е. Ф. Мищенко. На этом семинаре выступали многие известные специалисты по теории управления: Айзерман, Лернер, Розоноэр, Фельдбаум и др. В итоге Л. С. Понтрягин поставил общую задачу оптимального управления (близкую к той, которая рассмотрена выше). Необходимые условия минимума в ней доказал В. Г. Болтянский.

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.

Рассмотрим векторный вариант простейшей задачи вариационного исчисления

$$J(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))dt \to \mathrm{extr}, \quad x(t_i) = x_i, \quad i = 0, 1, \quad (P_9)$$

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

Напомним, что функция $x(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n)$ называется *допустимой в задаче* (P_9), 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 (максимумом)* в задаче (P_9), 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 $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$ $(J(x(\cdot)) \leq J(\widehat{x}(\cdot)))$.

Слабый локальный экстремум — это слабый локальный минимум, либо слабый локальный максимум.

В вариационном исчислении, наряду со слабым экстремумом, рассматривают еще и сильный локальный экстремум, где близость функций измеряется в пространстве $C([t_0, t_1], \mathbb{R}^n)$. Точнее говоря, обозначим через $PC^1([t_0, t_1], \mathbb{R}^n)$ пространство кусочно-непрерывно-дифференцируемых вектор-функций на $[t_0, t_1]$. Допустимая функция $x(\cdot) \in PC^1([t_0, t_1], \mathbb{R}^n)$ в задаче (P_9) определяется аналогично предыдущему. Скажем, что допустимая функция $\widehat{x}(\cdot)$ доставляет *сильный локальный минимум (максимум)* в задаче (P_9), если существует такое $\varepsilon > 0$, что для любой допустимой функции $x(\cdot)$, для которой $\|x(\cdot) - \widehat{x}(\cdot)\|_{C([t_0,t_1],\mathbb{R}^n)} < \varepsilon$ выполняется неравенство $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$ $(J(x(\cdot)) \leq J(\widehat{x}(\cdot)))$.

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

Заметим, что если функция $\widehat{x}(\cdot)$ доставляет сильный экстремум в задаче (P_9) и при этом $\widehat{x}(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n)$, то $\widehat{x}(\cdot)$ является и слабым экстремумом в этой задаче. Действительно, пусть $\varepsilon > 0$ такое, что как только $x(\cdot) \in PC^1([t_0, t_1], \mathbb{R}^n)$ и $\|x(\cdot) - \widehat{x}(\cdot)\|_{C([t_0,t_1],\mathbb{R}^n)} < \varepsilon$, то, скажем, $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$. Если теперь $x(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n)$ и $\|x(\cdot) - \widehat{x}(\cdot)\|_{C^1([t_0,t_1],\mathbb{R}^n)} < \varepsilon$, то так как, в частности, $x(\cdot) \in PC^1([t_0, t_1], \mathbb{R}^n)$ и $\|x(\cdot) - \widehat{x}(\cdot)\|_{C([t_0,t_1],\mathbb{R}^n)} \leq \|x(\cdot) - \widehat{x}(\cdot)\|_{C^1([t_0,t_1],\mathbb{R}^n)} < \varepsilon$, получаем, что $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$.

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

Теперь сформулируем условия Лежандра, Якоби и Вейерштрасса. Пусть $\widehat{x}(\cdot)$ — *экстремаль* (или *стационарная точка*) задачи (P_9), 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 $t \in [t_0, t_1]$.[5]

Пусть, далее, $h(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n)$, $h(t_0) = h(t_1) = 0$. 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

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

и дифференцируя еще раз, получаем, что

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

($\dot{h}^T(t)\widehat{L}_{x\dot{x}}(t)h(t) = h^T(t)\widehat{L}_{\dot{x}x}(t)\dot{h}(t)$ в силу того, что $\widehat{L}_{\dot{x}x}(t) = (\widehat{L}_{x\dot{x}}(t))^T$).

Этот функционал (как функцию от $h(\cdot)$) обозначим через $Q(h(\cdot))$ и рассмотрим задачу

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

Уравнение Эйлера для данной задачи имеет вид

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

Это уравнение называется *уравнением Якоби* исходной задачи (P_9).

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 $h(t_0) = h(\tau) = 0$.

[5]Запись $\widehat{L}_{\dot{x}\dot{x}}(t) \geq 0$ и $\widehat{L}_{\dot{x}\dot{x}}(t) > 0$, как и раньше, означает соответственно неотрицательную и положительную определенность симметричной матрицы $\widehat{L}_{\dot{x}\dot{x}}(t)$.

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

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

Пусть L — интегрант в задаче (P_9). Если L — дифференцируемая функция на некотором открытом множестве $G \times \mathbb{R}^n$, где G — открытое подмножество $\mathbb{R} \times \mathbb{R}^n$, то функция $\mathcal{E}(t, x, \dot{x}, u) = L(t, x, u) - L(t, x, \dot{x}) - L_{\dot{x}}(t, x, \dot{x}) \cdot (u - \dot{x})$, определенная на $G \times \mathbb{R}^n$, 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 $u \in \mathbb{R}^n$ и $t \in [t_0, t_1]$.

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 $u \in \mathbb{R}^n$ и $t \in [t_0, t_1]$.

При доказательстве необходимых условий в задаче (P_9) нам понадобится один несложный технический результат, который приводим без доказательства (см. ...)

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

$$\begin{aligned} &\inf\{\, J(x(\cdot)) \mid x(\cdot) \in PC^1([t_0, t_1], \mathbb{R}^n),\ x(t_0) = x_0,\ x(t_1) = x_1 \,\} = \\ &\quad = \inf\{\, J(x(\cdot)) \mid x(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n),\ x(t_0) = x_0,\ x(t_1) = x_1 \,\}. \end{aligned}$$

Следующая теорема — основной результат данного параграфа.

Теорема 9 (Необходимые условия минимума в простейшей задаче вариационного исчисления). *A) Пусть в задаче* (P_9) *интегрант* L *непрерывно дифференцируем на открытом подмножестве* $G \times \mathbb{R}^n$, *где* G *— открытое подмножество* $\mathbb{R} \times \mathbb{R}^n$. *Если допустимая функция* $\widehat{x}(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n)$ *доставляет сильный локальный минимум в* (P_9), *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*) *для всех* $t \in [t_0, t_1]$ *и* $u \in \mathbb{R}^n$ *выполнено условие Вейерштрасса*

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

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

B) *Пусть в задаче* (P_9) *интегрант* L *дважды непрерывно дифференцируем на открытом подмножестве* $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n$. *Если допустимая функция* $\widehat{x}(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n)$ *доставляет слабый локальный минимум в* (P_9), *то для всех* $t \in [t_0, t_1]$ *выполнено уравнение Эйлера, условие Лежандра и условие Якоби.*

Доказательство. *A*) Запишем задачу (P_9) как задачу оптимального управления:

$$\int_{t_0}^{t_1} L(t, x(t), u(t))\, dt \to \min, \quad \dot{x} = u, \quad x(t_0) = x_0, \ \ x(t_1) = x_1. \qquad (P_9')$$

Легко видеть, что $\widehat{x}(\cdot)$ доставляет сильный минимум в (P_9) 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_9').

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

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

и условие минимума по $u(\cdot)$:

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

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 $f\colon u \to L(t, \widehat{x}(t), u) - p(t) \cdot u$ на $\mathbb{R}^n$ 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 $L(t, \widehat{x}(t), u) - L_{\dot{x}}(t, \widehat{x}(t), \dot{\widehat{x}}(t)) \cdot u \geq L(t, \widehat{x}(t), \dot{\widehat{x}}(t)) - L_{\dot{x}}(t, \widehat{x}(t), \dot{\widehat{x}}(t)) \cdot \dot{\widehat{x}}(t)$ для всех $u \in \mathbb{R}^n$ и $t \in [t_0, t_1]$ или, что тоже $L(t, \widehat{x}(t), u) - L(t, \widehat{x}(t), \dot{\widehat{x}}(t)) - L_{\dot{x}}(t, \widehat{x}(t), \dot{\widehat{x}}(t)) \cdot (u - \dot{\widehat{x}}(t)) \geq 0$, т. е. выполнено условие Вейерштрасса.

B) 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: $\varphi''(0) \geq 0$. Чтобы не усложнять запись, ниже считаем $n = 1$. 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

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

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

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

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

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

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

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

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

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

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

и условие минимума по $u(\cdot)$:

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

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

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

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

По определению $\widetilde{h}(t) = 0$, если $t \geq \tau$ и поэтому из (vi) вытекает, что $p(\tau+0) = 0$. Но функция $p(\cdot)$ непрерывна и поэтому (снова из (vi)) 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. $\square$

10. Теория поля и достаточные условия сильного экстремума в простейшей задаче вариационного исчисления

В том параграфе рассматривается фрагмент теории поля и достаточных условий экстремума в вариационном исчислении на примере простейшей задачи.

Пусть $\{x(\cdot,\lambda)\}$ — семейство экстремалей простейшей задачи вариационного исчисления (P_9) (т. е., напомним, для функций $x(\cdot,\lambda)\in C^1([t_0,t_1],\mathbb{R}^n)$ выполняется уравнение Эйлера) и параметр λ принадлежит некоторому открытому множеству в $\mathbb{R}^n$.

Пусть $\widehat{x}(\cdot)$ — фиксированный элемент данного семейства экстремалей. Будем говорить, что функция $\widehat{x}(\cdot)$ *окружена полем экстремалей* $\{x(\cdot,\lambda)\}$, если существует открытое множество G, содержащее множество график $\{\,(t,\widehat{x}(t))\in\mathbb{R}^{n+1}\mid t\in[t_0,t_1]\,\}$, что для любой точки $(\tau,\xi)\in G$ существует единственная экстремаль семейства, проходящая через эту точку, т. е. существует единственное $\lambda=\lambda(\tau,\xi)$ такое, что $x(\tau,\lambda(\tau,\xi))=\xi$. При этом, для каждого $t\in[t_0,t_1]$ отображение $\lambda\mapsto x(t,\lambda)$ и отображение $(\tau,\xi)\mapsto\lambda(\tau,\xi)$ непрерывно дифференцируемы.

Функция $u\colon G\to\mathbb{R}^n$, $u(\tau,\xi)=\dfrac{d}{dt}x(t,\lambda(\tau,\xi))|_{t=\tau}$, называется *функцией наклона поля.*

Если существует такая точка (t_*,x_*), что $x(t_*,\lambda)=x_*$ для всех λ, то семейство $\{x(\cdot,\lambda)\}$ называется *центральным полем экстремалей* (с центром в (t_*,x_*)).

Теорема (Достаточные условия сильного минимума в простейшей задаче)**.** *Пусть $\widehat{x}(\cdot)$ — допустимая экстремаль в простейшей задаче (P_9) окружена в области G центральным полем экстремалей с центром в точке $(t_*,\widehat{x}(t_*))$, $t_*<t_0$. Интегрант L непрерывно дифференцируем на $G\times\mathbb{R}^n$ и функция $\dot{x}\mapsto L(t,x,\dot{x})$ выпукла на $\mathbb{R}^n$. Тогда $\widehat{x}(\cdot)$ — сильный локальный минимум в задаче (P_9).*

Доказательство. Снова, для простоты, считаем, что $n=1$. Рассмотрим функцию $S\colon G\to\mathbb{R}$, определенную равенством

$$S(\tau,\xi)=\int_{t_*}^{\tau}L(t,x(t,\lambda(\tau,\xi)),\dot{x}(t,\lambda(\tau,\xi)))\,dt.$$

Она называется *S-функцией центрального поля* $\{x(\cdot,\lambda)\}$. Эта функция непрерывно дифференцируема на G как суперпозиция непрерывно дифференцируемых функций. Подсчитаем ее частные производные. Имеем

$$\begin{aligned}\frac{\partial S(\tau,\xi)}{\partial\tau}&=L(\tau,x(\tau,\lambda(\tau,\xi)),\dot{x}(\tau,\lambda(\tau,\xi)))+\\&+\int_{t_*}^{\tau}(L_x(t,x(t,\lambda(\tau,\xi)),\dot{x}(t,\lambda(\tau,\xi)))x_\lambda(t,\lambda(\tau,\xi))\lambda_\tau(\tau,\xi)+\\&+L_{\dot{x}}(t,x(t,\lambda(\tau,\xi)),\dot{x}(t,\lambda(\tau,\xi)))\dot{x}_\lambda(t,\lambda(\tau,\xi))\lambda_\tau(\tau,\xi))\,dt.\end{aligned}$$

Интегрируя по частям во втором слагаемом в интеграле и учитывая, что функции $x(\cdot,\lambda)$ — экстремали, получим

$$\frac{\partial S(\tau,\xi)}{\partial\tau}=L(\tau,x(\tau,\lambda(\tau,\xi)),\dot x(\tau,\lambda(\tau,\xi)))+ \\ +L_{\dot x}(t,x(t,\lambda(\tau,\xi)),\dot x(t,\lambda(\tau,\xi)))x_\lambda(t,\lambda(\tau,\xi))\lambda_\tau(\tau,\xi))|_{t_*}^{\tau}.$$

Дифференцируя тождество $x(\tau,\lambda(\tau,\xi))=\xi$ по τ, получаем, что $\dot x(\tau,\lambda(\tau,\xi))+x_\lambda(\tau,\lambda(\tau,\xi))\lambda_\tau(\tau,\xi)=0$. Теперь учитывая, что по определению $\dot x(\tau,\lambda(\tau,\xi))=u(\tau,\xi)$ и что $x_\lambda(t_*,\lambda(\tau,\xi))=0$ (так как поле центрально), приходим к равенству

$$\frac{\partial S(\tau,\xi)}{\partial\tau}=L(\tau,\xi,u(\tau,\xi))-L_{\dot x}(\tau,\xi,u(\tau,\xi))u(\tau,\xi).$$

Аналогичные вычисления показывают, что

$$\frac{\partial S(\tau,\xi)}{\partial\xi}=\int_{t_*}^{\tau}(L_x(t,x(t,\lambda(\tau,\xi)),\dot x(t,\lambda(\tau,\xi)))x_\lambda(t,\lambda(\tau,\xi))\lambda_\xi(\tau,\xi)+ \\ +L_{\dot x}(t,x(t,\lambda(\tau,\xi)),\dot x(t,\lambda(\tau,\xi)))\dot x_\lambda(t,\lambda(\tau,\xi))\lambda_\xi(\tau,\xi))\,dt= \\ =L_{\dot x}(t,x(t,\lambda(\tau,\xi)),\dot x(t,\lambda(\tau,\xi)))x_\lambda(t,\lambda(\tau,\xi))\lambda_\xi(\tau,\xi))|_{t_*}^{\tau}.$$

Дифференцируя тождество $x(\tau,\lambda(\tau,\xi))=\xi$ по ξ, получаем $x_\lambda(t,\lambda(\tau,\xi))\lambda_\xi(\tau,\xi)=1$ и учитывая снова, что $\dot x(\tau,\lambda(\tau,\xi))=u(\tau,\xi)$ и $x_\lambda(t_*,\lambda(\tau,\xi))=0$, приходим к соотношению

$$\frac{\partial S(\tau,\xi)}{\partial\xi}=L_{\dot x}(\tau,\xi,u(\tau,\xi)).$$

Пусть $x(\cdot)\in PC^1([t_0,t_1])$ — допустимая функция в задаче (P_9), график которой принадлежит G.[6] Рассматривая функцию S на графике $x(\cdot)$, получаем функцию $t\mapsto S(t,x(t))$ на $[t_0,t_1]$. Ясно, что эта функция принадлежит $PC^1([t_0,t_1])$ и, за исключением конечного числа точек, будем иметь (учитывая выражения для частных производных функции S)

$$S'(t,x(t))=L(t,x(t),u(t,x(t)))-L_{\dot x}(t,x(t),u(t,x(t)))\,u(t,x(t))+ \\ +L_{\dot x}(t,x(t),u(t,x(t)))\,\dot x(t). \qquad (i)$$

Если $x(\cdot)=\widehat x(\cdot)$, то $u(t,\widehat x(t))=\dot{\widehat x}(t)$ и тогда

$$S'(t,\widehat x(t))=L(t,\widehat x(t),\dot{\widehat x}(t)). \qquad (ii)$$

Теперь, используя (ii), а затем (i), получим

$$J(x(\cdot))-J(\widehat x(\cdot))=\int_{t_0}^{t_1}L(t,x(t),\dot x(t))\,dt-\int_{t_0}^{t_1}L(t,\widehat x(t),\dot{\widehat x}(t))\,dt= \\ =\int_{t_0}^{t_1}L(t,x(t),\dot x(t))\,dt-\int_{t_0}^{t_1}S'(t,\widehat x(t))\,dt=\int_{t_0}^{t_1}L(t,x(t),\dot x(t))\,dt- \\ -(S(t_1,x_1)-S(t_0,x_0))=\int_{t_0}^{t_1}L(t,x(t),\dot x(t))\,dt- \\ -\int_{t_0}^{t_1}S'(t,x(t))\,dt=\int_{t_0}^{t_1}(L(t,x(t),\dot x(t))-L(t,x(t),u(t,x(t)))- \\ -(\dot x(t)-u(t,x(t)))L_{\dot x}(t,x(t),u(t,x(t)))\,dt=\int_{t_0}^{t_1}\mathcal{E}(t,x(t),u(t,x(t)),\dot x(t))\,dt.$$

---

[6] Нетрудно проверить, что множество $\{\,x(\cdot)\in C([t_0,t_1])\mid (t,x(t))\in G,\ t\in[t_0,t_1]\,\}$ открыто в $C([t_0,t_1])$ и тем самым есть окрестность $\widehat x(\cdot)$ в $C([t_0,t_1])$.

Формула $J(x(\cdot)) - J(\widehat{x}(\cdot)) = \int_{t_0}^{t_1} \mathcal{E}(t, x(t), u(t, x(t)), \dot{x}(t))\, dt$ называется *основной формулой Вейрштрасса.*

Поскольку функция $\dot{x} \mapsto L(t, x, \dot{x})$ выпукла на $\mathbb{R}$ для любых $(t, x) \in G$, то $\mathcal{E}(t, x(t), u(t, x(t)), \dot{x}(t)) \geq 0$ для всех $t \in [t_0, t_1]$ и значит, $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$. $\square$

Замечание. Обозначим $\mathcal{H}(\tau, \xi, p) = p \cdot u(\tau, \xi) - L(\tau, \xi, u(\tau, \xi))$. Тогда из определений сразу следует, что на G функция S удовлетворяет следующему уравнению в частных производных

$$\frac{\partial S(\tau, \xi)}{\partial \tau} + \mathcal{H}\left(\tau, \xi, \frac{\partial S(\tau, \xi)}{\partial \xi}\right) = 0,$$

которое называется *уравнением Гамильтона–Якоби.*

Вопросы к экзамену

1. Дифференцируемость, непрерывная дифференцируемость и строгая дифференцируемость конечномерных отображений. Конечномерные теоремы отделимости (доказательство второй теоремы отделимости).

2. Модифицированный метод Ньютона и разрешимость конечномерной системы нелинейных уравнений.

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

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

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

6. Правило множителей Лагранжа для конечномерной гладкой задачи с ограничениями, задаваемыми равенствами и неравенствами (формулировка теоремы Каруша–Джона).

7. Необходимые и достаточные условия минимума второго порядка для конечномерной гладкой задачи с ограничениями, задаваемыми равенствами и неравенствами (формулировка теоремы Левитина–Милютина–Осмоловского).

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

9. Уравнение Эйлера для простейшей задачи классического вариационного исчисления и его интегралы. Необходимые условия слабого экстремума для задачи Больца.

10. Необходимые и достаточные условия минимума максимума конечного числа гладких функций многих переменных.

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

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

13. Принцип максимума Понтрягина для понтрягинской задачи оптимального управления

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

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

17. Поле экстремалей. Достаточные условия сильного минимума для простейшей задачи вариационного исчисления. Основная формула Вейерштрасса. Уравнение Гамильтона–Якоби.