ОБЩИЙ ПРИНЦИП ЛАГРАНЖА

А.В. Дмитрук

Центральный экономико-математический институт РАН,
Кафедра оптимального управления факультета ВМК МГУ

(Совместная работа с Н.П. Осмоловским)

Конференция по теории приближений, теории экстремума
и оптимальному восстановлению,
посвященная 80-летию Г.Г. Магарил–Ильяева
13 июня 2024 года

Общая задача на экстремум с ограничениями равенства и неравенства:

$$
\begin{aligned}
&F_0(x) \to \min, \\
&f_i(x) \in K_i\,, \qquad i = 1, \ldots, \nu, \\
&g(x) \,=\, 0.
\end{aligned}
\tag{1}
$$

Здесь $\quad x \in X, \qquad F_0 : X \to \mathbf{R}, \quad f_i : X \to Z_i\,, \quad g : X \to Y,$

$X,\ Y,\ Z_i$ – банаховы пространства, $\quad i = 1, \ldots, \nu,$

$K_i \subset Z_i$ – замкнутые выпуклые конусы с непустой внутренностью.

Если все $Z_i = \mathbf{R}$, а $K_i = \mathbf{R}_-$, то это обычная *задача нелинейного программирования* с неравенствами $f_i(x) \leqslant 0$.

В общем случае каждое включения $f_i(x) \in K_i$ может задаваться бесконечным числом скалярных неравенств.

С другой стороны, все ограничения неравенства $f_i(x) \in K_i$ можно записать как одно ограничение $f(x) \in K$, если задать отображения

$$f : X \to Z = Z_1 \times \ldots \times Z_\nu, \qquad f(x) = (f_1(x), \ldots, f_\nu(x)).$$

Однако мы пока оставим задачу (1), чтобы сохранить визуальную связь с традиционными постановками.

Различные варианты и частные случаи этой задачи рассматривались многими авторами, начиная с конца 1950-х годов (Ф.Джон, Л.Гурвиц, Р.В.Гамкрелидзе и Г.Л.Харатишвили, Б.Н.Пшеничный, А.Я.Дубовицкий и А.А.Милютин, А.Д.Иоффе и В.М.Тихомиров, В.А.Якубович и А.С.Матвеев, П.П.Варайя, И.Нагахиса и И.Сакава, С.Курсьюч, Х.Маурер и Й.Цове, Д.О.Норрис, Б.Х.Порсио, Р.Т.Рокафеллар, Е.В.Тамминен, Дж.Ян, и др.). Однако точно такой постановки, кажется, не было. Она охватывает широкий класс задач оптимизации. Не претендуя на абсолютную новизну, дадим здесь необходимое условие локального минимума, которое очень просто формулируется и поэтому удобно для применения в конкретных задачах.

Типичное приложение – задачи оптимального управления с фазовыми ограничениями $\Phi(t, x(t)) \leqslant 0$, или со смешанными ограничениями $\varphi(t, x(t), u(t)) \leqslant 0$, которые можно трактовать как принадлежность к конусам неположительных функций в пространствах C и L_∞ соответственно.

Предположения

1) Целевая функция F_0 и отображения f_i дифференцируемы по Фреше в точке $x_0 \in X$;

оператор g строго дифференцируем в x_0 (предположения гладкости),

2) образ производной $g'(x_0)$ замкнут в Y (ослабленная регулярность ограничения равенства).

Хотя все отображения дифференцируемы, задача (1) не является стандартной гладкой задачей. Ее называют *полубесконечной*, (а лучше бы назвать *полугладкой*).

Пусть $K_i^0 := \{z_i^* \in Z_i^* : \langle z_i^*, K_i \rangle \leqslant 0\}$ есть полярный конус к K_i, $i = 1, \ldots, \nu$.

Правило множителей Лагранжа

Теорема 1. Пусть x_0 есть точка локального минимума в задаче (1). Тогда найдутся множители Лагранжа $\alpha_0 \geqslant 0$, $z_i^* \in Z_i^*$, $i = 1, \ldots, \nu$, и $y^* \in Y^*$, не все равные нулю, где $z_i^* \in K_i^0$ и $\langle z_i^*, f_i(x_0)\rangle = 0$, $i = 1, \ldots, \nu$ (каждый z_i^* есть внешняя нормаль к конусу K_i в точке $f_i(x_0)$), такие что функция Лагранжа

$$\mathcal{L}(x) = \alpha_0 F_0(x) + \sum_{i=1}^{\nu} \langle z_i^*, f_i(x)\rangle + \langle y^*, g(x)\rangle$$

стационарна в x_0:

$$\mathcal{L}'(x_0) = \alpha_0 F_0'(x_0) + \sum_{i=1}^{\nu} z_i^* f_i'(x_0) + y^* g'(x_0) = 0. \tag{2}$$

Это равенство называется уравнением Эйлера–Лагранжа.

а) Если все $Z_i = \mathbf{R}$ и $K_i = \mathbf{R}_-$, то $K_i^0 = \mathbf{R}_+$, и мы имеем классический результат со скалярными множителями $z_i^* = \alpha_i \geqslant 0$.

б) Равенства $\langle z_i^*, f_i(x_0)\rangle = 0, \quad i = 1, \ldots, \nu,$ есть условия дополняющей нежесткости. Если индекс i неактивный, т.е. $f_i(x_0) \in \operatorname{int} K_i$, то в данной точке автоматически внешняя нормаль $z_i^* = 0$.

в) В задачах с фазовыми и смешанными ограничениями

$$\Phi_i(t, x(t)) \leqslant 0 \qquad \text{и} \qquad \varphi_j(t, x(t), u(t)) \leqslant 0$$

в качестве множителей z_i^* получаем неотрицательную меру Лебега–Стилтьеса $d\mu_i(t)$, порожденную неубывающей функцией $\mu_i(t)$, и функционал $z_j^* \in L_\infty^*$, который в некотором *регулярном случае* задается интегрируемой функцией $\lambda_j(t) \geqslant 0$, и тогда функция Лагранжа содержит члены

$$\sum_i \int_{t_0}^{t_1} \Phi_i(t, x(t))\, d\mu_i(t) \; + \; \sum_j \int_{t_0}^{t_1} \varphi_j(t, x(t), u(t))\, \lambda_j(t)\, dt.$$

Доказательство проводится по схеме Дубовицкого–Милютина с использованием стандартных фактов функционального анализа:

теоремы Дубовицкого–Милютина о непересечении выпуклых конусов,

теоремы Фаркаша–Минковского о сопряженном к линейному прообразу выпуклого конуса,

леммы об аннуляторе ядра сюрьективного оператора,

теоремы Люстерника об оценке расстояния до множества нулей нелинейного оператора.

Теорема Дубовицкого–Милютина.

Пусть $\Omega_1, \ldots, \Omega_m, \Omega$ — непустые выпуклые конусы, из которых первые m открыты. Тогда пересечение всех этих конусов пусто в том и только том случае, когда существуют функционалы $x_1^* \in \Omega_1^*, \ldots, x_m^* \in \Omega_m^*$, $x^* \in \Omega^*$, не все равные нулю, сумма которых равна нулю:

$$x_1^* + \ldots + x_m^* + x^* = 0. \tag{3}$$

(Абстрактное уравнение Эйлера–Лагранжа.)

Фактически это эквивалент теоремы Хана–Банаха об отделимости.

Пусть X, Y — топологические векторные пространства,
$A : X \to Y$ — линейный непрерывный оператор.

Теорема Фаркаша–Минковского.
Пусть $K \subset Y$ есть выпуклый конус с непустой внутренностью,
$\Omega = \{x \in X : \; Ax \in K\} = A^{-1}K$ — его прообраз.
Предположим, что $\exists\, x_0$ такой что $Ax_0 \in \operatorname{int} K$.

Тогда $x^* \in \Omega^*$ в том и только том случае, когда существует $y^* \in K^*$ такой что $x^* = A^* y^*$. Другими словами, $\Omega^* = A^* K^*$.

(В "физических" обозначениях $x^* = y^* A$ и $\Omega^* = K^* A$.)

Пусть X и Y — банаховы пространства, $G \subset X$ — открытое множество, и $g : G \to Y$ — оператор, строго дифференцируемый в точке $x_0 \in G$. Пусть $g(x_0) = 0$.

Теорема Люстерника.
Пусть $g'(x_0)X = Y$ (условие регулярности). Тогда найдется окрестность $\mathcal{O}(x_0)$ и константа C, такие что для любого $x \in \mathcal{O}(x_0)$ существует поправка $h \in X$, такая что

$$g(x+h) = 0 \quad \text{и} \quad \|h\| \leqslant C\,\|g(x)\|. \tag{4}$$

Схема Дубовицкого–Милютина состоит из двух шагов.

Шаг 1: переход от локального минимума к несовместности системы конусных аппроксимаций.

Шаг 2: эта несовместность переписывается в виде уравнения Эйлера–Лагранжа.

Несовместность линеаризованной системы

Для каждого i введем конус внутренних направлений к конусу K_i в точке $z_i = f_i(x_0)$:

$$C_i = \{\bar z \in Z_i \mid \quad \exists\, \alpha > 0 : \quad f_i(x_0) + \alpha\, \bar z \in \operatorname{int} K_i\,\}.$$

Нетрудно видеть, что $C_i = \operatorname{int} K_i - \mathbf{R}_+ f_i(x_0) = \operatorname{int} K_i + \mathbf{R}\, f_i(x_0)$,
и этот конус непустой, выпуклый и открытый.
Его сопряженный есть

$$C_i^* \;=\; \{z_i^* \in K_i^* \;:\; \langle z_i^*, f_i(x_0)\rangle = 0\,\}\,.$$

Если i неактивный, т.е. $f_i(x_0) \in \operatorname{int} K_i$, то $C_i = Z_i$ и $C_i^* = \{0\}$.

Без нарушения общности считаем $F_0(x_0) = 0$.

Рассмотрим системы конусных аппроксимаций в точке x_0:

$$\left.\begin{aligned} & F_0'(x_0)\,\bar{x} < 0, \\ & f_i'(x_0)\,\bar{x} \in C_i\,, \quad i = 1, \ldots, \nu, \\ & g'(x_0)\,\bar{x} \, = \, 0. \end{aligned}\right\} \tag{5}$$

Лемма. Пусть в x_0 имеется локальный минимум, и $g'(x_0)X = Y$ (условие Люстерника). Тогда система (5) несовместна.

Доказательство. Допустим, такой $\bar{x}$ есть. При $\varepsilon \to 0+$ имеем $g(x_0 + \varepsilon\bar{x}) = g(x_0) + g'(x_0)\varepsilon\bar{x} + o(\varepsilon) = o(\varepsilon)$, и тогда по теореме Люстерника об оценке расстояния существует поправка $r_\varepsilon = o(\varepsilon)$, такая что $g(x_0 + \varepsilon\bar{x} + r_\varepsilon) = 0$.

Для любого $i = 1, \ldots, \nu$ имеем $f_i(x_0) \in K_i$ и $f_i(x_0) + \alpha f_i'(x_0)\bar{x} \in \operatorname{int} K_i$ при некотором $\alpha > 0$. Отсюда в силу выпуклости K_i, при малых $\varepsilon > 0$

$$f_i(x_0 + \varepsilon\bar{x} + r_\varepsilon) \;=\; f_i(x_0) + \varepsilon f_i'(x_0)\bar{x} + o(\varepsilon) \in \operatorname{int} K_i\,.$$

Аналогично получаем $F_0(x_0 + \varepsilon\bar{x} + r_\varepsilon) < F_0(x_0)$ для малых $\varepsilon > 0$, так как $F_0'(x_0)\bar{x} < 0$ и $\|r_\varepsilon\| = o(\varepsilon)$.

Итак, при малых $\varepsilon > 0$ точка $x_0 + \varepsilon\bar{x} + r_\varepsilon$ удовлетворяет всем ограничениям задачи и дает меньшее значение функционала, чем x_0, что противоречит локальному минимуму в x_0. $\square$

Переход к уравнению Эйлера–Лагранжа

Рассмотрим основной, невырожденный случай, когда $g'(x_0)X = Y$, кроме того, $F_0'(x_0) \neq 0$ и $\operatorname{Im} f_i'(x_0) \cap C_i \neq \varnothing$ для всех $i = 1, \ldots, \nu$, т.е. все конусы в системе (5) непусты.

По теореме Дубовицкого–Милютина о конусах найдется опорный функционал x_0^* к полупространству $\Omega_0 = \{\bar{x} : F_0'(x_0)\bar{x} < 0\}$, опорные функционалы x_i^* к конусам $\Omega_i = \{\bar{x} :\ f_i'(x_0)\bar{x} \in C_i\},\ i = 1, \ldots, \nu$, и опорный функционал x^* к подпространству $\Omega = \{\bar{x} : g'(x_0)\bar{x} = 0\}$, не все равные нулю, сумма которых равна нулю:

$$x_0^* + x_1^* + \ldots + x_\nu^* + x^* = 0.$$

Как известно, $x_0^* = -\alpha_0 F_0'(x_0)$ при некотором $\alpha_0 \geqslant 0$,
по теореме Фаркаша–Минковского каждый $x_i^* = z_i^* f_i'(x_0)$
при некотором $z_i^* \in K_i^*$ таком, что $\langle z_i^*, f_i(x_0)\rangle = 0$,
и по лемме об аннуляторе $x^* = y^* g'(x_0)$ при некотором $y^* \in Y^*$.

Таким образом,

$$-\alpha_0 F_0'(x_0) + \sum_{i=1}^{\nu} z_i^* f_i'(x_0) + y^* g'(x_0) = 0.$$

Ясно, что $\alpha_0 + \sum \|z_i^*\| + \|y^*\| > 0$ (иначе все опорные функционалы равны нулю).

Заменив все $z_i^* \mapsto -z_i^* \in K_i^0$, и $y^* \mapsto -y^*$, мы получаем уравнение Эйлера–Лагранжа в требуемом виде (18).

Вырожденные случаи, когда один из конусов Ω_i пуст (т.е. $F_0'(x_0) = 0$ или $\operatorname{Im} f_i'(x_0) \cap C_i = \emptyset$), или $g'(x_0)X \neq Y$, рассматриваются еще проще. Теорема доказана. $\square$

Отметим, что конусы K_i могут быть заданы неравенствами $\varphi_i(z) \leqslant 0,$ где все φ_i – сублинейные функции на пространствах Z_i , которые могут принимать значения < 0.

Т.е. задача (1) может иметь такой эквивалентный вид:

$$F_0(x) \to \min, \qquad \varphi_i(f_i(x)) \leqslant 0, \quad i = 1, \ldots, \nu, \qquad g(x) = 0. \tag{6}$$

Тогда конусы внешних нормалей $N_i(z) = \mathbf{R}_+ \, \partial\varphi_i(z)$, поэтому теперь в функции Лагранжа $z_i^* = \alpha_i \, \hat{z}_i^*$, где $\alpha_i \geqslant 0$, а $\hat{z}_i^* \in \partial\varphi_i(f_i(x_0))$ и $\langle \hat{z}_i^*, \, f_i(x_0)\rangle = 0$, т.е.

$$\mathcal{L}(x) = \alpha_0 F_0(x) + \sum_{i=1}^{\nu} \alpha_i \, \langle \hat{z}_i^*, \, f_i(x)\rangle + \langle y^*, \, g(x)\rangle, \tag{7}$$

и по-прежнему справедливо уравнение Эйлера–Лагранжа: $\mathcal{L}'(x_0) = 0$.

Здесь уже надо учитывать условия дополняющей нежесткости:

$$\alpha_i \, \varphi_i(f_i(x_0)) = 0, \qquad i = 1, \, \ldots, \, \nu,$$

а нетривиальным должен быть набор $(\alpha_0, \, \alpha_1, \, \ldots, \, \alpha_\nu, \, y^*)$.

Некоторые обобщения задачи (1)

Вместо конусов K_i могут фигурировать произвольные выпуклые множества Q_i с непустой внутренностью, т.е. задача может иметь вид

$$F_0(x) \to \min, \qquad f_i(x) \in Q_i\,, \quad i = 1, \ldots, \nu, \qquad g(x) = 0. \tag{8}$$

В этом случае функция Лагранжа по-прежнему есть

$$\mathcal{L}(x) = \alpha_0 F_0(x) + \sum_{i=1}^{\nu} \langle z_i^*, f_i(x) \rangle + \langle y^*, g(x) \rangle\,,$$

где $z_i^* \in N_i(f_i(x_0))$ – внешние нормали к мн-вам Q_i в точках $f_i(x_0)$.

Доказательство необходимого условия $\mathcal{L}'(x_0) = 0$ остается тем же самым.

Если Q_i заданы неравенствами $\varphi_i(z) \leqslant 0$, где все φ_i – выпуклые функции, которые могут принимать значения < 0, то, как и прежде, $N_i(z) = \mathbf{R}_+ \partial\varphi_i(z)$, т.е. функция Лагранжа имеет тот же вид (7), где по-прежнему $\alpha_i \geqslant 0$, $\hat{z}_i^* \in \partial\varphi_i(f_i(x_0))$, условия нетривиальности и дополняющей нежесткости прежние, и отсутствуют лишь соотношения $\langle \hat{z}_i^*,\, f_i(x_0)\rangle = 0$, справедливые в случае конусов.

Наконец, целевая функция F_0 так же, как и остальные, может иметь вид $F_0(x) = \varphi_0(f_0(x))$, где $f_0 : X \to Z_0$ есть гладкое отображение, а φ_0 – выпуклая функция на пространстве Z_0 .

В этом случае первый член в функции Лагранжа вместо $\alpha_0\, F_0(x)$ будет, подобно другим, $\alpha_0 \langle \hat{z}_0^*,\, f_0(x)\rangle$, где $\alpha_0 \geqslant 0$, а $\hat{z}_0^* \in \partial\varphi_0(f_0(x_0))$.

Отметим также, что эта постановка содержит как частный случай стандартную задачу выпуклого программирования: если все пространства $Z_i = X$, все $f_i = \mathrm{Id}$, а оператор g – аффинный, то мы имеем задачу

$$\varphi_0(x) \to \min, \qquad \varphi_i(x) \leqslant 0, \quad i = 1, \dots, \nu, \qquad Ax + b = 0. \tag{9}$$

Здесь

$$\mathcal{L}(x) = \alpha_0 \langle \hat{z}_0^*, x \rangle + \sum_{i=1}^{\nu} \alpha_i \langle \hat{z}_i^*, x \rangle + \langle y^*, Ax + b \rangle , \tag{10}$$

и уравнение Эйлера–Лагранжа

$$\mathcal{L}'(x_0) = \alpha_0 \, \hat{z}_0^* + \sum_{i=1}^{\nu} \alpha_i \, \hat{z}_i^* + y^* A = 0 \tag{11}$$

есть не что иное, как *субдифференциальная форма теоремы Каруша–Куна–Таккера*.

В этом “чисто выпуклом” случае функция Лагранжа (10) есть по сути опорная функция к “стандартной” функции Лагранжа задачи (7):

$$\widetilde{\mathcal{L}}(x) = \alpha_0\, \varphi_0(x) + \sum_{i=1}^{\nu} \alpha_i\, \varphi_i(x) + \langle y^*, Ax + b \rangle \,,$$

и равенство (9) означает, что $\partial \widetilde{\mathcal{L}}(x_0) \ni 0,$ т.е. что функция $\widetilde{\mathcal{L}}(x)$ (в силу ее выпуклости) достигает минимума по всему пространству X в данной точке x_0.

Некоторый недостаток теоремы в этом случае (как и у всех субдифференциальных условий) в том, что она работает только в пространствах с топологией, тогда как теорема о минимуме функции Лагранжа справедлива в произвольном векторном пространстве.

Вернемся на минуту к случаю задачи со скалярными неравенствами:

$$f_0(x) \to \min, \qquad f_i(x) \leqslant 0\,, \quad i = 1, \ldots, \nu, \qquad g(x) = 0.$$

Б.н.о. считаем, что в точке x_0 все $f_i(x_0) = 0, \quad i = 0, 1, \ldots, \nu.$

В конце 1970-х годов Левитин–Милютин–Осмоловский ввели следующее понятие: в точке x_0 выполнена $s-$необходимость, если существует окрестность $\mathcal{O}(x_0)$, в которой система строгих неравенств и равенства

$$f_0(x) < 0, \qquad f_i(x) < 0\,, \quad i = 1, \ldots, \nu, \qquad g(x) = 0. \qquad (12)$$

не имеет решений.

Ясно, что это есть необходимое условие локального минимума. Более того, это самое сильное необходимое условие из тех, которые симметричны относительно функционала и ограничений неравенства.

Так как все f_i теперь равноправны, можно не выделять f_0 , а рассматривать систему

$$f_i(x) < 0\,, \quad i = 1, \ldots, \nu, \qquad g(x) = 0, \tag{13}$$

и находить необходимые условия для ее локальной несовместности. Фактически, этим мы и занимались.

Обобщая эту ситуацию, будем рассматривать локальную несовместность системы

$$f_i(x) \in \operatorname{int} K_i\,, \quad i = 1, \ldots, \nu, \qquad g(x) = 0 \tag{14}$$

для наших прежних отображений $f_i : X \to Z_i$.

Теорема 2. Пусть в точке x_0 система (14) локально несовместна. Тогда найдутся множители Лагранжа $z_i^* \in Z_i^*$, $i = 1, \ldots, \nu$, и $y^* \in Y^*$, не все равные нулю, где $z_i^* \in K_i^0$ и $\langle z_i^*, f_i(x_0)\rangle = 0$, $i = 1, \ldots, \nu$ (каждый z_i^* есть внешняя нормаль к конусу K_i в точке $f_i(x_0)$), такие что функция Лагранжа

$$\mathcal{L}(x) = \sum_{i=1}^{\nu} \langle z_i^*, f_i(x)\rangle + \langle y^*, g(x)\rangle$$

стационарна в x_0:

$$\mathcal{L}'(x_0) = \sum_{i=1}^{\nu} z_i^* f_i'(x_0) + y^* g'(x_0) = 0. \tag{15}$$

Можно рассматривать локальную несовместность системы

$$f_i(x) \in \operatorname{int} Q_i\,, \quad i = 1, \ldots, \nu, \qquad g(x) = 0 \tag{16}$$

для произвольных выпуклых множеств Q_i с непустой внутренностью.

В этом случае функция Лагранжа по-прежнему имеет вид

$$\mathcal{L}(x) = \sum_{i=1}^{\nu} \langle z_i^*, f_i(x) \rangle + \langle y^*, g(x) \rangle\,,$$

где $z_i^* \in N_i(f_i(x_0))$ — внешние нормали к мн-вам Q_i в точках $f_i(x_0)$.

Теорема 2 по-прежнему справедлива.

Наконец, как уже было сказано, можно считать, что имеется лишь одно отображение $f : X \to Z$, одно выпуклое множество Q с непустой внутренностью, и рассматривается система

$$f(x) \in \operatorname{int} Q\,, \qquad g(x) = 0. \tag{17}$$

Здесь справедлива

Теорема 3. Пусть в точке x_0 система (17) локально несовместна. Тогда найдутся множители Лагранжа $z^* \in Z^*$ и $y^* \in Y^*$, не все равные нулю, где z^* есть внешняя нормаль к множеству Q в точке $f(x_0)$, такие что функция Лагранжа

$$\mathcal{L}(x) = \langle z^*, f(x)\rangle + \langle y^*, g(x)\rangle$$

стационарна в x_0:

$$\mathcal{L}'(x_0) = z^* f'(x_0) + y^* g'(x_0) = 0. \tag{18}$$

Все предыдущие теоремы вытекают из этой последней.

Литература

- А.Я. Дубовицкий, А.А. Милютин. Задачи на экстремум при наличии ограничений // ЖВМ и МФ — 1965. Т. 5, № 3. С. 395–453.
- А.А. Милютин, А.В. Дмитрук, Н.П. Осмоловский. *Принцип максимума в оптимальном управлении.* Мехмат МГУ, 2004 (168 с.)
- A.V. Dmitruk, N.P. Osmolovskii. A General Lagrange Multipliers Theorem and Related Questions // "Control Systems and Math. Methods in Economics" (G. Feichtinger et al. eds.), *Lecture Notes in Economics and Mathematical Systems, v. 687,* p. 165–194, Springer 2018.
- А.В. Дмитрук, Н.П. Осмоловский. Правило множителей Лагранжа для общей задачи на экстремум с бесконечным числом ограничений // Материалы 30-й КРОМШ, Симферополь, 2019, с. 213–216.
- A.V. Dmitruk, N.P. Osmolovskii. Lagrange Multipliers Rule for a General Optimization Problem with an Infinite Number of Constraints // Recent Advances of the Russian Operations Research Society (A. Vasin and F. Aleskerov eds.), Cambridge Scholars Publishing, 2020, p. 212–232.

СПАСИБО ЗА ВНИМАНИЕ!