Наш семинар: математические сюжеты

О сумме логарифмически выпуклых функций

Академик РАН Д. В. Аносов

Предлагается простое доказательство логарифмической выпуклости суммы логарифмически выпуклых функций.

1. В этой статье все функции принимают вещественные значения и определены на замкнутом, полуоткрытом или открытом интервале I, который может быть конечным или бесконечным (включая и случай всей числовой прямой $\mathbb{R}$); иными словами, I — связное подмножество $\mathbb{R}$.

Как известно, функция f называется выпуклой, если для любого отрезка $[a,b] \subset I$ график ограничения $f\big|_{[a,b]}$ функции f на этот отрезок лежит ниже отрезка, соединяющего концы $(a; f(a))$ и $(b; f(b))$ этого графика. Подробнее и точнее это сокращённое выражение, которым я буду пользоваться и далее, означает, что для любого $x \in [a,b]$ точка $(x; f(x))$ графика лежит не выше точки указанного отрезка с той же абсциссой. Аналитически данное условие записывается так:

$$f((1-t)a+tb) \leqslant (1-t)f(a)+tf(b) \qquad \text{при всех } t \in [0,1]. \tag{1}$$

Функция f называется логарифмически выпуклой, если она принимает только положительные значения и её логарифм $\log f(x)$ является выпуклой функцией. (Выбор основания логарифмов здесь не играет роли, ибо $\log_{c_1} f = \log_{c_1} c_2 \log_{c_2} f$. Но, конечно, подразумевается, что основание > 1.) Это же можно выразить так: для любого $[a,b] \subset I$

$$f((1-t)a+tb) \leqslant f(a)^{1-t} f(b)^t \qquad \text{при всех } t \in [0,1]. \tag{2}$$

На геометрическом же языке можно сказать, что график $f\big|_{[a,b]}$ лежит ниже графика экспоненциальной функции, принимающей те же самые значения, что и f, в концах отрезка. (Под экспоненциальной функцией сейчас понимается не только «чистая экспонента» $e^{\beta x}$, но и та же экспонента с положительным постоянным множителем, т.е. функция вида $\alpha e^{\beta x}$, где $\alpha > 0$.) Графики ограничений всевозможных экспонент на всевозможные $[a, b]$ образуют некоторое семейство дуг $\mathcal{E}$ в полуплоскости $\mathbb{R}\times(0,\infty)$. Для любых двух точек этой полуплоскости с различными абсциссами имеется ровно одна дуга из $\mathcal{E}$ с концами в этих точках. Логарифмически выпуклые функции суть те, для которых графики их ограничений на всевозможные $[a, b] \subset I$ лежат ниже дуг из $\mathcal{E}$ с теми же концами. В определении выпуклых функций аналогичную роль играет семейство $\mathcal{L}$ всех невертикальных отрезков (графиков линейных функций) в $\mathbb{R}^2$ (при этом семейство $\mathcal{L}$ обладает в $\mathbb{R}^2$ свойством, аналогичным указанному выше свойству семейства $\mathcal{E}$ в $\mathbb{R}\times(0,\infty)$: любые две точки плоскости с различными абсциссами являются концами ровно одного отрезка из $\mathcal{L}$).

Каждая линейная функция выпукла, а каждая экспонента — логарифмически выпукла. Кроме того, каждая экспонента выпукла (это простое упражнение по анализу). Отсюда сразу следует, что каждая логарифмически выпуклая функция f выпукла (график $f\big|_{[a,b]}$ лежит под графиком соответствующей экспоненты, а тот — под соответствующим отрезком). Для экспоненты $f = \alpha e^{\beta x}$ с $\beta \neq 0$ любая точка графика $f\big|_{[a,b]}$, кроме его концов, лежит строго ниже точки соответствующего отрезка с той же абсциссой. Линейные функции, кроме констант, не являются логарифмически выпуклыми. Пересечение $\mathcal{L} \cap \mathcal{E}$ состоит из горизонтальных отрезков.

Совершенно очевидно, что сумма выпуклых функций выпукла; отсюда следует, что произведение логарифмически выпуклых функций логарифмически выпукло. Отнюдь не столь очевидна

Теорема 1. *Если f, g логарифмически выпуклы, то $f + g$ тоже логарифмически выпукла.*

Эта теорема давно известна, но излагаемое ниже её доказательство вполне может быть новым; довольно уверенно можно сказать, что в литературе на русском языке оно не публиковалось. Докажем сперва другую теорему.

Теорема 2. *Функция f логарифмически выпукла тогда и только тогда, когда при всех $\alpha \in \mathbb{R}$ функции $e^{\alpha x} f(x)$ выпуклы.*

Доказательство теоремы 2. В одну сторону утверждение очевидно: $\log(e^{\alpha x} f(x)) = \alpha x + \log f(x)$ отличается от $\log f(x)$ на линейную функцию, так что $\log(e^{\alpha x} f(x))$ и $\log f(x)$ выпуклы или не выпуклы одновременно. А если функция $e^{\alpha x} f(x)$ логарифмически выпукла, то она тем более выпукла.

Обратно, пусть дано, что при всех $\alpha \in \mathbb{R}$ функции $e^{\alpha x} f(x)$ выпуклы. Пусть $[a, b] \subset I$. Условие (2) выполняется тогда и только тогда, когда аналогичное условие выполняется для функции $e^{\alpha x} f(x)$ (ведь $e^{\alpha((1-t)a+tb)} = (e^{\alpha a})^{1-t}(e^{\alpha b})^t$; можно также воспользоваться логарифмированием). Существует такое α, что $e^{\alpha a} f(a) = e^{\alpha b} f(b)$ (именно,

$$\alpha = \frac{1}{b-a} \ln \frac{f(a)}{f(b)}; \tag{3}$$

можно также сослаться на указанное выше свойство семейства $\mathcal{E}$, взять дугу, соединяющую концы графика функции $f\big|_{[a,b]}$, и разделить f на соответствующую экспоненту). Так как концы дуги $y = e^{\alpha x} f(x), \quad x \in [a, b]$ лежат на одинаковой высоте, то соединяющая их дуга из $\mathcal{E}$ — это горизонтальный отрезок. А раз функция $e^{\alpha x}f(x)$ выпукла, то её график лежит под этим отрезком. Здесь последний выступает как дуга из $\mathcal{L}$, но он же является и дугой из $\mathcal{E}$, соединяющей концы этого графика. Вот и выходит, что для $e^{\alpha x}f(x)$ выполняется аналог условия (2). Аналитически то же выражается формулами

$$e^{\alpha((1-t)a+tb)}f((1-t)a+tb) \leqslant (1-t)e^{\alpha a}f(a) + te^{\alpha b}f(b) = e^{\alpha a}f(a) = \\ = \left(e^{\alpha a}f(a)\right)^{1-t}\left(e^{\alpha b}f(b)\right)^{t},$$

где сперва использована выпуклость $e^{\alpha x}f(x)$, а затем — что $e^{\alpha a}f(a) = e^{\alpha b}f(b)$.

Теорема 2 даёт своего рода характеризацию логарифмической выпуклости с помощью выпуклости. Для логарифмической выпуклости функции f необходима её выпуклость, но этого далеко не достаточно (линейные функции). Для доказательства логарифмической выпуклости надо как бы подвергнуть f бесконечному числу «тестов» — она должна удовлетворять (1) (со всевозможными a, b) не только сама, но и после умножения на экспоненту. Из доказательства видно, почему это так — при одном из этих «тестов» (отвечающем (3)) условие (1) для $e^{\alpha x}f(x)$ оказывается совпадающим с условием (2).

Доказательство теоремы 1. Мы хотим доказать, что в предположениях этой теоремы функция $e^{\alpha x}(f(x)+g(x))$ выпукла. Но она является суммой двух логарифмически выпуклых и тем более выпуклых функций $e^{\alpha x}f(x)$, $e^{\alpha x}g(x)$.

2. Здесь я остановлюсь на имеющемся в литературе доказательстве теоремы 1.

Когда f дважды дифференцируема, то условие её выпуклости состоит в том, что всюду $f'' \geqslant 0$, а условие логарифмической выпуклости — в том, что всюду $f''f - f'^2 \geqslant 0$. Для дважды дифференцируемых логарифмически выпуклых f, g утверждение теоремы 1 состоит в том, что всюду $(f+g)''(f+g) - (f'+g')^2 \geqslant 0$. Оно сводится к такому чисто алгебраическому утверждению: если $c, \gamma > 0$, $ac \geqslant b^2$ и $\alpha\gamma \geqslant \beta^2$, то $(a+\alpha)(c+\gamma) \geqslant (b+\beta)^2$. (Мы принимаем $c = f$ и $\gamma = g$, поэтому c, γ положительны.) Для доказательства последнего достаточно заметить, что неравенство $ac \geqslant b^2$ является при $c > 0$ необходимым и достаточным условием неотрицательности квадратного трёхчлена $ax^2 + 2bx + c$ при всех x. После того как мы перефразируем аналогичным образом неравенства $\alpha\gamma \geqslant \beta^2$ и $(a+\alpha)(c+\gamma) \geqslant (b+\beta)^2$, остаётся заметить, что сумма неотрицательных трёхчленов неотрицательна. Можно и не обращаться к квадратным трёхчленам, а рассуждать так. Если $a = 0$ или $\alpha = 0$, то утверждение тривиально (почему?). Если же $a \neq 0$ и $\alpha \neq 0$, то

$$(a+\alpha)(c+\gamma) - (b+\beta)^2 = \underline{ac} + \alpha c + a\gamma + \underline{\underline{\alpha\gamma}} - \underline{b^2} - 2b\beta - \underline{\underline{\beta^2}} \geqslant \\ \geqslant \alpha c + a\gamma - 2b\beta = a\alpha\left(\frac{c}{a} + \frac{\gamma}{\alpha}\right) - 2b\beta \geqslant a\alpha\left(\left(\frac{b}{a}\right)^2 + \left(\frac{\beta}{\alpha}\right)^2\right) - 2b\beta = \\ = \frac{\alpha}{a}b^2 + \frac{a}{\alpha}\beta^2 - 2b\beta = \left(\sqrt{\frac{\alpha}{a}}b\right)^2 + \left(\sqrt{\frac{a}{\alpha}}\beta\right)^2 - 2\left(\sqrt{\frac{\alpha}{a}}b\right)\left(\sqrt{\frac{a}{\alpha}}\beta\right) = \\ = \left(\sqrt{\frac{\alpha}{a}}b - \sqrt{\frac{a}{\alpha}}\beta\right)^2 \geqslant 0.$$

Здесь сперва использовано, что подчёркнутые члены с плюсом не меньше подчёркнутых членов с минусом, затем — что неравенство $ac \geqslant b^2$ влечёт положительность a (ввиду $a \neq 0$, $c > 0$) и эквивалентно неравенству $\frac{c}{a} \geqslant \frac{b^2}{a^2}$, а также аналогичное соображение для α, β, γ.

В общем случае можно вывести теорему 1 из того же самого алгебраического утверждения. Непосредственно с его помощью доказывается, что если $f, g > 0$,

$$f\left(\frac{a+b}{2}\right)^2 \leqslant f(a)f(b) \tag{4}$$

и аналогично для g, то и

$$\left(f\left(\frac{a+b}{2}\right) + g\left(\frac{a+b}{2}\right)\right)^2 \leqslant (f(a)+g(a))(f(b)+g(b)). \tag{5}$$

Это (меняя обозначения) снова сводится к тому, что

$$\text{из} \quad c, \gamma > 0, \ ac \geqslant b^2, \ \alpha\gamma \geqslant \beta^2 \quad \text{следует} \quad (a+\alpha)(c+\gamma) \geqslant (b+\beta)^2$$

(только в данном случае можно с самого начала считать, что все числа $a, b, c, \alpha, \beta, \gamma$ больше 0). Надо, конечно, пояснить, почему частного случая (2) (с $f+g$ вместо f), отвечающего $t = \frac{1}{2}$, достаточно для заключения о логарифмической выпуклости $f+g$. Опять меняя обозначения, мы приходим к вопросу: можно ли из частного случая (1), отвечающего $t = \frac{1}{2}$, т.е. из того, что

$$2f\left(\frac{a+b}{2}\right) \leqslant f(a) + f(b) \quad \text{при } [a,b] \subset I, \tag{6}$$

сделать вывод о выпуклости f?

Из (6) легко следует, что

$$f((1-t)a+tb) \leqslant (1-t)f(a) + tf(b) \quad \text{при двоично-рациональных } t \in [a,b].$$

Перейти к любым $t \in [a,b]$ можно, если известно, что f непрерывна. Точнее, достаточно, чтобы f была непрерывна во внутренних точках I, а в конце A интервала I, если $A \in I$, достаточно выполнения условия

$$\varlimsup_{x \to A} f(x) \leqslant f(A). \tag{7}$$

Но сравнительно легко доказать, что выпуклая функция непрерывна внутри I, а если I содержит какой-нибудь свой конец A, то в нём выполняется (7). Значит, то же верно и для логарифмически выпуклой функции. Получается, что в условиях теоремы 1 f и g можно считать непрерывными внутри I и удовлетворяющими условию (7) в конце интервала I, если этот конец содержится в I. Стало быть, то же самое справедливо и для $f+g$. А тогда нам достаточно (5).

Ради полноты я приведу доказательство непрерывности f внутри I и свойства (7) для выпуклой f. Достаточно доказать два утверждения:

а). Если $[a,b] \subset I$, то $f(a) \geqslant \varlimsup\limits_{x \to a, x \in [a,b]} f(x)$ и аналогично для $f(b)$.

б). Если a — внутренняя точка I, то $f(a) \leqslant \varliminf\limits_{x \to a} f(x)$.

К а). Для $x \in [a,b]$ имеем $x = (1-t(x))a + t(x)b$, где $t = \frac{x-a}{b-a} \to 0$ при $x \to a$.

А $f(x) \leqslant (1-t(x))f(a)+t(x)f(b)$. Правая часть при $x\to a$ стремится к $f(a)$, а верхний предел левой есть $\overline{\lim\limits_{x\to a, x\in[a,b]}} f(x)$.

К б). При достаточно малых h точки $a\pm h\in I$. При этом $2f(a)\leqslant f(a-h)+f(a+h)$. Следовательно,

$$2f(a)\leqslant \varlimsup_{h\to 0} f(a-h)+\varliminf_{h\to 0} f(a+h)\leqslant f(a)+\varliminf_{x\to a} f(x)$$

(здесь использовано, что ввиду а) $\varlimsup\limits_{h\to 0} f(a-h)\leqslant f(a)$). Отсюда получается б).

Вместе взятые, рассуждения п. 2 заметно длиннее, чем п. 1. Однако в любом случае, говоря о выпуклых функциях, надо остановиться на их непрерывности, так что эту часть п. 2 можно не учитывать при оценке длины доказательства теоремы 1. А тогда данное в п. 2 доказательство этой теоремы оказывается не длиннее, чем в п. 1. Но оно кажется формальнее: какое отношение вспомогательные квадратные трёхчлены имеют к нашей задаче? А если обходиться без них, то вообще получается какой-то алгебраический трюк. Мы не видим «движущих пружин» доказательства. В п. 1 же наши рассуждения были непосредственно связаны с геометрией задачи. Всё это отдаёт субъективизмом, но думаю, что многие (как и я) найдут рассуждения п. 1 более прозрачными.

Доказательство, приведённое в п. 2, имеется в книжке Э. Артина о Г-функции [1], а также у И. И. Привалова [2] и Н. Бурбаки [3]. В последней оно приводится (только для дважды дифференцируемых функций, что по существу не проще) при изложении свойств Г-функции, воспроизводящем текст Артина. В теореме 1 легко перейти от суммы к интегралу, что доставляет логарифмическую выпуклость Г-функции. Это свойство несколько облегчает изучение последней, на что впервые обратили внимание Г. Бор и Й. Моллеруп в своём учебнике анализа [4] (пп. 39, 41, а также пример 6 на с. 161. Впоследствии этот учебник переиздавался, но в новых изданиях — по крайней мере тех, которые я видел, — о логарифмической выпуклости Г-функции не говорилось. Возможно, авторы сочли, что после выхода книжки Артина писать об этом незачем). В [4] не было утверждения о логарифмической выпуклости суммы, а сразу доказывалась логарифмическая выпуклость интеграла, что делалось так же, как позднее в [1] доказывалось первое утверждение. Выделение первого утверждения как естественно предшествующего второму — это, по-видимому, усовершенствование Артина.

Сомнительно, чтобы вопрос о логарифмической выпуклости (как в общем случае суммы логарифмически выпуклых функций или интеграла от таковых, так и в специальном случае Г-функции) мог возникнуть в XIX веке, так что рассуждения, имеющиеся у Бора – Моллерупа – Артина, скорее всего, им и принадлежат.

И. И. Привалов сообщает, что он узнал доказательство теоремы 1 от А. И. Плеснера. Последний незадолго до того эмигрировал в СССР из Германии, где он вполне мог узнать это доказательство от самого Артина или из его книги.

Теорема 2 принадлежит П. Монтелю [5]. Его доказательство отлично от нашего. Та же теорема имеется в [2]. Привалов ссылается на [5], но приводит другое доказательство. В конечном счёте его основная идея — та же, что и в п. 1

(подбор подходящей экспоненты — такой, при которой неравенства (1) и (2) для $e^{\alpha x} f(x)$ становятся эквивалентными), но оформлена она иначе и, по-моему, менее прозрачно.

Психологическая загадка: почему, имея теорему 2, Привалов не вывел из неё теорему 1? Мне кажется, дело в том, что в основной части [2] большую роль играет оператор, который как бы заменяет оператор Лапласа (для более общих функций) и который теперь называют «оператором Привалова». А в главе о выпуклых функциях Привалов в порядке подготовки делал упор на использование аналогичного (но более простого) разностного оператора и связанные с ним обстоятельства типа принципа максимума. По-видимому, при написании этой главы он менее заботился о различных взаимосвязях, не имеющих отношения к этому оператору (хотя всё-таки счёл обязательным привести теорему 1, доказываемую без его помощи).

Список литературы

[1] *Артин Э.* Введение в теорию гамма-функции. М.-Л.: ГТТИ, 1934.

[2] *Привалов И. И.* Субгармонические функции. М.-Л.: ОНТИ НКТП, 1937.

[3] *Бурбаки Н.* Функции действительного переменного. М.: Наука, 1965.

[4] *Bohr H., Mollerup J.* Laerebog i Matematisk Analyse, t. III. Kφbenhavn: Jul. Gjellerups Forlag, 1922.

[5] *Montel P.* Sur les fonctions convexes et les fonctions sousharmoniques. J. de math. pures et appl., 1928, t. 7, v. 1, 29-60.