МАТЕМАТИЧЕСКОЕ ПРОСВЕЩЕНИЕ

Третья серия

выпуск 10

Москва
Издательство МЦНМО
2006

УДК 51.009
ББК 22.1
М34

Издание осуществлено при поддержке РФФИ
(*издательский проект* № 05–01–14083).

Редакционная коллегия

Адрес редакции:
119002, Москва, Б. Власьевский пер., д. 11, к. 301
(с пометкой «Математическое просвещение»)
Email: matpros@mccme.ru   Web-page: www.mccme.ru/free-books

М34 **Математическое просвещение.** Третья серия, вып. 10. —
М.: МЦНМО, 2006. — 288 с.
ISBN 5-94057-227-8

В сборниках серии «Математическое просвещение» публикуются материалы о проблемах современной математики, изложенные на доступном для широкой аудитории уровне, заметки по истории математики, обсуждаются проблемы математического образования.

УДК 51.009
ББК 22.1

ISBN 5-94057-227-8

Содержание

Владимиру Михайловичу Тихомирову 70 лет

В 2005 году отмечался юбилей академика Российской академии естественных наук, профессора Московского государственного университета, заведующего кафедрой оптимального управления и главного редактора нашего «Математического просвещения», Владимира Михайловича Тихомирова. Трудно поверить, что этому красивому, моложавому и полному энергии человеку уже семьдесят. Владимир Михайлович продолжает активно и творчески работать как математик-исследователь, читать лекции не только в МГУ, но и в Независимом московском университете, на летней школе в Дубне и (в текущем семестре) в Бременском университете, вдохновенно вести многостороннюю общественную работу.

Редколлегия МП, преподаватели НМУ и Московского центра непрерывного математического образования желают Владимиру Михайловичу в первую очередь доброго здоровья. Нет нужды явно высказывать принятые в таких случаях прочие пожелания — зная Владимира Михайловича, мы уверены, что и «дальнейшие творческие успехи», и «долгие годы счастливой и плодотворной жизни» обязательно приложатся, было бы здоровье.

* * * * * *

Те, кто учился на мехмате МГУ в пятидесятые и шестидесятые годы прошлого столетия, прекрасно помнят студента, а затем аспиранта и ассистента Володю Тихомирова — одного из наиболее заметных людей на факультете. Необыкновенно красивый, стремительный и открытый человек, он считался одним из самых талантливых математиков своего поколения и был, одновременно, одним из наиболее активных общественных деятелей мехмата.

Его большая организационная работа по комсомольской линии не вызывала, однако, ни осуждения, ни раздражения студенческой элиты этого периода, в целом настроенной скептически или антагонистически к тем, кто сотрудничал с коммунистической властью. Дело в том, что в своей общественной деятельности Володя не стремился к выполнению указаний идущих сверху, а наоборот, пользовался существующими структурами для того, чтобы воплощать в жизнь полезные инициативы идущие от него и его друзей — тех, кому исполнилось двадцать к началу «хрущевской оттепели». Разумеется, Володя не занимался комсомольской работой с целью сделать карьеру — впрочем, ему бы это не удалось: партийной организации факультета очень не нравились его честность и независимость, «своим» они никак не могли его признать.

Но главным в этот период жизни Владимира Михайловича — как, впрочем, и во всей его дальнейшей жизни — была научная работа. Ему в этом отношении сразу повезло: уже на младших курсах на него обратил внимание Андрей Николаевич Колмогоров, и Володя стал его учеником и соратником. Первые работы юного Тихомирова, сразу принесшие ему известность, были выполнены под руководством А. Н. Колмогорова, да и в дальнейшем влияние его учителя было огромным. Так, три из пяти сюжетов математического творчества Тихомирова, описанных в статье Г. Г. Магарил-Ильяева (см. сс. 8–22 ниже), продолжают и развивают классические работы его учителя. Стоит отметить, что Колмогоров оказал огромное влияние на Владимира Михайловича не только в научном плане, но и в человеческом — и в смысле жизненной философии, и в отношении к культуре.

Вся дальнейшая жизнь Владимира Михайловича оказалась связанной с механико-математическим факультетом МГУ. Здесь он сразу после аспирантуры защитил кандидатскую диссертацию, в 36 лет защитил докторскую, стал доцентом и профессором, заведующим кафедрой. Кафедра оптимального управления была создана при его активном участии в 1966 году. Первым заведующим кафедрой стал академик В. А. Трапезников, однако с самого основания кафедрой реально руководил Владимир Михайлович. В 1989 он возглавил кафедру не только де-факто, но и де-юре.

Кафедра оптимального управления мехмата МГУ — любимое детище Владимира Михайловича. Он сумел привлечь к ней весьма квалифицированных специалистов, в том числе своих лучших учеников, создать доброжелательную творческую атмосферу, взяв на себя все трудности административной работы в непростые прошедшие и нынешние времена.

Все эти годы Владимир Михайлович продолжал заниматься общественной работой, как на факультете, так и вне него. Из разнообразных направлений этой деятельности — неизменно полезной и достойной уважения — мы отметим здесь лишь одно: работу по сохранению наследия московской математической школы (и конкретно научного наследия А. Н. Колмогорова) и по истории математики вообще. Многие помнят замечательные вечера памяти различных московских ученых, которое он организовывал и проводил в Центральном доме ученых, а читатели «Математического просвещения» знают эту сторону его деятельности по его историческим статьям, опубликованным в нашем сборнике:

об А. Н Колмогорове, Л. С. Понтрягине (№2), А. О. Гельфонде, Л. Г. Шнирельмане (№4), Л. А. Люстернике (№5), А. С. Кронроде, Л. В. Канторовиче (№6), И. М. Гельфанде (№8), И. Ф. Шарыгине (№9).

Владимир Михайлович прекрасный семьянин. Те, кому посчастливилось бывать в его гостеприимном доме, радуются царящей в нем теплой атмосфере и общением с женой Владимира Михайловича Наташей и с их двумя дочерьми.

Выступая на весьма нетрадиционном юбилейном вечере, посвященном его семидесятилетию, сценарий которого он сам и придумал, Владимир Михайлович отметил, что его жизнь была удивительно счастливой. А мы можем добавить — счастливой и яркой, и пожелать самим себе как можно чаще с ней соприкасаться.

Пять сюжетов о творчестве Владимира Михайловича Тихомирова

Г. Г. Магарил-Ильяев

Математическое творчество Владимира Михайловича Тихомирова очень разнообразно, но основные интересы его связаны с теорией приближений, теорией экстремальных задач и выпуклым анализом. В своей деятельности Владимир Михайлович всегда руководствуется идеей, что все значительное в математике основано на небольшом числе общих фундаментальных принципов, а конкретные результаты получаются на основе хорошо развитых исчислений. Здесь будет представлено несколько сюжетов, рассказывающих о проблемах, которыми занимался Владимир Михайлович в разные времена и где его влияние было существенным в становлении и/или развитии соответствующей тематики. Этих сюжетов пять, и это связано, в частности, с тем, что Владимир Михайлович любит, чтобы в любом списке (а он их составляет часто и по самым разным поводам) число пунктов было кратно пяти.

1. Поперечники множеств. В 1936 году А. Н. Колмогоров ввел понятие поперечника — величину, которая характеризует наилучшее приближение данного множества пространствами фиксированной размерности. Определение поперечника таково. Пусть X — нормированное пространство, C — подмножество X и n — неотрицательное целое число. Величина

$$d_n(C,X)=\inf_{L_n}\sup_{x\in C}\inf_{\xi\in L_n}\|x-\xi\|_X,$$

где первая нижняя грань берется по всем подпространствам L_n в X размерности не выше n, называется *n-поперечником по Колмогорову множества C в X*.

Для того, чтобы лучше почувствовать смысл этого понятия, рассмотрим следующий пример. Пусть X — плоскость с евклидовой нормой $|\cdot|$ и C — эллипс (см. рис. 1). Мы хотим наилучшим образом приблизить C одномерными подпространствами (т. е. прямыми L, проходящими через ноль). Это можно понимать так. Фиксируем какую-нибудь прямую L и пусть $x\in C$. Расстоянием от x до L в метрике X называется величина $d(x,C,X)=\inf_{\xi\in L}|x-\xi|$, а величина $d(C,L,X)=\sup_{x\in C}d(x,L,X)$

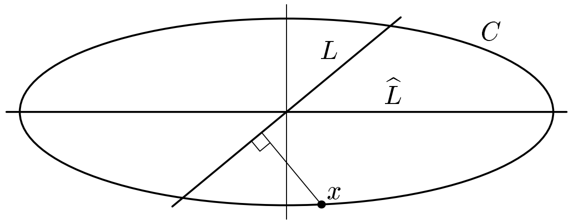


Рис. 1.

— уклонением C от L в метрике X. Нас интересует то подпространство $\widehat{L}$, на котором это уклонение минимально. Мы говорим тогда, что $\widehat{L}$ осуществляет наилучшее приближение эллипса C одномерными подпространствами. Очевидно, что $\widehat{L}$ совпадает с осью абсцисс и минимальное уклонение равно длине малой полуоси эллипса, что в соответствии с определением выше и есть 1-поперечник по Колмогорову эллипса C в евклидовой норме.

Появление работы Колмогорова ознаменовало начало нового этапа развития теории приближений, связанного с поисками наилучшего метода приближения данного функционального класса. С 1936 года и вплоть до шестидесятых годов практически не было публикаций на тему о поперечниках. Но затем, под влиянием исследований, связанных с проблемой Гильберта о суперпозициях функций и модной тогда теорией информации, эта тематика стала активно развиваться.

В этот период В. М. Тихомиров публикует ряд работ, которые стали определяющими для ее дальнейшего развития. Он ввел ряд величин, которые с разных точек зрения характеризуют аппроксимативные возможности данного множества (проекционный поперечник, линейный поперечник, поперечник по Гельфанду, по Бернштейну и др.) и получил ряд конкретных результатов для них, основываясь на разработанных им же методах, которые впоследствии многократно применялись и обобщались в десятках работ. Определенный итог этого этапа своей деятельности Владимир Михайлович подвел в монографии [1].

Здесь мы сформулируем один результат В. М. Тихомирова, который дает общий подход к оценкам поперечников снизу и покажем на примере, как он применяется.

Теорема (Тихомирова о поперечнике шара). *Пусть X — нормированное пространство, L_{n+1} — $(n+1)$-мерное подпространство X, $\gamma > 0$ и $\gamma BL_{n+1} = \{x \in L_{n+1} \mid \|x\|_X \leqslant \gamma\}$. Тогда*

$$d_n(\gamma BL_{n+1}, X) = \gamma.$$

Рассмотрим следующий пример. Пусть $X = C([0,1])$ — пространство непрерывных функций $x(\cdot)$ на отрезке $[0,1]$ с нормой $\|x(\cdot)\|_{C([0,1])} = \max_{t\in[0,1]}|x(t)|$, $C = W^1_\infty([0,1])$ — класс функций $x(\cdot)$ на $[0,1]$, удовлетворяющих условию Липшица (с константой единица), т. е. $|x(t) - x(t')| \leqslant |t - t'|$ для всех $t, t' \in [0,1]$. Покажем, что для любого $n \geqslant 1$

$$d_n(W^1_\infty([0,1]), C([0,1])) = \frac{1}{2n}\,.$$

Оценка снизу. Обозначим через L_{n+1} пространство ломаных на отрезке $[0,1]$ с изломами в точках $t_k = k/n$, $k = 1,\dots,n-1$. Легко проверить, что размерность L_{n+1} равна $n+1$ и что если $x(\cdot) \in L_{n+1}$ и максимум модулей угловых коэффициентов звеньев ломаной $x(\cdot)$ не превосходит единицы, то $x(\cdot) \in W^1_\infty([0,1])$. В частности, если $\|x(\cdot)\|_{C([0,1])} \leqslant 1/2n$, то нетрудно видеть, что этот максимум модулей не превосходит единицы и значит, $(1/2n)BL_{n+1} \subset W^1_\infty([0,1])$. Отсюда, учитывая очевидное свойство поперечника: если $C_1 \subset C_2$, то $d_n(C_1, X) \leqslant d_n(C_2, X)$, будем иметь по теореме Тихомирова о поперечнике шара:

$$d_n(W^1_\infty([0,1]), C([0,1])) \geqslant d_n\left(\frac{1}{2n}BL_{n+1}, C([0,1])\right) = \frac{1}{2n}\,.$$

Оценка сверху. Обозначим через L_n подпространство ломаных с изломами в точках $\tau_k = (2k-1)/2n$, $k = 1,\dots,n-1$, и постоянных на отрезках $[0,\tau_1]$ и $[\tau_{n-1},1]$. Легко видеть, что L_n — n-мерное пространство. Пусть $x(\cdot) \in W^1_\infty([0,1])$. Сопоставим этой функции ломаную $\xi(\cdot) \in L_n$ такую, что $\xi(\tau_k) = x(\tau_k)$, $k = 1,\dots,n-1$. Простой подсчет показывает, что $\|x(\cdot) - \xi(\cdot)\|_{C([0,1])} \leqslant 1/2n$. Тогда расстояние от $x(\cdot)$ до L_n в метрике $C([0,1])$ тем более не превосходит $1/2n$, т. е. $d(x(\cdot), L_n, C([0,1])) \leqslant 1/2n$. Переходя слева в этом неравенстве к верхней грани по всем $x(\cdot) \in W^1_\infty([0,1])$, а затем к нижней грани по всем подпространствам размерности не выше n, получаем, что

$$d_n(W^1_\infty([0,1]), C([0,1])) \leqslant \frac{1}{2n}\,.$$

Вместе с оценкой снизу это доказывает нужное утверждение.

2. Средняя размерность и наилучшие приближения на некомпактных многообразиях. В теории приближений, начиная с работ С. Н. Бернштейна, изучаются приближения отдельных функций и классов функций, заданных на всей прямой $\mathbb{R}$. Сам Бернштейн ввел для этого некий аналог пространства тригонометрических полиномов, а именно, пространство $B_\sigma(\mathbb{R})$, $\sigma > 0$, которое представляет собой сужения целых функций f на $\mathbb{R}$, удовлетворяющих условию: для любого $\varepsilon > 0$ найдется

такое число $C_\varepsilon = C_\varepsilon(f) \geqslant 0$, что для всех $z \in \mathbb{C}$ выполняется неравенство $|f(z)| \leqslant C_\varepsilon \exp(\sigma + \varepsilon)|z|$. Эквивалентное определение $B_\sigma(\mathbb{R})$ таково: это совокупность функций f на $\mathbb{R}$, у которых преобразование Фурье (как обобщенной функции) сосредоточено на отрезке $[-\sigma, \sigma]$. В последующие годы появились другие средства приближения классов функций на прямой, например, пространства сплайнов и вейвлетов. Кусочно-постоянные функции и ломаные — простейшие примеры пространств сплайнов. Эти пространства и пространства $B_\sigma(\mathbb{R})$ бесконечномерны, а стандартные классы функций (скажем, соболевские классы функций на прямой), для приближения которых они используются — некомпактны. Как в этой ситуации сравнивать различные средства приближения?

Один их возможных подходов основан на понятии средней размерности пространства, истоки которого восходят к работам К. Шеннона по теории информации, где он дал определение «энтропии на единицу времени» случайного сигнала на прямой с ограниченным спектром (т. е. реализации этого сигнала как раз являются элементами $B_\sigma(\mathbb{R})$). В 1956 году А. Н. Колмогоров модифицировал это определение для подпространств обычных (не случайных) функций.

Первый результат в этом направлении — энтропия на единицу времени ограниченных функций из $B_\sigma(\mathbb{R})$ — был получен В. М. Тихомировым [2]. Впоследствии Владимир Михайлович ввел характеристику подпространства, аналогичную колмогоровской, но отправляющуюся не от энтропии, а от поперечника по Колмогорову, которая и получили название средней размерности пространства. Суть дела такова. Начнем с простой ситуации. Пусть L^h — пространство ломаных на $\mathbb{R}$ с изломами в точках $\{kh\}$, $k \in \mathbb{Z}$, $h > 0$. Для любого $T > 0$ обозначим через L_T^h сужение пространства L^h на отрезок $[-T, T]$. Легко проверить, что размерность L_T^h равна $2T/h + 3$, если T/h — целое, и $2[T/h] + 3$ (где $[T/h]$ — целая часть T/h), если T/h — нецелое, и отсюда легко следует, что $\lim_{T\to\infty}(\dim L_T^h)/2T = h^{-1}$. Это число и называется средней размерностью (или размерностью на единицу времени) пространства L^h (заметим, что мы, очевидно, получим тот же результат, если будем рассматривать, например, только ограниченные ломаные). Эта процедура (с заменой в последней формуле предела на нижний предел) позволяет определить среднюю размерность любого пространства, сужение которого на $[-T, T]$ конечномерно для всех $T > 0$.

Ситуация несколько сложнее, если сужение подпространства на $[-T, T]$ не конечномерно. В этом случае поступаем следующим образом. Пусть L — подпространство, скажем, пространства $C^b(\mathbb{R})$ ограниченных непрерывных функций $x(\cdot)$ на прямой с нормой $\|x(\cdot)\|_{C^b(\mathbb{R})} = \sup_{t\in\mathbb{R}} |x(t)|$ и $BC^b(\mathbb{R})$ — единичный шар в $C^b(\mathbb{R})$. Обозначим через $(L \cap BC^b(\mathbb{R}))_T$ —

сужение множества $L \cap BC^b(\mathbb{R})$ на отрезок $[-T, T]$ и предположим, что это сужение компактно в $C([-T, T])$ (пространство непрерывных функций на $[-T, T]$ с обычной нормой) для каждого $T > 0$. Тогда его можно с любой точностью приблизить конечномерными подпространствами и значит, для любого $\varepsilon > 0$ конечна величина

$$K_\varepsilon(T, L, C^b(\mathbb{R})) = \min\{n \in \mathbb{Z}_+ \mid d_n((L \cap BC^b(\mathbb{R}))_T, C([-T, T])) < \varepsilon\},$$

где $d_n((L \cap BC^b(\mathbb{R}))_T, C([-T, T]))$ — n-поперечник по Колмогорову множества $(L \cap BC^b(\mathbb{R}))_T$ в метрике $C([-T, T])$. Ясно, что $K_\varepsilon(T, L, C^b(\mathbb{R}))$ — минимальная размерность подпространства, аппроксимирующего множество $(L \cap BC^b(\mathbb{R}))_T$ с точностью до ε.

Легко проверить, что функция $\varepsilon \to K_\varepsilon(T, L, C^b(\mathbb{R}))$ не возрастает при каждом $T > 0$.

Средней размерностью пространства L в $C^b(\mathbb{R})$ назовем величину

$$\overline{\dim}(L, C^b(\mathbb{R})) = \lim_{\varepsilon \to 0} \liminf_{T \to \infty} \frac{K_\varepsilon(T, L, C^b(\mathbb{R}))}{2T}.$$

Легко проверить, что если сужение L на $[-T, T]$ конечномерно, то данное определение равносильно определению, данному выше.

В терминах средней размерности результат В. М. Тихомирова, о котором говорилось выше, звучит так

$$\overline{\dim}(B_\sigma(\mathbb{R}) \cap C^b(\mathbb{R}), C^b(\mathbb{R})) = \frac{\sigma}{\pi}.$$

Пусть $\nu > 0$. Тогда отсюда и из предыдущего следует, что если h и σ таковы, что $1/h = \sigma/\pi = \nu$, то средние размерности пространства ограниченных ломаных и пространства ограниченных функций из $B_\sigma(\mathbb{R})$ равны ν. Рассматривая все пространства функций на прямой, у которых средняя размерность не превосходит ν, можно говорить о выборе наилучшего среди них для приближении того или иного класса функций на прямой. Точнее говоря, можно определить средний ν-поперечник по Колмогорову и другие средние поперечники и поставить те же вопросы, что и в компактной ситуации: точные и асимптотически точные значения этих поперечников, экстремальные пространства и т. п. Это определило новое направление в теории приближений — наилучшие приближения некомпактных классов функций. Здесь получено много интересных результатов. В настоящее время наиболее активно эта тематика развивается в Китае.

3. Принцип Лагранжа. В 1897 году Лагранж высказал следующий 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 функции, о которой говорилось, функции, задающие уравнения связи, умноженные на неопределенные множители, и искать затем максимум и минимум построенной суммы, как если бы переменные были независимы. Полученные уравнения, присоединенные к уравнениям связи, послужат для определения всех неизвестных».

Обдумывая необходимые условия экстремума в различных задачах, Владимир Михайлович пришел к выводу, что этот принцип (если придать ему чуть более расширенное толкование) имеет универсальный характер: практически все необходимые условия экстремума могут быть формально выведены, пользуясь рецептом Лагранжа.

Слова Лагранжа относились к задачам вида

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

где extr означает либо максимум, либо минимум. Согласно рекомендациям надо составить функцию (функцию Лагранжа) данной задачи: $\mathcal{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x)$ (разумно ставить множитель и у f_0; это некоторое продвижение со времен Лагранжа, но, правда, единственное), где $\lambda = (\lambda_0, \ldots, \lambda_m)$ — набор «неопределенных множителей» (множителей Лагранжа). Далее надо искать максимум и минимум функции $\mathcal{L}$ по x «как если бы переменные были независимы», для чего сначала надо выписать необходимые условия экстремума: равенство нулю частных производных. Если f_i — функции n переменных, то мы получим n уравнений. Относительно множителей λ_i эти уравнения однородны и поэтому один из них можно считать, скажем, равным единице. В результате имеем $n+m$ уравнений (m уравнений связи) для нахождения $n+m$ неизвестных. Точный результат (правило множителей Лагранжа) звучит так: *если функции f_i, $i = 0, 1, \ldots, m$, непрерывно дифференцируемы в некоторой окрестности точки $\widehat{x}$ и в этой точке достигается локальный экстремум, то найдется такой ненулевой набор множителей Лагранжа $\lambda = (\lambda_0, \ldots, \lambda_m)$, что $\widehat{x}$ удовлетворяет соотношению $\mathcal{L}_x(\widehat{x}, \lambda) = 0$, т. е. $\sum_{i=0}^{m} \lambda_i f_i'(\widehat{x}) = 0$, где $f_i'(\widehat{x})$ — производные (градиенты) функций f_i, $0 \leqslant i \leqslant m$, в точке $\widehat{x}$.*

Рассмотрим теперь так называемую задачу оптимального управления

$$\int_{t_0}^{t_1} f(t, x(t), u(t))\, dt \to \min, \quad \dot{x} = \varphi(t, x, u), \quad u(t) \in U,$$
$$x(t_0) = x_1, \quad x(t_1) = x_1, \quad (3.1)$$

и попробуем понять вид необходимых условий минимума в этой задаче, следуя формально рекомендациям Лагранжа.

В этой задаче f и φ — функции трех переменных, U — некоторое множество на прямой. Переменная u называются управлением, а x — фазовой переменной. Интерес к такого сорта задачам возник в пятидесятые годы прошлого века в ответ на запросы практики: требовалось оптимально (в том или ином смысле) управлять различными процессами, учитывая естественную ограниченность ресурсов (материальных, энергетических и т. п.). В нашем случае процесс описывается дифференциальным уравнением $\dot{x} = \varphi(t, x, u)$ (на «вход» подается управление $u(\cdot)$, на «выходе» получаем $x(\cdot)$). Мы хотим найти такое управление $\widehat{u}(\cdot)$, чтобы соответствующая фазовая переменная $\widehat{x}(\cdot)$ в начальный и конечный моменты времени t_0 и t_1 принимала значения x_0 и x_1, чтобы $\widehat{u}(\cdot)$ для каждого $t \in [t_0, t_1]$ не выходило за пределы множества U (отражающего ограниченность наших возможностей) и, наконец, чтобы это управление было оптимально в том смысле, что интеграл на паре $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ принимал минимальное значение.

Необходимые условия минимума в подобной задаче были получены в пятидесятые годы и получили название принципа максимума Понтрягина. Это одно из наиболее ярких достижений теории экстремальных задач.

Будем смотреть на задачу (3.1) как на задачу минимизации функции двух переменных $x(\cdot)$ и $u(\cdot)$, удовлетворяющих соответствующим ограничениям. Ограничение $\dot{x}(t) - \varphi(t, x(t), u(t)) = 0$, $t \in [t_0, t_1]$, можно воспринимать как континуум равенств, каждое из которых надо умножить на «неопределенный множитель», который обозначим (следуя традиции) $p(t)$ и «сложить», т. е. проинтегрировать. Таким образом, функция Лагранжа задачи (3.1) имеет вид

$$\mathcal{L}(x(\cdot), u(\cdot), \lambda) = \int_{t_0}^{t_1} (\lambda_0 f(t, x(t), u(t)) + p(t)(\dot{x}(t) - \varphi(t, x(t), u(t))))\,dt + \\ + \mu_0(x(t_0) - x_0) + \mu_1(x(t_1) - x_1),$$

где $\lambda = (\lambda_0, p(\cdot), \mu_0, \mu_1)$ — набор множителей Лагранжа. Пусть $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ — решение задачи. Выпишем необходимые условия минимума функции Лагранжа в этой точке отдельно по $x(\cdot)$ и по $u(\cdot)$. По $x(\cdot)$ — это так называемая задача Больца — стандартная задача классического вариационного исчисления. Если обозначить через L подынтегральную функцию, то необходимые условия в этой задаче имеют вид

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0 \quad \Leftrightarrow \quad \dot{p}(t) = -p(t)\widehat{\varphi}_x(t) - \lambda_0 \widehat{f}_x(t) \tag{3.2}$$

(где $\widehat{L}_{\dot{x}}(t) = L_{\dot{x}}(t, \widehat{x}(t), \dot{\widehat{x}}(t), \widehat{u}(t), \lambda_0, p(t))$ и аналогично для остальных функций с крышкой) и

$$\widehat{L}_{\dot{x}}(t_0) = \mu_0, \ \widehat{L}_{\dot{x}}(t_1) = -\mu_1 \quad \Leftrightarrow \quad p(t_0) = \mu_0, \ p(t_1) = -\mu_1. \tag{3.3}$$

Нетрудно проверить, что если $\widehat{u}(\cdot)$ доставляет минимум по $u(\cdot)$ функции Лагранжа, то необходимо (и достаточно), чтобы в каждой точке t, где функция $\widehat{u}(\cdot)$ непрерывна, функция L достигала минимума по $u \in U$ в точке $\widehat{u}(t)$, т. е.

$$\min_{u\in U}(\lambda_0 f(t,\widehat{x}(t),u) - p(t)\varphi(t,\widehat{x}(t),u)) = \\ = \lambda_0 f(t,\widehat{x}(t),\widehat{u}(t)) - p(t)\varphi(t,\widehat{x}(t),\widehat{u}(t)). \quad (3.4)$$

Можно записать это выражение в форме максимума, поменяв знаки перед $\lambda_0 f$ и $p\varphi$ на противоположные.

Соотношения (3.2), (3.3) и (3.4) (в форме максимума) и есть принцип максимума Понтрягина для данной задачи. Точную формулировку не будем приводить, но суть ее (как и в правиле множителей Лагранжа) в том, что существует такой ненулевой набор множителей Лагранжа λ, что выполняются условия (3.2)–(3.4).

Решение конкретных задач на основе принципа Лагранжа происходит, как правило, по следующей схеме: выписываются необходимые условия экстремума, потом они анализируются и в результате находят «подозреваемый» на экстремум объект (скажем, функцию). После этого проверяют, что найденный объект действительно есть решение данной задачи (в следующем сюжете будет рассмотрен пример). Важно при этом отметить, что зачастую нам нужна лишь структура необходимых условий, для чего вполне достаточно владения принципом Лагранжа, а не знания соответствующего точного результата. Более того, иногда такого результата просто нет, а принцип Лагранжа дает правильный ориентир для решения задачи.

В течении многих лет Владимир Михайлович пропагандирует принцип Лагранжа как эффективный прием для решения самых различных экстремальных задач. К сожалению, пока его идея: «Почти все экстремальные задачи устроены одинаково и решаются стандартно» еще не «овладела массами». До сих пор появляются работы, посвященные решению той или иной экстремальной задачи, которые суть упражнения на применение принципа Лагранжа. Особенно важно понимание принципа Лагранжа для «потребителей» теории экстремума (инженеров, экономистов, управленцев), которым эта теория экстремума преподносится как набор отдельных рецептов для решения задач того или иного типа.

Различные вопросы теории экстремума, в частности, принцип Лагранжа Владимир Михайлович продумывал со многими своими учениками и коллегами, и в первую очередь здесь следует назвать Александра Давидовича Иоффе и Владимира Михайловича Алексеева.

Систематическое изложение необходимых условий экстремума в различных экстремальных задачах с точки зрения принципа Лагранжа можно найти в монографии [3].

4. Неравенства для производных. Неравенства вида

$$\|x^{(k)}(\cdot)\|_{L_q(T)} \leqslant K\|x(\cdot)\|^{\alpha}_{L_p(T)}\|x^{(n)}(\cdot)\|^{\beta}_{L_r(T)}\,, \tag{4.1}$$

где $0 \leqslant k < n$, $1 \leqslant p,q,r \leqslant \infty$, $T = \mathbb{R}_+$ или R и $\alpha \geqslant 0$, $\beta \geqslant 0$ (при этом α и β однозначно определяются параметрами k,n,p,q,r), называют *неравенствами колмогоровского типа* или *неравенствами Ландау – Колмогорова*. Такое название связано с тем, что в 1913 году Э. Ландау доказал неравенство (4.1) для случая $k=1$, $n=2$, $q=p=r=\infty$ и $T=\mathbb{R}_+$ и нашел наилучшую (т. е. наименьшую из возможных) констант K в этом неравенстве, которая оказалась равной 2, а в 1938 году А. Н. Колмогоров нашел точную константу в неравенстве (4.1) для любых целых k и n, $1 \leqslant k < n$, $q=p=r=\infty$ и $T=\mathbb{R}$.

Неравенства вида (4.1) привлекали внимание многих известных математиков (Адамар, Надь, Харди, Литтльвуд, Полиа и др.). Всплеск интереса к этим неравенствам произошел в шестидесятые годы прошлого века в связи с задачей о наилучшем приближении неограниченного оператора (обычно оператора дифференцирования) ограниченными операторами, поставленной С. Б. Стечкиным. Был получен ряд новых точных результатов. Неравенства (4.1) играют важную роль в различных вопросах анализа и теории приближений.

Нахождение точной константы в таком неравенстве равносильно решению следующей экстремальной задачи

$$\|x^{(k)}(\cdot)\|_{L_q(T)} \to \max, \quad \|x(\cdot)\|_{L_p(T)} \leqslant \gamma_1,\ \|x^{(n)}(\cdot)\|_{L_r(T)} \leqslant \gamma_2 \tag{4.2}$$

для любых $\gamma_1 > 0$ и $\gamma_2 > 0$.

В начале семидесятых годов В. М. Тихомиров поставил общую задачу

$$\|D^{\alpha_0}x(\cdot)\|_{L_{p_0}(T)} \to \max, \quad \|D^{\alpha_i}x(\cdot)\|_{L_{p_i}(T)} \leqslant \gamma_i,\ i=1,\dots,m$$

(где D^{α} — оператор (вообще говоря) дробного дифференцирования, T — не только прямая или полупрямая, но, скажем, отрезок или все пространство $\mathbb{R}^n$) и предложил исследовать ее, опираясь на общие принципы теории экстремума. При этом основная цель, которую преследовал Владимир Михайлович, заключалась в том, чтобы испытать возможности этой теории, так как при различных p_i моделируется практически весь спектр задач, которые в ней изучаются (задачи вариационного исчисления, оптимального управления, выпуклые задачи, задачи с фазовыми ограничениями и т. д.). Не последнюю роль играла, конечно, и эстетическая сторона дела: точное решение красивой экстремальной задачи — всегда немалое удовольствие.

Через некоторое время выяснилось, что фактически все решенные к тому времени задачи вида (4.2) представляют собой, по сути дела, упражнения на применение принципа Лагранжа. Более того, такой естественный подход (решение экстремальной задачи методами теории экстремума) позволил сразу увидеть возможные обобщения полученных результатов, продвинуться в решении новых задач и увидеть связи их с другими задачами теории приближений.

В качестве примера найдем точную константу в одной достаточно простой ситуации, но при этом важно заметить, что схема рассуждений будет такой же и при доказательстве любого другого неравенства, где можно получить точный ответ аналитически.

Докажем точное неравенство

$$\|x(\cdot)\|_{C^b(\mathbb{R}_+)} \leqslant \sqrt{2}\,\|x(\cdot)\|^{1/2}_{L_2(\mathbb{R}_+)}\|\dot{x}(\cdot)\|^{1/2}_{L_2(\mathbb{R}_+)}\,, \tag{4.3}$$

справедливое для всех локально абсолютно непрерывных (т. е. абсолютно непрерывных на каждом отрезке) функций $x(\cdot) \in L_2(\mathbb{R}_+)$, у которых первая производная $\dot{x}(\cdot)$ также принадлежит $L_2(\mathbb{R}_+)$. Обозначим это пространство через $\mathcal{W}^1_2(\mathbb{R}_+)$.

Рассмотрим на $\mathcal{W}^1_2(\mathbb{R}_+)$ экстремальную задачу

$$x(0) \to \max, \quad \int_{\mathbb{R}_+} x^2(t)\,dt \leqslant 1,\ \int_{\mathbb{R}_+} \dot{x}^2(t)\,dt \leqslant 1. \tag{4.4}$$

Если мы найдем решение $\widehat{x}(\cdot)$ этой задачи, на котором ограничения обращаются в равенства, то докажем соответствующее точное неравенство. Действительно, пусть $x(\cdot) \in \mathcal{W}^1_2(\mathbb{R}_+)$ и $x(\cdot) \neq 0$. Тогда легко проверить, что функция $y(t) = ax(bt)$, где $a = (\|x(\cdot)\|_{L_2(\mathbb{R}_+)}\|\dot{x}(\cdot)\|_{L_2(\mathbb{R}_+)})^{-1/2}$ и $b = \|x(\cdot)\|_{L_2(\mathbb{R}_+)}/\|\dot{x}(\cdot)\|_{L_2(\mathbb{R}_+)}$ является допустимой в (4.4) и поэтому $y(0) = ax(0) \leqslant \widehat{x}(0)$, или

$$x(0) \leqslant \widehat{x}(0)\|x(\cdot)\|^{1/2}_{L_2(\mathbb{R}_+)}\|\dot{x}(\cdot)\|^{1/2}_{L_2(\mathbb{R}_+)}\,. \tag{4.5}$$

Отсюда следует, что справедливо такое неравенство

$$\|x(\cdot)\|_{C^b(\mathbb{R}_+)} \leqslant \widehat{x}(0)\|x(\cdot)\|^{1/2}_{L_2(\mathbb{R}_+)}\|\dot{x}(\cdot)\|^{1/2}_{L_2(\mathbb{R}_+)}\,. \tag{4.6}$$

В самом деле, если для некоторого $\overline{x}(\cdot) \in \mathcal{W}^1_2(\mathbb{R}_+)$ это неравенство не выполняется, то найдется такое $\tau \in \mathbb{R}_+$, что $\overline{x}(\tau)$ будет больше правой части в (4.6) при $x(\cdot) = \overline{x}(\cdot)$. Но на функции $x(t) = \overline{x}(t-\tau)$ значение правой части в (4.6), очевидно, не больше, чем на $\overline{x}(\cdot)$, а так как $x(0) = \overline{x}(\tau)$, то приходим к противоречию с неравенством (4.5).

Итак, для доказательства (4.3) осталось найти решение задачи (4.4) и убедиться, что его значение в нуле равно $\sqrt{2}$. Функция Лагранжа задачи (4.4) имеет вид

$$\mathcal{L}(x(\cdot),\lambda) = \lambda_0 x(0) + \lambda_1\Big(\int_{\mathbb{R}_+} x^2(t)\,dt - 1\Big) + \lambda_2\Big(\int_{\mathbb{R}_+} \dot{x}^2(t)\,dt - 1\Big),$$

где $\lambda = (\lambda_0, \lambda_1, \lambda_2)$. Согласно принципу Лагранжа, если $\widehat{x}(\cdot)$ — решение (4.4), то найдется такой ненулевой вектор λ, что функция Лагранжа удовлетворяет необходимым условиям минимума в точке $\widehat{x}(\cdot)$. Задача минимизации функции Лагранжа — это задача Больца (о которой говорилось уже в предыдущем сюжете), правда, на бесконечном промежутке. Не обращая на это внимания, выпишем формально необходимые условия минимума. Они заключаются в том, что должно выполняться уравнение Эйлера:

$$-2\lambda_2\ddot{\widehat{x}}(t) + 2\lambda_1\widehat{x}(t) = 0 \tag{4.7}$$

и условие трансверсальности:

$$2\lambda_2\dot{\widehat{x}}(0) = \lambda_0. \tag{4.8}$$

Проанализируем полученные соотношения. Во-первых, λ_1 и λ_2 не равны нулю, ибо, если $\lambda_1 = 0$, то из (4.7) следует, что либо $\lambda_2 = 0$, либо $\ddot{\widehat{x}}(\cdot) = 0$. В первом случае из (4.8) вытекает, что $\lambda_0 = 0$, т. е. все множители Лагранжа равны нулю, что невозможно. Во втором случае получаем, что $\widehat{x}(t) = at + b$. Но эта функция не принадлежит $L_2(\mathbb{R}_+)$, если $ab \neq 0$, а если $a = b = 0$, то $\widehat{x}(\cdot) = 0$, что, очевидно, тоже не так по смыслу задачи. Аналогичные рассуждения показывают, что и $\lambda_2 \neq 0$.

Далее, λ_1 и λ_2 должны быть одного знака, так как в противном случае никакое нетривиальное решение (4.7) не принадлежит $L_2(\mathbb{R}_+)$. Поскольку в соотношениях (4.7) и (4.8) множители Лагранжа определены с точностью до ненулевого сомножителя, то можно считать, что $\lambda_1 > 0$ и $\lambda_2 > 0$. Тогда из (4.7) следует, что

$$\widehat{x}(t) = C_1 e^{-\sqrt{\lambda_1/\lambda_2}\,t} + C_2 e^{\sqrt{\lambda_1/\lambda_2}\,t}.$$

Но $\widehat{x}(\cdot) \in L_2(\mathbb{R}_+)$, поэтому $C_2 = 0$ и тем самым $\widehat{x}(t) = C_1 e^{-\sqrt{\lambda_1/\lambda_2}\,t}$. Так как $\widehat{x}(\cdot) \neq 0$, то отсюда вытекает, что $\dot{\widehat{x}}(0) \neq 0$ и по смыслу задачи ясно, что $\dot{\widehat{x}}(0) < 0$. Тогда из (4.8) получаем, что $\lambda_0 < 0$ и можем считать, что $\lambda_0 = -1$.

Константы C_1, λ_1 и λ_2 найдем из (4.8) и соотношений

$$\int_{\mathbb{R}_+} \widehat{x}^2(t)\,dt = 1, \quad \int_{\mathbb{R}_+} \dot{\widehat{x}}^2(t)\,dt = 1.$$

Элементарный подсчет показывает, что

$$\widehat{x}(t) = \sqrt{2}e^{-t}, \quad \lambda_1 = \lambda_2 = \frac{1}{2\sqrt{2}}.$$

Ясно, что $\widehat{x}(\cdot) \in \mathcal{W}_2^1(\mathbb{R}_+)$.

Докажем теперь, что функция $\widehat{x}(\cdot)$ действительно является решением задачи (4.4). Она удовлетворяет уравнению (4.7). Умножим обе его части на $x(\cdot) \in \mathcal{W}_2^1(\mathbb{R}_+)$, проинтегрируем по $\mathbb{R}_+$, затем проинтегрируем в первом слагаемом по частям[1)] и подставим вместо $\widehat{x}(\cdot)$, λ_1, λ_2 их выражения. Тогда получим, что для всех $x(\cdot) \in \mathcal{W}_2^1(\mathbb{R}_+)$ имеет место тождество

$$x(0) = \int_{\mathbb{R}_+} e^{-t} x(t)\, dt - \int_{\mathbb{R}_+} e^{-t} \dot{x}(t)\, dt.$$

(справедливость которого легко проверить и непосредственно).

Пусть $x(\cdot)$ — допустимая функция в задаче (4.4). По неравенству Коши – Буняковского

$$x(0) \leqslant \left(\int_{\mathbb{R}_+} e^{-2t}\, dt \right)^{1/2} \|x(\cdot)\|_{L_2(\mathbb{R}_+)} + \\ + \left(\int_{\mathbb{R}_+} e^{-2t}\, dt \right)^{1/2} \|\dot{x}(\cdot)\|_{L_2(\mathbb{R}_+)} \leqslant \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} = \sqrt{2}.$$

Но $\widehat{x}(0) = \sqrt{2}$ и следовательно, $\widehat{x}(\cdot)$ — решение задачи (4.4). Неравенство (4.3) доказано.

Относительно подхода, основанного на принципе Лагранжа, к задачам о неравенствах для производных и их взаимосвязям с другими задачами теории приближений см. в [4].

5. Выпуклый анализ. Выпуклый анализ — раздел математики, где изучают выпуклые множества, выпуклые функции и выпуклые экстремальные задачи.

Работы Владимира Михайловича по выпуклому анализу очень разнообразны и связаны как собственно с развитием этой дисциплины, так и с приложениями ее к задачам теории приближений и теории экстремума. Но важно и то, что под влиянием этих работ сформировался определенный взгляд на предмет выпуклого анализа, который оказался весьма плодотворным, особенно для приложений. Суть этого взгляда состоит в том, что основное содержание выпуклого анализа — это соотношения двойственности для некоторого набора операторов и порожденное ими выпуклое исчисление. Поясним сказанное на примере оператора сопряжения для конусов и затем применим это к вопросу о критериях существования решений систем линейных неравенств.

[1)] Учитывая, что функции из $\mathcal{W}_2^1(\mathbb{R}_+)$ ограничены. Действительно, $\int_0^t x(\tau)\,\dot{x}(\tau)\,d\tau = x^2(\tau)|_0^t - \int_0^t x(\tau)\,\dot{x}(\tau)\,d\tau$. Следовательно, $2\int_0^t x(\tau)\,\dot{x}(\tau)\,d\tau = x^2(t) - x^2(0)$. Отсюда, применяя неравенство Коши – Буняковского, получим неравенство $x^2(t) \leqslant x^2(0) + 2\left(\int_{\mathbb{R}_+} x^2(\tau)\,d\tau\right)^{1/2} \left(\int_{\mathbb{R}_+} \dot{x}^2(\tau)\,d\tau\right)^{1/2}$ для всех $t \in \mathbb{R}_+$.

Обозначим через $\mathbb{R}^n$ обычное евклидово пространство всех вектор-столбцов $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, а через $(\mathbb{R}^n)'$ — евклидово пространство вектор-строк $y = (y_1, \dots, y_n)$. Если $y \in (\mathbb{R}^n)'$, то отображение $x \mapsto y{\cdot}x = \sum_{i=1}^n y_i x_i$ есть, очевидно, линейный функционал и любой линейный функционал на $\mathbb{R}^n$ может быть представлен в таком виде. Таким образом, мы отождествляем $(\mathbb{R}^n)'$ с пространством всех линейных функционалов на $\mathbb{R}^n$ и говорим, что $(\mathbb{R}^n)'$ — двойственное пространство к $\mathbb{R}^n$.

Сопоставим каждому выпуклому конусу $C \in \mathbb{R}^n$ конус $C^* \in (\mathbb{R}^n)'$ (который называется сопряженным конусом к C) по формуле: $C^* = \{\, y \in (\mathbb{R}^n)' \mid y \cdot x \geqslant 0,\ \forall\, x \in C \,\}$. Легко видеть, что это выпуклый конус (и даже замкнутый). Соответствующий оператор обозначим «$*$». Повторное применение этого оператора приводит к конусу $C^{**} = (C^*)^* = \{\, x \in \mathbb{R}^n \mid y \cdot x \geqslant 0,\ \forall\, y \in C^* \,\}$. Элементарно проверяется, что $C \subset C^{**}$.

Соотношение двойственности для данного оператора состоит в том, что *если C — выпуклый замкнутый конус, то*

$$C^{**} = C. \tag{5.1}$$

Доказательство есть простое следствие теоремы отделимости точки от замкнутого выпуклого множества.

Если C_1 и C_2 — выпуклые конусы, то, очевидно, $C_1 + C_2$ и $C_1 \cap C_2$ — также выпуклые конусы. Далее, если C — выпуклый конус в $\mathbb{R}^n$ и $A\colon \mathbb{R}^n \to \mathbb{R}^m$ — линейный оператор (который будем отождествлять с его матрицей размера $m \times n$ в стандартных базисах $\mathbb{R}^n$ и $\mathbb{R}^m$ и обозначать то же буквой), то образ (прообраз) C при действии A, который обозначим AC (CA), есть снова выпуклый конус в $\mathbb{R}^m$ $(\mathbb{R}^n)$.

Выпуклое исчисление для данного случая — это формулы для сопряженных конусов только что определенных операций. Для того, чтобы их выписать определим еще оператор $A^*\colon (\mathbb{R}^m)' \to (\mathbb{R}^n)'$ — сопряженный к A, действующий по правилу $A^*y = yA$. Итак, формулы выглядят так:

$$\begin{array}{llll} (a) & (C_1 \cap C_2)^* = C_1^* + C_2^* \qquad & (b) & (C_1 + C_2)^* = C_1^* \cap C_2^* \\ (c) & \qquad (AC)^* = C^* A^* & (d) & \qquad (CA)^* = A^* C^*. \end{array}$$

Формулы (b) и (c) справедливы без каких-либо предположений (и проверяются без труда). Для справедливости формул (a) и (d) требуются дополнительные предположения, но они заведомо выполняются, если считать, что конусы полиэдральны, т. е. являются пересечениями конечного числа замкнутых полупространств. Можно проверить, что определенные выше операции переводят полиэдральные конусы в полиэдральные.

Перейдем к приложениям. Если A — матрица размера $m \times n$ и $b \in \mathbb{R}^m$, то системы m линейных уравнений и неравенств с n неизвестными могут быть записаны так: $Ax = b$ и $Ax \leqslant b$, где неравенство понимается покоординатно.

Первое, что мы выясним — это условия существования неотрицательного решения у системы $Ax = b$ или, другими словами, когда $b \in A\mathbb{R}^n_+$, где $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \geqslant 0\}$. Ясно, что $\mathbb{R}^n_+$ — полиэдральный конус. В силу соотношения (5.1), формулы (c) и очевидного равенства $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$, будем иметь

$$A\mathbb{R}^n_+ = ((A\mathbb{R}^n_+)^*)^* = (\mathbb{R}^n_+ A^*)^*,$$

т. е. $b \in A\mathbb{R}^n_+$ в том и только в том случае, когда $y \cdot b \geqslant 0$ для тех $y \in (\mathbb{R}^m)'$, для которых $A^* y \geqslant 0$.

Это известная теорема Минковского – Фаркаша.

Следующий вопрос: какие условия существования решения у системы неравенств $Ax \leqslant b$? Другими словами, когда $b \in A\mathbb{R}^n + \mathbb{R}^m_+$? В силу (5.1), ($b$), ($c$) и того, что $(\mathbb{R}^n)^* = \{0\}$, будем иметь

$$A\mathbb{R}^n + \mathbb{R}^m_+ = ((A\mathbb{R}^n + \mathbb{R}^m_+)^*)^* = ((\mathbb{R}^n)^* A^* \cap (\mathbb{R}^m_+)^*)^* = (0A^* \cap \mathbb{R}^m_+)^*.$$

Но $0A^*$ — это ядро A^* и поэтому $b \in A\mathbb{R}^n + \mathbb{R}^m_+$ в том и только в том случае, когда $y \cdot b \geqslant 0$ для тех $y \in (\mathbb{R}^m)'$, для которых $y \geqslant 0$ и $A^* y = 0$.

Это также известная теорема Ки Фаня (Фань Цзи).

Наконец естественно поставить вопрос о существовании неотрицательного решения у системы неравенств $Ax \leqslant b$, т. е. когда $b \in A\mathbb{R}^n_+ + \mathbb{R}^m_+$? В силу (5.1), ($b$) и ($c$) получим

$$A\mathbb{R}^n_+ + \mathbb{R}^m_+ = ((A\mathbb{R}^n_+ + \mathbb{R}^m_+)^*)^* = ((\mathbb{R}^n_+)^* A^* \cap (\mathbb{R}^m_+)^*)^* = (\mathbb{R}^n_+ A^* \cap \mathbb{R}^m_+)^*$$

и таким образом, $b \in A\mathbb{R}^n_+ + \mathbb{R}^m_+$ в том и только в том случае, когда $y \cdot b \geqslant 0$ для тех $y \in (\mathbb{R}^m)'$, для которых $y \geqslant 0$ и $A^* y \geqslant 0$.

Это утверждение известно как теорема Гейла.

Если выпуклый конус — подпространство, то сопряженный конус — это аннулятор данного подпространства и соответствующее выпуклое исчисление позволяет получить критерии существования решений систем линейных уравнений.

Многие факты теории приближений и теории экстремума могут быть получены как следствия соотношений двойственности и формул выпуклого исчисления для тех или иных операторов, переводящих выпуклые объекты (выпуклые множества и выпуклые функции) в себя.

В монографиях [5]–[8] отражены основные воззрения Владимира Михайловича на предмет выпуклого анализа и его взаимосвязи с анализом, геометрией, теорией экстремума и теорией приближений.

Список литературы

[1] Тихомиров В. М. *Некоторые вопросы теории приближений.* Изд-во МГУ, 1976. 304 с.

[2] Тихомиров В. М. *Об ε-энтропии некоторых классов аналитических функций* // ДАН СССР, т. 117, №2, 1957, с. 191–194.

[3] Алексеев В. М., Тихомиров В. М., Фомин С. В. *Оптимальное управление.* Наука, Москва, 1979. 429 с.

[4] Магарил-Ильяев Г. Г., Тихомиров В. М. *О неравенствах для производных колмогоровского типа* // Мат. сборник, т. 188, №12, 1997, с. 73–106.

[5] Иоффе А. Д., Тихомиров В. М. *Теория экстремальных задач.* Наука, Москва, 1974. 479 с.

[6] Тихомиров В. М. *Выпуклый анализ.* Итоги науки и техники. Серия Современные проблемы математики. Фундаментальнык направления. ВИНИТИ, Москва, 1987, с. 5–101.

[7] Магарил-Ильяев Г. Г., Тихомиров В. М. *Выпуклый анализ и его приложения.* УРСС, Москва, 2003 (2-ое изд.), 176 с.

[8] Magaril-Il'yaev G. G., Tikhomirov V. M. *Convex Analysis: Theory and Applications.* AMS, Translations of Mathematical Monographs, vol. 222, 2003, 183 p.

Г. Г. Магарил-Ильяев, Московский государственный институт радиотехники, электроники и автоматики (технический университет)

Проблемы математического образования

Алгебра, анализ и дифференциальные уравнения (синтетический курс)

В. М. Тихомиров

Введение

Как надо учить математике и чему? Давайте поразмышляем.

Конечно, стоит воспользоваться богатейшим опытом преподавания математики на механико-математическом факультете МГУ. Начну с воспоминаний.

Я поступил на механико-математический факультет МГУ в 1952 году. Математическое отделение мехмата состояло в ту пору из восьми кафедр: анализа (заведующим кафедрой был тогда А. Я. Хинчин), высшей геометрии и топологии (П. С. Александров), алгебры (А. Г. Курош), дифференциальной геометрии (В. Ф. Каган), дифференциальных уравнений (И. Г. Петровский), теории функций и функционального анализа (Д. Е. Меньшов), теории вероятностей (А. Н. Колмогоров), теории чисел (А. О. Гельфонд).

Важную роль в преподавании играли обязательные курсы. Вот список обязательных курсов, из которых, в основном, складывалось мое математическое образование: математический анализ (лекторы Л. А. Тумаркин и А. И. Маркушевич), аналитическая геометрия (П. С. Александров), алгебра (А. Г. Курош), обыкновенные дифференциальные уравнения (В. В. Немыцкий), дифференциальная геометрия (П. К. Рашевский), анализ III (А. Н. Колмогоров), теория функций комплексного переменного (А. О. Гельфонд), дифференциальные уравнения с частными производными (О. А. Олейник), вариационное исчисление (И. М. Гельфанд), теория вероятностей (Ю. В. Прохоров).

Я недаром привел списки заведующих кафедрами и лекторов. По-моему, они свидетельствуют о том, что мехмат МГУ в ту пору был несравненным по концентрации выдающихся ученых среди всех остальных университетов мира. В подтверждение этого приведу слова В. И. Арнольда: «Плеяда великих математиков, собранных на одном факультете, представляет собой явление исключительное, и мне не приводилось встречать ничего подобного нигде.» (В. И. Арнольд, «Избранное–60». М.: Фазис, 1997, с. 714.) И образование, которое давалось в те годы, было очень высокого ранга. Это образование закладывалось не только на обязательных курсах. Знания, полученные на них, подкреплялись практическими семинарами, которые вели опытнейшие преподаватели-доценты, имевшие многолетний стаж работы. Назову некоторых: Н. Д. Айзенштат, В. Б. Демидович, С. А. Гальперн, З. М. Кишкина, А. С. Пархоменко, И. В. Проскуряков и другие. Но главной особенностью мехмата МГУ того времени было изобилие специальных курсов и специальных семинаров. Среди семинаров, сыгравших выдающуюся роль в истории отечественной математики — топологический кружок П. С. Александрова, семинар по теории вероятностей А. Н. Колмогорова и А. Я. Хинчина, семинар по теории функций Н. К. Бари и Д. Е. Меньшова, семинар по комплексному анализу М. А. Лаврентьева и А. И. Маркушевича, семинар по уравнениям с частными производными И. Г. Петровского, С. Л. Соболева и А. Н. Тихонова, семинар по всей математике И. М. Гельфанда и многие, многие другие.

Но вернемся к обязательным курсам. Они, как это можно усмотреть из сопоставлений их названий с названиями кафедр, распределялись, как правило, по кафедрам. Кафедры дифференциальных уравнений и теории функций и функционального анализа имели два курса (первая — по обыкновенным уравнениям и уравнениям с частными производными, вторая — по комплексному переменному и вариационному исчислению). Курс «Анализ III», читавшийся Колмогоровым, был особенным. Это был *синтетический курс*, соединявший в себе элементы теории функций, функционального анализа и интегральных уравнений (раньше эти дисциплины читались порознь).

* * * * * *

В течение последних двух лет мне довелось преподавать в двух необычных университетах — Московском независимом университете и Международном университете Бремена. В этой статье отражены некоторые идеи, которые возникли у меня в связи с этим преподаванием.

Сложившаяся у нас в России система университетского преподавания математики (да и многого другого) построена по так сказать «федеративному» принципу, когда математика рассматривается, как страна, разделенная на фактически независимые штаты: Анализа, Алгебры, Геометрии, Дифференциальных уравнений, Теории функций и функционального анализа, Теории вероятностей...

В университете Бремена введен особый курс, который я не встречал в других местах. Он называется “Perspectives of Mathematics”. Это курс без фиксированной программы, читаемый по очереди последовательно всеми профессорами университета для студентов, получивших уже предварительные знания по анализу и линейной алгебре. И мне пришла в голову идея прочитать *синтетический курс* по всей той математике, которой меня учили на первых двух курсах мои учителя пятьдесят с лишним лет тому назад. При этом я исходил из идеи, что математика — это единая страна, единое унитарное государство, и потому все можно воссоединить в едином цикле лекций. Здесь представлен конспект четырех лекций этого курса.

В первой лекции речь идет о решении линейных уравнений, во второй — о кониках и квадриках, затем о началах анализа и, наконец, о дифференциальных уравнениях. Каждая лекция подразделяется на четыре части. В «нулевой» части речь идет о предыстории, это «гуманитарный» фрагмент, не требующий каких-то математических знаний. Он еще адресован и дедушкам с бабушками, которые что-то хотят преподать своим внукам. В следующей части мы вкратце касаемся истоков теории, и там мы опираемся, в основном, на школьные знания. Затем речь идет о развитии темы; она частично бывает представлена в математическом вузовском образовании и на первых курсах университетов. В последней части доказываются некоторые фундаментальные результаты, читаемые, да и то далеко не всегда, на старших курсах университетов. В конце статьи вкратце обсуждаются приложения рассмотренных в ней вопросов к другим разделам математики и к естествознанию.

В начальных частях лекций речь заходит о знаменитых трудах, таких как папирус Райнда — древнейший задачник, известный в истории науки, «Коники» Аполлония, “Nova methodus” Лейбница (первая опубликованная статья по анализу) и “Method of Fluctions” Ньютона, далее упоминаются труды Эйлера, Лагранжа и Коши. В заключительных частях доказываются альтернатива Фредгольма, теорема Гильберта – Шмидта о симметрических операторах, обобщенная теорема Люстерника об обратном отображении, теоремы Коши существования, единственности, непрерывной и дифференцируемой зависимости решений дифференциальных уравнений.

Основной части предшествует раздел, содержащий предварительные сведения, в котором вводятся основные пространства, в которых предстоит действовать, а также формулируются три принципа существования и три принципа линейного анализа. Доказательства некоторых из этих принципов и некоторые прямые следствия из них приведены в приложении. В самом конце статьи мы возвращаемся к вопросу о том, чему и как надо учить математике.

Предварительные сведения

Здесь собраны основные понятия, а также некоторые факты общей топологии и линейного анализа, которые нам понадобятся далее в заключительных частях лекций. Я рассчитываю на читателей, имеющих какое-то отношение к математике, в частности, на школьников, ей интересующихся. Этой категории читателей исходные понятия в той или иной мере знакомы, так что к материалу этого раздела они могут обращаться по мере необходимости при чтении статьи (или ее отдельных частей, которые в значительной степени независимы друг от друга). И потому все предварительные сведения набраны петитом.[1)]

Стоит также предупредить читателя, что часть утверждений оставлена без доказательства, или дается лишь основная идея доказательства. Такие пропущенные детали рассуждений читателю рекомендуется восстановить самостоятельно.

Пространства и основные понятия

Основным, важнейшим пространством для нас явится совокупность вещественных чисел $\mathbb{R}$. Вещественные числа можно складывать, вычитать и делить (нельзя, впрочем, делить на нуль). Вещественные числа будут для нас примером векторного, топологического, метрического и банахова пространств, определения которых нас ждут впереди.

Кроме $\mathbb{R}$ мы постоянно будем сталкиваться с пространствами $\mathbb{R}^n$ и $(\mathbb{R}^n)'$.

Элементы $\mathbb{R}^n$ это столбцы $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ из n вещественных чисел; весь такой столбец мы будем обозначать одной буквой x и писать при этом $x \in \mathbb{R}^n$. Такие элементы будем называть *векторами*. Если $x \in \mathbb{R}^n$, то через x^T обозначается строка $(x_1, \dots, x_n)$. Столбцы можно покоординатно складывать и умножать на вещественные числа: если $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ и $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, то $x + y = \begin{pmatrix} x_1+y_1 \\ \vdots \\ x_n+y_n \end{pmatrix}$, а если $\alpha \in \mathbb{R}$, то $\alpha x = \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{pmatrix}$.

[1)] Сведения, содержащиеся в этом разделе, изложены в большинстве учебных книг по математике. Хотел бы порекомендовать читателю статьи «Банахово пространство», «Векторное пространство», «Гильбертово пространство», «Компактное пространство», «Компактный оператор», «Метрическое пространство», «Топологическое пространство» в кн. «Математическая Энциклопедия», 5тт., М.: Советская энциклопедия, 1977.

Кроме пространства $\mathbb{R}^n$ векторов-столбцов будем рассматривать пространство $(\mathbb{R}^n)'$ векторов-строк. $\mathbb{R}^n$ и $(\mathbb{R}^n)'$ доставляют примеры векторных пространств, т.е. множеств X с двумя операциями — сложением элементов и умножением элементов на вещественные числа с естественными аксиомами (коммутативности, ассоциативности, наличия нуля и обратного элемента для сложения и аксиом $(\alpha\beta)x = \alpha(\beta x)$, $(\alpha+\beta)x = \alpha x + \beta x$ и $\alpha(x+y) = \alpha x + \alpha y$, связывающих друг с другом операции сложения и умножения на число).

Помимо векторов нам встретятся таблицы размеров $n \times m$, называемые *матрицами*:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}.$$

Векторы из $\mathbb{R}^n$ можно воспринимать, как матрицы из одного столбца и n строк, векторы из $(\mathbb{R}^n)'$ — как матрицы из одной строки и n столбцов.

Определяется операция умножения матрицы A на вектор. Произведение матрицы A на вектор $x \in \mathbb{R}^n$ равно вектору $Ax = \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix}$; произведение вектора $y \in (\mathbb{R}^m)'$ на A — это вектор $yA = (\sum_{j=1}^m a_{j1}y_j, \dots, \sum_{j=1}^m a_{jn}y_j)$.

Если $x \in (\mathbb{R}^n)'$, а $y \in \mathbb{R}^n$, то через $x \cdot y$ обозначают сумму $x_1y_1 + \ldots + x_ny_n$ (это соответствует правилу умножения матриц). Величину $\sqrt{x^T \cdot x}$ обозначают $|x|$ и называют *модулем* вектора x.

Пространства $\mathbb{R}$, $\mathbb{R}^n$ и $(\mathbb{R}^n)'$ доставляют примеры топологических пространств. Вот что это такое. Пусть задано множество X и в нем система подмножеств τ, обладающая свойствами: a) пустое множество и всё X принадлежат τ, b) объединение любой совокупности множеств из τ и пересечение конечного числа множеств из τ принадлежат τ. Такая пара (X, τ) называется *топологическим пространством*.

Множества из τ называются *открытыми*, дополнения к открытым множествам называются *замкнутыми множествами*. Подмножество топологического пространства называется *компактом*, если из любого его открытого покрытия можно выделить конечное подпокрытие. Примером компакта на $\mathbb{R}$ может служить отрезок $[a, b]$, $-\infty < a < b < \infty$.

Нетрудно доказать, что пространства $\mathbb{R}^n$ и $(\mathbb{R}^n)'$, в которых множество называется открытым, если вместе с каждой своей точкой ξ оно содержит открытый шар $U(\xi, \varepsilon) = \{x \in \mathbb{R}^n \mid |x - \xi| < \varepsilon\}$ с центром в ξ радиуса ε при некотором $\varepsilon > 0$ (зависящем от ξ), становятся топологическими пространствами.

Важным примером топологических пространств являются метрические пространства.

Метрическое пространство — это множество X, на котором определена функция расстояния, сопоставляющая паре $(x_1, x_2) \in X \times X$ неотрицательное число $d(x_1, x_2)$ (расстояние между x_1 и x_2), равное нулю, если и лишь если $x_1 = x_2$, обладающее свойством симметрии $d(x_1, x_2) = d(x_2, x_1)$ (для любых x_1 и x_2) и удовлетворяющее неравенству треугольника: $d(x_1, x_3) \leqslant d(x_1, x_2) + d(x_2, x_3)$ для любых x_1, x_2, x_3.

Метрическое пространство становится топологическим, если открытые множества в нем определять, как множества, которые вместе с каждой своей точкой ξ содержат *открытый шар* $U(\xi, \varepsilon) = \{x \mid d(x, \xi) < \varepsilon\}$ (с центром в ξ радиуса ε при некотором $\varepsilon > 0$, зависящем от ξ).

Примерами метрических пространств могут служить прямая $\mathbb{R}$ (с $d(x, y) = |x - y|$) и n-мерные пространства $\mathbb{R}^n$ или $(\mathbb{R}^n)'$ с $d(x, y) = |x - y| = (\sum_{i=1}^n (x_i - y_i)^2)^{1/2}$.

Частным случаем метрических пространств являются *нормированные* пространства. Это векторные пространства $(X, \|\cdot\|_X)$, в которых определена норма $\|\cdot\|_X : X \to \mathbb{R}_+$, равная нулю лишь на нулевом элементе и удовлетворяющая аксиомам: $\|\alpha x\|_X = |\alpha|\,\|x\|_X$, для любых $\alpha \in \mathbb{R}$, $x \in X$ и $\|x + x'\|_X \leqslant \|x\|_X + \|x'\|_X$ для любых $x, x' \in X$. Метрика в нормированном пространстве определяется равенством $d(x, x') = \|x - x'\|_X$.

Геометрические рассмотрения естественно приводят к определению n-мерного *евклидова* пространства векторов-строк $x = (x_1, \ldots, x_n)$, оснащенного *скалярным произведением* $\langle x, y\rangle = \sum_{k=1}^n x_k y_k$ (это пространство будем обозначать $\mathbf{E}^n$). Оно становится банаховым, если норму определить как $\|x\|_{\mathbf{E}^n} = \sqrt{\langle x, x\rangle}$. Векторное пространство $(X, \langle\cdot,\cdot\rangle)$ со скалярным произведением $\langle\cdot,\cdot\rangle$, удовлетворяющим аксиомам неотрицательности ($\langle x, x\rangle \geqslant 0$ для любого $x \in X$, причем равенство достигается лишь на нулевом векторе), симметрии ($\langle x, x'\rangle = \langle x', x\rangle$ для любых $x, x' \in X$) и билинейности ($\langle \alpha x + \alpha' x', x''\rangle = \alpha\langle x, x''\rangle + \alpha'\langle x', x''\rangle$) называется *предгильбертовым*.

Объектом изучения будут далее и непрерывные функции на топологических или метрических пространствах.

Понятие функции или отображения (как и понятие множества) относится к числу изначальных понятий в математике. Если даны два множества X и Y, то отображением из X в Y называется соответствие f, при котором каждому элементу x множества X соотнесено либо пустое множество, либо один элемент $f(x)$ множества Y. При этом пишут: $f\colon X \to Y$. (Иначе можно сказать, что функция $f\colon X \to Y$ определяется подмножеством пар (x, y) (*декартова произведения* $X \times Y$ *всех пар* (x, y)), при котором каждому $x \in X$ соответствует либо пустое множество, либо одна пара (x, y); совокупность пар $(x, f(x))$ называется *графиком* f). Отображения из X в $\mathbb{R}$ называют *функционалами*, а если $X = \mathbb{R}^n$, то *функциями*. Подмножество $\operatorname{dom} f \subset X$ тех x, для которых $f(x) \neq \varnothing$, — *область определения* f; если $A \subset X$, то $\{y \in Y \mid \exists\, x \in A\colon y = f(x)\}$ — *образ* $A \subset X$ при отображении f (обозначается $f(A)$), а $\{x \in X \mid f(x) \in B\}$ — *прообраз* $B \subset Y$ (обозначается $f^{-1}(B)$). Если $f(A) = Y$, то отображение f называют *сюръективным*.

Пусть (X_1, τ_1) и (X_2, τ_2) — топологические пространства. Отображение $f\colon X_1 \to X_2$ называется *непрерывным*, если прообраз любого множества из τ_2 принадлежит τ_1 («прообраз открытого множества открыт»). Функция $f\colon X \to \mathbb{R}$ (где X — топологическое пространство) называется *полунепрерывной снизу* (*сверху*), если ее лебеговские множества $\mathcal{L}_\alpha(f) = \{x \mid f(x) \leqslant \alpha\}$ ($\mathcal{M}_\alpha(f) = \{x \mid f(x) \geqslant \alpha\}$) замкнуты при любом $\alpha \in \mathbb{R}$.

В случае метрического пространства определение непрерывной функции равносильно более привычному (с ε и δ).

Пусть (X, d) — метрическое пространство и f — функция на X. Она называется *непрерывной в точке* $\xi \in X$, если для всякого $\varepsilon > 0$ найдется $\delta > 0$ такое, что если $d(x, \xi) < \delta$, то $|f(x) - f(\xi)| < \varepsilon$. Если f непрерывна в каждой точке множества $A \subset X$, она называется *непрерывной на* A. Отображение $F\colon X \to X$ метрического пространства (X, d) в себя называется *сжимающим*, если существует такое число θ, $0 < \theta < 1$, что $d(F(x), F(x')) \leqslant \theta d(x, x')$ для любых x, $x' \in X$. Множество в метрическом пространстве (X, d) называется *всюду плотным* в X, если его замыкание совпадает с X; множество $A \subset X$ называется *нигде не плотным в* X, если каждое открытое подмножество в X содержит открытый шар, не пересекающийся с A.

Важнейшим метрическим бесконечномерным пространством является пространство $C([a, b])$ непрерывных вещественных функций на отрезке $[a, b]$, где расстояние определяется по формуле: $d(f, g) = \max_{x \in [a,b]} |f(x) - g(x)|$.

Пусть (X, d) — метрическое пространство. Последовательность его элементов $\{x_n\}_{n \in \mathbb{N}}$ называется *сходящейся*, если существует элемент x такой, что $\lim\limits_{n\to\infty} d(x_n, x) = 0$; она называется *фундаментальной, если для любого* $\varepsilon > 0$ *найдется номер* $N \in \mathbb{N}$

такой, что если $n, m > N$, то $d(x_m, x_n) < \varepsilon$. Очевидно, что любая сходящаяся последовательность фундаментальна.

Метрическое пространство (X, d) называется *полным*, если в нём любая фундаментальная последовательность является сходящейся. Примерами полных метрических пространств являются прямая $\mathbb{R}$, пространства $\mathbb{R}$ и $(\mathbb{R}^n)'$ и пространство $C([a, b])$.

Полные относительно метрики $d(x, x') = \|x - x'\|_X$ нормированные пространства называются *банаховыми*. Если X — нормированное пространство, то совокупность X^* непрерывных линейных функционалов на нем является банаховым пространством. Действие линейного функционала $x^* \in X^*$ на элемент $X \in X$ обозначим $\langle x^*, x\rangle$.

Обобщением евклидовых пространств $\mathbf{E}^n$ является гильбертово пространство — векторное пространство со скалярным произведением, полное относительно метрики $d(x, y) = \sqrt{\langle x - y, x - y\rangle}$. Гильбертово пространство является метрически изоморфным со своим сопряженным.

И в заключение дадим определение основного понятия дифференциального исчисления — понятия производной. Пусть X — одно из пространств, о которых речь шла выше: либо $\mathbb{R}$, либо $\mathbb{R}^n$, либо банахово пространство $(X, \|\cdot\|_X)$, U — открытое множество, содержащее точку $\widehat{x}$ и $f\colon U \to \mathbb{R}$ – вещественная функция. Говорят, что функция f дифференцируема в точке $\widehat{x}$, если найдется число a, соответственно — вектор $a \in (\mathbb{R}^n)'$ или непрерывный линейный функционал a на пространстве X такие, что $f(\widehat{x} + x) = f(\widehat{x}) + a[x] + r(x)$, где $a[x]$ в случае $\mathbb{R}$ это просто произведение чисел a и x, в случае $\mathbb{R}^n$ — произведение $a \cdot x$ вектора $a \in (\mathbb{R}^n)'$ на вектор $a \in \mathbb{R}^n$, в банаховом случае — это результат действия линейного функционала a на элемент x (т. е. $\langle a, x\rangle$), а $r(x) = o(x)$, а это означает, что $\lim\limits_{\|x\|\to 0} \dfrac{|r(x)|}{\|x\|} = 0$ (где $\|x\|$ в конечномерном случае означает $|x|$, а в бесконечномерном — $\|x\|_X$.

Принципы существования

I. Принцип компактности Вейерштрасса – Лебега – Бэра. *Полунепрерывная снизу на компакте функция достигает своей нижней грани.*

II. Принцип разреженности Бэра. *Полное метрическое пространство не представимо как объединение счетного числа нигде не плотных множеств.*

III. Принцип сжимающих отображений Пикара – Каччиополли – Банаха. *Сжимающее отображение полного метрического пространства в себя имеет неподвижную точку.* Отсюда легко выводится, что если некоторая степень отображения является сжимающей, то само отображение имеет единственную неподвижную точку.

Три принципа линейного анализа

Определение. Говорят, что два множества A и B в нормированном пространстве X *строго отделимы*, если существуют непрерывный линейный функционал $x^* \in X^*$ и число $\varepsilon > 0$ такие, что $\min_{x\in A}\langle x^*, x\rangle \geqslant \max_{x\in B}\langle x^*, x\rangle + \varepsilon$.

Принцип отделимости (следствие принципа Хана – Банаха). *Замкнутое выпуклое подмножество банахова пространства строго отделимо от точки, ему не принадлежащей.* Следствием этого результата является факт нетривиальности аннулятора $L^\perp = \{x^* \in X^* \mid \forall x \in L\ \langle x^*, x\rangle = 0\}$: *если L — замкнутое подпространство X, не совпадающее с X, то $L^\perp \neq \{0\}$.*

Теорема о правом обратном (следствие принципа открытости Банаха). *Пусть X и Y — банаховы пространства, $\Lambda\colon X \to Y$ — линейный, непрерывный, сюръективный оператор. Тогда существуют такие отображение $R\colon Y \to X$ и константа C, что $\Lambda R y = y$, $\|R(y)\|_X \leqslant C\|y\|_Y$ для всех $y \in Y$.*

Определение. Топология в пространстве X^*, сопряженном к нормированному пространству X, порожденная множествами $\Pi(x,\alpha) = \{x^* \mid \langle x^*, x\rangle < \alpha\}$, $x \in X$, $\alpha > 0$ и их конечными пересечениями, называется *слабой*.

Принцип компактности Банаха – Алаоглу. *Ограниченное выпуклое замкнутое подмножество пространства, сопряженного к банахову пространству, компактно в слабой топологии.*

1. Теория линейных уравнений

1.0. Задача из папируса Райнда и решение одного уравнения с одним неизвестным

Древнейшей рукописью, в которой обсуждаются математические вопросы, считается так называемый «Папирус Райнда» (названный по имени его владельца, египтолога Г. Райнда), написанный, как полагают, 4000 лет тому назад. В нем содержатся решения арифметических задач. Вот задача из этого папируса:

> *Число и седьмая часть этого числа в сумме равны девятнадцати. Чему равно число?*

(Если читателю захочется узнать, как египтянин должен был решать эту задачу, он может прочитать об этом в книге Г. Вилейтнера «Хрестоматия по истории науки», М.–Л.: ОНТИ, 1935, с. 14).

Для решения этой задачи надо составить одно уравнение с одним неизвестным:

$$Ax = b, \tag{1.1}$$

в котором $A = 1 + \frac{1}{7} = \frac{8}{7}$ и $b = 19$ известны, а x надо найти. О том, как решать уравнения с одним неизвестным, многие узнают еще до школы: для того, чтобы решить уравнение (1.1), надо поделить b на A (и тогда найдется $x = \frac{b}{A}$). В задаче из папируса Райнда $x = 19 : \frac{8}{7} = 16\frac{5}{8}$.

1.1. «Семь частей искусства математики» и решение системы двух уравнений с двумя неизвестными

В китайском учебнике «Семь частей искусства математики», написанном во время династии Хань свыше восемнадцати веков тому назад, есть задача: *В клетке фазаны и кролики. У них вместе* 35 *голов и* 94 *ноги. Сколько в клетке фазанов и сколько кроликов?*

Решать такие задачи нас учат в школе. В мое время детей учили решать их двумя способами: арифметически (в четвертом и пятом классе) и алгебраически (начиная с шестого). При арифметическом решении надо было рассуждать, скажем, так. Если бы вместо кроликов были бы куры, у них с фазанами вместе было бы семьдесят ног, на двадцать четыре ноги меньше. Эти двадцать четыре ноги получились за счет того, что кроликов было двенадцать, а фазанов, следовательно, двадцать три. Проверка: у двадцати трех фазанов сорок шесть ног, у двенадцати кроликов — сорок восемь, а всего девяносто четыре. Задача решена правильно.

Чтобы решить задачу «алгебраически», надо составить два уравнения с двумя неизвестными. Если число фазанов обозначить через x_1, а число кроликов через x_2, мы приходим к уравнениям: $x_1+x_2 = 35$, $2x_1+4x_2 = 94$.

В общем виде два уравнения с двумя неизвестными записываются так: $a_{11}x_1 + a_{12}x_2 = b_1$, $a_{21}x_1 + a_{22}x_2 = b_2$, где a_{ij}, $1 \leqslant i,j \leqslant 2$, и b_i, $i = 1,2$, известны, а x_1 и x_2 надо найти. Эту систему уравнений можно записать в виде (1.1), только A будет уже не числом, а таблицей из четырех чисел $A = \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right)$ (называемой *матрицей*). При этом x и b — это тоже не числа, а векторы — пары чисел, расположенные в столбец $x = \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$, $b = \left(\begin{smallmatrix} b_1 \\ b_2 \end{smallmatrix}\right)$. Для матрицы A и вектора x можно определить произведение Ax: надо первый элемент верхней строки умножить на x_1, а второй на x_2, получив первый элемент вектора произведения, а потом то же проделать со второй строкой. Таким образом, произведение матрицы A на вектор x приводит к вектору $Ax = \left(\begin{smallmatrix} a_{11}x_1+a_{12}x_2 \\ a_{21}x_1+a_{22}x_2 \end{smallmatrix}\right)$, и равенство (1.1) превращается в систему:

$$Ax = b \;\Leftrightarrow\; a_{11}x_1 + a_{12}x_2 = b_1,\; a_{21}x_1 + a_{22}x_2 = b_2. \tag{1.1$'$}$$

Школьников учат решать подобные системы методом исключения неизвестных. Например, для того, чтобы решить задачу про фазанов и кроликов, разумно из первого уравнения выразить x_2 через x_1 и, подставив во второе уравнение, найти x_1. Так или примерно так решают линейные уравнения и на практике (где иногда встречаются задачи с миллионами неизвестных). Но есть путь, имеющий теоретический интерес, который ведет к явному выражению решений не только двух уравнений с двумя неизвестными, но любого числа n уравнений с n неизвестными через «определители». Продемонстрируем его сначала на примере двух уравнений с двумя неизвестными.

Обозначив $a^1 = \left(\begin{smallmatrix} a_{11} \\ a_{21} \end{smallmatrix}\right)$, $a^2 = \left(\begin{smallmatrix} a_{12} \\ a_{22} \end{smallmatrix}\right)$ и $b = \left(\begin{smallmatrix} b_1 \\ b_2 \end{smallmatrix}\right)$, изобразим эти векторы на плоскости в декартовой системе координат (см. рис. 1). Пусть $V(a^1, a^2)$ означает (ориентированную) площадь параллелограмма, порожденного векторами a^1 и a^2. Функция $(a^1, a^2) \to V(a^1, a^2)$ обладает следующими очевидными свойствами:

a) если $e^1 = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, $e^2 = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$, то $V(e^1, e^2) = 1$;

b) $V(e^1, e^2) = -V(e^2, e^1)$;

c) $V(\alpha_1 a^1 + \bar{\alpha}^1 \bar{a}^1, a^2) = \alpha_1 V(a^1, a^2) + \bar{\alpha}^1 V(\bar{a}^1, a^2)$.

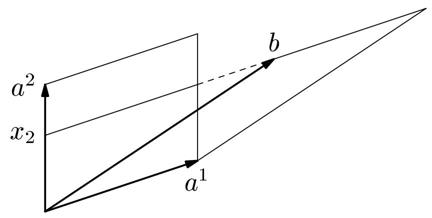


Рис. 1.

Эти свойства приводят к тому, что

$$V = V(a^1, a^2) = V(a_{11}e^1 + a_{21}e^2, a_{12}e^1 + a_{22}e^2) = \\ = (a_{11}a_{22} - a_{12}a_{21})V(e^1, e^2) = a_{11}a_{22} - a_{12}a_{21}.$$

Данное выражение называется *детерминантом* (*или определителем*) *матрицы* $A = \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right)$, обозначаемым $\det A$. Обозначим $V_1 = V(b, a^2)$ и $V_2 = V(a^1, b)$. Уравнение (1.1′) можно переписать в виде $a^1 x_1 + a^2 x_2 = b$. Из рис. 1 можно усмотреть, что $x_2 = \dfrac{V_2}{V}$ и аналогично, что $x_1 = \dfrac{V_1}{V}$, причем оба соотношения сразу выводятся из самих свойств a)–c).

В задаче про фазанов и кроликов $V = 2$, $V_1 = 46$, $V_2 = 24$, откуда получается, что в клетке 23 фазана и 12 кроликов. Это решение сложнее, чем решение исключением неизвестных, но оно, повторюсь, важно для многих теоретических рассмотрений.

Расскажем теперь о системах линейных уравнений со многими переменными. Теория таких уравнений начала складываться в 18 веке, а завершилась в девятнадцатом. С этой теорией знакомят студентов во многих вузах инженерного, экономического и многих других профилей, где изучают математику (и, разумеется, в университетах).

1.2. Системы n линейных уравнений с n переменными

Систему n уравнений с n неизвестными также можно записать в виде (1.1), только A будет уже матрицей размера $n \times n$:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix},$$

x и b — это n-мерные векторы-столбцы

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \qquad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix},$$

а произведение матрицы A на вектор x, определяемое аналогично двумерному случаю, приводит к вектору

$$Ax = \begin{pmatrix} a_{11}x_1 + \ldots + a_{1n}x_n \\ \vdots \\ a_{n1}x_1 + \ldots + a_{nn}x_n \end{pmatrix},$$

и тогда равенство (1.1) превращается в систему:

$$Ax = b \quad \Leftrightarrow \quad \sum_{j=1}^{n} a_{1j}x_j = b_1, \ldots, \sum_{j=1}^{n} a_{nj}x_j = b_n, \tag{1.2}$$

Попробуем решить систему (1.2) методом, обобщающим прием, только что продемонстрированный в предыдущем разделе, когда n равнялось двум.

Обозначим $a^1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix}, \ldots, a^n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{nn} \end{pmatrix}$ и $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$.

Введем функцию $V\colon (\mathbb{R}^n)^n \to \mathbb{R}$, $V = V(a^1, \ldots, a^n)$ «объема» системы векторов $\{a^j\}_{j=1}^n$. Двумерные и трехмерные примеры показывают, что разумно потребовать от этой функции выполнения таких условий:

a) условия нормировки, согласно которому объем единичного куба должен равняться единице: $V(e^1, \ldots, e^n) = 1$, где $e^1 = \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e^n = \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix}$ — стандартный базис в $\mathbb{R}^n$;

b) условия «антисимметричности»:

$$V(e^1, \ldots, e^{i-1}, e^i, e^{i+1}, e^{i+2}, \ldots, e^n) = -V(e^1, \ldots, e^{i-1}, e^{i+1}, e^i, e^{i+2}, \ldots, e^n);$$

c) условия линейности по каждому векторному аргументу, для этого достаточно потребовать, чтобы

$$V(\alpha a^1 + \bar{\alpha}\bar{a}^1, a^2, \ldots, a^n) = \alpha V(a^1, a^2, \ldots, a^n) + \bar{\alpha} V(\bar{a}^1, a^2, \ldots, a^n).$$

Также, как в двумерном случае, из приведенных аксиом однозначно выводится выражение для объема параллелепипеда, порожденного векторами a^1, a^2 и a^3 в трехмерном случае:

$$V(a^1, a^2, a^3) =$$
$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{32}a_{23}.$$

Продолжая по индукции, приходим к такому результату: *существует единственная функция $V = V(a^1, \ldots, a^n)$, которая определяется аксиомами* a)-c). Эта функция равна сумме $\sum_P (-1)^{\operatorname{sign} P} \prod_{i=1}^n a_{iP(i)}$, взятой по всем перестановкам P первых n чисел $\{1, 2, \ldots, n\}$ со знаком, равным $(-1)^{\operatorname{sign} P}$, где число $\operatorname{sign} P$ равно нулю, если переход от $\{1, 2, \ldots, n\}$

к $\{P(1), P(2), \ldots, P(n)\}$ требует четного числа транспозиций соседних элементов, и единице, если требуется нечетное число транспозиций (скажем, член $a_{21}a_{32}a_{13}$ имеет знак $+$, ибо переход от (312) к (123) требует двух транспозиций: (312) $\to$ (132) $\to$ (323)). Выписанная функция удовлетворяет аксиомам a)–c). Она и называется *детерминантом или определителем* матрицы A и обозначается $\det A$. Для тренировки читатель, который сталкивается с этим понятием впервые, может сосчитать какой-нибудь определитель четвертого порядка или выписать ответ в общем виде (придется просуммировать выражение из двадцати четырех слагаемых).

Основной теоретический результат, касающийся систем n линейных уравнений с n неизвестными — правило Крамера[2] — состоит в следующем:

Предложение 1.1. a) *Если определитель системы* $V = V(a^1, \ldots, a^n)$ *отличен от нуля, то уравнение* $Ax = b$ *разрешимо и решение системы имеет вид* $x_i = \dfrac{V_i}{V}$, *где* $V_i = V(a^1, \ldots, a^{i-1}, b, a^{i+1} \ldots, a^n)$.

b) *Имеет место альтернатива: либо уравнение* $Ax = b$ *однозначно разрешимо для любого* $b \in \mathbb{R}^n$, *либо однородная система* $Ax = 0$ *имеет ненулевое решение.*

Доказательство. Пусть задана матрица $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$, у которой $\det A = V(a^1, \ldots, a^n) \neq 0$. Из геометрического смысла определителя (и, разумеется, из его аксиоматических свойств) следует, что если вместо a^j подставить в $V(a^1, \ldots, a^n)$ вектор a^k, $k \neq j$, то получится нуль. Пусть V_{ij} — объем системы векторов $\{a^1, \ldots, a^{i-1}, e^j, a^{i+1}, \ldots, a^n\}$. Если помножить V_{ij} на a_{ij} и раскрыть выражение для определителя V_{ij}, то соберутся все члены определителя матрицы A, которые содержат в произведении элемент a_{ij}. Теперь, если просуммировать $\sum_{i=1}^n a_{ij}V_{ij}$, то соберутся все члены определителя матрицы A; если же взять сумму $\sum_{i=1}^n a_{ik}V_{ij}$ для $k \neq j$, то получится определитель матрицы с двумя одинаковыми столбцами a^k, т. е. нуль. Отсюда вытекает, что $e^j = \sum_{i=1}^n \dfrac{V_{ij}}{V} a^i$, где e^j — j-й базисный вектор, а V_{ij} — объем системы векторов $\{a^1, \ldots, a^{i-1}, e^j, a^{i+1}, \ldots, a^n\}$, а значит следует, что система уравнений (1.2) разрешима для любого $b = \sum_{j=1}^n b_j e^j$. При этом однородная система не может иметь ненулевого решения, ибо это несовместимо с неравенством нулю определителя. Заодно мы доказали правило Крамера, ибо $b = \sum_{j=1}^n b_j e^j = \sum_{i,j=1}^n b_j \dfrac{V_{ij}}{V} a^j = \sum_{j=1}^n \dfrac{V_j}{V} a^j$.

[2] Габриэль Крамер (1704–1752) — швейцарский математик; установил и опубликовал (в 1750 году) правило решения n линейных уравнений с n неизвестными, заложив основы теории определителей.

Предположим теперь, что система (1.2) разрешима для любого вектора b, а однородная система имеет ненулевое решение x^1. Обозначим через L_1 совокупность всех решений однородной системы. Решим уравнение $Ax^2 = x^1$. Тогда $x^2 \in L_2 \setminus L_1$, где L_2 — совокупность всех решений $A^2x = 0$. Решим уравнение $Ax_3 = x_2$ и далее будем поступать аналогично, и в результате (вроде бы) построим любое число линейно независимых векторов $\{x^k\}_{k\in\mathbb{N}}$. Но это невозможно, ибо $V(x^1, \dots, x^n) \neq 0$ (что вытекает из линейной независимости) и тогда по доказанному все остальные векторы должны выражаться через $\{x^m\}_{m=1}^n$. $\square$

Проведенное только что рассуждение является базисным в доказательстве бесконечномерного аналога доказанной теоремы, известного как альтернатива Фредгольма. С этой альтернативой (да и то не всегда) знакомят лишь студентов университетов (в то время, как правило Крамера проходят всюду — оно является существенной компонентой курса линейной алгебры[3])).

1.3. Альтернатива Фредгольма

Определение. Линейный оператор $\Lambda\colon X \to X$, где $(X, \|\cdot\|_X)$ — банахово пространство, называется *компактным*, если он отображает единичный шар X в компактное множество.

Теорема 1 (альтернатива Фредгольма). *Пусть X это банахово пространство, $\Lambda\colon X \to X$ компактный линейный оператор и $A = I - \Lambda$, где I — единичный оператор. Тогда имеет место альтернатива: либо уравнение $Ax = b$ однозначно разрешимо для любого $b \in X$, либо однородная система $Ax = 0$ имеет ненулевое решение.*

Доказательство. Если $\operatorname{Ker} A \neq \{0\}$, тогда неоднозначность следует уже для $b = 0$. Покажем, что если уравнение $Ax = b$ разрешимо для любого b, то $\operatorname{Ker} A = \{x \in X \mid Ax = 0\} = \{0\}$. Допустим, что это не так и пусть $x_1 \neq 0$ и $x_1 \in \operatorname{Ker} A$. Найдем $x_2 = Ax_1$. Тогда $x_2 \in \operatorname{Ker} A^2 \setminus \operatorname{Ker} A$. Продолжая этот процесс, найдем цепь строго вложенных друг в друга подпространств $\operatorname{Ker} A \subset \operatorname{Ker} A^2 \ldots \subset \operatorname{Ker} A^n \subset \ldots$ Легко доказывается, что в каждом банаховом пространстве для любого подпространства, не совпадающего со всем пространством, найдется вектор единичной длины, находящийся на расстоянии, сколь угодно близком к единице от этого подпространства. Выберем в каждом подпространстве $\operatorname{Ker} A^n$ элемент x_n единичной нормы, находящийся на расстоянии $\geqslant 1/2$ от подпространства

[3]) Следует сказать при этом, что правилом Крамера в практических расчетах никто не пользуется.

$\operatorname{Ker} A^{n-1}$. Если $m > n$, то нетрудно проверить, что $z = x_n - Ax_n - Ax_m \in \operatorname{Ker} A^{n-1}$ и значит, $\|\Lambda x_m - \Lambda x_n\|_X = \|z - x_m\|_X \geqslant 1/2$, что невозможно ввиду компактности Λ. $\square$

Доказанная теорема имеет многочисленные приложения в анализе. О них мы поговорим в конце. Альтернатива Фредгольма была доказана шведским математиком Иваром Фредгольмом (1866–1927) в 1900 году для специального класса линейных уравнений (так называемых интегральных уравнений Фредгольма второго рода вида $x(t) - \int_a^b K(t,\tau)x(\tau)d\tau = b(t)$ с непрерывными K и b, рассматриваемых в пространстве $C([a,b])$). В работе Фредгольма рассматривалось также и транспонированное уравнение $\xi(\tau) - \int_a^b K(\tau,t)\xi(t)dt = \beta(\tau)$ и доказывалась вторая теорема Фредгольма, о том, что если для основного уравнения имеет место первый случай альтернативы, то он имеет место и для транспонированного уравнения, число же линейно независимых решений основного и транспонированного решений конечно и одинаково (см. И. Г. Петровский, «Лекции по теории интегральных уравнений», М.: УРСС, 2003; в этой книге теорема доказывается методом вырожденных ядер). Сформулированную выше первую часть второй теоремы Фредгольма легко можно вывести из результата, доказанного нами.

Альтернатива Фредгольма явилась итогом больших усилий многих математиков второй половины 19 века, и этот результат воспринимался тогда как выдающееся событие, венчающее теорию линейных уравнений.

* * * * * *

Вот мы и совершили свое первое четырехступенное восхождение к одной из вершин математической науки. Повторюсь: первая ступень нашего «восхождения» преодолевается обычно еще до школы, вторая — на уроках арифметики и алгебры в школе, третья — в вузах, где преподают математику, четвертая — в университетах.

Нами была предпринята попытка преодолеть весь этот путь — от папируса Райнда к началу двадцатого века — в одной лекции. Удалась ли эта попытка, судить читателю.

2. Теория квадратичных функций. Коники и квадрики

Этот раздел мы начинаем не с одномерного, а с двумерного случая, с плоскости, с плоской геометрии, и это уведет нас сначала в мир античности. А далее нам останется сделать всего два шага — в n-мерный, а затем в бесконечномерный мир.

2.0. Несколько слов о «Кониках» Аполлония

Аполлоний (ок. 250 до н.э. – ок. 170 до н.э.), наряду с Евклидом и Архимедом, — один из трех величайших математиков древности. Его многотомный труд «Коники» (конические сечения) (увы, не полностью дошедший до нас), является безусловной вершиной античной математики. Там были исследованы плоские сечения прямого кругового конуса — эллипсы, гиперболы и параболы, — такие имена дал этим кривым Аполлоний. Об Аполлонии и его великом труде читатель может узнать из замечательной книги Б. А. Розенфельда «Аполлоний Пергский», М.: МЦНМО, 2004.

Многое в труде Аполлония поражает и нашего современника. Например, в пятой книге Аполлоний описывает кривую, разделяющую две плоские области, из точек одной из которых можно провести две нормали к эллипсу, из точек другой — четыре (а на самой кривой — три). Ныне эта кривая (огибающая семейства нормалей к эллипсу) известна, как астроида. Аполлонию удалось описать астроиду, уравнение которой $\left(\frac{x_1}{a_1}\right)^{2/3}+\left(\frac{x_2}{a_2}\right)^{2/3}=1$ было выписано в семнадцатом веке. Как всего этого можно добиться, не располагая средствами алгебры, чисто геометрически, — непостижимо уму нашего современника.

Но мы-то в школе проходим алгебру, и нам будет полегче изучать кривые Аполлония с ее помощью.

2.1. Декарт и теория плоских квадрик

В семнадцатом веке творили два великих французских ученых — Рене Декарт (1596–1650) и Пьер Ферма (1601–1665).

Им принадлежит фундаментальная для всего дальнейшего развития математики идея *арифметизации геометрии*. Они построили арифметические модели плоскости и пространства. Благодаря этой идее, в частности, геометрические задачи оказалось возможным решать алгебраическими средствами.

Напомним о том, что такое декартова модель плоскости. Нас в школе учат, как провести перпендикулярные прямые. Мысленно проведем их и выберем на каждой из них (одинаковый) масштаб. Тогда каждой точке плоскости соотнесется пара вещественных чисел (x_1, x_2) — ее координаты, а каждой геометрической фигуре — некоторое подмножество таких пар. Две основных фигуры геометрии древних — прямые и окружности — описываются так: прямые — линейными уравнениями $a_1x_1 + a_2x_2 = b$, а окружности — специальными квадратичными уравнениями: $(x_1-\alpha_1)^2+(x_2-\alpha_2)^2=r^2$.

Если зафиксировать декартову систему координат на плоскости, то точки плоскости можно назвать по-другому — векторами. Их можно складывать по правилу параллелограмма и можно умножать на вещественные числа. В арифметической модели Декарта сложение двух векторов $x = (x_1, x_2)$ и $x' = (x'_i, x'_2)$ приводит к вектору $x + x' = (x_1 + x'_1, x_2 + x'_2)$, а умножение вектора $x = (x_1, x_2)$ на число a дает вектор $ax = (ax_1, ax_2)$. Определим скалярное произведение $\langle x, x' \rangle$ векторов $x = (x_1, x_2)$ и $x' = (x'_1, x'_2)$ равенством $\langle x, x' \rangle = x_1x'_1 + x_2x'_2$. Пространство со скалярным произведением называют в конечномерном случае *евклидовым пространством*, а в бесконечномерном (при условии полноты) — *гильбертовым*. Мы построили таким образом двумерное евклидово пространство, которое обозначим $\mathbf{E}^2$.

Если $\langle x, x' \rangle = 0$ (т. е. скалярное произведение векторов равно нулю), векторы x и x' из $\mathbf{E}^2$ называют *ортогональными*, а вектор $x \in \mathbf{E}^2$ такой, что $\langle x, x \rangle = 1$, называют *единичным*. Эти термины находятся в соответствии с геометрическим взглядом на евклидову плоскость, в которой длина вектора $x = (x_1, x_2)$, согласно теореме Пифагора, равна $|x| = \sqrt{x_1^2 + x_2^2} = \sqrt{\langle x, x \rangle}$. При этом скалярное произведение векторов, как нетрудно подсчитать, равно произведению длин векторов на косинус угла между ними.

Посмотрим теперь как алгебраически описываются на евклидовой плоскости конические сечения Аполлония. В трехмерном пространстве троек (x_1, x_2, x_3) прямой круговой конус задается уравнением $x_1^2 + x_2^2 - x_3^2 = 0$. Рассмотрим его сечения плоскостями $x_3 = \alpha_1 x_1 + \alpha_2 x_2 + \beta$. Подставив в это равенство вместо x_3 выражение $\alpha_1 x_1 + \alpha_2 x_2 + \beta$, после возведения в квадрат и приведения подобных получим такое уравнение: $Q(x) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + 2b_1x_1 + 2b_2x_2 + c = 0$. Функции $x \mapsto Q(x)$ называют *квадратическими*, функции $x \mapsto q(x) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = \langle Ax, x \rangle$[4], где $A = \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right)$, $a_{21} = a_{12}$, называют *квадратичными формами*. Матрицу A со свойством $a_{21} = a_{12}$ называют *симметрической*. Каждая симметрическая матрица порождает с одной стороны объект алгебры и анализа — квадратичную форму (т. е. функцию двух переменных $x \mapsto \langle Ax, x \rangle$), а с другой — геометрический объект — преобразование плоскости: $x \mapsto Ax$, когда точка $x = (x_1, x_2)$ плоскости переходит в точку $Ax = (a_{11}x_1 + a_{12}x_2, a_{21}x_1 + a_{22}x_2)$. Это преобразование обладает следующим свойством: $\langle Ax, x' \rangle = \langle x, Ax' \rangle$. Тогда говорят, что A задает симмерический оператор, преобразующий плоскость в себя.

Коники в арифметической модели Декарта — это линии уровня квадратичных функций, иначе говоря, это множества точек, удовлетворяющих уравнениям $Q(x) = 0$, где $Q(x) = \langle Ax, x \rangle + 2\langle b, x \rangle + c$ — квадратичная

[4] Правильнее было бы писать (в соответствии с предыдущим разделом) $\langle (Ax^T)^T, x \rangle$, но мы здесь при рассмотрении конечномерного евклидова случая позволяем себе более простую запись.

функция, и возникает задача привести саму функцию и изображаемую ею кривую к более «удобному» виду. Сделаем это в важнейшем, невырожденном случае, когда матрица A обратима (т.е. $\det A \neq 0$).

Сначала простейшей заменой переменных $x = y + a \Leftrightarrow x_1 = y_1 + a_1,\ x_2 = y_2 + a_2$ попробуем избавиться от линейных членов. Имеем: $0 = \langle Ax, x\rangle + 2\langle b, x\rangle + c = \langle A(y+a), (y+a)\rangle + 2\langle b, (y+a)\rangle + c$. Раскрывая скобки и приводя подобные члены, получим равенство $\langle Ay, y\rangle + 2\langle b + Aa, y\rangle + c = 0$. Видно, что если положить $\widehat{a} = -A^{-1}b$ и $c' = c + 2\langle b, \widehat{a}\rangle$, получим, что в новых координатах уравнение кривой будет иметь вид: $\langle Ay, y\rangle + c' = 0$.

Продолжим наши алгебро-аналитические рассмотрения, а потом вскроем их геометрический смысл. Имеет место

Предложение 2.1. *Существует пара единичных ортогональных векторов e^1 и e^2, относительно которых невырожденная квадратичная форма $q(y) = \langle Ay, y\rangle$ приобретает диагональный вид*:

$$q(y) = \lambda_1 \langle y, e^1\rangle^2 + \lambda_2 \langle y, e^2\rangle^2,$$

где λ_1, $\lambda_2 \neq 0$.

Доказательство. Рассмотрим экстремальную задачу: *найти максимум функции q* при условии, что $y_1^2 + y_2^2 = 1$ (не ограничив себя в общности, можно считать, что среди значений q есть положительные). Из теоремы Вейерштрасса о достижении максимума непрерывной функции на компакте (см. с. 29) вытекает существование решения этой задачи (обозначим его e^1) (см. рис. 2). Теперь надо применить правило множителей Лагранжа. Если читатель не знает, что это такое, ему следует сначала прочитать следующий раздел, где это правило доказывается и в двумерном, и в многомерном, и в бесконечномерном случае. В применении к нашей задаче правило состоит в том, что найдется такой множитель Лагранжа λ_1, что производная по x функции Лагранжа

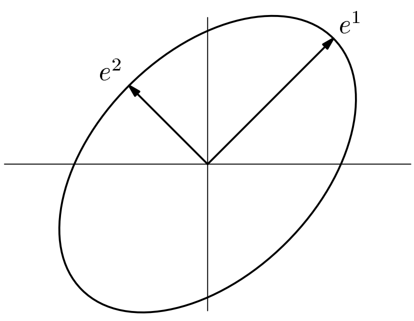


Рис. 2.

$\mathcal{L}(x, \lambda_1) = -\langle Ax, x\rangle + \lambda_1\langle x, x\rangle$ равна нулю (о производных функций двух и многих переменных см. с. 29; простые выкладки показывают, что производная функции $x \mapsto \langle Ax, x\rangle$ равна $2Ax$, функции $\langle b, x\rangle$ равна b, производная константы равна нулю). Отсюда вытекает, что $Ae^1 = \lambda_1 e^1$ и при этом λ_1 — это максимум в задаче (ибо $\langle Ae^1, e^1\rangle = \lambda_1 > 0$). Пусть теперь e^2 — единичный вектор, ортогональный e^1 (для нахождения такого вектора надо решить уравнения $\langle x, e^1\rangle = 0$, $\langle x, x\rangle = 1$). Тогда $\langle Ae^2, e^1\rangle = \langle e^2, Ae^1\rangle = \lambda_1\langle e^2, e^1\rangle = 0$, т. е. Ae^2 и e^2 пропорциональны. Обозначим коэффициент пропорциональности через λ_2. Значит, $y = \langle y, e^1\rangle e^1 + \langle y, e^2\rangle e^2$, откуда и следует, что $q(y) = \lambda_1\langle y, e^1\rangle^2 + \lambda_2\langle y, e^2\rangle^2$, λ_1, $\lambda_2 \neq 0$. □

Нетрудно понять, что вектор $\lambda_2 e^2$ является решением задачи о минимуме функции q при условии, что $y_1^2 + y_2^2 = 1$.

Векторы e^i, $i = 1, 2$, построенные нами (обладающие свойством $Ae^i = \lambda_i e^i$), называются *собственными векторами оператора* A. Доказанный нами результат можно переформулировать так: *каждый симметрический оператор на евклидовой плоскости обладает базисом из собственных векторов.*

Теперь переведем всё это на язык геометрии. Моделью куска плоскости может служить поверхность стола, за которым нам приходится работать. Мысленно нарисуем на этом столе декартову систему координат. Но нам потребуется еще один экземпляр евклидовой плоскости. Письменные столы иногда накрывают прозрачным стеклом. Вообразим себе такой прозрачный бесконечно тонкий и в то же время бесконечно жесткий через все проникающий кусок плоскости. Снимем его с нашего письменного стола и поднесем к конусу (расположенному также где-то недалеко от нас). Пересечем этим куском плоскости все образующие конуса, обведем цветным карандашом по контуру сечения и обозначим точку на нашем куске плоскости, где он пересекается с центральной осью конуса. На нашем прозрачном куске плоскости образовалась кривая, которую Аполлоний назвал эллипсом, и внутри этого эллипса расположился его центр. Положим теперь опять наш кусок плоскости на плоскость нашего письменного стола, прорисуем контур эллипса с прозрачного куска на плоскость стола и вернемся к аналитическим рассмотрениям. Уравнение нарисованного эллипса в системе координат на столе, имеет вид $\langle Ax, x\rangle + 2\langle b, x\rangle + c = 0$. Первое, что мы сделали, избавились от линейных членов. На языке геометрии это означает, что мы поместили новую декартову систему координат в центр эллипса (направив оси параллельно изначальным). А далее, взглянув на эллипс с новой системой координат (см. рис. 2), помещенной в его центр, мы заподозрили, что прямая, проходящая через максимально удаленные относительно центра точки и прямая, проходящая через минимально удаленные от центра точки, перпендикулярны друг другу.

Этот факт мы и доказали с помощью правила множителей Лагранжа. Повернув оси так, чтобы они пошли по прямым, соединяющим максимально и минимально удаленные точки, в новых осях получили уравнение $\lambda_1\langle y,e^1\rangle^2+\lambda_2\langle y,e^2\rangle^2=-c'$. В случае эллипса $-\lambda_i/c'>0,\ i=1,2$. А если бы мы сразу наш эллипс, нарисованный на прозрачной плоскости, «привезли» по столу и расположили в старой системе координат пустив его оси по осям координат, то получили бы уравнение $\left(\frac{x_1^2}{a_1^2}\right)+\left(\frac{x_2^2}{a_2^2}\right)=1$. Если же знаки $-\lambda_i/c'$ противоположны, то получили бы уравнение гиперболы: $\left(\frac{x_1^2}{a_1^2}\right)-\left(\frac{x_2^2}{a_2^2}\right)=1$.

Мы доказали, что невырожденная (и не являющаяся точкой $x_1^2+x_2^2=0$ или пустым множеством $x_1^2+x_2^2=-1$) кривая второго порядка — либо эллипс, либо гипербола.

2.2. Многомерный случай

Рассмотрим пространство векторов $x=(x_1,\dots,x_n)$ и определим скалярное произведение $\langle x,x'\rangle$ векторов $x=(x_1,\dots,x_n)$ и $x'=(x_1',\dots,x_n')$ равенством $\langle x,x'\rangle=\sum_{k=1}^n x_kx_k'$. Если $\langle x,x'\rangle=0$, векторы x и x' называют ортогональными, вектор x такой, что $\langle x,x\rangle=1$, называют единичным. Пусть $A=(a_{ij})_{1\leqslant i,j\leqslant n}$, $a_{ij}=a_{ji}$, — симметрическая матрица, а $q(x)=\sum_{i,j=1}^n a_{ij}x_ix_j$ — соответствующая ей квадратичная форма. Имеет место аналог предложения 2.1:

Предложение 2.2. *Существует базис из единичных ортогональных векторов* $\{e^k\}_{k=1}^n$, *относительно которых невырожденная квадратичная форма* $q(x)=\sum_{i,j=1}^n a_{ij}x_ix_j$ *приобретает вид*:

$$q(x)=\sum_{k=1}^n\lambda_k\langle x,e^k\rangle^2,\quad \lambda_k\in\mathbb{R}.$$

Векторы e^k, $1\leqslant k\leqslant n$, являются *собственными векторами* оператора A, так что можно сказать, что в $\mathbb{R}^n$ имеется ортонормированный базис из собственных векторов оператора A.

Этот результат принадлежит одному из выдающихся математиков 19 века — Карлу Якоби (1804–1851). Доказательство его совершенно аналогично двумерному, но мы его не приводим, ибо сам результат является частным случаем бесконечномерного обобщения, к рассмотрению которого мы переходим.

2.3. Бесконечномерный случай. Теорема Гильберта – Шмидта

Определение. Пусть $(X, \langle\cdot,\cdot\rangle)$ — гильбертово пространство. Линейный оператор $A\colon X \to X$ называется *симметричным* если $\langle Ax, x'\rangle = \langle x, Ax'\rangle$ для любых $x, x' \in X$.

Теорема 2. *Пусть $(X, \langle\cdot,\cdot\rangle)$ — сепарабельное гильбертово пространство и $A\colon X \to X$ — компактный симметричный линейный оператор. Тогда в X существует базис $\{e^i\}$ собственных векторов оператора A.* (Гильбертово пространство называется *сепарабельным*, если оно содержит счетное всюду плотное множество).

Этот результат был получен Гильбертом (1906) в пространстве l_2 бесконечномерных векторов $x = (x_1, \ldots, x_n, \ldots)$ для которых $\sum_{k\in\mathbb{N}} x_k^2 < \infty$. В этом пространстве вводится скалярное произведение по естественной формуле $\langle x, x'\rangle = \sum_{k\in\mathbb{N}} x_k x'_k$, а оператор A задается бесконечной матрицей $(a_{ij})_{i,j\in\mathbb{N}}$. Затем результат Гильберта был обобщен его учеником Шмидтом на общий случай гильбертова пространства. (Существует легенда, что когда Шмидт рассказывал об этой теореме на семинаре, Гильберт спросил: «Шмидт, о каком-таком гильбертовом пространстве Вы говорите? Я не понимаю»).

Доказательство. Рассмотрим экстремальную задачу о максимуме модуля квадратичной формы на единичном шаре:

$$|\langle Ax, x\rangle| \to \max, \quad \langle x, x\rangle \leqslant 1. \tag{1}$$

Из компактности оператора A следует непрерывность функции $x \mapsto \langle Ax, x\rangle$. Из принципа существования Банаха – Алаоглу (см. с. 30; надо учитывать при этом, что гильбертово пространство совпадает со своим сопряженным, а единичный шар — выпуклое замкнутое ограниченное множество) вытекает, что решение e^1 задачи существует. Из правила множителей Лагранжа следует, что найдется такое число λ_1, что производная по x функции Лагранжа $\mathcal{L}(x, \lambda_1) = -\langle Ax, x\rangle + \lambda_1\langle x, x\rangle$ равна нулю, что приводит к соотношению $Ae^1 = \lambda_1 e^1$. Рассмотрим новую экстремальную задачу

$$|\langle Ax, x\rangle| \to \max, \quad \langle x, x\rangle \leqslant 1,\ \langle x, e^1\rangle = 0. \tag{2}$$

Применение принципа существования и правила множителей Лагранжа приводит к равенству $Ae^2 = \lambda_2 e^2$. Продолжая процесс дальше, будем получать систему собственных чисел, стремящуюся к нулю (из-за компактности оператора). Затем надо взять пространство, ортогональное всем построенным e^k с не равными нулю собственными числами, и выбрать в нем ортонормированный базис (с нулевыми собственными значениями). Это приводит к искомому базису в X. $\square$

Читателю, знакомящемуся с этой теоремой впервые, рекомендуется провести подробное доказательство теоремы в трехмерном случае.

* * * * * *

Приложения доказанной теоремы совершенно неисчислимы — к решению смешанных задач для гиперболических уравнений, краевых задач для параболических уравнений, к гармоническому анализу, теории представлений, квантовой механике и т. д., и т. п. Кое-какие приложения мы обсудим далее.

* * * * * *

Так завершилась наша вторая экскурсия от Аполлония к Гильберту и Шмидту.

3. Начала математического анализа

3.0. Ньютон и Лейбниц — рождение анализа

Первой публикацией по анализу была коротенькая заметка Лейбница, появившаяся в 1684 году в журнале "Acta Eruditotum" — одном из первых научных журналов в истории науки. Статья была озаглавлена так: "Nova metodus pro maximis et minimis itemque tangentibus, quae nec fractas nec irrationales quantitates moratur, et singulare proillis calculus gemes" (Новый метод про максимумы и минимумы и определение касательных, не изменяющий в случае дробных или иррациональных величин и специальный способ исчисления к нему; выдержки из нее можно прочитать в упомянутой в первом разделе книге Вилейтнера на с. 272). В заглавии встречается слово "calculus" — *исчисление*. С чуть более поздней поры так и повелось: "differential and integral calculus" — дифференциальное и интегральное исчисление. Название заметки свидетельствует о стимулах, которые руководили ее автором при создании анализа: это — задачи на экстремум (т. е. на максимум и минимум) и проблемы геометрии.

Но необходимо сказать, что двумя десятилетиями ранее к основным концепциям дифференциального исчисления совсем с другой стороны пришел Ньютон. В 1671 году он представил рукопись работы "The Method of Fluctions and infinite Series with its Applications to the Geometry of Curves-Lines" (Метод флюксий и бесконечных рядов с приложением к геометрии кривых — см. И. Ньютон, «Математические работы», М.-Л., ОНТИ, 1937), сохранившей свое значение и в наше время.

Не указываю ни имен, ни лет жизни этих двух великих ученых — это должен знать каждый, не пишу об их национальной принадлежности, ибо прежде всего они принадлежат всему человечеству.

Ньютон с самого начала строил дифференциальное и интегральное исчисление, как аппарат естествознания. (О том, как устроен ньютонов мир, читатель может узнать из его “Principia” — «Математических началах натуральной философии» (1687) (см. Собрание сочинений трудов А. Н. Крылова, том VII, где помещен перевод «Математических начал»; Лагранж назвал это сочинение Ньютона «величайшим из произведений человеческого ума»).

«Для пояснения искусства анализа, — писал Ньютон, — остается привести некоторые примеры задач [...]. 1) Пусть длина пути известна. Нужно узнать скорость в данный момент времени. 2) Пусть известна скорость движения. Надо узнать длину пройденного пути.» (См. И. Ньютон, «Математические работы», М.-Л., ОНТИ, с. 45.)

И ведь действительно, если попытаться осмыслить понятие «скорости в данный момент» объекта, движущегося неравномерно, то неизбежно приходишь к идее *предела средней скорости за малый промежуток времени при стремлении этого промежутка к нулю*. Иначе говоря, если объект движется по прямолинейной дороге неравномерно, и $x(t)$ — расстояние его от какой-то начальной точки в момент времени t, то скорость $v(t)$ в момент t это предел отношения средних скоростей за малый промежуток времени при стремлении этого промежутка к нулю, т. е. $\lim \frac{x(t+\Delta t)-x(t)}{\Delta t}$ при Δt стремящемся к нулю. Эту величину стали обозначать $x'(t)$ (а Ньютон обозначал $\dot{x}(t)$).

(И само понятие предела Ньютон осознавал абсолютно отчетливо. Вот его слова: «Количества [...], которые в продолжение любого конечного времени постоянно стремятся к равенству и ранее конца этого времени приближаются друг к другу ближе, нежели на любую заданную величину, будут в пределе равны.» Это почти определение предела по Коши, которому нас учат на первом курсе.)

Большинство основных фактов дифференциального исчисления имеют физическую или геометрическую интерпретацию.

Скажем, вот вы идете со скоростью v по вагону поезда, движущегося со скоростью V. Тогда ваша скорость относительно Земли равна $V+v$. Так иллюстрируется формула дифференциального исчисления: *производная суммы равна сумме производных*.

«Когда величина достигает наибольшего или наименьшего значения, в этот момент она не течет ни вперед, ни назад» (см. «Математические работы», с. 73). Это высказывание на физическом языке выражает то, что в курсах анализа называют теоремой Ферма: *в точке экстремума производная равна нулю* (о том, что сказано по этому поводу самим Ферма см. в книге Вилейтнера на с. 256). Если вы куда-то ездили по прямолинейной дороге и вернулись обратно, то в какой-то момент вы оказались в наиболее удаленной точке от начала. Там вы «не текли ни назад, ни вперед», ваша скорость равнялась нулю. Это — физическая интерпретация теоремы Ролля: *если функция непрерывна на отрезке, дифференцируема внутри него и на концах отрезка принимает одинаковые значения, то в некоторой внутренней точке ее производная равна нулю.*

Но довольно о дифференциальном исчислении. Скажем несколько слов об обратной задаче — как найти путь по скорости. И снова, если задуматься об этом, то никакой другой возможности не придумать, как только разделить время на мелкие кусочки и считать скорость на этих маленьких участках постоянной, равной значению ее в какой-то промежуточной точке, а потом уменьшать разбиение и переходить к пределу, который обозначают: $\int_{t_0}^{t_1} v(t)dt$ (t_0 — начальная точка движения, t_1 — конечная, $v(t)$ — скорость в момент t) и называют *определенным интегралом* от скорости $v(\cdot)$ в пределах от t_0 до t_1.

Подведем итог: производная пути по времени $x'(t)$ — это скорость $v(t)$, а определенный интеграл, восстанавливающий пройденный путь $x(t)$ по скорости — это площадь под графиком скорости. И если теперь сопоставить все сказанное, то мы приходим к такому результату:

$$x(t_1) - x(t_0) = \int_{t_0}^{t_1} v(t)dt = \int_{t_0}^{t_1} x'(t)dt.$$

Это не что иное, как знаменитая формула Ньютона – Лейбница, которую иногда называют *основной формулой интегрального исчисления.* Она связывает два исчисления друг с другом.

А теперь перейдем к обсуждению одной из важнейших теорем анализа — теоремы об обратном отображении. Здесь мы собираемся повторить (и даже с некоторым усилением) то, что уже дважды делали: сначала мы сформулируем теорему в одномерном случае, затем в многомерном, и наконец, в бесконечномерном случае, а доказывать все эти результаты будем одновременно, и читатель убедится в том, что в бесконечномерном случае доказательство не отличается от одномерного, надо только знать о «бесконечномерности» нечто просто формулируемое, но фундаментальное.

Пусть $X = Y = \mathbb{R}$, или $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$, $m \leqslant n$, или $(X, \|\cdot\|_X)$ и $(Y, \|\cdot\|_Y)$ — банаховы пространства, U — окрестность нуля в X, $F\colon U \to Y$, $F(0_X) = 0_Y$.

Определение. Говорят, что F *строго дифференцируема в начале координат пространства* X (и пишут $F \in SD^1(0)$) если существует число $F'(0)$ (соответственно — матрица $F'(0)$ или в самом общем случае — линейный непрерывный оператор $F'(0)\colon X \to Y$) такие, что для любого $\varepsilon > 0$ существует такое $\delta > 0$, что $\|F(x') - F(x) - F'(0)(x' - x)\|_Y < \varepsilon\|x' - x\|_X$, если только $\|x\|_X < \delta$, $\|x'\|_X < \delta$, где в конечномерном случае $\|x\|_X = |x|$ и $\|y\|_Y = |y|$.

Отметим, что строго дифференцируемая в $\widehat{x}$ функция дифференцируема в точке $\widehat{x}$ и непрерывна в некоторой окрестности $\widehat{x}$. Не всякая дифференцируемая функция строго дифференцируема, это показывает пример функции $F(x) = x^2D(x)$, где $D(x)$ — функция Дирихле, равная нулю в иррациональных точках и единице в рациональных. (Функция F непрерывна и дифференцируема в нуле и разрывна всюду, кроме нуля.)

3.1. Теорема об обратной функции в одномерном случае

Теорема 3 а) об обратной функции. *Пусть U — окрестность нуля в $\mathbb{R}$, $F\colon U \to \mathbb{R}$ — строго дифференцируемая в нуле функция, причем $F'(0) \neq 0$. Тогда существуют такие числа $\varepsilon > 0$, $\delta > 0$ и $K > 0$ что для любого числа $|y| < \delta$ найдется (единственное) число $x = x(y)$, $|x| < \varepsilon$ такое, что $F(x) = y$ и $|x| \leqslant K|y|$.*

Теорема 3 а) восходит к Ньютону. Он изложил (на примере решения уравнения $y^3 - 2y - 5 = 0$) метод вычислений корня этого уравнения, ныне всем известный как *метод Ньютона*, в своей работе «Анализ с помощью уравнений...» («Математические работы», с. 9).

Ньютон был замечательным вычислителем. В середине семидесятых годов семнадцатого века начал тлеть приоритетный спор между Ньютоном и Лейбницем об открытии анализа. С запросом по поводу того, что известно Ньютону о математическом анализе, обратился к нему ученый секретарь Королевского общества Ольденбург. Ньютон несколько раз отвечал своему респонденту. Во втором письме Ольденбургу, отправленном “24 октября 1676 г. от Рождества Христова” (и подлежащем, как указывает Ньютон, быть сообщенным Лейбницу, о нем мы еще будем иметь повод вспомнить), Ньютон пишет, что «мне прямо стыдно признаться, до какого числа знаков я довел на досуге эти вычисления» (речь шла о вычислении логарифмов — см. «Математические работы», с. 237).

Ниже мы излагаем менее быстрый, чем собственно метод Ньютона, но более удобный метод решения нелинейных уравнений.

Нам надлежит решить уравнение $F(x) = y$. Будем искать его итеративно $x_n = x_{n-1} + \dfrac{1}{F'(0)}(y - F(x_{n-1}))$, $n \in \mathbb{N}$, $x_0 = 0$. Эта процедура (называемая *модифицированным методом Ньютона*), изображена на рис. 3. (В методе Ньютона $x_n = x_{n-1} + \dfrac{1}{F'(x_{n-1})}(y - F(x_{n-1}))$, $n \in \mathbb{N}$, $x_0 = 0$). Оставим обоснование сходимости модифицированного метода Ньютона к $x(y)$ до последнего раздела, где будем доказывать теорему одновременно в одномерном, многомерном и бесконечномерном случаях.

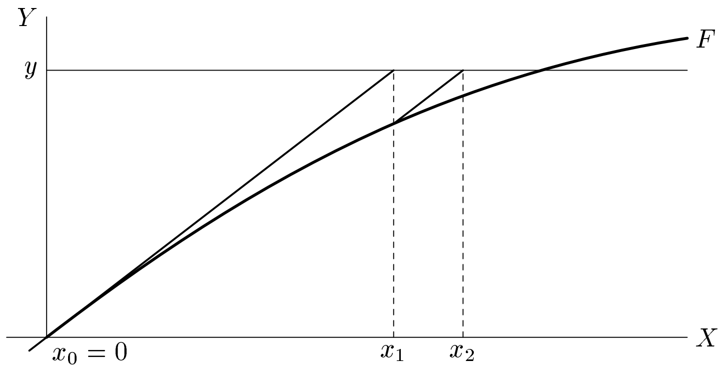


Рис. 3.

3.2. Многомерный случай

Рассмотрим частный вариант многомерного случая, когда $n = m$. Пусть требуется решить систему из n нелинейных уравнений

$$f_1(x_1, \dots, x_n) = y_1, \dots, f_n(x_1, \dots, x_n) = y_n.$$

Введем, как мы не раз уже делали, сокращенные обозначения:

$$F(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Тогда наша задача примет такой вид: решить уравнение $F(x) = y$. Решение будем искать точно также — модифицированным методом Ньютона: $x_n = x_{n-1} + (F'(0))^{-1}(y - F(x_{n-1}))$, $n \in \mathbb{N}$, $x_0 = 0$, где $(F'(0))^{-1}$ означает обратную матрицу. Доказательство проведем в следующем разделе.

3.3. Общий случай (одномерный, конечномерный и бесконечномерный; правило множителей Лагранжа)

Пусть $X = \mathbb{R}^n$, а $Y = \mathbb{R}^m$ и $A\colon X \to Y$ — линейный сюръективный оператор. Тогда существует правый обратный оператор $R\colon Y \to X$ такой, что $ARy = y$ и $|R(y)| \leqslant C|y|$ (где C — некоторая константа). Действительно, надо взять базис $\{e^i\}_{i=1}^m$ в Y, элементы f^i, $1 \leqslant i \leqslant m$, из прообраза $A^{-1}e^i$ и для $y = \sum_{i=1}^m y_i e^i$ положить $R(y) = \sum_{i=1}^m y_i f^i$. (Тогда $|R(y)| \leqslant \sum_{i=1}^m |y_i| \max_{1\leqslant i\leqslant m} |f^i| \leqslant \sqrt{m} \max_{1\leqslant i\leqslant m} |f^i||y|$).

Если же $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ — банаховы пространства и $A\colon X \to Y$ — линейный, непрерывный сюръективный оператор, то существование правого обратного оператора с приведенными выше свойствами есть факт, равносильный принципу открытости Банаха (см. с. 29 и приложение).

Теорема 3 в) (об обратном отображении). *Пусть $X = Y = \mathbb{R}$, соответственно $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$, $m \leqslant n$, или в самом общем случае $(X, \|\cdot\|_X)$ и $(Y, \|\cdot\|_Y)$ — банаховы пространства, U — окрестность нуля в X, $F\colon U \to Y$, $F \in SD^1(0)$, $F(0_X) = 0_Y$ и более того $F'(0)X = Y$ (в случае $X = Y = \mathbb{R}$ это означает, что $F'(0) \neq 0$, в случае $X = Y = \mathbb{R}^n$, что* $\det F'(0) \neq 0$ *(см. 1.2)). Тогда существуют такие $\varepsilon > 0$, $\delta > 0$ и $K > 0$, что для любого числа $|y| < \delta$ (соответственно, вектора $|y| < \delta$ в конечномерном случае или элемента $\|y\|_Y < \delta$ в банаховом случае) существует число $x = x(y)$, $|x| < \varepsilon$ (соответственно — вектор $x = x(y)$, $|x| < \varepsilon$ или элемент $x = x(y) \in X$, $\|x\|_X < \varepsilon$) такие, что $F(x) = y$ и $|x| \leqslant K|y|$ в конечномерном и $\|x\|_X \leqslant K\|y\|_Y$ в банаховом случае.*

Доказательство. Пусть в одномерном случае R — это число $\dfrac{1}{F'(0)}$, в многомерном случае, если $m = n$, — матрица $(F'(0))^{-1}$, а в остальных случаях — правый обратный для $F'(0)$ оператор. Определение строгой дифференцируемости влечет за собой существование такого $\delta > 0$, что если $\|x\|_X < \delta$, $\|x'\|_X < \delta$, то выполняется неравенство

$$\|F(x') - F(x) - F'(0)(x' - x)\|_Y < \frac{1}{2C}\|x' - x\|_X, \qquad (i)$$

где C — константа из теоремы о правом обратном (см. с. 29). Пусть при этом $U_X(0,\delta) = \{x \in X \mid \|x\|_X < \delta\} \subset U$ и F непрерывна в $U_X(0,\delta)$. Выбрав $\varepsilon < \dfrac{\delta}{8C}$, рассмотрим итеративную процедуру модифицированного метода Ньютона:

$$x_{n+1} = x_n + R(y - F(x_n)), \quad n \geqslant 0,\ x_0 = 0, \qquad (ii)$$

где в конечномерном случае $\|\cdot\| = |\cdot|$. Покажем, что все x_j, $j \in \mathbb{N}$, лежат в $U_X(0,\delta)$. Применим метод математической индукции. Имеем: $\|x_1\|_X \leqslant$ $\leqslant C\|y\|_Y < \delta$, значит, $x_1 \in U_X(0,\delta)$. Пусть $x_k \in U_X(0,\delta)$, $1 \leqslant k \leqslant n$. Тогда получим

$$-y + F(x_{k-1}) - F'(0)(x_k - x_{k-1}) \overset{(ii)}{=} 0, \quad 1 \leqslant k \leqslant n \qquad (iii)$$

и следовательно,

$$\begin{aligned} \|x_{n+1} - x_n\|_X &\overset{\operatorname{def} R}{\leqslant} C\|y - F(x_n)\|_Y \overset{(iii)}{=} \\ = C\|y - F(x_n) - y + F(x_{n-1}) &+ F'(0)(x_n - x_{n-1})\|_Y \overset{(i)}{\leqslant} 1/2\|x_n - x_{n-1}\|_X \leqslant \\ &\leqslant 1/4\|x_{n-1} - x_{n-2}\|_X \leqslant \ldots \leqslant 1/2^{n-1}\|x_1\|_X\,. \end{aligned} \qquad (iv)$$

Из (iv) и неравенства треугольника следует, что

$$\|x_{n+1}\|_X < 2\|x_1| < \delta, \tag{v}$$

т. е. элементы x_n определены для всех n, а $\|x_n - x_{n+m}\|_X \leqslant 1/2^{n-1}\|x_1\|_X$, откуда вытекает, что $\{x_n\}_{n\in\mathbb{N}}$ — фундаментальная последовательность. Переход к пределу в (ii) (существующий из-за непрерывности F в $U(0_X,\delta)$) приводит к равенству $F(x(y)) = y$, а переход к пределу в (v), учитывая, что $\|x_1\|_X \leqslant C\|y\|_Y$, обеспечивает неравенство $\|x(y)\|_X \leqslant K\|y\|_Y$ с $K = 2C$. □

Приведем одно из важнейших следствий этой теоремы: правило множителей Лагранжа.

Пусть X и Y это $\mathbb{R}^2$ и $\mathbb{R}$ или $\mathbb{R}^n$ и $\mathbb{R}^m$, $m < n$, или банаховы пространства $(X, \|\cdot\|_X)$ и $(Y, \|\cdot\|_Y)$, U — окрестность точки $\widehat{x}$ в X, $f_0\colon U \to \mathbb{R}$, $F\colon U \to Y$. Рассмотрим задачу о нахождении точек экстремума (т. е. максимума или минимума функции f_0 при ограничении типа равенства $F = 0$):

$$f_0(x) \to \text{extr}, \quad F(x) = 0. \tag{P}$$

Говорят, что элемент $\widehat{x}$ доставляет локальный минимум (максимум) задаче (P) (и пишут $\widehat{x} \in \text{locmin (locmax)}$), если существует такая окрестность V точки $\widehat{x}$, что для точки $x \in V$, для которой $F(x) = 0$, выполнено неравенство $f_0(x) \geqslant f_0(\widehat{x})$, $(f_0(x) \leqslant f_0(\widehat{x}))$.

Если $X = \mathbb{R}^2$, $Y = \mathbb{R}$, положим $F(x) = f_1(x)$ и $\mathcal{L}(x,\lambda) = f_0(x) + \lambda f_1(x)$ (здесь λ —число); если $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$ положим $F(x) = (f_1(x), \ldots, f_m(x))^T$ и $\mathcal{L}(x,\lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)$ (здесь $\lambda = (\lambda_1, \ldots, \lambda_m)$ — вектор); если $(X, \|\cdot\|_X)$ и $(Y, \|\cdot\|_Y)$ — банаховы пространства, положим $\mathcal{L}(x,\lambda) = f_0(x) + \langle\lambda, F(x)\rangle$, где λ — элемент сопряженного пространства Y^*, а $\langle\cdot,\cdot\rangle$ — стандартная билинейная форма на $Y^* \times Y$ ($\langle\lambda, x\rangle$ — это значение на элементе x линейного функционала λ). Написанные выражения называются *функцией Лагранжа* задачи (P).

Следствие (правило множителей Лагранжа). *Пусть в* (P) f_0, $F \in SD^1(0)$, $F'(0)X = Y$. *Если* $\widehat{x}$ *локальный экстремум* (P), *тогда найдется множитель Лагранжа* λ (*число, вектор или элемент пространства* Y^*, *сопряженного с* Y), *такой, что*

$$\mathcal{L}_x(\widehat{x},\lambda) = 0. \tag{3.1}$$

Расшифруем сказанное в конечномерных случаях. Если $X = \mathbb{R}^2$, $Y = \mathbb{R}$, то соотношение (3.1) означает, что $f_0'(\widehat{x}) + \lambda f_1(\widehat{x})$, т. е., что градиенты $f_0'(\widehat{x})$ и $f_1'(\widehat{x})$ пропорциональны (и если понимать геометрический смысл градиента функции на плоскости, то это утверждение становится почти очевидным), а если $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$, то (3.1) означает, что

$f_0'(\widehat{x})+\sum_{i=1}^m f_i'(\widehat{x})=0$, т.е., что градиент $f_0'(\widehat{x})$ является линейной комбинацией градиентов ограничений.

Доказательство. Доказательства следствия во всех случаях сходны и просты. Если бы в случае $X=\mathbb{R}^2$, $Y=\mathbb{R}$ градиенты не были бы пропорциональны, то к отображению $\Phi(x)=(f_0(x),f_1(x))$ можно было бы применить теорему об обратном отображении и найти элемент x_α такой, что $\Phi(x)=(\alpha,0)$, для которого $f_0(x_\alpha)=\alpha$, $f_1(x_\alpha)=0$ и $|x_\alpha|\leqslant K|\alpha|$, а это противоречит тому, что $\widehat{x}\in \operatorname{locextr}$. Совершенно аналогично приходим к противоречию во втором случае, если градиент $f_0'(\widehat{x})$ не является линейной комбинацией градиентов ограничений (и градиенты ограничений линейно независимы). А в бесконечномерном случае либо $\Phi'(\widehat{x})X=Y\times\mathbb{R}$, либо нет. В первом случае применяем теорему об обратном отображении и приходим к противоречию, а во втором надо воспользоваться нетривиальностью аннулятора подпространства $\Phi'(\widehat{x})X\subset Y$. □

В бесконечномерном случае правило множителей Лагранжа было доказано Л. А. Люстерником.

* * * * * *

Наше третье путешествие длилось более, чем два с половиной столетия с 1671 года, когда Ньютон представил «Метод флюксий» по 1934 г., когда Л. А. Люстерник опубликовал теорему о правиле множителей Лагранжа в банаховом случае. Но наш следующий путь будет короче, хотя он будет менее элементарен.

4. Дифференциальные уравнения

4.0. О пользе решения дифференциальных уравнений

В 17 веке было принято зашифровывать свои основные мысли в анаграммах. В том самом втором письме Ольденбургу, о котором мы упоминали в предыдущем разделе, Ньютон пишет:

> «Сущность этих действий [...] я лучше передам в следующем скрытом виде: 6accdae13eff7i3l9n4o4qrr4s9t12vx».

Текст анаграммы, раскрытый им впоследствии, таков: Data aequatione quotcunque fluentes quantitates involvente fluxiones invenire et vice versa. Вот перевод этой фразы: *по данному уравнению, содержащему флюенты, найти флюксии и обратно.* Термины «флюента» и «флюксия» были введены самим Ньютоном. Флюента по Ньютону — изменяющаяся величина, сейчас наиболее точно переводимая словом «функция». Флюксия — это скорость изменения флюенты, так что по сути дела в своей анаграмме Ньютон повторяет то, о чем мы уже говорили: суть анализа в нахождении скорости по пути и наоборот.

Один из крупнейших математиков современности — В. И. Арнольд — предложил такой вольный перевод анаграммы Ньютона: «*Полезно решать дифференциальные уравнения*» (В. И. Арнольд, «Избранное–60». М.: Фазис, 1997, с. 319.). Воспользуемся этим «полезным советом» Ньютона – Арнольда, тем более, что главную концепцию «Principia» можно выразить словами: «Мир управляется дифференциальными уравнениями».

Пусть $f = f(t,x)\colon \mathbb{R}^{n+1} \to \mathbb{R}^n$ $(t \in \mathbb{R},\, x \in \mathbb{R}^n)$. Далее рассматриваются дифференциальные уравнения вида $\dot x = f(t,x)$. Задача: *решить дифференциальное уравнение* $\dot x = f(t,x)$ *с начальным условием* $x(t_0) = x_0$ называется *задачей Коши*.

4.1. Линейные дифференциальные уравнения с постоянными коэффициентами

Начнем с простейших уравнений. Простейшее неоднородное дифференциальное уравнение $\dot x = v$, $x(t_0) = x_0$ обсуждалось в предыдущем разделе. Его решение дается интегралом $x(t) = x_0 + \int_{t_0}^t v(s)ds$. Здесь мы начнем с простейшего *однородного* уравнения: $\dot x = x$. Для решения задачи Коши $\dot x = x$, $x(0) = 1$ этого уравнения, рассмотрим отображение $F\colon C([-a,a]) \to C([-a,a])$, $F(x(\cdot))(t) = 1 + \int_0^t x(s)ds$. Нетрудно показать, что некоторая степень F^n этого отображения (зависящая от a) является сжимающей, и значит, в силу следствия из принципа сжимающих отображений и полноты пространства $C([-a,a])$, существует единственный элемент $\widehat{x}(\cdot)$ такой, что $F(\widehat{x}(\cdot))(t) = \widehat{x}(t)$, получаемый итерациями $x_{n+1}(t) = F(x_n(\cdot))(t)$, $n \geqslant 0$, $x_0(\cdot)$ — произвольный элемент.

Если начать итеративную процедуру с функции $x_0(t) = 1$, мы получим $x_1(t) = 1+t$, $x_2(t) = 1+t+\dfrac{t^2}{2}, \ldots, x_n(t) = \sum_{k=0}^n \dfrac{t^k}{k!}$. Эти суммы равномерно сходятся к функции $t \to e^t$.

Решение задачи Коши $\ddot x + x = 0$, $x(0) = 0$, $\dot x(0) = 1$ для уравнения гармонического осциллятора (выражающего второй закон Ньютона для частицы, притягивающейся к началу согласно закону Гука) сводится к нахождению неподвижной точки оператора $F(x(\cdot))(t) = t - \int_0^t (t-s)x(s)ds$. Стартуя от $x_0(t) = t$, приходим к ряду для синуса (убедитесь в этом самостоятельно).

Для уравнения $\dot x = A(t)x$, где $A(\cdot)$ — непрерывная функция, итеративная процедура приведет к решению $x(t) = Ce^{\int_{t_0}^t A(s)ds}$. Для неоднородного уравнения $\dot x = A(t)x + a(t)$ выпишется тот же ответ, что и методом Лагранжа вариации постоянных. Систему двух уравнений: $\dot x_1 = a_{11}x_1 + a_{12}x_2$, $\dot x_2 = a_{21}x_1 + a_{22}x_2$, $x_1(t_0) = \xi_1$, $x_2(t_0) = \xi_2$ с постоянными коэффициентами можно свести к одному уравнению второго порядка: $\ddot x + p\dot x + q = 0$. Если корни λ_i, $i = 1, 2$, различны, то общее решение $x(t, C_1, C_2) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}$, а если корни одинаковы и уравнение не распадается, то общее решение таково: $x(t, C_1, C_2) = e^{\lambda t}(C_1 + C_2 t)$.

Эти результаты принадлежат, в основном, Эйлеру.

4.2. Общая теория линейных дифференциальных уравнений

Для решения задачи Коши

$$\dot{x} = A(t)x + y(t), \quad x(t_0) = x_0, \tag{4.1}$$

где $A(\cdot)$ — непрерывная размера $n \times n$ матричная функция, а $y(\cdot)$ — непрерывная n-мерная вектор-функция, снова рассмотрим отображение $F\colon C([-a,a],\mathbb{R}^n) \to C([-a,a],\mathbb{R}^n)$, $F(x(\cdot))(t) = x_0 + \int_0^t (A(s)x(s) + y(s))\,ds$. Показывается, что некоторая степень F является сжимающим отображением. Из принципа сжимающих отображений следует существование неподвижной точки, что доказывает глобальную теорему существования задачи Коши для линейных дифференциальных уравнений, согласно которой на всем интервале $(-a, a)$ существует и единственно решение задачи (4.1).

4.3. Общая теория дифференциальных уравнений (существование, непрерывная и дифференцируемая зависимость решений)

Теорема 4 а) (локальная теорема существования решений). *Пусть на открытом множестве $G \subset \mathbb{R} \times \mathbb{R}^n$ функция $f\colon G \to \mathbb{R}^n$, $f = f(t,x)$, удовлетворяет условию Липшица относительно x: $|f(t,x) - - f(t,x')| \leqslant M|x - x'|$ для любых x, $x' \in G$, и пусть компакт $\mathcal{K}$ принадлежит G. Тогда для любой точки $(\widehat{t}, \widehat{x}) \in \mathcal{K}$ существуют такие число $\delta > 0$ и окрестность V точки $\widehat{x}$, что для любой точки (t_0, x_0), удовлетворяющей условиям $|t_0 - \widehat{t}| < \delta$ и $x_0 \in V$, существует единственное решение $t \mapsto x(t, t_0, x_0)$ задачи Коши:*

$$\dot{x} = f(t,x), \quad x(t_0) = x_0, \tag{4.3}$$

определенное на отрезке $[t_0 - \delta, t_0 + \delta]$ и непрерывное по совокупности переменных.

в) (о непрерывной зависимости решения от параметров). *Пусть $G \subset \mathbb{R} \times \mathbb{R}^n$ — открытое множество, $\mathcal{A}$ — топологическое пространство, $f\colon G \times \mathcal{A} \to \mathbb{R}^n$ — функция переменных (t, x, α), $t \in \mathbb{R}$, $x \in \mathbb{R}^n$, $\alpha \in \mathcal{A}$, непрерывна по совокупности переменных и строго дифференцируема по x равномерно по α в произведении $[t_0, t_1]$ на окрестность W точки $(\widehat{x}, \widehat{\alpha})$ (расположенном в $G \times \mathcal{A}$). Тогда существует такая окрестность*

$\mathcal{A}_0 \subset \mathcal{A}$, *что решение задачи Коши*

$$\dot{x} = f(t, x, \alpha), \quad x(t_0) = x_0$$

определено на отрезке $[t_0, t_1]$ *и непрерывно зависит от параметра* α.

Строгая дифференцируемость, равномерная по параметру α, означает, что для любого $\varepsilon > 0$ существует окрестность $W_0 \subset W$ такая, что $|f(t, x, \alpha) - f(t, x', \alpha') - f_x(t, \widehat{x}(t), \widehat{\alpha})(x - x')| < \varepsilon|x - x'|$, если $t \in [t_0, t_1]$, (x, α) и (x', α) принадлежат W_0.

c) (О ЗАВИСИМОСТИ РЕШЕНИЙ ОТ ПАРАМЕТРОВ И ГРАНИЧНЫХ УСЛОВИЙ). *Пусть* $\mathcal{U} \subset \mathbb{R}^m$ *— окрестность нуля в* $\mathbb{R}^m$ *и* $f\colon [t_0, t_1] \to G \times \mathcal{U}$ *— функция, непрерывно дифференцируемая по совокупности переменных. Если* $\widehat{x}(\cdot)$ *— решение задачи Коши, то существует* $\delta > 0$ *и для любого* $|u| < \delta$ *— решение* $x(\cdot, u, w)$ *задачи Коши* $\dot{x} = f(t, x, u)$, $x(t_0) = w$, *непрерывно-дифференцируемо зависящее от граничного условия* w *и параметра* u.

Дадим набросок доказательств теорем a) и b).

a) Рассмотрев отображение $F(x(\cdot))(t) = x(t) - x_0 - \int_{t_0}^{t} f(s, x(s))ds$, убеждаемся, что при малом d оно удовлетворяет условию сжимаемости в $C([t_0 - d, t_0 + d], \mathbb{R}^n)$, и теорема a) следует из принципа сжимающих отображений (и полноты пространства $C([t_0 - d, t_0 + d], \mathbb{R}^n)$.

Для доказательства теоремы b) надо применить метод Ньютона к отображению $F((x(\cdot), \alpha))(t) = \widehat{x}(t) + x(t) - x_0 - \int_{t_0}^{t} f(s, \widehat{x}(s) + x(s), \alpha)ds$ из пространства $C([t_0, t_1], \mathbb{R}^n)$ в пространство $C^1([t_0, t_1], \mathbb{R}^n)$ для решения уравнения $F = 0$, начиная с функции $x_0(t) \equiv 0$. Основное условие применимости метода Ньютона о сюръективности отображения F выполняется, ибо производная этого отображения $F'(0, \widehat{\alpha}))(t) = x(t) - \int_{t_0}^{t} f_x(s, \widehat{x}(s), \widehat{\alpha})x(s)ds$ отображает пространство $C([t_0, t_1], \mathbb{R}^n)$ на пространство $C^1([t_0, t_1], \mathbb{R}^n)$ в силу теоремы существования для линейных систем.

Теорема c) также доказывается прямым применением метода Ньютона.

* * * * * *

Остановимся здесь, чтобы подвести некоторые итоги. Как связаны четыре наши теоремы с окружающим нас миром?

Ньютон научил нас тому, что эволюционные процессы и явления описываются (обыкновенными) дифференциальными уравнениями. Второй закон Ньютона описывает движения взаимодействующих частиц. Теорема существования и единственности решения задачи Коши (теорема 4) выражает идею лапласовского детерминизма: если бы Некто мог в одно мгновение узнать все скорости и положения частиц Мира, он знал бы о том, что было в прошлом и узнал бы, все, что случится в будущем. Это вольный пересказ того, о чем не раз говорил Лаплас. Два столетия идея предопределенности была одной из философских доминант (совсем недавно, каких-то полстолетия тому назад, все дороги вели к коммунизму). И лишь недавно в период *современной математики* обнаружилось, что чаще всего детерминированные процессы в итоге запутываются и скорее напоминают хаотическое движение. Но это уже другая история.

Эйлер писал как-то, что в мире не происходит ничего, в чем не был бы виден смысл какого-нибудь максимума или минимума. Среди прочего имелись в виду *вариационные принципы* в естествознании. Соответствующая математическая теория опирается на правило множителей Лагранжа (см. следствие из теоремы 3). И мы этим следствием подошли совсем близко к тому, чтобы охватить основы вариационной теории и ее приложений к естествознанию.

Тяжелый удар «предопределенности» примерно столетие тому назад был нанесен зарождавшейся квантовой механикой. В двадцатые годы родились две теории — Гейзенберга и Шрёдингера, не слишком похожие друг на друга. И вот ведь какая случилась удача: один из молодых людей, кто имел контакты с творцами квантовой механики, а именно Макс Борн, когда-то в Геттингене слушал Гильберта и Шмидта про симметрические операторы. Он-то и понял, что в рассмотрениях этих операторов в гильбертовом пространстве и лежит родство двух ветвей теории. Такую роль (среди многих иных) сыграла наша теорема 2. А еще с ее помощью решаются смешанные задачи для гиперболических уравнений (см. О. А. Олейник, «Лекции об уравнениях с частными производными», М.: БИНОМ, 2005), краевые задачи теории параболических уравнений.

Помимо частиц существуют тела — струны, балки, мембраны и все такое прочее. И часто случается так, что подобно тому, как решение обыкновенного дифференциального уравнения сводится к интегрированию, так решение уравнений с частными производными описываются интегральными уравнениями (см. упоминавшуюся книгу Петровского по интегральным уравнениям). И здесь важнейшую роль играет теорема 1. И все это лишь доля той пользы, которую могут принести эти теоремы. Так что мы их доказывали не зря.

Приложение. Принципы существования

Принцип компактности Вейерштрасса – Лебега – Бэра. *Функция, полунепрерывная снизу на компакте, принимает свое минимальное значение.*

Доказательство. Обозначим $U_n := X \setminus \mathcal{L}_n(f)$, $n \in \mathbb{Z}$. Очевидно, что $\dots \subset U_n \subset U_{n-1} \subset \dots$. Из определения полунепрерывности снизу следует, что U_n открыты, т.е. $\{U_n\}_{n\in\mathbb{Z}}$ — открытое покрытие X. Тогда из определения компактности получаем, что существует такое число $m \in \mathbb{Z}$, что $X = U_m$, т.е. f ограничена снизу. Пусть $\mu = \inf f$ на X. Если $f(x) \neq \mu$, обозначим $V_n := X \setminus \mathcal{L}_{\mu+(1/n)}(f)$, $n \in \mathbb{N}$. Из определений полунепрерывности снизу и нижней грани следует, что $\{V_n\}_{n\in\mathbb{N}}$ — открытое покрытие X. Значит, в силу компактности, найдется такое число $s \in \mathbb{N}$, что $X = V_s$, т.е. $f > \mu + 1/s$. Противоречие с тем, что $\mu = \inf f$. □

Принцип сжимающих отображений Пикара – Каччиополли – Банаха. *Сжимающее отображение полного метрического пространства в себя имеет единственную неподвижную точку.*

Доказательство. Взяв произвольный элемент $x_0 \in X$, положим $x_n = F(x_{n-1})$, $n \in \mathbb{N}$. Тогда

$$\begin{aligned} d(x_{n+m}, x_m) &\leqslant d(x_{n+m}, x_{n+m-1}) + d(x_{n+m-1}, x_{n+m-2}) + \ldots + d(x_{m+1}, x_m) = \\ &= d(F(x_{n+m-1}), F(x_{n+m-2})) + \ldots + d(F(x_m), F(x_{m-1})) \leqslant \\ &\leqslant d(x_1, x_0)(\theta^{n+m-1} + \ldots + \theta^{m-1}). \end{aligned}$$

Здесь первое неравенство — это неравенство треугольника, а второе неравенство следует из сжимаемости. Полученная оценка показывает, что последовательность $\{x_n\}_{n\in\mathbb{N}}$ фундаментальна и сходится к элементу $\widehat{x} \in X$. Имеем: $d(F(\widehat{x}), x_{n+1}) = d(F(\widehat{x}), F(x_n)) \leqslant \theta d(\widehat{x}, x_n) \to 0$, значит, $d(F(\widehat{x}), \widehat{x}) = 0$, что равносильно $\widehat{x} = F(\widehat{x})$. Единственность легко следует из сжимаемости. □

* * * * * *

И в заключение несколько слов о том, как, по моему мнению, и чему надо учить математике. Я бы выделил пять стадий — четыре, как в каждой из моих лекций, и еще пятую. До школы и в начальной школе разумно научить дитя считать и дать импульс к размышлению. Мои начальные фрагменты лекций призваны дать такие импульсы к размышлениям.

Компьютерам дети научаются сами, но логику, понимание того, что есть научная истина, человек (пока еще) не может извлечь из машины, его надо этому учить, начиная со школы.

На каком-то уровне каждому надо сделать выбор своей жизненной дороги — гуманитарной или научно-технической. Во втором случае разумно приступить к изучению математического анализа. Вторые фрагменты моих лекций обращены к тем, кто хочет узнать о математике, как о предмете, соответствующем выбору их жизненного пути.

На первых курсах технических, экономических и естественно-научных институтов и университетов надо освоить технические основы нашей науки (начала исчислений, линейной алгебры, теории дифференциальных уравнений, комплесного анализа и теории вероятностей). В третьих разделах моих лекций рассказывается о началах линейной алгебры, анализа и дифференциальных уравнений.

А потом, в институтах с хорошей математикой и в университетах я предложил бы читать некий сводный курс математики, где все выстраивалось бы в единую гармоничную картину. Из фрагментов такого курса состоят четвертые части моих лекций, в которых я постоянно подчеркивал, что дистанция между истоком и так сказать «устьем» в каждой нашей теме была невелика. По-видимому, потоки человеческой мысли слишком причудливы и потому так долго текли от своего начала к естественному концу.

Вот вам четыре стадии: начальная школа, средняя и специализированная школа, пропедевтическое институтское или университетское образование и сводный курс математики.

А пятая — обещанная — часть? Пятая часть — *современная математика.* Как учить современной математике и чему? Вот вопрос. Давайте поразмышляем. Но в другой раз.

В. М. Тихомиров, мехмат МГУ
E-mail: `tikhomir@mccme.ru`

Олимпиады и математика

А. Б. Скопенков

Перед учителями и руководителями кружков, занимающимися с сильными школьниками, встает вопрос: как подготовить школьников к олимпиадам или к «серьезной» математике? Некоторые думают, что для первого надо прорешивать задачи последних олимпиад, для второго надо читать научную литературу, и что ввиду принципиальной разницы первого и второго бессмысленно пытаться достичь и того, и другого. Автор этой заметки придерживается распространенного мнения о том, что эти подходы недостаточно эффективны и приводят к вредным «побочным эффектам»: школьники либо чрезмерно увлекаются *спортивным* элементом в решении задач, либо изучают *язык* высшей математики вместо ее содержательной стороны.

Мне кажется, что основу математического образования сильного ученика должно составлять *решение и обсуждение мотивированных для ученика задач, в процессе которых он знакомится с важными математическими идеями и теориями.* Это одновременно подготовит школьника и к математической науке, и к олимпиадам, и не нанесет вред его развитию в целом; это будет более эффективно и для достижения успеха только в олимпиадах или только в науке (если не учитывать большого количества других факторов, кроме разумной организации занятий).

Как и при естественном развитии самой математики, каждая следующая задача должна быть мотивирована либо практикой, либо уже решенными задачами. Поэтому ученик, занимающийся «мотивированной для него» математикой (обычно более элементарной, но обычно содержательной и потому обычно сложной) вместо «немотивированной для него» математики (обычно менее элементарной, но обычно языковой и потому обычно тривиальной), имеет преимущество в дальнейшей учебе и научной работе.

Олимпиадных задач очень много, большинство из них интересны школьнику и среди них много математически содержательных. Такие задачи могут составить основу изучаемого материала. Однако решение олимпиадных задач без изучения математических идей и теорий недостаточно эффективно даже для «чистой» подготовки к олимпиадам (на долгих — год и более — промежутках времени, как и вообще решение сиюминутных задач без фундаментального развития). Кроме того, большинству людей легче достичь успеха на олимпиадах (как и в других видах творческой деятельности) в том случае, когда они не считают успех главной целью. Сложную задачу легче решить, если спокойно думать о самой задаче, а не о награде, которая последует за ее решением. Поэтому школьник, мотивированный более высокой целью, чем успех на олимпиаде, имеет на этой олимпиаде психологическое преимущество.

Как удачно подобрать задачи для обучения учеников? Этот вопрос учителя задавали себе последние десять тысяч лет [2, предисловие], [4], [5, с. 26–33], [7, предисловие]. Расскажу и я о части своего практического опыта, находящегося в русле опыта приведенных авторитетов.

В Московском центре непрерывного математического образования (МЦНМО) под руководством автора этой заметки проходит кружок «Олимпиады и Математика». Моими ассистентами в разное время были и являются А. Акопян, А. Засорин, Д. Пермяков, С. Спиридонов и И. Шнурников. Это студенты механико-математического факультета Московского государственного университета, в прошлом победители Международных и Всероссийских олимпиад школьников. Большинство из них отличники, некоторые из них уже являются авторами научных работ. Аналогичный кружок «Математический Семинар» я веду в физико-математической школе-интернате им. А. Н. Колмогорова (СУНЦ МГУ) с 1994 года (до 2001 года совместно с В. Н. Дубровским).

Участвовать в кружке «Олимпиады и Математика» имеет право любой желающий. Однако уровень занятий довольно высок; большинство участников нашего кружка — ученики 8–11 классов, которые имеют шанс пройти на Всероссийскую олимпиаду.

Активное участие в кружке требует затрат времени и сил, поэтому его желательно согласовать с родителями и учителями. На занятиях кружка школьники учатся решать интересные задачи, подобранные так, что в процессе их решения и обсуждения ученики знакомятся с важными математическими идеями и теориями. В начале каждой темы решаются и разбираются в основном задачи, предлагавшиеся ранее на олимпиадах (или аналогичные таковым). А в конце дело часто доходит до *задач для исследования*. Мы уделяем много времени *индивидуальным* занятиям, разбирая лично с каждым школьником его решения и давая ему подсказки и/или дополнительные задачи, а также занимаемся со школьниками, которые решают исследовательские задачи (и выступают со своими результатами на конференциях школьников). Более подробно о задачах для исследования см., например, [6]. Занятия кружка объединяются в циклы из 1–3 занятий, связанных общей темой или идеей. Разные циклы почти независимы друг от друга, а их аннотации объявляются заранее (поэтому можно изучать только те циклы, которые школьнику наиболее интересны). Некоторые материалы кружка опубликованы в журналах «Квант», «Математическое просвещение», «Математическое образование» и “Quantum”. В хорошую погоду занятия кружка часто проводятся с выездом на природу.

Важной составляющей кружка являются выездные школы. Занятия на этих школах ведут замечательные математики и учителя, там превосходно организован быт и досуг школьников. Приглашаются туда наиболее активные участники кружка, а также те победители Московской и Всероссийской олимпиад, которые живут в Москве или ближнем Подмосковье и не занимаются в кружке.

Подчеркну, что успешное участие в кружке не учитывается при формировании команды Москвы на Всероссийскую Олимпиаду. Но, конечно, оно поможет успешно выступить на любой олимпиаде.

Приведу пример двух циклов задач из разбиравшихся на кружке. Первый цикл подводит школьника к открытию леммы Бернсайда. Заметим, что в существующих учебниках (в том числе для школьников) изучение приложений леммы Бернсайда начинается с немотивированных понятий и теории. Каждая следующая серия выдавалась школьникам после разбора предыдущей. Второй цикл подводит школьника к понятию цепной дроби на примере алгоритма вычисления количества целых точек «под прямой $y = \alpha x$». Другие циклы можно найти в интернете на странице `www.mccme.ru/circles/oim`. (Многие материалы на этой странице не претендуют на оригинальность. Не все материалы там составлены мной; имена других составителей указываются.

Благодарю за замечания и обсуждения, способствовавшие улучшению настоящей заметки, М. Н. Вялого, В. Н. Добровольскую и ее научного руководителя А. А. Русакова, а также всех участников кружка «Олимпиады и Математика».

Комбинаторика классов эквивалентности

1-я серия.

1. Сколькими способами можно раскрасить грани куба в красный и серый цвета? Раскраски, совмещающиеся в (трехмерном) пространстве, считаются одинаковыми.

2. Найдите число раскрасок карусели из n вагончиков в a цветов (т. е. количество раскрасок вершин правильного n-угольника в a цветов, если раскраски, совмещающиеся поворотом, неотличимы) для
(a) $n = 5$; (b) $n = 4$; (c) $n = 6$.

3. Найдите количество закрасок в a цветов

(a) правильного треугольника, разбитого средними линиями на 4 равных треугольника.

(b) то же для 9 треугольников.

(c) квадрата, разбитого средними линиями на 4 равных квадрата.

(d) то же для 9 равных квадратов.

4.* Найдите число замкнутых ориентированных связных ломаных с вершинами в вершинах данного правильного p-угольника (где p — простое). Ломаные, совмещающиеся поворотом, неотличимы.

2-я серия.

Аналогично придуманному вами способу можно решить задачу 2 для произвольного n, однако решение будет громоздким. Приведем более простой (для «очень непростых» n) способ на примере решения задачи 2c.

Обозначим через X искомое число раскрасок. Назовем *занумерованной раскраской* раскраску карусели из шести *занумерованных* вагончиков. Тогда всего есть a^6 занумерованных раскрасок. Для занумерованной раскраски α обозначим через $[\alpha]$ соответствующую незанумерованную раскраску.

Посчитаем двумя способами число P пар (α, n), в которых $n \in \{0, 1, 2, 3, 4, 5\}$ и α — занумерованная раскраска, переходящая в себя при повороте на $2\pi n/6$. Пусть $\operatorname{stab}\alpha$ — число тех $n \in \{0, 1, 2, 3, 4, 5\}$, для которых поворот на $2\pi n/6$ переводит занумерованную раскраску α в себя. Тогда

$$P = \sum_{\alpha} \operatorname{stab}\alpha = \sum_{\text{раскраскам } \beta} \sum_{\{\alpha:[\alpha]=\beta\}} \operatorname{stab}[\alpha] = \sum_{\text{раскраскам } \beta} \operatorname{stab}\beta \frac{6}{\operatorname{stab}\beta} = 6X.$$

Предпоследнее равенство выполнено, поскольку

– для занумерованных раскрасок α_1 и α_2, переводящихся друг в друга поворотами, $\operatorname{stab}\alpha_1 = \operatorname{stab}\alpha_2$ (эти равные числа обозначаются $\operatorname{stab}[\alpha_1]$), и

– количество занумерованных раскрасок, получающихся поворотами из данной занумерованной раскраски α, равно $\dfrac{6}{\operatorname{stab}\alpha}$.

5. Докажите последнее утверждение.

С другой стороны, поворот на $2\pi n/6$ переводит в себя ровно $a^{(n,6)}$ закрепленных раскрасок, поэтому $P = \sum\limits_{n=0}^{5} a^{(n,6)}$. Итак,

$$X = \frac{1}{6}(a^6 + 2a + 2a^2 + a^3).$$

6. Найдите число

(a) раскрасок карусели из n вагончиков в a цветов.

(b) a-цветных ожерелий из n бус.

7. Сколькими способами можно раскрасить в a цветов грани (a) правильного тетраэдра; (b) куба? Раскраски, совмещающиеся вращением (трехмерного) пространства, считаются одинаковыми.

8. Сколько существует различных (т. е. неизоморфных) (не)ориентированных графов с 4 вершинами? А с 5?

9.* Отображения $\{0,1\}^n \to \{0,1\}$ (т. е. функции алгебры логики от n переменных) называются *конгруэнтными*, если они становятся равными после переименования переменных.

(a) Найдите число b_n функций алгебры логики от n переменных с точностью до конгруэнтности.

(b) Докажите, что существует $\lim\limits_{n\to\infty} n!\,b_n/2^{2^n}$ и найдите этот предел.

10. Сформулируйте общую теорему, которую можно было бы применять вместо повторения приведенного решения задачи 2c.

3-я серия. Вот ответ.

Лемма Бернсайда. *Пусть заданы конечное множество M и семейство $\{g_1 = \mathrm{id}\, M,\ g_2, \ldots, g_n\}$ преобразований этого множества, замкнутое относительно композиции и взятия обратного элемента. Назовем элементы множества M* эквивалентными, *если один можно перевести в другой одним из данных преобразований. Тогда число классов эквивалентности равно $\dfrac{1}{n}\sum\limits_{k=1}^{n} \mathrm{fix}(g_i)$, где $\mathrm{fix}(g_i)$ — число элементов множества M, которые преобразование g_i переводит в себя.*

(Чтобы сделать этот и следующие результаты менее доступными, их обычно формулируют и доказывают на языке теории абстрактных групп.)

Назовем *(конечной) группой* конечное семейство преобразований (т. е. перестановок) некоторого множества, замкнутое относительно композиции и взятия обратного элемента. Дальнейшие задачи особенно интересны тем, кто уже решал задачи про перестановки.

11. (a) Любая группа содержит тождественное преобразование.

(b) Если число n преобразований в группе G простое, то для любого нетождественного преобразования $g \in G$ имеем $G = \{g, g^2, \ldots, g^n = \mathrm{id}\}$.

Если в группе G найдется преобразование g, для которого $G = \{g, g^2, \ldots, g^n = \mathrm{id}\}$, то группа G называется *циклической*. Группа G называется *бициклической* (этот термин не общепринят), если для некоторых целых

$p > 1$ и $q > 1$ и преобразований $g, h \in G$ выполнено $gh = hg$, $g^p = \mathrm{id}$, $h^q = \mathrm{id}$ и $G = \{g^k h^l \mid 1 \leqslant k \leqslant p, 1 \leqslant l \leqslant q\}$.

12. Приведите пример бициклической группы из 4 элементов.

13. Любая ли группа из n преобразований является циклической или бициклической для (a) $n = 4$; (b) $n = 6$; (c) $n = 8$; (d)* $n = 9$; (e) $n = 10$; (f)* $n = 15$?

14.* *Теорема Лагранжа.* Если H — подгруппа группы G (т. е. подмножество группы G, также являющееся группой), то $\#G$ делится на $\#H$.

15.* *Теоремы Силова.* Пусть p — простое, n делится на p^k и не делится на p^{k+1}, а G — группа из n элементов. Докажите, что

(a) В G имеется подгруппа из p^k элементов.

(b) Число таких подгрупп сравнимо с 1 по модулю p.

(c) Для любых двух таких подгрупп H и H' найдется такое преобразование $g \in G$, что $H' = \{ghg^{-1} \mid h \in H\}$.

(d) Если p и q — простые, $p < q$ и $q - 1$ не делится на p, то любая группа из pq преобразований является циклической или бициклической.

Целые точки под прямой

Эти задачи подводят школьника к алгоритму вычисления суммы $f_\alpha(n) = \sum\limits_{k=1}^{n} [\alpha k]$, т. е. количества целых точек «под прямой $y = \alpha x$». Через $[x]$ и $\{x\}$ обозначаются целая и дробная части числа x, соответственно. Латинские буквы обозначают целые числа. Алгоритм строится в задачах 2abc, задача 1 полезна «для разогрева».

1. Найдите $f_\alpha(n)$ для (a) целого α; (b) целого 2α; (c) целого 3α; (d) $\alpha = a/n$ для данных целых a и n; (e) Найдите $\lim\limits_{n\to\infty} \dfrac{f_\alpha(n)}{n^2}$.

2. Докажите, что

(a)
$$f_\alpha(n) = f_{\{\alpha\}}(n) + \frac{1}{2}[\alpha]n(n+1).$$

(b)
$$f_\alpha(n) + f_{\frac{1}{\alpha}}([n\alpha]) - [\frac{n}{q}] = n[n\alpha],$$

где q — знаменатель несократимой дроби, представляющей число α, если α рационально, и $q = \infty$ (т. е. $[\frac{n}{q}] = 0$), если α иррационально.

Указание: посчитайте количество целых точек в прямоугольнике $1 \leqslant x \leqslant n$, $1 \leqslant y \leqslant [n\alpha]$.

(c) Найдите алгоритм вычисления суммы $f_\alpha(n)$. Указание:

$$f_{2/3}(n) = n[\frac{2n}{3}] + [\frac{n}{3}] - f_{3/2}([\frac{2n}{3}]);$$

$$f_{3/2}([\frac{2n}{3}]) = \frac{1}{2}[\frac{2n}{3}]([\frac{2n}{3}]+1) + f_{1/2}([\frac{2n}{3}]);$$
$$f_{1/2}([\frac{2n}{3}]) = [\frac{2n}{3}][\frac{n}{3}] + [\frac{n}{3}] - f_2([\frac{n}{3}]),$$

поскольку $[[x]/n] = [x/n]$ для целого $n > 0$ и, значит, $[\frac{[\frac{2n}{3}]}{2}] = [\frac{n}{3}]$;

$$f_2([\frac{n}{3}]) = [\frac{n}{3}]([\frac{n}{3}]+1).$$

(d) Найдите алгоритм вычисления суммы $\sum\limits_{k=1}^{n}\{\alpha k\}$.

Замечания. Частный случай равенства 2b (на котором основан предлагаемый алгоритм) для положительных нечетных взаимно простых чисел $p < q$, $\alpha = p/q$ и $n = (q-1)/2$ (тогда $[n\alpha] = (p-1)/2$) появляется при доказательстве закона взаимности квадратичных вычетов [1, вопрос 16 к главе II и V.2.l]; доказательство общего случая аналогично. Сумма из 2d вычислена (более громоздким способом, чем предложенный здесь) в [3].

Список литературы

[1] Виноградов И. М. *Основы теории чисел.* М.: Наука, 1972.

[2] Генкин С. А., Итенберг И. В., Фомин Д. В. *Ленинградские математические кружки.* Киров, 1994.

[3] Добровольская В. Н. *Неполные суммы дробных долей* // Чебышевский сборник, 2004. Т. 5, №2(10). С. 42–48.

[4] Судзуки Д. *Основы дзэн-буддизма. Наука Дзэн — ум дзэн.* Киев, 1992.

[5] Платон. *Федон*, в кн. Федон, Пир, Федр, Парменид. М.: Мысль, 1999.

[6] Скопенков А. *Исследовательские задачи для школьников* // «Тезисы докладов конференции, посвященной 100-летию С. М. Никольского». М.: МИ РАН, 2005.
Электронный вариант: `www.mccme.ru/circles/oim/100.ps`

[7] Шклярский Д. О., Ченцов Н. Н., Яглом И. М. *Избранные задачи и теоремы элементарной математики. Арифметика и алгебра.* М.: Физматлит, 2001.

А. Б. Скопенков, кафедра дифференциальной геометрии и приложений, механико-математический факультет, Московский государственный университет; Независимый московский университет; Московский институт открытого образования.
E-mail: `skopenko@mccme.ru`

Математический мир

I H É S

А. Б. Сосинский

Институт Высших Научных Исследований под Парижем

Пятьдесят лет тому назад, в середине прошлого века, французский промышленник русского происхождения Лев Мощан (Léon Motchane) увлекся идеей создать во Франции аналог знаменитого американского IAS[1]. В 1958 году такой центр, названный IHÉS[2], был создан, и Лев Мощан стал его первым директором. Об истории и о сегодняшнем дне этого уникального научного центра наш рассказ.

[1]Institute for Advanced Studies – Принстонский Институт Высших Исследований.

[2]Institut des Hautes Études Scientifiques – Институт Высших Научных Исследований, сегодня находящийся в Bures-sur-Yvette под Парижем.

Немного об истории IHÉS

Наверное, стоит сказать несколько слов о создателе IHÉS. Лев Мощан родился в Санкт-Петербурге в 1900 году. Эмигрировав из России во время гражданской войны, в 1918 году он оказался в Швейцарии практически без средств к существованию. Тогда он не имел возможности получить научное образование, к которому с юных лет стремился. Пришлось сразу зарабатывать на жизнь, сначала в Швейцарии, а затем во Франции, куда Мощан перебрался в 1924 году. И он в этом преуспел: в послевоенные годы Léon Motchane уже был крупным французским промышленником. Будучи обеспеченным человеком, он вернулся к своей юношеской любви к математике и, представьте себе, в возрасте 54 лет (!), защитил диссертацию у выдающегося французского математика Гюстава Шоке (Gustave Choquet). После этого он и загорелся мечтой создать во Франции научный центр наподобие IAS, и приложил к этой цели свои незаурядные деловые и организаторские способности.

Создать IHÉS удалось не сразу. Несколько лет ушло на выбивание денег (капитала самого Мощана, конечно же, для этой цели было мало) и на уговаривание властей поддержать затею. Важную роль сыграла поддержка Роберта Оппенгеймера (Robert Oppenheimer), знаменитого американского физика, в то время директора IAS, и Жана Дьёдонне, одного из основателей группы Бурбаки[3)] и, пожалуй, самого активного французского математика середины XX века. Наконец, в 1958 году, IHÉS открыл свои офисы, первоначально расположенные в центре Парижа, и туда начали приезжать на один, два, или три месяца ведущие математики всего мира. Определилась административная структура и принципы научной организации центра. Были назначены первые два постоянных профессора IHÉS: ими стали Дьёдонне и великий французский математик А. Гротендик (Alexandre Grothendieck).

Важной вехой в становлении IHÉS стало приобретение в местечке Бюр-сюр-Иветт (двадцать пять минут от центра Парижа на загородном метро) десятигектарного лесного участка Буа Мари, куда переехали все службы центра и где он находится по сей день. Уже в семидесятые годы к этой недвижимости прибавился другой участок, Ормай (Résidence Ormaille), в десяти минутах хода от Буа Мари; в Ормай расположены квартиры приезжающих ученых.

В 1974 году Лев Мощан ушел на пенсию, его на должности директора сменил голландский математик Н. Куйпер (Nicolas Kuiper), его в свою очередь заменил знаменитый французский геометр М. Берже (Marcel

---

[3)] О группе Бурбаки можно прочитать в «Математическом Просвещении», вып. 2, 1998; с. 4–12.

Berger). С 1994 года французский математик Ж.-П. Бургиньон (Jean-Pierre Bourguignon) — директор IHÉS.

Как организован IHÉS

В IHÉS совсем небольшой штат постоянных работников: директор, несколько лиц, занимающих административные и секретарские должности, работники бытовых служб и всего шесть постоянных профессоров. Основную же часть работающих в центре людей составляют приглашенные ученые: математики и физики-теоретики достаточно высокого международного уровня, подавшие индивидуальные заявки и приглашенные в IHÉS его Ученым Советом (возглавляемым директором и составленным в основном из постоянных профессоров).

В настоящее время постоянные профессора IHÉS следующие: математики – француз А. Конн (Alain Connes), русские Михаил Громов и Максим Концевич, француз Л. Лафорг (Laurent Lafforgue) и физики — француз Т. Дамур (Thibault Damour) и россиянин Никита Некрасов. Все математики (кроме Громова) — лауреаты высшей математической награды — премии Филдса; Громов же считается крупнейшим геометром XX века (впрочем, у него есть прекрасные работы и по алгебре, и по анализу). Примечательно, что среди шести названных выше профессоров — трое русских. Это совсем не связано с происхождением основателя IHÉS, а просто, как мне кажется, объясняется общепризнанным высоким уровнем российской математической школы (в данном случае — школы В. А. Рохлина в Санкт-Петербурге и И. М. Гельфанда в Москве), а также печальным состоянием «заботы» властей об отечественной науке, фактически вынуждающей лучших молодых научных талантов к эмиграции (и Громов, и Концевич — французские граждане, и я не сомневаюсь, что вскоре таковым станет Некрасов).

В отличие от других подобных центров, выбор приглашенных ученых (они приезжают на месяц или два, в виде исключения и на бóльшие сроки) сознательно носит стихийный характер: не проводятся тематические конференции, не организуются семестры или триместры по определенным разделам математики или теоретической физики — приезжающие ученые встречают там в основном незнакомых коллег, работающих в смежных (или далеких) областях. Этот необычный организационный принцип, восходящий, видимо, к основателю IHÉS Льву Мощану, доказал свою эффективность. Руководители центра считают неожиданные встречи ученых, пусть далеких друг от друга по тематике, более плодотворными, более взрывными, чем общение специалистов и так знакомых с работами и методами друг друга. В IHÉS постоянно работают два или три семинара, где можно выступить или послушать коллег.

Разумеется, имеется хорошая компьютерная поддержка работы приезжих. Есть неплохая математическая библиотека прямо в Буа Мари, и отличная — в университете Орсэ (Université Paris-Sud à Orsay). Есть удобный способ публикации — так называемые препринты IHÉS (которые в последние годы превратились из бумажных в чисто электронные). Словом, созданы идеальные условия для научной работы — живи, мол, и твори!

Жизнь в IHÉS

Что вас ждет, если вас пригласят в IHÉS?

Приехав в Бюр в качестве приглашенного профессора, вы получите отдельный офис в основном здании в Буа-Мари и удобную квартиру в Ормай (если вы с семьей и детьми — коттедж). В один из первых дней вы скорей всего сделаете доклад на одном из постоянно действующих семинаров, тем самым познакомитесь с частью других находящихся в это время в IHÉS математиков и физиков.[4)]

Жизнь приезжающих ученых ничем не регламентирована. Вы можете не приходить в свой офис, гулять целыми днями по огромному лесному участку Буа Мари или пропадать в Париже, от культурных соблазнов которого никто вас отговаривать не будет (разве что совесть). Но вскоре вы узнаете, что хоть и нет жестких правил, зато есть традиции, с которыми следует считаться.

Первая и главная из них — обед. На него принято ходить. Происходит он в час дня в столовой рядом с административным зданием. Столовая чем-то напоминает монашескую трапезную: очень длинные деревянные столы без скатертей, вдоль них скамейки. Обед идет неторопливо. Меню меняется изо дня в день, но на данный день фиксировано. Неизменно подается неплохое столовое вино. За каждым из столов одновременно идут два или три разговора на математические и околоматематические темы. Один из моих французских коллег как-то мне сказал, что, по его оценкам, за этими столами за последние 50 лет родилось больше теорем, чем в любом другом месте в мире. Может быть, это и преувеличение, но, во

[4)] Я помню, какое впечатление все это произвело на меня, когда я впервые оказался в Бюре в 1992 году. Мне дали офис на первом этаже между двумя филдсовскими лауреатами — Конном и Р. Томом (René Thom), на моем первом докладе активно участвовали в дискуссии и Берже, и Том, и Громов. Я вначале чувствовал себя явно не на месте, эдаким самозванцем, обманом проникшим в круг избранных. Но вскоре непринужденная обстановка IHÉS меня успокоила и я получил от того пребывания много удовольствия и пользы. А месяц спустя, уже в Москве, вновь вспомнил подробности жизни в Ормай и расхохотался, когда Яков Григорьевич Синай, только что вернувшийся из Бюра, мне сказал: «Лёша, а я месяц спал в твоей постели» (его поселили в той же квартире, что и меня).

всяком случае, бюровские обеды бесспорно являются сильным катализатором научного творчества и объединения ученых.

Вот только один пример. В 1998 году Максим Концевич рассказывал сидящим за его столом в трапезной IHÉS про занятия Наполеона математикой и о так называемой «теореме Наполеона», будто бы впервые доказанной будущим французским императором. Его с интересом слушал Конн, который об этом ничего не знал, и попросил сформулировать теорему. Концевич по ошибке вместо теоремы Наполеона сформулировал похожую, но значительно более трудную теорему Морли. Конн призадумался над доказательством, которое ему далось не сразу, но, видимо, подгоняемый мыслью — если Наполеон смог, то неужели я не смогу? — наконец нашел доказательство. Его можно прочитать в 9-м выпуске «Математического просвещения», 2005, с. 100–104. Кстати, во введении к этой статье Конн отмечает важную роль обедов в IHÉS.

Менее обязательной, но тоже важной традицией является ежедневное чаепитие — файв-о-клок — происходящее, разумеется, в 17:00 в одной из гостиных административного здания, рядом с офисом директора. Чай или кофе, пирожные, печенья, разбросанные по большой уютной комнате с камином журнальные столики, диваны и кресла, пара небольших досок с мелом. Кто-то сидит, кто-то стоит или расхаживает, у одной из досок собралось два-три человека, оживленно обсуждающие написанные на ней математические каракули. Здесь тоже, быть может, рождается очередное математическое открытие.

С камином связана следующая история, произошедшая лет шесть тому назад в этой гостиной. Тогда я чувствовал себя уже своим человеком в непринужденной обстановке IHÉS и был не прочь даже пошутить, как говорится, на грани фола. Находясь на чаепитии, я обратил внимание на металлическую модель многогранника Коннели, лежащую рядом с камином, явно в роли мехов. В центре гостиной директор IHÉS Бургиньон беседовал с П. Картье (Pierre Cartier), знаменитым французским математиком, одним из первых активистов Бурбаки. Я подошел к ним и очень громко, чтобы было слышно во всей гостинице, заявил:

– Что это за безобразие у вас лежит в качестве мехов рядом с камином?! Его использование с этой целью бросает тень на квалификацию дирекции этого института и на компетентность его научных сотрудников!

– Как, Сосинский, неужели ты не узнал многогранник Коннели, — налетел хорошо знающий меня Картье, — это же знаменитый пример непрерывно изгибаемого полиэдра, который...

– Знаю, знаю, но вы же хотите его использовать в качестве мехов!

– Ну и что?

– А то, что московский математик Иджат Сабитов недавно доказал, что объем таких многогранников не изменяется при изгибании, поэтому сколько ни качай эти «мехи», никакого дуновения воздуха ты не создашь. Вот я и говорю про квалификацию дирекции и ученых...

Характерно для IHÉS, что ни Картье, ни Бургиньон не обиделись на мою выходку, очень смеялись, а потом просили меня рассказать работу Сабитова на семинаре...

Конечно, открытия часто происходят и более традиционном образом — в непринужденных парных беседах за доской в офисах, или в размеренных прогулках по аллеям холмистой территории леса Буа-Мари.

Вечером здесь быстро все затихает — математики, в большинстве своем «совы», бдят допоздна за своими компьютерами. Отвлекающих моментов вечером в тихом Бюр-сюр-Иветт просто нет никаких. Днем, правда, любители целлулоидного шарика гоняют его по пинг-понговому столу, а иногда в Ормай происходят волейбольные сражения, особенно захватывающие, когда их организует Александр Александрович Кириллов, непревзойденный по этому делу мастер.

Ваш срок — месяц или два — пролетает незаметно. Слишком быстро приходится возвращаться к обычной жизни у себя дома...

Засилье русскоговорящих

В последний приезд в IHÉS (не будучи приглашенным профессором, я находился в Париже проездом после конференции в Неаполе и заехал побеседовать с Громовым), я подсчитал (по постоянно обновляемой адресной книжке Ормай) процент русскоговорящих среди приглашенных профессоров: их оказалось около двадцати пяти человек из шестидесяти с лишним — около тридцати процентов! Это объясняется уже отмеченной мной высокой репутацией наших (нынешних и бывших) сограждан-математиков, а также их желанием на родном языке пообщаться с такими математическими звездами, как Громов и Концевич. Как заметил один мой остроумный московский коллега: «мы экспортируем шахматистов, нефть и газ, хоккеистов и математиков».

Большинство из русскоговорящих приезжают не из России, а из США, или Израиля, или Германии — из той страны, куда их забросила их эмигрантская судьба. Большинство не испытывает особой ностальгии по покинутому ими советскому государству, с тех пор канувшему в Лету, они уже стали гражданами другой страны, но профессионального или душевного общения на родном языке — ой как не хватает.

К сожалению, это обстоятельство приводит к отрицательному явлению в нынешней жизни IHÉS. Русскоязычные математики держатся несколько обособлено и общаются преимущественно друг с другом (да еще на непонятном русском языке!), что вызывает вполне обоснованное раздражение многих коллег. Пока попытки Конна, Лафforга и даже Громова как-то уладить эту проблему к успеху не приводят.

Несомненно, приятно осознавать значимость российской математической школы, как и приятно оказаться под Парижем в «привилегированном советском научном доме отдыха» (по меткому выражению одного моего коллеги). Но, конечно, IHÉS нельзя считать центром, созданным выходцем из России для выходцев из России — это уникальное место встречи лучших математиков (и математических физиков) мира, со своими своеобразными традициями и удивительной научной аурой.

А. Б. Сосинский, ВКМ НМУ
E-mail: `abs@diffiety.ac.ru`

Наш семинар: математические сюжеты

Четыре алгоритмических лица случайности

В. А. Успенский

Введение

Если кто-либо скажет нам, что он подбросил «честную» монету двадцать раз и, обозначив герб единицей, а решетку нулем, получил такой результат:

$$10001011101111010000 \tag{I}$$

или такой:

$$01111011001101110001, \tag{II}$$

мы вряд ли будем удивлены. Однако если нам скажут, что результат бросаний был таким:

$$00000000000000000000 \tag{III}$$

или таким:

$$01010101010101010101 \tag{IV}$$

— мы будем поражены и вообще не поверим или же усомнимся в корректности эксперимента. Возникает вопрос: почему?

По-видимому, цепочки (I) и (II) воспринимаются как случайные, а цепочки (III) и (IV) — как неслучайные.

Но что означают слова «воспринимается как случайная»? Классическая теория вероятностей не дает ответа на этот важный вопрос. Не столь редко можно услышать следующее объяснение: вероятность каждой из цепочек (III) и (IV) слишком мала, она равна 2^{-20}, что меньше одной миллионной. Но ведь ровно такую же вероятность имеют цепочки (I) и (II).

Математическое просвещение, сер. 3, вып. 10, 2006(71–108)

Три важных замечания.

Во-первых, впечатление случайности зависит от распределения вероятностей. Если одна сторона монеты тяжелее другой или если человек в процессе бросания научается подкидывать монету так, чтобы она падала на нужную сторону, то появление цепочек (III) или (IV) может стать вполне ожидаемым. Вначале — для простоты — мы будем заниматься лишь независимыми бросаниями с вероятностями одна вторая для герба и одна вторая для решетки.

Во-вторых, говорить о случайности имеет смысл лишь в применении к очень длинным цепочкам. Бессмысленно спрашивать, которая из цепочек 00, 01, 10, 11 более случайна, чем другие.

В-третьих, точной границы между случайными и неслучайными (в интуитивном смысле) цепочками нет и не может быть. Ведь если в случайной цепочке заменить один знак, она остается случайной. Но, заменяя много раз, мы от любой цепочки можем прийти к (III) или (IV). Это известный парадокс кучи.

Итак, следует рассматривать только очень длинные цепочки, а в идеале — бесконечные. (Вообще, бесконечность — это такое полезное приближение сверху к очень большому конечному.) Бесконечные цепочки принято называть *последовательностями.* Оказывается, последовательности уже можно довольно осмысленно разделить на случайные и неслучайные. Иными словами, можно не без успеха пытаться найти строгое математическое определение для понятия „случайная последовательность нулей и единиц“. В настоящем очерке мы изложим результаты таких попыток, предпринятых различными авторами. Однако следует честно признать, что для практических приложений интерес представляют именно конечные случайные цепочки и потому идеализация, происходящая при переходе к цепочкам бесконечной длины, неизбежно сопряжена с «отрывом от жизни». Впрочем, аналогичный отрыв возникает и при изучении совокупности *всех* конечных цепочек, поскольку в жизни встречаются лишь цепочки ограниченной длины, а в очень длинных случайных цепочках возникают такие эффекты, которые могут и не соответствовать наивным представлениям о случайности.

Сделав это признание, приступим к изложению.

Нам будут полезны некоторые термины и обозначения.

Всякая конечная цепочка $\langle i_1, \ldots, i_n \rangle$ нулей и единиц называется также *двоичным словом*, а число n — *длиной* этого слова. Длина слова x обозначается так: $|x|$. Длина слова может быть равна нулю; слово длины нуль ничего не содержит и называется *пустым.* Пустое слово обозначается буквой Λ.

Множество всех двоичных слов обозначается буквой Ξ. Множество всех последовательностей нулей и единиц обозначается буквой Ω.

Для последовательности $\langle a_1, a_2, a_3, a_4, \ldots \rangle$ каждое из двоичных слов $\langle a_1, a_2, \ldots, a_n \rangle$ является ее *началом*. Пусть $x \in \Xi$ — двоичное слово. Множество всех тех последовательностей из Ω, для которых x служит началом, обозначим через Ω_x. Каждое множество вида Ω_x, где $x \in \Xi$, принято называть *шаром*. Шару Ω_x приписывается его *объём* $\mathbf{v}(x)$, равный $2^{-|x|}$.

Содержательно, каждую последовательность из Ω мы рассматриваем как серию результатов бросаний монеты. Еще раз подчеркнем, что для наглядности мы ограничиваемся на первых порах ситуацией, когда результаты всех бросаний равновероятны. На математическом языке это означает, что на Ω задано так называемое *равномерное распределение вероятностей* — а это, в свою очередь, означает следующее: для каждого шара вероятность того, что случайно выбранная последовательность из Ω попадет в этот шар, равна объему этого шара.

Наша цель — попытаться выделить из Ω некоторое точно описанное подмножество, претендующее на роль множества всех случайных последовательностей. Традиционная теория вероятностей не только не приближается к решению этой задачи, но даже не может ее сформулировать в своих терминах. На помощь приходит теория алгоритмов. Может показаться парадоксальным, что понятие случайности уточняется на основе такого чуждого случайности понятия, как алгоритм, — тем не менее дело обстоит именно так: все известные до сих пор определения случайности индивидуального объекта (в нашем примере — индивидуальной последовательности нулей и единиц) опираются на понятие алгоритма.

Чтобы найти требуемое определение, поступают так. Формулируют некое характеристическое свойство, которым обладают случайные (в неформальном, интуитивном смысле) последовательности. А затем последовательности, обладающие этим свойством, и объявляют — по определению — случайными.

Какими же свойствами обладает случайная последовательность нулей и единиц?

Во-первых, она *частотоустойчива*. Вот что это означает для того простейшего случая, когда нули и единицы равновероятны: частота нулей, как и частота единиц, стремится к одной второй. (*Частота нулей* — это их доля в начальном отрезке последовательности.) Но более того: в случайной последовательности указанная устойчивость частот выполняется не только для последовательности в целом, но и для любой ее законной, разумной подпоследовательности.

Во-вторых, она *хаотична*. Это означает, что она сложно устроена и не может иметь разумного описания. Психологическому эксперименту, с которого мы начали, Колмогоров дал такое объяснение. Цепочки (I) и (II) потому воспринимаются как случайные, что они сложны, их устройство нельзя коротко описать. А вот цепочки (III) и (IV) имеют простое, легко описываемое устройство.

В-третьих, она *типична*. Это означает, что она принадлежит любому разумному большинству.

В-четвертых, она *непредсказуема*. Это означает, что играя против нее на деньги (т.е. пытаясь угадать члены последовательности и делая ставки), последовательность невозможно обыграть, какой бы разумной стратегией ни пользоваться.

Слово «разумный», встречающееся в объяснениях перечисленных четырех свойств, конечно, нуждается в уточнении. Теория алгоритмов как раз и предлагает такие уточнения, наполняя это слово точным смыслом — своим для каждого из наших четырех свойств. Тем самым возникают четыре алгоритмических свойства: *частотная устойчивость*, *хаотичность*, *типичность*, *непредсказуемость*. Каждое из них представляет свое собственное алгоритмическое лицо случайности, и каждое из них с большими или меньшими основаниями может претендовать на роль строгого математического определения понятия случайности. Можно сказать и так: возникают четыре точно очерченных класса последовательностей, каждый из которых претендует на то, чтобы служить «истинным» классом случайных последовательностей; некоторые из этих претензий более оправданы, чем другие.

Лицо Первое: Частотоустойчивость и стохастичность

По-видимому, одним из первых поставил вопрос о том, что такое отдельно взятая случайная последовательность, замечательный немецкий математик Рихард фон Мизес в начале XX века — в 1919 г. Во всяком случае, именно он первым предложил сравнительно удачное (хотя и нестрогое) определение, послужившее отправной точкой для дальнейшего развития.

Мизес исходил из того, что случайной последовательности должна быть присуща устойчивость частот. А именно, доля единиц (как и доля нулей) в начальном отрезке случайной последовательности должна стремиться к одной второй при неограниченном увеличении длины начального отрезка. Но этого недостаточно. Например, этим свойством устойчивости частот обладает заведомо неслучайная последовательность

$$0, 1, 0, 1, 0, 1, 0, 1, \dots .$$

Очевидным образом необходимо, чтобы устойчивостью частот обладала не только вся последовательность целиком, но и ее подпоследовательности. Однако устойчивостью частот не могут обладать *все* подпоследовательности. В самом деле, возьмем самую что ни на есть случайную последовательность и отберем в подпоследовательность те ее члены, которые равны нулю... Значит, подпоследовательность, в которой должна соблюдаться устойчивость частот, должна быть *разумной*, или *допустимой*. Например, если отобрать в подпоследовательность все те члены, чьи номера суть простые числа, или же все те члены, которые непосредственно следуют за нулевыми, то полученная в каждом из этих двух вариантов подпоследовательность является допустимой.

Последовательности, в которых устойчивость частот соблюдается во всех ее допустимых подпоследовательностях, Колмогоров предложил называть *стохастическими*.

Сам Мизес объяснял, какие последовательности следует считать допустимыми, в расплывчатых терминах. Первое уточнение предложил в 1940 г. американский математик Алонзо Чёрч — один из создателей теории алгоритмов. Чёрч предложил, чтобы допустимые подпоследовательности строились на основе определенных алгоритмов. Последовательности, в которых устойчивость частот наблюдается во всех допустимых по Чёрчу подпоследовательностях, получили название *стохастических по Чёрчу*. Однако оказалось, что определение Чёрча слишком широко: существует, например, стохастическая по Чёрчу последовательность, перестающая быть таковой после вычислимой перестановки ее членов.

В 1963 г. Колмогоров усовершенствовал конструкцию Чёрча, предложив более обширный класс допустимых алгоритмов отбора членов исходной последовательности в подпоследовательность и тем самым заведомо не меньший (а на деле более обширный) класс допустимых подпоследовательностей; в частности, колмогоровский алгоритм разрешает членам подпоследовательности иметь иной порядок следования, чем в исходной последовательности. Поэтому класс всех последовательностей, *стохастических по Колмогорову* (т.е. таких, у которых устойчивость частот наблюдается во всех допустимых по Колмогорову подпоследовательностях), является подклассом класса последовательностей, стохастических по Чёрчу, — на самом деле даже строгим подклассом, т.е. не совпадающим с объемлющим классом. Для дальнейших ссылок класс всех последовательностей, стохастических по Колмогорову, обозначим буквой **S**.

Выяснилось, что класс **S** также слишком широк. Оказалось, например, что может быть построена такая стохастическая по Колмогорову последовательность, в каждом начальном отрезке которой единиц больше, чем нулей; а это противоречит как интуиции, так и законам теории вероятностей (например, закону возвратности состояний при случайном блуждании). Таким образом, даже наиболее продвинутое из известных на сегодняшний день математически строгих уточнений идей Мизеса не дает полного отражения интуитивного представления о случайности (хотя и это неполное отражение может оказаться полезным).

Для большей ясности приведем конструкции Чёрча и Колмогорова. Центральным для обеих конструкций является указание тех правил, согласно которым из членов рассматриваемой последовательности составляются допустимые подпоследовательности. Поскольку каждое такое правило производит отбор членов последовательности для включения их в допустимую подпоследовательность, сами эти правила принято называть *допустимыми правилами отбора*.

Для наглядности представим себе, что члены исследуемой последовательности написаны на картах. Карты лежат одна за другой, лицевой стороной вниз, так что мы не видим, что на них написано. Наша цель — отобрать некоторые из этих карт и составить из них другую последовательность — допустимую подпоследовательность. (Как мы увидим, в случае колмогоровской конструкции термин «подпоследовательность» имеет более широкий смысл, чем это обычно принято.) Допустимое правило отбора представляет собою алгоритм, который на каждом этапе построения подпоследовательности предписывает, во-первых, которую из карт надлежит открыть и, во-вторых, следует или нет включать эту открываемую карту в подпоследовательность. Входом алгоритма служит информация о всех уже открытых к этому моменту картах. Может случиться, что алгоритм отберет лишь конечное число членов исходной последовательности; в этом случае считается, что никакой допустимой подпоследовательности не образовалось. (Напомним, что исходная последовательность признаётся стохастической при наличии устойчивости частот в любой ее допустимой подпоследовательности.)

Переходим к более точному описанию.

Функция называется *вычислимой*, коль скоро существует *вычисляющий* ее алгоритм. При этом говорят, что алгоритм *вычисляет* функцию f, коль скоро он перерабатывает всякий ее аргумент x, на котором функция определена, в соответствующее значение $f(x)$ и не выдает никакого результата в применении ко всякому такому аргументу, на котором функция не определена.

Исследуемую на устойчивость последовательность обозначим $a_1, a_2, a_3, \ldots$, так что на n-й карте записана цифра a_n, равная 0 или 1.

Допустимое по Чёрчу правило отбора является произвольной вычислимой функцией G, определенной на множестве Ξ всех двоичных слов и принимающей одно из двух значений: *Да* и *Нет*. Карты открываются последовательно, одна за другой, начиная с первой, и каждый раз — до открытия очередной карты — решается вопрос, включать ли эту карту в подпоследовательность. Вопрос этот решается следующим образом. Пусть уже открыты n карт, на которых записаны, соответственно, цифры $a_1, \ldots, a_n$. Если $G(a_1, \ldots, a_n)$ есть *Нет*, то следующая, $(n+1)$-я карта не включается в подпоследовательность. Если $G(a_1, \ldots, a_n)$ есть *Да*, то следующая, $(n+1)$-я карта включается в подпоследовательность в качестве очередного члена. Включать или не включать в подпоследовательность член a_1, зависит от значения $G(\Lambda)$. Таким образом, правилу G отвечает бесконечная подпоследовательность (в случае, если таковая образовалась) $a_{n(1)}, a_{n(2)}, a_{n(3)}, \ldots$, где числа $n(1)$, $n(2)$, $n(3)$, ... образуют возрастающую последовательность, составленную из всех чисел n, для которых $G(a_1, \ldots, a_{n-1}) =$ *Да*.

Изложению конструкции Колмогорова предпошлем следующее обобщение понятия подпоследовательности. *Обобщенной подпоследовательностью* последовательности $a_1, a_2, \ldots, a_n, \ldots$ назовем всякую последовательность вида

$$a_{\varphi(1)}, a_{\varphi(2)}, \ldots, a_{\varphi(k)}, \ldots,$$

для которой выполнено условие

$$i < j \implies \varphi(i) \neq \varphi(j).$$

Для обычной подпоследовательности (которая является частным случаем обобщенной) это условие заменено на более сильное — с правой частью $\varphi(i) < \varphi(j)$.

Каждое допустимое по Колмогорову правило отбора имеет целью создать некоторую обобщенную подпоследовательность исходной последовательности. Мы потому написали «имеет целью создать», а не «создает», что процесс создания может прерваться, и тогда в результате применения правила возникает не бесконечная последовательность, а *кортеж* (т. е. конечный набор), составленный из членов исходной последовательности. Для стохастичности по Колмогорову требуется устойчивость частот в каждой из допустимых, т. е. образованных каким-либо допустимым по Колмогорову правилом, обобщенных подпоследовательностей.

Само колмогоровское (т. е. допустимое по Колмогорову) правило состоит из двух вычислимых функций: F и G. Первая служит для образования некой вспомогательной обобщенной подпоследовательности. Окончательная же обобщенная подпоследовательность исходной последовательности строится при помощи функции G в качестве обычной подпоследовательности вспомогательной обобщенной подпоследовательности. Аргументами функций F и G служат двоичные слова, так что каждая из этих функций определена на своем подмножестве множества Ξ. Значениями функции F служат целые положительные числа, у функции G два возможных значения: *Да* и *Нет*. Сперва строится последовательность натуральных чисел $n(1), n(2), n(3), \ldots$:

$$n(1) = F(\Lambda),\ n(2) = F(a_{n(1)}),\ \ldots,\ n(k+1) = F(a_{n(1)}, \ldots, a_{n(k)}).$$

Построение этой последовательности прекращается (и возникает конечный кортеж), как только наступает один из трех случаев:

- ▷ значение $F(a_{n(1)}, \dots, a_{n(k)})$ не определено;
- ▷ значение $G(a_{n(1)}, \dots, a_{n(k)})$ не определено;
- ▷ значение $F(a_{n(1)}, \dots, a_{n(k)})$ совпадает с одним из чисел $n(1), \dots, n(k)$.

Если же остановки в построении не произошло и возникла бесконечная последовательность номеров $n(1), n(2), n(3), \dots$, то далее строится вспомогательная обобщенная подпоследовательность $a_{n(1)}, a_{n(2)}, a_{n(3)}, \dots$. Из нее, наконец, отбираются — в порядке возрастания k — те ее члены $a_{n(k)}$, для которых выполнено равенство $G(a_{n(1)}, \dots, a_{n(k-1)}) =$ *Да*.

Лицо Второе: Хаотичность

Вернемся к примерам конечных цепочек из Введения и вспомним объяснение, предложенное Колмогоровым: цепочки (I) и (II) воспринимаются как случайные, потому что они *сложны*; цепочки (III) и (IV) воспринимаются как неслучайные, потому что они *просты*. По-видимому, мы ожидаем, что результат случайного эксперимента окажется сложным, и удивляемся, когда получаем что-то простое.

Некоторые объекты можно квалифицировать как большие или маленькие, другие — как тяжелые или легкие, третьи — как сложные или простые. В начале 60-х годов Колмогоров наметил математическую теорию, позволяющую оценивать сложность объектов. Сейчас эта теория называется ТЕОРИЕЙ КОЛМОГОРОВСКОЙ СЛОЖНОСТИ.

В основе теории колмогоровской сложности лежит следующая простая и естественная идея:

сложность объекта измеряется длиной его кратчайшего описания.

В самом деле, каждый объект может иметь сколь угодно сложные описания, но сложный объект невозможно описать коротко.

Пусть Y — множество всех объектов, которые мы рассматриваем, а X — множество всех мыслимых описаний этих объектов. Через $|x|$ обозначим длину описания x. Сложность объекта y обозначим $\mathrm{Comp}\,(y)$. В соответствии со сказанным,

$$\mathrm{Comp}\,(y) = \min_x \{|x| : x \text{ есть описание для } y\}.$$

Если объект y неописуем, т.е. для него не существует описания, то, естественно, его сложность равна бесконечности.

Разумеется, длины надо мерить единообразным способом, чтобы не получилось, что описание одного и того же объекта имеет на китайском языке длину единица (один иероглиф), а на русском — сорок (сорок букв). Поэтому все описания кодируются в двоичном алфавите — в виде двоичных цепочек. Вот эти-то двоичные коды мы и будем отныне считать описаниями, так что отныне $X = \Xi$.

Множество всех таких пар $\langle x, y\rangle$, что x служит описанием для y, естественно называть *языком описания*.

Заметим, что и один и тот же объект может иметь в данном языке много описаний, и одно и то же описание может служить описанием многих объектов — таково, например, описание: «двоичное слово, состоящее только из нулей» (а, скажем, выражение «двоичное слово» годится в качестве описания для любого двоичного слова).

Всё сказанное носило предварительный характер, чтобы оправдать нижеследующее формальное изложение.

В декартовом произведении $\Xi \times Y$ рассматриваем произвольное подмножество E, которое называем *языком описания*. Если $\langle x, y\rangle \in E$, будем говорить, что x является *описанием* объекта y. *Сложность* Comp_E объекта y *относительно* языка E определяется так:

$$\mathrm{Comp}_E(y) = \min_x\{|x| : \langle x, y\rangle \in E\}.$$

Напомним, что минимум по пустому множеству принято считать равным бесконечности.

Для языка $E = \Xi \times Y$, в котором каждое x служит описанием для каждого y, сложность любого объекта y равна нулю, поскольку одним из описаний этого y является пустое слово; такой язык мыслим, но не будет встречаться в тех семействах языков, которые мы будем рассматривать.

Представим себе два языка, причем описание какого-либо объекта на втором языке происходит путем удвоения описания этого же объекта на первом языке. Ясно, что второй язык «хуже» первого. Предпочтительнее те языки, которые в состоянии давать более короткие описания.

Будем говорить, что язык A *не хуже* языка B и писать $A \leqslant B$, если существует такая константа c, что для всякого y справедливо неравенство $\mathrm{Comp}_A(y) < \mathrm{Comp}_B(y) + c$, т. е. если

$$\exists c\, \forall y\ \mathrm{Comp}_A(y) < \mathrm{Comp}_B(y) + c.$$

Если принять, что для так называемых естественных языков, т. е. для тех, которыми пользуется человечество, могут быть сформулированы правила перевода с одного языка на другой, то становится очевидным, что каждый из этих языков не хуже другого. Ведь, скажем, турецкое описание какого-либо объекта можно составить так: взять произвольное японское описание этого объекта и дополнить его правилами японо-турецкого перевода. И турецкое, и японское описания, и правила перевода считаются закодированными в виде двоичных слов. Поэтому длина такого турецкого описания равна длине японского описания плюс не зависящая от выбора объекта длина правил перевода. В качестве японского описания выберем кратчайшее. В итоге получаем, что турецкий язык не хуже японского.

Встает вопрос о выборе оптимального языка — такого, который не хуже любого другого. Пусть $\mathcal{L}$ — некоторое языково́е семейство, т. е. попросту некоторое множество языков. Язык A из этого семейства $\mathcal{L}$ называется *оптимальным* (для $\mathcal{L}$), если он не хуже любого другого языка из этого семейства, т. е. если

$$\forall B \in \mathcal{L}\ A \leqslant B.$$

Если оптимальный язык существует, то именно с его помощью следует измерять сложность. Сложность объекта относительно одного из оптимальных языков называется *алгоритмической энтропией* этого объекта. Энтропию и рассматривают как окончательную искомую меру сложности — в рамках заданного языково́го семейства.

Для некоторых важных языковы́х семейств имеет место теорема о существовании оптимального языка, а тем самым и энтропии. Эта теорема называется *теоремой Соломонова – Колмогорова.*

Для данного семейства может существовать много оптимальных языков и тем самым много энтропий. Однако, в силу определения оптимальности, любые две энтропии (взятые для одного и того же фиксированного языково́го семейства) различаются не более чем на аддитивную константу. Иными словами, если A и B суть два оптимальных для $\mathcal{L}$ языка, то существует такая константа c, что для всех y

$$|\operatorname{Comp}_A(y) - \operatorname{Comp}_B(y)| \ < \ c.$$

Замечание. Конечно, неприятно, что алгоритмическая энтропия, претендующая на роль «истинной сложности» определяется не однозначно, а всего лишь с точностью до аддитивной добавки, не превышающей константы. Однако попытки найти среди энтропий «наиболее оптимальную» пока что ни к чему хорошему не привели. С другой стороны, ведь, скажем, и длина, и вес также определены не однозначно, а с точностью до мультипликативной константы, зависящей от выбранной единицы измерения (10 см=100 мм; 2 кг=2000 г; и т. п.).

Из уважения к Колмогорову алгоритмическую энтропию обозначают обычно буквой K; иногда к этой букве K добавляют еще вторую букву, указывающую то языково́е семейство, к которому относится рассматриваемая энтропия. Если K' и K'' суть две энтропии, относящиеся к одному и тому же языково́му семейству, то, как отмечалось,

$$|\mathrm{K}' - \mathrm{K}''| \ < \ c.$$

Колмогоров не только сформулировал понятие алгоритмической энтропии объекта, но и применил его к исследованию случайных последовательностей. Наблюдение Колмогорова состояло в том, что в случайной последовательности энтропия начального отрезка растет достаточно быстро при неограниченном увеличении длины этого отрезка. (Ничего не поделаешь, случайная последовательность может начинаться с миллиона нулей — но и этот миллион ничто перед бесконечностью, и если взять все начальные отрезки в совокупности, то их энтропия неизбежно будет расти.)

Итак, в качестве описываемых объектов мы будем рассматривать двоичные цепочки — такие, как (I), (II), (III), (IV) и т. п. Поэтому отныне $Y = \Xi$.

Если язык содержит пару $\langle z, z \rangle$, то это означает попросту, что цепочка z описывает самоё себя. Язык D, состоящий в точности из пар такого вида, называется *диагональным* (термин из математики), или *автонимным* (термин из лингвистики). Для такого языка $\mathrm{Comp}_D(y) = |y|$. Ограничимся языковы́ми семействами, содержащими автонимный язык (таким будет, в частности, семейство монотонных языков, описанное ниже). Тогда для каждой энтропии K, соответствующей этому семейству, и подходящей константы c будет выполнено неравенство:

$$\mathrm{K}(y) < |y| + c.$$

Таким образом, если пренебречь аддитивной константой, максимально возможное значение для энтропии цепочки равно длине цепочки. Колмогоров предположил, что у случайной последовательности энтропия начального отрезка достигает этого максимума, то есть равна длине этого отрезка, — опять-таки, при пренебрежении аддитивной константой. В этом состояла основная идея Колмогорова относительно хаотичности.

Итак, фиксируем некоторое языково́е семейство и одну из соответствующих этому семейству энтропий K. Назовем последовательность

$$a_1, a_2, \ldots, a_n, \ldots$$

хаотической, если существует такая константа c, что

$$\mathrm{K}(a_1, a_2, \ldots, a_n) > n - c.$$

Очевидно, свойство хаотичности не зависит от выбранной энтропии, а только от выбранного семейства.

Оказалось, что при подходяще выбранном языковом семействе строгое понятие хаотичности хорошо отражает интуитивное представление о случайности.

Создавая теорию сложности объектов, Колмогоров придал соотношению „описание – объект“ алгоритмический характер. Следуя Колмогорову, ограничимся *перечислимыми* языками. Понятие *перечислимого* *множества* — это одно из основных понятий теории алгоритмов (да и математики в целом). Этому понятию можно дать такое наглядное объяснение. Представим себе безостановочно работающий принтер, последовательно печатающий слова. После напечатания слова принтер делает пробел, так что слова отделены одно от другого. Перед напечатанием очередного слова принтер берёт время на размышление. Это время может оказаться бесконечным, и тогда слово не печатается вовсе; в таком случае напечатанным будет лишь конечное множество слов — в частности, пустое множество, если принтер в самом начале своей работы задумался навсегда. Так вот, множество всех когда-либо напечатанных таким принтером слов окажется перечислимым — и всякое перечислимое множество может быть получено таким образом при подсоединении принтера к «идеальному компьютеру». Перечислимо и множество теорем любой формальной теории. Здесь нет места разъяснять, ни что такое идеальный компьютер, ни что такое формальная теория. Но определение понятия „перечислимое множество“ мы сейчас приведем.

Для наглядности начнем с термина «счетное множество». Термин этот имеет два варианта значения. В первом, более узком (и более распространенном) значении счетное множество — это такое множество, которое можно поставить во взаимно однозначное соответствие с натуральным рядом. Во втором, более широком значении счетное множество — это такое множество, которое можно поставить во взаимно однозначное соответствие с каким-либо начальным отрезком натурального ряда. Напомним, что *начальным отрезком натурального ряда* называется произвольное множество M натуральных чисел, содержащее вместе со всяким своим элементом и все меньшие числа. Таким образом, и весь натуральный ряд, и пустое множество являются начальными отрезками. Поэтому при втором понимании счетности все конечные множества, в том числе пустое множество, оказываются счетными. Именно это второе, более широкое понимание счетности удобно для наших целей; его мы и примем. А тогда счетному множеству можно дать и такое определение: *множество называется счетным, если оно либо пусто, либо может быть расположено в последовательность* (*т. е. является множеством членов какой-либо последовательности*). Например, конечное множество $\{a, b, c\}$ можно расположить в последовательность $a, b, c, c, c, c, \ldots$. Заменяя в приведённом определении счетного множества термин «последовательность» на термин «вычислимая последовательность», мы получаем определение перечислимого множества. Что касается определения термина «вычислимая последовательность», то оно сейчас будет дано.

Последовательность $w_1, w_2, \ldots, w_n, \ldots$ называется *вычислимой*, коль скоро существует алгоритм, вычисляющий ее n-й член w_n по его номеру n.

Поэтому понятие вычислимой последовательности называют *алгоритмическим аналогом* или *эффективным аналогом* понятия последовательности, а понятие перечислимого множества — алгоритмическим аналогом или эффективным аналогом понятия счетного множества. Перечислимые множества называются также *эффективно счетными*. Повторим определение: *множество называется перечислимым, или эффективно счетным, если оно либо пусто, либо может быть расположено в вычислимую последовательность* (*т. е. является множеством членов какой-либо вычислимой последовательности*).

Все рассматриваемые нами языки суть подмножества декартова произведения $\Xi \times \Xi$ и, следовательно, счетны. Идеология Колмогорова состояла в том, чтобы рассматривать только эффективно счетные (они же — перечислимые) языки.

Окончательно надлежащий выбор языково́го семейства произвел колмогоровский ученик Леонид Левин. Именно, в 1973 г. он ввел в рассмотрение семейство монотонных языков и изучил соответствующее этому семейству понятие хаотичности. Приведем необходимые определения.

Будем говорить, что двоичные слова u и v *согласованы* и писать $u \approx v$, если одно из этих слов есть начало другого.

Язык E называется *монотонным*, если он перечислим и выполнено условие:

$$[\langle x_1, y_1\rangle \in E \;\&\; \langle x_2, y_2\rangle \in E \;\&\; (x_1 \approx x_2)] \implies [y_1 \approx y_2].$$

Всякую последовательность, являющуюся хаотической для семейства монотонных языков, будем называть просто *хаотической*.

Алгоритмическая энтропия, соответствующая семейству монотонных языков, называется *монотонной энтропией* и обозначается KM. Условие хаотичности последовательности $\langle a_1, a_2, a_3, \ldots\rangle$ можно записать так:

$$\exists c\, \forall n\, \mathrm{KM}(a_1, a_2, \ldots, a_n) > n - c.$$

Класс всех хаотических последовательностей обозначается буквой **C**.

Считается, что строгое понятие хаотической последовательности может служить хорошим отражением интуитивного понятия случайной последовательности. К тому есть два основания.

Во-первых, для каждой отдельно взятой хаотической последовательности выполняются основные законы теории вероятностей.

Во-вторых, класс хаотических последовательностей совпадает с другим «претендентом» на роль строгого аналога расплывчатого класса случайных последовательностей — а именно, с классом **T** всех типических последовательностей (о нём будет рассказано далее):

$$\mathbf{C} = \mathbf{T}.$$

Поэтому хаотические последовательности можно было бы называть *типическо-хаотическими* или *хаотическо-типическими*, а сам класс обозначать **TC** или **CT**. Этот класс у́же класса **S** всех последовательностей, стохастических по Колмогорову (последний, как уже́ отмечалось, слишком широк):

$$\mathbf{TC} \subset \mathbf{S}, \quad \mathbf{TC} \neq \mathbf{S}.$$

Лицо Третье: Типичность

Сказать про какой-либо объект, что он типичен — это значит сказать, что он принадлежит к любому подавляющему большинству. Например, типичный человек, встреченный на московской улице, имеет рост менее двух метров (т. е. принадлежит к подавляющему большинству людей, имеющих рост менее двух метров), возраст более трех лет (т. е. принадлежит к подавляющему большинству людей, имеющих возраст более трех лет) и т. д. Конечно, любое из большинств, о которых идет речь, должно быть разумным: ведь какой объект ни возьми, он заведомо не принадлежит к подавляющему большинству, образованному всеми остальными, отличными от данного, объектами.

Мы исходим из интуитивного представления о том, что случайный объект должен обладать свойством типичности. Наша цель — дать строгое математическое определение этого свойства для последовательностей нулей и единиц при равномерном распределении вероятностей на пространстве Ω таких последовательностей. Последовательности, удовлетворяющие этому строгому определению, мы будем называть *типическими*, оставляя слово *типичный* для употребления на интуитивном уровне. В силу сказанного выше, нам надо определить сперва, что есть подавляющее большинство последовательностей, а затем — какие большинства следует считать разумными. А тогда класс типических последовательностей автоматически определится как пересечение всех разумных большинств.

Оставляя термин «подавляющее большинство» для интуитивного употребления, мы будем говорить о *больших* множествах последовательностей. Дополнение к большому множеству до всего Ω будем называть *малым*. Очевидно, достаточно определить, что такое малое множество, — а тогда большие множества определятся как дополнения к малым.

Итак, пусть дано какое-то множество Q последовательностей, $Q \subset \Omega$. Мы хотим определить, когда его следует считать *малым*. В терминах теории вероятностей мы сказали бы, что Q малое, коль скоро вероятность попадания в Q равна нулю. В терминах теории меры мы сказали бы, что Q малое, коль скоро оно имеет меру нуль. Мы, однако, постараемся определить, что такое малое множество в более простых терминах.

Множество Q называется *малым*, если его можно покрыть шарами, сумма объемов которых сколь угодно мала. Определению малого множества можно дать такую формулировку: множество Q называется *малым*, если для каждого натурального числа m найдется такая последовательность двоичных слов $\langle x(1), x(2), \ldots, x(n), \ldots\rangle$, что

$$Q \subset \bigcup_n \Omega_{x(n)};$$

$$\sum_n \mathbf{v}(\Omega_{x(n)}) = \sum_n 2^{-|x(n)|} < \frac{1}{m}.$$

Очевидно, каждая отдельная последовательность из Ω образует малое множество, а потому понятие большого множества не может претендовать на роль уточнения понятия *разумного* большинства. Пересечение всех больших множеств пусто.

«Разумность» вводится путем следующей корректировки определения.

Во-первых, потребуем, чтобы упоминаемая в определении последовательность двоичных слов $\langle x(1), x(2), \ldots, x(n), \ldots\rangle$ была вычислимой. Другими словами, мы требуем, чтобы существовал алгоритм, вычисляющий ее n-й член x_n по его номеру n.

Во-вторых, потребуем, чтобы такая вычислимая последовательность не просто существовала для каждого m, но строилась бы по этому m *эффективно*. Слово «эффективно» означает „с помощью алгоритма“. Здесь необходимы разъяснения. Дело в том, что алгоритм, получающий на вход m и выдающий на выходе требуемую последовательность, невозможен просто потому, что последовательность — это бесконечный объект, а алгоритмы оперируют лишь с конечными объектами. Однако в нашем случае, поскольку последовательность вычислима, то у нее есть алгоритм, который ее вычисляет — даже очень много таких алгоритмов. Алгоритмы (в другой системе терминов — программы алгоритмов) являются конечными объектами и потому вполне осмысленно говорить об алгоритме, который по числу m, поступившему на его вход, выдает на выходе один из алгоритмов (в другой системе терминов — одну из программ), вычисляющих последовательность $\langle x(1), x(2), \ldots, x(n), \ldots\rangle$.

Внеся в определение малого множества эти два добавления, мы получаем определение *эффективно малого* множества — а тем самым и определение *эффективно большого* множества. Пересечение всех эффективно больших множеств оказывается непустым. Более того, оно само является эффективно большим. Это наименьшее среди эффективно больших множеств и есть искомое множество $\mathbf{T}$ всех *типических* последовательностей.

Типические последовательности называют также *случайными по Мартин-Лёфу* — по имени ученика Колмогорова, замечательного шведского математика Пера Мартин-Лёфа, который в 1966 г. сформулировал только что изложенное определение типичности в качестве строгого уточнения понятия случайности.

Как уже говорилось, класс **T** всех типических последовательностей совпадает с классом **C** всех хаотических последовательностей:

$$\mathbf{T} = \mathbf{C}.$$

Поэтому, как уже говорилось, типические последовательности можно именовать также *хаотическо-типическими* или *типическо-хаотическими*, а сам класс всех таких последовательностей обозначать **CT** или **TC**.

Как мы уже знаем,

$$\mathbf{CT} \subset \mathbf{S}, \quad \mathbf{CT} \neq \mathbf{S}.$$

Лицо Четвертое: Непредсказуемость

Вот начало «Двенадцати стульев» Ильфа и Петрова: «В уездном городе N было так много парикмахерских заведений и бюро похоронных процессий, что казалось, жители города рождаются лишь затем, чтобы побриться, остричься, освежить голову вежеталем и умереть». Суждение остается верным, если заменить N на Москву, парикмахерские на залы игровых автоматов, похоронные бюро на казино, а цель рождения на «играть». Этот печальный факт, однако, позволяет наглядно объяснить то свойство случайных последовательностей, которое мы называем непредсказуемостью.

Интуитивно ясно, что всякая случайная последовательность является *непредсказуемой* в том смысле, что в каком бы порядке мы ни выбирали ее члены, знание значений уже выбранных членов не позволяет предсказать значение того следующего члена, который мы намереваемся выбрать. Таким образом, Казино, обладающее такой последовательностью и предлагающее Игроку угадывать ее члены и делать при этом денежные ставки, не разорится; говоря более точно, Казино уверено, что никакой Игрок не может обладать такой стратегией игры, которая приведет к разорению Казино, каким бы капиталом оно ни обладало.

Непредсказуемость какой-либо последовательности, таким образом, определяется в терминах игры, которую Игрок ведет против обладающего этой последовательностью Казино, или, для краткости — *против данной последовательности.*

Итак, представим себе что Игрок приходит в Казино. Каждая из сторон — и Казино, и Игрок — обладает своим начальным капиталом. Казино располагает некоторой фиксированной, но неизвестной Игроку последовательностью нулей и единиц, и предлагает Игроку предсказывать ее члены — необязательно в монотонном порядке их следования и даже необязательно все ее члены.

Для наглядности представим себе, что члены последовательности написаны на картах, которые лежат рубашками вверх, так что Игрок не видит, что там написано. Последовательность предстает перед Игроком в виде бесконечного ряда таких карт. Игра состоит в том, что Игрок на каждом своем ходу указывает ту карту, которая должна быть открыта, одновременно предсказывая значение, которое обнаружится на этой карте, и объявляя размер денежной ставки. Если предсказание окажется правильным, Казино выплачивает Игроку сумму ставки, если неправильным — Игрок выплачивает эту сумму Казино. Считается, что Игрок *выиграл*, если он сумел разорить Казино. Разумеется, если Игроку открыт неограниченный кредит, он всегда может разорить Казино, удваивая ставки. Но Игра идет на наличные, так что величина ставки ограничена текущим капиталом Игрока.

Последовательность называется *предсказуемой*, если существует выигрывающий алгоритм игры. *Выигрывающим* мы называем алгоритм со следующим свойством: каким бы начальным капиталом ни обладало Казино, оно рано или поздно будет разорено, если Игрок применит этот алгоритм. Последовательность называется *непредсказуемой*, если она не является предсказуемой.

На математическом языке ситуация описывается так.

Рассматривается бесконечная последовательность нулей и единиц

$$\mathbf{a} = \langle a_1, a_2, a_3, \ldots \rangle.$$

При каждом ходе Игрока возникает тройка чисел

$$\langle n, i, v \rangle,$$

где

$$n \in \mathbb{N}, \quad i \in \{0,\ 1\}, \quad v \in \mathbb{Q}, \quad v \geqslant 0.$$

Содержательно: натуральное число n есть номер того члена последовательности, на который делается ставка; i есть предсказываемое значение этого члена; неотрицательное рациональное число v есть размер ставки. Ходы делаются друг за другом, начиная с первого; тройка, возникающая на k-м ходу, обозначается $\langle n(k), i(k), v(k) \rangle$. Более математически грамотно было бы сказать, что каждый ход *есть* тройка чисел, и что ходы не делаются, а *предъявляются*.

Капитал Игрока перед k-м ходом обозначается $V(k-1)$. Без ограничения общности можно считать, что начальный капитал Игрока равен единице: $V(0) = 1$.

После каждого хода капитал Игрока меняется (если только ставка не была нулевой). А именно:

- ▷ если $i(k) = a_{n(k)}$ (Игрок угадал), то $V(k) = V(k-1) + v(k)$;
- ▷ если $i(k) \neq a_{n(k)}$ (Игрок не угадал), то $V(k) = V(k-1) - v(k)$.

И еще два прибавления к сказанному.

Во-первых, ходы бывают *корректные* и *некорректные*, и чтобы игра продолжалась, необходимо, чтобы ход был корректным. А именно, k-й ход считается *корректным*, если выполнены оба нижеследующие требования:

1) номер открываемой карты *корректен*; это значит, что подлежащая открытию карта не была уже открыта ранее, т. е. $n(k)$ не совпадает ни с одним из чисел $n(1), \ldots, n(k-1)$;

2) делаемая Игроком ставка *корректна*; это значит, что она меньше его текущего капитала, т. е. $v(k) < V(k-1)$.

Если же хотя бы одно из этих требований не выполнено, ход считается *некорректным*.

Основное правило остановки. Если Игрок совершает некорректный ход, игра останавливается. При этом Игрок остается при имеющемся у него к данному моменту капитале — и тем самым заведомо не выигрывает.

Во-вторых, не исключается возможность того, что Игрок вообще не делает очередного хода (даже самого первого хода!), и в этом случае Игрок также остается при имеющемся у него к данному моменту капитале и, как и в случае некорректного хода, заведомо не выигрывает. Однако в этом случае — в отличие от случая некорректного хода — мы избегаем выражения «игра останавливается». Дело в том, что ситуацию неделания хода можно наглядно представить себе следующим образом. Перед каждым своим ходом Игрок решает, какой ход ему следует сделать. Решение требует обдумывания, и Игрок берет время на обдумывание. Время это ничем не ограничено, и процесс размышления может затянуться до бесконечности. В течение времени обдумывания хода капитал Игрока не меняется. Поэтому если Игрок ни за какое конечное время не принимает решения о своем ходе, его капитал застывает. В этом случае выигрыша Игрока произойти не может. Однако в этом случае (в отличие от случая некорректного хода) не наступает момента остановки игры — никогда не поступает сигнала об остановке. Таким образом, у игры есть три возможных сценария развития: 1) Игрок делает бесконечное число ходов; 2) Игрок делает лишь конечное число ходов, и причиной этого служит то, что был сделан некорректный ход; 3) Игрок делает лишь конечное число ходов, и причиной этого служит то, что на каком-то этапе игры Игрок не в состоянии прийти к решению об очередном ходе. Разумеется, сказанное носит иллюстративный характер, и математическое описание не включает в себя ссылку на такие понятия, как „решать“, „обдумывать“ и т. п.

По определению, Игрок *выигрывает* (при игре против $\mathbf{a}$), если

$$\sup_k V(k) = +\infty,$$

т. е. если

$$\forall W\ \exists k\ V(k) > W.$$

Содержательно это означает, что Игрок разоряет Казино, каким бы исходным капиталом W Казино ни обладало. Очевидно, что Игрок в состоянии выиграть лишь при условии, что всякий раз, когда ему предстоит делать ход, он его делает и этот ход оказывается корректным.

Как выглядит игра, мы описали. Перейдем теперь к понятию системы игры, или *стратегии*. Смысл стратегии в том, чтобы избавить Игрока от необходимости самостоятельно принимать решения: стратегия берет эту функцию на себя. Стратегия есть правило, для каждого хода указывающее Игроку, какой на этом ходу он должен сделать ход (т. е. какую тройку предъявить). Разумеется, стратегия выдает такое указание лишь в том случае, если ход должен быть сделан. Выше уже отмечалась возможность того, что никакого хода не делается; в этом случае, естественно, стратегия не выдает никакого указания. При указании хода стратегия опирается на всю предшествующую историю игры. История же игры состоит из всех уже сделанных к рассматриваемому моменту ходов и из всех ставших уже известными членов последовательности. Мы лишь потому не включаем в историю игры информацию о капитале Игрока на каждый момент, что эта информация легко вычисляется из только что перечисленных сведений.

Таким образом, историю игры перед k-м ходом можно записать в виде таблицы

$$\begin{array}{cccc} n(1) & n(2) & \ldots & n(k-1) \\ i(1) & i(2) & \ldots & i(k-1) \\ v(1) & v(2) & \ldots & v(k-1) \\ a_{n(1)} & a_{n(2)} & \ldots & a_{n(k-1)} \end{array}$$

(Если $k = 0$, таблица пуста.)

На математическом языке стратегия есть функция, которая каждой подобной таблице (в том числе пустой) либо ничего не ставит в соответствие, либо ставит в соответствие некоторый ход, т. е. тройку $\langle n, i, v\rangle$. Под «каждой подобной таблицей» мы понимаем отнюдь не только такую таблицу, которая отражает реальное течение игры, а произвольную таблицу, в которой в первой строке стоят положительные целые числа, в третьей — неотрицательные рациональные числа, а во второй и четвертой — нули и единицы.

Если таблица реально встретилась в процессе игры (в качестве истории игры на каком-то этапе) и если задана стратегия, то первые три строки таблицы однозначно восстанавливаются по ее четвертой строке. В самом деле, применение стратегии к пустой таблице дает нам первый ход $\langle n(1), i(1), v(1)\rangle$. Тем самым — поскольку четвертая строка предполагается известной — мы получаем историю игры перед вторым ходом в виде таблицы

$$\begin{array}{c} n(1) \\ i(1) \\ v(1) \\ a_{n(1)} \end{array}$$

Теперь к этой таблице снова применяем стратегию, получаем второй ход $\langle n(2), i(2), v(2)\rangle$ и таблицу

$$\begin{array}{cc} n(1) & n(2) \\ i(1) & i(2) \\ v(1) & v(2) \\ a_{n(1)} & a_{n(2)} \end{array}$$

И так далее.

Сказанное дает нам право при определении стратегии брать в качестве аргумента не всю таблицу в целом, а лишь ее четвертую строку. Заметим, что в этой четвертой строке стоит двоичное слово, т. е. элемент множества Ξ. Стратегия должна, имея этот элемент, или не выдавать ничего, или выдавать ход, который есть тройка, т. е. элемент декартова произведения $\mathbb{N}\times\{0,1\}\times\mathbb{Q}^{\geqslant 0}$. Здесь символом $\mathbb{Q}^{\geqslant 0}$ обозначено множество всех неотрицательных рациональных чисел.

Мы приходим к окончательному определению понятия стратегии: *стратегия* есть отображение некоторого подмножества множества Ξ всех двоичных слов в декартово произведение $\mathbb{N} \times \{0,1\} \times \mathbb{Q}^{\geqslant 0}$:

$$\Xi \longrightarrow \mathbb{N} \times \{0,1\} \times \mathbb{Q}^{\geqslant 0}.$$

Для наших целей особый интерес представляет случай, когда указанное отображение задается каким-то алгоритмом. Поясним, что это значит. Пусть $\mathbf{A}$ — алгоритм, на вход которого могут подаваться элементы из множества X, а на выходе получаются элементы из множества Y. В множестве X выделяется *область результативности* алгоритма $\mathbf{A}$, состоящая из тех и только тех элементов, в применении к которым $\mathbf{A}$ дает результат. Алгоритм $\mathbf{A}$ *задает* следующее отображение: область определения этого отображения совпадает с областью результативности алгоритма, и для каждого элемента этой области значение отображения на этом элементе совпадает с тем результатом, который получается при применении алгоритма к этому элементу.

Если стратегия задана каким-то алгоритмом, она называется *вычислимой*. (Если подать на вход задающего стратегию алгоритма такую историю игры, для которой очередной ход не определен, то алгоритм работает на этом входе бесконечно долго, не приходя ни к какому результату, но и не выдавая сообщение об отсутствии такового. Эта бесконечная работа алгоритма как раз и происходит за то взятое Игроком на обдумывание бесконечное время, о котором говорилось выше.)

Стратегия называется *выигрывающей для последовательности* **a**, если Игрок, применяющий эту стратегию в игре против **a**, выигрывает.

Последовательность называется *предсказуемой*, если для нее существует выигрывающая вычислимая стратегия, и *непредсказуемой* в противном случае.

Класс всех непредсказуемых последовательностей обозначается буквой **U**.

Известно, что всякая непредсказуемая последовательность является стохастической по Колмогорову (принадлежит классу **S**) и что всякая типическо-хаотическая последовательность (последовательность класса **CT**) непредсказуема:

$$\mathbf{CT} \subset \mathbf{U} \subset \mathbf{S}.$$

Известно также, что класс стохастических по Колмогорову последовательностей существенно шире класса непредсказуемых:

$$\mathbf{S} \neq \mathbf{U}.$$

Открытым остается вопрос о совпадении классов хаотических (он же класс типических) и непредсказуемых последовательностей:

$$\mathbf{CT} = \mathbf{U}??$$

Эта важная проблема ждет своего решения.

О безостановочных стратегиях. Если игра никогда не останавливается, она называется *безостановочной*. Стратегия называется *безостановочной*, если какова бы ни была последовательность, применение против нее этой стратегии приводит к безостановочной игре. В определении предсказуемости можно ограничиться безостановочными стратегиями и дать такую равносильную формулировку: последовательность называется *предсказуемой*, если для нее существует выигрывающая вычислимая безостановочная стратегия. Чтобы убедиться в равносильности, достаточно объяснить, как можно алгоритм **A**, задающий выигрывающую стратегию, переделать в алгоритм **B**, задающий выигрывающую безостановочную стратегию. Такая переделка осуществляется весьма просто. Сперва, по поступившему на вход алгоритма **A** двоичному слову восстанавливается история игры, что позволяет знать как номера всех открытых ранее карт, так и величину текущего капитала Игрока. Затем всякий ход, предписываемый исходным алгоритмом **A**, проверяется на корректность, и если он оказывается некорректным, то новый алгоритм **B** никакого хода не выдает, а объявляется не определенным на указанном двоичном слове.

Обобщение на вычислимые распределения вероятностей

Предварительные замечания

До сих пор, для простоты и большей наглядности, мы ограничивались ситуацией, когда на пространстве Ω всех двоичных последовательностей было задано *равномерное распределение вероятностей*. Все основные идеи были видны на примере этой простейшей ситуации. Для полноты картины мы намерены рассмотреть теперь общую ситуацию, когда на пространстве Ω задано произвольное *вычислимое распределение вероятностей*. Что это такое, будет разъяснено позже. А сейчас мы хотим сказать несколько слов, направленных на то, чтобы сделать изложение доступным и такому Читателю, который не знаком с общим понятием *распределения вероятностей*, или *вероятностной меры*.

Говорят, что на множестве M задана *мера* μ, коль скоро, во-первых, выделен некоторый класс подмножеств множества M, называемых *измеримыми*, и, во-вторых, каждому измеримому подмножеству A отнесено неотрицательное число $\mu(A)$, называемое мерой этого подмножества. При этом требуется, чтобы выполнялись некоторые аксиомы, приводить которые мы здесь не будем; отметим только, что следствием этих аксиом является такой факт: объединение непересекающихся измеримых множеств измеримо и его мера равна сумме мер этих множеств. Если $\mu(M) = 1$, мера называется *вероятностной*, или *распределением вероятностей*. В этом случае $\mu(A)$ содержательно трактуется как вероятность того, что случайно выбранный элемент множества M попадает в A.

Каждая мера на Ω характеризуется мерами шаров. Так, для равномерного распределения (и только для него) $\forall x \in \Xi\ \mu(\Omega_x) = 2^{-|x|}$.

Равномерное распределение вероятностей на пространстве Ω отвечает тому сценарию, когда нули и единицы возникают в последовательности с равными вероятностями. Ближайшим обобщением равномерного распределения является *распределение Бернулли*, или *бернуллиевское распределение* (называемое также *биномиальным*), при котором нули и единицы возникают с вероятностью p для единицы и вероятностью $1-p$ для нуля; это число p условимся называть *параметром* бернуллиевского распределения. Если параметр равен одной второй, получаем равномерное распределение. Говоря формально, распределение Бернулли с параметром p задается формулой $\mu(\Omega_x) = p^k(1-p)^{|x|-k}$, где k — количество единиц в слове x.

Следующим обобщением служит класс распределений, которые мы будем называть *квазибернуллиевскими*. Пусть дана последовательность действительных чисел $\mathbf{p} = \langle p(1), p(2), \ldots, p(k), \ldots \rangle$, $0 \leqslant p(k) \leqslant 1$. Про распределение μ будем говорить, что оно *квазибернуллиевское с параметром* $\mathbf{p}$, коль скоро для всякого двоичного слова $x = \langle x_1, \ldots, x_n \rangle$ имеет место равенство $\mu(\Omega_x) = \prod_{i=1}^{i=n} r_i$, где $r_i = p(i)$ при $x_i = 1$ и $r_i = 1 - p(i)$ при $x_i = 0$. На содержательном уровне это означает, что единицы и нули появляются с переменной вероятностью, зависящей только от номера члена. Для ясности: бернуллиевское распределение с параметром p является квазибернуллиевским с параметром $\langle p, p, \ldots, p, \ldots \rangle$.

В этой главке мы укажем как определения стохастичности, хаотичности, типичности и непредсказуемости должны быть расширены для общей ситуации *вычислимого распределения вероятностей* (что это такое, будет разъяснено ниже). Объявим наперёд, что при этих расширенных определениях для произвольного вычислимого распределения μ сохраняются те же соотношения, которые были выписаны выше для случая равномерного распределения:

$\mathbf{C}(\mu) = \mathbf{T}(\mu) \subset \mathbf{U}(\mu) \subset \mathbf{S}(\mu)$;

$\mathbf{S}(\mu) \neq \mathbf{U}(\mu)$ (при условии, что мера любого шара положительна).

Здесь через $\mathbf{C}(\mu)$, $\mathbf{T}(\mu)$, $\mathbf{U}(\mu)$, $\mathbf{S}(\mu)$ соответственно обозначены те классы последовательностей, хаотических, типических, непредсказуемых, стохастических по Колмогорову относительно распределения μ, кои будут определены ниже. (В наших прежних обозначениях $\mathbf{C} = \mathbf{C}(\eta)$, $\mathbf{T} = \mathbf{T}(\eta)$, $\mathbf{U} = \mathbf{U}(\eta)$ и $\mathbf{S} = \mathbf{S}(\eta)$, где η есть равномерное распределение.)

Эта главка — для любителей обобщений, и мы призываем уважаемого Читателя подумать, стоит ли ее читать. Во-первых, она несколько труднее предыдущих главок. Во-вторых, сама задача о формулировке строгого определения случайной последовательности делается тем менее ясной, чем более обширным становится класс рассматриваемых распределений. Ведь само представление о случайности индивидуальной последовательности имеет сколько-нибудь ясный интуитивный смысл лишь для простейших распределений вероятностей — таких, как равномерное и его ближайшие обобщения.

Вычислимые меры и вычислимые распределения

Хотелось бы называть меру μ на пространстве Ω вычислимой, если существует такой алгоритм, который по каждому двоичному слову x дает меру $\mu(\Omega_x)$ шара Ω_x. Однако алгоритмы не могут иметь дело с действительными числами, а только с числами целыми, рациональными и т. п. — строго говоря, даже не с числами, а с их именами в виде слов в каком-либо фиксированном конечном алфавите. (*Словом в* данном алфавите называется любая конечная цепочка, составленная из букв этого алфавита.) Дать же имена всем действительным числам невозможно, поскольку для каждого конечного алфавита все слова в нём образуют лишь счетное множество. Поэтому можно требовать лишь наличие алгоритма, дающего не самоё меру шара, а сколь угодно точное рациональное приближение к ней.

Окончательное определение таково. Меру μ называют *вычислимой*, коль скоро существует алгоритм, который для каждого двоичного слова x и каждого положительного рационального числа ε выдает такое рациональное число, которое отличается от меры $\mu(\Omega_x)$ шара Ω_x не более, чем на ε.

Иногда в определение вычислимости включают еще и требование о существовании такого алгоритма, который для каждого двоичного слова x определяет, имеет ли место равенство $\mu(\Omega_x) = 0$. Вычислимые меры, удовлетворяющие этому дополнительному требованию (которое мы не включаем в определение вычислимости), будем называть *сильно вычислимыми*. Заметим. что из вычислимости меры не вытекает, как могло бы показаться, ее сильная вычислимость.

Важным подклассом класса сильно вычислимых мер является класс вычислимо-рациональных мер. Условимся называть меру *вычислимо-рациональной*, если мера каждого шара есть рациональное число и существует алгоритм, вычисляющий эту меру по заданному шару, т. е., на более точном языке, алгоритм, который для каждого двоичного слова x выдает дробь, выражающую меру $\mu(\Omega_x)$ шара Ω_x. Заметим для ясности, что если мера вычислима, а ее значения на шарах рациональны, то отсюда еще не следует, что она вычислимо-рациональна: умение находить для данного рационального числа сколь угодно точные рациональные приближения к нему еще не означает умения выписать само это число в виде отношения двух целых чисел!

Теперь нет нужды отдельно определять, что такое вычислимое распределение вероятностей, — это просто-напросто вычислимая вероятностная мера. Именно на вычислимые распределения естественно обобщаются многие конструкции и факты, изложенные нами для равномерного распределения. В частности, для произвольного вычислимого распределения верны как *теорема Мартин-Лёфа*, утверждающая, что пересечение всех эффективно больших множеств само является эффективно большим, так и *теорема Левина*, утверждающая, что типичность равносильна хаотичности, определенной на основе монотонной энтропии.

Стохастичность

Для какой-либо последовательности $\mathbf{e}$ ее k-й член обозначаем e_k или, дабы избежать многоэтажных индексов, $e(n)$.

В случае равномерного распределения стохастичность последовательности понималась как *глобальная устойчивость частот*, т. е. как устойчивость частот в *каждой* из допустимых подпоследовательностей. Допустимые же подпоследовательности возникали путем применения допустимых по Колмогорову правил отбора. В ситуации произвольной вероятностной меры общая схема сохраняется, только устойчивость частот заменяется на некоторое более общее свойство, а именно — на так называемый *закон больших чисел*.

Определение стохастичности для бернуллиевского распределения с параметром p очевидно: надо потребовать, чтобы в каждой допустимой подпоследовательности соблюдалась устойчивость частот с одной и той же предельной частотой p. Иными словами, доля единиц в начальном отрезке всякой допустимой подпоследовательности должна, при неограниченном удлинении отрезка, стремится к p. Случаи, когда $p = 0$ или $p = 1$ являются вырожденными; для первого из них не считается стохастической последовательность, в которой встречается хотя бы одна единица, для второго — последовательность, в которой встречается хотя бы один ноль.

Фон Мизес понимал вероятность как предельную частоту. Поэтому его идеология не простирается за пределы бернуллиевских распределений. Мостом к общему случаю произвольного распределения служат распределения квазибернуллиевские. Ими и займемся.

Ясно, что для квазибернуллиевского распределения предельной частоты в подпоследовательности может и не быть, а если она и есть, то, вообще говоря, в каждой подпоследовательности — своя. Приходится поэтому говорить о стохастичности относительно данного правила отбора. Пусть $\mathbf{p}$ — параметр квазибернуллиевского распределения. Прежде всего заявим, что не считается стохастической последовательность, в которой значение хотя бы одного ее члена имеет нулевую вероятность. Или, говоря на формальном языке: последовательность $\mathbf{a}$ не считается стохастической, если существует такое k, что либо $a(k) = 0$ и $p(k) = 1$, либо $a(k) = 1$ и $p(k) = 0$. Для остальных последовательностей определение таково. Последовательность $\mathbf{a}$ будем называть *стохастической относительно правила отбора* Θ, коль скоро ее обобщенная подпоследовательность

$$\mathbf{b} = \langle a(m_1), a(m_2), \dots, a(m_k), \dots \rangle,$$

полученная по этому правилу, удовлетворяет требованию закона больших чисел:

$$\frac{a(m_1)+\dots+a(m_k)}{k} - \frac{p(m_1)+\dots+p(m_k)}{k} \to 0$$

при $k \to \infty$. Далее, последовательность $\mathbf{a}$ называется *стохастической* (*по Колмогорову*), если для любого колмогоровского правила, приводящего к бесконечной обобщенной подпоследовательности, она является стохастической относительно этого правила. (Замечание. Слово «бесконечной» в предыдущей фразе избыточно, так как всякая обобщенная подпоследовательность является последовательностью, а следовательно, бесконечна по определению. Тем не менее мы будем иногда использовать это формально избыточное упоминание о бесконечности, дабы подчеркнуть, что имеется в виду именно последовательность, а не кортеж.)

Чтобы объяснить, как понятие стохастичности по Колмогорову обобщается на более широкий класс вероятностных распределений, введем несколько обозначений.

Пусть $n(1), n(2), \dots, n(k)$ суть натуральные числа и пусть $i(1), i(2), \dots, i(k) \in \{0,1\}$. Через $A^{n(1),\dots,n(k)}_{i(1),\dots,i(k)}$ обозначим множество всех таких последовательностей $\mathbf{a} \in \Omega$, у которых

$$a_{n(1)} = i(1),\ a_{n(2)} = i(2),\ \dots,\ a_{n(k)} = i(k). \qquad (*)$$

Дробь

$$\mu(A^{n(1),\ \dots,\ n(k),\ m}_{i(1),\ \dots,\ i(k),\ 1}) / \mu(A^{n(1),\ \dots,\ n(k)}_{i(1),\ \dots,\ i(k)})$$

обозначим символом $\mu\left(\begin{array}{c|ccc} m & n(1), & \dots, & n(k) \\ 1 & i(1), & \dots, & i(k) \end{array}\right)$; это есть условная вероятность того, что m-й член последовательности $\mathbf{a}$ будет равен 1 при условии $(*)$; эта величина не определена, коль скоро знаменатель дроби обращается в нуль.

Рассмотрим два произвольных, но фиксированных объекта: последовательность $\mathbf{a} \in \Omega$ и допустимое по Колмогорову правило отбора Θ. Наша цель — придать смысл высказыванию «последовательность $\mathbf{a}$ стохастична относительно правила Θ». Процесс, посредством которого правило Θ отбирает из последовательности $\mathbf{a}$ члены обобщенной подпоследовательности $\mathbf{b}$, состоит из двух этапов. На первом этапе строится вспомогательная обобщенная подпоследовательность $\mathbf{c}$, из некоторых членов которой на втором этапе и строится $\mathbf{b}$. Более подробно,

$$\mathbf{c} = \langle a_{n(1)}, a_{n(2)}, \dots, a_{n(k)}, \dots \rangle,$$

где номер $n(k)$ вычисляется алгоритмически по кортежу $\langle a_{n(1)}, a_{n(2)}, \dots, a_{n(k-1)} \rangle$. Имея на входе этот же кортеж, правило Θ решает, включать или нет член $a_{n(k)}$ в окончательную последовательность $\mathbf{b}$. Таким образом,

$$\mathbf{b} = \langle a(n(k_1)), a(n(k_2)), \dots, a(n(k_j)), \dots \rangle.$$

На каждом из обоих этапов может наступить момент, когда правило Θ не даст никакого результата, т. е. не выдаст номера на первом этапе или не примет решения на втором этапе. Если такое произойдет, последовательность **b** окажется конечной — т. е., говоря более строго, окажется не последовательностью, а кортежем. В этом случае никакого требования к **b** не выдвигается и, говоря формально, **a** признаётся стохастической относительно Θ. Если же **b** оказалось бесконечной, то для признания последовательности **a** стохастической относительно Θ требуется выполнение нижеследующего свойства последовательности **a**.

Обозначим через r_j величину

$$\mu\left(\begin{array}{c|ccc} n(k_j) & n(1), & n(2)\ldots, & n(k_j-1) \\ 1 & a(n(1)), & a(n(2))\ldots, & a(n(k_j-1)) \end{array}\right).$$

Рассмотрим разность

$$\delta_j \;=\; \frac{r_1+r_2+\cdots+r_j}{j} - \frac{a(n(k_1))+a(n(k_2))+\cdots+a(n(k_j))}{j}\,.$$

Величина δ_j не определена, если не определена хотя бы одна из величин $r_1,\ldots,r_j$. Скажем, что **b** *подчиняется закону больших чисел*, если все величины δ_j определены и $\delta_j \to 0$ при $j\to\infty$. Последовательность **a** назовем *стохастической относительно правила* Θ, если обобщенная подпоследовательность **b**, полученная из **a** согласно правилу Θ, подчиняется закону больших чисел.

Наконец, последовательность **a** называется *стохастической* (*по Колмогорову*), если для любого допустимого по Колмогорову правила отбора она является стохастической относительно этого правила — при условии, что это правило строит бесконечную обобщенную подпоследовательность (а не конечный кортеж).

Заметим, что в самом определении вычислимость меры не используется. Однако без предположения о вычислимости едва ли возможно сравнивать частотоустойчивость (она же стохастичность) с другими алгоритмическими лицами случайности.

Хаотичность

Определение последовательности, хаотической относительно меры μ, таково. Последовательность $\mathbf{a} = \langle a_1, a_2, a_3, \ldots\rangle$ называется *хаотической относительно* μ, если существует такая (зависящая от меры μ) константа c, что для всех n выполняется неравенство

$$\mathrm{KM}(a_1,a_2,\ldots,a_n) > -\log\mu(\Omega_{a_1,a_2,\ldots,a_n}) - c,$$

где, как всегда, знак логарифма без нижнего индекса означает логарифм по основанию два.

Мотивировка этого определения, имеющего содержательный смысл для вычислимых вероятностных мер, такая же, как и в случае равномерного распределения вероятностей. Дело в том, что для всякой вычислимой вероятностной меры выполняется соотношение:

$$\exists c\ \forall x \in \Xi \quad \mathrm{KM}(x) < -\log \mu(\Omega_x) + c.$$

Типичность

Понятие типичности обобщается на произвольную меру достаточно очевидным образом — надо просто в том определении типической последовательности, которое давалось для случая равномерного распределения, заменить объём шара на его меру.

Сперва определяется понятие эффективно малого множества. Множество $Q \subset \Omega$ называется *эффективно малым относительно меры* μ, коль скоро существует алгоритм $\mathbf{A}$, удовлетворяющий нижеследующему требованию. При поступлении на вход алгоритма $\mathbf{A}$ натурального числа m на его выходе возникает алгоритм построения такой последовательности двоичных слов $\langle x(1), x(2), \ldots, x(n), \ldots \rangle$ (т. е. алгоритм вычисления такой функции $n \mapsto x(n)$), что

$$Q \subset \bigcup_n \Omega_{x(n)};$$

$$\sum_n \mu(\Omega_{x(n)}) < \frac{1}{m}\,.$$

Далее, *эффективно большое относительно меры* μ множество определяется как дополнение (до Ω) к эффективно малому.

Для вычислимых мер имеет место *теорема Мартин-Лёфа*: объединение всех эффективно малых множеств является эффективно малым, а пересечение всех эффективно больших — эффективно большим. Существующее согласно этой теореме наименьшее эффективно большое множество называют *конструктивным носителем меры* μ. В случае, если μ — вычислимое распределение вероятностей, последовательности, принадлежащие конструктивному носителю распределения μ, и называют *типическими* относительно этого распределения. Таким образом, $\mathbf{T}(\mu)$ есть не что иное как конструктивный носитель распределения μ.

Непредсказуемость

Здесь мы укажем, что́, при переходе к произвольным распределениям, нужно добавить к определениям, сформулированным для равномерного случая. Таких добавлений будет два: некий мультипликативный коэффициент (для равномерного распределения он равен единице и потому не нужен) и дополнительное правило остановки (в случае равномерного распределения оно потому не нужно, что не возникает такой ситуации, при которой оно могло бы быть применено).

Распределение вероятностей влияет на правило изменения капитала Игрока после очередного хода. Если Игрок не угадал, его капитал уменьшается на сумму сделанной им ставки — так же, как и в случае равномерного распределения. Но если Игрок угадал, то прирост его капитала равняется ставке, умноженной на некоторый коэффициент. Этот коэффициент сравнительно велик, если вероятность угадать была низка, и сравнительно мал, если вероятность угадать была высока. Для равномерного распределения вероятность угадать всегда равна одной второй, а коэффициент всегда равен единице. Точная формулировка правила прироста капитала будет сейчас изложена.

Для последовательности $\mathbf{a}$ ее k-й член обозначаем a_k, для последовательности $\mathbf{a}'$ ее k-й член обозначаем a'_k, и т. д.

Ход, делаемый на j-м ходу, есть тройка $\langle n(j), i(j), v(j)\rangle$.

Пусть последовательность, которой располагает Казино, есть $\mathbf{a}$. Положим

$$A(k-1) = \{\mathbf{a}' \in \Omega : a'_{n(j)} = a_{n(j)} \text{ при } j = 1, 2 \ldots, k-1\}$$

(так что $A(0) = \Omega$),

$$A_i(k) = \{\mathbf{a}' \in A(k-1) : a'_{n(k)} = i\} \text{ для } i = 0, 1.$$

Разумеется, эти обозначения имеют смысл в предположении, что все участвующие в их определениях номера $n(l)$ определены. Полезно заметить, что

$$\Omega = A(0) \supset A(1) \supset A(2) \supset \ldots; \tag{1}$$

$$1 = \mu(A(0)) \geqslant \mu(A(1)) \geqslant \mu(A(2)) \geqslant \ldots; \tag{2}$$

$$\text{если Игрок на } k\text{-м ходу угадал, то } \; i(k) = a_{n(k)},\; A_{i(k)}(k) = A(k); \tag{3}$$

$$\text{если Игрок на } k\text{-м ходу не угадал, то } \; i(k) \neq a_{n(k)},\; A_{1-i(k)}(k) = A(k); \tag{4}$$

$$A(k-1) = A_0(k) \cup A_1(k). \tag{5}$$

В случае, если $i(k) = a_{n(k)}$ (прогноз Игрока на его k-м ходе оправдался — он угадал), капитал Игрока увеличивается по формуле

$$V(k) = V(k-1) + v(k)\mu(A_{1-i(k)}(k))/\mu(A_{i(k)}(k)). \tag{6}$$

Наша формула (6) гарантирует «честность» игры: математическое ожидание выигрыша, то есть прироста капитала, за один ход равно нулю. Однако на пути применения этой формулы нас подстерегает неприятность, невозможная для равномерного распределения — как и для любой позитивной меры (мера называется *позитивной*, если мера любого шара положительна). Неприятность заключается в том, что стоящая в знаменателе вероятность $\mu(A_{i(k)}(k))$ может оказаться равной нулю. Эта неприятность устраняется путем введения Дополнительного правила остановки, которое мы сейчас приведем. В случае равномерного распределения это правило было излишним, поскольку предусмотренная в нём ситуация не могла встретиться.

Дополнительное правило остановки. Оно применяется тогда, когда в ходе игры *впервые* наступает такая ситуация, что $\mu(A(k)) = 0$ (просим Читателя взглянуть на формулу (2)). Пусть $\mu(A(k-1)) \neq 0$, $\mu(A(k)) = 0$. Последний ход, который был сделан, имел номер k. Делая этот ход, Игрок объявил свой прогноз $i(k)$. Если прогноз оказался верным, то есть оказалось, что $i(k) = a_{n(k)}$, игра останавливается, а капитал Игрока считается возросшим до бесконечности: $V(k) = +\infty$. Таким образом, в этом случае Игрок объявляется выигравшим. Если же прогноз оказался неверным, т. е. оказалось, что $i(k) \neq a_{n(k)}$ (или, что то же самое, $1 - i(k) = a_{n(k)}$), игра опять-таки останавливается, но с тем капиталом Игрока $V(k)$, который у него был к этому моменту. Таким образом, в этом случае Игрок не выигрывает.

Ясно, что проблема с нулевым знаменателем в формуле (6) устраняется этим правилом. Действительно, формула (6) применяется лишь в случае, когда $i(k) = a_{n(k)}$. В этом случае, согласно (3), $A_{i(k)}(k) = A(k)$. Поэтому интересующий нас знаменатель может оказаться нулевым лишь в ситуации, когда $\mu(A(k)) = 0$. Но именно эта ситуация как раз и регулируется не формулой (6), а нашим Дополнительным правилом. (Можно, впрочем, считать, что и формулой (6), — если разрешить деление на нуль положительного числа $\mu(A_{1-i(k)}(k))$ и принять бесконечность в качестве результата такого деления.)

Определения стратегии, вычислимой стратегии, безостановочной стратегии, выигрывающей стратегии остаются, с учетом сделанных добавлений, для общего случая теми же, как и для частного случая равномерного распределения.

О безостановочных стратегиях. В заключение прокомментируем те предусмотренные правилами ситуации, когда игра останавливается. Заметим, что остановки не включены в понятие стратегии. Стратегия лишь указывает очередной ход или не указывает ничего, если таковой не определен; но в последнем случае игра не останавливается, а просто Игрок задумывается навечно — для вычислимой стратегии это означает, в математических терминах, что алгоритм указания хода работает безостановочно, не приходя ни к какому результату. Остановки же регулируются своими собственными правилами, «внешними» по отношению к стратегии. К сожалению, в общем случае вычислимого распределения наши правила остановки не содержат алгоритма их применения. Более того, алгоритм, позволяющий в любой момент определять, следует или не следует останавливать игру, в общем случае невозможен. Конкретно, для Основного правила речь идет о проверке неравенства, связывающего ставку и капитал, а для Дополнительного правила — о проверке положительности меры некоторого множества. Разумеется, если мера сильно вычислима, сложность со второй из указанных проверок отсутствует. Более того, для всякой вычислимо-рациональной (в частности, равномерной) меры очевидны алгоритмы обеих проверок, но, повторяем, подобные алгоритмы могут и отсутствовать в случае более сложно устроенной вычислимой меры.

Ввиду сказанного представляет интерес утверждение о том, что при определении понятия предсказуемости можно ограничиться одними только безостановочными стратегиями. Именно для обоснования этого утверждения нам и потребуется предположение о вычислимости меры (отметим, что до сих пор оно не использовалось). В этом предположении мы сейчас укажем, каким способом алгоритм **A**, задающий выигрывающую стратегию, переделывается в алгоритм **B**, задающий выигрывающую безостановочную стратегию.

Итак, нам задан алгоритм **A**, на вход которого поступает двоичное слово $x \in \Xi$, а на выходе, в качестве результата, либо не появляется ничего, либо появляется очередной ход. Содержательно, двоичные цифры, составляющие это слово, суть значения уже открытых к этому моменту членов последовательности. Как мы знаем, по слову x восстанавливается вся предыдущая история игры, т. е. все числа $n(k)$, $i(k)$, $v(k)$, $V(k)$ при $k \leqslant s = |x|$. Разумеется, эти числа восстанавливаются, если только они определены (кстати, можно всегда предполагать, что все они определены, если только определен и корректен s-й ход, каковой есть тройка $\langle n(s), i(s), v(s) \rangle$). Чтобы подчеркнуть зависимость указанных чисел от x, будем писать $n(x,k)$, $i(x,k)$, $v(x,k)$, $V(x,k)$; ведь когда мы ранее писали просто $n(k)$, $i(k)$ и т. д., мы предполагали фиксированной ту последовательность, против которой идет игра и «частью» которой является слово x. Вместе с указанными числами возникают, при всех $k \leqslant s$, и те множества, которые для фиксированной последовательности мы обозначали ранее как $A(k)$; теперь мы будем обозначать их $A(x,k)$.

Наша цель — переделать алгоритм **A** в алгоритм **B**, никогда не приводящий к ситуации, требующей остановки игры и такой, что если **A** был выигрывающим, то и **B** — выигрывающий. Для вычислимо-рациональной меры переделка очевидна: получив на вход двоичное слово x, проверяем, не приводит ли ход, указанный для этого слова алгоритмом **A**, к ситуации остановки; если не приводит, объявляем, что **B** указал нам этот ход; если приводит, объявляем, что **B** на x не дает результата.

В общем же случае произвольной вычислимой меры переделка **A** в **B** требует более изощренного приёма. Новый алгоритм **B** будет работать так. В случае, если ход, указанный алгоритмом **A**, не приводит к ситуации остановки, этот ход — как и для вычислимо-рациональной меры — объявляется результатом алгоритма **B**. Однако в случае, если в результате применения **A** к x возникает ситуация остановки, новый алгоритм **B** не будет давать никакого результата в применении к x. Тем самым окажется, что **B** задает безостановочную стратегию.

Для наглядности представим себе, что **A** и **B** работают параллельно. Напомним, что **A** нам задан, **B** мы строим. Пусть на вход обоих алгоритмов поступило слово x, $|x| = s$. Если **A** не дает результата в применении к этому слову, то и **B** объявляется не дающим результата. Если же **A** в применении к x выдает очередной ход, то алгоритм **B** пытается прежде всего проверить, не возникла ли ситуация, подпадающая под действие одного из правил остановки. Проверка корректности номера не вызывает затруднений. Если номер $n(x, s)$ открываемой карты оказался некорректным, алгоритм **B** объявляется не дающим результата на входе x; капитал Игрока в этом случае перестает меняться, он застывает. Если же номер корректен, включается проверка корректности ставки.

С проверкой корректности ставки дело обстоит сложнее. Как уже отмечалось, может и не существовать алгоритма, правильно отвечающего на вопрос, меньше ли делаемая ставка текущего капитала Игрока. Однако не всё так плохо. Прежде чем сделать ставку, Игроку следует убедиться, что намеченная ставка меньше текущего капитала, и только в этом случае ее объявить. Изложим новый алгоритм **B** поведения Игрока более точно. Мы опираемся на следующий факт: если мера μ вычислима, то можно предъявить такой вспомогательный алгоритм **C** (зависящий от **A**), который для каждого x пытается проверить, выполняется ли неравенство $v(x, s) < V(x, s)$, и достигает при этом следующего эффекта: алгоритм приходит к какому-то результату (скажем, к результату *Да*), если указанное неравенство выполняется, и работает бесконечное время, не приходя ни к какому результату, если выполняется противоположное неравенство $v(x, s) \geqslant V(x, s)$. Собираясь сделать свой s-й ход и получив от алгоритма **A** рекомендацию о ставке $v(x, s)$, Игрок, прежде чем сделать ход, включает алгоритм **C** и ждет результата. Никакого хода не делается, пока **C** не предъявит ему свой результат. Таким образом, алгоритм **B** не выдаст никакого хода, если **A** предложил некорректный ход; в этом случае капитал Игрока застывает. Если же работа алгоритма **C** заканчивается, алгоритм **B** рекомендует ту же ставку, что и **A**, и приступает к проверке того, не следует ли применить Дополнительное правило остановки.

Алгоритмическая ситуация с Дополнительным правилом похожа на ситуацию с правилом остановки, делаемой по причине некорректности ставки. Алгоритм, распознающий, выполняется или нет равенство $\mu(A(x,s)) = 0$, легко строится в случая сильно вычислимой меры, поскольку множество $A(x,s)$ получается в результате объединения конечного числа шаров, указываемых алгоритмически по паре $\langle x, s\rangle$. В общем же случае вычислимой меры такого алгоритма может и не быть. Существует, однако, более слабый алгоритм, которого оказывается достаточно. Именно, если мера μ вычислима, то можно предъявить такой вспомогательный алгоритм **D** (зависящий от **A**), который для каждого x пытается проверить, выполняется ли неравенство $\mu(A(x,s)) \neq 0$, и достигает при этом следующего эффекта: алгоритм приходит к какому-то результату (скажем, к результату *Да*), если указанное неравенство выполняется, и работает бесконечное время, не приходя ни к какому результату, если выполняется равенство $\mu(A(x,s)) = 0$. Действия алгоритма **B** в связи с проверкой Дополнительного правила описываются — в терминах поведения Игрока — следующим образом. Игрок включает алгоритм **D** и ждет результата. Дождавшись, он делает ход, применяя алгоритм **A**. Но он не делает никакого хода, пока **D** не предъявит ему свой результат. Таким образом, алгоритм **B** выдаст тот же ход, что и **A**, если не возникло ситуации, подпадающей под Дополнительное правило, и не выдаст никакого хода, если такая ситуация возникла. Если ход происходит, то капитал Игрока меняется на общих основаниях — уменьшается на величину ставки или увеличивается согласно формуле (6), причем в этом случае опасный знаменатель заведомо не равен нулю. В случае же неделания хода капитал Игрока, согласно нашим правилам, застывает, если последний прогноз был неверен, и объявляется бесконечным, если этот прогноз был верен.

Только что описанное поведение капитала Игрока можно следующим образом описать в терминах работы алгоритма **D**. Если прогноз неверен ($i(x) \neq a_{n(x)}$), то пока **D** работает, капитал Игрока остается застывшим. Если же прогноз верен ($i(x) = a_{n(x)}$), то с каждым шагом работы алгоритма **D** капитал Игрока возрастает на единицу. Таким образом, капитал будет возрастать неограниченно (и, следовательно, Игрок выиграет), если **D** никогда не закончит работу. Но если и когда алгоритм **D** достигает результата, рост капитала Игрока прекращается. Более того, в этом случае весь произошедший за время работы алгоритма **D** прирост капитала Игрока аннулируется, после чего этот капитал прирастает по формуле (6). И еще: никакое значение капитала, промежуточное между делаемыми ходами (т. е. возникающее во время *завершающейся* работы алгоритма **D**), не должно включаться в множество тех значений капитала, бесконечность супремума коих влечет выигрыш Игрока (в противном случае у Игрока появилась бы дополнительная возможность выигрыша за счет того, что длительности междуходовых задержек могут оказаться неограниченными в их совокупности).

Наше новое описание поведения капитала Игрока, предполагающее рост капитала во время работы алгоритма **D** с возможным последующим обнулением прироста, может вызвать у просвещенного Читателя два законных замечания.

Первое замечание состоит в том, что мы фактически изменили правила игры — в той их части, в которой говорится о порядке изменения капитала. Согласившись с этим упреком Читателя, ответим следующее. Действительно, есть две системы правил, старая и новая — а тем самым и два определения игры. Но в отношении предсказуемости обе игры равносильны: всякая последовательность, предсказуемая при одной из наших двух систем правил, предсказуема и при другой.

Второе замечание состоит в том, что по новым правилам поведение капитала зависит от вспомогательного алгоритма **D**, а сам этот алгоритм, как показывает анализ, зависит не от меры μ как таковой, а от задающего ее алгоритма — т. е. от того алгоритма, присутствующего в определении вычислимой меры, который для любого шара находит сколь угодно точные приближения к мере этого шара. Возникает естественный вопрос, не может ли случиться, что при новых правилах и понятие предсказуемости окажется зависящим не от самой меры, а от задающего ее алгоритма. Ответ: нет, не может. Ведь течение игры при одном задающем меру алгоритме может отличаться от течения игры при другом алгоритме, задающем ту же меру, лишь длительностью интервалов между ходами; что же касается капитала Игрока, то разница может быть лишь в том, сколь велик окажется тот прирост капитала, который подлежит аннулированию.

История и библиография

[1] А. Н. Колмогоров, В. А. Успенский. *Алгоритмы и случайность* // Теория вероятностей и ее применения, 1987. Т. 32, вып. 3, с. 425–455.

[2] В. А. Успенский, А. Л. Семёнов. *Теория алгоритмов: основные открытия и приложения.* — М.: Физматлит, 1987. — 288 с.

[3] В. А. Успенский, А. Л. Семёнов, А. Х. Шень. *Может ли индивидуальная последовательность нулей и единиц быть случайной?* // Успехи математических наук, 1990. Т. 45, вып. 1, с. 105–162.

[4] V. A. Uspensky, A. Shen. *Relations between varieties of Kolmogorov complexities* // Mathematical Systems Theory, 1996. Vol. 29, no.3, p. 271–292.

[5] An. A. Muchnik, A. L. Semenov, V. A. Uspensky. *Mathematical metaphysics of randomness* // Theoretical Computer Science, 1998. Vol. 207, p. 263–317.

[6] А. Шень. *О соотношениях между различными алгоритмическими определениями случайности* // Доклады Академии наук СССР, 1988. Т. 302, № 3, с. 548–552.

[7] В. В. Вьюгин. *Алгоритмическая энтропия (сложность) конечных объектов и ее применение к определению случайности и количества информации* // Семиотика и информатика. Выпуск 16. — М: ВИНИТИ, 1981. — С. 14–43.

[8] M. Li, P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* — New York e. a.: Springer-Verlag, 1993. — xx+546 pp., 38 illustrations.;
Second edition. — New York e. a.: Springer-Verlag, 1997. — xx+637pp., 41 illustrations.

Сам этот список из восьми названий никоим образом не претендует на полноту. Однако в этих публикациях (особенно в [8]) можно найти дальнейшие ссылки, и вкупе с этими ссылками некое подобие полноты уже достигается.

В книге [2] следует обратить внимание на § 2.6 «Приложения к теории вероятностей: определения случайной последовательности». Следует также иметь в виду, что терминология этой книги несколько архаична: хаотические последовательности называются там *случайными по Колмогорову*, а типические — *случайными по Мартин-Лёфу*. Частотоустойчивые, они же стохастические, последовательности называются в [2] *случайными по Мизесу*, стохастические по Чёрчу — *случайными по Мизесу – Чёрчу*, стохастические по Колмогорову — *случайными по Мизесу – Колмогорову – Лавлэнду* (а в [3] — *стохастическими по Колмогорову – Лавлэнду*; Д. Лавлэнд ⟨D. Loveland⟩ открыл эти последовательности независимо от Колмогорова, хотя и позже: соответствующая статья Колмогорова была опубликована в 1963 г., статья Лавлэнда — в 1966 г.) Что касается непредсказуемых последовательностей — в том точном понимании, как было сформулировано выше, — то они в указанной книге отсутствуют, поскольку появились в литературе лишь в 1998 г. в статье [5].

Пример стохастической по Чёрчу последовательности, перестающей быть таковой после подходяще выбранной вычислимой перестановки ее членов, опубликовал в 1966 г. Д. Лавлэнд. Значение этого примера состоит не только в том, что он показывает неадекватность определения Чёрча, но и в том, что он открывает новое качество случайности, интуитивно очевидное, но ранее не замеченное: случайная последовательность должна оставаться случайной после любой вычислимой (т. е. задаваемой каким-то алгоритмом) перестановки ее членов.

Теория сложности объектов (в дополнение к уже развивавшейся в то время теории сложности вычислений) была основана Колмогоровым на его семинарах в Московском университете в начале 60-х годов XX века и имела своей главной целью перестройку понятийной базы теории информации (на основе представления о том, что чем сложнее объект, тем больше информации он содержит) с последующим приложением к теории случайности. В опубликованной в 1969 г. статье Колмогоров писал:

> 1) основные понятия теории информации должны и могут быть обоснованы без помощи обращения к теории вероятностей и так, что понятия «энтропия» и количество информации оказываются применимы к индивидуальным объектам;
>
> 2) введенные таким образом понятия теории информации могут лечь в основу новой концепции случайного, соответствующей естественной мысли о том, что случайность есть отсутствие закономерности.

Предложение измерять сложность объекта длиной его кратчайшего описания и понятие оптимального языка были изложены Колмогоровым в статье 1965 г.; за год до того сходные идеи были опубликованы американским исследователем Рэем Соломоновым ⟨Ray Solomonoff⟩, о работах которого Колмогоров узнал лишь позже. Ввиду этого теорему о существовании оптимального языка мы называем *теоремой Соломонова – Колмогорова*. В эти же годы (т. е. в середине 60-х годов XX в.) Колмогоров высказывает на своих семинарах предположение о том, что быстрота роста энтропии начальных отрезков последовательности может служить критерием случайности рассматриваемой последовательности. Однако введенное им в рассмотрение языково́е семейство приводило к энтропии, непригодной для поставленной цели. Как уже говорилось выше, в главке о хаотичности, годное для этой цели семейство обнаружил в 1973 г. Леонид Левин, введя в рассмотрение монотонную энтропию.

Типические последовательности (под названием “random”, т. е. «случайные») были, как уже отмечалось в главке о типичности, открыты в 1966 г. Пером Мартин-Лёфом ⟨Per Martin-Löf⟩.

Утверждение о том, что класс последовательностей, стохастических по Колмогорову, шире класса типических (они же — хаотические) последовательностей, доказал Александр Шень (см. [6], а также [3, п. 6.2.4]).

Пусть K — энтропия в каком-то из вариантов этого понятия. (В литературе исследованы не менее шести таких вариантов: простая, априорная, монотонная, процессная, префиксная и энтропия разрешения; все эти варианты существенно различны в том точном смысле, что разность энтропий, принадлежащих любым двум из перечисленных вариантов, неограничена.) Скажем, что последовательность $a_1, a_2, a_3, \ldots, a_n \ldots$ является *хаотической относительно* K (при равномерном распределении!), если

$$\exists c\, \forall n\, \mathrm{K}(a_1, a_2, a_3, \ldots, a_n) > n - c.$$

(Заметим для ясности, что ни для простой энтропии, ни для энтропии разрешения последовательностей с таким свойством нет вовсе, а для каждой из остальных четырех энтропий хаотичность оказывается равнообъемной типичности.)

В той же статье Левина 1973 г., в которой была введена монотонная энтропия, содержалось и доказательство того факта, что понятие хаотичности относительно монотонной энтропии равнообъемно понятию типичности. Независимо от Левина в том же 1973 г. Клаус-Петер Шнорр ⟨Claus-Peter Schnorr⟩ ввел в рассмотрение свой вариант энтропии, так называемую *процессную энтропию* (у Шнорра — *process complexity*), и доказал (независимо от Левина, но очень похожим способом), что понятие хаотичности относительно процессной энтропии также равнообъемно понятию типичности. Хотя процессная энтропия и монотонная энтропия существенно различаются (как показал Владимир Вьюгин [7, с. 35, строки 6–4 снизу], их разность не ограничена никакой константой) и хотя впоследствии сам Шнорр перестал пользоваться своей процессной энтропией, фактически от нее отказавшись, вышеуказанную *теорему Левина* иногда называют *теоремой Левина – Шнорра*.

Префиксную энтропию ввел в 1974 г. Левин и годом позже (но независимо от Левина) Грегори Чэйтин ⟨Gregory J. Chaitin⟩ — см. его статью A theory of program size formally identical to information theory // Journal of the Association of Computing Machinery, 1975, v. 22, no. 3, p. 329–340). В той же статье Чэйтин ввел понятие хаотичности относительно префиксной энтропии и объявил (без доказательства), что этот вариант хаотичности равносилен типичности; первое опубликованное доказательство этого факта появилось в статье Вьюгина [7]: следствие 3.2 на с. 38. *Префиксная энтропия* может быть определена как энтропия для семейства префиксных языков. Язык E называется *префиксным*, если он перечислим и выполнено условие:

$$[\langle x_1, y_1\rangle \in E \ \&\ \langle x_2, y_2\rangle \in E \ \&\ (x_1 \approx x_2)] \Longrightarrow [y_1 = y_2].$$

Заметим еще, что в литературе термин «сложность» (“complexity”) часто употребляется в смысле „энтропия“, т.е. в смысле „сложность относительно оптимального языка“.

Непредсказуемые последовательности (в смысле настоящей статьи) впервые возникли весной 1991 г. в совместном докладе под названием “Randomness and Lawlessness” («Случайность и беззаконность»), который Андрей Мучник, Алексей Семёнов и автор этих строк сделали на конференции в Калифорнии. Конференция была посвящена основаниям теории случайности и проходила с 4 по 7 марта в Институте математических исследований в социальных науках (Institute for Mathematical Studies in the Social Sciences) Станфордского университета. Содержание доклада было опубликовано в 1998 г. в виде статьи [5]. В этой статье приведены, в частности, теоремы о соотношении непредсказуемости с другими алгоритмическими лицами случайности.

Отметим, что определение непредсказуемости в нашем настоящем очерке отличается одной деталью от определения в [5]. Именно, в [5] для корректности ставки требовалось выполнение не строгого неравенства $v(k) < V(k-1)$, а более слабого нестрогого неравенства $v(k) \leqslant V(k-1)$. Класс непредсказуемых последовательностей остается одним и тем же при обоих пониманиях корректности. Замена в определении корректности нестрогого неравенства на строгое вызвана двумя причинами. Во-первых, сама игра становится более содержательной: ведь в случае равенства ставки текущему капиталу стоит Игроку ошибиться в своем предсказании, как его капитал обнулится, и он будет вынужден в дальнейшем делать лишь нулевые ставки. Во-вторых, именно строгое неравенство позволяет при переносе определений с рационально-вычислимых мер (каковые только и рассматривались в [5]) на любые вычислимые меры, сохранить эффективность действий игрока. Ведь каждый раз Игрок должен удостоверяться в корректности хода. Получить же такое удостоверение алгоритмическим путем возможно лишь в варианте строгого неравенства: существует алгоритм, в том и только в том случае дающий положительный ответ на вопрос «$v(k) < V(k-1)$?», когда и в самом деле $v(k) < V(k-1)$; не существует алгоритма, в том и только в том случае дающего положительный ответ на вопрос «$v(k) \leqslant V(k-1)$?», когда и в самом деле $v(k) \leqslant V(k-1)$.

Сама идея о связи случайности с невозможностью гарантированного выигрыша достаточно очевидна: еще фон Мизес, не давая точных формулировок, говорил о «невозможности системы игры». Впоследствии встречались и строгие определения, однако воспроизводимая здесь (с косметическим ремонтом) формулировка из [5] кажется более близкой к интуитивному представлению о случайности. Дело в том, что предшествующие игровые определения либо использовали стратегии, в которых вычислимость (означающая наличие алгоритма, указывающего игроку его очередной ход) заменялась на другое (хотя и связанное с понятием алгоритма, но, видимо, менее естественное) требование, либо заведомо не приводили к классу последовательностей, равнообъемному классу типическо-хаотических последовательностей. Для определения непредсказуемости из [5] остается надежда на указанную равнообъемность. Ясно, что чем бо́льшим количеством «лиц случайности» характеризуется какой-то точно очерченный класс последовательностей, тем обоснованнее право этого класса служить формальным аналогом расплывчатого интуитивного представления о случайности.

В. А. Успенский, мехмат МГУ

Алгебра многогранников

Г. Ю. Панина

1. Введение

Статья предназначена старшеклассникам, целенаправленно занимающимся математикой, студентам и преподавателям. Кроме того, она может быть рекомендована математикам-профессионалам в качестве легкого чтения.

Для удобства читателя мы используем двухуровневую систему: (меньшая) часть текста выделена мелким шрифтом — этот более сложный материал может быть пропущен начинающими.

А вот задачи пропускать нельзя — над ними нужно подумать (в Приложении приведены ответы и указания к некоторым задачам).

Совсем недавно была придумана замечательная вещь — алгебра многогранников. Как это иногда бывает с действительно важными конструкциями, ее придумали одновременно и независимо разные люди — Александр Пухликов совместно с Аскольдом Хованским и Питер МакМаллен (P. McMullen). И как это часто бывает, алгебра многогранников имеет отношение к разным областям математики.

Например, есть такой вариант задачи о равносоставленности: многогранник разрешается разрезать на куски, а затем, передвигая куски с помощью параллельных переносов, составить новый многогранник. (Звучит почти как третья проблема Гильберта, но смысл здесь иной). Алгебра многогранников, которую мы построим, помогает найти критерий равносоставленности для этой задачи.

С другой стороны, алгебра многогранников тесно связана с алгебраической геометрией торических многообразий. Эта «высокая наука» может быть переведена на простой и знакомый язык многогранников (так же, как делаются переводы с одного языка на другой). В соответствующем «словаре» термин «многогранник» переводится как «очень обильный пучок». «Сложение многогранников по Минковскому» переводится как «тензорное произведение пучков». Известно, что очень обильные пучки можно не только тензорно перемножать, но и делить друг на друга — они порождают группу Пикара. Если перевести последнюю фразу на язык многогранников, то получится следующее утверждение.

Выпуклые многогранники можно не только складывать, но и вычитать по Минковскому — они порождают группу виртуальных многогранников. Что это значит геометрически, будет рассказано в параграфе 6.

Более того, мы покажем, что выпуклые многогранники порождают еще более богатую структуру — градуированную алгебру (это и есть алгебра многогранников).

Говоря простым языком, по своим алгебраическим свойствам многогранники похожи на многочлены.

На примере этой красивой конструкции вы увидите, как в математике действует старая детская поговорка: «Если нельзя, но очень хочется, то можно.»

Например, нельзя вычесть по Минковскому из маленького многогранника большой. Но если придумать как, то можно.

Выражения «логарифм треугольника» и «экспонента тетраэдра» — бессмыслица. Но если проявить фантазию, то они станут осмысленными и позволят раскладывать многогранники в сумму однородных компонент.

Изложение отнюдь не претендует на полноту. Наша задача — дать общее представление о конструкции и методах, попутно научив читателя некоторым общим математическим приемам.

2. Сложение по Минковскому

Все геометрические объекты, которые мы рассматриваем, живут в трехмерном или двумерном вещественном пространстве — выше мы подниматься не будем.

В пространствах большей размерности дело обстоит точно так же и не возникает никаких дополнительных эффектов.

Считая фиксированным начало координат O, для удобства мы отождествляем точку пространства с ее радиус-вектором.

Выпуклым многогранником (в $\mathbb{R}^2$ или $\mathbb{R}^3$) называется выпуклая оболочка непустого конечного множества точек.

Таким образом, точка или отрезок тоже считаются многогранниками, а открытый квадрат (квадрат без границы) — нет.

Наша цель — организовать многогранники в алгебраическую структуру, т. е. научиться производить с ними алгебраические действия, подчиненные обычным аксиомам.

Первое алгебраическое действие — суммирование по Минковскому.

Здесь имеется исторически сложившаяся путаница: скоро мы увидим, что это вовсе не сложение, а умножение. Поэтому мы используем для этой операции знак $\otimes$, которым обычно обозначают тензорное произведение и (иногда) свертку.

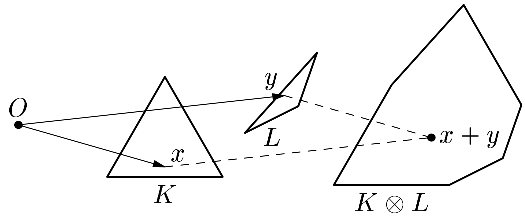


Рис. 1.

Определение 1. *Суммой Минковского* (или векторной суммой) двух выпуклых многогранников K и L называется множество точек

$$K \otimes L = \{x + y \mid x \in K, y \in L\}.$$

Задача 1. Найти сумму Минковского двух параллельных отрезков.

Задача 2. Найти сумму Минковского двух непараллельных отрезков.

Задача 3. Что произойдет с суммой Минковского, если подвинуть одно из слагаемых параллельным переносом?

Задача 4. Найти сумму Минковского треугольника и отрезка, лежащих в одной плоскости.

Задача 5. Найти сумму Минковского треугольника и отрезка, не лежащих в одной плоскости.

Задача 6. Что произойдет с суммой Минковского, если изменить начало координат?

Задача 7. Выпуклой оболочкой каких точек является сумма Минковского двух многогранников?

Задача 8. Убедитесь, что $K \otimes K$ есть многогранник K, гомотетично растянутый в 2 раза.

Следующее предложение почти очевидно.

Предложение 1.

- ▷ *Операция $\otimes$ коммутативна.*
- ▷ *Сумма Минковского содержит каждое из слагаемых после подходящего параллельного переноса.*
- ▷ *Сумма Минковского двух выпуклых многогранников — выпуклый многогранник.*
- ▷ *Для $\mathbf{1} = \{O\}$ и любого выпуклого многогранника K, справедливо: $K \otimes \mathbf{1} = K$.*

Последнее утверждение означает что у нас имеется нейтральный элемент **1** — многогранник, состоящий из одной точки O.

Внимательный читатель уже заметил, что мы приближаемся к понятию «абелева группа», но — увы! — второе утверждение предложения не дает никакой надежды на существование обратного, так как при сложении по Минковскому слагаемые не уменьшаются. Поэтому в классе выпуклых многогранников обратимы только одноточечные многогранники.

Разумеется, полугруппа выпуклых многогранников стандартным образом вкладывается в свою группу Гротендика, т. е. группу формальных выражений вида $K \otimes L^{-1}$. Наша задача — придать геометрический смысл этим выражениям. Это будет сделано на языке виртуальных многогранников (п. 6) и на двойственном языке опорных функций (п. 3.)

3. Опорная функция и веер выпуклого многогранника

Определение 2. Опорной функцией выпуклого многогранника K называется функция $h_K(x)$, заданная на $\mathbb{R}^2$ (или $\mathbb{R}^3$, в зависимости от того, где живет K) формулой

$$h_K(x) = \max_{y \in K}(x, y).$$

(скобка обозначает скалярное произведение радиус-векторов).

Для вектора x максимум этого скалярного произведения достигается в точках пересечения многогранника K и прямой, касающейся K и ортогональной x (см. рис. 2).

Задача 9. Вообще-то, существуют две прямые с таким свойством. Какую именно мы берем?

Вот очевидные свойства опорной функции.

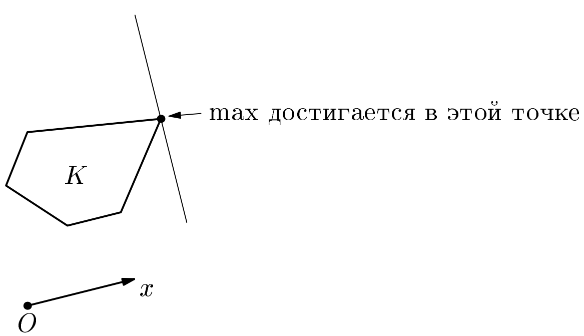


Рис. 2.

Предложение 2.

1. *В начале координат опорная функция равна нулю*: $h_K(O) = 0$.
2. *Для* $\lambda > 0$ *имеем* $h_K(\lambda x) = \lambda h_K(x)$.
3. *Опорная функция одноточечного многогранника — линейная функция. В частности, это означает, что график опорной функции выпуклого многогранника из* $\mathbb{R}^2$ *есть плоскость в трехмерном пространстве.*
4. *Любая линейная функция, принимающая значение* 0 *в начале координат, является опорной функцией некоторого одноточечного многогранника.*

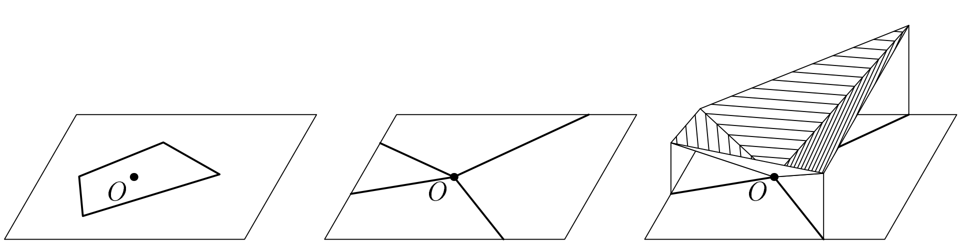
Задача 10. Докажите, что опорная функция суммы Минковского равна сумме опорных функций слагаемых:

$$h_{K \otimes L} = h_K + h_L.$$

Рассмотрим плоский многоугольник K. При фиксированном x скалярное произведение (см. определение опорной функции) достигает своего максимума в некоторой вершине A многоугольника K. Если мы будем потихоньку поворачивать радиус-вектор x, то некоторое время максимум будет достигаться в той же вершине A, затем в некоторый момент — на целом ребре, примыкающем к вершине A, а затем перепрыгнет на соседнюю вершину. Это значит, что опорная функция многогранника «склеена» из кусочков опорных функций вершин, т. е. из кусков линейных функций.

Пространство $\mathbb{R}^2$ оказывается разбито на плоские конуса (это то же самое, что плоские углы) с общей вершиной в начале координат так, что на каждом из конусов опорная функция линейна.

Тем самым, график опорной функции склеен из плоских кусочков, причем места склейки — лучи, ортогональные ребрам K (рис. 3).



многоугольник K веер K график опорной функции K

Рис. 3.

Что же происходит в трехмерном пространстве, какова опорная функция трехмерного многогранника K? Принципиально здесь всё то же самое, только картинка получается богаче: $\mathbb{R}^3$ оказывается разбито на многогранные конуса с общей вершиной в начале координат так, что на каждом из конусов опорная функция линейна.

Определение 3. Это разбиение называется *веером* многогранника K.

Удобно рисовать не сам веер, а его пересечение с единичной сферой с центром в точке O (*сферический веер* многогранника K). Мы получим разбиение сферы на сферические многогранники (*клетки веера*).

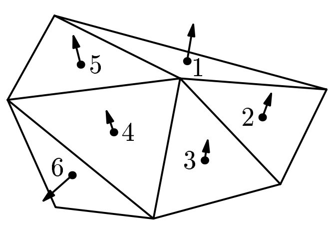
Имея многогранник, легко построить его сферический веер:

1. Отметьте на сфере концы внешних нормалей граней многогранника K.
2. Соедините на сфере каждые две полученные точки кратчайшим отрезком (куском большого круга), если соответствующие грани делят ребро.

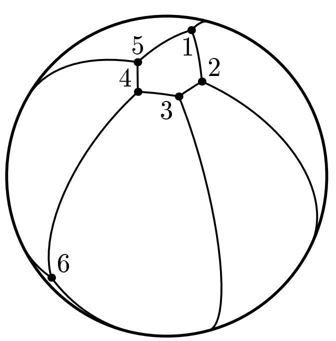
(Докажите, что в результате этой процедуры действительно получается сферический веер.)

Обратите внимание на *комбинаторную двойственность* многогранника и его веера:

- ▷ вершинам веера соответствуют грани многогранника,
- ▷ ребрам веера соответствуют ребра многогранника,
- ▷ клеткам веера соответствуют вершины многогранника.



K



веер K

Рис. 4.

Задача 11. Постройте сферический веер куба, тетраэдра, треугольника (лежащего в трехмерном пространстве), отрезка (лежащего в трехмерном пространстве).

Теорема 1.

1. *Опорная функция выпуклого многогранника выпукла. Это значит, что для* $0 \leqslant \lambda \leqslant 1$
$$h_K(\lambda x_1 + (1-\lambda)x_2)) \leqslant \lambda h_K(x_1) + (1-\lambda)h_K(x_2),$$
или (для тех, кто любит картинки больше формул) что график функции h_K *— выпуклая вниз поверхность.*

2. *Каждая непрерывная выпуклая функция* h*, принимающая значение* 0 *в начале координат, которая кусочно линейна относительно некоторого веера (т. е. разбиения пространства на конуса), является опорной функцией некоторого выпуклого многогранника. (Множество таких функций обозначим через* $\mathcal{H}$*.)*

3. *Опорная функция одноточечного многогранника — линейная функция. В частности, это означает, что график опорной функции точки из* $\mathbb{R}^2$ *есть плоскость в трехмерном пространстве.*

Поясним пункт 2. Восстановить многогранник легко: те линейные функции, из которых склеена h, дают нам координаты вершин.

При сложении по Минковскому выпуклых многогранников веера дробятся: чтобы нарисовать сферический веер суммы, нужно на одной и той же сфере нарисовать веера обоих слагаемых.

Мы поняли, что грубо говоря, выпуклый многогранник — это то же самое, что и его опорная функция: эти два объекта однозначно определяют друг друга, причем сложить два многогранника по Минковскому — то же самое, что сложить их опорные функции.

Официально это звучит так: полугруппа выпуклых многогранников канонически изоморфна полугруппе выпуклых кусочно-линейных функций $\mathcal{H}$. Это влечет изоморфизм соответствующих групп Гротендика.

Обращать (и вычитать) по Минковскому выпуклые многогранники мы пока не умеем. Но зато можно запросто брать со знаком «минус» (и вычитать друг из друга) опорные функции. Что при этом мы получим? Разность двух функций из $\mathcal{H}$ останется, разумеется непрерывной и кусочно-линейной относительно некоторого веера. А вот свойство выпуклости пропадет. Забавная ситуация: мы пока не знаем, что за объект «обратный по Минковскому к кубу», но уже поняли, чему равна его опорная функция. Хотелось бы придать этому объекту геометрический смысл. Мы это сделаем, но в несколько ходов.

4. Интегрирование по эйлеровой характеристике

Эйлерова характеристика — бесспорный лидер среди математических понятий по соотношению «простота – полезность». Об этом знает каждый, кому приходилось причесывать ежика, раскрашивать карты или учиться отличать тор от кренделя. Оказывается, по эйлеровой характеристике можно интегрировать, обращаясь с ней как с мерой.

Определение 4. Множество точек будем называть *многогранным*, если его можно получить конечным числом операций объединения, пересечения и дополнения из некоторого (конечного) набора выпуклых многогранников.

Определение 5. *Клеточным разбиением* многогранного множества M называется разбиение M на непересекающиеся многогранные подмножества (*клетки разбиения*) такие, что каждая клетка есть одно из следующих множеств:

0: точка (0-мерная клетка);

1: открытый отрезок (отрезок с выброшенными концами) (1-мерная клетка);

2: открытый двумерный многоугольник (многоугольник с удаленной границей) (2-мерная клетка);

3: открытый трехмерный многогранник (3-мерная клетка).

Определение 6. *Эйлеровой характеристикой* клеточного разбиения множества M называется

$$\chi(M) = (\text{число 0-мерных клеток}) - (\text{число 1-мерных клеток}) + \\ + (\text{число 2-мерных клеток}) - (\text{число 3-мерных клеток}).$$

Хотя одно и то же множество можно по-разному разбить на клетки, значение $\chi(M)$ от этого не меняется. (Это далеко не тривиальный факт, который мы оставляем без доказательства.)

Очевидное, но очень важное свойство эйлеровой характеристики — аддитивность:

$$\chi(M_1 \cup M_2) = \chi(M_1) + \chi(M_2) - \chi(M_1 \cap M_2).$$

Интегрировать нам предстоит только очень просто устроенные функции.

Определение 7. Пусть $f\colon M \to \mathbb{Z}$ — функция, заданная на некотором многогранном множестве M, такая, что f принимает конечное множество целочисленных значений, причем прообраз каждого числа —

многогранное множество. Такие функции мы будем называть *многогранными*, а множество всех многогранных функций, заданных на всем пространстве, будем обозначать через $\mathcal{M}$.

Положим по определению

$$\int_M f(x)\,d\chi(x) = \sum_{a\in\mathbb{Z}} a\cdot\chi(f^{-1}(a)).$$

Это и есть *интеграл функции по эйлеровой характеристике.*

И всё: нет никаких верхних и нижних интегральных сумм, никаких переходов к пределу, никаких борелевых множеств (всех атрибутов интеграла Лебега).

Зато есть привычные свойства интеграла (докажите!):

Теорема 2.

1. *Интеграл суммы функций есть сумма интегралов слагаемых*
$$\int_M (f(x)+g(x))\,d\chi(x) = \int_M f(x)\,d\chi(x) + \int_M g(x)\,d\chi(x).$$
2. *Интеграл по объединению непересекающихся множеств есть сумма интегралов по каждому из них*
$$\int_{M\cup L} f(x)\,d\chi(x) = \int_M f(x)\,d\chi(x) + \int_L f(x)\,d\chi(x).$$
3. (*Теорема Фубини.*) *Можно интегрировать сначала по каждому слою, а затем — полученную функцию проинтегрировать по слоям*:
$$\int f(x,y)\,d\chi(x,y) = \int\Big(\int f(x,y)\,d\chi(x)\Big)d\chi(y).$$

5. От многогранников к функциям. Сложение по Минковскому как свертка по эйлеровой характеристике

Определение 8. *Сверткой* по эйлеровой характеристике двух многогранных функций f и g называется функция $f\otimes g$, заданная формулой

$$f\otimes g(x) = \int f(x-y)g(y)\,d\chi(y).$$

С каждым выпуклым многогранником K естественно связана многогранная функция I_K — его характеристическая функция:

$$I_K = \begin{cases} 1, & \text{если } x\in K;\\ 0, & \text{иначе.}\end{cases}$$

Предложение 3. *Пусть K и L — выпуклые многогранники. Тогда свертка I_K и I_L есть характеристическая функция суммы Минковского K и L:*

$$I_K \otimes I_L = I_{K \otimes L}.$$

Иными словами, сложить два многогранника — то же самое, что свернуть их характеристические функции по эйлеровой характеристике.

Поэтому мы можем позволить себе вольность обозначать выпуклый многогранник K и его характеристическую функцию I_K одним и тем же символом K.

Доказательство. Нам нужно доказать, что

$$\int I_K(x-y)I_L(y)\, d\chi(y) = \begin{cases} 1, & \text{если } x \in K \otimes L; \\ 0, & \text{иначе.} \end{cases}$$

Под знаком интеграла стоит характеристическая функция пересечения двух выпуклых многогранников. Следовательно, интеграл при фиксированном x равен нулю, если эти многогранники не пересекаются (и сама функция — тождественный ноль), и равен единице, если подынтегральная функция — не тождественный ноль. Осталось заметить, что равенство $I_K(x-y)I_L(y) = 1$ справедливо при некотором y тогда и только тогда, когда x представим в виде суммы точек $x-y$ и y из многогранников K и L. □

6. Обращение по Минковскому. Группа виртуальных многогранников. Выворачивание наизнанку

С этого момента выпуклый многогранник — не только множество точек, но и многогранная функция.

Это обстоятельство и позволяет обращать выпуклые многогранники (относительно $\otimes$), но не в классе выпуклых многогранников (что, как мы убедились, невозможно), а в классе многогранных функций. А именно, верна теорема об обращении по Минковскому:

Теорема 3. *Для любого выпуклого многогранника K существует многогранная функция K^{-1} такая, что*

$$K \otimes K^{-1} = \mathbf{1}.$$

Эта функция устроена просто (см. рис. 5), она принимает только два значения:

$$K^{-1} = \begin{cases} (-1)^{\dim K}, & \textit{если } x \in \operatorname{Int}(S(K)); \\ 0, & \textit{иначе.} \end{cases}$$

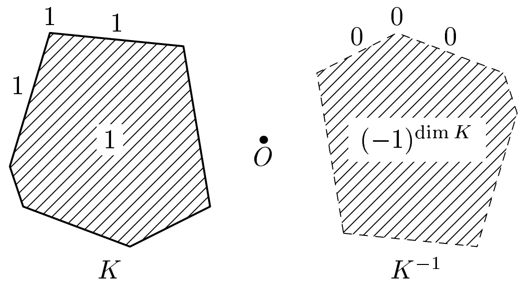


Рис. 5.

Поясним обозначения: $\dim K$ — размерность многогранника, S — центральная симметрия относительно начала координат, $\operatorname{Int} K$ — многогранник K без границы (от слова “interior” — внутренность).

Задача 12. Убедитесь, что теорема верна
– для отрезков на прямой (одномерных многогранников),
– для многоугольников на плоскости.
(После этого для трехмерных многогранников она станет очевидна.)

Задача 13. Прежде, чем читать дальше, подумайте, как определить разность по Минковскому двух выпуклых многогранников.

Определение здесь естественное:

Определение 9. Разность Минковского двух выпуклых многогранников $K \otimes L^{-1}$ есть свертка функций K и L^{-1}.

В отличие от обратного к выпуклому многограннику (который, грубо говоря, не сильно отличается от выпуклого многогранника), разности Минковского гораздо более разнообразны. Например, разность двух квадратов (один из них — повернутая копия другого) — правильная звезда с 8 лучами (см. рис. 6). На рисунке указаны значения этой функции, но не все — из-за нехватки места.

Забавно проследить выворачивание наизнанку (см. рис. 7): возьмем трапецию и вычтем из нее отрезок, параллельный основанию. Будем постепенно увеличивать длину вычитаемого отрезка. Пока отрезок маленький, у нас будет получаться (выпуклая) трапеция. Когда отрезок достигнет длины меньшего основания, разность выродится в треугольник. Затем появится растущий отрицательный кусок, а положительная часть будет уменьшаться, пока не исчезнет совсем.

Рассмотрим множество $\mathcal{P}$ всех многогранных функций, представимых в виде $K \otimes L^{-1}$. Операция свертки по Минковскому превращает $\mathcal{P}$ в абелеву группу. Действительно, мы уже знаем, что такое нейтральный

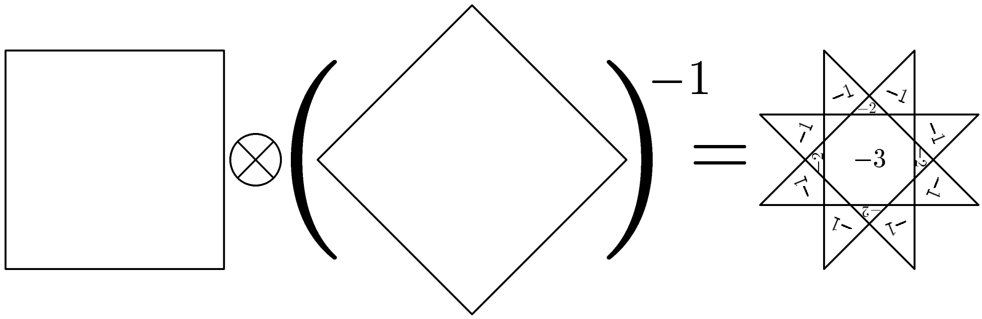


Рис. 6.

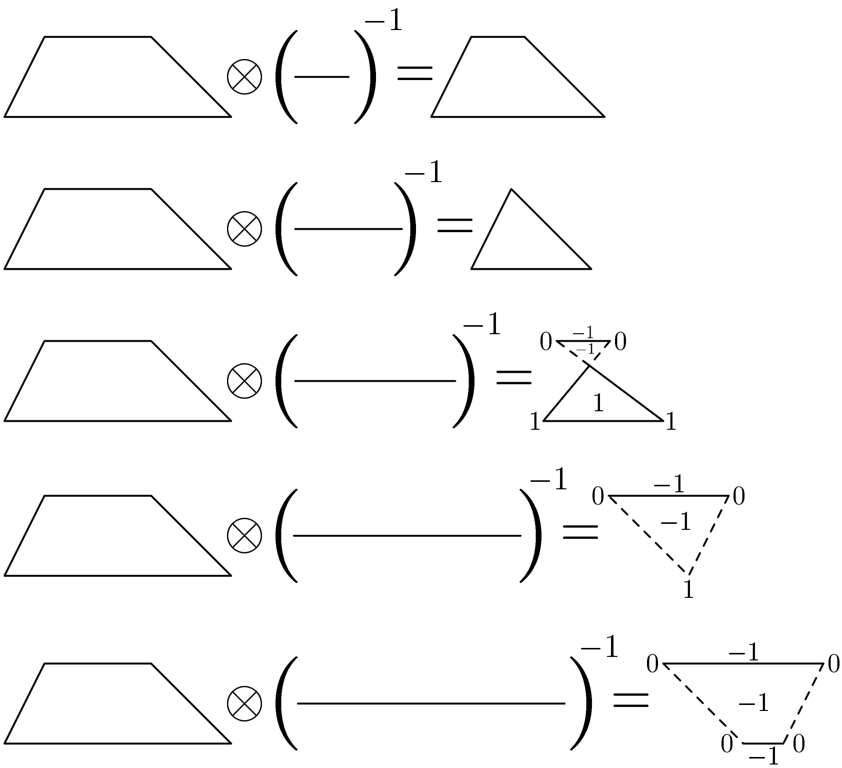


Рис. 7.

элемент, умеем обращать выпуклые многогранники (а значит, и выражения вида $K \otimes L^{-1}$).

Определение 10. $\mathcal{P}$ называется *группой виртуальных многогранников*.

7. Веера виртуальных многогранников

Что такое опорная функция виртуального многогранника, мы поняли раньше, чем поняли, что такое сам виртуальный многогранник.

Группа $\mathcal{P}$ изоморфна группе непрерывных функций, кусочно-линейных относительно некоторого веера.

Как может выглядеть сферический веер виртуального многогранника $K \otimes L^{-1}$? Нанесем на сферу линии излома функций h_K и h_L. Когда мы рассмотрим линии излома функции $h_K - h_L$, некоторые отрезки могут исчезнуть, здесь не будет простого измельчения, как при сложении выпуклых многогранников, и клетки полученного веера вполне могут оказаться невыпуклыми.

Вот красивый и важный пример.

Пример 1. Гиперболический тетраэдр. Сферический «Инь и Янь».

Возьмем правильный тетраэдр Δ, отметим два противоположных ребра и вычтем из Δ отмеченные рёбра. Веер полученного виртуального многогранника (проверьте!) состоит из четырех равных невыпуклых частей (см. рис. 8).

С выпуклым многогранником естественно связана выпуклая многогранная поверхность — объединение его граней.

С некоторыми невыпуклыми многогранными поверхностями (которым не запрещается иметь самопересечения) можно связать виртуальный многогранник.

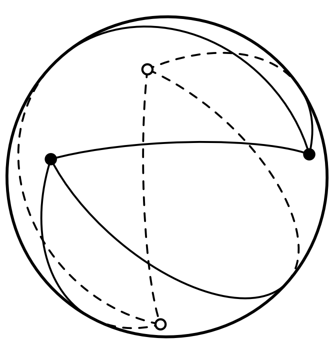


Рис. 8.

Это делается поэтапно согласно следующему алгоритму. Может получиться так, что очередной шаг невыполним. Это означает, что с данной поверхностью невозможно связать виртуальный многогранник.

Может получиться так, что выполнить некоторый шаг можно по-разному. Значит, виртуальных многогранников, связанных с этой поверхностью, несколько. (Например, с тетраэдром можно связать 52 различных виртуальных многогранника.)

Шаг 1. Для каждой грани поверхности зафиксируем ортогональный ей вектор.

Шаг 2. Нанесем концы всех этих векторов на единичную сферу. Полученные точки должны оказаться различными.

Шаг 3. Построим веер будущего виртуального многогранника. Соединим полученные точки геодезическими отрезками (не обязательно кратчайшими!) по следующему правилу: полученная картинка должна быть комбинаторно двойственна поверхности. В частности, мы соединяем две точки тогда, когда соответствующие грани делят ребро.

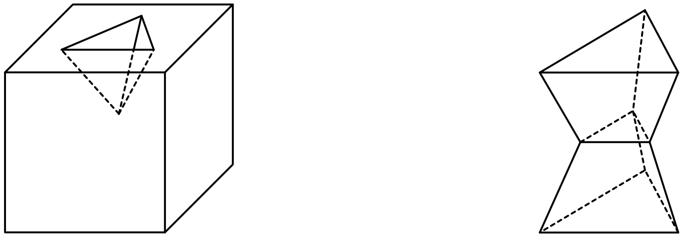
Шаг 4. Построим теперь кусочно-линейную функцию h. Она будет склеена из кусков опорных функций вершин поверхности. Возьмем клетку полученного веера, натянем на нее конус и положим функцию h на клетке равной опорной функции той вершины нашей поверхности, которая соответствует этой клетке. (То же самое происходит и в выпуклом случае). Проделав так со всеми клетками, получим кусочно-линейную функцию.

Шаг 5. Представим h в виде разности двух выпуклых функций $h = h_1 - h_2$. (Почему это возможно — отдельная задача. Решите ее.) Функции h_1 и h_2 соответствуют выпуклым многогранникам K_1 и K_2. Искомый виртуальный многогранник равен $K_1 \otimes K_2^{-1}$.

Задача 14. Каким поверхностям (рис. 9) соответствуют виртуальные многогранники?

8. От группы к алгебре. Инварианты Хадвигера на плоскости

На множестве всех многогранных функций зададим структуру алгебры над полем рациональных чисел. Для этого надо описать три операции, удовлетворяющие стандартным аксиомам, — сложение двух функций, умножение двух функций и умножение функции на рациональное число.



куб с тетраэдральной ямкой склейка двух усеченных тетраэдров

Рис. 9.

1. Сложение в алгебре — это просто обычное сложение функций:

$$(f+g)(x) = f(x) + g(x).$$

2. Роль умножения играет свертка по эйлеровой характеристике $f \otimes g$.
3. Умножать функцию на целое число нужно обычным способом:

$$(a \cdot f)(x) = a \cdot f(x).$$

А вот как надо умножать на рациональное число, объясним чуть позже — это и есть самое интересное и нетривиальное.

В нашей алгебре есть нейтральный элемент $\mathbf{1}$ — характеристическая функция начала координат.

Задача 15. Проверьте, что введенные операции удовлетворяют аксиомам алгебры:

1. $f \otimes (g+h) = f \otimes g + f \otimes h$.
2. $f \otimes g = g \otimes f$.

Подсказка: достаточно убедиться в справедливости аксиом для случая, когда f и g — характеристические функции выпуклых многогранников.

Наконец, важный завершающий шаг построения алгебры — факторизация по параллельным переносам. Условимся считать в нашей алгебре два многогранника (= их характеристические функции) равными, если они отличаются на параллельный перенос.

Мы факторизуем кольцо многогранных функций по всем соотношениям вида $K - tK = 0$, где K — выпуклый многогранник, а t — параллельный перенос.

Заметим, что при этом автоматически равными между собой становятся многие многогранные функции. Каждую многогранную функцию можно разрезать на куски, затем параллельно перенести каждый кусок (для разных кусков можно использовать разные параллельные переносы!), и полученная многогранная функция окажется равной исходной.

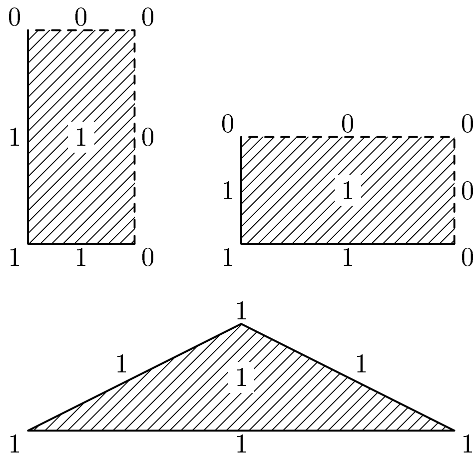


Рис. 10.

Задача 16. Какие из многогранных функций, изображенных на рис. 10, равны между собой в алгебре многогранников?

Здесь уместно сделать отступление. Идея разрезания многогранников на куски и складывания из кусков чего-то нового знакома тем, кто знает о знаменитой третьей проблеме Гильберта.

Третья проблема Гильберта отличается от того, что происходит в алгебре многогранников:

во-первых, куски разрешается передвигать, пользуясь любыми движениями плоскости, а не только параллельными переносами,

во-вторых, в алгебре многогранников учитываются куски всех размерностей (точки и отрезки не равны нулю), тогда как в третьей проблеме Гильберта ими пренебрегают.

После факторизации по параллельным переносам в алгебре многогранников стало трудно различать функции: после разрезания на куски и перемешивания картинка может измениться неузнаваемо.

Для распознавания равных многогранных функций на плоскости введем следующие инварианты.

Два из них простые.

Положим
$$h_0(f) = \int f(x)\, d\chi(x)$$
и
$$h_2(f) = \int f(x)\, dx.$$
Эти величины называются *нулевым и вторым инвариантом Хадвигера* многогранной функции f.

Во втором интеграле интегрирование обычное, по мере Лебега. Величину $h_2(f)$ естественно называть площадью функции f.

Введем более сложный инвариант, вернее целое семейство таковых. При фиксированном единичном векторе ξ определим *длину ребра с нормалью* ξ многогранной функции в два приема:

1. Пусть K — выпуклый многогранник. Положим по определению $h_1(K, \xi)$ равным длине ребра K, внешняя нормаль которого равна ξ.

2. Произвольную многогранную функцию f разложим на выпуклые многогранники $f = \sum_i a_i K_i$ и положим $h_1(f, \xi) = \sum_i a_i h_1(K_i, \xi)$.

Функция h_1 называется *первым инвариантом Хадвигера*.

Задача 17*. Докажите, что это определение корректно: значение h_1 не зависит от представления функции f в виде линейной комбинации многогранников.

Теорема 4. *Система инвариантов Хадвигера полна, т. е. две многогранные функции f и g, заданные на $\mathbb{R}^2$, равны (после факторизации) тогда и только тогда, когда*

$$\begin{aligned} &h_0(f) = h_0(g), \\ &h_2(f) = h_2(g), \\ &h_1(f, \xi) = h_1(g, \xi), \quad \textit{для любого } \xi. \end{aligned}$$

Задача 18*. Докажите эту теорему.

В пространствах большей размерности ситуация аналогичная: там тоже определена система инвариантов Хадвигера, являющаяся полной. При этом инвариант h_k также «отвечает» за грани размерности h_k многогранной функции, но роль параметра ξ играет не единичный вектор, а флаг.

Задача 19. Вычислите инварианты Хадвигера многогранных функций, изображенных на рисунке 10.

9. Пытаемся умножать на дроби

Чтобы множество многогранных функций превратить в алгебру над $\mathbb{Q}$, надо задать операцию умножения многогранных функций на рациональные числа.

При этом надо обеспечить выполнение аксиом дистрибутивности

$$q \cdot (f + g) = q \cdot f + q \cdot g$$

и ассоциативности

$$(qr) \cdot f = q \cdot (r \cdot f).$$

Как уже говорилось, умножить многогранную функцию на целое число легко — нужно просто использовать поточечное умножение:

$$(n \cdot f)(x) = nf(x).$$

С дробями дело обстоит сложнее. Пусть f — многогранная функция.

Попробуем определить многогранную функцию $1/n \cdot f$ для натурального числа n (этого достаточно для умения умножать на любые рациональные числа).

Мы должны найти такую многогранную функцию g, что в нашей алгебре выполнено равенство $n \cdot g = f$.

Просто умножить f на $1/n$ нельзя: у многогранных функций по определению все значения целые.

Более того, искомая функция g может вообще не существовать: заметим, что

$$\int f(x)\, d\chi(x) = n \int g(x)\, d\chi(x),$$

т. е. $\int f(x)\, d\chi(x)$ кратно n, что не всегда верно.

Значит, у нас есть шанс научиться умножать на произвольные дроби лишь те функции, у которых $\int f(x)\, d\chi(x) = 0$.

Начнем с примеров.

Задача 20. Найти $1/2 \cdot f$, где f — характеристическая функция
– полуоткрытого отрезка;
– квадрата с выкинутыми двумя смежными сторонами;
– треугольника с выкинутой вершиной;
– тетраэдра с выкинутой вершиной.

Задача 21. Найти $1/n \cdot f$, где f — характеристическая функция
– полуоткрытого отрезка;
– параллелограмма с выкинутыми двумя смежными сторонами;
– треугольника с выкинутой вершиной;
– тетраэдра с выкинутой вершиной.

Чтобы умножить на $1/n$ произвольную многогранную функцию f, для которой $\int f(x)\, d\chi(x) = 0$, надо проделать следующее.

- ▷ Представить f в виде линейной комбинации полуоткрытых отрезков, треугольников с выкинутой вершиной, и тетраэдров с выкинутой вершиной.
- ▷ Умножить каждое из слагаемых на $1/n$.
- ▷ Воспользоваться дистрибутивностью.

Таким образом, на множестве многогранных функций мы почти ввели структуру алгебры над полем рациональных чисел. Слово «почти» относится к тому факту, что функции с ненулевым значением $\int f(x)\, d\chi(x)$ нельзя умножить на произвольную дробь.

10. Логарифм и экспонента

На первый взгляд, это совершенно дикая нереализуемая идея: мы с большим трудом справились с умножением на рациональные числа, а здесь участвует трансцендентное число e, и вообще неясно, как возвести число в степень «треугольник».

На самом деле введение логарифма и экспоненты — хороший пример того, что часто общее бывает проще частного. Этот прием будет работать не только в алгебре многогранников, но и в любой другой градуированной алгебре с обрывающейся градуировкой.

Пусть f — многогранная функция такая, что $\int f(x)\,d\chi(x) = 0$. Определим логарифм $f+1$, минуя число e. Из курса анализа известно, что при фиксированном x степенной ряд

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}$$

сходится к $\ln(x+1)$. (Читателю можно не вспоминать подробности вопросов сходимости рядов, нам важен лишь его вид).

Экспонента тоже представима в виде ряда:

$$\exp x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Что если попытаться подставить в эти ряды функцию f? Мы уже умеем перемножать многогранные функции и складывать их, но как быть с переходом к пределу?

А здесь нас ждет подарок судьбы (доказательство мы опустим, подарок так подарок):

Предложение 4. *Для выпуклого многогранника K*

$$(K-1)^4 = 0.$$

Значит, вместо бесконечного ряда у нас останется простая конечная сумма.

Определение 11. Пусть f — многогранная функция такая, что

$$\int f(x)\,d\chi(x) = 0.$$

Положим по определению

$$\ln(f+1) \stackrel{\text{def}}{=} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{f^n}{n}, \qquad \exp f \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} \frac{f^n}{n!}\,.$$

Введенные нами логарифм и экспонента обладают знакомыми со школы свойствами:

Теорема 5. *Пусть многогранные функции f и g таковы, что*

$$\int f(x)\,d\chi(x) = \int g(x)\,d\chi(x) = 0.$$

Тогда

1. $\ln((f+1)\otimes(g+1)) = \ln(f+1) + \ln(g+1)$.
2. $\exp(f+g) = \exp f \otimes \exp g$.
3. $\ln(\exp f) = f$.
4. $\exp(0) = \mathbf{1}$.

Задача 22. Докажите эту теорему. Это надо делать в лоб, как в пятом классе на уроках алгебры: прямо подставить степенные ряды, раскрыть скобки, привести подобные члены и не забыть при этом, что большие степени обнуляются.

11. Градуировка в алгебре многогранников

Мы покажем, что построенная нами алгебра похожа на алгебру многочленов. Подобно тому, как многочлен естественным образом представим в виде суммы одночленов разных степеней, многогранные функции представимы в виде суммы так называемых однородных элементов. При этом однородные многогранные функции, так же как и однородные многочлены, «ловятся» гомотетиями.

Начнем с элементарного замечания. Пусть $f(x)$ — многочлен. Рассмотрим новый многочлен $(2)f(x)$, полученный гомотетичным сжатием оси x:

$$(2)f(x) = f(2x)$$

При этом свободный член нового многочлена остается прежним, коэффициент при x увеличивается в 2 раза, коэффициент при x^2 увеличивается в 4 раза, и так далее.

Поэтому одночлены среди многочленов легко распознать с помощью растягивания гомотетией оси x: например, одночлен третьей степени — это такой многочлен, который при действии (2) увеличивается в 8 раз.

Иными словами, справедлива очевидная теорема:

Теорема 6. *В алгебре многочленов $\mathcal{F}$ от одной переменной можно выделить (непересекающиеся) подмножества $\mathcal{F}_0$, $\mathcal{F}_1$, $\mathcal{F}_2$, ... такие, что*

1. *Каждый многочлен f единственным образом представим в виде $f = \sum_i f_i$, $f_i \in \mathcal{F}_i$.*
2. *Если $f_i \in \mathcal{F}_i$, $f_j \in \mathcal{F}_j$, то $f_i \cdot f_j \in \mathcal{F}_{i+j}$.*
3. *Действие (2) увеличивает элементы $\mathcal{F}_i$ в 2^i раз.*

(В таких случаях говорят, что на множестве многочленов есть структура градуированной алгебры.)

Множество $\mathcal{F}_i$ называются однородными компонентами степени i алгебры многочленов.

На множестве многогранных функций есть аналогичное действие гомотетии: положим $(2)f(x) = f(x/2)$. Замечательный и нетривиальный факт состоит в том, что для алгебры многогранников также имеет место аналог теоремы об однородных компонентах.

Теорема 7. *В алгебре многогранников $\mathcal{M}$ можно выделить (непересекающиеся) подмножества $\mathcal{M}_i$, $i = 0, 1, 2, 3$ такие, что*

1. *Каждая многогранная функция f единственным образом представима в виде $f = \sum_i f_i$; $f_i \in \mathcal{M}_i$.*
2. *Если $f_i \in \mathcal{M}_i$, $f_j \in \mathcal{M}_j$, то $f_i \cdot f_j \in \mathcal{M}_{i+j}$.*
3. *Действие (2) увеличивает многогранную функцию $f_i \in \mathcal{M}_i$ в 2^i раза: $f \in \mathcal{M}_i$ равносильно тому, что $(2)f = 2^i f$.*

Множество $\mathcal{M}_i$ называется однородной компонентой алгебры многогранников степени i.

Элементы $\mathcal{M}_i$ являются аналогами одночленов степени i.

Доказательство. Попробуем понять, что из себя представляют однородные элементы алгебры многогранников (аналоги одночленов в алгебре многочленов). Пусть K — выпуклый многогранник. Заметим, что

$$(2)\ln K = 2 \cdot \ln K.$$

Действительно,

$$2 \cdot \ln K = \ln(K \otimes K) = \ln((2)K) = (2)\ln K.$$

Задача 23. Докажите, что

$$(2)(\ln K \otimes \ln K) = 4 \cdot (\ln K \otimes \ln K)$$

и что

$$(2)((\ln K)^i) = 2^i \cdot ((\ln K)^i)$$

Наконец отметим, что $(2)\mathbf{1} = \mathbf{1}$. (Действительно, точка при растягивании не меняется).

Поэтому естественно задать $\mathcal{M}_i$ как множество всех линейных комбинаций функций вида $(\ln K)^i$, где K — выпуклый многогранник.

Нам надо научиться раскладывать многогранную функцию f в сумму однородных членов. Для этого достаточно проделать это для выпуклого многогранника K.

Положим $p = \ln K$. Тогда

$$K = \exp(\ln K) = \sum_{i=0}^{4} \frac{p^i}{i!}.$$

Это и есть искомое разложение на однородные компоненты.

Приложение. Ответы и указания

Задача 1. Сумма Минковского есть отрезок суммарной длины.

Задача 2. Параллелограмм.

Задача 4. Пятиугольник.

Задачи 6 и 3. Сумма останется той же с точностью до параллельного переноса.

Задача 11. Веер тетраэдра изображен на рис. 11.

Веер отрезка — разбиение сферы большим кругом (ортогональным отрезку) на две полусферы.

Веер куба — разбиение сферы тремя большими кругами. В этом можно убедиться, пользуясь алгоритмом, но проще заметить, что куб есть сумма Минковского трех отрезков, веера которых мы уже знаем.

Задача 14. Виртуальный многогранник существует только для второй поверхности, см. рис. 12.

Задача 16. Первые две.

Задача 20.

– Полуоткрытый отрезок равен сумме двух полуоткрытых отрезков половинной длины.

– Такой полуоткрытый параллелограмм легко разбить на две равные части разрезом, параллельным одной из сторон.

– Представьте этот полуоткрытый треугольник как объединение двух (гомотетично в 2 раза меньших) полуоткрытых треугольников и параллелограмма.

– Аналогично предыдущему пункту, представьте тетраэдр в виде объединения двух (гомотетично в 2 раза меньших) тетраэдров и двух призм, с треугольным основанием и параллелограммом в основании. Призмы режутся пополам легко.

Задача 21. См. указания к предыдущей задаче.

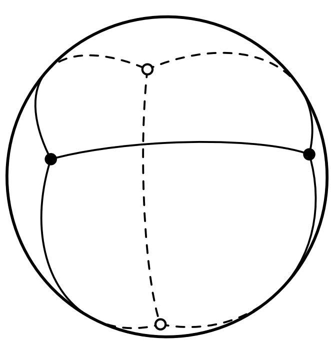


Рис. 11.

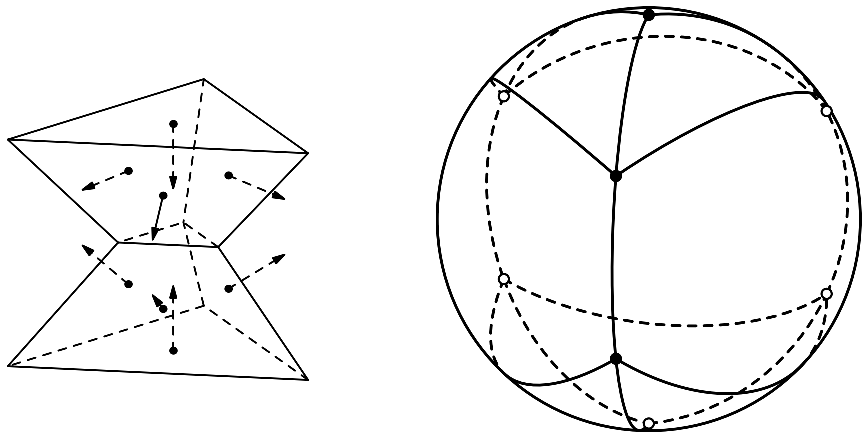


Рис. 12.

Список литературы

[1] Пухликов А., Хованский А. *Конечно-аддитивные меры виртуальных многогранников* // Алгебра и Анализ, 1992. Т. 4, №2. С. 161–185.

[2] McMullen P. *The polytope algebra* // Adv. Math., 1989. V. 78, no 1. P. 76–130.

[3] Panina G. *New counterexamples to A.D. Alexandrov's hypothesis* // Adv. in Geometry, 2005. V. 5. P. 301–317.

Панина Гаянэ Юрьевна, Санкт-Петербургский институт информатики и автоматизации РАН. С. Петербург 199178, 14 линия В.О. 39
факс (812)3284450
д. тел. (812)5504571
e-mail: `panina@iias.spb.su`

О сумме углов многогранника

И. В. Изместьев*

Преподнося сюрприз
Суммой своих углов,
Вещь выпадает из
Нашего мира слов.

Иосиф Бродский,
Натюрморт VII

1. Введение

В этой статье мы докажем ряд соотношений, связывающих величины внутренних углов многогранника. Самое известное из них (к сожалению, менее известное, чем оно того заслуживает) — формула Грама. Для трехмерного многогранника она гласит:

Теорема 1.

$$2\sum_e \alpha_e - \sum_v \alpha_v = 2\pi(F-2), \qquad (1)$$

где α_e — двугранный угол при ребре e, α_v — телесный угол при вершине v, F — количество граней многогранника, суммирование ведется по всем вершинам и ребрам.

Например, куб имеет 12 ребер с углами по $\pi/2$, 8 вершин с углами также по $\pi/2$ и 6 граней. Подстановка этих величин в формулу (1) дает верное тождество.

Формула Грама обобщается на многогранники более высокой размерности. Для этого нам надо определить величину угла при каждой грани многогранника. Можно поступить традиционно: если плоский (и вместе с ним двугранный) угол измеряется длиной дуги, а телесный площадью области, высекаемой на единичной сфере, то полный k-мерный угол следует положить равным площади единичной k-мерной сферы. Однако иногда удобно углы нормировать, приняв величину полного угла за единицу. В нашем случае это приводит к элегантному равенству.

*Работа выполнена в рамках проекта «Полиэдральные поверхности», финансируемого DFG (Германским исследовательским обществом).

Теорема 2 (Формула Грама).

$$\sum_{k=0}^{n}(-1)^k\varphi_k = 0, \tag{2}$$

где φ_k — сумма (нормированных) величин углов при k-мерных гранях данного n-мерного многогранника.

Дадим более строгое определение величины угла. Пусть P — n-мерный многогранник, Q — его k-мерная грань. Выберем точку x внутри Q и рассмотрим шар B малого радиуса с центром в x. Величина угла многогранника P при грани Q полагается равной доле объема шара B, находящейся внутри многогранника P:

$$\varphi_Q(P) = \frac{\operatorname{vol}(B\cap P)}{\operatorname{vol}(B)}\,.$$

Эквивалентным образом можно рассмотреть малую сферу с центром в x и определить, какая доля ее площади находится внутри P.

Данное определение проясняет смысл слагаемых φ_{n-1} и φ_n в формуле (2). Следуя ему, для всех углов при $(n-1)$-мерных гранях мы получаем величину $1/2$. Кроме того, мы уславливаемся считать сам многогранник своей (единственной) n-мерной гранью, откуда по определению находим $\varphi_n = 1$. Теперь читатель без труда может установить, что формула (1) является частным случаем формулы (2).

Кроме формулы Грама, мы докажем и другие соотношения между величинами углов многогранника, формулы Дена – Соммервиля. Они появляются в размерностях начиная с 4, и мы отложим их формулировку до более удобного момента.

* * * * * *

Для полноты изложения дадим определения многогранника и его граней. Ниже приведены также некоторые соглашения, используемые в дальнейшем.

Определение 1. *Выпуклым многогранником* в $\mathbb{R}^n$ называется ограниченное пересечение конечного числа полупространств. Многогранник называется *n-мерным*, если он не содержится в подпространстве $\mathbb{R}^m \subset \mathbb{R}^n$ меньшей размерности.

Все многогранники в статье будут предполагаться выпуклыми. Как правило, мы также предполагаем, что размерность многогранника совпадает с размерностью объемлющего пространства. Это важно, например, при определении величины угла при грани.

Определение 2. *Опорной плоскостью* многогранника $P \subset \mathbb{R}^n$ называется гиперплоскость (подпространство размерности $n-1$), имеющая непустое пересечение с P и такая, что P лежит по одну сторону от нее. *Гранью* многогранника P называется пересечение P с любой его опорной плоскостью. Иногда к граням относят также сам многогранник P и/или пустое множество.

Каждая грань многогранника сама является многогранником и имеет определенную размерность. (Размерность пустого множества удобно положить равной -1.) Грани размерности 0 — это вершины. Грани размерности $n-1$ в n-мерном многограннике называются его фасетами. Многогранник может быть представлен как пересечение полупространств, граничные гиперплоскости которых задаются фасетами многогранника. Также многогранник является выпуклой оболочкой своих вершин. «Самый маленький» n-мерный многогранник в $\mathbb{R}^n$ — это симплекс, имеющий $n+1$ фасету и $n+1$ вершину. Если мы рассмотрим выпуклую оболочку конечного числа точек в общем положении, то все грани полученного многогранника, за исключением, быть может, его самого, будут симплексами (пример: октаэдр, с пошевеленными вершинами, если мы настаиваем на общем положении). Такой многогранник называется симплициальным. О симплициальных многогранниках пойдет речь в конце этой статьи.

Мы будем использовать без доказательства формулу Эйлера для n-мерного многогранника. Если мы обозначим через f_k, $k = 0, 1, \ldots, n$ количество k-мерных граней данного n-мерного многогранника P, то эта формула гласит:

$$\sum_{k=0}^{n} (-1)^k f_k = 1. \tag{3}$$

Поскольку $f_n = 1$ (многогранник является своей единственной n-мерной гранью), эту формулу часто записывают в виде

$$\sum_{k=0}^{n-1} (-1)^k f_k = 1 + (-1)^{n-1}.$$

В одном из доказательств (доказательство Шепарда в параграфе 3) мы применим формулу Эйлера в более общей ситуации, когда рассматривается разбиение многогранника на более мелкие многогранники, и f_k обозначает количество k-мерных частей разбиения.

Более близко с восхитительным и до сих пор загадочным миром многогранников читатель может познакомиться по книге [8].

2. Формула Грама для трехмерного многогранника

В этом параграфе мы докажем теорему 1. Для этого нам понадобится вспомогательное утверждение:

Лемма 1. *Для сферического треугольника Δ на сфере единичного радиуса выполняется равенство*

$$S(\Delta) = \alpha + \beta + \gamma - \pi, \tag{4}$$

где $S(\Delta)$ — площадь треугольника, а α, β и γ — величины его углов.

Доказательство. По определению, сферический треугольник ограничен дугами больших окружностей. Иначе говоря, Δ является пересечением трех полусфер таких, что их граничные окружности не пересекаются в одной точке. Пусть H_1, H_2, H_3 — эти полусферы. Тогда имеем

$$H_1 \cap H_2 \cap H_3 = \Delta,$$
$$H_1 \cup H_2 \cup H_3 = \mathbb{S}^2 \setminus \Delta'.$$

Здесь $\mathbb{S}^2$ — сфера, Δ' — треугольник, центрально-симметричный треугольнику Δ относительно центра сферы (рис. 1). Рассматривая части, на которые делят сферу границы областей H_1, H_2 и H_3, нетрудно доказать равенство

$$\begin{aligned} S(H_1 \cup H_2 \cup H_3) \quad = \quad & S(H_1) + S(H_2) + S(H_3) - S(H_1 \cap H_2) - \\ & - S(H_2 \cap H_3) - S(H_3 \cap H_1) + S(H_1 \cap H_2 \cap H_3) \end{aligned}$$

(формула включения-исключения). Однако $S(H_i) = 2\pi$ для $i = 1, 2, 3$, а $S(H_i \cap H_j)$ равняется удвоенному углу треугольника Δ при вершине,

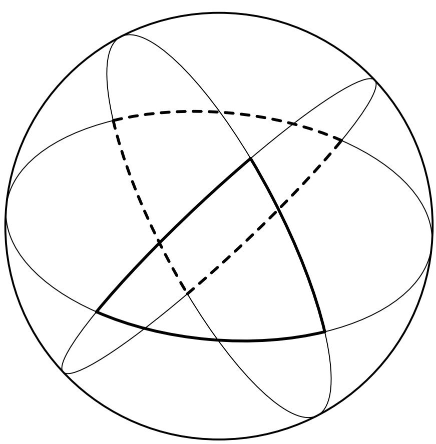


Рис. 1. *Сферический треугольник и его антипод*

лежащей на границе полусфер H_i и H_j. Подставляя эти и полученные выше равенства, получаем

$$4\pi - S(\Delta') = 3 \cdot 2\pi - 2\alpha - 2\beta - 2\gamma + S(\Delta).$$

Поскольку $S(\Delta') = S(\Delta)$, отсюда вытекает равенство (4). □

Задача 1. Докажите, что площадь выпуклого сферического n-угольника M на единичной сфере вычисляется по формуле

$$S(M) = \sum_{i=1}^{n} \alpha_i - (n-2)\pi,$$

где α_i, $i = 1, \ldots, n$ — величины углов многоугольника. (*Подсказка:* Разрежьте многоугольник на треугольники.)

Доказательство теоремы 1. Для каждой вершины v многогранника возьмем сферу малого радиуса с центром в точке v. Пересечение этой сферы с многогранником является сферическим многоугольником. Увеличим сферу так, чтобы ее радиус стал равным единице и обозначим полученный сферический многоугольник через M_v. По задаче 1 имеем

$$S(M_v) = \sum_{i=1}^{n_v} \alpha_i(v) - (n_v - 2)\pi,$$

где n_v — количество сторон многоугольника M_v, а $\alpha_i(v)$ — величины его углов. При этом $\alpha_i(v)$ равны, как нетрудно видеть, величинам двугранных углов при ребрах, исходящих из вершины v, а $S(M_v)$ равняется по определению величине α_v телесного угла при v. Просуммировав все полученные равенства, получаем:

$$\sum_v \alpha_v = \sum_v \sum_{e \supset v} \alpha_e - \pi \sum_v (n_v - 2).$$

Поскольку каждое ребро e содержит две вершины, двугранный угол α_e появляется в правой части равенства ровно два раза. Точно так же имеем $\sum_v n_v = 2E$, где E — число ребер многогранника. Таким образом, равенство переписывается в виде

$$\sum_v \alpha_v = 2 \sum_e \alpha_e - 2\pi(E - V),$$

где V обозначает число вершин многогранника. Принимая во внимание формулу Эйлера $V - E + F = 2$, мы приходим к равенству (1). □

Доказательство леммы 1 может быть обобщено на более высокие размерности, как это предлагается сделать читателю в следующей задаче.

Задача 2. Пусть $\mathbb{S}^n$ — единичная сфера размерности n (множество точек на расстоянии 1 от данной точки в $(n+1)$-мерном пространстве).

Сферическим симплексом называется пересечение $(n+1)$-й полусферы такое, что ни одна точка сферы не является их общей граничной точкой. Пусть Δ — симплекс на $\mathbb{S}^n$. Докажите равенство

$$\sum_{k=0}^{n}(-1)^k\varphi_k = (1+(-1)^n)\frac{\mathrm{vol}(\Delta)}{\mathrm{vol}(\mathbb{S}^n)}\,. \tag{5}$$

Здесь vol обозначает объем[1)], φ_k — сумма нормированных величин углов при k-мерных гранях симплекса Δ. Углы при гранях сферического многогранника измеряются аналогично углам евклидова многогранника, смотри введение.

Формула (5) называется формулой Грама для сферического симплекса. Заметьте, что в нечетных размерностях правая часть в ней равна 0. Таким образом, данная формула не позволяет выразить объем трехмерного сферического симплекса через величины его углов.

Попытка обобщить данное выше доказательство формулы Грама на высшие размерности наталкивается на серьезные трудности (осознайте, какие!). Поэтому мы должны пойти другим путем.

3. Два доказательства формулы Грама в высших размерностях

Доказательство Соммервиля

Первое доказательство формулы (2) для выпуклого n-мерного многогранника P при произвольном n было предложено Соммервилем в [6]. Оно состоит из двух этапов.

Лемма 2. *Формула Грама верна для симплекса.*

Доказательство. Заметим, что формула Грама для сферического симплекса (5) верна на сфере произвольного радиуса: при гомотетии величины углов симплекса не меняются, отношение объема симплекса к объему сферы также остается неизменным. Идея доказательства состоит в том, чтобы рассмотреть данный евклидов симплекс как сферический на сфере бесконечного радиуса. Объем такой сферы бесконечен, поэтому правая часть равенства (5) обращается в 0.

Более строго: поместим n-мерное пространство $\mathbb{R}^n$ вместе с лежащим в нем евклидовым симплексом $\Delta_{\mathbb{R}}$ в $(n+1)$-мерное пространство $\mathbb{R}^{n+1}$. Рассмотрим сферу радиуса R в $\mathbb{R}^{n+1}$, касающуюся $\mathbb{R}^n$ вблизи симплекса, скажем, в одной из его точек. Спроецируем симплекс на сферу при

[1)] Для подмножества сферы $\mathbb{S}^n$ мы говорим о его объеме, а не о площади, поскольку все рассматриваемые здесь множества, и прежде всего симплекс Δ, содержатся в $\mathbb{S}^n$; пространство $\mathbb{R}^{n+1}$ играет вспомогательную роль.

помощи проекции из ее центра. Гиперплоскости в $\mathbb{R}^n$ перейдут при проекции в большие гиперсферы (точнее, в их половины), поэтому образом $\Delta_{\mathbb{R}}$ будет сферический симплекс, который мы обозначим $\Delta_{\mathbb{S}}$. Теперь будем увеличивать радиус сферы, следя при этом, чтобы она по-прежнему касалась пространства $\mathbb{R}^n$ в одной из точек симплекса. Ясно, что при этом величины углов симплекса $\Delta_{\mathbb{S}}$ стремятся к величинам соответствующих углов симплекса $\Delta_{\mathbb{R}}$, объем $\Delta_{\mathbb{S}}$ стремится к объему $\Delta_{\mathbb{R}}$, а объем сферы стремится к бесконечности. Поэтому в пределе формула (5) переходит в формулу (2). □

Второй этап доказательства состоит в разбиении многогранника P на симплексы и выводе формулы Грама для целого P из формул для его частей. То специальное подразделение, которое мы будем рассматривать, получается путем последовательного разбиения граней многогранника.

Пусть Q — грань многогранника P. Выберем внутри Q произвольную точку c_Q и построим пирамиды с вершиной c_Q над всеми гранями, содержащимися в Q. Эта операция называется *звездным разбиением* многогранника P в грани Q. Гранями получающегося подразделения называются все грани многогранника P, отличные от Q (и, таким образом, не затронутые разбиением), а также все получившиеся пирамиды и грани этих пирамид.

Будем осуществлять звездные разбиения в гранях многогранника P в порядке возрастания размерности: сначала подразделим по очереди все ребра, затем все двумерные грани, и т. д. В конце произведем разбиение в самом P, выбрав точку c_P в его внутренности. Рисунок 2 иллюстрирует получающиеся разбиения в трехмерном случае (здесь сначала разбиты одновременно все ребра, затем все грани, последний шаг с точкой c_P опущен). После каждого шага мы получаем новое, более мелкое подразделение P. Звездное разбиение в грани текущего подразделения определяется

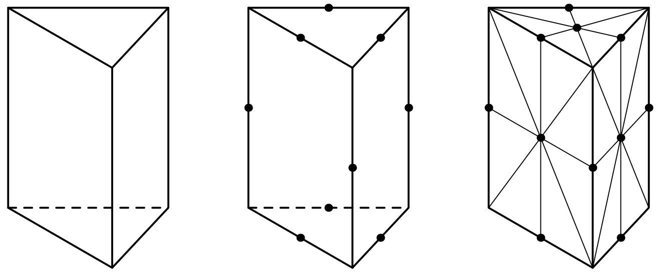


Рис. 2. *Барицентрическое подразделение как результат последовательных звездных разбиений*

так же, как и выше в случае многогранника. Нормированная величина угла при грани подразделения определяется тоже аналогично величине угла при грани многогранника.

Покажем, что альтернированная сумма величин углов при гранях не меняется при звездном разбиении. Пусть φ_k — сумма величин углов при k-мерных гранях подразделения, полученного в результате нескольких первых шагов, φ'_k — аналогичная сумма после звездного разбиения в очередной грани Q. Обозначим через $f_k(Q)$ количество k-мерных граней текущего подразделения, содержащихся в грани Q, и через α — угол многогранника P при грани Q. Тогда имеют место равенства:

$$\begin{aligned} &\varphi'_l = \varphi_l && \text{при } l > k, && \varphi'_k = \varphi_k + \alpha(f_{k-1}(Q) - 1),\\ &\varphi'_l = \varphi_l + \alpha f_{l-1}(Q) && \text{при } k > l > 0, && \varphi'_0 = \varphi_0 + \alpha. \end{aligned}$$

Суммируя с чередующимися знаками, получаем равенство

$$\sum_{k=0}^{n}(-1)^k\varphi'_k = \sum_{k=0}^{n}(-1)^k\varphi_k + \\ + \alpha(1 - f_0(Q) + \dots + (-1)^{k-1}f_{k-2}(Q) + (-1)^k(f_{k-1}(Q) - 1)).$$

Так как $f_0(Q) = 1$, из формулы Эйлера для многогранника Q следует, что коэффициент при α равен 0. Таким образом,

$$\sum_{k=0}^{n}(-1)^k\varphi'_k = \sum_{k=0}^{n}(-1)^k\varphi_k\,.$$

Теперь подсчитаем сумму $\sum_{k=0}^{n}(-1)^k\varphi_k$ после последнего разбиения. Нетрудно доказать индукцией по размерности n многогранника P, что n-мерные грани получившегося подразделения — симплексы с наборами вершин $(c_{Q^0}, c_{Q^1}, \dots, c_{Q^n})$, где Q^k — одна из k-мерных граней многогранника P и $Q^k \subset Q^{k+1}$ (такое подразделение называется барицентрическим, если в качестве точки c_Q выбирается центр масс грани Q). Каждая грань подразделения является гранью одного или нескольких из таких симплексов. Кроме того, ясно, что угол при каждой грани равен сумме углов при ней всех симплексов, которым она принадлежит. Поэтому имеет место равенство

$$\sum_{k=0}^{n}(-1)^k\varphi_k = \sum_{\Delta}\sum_{k=0}^{n}(-1)^k\varphi_k^{\Delta}\,,$$

где суммирование ведется по всем n-мерным симплексам подразделения, а φ_k^{Δ} обозначает сумму углов при k-мерных гранях симплекса Δ. Поскольку каждая из внутренних сумм равна 0 по ранее доказанному, имеем $\sum_{k=0}^{n}(-1)^k\varphi_k = 0$ для получившегося подразделения, а значит и для многогранника P. Формула Грама доказана.

* * * * * *

Через 40 лет после публикации доказательство Соммервиля было несправедливо сочтено ошибочным. Вероятно, причиной послужило то, что Соммервиль употреблял термин «многогранник» для подразделений, возникающих в процессе доказательства, и это не удовлетворяло повысившимся стандартам строгости. В результате появились новые доказательства формулы Грама. Одно из них, опубликованное Шепардом в [7] отличается особой элегантностью, и мы его здесь приведем.[2)]

Доказательство Шепарда

Вернемся к определению величины угла $\varphi_Q(P)$ многогранника P при грани Q. Грани Q можно сопоставить конус L_Q, ограниченный гиперплоскостями фасет, содержащих Q. Конус L_P — это все пространство $\mathbb{R}^n$, конус фасеты — полупространство. Если c_Q — произвольная точка внутри грани Q, то мы имеем

$$\varphi_Q(P) = \frac{\mathrm{vol}(B \cap L_Q)}{\mathrm{vol}(B)},$$

где B — шар произвольного радиуса с центром в c_Q. Выбрав по точке c_Q в каждой грани Q, перенесем каждый конус L_Q так, чтобы точка c_Q попала в начало координат. Тогда формула Грама равносильна равенству

$$\sum_Q (-1)^{\dim Q} \mathrm{vol}(B \cap (L_Q - c_Q)) = 0, \tag{6}$$

где B на этот раз обозначает шар с центром в начале координат. Для доказательства последнего равенства достаточно показать, что для (почти) любого луча, исходящего из начала координат, сумма ± 1, соответствующих тем конусам, внутри которых содержится этот луч, равна 0. Иными словами, покрытие шара B конусами $L_Q - c_Q$, взятыми с соответствующими знаками, имеет в почти любой точке кратность 0.

Рассмотрим луч x, не параллельный ни одной из фасет многогранника P. Точки шара B, не лежащие на таких лучах, образуют множество нулевого объема, поэтому достаточно рассмотреть лучи, удовлетворяющие этому условию. Спроецируем многогранник P в направлении x на плоскость, ортогональную x. Обозначим проекцию через P^x. Нетрудно видеть, что луч x содержится в конусе $L_Q - c_Q$ в том и только в том случае, если $Q = P$ или если грань Q «освещена» лучами, параллельными x (грани, проекции которых лежат на границе P^x, освещенными не считаются). Рисунок 3 иллюстрирует ситуацию в трехмерном случае.

[2)] Доказательство Шепарда содержится также в переводе на английский язык книги В. В. Прасолова и В. М. Тихомирова «Геометрия» [5]. Готовится ее русское переиздание.

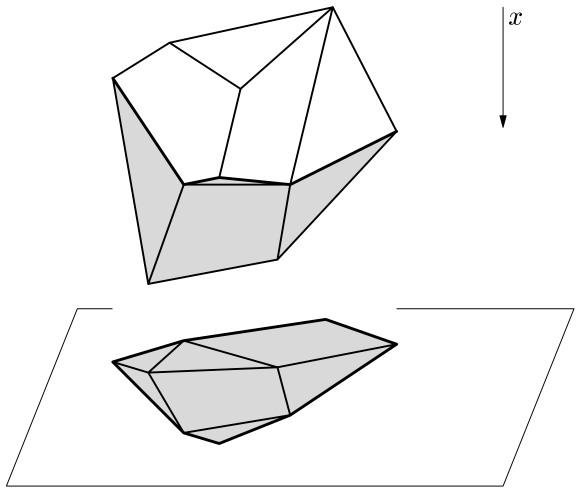


Рис. 3. *Свет и тень. Ребра, по которым свет «скользит», считаются затененными*

Обозначим через f_k^x количество освещенных k-мерных граней, положив $f_n^x = 1$. Требуется доказать равенство $\sum_{k=0}^n (-1)^k f_k^x = 0$.

Для этого рассмотрим затененные грани. В размерности k их количество равно $f_k - f_k^x$. В то же время их проекции образуют разбиение $(n-1)$-мерного многогранника P^x. Поэтому по формуле Эйлера имеем

$$\sum_{k=0}^{n} (-1)^k (f_k - f_k^x) = 1.$$

Вычитая это равенство из формулы Эйлера $\sum_{k=0}^n (-1)^k f_k = 1$ для многогранника P, приходим к требуемому утверждению. Таким образом, равенство (6) доказано, а вместе с ним и формула Грама.

4. Соотношения Дена – Соммервиля

Вернемся к сферическому симплексу размерности n:

$$\Delta = \cap_{i=0}^n H_i, \tag{7}$$

где $H_0, H_1, \ldots, H_n$ — полусферы n-мерной сферы $\mathbb{S}^n$. Формула Грама (5) для сферического симплекса является следствием равенства объемов симплекса Δ и симметричного ему симплекса Δ' (см. доказательство леммы 1). Однако границы полусфер H_i делят сферу на множество других областей. При этом нетрудно заметить, что суммарный объем частей, k-кратно покрываемых полусферами H_i, равен суммарному объему частей, покрытых $n+1-k$ раз. Эти равенства приводят к целому набору новых соотношений между углами симплекса, называемых (метрическими) соотношениями Дена – Соммервиля. Кроме того, соотношения Дена – Соммервиля распространяются на величины углов симплициального многогранника. Но наибольшую известность они получили как соотношения между количеством граней данной размерности многогранника такого вида. Обо всем этом и пойдет речь в данном параграфе.

Обозначим через S_i граничную сферу полусферы H_i. Грани симплекса (7) допускают следующее описание. Пусть I — подмножество множества индексов $\{0, 1, \ldots, n\}$, не совпадающее со всем множеством. Положим

$$F_I = (\cap_{i \in I} S_i) \cap \Delta.$$

Тогда F_I — грань симплекса Δ, имеющая размерность $n - |I|$. Грани F_I мы сопоставим ее *лунку*

$$H_I = \cap_{i \in I} H_i$$

(сравните с определением конуса грани евклидова многогранника в предыдущем параграфе). Нетрудно доказать, что отношение объема лунки к объему сферы есть величина угла при соответствующей грани. Таким образом, имеем

$$\varphi_l = \frac{1}{\mathrm{vol}(\mathbb{S}^n)} \sum_{|I| = n-l} \mathrm{vol}(H_I).$$

Рассмотрим теперь области, на которые разбивают сферу $\mathbb{S}^n$ ее подсферы S_i. Каждому множеству $J \subset \{0, 1, \ldots, n\}$ отвечает область

$$G_J = (\cap_{j \in J} H_j) \cap (\cap_{j \notin J} \overline{H_j}),$$

где $\overline{H_j}$ обозначает полусферу, противоположную H_j. Ясно, что симметрия относительно центра сферы переводит область G_J в область $G_{\{0,1,\ldots,n\} \setminus J}$. Поэтому для величин

$$\psi_k = \frac{1}{\mathrm{vol}(\mathbb{S}^n)} \sum_{|J| = k} \mathrm{vol}(G_J)$$

(доля объема сферы, покрытого полусферами H_i ровно k раз) имеют место равенства

$$\psi_k = \psi_{n+1-k}\,.$$

Если мы сможем выразить величины $\{\psi_k\}$ через набор величин $\{\varphi_l\}$, то получатся соотношения между величинами углов симплекса Δ.

Включение $G_J \subset H_i$ имеет место тогда и только тогда, когда $i \in J$. Поэтому

$$G_J \subset H_I \ \Leftrightarrow\ J \supset I.$$

Следовательно, если $|J| = k$, то лунки H_I, отвечающие $(n-l)$-мерным граням, покрывают данную область G_J с кратностью $\binom{k}{l}$. Это наблюдение позволяет выразить суммарный объем $(n-l)$-лунок через объемы областей G_J следующим образом:

$$\varphi_{n-l} = \sum_{k=l}^{n+1} \binom{k}{l} \psi_k \,. \qquad (8)$$

(Например, для $l = n$ имеем $\varphi_0 = \psi_n + (n+1)\psi_{n+1}$.)

Заметим, что $\psi_{n+1} = \operatorname{vol}(\Delta)/\operatorname{vol}(\mathbb{S}^n)$. Для удобства введем дополнительную величину

$$\varphi_{-1} = \frac{\operatorname{vol}(\Delta)}{\operatorname{vol}(\mathbb{S}^n)} \,.$$

Тогда равенство (8) выполняется при всех l от 0 до $n+1$ и мы получаем систему из $(n+2)$ уравнений относительно неизвестных $\psi_0, \psi_1, \ldots, \psi_{n+1}$. Ее можно решить путем последовательного исключения неизвестных, начиная с ψ_{n+1}. Однако есть прием, позволяющий решить систему быстрее и записать соотношения между углами в более запоминающейся форме.

Умножим каждое из равенств (8) на t^l и просуммируем по l от 0 до $n+1$. В результате получим равенство

$$\sum_{l=0}^{n+1} \varphi_{n-l} t^l = \sum_{k=0}^{n+1} \psi_k (t+1)^k.$$

Производя замену $t \to t-1$ и сравнивая коэффициенты при соответствующих степенях t, приходим к равенствам

$$\psi_k = \sum_{l=k}^{n+1} (-1)^{l-k} \binom{l}{k} \varphi_{n-l} \quad \text{при } k = 0, 1, \ldots, n+1. \qquad (9)$$

Получающиеся отсюда при подстановке в равенства $\psi_k = \psi_{n+1-k}$ соотношения между углами сферического симплекса могут быть перенесены на евклидов случай. Достаточно точно так же, как мы это делали при доказательстве леммы 2, приблизить евклидов симплекс сферическим на сфере большого радиуса. В результате величина φ_{-1}, равная отношению объема симплекса к объему сферы, обнуляется.

Таким образом, нами доказана

Теорема 3 (Соотношения Дена – Соммервиля для симплекса). *Пусть φ_k, $k = 0, \ldots, n$, — сумма величин углов при k-мерных гранях n-мерного симплекса Δ (сферического или евклидова). Введем также обозначение*

$$\varphi_{-1} = \begin{cases} \dfrac{\operatorname{vol}(\Delta)}{\operatorname{vol}(\mathbb{S}^n)} & \textit{в сферическом случае,} \\ 0 & \textit{в евклидовом случае.} \end{cases}$$

Рассмотрим многочлен

$$\varphi(t) = \sum_{k=0}^{n+1} \varphi_{n-k} t^k$$

и положим $\varphi(t-1) = \sum_{k=0}^{n+1} \psi_k t^k = \psi(t)$*. Тогда имеет место равенство*

$$t^{n+1}\psi(t^{-1}) = \psi(t). \tag{10}$$

Или, в более явном виде,

$$\psi_k = \psi_{n+1-k} \quad \textit{при } k = 0, 1, \ldots, \lfloor \tfrac{n}{2} \rfloor\,. \tag{11}$$

Наконец, в терминах величин φ_k*:*

$$\sum_{l=k}^{n+1} (-1)^{l-k} \binom{l}{k} \varphi_{n-l} = \sum_{l=0}^{k} (-1)^{k-l} \binom{n+1-l}{k-l} \varphi_{l-1}\,, \tag{12}$$

также при $k = 0, 1, \ldots, \lfloor n/2 \rfloor$*.*

Проанализируем, сколько новой информации содержится в соотношениях Дена – Соммервиля. Можно показать, что $\lfloor (n+2)/2 \rfloor$ равенств (12) линейно независимы.[3] Однако соотношение для $k = 0$ равносильно равенству $\psi_0 = \psi_{n+1}$ и тем самым формуле Грама. Кроме того, из формулы (8) при $l = 0, 1$ и соотношений (11) следует:

$$\varphi_{n-1} = \sum_{k=1}^{n+1} k\psi_k = \frac{n+1}{2} \sum_{k=0}^{n+1} \psi_k = \frac{n+1}{2} \varphi_n\,,$$

что вполне естественно, поскольку по определению $\varphi_n = 1$ и $\varphi_{n-1} = (n+1)/2$ (у симплекса $n+1$ фасета, и угол при каждой равен $1/2$). Таким образом, после эквивалентных преобразований из системы (12) можно исключить одно тривиальное равенство, а одно из равенств (формула Грама) было нам уже известно. Итого мы имеем $\lfloor n/2 \rfloor - 1$ новых нетривиальных соотношений. В «осязаемых» случаях $n = 2$ и $n = 3$ мы, таким образом, не получаем ничего нового; в размерностях 4 и 5 имеется по одному новому линейному соотношению. Вместе с формулой Грама получаются следующие системы равенств:

$$\begin{cases} \varphi_0 - \varphi_1 + \varphi_2 - \dfrac{3}{2} = 2\varphi_{-1}\,, \\ \qquad 2\varphi_1 - 3\varphi_2 + 5 = 0 \end{cases} \quad \text{при } n = 4,$$

$$\begin{cases} \varphi_0 - \varphi_1 + \varphi_2 - \varphi_3 + 2 = 0, \\ \qquad\qquad \varphi_2 - 2\varphi_3 + 5 = 0 \end{cases} \quad \text{при } n = 5.$$

---

[3] Для того, чтобы в этом убедиться, удобно перейти от переменных φ_l к переменным ψ_k.

Теорему (3) можно обобщить на случай симплициального многогранника. Напомним, что многогранник размерности n называется *симплициальным*, если все его грани, за исключением быть может его самого, являются симплексами.

Теорема 4 (Соотношения Дена – Соммервиля для симплициального многогранника). *Пусть P — n-мерный симплициальный многогранник (сферический или евклидов). Обозначим через f_k, $k = 0, 1, \ldots, n$, количество его k-мерных граней, через φ_k сумму нормированных величин углов при k-мерных гранях. Кроме того, положим $f_{-1} = 1$ и*

$$\varphi_{-1} = \begin{cases} \dfrac{\mathrm{vol}(P)}{\mathrm{vol}(\mathbb{S}^n)} & \text{в сферическом случае,} \\ 0 & \text{в евклидовом случае.} \end{cases}$$

Рассмотрим многочлены

$$\varphi(t) = \sum_{k=0}^{n+1} \varphi_{n-k} t^k,$$

$$f(t) = \sum_{k=0}^{n} f_{n-k-1} t^k$$

и положим

$$\varphi(t-1) = \sum_{k=0}^{n+1} \psi_k t^k = \psi(t),$$

$$f(t-1) = \sum_{k=0}^{n} h_k t^k = h(t).$$

Тогда имеет место равенство

$$t^{n+1}(\psi(t^{-1}) + h(t^{-1}) - 1) = \psi(t) + h(t) - 1. \tag{13}$$

В более явном виде,

$$\begin{aligned} \psi_k - \psi_{n+1-k} &= h_{n+1-k} - h_k \quad \text{при } k = 1, \ldots, \lfloor \tfrac{n}{2} \rfloor, \\ \psi_0 - \psi_{n+1} &= h_{n+1} - h_0 + 1. \end{aligned} \tag{14}$$

Заметим, что обе части последнего равенства в (14) равны 0 по формулам Эйлера и Грама.

Доказательство. Выберем внутри многогранника точку c_P и разобьем P на n-мерные симплексы — пирамиды с вершиной c_P и фасетами многогранника в качестве оснований. Подсчитаем суммы величин углов всех симплексов при гранях данной размерности. Углы при вершинах многогранника составлены из углов при вершинах симплексов, при этом остаются еще углы симплексов при их общей вершине c_P. Вместе они составляют полный угол, поэтому имеем

$$\sum_{\Delta} \varphi_0(\Delta) = \varphi_0 + 1.$$

Вообще, для любого k от 0 до $n-1$ углы при k-мерных гранях многогранника составлены из углов при k-мерных гранях симплексов разбиения. Каждая из не учтенных при этом k-мерных граней симплексов натянута на вершину c_P и некоторую $(k-1)$-мерную грань многогранника. Соответствующие углы группируются в f_{k-1} полных углов. Таким образом, имеют место равенства

$$\sum_{\Delta} \varphi_k(\Delta) = \varphi_k + f_{k-1} \quad \text{при } k = 0, \ldots, n-1.$$

Кроме того,

$$\sum_{\Delta} \varphi_n(\Delta) = f_{n-1}\,,$$
$$\sum_{\Delta} \varphi_{-1}(\Delta) = \varphi_{-1}\,.$$

Обозначая φ-многочлен для симплекса Δ через $\varphi^{\Delta}(t)$, приходим к равенству

$$\sum_{\Delta} \varphi^{\Delta}(t) = \varphi(t) + f(t) - 1.$$

Теорема 4 получается теперь применением теоремы 3 к каждому из симплексов Δ. □

Задача 3. Вычислите многочлен $h(t)$ для n-мерного симплекса.

Оказывается, коэффициенты многочлена h также обладают свойством симметрии, подобным равенствам $\psi_k = \psi_{n+1-k}$ для многочлена ψ. Эти соотношения более широко известны, чем соотношения между углами симплициального многогранника, и именно их имеют в виду, говоря о соотношениях Дена – Соммервиля.

Теорема 5 (Комбинаторные соотношения Дена – Соммервиля). *Пусть P — n-мерный симплициальный многогранник. Обозначим через f_k, $k = 0, 1, \ldots, n$, количество его k-мерных граней и положим $f_{-1} = 1$. Рассмотрим многочлен*

$$f(t) = \sum_{k=0}^{n} f_{n-k-1} t^k$$

и положим

$$f(t-1) = \sum_{k=0}^{n} h_k t^k = h(t).$$

Тогда имеет место равенство

$$t^n h(t^{-1}) = h(t). \tag{15}$$

В более явном виде,

$$h_k = h_{n-k}.$$

Доказательство. Воспользуемся теоремой 4. Поскольку нас интересуют только количества граней многогранника, мы можем среди многогранников данного комбинаторного типа выбрать такой, величины углов которого легко вычислить или оценить. В случае удачного выбора из соотношений теоремы 4 будут следовать нужные нам соотношения.

Пусть P — данный евклидов многогранник в $\mathbb{R}^n$. Рассмотрим n-мерную сферу, касающуюся пространства $\mathbb{R}^n$ во внутренней точке многогранника P, и спроецируем P на сферу из ее центра. Если радиус сферы r мал, то получающийся при проекции многогранник $P_{\mathbb{S}}$ занимает почти всю полусферу, обращенную к пространству $\mathbb{R}^n$. Кроме того, можно показать, что величины углов многогранника P при всех гранях размерностей от 0 до $n-1$ стремятся к $1/2$ при стремлении радиуса сферы к 0. Таким образом,

$$\lim_{r\to 0} \varphi_k = \frac{1}{2} f_k \quad \text{при } k = 0, \ldots, n-1.$$

Заметим также, что $\lim_{r\to 0} \varphi_{-1} = 1/2$ и $\varphi_n = 1$ независимо от величины радиуса. Отсюда получаем

$$\lim_{r\to 0}(\varphi(t) + f(t) - 1) = \frac{t}{2} f(t) + 1 + f(t) - 1 = \left(\frac{t}{2} + 1\right) f(t),$$

где сходимость понимается как сходимость последовательностей коэффициентов многочленов. Следовательно,

$$\lim_{r\to 0}(\psi(t) + h(t) - 1) = \frac{t+1}{2} h(t).$$

Тогда из равенства (13) следует

$$t^{n+1} \frac{t^{-1}+1}{2} h(t^{-1}) = \frac{t+1}{2} h(t),$$

откуда напрямую следует равенство (15). □

Комбинаторные соотношения Дена – Соммервиля были впервые доказаны Соммервилем в [6], но независимо от метрических соотношений теоремы 4. Необходимо упомянуть, что исследования Соммервиля были вдохновлены работой Дена [2]. Ден доказал теорему 3 для $n = 4$ и $n = 5$, указав также на возможность обобщения на случай произвольного n. Метрические соотношения Дена – Соммервиля доказаны также в статье [4] для «квазисимплициальных» многогранников — n-мерных многогранников, у которых все грани размерности меньшей либо равной $n-2$ являются симплексами.

5. Другие соотношения между углами

Кроме внутреннего угла при грани Q многогранника P, можно рассмотреть также и внешний угол. Для этого определим нормальный конус $N_Q(P)$ грани Q относительно P следующим образом:

$$N_Q(P) = \left\{ \sum a_i \boldsymbol{n}_i \mid a_i \geqslant 0 \right\},$$

где $\boldsymbol{n}_i$ — внешние нормали к фасетам, содержащим грань Q. Эквивалентное определение: возьмем точку x во внутренности грани Q и рассмотрим все точки пространства $\mathbb{R}^n$, для которых x — ближайшая точка многогранника P. Эти точки составляют множество $x+N_Q(P)$. Нетрудно видеть, что размерность конуса $N_Q(P)$ равна $\dim P - \dim Q$. В частности, нормальные конусы вершин имеют полную размерность. Из второго определения внешнего угла нетрудно увидеть, что нормальные конусы вершин покрывают все пространство, пересекаясь лишь по конусам меньшей размерности. Поэтому, если мы определим величину внешнего угла при грани Q стандартным образом:

$$\nu_Q(P) = \frac{\mathrm{vol}(B \cap N_Q)}{\mathrm{vol}(B)},$$

то получим равенство

$$\sum_v \nu_v(P) = 1, \tag{16}$$

где суммирование ведется по всем вершинам многогранника P. В случае трехмерного многогранника формула (16) является дискретным аналогом формулы Гаусса – Бонне для сферы.

Наконец, упомянем соотношения, вовлекающие внутренние и внешние углы одновременно. На этот раз пусть P — полиэдральный конус (пересечение конечного числа полупространств, проходящих через начало координат). Грани конуса P, а также величины внутренних и внешних углов при гранях определяются очевидным образом. Питер Макмаллен доказал в [3] равенства

$$\sum_Q \varphi_0(Q)\nu_Q(P) = 1,$$
$$\sum_Q (-1)^{\dim Q}\varphi_0(Q)\nu_Q(P) = 0,$$
$$\sum_Q (-1)^{\dim Q}\nu_0(Q)\varphi_Q(P) = 0,$$

где суммирование ведется каждый раз по всем граням конуса P, а начало координат рассматривается как грань каждого из конусов. Отметим, что на самом деле полиэдральный конус — это то же самое, что сферический многогранник (рассмотрим его пересечение со сферой S с центром в начале координат). При этом величинам внутренних углов $\varphi_0(Q)$ отвечают объемы граней $S \cap P$, а внешние углы остаются внешними углами. Если сложить первое и второе равенства Макмаллена, то получится соотношение, вовлекающее только грани четной размерности сферического многогранника. Это соотношение является дискретным аналогом формулы Гаусса – Бонне для высших размерностей.

Доказательства формулы Гаусса – Бонне и формулы Грама приведены в главе 7 книги [1]. Там рассматривается также случай многогранников в пространстве Лобачевского.

Автор признателен Э. Б. Винбергу, М. Н. Вялому и В. В. Прасолову за ценные замечания.

Список литературы

[1] Алексеевский Д. В., Винберг Э. Б., Солодовников А. С. *Геометрия пространств постоянной кривизны.* Итоги науки и техн. Совр. проблемы матем. Фунд. напр., т. 29. М.: ВИНИТИ, 1988.

[2] Dehn M. *Die Eulersche Formel im Zusammenhang mit dem Inhalt in der nichteuklidischen Geometrie* // Math. Ann., 1906. B. 61. S. 561–586.

[3] McMullen P. *Non-linear angle-sum relations for polyhedral cones and polytopes* // Math. Proc. Cambridge Philos. Soc., 1975. Vol. 78, no. 2. P. 247–261.

[4] Perles M. A., Shephard G. C. *Angle sums of convex polytopes* // Math. Scand., 1967. Vol. 21. P. 199–218.

[5] Prasolov V. V., Tikhomirov V. M. *Geometry.* Translations of Mathematical Monographs, 200. American Mathematical Society, Providence, RI, 2001.

[6] Sommerville D. M. Y. *The relations connecting the angle-sums and volume of a polytope in space of n dimensions* // Proceedings Royal Soc. London, Ser. A, 1927. Vol. 115. P. 103–119.

[7] Shephard G. C. *An elementary proof of Gram's theorem for convex polytopes* // Can. J. Math., 1967. Vol. 19. P. 1214–1217.

[8] Ziegler G. M. *Lectures on polytopes*. Graduate Texts in Mathematics. 152. Berlin: Springer-Verlag, 1995.

Изместьев Иван В.,
Institut für Mathematik MA 8-3,
Technische Universität Berlin,
Str. des 17. Juni 136
D-10623 Berlin
Germany
e-mail: `izmestie@math.tu-berlin.de`

Обобщенная теорема Ван дер Вардена

В. О. Бугаенко

Введение

В двадцатых годах прошлого века внимание математиков привлекла задача с элементарной формулировкой, решение которой длительное время найти не удавалось. Вот эта задача.

> *Пусть множество целых чисел раскрашено в конечное число цветов. Тогда найдется арифметическая прогрессия сколь угодно большой конечной длины, члены которой окрашены в один цвет.*

После упорных усилий задачу удалось решить молодому голландскому математику Б. Л. Ван дер Вардену. Решение оказалось элементарным, но достаточно сложным. История этого доказательства приведена в книге А. Я. Хинчина [3], а в изложении самого Ван дер Вардена — в дополнении к книге Р. Грэхема [1]. В обеих этих книгах можно также найти доказательство теоремы.

Как часто бывает в математике, чтобы упростить решение задачи, нужно сформулировать ее для более общего случая. Первым шагом к доказательству самого Ван дер Вардена (и, по видимому, ко всем известным доказательствам) была догадка о том, что доказывать нужно более сильное утверждение, а именно: предполагать, что раскрашивается не вся числовая прямая, а лишь некоторый ее конечный кусок, размер которого зависит от количества цветов раскраски и длины прогрессии.

Мы обобщим задачу сразу в нескольких направлениях, главным из которых будет выход из числовой прямой на плоскость. Вместо раскраски множества целых чисел рассмотрим раскраску целочисленной решетки на плоскости в конечное число цветов. Выбрав в этой решетке *конечную фигуру M* (конечное множество точек), мы будем доказывать существование подобной ей одноцветной фигуры. Наше доказательство будет в основных идеях повторять доказательство Ван дер Вардена, но приобретет геометрическую наглядность, утерянную в вырожденном одномерном случае.

Сформулируем другие обобщения, которые будут сделаны. Во-первых, ясно, что перейдя от размерности один к размерности два, можно двинуться и дальше к случаю решетки в пространстве любой размерности. Во-вторых, условие подобия можно заменить на более сильное — гомотетичность с целым положительным коэффициентом. (В дальнейшем, говоря о гомотетии, мы всегда будем иметь в виду гомотетию с целым положительным коэффициентом.) Достаточно изящное доказательство такого обобщения теоремы было приведено П. Андерсоном [4], который, ссылаясь на Р. Радо, приписывает исходное доказательство Г. Грунвальду. Затем доказательство Андерсона было пересказано на русском языке В. В. Прасоловым [2].

Мы пойдем несколько дальше и будем предполагать, что «объектом раскраски» может служить не только решетка, но и всё пространство (при этом рассматриваемая фигура M может не вкладываться ни в какую решетку в этом пространстве). Вообще говоря, такое обобщение можно вывести из теоремы для решетки. Предлагаем читателю сделать это самостоятельно (см. упражнение 4). Мы же, модернизировав доказательство Андерсона, добьемся того, чтобы оно сразу годилось для случая не только решетки, но и всего пространства.

Аналогично уже упомянутому усилению исходной (одномерной) теоремы Ван дер Вардена, будем предполагать в условии, что красится не всё пространство, а лишь конечная фигура в нём (зависящая от данной в условии фигуры и количества цветов раскраски).

Обозначим через L объект раскраски — либо пространство любой размерности, либо целочисленную решетку в пространстве. Будем говорить, что фигура $\widehat{M} \subset L$ является *монохроматической накрывающей ранга* k фигуры $M \subset L$, если для любой раскраски пространства (или решетки) L в k цветов существует одноцветная фигура $F \subset \widehat{M}$, гомотетичная M. Заметим, что образы монохроматической накрывающей при сдвиге и гомотетии также являются монохроматическими накрывающими той же фигуры того же ранга. Теперь мы можем сформулировать основную теорему.

Теорема. *Для любой конечной фигуры $M \subset L$ и любого натурального числа k существует ее конечная монохроматическая накрывающая ранга k.*

Частные случаи

В этом параграфе мы докажем несколько частных случаев основной теоремы. Начав с простейшего случая теоремы Ван дер Вардена для двух цветов и прогрессии из трех членов, мы обобщим его в трех направлениях: выход на плоскость, увеличение количества цветов и увеличение количества точек. В результате мы получим доказательство того, что

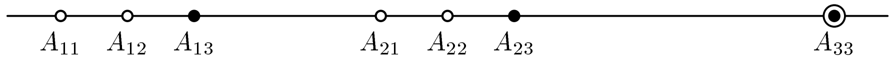


Рис. 1.

в раскрашенной в два цвета целочисленной решетке на плоскости найдется одноцветный квадрат. Тем самым, с одной стороны, будет решена непростая задача, представляющая самостоятельный интерес, а с другой стороны, будут представлены все основные идеи, необходимые для доказательства теоремы в общем случае, которое будет приведено в следующем параграфе. Иллюстрациями к нему смогут служить нижеприведенные примеры. Желающие сразу перейти к общему случаю, могут этот параграф пропустить.

1. *В раскрашенном в два цвета множестве целых чисел найдутся три одноцветных числа, образующие арифметическую прогрессию.*

Это простая задача, допускающая много различных решений. Мы приведем решение, которое станет основой для дальнейшего обобщения.

Рассмотрим на числовой прямой (тем самым, вместо чисел будем говорить о точках) девять троек точек вида $(x, x+1, x+2)$ (например, для $1 \leqslant x \leqslant 9$). Существует $2^3 = 8$ способов раскраски такой тройки точек, поэтому, согласно принципу Дирихле, какие-то две из рассматриваемых троек раскрашены одинаково. С другой стороны, в каждой тройке найдутся две одноцветные точки. Таким образом, мы получили четыре точки A_{11}, A_{12}, A_{21} и A_{22}, раскрашенные одинаково (для определенности скажем, в белый цвет), причем $\overrightarrow{A_{11}A_{12}} = \overrightarrow{A_{21}A_{22}}$ (рис. 1). Отметим также точки A_{13} и A_{23} так, чтобы тройки A_{11}, A_{12}, A_{13} и A_{21}, A_{22}, A_{23} образовывали арифметические прогрессии. Если какая-то из этих двух точек белая, то искомая тройка найдена. Если же они обе черные, то рассмотрим точку A_{33}, образующую арифметическую прогрессию с точками A_{13} и A_{23}. Если точка A_{33} черная, то одноцветной будет тройка точек (A_{13}, A_{23}, A_{33}), а если белая, то тройка (A_{11}, A_{22}, A_{33}).

Координаты всех точек приведённой выше конструкции лежат в интервале между 1 и 21, поэтому монохроматической накрывающей ранга 2 для трех последовательных целых точек на прямой является множество из 21 последовательной целой точки.

2. *В раскрашенной в два цвета целочисленной решетке на плоскости найдется одноцветный равнобедренный прямоугольный треугольник.*

Рассмотрим прямую, параллельную оси решетки, и на ней найдем четыре одноцветные точки A_{11}, A_{12}, A_{21} и A_{22}, расположенные так же, как и в предыдущем примере. Точки A_{13}, A_{23} и A_{33} выберем так, чтобы треугольники $A_{11}A_{12}A_{13}$, $A_{21}A_{22}A_{23}$ и $A_{13}A_{23}A_{33}$ были прямоугольными

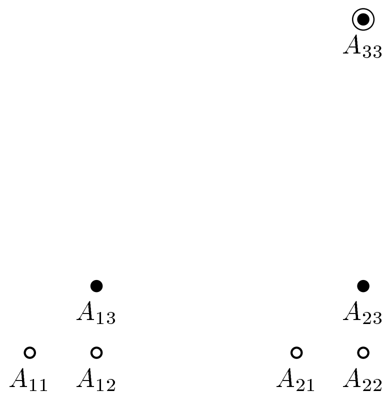


Рис. 2.

равнобедренными (рис. 2). Дальнейшее доказательство повторяет рассуждения из п. 1. Все построенные точки лежат внутри квадрата 11×11, поэтому монохроматической накрывающей ранга 2 для равнобедренного прямоугольного треугольника будет множество точек целочисленной решетки, заполняющих квадрат 11×11.

Замечание. Нетрудно заметить, что случай произвольного треугольника фактически не отличается от разобранного. Действительно, условие теоремы инвариантно относительно аффинных преобразований, а все треугольники аффинно эквивалентны. Следует лишь уточнить в условии, что в качестве решетки нужно взять целочисленную решетку в некоторой косоугольной системе координат.

3. *В раскрашенной в три цвета целочисленной решетке на плоскости найдется одноцветный равнобедренный прямоугольный треугольник.*

Рассмотрим на плоскости достаточно большой (смысл слов «достаточно большой» будет уточнен ниже) квадрат, «нижняя» сторона которого лежит на оси абсцисс, и его всевозможные сдвиги на целочисленные векторы вдоль этой оси. Среди полученных квадратов найдутся два одинаково окрашенных, поскольку количество всевозможных раскрасок квадрата конечно. Обозначим эти квадраты Φ_1 и Φ_2. Аналогично, внутри квадрата Φ_1 найдем два одинаково раскрашенных квадрата Φ_{11} и Φ_{12} меньшего размера (также имеющие по стороне на оси абсцисс). Соответствующие им квадраты внутри квадрата Φ_2 обозначим Φ_{21} и Φ_{22} соответственно. На стороне квадрата Φ_{11}, лежащей на оси абсцисс, найдем две одноцветные точки A_{111} и A_{112} (условием их существования и определяются размеры всех упоминаемых квадратов, в частности, квадрат Φ_{11} должен иметь размер 4×4). Соответственно одноцветные пары

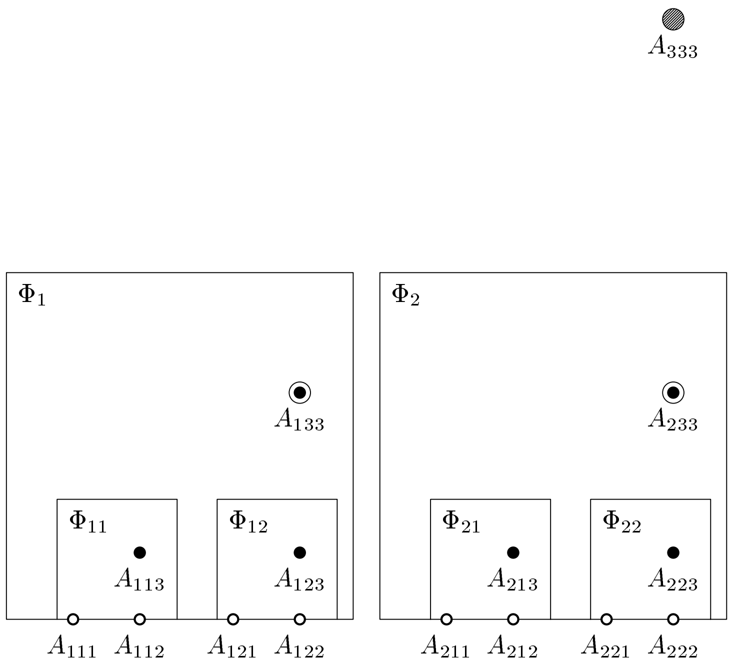


Рис. 3.

(A_{121}, A_{122}), (A_{211}, A_{212}) и (A_{221}, A_{222}) будут внутри квадратов Φ_{12}, Φ_{21} и Φ_{22} (рис. 3). Построим точки $A_{i_1 i_2 3}$ $(1 \leqslant i_1, i_2 \leqslant 2)$ так, чтобы треугольники $A_{i_1 i_2 1} A_{i_1 i_2 2} A_{i_1 i_2 3}$ были равнобедренными прямоугольными. Далее строим точки $A_{i_1 33}$ $(1 \leqslant i_1 \leqslant 2)$ так, чтобы треугольники $A_{i_1 13} A_{i_1 23} A_{i_1 33}$ были равнобедренными прямоугольными. Наконец, находим точку A_{333}, чтобы треугольник $A_{133} A_{233} A_{333}$ был равнобедренным прямоугольным. Получившаяся конструкция из 15 точек $A_{i_1 i_2 i_3}$ содержит 8 точек, индексы которых равны 1 или 2 (их мы назовем *основными*), и 7 точек, имеющих хотя бы один индекс, равный 3 (а их назовем *добавочными*). Все основные точки одноцветные. Если хотя бы одна из добавочных точек того же цвета, то одноцветный треугольник легко находится. В противном случае каждая из них покрашена в один из оставшихся двух цветов, при этом к конструкции из добавочных точек можно применить рассуждение из п. 2.

Замечание. Аналогичным образом наращивая приведённую конструкцию, можно доказать, что одноцветный равнобедренный прямоугольный треугольник существует при раскраске решетки в любое конечное число цветов. Мы будем использовать этот факт в п. 4, хотя строгое его доказательство отложим до общего случая.

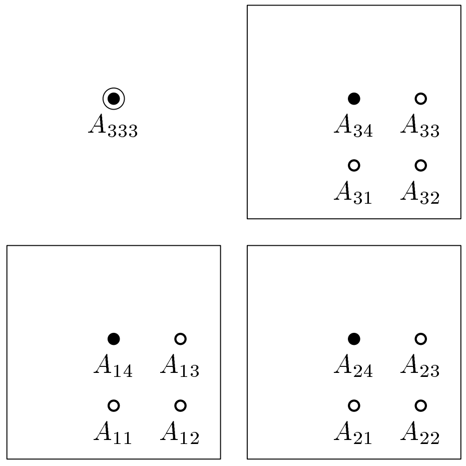


Рис. 4.

4. *В раскрашенной в два цвета целочисленной решетке на плоскости найдется одноцветный квадрат.*

Рассмотрим новую раскраску решетки, получаемую из данной следующим образом. Цвет точки определяется цветом квадрата 11×11 с левым нижним углом в ней. Количество цветов такой раскраски равно 2^{121}. Найдется одноцветный (в смысле этой раскраски) равнобедренный прямоугольный треугольник. Это значит, что существуют три равных одинаково окрашенных квадрата 11×11, расположенных в виде равнобедренного прямоугольного треугольника. В любом из них, согласно п. 2, найдется одноцветный равнобедренный прямоугольный треугольник. Выберем такой треугольник в одном из квадратов, в остальных рассмотрим треугольники, соответствующие ему при наложении квадратов. Мы получаем три одноцветных (белых) равнобедренных прямоугольных треугольника $A_{11}A_{12}A_{13}$, $A_{21}A_{22}A_{23}$ и $A_{31}A_{32}A_{33}$ (рис. 4). Дополним каждый из них до квадрата точками A_{14}, A_{24} и A_{34} соответственно. Если одна из этих точек белая, то искомый одноцветный квадрат найден. Если же все они черные, то рассмотрим точку A_{44}, дополняющую до квадрата треугольник $A_{14}A_{24}A_{34}$. Если эта точка черная, то искомым одноцветным квадратом будет $A_{14}A_{24}A_{34}A_{44}$, а если белая — то $A_{11}A_{22}A_{33}A_{44}$.

Доказательство основной теоремы

Перейдем к доказательству основной теоремы, сформулированной на с. 152. Будем использовать индукцию по количеству точек фигуры M, которое мы обозначим n.

База индукции ($n = 2$) очевидна. Действительно, пусть множество M состоит из двух точек. На соединяющей их прямой отложим $k+1$ точку так, чтобы расстояния между соседними точками были равны расстоянию между точками фигуры M. Согласно принципу Дирихле, среди этих точек найдутся две одноцветные. Они и будут составлять фигуру, гомотетичную M. Значит, построенное множество из $k+1$ точки и будет монохроматической накрывающей ранга k.

Шаг индукции. Пусть мы имеем натуральное число k и фигуру $M = \{M_1, M_2, \ldots, M_{n+1}\}$, содержащую $n+1$ точку. Зафиксируем в L точку O, которую будем называть началом координат. Радиус-векторы точек фигуры M обозначим $\vec{v}_i = \overrightarrow{OM_i}$. Согласно предположению индукции, для фигуры $M' = \{M_1, M_2, \ldots, M_n\}$, содержащей n точек, монохроматическая накрывающая любого ранга (а не только ранга k) существует.

При наличии начала координат любая конечная фигура может быть задана набором векторов, а именно, множеством радиус-векторов ее точек. Будем называть конечную фигуру в пространстве с фиксированным началом координат *шаблоном*. Шаблон может быть «приложен» к любой точке плоскости. Эта операция означает его сдвиг на вектор, соединяющий начало координат с точкой приложения. Всякий шаблон из n точек в пространстве, раскрашенном в k цветов, естественным образом порождает новую раскраску в k^n цветов. Цвет каждой точки определяется набором цветов (в исходной раскраске) приложенного к ней шаблона.

Выберем в L конечную последовательность шаблонов $\Phi^{(j)}$ ($j = 0, 1, \ldots, k$) следующим образом. Шаблон $\Phi^{(0)}$ является *тривиальным* (состоит из одной точки — начала координат). При $j \geqslant 1$ шаблон $\Phi^{(j)}$ строим (используя уже построенный шаблон $\Phi^{(j-1)}$) следующим образом. Сначала определим фигуру $\Phi''^{(j)}$ как монохроматическую накрывающую ранга $k^{|\Phi^{(j-1)}|}$ фигуры M' (знак модуля означает количество точек, составляющих фигуру). Затем построим фигуру $\Phi'^{(j)} = \bigcup\limits_{A \in \Phi''^{(j)}} \Phi_A^{(j-1)}$ — объединение шаблонов $\Phi^{(j-1)}$, приложенных ко всем точкам фигуры $\Phi''^{(j)}$. Наконец расширим $\Phi'^{(j)}$ до $\Phi^{(j)}$ так, чтобы для любого гомотетичного образа $\{A_1, A_2, \ldots, A_{n+1}\}$ множества M, для которого $\{A_1, A_2, \ldots, A_n\} \subset \Phi'^{(j)}$, было выполнено $A_{n+1} \in \Phi^{(j)}$.

Докажем, что $\Phi^{(k)}$ и будет монохроматической накрывающей ранга k фигуры M. Для этого построим последовательность фигур C_j ($j = 0, 1, \ldots, k$), которые мы назовем *конструкциями*. Каждая конструкция C_j состоит из точек $A_{i_1 i_2 \ldots i_j}$ (иными словами, точки конструкции C_j мы будем обозначать буквой A с j индексами). Индексы изменяются от 1 до $n+1$, но со следующим ограничением, которое мы назовем *условием мажорирования*: если какой-либо индекс равен $n+1$, то и все последующие также равны $n+1$. Точки, все индексы которых не превосходят n, будем называть *основными*, а имеющие индекс равный $n+1$ — *добавочными*. При построении конструкции мы будем добиваться выполнения *условия изотропности*: чтобы шаблон $\Phi^{(k-j)}$ был окрашен одинаково, будучи приложенным ко всем основным точкам фигуры C_j.

Построение будем опять осуществлять по индукции. Конструкция C_0 состоит из одной точки A (без индексов). Условие изотропности для одноточечной конструкции, очевидно, выполняется. Без ограничения общности можно считать, что точка A совпадает с началом координат O. Опишем построение конструкции C_j в предположении, что C_{j-1} уже построена. К каждой точке конструкции C_{j-1} приложим шаблон $\Phi^{(k-j+1)}$. Для всех основных точек эти шаблоны окрашены одинаково, поскольку по предположению индукции условие изотропности для C_{j-1} выполняется. Сузим один из этих шаблонов (приложенный к некоторой выбранной основной точке T) до шаблона $\Phi'^{(k-j+1)}$, а в нём найдем одноцветную с точки зрения раскраски, порожденной шаблоном $\Phi^{(k-j)}$ (это обеспечит изотропность конструкции на следующем шаге), гомотетичную M' фигуру. Затем получившуюся фигуру дополним до фигуры, гомотетичной M, принадлежащей исходному несуженному шаблону $\Phi^{(k-j+1)}$. Вектор, соединяющий точку приложения шаблона с центром гомотетии, обозначим u_j, а коэффициент гомотетии обозначим λ_j. Перенесем параллельно полученную фигуру всевозможными сдвигами, переводящими выбранную точку T во все остальные точки конструкции C_{j-1} (в том числе и добавочные). Таким образом, каждая точка конструкции C_{j-1} породит $n+1$ дочернюю точку. Обозначим их, приписывая к обозначению соответствующей материнской точки очередной j-й индекс, изменяющийся от 1 до $n+1$. Тогда векторы, соединяющие точки конструкции C_{j-1} с дочерними точками, выражаются формулой

$$\overrightarrow{A_{i_1 i_2 \dots i_{j-1}} A_{i_1 i_2 \dots i_{j-1} i_j}} = \vec{u}_j + \lambda_j \vec{v}_{i_j},$$

откуда легко получить формулу для радиус-вектора любой точки:

$$\overrightarrow{OA_{i_1 i_2 \dots i_j}} = (\vec{u}_1 + \vec{u}_2 + \dots + \vec{u}_j) + \lambda_1 \vec{v}_{i_1} + \lambda_2 \vec{v}_{i_2} + \dots + \lambda_j \vec{v}_{i_j}. \qquad (1)$$

Наконец, из всех полученных точек $A_{i_1 i_2 \dots i_j}$ сотрем те, чей набор индексов не удовлетворяет условию мажорирования. В результате мы получим искомую конструкцию C_j.

Нам понадобятся три вспомогательных утверждения.

Лемма 1. *Точка $A_{i_1 i_2 \dots i_k}$ при любом j принадлежит шаблону $\Phi^{(j)}$, приложенному к точке $A_{i_1 i_2 \dots i_{k-j}}$. В частности ($j = k$), все точки конструкции C_k принадлежат $\Phi^{(k)}$.*

Доказательство. Применим индукцию по j. База ($j = 0$) очевидна: любая точка принадлежит приложенному к ней тривиальному шаблону. Шаг индукции следует из того, что шаблон $\Phi^{(j)}$, приложенный к точке $A_{i_1 i_2 \dots i_{k-j}}$ содержит в себе объединение фигур, полученных приложением шаблона $\Phi^{(j-1)}$ к нескольким точкам, в том числе и к $A_{i_1 i_2 \dots i_{k-j+1}}$.

Лемма 2. *Пусть среди точек $A_{i_1 i_2 \dots i_k}$ выбрано упорядоченное множество из $n+1$ точки так, что для каждого индекса выполнено одно из двух: либо он одинаков для всех точек множества, либо для каждой точки он равен ее номеру. Тогда выбранное множество гомотетично M.*

Доказательство. Обозначим выбранные точки $F_1, F_2, \dots, F_{n+1}$. Пусть индексы (в исходных обозначениях), для которых выполняется второй из указанных в условии случаев, имеют номера $j_1, j_2, \dots, j_r$. Тогда из формулы (1) следует

$$\overrightarrow{F_sF_t} = \overrightarrow{OF_t} - \overrightarrow{OF_s} = \sum_{i=1}^{r} \lambda_{j_i}(\vec{v}_t - \vec{v}_s) = \left(\sum_{i=1}^{r} \lambda_{j_i}\right) \overrightarrow{M_sM_t}.$$

А это и означает, что выбранные точки составляют фигуру, гомотетичную M с коэффициентом $\sum\limits_{i=1}^{r} \lambda_{j_i}$.

Лемма 3. *Изменение не равного $n+1$ значения любого индекса любой точки на другое значение, также не равное $n+1$, не меняет ее цвета.*

Доказательство. Точки $A_{i_1 \dots i'_j \dots i_k}$ и $A_{i_1 \dots i''_j \dots i_k}$, отличающиеся лишь в j-м индексе являются соответствующими точками шаблона $\Phi^{(k-j)}$, приложенного к точкам $A_{i_1 \dots i'_j}$ и $A_{i_1 \dots i''_j}$. Но по условию изотропности эти шаблоны раскрашены одинаково, значит и соответствующие их точки одноцветны.

Перейдем к завершению доказательства основной теоремы. Согласно лемме 1, шаблон $\Phi^{(k)}$ содержит в себе конструкцию C_k. Из леммы 3 следует, что все основные точки этой конструкции окрашены в один цвет. Назовем этот цвет первым. Рассмотрим некоторую добавочную точку $A_{i_1 i_2 \dots i_k}$. Введем для нее альтернативное обозначение F_{n+1}, и построим n основных точек F_s ($s = 1, 2, \dots, n$), получаемых из F_{n+1} заменой в индексах ее исходного выражения $A_{i_1 i_2 \dots i_k}$ всех чисел $n+1$ на s. По лемме 2, полученное множество точек $F = \{F_1, F_2, \dots, F_{n+1}\}$ гомотетично M. Значит, если хотя бы одна добавочная точка окрашена в первый цвет, то искомое одноцветное множество найдено. В противном случае рассмотрим множество таких добавочных точек, у которых только последний индекс равен $n+1$. Опять же, по лемме 3 они все окрашены в один цвет;

назовем его вторым. Рассуждая аналогично, получим, что если во второй цвет окрашена хотя бы одна из точек, у которой числу $n+1$ равны два последних индекса, то утверждение доказано. Продолжая эти рассуждения (строго говоря, применяя индукцию), мы приходим к выводу, что утверждение доказано, если при некотором $s \leqslant k$ в один из первых s цветов (порядок на множестве цветов возникает при этом естественным образом) окрашена хотя бы одна точка, у которой последние s индексов равны $n+1$. Последнее же, очевидно, справедливо при $s = k$, поскольку точка $A_{n+1,n+1,\ldots,n+1}$ окрашена в один из данных k цветов.

Автор благодарит В. А. Клепцына и М. Н. Вялого за полезные обсуждения.

Упражнения

1. Покажите, что теорема Ван дер Вардена является частным случаем доказанной теоремы.

2. Можно ли при доказательстве теоремы в качестве базы индукции взять более простой случай $n = 1$?

3. Некоторые из точек $A_{i_1 i_2 \ldots i_k}$ из приведённой в доказательстве конструкции могут совпадать. Покажите, что и в этом случае доказательство остается в силе.

4. В предположении, что теорема доказана только для случая целочисленной решетки, приведите короткое рассуждение, доказывающее ее в полной общности.

Указание: решетка большей размерности может быть вложена (например, с помощью проецирования) в пространство меньшей размерности.

Список литературы

[1] *Р. Грэхем.* Начала теории Рамсея. М. Мир. 1984.

[2] *В. В. Прасолов.* Рассказы о числах, многочленах и фигурах. М. Фазис. 1997.

[3] *А. Я. Хинчин.* Три жемчужины теории чисел. М. Наука. 1979.

[4] *P. G. Anderson.* A generalizatoion of Baudet's conjecture (Van der Waerden theorem). Amer. Math. Monthly, 1976, **83**, 359–361.

В. О. Бугаенко, Московский независимый университет

E-mail: `bugaenko@mccme.ru`

Уникальносоставленные фигуры

А. М. Петрунин, С. Е. Рукшин

В этой статье мы рассматриваем отношение равносоставленности на множестве выпуклых плоских фигур.

Первым существенным результатом в этом направлении было доказательство Ф. Бойяи и П. Гервином (1832) следующего утверждения: *два многоугольника, имеющие равные площади, равносоставлены* (*т. е. один из них можно разрезать на конечное число многоугольных частей, переложив которые, можно получить другой многоугольник*).

С проблемой равносоставленности трехмерных тел связана знаменитая третья проблема Гильберта [1]. Как показал в 1900 г. Макс Ден, даже для тетраэдров равенство объемов не влечет их равносоставленности.

В дальнейшем теорема Бойяи – Гервина и другие результаты, связанные с третьей проблемой Гильберта, обобщались в различных направлениях [1].

Так, в математическом фольклоре известно следующее утверждение: круг не равносоставлен никакому многоугольнику. В процессе обобщения теоремы Бойяи – Гервина на случай *круговых многоугольников* [7], граница которых состоит из отрезков и дуг окружностей, удалось усилить это утверждение, доказав, что *круг не равносоставлен никакой выпуклой фигуре, отличной от круга.* Такие выпуклые фигуры, не равносоставленные никакой другой выпуклой фигуре, мы будем называть уникальносоставленными. В этой статье мы дадим полное описание таких фигур.

Эти результаты были получены нами почти 20 лет назад [6]. Знакомство с замечательной работой [12] позволило существенно упростить доказательство.

Для понимания большей части статьи потребуется только знание доказательства теоремы Бойяи – Гервина, которое можно найти, например, в книжках Болтянского [1] или [2].

Для упрощения восприятия мы доказываем последовательно несколько теорем, каждая следующая из которых обобщает предыдущую.

1. Обозначения и определения

F будет обозначать плоскую и чаще всего выпуклую фигуру, $S(F)$ обозначает площадь F, ∂F обозначает кривую, ограничивающую F.

Определение 1. Две плоских фигуры F и G называются *равносоставленными* (обозначение $F \sim G$), если одну из них можно разрезать отрезками прямой на конечное число частей и составить из этих частей другую.

Можно рассматривать и другие определения равносоставленности. Вместо разрезов по отрезкам прямой можно разрешить разрезы вдоль спрямляемых жордановых кривых, или вообще вдоль произвольных жордановых кривых. В общем случае эти определения дают различные классы равносоставленных фигур. Но, как было доказано в [12], для *выпуклых* фигур эти определения равносильны, т. е. если две *выпуклые* фигуры «равносоставлены» по одному из новых определений, то они равносоставлены и в смысле определения 1.

Равносильность этих определений является достаточно глубоким фактом, и именно она позволила существенно упростить первоначальные авторские доказательства: удалось отказаться от топологических подробностей и оставить только наглядные геометрические рассуждения, изложенные в настоящей статье.

К совершенно иным эффектам приводит теоретико-множественное определение равносоставленности. Назовем *равноразложимыми* два множества точек X и Y, если существует разбиение X на непересекающиеся множества $A_1, A_2, \ldots, A_n$ и Y на $B_1, B_2, \ldots, B_n$ такие, что множества B_i получаются из соответствующих A_i подходящими движениями $f_i\colon B_i = f(A_i)$. Знаменитый парадокс Банаха – Тарского состоит в том, что шар в $\mathbb{R}^3$ равноразложим с объединением двух непересекающихся шаров того же радиуса (см. [11]). Оказывается, что вообще любые два множества в $\mathbb{R}^3$, имеющие непустую внутренность, равноразложимы. Отсылаем заинтересованного читателя к книге [10].

В дальнейшем описании уникальносоставленных фигур мы будем понимать равносоставленность исключительно в смысле определения 1.

Определение 2. Выпуклая фигура F называется *уникальносоставленной*, если любая выпуклая фигура, равносоставленная F, является конгруэнтной F.

Как мы уже упоминали, к уникальносоставленным фигурам относится круг. В [12] доказана уникальносоставленность эллипса.

Определение 3. Две кривые (или два набора кривых) α и β называются *равносоставленными* (или $\alpha \sim \beta$), если первую(ый) можно разбить на конечное число дуг и составить из них вторую(ой).

Определение 4. Две кривые α и β называются *стабильно равносоставленными* (или $\alpha \approx \beta$), если из них можно исключить конечное число отрезков прямой так, что оставшиеся два набора кривых равносоставлены.

Следующее утверждение является простым обобщением теоремы Бойяи – Гервина, утверждающей, что равновеликие многоугольники равносоставлены, (см. [1] или [2]):

Утверждение 5. *Две выпуклые фигуры F и G равносоставлены тогда и только тогда, когда их площади равны и кривые, их ограничивающие, стабильно равносоставлены. Или*

$$F \sim G \;\Leftrightarrow\; S(F) = S(G) \text{ и } \partial F \approx \partial G.$$

Доказательство. Необходимость: совокупность всех кривых, ограничивающих куски разбиения, состоит из дуг первой кривой плюс какое-то количество «внутренних» отрезков прямой, и из тех же кусков состоит и вторая фигура. Таким образом, если из границы F вырезать отрезки прямой, которые соответствуют «внутренним» отрезкам разбиения G, и аналогично поступить с границей G, то оставшиеся куски границ будут равносоставлены.

Достаточность: рассмотрим разбиение границ F и G как в определении стабильной равносоставленности. Если к каждой такой дуге фигур F и G провести хорду, то образовавшиеся горбушки (сегменты) для равных дуг равны. Так как $S(F) = S(G)$, то и площади оставшихся многоугольников равны, а значит, они равносоставлены по теореме Бойяи – Гервина. □

2. Уникальносоставленность круга

Теорема 6. *Круг уникальносоставлен.*

Для доказательства нам потребуется следующая лемма:

Лемма 7. *Если граница выпуклой фигуры F состоит из конечного числа отрезков прямой и дуг окружностей радиуса R с общей угловой мерой $360°$, то она является R-окрестностью выпуклого многоугольника.*

В частности, F содержит в себе круг радиуса R.

Сначала мы покажем, что теорема 6 действительно следует из леммы:

Доказательство теоремы 6. Обозначим через K круг радиуса R. Пусть F есть выпуклая фигура, такая что $K \sim F$. Тогда из утверждения 5 мы имеем: $S(F) = S(K)$ и граница F *стабильно равносоставлена* окружности радиуса R. В частности, F удовлетворяет лемме 7. Таким образом, F содержит в себе копию K. Так как $S(F) = S(K)$, получаем $K \cong F$. □

В доказательстве леммы мы будем пользоваться следующим фактом, который предоставляем доказать читателю в качестве упражнения:

Упражнение 8. Пусть $A_1A_2\ldots A_n$ есть замкнутая ломаная, которая при обходе по ней поворачивает всё время против часовой стрелки, и общий угол поворота равен $360°$ — тогда ломаная $A_1A_2\cdots A_n$ ограничивает выпуклый многоугольник.

Доказательство леммы 7. Пусть α есть выпуклая кривая стабильно равносоставленная окружности радиуса R. Заметим, что в этом случае кривая не имеет угловых точек. Действительно, при обходе вокруг кривой необходимо повернуть на $360°$, и ровно на этот же угол мы поворачиваем, пройдя по всем дугам окружности; таким образом, на углы не остается места (т. е., если бы имелась еще и угловая точка, то общий угол поворота был бы больше чем $360°$, что невозможно). Таким образом, кривая α состоит из поочередно сменяющихся дуг окружностей и отрезков прямых, так что отрезки и дуги продолжают друг друга в том же направлении (см. рис. 1). Для каждого отрезка прямой в α рассмотрим его параллельный перенос в перпендикулярном направлении внутрь фигуры на расстояние R. При этом концы отрезков, соседние через дугу, перейдут в одну точку, и мы получим замкнутую ломаную из параллельных переносов всех отрезков α. Звенья этой ломаной поворачивают в одну и ту же сторону (так же как и α), и, так же как у α, общий угол поворота будет $360°$. Воспользовавшись упражнением 8, мы получаем, что эта ломаная ограничивает выпуклый многоугольник. Обозначим его M, и из построения легко видеть, что сама фигура F есть множество точек на расстоянии R от M. □

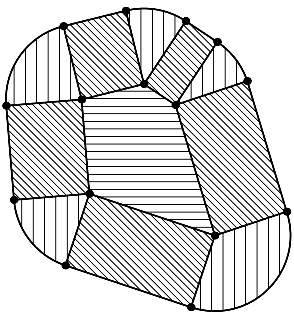


Рис. 1.

3. Уникальносоставленность линзы

Следующий факт имеет несколько более сложное доказательство:

Теорема 9. *Пересечение двух кругов одинакового радиуса есть уникальносоставленная фигура.*

Для краткости, давайте назовем пересечение двух кругов радиуса R линзой и будем обозначать ее L_ω, где ω есть угловая мера дуг, ее ограничивающих (в частности, $L_{2\pi}$ есть круг радиуса R). С помощью точно такого же рассуждения, что и в доказательстве теоремы 6, теорема 9 сводится к следующей лемме:

Лемма 10. *Если выпуклая фигура F имеет границу, стабильно равносоставленную границе линзы L_ω, то $S(F) \geqslant S(L_\omega)$. Более того, если $S(F) = S(L_\omega)$, то $F \cong L_\omega$.*

Доказательство будет проведено в два шага. Сначала мы докажем следующее, более слабое утверждение, а уже потом приступим к общему случаю.

Лемма 11. *Если центрально-симметричная выпуклая фигура F имеет границу, стабильно равносоставленную границе линзы L_ω, то $S(F) \geqslant S(L_\omega)$. Более того, если $S(F) = S(L_\omega)$, то $F \cong L_\omega$.*

Доказательство. В доказательстве мы воспользуемся двумя процедурами: «вырезанием параллелограмма» (см. рис. 2) и «четырехшарнирным методом» (см. рис. 3). Первый (примененный несколько раз) заменит фигуру F на фигуру F_1 с теми же свойствами, что и F, но с меньшей площадью и без отрезков прямых на границе. Второй (также примененный несколько раз) заменит уже полученную F_1 на F_2 с теми же свойствами, что и F_1, но с меньшей площадью и с не более чем двумя угловыми точками на границе (это немедленно влечет, что $F_2 \cong L_\omega$).

Предположим, на границе F есть отрезок прямой AB, и пусть $A'B'$ обозначает центрально-симметричный ему отрезок. Тогда из F можно вырезать параллелограмм $ABA'B'$ и из оставшихся двух кусков составить центрально-симметричную выпуклую фигуру F'. При этом, очевидно, $S(F') < S(F)$ и $\partial F' \approx \partial F$ (т. е. их границы стабильно равносоставлены). Повторив эту операцию столько раз, сколько возможно, выбирая каждый раз новую пару отрезков, мы получаем центрально-симметричную выпуклую фигуру F_1 без отрезков прямой на границе такую, что $S(F_1) \leqslant S(F)$ и $\partial F_1 \approx \partial F$.

Предположим, на границе F' есть две пары центрально-симметричных угловых точек A, A' и B, B'. Разрезав F' отрезками AB, BA', $A'B'$ и $B'A$,

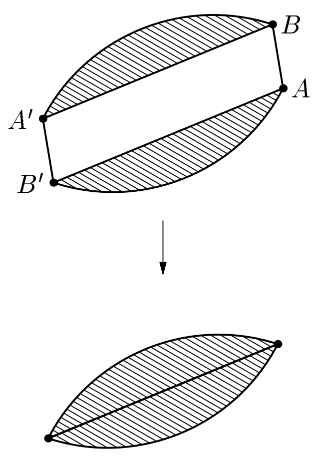


Рис. 2. *Вырезание параллелограмма*

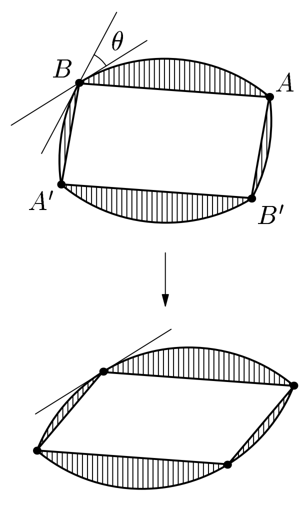


Рис. 3. *Четырехшарнирный метод*

мы получаем параллелограмм $ABA'B'$ и четыре горбушки (см. рис. 3). Не умаляя общности, можно считать, что угол $\angle ABA'$ тупой. Пусть θ есть внешний угол границы F при B. Представьте, что в вершины параллелограмма вставлены шарниры, а стороны сделаны из жесткого материала, при этом горбушки жестко приделаны к сторонам. Давайте повернем сторону AB вокруг B на угол θ так, чтобы угол $\angle ABA'$ увеличился. Тогда площадь параллелограмма $ABA'B'$, а значит и площадь фигуры F_1', образованной параллелограммом и четырьмя горбушками, уменьшится; при этом точки B и B' перестанут быть угловыми.

Повторив эту процедуру столько раз, сколько возможно, выбирая каждый раз новую пару пар центрально-симметричных угловых точек, мы получим фигуру F'' такую, что:

а) F'' имеет не более одной пары угловых точек;

б) F'' центрально-симметрична;

в) $S(F'') \leqslant S(F')$;

г) граница F'' равносоставлена границе F' (в частности, $\partial F'' \approx \partial F$ и у F'' нет отрезков прямой на границе).

Из этих условий нетрудно видеть, что $F'' \cong L_\omega$. Таким образом,

$$S(F) \geqslant S(F') \geqslant S(F'') = S(L_\omega).$$

Отметим, что F'' была получена из F как результат последовательности шагов, в каждом из которых площадь уменьшалась. Значит, в случае равенства $S(F) = S(L_\omega)$ мы вовсе не делали ни одного шага, то есть $F \cong F' \cong F'' \cong L_\omega$. $\square$

Доказательство леммы 10. Пусть F есть фигура описанного типа. Проведем две параллельные прямые a и b так, что вся F лежит в полосе между ними, и обе прямые касаются F. Обозначим через A и B точки пересечения F соответственно с a и b (если прямая соприкасается с F вдоль отрезка, то следует выбрать произвольную точку на отрезке). Пусть α и β обозначают соответственно общую угловую величину всех дуг окружностей на участках границы F от A до B и от B до A против часовой стрелки.

Упражнение 12. Покажите, что пару A, B можно выбрать с дополнительным свойством, что $\alpha = \beta$.

Подсказка. Для этого надо просто прокрутить пару A, B непрерывно вокруг F, так, чтобы A перешло бы в B, а B в A, и воспользоваться тем, что непрерывная функция на отрезке принимает все промежуточные значения. (Мы советуем посмотреть статью Болтянского и Савина [3], в которой обсуждаются геометрические задачи, в решении которых применяются схожие идеи.)

Теперь мы можем считать, что $\alpha = \beta = \omega/2$. Разрежем F отрезком AB на две части F_- и F_+. Из двух копий F_- можно составить центрально-симметричную фигуру $\widetilde{F}_-$. Аналогично поступим с F_+ и получим фигуру $\widetilde{F}_+$.

Согласно лемме 11 каждая из фигур $\widetilde{F}_-$ и $\widetilde{F}_+$ имеет площадь не меньше, чем линза L_ω. А значит

$$S(F) = S(F_+) + S(F_-) = \frac{1}{2}(S(\widetilde{F}_+) + S(\widetilde{F}_-)) \geqslant S(L_\omega)$$

(см. также рис. 4).

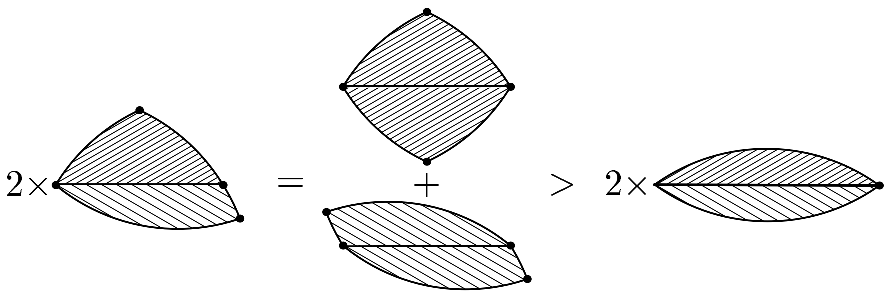


Рис. 4.

Отметим, что равенство может достигаться только в случае, если обе фигуры $\widetilde{F}_-$ и $\widetilde{F}_+$ конгруэнтны L_ω, а значит F имеет не более двух угловых точек и не имеет отрезков прямой на границе, а значит $F \cong L_\omega$. □

4. Круговые многоугольники

Мы приступаем к изучению круговых многоугольников. Начнем с определений:

Определение 13. Выпуклая фигура называется *круговым многоугольником*, если ее граница равносоставлена конечному набору дуг окружностей и отрезков прямой.

Определение 14. Круговой многоугольник M называется *овальным* (или *овалом*), если он обладает следующими свойствами:

а) не имеет отрезков прямой на границе;

б) имеет две перпендикулярные оси симметрии a и b (их точки пересечения с границей M будем обозначать через A, A' и B, B', как показано на рис. 5);

в) если граница M имеет угловые точки, то они находятся на оси a, т. е. в точках A и A';

г) радиусы дуг окружностей, ограничивающей M, монотонно возрастают от A до B.

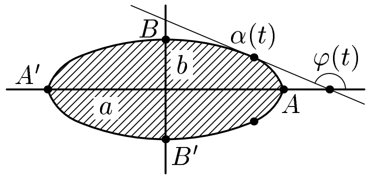


Рис. 5. *Овал*

Легко видеть, что круговой многоугольник овален тогда и только тогда, когда овальна его круговая окрестность.

Теорема 15. *Круговой многоугольник уникальносоставлен тогда и только тогда, когда он овален.*

В следующем разделе мы обобщим этот результат на случай всех выпуклых фигур.

Нетрудно видеть, что в классе круговых многоугольников со стабильно равносоставленной границей есть единственный, с точностью до конгруэнтности, овал. Как и в случаях круга и линзы, теорема 11 для круговых многоугольников немедленно следует из аналога леммы 10:

Лемма 16. *Если круговой многоугольник M имеет границу, стабильно равносоставленную границе овала V, то*

$$S(M) \geqslant S(V).$$

Более того, если $S(M) = S(V)$, то $M \cong V$.

Таким образом, овалы обладают следующим экстремальным свойством: среди всех круговых многоугольников со стабильно равносоставленной границей овалы имеют наименьшую площадь.

Как и в случае с линзой, доказательство будет проведено в два шага; сначала мы докажем аналогичное более слабое утверждение:

Лемма 17. *Если центрально-симметричный круговой многоугольник M имеет границу, стабильно равносоставленную границе овала V, то*

$$S(M) \geqslant S(V).$$

Более того, если $S(M) = S(V)$, то $M \cong V$.

Доказательство леммы 17 проходит индукцией по количеству дуг различных радиусов на границе M, при этом мы считаем угловые точки за дуги нулевого радиуса. Доказательство для круга можно принять за базу индукции, а доказательство для линзы можно рассматривать как часть первого шага в этой индукции (на границе линзы присутствуют дуги радиуса R и радиуса 0). Тем не менее, в доказательстве участвует еще пара дополнительных идей, которые нам не были нужны раньше.

В частности, нам потребуются следующие утверждения, доказательство которых мы предоставляем читателю:

Упражнение 18. Пусть F и G суть выпуклые фигуры с равносоставленной границей, а F_R и G_R — их R-окрестности; тогда $F \sim G$ тогда и только тогда, когда $F_R \sim G_R$.

Упражнение 19. Если выпуклая фигура F имеет границу, равносоставленную границе овала V, то диаметр F не превосходит диаметра V. Более того, в случае равенства диаметров, F конгруэнтна V.

Доказательство леммы 17. Пусть M есть круговой многоугольник с n различными радиусами дуг на границе.

Как и в случае линзы, применив несколько раз «вырезание параллелограмма», мы сводим задачу к случаю, когда на границе M отсутствуют отрезки прямой. В частности, мы можем считать что граница M, уже не только стабильно равносоставлена, но и равносоставлена границе V.

Если на границе кругового многоугольника M нет угловых точек, то существует круговой многоугольник M', такой что $M = M'_R$ (напомним, что M'_R обозначает R-окрестность M'). Таким образом, из упражнения 18 следует, что достаточно доказать лемму для случая, когда на границе M есть угловые точки.

Если у M более одной пары угловых точек, то можно применить несколько раз «Четырехшарнирный метод» и свести задачу к случаю, когда на границе M есть ровно две угловые точки.

Далее, пусть A, A' есть единственная пара угловых точек M, пусть α есть внешний угол при каждой из этих точек, и r_1 есть минимальный ненулевой радиус дуг на границе M. Разрежем границу M на две дуги по точкам A и A' и продолжим оба конца первой из дуг двумя дугами радиуса r_1 с угловой величиной α. К полученной дуге можно приставить вторую из дуг от A до A' до образования выпуклой замкнутой кривой без угловых точек (рис. 6). Обозначим через $\bar{M}$ фигуру, ограниченную полученной кривой.

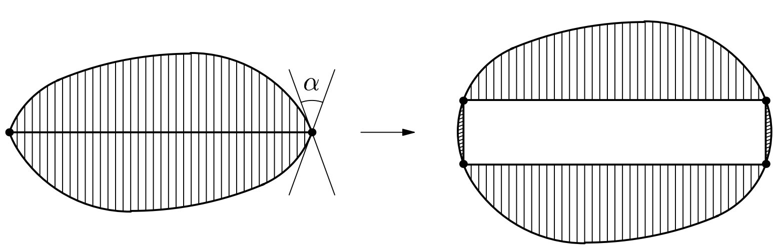


Рис. 6.

Проделаем ту же операцию с овалом V, получим овал $\bar{V}$ с границей равносоставленной $\bar{M}$.

Заметим что при этом

$$S(\bar{V}) - S(V) \geqslant S(\bar{M}) - S(M),$$

причем равенство достигается только если M конгруэнтна V.

Действительно: V получается из $\bar{V}$ добавлением двух круговых сегментов и прямоугольника с одной стороной, равной диаметру $\bar{V}$, и другой, равной $2r_1 \sin(\alpha/2)$. С другой стороны, M получается из $\bar{M}$ добавлением тех же двух сегментов и параллелограмма с одной стороной, не превосходящей диаметра $\bar{V}$ (в силу упражнения 19), и другой стороной, также равной $2r_1 \sin(\alpha/2)$.

Теперь мы получили круговой многоугольник $\bar{M}$ с $n-1$ различными радиусами дуг на границе. По предположению индукции,

$$S(\bar{M}) \geqslant S(\bar{V}).$$

Сложив эти неравенства, получаем

$$S(M) \geqslant S(V).$$

Случай равенства возможен, только если диаметр M равен диаметру V. Значит, в силу упражнения 19, $M \cong V$. □

Следующее доказательство очень похоже на доказательство леммы 10 для случая линзы, но в нём участвует еще один забавный трюк.

Доказательство леммы 16. Пусть M есть круговой многоугольник. Пусть A и B суть точки на границе M, такие, что через них можно провести параллельные прямые так, что вся M будет лежать в полосе между ними. Разрежем M отрезком AB на две части M_- и M_+. Из двух копий M_- составим центрально-симметричную выпуклую фигуру $\widetilde{M}_-$. Аналогично поступим с M_+ и получим фигуру $\widetilde{M}_+$. Давайте обозначим через $\widetilde{V}_+$ и $\widetilde{V}_-$ овалы с границей, стабильно равносоставленной соответственно $\widetilde{M}_+$ и $\widetilde{M}_-$.

Диаметры овалов $\widetilde{V}_+$ и $\widetilde{V}_-$ непрерывно зависят от выбора точек A и B. Таким образом, прокручиванием A и B вокруг M, также как в доказательстве для линзы, можно добиться такого выбора A и B, что диаметры $\widetilde{V}_-$ и $\widetilde{V}_+$ совпадут.

Разрезав $\widetilde{V}_-$ двумя осями симметрии, мы получим четыре четвертинки, конгруэнтные, скажем $\widetilde{V}_{-1/4}$. Аналогично поступим с $\widetilde{V}_-$, получим четыре четвертинки, конгруэнтные $\widetilde{V}_{+1/4}$.

Из двух копий $\widetilde{V}_{-1/4}$ и двух копий $\widetilde{V}_{+1/4}$ можно составить центрально-симметричный круговой многоугольник N, как показано на рис. 7:

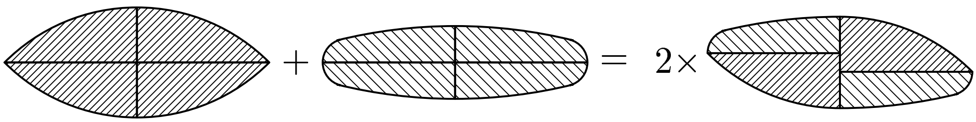


Рис. 7.

Применив лемму 17 к N, получаем:

$$S(M) = S(M_+) + S(M_-) = \frac{1}{2}(S(\widetilde{M}_+) + S(\widetilde{M}_-)) \geqslant$$
$$\geqslant \frac{1}{2}(S(\widetilde{V}_+) + S(\widetilde{V}_-)) = S(N) \geqslant S(V).$$

Остается проверить, что равенство в обоих этих неравенствах влечет конгруэнтность M и V, что очевидно.

Мы доказали, что любой овальный круговой многоугольник является уникальносоставленной фигурой. Остается еще проверить, что не существует других уникальносоставленных круговых многоугольников. Применив утверждение 5, мы получаем, что каждый круговой многоугольник равносоставлен набору из овала и квадрата. Таким образом, достаточно предъявить две неконгруэнтные выпуклые фигуры, каждая из которых равносоставлена нашему набору из квадрата и овала. Возможность этого построения очевидно следует из теоремы Бойяи – Гервина. □

5. Набросок доказательства общего случая

Это наиболее технический и наименее геометрический раздел статьи. Для его понимания потребуется знание в некотором объеме теории функций вещественного переменного (достаточно первых четырех глав книги [4]), а также элементарных свойств метрики Хаусдорфа (которую мы обозначаем d_H), для компактных подмножеств плоскости (см. например, [8]).

Пусть $\alpha\colon [0,T) \to \mathbb{R}^2$ есть граница выпуклой фигуры, параметризованная длиной в направлении против часовой стрелки. У α определена правая производная $\alpha^+(t)$. Определим угол поворота кривой $\varphi(t)$ как угол от $\alpha^+(0)$ до $\alpha^+(t)$ (измеряемый от 0 до 2π против часовой стрелки). Из выпуклости α следует, что $\varphi(t)$ монотонно возрастающая функция. Это дает возможность определить «верхнюю кривизну» $k(t)$ как верхний предел

$$k(t_0) = \varlimsup_{t\to t_0} \frac{\varphi(t) - \varphi(t_0)}{t - t_0}.$$

Так как возрастающая функция имеет производную в почти всех точках, мы получаем, что $k(t)$ конечна для почти всех t.

Определим «нижний радиус кривизны» через $R(t) = 1/k(t)$ (считая $0 = 1/\infty$). Это дает возможность обобщить определение овала (определение 14) на случай произвольной выпуклой фигуры F. Следует только поменять пункт *г*) на следующий:

г̌) нижний радиус кривизны кривой, ограничивающей F, монотонно возрастают от A до B.

Это свойство можно переформулировать так: если вы едете на машине от A до B по границе F, то вам придется всё время крутить руль только вправо. Ваши колёса всё время будут повернуты влево, иначе F была бы не выпуклой, но сам руль вам придется крутить вправо, уменьшая угол между передними и задними колесами.

В случае, когда граница F — гладкая кривая, условие *г̌*) совпадает с монотонным возрастанием обычного радиуса кривизны от A до B.

Мы, наконец, готовы сформулировать главный результат этой статьи:

Теорема 20. *Выпуклая фигура уникальносоставлена тогда и только тогда, когда она овальна.*

В доказательстве нам потребуется понятие «профиль выпуклой фигуры». Пусть F есть выпуклая фигура. Обозначим через $\rho_F(r)$ угловую меру участка границы F с нижним радиусом кривизны $\leqslant r$. Функция $\rho_F\colon \mathbb{R}_+ \to [0, 2\pi]$ называется профилем фигуры F.

Например, профиль круга радиуса R есть

$$\rho_K(r) = \begin{bmatrix} 0 & \text{если} & r < R \\ 2\pi & \text{если} & r \geqslant R, \end{bmatrix}$$

а профиль любого многоугольника тождественно равен 2π.

В случае круговых многоугольников профиль границы есть ступенчатая функция. Равенство профилей границ круговых многоугольников равносильно стабильной равносоставленности их границ. В общем случае это уже не так. Тем не менее, «равенство профилей» выпуклых фигур работает почти так же, как «стабильная равносоставленность границ», т. е. верен следующий аналог утверждения 1:

Утверждение 21. *Следующие два свойства пары выпуклых фигур F и G равносильны:*

(а) F и G имеют равные профили границ и равную площадь,

(б) для любого $\varepsilon > 0$ найдется пара фигур F' и G', ε-близких по Хаусдорфу соответственно к F и G и таких, что $F \sim G'$ и $G \sim F'$.

Кроме того, легко видеть, что в классе выпуклых фигур с равным профилем (не равным тождественно 2π) найдется единственный, с точностью до конгруэнтности, овал.

Также как в случае круговых многоугольников, теорема 20 следует из экстремального свойства овалов, аналогичного лемме 16:

Лемма 22. *Если выпуклая фигура F имеет профиль, равный профилю овала V, то $S(F) \geqslant S(V)$. Более того, если $S(F) = S(V)$, то $F \cong V$.*

Набросок доказательства. Произвольную выпуклую фигуру F можно приблизить последовательностью круговых многоугольников M_i, так что профили M_i сходятся к профилю F. (Это почти равносильно тому, что измеримую функцию на отрезке можно приблизить (по мере) ступенчатой функцией.)

Пусть V_i обозначают овалы, соответствующие M_i. Нетрудно видеть, что V_i сходятся по Хаусдорфу к некоторому овалу V. Более того, полученный овал V имеет тот же профиль, что F. Из леммы 16 получаем $S(V_i) \leqslant S(M_i)$ и, перейдя к пределу, $S(V) \leqslant S(F)$. Остается только показать, что в случае равенства V и F конгруэнтны. Для этого надо слегка уточнить лемму 16:

Лемма 23. *Для любого $\varepsilon > 0$ существует $\delta > 0$ такое, что, если M есть круговой многоугольник, и V есть соответствующий ему овал, то $d_H(V, M) > \varepsilon h$ влечет $S(M) - S(V) > \delta h^2$, где h обозначает ширину овала V.*

Доказательство состоит из той же последовательности шагов, что и доказательство сильного варианта леммы 3. □

6. Продвинутые упражнения

1. Аффинная уникальносоставленность. Понятие равносоставленности можно также рассматривать для различных групп преобразований плоскости, как, например, для группы параллельных переносов или группы подобий. Довольно интересным случаем является группа $SL(2, \mathbb{R})$ эквиаффинных преобразований, т. е. аффинных преобразований, сохраняющих площадь. В этом случае, такими аффинно-уникальносоставленными фигурами являются только эллипсы.

Для решения этого упражнения вам понадобится понятие «аффинной длины», (см., например, книгу Фейеша Тота [9].)

2. Уникальносоставленные наборы фигур. Уникальносоставленность можно обобщить на наборы из n выпуклых фигур: *набор из n выпуклых фигур называется уникальносоставленным, если из равносоставленности этого набора второму набору из n выпуклых фигур следует, что каждая из фигур первого набора конгруэнтна одной из фигур второго.*

Набор из n выпуклых фигур уникальносоставлен тогда и только тогда, когда он состоит из конгруэнтных между собой овалов.

3. Уникальносоставленные тела.

а) Докажите, что для любого выпуклого тела в $\mathbb{R}^3$ найдется произвольно близкое (в метрике Хаусдорфа) уникальносоставленное выпуклое тело.

б) Попробуйте найти точный смысл слов «почти все» и доказать, что «почти все выпуклые тела в трехмерном пространстве являются уникальносоставленными». Для этого вам придется узнать, что такое категория Бэра; это можно сделать, почитав книжку [5].

4. Равносоставленность неограниченных фигур. Эту задачу предложил нам А. В. Гиль.

Равносоставленность можно определить для неограниченных фигур, допуская разрезания вдоль конечного числа лучей и отрезков прямых.

а) Сформулируйте и докажите аналог теоремы Бойяи – Гервина для *бесконечных многоугольников*, т. е. для фигур, ограниченных конечным числом лучей и отрезков. Иначе говоря, найдите необходимое и достаточное условие равносоставленности двух бесконечных многоугольников.

б) Докажите, что если бесконечная выпуклая фигура F уникальносоставлена, то из каждой ее точки исходит ровно один луч, целиком содержащийся в F.

в) Опишите все бесконечные уникальносоставленные выпуклые фигуры.

Авторы выражают благодарность А. В. Гилю, обратившему их внимание на класс задач, связанных с равносоставленностью неограниченных фигур, и Ф. В. Петрову за полезные обсуждения и редактирование окончательного варианта текста.

Список литературы

[1] Болтянский В. Г. *Третья проблема Гильберта*. М.: Наука, 1977.

[2] Болтянский В. Г. *Равновеликие и равносоставленные фигуры*. М.: Гостехиздат, 1956. «Популярные лекции по математике», вып. 22.

[3] Болтянский В. Г., Савин А. П. *Соображения непрерывности и крах гипотезы Борсука* // Квант №3, 1994, с. 3–7.

[4] Натансон И. П. *Теория функций вещественной переменной*. М.: Наука, 1974.

[5] Окстоби Дж. *Мера и категория*. М.: Мир, 1974.

[6] Петрунин А. М., Рукшин С. Е. *Уникальносоставленные выпуклые фигуры* // IX Всесоюзная геометрическая конференция. Кишинев: Штиинца, 1988. С. 242–243.

[7] Петрунин А. М., Рукшин С. Е. *О равносоставленности круговых многоугольников* // Задачи геометрии в целом для погруженных многообразий. СПб: РГПУ, 1991. С. 111–118.

[8] Скворцов В. А. *Примеры метрических пространств*. М.: МЦНМО, 2002. Библиотека «Математическое просвещение», вып. 16.

[9] Фейеш Тот Л. *Расположения на плоскости, на сфере и в пространстве*. М.: Физматлит, 1958.

[10] Ященко И. В. *Парадоксы теории множеств*. М.: МЦНМО, 2002. Библиотека «Математическое просвещение», вып. 20.

[11] Banach S., Tarski A. *Sur la decomposition des ensembles de points en parties respectivement congruentes* // Fundamenta Mathematicae, 1924. Vol. 6. P. 244–277.

[12] Dubins L., Hirsch M., Karush J. *Scissor congruence* // Israel J. Math., 1963. Vol. 1. P. 239–247.

[13] Rukshin S. *Some remarks about Hilbert' third problem and equidecomposable convex figures* // Bull. Math. de la Soc. Sci. Math. de Roumanie, 1998. Vol. 41(89), No 4. P. 285–288.

А. М. Петрунин, Penn State University, USA

С. Е. Рукшин, РГПУ им. Герцена, Городской математический центр для одаренных школьников, Санкт-Петербург
E-mail: `serger@twell.ru`

Примеры трансцендентных чисел

А. Каибханов А. Скопенков*

Wenn Sie unterrichten und davon ausgehen, dass sich parallele Linien im Unendlichen berühren, ergibt sich doch, das müssen Sie zugeben, so etwas wie Transzendenz.

G. Grass, *Katz und Maus*

Если Вы учите, что параллельные линии пересекаются в бесконечности, то Вы всё же должны согласиться: получается нечто непостижимое.

Г. Грасс, *Кошка и мышка*

Приводится простое доказательство трансцендентности числа Лиувилля и новое простое доказательство трансцендентности числа Малера.

Введение

Число x называется *трансцендентным*, если оно не является корнем алгебраического уравнения

$$a_t\lambda^t + a_{t-1}\lambda^{t-1} + \dots + a_1\lambda + a_0 = 0$$

с целыми коэффициентами $a_t \neq 0,\ a_{t-1}, \dots, a_0$.

В университете или даже в старших классах изучается теоретико-множественное доказательство существования трансцендентных чисел [2, Гл. 2, §6]. Это доказательство не дает конкретного примера трансцендентного числа. Приведение *явных* примеров трансцендентных чисел и доказательство их трансцендентности является более трудным материалом, который не всегда входит даже в программу университетского

*А. Скопенков поддержан Российским Фондом Фундаментальных Исследований, гранты номер 05-01-00993 и 04-01-00682, грантами Президента РФ НШ-1988.2003.1 и МД-3938.2005.1, а также стипендией П. Делиня, основанной на его Премии Бальзана 2004 года.

курса. В данной заметке мы приведем такие примеры и доказательства их трансцендентности, которые будут понятны даже старшеклассникам.

Этот абзац предназначен читателям, не знакомым с трансцендентными числами. Ясно, что любое рациональное число не является трансцендентным. Число y называется *алгебраическим*, если оно не трансцендентно. Поэтому алгебраические числа — нечто промежуточное между рациональными числами и произвольными вещественными числами. Зачем нужно доказывать трансцендентность чисел? Одной из мотивировок является то, что всякое *построимое* число (т. е. число, которое может быть построено с помощью циркуля и линейки) является алгебраическим. Таким образом, любое трансцендентное число не построимо [2, Гл. 2, §6 и Гл. 3, §3].

Первый явный пример трансцендентного числа был приведен Жозефом Лиувиллем в 1835 г. [2, Гл. 2, §6].

Теорема Лиувилля. *Число* $\lambda = \sum\limits_{n=0}^{\infty} 2^{-n!}$ *трансцендентно.*

Лиувилль доказал также более общий результат. Приводимая ниже формулировка общей теоремы Лиувилля используется в настоящем тексте *только* для мотивировки нижеследующей теоремы Малера.

Общая теорема Лиувилля. *Пусть z — алгебраическое число, являющееся корнем многочлена степени $n > 1$. Тогда z не может быть приближено рациональным числом $\frac{p}{q}$ с точностью лучшей, чем $\frac{1}{q^{n+1}}$; другими словами, при достаточно больших целых q непременно выполняется неравенство* $\left|z - \frac{p}{q}\right| > \frac{1}{q^{n+1}}$. [2, Гл. 2, §6].

В 1929 г. Курт Малер доказал трансцендентность следующего числа [4].

Теорема Малера. *Число* $\mu = \sum\limits_{n=0}^{\infty} 2^{-2^n}$ *трансцендентно.*

Трансцендентность этого *числа Малера* не следует из общей теоремы Лиувилля, а также из теорем Туэ, Зигеля и Рота [2, Гл. 2, §6], [3]. В работе [4] был получен более общий результат. Доказательство в [4] (так же как и в [5]) неэлементарно и длинно (ср. [1]).

Главный результат данной заметки — *короткое элементарное доказательство трансцендентности числа Малера* (основанное на двоичной записи). Видимо, это доказательство является новым. Оно представлено в разделе «Доказательство теоремы Малера» и не использует остальной части заметки. Но для удобства читателей мы представляем некоторые идеи этого доказательства в разделах «Идея простого доказательства теоремы Малера» и «Основное наблюдение для доказательства теоремы

Малера». Наши идеи дают более общий результат, который приводится в последнем разделе.

Мы также приведем короткое элементарное доказательство трансцендентности числа Лиувилля. Это доказательство представлено в следующем разделе и хорошо известно специалистам, однако мы не нашли его в опубликованном виде.

Данная заметка была представлена в 2002 г. А. Каибхановым на международной конференции Intel ISEF (США, Луисвилль), а также И. Никокошевым и А. Скопенковым на Летней Конференции Турнира Городов (Россия, Белорецк). Мы выражаем благодарность А. Галочкину и Д. Лешко за полезные обсуждения.

Доказательство теоремы Лиувилля

Обозначим $\lambda_s = \sum\limits_{n=0}^{s} 2^{-n!}$.

Сначала докажем, что *число Лиувилля λ иррационально*. Предположим, напротив, что существует линейный многочлен $f(x) = bx + c$ с целыми коэффициентами $b \neq 0$ и c такой, что $f(\lambda) = 0$. Заметим, что это уравнение имеет только один корень, значит $f(\lambda_s) \neq 0$ для некоторого $s > |b|$. Мы получим противоречие из следующих неравенств для некоторого $s > |b|$:

$$2^{-s!} \leqslant |f(\lambda_s)| = |f(\lambda) - f(\lambda_s)| = |b| \cdot (\lambda - \lambda_s) < 2|b| \cdot 2^{-(s+1)!}.$$

Первое неравенство верно, так как $f(\lambda_s) \neq 0$ может быть представлено как дробь со знаменателем $2^{s!}$. Последнее неравенство верно, так как

$$0 < \lambda - \lambda_s < 2^{-(s+1)!} \sum_{n=0}^{\infty} 2^{-n} = 2 \cdot 2^{-(s+1)!}.$$

Теперь докажем, что *число Лиувилля λ не является квадратичной иррациональностью*, т. е. не является корнем квадратного уравнения $f(x) = ax^2 + bx + c = 0$ с целыми коэффициентами $a \neq 0$, b и c. Предположим, напротив, что λ является корнем такого уравнения. Так как квадратное уравнение имеет не более двух корней, то $f(\lambda_s) \neq 0$ для достаточно больших s. Тогда для достаточно больших s мы получим противоречие из следующих неравенств:

$$2^{-2s!} \leqslant |f(\lambda_s)| = |f(\lambda) - f(\lambda_s)| = (\lambda - \lambda_s) \cdot |a(\lambda + \lambda_s) + b| < \\ < (2|a|\lambda + |b|) \cdot 2 \cdot 2^{-(s+1)!}.$$

Первое неравенство верно, так как $f(\lambda_s) \neq 0$ может быть представлено как дробь со знаменателем $2^{2s!}$. Последнее неравенство доказывается аналогично случаю линейного многочлена.

Наконец, приведем доказательство *трансцендентности числа Лиувилля* λ. Предположим, напротив, что число λ является корнем алгебраического уравнения $f(x) = a_t x^t + a_{t-1}x^{t-1} + \dots + a_1 x + a_0 = 0$ с целыми коэффициентами $a_0, \dots, a_{t-1}, a_t \neq 0$. Так как это уравнение имеет лишь конечное число корней, то $f(\lambda_s) \neq 0$ для достаточно больших s. Тогда для достаточно больших s мы получим противоречие из следующих неравенств:

$$2^{-ts!} \leqslant |f(\lambda_s)| = |f(\lambda) - f(\lambda_s)| = (\lambda - \lambda_s) \cdot \Big| \sum_{0 \leqslant i < n \leqslant t} a_n \lambda^{n-1-i} \lambda_s^i \Big| < C \cdot 2^{-(s+1)!},$$

где C не зависит от s (но зависит от коэффициентов многочлена f). Первое неравенство верно, так как $f(\lambda_s) \neq 0$ может быть представлено как дробь со знаменателем $2^{ts!}$. Последнее неравенство доказывается аналогично случаю многочленов первой и второй степени. $\square$

Идея простого доказательства теоремы Малера

Мы продемонстрируем идею доказательства на следующем примере. Мы докажем, что число Лиувилля

$$\nu = \sum_{n=0}^{\infty} 10^{-2^n} = 0{,}11010001000000010\dots$$

не является *квадратичной иррациональностью*, т.е. корнем квадратного уравнения $x^2 + bx + c = 0$ с целыми коэффициентами b и c. (Аналогично доказывается, что число Малера μ не является квадратичной иррациональностью.) Рассмотрим десятичную запись числа $-b\nu - c$ для некоторых целых b и c одного знака (случай различных знаков доказывается аналогично). Рассмотрим ненулевые цифры в этой десятичной записи, расположенные достаточно далеко от запятой. Ясно, что они образуют «сгустки» около позиций с номерами 2^n: каждый «сгусток» представляет число $|b|$. Например для $b = -17$ мы имеем следующее:

$$17\nu - c = \dots{,}87170017000000170\dots017\dots.$$

С другой стороны,

$$\nu^2 = \sum_{k,l=0}^{\infty} 10^{-2^k-2^l} = 0{,}0121220\dots12202000200000012\dots.$$

Таким образом, в десятичной записи числа ν^2 ненулевые цифры расположены около позиций с номерами $2^k + 2^l$. Но для достаточно больших l и $k = 2l$ на этих же позициях числа $-b\nu - c$ стоят нули. Следовательно $\nu^2 \neq -b\nu - c$. Значит, ν не является квадратичной иррациональностью.

Основное наблюдение для доказательства теоремы Малера

Мы продемонстрируем одну из наших главных идей на следующем примере. Мы докажем, что *число* $E = \sum\limits_{n=0}^{\infty} \frac{a_n}{n!}$ *иррационально для любой ограниченной последовательности* $\{a_n\}_{n=0}^{\infty}$ *ненулевых целых чисел.* (Заметим, что $e = \sum\limits_{n=0}^{\infty} \frac{1}{n!}$.)

Предположим, напротив, что E рационально. Тогда существуют сколь угодно большие q, для которых

$$E = \frac{p}{q} = \sum_{n=0}^{q} \frac{a_n}{n!} + \frac{a_{q+1}}{(q+1)!} + \sum_{n=q+2}^{\infty} \frac{a_n}{n!}\,.$$

Поэтому

$$p(q-1)! - \sum_{n=0}^{q} \frac{q!a_n}{n!} = \frac{a_{q+1}}{q+1} + \sum_{n=q+2}^{\infty} \frac{q!a_n}{n!}$$

целое число. Но это невозможно, так как для достаточно больших q имеем

$$\left|\sum_{n=q+2}^{\infty} \frac{q!a_n}{n!}\right| \leqslant C \cdot \sum_{n=q+2}^{\infty} \frac{q!}{n!} < C \cdot \sum_{n=2}^{\infty} \frac{1}{(q+1)^n} = \frac{C}{(q+1)q} < \frac{1}{q+1} \leqslant \left|\frac{a_{q+1}}{q+1}\right| < \frac{1}{2}\,.$$

□

Доказательство теоремы Малера

Возьмем многочлен $f(x) = a_t x^t + a_{t-1}x^{t-1} + \dots + a_1 x + a_0$ с целыми коэффициентами $a_t \neq 0,\ a_{t-1}, \dots, a_0$. Выберем достаточно большое число p и положим $s = 2^p(2^t - 1) - 2^{p-1}$. (В двоичной записи $s = 1\ldots1010\ldots0$ с t единицами и p нулями). Достаточно доказать, что

дробная часть $\{2^s f(\mu)\}$ *не равна нулю.*

(Это число является «s-хвостом» десятичной записи числа $f(\mu)$: его двоичное разложение получено из двоичного разложения числа $f(\mu)$ отбрасыванием всех знаков слева от s-го знака после запятой.)

Раскрывая скобки, получим

$$\mu^q = \left(\sum_{n=0}^{\infty} 2^{-2^n}\right)^q = \sum_{n=0}^{\infty} d_n(q) 2^{-n},$$

где $d_n(q)$ есть количество упорядоченных представлений числа n в виде суммы q степеней двойки (не обязательно различных степеней). Другими словами, $d_n(q)$ является количеством упорядоченных наборов $(w_1, \dots, w_q)$ длины q таких, что $n = 2^{w_1} + \dots + 2^{w_q}$ (возможно $w_i = w_j$). Например, $d_3(2) = 2$, поскольку $3 = 2^0 + 2^1 = 2^1 + 2^0$ и $d_0(0) = 1$.

Имеем

$$f(\mu)=\sum_{n=0}^{\infty} d_n 2^{-n}, \quad \text{где} \quad d_n = a_t d_n(t) + a_{t-1} d_n(t-1) + \dots + a_0 d_n(0).$$

Ясно, что $d_n(q)=0$ тогда и только тогда, когда n имеет более q единиц в двоичной записи. Обозначим через U_q множество чисел n, имеющих не более q единиц в двоичной записи. Тогда $d_n=0$ для $n\notin U_q$.

Число $s_1 = 2^p(2^t-1)$ — наименьшее число из U_t, большее, чем s (в двоичной записи $s_1 = 1\dots10\dots0$ с t единицами и p нулями). Число $s_2 = 2^{t+p}$ — наименьшее число из U_t, большее, чем s_1. Следовательно

$$\{2^s f(\mu)\} = \left\{\sum_{n=0}^{\infty} d_n 2^{s-n}\right\} = \left\{d_{s_1}2^{s-s_1} + \sum_{n=s_2}^{\infty} d_n 2^{s-n}\right\}.$$

Тогда

$$\{2^s f(\mu)\}\neq 0, \quad \text{поскольку} \quad \left|\sum_{n=s_2}^{\infty} d_n 2^{s-n}\right| \overset{(1)}{<} 2^{s-s_1}|d_{s_1}| \overset{(2)}{<} 1/2.$$

Для доказательства неравенств (1) и (2) нам понадобится следующая лемма, которую мы докажем позже.

Лемма о представлении. *Количество $d_n(q)$ упорядоченных представлений числа n в виде суммы q степеней двойки не превосходит* $(q!)^2$.

Из леммы о представлении получаем, что существует число $D=D(f)$ такое, что $|d_n|\leqslant D$ для каждого n. Следовательно неравенство (2) верно, так как $|d_{s_1}|2^{s-s_1}\leqslant D\cdot 2^{-2^{p-1}} < 1/2$ для достаточно больших p. Неравенство (1) верно, так как для достаточно больших p (аналогично утверждению о $\lambda-\lambda_s$ для числа Лиувилля)

$$\left|\sum_{n=s_2}^{\infty} d_n 2^{s-n}\right| \leqslant D\sum_{n=s_2}^{\infty} 2^{s-n} = D\cdot 2^{s+1-s_2} = D\cdot 2^{s-s_1+1-2^p} <$$
$$< 2^{s-s_1} \leqslant 2^{s-s_1}|d_{s_1}|.$$

Здесь последнее неравенство верно, так как $s_1\in U_t - U_{t-1}$, и $d_{s_1}(q)=0$ для $q<t$, следовательно $d_{s_1} = a_t d_{s_1}(t)\neq 0$. □

Доказательство леммы о представлении. Для $q=0$ имеем $d_0(0)=1\leqslant 0!^2$. Для $q\geqslant 1$ имеем $d_n(q) = \sum\limits_{r=0}^{\infty} d_{n-2^r}(q-1)$. Следовательно, при помощи индукции по q получаем, что достаточно доказать следующее утверждение:

для каждого $n\in U_q$ существует не более q^2 целых $r\geqslant 0$ таких, что $n-2^r\in U_{q-1}$.

Рассмотрим двоичное представление числа $n \in U_q$:

$$n = 2^{w_k} + 2^{w_{k-1}} + \dots + 2^{w_1}, \quad \text{где} \quad w_k > w_{k-1} > \dots > w_1 \geqslant 0 \quad \text{и} \quad k \leqslant q.$$

Обозначим $w_0 = -1$. Так как $n - 2^r < 0$ для $r > w_k$, то достаточно доказать, что

для каждого $i = 1, 2, ..., k$ *существует не более* q *целых* $r \in [w_{i-1} + 1, w_i]$ *таких, что* $n - 2^r \in U_{q-1}$.

Следовательно, достаточно доказать, что

для каждого $i = 1, 2, ..., k$ *и* $r \in [w_{i-1} + 1, w_i - q]$ *имеем* $n - 2^r \notin U_{q-1}$.

Чтобы доказать это утверждение, заметим, что $2^{w_i} - 2^r = 2^{w_i - 1} + 2^{w_i - 2} \dots + 2^r$ есть сумма $w_i - r \geqslant q$ различных степеней двойки, каждая из которых больше $2^{w_{i-1}}$ и меньше $2^{w_i} < 2^{w_{i+1}}$. Поэтому число $n - 2^r$ представляется в виде суммы более чем $q - 1$ различных степеней двойки. Значит, $n - 2^r \notin U_{q-1}$. $\square$

Обобщение

Предположим, что дана строго возрастающая последовательность $\{a_n\}$ натуральных чисел.

Пусть дано натуральное число q. Натуральное число m называется *q-представимым*, если m можно представить в виде суммы не более чем q членов последовательности $\{a_i\}$: $m = a_{i_1} + a_{i_2} + \dots + a_{i_q}$. Эти члены не обязательно различны (например, число $m = a_2 + a_2$ является 2-представимым).

Последовательность $\{a_i\}$ называется *q-разреженной*, если для любого целого M существует три последовательных q-представимых числа a, b и c, промежутки между которыми больше M (т. е. таких, что $b - a > M$ и $c - b > M$).

Последовательность $\{a_i\}$ называется *разреженной*, если она q-разреженная для любого q.

Например, последовательность $a_i = i$ всех целых положительных чисел является 1-представимой, следовательно, эта последовательность не 1-разрежена и тем более не разрежена. При доказательстве теоремы Малера было доказано, что последовательность 2^i является разреженной.

Последовательность $\{a_n\}$ натуральных чисел называется *q-рыхлой*, если количество способов представления числа n в виде суммы q членов этой последовательности (не обязательно различных) не превосходит некоторой константы C_q, не зависящей от n (но, возможно, зависящей от q). Сформулируем это немного по-другому. Для любых натуральных q и n обозначим через $d_n(q)$ количество представлений числа n в виде суммы q (не обязательно различных) членов последовательности $\{a_i\}$ с учетом порядка. Другими словами, $d_n(q)$ — это количество упорядоченных наборов $(a_{i_1}, \dots, a_{i_q})$ с $n = a_{i_1} + \dots + a_{i_q}$. Последовательность $\{a_i\}$ называется *q-рыхлой*, если существует число C_q такое, что $d_n(q) < C_q$ при любом n.

Последовательность $\{a_i\}$ называется *рыхлой*, если она является q-рыхлой для любого натурального q.

Например последовательность $a_i = i$ не рыхлая и даже не 2-рыхлая. Действительно, натуральное число n имеет не менее $n/2-1$ представлений в виде суммы двух натуральных чисел, таким образом $d_n(2) \geqslant n/2 - 1$. Из леммы о представлении вытекает, что последовательность 2^i является рыхлой.

Аналогично приведённому доказательству теоремы Малера можно доказать следующий результат.

Теорема. *Если последовательность $\{a_n\}$ рыхлая и разреженная, то число $\sum\limits_{n=0}^{\infty} 2^{-a_n}$ трансцендентно.*

Было бы интересно обобщить наше доказательство трансцендентности числа Малера и приведённой теоремы до достаточного условия трансцендентности числа, включающего два «хороших» приближения этого числа.

Авторы не знают, являются ли следующие числа трансцендентными (или квадратичными иррациональностями):

(a) $\sum\limits_{n=0}^{\infty} d_n 2^{-2^n}$ для любой ограниченной последовательности d_n целых положительных чисел.

(b) $\sum\limits_{n=0}^{\infty} d_n 2^{-2^n}$ для любой последовательности $0 \leqslant d_n \leqslant n$.

(c) $\sum\limits_{n=0}^{\infty} (-1)^n 2^{-2^n}$.

(d) $\sum\limits_{n=0}^{\infty} \varepsilon_n 2^{-2^n}$ для любой последовательности $\varepsilon_n \in \{\pm 1\}$.

(e) $\sum\limits_{n=0}^{\infty} 2^{-[1{,}1^n]}$.

(f) $\sum\limits_{n=0}^{\infty} 2^{-f_n}$, где $f_{n+2} = f_{n+1} + f_n$, $f_0 = f_1 = 1$ — последовательность Фибоначчи.

Список литературы

[1] Галочкин А. *О мере трансцендентности значений функций, удовлетворяющих некоторым функциональным уравнениям* // Мат. Заметки, 1980. Т. 27, №2. С. 175–183.

[2] Курант Р., Роббинс Г. *Что такое математика*? М.: МЦНМО, 2001.

[3] Фельдман Н. *Алгебраические и трансцендентные числа* // Квант, №7, 1983. С. 2–7.

[4] Mahler K. *Arithmetische Eigenschaften der Lösungen einer Klasse von Funktionalgleichungen* // Mathematische Annalen, 1929. Bd. 1. S. 342–366.

[5] Nishioka K. *Mahler functions and transcendence.* Lecture Notes in Math., 1631. Berlin – New York, 1996.

А. Б. Скопенков, кафедра дифференциальной геометрии и приложений, механико-математический факультет, Московский государственный университет, Москва, 119992, Россия; Независимый московский университет, Б. Власьевский пер., 11, Москва, 119002, Россия.
E-mail: `skopenko@mccme.ru`

А. Каибханов, кафедра дифференциальной геометрии и приложений, механико-математический факультет, Московский государственный университет, Москва, 119992, Россия.
E-mail: `kaib@mail.ru`

Формула для объема пересечения куба с полупространством

А. Я. Белов

В свое время я исследовал разбиение массива горных пород системами трещин. Каждая такая система моделировалась системой параллельных равноотстоящих плоскостей. Случай трех систем тривиален — получается разбиение пространства на равные параллелепипеды. К сожалению разбиения пространства бо́льшим числом систем (мне кажется) почти не рассматривались (хотя такие разбиения должны иметь отношения к квазикристаллам). На практике наблюдался случай четырех систем прямых. Меня интересовало распределение блоков по объемам. С точностью до аффинного преобразования можно считать, что первые три системы разбивают пространство на единичные кубы, четвертая система их доразбивает. Так я пришел к формуле объема пересечения куба с полупространством, которая имеет красивое n-мерное обобщение.

Рассмотрим единичный куб K: $K = \{(x_1, \ldots, x_n) \mid 0 \leqslant x_i \leqslant 1\}$ и полупространство P_d, задаваемое неравенством
$$\sum a_i x_i \leqslant d.$$

Выберем направления осей координат так, чтобы все коэффициенты a_i стали неотрицательными. Для этого поступим так.

Если только один коэффициент a_i отрицательный, то сделаем замену $x_i \to 1 - x_i$; при этом a_i заменится на $-a_i$, куб K перейдет в себя, а свободный член d станет равным $d + a_i$. Если же отрицательных коэффициентов a_i больше одного, поступим аналогично: для множества $I \subset \{1, \ldots, n\}$ индексов i, заданных условием $a_i < 0$, сделаем ту же подстановку $x_i \to 1 - x_i$, оставив все остальные координаты x_i без изменений. Как и раньше, куб K перейдет в себя, а полупространство P_d — в полупространство $P'_d = \{(x_1, \ldots, x_n) \mid \sum_{i \in I} a'_i x_i \leqslant d'\}$. После такого преобразования объемы пересечений $K \cap P_d$ и $K \cap P'_d$ будут одинаковыми, $d' = d + \sum_{i \in I} a_i$, а все коэффициенты $a'_i = |a_i|$ станут неотрицательными, чего мы и добивались.

Далее, если $a_i = 0$ для некоторого i, то полупространство P_d параллельно i-й оси координат и n-мерный объем $K \cap P_d$ равен $(n-1)$-мерному объему пересечения P_d с соответствующей гранью куба, так что ситуация сводится к меньшей размерности. В дальнейшем считаем поэтому, что $a_i > 0$ при всех $i = 1,\ 2,\ \ldots,\ n$.

Разделив обе части неравенства $\sum_{i\in I} a_i x_i \leqslant d$ на $\sqrt{\sum_{i\in I} a_i^2}$, приходим к случаю, когда $\sqrt{\sum_{i\in I} a_i^2} = 1$. В этом случае $V_n(d)$ — объем пересечения $K \cap P_d$ — задается равенством:

$$V_n(d) = \frac{1}{n!\prod_{i=1}^n a_i}\Bigg(\sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|}\Big(d - \sum_{i\in I} a_i\Big)_+^n\Bigg). \tag{1}$$

Здесь все a_i положительны, $x_+ = \max(x, 0)$, а $|I|$ — число элементов множества I; если $I = \varnothing$, то полагаем $|I| = 0$ и $\sum_{i\in I} a_i = 0$.

В частности, если $d \leqslant 0$, то $V_n(d) = 0$, а если $d \geqslant \sum_{i=1}^n a_i$, то $V_n(d) = 1$.

В размерностях $n = 1$, 2 и 3 формулы для объема превращаются в такие. Если $n = 1$, то длина пересечения единичного отрезка и луча $x \leqslant d$ равна

$$V_1(d) = d_+ - (d-1)_+.$$

В двумерном и трехмерном случаях правая часть принимает вид:

$$V_2(d) = \frac{1}{2}(a_1a_2)\cdot\Big(d_+^2 - (d-a_1)_+^2 - (d-a_2)_+^2 + (d-a_1-a_2)_+^2\Big)$$

и

$$\begin{aligned} V_3(d) = \frac{1}{6}(a_1a_2a_3)^{-1}\cdot\Big(&d_+^3 - (d-a_1)_+^3 - (d-a_2)_+^3 - (d-a_3)_+^3 + \\ &+ (d-a_1-a_2)_+^3 \ + \ (d-a_2-a_3)_+^3 \ + \ (d-a_3-a_1)_+^3 \ + \\ &- (d-a_1-a_2-a_3)_+^3\Big), \end{aligned}$$

соответственно.

Из формулы (1) можно получить формулу для площади сечения куба границей полупространства P_d. Для этого заметим, что число d есть расстояние от начала координат O до границы полупространства P_d, а разность объемов $V(d+\Delta d) - V(d) = S(d)\Delta d + o(\Delta)$, где $S(d)$ — искомая площадь сечения куба. Продифференцировав равенство (1) по параметру d, получаем следующую формулу для площади сечения:

$$S(d) = \frac{1}{(n-1)!\prod_{i=1}^n a_i}\left(\sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|}\Big(d - \sum_{i\in I} a_i\Big)_+^{n-1}\right). \tag{2}$$

Равенство нулю $S(d)$ при $d \geqslant \sum a_i$ (т.е. когда граница ∂P_d полупространства не пересекает куба K) означает забавное комбинаторное тождество:

$$\sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|}\Big(d - \sum_{i\in I} a_i\Big)^{n-1} \equiv 0. \tag{3}$$

(Нижние значки «+» сняты, ибо все члены положительны, поэтому левая часть — многочлен от переменных d и a_i, где $i = 1, \ldots, n$. Равенство нулю многочлена в открытой области влечет тождественное равенство нулю).

Доказательство формулы (1)

Отметим, что выражение $\frac{1}{n!}\Big(\prod_{i=1}^n a_i\Big)^{-1} \cdot d_+^n$ характеризует объем пересечения P_d с положительным ортантом

$$Q = \{(x_1, \ldots, x_n) \mid x_i \geqslant 0\}.$$

Величина

$$\frac{1}{n!}\Big(\prod_{i=1}^n a_i\Big)^{-1} \left(d - \sum_{i=1}^n a_i f_i\right)_+^n$$

равна объему пересечения P_d со сдвигом $Q^{\vec{f}}$ положительного ортанта Q на вектор $\vec{f} = (f_1, \ldots, f_n)$ (т.е. с областью $\{(x_1, \ldots, x_n) \mid x_i \geqslant f_i\}$). Пусть V_I^d — объем пересечения P_d с областью

$$U_I = \{(x_1, \ldots, x_n) \mid x_i \geqslant 0,\ i \notin I,\ x_i \geqslant 1,\ i \in I\}.$$

В частности, $U_\varnothing = \{(x_1, \ldots, x_n) \mid x_j \geqslant 0\}$. Тогда формула (1) примет вид:

$$V_n(d) = \sum_{I \subseteq \{1, \ldots, n\}} (-1)^{|I|} V_I^d.$$

Пусть $Q_i = \{(x_1, \ldots, x_n) \mid x_j \geqslant 0,\ j \neq i,\ x_i \geqslant 1\}$. Тогда $U_I = \cap_{i \in I} Q_i$ и формула (1) окажется частным случаем (когда $\Phi = P_d$) формулы

$$\mathrm{Vol}(K \cap \Phi) = \mathrm{Vol}(\Phi \cap U_\varnothing) + \sum_{\substack{1 \leqslant k \leqslant n,\\ 1 \leqslant i_1 < \cdots < i_k \leqslant n}} (-1)^k \cdot \mathrm{Vol}(\Phi \cap Q_{i_1} \cap \ldots \cap Q_{i_k}). \quad (4)$$

Здесь $\mathrm{Vol}(M)$ — объем множества M, Φ — произвольная фигура, пересечение которой с любым сдвигом положительного ортанта ограничено.

Докажем равенство (4). Единичный куб K дополняется до Q объединением множеств Q_i. То же верно и для пересечения соответствующих множеств с Φ. Пусть $\Phi_j = \Phi \cap Q_j$. Тогда $\Phi \cap Q_{i_1} \cap \ldots \cap Q_{i_k} = \Phi_{i_1} \cap \ldots \cap \Phi_{i_k}$, и достаточно проверить, что величина

$$\sum_{k=1}^n (-1)^{k+1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \mathrm{Vol}(\Phi_{i_1} \cap \ldots \cap \Phi_{i_k}).$$

есть объем объединения $\cup_j \Phi_j$. Но это есть частный случай формулы включения–исключения.

Напомним эту формулу. Пусть U_i – семейство множеств, $|U|$ означает объем множества U. Тогда

$$|\cup U_i| = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1\leqslant i_1<\dots<i_k\leqslant n} |U_{i_1} \cap \ldots \cap U_{i_k}|.$$

(В дискретном случае обычно полагают в качестве $|U|$ число элементов. Тогда получается формула из обычной комбинаторики (см. [1].))

Смысл этой формулы таков: чтобы подсчитать число элементов (объем) объединения, надо просуммировать количества элементов (объемы) множеств. Но тогда попарные пересечения окажутся добавленными лишний раз. Их надо вычесть. Но тогда надо сделать поправку на тройные пересечения, и т. д.

Замечания.

1. С каждым множеством U можно связать *характеристическую функцию* ξ_U: $\xi_U(x) = 1$ при $x \in U$, $\xi_U(x) = 0$ при $x \notin U$. Объем множества U равен $\int \xi_U$. Пусть $\Delta_i \colon g(x) \to g - T_{\vec{e}_i}(g)$, где $T_{\vec{f}}(g)(x) = g(x + \vec{f})$. Если $L = [0,1]$, $R = [0,\infty)$ то $\xi_L = \Delta_1(\xi_R)$. Характеристическая функция куба ξ_K получается из характеристической функции положительного ортанта Q применением оператора $\prod_{i=1}^{n} \Delta_i$. Но

$$\prod_{i=1}^{n} \Delta_i(\xi_Q) = (1 - T_{\vec{e}_1}) \cdot \ldots \cdot (1 - T_{\vec{e}_n})(\xi_Q) = \\ = \sum_{\substack{0\leqslant k\leqslant n,\\ 1\leqslant i_1<\dots<i_k\leqslant n}} (-1)^k T_{\vec{e}_{i_1}}(\xi_Q) \cap \ldots \cap T_{\vec{e}_{i_k}}(\xi_Q).$$

Отсюда также следует формула (4).

2. Аналогичным образом с помощью включения–исключения доказывается поляризационная формула:

$$(x_1 + \dots + x_n)^n - \sum_i (x_1 + \dots + \widehat{x_i} + \dots + x_n)^n + \dots + \\ + (-1)^k \sum_{i_1<\dots<i_k} (x_1 + \dots + \widehat{x_{i_1}} + \dots + \widehat{x_{i_k}} + \dots + x_n)^n + \\ + \dots + (-1)^n \sum x_i^n = \sum_{\sigma\in S_n} x_{\sigma(1)} \cdot \ldots \cdot x_{\sigma(n)}.$$

Если переменные x_i коммутируют, то правая часть в этой формуле равна $n!\big(x_1 \cdot \ldots \cdot x_n\big)$.

3. В школе нас приучают избегать крупных формул. Однако правильно организованая формула (своего рода иероглиф) оказывается красивой! А красота служит проверке правильности.

Автор признателен Г. А. Гальперину, побудившему его написать данную заметку и обратившему внимание на задачу о площади сечения.

Список литературы

[1] Виленкин Н. Я. *Комбинаторика*. М.: Наука, 1967.

[2] Белов А. Я. *Статистическая геометрия и равновесие блочных массивов*. Дисс. на соиск. уч. степ. канд. физ. мат. наук. М.: МГИ, 1991.

А. Я. Белов, МИОО
E-mail: `kanel@mccme.ru`

О представлении чисел в виде суммы двух квадратов

М. Н. Вялый

Какие целые числа являются суммами двух квадратов? Ответ на этот вопрос известен: число представляется в виде суммы двух квадратов тогда и только тогда, когда в его разложение на простые множители каждое простое число вида $4k + 3$ входит в четной степени.

Как найти представление числа n в виде суммы двух квадратов? Конечно, можно перебрать все пары $0 \leqslant x, y \leqslant n$ и проверить для каждой пары равенство $x^2 + y^2 = n$. Нельзя ли придумать алгоритм, который решает эту задачу быстрее?

Эффективных алгоритмов для представления числа в виде суммы двух квадратов пока нет. Но задача значительно упрощается, если известен корень из -1 по модулю n. В этом случае можно обойтись без перебора.

Алгоритм Эрмита – Серре.

1. Вход: n, z, причем $z^2 \equiv -1 \pmod{n}$.
2. Полагаем $a_0 = n$, $a_1 = z$.
3. К паре чисел a_0, a_1 применяем алгоритм Евклида, т.е. вычисляем последовательность остатков от деления a_i на a_{i+1}:
$$a_i = q_i a_{i+1} + a_{i+2}\,.$$
4. Останавливаемся в тот момент, когда два последовательных остатка не превосходят $\sqrt{n}$: $a_t \leqslant \sqrt{n}$, $a_{t+1} \leqslant \sqrt{n}$. Это и есть нужная пара чисел:
$$a_t^2 + a_{t+1}^2 = n.$$

Числа в последовательности остатков, возникающей в алгоритме Евклида, быстро убывают. Поэтому вычисления по алгоритму Эрмита – Серре занимают гораздо меньше времени, чем перебор всех возможностей.

Откуда берется связь между алгоритмом Евклида и представлениями числа в виде суммы двух квадратов? Есть два, содержательно очень близких, способа обосновать алгоритм Эрмита – Серре. Первый основан на связи между алгоритмом Евклида и цепными дробями и использует симметричность разложения в цепную дробь числа n/z [2]. Второй предложен А. ван дер Портеном [4] (см. также [3]) и использует симметричные разложения матриц размера 2×2. Ниже это доказательство излагается подробно.

Мы опишем связь между алгоритмом Евклида и матрицами размера 2×2 с помощью функции «скобки», которая определена на последовательностях натуральных (положительных целых) чисел следующими рекуррентными соотношениями:

$$[\,] = 1, \quad [q] = q, \quad [q_0, q_1, q_2, \ldots, q_n] = q_0[q_1, q_2, \ldots, q_n] + [q_2, \ldots, q_n]. \tag{1}$$

Например, $[q_0, q_1] = q_0[q_1] + [\,] = q_0q_1 + 1$.

Непосредственно из определения видно, что эта функция положительна на любой последовательности $q_0, q_1, q_2, \ldots, q_n$ положительных целых чисел. Поэтому

$$[q_0, q_1, \ldots, q_n] > [q_1, \ldots, q_n]. \tag{2}$$

Таким образом, из (1) следует, что применение алгоритма Евклида к паре чисел $[q_0, q_1, q_2, \ldots, q_n]$, $[q_1, q_2, \ldots, q_n]$ дает последовательность остатков $a_i = [q_i, \ldots, q_n]$, при этом последовательностью частных будет $q_0, \ldots, q_n$. Отсюда вытекает следующая

Лемма 1. *Представление пары взаимно простых чисел a, b в виде*

$$a = [q_0, q_1, q_2, \ldots, q_n], \quad b = [q_1, q_2, \ldots, q_n] \tag{3}$$

однозначно определено.

Заметим также, что та же самая последовательность $q_0, q_1, q_2, \ldots, q_n$ возникает при разложении a/b в цепную дробь.

Шаг алгоритма Евклида можно представить в матричном виде

$$\begin{pmatrix} a_i \\ a_{i+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -q_{i-1} \end{pmatrix} \begin{pmatrix} a_{i-1} \\ a_i \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} a_{i-1} \\ a_i \end{pmatrix} = \begin{pmatrix} q_{i-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_i \\ a_{i+1} \end{pmatrix}.$$

Мы будем использовать произведения матриц второго вида. Матричные элементы таких произведений легко выражаются через скобки $[\cdot]$ с помощью матричного равенства

$$\begin{pmatrix} [q_0, \ldots, q_n] & [q_0, \ldots, q_{n-1}] \\ [q_1, \ldots, q_n] & [q_1, \ldots, q_{n-1}] \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} [q_1, \ldots, q_n] & [q_1, \ldots, q_{n-1}] \\ [q_2, \ldots, q_n] & [q_2, \ldots, q_{n-1}] \end{pmatrix}, \tag{4}$$

которое проверяется прямым вычислением. Из (4) по индукции следует, что

$$\begin{pmatrix} [q_0, \ldots, q_n] & [q_0, \ldots, q_{n-1}] \\ [q_1, \ldots, q_n] & [q_1, \ldots, q_{n-1}] \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix}. \tag{5}$$

Из (5) и леммы 1 следует единственность представления матрицы 2×2 вида

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \qquad (6)$$

(конечно, такое представление существует не всегда).

Нам потребуется следующая лемма.

Лемма 2. *Пусть $ad = b^2 + 1$, $a > b \geqslant d > 0$. Тогда существует единственное натуральное q такое, что*

$$\begin{pmatrix} a & b \\ b & d \end{pmatrix} = \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a' & b' \\ b' & d' \end{pmatrix} \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} \qquad (7)$$

и либо $a' > b' \geqslant d' > 0$, либо $b' = 0$, $a' = d' = 1$.

Доказательство. Легко проверить, что из (7) следуют равенства $a' = d$, $b' = b - qd$, $d' = (a - qb) - q(b - qd)$:

$$\begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} d & b - qd \\ b - qd & (a - qb) - q(b - qd) \end{pmatrix} \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} = \\ = \begin{pmatrix} b & a - qb \\ d & b - qd \end{pmatrix} \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ b & d \end{pmatrix}. \qquad (8)$$

Поэтому условия $a' > b' > 0$ означают $d > b - qd > 0$. Таким образом, q должно быть частным от целочисленного деления b на d.

Рассмотрим вначале случай, когда b не делится на d: $b = qd + r$, $0 < r < d$. Поскольку $ad - b^2 = 1$,

$$\det \begin{pmatrix} a' & b' \\ b' & d' \end{pmatrix} = 1 \qquad (9)$$

(эта матрица получается умножением матрицы с определителем 1 на две матрицы с определителем -1). Поэтому из $a' = d > 0$, $b' = r > 0$ следует, что и $d' > 0$. Кроме того, $a' = d > r = b'$. Осталось проверить неравенство $r = b' \geqslant d'$. Если $d' > b'$, то $a'd' \geqslant (b' + 1)^2 > b'^2 + 1$, что противоречит (9).

Теперь рассмотрим случай, когда b делится на d. Тогда

$$\begin{pmatrix} a' & b' \\ b' & d' \end{pmatrix} = \begin{pmatrix} a' & 0 \\ 0 & d' \end{pmatrix}.$$

Определитель этой (диагональной) матрицы равен 1, поэтому $a' = d' = 1$.

Заметим, что лемму 2 можно применять индуктивно: полученная матрица $\begin{pmatrix} a' & b' \\ b' & d' \end{pmatrix}$ либо единичная, либо тоже удовлетворяет условию леммы. При последовательном применении леммы 2 внедиагональные элементы убывают, как видно из (8). Поэтому рано или поздно матрица станет единичной. В результате такого процесса получается *симметричное разложение* исходной матрицы, имеющее вид

$$\begin{pmatrix} a & b \\ b & d \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \tag{10}$$

Теперь вернемся к обоснованию алгоритма Эрмита – Серре. Без ограничения общности можно считать, что $z < n/2$. Тогда $k = (z^2+1)/n \leqslant z$. Поэтому из леммы 2 следует существование симметричного разложения для матрицы

$$\begin{pmatrix} n & z \\ z & k \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{11}$$

Обозначим

$$A = \begin{pmatrix} x & u \\ y & v \end{pmatrix} = \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Тогда

$$\begin{pmatrix} n & z \\ z & k \end{pmatrix} = A^t A = \begin{pmatrix} x & y \\ u & v \end{pmatrix} \begin{pmatrix} x & u \\ y & v \end{pmatrix}.$$

Значит, $n = x^2 + y^2$. Так как разложение вида (6) единственно, то из (5) заключаем, что $x = a_{n+1}$, $y = a_{n+2}$, где a_i — остатки в алгоритме Евклида, примененном к паре $a_0 = n$, $a_1 = z$. Ясно, что $a_{n+k}^2 \leqslant n$ при $k \geqslant 1$. Для завершения обоснования алгоритма Эрмита – Серре осталось проверить, что $a_i^2 > n$ при $i \leqslant n$. Для этого достаточно проверить, что $a_n = q_n a_{n+1} + a_{n+2} > \sqrt{n}$. И действительно,

$$a_n^2 = (q_n x + y)^2 \geqslant (x+y)^2 = n + 2xy > n.$$

Замечания.

1. У цепных дробей есть естественная геометрическая интерпретация (см., например, [1].) Интересно было бы выразить явно геометрический смысл леммы 2.

2. Из (11) можно вывести упоминавшийся выше факт, что разложение n/z в цепную дробь симметрично.

3. В книге Дэвенпорта [2] приведено несколько способов разложения числа в сумму двух квадратов. Все они так или иначе используют разложения в цепные дроби и неэффективны в смысле теории вычислительной сложности: время работы соответствующих алгоритмов экспоненциально зависит от длины входа (которая примерно равна логарифму числа, для которого находится представление в виде суммы двух квадратов).

4. В случае простого n имеет место замечательная формула Гаусса, доказанная Коши и Якобшталем. Пусть $n = 4k + 1$ — простое,

$$x \equiv \frac{1}{2}\binom{2k}{k} \pmod n, \quad y \equiv (2k)!x \pmod n, \quad |x|, |y| < n/2.$$

Тогда $x^2 + y^2 = n$. Несмотря на «явный» вид формулы Гаусса, она также не приводит к эффективному алгоритму.

К сожалению, в [2] не приводится доказательства формулы Гаусса, а лишь сообщается, что известные доказательства не очень просты. Возможно, кто-нибудь из читателей «Математического просвещения» придумает простое доказательство этого факта?

Автор благодарен Ю. А. Флёрову, заинтриговавшего его вопросом об алгоритмах разложения числа в сумму квадратов, а также И. Богданову, В. Бугаенко, В. Прасолову за интерес к статье и ценные замечания.

Список литературы

[1] Арнольд В. И. *Цепные дроби.* Биб-ка Матем. просв., вып. 14. М.: МЦНМО, 2001.

[2] Дэвенпорт Г. *Высшая арифметика.* М.: Наука, 1965.

[3] Butcher J. C. *MATHEMATICAL MINIATURE 14: Sums of two squares revisited.*
www.math.auckland.ac.nz/~ butcher/miniature/miniature14.pdf

[4] van der Poorten A. J. *The Hermite–Serret Algorithm and* $12^2 + 33^2$ // Cryptography and Computational Number Theory. Springer Verlag, 2001. P. 129–136.

Вялый М. Н., ВЦ РАН, ВКМ НМУ
e-mail: `vyalyi@mccme.ru`

Периодические последовательности

М. Ш. Цаленко

Остатки от деления членов последовательностей, определяемых рекуррентными соотношениями, на данное число k образуют периодические последовательности. Примерами могут служить арифметические и геометрические прогрессии, степенные выражения n^m, в которых n пробегает значения $0, 1, 2, \dots$, а m является фиксированным положительным целым числом, числа Фибоначчи и многие другие последовательности. Нахождение периодов некоторых последовательностей является основной темой этой статьи.

Материал этой статьи может быть использован в математических кружках. Он также полезен и доступен учащимся старших классов, интересующимся математикой и принимающим участие в математических соревнованиях.

1. Рекуррентные последовательности

Мы будем рассматривать последовательности целых чисел или, другими словами, функции неотрицательного целого аргумента, принимающие целые значения. По традиции значение такой функции a на числе n (n-й элемент последовательности) обозначается a_n.

Последовательность называется *рекуррентной*, если она задается формулой

$$a_{n+k} = c_1 a_{n+k-1} + c_2 a_{n+k-2} + \dots + c_k a_0,$$

где $c_1, \dots, c_k$ — фиксированные числа.

Проиллюстрируем это определение на нескольких примерах. Простейший пример — арифметическая прогрессия, которая задается соотношением $a_{n+1} = a_n + d$. Ее можно задать и по-другому, исключив d: $a_{n+2} - 2a_{n+1} + a_n = 0$. Аналогичную формулу можно написать и для последовательности квадратов $a_n = n^2$: $a_{n+3} - 3a_{n+2} - 3a_{n+1} + a_n = 0$.

Задача. Докажите, что для любого k последовательность $a_n = n^k$ удовлетворяет рекуррентному соотношению

$$\sum_{i=0}^{k} (-1)^i \binom{k}{i} a_{n+i} = 0.$$

Рекуррентными являются и многие другие хорошо известные виды последовательностей, скажем, геометрические прогрессии $a_{n+1} = qa_n$ или числа Фибоначчи $a_{n+2} = a_{n+1} + a_n$, $a_0 = 0$, $a_1 = 1$. Некоторые рекуррентные последовательности периодичны, например, последовательность

$$a_{n+2} = a_{n+1} - a_n, \quad a_0 = 2,\ a_1 = 1$$

периодична с периодом 6 (проверьте!).

Упражнение. Проверьте, что остатки, образуемые числами Фибоначчи при делении на 2, 3, 4 соответственно образуют периодические последовательности с периодами 3, 8, 12.

2. Сравнения по модулю

Говорят, что целые числа a и b *сравнимы по модулю* $n > 1$, если при делении на n они дают одинаковый остаток (обозначение $a \equiv b \pmod{n}$). Отношение сравнимости по любому модулю является отношением эквивалентности. Более того, сумма остатков сравнима с остатком от суммы и произведение остатков сравнимо с остатком от произведения. Классы эквивалентности по модулю n называются классами вычетов или просто вычетами по модулю n. Вычеты можно складывать и умножать по следующим правилам

$$[a] + [b] = [a + b], \qquad [a] \cdot [b] = [ab].$$

Множество целых чисел $\mathbb{Z}$ и множества вычетов $\mathbf{Z}_n$ называются в алгебре кольцами, а сопоставление числу a класса вычетов $[a]$, которому принадлежит остаток от деления a на n, является гомоморфизмом кольца целых чисел $\mathbb{Z}$ на кольцо вычетов $\mathbf{Z}_n$.

Вычет $[a]$ называется *делителем нуля* в $\mathbf{Z}_n$, если существует такой вычет $[b]$, что $[a] \cdot [b] = [0]$. Вычет $[a]$ называется *обратимым* в $\mathbf{Z}_n$, если существует такой вычет $[b]$, что $[a] \cdot [b] = [1]$. Можно проверить, что каждый вычет либо является делителем нуля, либо обратим. Обратимыми являются в точности те вычеты $[a]$, для которых a взаимно просто с n. В доказательстве этого факта полезно использовать следующее свойство наибольшего общего делителя (a, b) чисел a и b: (a, b) *является наименьшим положительным числом, которое можно представить в виде целочисленной линейной комбинации* $ua + vb$ *чисел* a *и* b.

3. Периодические последовательности

Назовем последовательность $a_0, \dots, a_n, \dots$ *периодической*, если существует такое целое число $T > 0$, что

$$a_{n+T} = a_n \text{ для всех } n \text{ бóльших некоторого } n_0. \tag{1}$$

Наименьшее возможное число T, для которого выполняется (1), назовем *периодом* последовательности a_n.

Используя свойство наибольшего общего делителя, упомянутое выше, и определение периодической последовательности, нетрудно доказать следующий факт:

Теорема 1. *Если для последовательности a_n существуют два положительных числа T_1 и T_2, для которых выполняется* (1), *то для $D = (T_1, T_2)$ также выполняется* (1).

Поэтому все числа T, для которых выполняется (1), кратны периоду последовательности.

Теорема 2. *Пусть последовательность a_n задается рекуррентным соотношением $a_{n+t} = c_1 a_{n+t-1} + c_2 a_{n+t-2} + \dots + c_t a_0$, в котором все коэффициенты $c_1, \dots, c_t$ — целые числа. Если $a_0, a_1, \dots, a_{t-1}$ — целые числа, то все члены последовательности являются целыми числами, остатки от деления которых на любое целое число образуют периодическую последовательность.*

Прежде чем доказывать теорему 2, разберем случай геометрической прогрессии a^n. Рассмотрим остатки степеней a по модулю k. Последовательность $[a]^n = [a^n]$ обязательно содержит одинаковые члены (так как вычетов по модулю k конечное число). Значит, выполняется равенство

$$[a]^{m_1} = [a]^{m_2} \tag{2}$$

при $m_1 \neq m_2$. Без ограничения общности считаем, что $T = m_1 - m_2 > 0$. Умножая равенство (2) на $[a]^{n-m_2}$, убеждаемся, что $[a]^{n+T} = [a]^n$ при $n \geqslant m_2$. Поэтому последовательности остатков от деления a^n на k периодическая. Подробнее эти последовательности и их периоды обсуждаются ниже в разделе 5.

Пример последовательности $1, 2, 2^2, \dots$ и модуля 4 показывает, что в общем случае периодическое повторение не обязано начинаться с самого начала последовательности: в данном случае последовательность остатков есть $1, 2, 0, 0, \dots$

Доказательство теоремы 2. Оно обобщает приведенное выше рассуждение об остатках членов геометрической прогрессии.

Для остатков членов рекуррентной последовательности также выполняется рекуррентное соотношение

$$[a_{n+t}] = [c_1] \cdot [a_{n+t-1}] + [c_2] \cdot [a_{n+t-2}] + \dots + [c_t] \cdot [a_0],$$

поэтому любые t последовательных элементов рекуррентной последовательности однозначно определяют следующее число в последовательности. Но наборов из t остатков (кортежей длины t) от деления на k

конечное число — k^t. Значит, какой-то набор встретится дважды, после чего элементы последовательности станут периодически повторяться.

4. Степенны́е последовательности

Последовательности $a_n = n^m$ являются рекуррентными, так что последовательности остатков $r_n = n^m \pmod k$ периодичны в силу теоремы 2.

Лемма 1. *a) Период T последовательности $r_n = n^m \pmod k$ является делителем k.*

b) Множество простых делителей периода T совпадает с множеством простых делителей k.

c) Сумма $mT + \binom{m}{2}T^2 + \dots + \binom{m}{m-1}T^{m-1}$ делится на k.

Доказательство. a) Для любого $i = 0, 1, 2, \dots$ имеем:

$$(i+k)^m = i^m + \binom{m}{1}i^{m-1}k + \binom{m}{2}i^{m-2}k^2 + \dots + k^m \equiv i^m \pmod k.$$

По следствию теоремы 1 период T является делителем k. Поэтому из написанного выше сравнения следует, что $(i+T)^m \equiv i^m \pmod k$ для всех $i \geqslant 0$.

b) Если T является периодом, то, в частности, при $i = 0$ имеем:

$$0^m = 0 \equiv (0+T)^m = T^m \pmod k.$$

Следовательно, число T^m делится на k и поэтому среди простых делителей T должны находиться все простые делители k. Из a) видно, что верно и обратное утверждение. Поэтому множества простых делителей k и T совпадают.

c) Если T — период, то при $i = 1$ имеем:

$$1 = 1^m \equiv (1+T)^m = 1 + mT + + \binom{m}{2}T^2 + \dots + \binom{m}{m-1}T^{m-1} + T^m \pmod k.$$

Сокращая 1 в обеих частях этого сравнения и $T^m \equiv 0 \pmod k$ в правой части, получаем нужное сравнение.

Теорема 3. *Остатки от деления членов последовательности полных квадратов на целое число $k > 1$ образуют последовательность, период T которой равен k, если k не делится на 4, и равен $k/2$, если k делится на 4.*

Доказательство. В силу утверждения a) леммы 1 период T является делителем k. Если $T < k$, то $T \leqslant k/2$. Поэтому $2T \leqslant k$. В силу утверждения c) леммы 1 $2T \geqslant k$, т. е. $2T = k$. Следовательно, число k четно. В силу

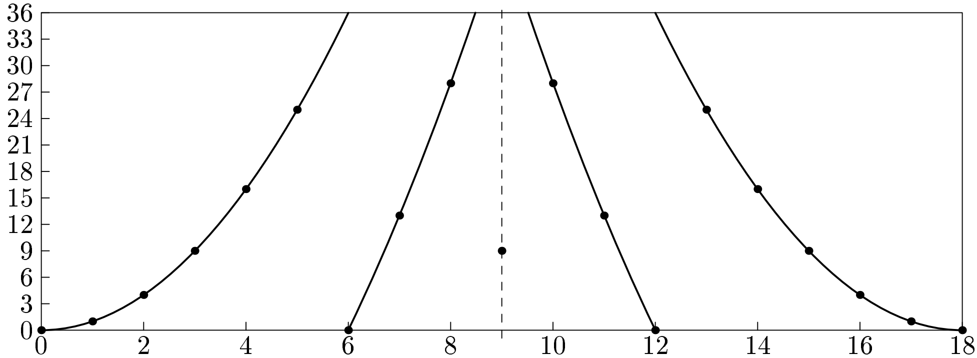


Рис. 1. $x^2 \pmod{18}$

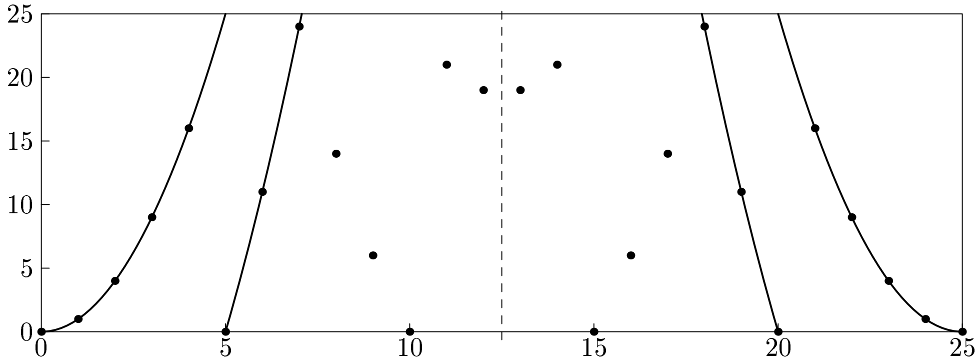


Рис. 2. $x^2 \pmod{25}$

утверждения b) леммы 1 период тоже является четным числом. Поэтому k делится на 4, а период T равен $k/2$. Во всех остальных случаях $T = k$.

Задача. Про многочлен $P(x)$ с целыми коэффициентами известно, что $P(1) = 2$. Докажите, что $P(7)$ не является полным квадратом.

Перейдем теперь к общему случаю последовательности $a_n = n^m$, $m > 2$. Как уже отмечалось, всякая последовательность a_n является функцией аргумента, принимающего неотрицательные целые значения. График такой функции состоит из дискретного множества точек. Как видно из сравнений $n^{2m} \equiv (-n)^{2m} \equiv (T - n)^{2m} \pmod{k}$, в случае степенны́х последовательностей с четным показателем часть графика, соответствующая периоду (от 0 до T) симметрична относительно прямой $x = T/2$. На рис. 1 и 2 можно увидеть эту симметрию (куски парабол добавлены «для красоты»).

Теперь построим *наименьшее* число T, множество простых делителей которого совпадает с множеством простых делителей числа k и mT делится на k.

Пусть $k = p_1^{\alpha_1} p_2^{\alpha_2} \cdot \ldots \cdot p_s^{\alpha_s}$, где $p_1, p_2, \ldots, p_s$ — различные простые числа, а показатели степеней $\alpha_1, \ldots, \alpha_s$ больше нуля. Представим m в виде $m = p_1^{\beta_1} p_2^{\beta_2} \cdot \ldots \cdot p_s^{\beta_s} m_1$, допуская для показателей степеней $\beta_1, \ldots, \beta_s$ значение 0. Положим

$T(k,m) = p_1^{\gamma_1} p_2^{\gamma_2} \cdot \ldots \cdot p_s^{\gamma_s}$, где $\gamma_i = \alpha_i - \beta_i$, если $\alpha_i > \beta_i$, и $\gamma_i = 1$, если $\alpha_i \leqslant \beta_i$.

Ясно, что число $T(k,m)$ является наименьшим делителем k, обладающим указанными выше свойствами. Обозначим $d(k,m) = k/T(k,m)$. Легко видеть, что m делится на $d(k,m)$ и

$d(k,m) = p_1^{\delta_1} p_2^{\delta_2} \cdot \ldots \cdot p_s^{\delta_s}$, где $\delta_i = \beta_i$, если $\alpha_i > \beta_i$, и $\delta_i = \alpha_i - 1$, если $\alpha_i \leqslant \beta_i$,

Теорема 4. *Период последовательности* $r_n = n^m \pmod{k}$ *является делителем числа* $T(k,m)$.

Доказательство. Для упрощения записи вместо $T(k,m)$ и $d(k,m)$ будем писать T и d. В силу следствия теоремы 1 для доказательства теоремы достаточно показать, что для любого целого неотрицательного i выполняется $(i+T)^m \equiv i^m \pmod{k}$. Так как

$$(i+T)^m = i^m + \binom{m}{1} T i^{m-1} + \cdots + \binom{m}{l} T^l i^{m-l} + \cdots + T^m,$$

то достаточно доказать, что все числа $\binom{m}{l} T^l$ делятся на $k = dT$ при $1 \leqslant l \leqslant m$.

Вначале рассмотрим случай $l < m$. Обозначим $q = m/d$. Так как

$$\binom{m}{l} T^l = \frac{m(m-1) \cdot \ldots \cdot (m-l+1)}{l \cdot (l-1)!} T^l = (dT) \binom{m-1}{l-1} \cdot \frac{q}{l} T^{l-1},$$

достаточно доказать, что $A = \binom{m-1}{l-1} \cdot \frac{q}{l} T^{l-1}$ — целое. Как показывает разложение

$$\binom{m}{l} = d \cdot \binom{m-1}{l-1} \cdot \frac{q}{l},$$

в знаменатель представления A в виде несократимой дроби могут входить только делители k (если p не является делителем k, то p также не является делителем d). Но если k делится на p, то T^{l-1} делится по крайней мере на p^{l-1}. С другой стороны, p может входить в разложение l на простые множители не более, чем $\lfloor \log_p l \rfloor$ раз (здесь $\lfloor\ \rfloor$ обозначает целую часть числа). Из неравенства $\log_p l \leqslant \log_2 l \leqslant l-1$ следует, что число A — целое.

Остается показать, что T^m делится на k. Если p — простой делитель k, на который m не делится, то p входит в T в той же степени, что и в k. Если же p — общий простой делитель k и m, то p входит в T по крайней мере один раз.

Если $\alpha \leqslant \beta$ (для упрощения записи индексы опущены), то p входит в T в первой степени. Тогда p входит в T^m по крайней мере p^β раз. Так как $p \geqslant 2$ и $\beta \geqslant \alpha$, то $p^\beta \geqslant 2^\beta > \beta \geqslant \alpha$, и T^m делится на p^α.

Если $\alpha > \beta$, то p входит в T по крайней мере $(\alpha-\beta)p^\beta$ раз. Покажем, что $(\alpha-\beta)p^\beta \geqslant \alpha$. Так как $\beta \leqslant p^\beta - 1$, то $1/\beta \geqslant 1/(p^\beta-1)$. Учитывая, что $\alpha - \beta \geqslant 1$, получаем

$$1 + \frac{\alpha-\beta}{\beta} \geqslant 1 + \frac{1}{p^\beta - 1} \quad \Leftrightarrow \quad \frac{\alpha}{\beta} \geqslant \frac{p^\beta}{p^\beta - 1} \quad \Leftrightarrow \quad (\alpha-\beta)p^\beta \geqslant \alpha.$$

Тем самым доказано, что T^m делится на k.

Следствие 1. *Если $k = p_1 p_2 \cdot \ldots \cdot p_s$, где $p_1, p_2, \ldots, p_s$ — различные простые множители, то при любом m период последовательности $a_n = n^m \pmod{k}$ равен k.*

Следствие 2. *Если k и m взаимно просты, то период последовательности $a_n = n^m \pmod{k}$ равен k.*

Доказательство. Предположим, что период T меньше k. Множество простых делителей у T такое же, как у k (лемма 1b). Поэтому без ограничения общности можно считать, что $k = p^2 q$ и T является делителем $T_1 = pq$. Тогда $i^m \equiv (i + T_1)^m \pmod{k}$, а так как T_1^2 делится на k, то

$$1^m \equiv (1 + T_1)^m = 1 + mT_1 + \binom{m}{2} T_1^2 + \cdots + T_1^m \equiv 1 + mT_1 \pmod{k}.$$

Значит, mT_1 делится на k. Поскольку k и m взаимно просты, то и T_1 делится на k. Приходим к противоречию, так как $T_1 < k$.

В том случае, когда m и k имеют общие множители, период может оказаться меньше $T(k,m)$, как показывает следующий пример. Если $k = 32 = 2^5$ и $m = 8 = 2^3$, то $T(32,8) = 4$. Однако период $T = 2$, как показывают вычисления:

n	0	1	2	3	4
r_n	0	1	0	1	0

Единственное неочевидное равенство в этой таблице: $3^8 \equiv 1 \pmod{32}$. Оно легко усматривается из разложения

$$3^8 - 1 = (3-1)(3+1)(3^2+1)(3^4+1).$$

Все множители в правой части четны, а $3+1$ делится на 4.

Следствие 3. *Если* $k = p_1^{\alpha_1} p_2^{\alpha_2} \cdot \ldots \cdot p_s^{\alpha_s}$, $m = p_1^{\beta_1} p_2^{\beta_2} \cdot \dots \cdot p_s^{\beta_s}$, $\beta_i \geqslant \alpha_i - 1$ *для всех* $i = 1, 2, \ldots, s$, *то период последовательности* $a_n = n^m \pmod k$ *равен* $p_1 p_2 \cdot \dots \cdot p_s$.

5. Геометрические прогрессии

Из рассмотрения степенны́х функций может показаться, что период всегда связан с модулем k. Однако уже для случая показательных функций (или геометрических прогрессий) $a_n = m^n$, $m > 1$, период не имеет очевидной связи с модулем.

Поскольку следующий член геометрической прогрессии однозначно определяется предыдущим, периодом a_n будет такое наименьшее $T > 0$, что

$$m^i (m^T - 1) \equiv 0 \pmod k \tag{3}$$

при всех достаточно больших i.

Вначале рассмотрим случай, когда m взаимно просто с k. В этом случае T совпадает с *порядком* $[m]$ в кольце $\mathbf{Z}_k$, т. е. наименьшим положительным числом r, для которого $[m]^r = [1]$. Напомним, что порядок является делителем количества чисел, меньших k и взаимно простых с k. Это последнее число обозначается $\varphi(k)$ и называется *функцией Эйлера*. Функция Эйлера обладает свойством *мультипликативности*: для взаимно простых x, y выполняется $\varphi(xy) = \varphi(x)\varphi(y)$. Из свойства мультипликативности легко вывести *формулу Эйлера*

$$\varphi(p_1^{\alpha_1} p_2^{\alpha_2} \cdot \ldots \cdot p_s^{\alpha_s}) = (p_1^{\alpha_1} - p_1^{\alpha_1 - 1})(p_2^{\alpha_2} - p_2^{\alpha_2 - 1}) \cdot \ldots \cdot (p_s^{\alpha_s} - p_s^{\alpha_s - 1}),$$

где p_i — различные простые числа.

Поскольку T делит $\varphi(k)$, то выполняется сравнение $m^{\varphi(k)} \equiv 1 \pmod k$, которое называется *теоремой Эйлера*. Если k — простое, то $\varphi(k) = k - 1$ и теорема Эйлера превращается в *малую теорему Ферма*: $m^{k-1} \equiv 1 \pmod k$.

Все приведенные выше факты хорошо известны, и их можно найти во многих книгах по теории чисел, скажем, в книге Н. Б. Алтуфовой, А. В. Устинова «Алгебра и теория чисел. Сборник задач для математических школ.», М.: МЦНМО, 2002.

Общий случай легко сводится к уже рассмотренному. Для любого целого $k > 1$ положим $k = p_1^{\alpha_1} p_2^{\alpha_2} \cdot \ldots \cdot p_s^{\alpha_s} k_1$, где $p_1, p_2, \ldots, p_s$ — все общие простые делители m и k, а $\alpha_1, \alpha_2, \ldots, \alpha_s$ — наибольшие степени, в которых эти делители входят в разложение k на простые множители. В дальнейшем произведение $p_1^{\alpha_1} p_2^{\alpha_2} \cdot \ldots \cdot p_s^{\alpha_s}$ будет обозначаться через $b(m, k)$ или просто через b. Число k_1 равно $k / b(m, k)$.

Теорема 5. *Период последовательности остатков* $r_n = m^n \pmod{k}$, *где* m *и* k — *фиксированные целые числа, большие* 1, *равен порядку* $[m]$ *в* $\mathbf{Z}_{k_1}$.

Доказательство. Поскольку $b(m,k)$ является произведением степеней простых чисел, входящих в m, все достаточно большие степени m делятся на $b(m,k)$. Но тогда из сравнения (3) заключаем, что $m^T - 1 \equiv 0 \pmod{k_1}$.

Проиллюстрируем утверждение теоремы 5 на следующем примере. Пусть $k = 12$. Если $m = 2$, то $k_1 = 4$ и $b = 4$. Тогда период последовательности остатков, получающихся при делении чисел 2^n на 12, равен 2, так как вычет $[2]$ имеет порядок 2 в $\mathbf{Z}_3$. Если $m = 3$, то $k_1 = 4$ и $b = 3$, и тогда период последовательности $3^n \pmod{12}$ также равен 2, так как вычет $[3]$ имеет порядок 2 в $\mathbf{Z}_4$.

Отметим, что теорема 5 объясняет, в частности, тот факт, что критерии делимости на разные числа k имеют разную сложность: сложность критерия определяется величиной порядка вычета $[10]$ в $\mathbf{Z}_{k_1}$. Например, $[10]$ имеет порядок 2 в $\mathbf{Z}_{11}$ и порядок 6 в $\mathbf{Z}_7$.

6. Числа Фибоначчи

Напомним, что последовательность чисел Фибоначчи определяется рекуррентным соотношением

$$a_{n+2} = a_{n+1} + a_n$$

и начальными условиями $a_0 = 0$, $a_1 = 1$. В силу теоремы 2 для любого k остатки образуют периодическую последовательность. Заметим, что остатки чисел Фибоначчи удовлетворяют тому же рекуррентному соотношению по любому модулю. Это позволяет упростить вычисление членов последовательности остатков. Кроме того, любые два соседних члена последовательности чисел Фибоначчи полностью определяют всю последовательность: если известны члены a_m, a_{m+1} то, используя рекуррентное соотношение, можно найти не только a_{m+2}, a_{m+3}, …, но и a_{m-1}, a_{m-2}, …, поскольку из $a_{m+1} = a_m + a_{m-1}$ следует $a_{m-1} = a_{m+1} - a_m$.

Из последнего замечания сразу следуют два важных вывода:

а) *остатки образуют периодическую последовательность, первый период которой начинается с членов* a_0 *и* a_1;

б) *существует бесконечно много чисел Фибоначчи, которые делятся на любое заданное* k.

Последнее утверждение отмечено также в книге Н. Н. Воробьёва «Числа Фибоначчи».

Заметим, что *любые два соседних члена последовательности чисел Фибоначчи взаимно просты*. Это утверждение устанавливается с помощью метода математической индукции.

Обозначим через m номер второго числа Фибоначчи, которое делится на k (первым является число $a_0 = 0$). Число a_{m-1} взаимно просто с a_m в силу сказанного выше и тем самым взаимно просто с k, так как a_m согласно нашему выбору делится на k. Значит, вычет $[a_{m-1}]$ обратим в $\mathbf{Z}_k$, обозначим через p его порядок.

Теорема 6. *В обозначениях, введенных выше, период последовательности остатков, получающихся при делении чисел Фибоначчи на $k > 1$, равен pm.*

Доказательство. Первый период начинается числами 0, 1 и должен кончаться числом 1, так как $1 + 0 = 1$. Обозначим через r остаток от деления a_{m-1} на k. Тогда члены с номерами $m-1$, m, $m+1$, $m+2$ в последовательности остатков выглядят следующим образом:

$$r,\ 0,\ r,\ r,\ 2r \pmod{k}.$$

Значит, остаток с номером $m + i$, $0 \leqslant i < m$ получается из остатка с номером i умножением на r и последующим взятием остатка по модулю k. Поэтому перед третьим нулем в последовательности остатков появится остаток от деления r^2 на k. Продолжая дальше, можно заметить, что после третьего нуля начальный отрезок последовательности умножается на r^2 (с последующим взятием остатка по модулю k) и т. д. Следовательно, 1 появится первый раз перед нулем в последовательности остатков, когда перед нулем окажется $[r^p]$. Так как второму нулю предшествуют m чисел и эти числа умножаются последовательно на r, r^2, $\ldots r^{p-1}$, то общее число чисел в периоде равно mp, что и утверждалось.

Покажем на примерах, как вычисляется период последовательности остатков чисел Фибоначчи, получающихся при делении на число k.

Пусть $k = 3$. Второй член в последовательности чисел Фибоначчи, который делится на 3, имеет номер 4. Предыдущий член равен 2. Так как $2^2 \equiv 1 \pmod{3}$, то порядок этого члена равен 2. Следовательно, по теореме 6 период равен $4 \cdot 2 = 8$.

Для вычисления периода необязательно знать второе число a_m, которое делится на k. Достаточно знать только остаток r_{m-1} от деления a_{m-1} на k, поскольку остатки связаны тем же рекуррентным соотношением, что и сами числа Фибоначчи. Продемонстрируем сказанное для числа $k = 10$. Вычисляем последовательность остатков:

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
r_i	0	1	1	2	3	5	8	3	1	4	5	9	4	3	7	0

Следовательно, $m = 15$ и осталось найти порядок $[7]$ в кольце $\mathbf{Z}_{10}$. Так как 7^2 заканчивается на 9, 7^3 — на 3, 7^4 — на 1, порядок равен 4. Поэтому период равен $15 \cdot 4 = 60$.

Приведем таблицу значений периода для $k = 2, \ldots, 15$. В этой таблице, как и выше, — m означает номер второго члена в последовательности Фибоначчи, который делится на заданное k, a_m — соответствующее число Фибоначчи, p — порядок $[a_{m-1}]$ в $\mathbf{Z}_k$, $T(k)$ — период.

k	m	a_m	p	$T(k) = mp$
2	3	2	1	$3 \cdot 1 = 3$
3	4	3	2	$4 \cdot 2 = 3$
4	6	8	2	$6 \cdot 2 = 12$
5	5	5	4	$5 \cdot 4 = 20$
6	12	144	2	$12 \cdot 2 = 24$
7	8	21	2	$8 \cdot 2 = 16$
8	6	8	2	$6 \cdot 2 = 12$
9	12	144	2	$12 \cdot 2 = 24$
10	15	610	4	$15 \cdot 4 = 60$
11	10	55	1	$10 \cdot 1 = 10$
12	12	144	2	$12 \cdot 2 = 24$
13	7	13	4	$7 \cdot 4 = 28$
14	24	46368	2	$24 \cdot 2 = 48$
15	20	6765	2	$20 \cdot 2 = 40$

Автор глубоко благодарен Г. А. Гальперину, тщательно прочитавшему первоначальный вариант рукописи и сделавшему многочисленные замечания, которые были учтены при доработке текста статьи.

Michael Tsalenko
E-mail: `mtsalenko@sbcglobal.net`

Конкурсы и олимпиады

Студенческий конкурс решения задач 2004–2005 гг.

О правилах конкурса

Вниманию студентов университета предлагается домашний конкурс. Подобные конкурсы проводились неоднократно в прошлом. (Например, о конкурсе 2001–2002 гг. сообщалось в «Математическом просвещении», вып. 7, 2003, с. 177–181.) Среди их участников есть как недавние, так и очень давние выпускники мат-меха. Идея конкурса (в отличие от олимпиад) в решении задач, в основном, исследовательского характера, которые не решаются одним приемом, а требуют серьезных и неспешных раздумий. Конкурс проводится под эгидой Санкт-Петербургского математического общества. Общество образует жюри и выделяет денежные призы для награждения победителей.

В конкурсе могут принимать участие студенты всех курсов. Итоги среди первокурсников подводились отдельно. Допускались коллективные (не более трех человек) работы.

Ниже приводятся условия задач конкурса 2004–2005 гг. В предлагаемом списке имеются задачи различной трудности, от сравнительно несложных до нерешенных проблем. Участники могут выбрать задачи по вкусу и подавать решения *любого набора* задач, *даже одной*. Принимаются решения частей задач, состоящих из нескольких пунктов.

Условия задач

Задача 1. Дано множество $A \subset \mathbb{R}$ положительной меры Лебега. Функция $f\colon \mathbb{R} \to \mathbb{R}$ такова, что при всех $x \in \mathbb{R}$ и при всех $a \in A$

$$f(x-a) = f(x) + f(a) - 1.$$

Докажите, что $f(x) \equiv 1$.

Задача 2. Функция $f\colon (-1,1) \to \mathbb{R}$ бесконечно дифференцируема и для любого $k = 0, 1, 2, \ldots$ точка $x = 0$ — единственный корень производной $f^{(k)}(x)$. Докажите, что

$$\forall \varepsilon > 0 \quad |f'|^{1+\varepsilon} = o(f) \quad \text{при} \quad x \to 0\,.$$

Задача 3. Даны числа $0 < a < b$. Убывающая непрерывная функция $y(x)\colon [0,1] \to [0,1]$ определена условием $x^a - x^b = y^a - y^b$. Докажите, что

$$\int_0^1 \frac{\ln y}{x}\,dx = -\frac{\pi^2}{3ab}\,.$$

Задача 4. В некоторых узлах целочисленной решетки на плоскости расположены точечные источники света, а в некоторых — точечные заслонки. Источники света образуют квадратную подрешетку со стороной M, а заслонки — квадратную подрешетку со стороной N (стороны всех решеток параллельны координатным осям). Источник и заслонка не могут находиться в одной точке.

В одной из незанятых точек плоскости находится наблюдатель. Наблюдатель видит источник света, если на соединяющем их отрезке нет ни одной заслонки. При каких отношениях M/N расположение решеток и наблюдателя может быть таким, что наблюдатель

а) не видит ни одного источника света?

б) не видит ни одного источника света внутри некоторого ненулевого угла с вершиной в точке, где он находится?

Задача 5. Обозначим через T единичную окружность $T = \{z \in \mathbb{C} : |z| = 1\}$. Докажите, что существует такая измеримая функция $f\colon T \to T$, что для любой интегрируемой функции $h\colon T \to T$ (можно ограничиться непрерывными h) найдется такая возрастающая последовательность натуральных чисел n_k, что

$$\int_T |f^{n_k}(t) - h(t)|\,dt \to 0 \qquad \text{при} \qquad k \to +\infty.$$

(Здесь f^{n_k} — обычное возведение в степень.)

Задача 6. а) Дан сферический треугольник со сторонами a, b, c и противолежащими углами α_s, β_s, γ_s. Углы треугольника со сторонами a, b, c на плоскости обозначим α, β, γ. Докажите, что

$$\alpha_s - \alpha \leqslant (\beta_s - \beta) + (\gamma_s - \gamma).$$

б) Верно ли аналогичное неравенство для треугольников на двух сферах разных радиусов? Для плоскости и плоскости Лобачевского?

Задача 7. а) Дан вещественный многочлен $P(x,y)$, имеющий в точке $(0,0)$ изолированный нуль (т.е. $P(0,0)=0$ и $P(x,y)\neq 0$ в некоторой проколотой окрестности точки $(0,0)$). Докажите, что существуют такие натуральное n и вещественное $C>0$, что в некоторой окрестности нуля верно неравенство

$$|P(x,y)| \geqslant C(x^2+y^2)^n. \qquad (*)$$

б) Найдите оценки наименьшего показателя $n=n(k)$, $k\to+\infty$, такого что неравенство $(*)$ имеет место для всех многочленов P степени k.

Задача 8. Рассмотрим n-значные числа в k-ичной системе счисления (первой цифрой может быть и 0). Назовем число хорошим, если любые две его соседние цифры отличаются не более чем на 1. Для каждого n-значного числа A определим $f(A)$ как наименьшее число цифр, которое необходимо в нем зачеркнуть, чтобы получившееся число стало хорошим.

а) Докажите, что для некоторого $c>0$ наблюдается следующее явление концентрации:

$$\frac{1}{k^n}\sum_{A}|f(A)/n-c|\to 0, \qquad \text{при} \quad n\to+\infty,$$

где сумма берется по всем n-значным числам.

б) Докажите, что для $k=3$ константа c равна $1/3-\sqrt{5}/15$.

в) Получите как можно лучшие оценки для c при $k>3$.

Задача 9. Пусть функция f непрерывна в замкнутом единичном круге D и тождественно равна нулю на его границе. Известно, что в точках a и b, принадлежащих $\operatorname{Int} D$, она имеет локальные максимумы.

а) Докажите, что если $f\in C^2(D)$, то в $\operatorname{Int} D$ существует еще одна (отличная от a и b) стационарная точка f.

б) Верно ли заключение пункта а), если $f\in C^1(\operatorname{Int} D)$?

Задача 10. Замкнутые несамопересекающиеся гладкие кривые в $\mathbb{R}^2$ с кривизной меньше 1 будем называть *плавными*.

а) Докажите, что в круг радиуса 2.1 нельзя вписать плавную кривую длины больше 100. Найдите более разумную оценку сверху длины плавной кривой, которую можно поместить в этот круг.

б) Докажите, что в круг радиуса 2.2 можно вписать плавную кривую сколь угодно большой длины.

в) Докажите, что существует такое число $\ell>2\pi$, что в круг радиуса 2.2 нельзя вписать плавную кривую длины ℓ.

г) Верно ли, что любые две плавные кривые, лежащие в круге большого радиуса, гомотопны в классе плавных кривых?

Задача 11. Обозначим через c_n последовательность чисел Каталана: $c_n=\frac{1}{n+1}\binom{2n}{n}$. Докажите, что

$$\det\begin{pmatrix} c_0 & c_1 & c_2 & \dots & c_n\\ c_1 & c_2 & c_3 & \dots & c_{n+1}\\ c_2 & c_3 & \dots & \dots & c_{n+2}\\ \dots & \dots & \dots & \dots & \dots\\ c_n & c_{n+1} & c_{n+2} & \dots & c_{2n}\end{pmatrix}=1.$$

Задача 12. а) Докажите, что для любых $m, n \in \mathbb{N}$ существует такая непрерывная функция $f\colon S^{n-1} \to \mathbb{R}^m$ и точки $x_1, x_2, \ldots, x_{n-m+2}$ на сфере S^{n-1}, что не существует поворота $U \in SO(n)$, для которого

$$f(Ux_1) = f(Ux_2) = \cdots = f(Ux_{n-m+2}).$$

б) Докажите, что для любой непрерывной функции $f\colon S^2 \to \mathbb{R}^2$ и для любых точек x_1 и x_2 на сфере S^2, существует поворот $U \in SO(2)$, для которого $f(Ux_1) = f(Ux_2)$.

в) Докажите, что для любой непрерывной функции $f\colon S^2 \to \mathbb{R}$ найдутся три точки x_1, x_2 и x_3 на сфере S^2, являющиеся концами ортов, для которых $f(x_1) = f(x_2) = f(x_3)$.

г) Докажите, что для любой непрерывной функции $f\colon S^2 \to \mathbb{R}$ и для любых трех точек x_1, x_2 и x_3 на сфере S^2, существует поворот $U \in SO(3)$, для которого $f(Ux_1) = f(Ux_2) = f(Ux_3)$.

д) Докажите еще для каких-нибудь m, n и конфигураций из $k = n - m + 1$ точек $x_1, \ldots, x_k$ на сфере S^{n-1} и любой непрерывной функции $f\colon S^{n-1} \to \mathbb{R}^m$, что существует поворот $U \in SO(n)$, для которого

$$f(Ux_1) = f(Ux_2) = \cdots = f(Ux_k).$$

Задача 13. а) Назовем выпуклый центрально симметричный многогранник в $\mathbb{R}^n$ *славным*, если на сфере S^{n-1} существует борелевская мера μ, такая что при всех $x \in \mathbb{R}^n$

$$\|x\| = \int |\langle x, y\rangle|\, d\mu(y),$$

где через $\|\cdot\|$ обозначена норма, порождаемая этим многогранником. Докажите, что многогранник славен тогда и только тогда, когда выполняется принцип усреднения: для любых точек $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ и любых вещественных чисел $\lambda_1, \lambda_2, \ldots, \lambda_k$ из условия, что неравенство $\sum \lambda_i |\langle x_i, y\rangle| \geqslant 0$ справедливо для любого $y \in \mathbb{R}^n$, следует неравенство $\sum \lambda_i \|x_i\| \geqslant 0$.

б) Останется ли в конечномерном случае утверждение верным, если вместо многогранника взять любое центрально симметричное выпуклое тело?

Задача 14. Пусть F — распределение на $[0,1]$, σ^2 — его дисперсия, а μ_4 — четвертый центральный момент. Докажите, что

$$\mu_4 + 3\sigma^4 \leqslant \sigma^2,$$

причем равенство возможно только для вырожденного распределения или распределения Бернулли.

Обобщение перестановочного неравенства и монгольское неравенство

Л. В. Радзивиловский

После прочтения этой статьи вам станут очевидными доказательства многих неравенств. Все они будут вытекать в одну-две строчки из одного простого принципа, который объясняется в этой статье. Вот примеры того, что станет очевидным:

1. $x^5+y^5+z^5 \geqslant x^3y^2+y^3x^2+z^3x^2$ (здесь и далее числа положительны).
2. $\dfrac{x}{1+x^2}+\dfrac{y}{1+y^2}+\dfrac{z}{1+z^2} \leqslant \dfrac{x}{1+y^2}+\dfrac{y}{1+z^2}+\dfrac{z}{1+x^2}$.
3. $(a_1+a_2+\dots+a_n)\big(\dfrac{1}{a_1}+\dfrac{1}{a_2}+\dots+\dfrac{1}{a_n}\big) \geqslant n^2.$
4. $(a_1^3+a_2^3+\dots+a_n^3)(a_1+a_2+\dots+a_n) \leqslant n(a_1^4+a_2^4+\dots+a_n^4).$
5. $\left(1+\dfrac{a_1^2}{a_2}\right)\cdot\left(1+\dfrac{a_2^2}{a_3}\right)\cdot\ldots\cdot\left(1+\dfrac{a_n^2}{a_1}\right) \geqslant (1+a_1)\cdot(1+a_2)\cdot\ldots\cdot(1+a_n).$ (16-й Турнир Городов, осень, основной вариант, старшие классы, задача №4, Л. Д. Курляндчик).
6. $\sqrt{x+2^x}+\sqrt{y+2^y}+\sqrt{z+2^z} \leqslant \sqrt{y+2^x}+\sqrt{z+2^y}+\sqrt{x+2^z}.$
7. Дана таблица $n\times n$, заполненная числами по следующему правилу: в клетке, стоящей в i-й строке и j-м столбце таблицы, записано число $(i+j-1)-1$. В таблице зачеркнули n чисел таким образом, что никакие два зачеркнутых числа не находятся в одном столбце или в одной строке. Докажите, что сумма зачеркнутых чисел не меньше 1. (13й Турнир Городов, весна, основной вариант, старшие классы, задача №3, С. Иванов)

Кроме того, мы докажем монгольское неравенство. Это уже займет не одну строчку, но всё равно будет довольно прозрачно. Об истории монгольского неравенства и о двух его сложных доказательствах написано в статье А. И. Храброва [4].

Я намеренно не формулирую факты кратчайшим способом, а предпочитаю «повозиться» и рассказать всё в том порядке, в котором я это осознавал. Мы начнем с частных случаев и перейдем к общим теоремам, а не наоборот. Конечно, жаль бумагу, байты и леса, но именно «возня» способствует развитию интуиции. Чтоб не отнимать у читателей удовольствия «повозиться», я опустил много деталей. Поэтому, чтобы понять этот текст, вам понадобится ручка и бумажка.

Перестановочное неравенство

Перестановкой (из n элементов) мы называем биективную функцию из множества чисел $\{1, 2, 3, \ldots, n\}$ в себя. Слово «биективная» означает, что во всякое число под действием этой функции переходит ровно одно число. Обозначать перестановку я буду греческой буквой σ (сигма). Числа $\sigma(1)$, $\sigma(2)$, …, $\sigma(n)$ — это те же числа 1, 2, …, n, но переставленные в другом порядке.

Рассмотрим два набора чисел: $a_1, a_2, \ldots, a_n$ и $b_1, b_2, \ldots, b_n$. Будем говорить, что наборы $a_1, a_2, \ldots, a_n$ и $b_1, b_2, \ldots, b_n$ упорядочены «одинаково», если наибольшему числу в первом наборе соответствует наибольшее во втором (т. е. индексы у них одинаковые), второму по величине числу в первом наборе соответствует второе по величине, и т. д., а наименьшему соответствует наименьшее. Например, если $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_n$ и $b_1 \geqslant b_2 \geqslant \cdots \geqslant b_n$, то наборы одинаково упорядочены. Другой пример: наборы a, b, c и a^3, b^3, c^3 одинаково упорядочены. Будем говорить, что наборы обратно упорядочены, если наибольшему числу в первом наборе соответствует наименьшее во втором, второму по величине в первом наборе соответствует второе с конца во втором и т. д., а наименьшему соответствует наибольшее. Например, если $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_n$ и $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n$, то наборы обратно упорядочены. Наборы a, b, c и $1/a, 1/b, 1/c$ также обратно упорядочены.

Для любых двух наборов $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_n$ и $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n$ существует такая «самая правильная» перестановка σ, что наборы $a_1, a_2, \ldots, a_n$ и $b_{\sigma(1)}, b_{\sigma(2)}, \ldots b_{\sigma(n)}$ одинаково упорядочены. Аналогично, для любых двух наборов можно найти и «самую неправильную» перестановку σ, при которой наборы $a_1, a_2, \ldots, a_n$ и $b_{\sigma(1)}, b_{\sigma(2)}, \ldots b_{\sigma(n)}$ обратно упорядочены.

Какое отношение имеет всё это к доказательству неравенств? А вот какое:

Теорема 1 (перестановочное неравенство). Рассмотрим все возможные перестановки из n элементов. Тогда значение выражения

$$a_1 b_{\sigma(1)} + a_2 b_{\sigma(2)} + \cdots + a_n b_{\sigma(n)}$$

будет самым большим, когда числа упорядочены одинаково, и самым маленьким, когда числа обратно упорядочены.

Иначе говоря, значения, полученные при остальных перестановках, не больше значения, полученного при «самой правильной», и не меньше значения, полученного при «самой неправильной» перестановке.

Например, из теоремы следует, что $a/b + b/c + c/a \geqslant 3$: ведь a, b, c и $1/a, 1/b, 1/c$ обратно упорядочены, а значит, правая часть $a/a + b/b + c/c$ получена при самой неправильной перестановке и потому дает наименьшее значение.

Другие примеры дают неравенства 1 и 2 из списка в начале статьи. Конечно, подобные неравенства можно доказать и другими методами. Но перестановочное неравенство позволяет единообразно доказывать много неравенств.

Это неравенство очень впечатлило меня в первый же раз, когда я с ним столкнулся, своей мощью и общностью. Узнал я его в 11-м классе школы Шевах-Мофет (Тель-Авив) на очень интересном уроке по неравенствам. Проводил этот урок Михаил Розенберг (учитель этой же школы).

Перестановочное неравенство справедливо для любых действительных чисел (нет ограничения положительности).

Интуитивно легко почувствовать правильность перестановочного неравенства: если вы хотите получить число побольше, то приставьте к бо́льшим числам бо́льшие коэффициенты при суммировании. Это выгодней, чем тратить большие коэффициенты на маленькие числа.

Строго доказать перестановочное неравенство также несложно.

Лемма. Пусть $a_1 > a_2$, $b_1 > b_2$. Тогда $a_1 b_1 + a_2 b_2 > a_1 b_2 + a_2 b_1$.

Доказательство. Перенесем всё в левую часть:

$$a_1 b_1 + a_2 b_2 - a_1 b_2 - a_2 b_1 > 0,$$
$$(a_1 - a_2)(b_1 - b_2) > 0.$$

Но произведение двух положительных чисел положительно, и полученное неравенство равносильно исходному.

Доказательство перестановочного неравенства. Возьмем два набора чисел. Если есть такая пара индексов i, k, что порядок неправильный (например, $a_i > a_k$, $b_k > b_i$), то мы поменяем их местами. При этом сумма возрастет (согласно лемме). Будем повторять такие операции, пока это возможно (в «компьютерной науке» этот процесс называется «сортировка»). Если это уже невозможно (сортировка закончилась), то легко понять, что порядок стал правильным. Но мы начинали с произвольного порядка и несколько раз увеличивали наше число! Значит, правильный порядок соответствует самому большому значению.

Аналогично доказывается обратное направление перестановочного неравенства. Просто надо запустить сортировку наоборот: каждый раз, если есть хоть одна правильная пара индексов, нужно сделать из нее неправильную. Согласно лемме число будет постоянно убывать. В конце мы получим самый неправильный порядок и значение будет меньше. А ведь мы начинали с произвольного порядка.

Упражнение. Проследите доказательство внимательно и поймите, когда в перестановочном неравенстве достигается равенство.

Упражнение. В неравенствах 1 и 2 в начале статьи мы потребовали, чтобы числа были положительными. Действительно ли это нужно? Где «сломается» доказательство, если мы «запустим» в него произвольные действительные числа?

Теорема 2 (неравенство Чебышёва). Пусть наборы $a_1, a_2, \ldots, a_n$ и $b_1, b_2, \ldots, b_n$ упорядочены одинаково. Тогда

$$n(a_1b_1 + a_2b_2 + \cdots + a_nb_n) \geqslant (a_1 + \cdots + a_n)(b_1 + \cdots + b_n).$$

Если же они обратно упорядочены, то выполняется неравенство с противоположным знаком.

Вообще-то неравенством Чебышёва правильнее называть более общее неравенство (см. [3]), но это название уже прижилось (см., скажем, [2]).

Доказательство. В первом случае

$$\begin{aligned} a_1b_1 + a_2b_2 + \cdots + a_nb_n &\geqslant a_1b_1 + a_2b_2 + \cdots + a_nb_n, \\ a_1b_1 + a_2b_2 + \cdots + a_nb_n &\geqslant a_1b_2 + a_2b_3 + \cdots + a_nb_1, \\ &\ldots \\ a_1b_1 + a_2b_2 + \cdots + a_nb_n &\geqslant a_1b_n + a_2b_1 + \cdots + a_nb_{n-1}. \end{aligned}$$

Суммируя всё это, получаем искомый результат. Если же наборы обратно упорядочены, то знаки неравенства перевернутся во всех слагаемых и в сумме.

В качестве немедленных следствий получаем неравенства 3 и 4 из списка в начале статьи и многие другие.

Поменяем сложение на умножение

Рассмотрим неравенство 5 из списка в начале статьи:

$$\left(1 + \frac{a_1^2}{a_2}\right) \cdot \left(1 + \frac{a_2^2}{a_3}\right) \cdot \ldots \cdot \left(1 + \frac{a_n^2}{a_1}\right) \geqslant (1 + a_1) \cdot (1 + a_2) \cdot \ldots \cdot (1 + a_n).$$

Как уже говорилось, эта задача предлагалась на 16-м Турнире Городов, в котором я участвовал. Мне показалось естественным умножить обе части на произведение всех чисел. Получим:

$$(a_2 + a_1^2) \cdot (a_3 + a_2^2) \cdot \ldots \cdot (a_1 + a_n^2) \geqslant (a_1 + a_1^2) \cdot (a_2 + a_2^2) \cdot \ldots \cdot (a_n + a_n^2).$$

Хотя я не смог тогда доказать это неравенство, у меня возникло ощущение, что «это всегда так». К тому времени я еще не знал перестановочного неравенства. Некоторые читатели уже, наверное, догадались, как его применить в этой задаче.

Теорема 3. Рассмотрим два набора положительных чисел $a_1, a_2, \ldots, a_n$ и $b_1, b_2, \ldots, b_n$ и все возможные перестановки чисел второго набора. Значение выражения

$$\prod \left(a_i + b_{\sigma(i)}\right)$$

будет самым большим, когда наборы $a_1, a_2, \ldots, a_n$ и $b_{\sigma(1)}, b_{\sigma(2)}, \ldots, b_{\sigma(n)}$ обратно упорядочены, а самым маленьким — когда эти наборы одинаково упорядочены.

Замечание. Вместо условия положительности чисел достаточно потребовать, чтобы для всяких двух индексов i, k число $a_i + b_k$ было положительно.

Обратите также внимание, что знаки неравенства в теореме 3 по сравнению с перестановочным неравенством перевернуты. Объяснение этому будет дано ниже.

Теорема следует из леммы, которая есть ее частный случай для $n = 2$:

Лемма. Пусть $a_1 > a_2 > 0$, $b_1 > b_2 > 0$. Тогда $(a_1 + b_1)(a_2 + b_2) < < (a_1 + b_2)(a_2 + b_1)$.

Упражнение. Доказать лемму и теорему по аналогии с перестановочным неравенством.

Упражнение. Вывести отсюда неравенство

$$(a_2 + a_1^2) \cdot (a_3 + a_2^2) \cdot \ldots \cdot (a_1 + a_n^2) \geqslant (a_1 + a_1^2) \cdot (a_2 + a_2^2) \cdot \ldots \cdot (a_n + a_n^2).$$

Я подозреваю, что автор задачи 5 имел в виду другое ее решение, быть может, использующее неравенство $1 + \dfrac{a^2}{b} \geqslant \dfrac{(1+a)^2}{1+b}$. Пусть, скажем, нас попросили доказать

$$\left(1 + \frac{a_1^5}{a_2^3}\right) \cdot \left(1 + \frac{a_2^5}{a_3^3}\right) \cdot \ldots \cdot \left(1 + \frac{a_n^5}{a_1^3}\right) \geqslant (1 + a_1^2) \cdot (1 + a_2^2) \cdot \ldots \cdot (1 + a_n^2).$$

Упражнение. (а) Докажите это неравенство через перестановочное неравенство.

(б) А теперь через обобщение неравенства $1 + \dfrac{a^2}{b} \geqslant \dfrac{(1+a)^2}{1+b}$. Какое обобщение? Сами догадайтесь!

Разумеется, пункт (а) проще и требует меньше изобретательности. Но мы можем написать и более сложное неравенство, которое из теоремы 3 очевидно, а доказать его напрямую совсем тяжело.

Теперь возьмемся за задачку 7 из списка в начале статьи. Зачеркнуть числа так, как сказано в условии задачи, — это то же самое, что расставить на доске n ладей, которые не бьют друг друга. Но расставить ладьи — это ведь, по сути дела, выбрать перестановку. А какая перестановка даст самое маленькое число? Самая неправильная.

Итак:

Теорема 4. Рассмотрим два набора положительных чисел $a_1, a_2, \ldots, a_n$ и $b_1, b_2, \ldots, b_n$ и все возможные перестановки чисел второго набора. Значение выражения

$$\sum \frac{1}{a_i + b_{\sigma(i)}}$$

будет самым большим, когда наборы $a_1, a_2, \ldots, a_n$ и $b_{\sigma(1)}, b_{\sigma(2)}, \ldots, b_{\sigma(n)}$ одинаково упорядочены, а самым маленьким — когда эти наборы обратно упорядочены.

И теорема, и задача о таблице легко следуют из следующей леммы.

Лемма. $\dfrac{1}{k+x} + \dfrac{1}{m+x} \geqslant \dfrac{1}{x} + \dfrac{1}{k+m+x}$.

Все детали доказательства леммы, теоремы и задачи про таблицу предоставляются читателю. Они не содержат новых нетривиальных идей.

Есть еще один способ решить задачу про таблицу, воспользовавшись известным неравенством

$$(a_1 + a_2 + \cdots + a_n)\Big(\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}\Big) \geqslant n^2.$$

Оно следует из теоремы 2. Но его можно доказать и через неравенство Коши-Буняковского, и с помощью двух неравенств Коши, и другими способами. Впрочем, мы отвлеклись от основной идеи.

Основная идея же, как наверняка уже сообразили самые догадливые из читателей, состоит в том, что хотя теоремы 1, 3, 4 и выглядят очень общими, но все они являются узкими частными случаями чего-то гораздо более общего. И у этого «гораздо более общего» масса таких частных случаев. Но для того чтобы сформулировать это «гораздо более общее», необходимо знать, что такое выпуклая функция.

Некоторые читатели, наверное, знакомы с выпуклыми функциями, но на всякий случай мы напомним определение и основные свойства.

Выпуклые и вогнутые функции

Надграфик функции $f(x)$ — это множество таких точек (x, y), для которых $y \geqslant f(x)$ (т. е. они находятся «над графиком»).

Выпуклая функция — это такая функция, у которой надграфик — выпуклое множество. Иначе говоря, если соединить две точки на графике или над графиком отрезком, то весь отрезок будет лежать в надграфике.

Достаточно потребовать этого для точек, принадлежащих графику. Иными словами, функция должна лежать под «хордой». Можно записать это алгебраически: если $0 < t < 1$, то

$$t \cdot f(x_1) + (1-t) \cdot f(x_2) \geqslant f(t \cdot x_1 + (1-t) \cdot x_2).$$

Вогнутая функция — это такая функция, что, наоборот, подграфик (который определяется аналогично надграфику) является выпуклым множеством, т. е.

$$t \cdot f(x_1) + (1-t) \cdot f(x_2) \leqslant f(t \cdot x_1 + (1-t) \cdot x_2).$$

Функция может быть выпуклой или вогнутой как на всей прямой, так и на каком-то отрезке или луче.

Слова «выпуклая» и «вогнутая» очень похожи. Чтобы не запутаться, какая функция выпуклая, а какая вогнутая, можно думать о веселых и грустных функциях и вспоминать рот у нужного смайлика (см. рис. 1).

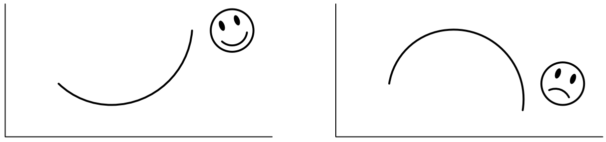
Примеры выпуклых функций: e^x, a^x, $|x|$, x^2, x^{2k} (k — целое положительное) на всей прямой, $\operatorname{tg}(x)$ на промежутке $[0; \pi/2)$, $1/x$ и x^n (n — целое) на положительных числах.

Примеры вогнутых функций: $\ln(x)$ и $\log_a(x)$ на положительных числах, $\sin(x)$ на отрезке $[0; \pi]$, $\cos(x)$ на отрезке $[-\pi/2; \pi/2]$, $\operatorname{arctg}(x)$ на положительных числах.

Вообще, функция x^α на положительных числах является вогнутой, если α лежит между 0 и 1, и выпуклой в противном случае.

Перечислим основные свойства выпуклых функций, начиная с определения. (Всё, что можно сказать о выпуклых функциях, можно сказать и о вогнутых. При этом нужно перевернуть знак неравенства.)

1. Надграфик — выпуклое множество.



а) б)

Рис. 1. *а) веселая (выпуклая) и б) грустная (вогнутая) функции*

2. Функция лежит под хордой: если $0 < t < 1$, то
$$t \cdot f(x_1) + (1-t) \cdot f(x_2) \geqslant f(t \cdot x_1 + (1-t) \cdot x_2).$$
3. Функция лежит над продолжением хорды, а именно, если $t < 0$ или $1 < t$, то
$$t \cdot f(x_1) + (1-t) \cdot f(x_2) \leqslant f(t \cdot x_1 + (1-t) \cdot x_2).$$
4. $\dfrac{f(x_1)+f(x_2)}{2} \geqslant f\Big(\dfrac{x_1+x_2}{2}\Big)$. Для непрерывной функции этого достаточно, чтобы функция была выпуклой, а в общем случае — нет. Но для любой выпуклой функции это свойство выполняется — достаточно подставить $t = 1/2$.
5. $\dfrac{k \cdot f(x_1) + m \cdot f(x_2)}{k+m} \geqslant f\Big(\dfrac{k \cdot x_1 + m \cdot x_2}{k+m}\Big)$. Здесь k, m — натуральные числа. Обобщение пункта 4 и частный случай пункта 2. На самом деле это свойство выводится (но сложным способом) из пункта 4.
6. $\dfrac{n \cdot f(x_1) - m \cdot f(x_2)}{n-m} \leqslant f\Big(\dfrac{n \cdot x_1 - m \cdot x_2}{n-m}\Big)$. Здесь $n > m$ — натуральные числа. Частный случай пункта 3.
7. Если у выпуклой функции есть производная, то функция лежит над касательной, т. е.
$$f(x) \geqslant f(x_0) + f'(x_0) \cdot (x - x_0).$$
8. Если у функции есть вторая производная, то она неотрицательна. Значит, если есть вторая производная и она отрицательна (хотя бы в одной точке), то функция уже не является выпуклой. С помощью этого свойства проще всего проверять выпуклость.
9. Неравенство Йенсена:
$$\frac{f(x_1)+f(x_2)+\dots+f(x_n)}{n} \geqslant f\Big(\frac{x_1+x_2+\dots+x_n}{n}\Big).$$
Центр тяжести многоугольника, вершины которого лежат на графике функции, находится над графиком. А куда ему деться? Надграфик же выпуклый.
10. Неравенство Йенсена с положительными весами:
$$\frac{w_1 f(x_1)+w_2 f(x_2)+\dots+w_n f(x_n)}{w_1+w_2+\dots+w_n} \geqslant f\Big(\frac{w_1x_1+w_2x_2+\dots+w_nx_n}{w_1+w_2+\dots+w_n}\Big).$$
11. $\dfrac{f(a)+f(b)}{2} \geqslant \dfrac{\int\limits_a^b f(x)\,dx}{b-a} \geqslant f\Big(\dfrac{a+b}{2}\Big)$.

Чтобы было интересней, нужно самостоятельно проверить, какие функции являются выпуклыми, а какие — вогнутыми (а не верить автору на слово), и что вытекает из этого и из перечисленных выше свойств.

К примеру, неравенство Йенсена для экспоненты или логарифма становится неравенством Коши (среднее арифметическое не меньше, чем среднее геометрическое). Из всех n-угольников, вписанных в окружность, наибольшие площадь и периметр — у правильного (это следует из неравенства Йенсена для синуса). А из всех n-угольников, описанных вокруг данной окружности, наименьшие площадь и периметр — у правильного (это следует из неравенства Йенсена для тангенса).

Из пункта 7 следует, что $e^x \geqslant x+1$, или, например, что $e^x \geqslant ex$. Отсюда можно понять, что больше: e^π или π^e. Ведь $e^{\pi/e-1} \geqslant \pi/e$.

Интересно также подставить x^n или e^x в пункт 11. Вообще, мы знаем много выпуклых и вогнутых функций, а у них много интересных свойств, и отсюда можно получить много замечательных неравенств.

Доказательства всех свойств, их формулировку и доказательство для вогнутых функций, множество частных случаев для конкретных функций и различные следствия из них читатель может получить сам в качестве упражнения. Чем больше времени вы потратите на это упражнение, тем больше оно доставит вам радости.

Кроме того, читатель должен сам понять, когда в неравенствах достигается равенство и чем отличается строго выпуклая функция (такая, как x^2) от нестрого выпуклой (такой, как $|x|$).

Общее неравенство

Итак, мы готовы сформулировать и доказать обобщение теорем 1, 3 и 4.

Теорема 5. Пусть f — непрерывная функция. Следующие условия эквивалентны.

(а) Значение выражения $f(a_1+b_{\sigma(1)})+f(a_2+b_{\sigma(2)})+\dots+f(a_n+b_{\sigma(n)})$ будет самым большим, когда числа упорядочены одинаково.

(б) Значение выражения $f(a_1+b_{\sigma(1)})+f(a_2+b_{\sigma(2)})+\dots+f(a_n+b_{\sigma(n)})$ будет самым маленьким, когда числа обратно упорядочены.

(в) Если $a_1>a_2$, $b_1>b_2$, то $f(a_1+b_1)+f(a_2+b_2) \geqslant f(a_1+b_2)+f(a_2+b_1)$.

(г) f выпукла.

Ну и конечно, то же самое с обратным знаком.

Теорема 5′. Пусть f — непрерывная функция. Следующие условия эквивалентны.

(а) Значение выражения $f(a_1+b_{\sigma(1)})+f(a_2+b_{\sigma(2)})+\dots+f(a_n+b_{\sigma(n)})$ будет самым большим, когда числа обратно упорядочены.

(б) Значение выражения $f(a_1+b_{\sigma(1)})+f(a_2+b_{\sigma(2)})+\dots+f(a_n+b_{\sigma(n)})$ будет самым маленьким, когда числа упорядочены одинаково.

(в) Если $a_1>a_2$, $b_1>b_2$, то $f(a_1+b_1)+f(a_2+b_2)\leqslant f(a_1+b_2)+f(a_2+b_1)$.

(г) f вогнута.

Если функция выпукла/вогнута только на каком-то промежутке, а не на всей прямой, нужно потребовать во всех пунктах предыдущих теорем, чтобы для всяких двух индексов i, k величина a_i+b_k принадлежала промежутку, на котором функция выпукла/вогнута.

Доказательство. Пункт (в) является частным случаем пункта (а) и частным случаем пункта (б) при $n=2$.

Но и обратно: из пункта (в) выводятся пункты (а) и (б) при помощи сортировки. Это мы уже делали.

Осталось установить эквивалентность пунктов (в) и (г).

Обозначим $x=a_2+b_2$, $k=a_1-a_2$, $m=b_1-b_2$. Тогда формула (в) приобретает вид $f(x+k+m)+f(x)\geqslant f(x+k)+f(x+m)$.

Подставим $k=m$ и получим одно из свойств выпуклой функции (номер 4 в нашем списке). Оно вообще-то слабее выпуклости, но для непрерывных функций эти два свойства равносильны. Значит, из (в) следует (г).

Чтобы из (г) вывести (в), нам понадобится взглянуть на картинку (рис. 2).

На графике отмечено 4 точки:

$$(x,f(x)),\quad (x+k,f(x+k)),\quad (x+m,f(x+m)),\quad (x+k+m,f(x+k+m)).$$

Две средние точки лежат под хордой, соединяющей две крайние точки $(x,f(x))$ и $(x+k+m,f(x+k+m))$ — ведь функция лежит под хордой. Значит, середина отрезка, соединяющего две средние точки — а она имеет координаты $\left(\dfrac{x+k+x+m}{2},\dfrac{f(x+k)+f(x+m)}{2}\right)$, — лежит под отрезком, соединяющим две крайние точки. Середина верхнего отрезка

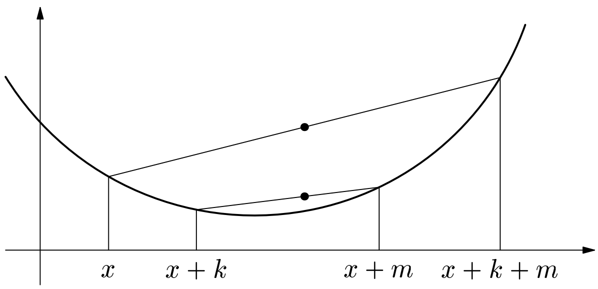


Рис. 2.

$\left(\frac{x+k+x+m}{2}, \frac{f(x+k+m)+f(x)}{2}\right)$ имеет такую же абсциссу. Но она выше, чем середина нижнего отрезка! Значит,

$$\frac{f(x+k+m)+f(x)}{2} \geqslant \frac{f(x+k)+f(x+m)}{2},$$

что и требовалось доказать.

Полностью аналогично доказывается теорема $5'$.

Упражнение. Подставьте выпуклую функцию e^x в теорему 5 и получите частный случай теоремы 1, когда числа в наборах положительны. Выведите общий случай из этого частного.

Упражнение. Подставьте вогнутую функцию $\ln(x)$ в теорему 5 и получите теорему 3.

Вот почему знаки неравенств в теоремах 1 и 3 разные: экспонента «веселая», а логарифм «грустный».

Упражнение. Подставьте $1/x$ в теорему 5 и получите теорему 4.

Упражнение. Подставьте функцию $\sqrt{x}$ (она грустная) в теорему 5 и получите новую теорему.

Упражнение. Поиграйте с другими функциями и получите еще много разных неравенств.

Монгольское неравенство

Пусть $a_1 \geqslant a_2 \geqslant \dots \geqslant a_n$. Тогда

$$\frac{a_1+a_2}{2} \cdot \frac{a_2+a_3}{2} \cdot \ldots \cdot \frac{a_n+a_1}{2} \leqslant$$
$$\leqslant \frac{a_1+a_2+a_3}{3} \cdot \frac{a_2+a_3+a_4}{3} \cdot \ldots \cdot \frac{a_{n-1}+a_n+a_1}{3} \cdot \frac{a_n+a_1+a_2}{3}.$$

Условие $a_1 \geqslant a_2 \geqslant \dots \geqslant a_n$ нельзя отбросить. Например, последовательность 100, 1, 100, 1, 100, 1, … не удовлетворяет этому неравенству.

В упоминавшейся выше статье [4] рассказана история этого неравенства и даны два сложных доказательства. Одно из них фактически доказывает обобщение монгольского неравенства.

Возьмем логарифм от обеих частей:

$$\ln\left(\frac{a_1+a_2}{2}\right) + \ln\left(\frac{a_2+a_3}{2}\right) + \dots + \ln\left(\frac{a_n+a_1}{2}\right) \leqslant$$
$$\leqslant \ln\left(\frac{a_1+a_2+a_3}{3}\right) + \ln\left(\frac{a_2+a_3+a_4}{3}\right) + \cdots$$
$$\cdots + \ln\left(\frac{a_{n-1}+a_n+a_1}{3}\right) + \ln\left(\frac{a_n+a_1+a_2}{3}\right).$$

Мы утверждаем, что это правильно не только для логарифма, но и для любой грустной функции. А для любой веселой функции правильно такое же неравенство, но с противоположным знаком.

Так вот, в статье Храброва это доказывается через неравенство Караматы, которое является общим неравенством для выпуклых функций. У нас есть свое, не менее общее, неравенство для выпуклых функций, а именно теорема 5. Значит, можно попробовать вывести обобщение монгольского неравенства из теоремы 5.

Это примерно то же самое, что и вывести само монгольское неравенство из теоремы 3.

Мы для наглядности разберем частный случай (вывод монгольского неравенства из теоремы 3), а общий случай делается точно так же.

Доказательство монгольского неравенства. Пусть

$$x_1 = \frac{a_1}{3} + \frac{a_2}{6}\,,\ x_2 = \frac{a_2}{3} + \frac{a_3}{6}\,,\ \ldots,\ x_{n-1} = \frac{a_{n-1}}{3} + \frac{a_n}{6}\,,\ x_n = \frac{a_n}{3} + \frac{a_1}{6}\,,$$
$$y_1 = \frac{a_1}{6} + \frac{a_2}{3}\,,\ y_2 = \frac{a_2}{6} + \frac{a_3}{3}\,,\ \ldots,\ y_{n-1} = \frac{a_{n-1}}{6} + \frac{a_n}{3}\,,\ y_n = \frac{a_n}{6} + \frac{a_1}{3}\,.$$

Тогда монгольское неравенство приобретает вид

$$(x_1 + y_1)\cdot(x_2 + y_2)\cdot\ldots\cdot(x_n + y_n) \leqslant (x_1 + y_2)\cdot(x_2 + y_3)\cdot\ldots\cdot(x_n + y_1)$$

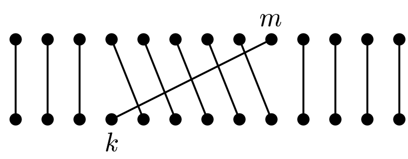
Заметим, что $x_1 \geqslant x_2 \geqslant x_3 \geqslant \cdots \geqslant x_{n-1}$ и $y_1 \geqslant y_2 \geqslant y_3 \geqslant \cdots \geqslant y_{n-1}$.

Но, к сожалению, мы не можем сказать, что $x_{n-1} \geqslant x_n$ или $y_{n-1} \geqslant \geqslant y_n$, — это просто неверно. Мы можем только сказать, что $x_1 \geqslant x_n$, а $y_{n-1} \leqslant y_n$.

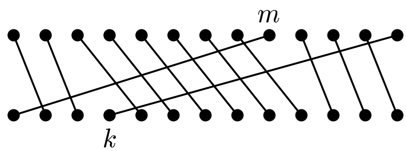
Если мы отсортируем наборы чисел по убыванию, то число x_n займет в списке иксов место m (не первое), а число y_n займет в списке игреков место k (не последнее).

Заметим, что $x_i \geqslant y_i$ при $i \leqslant n-1$, но $x_n \leqslant y_n$. Значит, место x_n в списке иксов не ближе к началу, чем место y_n в списке игреков, т. е. $k \leqslant m$.

Перестановка, представляющая левую часть, будет выглядеть так:



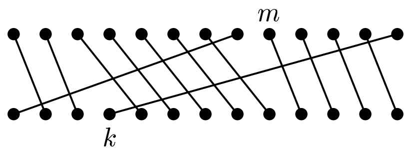
(самый большой с самым большим, второй со вторым и т. д., но номер m с номером k). А перестановка из правой части выглядит вот так (проверьте):



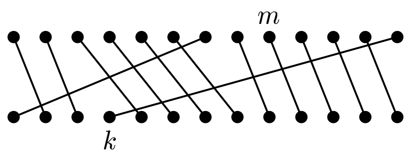
И ни одна из этих перестановок не является самой правильной или самой неправильной. Значит, монгольское неравенство не следует из теоремы 3. Увы. Мы похвастались, что из наших общих теорем следует монгольское неравенство, а оно из них не следует.

Но мы не сдадимся. Вспомним не формулировку, а доказательство теоремы 3, которое, как отмечено выше, аналогично доказательству теоремы 1. Там мы «улучшали» перестановку, и значение менялось монотонно.

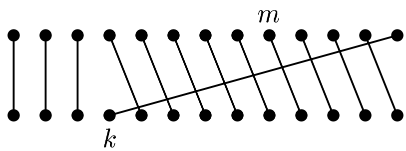
Во второй перестановке верхнее m связано с нижним 1, а верхнее $m-1$ связано с нижним $m+1$. Это «неправильно». Давайте поменяем их местами.



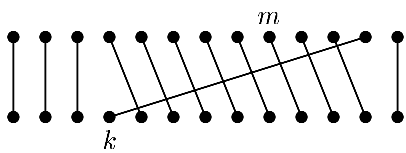
Теперь верхнее $m-1$ связано с нижним 1, а верхнее $m-2$ связано с нижним m. Это опять «неправильно». Давайте поменяем и их местами.



И так далее, пока самый большой игрек не окажется в паре с самым большим иксом.



Теперь k-й по величине игрек оказывается в паре с самым маленьким иксом. А последний игрек в паре с предпоследним иксом. Улучшим это.



Сейчас k-й игрек оказывается в паре с $(n-1)$-м иксом, а $(n-1)$-й игрек в паре с $(n-2)$-м иксом. Улучшим и это. Так за несколько шагов придем к первой перестановке.

Итак, хотя первая перестановка и не «самая правильная», а вторая не «самая неправильная», но первая перестановка «правильнее», чем вторая. А логарифм — функция грустная, значит, левая часть неравенства меньше правой. Что и требовалось доказать.

(Мораль такова: нужно знать не только формулировки теорем, но и их доказательства).

Функция $\ln(1+e^x)$

Вспомним неравенство номер 5 из списка в начале статьи:

$$\left(1+\frac{a_1^2}{a_2}\right)\cdot\left(1+\frac{a_2^2}{a_3}\right)\cdot\ldots\cdot\left(1+\frac{a_n^2}{a_1}\right)\geqslant(1+a_1)\cdot(1+a_2)\cdot\ldots\cdot(1+a_n).$$

Перед тем как его доказывать, мы умножили его на знаменатель. Но, как заметил Илья Гринглаз, его можно доказать и напрямую. А именно, $\prod\left(1+b_{\sigma(i)}/a_i\right)$ достигает минимума при самой правильной, а максимума при самой неправильной перестановке. Это тоже обобщение нашего неравенства, и оно эквивалентно теореме 3. Обозначая $c_i=1/a_i$, приходим к выводу, что $\prod\left(1+c_i\cdot b_{\sigma(i)}\right)$ достигает максимума при одинаковом, а минимума — при обратном порядке.

Этот факт, в свою очередь, эквивалентен выпуклости функции $\ln(1+e^x)$.

Выпуклость функции $\ln(1+e^x)$ эквивалентна неравенству Йенсена для этой функции, т. е.

$$\ln\left(1+e^{\sum x_i/n}\right)\leqslant\frac{\sum\ln(1+e^{x_i})}{n}.$$

Применим экспоненту и получим $1+\sqrt[n]{\prod a_i}\leqslant\sqrt[n]{\prod(1+a_i)}$, где $a_i=e^{x_i}$. А это — неравенство Гюйгенса (о неравенстве Гюйгенса см. [1]). Его легко доказать напрямую. Если возвести неравенство в n-ю степень, то оно легко разобьется в сумму неравенств, которые следуют из неравенства Коши.

Так неожиданно оказывается, что теорема 3 эквивалентна неравенству Гюйгенса. Мы не станем доказывать утверждения этого параграфа, а оставим их читателям в качестве упражнений.

Список литературы

[1] Балк М. Б., Паравян Н. А. *Неравенства Гюйгенса и их применение* // Математика в школе, 1974. №2. С. 70–74.

[2] Маршалл А., Олкин И. *Неравенства: теория мажоризации и ее приложения.* М.: Мир, 1983.

[3] Харди Г. Г., Литтльвуд Дж. Е., Полиа Г. *Неравенства.* М.: ИЛ, 1948.

[4] Храбров А. И. *Вокруг монгольского неравенства* // Математическое просвещение, сер. 3, 2003. Вып. 7. С. 149–162.

Л. В. Радзивиловский, алгоритмист, Aspectus, Petah-Tikva, Israel.

E-mail: `levr78@hotmail.com`
ICQ: 129069668

Избранные задачи 27 Турнира Городов

С. А. Дориченко

В этой небольшой заметке приводятся условия и решения нескольких красивых задач, предлагавшихся в октябре 2005 года участникам осеннего основного тура 27 Турнира городов.

По краю многоугольного стола ползут два муравья. Все стороны стола длиннее 1 *м, а расстояние между муравьями всегда ровно* 10 *см. Сначала оба муравья находятся на одной из сторон стола.*
а) *Пусть стол выпуклый. Всегда ли муравьи смогут проползти по краю стола так, чтобы в каждой точке края побывал каждый из муравьев?*
б) *Пусть стол не обязательно выпуклый. Всегда ли муравьи смогут проползти по краю стола так, чтобы на краю не осталось точек, в которых не побывал ни один из муравьев?* (А. В. Акопян)

Ответ в обоих пунктах отрицательный, хотя многие школьники пытались доказать обратное.

Обозначим первого муравья буквой P, а второго — буквой Q. По условию длина отрезка PQ всегда равна 10 см.

а) Рассмотрим ромб $ABCD$, где $AC = 5$ см, $BD = 2$ м, изначально муравьи находятся на одной из сторон. Для удобства объяснения расположим ромб так, чтобы диагональ BD была горизонтальной.

Посмотрим, как может двигаться отрезок PQ. Пусть сначала P левее Q. Если бы в результате некоторого движения муравьев P оказался правее Q, то в процессе этого движения отрезок PQ в какой-то момент должен был быть вертикальным (что ясно из непрерывности движения). Но это невозможно: самый длинный вертикальный отрезок с концами на

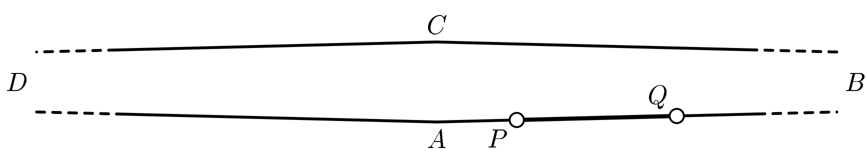


Рис. 1. *Муравьи ползают по краю выпуклого стола*

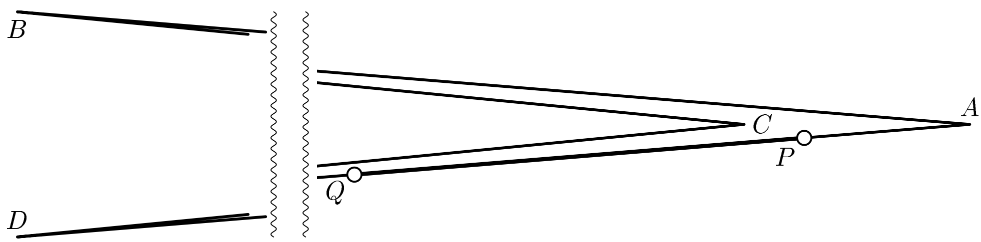


Рис. 2. *Муравьи ползают по краю невыпуклого стола*

сторонах ромба — это AC, а его длина меньше 10 см. Значит P всегда будет левее Q, откуда Q никогда не попадет в левую вершину ромба (а P — в правую).

б) Рассмотрим невыпуклый четырехугольник $ABCD$, похожий на узкую и вытянутую букву V, повернутую на $90°$ против часовой стрелки: диагональ AC горизонтальна и имеет длину 5 см, A правее C, диагональ BD вертикальна и тоже имеет длину 5 см; $CB = CD = 2$ м (точки A и C обе правее B).

Пусть изначально муравьи находятся на стороне AD, P правее Q. Как и в пункте а), отрезок PQ никогда не будет вертикальным. Поэтому муравей P снова всегда будет правее Q.

Докажем, что ни один из муравьев не сможет попасть в точку C.

Если бы один муравей попал в точку C, другой в этот момент был бы левее его (так как самая дальняя от C точка не левее ее — это точка A, но $AC < 10$ см). Поэтому муравей Q не сможет оказаться в C: ведь он тогда был бы правее P. А муравей P не может оказаться в C, поскольку не может покинуть сторон AD и AB: попасть на две оставшиеся стороны он мог бы только через одну из вершин B или D, но в этот момент Q оказался бы правее P.

Еще один интересный вопрос не был включен в олимпиаду из-за его сложности: *всегда ли муравьи смогут проползти по краю выпуклого стола так, чтобы на краю не осталось точек, в которых не побывал ни один из муравьев?*

А во второй половине прошлого века на семинаре Е. М. Ландиса в Московском Государственном Университете разбирался следующий, близкий по теме, вопрос:

На прямой дана непрерывная функция f, равная нулю вне некоторого отрезка $[a, b]$. На оси абсцисс слева от отрезка $[a, b]$ расположен отрезок l. Всегда ли можно передвинуть его в координатной плоскости так, чтобы его концы все время оставались на графике функции f и двигались бы непрерывно, а в итоге l оказался бы на оси абсцисс справа от отрезка $[a, b]$?

Предлагаем желающим читателям ответить на эти вопросы самостоятельно!

У Карлсона есть 1000 *банок с вареньем. Банки не обязательно одинаковые, но в каждой не больше, чем* $1/100$ *часть всего варенья. На завтрак Карлсон может съесть поровну варенья из любых* 100 *банок. Докажите, что Карлсон может действовать так, чтобы за некоторое количество завтраков съесть всё варенье.* (Д. Мусатов)

Задача была предложена ученикам 8 и 9 классов, но оказалась очень сложной: пока ее решил лишь один участник (к моменту написания этой заметки были проверены не все работы). Приведем решение, написанное членом жюри Турнира Городов А. В. Николаевым.

Предположим, Карлсон умеет съедать все варенье, если вместо $1/100$ и 100 в условии указано $1/99$ и 99 (общее количество банок неважно, главное, что их достаточно много, не меньше 100). Будем говорить про эти задачи соответственно «задача-100» и «задача-99». Объясним, как тогда действовать Карлсону в случае «задачи-100».

Он мысленно делит все банки варенья на самую большую (по количеству варенья) и остальные. Заметим, что для «остальных» банок выполняется условие «задачи-99» (ведь в «остальных» банках не меньше, чем $99/100$ всего варенья, и в каждой из этих банок не более $1/100$ всего варенья, то есть не более $(1/100) \cdot (100/99) = 1/99$ от количества варенья в «остальных» банках).

Поэтому Карлсон мог бы действовать так: съесть из «остальных» банок все варенье по алгоритму «задачи 99», на каждом шаге беря набор 99 банок из «остальных» и добавляя все время одну и ту же сотую банку — «самую большую». Для того, чтобы варенье в «самой большой» банке и «остальных» кончилось одновременно, необходимо, чтобы в ней было ровно в 99 раз меньше варенья, чем суммарно в «остальных» банках (поскольку, разумеется, из нее каждый раз съедается в 99 раз меньше, чем из «остальных»). То есть необходимо, чтобы в «самой большой» банке изначально была ровно $1/100$ доля от общего количества варенья.

Объясним, как Карлсону добиться этого. Пусть в «самой большой» банке меньше $1/100$ общего количества варенья. Карлсон будет выбирать 100 непустых банок из «остальных» и съедать из них некоторое количество варенья. Доля варенья в «самой большой» банке при этом будет увеличиваться. Покажем, как ему действовать, чтобы гарантированно довести эту долю до $1/100$. Если количество варенья в самой маленькой банке (из выбранных ста) позволяет съесть часть варенья так, чтобы доля «самой большой» банки стала равна $1/100$, он так и делает. Иначе съедает все варенье из самой маленькой банки, уменьшая количество непустых банок. Когда он остановится? Либо когда добьется требуемого, либо когда непустых банок среди «остальных» станет меньше 100. Но в последнем случае доля «самой большой» точно не меньше $1/100$, т.е. Карлсон должен был остановиться раньше.

Итак, сначала Карлсон подготавливает банки: добивается, чтобы в «самой большой» банке была ровно $1/100$ доля от общего количества варенья (при этом условие «задачи 100», очевидно, будет по-прежнему выполнено). Затем, используя алгоритм «задачи-99», съедает все варенье, как написано выше.

Для завершения решения осталось заметить, что мы научились сводить «задачу-100» к «задаче-99», но точно также можно свести ее теперь к «задаче-98», ту в свою очередь к «задаче-97», и т. д. Ну а «задача-1» совершенно очевидна.

На окружности расставлено несколько положительных чисел, каждое из которых не больше 1. *Докажите, что можно разделить окружность на три дуги так, что суммы чисел на соседних дугах будут отличаться не больше, чем на* 1. (*Замечание. Если на дуге нет чисел, то сумма на ней считается равной нулю.*) (М. И. Малкин)

Верна и более общая задача: в тех же условиях можно для любого натурального числа n разделить окружность на n дуг так, что суммы чисел на соседних дугах будут отличаться не больше, чем на 1.

Сначала разобъем окружность на n дуг произвольным образом. Пусть найдутся две соседние дуги, суммы чисел на которых отличаются больше, чем на 1. Тогда передвинем границу этих дуг так, чтобы на дуге с большей суммой стало на одно число меньше, а на дуге с меньшей суммой — на это же одно число больше. Если и после такого изменения найдутся соседние дуги с разностью, большей 1, снова сделаем аналогичное изменение, и т.д. Остается вопрос: приведет ли наш алгоритм когда-нибудь к цели, или он может продолжаться бесконечно долго?

Автор задачи М. И. Малкин придумал простое и красивое доказательство корректности алгоритма. Занумеруем дуги (числами от 1 до n) и рассмотрим сумму $x_1^2+\ldots+x_n^2$, где x_i — сумма чисел на i-й дуге. Пусть, например, $x_{i+1}-x_i>1$. Применив наш алгоритм, получим две новые соседние дуги с суммами $x_{i+1}-a$ и x_i+a. При этом разница между старой суммой квадратов и новой равна $x_{i+1}^2+x_i^2-(x_{i+1}-a)^2-(x_i+a)^2 = 2a(x_{i+1}-x_i-a)>0$, поскольку $x_{i+1}-x_i>1$ и $0\leqslant a<1$. Значит, сумма квадратов уменьшилась. Но сумма квадратов может принимать лишь конечное число значений (поскольку есть лишь конечное число способов разбить числа на окружности на n частей) и значит не может уменьшаться бесконечное число раз. Поэтому в конце концов мы придем к искомому разделению.

Точно так же доказывается следующий факт: для любого натурального n любой конечный набор положительных чисел, не превосходящих 1, можно разбить на n частей (среди которых могут быть пустые, с нулевой суммой) так, что суммы чисел в любых двух частях будут отличаться не больше, чем на 1.

История очередной задачи очень интересна. В журнале «Квант» №8 за 1983 год И. Ф. Шарыгин опубликовал статью «Вокруг биссектрисы», в которой сформулировал забавный вопрос, мало известный тогда даже среди любителей геометрии: *Про данный треугольник известно, что треугольник, образованный основаниями его биссектрис, является равнобедренным. Можно ли утверждать, что и данный треугольник равнобедренный*? И. Ф. Шарыгин доказал, что утверждать это нельзя (подробности см. например в его книге «Геометрия. Планиметрия. 9–11 классы», Дрофа, 2001, задача 500), но не сумел построить конкретного примера неравнобедренного треугольника (т. е. точно указать величины всех его углов) с таким экзотическим свойством. Недавно пример удалось построить С. И. Токареву, причем треугольник оказался хорошо известным любителям геометрии. Так возникла следующая задача:

Дан треугольник ABC, AA_1, BB_1 и CC_1 — его биссектрисы. Известно, что величины углов A, B и C относятся как $4:2:1$. Докажите, что $A_1B_1 = A_1C_1$. (С. И. Токарев)

Опишем вокруг треугольника ABC окружность. Углы нашего треугольника равны $4\pi/7$, $2\pi/7$ и $\pi/7$; поэтому, разделив окружность на 7 равных дуг, начиная с точки A, получим вписанный в окружность правильный семиугольник $ABXYZCT$.

Заметим сначала, что AA_1 — биссектриса угла A, и значит проходит через точку Y — середину дуги BC. Аналогично BB_1 — биссектриса угла B, и значит проходит через точку T — середину дуги AC.

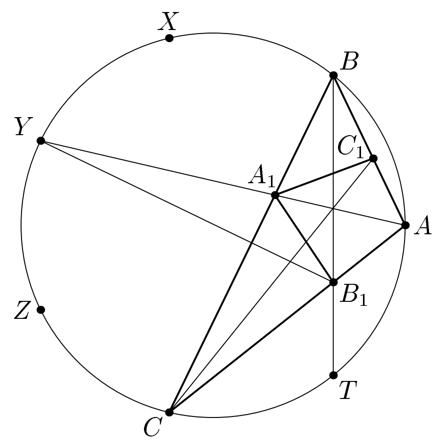


Рис. 3.

Докажем, что $\triangle BCC_1 = \triangle AYB_1$. Действительно, $BC = AY$ из симметрии; $\angle C_1BC = 2\pi/7 = \angle B_1AY$ (опираются на одинаковые дуги), и $\angle BCC_1 = \pi/14 = \angle AYB_1$ (первое равенство верно, так как CC_1 — биссектриса угла C, второе — так как YB_1 — биссектриса угла AYT из симметрии). Тем самым треугольники равны по второму признаку. А значит $YB_1 = CC_1$.

Но тогда треугольники YA_1B_1 и CA_1C_1 равны по первому признаку ($\angle A_1YB_1 = \pi/14 = \angle A_1CC_1$, $A_1Y = A_1C$ из симметрии, $YB_1 = CC_1$ по доказанному). Откуда $A_1B_1 = A_1C_1$, что и требовалось доказать.

С. А. Дориченко, Турнир Городов, НМУ, школы №57 и №179 г.Москвы
E-mail: `dsa@mccme.ru`

По мотивам задачника «Математического просвещения»

В этом номере мы приводим решение задачи 5.9 из задачника «Математического просвещения». Эта задача связана с дискретным аналогом *гармонических функций*, т. е. функций, удовлетворяющих уравнению Лапласа

$$\Delta f = 0.$$

Гармонической на целочисленной решетке $\mathbb{Z}^n$ называется такая функция $f\colon \mathbb{Z}^n \to \mathbb{R}$, значение которой в каждой точке равно среднему арифметическому от значений соседей:

$$f(x_1, \ldots, x_n) = \frac{1}{2n} \sum_{i=1}^{n} \Big(f(x_1, \ldots, x_i - 1, \ldots, x_n) + f(x_1, \ldots, x_i + 1, \ldots, x_n) \Big).$$

Супергармонической на целочисленной решетке $\mathbb{Z}^n$ называется такая функция, для которой вместо написанного выше равенства выполняется неравенство $\geqslant$ в каждой точке $(x_1, \ldots, x_n)$.

В непрерывном случае для гармонических функций справедлива теорема Лиувилля: *всякая ограниченная на* $\mathbb{R}^n$ *гармоническая функция постоянна*. Достаточно даже потребовать односторонней ограниченности. Аналогичное утверждение выполняется и для функций, гармонических на решетке: *если гармоническая на решетке функция неотрицательна, то она постоянна*. Частными случаями этого утверждения являются пункты б) и в) задачи 5.9[1]. Его доказательство, найденное германским школьником П. Шольце, приводится ниже. Другое доказательство этого факта, использующее бесконечномерную выпуклую геометрию, намечено в книге Е. Б. Дынкина, А. А. Юшкевича «Теоремы и задачи о процессах Маркова», М.: Наука, 1966. В этой замечательной книге объясняется связь между гармоническими и супергармоническими функциями на решетке и случайными блужданиями. В частности, на одномерной и двумерной решетках *все супергармонические неотрицательные функции постоянны*. Начиная с размерности 3, существуют непостоянные супергармонические неотрицательные функции. (Разница между малыми и большими

[1] Отметим также, что пункт б) можно найти в задачнике второй серии сборников «Математическое просвещение» (вып. 3, 1958, с. 269, задача 20).

размерностями обусловлена в данном случае возвратностью или невозвратностью случайного блуждания.)

В статье И. И. Богданова и Г. Р. Челнокова приводится решение пункта а) задачи 5.9. Там речь идет о естественном обобщении понятия супергармонических функций на одномерной решетке. Среди таких обобщенных супергармонических функций всюду неотрицательными являются только константы, как и в предыдущих случаях. Заметим, что этот результат также можно получить, изучая случайные блуждания на решетке.

Редколлегия «Математического просвещения» планирует продолжить публикации на эту тему.

Тем, кто заинтересовался дискретными гармоническими функциями, советуем также прочитать статью А. И. Храброва «Дискретные гармонические функции» в сборнике «Задачи Санкт-Петербургской олимпиады школьников по математике 2005 года», СПб: Невский диалект, БХВ-Петербург, 2005; с. 112–145.

Обобщенные супергармонические последовательности

И. И. Богданов Г. Р. Челноков

В этой заметке приводится решение задачи 5.9а) из задачника «Математического просвещения». Напомним формулировку:

> В клетках бесконечной клетчатой ленты записаны положительные числа. Известно, что каждое число не меньше среднего арифметического трех соседей слева и трех справа. Докажите, что числа равны.

Формально числа, записанные в клетках бесконечной клетчатой ленты, можно рассматривать либо как функции $\mathbb{Z} \to \mathbb{R}$ из множества целых чисел в множество действительных чисел, либо как бесконечные в обе стороны последовательности действительных чисел

$$\ldots, f_{-n}, \ldots, f_{-1}, f_0, f_1, \ldots, f_n, \ldots$$

Мы выбираем второй способ и далее говорим только о последовательностях.

Определение. Пусть F — непустое конечное множество целых чисел, симметричное относительно нуля, причем числа в F взаимно просты в совокупности. Последовательность a_n назовем *F-супергармонической*, если для любого n выполняется неравенство

$$a_n \geqslant \frac{1}{|F|} \sum_{f \in F} a_{n+f} \, .$$

Например, $\{-1, +1\}$-супергармонические последовательности — это то же самое, что супергармонические на $\mathbb{Z}^1$ функции.

Теорема. *Если F-супергармоническая последовательность неотрицательна, то она постоянна.*

Задача 5.9а) является частным случаем этой теоремы, в котором $F = \{-3, -2, -1, 1, 2, 3\}$. Для простоты изложения мы приводим только доказательство этого частного случая. Общее доказательство получается аналогично.

Бесконечные в обе стороны последовательности чисел образуют векторное пространство относительно операций покомпонентного сложения и умножения всех членов на константу. Зададим на этом векторном пространстве линейный оператор X сдвига влево: $X(a_n) = (b_n)$, где $b_n = a_{n+1}$.

Нас будут интересовать многочлены от X относительно стандартных операций: произведение операторов есть их композиция, например, оператор X^2 переводит последовательность (c_n) в (c_{n+2}); сложение операторов переводит последовательность в сумму ее образов, например, оператор X^2+2X+1 переводит последовательность (c_n) в $(c_{n+2}+2c_{n+1}+c_n)$; число λ понимается как оператор умножения каждого члена последовательности на λ. Каждый такой многочлен является линейным оператором.

Напомним, что в этом случае композиция линейных операторов совпадает с произведением многочленов, т.е. композиция операторов $P(X)$ и $Q(X)$ есть оператор $PQ(X)$. Например, $X^2+3X+2 = (X+1)(X+2)$, т.е., подействовав на последовательность сначала оператором $X+1$, а затем $X+2$ (или наоборот), мы получим тот же результат, как при действии оператором X^2+3X+2.

На этом языке удобно записываются многие понятия. Так, множество линейных рекуррент с уравнением $u_{n+k} = a_0u_n + a_1u_{n+1} + \dots + a_{k-1}u_{n+k-1}$ — это множество последовательностей, для которых

$$X^k(u_n) = (a_0 + a_1X + \dots + a_{k-1}X^{k-1})(u_n),$$

или, проще говоря, просто множество всех последовательностей, обнуляемых оператором $X^k - a_{k-1}X^{k-1} - \dots - a_1X - a_0$, т.е. его ядро. Отсюда, раскладывая этот многочлен на линейные сомножители, нетрудно получить общую формулу линейной рекурренты.

Условие задачи 5.9а) формулируется теперь в следующем виде. Дана такая последовательность (a_n) положительных чисел, что

$$6a_{n+3} \geqslant a_n + a_{n+1} + a_{n+2} + a_{n+4} + a_{n+5} + a_{n+6},$$

или, другими словами, применение многочлена

$$P(X) = 1 + X + X^2 - 6X^3 + X^4 + X^5 + X^6$$

к последовательности (a_n) дает неположительную последовательность. Требуется доказать, что (a_n) есть константа.[1)]

Заметим, что $P(x) = (x-1)^2Q(x)$, причем все коэффициенты $Q(x)$ неотрицательны.[2)] Действительно, достаточно проверить этот факт для

[1)]При доказательстве теоремы многочлен $P(X)$ нужно заменить на соответствующий множеству F возвратный многочлен (F-многочлен).

[2)]Аналогичное утверждение справедливо и для произвольных F-многочленов, доказательство повторяется почти дословно.

многочленов $x^{k+r}-2x^k+x^{k-r}$, потому что $P(x)=(x^6-2x^3+1)+(x^5-2x^3+x)+(x^4-2x^3+x^2)$. Для многочленов указанного вида утверждение очевидно:

$$x^{k+r}-2x^k+x^{k-r}=x^{k-r}(x^r-1)^2=(x-1)^2x^{k-r}(x^{r-1}+x^{r-2}+\dots+1)^2.$$

Далее, корни $Q(z)$ по модулю не равны 1, т. е. $Q(z)\neq 0$ при любом комплексном z, $|z|=1$. Действительно, $|P(z)|\geqslant|6z^3|-1-|z|-|z^2|-|z^4|-|z^5|-|z^6|=0$, причем равенство может достигаться лишь когда числа 1 и z имеют равные аргументы, т. е. $z=1$. Но $Q(1)\neq 0$ из положительности коэффициентов.[3)]

Обозначим $Q(X)(a_n)=(b_n)$. Тогда элементы (b_n) получаются из элементов (a_n) линейной комбинацией с положительными коэффициентами (а именно, с коэффициентами многочлена Q); таким образом, последовательность (b_n) положительна. По условию, последовательность $(d_n)=(X-1)^2(b_n)=P(X)(a_n)$ неположительна. Покажем, что (b_n) постоянная. Пусть $(c_n)=(X-1)(b_n)$, тогда $(d_n)=(X-1)(c_n)$. По условию получаем $d_n=c_{n+1}-c_n\leqslant 0$, т. е. (c_n) не возрастает. Предположим, что существует $c_k\neq 0$. Если $c_k<0$, то и все $c_l\leqslant c_k<0$ при $l>k$. Тогда $b_{l+1}-b_l\leqslant c_k$ при $l>k$, т. е. b_l с возрастанием l на каждом шаге убывает хотя бы на $|c_k|$; это невозможно, так как (b_n) положительна. Аналогично, если $c_k>0$, то при всех $l<k$ с уменьшением l величина b_l уменьшается хотя бы на c_k, что невозможно. Итак, $c_k\equiv 0$, что и означает, что (b_n) постоянная.

Так как коэффициенты q_i многочлена $Q(x)$ и члены последовательности (a_n) положительны, то $b_n=q_0a_n+q_1a_{n+1}+\dots\geqslant q_0a_n$, т. е. $a_n\leqslant b_n/q_0=b_0/q_0$, и последовательность (a_n) ограничена. Далее, заметим, что применение оператора $Q(X)$ к постоянной последовательности умножает ее на $Q(1)$. Поскольку последовательность $Q(X)(a_n)$ постоянная и $Q(1)\neq 0$, то из (a_n) можно вычесть постоянную последовательность $c=b_0/Q(1)$ и получить такую (ограниченную!) последовательность $a'_n=a_n-c$, что $Q(X)(a'_n)=(0)$.

Решение хотелось бы закончить так. Мы получили, что (a'_n) есть линейная рекуррента, у характеристического уравнения которой нет корней, по модулю равных 1. Применив общую формулу линейной рекурренты, получаем, что (a'_n) есть линейная комбинация последовательностей вида $(R(n)\alpha^n)$, где α пробегает множество корней $Q(x)$, а $R(n)$ — многочлен. Казалось бы, поскольку $|\alpha|\neq 1$, то такая последовательность

[3)]При доказательстве теоремы аналогичное утверждение нужно доказать для произвольного F-многочлена. В этом доказательстве существенно используется условие, что числа из F взаимно просты.

Дальнейшие рассуждения опираются на указанные свойства F-многочленов, а в остальном от вида F они не зависят.

ограничена лишь когда $a'_n \equiv 0$. Это и в самом деле верно, но известные нам доказательства основаны на достаточно тонких рассуждениях с применением теоремы Кронекера.

Поэтому завершим доказательство несколько иначе (и проще).

Мы должны показать, что, применяя $Q(X)$ к ограниченной ненулевой последовательности, нельзя получить (0). Разложим $Q(x)$ на линейные множители; тогда достаточно доказать это для одного такого множителя $X-\alpha$, где $|\alpha| \neq 1$ (действительно, поскольку этот оператор переводит ограниченные последовательности в ограниченные, то по индукции сразу получается требуемое). Однако $X-\alpha$, очевидно, обнуляет лишь последовательности вида $u_n = \alpha^n u_0$, которые при $|\alpha| \neq 1$ и $u_0 \neq 0$ не ограничены, что и требовалось доказать.

Замечание. Похожими методами легко решается, например, задача 7.11 из задачника «Математического просвещения». Напомним ее условие:

> Все комплексные корни уравнения $A_0X^n + A_1X^{n-1} + \dots + A_n = 0$ по модулю строго меньше 1. Последовательность $\{v_k = A_0u_{k+n} + A_1u_{k+n-1} + \dots + A_nu_k\}$ — сходится. Докажите, что последовательность $\{u_k\}$ тоже сходится.

Достаточно понять, что на классе ограниченных последовательностей верно операторное равенство $(1-aX)(1+aX+a^2X^2+\dots) = 1$, если $|a| < 1$.

И. И. Богданов, МФТИ
e-mail: iibogdanov@yahoo.com

Г. Р. Челноков

О неотрицательных гармонических функциях на решетке

П. Шольце

Гармонической на целочисленной решетке $\mathbb{Z}^n$ называется такая функция $f\colon \mathbb{Z}^n \to \mathbb{R}$, значение которой в каждой точке равно среднему арифметическому от значений соседей:

$$f(x_1,\ldots,x_n)=\frac{1}{2n}\sum_{i=1}^{n}\Bigl(f(x_1,\ldots,x_i-1,\ldots,x_n)+f(x_1,\ldots,x_i+1,\ldots,x_n)\Bigr).$$

Данная заметка посвящена доказательству следующего утверждения, из которого, в частности, следуют пункты б) и в) задачи 5.9 задачника «Математического просвещения».

Теорема. *Гармоническая на решетке неотрицательная функция постоянна.*

Аналогично одномерному случаю (см. предыдущую статью И. И. Богданова и Г. Р. Челнокова) мы введем операторы сдвига X_1, …, X_n, действующие на наших функциях по следующему правилу:

$$(X_i f)(x_1,\ldots,x_n)=f(x_1,\ldots,x_i+1,\ldots,x_n).$$

Таким образом, X_i «сдвигает» расстановку чисел в узлах решетки на 1 по i-й координате. Условие гармоничности функции записывается с помощью операторов сдвига как

$$(X_1+X_1^{-1}+X_2+X_2^{-1}+\cdots+X_n+X_n^{-1})f=2nf$$

или

$$(X_1+X_1^{-1}+X_2+X_2^{-1}+\cdots+X_n+X_n^{-1}-2n)f=0,$$

где 0 — тождественно нулевая расстановка.

Операторы сдвига коммутируют. Поэтому можно рассматривать многочлены от операторов X_i, X_i^{-1} с действительными коэффициентами, понимая умножение операторов как композицию, а сложение и умножение на число — как сложение и умножение на число функций $\mathbb{Z}\to\mathbb{R}^n$.

Перевод и редактирование И. И. Богданова.

Заметим, что, применяя многочлен от сдвигов X_i, X_i^{-1} к гармонической функции, мы опять получаем гармоническую функцию. При этом, если коэффициенты многочлена и исходная функция неотрицательны, то и полученная функция неотрицательна.

Мы будем вести доказательство индукцией по n. Доказательство для $n = 1$ тривиально, поскольку одномерная гармоническая функция линейна, а ограниченная линейная функция постоянна.

Для шага индукции достаточно доказать, что функция f из условия теоремы удовлетворяет неравенству

$$2(n-1)f \geqslant (X_2 + X_2^{-1} + \dots + X_n + X_n^{-1})f \tag{1}$$

(подобные неравенства мы всегда будем воспринимать поточечно; в частности, данное неравенство означает, что значение левой функции в каждой точке не меньше, чем соответствующее значение правой). Действительно, тогда из симметрии выполняются неравенства такого же вида с заменой индекса 1 на i; просуммировав по всем i, мы получим, что

$$(2n - X_1 - X_1^{-1} - X_2 - X_2^{-1} - \dots - X_n - X_n^{-1})f \geqslant 0.$$

Поскольку это неравенство обращается в равенство, то обращались в равенство и все слагаемые, т.е. при любом j

$$\Big(\sum_{i\neq j}(X_i + X_i^{-1})\Big)f = 2(n-1)f.$$

Зафиксируем произвольное $a \in \mathbb{Z}$. Тогда функция

$$f_1\colon (x_2, \dots, x_n) \mapsto f(a, x_2, \dots, x_n)$$

является неотрицательной гармонической функцией на $(n-1)$-мерной решетке, по предположению индукции она постоянна; по тем же причинам и функция

$$f_2\colon (x_1, x_3, \dots, x_n) \mapsto f(x_1, a, x_2, \dots, x_n)$$

постоянна. А тогда и функция f также постоянна.

Осталось доказать неравенство (1).

Введем многочлен $m(x_2, \dots, x_n) = (2n-2) - x_2 - x_2^{-1} - \dots - x_n - x_n^{-1}$ и обозначим через M соответствующий ему оператор $m(X_2, \dots, X_n)$. Нам нужно доказать, что $Mf \geqslant 0$. Заметим, что $(Mf)(x_1, x_2, \dots, x_n)$ зависит только от значений функции при данном фиксированном x_1.

Сумма сдвигов вдоль первой координаты $(X_1 + X_1^{-1})f$ выражается через сдвиги f вдоль остальных координат; иначе говоря, сумма двух значений f в точках, симметричных относительно плоскости $x_1 = a$, выражается через значения f в этой плоскости. Выясним, как именно она выражается.

Введем многочлены

$$p_0(t) = 2, \quad p_1(t) = 2 + t, \quad p_{k+2}(t) = (2+t)p_{k+1}(t) - p_k(t).$$

Положим $P_k = p_k(M)$.

Лемма 1. *$P_k f = (X_1^k + X_1^{-k})f$ при любом k. Как следствие, $P_k f \geqslant 0$.*

Доказательство. Индукция по n. Для $n = 0, 1$ утверждение очевидно. Для остальных значений n из определения гармонической функции получаем

$$\begin{aligned}
&(X_1^k + X_1^{-k})f = \\
&\quad = (X_1^{k-1}(X_1 + X_1^{-1}) - X_1^{k-2} + X_1^{-k+1}(X_1 + X_1^{-1}) - X_1^{-k+2})f = \\
&\quad = -(X_1^{k-2} + X_1^{k-1}(X_2 + X_2^{-1} + \dots + X_n + X_n^{-1} - 2n))f - \\
&\quad - (X_1^{-k+2} + X_1^{-k+1}(X_2 + X_2^{-1} + \dots + X_n + X_n^{-1} - 2n))f = \\
&\quad = (X_1^{k-1} + X_1^{-k+1})(2+M)f - (X_1^{k-2} + X_1^{2-k})f = \\
&\qquad = ((2+M)P_{k-1} - P_{k-2})f = P_k f,
\end{aligned}$$

что и требовалось доказать. □

Замечание. Из доказательства леммы следует, что любую функцию, заданную на плоскости $x_1 = 0$, можно продолжить гармоническим образом на всю область определения: она восстанавливается из доказанной леммы, *если предположить*, что она симметрична относительно этой же плоскости. При этом, очевидно, она не обязана быть неотрицательной, даже если исходные данные были неотрицательными.

Теперь мы временно забудем про операторы P_k и выясним несколько свойств многочленов $p_k(t)$.

Лемма 2. *$p_k(2\cos\alpha - 2) = 2\cos(k\alpha)$.*

Доказательство. Индукция по k. При $k = 0$ и $k = 1$ это очевидно. Используя определение многочленов $p_k(t)$, получаем

$$\begin{aligned}
&p_{k+2}(2\cos\alpha - 2) = 2\cos\alpha\, p_{k+1}(2\cos\alpha - 2) - p_k(2\cos\alpha - 2) = \\
&\quad = 2(2\cos\alpha\cos(k+1)\alpha - \cos k\alpha) = 2(\cos(k+2)\alpha + \cos k\alpha - \cos k\alpha) = \\
&\qquad = 2\cos(k+2)\alpha
\end{aligned}$$

из предположения индукции. □

Замечание. Из леммы 2 сразу вытекает, что $p_k(t) = 2T_k\big(\frac{t}{2} + 1\big)$, где $T_k(t)$ — многочлен Чебышёва степени k.

Лемма 3. *$p_k(t) = 2 + (t - t_1) \cdot \ldots \cdot (t - t_k)$, где $t_j = 2\cos\frac{2j\pi}{k} - 2$. Как следствие, $p_k(t) \geqslant 2$ при $t \geqslant 0$.*

Доказательство. Согласно лемме 2, корнями многочлена $p_k(t)-2$ являются $t_j = 2\cos\dfrac{2i\pi}{k}-2$, $i=0,1,\ldots,k-1$. Некоторые числа встречаются среди списка t_j дважды. Однако, если $d=t_r=t_j$, то $r+j=k$, и $-4<d<0$; заметим, что при $t\in[-4,0]$ выполняется неравенство

$$|p(t)| = 2\left|\cos\left(k\arccos\frac{t+2}{2}\right)\right| \leqslant 2,$$

а $p(d)-2=0$, поэтому d — внутренняя точка локального максимума $p_k(t)-2$ и, как следствие, его корень кратности 2. Таким образом, мы насчитали k корней многочлена $p_k(t)-2$; кроме того, его старший коэффициент равен 1 (из определения). Тогда $p_k(t)=2+(t-t_1)\cdot\ldots\cdot(t-t_k)$. При этом все t_i неположительны, откуда следует второе утверждение леммы. □

Пусть r — наименьшее неотрицательное число такое, что неравенство $(M+r)g\geqslant 0$ выполняется для *любой* неотрицательной гармонической функции g. Заметим, что $(M+2)g=(X_1+X_1^{-1})g\geqslant 0$ из леммы 1; поэтому r существует и не превосходит 2. Требуемое неравенство (1) эквивалентно тому, что $r=0$.

Предположим, что $r>0$. Разделим многочлен $p_k(t)$ с остатком на $(t-2(n-1))(t+r)$:

$$p_k(t) = (t-2(n-1))(t+r)q_k(t)+r_k(t), \quad \deg r_k \leqslant 1. \tag{2}$$

Лемма 4. *$q_k(t)$ можно представить в виде*

$$q_k(t) = \sum_{l=0}^{k-2} A_l p_l(t),$$

где $A_l\geqslant 0$.

Доказательство. Подставим в (2) $t=2\cos\alpha-2$ и положим $z=e^{i\alpha}$. Тогда $t=z+z^{-1}-2$, а $p_k(t)=z^k+z^{-k}$. Соответственно, частное q_k принимает вид

$$\begin{aligned} q_k(z+z^{-1}-2) &= \frac{z^k+z^{-k}-r_k(z+z^{-1}-2)}{(z+z^{-1}-2n)(z+z^{-1}+(r-2))} = \\ &= A_{k-2}(z^{k-2}+z^{-k+2})+A_{k-3}(z^{k-3}+z^{-k+3})+\cdots+A_0, \end{aligned} \tag{3}$$

поскольку $q_k(z+z^{-1}-2)$ симметрично относительно замены $z\mapsto z^{-1}$. Заметим, что тогда $q_k(t)=A_{k-2}p_{k-2}(t)+\cdots+A_0$, ибо для любого t найдется такое (возможно, комплексное) z, что $t=z+z^{-1}-2$. Поэтому нам надо доказать неотрицательность коэффициентов A_l.

При $z\to\infty$ имеем

$$q_k(z+z^{-1}-2) = A_{k-2}z^{k-2}+A_{k-3}z^{k-3}+\cdots+A_1z+A_0+o(1).$$

С другой стороны,

$$q_k(z+z^{-1}-2)=$$
$$=\frac{z^k}{(z+z^{-1}-2n)(z+z^{-1}+(r-2))}+\frac{z^{-k}-r_k(z+z^{-1}-2)}{(z+z^{-1}-2n)(z+z^{-1}+(r-2))}=$$
$$=\frac{z^k}{(z+z^{-1}-2n)(z+z^{-1}+(r-2))}+o(1),$$

поскольку $r_k(t)$ не более, чем линеен.

Значит, при $z\to\infty$

$$\frac{z^k}{(z+z^{-1}-2n)(z+z^{-1}+(r-2))}=A_{k-2}z^{k-2}+A_{k-3}z^{k-3}+\dots+A_1z+A_0+o(1).$$

Подставив $x=1/z$ и домножив на x^{k-2}, получаем при $x\to 0$

$$\frac{1}{(x^2+1-2nx)(x^2+1+(r-2)x)}=A_{k-2}+A_{k-3}x+\dots+A_1x^{k-3}+A_0x^{k-2}+$$
$$+o(x^{k-2}),$$

т. е. A_l есть коэффициенты ряда Маклорена функции

$$\frac{1}{(x^2+1-2nx)(x^2+1+(r-2)x)}=$$
$$=\frac{x^2-2x+1}{x^2-2nx+1}\cdot\frac{1}{(x^2-2x+1)(x^2+(r-2)x+1)}=u(x)v(x).$$

Достаточно доказать, что ряд Маклорена каждого из сомножителей неотрицателен. Пусть $\nu_1>1$, $\nu_2=1/\nu_1$ — корни многочлена $x^2-2nx+1$, тогда

$$u(x)=\frac{x^2-2x+1}{x^2-2nx+1}=1+\frac{(2n-2)x}{(1-\nu_1x)(1-\nu_2x)}=$$
$$=1+\frac{(2n-2)x}{\nu_1-\nu_2}\left(\frac{\nu_1}{1-\nu_1x}-\frac{\nu_2}{1-\nu_2x}\right)=1+\frac{2n-2}{\nu_1-\nu_2}\left(\sum_{l=1}^{\infty}\nu_1^lx^l-\sum_{l=1}^{\infty}\nu_2^lx^l\right)=$$
$$=1+\frac{2n-2}{\nu_1-\nu_2}\sum_{l=1}^{\infty}(\nu_1^l-\nu_2^l)x^l,$$

где все коэффициенты, очевидно, положительны.

Рассмотрим $v(x)$. Заметим, что у ряда

$$\frac{1}{1-2x+x^2}=1+2x+3x^2+\dots$$

все коэффициенты положительны. Тогда, если $r=0$, то сомножитель

$v(x)$, равный квадрату этого ряда, положителен. Если же $r > 0$, то

$$xv(x) = \frac{x}{(x^2+1-2x)(x^2+1+(r-2)x)} = \frac{1}{r}\left(\frac{1}{1-2x+x^2} - \frac{1}{1+(r-2)x+x^2}\right).$$

Аналогично проведенным выше вычислениям, имеем

$$rxv(x) = \sum_{l=0}^{\infty}(l+1)x^l - \frac{1}{\zeta-\zeta^{-1}}\sum_{l=0}^{\infty}(\zeta^{l+1}-\zeta^{-l-1})x^l = \\ = \sum_{l=0}^{\infty}\left(l+1-\frac{\zeta^{l+1}-\zeta^{-l-1}}{\zeta-\zeta^{-1}}\right)x^l,$$

где ζ, ζ^{-1} — корни трехчлена $x^2+1+(r-2)x$ (они при $0 < r \leqslant 2$ комплексно сопряжены). Осталось заметить, что

$$\left|\frac{\zeta^{l+1}-\zeta^{-l-1}}{\zeta-\zeta^{-1}}\right| = |\zeta^l+\zeta^{l-2}+\dots+\zeta^{-l}| \leqslant l+1,$$

поэтому для (действительных!) коэффициентов ряда $rxv(x)$ верно неравенство

$$l+1-\frac{\zeta^{l+1}-\zeta^{-l-1}}{\zeta-\zeta^{-1}} \geqslant 0,$$

а значит, и коэффициенты ряда $v(x)$ неотрицательны. □

Рассмотрим операторы $Q_k = q_k(M)$, $R_k = r_k(M)$ и произвольную неотрицательную гармоническую функцию g.

Лемма 5. *$R_kg \geqslant 0$.*

Доказательство. Имеем

$$R_kg = P_kg + (M+r)(2n-2-M)Q_kg = \\ = P_kg + (M+r)(X_2+X_2^{-1}+\dots+X_n+X_n^{-1})Q_kg.$$

Заметим, что $P_ig \geqslant 0$ из леммы 1, а поэтому $Q_kg \geqslant 0$ из леммы 4. Тогда применение оператора $(M+r)$ к гармонической неотрицательной функции $(X_2+X_2^{-1}+\dots+X_n+X_n^{-1})Q_kg$ дает снова неотрицательную функцию. Отсюда следует утверждение леммы. □

Завершим доказательство теоремы. Итак, мы предположили, что $r > 0$. Тогда $0 < \varphi = \arccos\frac{2-r}{2} \leqslant \pi$, и существует такое k, что $\pi/2 < k\varphi \leqslant \pi$. При этом k имеем $p_k(-r) = 2\cos k\varphi < 0$ согласно лемме 2. Кроме того, из леммы 3 получаем $p_k(2n-2) \geqslant 2$.

Подставив в (2) $t = 2(n-1)$ и $t = -r$, получаем
$$r_k(2(n-1)) = p_k(2(n-1)), \quad r_k(-r) = p_k(-r),$$
откуда
$$r_k(t) = p_k(2(n-1)) + (x - 2(n-1))\frac{p_k(2(n-1)) - p_k(-r)}{2(n-1)+r}.$$
Тогда
$$R_k g = \frac{p_k(2(n-1)) - p_k(-r)}{2(n-1)+r} Mg + \\ + \left(p_k(2(n-1)) - 2(n-1)\frac{p_k(2(n-1)) - p_k(-r)}{2(n-1)+r}\right) g \geqslant 0,$$
причем коэффициент при M в этом выражении положителен. Разделив на этот коэффициент, получаем
$$\left(M + p_k(2(n-1))\frac{2(n-1)+r}{p_k(2(n-1)) - p_k(-r)} - 2(n-1)\right) g \geqslant 0.$$
Однако
$$p_k(2(n-1))\frac{2(n-1)+r}{p_k(2(n-1)) - p_k(-r)} - 2(n-1) = \\ = (2(n-1)+r)\frac{p_k(2(n-1))}{p_k(2(n-1)) - p_k(-r)} - 2(n-1) < \\ < 2(n-1) + r - 2(n-1) = r;$$
неравенства выполняются, ибо $p_k(2(n-1)) \geqslant 2 > 0 > p_k(-r)$. Получаем противоречие с выбором r. Значит, $r = 0$, откуда следует неравенство (1), из которого, как было показано выше, и следует теорема.

Peter Scholze
e-mail: `Peter.Scholze@web.de`

О прямых Симсона, кривой Штейнера и кубике Мак-Кэя

Е. Д. Куланин

В статье рассматриваются вопросы, затронутые в [3], а также приводится решение задачи 8.4 из задачника «Математического просвещения».

1. Вводные замечания

Напомним сначала некоторые определения. Пусть P — точка в плоскости треугольника ABC, A_1, B_1, C_1 — основания перпендикуляров, опущенных из точки P на прямые BC, CA, AB соответственно. Тогда треугольник $A_1B_1C_1$ называется педальным треугольником точки P относительно треугольника ABC. Прямые, проходящие через вершину данного угла и симметричные относительно биссектрисы этого угла, называются изогональными относительно этого угла. Окружность, на которой лежат середины сторон треугольника, основания его высот и середины отрезков от вершин до точки пересечения высот (ортоцентра), называется окружностью Эйлера или окружностью девяти точек этого треугольника.

Приведем формулировки тех теорем из [3], которые понадобятся нам в этой статье. Желающие могут найти их доказательства в [3], для чего в скобках приводятся номера этих теорем, которые они имеют в [3].

Теорема 1 (7). Пусть P — точка, не лежащая на описанной окружности треугольника ABC. Тогда прямые, изогональные соответственно прямым PA, PB, PC относительно углов A, B, C этого треугольника, пересекаются в одной точке P'.

Точки P и P' называются изогонально сопряженными точками относительно треугольника ABC или просто изогональными точками.

В дальнейшем мы покажем, что если точка P лежит на описанной окружности треугольника ABC, то прямые, изогональные прямым PA, PB, PC относительно углов A, B, C, параллельны. Легко видеть, что центр описанной окружности треугольника и точка пересечения его высот изогонально сопряжены.

Математическое просвещение, сер. 3, вып. 10, 2006(243–264)

Теорема 2 (9). Пусть P и P' — точки, изогонально сопряженные относительно треугольника ABC, а $A_1B_1C_1$ и $A_1'B_1'C_1'$ — педальные треугольники этих точек. Тогда вершины треугольников $A_1B_1C_1$ и $A_1'B_1'C_1'$ лежат на одной окружности.

Так как центр O описанной окружности треугольника и точка H пересечения его высот изогональны, то из теоремы 2 сразу же вытекает, что середины сторон произвольного треугольника и основания его высот лежат на одной окружности. Обозначим через AH_1, BH_2, CH_3 высоты непрямоугольного треугольника ABC. Тогда треугольники AH_2H_3, BH_3H_1, CH_1H_2 подобны треугольнику ABC с коэффициентами подобия $|\cos A|$, $|\cos B|$, $|\cos C|$ соответственно.

Таким образом, треугольники AH_2H_3, BH_3H_1, CH_1H_2 являются уменьшенными копиями треугольника ABC, поэтому можно рассматривать точки, одинаково расположенные относительно подобных треугольников ABC, AH_2H_3, BH_3H_1, CH_1H_2.

Теорема 3 (10). Пусть A_1, B_1, C_1 — проекции точки P на стороны (или продолжения сторон) BC, CA, AB треугольника ABC; P_1, P_2, P_3 — точки, симметричные P относительно середин сторон B_1C_1, C_1A_1, A_1B_1 треугольника $A_1B_1C_1$; H_1, H_2, H_3 — основания высот AH_1, BH_2, CH_3 треугольника ABC. Тогда точки P, P_1, P_2, P_3 одинаково расположены по отношению к треугольникам ABC, AH_2H_3, BH_3H_1, CH_1H_2 соответственно.

Следствие 1. Пусть A_1, B_1, C_1 — проекции точки P на стороны (или продолжения сторон) BC, CA, AB треугольника ABC; H_1, H_2, H_3 — основания высот AH_1, BH_2, CH_3 треугольника ABC. Тогда точки P_1, P_2, P_3, одинаково расположенные с точкой P по отношению к треугольникам AH_2H_3, BH_3H_1, CH_1H_2, ABC, совпадают с ортоцентрами треугольников AB_1C_1, BC_1A_1, CA_1B_1.

Следствие 2 (5). Пусть H_1, H_2, H_3 — основания высот AH_1, BH_2, CH_3 треугольника ABC; точки P, P_1, P_2, P_3 одинаково расположены по отношению к треугольникам ABC, AH_2H_3, BH_3H_1, CH_1H_2 соответственно. Тогда треугольник $P_1P_2P_3$ равен педальному треугольнику $A_1B_1C_1$ точки P относительно треугольника ABC, причем стороны треугольников $P_1P_2P_3$ и $A_1B_1C_1$ соответственно параллельны.

Теорема 4 (11). Пусть P — произвольная точка в плоскости треугольника ABC, не лежащая на его описанной окружности; AH_1, BH_2, CH_3 — его высоты, пересекающиеся в точке H; E_1, E_2, E_3 — середины отрезков AH, BH, CH; P_1, P_2, P_3 — точки, одинаково расположенные с P относительно треугольников AH_2H_3, BH_3H_1, CH_1H_2, ABC соответственно. Тогда прямые E_1P_1, E_2P_2, E_3P_3 пересекаются в такой точке K

окружности Эйлера треугольника ABC, через которую проходит описанная окружность педального треугольника точки P относительно треугольника ABC.

2. Прямые Симсона

Рассмотрим теперь конфигурацию, изображенную на рис. 1. Проведем через ортоцентр H треугольника ABC прямую, которая пересечет описанные окружности треугольников AH_2H_3, BH_3H_1, CH_1H_2 в точках P_a, P_b, P_c соответственно.

Так как $\angle P_aHH_2 = \angle P_cHH_2 = \angle P_bHB$, а $\angle AH_2H_3 = \angle CH_2H_1 = \angle H_1BH_3$, то точки P_a, P_b, P_c одинаково расположены относительно подобных треугольников $AH_2H_3, BH_3H_1, CH_1H_2$. Пусть P — точка, одинаково расположенная с точками P_a, P_b, P_c относительно треугольников ABC, AH_2H_3, BH_3H_1, CH_1H_2; H_a, H_b, H_c — основания перпендикуляров, опущенных из точки P на прямые BC, CA, AB соответственно. Тогда согласно следствию 1 точки P_b, P_c, P_a совпадают с ортоцентрами

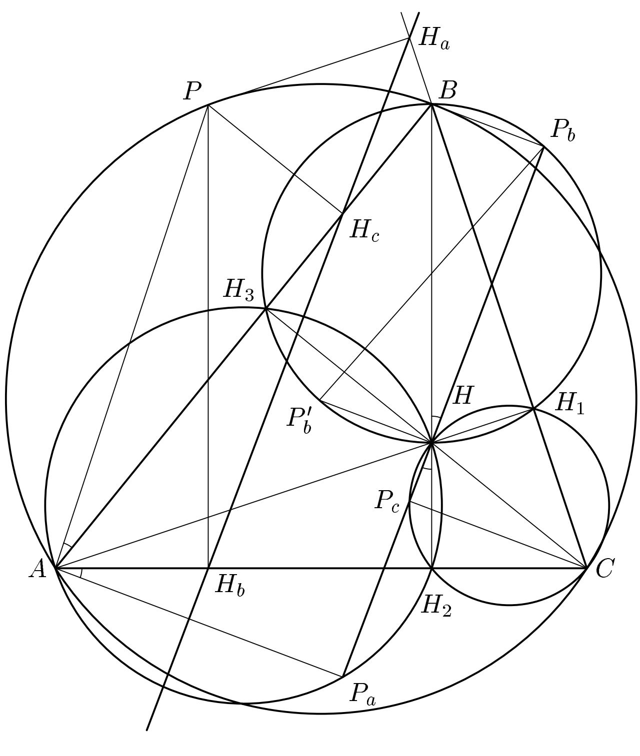


Рис. 1.

треугольников H_aBH_c, H_aCH_b, H_cAH_b соответственно, т. е. $BP_b \perp H_aH_c$, $CP_c \perp H_aH_b$, $AP_a \perp H_bH_c$. Но $\angle BP_bH = \angle CP_cH = \angle AP_aH = 90^\circ$ как вписанные углы, опирающиеся на диаметры описанных окружностей треугольников BH_3H_1, CH_1H_2, AH_2H_3, поэтому $H_aH_c \parallel P_bH$, $H_aH_b \parallel P_cH$, $H_bH_c \parallel P_aH$. Вспомнив, что точки P_a, P_b, P_c, H лежат на одной прямой, получаем, что и точки H_a, H_b, H_c также лежат на одной прямой. Кроме того, $H_aH_b = P_aP_b$, $H_bH_c = P_bP_c$, $H_cH_a = P_cP_a$ как противоположные стороны параллелограммов $P_aH_bH_aP_b$, $P_cH_bH_cP_b$, $P_aH_cH_aP_c$. Таким образом, следствие 2 справедливо и в случае вырожденного педального треугольника. Итак, мы пришли к следующему результату: основания перпендикуляров, опущенных из точки, взятой на описанной окружности треугольника, на его стороны или продолжения сторон, лежат на одной прямой.

Эта прямая называется прямой Симсона.

Теорема 5. Прямая Симсона точки P относительно треугольника ABC проходит через середину отрезка PH, где H — ортоцентр треугольника ABC.

Доказательство. Вернемся к нашей конфигурации, изображенной на рис. 1. Мы установили, что прямая $P_aP_bP_c$ параллельна прямой Симсона точки P, одинаково расположенной с точками P_a, P_b, P_c относительно треугольников ABC, AH_2H_3, BH_3H_1, CH_1H_2 соответственно, поэтому согласно теореме 3 точки P_a, P_b, P_c симметричны точке P относительно середин отрезков H_aH_b, H_bH_c, H_aH_b. Другими словами, прямая Симсона $H_bH_cH_a$ точки P проходит через середины отрезков PP_a, PP_b, PP_c, а, значит, и через середину отрезка PH (напомним, что прямая $P_aP_bP_c$ проходит через H). $\square$

Далее, пусть P_b' — точка, диаметрально противоположная P_b относительно описанной окружности треугольника BH_3H_1. Тогда $\angle P_b'HP_b = 90^\circ$, т. е. $P_b'H \perp P_bH$, но $P_b'H$ параллельна прямой Симсона точки P', диаметрально противоположной точке P относительно описанной окружности треугольника ABC.

Таким образом, получено

Предложение 1. Прямые Симсона диаметрально противоположных точек перпендикулярны.

Пусть точка P движется против часовой стрелки по описанной окружности треугольника ABC с постоянной угловой скоростью ω. Тогда угловые меры дуг BP и BP_b равны и $\angle BHP_b = \frac{1}{2}\smile BP_b = \frac{1}{2}\omega t$. Это означает, что прямая $P_aP_cHP_b$ вращается вокруг точки H по часовой стрелке с постоянной угловой скоростью $\omega/2$. Прямая Симсона точки P, параллельная прямой P_aP_b, также вращается с угловой скоростью $\omega/2$ по часовой стрелке вокруг некоторого переменного центра вращения. Другими словами, если радиус OP вращается с постоянной угловой скоростью ω, то прямая Симсона точки P вращается с постоянной угловой скоростью $-\omega/2$.

Предложение 2. Прямая Симсона точки P, лежащей на описанной окружности треугольника ABC, перпендикулярна прямым, симметричным прямым PA, PB, PC относительно биссектрис углов A, B, C треугольника ABC соответственно.

Доказательство. Проведем отрезки AP_a, BP_b, CP_c. Так как отрезки AH, BH, CH являются диаметрами описанных окружностей треугольников AH_2H_3, BH_3H_1, CH_1H_2, то $\angle AP_aH = \angle BP_bH = \angle CP_cH = 90^\circ$ и поэтому отрезки AP_a, BP_b, CP_c параллельны. Но поскольку точки P, P_a, P_b, P_c одинаково расположены относительно треугольников ABC, AH_2H_3, BH_3H_1, CH_1H_2, то прямые AP_a, BP_b, CP_c симметричны прямым AP, BP, CP относительно биссектрис углов A, B, C треугольника ABC. Осталось только вспомнить, что прямая $P_aP_cHP_b$ параллельна прямой Симсона точки P. □

Использовав определение изогональных прямых, предложение 2 можно переформулировать следующим образом:

Предложение 3. Прямая Симсона точки P, лежащей на описанной окружности треугольника ABC, перпендикулярна прямым, изогональным прямым PA, PB, PC относительно углов A, B, C треугольника ABC.

Заметим, что попутно мы доказали, что для точки P описанной окружности треугольника ABC прямые, изогональные прямым PA, PB, PC относительно углов A, B, C треугольника ABC, параллельны. Будем считать, что параллельные прямые пересекаются в бесконечно удаленной точке. Тогда можно сказать, что точка P', изогональная точке P, лежащей на описанной окружности треугольника ABC, бесконечно удалена. Верно и обратное, т. е. точка P, изогональная бесконечно удаленной точке P', относительно треугольника ABC, лежит на описанной окружности этого треугольника. Другими словами, справедливо

Предложение 4. Прямые ℓ_a, ℓ_b, ℓ_c, изогональные параллельным прямым ℓ'_a, ℓ'_b, ℓ'_c относительно треугольника ABC, пересекаются в точке, лежащей на описанной окружности этого треугольника.

Непосредственным следствием предложения 4 является

Теорема 6. Пусть прямая ℓ, проходящая через центр описанной окружности треугольника ABC, одинаково расположена с прямыми ℓ_a,

ℓ_b, ℓ_c относительно треугольников ABC, AH_2H_3, BH_3H_1, CH_1H_2. Тогда прямые ℓ_a, ℓ_b, ℓ_c пересекаются в точке, лежащей на окружности Эйлера этого треугольника.

Доказательство. Пусть E — центр окружности Эйлера, а H — ортоцентр треугольника ABC; E_1, E_2, E_3 — середины отрезков AH, BH, CH. Проведем через точку E прямую ℓ_1, параллельную ℓ, а через точки E_1, E_2, E_3 — прямые ℓ_a, ℓ_b, ℓ_c, одинаково расположенные с прямой ℓ относительно треугольников ABC, AH_2H_3, BH_3H_1, CH_1H_2. Поскольку треугольники AH_2H_3, BH_3H_1, CH_1H_2 при симметрии относительно соответствующих биссектрис треугольника ABC переходят в треугольники, гомотетичные треугольнику ABC, то прямые ℓ_a, ℓ_b, ℓ_c изогональны относительно треугольника $E_1E_2E_3$ прямым ℓ'_a, ℓ'_b, ℓ'_c, параллельным прямой ℓ_1 и проходящим через вершины треугольника $E_1E_2E_3$. Поэтому в силу предложения 4 прямые ℓ_a, ℓ_b, ℓ_c пересекаются в точке, лежащей на описанной окружности треугольника $E_1E_2E_3$, совпадающей с окружностью Эйлера треугольника ABC. □

Теорема 7. Пусть P — произвольная точка описанной окружности треугольника ABC; AH_1, BH_2, CH_3 — его высоты, пересекающиеся в точке H; E_1, E_2, E_3 — середины отрезков AH, BH, CH; P_1, P_2, P_3 — точки, одинаково расположенные с P относительно треугольников AH_2H_3, BH_3H_1, CH_1H_2, ABC соответственно. Тогда прямые E_1P_1, E_2P_2, E_3P_3 пересекаются в такой точке K окружности Эйлера треугольника ABC, через которую проходит прямая Симсона точки P относительно треугольника ABC.

Доказательство. Обозначим через A_1, B_1, C_1 проекции точки P на стороны BC, CA, AB треугольника ABC, $\angle CAB = \alpha$, $\angle ABC = \beta$, $\angle BCA = \gamma$ (рис. 2). Так как P_2 и P_3 — ортоцентры треугольников A_1BC_1 и B_1CA_1 (см. следствие 1), то

$$\angle P_2A_1P_3 = \angle P_2A_1C + \angle P_3A_1C = 90^\circ - \angle C_1P_2A_1 + 90^\circ - \gamma = \\ = 180^\circ - \angle C_1P_2A_1 - \gamma.$$

Но $\angle C_1P_2A_1 = \angle C_1PA_1$ как противоположные углы параллелограмма $PA_1P_2C_1$, а $\angle C_1PA_1 + \angle C_1BA_1 = 180^\circ$, поскольку четырехугольник PA_1BC_1 вписанный ($\angle PA_1B + \angle PC_1B = 90^\circ + 90^\circ = 180^\circ$). Поэтому

$$\angle P_2A_1P_3 = (180^\circ - \angle C_1P_2A_1) - \gamma = \angle C_1BA_1 - \gamma = \\ = 180^\circ - \beta - \gamma = \alpha = \angle E_2KE_3 = \angle P_2KP_3.$$

Из равенства углов P_2KP_3 и $P_2A_1P_3$ следует, что точки P_2, K, A_1, P_3 лежат на одной окружности. Аналогично показывается, что точки P_3, B_1, P_1, K также лежат на одной окружности.

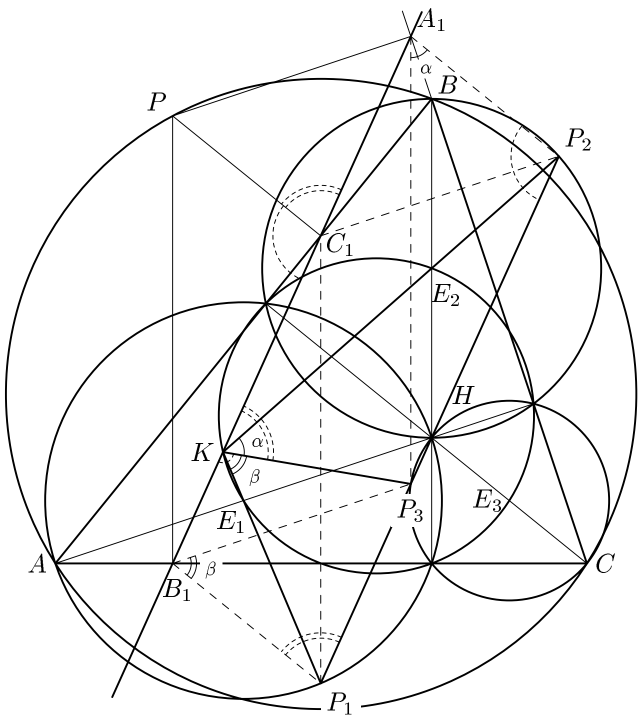


Рис. 2.

Далее, так как точка C_1 лежит на прямой Симсона A_1B_1, то $\angle B_1C_1P+$ $+\angle PC_1A_1 = 180°$, но $\angle B_1C_1P = \angle A_1P_2P_3$, $\angle A_1C_1P = \angle B_1P_1P_3$ как углы с соответственно параллельными сторонами, а из вписанных четырехугольников $A_1P_2P_3K$ и $B_1P_1P_3K$ находим

$$\angle P_3KA_1 = 180° - \angle A_1P_2P_3 = 180° - \angle B_1C_1P = \angle PC_1A_1,$$
$$\angle P_3KB_1 = 180° - \angle B_1P_1P_3 = 180° - \angle A_1C_1P = \angle PC_1B_1.$$

Окончательно получаем $\angle B_1KP_3 + \angle P_3KA_1 = \angle PC_1A_1 + \angle PC_1B_1 = 180°$, откуда следует, что точка K лежит на прямой Симсона A_1B_1. □

Прямые Симсона можно считать описанными окружностями бесконечно большого радиуса вырожденных педальных треугольников, поэтому теорема 7 является аналогом теоремы 4, и, более того, эти теоремы можно объединить в одну:

Теорема 8. Пусть P — произвольная точка в плоскости треугольника ABC; AH_1, BH_2, CH_3 — его высоты, пересекающиеся в точке H; E_1, E_2, E_3 — середины отрезков AH, BH, CH; P_1, P_2, P_3 — точки, одинаково расположенные с P относительно треугольников AH_2H_3, BH_3H_1,

CH_1H_2, ABC соответственно. Тогда прямые E_1P_1, E_2P_2, E_3P_3 пересекаются в такой точке K окружности Эйлера треугольника ABC, через которую проходит описанная окружность педального треугольника точки P относительно треугольника ABC.

Теорему 8 можно переформулировать следующим образом:

Теорема 9. Пусть P — произвольная точка в плоскости треугольника ABC; AH_1, BH_2, CH_3 — его высоты, пересекающиеся в точке H; E_1, E_2, E_3 — середины отрезков AH, BH, CH; O — центр описанной окружности треугольника ABC; ℓ_1, ℓ_2, ℓ_3 — прямые, параллельные прямой OP и проходящие через точки E_1, E_2, E_3 соответственно; ℓ_1', ℓ_2', ℓ_3' — прямые, изогональные ℓ_1, ℓ_2, ℓ_3 относительно треугольника $E_1E_2E_3$. Тогда прямые ℓ_1', ℓ_2', ℓ_3' пересекаются в такой точке K окружности Эйлера треугольника ABC, через которую проходит описанная окружность педального треугольника точки P относительно треугольника ABC.

Теорема 10. Прямые Симсона диаметрально противоположных точек P и P' перпендикулярны и пересекаются в точке, лежащей на окружности Эйлера треугольника ABC, причем вторые точки пересечения этих прямых с окружностью Эйлера являются концами ее диаметра, параллельного PP'.

Доказательство. Согласно теореме 7 прямые Симсона диаметрально противоположных точек P и P' пересекаются в точке K, лежащей на окружности Эйлера треугольника ABC.

Но, как было установлено выше (см. предложение 1), прямые Симсона диаметрально противоположных точек перпендикулярны, поэтому K_1K_1' — диаметр окружности Эйлера, где K_1 и K_1' — вторые точки пересечения прямых Симсона точек P и P' с окружностью Эйлера (рис. 3).

Рассмотрим прямую Симсона ℓ точки P. Точка K пересечения этой прямой с окружностью Эйлера совпадает с точкой пересечения прямых E_1P_1, E_2P_2, E_3P_3, одинаково расположенных с прямой OP относительно треугольников AH_2H_3, BH_3H_1, CH_1H_2, ABC (O — как всегда центр описанной окружности треугольника ABC). Вторая точка K_1 пересечения прямой Симсона ℓ точки P с окружностью Эйлера совпадает согласно теореме 5 с серединой отрезка PH, где H — ортоцентр треугольника ABC.

Аналогично, точка K_1' совпадает с серединой отрезка $P'H$, и, таким образом, диаметр K_1K_1' окружности Эйлера является средней линией треугольника PHP' и, следовательно, параллелен диаметру PP'. □

Замечание. Пусть P' — бесконечно удаленная точка, изогональная точке P. Тогда точка K_1, совпадающая с серединой отрезка PH, изогонально сопряжена с P' относительно треугольника $E_1E_2E_3$ (см.

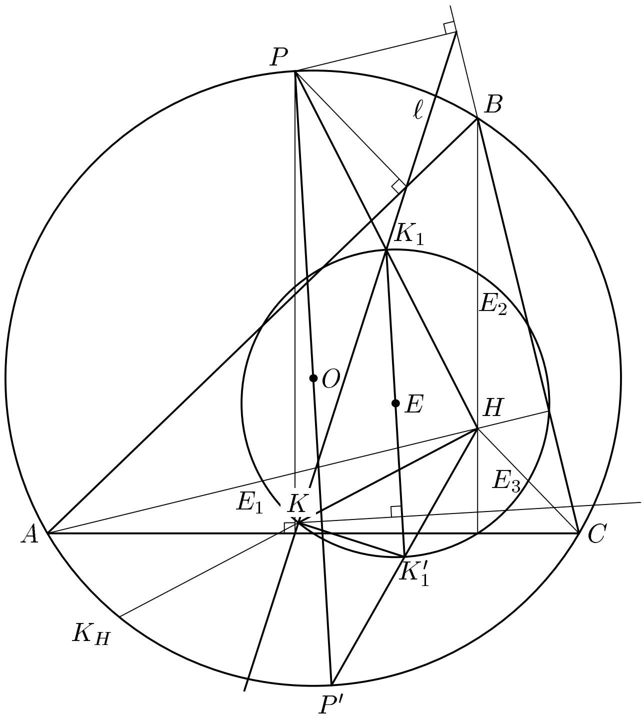


Рис. 3.

предложения 2–4). Тогда, если считать прямую Симсона ℓ точки P общей педальной окружностью точек P и P', то теорема 9 будет справедлива и для бесконечно удаленной точки P'.

Напомним, что точкой Тебо называется такая точка окружности Эйлера треугольника ABC, в которой пересекаются прямые Эйлера треугольников AH_2H_3, BH_1H_3, CH_1H_2, где H_1, H_2, H_3 — основания высот треугольника ABC.

Следствие 3. Пусть прямая Эйлера треугольника ABC пересекает его описанную окружность в точках P и P'. Тогда прямые Симсона точек P и P' пересекаются в точке Тебо треугольника ABC.

Следствие 4. Прямые Симсона концов диаметра описанной окружности треугольника, проходящего через центр его вписанной окружности, пересекаются во внутренней точке Фейербаха этого треугольника, а прямые Симсона концов диаметров, проходящих через центры вневписанных окружностей, пересекаются во внешних точках Фейербаха этого треугольника.

Теорема 11. Пусть P и P' — диаметрально противоположные точки описанной окружности треугольника ABC, K — точка пересечения прямых Симсона точек P и P', лежащая на окружности Эйлера треугольника ABC, K_1 и K_1' — вторые точки пересечения прямых Симсона точек P и P' с окружностью Эйлера, K_H — точка, симметричная ортоцентру H относительно точки K.

Тогда прямая Симсона точки K_H перпендикулярна диаметру K_1K_1' окружности Эйлера треугольника ABC.

Доказательство. Вернемся к рис. 3. Пусть K — точка окружности Эйлера треугольника ABC, а K_H — точка, гомотетичная K с центром гомотетии H и коэффициентом гомотетии 2.

Как мы уже говорили, прямые KE_1, KE_2, KE_3 изогональны диаметральной прямой PP'. Поэтому и параллельные им прямые K_HA, K_HB, K_HC также изогональны этой прямой. Но, как мы знаем (см. предложение 2), прямая Симсона точки P перпендикулярна параллельным прямым, изогональным прямым PA, PB, PC. Отсюда следует, что прямая Симсона точки K_H перпендикулярна диаметру PP', а, значит, и параллельному ему диаметру K_1K_1' окружности Эйлера треугольника ABC. □

Следствие 5. Пусть T_H — точка, симметричная ортоцентру относительно точки Тебо произвольного треугольника. Тогда прямая Симсона точки T_H перпендикулярна прямой Эйлера этого треугольника.

Следствие 6. Прямая Симсона точки, симметричной ортоцентру произвольного треугольника относительно его внутренней точки Фейербаха, перпендикулярна прямой, проходящей через центры вписанной и описанной окружностей этого треугольника.

Докажем теперь следующее утверждение:

Предложение 5. Пусть треугольники ABC и $A'B'C'$ с ортоцентрами H и H' имеют общую окружность Эйлера.

Тогда, если какие-либо две взаимно перпендикулярные прямые Симсона треугольника ABC совпадают с какими-либо двумя взаимно перпендикулярными прямыми Симсона треугольника $A'B'C'$, то совпадают и все остальные прямые Симсона этих треугольников.

Доказательство. Как следует из теоремы 10, совпадающие прямые Симсона P_1K и P_2K треугольников ABC и $A'B'C'$ являются прямыми Симсона концов параллельных диаметров описанных окружностей треугольников ABC и $A'B'C'$, получающихся из диаметра P_1P_2 их общей окружности Эйлера гомотетиями с центрами в точках H и H' и коэффициентом 2 (рис. 4).

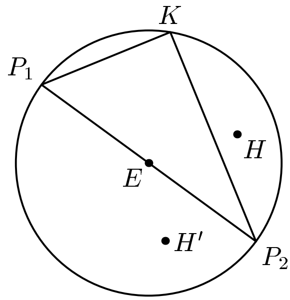


Рис. 4.

Начнем вращать эти параллельные диаметры вокруг центров соответствующих окружностей в одну и ту же сторону с одинаковой угловой скоростью, тогда с той же угловой скоростью и в ту же сторону будет вращаться и соответствующий этим диаметрам диаметр P_1P_2 общей окружности Эйлера треугольников ABC и $A'B'C'$. Прямые Симсона концов вращающихся диаметров будут проходить через середины отрезков, соединяющих эти концы с точками H и H' (для концов диаметра описанной окружности треугольника ABC — с H, а для концов второго диаметра — с H'), т.е. через точки P_1 и P_2.

Но, как было установлено ранее, прямые Симсона концов параллельных диаметров будут вращаться вокруг точек P_1 и P_2 в противоположную сторону с угловой скоростью в два раза меньшей скорости вращения параллельных диаметров.

Итак, прямые Симсона концов параллельных диаметров описанных окружностей треугольников ABC и $A'B'C'$ в начальный момент времени совпадают, а затем начинают вращаться вокруг одних и тех же точек P_1 и P_2 в одну и ту же сторону с одинаковыми угловыми скоростями и поэтому совпадают в любой момент времени.

Более точно: пусть O и O' — центры описанных окружностей треугольников ABC и $A'B'C'$. Тогда, если для двух точек P и P', лежащих на описанных окружностях треугольников ABC и $A'B'C'$, векторы OP и $O'P'$ сонаправлены, то прямые Симсона точек P и P' совпадают. $\square$

3. Касание прямых Симсона с окружностью Эйлера данного треугольника

Напомним следующий факт, являющийся непосредственным следствием теорем 2 и 9.

Пусть P и P' — изогонально сопряженные точки относительно треугольника ABC. Тогда, если прямая PP' проходит через центр описанной окружности этого треугольника, то общая педальная окружность точек P и P' касается окружности Эйлера треугольника ABC.

В случае, когда P лежит на описанной окружности треугольника ABC, прямые, изогональные прямым PA, PB, PC, параллельны, т.е. точка P' бесконечно удалена, а общая педальная окружность точек P и P' вырождается в прямую Симсона точки P. Как мы уже знаем (см. замечание после теоремы 10), бесконечно удаленной точке P' соответствует точка пересечения прямой Симсона точки P с окружностью Эйлера треугольника ABC, совпадающая с серединой отрезка PH, где H — ортоцентр треугольника ABC.

Вторая точка пересечения прямой Симсона точки P с окружностью Эйлера получится в результате пересечения прямых E_1P_1, E_2P_2, E_3P_3, изогональных прямой OP относительно углов E_1, E_2, E_3 треугольника $E_1E_2E_3$ (здесь, как и раньше, E_1, E_2, E_3 — точки Эйлера, т.е. середины отрезков AH, BH, CH). Прямая Симсона точки P будет касаться окружности Эйлера треугольника ABC в том и только в том случае, когда эти две точки совпадают, а для этого необходимо и достаточно, чтобы прямые, изогональные прямым AP, BP, CP относительно углов A, B, C треугольника ABC, были параллельны OP.

Другими словами, прямая OP должна проходить через бесконечно удаленную точку P', т.е. и в этом случае прямая PP' проходит через центр O описанной окружности треугольника ABC.

Теорема 12. Среди всех прямых Симсона произвольного треугольника имеется ровно три, которые касаются окружности Эйлера этого треугольника. Точки касания являются вершинами равностороннего треугольника. Если α, β, γ — величины углов данного треугольника ($\alpha < \beta < \gamma$), то стороны этого равностороннего треугольника образуют с соответствующими сторонами данного треугольника углы, равные $(\gamma - \alpha)/3$, $(\beta - \alpha)/3$, $(\gamma - \beta)/3$.

Доказательство. Пусть AL — биссектриса угла BAC, N — точка, диаметрально противоположная P, K — точка пересечения AL и PN, $1/2\smile AP = \varphi_1$ (рис. 5). Прямые AP и OP изогональны относительно угла BAC тогда и только тогда, когда $\angle PAK = \angle PKA$. Поскольку PN — диаметр, то $\angle PAN = 90^\circ$ и $\angle APK = \angle APN = 90^\circ - \angle ANP = 90^\circ - \varphi_1$ и, предполагая, что $\angle PAK = \angle PKA$, найдем

$$\angle PAK = \frac{1}{2}(180^\circ - \angle APK) = \frac{1}{2}(180^\circ - (90^\circ - \varphi_1)) = \\ = \frac{1}{2}(90^\circ + \varphi_1) = 45^\circ + \frac{1}{2}\varphi_1 .$$

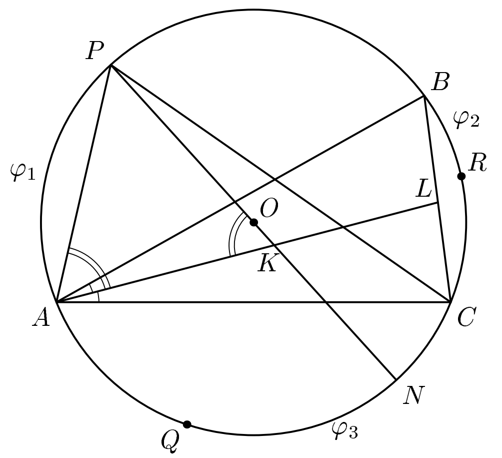


Рис. 5.

С другой стороны,

$$\angle PAK = \angle PAB + \angle BAL = \frac{1}{2}\smile PB + \frac{1}{2}\angle BAC = \\ = \frac{1}{2}\smile APB - \frac{1}{2}\smile AP + \frac{1}{2}\alpha = \gamma - \varphi_1 + \frac{1}{2}\alpha.$$

Поэтому $45^\circ + \varphi_1/2 = \gamma - \varphi_1 + \alpha/2$, откуда $3\varphi_1/2 = \gamma + \alpha/2 - 45^\circ$, $\varphi_1 = 2/3(\gamma + \alpha/2) - 30^\circ = 2/3\gamma + \alpha/3 - 30^\circ$.

Точно так же найдем еще две точки R и Q, для которых их прямые Симсона касаются окружности Эйлера треугольника ABC. Они определяются условиями $\varphi_2 = 1/2\smile BK = 2/3\alpha + \beta/3 - 30^\circ$, $\varphi_3 = 1/2\smile CQ = 2/3\beta + \gamma/3 - 30^\circ$ (как всегда, считаем, что $\alpha < \beta < \gamma$).

Найдем углы треугольника PQR:

$$\angle PQR = \frac{1}{2}\smile PR = \frac{1}{2}\smile PB + \frac{1}{2}\smile BR = \gamma - \varphi_1 + \varphi_2 = \\ = \gamma - (2/3\gamma + 1/3\alpha - 30^\circ) + 2/3\alpha + \beta/3 - 30^\circ = \\ = \gamma/3 + \alpha/3 + \beta/3 = (\alpha + \beta + \gamma)/3 = 180^\circ/3 = 60^\circ.$$

Аналогично, $\angle PRQ = \angle QPR = 60^\circ$.

Итак, мы получили, что треугольник PQR равносторонний. Поскольку точки касания P_1, Q_1, R_1 прямых Симсона точек P, Q, R с окружностью Эйлера треугольника ABC являются серединами отрезков PH, QH, RH, где H — ортоцентр треугольника ABC, то треугольники PQR и $P_1Q_1R_1$ гомотетичны и, следовательно, треугольник $P_1Q_1R_1$ также равносторонний.

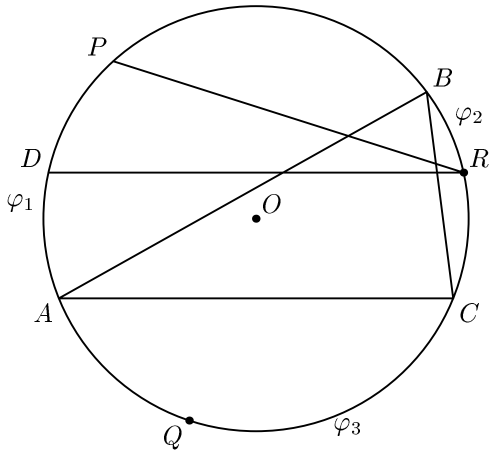


Рис. 6.

Найдем угол между сторонами PR и AC треугольников PQR и ABC. Для этого проведем через точку R хорду RD, параллельную AC (см. рис. 6). Тогда

$$\angle PRD = \frac{1}{2}\smile PD = \frac{1}{2}\smile AP - \frac{1}{2}\smile AD = \frac{1}{2}\smile AP - \frac{1}{2}\smile CR =$$
$$= \frac{1}{2}\smile AP - (\frac{1}{2}\smile BC - \frac{1}{2}BK) = \varphi_1 - (\alpha - \varphi_2) =$$
$$= 2/3\gamma + 1/3\alpha - 30^\circ - (\alpha - (2/3\alpha + \beta/3 - 30^\circ)) =$$
$$= 2/3\gamma + \alpha/3 - 30^\circ - \alpha/3 + \beta/3 - 30^\circ =$$
$$= (\gamma+\alpha+\beta)/3 + \gamma/3 - \alpha/3 - 60^\circ = 180^\circ/3 + (\gamma-\alpha)/3 - 60^\circ = (\gamma-\alpha)/3.$$

Аналогично, углы между сторонами QR и AB, PQ и BC равны соответственно $(\beta-\alpha)/3$ и $(\gamma-\beta)/3$.

Заметим, что в ходе доказательства теоремы 12 установлено, что на описанной окружности каждого треугольника существуют ровно три точки, прямые Симсона которых касаются окружности Эйлера треугольника, причем эти точки совпадают с вершинами правильного треугольника. Данное утверждение составляет содержание задачи 8.4 задачника «Математического просвещения» (вып. 9, 2004, с. 246). Таким образом, попутно мы привели решение этой задачи.

Напомним, что геометрическое место точек P таких, что прямая PP', где P' — точка, изогональная P, проходит через центр описанной окружности треугольника, называется кубикой Мак-Кэя этого треугольника. Из теорем 2 и 9 следует, что кубика Мак-Кэя треугольника совпадает

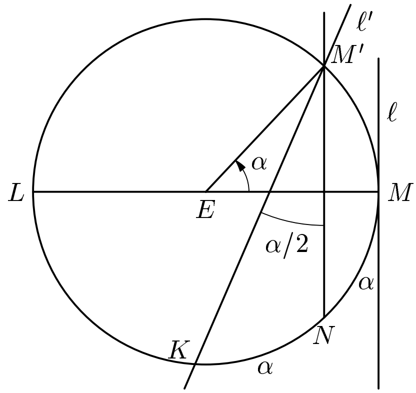


Рис. 7.

с геометрическим местом точек P, обобщенная педальная окружность которых касается окружности Эйлера этого треугольника.

Так как прямые Симсона точек P, Q, R где P, Q, R — вершины равностороннего треугольника, лежащие на описанной окружности треугольника ABC, касаются окружности Эйлера этого треугольника, то точки P, Q, R принадлежат кубике Мак-Кэя треугольника ABC. Таким образом, кубика Мак-Кэя пересекает описанную окружность треугольника в трех точках, являющихся вершинами равностороннего треугольника.

Рассмотрим теперь одну из прямых Симсона, касающихся окружности Эйлера данного треугольника в точке M (рис. 7). Пусть эта прямая Симсона ℓ является прямой Симсона точки P, лежащей на описанной окружности треугольника ABC, и пусть радиус OP повернулся против часовой стрелки на угол α так, что точка P приняла положение P'.

Тогда в силу гомотетичности описанной окружности треугольника ABC и окружности Эйлера этого треугольника с центром в точке пересечения высот H прямая Симсона ℓ' точки P' будет проходить через точку M' окружности Эйлера такую, что $\smile MM' = \alpha$. При этом, как уже неоднократно говорилось, прямая ℓ' повернется вокруг точки M' на угол $\alpha/2$, т. е., если провести хорду $M'N$, параллельную ℓ, то $\angle KM'N = \alpha/2$ и $\smile KN = \alpha$. Дуги $M'M$ и MN симметричны относительно диаметра LM, поэтому $\smile NM = \smile MM' = \alpha$ и $\smile KM = \smile KN + \smile NM = \alpha + \alpha = 2\alpha$.

Итак, справедливо

Предложение 6. Пусть M — точка касания прямой Симсона произвольного треугольника с его окружностью Эйлера. Тогда произвольная прямая Симсона этого треугольника пересекает его окружность Эйлера в точках K и Q таких, что точки K и Q лежат по разные стороны от диаметра, проходящего через точку M и $\smile KM = 2\smile MQ$.

4. О кубике Мак-Кэя и кривой Штейнера

В этом пункте нам понадобится следующее утверждение из [3].

Теорема 13 (12). *Изогональным образом прямой ℓ, проходящей через центр описанной окружности треугольника ABC и не содержащей его вершин, является равносторонняя гипербола, описанная около этого треугольника, причем асимптотами этой гиперболы являются прямые Симсона точек пересечения прямой ℓ с описанной окружностью треугольника ABC.*

Рассмотрим пару взаимно перпендикулярных прямых Симсона, одна из которых касается окружности Эйлера данного треугольника ABC, а вторая проходит через центр этой окружности (рис. 8). Тогда эти прямые являются асимптотами равносторонней гиперболы Γ, описанной около треугольника ABC и совпадающей с изогональным образом прямой ℓ, проходящей через центр O описанной окружности треугольника ABC параллельно той асимптоте, на которой лежит центр окружности Эйлера этого треугольника. Прямая ℓ проходит через точку P пересечения гиперболы Γ с описанной окружностью треугольника ABC, центр O этой окружности и бесконечно удаленную точку P', изогональную точке P, поэтому точки P и P' принадлежат кубике Мак-Кэя треугольника ABC. Если вращать прямую ℓ вокруг точки O по часовой стрелке, то прямая

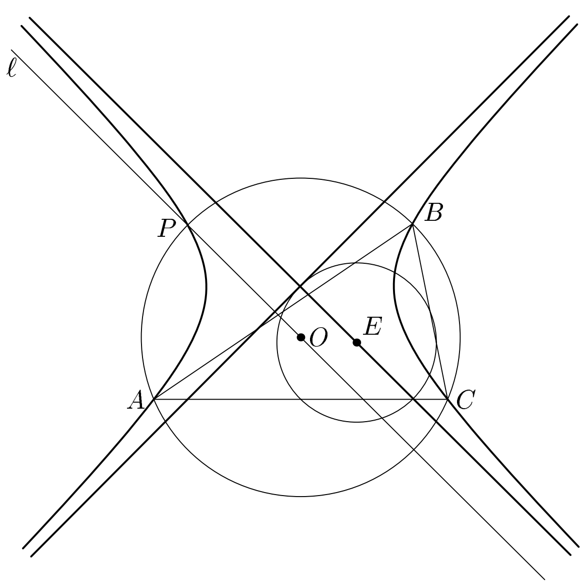


Рис. 8.

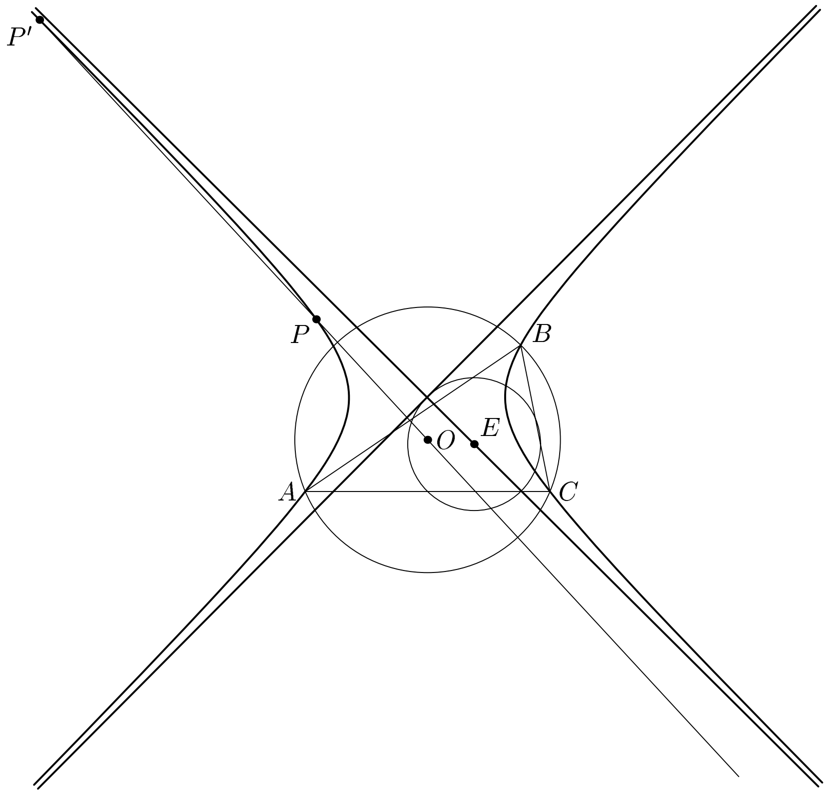


Рис. 9.

Симсона точки P начнет вращаться против часовой стрелки (см. начало пункта 2), соответственно против часовой стрелки будет вращаться и перпендикулярная ей прямая.

Отсюда следует, что прямая ℓ пересечет гиперболу Γ, являющуюся ее изогональным образом, в двух изогональных точках P и P' (рис. 9). Так как прямая ℓ проходит через центр описанной окружности треугольника ABC, то это означает, что точки P и P' принадлежат кубике Мак-Кэя этого треугольника. Но после бесконечно малого поворота исходной прямой ℓ вокруг точки O точка P' будет мало отличаться от бесконечно удаленной точки, через которую проходит исходная гипербола Γ, а также ее асимптота, содержащая центр окружности Эйлера треугольника ABC и точку касания этой окружности с прямой Симсона точки P. Из приведенных рассуждений вытекает

Теорема 14. Кубика Мак-Кэя произвольного треугольника имеет три асимптоты, которые пересекаются в центре окружности Эйлера этого треугольника и проходят через три точки касания окружности Эйлера с прямыми Симсона.

Рассмотрим теперь равностороннюю гиперболу Γ и из произвольной точки I этой гиперболы как из центра проведем окружность, проходящую через центр O гиперболы Γ (рис. 10). Пусть D — точка гиперболы, диаметрально противоположная I, а окружность с центром I и радиусом ID пересекает ту ветвь гиперболы, на которой лежит ее центр I, в точках A и B, C — вторая точка пересечения этой окружности с другой ветвью гиперболы Γ.

Покажем, что треугольник ABC правильный. Опустим из точки D на прямую AB перпендикуляр, который пересечет гиперболу в точке H. Тогда H — ортоцентр треугольника ADB и точка H', симметричная H относительно центра O гиперболы, лежит на этой гиперболе. С другой стороны, поскольку центр O гиперболы Γ лежит на окружности Эйлера треугольника ADB, вписанного в эту гиперболу (см. лемму 1 в [3]), а описанная окружность этого треугольника гомотетична окружности Эйлера с центром H и коэффициентом 2, то точка H' совпадает с точкой пересечения описанной окружности треугольника ADB с гиперболой Γ, т.е. с точкой C. Но тогда диагонали четырехугольника $ICDH$ делятся точкой пересечения O пополам и, таким образом,

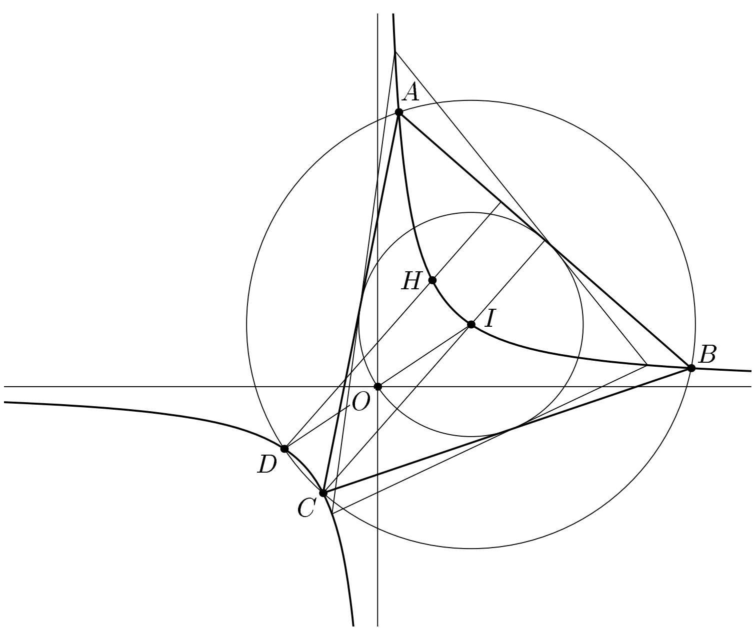


Рис. 10.

$ICDH$ — параллелограмм. Так как $CI \parallel DH$, а $DH \perp AB$, то и $CI \perp AB$. Поскольку I лежит на гиперболе Γ, отсюда вытекает, что I — ортоцентр треугольника ABC. Вспомнив, что I — центр описанной окружности треугольника ABC, получим, что в этом треугольнике ортоцентр совпадает с центром описанной окружности, и, следовательно, треугольник ABC — равносторонний.

Итак, существует треугольник, вписанный в гиперболу Γ и описанный около окружности I, поэтому по теореме Понселе (см. [4, с. 166]) таких треугольников существует бесконечно много: за вершину A можно взять любую точку гиперболы, лежащую вне окружности I и провести через A прямые до пересечения с гиперболой в точках B и C. Тогда прямая BC будет касаться окружности I (рис. 10).

Теорема 15. Пусть I — окружность с центром на равносторонней гиперболе Γ, проходящая через центр F этой гиперболы. Тогда асимптоты кубик Мак-Кэя всех треугольников, вписанных в гиперболу Γ и описанных около окружности I, соответственно параллельны, а их точки пересечения лежат на прямой FI.

Доказательство. Рассмотрим треугольник ABC, вписанный в гиперболу Γ и описанный около окружности I. Поскольку окружность Эйлера треугольника ABC проходит через центр F гиперболы Γ одновременно со вписанной окружностью I этого треугольника, то по теореме Фейербаха эти окружности касаются в точке F и, таким образом, точка F совпадает с одной из точек Фейербаха треугольника ABC. Отсюда следует, что центр E окружности Эйлера треугольника ABC лежит на прямой FI, но по теореме 14 центр окружности Эйлера треугольника является точкой пересечения асимптот кубики Мак-Кэя этого треугольника.

Для любого треугольника ABC, вписанного в гиперболу Γ и описанного около окружности I, гипербола Γ является изогональным образом прямой OI, где O — центр описанной окружности треугольника ABC. Так как прямая OI касается гиперболы Γ в точке I, то это означает, что центры описанных окружностей треугольников ABC лежат на касательной к гиперболе Γ в точке I, а асимптоты этой гиперболы совпадают с прямыми Симсона точек пересечения этой касательной с описанными окружностями треугольников ABC.

Применим к треугольнику ABC гомотетию с центром F, переводящую окружность Эйлера этого треугольника в окружность I. Тогда окружность Эйлера треугольника $A_1B_1C_1$, полученного в результате этой гомотетии, совпадет с окружностью I, являющейся окружностью Эйлера правильного треугольника, вписанного в гиперболу Γ и описанного около окружности I. Так как, кроме того, асимптоты гиперболы будут прямыми Симсона обоих треугольников, то согласно предложению 5 совпадут и остальные соответствующие прямые Симсона этих треугольников. Отсюда следует, что совпадут также точки касания прямых Симсона этих треугольников с их общей окружностью Эйлера и по теореме 14 совпадут асимптоты их кубик Мак-Кэя. Поскольку треугольники ABC получаются из треугольников $A_1B_1C_1$ гомотетией с центром F, то асимптоты их кубик Мак-Кэя соответственно параллельны. □

Кривой Штейнера называется трехрогая гипоциклоида, т.е. кривая, описываемая точкой M окружности радиуса r, которая катится по окружности радиуса $3r$ без скольжения, все время касаясь ее внутренним образом.

Теорема 16. Пусть S и S_1 — концентрические окружности радиусов $3r$ и r соответственно, K — кривая Штейнера с вершинами на окружности S, касающаяся окружности S_1 в точке M. Тогда касательная к кривой Штейнера пересекает окружность S_1 в двух точках N и P таких, что $\smile PM = 2\smile NM$.

Доказательство. Обозначим через S_2 окружность радиуса r, касающуюся окружности S_1 внешним образом в точке M и окружности S внутренним образом в точке M_1 (рис. 11). Кривая Штейнера, указанная в условии, совпадает с траекторией точки M окружности S_2 при качении

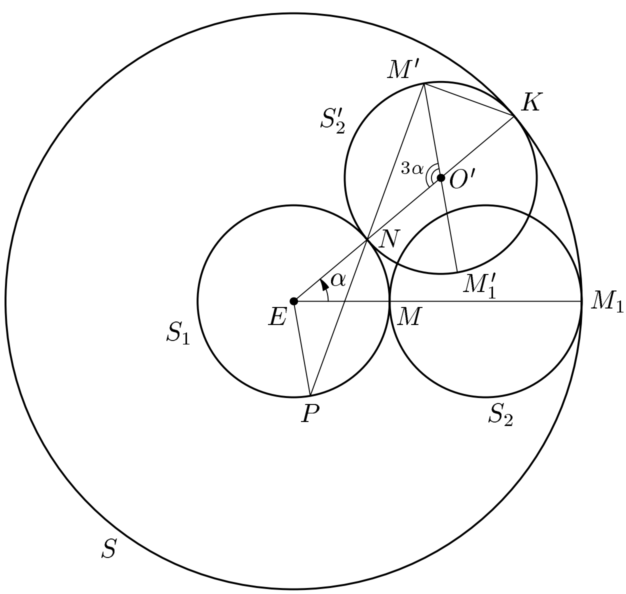


Рис. 11.

внутренним образом окружности S_2 по окружности S без скольжения. Предположим, что окружность S_2 после поворота приняла положение S_2' и стала касаться окружностей S_1 и S в точках N и K соответственно, при этом $\angle KEM_1 = \alpha$, где E — центр окружностей S и S_1, а точки M и M_1 перешли в M' и M_1'.

Так как K — центр мгновенного вращения окружности S_2', то вектор скорости точки M', направленный по касательной к ее траектории, совпадающей с кривой Штейнера, перпендикулярен прямой $M'K$. Но $\angle NM'K = 90^\circ$, как вписанный угол окружности S_2', опирающийся на ее диаметр NK, поэтому прямая $M'N$ является касательной к кривой Штейнера в точке M'. Пусть P — точка пересечения касательной $M'N$ с окружностью S_1, а O' — центр окружности S_2'. Из-за отсутствия скольжения дуга KM_1 окружности S равна дуге KM_1' окружности S_2': $\smile KM_1 = \smile KM_1'$ или $\alpha \cdot 3r = \beta \cdot r$, откуда $\beta = \angle KO'M_1' = 3\alpha = \angle M'O'N$.

Далее, из равенства равнобедренных треугольников PEN и $M'O_1N$ выводим, что $\angle PEN = \angle M'O'N = 3\alpha$, откуда получаем: $\angle PEM = \angle PEN - \angle MEN = 3\alpha - \alpha = 2\alpha$ и, таким образом, $\smile PM = 2\smile NM$. $\square$

Из теоремы 12, предложения 6 и теоремы 16 вытекает

Теорема 17. Огибающей прямых Симсона произвольного треугольника является кривая Штейнера, касающаяся его окружности Эйлера в трех точках.

На рис. 12 изображены треугольник, его описанная окружность, окружность Эйлера и кривая Штейнера этого треугольника.

В свою очередь из теорем 14 и 17 выводится

Следствие 7. Асимптоты кубики Мак-Кэя произвольного треугольника совпадают с осями кривой Штейнера этого треугольника.

Обозначим, как всегда, через H ортоцентр треугольника ABC. Тогда окружность Эйлера треугольника ABC проходит через середины сторон треугольников ABC, ABH, BCH, CAH и поэтому эти треугольники имеют общую окружность Эйлера.

Отметим без доказательства, что треугольники ABC, ABH, BCH, CAH имеют также общую кривую Штейнера, а их кубики Мак-Кэя — соответственно общие асимптоты.

Приведем в заключение еще один интересный факт. Как известно, прямые, которые делят угол на три равные части, называются трисектрисами этого угла. Теорема Морлея утверждает, что трисектрисы углов треугольника, прилежащие к его сторонам, пересекаются в трех точках, являющихся вершинами равностороннего треугольника (более подробно о теореме Морлея см. [2] и [1]). Назовем этот треугольник треугольником

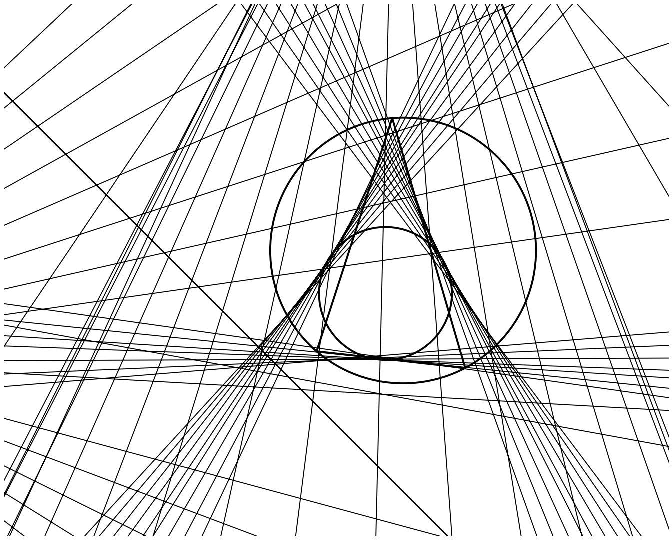


Рис. 12.

Морлея. Оказывается, что для любого треугольника стороны его треугольника Морлея и равностороннего треугольника, вершины которого совпадают с вершинами кривой Штейнера исходного треугольника, соответственно параллельны.

Список литературы

[1] Конн А. *Новое доказательство теоремы Морли* // Математическое просвещение. Третья серия. Вып. 9. 2005.

[2] Куланин Е. Д. *Вокруг теоремы Морлея* // Приложение «Математика» к газете «Первое сентября», №24–25, 1995.

[3] Куланин Е. Д. *Об описанных окружностях чевианных и педальных треугольников и некоторых кривых, связанных с треугольником* // Математическое просвещение. Третья серия. Вып. 9. 2005.

[4] Прасолов В. В., Тихомиров В. М. *Геометрия*. М.: МЦНМО, 1997.

Е. Д. Куланин, Московский городской психолого-педагогический университет
e-mail: `lucas03@mail.ru`

Об одном свойстве интегрируемой функции

М. Аппельбаум В. Журавлёв П. Самовол

В этой заметке приводится решение задачи 9.10 из задачника «Математического просвещения».

На математических соревнованиях самого разного уровня встречаются задачи с несколько парадоксальной формулировкой. Например, в книге [1, с. 105] находим:

Задача 1. Один человек каждый месяц записывал свой доход и расход. Может ли быть так, что за любые пять идущих подряд месяцев его общий доход превышал доход, а в целом за год его доход превысил расход?

(В формулировке А. А. Егорова задача звучит более интригующе: «как долго может законно существовать бизнес-сообщество, если по итогам работы за любые 5 месяцев у сообщества есть прибыль, но за год в налоговую инспекцию сообщество подаёт отчёт о том, что прибыли нет и налоги платить не из чего...»)

Второй пример предлагался на Международной Математической Олимпиаде (IMO) 1977 года в Белграде (6 очков), см. [2, с. 5, №19.2]:

Задача 2. В конечной последовательности действительных чисел сумма любых семи идущих подряд членов отрицательна, а сумма любых одиннадцати идущих подряд членов положительна. Найти набольшее число членов данной последовательности.

В Шестнадцатом Международном Турнире городов (1994–1995, осенний тур) находим задачу:

Задача 3. (А. Канель-Белов) Периоды двух последовательностей — 7 и 13. Какова максимальная длина начального куска, который может у них совпадать? (Период последовательности $\{a_n\}$ — это наименьшее натуральное число p такое, что для любого номера n выполняется равенство $a_n = a_{n+p}$).

Математическую идею всех данных задач можно сформулировать так.

Для последовательности действительных чисел, записанных в строчку, выполняются следующие условия $(*)$:

1. сумма любых m идущих подряд членов отрицательна;

2. сумма любых n идущих подряд членов положительна.

Чему равно $N_{\max}$ — максимальное число членов данной последовательности?

В этой статье мы рассмотрим ещё одну версию данной проблемы.

Задача 4. Существует ли такая непрерывная функция $y = f(x)$, что любой определённый интеграл от этой функции по любому отрезку длины $m = 3$ отрицателен, а по любому отрезку длины $n = 5$ — положителен?

Такие функции существуют. Аналитическое выражение одной из них и примерная схема ее графика приведены ниже на рисунке 1. График симметричен относительно прямой $x = 3$. Каждая из частей графика также симметрична: первая относительно прямой $x = 1{,}5$, вторая (правая) часть симметрична относительно прямой $x = 4,5$. Кроме того, для $x \in [0; 3]$ выполняется равенство $f(x) = f(x+3)$ (периодичность).

Поэтому

$$\int\limits_a^{a+3} f(x)\,dx = \int\limits_a^3 f(x)\,dx + \int\limits_3^{a+3} f(x)\,dx = \int\limits_a^3 f(x)\,dx + \int\limits_0^a f(x)\,dx = \int\limits_0^3 f(x)\,dx.$$

(Здесь и далее мы предполагаем, что равенства выполняются для тех значений параметров, при которых отрезки интегрирования попадают в область определения функции.)

$$f(x) = \begin{cases} -6, \text{ при } 0 \leqslant x < 1, \\ 180x - 186, \text{ при } 1 \leqslant x < 1{,}1, \\ 12, \text{ при } 1{,}1 \leqslant x < 1{,}9, \\ -180x + 354, \text{ при } 1{,}9 \leqslant x < 2, \\ -6, \text{ при } 2 \leqslant x < 3, \\ -6, \text{ при } 3 \leqslant x < 4, \\ 180x - 726, \text{ при } 4 \leqslant x < 4{,}1, \\ 12, \text{ при } 4{,}1 \leqslant x < 4{,}9, \\ -180x + 894, \text{ при } 4,9 \leqslant x < 5, \\ -6, \text{ при } 5 \leqslant x \leqslant 6. \end{cases}$$

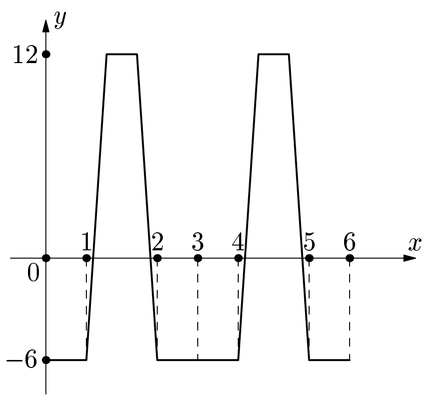


Рис. 1.

С другой стороны, функция $f(x)$ принимает одно и то же значение -6 на отрезках $[0;1]$ и $[5;6]$. Поэтому

$$\int\limits_{a}^{a+5} f(x)\,dx = \int\limits_{0}^{5} f(x)\,dx \qquad (0 \leqslant a \leqslant 1).$$

Интеграл от линейной функции равен произведению длины отрезка интегрирования на полусумму значений функции в концах отрезка интегрирования. Поэтому

$$\int\limits_{0}^{3} f(x)\,dx = 1\cdot(-6) + 0{,}1\cdot\frac{-6+12}{2} + 0{,}8\cdot 12 + 0{,}1\cdot\frac{-6+12}{2} + 1\cdot(-6) = \\ = -12 + 9{,}6 + 2\cdot 0{,}3 = -1{,}8 < 0,$$

а

$$\int\limits_{0}^{5} f(x)\,dx = -1{,}8 + 1\cdot(-6) + 2\cdot 0{,}3 + 9{,}6 = 2{,}4 > 0.$$

Таким образом, для данной функции любой определенный интеграл по отрезку длины 3 отрицательный, а по отрезку длины 5 — положительный.

Докажем следующую теорему.

Теорема 1. *Интегрируемая на отрезке $[0,c]$ функция $f(x)$ такова, что любой интеграл по отрезку длины n положителен, а по отрезку длины m отрицателен, $c > m > n > 0$. Тогда*

a) $c < m + n$;

b) если m и n соизмеримы, т. е. $m/n = q/p$, где p, q — взаимно простые целые числа, то $c < m + n - m/q$.

Доказательство a): предположим, что $c \geqslant m+n$. Обозначим через $F(x)$ первообразную функции $f(x)$, а через $\Phi(x)$ — первообразную функции $F(x)$. Используя формулу Ньютона – Лейбница и учитывая, что интеграл от положительной функции положителен, от отрицательной отрицателен, получаем противоречие:

$$0 < \int\limits_{0}^{m}\left[\int\limits_{x}^{x+n} f(y)\,dy\right]dx = \int\limits_{0}^{m}(F(x+n) - F(x))\,dx = \\ = \Phi(m+n) - \Phi(m) - \Phi(n) + \Phi(0) = \\ \int\limits_{0}^{n}(F(y+m) - F(y))\,dy = \int\limits_{0}^{n}\left[\int\limits_{y}^{y+m} f(x)\,dx\right]dy < 0.$$

Итак, $c < m + n$.

Доказательство b). Полагаем $d = m/q = n/p$, так что $m = qd$, $n = pd$. Обозначим

$$S_{d,k} = \int\limits_{(k-1)d}^{kd} f(x)\,dx, \qquad k \in \mathbb{N}, d > 0, kd \leqslant c.$$

Рассмотрим последовательность

$$S_{d,1}, S_{d,2}, S_{d,3}, \ldots, S_{d,p+q-1}. \tag{2}$$

Согласно условию теоремы эта последовательность удовлетворяет следующим свойствам:

1. сумма любых подряд идущих p ее членов положительна;

2. сумма любых подряд идущих q ее членов отрицательна.

Но, как следует из приводимой ниже леммы, последовательностей с такими свойствами не существует (максимальная длина такой последовательности равна $p + q - (p, q) - 1$). Приходим к противоречию.

Теперь сформулируем и докажем упомянутую лемму.

Лемма 1. *Будем говорить, что последовательность действительных чисел удовлетворяет свойству $*(n, m)$, если сумма любых n идущих подряд членов последовательности положительна, а сумма любых m идущих подряд членов последовательности отрицательна.*

Если последовательность a_i, $1 \leqslant i \leqslant N$ удовлетворяет свойству $*(n, m)$, то $N \leqslant m + n - d - 1$, где $d = (m, n)$ — наибольший общий делитель чисел n и m.

Доказательство. Без ограничения общности считаем, что $m > n$. Очевидно, что при выполнении условия леммы числа m и n не могут быть кратны друг другу. Поэтому $n > d = (m, n)$. Обозначим $n = n_1 d$, $m = m_1 d$, $(m_1, n_1) = 1$.

Будем доказывать лемму от противного. Предположим, что в последовательности $n + m - d = d(m_1 + n_1 - 1)$ членов. Разобъем последовательность на $m_1 + n_1 - 1$ групп по d чисел в каждой группе. Согласно условию леммы получаем, что сумма чисел в любых m_1 группах положительна, а сумма чисел в любых n_1 группах — отрицательна. По сути дела мы свели доказательство к частному случаю леммы, когда n, m взаимно просты. Поэтому в дальнейшем рассуждении мы считаем, что $(n, m) = 1$ и говорим о членах последовательности длины $n + m - 1$.

Рассмотрим любые $m - n$ подряд идущих членов последовательности. Кроме них последовательность содержит $(m + n - 1) - (m - n) = 2n - 1$ членов. Так как число $2n - 1$ нечетное, то при любом разбиении этого числа на два слагаемых одно из них будет не меньше n. Так что слева или справа от выбранной группы из $m-n$ членов находится еще не менее n членов.

Сумма любых n подряд идущих членов последовательности положительна. Если добавить n членов к выбранной группе из $m-n$ членов последовательности, то сумма полученного набора из m подряд идущих членов последовательности будет отрицательна. Поэтому сумма выбранных $m-n$ членов должна также быть отрицательной.

Итак, для взаимно простых m, n мы показали, что если последовательность удовлетворяет свойству $*(n,m)$, то она также удовлетворяет свойству $*(n,m-n)$. Далее действуем аналогично, заменяя бо́льшее из чисел m, n на их разность. Получаем такую же последовательность пар чисел, как в алгоритме Евклида. Поскольку исходные числа n, m — взаимно простые, приходим к тому, что последовательность удовлетворяет свойству $*(q,1)$, что невозможно (если все члены последовательности отрицательны, то любая их сумма также будет отрицательной). $\square$

Замечание 1. Таким образом, мы видим, что описанная в начале статьи математическая идея может принимать самые разные формы. Вместе с тем переход в «интегральную тематику» позволяет рассмотреть случай несоизмеримых пределов интегрирования.

Оценки в теореме 1 нельзя улучшить. Будем говорить, что функция $f(x)$, определенная на отрезке $[0;c]$, удовлетворяет интегральному свойству $*(n,m)$, если интеграл от $f(x)$ по любому отрезку длины n положителен, а по любому отрезку длины m отрицателен.

Теорема 2. 1) *Если m и n соизмеримы, т. е. $m/n = q/p$, где p, q — взаимно простые целые числа, то для любого $\varepsilon > 0$, $m+n-m/q > \varepsilon$, можно построить интегрируемую функцию $f(x)$ на отрезке $[0,m+n-m/q-\varepsilon]$, удовлетворяющую интегральному свойству $*(m,n)$.*

2) *Если m и n несоизмеримы, то для любого $\varepsilon > 0$, $m+n > \varepsilon$, можно построить интегрируемую функцию $f(x)$ на отрезке $[0,m+n-\varepsilon]$, удовлетворяющую интегральному свойству $*(m,n)$.*

Для доказательства теоремы мы будем строить функции, удовлетворяющие усиленному интегральному свойству $*(m,n)$. А именно, *функцией типа* (L,m,S_m,n,S_n) назовем функцию, определенную на отрезке $[0;L]$, и такую, что интеграл от этой функции по любому отрезку длины m равен S_m, а по любому отрезку длины n равен S_n.

Замечание 2. Если построена функция типа (L,m,S_m,n,S_n) такая, что L, m, n — целые, $S_m < 0$, $S_n > 0$, то интегралы от этой функции по отрезкам $[k;k+1]$ предоставляют пример «парадоксальной» последовательности (см. [4]). В частности, из теоремы 2 следует, что оценка в лемме 1 точна. Приводимое ниже доказательство теоремы 2 является прямым обобщением алгоритма построения «парадоксальных» последовательностей из [4].

Лемма 2. *Пусть $L < m + n$ и для любых A, B существует функция типа (L, m, A, n, B), $m > n$. Тогда для любых A, B существуют функции типов $(L + m, m, A, n + m, B)$ и $(L + n, m, A, n + m, B)$.*

Доказательство. Пусть f имеет тип (L, m, A, n, B). Тогда функция

$$g(x) = \begin{cases} f(x), & \text{если } 0 \leqslant x < m, \\ f(x-m), & \text{если } m \leqslant x < L + m, \end{cases}$$

имеет тип $(L + m, m, A, n + m, A + B)$. Действительно, если $b - a = m$, $a < m$, то

$$\int\limits_a^b g(x)\,dx = \int\limits_a^m g(x)\,dx + \int\limits_m^b g(x)\,dx = \int\limits_a^m f(x)\,dx + \int\limits_0^{b-m} f(x)\,dx = A,$$

а если $b - a = n + m$, то из $L < m + n$ следует, что $a < m$, а значит

$$\int\limits_a^b g(x)\,dx = \int\limits_a^m f(x)\,dx + \int\limits_0^{b-m} f(x)\,dx = \int\limits_a^m f(x)\,dx + \int\limits_0^{a+n} f(x)\,dx =$$

$$= \int\limits_0^a f(x)\,dx + \int\limits_a^m f(x)\,dx + \int\limits_a^{a+n} f(x)\,dx = A + B.$$

Докажем, что функция

$$g(x) = \begin{cases} f(x-n+m), & \text{если } 0 \leqslant x < n, \\ f(x-n), & \text{если } n \leqslant x < L + n, \end{cases}$$

имеет тип $(L + n, m, A, n + m, A + B)$. Если $b - a = m$, $a < n$, то

$$\int\limits_a^b g(x)\,dx = \int\limits_a^n g(x)\,dx + \int\limits_n^b g(x)\,dx = \int\limits_{a+m-n}^m f(x)\,dx + \int\limits_0^{b-n} f(x)\,dx = A,$$

а если $b - a = n + m$, то из $L < m + n$ следует, что $a < n$, а значит

$$\int\limits_a^b g(x)\,dx = \int\limits_{a+m-n}^m f(x)\,dx + \int\limits_0^{b-n} f(x)\,dx = \int\limits_{a+m-n}^m f(x)\,dx + \int\limits_0^{a+m} f(x)\,dx =$$

$$= \int\limits_{a+m-n}^{a+m} f(x)\,dx + \int\limits_{a+m}^m f(x)\,dx + \int\limits_0^{a+m} f(x)\,dx = B + A.$$

Теперь лемма вытекает из того факта, что отображение $(A, B) \mapsto (A, A + B)$ взаимно однозначно. $\square$

Доказательство теоремы 2. Для $n < m < L = 2n - \delta$ зададим функцию f типа (L, m, A, n, B) явной формулой:

$$f(x) = \begin{cases} b, & \text{если } 0 \leqslant x < n - \delta \text{ или } n \leqslant x \leqslant L, \\ a, & \text{если } n - \delta \leqslant x < n. \end{cases} \tag{3}$$

Интеграл от f по любому отрезку длины n равен $B = \delta a + (n - \delta)b$, а по любому отрезку длины m равен $A = \delta a + (m - \delta)b$. Получающаяся система линейных уравнений невырождена и поэтому имеет единственное решение при любых A, B.

Преобразования, описанные в лемме 2, одинаково меняют сумму $m+n$ и длину L. Для функции из (3) разность $m + n - L$ равна $m - n + \delta$.

Поэтому, многократно применяя лемму 2 и используя функцию из (3), получаем следующее утверждение: если пара (m, n) переходит в пару (m_ℓ, n_ℓ), $n_\ell < m_\ell < 2n_\ell - \varepsilon$, после нескольких преобразований вида

$$(m, n) \mapsto (m - n, n) \text{ или } (m, n) \mapsto (m, m - n), \tag{4}$$

то существуют функции типа $(n+m-(m_\ell-n_\ell)-\varepsilon, n, A, m, B)$ для любых A, B и $0 < \varepsilon < n + m - (m_\ell - n_\ell)$.

Применим это рассуждение для доказательства пункта 1) теоремы. Положим $d = m/q = n/p$, очевидно, что $n = pd$ и $m = qd$. Выберем натуральное число N, такое что $0 < d/N < \varepsilon$. Рассмотрим натуральные числа $m_1 = Nq$ и $n_1 = Np$. Очевидно, что $N = (m_1, n_1)$. Построим функцию $y = f_1(x)$ типа $(m_1 + n_1 - N - 1, m_1, A, n_1, B)$. В качестве искомой достаточно рассмотреть функцию $y = f(x) = f_1(dx/N)$, определенную на отрезке

$$\left[0; \frac{m_1 + n_1 - N - 1}{N} d\right] = \left[0; m + n - \frac{m}{q} - \frac{d}{N}\right] \supset \left[0; m + n - \frac{m}{q} - \varepsilon\right].$$

Аналогично доказывается пункт 2) теоремы. В этом случае в силу несоизмеримости m, n последовательность преобразований (4) продолжается неограниченно:

$$(m, n) \mapsto (m_1, n_1) \mapsto (m_2, n_2) \mapsto \ldots \mapsto (m_\ell, n_\ell) \mapsto \ldots,$$

причем $m_\ell + n_\ell \to 0$ при $\ell \to \infty$ и сколь угодно часто выполняется неравенство $n_\ell < m_\ell < 2n_\ell$. Выбрав такое ℓ, что $m_\ell+n_\ell < \varepsilon/2$ и $n_\ell < m_\ell < 2n_\ell$, заключаем, что так что существуют функции типа $(n + m - \varepsilon, n, A, m, B)$ для любых A, B и $0 < \varepsilon < n + m$. $\square$

Замечание 3. Построенные в доказательстве ступенчатые функции можно без труда превратить в непрерывные, удовлетворяющие аналогичным условиям.

В заключение предлагаем обобщение на многомерный случай как задачу для самостоятельного исследования.

Задача. Для каких областей Φ_1, Φ_2, Φ_3 существует непрерывная функция f с областью определения Φ_3 и такая, что интеграл от f по любому параллельному переносу Φ_1, лежащему в Φ_3, положителен, а интеграл от f по любому параллельному переносу Φ_2, лежащему в Φ_3, отрицателен. Рассмотрите случай, когда Φ_1 — прямоугольник, а Φ_2 — круг.

Авторы выражают благодарность профессору Канель-Белову (Московский Институт Открытого Образования) за участие в обсуждении этой проблемы.

Список литературы

[1] Васильев Н. Б., Гутенмахер В. Л., Раббот Ж. М., Тоом А. Л. *Заочные математические олимпиады.* М.: Наука, 1997. С. 105, 108–109.

[2] *Международные математические олимпиады.* Сост. А. А. Фомин, Г. М. Кузнецова. М.: Дрофа, 2000. С. 5, 34–35.

[3] Произволов В., Спивак А. *Усреднение по окружности* // Квант, 1998. №1, с. 29–31.

[4] Самовол П., Аппельбаум М., Жуков А. *Как построить парадоксальный пример* // Квант, 2005. №1, с. 35–37.

[5] Международный Турнир городов, 1994–1995 (Осенний тур). `http://www.turgor.ru/16/turnir16.php#turnir16otm`

Dr. Peter Samovol, Ben-Gurion University of Negev, Beer-Sheva, Israel Kaye Academic College of Education, Beer-Sheva, Israel
E-mail: `Pet12@012.net.il`

Dr. Mark Applebaum, Kaye Academic College of Education, Beer- Sheva, Israel
E-mail: `Amark@012.net.il`

Журавлёв Валерий Михайлович (к.ф.-м.н.), ОАО «Сибирско-Уральская нефтегазохимическая компания», г. Москва, Россия
E-mail: `Zhuravlev@sibur.ru`

Комментарий к статье М. Аппельбаума, В. Журавлёва и П. Самовола

А. Я. Канель

Задача интересна и для случая нестрогих неравенств. В этом случае картина такова. Рассмотрим непрерывную функцию $f(x)$, определённую на отрезке длины γ, интеграл от которой по любому отрезку длины α неположителен, а по любому отрезку длины β — неотрицателен.

Если $\alpha/\beta \notin \mathbb{Q}$, и $\gamma \geqslant \alpha+\beta$, то $f(x)\equiv 0$. (Это следует из приведенного в статье доказательства.) (А если $\gamma < \alpha+\beta$, то содержательные примеры существуют.)

Пусть теперь величины α и β соизмеримы. Тогда для некоторого $\delta > 0$, $\alpha = n\delta$, $\beta = n\delta$, $(m,n) = 1$. Из результатов статьи следует, что если $\gamma \geqslant \alpha+\beta$, то интеграл от f по любому отрезку длины δ равен нулю, а функция f периодична на своей области определения с периодом δ.

Случай $\gamma \leqslant \alpha+\beta-\delta$ покрывается результатами статьи, пусть $\gamma = \alpha+\beta-\delta+\varepsilon$, $\varepsilon < \delta$.

Тогда можно утверждать только, что интеграл от f по любому сдвигу начального отрезка длины δ на величину $k\delta+\tau$, где $\tau \leqslant \varepsilon$, равен нулю. При этом можно построить примеры функции с ненулевым интегралом для каждого такого сдвига.

Нам пишут

Более короткие решения задач из задачника «Математического просвещения»

И. И. Богданов

Задача 1.6б). (Автор решения — А. Бадзян.)

Предположим противное. Пусть z_0 — корень $P(x)$ аргумента α. Тогда $\operatorname{Arg} z_0^n \in (0, \pi)$ при любом $0 < n < \pi/\alpha$, т. е. $\operatorname{Im} z_0^n > 0$. Так как $Q(x)$ — неконстантный многочлен степени, меньшей π/α с неотрицательными коэффициентами, то $\operatorname{Im} Q(z_0) > 0$, т. е. $Q(z_0) \neq 0$ и $Q(x)$ не делится на $P(x)$. Противоречие.

Задача 7.3. (Автор решения — И. Богданов.)

Пусть $P(x)$ — минимальный многочлен матрицы AA^T, тогда

$$A^T P(AA^T) A = Q(A^T A) = 0,$$

где $Q(x) = xP(x)$. Таким образом, если λ — ненулевое собственное значение $A^T A$, то $Q(\lambda) = 0$, а, следовательно, и $P(\lambda) = 0$, т. е. λ — собственное значение AA^T. Обратное доказывается аналогично.

* * * * * *

От редколлегии

Напомним условие задачи 1.6: а) Дан многочлен $P(X)$. Для любого $X > 0$: $P(X) > 0$. Доказать, что что $P = Q/T$, где Q и T — многочлены с неотрицательными коэффициентами. б)* пусть P — квадратный трехчлен, α — аргумент его комплексного корня. Тогда степень Q не меньше $2\pi/\alpha$.

Решение А. Бадзяна оказалось для нас неожиданностью: ранее опубликованное решение задачи 1.6.б) занимает свыше двух страниц.

Мы призываем читателей присылать свои решения: ясные, понятные, с красивыми идеями и, по возможности, краткие (плохие решения мы и сами напишем).

Если Вы решили не буквально ту задачу, которая была опубликована в «Задачнике», а близкую к ней, то всё равно присылайте решение. Это может быть интересно само по себе, а кроме того, в условиях задач встречались ошибки.

Задача о собственных числах

А. К. Ковальджи

Приведем формулировку и решение задачи 7.3 из задачника «Математического просвещения».

Задача (теорема). Пусть A — прямоугольная матрица, A' — транспонированная матрица, тогда матрицы AA' и $A'A$ имеют одинаковые наборы ненулевых собственных чисел с одинаковыми кратностями.

Доказательство. Пусть $AA'v = \lambda v$, где v — собственный вектор, λ — собственное число. Тогда $A'AA'v = \lambda A'v$, $(A'A)(A'v) = \lambda(A'v)$. Значит, каждое собственное число матрицы AA' является собственным числом матрицы $A'A$ и наоборот. Докажем, что совпадают кратности чисел, исходя из идеи непрерывного изменения.

Рассмотрим матрицу B такого же размера как A, максимальный минор которой — диагональная матрица с различными элементами, а остальные элементы — нули. Все ненулевые собственные числа матрицы BB' различны. Рассмотрим матрицу $C = (1-\alpha)B + \alpha A$, где α — параметр, $0 \leqslant \alpha \leqslant 1$.

При непрерывном изменении α от 0 до 1 собственные числа CC' непрерывно меняются от набора собственных чисел BB' до набора собственных чисел BB', причем, при $\alpha = 0$ все собственные числа различны, а при $\alpha = 1$ они «склеиваются» и получают кратности.

Поскольку между различными собственными числами CC' и $C'C$ существует взаимно однозначное соответствие, и они меняются непрерывно, то при каждой «склейке» значений собственных чисел будет сохраняться равенство их кратностей. Теорема доказана.

Замечание. Эта теорема важна для прикладных задач, например для метода главных компонент, поскольку матрица AA' может оказаться очень большого размера, и, чтобы напрямую вычислить ее собственные числа и векторы, не хватит ресурсов компьютера, а матрица $A'A$ — может оказаться маленького размера, тогда, вычислив ее собственные числа λ_i и векторы v_i, мы легко найдем собственные числа и векторы AA', равные λ_i и Av_i.

Исправления к статье А. Скопенкова «Вокруг критерия Куратовского планарности графов»

А. Скопенков

В упомянутую статью (№9, сс. 116–128) следует внести следующие исправления.

С. 119, строка 3. Перед словом «Предположим» следует добавить «Будем рассматривать графы с петлями и кратными ребрами.»

С. 119, строка 5. Вместо «с минимальным числом ребер» следует читать «с минимальным числом ребер, не содержащий изолированных вершин.»

С. 125, строка 7 снизу. Вместо

«Рассмотрим объединение $\bar{N}$ двумерных граней 2-полиэдра N. Поскольку каждый из графов K_5 и $K_{3,3}$ вложим и в тор, и в лист Мебиуса, то $\bar{N}$ есть несвязное объединение дисков. Заменим каждый диск на „колесо“.»

следует читать

«Так как N не содержит зонтика, то окрестность любой точки в N является объединением дисков и отрезков, склеенных за одну точку (рис. 1.a). Если этих дисков больше одного, то заменим эту окрестность на изображенную на рис. 1.b. Обратное преобразование является стягиванием „звезды с несколькими лучами“ и поэтому сохраняет планарность. Значит, достаточно доказать планарность для полученного указанной заменой 2-полиэдра. Рассмотрим объединение $\bar{N}$ его двумерных граней.

(a) (b)

Рис. 1.

Тогда окрестность любой точки в $\bar{N}$ является диском. Значит, по теореме классификации поверхностей $\bar{N}$ является несвязным объединением сфер с ручками, пленками Мёбиуса и дырками. Поскольку каждый из графов K_5 и $K_{3,3}$ вложим и в тор с дыркой, и в лист Мебиуса, то $\bar{N}$ есть несвязное объединение дисков с дырками. Заменим каждый из этих дисков с дырками на граф с рис. 2.

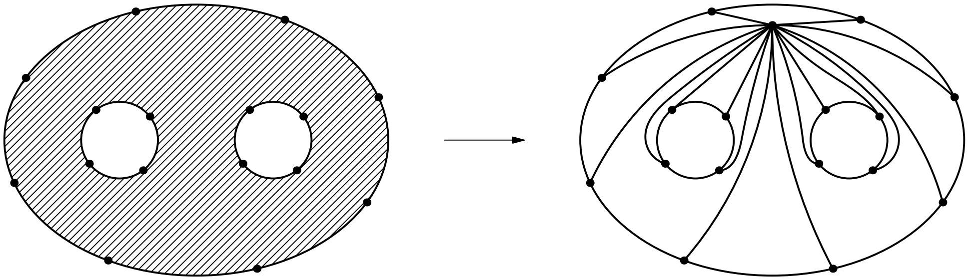


Рис. 2.

Автор приносит извинения за допущенную небрежность.

Задачный раздел

В этом разделе вниманию читателей предлагается подборка задач разной степени сложности, в основном трудных. Некоторые из этих задач (не обязательно самые сложные!) требуют знания «неэлементарной» математики — анализа, линейной алгебры и т. п.

Составителям этой подборки кажется, что предлагаемые ниже задачи окажутся интересными как для сильных школьников, интересующихся математикой, так и для студентов-математиков.

Мы обращаемся с просьбой ко всем читателям, имеющим свои собственные подборки таких задач, присылать их в редакцию. И, разумеется, мы с удовольствием будем публиковать свежие авторские задачи.

В скобках после условия задачи приводится фамилия автора (уточнения со стороны читателей приветствуются). Если автор задачи нам неизвестен, то в скобках указывается «фольклор».

1. В вершинах куба написаны числа. Вместо каждого числа записывают среднее арифметическое чисел, стоящих в трех соседних вершинах (числа заменяют одновременно). После 10 таких операций в вершинах оказались исходные числа. Найдите все такие числа. (*А. К. Ковальджи*)

2. Кривая $\mathcal{C}$ задана в $\mathbb{R}^4$ параметрически уравнениями $x_i = P_i(\tau)$, $i = 1, 2, 3, 4$, где многочлены $P_i(\tau)$ имеют степень 3. Докажите, что $\mathcal{C}$ живет в трехмерной плоскости (т. е. найдется 3-мерное аффинное подпространство, которое ее содержит). (*А. Я. Белов*)

3. Рассмотрим циклические слова с отмеченным началом из 0 и 1 с четным числом нулей. Припишем единицам слова знаки $+$ и $-$ так, чтобы знаки соседних единиц совпадали или отличались в зависимости от четности числа нулей между ними (например, идущие подряд единицы берутся с одинаковым знаком). *Сигнатурой слова* назовем абсолютную величину алгебраической суммы единиц. Докажите, что число слов длины n и сигнатуры $n-2k$ равно числу сочетаний из n по k при $n-2k>0$ и вдвое меньше при $n-2k=0$. (*Э. Б. Винберг*)

4. На плоскости проведены n систем равноотстоящих прямых; i-я система состоит из всех прямых вида $a_i x + b_i y = c_i + k$, $k \in \mathbb{Z}$. При

этом никакие три прямых не пересекаются в одной точке, и никакие две системы не параллельны. Эти системы разбивают плоскость на многоугольники. Пусть S — средняя площадь многоугольника, S_{ij} — площадь параллелограмма решетки, порожденной i-й и j-й системами. Докажите, что

$$S^{-1} = \sum_{i<j} S_{ij}^{-1}$$

(Средняя площадь многоугольника — это величина $S = \lim\limits_{t\to+\infty} S_t/N_t$, где S_t — общая площадь всех многоугольников разбиения, целиком содержащихся в круге радиуса t с центром в начале координат, N_t — количество этих многоугольников.) (*А. Я. Канель*)

5. На берегу круглого острова Гдетотам расположено n деревень, в каждой живут борцы. Был проведен турнир, в котором каждый борец встретился со всеми борцами из всех других деревень. Деревня А считается *сильнее* деревни Б, если хотя бы $(1-\alpha)$-я часть поединков между борцами из этих деревень заканчивается победой борца из деревни А. (У всех борцов разная сила, и в поединке всегда побеждает сильнейший.) Выяснилось, что каждая деревня сильнее следующей за ней по часовой стрелке. Докажите, что $\alpha \leqslant \alpha_0 = \dfrac{1}{4\cos^2\dfrac{\pi}{n+2}}$, причем число α_0 нельзя заменить на меньшее.
(*И. Богданов*)

6. Внутри выпуклого четырехугольника взята точка, равноудаленная от противоположных сторон. Оказалось, что она лежит на прямой, соединяющей середины диагоналей. Докажите, что четырехугольник является либо вписанным, либо описанным, либо трапецией. (*А. Заславский*)

7. От прямоугольника отрезают квадрат, а с оставшимся прямоугольником производят ту же процедуру. Может ли последовательность отношений сторон у этих прямоугольников быть периодической, если одна из сторон исходного прямоугольника равна 1, а другая равна а) $\sqrt{2}$, б) $\sqrt[3]{2}$, в) $\sqrt{2005}$. (*В. А. Уфнаровский, А. Я. Белов*)

8. Можно ли покрасить сферу белой и красной красками так, чтобы любые три исходящих из центра сферы взаимно перпендикулярных луча пересекали ее в одной красной и двух белых точках?
(*S. Kochen, M. Specker*)

9. G — группа порядка $2^n(2k+1)$, содержащая элемент порядка 2^n. Докажите, что множество элементов нечетного порядка является подгруппой. (*Заочный конкурс памяти Кирилла Дочева*)

10. Среди k монет есть одна фальшивая, причем неизвестно, легче она или тяжелее. За какое минимальное число взвешиваний можно определить фальшивую монету на чашечных весах без гирь, если при этом а) требуется узнать, легче она или тяжелее; б) не требуется узнать это. (*А. М. Яглом, И. М. Яглом*)

11. Последовательность непрерывных функций (F_n), где n-я функция зависит от n переменных, называется *средней*, если она удовлетворяет следующим свойствам:

    1) Для любого натурального n функция F_n симметрична, однородна (т. е. при перестановке переменных значение F_n не меняется и
$$F_n(\lambda \cdot x_1, \ldots, \lambda \cdot x_n) = \lambda \cdot F_n(x_1, \ldots, x_n)$$
для любых чисел $\lambda, x_1, \ldots, x_n$), кроме того $F_n(x, \ldots, x) = x$ для любого x, $F_2(1,0) = 1/2$.

    2) Для любых натуральных n и k равенство
$$F_{n+k}(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+k}) = F_{n+k}(x_1, \ldots, x_n, Y, \ldots, Y)$$
где $Y = F_k(x_{n+1}, \ldots, x_{n+k})$, выполнено для любых чисел $x_1, \ldots, x_{n+k}$.

    Докажите, что функция F_n есть среднее арифметическое:
$$F_n(x_1, \ldots, x_n) = \frac{x_1 + \cdots + x_n}{n}.$$
(*А. Я. Канель*)

12. Квадрат разбит на треугольники равной площади. Докажите, что их число четно. (*Фольклор*)

Решения задач из предыдущих выпусков

8.2. Условие. Найти дискриминант многочлена $P(x) = x^{2003} + x + 1$.

Решение. Рассмотрим более общую задачу: найти дискриминант многочлена $P_t(x) = x^{2003} + x + t$. Сначала заметим, что дискриминант равен результанту многочлена и его производной, и по известной формуле (см., например, Б. Л. ван дер Варден, «Алгебра») результант многочленов $a_0x^n + a_1x^{n-1} + \ldots + a_n$ и $b_0x^m + b_1x^{m-1} + \ldots + b_m$ равен определителю

$$\begin{vmatrix}
a_0 & a_1 & \ldots & a_n & 0 & \ldots\ldots\ldots\ldots & 0 & 0 \\
0 & a_0 & a_1 & \ldots & a_n & 0 & \ldots\ldots\ldots\ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots\ldots\ldots\ldots & & & & 0 & a_0 & a_1 \ \ldots & a_n \\
b_0 & b_1 & \ldots\ldots & b_m & 0 & \ldots\ldots & 0 & 0 \\
0 & b_0 & b_1 & \ldots\ldots & b_m & 0 & \ldots\ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots\ldots\ldots\ldots & & & 0 & b_0 & b_1 & \ldots\ldots & b_m
\end{vmatrix}.$$

Поскольку $P'(x) = 2003x^{2002}+1$, видно, что в нашем случае результант есть многочлен от t степени 2002 и старшим коэффициентом 2003^{2003}. Вычислим этот многочлен. Для этого посмотрим, при каких t у P и P' есть общий корень. Корни P' суть $x_k = 2003^{-1/2002} \exp\big(\frac{\pi i(2k+1)}{2002}\big)$, $k = 1, \ldots, 2002$. Поэтому искомые значения t суть $t_k = -x_k - x_k^{2003} = -\frac{2002}{2003}x_k$. Многочлен с корнями t_k есть $\prod_{k=1}^{2002}(t-t_k) = t^{2002} + \frac{2002^{2002}}{2003^{2003}}$. Поэтому искомый многочлен получается из него умножением на число (раз он той же степени). Мы знаем его старший коэффициент, поэтому осталось выписать ответ: дискриминант многочлена $P_t(x)$ равен $2003^{2003}t^{2002} + 2002^{2002}$, а дискриминант многочлена $P(x)$ получается при $t = 1$.

Ответ: $2003^{2003} + 2002^{2002}$. *(В. В. Доценко)*

8.3. Условие. Можно ли круг с двумя дырками отобразить в себя без неподвижных точек?

Решение. В условии задачи допущена ошибка. Вопрос состоит в существовании непрерывного взаимнооднозначного отображения (гомеоморфизма).

Круг с двумя дырками гомеоморфен сфере с тремя дырками. Можно считать, что эти дырки имеют небольшой размер (скажем, 1°) и расположены на экваторе сферы так, что расстояния между их центрами одинаковы и равны 120°.

Гомеоморфизм без неподвижных точек можно теперь описать так: это композиция поворота на 120° вокруг полярной оси и симметрии относительно плоскости экватора. Поскольку полушария меняются местами, то вне экватора неподвижных точек нет. С другой стороны, ограничение этого отображения на экватор является поворотом на 120°, у которого неподвижных точек тоже нет.

Замечание. Данная конструкция обобщается для круга с $n \geqslant 2$ дырками.

(*М. Л. Концевич*)

8.4. Условие. Доказать, что на описанной окружности каждого треугольника существует ровно три точки, для которых соответствующие прямые Симсона касаются окружности девяти точек треугольника, причем эти точки являются вершинами правильного треугольника.

Решение. Для решения нам понадобятся следующие два факта о прямых Симсона.

1°. Если H — точка пересечения высот треугольника ABC, P — точка на его описанной окружности, то прямая Симсона точки P проходит через середину отрезка PH (рис. 1).

2°. Если P_1 и P_2 — точки описанной окружности треугольника ABC, s_1 и s_2 — соответствующие им прямые Симсона, то ориентированный угол между прямыми s_1 и s_2 (измеренный от s_1 к s_2) равен половине ориентированного угла P_2OP_1 (O — центр описанной окружности треугольника ABC), рис. 2.

Используя эти утверждения (их доказательства можно найти, например, в книге В. В. Прасолова «Задачи по планиметрии», задачи 5.92 и 5.96), получить решение задачи очень легко.

А именно, поскольку окружность Эйлера треугольника получается из его описанной окружности гомотетией с центром H и коэффициентом $1/2$, то, согласно утверждению 1°, любая прямая Симсона имеет хотя бы одну общую точку с окружностью Эйлера (на рис. 3 эта точка — середина отрезка PH — обозначена через L). Чтобы прямая Симсона касалась окружности Эйлера, эта точка должна быть их единственной общей точкой, т. е. прямая Симсона должна быть перпендикулярна отрезку EL (E — центр окружности Эйлера). Рассмотрим теперь треугольник OHP; поскольку $HE = EO$, отрезок EL является его средней линией, тем самым, $EL \parallel OP$. Значит, прямая Симсона касается окружности Эйлера тогда и только тогда, когда она перпендикулярна отрезку OP.

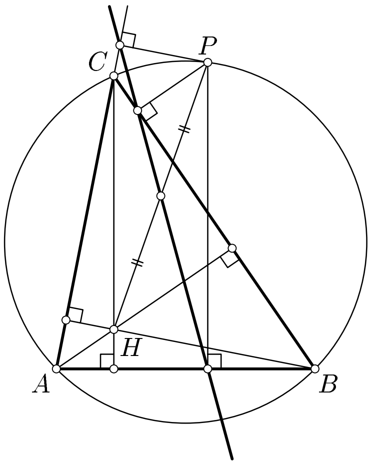


Рис. 1.

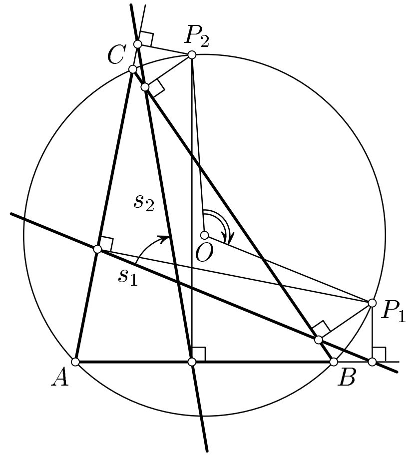


Рис. 2.

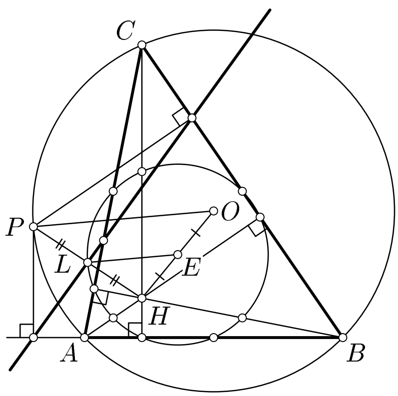


Рис. 3.

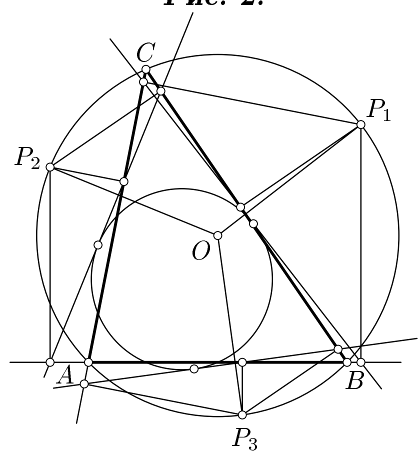


Рис. 4.

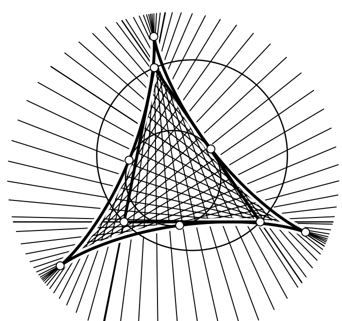


Рис. 5.

Используем утверждение 2°: при проходе точкой P всей описанной окружности (т. е. когда отрезок OP поворачивается вокруг O на 360°) прямая Симсона поворочивается на -180°. При этом условие перпендикулярности выполняется ровно трижды, причём между соответствующими положениями точками P (на рис. 4 они обозначены P_1, P_2, P_3) отрезок OP поворачивается на 120°. Так что $P_1P_2P_3$ — равносторонний треугольник.

Замечания. 1. Пусть вершинам треугольника ABC соответствуют комплексные числа a, b, c, а центру его описанной окружности — 0. Можно доказать, что тогда точкам P_1, P_2, P_3 соответствуют три значения $\sqrt[3]{-abc}$.

2. Утверждение задачи основано на том, что огибающей прямых Симсона данного треугольника является дельтоида (гипоциклоида с тремя остриями), описанная вокруг его окружности Эйлера, рис. 5.

(*М. Ю. Панов*)

9.4. Условие. Дана последовательность $\{a_k\}_{k=1}^{\infty}$, такая что $a_1 = 1$, $a_k = a_{k-1} + a_{[k/2]}$ при $k > 1$. Докажите, что ни один ее член не делится на 4.

Первое решение. Пусть $n = 2^k m$, где m — нечетно. Докажем индукцией по n, что

1. если k нечетно, то $a_n \equiv 2 \pmod 4$;
2. если k четно и количество цифр в двоичной записи числа n нечетно, то $a_n \equiv 1 \pmod 4$;
3. если k четно и количество цифр в двоичной записи числа n четно, то $a_n \equiv 3 \pmod 4$.

Последние два случая можно объединить в один: если k четно, то $a_n \equiv 2s(n) - 1 \pmod 4$, где $s(n)$ — количество цифр в двоичной записи числа n.

База $n = 1$ очевидна.Переход: пусть это верно для $1, 2, \ldots, n-1$, докажем для $n = 2^k m$.

1. Если k нечетно, то по предположению индукции $a_n \equiv 2s(n-1) - 1 + 2s(n/2) - 1 \pmod 4$. Учитывая, что $s(n/2) = s(n)$, $s(n-1) = s(n) + k - 1$, получаем

$$2s(n-1) - 1 + 2s(n/2) - 1 = 2s(n) - 1 + 2s(n) - 1 + 2k - 2 = 2k \equiv 2 \pmod 4.$$

2. Если k четно и $k > 0$, то предположению индукции $a_n \equiv 2s(n-1) - 1 + 2 \pmod 4$. Из $s(n-1) = s(n) + k - 1$ получаем

$$2s(n-1) - 1 + 2 = 2s(n) + 2k - 1 \equiv 2s(n) - 1 \pmod 4.$$

Если $k = 0$, то полагаем $n - 1 = 2^\ell s$, где s нечетно.

Если $\ell-1$ нечетно, то из $a_n \equiv 2+2s(n-1)-1 \pmod 4$ и $s(n-1)=s(n)-1$ получаем $2+2s(n-1)-1=2s(n)-1$.

Если $\ell-1$ четно, то из $a_n \equiv 2+2s([(n-1)/2])-1 \pmod 4$ и $s([(n-1)/2])=s(n)-1$ получаем $2+2s([(n-1)/2])-1=2s(n)-1$.

Индуктивный переход доказан. Значит, a_n не делится на 4.

(*А. Бадзян*)

Второе решение. Доказательство того, что a_n не делится на 4, следует из трех фактов.

1. Если n нечетное, то a_n нечетное.

Действительно, при $k>0$ имеем $a_{2k+1}=a_{2k}+a_k=a_{2k-1}+2a_k$, так что a_{2k+1} и a_{2k-1} имеют одинаковую четность. Поскольку a_1 нечетно, то и все остальные a_{2k+1} также нечетны.

2. При $k>0$ выполняется $a_{4k} \equiv a_k \pmod 4$.

Для $k=1$ это сравнение выполняется. Для остальных k его можно проверить по индукции:

$$
\begin{aligned}
a_{4k+4} &= a_{4k+3}+a_{2k+2} = a_{4k+2}+2a_{2k+1}+a_{k+1} = \\
&= a_{4k+1}+3a_{2k+1}+a_{k+1} = a_{4k}+3a_{2k+1}+a_{2k}+a_{k+1} = \\
&= a_{4k}+4a_{2k+1}+a_{k+1}-a_k,
\end{aligned}
$$

поэтому $a_{4k+4}-a_{k+1} \equiv a_{4k}-a_k \pmod 4$.

3. Если n четно, но не делится на 4, то для a_n справедливо то же самое: оно четно и не делится на 4.

Из п. 2 получаем:

$$
\begin{aligned}
a_{4k+2} &= a_{4k+1}+a_{2k+1} = a_{4k}+a_{2k}+a_{2k+1} = \\
&= (a_{4k}-a_k)+2a_{2k+1} \equiv 2a_{2k+1} \pmod 4,
\end{aligned}
$$

далее применим утверждение п. 1.

Теперь утверждение задачи можно доказать от противного: если n — наименьшее число, для которого a_n делится на 4, то n должно делиться на 4 (в силу пп. 1 и 3), но тогда в силу п. 2 $a_{n/4}$ также должно делиться на 4. Получили противоречие, которое означает, что все a_n не делятся на 4. (*А. Зелевинский*)

Новые издания

Книги издательства МЦНМО (2005 г.):

В. И. Арнольд. **Задачи семинара 2003–2004.** 56 с.

М. Бакингэм, К. Кофман. **Сначала надо нарушить все правила! Что лучшие в мире менеджеры делают по-другому?** Пер. с англ. Д. Д. Мухиной. 328 с.

А. Б. Богатырёв. **Экстремальные многочлены и римановы поверхности.** 176 с.

Д. Ван, Ч. Ли, Ш.-Н. Чоу. **Нормальные формы и бифуркации векторных полей на плоскости.** Пер. с английского под редакцией Ю. С. Ильяшенко. 416 с.

Дж. У. Вик. **Теория гомологий. Введение в алгебраическую топологию.** Пер. с англ. П. А. Колгушкина. 288 с.

Б. П. Гейдман, И. Э. Мишарина. **Методические рекомендации по работе с комплектом учебников «Математика. 3 класс».** 136 с.

«Глобус» Общематематический семинар. Под ред. М. А. Цфасмана и В. В. Прасолова. Выпуск 2. 216 с.

Н. В. Горбачев. **Сборник олимпиадных задач по математике.** 560 с.

В. А. Гордин. **Как это посчитать? Обработка метеорологической информации на компьютере. Идеи, методы, алгоритмы, задачи.** 280 с.

Ю. С. Ильяшенко. **Аттракторы и их фрактальная размерность.** 16 с.

А. Б. Каток, Б. Хасселблат. **Введение в теорию динамических систем с обзором последних достижений.** Пер. с англ. под ред. А. С. Городецкого. 464 с.

В. Г. Кац. **Вертексные алгебры для начинающих.** Пер. с англ. И. М. Парамоновой. 200 с.

В. Г. Кац, П. Чен. **Квантовый анализ.** Пер. с англ. Ф. Ю. Попеленского и Ж. Г. Тотровой. 128 с.

Математика и безопасность информационных технологий. Материалы конференции в МГУ 28-29 октября 2004 г. 384 с.

Т. Мива, М. Джимбо, Э. Датэ. **Солитоны: дифференциальные уравнения, симметрии и бесконечномерные алгебры.** Пер. с англ. С. З. Пакуляк. 112 с.

Н. М. Мишачев, Я. М. Элиашберг. **Введение в h-принцип.** Пер. с англ. Н. М. Мишачева. 232 с.

С. П. Новиков, И. А. Тайманов. **Современные геометрические структуры и поля.** 584 с.

А. М. Романов. **Занимательные вопросы по астрономии и не только.** 415 с.

А. Б. Сосинский. **Узлы. Хронология одной математической теории.** 112 с.

В. Г. Тураев. **Введение в комбинаторные кручения.** Пер. с англ. С. К. Ландо под ред. В. А. Васильева. 136 с.

Дж. Харрис. **Алгебраическая геометрия. Начальный курс.** Пер. с англ. под ред. Ф. Л. Зака. 400 с.

Г. С. Шаров, А. М. Шелехов, М. А. Шестакова. **Сборник задач по дифференциальной геометрии.** 112 с.

А. Шень. **Логарифм и экспонента.** 24 с.

А. Шень. **Простые и составные числа.** 16 с.

С. А. Шестаков. **Векторы на экзаменах.** 112 с.

XXVII Турнир им. М. В. Ломоносова 26 сентября 2004 года. Задания, решения, комментарии. Сост. А. К. Кулыгин. 192 с.

Серия «Библиотека „Математическое просвещение“»

Вып. 30. Ю. П. Соловьёв. **Неравенства.** 16 с.

Вып. 31. В. Ю. Протасов. **Максимумы и минимумы в геометрии.** 56 с.

Вып. 32. А. В. Хачатурян. **Геометрия Галилея.** 32 с.

* * * * * *

По вопросам приобретения этих книг обращаться по адресу: 119002, Москва, Большой Власьевский переулок, дом 11, магазин «Математическая книга».

Тел.: (495)-241-7285, факс: (495)-291-6501, e-mail: biblio@mccme.ru

Опечатки, замеченные в №9

Страница,	Строка	Напечатано	Следует читать
30,	1,2 снизу	следует вычеркнуть	
141,	формула (29)	+ (первые 3 вхождения)	·
157,	левый верхний рис.	неправильный	правильный см. на сайте
224,	5 сверху	А. В. Спивак	

Редактор В. В. Ященко
Подготовка оригинал-макета: $\LaTeX 2\varepsilon$,
METAPOST, М. Н. Вялый

Подписано в печать 13.02.2006 г. Формат $70 \times 100/16$.
Бумага офсетная №1. Печать офсетная. Печ. л. 18,0. Тираж 1000.

Издательство Московского Центра
непрерывного математического
образования
119002, Москва, Большой Власьевский пер., 11. Тел. (495) 241 74 83

Отпечатано с готовых диапозитивов в ГУП «Облиздат»
248640, г. Калуга, пл. Старый Торг, 5
Заказ №