СОДЕРЖАНИЕ И ТЕХНОЛОГИИ МАТЕМАТИЧЕСКОГО ОБРАЗОВАНИЯ В ВУЗЕ

УДК 517.27

НАГЛЯДНОЕ ГЕОМЕТРИЧЕСКОЕ ОБОСНОВАНИЕ МЕТОДА МНОЖИТЕЛЕЙ ЛАГРАНЖА

А. В. Бегунц, В. В. Галатенко, Д. В. Горяшин

Механико-математический факультет МГУ имени М. В. Ломоносова
Россия, 119991, г. Москва, Ленинские горы, Главное здание МГУ;
e-mail: ab@rector.msu.ru, vgalat@imscs.msu.ru, goryashin@mech.math.msu.su

Описывается наглядное обоснование метода множителей Лагранжа, являющегося основным методом для отыскания условных эстремумов функций. Достоинством описываемого обоснования является сохранение симметричности переменных, которая нарушается при традиционном доказательстве теоремы о необходимом условии условного экстремума.

Ключевые слова: математическое образование, условный экстремум, метод множителей Лагранжа, градиент.

Основным методом нахождения условного экстремума функций нескольких переменных при ограничениях типа равенств является метод множителей Лагранжа. Доказательство этого метода, традиционно излагаемое в учебной литературе (см., например, [1, гл. 6, § 3; 2, гл. 14, § 11]) и в лекционных курсах (см., например, [3, гл. 14]), оказывается довольно трудным для студентов: симметричные по своей природе условия доказываются рассуждением, в котором симметричность нарушается, и значительная часть студентов не чувствует естественности происхождения необходимых условий условного экстремума. По этой причине метод множителей Лагранжа воспринимается многими студентами только на «алгоритмическом» уровне, что приводит к бездумному его применению и к ошибкам, с этим связанным.

В настоящей заметке мы приведём наглядные геометрические рассуждения, показывающие естественность метода множителей Лагранжа и объясняющие природу самих этих множителей. Приводимые рассуждения можно или предпослать строгому доказательству (например, при работе со студентами математических специальностей и направлений подготовки), или же предложить в качестве идеи доказательства (например, студентам технических и естественнонаучных специальностей).

Важной методической особенностью предлагаемых рассуждений является отсутствие привязки к какой-либо системе координат. Этот факт особенно полезно подчеркнуть при работе со студентами естественнонаучных специальностей, например, физиками и механиками, указав, что в широком спектре прикладных задач функция (скалярное поле, например, поле температур) — это объективная данность, а система координат — инструмент, искусственно вводимый для работы с этой функцией (полем).

Базовые понятия, которые будут использоваться в излагаемом рассуждении, включают поверхность уровня, касательную (гипер)плоскость, производную по направлению и градиент функции нескольких переменных. В

случае функций трёх переменных эти понятия максимально естественны для восприятия: например, гиперплоскость будет стандартной двумерной плоскостью в трёхмерном пространстве.

Следует подчеркнуть идейную близость предлагаемых нами рассуждений к рассуждениям и пояснениям из книг [4, гл. 8, § 7; 5, гл. 5, § 43], также сопровождаемым наглядными иллюстрациями, но для случая функций двух переменных. В этом случае роль касательной плоскости играет касательная прямая. Рассмотрение трёхмерного случая представляется методологически более оправданным, так как при сохранении наглядности демонстрирует возможность выхода из анализируемой точки с сохранением условия связи по континууму направлений, а не исключительно по двум взаимно противоположным направлениям вдоль касательной прямой. В этом смысле случай трёх переменных (и одного условия связи) представляется более близким к общей ситуации, чем случай двух переменных, так как подчеркивает разницу между утверждениями «существует пара направлений» и «для любой пары направлений».

Начнём с того, что рассмотрим случай функции трёх переменных $f(x, y, z)$ и одного условия связи $\varphi(x, y, z) = 0$. Будем считать, что уравнение $\varphi(x, y, z) = 0$ задаёт гладкую поверхность Φ, точка (x_0, y_0, z_0) принадлежит этой поверхности и в окрестности этой точки функция f непрерывно дифференцируема (последнее требование можно существенно ослабить с сохранением схемы рассуждений, но в рамках текущей заметки это представляется излишним). Функция Лагранжа в этом случае имеет вид

$$L(x, y, z) = L_\lambda(x, y, z) = f(x, y, z) - \lambda\varphi(x, y, z)$$

(часто последний член в функции Лагранжа вводится со знаком плюс — см., например, [1, гл. 6, § 3] или [5, гл. 5, § 43], но на суть это не влияет: один вариант сводится к другому изменением знака множителя Лагранжа λ или же переходом от условия связи $\varphi(x, y, z) = 0$ к эквивалентному условию $-\varphi(x, y, z) = 0$). Необходимым условием наличия в точке (x_0, y_0, z_0) условного экстремума является существование такого λ, для которого в этой точке одновременно выполняются равенства

$$\frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} - \lambda\frac{\partial \varphi}{\partial x} = 0, \qquad \frac{\partial L}{\partial y} = \frac{\partial f}{\partial y} - \lambda\frac{\partial \varphi}{\partial y} = 0, \qquad \frac{\partial L}{\partial z} = \frac{\partial f}{\partial z} - \lambda\frac{\partial \varphi}{\partial z} = 0.$$

Одна из стандартных эквивалентных переформулировок этого условия заключается в том, что для наличия в точке (x_0, y_0, z_0) условного экстремума необходима коллинеарность в этой точке градиентов $\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$ и $\nabla\varphi = \left(\frac{\partial \varphi}{\partial x}, \frac{\partial \varphi}{\partial y}, \frac{\partial \varphi}{\partial z}\right)$.

Объясним наглядно, почему при нарушении условия коллинеарности этих градиентов точка (x_0, y_0, z_0) не может являться точкой условного экстремума. Самодостаточной иллюстрацией приводимого далее рассуждения является рис. 1.

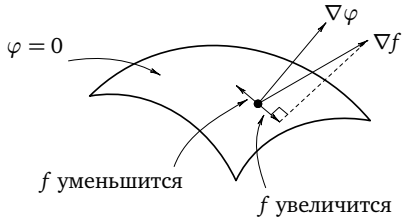

Рис. 1

В самом деле, в силу гладкости поверхности Φ имеем $\nabla\varphi(x_0, y_0, z_0) \neq 0$, а условие неколлинеарности дополнительно обеспечивает, что и $\nabla f(x_0, y_0, z_0) \neq 0$. Плоскость π, проходящая через точку (x_0, y_0, z_0) и ортогональная вектору $\nabla\varphi(x_0, y_0, z_0)$, является касательной для поверхности Φ. Эта плоскость описывает все возможные направления выхода из точки (x_0, y_0, z_0), не нарушающие условия связи (более формально это можно описать в терминах векторов скорости гладких кривых, лежащих на Φ и проходящих через точку (x_0, y_0, z_0)). Аналогично плоскость, проходящая через точку (x_0, y_0, z_0) и ортогональная вектору $\nabla f(x_0, y_0, z_0)$, является касательной для поверхности уровня функции f. Для любого направления, образующего с $\nabla f(x_0, y_0, z_0)$ острый угол, выход из точки (x_0, y_0, z_0) вдоль этого направления приводит к локальному увеличению значений функции f, так как её производная по этому направлению положительна (более формально это утверждение также формулируется в терминах гладких кривых, проходящих через точку (x_0, y_0, z_0), и функций одной переменной, являющихся ограничениями функции f на эти кривые). Наоборот, для любого направления, образующего с вектором $\nabla f(x_0, y_0, z_0)$ тупой угол, выход из точки (x_0, y_0, z_0) вдоль этого направления приводит к локальному уменьшению значений функции f.

Поскольку градиенты ∇f и $\nabla\varphi$ в точке (x_0, y_0, z_0) неколлинеарны, проекция $\nabla f(x_0, y_0, z_0)$ на плоскость π, ортогональную вектору $\nabla\varphi(x_0, y_0, z_0)$, является ненулевой и образует с вектором $\nabla f(x_0, y_0, z_0)$ острый угол. Таким образом, эта проекция представляет собой «разрешённое» направление выхода из точки (x_0, y_0, z_0) (то есть направление выхода без нарушения условия связи), при котором значения функции f локально увеличиваются. Следовательно, точка (x_0, y_0, z_0) не может быть точкой условного максимума. Но, аналогично, направление, противоположное этой проекции, представляет собой разрешённое направление выхода из точки (x_0, y_0, z_0), образующее с вектором $\nabla f(x_0, y_0, z_0)$ тупой угол. При выходе по этому направлению значения функции f локально уменьшаются и, следоватеьно, точка (x_0, y_0, z_0) не может быть точкой условного минимума.

Приведённые рассуждения можно перенести на большее число уравнений связи, а также на функции иного числа переменных. Например, рассмотрим добавление ограничения $\psi(x, y, z) = 0$, также задающего гладкую поверхность, содержащую точку (x_0, y_0, z_0), и имеющего в этой точке градиент $\nabla\psi$, неколлинеарный $\nabla\varphi(x_0, y_0, z_0)$. Разрешёнными направлениями выхода из (x_0, y_0, z_0) (с точки зрения сохранения условий связи) являются все направления, ортогональные одновременно векторам $\nabla\varphi(x_0, y_0, z_0)$ и $\nabla\psi(x_0, y_0, z_0)$. Если (x_0, y_0, z_0) является точкой условного экстремума, то все разрешённые направления должны быть ортогональны $\nabla f(x_0, y_0, z_0)$: в противном случае найдется пара разрешённых направлений, являющихся противоположными и образующих с $\nabla f(x_0, y_0, z_0)$ острый и тупой угол соответственно. Выход по первому направлению будет приводить к локальному увеличению значений f,

а по второму — к локальному уменьшению значений, что в совокупности противоречит предположению о наличии условного экстремума. Таким образом,

$$\nabla f(x_0, y_0, z_0) \perp \langle \nabla\varphi(x_0, y_0, z_0), \nabla\psi(x_0, y_0, z_0)\rangle^{\perp},$$

где угловые скобки обозначают линейную оболочку заключённых в них векторов, а показатель $\perp$ у линейного пространства обозначает его ортогональное дополнение. Другими словами,

$$\nabla f(x_0, y_0, z_0) \in \langle \nabla\varphi(x_0, y_0, z_0), \nabla\psi(x_0, y_0, z_0)\rangle,$$

то есть существуют такие величины λ_1 и λ_2 (как раз и представляющие собой множители Лагранжа), что

$$\nabla f(x_0, y_0, z_0) = \lambda_1 \nabla\varphi(x_0, y_0, z_0) + \lambda_2 \nabla\psi(x_0, y_0, z_0).$$

Следует особо подчеркнуть, что условие ортогональности ∇f и разрешённых (с точки зрения сохранения условий связи) направлений выхода из рассматриваемой точки эквивалентно тому, что все разрешённые направления лежат в касательной плоскости к поверхности уровня функции f (содержащей все направления, ортогональные ∇f), а это в точности и означает касание поверхности уровня функции f и многообразия, задаваемого условиями связи. Таким образом, на геометрическом уровне необходимым условием экстремума (в гладком случае) является касание.

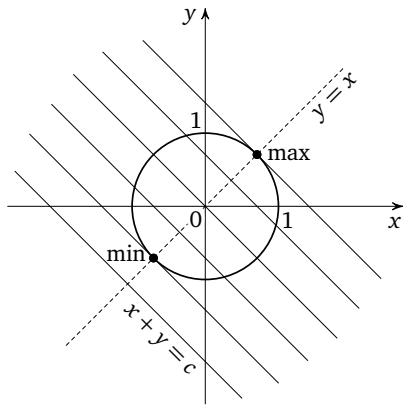

Рис. 2

Проиллюстрируем это следующим примером. Исследуем на экстремум функцию $f(x, y) = x + y$ при условии $x^2 + y^2 = 1$. Линии уровня функции f представляют собой прямые $x + y = c$ (см. рис. 2). Ровно две из них касаются окружности $x^2 + y^2 = 1$, при этом большему значению $c = \sqrt{2}$ соответствует точка $x = y = \dfrac{1}{\sqrt{2}}$, а меньшему значению $c = -\sqrt{2}$ — точка $x = y = -\dfrac{1}{\sqrt{2}}$. В данном случае мы имеем дело с функцией, непрерывной на компакте, а значит, достигающей на нем наибольшего и наименьшего значений. Из этого сразу следует, что первая из найденных точек является точкой условного максимума, а вторая — точкой условного минимума, и дополнительных обоснований этого не требуется.

В качестве идеи ещё одного способа рассуждений добавим, что в рассмотренном примере $\nabla f \equiv (1; 1)$, а $\nabla\varphi(x, y) = (2x; 2y)$, поэтому равенство $\nabla f = \lambda\nabla\varphi$ возможно лишь на прямой $x = y$, которая пересекает окружность $x^2 + y^2 = 1$ как раз в двух точках, найденных выше.

Наконец, можно проверить полученный результат традиционным способом. Составим функцию Лагранжа: $L = x + y - \lambda(x^2 + y^2 - 1)$. Точки условного

экстремума должны удовлетворять системе уравнений

$$\begin{cases} 1 - 2\lambda x = 0, \\ 1 - 2\lambda y = 0, \\ x^2 + y^2 = 1. \end{cases}$$

Получаем либо $x = y = \dfrac{1}{\sqrt{2}}$, $\lambda = \dfrac{1}{\sqrt{2}}$, либо $x = y = -\dfrac{1}{\sqrt{2}}$, $\lambda = -\dfrac{1}{\sqrt{2}}$. Несмотря на то, что, как указывалось выше, для выяснения характера найденных точек достаточно найти в них значение функции f, мы из методических целей составим второй дифференциал функции Лагранжа. Будем рассматривать одновременно обе точки, зафиксировав в каждой из них значение λ. Тогда получим $d^2L = -2\lambda(dx^2 + dy^2)$. Это означает, что точка, которой соответствует положительное значение λ, является условным максимумом, а точка, которой соответствует отрицательное значение λ — условным минимумом.

Подчеркнём, что именно равенство $\nabla f = \lambda \nabla \varphi$ обеспечивает возможность использования в качестве необходимого условия условного экстремума функции $f(x, y, z)$ при условии связи $\varphi(x, y, z) = 0$ необходимого условия безусловного экстремума (равенства нулю частных производных по всем переменным) функции Лагранжа $L(x, y, z) = f(z, y, z) - \lambda\varphi(x, y, z)$, рассматриваемой при данном значении множителя Лагранжа. Это же относится и к случаю иного количества переменных или уравнений связи. Здесь явно видна разница между переменными и множителями Лагранжа, которые следует рассматривать как параметры — несмотря на удобство считать функцию Лагранжа зависящей от них как от переменных (это позволяет записать уравнение связи в виде равенства нулю частной производной L по λ). Особенно ясно это проявляется при проверке достаточного условия условного экстремума, а именно, при нахождении второго дифференциала функции Лагранжа, при котором L рассматривается как функция именно от переменных (x, y, z), при этом значение λ остается фиксированным для исследуемой на условный экстремум точки (подробнее см., например, [6, гл. 8, § 8.8]).

Предложенная схема обоснования метода множителей Лагранжа позволяет наряду с задачей поиска экстремумов при ограничениях типа равенств обсудить аналогичную задачу для случая ограничений типа неравенств. В качестве одного из вариантов рассмотрим задачу нахождения наибольшего и наименьшего значения функции f на множестве, заданном неравенствами вида $\varphi_k \geqslant 0$ $(k = 1, \ldots, m)$. Задачи такого типа являются стандартными в различных прикладных областях — в качестве примера можно привести задачу квадратичной оптимизации, возникающую в одном из классических методов машинного обучения — методе опорных векторов [7].

Подход к поиску условных экстремумов при ограничениях типа неравенств, стандартно обсуждаемый в рамках курса математического анализа, состоит в следующем: отдельно рассмотреть внутренность разрешённого множества точек ($\varphi_k > 0$, $k = 1, \ldots, m$) и осуществить там поиск безусловных экстремумов, и отдельно рассмотреть границу, осуществив на ней стандартный поиск условных экстремумов с ограничениями типа равенств. Послед-

нее рассмотрение, в случае $m > 1$, также разбивается на отдельные подслучаи, определяемые тем, какие из ограничений реализуются в граничном виде ($\varphi_k = 0$), а какие — нет ($\varphi_k > 0$).

Оказывается, что по крайней мере на теоретическом уровне все эти случаи могут быть записаны в виде единой системы уравнений. Более конкретно, согласно теории Куна – Таккера (см., например, [8, гл. 6]), для решения сформулированной задачи можно рассмотреть функцию Лагранжа

$$L(x_1, \dots x_n) = f(x_1, \dots x_n) - \sum_{k=1}^{m} \lambda_k \varphi_k(x_1, \dots x_n)$$

и для поиска возможных точек экстремума составить систему уравнений, приравняв к нулю производные по переменным, но заменив условия связи (неформально — совпадающие с условиями равенства нулю производных по λ_k) так называемыми *условиями дополняющей нежёсткости*

$$\lambda_k \varphi_k(x_1, \dots x_n) = 0 \quad (k = 1, \dots, m).$$

Суть условий дополняющей нежёсткости может быть очень легко пояснена. Если рассматриваемая точка лежит на границе, т. е. $\varphi_k(x_1, \dots x_n) = 0$, то соответствующее слагаемое $\lambda_k \varphi_k$ участвует в функции Лагранжа так же, как и в случае классических ограничений типа равенств. Если же точка не лежит на этой границе, т. е. $\varphi_k(x_1, \dots x_n) > 0$, то это неравенство выполняется и в целой окрестности. Следовательно, в проводимом локально анализе ограничение $\varphi_k \geqslant 0$ автоматически выполняется и может не учитываться, что и достигается удалением φ_k из функции Лагранжа за счёт равенства $\lambda_k = 0$.

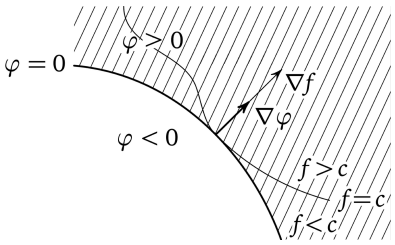

Рис. 3

Обсудим смысл знака множителя Лагранжа λ в случае одного ограничения $\varphi \geqslant 0$. Составим функцию Лагранжа $L = f - \lambda\varphi$ и рассмотрим точку, лежащую на границе ($\varphi = 0$) и удовлетворяющую необходимому условию условного экстремума $\nabla f = \lambda \nabla \varphi$. Если $\lambda > 0$, то ∇f и $\nabla \varphi$ сонаправлены, и при сдвиге в этом направлении имеет место локальное увеличение значений и функции f, и функции φ (см. рис. 3). Значит, можно так сдвигать точку внутрь «разрешённого» множества (то есть множества, задаваемого неравенством $\varphi \geqslant 0$), что при этом значение f увеличивается. Это означает, что рассматриваемая точка для функции f не может быть точкой условного максимума (при ограничении типа неравенств). Аналогично, если $\lambda < 0$, то соответствующая точка не может быть точкой условного минимума.

Таким образом, при наличии ограничений типа неравенств для отыскания наибольших (наименьших) значений функции можно учитывать дополнительное необходимое условие $\lambda_k \leqslant 0$ ($\lambda_k \geqslant 0$), сокращая перебор.

Проиллюстрируем это на простом примере, уже обсуждавшемся выше. Рассмотрим задачу поиска точки глобального минимума функции

$f(x,y) = x + y$ на множестве $1 - x^2 - y^2 \geqslant 0$. Запишем функцию Лагранжа: $L = x + y - \lambda(1 - x^2 - y^2)$. Необходимые условия экстремума принимают вид

$$\begin{cases} 1 + 2\lambda x = 0, \\ 1 + 2\lambda y = 0, \\ \lambda(1 - x^2 - y^2) = 0. \end{cases}$$

При $\lambda = 0$ решений нет, поэтому и наибольшее, и наименьшее значения достигаются на границе. При $\lambda \neq 0$, решая систему, находим $x = y = -\dfrac{1}{2\lambda}$, $\dfrac{1}{2\lambda^2} = 1$. Поскольку при поиске минимума достаточно ограничиться случаем $\lambda > 0$, получаем единственное значение $\lambda = \dfrac{1}{\sqrt{2}}$ и, соответственно, единственную точку $x = y = -\dfrac{1}{2\lambda}$. С учетом компактности, гарантирующей, что точка минимума существует, можно сразу утверждать, что найденная точка ей и является.

В качестве несложного вопроса предлагаем читателю подумать над переносом связи между характером экстремума и знаком множителя Лагранжа на случай большего числа ограничений типа неравенств. Например, можно ли в случае двух ограничений $\varphi_1 \geqslant 0$, $\varphi_2 \geqslant 0$ утверждать, что в точке условного максимума ни один из множителей Лагранжа λ_1 и λ_2 не может быть положительным, а в точке условного минимума ни один из множителей Лагранжа не может быть отрицательным (и, соответственно, если множители Лагранжа имеют противоположные знаки, то рассматриваемая точка заведомо не является точкой условного экстремума)?

Подводя итог, сформулируем основную мысль данной заметки. Авторам представляется важным независимо от специфики направления обучения не сводить объяснение метода множителей Лагранжа исключительно к технической и алгоритмической сторонам. Приведённые рассуждения, являясь достаточно краткими и, следовательно, не требующими в рамках курса значительного времени на их обсуждение, могут помочь студентам разных специальностей не только лучше понять сам метод множителей Лагранжа, но и почувствовать общие идеи, стоящие за этим методом.

ЛИТЕРАТУРА

1. Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления. Том I. — М.: Физматлит, 2001. 616 с.
2. Архипов Г. И., Садовничий В. А., Чубариков В. Н. Лекции по математическому анализу. — М.: Дрофа, 2004. 640 с.
3. Теляковский С. А. Курс лекций по математическому анализу. Семестр II // Лекц. курсы НОЦ. 2011. Т. 17. С. 3–192.
4. Зорич В. А. Математический анализ. Часть I. — М.: МЦНМО, 2017. 576 с.
5. Кудрявцев Л. Д. Курс математического анализа. Том II. — М.: Высшая школа, 1981. 584 с.

6. Виноградова И. А., Олехник С. Н., Садовничий В. А. Математический анализ в задачах и упражнениях: В 3-х т. Том 1: Дифференциальное и интегральное исчисление. Новое изд. — М.: Изд-во Московского университета; МЦНМО, 2017. 412 с.
7. Cortes C., Vapnik V. Support-Vector Networks // Mach. Learn. 1995. Vol. 20, issue 3. P. 273–297.
8. Хедли Дж. Нелинейное и динамическое программирование. — М.: Мир, 1967. 506 с.

Поступила 30.08.2017

A DEMONSTRATIVE GEOMETRIC REASONING FOR THE METHOD OF LAGRANGE MULTIPLIERS

A. V. Begunts, V. V. Galatenko, D. V. Goryashin

We describe a demonstrative scheme of reasoning for the method of Lagrange multipliers, which is the basic method for finding constrained extrema of functions. Along with intuitive clearness, the described scheme has a benefit of preserving symmetry of variables, which is violated in the conventional proof of a theorem on conditions necessary for constrained extremum.

Keywords: mathematics education, constrained extremum, the method of Lagrange multipliers, gradient.