Лицей "Вторая школа" имени В. Ф. Овчинникова

11 класс, 2024–2025 учебный год
Математический профиль

Бибиков Павел Витальевич

Оглавление

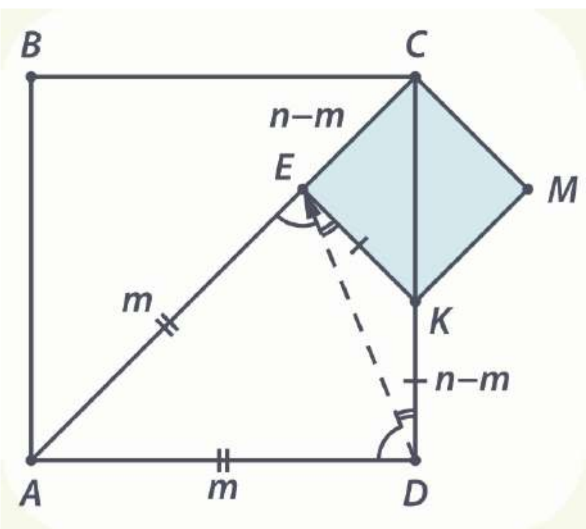
МАТЕРИАЛЫ МАТЕМАТИЧЕСКОГО ПРОФИЛЯ

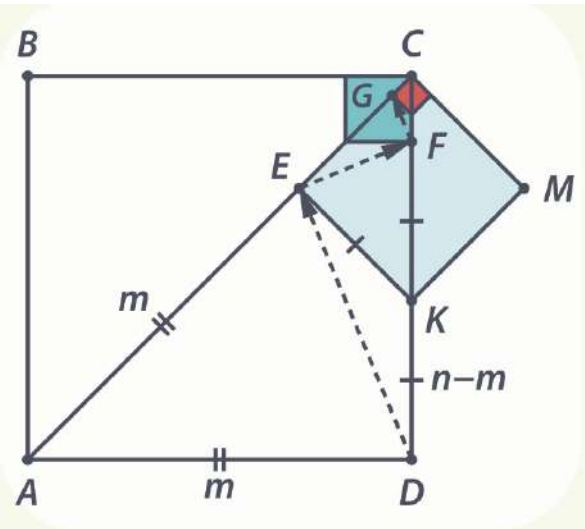
Цепные дроби

Широко известно высказывание Пифагора: «*Все сущее есть число*!» Поскольку на тот момент были известны лишь рациональные числа, то по сути Пифагор провозгласил, что «все сущее есть *рациональное* число». Это был не просто слоган, на этом основывалась вся философия учения Пифагора, его школа, его влияние в обществе. По иронии судьба именно Пифагору довелось опровергнуть свое же учение, на котором была построена его школа. А именно, Пифагор обнаружил, что диагональ квадрата со стороной 1 *не является числом*, точнее говоря, не является рациональным числом.

Замечание. В современной нотации это означает, что уравнение $m\sqrt{2} = n$ не имеет решений в натуральных числах. В самом деле, предположим, что натуральное решение (m, n) существует. Тогда можно считать, что эти числа взаимно просты (в противном случае мы просто сократим их на их НОД, что не изменит наше равенство). Возведем это равенство в квадрат: $2m^2 = n^2$. Значит, число n^2, а вместе с ним и n четно: $n = 2n_1$. Подставляя это в наше уравнение, получаем $m^2 = 2n_1^2$, откуда следует, что и m четно: $m = 2m_1$. Но тогда числа m и n оба четны, что невозможно по нашему предположению.

Доказательство Пифагора было совершенно другим — геометрическим. Рассмотрим квадрат со стороной 1 и диагональю $\sqrt{2}$. Предположим, что существует такой отрезок, который ровно m раз умещается на стороне квадрата и n раз на его диагонали. Отложим теперь сторону квадрата на его диагонали. На рисунке показано, как на диагонали AC квадрата $ABCD$ отложен отрезок AE, равный стороне квадрата. Диагональ квадрата больше его стороны, поскольку в треугольнике ACD гипотенуза больше катета. Остаток CE диагонали будет составлять $n - m$ единиц, значит, его длина в нашей системе тоже будет выражаться целым числом.





Построим теперь на отрезке CE меньший квадрат $CEKM$. Его вершина K окажется на стороне CD первого квадрата, поскольку в любом квадрате диагональ образует со стороной угол 45°. Легко видеть, что треугольники AED и EKD равнобедренные, следовательно, отрезок KD будет равен стороне квадрата EK и его длина тоже будет составлять целое число $n - m$ единиц.

А теперь давайте запустим процесс бесконечного спуска. Сделаем ту же самую операцию с меньшим квадратом: отложим на его диагонали CK отрезок KF, равный стороне этого квадрата. Тогда остаток CF этой диагонали снова будет состоять из целого числа единиц длины. Ведь и диагональ и сторона этого квадрата были целыми числами. Потом на отрезке CF построим третий квадрат, снова отложим на его диагонали сторону, построим на полученном остатке этой диагонали четвёртый квадрат, с ним сделаем то же самое, и так далее. Процесс этот будет бесконечным, ведь на диагонали любого квадрата всегда можно отложить его сторону. При этом стороны всех построенных нами квадратов будут целыми числами. Но такого, очевидно, быть не может — противоречие.

Эта ошибка могла стоить Пифагору очень дорого: ведь получалось, что вся его теория, все его учение, вся его школа — просто обман! А раз так, то, обнародовав это (великое) открытие, Пифагор тем самым погубил бы и себя, и свою школу и своих учеников. именно поэтому Пифагор *засекретил* это открытие и по легенде даже убил своего ученика Гиппаса, попытавшегося нарушить запрет Пифагора и сообщить об этом открытии людям.

Тем не менее, существование иррациональных чисел все же стало известно (причем весьма скоро после смерти Пифагора), поэтому требовалось их изучать. Интересно отметить, что даже строгое определение вещественного числа достаточно непросто (хотя, по-видимому, оно было известно Евдоксу), да и, прямо скажем, что работать на практике гораздо удобнее с числами рациональными.

Поясним кратко, из-за чего возникают существенные сложности. Во-первых, дадим собственно определение вещественного числа. Вещественное (или действительное) число — это *бесконечная десятичная дробь*. Казалось бы, не очень сложно. Но вот вопрос: если дробь *бесконечна*, то как можно проводить операции с такими объектами? Например, как определить сумму чисел $\sqrt{2}$ и $\sqrt{3}$? Если бы мы складывали целые числа, мы бы выписали их в столбик и считали бы *справа налево*, с самого маленького разряда. А здесь нет этого «права», поскольку дробь бесконечна!

Замечание. Относительно недавно, в начале XX века, люди поняли, что право ничем не лучше лева, и если вещественные числа представляют собой бесконечные *вправо* последовательности цифр, то почему бы не рассмотреть бесконечные *влево* последовательности цифр? Это, казалось бы, простое наблюдение привело в возникновению новой области математики, которая называется p-адическим анализом.

Более того, возникает и другой вопрос. Когда мы определяем число $\sqrt{2}$, мы говорим, что «это положительное вещественное число, квадрат которого равен 2». Но почему такое число *существует*? Ведь вещественного числа, квадрат которого равен -1, нет! Почему же найдется число с квадратом, равным 2?!

И хотя на все эти вопросы можно дать совершенно строгие ответы (что и было сделано Евдоксом), возникает естественное желание работать именно с рациональными числами, поскольку с ними всех этих проблем удается избежать. Поэтому возникает важная *практическая* задача: как приблизить иррациональное число рациональным с достаточно большой точностью?

С одной стороны, это очень легко. Например, можно взять десятичную запись числа $\sqrt{2} = 1{,}4142135\ldots$, оборвать ее в некотором месте (например, после цифры 5) и записать приближение

$$\sqrt{2} \approx \frac{14142135}{10000000} = \frac{2828427}{2000000}.$$

Точность, конечно, высокая — аж 7 точных знаков после запятой. Есть два «но». Во-первых, чтобы получить такое приближение, нам нужно каким-то образом *вычислить* десятичную запись числа $\sqrt{2}$ с достаточной точностью. А как быть, если вычисления провести не удастся? И во-вторых, согласитесь, работать с дробью, у которой числитель и знаменатель порядка миллиона, тоже неудобно. Хотелось бы приближать число не просто абы какими дробями, а с возможно небольшими числителем и знаменателем. Возможно ли это сделать?

Оказывается, да! Приведем здесь три вопроса, в попытках ответить на которые, по сути, и возникла *теория цепных дробей* — математическая теория, соединяющая в себе алгебру и геометрию и дающая исчерпывающий ответ на наш вопрос.

Число π

Число π, являющееся отношением длины окружности к ее диаметру, является, возможно, самой важной и самой известной математической константой. Достаточно точное вычисление этой константы практически необходимо, например, при астрономических расчетах или вычислении расстояний между точками на поверхности Земли. То, что π чуть больше тройки, знали уже в Древнем Египте. А вот историю вычисления следующих знаков числа π можно отслеживать, сравнивая уровень культуры и достижений той цивилизации, которая сумела продвинутся в этом вычислении. Первый значимый результат был получен Архимедом, и его знаменитое приближение 22/7 до сих пор используется при приближенных расчетах. Оно дает два точных знака числа π после запятой.

Следующий значимый шаг был сделан спустя более чем тысячу лет после Архимеда в Китае: китайские математики нашли приближение числа π с точностью в шесть знаков после запятой: $\pi \approx 355/113$.

История дальнейших вычислений числа π тесно связана с *рядами*. Например, И. Ньютон открыл следующее замечательное равенство:

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \ldots$$

Однако попытки вычислить с помощью этого равенство хотя бы два точных знача числа π скорее всего, будут неудачны. Дело в том, что ряд сходится очень медленно, поскольку знаменатели медленно растут. Чтобы посчитать два точных знака, придется просуммировать дроби со знаменателем, меньшим 100 (это как минимум!), что, разумеется, почти нереально сделать без компьютера.

Но были найдены и другие ряды, сходящиеся гораздо быстрее. Например, ряд

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{9801} \sum_{k=0}^{\infty} \frac{(4k)!(1103 + 26390k)}{(k!)^4 396^{4k}},$$

найденный С. Рамануджаном, дает на каждом новом члене по 14 следующих точных знаком числа π! А алгоритм $\pi \approx \frac{(a_n+b_n)^2}{4t_n}$, где $a_0 = 1$, $b_0 = 1/\sqrt{2}$, $t_0 = 1/4$, $p_0 = 1$ и

$$a_{n+1} = \frac{a_n + b_n}{2}, \quad b_{n+1} = \sqrt{a_n b_n}, \quad t_{n+1} = t_n - p_n(a_n - a_{n+1})^2, \quad p_{n+1} = 2p_n,$$

на каждом шаге учетверяет число точных знаков числа π. Сейчас такие алгоритмы используются для тестирования работы суперкомпьютеров.

Календарь

Еще одной задачей, в которой необходимо приблизить иррациональное число рациональным, является разработка достаточно точного и удобного календаря. Все знают, что Земля совершает один оборот вокруг Солнца за 365 дней. Однако на самом деле времени требуется чуть больше: 365,2421988 … дней. Это число нужно приблизить рациональной дробью.

В юлианской системе (разработанной группой александрийских астрономов во главе с Созигеном примерно в 45 г. до н.э. и названной в честь Юлия Цезаря) приближение было совсем грубым: $365 + \frac{1}{4}$ (хотя все же оно было много точнее имевшейся до этого римской системы, где один год полагался равным 355 дням!). Из-за этого достаточно быстро (быстро по меркам астрономическим; примерно за 1500 лет…) накопилась ошибка примерно в 10 дней. (На самом деле ошибка в 1 день накапливается в юлианском календаре за 128 лет.) В результате в 1582 г. был введен григорианский календарь (названный в честь Папы Римского Григория XIII), который давал точность приближения уже в 3 знака после запятой:

$$365{,}2425 = 365 + \frac{1}{4} - \frac{1}{100} + \frac{1}{400}.$$

Эта формула уточняет правило определения високосных лет: високосным является каждый четвертый год, если он только не является сотым. Однако каждый четырехсотый год по-прежнему объявляется високосным. Например, 2000 год был високосным. Точность григорианского календаря в десятки раз выше точности календаря юлианского: ошибка в 1 день накапливается за 3280 лет! Кстати говоря, в России григорианский календарь был введен лишь в 1918 г., и к том времени ошибка юлианского календаря составляла уже 13 дней. Поэтому мы и празднуем Старый Новый год 13 января.

Однако такие поправки (в которых еще нужно помнить) могут показать неудобными. На самом деле они таковы из-за того, что мы приближаем десятичную дробь очень топорно, как до этого приближали $\sqrt{2}$. Оказывается, что приблизить дробную часть числа 365,2421988 … можно более просто и точно. Вот несколько первых таких приближений:

$$\frac{1}{4}, \quad \frac{7}{29}, \quad \frac{8}{33}, \quad \frac{31}{128}, \quad \frac{132}{545}, \dots$$

Первая поправка породила юлианский календарь. Вторая поправка никогда не использовалась, поскольку почти ничем не отличалась от более точной третьей поправки, предложенной О. Хайамом в XI в. и положившей начало персидскому календарю. Суть его в том, что на каждые 33 года должно приходиться 8 високосных. Этого можно добиться так: семь раз считать високосным каждый четвертый год, а восьмой раз — *пятый* (т.к. $7 \cdot 4 + 5 = 33$). Этот календарь, во-первых, довольно удобен, а во-вторых, почти в полтора раза точнее григорианского: ошибка в 1 день набегает лишь за 4500 лет (!!!).

Как же получилось так: что григорианский календарь, учитывающий поправки порядка 100, оказался хуже персидского календаря, где знаменатель равен всего-навсего 33?! Ответ на этот вопрос мы получим, изучая цепные дроби.

Небесная механика

Цепные дроби появились в астрономических исследованиях не только при создании календаря, но и при вычислении затмений, движения планет и других периодичностей, которые появляются в небесной механике. При описании соизмеримости частот различных периодических движений, например кеплеровских движений планет, астрономы встретились с необходимостью знать хорошие рациональные приближения к этим, вообще говоря, иррациональным числам. При этом особенное значение имело, насколько хорошо можно приблизить число, вообще говоря иррациональное, рациональной дробью с не очень большим знаменателем. Слишком близкое приближение называется *резонансом* и может привести к сильному возмущению одной планетой движения другой.

Рассмотрим такую модель. Пусть две планеты (например, Юпитер и Сатурн) вращаются вокруг Солнца по концентрическим окружностям в одну сторону. Если отношение периодов их обращений вокруг Солнца с большой точностью равно рациональному числу с небольшими числителем и знаменателем (в реальности это отношение примерно равно $2 : 5$), то эти две планеты будут оказываться на маленьком расстоянии (минимальном возможном) друг от друга вблизи трёх (т.к. $3 = 5 - 2$) фиксированных точек. При маленьком расстоянии, как известно, наибольшая гравитация, так что орбиты обеих планет будут испытывать сильные деформации лишь в трёх направлениях. Планеты при этом как бы «сталкивают» друг друга с орбит. Это так называемые неравенства в движении Юпитера и Сатурна, которые имеют период около 800 лет.

Совсем другое дело, если отношение периодов обращений планет с большой точностью — рациональное число с большим знаменателем. Пусть оно равно, скажем, 151/700. Тогда «точек большой гравитации» $700 - 151 = 549$, и взаимное влияние («сталкивание») планет более «размазанное».

От чего зависят размеры рациональной дроби, которая достаточно точно приближает отношение периодов? Оказывается, что они зависят от так называемых *неполных частных* цепной дроби, в которую раскладывается отношение периодов. Поэтому астрономы очень рано (этим интересовались еще Ньютон и Кеплер) поставили себе вопрос, какие же практически величины этих неполных частных. Если какое-нибудь неполное частное, очень велико, скажем миллион, то приближение, которое получится, если оборвать дробь этим неполным частным, будет колоссально точным. Если же оно, например, всего только 2, то погрешность будет довольно большой.

Поэтому вопрос о том, возрастают ли эти коэффициенты, и с какой скоростью они возрастают, имеет реальное астрономическое значение для судьбы Вселенной, для судьбы Солнечной системы. Первое математическое исследование этого важного вопроса принадлежало, вероятно, астроному Х. Гильдену, который опубликовал его в докладах Парижской академии наук в 1888 году. По-видимому, это была экспериментальная работа, потому что астрономы исследовали отношения частот различных планет, больших и малых, и знали первые неполные частные отношений этих частот. Теорема, которая дала окончательный ответ на этот вопрос, называется теоремой Кузьмина, и к ней мы также обратимся в процессе изучения цепных дробей.

Цепные дроби для рациональных чисел

Теорию цепных дробей можно излагать как алгебраически, так и геометрически. Мы рассмотрим оба этих подхода, что позволит нам, с одной стороны, по-новому взглянуть на уже известные факты, а с другой — покажет взаимодействие алгебры и геометрии. Мы начнем изучение цепных дробей с наиболее простого случая рациональных чисел. Освоившись с аппаратом цепных дробей на этом уровне, мы затем будем готовы перенести наши конструкции на более трудные и серьезные вопросы, связанные с иррациональными числами.

По сути цепная дробь для рациональной дроби a/b со взаимно простыми числителем и знаменателем есть не что иное как алгоритм Евклида, записанный в несколько непривычной форме. Проще всего это продемонстрировать на конкретном примере.

Применим алгоритм Евклида для вычисления НОДа чисел 42 и 31. Проводя последовательно деления с остатком, получаем следующую цепочку равенств:

$$\begin{aligned} 42 &= 1 \cdot 31 + 11, \\ 31 &= 2 \cdot 11 + 9, \\ 11 &= 1 \cdot 9 + 2, \\ 9 &= 4 \cdot 2 + 1. \end{aligned}$$

Теперь разделим каждое равенство на делитель, который в нем присутствует (например, первое равенство нужно разделить на 31, второе — на 11, и т.д.):

$$\begin{aligned} \frac{42}{31} &= 1 + \frac{11}{31}, \\ \frac{31}{11} &= 2 + \frac{9}{11}, \\ \frac{11}{9} &= 1 + \frac{2}{9}, \\ \frac{9}{2} &= 4 + \frac{1}{2}. \end{aligned}$$

Заметим, что в левых и правых частях встречаются *обратные дроби*: например, 11/31 и 31/11. Давайте подставим каждую дробь, начиная с последней, в предыдущее равенство. В итоге мы получим следующее выражение:

$$\frac{42}{31} = 1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{2}}}}.$$

Выражение правой части и называется *цепной дробью* числа 31/11. Именно эти выражения мы будем изучать в данной главе.

Теперь опишем процесс построения произвольной цепной дроби положительного рационального числа a/b. Этот процесс будет быстрее, нежели тот, с помощью которого мы пришли к самому понятию цепной дроби. А именно, возьмем дробь a/b и выделим в ней целую часть $q_0 = [a/b]$. Оставшаяся дробная часть r_0/b будет меньше 1. Теперь сделаем «двойной переворот» этой дробной части:

$$\frac{a}{b} = q_0 + \frac{r_0}{b} = q_0 + \frac{1}{b/r_0}.$$

Далее, возьмем дробь b/r_0 и применим к ней ту же самую процедуру: выделим целую часть q_1, возьмем остаток, перевернем его, и т.д. В итоге мы получим следующее представление нашей исходной дроби:

$$\frac{a}{b} = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \cfrac{1}{q_3 + \ldots + \cfrac{1}{q_{m-1} + \cfrac{1}{q_m}}}}}.$$

Здесь $q_0 \in \mathbb{N} \cup \{0\}$ — натуральное число или 0, $q_i \in \mathbb{N}$ при $i \geqslant 1$, причем $q_m > 1$.

Возникает два вопроса. Во-первых, как разложить в цепную дробь иррациональные числа? Во-вторых, если у нас есть цепная дробь, как быстро вычислить соответствующее ей рациональное число? Первый вопрос является чрезвычайно сложным, и мы вернемся к нему позже, когда немного привыкнем к цепным дробям. Сейчас мы получим детальный ответ на второй вопрос.

В ходе наших рассуждений мы совершенно не будем использовать тот факт, что цепная дробь

$$q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \ldots + \cfrac{1}{q_{m-1} + \cfrac{1}{q_m}}}},$$

которую мы изучаем, состоит из натуральных чисел. мы будем работать с величинами q_0, ..., q_m как с обычными переменными, поэтому многие факты, которые мы сейчас докажем, останутся справедливыми при замене натуральных чисел на, скажем, многочлены.

Давайте начнем с рассмотрения примеров преобразования нашей цепной дроби в рациональную для маленьких значений m. Простыми вычислениями можно убедиться, что

$$q_0 + \frac{1}{q_1} = \frac{q_0 q_1 + 1}{q_1},$$

$$q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2}} = \frac{q_0 q_1 q_2 + q_0 + q_2}{q_1 q_2 + 1},$$

$$q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \cfrac{1}{q_3}}} = \frac{q_0 q_1 q_2 q_3 + q_0 q_1 + q_0 q_3 + q_2 q_3 + 1}{q_1 q_2 q_3 + q_1 + q_3}.$$

Какие закономерности можно обнаружить, глядя на эти формулы? Например, видно, что числитель каждой дроби похож на знаменатель следующей за ней дроби, только в знаменателе все индексы сдвинуты на 1. Это несложно доказать: для этого достаточно заметить, что цепная дробь

$$q_1 + \cfrac{1}{q_2 + \ldots + \cfrac{1}{q_{m-1} + \cfrac{1}{q_m}}}$$

содержится в исходной цепной дроби, а числитель этой окажется знаменателем содержащей ее следующей цепной дроби.

Давайте для удобства обозначим числитель и знаменатель нашей цепной дроби

$$q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \ldots + \cfrac{1}{q_{m-1} + \cfrac{1}{q_m}}}}$$

через $A_m = [q_0, q_1, \ldots, q_m]$ и $B_m = [q_1, \ldots, q_m]$. Обратите внимание: в обозначениях мы отражаем тот факт, что знаменатель числитель каждой дроби вычисляется так же, как знаменатель следующей за ней дроби с учетом сдвига индексов на 1. Наша цель — получить рекуррентные формулы для A_m и B_m.

Прежде всего заметим, что верны следующие (хотя и не удобные на практике) формулы:

$$[q_0, \ldots, q_m] = q_0[q_1, \ldots, q_m] + [q_2, \ldots, q_m]. \tag{1}$$

В самом деле,

$$\frac{[q_0, \dots, q_m]}{[q_1, \dots, q_m]} = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \dots + \cfrac{1}{q_{m-1} + \cfrac{1}{q_m}}}} = q_0 + \frac{[q_2, \dots, q_m]}{[q_1, \dots, q_m]},$$

откуда и следует формула (1).

Почему эта формула неудобна для практических вычислений? Дело в том, что для вычисления величины $[q_0, \dots, q_m]$ нам потребуется вычислить также величины $[q_1, \dots, q_m]$ и $[q_2, \dots, q_m]$, а для них в свою очередь — $[q_3, \dots, q_m]$, и т.д. Поэтому процесс будет слишком долгим. Хотелось бы побыстрее…

1. **(a) Правило Эйлера.** Величина $[q_0, \dots, q_m]$ строится следующим образом. Сначала считается произведение всех чисел q_0, …, q_m, затем — произведения всех чисел, которые остаются при выбрасывании из этого набора одной пары рядом стоящих чисел, затем — при выбрасывании двух пар, и т.д. В случае, если выброшены все элементы, то произведение по определению полагается равным 1. Сумма этих произведений и равна $[q_0, \dots, q_m]$.
   (b) Докажите, что имеет место следующая симметрия:

   $$[q_0, q_1, \dots, q_{m-1}, q_m] = [q_m, q_{m-1}, \dots, q_1, q_0].$$

2. Докажите следующие рекуррентные формулы: $A_m = q_m A_{m-1} + A_{m-2}$, $B_m = q_m B_{m-1} + B_{m-2}$ (здесь $A_{-1} = B_0 = 1$ и $A_{-2} = B_{-1} = 0$).

3. Вычислите цепные дроби со следующими наборами неполных частных: **(a)** $[3, 7, 16]$; **(b)** $[1, 2, 2, 2, 2]$; **(c)** $[2, 1, 2, 1, 1, 4, 2]$.

4. **(a)** Докажите, что $A_m B_{m-1} - A_{m-1} B_m = (-1)^{m-1}$.
   (b) Докажите, что числа A_m и B_m взаимно просты.

5. Пусть n — натуральное число и a_1, …, a_{n-1} — произвольные вещественные числа. Определим последовательности u_0, …, u_n и v_0, …, v_n следующими соотношениями: $u_0 = u_1 = v_0 = v_1 = 1$, $u_{k+1} = u_k + a_k u_{k-1}$, $v_{k+1} = v_k + a_{n-k} v_{k-1}$. Докажите, что $u_n = v_n$.

Матрицы и цепные дроби

Алгоритм Евклида и построение цепной дроби рационального числа $\frac{a}{b}$ допускает матричное описание. Для этого запишем равенство $a = bq_0 + r_0$ и рассмотрим две дроби $\frac{a_0}{b_0} = \frac{a}{b}$ и $\frac{a_1}{b_1} = \frac{b}{r_0}$. Ясно, что $a_0 = q_0 a_1 + b_1$ и $b_0 = a_1$. Эти равенства можно записать в следующем виде: $\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}$. Продолжая записывать такие равенства для получающихся дробей $\frac{a_k}{b_k}$, получаем следующую формулу:

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \ldots \begin{pmatrix} q_m & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Определение. *Матрицей цепной дроби* называется матрица

$$M^+(q_0, \ldots, q_m) = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \ldots \begin{pmatrix} q_m & 1 \\ 1 & 0 \end{pmatrix}.$$

1. Вычислите матрицу цепной дроби числа $\frac{10}{7}$.

2. Докажите, что $M^+(q_0, \ldots, q_m) \begin{pmatrix} 0 \\ 1 \end{pmatrix} = M^+(q_0, \ldots, q_{m-1}) \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

3. **(a)** Рассмотрим цепную дробь для числа $\frac{a}{b}$, и пусть $A_k = [q_0, q_1, \ldots, q_k]$ и $B_k = [q_1, \ldots, q_k]$, где $k = 0, 1, \ldots, m$ и $B_0 = 1$. Докажите, что $M^+(q_0, \ldots, q_m) \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} A_{m-1} \\ B_{m-1} \end{pmatrix}$.

(b) Докажите, что $M^+(q_0, q_1, \ldots, q_m) = \begin{pmatrix} A_m & A_{m-1} \\ B_m & B_{m-1} \end{pmatrix}$.

(c) Докажите, что $\begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \ldots \begin{pmatrix} q_k & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} A_k & A_{k-1} \\ B_k & B_{k-1} \end{pmatrix}$.

4. **(a)** Докажите свойства чисел A_k и B_k из задач 2, 3ab, 1b (именно в таком порядке) с помощью матрицы цепной дроби.

(b) Решите задачу 4 с помощью матриц.

Геометрическая интерпретация цепных дробей

После арифметического знакомства с теорией цепных дробей мы теперь дадим ее геометрическое представление. Оказывается, что многие результаты, полученные нами в предыдущем разделе, допускают простую геометрическую интерпретацию, зачастую уже известную нам. Более того, именно геометрическая теория цепных дробей в настоящий момент является предметом активного изучения, поскольку в таком виде она допускает *многомерное обобщение*, с которым мы кратко познакомимся позднее.

Сейчас мы будем раскладывать в цепную дробь не только рациональные, но и произвольные вещественные числа, большие 1. Сначала опишем алгераический процесс разложения числа $\alpha > 1$ в цепную дробь. Для этого положим $q_0 = [\alpha]$ и $\{\alpha\} = \alpha - q_0$ — дробная часть. Теперь $\alpha = q_0 + \dfrac{1}{\alpha_1}$, где $\alpha_1 = \dfrac{1}{\alpha}$. Повторяя процесс для числа α_1, мы получаем следующее число q_1, и т.д.

Самым трудным и самым важным для нас будет проинтерпретировать процесс построения цепной дроби на языке геометрии. У нас есть неплохой задел для этого: геометрия целочисленной решетки $\mathbb{Z}^2$. Сейчас мы увидим, как уже известные нам результаты из этой области находят свое применение в теории цепных дробей.

Рассмотрим на координатной плоскости прямую ℓ, заданную уравнением $y = \alpha x$. Обозначим через $\bar{\imath}$ и $\bar{\jmath}$ базисные векторы. Тогда нашу прямую можно задать с помощью *направляющего вектора* $\bar{a}_0$, т.е. вектора, коллинеарного прямой. Поскольку таких векторов много (они отличаются друг от друга умножением на константу), нормируем его так, чтобы координата $\bar{a}_0$ по оси абсцисс была бы равна 1. Тогда $\bar{a}_0 = \bar{\imath} + \alpha\bar{\jmath}$.

Заметим, что цепная дробь строится по следующему алгоритму. Сначала мы выделяем целую часть числа, затем рассматриваем перевернутую дробную часть и делаем для нее то же самое. Значит, нам нужно научиться выделять целую часть и переворачивать дробную часть геометрически.

Выделение целой части выглядит следующим образом. Рассмотрим наибольшее $q_0 \in \mathbb{N} \cup \{0\}$, такое, что вектор $\bar{e}_0 = \bar{\imath} + q_0\bar{\jmath}$ лежит ниже прямой ℓ. Несложно видеть, что $q_0 = [\alpha]$, поскольку угловой коэффициент α прямой ℓ должен быть чуть больше углового коэффициента q_0 вектора $\bar{e}_0$. Поэтому выделить целую часть числа — значит поднять вектор оси абсцисс с помощью наибольшего возможного целого вектора, пропорционального вектору с оси ординат.

Теперь самое сложное. Как перевернуть дробную часть? Сначала ее нужно выделить. Для этого рассмотрим новую систему координат, образованную векторами $\bar{e}_0$ и $\bar{\jmath}$ (взятыми именно в таком порядке, это важно!). Прямая ℓ никак не поменяется, однако ее уравнение, безусловно, станет другим. Каким именно? Чтобы понять это, удобно следить за образующим вектором $\bar{a}_0$. Разложим его по векторам $\bar{e}_0$ и $\bar{\jmath}$:

$$\bar{a}_0 = \bar{\imath} + \alpha\bar{\jmath} = \bar{\imath} + (q_0 + \{\alpha\})\bar{\jmath} = (\bar{\imath} + q_0\bar{\jmath}) + \{\alpha\}\bar{\jmath} = \bar{e}_0 + \{\alpha\}\bar{\jmath}.$$

Значит, уравнение нашей прямой ℓ в новой системе координат $\{\bar{e}_0, \bar{\jmath}\}$ выглядит так: $y = \{\alpha\}x$.

Вот мы и выделили дробную часть. Остался последний шаг: нужно ее перевернуть. Этот прием должен быть знаком нам из курса алгебры: нужно просто поменять координаты местами…

Рассмотрим теперь систему координат $\{\bar{\jmath}, \bar{e}_0\}$! Посмотрим, как изменится уравнение нашей прямой ℓ. Для этого опять же проследим за направляющим вектором $\bar{a}_0$. Теперь он будет разложен так: $\bar{a}_0 = \{\alpha\}\bar{\jmath} + \bar{e}_0$. По нашему соглашению мы выбираем такую нормировку направляющего вектора, при которой *первая* его координата равна 1. Поэтому вместо вектора $\bar{a}_0$ нужно взять вектор

$$\bar{a}_1 = \frac{1}{\{\alpha\}}\bar{a}_0 = \bar{\jmath} + \frac{1}{\{\alpha\}}\bar{e}_0.$$

Тем самым уравнение прямой ℓ теперь выглядит так: $y = \alpha_1 x$, где $\alpha_1 = 1/\{\alpha\}$. Вот и переворот!

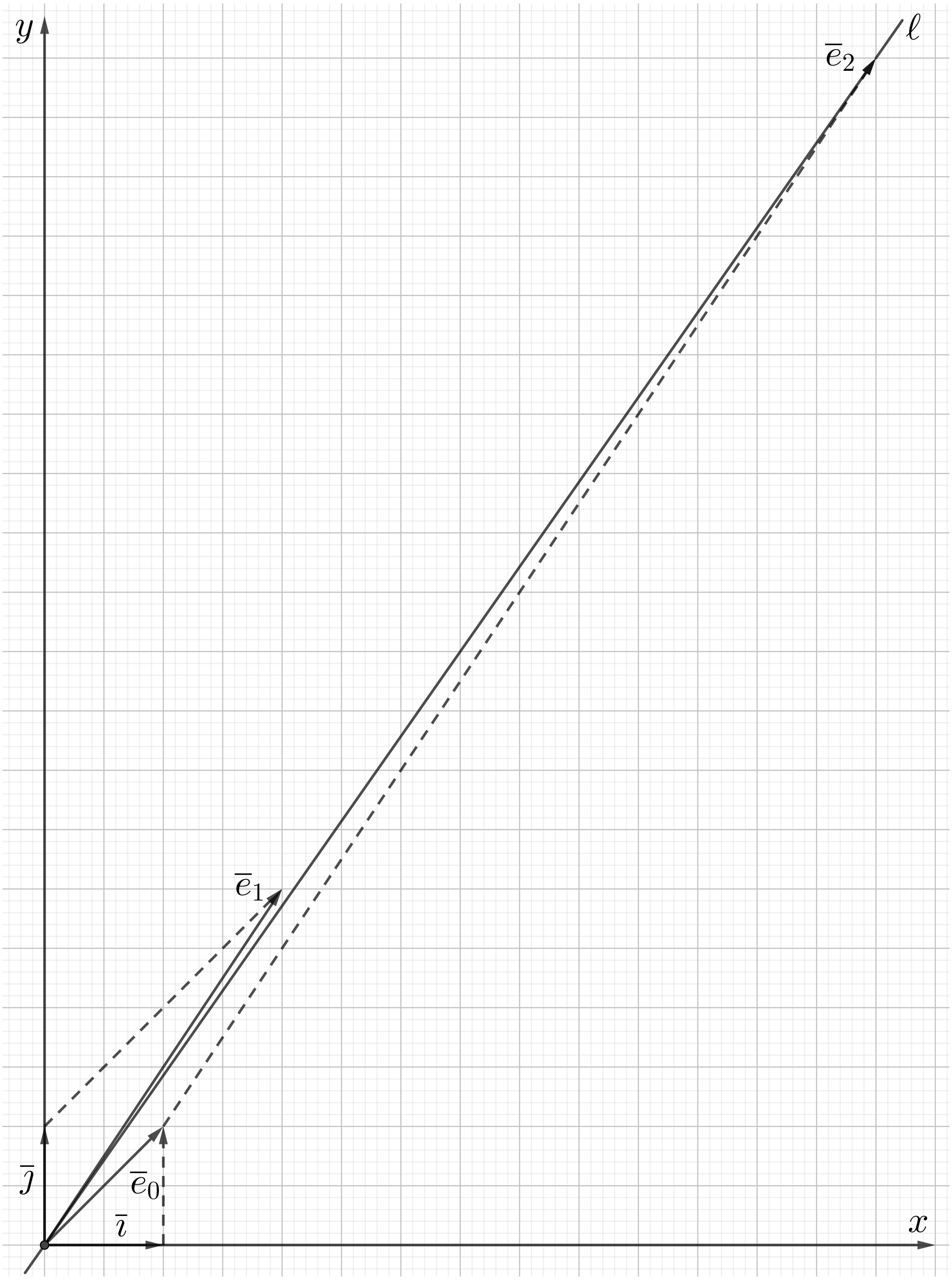
Осталось лишь зациклить процедуру: применить те же действия к вектору $\bar{a}_1$, потом к $\bar{a}_2$, и т.д. В результате мы найдем числа q_0, q_1, q_2, …, которые и будут неполными частными цепной дроби для числа α.

Полученный алгоритм называется *алгоритмом вытягивания носов*, поскольку векторы $\bar{e}_0, \bar{e}_1$, …становятся все длиннее и длиннее. Если в какой-то момент вектор $\bar{e}_k$ станет коллинеарен нашей прямой ℓ, то процесс построения цепной дроби завершится, что, как мы помним, возможно лишь в случае рационального числа α. В противном случае «носы» $\bar{e}_k$ будут подходить к прямой ℓ все ближе и ближе, но никогда не лягут на нее.

Чтобы лучше понять алгоритм вытягивания носов, полезно будет провести его с помощью рисунка ниже, где показан процесс для $\alpha = \dfrac{10}{7}$. Отметим, однако, что наиболее интересные результаты с помощью этого алгоритма удается получить в случае, когда число α иррационально.

Картинка из алгоритма вытягивания носов может быть получена сразу, а не пошагово, следующим образом. Рассмотрим на нашей координатной плоскости прямую ℓ и целочисленную решетку $\mathbb{Z}^2$. Будем рассматривать лишь те точки этой решетки, которые лежат в первой координатной четверти. Тогда прямая ℓ делит эту решетку на две части. Если угловой коэффициент α этой прямой иррационален, то на самой прямой ℓ нет целых точек. Тогда возьмем все целые точки, лежащие сверху от прямой ℓ и обозначим через $\mathcal{P}_+$ *выпуклую оболочку* этого множества точек.

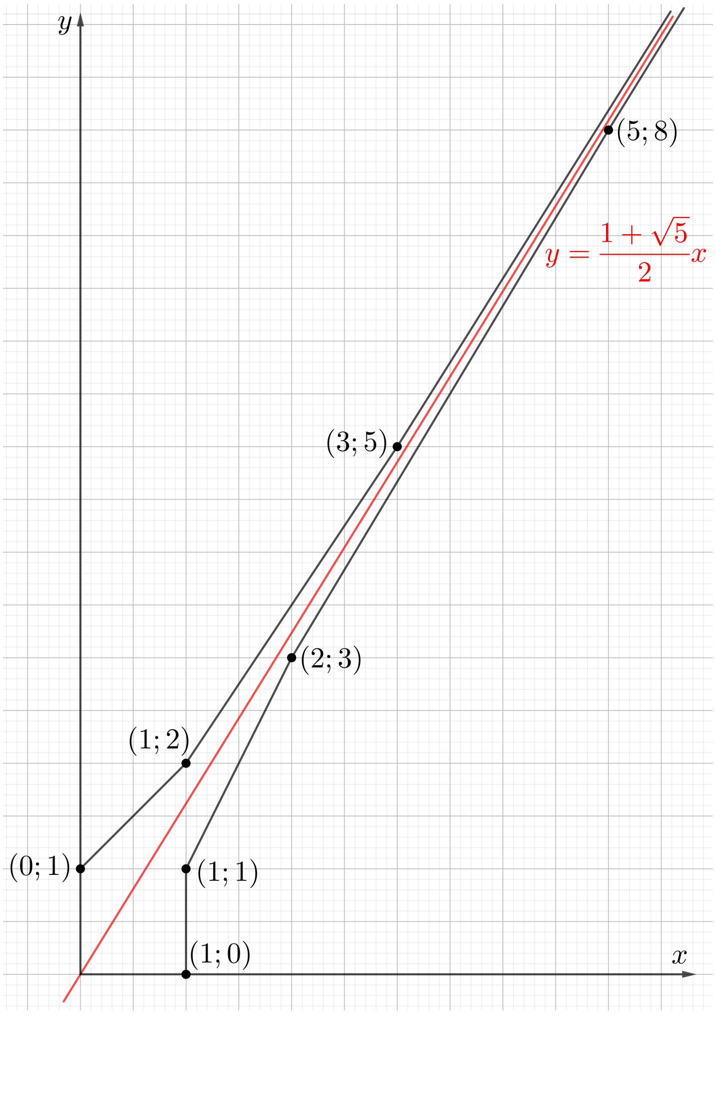
Поясним, что значит «выпуклая оболочка». Неформально говоря, представить себе это можно следующим образом. Вобьем колышки в узлах нашей целочисленной решетки и попробуем натянуть забор, ограничивающие верхнюю часть этой решетки. Те колышки, через которые пройдет забор, и будут концами векторов $\bar{e}_m$ (а количества целые точек на отрезках, соединяющих соседние колышки, равны неполным частным q_m; см. текст далее).

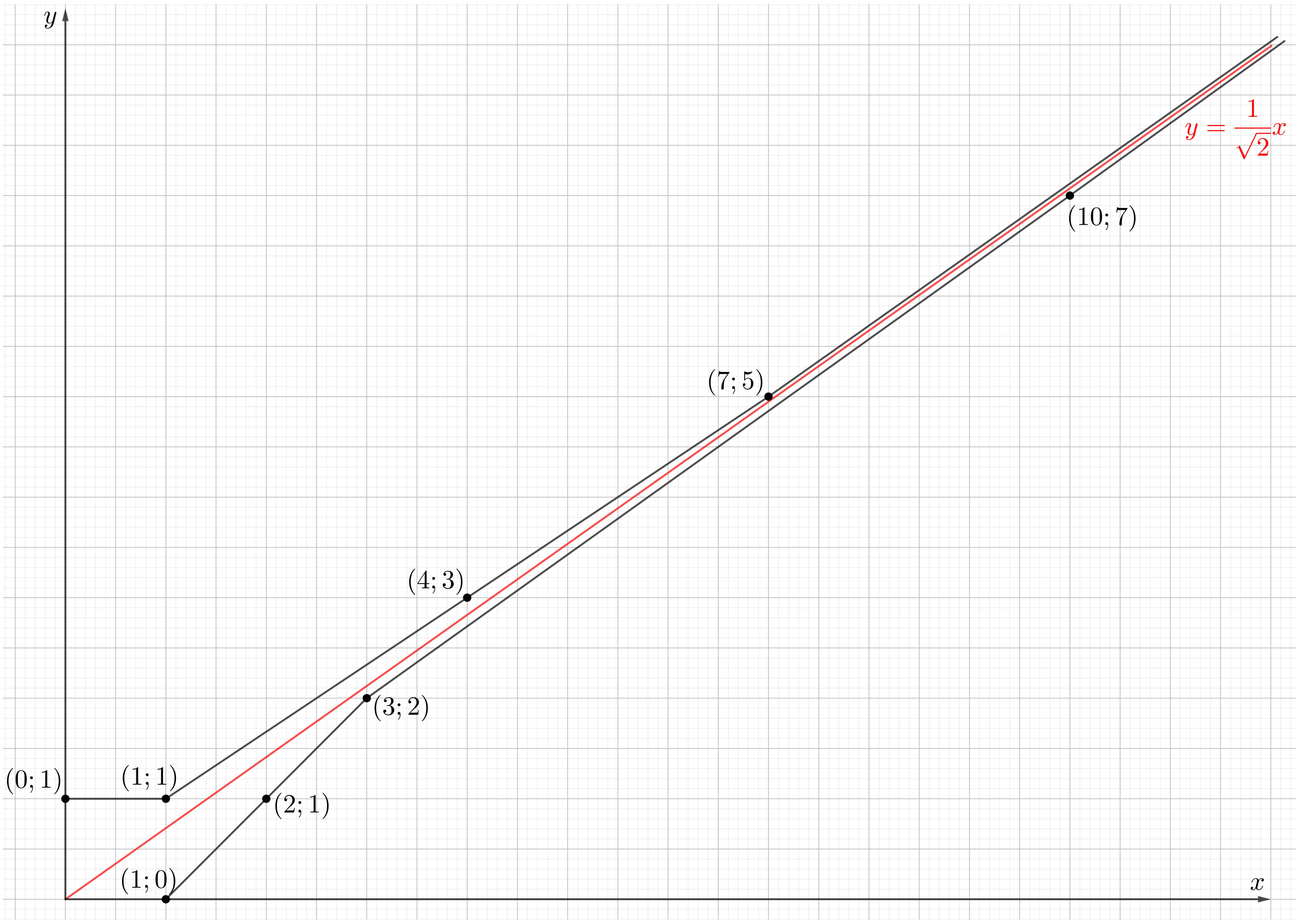


y
ℓ
$\overline{e}_2$
$\overline{e}_1$
$\overline{\jmath}$
$\overline{e}_0$
$\overline{\imath}$
x

Формально говоря, *выпуклая оболочка* множества (конечного или бесконечного) точек на плоскости — это наименьшее (относительно включения) выпуклое множество, содержащее все точки нашего множества. Или по-другому: выпуклая оболочка — это пересечение всех выпуклых фигур, содержащих наше множество точек.

Если аналогичную конструкцию проделать с нижней половиной целых точек, то мы получим вторую выпуклую оболочку $\mathcal{P}_-$. Целые точки, лежащие на границах $\mathcal{P}_+$ и $\mathcal{P}_-$, можно мыслить как элементы цепной дроби. А именно, числа q_i равны количеству частей, на которые делится отрезок между концами векторов $\bar{e}_{k-2}$ и $\bar{e}_k$ узлами решетки $\mathbb{Z}^2$ (при этом $\bar{e}_{-2} = \bar{\imath}$ и $\bar{e}_{-1} = \bar{\jmath}$). В самом деле, ведь вектор $\bar{e}_k$ получается из вектора $\bar{e}_{k-2}$ поднятием (прибавлением) его на вектор $q_k\bar{e}_{k-1}$ к прямой ℓ. Так вот звено нашего паруса, соединяющее концы векторов $\bar{e}_{k-2}$ и $\bar{e}_k$, и равно вектору $q_k\bar{e}_{k-1}$! Значит, на этом звене ровно q_k частей (равных вектору $\bar{e}_{k-1}$, на котором, кстати говоря, нет целых точек, кроме конца и начала).





Прямо скажем, что геометрическая конструкция цепных дробей достаточно трудна. Получим ли мы что-нибудь интересное за наши труды? Оказывается, что получим мы очень многое! Фактически *все* результаты, которые были ранее доказаны нами чисто арифметически, сейчас мы получим бесплатно, просто как следствие нашей конструкции. так что сложности, с которыми мы столкнулись в процессе геометрического построения цепной дроби, оправданы: ведь в этом построении содержится вся их теория!

1. Запишите разложение вектора $\overline{e}_m$ по базису $\{\overline{\imath}, \overline{\jmath}\}$.

2. Докажите, что $M^+(q_0, q_1, \dots, q_m)\imath = \begin{pmatrix} B_m \\ A_m \end{pmatrix}$ и $M^+(q_0, q_1, \dots, q_m)\jmath = \begin{pmatrix} B_{m-1} \\ A_{m-1} \end{pmatrix}$.

3. Докажите, что $A_m B_{m-1} - A_{m-1} B_m = (-1)^{m-1}$.

4. Через два узла клетчатой бумаги провели параллельные прямые. Докажите, что в замкнутой полосе бесконечно много узлов клетчатой бумаги.

5. Рассмотрим целочисленную решетку $\mathbb{Z}^2$. Построим на ней бесконечную ломаную следующим образом. Из некоторой точки проведем горизонтальный отрезок направо, или вертикальный отрезок наверх. После m вертикальных и n горизонтальных шагов продолжим нашу ломаную периодически, где период — это построенный вначале фрагмент. Теперь рассмотрим следующую операцию «загибания уголка» в стартовом фрагменте: $\ulcorner \longrightarrow \lrcorner$. Затем полученный стартовый фрагмент продолжим периодически до новой ломаной. При каких m и n существует ломаная, такая, что, загнув у нее некоторый уголок, мы получим ломаную, отличающуюся от исходной параллельным переносом?

Геометрия на клетчатой бумаге

Герман Вейль Сказал: «За душу каждого математика борются ангел чистой геометрии и дьявол абстрактной алгебры.» Но иногда эти двое действуют сообща. И тогда возникает Красота. Тема, которую мы рассмотрим в этом разделе, является ярким примером такого совместного взаимодействия алгебры и геометрии. Называется она *геометрия чисел*. Несмотря на свою красоту и наглядность, эта область математики родилась относительно недавно: в начале XX века. Ее основоположником был Г. Минковский, один из выдающихся математиков первой половины XX века. С другой стороны, некоторые результаты этой теории, связанные с приближением иррациональных чисел, были известны еще Дирихле и Лиувиллю.

Зафиксируем основные понятия, которыми мы будем пользоваться на протяжении всего этого раздела.

Определение. *Целочисленной решеткой* называется множество точек с целыми координатами на плоскости. Обозначается целочисленная решетка через $\mathbb{Z}^2$. Элементы этой решетки называются *узлами*.
Многоугольник называется *целым*, если его вершины лежат в узлах целочисленной решетки.

Оказывается, что многие *геометрические* свойства целых многоугольников легко и просто формулируются в *алгебраических* терминах, и наоборот, ряд алгебраических формул быстро и просто доказывается с помощью геометрии, связанной с такими многоугольниками. Во второй половине XX века эти результаты привели к появлению множества новых теорем, связанных с самыми разными областями математики. Основную роль в появлении этих результатов сыграли выдающийся российский математик А. Хованский и его ученики.

Мы начнем наш рассказ целых многоугольниках с известного софизма. Рассмотрим рис. 1 (слева). На нем изображен квадрат со стороной, равной восьми клеткам, который разрезан на четыре части: два треугольника и две трапеции. Если переложить эти части, как показано на рис. 1 (справа), то из них удается собрать прямоугольник со сторонами 5 и 13. Но площадь исходного квадрата равна 64, а площадь получившегося четырехугольника равна 65. Каким образом возникла лишняя клетка?!

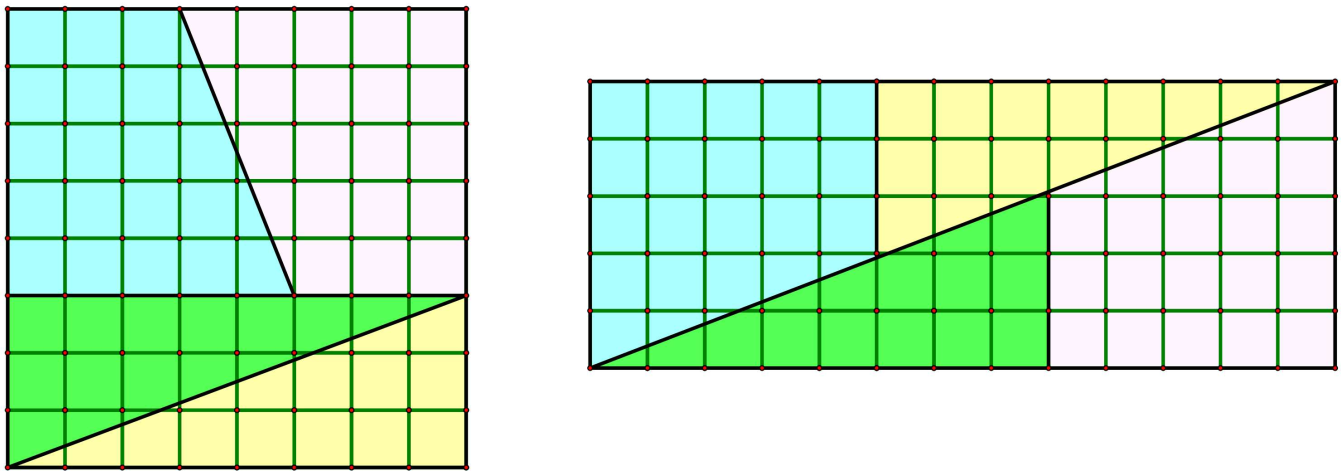


Рис. 1:

Ответ на этот вопрос содержится на рис. 2. Все дело в том, что в прямоугольнике вершины трапеций и треугольников не лежат на его диагонали! Видно, что между сторонами этих трапеций и параллелограммов остается небольшой зазор, площадь которого как раз и равна той самой «лишней» клетке. Более того, ясно, что фигура, образующая этот зазор, является параллелограммом (поскольку ее боковые стороны попарно равны).

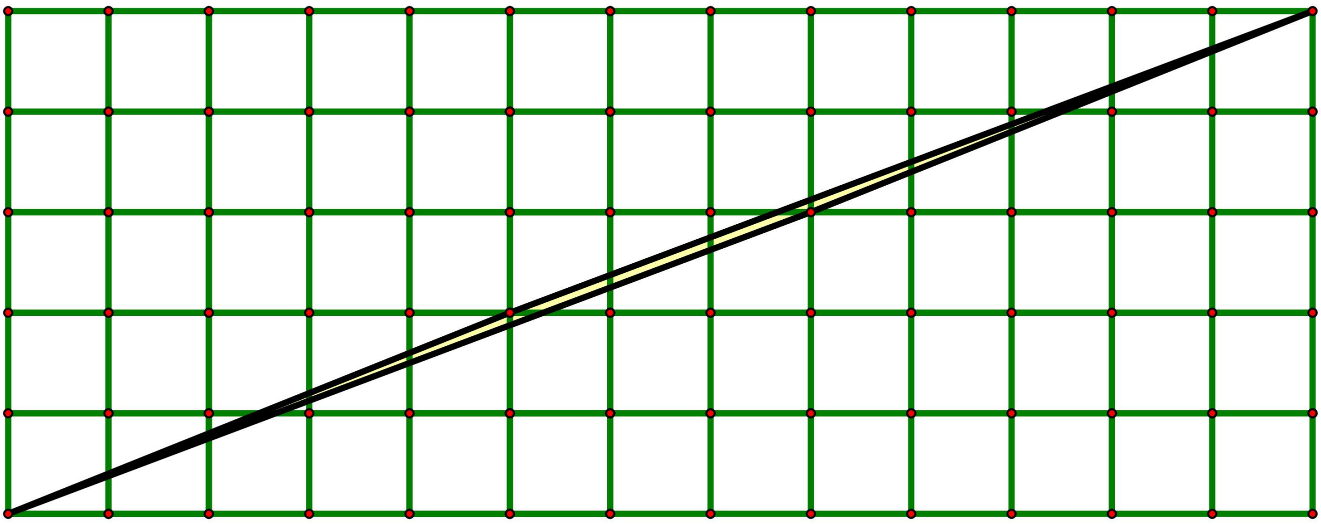


Рис. 2:

Софизм этот действительно впечатляет. Можно ли придумать другие подобные разрезания? Какие целочисленные точки лежат столь близко с наклонной прямой, что их не удается отличить на глаз? Что собой представляют параллелограммы, натянутые на такие точки? Всегда ли их площадь равна 1, или можно сделать их еще меньше? Ответам на эти вопросы мы и посвятим данный раздел.

Мы начнем с ответа на вопрос о площади «узких» параллелограммов. Как формализовать «узость»? Проще всего это сделать так: будем рассматривать параллелограммы, не содержащие ни внутри, ни на границе целых точек (за исключением вершин). Оказывается, что площади таких параллелограммов уменьшить нельзя! Более того, их площади вообще не меняются

Теорема о фундаментальном параллелограмме. *Целый параллелограмм не содержит на внутри, ни на границе узлов решетки (за исключением вершин) если и только если его площадь равна 1.*

Перед тем, как переходить к доказательству, сделаем несколько замечаний. Во-первых, нам потребуется следующий простой способ построения параллелограммов. Рассмотрим два вектора, выходящих из одной точки (узла решетки). Тогда четырехугольник, построенный на этой общей точке, двух концах этих векторов и конце их *суммы*, будет параллелограммом. В самом деле, это есть не что иное как правило сложения векторов! Поэтому, говоря о параллелограмме, часто бывает удобно мыслить его как пару векторов. Мы в дальнейшем тоже будем часто придерживаться этой точки зрения.

Во-вторых, поясним, почему в названии теоремы сказано «фундаментальный параллелограмм». Здесь речь идет о важном понятии *базиса*. Мы знаем, что по определению *базисом* $\{\bar{e}_1, \bar{e}_2\}$ *на плоскости* называется пара неколлинеарных (т.е. непропорциональных) векторов. *Теорема о базисе* утверждает, что любой вектор $\bar{v}$ на плоскости можно представить в виде линейной комбинации векторов базиса: $\bar{v} = x\bar{e}_1 + y\bar{e}_2$, при этом числа x и y называются *координатами* вектора $\bar{v}$ в базисе $\{\bar{e}_1, \bar{e}_2\}$. Отметим, что базисов существует бесконечное множество, и в разных базисах один и тот же вектор имеет *разные* координаты. Смысл базиса заключается в том, что часто удается свести рассуждения к рассмотрению только *двух* векторов базиса вместо рассмотрения *бесконечного* количества векторов с плоскости.

А теперь попробуем породить не всю плоскость, а только *целочисленную решетку* $\mathbb{Z}^2$. Хотелось бы отразить специфику, связанную с целыми числами, фигурирующими в определении решетки. Для этого будем рассматривать не произвольные линейные комбинации векторов $\bar{e}_1$ и $\bar{e}_2$, а лишь *целочисленные*, т.е. комбинации вида $n\bar{e}_1 + m\bar{e}_2$, где $n, m \in \mathbb{Z}$. Возникает вопрос: верно ли, что произвольные целые векторы $\bar{e}_1$ и $\bar{e}_2$ образуют *базис решетки*, т.е. любой вектор является их целочисленной линейной комбинацией? Легко сообразить, что нет: векторы $\bar{e}_1 = (1, 0)$ и $\bar{e}_2 = (0, 2)$ не смогут породить вектор $\bar{v} = (0, 1)$, поскольку $\bar{v} = (1/2)\bar{e}_2$. А какие же векторы образуют базис целочисленной решетки?! Ответ на этот вопрос дает следующая теорема.

Теорема о базисе целочисленной решетки. *Целочисленные векторы* $\{e_1, e_2\}$ *образуют базис целочисленной решетки* $\mathbb{Z}^2$ *тогда и только тогда, когда параллелограмм, построенный на этих векторах, не содержит узлов решетки (или, что то же самое, имеет площадь 1).*

Именно поэтому целый параллелограмм, который не содержит узлов решетки, называется *фундаментальным*: он порождает эту решетку! Доказать эту теорему вы сможете после доказательства теоремы о фундаментальном параллелограмме.

Замечание. Прежде чем переходить к строгому доказательству теоремы о фундаментальном параллелограмме, мы предпошлем ему более простое и наглядное (но менее строгое) рассуждение, которое важно тем, что *объясняет*, почему площадь фундаментального параллелограмма действительно равна 1.

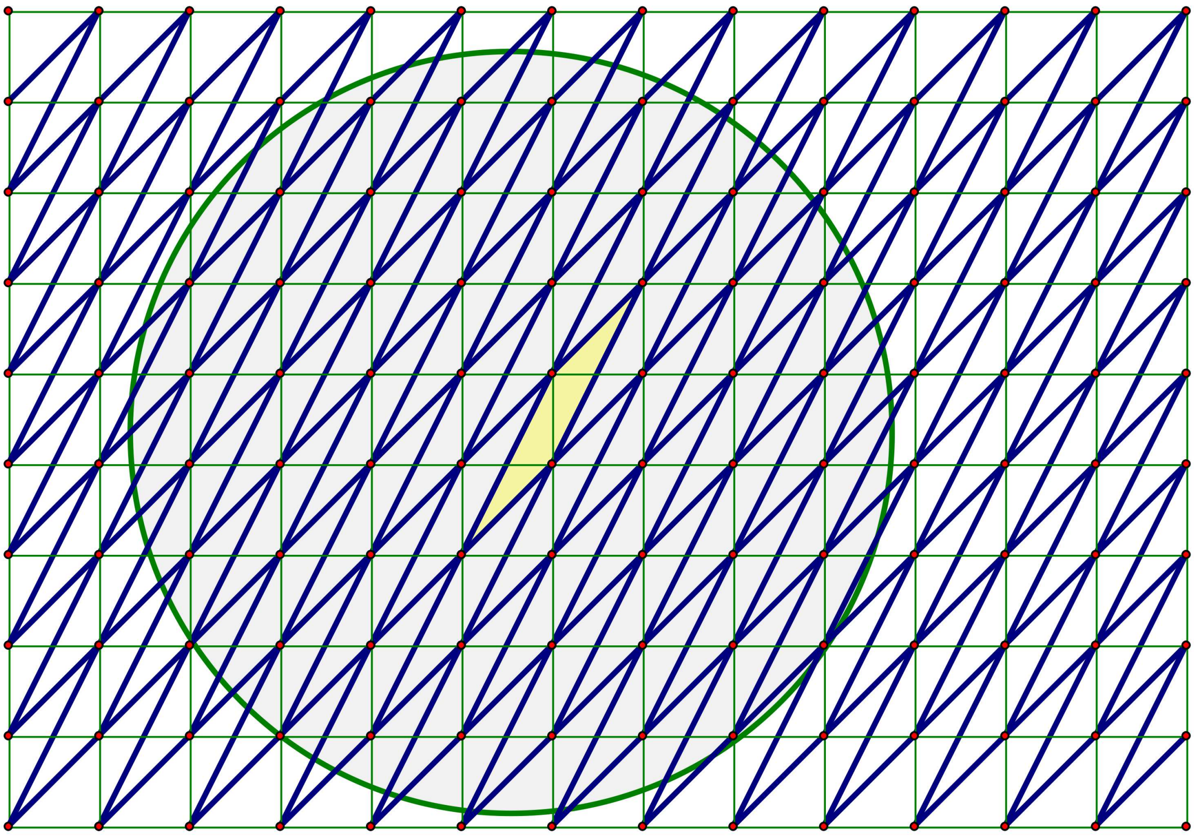


Рис. 3:

Сдвигая наш параллелограмм на всевозможные комбинации векторов, на которые он натянут, мы можем покрыть всю плоскость равными параллелограммами, подобно тому, как её покрывали единичные квадратики, образованные линиями координатной сетки (рис. 3). Возьмем кусок плоскости большой площади A и посчитаем, сколько в нем, с одной стороны, наших параллелограммов, с другой стороны, целых точек. Пусть площадь параллелограмма равна S, тогда, если площадь A очень велика, число параллелограммов приблизительно равно A/S (этот кусок плоскости не обязательно состоит из целых параллелограммов, поэтому равенство будет неточным; впрочем, можно взять кусок, состоящий из целых параллелограммов, тогда получится точное равенство). Понятно, что число целых точек примерно равно A. Посчитаем теперь число целых точек в нашей области другим способом. На каждый параллелограмм приходится 4 точки (его вершины), но при этом мы считаем каждую вершину 4 раза, и если мы посчитаем число всех вершин всех параллелограммов, то получится в 4 раза больше, чем число всех целых точек вообще. Поэтому целых точек и параллелограммов одинаковое количество. Получается, что $A \approx A/S$ при очень большом A. Значит, $S = 1$.

Доказательство теоремы о фундаментальном параллелограмме. Доказательство теоремы о фундаментальном параллелограмме является ярким примером взаимодействия алгебры и геометрии. Казалось бы, алгебраическое рассуждение на координатной плоскости мы проинтерпретируем геометрически, что позволит по-новому взглянуть на уже хорошо знакомые нам факты.

Прием, который используется для доказательства, называется *алгоритм сплющивания носов* (в литературе обычно упоминается обратный *алгоритм вытягивания носов*, он в некотором смысле обратен тому алгоритму, который мы будем использовать сейчас). Рассмотрим произвольный параллелограмм, не содержащий узлов. Если попробовать нарисовать такой параллелограмм на листке бумаги, то видно, что он должен быть очень узким. Идея заключается в том, чтобы «сплющить» этот параллелограмм в более простой (и, как следствие, более «широкий»), сохраняя его площадь. Попробуем сделать это.

Без ограничения общности можно считать, что одна из вершин нашего параллелограмма находится в начале координат (этого всегда можно добиться параллельным переносом). Рассмотрим сторону параллелограмма, выходящую из начала координат и вершину (a, b) этой стороны, отличную от начала координат. Мы предъявим цепочку преобразований, которые охраняют площади и композиция которых переведет эту вершину в вершину $(1, 0)$ (это и означает «сплющивание параллелограмма»). На самом деле такая цепочка преобразований нам уже известна и встречалась в курсе теории чисел. Речь идет об *алгоритме Евклида* поиска наибольшего общего делителя.

Напомним, что алгоритм Евклида поиска НОД(a, b) заключается в последовательных делениях с остатком. Сначала мы делим a на b с остатком: $a = bq + r$, где $0 \leqslant r < |b|$. Затем делим b на r с остатком, и т.д. Когда при делении очередных двух чисел остаток окажется равным нулю, алгоритм Евклида завершится. Теперь нам потребуется геометрическая версия этого алгоритма.

Рассмотрим преобразование $\mathcal{A}$ плоскости $\mathbb{R}^2$, которое каждой точке с координатами (x, y) ставит в соответствие точку с координатами $(y, x - qy)$ (напомним, что число q — это частное от деления a на b). В математической нотации это записывается так:

$$\mathcal{A} : \mathbb{R}^2 \to \mathbb{R}^2, \quad \mathcal{A} : (x, y) \mapsto (y, x - qy).$$

1. **(a)** Докажите, что преобразование $\mathcal{A}$ сохраняет начало координат и является линейным, т.е. $\mathcal{A}(\overline{v}_1 + \overline{v}_2) = \mathcal{A}\overline{v}_1 + \mathcal{A}\overline{v}_2$ для любых векторов $\overline{v}_1$ и $\overline{v}_2$.
   (b) Докажите, что преобразование $\mathcal{A}$ переводит прямые в прямые.
   (c) Найдите *инвариантную прямую* преобразования $\mathcal{A}$, т.е. такую прямую ℓ, что $\mathcal{A}\ell = \ell$.
   (d) Нарисуйте образ фундаментального параллелограмма, построенного на векторах $(8, 3)$ и $(5, 2)$.
   (e) Найдите обратное преобразование $\mathcal{A}^{-1}$.

2. Докажите, что преобразование $\mathcal{A}$ сохраняет площади фигур (такие преобразования называются *симплектическими*).

3. Используя преобразование $\mathcal{A}$, докажите теорему о фундаментальном параллелограмме.

4. Докажите, что если площадь целого параллелограмма равна 1, то он является фундаментальным.

5. Докажите, что целочисленные векторы образуют базис решетки $\mathbb{Z}^2$ если и только если построенный на них параллелограмм является фундаментальным.

6. **(a)** Рассматривается целый треугольник, внутри которого есть ровно одна целая точка, а на границе нет ни одной целой точки кроме вершин. Докажите, что целая точка внутри него — это точка пересечения медиан.
   (b) Рассматривается целый треугольник, внутри которого есть ровно две целые точки (при этом целые точки могут лежать и на сторонах). Докажите, что прямая, проходящая через эти точки, либо проходит через вершину треугольника, либо параллельна одной из его сторон.

7. **Ряды Фарея.** Зафиксируем произвольное натуральное число n и рассмотрим все несократимые дроби a/b, такие, что $0 \leqslant a \leqslant b \leqslant n$. Расположим их в порядке возрастания. Получившийся набор дробей и называется *рядом Фарея*. Например, для $n = 6$ он выглядит так:

$$\frac{0}{1}, \ \frac{1}{6}, \ \frac{1}{5}, \ \frac{1}{4}, \ \frac{1}{3}, \ \frac{2}{5}, \ \frac{1}{2}, \ \frac{3}{5}, \ \frac{2}{3}, \ \frac{3}{4}, \ \frac{4}{5}, \ \frac{5}{6}, \ \frac{1}{1}.$$

   Пусть $a/b < c/d$ — соседние дроби в ряду Фарея. Докажите, что $|ad - bc| = 1$.

Подходящие дроби и приближения

Ранее мы уже отмечали, что одной из мотивировок рассмотрения цепных дробей является тот факт, что если оборвать эти дроби, мы получим очень точное приближение числа, которое мы раскладываем в цепную дробь. Формализуем это наблюдение.

1. Докажите, что
$$\frac{A_0}{B_0} < \frac{A_2}{B_2} < \ldots \leqslant \alpha \leqslant \ldots < \frac{A_3}{B_3} < \frac{A_1}{B_1},$$
причем центральные неравенства строгие, если α иррационально.

2. Докажите, что $\left|\alpha - \frac{A_m}{B_m}\right| < \left|\alpha - \frac{A_{m+1}}{B_{m+1}}\right|$. (Иначе говоря, каждая подходящая дробь приближает число α точнее, чем предыдущая.)

3. **Теорема Дирихле о приближениях.** Пусть A_m/B_m — подходящая дробь иррационального числа α. Докажите, что $\left|\alpha - \frac{A_m}{B_m}\right| < \frac{1}{B_m^2}$.

4. Докажите, что для каждого иррационального числа α найдется бесконечно много дробей x/y, для которых выполнено неравенство $\left|\alpha - \frac{x}{y}\right| < \frac{1}{2y^2}$.

5. **(a)** Докажите, что для числа $\alpha = \sqrt{2}$ неравенство $\left|\alpha - \frac{x}{y}\right| < \frac{1}{y^3}$ имеет лишь конечное число решений в натуральных числах.

   (b) Докажите, что для числа $\alpha = \frac{\sqrt{5}+1}{2}$ и для любого $k > \sqrt{5}$ неравенство $\left|\alpha - \frac{x}{y}\right| < \frac{1}{ky^2}$ имеет лишь конечное число решений в натуральных числах.

6. Докажите, что существует такая ограниченная последовательность положительных вещественных чисел $\{x_n\}$, что для любых n и m выполнено неравенство $|x_n - x_m| > 1/|n-m|$.

Теорема Лагранжа – 1

Мы выяснили, что любое действительное число α помимо стандартной десятичной записи обладает еще одним представлением в виде цепной дроби. Для рациональных чисел α цепная дробь будет конечной, а для иррациональных — бесконечной. В прошлых разделах мы много внимания уделяли свойствам *подходящих дробей* (т.е. цепных дробей, оборванных после какого-то неполного частного), однако практически ничего не знаем о бесконечных цепных дробях в целом.

На самом деле про структуру бесконечных цепных дробей известно не так уж и много. Фактически на сегодняшний день известно только два значимых и ярких результата. Первых из них называется теоремой Лагранжа и будет рассмотрен в этом разделе, а второй — теоремой Гаусса–Кузьмина, и к нему мы перейдем в следующем разделе.

Начнем с того, что проведем следующую аналогию между десятичными дробями и цепными. Какие десятичные дроби устроены наиболее просто? Конечные дроби рассматривать не будем, они слишком просты. Более сложными являются *периодические десятичные дроби*, т.е. дроби вида

$$a_0{,}a_1a_2\ldots a_k(b_1b_2\ldots b_n).$$

Несложно доказать, что периодические десятичные дроби соответствуют рациональным числам, и наоборот, каждое рациональное число порождает периодическую десятичную дробь (возможно, с нулевым периодом).

Возникает естественный вопрос: а каким числам соответствуют *периодические цепные дроби*? Ответ на этот вопрос и дает теорема Лагранжа.

Перед тем как переходить к этой теореме, проведем несколько экспериментов. А именно, разложим несколько иррациональных чисел в цепные дроби. Начнем с числа $\sqrt{2}$. Как мы помним, разложение в цепную дробь основано на двух операциях: выделении целой части и перевороте дробно части. Для того, чтобы можно было выделить целую часть из перевернутой дробной, удобно воспользоваться домножением на сопряженное:

$$\sqrt{2} = 1 + \cfrac{1}{\cfrac{1}{\sqrt{2}-1}} = 1 + \cfrac{1}{\sqrt{2}+1} = 1 + \cfrac{1}{2 + \cfrac{1}{\cfrac{1}{\sqrt{2}-1}}}.$$

Как видно, на втором и третьем шаге получаются одинаковые дробные части. Значит, неполные частные для них совпадают, и мы можем написать окончательный ответ:

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \ldots}}}.$$

Как видно, цепная дробь оказалась *периодической*, т.е. ее неполные частные, начиная с первого, повторяются! Это тем более удивительно, что *десятичная* запись числа $\sqrt{2}$ не обладает никакими хорошими свойствами:

$$\sqrt{2} = 1{,}41421356\ldots$$

Для удобства такую периодическую цепную дробь записывают так: $\sqrt{2} = [1;\overline{2}]$: точка с запятой отделяет целую часть числа от дробной, а черта ставится над периодом.

1. Разложите в периодические цепные дроби следующие числа: **(a)** $\sqrt{3}$; **(b)** $\sqrt{26}$; **(c)** $\dfrac{1+\sqrt{5}}{2}$; **(d)** $\dfrac{3+\sqrt{5}}{2}$.

2. Докажите, что **(a)** $\sqrt{n^2+1} = [n;\overline{2n}]$; **(b)** $\sqrt{n^2-1} = [n-1;\overline{1,2(n-1)}]$.

Естественно попробовать разложить в цепные дроби и другие числа вида $\sqrt{N}$, где $N \in \mathbb{N}$ — не точный квадрат. Если это сделать, то получается следующая таблица.

N	цепная дробь	N	цепная дробь	N	цепная дробь
2	$[1;\overline{2}]$	14	$[3;\overline{1,2,1,6}]$	26	$[5;\overline{10}]$
3	$[1;\overline{1,2}]$	15	$[3;\overline{1,6}]$	27	$[5;\overline{5,10}]$
5	$[2;\overline{4}]$	17	$[4;\overline{8}]$	28	$[5;\overline{3,2,3,10}]$
6	$[2;\overline{2,4}]$	18	$[4;\overline{4,8}]$	29	$[5;\overline{2,1,1,2,10}]$
7	$[2;\overline{1,1,1,4}]$	19	$[4;\overline{2,1,3,1,2,8}]$	30	$[5;\overline{2,10}]$
8	$[2;\overline{1,4}]$	20	$[4;\overline{2,8}]$	31	$[5;\overline{1,1,3,5,3,1,1,10}]$
10	$[3;\overline{6}]$	21	$[4;\overline{1,1,2,1,1,8}]$	32	$[5;\overline{1,1,1,10}]$
11	$[3;\overline{3,6}]$	22	$[4;\overline{1,2,4,2,1,8}]$	33	$[5;\overline{1,2,1,10}]$
12	$[3;\overline{2,6}]$	23	$[4;\overline{1,3,1,8}]$	34	$[5;\overline{1,4,1,10}]$
13	$[3;\overline{1,1,1,1,6}]$	24	$[4;\overline{1,8}]$	35	$[5;\overline{1,10}]$

Глядя на эту таблицу, можно обнаружить множество замечательных свойств периодических разложений. Перечислим лишь некоторые из них.

- Цепные дроби для $\sqrt{N}$ периодические, причем период начинается сразу после целой части.
- Последнее число периода равно удвоенной целой части.
- Остальной период симметричен, т.е. последнее число периода равно первому, предпоследнее второму, и т.д.

Эти факты мы докажем как следствие из основной теоремы о периодических цепных дробях.

Теорема Лагранжа. Число α раскладывается в периодическую цепную дробь если и только если оно является квадратичной иррациональностью, т.е. иррациональным корнем уравнения вида $ax^2 + bx + c = 0$, где $a, b, c \in \mathbb{Z}$ — произвольные целые числа.

Мы дадим сразу два доказательства этой теоремы — алгебраическое (из которого, в частности, будут следовать описанные выше закономерности для цепных дробей $\sqrt{N}$) и геометрическое (которое прояснит суть явления и позволит обобщить понятие цепной дроби в размерности выше 2). Однако перед тем как переходить к доказательствам, мы сейчас разберем простую часть этой теоремы. А именно, мы докажем, что если цепная дробь для числа α периодична, то α является корнем некоторого квадратного уравнения с целыми коэффициентами.

Начнем со случая, когда число α имеет вид $\alpha = [\overline{q_0; q_1, \dots, q_n}]$, т.е. период начинается сразу с цело части. В это случае число α удовлетворяет следующему уравнению:

$$\alpha = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \dots + \cfrac{1}{q_n + \cfrac{1}{\alpha}}}}.$$

Это соотношение можно переписать в виде $\alpha = \dfrac{\alpha A_n + A_{n-1}}{\alpha B_n + B_{n-1}}$ (здесь мы воспользовались рекуррентными формулами (**??**); для того, чтобы их применить, достаточно положить $q_{n+1} = \alpha$; напомним, что при доказательстве этих рекуррентных формул мы не пользовались целочисленностью величин q_i). Таким образом, число α является корнем квадратного уравнения $B_n\alpha^2 - (A_n - B_{n-1})\alpha - A_{n-1} = 0$.

Посчитаем, например, число $\alpha = [\overline{1; 2, 3, 4}]$. Вычислим последовательно величины A_k и B_k:

$$A_0 = 1, \quad A_1 = 3, \quad A_2 = 10, \quad A_3 = 43,$$
$$B_0 = 1, \quad B_1 = 2, \quad B_2 = 7, \quad B_3 = 30.$$

Таким образом, мы получаем уравнение $\alpha = \dfrac{43\alpha + 10}{30\alpha + 7}$, решая которое, находим $\alpha = \frac{9+\sqrt{96}}{15}$.

Общий случай легко сводится к уже рассмотренному. В самом деле, достаточно сначала записать периодическую часть в виде квадратичной иррациональности, а затем преобразовать получившуюся конечную цепную дробь.

3. Вычислите явно следующие числа: **(a)** $[\overline{4; 3}]$, **(b)** $[\overline{4; 3, 2}]$, **(c)** $[4; \overline{3}]$, **(d)** $[4; \overline{3, 2}]$.

В заключение мы вернемся к чисто периодическому случаю. Заметим, что возникающее квадратное уравнение имеет два решения, одно из которых отрицательно и, очевидно, нам не подходит. Однако это отрицательное решение будет играть существенную роль в ходе доказательства теоремы Лагранжа, поэтому остановимся на нем подробнее.

Пусть α — положительный корень уравнения $B_n\alpha^2 - (A_n - B_{n-1})\alpha - A_{n-1} = 0$. Оказывается, что полезно рассматривать и второй корень α' этого уравнения (он называется *алгебраически сопряженным с* α). И вот почему.

Предложение. Если $\alpha = [\overline{q_0; q_1, \dots, q_n}]$, то $\alpha' = -1/[\overline{q_n; q_{n-1}, \dots, q_0}]$. Кроме того, $(\alpha')' = \alpha$.

Доказательство. Рассмотрим число $\beta = [\overline{q_n, q_{n-1}, \dots, q_0}]$. Докажем, что число β удовлетворяет уравнению $\beta = \frac{\beta A_n + B_n}{\beta A_{n-1} + B_{n-1}}$. Если нам это удастся, то тогда соответствующее квадратное уравнение для β имеет вид $A_{n-1}\beta^2 - (A_n - B_{n-1})\beta - B_n = 0$, откуда получаем, что $A_{n-1}(-1/\beta)^2 - (B_{n-1} - A_n)(-1/\beta) - B_n = 0$. А это и означает, что $\alpha' = -1/\beta$ (т.е. $-1/\beta < 0 < \alpha$).

Посчитаем числитель цепной дроби для числа

$$\beta = q_n + \cfrac{1}{q_{n-1} + \cfrac{1}{\dots + \cfrac{1}{q_0 + \cfrac{1}{\beta}}}}.$$

Применяя рекуррентные формулы для чисел A_m и B_m и пользуясь симметричностью скобки $[q_n, q_{n-1}, \dots, q_0]$, имеем:

$$A_{n+1} = \beta[q_n, \dots, q_0] + [q_n, \dots, q_1] = \beta[q_0, \dots, q_n] + [q_1, \dots, q_n] = \beta A_n + B_n.$$

Для знаменателя вычисления аналогичны. Тем самым наше предложение доказано.

Таким образом, число α' получается из числа α обращением его периода. Такое странное на первый взгляд действие понадобится нам в следующем разделе при доказательстве периодичности квадратичной иррациональности. Еще одним удивительным обстоятельством является следующий факт. Ясно, что если $\alpha > 1$, то $-1 < \alpha' < 0$. Оказывается, что это нехитрое свойство сопряженных чисел полностью описывает те квадратичные иррациональности, для которых цепная дробь *чисто периодична*, т.е. период ее начинается с целой части. Поэтому зафиксируем следующее определение.

Определение. Квадратичная иррациональность α называется *приведенной*, если $-1 < \alpha' < 0$.

Теорема Лагранжа – 2

Именно в доказательстве этого последнего утверждения и состоит основная конструкция доказательства теоремы Лагранжа. Поэтому сформулируем это утверждение в виде отдельной теоремы.

Теорема Галуа. Пусть $\alpha > 1$ — приведенная квадратичная иррациональность. Тогда цепная дробь для α чисто периодическая.

Замечание. Эварист Галуа — один из самых талантливых французских математиков (наряду с Анри Пуанкаре), в своих работах фактически создавший современную алгебру и погибший на дуэли в возрасте 20 лет. Галуа жил впервой половине XIX века, в то время как Жозеф Луи Лагранж, чью теорему мы хотим доказать — математик XVIII века. Каким образом Лагранж доказывал свою теорему, если он не знал теоремы Галуа? По-видимому, его рассуждения были более общими и в конечном счете могли бы привести и к доказательству теоремы Галуа. Однако Лагранж не выделил этого доказательства и, возможно, даже не обратил внимания на данный факт. Поэтому сейчас этот результат носит имя Галуа.

Доказательство. Идея доказательства теоремы Галуа формулируется достаточно просто: для доказательства периодичности необходимо организовать процесс (более сложный, нежели простое вычисление неполных частных) и доказать, что он *зациклится*. В теории чисел это обычно происходит из-за *конечности* множества возможных значений величин, получаемых в ходе этого процесса. Однако техническая реализация этой идеи достаточно громоздка и нетривиальна (как, впрочем, и подавляющее большинство других результатов в теории чисел).

Итак, рассмотрим приведенную квадратичную иррациональность α, являющуюся решением квадратного уравнения $a\alpha^2 + b\alpha + c = 0$ с целыми коэффициентами. Тогда, решая это уравнение, мы получаем

$$\alpha = \frac{P+\sqrt{D}}{Q} \quad \text{и} \quad \alpha' = \frac{P-\sqrt{D}}{Q},$$

где P, D и Q — некоторые целые числа. Наша ближайшая цель — с помощью прямых вычислений кое-что узнать про эти числа.

1. Докажите, что числа P, Q и D обладают следующими свойствами:
- **(a)** $P^2 - D \,\vdots\, Q$,
- **(b)** $P, Q > 0$,
- **(c)** $P < \sqrt{D}$,
- **(d)** $Q < 2\sqrt{D}$.

2. Пусть $\alpha = q_0 + \dfrac{1}{\alpha_1}$, где $q_0 = [\alpha] \in \mathbb{N}$ и $\alpha_1 > 1$. Докажите, что если α — приведенная квадратичная иррациональность, то и α_1 — тоже приведенная квадратичная иррациональность, причем если $\alpha = \dfrac{P+\sqrt{D}}{Q}$, то $\alpha_1 = \dfrac{P_1+\sqrt{D}}{Q_1}$.

3. Для всех $n \geqslant 0$ положим $\alpha_n = q_n + \dfrac{1}{\alpha_{n+1}}$, где $q_n = [\alpha_n]$ и $\alpha_0 = \alpha$. Докажите, что последовательность $\alpha_0, \alpha_1, \alpha_2, \ldots$является периодической.

4. Докажите, что последовательность $\alpha_0, \alpha_1, \alpha_2, \ldots$является чисто периодической. (*Указание*: рассмотрите последовательность $\beta_n = -\dfrac{1}{\alpha_n'}$.)

Перед тем как переходить к теореме Лагранжа (из теоремы Галуа она следует буквально в одно действие), мы сначала обоснуем те гипотезы, которые мы сформулировали про цепные дроби чисел вида $\sqrt{N}$, поскольку цепные дроби этих чисел очень похожи на чисто периодические (и, прямо скажем, легко в них превращаются).

Следствие. Имеет место равенство $\sqrt{N} = [q_0; \overline{q_1, q_2, \ldots, q_2, q_1, 2q_0}]$.

Доказательство. Идея заключается в том, чтобы воспользоваться предложением и развернуть период чисел α_n в обратную сторону. Но для начала нам необходимо свести иррациональность $\sqrt{N}$ к приведенной, чтобы воспользоваться теоремой Галуа. Для этого рассмотрим число $\alpha = \sqrt{N} + q_0$. Несложно показать, что оно уже является приведенной квадратичной иррациональностью, т.к. $\alpha' = q_0 - \sqrt{N} \in (-1; 0)$. Поэтому по теореме Галуа $\alpha = [\overline{2q_0; q_1, q_2, \ldots, q_n}]$.

Теперь воспользуемся предложением и рассмотрим число $[\overline{q_n, \ldots, q_1, 2q_0}] = -\dfrac{1}{\alpha'}$. Однако с другой стороны

$$-\frac{1}{\alpha'} = \frac{1}{\sqrt{N} - q_0} = [\overline{q_1, q_2, \ldots, q_n, 2q_0}].$$

Сравнивая два получившихся разложения числа $-\dfrac{1}{\alpha'}$ в цепную дробь, получаем, что $q_n = q_1$, $q_{n-1} = q_2$, и т.д., что и требовалось доказать.

Теперь мы готовы доказать саму теорему Лагранжа. Напомним, что нам осталось доказать следующее утверждение: *каждая квадратичная иррациональность раскладывается в периодическую цепную дробь.*

Доказательство теоремы Лагранда. Идея доказательства состоит в том, чтобы показать, что число α_{n+1} будет приведенной квадратичной иррациональностью для достаточно большого n. Тогда цепная дробь для числа α_{n+1} будет чисто периодической по теореме Галуа, а цепная дробь для исходного числа α будет отличаться от нее конечным набором неполных частных, стоящих перед периодом.

Возьмем квадратичную иррациональность и запишем ее в виде $\alpha = \dfrac{\alpha_{n+1}A_n + A_{n-1}}{\alpha_{n+1}B_n + B_{n-1}}$ (это не что иное как формулы (**??**)). Теперь заметим, что и α, и α_{n+1} являются квадратичными иррациональностями, а числа A_n, B_n, A_{n-1} и B_{n-1} — целые. Отсюда следует, что аналогичное соотношение выполняется для сопряженных иррациональностей (подробно мы это уже доказывали в ходе доказательства теоремы Галуа):

$$\alpha' = \frac{\alpha'_{n+1}A_n + A_{n-1}}{\alpha'_{n+1}B_n + B_{n-1}} \quad \Rightarrow \quad \alpha'_{n+1} = -\frac{B_{n-1}\alpha' - A_{n-1}}{B_n\alpha' - A_n} = -\frac{B_{n-1}}{B_n} \cdot \frac{\alpha' - \frac{A_{n-1}}{B_{n-1}}}{\alpha' - \frac{A_n}{B_n}}.$$

Устремим теперь n к бесконечности. Как мы помним, дроби $\frac{A_{n-1}}{B_{n-1}}$ и $\frac{A_n}{B_n}$ стремятся (и очень быстро) к α, поэтому вторая дробь в последнем равенстве стремится к 1. Более того, вспоминая, что соседние дроби $\frac{A_{n-1}}{B_{n-1}}$ и $\frac{A_n}{B_n}$ приближают число α с разных сторон (если одна из них больше α, то вторая обязательно меньше), то можно выбрать достаточно большое n, при котором вторая дробь будет меньше 1. Учитывая, что $B_n > B_{n-1} > 0$, получаем, что $-1 < \alpha'_{n+1} < 0$, т.е. число α_{n+1} является приведенной квадратичной иррациональностью. Осталось применить для нее теорему Галуа. Тем самым теорема Лагранжа доказана.

В заключение отметим, что до сих пор не существует удовлетворительного исследования цепных дробей для иррациональных чисел, отличных от квадратичных иррациональностей. Ранее мы уже выписывали цепную дробь для числа π:

$$\pi = [3; 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 2, 1, \ldots].$$

До сих пор неизвестна структура неполных частных для этого числа. Также неизвестна структура неполных частных для числа $\sqrt[3]{2} = [1; 3, 5, 1, 1, 4, \ldots]$, хотя некоторой надеждой на возможность изучения кубических иррациональностей является *теория многомерных цепных дробей*, о которой мы поговорим ниже.

Пожалуй, единственным интересным исключением здесь является число $e = 1 + 1/1! + 1/2! + 1/3! + \ldots$, с которым мы встречались при работе с производящими функциями. А именно, Леонард Эйлер показал, что

$$e = [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots],$$

т.е. цепная дробь «почти периодична»: повторяется кусок $(1, 1, 2n)$, правда, с возрастающими n. Еще одним примером являются также связанные с e числа

$$\frac{e^{2/k} - 1}{e^{2/k} + 1} = [0; k, 3k, 5k, 7k, 9k, \ldots].$$

Кроме того, имеет место равенство

$$\operatorname{tg} 1 = [1; 1, 1, 3, 1, 5, 1, 7, 1, 9, \ldots].$$

Но описания достаточно широкого класса чисел, для которых цепные дроби обладают подобными закономерностями, до сих пор неизвестно.

Теорема Лагранжа – 3

Алгебраическое доказательство теоремы Лагранжа, данное нами в предыдущем листке, очень типично для теоретико-числовых результатов: рассуждения длинные, технические и…совершенно непонятные! Да, они позволили нам доказать несколько красивых фактов (теорему Галуа и теорему о цепной дроби для чисел $\sqrt{N}$), но сама природа периодичности пока что остается загадкой! В этом разделе мы дадим другое, геометрическое доказательство теоремы Лагранжа, которое, хоть и не позволит нам вывести из него много следствий, но зато объяснит природу возникновения периодичности квадратичных иррациональностей.

Перед тем, как переходить к самому доказательству, мы предварительно продемонстрируем две основные его идеи на более простых фактах: *малой теореме Ферма* и *теореме о фундаментальном параллелограмме.*

Малая теорема Ферма

Напомним, что малая теорема Ферма утверждает следующее: *для любого простого p и $a \neq 0$, $a \in \mathbb{Z}_p$ имеет место сравнение $a^{p-1} \equiv_p 1$.* Как видно, эта теорема фактически тоже утверждает нечто про периодичность. Здесь имеется в виду периодичность геометрической прогрессии $\{a^k\}$ по модулю простого p (более того, мы даже получаем информацию о длине этого периода: он является делителем числа $p-1$). Сейчас мы дадим геометрическое доказательство этой теоремы.

Для этого организуем все ненулевые остатки по модулю p в прямоугольную таблицу. Сделаем мы это следующим образом. Рассмотрим всевозможные ненулевые остатки по модулю p и посмотрим, в какие элементы они переходят при умножении на $a \neq 0$. Начнем с самого простого остатка, равного единице. Получаем цепочку

$$1 \to a \to a^2 \to \ldots \to a^{T-1} \to a^T = 1.$$

Если $T = p-1$, то теорема доказана. В противном случае выберем ненулевой элемент $a_1 \in Z_p$, которого нет в этой цепочке, и рассмотрим последовательность

$$a_1 \to aa_1 \to a^2a_1 \to \ldots \to a^{T-1}a_1 \to a^Ta_1 = a_1.$$

Ее длина, очевидно, равна T, поскольку умножение на a_1 переводит первую последовательность во вторую, а деление на a_1 — вторую последовательность в первую. Продолжим этот процесс, пока не исчерпаем все ненулевые элементы кольца $\mathbb{Z}_p$. Получаем следующую таблицу:

1	a	a^2	……	a^{T-1}
a_1	aa_1	a^2a_1	……	$a^{T-1}a_1$
…	…	…	……	…
a_n	aa_n	a^2a_n	……	$a^{T-1}a_n$

В этой таблице $p-1$ элементов. В каждой строке T элементов, так как любая строка получается из первой умножением на a_i. Поэтому таблица является *прямоугольной* размера $T \times n$. Откуда следует, что

$$a^T \equiv 1 \pmod{p} \quad \Rightarrow \quad a^{T\times} = a^{p-1} \equiv_p 1.$$

Например, при $a = 4$, $p = 11$ наша таблица выглядит так:

1	4	5	9	3
2	8	10	7	6

Таким образом, малая теорема Ферма (а вместе с ней и периодичность геометрических прогрессий по простому модулю) является *геометрическим* фактом: она следует из прямоугольности таблиц остатков (являющихся диаграммами Юнга перестановок, соответствующих линейной функции $f(x) = ax$). А эта прямоугольность в свою очередь следует из существования *преобразования* (умножения на подходящий элемент), переводящего строки этой таблицы друг в друга, что влечет равенство длин этих строк.

Цепные дроби

Идея геометрического доказательства периодичности неполных частных цепных дробей в некоторой степени аналогична: нужно найти преобразование, которое будет сдвигать неполные частные и тем самым породит период. Оказывается, что для его нахождения нужно использовать геометрическую интерпретацию цепной дроби в виде паруса, т.е. выпуклой оболочки множества целых точек, лежащих выше или ниже прямой $y = \alpha x$. Но откуда взять преобразование, которое будет «сдвигать этот парус по себе»? Для этого нам потребуется вернуться назад в процессе построения цепной дроби и начать не с числа α и не с прямой, которая им задается, а с...того-самого преобразования, которое и породит период!

При доказательстве теоремы о фундаментальном параллелограмме мы впервые столкнулись с *линейным оператором* $\mathcal{A} : (x, y) \to (y, x - qy)$. Этот оператор являлся преобразованием евклидовой плоскости, отличным от движения, подобия или инверсии, но тем не менее также обладающим хорошими геометрическими свойствами. Например, этот оператор переводил прямые в прямые (что для нас сейчас будет особенно важно), а также сохранял площади (это нам сейчас не понадобится). Можно рассмотреть и другие линейные операторы; в общем виде они имеют вид $\mathcal{A} : (x, y) \to (ax + by, cx + dy)$. Мы будем рассматривать только линейные операторы, коэффициенты которых, во-первых, являются целыми числами, а во-вторых, удовлетворяют соотношению $ad - bc = \pm 1$, т.е. они также будут сохранять площади. На самом деле нам будет существенно не сохранение площади, а сохранение *целочисленной решетки* $\mathbb{Z}^2$, на которой и строится цепная дробь согласно алгоритму вытягивания носов. Согласно теореме о фундаментальном параллелограмме, это соотношение является необходимым и достаточным для того, чтобы оператор $\mathcal{A}$ биективно переводил бы решетку $\mathbb{Z}^2$ в себя.

Оказывается, с каждым линейным оператором (ну, почти с каждым...) можно связать две прямые, для которых в свою очередь можно построить паруса. Т.е. начинать нужно не с прямой, а с оператора!

Ну что ж, начнем. Зафиксируем некоторые линейный оператор, например, оператор $\mathcal{A}: (x, y) \to (x + 2y, x + y)$. Попробуем найти для этого оператора *неподвижную прямую*, т.е. такую прямую ℓ, что $\mathcal{A}\ell = \ell$ (такая прямая называется *собственной прямой оператора* $\mathcal{A}$). Для этого рассмотрим матрицу $A = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$ линейного оператора $\mathcal{A}$ и найдем его *собственный вектор* $\overline{v} = \binom{1}{\alpha}$, т.е. такой вектор, что $\mathcal{A}\overline{v} = \lambda\overline{v}$ (напомним, что *собственный вектор* — это такой вектор, на котором оператор действует как гомотетия).

Для этого запишем уравнение $(A - \lambda E)\overline{v} = \overline{0}$ и приравняем к нуля определитель матрицы $A - \lambda E$: получается уравнение $(1 - \lambda)^2 - 1 \cdot 2 = 0$, откуда находим $\lambda_{1,2} = 1 \pm \sqrt{2}$ и $\alpha_{1,2} = \mp 1/\sqrt{2}$. Таким образом, у нашего линейного оператора $\mathcal{A}$ есть две инвариантные прямые $\ell_i = \{y = \alpha_i x\}$.

Замечание. Обратите внимание, что $\det A = -1 = \lambda_1 = \lambda_2$. Вообще, произведение собственных чисел матрицы A равно ее определителю. Это простое следствие теоремы Виета.

Итак, под действием нашего линейного оператора $\mathcal{A}$ одна прямая растягивается, а другая сжимается, причем коэффициенты растяжения и сжатия одинаковы по модулю, потому что $\mathcal{A}$ сохраняет площадь. Такое преобразование называется *гиперболическим поворотом*, и вот почему. Введем новую систему координат, взяв одну из этих прямых за ось Ou, а другую — за ось Ov. Тогда линейный оператор $\mathcal{A}$ сохраняет *гиперболы*, заданные в этой системе координат уравнением $uv = \text{const}$ при разных значениях константы, потому что одна из координат u и v увеличивается во сколько-то раз, а другая во столько же раз уменьшается (при этом гипербола $uv = c$ переходит в гиперболу $uv = -c$).

Возьмем теперь множества всех целых точек, которые расположены в положительном квадранте выше и ниже прямой ℓ_1 (поскольку угловой коэффициент этой прямой иррационален, на самой прямой ℓ_1 целых точек, кроме начала координат, нет) и рассмотрим их выпуклые оболочки $\mathcal{P}_+$ и $\mathcal{P}_-$. При преобразовании $\mathcal{A}$ целые точки, расположенные выше прямой ℓ_1, переходят в целые точки, расположенные ниже этой прямой, и наоборот. Поэтому $\mathcal{A}(\mathcal{P}_+) \subseteq \mathcal{P}_-$ и $\mathcal{A}(\mathcal{P}_-) \subseteq \mathcal{P}_+$. Отсюда следует, что геометрические характеристики наших парусов $\mathcal{P}_\pm$ образуют периодические последовательности, поскольку они переводятся в себя оператором $\mathcal{A}$. В частности, так называемые *целочисленные длины* звеньев парусов (т.е. количество отрезков, на которые звено делится лежащими на нем целыми точками) периодичны. С другой стороны, эти целочисленные длины как раз и являются неполными частными. Вот и все! Вот откуда берется периодичность цепных дробей квадратичных иррациональностей! Сдвиг неполных частных осуществляет линейный оператор, который фактически и строит нашу цепную дробь.

Заметим, что, вообще говоря, теорема Лагранжа доказана нами не для всех квадратичных иррациональностей, а лишь для тех, которые являются собственными значениями некоторого линейного оператора $\mathcal{A}: (x, y) \to (ax + by, cx + dy)$ с целочисленными коэффициентами a, b, c, d, удовлетворяющими соотношению $ad - bc = \pm 1$. Выясним, насколько широк класс таких квадратичных иррациональностей. Для этого выразим соответствующие угловые коэффициенты α собственных прямых через a, b, c, d. Записывая систему уравнений

$$\begin{cases} cx + dy = \alpha(ax + by), \\ \qquad\quad y = \alpha x, \end{cases}$$

и решая ее, находим

$$\alpha = \frac{d - a \pm \sqrt{(d-a)^2 + 4bc}}{2b} = \frac{d - a \pm \sqrt{(a+d)^2 - 4\epsilon}}{2b},$$

где $\epsilon = ad - bc = \pm 1$.

Замечание. В ходе вычислений для числа α у нас под корнем встретились два важных числа, связанных с линейным оператором $\mathcal{A}$: уже знакомый нам *определитель* $\det \mathcal{A} = ad - bc = \pm 1$ и новый для нас *след* $\operatorname{tr} \mathcal{A} = a + d$. Роль этих констант состоит в том, что они являются *инвариантами* линейного оператора (как и собственные числа) относительно линейных замен координат.

Оказывается, что иррациональности $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$, $\sqrt{6}$ удается реализовать подобным образом:

$\sqrt{2}$	$\frac{\sqrt{3^2-1}}{2}$
$\sqrt{3}$	$\sqrt{2^2-1}$
$\sqrt{5}$	$\sqrt{2^2+1}$
$\sqrt{6}$	$\frac{\sqrt{5^2-1}}{2}$

Возможно, любую квадратичную иррациональность можно реализовать подобным образом, однако я не знаю, так ли это (по-видимому, это верно). Мы сейчас даже не будем восполнять этот пробел, чтобы не отодвигать на второй план геометрическую суть периодичности, связанную с действием линейного оператора на паруса, определяемые этим оператором.

Уравнение Пелля

В заключение раздела о периодичности цепных дробей квадратичных иррациональностей мы рассмотрим один из самых ярких примеров применения этой периодичности: решение *уравнения Пелля*. Уравнением Пелля называется диофантово уравнение

$$x^2 - Ny^2 = 1, \quad \text{где } N \text{ — не точный квадрат.}$$

(Такое ограничение накладывается во избежание тривиального решения, связанного с разложением левой части на множители по разности квадратов.)

Разрозненные упоминания об уравнении Пелля встречаются на протяжении всей истории математики. Наиболее известна так называемая *задача о скоте*, предложенная древнегреческим ученым Архимедом Эратосфену. Эта задача содержит восемь неизвестных (число быков в стадах бога Юпитера), которые связаны между собой линейными соотношениями, а также двум дополнительным условиям, согласно которым некоторые числа должны быть точными квадратами. После всех алгебраических преобразований задача сводится к решению следующего уравнения Пелля: $x^2 - 4729494y^2 = 1$. Наименьшее решение этого уравнения (по-видимому, впервые найденное лишь в 1880 г. Амтором) превышает 10^{41} (!!!). Скорее всего, ни Архимед, ни кто-либо еще из древнегреческих математиков не могли обладать полным решением этой задачи, но сам факт ее постановки свидетельствует о том, что некоторые сведения об уравнении Пелля у них все же были.

Оказывается, что существование решений уравнения Пелля тесно связано с геометрическим доказательством теоремы Лагранжа о периодичности цепных дробей квадратичных иррациональностей. Напомним, что для такого доказательства мы фиксировали линейный оператор $\mathcal{A}$, матрица которого имела целые коэффициенты и определитель ± 1. Этот оператор действовал на плоскости $\mathbb{R}^2$ линейными заменами координат, сохраняя при этом целочисленную решетку $\mathbb{Z}^2$. Для такого оператора можно найти *инвариантные прямые* $\ell_{1,2} = \{y = \alpha_{1,2}x\}$, которые переходят в себя под действием этого оператора: $\mathcal{A}(\ell_{1,2}) = \ell_{1,2}$. Эти прямые разбивают всю плоскость на четыре угла. Мы рассматриваем множество целых точек, лежащих в каждом угле, и строим его *выпуклую оболочку* $\mathcal{P}^{\pm}_{1,2}$, являющуюся бесконечной ломаной, неограниченно приближающейся к прямым $\ell_{1,2}$. Тогда линейный оператор $\mathcal{A}$ сохраняет эти выпуклые оболочки, сдвигая их по себе. С другой стороны, парус несет в себе всю информацию о неполных частных цепной дроби, соответствующей числам $\alpha_{1,2}$: неполные частные равны целочисленным длинам отрезков ломаных, которые образуют парусы. Поэтому наличие периода у таких цепных дробей объясняется существованием линейного оператора, который сдвигает неполные частные, переводя один фрагмент цепной дроби в следующий.

В рамках этого доказательства остается невыясненным следующий вопрос: какие именно числа $\alpha_{1,2}$ могут быть реализованы как угловые коэффициенты инвариантных прямых? Ясно, что такие числа являются корнями некоторого квадратного уравнения, но оно имеет определенный вид, и непонятно, можно ли реализовать произвольную квадратичную иррациональность как угловой коэффициент такой инвариантной прямой.

Наше изучение уравнения Пелля начнется именно с такого вопроса. На самом деле мы докажем, что в качестве угловых коэффициентов $\alpha_{1,2}$ всегда можно реализовать числа $\pm 1/\sqrt{N}$, где число N отлично от полного квадрата.

Для этого рассмотрим на плоскости $\mathbb{R}^2$ множество точек (x, y), чьи координаты удовлетворяют уравнению Пелля $x^2 - Ny^2 = 1$. Оказывается, что это множество точек является гиперболой. Мы знаем, что стандартное уравнение гиперболы имеет вид $y = k/x$, или $xy = k$. Однако гиперболу можно задать и другим уравнением. Для этого возьмем простейшую гиперболу $xy = 1$ и *повернем ее* на угол 45° по часовой стрелке. С точки зрения координат такой поворот является *линейным оператором*, имеющим следующий вид: $\mathcal{R}: (x, y) \to ((y - x)/\sqrt{2}, (y + x)/\sqrt{2})$ (такой оператор еще называется ортогональным). Поэтому наша гипербола перейдет в гиперболу, заданную уравнением $y^2 - x^2 = 2$. Вот откуда берутся разности квадратов!

Наша гипербола имеет асимптоты, задающиеся уравнением $x^2 - Ny^2 = 0$, т.е. прямые $\ell_\pm = \{y = \pm x/\sqrt{N}\}$ (это стандартный прием, позволяющий находить асимптоты или, наоборот, строить возмущенные кривые: нужно взять несколько прямых ℓ_1, ..., ℓ_k и их объединение, заданное уравнением $\ell_1 \cdot \ldots \cdot \ell_k = 0$, а затем возмутить это уравнение, заменив число 0 в правой части на маленькое ненулевое число ε). Теперь мы хотим найти линейный оператор $\mathcal{H}_N$, который имел бы целые координаты, определитель 1 и который сохранял бы эту гиперболу.

Чтобы найти такой оператор, воспользуемся следующим соображением. С точки зрения проективной геометрии гиперболы представляет собой овал (т.е. замкнутую кривую) на проективной плоскости (этот овал пересекает бесконечно удаленную прямую в двух точках, поэтому на плоскости обычной мы видим два несвязных куска). Овал похож на окружность, для которой легко придумать геометрическое преобразование, которое переводит ее в себя — это просто поворот вокруг центра окружности. Давайте сначала научимся описывать такой поворот окружности в координатах, а затем обобщим его на случай нашей гиперболы.

начнем с поворота простейшей окружности, заданной на координатной плоскости уравнением $x^2 + y^2 = 1$. Точка (x, y) такой окружности может быть задана углом φ: $(x, y) = (\cos\varphi, \sin\varphi)$. Тогда поворот окружности — это просто изменение угла φ у каждой точки, т.е. сдвиг $\varphi \mapsto \varphi + \alpha$ для некоторого фиксированного угла α. Тогда оператор поворота $\mathcal{R}$ действует следующим образом:

$$\mathcal{R}: (x, y) = (\cos\varphi, \sin\varphi) \mapsto (x', y') = (\cos(\varphi + \alpha), \sin(\varphi + \alpha)).$$

Заметим, что $x' = \cos(\varphi + \alpha) = \cos\varphi\cos\alpha - \sin\varphi\sin\alpha = x\cos\alpha - y\sin\alpha$ и $y' = \sin(\varphi + \alpha) = \cos\varphi\sin\alpha + \sin\varphi\cos\alpha = x\sin\alpha + y\cos\alpha$. Это означает, что матрица R линейного оператора $\mathcal{R}$ имеет следующий вид: $R = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$. Таким образом, мы нашли матрицу оператора поворота, сохраняющего единичную окружность (и пучок концентрических окружностей с центром в точке $(0, 0)$).

Рассмотрим теперь поворот эллипса. Эллипс в декартовой системе координат задается уравнением $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$, поэтому его точки можно параметризовать следующим образом: $(x, y) = (a\cos\varphi, b\sin\varphi)$. Получается, что поворот эллипса задается оператором

$$\mathcal{R}' : (x, y) = (a\cos\varphi, b\sin\varphi) \mapsto (x', y') = (a\cos(\varphi+\alpha), b\sin(\varphi+\alpha)).$$

Применяя формулы косинуса суммы и синуса суммы, получаем следующую матрицу поворота эллипса: $R' = \begin{pmatrix} \cos\alpha & -(a/b)\sin\alpha \\ (b/a)\sin\alpha & \cos\alpha \end{pmatrix}$.

Перейдем к повороту гиперболы. Для начала рассмотрим случай прямоугольной гиперболы $x^2 - y^2 = 1$ (это соответствует случаю $N = 1$ в уравнении Пелля). Прежде всего нам необходимо параметризовать точки (x, y) этой гиперболы так же, как мы параметризовывали точки окружности и эллипса. Для этого необходимо использовать *гиперболический косинус* и *гиперболический синус*: $\operatorname{ch} t = \frac{a^t + e^{-t}}{2}$ и $\operatorname{sh} t = \frac{e^t - e^{-t}}{2}$. Смысл ввода таких величин заключается в следующем. Уравнение гиперболы $x^2 - y^2 = 1$ переходит в уравнение окружности $x^2 + y^2 = 1$ при замене координат $(x, y) \mapsto (x, iy)$, где $i = \sqrt{-1}$ — мнимая единица. Попробуем сделать подстановку $t = i\varphi$ в параметризации гиперболы. Вспоминая, что $e^(i\varphi) = \cos\varphi + i\sin\varphi$, сразу получаем, что $\operatorname{ch}(i\varphi) = \cos\varphi$ и $\operatorname{sh}(i\varphi) = i\sin(\varphi)$. Поэтому при такой подстановке точка гиперболы (x, y) действительно перейдет в точку окружности (x, iy).

Основное тригонометрическое тождество $\cos^2\varphi + \sin^2\varphi = 1$ в случае перехода к гиперболическим функциям принимает следующий вид: $\operatorname{ch}^2 t - \operatorname{sh}^2 t = 1$. Поэтому точка $(x, y) = (\operatorname{ch} t, \operatorname{sh} t)$ действительно лежит на гиперболе и пробегает всю ее верхнюю правую ветку.

Осталось воспроизвести вычисления, которые мы совершали при получении матрицы поворота. Сейчас нам необходимо совершить сдвиг $t \mapsto t + a$ для фиксированной константы a, и тогда мы получим оператор

$$\mathcal{H} : (x, y) = (\operatorname{ch} t, \operatorname{sh} t) \mapsto (x', y') = (\operatorname{ch}(t+a), \operatorname{sh}(t+a).$$

Поскольку $x' = \operatorname{ch}(t+a) = \operatorname{ch} t \operatorname{ch} a + \operatorname{sh} t \operatorname{sh} a = x\operatorname{ch} a + y\operatorname{sh} a$ и $y' = \operatorname{sh}(t+a) = \operatorname{ch} t \operatorname{sh} a + \operatorname{sh} t \operatorname{ch} a = x\operatorname{ch} a + y\operatorname{sh} a$, матрица оператора гиперболического поворота имеет следующий вид: $H = \begin{pmatrix} \operatorname{ch} a & \operatorname{sh} a \\ \operatorname{sh} a & \operatorname{ch} a \end{pmatrix}$.

Нам остается подправить нормировку матрицы гиперболического поворота для гиперболы $x^2 - Ny^2 = 1$. Параметризация этой гиперболы имеет вид $(x, y) = (\operatorname{ch} t, (1/\sqrt{N})\operatorname{sh} t)$. По аналогии с подправлением матрицы оператора поворота эллипса, получаем следующий оператор гиперболического поворота $\mathcal{H}_N$ с матрицей $H_N = \begin{pmatrix} \operatorname{ch} a & \sqrt{N}\operatorname{sh} a \\ (1/\sqrt{N})\operatorname{sh} a & \operatorname{ch} a \end{pmatrix}$.

Таким образом, выбрав в качестве стартового оператора оператор гиперболического поворота $\mathcal{H}$, мы получим оператор, сохраняющий прямые $y = \pm x/\sqrt{N}$. Теперь нам осталось добиться того, чтобы этот оператор сохранял бы целочисленную решетку $\mathbb{Z}^2$: в этом случае он будет в точности тем искомым линейным оператором, который будет сохранять парусы и сдвигать неполные частные цепной дроби числа $1/\sqrt{N}$, порождая период.

Обозначим для удобства $\operatorname{ch} a = x_1$ и $(1/\sqrt{N}) \operatorname{sh} a = y_1$. Тогда матрица оператора гиперболического поворота примет следующий вид: $H_N = \begin{pmatrix} x_1 & Ny_1 \\ y_1 & x_1 \end{pmatrix}$. Значит, если числа x_1 и y_1 будут целыми, мы добьемся цели. Какому условию должны удовлетворять числа x_1 и y_1? Единственное условие — выполнение соотношения $\operatorname{ch}^2 a - \operatorname{sh}^2 a = 1$, которое переписывается в виде $x_1^2 - Ny_1^2 = 1$.

Таким образом, мы приходим к следующему утверждению.

> Прямые $y = \pm x/\sqrt{N}$ являются инвариантными прямыми линейного оператора, матрица которого имеет вид $H_N = \begin{pmatrix} x_1 & Ny_1 \\ y_1 & x_1 \end{pmatrix}$, где x_1 и y_1 — целые решения уравнения Пелля $x^2 - Ny^2 = 1$.

По сути своей задача об отыскании натуральных решений уравнения Пелля сводится к задаче поиска натуральных точек на гиперболе. Легко видеть, что на гиперболе $x^2 - Ny^2 = 1$ всегда есть очевидная целая точка $(1, 0)$. Однако она соответствует неинтересному тождественному оператору $\mathcal{E}\colon\ (x, y) \mapsto (x, y)$ с единичной матрицей $E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Поэтому мы будем искать нетривиальные натуральные решения.

Начнем с того, что попробуем угадать простейшие решения этого уравнения при маленьких N. Составим следующую таблицу:

N	x	y
2	3	2
3	2	1
5	9	4
6	5	2
7	8	3
8	3	1
10	19	6
11	10	3

Какие закономерности можно увидеть, глядя на эту таблицу? Например, можно заметить, что если взять решения (x_1, y_1) и (x_2, y_2) для соседних N, то тогда $x_1y_2 - x_2y_1 = (x_1, x_2) \cdot (y_1, y_2)$ (в правой части написано произведение НОДов). Доказательства этому удивительному факту пока что нет…

А еще можно попробовать посчитать цепные дроби для величин x/y и $\sqrt{N}$ (недаром ведь мы ими занимались!). И вот какие чудеса обнаруживаются:

$\sqrt{N}$	x/y
$\sqrt{2}=[1;\overline{2}]$	$\frac{3}{2}=[1;2]$
$\sqrt{3}=[1;\overline{1,2}]$	$\frac{2}{1}=[2]$
$\sqrt{5}=[2;\overline{4}]$	$\frac{9}{4}=[2;4]$
$\sqrt{6}=[2;\overline{2,4}]$	$\frac{5}{2}=[2;2]$
$\sqrt{7}=[2;\overline{1,1,1,4}]$	$\frac{8}{3}=[2;1,2]$
$\sqrt{8}=[2;\overline{1,4}]$	$\frac{3}{1}=[3]$
$\sqrt{10}=[3;\overline{6}]$	$\frac{19}{6}=[3;6]$
$\sqrt{11}=[3;\overline{3,6}]$	$\frac{10}{3}=[3;3]$

Что можно сказать теперь? Видно, что решения уравнения Пелля получаются обрывом цепной дроби для числа $\sqrt{N}$ в подходящем месте. Сам по себе этот факт ожидаем. В самом деле, перепишем уравнение Пелля в следующем виде:

$$\left(\frac{x}{y}\right)^2 - N = \frac{1}{y^2}.$$

Видно, что дробь x/y должна быть подходящей дробью для $\sqrt{N}$, поскольку она достаточно точно приближает его. На самом деле это несложно доказать и строго. Заметим, что $x + \sqrt{N}y > x + 1 \cdot y = 2y$, поэтому

$$1 = x^2 - Ny^2 = (x-\sqrt{N}y)(x+\sqrt{N}y) > (x-\sqrt{N}y)2y \quad \Rightarrow \quad \frac{x}{y} - \sqrt{N} < \frac{1}{2y^2}.$$

Значит, для нахождения решения уравнения Пелля нужно разложить число $\sqrt{N}$ в периодическую цепную дробь и оборвать этот период в подходящем месте. Но в каком?! Ответ на этот вопрос дает следующая теорема.

Теорема. Пусть $\sqrt{N} = [q_0;\overline{q_1,q_2,\dots,q_n,2q_0}]$. Тогда (A_n,B_n) — наименьшее натуральное решение уравнения Пелля $x^2 - Ny^2 = (-1)^{n-1}$.

Доказательство. Пусть $\alpha = q_0 + \sqrt{N} = [2q_0;\overline{q_1,q_2,\dots,q_n}]$. Тогда если A_{n-1}/B_{n-1} и A_n/B_n — подходящие дроби для $\sqrt{N}$, то по рекуррентным формулам имеем

$$\sqrt{N} = \frac{\alpha A_n + A_{n-1}}{\alpha B_n + B_{n-1}}.$$

Домножая на знаменатель дроби и преобразовывая, получаем

$$\sqrt{N}(q_0B_n + B_{n-1} - A_n) = q_0A_n + A_{n-1} - NB_n.$$

Поскольку число $\sqrt{N}$ является иррациональным, то

$$q_0B_n + B_{n-1} = A_n \quad \text{и} \quad q_0A_n + A_{n-1} = NB_n.$$

Выразим отсюда A_{n-1} и B_{n-1} и подставим их в формулу $A_nB_{n-1} - A_{n-1}B_n = (-1)^{n-1}$:

$$(-1)^{n-1} = A_n(A_n - q_0B_n) - B_n(NB_n - q_0A_n) = A_n^2 - NB_n^2.$$

Таким образом, пара (A_n, B_n) действительно является решением уравнения $x^2 - Ny^2 = (-1)^{n-1}$.

Решение уравнения Пелля нами практически получено. Осталось лишь убрать неопределенность со знаком в формулировке теоремы. Если n нечетно, то справа получается 1, и тем самым уравнение Пелля решено. Если же n четно, то нужно просто продлить период на один цикл. В таком случае его новая длина станет равной $n + 1 + n = 2n + 1$, поэтому мы снова получим решение уравнения Пелля.

Возьмем, к примеру, $N = 29$. Разложение $\sqrt{29}$ в цепную дробь имеет следующий вид: $\sqrt{29} = [5; \overline{2, 1, 1, 2, 10}]$. Подходящие дроби равны

$$\frac{5}{1}, \quad \frac{11}{2}, \quad \frac{16}{3}, \quad \frac{27}{5}, \quad \frac{70}{13}.$$

Поскольку длина периода равна 4, то пара $(70, 13)$ является решением уравнения $x^2 - 29y^2 = -1$. Чтобы получить решение уравнения $x^2 - 29y^2 = 1$, нужно взять еще один цикл:

$$\frac{727}{135}, \quad \frac{1524}{283}, \quad \frac{2251}{418}, \quad \frac{3775}{701}, \quad \frac{9801}{1820}.$$

Значит, пара $(9801, 1820)$ и будет наименьшим (!) решением нашего уравнения Пелля.

Надо сказать, что уравнением Пелля связано много еще не решенных проблем. Например, до сих пор неизвестно точное описание чисел N, длина периода цепных дробей для $\sqrt{N}$ у которых нечетна.

Таким образом, мы доказали, что любое уравнение Пелля $x^2 - Ny^2 = 1$ имеет нетривиальное натуральное решение. Таким образом, для любого N, отличного от точного квадрата, существует линейный оператор $\mathcal{H}_N$, сохраняющий прямые $y = \pm x/\sqrt{N}$ и целочисленную решетку $\mathbb{Z}^2$. Таким образом, мы получаем доказательство периодичности цепных дробей всех квадратичных иррациональностей вида $1/\sqrt{N}$.

Следующий вопрос таков: а как получить все решения уравнения Пелля? Один из возможных способов — брать все новые и новые подходящие дроби для $\sqrt{N}$, обрывающиеся перед удвоенной целой частью. Однако можно поступить проще. Заметим, что если (x_1, y_1) — некоторое нетривиальное решение уравнения Пелля, то по нему можно построить линейный оператор $\mathcal{H}$, являющийся гиперболическим поворотом и имеющий матрицу $H = \begin{pmatrix} x_1 & Ny_1 \\ y_1 & x_1 \end{pmatrix}$. Заметим, что $H\binom{1}{0} = \binom{x_1}{y_1}$, т.е. оператор $\mathcal{H}$ сдвигает решение $(1, 0)$ в решение (x_1, y_1). Ясно, что применив этот оператор уже к решению (x_1, y_1), мы получим новое решение (x_2, y_2). И вообще, для любого натурального k пара $\binom{x_k}{y_k} = H^k\binom{1}{0}$ является решением уравнения Пелля $x^2 - Ny^2 = 1$. Легко описать последовательности $\{x_k\}$ и $\{y_k\}$ рекуррентно:

$$x_k = x_1 x_{k-1} + N y_1 y_{k-1}, \quad y_k = y_1 x_{k-1} + x_1 y_{k-1}.$$

Осталось доказать, что у уравнения Пелля нет других натуральных решений. Для этого предположим, что такое решение (x', y') существует. Рассмотрим пару соседних решений

(x_k, y_k) и (x_{k+1}, y_{k+1}), «зажимающую» решение (x', y'), т.е. такое, что $x_k \leqslant x' \leqslant x_{k+1}$. А теперь будем применять к этой тройке решений обратный оператор $\mathcal{H}^{-1}$. Этот оператор будет не увеличивать, а уменьшать наши решения, сдвигая их все ближе и ближе к тривиальному решению $(x_0, y_0) = (1, 0)$. Поэтому, сдвинув (x_k, y_k) до тривиального решения, а (x_{k+1}, y_{k+1}) — до ближайшего натурального (x_1, y_1), мы получаем противоречие, поскольку решение, полученное из (x', y'), будет еще меньше.

Замечание. У оператора $\mathcal{H}$ есть еще один смысл: это не что иное как оператор умножения элементов кольца $\mathbb{Z}[\sqrt{N}]$ на $x_1 + \sqrt{N}y_1$. Такой взгляд дает следующие формулы для чисел x_k, y_k:

$$x_k + \sqrt{N}y_k = (x_1 + \sqrt{N}y_1)^k.$$

Об этом мы поговорим позднее.

В заключение отметим, что задача описания целых или рациональных точек на алгебраических кривых (т.е. кривых, заданных многочленами) еще очень далека от завершения. Рациональные точки на кривых степени 3 (такие кривые называются эллиптическими) были описаны Б. Мазуром в 60-е гг. XX в., и этот результат стал одним из наиболее значимых в теории чисел за всю историю. Для кривых более высокой степени практически ничего неизвестно…

Замечание. Отметим, впрочем, теорему Фалтингса, утверждающую, что на алгебраических кривых рода $g > 1$ существует лишь конечное число целых точек. Отсюда, в частности, следует «слабая версия» Великой теоремы Ферма: уравнение $x^n + y^n = z^n$ при $n \geqslant 3$ имеет лишь конечное (возможно, нулевое) число взаимно простых решений в натуральных числах. Тот факт, что это уравнение на самом деле не имеет натуральных решений, был доказан Э. Уайлсом в 1995 г. (некоторые результаты были получены им в соавторстве с Р. Тейлором).

Еще об уравнении Пелля – 2

Определение. *Уравнением Пелля* называется диофантово уравнение $x^2 - dy^2 = 1$, где d — не точный квадрат. *Уравнением типа Пелля* называется диофантово уравнение $x^2 - dy^2 = r$, где d — не точный квадрат. *Минимальным решением уравнения типа Пелля* называется натуральное решение (x_1, y_1), в котором число x_1 минимально.

Наша цель — описать все решения уравнения Пелля. Кстати, вопрос существования решения (хотя бы одного) у произвольного уравнения типа Пелля до сих пор открыт. Рассмотрим множество $\mathbb{Z}[\sqrt{d}]$, состоящее из всех чисел вида $x + y\sqrt{d}$, где x, y — целые, а d — натуральное, не являющееся полным квадратом. Для каждого числа $x + y\sqrt{d}$ определим *норму* $N(x + y\sqrt{d}) = x^2 - dy^2$. Решить уравнение Пелля $\Leftrightarrow$ найти все числа с нормой 1 в кольце $\mathbb{Z}[\sqrt{d}]$.

1. **(a)** Докажите *теорему Дирихле*: для любого вещественного иррационального числа ξ и натурального N существует такое целое число a и натуральное число b, что $b \leqslant N$ и $| b\xi - a | \leqslant \dfrac{1}{N}$.
   (b) Докажите, что существует такое целое число n, что уравнение $x^2 - dy^2 = n$ имеет бесконечно много решений в целых числах. (*Указание*: возьмите $\xi = \sqrt{d}$ и оцените норму $a^2 - db^2$ с помощью теоремы Дирихле.)
   (c) Докажите, что если $x_1 \equiv_n x_2$ и $y_1 \equiv_n y_2$, где $n = N(x_2 + y_2\sqrt{d})$, то число $x_1 + y_1\sqrt{d}$ делится на число $x_2 + y_2\sqrt{d}$ в кольце $\mathbb{Z}[\sqrt{d}]$.
   (d) Докажите, что у уравнения Пелля $x^2 - dy^2 = 1$ существует решение в натуральных числах.
   (e) Пусть (a, b) — минимальное решение уравнения Пелля $x^2 - dy^2 = 1$. Докажите, что если $z = a + b\sqrt{d}$, то все натуральные решения (x_n, y_n) уравнения Пелля имеют вид $z^n = x_n + y_n\sqrt{d}$, причем других решений нет. Иначе говоря, операция $(x, y) \mapsto (ax + dby, ay + bx)$ переводит решение в следующее за ним решение. Геометрически это есть ни что иное как сдвиг гиперболы $x^2 - dy^2 = 1$ (*Указание*: обратите внимание, что сдвигаться по гиперболе можно в обе стороны.).

2. Найдите все решения в целых числах у следующих уравнений:
   (a) $x^2 - 5y^2 = 1$,
   (b) $x^2 - xy - y^2 = 1$,
   (c) $x^2 - 5y^2 = 7$,
   (d) $x^2 - 3y^2 = 13$.

3. При каких простых p уравнение $x^2 - py^2 = -1$ имеет решения?

4. Докажите, что уравнение $x^3 + y^3 + z^3 + t^3 = 1999$ имеет бесконечно много решений в целых числах.

Еще об уравнении Пелля – 3

Мысль. Уравнение Пелля возникает на практике в следующих ситуациях. Во-первых, это мощный инструмент для разных конструктивов. Во-вторых, это хороший способ разложить на множители число x^2+1 (хоть и не при всех x). В-третьих, иногда могут оказаться полезными конструкции, связанные с поиском решения уравнения Пелля: формулы $(a+b\sqrt{d})^n = x_n + y_n\sqrt{d}$, рекуррентный процесс $(x, y) \mapsto (ax + dby, ay + bx)$ и т.д.

1. Докажите, что существует бесконечно много натуральных n, таких, что $n^2 + 1 \mid n!$.

2. Найти все целые a, при которых уравнение $x^2+axy+y^2 = 1$ имеет бесконечно много целых решений.

3. Пусть $p(n)$ — наибольший простой делитель числа $n^2 + 1$.
(a) Докажите, что существует бесконечно много троек (a, b, c) натуральных чисел, таких, что $p(a) = p(b) = p(c)$.
(b) Докажите, что существует бесконечно много натуральных n таких, что у числа $n^2 + 1$ найдутся два положительных делителя, разность которых равна n.

4. Докажите, что существует бесконечно много натуральных чисел a, b, c, образующих арифметическую прогрессию, таких, что числа $ab + 1$, $bc + 1$ и $ca + 1$ являются точными квадратами.

5. Докажите, что при нечетных натуральных n выражение

$$\left[\frac{(3+2\sqrt{2})^n - (3-2\sqrt{2})^n}{4}\right]$$

является полным квадратом.

6. Докажите, что можно провести бесконечно много прямых общего положения так, чтобы для любой прямой все расстояния между точками ее пересечения с остальными прямыми были бы натуральными числами.

7. Докажите, что для любого натурального $n > 2$ найдутся единственные нечетные числа x, y такие, что $x^2 + 7y^2 = 2^n$.

8. Пусть a — натуральное число, не являющееся точным квадратом. Обозначим через A множество всех натуральных чисел вида $\dfrac{x^2 - a}{x^2 - y^2}$ при всевозможных целых x и y, где $x > \sqrt{a}$. Через B обозначим множество всех натуральных чисел, представимых в виде $\dfrac{x^2 - a}{x^2 - y^2}$ при всевозможных целых x и y, где $0 \leqslant x < \sqrt{a}$. Докажите, что $A = B$.