Вариационное исчисление и оптимальное управление (курс лекций)

Г. Г. Магарил-Ильяев

Содержание

Введение

Какие причины побуждают решать задачи на максимум и минимум? Интерес к задачам на максимум и минимум (или короче, к задачам на экстремум или экстремальным задачам) проявился уже на заре развития математики. Первая причина такого интереса лежит в области психологии. Любознательность и стремление к совершенству, желание "дойти до самой сути" были присущи человеку во все времена, и это являлось и является стимулом к поиску 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

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

более ранними размышлениями Галилея на эту тему). Историю развития вариационного исчисления принято отсчитывать именно с этого времени — года решения задачи о брахистохроне.

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

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. Дифференцируемость, строгая дифференцируемость и субдифференцируемость. Конечномерные теоремы отделимости

В этом пункте напоминаются необходимые для дальнейшего факты, связанные с пространством $\mathbb{R}^n$, с дифференциальными свойствами отображений из $\mathbb{R}^n$ в $\mathbb{R}^m$ и 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 $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$. 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 $f_1, \ldots, f_m$, то они определяют отображение $F\colon A \to \mathbb{R}^m$ по правилу $F(x) = (f_1(x), \ldots, f_m(x))^T$, $x \in A$. В этом смысле многие вопросы, связанные с отображениями можно свести к соответствующим вопросам для функций. Однако часто удобно (и полезно, имея в ввиду обобщения на бесконечномерный случай) работать именно с отображениями.

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

для любой последовательности $\{x_k\}$ точек из A, сходящейся к $\widehat{x}$ последовательность $\{F(x_k)\}$ сходится к $F(\widehat{x})$.

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

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.

Дифференцируемость и непрерывная дифференцируемость. Пусть U — окрестность точки $\widehat{x} \in \mathbb{R}^n$. Говорят, что функция $f : U \to \mathbb{R}$ *дифференцируема в точке* $\widehat{x}$, если существует линейный функционал на $\mathbb{R}^n$, т. е. вектор $a = (a_1, \ldots, a_n) \in (\mathbb{R}^n)^*$ 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 $f'(\widehat{x})$.

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 $\partial f(\widehat{x})/\partial x_1, \ldots, \partial f(\widehat{x})/\partial x_n$. 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

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

этой точке. Соответствующих примеров множество. Скажем, функция на $\mathbb{R}^2$, равная единице на множестве $\{\,(x_1,x_2)\mid x_1x_2=0\,\}$ 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, $i=1,\dots,m$, и отображение $F\colon U\to\mathbb{R}^m$ определено по правилу: $F(x)=(f_1(x),\dots,f_m(x))^T$, то легко проверить, что F дифференцируемо в точке $\widehat{x}$ тогда и только тогда, когда функции f_i, $1\le i\le m$, дифференцируемы в $\widehat{x}$. При этом строки матрицы $F'(\widehat{x})$ суть векторы $f_1'(\widehat{x}),\dots,f_k'(\widehat{x})$. 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 $f'\colon U\to(\mathbb{R}^n)^*$, 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 $F'\colon U\to\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$, 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).

Строгая дифференцируемость. 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 $x \mapsto F(x) - F'(\widehat{x})x$.

Субдифференцируемость. Пусть $x, y \in \mathbb{R}^n$. Множество $[x, y] = \{z \in \mathbb{R}^n \mid z = (1-\alpha)x + \alpha y,\ 0 \le \alpha \le 1\}$ называется *отрезком* (*соединяющим точки x и y*). Непустое множество $A \subset \mathbb{R}^n$ называется *выпуклым,* если вместе с любыми двумя своими точками x и y оно содержит отрезок $[x, y]$.

Пустое множество выпукло по определению.

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

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

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

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

Собственная функция f на $\mathbb{R}^n$ выпукла тогда и только тогда, когда для любых $x_1, x_2 \in \mathbb{R}^n$ 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, 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 $n = 1$).

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.

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. $x^* \cdot x \geq x^* \cdot \widehat{x}$ для любого $x \in A$. Предположим противное, что существует такой элемент $x_0 \in A$, что $x^* \cdot x_0 < x^* \cdot \widehat{x}$. Поскольку A выпукло, то $(1-t)\widehat{x} + tx_0 \in A$ при $0 \leq t \leq 1$ и при малых t (учитывая, что $x^* \cdot (x_0 - \widehat{x}) < 0$) будем иметь

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

Величина справа не превосходит r^2, следовательно, точки $(1-t)\widehat{x} + tx_0$ принадлежат множеству A_1, но это противоречит тому, что $\widehat{x}$ — минимум функции f на этом множестве. Итак, $x^* \cdot b < x^* \cdot \widehat{x} \leq x^* \cdot x$ для всех $x \in A$, или $x^* \cdot b < \inf_{x \in A} x^* \cdot x$, т. е. A и b строго отделимы. $\square$

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. Понятие субдифференциала принадлежит Фенхелю (1951 г.). Теорема отделимости точки от выпуклого множества доказана Минковским (1910 г.).

2. Модифицированный метод Ньютона и разрешимость конечномерной системы нелинейных уравнений

В этом параграфе будет доказан второй основополагающий для построения теории экстремума результат — теорема о разрешимости системы нелинейных уравнений.

И. Ньютон неоднократно излагал способ нахождения корня нелинейного уравнения $F(x) = 0$ для функции одного переменного. В частности, в известном письме к Г. Ольденбургу (1676 г.) он демонстрирует его на примере решения уравнения $x^3 - 2x - 5 = 0$. Этот способ известен как "метод Ньютона". Для нахождения решения уравнения $F(x) = y$ он заключается в построении последовательности $\{x_k\}$, где x_k находится как решение линейного уравнения $F'(x_{k-1})(x - x_{k-1}) = y - F(x_{k-1})$, т. е. $x_k = x_{k-1} + (F'(x_{k-1}))^{-1}(y - F(x_{k-1}))$, $\;k \in \mathbb{N}$. При этом, x_0 стараются выбрать так, чтобы $F(x_0)$ было достаточно близко к y.

Мы для решения нелинейного уравнения $F(x) = y$ ($x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$) будем использовать *модифицированный метод Ньютона*. Он основан на решении линейного уравнения $\Lambda x = y$, где $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

Напомним, что оператор $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ называется *сюръективным*, если $\Lambda\mathbb{R}^n = \mathbb{R}^m$, т. е. для любого $y \in \mathbb{R}^m$ существует решение уравнения $\Lambda x = y$.

Лемма (о правом обратном, или о разрешимости системы линейных уравнений)**.** *Пусть $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ — сюръективный оператор. Тогда существуют непрерывное отображение $R\colon \mathbb{R}^m \to \mathbb{R}^n$ и константа $\gamma > 0$ такие, что $\Lambda R(y) = y$ и $|R(y)| \le \gamma|y|$ для всех $y \in \mathbb{R}^m$.*

Доказательство. Пусть $e_1, \ldots, e_m$ — 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 $\gamma = \sum_{i=1}^m |f_i|$. □

Теорема 2 (о разрешимости системы нелинейных уравнений). *Пусть U — окрестность точки $\widehat{x} \in \mathbb{R}^n$, отображение $F\colon U \to \mathbb{R}^m$ строго дифференцируемо в $\widehat{x}$ и $F'(\widehat{x})$ — сюръективный оператор. Тогда существуют окрестность W точки $F(\widehat{x})$ и константа $K > 0$ такие, что для каждого $y \in W$ найдется элемент $x(y) \in U$ такой, что $F(x(y)) = y$ и $|x(y) - \widehat{x}| \le K|y - F(\widehat{x})|$.*

Доказательство. Обозначим для краткости $\Lambda = F'(\widehat{x})$. Оператор Λ удовлетворяет условиям леммы о правом обратном. Пусть соответствующие отображение R и константа γ из этой леммы. Пусть, далее, $\varepsilon > 0$ таково, что $\varepsilon\gamma < 1$. Согласно определению строгой дифференцируемости F в точке $\widehat{x}$ найдется $\delta = \delta(\varepsilon) > 0$ такое, что для всех $x, x' \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \subset U$ справедливо неравенство

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

где $\theta = \varepsilon\gamma$.

Положим $W = U_{\mathbb{R}^m}(F(\widehat{x}), \delta(1-\theta)/\gamma)$ и пусть $y \in W$. Рассмотрим последовательность (модифицированный метод Ньютона)

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

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

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

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

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

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

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

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

т. е. $x_{k+1}\in U_{\mathbb{R}^n}(\widehat{x},\delta)$ и тем самым $x_k\in U_{\mathbb{R}^n}(\widehat{x},\delta)$ для всех $k\ge 0$.

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

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

Положим $x(y)=\lim_{k\to\infty}x_k$. Тогда из (v) следует, что $x(y)\in U_{\mathbb{R}^n}(\widehat{x},\delta)\subset U$. Из (i) вытекает (как уже отмечалось) непрерывность F на $U_{\mathbb{R}^n}(\widehat{x},\delta)$. Переходя к пределу в (iii) при $k\to\infty$, получаем, что $F(x(y))=y$, а переход к пределу в (iv) приводит к оценке $|x(y)-\widehat{x}|\le K|y-F(\widehat{x})|$, где $K=\gamma/(1-\theta)$. □

Замечания.

1. В лемме о правом обратном можно явным образом предъявить один из правых обратных: $R(y)=\Lambda^T(\Lambda\Lambda^T)^{-1}$, где Λ^T — матрица, транспонированная к Λ ($\Lambda\Lambda^T$ — обратимая матрица, как нетрудно проверить). Но преимущество приведенного в лемме доказательства заключается в том, что оно легко обобщается на другие ситуации.

2. Из приведенного доказательства теоремы видно, что строгая дифференцируемость — естественное требование гладкости при нахождении решения нелинейного уравнения с помощью модифицированного метода Ньютона.

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 для конечномерных гладких задач без ограничений и правило множителей Лагранжа для конечномерных гладких задач с ограничениями, задаваемые равенствами

В этом пункте будут доказаны изначальные результаты теории экстремума — теорема Ферма (необходимое условие экстремума для гладких задач без ограничений) и правило множителей

Лагранжа для гладких задач с ограничениями, задаваемыми равенствами. Как будет видно из дальнейшего, 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.

3.1. Гладкие задачи без ограничений.

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

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

$$f(x) \to \mathrm{extr}, \quad x \in U. \qquad (P_2)$$

С точки зрения общей задачи (P_1) здесь $U = C = X$, т. е. (P_2) — задача без ограничений.

Теорема (Ферма для 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.

2. Задачи с ограничениями, задаваемые равенствами.

Можно высказать следующий общий принцип. Если ищется
максимум или минимум некоторой функции многих
переменных при условии, что между этими переменными
имеется связь, задаваемая одним или несколькими
уравнениями, нужно прибавить к минимизируемой
функции функции, задающие уравнения связи,

умноженные на неопределенные множители, и искать затем максимум или минимум построенной суммы, как если бы переменные были независимы. Полученные уравнения, присоединенные к уравнениям связи, послужат для определения всех неизвестных

Ж. Л. Лагранж

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

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

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 $\lambda_0, \lambda_1, \ldots, \lambda_m$ — *множителями Лагранжа*, а вектор $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m)$ — *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, $i = 0, 1, \ldots, m$, 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 $\lambda_0 \ne 0$.

Доказательство. Если векторы $f_0'(\widehat{x}), f_1'(\widehat{x}), \ldots, f_m'(\widehat{x})$ линейно зависимы, то это и есть утверждение теоремы. Покажем, что если эти векторы линейно независимы, то применяя теорему о разрешимости, придем к противоречию с тем, что $\widehat{x}$ — локальный экстремум.

Действительно, определим отображение $F\colon V \to \mathbb{R}^{m+1}$ по формуле $F(x) = (f_0(x), f_1(x), \ldots, f_m(x))^T$. Оно строго дифференцируемо в точке $\widehat{x}$ (в силу строгой дифференцируемости в $\widehat{x}$ функций $f_0, f_1, \ldots, f_m$) и $F'(\widehat{x})(\mathbb{R}^n) = \mathbb{R}^{m+1}$, так как строки матрицы $F'(\widehat{x})$ суть векторы $f_0'(\widehat{x}), f_1'(\widehat{x}), \ldots, f_m'(\widehat{x})$. По теореме 2 найдутся окрестность W точки $F(\widehat{x}) = (f_0(\widehat{x}), 0, \ldots, 0)^T$ и константа $K > 0$ такие, что для всех $y \in W$ существует вектор $x(y) \in U$, для которого $F(x(y)) = y$ и $|x(y) - \widehat{x}| \le K|y - F(\widehat{x})|$. Если ν достаточно мало по модулю, то точки $y_\nu = (f_0(\widehat{x}) + \nu, 0, \ldots, 0)^T$, очевидно, принадлежат W. Тогда $F(x(y_\nu)) = y_\nu$, или $(f_0(x(y_\nu)), f_1(x(y_\nu)), \ldots, f_m(x(y_\nu)))^T = (f_0(\widehat{x}) + \nu, 0, \ldots, 0)^T$ и $|x(y_\nu) - \widehat{x}| \le K|y_\nu - F(\widehat{x})| = K|\nu|$, т. е. в любой окрестности $\widehat{x}$ находятся допустимые в задаче (P_3) точки $x(y_\nu)$ и при этом, $f_0(x(y_\nu)) < f_0(\widehat{x})$ $(f_0(x(y_\nu)) > f_0(\widehat{x}))$, если $\nu < 0$ $(\nu > 0)$ в противоречие с тем, что $\widehat{x}$ — локальный экстремум.

Последнее утверждение теоремы проверяется без труда. $\square$

Замечания.

1. Стандартные требования гладкости функций в классическом правиле множителей Лагранжа — их непрерывная дифференцируемость в окрестности точки $\widehat{x}$. В силу сказанного выше, это правило есть следствие данной теоремы.

2. Насколько существенны требования строгой дифференцируемости функций в теореме 3? Можно ли обойтись только дифференцируемостью этих функций в точке $\widehat{x}$? Оказывается нет, нужна еще непрерывность функций $f_1, \ldots, f_m$ в окрестности точки $\widehat{x}$ (которая в нашем случае следует из строгой дифференцируемости). В противном случае теорема не верна (соответствующий пример см. ...). От функции же f_0 достаточно требовать, только дифференцируемость.

Исторический комментарий. Правило множителей (вынесенное выше в эпиграф) было сформулировано в мемуаре Лагранжа Théorie des fonctions analytiques, Paris, 1797.

В этом параграфе упоминались три великих имени.

Пьер Ферма (1601 – 1665) оказал огромное влияние на последующее развитие математики. Его вклад в анализ, естествознание, теорию чисел и геометрию, где он был почти всюду первопроходцем, сохранил свое значение до сей поры.

Исаак Ньютон (1643 – 1727) — величайший ученый всех времен. Он заложил основы точного естествознания, разработав язык, на котором формулируются законы природы и начал построение Макромира.

Жозеф Луи Лагранж (1736 – 1813), наряду с Л. Эйлером, был величайшим ученым 18 века. Его имя более других будет фигурировать в этой брошюре.

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

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

называется *задачей с ограничениями, задаваемые равенствами 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 (Правило множителей Лагранжа для задачи (P_4)). *Если точка $\widehat{x} \in U$ является локальным минимумом в задаче (P_4) и функции f_i, $i = 0, 1, \ldots, m$, 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*).

Для доказательства этого результата нужна некоторая модификация теоремы о разрешимости системы нелинейных уравнений.

Напомним, что множество $C \subset \mathbb{R}^n$ называется *конусом*, если с каждым своим элементом x оно содержит и элемент αx для любого $\alpha > 0$. Нетрудно проверить, что конус C является выпуклым тогда и только тогда, когда $C + C = C$, или тогда и только тогда, когда он содержит все элементы вида $\alpha_1 x_1 + \ldots + \alpha_s x_s$, где $x_i \in C$, $\alpha_i > 0$, $1 \le i \le s$, и $\sum_{i=1}^{s} \alpha_i \ne 0$.

Лемма (о правом обратном для конуса). *Пусть $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, C — выпуклый конус в $\mathbb{R}^n$ и $\Lambda(C) = \mathbb{R}^m$. Тогда существуют отображение $R\colon \mathbb{R}^m \to C$ и константа $\gamma > 0$ такие, что $\Lambda R(y) = y$ и $|R(y)| \le \gamma |y|$ для всех $y \in \mathbb{R}^m$.*

Доказательство. Пусть $e_1, \ldots, e_m$ — базис в $\mathbb{R}^m$. По предположению найдутся такие $f_i \in C$, $1 \le i \le 2m$, что $\Lambda f_i = e_i$ и $\Lambda f_{m+i} = -e_i$, $1 \le i \le m$. Для каждого $y = \sum_{i=1}^{m} y_i e_i \in \mathbb{R}^m$ положим $R(y) = \sum_{i=1}^{m} |y_i| g_i$, где $g_i = f_i$, если $y_i \ge 0$ и $g_i = f_{m+i}$, если $y_i < 0$. Ясно, что $R(y) \in C$, $\Lambda(R(y)) = \sum_{i=1}^{m} |y_i| \Lambda g_i = \sum_{i=1}^{m} y_i e_i = y$ и $|R(y)| \le \sum_{i=1}^{m} |y_i||g_i| \le \max_{1 \le i \le m} |y_i| \sum_{i=1}^{2m} |f_i| \le \gamma |y|$, где $\gamma = \sum_{i=1}^{2m} |f_i|$. $\square$

Теорема $2'$ (о разрешимости системы нелинейных уравнений для конуса) *Пусть U — окрестность точки $\widehat{x}$ в $\mathbb{R}^n$, C — выпуклый замкнутый конус в $\mathbb{R}^n$, отображение $F\colon U \to \mathbb{R}^m$ строго дифференцируемо в $\widehat{x}$ и $F'(\widehat{x})(C) = \mathbb{R}^m$. Тогда существуют окрестность*

W точки $F(\widehat{x})$ и константа $K > 0$ такие, что для каждого $y \in W$ найдется элемент $x(y) \in U \cap (\widehat{x} + C)$ такой, что $F(x(y)) = y$ и $|x(y) - \widehat{x}| \leq K|y - F(\widehat{x})|$.

Доказательство. Рассуждения здесь точно те же, что и в теореме 2; надо только убедиться в том, что $x(y) \in U \cap (\widehat{x} + C)$ для всех $y \in W$. Действительно, $x_0 = \widehat{x} \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \cap (\widehat{x} + C)$. Пусть $x_s \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \cap (\widehat{x} + C)$, $1 \leq s \leq k$. Покажем, что $x_{k+1} \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \cap (\widehat{x} + C)$. То, что $x_{k+1} \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$ доказано. Так как $x_k \in \widehat{x} + C$, то из (ii) и в силу выпуклости C следует, что $x_{k+1} \in x_k + C \subset \widehat{x} + C + C = \widehat{x} + C$, т. е. $x_{k+1} \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \cap (\widehat{x} + C)$ и значит, вся последовательность принадлежит этому множеству. Вследствие замкнутости C получаем, что $x(y) \in U_{\mathbb{R}^n}(\widehat{x}, \delta) \cap (\widehat{x} + C) \subset U \cap (\widehat{x} + C)$. □

Доказательство теоремы 4. Заметим сначала, что утверждение (c) можно считать выполненным всегда. В самом деле, отбросим те ограничения среди неравенств, для которых $f_i(\widehat{x}) < 0$. Тогда $\widehat{x}$ будет локальным экстремумом и в новой задаче. Если для этой задачи доказаны утверждения (a) и (b), то (c) выполняется автоматически. Дополнив найденный набор множителей Лагранжа нулевыми компонентами, соответствующие тем номерам, где $f_i(\widehat{x}) < 0$, получим утверждения (a), (b) и (c) для исходной задачи.

Рассмотрим отображение $F\colon U \times \mathbb{R}^{m'+1} \to \mathbb{R}^{m+1}$, определенное по правилу: $F(x, u) = (f_0(x) + u_0, f_1(x) + u_1, \ldots, f_{m'}(x) + u_{m'}, f_{m'+1}(x), \ldots, f_m(x))^T$, где $u = (u_0, u_1, \ldots, u_{m'})^T$. Оно, очевидно, строго дифференцируемо в точке $(\widehat{x}, 0)$. Возможно одно из двух: либо $F'(\widehat{x}, 0)(\mathbb{R}^n \times \mathbb{R}^{m'+1}_+)$ совпадает со всем пространством $\mathbb{R}^{m+1}$, либо нет. Покажем, что первая возможность приводит к противоречию с тем, что $\widehat{x}$ — локальный минимум. В самом деле, в этом случае выполнены условия теоремы о разрешимости нелинейных уравнений для конуса, в которой вместо $\mathbb{R}^n$, $\mathbb{R}^m$, U и $\widehat{x}$ надо рассмотреть $\mathbb{R}^n \times \mathbb{R}^{m'+1}$, $\mathbb{R}^{m+1}$, $U \times \mathbb{R}^{m'+1}$ и $(\widehat{x}, 0)$, а $C = \mathbb{R}^n \times \mathbb{R}^{m'+1}_+$. Так как условие (c) выполнено, то $F(\widehat{x}, 0) = (f_0(\widehat{x}), 0, \ldots, 0)^T$. Поэтому для всех достаточно малых $\varepsilon > 0$ точки $y_\varepsilon = (f_0(\widehat{x}) - \varepsilon, 0 \ldots, 0)^T$ принадлежат W. Тогда найдутся точки $(x(y_\varepsilon), u(y_\varepsilon)) \in (U \times \mathbb{R}^{m'+1}) \cap ((\widehat{x}, 0) + \mathbb{R}^n \times \mathbb{R}^{m'+1}_+) = U \cap \mathbb{R}^{m'+1}_+$ такие, что $F(x(y_\varepsilon), u(y_\varepsilon)) = y_\varepsilon$, или $f_0(x(y_\varepsilon)) + u_0(y_\varepsilon) = f_0(\widehat{x}) - \varepsilon, f_1(x(y_\varepsilon)) + u_1(y_\varepsilon) = 0, \ldots, f_{m'}(x(y_\varepsilon)) + u_{m'}(y_\varepsilon) = 0, f_{m'+1}(x(y_\varepsilon)) = 0, \ldots, f_m(x(y_\varepsilon)) = 0$. Так как $u_i(x(y_\varepsilon)) \geq 0$, $i = 1, \ldots, m'$, то $f_i(x(y_\varepsilon)) \leq 0$, $i = 1, \ldots, m'$ и тем самым точки $x(y_\varepsilon)$ допустимы в задаче (P_4). Кроме того $|x(y_\varepsilon) - \widehat{x}| \leq |(x(y_\varepsilon), u(y_\varepsilon)) - (\widehat{x}, 0)| \leq K|y_\varepsilon - F(\widehat{x}, 0)| = K\varepsilon$ и $f_0(x(y_\varepsilon)) \leq f_0(x(y_\varepsilon)) + u_0(y_\varepsilon) = f_0(\widehat{x}) - \varepsilon < f_0(\widehat{x})$. Таким образом, в любой окрестности $\widehat{x}$ существуют допустимые в задаче (P_4) точки, в которых значение f_0 меньше, чем в $\widehat{x}$. Это противоречит тому, что $\widehat{x}$ — локальный минимум.

Итак, $F'(\widehat{x}, 0)(\mathbb{R}^n \times \mathbb{R}^{m'+1}_+) \neq \mathbb{R}^{m+1}$. Ясно, что $C_1 = F'(\widehat{x}, 0)(\mathbb{R}^n \times \mathbb{R}^{m'+1}_+)$ — выпуклый конус в $\mathbb{R}^{m+1}$. Пусть $y_0 \notin C_1$. По первой теореме отделимости (теорема 1.2) найдется ненулевой вектор $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in (\mathbb{R}^{m+1})^*$ такой, что $\overline{\lambda} \cdot y_0 \leq \inf_{y \in C_1} \overline{\lambda} \cdot y$. Отсюда легко следует, что $\overline{\lambda} \cdot y \geq 0$ для всех $y \in C_1$. Это равносильно тому, что

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

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

Исторический комментарий. Задачи с неравенствами стали играть важную роль после работы Л. В. Канторовича 1939 г., в которой он показал, что многие задачи экономического содержания формализуются как экстремальные задачи с ограничениями, задаваемыми линейными неравенствами. Подобным задачам будет посвящен отдельный параграф. Задача (P_4) впервые была рассмотрена в диссертации В. Каруша, защищенной в 1939 г. в Чикаго, но результат остался неизвестным за пределами узкого круга математиков. В 1948 г. теорема была передоказана Ф. Джоном, после чего она стала достоянием широкого круга специалистов по теории экстремума.

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 $f_0\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ — 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 $f_0(\widehat{x}) \le f_0(x)$ для всех

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

Напомним, что понятие субдифференциала выпуклой функции было определено в §1.

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 $f_0(x) - f_0(\widehat{x}) \ge 0 = 0 \cdot x$ 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 $f_0(x) - f_0(\widehat{x}) \ge 0 \cdot x = 0$, т. е. $f_0(x) \ge f_0(\widehat{x})$ для любого $x \in \mathbb{R}^n$. □

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

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

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

С задачей (P_5) связывается функция Лагранжа $\mathcal{L}\colon \mathbb{R}^n \times (\mathbb{R}^{m+1})^* \to \mathbb{R}$, $\mathcal{L}(x, \overline{\lambda}) = \sum_{i=0}^m \lambda_i f_i(x)$, 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 $\overline{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_m)$, что выполняются*

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

Если существует допустимая в (P_5) точка $\widehat{x}$ и набор множителей Лагранжа $\overline{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_m)$, 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 $f_i(\overline{x}) < 0$, $1 \le i \le m$, то $\lambda_0 \ne 0$ (условие Слейтера).

Доказательство. Пусть $\widehat{x}$ — решение задачи (P_5). Рассмотрим множества $B = \{\, b = (b_0, b_1, \dots, b_m)^T \in \mathbb{R}^{m+1} \mid \exists x \in A : f_0(x) - f_0(\widehat{x}) \le b_0,\ f_i(x) \le b_i,\ i = 1, \dots, m \,\}$ и $C = \{\, c \in \mathbb{R}^{m+1} \mid c = (c_0, 0, \dots, 0)^T,\ c_0 < 0 \,\}$. Покажем, что эти множества непусты, выпуклы и $B \cap C = \emptyset$.

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

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

Согласно первой теореме отделимости (теорема 1.1) найдется такой ненулевой вектор $\overline{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_m) \in (\mathbb{R}^{m+1})^*$, что $\sup_{c \in C} \sum_{i=0}^m \lambda_i c_i \le \inf_{b \in B} \sum_{i=0}^m \lambda_i b_i$, или $\sup_{c_0 < 0} \lambda_0 c_0 \le \inf_{b \in B} \sum_{i=0}^m \lambda_i b_i$. Отсюда следует, что

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

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

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

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

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

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

Деля на λ_0, получаем требуемое.

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

6. 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

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

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

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

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:

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

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

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

справедливому для всех функций $x(\cdot) \in C^1([t_0, t_1])$, для которых $x(t_0) = x(t_1) = 0$. В частности, $\int_{t_0}^{t_1} \dot{x}(t)\, dt = 0$. В силу последнего, равенство (ii) остается верным, если к первому сомножителю в интеграле прибавить любую константу. Выберем константу c так, чтобы $\int_{t_0}^{t_1} (-p(t) + \widehat{L}_{\dot{x}}(t) - c)\, dt = 0$. Тогда подставляя в аналог равенства (ii), где выражение в скобках равно $-p(\cdot) + \widehat{L}_{\dot{x}}(\cdot) - c$, функцию $x(\cdot)$, для которой $\dot{x}(\cdot) = -p(\cdot) + \widehat{L}_{\dot{x}}(\cdot) - c$, 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 $x_\alpha(\cdot)$ — то же, что и в пункте (a). Рассуждая точно также как и выше, приходим к соотношению

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

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 (iii) и интегрируя по частям, получаем, что

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

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 (iv), приходим к равенству

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

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

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 $\ddot{\widehat{x}}(\cdot)$.

[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.

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

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

называется *задачей Лагранжа классического 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] $\Gamma(x(\cdot), u(\cdot)) = \{\, (t, x(t), u(t)) \mid t \in [t_0, t_1] \,\} \subset G$, $\dot{x}(t) = \varphi(t, x(t), u(t))$ для всех $t \in [t_0, t_1]$ и $x(t_i) = x_i$, $i = 0, 1$.

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

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).

Свяжем с задачей (P_7) *функцию Лагранжа*

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

где λ_0 и $p\colon [t_0, t_1] \to (\mathbb{R}^n)^*$ — *множители Лагранжа,* $\overline{\lambda} = (\lambda_0, p(\cdot))$ — *набор множителей Лагранжа.* Функцию под знаком интеграла (интегрант) обозначим через $L = L(t, x, \dot{x}, u, \lambda_0, p) = \lambda_0 f(t, x, u) + p \cdot (\dot{x} - \varphi(t, x, u))$.

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: $\widehat{f}_x(t) = f_x(t, \widehat{x}(t), \widehat{u}(t))$ и

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

аналогично для частной производной по u, частных производных отображения φ и интегранта L.

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 экстремум 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 $\widehat{L}_{\dot{x}}(\cdot) \in C^1([t_0, t_1])$, 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 2 о разрешимости системы нелинейных уравнений и теоремы 1.2 (второй теоремы отделимости), понадобится еще стандартная теорема о дифференцируемой зависимости решений дифференциальных уравнений от параметров, которую здесь приводим без доказательства и в удобной для нас форме.

Теорема (о дифференцируемой зависимости). *Пусть $u_i(\cdot) \in C([t_0, t_1], \mathbb{R}^r)$, $i = 1, \ldots, k$, $\overline{u}(\cdot) = (u_1(\cdot), \ldots, u_k(\cdot))$ и $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — допустимая пара в задаче (P_7). Тогда найдется такая окрестность нуля V в $\mathbb{R}^k$, что для любого $\overline{\alpha} = (\alpha_1, \ldots, \alpha_k)^T \in V$ существует единственное решение $x(\cdot, \overline{\alpha}, \overline{u}(\cdot))$ задачи Коши: $\dot{x} = \varphi(t, x, \widehat{u}(t) + \sum_{i=1}^k \alpha_i u_i(t))$, $x(t_0) = x_0$, определенное на всем отрезке $[t_0, t_1]$. Кроме того, $x(\cdot, \overline{\alpha}, \overline{u}(\cdot)) \to \widehat{x}(\cdot)$ при $\overline{\alpha} \to 0$ в метрике $C^1([t_0, t_1], \mathbb{R}^n)$, для каждого $t \in [t_0, t_1]$ отображение $\overline{\alpha} \mapsto x(t, \overline{\alpha}, \overline{u}(\cdot))$ непрерывно дифференцируемо на V и его частные производные в нуле на отрезке $[t_0, t_1]$ удовлетворяют дифференциальным уравнениям*

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

Доказательство теоремы 7. Пусть $u(\cdot) \in C([t_0, t_1], \mathbb{R}^r)$, $\alpha \in \mathbb{R}$ и $u(\cdot, \alpha, u(\cdot)) = \widehat{u}(\cdot) + \alpha u(\cdot)$. Согласно теореме о дифференцируемой зависимости, для всех достаточно малых по модулю α существует единственное решение $x(\cdot, \alpha, u(\cdot))$ задачи Коши: $\dot{x} = \varphi(t, x, u(t, \alpha, u(\cdot))$, $x(t_0) = x_0$. Если $t \in [t_0, t_1]$, то производную отображения $\alpha \mapsto x(t, \alpha, u(\cdot))$ в нуле обозначим $y(t, u(\cdot))$.

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

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

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

— локальный экстремум. Действительно, пусть данное предположение верно. Тогда существуют такие $u_i(\cdot) \in C([t_0,t_1],\mathbb{R}^r)$, $i = 1,\dots,n+1$, что векторы $(y_0(u_i(\cdot)), y(t_1,u_i(\cdot)))$, $i = 1,\dots,n+1$, образуют базис в $\mathbb{R}^{n+1}$. Положим $\overline{u}(\cdot) = (u_1(\cdot),\dots,u_{n+1}(\cdot))$, $\overline{\alpha} = (\alpha_1,\dots,\alpha_{n+1})^T$ и $u(\cdot,\overline{\alpha},\overline{u}(\cdot)) = \widehat{u}(\cdot) + \sum_{i=1}^{n+1}\alpha_i u_i(\cdot)$. В силу теоремы о дифференцируемой зависимости для достаточно малых $\overline{\alpha}$ существует решение $x(\cdot,\overline{\alpha},\overline{u}(\cdot))$ задачи Коши: $\dot{x} = \varphi(t,x,u(t,\overline{\alpha},\overline{u}(\cdot)))$, $x(t_0) = x_0$ и, снова, в силу стандартных фактов, для малых $\overline{\alpha}$ функция $\overline{\alpha} \mapsto J(x(\cdot,\overline{\alpha},\overline{u}(\cdot)), u(\cdot,\overline{\alpha},\overline{u}(\cdot)))$ определена и непрерывно дифференцируема. Таким образом, существует окрестность V нуля в $\mathbb{R}^{n+1}$, что корректно определено отображение $F\colon V \to \mathbb{R}^{n+1}$, сопоставляющее $\overline{\alpha} \in V$ вектор $(J(x(\cdot,\overline{\alpha},\overline{u}(\cdot)), u(\cdot,\overline{\alpha},\overline{u}(\cdot))), x(t_1,\overline{\alpha},\overline{u}(\cdot)))$. Это отображение непрерывно дифференцируемо на V и его производная (матрица Якоби) в нуле невырождена, поскольку ее столбцы суть векторы $(y_0(u_i(\cdot)), y(t_1,u_i(\cdot)))$, $i = 1,\dots,n+1$.

Теперь применяя обычную теорему об обратной функции или теорему 2 о разрешимости системы нелинейных уравнений (условия, которой выполнены, поскольку отображение F непрерывно дифференцируемо на V 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 экстремум. Действительно, $F(0) = (J(\widehat{x}(\cdot),\widehat{u}(\cdot)), x_1)$. Для достаточно малых по модулю ν вектор $y_\nu = (J(\widehat{x}(\cdot),\widehat{u}(\cdot))+\nu, x_1)$ принадлежит соответствующей окрестности $F(0)$ (окрестность W в теореме 2). Тогда существует вектор $\overline{\alpha}(y_\nu) \in V$ такой, что $F(\overline{\alpha}(y_\nu)) = y_\nu$, или $J(x(\cdot,\overline{\alpha}(y_\nu),\overline{u}(\cdot)), u(\cdot,\overline{\alpha}(y_\nu),\overline{u}(\cdot))) = J(\widehat{x}(\cdot),\widehat{u}(\cdot)) + \nu$, $x(t_1,\overline{\alpha}(y_\nu),\overline{u}(\cdot)) = x_1$ и, кроме того, $|\overline{\alpha}(y_\nu)| \le K|\nu|$. Из этих соотношений следует (в силу того, что $x(\cdot,\overline{\alpha},\overline{u}(\cdot)) \to \widehat{x}(\cdot)$ в $C^1([t_0,t_1],\mathbb{R}^n)$ и, очевидно, $u(\cdot,\overline{\alpha},\overline{u}(\cdot)) \to \widehat{u}(\cdot)$ в $C([t_0,t_1],\mathbb{R}^r)$ при $\overline{\alpha} \to 0$), что в любой окрестности пары $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ найдется допустимая в задаче (P_7) точка, в которой значение функционала больше (при $\nu > 0$) или меньше (при $\nu < 0$), чем в $(\widehat{x}(\cdot),\widehat{u}(\cdot))$, в противоречие с тем, что $(\widehat{x}(\cdot),\widehat{u}(\cdot))$ — локальный экстремум.

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

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

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

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

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

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

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

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

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 (максимумом)* 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 $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$ $(J(x(\cdot)) \leq J(\widehat{x}(\cdot)))$. *Локальный экстремум* — это либо локальный минимум, либо локальный максимум.

Следствие 1 (Необходимые условия экстремума 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 экстремум 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 $x=x_1,\ \dot{x}_1=x_2,\dots,\dot{x}_{n-1}=x_n,\ \dot{x}_n=u$, задачу (P_7') можно записать как задачу Лагранжа

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

Простая проверка показывает, что если $\widehat{x}(\cdot)$ — локальный экстремум в задаче (P_7'), то вектор-функция $(\widehat{x}_1(\cdot),\dots,\widehat{x}_n(\cdot),\widehat{u}(\cdot))^T$ — локальный экстремум в данной задаче. Согласно общей теореме найдутся такие множители Лагранжа λ_0 и $p(\cdot)=(p_1(\cdot),\dots,p_n(\cdot))\in C^1([t_0,t_1],(\mathbb{R}^n)^*)$, не равные одновременно нулю, что для функции Лагранжа

$$\mathcal{L}=\int_{t_0}^{t_1}\biggl(\lambda_0L(t,x_1(t),\dots,x_n(t),u(t))+\sum_{k=1}^{n-1}p_k(t)(\dot{x}_k(t)-x_{k+1}(t))+$$
$$+p_n(t)(\dot{x}_n(t)-u(t))\biggr)\,dt$$

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. □

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

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

называется *изопериметрической задачей.*

Функция $x(\cdot)\in C^1([t_0,t_1],\mathbb{R}^n)$ называется *допустимой в задаче* (P_7''), если $\Gamma(x(\cdot))=\{(t,x(t),\dot{x}(t))\mid t\in[t_0,t_1]\}\subset G$, $\int_{t_0}^{t_1}f_i(t,x(t),\dot{x}(t))\,dt=\alpha_i$, $1\le i\le m$, и $x(t_i)=x_i$, $i=0,1$.

Допустимая функция $\widehat{x}(\cdot)$ называется *локальным минимумом (максимумом)* 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 $J(x(\cdot)) \geq J(\widehat{x}(\cdot))$ $(J(x(\cdot)) \leq J(\widehat{x}(\cdot)))$. *Локальный экстремум* — это либо локальный минимум, либо локальный максимум.

По аналогии с задачей с ограничениями типа равенств из §2 свяжем с задачей (P_7'') следующую функцию Лагранжа $L(t,x,\dot{x},\overline{\lambda}) = \sum_{i=0}^m \lambda_i f_i(t,x,\dot{x})$, где $\overline{\lambda} = (\lambda_0,\lambda_1,\ldots,\lambda_m)$ — набор множителей Лагранжа. Естественно ожидать, если $\widehat{x}(\cdot)$ — локальный экстремум, то для некоторого набора $\overline{\lambda}$ в $\widehat{x}(\cdot)$ выполнены необходимые условия экстремума в задаче $L(t,x(t),\dot{x}(t),\overline{\lambda}) \to \text{extr}$, $x(t_i) = x_i$, $i = 0,1$, которая есть простейшая задача вариационного исчисления, т. е. должно выполняться уравнение Эйлера. Точная формулировка такова.

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 экстремум 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

$$\int_{t_0}^{t_1} g_0(t,z(t),u(t))\,dt \to \text{extr}, \quad \dot{z} = \varphi(t,z,u), \quad x(t_0) = x_0, \quad x(t_1) = x_1,$$
$$y_i(t_0) = 0, \quad y_i(t_1) = \alpha_i, \quad i = 1,\ldots,m, \quad (i)$$

где $z = (x_1,\ldots,x_n,y_1,\ldots,y_m)^T$, $u = (u_1,\ldots,u_n)^T$, координатная запись дифференциальной связи $\dot{z} = \varphi(t,x,u)$ имеет вид: $\dot{x}_1 = u_1,\ldots,\dot{x}_n = u_n, \dot{y}_1 = f_1(t,x,u),\ldots,\dot{y}_m = f_m(t,x,u)$ и $g_0(t,z,u) = f_0(t,x,u)$. Несложная проверка показывает, что если $\widehat{x}$ — локальный экстремум в задаче (P_7''), то $(\widehat{z}(\cdot),\widehat{u}(\cdot))$, где $\widehat{x}(\cdot) = (\widehat{x}_1(\cdot),\ldots,\widehat{x}_n(\cdot),0,\ldots,0)^T$, — локальный экстремум в данной задаче.

Согласно основной теореме 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

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

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

Исторический комментарий. Как уже говорилось, вариационное исчисление зародилось в 18 веке в трудах Эйлера и Лагранжа. Эйлер доказал необходимые условия экстремума и в изопериметрической задачи, и в задаче со старшими производными. Лагранж изучал более общую задачу, ... 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 все эти результаты (лемму о пакете иголок не доказываем).

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). Также определяется дифференцируемость F в любой точке $x \in U \cap (\widehat{x}+\mathbb{R}^n_+)$ (если x принадлежит внутренности $U \cap (\widehat{x}+\mathbb{R}^n_+)$, то это обычная производная). Можно говорить о непрерывной дифференцируемости F на $U \cap (\widehat{x}+\mathbb{R}^n_+)$, которая (если F задается m функциями) равносильна непрерывности частных производных этих функций на $U \cap (\widehat{x} + \mathbb{R}^n_+)$ и строгая дифференцируемость отображения F в $\widehat{x}$ также вытекает (по теореме о среднем) из непрерывной дифференцируемости F на $U \cap (\widehat{x} + \mathbb{R}^n_+)$.

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*

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

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

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

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 $|x(y) - \widehat{x}| \le K|y - F(\widehat{x})|$.*

Для доказательства надо, фактически, дословно повторить рассуждения из аналогичного утверждения в §4 с $C = \mathbb{R}^n_+$.

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 $\alpha \ge 0$. 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 $(\tau_i,v_i)\in T\times U$, $\tau_1\le\ldots\le\tau_k$, 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 $\mathcal{N}=\{\,(\tau_i,v_i),\ 1\le i\le k\,\}$ 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

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

или $p(\tau)\varphi(\tau,\widehat{x}(\tau),v) - \lambda_0 f(\tau,\widehat{x}(\tau),v) \le p(\tau)\varphi(\tau,\widehat{x}(\tau),\widehat{u}(\tau)) - \lambda_0 f(\tau,\widehat{x}(\tau),\widehat{u}(\tau))$. 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 $J(x(\cdot)) \ge J(\widehat{x}(\cdot))$ $(J(x(\cdot)) \le J(\widehat{x}(\cdot)))$.

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

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

[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 $\widehat{L}_{\dot{x}\dot{x}}(t)$.

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

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

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

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

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 в §3. Доказательство, в терминах данного параграфа, заключается в том, что если $\widehat{x}$ — локальный минимум, то ноль есть локальный минимум для функции φ, определенной соотношением $(*)$ и тогда необходимо $\varphi'(0) = 0$. Расшифровка этого условия и приводит к уравнению Эйлера.

Теперь докажем условие Лежандра, расшифровывая необходимое условие минимума второго порядка: $\varphi''(0) \geq 0$. Согласно формуле для $\varphi''(0)$, выписанной выше, данное условие равносильно тому, что $Q(h(\cdot)) \geq 0$ для всех $h(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n)$ таких, что $h(t_0) = h(t_1) = 0$. Это означает, что функция $\widehat{h}(\cdot) = 0$ есть слабый локальный минимум в задаче

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

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, учитывая, что $\overline{h}^T(t)\widehat{L}_{\dot{x}x}(t)\dot{\overline{h}}(t) = \dot{\overline{h}}^T(t)\widehat{L}_{x\dot{x}}(t)\overline{h}(t)$ и интегрируя по частям ($\overline{h}(t_0) = \overline{h}(\tau) = 0$), получим

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

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

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

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

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

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

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

и условие минимума по $u(\cdot)$:

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

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 $u \mapsto \widetilde{h}^T(t)\widehat{L}_{\dot{x}x}(t)u + (1/2)u^T\widehat{L}_{\dot{x}\dot{x}}(t)u - p(t)\cdot u$ достигает минимума в точке $\dot{\widetilde{h}}(t)$. Следовательно, по теореме Ферма, ее производная в этой точке равна нулю:

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

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

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 такая окрестность G ее графика $\{\,(t,\widehat{x}(t)) \in \mathbb{R}^{n+1} \mid t \in [t_0,t_1]\,\}$, 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 $x(\tau,\lambda(\tau,\xi)) = \xi$.

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 (о построении центрального поля экстремалей)**.** *Пусть $\widehat{x}(\cdot) \in C^2([t_0,t_1],\mathbb{R}^n)$ — экстремаль в простейшей задачи классического вариационного исчисления, интегрант $L \in C^3(U,\mathbb{R}^{2n+1})$, где U — окрестность расширенного графика $\Gamma(\widehat{x}(\cdot)) = \{\,(t,\widehat{x}(t),\dot{\widehat{x}}(t)) \in \mathbb{R}^{2n+1} \mid t \in [t_0,t_1]\,\}$ и на $\widehat{x}(\cdot)$ выполнены усиленные условия Лежандра и Якоби. Тогда $\widehat{x}(\cdot)$ можно окружить центральным полем экстремалей.*

Доказательство. Проведем его для случая $n = 1$, чтобы не усложнять запись. Так как выполнено усиленное условие Лежандра, то найдется такая окрестность $U_1 \subset U$ расширенного графика $\Gamma(\widehat{x}(\cdot))$, что $L_{\dot{x}\dot{x}}(t,x,\dot{x}) > 0$ для всех $(t,x,\dot{x}) \in U_1$. Положим $\Phi(t,x,\dot{x}) = L^{-1}_{\dot{x}\dot{x}}(t,x,\dot{x})(L_x(t,x,\dot{x}) - L_{\dot{x}t}(t,x,\dot{x}) - L_{\dot{x}x}(t,x,\dot{x})\dot{x})$ и рассмотрим дифференциальное уравнение на $[t_0,t_1]$:

$$\ddot{x} = \Phi(t,x,\dot{x}). \qquad (i)$$

Функция $\widehat{x}(\cdot)$ является его решением. Действительно, поскольку $\widehat{x}(\cdot)$ — экстремаль, то

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

и деля обе части уравнения на $\widehat{L}_{\dot{x}\dot{x}}(t)$, получаем, что $\widehat{x}(\cdot)$ — решение (i).

Записывая (i) в виде системы $\dot{x} = y$, $\dot{y} = \Phi(t,x,\dot{x})$, воспользуемся стандартными теоремами из дифференциальных уравнений. По локально теореме существования и единственности найдется такое $\varepsilon > 0$, что $\widehat{x}(\cdot)$ продолжаемо на отрезок $[t_0-\varepsilon, t_1+\varepsilon]$. Далее, согласно глобальной теореме существования и единственности и дифференцируемой зависимости от начальных данных, для любой точки $t_* \in (t_0-\varepsilon, t_0)$ и любого достаточно малого по модулю λ существует единственное решение $x(\cdot,t_*,\lambda)$ уравнения (i) с начальными условиями: $x(t_*) = \widehat{x}(t_*)$, $\dot{x}(t_*) = \dot{\widehat{x}}(t_*) + \lambda$, определенное на всем отрезке $[t_0-\varepsilon, t_1+\varepsilon]$, и при этом, отображение $(t,t_*,\lambda) \mapsto x(t,t_*,\lambda)$ непрерывно дифференцируемо на $(t_0-\varepsilon, t_1+\varepsilon) \times (t_0-\varepsilon, t_0) \times V$, где V — некоторая окрестность нуля.

Функции $x(\cdot,t_*,\lambda)$ являются экстремалями. Действительно, $x(\cdot,t_*,\lambda)$ удовлетворяет (i) и тогда используя (ii) (но “справа налево”), получаем, что $x(\cdot,t_*,\lambda)$ удовлетворяет уравнению Эйлера:

$$-\frac{d}{dt}L_{\dot{x}}(t,x(t,t_*,\lambda),\dot{x}(t,t_*,\lambda)) + L_x(t,x(t,t_*,\lambda),\dot{x}(t,t_*,\lambda)) = 0. \qquad (iii)$$

Покажем, что $\widehat{x}(\cdot)$ окружена центральным полем экстремалей $x(\cdot,t_*,\lambda)$ с центром в $(t_*,\widehat{x}(t_*))$, если t_* достаточно близко к t_0.

Пусть фиксированы некоторые $t_* \in (t_0-\varepsilon, t_0)$ и $\bar{t} \in [t_0,t_1]$. В окрестности $(\bar{t},0)$ рассмотрим отображение Ψ, определенное формулой $\Psi(t,\lambda) = (t,x(t,t_*,\lambda))$ (для достаточно малых по модулю λ). Так как

$$\Psi'(\bar{t},0) = \begin{pmatrix} 1 & 0 \\ x_t(\bar{t},t_*,0) & x_\lambda(\bar{t},t_*,0) \end{pmatrix},$$

то это отображение обратимо, если $x_\lambda(\bar{t}, t_*, 0) \neq 0$. В этом случае по теореме об обратной функции найдется такая окрестность точки $\Psi(\bar{t}, 0) = (\bar{t}, \widehat{x}(\bar{t}))$, что для любой точки (τ, ξ) из этой окрестности существуют единственные $t(\tau, \xi)$ и $\lambda(\tau, \xi)$, для которых справедливо равенство

$$\Psi(t(\tau, \xi), \lambda(\tau, \xi)) = (\tau, \xi) \ \Rightarrow \ x(\tau, t_*, \lambda(\tau, \xi)) = \xi.$$

Если $x_\lambda(\bar{t}, t_*, 0) \neq 0$ для любого $\bar{t} \in [t_0, t_1]$, то график функции $\widehat{x}(\cdot)$, являющийся компактом, покрывается окрестностями точек $\Psi(\bar{t}, 0) = (\bar{t}, \widehat{x}(\bar{t}))$ с отмеченными свойствами. Выбирая конечное подпокрытие, получаем окрестность $G \subset U_1$ графика $\widehat{x}(\cdot)$, обладающую тем свойством, что для любой точки $(\tau, \xi) \in G$ существует единственная точка $\lambda(\tau, \xi)$ такая, что $x(\tau, t_*, \lambda(\tau, \xi)) = \xi$ и $x(t_*, t_*, \lambda) = \widehat{x}(t_*)$, т. е. построено центральное поле экстремалей. При этом, отображение $(\tau, \xi) \mapsto \lambda(\tau, \xi)$ дважды непрерывно дифференцируемо.

Итак, осталось показать, что $x_\lambda(t, t_*, 0) \neq 0$ для любого $t \in [t_0, t_1]$. Дифференцируя тождество (iii) по λ в нуле и обозначая через $h(t, t_*) = x_\lambda(t, t_*, 0)$, будем иметь

$$-\frac{d}{dt}(\widehat{L}_{\dot{x}\dot{x}}(t)\dot{h}(t, t_*) + \widehat{L}_{\dot{x}x}(t)h(t, t_*)) + \widehat{L}_{x\dot{x}}(t)\dot{h}(t, t_*) + \widehat{L}_{xx}(t)h(t, t_*) = 0. \qquad (iv)$$

Таким образом, $h(\cdot, t_*)$ есть решение уравнения (iv) (являющееся уравнением Якоби) с начальными условиями: $h(t_*, t_*) = 0$ и $\dot{h}(t_*, t_*) = 1$ в силу того, что $x(t_*, t_*, \lambda) = \widehat{x}(t_*)$ и $\dot{x}(t_*, t_*, \lambda) = \widehat{x}(t_*) + \lambda$.

Пусть $h(\cdot, t_0)$ — решение уравнения Якоби на $[t_0, t_1]$ с начальными условиями: $h(t_0, t_0) = 0$ и $\dot{h}(t_0, t_0) = 1$. Поскольку выполнено усиленное условие Якоби, то $h(\cdot, t_0)$ не обращается в ноль на $(t_0, t_1]$. Функция $h(\cdot, t_0)$ продолжается (по локальной теореме существования) на некоторый отрезок $[t_0 - \varepsilon_0, t_1]$ (можно считать, что $\varepsilon_0 \leq \varepsilon$). Согласно теореме о непрерывной зависимости от начальных данных, если $t_* \in (t_0 - \varepsilon_0, t_0]$, $t_* \to t_0$, то $h(\cdot, t_*) \to h(\cdot, t_0)$ в $C^1([t_0 - \varepsilon_0, t_1])$. Отсюда и из того, что $\dot{h}(t_0, t_0) = 1$ и $h(\cdot, t_0) \neq 0$ на $(t_0, t_1]$ будет следовать, что $x_\lambda(\cdot, t_*, 0) = h(\cdot, t_*) \neq 0$ на $(t_*, t_1]$ (и тем более на $[t_0, t_1]$) для t_* достаточно близких к t_0. □

Теорема (Достаточные условия сильного минимума в простейшей задаче)**.** *Пусть $\widehat{x}(\cdot) \in C^2([t_0, t_1], \mathbb{R}^n)$ — допустимая экстремаль в простейшей задаче, V — окрестность графика $\widehat{x}(\cdot)$ и интегрант $L \in C^3(V \times \mathbb{R}^n)$. Пусть, далее, на $\widehat{x}(\cdot)$ выполнены усиленные условия Лежандра и Якоби и для любых $(t, x) \in V$ функция $\dot{x} \mapsto L(t, x, \dot{x})$ выпукла на $\mathbb{R}^n$. Тогда $\widehat{x}(\cdot)$ — сильный локальный минимум в простейшей задаче.*

Доказательство. Снова, для простоты, считаем, что $n = 1$. Ясно, что выполнены условия теоремы о построении поля экстремалей и поэтому в некоторой окрестности G графика $\widehat{x}(\cdot)$ существует центральное поле экстремалей $\{x(\cdot, \lambda)\}$ с центром в точке $(t_*, \widehat{x}(t_*))$, $t_* < t_0$. 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

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

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

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

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)

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

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

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

[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 $C([t_0,t_1])$.

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

$$J(x(\cdot)) - J(\widehat{x}(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))\, dt - \int_{t_0}^{t_1} L(t, \widehat{x}(t), \dot{\widehat{x}}(t))\, dt =$$

$$= \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))\, dt - \int_{t_0}^{t_1} S'(t, \widehat{x}(t))\, dt = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))\, dt -$$

$$- (S(t_1, x_1) - S(t_0, x_0)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))\, dt -$$

$$- \int_{t_0}^{t_1} S'(t, x(t))\, dt = \int_{t_0}^{t_1} (L(t, x(t), \dot{x}(t)) - L(t, x(t), u(t, x(t))) -$$

$$- (\dot{x}(t) - u(t, x(t))) L_{\dot{x}}(t, x(t), u(t, x(t)))\, dt = \int_{t_0}^{t_1} \mathcal{E}(t, x(t), u(t, x(t)), \dot{x}(t))\, dt.$$

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.*

11. Выпуклая двойственность. Теорема Фенхеля–Моро. Двойственность в линейном программировании

В вопросах, связанных с изучением выпуклых объектов (выпуклых множеств, выпуклых функций и выпуклых экстремальных задач), важную роль играет феномен двойственности. Он заключается в том, что каждому выпуклому объекту можно сопоставить двойственный выпуклый объект, который тесно связан с исходным и совместное их исследование, как правило, бывает весьма плодотворным.

Двойственность выпуклых множеств. В основе этой двойственности лежит следующее наблюдение. Выпуклое замкнутое множество (в $\mathbb{R}^n$) допускает двойное описание: первое (на языке исходного пространства) есть просто определение выпуклости и замкнутости, второе (на языке двойственного пространства $(\mathbb{R}^n)^*$) есть следствие теоремы 1.2 и состоит в том, что это множество является пересечением всех полупространств его содержащих. Этому можно придать более наглядную аналитическую форму. Действительно, пусть A — выпуклое замкнутое множество. Для любого $x^* \in (\mathbb{R}^n)^*$, очевидно, справедливо включение $A \subset \{\, x \in \mathbb{R}^n \mid x^* \cdot x \leq \sup_{x \in A} x^* \cdot x \,\}$. Следовательно, $A \subset \bigcap_{x^* \in X^*} \{\, x \in \mathbb{R}^n \mid x^* \cdot x \leq$

$\sup_{x\in A} x^* \cdot x\,\}$. Обозначим множество справа через B и покажем, что $A = B$. Если это не так, то найдется элемент $x_0 \in B \setminus A$. Тогда по теореме 1.2 существует элемент $x_0^* \in (\mathbb{R}^n)^*$ такой, что $\sup_{x\in A} x_0^* \cdot x < x_0^* \cdot x_0$. Но это противоречит тому, что $x_0 \in B$ и поэтому $A = B$.

Обозначим через sA функцию на $(\mathbb{R}^n)^*$, определенную равенством $sA(x^*) = \sup_{x\in A} x^* \cdot x$. Ясно, что это выпуклая функция. Она называется *опорной функцией множества* A. Субдифференциал этой функции в нуле, по определению, имеет вид: $\partial sA = \{\, x \in \mathbb{R}^n \mid x^* \cdot x \le sA(x^*),\ \ \forall\, x^* \in (\mathbb{R}^n)^*\,\}$. Таким образом, равенство $A = B$ может быть записано так

$$A = \partial sA,$$

которое можно интерпретировать следующим образом. Выпуклому замкнутому множеству A сопоставляется двойственный объект — его опорная функция, а само множество восстанавливается по этой функции с помощью операции субдифференцирования.

Двойственность выпуклых функций. Напомним, что функция $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ называется выпуклой, если ее надграфик $\operatorname{epi} f = \{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R} \mid \alpha \ge f(x),\ f(x) < +\infty\,\}$ — выпуклое множество. Функция $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ называется *замкнутой*, если ее надграфик — замкнутое множество. Выясним теперь, как аналитически можно интерпретировать тот факт, что надграфик выпуклой замкнутой функции есть пересечение всех содержащих его полупространств.

Напомним, что функция $x \mapsto x^* \cdot x + \alpha$ на $\mathbb{R}^n$, где $x^* \in (\mathbb{R}^n)^*$ и $\alpha \in \mathbb{R}$, называется аффинной.

Теорема (о поточечной верхней грани аффинных функций). *Функция $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ является выпуклой и замкнутой тогда и только тогда, когда она есть поточечная верхняя грань некоторого семейства аффинных функций.*

Доказательство. Если f — поточечная верхняя грань семейства аффинных функций, то ее надграфик есть пересечение надграфиков этих функций, которые, очевидно, выпуклы и замкнуты и поэтому функция f выпукла и замкнута.

Обратно, пусть функция f выпукла и замкнута. Если $f(x) \equiv +\infty$, то она есть, например, поточечный предел констант. Пусть функция f не равна тождественно $+\infty$ и $x_0 \in \mathbb{R}^n$, $\alpha_0 \in \mathbb{R}$ такие, что $\alpha_0 < f(x_0)$. Ясно, что $(x_0, \alpha_0) \notin \operatorname{epi} f$. По теореме 1.2 найдется ненулевой вектор $(x^*, \gamma) \in (\mathbb{R}^n \times \mathbb{R})^*$ такой, что

$$\sup_{(x,\alpha)\in \operatorname{epi} f} (x^* \cdot x + \gamma\alpha) < x^* \cdot x_0 + \gamma\alpha_0. \qquad (i)$$

Заметим, что $\gamma \le 0$, ибо в противном случае, увеличивая α, пришли бы к противоречию с неравенством (i).

Пусть $f(x_0) < +\infty$. Подставляя точку $(x_0, f(x_0)) \in \operatorname{epi} f$ в (i), получаем, что $\gamma(\alpha_0 - f(x_0)) > 0$. Но $\alpha_0 - f(x_0) < 0$ и поэтому в данном случае $\gamma < 0$. Можно считать, что $\gamma = -1$ (деля, если необходимо, обе части неравенства (i) на $-\gamma$). Тогда, обозначая через c верхнюю грань в (i), это неравенство можно записать в виде двух неравенств

$$x^* \cdot x - c \le \alpha, \qquad x^* \cdot x_0 - c > \alpha_0, \quad \forall\ (x, \alpha) \in \operatorname{epi} f. \tag{ii}$$

Из первого при $\alpha = f(x)$ следует, что для аффинной функции $x \mapsto p(x) = x^* \cdot x - c$ выполняется неравенство $p(x) \le f(x)$ для всех $x \in \mathbb{R}^n$. Из второго, что $\alpha_0 < p(x_0)$ и значит, $\alpha_0 < p(x_0) \le f(x_0)$. Выбирая α_0 близко к $f(x_0)$, получаем, что во всех точках, где функция f конечна, она есть поточечный предел аффинных функций и при этом, не превосходящих f.

Пусть теперь $f(x_0) = +\infty$. Для доказательства теоремы в этом случае надо для любого $\alpha_0 \in \mathbb{R}$ построить аффинную функцию, значение которой в точке x_0 больше α_0. Пусть $\alpha_0 \in \mathbb{R}$. Если в (i) $\gamma < 0$ (как и выше, считаем тогда, что $\gamma = -1$), то из (ii) вытекает, что $p(x_0) = x^* \cdot x_0 - c > \alpha_0$ и все доказано. Если же $\gamma = 0$ (отделяющая гиперплоскость "вериткальна"), то (i) запишется так

$$x^* \cdot x - c \le 0, \qquad x^* \cdot x_0 - c > 0, \quad \forall\ (x, \alpha) \in \operatorname{epi} f. \tag{iii}$$

По доказанному существует аффинная функция p, которая всюду не превосходит f. Для каждого $\mu > 0$ рассмотрим аффинную функцию $x \mapsto p_\mu = p(x) + \mu(x^* \cdot x - c)$. Из первого соотношения в (iii) следует, что эта функция также всюду не превосходит f, а из второго, что $p_\mu(x_0) = p(x_0) + \mu(x^* \cdot x - c) > \alpha_0$ для достаточно больших μ. Итак, f есть поточечная верхняя грань семейства аффинных функций, ее не превосходящих. □

Придадим теперь этому утверждению аналитическую форму. Для этого введем некоторые понятия.

Пусть $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$. Функция $f^*\colon (\mathbb{R}^n)^* \to \overline{\mathbb{R}}$, определенная равенством $f^*(x^*) = \sup_{x \in \mathbb{R}^n}(x^* \cdot x - f(x))$, называется *сопряженной функцией к* f, а функция $f^{**}\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, определенная по правилу $f^{**}(x) = \sup_{x \in (\mathbb{R}^n)^*}(x^* \cdot x - f^*(x^*))$, называется *второй сопряженной функцией к* f.

Ясно, что функции f^* и f^{**} выпуклы и замкнуты, как верхнии грани аффинных функций. Из определений легко следует, что всегда $f^{**}(x) \le f(x)$ для всех $x \in \mathbb{R}^n$.

Следствие (Теорема Фенхеля–Моро)**.** *Функция* $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *выпукла и замкнута в том и только в том случае, когда* $f^{**} = f$.

Доказательство. Если $f^{**} = f$, то f выпукла и замкнута, так как таковой является f^{**}. Пусть f выпукла и замкнута. Если $f(x) \equiv +\infty$, то, очевидно, $f^{**}(x) \equiv +\infty$. Пусть $f \ne +\infty$. По доказанному существует такая аффинная функция $x \mapsto x^* \cdot x - \alpha$, что $x^* \cdot x - \alpha \le f(x)$ для всех $x \in \mathbb{R}^n$ или, равносильно, $\alpha \ge \sup_{x \in \mathbb{R}^n}(x^* \cdot x - f(x)) = f^*(x^*)$. Так как f — верхняя таких функций, то для всех $x \in \mathbb{R}^n$

$$f(x) = \sup_{\substack{x^* \in (\mathbb{R}^n)^*,\ \alpha \in \mathbb{R} \\ \alpha \ge f^*(x^*)}} (x^* \cdot x - \alpha). \tag{i}$$

Поскольку f не равна тождественно $+\infty$, то f^* нигде не обращается в $-\infty$ и поэтому в (i) вместо α можно взять $f^*(x^*)$. Следовательно,$f(x) = \sup_{x^* \in (\mathbb{R}^n)^*}(x^* \cdot x - f^*(x^*)) = f^{**}(x)$ для всех $x \in \mathbb{R}^n$. □

Двойственность экстремальных задач. Применим теорему Фенхеля–Моро к построению двойственных задач. Пусть $f\colon \mathbb{R}^n \to$

$\overline{\mathbb{R}}$. Рассмотрим задачу

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

Включим ее в серию “подобных” ей задач (или, как говорят, “возмутим” данную задачу). Точнее говоря, пусть функция $F\colon \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ такова, что $F(x, 0) = f(x)$ для всех $x \in \mathbb{R}^n$. Каждому $y \in \mathbb{R}^m$ сопоставим задачу

$$F(x, y) \to \min, \quad x \in \mathbb{R}^n. \tag{$\mathcal{P}_y$}$$

Семейство таких задач называется *возмущением задачи* $\mathcal{P}$, а функция $S\colon \mathbb{R}^m \to \overline{\mathbb{R}}$, сопоставляющая $y \in \mathbb{R}^m$ значение задачи $\mathcal{P}_y$, называется *S-функцией* данного семейства. Ясно, $S(0)$ — значение исходной задачи $(\mathcal{P})$.

Как уже было отмечено, $S^{**}(0) \le S(0)$, а если S-функция выпукла и замкнута, то по теореме Фенхеля–Моро $S^{**}(0) = S(0)$. Выпишем задачу, значением которой является величина $S^{**}(0)$. По определению $S^{**}(0) = \sup_{y^* \in (\mathbb{R}^m)^*}(-S^*(y^*))$. Далее

$$S^*(y^*) = \sup_{y \in \mathbb{R}^m}(y^* \cdot y - \inf_{x \in \mathbb{R}^n} F(x, y)) = \sup_{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}^m}}(0 \cdot x + y^* \cdot y - F(x, y)).$$

Справа стоит сопряженная функция к F в точке $(0, y^*)$. Таким образом, задача, значение которой равно $S^{**}(0)$ имеет вид

$$-F^*(0, y^*) \to \max, \quad y^* \in (\mathbb{R}^m)^*. \tag{$\mathcal{P}^*$}$$

Задача $(\mathcal{P}^*)$ называется *двойственной задачей к* $(\mathcal{P})$ (относительно заданного возмущения).

Задача линейного программирования. Выпишем двойственную задачу к задаче линейного программирования.

Пусть $c^* \in (\mathbb{R}^n)^*$, A — матрица размера $m \times n$ и $b \in \mathbb{R}^m$. Задачу

$$c^* \cdot x \to \min, \quad Ax \ge b, \quad x \ge 0^7 \tag{LP}$$

называют *задачей линейного программирования* (в нормальной форме).

В терминах общей постановки здесь $f(x) = c^* \cdot x$, когда $Ax \ge b$, $x \ge 0$ и $f(x) = +\infty$ в противном случае.

Выпишем двойственную задачу к (LP) относительно возмущения

$$c^* \cdot x \to \min, \quad Ax \ge b + y, \quad x \ge 0,$$

[7] Неравенства понимаются покоординатно.

где $y \in \mathbb{R}^m$ (т. е. $F(x,y) = c^* \cdot x$, когда $Ax \geq b + y$, $x \geq 0$ и $F(x,y) = +\infty$ в противном случае). Имеем

$$F(0,y^*) = \sup_{\substack{x\in\mathbb{R}^n \\ y\in\mathbb{R}^m}} (y^* \cdot y - F(x,y)) = \sup_{\substack{x\in\mathbb{R}^n \\ y\in\mathbb{R}^m}} (y^* \cdot y - \inf_{\substack{Ax\geq b+y \\ x\geq 0}} c^* \cdot x) =$$

$$= \sup_{\substack{Ax\geq b+y \\ x\geq 0,\ y\in\mathbb{R}^m}} (y^* \cdot y - c^* \cdot x) = \begin{cases} \sup_{x\geq 0}(y^* \cdot (Ax - b) - c^* \cdot x), & \text{если } y^* \geq 0; \\ +\infty, & \text{если не так.} \end{cases}$$

Далее

$$\sup_{x\geq 0}((y^*A - c^*) \cdot x - y^* \cdot b) = \begin{cases} -y^* \cdot b, & \text{если } y^*A - c^* \leq 0; \\ +\infty, & \text{если не так.} \end{cases}$$

Таким образом,

$$F(0,y^*) = \begin{cases} -y^* \cdot b, & \text{если } y^*A \leq c^*,\ \ y^* \geq 0; \\ +\infty, & \text{если не так} \end{cases}$$

и следовательно, двойственная задача имеет вид

$$y^* \cdot b \to \max, \quad y^*A \leq c^*, \quad y^* \geq 0. \qquad (LP^*)$$

Сформулируем в конце две теоремы, касающиеся существования решения в задаче линейного программирования и двойственных связей между задачами (LP) и (LP^*).

Теорема (существования). *Если в задаче (LP) $((LP^*))$ множество элементов не пусто и ее значение конечно, то эта задача имеет решение.*

Теорема (двойственности). *Для задач (LP) и (LP^*) справедлива следующая альтернатива: либо значения этих задач конечны и равны и в каждой из них существует решение, либо в одной из них множество допустимых элементов пусто , а в другой или множество допустимых элементов пусто, или ее значение бесконечно.*

В первом случае допустимые элементы $\widehat{x}$ и $\widehat{y}^*$ в задачах (LP) и (LP^*) являются их решениями тогда и только тогда, когда $c \cdot \widehat{x} = \widehat{y}^* \cdot b$.

Доказательство. Пусть x — допустимый элемент в задаче (LP). Тогда $c \cdot x \geq \widehat{y}^* Ax \geq \widehat{y}^* \cdot b = c \cdot \widehat{x}$ и значит, $\widehat{x}$ — решение этой задачи. Аналогично проверяется, что $\widehat{y}^*$ — решение (LP^*). □

12. Теорема существования Тонелли

Теорема Тонелли — это теорема о существовании решения в вариационных задачах. Здесь мы ограничимся рассмотрением простейшей задачи вариационного исчисления. Для доказательства теоремы понадобятся некоторые стандартные факты из функционального анализа. Напомним их.

Пусть X — нормированное пространство и X^* — его сопряженное, которое также является нормированным пространством с нормой $\|x^*\|_{X^*} = \sup_{\|x\|_X \le 1} |\langle x^*, x\rangle|$, где $\langle x^*, x\rangle$ — значение линейного функционала $x^* \in X^*$ на элементе $x \in X$.

Каждому $x \in X$ можно сопоставить линейный функционал на X^* по правилу: $x^* \mapsto \langle x^*, x\rangle$. *Слабой* топологией на X^** называется слабейшая топология, в которой эти функционалы непрерывны для всех $x \in X$. Отсюда следует, что если x_n^* сходится к x^* при $n \to \infty$ в слабо* топологии, то $\langle x_n^*, x\rangle \to \langle x^*, x\rangle$ при $n \to \infty$ для любого $x \in X$.

Теорема (Банаха–Алаоглу). *Ограниченные замкнутые множества в X^* компактны в слабо* топологии.*

Если пространство X сепарабельно (т. е. содержит счетное всюду плотное множество), то слабо* топология метризуема на слабо* компактных множествах. Это равносильно тому, что из любой последовательности элементов такого множества можно выделить подпоследовательность, слабо* сходящуюся к элементу данного множества.

Теорема (Арцела). *Для того, чтобы замыкание некоторого множества функций из $C([t_0, t_1])$ было компактом необходимо и достаточно, чтобы функции этого множества были равномерно ограничены и равностепенно непрерывны.*

Функция $x(\cdot)$ называется *абсолютно непрерывной* на отрезке $[t_0, t_1]$, если для любого $t \in [t_0, t_1]$ справедливо представление $x(t) = x_0 + \int_{t_0}^{t} z(\tau)\,d\tau$, где $z(\cdot)$ — суммируемая функция на $[t_0, t_1]$. При этом, $\dot{x}(t) = z(t)$ для п. в. $t \in [t_0, t_1]$.

Пусть $1 \le p < \infty$. Обозначим через $L_p([t_0, t_1])$ совокупность таких измеримых функций $x(\cdot)$ на отрезке $[t_0, t_1]$, что $\int_{t_0}^{t_1} |x(t)|^p\,dt < \infty$. Это нормированное пространство с нормой $\|x(\cdot)\|_p = \left(\int_{t_0}^{t_1} |x(t)|^p\,dt\right)^{1/p}$. Если $p > 1$ и $1/p' + 1/p = 1$, то пространство $L_{p'}([t_0, t_1])$ является сопряженным к $L_p([t_0, t_1])$ (в том смысле, что отображение, сопоставляющее $y(\cdot) \in L_{p'}([t_0, t_1])$ линейный непрерывный функционал $x(\cdot) \mapsto \int_{t_0}^{t_1} y(t)x(t)\,dt$ на$L_p([t_0, t_1])$, осуществляет изометрический изоморфизм $L_{p'}([t_0, t_1])$ и $(L_p([t_0, t_1]))^*$). Пространства $L_p([t_0, t_1])$ сепарабельны.

Пусть $1 < p < \infty$, $x(\cdot) \in L_p([t_0, t_1])$ и $y(\cdot) \in L_{p'}([t_0, t_1])$. Тогда справедливо *неравенство Гельдера*

$$\int_{t_0}^{t_1} x(t)y(t)\, dt \le \|x(\cdot)\|_p \|y(\cdot)\|_{p'}.$$

Пусть $1 \le p < \infty$. Обозначим через $W_p^1([t_0, t_1])$ пространство абсолютно непрерывных функций $x(\cdot)$ на $[t_0, t_1]$, для которых $\dot{x}(\cdot) \in L_p([t_0, t_1])$.

Рассмотрим простейшую задачу вариационного исчисления

$$J(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))\, dt \to \min, \quad x(t_0) = x_0, \quad x(t_1) = x_1. \tag{1}$$

Будем искать решение этой задачи на множестве абсолютно непрерывных функций $x(\cdot)$ на $[t_0, t_1]$ таких, что функция $t \mapsto f(t, x(t), \dot{x}(t))$ интегрируема.

Теорема (Тонелли)**.** *Пусть функция* $L\colon \mathbb{R}^3 \to \mathbb{R}$ *непрерывна на* $\mathbb{R}^3$ *вместе со своей частной производной по* $\dot{x}$*, для любых* t *и* x *функция* $\dot{x} \mapsto L(t, x, \dot{x})$ *выпукла и существуют такие константы* $\alpha > 0$*,* β *и* $p > 1$*, что* $L(t, x, \dot{x}) \ge \alpha|\dot{x}|^p + \beta$ *для всех* $(t, x, \dot{x})$*. Тогда в* (1) *существует решение, принадлежащее* $W_p^1([t_0, t_1])$ (*которое, в частности, абсолютно непрерывно*).

Доказательство. Пусть $\{x_n(\cdot)\}$ — минимизирующая последовательность в (1) и функции $x_n(\cdot)$, $n \in \mathbb{N}$, абсолютно непрерывны. Можно считать, что числовая последовательность $\{J(x_n(\cdot))\}$ монотонно убывает. Учитывая это обстоятельство, будем иметь после интегрирования оценки на рост L в условии теоремы

$$\int_{t_0}^{t_1} |\dot{x}_n(t)|^p\, dt \le \frac{1}{\alpha}\left(\int_{t_0}^{t_1} L(t, x_n(t), \dot{x}_n(t))\, dt - \beta(t_1 - t_0)\right) \le$$
$$\le \frac{1}{\alpha}\left(\int_{t_0}^{t_1} L(t, x_1(t), \dot{x}_1(t))\, dt - \beta(t_1 - t_0)\right) = C. \tag{i}$$

Покажем, что функции $x_n(\cdot)$, $n \in \mathbb{N}$, равномерно ограничены и равностепенно непрерывны. Действительно, так как

$$x_n(t) = x_0 + \int_{t_0}^{t} \dot{x}_n(\tau)\, d\tau \tag{ii}$$

для любого $t \in [t_0, t_1]$, то отсюда, применяя неравенство Гельдера и используя (i), получаем

$$|x_n(t)| \le |x_0| + \int_{t_0}^{t_1} |\dot{x}_n(\tau)|\, d\tau \le |x_0| + \left(\int_{t_0}^{t_1} |\dot{x}_n(\tau)|^p\, d\tau\right)^{1/p} \left(\int_{t_0}^{t_1} d\tau\right)^{1/p'} \le$$
$$\le |x_0| + C^{1/p}(t_1 - t_0)^{1/p'}.$$

Далее, по тем же соображениям, имеем для любого $t \in [t_0, t_1]$ и h таких, что $t + h \in [t_0, t_1]$

$$|x_n(t + h) - x_n(t)| \le \int_{t}^{t+h} |\dot{x}_n(\tau)|\, d\tau \le C^{1/p} h^{1/p'}.$$

Следовательно, по теореме Арцела из последовательности $\{x_n(\cdot)\}$ можно выделить подпоследовательность, сходящуюся в $C([t_0,t_1])$ к некоторой функции $\widehat{x}(\cdot)$. Будем считать, что сама последовательность $\{x_n(\cdot)\}$ сходится к этой функции.

В силу (i) последовательность $\{\dot{x}_n(\cdot)\}$ принадлежит замкнутому шару в $L_p([t_0,t_1])$ с центром в нуле радиуса $C^{1/p}$. Пространство $L_p([t_0,t_1])$ является сопряженным к сепарабельному пространству $L_{p'}([t_0,t_1])$ и поэтому по теореме Банаха–Алаоглу из последовательности можно выделить подпоследовательность, слабо* сходящуюся к некоторой функции $z(\cdot)\in L_p([t_0,t_1])$. Будем считать, что сама последовательность $\{\dot{x}_n(\cdot)\}$ слабо* сходится к $z(\cdot)$.

Для каждого $t\in[t_0,t_1]$ характеристическая функция $\chi_t(\cdot)$ отрезка $[t_0,t]$, очевидно, принадлежит $L_{p'}([t_0,t_1])$ и поэтому по неравенству Гельдера линейный функционал $x(\cdot)\mapsto\int_{t_0}^t x(\tau)\,d\tau$ непрерывен на $L_p([t_0,t_1])$. Тогда из того, что последовательность $\{\dot{x}_n(\cdot)\}$ слабо* сходится к $z(\cdot)$, а $\{x_n(\cdot)\}$ равномерно сходится к $\widehat{x}(\cdot)$, получаем, переходя к пределу в (ii), что

$$\widehat{x}(t)=x_0+\int_{t_0}^t z(\tau)\,d\tau.$$

Таким образом, функция $\widehat{x}(\cdot)$ абсолютно непрерывна, $\dot{\widehat{x}}(t)=z(t)$ для п. в. $t\in[t_0,t_1]$ и так как $z(\cdot)\in L_p([t_0,t_1])$, то $\widehat{x}(\cdot)\in W_p^1([t_0,t_1])$.

Покажем, что $\widehat{x}(\cdot)$ — решение задачи (1). Используя элементарные оценки и то, что из условия выпуклости следует неравенство $L(t,x_n(t),\dot{x}_n(t))-L(t,x_n(t),\dot{\widehat{x}}(t))\ge(\dot{x}_n(t)-\dot{\widehat{x}}(t))L_{\dot{x}}(t,x_n(t),\dot{\widehat{x}}(t))$ для всех $t\in[t_0,t_1]$ и $n\in\mathbb{N}$ (см. §9), будем иметь

$$\begin{gathered}
J(x_n(\cdot))-J(\widehat{x}(\cdot))=\int_{t_0}^{t_1}(L(t,x_n(t),\dot{x}_n(t))-L(t,x_n(t),\dot{\widehat{x}}(t)))\,dt+\\
+\int_{t_0}^{t_1}(L(t,x_n(t),\dot{\widehat{x}}(t))-L(t,\widehat{x}(t),\dot{\widehat{x}}(t)))\,dt\ge\int_{t_0}^{t_1}(\dot{x}_n(t)-\dot{\widehat{x}}(t))L_{\dot{x}}(t,x_n(t),\dot{\widehat{x}}(t))\,dt+\\
+\int_{t_0}^{t_1}(L(t,x_n(t),\dot{\widehat{x}}(t))-L(t,\widehat{x}(t),\dot{\widehat{x}}(t)))\,dt=\int_{t_0}^{t_1}((\dot{x}_n(t)-\dot{\widehat{x}}(t))(L_{\dot{x}}(t,x_n(t),\dot{\widehat{x}}(t))-\\
-L_{\dot{x}}(t,\widehat{x}(t),\dot{\widehat{x}}(t))))\,dt+\int_{t_0}^{t_1}(\dot{x}_n(t)-\dot{\widehat{x}}(t))L_{\dot{x}}(t,\widehat{x}(t),\dot{\widehat{x}}(t))\,dt+\\
+\int_{t_0}^{t_1}(L(t,x_n(t),\dot{\widehat{x}}(t))-L(t,\widehat{x}(t),\dot{\widehat{x}}(t)))\,dt. \qquad (iii)
\end{gathered}$$

Первый интеграл справа стремится к нулю при $n\to\infty$. Действительно, последовательность $\{\dot{x}_n(\cdot)-\dot{\widehat{x}}(\cdot)\}$ равномерно ограничена в $L_p([t_0,t_1])$, а второй сомножитель равномерно на $[t_0,t_1]$ стремится к нулю при $n\to\infty$ (так как $x_n(\cdot)$ сходится равномерно к $\widehat{x}(\cdot)$ и функция $L_{\dot{x}}$ непрерывна). Применяя теперь неравенство Гельдера, получаем, что интеграл сходится к нулю. Второй интеграл стремится к нулю, поскольку последовательность $\{\dot{x}_n(\cdot)\}$ слабо* сходится к $\dot{\widehat{x}}(\cdot)$. Третий интеграл сходится к нулю вследствие того, что подынтегральное выражение равномерно стремиться к нулю. Тогда, обозначая через S значение задачи (1), получаем, переходя к пределу в (iii), что $S-J(\widehat{x}(\cdot))\ge 0$, т. е. $S=J(\widehat{x}(\cdot))$ и значит, $\widehat{x}(\cdot)$ — решение данной задачи. □

Избранные задачи

1. Аэродинамическая задача Ньютона

В “Началах натуральной философии” (1687) Ньютон описал некую среду (типа разреженного воздуха), поставил и решил вопрос о том, когда *“тело, образующееся при вращении кривой вокруг оси при движении в упомянутой среде* [...] *будет испытывать меньшее сопротивление, нежели всякое иное тело вращения при той же высоте и наибольшей ширине”*.

Формализованная постановка этой задачи, как задачи оптимального управления, имеет вид

$$\int_0^{t_1} \frac{tdt}{1+u^2(t)} \to \min, \quad \dot{x}=u, \quad x(0)=0, \quad x(t_1)=x_1, \quad u \geq 0, \quad (1)$$

где $x_1 > 0$ и функция $x(\cdot)$ — абсолютно непрерывна.

Функция Лагранжа такова

$$\mathcal{L}(x(\cdot), u(\cdot), \overline{\lambda}) = \int_0^{t_1} L(t, x(t), \dot{x}(t), u(t), p(t), \lambda_0) dt,$$

где $L(t, x, \dot{x}, u, p, \lambda_0) = \lambda_0 t/(1+u^2) + p(\dot{x}-u)$ и $\overline{\lambda} = (\lambda_0, p(\cdot))$.

Согласно принципу максимума должно выполняться уравнение Эйлера по x:

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0 \;\Leftrightarrow\; -\dot{p} = 0 \;\Leftrightarrow\; p = \mathrm{const} \qquad (i)$$

и условие минимума u:

$$\min_{u \geq 0} \left(\frac{\lambda_0 t}{1+u^2} - pu \right) = \frac{\lambda_0 t}{1+\widehat{u}^2(t)} - p\widehat{u}(t). \qquad (ii)$$

Проверим,что $\lambda_0 \neq 0$. Если $\lambda_0 = 0$, то необходимо $p \neq 0$, так как множители Лагранжа не могут быть все нулевые. Но, если $p \neq 0$, то из (ii) следует, что $\widehat{u}(t) = 0$ п. в. и тогда $\widehat{x}(t_1) = \int_0^{t_1} \widehat{u}(t)dt = 0$ в противоречие с предположением. Итак, $\lambda_0 \neq 0$ и мы можем считать, что $\lambda_0 = 1$.

Отметим еще, что $p < 0$. Действительно, если $p \geq 0$, то при любом $t \geq 0$ функция $u \to f(t, u) = t/(1+u^2) - pu$ монотонно убывает и поэтому соотношение (ii) не может выполняться.

Легко видеть, что при малых t функция $f(t, \cdot)$ достигает минимума в нуле и что это происходит до того момента τ, когда значение в нуле становится равным второму, положительному минимуму этой функции, который достигается в точке $\widehat{u}(\tau) > 0$ (под $\widehat{u}(\tau)$ понимаем $\widehat{u}(\tau+0)$). Таким образом, τ определяется из условий: $f_u(\tau, \widehat{u}(\tau)) = 0$ и $f(\tau, \widehat{u}(\tau)) = f(\tau, 0) = \tau$, что приводит к уравнениям:

$$p = -\frac{2\tau\widehat{u}(\tau)}{(1+\widehat{u}(\tau)^2)^2}, \quad \tau = \frac{\tau}{1+\widehat{u}^2(\tau)} - p\widehat{u}(\tau). \qquad (iii)$$

Подставив одно из уравнений (iii) в другое, убеждаемся, что $\widehat{u}(\tau) = 1,\ \tau = -2p$.

После момента τ (момента излома управления) управление определяется как единственный корень уравнения $f_u(t,u)=0$, из которого следует, что

$$t=-\frac{p(1+u^2)^2}{2u}=-\frac{p}{2}\left(\frac{1}{u}+2u+u^3\right).$$

При этом

$$\frac{dx}{du}=\frac{dx}{dt}\frac{dt}{du}=u\frac{dt}{du}=-\frac{p}{2}\left(-\frac{1}{u}+2u+3u^3\right).$$

Интегрируя это уравнение с учетом того, что $x(\tau)=0$ и $\widehat{u}(\tau)=1$, приходим к семейству решений в параметрической форме:

$$x(t,p)=-\frac{p}{2}\left(\ln\frac{1}{u}+u^2+\frac{3}{4}u^4\right)+\frac{7}{8}p,\quad t=-\frac{p}{2}\left(\frac{1}{u}+2u+u^3\right),$$

где $p<0$ и p определяется из условия $x(t_1)=x_1$. Эту кривую называют *кривой Ньютона.*

Убедимся, что эта кривая доставляет абсолютный минимум в задаче. Пусть $x(\cdot)$ — допустимая функция в исходной задаче, т. е. она кусочно непрерывно дифференцируема, $x(0)=0$ и $x(t_1)=x_1$. Тогда в силу (ii) $(\lambda_0=1)$

$$\frac{t}{1+\dot{x}^2(t)}-p\dot{x}(t)\geq\frac{t}{1+\widehat{u}^2(t)}-p\widehat{u}(t).$$

Интегрируя это неравенство с учетом того, что $\int_0^{t_1}\dot{x}(t)dt=\int_0^{t_1}\dot{\widehat{x}}(t)dt=x_1$, получаем требуемое:

$$\int_0^{t_1}\frac{tdt}{1+\dot{x}^2(t)}\geq\int_0^{t_1}\frac{tdt}{1+\dot{\widehat{x}}^2(t)}.$$

Ответ. *Решение задачи Ньютона равно нулю до некоторого момента τ, определяемого граничным условием на правом конце, а далее оно идет по кривой Ньютона.*

2. Задача о гармоническом осцилляторе.

Согласно закону Гука, сила притяжения материальной точки пружиной пропорциональна уклонению точки. Второй закон Ньютона утверждает, что движение точки под воздействием силы пружины, описывается дифференциальным уравнением: $m\ddot{x}=F$ и тем самым $m\ddot{x}=-kx$. Соответствующий закон движения (как и вообще все законы классической механики) является следствием *принципа стационарного действия*, согласно которому траектория $t\mapsto x(t)$ материальной точки является стационарной точкой *функционала действия* $x(\cdot)\mapsto\int(m\dot{x}^2(t)/2-kx^2(t)/2)\,dt$.

Мы рассмотрим следующую задачу

$$\int_0^T (\dot{x}^2(t) - x^2(t))\, dt \to \min, \quad x(0) = x(T) = 0, \tag{2}$$

где $T > 0$.

Уравнению Эйлера имеет вид: $\ddot{x} + x = 0$. Его общее решение: $x(t) = C_1 \cos t + C_2 \sin t$. Если $T \neq k\pi, k \in \mathbb{N}$, то единственная экстремаль — тождественный нуль. Если $T = k\pi$, то имеется однопараметрическое семейство экстремалей $x(t, C) = C \sin t$.

Пусть $\widehat{x}(\cdot)$ — допустимая экстремаль. На этой экстремали выполнено усиленное условие Лежандра, так как $L_{\dot{x}\dot{x}} = 2$ и условие Вейерштрасса, поскольку функция $\dot{x} \mapsto \dot{x}^2 - x^2$ выпукла. Уравнение Якоби имеет тот же вид, что и уравнение Эйлера: $\ddot{h} + h = 0$.

Если $T > \pi$, то $\widehat{x}(\cdot)$ не может доставлять ни слабый, ни сильный минимум, поскольку на $(0, T)$ точка π является сопряженной к нулю и значит, не выполнено необходимое условие Якоби.

Если $T < \pi$, то для экстремали $\widehat{x}(\cdot) = 0$ выполнены, очевидно, все условия теоремы о достаточных условиях сильного минимума и, следовательно, $\widehat{x}(\cdot) = 0$ — сильный минимум.

Пусть $T = \pi$. Здесь можно поступить следующим образом. Рассмотрим задачу (2) на отрезке $[\varepsilon, T]$, где $0 < \varepsilon < T < \pi$. Тогда экстремаль $\widehat{x}(\cdot) = 0$ на этом отрезке окружена центральным полем экстремалей $\{C \sin t\}$ и она, конечно, по теореме о достаточных условиях есть сильный минимум. В данном случае легко выписать в явном виде основную формулу Вейерштрасса. Действительно, функция наклона поля $u(\tau, \xi) = \xi \operatorname{ctg} \tau$ и поэтому для любой допустимой экстремали $x(\cdot)$ имеем $\int_\varepsilon^T (\dot{x}^2(t) - x^2(t))\, dt = \int_\varepsilon^T (\dot{x}(t) - x(t) \operatorname{ctg} t)^2\, dt$. Справедливость данного равенства легко проверить и непосредственно (раскрывая в правой части квадрат и интегрируя по частям), и более того, оно остается верным, когда $\varepsilon = 0$, а $T = \pi$ и, следовательно, $\widehat{x}(\cdot) = 0$ — решение задачи (2) при $T = \pi$.

Отметим, что тождественный нуль на $[0, T]$ при $T \leq \pi$ есть решение задачи (2) следует из установленного равенства без обращения к общим результатам.

Отметим также, что если $T > \pi$, то подставляя в минимизируемый интеграл допустимую функцию $\alpha\overline{x}(\cdot)$, где $\overline{x}(t) = \sin \frac{\pi}{T} t$ и $\alpha \in \mathbb{R}$, легко проверить, что этот интеграл меньше $-\alpha^2$. Отсюда следует, в силу произвольности α, что значение задачи равно $-\infty$.

Ответ. *При $0 < T \leq \pi$ решением задачи о гармоническом осцилляторе является тождественный нуль. При $T > \pi$ значение задачи равно $-\infty$.*

3. Задача о быстродействии.

Эта задача была поставлена Фельдбаумом (1952) и заключается в *"определении характера движения в задаче о наискорейшей остановке в заданной точке управляемого транспортного средства, двигающегося из начальной точки с определенной скоростью (в начальный момент) по горизонтальным рельсам без трения и снабженного двигателем, развивающим силу, не превосходящую определенного предела"*.

1. *Формализация.* Пусть m — масса тележки , x_0 — ее начальная координата и v_0 — начальная скорость . Внешнюю силу (силу тяги) обозначим через u, а текущую координату тележки через $x(t)$. По закону Ньютона $m\ddot{x} = u$. Ограничение на тягу зададим в виде $u \in [u_1, u_2]$. Без ограничения общности можно считать, что $u_1 = -1$, $u_2 = 1$ и $m = 1$. Это приводит к следующей формализации задачи:

$$T \to \min, \quad m\ddot{x} = u, \quad |u| \le 1, \quad x(0) = x_0, \quad \dot{x}(0) = v_0,$$
$$x(T) = \dot{x}(T) = 0, \quad (3)$$

или в форме задачи оптимального управления (обозначая $x = \xi_1$, $v_0 = \xi_2$)

$$\int_0^T dt \to \min, \quad \dot{x}_1 = x_2, \quad \dot{x}_2 = u, \quad |u| \le 1,$$
$$x_1(0) = \xi_1, \quad x_2(0) = \xi_2, \quad x_1(T) = x_2(T) = 0.$$

Функция Лагранжа имеет вид

$$\mathcal{L} = \int_0^T \left(\lambda_0 + p_1(\dot{x}_1 - x_2) + p_2(\dot{x}_2 - u)\right) dt +$$
$$+ \lambda_1(x_1(0) - \xi_1) + \lambda_2(x_2(0) - \xi_2) + \lambda_3 x_1(T) + \lambda_4 x_2(T)$$

Согласно принципу максимума получаем условия стационарности поx:

$$-\frac{d}{dt}L_{\dot{x}_i} + L_{x_i} = 0,\ i = 1,2 \Rightarrow \dot{p}_1 = 0,\ \dot{p}_2 = -p_1 \Rightarrow p_2(t) = Ct + C_1;$$

условия трансверсальности:

$$p_1(0) = \lambda_1, \quad p_2(0) = \lambda_2, \quad p_1(T) = -\lambda_3, \quad p_2(T) = -\lambda_4;$$

условие минимума по u:

$$\min_{|u|\le 1}(-p_2(t)u) = -p_2(t)\widehat{u}(t) \Rightarrow \widehat{u}(t) = \operatorname{sign} p_2(t) \ \text{ при } \ p_2(t) \ne 0;$$

условие стационарность по T:

$$\mathcal{L}_T = 0 \Rightarrow \lambda_0 + \lambda_3\dot{x}_1(T) + \lambda_4\dot{x}_2(T) = 0.$$

Учитывая, что $\dot{x}_1(T) = 0$, $\lambda_4 = -p_2(T)$, $p_2(T)\widehat{u}(T) = |p_2(T)|$, получаем $\lambda_0 = |p_2(T)|$. Если допустить, что $\lambda_0 = 0$, то $p_2(T) = 0$, причем $|p_2(t)| \not\equiv 0$ (ибо иначе все множители Лагранжа были бы нули). Значит, $p_2(t) = C(t - T)$ а $\widehat{u}(t) \equiv +1$, либо $\widehat{u}(t) \equiv -1$. Множество начальных условий, соответствующих таким управлениям, описывается уравнением $\xi_2 = -\sqrt{2\xi_1}$, если $\xi_1 \geq 0$ и $\xi_2 = \sqrt{-2\xi_1}$, если $\xi_1 \leq 0$. Если же ξ_1 и ξ_2 не связаны этим соотношением, то $\lambda_0 \neq 0$ и полагаем $\lambda_0 = 1$. Тогда из из условий трансверсальности следует, что $|p_2(T)| = 1$, т. е. $p_2(t)$ равно либо $p_2^+(t) = C(t-T)+1$, либо $p_2^-(t) = C(t-T)-1$. Этим возможностям соответствуют такие управления (для некоторого $\tau \in (0,T)$): $u^+(t) = -1$, если $0 \leq t \leq \tau$ и $u^+(t) = 1$, если $\tau < t \leq T$, и $u^-(t) = 1$, если $0 \leq t \leq \tau$ и $u^-(t) = -1$, если $\tau < t \leq T$. Для тех значений t, где $\widehat{u}(t) = 1$, $\dot{x}_2 = 1 \Rightarrow \dot{x}_1 = t + C' \Rightarrow x_1 = \frac{t^2}{2} + C't + C'' = \frac{x_2^2}{2} + C$.

Аналогично получаем, что для тех значений t, для которых $\widehat{u}(t) = -1$, фазовая траектория — кусок параболы $x_1 = -\frac{x_2^2}{2} + C$. Это определяет движение траекторий по фазовой плоскости (см. рис. в АТФ).

Покажем, что найденное решение $\widehat{x}(\cdot) = \widehat{x}_1(\cdot)$, отвечающее заданной начальной точке (x_0, v_0), действительно доставляет решение задаче (3).

Предположим, что некоторая функция $x(\cdot)$ определена на отрезке $[0,\tilde{T}]$, имеет кусочно непрерывную вторую производную и $x(0) = x_0$, $\dot{x}(0) = v_0$, $x(\tilde{T}) = \dot{x}(\tilde{T}) = 0$, причем $\tilde{T} \leq T$. При $\tilde{T} < T$ мы доопределим $x(\cdot)$, положив $x(t) \equiv 0$ при $t \in [\tilde{T}, T]$. После этого обе функции $x(\cdot)$ и $\widehat{x}(\cdot)$ будут определены на одном и том же отрезке $[0,T]$ и будут иметь одинаковые граничные условия: $x(0) = \widehat{x}(0) = x_0$, $\dot{x}(0) = \dot{\widehat{x}}(0) = v_0$, $x(T) = \widehat{x}(T) = \dot{x}(T) = \dot{\widehat{x}}(T) = 0$.

Покажем, что если $|\ddot{x}| \leq 1$, то $x(\cdot) = \widehat{x}(\cdot)$ и, в частности, неравенство $\tilde{T} < T$ невозможно. Тем самым будет доказана оптимальность функции $\widehat{x}(\cdot)$. Ввиду симметрии задачи ограничимся случаем, когда вначале управление $u = \ddot{x}$ положительно. Если $|\ddot{x}| \leq 1$, то, интегрируя дважды неравенство $\ddot{x}(t) \leq 1$ и учитывая граничные условия, получаем

$$\widehat{x}(\tau) - x(\tau) = \int_0^\tau \int_0^t (1 - \ddot{x}(s))\, ds\, dt \geq 0, \tag{i}$$

причем равенство здесь возможно, только если во всех точках непрерывности $\ddot{x}(s) \equiv 1$, а тогда $x(t) \equiv \widehat{x}(t)$, $t \in [0,\tau]$.

Аналогично, интегрируя дважды неравенство $\ddot{x}(t) \geq -1$, получаем

$$\widehat{x}(\tau) - x(\tau) = \int_\tau^T \int_t^T (-1 - \ddot{x}(s))\, ds\, dt \leq 0, \tag{ii}$$

причем и здесь равенство возможно лишь, если $\ddot{x}(s) \equiv -1$ и $x(t) \equiv \widehat{x}(t)$, $t \in [\tau, T]$.

Сравнивая (i) и (ii), находим, что $x(\tau) = \widehat{x}(\tau)$, а тогда $x(t) \equiv \widehat{x}(t)$, $t \in [0,T]$.

Ответ. *Оптимальный процесс находится из условия* $|\ddot{x}(t)| = 1$ *во всех точках, кроме, может быть, одной, при переходе через которую* $\ddot{x}(\cdot)$ *меняет знак.*

Вопросы к экзамену

1. Дифференцируемость, непрерывная дифференцируемость и строгая дифференцируемость конечномерных отображений. Субдифференциал. 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 и правило множителей Лагранжа для конечномерной гладкой задачи с ограничениями, задаваемыми равенствами.

4. Модифицированный метод Ньютона и разрешимость конечномерной системы нелинейных уравнений для конуса.

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

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

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

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

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

10. Принцип максимума Понтрягина для понтрягинской задачи оптимального управления

11. Условия Лежандра, Якоби и Вейерштрасса. Необходимые условия сильного минимума для простейшей задачи вариационного исчисления.

12. Условия Лежандра, Якоби и Вейерштрасса. Необходимые условия слабого минимума для простейшей задачи вариационного исчисления.

13. Поле экстремалей. Теорема о построении центрального поля экстремалей.

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

15. Двойственность выпуклых объектов. Теорема Фенхеля–Моро.

16. Построение двойственной задачи к данной (относительно заданного возмущения). Двойственность в линейном программировании. Формулировка теорем существования и двойственности.

17. Теорема Тонелли существования решения в простейшей задаче вариационного исчисления.

18. Аэродинамическая задача Ньютона.

19. Задача о гармоническом осцилляторе.

20. Задача о быстродействии.