Вневписанные окружности и дюжины точек.

(Представляется В. Филимоновым и А. Заславским.)

О появлении этой серии задач.

Всё началось со следующей задачи, которую мне сообщил Д. Терёшин.

Задача (Д. Терёшин). Рассмотрим треугольник ABC и две его вневписанные окружности, одна из которых касается стороны AC в точке K и продолжений сторон AB и BC в точках L иM, а другая — стороны AB в точке P и продолжений сторон AC и BC в точках Q иR. Докажите, что точка пересечения X прямых LM и QR лежит на высоте (проведённой из вершины A) треугольника ABC.

Сходу геометрического решения этой задачи найти не удалось, работали лишь вычислительные способы. Некоторые наблюдения добавили интригу в этот сюжет. Оказалось, что также точка пересечения Y прямых KM и PR лежит на высоте треугольника ABC, а кроме того, длины отрезков AY и AX равны соответственно радиусу вписанной окружности и вневписанной окружности, касающейся стороны BC. Были обнаружены другие многочисленные факты, обнаружилась связь с известными трудными задачами олимпиад (некоторые из них присутствуют в серии). Желание получить чисто геометрические объяснения этих фактов побудило рассмотреть более подробно точки касания сторон с вписанной и вневписанными окружностями, и прямые, их соединяющие. Эти четыре окружности обладают различиями, связанными с геометрическим расположением: скажем, вписанная окружность всегда меньше вневписанной, вписанная расположена внутри, а вневписанная — вне треугольника. Однако, эти окружности имеют общие глубокие свойства: каждая из них касается трёх прямых, содержащих стороны треугольника, центр каждой из них лежит на пересечении трёх биссектрис углов треугольника (если под биссектрисой понимать внутреннюю либо внешнюю биссектрису), и как правило, наличие некоторого свойства у одной из окружностей влечёт наличие этого свойства (или аналога этого свойства) и у других. Поэтому вписанная и вневписанные окружности в некотором смысле равноправны по отношению к данному треугольнику, и для понимания некоторых важных геометрических фактов понадобилось одновременное рассмотрение всей четвёрки окружностей. Этим объясняется введение не вполне стандартных, но «равноправных» обозначений (см. ниже).

Разделы A, B, C этой серии появились в результате работы автора текста совместно с И. Богдановым, раздел D добавлен А. Заславским. Благодарим А. Акопяна и В. Протасова за проявленное внимание и замечания.

П. Кожевников

Обозначения.

Все рассмотрения происходят в произвольном неравнобедренном треугольнике ABC. На протяжении всей серии придерживаемся следующих обозначений для объектов, связанных с треугольником ABC.

R, p — радиусы описанной и вписанной окружности, полупериметр;

a, b, c — длины сторон BC, CA, AB;

A', B', C' — середины сторон BC, CA, AB;

AH_a, BH_b, CH_c — высоты, H — ортоцентр треугольника ABC;

Ω — описанная окружность, O — её центр;

ω_0 — вписанная окружность, I_0 — её центр; ω_1, ω_2, ω_3 — вневписанные окружности (касающиеся соответственно сторон BC, CA, AB), I_1, I_2, I_3 — их центры, r_i — радиусы окружностей ω_i;

I'_0, I'_1, I'_2, I'_3 — центры вписанной и вневписанных окружностей треугольника $A'B'C'$.

В предложенных обозначениях присутствует следующая симметрия: Заметим, что 6 прямых I_iI_j $(i \neq j)$— внешние и внутренние биссектрисы треугольника ABC. Поэтому четвёрка точек I_0, I_1, I_2, I_3 — ортоцентрическая, и треугольник ABC — ортотреугольник (то есть треугольник с вершинами в основаниях высот любого из четырёх треугольников $I_0I_1I_2$, $I_1I_2I_3$, $I_2I_3I_0$, $I_3I_0I_1$). При этом точки A, B, C однозначно соответствуют разбиениям множества из четырёх индексов $\{0, 1, 2, 3\}$ на пары: $A = I_0I_1 \cap I_2I_3$, $B = I_0I_2 \cap I_1I_3$, $C = I_0I_3 \cap I_1I_2$.

Серия A: Первая дюжина: точки касания

Пусть A_i, B_i, C_i $(i = 0, 1, 2, 3)$ — точки касания окружности ω_i с прямыми BC, CA, AB соответственно (на рис. A — красные точки, их 12 штук).

A1. A_0 и A_1, а также A_2 и A_3 симметричны относительно A', причём $A_0A_3 = A_1A_2 = c$, $A_0A_2 = A_1A_3 = b$, $A'A_0 = A'A_1 = \dfrac{|b-c|}{2}$, $A'A_2 = A'A_3 = \dfrac{b+c}{2}$. (Аналогично — симметрия относительно B' и C'.)

A2. а) Прямые AA_i, BB_i, CC_i пересекаются в одной точке.

б) Прямые AA_1, BB_2, CC_3 пересекаются в одной точке. (Аналогично, тройки прямых AA_0, BB_3, CC_2; AA_2, BB_1, CC_0; AA_3, BB_0, CC_1 пересекаются в одной точки или параллельны.)

A3. Радикальные оси пар окружностей ω_i и ω_j — внутренние и внешние биссектрисы углов треугольника $A'B'C'$. (Найдите радикальные центры всевозможных троек из окружностей ω_0, ω_1, ω_2, ω_3.)

A4#. Среди окружностей, касающихся тройки окружностей ω_1, ω_2, ω_3, есть три окружности, проходящие через точку I_0'. (Докажите аналогичное утверждение для других троек окружностей.)

A5#. $AA_1 \parallel I_0A'$ (аналогично $AA_0 \parallel I_1A'$, $AA_2 \parallel I_3A'$, $AA_3 \parallel I_2A'$ и т. д.).

A6#. Прямые I_0A_1, I_1A_0, I_2A_3, I_3A_2 пересекаются в одной точке. Что это за точка?

Серия B: Вторая дюжина: «фокусы»

Обозначим точки пересечения $B_{01} = B_{10} = A_0B_0 \cap A_1B_1$, $B_{23} = B_{32} = A_2B_2 \cap A_3B_3$. (Здесь $A_0B_0 \cap A_1B_1$ — это именно B_{01}, а не A_{01}, так как A соответствует разбиению множества индексов $\{0, 1, 2, 3\}$ на пары 0, 1 и 2, 3.) Аналогично определим точки все 12 точек: A_{ij}, где $i \in \{0, 1\}$, $j \in \{2, 3\}$ (везде полагаем $A_{ij} = A_{ji}$); B_{ij}, где $i \in \{0, 2\}$, $j \in \{1, 3\}$; C_{ij}, где $i \in \{0, 3\}$, $j \in \{1, 2\}$. На рис. B1, B2 все эти точки — фиолетовые точки.

Докажите следующие утверждения.

B0. Докажите, что угол $B_2B_{23}B_3$ — прямой (то же для аналогичных углов).

B1. Точки B_{23}, C_{23}, A_2, A_3 лежит на одной окружности. Найдите центр этой окружности. (Аналогично точки B_{01}, C_{01}, A_0, A_1 лежат на одной окружности, и т. д., таким образом получается, что красные и фиолетовые точки расположены на шести окружностях.)

B2. A_{ij} лежат на средней линии $B'C'$ (аналогично, B_{ij} и C_{ij} лежат на средних линиях, таким образом, 12 фиолетовых точек расположены по 4 на трёх прямых $A'B'$, $B'C'$, $C'A'$).

B3. A_{13} (и аналогично A_{02}, B_{01}, B_{23} лежат на окружности с диаметром AB, (причём $A_{02}B_{01}A_{13}B_{23}$ — прямоугольник.) (Таким образом, фиолетовые точки — расположены по 4 на трёх окружностях с диаметрами BC, CA, AB).

B4. Выразите длины $A_{02}A_{03}$ и т.д. через a, b, c.

B5. Точка A_{ij} лежит на прямой I_iI_j, причём A_{ij} является проекцией точки A на прямую I_iI_j. (таким образом, 12 фиолетовых точек лежат по две на шести биссектрисах углов треугольника ABC).

B6. Точки A_{02} и C_{02} — фокусы окружностей ω_0 и ω_2 (то есть A_{02} и C_{02} – пара точек, инверсных относительно каждой из этих двух окружностей). (Таким образом, фиолетовые точки разбиваются на 6 пар фокусов; отсюда, в частности, следует, что внутри каждой из окружностей ω_i лежит ровно три фиолетовые точки).

B7#. Найдите радикальные центры восьми троек таких окружностей с разными центрами из задачи B1.

B8#. Шестёрка точек $A_{03}A_{02}C_{02}C_{23}B_{23}B_{03}$ лежит на одной окружности (имеются ещё три аналогичные окружности). Найдите центры этих окружностей. Выразите их радиусы через элементы треугольника ABC.

B9#. A_{02} и A_{13} — центры соответственно вписанной и вневписанной, либо двух вневписанных окружностей для треугольника $B'H_aH_b$.

Серия C: Третья дюжина: «пересечения — на высотах»

Положим $A_{(3)} = A_0C_0 \cap A_1B_1$ (Здесь $A_0C_0 \cap A_1B_1$ — это именно $A_{(3)}$, а не $A_{(2)}$, так как точке C соответствует разбиению индексов на пары 0, 3 и 1, 2, и индекс 3 — это второй индекс из пары, содержащей 0). Аналогично, $A_{(2)} = A_0B_0 \cap A_1C_1$, $A_{(0)} = A_2B_2 \cap A_3C_3$, $A_{(1)} = A_2C_2 \cap A_3B_3$, и точно так же вводятся точки $B_{(i)}$ и $C_{(i)}$ — всего 12 точек, они отмечены зелёным на рис. C.

Докажите следующие утверждения.

C1. Точки $A_{(i)}$ лежат на прямой AH_a (и аналогично для точек $B_{(i)}$ и $C_{(i)}$, таким образом 12 зелёных точек лежат по 4 точки на каждой из высот треугольника ABC).

C2. Отрезок $AA_{(i)}$ равен по длине r_i.

C3. Прямая $A_{(i)}A_i$ параллельна одной из биссектрис угла A.

C4. Докажите, что прямые $A_{(1)}A_1$, $B_{(2)}B_2$ и $C_{(3)}C_3$ пересекаются в одной точке. (Аналогично, имеются ещё три тройки прямых, пересекающихся в одной точке: $A_{(0)}A_0$, $B_{(3)}B_3$ и $C_{(2)}C_2$; $A_{(3)}A_3$, $B_{(0)}B_0$ и $C_{(1)}C_1$; $A_{(2)}A_2$, $B_{(1)}B_1$ и $C_{(0)}C_0$.)

C5. Треугольники $A_1B_2C_3$ и $A_{(0)}B_{(0)}C_{(0)}$ центрально симметричны. Найдите их центр симметрии. (Аналогично, пары треугольников $A_0B_3C_2$ и $A_{(1)}B_{(1)}C_{(1)}$, $A_3B_0C_1$ и $A_{(2)}B_{(2)}C_{(2)}$, $A_2B_1C_0$ и $A_{(3)}B_{(3)}C_{(3)}$ центрально симметричны.)

C6. Описанные окружности треугольников $A_{(1)}B_{(2)}C_{(3)}$, $A_{(0)}B_{(3)}C_{(2)}$, $A_{(3)}B_{(0)}C_{(1)}$, $A_{(2)}B_{(1)}C_{(0)}$ имеют общий центр (таким образом зелёные точки расположены по три на четырёх концентрических окружностях). Найдите общий центр этих четырёх окружностей.

C7.# Выразите длины отрезков AH, BH, CH через радиусы r_i.

C8.# Выразите радиус описанной окружности треугольника $A_{(1)}B_{(2)}C_{(3)}$ через R и r_0. (Аналогичным образом, выразите радиусы окружностей из задачи C6.)

C9.# Прямая I_iA' проходит через $A_{(i)}$ (аналогично I_iB' проходит через $B_{(i)}$, I_iC' проходит через $C_{(i)}$).

Серия D: Четвертая дюжина.

Положим $C_0^* = A_0B_0 \cap A_1B_2$, и аналогично введём 12 точек A_i^*, B_i^*, C_i^* (на рис. D они покрашены синим). (Построение этих точек легко описывается следующим образом: Возьмём одну из окружностей, например ω_0. Возьмём точки её касания с двумя сторонами, например A_0, B_0. Возьмём точки касания этих же сторон с двумя другими окружностями, которые симметричны выбранным ранее относительно соответствующих середин, в данном случае A_1, B_2. Построим точку пересечения прямых, соединяющих две выбранных пары точек касания).

Докажите следующие утверждения.

D1. Стороны треугольника $A_i^*B_i^*C_i^*$ проходят через вершины ABC.

D2. Проведём через C_i^* произвольную прямую и найдем точки A'', B'' её пересечения со сторонами BC, AC. Тогда прямые $A''B_i^*$, $B''A_i^*$ пересекаются в некоторой точке C'' стороны AB.

D3. Прямые AA'', BB'', CC'' пересекаются в одной точке, изогонально сопряжённая к которой лежит на прямой OI_i.

D4. Окружность $A''B''C''$ проходит через точку Фейербаха F_i. Наверное, есть ещё какие-то свойства.

D5. Четыре синие точки, обозначенные одной буквой, лежат на одной прямой — соответствующей стороне ортотреугольника.

D6. a) Треугольники $A_i^*B_i^*C_i^*$ и ABC перспективны (то есть прямые, соединяющие соответствующие вершины этих треугольников, пересекаются в одной точке).

b) Попытайтесь отыскать какие-либо соотношения между четырьмя центрами перспективы.

D7. (Обобщение задачи D4) Рассмотрим произвольную точку C^{**} на прямой H_aH_b. Проведём через C^{**} произвольную прямую и найдем точки A'', B'' её пересечения со сторонами BC, AC. Пусть P — точка пересечения прямых AA'' и BB'', а C'' — точка пересечения CP и AB. Тогда описанные окружности всех треугольников $A''B''C''$ имеют общую точку.

Вневписанные окружности и дюжины точек.

Указания, решения, комментарии.

Серия A: Дюжина точек касания

A1. Следует из подсчёта отрезков касательных, например, $2AB_1 = AB_1 + AC_1 = AB + BA_0 + AC + CA_0 = 2p$, откуда $B'B_1 = p - \frac{b}{2} = \frac{a+c}{2}$. (См. также замечание к задаче B5.)

A2. Следует из теоремы Чевы (используется равенство отрезков касательных).

A3. Из задачи A1 следует, что точка A' имеет равные степени относительно окружностей ω_2 и ω_3, значит A' лежит на их радикальной оси. Кроме того, эта радикальная ось перпендикулярна линии центров I_2I_3, то есть параллельна (внутренней) биссектрисе угла BAC или угла $B'A'C'$. Таким образом, радикальная ось — биссектриса угла $B'A'C'$. Искомые радикальные центры — точки I'_0, I'_1, I'_2, I'_3.

A4#. (Эта задача формулировалась как гипотеза в докладе К. Кузнецовой (Великие Луки) на конференции школьников «Старт в науку — 2009»)

Из задачи A3 следует, что существует инверсия с центром I'_0, переводящая каждую из окружностей ω_1, ω_2, ω_3 в себя. При этой инверсии прямые AB, BC, CA перейдут в окружности, проходящие через I'_0 и касающиеся окружностей ω_1, ω_2, ω_3.

A5#. Гомотетия с центром A, переводящая ω_1 в ω_0, переводит диаметр KA_1 в диаметр A_0L. Таким образом, прямая AA_1 совпадает с прямой LA_1. После этого наблюдения утверждение задачи следует из того, что I_0A' — средняя линий треугольника A_0LA_1.

A6#. В обозначениях из решения предыдущей задачи: треугольники A_1LA_0 и A_1AH_a гомотетичны (с центром A_1), поэтому прямая A_1I_0 является медианой в треугольнике A_1AH_a, то есть проходит через середину высоты AH_a.

Серия B: Вторая дюжина: «фокусы»

B0. Следует из того, что прямые A_iB_i параллельны биссектрисам угла C (внешней или внутренней).

B1. Из задачи B0 следует, что $A_2B_{23} \perp A_3B_{23}$ и $A_2C_{23} \perp A_3C_{23}$, поэтому указанные 4 точки лежат на одной окружности с диаметром A_2A_3. Из A1 следует, что центр этой окружности — A' (а радиус равен $\frac{b+c}{2}$. Аналогично, точки B_{01}, C_{01}, A_0, A_1 лежат на одной окружности с центром A' (и радиусом $\frac{|b-c|}{2}$).

B2. В прямоугольном треугольнике $A_2B_{23}A_3$ имеем: $A'B_{23} = A'A_2$, (и равно $\frac{b+c}{2}$ — см. A2), поэтому равнобедренные треугольники $A_2A'B_{23}$ и A_2CB_2 гомотетичны, и $A'B_{23} \parallel AC$, то есть B_{23} лежит на средней линии $A'C'$.

Замечание. B_{23} также лежит на окружности с диаметром B_2B_3.

B3. Зная, длину $A'B_{23}$ (см. B1), легко найти $C'B_{23} = A'B_{23} = A'C' = \frac{c}{2}$, поэтому B_{23} лежит на окружности радиуса $\frac{c}{2}$ с центром C'. Для других точек типа B_{ij} подсчёт аналогичен.

B4. Из B2 легко получить: $A_{13}A_{12} = A_{13}C' + C'B' + B'A_{12} = \frac{c+a+b}{2} = p$; $A_{13}A_{03} = A_{13}A_{12} - A_{03}A_{12} = p - b$ (так как $A_{03}A_{12}$ — диаметр окружности из B2). Аналогично $A_{03}A_{02} = p - a$, $A_{02}A_{12} = p - c$.

В5. (Одна из возможных конфигураций этой задачи — в задаче 1.66 в задачнике Прасолова, см. также статью Протасова («Квант», № 4 — 2008); также см. задачу 255 из задачника Шарыгина 9 — 11, которую автор даже отмечает (случайно ли?) в предисловии.)

Из В1 и В2 имеем: $C'B_{23} \parallel AC$ и $C'B_{23} = C'A$, откуда $\angle B_{23}AC' = \angle C'AB_{23} = \angle B_{23}AB_3$, таким образом, AB_{23} — внешняя биссектриса угла BAC. Кроме того, из В2 следует, что $BB_{23} \perp AB_{23}$. Для других точек доказательство аналогично.

Замечание. Обратим внимание на множество параллелограммов на рисунке (стороны которых параллельны либо сторонам треугольника ABC, либо его биссектрисам). Скажем, из параллелограммов $A_3A_{13}A_{23}C$ и $BA_{13}A_{23}A_2$ видно геометрическое объяснение задачи А1.

В6. Треугольники $I_0A_{02}B_0$ и $I_0B_0C_{02}$ подобны (в подсчёте углов используем, что B_0C_{02} параллельна внешней биссектрисе угла ABC), откуда $I_0A_{02} \cdot I_0C_{02} = r_0^2$.

В7#. Искомые радикальные центры — это точки I_i, а также точки, симметричные им относительно центра описанной окружности треугольника ABC.

Например, из В6 следует, что $I_0A_{02} \cdot I_0C_{02} = I_0A_{03} \cdot I_0B_{03} = I_0B_{01} \cdot I_0C_{01} = r_0^2$, значит степени точки I_0 относительно окружностей, построенных на отрезках A_0A_1, B_0B_2, C_0C_2 как на диаметрах (см. В1), равны.

Далее, рассмотрим, например, окружности с диаметрами A_2A_3, B_1B_3 и C_1C_2. Точка I_3 лежит на радикальной оси первых двух окружностей, так как равные отрезки I_3A_2 и I_3B_1 являются касательными к этим окружностям. Кроме того, радикальная ось перпендикулярна линии центров этих окружностей, то есть средней линии треугольника ABC. Три таких прямые пересекаются в точке, симметричной I относительно O.

В8#. Это окружности с центрами I_i'.

В самом деле, пусть, скажем, X – проекция I_0' на $B'C'$. Тогда из В3 вытекает: $XA_{12} = XB' + B'A_{12} = \dfrac{p-b}{2} + \dfrac{b}{2} = \dfrac{p}{2}$. Тогда $I_0'A_{12}^2 = I_0'X^2 + XA_{12}^2 = \dfrac{r_0^2 + p^2}{4}$. Аналогично квадрат расстояния от точки I_0' до любой из точек A_{03}, A_{02}, C_{02}, C_{23}, B_{23}, B_{03} равен $\dfrac{r^2 + p^2}{4}$.

Таким же образом, доказывается, что окружность с центром I_1' имеет радиус $\dfrac{r_1^2 + (p-a)^2}{4}$ и т. д.

Замечание. На самом деле, нетрудно установить общий факт: три пары фокусов для трёх окружностей, центры которых не лежат на одной прямой, лежат на одной окружности (доказательство — упражнение на степень точки плюс тот факт, что радикальные оси должны пересекаться в одной точке).

Замечание. Это одна из окружностей семейства *Тукера* для треугольника $I_1I_2I_3$.

В9#. (См. также статью В. Протасова из «Кванта» № 4 — 2008, эта задача играет важную роль в доказательстве теоремы Фейербаха.) C_0A_{02} — биссектриса угла $AB'H_a$ (из симметрии). Рассмотрим окружность девяти точек, треугольник $B'H_aH_b$, вписанный в эту окружность, и точку C' — середину дуги H_aH_b. Так как (см. В3) $C'A_{02} = C'H_a = C'H_b$, то по теореме, обратной лемме о трезубце, получаем, что A_{02} — центр вписанной или вневписанной окружности для треугольника $B'H_aH_b$.

Серия С: Третья дюжина: «пересечения — на высотах»

С1-3. Из В5 следует, что $AA_{02}A_0A_{03}$ — параллелограмм (его стороны параллельны (внешним) биссектрисам углов CBA и ACB). Также $A_{(0)}A_{02}I_0A_{03}$ — параллелограмм (его стороны параллельны (внутренним) биссектрисам углов CBA и ACB). Поэтому $\overrightarrow{I_0A_0}$ и $\overrightarrow{A_{(0)}A}$ симметричны относительно середины отрезка $A_{02}A_{03}$. Отсюда вытекает С1 и С2. Так как $I_0A_0A_{(0)}A$ — параллелограмм, то $A_{(i)}A_i \parallel AI_0$.

C4. (Это задача Емельянова 10.7 с 5 этапа Всероссийской олимпиады 2002? года.) Из C3 вытекает, что эти прямые — высоты треугольника $A_{(1)}B_{(2)}C_{(3)}$.

Замечание. Можно показать, что точка пересечения указанных трёх прямых симметрична ортоцентру треугольника $A_0B_0C_0$ относительно точки I'_0.

C5. Покажем, что искомые центры симметрии — точки I'_i.

Так как радикальная ось делит пополам отрезки общих касательных, из задачи A3 вытекает, что прямые B_2C_2 $(= B_2C_{(0)})$ и B_3C_3 $(= C_3B_{(0)})$ симметричны относительно прямой $A'I'_0$, или относительно точки I'_0. Аналогично, прямые $A_1C_{(0)}$ и $C_3A_{(0)}$ симметричны относительно I'_0. Это означает, что соответствующие точки пересечения $C_{(0)}$ и C_3 симметричны относительно I'_0.

C6. Искомый центр — точка H.

Из C1 мы знаем, что например, $A_{(0)} = A_3C_3 \cap AH_a$ и $C_{(2)} = A_3C_3 \cap CH_c$. Так как A_3C_3 параллельна биссектрисе угла B, то A_3C_3 образует также равные углы с высотами AH_a и CH_c. Отсюда вытекает, что треугольник $HA_{(0)}C_{(2)}$ равнобедренный, то есть H равноудалена от $A_{(0)}$ и $C_{(2)}$.

C7-8. Радиусы описанных окружностей из задачи C6 равны $|\rho_i|$, где $\rho_0 = AH + r_1 = BH + r_2 = CH + r_3$, $\rho_1 = r_0 - AH = BH - r_3 = CH - r_2$, $\rho_2 = AH - r_3 = r_0 - BH = CH - r_1$, $\rho_3 = AH - r_2 = BH - r_1 = r_0 - CH$ (здесь AH и т. д. позволим быть отрицательными, если соответствующий угол треугольника тупой). Отсюда выражаем AH, BH, CH через радиусы r_i $AH = \dfrac{r_0 + r_1 + r_2 + r_3}{2} - r_1$, $BH = \dfrac{r_0 + r_1 + r_2 + r_3}{2} - r_2$, $CH = \dfrac{r_0 + r_1 + r_2 + r_3}{2} - r_3$.

Получаем: $\rho_0 = \dfrac{r_0 + r_1 + r_2 + r_3}{2}$, $\rho_1 = \dfrac{r_0 + r_1 - r_2 - r_3}{2}$, $\rho_2 = \dfrac{r_0 - r_1 + r_2 - r_3}{2}$, $\rho_3 = \dfrac{r_0 - r_1 - r_2 + r_3}{2}$, или с учётом соотношения $r_1 + r_2 + r_3 = 4R + r_0$ (см. задачник Прасолова 12.24), $\rho_0 = r_0 + 2R$, $\rho_1 = |r_1 - 2R|$, $\rho_2 = |r_2 - 2R|$, $\rho_3 = |r_3 - 2R|$.

C9.# Из задачи A5 вытекает, что, скажем, I_0A' пересекает высоту AH_a в точке S такой, что $\overrightarrow{AS} = \overrightarrow{I_0A_0}$, то есть в точке $A_{(0)}$ (см. задачи C1-2).

C10.# (задача предложена Д. Прокопенко) 1. Нетрудно видеть, что A — середина MN, поэтому $AA_{(0)}$ серединный перпендикуляр в треугольнике MA_0N.

2. Из задачи C3 следует, что $A_0A_{(0)}$ и A_0I_0 симметричны относительно биссектрисы угла MA_0N. Так как A_0I_0 — высота треугольника, то прямая $A_0A_{(0)}$ содержит центр описанной окружности треугольника MA_0N. Из 1 и 2 следует требуемое.

Ответ: ортоцентром треугольника A_0MN является точка, симметричная A_0 относительно I_0.

Задачи серии D представляют собой переформулировку утверждения теоремы Емельяновых, и их решения могут быть найдены в книге «Летние Конференции Турнира Городов. Избранные материалы. Выпуск 1.» МЦНМО, 2009.

Excircles and Dozens of Points.

(Presented by V. Filimonov and A. Zaslavsky.)

On the Origin of this Series of Problems.

The work on this series started from the problem posed by D. Tereshin.

Problem (D. Tereshin). Consider triangle ABC and its excircles: one of them touches the side AC at K and touches the lines AB and BC at L and M, the other touches the side AB at P and touches the lines AC and BC at Q and R. Prove that the intersection point X of the lines LM and QR lies on the altitude (passing through A) of the triangle ABC.

At once the geometrical solution was not found, only calculations work. Some observations made this problem more exciting. It appears that the intersection point Y of lines KM and PR lies on the altitude of triangle ABC, and the length of the segments AY and AX equal to the radii of incircle and excircle touching the side BC. Some other results were obtained, and some connections with known problems from olympiads were established. In search for the geometrical explanation of these results we tried to consider in details the touch points of the sides with incircle and excircles, and the lines joining these touch points.

The incircle and the excircles have some different properties (for example, the incircle is smaller than any of the excircles, the incircle lies inside the triangle while the excircle lies outside the triangle). Nevertheless, these four circles have deep common properties: each of them touches the three sidelines of the triangle, the center of each circle is the intersection of three angle bisectors (either internal or external). So as a rule, if one of the four circles has some property, then the others have an analogous property. That is why these four circles in some sense enjoy equal rights with respect to the original triangle. To understand some important geometrical results we need to consider all the four circles simultaneously. Thus we introduce some non-regular but symmetrical notation (see below).

The sections A, B, C of the project were made by the author of the text jointly with I. Bogdanov, the section D was added by A. Zaslavsky. Also A. Akopyan, D. Prokopenko, and V. Protassov had made many useful notes and additions.

P. Kozhevnikov

Excircles and Dozens of Points.

Notation.

In a fixed non-equilateral triangle ABC let us denote:

R, p — the radius of the circumcircle and semiperimeter;

a, b, c — lengths of BC, CA, AB;

A', B', C' — midpoints of BC, CA, AB;

AH_a, BH_b, CH_c — altitudes, H — orthocenter of triangle ABC;

Ω — circumcircle, O — circumcenter;

ω_0 — incircle, I_0 — incenter; ω_1, ω_2, ω_3 — excircles (touching segments BC, CA, AB, respectively), I_1, I_2, I_3 — centers of the excircles, r_i — radii of ω_i;

I'_0, I'_1, I'_2, I'_3 — centers of incircle and excircles of triangle $A'B'C'$.

The notation has the following symmetry: Note that 6 lines I_iI_j $(i \neq j)$ are internal and external bisectors of triangle ABC. Therefore the quadruple I_0, I_1, I_2, I_3 is orthocentric, and ABC is the orthotriangle (that is the triangle having feet of altitudes as vertices) for each of triangles $I_0I_1I_2$, $I_1I_2I_3$, $I_2I_3I_0$, $I_3I_0I_1$). To each of the points A, B, C we put into correspondence a partition of 4-element set $\{0, 1, 2, 3\}$ into two 2-element subsets: $A = I_0I_1 \cap I_2I_3$, $B = I_0I_2 \cap I_1I_3$, $C = I_0I_3 \cap I_1I_2$.

Also see the further notation

Series A: The First Dozen: Touch Points

Let A_i, B_i, C_i $(i = 0, 1, 2, 3)$ be touch points of ω_i and lines BC, CA, AB, respectively (see 12 red points in Fig. A).

Prove the following statements.

A1. A_0 и A_1 (and also A_2 and A_3) are symmetric with respect to A', moreover, $A_0A_3 = A_1A_2 = c$, $A_0A_2 = A_1A_3 = b$, $A'A_0 = A'A_1 = \dfrac{|b-c|}{2}$, $A'A_2 = A'A_3 = \dfrac{b+c}{2}$. (Similarly there is symmetry with respect to B' and C'.)

A2. a) AA_i, BB_i, CC_i are concurrent.

б) AA_1, BB_2, CC_3 are concurrent. (Similarly, triples of lines AA_0, BB_3, CC_2; AA_2, BB_1, CC_0; AA_3, BB_0, CC_1 are either concurrent or parallel.)

A3. Radical axis of pairs ω_i and ω_j are internal and external bisectors of triangle $A'B'C'$. (Find the radical centers of triples of circles ω_0, ω_1, ω_2, ω_3.)

A4. In the set of circles touching ω_1, ω_2, ω_3 there exist three circles passing through I'_0. (Formulate and prove the similar statement for the other triples of circles.)

A5. $AA_1 \parallel I_0A'$ (similarly $AA_0 \parallel I_1A'$, $AA_2 \parallel I_3A'$, $AA_3 \parallel I_2A'$, etc.).

A6. I_0A_1, I_1A_0, I_2A_3, I_3A_2 are concurrent. Determine the intersection point of these lines.

Series B: The Second Dozen: "Foci"

Let us denote $B_{01} = B_{10} = A_0B_0 \cap A_1B_1$, $B_{23} = B_{32} = A_2B_2 \cap A_3B_3$. (Here $A_0B_0 \cap A_1B_1$ is B_{01}, and not A_{01}, since A corresponds to the partition of $\{0,1,2,3\}$ into pairs $0,1$ and $2,3$.) Similarly define all 12 points: A_{ij} with $i \in \{0,1\}$, $j \in \{2,3\}$ (we put $A_{ij} = A_{ji}$); B_{ij} with $i \in \{0,2\}$, $j \in \{1,3\}$; C_{ij} with $i \in \{0,3\}$, $j \in \{1,2\}$. (See 12 violet points in Fig. B1, B2.)

Prove the following statements.

B0. $\angle B_2B_{23}B_3 = 90^\circ$ (similarly for the other angles).

B1. B_{23}, C_{23}, A_2, A_3 are concyclic.

Find the center of the circle passing through these points. (Similarly, B_{01}, C_{01}, A_0, A_1 are concyclic, etc., thus red and violet points belong to 6 circles.)

B2. A_{ij} lies on the midline $B'C'$

(similarly, B_{ij} and C_{ij} lie on midlines, thus 12 violet points lie on 3 lines $A'B'$, $B'C'$, $C'A'$).

B3. A_{13} (the same for A_{02}, B_{01}, B_{23}) lie on the circle with diameter AB,

(moreover, $A_{02}B_{01}A_{13}B_{23}$ is a rectangle.)

(Thus 12 violet points lie on 3 circles with diameters BC, CA, AB).

B4. Find the lengths $A_{02}A_{03}$, etc., in terms of a, b, c.

B5. A_{ij} lies on I_iI_j, moreover, A_{ij} is the projection of A to I_iI_j. (thus 12 violet points belong to 6 bisectors of triangle ABC).

B6. A_{02} and C_{02} are *foci* of ω_0 and ω_2 (*Foci* means that A_{02} and C_{02} is a pair of points inverse to each other with respect to each of two circles). (Thus 12 violet points are partitioned into 6 pairs of foci; in particular, from that it follows that each of ω_i contains exactly 3 violet points).

B7. Determine the radical centers of triples of circles from Problem B1 having distinct centers.

B8. Six points $A_{03}, A_{02}, C_{02}, C_{23}, B_{23}, B_{03}$ lie on a circle (also there exist 3 circles constructed in the same manner). Determine the centers of these circles. Find the radii of these circles in terms of a, b, c.

B9. A_{02} and A_{13} are either the centers of incircle and excircle or the centers of excircles, for the triangle $B'H_aH_b$.

Series C: The Third Dozen: Intersections on the Altitudes

Let $A_{(3)} = A_0C_0 \cap A_1B_1$ (Here $A_0C_0 \cap A_1B_1$ is $A_{(3)}$, and not $A_{(2)}$, since C corresponds to the partition of $\{0,1,2,3\}$ into pairs $0,3$ and $1,2$, here 3 belongs to the pair containing 0). Similarly, $A_{(2)} = A_0B_0 \cap A_1C_1$, $A_{(0)} = A_2B_2 \cap A_3C_3$, $A_{(1)} = A_2C_2 \cap A_3B_3$; in the same manner define $B_{(i)}$ and $C_{(i)}$ — totally 12 green points in Fig. C.

Prove the following statements.

C1. Points $A_{(i)}$ lie on the line AH_a (similarly for $B_{(i)}$ and $C_{(i)}$, thus 12 green points lie on the altitudes of triangle ABC).

C2. The length of $AA_{(i)}$ is equal to r_i.

C3. $A_{(i)}A_i$ is parallel to one of two bisectors of angle A.

C4. Prove that $A_{(1)}A_1$, $B_{(2)}B_2$, and $C_{(3)}C_3$ are concurrent. (Similarly, there exist three triples of concurrent lines: $A_{(0)}A_0$, $B_{(3)}B_3$, $C_{(2)}C_2$; $A_{(3)}A_3$, $B_{(0)}B_0$, $C_{(1)}C_1$; $A_{(2)}A_2$, $B_{(1)}B_1$, $C_{(0)}C_0$.)

C5. Triangles $A_1B_2C_3$ and $A_{(0)}B_{(0)}C_{(0)}$ are symmetric (with respect to a point). Define the center of symmetry. (Similarly, pair of triangles $A_0B_3C_2$ and $A_{(1)}B_{(1)}C_{(1)}$, $A_3B_0C_1$ and $A_{(2)}B_{(2)}C_{(2)}$, $A_2B_1C_0$ and $A_{(3)}B_{(3)}C_{(3)}$ are symmetric.)

C6. Triangles $A_{(1)}B_{(2)}C_{(3)}$, $A_{(0)}B_{(3)}C_{(2)}$, $A_{(3)}B_{(0)}C_{(1)}$, $A_{(2)}B_{(1)}C_{(0)}$ Have a common circumcenter (thus green points lie on 4 concentric circles). Define the common circumcenter.

C7. Find AH, BH, CH in terms of radii r_i.

C8. Find the radius of the circumcircle of triangle $A_{(1)}B_{(2)}C_{(3)}$ in terms of R and r_0. (Similarly, find the radii of the circles from Problem C6.)

C9. I_iA' passes through $A_{(i)}$ (similarly, I_iB' passes through $B_{(i)}$, I_iC' passes through $C_{(i)}$).

C10. Let l_a be a line passing through A and parallel to BC. $M = A_0C_0 \cap l_a$, $N = A_0B_0 \cap l_a$. Prove that $A_{(0)}$ is the circumcircle of the triangle A_0MN.

Determine the orthocenter of the triangle A_0MN.

Series D: The Fourth Dozen.

Let $C_0^* = A_0B_0 \cap A_1B_2$, and similarly define 12 blue points A_i^*, B_i^*, C_i^* (see Fig. D). (The description of these points is as follows: Take one of the circles, for example ω_0. Take its two touch points, say A_0, B_0. Take the touch points of these sides with two other circles that are symmetric to A_0, B_0 with respect to the midpoints of the sides — A_1, B_2. Take the intersection points of the lines joining pairs of these points.)

Prove the following statements.

D1. The sidelines of triangle $A_i^*B_i^*C_i^*$ pass through the vertices of triangle ABC.

D2. A line passing through C_i^* intersects BC, AC at A'', B'', respectively. Show that $A''B_i^*$ and $B''A_i^*$ intersect at some point C'' of the line AB.

D3. AA'', BB'', CC'' have a common point that is isogonally conjugate to some point of the line OI_i.

D4. The circumcircle of triangle $A''B''C''$ passes through the Feuerbach point F_i.

D5. Four blue points denoted by the same letter lie on a sideline of the orthotriangle.

D6. a) Triangles $A_i^*B_i^*C_i^*$ and ABC are perspective (i.e. the lines joining corresponding vertises of these triangles are concurrent)

b)Try to find some relations between four centers of perspective.

D7. (The generalization of the problem D4) Let C^{**} be a point on line H_aH_b. An arbitrary line passing through C^{**} intersects BC, AC at A'', B'', respectively. Let P be the point of intersection of lines AA'' and BB'', and C'' be the point of intersection of lines CP and AB. Then the circumcircles of all triangles $A''B''C''$ have the common point.

Tasks from series A, B, C marked # and also from series D were given to the paticipants after the intermediate finish.

Excircles and Dozens of Points.

Hints, Solutions, Comments.

Series A: The First Dozen: Touch Points

A1. Follows from the calculation of the tangent segments, for example, $2AB_1 = AB_1 + AC_1 = AB + BA_0 + AC + CA_0 = 2p$, hence $B'B_1 = p - \frac{b}{2} = \frac{a+c}{2}$. (Also see a comment on B5)

A2. Follows from Ceva Theorem using the equality of the segments of tangents).

A3. From A1 it follows that A' equal powers with respect to the circles ω_2 and ω_3, hence A' lies on the radical axis of these circles. This radical axis is perpendicular to I_2I_3, hence it is parallel to the bisector of the angle BAC (or $B'A'C'$). Thus this radical axis as a bisector of the angle $B'A'C'$. Hence radical centers of the triples of circles are I'_0, I'_1, I'_2, I'_3.

A4#. (This Problem was formulated in thesis of K. Kuznetsova (Velikie Luki) at the Conference "Start v Nauku — 2009")

From A3 it follows that there exists an inversion with center I'_0 that takes each of the circles ω_1, ω_2, ω_3 to itself. This inversion takes AB, BC, CA to the circles passing through I'_0 and touching ω_1, ω_2, ω_3.

A5#. The homothety with center A taking ω_1 to ω_0 takes diameter KA_1 to diameter A_0L. Thus AA_1 coincides to LA_1. I_0A' is a midline of the triangle A_0LA_1. This completes the solution.

A6#. Using the notation of the previous solution: triangles A_1LA_0 and A_1AH_a are homothetic (with center A_1), hence A_1I_0 is the median in triangle A_1AH_a passing through the midpoint of the altitude AH_a.

Series B: The Second Dozen: "Foci"

B0. The statement follows since A_iB_i is parallel to a bisector (either internal or external) of angle C.

B1. From B0 it follows that $A_2B_{23} \perp A_3B_{23}$ и $A_2C_{23} \perp A_3C_{23}$, hence 4 mentioned points lie on the circle with diameter A_2A_3. From A1 it follows that the center of this circle is A' (and radius equals to $\frac{b+c}{2}$. Similarly, points B_{01}, C_{01}, A_0, A_1 lie on the circle with center A' (and radius equals $\frac{|b-c|}{2}$.

B2. In a right-angled triangle $A_2B_{23}A_3$: $A'B_{23} = A'A_2$, ($= \frac{b+c}{2}$ — see A2), hence equilateral triangles $A_2A'B_{23}$ and A_2CB_2 are homothetic, and $A'B_{23} \parallel AC$, that means that B_{23} lies on the midline $A'C'$.

Note. B_{23} also lies on the circle with diameter B_2B_3.

B3. Determine the length $A'B_{23}$ (see B1), and obtain $C'B_{23} = A'B_{23} = A'C' = \frac{c}{2}$, hence B_{23} lies on the circle of radius $\frac{c}{2}$ with center C'. For the other points the calculation could be done in the same manner.

B4. From B2 it is easy to obtain: $A_{13}A_{12} = A_{13}C' + C'B' + B'A_{12} = \frac{c+a+b}{2} = p$; $A_{13}A_{03} = A_{13}A_{12} - A_{03}A_{12} = p - b$ (since $A_{03}A_{12}$ is a diameter of the circle from B2). Similarly, $A_{03}A_{02} = p - a$, $A_{02}A_{12} = p - c$.

B5. (One of the possible configurations — Problem 1.66 in the book of Prassolov, also see the article of V. Protassov ("Quant", № 4 — 2008); also see Problem 255 from the book of Sharygin 9 — 11, this Problem is specially mentioned in the Preface.)

From B1 and B2 it follows: $C'B_{23} \parallel AC$ и $C'B_{23} = C'A$, hence $\angle B_{23}AC' = \angle C'AB_{23} = \angle B_{23}AB_3$, thus AB_{23} is the external bisector of angle BAC. Moreover, from B2 it follows that $BB_{23} \perp AB_{23}$. Similarly for other points.

Comment. Note that Fig. contain many parallelograms (the sides of which are parallel either to the sides or to the bisectors of the triangles ABC). For example, taking parallelograms $A_3A_{13}A_{23}C$ and $BA_{13}A_{23}A_2$ we see another explanation of the equality from A1.

B6. Triangles $I_0A_{02}B_0$ and $I_0B_0C_{02}$ are similar (in the calculations of angles we use that B_0C_{02} is parallel to the external bisector of the angle ABC), hence $I_0A_{02} \cdot I_0C_{02} = r_0^2$.

B7#. The radical centers are points I_i and points symmetrical to them with respect to the circumcenter of triangle ABC.

For example, from B6 it follows that $I_0A_{02} \cdot I_0C_{02} = I_0A_{03} \cdot I_0B_{03} = I_0B_{01} \cdot I_0C_{01} = r_0^2$, hence I_0 has equal powers with respect to the circles with diameters A_0A_1, B_0B_2, C_0C_2 (see B1).

Then, consider, for instance, the circles with diameters A_2A_3, B_1B_3 and C_1C_2. The point I_3 lies on the radical axis of the first two circles, because the equal segments I_3A_2 and I_3B_1 are tangent lines to these circles. Moreover, the radical axis is perpendicular to the line joining the centers of the circles, i.e. the medial line of ABC. Three such lines intersect in the point symmetrical to I with respect to O.

B8#. These are the circles with centers I_i'.

Let X be the projection of I_0' to $B'C'$. Then from B3 it follows: $XA_{12} = XB' + B'A_{12} = \dfrac{p-b}{2} + \dfrac{b}{2} = \dfrac{p}{2}$. Further, $I_0'A_{12}^2 = I_0'X^2 + XA_{12}^2 = \dfrac{r_0^2 + p^2}{4}$. Similarly, the square of the distance from I_0' to each of the points A_{03}, A_{02}, C_{02}, C_{23}, B_{23}, B_{03} equals to $\dfrac{r^2 + p^2}{4}$.

In the same way it is proved that the radius of the circle with center I_1' equals to $\dfrac{r_1^2 + (p-a)^2}{4}$, etc.

Comment. The following general result holds: three pair of foci for three circles which centers are not collinear lie on a circle (the proof is an exercise on a power of a point with respect to a circle).

Comment. This circle is of so called *Tucker* circles for the triangle $I_1I_2I_3$.

B9#. (Also see the article of V. Protassov in "Quant" № 4 — 2008, this problem plays an important role in the proof of Feuerbach Theorem.) C_0A_{02} is a bisector of angle $AB'H_a$ (this follows from symmetry). Consider a nine-point circle, triangle $B'H_aH_b$ is inscribed to this circle, C' is a midpoint of the arc H_aH_b. Since (see B3) $C'A_{02} = C'H_a = C'H_b$ we have that A_{02} is a center of either inscribed or exscribed circle of triangle $B'H_aH_b$.

Series C: The Third Dozen: Intersections on the Altitudes

C1-3. From B5 it follows that $AA_{02}A_0A_{03}$ is a parallelogram (its sides are parallel to bisectors of angles CBA and ACB). Also $A_{(0)}A_{02}I_0A_{03}$ is a parallelogram (its sides are parallel to bisectors of angles CBA and ACB). Therefore $\overrightarrow{I_0A_0}$ and $\overrightarrow{A_{(0)}A}$ are symmetric with respect to the midpoint of the segment $A_{02}A_{03}$. This implies C1 and C2. Since $I_0A_0A_{(0)}A$ is a parallelogram, $A_{(i)}A_i \parallel AI_0$.

C4. (This is the Problem of Emelyanov No 10.7 from All-Russian Olympiad — 2002?.) From C3 it follows that these lines are the altitudes of the triangle $A_{(1)}B_{(2)}C_{(3)}$.

Note. One can show that the intersection point of these three lines is symmetric to the orthocenter of triangle $A_0B_0C_0$ with respect to I_0'.

C5. Show that the centers are points I_i'.

Radical axis bisects the segments of common tangents, hence from A3 it follows that B_2C_2 $(= B_2C_{(0)})$ and B_3C_3 $(= {}_3B_{(0)})$ are symmetruc with respect to the line $A'I_0'$, and also with respect to point I_0'. Similarly, $A_1C_{(0)}$ and ${}_3A_{(0)}$ are symmetric with respect to I_0'. This means that the corresponding points of intersection $C_{(0)}$ and C_3 are symmetric with respect to I_0'.

C6. The center is H.

From C1 we know that, for example, that $A_{(0)} = A_3C_3 \cap AH_a$ and $C_{(2)} = A_3C_3 \cap CH_c$. Since A_3C_3 is parallel to the bisector of the angle B, the angles between A_3C_3 and the altitudes AH_a and CH_c are equal. From that it follows that triangle $HA_{(0)}C_{(2)}$ is equilateral, hence H is equidistant from $A_{(0)}$ and $C_{(2)}$.

C7-8. The radii of the circumcircles from C6 equal $|\rho_i|$, where $\rho_0 = AH + r_1 = BH + r_2 = CH + r_3$, $\rho_1 = r_0 - AH = BH - r_3 = CH - r_2$, $\rho_2 = AH - r_3 = r_0 - BH = CH - r_1$, $\rho_3 = AH - r_2 = BH - r_1 = r_0 - CH$ (here AH, ect., could be negative if the corresponding angle of the triangle is obtuse). From this we have $AH = \dfrac{r_0 + r_1 + r_2 + r_3}{2} - r_1$, $BH = \dfrac{r_0 + r_1 + r_2 + r_3}{2} - r_2$, $CH = \dfrac{r_0 + r_1 + r_2 + r_3}{2} - r_3$. Further, $\rho_0 = \dfrac{r_0 + r_1 + r_2 + r_3}{2}$, $\rho_1 = \dfrac{r_0 + r_1 - r_2 - r_3}{2}$, $\rho_2 = \dfrac{r_0 - r_1 + r_2 - r_3}{2}$, $\rho_3 = \dfrac{r_0 - r_1 - r_2 + r_3}{2}$, and putting the relation $r_1 + r_2 + r_3 = 4R + r_0$ (see the book of Prassolov, Problem 12.24), $\rho_0 = r_0 + 2R$, $\rho_1 = |r_1 - 2R|$, $\rho_2 = |r_2 - 2R|$, $\rho_3 = |r_3 - 2R|$.

C9.# From A5 it follows that that I_0A' intersects the altitude AH_a at point S such that $\overrightarrow{AS} = \overrightarrow{I_0A_0}$, that is the point $A_{(0)}$ (see Problems C1-2).

C10.# (This Problems was proposed by D. Prokopenko) 1. It is easy to show that A is the midpoint of MN, hence $AA_{(0)}$ is the perpendicular bisector of MN.

2. From C3 it follows that $A_0A_{(0)}$ and A_0I_0 are symmetric with respect to the bisector of the angle MA_0N. Since A_0I_0 is the altitude of the triangle, $A_0A_{(0)}$ contains the circumcenter of triangle MA_0N.

Combining 1 and 2 we get the required statement.

The orthocenter of triangle A_0MN is the point symmetric to A_0 with respect to I_0.

The tasks of series D are the reformulation of the Emelyanovs' Theorem, and their solutions can be found in the book "Summer Conferences of the Tournament of Towns. Selected matherials. Volume 1." (MCCME, 2009, in Russian)

Магические графы

К. Кохась, Д. Ростовский

Определения и обозначения

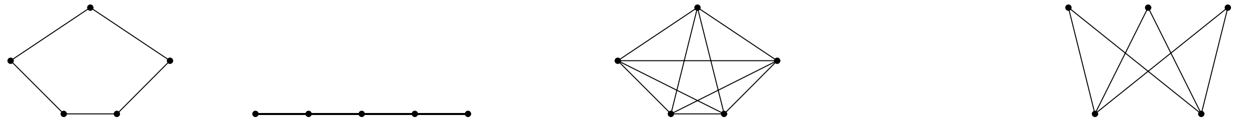
Все рассматриваемые графы не имеют изолированных вершин, кратных рёбер и петель.

Слова «цикл» и «путь» всюду означают *простой* цикл и *простой* путь в графе.

Граф называется *полумагическим*, если на его рёбрах можно расставить положительные числа (веса) так, что для каждой вершины сумма весов рёбер, выходящих из неё, равна одному и тому же числу s. Граф называется *магическим*, если возможна такая расстановка с попарно различными числами. Заметим, что в полумагическом графе висячая вершина может быть только концом изолированного ребра, причём в магическом графе такое ребро в графе может быть только одно.

Подграф F данного графа G называется его *скелетом*, если любая вершина G является вершиной одного из его рёбер. Скелет называется *1-2-скелетом*, если степень любой его вершины равна 1 или 2, причём степени вершин в каждой компоненте связности одинаковы. Иначе говоря, 1-2-скелет состоит из изолированных рёбер и непересекающихся простых циклов. Если зафиксирован 1-2-скелет F графа G, то все рёбра графа G делятся на три группы: принадлежащие *циклической* части F (обозначим её F_c); принадлежащие *линейной* части F (обозначим её F_ℓ), т.е. изолированные рёбра в F; наконец, вообще не принадлежащие F. Будем говорить, что 1-2-скелет *разделяет* рёбра e_1 и e_2, если эти два ребра лежат в разных группах. Иными словами, хотя бы одно из них должно принадлежать F, но не оба в F_c и не оба в F_ℓ.

Будем использовать обозначения: C_n — цикл из n рёбер ($n \geqslant 3$); P_n — путь из n рёбер; K_n — полный граф с n вершинами; $K_{m,n}$ — полный двудольный граф с долями по m и n вершин.



Цикл C_5 Путь P_5 Полный граф K_4 Полный двудольный граф $K_{2,3}$

Рис. 1. Некоторые стандартные графы

Прямым произведением $F \times G$ двух графов называется граф, у которого множество вершин есть множество всевозможных пар вида (v, w), где v — вершина F, w — вершина G. Вершины (v_1, w_1) и (v_2, w_2) соединены ребром, если либо $v_1 = v_2$ и в графе G есть ребро w_1w_2, либо $w_1 = w_2$ и в графе F есть ребро v_1v_2. *Удвоением* графа G будем называть граф $G \times P_1$. *Гантелей* будем называть граф, состоящий либо из двух нечётных циклов, пересекающихся ровно по одной вершине, либо из двух нечётных циклов, соединённых путём любой длины.

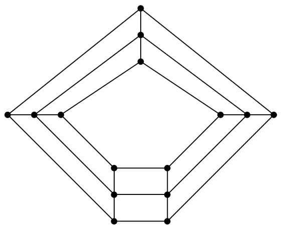


Рис. 2. Граф $C_5 \times P_2$

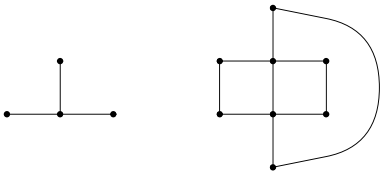


Рис. 3. Граф и его удвоение

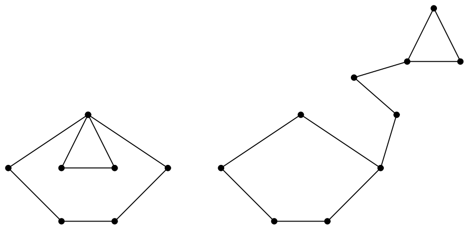


Рис. 4. Гантели

1 *Примеры*

1.1. Убедитесь, что магических графов меньше чем с 5 вершинами не существует, за исключением графа P_1 (одно ребро).

1.2. Докажите, что любой двудольный граф с нечётным числом вершин — не магический. А как у этих графов с полумагичностью?

1.3. Исследуйте следующие графы на полумагичность и магичность (ответы могут зависеть от n и m):
a) K_n; b) $K_{m,n}$; c) $P_n \times P_1$; d) $P_n \times P_m$ при n, $m > 1$; e) $C_n \times P_1$; f) $C_n \times P_m$, $n \geqslant 3$, $m > 1$;
g) цикл из $2n$ вершин, в котором противоположные вершины попарно соединены.

2 *Полумагические графы*

2.1. Докажите, что если полумагический граф G содержит чётный цикл, то в G найдётся полумагический скелет (т. е. скелет, который как самостоятельный граф является полумагичским графом), содержащий не все рёбра этого цикла.

2.2. Докажите, что если полумагический граф G содержит гантелю, то в G найдётся полумагический скелет, содержащий не все рёбра этой гантели.

2.3. Докажите, что в любом полумагическом графе можно выбрать 1-2-скелет.

2.4. Основная теорема о полумагических графах. Докажите, что граф тогда и только тогда является полумагическим, когда в нём любое ребро принадлежит некоторому 1-2-скелету.

Следующие задачи посвящены выяснению вопроса, в каких случаях граф содержит 1-2-скелет. Сначала разберёмся с 1-скелетами (в которых каждая вершина имеет степень 1). Будем называть граф *мягким*, если он не содержит 1-скелета, а в противном случае будем называть его *твёрдым*. Мягкий граф будем называть *насыщенным*, если при добавлении в него произвольного ребра он становится твёрдым. Например, полный граф с нечётным числом вершин — мягкий и насыщенный.

Пусть G — произвольный граф, S — некоторое множество его вершин. Через $G \setminus S$ обозначим граф, полученный удалением из G всех вершин множества S и их рёбер.

2.5. Пусть G — насыщенный мягкий граф, S — множество всех вершин в нём, каждая из которых соединена рёбрами со всеми остальными вершинами. Докажите, что компоненты связности графа $G \setminus S$ являются полными графами.

2.6. Основная теорема о мягких насыщенных графах. Граф G — мягкий и насыщенный тогда и только тогда, когда либо

a) G полный граф с нечётным числом вершин; либо

b) число вершин графа G чётно и в нём можно выделить такие непересекающиеся полные подграфы S_0, G_1, G_2, $\ldots$, G_k, где $k = |S_0| + 2$, что при всех i в каждом G_i число вершин нечётно и каждая вершина G_i соединена ребром со всеми вершинами S_0, и никаких других рёбер в графе нет.

2.7. Докажите, что граф G твёрдый тогда и только тогда, когда для каждого подмножества S множества вершин графа G граф $G \setminus S$ имеет не более $|S|$ нечётных компонент связности.

2.8. Докажите, что граф G обладает 1-2-скелетом в том и только том случае, если для каждого подмножества S множества вершин графа G граф $G \setminus S$ имеет не более $|S|$ изолированных вершин.

3 Магические графы

3.1. Докажите, что любой магический граф обладает двумя свойствами:

(1) Любое его ребро принадлежит какому-нибудь 1-2-скелету.

(2) Любая пара его рёбер разделяется каким-нибудь 1-2-скелетом.

3.2. Докажите обратное утверждение: любой граф, удовлетворяющий этим двум условиям, является магическим.

3.3. Граф G' получен из магического графа G добавлением нового ребра, причём это ребро принадлежит некоторому 1-2-скелету графа G'. Докажите, что граф G' — магический.

3.4. Граф G состоит из двух (неизоморфных) компонент связности, каждая содержит не меньше 3 вершин. Обе компоненты являются магическими графами. Верно ли, что граф G обязательно является магическим?

3.5. a) Если полумагический граф G не содержит изолированных рёбер, и для любого ребра e найдётся 1-2-скелет, циклическая часть которого не содержит e, то удвоение G — магический граф.

b) Пусть G — полумагический граф без изолированных рёбер, а H — произвольный граф без изолированных вершин и изолированных рёбер, то граф $G \times H$ — магический.

3.6. Дан граф G, в котором не менее 4 вершин. Граф G_1 получен добавлением к G одной новой вершины, которая соединена со всеми вершинами G. Докажите, что G_1 магический тогда и только тогда, когда G имеет 1-2-скелет и не имеет изолированных рёбер.

3.7. a) Если в графе $n \geqslant 5$ вершин и степени всех вершин не меньше $\frac{n}{2} + 1$, то граф магический.

b) Существуют неполумагические графы со сколь угодно большим числом вершин n, у которых минимальная степень вершины равна $n/2$.

3.8. Пусть G — связный магический граф с $n \geqslant 5$ вершинами и r рёбрами. Тогда $r > \frac{5}{4}n$.

3.9. Для каждого $n = 5$, 6, 7, 8 приведите пример связного магического графа с n вершинами и r рёбрами, где r — наименьшее натуральное число, удовлетворяющее неравенству $r > \frac{5}{4}n$.

3.10. Постройте такой граф для произвольного $n \geqslant 5$.

3.11. Докажите, что связный магический граф с n вершинами и r рёбрами существует для любой пары n, r, в которой $\frac{5}{4}n < r \leqslant \frac{n(n+1)}{2}$.

Промежуточный финиш

4 Однородные графы

Однородные графы степени 1 и 2 устроены исключительно примитивно и вопрос об их магичности решается очевидным образом. Поэтому ограничимся далее случаем степеней, не меньших 3.

Назовём *псевдоциклом* набор рёбер, образующий чётный цикл или гантелю (напомним, что оба цикла в гантеле — нечётные).

Рассмотрим некоторый чётный цикл. Расставим мысленно на его рёбрах попеременно числа 1 и -1, а на всех рёбрах, не входящие в этот цикл, — нули. Будем говорить, что два ребра *слабо разделяются* этим циклом, если они при этой расстановке получат разные веса. Аналогично, выбрав некоторую гантелю, расставим на ней числа ± 1 и ± 2 как на рис. 5 при $a = 1$, а на не вошедших в неё рёбрах расставим нули. Два ребра *слабо разделяются* этой гантелей, если они при этой расстановке получат разные веса. Наконец, будем говорить, что два ребра *слабо разделяются псевдоциклами*, если существует чётный цикл или гантеля, слабо разделяющая эти рёбра.

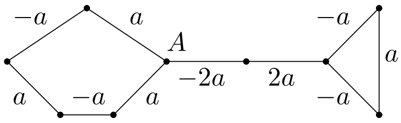


Рис. 5. Знакопеременные веса рёбер гантели

4.1. Докажите, что в однородном графе степени $d \geqslant 3$ любое ребро содержится в псевдоцикле.

4.2. Докажите, что однородный граф степени $d \geqslant 3$ является магическим тогда и только тогда, когда в нём любые два ребра слабо разделяются псевдоциклами.

4.3. Докажите следующую теорему: Пусть G — однородный граф степени $d \geqslant 3$, и $G_1, \ldots, G_k$ — его компоненты связности. Тогда G — магический граф тогда и только тогда, когда все G_i — магические графы.

Назовём индексом рёберной связности $\ell(G)$ графа G наименьшее число рёбер, которые необходимо из него выкинуть, чтобы он потерял связность.

4.4. Пусть G — связный однородный двудольный граф. Докажите, что его магичность или немагичность зависят только от величины $\ell(G)$ и проведите полное исследование этой зависимости.

5 *Добавления*

5.1. Добавление к задаче 1.3.а. Граф называется *супермагическим*, если на нём существует магическая расстановка, веса рёбер в которой — последовательные натуральные числа.

При каких n граф K_n является супермагическим?

5.2. Добавление к задаче 3.7. В графе 2009 вершин, степень каждой не меньше 1006. В графе удалили не более 500 рёбер. Докажите, что граф остался магическим.

Решения

1 Примеры

1.1. Если в графе на 4 вершинах 1 или 2 ребра, то в нём есть изолированные вершины. В любом графе с 3 или 4 рёбрами есть две смежные вершины степени 2, что противоречит магичности. Если рёбер 6, то это полный граф K_4, обсуждавшийся в задаче 1.3a). Наконец, если рёбер 5, то граф представляет собой цикл $ABCD$ с диагональю AC. Тогда $2s$ — т. е. сумма весов рёбер при вершинах A и C — это сумма весов всех рёбер графа с удвоенным весом ребра AC. С другой стороны, $2s$ — сумма весов рёбер при вершинах B и D, т. е. сумма весов всех рёбер, кроме AC. Но это значит, что ребро AC имеет нулевой вес, что запрещено.

1.2. Ответ: граф не полумагический. Пусть первая доля содержит k вершин, вторая — ℓ вершин, s — сумма весов рёбер у каждой вершины. Если граф полумагический, то сумма весов всех рёбер графа равна сумме весов рёбер, выходящих из вершин первой доли, т. е. ks, и она же равна сумме весов рёбер, приходящих во вторую долю, т. е. ℓs. Значит, $\ell = k$, что невозможно, если общее число вершин нечётно.

1.3. a) Ответ: граф всегда полумагический, магическим он является при $n = 2$ и при $n > 5$.

Полумагичность очевидна. Как и во всяком однородном графе, можно все веса взять равными единице.

При $n = 3$ граф не магический — это тоже очевидно.

При $n = 4$ имеем 4 вершины A, B, C, D. Допустим, что граф магический, пусть s — сумма весов рёбер, сходящихся в одной вершине. Тогда $2s$ — т. е. сумма весов рёбер при вершинах A и C — это сумма весов всех рёбер графа без веса ребра CD, но с удвоенным весом ребра AC. Делая аналогичный подсчёт для вершин B и D, находим, что веса рёбер AC и BD равны.

При $n > 5$ граф магический. Это можно установить следующим образом. Поскольку граф однороден, мы можем рассматривать произвольные (не обязательно положительные) веса рёбер. (В регулярном графе мы всегда можем сделать веса положительными, добавив ко всем весам одну и ту же большую положительную константу.) Опишем конструкцию построения магических меток однородного графа с помощью чётных циклов.

> Выпишем все чётные циклы, являющиеся подграфами нашего графа, и пронумеруем их числами от 1 до N (где N — их количество). Для k-го цикла в нашем списке назначим веса его рёбер — попеременно плюс и минус 3^k, эти веса поставим в качестве меток возле соответствующих рёбер. После того как мы просмотрели все циклы, сложим все метки, стоящие около каждого ребра.

Докажем, что полученная разметка рёбер графа магическая. Действительно, каждый цикл даёт нулевой суммарный вклад весов в каждую вершину, поэтому сумма весов каждой вершины равна нулю. Проверим, что все веса различны. Для каждого ребра выпишем список номеров тех циклов, в которые входит это ребро. Очевидно, что для любых двух рёбер графа существует чётный цикл, содержащий лишь одно из них. Следовательно, для любых двух рёбер списки номеров циклов не совпадают. Но тогда суммы весов, назначенные с помощью этих циклов, для разных рёбер попарно не равны. Это следует из того, что каждый такой суммарный вес можно трактовать как N-значное число в троичной системе счисления, в которой используются цифры 0 и ± 1. Несовпадение списков означает, что полученные числа различаются в каких-то разрядах троичной записи и поэтому не равны.

b) Ответ: граф полумагический только при $m = n$. При $m = n > 2$ он магический.

Для полумагичности необходимо, чтобы числа m и n были равны. *На балу каждая дама танцевала с пятью кавалерами, а каждый кавалер с пятью дамами. Докажите, что дам и кавалеров было поровну.* Ну или что-то в этом роде.

При $m = n$ граф однородный и потому полумагический. При $m = n = 2$ он не магический, это очевидно. А при $m = n > 2$ граф магический, в чём можно убедиться конструкцией аналогичной предыдущему решению.

c) Ответ: граф полумагический, но не магический.

Для полумагичности достаточно расставить на двух крайних рёбрах двойки, а на остальных единицы. Чтобы убедиться, что граф не магический, достаточно посмотреть на ребро, у которого обе вершины степени 2, и смежные с ним рёбра.

d) Ответ: граф магический, если m или n нечётны, и не полумагический, если и m, и n чётны.

При чётных n и m граф двудольный и имеет $(n+1)(m+1)$ вершин (нечётное число). По утверждению задачи 1.2 граф не может быть полумагическим.

Докажем, что при нечётном n и $m > 1$ граф магический.

Сначала рассмотрим случай $m = 2$. Стартовая полумагическая расстановка неотрицательных весов на графе показана на рис. 6: жирные рёбра имеют вес $2M$, пунктирные — вес 0, остальные рёбра имеют вес M, где M — большое число, которое мы выберем чуть позже. Это ещё не доказательство полумагичности, поскольку некоторые веса нулевые, а граф не однородный.

Теперь мы выполним основную конструкцию построения весов с помощью чётных циклов (см. решение задачи 1.3*a*), но с тремя поправками (всё-таки наш граф не однородный). Первая поправка состоит в том, что мы будем рассматривать не все циклы, а только 4-циклы (стороны клеточек). Вторая поправка состоит в том, что итоговый вес ребра мы положим равным сумме веса, назначенного с помощью основной конструкции, и веса этого же ребра

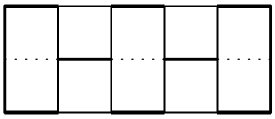


Рис. 6. Почти полумагические веса на графе $P_n \times P_2$

в стартовой расстановке. Наконец, третья поправка — назначая положительные и отрицательные веса рёбер каждого цикла (здесь у нас есть произвол, с какого знака начинать), мы будем следить, чтобы рёбра, имеющие нулевой вес в стартовой расстановке, всегда получали положительный вес. Наконец, выберем число M настолько большим, чтобы в результате выполнения всей этой конструкции все веса рёбер оказались бы положительными и различными. Полученная расстановка весов будет магической.

В случае, когда n — нечётно, $m > 2$, мы действуем аналогично. Стартовые расстановки весов показаны на рис. 7 (нечётная сторона вертикальна).

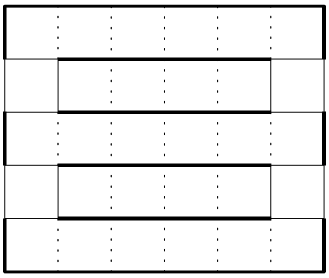


Рис. 7. Полумагические веса на графе $P_n \times P_m$

e) Ответ: граф всегда полумагический. Магическим он является лишь при чётном n.

Граф можно представлять себе как набор вершин и рёбер n-угольной призмы.

Пусть $n = 2k$. В графе есть очевидные циклы длины 4 (контуры граней) и два $2k$-цикла (контуры оснований). Здесь также выполнено свойство, что для любых двух рёбер найдётся чётный цикл, содержащий лишь одно из них. Таким образом, применима основная конструкция для однородных графов.

Пусть $n = 2k + 1$. Проверим, что граф не магический. Пусть d — сумма весов рёбер, сходящихся в вершине. Тогда, как нетрудно видеть, сумма весов всех рёбер графа равна nd. Обозначим наш граф-призму через $A_1A_2 \ldots A_{2k+1}B_1 \ldots B_{2k+1}$. Сумма весов рёбер, выходящих из вершин A_1, A_3, …, A_{2k+1}, B_2, B_4, …, B_{2k}, равна nd и при этом представляет собой сумму весов всех рёбер графа без ребра B_1B_{2n+1}, но с ребром A_1A_{2n+1}, учтённым дважды. Следовательно, веса рёбер A_1A_{2n+1} и B_1B_{2n+1} равны.

f) Ответ: граф магический. Решение аналогично 1.3d). Пусть A_1, …, A_{m+1} — вершины графа P_m. Стартовая полумагическая расстановка неотрицательных весов на графе $C_n \times P_m$ выглядит следующим образом: рёбра всех подграфов вида $C_n \times A_i$ имеют вес 2, остальные рёбра имеют вес 0. В качестве набора чётных циклов опять рассматриваем 4-циклы.

g) Ответ: граф всегда полумагический, а при нечётных n он ещё и магический.

Полумагический он, потому что однородный. При $n = 2$ это граф K_4, мы его обсуждали в задаче 1.3 a). При нечётном n граф магический, поскольку работает основная конструкция: есть хороший запас чётных циклов — 4-циклы, содержащие соседние диаметры, и $(n + 1)$-циклы вида «полукруг».

При чётном n граф не магический. Пусть $A_1A_2 \ldots A_nB_n \ldots B_2B_1$ — вершины цикла. Рёбра, выходящие из вершин A_i, B_i, где i пробегает все нечётные индексы, — это все рёбра графа, кроме A_nB_n, причём ребро A_1B_1 учтено дважды. Отсюда следует, что в любой полумагической расстановке весов противоположные рёбра $2n$-цикла имеют одинаковый вес.

2 *Полумагические графы*

2.1. Пусть a — минимальный вес ребра в данном чётном цикле. Обходя цикл, будем попеременно то уменьшать, то увеличивать на a вес рёбер цикла. В результате вес некоторых рёбер станет нулевым — сотрём их. Оставшийся граф с полученной расстановкой весов и будет искомым полумагическим скелетом.

2.2. Пусть A — вершина одного из нечётных циклов гантели, к которой прикреплена ручка гантели (или второй цикл, если ручки нет). Рассмотрим следующее назначение весов рёбер гантели. Будем обходить нечётный цикл, начиная с вершины A, и попеременно присваивать рёбрам веса $\pm a$. Вернувшись в вершину A, мы получим, что оба ребра данного цикла, сходящиеся в вершине A, имеют вес a. Продолжим движение по ручке, попеременно назначая веса её рёбер

$\mp 2a$. Дойдя до второго цикла, обойдём его, продолжая назначать веса $\pm a$. В результате мы получим полумагическое назначение весов с нулевой суммой в каждой вершине (см рис. 5).

Прибавим построенные веса к уже имеющимся весам рёбер гантели, причём подберём a так, чтобы все веса получились в результате неотрицательными и вес по крайней мере одного из рёбер стал равен нулю. Получится полумагическая разметка рёбер графа, причём все нулевые рёбра можно стереть (очевидно, изолированных вершин от этого появиться не может). Останется искомый полумагический скелет.

2.3. С помощью конструкций из решений задач 2.1, 2.2 мы можем последовательно уменьшать количество рёбер в графе, разрушая чётные циклы и гантели, и сохраняя при этом полумагичность. Заметим, что в силу полумагичности наш граф ни в какой момент не будет иметь висячих вершин (кроме вершин изолированных рёбер). Заметим также, что если в компонента связности графа имеет два нечётных цикла, то в ней можно найти чётный цикл или гантелю. Если же компонента содержит ровно один (нечётный) цикл и не имеет при этом висячих вершин, то ничего, кроме этого цикла, она содержать не может. Значит, в тот момент, когда все чётные циклы и гантели будут разрушены, граф будет представлять собой несколько изолированных рёбер плюс несколько изолированных (нечётных) циклов.

2.4. Проверим, что в полумагическом графе любое ребро принадлежит некоторому 1-2-скелету. Пусть G — любой из графов с минимальным числом рёбер, имеющий ребро e, не принадлежащее ни одному 1-2-скелету. Фиксируем полумагическую расстановку $\mathcal{W}$ весов на графе G. Возьмём произвольный 1-2-скелет и с помощью него построим ещё одну полумагическую расстановку весов $\mathcal{S}$: пусть каждое ребро из линейной части скелета имеет вес a, каждое ребро из циклической части — вес $a/2$, а рёбра, не входящие в скелет, (и в том числе e) имеют вес 0. Число a подберём таким образом, веса из расстановки $\mathcal{S}$ не превосходили соответствующих весов из расстановки $\mathcal{W}$ и чтобы хотя бы на одном ребре равенство достигалось. Теперь вычтем из весов расстановки $\mathcal{W}$ веса $\mathcal{S}$. Получится полумагическая расстановка весов, в которой не все веса равны нулю, так как вес ребра e не изменился. Если теперь стереть ребра нулевого веса, получится полумагический граф G', который является скелетом в G, содержит меньше рёбер, чем G, причём ребро e не принадлежит никакому 1-2-скелету G' (потому что «скелет моего скелета — мой скелет»). Это противоречит определению графа G. Следовательно, таких графов G не существует, чтд.

Теперь убедимся, что если в графе любое ребро принадлежит какому-нибудь 1-2-скелету, то граф — полумагический. Для каждого 1-2-скелета поставим на всех рёбрах его циклической части вес 1, а на всех изолированных рёбрах — вес 2. Тогда вклад этого скелета в каждую вершину будет одинаковым. Перебирая все 1-2-скелеты, просуммируем веса, полученные таким способом. Это и есть требуемая полумагическая расстановка весов.

2.5. Мы приводим решение из [1, § 3.1.2]. Пусть A, B, C — вершины из $G \setminus S$, причём B соединено ребром с A и C. Для доказательства утверждения задачи достаточно проверить, что в этом случае в граф G содержит ребро AC. Допустим, что это не так. По определению множества S, в графе G найдётся вершина D, не соединённая с B ребром. Если к графу G добавить ребро AC, то в силу насыщенности, полученный граф будет обладать 1-скелетом и ребро AC будет принадлежать этому скелету. Покрасим этот скелет в красный цвет. Аналогично при добавлении ребра BD найдём синий 1-скелет, содержащий ребро BD. Сейчас мы из этих двух скелетов соберём 1-скелет графа G и получим противоречие.

Объединим эти скелеты; кратности рёбер, которые оказались одновременно красными и синими, будем считать равными единице. Получим 1-2-скелет графа $G \cup AC \cup BD$. Очевидно, рёбра AC и BD принадлежат циклической части этого 1-2-скелета, причём все циклы в ней чётные, так как красные и синие рёбра в циклах чередуются.

Если рёбра AC и BD лежат в разных циклах, то искомый 1-скелет построить совсем легко: возьмём за основу красный 1-скелет и все красные рёбра того цикла, где лежит ребро AC, заменим на синие рёбра этого же цикла.

Пусть теперь рёбра AC и BD лежат в одном цикле γ. Начнём движение из вершины B по синему ребру BD и дальше вдоль цикла γ, пока не дойдём до вершины A или C. Пусть это будет A, эти случаи совершенно аналогичны. Поскольку красное ребро, начинающееся в вершине A, — это AC, мы в процессе движения пришли в A по синему ребру. Таким образом, пройденный путь из B в A начинается и кончается синим ребром. Возьмём тогда синий скелет, заменим все синие рёбра пройденного пути на красные, а также добавим ребро AB. Получится 1-скелет графа G.

2.6. Мы приводим решение из [1, § 3.1.2]. Если число вершин в насыщенном мягком графе G нечётно, то очевидно, что он полный. Пусть число вершин в G чётно и пусть S — множество всех вершин G, которые соединены со всеми остальными вершинами, s — их количество; G_1, G_2, …, G_k — компоненты связности графа $G \setminus S$. По утверждению предыдущей задачи мы знаем, что они являются полными графами.

Если в $G \setminus S$ являются нечётными не более s компонент, то 1-скелет находится легко. Рассмотрим тогда случай, когда в графе $G \setminus S$ не менее $s + 1$ нечётной компоненты, а с учётом того, что число вершин в G чётно — не менее $s + 2$ компонент. Если нечётных компонент оказалось больше $s + 2$, соединим любые две из них ребром, получится граф G_1, для которого верно, что граф $G_1 \setminus S$ имеет больше s нечётных компонент связности. В таком графе не может быть 1-скелетов (это очевидно, и к тому же следует из простой части утверждения задачи 2.7), что противоречит насыщенности графа G.

Итак, у графа G ровно $s + 2$ нечётные компоненты. По аналогичным соображениям у него не может быть при этом чётных компонент.

2.7. Это утверждение — классическая теорема Татта (W. Tutte) Мы приводим её доказательство, следуя изложению в [1, § 3.1.2].

Если в графе G нашлось такое множество вершин S, что в графе $G \setminus S$ больше $|S|$ нечётных компонент связности, то граф G мягкий. Это очевидно.

Проверим обратное утверждение. Допустим, что для каждого подмножества S множества вершин графа G граф $G \setminus S$ имеет не более $|S|$ нчётных компонент связности, но при этом граф G мягкий.

Число вершин графа G должно быть чётно, так как в противном случае при $S = \varnothing$ сразу получаем противоречие. Добавим к графу G несколько рёбер, чтобы получился насыщенный мягкий граф G'. Пусть S' — множество вершин, смежных с каждой вершиной G', s — их количество. Поскольку количество вершин в графе G' такое же как и в G, т. е. чётно, то по основной теореме о мягких насыщенных графах, граф $G' \setminus S'$ содержит $s+2$ нечётные компоненты (нам важно, что их больше s), каждая из которых — полный граф. Уберём те рёбра, которые мы добавили, делая граф насыщенным. Возможно, при этом некоторые компоненты графа $G' \setminus S'$ распадутся на части, но в любом случае хотя бы один из «осколков» нечётной компоненты будет нечётным и общее число нечётных компонент будет больше s. Таким образом, построенное множество S' опровергает основное обсуждаемое свойство графа G.

2.8. Пусть n — количество вершин графа G. Построим новый граф G' с $2n$ вершинами: каждой вершине v графа G соответствуют две вершины v' и v'' в G'; каждому ребру uv в графе G соответствуют два ребра в графе G' — $u'v''$ и $u''v'$ (других рёбер в G' нет). Ясно, что G' — двудольный граф, и количество его рёбер в два раза превосходит число рёбер в G.

Заметим, что существование 1-2-скелета в исходном графе равносильно тому, что в графе G' найдётся паросочетание из n рёбер. В самом деле, для каждого цикла $v_1v_2\dots v_\ell$, принадлежащего скелету, в графе G' присутствуют рёбра $v_1'v_2''$, $v_2'v_3''$, $\dots$, $v_\ell'v_1''$; аналогично, для изолированного ребра uv данного скелета в G' есть рёбра $u'v''$ и $v'u''$. Ясно, что все такие рёбра образуют полное паросочетание. Обратно, если дано полное паросочетание графа G', то по нему нетрудно построить 1-2-скелет в G. Например, рёбрам $u'v''$, $v'w''$, $w'z''$, $z'u''$ паросочетания соответствует цикл $uvwz$ в графе G, а рёбрам $u'v''$ и $v'u''$ — изолированное ребро uv в скелете.

Теперь рассмотрим условие о том, что для каждого подмножества S множества вершин графа G граф $G \setminus S$ имеет не более $|S|$ изолированных вершин. Сформулируем его для графа G'. Возьмём любой набор S вершин графа G. Что значит, что при их выкидывании вершина u осталась изолированной? Это значит, что в графе G' все соседи вершины u' лежат в множестве S''. Если после выкидывания набора S образовалось $k > |S|$ изолированных вершин, то в графе G' нарушается условие леммы Холла: у k вершин не более $|S|$ соседей, что меньше, чем k. Ясно, что верно и обратное. Таким образом, наше условие равносильно выполнению в графе G' леммы Холла, то есть, снова равносильно наличию в G' полного паросочетания.

3 Магические графы

3.1. (1) Этим свойством обладают все полумагические графы.

(2) Докажем более общий факт: если в полумагическом графе есть такой полумагический набор весов рёбер, в котором веса каких-то двух рёбер e_1 и e_2 не равны, то рёбра e_1 и e_2 разделяются 1-2-скелетом.

Это устанавливается аналогично решению задачи 2.4. Выберем минимальный граф; фиксируем ту расстановку, где веса не равны; отнимем подходящим образом веса у рёбер, принадлежащих скелету; получится меньший граф. Так как исходный граф мы выбрали минимальным, одно из рёбер e_1, e_2 должно было получить при этом нулевой вес и было стёрто. Тогда в оставшемся графе по утверждению задачи 2.4 второе из этих рёбер принадлежит некоторому 1-2-скелету, который будет также и скелетом в исходном графе и будет разделять рёбра e_1 и e_2.

3.2. Пронумеруем все 1-2-скелеты и для k-го скелета положим вес рёбер циклической части равным 3^k, а вес рёбер линейной части — $2 \cdot 3^k$. Теперь для каждого ребра найдём сумму его весов по всем содержащим его 1-2-скелетам. Получится полумагическая расстановка весов, которая является магической в силу единственности троичной записи натурального числа.

3.3. Следует из 3.2.

3.4. Ответ: нет, граф G может оказаться не магическим. Мы почерпнули этот пример в [5]. На рисунке 8 показаны два магических графа. Для любой магической расстановки весов ребра, нарисованные пунктиром, должны иметь вес $r/2$, где r — суммарный вес рёбер, сходящихся в одной вершине.

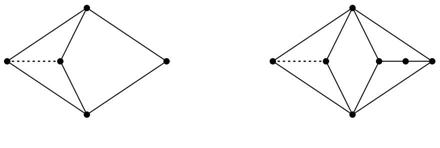


Рис. 8. Объединение магических графов — не всегда магический граф

3.5. а) Удвоение G^2 состоит из двух экземпляров графа G_1 и G_2 графа G и множества рёбер E между соответственными вершинами. Соответственные рёбра в компонентах G_1 и G_2 будем называть *параллельными.* Рёбра из множества E будем называть *вертикальными.* Подграф в G^2, состоящий из двух соответственных компонент в G_1 и G_2, назовём *дублированным.*

Сначала опишем конструкцию *поворота параллельных рёбер* в удвоенном графе. Пусть подграф H графа G^2 представляет собой объединение подграфов, лежащих в компонентах G_1 и G_2, таких что эти подграфы содержат параллельные ребра A_1B_1 и A_2B_2. Убёрем в подграфе H рёбра A_1B_1, A_2B_2 и добавим рёбра A_1A_2 и B_1B_2. Полученный подграф назовём H'. Будем говорить, что подграф H' получен из H с помощью поворота параллельных рёбер. Очевидно, что подграфы H и H' одновременно являются (или не являются) 1-2-скелетами.

Теперь докажем, что граф G^2 из условия задачи является магическим. Для этого применим критерий магичности — утверждение задач 3.1–3.2.

(1) Любое ребро принадлежит 1-2-скелету. Для рёбер из G_1 (и из G_2) это очевидно: в качестве скелета берём 1-2-скелет в G_1, содержащий это ребро, в объединении с его дублем в G_2. Для вертикальных рёбер следует взять поворот параллельных рёбер подходящего дублированного 1-2-скелета.

(2) Любая пара рёбер разделяются 1-2-скелетом. В случае, когда оба ребра e_1 и e_2 из G_1 (или оба из G_2), возьмём дублированный 1-2-скелет, содержащий ребро e_1. Если этот скелет не разделяет ребра e_1 и e_2, оба этих ребра принадлежат скелету. Тогда выполним поворот ребра e_2 и параллельного ему, получится скелет, разделяющий рёбра.

В случае, когда ребро e_1 из G_1, а ребро e_2 из G_2, возьмём в G_1 1-2-скелет, содержащий e_1 (он существует в силу утверждения задачи 2.4), а в G_2 — 1-2-скелет, не содержащий e_2 (существует по условию). Их объединение есть искомый разделяющий 1-2-скелет.

Если e_1 из G_1, а e_2 — вертикальное, подойдёт дублированный скелет, содержащий ребро e_1.

Наконец, если оба ребра — A_1A_2 и B_1B_2 — вертикальные, то поскольку в графе G не было изолированных рёбер, в G_1 найдётся ребро A_1X_1 (где $X_1 \neq B_1$) или B_1Y_1 (где $Y_1 \neq A_1$). Выберем дублированный скелет, содержащий это ребро, и повернём это ребро и параллельное ему.

b) Доказывается аналогично п. a).

3.6. Утверждение задачи мы взяли в [6]. Доказательство, приведённое там, опирается на критерий магичности графа, который не встречался в данной серии задач. Задача содержит два утверждение, сложным является утверждение «тогда» — *если G' — магический граф, то граф G имеет 1-2-скелет и не имеет изолированных рёбер и изолированных вершин.* Мы приводим доказательства этого утверждения, найденные участниками конференции.

Доказательство 1. Если бы в G была изолированная вершина, то в графе G' она оказалась бы висячей и граф G' не мог бы быть магическим. Если бы в G было изолированное ребро, то концы этого ребра в графе G' оказались бы смежными вершинами степени 2 и граф G' не мог бы быть магическим.

Допустим, что в G не существует 1-2-скелета.

Обозначим новую вершину графа G' через S. Возьмём какой-нибудь 1-2-скелет K графа G', можно считать, что все циклические компоненты в нём суть нечётные циклы. Рассмотрим компоненту этого скелета, содержащую вершину S. Эта компонента не может быть нечётным циклом, так как иначе при удалении из него вершины S мы могли бы разбить остальные вершины этого цикла на пары и вместе с остальными частями рассматриваемого скелета получили бы скелет G. Значит, эта компонента является изолированным ребром SA_1. Сейчас мы построим в графе G два множества вершин — $\mathcal{A} = \{A_1, \ldots, A_n\}$ и $\mathcal{B} = \{B_1, B_2, \ldots, B_n\}$, удовлетворяющих следующим условиям: все ребра A_iB_i $(1 \leqslant i \leqslant n)$ принадлежат скелету K, и все рёбра из вершин A_i, ведут в множество $\mathcal{B}$.

Для начала конструкции возьмём $\mathcal{A} = \{A_1\}$, и положим $B_1 = S$. Допустим, что уже построены множества $\mathcal{A} = \{A_1, \ldots, A_k\}$ и $\mathcal{B} = \{B_1, \ldots, B_k\}$, Допустим, что из множества $\mathcal{A}$ выходит какое-либо ребро, идущее вне $\mathcal{A} \cup \mathcal{B}$, скажем, A_kB_{k+1}. Вершина B_{k+1} принадлежит некоторой компоненте скелета K. Если это нечётный цикл, то мы легко можем перестроить скелет K, чтобы получился полноценный 1-2-скелет графа G, что невозможно.

> Для этого рассмотрим кратчайший путь от B_1 до B_{k+1}, идущий по вершинам $\mathcal{A} \cup \mathcal{B}$ и в котором вершины множеств $\mathcal{A}$ и $\mathcal{B}$ чередуются. Он имеет чётную длину. Выберем в нём все рёбра с чётным номером (последнее из них оканчивается вершиной B_{k+1}) и разобьём на пары все остальные вершины нечётного цикла.

Значит, можно считать, что вершина B_{k+1} принадлежит изолированному ребру $B_{k+1}A_{k+1}$ скелета K. Поместим тогда вершину B_{k+1} в множество $\mathcal{B}$, а вершину A_{k+1} — в множество $\mathcal{A}$.

Будем продолжать увеличивать множества $\mathcal{A}$ и $\mathcal{B}$ описанным образом, пока это возможно. В конце концов окажется, что из множества $\mathcal{A}$ все рёбра ведут только в $\mathcal{A} \cup \mathcal{B}$. Предположим, что две вершины A_i и A_j соединены ребром. Рассмотрим кратчайший путь между этими вершинами, в котором вершины из $\mathcal{A}$ и $\mathcal{B}$ чередуются (существование такого пути легко усмотреть из процесса построения пары множеств). Вместе с ребром A_iA_j он образует нечётный цикл, и тогда скелет K перестраивается в скелет графа G способом, аналогичным описанному выше.

Итак, требуемые множества $\mathcal{A}$ и $\mathcal{B}$ построены. Заметим теперь, что суммы весов всех вершин в этих множествах равны (ибо в них поровну вершин). С другой стороны, сумма весов всех вершин из $\mathcal{A}$ складывается из весов всех рёбер вида A_iB_j, в сумма весов вершин $\mathcal{B}$ — из тех же рёбер, а так же из рёбер вида B_1B_i (напомним, что вершина $B_1 = S$ соединена со всеми вершинами графа G)! Противоречие.

Доказательство 2 (А. Цыбышев). Рассмотрим магическую расстановку весов на рёбрах графа G'. Мысленно забудем про рёбра, выходящие из S, и будем временно рассматривать только рёбра графа G. Применим к ним алгоритм «избавления» от чётных циклов и гантелей, описанный в решении задач 2.1 и 2.2. В результате останется граф F с весами на рёбрах, в котором нет чётных циклов и гантелей, а сумма весов в каждой вершине такая же как в начале. Вернём обратно рёбра из вершины S — получится полумагический граф F'.

Если в графе F найдётся изолированная вершина A, то в графе F' вершина A будет висячей, что противоречит его полумагичности.

Пусть в F есть висячая вершина A, и пусть B — соседняя с ней вершина. Найдём в графе F' 1-2-скелет, содержащий ребро SB. Очевидно, он должен состоять из цикла $SABS$, а также других циклов и изолированных рёбер. Но тогда эти циклы, изолированные рёбра и ребро AB образуют 1-2-скелет графа G, что и требовалось.

Осталось разобрать случай, когда в F нет ни изолированных, ни висячих вершин. Поскольку в нём нет также чётных циклов и гантелей, все его компоненты — нечётные циклы. Но тогда они образуют искомый 1-2-скелет графа G.

Теперь докажем вторую часть утверждения задачи. Проверим, что если в графе G есть 1-2-скелет и нет изолированных рёбер, то граф G' удовлетворяет свойствам задачи 3.1 (и следовательно, магический). Обозначим новую вершину графа G' через A, а 1-2-скелет в графе G (любой, если их несколько) — через S.

1) Проверим, что каждое ребро G' принадлежит 1-2-скелету.

Случай а). Интересующее нас ребро BC лежит в графе G.

а1) Если ребро BC принадлежит линейной части скелета S, заменим в S ребро BC на треугольник ABC — получится скелет графа G', содержащий BC.

а2) Если ребро BC принадлежит циклической части скелета S, скажем, циклу $BCD\dots B$, заменим в S ребро CD на два ребра AC, AD — получится скелет графа G', содержащий BC (в слегка увеличенном цикле).

а3) Если ребро BC не принадлежит S и при этом вершины B и C принадлежат одной компоненте скелета — циклу $BD_1\dots D_pCE_1\dots E_qB$, сконструируем из этого цикла два новых: $BD_1\dots D_pCB$ и $AE_1\dots E_qA$ (при $q = 1$ второй цикл — это просто изолированное ребро), получится скелет графа G', содержащий BC.

а4) Если ребро BC не принадлежит S и при этом вершины B и C принадлежат разным компонентам скелета — $BB_1\dots B_pB$ и $CC_1\dots C_qC$, заменим их на один большой цикл $BB_1\dots B_pAC_1\dots C_qCB$.

Случай б). Интересующее нас ребро AB выходит из вершины A.

Если вершина B содержится в изолированном ребре BC 1-2-скелета, то заменим это ребро на цикл $ABCA$. Если же вершина B содержится в цикле $BD_1\dots D_pB$, то заменим его на цикл $ABD_1\dots D_pA$.

2) Проверим, что любые два ребра e и f разделяются 1-2-скелетами.

Случай а) Интересующие нас ребра принадлежат графу G.

а1) Одно из ребер, скажем, e, принадлежит циклической части скелета. Если ребро f принадлежит скелету, заменим цикл, в котором лежит ребро e, на увеличенный цикл, не содержащий ребра e (проходящий через вершину A, мы так делали в случае а2). Если ребро f не принадлежит скелету, заменим цикл, в котором лежит ребро e, на увеличенный цикл, содержащий ребро e.

а2) Одно из рёбер — e — лежит в линейной части скелета, а другое — тоже в линейной или вообще не принадлежит скелету. Добавим к скелету рёбра, соединяющие концы ребра e с вершиной A.

а3) Оба ребра не принадлежат скелету. В качестве разделяющего возьмём скелет, содержащий ребро e, построенный в первой части решения; при его построении добавлялись рёбра не принадлежащие графу G.

Случай б) Одно из рёбер лежит в G, другое — в G'. Мы оставляем читателю довести до конца этот несложный перебор. Следует помнить, что граф G имеет не менее 4 вершин и не имеет изолированных рёбер.

3.7. Мы взяли это утверждение в [3].

а) Возьмём любые рёбра e и f и докажем, что они разделяются некоторым 1-2-скелетом. Выкинем из графа две вершины — концы ребра f — и все рёбра, выходящие из них. В оставшемся графе $n-2$ вершины и степень каждой из них не меньше, чем $\frac{n}{2} - 1 = \frac{n-2}{2}$. Тогда, как известно, в этом графе найдётся цикл, проходящий по всем его вершинам (гамильтонов цикл). Этот цикл, вместе с ребром f, образует 1-2-скелет в исходном графе. Он разделяет рёбра e и f, так как f лежит в его линейной части, а e — нет.

b) Построим граф G на $n = 2k$ вершинах $X_1, \dots, X_k, Y_1, \dots, Y_k$, в котором проведены все рёбра вида X_iY_j и ребро Y_1Y_2. Степень каждой из вершин X_i не меньше $k = \frac{n}{2}$. Докажем, что G не является полумагическим графом.

Рассмотрим любой 1-2-скелет в G. Каждая из вершин X_i имеет в этом скелете либо одну, либо две смежные вершины среди Y_i. Поскольку вершин обеих типов поровну, то 1-2-скелет должен являться паросочетанием из k рёбер вида X_iY_j. Это значит, что ребро Y_1Y_2 не содержится ни в одном 1-2-скелете, т. е. G — не полумагический.

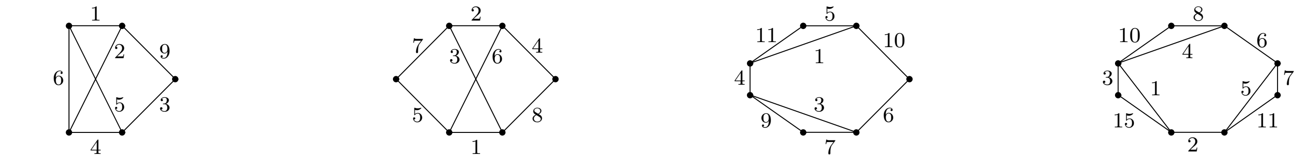
3.8. В магическом графе нет вершин степени 1, и более того, никакие две вершины степени 2 не соединены ребром. Пусть V — множество вершин степени 2 (возможно, пустое), а W — множество вершин степени 3 или больше. Обозначим сумму весов рёбер в каждой вершине через s. Сумма весов всех рёбер, выходящих из V, равна $s|V|$. С другой стороны, все эти рёбра имеют один из концов в W, поэтому сумма их весов не превосходит $s|W|$. Таким образом, $|V| \leqslant |W|$ (причём строго меньше, если внутри W есть хотя бы одно ребро). Далее, сумма степеней всех вершин не меньше, чем $2|V| + 3|W|$, поэтому в графе есть не меньше, чем $|V| + \frac{3}{2}|W|$ рёбер. Но $|V| + \frac{3}{2}|W| \geqslant \frac{5}{4}(|V| + |W|) = \frac{5}{4}n$, т. к. $|W| \geqslant |V|$.

Для того чтобы количество рёбер действительно равнялось $\frac{5}{4}n$, необходимо, чтобы не было рёбер с обоими концами в W, т.е. чтобы граф был двудольным. В этом случае $s|V| = s|W|$, т.е. $|V| = |W|$. Но количество рёбер между V и W, с одной стороны, равно $2|V|$, а с другой стороны, не меньше чем $3|W|$, т.е. $|V| \geqslant \frac{3}{2}|W|$, что невозможно. Таким образом, неравенство $r > \frac{5}{4}n$ доказано.

3.9. Пусть $n = 5$, 6, 7 или 8. На рисунке изображены магические графы с минимальным числом рёбер.

3.10. На рис. 10 a, b, c, e, f) изображены примеры магических графов с минимальным возможным количеством рёбер. Вид графа зависит от остатка от деления n на 4. При $n = 4k$ приведены два вида графов: двудольный и недвудольный, при $n = 4k + 2$ — только двудольный, в остальных двух случаях приведён пример недвудольного графа.

Доказательство магичности представленных графов состоит в рутинной проверке критерия магичности (задача 3.2). Мы не будем здесь делать эту проверку, но заметим, что есть обходной манёвр, который позволяет не делать такого перебора, и лишь немного не дотягивает до строгого доказательства, а именно, мы приведём магическую



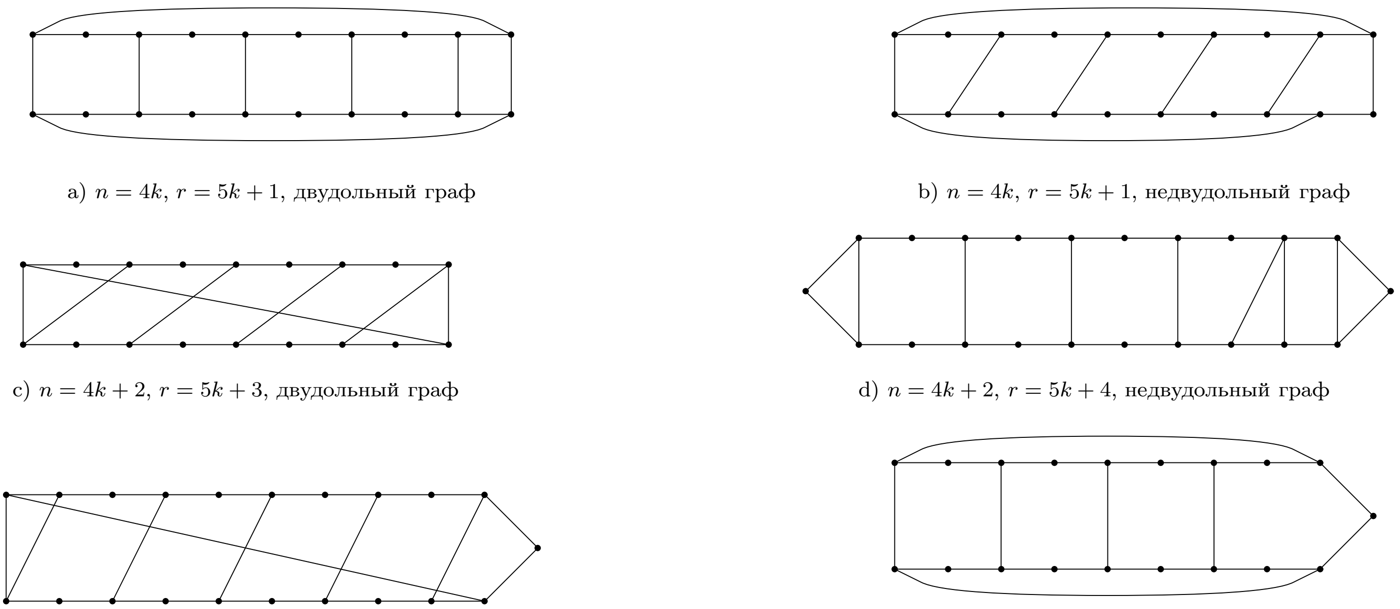
a) 5 вершин, 7 рёбер

b) 6 вершин, 8 рёбер

c) 7 вершин, 9 рёбер

d) 8 вершин, 11 рёбер

Рис. 9. Минимальные магические графы



a) $n = 4k$, $r = 5k + 1$, двудольный граф

b) $n = 4k$, $r = 5k + 1$, недвудольный граф

c) $n = 4k + 2$, $r = 5k + 3$, двудольный граф

d) $n = 4k + 2$, $r = 5k + 4$, недвудольный граф

e) $n = 4k + 1$, $r = 5k + 2$, недвудольный граф

f) $n = 4k + 3$, $r = 5k + 4$, недвудольный граф

Рис. 10. Примеры магических графов с минимальным числом рёбер

расстановку весов рёбер, в «достаточно типичном» случае. Мы ограничимся случаем $n = 4k + 3$, $r = 5k + 4$, $k = 2$; расстановка весов показана на рис 11.

3.11. Мы приводим решение по мотивам [4]. Пусть уже построен связный магический граф с n вершинами и r рёбрами, не являющийся полным. Если он недвудольный, то при добавлении к нему ещё одного (любого!) ребра он не утрачивает магичности.

Действительно, новое ребро e обязательно входит в некоторый цикл. Если этот цикл чётен, то припишем ребру e значение ε, а к остальным рёбрам цикла прибавим попеременно $\pm\varepsilon$, причём подберём ε так, чтобы все веса остались положительными и различными. Полученная расстановка весов на новом графе будет магической.

Пусть теперь ребро e входит в нечётный цикл. В силу недвудольности исходного графа, существует нечётный цикл, не содержащий e. Тогда e лежит в некоторой гантели (см. лемму в решении задачи 4.1.). И опять можно приписать ребру e вес ε, а к рёбрам гантели прибавить поправки $\pm\varepsilon$, $\pm 2\varepsilon$, чтобы расстановка осталась магической.

Таким образом, достаточно для каждого $n \geqslant 5$ построить «минимальный» недвудольный граф. Это было сделано в предыдущей задаче для $n \neq 4k + 2$ (см. рис. 10 b, e, f). Конструкции графов мы взяли в статье [4]. К сожалению, конструкция минимального недвудольного графа для $n = 4k+2$ в этой статье неверна. Кроме того, граф на рис. 10 b) при $n = 8$ не магический (в нём не разделяются 1-2-скелетами наклонное и нижнее ребро), пример магического графа при $n = 8$ показан на рис. 9 d), его изобрёл участник конференции А. Цыбышев. Мы не знаем, существует ли недвудольный магический граф с $4k + 2$ вершинами и $5k + 3$ рёбрами (при $k > 3$), поэтому для случая $5k + 3$ рёбер оставим двудольный пример, а конструкцию добавления ребер начнём с недвудольного графа, содержащего $5k + 4$ ребра. Этот недвудольный граф с $4k + 2$ вершинами и $5k + 4$ рёбрами показан на рис. 10 d). Пример расстановки весов на этом графе при $k = 3$ см. на рис. 12.

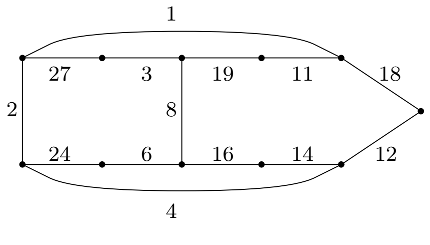


Рис. 11. $n = 4k + 3$, $r = 5k + 4$, $k = 2$

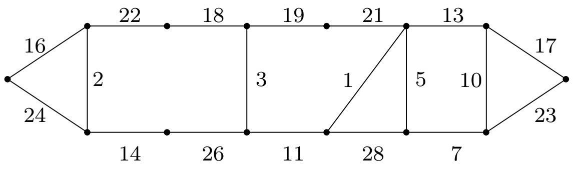


Рис. 12. $n = 4k + 2$, $r = 5k + 4$, $k = 3$

4 Однородные графы

4.1. Решение 1 (вокруг двудольности).

Лемма. Пусть в связном графе даны два нечётных цикла, один из которых содержит ребро e, а другой — нет. Тогда e содержится в чётном цикле или гантеле.

Доказательство. Если циклы не пересекаются или пересекаются по одной вершине, то ребро e, очевидно, содержится в гантеле. Рассмотрим случай, когда циклы пересекаются не менее чем по двум вершинам. Пусть X и Y — концевые вершины ребра e. Удалим ребро e из первого цикла, на оставшуюся часть этого цикла будем ссылаться как на путь XY. Пусть A и B — первая и последняя вершина пути XY, принадлежащие второму циклу, тогда отрезки пути XA и BY не пересекаются с циклом. Вершины A и B делят второй цикл на два пути разной чётности. Один из них дополняет пути XA и BY до нечётного пути $XABY$, который, вместе с ребром XY, образует чётный цикл.

Теперь обратимся к утверждению задачи. Не умаляя общности можно считать, что граф связный. Рассмотрим произвольное ребро e с концами A и B. Выкинем его из графа. Допустим сначала, что граф $G \setminus e$ распался на две компоненты связности. Поскольку степени всех вершин были больше 1, компоненты содержат более одной вершины. Ни одна из компонент не может быть двудольным графом. В самом деле, в двудольном графе сумма степеней вершин в обеих долях одинаковы; в нашей же компоненте сумма степеней в одной компоненте будет кратна d, а в другой (в той, куда попадёт конец ребра e) сумма степеней будет сравнима по модулю d с -1. Таким образом, в каждой компоненте есть нечётный цикл. Значит, ребро e содержится в гантеле.

Теперь предположим, что граф $G \setminus e$ связен. Рассмотрим произвольный путь из A в B в этом графе. Если он нечётен, то e содержится в чётном цикле. Пусть этот путь чётен (и, значит, e содержится в нечётном цикле). Тогда, если бы граф $G \setminus e$ был двудольным, то эти вершины попали бы в одну и ту же долю, что невозможно, ибо сумма степеней вершин в этой доле была бы сравнима по модулю d с -2, а в противоположной доле — кратна d. Значит, граф $G \setminus e$ не двудольный, а тогда найдётся нечётный цикл, не содержащий e. Осталось воспользоваться утверждением, приведённым в начале решения.

Решение 2. Это решение предложил участник конференции Алексей Цыбышев.

Расставим на всех рёбрах исходного однородного графа числа $1/d$, сумма в каждой вершине будет равна 1. Начнём проводить процесс, описанный в решении задачи 2.3, — избавляться от чётных циклов и гантелей, меняя соответствующим образом веса рёбер и откидывая нулевые рёбра. В итоге останется 1-2-скелет с полумагической расстановкой. Очевидно, что на его рёбрах стоят числа 1 и $1/2$. Но поскольку $1/d$ не равно ни 0, ни 1, ни $1/2$, любое ребро хоть раз изменило свой вес. Это значит, что любое ребро содержится в каком-нибудь псевдоцикле.

4.2. Будем называть расстановку ± 1 на чётных циклах и ± 1, ± 2 на гантелях, описанную в тексте условий, стандартной расстановкой на псевдоцикле.

Лемма. Пусть в графе задана расстановка чисел на рёбрах, причём веса всех рёбер ненулевые, а сумма в каждой вершине равна нулю. Тогда любое ребро содержится в чётном цикле или гантеле.

Доказательство леммы почти дословно повторяет решение задачи 4.1. Вместо количества рёбер нужно говорить о сумме их весов и пользоваться тем, что в двудольном графе сумма весов рёбер, выходящих из обеих долей, одинаковы.

1. Предположим, что однородный граф G — магический с суммой s в каждой вершине. Вычтем из веса каждого ребра число s/d, получится расстановка на рёбрах различных чисел с нулевой суммой в каждой вершине.

Выберем в графе G произвольное ребро ненулевого веса и согласно лемме найдём содержащий его псевдоцикл. Вычтем из весов рёбер этого псевдоцикла его стандартную расстановку, умноженную на такой коэффициент, чтобы вес данного ребра обнулился. Теперь выкинем из G все «нулевые» рёбра. Полученная разметка рёбер уменьшенного графа по прежнему обладает нулевой суммой в каждой вершине. Снова выберем в нём ребро и снова применим лемму, и т. д.

Количество рёбер в графе на каждом шагу уменьшается и рано или поздно все рёбра станут «нулевыми». Это будет означать, что исходная разметка рёбер графа G является «суммой» стандартных расстановок на псевдоциклах с подходящими коэффициентами. Поскольку любые два ребра G имеют разный вес, то для них найдётся псевдоцикл, вносящий в эти рёбра разный вклад. Это и означает, что он слабо разделяет эти два ребра.

2. Предположим теперь, что любые два ребра слабо разделяются псевдоциклами. Выпишем все псевдоциклы, и пронумеруем их числами от 1 до N (где N — их количество). Для k-го псевдоцикла назначим веса его рёбер, умножив его стандартную расстановку на 5^k. Теперь для каждого ребра сложим все назначенные ему веса и прибавим большую константу C, чтобы все веса стали положительными. Будем считать этот результат окончательным весом данного ребра. Полученная разметка рёбер графа — магическая: каждый псевдоцикл даёт нулевой суммарный вклад весов в каждую вершину; добавление C к каждому ребру изменяет сумму в вершине на dC. При этом веса всех рёбер различны. Действительно, так как целое число однозначно представляется в виде комбинации степеней пятёрки с коэффициентами $-2, -1, 0, 1, 2$, а у любых двух рёбер хотя бы в одном псевдоцикле коэффициенты при соответствующей степени пятёрки различны.

4.3. Эта теорема является непосредственным следствием критерия магичности однородных графов, изложенного в предыдущей задаче.

Докажем, что любые два ребра G слабо разделяются псевдоциклами. Если они лежат в одной компоненте, то это следует её магичности. Если же в разных, то в одной из них можно выбрать псевдоцикл, содержащий соответствующее ребро (задача 4.1.), он и будет разделять эти два ребра.

4.4. 1. Проверим сначала, что $\ell(G) \neq 1$. Если $\ell(G) = 1$, то при удалении одного ребра e граф распадается на две компоненты связности. Рассмотрим любую из них. Как и сам граф G, она является двудольным графом, причём степень одной её вершины равна $d-1$, а остальных — ровно d. Как уже обсуждалось (см. решение задачи 4.1.), этого не может быть.

2. Докажем, что если $\ell(G) \geqslant 3$, то G — магический. Возьмём любые два ребра e и f и докажем, что они слабо разделяются псевдоциклами. При выкидывании этих рёбер граф остаётся связным, поэтому найдётся цикл, содержащий e, но не содержащий f. В силу двудольности этот цикл чётен, и он разделяет e и f.

3. Докажем, что если $\ell(G) = 2$, то G — не магический. Пусть при выкидывании рёбер e и f граф теряет связность; тогда образуется ровно две компоненты связности, обозначим их V и W. Каждая из них является двудольным графом. Если рёбра e и f имеют общий конец, то в одной из компонент окажется вершина степени $d-2$, в то время как остальные её вершины имеют степень d — такой граф не может быть двудольным. Следовательно, $e = AB$ и $f = CD$ не имеют общих рёбер, и в одной компоненте содержатся вершины A, C, а в другой — вершины C и D, степени которых равны $d-1$. Значит, A и C (а также B и D) попадают в разные доли и все пути между ними нечётны.

Докажем, что рёбра e и f не могут слабо разделяться псевдоциклом. Нечётных циклов (а значит, и гантелей) в графе G вообще нет в силу его двудольности. Если же чётный цикл содержит, например, ребро e, то он содержит и ребро f, причём из выводов предыдущего абзаца следует, что между ними в цикле с каждой стороны расположено нечётное число ребер. Значит, этот цикл не может слабо разделять рёбра e и f.

References

[1] *Ловас Л., Пламмер М.* Прикладные задачи теории графов. М.: Мир, 1998.

[2] *Doob M.* Characterizations of regular magic graphs // J. Combin. Theory, ser. B. Vol. 25. 1978. P. 94–104.

[3] *Katerinis P.* Minimum degree, factors and magic graphs.

[4] *Trenkler M.* Number of vertices and edges of magic graphs // Ars Combinatoria. 2000. Vol. 55. P. 93–96.

[5] *Trenkler M.* Some results on magic graphs // Proceedings of the third Czechoslovak symposium on graph theory. Teubner-texte zur Mathematik. Bd. 59. Leipzig: Taubner Verlagsgellschaft, 1983. P. 328–332. arXiv:0906.1317v1.

[6] *Semaničová A.* Magic graphs having saturated vertex // Tatra Mt. Math. Publ. 2007. Vol. 36. P. 121-128.

Magic graphs

K. Kokhas, D. Rostovskiy

Definitions and notations

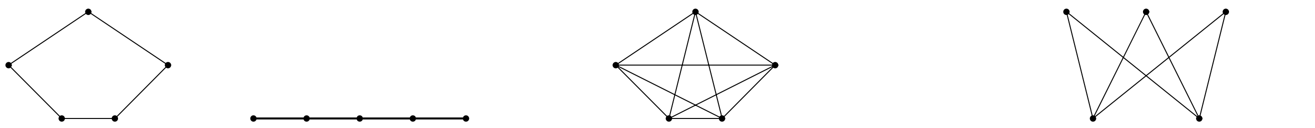
All the graphs under consideration are supposed to be without isolated vertices, multiple edges and loops.

The words “cycle” and “path” mean *simple* cycle and *simple* path in a graph.

For every edge of a graph we assign a positive number that we call a weight of this edge. A graph is called *semimagic* if it is possible to choose weights of its edges and a positive number s such that for each vertex the sum of weights of its edges equals to s. A graph is called *magic* if it possible to choose these weights to be pairwise different. Observe that a vertex of degree 1 in the semimagic graph is necessarily the endpoint of the isolated edge. A magic graph can contain at most 1 isolated edge.

A subgraph F of a given graph G is called a *skeleton*, if it contains all the vertices of G and none of them is isolated vertex in F. *1-2-skeleton* is a skeleton such that all its vertices have degree 1 or 2 and for each component the degrees of its vertices are the same. In other words 1-2-skeleton consists of isolated edges and simple cycles only. For each 1-2-skeleton we can split all the edges of the graph onto 3 groups: edges that belong to the *cyclic* part of F (we will denote it by F_c); edges that belong to the *linear* part of F, i. e. isolated edges in F (we will denote it by F_ℓ); and edges that do not belong to F. We say that 1-2-skeleton *separates* edges e_1 and e_2 if these two edges belong to different groups. In other words at least one of them belongs to F but at most one belongs to F_c and at most one belongs to F_ℓ.

We will use the following notations: C_n is the cycle with n edges ($n \geqslant 3$); P_n is the path with n edges; K_n is the complete graph with n vertices; $K_{m,n}$ is the complete bipartite graphs with parts of m and n vertices.



A cycle C_5 A path P_5 A complete graph K_5 A complete bipartite graph $K_{2,3}$

Figure 1: Some standard graphs

A direct product $F \times G$ of two graphs is the following graph. Its vertex set is the set of all pairs (v, w), where v is a vertex of F, w is a vertex of G. The vertices (v_1, w_1) and (v_2, w_2) are joined by an edge, if either $v_1 = v_2$ and G contains the edge w_1w_2, or $w_1 = w_2$ and F contains the edge v_1v_2. The graph $G \times P_1$ is called *the double* of graph G. *Dum-bell* is a graph consisting of either two odd cycles which share exactly one common vertex, or two odd cycles joined by a path of an arbitrary length.

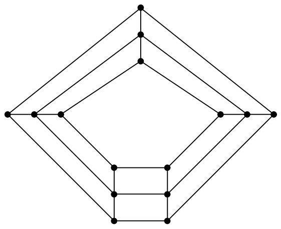


Figure 2: Graph $C_5 \times P_2$

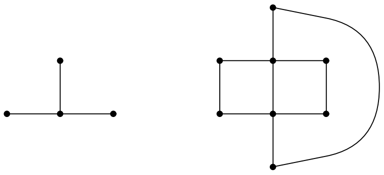


Figure 3: Graph and its double

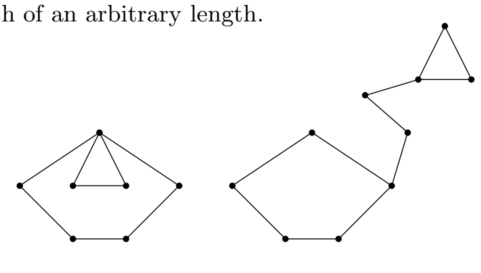


Figure 4: Dum-bells

1 *Examples*

1.1. Show that magic graphs with less than 5 vertices do not exist, except the graph P_1 (one edge).

1.2. Prove that a bipartite graph with odd number of vertices is non magic. Could it be semimagic?

1.3. Determine wether these graphs are semimigic or magic (the answers may depend on n and m)
a) K_n; b) $K_{m,n}$; c) $P_n \times P_1$; d) $P_n \times P_m$ при n, $m > 1$; e) $C_n \times P_1$; f) $C_n \times P_m$, $n \geqslant 3$, $m > 1$; g) cycle of $2n$ vertices, where every two opposite vertices are joined by edge.

2 *Semimagic graphs*

2.1. Prove that if a semimagic graph G contains an even cycle then G contains also a semimagic skeleton (i. e. the skeleton which is a semimagic graph itself) such that not all the edges of the cycle belong to this skeleton.

2.2. Prove that if a semimagic graph G contains a dum-bell then G contains also a semimagic skeleton such that not all the edges of the dum-bell belong to this skeleton.

2.3. Prove that each semimagic graph has 1-2-skeleton.

2.4. The main theorem about semimagic graphs. Prove that a graph is semimagic if and only if each of its edges belongs to some 1-2-sceleton.

In the following problems we find out when a graph contains 1-2-skeleton. We call a graph *soft* if it does not have 1-skeleton, and *solid* if it contains 1-skeleton. A soft graph is called *saturated* if it turns solid when an arbitrary edge has been added.

Let G be an arbitrary graph, S is an arbitrary set of its vertices. Denote by $G \setminus S$ the graph obtained by deletion of all the vertices of the set S and its edges.

2.5. Let G be a saturated soft graph, S be the set of all its vertices such that each of them is joined with all other vertices. Prove that all components of the graph $G \setminus S$ are complete graphs.

2.6. The main theorem about saturated soft graphs. A graph G is saturated and soft if and only if either

a) G is a complete graph with odd number of vertices, or

b) the number of vertices of G is even and we can split it onto complete graphs S_0, G_1, G_2, ..., G_k, where $k = |S_0| + 2$, such that for all i the number of vertices in G_i is odd and every vertex of G_i is joined with all the vertices of S.

2.7. Prove that a graph G is solid if and only if for each set S of vertices of G the graph $G \setminus S$ has at most $|S|$ odd components.

2.8. Prove that graph G contains 1-2-skeleton if and only if for each set S of vertices of G the graph $G \setminus S$ has at most $|S|$ isolated vertices.

3 Magic graphs

3.1. Prove that each magic graph has the following two properties:

(1) Every edge of the graph belongs to some 1-2-skeleton.

(2) Every two edges are separated by some 1-2-skeleton.

3.2. Prove the converse statement: if a graph has these two properties then it is magic.

3.3. Graph G' is obtained from magic graph G by adding a new edge and this new edge belongs to some 1-2-skeleton of graph G'. Prove that G' is magic.

3.4. Graph G consists of two (non isomorphic) components, each component has at least 3 vertices. Both components are magic graphs. Is it true that G is necessarily magic?

3.5. a) For each edge e in a semimagic graph G (without isolated edges) there exists a 1-2-skeleton, whose cyclic part does not contain e. Prove that the double of G is magic.

b) G is a semimagic graph without isolated edges, H is an arbitrary connected graph without isolated edges. Prove that $G \times H$ is a magic graph.

3.6. G is an arbitrary graph with at least 4 vertices. Graph G' is obtained by adding one more vertex to G, and this vertex is joined with all the "old" vertices of G. Prove that the graph G' is magic if and only if the graph G is without isolated edges and it has 1-2-skeleton.

3.7. a) Graph G has $n \geqslant 5$ vertices. The degrees of vertices of G are at least $\frac{n}{2} + 1$. Prove that g is a magic graph.

b) Prove that for any large n there exist non semimagic graph such that the minimal degree of its vertices equals to $[n/2]$.

3.8. G is a connected magic graph with $n \geqslant 5$ vertices and r edges. Prove that $r > \frac{5}{4}n$.

3.9. For $n = 5$, 6, 7, 8 construct a connected magic graph with n vertices and r edges, where r is the minimal integer that satisfies the inequality $r > \frac{5}{4}n$.

3.10. Construct an analogous graph for each $n \geqslant 5$.

3.11. Prove that there exists a connected magic graph with n vertices and r edges, if the pair (n, r) satisfies the inequality $\frac{5}{4}n < r \leqslant \frac{n(n+1)}{2}$.

Semifinal

4 Regular graphs

We will not discuss when regular graphs of degree 1 and 2 are magic. Below we will consider regular graphs of degree at least 3.

A pseudocycle is an even cycle or dum-bell (remind that both cycles in dum-bell are even).

Consider an even cycle. Put alternatively on its edges weights 1 and -1, let all other edges have weight 0. We say that two edges are *weakly separated* by the cycle if they have had different weights. Analogously, for each dum-bell, put the weights ± 1 and ± 2 on its edges as in fig. 5 (where $a = 1$), and let all other edges have weight 0. We say that two edges are *weakly separated* by the dum-bell if they have had different weights. Finally, we say that two edges are *weakly separated* by a pseudocycle if there exists an even cycle or a dum-bell that weakly separates these edges.

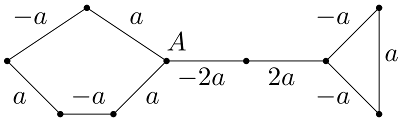


Figure 5: Alternative weights of dum-bell edges

4.1. Prove that every edge of the regular graph of degree $d \geqslant 3$ belongs to some pseudocycle.

4.2. Prove that the regular graph of degree $d \geqslant 3$ is magic if and only if any two of its edges are separated by pseudocycle.

4.3. Prove the following theorem. Let G be a regular graph of degree $d \geqslant 3$ and $G_1, \ldots, G_k$ be its components. Then G is magic if and only if all G_i are magic.

Index of edge connectivity $\ell(G)$ is the minimal number of edges of G that should be erased in order to obtain disconnected graph.

4.4. Let G be connected regular bipartite graph. Prove that the property "to be magic" or "to be non-magic" depends on $\ell(G)$ only and completely investigate this dependance.

5 Addendum

5.1. To the problem 1.3.a. Graph is called *supermagic* if its magic weights are consecutive positive integers.

For which n graph K_n is supermagic?

5.2. To the problem 3.7. A graph has 2009 vertices of degree at least 1006. At most 500 edges were deleted. Prove that the rest graph is still magic.

Solutions

1 *Examples*

1.1. If a graph with 4 vertices has 1 or 2 edges then it has isolated vertex. If it has 3 or 4 edges then it contains two adjacent vertices of degree 2 and hence it is non-magic. The graph with 6 edges is necessarily K_4, see problem 1.3a).

Finally, if it has 5 edges then it is isomorphic to the cycle $ABCD$ with the diagonal AC. Then the sum of weights of edges adjacent to vertices A and C equals $2s$. Geometrically, it is the sum of weights of all edges, where the weight of AC has multiplicity 2. The other way to obtain the sum $2s$ is to sum up the weights of edges adjacent to vertices B and D. This is a sum of all edges of the graph except AC. Therefore AC has zero weight, which is forbidden.

1.2. Answer: the graph is not semimagic. Let one part of the graph contains k vertices, the second part contains ℓ vertices, and let s be the sum of weights of all edges adjacent to the same vertex. If the graph is semimagic, then the sum of weights of edges adjacent to vertices of the first part equals ks, the sum of weights of edges adjacent to vertices of the second part equals ℓs, and both sums equals the sum of weights of all edges of the graph. Therefore $\ell = k$. This is impossible because the total number of vertices is odd.

1.3. a) Answer: the graph is always semimagic, it is magic for $n = 2$ and $n > 5$ only.

To show that the graph is semimagic take all weights equal to 1.

If $n = 3$ the graph is not magic, it is evident.

If $n = 4$ we have 4 vertices A, B, C, D. Assume that it is magic, let s be the sum of weights of all edges adjacent to the same vertex. Then $2s$ is the sum of all edges adjacent to vertices A and C, i.e. the sum of all edges of the graph except CD but with weight of AC counted twice. By the analogous consideration for vertices B and D we will obtain that the weights of AC and BD coincide.

If $n > 5$ the graph is magic. We will prove this by the following construction. Since the graph is regular, we may consider arbitrary weights (not necessarily positive), because in the regular graph we can make all the weights to be positive by adding a large positive constant. Let us describe the *main construction* of magic weights for the regular graphs by means of even cycles.

> Write out all even cycles that are contained in our graph and enumerate them by numbers from 1 to N (where N is the total number of these cycles). For any k put the weights $\pm 3^k$ alternatively on the edges of k-th cycle. After that for each edge sum up all the weights on it.

Let us check that this set of weights is magic. Indeed, the sum of weights of edges adjacent to every vertex equals 0, because the contribution of each cycle to this sum is 0. Now let us check that all weights are distinct. For each edge of the graph write out the list cycles which contain this edge. It is clear that for any two edges there exists an even cycle that contains one of these edges only. Therefore for any two edges their lists of cycles do not coincide. But then the sums of weights determined by the cycles are not equal. This is because the weights obtained by our construction may be regarded as N-digital ternary numbers in system (with base 3) with digits 0 and ± 1. Since all lists are distinct then all these ternary numbers are distinct also.

b) Answer: the graph is semimagic for $m = n$ only. For $m = n > 2$ it is magic.

The equality $m = n$ is necessary for the graph to be semimagic (see problem 1.2).

If $m = n$ the graph is regular and hence semimagic. If $m = n = 2$ it is evidently non-magic (see problem 1.1). And if $m = n > 2$, the graph is magic, due to construction from the previous solution.

c) Answer: the graph is semimagic but non-magic.

To show that it is semimagic it is sufficient to assign weights of all edges to be 1, except leftmost and rightmost edges of weight 2. The graph is non-magic because it contains adjacent vertices of degree 2.

d) Answer: the graph is magic if either m or n is odd. If both m and n are even, then the graph is not semimagic.

If both m and n are even, then the graph is bipartite with $(n + 1)(m + 1)$ vertices (odd number). This graph is not semimagic due to problem 1.2.

Now prove that the graph is magic for odd n and $m > 1$.

At first consider case $m = 2$. Consider the initial placement of nonnegative semimagic weights depicted on fig. 6: bold edges have weight $2m$, dashed edges have weight 0, all other edges have weight M, where M is a big number, that we will choose later. This set of weights is almost semimagic, but some weights here are zero and the graph is not regular.

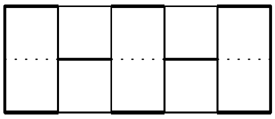


Figure 6: Almost semimagic weights on graph $P_n \times P_2$

Now we will perform the main construction of magic weights for the regular graphs by means of even cycles, but with 3 corrections (because our graph is not regular):

1) In the main construction we will consider 4-cycles only (i.e. the sides of cells).

2) The final weight of an edge will be equal to the sum of its initial weight and the weight obtained by the main construction.

3) When we assign positive and negative weights of edges in cycles, we choose plus sign for edges whose initial weight has been equal to zero.

Finally, choose M so big that all the final weights turn out to be positive and distinct. Then this set of weights will be magic.

Now consider a case n is odd, $m > 2$. We perform analogous actions. The initial weights are depicted on fig. 7 (odd number n corresponds to the vertical side of the picture).

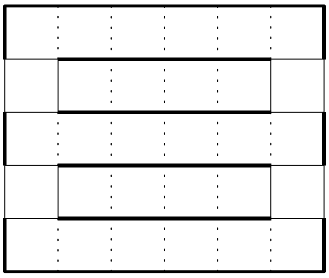


Figure 7: Almost semimagic weights on graph $P_n \times P_m$

e) Answer: the graph is always semimagic. It is magic for even n only.

We may realize this graph as edges of n-gonal prism.

Let $n = 2k$. The graph contains evident cycles of length 4 (sides of facets) and two $2k$-cycles (sides of bases). The property "for any two edges there is a cycle that contains one of them only" is satisfied. Therefore we can perform the main construction for regular graphs.

Let us prove that the graph is non-magic for $n = 2k + 1$. As usual let s be the sum of weights of edges adjacent to the same vertex. It is easy to see that the sum of weights of all edges equals nd. Denote our prism by $A_1A_2 \dots A_{2k+1}B_1 \dots B_{2k+1}$. The sum of weights of edges adjacent to vertices A_1, A_3, …, A_{2k+1}, B_2, B_4, …, B_{2k}, equals nd and can be interpreted as the sum of weights of all edges of the graph except B_1B_{2n+1} and with edge A_1A_{2n+1} counted twice. Hence, the weights of A_1A_{2n+1} and B_1B_{2n+1} are equal.

f) Answer: the graph is magic.

Solution is analogous to solution 1.3d). Let A_1, …, A_{m+1} be vertices of graph P_m. Initial placement of nonnegative weights on the graph looks as follows: all edges of subgraphs of the form $C_n \times A_i$ have weight 2, all other weights are 0. We apply the main construction for 4-cycles only.

g) Answer: the graph is always semimagic; it is magic for odd n.

It is semimagic because it is reguar. For $n = 2$ this graph is K_4, we discuss it in problem 1.3 a). For odd n it is magic because the main construction works (we have a good store of even cycles here: 4-cycles that contains subsequent diameters and $(n+1)$-cycles of the form "semicircle").

For even n the graph is non-magic. Let $A_1A_2 \dots A_nB_n \dots B_2B_1$ be vertices of the given cycle. The set of edges adjacent to all vertices A_i, B_i, where i runs over odd numbers, is the set of all edges of the graph except A_nB_n and with A_1B_1 counted twice. It follows that in any semimagic set of weights the opposite edges of the cycle have the same weight.

2 *Semimagic graph*

2.1. Let a be the minimal weight of the edges in the given cycle. We will move along the cycle and decrease and increase by a alternatively the weights of the edges of the cycle. After that erase all the edges with zero weight. The remaining graph together with the weights of its edges will be desired semimagic skeleton.

2.2. Let A be the vertex of degree 3 (or 4) of one of odd cycles of the dum-bell. We will bypass the dum-bell starting from the vertex A. First of all we will move along the odd cycle and assign the weights to its edges to be $\pm a$ alternatively. After return to the vertex A, we have two edges of weight a adjacent to A. Then we move along the handle of dum-bell and assign the weights of its edges to be $\mp 2a$ alternatively. After that we move along the second cycle assigning its edges weights $\pm a$ alternatively. We obtain semimagic weights with $s = 0$ (see fig. 5).

Now we are going to add new weights to old ones. For this choose the value of parameter a so that all the weights of the graph would be nonnegative and the weight of at least one of the edges would be equal to 0. We will obtain semimagic weights. After erasing all the edges with zero weight we obtain the desired skeleton.

2.3. The constructions of solutions 2.1, 2.2 allow us to decrease consequently the number of edges in the graph by destroying its even cycles and dum-bells. The graph will be semimagic during all these operations and therefore there will be no pendant vertices in it (except the endpoints of isolated edges). Observe that if a component of the graph contains two even cycles then it contains also an odd cycle or a dum-bell. And if a component contains exactly one (odd) cycle and does not contain pendant vertices then this component is exactly this odd cycle.

So, after destroying all even cycles and dum-bells we will obtain a graph consisting of several isolated edges and several (odd) cycles.

2.4. Let us check that every edge of a semimagic graph belongs to some 1-2-skeleton. Let G be a graph with minimal number of edges such that one of its edges, say, e does not belong to any 1-2-skeleton. Fix a set of semimagic weights $\mathcal{W}$ on graph G. Fix an arbitrary 1-2-skeleton and construct one more semimagic set of weights $\mathcal{S}$ as follows. Let each edge of the linear part of the skeleton has weight a, each edge of the cyclic part has weight $a/2$ and all other edges (including e) have weight 0. Choose the value of a so that all the weights of the set $\mathcal{S}$ do not exceed the corresponding weights in the set $\mathcal{W}$ and for at least one edge we have an equality. Now subtract from weights of $\mathcal{W}$ the weights of $\mathcal{S}$. We obtain a semimagic set of weights, where not all the weights are equal to 0, because the weight of edge e has not changed. Now remove all edges of zero weight. We obtain semimagic graph G' which is a skeleton of G. G' contains less edges than G and edge e does not belong to any 1-2skeleton of G', because "the skeleton of my skeleton is my skeleton". This is a contradiction with the definition of G. Therefore the graph G does not exist.

Now let us check that if every edge of the graph belongs to some 1-2-skeleton, then the graph is semimagic. For each 1-2-skeleton assign the weight of its cyclic edges be 1 and the weight of its isolated edges be 2. We obtain semimagic set of weights (but with zero weights). Let us sum up all these set of weights over all 1-2-skeletons. The result is the desired semimagic set of weights.

2.5. We take this problem and the following solution from [1, § 3.1.2]. Let A, B, C be vertices of $G \setminus S$ and B is joined with both A and C. It is sufficient to prove that graph G contains edge AC. Assume that this is not true. By the definition of the set S graph G contains the vertex D such that the edge BD does not belong to the graph. If we add edge AC to graph G, then the new graph has 1-skeleton (since graph G is saturated). It is clear that edge AC must belong to this skeleton. Color this skeleton in red. Analogously, if we add edge BD to graph G, we can find blue 1-skeleton containing edge BD. Now from these two skeletons we will construct a 1-skeleton of graph G and get a contradiction.

Consider the union of these skeletons; the edges which are red and blue simultaneously we will consider as usual (non-multiple) edges. Then this union is a 1-2-skeleton of graph $G \cup AC \cup BD$, and all its cycles are even because red and blue edges alternate.

It is clear that edges AC and BD are both in the cyclic part. If these edges belong to different cycles, then the desired 1-skeleton can be constructed as follows. Take the red skeleton and replace all red edges of the cycle that contains AC by blue edges of the same cycle. Now consider the second case, let edges AC and BD belong to cycle γ. Let us bypass cycle γ starting from vertex B and edge BD till we reach vertex A or C. Let it be A for definiteness. Since the red edge of vertex A is AC, we finish our movement by blue edge. Hence the path from B to A starts and finishes with blue edges. Take the blue skeleton, replace all blue edges of the path by red edges of the same path and add edge AB. We obtain 1-skelton of graph G.

2.6. We take this problem and the following solution from [1, § 3.1.2]. It is evident that a soft saturated graph with odd number of vertices is necessarily complete. Let the number of vertices in G be even; let S be the set of all vertices of G that are joined with all other vertices and s be the number of these vertices; let G_1, G_2, …, G_k be components of connectivity of graph $G \setminus S$. Due to the statement of the previous problem we know that they are all complete graphs.

If $G \setminus S$ has at most s odd components, construction of the 1-skeleton is trivial. Assume that $G \setminus S$ has at least $s+1$ odd components; taking into account parity of number of vertices of G, we conclude that $G \setminus S$ has at least $s+2$ odd components. If the number of odd components is greater than $s+2$, join any two of them by an edge. We obtain graph G_1 such that graph $G_1 \setminus S$ has more than s odd components of connectivity. There are no 1-skeletons in this graph (it is evident, it follows also from the easy part of the statement of problem 2.7), but this is impossible because graph G is saturated.

Thus, graph G has exactly $s+2$ odd components. It can not have even components due to analogous reasons.

2.7. This statement is classical Tutte theorem. The following proof is from [1, § 3.1.2].

If we can find the set of vertices S in graph G, such that graph $G \setminus S$ has more than $|S|$ odd components of connectivity, then graph G is soft. It is clear.

Check the converse statement. Assume that for any subset S of the set of vertices of G graph $G \setminus S$ has at most $|S|$ components of connectvity but at the same time graph G is soft.

The number of vertices of graph G is even because otherwise $S = \varnothing$ leads to the contradiction. Add several edges to graph G to obtain soft saturated graph G'. Let S' be set of vertices joined with every vertex of G', s be number of its elements. Since G' and G have equal (even) number of vertices due to main theorem about soft saturated graphs we have that graph $G' \setminus S'$ contains $s+2$ odd components, each of them is a complete graph. Now remove those edges we have add making graph saturated. It is possible that some components of graph $G' \setminus S'$ will fall to parts but at least one part of odd component will be odd and the total number of odd components will be grater than s. Thus, the set S' disproves the property of G under consideration.

2.8. Let n be the number of vertices of graph G. Construct a new graph G' with $2n$ vertices. For every vertex v in G take two vertices v' and v'' in G'; for every edge uv define two edges in G': $u'v''$ and $u''v'$ Then G' is a bipartite graph, that has twice as many edges as G.

Remark that the existence of 1-2-skeleton in G is equivalent to the existence of perfect matching in G'. Indeed, for each cycle $v_1v_2\dots v_\ell$ of the skeleton graph G' has edges $v_1'v_2''$, $v_2'v_3''$, …, $v_\ell'v_1''$; analogously for any isolated edge graph G' contains edges $u'v''$ and $v'u''$. It is clear that all these edges form a perfect matching. Conversely, for any perfect matching of graph G' it is not difficult to construct a 1-2-skeleton. For example, the edges $u'v''$, $v'w''$, $w'z''$, $z'u''$ of the perfect matching determine a cycle $uvwz$ of the graph G, and edges $u'v''$ and $v'u''$ determine an isolated edge uv of the skeleton.

As we know, for each set S of vertices of G the graph $G \setminus S$ has at most $|S|$ isolated vertices. Let us reformulate this property in terms of graph G'. Consider an arbitrary set S of vertices of graph G. What does it mean that after deletion of this set the vertex u becomes isolated? This means that in the graph G' all neighbours of u' belong to S''. If after removing set S we have $k > |S|$ isolated vertices then the conditions of Hall theorem is not satisfied in graph G because we found a set of k vertices that has at most $|S|$ neighbours (the last number is less then k). The converse is also true (i.e. if the conditions of Hall theorem is not satisfied, then the property under discussion holds). Therefore this property is equivalent to the conditions of Hall theorem in graph G' that is equivalent to existence of perfect matching in G'.

3 Magic graphs

3.1. (1) All the semimafic graphs have this property.

(2) We will prove more general fact: if a semimagic graph has a semimagic set of weights such that two edges, say e_1 and e_2, have distinct weights, then the edges e_1 and e_2 are separated by a 1-2-skeleton.

It can be done analogously to the solution of problem 2.4. Choose a minimal graph; fix a set of weights, where not all weights are equal; subtract by a suitable way the weights belonging to a skeleton; we will obtain smaller graph. Since the initial graph was minimal, one of edges e_1, e_2 must receive zero weight and should be removed. In the remaining graph the second edge due to statement of problem 2.4 belongs to some 1-2-skeleton, that sepapates these edges in the initial graph.

3.2. Let us enumerate all the 1-2-skeletons. Let the edges of cyclic part of k-th skeleton have weights 3^k, and edges of linear parts have weight $2 \cdot 3^k$. For each edge sum up all its weights over all 1-2-skeletons. The set of weights obtained is semimagic due to uniqueness of ternary notation of numbers.

3.3. It follows from 3.2.

3.4. Answer: no, graph G is not necessarily magic. Two magic graphs are depicted on the fig. 8 For any set of magic weights the edges denoted by dashed lines must have weight $s/2$.

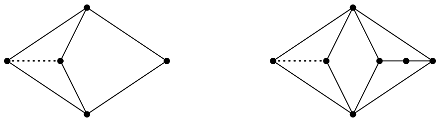


Figure 8: Union of magic graphs can be non-magic

3.5. a) The double G^2 consists of two copies G_1 and G_2 of the graph G and of the set of edges E between the corresponding vertices. The corresponding edges in parts G_1 and G_2 are called *parallel*. The edges from the set E are called *vertical*. A subgraph of G^2 consisting of two copies of some subgraph of G is called *duplicated*.

First of all describe a construction of *rotation of parallel edges* in G^2. Let subgraph H of graph G^2 be the union of two subgraphs in parts G_1 and G_2 (without vertical edges) such that these subgraphs contain parallel edges A_1B_1 and A_2B_2. Let us replace edges A_1B_1, A_2B_2 in subgraph H by edges A_1A_2 and B_1B_2. Denote the new subgraph by H'. We say that subgraph H' is obtained from H by the rotation of parallel edges. It is clear that both of H and H' are (or are not) 1-2-skeletons.

To prove that graph G^2 is magic let us apply criterion from problems 3.1–3.2.

(1) Every edge belongs to 1-2-skeleton. It is clear for edges from G_1 (and from G_2): duplicate the 1-2-skeleton containing this edge in G_1. For vertical edges choose suitable rotation of edges of appropriate duplicated 1-2-skeleton.

(2) Every two edges are separated by 1-2-skeleton.

- If both of edges e_1 and e_2 belong to G_1 consider a duplicated 1-2-skeleton containing e_1. If it does not separate e_1 and e_2, then both edges belong to the skeleton. By rotating edge e_2 and its parallel copy we obtain a separating skeleton.
- If e_1 belongs to G_1, and e_2 belongs to G_2, consider the union of 1-2-skeleton in G_1 containing e_1 (it exists due to the statement of problem 2.4), and 1-2-skeleton in G_2 that does not contain e_2 (it exists by the condition of the problem).
- If e_1 belongs to G_1 and e_2 is vertical consider a duplicated skeleton containing e_1.
- Finally, if both of edges A_1A_2 and B_1B_2 are vertical choose in G_1 an edge A_1X_1 (where $X_1 \neq B_1$) or B_1Y_1 (where $Y_1 \neq A_1$), this edge exists since G has no isolated edges. Consider a duplicated skeleton containing this edge and rotate this edge together with parallel edge.

b) Analogously to a).

3.6. We take the statement of the problem from [5]. The following solution was found by participants of the conference.

1. Check that if graph G' is magic then graph G has 1-2-skeleton and has no isolated vertices and edges. Isolated vertex in G corresponds to a pendant vertex in G'. Isolated edge in G corresponds to two adjacent vertices of degree 2. Both constructions are impossible in a magic graph.

Assume that G does not contain 1-2-skeleton.

Denote by S the new vertex of graph G'. Fix an arbitrary 1-2-skeleton in graph G', w.l.o.g. we may assume that all its cycles are odd. Consider the component of the skeleton that contains vertex S. If this component is an odd cycle, remove vertex S and split other vertices of the cycle on pairs. Together with other components of the skeleton they form a skeleton of graph G. Therefore we may assume that this component is an isolated edge SA_1. Now we will construct two sets of vertices $\mathcal{A} = \{A_1, \ldots, A_n\}$ and $\mathcal{B} = \{B_1, B_2, \ldots, B_n\}$ such that the edges A_iB_i $(1 \leqslant i \leqslant n)$ belong to skeleton K and all the vertices adjacent to vertices of the set $\mathcal{A}$ belong to $\mathcal{B}$.

Let $\mathcal{A} = \{A_1\}$, $B_1 = S$. Assume that the sets $\mathcal{A} = \{A_1, \ldots, A_k\}$ and $\mathcal{B} = \{B_1, \ldots, B_k\}$ have constructed already and there is an edge that joins some vertex of the set $\mathcal{A}$ with some vertex outside $\mathcal{A} \cup \mathcal{B}$, say A_kB_{k+1}. Vertex B_{k+1} belongs to some component of the skeleton K. If this component is an odd cycle we can easily reconstruct the skeleton K to obtain a 1-2-skeleton of graph G.

> To do this consider the shortest path in $\mathcal{A} \cup \mathcal{B}$ from B_1 to B_{k+1} such that the vertices of $\mathcal{A}$ and $\mathcal{B}$ alternate. The path has even length, choose all its even edges (the last of them has endpoint B_{k+1}) and split onto pairs all other vertices of the odd cycle.

Therefore we may assume that vertex B_{k+1} belongs to the isolated edge $B_{k+1}A_{k+1}$ of skeleton K. Then place vertex B_{k+1} to the set $\mathcal{B}$ and vertex $\mathcal{A}_{k+1}$ to the set $\mathcal{A}$.

We will increase sets $\mathcal{A}$ and $\mathcal{B}$ by this algorithm until it is possible. As a result we obtain that the set $\mathcal{A}$ is joined by edges with $\mathcal{A} \cup \mathcal{B}$ only. Assume there exists an edge A_iA_j consider the shortest path between these two vertices such that the vertices of $\mathcal{A}$ and $\mathcal{B}$ alternate in it (the existence of this path can be easily seen from the process of construction of sets $\mathcal{A}$ and $\mathcal{B}$). This path together with edge A_iA_j form an odd cycle. Then the skeleton K can be reconstructed to the skeleton of graph G as described above.

So we have sets $\mathcal{A}$ and $\mathcal{B}$. Since these sets have equal number of elements, the sums of weights of their vertices are equal. But the sum of weights of $\mathcal{A}$ equals the sum of weights of all edges A_iB_j while the sum of weights of $\mathcal{B}$ equals the sum of weights of all edges A_iB_j and edges of the form B_1B_i (remind that $B_1 = S$ is adjacent to all other vertices of graph G). We obtain a contradiction.

2. The proof of the converse statement — if graph G has 1-2-skeleton and does not contain isolated edges, then G' satisfies conditions of the problem 3.1 (and therefore it is magic) — is not difficult technical exercise. The skeletons that we need for edge separating can be constructed by a suitable transformation of skeleton in G.

3.7. a) Prove that an arbitrary two edges e and f can be separated by some 1-2-skeleton. Remove the endpoints of edge e (and all their edges) from the graph G. The remaining part of the graph has $n-2$ vertices of degree at least $\frac{n}{2} - 1 = \frac{n-2}{2}$. Then it is known that there is a Hamiltonian cycle in this graph (the cycle that passes trough all the vertices of the graph). This cycle together with edge f forms 1-2-skeleton that separates edges e and f.

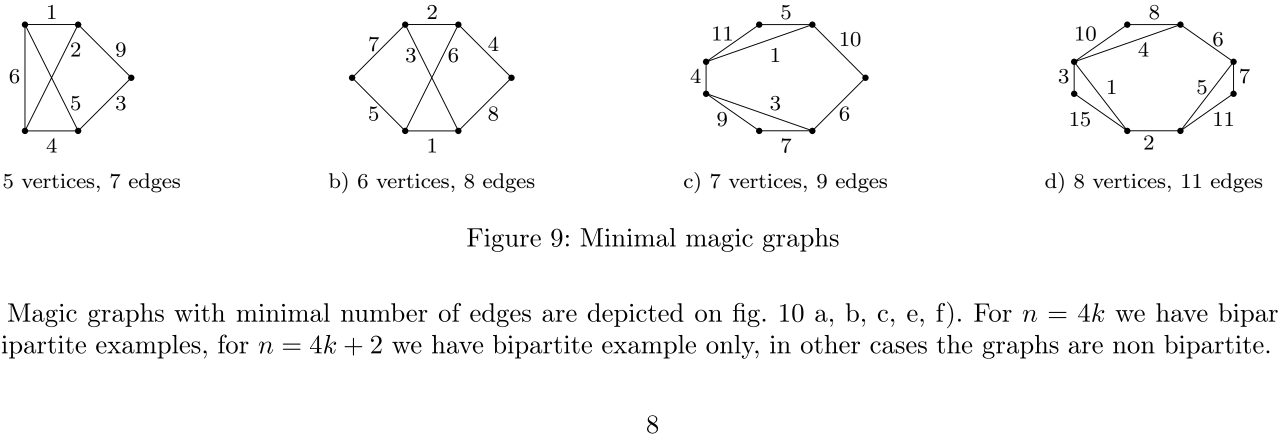
b) Consider the graph G with $n = 2k$ vertices X_1, …, X_k, Y_1, …, Y_k such that its set of edges consists of all edges X_iY_j and edge Y_1Y_2. The degree of each vertex X_i is at least $k = \frac{n}{2}$. Let us prove that G is non semimagic.

Consider an arbitrary 1-2-skeleton of G. In this skeleton each vertex X_i has one or two adjacent vertices among the vertices Y_i. Since we have equal number of vertices of both types, the 1-2-skeleton must be perfect matching. Therefore edge Y_1Y_2 does not belong to any 1-2-skeleton. Hence graph G is non-magic.

3.8. A magic graph has no vertex of degree 1 and any two vertices of degree 2 are not adjacent in it. Let V be the set of vertices of degree 2 (possibly, $V = \varnothing$), W be the set of vertices of degree at least 3. Let s be the sum of weights in each vertex. The sum of weights of edges that have an endpoint in V equals $s|V|$. The second endpoints of these edges belong to W, therefore this sum does not exceed $s|W|$. So, $|V| \leqslant |W|$. The sum of degrees of all the vertices is at least $2|V| + 3|W|$ hence the number of edges is not less than $|V| + \frac{3}{2}|W|$. But $|V| + \frac{3}{2}|W| \geqslant \frac{5}{4}(|V| + |W|) = \frac{5}{4}n$, because $|W| \geqslant |V|$.

The equality would be possible if there are no edges with both endpoints in W, i.e. in a bipartite graph. But in this case $s|V| = s|W|$, so $|V| = |W|$. Then the number of edges between V and W equals $2|V|$ (from the point of view of the set V) and in the same time it is at least $3|W|$. Hence $|V| \geqslant \frac{3}{2}|W|$ that is impossible. The inequality $r > \frac{5}{4}n$ is proven.

3.9. See fig.9.



a) 5 vertices, 7 edges

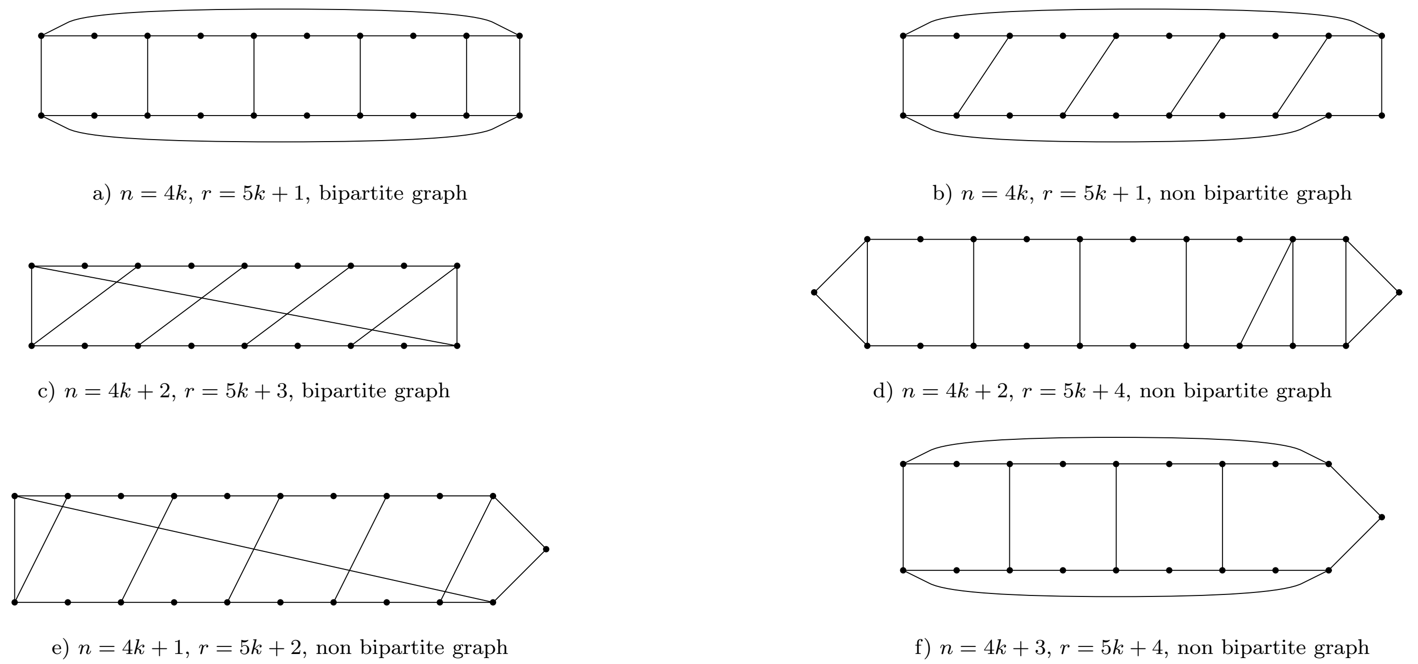
b) 6 vertices, 8 edges

c) 7 vertices, 9 edges

d) 8 vertices, 11 edges

Figure 9: Minimal magic graphs

3.10. Magic graphs with minimal number of edges are depicted on fig. 10 a, b, c, e, f). For $n = 4k$ we have bipartite and non bipartite examples, for $n = 4k + 2$ we have bipartite example only, in other cases the graphs are non bipartite.



a) $n = 4k$, $r = 5k + 1$, bipartite graph

b) $n = 4k$, $r = 5k + 1$, non bipartite graph

c) $n = 4k + 2$, $r = 5k + 3$, bipartite graph

d) $n = 4k + 2$, $r = 5k + 4$, non bipartite graph

e) $n = 4k + 1$, $r = 5k + 2$, non bipartite graph

f) $n = 4k + 3$, $r = 5k + 4$, non bipartite graph

Figure 10: Magic graphs with minimal number of vertices

The proof that the depicted graphs are magic consists of the routine verification that criterion from problems 3.2 holds. Instead of this this verification we show magic sets of weights for "typical case". Of course, this is not the proof but it follows that in concrete cases we really have magic graphs whose edges are separated in the spirit of the criterion. In general case the graphs will be magic too because the separation of its edges takes place "by the same reasons" as in this concrete examples. We restrict ourselves with case $n = 4k + 3$, $r = 5k + 4$, $k = 2$; see fig. 11.

3.11. This solution is a variation of [3]. Observe that if we add an arbitrary edge to non bipartite, connected (and non complete) magic graph then it remains be magic.

Indeed, the new edge e belongs to some cycle. If this cycle is even, assign the weight ε to this edge and the weight $\pm\varepsilon$ alternatively to all other edges of the cycle. Choose the value of ε so that all the weights remain positive and distinct. We obtain a magic set of weights.

If the cycle is odd choose another cycle that does not contain e (it exists because the graph is non bipartite). Then e belongs to some dum-bell (see lemma from solution 4.1) Once again, we can assign the weight ε to edge e and weights $\pm\varepsilon$, $\pm 2\varepsilon$ to other edges of dum-bell and obtain a magic set of weights.

Thus, to complete the solution it is sufficient for each $n \geqslant 5$ to construct "minimal" non-bipartite graph. It was done in the solution of the previous problem for $n \neq 4k + 2$ (see fig. 10 b, e, f). We found these examples in [3]. Unfortunately, the construction of minimal non-bipartite graph for $n = 4k + 2$ in this article is wrong. In addition, the graph on fig. 10 b) is not magic for $n = 8$ (it is impossible to separate by 1-2-skeletons the slanted and the lowest edges). Magic non bipartite graph for $n = 8$ is depicted on fig. 9 d), it was invented by A. Tsybyshev. We do not know wether non bipartite magic graph with $4k + 2$ vertices and $5k + 3$ edges ($k > 3$) exists, so for the case of $5k + 3$ edges we leave bipartite example, and begin our construction from non bipartite graph with $5k + 4$ edges. This non bipartite graph with $4k + 2$ vertices and $5k + 4$ edges is depicted on fig. 10 d). Magic weights on this graph for $k = 3$ see on fig. 12.

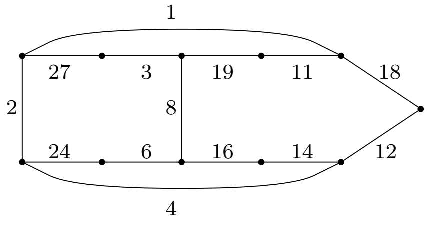


Figure 11: $n = 4k + 3$, $r = 5k + 4$, $k = 2$

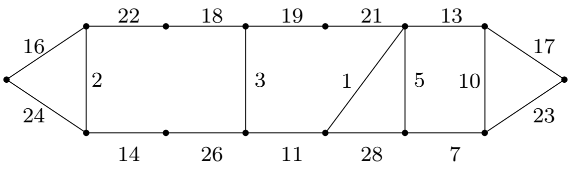


Figure 12: $n = 4k + 2$, $r = 5k + 4$, $k = 3$

4 Regular graphs

4.1. Solution 1.

Lemma. If a connected graph contains two odd cycles, and an edge e belongs to only one of them, than there exists an even cycle or a dum-bell that containпs edge e.

Proof. The statement is trivial if cycles do not intersect or intersect by one vertex.

Assume that intersection of the cycles contains at least two vertices. Let X and Y be the endpoints of edge e. Remove edge e from the first cycle, the remaining part of this cycle we will call the path XY. Let A and B be the first and the last vertex in the path XY that belong to the second cycle. Then the parts XA and BY of the path XY do not intersect the second cycle. Vertices A and B split the second cycle onto two paths whose number of vertices have different parity. Adding one of them to paths XA and BY we can obtain an odd path $XABY$. Together with edge XY it forms an even cycle.

Let us return to the statement of the problem. W.l.o.g. we may assume that the graph is connected. Remove an arbitrary edge $e = AB$ from the graph. Assume that graph $G \setminus e$ falls to two components of connectivity. Since the degrees of all vertices were greater than 1 these components contain more than one vertex. Either of components can not be a bipartite graph. This is because in the bipartite graph the sums of degrees of all vertices in parts are equal, but in our components one of sums is divisible by d and another (which contains vertex A or B) is not divisible by d. Therefore each component contains an odd cycle. Hence edge e belongs to some dum-bell.

Now assume that the graph $G \setminus e$ is connected. Consider an arbitrary path from A to B. If this path has odd number of edges then e is contained in an even cycle. Assume that this path is even (then e is contained in odd cycle). If graph $G \setminus e$ is bipartite then both vertices A and B are in the same part. But this is impossible because the sums of degrees in the parts are not equal: first sum is equivalent $-2 \pmod{d}$ and the second sum is divisible by d. Hence graph $G \setminus e$ is non bipartite and there exists an odd cycle that does not contain e. It remains to use lemma.

Solution 2 (by A. Tsybyshev). Assign weights $1/d$ to all edges of our regular graph. Then sum of weights in each vertex is equal to 1. Now start the process of destroying even cycles and dum-bells by changing weights and removing zero-weight edges that is described in the solution 2.3. As a result of this process we obtain a 1-2-skeleton with magic set of weights. Since we do not change the sum of weights in each vertex the sum of weights in each vertex is still equal to 1. Therefore the weights of edges are 1 and $1/2$. Since $d \neq 1,\ 1/2,\ 0$ we change the weight of every edge at least once. Hence each edge belongs to some pseudocycle.

4.2. The placement of numbers ± 1 on even cycles and ± 1, ± 2 on dum-bells that was described in section 4 (problems) we will call *standard* weights on pseudocycle.

Lemma. Let every edge of the graph has non zero (not necessarily positive) weight and sum of weights in each vertex is equal to 0. Then every edge is contained in even cycle or dum-bell.

The proof is analogous to the solution of the problem 4.1.

1. Assume that the regular graph G is magic and sum of weights in each vertex is equal to s. Subtract the number s/d from every weight. We obtain a placement of pairwise distinct numbers on edges of the graph with zero sum in each vertex.

Choose in the graph G an arbitrary edge of non zero weight and a pseudocycle that contains this edge (it exists due to lemma). Subtract from the weights of this pseudocycle the standard weights of the pseudocysle multiplyed by the appropriate coefficient in order to make the weight of chosen edge to be zero. Then remove all zero-weight edges. We obtain a placement with zero sum in each vertex. Then repeat this operation and so on.

In each step we decrease the number of edges therefore sooner or later all edges become zero-weight It means that the initial placement of numbers is a “sum” with appropriate coefficient of standard placements for pseudocycles. Since any two edges have distinct weights in the initial placement, there exists a pseudocycle such that its standard weights for this edges are distinct. By definition this pseudocycle weakly separates these edges.

2. Assume that every two edges are weakly separated by pseudocycles. Let us enumerate all pseudocycles. For k-th pseudocycle assign weights of its edges to be the standard placement multiplyed by 5^k. For each edge sum up its weights over all pseudocycles and after that add a large positive constant in order to make all weights positive. The obtained set of weights is magic.

4.3. It follows from the previuous problem.

4.4. 1. Let us check first that $\ell(G) \neq 1$. If $\ell(G) = 1$ then we can remove an edge e and obtain the graph with two components of connectivity. Each component itself is a magic graph, one of its vertices has degree $d-1$ and all others have degree d. This is impossible (see solution 4.1).

2. Prove that if $\ell(G) \geqslant 3$, then graph G is magic. Choose any two edges e and f and check that they are weakly separated by pseudocycles. After removing these edges the graph remains connected. Therefore there exists a cycle that contains e and does not contains f. Since the graph is bipartite this cycle is even and it separates edges e and f.

3. Prove that for $\ell(G) = 2$ graph G is non magic. If after removing edges e and f from graph G we obtain a disconnected graph then it has two components, say V and W, each of them is a bipartite graph. If the edges e and f have common endpoint then one of components has unique vertex of degree $d-2$ and other vertices of degree d. Such graph can not be bipartite. Therefore $e = AB$ and $f = CD$ have no common vertices, one component contains vertices A and C, another component contains vertices B and D of degree $d-1$. Hence B and D are in different parts of the component and all the paths that join these points have odd number of edges.

Now prove that edges e and f are not weakly separated by a pseudocycle. Since G is bipartite there are no even cycles (and dum-bells) in it. And all even cycles that contain both e and f do not separate these edges due to previous paragraph.

References

[1] *Ловас Л., Пламмер М.* Прикладные задачи теории графов. М.: Мир, 1998.

[2] *Doob M.* Characterizations of regular magic graphs // J. Combin. Theory, ser. B. Vol. 25. 1978. P. 94–104.

[3] *Trenkler M.* Number of vertices and edges of magic graphs // Ars Combinatoria. 2000. Vol. 55. P. 93–96.

[4] *Trenkler M.* Some results on magic graphs // Proceedings of the third Czechoslovak symposium on graph theory. Teubner-texte zur Mathematik. Bd. 59. Leipzig: Taubner Verlagsgellschaft, 1983. P. 328–332. arXiv:0906.1317v1.

[5] *Semaničová A.* Magic graphs having saturated vertex // Tatra Mt. Math. Publ. 2007. Vol. 36. P. 121-128.

Как считать слова?

Дмитрий Пионтковский, Максим Прасолов, Григорий Рыбников

1 Главная задача

Задача 1. *В словаре племени Винни–Пухов 100 слов. В фразах их языка возможны любые сочетания этих слов. Существуют два магических заклинания, "Земля стоит на великом крокодиле" и "Каждый вечер крокодил глотает солнце", которые вызывают ураган, и поэтому вслух можно произносить только такие фразы, в которых эти последовательности слов не встречаются*[1]*. Сколько всего фраз из двадцати слов можно произносить вслух?*

Задача 2. *У компьютера есть* 256 *различных команд. Существует одна последовательность из четырёх команд, после которой компьютер ломается. Программисты написали все возможные программы из семи команд. Сколько процентов из них не сломают компьютер?*[2]

Задача 3 (Главная Задача)**.** *Алфавит некоторого языка L состоит из N букв. Задано несколько слов $v_1, \dots, v_k$, которые называются* запретными *и в языке не употребляются. Слово (то есть ограниченная последовательность букв) называется* допустимым*, если никакая часть этого слова не является запретным словом. Сколько в языке L возможно допустимых слов из n букв?*

Задача 4. *Докажите, что задачи 1 и 2 сводятся к задаче 3.*

2 Как записывать ответ?

Зафиксируем какой-нибудь алфавит A из N букв (например, если $A = (a, b, c, \dots, z)$, то $N = 26$). *Словом* мы будем называть любую конечную последовательность букв алфавита A. *Подсловом* мы будем называть часть слова, состоящую из идущих подряд в этом слове букв.

Мы будем считать, что в каждом языке L есть ровно одно слово из нуля букв — *пустое* слово.

Мы будем считать, что запретные слова не содержатся друг в друге, т.е. никакое подслово запретного слова, кроме него самого, не является снова запретным. Кроме того, мы будем считать, что запретные слова состоят как минимум из двух букв, т.е. что пустое слово и отдельные буквы являются допустимыми словами. Напомним, что множество запретных слов конечно.

Задача 5. Свободным языком *алфавита A называется язык F_A, в котором вообще нет запретных слов. Докажите, что количество слов из n букв в этом языке равно N^n.*

Задача 6. *В языке B запретными являются все слова из двух различных букв. Докажите, что для любого натурального n количество допустимых слов из n букв в этом языке равно N.*

Пусть M — какое-нибудь множество слов. Обозначим через m_n количество слов в этом множестве, состоящих из n букв. *Рядом размеров* множества M называется бесконечная сумма

$$M(x) = m_0 + m_1 x + m_2 x^2 + m_3 x^3 + \dots$$

Такого вида бесконечные суммы (с произвольными числами в качестве коэффициентов m_n) мы будем кратко называть просто *рядами* (их полное название, которое мы не будем использовать — *формальные степенные ряды*).

Для каждого языка L его рядом размеров $L(x)$ называется ряд размеров множества допустимых слов. Например, для свободного языка F_A ряд размеров — это сумма геометрической прогрессии $F_A(x) = 1 + Nx + N^2x^2 + N^3x^3 + \dots$, а для языка B это $B(x) = 1 + Nx + Nx^2 + Nx^3 + \dots$

Задача 7. *Выпишите ряд размеров для языка над алфавитом $\{a, b\}$, в котором запретными являются слова aa и bb.*

[1] Даже если слова в других словарных формах.

[2] Подобная история в 1990-е гг произошла с первой версией микропроцессора *Pentium*.

3 Арифметика языков

Если множество M содержит только ограниченное количество слов, то его ряд размеров — это многочлен от переменной x. Для бесконечных множеств и ряды тоже бесконечные, но с ними можно производить разные арифметические операции, похожие на операции с многочленами, то есть складывать, вычитать, умножать друг на друга и на числа и даже иногда делить.

В определениях и задачах этого раздела $S = s_0 + s_1x + s_2x^2 + \dots$ и $R = r_0 + r_1x + r_2x^2 + \dots$ — два ряда, а L_1 и L_2 — какие-то два языка с разными алфавитами A_1 и A_2. Для определённости, мы будем считать, что алфавит A_1 состоит из заглавных букв, а алфавит A_2 — из строчных. Алфавит A — это объединение двух алфавитов A_1 и A_2, то есть в него входят и заглавные, и строчные буквы.

Определение 1. а) *Суммой рядов* R и S называется сумма

$$R + S = (s_0 + r_0) + (s_1 + r_1)x + (s_2 + r_2)x^2 + \dots$$

б) *Суммой языков* L_1 и L_2 называется язык $L_1 + L_2$ над алфавитом A, у которого множество допустимых слов — объединение множеств допустимых слов языков L_1 и L_2.

Задача 8. *Задайте язык $L_1 + L_2$ конечным множеством запретных слов.*

Задача 9. *Докажите, что если $L = L_1 + L_2$, то*

$$L(x) = L_1(x) + L_2(x) - 1.$$

Произведение рядов размеров определяется так же, как произведение многочленов.

Определение 2. *Произведением ряда R на одночлен ax^n* называется ряд

$$R \cdot ax^n = ar_0x^n + ar_1x^{n+1} + ar_2x^2x^{n+2} + \dots$$

Произведением рядов R и S называется сумма

$$R \cdot S = R \cdot s_0 + R \cdot s_1x + R \cdot s_2x^2 + \dots$$

Заметим, что эта бесконечная сумма рядов имеет смысл — складывая ряды почленно, мы получаем в качестве коэффициента при каждой степени x конечную сумму чисел.

Задача 10. *Докажите равенство*

$$(1 - x) \cdot (1 + x + x^2 + \dots) = 1.$$

Определение 3. *Произведением двух множеств слов M, N* называется множество MN всех слов вида mn, где m — слово из M, а n — слово из N.

Произведением двух языков L_1 и L_2 называется язык $L_1 \cdot L_2$ над алфавитом A, у которого множество допустимых слов является произведением множеств допустимых слов языков L_1 и L_2.

Задача 11. *Задайте язык $L_1 \cdot L_2$ конечным множеством запретных слов.*

Задача 12. *Докажите равенство*

$$L(x) = L_1(x) \cdot L_2(x).$$

Деление рядов не имеет аналога для языков, но позволяет сокращённо записывать их ряды размеров. Оно определяется по формуле, похожей на формулу суммы геометрической прогрессии.

Определение 4. Предположим, что ряд S начинается с единицы, то есть $s_0 = 1$, и $S = 1 + \overline{S}$, где $\overline{S} = s_1x + s_2x^2 + \dots$ Тогда *обратным рядом* называется ряд

$$\frac{1}{S} = 1 - \overline{S} + \overline{S}^2 - \overline{S}^3 + \dots$$

Частным от деления рядов R и S называется ряд

$$\frac{R}{S} = R - R \cdot \overline{S} + R \cdot \overline{S}^2 - R \cdot \overline{S}^3 + \dots$$

Частное двух рядов размеров языков может не соответствовать никакому языку, хотя бы потому, что в получившемся ряде могут появиться отрицательные коэффициенты.

Задача 13. *a) Докажите, что*

$$S \cdot \frac{R}{S} = R.$$

b) Докажите, что если $S \cdot T = R$, где ряд S начинается с единицы, то $T = \frac{R}{S}$.

Польза от операции деления рядов состоит в том, что многие бесконечные ряды можно записать с её помощью в виде конечного выражения — частного двух многочленов.

Задача 14. *а) Докажите, что*

$$F_A(x) = \frac{1}{1 - Nx}.$$

б) Запишите ряды размеров языков из задач 6 и 7 в виде частного двух многочленов.

Задача 15. *Докажите, что ряд размеров любого языка может быть записан в виде частного двух многочленов.*

Таким образом, ответ к Главной Задаче должен быть представим в виде частного двух многочленов.

4 Свободное слово

Задача 16. *Пусть L — язык над латинским алфавитом, в котором запретным является только слово "mouse". Найдите $L(x)$.*

Определение 5. Пусть a, b — два слова, из которых ни одно не является частью другого. Непустое слово c называется *зацеплением* слов a и b, если оно является окончанием слова a и в то же время началом слова b (например, слово "ко" — зацепление слов "молоко" и "корова").

Слово называется *свободным*, если у него нет никаких зацеплений с самим собой, кроме самого этого слова.

Задача 17. *Пусть в языке L над алфавитом A из N букв имеется только одно запретное слово — некоторое свободное слово из m букв. Докажите, что*

$$L(x) = \frac{1}{1 - Nx + x^m}.$$

Задача 18. *Решите задачу 2 в предположении, что последовательность, ломающая компьютер, является свободным словом.*

5 Преобразования слов

Определение 6. Пусть M и M' — два множества слов. Разобьём множество M на какие-нибудь две части K и L. Функция f из множества L в какое-нибудь подмножество I множества M' называется *преобразованием* множества M в множество M', если f сохраняет длину слова и является взаимно однозначным отображением из L в I.

В этом случае множество K называется *ядром* отображения f, а множество I — его *образом.*

Преобразование мы будем обозначать стрелкой: $M \Longrightarrow M'$.

Определение 7. Цепочка преобразований

$$M_1 \Longrightarrow M_2 \Longrightarrow \ldots \Longrightarrow M_n$$

называется *точной*, если ядро каждого следующего преобразования совпадает с образом предыдущего.

Задача 19. *Пусть L — язык над алфавитом A с множеством допустимых слов G и множеством всех не допустимых слов N. Постройте точную последовательность преобразований*

$$\emptyset \Longrightarrow N \Longrightarrow F_A \Longrightarrow G \Longrightarrow \emptyset,$$

где F_A — множество слов свободного языка, то есть всех слов над алфавитом A, а $\emptyset$ обозначает пустое множество.

Задача 20. *В ряд сидят по очереди мальчики и девочки, по 10 тех и других; последней сидит учительница. У детей есть конфеты, поровну в сумме у мальчиков и у девочек. Первый мальчик отдаёт все свои конфеты сидящей за ним девочке. Девочка съедает их, съедает из своих конфет столько же, а остаток отдаёт следующему мальчику. Тот тоже поступает точно так же, за ним — следующая девочка, и так далее. Последняя девочка отдаёт остаток своих конфет учительнице. Сколько ей достанется?*

Задача 21. *Пусть*
$$\emptyset \Longrightarrow M_1 \Longrightarrow M_2 \Longrightarrow \ldots \Longrightarrow M_n \Longrightarrow \emptyset$$
— точная цепочка преобразований.

а) Докажите, что если в каждом множестве M_i только конечное число m_i слов, то
$$m_1 + m_3 + m_5 + \dots = m_2 + m_4 + \dots$$

б) Докажите формулу для рядов размеров
$$M_1(x) + M_3(x) + M_5(x) + \dots = M_2(x) + M_4(x) + \dots$$

Определение 8. Множество слов M называется *свободным,* если никакое слово из M не является подсловом другого слова из этого множества, все слова в нём свободные и не имеют зацеплений между собой.

Задача 22. *Пусть L — язык над алфавитом A, у которого множество запретных слов B свободное. Обозначим через G множество его допустимых слов, через $\overline{G}$ — множество всех допустимых слов, кроме пустого. Постройте точную последовательность преобразований*
$$\emptyset \Longrightarrow B \cdot G \Longrightarrow A \cdot G \Longrightarrow \overline{G} \Longrightarrow \emptyset.$$

Задача 23. *Пусть L — язык над алфавитом A из N букв, у которого множество запретных слов B свободное. Докажите формулу*
$$L(x) = \frac{1}{1 - Nx + B(x)}.$$

Задача 24. *Докажите, что множество заклинаний в задаче 1 свободное, и решите её.*

Задача 25. *Найдите $L(x)$, если алфавит языка L — латинский, а запретные слова — это слова veni, vidi, vici.*

Определение 9. Пусть L — язык. *Простой сцепкой* называется слово $v = str$, где s, t, r — непустые слова, причём $g = st$ и $f = tr$ — запретные слова, и больше никаких запретных подслов в v нет. Конец r простой сцепки, остающийся после первого запретного слова g, называется её хвостом.

Задача 26. *Докажите, что множество запретных слов языка является свободным в том и только том случае, если в этом языке нет простых сцепок.*

Задача 27. *Пусть L — язык над некоторым алфавитом A с множеством запретных слов B и множеством простых сцепок S. Обозначим через G множество его допустимых слов, через $\overline{G}$ — множество всех допустимых слов, кроме пустого. Постройте точную последовательность преобразований*
$$S \cdot G \Longrightarrow B \cdot G \Longrightarrow A \cdot G \Longrightarrow \overline{G} \Longrightarrow \emptyset.$$

Задача 28. *Каким условиям должно удовлетворять множество запретных слов языка L, чтобы точную последовательность из задачи 27 можно было продолжить до последовательности*
$$\emptyset \Longrightarrow S \cdot G \Longrightarrow B \cdot G \Longrightarrow A \cdot G \Longrightarrow \overline{G} \Longrightarrow \emptyset$$
(такие языки назовём незапутанными)? Выведите формулу, которая выражала бы ряд размеров $L(z)$ незапутанного языка через число N букв в алфавите и ряды размеров множеств B и S.

Задача 29. *Вычислите ряд размеров для языка над алфавитом из трёх букв a, b, c с запретными словами abb, bbc, bac.*

Задача 30. *Вычислите ряд размеров для языка над алфавитом $A = \{x_1, \ldots, x_n, y_1, \ldots, y_n, z_1, \ldots, z_n\}$, в котором запретными являются все слова вида $x_i y_j$ и $y_j z_k$, где $1 \leq i, j, k \leq n$.*

Задача 31. *Докажите, если множество запретных слов незапутанного языка состоит только из одного слова, то это множество свободно.*

6 Ещё о свободных множествах

Задача 32. *Постройте бесконечное свободное множество в алфавите из двух букв.*

Задача 33. *Предположим, что множество запретных слов B языка L свободно, а алфавит содержит более одной буквы. Докажите, что множество допустимых слов этого языка бесконечно.*

Определение 10. Пусть $S = s_0 + s_1x + s_2x^2 + \dots$ и $R = r_0 + r_1x + r_2x^2 + \dots$ — два ряда. Если для любых коэффициентов s_k и r_k с одинаковыми номерами выполняется неравенство $s_k \geq r_k$, то будем говорить, что между рядами выполняется неравенство

$$S \geq R.$$

Задача 34. *Докажите, что если для рядов P, Q и R выполняются неравенства*

$$P \geq Q \text{ и } R \geq 0,$$

то

$$PR \geq QR.$$

Задача 35. *Предположим, что множества запретных слов B и B' двух языков L и L' над одним алфавитом A содержат одно и то же количество слов каждой длины, так что $B(z) = B'(z)$. Докажите, что если множество B свободно, то выполняется неравенство*

$$L'(z) \geq L(z),$$

причём равенство $L'(z) = L(z)$ достигается в том и только том случае, когда множество B' также является свободным.

Задача 36. *Известно, что алфавит состоит из двух букв, а множество B содержит не менее двух слов, одно из которых, слово w, имеет длину 2.*

a) Докажите, что множество B не является свободным.

b) Может ли оно быть свободным, если длина слова w равна 3?

Задача 37. *Известно, что алфавит состоит из n букв, а множество B состоит из g слов длины 2. Докажите, что если $g \leq n^2/4$, то множество B может быть выбрано свободным.*

Задача 38. *Докажите, что если $n = kd$ и $m \leq k^d(d-1)^{d-1}$, где числа d, k, m, n натуральные, то над алфавитом из n букв можно выбрать свободное множество, состоящее из m слов длины d.*

Задача 39. *а) Докажите, что если B — свободное множество над алфавитом из n букв, то выполняется неравенство*

$$\frac{1}{1 - nx + B(x)} \geq 1.$$

б) Верно ли обратное утверждение: если для некоторого натурального n и некоторого многочлена $p(x)$ с неотрицательными целыми коэффициентами и нулевым свободным членом выполняется неравенство

$$\frac{1}{1 - nx + p(x)} \geq 1,$$

то над алфавитом из n букв существует свободное множество B такое, что $B(x) = p(x)$?

Задача 40. *Пусть n — натуральное число, $p(x)$ — многочлен с неотрицательными целями коэффициентами и нулевым свободным членом. Докажите, что свободное множество B с рядом размеров $B(x) = p(x)$ существует в том и только том случае, когда существуют такие многочлены f и g с неотрицательными целыми коэффициентами такие без свободных членов, что*

$$(1-f)(1-g) \geq 1 - nx + p(x).$$

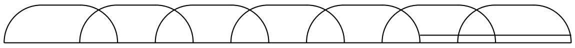
Задача 41. [3] *Придумайте условие, описывающее возможные ряды размеров множеств запрещённых слов незапутанных языков (подобно тому, как в задаче 40 охарактеризованы ряды размеров свободных множеств).*

7 Слова и цепи

Определение 11. Пусть L — язык. Цепями длины 1 называются все запретные слова, цепями длины 2 — все простые сцепки. Через эти цепи определяются ещё цепи длины 3, 4 и так далее. А именно, слово $v = str$ (где все слова s, t, r — непустые) называется *цепью длины* n, если его начало $g = st$ является цепью длины $n-1$, конец $f = tr$ — запретным словом, причём t является подсловом хвоста p цепи g, и никаких запретных подслов, кроме f, в конечном участке pr нет. *Хвостом* этой цепи называется слово r.

[3] Жюри не известны ни решение, ни даже ответ к этой задаче

Цепь выглядит примерно так (каждая дуга — это запретное слово в ней):



Длина цепи — это количество дуг. Зацепляются только соседние дуги (т. е. их пересечение – зацепление ненулевой длины), и выделенные два последних хвоста не содержат запретных слов, кроме последней дуги.

Например, если запретное слово — aba, то единственная цепь длины 1 — это aba, длины 2 — $ababa$, длины 3 — $abababa$, и так далее.

Задача 42. *Пусть в языке L запретными считаются слова “tournament”, “of”, “towns”. Выпишите все цепи длины n.*

Задача 43. Антицепь *длины n определяется так же, как и цепь длины n, но все слова языка L в определении 11 прочитывается “справа налево”, т. е. хвосты антицепей находятся слева, а начальная цепь длины $n-1$ — справа. Докажите, что множества цепей длины n и антицепей длины n совпадают.*

Задача 44. *Докажите, что никакая цепь длины n не содержит в качестве подслова никакую другую цепь длины n.*

Задача 45. *Докажите, что если слово имеет вид $w = gc$, где g — допустимое слово, а c — цепь, то в случае, если длина цепи c больше 1, слово w представимо в таком виде ровно двумя способами, причём длины цепей в этих представлениях различаются на единицу.*

Следующая задача даёт способ решения Главной Задачи.

Задача 46. *Пусть L — язык над алфавитом A. Обозначим через G множество его допустимых слов, через $\overline{G}$ — множество всех допустимых слов, кроме пустого. Пусть C_1 — множество цепей длины 1, C_2 — цепей длины 2, и так далее.*

Докажите формулу

$$L(x) = \frac{1}{1 - Nx + C_1(x) - C_2(x) + C_3(x) - \dots}$$

Задача 47. *Найдите ряд размеров для языка из задачи 42.*

Задача 48. *Найдите все возможные варианты ответов для задачи 2 в зависимости от того, какие именно команды ломают компьютер.*

Задача 49. *Назовём подслово c слова w* максимальной подцепью*, если w представимо в виде $w = gcu$, где g — допустимое слово, а c — цепь, причём для любого другого представления $w = gc'u'$ с другой цепью c' всегда слово c' — подслово слова c. Докажите, что любое недопустимое слово имеет ровно одну максимальную подцепь нечётной длины.*

Задача 50. *Пусть L — язык над алфавитом A, и A' — новый алфавит, полученный из A добавлением одной буквы. Пусть L' — язык над алфавитом A', в котором запретными являются все запретные слова языка L. Докажите формулу*

$$L'(x) = \frac{1}{\dfrac{1}{L(x)} - x}.$$

Задача 51. *Язык W называется* свободным произведением *языков L и L' над непересекающимися алфавитами A и A', если алфавит языка W есть объединение алфавитов A и A', а множество запретных слов — объединение множеств запретных слов языков L и L'. Выразите ряд размеров свободного произведения W через ряды размеров языков L и L'.*

Задача 52. *Предположим, что множество запретных слов языка L содержит только слова из двух букв. Рассмотрим другой язык M над тем же алфавитом, в котором запретными являются те и только те двухбуквенные слова, которые не являются запретными в языке L. Докажите равенство*

$$L(x)M(-x) = 1.$$

8 Дополнительные задачи

Задача 53. *Докажите, что свободное множество из m слов длины d в алфавите из $n = kd$ букв существует в том и только том случае, когда $m \leq k^d(d-1)^{d-1}$ (ср. задачу 38), если*

а) $d = 2$; *б) $d = 3$;* *в) $d > 3$.*

Определение 12. Язык называется *d-определённым*, если наибольшая из длин его запретных слов равна d. 2-определённый язык называется *квадратичным*.

Задача 54. *Квадратичные языки L и M из задачи 52 называются* двойственными *друг к другу (обозначение: $M = L^!$).*

а) Докажите, что $(L^!)^! = L$.

б) Найдите $(L_1 + L_2)^!$.

в) Опишите $(L_1 \cdot L_2)^!$.

Задача 55. *Пусть L — d-определённый язык. Определим новый язык $L^{(n)}$, в котором алфавитом являются все допустимые слова языка L длины n, а допустимые слова — все допустимые слова языка L, длина которых делится на n (выраженные через новые буквы).*

а) Докажите, что язык $L^{(n)}$ задаётся конечным набором запретных слов.

б) Всегда ли язык $L^{(n)}$ является d-определённым?

в) При каком наименьшем n язык $L^{(n)}$ гарантированно является квадратичным или свободным (вне зависимости от выбора d-определённого языкам L)?

Задача 56. *Для любого квадратичного языка L над алфавитом $x_1, \ldots, x_n$ определим ориентированный граф Γ_L следующим образом: его вершины — n точек, помеченные буквами $x_1, \ldots, x_n$, а ребро (стрелка) $x_i \to x_j$ проводится в том и только том случае, когда $x_i x_j$ — разрешённое слово. Обозначим через a_k количество допустимых слов из k букв. Докажите, что*

a) язык L конечный в том и только том случае, когда в графе Γ_L нет циклов;

b) язык L имеет полиномиальный рост (т. е. существуют два ненулевых многочлена p, q одной и той же степени d с положительным старшим коэффициентом такие, что $p(k) \geq a_k \geq q(k)$ для всех $k \geq 0$) в том и только том случае, когда в графе Γ_L есть цикл, но нет пересекающихся циклов;

c) язык L имеет экспоненциальный рост (т. е. для некоторых $c_1 > c_2 > 1$ и для всех k выполняются неравенства $c_1^k \geq a_k \geq c_2^k$) тогда и только тогда, когда в графе Γ_L есть хотя бы два пересекающихся цикла.

Задача 57. *Пусть L и $L^!$ — пара двойственных квадратичных языков. Возможно ли, что оба они имеют экспоненциальный рост?*

Задача 58. *Для любого d-определённого языка L над алфавитом $x_1, \ldots, x_n$ определим ориентированный граф Γ_L следующим образом: его вершины помечены всеми допустимыми словами длины $d-1$, а ребро (стрелка) $v \to w$ проводится в том и только том случае, когда при умножении слова v на некоторую букву x_i получается допустимое слово, последние $d-1$ букв которого составляют слово w. Докажите все три свойства a), b), c) из задачи 56 для построенного графа Γ_L.*

Определение 13. Пусть M — некоторое множество слов над алфавитом A. Слова u и v (над тем же алфавитом) называются *M-эквивалентными*, если для любого слова w слова uw и vw либо оба принадлежат M, либо оба не принадлежат M. Множество M называется *регулярным*, если найдётся такое натуральное число n, что в любом множестве из n слов найдутся два M-эквивалентных друг другу слова.

Задача 59. *Докажите, что множество допустимых слов любого языка регулярно.*

Определение 14. *Конечным автоматом* над алфавитом A называется ориентированный граф Γ с конечным множеством вершин V, причём
a) стрелки помечены буквами алфавита A, причём для любой буквы $a \in A$ из каждой вершины выходит ровно одна стрелка, помеченная a;
b) выделены *начальная вершина* $v_0 \in V$ и множество *принимающих вершин* $W \subseteq V$.

Будем воспринимать каждое слово над алфавитом A как инструкцию для путешествия по стрелкам конечного автомата (Γ, v_0, W): начинаем с начальной вершины, идём из неё по стрелке, помеченной первой буквой слова, дальше идём по стрелке, помеченной второй буквой, и т.д. Мы говорим, что автомат *принимает* слово, если соответствующий слову путь заканчивается в принимающей вершине.

Задача 60. *a) Докажите, что для любого регулярного множества M существует конечный автомат, принимающий слова из M и никаких больше.*

b) Докажите, что для любого конечного автомата множество принимаемых им слов регулярно.

Задача 61. *Докажите, что для любого регулярного множества M его ряд размеров может быть записан в виде частного двух многочленов.*

Задача 62. *Пусть M_w — множество всех допустимых слов языка L, оканчивающихся на фиксированное подслово w. Докажите, что ряд размеров множества M_w представим в виде частного двух многочленов.*

(До промежуточного финиша были предложены части 1–5, после промежуточного финиша добавлены части 6–8.)

Как считать слова?

Дмитрий Пионтковский, Максим Прасолов, Григорий Рыбников

Решения

1 Главная задача

1. См. задачу 24.

2. См. задачи 18 и 48.

3. Один из вариантов решения дан в задаче 46, а другой может быть получен с помощью задач 59 и 61.

4. В задаче 1 алфавит A состоит из $N = 100$ слов языка племени, роль слов языка L играют фразы языка племени Винни-Пухов, а запретные слова языка L — это два магических заклинания. В задаче 2 алфавит A состоит из $N = 256$ команд компьютера, а роль слов языка L играют программы. Единственное запретное слово — программа, ломающая компьютер.

2 Как записывать ответ?

5. Слово из m букв получается выбором на каждом из m мест любой из N букв. Перемножая количество возможностей на каждом месте, получаем N^m слов.

6. Если в допустимом слове первая буква x, то вторая тоже. Аналогично остальные. Значит, допустимое слово имеет вид $xx\dots x$, где x – одна из N букв алфавита. Следовательно допустимых слов из фиксированного количества букв N штук.

7. Пусть в допустимом слове первая буква a. Так как aa – запрещённое слово, то вторая буква данного допустимого слова b. Продолжаем рассуждения и получаем, что на нечётном месте стоит буква a, а на чётном – буква b. Если первая буква b, то на чётном месте стоит буква a, а на нечётном – b. Значит, ряд размеров данного языка таков: $1 + 2x + 2x^2 + 2x^3 + \dots$

3 Арифметика языков

8. Набор запретных слов таков: всевозможные слова из двух букв, в которых одна буква из первого алфавита, а другая – из второго, и запретные слова обоих языков. Очевидно, что допустимое слово каждого языка не содержит запретных слов в сумме языков. Если в не содержащем запретное подслово слове суммы языков первая буква принадлежит первому алфавиту, то вторая тоже, и аналогично остальные. Другими словами такое слово состоит из букв одного алфавита. Но тогда оно является допустимым в языке с этим алфавитом, а значит, допустимым в сумме языков.

9. Свободные члены рядов $L(x)$ и $L_1(x) + L_2(x) - 1$ равны единице. При $n > 0$ коэффициент при x^n ряда $L_1(x) + L_2(x) - 1$ равен сумме количеств слов длины n в языках L_1 и L_2, то есть количеству слов длины n в языке L, то есть коэффициенту при x^n в ряде $L(x)$.

10. Имеем

$$(1-x)(1+x+x^2+x^3+\dots) = 1 - x + (1-x)\cdot x + (1-x)\cdot x^2 + (1-x)\cdot x^3 + \dots =$$
$$= 1 - x + x - x^2 + x^2 - x^3 + x^3 - x^4 + \dots = 1,$$

что и требовалось.

11. Набор запретных слов таков: все слова из двух букв, в которых первая буква из второго алфавита, а вторая – из первого, и все запретные слова языков произведения. Рассмотрим допустимое слово произведения. В нём буквы второго алфавита следуют после букв первого алфавита, и это слово не содержит запретных слов языков-множителей. Значит, допустимое слово произведения не содержит указанных запретных слов. Теперь возьмём слово, не содержащее указанных запретных слов. В нём буквы второго алфавита следуют после букв первого алфавита, поэтому такое слово имеет вид w_1w_2, где w_1 – слово из букв первого алфавита, а w_2 – из букв второго. Слово w_1w_2 не содержит запретных слов языков-множителей, поэтому слова w_1 и w_2 допустимы в своих языках, то есть слово w_1w_2 допустимо в произведении языков.

12. Коэффициент при x^k в ряде $L_1(x)\cdot L_2(x) = L_1(x)\cdot n_0 + L_1(x)\cdot n_1x + \dots = (n_0\cdot m_0 + n_0m_1x + \dots) + (n_1m_0x + n_1m_1x^2 + \dots) + \dots$ равен $n_0m_k + n_1m_{k-1} + \dots + n_km_0$. Количество слов длины k в множестве слов $L_1\cdot L_2$ равно числу способов выбрать пару слов m из языка L_1 и n из языка L_2, суммарное количество букв которых равно k. Если в слове m i букв, то в слове n $k{-}i$ букв, а количество таких пар равно $m_i\cdot n_{k-i}$. Суммируя такие выражения для всех i, получаем $n_0m_k + n_1m_{k-1} + \dots + n_km_0$. Поэтому коэффициенты при x^k в рядах $L_1(x)\cdot L_2(x)$ и $L_1\cdot L_2(x)$ совпадают, следовательно сами ряды совпадают.

13.

а) В решении этой задачи нам потребуются тот факт, что стандартные свойства умножения и сложения многочленов (ассоциативность, коммутативность, дистрибутивность) выполняются и для рядов. Рассмотрим, например, ассоциативность умножения — $(P\cdot Q)\cdot R = P\cdot(Q\cdot R)$. Чтобы вычислить коэффициент при x^k в левой и правой части, достаточно его вычислить для рядов, у которых отброшены члены более высокой степени, то есть для многочленов. Поэтому указанное тождество для рядов следует из того же тождества для многочленов. Аналогично доказываются остальные свойства.

Заметим теперь, что поскольку ряд $\overline{S}$ начинается с первой степени x, в ряд $R\cdot\overline{S}^m$ не входят степени x от нулевой до $m-1$-й. Именно поэтому бесконечные суммы вида $R - R\cdot\overline{S} + R\cdot\overline{S}^2 - R\cdot\overline{S}^3 + \dots$ имеют смысл — для вычисления k-го коэффициента можно заменить эту сумму конечной. По той же причине для бесконечных сумм такого вида выполняется свойство дистрибутивности

$$R\cdot(S_1 + S_2 + S_3 + \dots) = R\cdot S_1 + R\cdot S_2 + R\cdot S_3 + \dots.$$

Имея это в виду, мы легко получаем тождество

$$(1+\overline{S})(1 - \overline{S} + \overline{S}^2 - \overline{S}^3 + \dots) = 1.$$

Отсюда

$$S\cdot\frac{R}{S} = S\cdot(R - R\cdot\overline{S} + R\cdot\overline{S}^2 - R\cdot\overline{S}^3 + \dots) = S\cdot R\cdot(1 - \overline{S} + \overline{S}^2 - \overline{S}^3 + \dots) =$$
$$= R\cdot(1+\overline{S})(1 - \overline{S} + \overline{S}^2 - \overline{S}^3 + \dots) = R.$$

б) Согласно утверждению а),

$$S\cdot(T - \frac{R}{S}) = S\cdot T - S\cdot\frac{R}{S} = R - R = 0$$

Предположим, что ряд $(T - \frac{R}{S})$ ненулевой. Так как ряд S начинается с единицы, то первый ненулевой коэффициент ряда $(T - \frac{R}{S})$ равен первому ненулевому коэффициенту ряда $S\cdot(T - \frac{R}{S})$. Значит, ряд $(T - \frac{R}{S})$ нулевой и $T = \frac{R}{S}$.

14.

а) Рассуждая аналогично задаче 10, получаем

$$F_A(x) = 1 + Nx + N^2x^2 + \dots = \frac{1}{1-Nx}.$$

б) Получаем

$$1 + Nx + Nx^2 + Nx^3 + \dots = -N + 1 + N\cdot(1 + x + x^2 + x^3 + \dots) = -N + 1 + \frac{N}{1-x} = \frac{1+(N-1)x}{1-x}$$

и

$$1 + 2x + 2x^2 + 2x^3 + \dots = 1 + 2x\cdot(1 + x + x^2 + x^3 + \dots) = 1 + \frac{2x}{1-x} = \frac{1+x}{1-x}.$$

15. Следует из задач 59 и 61.

4 Свободное слово

16. Как следует из задачи 17 ниже, получаем $L(x) = \frac{1}{1-26x+x^5}$. Можно и непосредственно получить эту формулу, проведя рассуждения, аналогичные решению задачи 17.

17. Пусть a_k — число допустимых слов длины k. Ясно, что $a_0 = 1$. Докажем, что при $k > 0$ выполнено рекуррентное соотношение $a_k = Na_{k-1} - a_{k-m}$ (мы считаем, что $a_k = 0$ при $k < 0$, так как слов отрицательной длины не существует).

Действительно, приписывая к началу каждого допустимого слова из $k-1$ буквы каждую букву алфавита, мы получаем Na_{k-1} слов, среди которых все допустимые слова длины k. Посмотрим, какие недопустимые слова длины k получаются таким образом, то есть имеют вид cg, где c — буква, а g — допустимое слово длины $k-1$. Ясно, что запретное подслово должно стоять в начале, то есть $cg = wf$, где w — запретное слово, а f — допустимое. Из того, что w — свободное слово, следует, что для любого допустимого слова f слово, получающееся отбрасыванием первой буквы в слове wf, допустимо (в противном случае w имело бы зацепление с самим собой). Поэтому множество всех слов вида cg, где c — буква, а g — допустимое слово длины $k-1$, является объединением двух непересекающихся множеств: множества допустимых слов длины k и множества слов вида wf, где f — допустимое слово длины $k-m$. Отсюда следует нужное нам рекуррентное соотношение.

Рассмотрим сумму соотношения $a_0 = 1$ и всех соотношений $a_k x^k = Na_{k-1}x^k - a_{k-m}x^k$ для $k = 1, 2, 3, \ldots$. Получится

$$L(x) = 1 + NxL(x) - x^m L(x).$$

Решая это уравнение относительно $L(x)$, получаем нужную нам формулу.

18 . Согласно задаче 17,

$$L(x) = \frac{1}{1-256x+x^4} = 1 + (256x - x^4) + (256x - x^4)^2 + (256x - x^4)^3 + \ldots$$

Коэффициент при x^7 равен $256^7 - 4 \cdot 256^3$. Поэтому вероятность поломки компьютера равна $\frac{4 \cdot 256^3}{256^7} = 4 \cdot 256^{-4}$, что примерно равно 10^{-10}.

5 Преобразования слов

19. Первая стрелка определена однозначно; вторая сопоставляет каждому слову из N то же самое слово, рассматриваемое как элемент F_A; третья сопоставляет каждому из оставшихся слов F_A его же, как элемент G; последняя стрелка тривиальна, как и первая.

20. Каждый из школьников съедает поровну конфет, бывших у мальчиков, и конфет, бывших у девочек. Последняя девочка доест все конфеты, бывшие у мальчиков. Поэтому и конфеты, бывшие у девочек, она тоже доест, и учительнице ничего не достанется.

21.
а) Пусть $M_{\text{odd}} = M_1 \cup M_3 \cup M_5 \cup \ldots$ и $M_{\text{even}} = M_2 \cup M_4 \cup M_4 \cup \ldots$. Каждое преобразование устанавливает взаимно-однозначное соответствие между некоторым подмножеством M_{odd} и M_{even}, причём, поскольку по краям стоят пустые множества, каждый элемент участвует ровно в одном из этих взаимно-однозначных соответствий. Поэтому во множествах M_{odd} и M_{even} поровну элементов.
б) Для каждого k множество $M_i^{(k)}$ слов из M_i длины k конечно; применяя к конечным множествам $M_i^{(k)}$ с данным k утверждение пункта а), получаем, что коэффициенты при x^k в левой и правой частях доказываемой формулы совпадают. Поскольку k — любое, это означает, что формула верна.

22. Проверим, что множество $A \cdot G$ является объединением двух непересекающихся множеств: множества $\overline{G}$ и множества $B \cdot G$. Доказательство этого опирается на то, что множество B — свободное, и почти буквально повторяет соответствующее рассуждение в решении задачи 17.

Теперь точная последовательность строится очевидным образом: первая и последняя стрелки тривиальны, вторая переводит каждый элемент $B \cdot G$ в себя (мы пользуемся тем, что $B \cdot G \subseteq A \cdot G$), а третья переводит в себя каждый из оставшихся элементов $A \cdot G$ (мы пользуемся тем, что $A \cdot G \setminus B \cdot G = \overline{G}$). В частности, ядром второго преобразования является пустое множество, а ядром третьего преобразования (и образом второго) — множество $D \cdot G$; образом третьего преобразования является $\overline{G}$.

23. По задаче 21б), из точной последовательности задачи 22 получаем

$$(B \cdot G)(x) + \overline{G}(x) = (A \cdot G)(x).$$

Заметим, что каждый элемент $A \cdot G$ записывается в виде ag, где $a \in A, g \in G$, однозначно. Поэтому $(A \cdot G)(x) = A(x)G(x)$. Далее, каждый элемент $B \cdot G$ записывается в виде bg, где $b \in B, g \in G$, однозначно (поскольку запрещённое слово не может быть подсловом другого запрещённого слова). Поэтому $(B \cdot G)(x) = B(x)G(x)$. Мы имеем $A(x) = Nx$, $G(x) = L(x)$, $\overline{G}(x) = G(x) - 1 = L(x) - 1$. Отсюда

$$B(x)L(x) + L(x) - 1 = NxL(x).$$

Решая это уравнение относительно $L(x)$, получаем требуемую формулу.

24. Обозначим встречающиеся в фразах слова буквами: "земля" — A, "стоять" — B, "на" — C, "великий" — D, "крокодил" — E, "каждый" — F, "вечер" — G, "глотать" — H, "солнце" — I. Тогда заклинаниям отвечают запретные слова "ABCDE" и "FGEHI". Эти слова свободны (потому что в каждом все буквы различны) и не имеют зацеплений друг с другом (потому что первая и последняя буквы второго слова не встречаются в первом). Следовательно, множество заклинаний свободно, и ряд размеров языка имеет вид

$$L(x) = \frac{1}{1 - 100x + 2x^5}.$$

Из этой формулы легко вывести (обращая рассуждения в решении задачи 17), что числа a_k (количество фраз из k слов) можно вычислить из начального условия $a_0 = 1$ и рекуррентного соотношения $a_k = 100a_{k-1} - 2a_{k-5}$. Вычисления приводят к ответу $a_{20} = 10^{40} - 32 \cdot 10^{30} + 264 \cdot 10^{20} - 448 \cdot 10^{10} + 16$.

25. В каждом запретном слове буква v встречается только на первом месте и все слова имеют длину 4, поэтому множество запретных слов свободно. По задаче 23, мы имеем

$$L(x) = \frac{1}{1 - 26x + 3x^4}.$$

26. Если множество запретных слов свободно, то, в частности, нет простых сцепок. Докажем обратное: если нет простых сцепок, то множество запретных слов свободно. Предположим противное: пусть множество запретных слов свободным не является. Тогда найдётся зацепление двух запретных слов, то есть найдутся такие три непустых слова s, t, r, что слова st и tr — запретные. Выберем такую тройку (s, t, r) так, чтобы она имела минимально возможную суммарную длину. Если это не простая сцепка, то str содержит запретное подслово w, отличное от st и tr. Заметим, что конец w не совпадает с концом tr, поскольку иначе либо w является подсловом tr, либо tr является подсловом w, что невозможно, так как никакое запретное слово не содержит другое запретное в качестве подслова. Аналогично, начало w не совпадает с началом st. Из тех же соображений, подслово w имеет общую часть как с подсловом s, так и с подсловом r. Обозначим через t' общую часть подслов st и w, через s' — остаток подслова st, а через r' — остаток подслова w. Эти слова непусты, их суммарная длина меньше суммарной длины слов s, t, r, а слова $s't' = st$ и $t'r' = w$ — запретные. Получено противоречие. Таким образом, из отсутствия простых сцепок следует, что множество запретных слов — свободное.

27. Будем строить преобразования последовательно, начиная с конца (с самой правой стрелки). Так как последнее множество пусто, то и область определения последнего преобразования — пустое множество. Поэтому предпоследнее преобразование имеет образом всё множество $\overline{G}$. Так как $\overline{G} \subseteq A \cdot G$, мы можем взять $\overline{G}$ в качестве области определения предпоследнего преобразования и считать, что каждый элемент $g \in \overline{G}$ переводится этим преобразованием в себя. Ядро этого преобразования состоит из не являющихся допустимыми слов вида ag, где a — буква и слово g является допустимым. Ясно, что для любого такого слова найдутся запретное слово w и допустимое слово f такие, что $ag = wf$ (мы уже использовали это соображение при решении задач 17 и 22). Это позволяет построить третью с конца стрелку (это преобразование также переводит каждый элемент своей области определения в себя). Рассмотрим ядро этого преобразования. Оно состоит из тех слов вида wf, где w — запретное, а f — допустимое, которые имеют вид av, где a — буква, а слово v допустимым не является. Выберем в v самое первое (начинающееся левее всех) запретное подслово u. Легко видеть, что подслово u слова $av = wf$ имеет общую часть с подсловом w и образует вместе с ним простую сцепку. Таким образом, ядро третьей с конца стрелки целиком содержится в $S \cdot G$, что позволяет построить преобразование $S \cdot G \Longrightarrow B \cdot G$ (также переводящее каждый элемент своей области определения в себя).

28. Из решения предыдущей задачи получаем, что язык является незапутанным тогда и только тогда, когда допустимы все слова вида rg, где r — хвост простой сцепки, а g — допустимое слово. Эквивалентное условие на множество запретных слов звучит так: не существует таких слов p, q, r, s, t, где слова p, q, s, t непусты, слова pq, qrs, st — запретные, причём $pqrs$ — простая сцепка.

Заметим, что любой элемент множества $S \cdot G$ однозначно представляется в виде произведения простой сцепки на допустимое слово (это легко следует из определения простой сцепки и того, что никакое запретное слово не является подсловом другого запретного слова). Аналогично решению задачи 23, для незапутанного языка L из точной последовательности мы получаем уравнение

$$S(x)L(x) + NxL(x) = B(x)L(x) + L(x) - 1,$$

откуда и получается искомая формула

$$L(x) = \frac{1}{1 - Nx + B(x) - S(x)}.$$

29. Простые сцепки имеют вид $abbc, abbac$, их хвосты — c, ac. Видно, что ни один из этих хвостов не оканчивается на непустое начало запретного слова. Поэтому язык является незапутанным. Соответственно, ряд размеров имеет вид

$$L(x) = \frac{1}{1 - 3x + 3x^3 - x^4 - x^5}.$$

30. Простые сцепки имеют вид $x_i y_j z_k$, где $1 \leq i, j, k \leq n$, их хвосты z_k. Так как ни одно запретное слово не начинается на z_k, язык является незапутанным. Соответственно, ряд размеров имеет вид

$$L(x) = \frac{1}{1 - 3nx + 2n^2x^2 - n^3x^3}.$$

31. Пусть w — единственное запретное слово, и пусть L — его длина. Предположим, что оно не является свободным, и pqr, $pq = qr = w$, — простая сцепка. Мы имеем $wr = pqr = pw$. Следовательно, конечный участок длины L у каждого из слов $wr = pw, wrr = pwr = ppw, wrrr = ppwr = pppw, \ldots$ равен w. Возьмём из этих слов первое, имеющее длину не меньше $2L$. Мы получим, что слово вида $rrr\ldots r$ имеет конечный участок, равный w. Но тогда у слова r есть непустое окончание, являющееся началом слова w. Следовательно наш язык не является незапутанным (см. решение задачи 28).

6 Ещё о свободных множествах

32. Например, если алфавит состоит из букв a и b, то свободным будет множество слов вида $a^n b^n ab$, где $n \geq 2$. Докажем это. Очевидно, что никакие два разных слова такого вида не являются подсловами друг друга. Осталось доказать, что между ними нет нетривиальных зацеплений (т.е. что любое зацепление есть просто зацепление слова с самим собой по всей длине). Действительно, если w — зацепление слов $a^n b^n ab$ и $a^m b^m ab$, то легко видеть, что w содержит не менее трёх букв. Как конец слова $a^n b^n ab$, оно имеет вид либо $b^k ab$, либо $a^k b^n ab$, где $1 \leq k \leq n$. Поскольку оно также должно быть началом слова $a^m b^m ab$, получаем, что $k = m = n$ и $w = a^n b^n ab = a^m b^m ab$ — тривиальное зацепление.

33. Лемма. Если $p(x) = 1 + p_1 x + p_n x^n$ — многочлен с единичным свободным членом степени $n \geq 1$, то ряд $f(x) = 1/p(x)$ не может быть многочленом (т.е. этот ряд содержит бесконечное множество ненулевых членов).

Доказательство леммы. Предположим, от противного, что ряд $f(x)$ — многочлен, т.е. $f(x) = f_0 + f_1 x + \ldots f_m x^m$, где старший коэффициент f_m ненулевой. Согласно задаче 13 а), получаем $1 = f(x)p(x) = 1 + (f_0 p_1 + f_1 p_0)x + \cdots + f_m p_n x^{m+n}$ — противоречие.

Перейдём к решению задачи. Согласно задаче 23, имеем

$$L(x) = \frac{1}{1 - Nx + B(x)}.$$

Если бы язык L был конечным, то ряд $L(x)$ был бы многочленом, что невозможно согласно доказанной лемме. Следовательно, множество допустимых слов языка бесконечно.

34. Очевидно, для любых рядов неравенство $A \geq B$ равносильно неравенству $A - B \geq 0$, т.е. условию, что коэффициенты ряда $A - B$ неотрицательны . Обозначим ряд $P - Q$ через $A = a_0 + a_1 x + a_2 x^2 \ldots$, а ряд R как $R = r_0 + r_1 x + r_2 x^2 \ldots$ Тогда произвольный n-й коэффициент ряда AR, вычисляемый по формуле $a_0 r_n + a_1 r_{n-1} + \cdots + a_n r_0$, представим в виде суммы неотрицательных числе, и потому сам неотрицательный. Это означает, что выполняется неравенство $AR \geq 0$, равносильное неравенству $PR - QR \geq 0$, или $PR \geq QR$.

35. По формуле из задачи 23,

$$L(x) = \frac{1}{1 - A(x) + B(x)}.$$

Согласно задаче 27, существует точная последовательность

$$\emptyset \Longrightarrow K \Longrightarrow B' \cdot G' \Longrightarrow A \cdot G' \Longrightarrow \overline{G}' \Longrightarrow \emptyset,$$

где K — ядро отображения $B \cdot G \Longrightarrow A \cdot G'$, а G' — множество допустимых слов языка L'. Из этой последовательности получаем (зад. 21) равенство

$$B'(x)G'(x) - A(x)G'(x) + G'(x) - 1 = K(x),$$

откуда (т.к. $B'(x) = B(x)$, $L'(x) = G'(x)$ и $K(x) \geq 0$)

$$L'(x)(B(x) - A(x) + 1) \geq 1.$$

Умножая это неравенство на ряд $L(x) \geq 0$, получаем (ввиду зад. 33)

$$L'(x)(B(x) - A(x) + 1) \cdot \frac{1}{1 - A(x) + B(x)} \geq L(x),$$

т. е.

$$L'(x) \geq L(x).$$

36. Пусть алфавит $A = \{a, b\}$. Обозначим второе слово через v. Очевидно, чтобы слова v и w не были свободными, необходимо, чтобы их начальные и конечные буквы были различные. Если при этом w начинается на одну букву, а v — на другую, то последняя буква слова v совпадает с первой буквой слова w, и эти слова зацепляются, так что множество B не будет свободным. Итак, осталось рассмотреть случай, когда оба слова v и w начинаются на одну букву (скажем, a), а заканчиваются на другую (b).

а) Имеем $w = ab$ и $v = a...b$. Очевидно, если первая буква b в слове v находится на k-м месте, то его подслово из букв на $(k-1)$-м и k-м местах равно w, так что множество B не свободно.

b) Ответ: нет. Пусть $w = aab$ (случай $w = abb$ аналогичен, с точностью до левой-правой симметрии и замены букв). Поскольку слово w не является подсловом в v, то в слове v за парой букв aa всегда следует ещё одна a. Так как слова v и w не зацепляются, то v не может начинаться на ab, т. е. оно начинается на aa. Следовательно, третья буква в v снова a, затем четвёртая и т. д., откуда $v = aa \dots a$ — противоречие.

37. Достаточно показать, что существует свободное множество B из $g = \left[n^2/4\right]$ двухбуквенных слов. Пусть $k = [n/2]$, т. е. $n = 2k$ или $n = 2k + 1$. Положим $B = \{x_i x_j | 1 \leq i \leq k, k + 1 \leq j \leq n\}$. Очевидно, оно свободно. Тогда в случае четного $n = 2k$ множество B насчитывает $k^2 = n^2/4$ элементов, а в случае нечётного $n = 2k+1$ оно содержит $k(k+1) = (n-1)(n+1)/4 = n^2/4 - 1/4 = \left[n^2/4\right]$ элементов, что и требовалось.

38. Достаточно показать, что существует свободное множество B из $m \leq k^d(d-1)^{d-1}$ слов длины d. Разобьём основной алфавит $A = \{x_1, \dots, x_n\}$ на два подмножества $P = \{x_1, \dots, x_k\}$ и $Q = \{x_{k+1}, \dots, x_n\}$. Тогда легко видеть, что множество $B = \{pq | p \in P, q \in F_Q$ — слово длины $d - 1\}$ — искомое.

39.
а) Согласно задаче 23, $\frac{1}{1-nx+B(x)} = L(x) \geq 1$.
б) Ответ: неверно. Например, в случае двухбуквенного алфавита A такого множества B не существует при $p(x) = x^3 + x^{10}$ (это доказано в задаче 36 b). Осталось убедиться, что ряд

$$f(x) = \frac{1}{1 - 2x + x^3 + x^{10}}$$

имеет неотрицательные коэффициенты (поскольку свободный член этого ряда единичный, это условие равносильно требуемому неравенству $f(x) \geq 1$). Оказывается, для этих коэффициентов справедливо даже более сильное неравенство $a_n \geq (3/2)^n$. Доказательство последнего неравенства можно провести по индукции, воспользовавшись рекуррентным соотношением $a_{n+10} = 2a_{n+9} - a_{n+7} - a_n$, которое следует из равенства $(1 - 2x + x^3 + x^{10})f(x) = 1$.

Другой подобный пример — случай $p(x) = 4x^6$, снова над алфавитом из двух букв.

40. Решение содержится в теореме 5.1 (эквивалентность A$\Longleftrightarrow$B) и предложении 5.6 в статье: David Anick, *Generic algebras and CW–complexes*, Proceedings of 1983 Conference on algebra, topology and K–theory in honor of John Moore. Princeton University, 1988, p. 247–331.

41. Вопрос остаётся открытым.

7 Слова и цепи

42. Слово “of” не участвует в построении цепей длины больше 1, потому что данные слова не начинаются на букву f и не заканчиваются на букву o. Буквы s нет в слове tournament, поэтому цепь может содержать слово towns только в конце. Буква t в слове tournament стоит лишь на первом и последних местах, поэтому это слово может зацепляться с собой только по одной букве. Исходя из сказанного, мы находим все цепи: tournament, of, towns; tournamentournament, tournamentowns;... Цепей длины больше 1 будет ровно две.

43. Рассмотрим цепь c длины n. От его правого конца можно строить антицепь по направлению к левому концу единственным образом. Будем присоединять звенья слева по очереди. Докажем по индукции, что начало i-го звена антицепи лежит между началом i-го справа звена цепи и концом $i{+}1$-го справа звена цепи. Для $i = 1$ утверждение верно, потому что в этом случае соответствующие звенья цепи и антицепи совпадают. Проверим базу также для $i = 2$. Второе звено антицепи лежит правее второго справа звена цепи, потому что между первым и вторым звеном нет запретных слов. Начало второго звена антицепи лежит левее конца третьего справа звена цепи, потому что правее третьего справа конца звена цепи нет запретных слов по определению цепи. Таким образом для $i = 2$ утверждение верно. Докажем шаг. По предположению индукции i-е справа звено цепи пересекает $i{-}1$-е звено антицепи, но $(i{+}1)$-е звено антицепи не пересекает $(i{-}1)$-го звена антицепи, поэтому $(i{+}1)$-е звено антицепи левее i-го справа звена цепи. Между концом $(i{+}2)$-го и началом i-го справа звена цепи не начинается никакое запретное слово, поэтому начало $(i{+}1)$-го звена антицепи левее конца $(i{+}2)$-го звена цепи. Так как $(i{-}1)$-е звено антицепи не левее $i{-}1$-го звена цепи, то $(i{+}1)$-е звено цепи не пересекает $(i{-}1)$-е звено антицепи, но пересекает i-е звено, поэтому $(i{+}1)$-е звено антицепи не левее $(i{+}1)$-го звена цепи.

44. Пусть цепь c' – подслово цепи c. Докажем по индукции, что i-е звено цепи c не правее i-го звена цепи c'. База $i = 1$ очевидна. Если первые звенья цепей совпадают, и одна является подсловом другой, то все остальные звенья также совпадают, и, в частности, если такие цепи имеют одинаковую длину, то они совпадают. Поэтому цепь c' начинается не с начала, а поэтому первое звено цепи c' правее или совпадает со вторым звеном цепи c. Если бы второе звено цепи c было правее первого звена цепи c', то тогда бы первое звено цепи c' можно было бы выбрать в качестве второго звена цепи c. Значит, второе звено цепи c левее второго звена цепи c', и база индукции верна для $i = 2$. Перейдём к шагу индукции. Если $(i{+}1)$-е звено цепи c' не пересекается с i-ым звеном цепи c, то $(i{+}1)$-е звено цепи c' правее $(i{+}1)$-го звена цепи c. Также добавленное звено цепи c' не пересекается с $(i{-}1)$-м звеном цепи c', а значит, по предположению индукции с $(i{-}1)$-м звеном цепи c. Поэтому, если $i{+}1$-е звено цепи c' пересекает i-е звено цепи c, то $i{+}1$-е звено цепи c всё равно не правее $(i{+}1)$-го звена цепи c', потому что иначе то звено можно было бы выбрать вместо этого. Шаг доказан. Значит, последние звенья цепей совпадают, но по прошлой задаче данная пара цепей является также парой антицепей одинаковой длины с совпадающими первыми звеньями, откуда они совпадают.

45. Пусть слово представлено в виде gc, где g – допустимое слово, а c – цепь длины не менее двух. По задаче 43 цепь c также является антицепью. Если при убирании первого звена цепи c появляется запретное слово, не зацепленное с укороченной цепью c, то антицепь c можно продолжить влево и получить новое представление $g'c'$, где длина цепи увеличилась. Если такого запретного слова не появилось, то антицепь c можно укоротить на самое левое звено и получить новое представление $g'c'$, в котором длина цепи уменьшилась. Если есть ещё одно представление $g''c''$, то заметим, что дуги антицепей c, c', c'' совпадают, а значит, длина каких-то двух цепей отличается минимум на два, но тогда допустимое слово перед короткой антицепью содержит самое левое звено более длинной антицепи, противоречие.

46. Для решения этой задачи мы применим задачу 21 для точной последовательности

$$\ldots \Longrightarrow C_{n+1} \cdot G \Longrightarrow C_n \cdot G \Longrightarrow C_{n-1} \cdot G \Longrightarrow \ldots C_1 \cdot G \Longrightarrow A \cdot G \Longrightarrow \bar{G}$$

Построим эту последовательность. Возьмём слово cg из $C_n \cdot G$. Хвост цепи c припишем в начало слова g и получим $c'g'$. Если g' допустимое слово, то слово $c'g'$ принадлежит $C_{n-1} \cdot G$. Если слово g' содержит запретное слово, то цепь c можно продолжить вправо до цепи c'', остаток обозначим g''. Тогда слово $c''g''$ принадлежит $C_{n+1} \cdot G$. Цепь можно единственным образом продолжать внутри слова, поэтому построенные отображения взаимно-однозначны на частях языков $C_n \cdot G$.

Стрелки $C_1 \cdot G \Longrightarrow A \cdot G \Longrightarrow \bar{G} \Longrightarrow \emptyset$ мы берём из задачи 27. По задаче 21

$$1 - L(x)(1 - Nx + C_1(x) - C_2(x) + C_3(x) - \ldots) = 0.$$

Откуда получаем требуемое.

47. По задаче 42 $C_1(x) = x^2 + x^5 + x^{10}, C_n(x) = x^{9(n-1)}(x^5 + x^{10})$. По формуле из задачи 46

$$L(x) = \frac{1}{1 - 26x + x^2 + (x^5 + x^{10})(1 - x^9 + x^{18} - \dots)} = \frac{1}{1 - 26x + x^2 + \frac{x^5+x^{10}}{1+x^9}} =$$
$$= \frac{1 + x^9}{1 - 26x + x^2 + x^5 + x^9 - 25x^{10} + x^{11}}$$

48. Рассмотрим четыре варианта запретного слова из четырёх букв с точностью до замены букв.

1) Запретное слово имеет вид aaaa. В этом случае $C_{2n} = x^{4n+1}, C_{2n-1} = x^{4n}$. По формуле задачи 46

$$L(x) = \frac{1}{1 - 256x + x^4 - x^5 + \dots} = \frac{1}{1 - 256x + \frac{x^4-x^5}{1-x^4}} = \frac{1 - x^4}{1 - 256x + 255x^5} =$$
$$= (1 - x^4)(1 + (256x - 255x^5) + (256x - 255x^5)^2 + \dots).$$

Коэффициент при x^7 равен $256^7 - 3 \cdot 256^2 \cdot 255 - 256^3 = 256^7 - 4 \cdot 256^3 + 3 \cdot 256^2$.

2) Запретное слово имеет вид abca и хотя бы две различные буквы. В этом случае $C_n = x^{3n+1}$.

$$L(x) = \frac{1}{1 - 256x + x^4 - x^7 + \dots} = 1 + (256x - x^4 + x^7 - \dots) + (256x - x^4 + x^7 - \dots)^2 + \dots$$

Коэффициент при x^7 равен $256^7 - 4 \cdot 256^3 + 1$

3) Запретное слово имеет вид abab. В этом случае $C_n = x^{2(n+1)}$.

$$L(x) = \frac{1}{1 - 256x + x^4 - x^6 + x^8(\dots)}$$

Коэффициент при x^7 равен $256^7 - 4 \cdot 256^3 + 2 \cdot 256$

4) Запретное слово свободно. А случай свободного запретного слова разобран в задаче 18.

49. Из задачи 46 следует, что

$$\frac{1}{1 - Nx} - L(x) = L(x) \cdot C_1(x) \cdot \frac{1}{1 - Nx} - L(x) \cdot C_2(x) \cdot \frac{1}{1 - Nx} + L(x) \cdot C_3(x) \cdot \frac{1}{1 - Nx} - \dots$$

(Бесконечная сумма имеет смысл, поскольку степени начальных членов слагаемых растут.) В левой части равенства стоит ряд размеров множества недопустимых слов. В правой части равенства стоит знакопеременная сумма рядов размеров множеств $L \cdot C_n \cdot F_A$. Определено отображение для каждого из этих языков в множество недопустимых слов, и наоборот, каждое недопустимое слово можно представить в виде gc_1u, где g – допустимое, c_1 – запретное слово. Но недопустимое слово может быть представлено в виде gc_nu, где g – допустимое, а c_n – цепь длины n, несколькими способами. Пусть для слова w число таких способов равно w_n. Тогда сумма чисел $(w_1 - w_2 + w_3 - w_4 + \dots)$ по всем словам длины k равна коэффициенту при x^k в правой части равенства выше, а значит, и в левой, что равно количество недопустимых слов длины k.

Заметим, что из представления слова w в виде gc_nu можно получить другое представление с длиной цепи на единицу меньше, отбросив хвост цепи c_n в остаток u. Поэтому два таких варианта сократятся в сумме $(w_1 - w_2 + w_3 - w_4 + \dots)$. Другими словами рассматриваемая сумма равна числу представлений слова w в виде gcu, где подслово c – это максимальная подцепь слова w, причём цепь c имеет нечётную длину.

Рассмотрим представление слова w в виде gcu, где c – максимальная цепь с самым правым последним звеном. По задаче 45 либо в слове gc есть максимальная цепь длины 1, либо оно представимо в виде $g'c'$, где g' – допустимо, а c' – цепь, которая либо длиннее, либо короче цепи c на одно звено. Заметим, что c' – также максимальная подцепь слова w. Одна из цепей c и c' имеет нечётную длину. Тем самым мы показали, что сумма $(w_1 - w_2 + w_3 - w_4 + \dots)$ не меньше единицы. Но сумма таких величин по всем недопустимым словам длины k равна их количеству. Значит, каждая величина равна единице, и в каждом недопустимом слове максимальных цепей нечётной длины ровно одна.

50. Применим формулу задачи 46.

$$L'(x) = \frac{1}{1 - (N + 1)x + C_1(x) - C_2(x)) + \dots} = \frac{1}{\frac{1}{L(x)} - x}$$

51. Применим формулу задачи 46.

$$W(x) = \frac{1}{1 - (N + N')x + (C_1(x) + C'_1(x)) - (C_2(x) + C'_2(x)) + \dots} = \frac{1}{\frac{1}{L(x)} + \frac{1}{L'(x)} - 1}$$

52. Допустимые слова языка M – это цепи языка L. Цепи длины n этих языков состоят из n+1-й буквы, поэтому $C_n(-x) = (-1)^{n+1}C_n(x)$. Имеем $M(-x) = 1 - Nx + C_1(x) - C_2(x) + \dots$, то есть $L(x)M(-x) = 1$.

8 Дополнительные задачи

53. Существование свободного множества при условии $m \leq k^d(d-1)^{d-1}$ доказано в задачах 37 и 38. Осталось доказать, что при $m > k^d(d-1)^{d-1}$ искомого свободного множества не существует.

а) Если задано $m > n^2/4$ слов из двух букв, то они могут составлять свободное множество S лишь при условии, что первая буква никакого слова не совпадает со второй буквой никакого другого слова, т.е. есть два непересекающихся множества букв, P и Q, элементы которых могут служить, соответственно, только начальными или только вторыми буквами слова из S. Обозначим $r = |P| + |Q| \leq n$ и $s = |P| \cdot |Q| \geq |S| = m > n^2/4$. По теореме Виета, натуральные числа $|P|$ и $|Q|$ являются корнями квадратного уравнения $x^2 - rx + s = 0$, дискриминант которого $D = r^2 - 4s$ отрицателен при выполнении указанных ограничений на r и s — противоречие.

б) Пусть B — свободное множество m слов длины 3. Поскольку в свободном множестве никакая первая буква слова не может быть также и последней буквой какого-либо слова, в алфавите A есть два непересекающихся подмножества X и Y, элементы которых встречаются, соответственно, только в начале и только в конце слов из B. Если есть буквы, которые не встречаются ни в конце, ни в начале слова, присоединим их произвольным образом к одному из множеств X и Y. Пусть, для определённости, число s элементов множества $X = \{x_1, \dots, x_s\}$ не превосходит количество элементов t множества $Y = \{y_1, \dots, y_t\}$. Каждый элемент множества B имеет вид $x_{i_1}x_{i_2}y_{j_1}$ или $x_{i_1}y_{j_1}y_{j_2}$, причём никакое финальное подслово вида xy элемента первого типа не может быть началом элемента второго типа. Проведём преобразование множества B, заменив для каждого финального подслова xy все слова первого типа, оканчивающиеся на него, на слова второго типа, по правилу $x_ixy \mapsto xyy_i$, где $1 \leq i \leq s \leq t$. Легко видеть, что при таком преобразовании никакие элементы не переходят в другие элементы множества B и никакие различные элементы не переходят в один и тот же, причём получившееся множество B' останется свободным. При этом в B' все элементы имеют вид $x_{i_1}y_{j_1}y_{j_2}$. Следовательно, количество элементов множества B' (равное по-прежнему числу m) не превосходит st^2, откуда

$$m \leq st^2 \leq (n-t)t^2 \leq \left(n - \frac{2n}{3}\right)\left(\frac{2n}{3}\right)^2 = \frac{4n^3}{27} = 4k^3,$$

что и требовалось.

в) Для доказательства потребуется следующая аналитическая

Лемма. Пусть $R(x) = 1 + a_1x + a_2x^2 + \dots$ — ряд с натуральными коэффициентами такой, что $R(x) = 1/p(x)$ для некоторого многочлена $p(x)$ с единичным свободным членом. Обозначим $R_n(x) = 1 + a_1x + a_2x^2 + \dots + a_nx^n$. Если для всех $x \in [0, x_0]$, где $x_0 > 0$, выполняется условие $p(x) \geq m > 0$, то для всех $n > 0$ справедливы неравенства $R_n(x_0) \leq 1/m$.

Не доказывая лемму, перейдём к решению задачи. Обозначим $s = mk^{-d} - (d-1)^{(d-1)}$; требуется доказать, что при $s > 0$ свободного множества не существует. Предположим, что это не так. Тогда, согласно задаче 39 а), ряд $1/p(x)$, где $p(x) = 1 - dkx + mx^d$, имеет натуральные коэффициенты (нулевых коэффициентов быть не может, поскольку этот ряд бесконечен согласно задаче 33). Отметим, что при $s > 0$ многочлен $p(x)$ положителен на отрезке $[0, 1]$ (доказательство: минимум этого многочлена на отрезке $[0, 1]$ достигается либо в концах отрезка, где $p(x)$ положителен, либо в такой точке x_0, в которой $p'(x_0) = 0$, т. е. при $x_0 = \frac{1}{k(d-1)}$; при этом $p(x_0) = sx_0^d > 0$). Это означает, что существует такое число $m > 0$, что $p(x) \geq m$ при $x \in [0, 1]$. Согласно лемме, это означает, что для всех n количество слов длины не выше n, равное $L_n(1)$, ограничено константой $1/m$.

54. а) По определению, запретными словами языка $L^!$ являются все двухбуквенные слова, не запретные в L. Следовательно, запретными словами языка $(L^!)^!$ будут все двухбуквенные слова, не запретные в $L^!$, т. е. в точности запретные слова языка L. Таким образом, и алфавиты, и наборы запретных слов языков L и $(L^!)^!$ совпадают, а потому и сами языки равны.

б) Поскольку множество запретных слов языка $M = (L_1 + L_2)^!$ есть объединение множеств запретных слов языков $L_1^!$ и $L_2^!$, а алфавит языка M представляет собой объединение их (непересекающихся) алфавитов, то язык M есть свободное произведение (см. определение в зад. 51) языков $L_1^!$ и $L_2^!$.

в) Запретными словами языка $(L_1 \cdot L_2)^!$ будут разрешённые двухбуквенные слова языков L_1 и L_2, а также слова вида aB, где a — буква из алфавита языка L_1, а B — буква из алфавита языка L_2. Это означает, что

$$(L_1 \cdot L_2)^! = L_2^! \cdot L_1^!.$$

55. Пусть w — слово длины nk (где $k \geq 1$) над алфавитом языка L, $w^{(n)}$ — соответствующее ему слово языка $L^{(n)}$. Разобьём w на подслова $w = w_1 \dots w_k$, каждое из которых соответствует букве языка $L^{(n)}$. Легко видеть, что слово w содержит какое-то запретное подслово u (состоящее, по определению, из не более чем d букв) в том и только том случае, когда в каком-то подслове $w' = w_p \dots w_{p+m-1}$ каждое из подслов w_i

либо содержится в подслове u, либо пересекается с ним, так что количество n-буквенных фрагментов m в подслове w' удовлетворяет неравенству $m \le s$, где

$$s = 2 + \left[\frac{d-2}{n}\right].$$

Таким образом, любое недопустимое слово $w^{(n)}$ языка $L^{(n)}$ содержит недопустимое подслово из не более чем s букв, т. е. язык $L^{(n)}$ задаётся конечным множеством запретных слов, причём длины запретных слов этого языка не превосходят s. Это доказывает утверждение а).
б) Ответ: не всегда.

Докажем, что при $d \ge 3$ и $n \ge 2$ длины запретных слов языка $L^{(n)}$ всегда меньше d; в частности, этот язык не может быть d-определённым, т. е. ответ на вопрос из п. б) отрицательный. Достаточно доказать неравенство $s < d$, или

$$2 + \frac{d-2}{n} < d.$$

Последнее неравенство равносильно неравенству $(d-2)(1-1/n) > 0$, которое, очевидно, верно при заданных ограничениях на d и n.
в) Ответ: $n = d-1$. По доказанному выше, язык $L^{(n)}$ является квадратичным или свободным (т. е. длины запретных слов не превосходят 2) при условии $s \le 2$, которое равносильно неравенству $2 + \frac{d-2}{n} < 3$, или $n > d-2$, т. е. $n \ge d-1$. Если же $n \le d-2$, то существуют такие d-определённые языки L, для которых язык $L^{(n)}$ имеет запретные слова из более чем трёх букв: примером служит язык L с трёхбуквенным алфавитом $\{a, b, c\}$ и единственным запретным словом abc^{d-2}.

56. См. решение задачи 58.

57. Ответ: да. Например, пусть A — алфавит из $n \ge 2$ букв. Рассмотрим язык $L = F_A \cdot F_A^!$. Поскольку язык F_A имеет экспоненциальный рост (для него в задаче 55 с) можно выбрать $c_1 = n+1$ и $c_2 = n$), причём $2F_A(x) \ge L(x) \ge F_A(x)$, то язык L также имеет экспоненциальный рост. Согласно задаче 53 в), имеем $L^! = (F_A^!)^! \cdot F_A^! = L$, так что оба языка L и $L^!$ имеют экспоненциальный рост.

58. Докажем сначала следующее утверждение (при решении задачи 56 без него можно обойтись).

Лемма. Пусть $a = \{a_0, a_1, a_2, \dots\}$ — последовательность натуральных чисел, в которой $a_0 = 1$ и для некоторого натурального N имеем $a_1 \ge 2, \dots, a_N \ge 2$. Тогда последовательность a имеет полиномиальный (соответственно, экспоненциальный) рост в том и только том случае, когда соответствующие неравенства из утверждений b) и c) выполняются для всех a_k при $k \ge N$.

Доказательство леммы. Пусть $M = \max\limits_{i \le N}\{a_i\}$. Очевидно, если для некоторых многочленов p, q степени d выполняются неравенства $p(k) \ge a_k \ge q(k)$ при $k \ge N$, то выполняются также и неравенства $p(k) + M \ge a_k \ge q(k) - M$ при всех k, что доказывает лемму в случае полиномиального роста. Аналогично, если $c_1^k \ge a_k \ge c_2^k$ при $k \ge N$, то $(M + c_1)^k \ge a_k \ge g^k$ при всех k, что полностью доказывает лемму.

Перейдём к решению задачи. Очевидно, все допустимые слова длины $\ge d-1$ получаются, если, начиная со слова в вершине графа, приписывать к нему справа буквы, которые прочитываются при прохождении какого-либо маршрута, начинающегося в этой вершине, причём разные слова соответствуют разным маршрутам. Очевидно, язык конечен тогда и только тогда, когда никакой маршрут не возвращается в начальную вершину, т.е. в графе нет циклов (что доказывает утверждение а)). Осталось рассмотреть случай, когда язык бесконечен и в графе есть цикл. В этом случае количество a_j слов длины $j \ge d$ равно числу маршрутов длины $j - d + 1$.

Предположим, что есть два пересекающихся цикла; пусть их длины суть d_1 и d_2, и v — общая вершина, исходящие из которой ребра различны для обоих циклов (скажем, отвечающие буквам x и y). Слова, которые прочитываются на ребрах k-звенных маршрутов, начинающихся в v и проходящим по каждому из циклов, различны, поэтому $a_k \ge 2$ при всех $k \ge 0$. Кроме того, для любого $j = (d-1) + q(d_1 + d_2) + r$, где $r < d_1 + d_2$ — остаток от деления числа $j - d + 1$ на $d_1 + d_2$, существует по крайней мере 2^q различных маршрутов длины $j - d + 1$ (на каждом из q шагов проходим оба цикла в произвольном порядке, а затем делаем r шагов в произвольном цикле), так что при $j \ge 2d$ имеем $a_j \ge 2^q = 2^{\left[\frac{j-d+1}{d_1+d_2}\right]} \ge c^j$, где $c = 2^{1/2(d_1+d_2)}$. Поскольку всегда $a_j \le n^j$, из доказанной леммы (при $N = 2g$) следует, что рост экспоненциальный.

Осталось разобрать случай, когда в графе Γ_L циклы есть, но они не пересекаются между собой. Достаточно проверить условия полиномиальности для количеств маршрутов $b_k = a_{k+d-1}$ длины k в графе Γ_L (т.к. если соответствующие неравенства выполняются для чисел b_k, то они выполняются и для a_k при $k \ge d-1$ при замене многочленов $p(x)$ и $q(x)$ на многочлены той же степени $p_1(x) = p(x+d-1)$ и $q_1(x) = q(x+d-1)$). Мы докажем, что каждый член последовательности b_k равен значению некоторого многочлена $b(k)$ с положительным старшим коэффициентом (будем называть такие последовательности полиномиальными).

Рассмотрим другой граф Γ'_L, вершинами которого служат циклы графа Γ_L и вершины графа Γ_L, не входящие в циклы (назовём их обособленными), а ребра соответствуют ребрам, соединяющим соответствующие компоненты (вершины или циклы) графа Γ_L. Очевидно, в графе Γ'_L циклов нет, т.е. множество маршрутов в нём конечно. Пусть Q^v — множество маршрутов в графе Γ'_L, исходящих из данной вершины v, и пусть q_k^v — количество маршрутов длины k, которое им соответствует в графе Γ_L. Поскольку $b_k = \sum_v q_k^v$, достаточно доказать, что последовательность $\{q_k^v\}$ для каждой вершины v полиномиальна. Воспользуемся индукцией по длине $D = D(v)$ максимального маршрута, исходящего из v. Если $D = 0$, то либо $q_k^v = 0$ при $k > 0$ (если v — обособленная вершина), либо $q_k^v = 1$ для всех k (если v — цикл), т. е. соответствующая последовательность всегда полиномиальна. Пусть теперь v — какая-то начальная вершина графа Γ'_L, из которой исходят r стрелок $a_1, \ldots, a_r$ к вершинам $v_1, \ldots, v_r$ (возможно, повторяющимся). По индукции, считаем $q_k^{v_i} = b_i(k)$ — многочлен с положительным старшим коэффициентом. Если v — обособленная вершина, то $q_k^v = \sum_{i=1}^r q_{k-1}^{v_i}$, т.е. эта последовательность полиномиальна как сумма полиномиальных последовательностей. Если же v — цикл, то перед переходом к вершинам по ребрам $a_1, \ldots, a_r$ возможно слово любой длины в цикле, поэтому $q_k^v = \sum_{i=1}^r \left(\sum_{j=1}^k q_{k-j}^{v_i}\right) = \sum_{i=1}^r \left(\sum_{j=1}^k b_i(k-j)\right)$ — сумма многочленов с положительными старшими коэффициентами. Утверждение доказано.

Примечание. Аналогичным образом можно определить тип роста любого регулярного множества. Для этого используется отвечающий этому множеству конечный автомат.

59. Обозначим множество допустимых слов через M. Пусть язык является d-определённым. Докажем, что каждое слово M-эквивалентно слову не более чем из d букв.

Действительно, если слово v не является допустимым, то какое слово к нему не припиши, результат не будет допустимым. Поэтому все недопустимые слова эквивалентны. В частности, любое из них эквивалентно какому-нибудь запретному слову, то есть слову длины не большей d.

Пусть слово u допустимо и имеет длину, большую d. Обозначим через v подслово слова u, состоящее из его последних d букв. Пусть w — произвольное слово. Если слово uw содержит запретное подслово, то это подслово содержится в vw, так как длина запретного подслова не больше d. Поэтому слова u и v эквивалентны.

Пусть в алфавите k букв. Тогда число слов длины не большей d не превосходит $(k+1)^d$. Положим $n = (k+1)^d + 1$. В любом наборе из n слов найдутся два, M-эквивалентные одному и тому же слову длины не большей d и, тем самым, эквивалентные друг другу. Поэтому множество допустимых слов регулярно.

60. а) Пусть S — максимальное множество слов, из которых никакие два не M-эквивалентны. Тогда любое другое слово эквивалентно какому-то слову из S. Построим конечный автомат. Возьмём S в качестве множества вершин графа. Для всех $s \in S$, $a \in A$, проведём из вершины s стрелку, помеченную a, в вершину, M-эквивалентную sa. В полученном графе назовём начальной вершиной ту, которая M-эквивалентна пустому слову, а принимающими вершинами — все слова из S, которые принадлежат M. Легко видеть, что построенный конечный автомат принимает те и только те слова, которые принадлежат M.
б) Любому слову отвечает путь по стрелкам конечного автомата. Ясно, что если для двух слов такие пути заканчиваются в одной вершине, то слова M-эквивалентны, где M — множество принимаемых автоматом слов. Поэтому в качестве n из определения регулярного множества можно взять число, на единицу большее числа вершин в автомате.

61. Рассмотрим конечный автомат (Γ, v_0, W), принимающий множество M. Для каждой вершины v этого конечного автомата обозначим через T_v множество слов, для которых соответствующие пути по графу Γ заканчиваются в v.

Далее, для каждой вершины v и каждой буквы a обозначим через $U(v,a)$ множество таких вершин u графа Γ, что из u в v идёт стрелка, помеченная a. Тогда выполнены следующие равенства:

$$T_{v_0}(x) = 1 + \sum_{a\in A} \sum_{u\in U(v_0,a)} xT_u(x) \tag{1}$$

и

$$T_v(x) = \sum_{a\in A} \sum_{u\in U(v,a)} xT_u(x) \tag{2}$$

для $v \neq v_0$.

Перенумеруем вершины графа Γ, начиная с v_0: $V = \{v_0, v_1, v_2 \ldots, v_k\}$. Заметим, что каждое из равенств (1), (2) можно воспринимать как уравнение вида

$$(1 + xP_i(x))T_{v_i}(x) = \sum_{j\neq i} xQ_{ij}(x)T_{v_j}(x) + R_i(x), \tag{3}$$

где $P_i(x), Q_{ij}(x), R_j(x)$ — некоторые известные многочлены, относительно неизвестных рядов $T_{v_0}(x), \ldots, T_{v_k}(x)$.

Попробуем решить уравнения (3). Выразим из последнего уравнения $T_{v_k}(x)$ через остальные неизвестные ряды,

$$T_{v_k}(x) = \sum_{j\neq k} x\frac{Q_{kj}(x)}{(1 + xP_k(x))}T_{v_j}(x) + \frac{R_k(x)}{(1 + xP_k(x))},$$

подставим это выражение вместо $T_{v_k}(x)$ в остальные уравнения, и домножим их все на $(1+xP_k(x))$. Мы получим уравнения того же вида, но число их (как и число неизвестных) станет на единицу меньше. Проделав то же самое для $T_{v_{k-1}}(x)$, $T_{v_{k-2}}(x)$, и т. д., мы получим в конце концов выражение для $T_{v_0}(x)$ в виде отношения двух многочленов. Подставив его в выражение для $T_{v_1}(x)$, получаем, что и этот ряд есть отношение двух многочленов. Продолжая этот процесс, получаем выражения того же вида для всех $T_{v_i}(x)$. Осталось заметить, что $M(x) = \sum_{v \in W} T_v(x)$.

62. Для каждого слова v обозначим через v^{opp} слово, состоящее из тех же букв в противоположном порядке. Для каждого множества слов M положим $M^{opp} = \{v^{opp} \mid v \in M\}$. Ясно, что $M^{opp}(x) = M(x)$ для любого M. Если L — язык с множеством запретных слов B, то через L^{opp} мы обозначаем язык с множеством запретных слов B^{opp}.

Вернёмся к задаче. Очевидно, что M_w^{opp} — это множество допустимых слов языка L^{opp}, начинающихся с подслова, равного w^{opp}. Это множество регулярно (доказательство аналогично решению задачи 59). Поэтому ряд $M_w(x) = M_w^{opp}(x)$ равен отношению двух многочленов.

На самом деле, множество M_w тоже регулярно, но доказательство этого заняло бы больше места.

How to count words?

Dmitri Piontkovski, Maxim Prasolov, Grigory Rybnikov

1 Main problem

Problem 1. *The language of Winnie-Pooh tribe has 100 words. All possible combinations of these words, in any order, are used as sentences of the language. The are two magic spells, "Earth stands on Great Crocodile" and "Every evening Crocodile swallows Sun", that cause tornado. That is why it is not allowed to pronounce sentences that contain the above sequences of words*[1]. *How many sentences of 20 words in this language are allowed?*

Problem 2. *A computer uses* 256 *commands. There is a sequence of four commands that breaks the computer. The programmers made all possible programs of* 7 *commands. Find the percentage of the programs that do not break the computer.*[2]

Problem 3 (Main Problem)**.** *The alphabet of a language L consists of N letters. Several words $v_1, \dots, v_k$ are called* forbidden *and are not used in the language. A word (that is, a finite sequence of letters) is called* admissible *if no part of it is a forbidden word. Find the number of admissible words of n letters in L.*

Problem 4. *Show that the Problems 1 and 2 are special cases of Problem 3.*

2 How to write down the answer?

Choose an alphabet A of N letters (for example, if $A = (a, b, c, \dots, z)$, then $N = 26$). By a *word* we will mean an arbitrary finite sequence of letters of the alphabet A. A part of a word is called its *subword.*

We assume that every language L has exactly one word of zero length, that is, an *empty* word.

We assume that distinct forbidden words are not subwords of each other. We also assume that each forbidden word has at least two letters, that is, the empty word and one-letter words are admissible. Recall that the set of forbidden words is finite.

Problem 5. *The* free language F_A *over the alphabet A is the language with no forbidden words. Prove that the number of the words of n letters in this language is equal to N^n.*

Problem 6. *Let B be the language whose forbidden words are all two-letter words with different letters. Prove that the number of admissible words of n letters in the language B is equal to N for any positive integer n.*

Let M be an arbitrary set of words. Let us denote by m_n the number of n-letter words in this set. The infinite sum

$$M(x) = m_0 + m_1 x + m_2 x^2 + m_3 x^3 + \dots$$

is called the *dimension series* of the set M. The infinite sums of such type (with arbitrary numbers as coefficients m_n) will be briefly referred to as *series* (their complete name, which will *not* be used here, is *formal power series*).

For any language L, by its dimension series $L(x)$ we will mean the dimension series of the set of admissible words. For example, for the free language F_A its dimension series is the geometric series $F_A(x) = 1+Nx+N^2x^2+N^3x^3+\dots$, and for the language B above we have $B(x) = 1 + Nx + Nx^2 + Nx^3 + \dots$

Problem 7. *Write down the dimension series for the language over the alphabet $\{a, b\}$ with forbidden words aa and bb.*

3 The arithmetics of languages

If a set M contains finitely many words, then its dimension series is a polynomial in the variable x. For infinite sets, their dimension series are infinite as well, but they allow various arithmetic operations similar to the operations over the polynomials, that is, addition, subtraction, multiplication by each other and by numbers, and even sometimes division.

In the definitions and problems of this section, $S = s_0 + s_1 x + s_2 x^2 + \dots$ and $R = r_0 + r_1 x + r_2 x^2 + \dots$ are two series, and L_1 and L_2 are two languages over alphabets A_1 and A_2 without common letters. We will assume that the alphabet A_1 consists of upper-case letters while the alphabet A_2 consists of lower-case ones. Let the alphabet A be the union of the alphabets A_1 and A_2, that is, A contains both upper-case and lower-case letters.

[1] Even if the words are in other forms

[2] Similar story happened in 1990s with the first version of *Pentium* microprocessor.

Definition 1. a) The *sum* of two series R and S is the series

$$R + S = (s_0 + r_0) + (s_1 + r_1)x + (s_2 + r_2)x^2 + \dots$$

b) The *sum* of two languages L_1 and L_2 is the language $L_1 + L_2$ over A whose set of admissible words is the union of the sets of admissible words of the languages L_1 and L_2.

Problem 8. *Define the language $L_1 + L_2$ by a finite set of forbidden words.*

Problem 9. *Prove that if $L = L_1 + L_2$, then*

$$L(x) = L_1(x) + L_2(x) - 1.$$

The product of two series is defined by the same way as the product of two polynomials.

Definition 2. *The product of a series R by a monomial ax^n* is the series

$$R \cdot ax^n = ar_0x^n + ar_1x^{n+1} + ar_2x^2x^{n+2} + \dots$$

The product of two series R and S is the sum

$$R \cdot S = R \cdot s_0 + R \cdot s_1x + R \cdot s_2x^2 + \dots$$

Note that this infinite sum of series is well-defined because the coefficient of every power of x is a finite sum of numbers.

Problem 10. *Prove that*

$$(1 - x) \cdot (1 + x + x^2 + \dots) = 1.$$

Definition 3. *The product of two sets of words M* and N is the set MN of all words of the form mn, where m is a word in M and n is a word in N.

The product of two languages L_1 and L_2 is the language $L_1 \cdot L_2$ over A whose set of admissible words is the product of the sets of admissible words of the languages L_1 and L_2.

Problem 11. *Define the language $L_1 \cdot L_2$ by a finite set of forbidden words.*

Problem 12. *Prove that*

$$L(x) = L_1(x) \cdot L_2(x).$$

The division of series has no version for languages, but it helps to write down their dimension series in a compact form. It is defined by a formula similar to the formula for the sum of an infinite geometric progression.

Definition 4. Suppose that a series S begins with the unit, that is, $s_0 = 1$, and $S = 1+\overline{S}$, where $\overline{S} = s_1x+s_2x^2+\dots$ Then its *inverse* is the series

$$\frac{1}{S} = 1 - \overline{S} + \overline{S}^2 - \overline{S}^3 + \dots$$

The *quotient* of two series R and S is the series

$$\frac{R}{S} = R - R \cdot \overline{S} + R \cdot \overline{S}^2 - R \cdot \overline{S}^3 + \dots$$

In general, the quotient of two dimension series can not be obtained as the dimension series for a language. For example, some of the coefficients of the quotient can be negative.

Problem 13. *a) Prove that*

$$S \cdot \frac{R}{S} = R.$$

b) Prove that if $S \cdot T = R$, where the series S begins with the unit, then $T = \frac{R}{S}$.

The use of the division of two series is that it helps to represent many infinite series by a finite formula, that is, a quotient of two polynomials.

Problem 14. *a) Prove that*

$$F_A(x) = \frac{1}{1 - Nx}.$$

b) Represent the dimension series from Problems 6 and 7 as a quotient of two polynomials.

Problem 15. *Prove that the dimension series of any language can be represented as a quotient of two polynomials.*

Thus, the answer to our Main Problem should be represented as a quotient of two polynomials.

4 Free word

Problem 16. *Let L be a language over the Latin alphabet with only one forbidden word "mouse". Find $L(x)$.*

Definition 5. Let a and b be words such that no one is a subword of each other. A nonempty word c is called an *overlap* of a and b if it is a beginning subword of a and, in the same time, the final subword of b (for example, the word "all" is an overlap of the words "ball" and "allow").

A word w is called *free* if it has no overlaps with itself except for the whole word w (e.g., the word "free" is free but the word "underground" is not).

Problem 17. *Suppose that in a language L over an alphabet A of N letters there is a single forbidden word, which is free and consists of m letters. Prove that*

$$L(x) = \frac{1}{1 - Nx + x^m}.$$

Problem 18. *Solve Problem 2 under the addition assumption that the sequence of commands breaking the computer is a free word.*

5 Transformations of words

Definition 6. Let M and M' be two sets of words. Let us divide the set M in two parts K and L. A function f mapping L to a subset I of M' is called a *transformation* of the set M to the set M' if f preserves lengths of words and is a one-to-one map of L onto I.

In this case, the set K is called the *kernel* of the transformation f and the set I is called the *image* of f.

A transformation will be denoted by an arrow: $M \Longrightarrow M'$.

Definition 7. A sequence of transformations

$$M_1 \Longrightarrow M_2 \Longrightarrow \ldots \Longrightarrow M_n$$

is called *exact* if the kernel of each subsequent transformation coincides with the image of the previous one.

Problem 19. *Let L be a language over an alphabet A, let G be the set of admissible words, and let N be the set of all non-admissible words. Construct an exact sequence of transformations*

$$\emptyset \Longrightarrow N \Longrightarrow F_A \Longrightarrow G \Longrightarrow \emptyset,$$

where F_A is the set of admissible words of the free language, that is, the set of all words over the alphabet A, and $\emptyset$ denotes the empty set.

Problem 20. *10 boys and 10 girls are sitting in a line so that boys' neighbors are girls and vice versa; their teacher is sitting next to them. Each of the children has some bonbons, and the total number of the boys' bonbons is equal to the total number of the girls' ones. The first boy gives all his bonbons to the girl sitting next to him. The girl eats all these bonbons, then she eats the same number of her own bonbons, and then she gives the rest of her bonbons to the next boy. He does the same (eats and gives the rest of bonbons to the next girl), then the next girl does the same, and so on. The last girl gives the rest of her bonbons to the teacher. How many bonbons does the teacher get?*

Problem 21. *Let*

$$\emptyset \Longrightarrow M_1 \Longrightarrow M_2 \Longrightarrow \ldots \Longrightarrow M_n \Longrightarrow \emptyset$$

be an exact sequence of transformations.

) Prove that if each set M_i consists of a finite number m_i of words, then

$$m_1 + m_3 + m_5 + \cdots = m_2 + m_4 + \ldots$$

b) Prove the following formula for the dimension series:

$$M_1(x) + M_3(x) + M_5(x) + \cdots = M_2(x) + M_4(x) + \ldots$$

Definition 8. A set M of words is called *free* if no word in M is a subword of another word in M, all words in M are free, and the words in M have no overlaps with each other.

Problem 22. *Let L be a language over an alphabet A, and let the set B of forbidden words of L be free. Denote the set of all admissible words by G and the set of all nonempty admissible words by $\overline{G}$. Construct an exact sequence of transformations*

$$\emptyset \Longrightarrow B \cdot G \Longrightarrow A \cdot G \Longrightarrow \overline{G} \Longrightarrow \emptyset.$$

Problem 23. *Let L be a language over an alphabet A of N letters, and let the set B of forbidden words of L be free. Prove the formula*

$$L(x) = \frac{1}{1 - Nx + B(x)}.$$

Problem 24. *Prove that the set of magic spells in Problem 1 is free, and solve the problem.*

Problem 25. *Find $L(x)$ provided that the alphabet of the language L is Latin and the forbidden words are* veni, vidi, vici.

Definition 9. Let L be a language. A *simple linkage* is a word $v = str$, where s, t, r are nonempty words such that the words $g = st$ and $f = tr$ are forbidden and there are no other forbidden subwords in v. The end r of the simple linkage (which is produced by cutting off the first forbidden subword g) is called the *tail* of v.

Problem 26. *Prove that the set of forbidden words of a language is free if and only if there are no simple linkages in it.*

Problem 27. *Let L be a language over an alphabet A, let B be its set of forbidden words, and let S be the set of all simple linkages. Denote the set of all admissible words by G and the set of all nonempty admissible words by $\overline{G}$. Construct an exact sequence of transformations*

$$S \cdot G \Longrightarrow B \cdot G \Longrightarrow A \cdot G \Longrightarrow \overline{G} \Longrightarrow \emptyset.$$

Problem 28. *Find the conditions on the set of forbidden words of a language L under which the exact sequence from Problem 27 could be extended to an exact sequence*

$$\emptyset \Longrightarrow S \cdot G \Longrightarrow B \cdot G \Longrightarrow A \cdot G \Longrightarrow \overline{G} \Longrightarrow \emptyset$$

(such languages are called non-tangled*). Give a formula to express the dimension series $L(z)$ of a non-tangled language in terms of the number N of letters and the dimension series of the sets B and S.*

Problem 29. *Find the dimension series of the language over the alphabet $\{a, b, c\}$ with forbidden words abb, bbc, bac.*

Problem 30. *Find the dimension series of the language over the alphabet $A = \{x_1, \ldots, x_n, y_1, \ldots, y_n, z_1, \ldots, z_n\}$, if the forbidden words are the words of the form $x_i y_j$ and $y_j z_k$, where $1 \leq i, j, k \leq n$.*

Problem 31. *Prove that if the set of forbidden words of a non-tangled language consists of a single word, then this set is free.*

6 Free sets revisited

Problem 32. *Construct an infinite free set over an alphabet of two letters.*

Problem 33. *Suppose that the set of forbidden words B of a language L is free and the alphabet has more than one letter. Prove that the set of admissible words of the language is infinite.*

Definition 10. Let $S = s_0 + s_1 x + s_2 x^2 + \ldots$ and $R = r_0 + r_1 x + r_2 x^2 + \ldots$ be two series. If the inequality $s_k \geq r_k$ holds for any k, then we say that the following inequality for the series holds:

$$S \geq R.$$

Problem 34. *Prove that if series P, Q, and R satisfy the inequalities*

$$P \geq Q \text{ and } R \geq 0,$$

then

$$PR \geq QR.$$

Problem 35. *Suppose that for every $d > 0$ the sets B and B' of forbidden words of the languages L and L' over the same alphabet A contain the same number of words of length d, so that $B(z) = B'(z)$. Prove that if the set B is free, then the inequality*

$$L'(z) \geq L(z)$$

holds; in addition, we have $L'(z) = L(z)$ if and only if the set B' is also free.

Problem 36. *Suppose that the alphabet consists of two letters and the set B contains at least two words, including a word w of length 2.*

a) Prove that the set B is not free.

b) Is it possible that B is free if w is of length 3?

Problem 37. *Suppose that an alphabet consists of n letters and B consists of g two-letter words. Prove that if $g \leq n^2/4$, then the set B may be chosen to be free.*

Problem 38. *Prove that if $n = kd$ and $m \leq k^d(d-1)^{d-1}$, where the numbers d, k, m, n are positive integers, then, over an alphabet of n letters, one can choose a free set consisting of m words of length d.*

Problem 39. *) Prove that, if B is a free set over an alphabet of n letters, then there is the following inequality*

$$\frac{1}{1 - nx + B(x)} \geq 1.$$

b) Is the converse true, that is, is it true that if the inequality

$$\frac{1}{1 - nx + p(x)} \geq 1$$

holds for a positive integer n and a polynomial $p(x)$ whose coefficients are positive integers and whose constant term is zero, then there exists a free set B over an alphabet of n letters, with $B(x) = p(x)$?

Problem 40. *Let n be a positive integer and let $p(x)$ be a polynomial with positive integer coefficients and zero constant term. Prove that there exists a free set B with dimension series $B(x) = p(x)$ if and only if there exist two polynomials f and g with nonnegative integer coefficients with $f(0) = g(0) = 0$ such that*

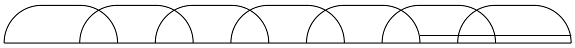
$$(1 - f)(1 - g) \geq 1 - nx + p(x).$$

Problem 41. [3] *Find a condition describing possible dimension series of the sets of forbidden words for non-tangled languages (like we described dimension series of free sets in problem 40).*

7 Words and chains

Definition 11. Let L be a language. *Chains of length one* are the forbidden words, and *chains of length 2* are the simple linkages. Next, one can define the chains of length 3, 4 etc. Namely, a word $v = str$ (where all the words s, t, r are nonempty) is called a *chain of length n* if its initial subword $g = st$ is a chain of length $n - 1$, the final subword $f = tr$ is a forbidden word, where t is a subword of the tail p of the chain g, and there are no other forbidden subwords but f in the final subword pr. The *tail* of the chain v is the word r.

A chain looks as follows (each arc denotes a forbidden subword in the chain):



The length of the chain is the number of arcs. The only overlaps are of neighboring arcs (and the overlaps of neighboring arcs are non-empty). The emphasized two final tails do not contain any forbidden subword but the last arc.

For example, if the only forbidden word is aba, then the only chain of length one is aba, the only chain of length two is $ababa$, the only one of length 3 is $abababa$, etc.

Problem 42. *Suppose that the forbidden words in a language L are the words "tournament", "of", "towns". Write up all chains of length n.*

Problem 43. Antichains *of length n are defined in the same way as chains of length n, with the only difference that we read words of L in Definition 11 "from right to left", i.e., the tail of an antichain is to the left, and the initial antichain of length $n - 1$ is to the right. Prove that the sets of length n chains and length n antichains coincide.*

Problem 44. *Prove that a chain of length n contains no other chain of length n as a subword.*

Problem 45. *Prove that if a word is decomposed as $w = gc$, where g is an admissible word and c is a chain, then, if in addition the length of c is greater than 1, w has exactly two decompositions of this form, and the lengths of the chains in these decompositions differ by 1.*

The next problem gives a way to solve the Main Problem.

Problem 46. *Let L be a language over an alphabet A. Let G be the set of its admissible words and $\overline{G}$ the set of all nonempty admissible words. Let C_1 be the set of chains of length one, C_2 the set of chains of length 2, and so on.*

Prove that

$$L(x) = \frac{1}{1 - Nx + C_1(x) - C_2(x) + C_3(x) - \dots}$$

[3] Neither a solution nor even an answer to this problem are known to the Jury

Problem 47. *Find the dimension series for the language in Problem 42.*

Problem 48. *Find all possible answers to Problem 2 depending on the form of the breaking sequence.*

Problem 49. *We say that a subword c of a word w is its* maximal subchain *if w can be decomposed as $w = gcu$, where g is an admissible word and c is a chain, and for any other decomposition $w = gc'u'$ with another chain c' the word c' is always a subword of c. Prove that any non-admissible word has a single maximal subchain of odd length.*

Problem 50. *Let L be a language over an alphabet A, and let A' be a new alphabet which extends A by one additional letter. Let L' be a language over A' with the same list of forbidden words as L. Prove that*

$$L'(x) = \frac{1}{\dfrac{1}{L(x)} - x}.$$

Problem 51. *A language W is called the* free product *of languages L and L' over disjoint alphabets A and A' if the alphabet of W is the union of the alphabets A and A' and the set of forbidden words is the union of the sets of forbidden words of L and L'. Express the dimension series of the free product W in terms of the dimension series of L and L'.*

Problem 52. *Suppose that all forbidden words of a language L are of two letters. Over the same alphabet, consider another language M whose forbidden words are all two-letter admissible words of L. Prove that*

$$L(x)M(-x) = 1.$$

8 Additional problems

Problem 53. *Prove that there exists a free set of m words of length d over an alphabet of $n = kd$ letters if and only if $m \leq k^d(d-1)^{d-1}$ (cf. Problem 38)*
a) for $d = 2$; *b) for $d = 3$;* *c) for $d > 3$.*

Definition 12. A language is said to be *d-defined* if the maximal length of its forbidden words is d. A 2-defined language is said to be *quadratic*.

Problem 54. *Quadratic languages L and M in Problem 52 are said to be* dual *to each other (notation: $M = L^!$).*
a) Prove that $(L^!)^! = L$.
b) Find $(L_1 + L_2)^!$.
c) Describe $(L_1 \cdot L_2)^!$.

Problem 55. *Let L be a d-defined language. Let us define a new language $L^{(n)}$ over the alphabet consisting of all length n admissible words of L as the language whose admissible words are all admissible words of L whose length is a multiple of n (rewritten in the new alphabet).*
a) Prove that $L^{(n)}$ is defined by finitely many forbidden words.
b) Is $L^{(n)}$ always d-defined?
c) For what minimal n, the language $L^{(n)}$ is necessarily either quadratic or free (for all d-defined languages L)?

Problem 56. *For any quadratic language L over the alphabet $x_1, \ldots, x_n$, let us define an oriented graph Γ_L as follows: it has n vertices labelled with $x_1, \ldots, x_n$, and there is an edge (an arrow) $x_i \to x_j$ if and only if the word x_ix_j is admissible. Denote the number of admissible words of length k by a_k. Prove that*
a) the language L is finite if and only if Γ_L has no cycles;
b) the language L has polynomial growth (i. e., there exist two nonzero polynomials p, q of the same degree d with positive leading coefficient such that $p(k) \geq a_k \geq q(k)$ for each $k \geq 0$) if and only if Γ_L has a cycle but has no intersecting cycles;
c) the language L has exponential growth (i. e., for some $c_1 > c_2 > 1$ and for all k, we have $c_1^k \geq a_k \geq c_2^k$) if and only if Γ_L has at least two intersecting cycles.

Problem 57. *Let L and $L^!$ be a pair of dual quadratic languages. Is it possible that both have exponential growth?*

Problem 58. *For any d-defined language L over the alphabet $x_1, \ldots, x_n$, we define the oriented graph Γ_L as follows: its vertices are labelled with all admissible words of length $d-1$, and there is an edge (arrow) $v \to w$ if and only if there is a letter x_i such that the word vx_i is admissible and the last $d-1$ letters of it constitute the word w. Prove all properties a), b), c) in Problem 56 for Γ_L.*

Definition 13. Let M be a set over an alphabet A. Words u and v (over the same alphabet) are said to be *M-equivalent* if, for any word w, the words uw and vw either both belong to M or neither of them belongs to M. The set M is said to be *regular* if there is a natural number n such that any set of n contains two M-equivalent words.

Problem 59. *Prove that the set of admissible words of any language is regular.*

Definition 14. A *finite automaton* over an alphabet A is an oriented graph Γ with a finite set of vertices V such that
a) the arrows are marked by the letters of the alphabet A, and for every vertex $v \in V$ and each letter $a \in A$, there is a unique arrow marked by a whose tail is v;
b) an *initial vertex* $v_0 \in V$ and a *set of approving vertices* $W \subseteq V$ are given.

Let us consider each word over the alphabet A as an instruction for a trip by arrows over the finite automaton (Γ, v_0, W), that is, we begin with the initial vertex, then go by the arrow marked by the first letter of the word, then follow the arrow marked by the second letter of the word, and so on. We say that the automaton *approves* a word if the path corresponding to the word ends with an approving vertex.

Problem 60. *a) Prove that for every regular set M there exists a finite automaton approving the words of M and no other words.*
b) Prove that for every finite automaton the set of approving words is regular.

Problem 61. *Prove that for every regular set M its dimension series can we represented as a quotient of two polynomials.*

Problem 62. *Let L be a language and M_w the set of all admissible words of L which have a final subword equal to a given word w. Prove that the dimension series of the set M_w can we represented as a quotient of two polynomials.*

Note. Parts 1–5 were suggested before the intermediate consideration of the problems. Parts 6–8 were added after the intermediate consideration of the problems.

How to count words?

Dmitri Piontkovski, Maxim Prasolov, Grigory Rybnikov

Solutions

1 Main problem

1. Cf. Problem 24.

2. Cf. Problems 18 and 48.

3. One version of the solution is given in problem 46; another version can be obtained using Problems 59 and 61.

4. In Problem 1, the alphabet A consists of $N = 100$ words of the tribe language. The phrases of the tribe language play the role of words of L, and the forbidden words of L are the two magic spells. In Problem 2, the alphabet A consists of $N = 256$ commands of the computer, and the programs play the role of words of L.The only forbidden word is the program of 4 commands that breaks the computer.

2 How to write down the answer?

5. To get an arbitrary word of m letters, one choose any of N letters in any of m places. Multiplying the numbers of possibilities in each place, we get N^m words.

6. If the first letter of an admissible word is x, then the second one is x as well. It follows that each admissible word has the form $xx \dots x$, where x is one of the N letters. Therefore, the number of admissible words of any given number of letters is N.

7. Assume that the first letter of an admissible word is a. Since aa is a forbidden word, then the second letter is b. Proceeding by the same way, we get a on the odd places and b on the even places. Similarly, if the first letter is b, then we get a on even places and b one odd ones. Thus, the dimension series of this language is $1 + 2x + 2x^2 + 2x^3 + \dots$

3 The arithmetics of languages

8. The collection of forbidden words is the following: all forbidden words of the both languages and all words of 2 letters such that the first letter is of the first alphabet and the second one is of the the second alphabet. Obviously, the admissible word of each language do not contain a subword which is forbidden in the sum of languages. Let w be a word of sum which does not contain subwords equal to the words described above. If its first letter is, say, in the first alphabet, then the subsequent letters are in the first alphabet as well, that is, each such a word consists of the letters of the same alphabet. It follows that w is admissible in the language of this alphabet, thus, it is admissible in the sum.

9. The initial terms of the series $L(x)$ and $L_1(x) + L_2(x) - 1$ are equal to 1. For $n > 0$, the coefficient of x^n in the series $L_1(x) + L_2(x) - 1$ is equal to the sum of the numbers of words of n letters in the languages L_1 and L_2, that is, the number of words of n letters in the language L, which is equal to the coefficient of x^n in the series $L(x)$.

10. We have

$$(1-x)(1+x+x^2+x^3+\dots) = 1 - x + (1-x)\cdot x + (1-x)\cdot x^2 + (1-x)\cdot x^3 + \dots =$$
$$= 1 - x + x - x^2 + x^2 - x^3 + x^3 - x^4 + \dots = 1,$$

as required.

11. The collection of forbidden words is the following: all forbidden words of two given languages and words of two letters such that their first letter is in the second alphabet and the second one is in the first alphabet. Consider an admissible word w of the product. In w, the letters of the second alphabet follow to the letters of the first one, therefore, w has the from w_1w_2, where w_1 is a word of the first language and w_2 is a word of the second one. The word w_1w_2 do not contain subwords which are forbidden in the languages-multipliers, therefore, w_1 and w_2 are admissible in their languages, that is, the word w_1w_2 is admissible in the product of the languages.

12. The coefficient of x^k in the series $L_1(x) \cdot L_2(x) = L_1(x) \cdot n_0 + L_1(x) \cdot n_1x + \dots = (n_0 \cdot m_0 + n_0m_1x + \dots) + (n_1m_0x + n_1m_1x^2 + \dots) + \dots$ is $n_0m_k + n_1m_{k-1} + \dots + n_km_0$. The number of words of length k in the set of words $L_1 \cdot L_2$ is equal to the number of possibilities to get a pair of words, m in L_1 and n in L_2, such that the total number of letters in these words is k. If the word m is of i letters, then the word n is of $k-i$ letters, so that the number of such pairs is equal to $m_i \cdot n_{k-i}$. Taking a sum of all such products for all i, one gets $n_0m_k + n_1m_{k-1} + \dots + n_km_0$. Thus, the coefficients of x^k in two series $L_1(x) \cdot L_2(x)$ and $L_1 \cdot L_2(x)$ coincide, therefore, the series itself coincide.

13. a) First let us show that the standard properties of addition and multiplication of polynomials (associativity, commutativity, distributivity) hold for series as well. For example, consider associativity relation for multiplication $(P \cdot Q) \cdot R = P \cdot (Q \cdot R)$. To compute the coefficient of x^k in the both sides of this relation, it is enough to make computations for the same series without terms of degree higher than k, i. e., for polynomials. Therefore, the relation for series follows from the same relation for polynomials. The other relations are proved in the same way.

Now note that, since the series $\overline{S}$ starts with x^1, the series $R \cdot \overline{S}^m$ has no terms of degree less than m. That is why infinite sums of the form $R - R \cdot \overline{S} + R \cdot \overline{S}^2 - R \cdot \overline{S}^3 + \dots$ make sense: to find the kth coefficient, the sum can be replaced by a finite one. For the same reason, the sums of this type satisfy the distributivity relation

$$R \cdot (S_1 + S_2 + S_3 + \dots) = R \cdot S_1 + R \cdot S_2 + R \cdot S_3 + \dots.$$

Having this in mind, we easily get

$$(1 + \overline{S})(1 - \overline{S} + \overline{S}^2 - \overline{S}^3 + \dots) = 1.$$

Hence

$$S \cdot \frac{R}{S} = S \cdot (R - R \cdot \overline{S} + R \cdot \overline{S}^2 - R \cdot \overline{S}^3 + \dots) = S \cdot R \cdot (1 - \overline{S} + \overline{S}^2 - \overline{S}^3 + \dots) =$$
$$= R \cdot (1 + \overline{S})(1 - \overline{S} + \overline{S}^2 - \overline{S}^3 + \dots) = R.$$

b) By the above, we have

$$S \cdot (T - \frac{R}{S}) = S \cdot T - S \cdot \frac{R}{S} = R - R = 0$$

Assume that the series $(T - \frac{R}{S})$ is nonzero. Since S starts with 1, the first nonzero coefficient of $(T - \frac{R}{S})$ is equal to the first nonzero coefficient of $S \cdot (T - \frac{R}{S})$. Therefore $T - \frac{R}{S} = 0$ and $T = \frac{R}{S}$.

14. a) Similarly to Problem 10, we obtain

$$F_A(x) = 1 + Nx + N^2x^2 + \dots = \frac{1}{1 - Nx}.$$

b) We have

$$1 + Nx + Nx^2 + Nx^3 + \dots = -N + 1 + N \cdot (1 + x + x^2 + x^3 + \dots) = -N + 1 + \frac{N}{1 - x} = \frac{1 + (N - 1)x}{1 - x}$$

and

$$1 + 2x + 2x^2 + 2x^3 + \dots = 1 + 2x \cdot (1 + x + x^2 + x^3 + \dots) = 1 + \frac{2x}{1 - x} = \frac{1 + x}{1 - x}.$$

15. The solution follows from Problems 59 and 61.

4 Free word

16. It follows from Problem 17 below that

$$L(x) = \frac{1}{1-26x+x^5}.$$

Also, one can directly obtain the above formula in a similar way as in the solution of Problem 17.

17. Let a_k be the number of admissible words of length k. Clearly, $a_0 = 1$. Let us prove the recurrent relation $a_k = Na_{k-1} - a_{k-m}$ for $k > 0$ (we have $a_i = 0$ for $i < 0$, since there are no words of negative length).

Indeed, by adding each letter of the alphabet to the beginning of each admissible word of length $k-1$, we obtain Na_{k-1} words, among which are all admissible words of length k. Let us find which non-admissible words of length k can be obtained in this way, i. e., can be written as cg, where c is a letter and g is an admissible word of length $k-1$. Clearly, the forbidden subword must stand at the beginning, so $cg = wf$, where w is the forbidden word and f is admissible. Since w is free, we conclude that, for any admissible word f, the word obtained from wf by cutting its first letter is admissible (otherwise w would have an overlap with itself). Therefore, the set of all words of the form cg, where c is a letter and g is an admissible word of length $k-1$, is the union of two non-intersecting sets: the set of all admissible words of length k and the set of all words of the form wf, where f is an admissible word of length $k-m$. Hence we get the recurrent relation.

Consider the sum of relation $a_0 = 1$ and all relations $a_k x^k = Na_{k-1}x^k - a_{k-m}x^k$ for $k = 1, 2, 3, \ldots$. We obtain

$$L(x) = 1 + NxL(x) - x^m L(x).$$

Solving this equation with respect to $L(x)$, we get the required formula.

18. According to Problem 17,

$$L(x) = \frac{1}{1-256x+x^4} = 1 + (256x - x^4) + (256x - x^4)^2 + (256x - x^4)^3 + \ldots$$

Here the coefficient of x^7 is $256^7 - 4 \cdot 256^3$. Thus, the probability of computer break is $\frac{4 \cdot 256^3}{256^7} = 4 \cdot 256^{-4}$, or, approximately, 10^{-10}.

5 Transformations of words

19. The first arrow is determined uniquely; the second one maps each word in N to the same word regarded as an element of F_A; the third one maps each of the remaining words in F_A to the same word as an element of G; the last arrow is as trivial as the first one.

20. Each of the children eats as many bonbons that belonged to boys as bonbons that belonged to girls. The last girl eats the last bonbons that belonged to boys. Thus she also eats the last bonbons that belonged to girls, and the teacher gets nothing at all.

21. a) Let $M_{\text{odd}} = M_1 \cup M_3 \cup M_5 \cup \ldots$ and $M_{\text{even}} = M_2 \cup M_4 \cup M_4 \cup \ldots$. Each transformation establishes a one-to-one correspondence between a subset of M_{odd} and a subset of M_{even}, besides, since both the rightmost and the leftmost sets are empty, each element participates in exactly one of these correspondences. Hence the sets M_{odd} and M_{even} have the same number of elements.
b) For each k, the set $M_i^{(k)}$ of words in M_i of length k is finite; by applying assertion a) to the finite sets $M_i^{(k)}$ with the same k, we conclude that the coefficient of x^k in the left-hand side of the formula is the same as in its right-hand side. Since k is arbitrary, it means that the formula is correct.

22. Let us verify that the set $A \cdot G$ is the union of two non-intersecting sets: the set $\overline{G}$ and the set $B \cdot G$. The proof, which is based on the fact that the set B is free, is almost literally the same as the corresponding reasoning in the solution of Problem 17.

Now it is easy to construct the required exact sequence: the first and the last arrows are trivial, the second one maps each element of $B \cdot G$ to itself (here we use that $B \cdot G \subseteq A \cdot G$), and the third one maps each of the remaining elements of $A \cdot G$ to itself (here we use that $A \cdot G \setminus B \cdot G = \overline{G}$). In particular, the kernel of the second transformation is empty, and the kernel of the third transformation, which is the same as the image of the second one, is $D \cdot G$; the image of the third transformation is $\overline{G}$.

23. By Problem 21b), the exact sequence in Problem 22 implies

$$(B \cdot G)(x) + \overline{G}(x) = (A \cdot G)(x).$$

Note that each element of $A \cdot G$ can be *uniquely* written as ag, where $a \in A, g \in G$. Hence $(A \cdot G)(x) = A(x)G(x)$. Further, each element of $B \cdot G$ can be *uniquely* written as bg, where $b \in B, g \in G$, (since no forbidden word is a subword of another forbidden word). Hence $(B \cdot G)(x) = B(x)G(x)$. We have $A(x) = Nx$, $G(x) = L(x)$, $\overline{G}(x) = G(x) - 1 = L(x) - 1$. Therefore,

$$B(x)L(x) + L(x) - 1 = NxL(x).$$

Solving this equation with respect to $L(x)$, we get the required formula.

24. Denote the words that occur in the spells by letters: "earth" — A, "stand" — B, "on" — C, "great" — D, "crocodile" — E, "every" — F, "evening" — G, "swallow" — H, "sun" — I. Then the spells correspond to forbidden words "ABCDE" and "FGEHI". These words are free (since all letters in each of them are distinct) and have no overlaps with each other (since both the first and the last letters of the second word do not occur in the first one). Therefore, the set of spells is free, and and the dimension series for the language is

$$L(x) = \frac{1}{1 - 100x + 2x^5}.$$

Using this formula, it is not hard to show (see the solution of Problem 17), that a_k (the number of sentences of k words) can be computed from the initial condition $a_0 = 1$ and the recurrent relation $a_k = 100a_{k-1} - 2a_{k-5}$. The computations provide us with the answer $a_{20} = 10^{40} - 32 \cdot 10^{30} + 264 \cdot 10^{20} - 448 \cdot 10^{10} + 16$.

25. Letter v occurs only as the first letter of each forbidden word and all forbidden words are of length 4. Hence the set of forbidden words is free. By Problem 23, we have

$$L(x) = \frac{1}{1 - 26x + 3x^4}.$$

26. If the set of forbidden words is free, then, in particular, there are no simple linkages. Let us prove that if there are no simple linkages, then the set of forbidden words is free. Assume the contrary, i. e., that the set of forbidden words is not free. Then there is an overlap of two forbidden words, that is, there exist three nonempty words s, t, r such that the words st and tr are forbidden. Choose such a triple (s, t, r) so that the length of str be minimal. If it is not a simple linkage, then str has a forbidden subword w other than st and tr. Note that the end of w is not he end of tr since otherwise either w would be a subword of tr or tr would be a subword of w, which is impossible as no forbidden word is a subword of another forbidden word. Similarly, the beginning of w is not the beginning of st. For the same reason, the subword w overlaps with both s and r. Denote the common part of st and w by t', the remaining part of st by s', and the remaining part of w by r'. These words are nonempty, the length of $s't'r'$ is less than the length of str, the words $s't' = st$ and $t'r' = w$ are forbidden. We obtain a contradiction. Thus, if there are no simple linkages, the set of forbidden words is free.

27. We construct the transformations starting from the end (from the rightmost arrow). Since the last set is empty, the domain of definition of the last transformation is also empty. Hence the image of the last but one transformation is the whole set $\overline{G}$. Since $\overline{G} \subseteq A \cdot G$, we can take $\overline{G}$ to be the domain of definition of the last but one transformation, and define the corresponding function to map each element $g \in \overline{G}$ to itself. The kernel of this transformation consists of all non-admissible words of the form ag, where a is a letter and g is an admissible word. It is readily seen that, for any word of this type, there exist a forbidden word w and an admissible word f such that $ag = wf$ (we already used similar reasoning in the solutions of Problems 17 and 22). So we can construct the third arrow from the right (this transformation also maps each element of its domain to itself). Consider the kernel of this transformation. It consists of those words of the form wf, where w is forbidden and f is admissible, which also have the form av, where a is a letter and v is a non-admissible word. Choose the leftmost forbidden subword u in v. Clearly, the subword u of the word $av = wf$ overlaps with the subword w and forms a simple linkage with it. Thus the kernel of the third arrow from the right is contained in $S \cdot G$. Hence it is possible to construct transformation $S \cdot G \Longrightarrow B \cdot G$ (which also maps each element of its domain to itself).

28. By the solution of the previous problem, we see that a language is non-tangled if and only if all words of the form rg, where r is the tail of a simple linkage and g is an admissible word, are admissible. An equivalent condition for the set of forbidden words writes as follows: there exist no such words p, q, r, s, t, where the p, q, s, t are nonempty, the words pq, qrs, st are forbidden, and $pqrs$ is a simple linkage.

Note that any element of the set $S \cdot G$ can be uniquely represented as the product of a simple linkage by an admissible word (it follows easily from the definition of simple linkage and the fact that no forbidden word is a subword of another forbidden word). In the same way as in the solution of Problem 23, for a non-tangled language L, we use the exact sequence to obtain the following equation:

$$S(x)L(x) + NxL(x) = B(x)L(x) + L(x) - 1,$$

whence follows the required formula

$$L(x) = \frac{1}{1 - Nx + B(x) - S(x)}.$$

29. Simple linkages are *abbc* and *abbac*, and their tails are c and ac. It is clear that none of these tails ends with the beginning of a forbidden word. Thus the language is non-tangled. Therefore, the dimension series is

$$L(x) = \frac{1}{1 - 3x + 3x^3 - x^4 - x^5}.$$

30. Simple linkages are $x_i y_j z_k$, where $1 \le i, j, k \le n$, their tails are z_k. Since no forbidden word starts with z_k, the language is non-tangled. Therefore, the dimension series is

$$L(x) = \frac{1}{1 - 3nx + 2n^2x^2 - n^3x^3}.$$

31. Let w be the unique forbidden word and let L be its length. Suppose that w is not free; let pqr, where $pq = qr = w$, is a simple linkage. We have $wr = pqr = pw$. Therefore, the last subword of length L in each of the words $wr = pw, wrr = pwr = ppw, wrrr = ppwr = pppw, \dots$ is equal to w. Take the first word in this sequence which has length at least $2L$. Then a word of the from $rrr \dots r$ has the final subword equal to w. But this means that the word r has a nonempty ending which is an initial subword of w. Therefore, the language under consideration is tangled (Cf. the solution of Problem 28).

6 Free sets revisited

32. For example, if the alphabet consists of the letters a and b, then the set of words $a^n b^n ab$, where $n \ge 2$, is free. Let us prove this. Obviously, no two words are subwords of each other. It remains to prove that there is no nontrivial overlap (i. e., each overlap is the letter-by-letter application of a word on itself). Let w is an overlap of the words $a^n b^n ab$ and $a^m b^m ab$. Then it is easy to see that w has at least three letters. Since w is an end of the word $a^n b^n ab$, it has the form either $b^k ab$ or $a^k b^n ab$, where $1 \le k \le n$. Because w is also a begin of the word $a^m b^m ab$, we get $k = m = n$ and $w = a^n b^n ab = a^m b^m ab$ is a trivial overlap.

33. Lemma. Let $p(x) = 1 + p_1 x + p_n x^n$ be a polynomial of degree $n \ge 1$. Then the series $f(x) = 1/p(x)$ cannot be a polynomial (i. e., this series has an infinite set of nonzero terms).

Proof of Lemma. Suppose (ad absurdum) that the series $f(x)$ is a polynomial, that is, $f(x) = f_0 + f_1 x + \dots f_m x^m$, where the leading coefficient f_m is nonzero. According to Problem 13 a), we have $1 = f(x)p(x) = 1 + (f_0 p_1 + f_1 p_0)x + \dots + f_m p_n x^{m+n}$, a contradiction.

Return to Problem 33. According to Problem 23, we have

$$L(x) = \frac{1}{1 - Nx + B(x)}.$$

If the language L was finite, the series $L(x)$ would be a polynomial, in contradiction with the above Lemma. It follows that the set of admissible words is infinite.

34. Obviously, for series A and B the inequality $A \ge B$ is equivalent to an inequality $A - B \ge 0$, which is equivalent to the condition that the coefficients of the series $A - B$ are nonnegative. Denote the series $P - Q$ by $A = a_0 + a_1 x + a_2 x^2 \dots$, and the series R by $R = r_0 + r_1 x + r_2 x^2 \dots$ Then an n-th coefficient of the series AR is given by the formula $a_0 r_n + a_1 r_{n-1} + \dots + a_n r_0$. So, a_n is a sum of nonnegative numbers, so that $a_n \ge 0$. This means that it holds an inequality $AR \ge 0$. Equivalently, we have $PR - QR \ge 0$, or $PR \ge QR$.

35. According to Problem 23,

$$L(x) = \frac{1}{1 - A(x) + B(x)}.$$

By Problem 27, there is an exact sequence

$$\emptyset \Longrightarrow K \Longrightarrow B' \cdot G' \Longrightarrow A \cdot G' \Longrightarrow \overline{G}' \Longrightarrow \emptyset,$$

where K is a kernel of the transformation $B \cdot G \Longrightarrow A \cdot G'$ and G' is the set of admissible words of the language L'. It follows from this exact sequence (Problem 21) that

$$B'(x)G'(x) - A(x)G'(x) + G'(x) - 1 = K(x),$$

therefore (since $B'(x) = B(x)$, $L'(x) = G'(x)$ and $K(x) \geq 0$),

$$L'(x)(B(x) - A(x) + 1) \geq 1.$$

Multiplying this by the series $L(x) \geq 0$, we get (using Problem 33)

$$L'(x)(B(x) - A(x) + 1) \cdot \frac{1}{1 - A(x) + B(x)} \geq L(x),$$

i. e.,

$$L'(x) \geq L(x).$$

36. Let $A = \{a, b\}$ be the alphabet. Denote the second word by v. Obviously, if the words v and w are free, then their initial and last letters differ. If, in addition, w the initial letters of the two words differ, then the last letter of v coincides with the first one of w, and , so that the set B is not free. So, it remains to consider the case when v and w begin with the same letter (say, a) and end with another one (b).

a) We have $w = ab$ and $v = a...b$. Obviously, if the first appearing of b in the word v is in k-th place, then the subword of v which consists of the $(k-1)$-th and k-th letters is w. It follows that B is not free.

b) Answer: no. Let $w = aab$ (the case $w = abb$ is analogous, up to the right–left symmetry and the interchanging the letters). Since the word w is not a subword of v, in v the letters that follows the pair of letters aa is again a. Since the words v and w have overlaps, v cannot begin with ab, i. e., v begins with aa. Therefore, the 3rd letter of v is a, as well as the 4th etc. It follows that $v = aa \ldots a$, a contradiction.

37. It is sufficient to show that there exist a free set B of $g = \lceil n^2/4 \rceil$ two-letter words. Let $k = [n/2]$, i. e., $n = 2k$ or $n = 2k+1$. Put $B = \{x_i x_j | 1 \leq i \leq k, k+1 \leq j \leq n\}$. Obviously, the set B is free. Then for an even $n = 2k$, the set B consists of $k^2 = n^2/4$ elements, and in the case of odd $n = 2k+1$ the set B is of $k(k+1) = (n-1)(n+1)/4 = n^2/4 - 1/4 = \lceil n^2/4 \rceil$ elements, as required.

38. It is sufficient to show that there exist a free set B of $m \leq k^d(d-1)^{d-1}$ words of length d. Let us divide the alphabet $A = \{x_1, \ldots, x_n\}$ by two subsets $P = \{x_1, \ldots, x_k\}$ and $Q = \{x_{k+1}, \ldots, x_n\}$. Then is es easy to see that the set $B = \{pq | p \in P, q \in F_Q \text{ — } \quad d-1\}$ is as needed.

39.

a) According to Problem 23, $\frac{1}{1-nx+B(x)} = L(x) \geq 1$.

b) Answer: no. For example, over a 2-letters alphabet A there is no such a set B with $p(x) = x^3 + x^{10}$ (it is shown in Problem 36 b). It remains to see that the series

$$f(x) = \frac{1}{1 - 2x + x^3 + x^{10}}$$

has nonnegative coefficients (since the initial term of the above series is 1, the above condition is equivalent to the required inequality $f(x) \geq 1$). We will show that the coefficients of the series satisfy the stronger inequality $a_n \geq (3/2)^n$. One can provide the proof of the last inequality by induction, using the reccurent relation $a_{n+10} = 2a_{n+9} - a_{n+7} - a_n$, which follows from the condition $(1 - 2x + x^3 + x^{10})f(x) = 1$.

Another similar example is the case $p(x) = 4x^6$, again over the two-letter alphabet.

40. The proof is given in Theorem 5.1 (equivalence A$\Longleftrightarrow$B) and Proposition 5.6 in the paper: David Anick, *Generic algebras and CW–complexes*, Proceedings of 1983 Conference on algebra, topology and K–theory in honor of John Moore. Princeton University, 1988, p. 247–331.

41. The question is still open.

7 Words and chains

42. The word “of” has not any overlaps with the other words because no words begin with the letter f and do not end with the letter o. There is no letter s in the word “tournament” so a chain can consist the word “towns” only as the last arc. A letter t is only on the first and the last position in the word “tournament”. So overlaps of the word “tournament” with itself are only “tournamentournament”.

Now we find all chains: tournament, of, towns; tournamenttournament, tournamenttowns;... There are two chains of the length n for $n > 1$.

43. From the arc on the right of a chain construct an antichain leftward arc by arc. We obtain a unique antichain of the length i if it exists. We prove by induction that the beginning of the ith arc of the antichain lies between the beginning of the ith from the right (numeration of arcs is from the right) and the end of the (i+1)th chain arcs. The first arcs of chain and antichain coincide, therefore the base of induction holds for $i = 1$. Now check the induction statement for $i = 2$. The second antichain arc is on the right from the second chain arc because there are not forbidden words between the first and the second antichain arcs. The beginning of the second antichain arc is on the left from the end of the third (from the right) chain arc because there is no forbidden words after the end of the third chain arc by the definition of a chain. So the base holds for $i = 2$. Then we prove the induction step. Suppose the induction statement is true for 1,2,...,i. By the induction assumption the ith chain arc intersects the (i−1)th antichain arc, but the (i+1)th antichain arc does not intersect the (i−1)th antichain arc, therefore the (i+1)th antichain arc is on the left from the ith chain arc. Between the end of the (i+2)th and the beginning of the ith chain arcs no forbidden words begin, therefore the beginning of the (i+1)th antichain arc is on the left from the end of the (i+2)th chain arc. The (i−1)th antichain is not on the left from the (i−1)th chain arc, so the (i+1) chain arc does not intersect the (i−1)th antichain arc, but intersects the ith arc. Therefore the (i+1)th antichain arc is not on the left from the (i+1)th chain arc. This finishes the proof.

44. Assume that a chain c' is a subword of a chain c. Let us prove by induction that the ith c-arc is not on the right from the ith c'-arc. The statement is obvious for $i = 1$. If the first arcs of considered chains coincide and the chains have the same length then their arcs coincide and so the whole chains are equal. Hence the chain c' does not start from the beginning of the word c. Between the first and the second c-arcs there is no forbidden words. Therefore the first c'-arc is not on the left from the second c-arc and then the second c'-arc is on the right from the second c-arc. So we proved the base for $i = 2$. Now we prove an induction step. We consider two cases.

The first case. The (i+1)th c'-arc does not intersect the ith c-arc. Then the (i+1)th c'-arc is on the right from the (i+1)th c-arc.

The second case. The (i+1)th c'-arc intersects the ith c-arc. The (i+1)th c'-arc does not intersect the (i−1)th c'-arc, therefore by induction hypothesis the (i+1)th does not intersect the (i−1)th c-arc. Summarizing observations we obtain that by definition of chain the (i+1)th c-arc is not on the right from the (i+1)th c'-arc.

The induction statement is proved and we must only consider the case when the last c- and c'- arcs coincide. In this case we note that these chains as antichains (by the previous problem) have the same arcs on the right and the same length. Therefore they coincide.

45. Assume that a word is decomposed as gc where g is admissible word and c is a chain of the length at least two. By the problem 43 the chain c is also an antichain. If you reduce the chain c by the beginning of the first arc and it will appear a forbidden word that is on the left from the reduced chain then the antichain c can be continued to the left and we obtain a new decomposition $g'c'$ where the chain length is increased by one. If a forbidden word does not appear then the antichain c can be reduced by the beginning of the arc on the left and we obtain a new decomposition $g''c''$ where the chain length is decreased. If the third decomposition $g''c''$ exists then note that the arcs of the antichains c, c', c'' are the same so there are two chains among them which lengths differ at least by two. But in this case the arc on the left of the longest antichain does not intersect the shortest antichain and we have a contradiction.

46. We will apply the problem 21 for an exact sequence

$$\ldots \Longrightarrow C_{n+1} \cdot G \Longrightarrow C_n \cdot G \Longrightarrow C_{n-1} \cdot G \Longrightarrow \ldots C_1 \cdot G \Longrightarrow A \cdot G \Longrightarrow \bar{G}$$

Let us construct this sequence. Choose a word cg from $C_n \cdot G$. Add a tail of the chain c to the beginning of the word g and obtain a decomposition $c'g'$. If g' is an admissible word then the word $c'g'$ belongs to $C_{n-1} \cdot G$. If g' contains a forbidden word then the chain c can be continued to the right to the chain c'', the rest of the word denote by g''. Then the word $c''g''$ belongs to $C_{n+1} \cdot G$. A chain can be continued in unique way in the word, so the constructed maps are biunique on the parts of $C_n \cdot G$.

The arrows $C_1 \cdot G \Longrightarrow A \cdot G \Longrightarrow \bar{G} \Longrightarrow \emptyset$ are the same as in the problem 27. By the problem 21,

$$1 - L(x)(1 - Nx + C_1(x) - C_2(x) + C_3(x) - \ldots) = 0.$$

So we obtain the formula.

47. By the problem 42 $C_1(x) = x^2 + x^5 + x^{10}, C_n(x) = x^{9(n-1)}(x^5 + x^{10})$. By the formula from the problem 46

$$L(x) = \frac{1}{1 - 26x + x^2 + (x^5 + x^{10})(1 - x^9 + x^{18} - \ldots)} = \frac{1}{1 - 26x + x^2 + \frac{x^5+x^{10}}{1+x^9}} =$$

$$= \frac{1 + x^9}{1 - 26x + x^2 + x^5 + x^9 - 25x^{10} + x^{11}}$$

48. Consider four cases of a forbidden word of four letters.

1) The forbidden word is of the form aaaa. In that case $C_{2n} = x^{4n+1}, C_{2n-1} = x^{4n}$. By the formula in the problem 46

$$L(x) = \frac{1}{1 - 256x + x^4 - x^5 + \ldots} = \frac{1}{1 - 256x + \frac{x^4 - x^5}{1-x^4}} = \frac{1 - x^4}{1 - 256x + 255x^5} =$$

$$= (1 - x^4)(1 + (256x - 255x^5) + (256x - 255x^5)^2 + \ldots).$$

Then a coefficient of x^7 equals to $256^7 - 3 \cdot 256^2 \cdot 255 - 256^3 = 256^7 - 4 \cdot 256^3 + 3 \cdot 256^2$.

2) The forbidden word has the form abca and at least two distinct letters. In that case $C_n = x^{3n+1}$.

$$L(x) = \frac{1}{1 - 256x + x^4 - x^7 + \ldots} = 1 + (256x - x^4 + x^7 - \ldots) + (256x - x^4 + x^7 - \ldots)^2 + \ldots$$

A coefficient of x^7 equals to $256^7 - 4 \cdot 256^3 + 1$

3) The forbidden word is of the form abab. In that case $C_n = x^{2(n+1)}$.

$$L(x) = \frac{1}{1 - 256x + x^4 - x^6 + x^8(\ldots)}$$

A coefficient of x^7 equals to $256^7 - 4 \cdot 256^3 + 2 \cdot 256$.

4) The forbidden word is free. This case was analyzed in the problem 18.

49. From the 46 it follows that

$$\frac{1}{1 - Nx} - L(x) = L(x) \cdot C_1(x) \cdot \frac{1}{1 - Nx} - L(x) \cdot C_2(x) \cdot \frac{1}{1 - Nx} + L(x) \cdot C_3(x) \cdot \frac{1}{1 - Nx} - \ldots$$

The left part of this equality is a dimension series of the set of inadmissible words. The right part of this equality is an alternative sum of dimension series of langugages $L \cdot C_n \cdot F_A$. One can define a map from these labguages to the set of inadmissible words and vice versa, every inadmissible word can be decomposed as gc_nu, where g is admissible, c_1 is a forbidden word. But one can decompose an inadmissible word as gc_nu in several ways. Denote the number of such decompositions of the word w as w_n. So then the sum of the numbers $(w_1 - w_2 + w_3 - w_4 + \ldots)$ for all inadmissible words of the length k equals to a coefficient of x^k in the right part of the equality behind – and therefore in the left part, that is equal to the amount of inadmissible words of the length k.

Note that from the decomposition of the word w as gc_nu one can obtain another decomposition, in which the chain length is decreased by one, by moving a tail of c_n to u. So two these decompositions will be cancel in the sum $(w_1 - w_2 + w_3 - w_4 + \ldots)$. In other words the sum $(w_1 - w_2 + w_3 - w_4 + \ldots)$ equals to the quantity of decompositions of the word w as gcu where subword c is a maximal chain of odd length.

Consider a decomposition of the word w as gcu where c is a maximal chain with the most right [1]arc. By the problem 45 in the word gc there is a maximal chain of the length one or it can be decomposed as $g'c'$ where g' is admissible and c' is a chain which length differs by one from the length of the chain c. In both cases there is a maximal chain of odd length in the word w. So we showed that $(w_1 - w_2 + w_3 - w_4 + \ldots)$ is at least one. But the sum of such quantities for all inadmissible words equals to their number. So then every such quantity equals to one and in every inadmssible word there is only one maximal chain of odd length.

50. Apply the formula from the problem 46.

$$L'(x) = \frac{1}{1 - (N+1)x + C_1(x) - C_2(x)) + \ldots} = \frac{1}{\frac{1}{L(x)} - x}$$

51. Apply the formula from the problem 46.

$$W(x) = \frac{1}{1 - (N + N')x + (C_1(x) + C_1'(x)) - (C_2(x) + C_2'(x)) + \ldots} = \frac{1}{\frac{1}{L(x)} + \frac{1}{L'(x)} - 1}$$

52. Admissible words of the language M are the chains of the language L. Chains of the length n consist of n+1 letters. Therefore $C_n(-x) = (-1)^{n+1}C_n(x)$. Thus we obtain $M(-x) = 1 - Nx + C_1(x) - C_2(x) + \ldots$, i.e. $L(x)M(-x) = 1$.

8 Additional problems

53. The existence of a free subset under the assumption $m \leq k^d(d-1)^{d-1}$ is established in Problems 37 and 38. It remains to show that no such set exists for $m > k^d(d-1)^{d-1}$.

[1] this means that any arc of any maximal chain is not on the right from the last arc of the chain c

a) If we are given $m > n^2/4$ words of length two, which form a free set S, then the first letter of any of these words cannot coincide with the last letter of another word, i. e., there are two disjoint subsets of letters, P and Q, whose elements may only serve as the first letter or the last letter, respectively, for a word in S. Let $r = |P| + |Q| \leq n$, and let $s = |P| \cdot |Q| \geq |S| = m > n^2/4$. By the Viet theorem, the numbers $|P|$ $|Q|$ are the roots of quadratic equation $x^2 - rx + s = 0$, whose discriminant $D = r^2 - 4s$ is negative under the above constrains on r and s; hence we get a contradiction.

b) Let B be a free set consisting of m words of length 3. Since no letter can be both the first and the last letter for words of the same free set, the alphabet A contains two disjoint subsets X and Y, whose element can serve as the first letter or the last letter, respectively, for a word in B. If there are letters that does not occur as the first or the last letter of a word in B, we add each of them to one of X and Y. Without loss of generality, we assume that the number s of elements of the set $X = \{x_1, \dots, x_s\}$ is no greater than the number t of elements of $Y = \{y_1, \dots, y_t\}$. Each element B has the form $x_{i_1}x_{i_2}y_{j_1}$ or $x_{i_1}y_{j_1}y_{j_2}$, and no final subword xy of an element of the first type can be the beginning of an element of the second type. Let us change the set B by replacing each word of the first type with a word of the second type according to the rule $x_ixy \mapsto xyy_i$, where $1 \leq i \leq s \leq t$. It is readily seen that such transformation maps no element of B to another element of B, no distinct elements are mapped to the same one, and the resulting set B' remains free. All elements of B' are of the form $x_{i_1}y_{j_1}y_{j_2}$. Therefore, the number of the elements of B' (which is still m) does not exceed st^2, whence

$$m \leq st^2 \leq (n-t)t^2 \leq \left(n - \frac{2n}{3}\right)\left(\frac{2n}{3}\right)^2 = \frac{4n^3}{27} = 4k^3,$$

as required.

c) The proof uses the following analytical

Lemma. Let $R(x) = 1 + a_1x + a_2x^2 + \dots$ be a series with positive integer coefficients. Suppose that $R(x) = 1/p(x)$ for a polynomial $p(x)$ with constant term 1. Let $R_n(x) = 1 + a_1x + a_2x^2 + \dots + a_nx^n$. If we have $p(x) \geq m > 0$ for all $x \in [0, x_0]$, where $x_0 > 0$, then inequality $R_n(x_0) \leq 1/m$ holds for each $n > 0$.

Omitting the proof of the Lemma, we pass to the solution of the Problem. Let $s = mk^{-d} - (d-1)^{(d-1)}$; we need to show that no free set exists if $s > 0$. Assume the contrary. Then, by Problem 39 a), the series $1/p(x)$, where $p(x) = 1 - dkx + mx^d$, has positive integer coefficients (there can be no zero coefficients, since the series is infinite by Problem 33). Note that the polynomial $p(x)$ is positive on the segment $[0, 1]$ whenever $s > 0$ (proof: the minimum of this polynomial on $[0, 1]$ is achieved either at an end of this segment, where $p(x)$ is positive, or at a point x_0 such that $p'(x_0) = 0$, i. e., at $x_0 = \frac{1}{k(d-1)}$; then $p(x_0) = sx_0^d > 0$). It means that there is a number $m > 0$ such that $p(x) \geq m$ for $x \in [0, 1]$. By Lemma, it follows that, for all n, the number of words of length at most n, which is $L_n(1)$, is bounded by the constant $1/m$.

54. a) By definition, the forbidden words of $L^!$ are all two-letter words that are not forbidden in L. Hence the forbidden words of $(L^!)^!$ are all two-letter words that are not forbidden in $L^!$, i. e. exactly all forbidden words of L. Thus the alphabets and the sets of forbidden words for languages L and $(L^!)^!$ are the same, hence the languages are equal.

b) Since the set of forbidden words for the language $M = (L_1 + L_2)^!$ is the union of the sets of forbidden words for the languages $L_1^!$ and $L_2^!$, and the alphabet of M is the union of their (disjoined) alphabets, the language M is the free product of $L_1^!$ and $L_2^!$ (see definition in Problem 51).

c) The forbidden words for $(L_1 \cdot L_2)^!$ are the admissible two-letter words of the languages L_1 and L_2 and all words of the form aB, where a is a letter of the alphabet of L_1 and B is a letter of the alphabet of L_2. Hence

$$(L_1 \cdot L_2)^! = L_2^! \cdot L_1^!.$$

55. Let w be a word of length nk (where $k \geq 1$) over the alphabet of L, and let $w^{(n)}$ be the corresponding word of $L^{(n)}$. Let us break w into subwords $w = w_1 \dots w_k$ each of which corresponds to a letter of the language $L^{(n)}$. It is readily seen that w has a forbidden subword u (which consists, by definition, of at most d letters) if and only if there is a subword $w' = w_p \dots w_{p+m-1}$ such that each w_i either is contained in the word u or overlaps with it, so that the number m of n-letter pieces in w' satisfies the inequality $m \leq s$, where

$$s = 2 + \left[\frac{d-2}{n}\right].$$

Thus any non-admissible word $w^{(n)}$ of $L^{(n)}$ contains a non-admissible word of length at most s, hence the language $L^{(n)}$ can be defined by a finite set of forbidden words, and the length of each forbidden word is no greater than s. This proves part a) of the problem.

b) Answer: not always.

Let us prove that the lengths of forbidden words of $L^{(n)}$ are less than d if $d \geq 3$ and $n \geq 2$; in particular, this language is not d-defined, which gives a negative answer to b). It suffices to prove the inequality $s < d$, or

$$2 + \frac{d-2}{n} < d.$$

The last inequality is clearly equivalent to inequality $(d-2)(1-1/n) > 0$, which is obvious under the given constrictions on d and n.

c) Answer: $n = d-1$.

By the above, the language $L^{(n)}$ is either quadratic or free (i. e., the lengths of forbidden words do not exceed 2) under assumption $s \le 2$, which is equivalent to inequality $2 + \frac{d-2}{n} < 3$, or $n > d-2$, i. e., $n \ge d-1$. On the other hand, if $n \le d-2$, then there exists a d-defined language L such that the language $L^{(n)}$ has forbidden words of more than three letters: for example, we can take the language L over the three-letter alphabet $\{a, b, c\}$ with one forbidden word abc^{d-2}.

56. See the solution of Problem 58.

57. Answer: yes.

For example, let A be an alphabet of $n \ge 2$ letters. Consider the language $L = F_A \cdot F_A^!$. Since F_A has exponential growth (in the statement of Problem 55 c) we can take $c_1 = n+1$ and $c_2 = n$), and since $2F_A(x) \ge L(x) \ge F_A(x)$, the language L also has exponential growth. By Problem 53 c), we have $L^! = (F_A^!)^! \cdot F_A^! = L$, thus both L and $L^!$ have exponential growth.

58. First we prove the following assertion (it is not necessary for solving Problem 56 only).

Lemma. Let $a = \{a_0, a_1, a_2, \dots\}$ be a sequence such that $a_0 = 1$ and the inequalities $a_1 \ge 2, \dots, a_N \ge 2$ for some positive integer N. Then the sequence a has polynomial (respectively, exponential) growth if and only if the corresponding inequalities in assertions b) and) hold for all a_k with $k \ge N$.

Proof of the Lemma. Let $M = \max\limits_{i \le N}\{a_i\}$. It is clear that if, for some polynomials p, q of degree d, the inequalities $p(k) \ge a_k \ge q(k)$ hold for $k \ge N$, the inequalities $p(k) + M \ge a_k \ge q(k) - M$ hold for all k. The Lemma for the case of polynomial growth follows. Similarly, if $c_1^k \ge a_k \ge c_2^k$ for $k \ge N$, then $(M + c_1)^k \ge a_k \ge g^k$ for all k, which completes the proof of the Lemma.

Let us pass to the solution of the problem. Clearly, to get all admissible words of length $\ge d-1$, we can do the following. We start with the word at a vertex of the graph. Then we go along a path that starts at this vertex, and each time we read a letter on an edge that we pass, we add this letter to the right of our word. Clearly, different words are obtain from different paths. It is readily seen that the language is finite if and only if no path returns to the initial vertex, that is, the graph has no cycles (it proves assertion a)). It remains to consider the case when the language is finite and there is a cycle in the graph. In this case the number a_j of words of length $j \ge d$ is equal to the number of paths of length $j - d + 1$.

Assume that there are two intersecting cycles; let their lengths be d_1 and d_2, and let v be their common vertex such that the edges issuing from it when we go along the two cycles are distinct (and correspond, say, to letters x and y). The words that we read on edges when we walk by paths of length k that start at v and go along each of the cycles are distinct, hence $a_k \ge 2$ for all $k \ge 0$. Moreover, for each $j = (d-1) + q(d_1 + d_2) + r$, where $r < d_1 + d_2$ is the remainder of division of $j - d + 1$ by $d_1 + d_2$, there exist at least 2^q distinct paths of length $j - d + 1$ (on each of q steps we go along both cycles in an arbitrary order, and then make r steps in an arbitrary cycle), thus for $j \ge 2d$ we have $a_j \ge 2^q = 2^{\left[\frac{j-d+1}{d_1+d_2}\right]} \ge c^j$, where $c = 2^{1/2(d_1+d_2)}$. Since always $a_j \le n^j$, the Lemma (for $N = 2g$) implies that the growth is exponential.

It remains to consider the case when the graph Γ_L has cycles, but they do not intersect each other. It suffices to verify the polynomiality condition for the number $b_k = a_{k+d-1}$ of paths of length k in the graph Γ_L (since if the corresponding inequalities hold for b_k, then they also hold for a_k for $k \ge d-1$ after the polynomials $p(x)$ and $q(x)$ are replaced with polynomials of the same degree $p_1(x) = p(x + d - 1)$ and $q_1(x) = q(x + d - 1)$). We will prove that each term of the sequence b_k is equal to the value of some polynomial $b(k)$ with positive highest coefficient (we say that such sequences are *polynomial*).

Let us consider another graph Γ'_L, whose vertices are the cycles of Γ_L and those vertices of Γ_L that belong to no cycle (the latter will be referred to as *isolated* vertices), and whose edges correspond to the edges that connect the corresponding components (cycles or isolated vertices) of Γ_L. It is clear that the graph Γ'_L has no cycles, i. e., the set of paths in it is finite. Let Q^v be the set of paths in Γ'_L that start at a given vertex v, and let q_k^v be the set of the corresponding paths of length k in Γ_L. Since $b_k = \sum_v q_k^v$, it suffices to show that the sequence $\{q_k^v\}$ is polynomial for each vertex v. We proceed by induction on the length $D = D(v)$ of a maximal path that starts at v. If $D = 0$, then either $q_k^v = 0$ for $k > 0$ (if v is an isolated vertex), or $q_k^v = 1$ for all k (if v is a cycle), thus the corresponding sequence is always polynomial. Let now v be an initial vertex of Γ'_L, from which r arrows $a_1, \dots, a_r$ issue to vertices $v_1, \dots, v_r$ (which are not necessarily distinct). By induction, we assume that $q_k^{v_i} = b_i(k)$ is a polynomial with positive highest coefficient. If v is an isolated vertex, then $q_k^v = \sum_{i=1}^r q_{k-1}^{v_i}$, hence this sequence is polynomial as the sum of polynomial sequences. On the other way, if v is a cycle, then, before we pass along one of the edges $a_1, \dots, a_r$ a word of any length is possible in the cycle, hence $q_k^v = \sum_{i=1}^r \left(\sum_{j=1}^k q_{k-j}^{v_i}\right) = \sum_{i=1}^r \left(\sum_{j=1}^k b_i(k-j)\right)$ is the sum of polynomials with positive highest coefficients. This completes the proof.

Note. It is possible to define the growth of any regular set in a similar way. To this end, the corresponding finite automaton is used.

59. Let M be the set of admissible words. Assume that the language is d-defined. Let us prove that each word is M-equivalent to a word of no more than d letters.

Indeed if a word v is non-admissible, then, for any word w, the word vw is also non-admissible. Hence all non-admissible words are equivalent. in particular, any of them is equivalent to a forbidden word, which is of length at most d.

Assume that u is an admissible word of length greater than d. By v denote the subword of u consisting of its last d letters. Let w be an arbitrary word. If the word uw has a forbidden subword, then this subword is contained in vw, since the length of any forbidden word is at most d. Hence the words u and v are equivalent.

Let the alphabet have k letters. Then the number of words of length at most d is no greater than $(k+1)^d$. Let $n = (k+1)^d + 1$. In any set of n words there exist two words that are M-equivalent to the same word of length at most d and, thus, to each other. Therefore, the set of admissible words is regular.

60. a) Let S be a maximal set of words in which no two words are M-equivalent. Then any other word is equivalent to one of S. Let us construct a finite automaton. Take S as the set of vertices of the graph. For all $s \in S$, $a \in A$, we draw an arrow marked by a from s to the vertex that is M-equivalent to sa. We say that the vertex which is M-equivalent to the empty word is the initial vertex of the automaton, and all elements of S which belong to M are the approving vertices of the automaton. It is easy to see that the automaton approves a word if and only if it belongs to M.
b) Any word determines a path along arrows of the finite automaton. Clearly, if to words determine paths ending at the same vertex, then these words are M-equivalent, where M is the set of all words approved by the automaton. Hence the number n in the definition of regular set can be taken to be one plus the number of the vertices of the automaton.

61. Consider a finite automaton (Γ, v_0, W) which approves the set M. For each vertex v of Γ, denote the set of words for which the corresponding paths in Γ end at v by T_v.

Further, for each vertex v and each letter a denote by $U(v,a)$ the set of all vertices u of Γ such that there is an arrow marked by a from u to v. Then the following relations hold:

$$T_{v_0}(x) = 1 + \sum_{a\in A} \sum_{u \in U(v_0,a)} xT_u(x) \tag{1}$$

and

$$T_v(x) = \sum_{a\in A} \sum_{u \in U(v,a)} xT_u(x) \tag{2}$$

for $v \neq v_0$.

Let us number the vertices of Γ, starting with v_0: $V = \{v_0, v_1, v_2 \dots, v_k\}$. Note that each of relations (1), (2) can be viewed as an equation of the form

$$(1 + xP_i(x))T_{v_i}(x) = \sum_{j\neq i} xQ_{ij}(x)T_{v_j}(x) + R_i(x), \tag{3}$$

where $P_i(x), Q_{ij}(x), R_j(x)$ are some given polynomials, with respect to unknown series $T_{v_0}(x), \dots, T_{v_k}(x)$.

Let us try to solve equations (3). We use the last equation to express $T_{v_k}(x)$ in terms of the rest series,

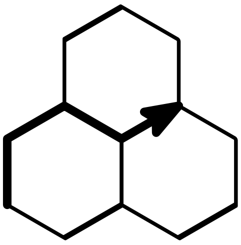
$$T_{v_k}(x) = \sum_{j\neq k} x\frac{Q_{kj}(x)}{(1+xP_k(x))}T_{v_j}(x) + \frac{R_k(x)}{(1+xP_k(x))},$$

substitute this expression instead of $T_{v_k}(x)$ into the remaining equations, and multiply them by $(1+xP_k(x))$. Thus we obtain equations of the same form, but their number (which is also the number of unknowns) decreases by one. By doing the same for $T_{v_{k-1}}(x)$, $T_{v_{k-2}}(x)$, etc., we obtain at last an expression of $T_{v_0}(x)$ as a quotient of two polynomials. By substituting it into the expression for $T_{v_1}(x)$, we find that this series is also a quotient of two polynomials. In this way we obtain the same for all $T_{v_i}(x)$. It remains to note that $M(x) = \sum_{v\in W} T_v(x)$.

62. For each word v, denote by v^{opp} the word consisting of the same letters in the opposite order. For any set of words M, we write $M^{opp} = \{v^{opp} \mid v \in M\}$. Clearly, $M^{opp}(x) = M(x)$ for any M. If L is a language whose set of forbidden words is B, then by L^{opp} we denote the language whose set of forbidden words is B^{opp}.

Let us return to our problem. It is clear that the set M_w^{opp} consists of all admissible words of L^{opp} which start with a subword equal to w^{opp}. This set is regular (the proof is similar to the solution of problem 59). Hence the series $M_w(x) = M_w^{opp}$ can be represented as a quotient of two polynomials.

The set M_w is also regular, but the proof of this fact would take more place.



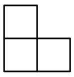
Замощения, раскраски и плиточные группы

А.Я.Белов-Канель, И.Иванов-Погодаев, А.Малистов, И.Митрофанов, М.Харитонов

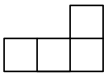
Задачи замощения очень часто становятся в центре различных сюжетов в математике. Очень часто такие задачи решаются с помощью раскрасок. Одной из целей настоящего проекта является изучение более мощного метода, связанного с применением понятий теории групп. Попутно мы изучим, чем по сути, является раскраска в новых терминах, а также рассмотрим вопросы, как замощения могут быть полезны в самой теории групп.

Первый цикл является предварительным, это несколько задач на замощение. Второй и третий циклы являются подготовительными, тут мы разрабатываем необходимую технику нового метода. Во втором вводятся некоторые полезные понятия и изучается связь слов и путей на графах. В третьем мы вводим основные понятия теории групп. В четвертом цикле мы применяем новую технику к плиточным замощениям.

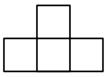
Цикл A. Замощения и раскраски



Уголки из 3 клеток



L-тетрамино



T-тетрамино

♦ **A1.** *При каких M и N прямоугольник $M\times N$ можно разбить на уголки из трех клеток?*

♦ **A2.** *При каких M, N и P прямоугольник $M\times N$ можно разбить на прямоугольники $P\times 1$?*

♦ **A3.** *При каких M, N и P прямоугольник $M\times N$ с одной добавленной клеткой можно разбить на прямоугольники $P\times 1$?*

♦ **A4.** *При каких M, N, A, B прямоугольник $M\times N$ можно разбить на прямоугольники $A\times B$?*

♦ **A5.** *При каких M и N прямоугольник $M\times N$ можно разбить на L-тетрамино?*

Следующая задача не решается с помощью методов раскраски, для ее решения нужно разработать более глубокие методы. Этим мы займемся в циклах B, C, D.

♦ **A6**.** *Докажите, что если прямоугольник $M\times N$ можно замостить с помощью T-тетрамино, то M и N делятся на* 4 .

Цикл B. Подготовительный материал: слова, графы и пути

Пусть есть некоторый конечный алфавит A: множество букв $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, **. . .** Из букв алфавита можно составлять слова — конечные последовательности букв, например, $\boldsymbol{abbc}$, $\boldsymbol{c}$, $\boldsymbol{cabcabccbb}$ и т.п. Будем называть *произведением* слов A и B слово, получающееся приписыванием слова B справа к слову A. Таким образом, алфавит задает множество слов, с определенной на нем операцией произведения.

N-ой степенью слова X называется слово, получающееся выписыванием слова X N раз подряд.

♦ **B1.** *Докажите, что если произведение слов U и V совпадает с произведением слов V и U, то и U, и V представляются в виде степеней какого-то слова A.*

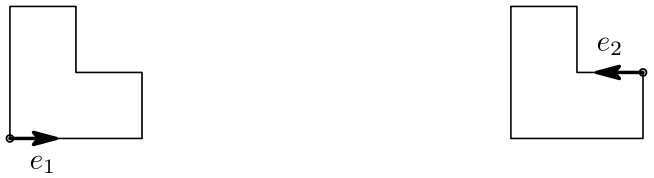
Полугруппой будем называть множество элементов с заданной на нем операцией произведения $*$ (результатом применения операции для двух элементов является элемент из того же множества) и выполненным для любых элементов $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ свойством $\boldsymbol{a*(b*c)=(a*b)*c}$. Это свойство называется свойством ассоциативности.

Примеры. Множество натуральных чисел образует полугруппу, относительно сложения. Множество рациональных чисел образует полугруппу относительно умножения, а относительно деления — нет, так как на ноль делить нельзя.

Задаваемое алфавитом множество слов является полугруппой относительно операции приписывания одного слова к другому. (Ассоциативность, очевидно, соблюдается.) Более точно, такое множество называется *свободной* полугруппой.

♦ **В2.** *Петя и Вася составляют различные слова используя буквы* $\boldsymbol{a, b, c, d, e, f}$. *В любом слове можно вычеркнуть любую из трех пар рядом стоящих букв (в любом порядке):* $\boldsymbol{a}$ *и* $\boldsymbol{b}$, $\boldsymbol{c}$ *и* $\boldsymbol{d}$, $\boldsymbol{e}$ *и* $\boldsymbol{f}$. *Также можно вставить любую из этих пар букв в любом месте слова. Например, слово* ***dacdbeaaf*** *можно преобразовать следующим образом:* ***dacdbeaaf*** $\rightarrow$ ***dabeaaf*** $\rightarrow$ ***deaaf*** $\rightarrow$ ***dfeeaaf***. *Докажите, что каждое слово можно такими операциями привести к виду, содержащему минимальное число букв, и этот вид не зависит от того, какие операции и в каком порядке применялись.*

Рассмотрим произвольное полимино. Расставим стрелки по его контуру, чтобы получился обход по часовой стрелке или против часовой стрелки. Сопоставим каждому полимино слово (последовательность букв), так чтобы стрелкам «вверх», «вниз», «вправо», «влево» отвечали буквы $\boldsymbol{U}$, $\boldsymbol{D}$, $\boldsymbol{R}$, $\boldsymbol{L}$ соответственно. При этом последовательность букв в слове должна соответствовать последовательности стрелок на контуре. Таким образом, каждому полимино будет соответствовать несколько слов, получающихся друг из друга с точностью до циклического сдвига.



RRULULDD *LULDDRRU*

Рисунок 1.

♦ **В3.** *Опишите множества слов, соответствующие домино из двух клеток и тримино-уголку из трех клеток и различным положениям этих плиток на клеточной плоскости.*

♦ **В4.** *Пусть* $\boldsymbol{M}$ *— множество слов, соответствующее различным положениям домино из двух клеток. Пусть с каждым словом из* $\boldsymbol{M}$ *можно проводить следующие операции:*

1. вставить или убрать рядом стоящие пары $\boldsymbol{R}$ *и* $\boldsymbol{L}$ *или* $\boldsymbol{U}$ *и* $\boldsymbol{D}$ *(как в задаче В2);*

2. приписать в начало слова букву $\boldsymbol{R}$, *а в конец — букву* $\boldsymbol{L}$, *или наоборот;*

3. приписать в начало слова букву $\boldsymbol{U}$, *а в конец — букву* $\boldsymbol{D}$, *или наоборот;*

4. приписывать друг к другу слова, получающиеся при таких преобразованиях.

Докажите, что тогда контур фигуры, разбиваемой на домино, можно получить при помощи этих преобразований.

Можно заметить, что: операция **1** соответствует добавлению или удалению пары путей «туда-обратно», операции **2** и **3** соответствуют возможности заменить слово на его циклический сдвиг, а операция **4** соответствует прикладыванию полимино друг к другу (с учетом того, что можно убрать куски путей «туда-обратно»). Это позволяет сформулировать необходимое условие замощения в терминах произведений слов.

♦ **В5.** *Пусть возможно замощение заданной конечной области набором полимино* $\boldsymbol{T}$. *Рассмотрим начальный набор слов, соответствующих набору* $\boldsymbol{T}$. *Докажите, что тогда слово, соответствующее контуру области, представляется как результат применения к этому начальному набору нескольких операций произведения и замены слова на его циклический сдвиг.*

♦ **В6.** *Приведите пример, когда слово контура получается в результате проведения операций из* $\boldsymbol{B5}$, *но замощение невозможно.*

Пусть теперь на плоскости есть многоугольник, разбитый на N маленьких многоугольников так, что никакая вершина многоугольника разбиения не лежит внутри стороны другого многоугольника. Рассмотрим замкнутые пути, произвольной длины, проходящие по сторонам маленьких многоугольников. Будем считать, что от добавления ребра «туда-обратно» путь не меняется. Выберем какую-нибудь вершину разбиения (вершину одного из многоугольников) — точку A. Произведением двух путей будем считать новый путь, получающийся, если сначала пройти первый «сомножитель», а потом второй.

♦ **B7.** *Рассмотрим бесконечное множество замкнутых путей начинающихся и заканчивающихся в точке A. Докажите, что есть конечный набор P замкнутых путей, начинающихся и заканчивающихся в точке A и таких, что любой замкнутый путь, проходящий через A представим в виде произведения конечного числа путей из P. Найдите минимальное число путей в таком наборе P.*

♦ **B8.** *Выберем другую вершину B, рассмотрим замкнутые пути, начинающиеся и заканчивающиеся в B. Докажите, что можно установить соответствие, такое что:*

1. Каждому пути через A будет соответствовать свой путь через B и наоборот;

2. Произведению двух путей из A будет соответствовать путь, являющийся произведением соответствующих этим сомножителям путей из B.

Цикл C. Группы

Группой будем называть множество элементов с заданной на нем операцией произведения $*$, удовлетворяющей следующим свойствам:

1. **Ассоциативность.** Для всех a, b, c выполнено $a * (b * c) = (a * b) * c$.
2. **Существование единицы.** Существует элемент e такой, что $ae = ea = a$ выполнено для всех a.
3. **Существование обратного элемента.** Для любого a существует элемент a^{-1}, такой что $aa^{-1} = a^{-1}a = e$.

Примеры. Множество целых чисел образует группу относительно сложения, обратным элементом является противоположный по знаку. Множество рациональных чисел (но без нуля) образует группу относительно умножения.

Группа подстановок. Три элемента можно расположить друг за другом шестью различными способами: 213, 321, 132, 312, 231, 123. Фактически, речь идет о преобразованиях трех элементов: можно поменять местами два из них (получится 213, 321 или 132), а можно сместить все три по циклу (получится 312 или 231). Можно также ничего не делать, тогда останется 123. То есть, у нас получается как бы шесть различных «действий». Если применить сначала одно такое действие, а потом — другое, результатом опять будет какое-то действие из этих шести. Например, если сначала поменять местами первые два элемента, а потом сместить все по кругу вправо на одну позицию, то получится 123 → 213 → 321. То есть, все равно, что поменять местами первый и третий элементы. Эти «действия» называются *подстановками.* Они образуют группу относительно операции последовательного применения. В этой группе шесть элементов, единицей является тождественное преобразование 123. Это минимальная группа, где есть элементы a и b, такие что $ab \neq ba$. Можно рассматривать группы подстановок S_n для разных натуральных n.

♦ **C0.** *Проверьте, что пути из задач B8 и B9 образуют группу.*

Группа называется *абелевой* или *коммутативной* если для любых элементов a, b выполнено $ab = ba$.

♦ **C1.** *Постройте некоммутативную группу из* 8 *элементов.*

♦ **C2.** *Докажите, что множество движений в пространстве, переводящих куб в себя, образует группу. Найдите число ее элементов.*

♦ **C3.** *Докажите, что множество слов в алфавите $\{a_1, {a_1}^{-1}, a_2, {a_2}^{-1}, \ldots, a_n, {a_n}^{-1}\}$ Образует группу относительно операции приписывания (с сокращением рядом стоящих обратных букв x и x^{-1}).*

Пусть G — группа. *Подгруппой* называется подмножество $H \in G$ элементов группы, такое, что если a, b лежат в H, то a^{-1}, b^{-1}, ba и ab тоже лежат в H. Подгруппа сама по себе тоже является группой, с той же операцией.

♦ **C4.** *Докажите, что каждый элемент в группе подстановок S_n есть произведение нескольких независимых циклов. Опишите все подгруппы группы подстановок из пяти элементов.*

Элементы $\boldsymbol{a}$ и $\boldsymbol{b}$ называются *сопряженными*, если существует такой элемент $\boldsymbol{x}$, что $\boldsymbol{xax^{-1} = b}$. Если $\boldsymbol{H}$ — подгруппа некоторой группы $\boldsymbol{G}$, то сопряженное множество $\boldsymbol{xHx^{-1}}$ также образует подгруппу $\boldsymbol{G}$, так как $\boldsymbol{xh_1x^{-1}xh_2x^{-1} = xh_1h_2x^{-1}}$.

♦ **C5.** *Найдите все сопряженные подгруппы в группе подстановок из пяти элементов.*

Часто одна и та же группа может быть описана различными способами. Например, элементами группы могут быть движения, переводящие многогранник в себя или подстановки из нескольких элементов, при этом само устройство группы будет одинаковым. Для описания «одинаковости» существует специальное понятие.

Группы $\boldsymbol{G}$ и $\boldsymbol{H}$ называются *изоморфными*, если между их элементами можно провести соответствие, обладающее следующими свойствами:

1. Для каждого элемента из $\boldsymbol{G}$ существует единственный соответствующий ему элемент из $\boldsymbol{H}$, и наоборот.

2. Если $\boldsymbol{g_1, g_2} \in \boldsymbol{G}$ соответствуют $\boldsymbol{h_1, h_2} \in \boldsymbol{H}$, то $\boldsymbol{g_1g_2}$ соответствует $\boldsymbol{h_1h_2}$.

♦ **C6.** *Найдите группу подстановок, изоморфную группе движений, сохраняющих куб.*

Пусть элементы некоторой группы раскрашены в несколько цветов так, что цвет произведения двух элементов зависит только от цветов сомножителей и не зависит от выбора элемента внутри цвета. То есть произведение элементов с цветами 1 и 2 всегда имеет один и тот же цвет, независимо от того, какие элементы с цветами 1 и 2 берутся.

♦ **C7.** *Докажите, что множество элементов с цветом, как у единицы, образует подгруппу. Такая подгруппа называется нормальной.*

Эквивалентное определение. Подгруппа $\boldsymbol{H} \in \boldsymbol{G}$ называется *нормальной*, если для любого элемента $\boldsymbol{g} \in \boldsymbol{G}$ (не обязательно принадлежащего $\boldsymbol{H}$) выполнено $\boldsymbol{gHg^{-1}} \in \boldsymbol{H}$. То есть, нормальная подгруппа сопряжена сама себе.

♦ **C8.** *Докажите эквивалентность определений.*

♦ **C9.** *Найдите все нормальные подгруппы в группе подстановок* $\boldsymbol{S_4}$.

♦ **C10.** *Рассмотрим некоторую группу* $\boldsymbol{G}$. *Пусть* $\boldsymbol{K}$ *множество элементов вида* $\boldsymbol{aba^{-1}b^{-1}}$ *(они называются коммутаторы). Пусть* $\boldsymbol{H}$ *— это всевозможные произведения элементов из* $\boldsymbol{K}$. *Докажите, что* $\boldsymbol{H}$ *является нормальной подгруппой.*

Цикл D. Плиточные группы

Множество клеток на квадратной решетке будем называть *связным*, если из любой его клетки в любую другую можно попасть, переходя из клетки в клетку по стороне.

Пусть $\boldsymbol{T}$ — множество клеток на квадратной решетке, дополнение к которому связно. Если периметр $\boldsymbol{T}$ можно обойти, не проходя по одному ребру дважды, будем называть $\boldsymbol{T}$ *плиткой.* Теперь мы будем рассматривать замощения с помощью плиток.

Рассмотрим множество слов, соответствующих конечному набору плиток. Пусть с этим множеством можно играть в игру, как в задаче B4, то есть брать слово, приписывать в любом месте взаимообратные буквы, заменять слово на его сопряженное, а также приписывать рядом уже получившиеся слова (то есть брать произведения слов).

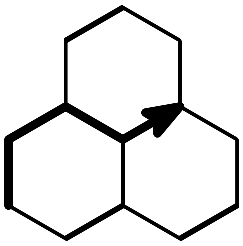
♦ **D1.** *Докажите, что полученные описанным образом слова образуют группу.*

Будем называть эту группу *плиточной группой* данного набора.

♦ **D2.** *Пусть* $\boldsymbol{C}$ *— множество замкнутых путей на квадратной решетке. Докажите, что множество соответствующих этим путям слов образует группу. Также докажите, что для любого набора плиточная группа является нормальной подгруппой в группе* $\boldsymbol{C}$.

♦ **D3.** *Докажите, что если область* $\boldsymbol{O}$ *замощается набором* $\boldsymbol{T}$, *то слово* $\boldsymbol{dO}$, *соответствующее периметру* $\boldsymbol{O}$, *лежит в плиточной группе* $\boldsymbol{T}$.

5



Замощения, раскраски и плиточные группы

♦ **E1.** Прямоугольник $m\times n$ разбит на домино. Проверьте, что какие-то две плитки образуют квадрат 2×2.

a) Каково минимальное возможное количество таких квадратиков?

b) Рассмотрим квадрат 2×2 из каких-то 2 домино. С этим квадратиком свяжем преобразование (Flip), переводящее вертикальное его разбиение в горизонтальное и наоборот. Докажите, что с помощью цепочки таких преобразований любое разбиение можно перевести в любое.

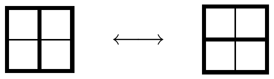


Рисунок 2.

c*) За какое минимально возможное количество таких преобразований любое разбиение можно перевести в любое?

Определение. В дальнейшем мы ещё столкнёмся с таким преобразованиями. Назовём их *флипами.* Флип — это такое преобразование, при котором выбирается кусок плоскости определённого вида и меняется расстановка плиток на нём.

♦ **E2.** Прямоугольник $m \times n$ разбит на плитки $1\times k$. Проверьте, что какие-то k из них образуют квадрат $k\times k$.

a) Каково минимальное возможное количество таких квадратиков?

b) Пусть прямоугольник разбит на плитки $1\times k$ и пусть какие-то k образуют квадрат $k\times k$. Аналогично, с этим квадратиком свяжем преобразование (Flip), переводящее вертикальное его разбиение в горизонтальное и наоборот. Докажите, что с помощью цепочки таких преобразований любое разбиение можно перевести в любое.

c*) За какое минимально возможное количество таких преобразований любое разбиение можно перевести в любое?

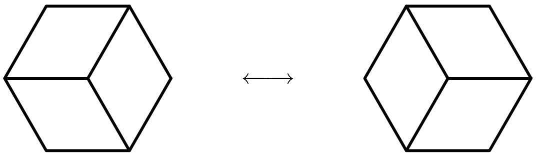


Рисунок 3.

♦ **E3.** a) Докажите, что центрально-симметричный выпуклый многоугольник разбивается на параллелограммы. Докажите обратное, т.е. что если выпуклый многоугольник разбивается на параллелограммы, то он центрально-симметричен.

b) Рассмотрим разбиение правильного $2n$-угольника с единичными сторонами на ромбы с единичными сторонами. Если какие-то три таких ромба образуют шестиугольник, то одно его разбиение можно заменить на другое (см. рис. 3), и назовем это преобразование *флипом.* Докажите, что с помощью цепочки флипов любое разбиение можно перевести в любое.

c) За какое минимально возможное количество флипов любое разбиение можно перевести в любое?

d) Какое минимальное число шестиугольников может наблюдаться при разбиении правильного $2n$-угольника с единичными сторонами на параллелограммы?

е) Рассмотрим разбиение правильного 6-угольника со стороной n на ромбы с единичными сторонами. Докажите, что с помощью цепочки флипов любое разбиение можно перевести в любое и найдите, за какое минимально возможное количество флипов любое разбиение можно перевести в любое.

Указание. Нарисуйте куб $n\times n\times n$ в виде кирпичной кладки. Что происходит с картинкой, если убирать кирпичи по одному?

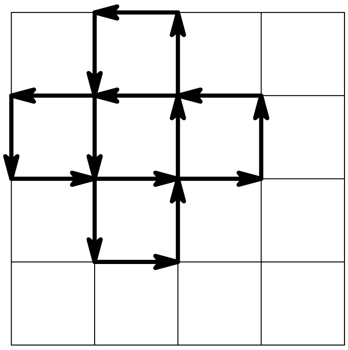


Рисунок 4.

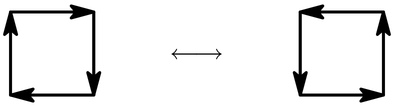


Рисунок 5.

♦ **E4.** Дана конечная прямоугольная решётка на клетчатой плоскости. У этой решётки у каждого ребра есть направление (см. рис. 4).

Условие 1. В каждую вершину входит то же количество рёбер, что и выходит.

Назовем *флипом* следующее преобразование: если все стрелки вокруг клетки направлены по часовой стрелке, поменяем их направление, и наоборот, если все стрелки вокруг клетки направлены против часовой стрелки, поменяем их направление. Докажите, что с помощью цепочки флипов (см. рис. 5) любое расположение стрелок, удовлетворяющее условию 1, можно перевести в любое расположение стрелок, удовлетворяющее условию 1.

Указание. Рассмотрите функцию $h(x)$, задаваемую следующим образом: для каждой клетки x, у которой есть ненаправленное ребро, $h(x) = 0$. Будем определять h индуктивно: если для x определена $h(x)$, а y — соседняя с x клетка по стороне, то $h(y)$ равна 0, если у клетки y есть ненаправленное ребро; 1, если при движении из x в y мы пересекаем направленное ребро, идущее слева направо относительно направления движения; -1, если при движении из x в y мы пересекаем направленное ребро, идущее справа налево относительно направления движения. Считайте известным, что эта функция определена корректно.

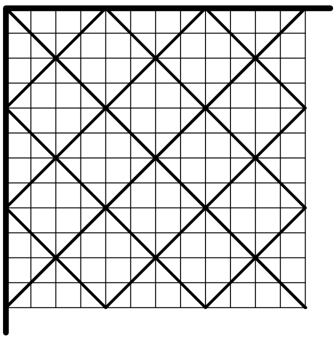


Рисунок 6.

♦ **E5.** Рассмотрим прямоугольник $m\times n$, который замостили плитками тетрамино. Рассмотрим следующую диагональную сетку прямых (см. рис. 6). Докажите, что любое ребро, проходящее по диагоналям 2 клеток, пересекает ровно одно тетрамино, а числа m и n делятся на 4.

♦ **E6.** Докажите, что с помощью флипов (см. рис. 7) любое разбиение можно перевести в любое.

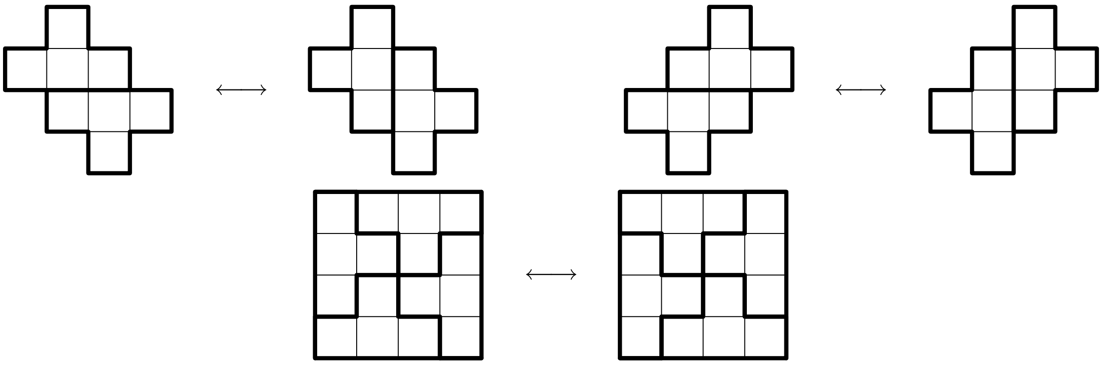


Рисунок 7.

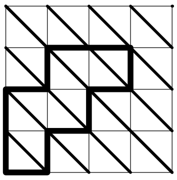


Рисунок 8.

♦ **E7.** Такие связные плитки, что любая прямая из показанных на рисунке 8 пересекает не более одной клетки плитки, назовём р-плитками. Пусть T_n — множество р-плиток, состоящих из n клеток, а $|T_n|$ — количество элементов этого множества. Найдите $|T_n|$.

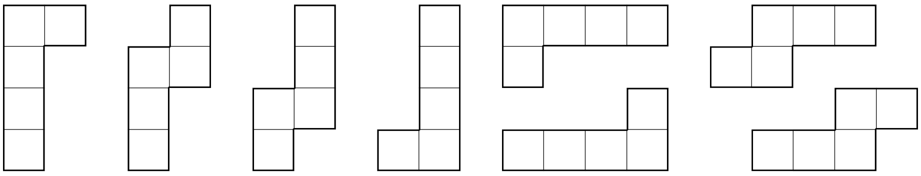


Рисунок 9.

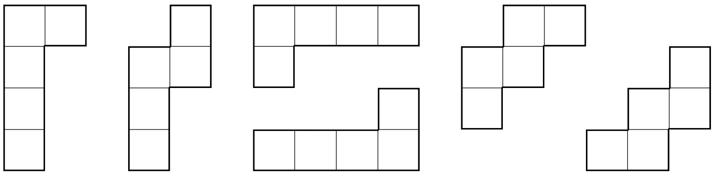


Рисунок 10.

♦ **E8.** Докажите, что если прямоугольник $a \times b$ можно покрыть р-плитками, показанными на а) рис. 9, b) рис. 10, то 10 делит ab.

♦ **E9.** Пусть D_n — фигура, указанная на рис. 11. Докажите, что если фигура D_n покрывается р-плитками с рис. 10, то $n \equiv 0, 4, 15, 19 \pmod{20}$.

♦ **E10.** Сколько способов замостить доминошками а) прямоугольник $2 \times m$? б) ацтекский бриллиант (Aztec diamond)? (см. рис. 12)

Нас интересует покрытие фигуры F в несколько слоёв плитками из некоторого набора T. Назовём *инвариантом* такую расстановку чисел в клетках, что сумма чисел, покрытых любой плиткой из набора T, делится на p. Инвариант *нетривиален*, если сумма чисел, покрытых фигурой F, не делится на p.

♦ **E11.** а) Докажите, что если нетривиальный инвариант существует, то нельзя покрыть F плитками из T так, что кратность покрытия каждой клетки сравнима с 1 по модулю p. б) Докажите, что если нетривиальных инвариантов нет, то такое покрытие существует.

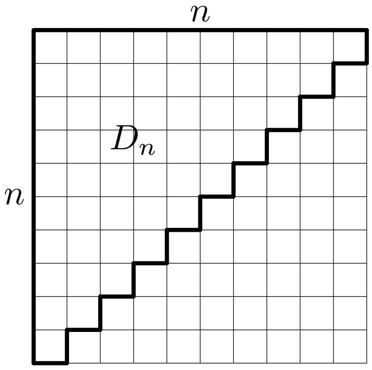


Рисунок 11.

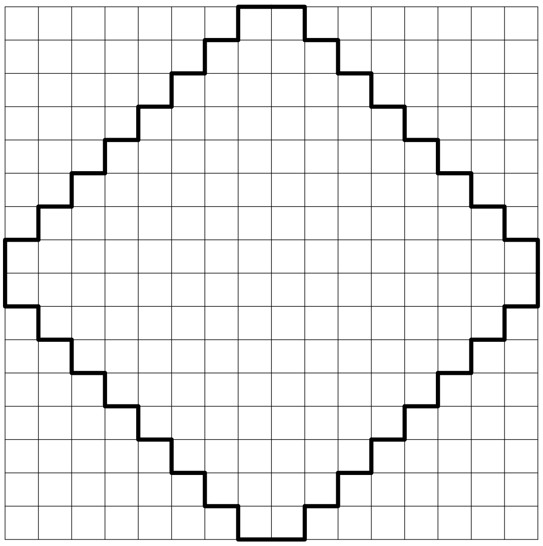


Рисунок 12.

♦ **E12.** Для каких m и n прямоугольник $m{\times}n$ можно покрыть уголками из 3 клеток так, чтобы каждая клетка была покрыта одинаковое число раз?

♦ **E13.** *Полуинвариантом* назовём такую расстановку чисел в клетках, что сумма чисел, покрытых любой плиткой из набора T, отрицательная, а сумма чисел, покрытых фигурой F, положительная. Докажите, что а) если полуинвариант существует, то покрыть фигуру F фигурами из T так, что каждая клетка покрыта одинаковое число раз, нельзя; б)если полуинвариантов нет, то такое покрытие существует.

♦ **E14.** Дано табло с лампочками и пульт с кнопками. Кнопка меняет состояние соединённых с ней лампочек на противоположное. а) Докажите, что количество возможных узоров является степенью двойки. б) Назовём *инвариантом* такой набор лампочек, что любая кнопка меняет состояние чётного числа лампочек из этого набора. Докажите, что если инвариантов нет, все лампочки можно погасить вне зависимости от начального состояния. в) Докажите, что если никакой инвариант не различает начального и конечного состояния, то можно осуществить переход из начального в конечное состояние.

♦ **E15.** Назовём набор флипов *полным*, если любое замощение области переводится любое замощение области с помощью цепочки из этого набора флипов. Существует ли набор плиток, для которого нельзя выбрать полного набора флипов?

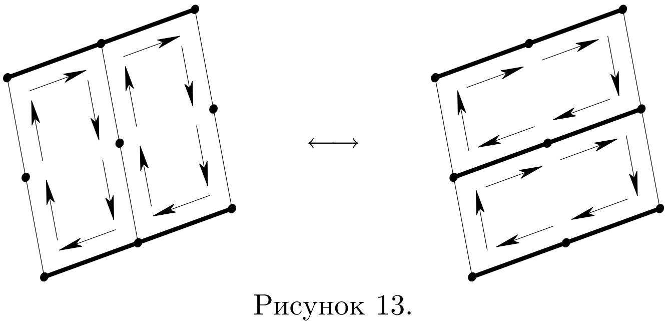


Рисунок 13.

♦ **E16*.** Плоский граф разбит на области , представляющие собой 6-угольники, внутри которых нет рёбер (можно считать их плитками домино 2×1). У каждого ребра две стороны, снабженные стрелками. Эти две стрелки имеют противоположную ориентацию. При этом каждая область ориентирована по часовой стрелке. Разрешается взять область из двух смежных 6-угольников и преобразовать её как указано на рисунке 13. Верно ли, что из любого такого разбиения плоского графа можно получить любое другое с помощью цепочки заданого набора флипов?

Замощения, раскраски и плиточные группы

♦ **F1.** Рассмотрим ориентированный граф, рёбра которого покрашены в два цвета, в каждую вершину входит красное и синее ребро, и выходят таких же цветов. Самосовмещением графа называется правило, которое сопоставляет каждой вершине графа какую-то другую вершину, при этом для любой вершины есть единственный прообраз. Пусть для любых двух вершин есть самосовмещение графа, сохраняющее ориентацию и цвета рёбер, переводящее первую вершину во вторую. Кроме того, если из вершины пройти три раза по синей стрелке, а потом — три раза по красной, то мы придём туда же, куда пришли бы, пройдя три раза по красной, а потом — три раза по синей. Докажите, что при замене числа «три» на «двадцать четыре» факт останется верным.

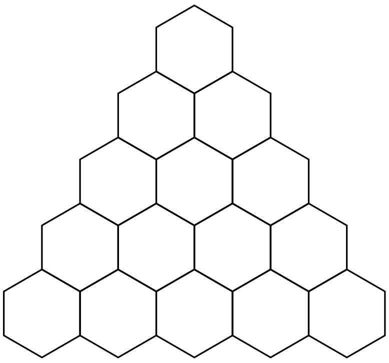


Рисунок 1. T_5

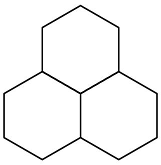


Рисунок 2. T_2

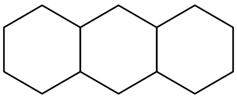


Рисунок 3. L_3

Определим область T_n как «треугольник» со стороной n, составленный из шестиугольников. Также определим L_n, как n шестиугольников, расположенных в ряд. (см. рис. 1–3)

♦ **F2.** Установите соответствие между фигурами на шестиугольной решетке и плитками на квадратной решетке. Расставьте стрелки двух цветов (a и b) на квадратной решетке так, чтобы слова, соответствующие плиткам L_n, давали замкнутые пути.

♦ **F3.** Придумайте раскраску (инвариант) новой квадратной решетки из F2, такую, что путям, соответствующим L_4, соответствуют нулевые значения, а фигурам T_n — нет.

♦ **F4.** Докажите, что T_n нельзя замостить фигурами L_3.

♦ **F5.** Найдите все значения n, при которых T_n можно замостить фигурами T_2.

Замощения, раскраски и плиточные группы

Решения цикла E

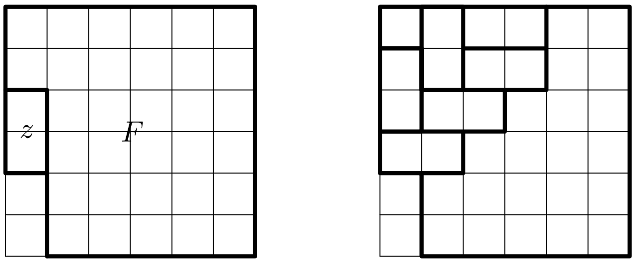


Рисунок 1.

Указание к решению E1b. Доказательство проводится по индукции. Пусть вертикальная сторона прямоугольника чётной длины. Докажем по индукции, что можно поставить все плитки домино вертикально для фигуры F на рис. 1 слева. Нам надо получить с помощью цепочки флипов плитку z. Пусть её нет и нельзя получить цепочкой флипов. Тогда имеем структуру на рис. 1 справа. Отсюда мы можем применить индукционное предположение для F без z.

Указание к решению E2b. Аналогично E1b.

Указание к решению E3. Нарисуйте куб $n{\times}n{\times}n$ в виде кирпичной кладки. Что происходит с картинкой, если убирать кирпичи по одному?

Указание к решению E4. Рассмотрите функцию $h(x)$, задаваемую следующим образом: для каждой клетки x, у которой есть ненаправленное ребро, $h(x) = 0$. Будем определять h индуктивно: если для x определена $h(x)$, а y — соседняя с x клетка по стороне, то $h(y)$ равна 0, если у клетки y есть ненаправленное ребро; 1, если при движении из x в y мы пересекаем направленное ребро, идущее слева направо относительно направления движения; -1, если при движении из x в y мы пересекаем направленное ребро, идущее справа налево относительно направления движения. Считайте известным, что эта функция определена корректно. Пусть расстановка стрелок A мажорирует расстановку стрелок B, если для любого x $h_A(x) \geqslant h_B(x)$. Рассмотрим расстановку C такую, что если один флип переводит её в расстановку D, то C не мажорирует D. Т.к. расстановок конечное число, такая C существует. Несложно доказать, что не существует такой точки x на C, что для всех точек y, соседних с ней по стороне, $h(x) \geqslant h(y)$. Тогда C определяется единственным образом. Значит, любую расстановку с помощью цепочки флипов можно перевести в любую другую.

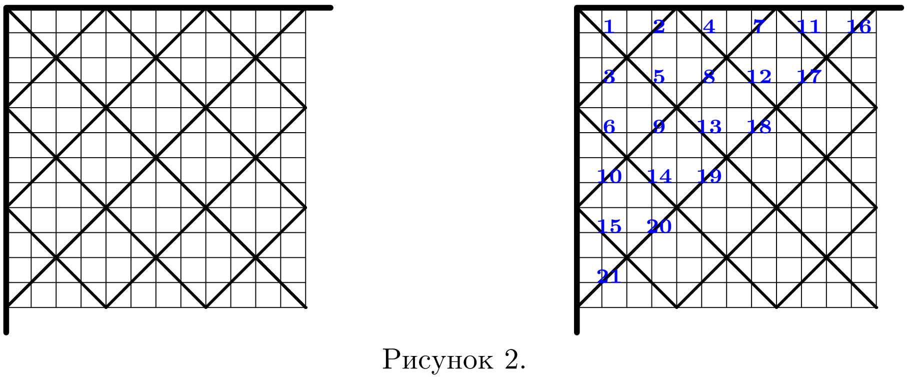


Рисунок 2.

Указание к решению E5. Пронумеруем диагонали как показано на рис. 2 справа. Пусть для всех диагоналей с номерами от 1 до n утверждение доказано. Пусть для диагонали с $(n+1)$-ым номером утверждение не верно. Перебором всех возможных случаев расположения Т-тетрамино, пересекающих $(n+1)$-ую диагональ, получаем противоречие. Как следствие доказательства получаем, что m и n делятся на 4.

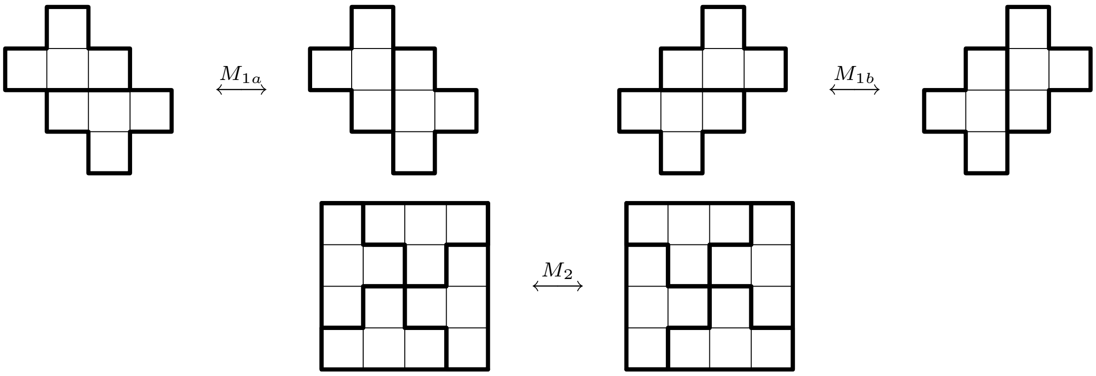


Рисунок 3.

Указание к решению E6. Рассмотрим диагональную сетку из $E5$. Из каждой диагонали направим ребро в сторону, где лежит бо́льшая часть тетрамино, содержащего эту сторону. Флипы M_{1a}, M_{1b} переводят любую расстановку Т-тетрамино A в B, если A и B соответствуют одинаковые расстановки стрелок. Для полученной сетки из направленных рёбер применим $E4$. M_2 и есть флип из задачи $E4$.

Ответ к E7: 2^{n-1}.

Ответ к E10a: $(m+1)$-ый член последовательности Фибоначчи $(1, 1, 2, 3, 5, 8, \dots)$.

Решения цикла F

Решение F1. Рассмотрим последовательность из двадцати четырех ребер a и двадцати четырех b. По условию, мы можем переставить три последних a и три первых b. Продолжая такие перестановки, мы добиваемся требуемого.

F2 и F3 являются подготовительными пунктами для F4 и F5. Мы представим указания к решению для F4 и F5.

Решение F4 и F5. Для решения задач F4 и F5 сначала представим эти задачи для квадратной решетки. Соответствующие фигуры представлены на рисунке. Область T_n при этом будет выглядеть как лестница (см. рисунок). Запишем слова для получившихся плиток. Итак, наша цель выяснить вопрос, принадлежат ли слова, соответствующие областям, в соответствующим плиточным группам (для L_3 и T_2).

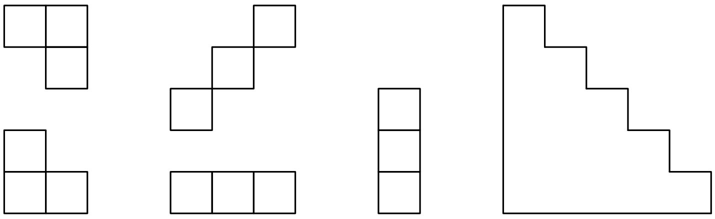


Рисунок 4.

Для этого мы воспользуемся новой идеей, предложенной Конвеем. Рассмотрим бесконечный ориентированный граф, в каждую вершину которого входит одно ребро A и одно B, и выходит также одно ребро A и одно B (см. рисунок). Непосредственно можно проверить, что словам для всех возможных положений L_3 и T_2 соответствуют замкнутые пути.

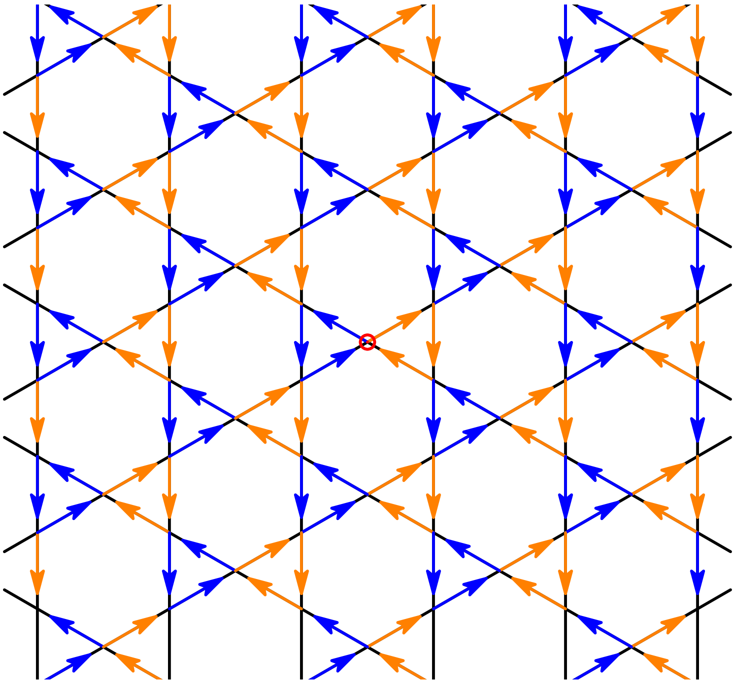
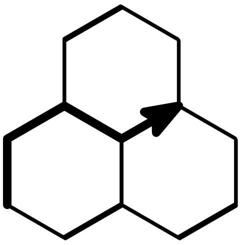


Рисунок 5.

Вычислим для путей, соответствующих положениям L_3 следующий инвариант: количество треугольных клеток, пройденных по часовой стрелке минус количество треугольных клеток, пройденных против часовой стрелки. Заметим, что это число складывается для произведений двух путей, а для L_3-путей оно равно нулю. Кроме того, инвариант не меняется при сопряжении. Значит, для всех слов из плиточной группы для L_3 этот инвариант равен нулю. В плитке T_n количество клеток должно делиться на три, в этом случае $n \equiv 0$ или $2 \pmod 3$. В этих случаях путям на графе для T_n будут соответствовать ненулевые значения инварианта. Значит, соответствующие слова не могут лежать в плиточной группе для L_3 и разбиение, указанное в задаче F4 невозможно.

Для F5 будем использовать другой инвариант: количество количество шестиугольных клеток, пройденных по часовой стрелке минус количество треугольных клеток, пройденных против часовой стрелки. Этот инвариант для различных положений T_2 (уже на квадратной решетки) будет давать значения 1 или -1. Также можно установить, что для слова T_n инвариант будет равен $\left[\frac{n+1}{3}\right]$. Допустим, что T_n разбивается на m плиток T_2. тогда $\left[\frac{n+1}{3}\right] = m \pmod 2$. Кроме того, так как в T_n ровно $n(n+1)/2$ клеток, получаем, что $m = n(n+1)/2 \pmod 2$. Следовательно, $\left[\frac{n+1}{3}\right] = \frac{(n+1)n}{2} \pmod 2$. Легко видеть, что данное соотношение не выполнено для $n \equiv 3, 5, 6$ или $9 \pmod{12}$. Учитывая, что $n \equiv 0$ или $2 \pmod 3$, осталось рассмотреть случаи $n \equiv 0, 2, 9$ или $11 \pmod{12}$. Для этих случаев T_n можно разбить на T_2. Построение этих разбиений оставляем читателю. Таким образом, разбиение, указанное в задаче F5 возможно только для $n \equiv 0, 2, 9$ или 11 (mod 12).



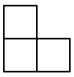
Pavements, colorings and tiling groups

**A.Belov-Kanel, I.Ivanov-Pogodaev, A.Malistov,
I.Mitrofanov, M.Kharitonov**

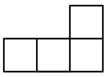
The problems of tiling often act as the centre of various mathematical matters. Very often such problems are solved with the help of coloring. One of the aims of this project is to study a more powerful method related to application of the notion of the group theory. Conjointly we will explore what, in essence, is the coloring in the new terms, and will also see how the tilings could be used in the group theory.

The first cycle is preliminary, it contains a few tiling related problems. The second and the third cycles are preparatory, during those we develop the necessary technique of the new method. During the second cycle some useful terms are introduced, and the relations between the words and the paths on graphs are studied. During the third cycle we introduce the basic terms of the group theory. In the fourth cycle we employ the new technique to the tiling problems.

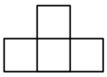
Section A. Tilings and colorings



3-cells corners



L-tetromino



T-tetromino

♦ **A1.** *Find all integers M and N such that a rectangle $M\times N$ can be cut into 3-cells corners?*

♦ **A2.** *Find all integers M, N, P such that a rectangle $M\times N$ can be cut into rectangles $P\times 1$?*

♦ **A3.** *Find all integers M, N, P such that a rectangle $M\times N$ with one additional cell can be cut into rectangles $P\times 1$?*

♦ **A4.** *Find all integers M, N, A, B such that a rectangle $M\times N$ can be cut into rectangles $A\times B$?*

♦ **A5.** *Find all integers M, N such that a rectangle $M\times N$ can be cut into L-tetromino?*

The following problem cannot be solved by colorings. We should develop some more powerful ideas to solve it. We shall do this in sections B, C, D.

♦ **A6**.** *Suppose that $M\times N$ rectangle can be tiled by T-tetramino. Prove that M and N are divisible by* 4.

Section B. Spade-work: words, graphs and paths

Let an alphabet A be a set of letters a, b, c, …. We can use these letters to arrange words which are finite sequences of these letters, for example, *abbc*, *c*, *cabcabccbb* etc. A word is called *product* of two words A and B if it is congruent with the word AB (B attached to the right of A). Hence, an alphabet determines the set of words with product operation.

A word written N times in a row $(\underbrace{XX\dots XX}_{N\ \text{times}})$ is called N-*power* of X and is denoted by X^N.

♦ **B1.** *Suppose that the product UV is congruent to the product VU. Prove that the words U and V are powers of some word A.*

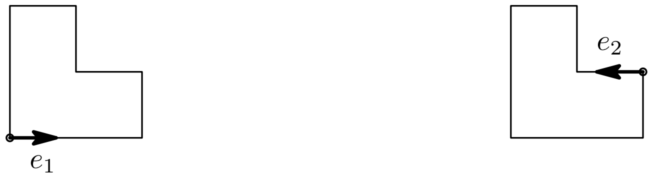
Consider a set of elements with some product operation $*$ (the product of two elements of the set also belongs to the set). This set is called a *semigroup* with respect to the $*$ operation if $\boldsymbol{a * (b * c) = (a * b) * c}$ holds for all elements $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$. This is *associative* property.

Examples. The set of integers is a semigroup with respect to the sum operation. The set of rational numbers is a semigroup with respect to the product operation. In this example we cannot change product by division because of that we cannot divide by zero.

The set of words determined by an alphabet is a semigroup with respect to the operation of attaching of one word to another. (It is clear that associative property holds.) More precisely, it is called *free* semigroup.

♦ **B2.** *Peter and John assembly various words using the letters* $\boldsymbol{a, b, c, d, e, f}$. *One can delete any of the neighboring pairs (any order)* $\boldsymbol{a}$ *and* $\boldsymbol{b}$, $\boldsymbol{c}$ *and* $\boldsymbol{d}$, $\boldsymbol{e}$ *and* $\boldsymbol{f}$ *from any word. Also, one can add any of these pairs into any part of any word. For example, we can transform the word* ***dacdbeaaf*** *by the following way:* ***dacdbeaaf*** $\longrightarrow$ ***dabeaaf*** $\longrightarrow$ ***deaaf*** $\longrightarrow$ ***dfeeaaf***. *Prove that any word can be transformed to the form containing the minimal number of letters. Prove that this form doesn't depend on a set of operations applied.*

Consider a polimino. Let us mark the edges on the boundary of this polimino with arrows placed clockwise (together) or counter-clockwise. Let us associate any polimino with the sequence of the letters on its boundary. For "up", "down", "right", "left" arrows we shall accordingly write $\boldsymbol{U}$, $\boldsymbol{D}$, $\boldsymbol{R}$, $\boldsymbol{L}$ letters. At the same time, the sequence of the letters in the word must correspond to the sequence of boundary arrows. Hence, there are several words correspond to each polimino. They can be transformed one to another by the cyclic shift.



$RRULULDD$ $LULDDRRU$

Figure 1.

♦ **B3.** *Describe the set of words corresponding to two-cells domino and three-cells corner trimino. (You should take into account various placements of the poliminoes on the plane.)*

♦ **B4.** *Let* $\boldsymbol{M}$ *be the set of words corresponding to the various placements of the two-cells domino on the plane. Suppose that we can apply the following operations to any word in* $\boldsymbol{M}$:
1. Add or delete neighboring pairs $\boldsymbol{R}$ *and* $\boldsymbol{L}$ *или* $\boldsymbol{U}$ *and* $\boldsymbol{D}$ *(in any order, as in B2);*
2. Add $\boldsymbol{R}$ *letter to the beginning of the word and* $\boldsymbol{L}$ *letter to the end of the word (and vice versa);*
3. Add $\boldsymbol{U}$ *letter to the beginning of the word and* $\boldsymbol{D}$ *letter to the end of the word (and vice versa);*
4. Attach obtained words to each other.

Let some region can be tiled by dominoes. Prove that the word corresponding to the boundary can be obtained by complex of these operations.

It is easy to see that: operation 1 corresponds to adding or deleting of pair of «there and back again» paths; operation 2 and 3 corresponds to the changing word by its cyclic shift; operation 4 corresponds to the attaching words to each other (using the fact that we can delete «there and back again» paths). Thus, we can formulate the necessary tiling condition using our language of word products.

♦ **B5.** *Suppose that we can tile the fixed finite region using the set of poliminoes* $\boldsymbol{T}$. *Consider the original set of words corresponding to* $\boldsymbol{T}$. *Prove that the word corresponding to the boundary is presented by the result of the several product and cyclic-shift operations.*

♦ **B6.** *Construct an example of the region which boundary can be obtained by the operations of* $\boldsymbol{B5}$, *but tiling cannot be done.*

Consider a polygon on the plane. Suppose that it is cut on N small polygons such that none of vertices of the small polygons belong to the inner part of other polygon edge. Consider closed paths of arbitrary length passing through the small polygons edges. Suppose that adding (or delete) the "there and back" edges doesn't change any path. Let point A be some vertex of a small polygon. If the path can be presented by pass through one path then do through the second one, then we say that our path is a product of two passed paths.

♦ **B7.** *Consider an infinite set of closed paths that start at point A. Prove that there exists a finite set P of closed paths from A such that any closed path starting at A can be presented by the finite product of the paths from P. Find the minimal number of paths in the set P.*

♦ **B8.** *Choose another vertex B. Consider closed paths starting at B. Prove that we can provide a correspondence such that the following properties hold:*

1. Every path starting at A corresponds to its own path staring at B.

2. If paths p_1 and p_2 starting at A correspond to the paths p'_1 and p'_2, then product p_1p_2 corresponds to the path $p'_1p'_2$.

Section C. Groups

A set with respect to the product operation $*$ is called *a group* if the following conditions hold:

1. Associative property. For any a, b, c holds $a * (b * c) = (a * b) * c$.
2. The existence of the unit element. There exists an element e such that the equality $ae = ea = a$ holds for any a in the group.
3. The existence of the inverse element. For any a in the group there exists an element a^{-1} such that $aa^{-1} = a^{-1}a = e$.

Examples. The set of integers is a group with respect to the sum operation. An opposite number is an inverse element. The set of rational numbers (without zero) is a group with respect to the product operation.

Substitution groups. There are six ways to place three elements in the row: 213, 321, 132, 312, 231, 123. In fact, we consider the transformations of the three elements: we can interchange two of them (so we obtain 213, 321 or 132), or we can shift all three elements by the cycle (so we obtain 312 or 231). Also we can do nothing so we stay with 123 in this case. It is easy to see that we have six different «actions». If we apply one of these actions and immediately another one, then we obtain the action from the six ones again. For example, if we interchange the first two elements and shift all three elements by the cycle to the right, then we obtain 123 → 213 → 321. Thus, we obtain the action of interchanging of the first and third elements. These «actions» are called *substitutions*. The set of substitutions is a group with respect to the consecutive applying operation. There are six elements in this group. The identical transformation 123 is the unit element. This group is the smallest group such that there holds $ab \neq ba$ for some a and b. We denote it by S_3. We can consider substitution group S_n for any integer n.

♦ **C0.** *Check that paths in the B8 and B9 form a group.*

A group is called *abelian* or *commutative* if $ab = ba$ holds for any a and b.

♦ **C1.** *Construct a nonabelian group consisting of 8 elements.*

♦ **C2.** *Prove that the set of motions in space that transfer the cube to itself is a group. Find the number of elements in this group.*

♦ **C3.** *Prove that the set of words in the alphabet $\{a_1, {a_1}^{-1}, a_2, {a_2}^{-1}, \ldots, a_n, {a_n}^{-1}\}$ is a group with respect to the attaching operation. (Also we can use cancellation of neighboring inverse letters x and x^{-1}.)*

Let G be a group. Suppose that H is a subset of G and if $a, b \in H$, then $a^{-1}, b^{-1}, ba, ab \in H$. Then we say that H is *a subgroup* of G. H is a group too, with respect to the same operation.

♦ **C4.** *Prove that every element in the substitution group S_n is a product of several independent cycles. Describe all subgroups in the substitution group S_5.*

Elements a and b are called *conjugate* if there exists an element x, such that $xax^{-1} = b$. If H is a subgroup, then *conjugate set* xHx^{-1} is also a subgroup, because of $xh_1x^{-1}xh_2x^{-1} = xh_1h_2x^{-1}$.

♦ **C5.** *Find all conjugate subgroups in the S_5 substitution group.*

A group often can be described by different methods. For example, elements of a group can be presented by the motions in space transferring polyhedron to itself, or by substitutions of several elements. The structure of a group be the same. There is a special notion to describe the «sameness» idea.

Groups $\boldsymbol{G}$ and $\boldsymbol{H}$ are called *isomorphic* if there exists a correspondence between $\boldsymbol{G}$ and $\boldsymbol{H}$ with the following properties:

1. There is an unique element $\boldsymbol{h} \in \boldsymbol{H}$ corresponding to any element $\boldsymbol{g} \in \boldsymbol{G}$, and vice versa;

2. If elements $\boldsymbol{h_1}, \boldsymbol{h_2} \in \boldsymbol{H}$ correspond to the elements $\boldsymbol{g_1}, \boldsymbol{g_2} \in \boldsymbol{G}$, then product $\boldsymbol{h_1 h_2}$ corresponds to $\boldsymbol{g_1 g_2}$.

♦ **C6.** *Suppose that $\boldsymbol{G}$ is a group of motions transferring the fixed cube to itself. Find a substitution group which isomorphic to $\boldsymbol{G}$.*

Suppose that the elements of a group are colored in several colors such that the color of product depends on colors of the elements only and not depends on the choose of elements of that color. Hence, the product of elements colored in 1 and 2 colors always has the same color (we can take any elements with color 1 and 2).

♦ **C7.** *Prove that the set of all elements having the same color as unit element is a subgroup. This subgroup is called* normal.

Equivalent definition. A subgroup $\boldsymbol{H} \in \boldsymbol{G}$ is called *normal* if for any $\boldsymbol{g} \in \boldsymbol{G}$ ($\boldsymbol{g}$ may be not in $\boldsymbol{H}$) the following holds: $\boldsymbol{gHg^{-1}} \in \boldsymbol{H}$. In other words, a normal subgroup is conjugate to itself.

♦ **C8.** *Prove this equivalence of the definitions.*

♦ **C9.** *Find all normal subgroups in the substitution group $\boldsymbol{S_4}$.*

♦ **C10.** *Consider a group $\boldsymbol{G}$. Let $\boldsymbol{K}$ be a set of elements $\boldsymbol{aba^{-1}b^{-1}}$. (they are called commutators). Suppose that $\boldsymbol{H}$ is the set of all possible products of elements in $\boldsymbol{K}$. Prove that $\boldsymbol{H}$ is a normal subgroup.*

Section D. Tiling groups

A set of cells on the square lattice is called *connected* if we can walk from any cell to another by passing thought cell edges.

Let $\boldsymbol{T}$ be a set of cells on the square lattice. Suppose that $\boldsymbol{T}$ have connected complement set. If we can walk through the boundary of $\boldsymbol{T}$ passing all the edges one time only, then we say that $\boldsymbol{T}$ is a *tile*. Now we shall consider the pavements with tiles.

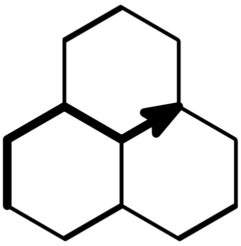
Consider a set of words corresponding to the finite tile set. Let us play with this set the game as in B4 problem: add or delete pairs of inverse letters, or change word by its conjugate, or attach words to each other (take words products).

♦ **D1.** *Prove that the set of all words that we obtain by such operations is a group with respect to the product.*

This group is called *a tiling group* of the given tile set.

♦ **D2.** *Let $\boldsymbol{C}$ be a set of all closed paths on the square lattice. Prove that the set of words corresponding to these paths is a group. Prove that for any tile set a tiling group is a normal subgroup in $\boldsymbol{C}$ group.*

♦ **D3.** *Suppose that region $\boldsymbol{O}$ can be tiled by tile set $\boldsymbol{T}$. Prove that the word $\boldsymbol{dO}$ corresponding to the boundary of $\boldsymbol{O}$ is in the tiling group of $\boldsymbol{T}$.*



Pavements, colorings and tiling groups

♦ **E1.** Consider a rectangle $m{\times}n$ tiled by dominoes. Check that there exists a subsquare $2{\times}2$ consisting of two dominoes.

a) Find the minimal number of these subsquares.

b) Consider a square $2{\times}2$ tiled by 2 dominoes. Let us define a *flip*: it is a transformation of a horizontal tiling to a vertical one and vice versa. Prove that for any tiling of an $m{\times}n$ rectangle there exists a sequence of such flips such that we can obtain any other tiling.

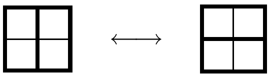


Figure 2.

c*) Find the minimal number of flips which is sufficient for transforming of any tiling to any other one.

Definition. Below we shall examine these flip transformations. To make a flip, we choose the region in the plane of fixed form and change the tiling into it.

♦ **E2.** Consider a rectangle $m{\times}n$ tiled by $1{\times}k$ tiles. Check that there exists a square $k{\times}k$ consisting of k $1{\times}k$ tiles.

a) Find the minimal number of these squares.

b) Consider a rectangle $k{\times}k$ tiled by $1{\times}k$ tiles. Similarly, a *flip* is a transformation of a horizontal tiling to a vertical one and vice versa. Prove that for any tiling of an $m{\times}n$ rectangle there exists a sequence of such flips such that we can obtain any other tiling.

c*) Find the minimal number of flips which is sufficient for transforming of any tiling to any other one.

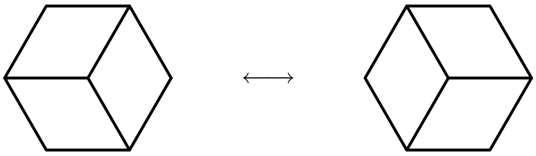


Figure 3.

♦ **E3.** a) Prove that a central-symmetric convex polygon can be tiled by parallelograms. Also prove the converse fact that if a convex polygon can be tiled by parallelograms then it is central-symmetric.

b) Consider a tiling of the regular $2n$-polygon (with unit edges) by rhombuses with unit edges. If there exist three such rhombuses forming a hexagon then we can change this tiling to another one (see picture 3). We shall call this transformation a *flip*. Prove that it's possible to obtain any tiling by some sequence of flips.

c) Find the minimal number of flips sufficient for transforming of any tiling to any other one.

d) Find the minimal number of hexagons used in tiling of a regular $2n$-polygon (with unit edges) by parallelograms.

e) Consider a tiling of the regular hexagon (with edges of length n) by rhombuses with unit edges. Prove that it is possible to obtain any tiling by some sequence of flips. Find the minimal number of flips sufficient for transforming of any tiling to any other one.

Note. Draw a cube $n{\times}n{\times}n$ using the picture of brick arrangement. How does the picture change when we remove the bricks one by one?

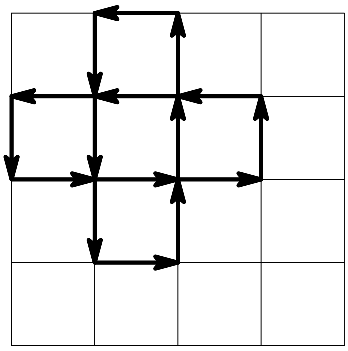


Figure 4.

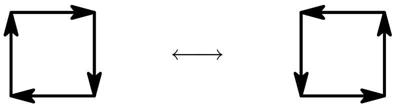


Figure 5.

♦ **E4.** Consider a finite rectangle grid on the square lattice. Suppose that any edge of this grid is oriented (see picture 4).

Condition 1. For any vertex, there are equal numbers of ingoing and outgoing edges.

Let us call by *flip* the following transformation. If all the arrows around the cell are oriented clockwise, then we change the directions of all arrows, and vice versa, if all the arrows are oriented counterclockwise, we do the same. Suppose that condition 1 holds for the placement of the arrows. Prove that this placement can be transformed by flips into any other placement satisfied condition 1.

Note. Consider the function $h(x)$, with the following condition hold: if cell x contains nondirected edge, then $h(x) = 0$. Let us define h by induction: if $h(x)$ is defined for some x, and y is a neighboring cell (by edge), then $h(y)$ is defined in the following way: $h(y) = 0$, if the cell y contains nondirected edge, $h(y) = 1$, if we intersect oriented edge which goes from the left to the right, and $h(y) = -1$, if we intersect directed edge which goes from the right to the left. You may assume that this function is well defined.

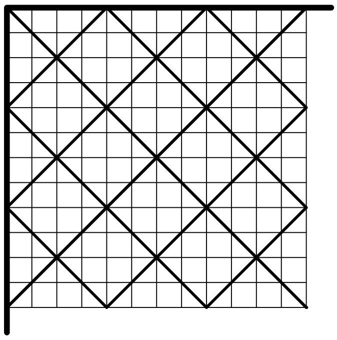


Figure 6.

♦ **E5.** Consider a rectangle $m{\times}n$ tiled by tetraminoes. Further consider the following diagonal net of lines (see the figure 6). Prove that each edge containing diagonals of two cells meets just one tetramino, and that m and n are divisible by 4.

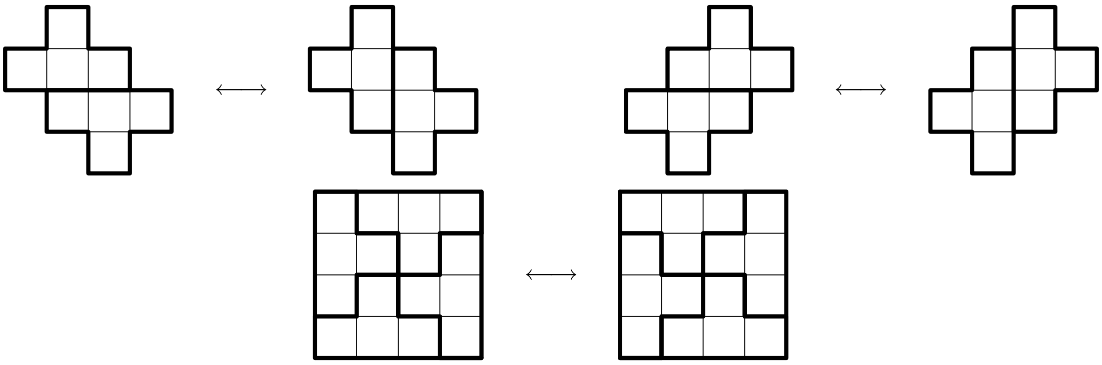


Figure 7.

⧫ **E6.** Prove that by sequences of flips (see the figure 7) each dissection can be transferred into each other.

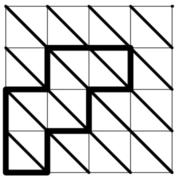


Figure 8.

⧫ **E7.** Connected tiles such that each line indicated in the figure 8 meets not more than one of its cells, will be called *p-tiles.* Let T_n be the set of p-tiles consisting of n cells, and $|T_n|$ be the number of elements in this set. Determine $|T_n|$.

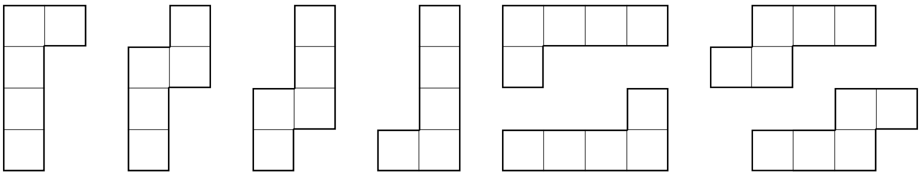


Figure 9.

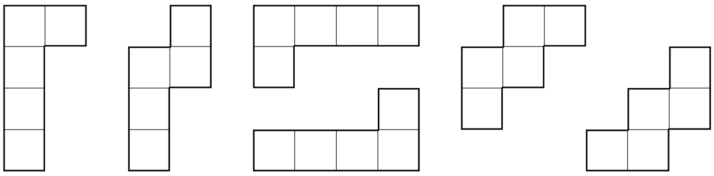


Figure 10.

⧫ **E8.** Prove that if a rectangle $a{\times}b$ can be covered by p-tiles shown at a) fig. 9, b) fig. 10, then 10 divides ab.

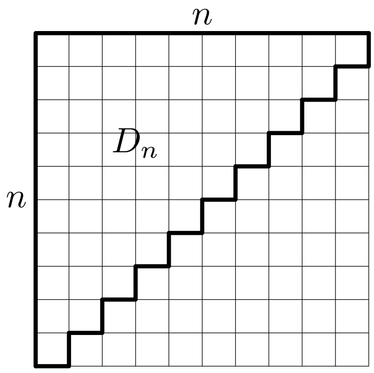


Figure 11.

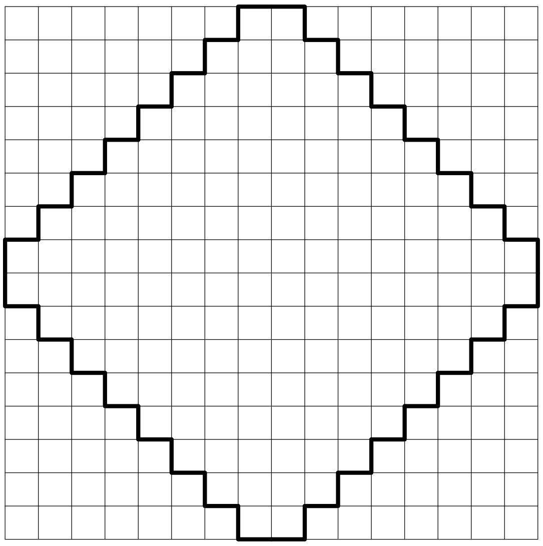


Figure 12.

⧫ **E9.** Let D_n be the figure indicated at fig. 11. Prove that if the figure D_n can be covered by p-tiles indicated at fig. 10 then $n \equiv 0, 4, 15, 19 \pmod{20}$.

⧫ **E10.** What is the number of possible tilings by dominoes for a) a rectangle $2{\times}m$? b) Aztec diamond? (see fig. 12)?

Consider a tiling (possibly with several layers) of a figure F by tiles from some set T. A distribution of numbers in cells will be called *an invariant* if the sum of numbers covered by any tile from the set T is divisible by p. An invariant is *nontrivial* if the sum of numbers covered by F is not divisible by p.

♦ **E11.** a) Prove that if a nontrivial invariant exists then F cannot be covered by tiles from T so that the multiplicity of covering of each cell equals 1 modulo p. b) Prove that if there are no nontrivial invariants then such a covering exists.

♦ **E12.** For which m and n a rectangle $m \times n$ can be covered by corner triminoes so that each cell is covered by equal number of triminoes?

♦ **E13.** *A semiinvariant* is a distribution of numbers in cells, such that the sum of numbers covered by each tile from the set T is negative and the sum of numbers covered by F is positive. Prove that a) if a semiinvariant exists then F cannot be covered by figures from T so that each cell is covered by equal number of figures; b) if there are no semiinvariants then such a covering does exist.

♦ **E14.** There are a board with lamps and a board with buttons. Pressing a button, we change the state of lamps connected with it, to the opposite state. a) Prove that the number of distributions possible for the states of lamps is a power of two. b) A set of lamps will be called an *invariant* if each button changes the state of an even number of lamps in it. Prove that if there are no invariants then all the lamps can be switched off independently of the initial state. c) Prove that if no invariant distinguishes the initial and the final states then a transfer from the initial state to the final state is possible.

♦ **E15.** A set of flips will be called *complete* if any pavement of the domain can be transferred to each one by some chain of flips from this set. Is it true that for some set of tiles there exists no complete set of flips?

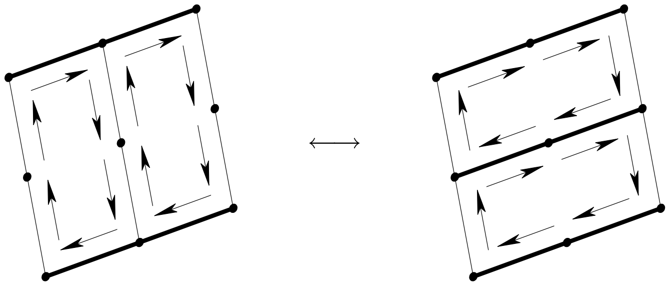
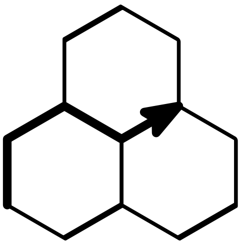


Figure 13.

♦ **E16*.** A plane graph is dissected into domains which are hexagons with no edges inside them (we may consider them as domino tiles 2×1). Each edge has two sides equiped by arrows. These two arrows have opposite directions and each hexagon has clockwise orientation. It is allowed to take a domain consisting of two adjacent hexagons and change it as is indicated at the figure 13. Is it true that every such dissection of a plane graph can be changed into each other by a chain of such flips?

9



Pavements, colorings and tiling groups

♦ **F1.** Consider an oriented graph whose edges are colored in two colors, and each vertex is the end of one red and one blue edge and the origin of similar edges. A self-coincidence of the graph is a rule which maps each vertex of the graph into some vertex so that each vertex has a single inverse image. Suppose that for each pair of vertices there exists a self-coincidence of the graph which saves orientation and colors of edges and maps the first vertex to the second one. If we pass a path such that first three edges are red and the remaining three edges are blue then we come to the same point as if first three edges were blue and the remaining three edges were red. Prove that the result remains true if 3 is replaced by 24.

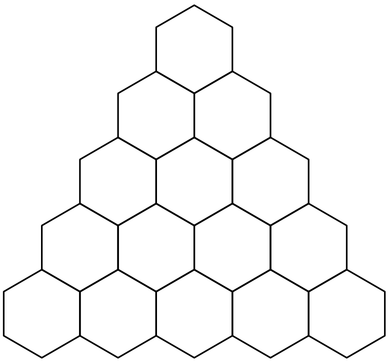


Figure 1. T_5

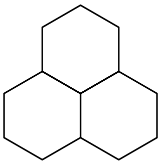


Figure 2. T_2

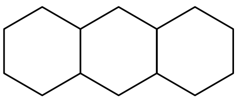


Figure 3. L_3

Define the domain T_n as a "triangle" formed of hexagons. Furthermore define L_n as n hexagon placed in a row (see fig. 1–3)

♦ **F2.** Find a correspondence between figures on a hexagon lattice and tiles on a square lattice. Place arrows of two colors (a and b) in the square lattice so that words corresponding to tiles L_n produce closed paths.

♦ **F3.** Construct a coloring (invariant) of a new square lattice from $F2$ such that paths L_4 correspond to zero values and figures T_n do not.

♦ **F4.** Prove that T_n cannot be tiled by figures L_3.

♦ **F5.** Find all values n such that T_n can be tiled by figures T_2.

Tilings of rectangles with T-tetrominoes

Michael Korn and Igor Pak

Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA, 02139

mikekorn@mit.edu, pak@math.mit.edu

August 26, 2003

Abstract

We prove that any two tilings of a rectangular region by T-tetrominoes are connected by moves involving only 2 and 4 tiles. We also show that the number of such tilings is an evaluation of the Tutte polynomial. The results are extended to a more general class of regions.

1 Introduction

The subject of tilings is a wonderful story that started as a collection of amateur problems (cf. [6]) and has now become an area of study in its own right, with numerous connections and applications to other fields: from group theory to topology, from enumerative combinatorics to probability. In the last decade various advanced methods have been developed which allowed some hard questions to be answered. This resulted in a structural approach to the study of tilings, which was presented in a recent survey [13] by the second author. The current paper carries out this approach to the very end for a special set of tiles. An unexpected combinatorial connection to the Tutte polynomial is a bonus and a delightful surprise.

In this paper we consider the set of T-tetrominoes, which has been studied earlier by Walkup in his curious paper [21]. His main result states that only rectangles of the form $4m \times 4n$ are tileable by T-tetrominoes. The proof in [21] is interesting but rather ad hoc. We explore further the structure of these tilings, combining Walkup's approach with several new direct bijections.

Our main result is local move connectivity of T-tetromino tilings for rectangular regions, resolving a conjecture in [13] in this case. This is done by introducing a new type of height function, and relating it to the tiling by means of two bijections. The height function technique for domino tilings was discovered by Thurston [19], and was used extensively in the recent literature to prove the local move connectivity for various sets of tiles (see e.g. [2, 8, 12, 17]).

Along the way we show that our new height functions enable us to define a lattice structure on all T-tetromino tilings of a rectangle. While this may seem a theoretical curiosity, in fact there are important applications of this result. There is a natural definition of a Markov chain on tilings (perform "random local moves") which translates into a nice Markov chain on height functions. Using the "coupling from the past" technique of Propp and Wilson and the fact that the height functions form a lattice (see [15, 14]) one can sample random T-tetromino tilings from an *exactly* uniform distribution (as opposed to a "nearly uniform" distribution usually obtained by the Markov chain approach). We refer to [16] for a full discussion of this approach and other examples of lattice structures on tilings.

Another classical problem for general tilings is their enumeration. Unfortunately, there seems to be little hope for a closed formula for the number of T-tetromino tilings of rectangles[1]. We show, however, that the

[1] For example, the 8×12 rectangle has an unpromising $1182 = 2 \cdot 3 \cdot 197$ tilings by T-tetrominoes.

number is an evaluation $T(3,3)$ of the *Tutte polynomial*, well-studied in the literature. Let us mention here that the Tutte polynomial is a fundamental invariant of graphs, and is related to a number of problems in discrete mathematics, computer science and several models in statistical mechanics. It has been extensively studied for various series of graphs as well as from a computational point of view (approximating its values is one of the challenges in the field). We refer to [22] for a beautifully written survey and a starting point for countless references.

Given the lack of a closed formula, one can ask whether the Markov chain approach can be used to efficiently *approximate* the number of tilings. In the past decade this idea has been developed into an important technique, first in the graph theoretic setting [7] and later in the tilings literature [9]. Without going into the details, the technique is based on showing rapid mixing of a Markov chain, and establishing a so-called self-reducibility property, the latter being usually much simpler. Roughly, one uses a Markov chain to sample a number of uniform tilings, collect statistics for certain patterns among the sampled tilings, and reduce the problem to a smaller similar problem. The self-reducibility allows such a reduction and keeps the errors relatively small. Unfortunately, we are unable to carry out this approach in full. We show the self-reducibility, but rapid mixing of our Markov chain goes beyond the scope of this work and is stated as a conjecture.

Returning back to combinatorics of the T-tetromino tilings, we answer a question as to what extent our results can be generalized from rectangular to other regions. We show that in fact all the bijective proofs go through for quadruplicated simply connected regions. As a corollary, we have the local move connectivity for such regions. We conclude by showing that if either condition on the regions is dropped, there is no local move property.

After the results of this paper were obtained, we learned of an alternative approach to the problem presented in a recent manuscript by Konstantin and Yuri Makarychev [10]. The authors showed that one can prove local move connectivity of T-tetromino tilings for rectangular regions by means of the so-called *ice graphs* and by using Eloranta's theorem (see [4]). We discovered that this approach can be combined with ours and one can define a height function, similar to ours. We present our findings in the appendix. For completeness and clarity of the exposition we start by recalling Makarychev's results (with independent proofs), define a new height function, and then proceed to prove the local move connectivity by a height function.

A few words about the structure of the paper. We start with definitions and basic results. In section 3 we state Walkup's result about the structure of T-tetromino tilings of a rectangle which will be an important technical tool for the rest of the paper. Then, in section 4, we define a new notion of chain graphs and show that they are in one-to-one correspondence with T-tetromino tilings. In section 5 we introduce the height functions, which we use in section 6 to prove local connectivity. We introduce a lattice structure on height functions in section 7. In section 8 we consider tilings of non-rectangular regions, proving local connectivity for quadruplicated simply connected regions. In section 9 we define two planar graphs which correspond to chain graphs, which enables us to obtain an enumerative formula for the number of T-tetromino tilings. Section 10 contains the description of the Markov chain $\mathcal{M}$ on T-tetromino tilings and proves the self-reducibility. We conclude with final remarks and the appendix outlining an alternative approach to local connectivity.

2 Main results

A *T-tetromino* is the figure formed by four unit squares arranged as shown in Figure 1. We make no distinction between the four possible orientations of the T-tetromino. A *tiling* of a region Γ with T-tetrominoes is an arrangement of T-tetrominoes which covers every square of Γ exactly once. An individual T-tetromino in such a tiling is called a *tile*. Let $\mathcal{T}_\Gamma$ denote the set of all possible tilings of Γ by T-tetrominoes.

In this paper, we will work exclusively with tilings by T-tetrominoes. Every reference to tilings or tiles refers to T-tetrominoes, even if this is not explicitly stated. We will chiefly be interested in tiling rectangular regions, although in section 8 we will see that much of what we prove also holds for a somewhat more general

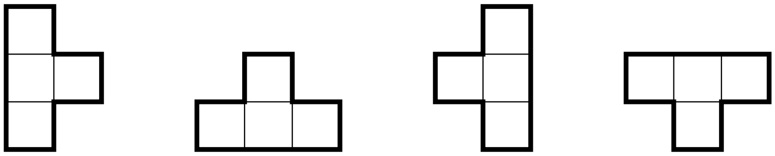


Figure 1: T-tetrominoes.

class of shapes.

Given a tiling of a region Γ, one can transform it into a new tiling by performing a "local move". A local move consists of picking up some (small) number of tiles, then re-filling that area with tiles in a different way. Two natural local moves for T-tetrominoes are shown in Figure 2. We call them the 2-move and the 4-move.

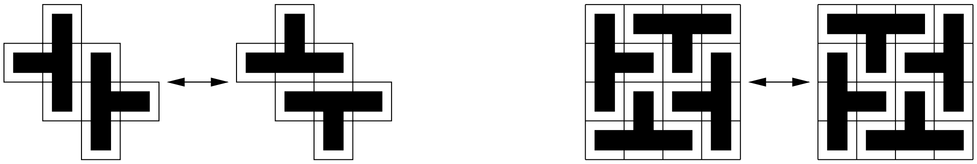


Figure 2: Local 2-move and local 4-move.

Suppose we are given a region Γ and a collection $\mathcal{L}$ of allowable local moves, and suppose that τ_1 and τ_2 are two different tilings of Γ. Let us say that τ_1 and τ_2 are *local-move equivalent* with respect to $\mathcal{L}$ if it is possible to transform τ_1 into τ_2 by performing a sequence of local moves from $\mathcal{L}$. This is an equivalence relation on the set of all tilings of Γ. A natural question to ask is whether all tilings of Γ are local-move equivalent. If so, we say that the region Γ has *local connectivity* with respect to $\mathcal{L}$. If $\mathcal{R}$ is a set of regions (the set of all rectangles, or the set of all simply-connected regions, for example), we say that $\mathcal{R}$ has a *local-move property* if there exists a finite set $\mathcal{L}$ of local moves such that all regions $\Gamma \in \mathcal{R}$ have local connectivity with respect to $\mathcal{L}$.

One of our main results is the following.

Theorem 1 *For tilings by T-tetrominoes, the set of all rectangles has a local-move property. Specifically, every rectangle* Γ *has local connectivity with respect to the 2-move and 4-move.*

This result was conjectured in [13] to hold for all simply connected regions. Later on, in section 8, we extend this theorem to a more general class of regions and show that the conjecture does not hold in full generality.

3 Tiling rectangles with T-tetrominoes

Without loss of generality, let Γ be a rectangle which is situated in the first quadrant of the Cartesian plane, with one corner at $(0, 0)$. Let a *type-A* point be a point whose coordinates are congruent mod 4 to (0,0) or (2,2), and let a *type-B* point be a point whose coordinates are congruent mod 4 to (0,2) or (2,0). A segment of length 1 is called a *cut* if there is no valid tiling of Γ in which a tile crosses that segment. A point is called *cornerless* if there is no valid tiling of Γ in which that point is one of the eight corners of a tile.

In [21], Walkup proves the following property of T-tetromino tilings of rectangles (see Figure 3).

Theorem 2 (Walkup) *If an* $m \times n$ *rectangle can be tiled by T-tetrominoes, then both* m *and* n *must be divisible by 4. Furthermore, all segments incident to type-A points are cuts, and all type-B points are cornerless.*

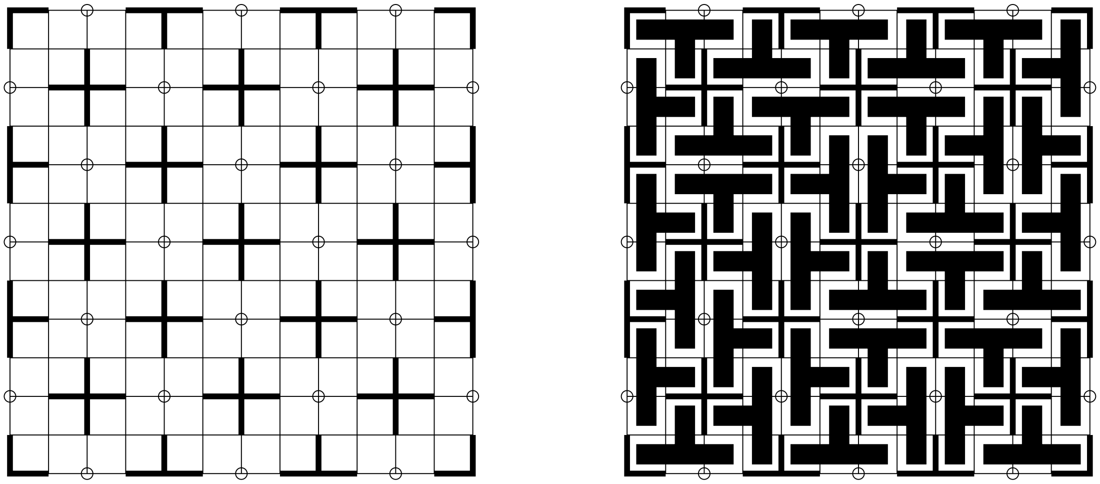


Figure 3: The dark lines are cuts. Circles are cornerless points.

From now on, we will only be concerned with rectangles having sides divisible by 4, since all other rectangles are untileable.

Define a *block* to be a 2×2 square whose corners have even coordinates. The following lemma is immediate by inspection from the structure of cuts and cornerless points.

Lemma 3 *In any tiling of a rectangle by T-tetrominoes, each tile contains three squares from one block and one square from an adjacent block. Similarly, each block contains three squares from one tile and one square from another tile.*

4 Chain graphs

Define an *antiblock* to be a 2×2 square whose corners have odd coordinates. Color the antiblocks white and gray in checkerboard fashion, so that antiblocks centered at type-A points are gray and those centered at type-B points are white.

For a $4m \times 4n$ rectangle Γ, let V_Γ be the set of points in Γ which have odd coordinates. Say that a directed graph on the vertices V_Γ is a *chain graph* if it satisfies the following properties:

- every edge connects vertices that are two units apart (either vertically or horizontally),
- every vertex has indegree 1 and outdegree 1, and
- every white antiblock contained in Γ borders exactly two edges of the graph, and these edges are non-adjacent.

Let $\mathcal{C}_\Gamma$ denote the set of all chain graphs of a region Γ.

Theorem 4 *For any $4m \times 4n$ rectangle Γ, we have $|\mathcal{C}_\Gamma| = |\mathcal{T}_\Gamma|$.*

The proof is based on an explicit bijection $\phi : \mathcal{T}_\Gamma \to \mathcal{C}_\Gamma$ defined as follows.

Let $\tau \in \mathcal{T}_\Gamma$ be a tiling. Notice that each vertex in V_Γ lies in the middle of some block. By Lemma 3, each tile in τ contains three squares from one block and one square from an adjacent block. Call these blocks the *primary* and *secondary* blocks of the tile respectively. For each tile, draw a directed edge from its primary block to its secondary block, and define $\phi(\tau)$ to be the directed graph which results (see Figure 4).

Theorem 4 follows immediately from the following Lemma.

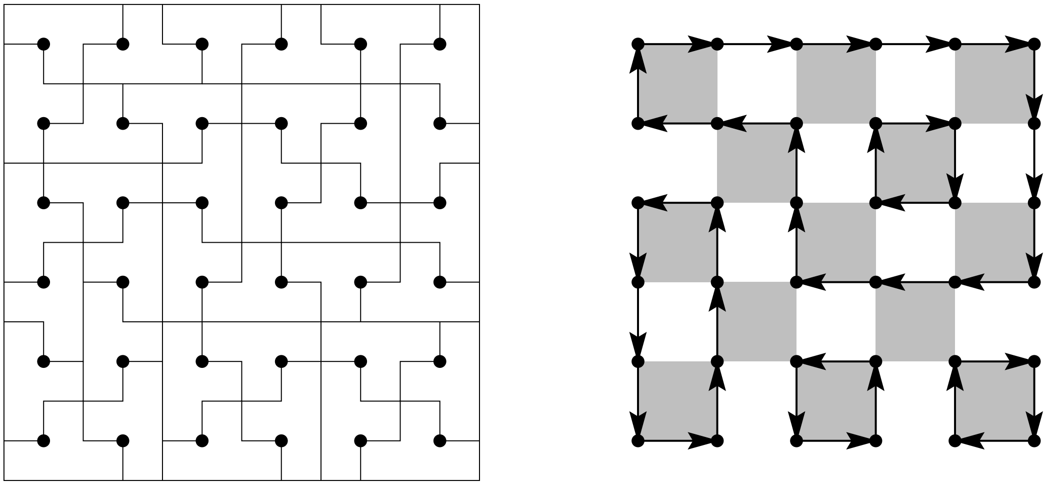


Figure 4: A tiling τ, and the chain graph $\phi(\tau)$.

Lemma 5 *For any $4m \times 4n$ rectangle Γ, the map ϕ defined above is a bijection between $\mathcal{T}_\Gamma$ and $\mathcal{C}_\Gamma$.*

Proof: First let us show that $\phi(\tau)$ is a chain graph. It is clear from the definition, and from Lemma 3, that every vertex will have indegree 1 and outdegree 1, and that edges will only connect vertices which are two units apart. As for the third restriction, consider a type-B point not on the boundary. Up to rotations and reflections, the tiles surrounding it must look like one of the two possibilities shown in Figure 5. Thus there will be exactly two edges bordering the associated white antiblock, and they will be non-adjacent. Hence $\phi(\tau)$ is a chain graph for all τ.

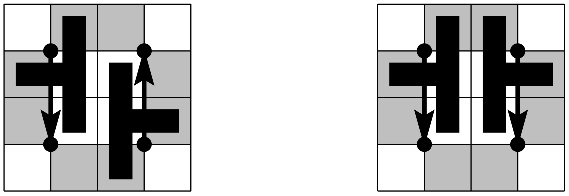


Figure 5: The two possibilities for a type-B point.

Notice that each tile corresponds to an edge in this graph. For each edge, there is only one possible tile placement which yields that edge and is consistent with the cuts and cornerless points. Hence the map ϕ is injective.

What remains to be shown is that every chain graph is equal to $\phi(\tau)$ for some tiling τ. As we just observed, for each edge there is only one possible tile placement that can yield that edge. So any chain graph will yield a collection of tile placements. It remains to be checked that these tiles cover all of Γ and do not overlap. Since each vertex has outdegree 1, the number of edges equals the number of blocks, so the total area of the tiles will equal the area of Γ. Thus it will be sufficient to verify that the tiles do not overlap.

Assume there are two tiles which overlap. Let us assume the overlap occurs in the block containing the squares A, B, D, and E (see Figure 6). Without loss of generality, we may take one of the tiles to be the one covering squares B, D, E, and F. Since each vertex has indegree 1 and outdegree 1, the tile which overlaps this one must contain only one square from this block, hence the overlap must occur at E. There are two possible tiles which cover E. First there is the tile which covers C, E, F, and G. If we have this, then the graph must contain both an edge and its opposite. This violates the rule about what a white antiblock may border. The other possibility is the tile which covers E, H, I, and J. In this case, the graph must contain two adjacent edges both on the same white antiblock, which again violates the constraint. Thus there can be no overlaps, which proves that every chain graph is $\phi(\tau)$ for some tiling τ. ■

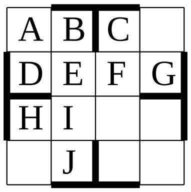


Figure 6: How an overlap may occur.

5 Height functions

Let us call a point having coordinates congruent mod 4 to (0,0) a *type-A0* point. Similarly, a point congruent to (2,2) will be called a *type-A1* point. (Points congruent to either (0,2) or (2,0) will still be called *type-B* points.)

For a $4m \times 4n$ rectangle Γ, let W_Γ be the set of points in Γ which have even coordinates. Let $\partial\Gamma$ denote the set of boundary point of Γ. Say that a function $f : W_\Gamma \to \mathbb{Z}$ is a *height function* if it satisfies the following properties:

- $f(x) = 0$ for all $x \in \partial\Gamma$,
- $f(x)$ is an even integer for all type-A0 points x,
- $f(x)$ is an odd integer for all type-A1 points x, and
- $|f(x) - f(y)| \leq 1$ whenever x and y are adjacent (at a distance of two units).

Let $\mathcal{H}_\Gamma$ denote the set of all height functions of a region Γ.

Theorem 6 *For any $4m \times 4n$ rectangle Γ, we have $|\mathcal{H}_\Gamma| = |\mathcal{T}_\Gamma|$.*

We define a map $\psi : \mathcal{C}_\Gamma \to \mathcal{H}_\Gamma$ as follows. Let $C \in \mathcal{C}_\Gamma$ be a chain graph. Define a function f° on the faces of C by the following rules. Let f° have the value 0 on the unbounded face of C. As we pass an edge of the graph, if the edge points to the right, let the value of f° increase by 1. (Similarly, if the edge points to the left, let the value of f° decrease by 1.) Now define $f : W_\Gamma \to \mathbb{Z}$ by letting $f(x)$ equal the value of f° on the face in which x lies (see Figure 7). Define $\psi(C)$ to be this function f.

Theorem 6 follows immediately from Theorem 4 and the following lemma.

Lemma 7 *For any $4m \times 4n$ rectangle Γ, the map ψ defined above is a bijection between $\mathcal{C}_\Gamma$ and $\mathcal{H}_\Gamma$.*

Proof: Let C be a chain graph, and let f be $\psi(C)$ as defined above. Let us first show that the function f is well-defined. If it is not, then there must exist some closed path through the faces of the graph such that the net change in the value of f° is non-zero. This means that upon going around this path counterclockwise, we cross more right-pointing edges than left-pointing edges, say. Therefore more edges leave the area enclosed by the path than enter that area. But this is impossible since every vertex has equal indegree and outdegree, so the net flow out of any region must be zero. Hence f is a well-defined function on W_Γ.

Next, let us verify that f is a valid height function. Points $x \in \partial\Gamma$ lie in the unbounded face of C, hence $f(x) = 0$ for such points. And if x and y are adjacent points, then they lie either in the same face of C or in adjacent faces of C, hence the difference between $f(x)$ and $f(y)$ is at most 1. Now let us verify the other two statements. As one travels from a type-A0 point x to another type-A0 point y which is 4 units away, one

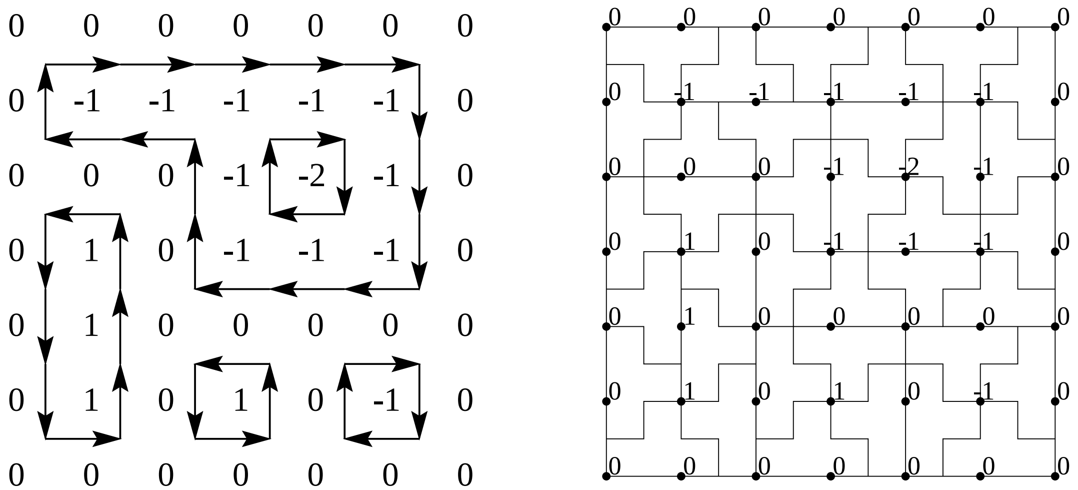


Figure 7: A chain graph C, and the function $f = \psi(C)$.

passes through the middle of a white antiblock (see Figure 8). In doing so, one crosses either 0 or 2 edges of C, hence the value of f° will have changed twice, or not at all, so $f(x)$ and $f(y)$ will have the same parity. Since (0,0) is a type-A0 point, and $f((0,0)) = 0$, it follows that $f(x)$ will be even for all type-A0 points x. By the same argument, all type-A1 points must have the same parity as each other. And (2,2) is a type-A1 point with $f((2,2)) = \pm 1$, so $f(x)$ will be odd for all type-A1 points x. Thus f is in fact a height function.

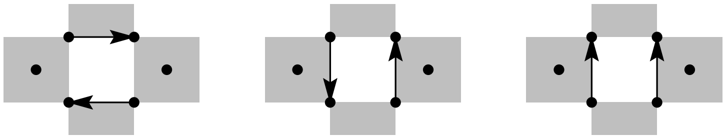


Figure 8: Two type-A0 points, and what might lie between them.

Given a height function $f = \psi(C)$, one can uniquely reconstruct the chain graph C by inserting directed edges in the places where the value of f increases or decreases. Hence ψ is an injective map. It remains to be shown that every height function f is equal to $\psi(C)$ for some valid chain graph C.

Take a height function f, and insert directed edges along the boundaries where the value of f increases or decreases. Call this graph C. Consider a vertex of C. To one corner of it, there is a type-A0 point x_0, on the opposite corner is a type-A1 point x_1, and the remaining two corners are type-B points y_0 and y_1. Since $f(x_0)$ is even, and $f(x_1)$ is odd, these values must differ by exactly 1. Without loss of generality, assume $f(x_0) = h$ and $f(x_1) = h + 1$. Then both $f(y_0)$ and $f(y_1)$ must be h or $h + 1$ as well. Up to rotations, the situation must look like one of the possibilities in Figure 9. Thus the vertex in question will have indegree 1 and outdegree 1.

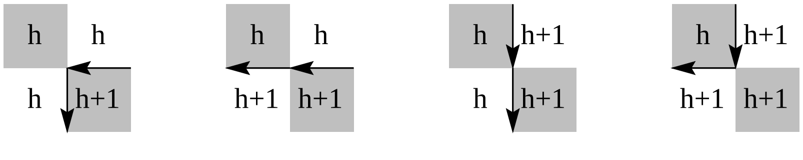


Figure 9: The possibilities for a vertex of C.

Now consider a type-B point y, which corresponds to a white antiblock. Let $f(y) = h$, and assume without loss of generality that h is even. If z_1 and z_2 are the two type-A0 points adjacent to y, then we must have $f(z_1) = f(z_2) = h$. If z_3 and z_4 are the two type-A1 points adjacent to y, then we must have $f(z_3) = h \pm 1$ and $f(z_4) = h \pm 1$, not necessarily the same (see Figure 10). So this white antiblock will border exactly two non-adjacent edges of C.

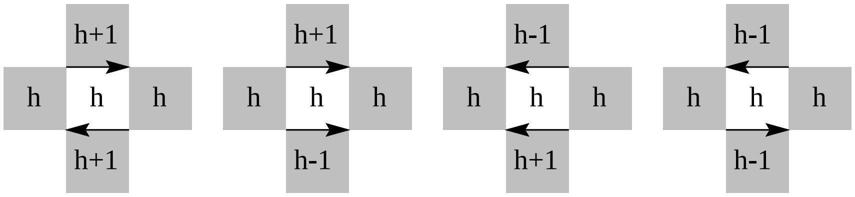


Figure 10: The possibilities for a white antiblock.

Hence the graph C constructed in this way from a height function f is indeed a chain graph, and $\psi(C) = f$. This completes the proof. ■

For ease of notation, define $\zeta(\tau) = \psi(\phi(\tau))$. For a $4m \times 4n$ rectangle Γ, the map ζ is the canonical bijection between $\mathcal{T}_\Gamma$ and $\mathcal{H}_\Gamma$.

Lemma 8 *Let Γ be a $4m \times 4n$ rectangle and let $\tau_1, \tau_2 \in \mathcal{T}_\Gamma$ be tilings of Γ. The tilings τ_1 and τ_2 differ by a 2-move if and only if the height functions $\zeta(\tau_1)$ and $\zeta(\tau_2)$ differ by 1 on some type-B point, and are the same everywhere else. The tilings τ_1 and τ_2 differ by a 4-move if and only if the height functions $\zeta(\tau_1)$ and $\zeta(\tau_2)$ differ by 2 on some type-A point, and are the same everywhere else.*

Proof: By inspection of the structure of cuts and cornerless points, one sees that the 2-move must be centered at a type-B point, and the 4-move must be centered at a type-A point. From Figure 11, one can see that if τ_1 and τ_2 differ by a 2-move, then the height functions $\zeta(\tau_1)$ and $\zeta(\tau_2)$ differ by 1 in their values on the corresponding type-B point. Similarly, if τ_1 and τ_2 differ by a 4-move, then the height functions $\zeta(\tau_1)$ and $\zeta(\tau_2)$ differ by 2 in their values on the corresponding type-A point.

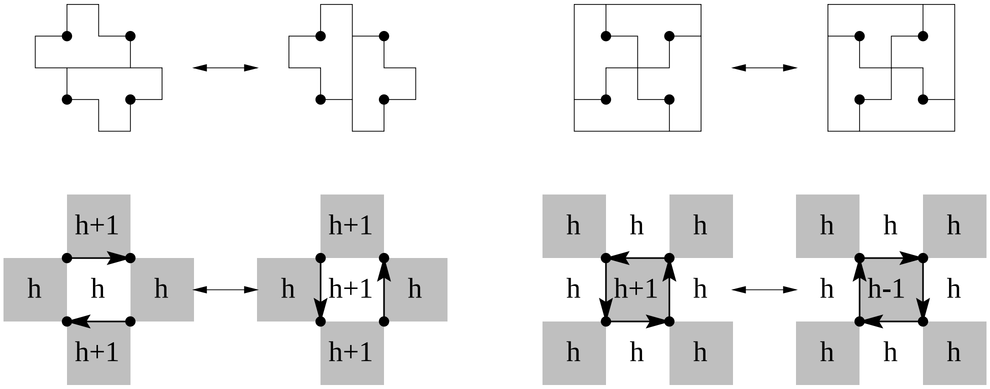


Figure 11: The 2-move and 4-move, and their effect on $\zeta(\tau)$.

As for the converse, suppose there are height functions f_1 and f_2 which are identical everywhere, except $f_1(y) = h$ and $f_2(y) = h + 1$ for some type-B point y. Thus the value of f_1 (or f_2) on the neighbors of y must be $h, h + 1, h$, and $h + 1$ (since they must alternate even and odd). Hence the picture must look like the bottom left of Figure 11, possibly rotated. Going backwards, we see what the chain graph and the tiling must then look like, and that in fact, $\zeta^{-1}(f_1)$ and $\zeta^{-1}(f_2)$ differ by a 2-move.

Similarly, suppose there are height functions f_1 and f_2 which are identical everywhere, except $f_1(x) = h+1$ and $f_2(x) = h - 1$ for some type-A point x. Thus $f_1(y) = f_2(y) = h$ for all neighbors y of x. Hence the picture must look like the bottom right of Figure 11. Going backwards, we see what the chain graph and the tiling must then look like, and that in fact, $\zeta^{-1}(f_1)$ and $\zeta^{-1}(f_2)$ differ by a 4-move. ■

For height functions $f_1, f_2 \in \mathcal{H}_\Gamma$, say that f_1 and f_2 differ by a 2-move (or 4-move) if the tilings $\zeta^{-1}(f_1)$ and $\zeta^{-1}(f_2)$ differ by a 2-move (or 4-move). By the previous Lemma, one can see that performing a 2-move on a height function f is equivalent to increasing or decreasing its value by 1 at some type-B point. Similarly, performing a 4-move is equivalent to increasing or decreasing the value of f by 2 at some type-A point. Of course, such moves may only be applied if the function that results is a valid height function.

6 Local connectivity from height functions

Theorem 1 will easily follow from the following lemma.

Lemma 9 *Let Γ be a $4m \times 4n$ rectangle, and let $f_1, f_2 \in \mathcal{H}_\Gamma$ be height functions. It is always possible to convert f_1 into f_2 by performing a sequence of 2-moves and 4-moves.*

Proof: For a $4m \times 4n$ rectangle Γ, let f_0 be the height function which is 1 on the type-A1 points of Γ, and 0 everywhere else. We would like to show that every height function f can be transformed into f_0. If every height function can be transformed into f_0, it follows that any height function can be transformed into any other. Suppose $f(x) > 1$ for some x. Let h be the largest value that f attains. Suppose there is a type-B point y which attains this value. Then f must take the values $h, h-1, h$, and $h-1$ on the neighbors of y. So we can perform a 2-move to change $f(y)$ to $h-1$ and still have a valid height function. We do this for all type-B points at which f attains the value h. Now look at any remaining (type A0 or A1) point x having $f(x) = h$. We must have $f(z) = h-1$ for the neighbors z of x, since there are no type-B points remaining for which $f(z) = h$. So we can perform a 4-move to change $f(x)$ to $h-2$. We do this for every point where f attains the value h. Now the largest value which appears is at most $h-1$, and we repeat the procedure until we have $f(x) \leq 1$ for all x.

We do a similar thing for points where $f(x) < 0$, increasing them until $f(x) \geq 0$ for all x. At this point, all points will have the value 0 or 1 (in particular, $f(x) = 0$ for all type-A0 points x, and $f(x) = 1$ for all type-A1 points). It just remains to set $f(y) = 0$ for all type-B points y, which can be done by a sequence of 2-moves. This finishes the procedure, proving the lemma. ■

7 The lattice structure on height functions

There is a natural partial order on $\mathcal{H}_\Gamma$. If $f_1, f_2 \in \mathcal{H}_\Gamma$ are height functions, we say $f_1 \leq f_2$ iff $f_1(x) \leq f_2(x)$ for all points x. This partial order can be extended to tilings—say $\tau_1 \leq \tau_2$ if $\zeta(\tau_1) \leq \zeta(\tau_2)$.

Theorem 10 *For any $4m \times 4n$ rectangle Γ, the poset P_Γ consisting of all tilings of Γ, with this order relation, is a distributive lattice.*

Proof: In order to prove that P_Γ is a lattice, we need to show that for height functions f_1 and f_2, there exists a unique greatest lower bound ("meet") α and least upper bound ("join") β. We define $\alpha(x) = \min\{f_1(x), f_2(x)\}$ and $\beta(x) = \max\{f_1(x), f_2(x)\}$, for all x. Clearly $\alpha \leq f_1$ and $\alpha \leq f_2$, and all other lower bounds are less than α. It just remains to be shown that α is a valid height function. Clearly the values of α on the boundary will be 0, and the type-A0 points will be even and the type-A1 points will be odd, because these properties hold for f_1 and f_2. As for adjacent values differing by at most 1, suppose x and y are adjacent points, and $\alpha(x) \geq \alpha(y) + 2$. Without loss of generality, assume $\alpha(y) = f_1(y)$. Then it would follow that $f_1(x) \geq \alpha(x) \geq \alpha(y) + 2 = f_1(y) + 2$, a contradiction. Therefore, α is a valid height function. The proof for β is analogous.

To prove that P_Γ is a distributive lattice, we need to verify the distributive laws: For height functions f, g, and h,

$$(f \vee g) \wedge (f \vee h) = f \vee (g \wedge h) \qquad \text{and} \qquad (f \wedge g) \vee (f \wedge h) = f \wedge (g \vee h).$$

For any x we have:

$$((f \vee g) \wedge (f \vee h))(x) = \min(\max(f(x), g(x)), \max(f(x), h(x)).$$

The functions min and max satisfy the distributive laws, so we have

$$\min(\max(f(x), g(x)), \max(f(x), h(x)) = \max(f(x), \min(g(x), h(x)) = (f \vee (g \wedge h))(x).$$

Hence

$$(f \vee g) \wedge (f \vee h) = f \vee (g \wedge h),$$

as desired. Note that changing the sign of functions switches the role of $\vee$ and $\wedge$, which implies the second distributive law. Therefore, P_Γ is a distributive lattice. ■

8 Non-rectangular regions

A quadruplicated simply connected region is a region which is formed by taking a simply-connected union of grid squares and dilating the figure by 4 in each direction. Let $\mathcal{Q}$ denote the set of all such regions. As we did for rectangles, we will assume that the corners of such a shape have coordinates which are congruent to (0,0) mod 4. Notice that $\mathcal{Q}$ contains all $4m \times 4n$ rectangles.

Theorem 11 *The second part of Theorem 2 holds for all regions* $\Gamma \in \mathcal{Q}$.

Proof: Suppose there exists a region $\Gamma \in \mathcal{Q}$ which can be tiled in a way which violates some of the supposed cuts and cornerless points. Let Γ' be the smallest $4m \times 4n$ rectangle which contains Γ. We can extend the tiling of Γ to a tiling of Γ' by adding tiled 4×4 squares to the part of Γ' which is not in Γ. This gives a tiling of Γ' which violates the necessary cuts and cornerless points, which contradicts Theorem 2. ■

As a result of this, all the above results for rectangles are also true for all $\Gamma \in \mathcal{Q}$. The proofs are the same as before.

The results do not hold if we drop the condition of being simply-connected. (Notice that the correspondence between chain graphs and height functions breaks down if the region is not simply connected, because points on the boundary of the region need not be on the unbounded face of the chain graph, so they may have nonzero height.) For example, Figure 12 shows a tiling of a non-simply connected region where neither the 2-move nor the 4-move can be applied.

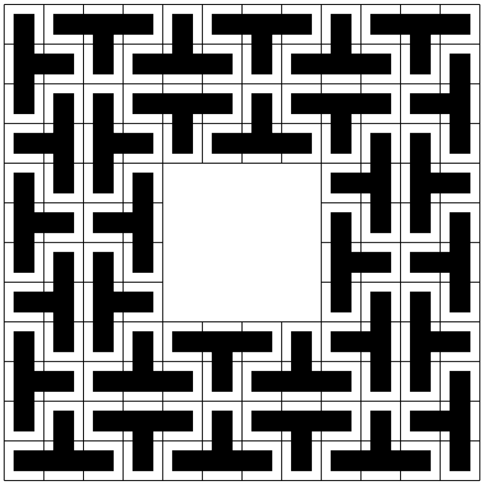


Figure 12: Tiling of a non-simply connected region.

Theorem 12 *Let $\mathcal{S}$ denote the set of all simply-connected regions. For tilings by T-tetrominoes, the set $\mathcal{S}$ does not have a local-move property.*

Proof: Let Δ_1 denote the region shown in Figure 13. It is straightforward to see that this region can be tiled in only two ways, namely the way shown and its mirror image. Since there are no intermediate tilings, and no tile is in the same place in both tilings, the only way for local connectivity to hold for this region is if we declare this entire transformation to be one local move.

In fact, we can generate infinitely many regions which admit only two tilings. Let Δ_k denote the region in Figure 14, where the total length of the region is $8k + 2$. As before, it can only be tiled in two ways, so in order to have local connectivity, the entire region must be considered to be a local move. No finite set of local moves can contain all of these, hence any finite set of local moves is insufficient to give local connectivity for these regions. ■

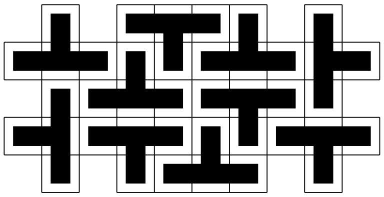


Figure 13: The region Δ_1.

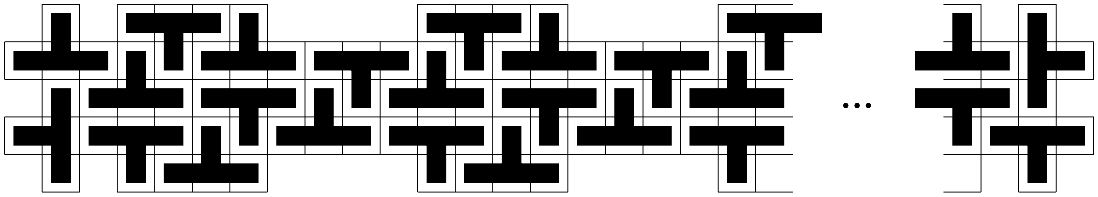


Figure 14: The region Δ_k.

9 Enumeration of tilings and the Tutte polynomial

For a region $\Gamma \in \mathcal{Q}$, define the graph G_Γ as follows. Include a vertex for each type-A1 point, and connect two vertices with an undirected edge if they are 4 units apart (vertically or horizontally). Similarly, define G^*_Γ by including a vertex for every type-A0 point, and again connecting those vertices which are 4 units apart. Note that when Γ is a $4m \times 4n$ rectangle, the graphs G_Γ and G^*_Γ are isomorphic to the $m \times n$ and $(m + 1) \times (n + 1)$ rectangular shape subgraphs of the square grid.

For a graph G, we let $V(G)$ and $E(G)$ denote the set of vertices and edges of G respectively. Let $c(G)$ denote the number of connected components of G. If $e \in E(G)$, let $G \backslash e$ be the graph formed by deleting e from G. Similarly, let G/e be the graph formed by contracting e in G.

The Tutte polynomial $T(G; x, y)$ is a polynomial in the variables x and y which is defined for undirected graphs G. Typically it is defined in terms of the following recursive formulas (see [22]):

- $T(G; x, y) = 1$ if G has no edges,
- $T(G; x, y) = y \cdot T(G \backslash e; x, y)$ if e is a loop,
- $T(G; x, y) = x \cdot T(G/e; x, y)$ if e is a cutedge,
- $T(G; x, y) = T(G \backslash e; x, y) + T(G/e; x, y)$ if e is neither a loop nor a cutedge.

Another equivalent definition of $T(G;x,y)$ is as follows. Let H be a spanning subgraph of G (that is, a subgraph of G which contains all the vertices of G). Then

$$T(G;x,y) = \sum_{H \subset G} (x-1)^{c(H)-c(G)}(y-1)^{c(H)+|E(H)|-|V(G)|}$$

where the sum is over all spanning subgraphs $H \subset G$.

Theorem 13 *For every $\Gamma \in \mathcal{Q}$, the number of T-tetromino tilings of Γ is equal to $2 \cdot T(G_\Gamma; 3, 3)$.*

To prove this, we introduce a few lemmas about spanning subgraphs of G_Γ and G^*_Γ.

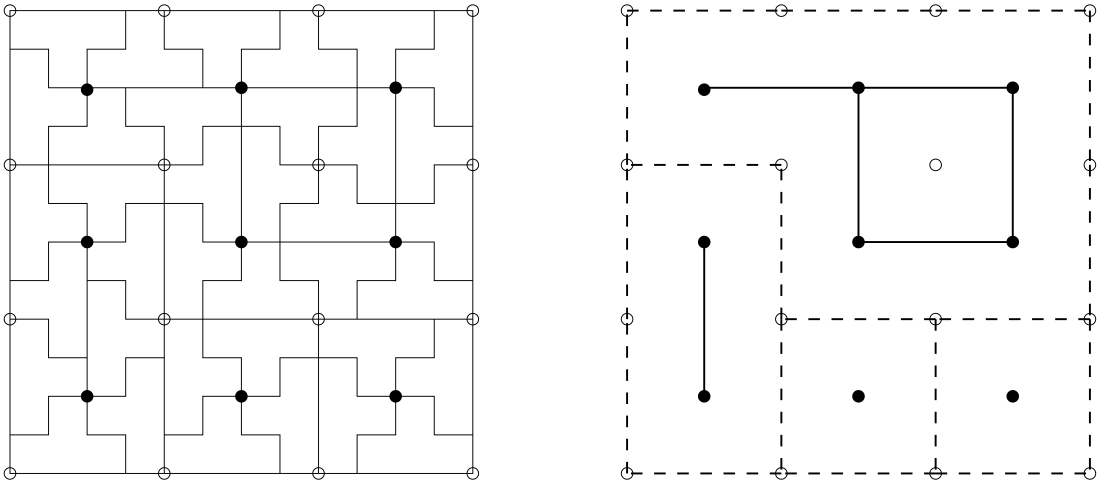


Figure 15: A tiling τ, and the graphs $\sigma(\tau)$ (solid lines) and $\sigma^*(\tau)$ (dotted lines).

Given a tiling τ of Γ, define $\sigma(\tau)$ to be the spanning subgraph of G_Γ which includes those edges which do not cross any tile. Similarly, define $\sigma^*(\tau)$ to be the spanning subgraph of G^*_Γ which includes those edges which do not cross any tile (see Figure 15).

Suppose H is a spanning subgraph of G_Γ. Define $\omega(H)$ to be the spanning subgraph of G^*_Γ consisting of those edges which do not cross any edge of H.

Lemma 14 *Fix $\Gamma \in \mathcal{Q}$ and a tiling $\tau \in \mathcal{T}_\Gamma$. Then $\omega(\sigma(\tau)) = \sigma^*(\tau)$. Furthermore, no edge of the chain graph $\phi(\tau)$ crosses an edge of either $\sigma(\tau)$ or $\sigma^*(\tau)$. Conversely, any edge of G_Γ or G^*_Γ which does not cross any edge of $\phi(\tau)$ is an edge of $\sigma(\tau)$ or $\sigma^*(\tau)$.*

Proof: Notice that the points where an edge of G_Γ and an edge of G^*_Γ intersect are precisely the type-B points in the interior of Γ. Consider any such point. Recalling Figure 5, observe that exactly one of the two edges which meet there will avoid crossing tiles of τ. Hence each such point is on an edge of either $\sigma(\tau)$ or $\sigma^*(\tau)$, but not both. So an edge of G^*_Γ is in $\sigma^*(\tau)$ if and only if no edge of $\sigma(\tau)$ crosses it. Hence $\sigma^*(\tau) = \omega(\sigma(\tau))$.

Recall that in $\phi(\tau)$, each edge corresponds to a tile; the edge connects the two blocks in which the tile lies. Edges of $\sigma(\tau)$ and $\sigma^*(\tau)$ run along block boundaries; an edge is present in these graphs if and only if no tile crosses that boundary. If no tile crosses that boundary, then no edge of $\phi(\tau)$ will either. Conversely, if no edge of $\phi(\tau)$ crosses a block boundary, then no tile crosses that boundary, hence that boundary will be an edge of $\sigma(\tau)$ or $\sigma^*(\tau)$. (See Figure 16.) ■

Corollary 15 *Suppose a region $\Gamma \in \mathcal{Q}$ and tilings $\tau_1, \tau_2 \in \mathcal{T}_\Gamma$ satisfy $\sigma(\tau_1) = \sigma(\tau_2)$. Then $\phi(\tau_1)$ and $\phi(\tau_2)$ are identical up to the orientation of the edges.*

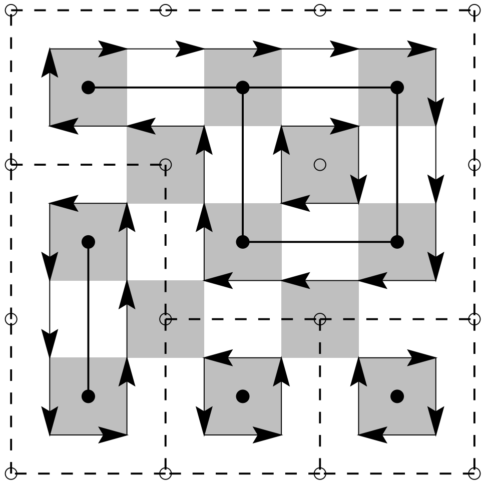


Figure 16: The graphs $\sigma(\tau)$ and $\sigma^*(\tau)$, and the chain graph $\phi(\tau)$.

Proof: Let $H = \sigma(\tau_1) = \sigma(\tau_2)$. For each white antiblock, there is exactly one edge of G_Γ which crosses it. The presence or absence of that edge in H determines which pair of edges along the white antiblock must be included in the corresponding chain graphs. This gives all the edges of the chain graphs, except those which do not border a complete white antiblock (ones near the boundary of the region). By inspection, one can see that all those edges must be included in order to have total degree 2 at each vertex of the chain graphs. ■

Lemma 16 *Let $\Gamma \in \mathcal{Q}$, and let H be a spanning subgraph of G_Γ. Then*

$$c(\omega(H)) \;=\; c(H) + |E(H)| - |V(G_\Gamma)| + 1.$$

Proof: We fix Γ and prove this by induction on the number of edges in H. If H has no edges, then $c(H) = |V(G_\Gamma)|$, so $c(H) + |E(H)| - |V(G_\Gamma)| + 1 = 1$, which is equal to $c(\omega(H))$, as required. Now assume that the result holds for all subgraphs $H \subset G_\Gamma$ with $|E(H)| < k$.

Consider a subgraph H with $|E(H)| = k$, and let $e \in E(H)$. First, suppose that e is a cutedge of H. Then $c(H\backslash e) = c(H) + 1$, $|E(H\backslash e)| = |E(H)| - 1$, and $c(\omega(H\backslash e)) = c(\omega(H))$. We conclude:

$$c(\omega(H)) = c(\omega(H\backslash e)) = c(H\backslash e) + |E(H\backslash e)| - |V(G_\Gamma)| + 1 = c(H) + |E(H)| - |V(G_\Gamma)| + 1.$$

Now suppose that e is not a cutedge of H. Then $c(H\backslash e) = c(H)$, $|E(H\backslash e)| = |E(H)| - 1$, and $c(\omega(H\backslash e)) = c(\omega(H)) - 1$. We have

$$c(\omega(H)) = c(\omega(H\backslash e)) + 1 = c(H\backslash e) + |E(H\backslash e)| - |V(G_\Gamma)| + 2 = c(H) + |E(H)| - |V(G_\Gamma)| + 1,$$

as desired. Therefore $c(\omega(H)) = c(H) + |E(H)| - |V(G_\Gamma)| + 1$ holds for all subgraphs $H \subset G_\Gamma$. ■

Suppose H is a spanning subgraph of G_Γ. Define $a(H) = 2c(H) + |E(H)| - |V(G_\Gamma)|$. Theorem 13 now follows from the following lemma.

Lemma 17 *Let Γ be a region in $\mathcal{Q}$. For every spanning subgraph $H \subset G_\Gamma$, there are exactly $2^{a(H)}$ tilings τ for which $\sigma(\tau) = H$.*

Proof: We need to show that for every spanning subgraph $H \subset G_\Gamma$, the corresponding (undirected) chain graph consists of $a(H)$ cycles. Each cycle can be oriented in two ways, hence we will get $2^{a(H)}$ valid chain graphs which correspond to H. Since chain graphs are in one-to-one correspondence with tilings, the result will follow.

Let C be a chain graph which corresponds to H. If C consists of k cycles, then it divides the plane into $k+1$ zones (possibly having holes). Each such zone is a maximal connected region on which the height function f is constant. Each zone must contain at least one type-A point, and thus must contain at least one vertex of H or $\omega(H)$. It cannot contain points from both H and $\omega(H)$, since the value of f is odd on the vertices of H and it is even on the vertices of $\omega(H)$. Observe that all vertices of H or $\omega(H)$ which live in the same zone are connected. Hence H and $\omega(H)$ have a total of $k+1$ connected components. Then $k = c(H) + c(\omega(H)) - 1 = 2c(H) + |E(H)| - |V(G_\Gamma)| = a(H)$, so the number of cycles in C is equal to $a(H)$, which proves the lemma. ■

10 Sampling of tilings

Let $\Gamma \in \mathcal{Q}$ be a quadruplicated simply-connected region. Define a Markov chain $\mathcal{M}$ whose states are T-tetromino tilings of Γ. Allow a transition from τ_1 to τ_2 if τ_1 and τ_2 differ by a 2-move or 4-move, with the probability of such a transition being $1/N$, where $N = |\Gamma|$ is the area of Γ. Observe that $N/2$ is larger than the maximum number of different local moves which can be applied to any one tiling. Now, let the probability of staying put in the state τ_1 be $1 - k/N \geq 1/2$, where k is the number of different local moves which can be applied to τ_1.

Observe that $\mathcal{M}$ is symmetric, and aperiodic since the probability of staying put is always $\geq 1/2$. Therefore, by Theorem 1, the Markov chain $\mathcal{M}$ is ergodic and converges to the uniform distribution on $\mathcal{T}_\Gamma$. The mixing time of $\mathcal{M}$ remains open, but we would like to make the following conjecture:

Conjecture 18 *The mixing time of the Markov chain $\mathcal{M}$ is polynomial in the area of Γ.*

We refer the reader to [1] for the various definitions of the mixing time of Markov chains and related results. Now, if the conjecture is true, we can use the Markov chain $\mathcal{M}$ to sample tilings $\tau \in \mathcal{T}_\Gamma$ from a nearly uniform distribution. Using the notion of self-reducibility (see introduction, [18]), we can use sampling to approximate $|\mathcal{T}_\Gamma|$. The self-reducibility of tilings follows from the following lemma.

Lemma 19 *Let $\Gamma \in \mathcal{Q}$, and consider a tiling $\tau \in \mathcal{T}_\Gamma$ chosen uniformly at random. Let S be the leftmost 4-by-4 square in the top row of Γ. Unless S is all of Γ, the probability that S is isolated in τ (covered by exactly 4 tiles) is at least 1/3 and at most 2/3.*

Proof: The 4-by-4 square S corresponds to a vertex s in G_Γ. Notice that because there is nothing to the left of S or above it, the vertex s must have degree 1 or 2 in G_Γ. The square S will be isolated if and only if no edge of $\sigma(\tau)$ is incident to s.

Case 1: Suppose s has degree 1 in G_Γ. Let e be the edge of G_Γ incident to s. Let H be a spanning subgraph of $G_\Gamma - \{s\}$. Let H_0 be the spanning subgraph of G_Γ which consists of just those edges in H, and let H_1 be the spanning subgraph of G_Γ which consists of those edges in H, plus e. Consider all tilings τ such that $\sigma(\tau)$ is either H_0 or H_1. We want to know what proportion of these tilings have $\sigma(\tau) = H_0$. Notice that $|E(H_0)| = |E(H_1)| - 1$, and $c(H_0) = c(H_1) + 1$. It follows that $a(H_0) = a(H_1) + 1$. So by Lemma 17, there will be twice as many tilings with $\sigma(\tau) = H_0$ as there are with $\sigma(\tau) = H_1$. This is true for any $H \subset G_\Gamma - \{s\}$. So upon picking a random tiling τ, the probability that e is present in $\sigma(\tau)$ is 1/3. So in this case, S is isolated with probability 2/3.

Case 2: Suppose s has degree 2 in G_Γ. Let e_1 and e_2 be the edges of G_Γ incident to s, and let t_1 and t_2 be the vertices adjacent to s along edges e_1 and e_2 respectively. Let H be a spanning subgraph of $G_\Gamma - \{s\}$. Let H_0 be the spanning subgraph of G_Γ which consists of just those edges in H, let H_1 be the graph which includes the edges of H plus e_1, let H_2 include the edges of H plus e_2, and let H_3 include the edges of H plus e_1 and e_2. Consider two subcases.

Subcase 2a: Suppose t_1 and t_2 are in different components of H. Notice that $|E(H_0)| = |E(H_1)| - 1 = |E(H_2)| - 1 = |E(H_3)| - 2$, and $c(H_0) = c(H_1) + 1 = c(H_2) + 1 = c(H_3) + 2$. So $a(H_0) = a(H_1) + 1 =$

$a(H_2) + 1 = a(H_3) + 2$. So among all tilings τ which come from one of these graphs, 4/9 of them will have $\sigma(\tau) = H_0$, 2/9 of them will have $\sigma(\tau) = H_1$, 2/9 of them will have $\sigma(\tau) = H_2$, and 1/9 of them will have $\sigma(\tau) = H_3$.

Subcase 2b: Suppose t_1 and t_2 are in the same component of H. In this case, $|E(H_0)| = |E(H_1)| - 1 = |E(H_2)| - 1 = |E(H_3)| - 2$, and $c(H_0) = c(H_1) + 1 = c(H_2) + 1 = c(H_3) + 1$. So $a(H_0) = a(H_1) + 1 = a(H_2) + 1 = a(H_3)$. So among all tilings τ which come from one of these graphs, 1/3 of them will have $\sigma(\tau) = H_0$, 1/6 of them will have $\sigma(\tau) = H_1$, 1/6 of them will have $\sigma(\tau) = H_2$, and 1/3 of them will have $\sigma(\tau) = H_3$.

Combining subcases 2a and 2b, we get the following. For any H, either 1/3 or 4/9 of the tilings which correspond to H will have S isolated. Hence when we sum over all possible graphs H, we find that between 1/3 and 4/9 of all tilings of Γ have S isolated, when s has degree 2 in G_Γ.

This proves the lemma. ■

11 Final remarks

We should mention that our chain graphs seem to be well known in the Statistical Physics literature under a name “fully-packed loop model on the square lattice”; in this case all loops have fugacity 2. We refer to [**?**] for an appearance of this model in Combinatorics literature, exact terminology and further references.

A number of questions remain for future study. First and foremost, it would be interesting to show that the mixing time of $\mathcal{M}$ is polynomial, resolving Conjecture 18. If the proof goes along similar lines as that in [9], it should lead to new interesting combinatorial notions of the “intermediate” height functions between the smallest and the largest (of a fixed region).

A related question would be to show hardness of approximation of the number of T-tetromino tilings of regions $\Gamma \in \mathcal{Q}$. For general regions and for planar bipartite regions #P results have been obtained for various evaluations of the Tutte polynomial (see [20, 22]), but for regions on a square grid much work is yet to be done (cf. [5]).

In a different direction, are there other tiling type problems which lead to, perhaps other, evaluations of the Tutte polynomial? Can the present construction be extended to coverings of certain perhaps complicated graphs with copies of $K_{1,3}$, so that their number is equal to $T(G; 3, 3)$ for general graphs G?

A more philosophical (and thus more difficult) question is to explain the meaning behind T-tetrominoes. What geometric properties of the T-tetrominoes force the rigid structure discovered by Walkup? Do all such structures imply the existence of height functions?

In general, are there other simple collections of tiles so that a certain rich collection of regions has a rigid structure of tilings, while general simply connected regions do not? Philosophically, this amounts to understanding the extent the boundary conditions control the structure of tilings in the middle. The example of domino tilings of Aztec diamonds comes to mind [3].

Finally, what can be said about the asymptotic behavior of the number a_n of T-tetromino tilings of a $4n \times 4n$ square? It is not hard to show that there exists a limit $c = \lim \log a_n/n^2$ as $n \to \infty$. It would be interesting to find upper and lower bounds on c similar to that in [11, 5].

Acknowledgements

We are grateful to the Makarychev brothers for sending us their preprint [10], and to David Karger, Jim Propp, and Van H. Vu for interesting conversations and helpful remarks. We are also thankful to referees for their comments and a suggestion to strengthen Theorem 10 to show that P_Γ is distributive.

The first author was supported by a National Defense Science and Engineering Graduate Fellowship. The second author was partially supported by NSA and NSF.

12 Appendix

12.1 The ice graph

Ice graphs are another type of directed graph which can be associated with a tiling. These graphs, and their associated height functions, provide another means of proving local connectivity for regions $\Gamma \in \mathcal{Q}$.

For a region $\Gamma \in \mathcal{Q}$, let B_Γ be the set of type-B points in Γ or $\partial\Gamma$. A directed graph on B_Γ is called an *ice graph* if it satisfies the following conditions:

- every two points which lie at opposite corners of the same block of Γ are connected with an edge, either one direction or the other, but not both, and
- every vertex has equal indegree and outdegree.

This notion has been explored by Eloranta [4] and others.

Let $\mathcal{I}_\Gamma$ denote the set of all ice graphs of a region Γ. Call a vertex *alternating* if it is incident to four edges which are oriented "in, out, in, out", in alternating order. Let $z(G)$ be the number of alternating vertices in an ice graph G.

In [10] the Makarychev brothers constructed a map $\mu : \mathcal{T}_\Gamma \to \mathcal{I}_\Gamma$ as follows.

For a tiling $\tau \in \mathcal{T}_\Gamma$, define a directed graph on B_Γ as follows. Observe that within each block, three squares belong to one T-tetromino, while one square, call it the *oddball*, belongs to a different T-tetromino. By inspection, we see that the oddball must be incident to a type-B point, rather than a type-A point. For each block, include a directed edge from the point next to the oddball square to the opposite corner of the block (see Figure 17). Define $\mu(\tau)$ to be the directed graph which results.

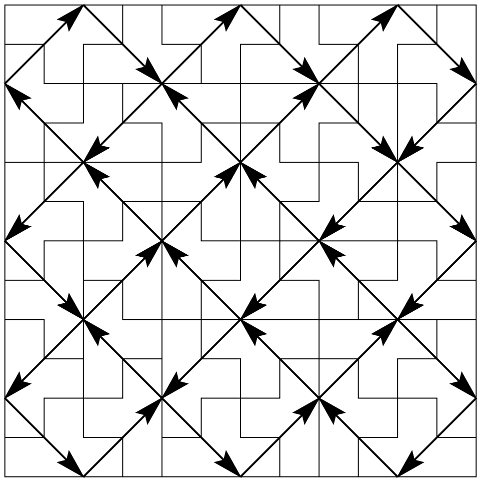


Figure 17: A tiling τ, and the ice graph $\mu(\tau)$.

Lemma 20 (K. and Y. Makarychev) *For any region $\Gamma \in \mathcal{Q}$, the map μ is a surjection from $\mathcal{T}_\Gamma$ to $\mathcal{I}_\Gamma$, in which every ice graph G is the image of $2^{z(G)}$ tilings.*

Sketch of proof: First let us show that $\mu(\tau)$ is an ice graph. Every edge connects two opposite corners of some block, so this graph will have edges in the correct places. Notice that each type-B point is adjacent to exactly two oddballs (recall Figure 5), unless the point is on $\partial\Gamma$, in which case it is adjacent to only one. Therefore, every vertex has equal indegree and outdegree. So $\mu(\tau)$ is in fact an ice graph.

Now we just need to show that every ice graph G comes from exactly $2^{z(G)}$ tilings. Take a vertex of G. If the vertex is on $\partial\Gamma$, there is only one way to place the tile which touches this vertex (see Figure 18).

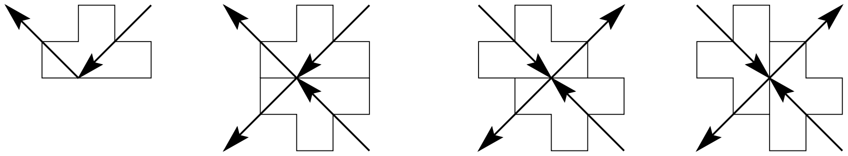


Figure 18: A boundary vertex, a nonalternating vertex, and the two options for an alternating vertex.

Similarly, if the vertex is not on the boundary, and not alternating, there is only one way to place the two tiles which touch this vertex. However, if the vertex is alternating, there are two ways to place the tiles around the vertex. The squares covered by the two tiles are the same in either case, so the decision of which one to use does not affect the rest of the tiling. Hence there are $2^{z(G)}$ ways to convert an ice graph G into a tiling. ■

Lemma 21 *If $\tau_1, \tau_2 \in \mathcal{T}_\Gamma$ are tilings such that $\mu(\tau_1) = \mu(\tau_2)$, then τ_1 and τ_2 are local-move equivalent.*

Sketch of proof: As we just saw, the only way in which these tilings may differ is in the way the tiles next to alternating points are arranged. Converting one such configuration into the other is done by performing a 2-move. Each tile is adjacent to only one type-B point, so these moves are disjoint and can be done independently of each other. So one can convert any such tiling into any other by a sequence of 2-moves. ■

12.2 Height on the ice graph

For a region $\Gamma \in \mathcal{Q}$, let A_Γ be the set of type-A points in Γ or $\partial\Gamma$. Say that a function $f : A_\Gamma \to \mathbb{Z}$ is an *ice-height function* if it satisfies the following conditions:

- $f(x) = 0$ for all points $x \in \partial\Gamma$, and
- $|f(x) - f(y)| = 1$ whenever x and y are adjacent (differ by 2 in each coordinate).

Let $\mathcal{J}_\Gamma$ denote the set of all ice-height functions of a region Γ.

Theorem 22 *For any region $\Gamma \in \mathcal{Q}$, we have $|\mathcal{J}_\Gamma| = |\mathcal{I}_\Gamma|$.*

We define a map $\nu : \mathcal{I}_\Gamma \to \mathcal{J}_\Gamma$ as follows. Let $G \in \mathcal{I}_\Gamma$ be an ice graph. Define a function f° on the faces of G by the following rules. Let f° have the value 0 on the unbounded face of G. As we pass an edge of the graph, if the edge is oriented left-to-right as we pass it, let the value of f° increase by 1. (Similarly, if the edge is oriented right-to-left, let the value of f° decrease by 1.) Now define $f : A_\Gamma \to \mathbb{Z}$ by letting $f(x)$ equal the value of f° on the face in which x lies (see Figure 19). Define $\nu(G)$ to be this function f.

Theorem 22 will follow from the following lemma.

Lemma 23 *For any region $\Gamma \in \mathcal{Q}$, the map ν is a bijection between $\mathcal{I}_\Gamma$ and $\mathcal{J}_\Gamma$.*

Proof: Let G be an ice graph, and let f be $\nu(C)$. The function f is well-defined for the same reason that the height function for the chain graph is well-defined—because every vertex has equal indegree and outdegree. It is clear that such a function meets the criteria for being an ice-height function.

From an ice-height function f, one can reconstruct the ice graph $G = \nu^{-1}(f)$ by directing every edge so the face with greater height is on the left. Since the net change in height going around any vertex is 0, every vertex will have equal indegree and outdegree, thus the graph so constructed will be a valid ice graph. ■

For ease of notation, define $\xi(\tau) = \nu(\mu(\tau))$. For a region $\Gamma \in \mathcal{Q}$, the map ξ is the canonical bijection between $\mathcal{T}_\Gamma$ and $\mathcal{J}_\Gamma$.

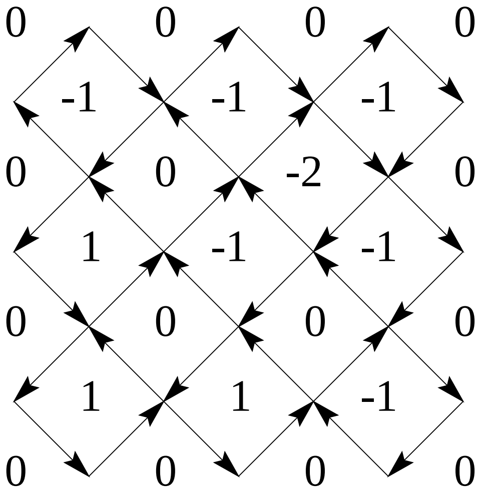


Figure 19: An ice graph G, and the function $f = \nu(G)$.

Lemma 24 *Let $\Gamma \in \mathcal{Q}$ and let $\tau_1, \tau_2 \in \mathcal{T}_\Gamma$ be tilings of Γ. If the tilings τ_1 and τ_2 differ by a 2-move, then $\xi(\tau_1) = \xi(\tau_2)$. If the tilings τ_1 and τ_2 differ by a 4-move, then $\xi(\tau_1)$ and $\xi(\tau_2)$ differ by 2 on some point, and are the same everywhere else. If f_1 and f_2 are ice-height functions which differ by 2 on some point and are the same everywhere else, then there exist tilings τ_1 and τ_2 such that $\xi(\tau_1) = f_1, \xi(\tau_2) = f_2$, and τ_1 and τ_2 differ by a 4-move.*

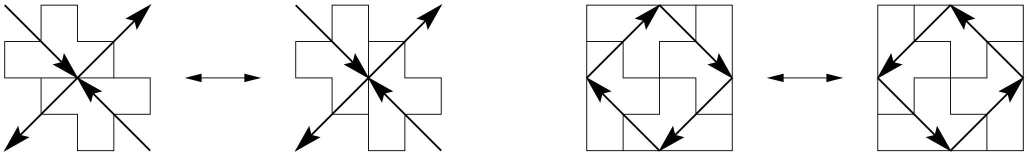


Figure 20: The effect of local moves on the ice graph.

Sketch of proof: A 2-move can only occur at an alternating type-B point, so if τ_1 and τ_2 differ by a 2-move, then $\mu(\tau_1) = \mu(\tau_2)$, so $\xi(\tau_1) = \xi(\tau_2)$ (see Figure 20).

If τ_1 and τ_2 differ by a 4-move, then $\mu(\tau_1)$ and $\mu(\tau_2)$ differ by the reversal of a directed 4-cycle, thus $\xi(\tau_1)$ and $\xi(\tau_2)$ will differ by 2 on the point inside that 4-cycle, and be the same everywhere else.

Now suppose f_1 and f_2 are ice-height functions such that $f_1(x) = h + 1$ and $f_2(x) = h - 1$, but $f_1 = f_2$ everywhere else. We must then have $f_1(y) = f_2(y) = h$ for the neighbors y of x. So x will be surrounded by a counterclockwise directed 4-cycle in the ice graph corresponding to f_1, and a clockwise directed 4-cycle in the ice graph corresponding to f_2. The problem is that a tiling which corresponds to f_1 may look like the left side of Figure 21. However, in such a case, there is always another tiling (which differs from the original by some 2-moves) such that a 4-move can be applied. ■

For ice-height functions $f_1, f_2 \in \mathcal{J}_\Gamma$, say that f_1 and f_2 differ by a 4-move if there exist tilings $\tau_1, \tau_2 \in \mathcal{T}_\Gamma$ which differ by a 4-move such that $\xi(\tau_1) = f_1$ and $\xi(\tau_2) = f_2$. By the previous Lemma, one can see that performing a 4-move on an ice-height function f is equivalent to increasing or decreasing its value by 2 at some point. Of course, such a move may only be applied if the function that results is a valid ice-height function.

Notice that for tilings τ_1 and τ_2, having $\xi(\tau_1)$ and $\xi(\tau_2)$ differ by a 4-move does not imply that τ_1 and τ_2 differ by a 4-move. However, it does imply that there exist tilings τ_1' and τ_2' which differ by a 4-move such

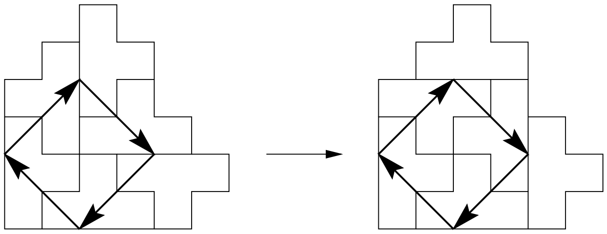


Figure 21: A tiling where a 4-move cannot be applied, and one where it can.

that $\xi(\tau_1') = \xi(\tau_1)$ and $\xi(\tau_2') = \xi(\tau_2)$. It then follows, by Lemmas 21 and 23, that τ_1 is local-move equivalent to τ_1' and τ_2 is local-move equivalent to τ_2'. Hence τ_1 and τ_2 will be local-move equivalent whenever $\xi(\tau_1)$ and $\xi(\tau_2)$ differ by a 4-move, or more generally, by a sequence of 4-moves.

Theorem 1 will now easily follow from the following Lemma.

Lemma 25 *Let $\Gamma \in \mathcal{Q}$, and let $f_1, f_2 \in \mathcal{J}_\Gamma$ be ice-height functions. It is always possible to convert f_1 into f_2 by performing a sequence of 4-moves.*

Proof: For any region, there will be a unique ice-height function f_0 whose value at each point is either 0 or 1. (Each face is either "even" or "odd", depending on how many steps from the exterior it is, thus each even face will have the value 0, and each odd face will have the value 1.) It will be sufficient to show that any ice-height function f can be transformed into f_0. Suppose $f(x) > 1$ for some point x. Let x be the point where f attains its largest value, call it h (if there are several possible points, choose any one). We must then have $f(y) = h - 1$ for the neighbors y of x. Thus we can perform a 4-move, and decrease $f(x)$ to $h - 2$. Repeat this process until f attains no values greater than 1. Now if there are points x where $f(x) < 0$, find the one where f attains its minimum. We can perform a 4-move to increase $f(x)$ by 2. We repeat this until f attains no values less than 0. Now $0 \leq f(x) \leq 1$ for all x, so we are done. ■

References

[1] D. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graphs.* monograph in preparation.

[2] T. Chaboud. Domino tiling in planar graphs with regular and bipartite dual. *Theoret. Comput. Sci.*, 159(1):137–142, 1996. Selected papers from the "GASCOM '94" (Talence, 1994) and the "Polyominoes and Tilings" (Toulouse, 1994) Workshops.

[3] H. Cohn, N. Elkies, and J. Propp. Local statistics for random domino tilings of the Aztec diamond. *Duke Math. J.*, 85(1):117–166, 1996.

[4] K. Eloranta. Diamond ice. *J. Statist. Phys.*, 96(5-6):1091–1109, 1999.

[5] M. E. Gegúndez, A. Márquez, and P. Revuelta. The Tutte polynomial on the square lattice. (preprint).

[6] S. W. Golomb. *Polyominoes.* Georges Allen and Unwin Ltd, London, 1966.

[7] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18(6):1149–1178, 1989.

[8] C. Kenyon and R. Kenyon. Tiling a polygon with rectangles. *Proc. 33-rd FOCS*, pages 610–619, 1992.

[9] M. Luby, D. Randall, and A. Sinclair. Markov chain algorithms for planar lattice structures. *SIAM J. Comput.*, 31(1):167–192 (electronic), 2001.

[10] K. Makarychev and Y. Makarychev. Proof of Pak's conjecture on tilings by T-tetrominoes (in Russian), `http://home.earthlink.net/~makarychev/tetromino/tetro.htm`.

[11] C. Merino and D. J. A. Welsh. Forests, colorings and acyclic orientations of the square lattice. *Ann. Comb.*, 3(2-4):417–429, 1999.

[12] C. Moore, I. Rapaport, and E. Remila. Tiling groups for wang tiles. *Proc. 13-th SODA*, pages 402–411, 2002.

[13] I. Pak. Tile invariants: New horizons. *Theor. Comp. Sci.* (to appear).

[14] J. Propp. Generating random elements of finite distributive lattices. *Electron. J. Combin.*, 4(2):Research Paper 15, approx. 12 pp. (electronic), 1997. The Wilf Festschrift (Philadelphia, PA, 1996).

[15] J. Propp and D. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, volume 9, pages 223–252, 1996.

[16] J. Propp and D. Wilson. Coupling from the past: a user's guide. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 181–192. Amer. Math. Soc., Providence, RI, 1998.

[17] E. Rémila. Tiling groups: new applications in the triangular lattice. *Discrete Comput. Geom.*, 20(2):189–204, 1998.

[18] Alistair Sinclair. *Algorithms for random generation and counting.* Progress in Theoretical Computer Science. Birkhäuser Boston Inc., Boston, MA, 1993.

[19] W. Thurston. Conway's tiling groups. *Amer. Math. Monthly*, 97(8):757–773, 1990.

[20] D. L. Vertigan and D. J. A. Welsh. The computational complexity of the Tutte plane: the bipartite case. *Combin. Probab. Comput.*, 1(2):181–187, 1992.

[21] D. W. Walkup. Covering a rectangle with T-tetrominoes. *Amer. Math. Monthly*, 72:986–988, 1965.

[22] D. J. A. Welsh. *Complexity: knots, colourings and counting*, volume 186 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge, 1993.

Note:

After the paper was finished there has been several subsequent developments. First, Michel Las Vergnas established a connection between our Theorem 13 and his results in the paper *"On the evaluation at* $(3,3)$ *of the Tutte polynomial of a graph"*, J. Combin. Theory Ser. B, vol. 45 (1988), 367–372. The authors were not aware of this paper, but the connections is in the same spirit as the appendix.

The authors later generalized and extended Theorem 13 in this paper to plane and ribbon graphs in the recent preprint *"Combinatorial evaluations of the Tutte polynomial"*, which is available from:
`http://www-math.mit.edu/~pak/research.html`

TILE INVARIANTS: NEW HORIZONS

Igor Pak

Department of Mathematics
MIT
Cambridge, MA 02139 USA
E-mail: pak@math.mit.edu

December 20, 2000

Abstract. Let $\mathbf{T}$ be a finite set of tiles. The group of invariants $\mathbb{G}(\mathbf{T})$, introduced by the author [P], is a group of linear relations between the number of copies of tiles in tilings of the same region. We survey known results about $\mathbb{G}$, the height function approach, the local move property, various applications and special cases.

Introduction

The problem of tileability of a region is very old, and in many instances computationally hard, even for small sets of tiles (see e.g. [MR,Ro]). The subject of this paper is different, although not unrelated. We study a *group of invariants* $\mathbb{G} = \mathbb{G}(\mathbf{T})$, associated with a set of tiles $\mathbf{T}$. This notion was introduced in [P], and further studied in [MuP,MoP]. The elements of $\mathbb{G}$ correspond to linear relations for the number of copies of tiles used in different tiling of every fixed region Γ. Turns out, this group has various nice properties, and in certain special cases can be fully computed.

In this paper we survey much of what is known about $\mathbb{G}$, the basic algebraic properties, some complexity results, as well as some applications and special cases. We describe some examples when coloring arguments do not suffice, while a different technique can be applied. A number of results never appeared before; their proofs will be sketched. We also include conjectures and open problems for further study.

Rather than define the group of invariants here, let us discuss a small but very interesting example of domino tilings, which was one of our motivations. Denote by τ_1, τ_2 the vertical and horisontal domino tiles, and let $\mathbf{T} = \{\tau_1, \tau_2\}$. Let Γ be a connected region on a square grid. The problem of tileability of Γ by $\mathbf{T}$ corresponds to finding a perfect matching in a dual graph, so it can be solved in polynomial time [LP].

Now, let A be a tiling of Γ by dominoes. Denote by $\alpha_1(A)$, $\alpha_2(A)$ the number of times tiles τ_1, τ_2 appear in A. Clearly, $\alpha_1(A) + \alpha_2(A) = |\Gamma|/2$, which follows

Key words and phrases. Polyomino tilings, tile invariants, Conway group, undecidability, height function.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

from the area consideration. Also, one can show that $\alpha_1(A) = \text{const}(\Gamma) \mod 2$, where the const depends only on the region Γ, and not on the tiling. This follows from a simple coloring argument [P]. We call the linear relations as above the *tile invariants.* In general, tile invariants are the linear relations of the type

$$(*) \quad c_1\, \alpha_1(A) + c_2\, \alpha_2(A) + \, \dots \equiv \text{const}(\Gamma) \mod m,$$

where the const(Γ) depends only on the region Γ, and not on the tiling A of Γ; $c_i \in \mathbb{Z}$, and $m = \infty$ is allowed. The group $\mathbb{G}(\mathbf{T})$ can be defined as the group of such invariants, with addition as a group operation (the precise definition will be given in section 1). In the case of dominoes, the group of invariants is $\mathbb{G}(\mathbf{T}) = \mathbb{Z} \times \mathbb{Z}_2$, generated by the two invariants described above.

Our goal is to determine the group of invariants, and compute it in some special cases. For example, as in the case of dominoes, tile invariants can often be derived from certain colorings of the squares. In section 1 we follow [P] and introduce the *group of valuations* $\mathbb{E} \subset \mathbb{G}$, closely related to the extended coloring arguments. As we mentioned above, in general not all tile invariants can be obtained by the extended coloring arguments. This difference can be underscored by the complexity results. We show that in general case computing $\mathbb{G}$ is NP-hard, and even undecidable when considered on the whole plane. At the same time, $\mathbb{E}$ can be determined in polynomial time (see section 3.)

Now, if the group $\mathbb{G}(\mathbf{T})$ is computed, one can use it to obtain criteria for tileability of regions Γ tileable by $\mathbf{T}$ with a proper subset $\mathbf{T}'$ of tiles. Indeed, in this case the number of times α_i the tiles $\tau_i \in \mathbf{T}'$ can occur in the tiling of Γ must satisfy a number of linear relations. Existence of integral solution of these relations gives a tileability criteria. This approach was pioneered in [CL] and later successfully used in [P] to obtain tileability results which cannot be proved by coloring arguments (see section 9.)

The difficulty with the group of invariants is proving that a suspected relation is indeed a tile invariant. At the moment we see only two ways of proving such a result. The first has to do with the *local move property.* Recall that one can obtain any domino tiling A_1 of a simply connected region Γ to any other domino tiling A_2 of Γ by a sequence of 2 × 2 moves (see e.g. [LP,T].) Now, in general, it suffices to check that a given relation is preserved by such moves. In fact, one can easily compute the whole group of invariants in this case (see section 4.)

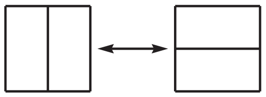


FIGURE 0.1. Local 2 × 2 move.

Unfortunately, very few sets of tiles have a finite number of local moves. For example, even for dominoes in three dimensions there exist infinitely many principally different simply connected regions which have exactly two domino tilings. In the other direction, even when we believe that there exist a finite number of local moves, even when we conjecture we know them all, the problem of proving this claim may be very hard.

The second and the most successful at the moment approach is based on the notion of height function, and was inspired by the Conway group [CL] and Thurston's article [T]. Roughly, Thurston defined a function from edges in the grid into a line, which maps tileable regions into loops. This approach is useful for proving local move property and finding new tile invariants [T,CL]. In the case of domino tilings, Thurston's height functions proves the connectivity of tilings by the 2×2 moves. It also gives a remarkable linear time algorithm for testing tileability of simply connected regions [Ch,F]. In sections 4, 5 we present general conditions for the technique to succeed.

While our exposition is somewhat brief due to the space limitations, we include a large number of examples and references when the techniques in the survey were successfully applied to various tiling problems. Among others, we present a final result of computation of the ribbon tile invariants [MoP], started earlier in [CL,MuP,P1] (see section 6). We also go at length to describe the Generalized Sperner's Lemma which can also be defined as a tile invariant for a special set of tiles (section 8.1). We conclude with the heuristic method for study of general set of tiles.

Many results are only stated in the main body of the paper. We sketch the proofs of new results in section 10.

Acknowledgements.

We would like to thank David Ingerman, Ezra Miller, Cris Moore, Roman Muchnik, Jim Propp and Richard Stanley for stimulating conversations and encouragement. Without them this work would never be written. We are also grateful to Mike Sipser, Dan Spielman and Santosh Vempala for help with complexity questions.

1. Basic definitions

The most general tiling problem can be formulated as follows. Let Λ be a finite or infinite set, and let $\mathcal{B}$ be a collection of finite subsets, which we call *regions*. Let '$\sim$' be an equivalence relation on $\mathcal{B}$. We will assume that '$\sim$' preserves size (the number of elements in the region). Finally, let $\mathbf{T}$ be a finite subset of $\mathcal{B}$ (the *set of tiles*). Denote by $\widetilde{\mathbf{T}}$ the set of regions $\tau \in \mathcal{B}$ such that $\tau \sim \tau' \in \mathbf{T}$. We assume that $\tau \nsim \tau'$, for all $\tau, \tau' \in \mathbf{T}$.

A typical example is a square grid $\Lambda = \mathbb{Z}^2$ with a set of simply connected regions $\mathcal{B}$ and translation equivalence '$\sim$'. Note that we view tiles here as subsets of squares, for example dominoes correspond to pairs of adjacent squares in the grid.

The problem of tileability by the set of tiles $\mathbf{T}$ is a decision whether a given set $\Gamma \in \mathcal{B}$ can be presented as a disjoint union of regions in $\widetilde{\mathbf{T}}$: $\Gamma = \sqcup \tau_i$, where $\tau_i \in \widetilde{\mathbf{T}}$ for all i. We denote such tilings by A and write $A \vdash \Gamma$. This problem is hard even in some very simple special cases, and will not be studied in this paper. Instead, we will study an abelian group $\mathbb{G}(\mathbf{T}, \mathcal{B})$ which can be defined as follows.

Let $\mathbf{T} = \{\tau_1, \dots, \tau_k\}$ be the set of tiles, where $k = |\mathbf{T}|$. For every tiling A of a region $\Gamma \in \mathcal{B}$ denote by $\alpha_i(A)$ the number of tiles $\tau \in A$ such that $\tau \sim \tau_i$. Now let

$$\mathbb{G}(\mathbf{T}, \mathcal{B}) = \mathbb{Z}^k / \mathbb{Z}\big\langle \big(\alpha_1(A) - \alpha_1(A'), \dots, \alpha_k(A) - \alpha_k(A')\big), \, \forall \Gamma \in \mathcal{B}, \, \forall A, A' \vdash \Gamma \big\rangle,$$

where on the right hand side we have a subgroup of k-vectors with A, A' any two tilings by $\mathbf{T}$ of the same region $\Gamma \in \mathcal{B}$. This is a *group of invariants*, the main subject of this paper. The elements of $\mathbb{G}(\mathbf{T}, \mathcal{B})$ are called *tile invariants.*

In general, $\mathbb{G}(\mathbf{T}, \mathcal{B})$ may depend heavily on the set of regions (all regions vs. simply connected regions) as well as a set of tiles (adding one tile may destroy most of the tile invariants). Note also that if $\mathcal{B}_1 \subset \mathcal{B}_2$, then $\mathbb{G}(\mathbf{T}, \mathcal{B}_1) \supset \mathbb{G}(\mathbf{T}, \mathcal{B}_2)$. Similarly, if $\mathbf{T}_1 \subset \mathbf{T}_2$, then

$$\mathbb{G}(\mathbf{T}_2, \mathcal{B}) \subset \mathbb{G}(\mathbf{T}_1, \mathcal{B}) \times \mathbb{Z}^{|\mathbf{T}_2| - |\mathbf{T}_1|}.$$

Define a *coloring group*

$$\mathbb{O}(\mathbf{T}) = \mathbb{Z}^\Lambda / \mathbb{Z}\langle x_1 + \cdots + x_r = 0, \; \forall \tau = \{x_1, \ldots, x_r\} \in \widetilde{\mathbf{T}} \rangle.$$

One can think of elements of $\mathbb{O}$ as of functions $f : \Lambda \to \mathbb{Z}$, such that $f(\Gamma) = \sum_{x \in \Gamma} f(x)$, and $f(\tau) = 0$ for all $\tau \in \widetilde{\mathbf{T}}$. The function f is called a *coloring map.* Before recently, coloring maps were the main tool to prove untileability [G]. Indeed, if $f(\Gamma) \neq 0$, this immediately implies that Γ is not tileable by $\mathbf{T}$. In this case we say that a *coloring argument* f rejects tileability of Γ. Let us add that any map $f : \Lambda \to G$, where G is abelian, can obtain from the above functions. In other words, if any coloring arguments $f : \Lambda \to G$ rejects tileability of Γ, for some abelian group G, it also rejects tileability for some $f : \Lambda \to \mathbb{Z}_m$.

Now, define an *extended coloring group*

$$\overline{\mathbb{O}}(\mathbf{T}) = \mathbb{Z}^\Lambda / \mathbb{Z}\langle x_1 + \cdots + x_r = y_1 + \cdots + y_r \rangle,$$

where $\tau = \{x_1, \ldots, x_r\}$, $\tau' = \{y_1, \ldots, r_r\}$, and $\tau \sim \tau' \in \widetilde{\mathbf{T}}$. Clearly, $\mathbb{O}(\mathbf{T}) \subset \overline{\mathbb{O}}(\mathbf{T})$. One can think of the elements of $\overline{\mathbb{O}}(\mathbf{T})$ as of functions $f : \Lambda \to \mathbb{Z}$, which are constant on equivalent tiles in $\widetilde{\mathbf{T}}$. We call such functions an *extended coloring maps.*

There is a natural map $\nu : \overline{\mathbb{O}}(\mathbf{T}) \to \mathbb{Z}^{\mathbf{T}}$ which maps the functions to their values on tiles in $\mathbf{T}$. We have $\mathbb{O}(\mathbf{T}) = \nu^{-1}(0)$. By definition, the value $f(\Gamma)$ of a function in $\overline{\mathbb{O}}(\mathbf{T})$ is independent on the tiling by $\mathbf{T}$, so ν extends to the quotient group $\mathbb{G}(\mathbf{T})$. Denote by $\mathbb{E}(\mathbf{T})$ the image of ν in $\mathbb{G}(\mathbf{T})$. We call $\mathbb{E}(\mathbf{T})$ the *group of valuations* of the set of tiles $\mathbf{T}$. From above,

$$\mathbb{E}(\mathbf{T}) \simeq \overline{\mathbb{O}}(\mathbf{T}) / \mathbb{O}(\mathbf{T}).$$

By definition, the subgroup $\mathbb{E}(\mathbf{T}) \subset \mathbb{G}(\mathbf{T})$ consists of all tile invariants which follow from the extended coloring maps.

Computing the coloring group and the group of valuations is of interest, so as to see which tileability criteria and which group invariants are "easy to obtain".

Unless stated otherwise, for the rest of the paper we will assume that $\Lambda \subseteq \mathrm{Z}^2$, where Z^2 denotes the square grid with elements - 1×1 squares. Denote by $\mathcal{B}$, $\mathcal{B}_{\mathrm{sc}}$, $\mathcal{B}_N$ the set of all regions, of all simply connected regions, and the set of regions in $N \times N$ square. The equivalence relation consists of parallel translations of the regions (no rotation or reflection is allowed). Let the set of tiles $\mathbf{T}$ consist of some k tiles, each of size $\leq R$. By abuse of notation, we use $\tau \in \mathbf{T}$ to denote $\tau \in \widetilde{\mathbf{T}}$.

The main questions of this paper can be stated as follows:

Group of Invariants Problem (GI) :
Given $\mathbf{T} \subset \mathrm{Z}^2$, compute $\mathbb{G}(\mathbf{T}, \mathcal{B})$ (or $\mathbb{G}(\mathbf{T}, \mathcal{B}_{\text{sc}})$, $\mathbb{G}(\mathbf{T}, \mathcal{B}_N)$).

Tileability Problem (T) :
Given $\mathbf{T} \subset \mathrm{Z}^2$, $\Gamma \in \mathcal{B}$ (or $\mathcal{B}_{\text{sc}}$, $\mathcal{B}_N$), decide whether Γ is tileable by $\mathbf{T}$.

Group of Valuations Problem (GV) :
Given $\mathbf{T} \subset \mathrm{Z}^2$, compute $\mathbb{E}(\mathbf{T})$.

Coloring Group Problem (CG) :
Given $\mathbf{T} \subset \mathrm{Z}^2$, compute $\mathbb{O}(\mathbf{T})$.

The last two problems are very much related, but we decided to separate them for convenience.

We say that a tile invariant is *finite* (*infinite*) if the order of the element in $\mathbb{G}$ is finite (infinite). Using definition $(*)$ in the introduction, the invariant is infinite if $m = \infty$. We will come back to tile invariants in the next section.

Remark 1.1 Much of this survey can be understood with conventional definitions of the tilings on a square grid. The point of this somewhat overgeneralized section was to introduce the general concepts and notation we use throughout the paper, as well as to prepare the reader to possible extensions and generalizations. While much of the results in the paper can be generalized by verbatim, we decided to keep the presentation simple for the sake of clarity. At the same time we hope that after reading this section the reader is fully equipped to generalize the results to any appropriate level.

Remark 1.2 One should keep in mind that the tile invariants were implicitly introduced in [CL] in order to obtain new tileability criteria. Although we downplay the connection in this paper, the results that are obtained in this direction can be judges as the most unexpected. See section 9 for for details.

2. Algebraic aspects

Fix a set of tiles $\mathbf{T} = \{\tau_1, \ldots, \tau_k\} \subset \mathrm{Z}^2$. Consider $\mathbb{G} = \mathbb{G}(\mathbf{T}, \mathcal{B})$. Since $\mathbb{G}$ is abelian, it can be presented as

$$\mathbb{G} \simeq \mathbb{Z}^r \times (\mathbb{Z}_2)^{m_2} \times (\mathbb{Z}_3)^{m_3} \times \cdots \times (\mathbb{Z}_{p^c})^{m_{p^c}} \times \ldots,$$

where $r \leq k$ is called the *free rank* of $\mathbb{G}$, denotes $\text{rk}(G)$, and $\mathbb{Z}^{\text{rk}} \subset \mathbb{G}$ is called the *free subgroup* of $\mathbb{G}$. Similarly, denote by $M = \sum_{q=p^c} m_q$ the *torsion rank* of $\mathbb{G}$, and $\mathbb{T} = (\mathbb{Z}_2)^{m_2} \times (\mathbb{Z}_3)^{m_3} \times \ldots \subset \mathbb{G}$ is called the *torsion subgroup* of $\mathbb{G}$. By construction, the torsion subgroup is always finite.

Proposition 2.1 *For N sufficiently large, we have* $\mathbb{G}(\mathbf{T}, \mathcal{B}_N) = \mathbb{G}(\mathbf{T}, \mathcal{B})$.

Sketch of proof. Consider a sequence of subgroups $\mathbb{G}_N = \mathbb{G}(\mathbf{T}, \mathcal{B}_N)$. Recall that $\mathbb{G}_N \supset \mathbb{G}_{N+1}$. By Hilbert Basis Theorem, this sequence stabilizes. □

Now let us turn to signed tilings and the coloring group. Denote by $\chi(\Gamma) \in \mathbb{R}^\Lambda$ the characteristic function of a region Γ. One can think of a tiling of Γ by $\mathbf{T}$ as of decomposition $\chi(\Gamma) = \chi(\tau) + \chi(\tau') + \dots$, where $\tau, \tau', \dots \in \mathbf{T}$. The signed tiling is similar decomposition, where each tile is used with a positive or negative sign. Note that the notion of the coloring argument extends to signed tilings as well.

Theorem 2.2 [P] *A region Γ has a signed tiling by $\mathbf{T}$ if and only if there is no coloring argument which would reject tileability.*

Sketch of proof. Note that signed tilings by $\mathbf{T}$ form a group $\mathbb{S}(\mathbf{T})$, with addition as an operation. By definition, we have $\mathbb{O}(\mathbf{T}) = \mathbb{Z}^{\mathbf{T}}/\mathbb{S}(\mathbf{T})$, which is a reformulation of the result. □

Similarly to the coloring arguments, consider the extended coloring arguments for signed tilings. Define $\mathbb{E}_\circ(\mathbf{T}) = \mathbb{E}(\mathbf{T} \cup -T)$, where $-\mathbf{T}$ contains the negative tiles $-\tau$, with $\chi_{-\tau} = -\chi_\tau$. We claim that

$$\mathbb{E}_\circ(\mathbf{T}) \simeq \mathbb{E}(\mathbf{T}).$$

Indeed, let $f : \Lambda \to Z$ be any extended coloring map. Since $\chi_{-\tau} + \chi_\tau = 0$, we have $f(-\tau) = -f(\tau)$ and thus $\mathbb{E}_\circ(\mathbf{T}) \subset \mathbb{E}(\mathbf{T})$. On the other hand, $\mathbb{E}(\mathbf{T}) \subset \mathbb{E}_\circ(\mathbf{T})$ since every extended coloring map by definition corresponds to an extended coloring map for signed tiles $\mathbf{T} \cup -\mathbf{T}$, and therefore defines a proper valuation on $\mathbf{T} \cup -\mathbf{T}$.

An interesting class of tile invariants are the *abelian invariants*, which are defined as tile invariants which remain invariants for signed tilings. Define *group of abelian invariants* $\mathbb{A}(\mathbf{T}) = \mathbb{G}(\mathbf{T} \cup -\mathbf{T})$. From above, we conclude that $\mathbb{E}(\mathbf{T}) \subset \mathbb{A}(\mathbf{T})$. In fact, this is an identity:

Theorem 2.3 $\mathbb{A}(\mathbf{T}) = \mathbb{E}(\mathbf{T})$. □

The real meaning of Theorem 2.3 can be seen in the following observation. If for some reason we have an abelian invariant, we can conclude that there exists a coloring map which defines it. In practice, finding such coloring map can be complicated. We leave the proof to the reader.

3. Complexity aspects

It is well known that the tileability problem is NP-complete when Γ is finite [GJ]. It is also undecidable when Γ is the whole plane [Be,Ri]. We shall prove that the similar situation holds for GI Problem. But first we need to state it as a decision problem.

GI-rank Problem: Given $\mathbf{T}$, r, decide whether $\operatorname{rk}\mathbb{G}(\mathbf{T}, \mathcal{B}) \geq r$.

Bounded GI-rank Problem: Given $\mathbf{T}$, r, N, decide whether $\operatorname{rk}\mathbb{G}(\mathbf{T}, \mathcal{B}_N) \geq r$.

Theorem 3.1 *The GI-rank Problem is undecidable. Similarly, the Bounded GI-rank Problem is* NP*-hard.*

The proof is given below in section 10. Roughly, Theorem 3.1 implies that computationally GI is intractible. A simple check shows that Theorem 3.1 extends to simply connected regions as well (i.e. computing the rank of $\mathbb{G}(\mathbf{T}, \mathcal{B}_{\text{sc}})$ is also undecidable). It seems likely that the proof can be modified to show that computing any of the exponents m_p in the torsion group is also undecidable.

Now, let us fix the set of tiles $\mathbf{T}$. Recall that $\operatorname{rk}(\mathbb{G}) \le |\mathbf{T}|$. Proposition 2.1 implies that the negative answer to the Bounded GI-rank Problem can be obtained by an exhaustive search for some finite $N = N(\mathbf{T})$. In other words, a sequence of Bounded GI-rank Problems is in co-NP (as N grows). The certificate for $\operatorname{rk}(\mathbb{G}) < r$ is a collection of $l > n - r$ bounded regions Γ_i, $1 \le i \le l$, and two collections of tilings $A_i, A_i' \vdash \Gamma_i$, such that

$$\operatorname{rk} \mathbb{Z}\big\langle\big(\alpha_1(A_i) - \alpha_1(A_i'), \dots, \alpha_k(A_i) - \alpha_k(A_i')\big), \, i = 1 \dots l\big\rangle \, > \, n - r.$$

In a way this makes it unlikely that there is a good generic way to establish the tile invariants for general sets of tiles. For example, if height functions exist for a given set of tiles, this puts the Bounded GI-rank Problem into NP. However, it is believed that an NP-hard problem cannot be in NP $\cap$ co-NP [GJ]. We will not attempt to formalize and extend this observation.

For the signed tilings, one can define the Signed Tileability Problem (ST) by analogy. Observe that Theorem 2.2 can be used now to establish the certificates for $\operatorname{rk}(\mathbb{O}) \ge r$, $m_p(\mathbb{O}) \ge m$. Using the logic as above one would conclude that ST and CG must have efficient solutions. This is true indeed.

Bounded CG-rank Problem: Given $\mathbf{T}$, r, N, decide whether $\operatorname{rk} \mathbb{O}(\mathbf{T}, \mathcal{B}_N) \ge \, r$.

Bounded GV-rank Problem: Given $\mathbf{T}$, r, N, decide whether $\operatorname{rk} \mathbb{E}(\mathbf{T}, \mathcal{B}_N) \ge \, r$.

Theorem 3.2 *Bounded CG-rank Problem and Bounded GV-rank Problem are in* P.

The proof is based on a simple reduction to a linear algebra problem, and is given in section 10. We believe that currently known algorithms for solving linear equations over the integers (see [BK,LLL,Sc]) can be used to determine the full groups $\mathbb{O}(\mathbf{T}, \mathcal{B}_N)$, $\mathbb{E}(\mathbf{T}, \mathcal{B}_N)$. Further, we conjecture that there exist an efficient algorithm for computing $\mathbb{O}(\mathbf{T}, \mathcal{B})$, $\mathbb{E}(\mathbf{T}, \mathcal{B})$. We hope to return to this problem in the future.

4. Height functions

There seem to be no general agreement as to what exactly is the method of height functions, especially when dimension increases. Here we present our personal approach with no attempt to justify it.

Suppose $\mathbf{T}$ is a fine set of tiles of the plane Z^2, or any other plane graph L with straight edges for that matter (for example L can be triangular of hexagonal lattice). Let V be a different plane, which will also be fixed. Suppose the edges of L are oriented, and there is a function $\varphi : L \to V$ which maps oriented edges into vectors in V. Also, let $\varphi(x, y) = -\varphi(y, x)$ for all edges $(x, y) \in L$ oriented from y

to x. Now, every path $x_1 \to x_2 \to x_3 \to \dots$ can be mapped to a path in V (up to translation): $v_1 \to v_2 \to v_3 \to \dots$, where $v_{i+1} - v_i = \varphi(x_i, x_{i+1})$. We think about the image of the path on a graph as a polygon in V with straight edges.

The function φ is called a *height function* if the following condition is satisfied:

($\star$) *For every simply connected region* Γ *tileable by a set of tiles* $\mathbf{T}$*, the image* $\varphi(\partial\Gamma)$ *is a closed loop.*

Here the boundary $\partial\Gamma$ is a closed path with any fixed starting point and oriented counterclockwise. We will always assume that there is a finite number of equivalence classes of values $\varphi(x, y)$ for all $(x, y) \in L$. The condition ($\star$) may seem difficult to check, so the following result helps to simplify it.

Theorem 4.1 *It suffices to check* ($\star$) *only for the tiles* $\tau \in \mathbf{T}$.

The theorem follows easily by induction from the following lemma of independent interest.

Lemma 4.2 *Let* $\Gamma \subset \mathbb{R}^2$ *be a simply connected region and is tiled by simply connected regions* $\tau_1, \dots, \tau_k$*. Then there exist* i *such that* $\Gamma - \tau_i$ *is also simply connected.*

Lemma 4.2 seems to be well known in geometric group theory, although we were unable to obtain any reference to that. In this context it was sketched in the pioneer paper [CL]. A simple proof can be found in [MP] (see also [Pr]).

Let us remark that in 3 and more dimensions Lemma 4.2 as stated is incorrect[1]. On the other hand, proof of Theorem 4.1 requires a result somewhat weaker that that in the lemma. For example, one can change the statement to *"there exist* $i_1, \dots, i_l$ *such that regions* $\tau_{i_1} \cup \dots \cup \tau_{i_l}$ *and* $\Gamma - \left(\tau_{i_1} \cup \dots \cup \tau_{i_l}\right)$ *are simply connected*[2]*"*. We do not believe that even this weaker condition holds. It would be interesting to find an explicit counterexample to that.

Now, once the height function is given, it can be used to prove certain tile invariants for the set of tiles $\mathbf{T}$, not unlike the extended coloring arguments. Indeed, consider any extended coloring argument $f : V \to G$ (G is abelian), where now we require the value $f(\varphi(\tau))$ to be invariant of the location of the τ on the plane. By construction, $f(\varphi(\Gamma))$ is always the sum of the $f(\varphi(\tau_i))$ and is independent of the tiling. Therefore the values $c_i = f(\varphi(\tau_i))$, $\tau \in \mathbf{T}$ define a tile invariant for $\mathbf{T}$.

Formally, denote by $\mathbb{E}_\varphi(\mathbf{T})$ the group of valuations of extended coloring arguments on V for the set of tiles $\varphi(\tau_i)$. Then

$$(**) \quad \mathbb{E}_\varphi(\mathbf{T}) \subseteq \mathbb{G}(\mathbf{T}, \mathcal{B}_{\text{sc}}).$$

This means that in certain cases when there exists a height function, one can obtain proofs of certain tile invariants by finding an appropriate extended coloring

[1] A counterexample is a family of six blocks which form a three dimensional cross shape figure, and is hard to disassemble. In this case no block can be removed without the remaining union of five blocks having a hole inside. Versions of this puzzle can be often found in toy stores.

[2] Actually, we need a slightly stronger condition on the intersection of the two simply connected parts.

argument in V. In other words, one can sometimes compute the whole group of invariants $\mathbb{G}(\mathbf{T}, \mathcal{B}_{\text{sc}})$.

We should note here that condition $(\star)$ does not necessarily imply that $\varphi(A)$, $A \vdash \Gamma$ is a tiling of Γ with tiles $\varphi(\tau_i)$[3]. Rather, we obtain a signed tiling of $\varphi(\Gamma)$. Still, the conclusion $(**)$ remains valid in view of results in section 3.

Let us emphasize once again, that the relationship

$$\textit{height functions} \quad \longleftrightarrow \quad \textit{tile invariants}$$

seem to go smoothly only on a plane. In principle, of course, neither Λ nor V have to be planar. There are several interesting example of the height functions when V is a line and dimension of Λ varies. We will come back to such examples in the next section. Let us note also that we don't seem to have any nontrivial example of two-dimensional height functions when Λ is not planar, and nothing at all when V is three and more - dimensional.

5. Local moves

5.1 One-dimensional height functions.

Let $\mathbf{T}$ be a finite set of tiles, $\mathcal{B}$ be any set of finite regions. We say that $\mathbf{T}$ satisfies *local move property with respect to* $\mathcal{B}$ if there exists a finite set of regions $\Gamma_1, \dots, \Gamma_\ell \in \mathcal{B}$, and two collections of tilings $A_i, A'_i \vdash \Gamma_i$, for all $1 \le i \le \ell$ (cf. section 3), such that

$(\diamond)$ *For every* $\Gamma \in \mathcal{B}$ *and two tilings* $A, A' \vdash \Gamma$, *there exists a sequence of tilings* $A = B_0 \to B_1 \to \dots \to B_t = A'$, *where the arrow* $X \to Y$ *is between two tilings which differ in a region* $\Gamma' \sim \Gamma_i$, *with the tilings* X, Y *restricted to* $\Gamma' \subset \Gamma$, *being the tilings* A_i *and* A'_i.

Theorem 5.1 *If* $\mathbf{T}$ *satisfies local move property with respect to* $\mathcal{B}$, *then the GI-rank Problem is in* P.

The main problem with the local move property is scarcity of the sets of tiles which have it and difficulty of proving it in this case. Most known approaches are more or less ad hoc, with a small exception of the height function approach. Again, there seem to be no consensus of how this should work in general. We describe here a version of it, following [T,Ch,ST].

Let $\Lambda \subset \mathbb{R}^d$ be a d-dimensional structure (set of lattice cubes, simplices, etc.) For every $\Gamma \subset \Lambda$ denote by $\widehat{\Gamma}$ the set of points $x \in \mathbb{R}^d$ inside Γ. Suppose $\varphi : \Lambda \to \mathbb{R}$ is a one dimensional height function, such that $\varphi : \tau \to \mathbb{R}$ can be defined at all points $x \in \widehat{\tau}$ (by using piecewise linearity, or otherwise). This defines a function $\varphi_A : \widehat{\Gamma} \to \mathbb{R}$ for every tiling $A \vdash \Gamma$. We say that $\varphi(A) \le \varphi(A')$, where $A, A' \vdash \Gamma$, if for all points $x \in \Gamma$ we have $\varphi_A(x) \le \varphi_{A'}(x)$. Finally, denote by '$\prec$' a partial linear order on tilings $A, A' \vdash \Gamma$:

$$A \prec A' \ \text{ if and only if } \ \varphi(A) \le \varphi(A').$$

[3] The tiles $\varphi(\tau_i) \subset V$ may also not be uniquely defined. The extended coloring argument f defined above must be constant on all such tiles though.

Note that a priori there could be incomparable tilings.

Now, suppose the "suspected" set of local moves

$$(\bigcirc)\quad \{(A_i \to A_i'), A_i, A_i' \vdash \Gamma_i, 1 \le i \le \ell\}$$

satisfied the following properties:

$(\bullet)$ *Either* $A_i \prec A_i'$ *or* $A_i \succ A_i'$ *for all* $1 \le i \le \ell$.

$(\bullet\bullet)$ *If* $x \in \widehat{\Gamma} - \partial\Gamma$, *is a local maximum of* φ_A, $A \vdash \Gamma$, *then there exists a local move* $A \to A'$ *such that* $A' \prec A$.

$(\bullet\bullet\bullet)$ *For all* $x \in \partial\Gamma$ *there exists a unique tile* τ_x, $\widehat{\tau} \ni x$, *such that if* x *is a local maximum of* φ_A, $A \vdash \Gamma$, *then* $A \ni \tau$.

Theorem 5.2 *Let* $\mathcal{B} = \mathcal{B}_{\text{sc}}$ *and* $d = 2$. *If* $(\bigcirc)$ *and a one-dimensional height function* φ *satisfies* $(\bullet) - (\bullet\bullet\bullet)$ *for all* $\Gamma \in \mathcal{B}$, *then* $\mathbf{T}$ *satisfies the local move property with respect to* $\mathcal{B}$, *with* $(\bigcirc)$ *as a set of local moves. Further, the maximum number* M *of local moves to be made satisfies* $M \le c\,|\Gamma|^2$, *where* $c = c(\mathbf{T})$ *does not depend on* Γ. *Finally, the Tileability Problem is in* P *in this case.*

To avoid problems related to generalizations of Lemma 4.2, the above result covers only the case $d = 2$. For $d \ge 3$ we need an additional geometric condition to compensate for absence of the Lemma. Formally, consider the following property:

$(\clubsuit)$ For every local maxima $x \in \partial\Gamma$, $\Gamma \in \mathcal{B}$ we always have $\Gamma - \tau_x = \Gamma' \sqcup \Gamma'' \sqcup \ldots$, where $\Gamma', \Gamma'', \ldots \in \mathcal{B}$.

It is easy to see that $\mathcal{B}_{\text{sc}}$ satisfies $(\clubsuit)$ for $d = 2$, so the following result is a generalization of Theorem 5.2.

Theorem 5.2′ *If in condition of Theorem 5.2 the property* $(\clubsuit)$ *is also satisfied, then conclusion of Theorem 5.2 holds for all* $d \ge 2$.

Note that the conclusion of Theorem 5.2 implies, by Theorem 5.1, that the GI Problem is also in P in this case. As we shall see, the examples include domino tilings, zonotopal tilings, etc. It would be interesting to find analogs of $(\bullet)$ for two-dimensional height function. This could positively resolve the connectivity conjecture for ribbon tilings.

Conjecture 5.3 *If* $\mathbf{T}$ *satisfies the local move property with respect to* $\mathcal{B}_{\text{sc}}$, *then Tileability Problem for regions* $\Gamma \in \mathcal{B}_{\text{sc}}$ *is in* P.

While we have only few known examples of the local moves property, the conjecture seem to hold. Theorem 5.2 seem to support the conjecture. Note that if $\Gamma \in \mathcal{B}$ is untileable, then $(\diamond)$ holds by default. Heuristicly, the conjecture suggests that for any set of local moves one should be able to define a "generalized one-dimensional height function", and apply the analog of the last part of Theorem 5.2.

5.2 Tiling Polytope.

Let us conclude this section with a polytopal interpretation of the local moves. Define *rational tilings* (cf. [SU]) to be decompositions $\chi(\Gamma) = \kappa\,\chi(\tau) + \kappa'\,\chi(\tau') + \ldots$, where $\tau, \tau', \dots \in \mathbf{T}$, $\kappa, \kappa', \dots \in \mathbb{Q}_+$.

Theorem 5.4 *Rational Tileability Problem is in* P.

Proof. Let '$\prec$' be a lexicographic order on Λ. For any $\tau \in \mathbf{T}$, denote by τ_x the unique tile $\sim \tau$, such that $x \prec y$ for all $y \in \tau_x$. In other words, let τ_x be the tile obtained by translation of τ such that x is the smallest element in τ_x.

Let $k = |\mathbf{T}|$. For any region $\Gamma \in \mathcal{B}$, consider a polytope $\mathbf{P}_\Gamma \subset \mathbb{R}^{k|\Gamma|} = \mathbb{R}\langle a_{x,\tau},\ x \in \Gamma, \tau \in \mathbf{T}\rangle$, defined by the following linear equations and inequalities:

$$\begin{cases} a_{x,\tau} \geq 0, \ \ \forall\, x \in \Gamma, \tau \in \mathbf{T}, \\ \displaystyle\sum_{x,\tau:\ \tau_x \ni y} a_{x,\tau} = 1, \ \forall\, y \in \Gamma. \end{cases}$$

Now, every rational point (a) in the polytope $\mathbf{P}_\Gamma$ corresponds to a rational tiling with $\kappa_{\tau_x} = a_{x,\tau}$. Since the system is rational, the rational tileability is equivalent to $\mathbf{P}_\gamma$ being empty or not. The latter can be determined in polynomial time (see e.g. [Sc]). □

Proposition 5.5 *Let* $\mathbf{P}_\Gamma$ *be the polytope defined in the proof of Theorem 5.4. Then the integer points in* $\mathbf{P}_\Gamma$ *correspond to the (usual) tilings of* Γ *with the set of tiles* $\mathbf{T}$. □

One can think of the points in $\mathbf{P}_\Gamma$ as of nonnegative real tilings of Γ. All the vertices are the rational tilings. Unfortunately, not all of them are integer (the usual) tilings. Denote by $\widehat{\mathbf{P}}_\Gamma \subset \mathbf{P}_\Gamma$ a convex hull of the integer points. We call $\widehat{\mathbf{P}}_\Gamma$ the *tiling polytope.* By definition, $\widehat{\mathbf{P}}_\Gamma$ is a $0-1$ polytope.

Let $A, A' \vdash \Gamma$. We say that a local move $A \to A'$ is *primitive* if for no $B \vdash \Gamma$ we can have two nonintersecting local moves $A \to B$ and $B \to A'$.

Theorem 5.6 *The primitive moves* $A \to A'$, *where* $A, A' \vdash \Gamma$, *are in one-to-one correspondence with edges in the tiling polytope* $\widehat{\mathbf{P}}_\Gamma$.

We should mention here that for large Γ the set of edges of the tiling polytope is much larger than the set of local moves described in the beginning. Indeed, while the local moves can be (and usually are) primitive moves, the minimal set of local moves is a very small subset of primitive moves which can be compositions of a number of (intersecting) local moves.

It is tempting to study the simplex method or other optimization problems on tiling polytopes. The difficulty is that the minimum number of linear relations and inequalities which define $\widehat{\mathbf{P}}_\Gamma$ is probably exponential in $|\Gamma|$ (it's superpolynomial unless P=NP).

5.3 Zonotopal tilings.

It was noted on many occasions that one can think of tilings by "lozenges" (analogues of dominoes in the triangular lattice) as of projection of the cubic surface, at least for certain nice simply connected regions. In fact, Thurston's height function coincides with the height of the surface in these cases (see [T,ST]). Let us briefly mention here that one can consider zonotopal tilings which extend this observation.

Let M be a finite set of vectors in $V = \mathbb{R}^d$ and suppose $\langle M \rangle = V$. Consider a polytope P_M defined as a Minkowski sum of elements in M (considered as intervals). Such polytopes are called *zonotopes*. Call *basis blocks* zonotopes P_B such that $B \subset M$, $\langle B \rangle = B = d$. Polyhedral subdivision of P_m into basis blocks are called *zonotopal tilings*. They have a number of interesting properties, in particular the basis blocks in every zonotopal tiling are in one to one correspondence with bases of a matroid M [BLSWZ,St,Z]. In fact, much of the information about P_M and zonotopal tilings can be obtained from from the (oriented) matroid structure of M (see references above).

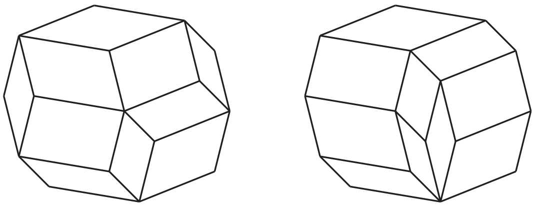


FIGURE 5.1. Two zonotopal tiling of a centrally symmetric 10-gon.

Among the most interesting properties of zonotopal tilings is (non)existence of a one-dimensional height functions. The latter correspond to the so-called 1-extensions of M (into $\mathbb{R}^{d+1}$). One can show that all zonotopal tilings that arise from every such extension are connected by "local moves" (in zonotopes generated by $d+1$ vectors). While 1-extensions of M may generate all tilings, all 1-extensions can make a graph of zonotopal tilings connected (there is a related notion of a coherent subdivision [GKZ,Z]). Still, there exist zonotopal tilings disconnected from the others. We refer to the above mentioned [BLSWZ,GKZ,St,Z] and the references therein.

6. Ribbon tiles

6.1 Basic definitions.

Let $\Lambda = \mathrm{Z}^2$ be the square grid. Let $x = (i, j) \in \Lambda$ be the square in Z^2 with i increasing downward and j increasing to the right. As before, let '$\sim$' be defined by translations.

Fix an integer $n \geq 2$. A region $\tau \in \mathcal{B}_{\mathrm{sc}}$ is called a *ribbon tile* if every diagonal $i - j = \text{const}$ contains at most one square of τ. Denote by $\mathbf{T}_n$ the set of ribbon tiles with n squares. It is easy to see that $|\mathbf{T}_n| = 2^{n-1}$, with tiles τ encoded by $\epsilon = (\epsilon_1, \ldots, \epsilon_{n-1})$, $\epsilon_i \in \{0, 1\}$ as follows. Start in the lower left corner of τ and move northeast; each upward move encode with 1, each right move with 0. Denote by τ_ϵ the tile as above, and by $\alpha_\epsilon(A)$ the number of times tile τ_ϵ occurs in a tiling A.

Define 2-moves to be the local moves which involve exactly two ribbon tiles. For description of all such moves see [P]. As observed by Adin [Ad], the total number of such moves is $\binom{|\mathbf{T}_n|}{2}$. This formula is somewhat misleading since not all pairs of ribbon tiles can form a 2-move, while some pairs can form it in several ways.

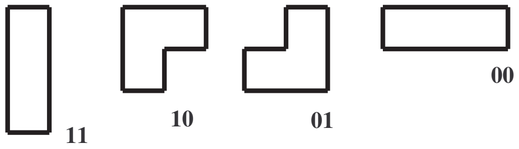


FIGURE 6.1. Ribbon trominoes.

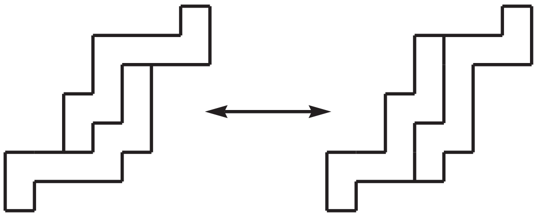


FIGURE 6.2. Example of 2-move for ribbon tiles.

The main object of this section is the successful computation of $\mathbb{G}(\mathbf{T}_n)$, and the local move property with respect to 2-moves. Note that there is an obvious *area invariant* which states that the total number of tiles τ is $|\Gamma|/n$.

6.2 Dominoes.

This is a classical example studied for decades (see e.g. [G,Ka,LP,TF]). Thurston [T]. defined an important one-dimensional height function φ which became a model for our generalization in section 5. Color the squares with two colors (black and white) in a checkerboard fashion. Orient all edges upward and to the right. The map φ is defined on edges in Z^2, and is $+1$ (-1) if the edge is moving counterclockwise (clockwise) around a black square.

One can show that the above height function with the set of 2-moves satisfies $(\bullet) - (\bullet\bullet\bullet)$. From here we obtain the local move property for 2-moves with respect to $\mathcal{B}_{\mathrm{sc}}$ as an immediate conclusion of Theorem 5.2. An elementary example shows that this does not hold for non simply connected regions. We should mention here that the result can be generalized to any planar regular graph with a bipartite dual graph [Ch]. Also, a careful look at the tileability algorithm reveals that it has cost $O(|\Gamma|)$, faster than other (general) matching algorithms [LP,Sc]. This result can be extended to non simply connected regions as well [F].

As mentioned in the introduction, the group of invariants $\mathbb{G}(\mathbf{T}_2) \simeq \mathbb{E}(\mathbf{T}_2) \simeq \mathbb{Z} \times \mathbb{Z}_2$ in this case.

6.3 Ribbon Trominoes.

The set of ribbon trominoes is the celebrated example, studied Conway and

Lagarias [CL][4]. They defined a two-dimensional height function φ which maps edges of the square lattice into a Cayley graph of a specially chosen group embedded in $\mathbb{R}^2$. The latter consists of hexagons and triangles. The sum of the winding numbers around centers of hexagons gives a nonabelian tile invariant:

$$\alpha_{01} - \alpha_{10} = \operatorname{const}(\Gamma).$$

One can conclude from here that the group of invariants $\mathbb{G}(\mathbf{T}_3) \simeq \mathbb{Z}^2$. On the other hand, direct computation shows that $\mathbb{E}(\mathbf{T}_3) \simeq \mathbb{Z} \times \mathbb{Z}_3$ [CL,P], so the infinite tile invariant above cannot be proved by means of coloring arguments.

The local move property for 2-moves with respect to $\mathcal{B}_{\text{sc}}$ remains open (see below). A special case was considered in [We] for the starecase shaped regions introduced in [CL] (see also [P]).

Before we conclude, let us mention here that the approach was later modified by Muchnik and the author [MuP] to prove that $\mathbb{G}(\mathbf{T}_4) \simeq \mathbb{Z}^2 \times \mathbb{Z}_2$. At the same time, $\mathbb{E}(\mathbf{T}_4) \simeq \mathbb{Z} \times \mathbb{Z}_4$ [P].

6.4 The general case.

It was recently shown in [MoP] that for all $n \geq 2$:

$$\mathbb{G}(\mathbf{T}_n, \mathcal{B}_{\text{sc}}) \simeq \begin{cases} \mathbb{Z}^m, & \text{if } n = 2m+1, \\ \mathbb{Z}^{m-1} \times \mathbb{Z}_2, & \text{if } n = 2m. \end{cases}$$

This proved the conjecture of the author [P], previously known only for $n \leq 4$. The main result of [P] is a similar result for a smaller set of regions $\mathbb{G}(\mathbf{T}_n, \mathcal{B}_{\text{rc}})$, where $\mathcal{B}_{\text{rc}})$ is the set of row convex regions. The author in [P] also found an explicit basis for the group:

$$\sum_{\epsilon:\ \epsilon_i=0,\, \epsilon_{n-i}=1} \alpha_\epsilon \ - \sum_{\epsilon:\ \epsilon_i=1,\, \epsilon_{n-i}=0} \alpha_\epsilon \ = \operatorname{const}(\Gamma), \ \ 1 \leq i < n/2,$$

and

$$\sum_{\epsilon:\ \epsilon_{n/2}=0} \alpha_\epsilon \ = \operatorname{const}(\Gamma) \mod 2, \ \ n = 2m.$$

On the other hand, it was shown in [P] that $\mathbb{E}(\mathbf{T}_n) \simeq \mathbb{Z} \times \mathbb{Z}_n$, and all tile invariants in the basis do not follow from the extended coloring arguments.

The technique used in [MoP] is notable since it used a new construction of the two-dimensional height function φ, which mapped the edges of the square lattice into $\{\omega^k, 0 \leq k \leq n-1\} \subset \mathbb{C}$, where $\omega = \exp(2\pi i/n)$. Then the authors take a signed area in $\mathbb{C}$ as a the generalized coloring argument. Remarkably, this single real-valued invariant contains all tile invariants presented above.

Denote by $\mathcal{B}_{\text{y}}$ and $\mathcal{B}_{\text{sy}}$ the set of regions with Young diagram and skew Young diagram shape (see e.g. [M,JK]). It was shown in [P] that $\mathbf{T}_n$ has local move property (for 2-moves) with respect to $\mathcal{B}_{\text{y}}$. The result, already more general than

[4] They actually considered one additional disconnected tile which we ignore. This set of tiles appeared after translation of the trominoes in hexagonal lattice into the square lattice [CL].

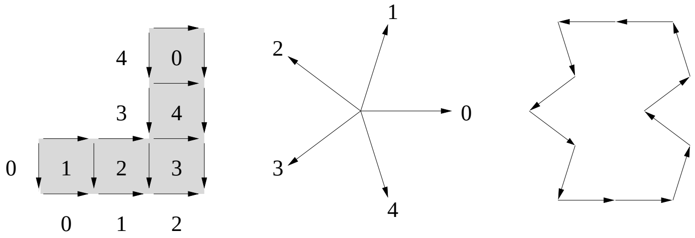


FIGURE 6.4. Ribbon tile $\tau = \tau_{0011}$, vectors ω^k, height function $\varphi(\tau)$.

[We], was later extended by the author to include $\mathcal{B}_{\text{sy}}$ (unpublished). Following [P], we conjecture the local move property with respect to all simply connected regions. The computation of $\mathbb{G}(\mathbf{T}_n, \mathcal{B}_{\text{sc}})$ and the height function arguments [MoP] seem to support the conjecture.

7. Small sets of tiles

7.1 T-tetrominoes.

It was shown in [Wa] that four rotations of T-tetromino can tile a $m \times n$ rectangle if and only if 4 divides both m and n. It is easy to see that the result cannot be proved by the coloring arguments. Nevertheless, no height function argument is known.

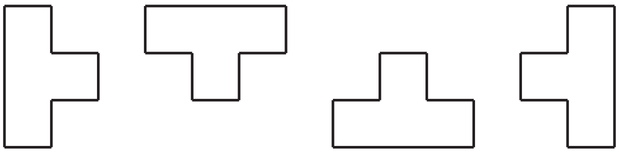


FIGURE 7.1. Four T-tetrominoes.

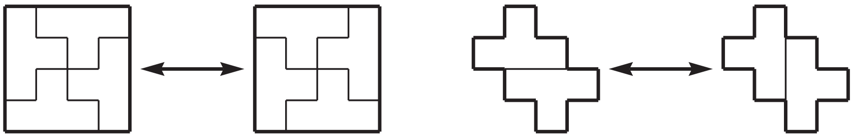


FIGURE 7.2. Local moves: 2-move and 4-move.

The set of tiles is of interest since it also seem to have a local move property. Observe that besides the 2-moves there is also a 4-move involving a reflection in a 4×4 square. We conjecture that these local move suffice. It seems that the combinatorial technique in [Wa] can be extended to prove the local move property with respect to rectangular regions.

7.2 Bars and Rectangular shapes.

Let $\mathbf{T}$ be a set of two "bars", i.e. of $m \times 1$ and $1 \times n$ rectangles. Claire and Rick Kenyon found a remarkable application of the height functions in this case [KK]. They introduced a tree-valued height function, and proved properties $(\bullet) - (\bullet\bullet\bullet)$ in this case. From here they deduced the local move connectivity (the only local move required is $A_1 \to A_2$, where $A_1, A_2 \vdash m \times n$ rectangle), obtain the general bound on the distance (it's $O(|\Gamma|^{3/2})$ in that case) and present a linear algorithm for testing tileability. The authors show that their analysis can be modified to rectangular regions $m \times n$ and $n \times m$. In particular, the authors present a quadratic algorithm for tileability and prove the local move property for 2×3 and 3×2 rectangles.

While the authors do not compute the group of invariants, it can be easily determined from either local move property or coloring arguments. Let us note that the polynomial algorithms for tileability exist only for simply connected regions, as in general case the problem is NP-complete [Ro] (see also [BJLS]).

7.3 L-trominoes.

Let $\mathbf{T}$ be the set of four rotations of L-trominoes. We showed in [P] that $\mathbb{G}(\mathbf{T}, \mathcal{B}) = \mathbb{E}(\mathbf{T}) = \mathbb{Z} \times \mathbb{Z}_3^2$. The proof involves some explicit coloring arguments.

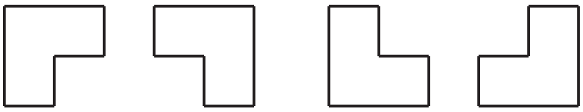


FIGURE 7.3. Four L-trominoes.

The set $\mathbf{T}$ has no local move property, as shown in [P]. There, we constructed large regions with exactly two tilings. Also, for general regions the tileability is NP-complete [MR]. It would be interesting to see if the same is true for simply connected regions. Let us mention here an old result that a $n \times n$ square with one square deleted can be tiled with $\mathbf{T}$ unless n is divisible by three [CJ].

7.4 Skew and square tetromino.

This example wa introduced by Propp, who found a very nice application of the height function approach [Pr]. The group of invariants $\mathbb{G}$ can be computed completely in this case, by using the coloring arguments and a nonabelian tile invariant presented in [Pr], which implies that $\text{rk}(\mathbb{G}) = 2$. There are two interesting features in this case. First, the authors makes a distinction between "odd" and "even" 2×2 squares. In principle, this can be done in other special cases, by taking a smaller group of translations. Still, this is by far the most interesting such example, as the infinite tile invariant becomes a finite tile invariant when odd and even squares are identified.

For the second feature, Propp in [Pr] defines a tile invariant as a signed area, refraining from the "winding number" approach in [CL]. This was the approach

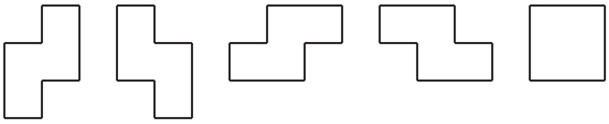


Figure 7.4. Skew and square tetromino.

continued in [MoP]. We hope the reader will enjoy this well written article and completes the computation of the full group of invariants as an exercise.

7.5 Dominoes again.

Let Γ be a simply connected region, and let k be a fixed integer. Consider all domino tilings of Γ with exactly k vertical domino. Recall that k can vary for different domino tilings, although its parity remains fixed. It was noted by Gupta [Gu] that sometimes one can make a connected graph $G(\Gamma, k)$ on these domino tilings by introducing 2×3 moves (see Figure 7.5). He showed that $G(\Gamma, k)$ is connected when Γ is a rectangle, the Aztec diamond, etc., but not in general case. We refer to [Gu] for the details.

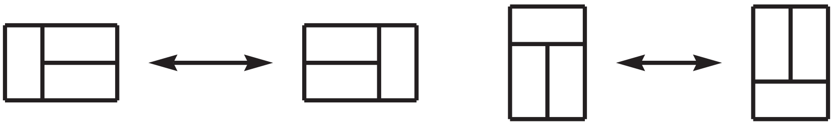


Figure 7.5. 2×3 moves.

In general, suppose $\mathbf{T}$ is a finite set of tiles and Γ is a tileable region. One can ask whether local connectivity exists for tilings $A \vdash \Gamma$ with given set of numbers $\alpha_i(A)$, defined as in the introduction. The work of Gupta suggests that certain nice sets of tiles and certain regions might satisfy this remarkable property.

7.6 More examples.

Consider the following two sets of tiles $\mathbf{T}_1$, $\mathbf{T}_2$. The first contains two rotations of T-tetromino and skew tetromino which fit into 2-row strip (see Figure 7.6). The second contains two rotations of T-pentamino, S-pentamino and skew tetromino, which fit into 3-row strip (see Figure 7.7). As before, we allow only translations of the tiles.

We are interested whether either or both sets have nonabelian tile invariants, local move property, height functions, etc. It is an exercise to establish these properties for regions which fit in 2-row and 3-row strip tiled by $\mathbf{T}_1$ and $\mathbf{T}_2$ respectively. Also, replacing skew tetrominoes with a square tetromino gives an interesting modification of $\mathbf{T}_2$. We challenge the reader to resolve these problems.

7.7 Other lattices.

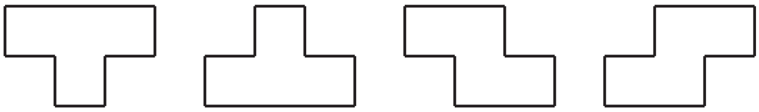


FIGURE 7.6. 2-row skew and T-tetrominoes.

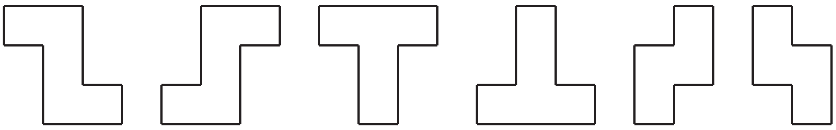


FIGURE 7.7. S and T - pentaminoes and skew tetrominoes.

It was realized rather early that tiling problems are of interest on other lattices as well [G]. The original question in [CL] comes from a hexagonal lattice, and the running example in [T] is the set of "lozenges", analogues of dominoes on a triangular lattice. A number of results for small sets of tiles on a triangular lattice was discovered recently by Rémila [Ré]. The author's approach is somewhat different from this article's main theme, and we strongly suggest it as a complimentary reading. Finally, a nice local connectivity result for squares-and-octagons was obtained by Gupta in [Gu].

8. Tilings in many dimensions

There is little known about tilings in many dimensions, although there seem to be no clear reason for that. As mentioned before, we do not know of any nonabelian tile invariant even for three-dimensional tiles. Without attempt to review the subject, let us present few examples that seem relevant.

8.1 Generalized Sperner's Lemma.

The Sperner's Lemma is the following classical result. Let Λ be a triangular lattice, Γ be a n-triangle with deleted three corner triangles. Color the vertices of the triangle with colors $\{0,1,2\}$, so that the sides are colored with 0, 1, 2 (clockwise). Then there exists a $(0,1,2)$ colored triangle. In fact, the number of $(0,1,2)$ triangles minus the number of $(0,2,1)$ triangles (reading colors clockwise) is always 1.

While the Sperner's Lemma is often associated with Brouwer's fixed point theorem (see e.g. [Sh]), its generalizations are easier to obtained in the context of the Stokes Theorem. We present here the Generalized Sperner's Lemma, which implies an abelian tile invariant for a certain set of tiles. While the generalization below is probably well known (and follows easily from Stokes Theorem) the interpretation of it in the language of tile invariants seems new and will be presented here along with a short proof of the lemma.

Let us state the Generalized Sperner's Lemma first in two, and then in all dimensions. Let Γ be any region on a triangular lattice colored with $\{0,1,2\}$. Denote

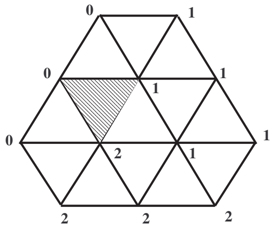


Figure 8.1. Sperner's Lemma.

by $\alpha_+(\Gamma)$ and $\alpha_-(\Gamma)$ the number of triangles with all three colors $(0, 1, 2)$, going clockwise and counterclockwise respectively. Then $\alpha_+ - \alpha_- = \mathrm{const}(\partial\Gamma)$, where $c = \mathrm{const}(\partial\Gamma)$ depends only on the coloring of the boundary. Note that we do not require Γ to be simply connected. The boundary $\partial\Gamma$ may be disconnected, but the coloring must be fixed on vertices of each connected component.

In general case, let Γ be any region in $V = \mathbb{R}^d$ with a fixed simplicial subdivision. Fix an orientation in $\mathbb{R}^d$ by taking a basis $(e_1, \ldots, e_d)$ in V. Consider any coloring of vertices of Γ with $d+1$ colors $\{0, 1, \ldots, d\}$. We say that Γ is $(d+1)$-*colored* in this case. We say that a simplex is *positive* (*negative*) if it is $(d+1)$-colored with basis $(\overrightarrow{01}, \overrightarrow{02}, \ldots, \overrightarrow{0d})$ having a positive (negative) volume, defined as a determinant of the corresponding linear transformation. Denote by $\alpha_+(\Gamma)$ and $\alpha_-(\Gamma)$ the number of positive and negative simplices in Γ, respectively. Then $\alpha_+ - \alpha_- = \mathrm{const}(\partial\Gamma)$, where the constant depends only on the coloring of $\partial\Gamma$, and not on the interior of Γ. Let us state this result as follows.

Theorem 8.1 (Generalized Sperner's Lemma) *Let Γ be a triangulated region in $\mathbb{R}^d$ with a fixed $(d+1)$-coloring of the boundary $\partial\Gamma$. Let A be a $(d+1)$-coloring of the interior vertices. Then*

$$\alpha_+(A) - \alpha_-(A) = \mathrm{const}(\partial\Gamma),$$

where const *depends only on the coloring of the boundary, and not on coloring A.*

Now, the lemma can be reduced to an infinite tile invariant for a special set of tiles. First, take the tiles to correspond to $(d+1)$-colorings by somewhat changing the boundaries around the vertices in a consistent way which depends on the color (cf. proof of Theorem 3.1). For example, a small simplex can be added to, or subtracted from the sides of a large simplex so that only simplices with the same "color" can fit together (see Figure 8.2). Denote by $\mathbf{T}$ this new set of tiles, corresponding to all possible $(d+1)$-colorings of vertices of d-dimensional simplices. In Figure 8.2 we exhibit one such two-dimensional tile corresponding to $(1, 2, 3)$-coloring.

Now notice that the "coloring" of the boundary uniquely defines the shape of the boundary. Thus the "colorings" of the interior vertices of Γ are in one-to-one correspondence with tilings of Γ with $\mathbf{T}$. Consider the tiles which correspond to $(d+1)$-colorings with distinct colors, with positive and negative orientation. Theorem 8.1 implies that the difference between the number of certain "positive"

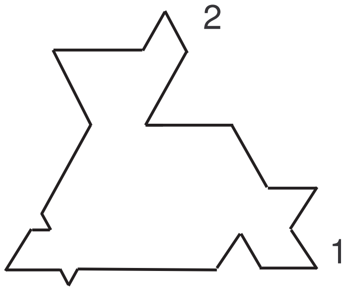


FIGURE 8.2. Modification of a 3-colored triangle.

and "negative" tiles is an fixed integer which depends on the boundary $\partial\Gamma$. We suggest the reader think through this simple, almost classical construction.

Let us note that from the proof (see section 10) it follows through verbatim that the infinite invariant defined in the lemma holds for signed tilings by $\mathbf{T}$ as well. Thus the tile invariant is abelian, and by Theorem 2.3 can be obtained by an extended coloring argument. Interestingly, this coloring argument is not obvious, and depends heavily on the way the set $\mathbf{T}$ is constructed.

Remark 8.2 The Sperner's Lemma has a number of variations, generalizations and applications. Let us first mention a similar in the spirit work [SS] where Sperner's Lemma is used to obtain relations for the volume(s) of simplices in tilings. The first d-dimensional version of the lemma can be found in [BC]. The cubical version, perhaps more acceptable for traditional tiling concepts, can be found in [Wo]. We refer to [Sh] for various application to fixed point results.

8.2 Parity check.

We will adopt the same notion of as in the previous subsection. Consider any triangular lattice $\Lambda \subset \mathbb{R}^d$, such that the dual graph is bipartite. In other words, we assume that the simplices are colored with black and white. An example is a regular partition of the cubic lattice with each cube partitioned into $d!$ simplices corresponding to permutations of basis vectors. The sign of the permutation then determines the color of the simplex.

Now consider colorings of vertices with m colors, $m \geq d$. We say that a simplex is *r-deficient* if it has exactly $(d+1-r)$ distinct colors of the vertices. Let Γ be any region in Λ with a fixed coloring of the boundary, and let A be any coloring of the interior vertices. Denote by $\rho_+(A)$ $(\rho_-(A))$ the number of black (white) 1-deficient simplices. Similarly, denote by $\alpha_+(A)$ $(\alpha_+(A))$ the number of black (white) 0-deficient simplices. Finally, let $\rho = \rho_+ - \rho_-$, $\alpha = \alpha_+ - \alpha_-$.

Theorem 8.3 *We have* $2\rho(A) + (d+1)\alpha(A) = \text{const}$*, where* $\text{const} = \text{const}(\Gamma)$ *depends only on the coloring of the boundary* $\partial\Gamma$ *and not on* A*.*

The proposition can be restated as an infinite abelian invariant of a certain set of tiles. We leave the details to the reader. As a bonus, the theorem implies that for odd d the total number of 1-deficient tiles has a fixed parity even when black and white tiles are indistinguishable. Even this is a nontrivial finite abelian invariant.

Let us conclude this part by presenting a special case when two independent tile invariants appear from such construction. This result is due to Moore and Newman, and it appeared in [MN]. We follow [Mo] in our presentation.

Consider any triangular lattice $\Lambda \subset \mathbb{R}^2$ with a bipartite dual graph. Fix a black/white coloring of triangles. Let Γ be a region in Λ with a fixed coloring of the boundary with colors $\{1,2,3,4\} = I$. Denote by $\rho_+(i,j,k)$ and $\rho_-(i,j,k)$ the number of black and white triangles colored with $i,j,k \in I$. Let

$$\begin{aligned}
\alpha_\pm &= \rho_\pm(1,1,2) + \rho_\pm(1,2,2) + \rho_\pm(3,4,4) + \rho_\pm(3,3,4),\\
\beta_\pm &= \rho_\pm(1,1,3) + \rho_\pm(1,3,3) + \rho_\pm(2,4,4) + \rho_\pm(2,2,4),\\
\gamma_\pm &= \rho_\pm(1,1,4) + \rho_\pm(1,4,4) + \rho_\pm(2,3,3) + \rho_\pm(2,2,3),\\
\alpha &= \alpha_+ - \alpha_-\,, \quad \beta = \beta_+ - \beta_-\,, \quad \gamma = \gamma_+ - \gamma_-\,.
\end{aligned}$$

Theorem 8.4 ([MN]) *We have* $\alpha(A) - \beta(A) = \mathrm{const}_1$, $\beta(A) - \gamma(A) = \mathrm{const}_2$, *where* $\mathrm{const}_1, \mathrm{const}_2$ *depend only on the coloring of the boundary* $\partial\Gamma$ *and not on* A.

We challenge the reader to obtain a proper generalization of the theorem to higher dimensions [Mo].

8.3 3-dimensional dominoes.

While dominoes on a square grid satisfy the local move property with respect to simply connected regions, this is no longer true for 3-dimensional dominoes. Heuristicly, in three dimensions there is enough space to make large simply connected "local moves". Formally, for any n there exist a simply connected region Γ with exactly two domino tilings $A_1, A_2 \vdash \Gamma$, so that the move $A_1 \to A_2$ involves at least n dominoes.

Indeed, consider a cycle of size $4\,n$ with a $(n-1) \times (n-1)$ square shaped hole inside. Think of the cycle lying in a (x,y) plane. Color this square with black and white colors in the usual checkerboard fashion. Fill this hole with dominoes pointing up or down (in the direction z), depending on whether the square is black or white. Now notice that there are exactly two domino tilings of this region Γ, as the positions of the vertical dominoes are fixed by the construction, and the only freedom we have is given by two possible tilings of the cycle. The move will involve $2\,n$ dominoes then, which proves the claim.

The construction naturally extends to tilings in any $d \geq 3$ dimensions. This makes it rather unlikely that there exists a one-dimensional height function as described in section 5.1. On the other hand, the tileability by dominoes is in P for any d (see [LP]).

Let us note that there are other generalizations of the 2-dimensional dominoes. For example, in three dimensions, one can consider $2 \times 2 \times 1$ blocks. The similar construction to the one above shows that there is no local move property with respect to the simply connected regions. It would be interesting to see if the tileability is also in P in this case (cf. [MR]).

8.4 Generalized ribbon tiles.

During the search of the nonabelian tiling arguments in many dimensions, one may ask as to whether some generalization ribbon tiles have any. Consider the obvious generalization, corresponding to connected d-dimensional tiles with at most one cube in every plane $L_c : x_1 + \ldots + x_d = c$. Denote by $\mathbf{T}_n^d$ the set of such tiles in d dimensions with n cubes. Note that $|\mathbf{T}_n^d| = d^{n-1}$. The problem of finding the tile invariant group $\mathbb{G}(\mathbf{T}_n^d, \mathcal{B}_{\mathrm{sc}})$ remains open in general case. Preliminary computations (for $d = 3$, $n = 3, 4$) suggest that $\operatorname{rk} \mathbb{G}(\mathbf{T}_n^3, \mathcal{B}_{\mathrm{sc}}) = 1$, i.e that there is no infinite nonabelian invariant in this case (area is clearly an infinite abelian invariant). We conjecture that $\operatorname{rk} \mathbb{G}(\mathbf{T}_n^d, \mathcal{B}_{\mathrm{sc}}) = 1$ for all $d \geq 3$. It is conceivable however, that the rank may increase if the set of regions is more restrictive. It would be interesting to find a nontrivial example of that.

9. Final Remarks

Let us begin by saying that in our opinion, papers [T], [CL] had a profound effect on the study of tilings, by introducing new techniques and methods into the field. The notion of tile invariants and the group of invariants [P] were inspired by [CL] and f-vectors in simple polytopes [Z]. Tile invariants have yet to become widely accepted. It is our goal here is to convince the reader that computing $\mathbb{G}(\mathbf{T})$ for various sets of tiles $\mathbf{T}$ is an important problem, which might lead to a better understanding of tilings.

To summarize this paper, me propose a new approach to the study of any fixed set of tiles $\mathbf{T}$. Fist, one can compute the coloring group $\mathbb{O}(\mathbf{T})$, an extended coloring group $\overline{\mathbb{O}}(\mathbf{T})$ and the group of valuations $\mathbb{E}(\mathbf{T})$ (cf. Theorem 3.2). Then one should attempt to determine $\mathbb{G}(\mathbf{T}, \mathcal{B}_{\mathrm{sc}})$ by computing $\mathbb{G}_N = \mathbb{G}(\mathbf{T}, \mathcal{B}_{\mathrm{sc}} \cap \mathcal{B}_N)$ for N large enough. If at some point $\mathbb{G}_N = \mathbb{E}(\mathbf{T})$, this implies that there are no nonabelian invariants (cf. Proposition 2.3), so the set $\mathbf{T}$ is not so interesting.

Suppose, on the other hand, that the calculations suggest existence of some nonabelian invariants in $\mathbb{G}(\mathbf{T})$. Then, one should check whether $\mathbf{T}$ satisfies local move property. If yes, attempt to find a one-dimensional height function which proves that (cf. Theorem 5.2). Then compute $\mathbb{G}(\mathbf{T})$ from local moves. If $\mathbf{T}$ does not satisfy the local move property, one should attempt to find nontrivial height functions φ, and compute groups $\mathbb{E}_\varphi(\mathbf{T}) \neq \mathbb{E}(\mathbf{T})$. Since $\mathbb{E}_\varphi \subset \mathbb{G}(\mathbf{T}, \mathcal{B}_{\mathrm{sc}})$, one might be able to compute the whole group of invariants that way (cf. section 6.3,4).

While Theorem 3.1 seem to suggest that the above prescription works only for special sets of tiles, we consider a success a proof of *any* nonabelian tile invariant or *any* local move property. The theory is still in the early stages of development, so even partial results are of interest.

Few words about the tileability applications. After all, tileability of the starecase shaped regions by the ribbon L-trominoes was the original motivation in [CL]. In general, suppose we are given two sets of tiles $\mathbf{T} \subset \mathbf{T}'$, and a fully computed tiling group $\mathbb{G}(\mathbf{T}', \mathcal{B})$. Now let $\Gamma \in \mathcal{B}$ be a region tileable by $\mathbf{T}'$. This determines all the constants $\operatorname{const}(\Gamma)$ for all tile invariants $(*)$. Now restriction of the tile invariants for $\mathbf{T}'$ to $\mathbf{T}$ gives a number of integer linear equations which may or may not have integer solutions. In the latter case the region is untileable by $\mathbf{T}$ (see [CL,P]).

From the point of view of tileability criteria, this seem like a weak approach. Indeed, in general, we need at least as many invariants as the number of tiles $|\mathbf{T}|$, and these tile invariants are hard to find and to prove. On the other hand, the integrality of solutions helps. In [P] we found several (un)tileability results in this direction. As a bonus, an easily computable coloring group $\mathbb{O}(\mathbf{T})$ can determine whether a certain tileability argument follows from the coloring argument. Or, as it was done in [CL], one can prove untileability of a Γ and then find a signed tilings of Γ by $\mathbf{T} \cup -\mathbf{T}$. By Theorem 2.2 one cannot prove untileability of Γ by the coloring arguments then.

There is a number of open problems that remain unresolved. Beside those mentioned earlier (Conjecture 5.3, questions about various small sets of tiles, etc.), let us stress again that we have yet to find an efficient algorithm for computing $\mathbb{E}(\mathbf{T})$ on the whole plane. It would be interesting to find other approaches to computing the group of invariants, besides the height functions, or find a reasoning why there cannot be any. It would be also very exciting to prove a local move property for some natural large set of tiles.

Let us conclude by saying that the local move property and one-dimensional height functions have important consequences in Statistical Physics and in study of Markov chains. Roughly, random application of local moves gives an easy way to sample random tilings; existence of the height function representation assists one in proving the rapid mixing. We refer to [BH,MN,PW,LRS,RY] for references and details.

10. Proof of Results

Proof of Theorem 3.2 (sketch).

We need to show that given N, $\mathbf{T} = \{\tau_1, \dots, \tau_k\}$, $|\tau_i| \le R$, one can solve Bounded CG-rank and Bounded GV-rank Problems in time polynomial in N, k, and R. Without loss of generality we will assume that $N \ge R$.

Denote by S the $N \times N$ square. Consider first a coloring group $\mathbb{O}(\mathbf{T}, \mathcal{B}_N)$. It is defined as $\mathbb{Z}^S$ quotient by the relations corresponding to translations of the tiles $\tau_i \in \mathbf{T}$ which lie in S. The rank of $\mathbb{O}$ is equal to the dimension of the corresponding real vector space (with the same integer linear equations).

There are at most N^2 translations of each tile, there are k tiles. In total, we need to calculate the rank of the system of at most $N^2 k$ equations with N^2 variables. This can clearly be done in polynomial time.

For the extended coloring group $\overline{\mathbb{O}}(\mathbf{T}, \mathcal{B}_N)$, we obtain a somewhat different set of equations. Fix one translation $\tau_i' \subset S$ of each tile $\tau_i \in \mathbf{T}$. Now, each translation τ_i'' gives an equation corresponding having to sum of the function on squares in τ_i'' equal to the sum of the function on squares in τ_i'. Again, we need to calculate the rank of the system of at most $N^2 k$ equations with N^2 variables.

Now, for the rank of the group of of valuations we have

$$\operatorname{rk} \mathbb{E}(\mathbf{T}, \mathcal{B}_N) = \operatorname{rk} \overline{\mathbb{O}}(\mathbf{T}, \mathcal{B}_N) - \operatorname{rk} \mathbb{O}(\mathbf{T}, \mathcal{B}_N).$$

This completes the proof. $\square$

Proof of Theorem 4.1 (sketch).

Use induction on the number of tiles in Γ to prove $(\star)$. The base is tautological. For the step of induction, consider τ from Lemma 4.2 such that $\Gamma' = \Gamma - \tau$ is simply connected. Fix a counterclockwise orientation on $\partial\tau$, $\partial\Gamma$, and $\partial\Gamma'$. Let $x \in \partial\Gamma$ be the starting point of the path P along the boundary. The paths P', R along the boundaries of Γ', τ are mapped into loops by inductive assumption. Observe that the intersection $P' \cup R$ will appear twice, once in each direction. On the other hand, $P = (P' - P' \cap R) \sqcup (R - P' \cap R)$. Adding the values of the height function φ along P as above, we obtain that P is also mapped into a loop. This completes the step of induction. □

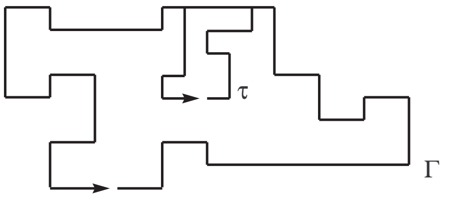


Figure 10.5. Simply connected regions Γ, τ and $\Gamma' - \tau$.

Proof of Theorem 5.1 (sketch).

We need to determine the group of invariants $\mathbb{G}(\mathbf{T}, \mathcal{B})$ in time polynomial in $k = |\mathbf{T}|$, ℓ, and $M = \max_i |\Gamma_i|$.

Indeed, tile invariants are precisely the maps $f : \mathbf{T} \to \mathbb{Z}$ which are invariant along the moves. In other words, we have

$$\mathbb{G}(\mathbf{T}) = \mathbb{Z}^{\mathbf{T}} / \mathbb{Z}\big\langle \big(\alpha_1(A_i) - \alpha_1(A_i'), \ldots, \alpha_k(A_i) - \alpha_k(A_i')\big),\ 1 \le i \le \ell \big\rangle.$$

Now, calculating all $\alpha_j(A_i)$ is polynomial in k, M. Proceed as in the proof of Theorem 3.2. Indeed, it remains now to determine rank of the system of ℓ linear equations (over $\mathbb{R}$). This can be done in polynomial time [Sc]. □

Proof of Theorem 5.2′ (sketch).

Denote by $\mathcal{A} = \mathcal{A}(\Gamma)$ the poset of all tilings $A \vdash \Gamma$, with '$\prec$' as an order relation. We claim that $\mathcal{A}$ has a minimum element A_0. Indeed, start with any tiling $A \vdash \Gamma$ and calculate φ_A. We claim that there exists a sequence of local moves from A to A_0. First, find any local maximum $x \in \widehat{\Gamma}$. If $x \notin \partial\Gamma$, then apply a local move $A \to A'$, and proceed by induction. If $x \in \partial\Gamma$, then both A, A_0 contain τ_x. Delete τ from Γ. Observe that we obtain either one region with smaller area or several smaller regions. Again proceed by induction. This proves the local connectivity property with respect to $\mathcal{B}$.

The second part follows from the following observation. Denote by A_I the largest element in $\mathcal{A}$. Then $M \le 2\Delta$, where Δ is the number of local moves from A_0 to A_I. Fix a value 0 of any point $z \in \partial\Gamma$. Let $\varphi_0 = \varphi_{A_0}$, $\varphi_I = \varphi_{A_I}$. Let h be the maximum value of φ on edges of Λ. Then for the maximum value H_I of φ_I we have $H_I \le h|\partial\Gamma| \le ch|\Gamma|$, where $0 \le c \le d2^d$. Similarly, for the smallest value H_0 of φ_0 we have $H \ge -ch|\Gamma|$.

Now, for every $A \vdash \Gamma$ define

$$\psi(A) = \int \varphi_A(x)\, d\mu,$$

where the integration is taken over $\widehat{\Gamma}$ and $d\mu$ is the usual euclidean measure on $\mathbb{R}^d$. We have

$$\psi(A_I) - \psi(A_0) \le \mu(\widehat{\Gamma})(H_I - H_0) \le c'|\Gamma|^2,$$

where c' is a constant which depends only on $\mathbf{T}$. Denote by δ the smallest change of ψ under the local move:

$$\delta = \min_{i=1}^{\ell} \left|\psi(A_i) - \psi(A_i')\right| > 0.$$

We conclude that $\Delta \le (c'/\delta)\,|\Gamma|^2 \le c''\,|\Gamma|^2$, which proves the claim.

For the last part, consider the following algorithm. Compute φ on $\partial\Gamma$. From above, the local maxima of $\varphi_0 = \varphi_{A_0}$ are on the boundary. Find a maximum value of $x \in \partial\Gamma$. This is clearly a local maximum of φ_0. Now delete τ_x from Γ and proceed accordingly. Eventually we either determine A_0 completely, or at some point we have to delete τ_x from Γ in an impossible situation. Since A_0 is unique, this implies untileability of Γ in that case. Note that the cost of the algorithm is $O(|\Gamma|^2\ell k)$. This completes the proof of the theorem. □

Proof of Theorem 5.6.

First, observe that tilings $A \vdash \Gamma$ correspond to vertices of $\mathbf{P}_\Gamma$. Indeed, suppose otherwise. By abuse of notation we can write this as $A = \beta_1 B_1 + \beta_2 B_2 + \ldots$, where $\beta_1, \beta_2, \dots \in \mathbb{R}_+$. But that means that zeroes of $(a_{x,\tau})$ on the left hand side correspond to zeroes on the right hand side, i.e. $B_1, B_2, \dots = A$. This proves the claim.

Similarly, consider two tilings $A_1, A_2 \vdash \Gamma$. Let

$$A_\lambda = \lambda\, A_1 + (1-\lambda)\, A_2 = \beta_1 B_1 + \beta_2 B_2 + \ldots,$$

where $0 < \lambda < 1$. The point A_λ lies on the interval $[A_1, A_2]$. By the observation above, only tiles that are in A_1, A_2 can appear in B_i. Therefore all tiles that lie in $A_1 \cap A_2$ must also appear in each of the B_i. On the other hand, a tile $\tau_x \in A_1$ must appear in B_i with the total weight λ. Having or not having τ_x splits the set of tilings B_i into two subsets. Since every element $y \in \Lambda$ must belong to some tile, the total set of tiles splits between tiles that contain and don't contain τ_x. Denote these sets of indices by I and J. The above implies that either every $B_i = A_1$, $i \in I$, every $B_j = A_2$, $j \in J$, or there exist B_i, B_j, $i \in I$, $j \in J$, such that $A_1 \to C$ and $C \to A_2$ are non-intersecting local moves (and the same is true for $A_1 \to D$ and $D \to A_2$). This completes the proof. □

Proof of Generalized Sperner's Lemma 8.1 (sketch).

Define an orientation of the $(d-1)$-dimensional simplices on the boundary to agree with orientation of $V = \mathbb{R}^d$. Formally, we say that a simplex on the boundary is positive (negative) if it is colored with d colors $\in \{0, 1, \dots, d\}$ and coloring the remaining vertex of a unique d-dimensional simplex in Γ with the remaining color would make this simplex positive (negative). Denote by $\beta_+(\partial\Gamma)$ and $\beta_-(\partial\Gamma)$ the number of positive and negative simplices on the boundary. A simplex (of any dimension) with repeated colors we call *neutral*.

Let us prove by induction that in conditions of the theorem we have:

$$\operatorname{const}(\partial\Gamma) = (d+1)\big(\beta_+(\partial\Gamma) - \beta_-(\partial\Gamma)\big).$$

First, let us prove the base of induction. Indeed, for a single positive (negative) d-dimensional simplex all $(d+1)$ simplices on the boundary are positive (negative). If the d-dimensional simplex is neutral, then the symmetry argument implies that const $= 0$ in this case.

For the step of induction, we can delete *any* d-dimensional simplex from Γ. Now observe that $\operatorname{const}(\partial\Gamma)$ is additive with respect to such division since the intersection of the boundaries is taken with opposite signs, and thus cancel each other (cf. proof of Theorem 4.1). We omit the easy details. □

Proof of Theorem 8.2.

Consider all 0-deficient $(d-1)$-dimensional simplices in Γ, i.e. $(d-1)$-dimensional faces with d distinct colors. Each such face is either on the boundary or is a boundary of one black and one white d-dimensional simplex. Denote by Δ the number of such faces. Denote by δ_+ (δ_-) the number of of such faces on the boundary, so that the adjacent simplex is black (white). By counting Δ separately, as a boundary of black or white squares, we obtain

$$\Delta = 2\rho_+ + (d+1)\alpha_+ + \delta_- = 2\rho_- + (d+1)\alpha_- + \delta_+.$$

Subtracting the sides in the last equality, we conclude

$$2\rho + (d+1)\alpha = \delta_+ - \delta_-.$$

This proves the result. □

References

[Ad] R. Adin, personal communication (1999).

[BLSWZ] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, G. Ziegler, *Oriented Matroids*, Cambridge U. Press, Cambridge, UK, 1999.

[Be] R. Berger, *The undecidability of the domino problem*, Mem. Amer. Math. Soc. **66** (1966).

[BJLS] F. Berman, D. Johnson, T. Leighton, P. Shor, *Generalized planar matching*, J. Algorithms **11** (1990), 153–184.

[BC] A. B. Brown, S. S. Cairns, *Strengthening of Sperner's lemma applied to homology theory*, Proc. Nat. Acad. Sci. U.S.A. **47** (1961), 113–114.

[BH] J. K. Burton, Jr., C. L. Henley, *A constrained Potts antiferromagnet model with an interface representation*, J. Phys. A **30** (1997), 8385–8413.

[Ch] T. Chaboud, *Domino tiling in planar graphs with regular and bipartite dual*, Theoret. Comput. Sci. **159** (1996), 137–142.

[CJ] I.-P. Chu, R. Johnsonbaugh, *Tiling deficient boards with trominoes*, Math. Mag. **59** (1986), 34–40.

[CL] J. H. Conway, J. C. Lagarias, *Tilings with polyominoes and combinatorial group theory*, J. Comb. Theory, Ser. A **53** (1990), 183–208.

[F] J. C. Fournier, *Tiling pictures of the plane with dominoes*, Discrete Math. **165/166** (1997), 313–320.

[GJ] M. Garey, D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-completeness*, Freeman, San Francisco, CA, 1979.

[GKZ] I. M. Gelfand, M. M. Kapranov, A. V. Zelevinsky, *Discriminants, resultants, and multidimensional determinants*, Birkhäuser, Boston, MA, 1994.

[G] S. Golomb, *Polyominoes*, Scribners, New York, 1965.

[GO] J. E. Goodman, J. O'Rourke, editors, *Handbook of Discrete and Computational Geometry*, CRC Press, Boca Raton, FL, 1997.

[Gu] D. Gupta, *Some tilings moves explored,* Ph.D. Thesis, MIT, 1998, pp. 135 pp..

[JK] G. James, A. Kerber, *The Representation Theory of the Symmetric Group*, Addison-Wesley, Reading, MA, 1981.

[Ka] P. W. Kastelyn, *The statistics of dimers on a lattice. I. The number of dimer arrangements on a quadratic lattice*, Phisica **27** (1961), 1209–1225.

[KK] C. Kenyon, R. Kenyon, *Tiling a polygon with rectangles*, Proc. 33rd Symp. Foundations of Computer Science (1992), 610–619.

[Ke] R. Kenyon, *A note on tiling with integer-sided rectangles*, J. Combin. Theory, Ser. A **74** (1996), 321–332.

[LLL] A. K. Lenstra, H. W. Lenstra, L. Lovász, *Factoring polynomials with rational coefficients*, Math. Ann. **261** (1982), 515–534.

[LP] L. Lovász, M. D. Plummer, *Matching theory*, North-Holland, Amsterdam, 1986.

[LRS] M. Luby, D. Randall, A. Sinclair, *Markov chain algorithms for planar lattice structures*, Proc. 36th Symp. Foundations of Computer Science (1995), 150–159.

[M] I. G. Macdonald, *Symmetric Functions and Hall Polynomials*, Oxford University Press, London, 1979.

[Mo] C. Moore, personal communication (2000).

[MN] C. Moore, M. E. J. Newman, *Height Representation, Critical Exponents, and Ergodicity in the Four-state Triangular Potts Antiferromagnet*, J. Stat. Phys. **99** (2000), 661–690.

[MoP] C. Moore, I. Pak, *Ribbon tile invariants from signed area*, J. Comb. Theory, Ser. A, to appear (2001).

[MR] C. Moore, J. M. Robson, *Hard tiling problems with simple tiles*, Discrete and Computational Geometry, to appear (2001).

[MuP] R. Muchnik, I. Pak, *On tilings by ribbon tetrominoes*, J. Comb. Theory, Ser. A **88** (1999), 188–193.

[P] I. Pak, *Ribbon tile invariants*, Trans. AMS **352** (2000), 5525–5561.

[Pr] J. Propp, *A pedestrian approach to a method of Conway, or, A tale of two cities*, Math. Mag. **70** (1997), 327–340.

[PW] J. Propp, D. Wilson, *Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics*, Random Structures and Algorithms **9** (1996), 223–252.

[RY] D. Randall, G. Yngve, *Random three-dimensional tilings of Aztec octahedra and tetrahedra: an extension of domino tilings*, Proc. SODA 2000, 636–645.

[Ré] E. Rémila, *Tiling groups: new applications in the triangular lattice*, Discrete Comput. Geom. **20** (1998), 189-204.

[Ri] R. M. Robinson, *Undecidability and nonperiodicity for tilings of the plane*, Invent. Math. **12** (1971), 177–209.

[Ro] J. M. Robson, *Sur le pavage de figure du plan par des barres*, in Actes des Journées Polyominos et Pavages (1991), 95–103.

[ST] N. C. Saldanha, C. Tomei, *An overview of domino and lozenge tilings*, Resenhas **2** (1995), 239–252.

[Sc] A. Schrijver, *Theory of Integer and Linear Programming*, John Wiley, New York, 1988.

[SU] E. R. Scheinerman, D. H. Ullman, *Fractional graph theory*, Wiley, New York, 1997.

[Si] M. Sipser, *Introduction to the Theory of Computation*, PWS Publishing Company, New York, 1997.

[Sh] Yu. A. Shashkin, *Fixed points*, AMS, Providence, RI, 1991.

[St] R. P. Stanley, *Enumerative Combinatorics,* Vol. 1, Wadsworth & Brooks/Cole, California, 1986.

[SS] S. K. Stein, S. Szabó, *Algebra and tiling*, Carus Mathematical Monographs, **25**, MAA, Washington, DC, 1994.

[TF] H. N. V. Temperley, M. E. Fisher, *Dimer problem in statistical mechanics – An exact result*, Philos. Mag. **6** (1961), 1061–1063.

[T] W. Thurston, *Conway's tiling group*, Amer. Math. Monthly **97** (1990), 757–773.

[Wa] D. W. Walkup, *Covering a rectangle with T-tetrominoes*, Amer. Math. Monthly **72** (1965), 986–988.

[We] D. C. West, *An elementary proof of two triangle-tiling theorems of Conway and Lagarias*, unpublished manuscript (1990), 6 pp.

[Wo] L. A. Wolsey, *Cubical Sperner lemmas as applications of generalized complementary pivoting*, J. Comb. Theory, Ser. A **23** (1977), 78–87.

[Z] G. Ziegler, *Lectures on Polytopes,* Graduate Texts in Mathematics 152, Springer, 1995.

Added in Print:

In the past year few advances has been made. First, Scott Sheffield resolved most of the open problems on ribbon tilings in "*Ribbon tilings and multidimensional height functions*", arXiv preprint math.CO/0107095. Among other things, he proved the local connectivity property, conjectured by the author in [P] (see section 6.4) and found a linear time algorithm for testing tileability.

Second, Cris Moore, Ivan Rapaport and Eric Remila defined a height function and proved a local connectivity property for the set of colored square tiles similar to that in section 8.2. Their paper "*Tiling groups for Wang tiles*" will appear in Proc. SODA'2002.

Finally, the author resolved affirmatively the question whether computing (unbounded) group $\mathbb{E}(\mathbf{T})$ is decidable ("*Computational complexity of tile invariants*", preprint, 2001.)

TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
S 0002-9947(00)02666-0
Article electronically published on August 8, 2000

RIBBON TILE INVARIANTS

IGOR PAK

ABSTRACT. Let $\mathbf{T}$ be a finite set of tiles, and $\mathcal{B}$ a set of regions Γ tileable by $\mathbf{T}$. We introduce a *tile counting group* $\mathbb{G}(\mathbf{T},\mathcal{B})$ as a group of all linear relations for the number of times each tile $\tau \in \mathbf{T}$ can occur in a tiling of a region $\Gamma \in \mathcal{B}$. We compute the tile counting group for a large set of *ribbon tiles*, also known as rim hooks, in a context of representation theory of the symmetric group.

The tile counting group is presented by its set of generators, which consists of certain new *tile invariants*. In a special case these invariants generalize the Conway-Lagarias invariant for tromino tilings and a height invariant which is related to computation of characters of the symmetric group.

The heart of the proof is the known bijection between rim hook tableaux and certain standard skew Young tableaux. We also discuss signed tilings by the ribbon tiles and apply our results to the tileability problem.

0. TRIVIA

Suppose we are given a set of the *tiles* on a plane. We are allowed to use translations of the tiles to arrange them in a geometric shape (each tile may occur several times). This arrangement is called *tiling* of that shape. One can ask whether a given region can be tiled by a given set of tiles, and if it can, how many different tilings there are.

For example, with a set of tiles shown in Figure 0.1 one can make four different tilings of the 4-by-6 rectangle. Two of them are shown in Figure 0.2. Now one can try to find a criterion for when you can tile a rectangle. Observe that each of these tiles alone can tile the whole plane.

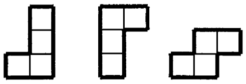


FIGURE 0.1.

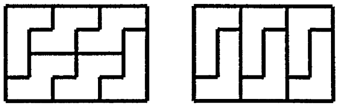


FIGURE 0.2.

Our personal favorite example is given by the set of tiles shown in Figure 0.3. One can show that there exists only one tiling of the fourth quadrant (see Fig. 0.4). The proof is left to the reader.

It turns out that there are certain nice sets of tiles for which it is not clear whether a given region can be tiled. Here is an example. Consider the 8 tiles shown in Figure 0.5. One can show that the 25-by-25 square cannot be tiled by these tiles.

Received by the editors December 12, 1997.

2000 *Mathematics Subject Classification.* Primary 05E10, 52C20; Secondary 05B45, 20C30.

Key words and phrases. Polyomino tilings, tile invariants, Conway group, rim (ribbon) hooks, Young diagrams, Young tableaux, rim hook bijection, symmetric group.

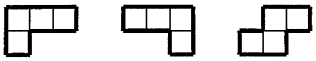


Figure 0.3.

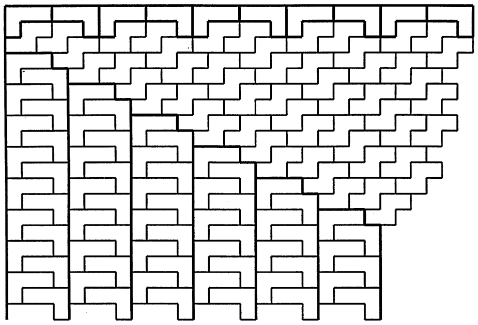


Figure 0.4.

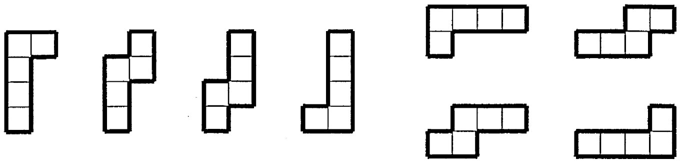


Figure 0.5.

Of course, this could be proved by an exhaustive search. In general, the following result holds.

Theorem 0.1. *If an a-by-b rectangle can be tiled by the tiles shown in Figure 0.5, then* $10 \mid a \cdot b$.

Another example. Consider 6 tiles shown in Figure 0.6. We again have

Theorem 0.2. *If an a-by-b rectangle can be tiled by the tiles shown in Figure 0.6, then* $10 \mid a \cdot b$.

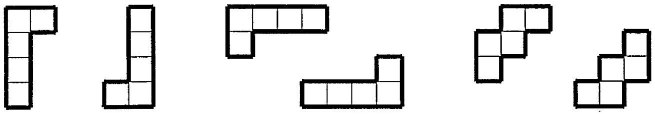


Figure 0.6.

Of course, an area argument shows that $5 \mid a \cdot b$.

Consider now a different region. Let Δ_N be a triangular shape as in Figure 0.7. One can check that Δ_{24} can be tiled by the tiles shown in Figure 0.6 while Δ_{25} cannot. Generally,

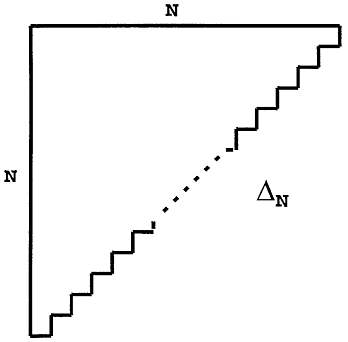


FIGURE 0.7.

Theorem 0.3. *If Δ_N can be tiled by the tiles shown in Figure 0.6, then* $N \equiv 0, 4, 15, 19 \pmod{20}$.

It turns out that all three theorems can be proved by use of the same kind of argument. Heuristically, the reason for untileability arises from the following question, completely different in nature:

- *Given a set of tiles and a tileable region, are there any linear relations for the number of times each tile occurs in a tiling?*

The rest of the paper explains the relevance of this question. All three theorems are proved in section 7.

1. INTRODUCTION

Let $\mathbb{Z}^2$ be a square lattice, and $\mathcal{R}$ the set of all compact simply connected regions in $\mathbb{Z}^2$. We think of these regions as disjoint unions of 1×1 squares. Sometimes they are called *polyominoes.* Fix a finite set of *tiles* $\mathbf{T} = \{\tau_1, \dots, \tau_N\}$, $\tau_i \in \mathcal{R}$, $i = 1, \dots, N$. Let tiles be invariant under translations. We say that a region $\Gamma \in \mathcal{R}$ is *tileable by* $\mathbf{T}$ if it can be presented as a disjoint union of the regions

$$\Gamma = \coprod_{1 \le j \le l} \tau_j',$$

where each region τ_j', $1 \le j \le l$ is a translation of some τ_{i_j}. Such a disjoint union is called a *tiling* s of Γ. Denote $\mathcal{S} = \mathcal{S}(\Gamma, \mathbf{T})$ a set of all tilings of Γ by the set of tiles $\mathbf{T}$.

Fix a set of tiles $\mathbf{T}$ and a region $\Gamma \in \mathcal{R}$. There are two basic questions one can ask:

- *Is* Γ *tileable by* $\mathbf{T}$*?*
- *If* Γ *is tileable by* $\mathbf{T}$*, what do the tilings look like?*

The first question is very classical and well understood (see e.g. [G]). It is usually not hard to find a tiling if Γ is tileable by $\mathbf{T}$, while proving the opposite can be extremely difficult. Except for ad hoc examples, there are basically two techniques for proving that a region cannot be tiled: coloring arguments and Conway group analysis (see [CL], [T]). Note also that the case when $\mathbf{T}$ contains a 1-by-1 square is trivial: every region is tileable.

While the first questions admits only a "yes" or "no" answer, the second question could be posed in many ways, each of them giving us some information about the

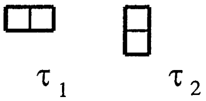


FIGURE 1.1.

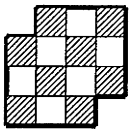


FIGURE 1.2.

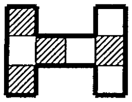


FIGURE 1.3.

structure of the tiling set $\mathcal{S}(\Gamma, \mathbf{T})$. One can find the following two questions in the literature (see e.g. [G], [CEP]):

- *How many tilings of* Γ *are there?*
- *What do random tilings* $s \in \mathcal{S}(\Gamma, \mathbf{T})$ *look like?*

It turns out that the answers to these questions depend heavily on the geometry of Γ, and can be very complicated even in very simple cases. In particular, finding the number of tilings $|\mathcal{S}(\Gamma, \mathbf{T})|$ is a more general problem than just finding whether a certain region has a tiling. In some cases this problem is known to be NP-complete, and probably cannot be solved by means other than exhaustive enumeration (see [GJ, p. 257]).

To avoid these difficulties we propose another approach to the problem. We fix only $\mathbf{T}$ and ask about properties of tilings of all regions at once. We would like to ask the following two questions:

- *Are there any relations for the number of times each tile occurs in a tiling of a given region?*
- *Is there a finite set of* local replacement rules *(we also call them* local moves *or just* moves*) such that for every region* $\Gamma \in \mathcal{R}$*, any tiling of* Γ *can be changed into any other tiling by a sequence of moves?*

Before we give formal definitions, let us illustrate what happens in the case of *dominoes*. Although small, this example will illustrate the variety of approaches as well as the complexity of a problem.

Let $\mathbf{T}_2$ be a set of two tiles: horizontal domino τ_1 and vertical domino τ_2 (see Fig. 1.1). It is easy to come up with a necessary condition for tileability (see e.g. [G]). Color the region in a checkerboard fashion. Since each domino must contain one black and one white square, the total number of black squares must be equal to the total number of white squares. For example, the region shown in Figure 1.2 cannot be tiled by dominoes since it has 8 black squares and only 6 white squares. Unfortunately, there exist untileable regions with an equal number of black and white squares (see e.g. Fig. 1.3). This means that we need a stronger condition for tileability.

Now suppose we have a region Γ that is known to be tileable. We want to compute the number of tilings Γ has. This turns out to be an interesting and nontrivial question. In the case of a 2-by-m rectangle the number of tilings is a Fibonacci number $F(m) = F(m-1) + F(m-2)$ (see Fig. 1.4). In the case of a

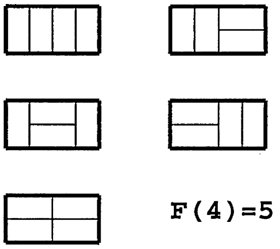


FIGURE 1.4.

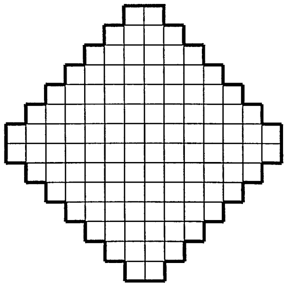


FIGURE 1.5. Aztec diamond A_7

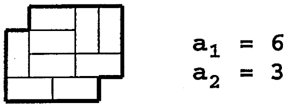


FIGURE 1.6.

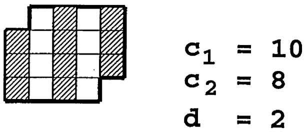


FIGURE 1.7.

general rectangle the problem was solved by Kastelyn and Temperley & Fisher (see [Ka], [TF]). In the case of an Aztec diamond (see Fig. 1.5) the domino tilings were enumerated by Elkies, Kuperberg, Larsen and Propp (see [EKLP]). Both results gave rise to many other questions about domino tiling (see e.g. [CEP]). In this work we do not further consider any numerical results of this type.

Let Γ be a region tileable by dominoes τ_1, τ_2. Consider $s \in \mathcal{S}(\Gamma, \mathbf{T}_2)$, a domino tiling of Γ. Suppose s consists of $a_1 = a_1(s)$ copies of the horizontal domino τ_1, and of $a_2 = a_2(s)$ copies of the vertical domino τ_2. Of course, $2\,(a_1 + a_2)$ is equal to the area $|\Gamma|$ of the region (see Fig. 1.6). This gives us the first relation. There is one more relation which is less obvious: $a_2 = \mathit{Const} \pmod 2$. To see this, let us color black every other column of the region Γ (see Fig. 1.7). Denote by c_1, c_2 the number of black and white regions respectively, and put $d = c_1 - c_2$. Since horizontal dominoes contain exactly one black and one white square, and vertical dominoes contain two squares of the same color, we immediately get $a_2 = d/2 \pmod 2$ (see Fig. 1.7).

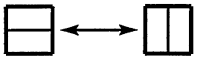


FIGURE 1.8.

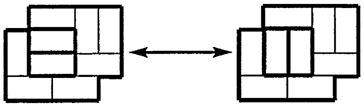


FIGURE 1.9.

Here is another way to look at the set of tilings $\mathcal{S}(\Gamma, \mathbf{T}_2)$. Let us allow the following *local replacement rules* (or simply *moves*): take two adjacent horizontal or vertical dominoes and flip them (see Fig 1.8). Of course, this move gives us a new tiling of Γ (see Fig 1.9). It is known that by a sequence of such moves we can get from any tiling $s \in \mathcal{S}(\Gamma, \mathbf{T}_2)$ of a simply connected region Γ to any other tiling $s' \in \mathcal{S}(\Gamma, \mathbf{T}_2)$ (see e.g. [T]). From this we immediately get $a_1(s) + a_2(s) =$

$a_1(s') + a_2(s')$ and $a_2(s) = a_2(s') \pmod 2$, since these identities are trivial for any single move. This also implies that for a general region there is no other relation for the numbers a_1, a_2 that does not follow from these two. We will use similar logic when proving our main results.

Now we are ready to introduce the *tile counting group* and *tile invariants*. Let $\mathbf{T} = \{\tau_1, \dots, \tau_N\}$ be a set of tiles. Denote by $\mathcal{R}_\mathbf{T} \subset \mathcal{R}$ set of regions tileable by $\mathbf{T}$. Let $\mathcal{B} \subset \mathcal{R}_\mathbf{T}$ be a fixed subset of tileable regions. Consider a tileable region $\Gamma \in \mathcal{B}$. We identify each tiling $s \in \mathcal{S}(\Gamma, \mathbf{T})$ with its multiset of tiles, $s \simeq \{\tau_{i_1}, \dots, \tau_{i_l}\}$. Of course, by doing so we lose some information about the geometric structure of the tilings, since there could be many tilings of Γ with the same multiset of tiles (see e.g. Fig. 1.4). As before, by $|\Gamma|$ we denote the area of Γ.

Let $\mathbb{Z}\langle\mathbf{T}\rangle$ be a group of formal integer linear combinations of $\mathbf{T}$. With each pair of tilings $s_1, s_2 \in \mathcal{S}(T, \Gamma)$ of a region $\Gamma \in \mathcal{B}$ we associate a relation:

$$(\tau_{i_1} + \dots + \tau_{i_l} = \tau_{j_1} + \dots + \tau_{j_r}) .$$

Let I be the linear span of such relations for all regions $\Gamma \in \mathcal{B}$ and for all pairs of tilings $s_1, s_2 \in \mathcal{S}(T, \Gamma)$. Define the *tile counting group* to be the quotient group

$$\mathbb{G}(\mathbf{T}; \mathcal{B}) = \mathbb{Z}\langle\mathbf{T}\rangle / I.$$

This will be the main object of our study. Since both groups in the quotient are abelian, one can think of a tile counting group $\mathbb{G}(\mathbf{T}; \mathcal{B})$ as a subgroup of $\mathbb{Z}\langle\mathbf{T}\rangle$. Thus it is reasonable to describe $\mathbb{G}(\mathbf{T}; \mathcal{B})$ by its set of independent generators (or the *basis*) given in $\mathbb{Z}\langle\mathbf{T}\rangle$.

For example, let $\mathbf{T}_2$ be a set of dominoes (see Fig. 1.1), and $\mathcal{B}$ a set of simply connected regions. The two tilings in Figure 1.8 correspond to the relation $2 \cdot \tau_1 = 2 \cdot \tau_2$. Since every domino tiling of a simply connected region can be obtained from every other domino tiling, I in this case is generated by the above relation. Therefore

$$\mathbb{G}(\mathbf{T}_2; \mathcal{B}) = \mathbb{Z}^2 / I \simeq \mathbb{Z} \times \mathbb{Z}_2.$$

The basis can be given as $\tau_1 + \tau_2$, $\tau_1 - \tau_2 \ \in \mathbb{Z}\langle\mathbf{T}_2\rangle$. Note that the second generator has order 2 as an element in $\mathbb{G}(\mathbf{T}_2; \mathcal{B})$, while it has infinite order as an element of $\mathbb{Z}\langle\mathbf{T}_2\rangle$.

Here is another way to describe the tile counting group. Let G be an abelian group, not necessarily finite. A map $f : \mathcal{B} \to G$ is called a *tile invariant* (or just an *invariant*) if for any tileable region $\Gamma \in \mathcal{B}$ and for any tiling $s \in \mathcal{S}(\Gamma, \mathbf{T})$ of it, $s \simeq \{\tau'_{i_1}, \dots, \tau'_{i_l}\}$, where τ'_j is a translation of a tile τ_j, we have

$$f(\Gamma) = f(\tau_{i_1}) + \dots + f(\tau_{i_l}).$$

The problem is to find all the tile invariants for a fixed set of tiles $\mathbf{T}$. Clearly a tile invariant is determined by its values on $\mathbf{T}$, so the problem of finding an invariant is equivalent to finding maps $f : \mathbf{T} \to G$ which can be extended to the set of all regions $\mathcal{B}$.

Let $\sum_{\tau \in \mathbf{T}} a(\tau) \in \mathbb{Z}\langle\mathbf{T}\rangle$ be an element of a tile counting group $\mathbb{G} = \mathbb{G}(\mathbf{T}; \mathcal{B})$. Suppose m is its order in $\mathbb{G}$ (m could be infinity). Then a map $f : \mathbf{T} \to \mathbb{Z}_m$, $m < \infty$, or $f : \mathbf{T} \to \mathbb{Z}$, $m = \infty$, defined by $f(\tau) = a(\tau) \pmod m$ or $f(\tau) = a(\tau)$, is a tile invariant, where by $\mathbb{Z}_m$ we mean the additive group of integers modulo m. Conversely, every tile invariant can be lifted to an element of the tile counting group. Thus the problem of computing the tile counting group $\mathbb{G}(\mathbf{T})$ is equivalent to describing all invariants. We say that tile invariants $f_1, f_2, \dots$ form an *independent*

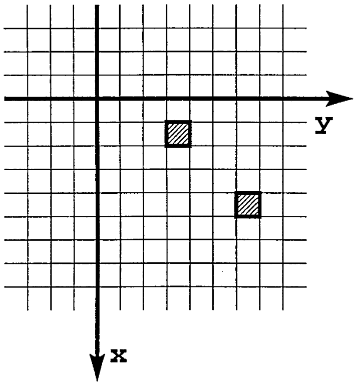


FIGURE 1.10.

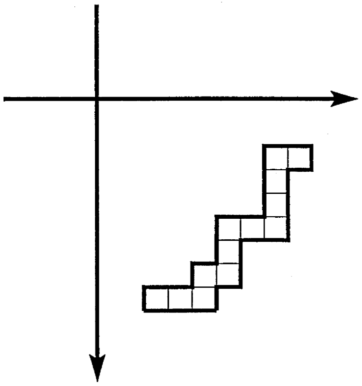


FIGURE 1.11.

basis of invariants if they correspond to an independent generating set in a tile counting group. An invariant $f : \mathbf{T} \to G$ is called *trivial* if $f(\tau) = 0$ for all $\tau \in \mathbf{T}$, where $0 \in G$ is the identity element. Otherwise the invariant is called *nontrivial.* An invariant $f : \mathbf{T} \to G$ is called *primitive* if $G \simeq \mathbb{Z}$ or $\mathbb{Z}_m$ for some m. For the rest of the paper we will be considering only primitive invariants.

Note that when $\mathcal{B} = \mathbf{T}$ every map $f : \mathcal{B} \to G$ is a tile invariant, i.e $\mathbb{G}(\mathbf{T}, \mathbf{T}) \simeq \mathbb{Z}^{|\mathbf{T}|}$. Generally, the bigger that our set of regions $\mathcal{B}$, the more equations we have on f, and the fewer tile invariants we get.

The obvious example of a nontrivial tile invariant is given by the area of tiles:

$$f_0 : \mathbf{T} \to \mathbb{Z}\,, \quad f_0(\tau_i) = |\tau_i|$$

which can be extended to all tileable regions: $f_0(\Gamma) = |\Gamma|$. This implies that the tile counting group has $\mathbb{Z}$ as a subgroup. In the case of domino tiles $\mathbf{T}_2$ we also get another invariant (see above):

$$f_* : \mathbf{T}_2 \to \mathbb{Z}_2\,, \quad f_*(\tau_1) = 0,\ f_*(\tau_2) = 1 \mod 2.$$

The main result of this paper is a description of a tile counting group for the following set of tiles.

Let the axes on a plane be as shown in Figure 1.10. We say that squares (i, j) and (i', j') lie on the same diagonal if $i - j = i' - j'$. For example, the two squares $(2, 4)$ and $(5, 7)$ lie on the same diagonal (see Fig. 1.10). A *ribbon tile* is a simply connected region with no two squares lying on the same diagonal. An example of a ribbon tile is shown in Figure 1.11. Denote by $\mathbf{T}_n$ the set of all ribbon tiles τ with n squares: $|\tau| = n$. Obviously, $\mathbf{T}_2$ is the set of dominos (see Fig 1.1). The sets $\mathbf{T}_3$, $\mathbf{T}_4$ and $\mathbf{T}_5$ are shown in Figures 1.12 – 1.14.

Note that $|\mathbf{T}_n| = 2^{n-1}$. Indeed, we can encode each ribbon tile by a sequence $(\varepsilon_1, \dots, \varepsilon_{n-1})$ of $n - 1$ zeroes and ones as follows. Call the lower left square the *starting square.* Begin with the starting square and move along the tile. Write **0** when going right, and write **1** when going up. See Figures 1.12 – 1.14 for these coding sequences for all tiles in $\mathbf{T}_3$, $\mathbf{T}_4$ and $\mathbf{T}_5$.

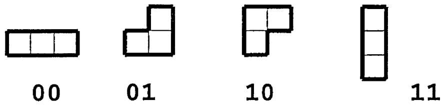


Figure 1.12.

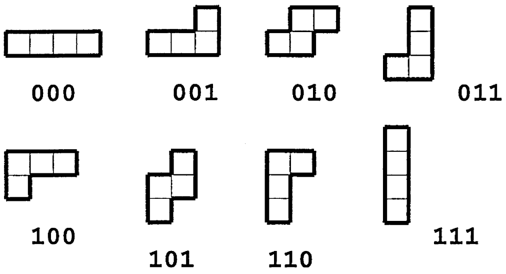


Figure 1.13.

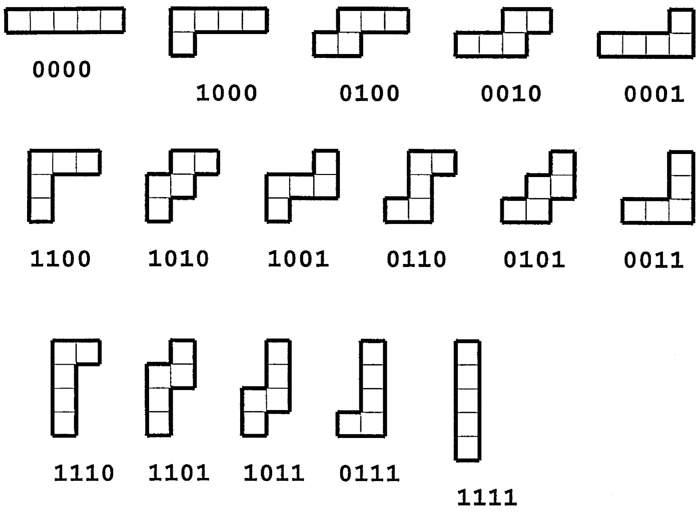


Figure 1.14.

Definition 1.1. Consider the sequence of maps $f_1, \dots , f_m : \mathbf{T}_n \to \mathbb{Z}$, $m = \lfloor \frac{n-1}{2} \rfloor$, defined as follows:

$$f_i(\varepsilon_1, \dots, \varepsilon_{n-1}) = \varepsilon_i - \varepsilon_{n-i}.$$

We call the map f_i the ***i*-convexity invariant**.

Definition 1.2. The constant map $f_0 : \mathbf{T}_n \to \mathbb{Z}$ defined as

$$f_0(\varepsilon_1, \dots, \varepsilon_{n-1}) = 1$$

is called the **area invariant**.

Note that the area invariant is designed to be 1 on a tile $\tau \in T_n$ rather than n. This is designed to simplify the statement of the main result (see below.)

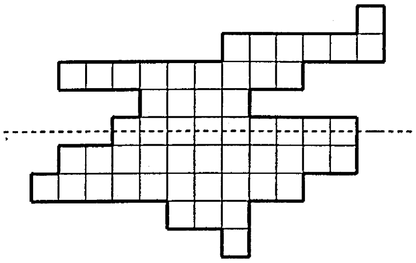


FIGURE 1.15.

Definition 1.3. If n is even, the map $f_* : \mathbf{T}_n \to \mathbb{Z}_2$ defined as

$$f_*(\varepsilon_1, \dots, \varepsilon_{n-1}) = \varepsilon_{n/2} \pmod 2$$

is called the **parity invariant**.

Before we state our main results, we need to specify the set of regions $\mathcal{B} \in \mathcal{R}_{\mathbf{T}_n}$. A region $\Gamma \in \mathcal{R}$ is called *row-convex* (*column-convex*) if every horizontal (vertical) line either intersects Γ in an interval or does not intersect it at all (see Fig. 1.15). Let $\mathcal{B}_{rc}$ be a set of tileable row-convex simply connected regions. The main result of this paper is the following theorem.

Theorem 1.4. *Let $\mathcal{B} = \mathcal{B}_{rc}$ be as above. Then:*

1) *When $n = 2m + 1$, $\mathbb{G}(\mathbf{T}_n, \mathcal{B}) \simeq \mathbb{Z}^{m+1}$ and the maps $f_0, f_1, \dots, f_m$ form an independent basis of invariants.*

2) *When $n = 2m$, $\mathbb{G}(\mathbf{T}_n, \mathcal{B}) \simeq \mathbb{Z}^m \times \mathbb{Z}_2$ and the maps $f_0, f_1, \dots, f_{m-1}, f_*$ form an independent basis of invariants.*

When $n = 2$ Theorem 1.4 says that the area and parity invariants form an independent basis. Analogously, when $n = 3$ Theorem 1.4 says that, aside from the area invariant f_0, there exists one other nontrivial tile invariant $f_1 : \mathbf{T} \to \mathbb{Z}$, where

$$f_1(\mathbf{10}) = 1\,, \quad f_1(\mathbf{01}) = -1\,, \quad f_1(\mathbf{00}) = f_1(\mathbf{11}) = 0$$

(see Fig. 1.12). In a different form this invariant was discovered by Conway and Lagarias in [CL] (see also [T]). To say that f_1 is an invariant is equivalent to saying that:

$$\#\mathbf{10} - \#\,\mathbf{01} = Const.$$

This means that the number of times the **10** tromino occurs in a tiling minus the number of times the **01** tromino occurs in the same tiling of a region Γ depends only on the region Γ, and not on the tiling.

Here is another nontrivial invariant that exists for all $n > 1$.

Definition 1.5. Consider the map $f_\bullet : \mathbf{T}_n \to \mathbb{Z}_2$ defined as follows:

$$f_\bullet(\varepsilon_1, \dots, \varepsilon_{n-1}) = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{n-1} \pmod 2.$$

We call $f_\bullet$ the **height invariant**.

The reason why $f_\bullet$ is called the *height invariant* can easily be seen from the picture. Consider the smallest rectangular box the ribbon tile τ can fit in (see Fig 1.16). Then $f_\bullet(\tau) = a - 1 \pmod 2$, where a is the height of the rectangle. This invariant was considered earlier in connection with certain characters of the

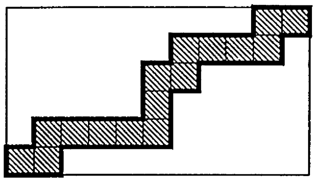


FIGURE 1.16.

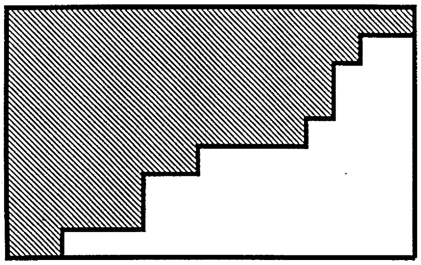


FIGURE 1.17.

symmetric group S_N (see [R], [JK], [S]). Namely, it corresponds to signs in the Murnaghan–Nakayama summation formula for computing the character values on the conjugacy classes (n^a), where $N = a \cdot n$ (see [JK], [M] for details). Observe that

$$f_\bullet = \begin{cases} f_1 + \dots + f_{m-1} + f_m \mod 2, & n = 2m+1, \\ f_1 + \dots + f_{m-1} + f_* \mod 2, & n = 2m. \end{cases}$$

This proves that $f_\bullet$ is indeed an invariant provided Theorem 1.4 holds.

Now let us say a few words about how Theorem 1.4 is proved. We shall present a finite set of moves which preserve the invariants but enable us to get from any tiling to any other. Formally, let $\mathcal{B}_y$ be the set of row- and column-convex regions such that when fit into the smallest possible box they contain the upper right, upper left and the lower left corner of the box (see Fig. 1.17).

Theorem 1.6. *Let $\mathcal{B} = \mathcal{B}_y$ be as above. For every $n > 1$ there is a finite set of at most $n\ 4^n$ moves such that any tiling of $\Gamma \in \mathcal{B}$ by $\mathbf{T}_n$ can be transformed by a sequence of moves to any other such a tiling.*[*]

When $n = 2$ we need only one move (see Fig 1.8). When $n = 3$ we already need 6 moves (see Fig 1.18). Together with Theorem 1.6, we prove this in section 3.

To finish the introduction, let us compare the definition of tile invariants with the *generalized coloring arguments* introduced by Conway and Lagarias (see [CL]).

Let G be an abelian group, not necessarily finite, and e its identity element. A map $f : \mathcal{R} \to G$ is called a *coloring map* if for every region $\Gamma \in \mathcal{R}$ we have

$$f(\Gamma) = f(x_1) + \dots + f(x_{|\Gamma|}),$$

where $x_1, \dots, x_{|\Gamma|}$ are the squares in Γ. Of course, f is defined by its values on all 1×1 squares.

[*] Ron Adin points out that the minimum number of local moves is exactly $\binom{|\mathbf{T}_n|}{2}$. The calculation uses our analysis in section 3. This gives $\binom{4}{2} = 6$ moves for $n = 3$, and $\binom{8}{2} = 28$ for $n = 4$.

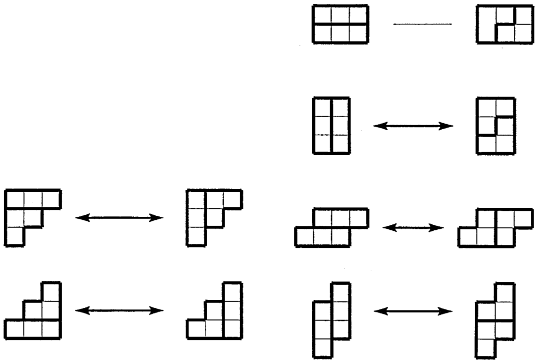


FIGURE 1.18.

A coloring map $f : \mathcal{R} \to G$ is called $\mathbf{T}$-*coloring* if $f(\Gamma) = e$ for all $\Gamma \in \mathcal{R}_{\mathbf{T}}$. In a sense the $\mathbf{T}$-coloring maps are complementary to tile invariants, which are defined only on tileable regions.

It is clear that to show that f is a $\mathbf{T}$-coloring map, all we need is to check that $f(\tau') = e$ for all translations τ' of a tile $\tau \in \mathbf{T}$. If this is the case, we say that f gives a *coloring argument* for a set of tiles $\mathbf{T}$. The idea is that we show that now $f(\Gamma) = e$ becomes a necessary condition for $\mathbf{T}$-tileability, which is usually easy to check. Various arguments can be found in [G].

To give an example, recall the argument we used to prove untileability by dominoes (see Fig. 1.2). It was, basically, a map $f : \mathcal{R} \to \mathbb{Z}$, defined by

$$f(i,j) = \begin{cases} 1\,, & \text{if } i+j = 1 \mod 2, \\ 0\,, & \text{if } i+j = 0 \mod 2. \end{cases}$$

The map f is an example of a $\mathbf{T}$-coloring map. Indeed, observe that, by definition, for both vertical and horizontal dominoes τ_1, τ_2 (see Fig. 1.1) we have $f(\tau_1) = f(\tau_2) = 0$.

It turns out that coloring arguments cannot be used to prove Theorems $0.1 - 0.3$. In particular, we have the following result.

Theorem 1.7. *Let* $\mathbf{T}$ *be the set of tiles shown in Figure 0.5. Consider a rectangle* $\Gamma = [5 \cdot a \times 5 \cdot b]$*, where* a *and* b *are odd. Then for any* $\mathbf{T}$*-coloring map* $f : \mathcal{R} \to G$ *we have* $f(\Gamma) = e$.

Recall that by Theorem 0.1 the region Γ in Theorem 1.7 is not $\mathbf{T}$-tileable. Thus Theorem 0.1 cannot be proved by the use of coloring arguments only. Theorem 1.7 and its analogs for other sets of tiles will be proved in section 8.

Let us note that coloring maps can be used not only to prove untileability but also to find some tile invariants. Here is how this can be done.

Let $\mathbf{T}$ be a set of tiles, and G an abelian group. Consider a map $g : \mathbf{T} \to G$. Suppose $f : \mathcal{R} \to G$ is a coloring map such that $f(\tau') = g(\tau)$ for all translations τ' of a tile $\tau \in \mathbf{T}$. Then there exists a tile invariant $\widehat{g} : \mathcal{R}_{\mathbf{T}} \to G$ such that $\widehat{g}(\tau) = g(\tau)$ for all $\tau \in \mathbf{T}$. We call this an *extended coloring argument* corresponding to the map f.

For example, let $\mathbf{T} = \mathbf{T}_2$ be the set of dominoes. Consider the coloring map $f : \mathcal{R} \to \mathbb{Z}_2$ defined by $f(i,j) = j \mod 2$. In a different form the map f was considered earlier (see Figure 1.7). We have $g(\tau_1) = 1$ and $g(\tau_2) = 0$ (see Figure 1.1), which proves that the map $f_\bullet$ (see Definition 1.5) is indeed an invariant if $n = 2$.

One can try to use the extended coloring argument to get all the tile invariants for a given set of tiles $\mathbf{T}$ and a set of regions $\mathcal{B} \subset \mathcal{R}_{\mathbf{T}}$. It is easy to see that this is impossible if the tile counting group $\mathbb{G}(\mathbf{T}_n, \mathcal{B}) \supsetneq \mathbb{G}(\mathbf{T}_n, \mathcal{R}_{\mathbf{T}})$, i.e. if there exists a tile invariant for the set of regions $\mathcal{B}$ that is not an invariant for the set of all $\mathbf{T}$-tileable regions $\mathcal{R}_{\mathbf{T}}$. It turns out that in the case of the ribbon tiles $\mathbf{T}_n$, $n > 2$, neither convexity nor parity invariants follow from extended coloring arguments. Analogously, for the height invariant we have the following result.

Theorem 1.8. *Let $\mathbf{T}_n$, $n > 1$, be a set of ribbon tiles. Then the height invariant $f_\bullet$ follows from the extended coloring argument if and only if $n = 2$.*

The proof of Theorem 1.8 is given in section 9. Of course, the "if" part is already proven. Incidentally, proving this theorem was the original goal of this work.

The rest of the paper is constructed as follows. In section 2 we define a rim hook correspondence, which is used in section 3 to prove Theorem 1.6. In section 4 we check that the coloring maps defined in Definitions $1.1 - 1.3$ are invariant under the local moves defined in Theorem 1.5. In section 5 we present a technique for working with tile invariants which enables us to extend the set of regions. In section 6 we show that there are no ribbon tile invariants other than those given in Theorem 1.4. Then we prove the main theorem itself.

The second part of the paper contains several applications of the main result. In section 7 we use ribbon tile invariants to find necessary conditions for tileability. We prove Theorems $0.1 - 0.3$ and a few other similar results. In section 8 we define and analyze signed tilings and prove Theorem 1.7 along with other related results. This section is motivated by the work [CL] of Conway and Lagarias, although we were able to avoid the use of combinatorial group theory. In section 9 we prove Theorem 1.8 and explore the connection between extended tile arguments and tile invariants. Finally, in section 10 we present several conjectures and open problems.

This work was done while the author was a postdoctoral fellow at MIT. The research was supported by a National Science Foundation Postdoctoral Research Fellowship.

2. The rim hook bijection

Let us recall some standard notation in combinatorics related to the representation theory of the symmetric group (see e.g. [JK], [M]).

A *partition* is a nonincreasing integer sequence $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l > 0$. With each partition λ we associate a region Γ_λ, called a *Young diagram* or a *Ferrers shape*, defined as follows:

$$\Gamma_\lambda = \{(i,j) \in \mathbb{Z}^2 \mid 1 \leq i \leq l,\, 1 \leq j \leq \lambda_i\}.$$

See Figure 2.1 for the Young diagram associated with the partition $(5,5,4,3,1)$. Denote $|\lambda| = \lambda_1 + \cdots + \lambda_l = |\Gamma_\lambda|$.

A *skew Young diagram* $\Gamma_{\lambda\setminus\mu}$ is the set theoretic difference of the Young diagrams associated with the partitions λ and μ:

$$\Gamma_{\lambda\setminus\mu} = \Gamma_\lambda \setminus \Gamma_\mu.$$

For example, the skew Young diagram $\Gamma_{(5,5,4,3,1)\setminus(3,2)}$ is shown in Figure 2.2.

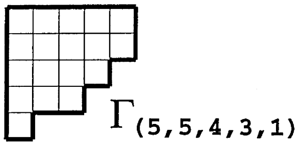


FIGURE 2.1.

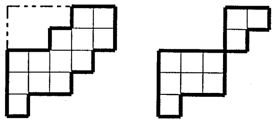
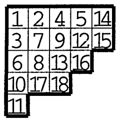


FIGURE 2.2.



1	2	4	5	14
3	7	9	12	15
6	8	13	16	
10	17	18		
11				

FIGURE 2.3.

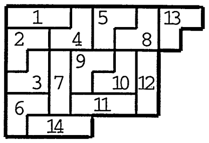


FIGURE 2.4.

To simplify the notation we will use λ to denote both the partition and the corresponding region Γ_λ, which we also call the Young diagram of shape λ. By $\mathcal{R}_y$ and $\mathcal{R}_{sy}$ we denote the set of all Young diagrams and the set of all skew Young diagrams, respectively.

By $\lambda \circ \mu$ we denote the skew Young diagram obtained as the disjoint union of the Young diagrams λ and μ where μ is located to the right and above λ. The example of $(3,3,1) \circ (2,1) = (5,4,3,3,1) \setminus (3,3)$ is shown in Figure 2.2.

A *Young tableau* is a Young diagram λ filled with integer numbers which increase in rows and columns (see Fig. 2.3). A Young tableaux is called *standard* if these numbers are $1,\dots,|\lambda|$. We can think of a Young tableau as a flag of Young diagrams $\emptyset = \lambda^1 \subset \lambda^2 \subset \ldots \subset \lambda^n = \lambda$. A *skew Young tableau* is defined analogously.

Recall that by $\mathbf{T}_n$ we denote a set of ribbon tiles with n squares (see Figs. 1.12-1.14). A *rim hook tableau* is a tiling of a Young diagram Γ_λ by ribbon tiles $\tau \in \mathbf{T}_n$ filled with numbers $1, 2, \dots, |\lambda|/n$ (squares in the same tile are filled with the same number), and such that squares of tiles with greater numbers are located either to the right or below squares of tiles with smaller numbers (see Fig. 2.4). Again we can think of a rim hook tableau as a flag of Young diagrams.

The *rim hook bijection* φ maps Young diagrams λ, $|\lambda| = m \cdot n$, tileable by $\mathbf{T}_n$ into n-tuples of Young diagrams $(\nu^1,\dots,\nu^n)$, $|\nu^1| + \cdots + |\nu^n| = m$. The bijection φ is designed in such a way that whenever we add a ribbon tile to the diagram λ on the outside, there exist i, $1 \ \le i \le n$, such that ν^i gets a square on the outside (see Fig. 2.5). If we think of rim hook tableaux as flag sequences of tileable Young diagrams, the rim hook bijection maps these flag sequences into the n-tuples of Young tableaux filled with numbers $1,\dots,m$ which are increasing in rows and columns in each tableau. It is known that this establishes a bijection between the rim hook tableau of a fixed tileable Young diagram λ and the n-tuples of Young tableaux with shapes $\varphi(\lambda) = (\nu^1,\dots,\nu^n)$ which are filled with numbers $1,\dots,m$, where $|\lambda| = m \cdot n$ (see e.g. [JK], [SW]).

The easiest way to understand the rim hook bijection is to look at Figure 2.6. Take a rim hook tableau tiled with 14 tiles $\tau_i \in \mathbf{T}_3$ and rotate it counterclockwise 135° degrees. Then project all hooks on the horizontal axis, preserving their labels and relative order. Split the "shadows" into three (n in the general case) separate sets of "shadows" depending on their horizontal coordinate $\bmod 3$. Then simply

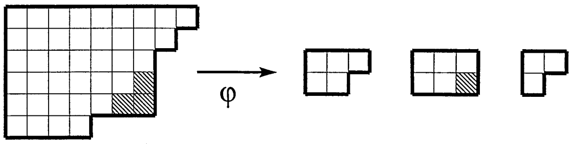


Figure 2.5.

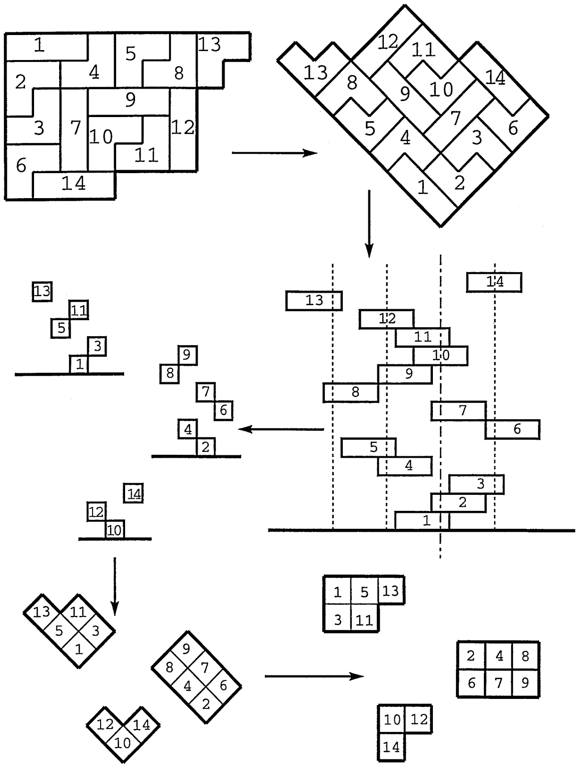


Figure 2.6.

shorten the shadows and reverse the procedure. At the end we get three Young tableaux filled with the numbers $1, \dots, 14$ (see Figure 2.6).

Theorem 2.1. *The map φ defined above is a one-to-one correspondence.*

The theorem goes back to Nakayama and Robinson (see [R], [JK]). In modern times it was rediscovered by Stanton and White (see [SW], [FS]) and is sometimes attributed to them.

Another way to think of the rim hook bijection is to say that it establishes a bijection between rim hook tableaux of shape λ and standard Young tableaux of the skew shape $\nu^1 \circ \cdots \circ \nu^n$. We shall use this interpretation in the next section. Various proofs and applications of the theorem can be found in [JK], [FS], [S].

3. Local moves

Let $\lambda \setminus \mu$ be a skew Young diagram. We define *local moves* on a set of standard Young tableaux of shape $\lambda \setminus \mu$ as follows. Take a pair of numbers i and $i+1$, $1 \le i < |\lambda|$, and exchange them if they lie in different rows and columns. We claim that, using these moves, one can start with any standard Young tableau of shape $\lambda \setminus \mu$ and get any other such tableau.

Formally, let $\Omega(\lambda \setminus \mu)$ be a graph with vertices all standard Young tableaux of shape $\lambda \setminus \mu$ and edges obtained by applying local moves. An example with $\lambda = (3,2)$ and $\mu = \emptyset$ is shown in Figure 3.1.

Theorem 3.1. *Let $\lambda \setminus \mu$ is a skew Young diagram. Then its graph $\Omega(\lambda \setminus \mu)$ is connected.*

This result is known and not hard to prove. Some generalizations and applications can be found in [BW], [BK].

Sketch of Proof. Introduce an orientation of edges of the graph $\Omega(\lambda \setminus \mu)$ by distinguishing situations when a local move exchanges i and $i+1$ with $i+1$ lying to the right and above i from those where $i+1$ lies to the left and below i (see Figure 3.1). Observe that the orientation is acyclic and has exactly one sink. This proves the result. □

Consider what happens if we apply the bijection φ to the vertices of a graph $\Omega(\lambda \setminus \mu)$ in Theorem 3.1. Fix a skew Young diagram $\nu = \nu^1 \circ \cdots \circ \nu^n$. Define *local moves* on a set of rim hook tableaux of shape $\lambda = \varphi^{-1}(\nu^1, \dots, \nu^n)$ by the image of the corresponding local moves on a standard skew Young tableaux. Observe that squares in a Young tableau of shape ν correspond to rim hooks in a rim hook tableau of shape λ. Thus the corresponding local moves on a set of rim hook tableaux will preserve all the rim hooks except two. Since these rim hooks have consequent labels, together they form a skew shape which has exactly two rim hook tableaux (see Figure 3.2). Note that the two rim hooks may lie far from each other, in which case the local move is just relabeling of their numbers. When $n = 1$ this is the only case that occurs.

Now we are ready to state an analog of Theorem 3.1. Denote by $\Omega_n(\lambda \setminus \mu)$ a graph with rim hook tableaux as vertices and edges connecting those pairs of tableaux which have the same set of all but two rim hooks.

Theorem 3.2. *Let λ be a Young diagram tileable by $\mathbf{T}_n$. Then its graph $\Omega_n(\lambda)$ is connected.*

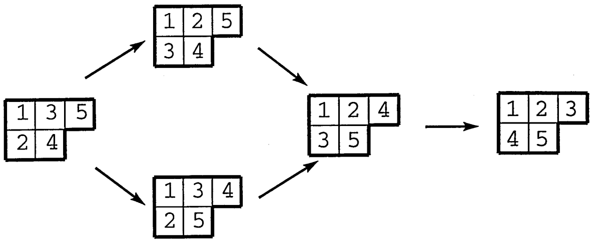


Figure 3.1.

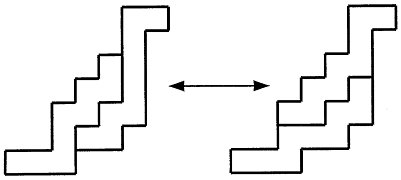


Figure 3.2.

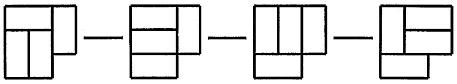


Figure 3.3.

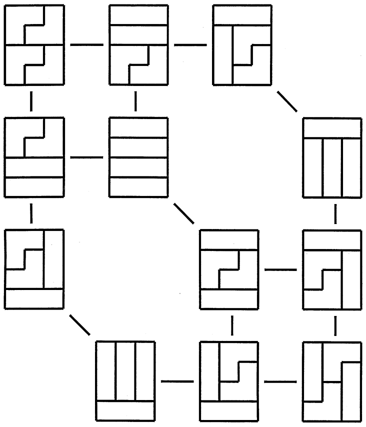


Figure 3.4.

Note that here we do not claim that $\Omega_n(\lambda \setminus \mu)$ is connected for any tileable skew Young diagram $\lambda \setminus \mu$. It is true and can be proved by the straightforward generalization of the rim hook bijection. We avoid the use of this natural generalization for the purpose of studying ribbon tile invariants.

Proof. Define $(\nu^1, \dots, \nu^n) = \varphi(\lambda)$ to be the image of λ under the rim hook correspondence. Consider $G = \Omega(\nu^1 \circ \cdots \circ \nu^n)$. The correspondence φ^{-1} maps vertices of G onto $\Omega_n(\lambda)$ and edges onto edges. In other words, $\varphi^{-1}(G)$ is a subgraph of $\Omega_n(\lambda)$. By Theorem 3.1 the graph G is connected. Therefore the graph $\Omega_n(\lambda)$ is also connected. □

Now let us make a graph on the ribbon tilings of λ. The idea is to erase the labels in the rim hook tableaux and connect those that were connected before.

Formally, consider a graph $\Theta_n(\lambda)$ with vertices being all ribbon tilings $s \in \mathcal{S}(\Gamma_\lambda, \mathbf{T})$ of a fixed shape λ. Define the edges to be the pairs of tilings that differ by exactly two tiles. Two examples of such graphs $\Theta_2(3,3,2)$ and $\Theta_3(3,3,3,3)$ are shown in Figure 3.3 and Figure 3.4 respectively.

Theorem 3.3. *Let λ be a Young diagram tileable by $\mathbf{T}_n$. Then the graph $\Theta_n(\lambda)$ is connected.*

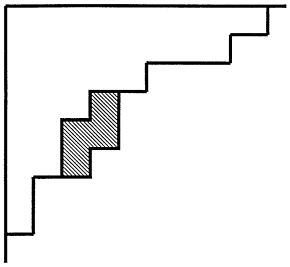


FIGURE 3.5.

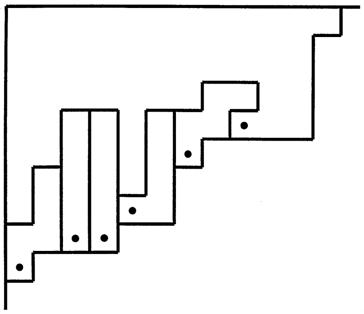


FIGURE 3.6.

Proof. Consider a map $\iota : \Omega_n(\lambda) \to \Theta_n(\lambda)$ which maps rim hook tableaux to ribbon tilings by erasing labels of tiles. By definition ι maps edges of $\Omega_n(\lambda)$ into edges of $\Theta_n(\lambda)$. Therefore, in order to prove that $\Theta_n(\lambda)$ is connected, all we need to show is that each vertex has a preimage. Indeed, if this is true, in order to find a path between two vertices in $\Theta_n(\lambda)$ we simply take their preimages, find a path between them in $\Omega_n(\lambda)$, and then map it back to $\Theta_n(\lambda)$.

In other words, the theorem in now reduced to the following lemma. □

Lemma 3.4. *Every ribbon tiling of a Young diagram λ admits a labeling which makes it a rim hook tableau of shape λ.*

Proof. We prove the lemma by induction on the number of squares $|\lambda|$. The base case is trivial. Fix a Young diagram λ. By the *border strip* of λ we mean the set of squares $(i, j) \in \lambda$ such that $(i + 1, j + 1) \notin \lambda$. A tile τ is called a *border tile* if it lies in the border strip (see Figure 3.5). We claim that every ribbon tiling of λ must contain at least one border tile. If we find such a tile τ, label it with the largest number. Then there are no tiles that lie to the right of or below τ, and we can proceed by induction with $\lambda \setminus \tau$.

In order to find a border ribbon, start with the lower left corner. It must belong to some tile. This tile has this square as its starting square. If it is not a border tile, find the first border square that is not in that tile. It must belong to some other tile. This tile also has this square as its starting square. Keep on doing so until we find the border tile. It must always exist, since the top right corner must also belong to some tile, and this square cannot be a starting square of any tile unless $n = 1$, in which case it is a border tile by definition (see Figure 3.6).

This proves the induction step together with the lemma. The lemma in turn implies Theorem 3.3. □

Proof of Theorem 1.6. Observe that the set or regions $\mathcal{B} = \mathcal{B}_y$ is exactly the set of all Young diagram shapes. Take the moves to be as described above. The number of different moves is bounded by the the number of pairs of ribbon tiles aligned to each other. The latter number is easily bounded by $n \cdot 4^n$, which proves the theorem. □

Finally, we would like to note that in the proof of Lemma 3.4 rim hooks need not be of the same length.

4. Tile invariants

Let $\mathcal{B} = \mathcal{B}_y$ be a set of tileable Young diagram shapes. In the previous section we showed that all the tilings of $\lambda \in \mathcal{B}_y$ can be obtained from each other by a finite set of moves. Here we show that the maps f_i, $0 \le i < n/2$, are constant along these moves. In other words, we shall prove Theorem 1.4 for the set of regions $\mathcal{B}_y$.

Let us look at the structure of the moves we introduced in section 3. Consider a large enough example, shown in Figure 3.2. Recall that each ribbon tile is encoded by a sequence $(\varepsilon_1, \dots, \varepsilon_{n-1})$ of $n-1$ zeroes and ones. In this notation our pair of tiles is mapped into a similar pair:

$$\mathbf{00110101}, \mathbf{01011110} \to \mathbf{00100101}, \mathbf{01010110}.$$

Subtracting the sequences as vectors, we get the vectors $(0,0,0,-1,0,0,0,0)$ and $(0,0,0,0,-1,0,0,0)$. In other words, the first tiles in a pair differ at the fourth place, where 1 becomes 0. Respectively, the second tiles in a pair differ at the fifth place, where 1 again becomes 0. Note that all tiles contain $n = 9$ squares, and $5 + 4 = 9$. We claim that this is a general observation.

Lemma 4.1. *In every move defined above one tile sequence changes from* 1 *to* 0 *(or from* 0 *to* 1*) at some place i while another has exactly the same change at place $n - i$.*

Proof. The proof is done by the following observation. Note that each skew Young diagram shape which corresponds to a move can be broken into three parts by the number of squares in a diagonal parallel to the line $y = -x$ (see Figure 4.1). After the move, the southwest and northeast part remain the same, while the middle part remain divided into two identical small ribbon tiles which get switched now. Therefore, the differences in tile sequences occur only in places where the southwest and northeast parts touch the middle part. If the southwest part was touching the upper of two small ribbon tiles, it now touches the lower one. This means that at place i the number in a sequence changed from 1 to 0 (see Figure 4.1). Respectively, the northeast part was touching the lower of two small ribbon tiles and now is touching the upper one. This means that at place $n-i$ the number in a sequence changed from 1 to 0.

The second case, when the southwest part was touching the upper of two small ribbon tiles before the move, and the lower tile after the move, is analogous. This proves the lemma. □

Now recall the definitions of the convexity invariants:

$$f_i(\varepsilon_1, \dots, \varepsilon_{n-1}) = \varepsilon_i - \varepsilon_{n-i},$$

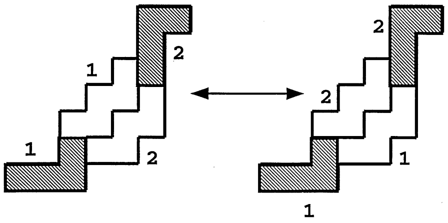


Figure 4.1.

where $1 \le i \le \left\lfloor \frac{n-1}{2} \right\rfloor$. Analogously, the parity invariants are defined by

$$f_*(\varepsilon_1, \dots, \varepsilon_{n-1}) = \varepsilon_m \mod 2,$$

where n is even and $m = n/2$.

Lemma 4.2. *The maps f_i and f_* are invariant under the local moves.*

Proof. The lemma follows easily from Lemma 4.1. Indeed, as we showed above, if a move changes tiles τ_1, τ_2 into tiles τ_1', τ_2', then

$$f_i(\tau_1') = f_i(\tau_1) \pm 1, \quad f_i(\tau_2') = f_i(\tau_2) \mp 1,$$

where the different sign in the second equation comes from the minus sign in $f_i = \varepsilon_i - \varepsilon_{n-i}$.

Therefore

$$f_i(\tau_1') + f_i(\tau_2') = f_i(\tau_1) + f_i(\tau_2),$$

which proves that the maps f_i are invariant under the local moves.

The case of a parity invariant is slightly different, since here we do not have opposite signs. Instead we have

$$f_*(\tau_1') = f_*(\tau_1) \pm 1, \quad f_*(\tau_2') = f_*(\tau_2) \pm 1.$$

Therefore

$$f_i(\tau_1') + f_i(\tau_2') = f_i(\tau_1) + f_i(\tau_2) \pm 2 = f_i(\tau_1) + f_i(\tau_2) \pmod 2,$$

which proves that the map f_* is invariant under the local moves. This finishes the proof of the lemma. □

Recall that by f_0 we denote the area invariant.

Corollary 4.3. *Let $\mathcal{B} = \mathcal{B}_y$ be a set of tileable Young diagram shapes. Then, when $n = 2m + 1$, the maps $f_0, f_1, \dots, f_m$ are the ribbon tile invariants. Analogously, when $n = 2m$, the maps $f_0, f_1, \dots, f_{m-1}, f_*$ are the ribbon tile invariants.*

Proof. By Lemma 4.2 the maps f_i and f_* are invariant under the local moves. By Theorem 3.3 we can get any ribbon tiling of a Young diagram from any other. Therefore these maps are indeed invariants on a set of tileable Young diagram shapes. This proves the corollary. □

Note that Corollary 4.3 proves only one part of Theorem 1.4 for the set of regions $\mathcal{B} = \mathcal{B}_y$. The second part, which states that these invariants form a basis, will be proven in section 6.

5. Increasing the set of regions

In this section we will generalize Corollary 4.3 from the set of Young diagram shapes to the set $\mathcal{B}_{rc}$ of all row-convex regions. As an intermediate step we use a set of all skew Young diagram shapes.

Our approach is based on the following general observation.

Lemma 5.1. *Let $\mathbf{T}$ be a set of tiles, and let $\mathcal{B}_1 \subset \mathcal{B}_2$ be two sets of $\mathbf{T}$-tileable regions. Suppose for each region $\Gamma_2 \in \mathcal{B}_2$ there is a region $\Gamma_1 \in \mathcal{B}_1$ such that $\Gamma_1 \supset \Gamma_2$ and $\Gamma_1 \setminus \Gamma_2$ is $\mathbf{T}$-tileable. Then, if $f : \mathcal{B}_2 \to G$ is an invariant on $\mathcal{B}_1$, it is also an invariant on $\mathcal{B}_2$.*

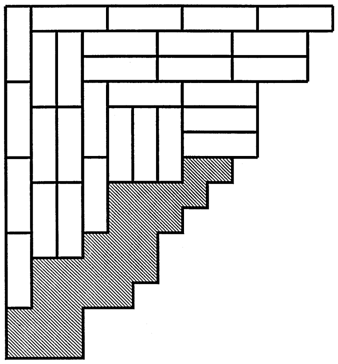


FIGURE 5.1.

Proof. We need to show that for any $\Gamma_2 \in \mathcal{B}_2$ all tilings $s \in \mathcal{S}(\Gamma_2, \mathbf{T})$ have the same G-value of f. We know that for any $\Gamma_2 \in \mathcal{B}_2$ there is a region $\Gamma_1 \in \mathcal{B}_1$, $\Gamma_1 \supset \Gamma_2$, such that $\Gamma_1 \setminus \Gamma_2$ is $\mathbf{T}$-tileable. Therefore the set of tilings $\mathcal{S}(\Gamma_2, \mathbf{T})$ is in a correspondence with a subset $\mathcal{S}' \subset \mathcal{S}(\Gamma_1, \mathbf{T})$ such that all the tilings $s \in \mathcal{S}'$ have the same fixed tiling of $\Gamma_1 \setminus \Gamma_2$. By definition the value of f is the same on all tilings of $\mathcal{S}(\Gamma_1, \mathbf{T})$. Since it is fixed on $\Gamma_1 \setminus \Gamma_2$, it must be the same on $\mathcal{S}(\Gamma_2, \mathbf{T})$. This finishes the proof. □

We call the set $\mathcal{B}_2$ of $\mathbf{T}$-tileable regions *reducible* to $\mathcal{B}_1$, if $\mathcal{B}_1 \subset \mathcal{B}_2$ and they satisfy the conditions of Lemma 5.1. Of course, if $\mathcal{B}_3$ is reducible to $\mathcal{B}_2$ and $\mathcal{B}_2$ is reducible to $\mathcal{B}_1$, then $\mathcal{B}_3$ is reducible to $\mathcal{B}_1$.

Lemma 5.2. *Let $\mathcal{B}_{sy}$ be the set of $\mathbf{T}_n$-tileable skew Young diagram shapes, and $\mathcal{B}_y$ the set of $\mathbf{T}_n$-tileable ordinary Young diagram shapes. Then $\mathcal{B}_{sy}$ is reducible to $\mathcal{B}_y \subset \mathcal{B}_{sy}$.*

Proof. Indeed, all we need to prove is that every skew Young diagram can be imbedded in an ordinary Young diagram so that their difference is tileable by the ribbon tiles. There is an easy way to do that just by using the horizontal and the vertical tiles.

The idea is shown in Figure 5.1. We start with the rightmost column of a skew Young diagram shape and move to the left. Whenever we move left, add on top a column of vertical tiles until they equal or exceed the column on the right. If they do exceed the column on the right, for each exceeding square add to the right a row of horizontal tiles until they equal or exceed the row below. In example shown in Figure 5.1 we do nothing for the first two columns. For the third column from the right we add one vertical tile and two horizontal. We add just one vertical tile for each of the next two columns. For the sixth column we are forced to add three vertical and two horizontal tiles, etc.

We stop when we are finished with the last column (the ninth in case of Figure 5.1). By construction we always have a Young diagram shape to the right of the building column. Therefore the resulting shape is also a Young diagram shape. This proves the lemma. □

Remark 5.3. In [Pa] we use this construction to define a generalization of the rim hook bijection for skew shapes. Note also that Lemma 5.2 can be generalized for

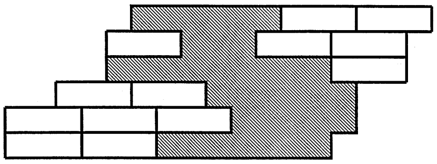


FIGURE 5.2.

any set of tiles $\mathbf{T}$ which contains a horizontal and a vertical tile, not necessarily of the same length.

Lemma 5.4. *Let $\mathcal{B}_{sy}$ be the set of $\mathbf{T}_n$-tileable skew Young diagram shapes, and $\mathcal{B}_{rc}$ the set of $\mathbf{T}_n$-tileable row-convex regions. Then $\mathcal{B}_{rc}$ is reducible to $\mathcal{B}_{sy} \subset \mathcal{B}_{rc}$.*

Proof. Indeed, all we need to prove is that every row-convex region can be imbedded in a skew Young diagram so that their difference is tileable by ribbon tiles. There is an easy way to do that just by using just horizontal tiles.

The idea is shown in Figure 5.2, and is similar to the idea used in Lemma 5.2. We start with the top row of our row-convex region and move to the bottom row. Each time we move down we add a row of horizontal tiles to the left so that they equal or exceed the row above. When get to the bottom we start adding rows of horizontal tiles to the right of the region in such a way that each row equals or exceeds the row below (see Figure 5.2). At the end we get a skew Young diagram shape, which proves the lemma. □

Corollary 5.5. *The statement of Corollary* 4.3 *holds for the set $\mathcal{B}_{rc}$ of row-convex $\mathbf{T}_n$-tileable regions.*

Proof. By Lemmas 5.2 and 5.4, $\mathcal{B}_{rc}$ is reducible to $\mathcal{B}_y$. By Lemma 5.1 this implies that every ribbon tile invariant for the set of regions $\mathcal{B}_y$ is also an invariant for the set of regions $\mathcal{B}_{rc}$. Together with Corollary 4.3, this proves the result. □

6. The tile counting group

Here we will prove that there are no invariants other than those which follow from convexity, parity and area invariants. Together with Corollary 5.5 this implies the main result of the paper, Theorem 1.4.

The idea is to show that every invariant is completely defined by its values on the horizontal and two-row ribbon tiles.

Denote $\tau_0 = \mathbf{0000} \ldots \mathbf{0}$, $\tau_i = \mathbf{0} \ldots \mathbf{010} \ldots \mathbf{0}$ ($\mathbf{1}$ is in the i-th place), $1 \le i \le n-1$. Let $\mathbf{B}_n \subset \mathbf{T}_n$ be the set of tiles τ_i, where $0 \le i \le m = \left\lfloor \frac{n}{2} \right\rfloor$.

Lemma 6.1. *Let $f_1, f_2 : \mathcal{B}_{rc} \to G$ be two ribbon tile invariants. We claim that if for every $\tau \in \mathbf{B}_n$ we have $f_1(\tau) = f_2(\tau)$, then $f_1 \equiv f_2$.*

Proof. We need to show that for every ribbon tile $\tau \in T_n$ we have $f_1(\tau) = f_2(\tau)$. This would immediately imply that $f_1 \equiv f_2$. We prove it by induction on the height ht of a ribbon tile (see the Introduction):

$$ht(\tau) = 1 + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{n-1}.$$

First we prove the base of the induction. If $ht(\tau) = 1$, then τ is a horizontal tile $\tau_0 \in \mathbf{B}_n$. If $ht(\tau) = 2$, then $\tau \in \mathbf{B}_n$ or $\tau = \tau_i$, $m < i \le n-1$. Observe that τ_i and

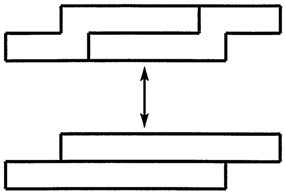


FIGURE 6.1.

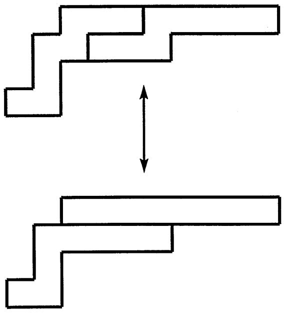


FIGURE 6.2.

τ_{n-i} form a two-row skew Young diagram shape which can also be divided into two horizontal tiles (see Figure 6.1). Therefore if invariants f_1, f_2 agree on $\mathbf{B}_n$ they must also agree on all the two-row ribbon tiles.

Now suppose the claim holds for all tiles $\tau \in \mathbf{T}_n$ with $ht(\tau) < k$. Let $\tau \in \mathbf{T}_n$ be a ribbon tile and $ht(\tau) = k$. The sequence corresponding to τ can be presented in the form $(\varepsilon_1, \dots, \varepsilon_i, 1, 0, 0, \dots, 0)$. The two-row ribbon tile τ_{n-i-1} can be aligned with the top two rows of τ to form a skew Young diagram shape (see Figure 6.2). This region can also be divided into a horizontal tile and a tile τ' with a sequence $(\varepsilon_1, \dots, \varepsilon_i, 0, 0, 0, \dots, 0)$ (see Figure 6.2). Note that $ht(\tau') = ht(\tau) - 1 = k - 1$. Therefore if invariants f_1, f_2 agree on all tiles $\tau \in \mathbf{T}_n$, $ht(\tau) < k$, they must also agree on all tiles $\tau \in \mathbf{T}_n$, $ht(\tau) = k$. This proves the induction step and finishes the proof of the lemma. □

Proof of Theorem 1.4. Corollary 5.5 implies that our maps are indeed invariants. All we need to prove now is that they are independent and generate the whole tile counting group.

To show independence, consider values our invariants take on $\mathbf{B}_n \subset \mathbf{T}_n$. The area invariant f_0 is a constant on all $\mathbf{B}_n$, including τ_0. The convexity invariant f_i is nonzero only on the tile τ_i. Analogously, the parity invariant is nonzero only on the tile τ_m, $n = 2\,m$. This immediately implies independence.

Now, Lemma 6.1 proves that a tile invariant is completely determined by its values on $\mathbf{B}_n$. This implies that when $n = 2\,m + 1$ there can be no invariants that are not generated by $f_0,\ f_1,\ \dots\ ,\ f_m$. Therefore $\mathbb{G}(\mathbf{T}; \mathcal{B}) \simeq \mathbb{Z}^{m+1}$.

We still have a little room left when $n = 2\,m$, since the parity invariant takes values in $\mathbb{Z}_2$ rather than in $\mathbb{Z}$. Recall that in the proof of Lemma 6.1 we showed that if f is an invariant, then

$$f(\tau_i) + f(\tau_{n-i}) = 2\, f(\tau_0)$$

(see Figure 6.1). When $i = m$ this gives $2 \cdot f(\tau_m) = 2 \cdot f(\tau_0)$, which implies that there can be no invariants that are not generated by $f_0,\ f_1,\ \dots\ ,\ f_{m-1}$ and f_*. Therefore $\mathbb{G}(\mathbf{T}; \mathcal{B}) \simeq \mathbb{Z}^m \times \mathbb{Z}_2$. This finishes the proof of Theorem 1.4 □

7. Applications to tileability

In this section we use ribbon tile invariants to give new tileability criteria for certain sets of tiles. Among the results we prove Theorems 0.1 − 0.3.

We use the following logic. Let $\mathbf{T} \subset \mathbf{T}_n$ be a subset of ribbon tiles. Suppose Γ is a region which is tileable by $\mathbf{T}_n$. Then we can use tile invariants to find diophantine equations for the number of times each tile $\tau \in \mathbf{T}_n$ occurs in the tilings. When restricted to a smaller set of tiles $\mathbf{T}$, sometimes these equations do not have an integer solution. This would imply that Γ is untileable by $\mathbf{T}$. Generally, having a solution for these equations becomes a *tileability test* which is easy to use in practice.

Here are a few examples when we apply the above logic successfully.

1. Let $\mathbf{T} \subset \mathbf{T}_4$ be the set of four tiles shown in Figure 7.1. We ask for which values (a, b) the rectangle $[a \times b]$ can be tiled by $\mathbf{T}$. Note that since $\mathbf{T}$ is asymmetric, we need to use the following notation: a is a height and b is a width of a rectangle.

Theorem 7.1. *Let* $\mathbf{T}$ *be the set of tiles in Figure 7.1. Then the rectangle* $[a \times b]$ *can be tiled by* $\mathbf{T}$ *if and only if* (a, b) *satisfies one of the following:*

1) $4|a$,
2) $8|b, \ \ a \geq 3$,
3) $2|a, \ \ 4|b$.

Proof. The tileability in cases $1) - 3)$ follows from the existence of tilings of the rectangles $[4 \times 1]$, $[2 \times 4]$ and $[3 \times 8]$ (see Figure 7.2).

To prove untileability in all the other cases we need several observations. First, it is obvious that no rectangle $[1 \times b]$ can be tiled.

Suppose now a is odd, $a \geq 3$. Then $4|b$ and the rectangle $[a \times b]$ can be tiled by a horizontal tile $[1 \times 4] \in \mathbf{T}_4$. Thus the height invariant (see Definition 1.5) $f_\bullet = f_1 + f_2 \pmod 2$, $f_\bullet : \mathcal{B} \to \mathbb{Z}_2$, has the value

$$f_\bullet([a \times b]) = 0.$$

However, $f_\bullet(\tau) = 1$ for all $\tau \in \mathbf{T}$. Therefore in order to be tileable, the rectangle $[a \times b]$ must contain an even number of tiles, i.e. $8|b$ and we are in case 2).

Now we need to show that no rectangle $[a \times b]$ with $a, b = 2 \mod 4$ can be tiled by $\mathbf{T}$. This can be done by coloring arguments (see the Introduction). We will do it in the next section (see Corollary 8.5). □

2. Let $\mathbf{T} \subset \mathbf{T}_5$ be the set of eight tiles shown in Figure 0.5. Note that in this case $\mathbf{T}$ is symmetric under the transposition. Theorem 0.1 claims that a rectangle $[a \times b]$ can be tiled by $\mathbf{T}$ only if $10|a \cdot b$. We claim that an even stronger statement is true.

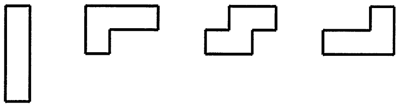


FIGURE 7.1.

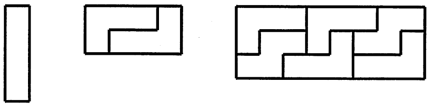


FIGURE 7.2.

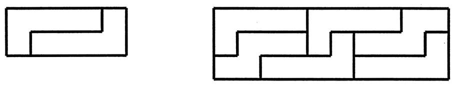


FIGURE 7.3.

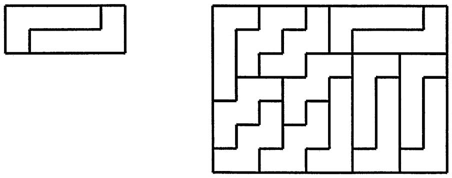


FIGURE 7.4.

Theorem 7.2. *Let* $\mathbf{T}$ *be the set of tiles in Figure 0.5. Then the rectangle* $[a \times b]$ *can be tiled by* $\mathbf{T}$ *if and only if* $10 | a \cdot b$ *and* $a, b > 1$.

Proof. The tileability follows immediately from the existence of tilings of the rectangles $[2 \times 5]$, $[3 \times 10]$ and of the rectangles $[5 \times 2]$, $[10 \times 3]$ transpose to them (see Figure 7.3).

The other direction is similar to the proof of Theorem 7.1. Observe that if a rectangle $[a \times b]$ can be tiled by $\mathbf{T}$, then either a or b can be divided by 5. But then $[a \times b]$ is tileable by either a horizontal tile τ_0 or a vertical tile τ_0', $\tau_0, \tau_0' \in \mathbf{T}_5$. Observe that

$$f_\bullet(\tau_0) = f_\bullet(\tau_0') = 0 \mod 2,$$

where $f_\bullet : \mathcal{B} \to \mathbb{Z}_2$ is a height invariant. Therefore $f_\bullet([a \times b]) = 0$ for any tileable rectangle $[a \times b]$. On the other hand, for any tile $\tau \in \mathbf{T}$ we have $f_\bullet(\tau) = 1$. Therefore there must be an even number of tiles in $[a \times b]$, i.e. $10 | a \cdot b$. This proves the theorem. □

3. Let $\mathbf{T} \subset \mathbf{T}_5$ be the set of six tiles shown in Figure 0.6. Note that in this case $\mathbf{T}$ is also symmetric under the transposition.

Theorem 7.3. *Let* $\mathbf{T}$ *be the set of tiles in Figure 0.5. Then the rectangle* $[a \times b]$ *can be tiled by* $\mathbf{T}$ *if and only if* $10 | a \cdot b$ *and* $a, b \neq 1, 3$.

Proof. The tileability follows immediately from the existence of tilings of the rectangles $[2 \times 5]$, $[7 \times 10]$ and of the rectangles $[5 \times 2]$, $[10 \times 7]$ that are transposed to them (see Figure 7.4).

To show the other direction, use the same reasoning as in the proof of Theorem 7.2. If a rectangle $[a \times b]$ can be tiled by $\mathbf{T}$, then its area must be divisible by 5 and therefore it is tileable by either a horizontal or a vertical tile in $\mathbf{T}_5$. Thus the 1-convexity invariant f_1 is 0 on any tileable $[a \times b]$. On the other hand, $f_1(\tau) = \pm 1$ for each $\tau \in \mathbf{T}$. Therefore $[a \times b]$ has to be tiled by an even number of tiles, and $10 | a \cdot b$. This finishes the proof of the theorem. □

Theorem 7.4. *The triangular shape* Δ_N *can be tiled by tiles shown in Figure 0.6 if and only if* $N \equiv 0, 4, 15, 19 \pmod{20}$.

Proof. First we prove the tileability part. Observe that Δ_4 and Δ_{15} are both tileable (see Figure 7.5). Now, if we have tilings of the regions Δ_M, Δ_N we can construct a

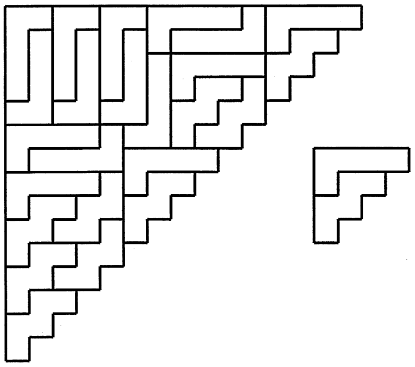


FIGURE 7.5.

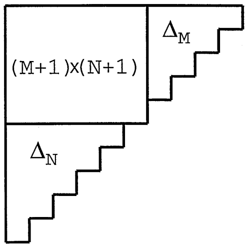


FIGURE 7.6.

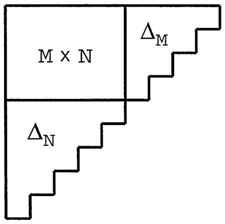


FIGURE 7.7.

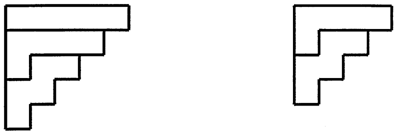


FIGURE 7.8.

tiling of the region Δ_{M+N+1} assuming the rectangle $[(M+1) \times (N+1)]$ is tileable. (see Figure 7.6). Analogously, if we have tilings of the regions Δ_M, Δ_N we can construct a tiling of the region Δ_{M+N} assuming the rectangle $[M \times N]$ is tileable (see Figure 7.7). Since both $[5 \times 16]$ and $[4 \times 15]$ are tileable (see Theorem 7.3), this gives us tilings of $\Delta_{20} = \Delta_{15+5+1}$ and $\Delta_{19} = \Delta_{15+5}$. Going further, if Δ_M is tileable, then Δ_{20+M} is also tileable. This covers all the values $N \equiv 0, 4, 15, 19 \pmod{20}$.

To prove the "only if" part, start by computing the area. We have

$$f_0(\Delta_N) = \frac{N\,(N+1)}{2}.$$

Since the area must be divisible by 5, this gives us $5|N\,(N+1)$ and $N \equiv 0, 4 \pmod 5$. Now, in each of these cases it is easy to see that Δ_N can be tiled by $\mathbf{T}_5$. Indeed, both Δ_4 and Δ_5 can be tiled by $\mathbf{T}_5$ (see Figure 7.8). Since any rectangle $[a \times b]$ is tileable by either a horizontal or a vertical tile in $\mathbf{T}_5$, we can use the construction in Figure 7.7 repeatedly and tile Δ_N for all $N \equiv 0, 4 \pmod 5$.

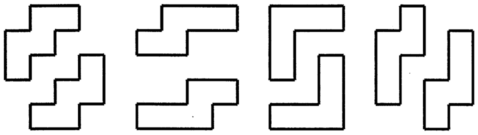


FIGURE 7.9.

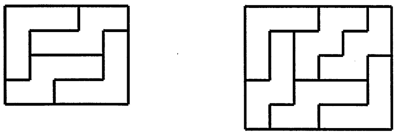


FIGURE 7.10.

Now we can compute the value of the 1-convexity invariant f_1 on Δ_N, $N \equiv 0, 4 \pmod 5$. We have

$$f_1(\Delta_4) = f_1(\Delta_5) = 0 \ , \ f_1(\Delta_{N+5}) = f_1(\Delta_N),$$

which gives us

$$f_1(\Delta_{5\,m+4}) = f_1(\Delta_{5\,m+5}) = 0.$$

Since f_1 takes only the values ± 1 on $\mathbf{T}$, this implies that in order to be tileable Δ_N must have an even number of tiles. Thus $2 | f_0 = \frac{N\,(N+1)}{10}$, and $20 | N\,(N+1)$. This immediately implies $N \equiv 0, 4, 15, 19 \pmod{20}$. □

4. Let $\mathbf{T} \subset \mathbf{T}_5$ be the set of eight tiles shown in Figure 7.9. As before, we solve the tileability problem for rectangular and triangular shapes.

Theorem 7.5. *Let* $\mathbf{T}$ *be the set of tiles in Figure 7.9. Then a rectangle* $[a \times b]$ *can be tiled by* $\mathbf{T}$ *if and only if* (a, b)*,* $a \le b$*, satisfies one of the following:*

1) $10 | a \cdot b$*,* $a \ge 8$*,*
2) $a = 4, 6,$ $5 | b$*,*
3) $a = 5,$ $2 | b$*.*

Proof. The tileability follows from the existence of tilings of the rectangles $[4 \times 5]$, $[5 \times 6]$ and of the rectangles $[5 \times 4]$, $[6 \times 5]$ that are transposed to them (see Figure 7.10). Indeed, then we can construct rectangles $[4 \times 5\,m]$, $[6 \times 5\,m]$, $[5 \times 2\,r]$, $r \ge 3$, which cover cases 2) and 3). An easy check shows that this also implies tileability in case 1).

Now let us prove the "only if" direction. First, we show that 10 must divide $a\,b$ in order for $[a \times b]$ to be tileable by $\mathbf{T} \subset \mathbf{T}_5$. Indeed, the area $f_0 = a \cdot b$ must be divisible by 5, which implies that $[a \times b]$ is tileable by either a horizontal or a vertical tile in $\mathbf{T}_5$. Now this implies that the 2-convexity invariant takes the value zero on a rectangle, while its value is ± 1 on each of the eight tiles in $\mathbf{T}$. Therefore $[a \times b]$ must contain an even number of tiles in order to be tileable by $\mathbf{T}$, which proves the claim.

Observe that since $10 | a \cdot b$, $a, b > 7$, covers by case 1) we are left with the cases when a or b is at most 7. It is easy to see that there are no tileable rectangle with $a = 2, 3$. An elaborate search of about a dozen beginnings shows that there are no

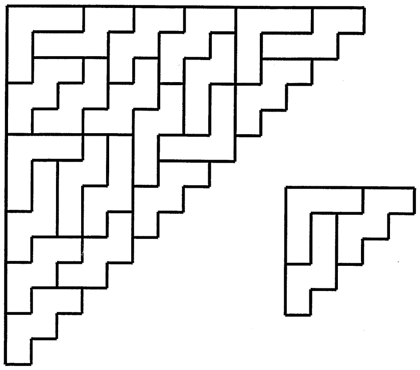


FIGURE 7.11.

tileable rectangle with $a = 7$. These cover all the untileable cases, and finishes the proof of the theorem. □

Theorem 7.6. *The triangular shape Δ_N can be tiled by tiles shown in Figure 7.9 if and only if* $N \equiv 0, 5, 14, 19 \pmod{20}$.

Proof. First we prove the tileability part. Observe that Δ_5 and Δ_{14} are both tileable (see Figure 7.11). Recall that by Theorem 7.5 the rectangles $[14 \times 5]$, $[15 \times 6]$ and $[20 \times N]$, $N \geq 5$, are all tileable by $\mathbf{T}$. Using the same construction as in the proof of Theorem 7.4 (see Figures 7.6 and 7.7), we get tilings on Δ_{19}, Δ_{20}. We also get a tiling of Δ_{20+N} from a tiling of Δ_N. This covers all the values $N \equiv 0, 5, 14, 19 \pmod{20}$ and proves the "if" part.

To prove the "only if" part, first recall that the area is divisible by 5 if and only if $N \equiv 0, 4 \pmod 5$, and in all these cases Δ_N is tileable by $\mathbf{T}_5$ (see the proof of Theorem 7.4). Compute the 2-convexity invariant $f_2(\Delta_N)$. We have (see Figures 7.7 and 7.8)

$$f_2(\Delta_4) = f_2(\Delta_5) = -1, \quad f_2(\Delta_{N+5}) = f_2(\Delta_N) - 1,$$

which gives us

$$f_2(\Delta_{5\,m+4}) = f_2(\Delta_{5\,m+5}) = -m.$$

Since f_2 takes only the values ± 1 on $\mathbf{T}$, this implies that in order to be tileable the number of tiles in $\Delta_{5\,m+4,5}$ must have the same parity as m. This immediately implies $N \equiv 0, 4, 15, 19 \pmod{20}$, and finishes the proof. □

8. Signed tilings and coloring arguments

Instead of ordinary tilings one can try using *signed tilings* (see [CL]), which are basically placements of tiles on a plane with weights +1 or −1 assigned to each of them. We say that they tile a region if the sum of the weights of the tiles is 1 for every square inside a region and 0 elsewhere.

Even with small sets of tiles it often happens that there are untileable regions which have signed tilings. For example, the region in Figure 1.3 has no ordinary domino tiling but has a signed domino tiling. Indeed, simply tile the rectangle $[3 \times 4]$ by dominoes and add two horizontal dominoes on the top and on the bottom with negative signs.

It turns out that signed tilings are easier to study because of their connection with coloring arguments. By analogy with tile invariants, a coloring map $f : \mathcal{R} \to G$ is *trivial* if $f(\Gamma) = 0$, for every $\gamma \in \mathcal{R}$, where 0 is an identity element in an abelian group G. Otherwise f is called *nontrivial.*

Theorem 8.1. *A region Γ has a signed tiling if and only if there is no abelian group G that has a nontrivial $\mathbf{T}$-coloring map $f : \mathcal{R} \to G$.*

Proof. Indeed, consider a group $\mathfrak{G}$ obtained as a free abelian group generated by elements $x_{i,j}$ that correspond to squares of a square grid. Let $\mathfrak{I} \subset \mathfrak{G}$ be a subgroup generated by sums $x_{i_1,j_1} + x_{i_2,j_2} + \dots$ that correspond to translations of tiles.

Observe that the $\mathbf{T}$-coloring maps correspond to the elements of the quotient group $\mathfrak{G}/\mathfrak{I}$. Therefore for every region Γ there exists a $\mathbf{T}$-coloring map $f : \mathcal{R} \to G$ such that $f(\Gamma) \neq 0$ unless the sum of squares of Γ lies in $\mathfrak{I}$. On the other hand, a region Γ has a signed tiling if and only if the sum of its squares lies in $\mathfrak{I}$. This proves the result. □

By analogy with the tile counting group we define a *coloring group* $\mathbb{O}(\mathbf{T}) = \mathfrak{G}/\mathfrak{I}$. It follows from the proof of the theorem that signed tilings and $\mathbf{T}$-coloring maps are basically dual to each other. However, it is convenient to separate them, since they give a different view of the subject.

It is important to note that whenever we have a $\mathbf{T}$-coloring map which proves that a certain region Γ is untileable, this also implies that Γ has no signed tiling. Therefore in the event when Γ is untileable but has a signed tiling (like the region in Figure 1.3 mentioned above), it also means that this fact cannot be proved by use of $\mathbf{T}$-coloring maps. Traditionally the coloring maps were a major instrument in proving untileability results (see [G]), so it is often desirable to check whether they can be used to prove any of our negative results. Below we will show that the results we obtained in the previous section in fact *cannot* be proved by use of coloring arguments.

Finally, let us point out where the difference between signed and ordinary tilings comes from. Instead of an abelian group G one can take a monoid M (commutative semigroup with an identity element). One can also define a semigroup morphism from all regions to M which is the identity on all tiles. It is not hard to prove a theorem similar to Theorem 8.1 which says that a region Γ has a tiling if and only if there exists no morphism which is not the identity on Γ (cf. [CL]). We skip the details.

Theorem 8.2. *Let $\mathbf{T}_n$ be a set of ribbon tiles, and let $\mathbb{O}(\mathbf{T}_n)$ be its coloring group. Then*

$$\mathbb{O}(\mathbf{T}_n) \simeq \mathbb{Z}^{n-1}.$$

Proof. Let Z be an abelian group generated by elements $z_0, z_1, \dots, z_{n-1}$. Define a coloring map

$$\zeta : \mathcal{R} \to Z/(z_0 + \dots + z_{n-1}),$$

which acts on generators as follows:

$$\zeta(x_{i,j}) = z_{(j-i \bmod n)}.$$

To prove that ζ is indeed a $\mathbf{T}_n$-coloring map, note that by definition ribbon tiles in $\mathbf{T}_n$ contain no two squares lying on the same diagonal. Therefore they must contain squares lying in n subsequent diagonals: $i, i+1, \dots, i+n-1$, where $i \in \mathbb{Z}$. Thus for all $\tau \in \mathbf{T}_n$

$$\zeta(\tau) = z_i + \cdots + z_n + z_1 + \cdots + z_{i-1} = 0$$

and ζ is indeed a $\mathbf{T}_n$-coloring map.

We can now prove that the coloring group for the set of ribbon tiles $\mathbf{T}_n$ is $\mathbb{Z}^{n-1}$. Let us prove that for any region Γ, $\zeta(\Gamma) = 0$, there exist a signed tiling of Γ. Then, by Theorem 8.1, this will imply that the coloring group is isomorphic to $Z/(z_0 + \cdots + z_{n-1}) \simeq \mathbb{Z}^{n-1}$, which proves the result.

Let d_1 be the difference of ribbon tiles **00** … **01** and **00** … **00** which have the same starting square. Observe that adding d_1 to a region adds a square $x_{i,j}$ and subtracts a square $x_{i+1,j+1}$ lying on the same diagonal. Let d_2 be the difference of two horizontal tiles **00** … **0**, one of them having the starting square $x_{i,j}$ and the other in $x_{i,j+1}$. By analogy with d_1, d_2 adds a square $x_{i,j}$ and subtracts a square $x_{i,j+n}$ lying on the same diagonal modulo n. Note that all the differences d_1 and d_2 are equivalent up to translation.

Now, let Γ be a region such that $\zeta(\Gamma) = 0$. We show that by adding enough differences d_1 and d_2 we can get an empty region. Indeed, by using d_2's we can move all the squares of Γ into the first n diagonals. Furthermore, by using d_2's we can move all the squares of Γ into the squares $x_{0,0}$, $x_{0,1}$, … , , $x_{0,n-1}$. Since $\zeta(\Gamma) = 0$, this means that we get the same weight m at each of these squares. Now subtract m copies of the horizontal tile. We get an empty region, which is exactly what we needed. This finishes the proof of the theorem. □

Note that if we use only d_1 without d_2 we get a sequence of numbers $\dots, m_{-1}$, m_0, m_1, $m_2, \dots$, which are numbers of squares in diagonals. This sequence gives rise to a coloring argument for the ordinary tilings that is more general than the $\mathbf{T}_2$-coloring map ζ. For example, it proves that the region in Figure 1.3 is untileable by dominoes. Indeed, the sequence is $1, 1, 2, 2, 2, 2, 1, 1$. The first two ones imply that there is a domino lying in the first two diagonals. Now, the two in the third place implies that there must be two dominoes that lie in the the third and fourth diagonal. Consequently, there must be two dominoes that lie in the the fifth and sixth diagonal. But this is impossible since the sixth number in the sequence is one. Therefore the region in Figure 1.3 is indeed untileable by dominoes.

When we restrict our set of tiles to a subset, all the coloring maps remain. However, a priori other coloring maps may appear. We show that in the cases considered in the previous section this does not happen.

Theorem 8.3. *For the set of tiles shown in Figure 7.1 the coloring group is $\mathbb{Z}^3$. For the sets of tiles shown in Figures 0.5, 0.6 and 7.9 the coloring group is $\mathbb{Z}^4$.*

Proof. All we need to show is that our sets of tiles generate differences d_1 and d_2. This would imply that their coloring group is the same as that of the ribbon tiles.

1) For the set of tiles in Figure 7.1, d_1 comes from subtracting **010** and **001** with the same starting square. The difference d_2 comes from subtracting **100** and **001** with starting squares $x_{i,j}$ and $x_{i-1,j}$ respectively.

2) For the set of tiles in Figure 0.5, d_1 comes from subtracting **0010** and **0001** with the same starting square. The difference d_2 comes from subtracting **0111** and **1110** with starting squares $x_{i,j}$ and $x_{i,j+1}$ respectively.

3) For the set of tiles in Figure 0.6, denote d_2 the difference between **0111** and **1110** with starting squares $x_{i,j}$ and $x_{i,j+1}$ respectively. It adds $x_{i,j}$ and subtracts

$x_{i-3,j+2}$. Analogously define a difference d_2' between **0001** and **1000** with starting squares $x_{i-1,j}$ and $x_{i,j}$ respectively. It subtracts $x_{i,j}$ and adds $x_{i-2,j+3}$. Now the difference d_1 comes from adding d_2 and d_2'.

4) For the set of tiles in Figure 7.9, d_1 comes from subtracting **0010** and **0011** with the same starting square. The difference d_2 comes from subtracting **0010** and **0100** with starting squares $x_{i,j}$ and $x_{i,j+1}$ respectively. □

Theorem 8.3 basically tells us that $\zeta(\Gamma) = 0$ is a criterion for a region Γ to have a signed tiling in each of those cases. Before we can conclude, we need the following technical result.

Lemma 8.4. 1) *Let* $n = 5$, $\Gamma = \Delta_N$. *Then* $\zeta(\Gamma) = e$ *if and only if* $N \equiv 0, 4 \pmod 5$.

2) *Let* $n = 5$, $\Gamma = [a \times b]$. *Then* $\zeta(\Gamma) = e$ *if and only if* $5 | a \cdot b$.

3) *Let* $n = 4$, $\Gamma = [a \times b]$. *Then* $\zeta(\Gamma) = e$ *if and only if* $4|a$ *or* $4|b$.

Proof. Parts 1) and 2) follow immediately from the area being divisible by 5 and existence of tilings by T_5 in all these cases (see the proof of Theorem 7.4).

Part 3) is analogous except for the case when $a, b \equiv 2 \pmod 4$. In this case a simple direct computation of zeta shows that $\zeta \neq e$. □

Corollary 8.5. 1) *Let* $n = 5$, *and let* $\mathbf{T}$ *be any of the sets of tiles shown in Figures 0.5, 0.6 and 7.9. Then* Δ_N *has a signed tiling by* $\mathbf{T}$ *if and only if* $N \equiv 0, 4 \pmod 5$.

2) *Let* $n = 5$, *and let* $\mathbf{T}$ *be any of the sets of tiles shown in Figures 0.5, 0.6 and 7.9. Then* $[a \times b]$ *has a signed tiling by* $\mathbf{T}$ *if and only if* $5 | a \cdot b$.

3) *Let* $n = 4$, *and let* $\mathbf{T}$ *be the set of tiles shown in Figure 7.1. Then* $[a \times b]$ *has a signed tiling by* $\mathbf{T}$ *if and only if* $4|a$ *or* $4|b$.

Proof. This follows immediately from Lemma 8.4 and Theorems 8.3 and 8.1. □

Corollary 8.6. *None of Theorems* 0.1 – 0.3 *can be proved by the coloring arguments.*

9. Extended coloring arguments and tile invariants

Recall the definition of a *coloring group*:

$$\mathbb{O}(\mathbf{T}) = \mathfrak{G}/\mathfrak{I},$$

where $\mathfrak{G}$ is a free abelian group generated by elements $x_{i,j}$ that correspond to squares of a square grid, and let $\mathfrak{I} \subset \mathfrak{G}$ be a subgroup generated by sums $x_{i_1,j_1} + x_{i_2,j_2} + \dots$ that correspond to translations of tiles.

By analogy, define an *extended coloring group* as follows:

$$\overline{\mathbb{O}}(\mathbf{T}) = \mathfrak{G}/\mathfrak{I}',$$

where $\mathfrak{I}' \subset \mathfrak{G}$ is a subgroup generated by relations

$$x_{i_1,j_1} + x_{i_2,j_2} + \dots = x_{p_1,q_1} + x_{p_2,q_2} + \dots,$$

where the tiles $\tau_1 = \coprod_k (i_k, j_k)$, $\tau_2 = \coprod_k (p_k, q_k)$ are translations of the same tile $\tau \in \mathbf{T}$. Of course, $\mathbb{O}(\mathbf{T}) \subset \overline{\mathbb{O}}(\mathbf{T})$.

By definition, every extended coloring argument for a set of tiles $\mathbf{T}$ corresponds to an element of the extended coloring group $\overline{\mathbb{O}}(\mathbf{T})$. In other words, there is a map $\nu : \overline{\mathbb{O}}(\mathbf{T}) \to \mathbb{G}(\mathbf{T})$. Since ν is homomorphic, the image $\mathbb{E}(\mathbf{T}) = \nu(\overline{\mathbb{O}}(\mathbf{T}))$ is a subgroup in the tile counting group $\mathbb{G}(\mathbf{T})$. Call this image $\mathbb{E}(\mathbf{T}) \subset \mathbb{G}(\mathbf{T})$ a *torsion group*.

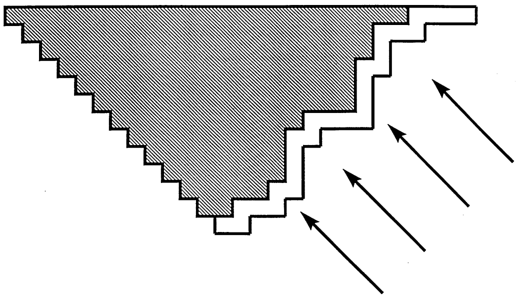


FIGURE 9.1.

Theorem 9.1. $\mathbb{E}(\mathbf{T}) \simeq \overline{\mathbb{O}}(\mathbf{T})/\mathbb{O}(\mathbf{T})$.

Proof. By definition, $\mathbb{O}(\mathbf{T})$ is the kernel of ν. This implies the result. □

Now we can show that in the case of ribbon tiles most of the invariants cannot be derived by the extended coloring arguments. We shall give a complete description of the torsion group $\mathbb{E}(\mathbf{T}_n)$ and compare it with the previously computed tile counting group $\mathbb{G}(\mathbf{T}_n) \supset \mathbb{E}(\mathbf{T}_n)$.

Definition 9.2. Let $\mathbf{T}_n$ be a set of ribbon tiles. A map $f_{\blacktriangledown} : \mathbf{T}_n \to \mathbb{Z}_n$ defined as

$$f_{\blacktriangledown}(\varepsilon_1, \dots, \varepsilon_{n-1}) = \sum_{i=1}^{n-1} \cdot\varepsilon_i \pmod n$$

is called the **shade invariant**.

An easy geometric interpretation of $f_{\blacktriangledown}$ is given in Figure 9.1. Imagine there is a wall behind our ribbon tile τ, and the light is coming from the southeast. Then $f_{\blacktriangledown}(\tau)$ is equal to the shaded area modulo n.

First, observe that $f_{\blacktriangledown}$ is a tile invariant. Indeed, if n is odd we have

$$f_{\blacktriangledown} = f_1 + 2\, f_2 + \dots + m\, f_m \pmod n,$$

where f_i is the i-convexity invariant and $n = 2\,m + 1$. Analogously, if n is even we have

$$f_{\blacktriangledown} = f_1 + 2\, f_2 + \dots + (m-1)\, f_{m-1} + f_*|^{\mathbb{Z}_n} \pmod n,$$

where $n = 2\,m$ and $g = f_*|^{\mathbb{Z}_n} \mod n$ is a parity invariant lifted to $\mathbb{Z}_{2\,m}$:

$$g(\varepsilon_1, \dots, \varepsilon_{n-1}) = m\varepsilon_m \pmod{2\,m}.$$

This gives $f_{\blacktriangledown} \in \mathbb{G}(\mathbf{T})$. Let us show that $f_{\blacktriangledown} \in \mathbb{E}(\mathbf{T})$, i.e. that the shade invariant can be obtained by the extended coloring argument. Indeed, consider a coloring map $f : \mathcal{R} \to \mathbb{Z}$, defined by $f(i,j) = i \mod n$. It is easy to see that $\nu(f) = f_{\blacktriangledown}$ and therefore $f_{\blacktriangledown} \in \mathbb{E}(\mathbf{T}_n)$. Analogously, the area invariant $f_0 = \nu(g) \in \mathbb{E}(\mathbf{T}_n)$, where $g : \mathcal{R} \to \mathbb{Z}$ is a coloring map defined by

$$f(i,j) = \begin{cases} 1\,, & i - j = 0 \mod n, \\ 0\,, & i - j \neq 0 \mod n. \end{cases}$$

We claim that except for f_0 and $f_{\blacktriangledown}$ and their linear combinations, no other nontrivial invariant can be obtained by the extended coloring argument.

Theorem 9.3. *Let $\mathbf{T}_n$ be a set of ribbon tiles, f_0 the area invariant and $f_\blacktriangledown$ the shade invariant. Then $\mathbb{E}(\mathbf{T}_n) \simeq \mathbb{Z} \times \mathbb{Z}_n$, and the maps f_0, $f_\blacktriangledown$ form an independent basis of invariants.*

Proof. We already showed that $f_0, f_\blacktriangledown \in \mathbb{E}(\mathbf{T}_n)$. Since they are independent, together they generate $\mathbb{Z} \times \mathbb{Z}_n$.

In the other direction, recall the computation of the coloring group $\mathbb{O}(\mathbf{T}_n) \simeq \mathbb{Z}^{n-1}$ given in Theorem 8.2. Let us compute $\overline{\mathbb{O}}(\mathbf{T}_n)$. By Theorem 9.1 this is all we need to find $\mathbb{E}(\mathbf{T})$.

Denote $\tau_0 = \mathbf{00} \ldots \mathbf{0}$, $\tau_1 = \mathbf{11} \ldots \mathbf{1}$ and $\tau_2 = \mathbf{10} \ldots \mathbf{0}$. Let us find all coloring maps $f : \mathcal{R} \to \mathfrak{G}$ (see the proof of Theorem 8.1), $f(i,j) = x_{i,j}$ such that the sums of squares in translations of τ_{0-2} are constant. Note here that a priori $\nu(f)$ does not have to be a tile invariant, since we do not check the relations for other ribbon tiles.

We claim that f is determined by values $x_{1,1}$, $x_{1,2}$, ... , $x_{1,n}$ and $x_{2,1}$. Indeed, translations of τ_0 and τ_1 give us $x_{i,j} = x_{i\pm n,j} = x_{i,j\pm n}$. Now, given $x_{i,j}, \ldots x_{i,j+n-2}$ and the value $f(\tau_2) = x_{1,1} + \cdots + x_{1,n-1} + x_{2,1}$, we get

$$x_{i+1,j} = f(\tau_2) - x_{i,j} - \cdots - x_{i,j+n-2}.$$

Therefore, given $x_{1,1}$, ... , $x_{1,n}$ and $x_{2,1}$, we first determine $x_{1,j}$, for all $j \in \mathbb{Z}$, then $x_{2,j}$, then $x_{3,j}$, etc. For the negative rows use $x_{i,j} = x_{i+n,j}$. This proves the claim.

Now, by taking a quotient $\overline{\mathbb{O}}(\mathbf{T}_n)/\mathbb{O}(\mathbf{T}_n)$ we can make all values $x_{1,1}$, $x_{1,2}$, ... , $x_{1,n-1}$ zero (see the proof of Theorem 8.2). Let $x_{1,n} = a$, $x_{2,1} = a + z$. The computations above give us $x_{i+1,j} = a + i\, z$ and $n\, z = 0$. Therefore $f = a \cdot f_0 + z \cdot f_\blacktriangledown$, and f_0, $f_\blacktriangledown$ generate the whole torsion group $\mathbb{E}(T_n)$. □

As a corollary, from Theorem 9.3 we immediately get Theorem 1.8. As we noted in the introduction, another way to prove Theorem 1.8 would be to find a tileable region $\Gamma \in \mathcal{R}_{\mathbf{T}_n} \setminus \mathcal{R}_{rc}$, $n \geq 3$, such that $f_\bullet$ is not constant on the set $\mathcal{S}(\Gamma, \mathbf{T}_n)$ of ribbon tilings. When $n = 3$ one such example is shown in Figure 9.2. The value of $f_\bullet$ is 1 on the first tiling and 0 on the second tiling. Note also that such a region Γ probably must have at least one hole inside (see Conjecture 10.1 in the next section).

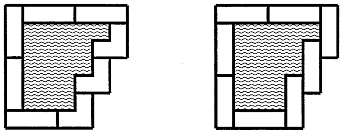


FIGURE 9.2.

10. CONCLUSION

Let us summarize the results in the paper and compare them with open questions.

The main result of the paper is a description of the tile counting group for a set of ribbon tiles. Note, however that we only considered row convex or column convex regions (which probably include all the interesting ones). However, Conway and Lagarias in [CL] were able to prove that for $n = 3$ the map f_1 is an invariant for all simply connected regions. Recently Muchnik and the author in [MP] used a similar technique to show that for $n = 4$ the maps f_1 and f_* are invariants for all simply connected regions. All the available evidence points to the following conjecture.*

Conjecture 10.1. *The i-convexity and parity maps are the group invariants for all simply connected regions.*

The major point of our proof is Theorem 1.6, which claims that there is a finite set of moves which can change any tiling of a given Young shape region to any other tiling. This result has been generalized by the author for any skew shape (see [Pa]). In fact we believe in the following conjecture.

Conjecture 10.2. *Let Γ be any simply connected region tileable by $\mathbf{T}_n$. Then the graph $\Theta_n(\Gamma)$ is connected.*

In other words, we claim that the moves defined in section 3 suffice. Of course, the results of section 4 imply that Conjecture 10.1 follows from Conjecture 10.2.

Let us move now to other sets of tiles. Unfortunately it is not always true that there exist a finite number of moves (or local replacement rules as they are also called). For example, let $\mathbf{T} \supset \mathbf{T}_3$ be the set of all trominoes (see [G] and Figure 10.1). There are infinitely many regions with exactly two tilings that are not local in any sense (see Figure 10.2). Even though there are no finite number of local moves, there still can be some invariants other than the area. Here is an example.

Let $\mathbf{T}$ be the set of four trominoes $\tau_1, \ldots, \tau_4$ in Figure 10.1. Let $\mathcal{B} = \mathcal{R}_{\mathbf{T}}$ be the set of all $\mathbf{T}$-tileable regions. Define maps $f_{1,2}, f_{2,3} : \mathbf{T} \to \mathbb{Z}_3$ as follows:

$$f_{1,2}(\tau_1) = f_{1,2}(\tau_2) = 1, \qquad f_{1,2}(\tau_3) = f_{1,2}(\tau_4) = 0,$$

$$f_{2,3}(\tau_2) = f_{2,3}(\tau_3) = 1, \qquad f_{2,3}(\tau_1) = f_{2,3}(\tau_4) = 0.$$

Theorem 10.3. *Let $\mathbf{T}$ and $\mathcal{B} = \mathcal{R}_{\mathbf{T}}$ be as above, and let f_0 be an area invariant. Then tile counting group*

$$\mathbb{G}(\mathbf{T}, \mathcal{B}) \simeq \mathbb{Z} \times \mathbb{Z}_3^2$$

and the maps f_0, $f_{1,2}$, $f_{2,3}$ form an independent basis of invariants.

Theorem 10.3 basically claims that there is one nontrivial tile invariant $f_{1,2}$, which can be stated as follows:

- For any convex region Γ the number of times *modulo 3* the tiles τ_1 and τ_2 occur in a tiling of Γ depends only on Γ.

Theorem 10.3 also claims that rotations of $f_{1,2}$ and the area invariant generate the whole tile invariant group.

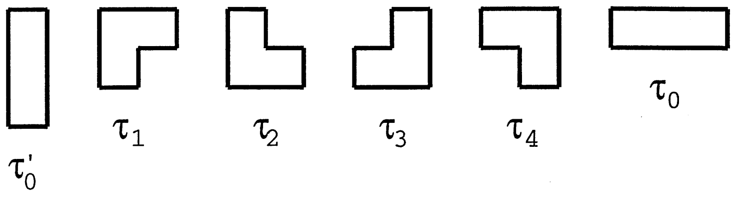


FIGURE 10.1.

[*] Conjecture 10.1 was recently established by C. Moore and the author (July 2000).

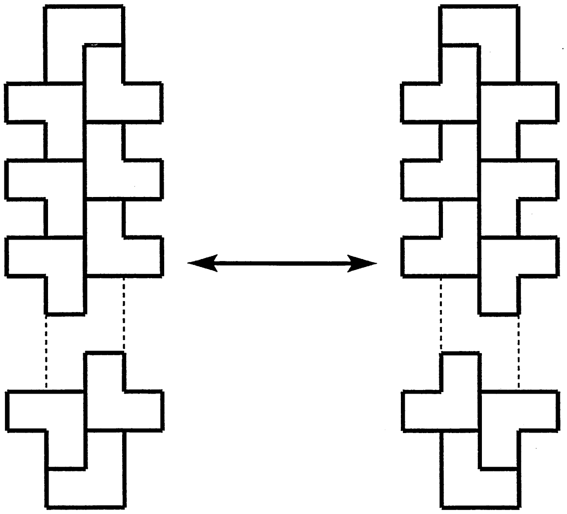


FIGURE 10.2.

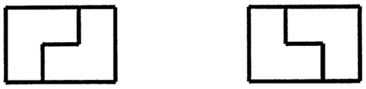


FIGURE 10.3.

Proof of Theorem 10.3. The idea is similar to the one used in section 6. By definition $\mathbb{G}(\mathbf{T};\mathcal{B}) = \mathbb{Z}^{\mathbf{T}}/I$, where I is the linear span or relations obtained from different tilings of the same region. Thus having enough relations that generate $I' \subset I$ will imply that $\mathbb{G}(\mathbf{T};\mathcal{B}) \subset \mathbb{Z}^{\mathbf{T}}/I'$. In this particular case two types of relations will suffice.

First, there is a relation obtained from the two tilings of $[2 \times 3]$ (see Figure 10.3):

$$\tau_1 + \tau_3 = \tau_2 + \tau_4.$$

Then there is another relation which comes from the two tilings in Figure 10.4:

$$4 \cdot \tau_4 + \tau_2 = 4 \cdot \tau_3 + \tau_1.$$

In combination with the first relation and rotations, this gives $3 \cdot \tau_1 = \cdots = 3 \cdot \tau_4$. Simple further computations show that the maps f_0, $f_{1,2}$, $f_{2,3}$ are independent and generate the whole tile counting group $\mathbb{G}(\mathbf{T}, \mathcal{R}_{\mathbf{T}})$. Therefore we have $\mathbb{G}(\mathbf{T}, \mathcal{R}_{\mathbf{T}}) \subset \mathbb{Z} \times \mathbb{Z}_3^2$.

Now it remains to prove that $f_{1,2}$ is indeed an invariant. This in turn would imply that $f_{1,2}$ is an invariant, and prove the theorem. This can be done by the tile extended coloring argument.

Let $g : \mathcal{R} \to \mathbb{Z}_3$ be a coloring map defined by $g(i,j) = j - i \pmod 3$. Observe that $g \in \overline{\mathbb{O}}(\mathbf{T})$. Compute the corresponding tile invariant $f = \nu(g) : \mathcal{R}_{\mathbf{T}} \to \mathbb{Z}_3$. We have

$$f(\tau_1) = f(\tau_3) = 0,\ f(\tau_2) = -1,\ f(\tau_2) = 1 \quad \text{mod } 3.$$

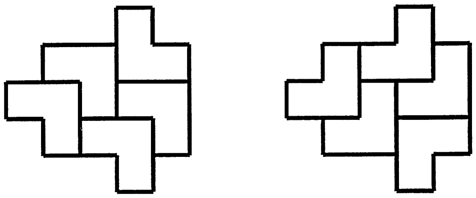


FIGURE 10.4.

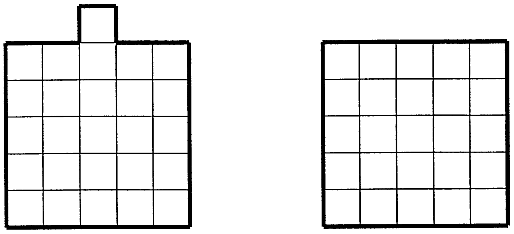


FIGURE 10.5.

By the symmetry, we also have another tile invariant $f' : \mathcal{R}_{\mathbf{T}} \to \mathbb{Z}_3$, given by

$$f'(\tau_2) = f'(\tau_4) = 0, f'(\tau_1) = -1,\ f'(\tau_3) = 1 \mod 3.$$

From this we get

$$f_{1,2} = f + f' - f_0 \mod 3.$$

Therefore both $f_{1,2}$ and $f_{2,3}$ are $\mathbf{T}$-invariants. This finishes the proof. □

Note that in the proof of Theorem 10.3 we used nothing but coloring arguments. Of course, with a smaller set of tileable regions, when the tile counting group becomes bigger, this would be impossible. That was the case with ribbon tiles. Indeed, in section 9 we showed that the height invariant cannot be extended to the set of all tileable regions, so the tile counting group $\mathbb{G}(\mathbf{T}_n, \mathcal{R}_{\mathbf{T}_n}) \subsetneq \mathbb{G}(\mathbf{T}_n, \mathcal{B}_{rc})$. Thus finding the tile counting group is probably hard in general unless all invariants follow from the extended coloring arguments.

Even in the case of all $\mathbf{T}$-tileable regions it is still possible to have invariants which do not follow from any extended coloring arguments. Indeed, consider the set $\mathbf{T}$ of two tiles shown in Figure 10.5. It is easy to see that *any* region either is untileable or has a unique tiling. On the other hand, extended coloring arguments can prove only the area invariant. Note that in this case every region has a signed tiling.

We believe that having $\mathbb{E}(\mathbf{T}) \simeq \mathbb{G}(\mathbf{T}, \mathcal{R}_{\mathbf{T}})$, is a rather rare event. However it might occur for certain nice sets of tiles such as finite sets of rectangles. Without making a precise conjecture, let us state the following problem.

Problem 10.4. Let $\mathbf{T}$ be a finite set of rectangles, and let $\mathcal{B}$ be the set of convex regions tileable by $\mathbf{T}$. Find tile invariant group $\mathbb{G}(\mathbf{T}; \mathcal{B})$ and the torsion group $\mathbb{E}(\mathbf{T})$.

For example, let $\mathbf{T}$ consist of two rectangles $[2 \times 1]$ and $[1 \times 3]$. It is not hard to see that $\mathbb{G}(\mathbf{T};\mathcal{B}) \simeq \mathbb{E}(\mathbf{T}) \simeq \mathbb{Z} \times \mathbb{Z}_5$. There is some recent literature relevant to the problem (see [Ke] for details).

In this paper we always considered tiles to be identical if they can be obtained by translation. Following some recent literature (see e.g. [Pr]), one can try to distinguish between ribbon tiles with different starting points. It seems to us that at least theoretically the whole analysis of the paper can be generalized for this case, though some computations may become complicated. We challenge the reader to find tileability applications of these generalizations.

Our proof of the main theorem was based on an ad hoc method which probably cannot be generalized in full for other sets of tiles. The heart of the proof is the rim hook bijection. There are shifted, tree, and skew analogs of this bijection (see [FS] and [Pa]), but they all can be reduced to the original bijection in one way or another.

Problem 10.5. Find a three-dimensional analog of the rim hook bijection.

Of course, there are infinitely many open questions and problems, but the reader is probably too tired already to be bothered by whatever is left.

Acknowledgments

I would like to express my deep gratitude to Jim Propp for interest in my work, for encouragement, for reading the first draft of this paper, and for useful suggestions. Conversations with Richard Stanley and his comments were very helpful. Nathan Lulov asked me the initial question which later became Theorem 1.8.

References

[BW] A. Bjorner, M. Wachs, *Generalized quotients in Coxeter groups.*, Trans. Amer Math. Soc. **308** (1988), 1–37. MR **89c:**05012

[BK] A. Berenstein, A. Kirillov, *Groups generated by involutions, Gelfand-Tsetlin patterns, and combinatorics of Young tableaux*, St. Petersburg Math. J. **7** (1996), 77–127. MR **96e:**05178

[CEP] H. Cohn, N. Elkies, J. Propp, *Local statistics for random domino tilings of the Aztec diamond*, Duke Math. J. **85** (1996), 117–166. MR **97k:**52026

[CL] J. H. Conway, J. C. Lagarias, *Tilings with polyominoes and combinatorial group theory*, J. Comb. Theory, Ser. A **53** (1990), 183–208. MR **91a:**05030

[EKLP] N. Elkies, G. Kuperberg, M. Larsen and J. Propp, *Alternating sign matrices and domino tilings.* I, II, J. Alg. Comb. **1** (1992), 111–132, 219–234. MR **94f:**52035, MR **94f:**52036

[FS] S. Fomin, D. Stanton, *Rim hook lattices*, St. Petersburg Math. J. **9** (1998), 1007–1016. MR **99c:**05202

[GJ] M. Garey, D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-completeness*, Freeman, San Francisco, CA, 1979. MR **80g:**68056

[G] S. Golomb, *Polyominoes*, Scribners, New York, 1965. MR **95k:**00006 (later ed.)

[JK] G. James, A. Kerber, *The Representation Theory of the Symmetric Group*, Addison-Wesley, Reading, MA, 1981. MR **83k:**20003

[Ka] P. W. Kastelyn, *The statistics of dimers on a lattice. I. The number of dimer arrangements on a quadratic lattice*, Physica **27** (1961), 1209–1225.

[Ke] R. Kenyon, *A note on tiling with integer-sided rectangles*, J. Combin. Theory, Ser. A **74** (1996), 321–332. MR **97c:**52045

[M] I. G. Macdonald, *Symmetric Functions and Hall Polynomials*, Oxford University Press, London, 1979. MR **84g:**05003

[MP] R. Muchnik, I. Pak, *On tilings by ribbon tetrominoes*, J. Combin. Theory, Ser. A **88** (1999), 199–193. CMP 2000:01

[Pa] I. Pak, *A generalization of the rim hook bijection for skew shapes*, preprint, 1997.

[Pr] J. Propp, *A pedestrian approach to a method of Conway, or, A tale of two cities*, Math. Mag. **70** (1997), 327–340. MR **98m:**52031

[R] G. de B. Robinson, *Representation Theory of the Symmetric Group*, Edinburgh University Press and Univ. of Toronto Press, 1961. MR **23:**A3182

[S] R. P. Stanley, *Enumerative Combinatorics.* Vol. 2, Cambridge Univ. Press, 1999. CMP 99:09

[SW] D. Stanton, D. White, *A Schensted algorithm for rim hook tableaux*, J. Comb. Theory, Ser. A **40** (1985), 211–247. MR **87c:**05014

[TF] H. N. V. Temperley, M. E. Fisher, *Dimer problem in statistical mechanics – An exact result*, Philos. Mag. **6** (1961), 1061–1063. MR **24:**B2436

[T] W. Thurston, *Conway's tiling group*, Amer. Math. Monthly **97** (1990), 757–773. MR91k:52028

Department of Mathematics, Yale University, New Haven, Connecticut 06520-8283
E-mail address: paki@math.yale.edu
Current address: Department of Mathematics, MIT, Cambridge, Massachusetts 02139
E-mail address: paki@math.mit.edu

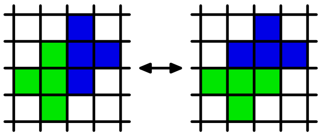
Доказательство гипотезы Пака о системе локальных ходов для замощений прямоугольников фигурами Т-тетромино

Константин Макарычев[*]
Юрий Макарычев[†]

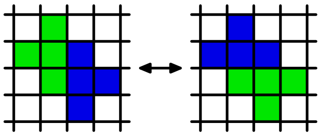
Аннотация

В этой работе мы докажем гипотезу Игоря Пака о замощениях односвязной области фигурами тетромино для случая замощений прямоугольника. Проблема была независимо решена в работе Михаила Корна и Игоря Пака [KP].

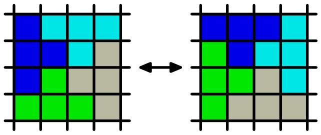
Гипотеза Пака [Pak]. Любое замощение односвязной области фигурами Т-тетромино можно перевести в любое другое последовательностью локальных ходов следующего вида:



M1A



M1B



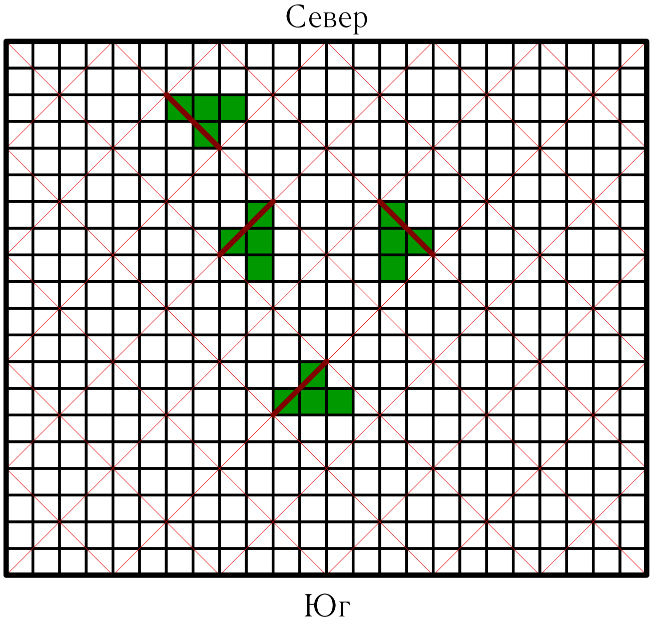
M2

[*] Принстонский Университет, kmakaryc@cs.princeton.edu
[†] Принстонский Университет, ymakaryc@cs.princeton.edu

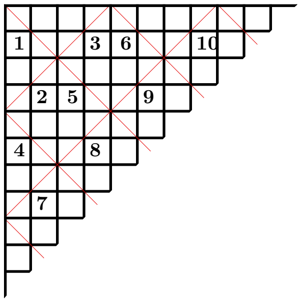
Доказательство состоит из нескольких частей. Сперва мы исследуем произвольные замощения прямоугольника тетромино, затем сведем задачу замощения (прямоугольника) к двойственной задаче (six-vertex model). И, наконец, воспользовавшись теоремой [Eloranta] о локальных ходах в six-vertex model, получим требуемый результат.

Для удобсва мы будем использовать следующие названия для направлений: *север* (верх), *юг* (низ), *запад* (лево), *восток* (право).

Рассмотрим произвольный прямоугольник $4n \times 4m$ и его произвольное замощение. Проведем в нем диагонали с шагом в 4 клетки (см. рисунок). Таким образом мы разделили прямоугольник на квадраты (наклоненные на 45°). Для краткости будем называть стороны этих квадратов — ребрами. Скажем, что ребро регулярно (в данном замощении), если оно пересекает ровно одно тетромино. На следующем рисунке нарисованы 4 регулярных ребра (красным цветом обозначены ребра, темно-красным — ребра, пересекающие тетромино). Назовем замощение регулярным, если все ребра в этом замощении регулярны. Мы хотим доказать, что все замощения регулярны.



Занумеруем ребра, начиная с северо-западного (верхнего левого) угла, двигаясь с юго-запада на северо-восток:



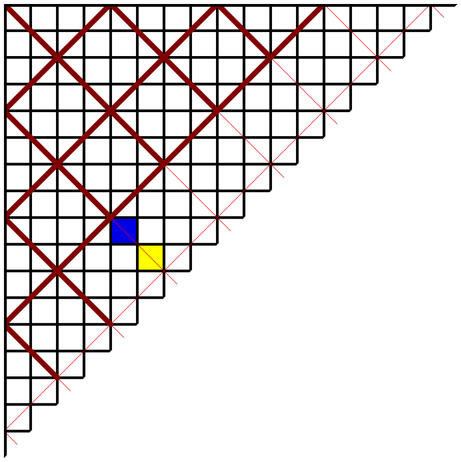
Лемма 1. Пусть P замощение, в котором первые n ребер регулярны, тогда $(n+1)$-ое ребро также регулярно.

Доказательство. Рассмотрим $(n+1)$-ое ребро в замощении P и покажем, что оно регулярно. Для этого предположим противное: $(n+1)$-ое ребро не регулярно, т.е. через две клетки ребра проходят различные тетромино.

Это ребро может быть «вертикальным» — идти с северо-запада на юго-восток или «горизонтальным» идти с юго-запада на северо восток. Мы рассмотрим эти два случая отдельно.

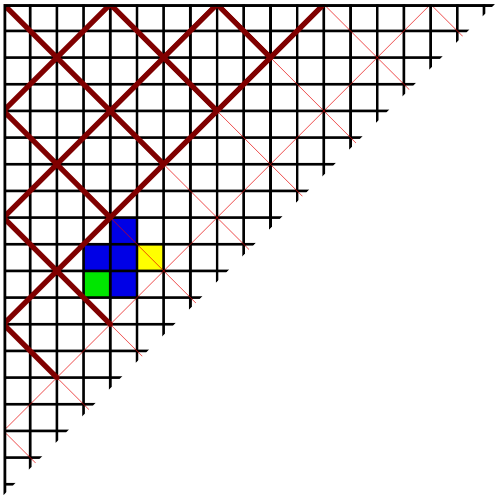
Вертикальное ребро

На следующих рисунках темно-красными жирными линиями выделены первые n ребер (про эти ребра нам известно, что они регулярные), $(n+1)$-ое ребро покрашено в синий и желтые цвета.



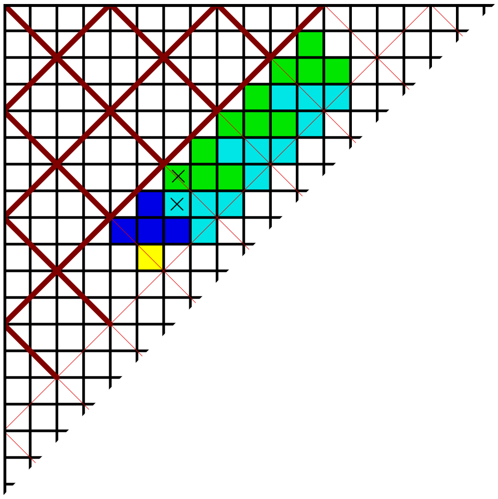
Рассмотрим тетромино, проходящее через синюю клетку (см. рисунок). Оно не пересекает регулярные ребра и не проходит через желтую клетку (вторую клетку ребра). Значит, вообще говоря, возможны следующие варианты:

вариант V1:



В этом случае зеленую клетку не может содержать ни одна фигура тетромино (важно, что на юго-западе от этой клетки регулярное ребро, которое содержит клетки одного и того же тетромино).

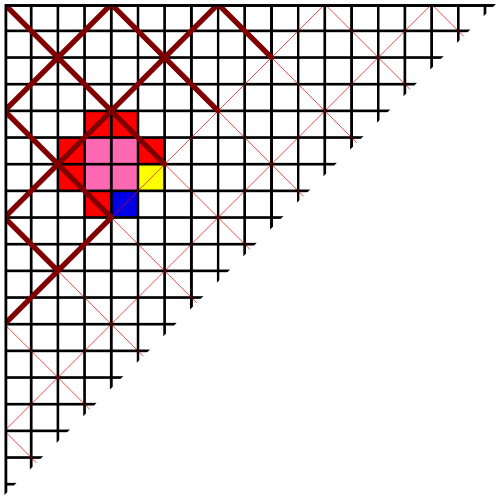
вариант V2:



Если тетромино расположено, как на рисунке выше, посмотрим какие тетромино проходят через клетки, помеченные крестиками. Единственная возможность (учитывая то, что регулярные (жирные) ребра тетромино может пересекать только по двум клеткам) — это тетромино, нарисованные голубым и зеленым цветом. Но зеленое тетромино стоит точно так же, как и синее поэтому, рассматривая тетромино на северо-востоке, мы получаем последовательность голубых и зеленых тетромино. Однако эта последовательность должна пересечь вертикальную или горизонтальную (восточную или северную соответственно) стенку прямоугольника. Значит такое замощение невозможно.

Горизонтальное ребро

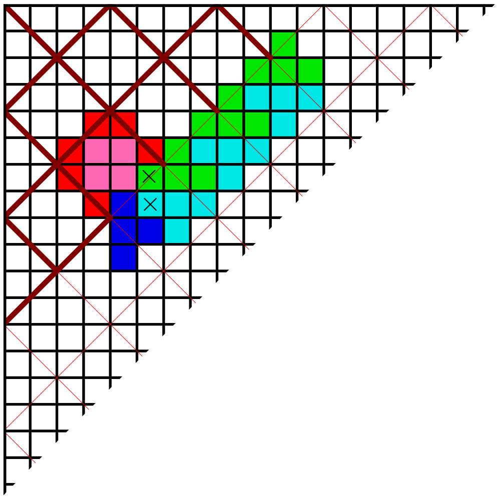
Заметим, что любое тетромино, пересекающее розовый квадрат (см. рисунок), пересекает его по двум клеткам. Действительно, если это тетромино пересекает одно из трех регулярных ребер красного квадрата, то его две клетки лежат на этом регулярном ребре, а остальные две должны лежать по одну сторону от ребра, т.е. обе лежат в красном квадрате. Т.к. все клетки красного квадрата покрыты каким-нибудь тетромино, а тетромино, проходящие через регулярные ребра, покрывают четное (0 или 2) число клеток квадрата, то и тетромино, проходящее через четвертое ребро, должно покрывать четное число клеток (хотя через четвертое (нерегулярное) ребро и проходят две тетромино, ясно, что лишь одно из них может пересекать розовый кварат).



Как и в случае с вертикальным ребром, рассмотрим тетромино проходящую через синюю клетку. Видно, что если оно пересекает розовый квадрат по двум клеткам, то оно содержит и желтую клетку, а значит это ребро регулярно.Таким образом это тетромино не может содержать клеток розового квадрата.

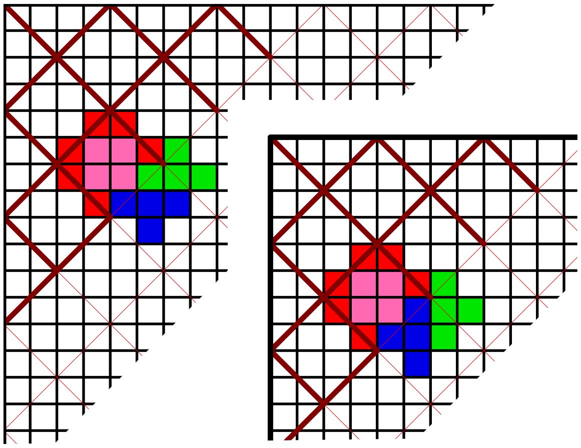
Рассмотрим оставшиеся варианты расположения этого тетромино (их два):

вариант H1:



Если синее тетромино расположено как на рисунке, то мы получаем случай аналогичный V2. Рассматривая тетромино проходящие через клетки, обозначенные крестиками, мы видим, что продолжение замощения на северо-восток определяется однозначно (как на рисунке). Однако такое продолжение должно пересечь северную или восточную границу.

вариант H2:



Если тетромино расположено, как на этом рисунке, то мы сделаем изображенную здесь замену (ход) и придем к случаю H1. (Т.е. мы изменим наше замощение и по пункту H1 увидим, что оно невозможно).

Таким образом, мы показали, что $(n+1)$-ое ребро регулярное.

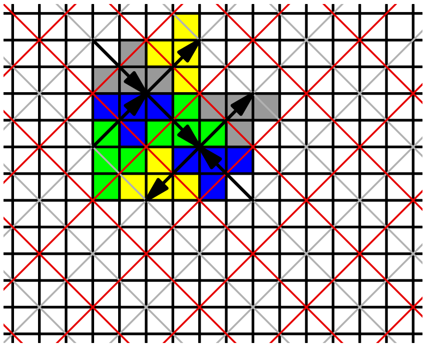
Ч.т.д.

Применяя лемму 1, по индукции, мы получаем:

Теорема 1. Любое замощение прямоугольника $4n \times 4m$ фигурами тетромино — регулярно.

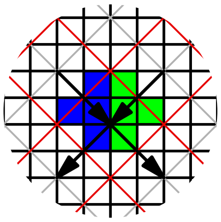
Замечание: из этой теоремы легко следует теорема [Walkup], что фигурами тетромино можно замостить только прямоугольники вида $4n \times 4m$. Действительно, предположим, что прямоугольник $k \times l$ (k или l не делится на 4) можно замостить тетромино. Повторяя это замощение несколько раз, мы можем замомстить пряугольник $4k \times 4l$. К этому замощению мы уже можем применить теорему 1. Т.е. это замощение регулярно. Теперь если мы рассмотрим изначальное замощение, оно будет так же регулярно, но из-за «граничного» эффекта это невозможно — в северо-восточном углу ребром будет отрезана область, которую невозможно замостить.

Теперь посмотрим как покрываются квадраты, ограниченные ребрами. Как мы уже знаем, каждому ребру квадрата соответствует одно тетромино. Оно может лежать в одном из двух смежных квадратов. Мы будем говорить, что оно направлено из квадрата, не содержащего это тетромино, в квадрат, его содержащий. Таким образом каждому замощению соответствует расстановка стрелок на ребрах «двойственной» решетки (она на рисунке нарисована серым цветом) с вершинами в центрах квадратов. Причем в каждую вершину направлено ровно две стрелки.

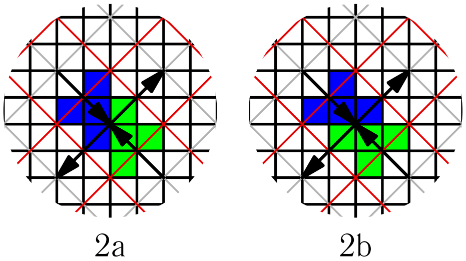


Наоборот, легко видеть, что каждой расстановке стрелок (при которой в каждую вершину двойственной решетки смотрят ровно две стрелки) соответствует какое-то замощение тетромино. Это достаточно проверить локально: нужно показать, что любой комбинации входящих стрелок в центр квадрата соответствует какое-то замощение самого квадрата. Причем всего различных вариантов вхождения стрелок два (с точностью до вращения):

1. Соседние стрелки смотрят во внутрь:



2. Диаметрально противоположенные стрелки смотрят во внутрь. Этому варианту соответствуют два замощения (которые переводятся друг в друга преобразованиями M1A и M1B):

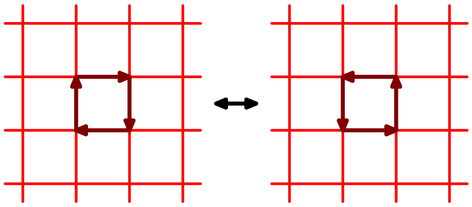


Чтобы задача растановки стрелок стала эквивалентной задаче замощения фигурами тетромино, отождествим замощения отличающиеся только на преобразования M1A и M1B. Таким образом образом мы свели задачу замощения к следующей задаче:

Задача «Six-vertex model». Дана решетка, на ее ребрах нужно расставить стрелки так, чтобы в каждую вершину входило ровно две.

Докажем следующую теорему.

Теорема 2 [Eloranta]. Пусть дана решетка в односвязной области. Тогда любую расстановку стрелок можно перевести в любую другую с помощью последовательности локальных ходов, обращающих стрелки цикла вокруг одной клетки.

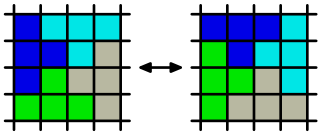


Доказательство. Введем функцию высоты $h(x)$ на клетках решетки. При переходе от одной клетки к другой h будет увеличиваться на 1, если мы пересекаем стрелку идущую слева направо (относительно хода нашего движения), и уменьшаться на 1 в противном случае. Эта функция определена корректно, т.к. если мы пройдем по замкнутому циклу число стрелок, выходящих из цикла, будет равно числу стрелок, входящих в цикл (это следует из того, что в каждую вершину входит ровно две стрелки и выходит столько же, значит верен закон сохранения: в каждый цикл «втекает» столько же стрелок сколько и «вытекает»). На границе положим $h(x)$ равной 0.

Рассмотрим произвольную расстановку стрелок. Приведем ее локальными ходами к одному из локальных минимумов (одна расстановка стрелок меньше или равна другой, если функция высоты первой меньше или равна функции высоты второй в каждой клетке). Покажем, что функция высоты полученой раскраски не содержит локальных максимумов. Действительно, если клетка клетка s — локальный максимум, то ее должен был бы обходить цикл (т.к. во всех направлениях $h(x)$ убывает), обратив направление этого цикла мы уменьшим значение $h(s)$ на 2 и не изменим значения в других точках. Итак, на границе $h(x)$ равна 0, а внутри области $h(x)$ не имеет локальных максимумов, значит внутри $h(x)$ строго отрицательна. Отсюда, следует, что на границе все стрелки обращены по часовой стрелке. Аналогично, внутри стрелки также образуют концентрические циклы. Таким оброзом мы показали, что все расстановки стрелок, являющиеся локальными минимумами, одинаковые.

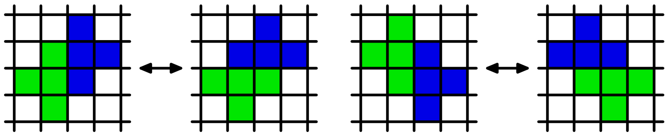
Ч.т.д.

Теперь перейдем от двойственной модели к модели замощения фигурами тетромино. Легко видеть, что обращению цикла соответствует следующий локальный ход:



M2

Таким образом этот ход и два хода, по которым мы отождествили замощения:



M1A M1B

образуют систему локальных ходов для замощения прямоугольника фигурами тетромино.

Итак, гипотеза Пака доказана в частном случае для прямоугольников.

Список литературы

[Eloranta] Kari Eloranta, Diamond Ice, Jour. of Stat. Phys., 1999

[KP] Michael Korn, Igor Pak, Tilings of rectangles with T-tetrominoes, http://www-math.mit.edu/~pak/ttet11.pdf

[Pak] Igor Pak, Tile Invariants: New Horizons, 2001

[Walkup] D. W. Walkup, Covering a rectangle with T-tetrominoes, Amer. Math. Monthly 72 (1965), 886-988

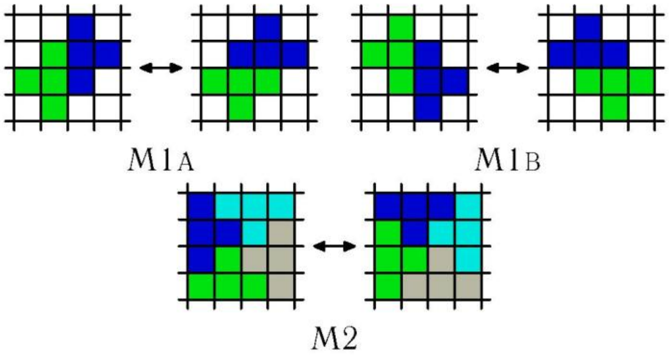
Доказательство гипотезы Пака о системе локальных ходов для замощений прямоугольников фигурами T-тетромино

Константин Макарычев*
Юрий Макарычев†

Аннотация

В этой работе мы докажем гипотезу Игоря Пака о замощениях односвязной области фигурами тетромино для случая замощений прямоугольника. Проблема была независимо решена в работе Михаила Корна и Игоря Пака [KP].

Гипотеза Пака [Pak]. Любое замощение односвязной области фигурами T-тетромино можно перевести в любое другое последовательностью локальных ходов следующего вида:

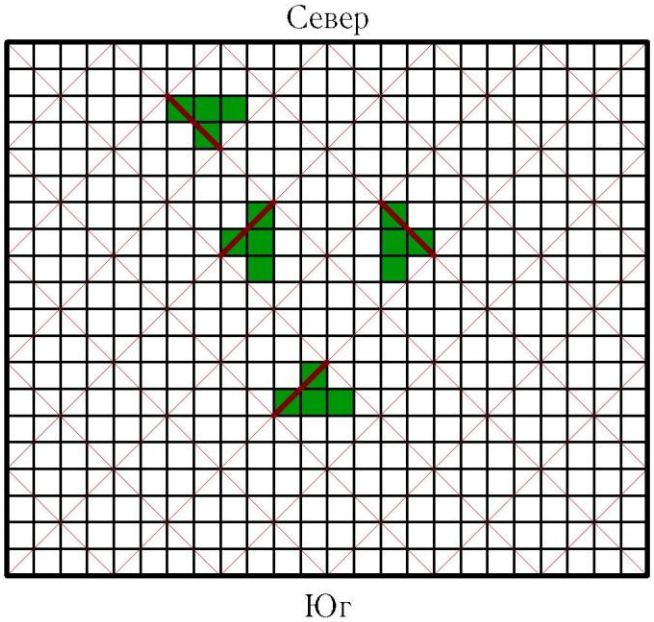


*Принстонский Университет, kmakaryc@cs.princeton.edu
†Принстонский Университет, ymakaryc@cs.princeton.edu

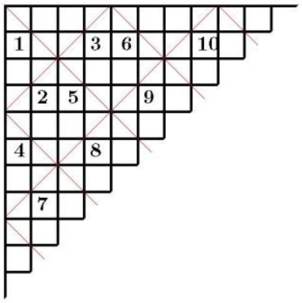
Доказательство состоит из нескольких частей. Сперва мы исследуем произвольные замощения прямоугольника тетромино, затем сведем задачу замощения (прямоугольника) к двойственной задаче (six-vertex model). И, наконец, воспользовавшись теоремой [Eloranta] о локальных ходах в six-vertex model, получим требуемый результат.

Для удобсва мы будем использовать следующие названия для направлений: *север* (верх), *юг* (низ), *запад* (лево), *восток* (право).

Рассмотрим произвольный прямоугольник $4n \times 4m$ и его произвольное замощение. Проведем в нем диагонали с шагом в 4 клетки (см. рисунок). Таким образом мы разделили прямоугольник на квадраты (наклоненные на 45°). Для краткости будем называть стороны этих квадратов — ребрами. Скажем, что ребро регулярно (в данном замощении), если оно пересекает ровно одно тетромино. На следующем рисунке нарисованы 4 регулярных ребра (красным цветом обозначены ребра, темно-красным — ребра, пересекающие тетромино). Назовем замощение регулярным, если все ребра в этом замощении регулярны. Мы хотим доказать, что все замощения регулярны.



Занумеруем ребра, начиная с северо-западного (верхнего левого) угла, двигаясь с юго-запада на северо-восток:



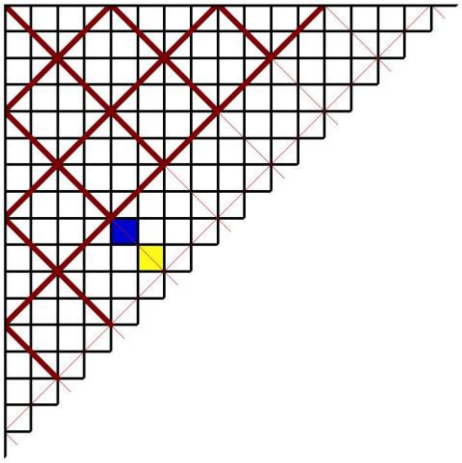
Лемма 1. Пусть P замощение, в котором первые n ребер регулярны, тогда $(n+1)$-ое ребро также регулярно.

Доказательство. Рассмотрим $(n+1)$-ое ребро в замощении P и покажем, что оно регулярно. Для этого предположим противное: $(n+1)$-ое ребро не регулярно, т.е. через две клетки ребра проходят различные тетромино.

Это ребро может быть «вертикальным» — идти с северо-запада на юго-восток или «горизонтальным» идти с юго-запада на северо восток. Мы рассмотрим эти два случая отдельно.

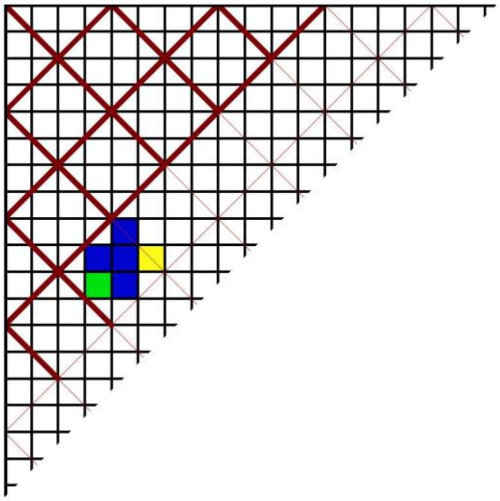
Вертикальное ребро

На следующих рисунках темно-красными жирными линиями выделены первые n ребер (про эти ребра нам известно, что они регулярные), $(n+1)$-ое ребро покрашено в синий и желтые цвета.



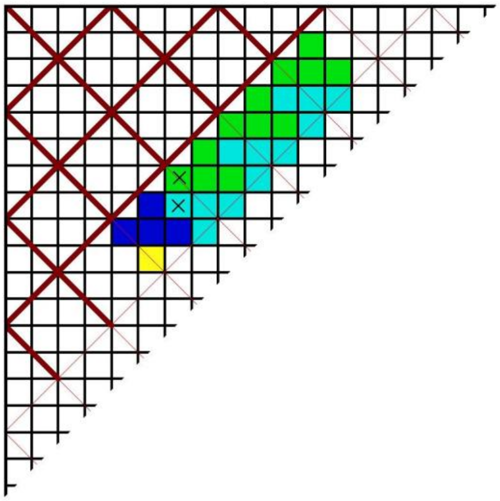
Рассмотрим тетромино, проходящее через синюю клетку (см. рисунок). Оно не пересекает регулярные ребра и не проходит через желтую клетку (вторую клетку ребра). Значит, вообще говоря, возможны следующие варианты:

ВАРИАНТ V1:



В этом случае зеленую клетку не может содержать ни одна фигура тетромино (важно, что на юго-западе от этой клетки регулярное ребро, которое содержит клетки одного и того же тетромино).

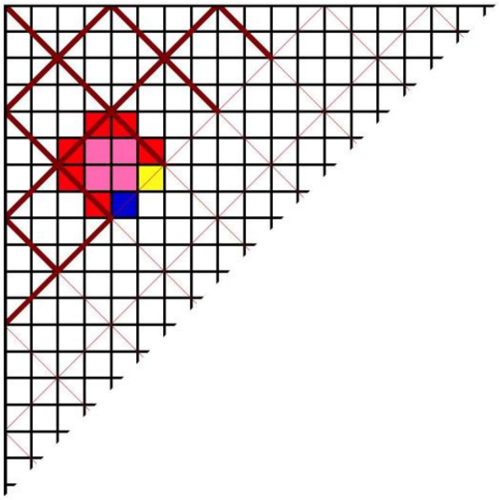
ВАРИАНТ V2:



Если тетромино расположено, как на рисунке выше, посмотрим какие тетромино проходят через клетки, помеченные крестиками. Единственная возможность (учитывая то, что регулярные (жирные) ребра тетромино может пересекать только по двум клеткам) — это тетромино, нарисованные голубым и зеленым цветом. Но зеленое тетромино стоит точно так же, как и синее поэтому, рассматривая тетромино на северо-востоке, мы получаем последовательность голубых и зеленых тетромино. Однако эта последовательность должна пересечь вертикальную или горизонтальную (восточную или северную соответственно) стенку прямоугольника. Значит такое замощение невозможно.

Горизонтальное ребро

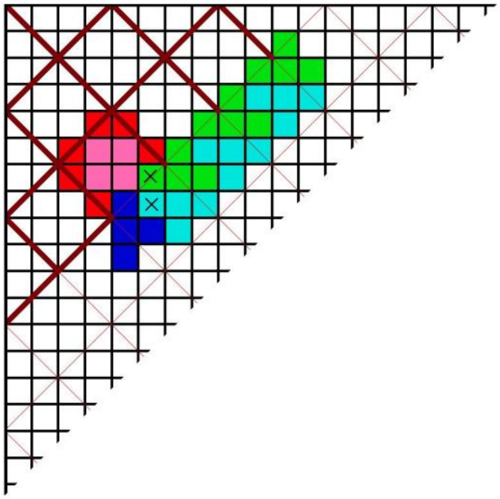
Заметим, что любое тетромино, пересекающее розовый квадрат (см. рисунок), пересекает его по двум клеткам. Действительно, если это тетромино пересекает одно из трех регулярных ребер красного квадрата, то его две клетки лежат на этом регулярном ребре, а остальные две должны лежать по одну сторону от ребра, т.е. обе лежат в красном квадрате. Т.к. все клетки красного квадрата покрыты каким-нибудь тетромино, а тетромино, проходящие через регулярные ребра, покрывают четное (0 или 2) число клеток квадрата, то и тетромино, проходящее через четвертое ребро, должно покрывать четное число клеток (хотя через четвертое (нерегулярное) ребро и проходят две тетромино, ясно, что лишь одно из них может пересекать розовый кварат).



Как и в случае с вертикальным ребром, рассмотрим тетромино проходящую через синюю клетку. Видно, что если оно пересекает розовый квадрат по двум клеткам, то оно содержит и желтую клетку, а значит это ребро регулярно.Таким образом это тетромино не может содержать клеток розового квадрата.

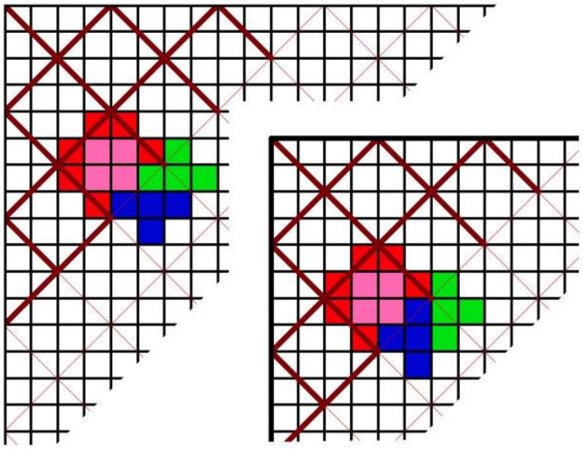
Рассмотрим оставшиеся варианты расположения этого тетромино (их два):

ВАРИАНТ H1:



Если синее тетромино расположено как на рисунке, то мы получаем случай аналогичный V2. Рассматривая тетромино проходящие через клетки, обозначенные крестиками, мы видим, что продолжение замощения на северо-восток определяется однозначно (как на рисунке). Однако такое продолжение должно пересечь северную или восточную границу.

ВАРИАНТ H2:



Если тетромино расположено, как на этом рисунке, то мы сделаем изображенную здесь замену (ход) и прийдем к случаю H1. (Т.е. мы изменим наше замощение и по пункту H1 увидим, что оно невозможно).

Таким образом, мы показали, что $(n+1)$-ое ребро регулярное.

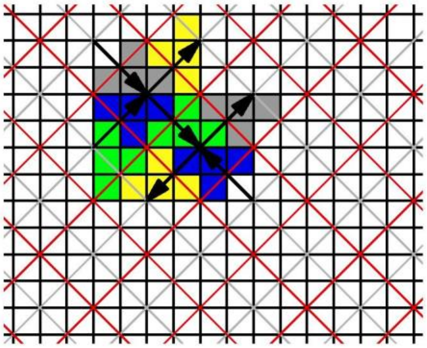
Ч.т.д.

Применяя лемму 1, по индукции, мы получаем:

Теорема 1. *Любое замощение прямоугольника $4n \times 4m$ фигурами тетромино — регулярно.*

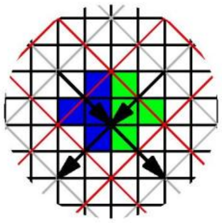
Замечание: из этой теоремы легко следует теорема [Walkup], что фигурами тетромино можно замостить только прямоугольники вида $4n \times 4m$. Действительно, предположим, что прямоугольник $k \times l$ (k или l не делится на 4) можно замостить тетромино. Повторяя это замощение несколько раз, мы можем замомстить пряугольник $4k \times 4l$. К этому замощению мы уже можем применить теорему 1. Т.е. это замощение регулярно. Теперь если мы рассмотрим изначальное замощение, оно будет так же регулярно, но из-за «граничного» эффекта это невозможно — в северо-восточном углу ребром будет отрезана область, которую невозможно замостить.

Теперь посмотрим как покрываются квадраты, ограниченные ребрами. Как мы уже знаем, каждому ребру квадрата соответствует одно тетромино. Оно может лежать в одном из двух смежных квадратов. Мы будем говорить, что оно направлено из квадрата, не содержащего это тетромино, в квадрат, его содержащий. Таким образом каждому замощению соответствует расстановка стрелок на ребрах «двойственной» решетки (она на рисунке нарисована серым цветом) с вершинами в центрах квадратов. Причем в каждую вершину направлено ровно две стрелки.

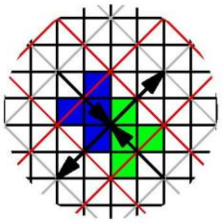


Наоборот, легко видеть, что каждой расстановке стрелок (при которой в каждую вершину двойственной решетки смотрят ровно две стрелки) соответствует какое-то замощение тетромино. Это достаточно проверить локально: нужно показать, что любой комбинации входящих стрелок в центр квадрата соответствует какое-то замощение самого квадрата. Причем всего различных вариантов вхождения стрелок два (с точностью до вращения):

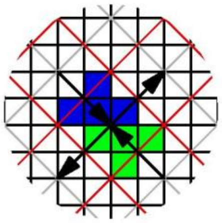
1. Соседние стрелки смотрят во внутрь:



2. Диаметрально противоположенные стрелки смотрят во внутрь. Этому варианту соответствуют два замощения (которые переводятся друг в друга преобразованиями M1A и M1B):



2a



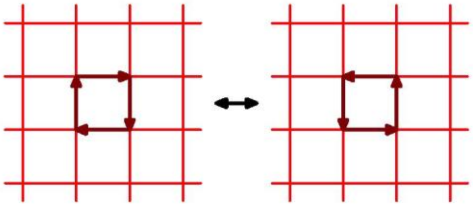
2b

Чтобы задача растановки стрелок стала эквивалентной задаче замощения фигурами тетромино, отождествим замощения отличающиеся только на преобразования M1A и M1B. Таким образом образом мы свели задачу замощения к следующей задаче:

Задача «Six-vertex model». Дана решетка, на ее ребрах нужно раставить стрелки так, чтобы в каждую вершину входило ровно две.

Докажем следующую теорему.

Теорема 2 [Eloranta]. Пусть дана решетка в односвязной области. Тогда любую расстановку стрелок можно перевести в любую другую с помощью последовательности локальных ходов, обращающих стрелки цикла вокруг одной клетки.

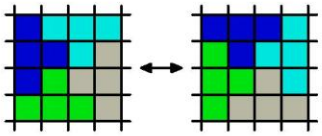


Доказательство. Введем функцию высоты $h(x)$ на клетках решетки. При переходе от одной клетки к другой h будет увеличиваться на 1, если мы пересекаем стрелку идущую слева направо (относительно хода нашего движения), и уменьшаться на 1 в противном случае. Эта функция определена корректно, т.к. если мы пройдем по замкнутому циклу число стрелок, выходящих из цикла, будет равно числу стрелок, входящих в цикл (это следует из того, что в каждую вершину входит ровно две стрелки и выходит столько же, значит верен закон сохранения: в каждый цикл «втекает» столько же стрелок сколько и «вытекает»). На границе положим $h(x)$ равной 0.

Рассмотрим произвольную расстановку стрелок. Приведем ее локальными ходами к одному из локальных минимумов (одна расстановка стрелок меньше или равна другой, если функция высоты первой меньше или равна функции высоты второй в каждой клетке). Покажем, что функция высоты полученой раскраски не содержит локальных максимумов. Действительно, если клетка клетка s — локальный максимум, то ее должен был бы обходить цикл (т.к. во всех направлениях $h(x)$ убывает), обратив направление этого цикла мы уменьшим значение $h(s)$ на 2 и не изменим значения в других точках. Итак, на границе $h(x)$ равна 0, а внутри области $h(x)$ не имеет локальных максимумов, значит внутри $h(x)$ строго отрицательна. Отсюда, следует, что на границе все стрелки обращены по часовой стрелке. Аналогично, внутри стрелки также образуют концентрические циклы. Таким оброзом мы показали, что все расстановки стрелок, являющиеся локальными минимумами, одинаковые.

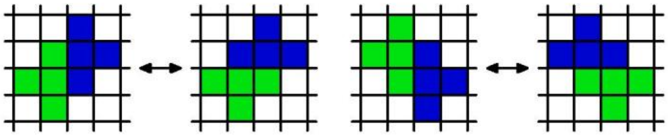
Ч.т.д.

Теперь перейдем от двойственной модели к модели замощения фигурами тетромино. Легко видеть, что обращению цикла соответствует следующий локальный ход:



M2

Таким образом этот ход и два хода, по которым мы отождествили замощения:



M1A M1B

образуют систему локальных ходов для замощения прямоугольника фигурами тетромино.

Итак, гипотеза Пака доказана в частном случае для прямоугольников.

Список литературы

[Eloranta] Kari Eloranta, Diamond Ice, Jour. of Stat. Phys., 1999

[KP] Michael Korn, Igor Pak, Tilings of rectangles with T-tetrominoes, http://www-math.mit.edu/~pak/ttet11.pdf

[Pak] Igor Pak, Tile Invariants: New Horizons, 2001

[Walkup] D. W. Walkup, Covering a rectangle with T-tetrominoes, Amer. Math. Monthly 72 (1965), 886-988

JOURNAL OF COMBINATORIAL THEORY, Series A 53, 183–208 (1990)

Tiling with Polyominoes and Combinatorial Group Theory

J. H. CONWAY

Princeton University, Princeton, New Jersey

AND

J. C. LAGARIAS

AT&T Bell Laboratories, Murray Hill, New Jersey

Communicated by Andrew Odlyzko

Received May 3, 1988

When can a given finite region consisting of cells in a regular lattice (triangular, square, or hexagonal) in $\mathbb{R}^2$ be perfectly tiled by tiles drawn from a finite set of tile shapes? This paper gives necessary conditions for the existence of such tilings using *boundary invariants*, which are combinatorial group-theoretic invariants associated to the boundaries of the tile shapes and the regions to be tiled. Boundary invariants are used to solve problems concerning the tiling of triangular-shaped regions of hexagons in the hexagonal lattice with certain tiles consisting of three hexagons. Boundary invariants give stronger conditions for nonexistence of tilings than those obtainable by weighting or coloring arguments. This is shown by considering whether or not a region has a *signed tiling*, which is a placement of tiles assigned weights 1 or -1, such that all cells in the region are covered with total weight 1 and all cells outside with total weight 0. Any coloring (or weighting) argument that proves nonexistence of a tiling of a region also proves nonexistence of any signed tiling of the region as well. A partial converse holds: if a simply connected region has no signed tiling by simply connected tiles, then there is a generalized coloring argument proving that no signed tiling exists. There exist regions possessing a signed tiling which can be shown to have no perfect tiling using boundary invariants. © 1990 Academic Press, Inc.

1. INTRODUCTION

Packing, covering, and tiling problems are among the most basic combinatorial problems. Here we consider problems concerning the possibility or impossibility of tiling finite regions of a regular lattice tiling of $\mathbb{R}^2$ by translations of a finite set of (lattice) tiles.

0097-3165/90 $3.00

There are three regular lattice tilings of $\mathbb{R}^2$, which are the triangular lattice, square lattice, and hexagonal lattice, pictured in Fig. 1.1. Each of these tilings divides $\mathbb{R}^2$ into *cells*, and any cell can be obtained from any other cell by a translation. A *lattice figure* or *region*, is a finite union of (closed) cells that is connected. Lattice figures for the three types of lattices are called *polyiamonds*, *polyominoes*, and *polyhexes*, respectively. Two lattice figures are *equivalent* if one can be obtained from the other by a translation. They are *congruent* if one can be obtained from the other by a Euclidean motion, which includes rotations and reflections. A (lattice) *tile* is a simply connected lattice figure. A set Σ of lattice figures *tiles* a region R if R can be covered with translates of figures in Σ such that each cell in R is covered by exactly one lattice figure.

Tiling problems on lattices are in general computationally difficult problems. Consider the following two problems:

PLANE TILING PROBLEM

Instance. A finite set Σ of tiles.

Question. Does Σ tile the whole lattice?

FINITE TILING PROBLEM

Instance. A region R and a finite set Σ of tiles.

Question. Does Σ tile R?

The Plane Tiling Problem is undecidable, as can be shown by a suitable encoding of the undecidable Wang Tiling Problem (also called the Domino Problem, see [2; 24; 14, Chap. 11]), in which each colored edge of a colored square (Wang tile) is replaced with an appropriately serrated edge following the lattice edges. The Finite Tiling Problem is clearly decidable by exhaustive enumeration and is in the computational complexity class NP because if a tiling exists it can be nondeterministically "guessed." However, it is NP-complete, as may be shown using an encoding of Square Tiling (see [8, p. 257]) again obtained using tiles with serrated edges. Consequently it is unlikely that there exists a polynomial time algorithm to solve

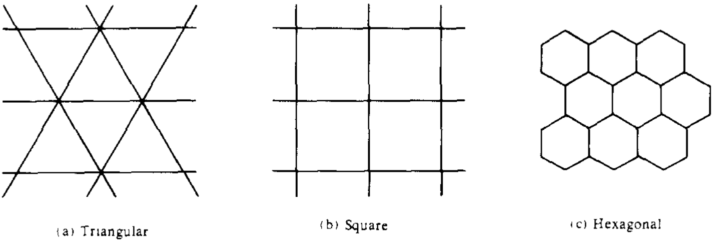


FIG. 1.1. Regular lattice tilings of $\mathbb{R}^2$.

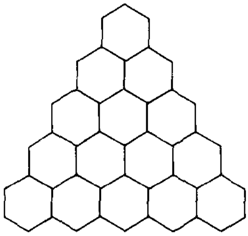


FIG. 1.2. Triangular region T_5.

the Finite Tiling Problem. Special methods do exist which can often be used to prove nonexistence of tilings of regions with a single tile. These include coloring and weighting arguments among others [3–6; 8–13; 16–20; 26].

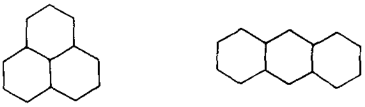
In view of the difficulty of the general Finite Tiling Problem, it is not too surprising that even apparently simple-looking tiling problems can prove difficult to solve. This paper arose from considering the following sets of tiling problems on the hexagonal lattice. Let T_N denote the triangular array of cells in the hexagonal lattice having $\binom{N+1}{2}$ cells pictured in Fig. 1.2. The *triangle tiling by triangles problem* is to decide: for which values of N can T_N be tiled by congruent copies of the triangular tile T_2 pictured in Fig. 1.3a? The *triangle tiling by lines problem* is to decide: for which values of N can T_N be tiled by congruent copies of the three-in-line tile L_3 pictured in Fig. 1.3b? In these problems one permits tiles to be rotated or reflected. In terms of equivalence classes of tiles the first problem above allows tiling by two inequivalent tiles and the second problem allows tiling by three inequivalent tiles, as pictured in Fig. 1.4.

These two tiling problems have the following answers.

THEOREM 1.1. *The triangular region T_N in the hexagonal lattice can be tiled by congruent copies of the triangular tile T_2 if and only if*

$$N \equiv 0,\ 2,\ 9,\ \textit{or}\ 11 \pmod{12}.$$

THEOREM 1.2. *It is impossible to tile the triangular region T_N in the hexagonal lattice with congruent copies of the three-in-line tile L_3.*



(a) Triangular tile T_2 (b) Three-in-line tile L_3

FIG. 1.3. Tiles for triangle tiling problems.

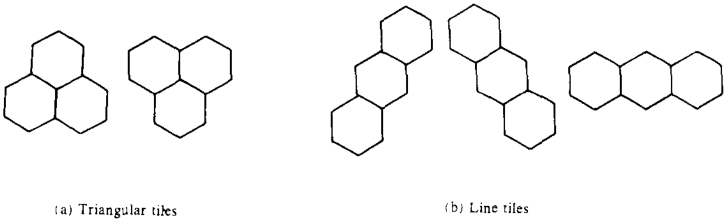


FIG. 1.4. Tile sets of translation-inequivalent tiles.

To solve these problems, we introduce combinatorial group-theoretic invariants associated to the boundaries of the tiles and the region to be tiled; we call these *boundary invariants*. Section 2 defines these invariants and shows that for a simply connected region R a necessary condition for a tiling by tiles in a set Σ to exist is that the combinatorial boundary of the region R be contained in a group $T(\Sigma)$ generated by the boundaries of the tiles in Σ (Theorem 2.1). This group-theoretic criterion seems in general no easier to verify than to solve the original problem. It can, however, be successfully applied to the case of the two triangle tiling problems, using group-theoretic properties specific to these problems. This is done in Section 3.

These solutions to the two triangle tiling problems are somewhat complicated, and it is reasonable to ask if simpler solutions exist. We investigate the relation between boundary invariants and other known necessary conditions for a tiling to exist. A region R has a *signed tiling* using tiles from a set Σ if there exists a placement of a finite number of such tiles, possibly overlapping, with each such tile assigned a weight of $+1$ and -1, such that for each cell in R the sum of the weights of tiles covering that cell add up to $+1$, while for each cell outside R the sum of the weights covering that cell is 0. Clearly a necessary condition for a tiling to exist is that a signed tiling exist. It is easy to determine when signed tilings exist for the triangle tiling problems.

THEOREM 1.3. *The triangular region T_N in the hexagonal lattice has a signed tiling by congruent copies of the triangular tile T_2 if and only if*

$$N \equiv 0 \text{ or } 2 \pmod 3.$$

THEOREM 1.4. *The triangular region T_N in the hexagonal lattice has a signed tiling by congruent copies of the three-in-line tile L_3 if and only if*

$$N \equiv 0 \text{ or } 8 \pmod 9.$$

These results are proved in Section 4.

Section 5 studies a notion of *generalized coloring argument* which includes known coloring and weighting arguments as special cases. Any generalized coloring argument that proves nonexistence of a tiling also proves nonexistence of a signed tiling (see Theorem 5.2). In view of the theorems above we immediately obtain the following consequence.

THEOREM 1.5. *It is impossible to solve the triangle tiling problems by a generalized coloring argument.*

This result gives a sense in which the two triangle tiling problems above do not have a simple solution.

Another interesting example is provided by a result of Walkup [26] showing that an $r \times s$ rectangle can be perfectly tiled by T-tetrominoes if and only if $r \equiv s \equiv 0 \pmod 4$. It can be checked that such rectangles have signed tilings by T-tetrominoes if and only if $rs \equiv 0 \pmod 8$. Hence this problem also cannot be solved by a generalized coloring argument. Walkup's ingenious argument is special to the T-tetromino; its relation to the combinatorial group theory approach of this paper is not obvious.

The boundary invariants defined in Section 2 can in principle be defined for tilings on finite subregions of any periodic tiling of $\mathbb{R}^2$ or of hyperbolic space $\mathbb{H}^n$.

We are indebted to Peter Doyle, Roger Lyndon, and Hugh Montgomery for helpful comments.

2. Boundary Invariants: The Tile Group

Boundary invariants can be defined for any regular lattice; for simplicity we treat only the case of the square lattice. The triangle tiling problems described in Section 1 for the hexagonal lattice can be translated into mathematically equivalent tiling problems on the square lattice, see Section 3.

The square lattice in $\mathbb{R}^2$ consists of lattice points, edges, and cells. A *lattice point* is a member of $\mathbb{Z}^2$. Two lattice points are *neighbors* if they are at distance one from each other, so each lattice point has exactly four neighbors. An *edge* is a line segment connecting two neighboring lattice points; it is either *horizontal* or *vertical*. A *cell* is the set of all points making up the interior and boundary of a square of area one having its four vertices at lattice points.

A (*directed*) *path* P in the square lattice consists of a sequence of directed edges specified by a sequence of lattice points $\{(x_i, y_i): 0 \leqslant i \leqslant n\}$, where the ith directed edge connects (x_{i-1}, y_{i-1}) to (x_i, y_i). It is *closed* if $(x_0, y_0) = (x_n, y_n)$. A directed path is *simple* if no edge appears twice and if it does not cross itself, where we say a path *crosses itself* if there is

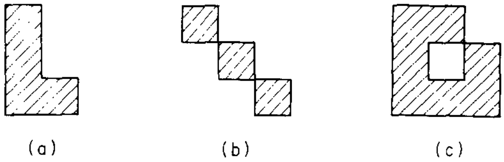


FIG. 2.1. Arrangements of cells, (a) and (b) are simply connected, (c) is not.

$0<i<n$ and $j \neq i$ with $(x_i, y_i) = (x_j, y_j)$ and the two edges from (x_{i-1}, y_{i-1}) to (x_{i+1}, y_{i+1}) consist of either two horizontal or two vertical edges.

A *region* R is a finite connected set of closed cells. The *topological boundary* ∂R of R is an (unordered) set of directed edges found as follows. The topological boundary ∂C of a cell C consists of its four edges, oriented counterclockwise. The boundary of ∂R is formed by taking the set of all edges in ∂C for all cells C in R, and discarding any edges that occur twice with opposite orientations. A region R is *simply connected* if its complement $\bar{R} = \mathbb{R}^2 - R$ is connected and if its boundary edges can be ordered to form a simple closed path. (This definition coincides with R being simply connected in the usual topological sense [23, p. 144].) Some examples illustrating these definitions are pictured in Fig. 2.1.

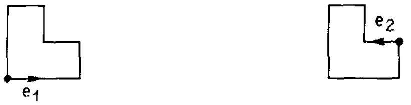
A simple closed path bounding a simply connected region R is uniquely specified by its first edge $\mathbf{e}$; we call such a path an *oriented boundary of R with leading edge* $\mathbf{e}$ and denote it by $\partial R(\mathbf{e})$. The first vertex in $\partial R(\mathbf{e})$ is called the *base point* of $\partial R(\mathbf{e})$. Some examples are shown in Fig. 2.2.

An *n-tile* is a simply connected region consisting of n cells. The notion of n-tile differs slightly from *n-omino* in that an n-tile may possibly be disconnected by removal of a single point while an n-omino may not, and n-ominos are required to be connected but not necessarily simply connected.

A *tile type* consists of the set of all translations of a tile.

A *tiling problem* consists of a region R and a set Σ of tile types. A region R can be *covered* or *tiled* by Σ if there exists a set of tiles in Σ that cover each cell of R exactly once.

We describe directed paths in the square lattice by words in the free group $\mathbf{F} = \langle A, U \rangle$ on two generators (where $A =$ "across," $U =$ "up"). To



$\partial R(e_1) = U^{-2}A^{-1}UA^{-1}UA^2$ $\partial R(e_2) = UA^2U^{-2}A^{-1}UA^{-1}$

FIG. 2.2. Oriented boundaries and associated words in free group.

the path $P=\{(x_i, y_i): 0 \leqslant i \leqslant n\}$ we assign the word $W=W(P)$ in **F** given by

$$W=G_n G_{n-1} \cdots G_1,$$

where

$$G_i=\begin{cases} A & \text{if } (x_i, y_i)=(x_{i-1}+1, y_{i-1}), \\ A^{-1} & \text{if } (x_i, y_i)=(x_{i-1}-1, y_{i-1}), \\ U & \text{if } (x_i, y_i)=(x_{i-1}, y_{i-1}+1), \\ U^{-1} & \text{if } (x_i, y_i)=(x_{i-1}, y_{i-1}-1). \end{cases}$$

Figure 2.2 gives the words associated to the oriented boundaries with specified base points for the regions pictured.

There is an obvious mapping in the reverse direction which assigns to each word W in **F** the directed path $P(W)$ starting from the fixed base point $(0, 0)$ in $\mathbb{Z}^2$ obtained by reading the word W from right to left, and one clearly has $W(P(W))=W$.

Given an oriented boundary $\partial R(\mathbf{e})$ of a simply connected region R we let $\partial R(\mathbf{e})$ also stand for the word $W(\partial R(\mathbf{e}))$ in **F**. The words

$$\{\partial R(\mathbf{e}): \mathbf{e} \text{ a counterclockwise oriented edge of } \partial R\}$$

are cyclic permutations of each other, hence are all conjugate in **F**. For example for the regions in Fig. 2.2,

$$\partial R(\mathbf{e}_2)=(UA^2)\, \partial R(\mathbf{e}_1)(UA^2)^{-1}.$$

The *combinatorial boundary* $[\partial R]$ of a simply connected region R is the conjugacy class in **F** containing all the oriented boundaries $\partial R(\mathbf{e})$ of R, i.e.,

$$[\partial R]=\{W\, \partial R(\mathbf{e})\, W^{-1}: W \in \mathbf{F}\}.$$

In what follows we use standard terminology in combinatorial group theory: $\langle W_1, W_2, ...\rangle$ denotes the subgroup of a free group **F** generated by the words W_i, for any subgroup **G** of **F** let $N(\mathbf{G})$ denote the smallest normal subgroup in **F** containing **G**, and let $[\mathbf{G}:\mathbf{G}]$ denote the commutator subgroup of **G**, i.e., the group generated by the commutators $W_1 W_2 W_1^{-1} W_2^{-1}$ for all $W_1, W_2 \in \mathbf{G}$.

The *cycle group* **C** is the subgroup of the free group **F** consisting of all words associated to closed directed paths in the square lattice. The combinatorial boundary of any simply connected region is contained in the cycle group **C**. In Section 5 we show that the cycle group is the commutator subgroup $[\mathbf{F}:\mathbf{F}]$ of **F**, hence is a normal subgroup of **F**, and in fact it can be shown that $\mathbf{C}=N(\langle AUA^{-1}U^{-1}\rangle)$.

We assign to a set of tiles $\Sigma = \{R_i\}$ a subgroup of $\mathbf{F}$ that contains all the boundaries of the tiles. The *tile group* $\mathbf{T}(\Sigma)$ is the smallest normal subgroup of $\mathbf{F}$ containing the combinatorial boundaries $[\partial R_i]$ of all tiles in Σ, i.e.,

$$\mathbf{T}(\Sigma) = N(\langle \partial R_i(\mathbf{e}_i) \colon 1 \leqslant i \leqslant m \rangle) = \langle W \, \partial R_i(\mathbf{e}_i) \, W^{-1} \colon W \in \mathbf{F},\ 1 \leqslant i \leqslant m \rangle.$$

Here $\partial R_i(\mathbf{e}_i)$ is an oriented boundary of R_i.

The tile group $\mathbf{T}(\Sigma)$ is contained in the cycle group $\mathbf{C}$ and is certainly a normal subgroup of $\mathbf{C}$. We call the quotient group

$$\mathbf{h}(\Sigma) = \mathbf{C}/\mathbf{T}(\Sigma)$$

the *tile homotopy group*. This name is suggested by analogy with the first homotopy group, based on the observation that $\mathbf{C}$ consists of the set of (allowable) closed paths in the lattice, while (roughly speaking) $\mathbf{T}(\Sigma)$ represents the paths that can be deformed to the empty path by picking up or laying down tiles.

The basic invariant that we assign to a region R to be tiled with a set of tiles Σ is its combinatorial boundary $[\partial R]$ viewed as a conjugacy class in the tile homotopy group $\mathbf{C}/\mathbf{T}(\Sigma)$.

THEOREM 2.1. *A necessary condition that a simply connected region R have a tiling by tiles in a set Σ is that the combinatorial boundary $[\partial R]$ of R be contained in the tile group* $\mathbf{T}(\Sigma)$.

It requires some care to give a completely rigorous proof of this result. Here we sketch a proof indicating the essential ideas, omitting proofs of some facts about 2-dimensional topology that can be proved along the lines of [23, Chaps. 5, 6].

Proof (Sketch). We must show that if R has a tiling in Σ then $[\partial R] \subseteq \mathbf{T}(\Sigma)$. Since $\mathbf{T}(\Sigma)$ is a normal subgroup of $\mathbf{F}$ it suffices to show that some oriented boundary $\partial R(\mathbf{e})$ of R is in $\mathbf{T}(\Sigma)$.

The proof is by induction on the number of tiles in a tiling by Σ. The result is clear when R is tiled by a single tile in Σ. So suppose $\mathcal{T}$ is a tiling of R with $k \geqslant 2$ tiles.

CLAIM. *There exists a decomposition $R = R^* \cup R^{**}$ such that R^*, R^{**} are both nonempty simply connected regions which can be tiled by Σ, and there are directed edges $\mathbf{e}_1$ of ∂R^*, $\mathbf{e}_2$ of ∂R^{**} so that*

$$\partial R(\mathbf{e}_1) = \partial R^{**}(\mathbf{e}_2) \, \partial R^*(\mathbf{e}_1).$$

The claim immediately completes the induction step, because $\partial R^*(\mathbf{e}_1)$, $\partial R^{**}(\mathbf{e}_2) \in \mathbf{T}(\Sigma)$ by the induction hypothesis, hence $\partial R(\mathbf{e}_1) \in \mathbf{T}(\Sigma)$.

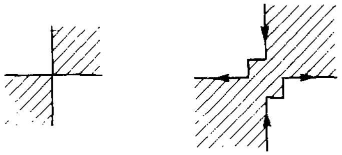


FIG. 2.3. Thickening.

Proof of Claim. First observe that the simple connectivity of R means essentially that it is topologically a disk with a simple closed curve as topological boundary. This is not literally true because R may have separating vertices, but becomes true if R is enlarged by adding two extra small squares of size ε around each separating vertex and deforming ∂R appropriately, see Fig. 2.3. (This process is called *thickening* in [23, p. 142].)

In the following argument we describe simply connected regions as though they were disks with Jordan curve boundaries, and the argument carries over to the general case by thickening.

Pick any tile S in $\mathscr{T}$ such that ∂R and ∂S have an edge in common. Then, since ∂R and ∂S are Jordan curves and $S \subseteq R$, one has joint partitions of ∂R and ∂S as

$$\partial R = \partial R_1 \cup \cdots \cup \partial R_{2j},$$

$$\partial S = \partial S_1 \cup \cdots \cup \partial S_{2j},$$

in which all ∂R_i and ∂S_i are simple paths, each $\partial R_i \neq \varnothing$ is a set of consecutive edges of ∂R, $\partial R_{2i+1} = \partial S_{2i+1}$ and $\partial S_{2i} \cap \partial R = \varnothing$. Figure 2.4 illustrates such a decomposition—the first edge of ∂R_1 is a common edge of ∂R and ∂S. (In this definition ∂S_i and ∂R are treated as sets of edges. In fact ∂S_{2i} and ∂R treated as point sets may have isolated vertices in common, see Fig. 2.4.) Note that this definition allows $\partial S_{2i} = \varnothing$, and this may sometimes occur, see Fig. 2.5b.

Now let ∂R^* denote ∂R_2 together with the reversal of all edges in ∂S_2. Then ∂R^* is a simple closed path and encloses a nonempty region R^* that

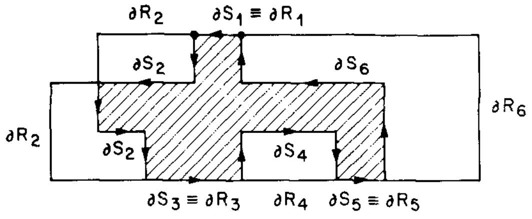
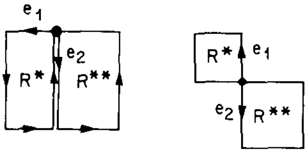


FIG. 2.4. Boundary of region R containing the tile S.



Case (a) Cases (b) and (c)

FIG. 2.5. Combining tile boundaries.

is simply connected. Let $R^{**} = R - R^*$. Then R^{**} has the simple closed path

$$\partial R^{**} = \partial S_1 \cup \partial S_2 \cup \partial S_3 \cup \partial R_4 \cup \partial S_5 \cup \cdots \cup \partial R_{2j}$$

as boundary, hence is simply connected. Now the tile S separates all the cells in R^* from the cells in $R^{**} - S$, hence all tiles in the tiling $\mathscr{T} - \{S\}$ of $R - S$ lie either in R^* or R^{**}, so $\mathscr{T}$ gives tilings $\mathscr{T}^*$ of R^* and $\mathscr{T}^{**}$ of R^{**}.

Finally, we observe that

$$\partial R(\mathbf{e}_1) = \partial R^{**}(\mathbf{e}_2)\, \partial R^*(\mathbf{e}_1),$$

where $\mathbf{e}_1$ is the first edge in ∂R_2, provided $\mathbf{e}_2$ is chosen suitably. The choice is: $\mathbf{e}_2$ is the first edge in ∂S_2 if $\partial S_2 \neq \varnothing$ (case (a)). Otherwise if $\partial S_3 \neq \varnothing$ then $\mathbf{e}_2$ is the first edge in ∂S_3 (case (b)), while if S_3 does not exist then $\mathbf{e}_2$ is the first edge in ∂S_1 (case (c)). These cases are illustrated in Fig. 2.5; case (c) occurs when $R^{**} = S$. This proves the claim. ▌

Theorem 2.1 provides a necessary condition for a perfect tiling to exist, hence serves as a criterion for proving nonexistence of perfect tilings. In general this theorem trades one hard problem for another. However in the special circumstances of the triangle tiling problems of Section 1, this criterion can be successfully applied.

3. TRIANGLE TILING PROBLEMS

The triangle tiling problems of Section 1 are easily converted to equivalent tiling problems on the square lattice. The region to be tiled becomes a "staircase" pictured in Fig. 3.1. The tile sets Σ_1 and Σ_2 for the two tiling problems are pictured in Fig. 3.2. Figure 3.2 gives a representative word for the combinatorial boundary $[\partial R]$ of each of the tiles pictured. A representative word for the boundary of the "staircase" region T_N is

$$\partial T_N = A^N U^{-N} (A^{-1} U)^N. \tag{3.1}$$

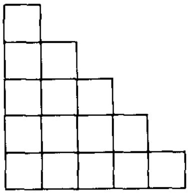
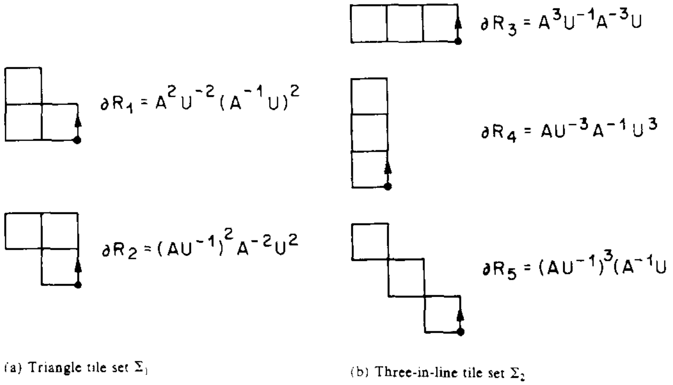


FIG. 3.1. Staircase region T_5.

The nonexistence parts of the proofs of Theorems 1.1 and 1.2 apply the criterion of Theorem 2.1: in these cases the boundary $[\partial T_N]$ is not contained in the appropriate tile group $\mathbf{T}(\Sigma_i)$. The proofs use a group-theoretic argument exploiting the special character of the tile group involved, due to the first author. One computes invariants associated to a special subgroup $\mathbf{H}$ of the free group $\mathbf{F} = \mathbf{F}_2$, defined below. The group $\mathbf{H}$ contains $[\partial T_N]$ and the tile groups $\mathbf{T}(\Sigma_i)$ of the two problems and has easily computable invariants, which are a consequence of the fact that the quotient group $\mathbf{F}/\mathbf{H}$ has a planar Cayley diagram.

Recall that the *Cayley diagram* $\mathscr{G}(\mathbf{F}_g/\mathbf{K})$ (also called the *group diagram*, *graph*, or *color diagram*) is a graph with directed labelled edges associated to a presentation of a quotient group $\mathbf{G} = \mathbf{F}_g/\mathbf{K}$ of the free group $\mathbf{F}_g$ on g generators, where $\mathbf{K}$ is a normal subgroup of relations. In the Cayley diagram of $\mathbf{G}$ each vertex corresponds to an element W of $\mathbf{G}$, and for each generator S_i of $\mathbf{F}_g$ there is a directed edge labelled i from vertex W to vertex $S_i W$. In particular every vertex in a Cayley diagram has $2g$ edges incident on it, with g edges directed inwards and g edges directed outwards.

The subgroup $\mathbf{K}$ of relations defining $\mathbf{G} = \mathbf{F}_g/\mathbf{K}$ has a simple characterization in terms of its Cayley diagram. Let $\mathscr{G}(\mathbf{F}_g/\mathbf{K})$ denote the *undirected* labelled graph obtained from $\mathscr{G}(\mathbf{F}_g/\mathbf{K})$ by ignoring the directions on the


$\partial R_3 = A^3U^{-1}A^{-3}U$
$\partial R_1 = A^2U^{-2}(A^{-1}U)^2$
$\partial R_4 = AU^{-3}A^{-1}U^3$
$\partial R_2 = (AU^{-1})^2A^{-2}U^2$
$\partial R_5 = (AU^{-1})^3(A^{-1}U$

(a) Triangle tile set Σ_1

(b) Three-in-line tile set Σ_2

FIG. 3.2. Tile sets for triangle tiling problems.

edges. Associate to any word $W = G_k^{\varepsilon_k} G_{k-1}^{\varepsilon_{k-1}} \cdots G_1^{\varepsilon_1}$ in the free group $\mathbf{F}_g$ (where the G_i are generators and each $\varepsilon_i = \pm 1$) a directed path on the edges of the undirected graph $\mathscr{G}(\mathbf{F}_g/\mathbf{K})$ starting from the identity vertex I which at the ith step follows a directed edge from the vertex labelled $W_i = G_i^{\varepsilon_i} G_{i-1}^{\varepsilon_{i-1}} \cdots G_1^{\varepsilon_1}$ to $W_{i+1} = G_{i+1}^{\varepsilon_{i+1}} W_i$ along the unique edge labelled i between W_i and W_{i+1}. Then a word W is in $\mathbf{K}$ if and only if it corresponds to a closed path in $\mathscr{G}(\mathbf{F}_g/\mathbf{K})$ starting from I.

The special subgroup $\mathbf{H}$ of $\mathbf{F}_2$ is defined by the property that it has associated quotient group $\mathbf{G} = \mathbf{F}_2/\mathbf{H}$ whose (undirected) Cayley diagram $\mathscr{G}(\mathbf{F}_2/\mathbf{H})$ is the infinite planar graph that tiles the plane with hexagons and triangles as pictured in Fig. 3.3. The shaded vertex denotes the identity element, and if $\mathbf{F}_2 = \langle A, U \rangle$ then A-generator edges border triangles labelled A and similarly for indicate U-generator edges. The graph $\mathscr{G}(\mathbf{F}_2/\mathbf{H})$ is the boundary of a lattice tesselation of the plane by equilateral triangles and hexagons. The group $\mathbf{G}$ is isomorphic to one of the 17 plane crystallographic groups (the one labelled **p3** in [6, p. 49]), and the subgroup of relations $\mathbf{H}$ is given by

$$\mathbf{H} = N(\langle A^3, U^3, (U^{-1}A)^3 \rangle). \tag{3.2}$$

In the sequel we take as the definition of $\mathbf{H}$ that its elements correspond to closed paths in the undirected Cayley diagram $\mathscr{G}(\mathbf{F}_2/\mathbf{H})$; the explicit characterization (3.2) of $\mathbf{H}$ is never used.

The relevance of $\mathbf{H}$ to the triangle tiling problems is established by the following claim.

CLAIM. *The tile groups* $\mathbf{T}(\Sigma_1)$, $\mathbf{T}(\Sigma_2)$ *and the combinatorial boundaries* $[\partial T_N]$ *for* $N \equiv 0$ *or* 2 (mod 3) *are all contained in* $\mathbf{H}$.

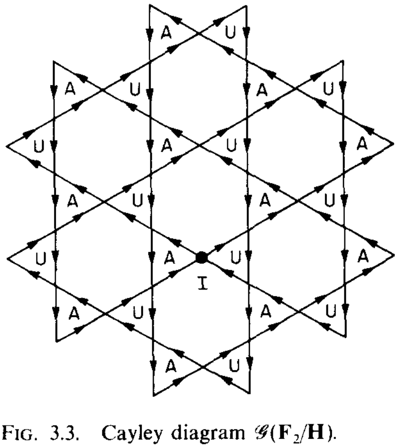


FIG. 3.3. Cayley diagram $\mathscr{G}(\mathbf{F}_2/\mathbf{H})$.

Proof of Claim. Since $\mathbf{H}$ is a normal subgroup of $\mathbf{F}_2$, it suffices to check that representative generators of $\{[\partial R]: R \in \Sigma_i\}$ and of $[\partial T_N]$ are in $\mathbf{H}$. To do this, one checks that such generators give closed paths starting from I in $\mathscr{G}(\mathbf{F}_2/\mathbf{H})$. This is easily done for the boundaries ∂R given in Fig. 3.2. It remains to check ∂T_N. To do this, one observes first that A^3, U^3 and $(A^{-1}U)^3$ are in $\mathbf{H}$. Next, these relations imply that $\partial T_N = A^N U^{-N} (A^{-1}U)^N$ is in $\mathbf{H}$ provided that ∂T_i is in $\mathbf{H}$ for $N \equiv i \pmod 3$, and one easily checks that ∂T_2, ∂T_3 are both in $\mathbf{H}$. ▌

The planar nature of the Cayley diagram $\mathscr{G} = \mathscr{G}(\mathbf{F}_2/\mathbf{H})$ gives rise to a large class of group-theoretic invariants associated to elements of $\mathbf{H}$, which consist of the winding numbers of the paths associated to elements of $\mathbf{H}$ about the hexagonal and triangular cells in the plane of the Cayley diagram. Let s be a cell (either hexagonal or triangular) in this tiling and let x_s be a point in the interior of s. The *winding number* (or *index*) $w(P; s)$ of a closed directed path $\mathbf{P}$ in $\mathscr{G}$ around s counts the number of times $\mathbf{P}$ encloses the cell s in the counterclockwise direction and is given by

$$w(\mathbf{P}; s) = \frac{1}{2\pi i} \oint_P \frac{1}{z - x_s}\, dz. \tag{3.3}$$

This quantity $w(P; s)$ is well defined independent of the choice of point x_s in s, and is *additive* in the sense that for two closed paths P_1 and P_2 starting at the same point W in $\mathscr{G}$ one has

$$w(P_2 P_1; s) = w(P_1; s) + w(P_2; s). \tag{3.4}$$

These facts about winding numbers in $\mathbb{R}^2$ are proved in basic texts on complex analysis, cf. [1, pp. 114–118; 15, pp. 233–241].

The winding number $w(\ ; s)$ induces a map $w(\ ; s)\colon \mathbf{H} \to \mathbb{Z}$ which assigns to a word $V \in \mathbf{H}$ the value $w(P(V); s)$ where $P(V)$ is the closed directed path in $\mathscr{G}$ associated to V, and (3.4) shows that this mapping is a homomorphism. Let S be any finite or infinite set of cells in $\mathscr{G}$ and let

$$w(P; S) = \sum_{s \in S} w(P; s).$$

This is well defined, since any closed path in the Cayley graph encloses a finite number of cells and it is clear that $w(\ ; S)\colon \mathbf{H} \to \mathbb{Z}$ is a homomorphism.

Now we use these invariants to solve the triangle tiling problems.

Proof of Theorem 1.1. Since all tiles in Σ_1 cover three cells, the region T_N cannot be tiled unless the number of cells $N(N+1)/2$ in T_N is a multiple of 3, hence $N \equiv 0$ or $2 \pmod 3$.

We will show that ∂T_N is not in the tile group $\mathbf{T}(\Sigma_1)$ when $N \equiv 3, 5, 6,$ or 8 (mod 12), and hence that no tiling exists in these cases by Theorem 2.1. Consider the homomorphism $\phi\colon \mathbf{H} \to \mathbb{Z}$ with $\phi(V) = w(V; S)$, where S is the set of all hexagons in the Cayley diagram $\mathcal{G} = \mathcal{G}(\mathbf{F}_2/\mathbf{H})$. Using the boundaries in Fig. 3.2a it is easy to calculate that $\phi(\partial R_1) = 1$, $\phi(\partial R_2) = -1$. The translation-invariance of the Cayley graph $\mathcal{G}$ allows one to see that

$$\phi(W\,\partial R_i W^{-1}) = \phi(\partial R_i), \qquad \text{for} \quad i = 1, 2,$$

for any word W in the free group. Hence

$$\phi([\partial R_1]) = 1, \qquad \phi([\partial R_2]) = -1.$$

We know that $\partial T_N \in \mathbf{H}$, and (3.1) yields

$$\phi(\partial T_N) = \left[\frac{N+1}{3}\right]. \tag{3.5}$$

Suppose that ∂T_N is in the tile group $\mathbf{T}(\Sigma_1)$, in which case there exists an integer m and words W_i such that

$$\partial T_N = \prod_{i=1}^{m} W_i (\partial R_{k_i})^{\varepsilon_i} W_i^{-1}, \tag{3.6}$$

where each $k_i = 1$ or 2 and each $\varepsilon_i = 1$ or -1. Then

$$\phi(\partial T_N) = \sum_{i=1}^{m} \phi(W_i (\partial R_{k_i})^{\varepsilon_i} W_i^{-1}) = \sum_{i=1}^{m} \varepsilon_i \phi(\partial R_{k_i}) \equiv m \ (\text{mod } 2). \tag{3.7}$$

Next we introduce a second homomorphism $\psi\colon \mathbf{H} \to \mathbb{Z}$ which views a word in $\mathbf{H}$ as defining a closed directed path in the square lattice $\mathbb{Z}^2$ starting at $(0, 0)$ as in the proof of Theorem 2.1 and which associates to each such path the sum of its winding numbers about all cells in the square lattice. That is, for a tile R the mapping $\psi(\partial R)$ counts the number of cells covered by the tile, so that, for example, one has

$$\psi(W\,\partial R_i W^{-1}) = \psi(\partial R_i) = 3 \qquad \text{for} \quad i = 1, 2,$$

for all W in the free group $\mathbf{F}_2$, and one has

$$\psi(\partial T_N) = \binom{N+1}{2}. \tag{3.8}$$

Now the hypothesis (3.6) gives

$$\psi(\partial T_N) = \sum_{i=1}^{m} \psi(W_i (\partial R_{k_i})^{\varepsilon_i} W_i^{-1}) = \sum_{i=1}^{m} \varepsilon_i \psi(\partial R_{k_i}) \equiv m \ (\text{mod } 2). \tag{3.9}$$

Combining (3.5), (3.7), (3.8), and (3.9) yields

$$\left[\frac{N+1}{3}\right] \equiv \binom{N+1}{2} \pmod 2,$$

which is a necessary condition for ∂T_N to be in the tile group $T(\Sigma_1)$. Both sides of this congruence are periodic (mod 12), and it is easily checked that this congruence does not hold for $N \equiv 3$, 5, 6, and 8 (mod 12), proving that $\partial T_N \notin \mathbf{T}(\Sigma_1)$ in these cases.

It is easy to construct tilings for $N \equiv 0$, 2, 9, or 11 (mod 12). We leave it to the reader to construct such tilings for T_2, T_9, T_{11}, and T_{12}. One then proceeds by induction on K, constructing tilings for T_{12K+L} for $L=2$, 9, 11, and 12 from that for T_{12K} using the scheme pictured in Fig. 3.4, noting that since a 2×3 rectangle can be tiled, so can a 5×6 rectangle and an 11×12 rectangle, whence an $L \times 12K$ rectangle can be tiled.

Proof of Theorem 1.2. Since all tiles in Σ_2 cover three cells, one must have $N \equiv 0$ or 2 (mod 3) as above. We will show that ∂T_N is not in the tile group $\mathbf{T}(\Sigma_2)$ in all cases, so that a tiling of T_N is impossible by Theorem 2.1. To do this, consider the homomorphism $\phi\colon \mathbf{H} \to \mathbb{Z}$ which counts the sum of all winding numbers around all triangles labelled U in the Cayley diagram $\mathscr{G}(\mathbf{F}_2/\mathbf{H})$ in Fig. 3.3. One easily calculates using the boundaries in Fig. 3.2b that

$$\phi(\partial R_3) = \phi(\partial R_4) = \phi(\partial R_5) = 0. \tag{3.10}$$

As in the previous proof one has

$$\phi(W\,\partial R_i W^{-1}) = \phi(\partial R_i) \qquad \text{for} \quad i = 3, 4, 5,$$

for all W in the free group. A computation using (3.1) in the Cayley diagram yields

$$\phi(\partial T_N) = \left[\frac{N+1}{3}\right]. \tag{3.11}$$

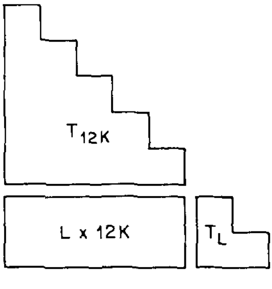


FIG. 3.4. Tiling of T_{12K+L} for $L = 2$, 9, 11, 12.

Suppose that ∂T_N were in the tile group $\mathbf{T}(\Sigma_2)$, so that

$$\partial T_N = \prod_{i=1}^{m} W_i(\partial R_{k_i})^{\varepsilon_i} W_i^{-1},$$

where each $W_i \in \mathbf{F}_2$, each $k_i \in \{3, 4, 5\}$ and each $\varepsilon_i = 1$ or -1. Then

$$\phi(\partial T_N) = \sum_{i=1}^{m} \varepsilon_i \phi(\partial R_{k_i}) = 0,$$

by (3.10). This contradicts (3.11) for $N \geqslant 2$, and this contradiction proves that ∂T_N is not in $\mathbf{T}(\Sigma_2)$ for $N \equiv 0$ or 2 (mod 3). ▌

A wide variety of related tiling problems can be solved using invariants associated to groups $\mathbf{H}$ for which $\mathbf{F}_g/\mathbf{H}$ has a planar Cayley diagram. See Thurston [25] for extensions of this approach and more examples.

4. Triangle Tiling Problem: Signed Tilings

Recall that a *signed tiling* of a region R by tiles from a set Σ consists of placements of a finite set of tiles, each assigned a weight of 1 or -1, such that for each cell in R the sum of the weights of the tiles covering this cell is 1 and for each cell not in R the sum of the weights covering this cell is 0.

Proof of Theorem 1.3. Since each tile in a signed tiling covers ± 3 cells (taking weights into account), the number of cells in T_N must be $\equiv 0$ (mod 3). Since T_N has $\binom{N+1}{2}$ cells, this requires $N \equiv 0$ or 2 (mod 3).

It suffices to exhibit signed tilings for $N \equiv 3$, 5, 6, 8 (mod 12). Let $N = 12K + L$ where $L = 3$, 5, 6, or 8. Then the triangular region T_N may be decomposed as pictured in Fig. 3.4 into regions T_{12K}, T_L and a rectangular region $L \times 12K$. The region T_{12K} may be tiled by Theorem 1.1, and the $L \times 12K$ rectangular region can also be tiled by congruent copies of the triangular tile T_3, by observing that a 3×2 rectangle can be so tiled, as can 5×6, 6×6, 8×6 rectangles. Hence to prove the theorem it suffices to find signed tilings of T_3, T_5, T_6, and T_8. Such tilings are easy to find. A signed tiling for T_3 is pictured in Fig. 4.1; signed tilings for T_5, T_6, and T_8 are left as exercises for the reader. ▌

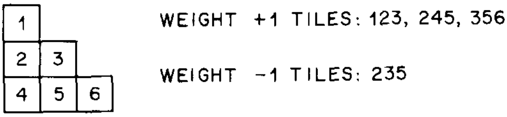


Fig. 4.1. Signed tiling of T_3 by triangle tiles.

Proof of Theorem 1.4. We first show that

$$N \equiv 0 \text{ or } 8 \pmod 9)$$

is a necessary condition for a signed tiling to exist. Number consecutively the horizontal rows of cells in the staircase region T_N in the square lattice so that row j contains exactly j cells. The tile set Σ_2 consists of three tiles labelled R_3, R_4, R_5 in Fig. 3.2, which we relabel A, B, and C, respectively, for notational convenience. A placement of tile A always has three cells in a single row, while B and C always have one cell in each of three contiguous rows. Suppose that a signed tiling exists for T_N and for this tiling let $n_{BC}^{+}(j)$ (resp. $n_{BC}^{-}(j)$) count the number of tiles of type B or C having weight $+1$ (resp. -1) which contain one cell in each of rows j, $j+1$, and $j+2$, and set

$$n_{BC}(j) = n_{BC}^{+}(j) - n_{BC}^{-}(j).$$

By counting the number of tiles covered in row j by this signed tiling one finds that

$$n_{BC}(j-2) + n_{BC}(j-1) + n_{BC}(j) \equiv j \pmod 3), \qquad 1 \leqslant j \leqslant N, \tag{4.1a}$$

$$n_{BC}(j-2) + n_{BC}(j-1) + n_{BC}(j) \equiv 0 \pmod 3), \qquad j \leqslant 0 \text{ or } j > N. \tag{4.1b}$$

Since a signed tiling is finite, there is a positive integer k such that all tiles are in rows $-k$ to $+k$. Hence $n_{BC}(-k-1) = n_{BC}(-k-2) = 0$ and applying the congruences for $j = -k, -k+1, ..., 0$ successively one obtains

$$n_{BC}(j) \equiv 0 \pmod 3), \qquad -k \leqslant j \leqslant 0.$$

Similarly starting from $n_{BC}(k+1) = n_{BC}(k+2) = 0$ and working backwards using $j = k+2, k+1, k, ..., N+3$ successively one obtains

$$n_{BC}(j) \equiv 0 \pmod 3), \qquad N+1 \leqslant j \leqslant k. \tag{4.2}$$

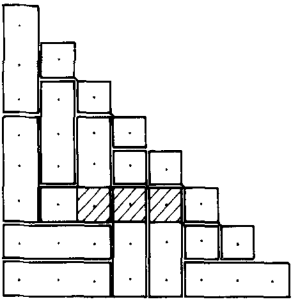


FIG. 4.2. Signed tiling of T_8 by three-in-line tiles. (A weight -1 tile is placed on the shaded squares.)

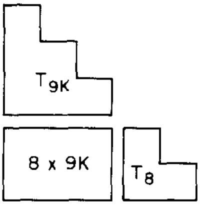


FIG. 4.3. Signed tiling of T_{9K+8} by three-in-line tiles.

Now working forwards using the congruences for $j = 1, 2, ..., N$ one finds for $1 \leqslant j \leqslant N$ that $n_{BC}(j)$ (mod 3) is periodic with period 9 and takes the values (1, 1, 1, 2, 2, 2, 0, 0, 0) for $j = (1, 2, 3, 4, 5, 6, 7, 8, 9)$, respectively. But (4.2) implies that $n_{BC}(N-1) \equiv n_{BC}(N) \equiv 0 \pmod 3$; this is impossible unless $N \equiv 0$ or $8 \pmod 9$.

It remains to construct signed tilings for $N \equiv 0$ or $8 \pmod 9$. A signed tiling for $N = 8$ is easy to find, and one is given in Fig. 4.2. Signed tilings for $N = 9K$, $9K + 8$ can be constructed by induction on K. Given a signed tiling for $9(K-1)+8$ we obtain one for $9K$ by tiling the last row with tiles of type R_3. Then given a signed tiling of T_{9K} we may subdivide T_{9K+8} as pictured in Fig. 4.3, and use the signed tilings of T_{9K} and T_8 provided by the induction hypothesis together with a tiling of the $8 \times 9K$ rectangular region with R_3 tiles. This completes the construction. ▌

5. Generalized Coloring Arguments and Tile Homology

Many tiling problems have been resolved using arguments involving colorings or weightings of the cells of the underlying lattice. We show that such arguments have a natural interpretation in terms of boundary invariants and that the strongest such arguments are equivalent to detecting the existence of signed tilings.

Consider the square lattice with its associated free group $\mathbf{F} = \langle A, U \rangle$ and cycle group $\mathbf{C} = [\mathbf{F} : \mathbf{F}]$. Coloring or weighting arguments correspond to additive invariants assigned to cells of the square lattice. Part (iii) of the following theorem shows that a natural group encoding such invariants is the maximal abelian quotient group $\mathbf{A}_0 = \mathbf{C}/[\mathbf{C} : \mathbf{C}]$, which we call the *cell group*.

THEOREM 5.1. (i) *The cycle group* $\mathbf{C}$ *consists of all words* W *such that* $P(W)$ *is a closed directed path in* $\mathbb{Z}^2$, *i.e.*, $\mathbf{C} = [\mathbf{F} : \mathbf{F}]$.

(ii) *The group* $[\mathbf{C} : \mathbf{C}]$ *consists of all words* W *such that* $P(W)$ *is a closed directed path in* $\mathbb{Z}^2$ *with winding number* 0 *around every cell in* $\mathbb{Z}^2$. *Consequently,* $[\mathbf{C} : \mathbf{C}]$ *is a normal subgroup of* $\mathbf{F}$.

(iii) *The group* $\mathbf{A}_0 = \mathbf{C}/[\mathbf{C}:\mathbf{C}]$ *is a direct sum of a countable number of copies of* $\mathbb{Z}$, *which are in one-to-one correspondence with the cells* c_{ij} *of the lattice* $\mathbb{Z}^2$. *The projection map* $\pi_{i,j}: \mathbf{C} \to \mathbb{Z}$ *onto the* c_{ij}*th* $\mathbb{Z}$*-summand of* $\mathbf{A}_0$ *is given by the winding number* $w(P(W); c_{ij})$.

We defer the proof of this theorem to the end of this section, in order to proceed directly to the discussion of coloring arguments.

A *generalized coloring map* is any homomorphism $\phi: \mathbf{C} \to \mathbf{A}$, where $\mathbf{A}$ is an abelian group. A *generalized coloring argument* uses a generalized coloring map ϕ to show that a simply connected region R cannot be tiled by tiles in a set Σ by showing that the image of the combinatorial boundary $[\partial R]$ under ϕ is not contained in the image of the tile group $\mathbf{T}(\Sigma)$ under ϕ. Since all such homomorphisms ϕ can be factored as the projection $\pi: \mathbf{C} \to \mathbf{A}_0 = \mathbf{C}/[\mathbf{C}:\mathbf{C}]$ composed with a homomorphism $\tilde{\phi}: \mathbf{A}_0 \to \mathbf{A}$, the strongest generalized coloring map is the projection π onto the cell group $\mathbf{A}_0$.

We justify the name "generalized coloring argument" by showing how the coloring argument in Golomb [9] can be formulated in terms of a generalized coloring map. It is well known that the checkerboard with two opposite corners removed ("mutilated checkerboard") pictured in Fig. 5.1 cannot be tiled with dominoes. To prove this one colors the checkerboard in a checkerboard pattern. Depending on where the mutilated checkerboard is placed on the lattice, it covers either 30 black squares and 32 white squares or 32 black squares and 30 white squares. Since each domino in a tiling covers one square of each color, any perfectly tiled region must contain the same number of squares of each color, hence the mutilated checkerboard cannot be tiled.

To obtain an equivalent generalized coloring argument one colors the cells of the lattice $\mathbb{Z}^2$ in a checkerboard pattern with the cell c_{ij} having lower left corner (i, j) being colored black if $i + j \equiv 0 \pmod 2$ and white if $i + j \equiv 1 \pmod 2$. Now define a map $\phi: \mathbf{C} \to \mathbb{Z} \oplus \mathbb{Z}$ given by $\phi = \phi_1 \oplus \phi_2$, where $\phi_1(W)$ counts the sum of the winding numbers of the closed path $P(W)$ about all cells c_{ij} with $i + j \equiv 0 \pmod 2$ (the "white" cells) and $\phi_2(W)$

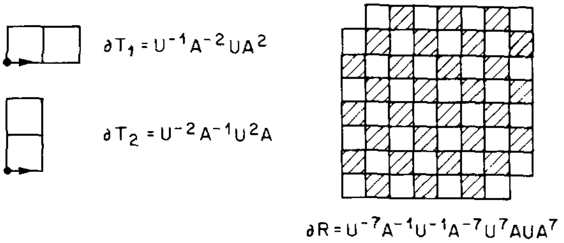

$\partial T_1 = U^{-1}A^{-2}UA^2$
$\partial T_2 = U^{-2}A^{-1}U^2A$
$\partial R = U^{-7}A^{-1}U^{-1}A^{-7}U^7AUA^7$

FIG. 5.1. Mutilated checkerboard and dominoes.

denotes the sum of the winding numbers of the closed path $P(W)$ about all cells c_{ij} with $i+j \equiv 1 \pmod 2$ (the "black cells"). The mutilated checkerboard R has boundary $\partial R = U^{-7}(A^{-1}U^{-1})A^{-7}U^{7}(AU)A^{7}$ while

$$\mathbf{T}(\Sigma) = N(\langle U^{-1}A^{-2}UA^{2}, U^{-2}A^{-1}U^{2}A\rangle),$$

see Fig. 5.1. Now $\phi([\partial R]) = \{(30, 32), (32, 30)\}$ while $\phi(\mathbf{T}(\Sigma)) = \{(n, n) \colon n \in \mathbb{Z}\}$, which shows that R cannot be tiled by dominoes.[1]

Other coloring and weighting arguments used in [5, 7, 9, 10, 13, 17] can be framed in terms of generalized coloring maps in similar fashion.

The information about nonexistence of tilings given by any generalized coloring map $\phi \colon \mathbf{C} \to \mathbf{A}$ is completely expressed in terms of the tile homotopy group, using the quotient map $\bar{\phi} \colon \mathbf{C}/\mathbf{T}(\Sigma) \to \tilde{\mathbf{A}} = \mathbf{A}/\phi(\mathbf{T}(\Sigma))$ induced by factoring out the tile group $\mathbf{T}(\Sigma)$; indeed $\phi([\partial R])$ is contained in $\phi(\mathbf{T}(\Sigma))$ if and only if $\bar{\phi}([\partial R])$ consists of the identity element in $\tilde{\mathbf{A}}$. Conversely, any homomorphism $\bar{\phi}$ from the tile homotopy group $\mathbf{h}(\Sigma)$ into an abelian group $\tilde{\mathbf{A}}$ arises from the generalized coloring map $\phi \colon \mathbf{C} \to \tilde{\mathbf{A}}$ given by $\phi = \bar{\phi} \circ \bar{\pi}$, where $\bar{\pi} \colon \mathbf{C} \to \mathbf{C}/\mathbf{T}(\Sigma)$ is the natural projection. Thus we may equally well consider generalized coloring arguments as specified by homomorphisms $\bar{\phi}$ from the tile homotopy group $\mathbf{h}(\Sigma)$ to abelian groups $\tilde{\mathbf{A}}$.

In this new context the maximal information available about tilings is given by the map $\pi_s \colon \mathbf{h}(\Sigma) \to \mathbf{H}(\Sigma)$, where $\mathbf{H}(\Sigma)$ is the maximal abelian quotient group of $\mathbf{h}(\Sigma)$.[2] We call $\mathbf{H}(\Sigma)$ the *tile homology group*, by analogy with the well-known fact that the first homology group is the maximal abelian quotient group of the first homotopy group. Using the projection $\bar{\pi} \colon \mathbf{C} \to \mathbf{C}/\mathbf{T}(\Sigma)$ we have $\mathbf{H}(\Sigma) = \mathbf{C}/\mathbf{B}(\Sigma)$, where $\mathbf{B}(\Sigma)$ is the kernel of $\pi_s \circ \bar{\pi}$. We call $\mathbf{B}(\Sigma)$ the *tile boundary group*. $\mathbf{B}(\Sigma)$ is the smallest normal subgroup of $\mathbf{C}$ containing $\mathbf{T}(\Sigma)$ and $[\mathbf{C} \colon \mathbf{C}]$. We claim that

$$\mathbf{B}(\Sigma) = \mathbf{T}(\Sigma)[\mathbf{C} \colon \mathbf{C}], \tag{5.1}$$

and that $\mathbf{B}(\Sigma)$ is a normal subgroup of $\mathbf{F}$. The inclusion $\mathbf{T}(\Sigma)[\mathbf{C} \colon \mathbf{C}] \subseteq \mathbf{B}(\Sigma)$ is clear. To prove the other inclusion, note that $\mathbf{T}(\Sigma)[\mathbf{C} \colon \mathbf{C}]$ is a normal subgroup of $\mathbf{F}$ (hence of $\mathbf{C}$) using the general fact that $\mathbf{G}_1\mathbf{G}_2 = \{g_1 g_2 \colon g_1 \in \mathbf{G}_1, g_2 \in \mathbf{G}_2\}$ is a normal subgroup of a group $\mathbf{G}$ whenever $\mathbf{G}_1$ and $\mathbf{G}_2$ are both normal subgroups of $\mathbf{G}$. Since $\mathbf{B}(\Sigma)$ is the smallest normal subgroup of $\mathbf{C}$ containing $\mathbf{T}(\Sigma)$ and $[\mathbf{C} \colon \mathbf{C}]$, it follows that $\mathbf{B}(\Sigma) \subseteq \mathbf{T}(\Sigma)[\mathbf{C} \colon \mathbf{C}]$, proving the claim.

[1] Since homomorphisms map conjugacy classes to conjugacy classes, one may ask why the image $\phi([\partial R])$ is not a single element in the *abelian* group $\mathbb{Z} \oplus \mathbb{Z}$. This is because $[\partial R]$ is actually an $\mathbf{F}$-conjugacy class, so its image is actually a conjugacy class in the *nonabelian* group $\mathbf{F}/[\mathbf{C} \colon \mathbf{C}]$.

[2] The map π_s is exactly the quotient map $\pi_s \colon \mathbf{C}/\mathbf{T}(\Sigma) - \mathbf{A}_0/\pi(\mathbf{T}(\Sigma))$ induced from the strongest generalized coloring map $\pi \colon \mathbf{C} \to \mathbf{A}_0 = \mathbf{C}/[\mathbf{C}, \mathbf{C}]$, as is easily checked.

The discussion above shows that the maximal information about nonexistence of tilings obtainable by a generalized coloring argument concerns whether or not the combinatorial boundary $[\partial R]$ is in the tile boundary group $B(\Sigma)$. Now we show that this condition is a necessary and sufficient condition for a signed tiling to exist.

THEOREM 5.2. *For a simply connected region* R *and set of tiles* Σ *the following conditions are equivalent*:

(i) *R has a signed tiling using tiles in* Σ.

(ii) *The combinatorial boundary* $[\partial R]$ *is in the tile boundary group* $\mathbf{B}(\Sigma)$.

Proof. (i) ⇒ (ii). Suppose that R has a signed tiling. Place R on $\mathbb{Z}^2$ so that it has an oriented boundary ∂R with base point $(0,0)$. Let $\{(T_i, \varepsilon_i)\colon 1 \leqslant i \leqslant k\}$ denote the signed tiling of R, with $\varepsilon_i = 1$ or -1 being the sign of the tile T_i. Let ∂T_i denote an oriented boundary of tile T_i with base point $(0,0)$, and let W_i be an oriented path from $\mathbf{m}_i$ to $(0,0)$, where $\mathbf{m}_i$ is the basepoint where the tile T_i is placed. Consider the word

$$W = (\partial R)^{-1} \prod_{i=1}^{k} (W_i (\partial T_i)^{\varepsilon_i} W_i^{-1}).$$

We claim that $P(W)$ is a closed path which has winding number 0 about all cells in $\mathbb{Z}^2$, so that by Theorem 5.1(ii), $W \in [\mathbf{C}:\mathbf{C}]$. To see this, we note that $P((\partial R)^{-1})$ has winding number -1 about all cells in R and winding number 0 elsewhere, while $P(W_i(\partial T_i)^{\varepsilon_i} W_i^{-1})$ has winding number ε_i about all cells in T_i and winding number 0 elsewhere, so the claim follows by definition of a signed tiling. Thus $W \in [\mathbf{C}, \mathbf{C}]$ and

$$\partial R = \left(\prod_{i=1}^{k} (W_i (\partial T_i)^{\varepsilon_i} W_i^{-1}) \right) W^{-1}$$

is expressed as an element of $\mathbf{T}(\Sigma)[\mathbf{C}:\mathbf{C}]$, which is $\mathbf{B}(\Sigma)$. Since $\mathbf{B}(\Sigma)$ is a normal subgroup of $\mathbf{F}$, the conjugacy class $[\partial R] \subseteq \mathbf{B}(\Sigma)$.

(ii) ⇒ (i). Place R so that it has an oriented boundary ∂R with base point $(0,0)$. Since $\partial R \in \mathbf{B}(\Sigma)$ and $\mathbf{B}(\Sigma) = \mathbf{T}(\Sigma)[\mathbf{C}:\mathbf{C}]$, one has

$$\partial R = \left(\prod_{i=1}^{k} (W_i (\partial T_i)^{\varepsilon_i} W_i^{-1}) \right) W^{-1},$$

where ∂T_i are oriented boundaries of tiles in Σ with basepoint $(0,0)$, ε_i takes values ± 1, and $W \in [\mathbf{C}:\mathbf{C}]$. Now we can reverse the previous argument. By Theorem 5.1(ii) the path $P(W)$ associated to the word

$$W = (\partial R)^{-1} \left(\prod_{i=1}^{k} (W_i (\partial T_i)^{\varepsilon_i} W_i^{-1}) \right)$$

has winding number 0 about all cells. Computing the winding numbers of $P((\partial R)^{-1})$ and $P(W_i(\partial T_i)^{\varepsilon_i} W_i^{-1})$ about each cell shows that $\{(T_i, \varepsilon_i)\}$ is a signed tiling of R. ▌

Theorem 1.5 follows easily from this result.

Proof of Theorem 1.5. Theorem 5.3 and the discussion preceding it show that a generalized coloring argument can only prove the nonexistence of a tiling by proving the nonexistence of a signed tiling. Since we have shown that both triangle tiling problems have instances having a signed tiling but no perfect tiling, these problems cannot be solved by any generalized coloring argument. ▌

We have obtained the following hierarchy of successively weaker tiling invariants.

THEOREM 5.3. *Let R be a simply connected region and Σ a set of tiles. Consider the conditions*:

(H1) *R can be tiled using tiles in Σ.*

(H2) *$[\partial R]$ is in the tile group $\mathbf{T}(\Sigma)$.*

(H3) *$[\partial R]$ is in the tile boundary group $\mathbf{B}(\Sigma)$.*

Then (H1) $\Rightarrow$ (H2) $\Rightarrow$ (H3). *These implications are not reversible in general.*

Proof. The assertion (H1) $\Rightarrow$ (H2) is exactly Theorem 2.1. (H2) $\Rightarrow$ (H3) is immediate.

To show (H2) $\not\Rightarrow$ (H1) let the tile set Σ consist of a 2×2 square and a 3×3 square. Consider the L-shaped region R obtained by removing a 2×2 square from the upper right corner of a 3×3 square. It is clear that $[\partial R] \in \mathbf{T}(\Sigma)$ and that R cannot be tiled by translates of these two tiles.

The implication (H3) $\not\Rightarrow$ (H2) follows from the triangle tiling by lines problem. By Theorem 1.4 a signed tiling exists for $N \equiv 0$ or 8 (mod 9), and (H3) then holds by Theorem 5.2. The proof of Theorem 1.2 shows that (H2) does not hold in this case. ▌

By using semigroups instead of groups one can obtain a necessary and sufficient condition for a tiling to exist. The *tile semigroup* $\mathbf{T}^+(\Sigma)$ is defined to be the subsemigroup of the free group $\mathbf{F}$ generated by the conjugacy classes $\{[\partial T]: T \in \Sigma\}$.

THEOREM 5.4. *Let R be a simply connected region and Σ a set of tiles. The following conditions are equivalent*:

(i) *R can be tiled by tiles in Σ.*

(ii) *$[\partial R]$ is contained in the tile semigroup $\mathbf{T}^+(\Sigma)$.*

Proof. (i) $\Rightarrow$ (ii). The proof of Theorem 2.1 actually shows this. (ii) $\Rightarrow$ (i). Using tiles with basepoints, if $[\partial R] \subseteq \mathbf{T}^+(\Sigma)$, then ∂R can be expressed as

$$\partial R = \prod_{i=1}^{k} W_i(\partial T_i) W_i^{-1},$$

from which a tiling of R can be directly read off, using winding numbers around cells of R. ∎

Now we give the proof that was deferred.

Proof of Theorem 5.1. (i) Let $\mathbf{C}_0$ consist of all words W such that $P(W)$ is a closed directed path in $\mathbb{Z}^2$. $\mathbf{C}_0$ is clearly a normal subgroup of $\mathbf{F}$.

We first show that $[\mathbf{F}:\mathbf{F}] \subseteq \mathbf{C}_0$. Expressing a word W in the generators A, U, A^{-1}, U^{-1} as a directed path it is easy to see this path is closed iff

$$\# \text{ occurrences}(A) = \# \text{ occurrences}(A^{-1}),$$

$$\# \text{ occurrences}(U) = \# \text{ occurrences}(U^{-1}).$$

All commutators $ABA^{-1}B^{-1}$ have this property, hence $[\mathbf{F}:\mathbf{F}] \subseteq \mathbf{C}_0$.

Now we prove $\mathbf{C}_0 \subseteq [\mathbf{F}:\mathbf{F}]$. Let W be a word representing an element of $\mathbf{C}_0$. We assign to each word W an invariant (n, k, l), where n is the length of the word, k is the maximum value $i^2 + j^2$ of any vertex $(i, j) \in \mathbb{Z}^2$ visited by the path $P(W)$, and l denotes the number of vertices (i, j) with $i^2 + j^2 = k$ (counted with multiplicity) that are visited by the path $P(W)$. Note that k and l are both less than n^2. We proceed by induction on triples (n, k, l) ordered lexicographically. The base case is $(0, 0, 0)$, which is the identity. For the induction step, if W contains any adjacent pairs of generators GG^{-1} we may cancel them and decrease its length. If this is not the case, the path $P(W)$ corresponding to W traverses no edge twice in succession. Let (i, j) be a vertex with $i^2 + j^2 = k$ visited by $P(W)$. If (i, j) is in the first quadrant, then either $W = W_2 A^{-1} U W_1$ with UW_1 visiting vertex (i, j) or $W = W_2 U^{-1} A W_1$ with AW_1 visiting vertex (i, j), as pictured in Fig. 5.2.

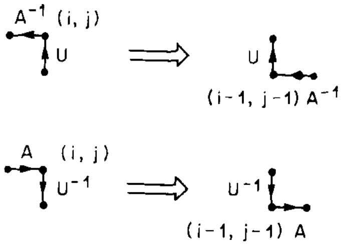


FIG. 5.2. Shortening a word in the first quadrant.

In the first case the word $\tilde{W} = W_2 U A^{-1} W_1$ has a lexicographically smaller value $(n, k, l-1)$ or $(n, k^1, *)$, and in the group **F** one has

$$W = (W_2 A^{-1} U A U^{-1} W_2^{-1}) \tilde{W},$$

where $W_2 A^{-1} U A U^{-1} W_2^{-1}$ is a conjugate of a commutator, so it is in $[\mathbf{F}:\mathbf{F}]$. By the induction hypothesis, $\tilde{W}$ is in $[\mathbf{F}:\mathbf{F}]$, hence so is W. In the second case above we use $\tilde{W} = W_2 A U^{-1} W_1$ and the same argument. Similar arguments work for (i, j) in the other three quadrants, completing the induction step, and proving $\mathbf{C}_0 = [\mathbf{F}:\mathbf{F}] = \mathbf{C}$.

(ii) Let $\mathbf{C}_1$ consist of all words W such that $P(W)$ is a closed path with winding number 0 about all cells. $\mathbf{C}_1$ is clearly a normal subgroup of **F**. We must show $\mathbf{C}_1 = [\mathbf{C}:\mathbf{C}]$.

We show $[\mathbf{C}:\mathbf{C}] \subseteq \mathbf{C}_1$. Since winding numbers are additive, if $W_1, W_2 \in \mathbf{C}$ then both they and their inverses correspond to closed paths, whence

$$\begin{aligned} &w(W_1 W_2 W_1^{-1} W_2^{-1}; c_{ij}) \\ &\quad = w(W_1; c_{ij}) + w(W_2; c_{ij}) + w(W_1^{-1}; c_{ij}) + w(W_2^{-1}; c_{ij}) = 0, \end{aligned}$$

for all cells c_{ij}.

We show $\mathbf{C}_1 \subseteq [\mathbf{C}:\mathbf{C}]$ by induction on the invariant (n, k, l) ordered as in the previous argument. The base case is the empty word, identified with the identity element of **F**. For the induction step, let W have value (n, k, l). If W contains any adjacent pairs of generators of the form GG^{-1}, we may cancel them and complete the induction step. Otherwise let (i, j) with $i^2 + j^2 = k$ be a vertex visited by the path corresponding to W. For the subsequent argument we relabel the cells so that c_{ij} denotes the cell whose vertex *furthest* from the origin $(0, 0)$ is (i, j). We examine all the visits of the path of W to (i, j). Suppose (i, j) is in the first quadrant. At each visit the path either arrives at this vertex from $(i, j-1)$ and exits to $(i-1, j)$ via $A^{-1}U$, or else arrives from $(i-1, j)$ and exits to $(i, j-1)$ via $U^{-1}A$, as in Fig. 5.2. Now we compute the winding number $w(W; c_{ij})$ using the argument principle, as in [15]. Since the path never crosses the line $i+j=k$, one has

$$w(W; c_{ij}) = \#\ \text{occurrences}(A^{-1}U) - \#\ \text{occurrences}(U^{-1}A),$$

where this sum is over visits to (i, j) only. Since this winding number is zero, there must be at least one visit of each kind, and one has $W = W_3 A^{-1} U W_2 U^{-1} A W_1$ or $W = W_3 U^{-1} A W_2 A^{-1} U W_1$, where W_1, W_2, W_3 are possibly empty words, and the path of W visits (i, j) in the middle of $A^{-1}U$ and of $U^{-1}A$. In the first case, let $\tilde{W} = W_3 U A^{-1} W_2 A U^{-1} W_1$, which has invariant either $(n, k, l-2)$ or $(n, k^1, *)$, and note that as words in F one has

$$W = (W_3 A^{-1} U W_2 U^{-1} A)(U A^{-1} W_2^{-1} A U^{-1} W_3^{-1}) \tilde{W}.$$

Calling the right side of this expression $Z\tilde{W}$, one finds after inserting suitable words of the form DD^{-1} that

$$Z = MNM^{-1}N^{-1},$$

where $M = W_3 A^{-1} U W_2 U^{-1} A W_3^{-1}$ and $N = W_3 U A^{-1} U^{-1} A W_3^{-1}$. Since M and N yield closed paths, it follows that Z is in $[\mathbf{C}:\mathbf{C}]$. By the induction hypothesis $\tilde{W}$ is in $[\mathbf{C}:\mathbf{C}]$, hence so is W. Similar arguments work in the second case $W = W_3 U^{-1} A W_2 A^{-1} U W_1$ and for (i, j) in the other three quadrants. This completes the induction step showing $\mathbf{C}_1 \subseteq [\mathbf{C}:\mathbf{C}]$, and (ii) is proved.

(iii) Define a homomorphism $\pi = \bigoplus_{i,j} \pi_{i,j}$ from $\mathbf{C}$ to $\bigoplus_{(i,j)} \mathbb{Z}$ by $\pi_{i,j} = w(P(W); c_{ij})$. This map is well defined by part (i) and its kernel is $[\mathbf{C}:\mathbf{C}]$ by part (ii). Hence its image is isomorphic to $\mathbf{C}/[\mathbf{C}:\mathbf{C}]$. ∎

References

1. L. V. Ahlfors, "Complex Analysis," 2nd ed., McGraw-Hill, New York, 1966.
2. R. Berger, The undecidability of the domino problem, *Mem. Amer. Math. Soc.* **66** (1966).
3. R. Brualdi and T. H. Foregger, Packing boxes with harmonic bricks, *J. Combin. Theory Ser. B* **17** (1974), 81–114.
4. R. Brualdi and T. H. Foregger, Some hypergraphs and packing problems associated with matrices of 0's and 1's, *J. Combin. Theory Ser. B* **17** (1974), 115–123.
5. N. G. de Bruijn, Filing boxes with bricks, *Amer. Math. Monthly* **76** (1964), 37–40.
6. H. S. M. Coxeter and W. O. J. Moser, "Generators and Relations for Discrete Groups," Springer-Verlag, Berlin, 1957.
7. M. Gardner, Mathematical games, *Sci. Amer.* **211** (1964), 124–130; **213** (1965), 96–104; **216** (1967), 124–132; **233** (1975), 112–117; **234** (1976), 122–140; **241** (1976), 119–123.
8. M. Garey and D. S. Johnson, "Computers and Intractability: A Guide to the Theory of NP-completeness," Freeman, San Francisco, 1979.

9. S. Golomb, Checker boards and polyominoes, *Amer. Math. Monthly* **61** (1954), 675–682.
10. S. Golomb, Covering a rectangle with L-tetrominoes, *Amer. Math. Monthly* **70** (1963), 760–761.
11. S. Golomb, Replicating figures in the plane, *Math. Gazette* **48** (1964), 403–412.
12. S. Golomb, "Polyominoes," Scribners, New York, 1965.
13. S. Golomb, Tiling with polyominoes, *J. Combin. Theory* **1** (1966), 280–296.
14. B. Grunbaum and G. C. Shepard, "Tilings and Patterns," Freeman, New York, 1987.
15. P. Henrici, "Applied and Computational Complex Analysis," Vol. I, Wiley, New York, 1974.
16. G. Katona and D. Szasz, Matching problems, *J. Combin. Theory* **10** (1971), 60–92.
17. J. B. Kelly, Polynomials and polyominoes, *Amer. Math. Monthly* **73** (1966), 464–471.
18. D. Klarner, A packing theory, *J. Combin. Theory* **8** (1970), 272–278.
19. D. Klarner, Brick-packing puzzles, *J. Recreational Math.* **6** (1973), 112–117.
20. D. Klarner and F. Göbel, Packing boxes with congruent figures, *Kon. Ned. Acad. Wetensch. Ser. A* **72** (1969), 465–472.
21. R. C. Lyndon and P. E. Schupp, "Combinatorial Group Theory," Springer-Verlag, New York, 1977.
22. W. Magnus, A. Karrass, and D. Solitar, "Combinatorial Group Theory," Interscience, New York, 1966 (Dover reprint 1976).
23. M. H. A. Newman, "Topology of Plane Sets of Points," Cambridge Univ. Press, Cambridge, 1951.
24. R. M. Robinson, Undecidability and nonperiodicity of tilings in the plane, *Inv. Math.* **12** (1971), 177–209.
25. W. Thurston, Conway's tiling groups, *Amer. Math. Monthly* **95** (1990), Special Geometry Issue, to appear.
26. D. Walkup, Covering a rectangle with T-tetrominoes, *Amer. Math. Monthly* **72** (1965), 986–988.

RIBBON TILE INVARIANTS FROM SIGNED AREA

Cristopher Moore

Computer Science Department and Department of Physics and
Astronomy, University of New Mexico, Albuquerque NM 87131
moore@cs.unm.edu

Igor Pak

Department of Mathematics, MIT,
Cambridge, MA 02139
pak@math.mit.edu

May 17, 2001

Abstract. Ribbon tiles are polyominoes consisting of n squares laid out in a path, each step of which goes north or east. Tile invariants were first introduced in [P1], where a full basis of invariants of ribbon tiles was conjectured. Here we present a complete proof of the conjecture, which works by associating ribbon tiles with certain polygons in the complex plane, and deriving invariants from the signed area of these polygons.

1. Introduction

Polyomino tilings have been an object of attention of serious mathematicians as well as amateurs for many decades [G]. Recently, however, the interest in tiling problems has grown as some important ideas and techniques have been introduced. In [P1], the second author introduced a *tile counting group*, which appears to encode a large amount of information concerning the combinatorics of tilings. He made a conjecture on the group structure, and obtained several partial results. A special case of the conjecture was later resolved in [MP]. In this paper we continue this study and complete the proof of the conjecture.

Consider the set of *ribbon tiles* $\mathbf{T}_n$, defined as connected n-square tiles with no two squares in the same diagonal $x + y = c$ (as in the figures below). It is easy to see that $|\mathbf{T}_n| = 2^{n-1}$, as each tile can be associated with a path of length $n - 1$ in the square lattice, each step of which goes east or north. Recording these moves by $\mathbf{0}$ and $\mathbf{1}$ respectively, we obtain a sequence $\varepsilon = (\varepsilon_1, \dots, \varepsilon_{n-1}) \in \{0, 1\}^{n-1}$, which uniquely encodes a ribbon tile. We will refer to this tile as τ_ε.

Key words and phrases. Polyomino tilings, tile invariants, Conway group, height representation.

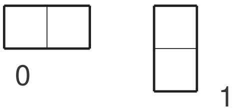


FIGURE 1. Two dominoes.

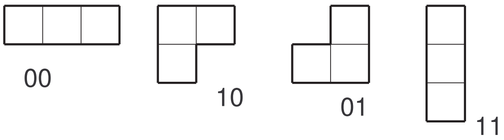


FIGURE 2. Four ribbon trominoes.

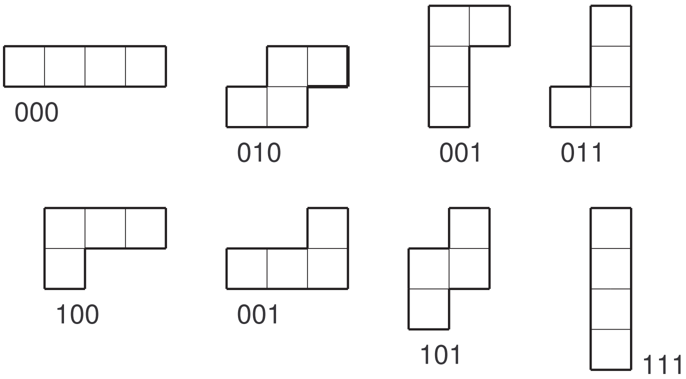


FIGURE 3. Eight ribbon tetrominoes.

Now, let Γ be a finite simply connected region, and let ν be a tiling of Γ by ribbon tiles in $\mathbf{T}_n$, $n \geq 2$. We denote by $a_\varepsilon(\nu)$ the number of times the ribbon tile τ_ε is used in ν.

Conjecture 1.1 [P1] *Let Γ and ν be as above. Then for every i, $1 \leq i < n/2$, we have:*

$$\sum_{\varepsilon:\ \varepsilon_i=0,\,\varepsilon_{n-i}=1} a_\varepsilon(\nu) \ - \sum_{\varepsilon:\ \varepsilon_i=1,\,\varepsilon_{n-i}=0} a_\varepsilon(\nu) \ = \ c_i(\Gamma),$$

where the $c_i(\Gamma)$ depend only on Γ and are independent of the tiling ν of Γ. Furthermore, when n is even, we have:

$$\sum_{\varepsilon:\ \varepsilon_{n/2}=1} a_\varepsilon(\nu) \ = \ c_*(\Gamma) \bmod 2,$$

where $c_*(\Gamma)$ is also independent of ν.

The main result of the paper is a proof of this conjecture for all $n \geq 2$:

Theorem 1.2 *Conjecture 1.1 holds for tilings by ribbon tiles $\mathbf{T}_n$ for all $n \geq 2$, and for all simply connected regions Γ.*

A few words about the history of this conjecture. For $n = 2$, it implies that for every domino tiling of Γ, the parity of the number of vertical dominoes is always the same. This, in fact, holds for every region, not just the simply connected ones, and follows from a folklore coloring argument (see [G,P1] for details).

For $n = 3$, the conjecture gives only one relation:

$$a_{\mathbf{01}}(\nu) - a_{\mathbf{10}}(\nu) = c_1(\Gamma).$$

This is the celebrated Conway-Lagarias relation for trominoes [CL]. Recently, the conjecture was established for $n = 4$ [MP], using a combinatorial technique similar to [CL]. In this notation, it was shown in [MP] that:

$$\begin{aligned} a_{\mathbf{001}} + a_{\mathbf{011}} - a_{\mathbf{101}} - a_{\mathbf{111}} &= c_1(\Gamma), \\ a_{\mathbf{010}} + a_{\mathbf{011}} + a_{\mathbf{110}} + a_{\mathbf{111}} &= c_*(\Gamma) \mod 2. \end{aligned}$$

It was shown in [CL], in a certain rigorous sense, that even for $n = 3$, the conjecture can't be proved by means of coloring arguments. This was extended by the second author to all $n \geq 4$ [P1]. It was observed in [P1], that for $n = 3$ there exists a non-simply connected region for which the relations in the conjecture do not hold. Thus, there is little hope of generalizing the conjecture to all regions.

The conjecture originated in [P1], where the author considered only row (or column) convex regions Γ, and proved the linear relations in Conjecture 1.1 for all such Γ [P1, Theorem 1.4]. The technique used a connection with combinatorics of Young tableaux which could not be extended to all simply connected regions (see [P1] for details). The author in [P1] also showed that the linear relations in the conjecture are the only relations which can occur between the $a_\varepsilon(\nu)$, even for this smaller set of regions (see section 2 below).

About the proof technique: We use notion of tile invariants, introduced in [P1], but here we define new real-valued invariants, which we call *adèle invariants.* As it turns out, these invariants imply all the integer-valued invariants that we need to establish. We then show the validity of the adèle invariants by presenting them as a signed area of a certain polygon corresponding to each tile. These two results together imply Theorem 1.2.

The rest of the paper is structured as follows. In section 2 we introduce tile invariants and compute the tile counting group based on Theorem 1.2. Much of the material follows [P1], so we present only sketches of the proofs for completeness. In section 3, we define and study the adèle invariants. Small examples are computed in section 4. We exhibit the relationship between the adèle invariants and integer invariants in section 5. This completes the proof of Theorem 1.2. We conclude with final remarks in section 6.

2. Tile invariants

Let us start by defining tilings and tile invariants. Let Λ be a set of (closed) squares of a square grid $\mathbb{Z}^2$ on a plane. A *region* is a finite subset $\Gamma \subset \Lambda$. Region $\Gamma \subset \Lambda$ is called simply connected if its boundary $\partial\Gamma$ is connected. We say that two regions Γ and Γ' are *equivalent*, denoted $\Gamma \sim \Gamma'$, if Γ is a parallel translation of Γ' (rotations and reflections are not allowed). Let $\widetilde{\Gamma} = \{\Gamma' : \Gamma' \sim \Gamma\}$ be the set of regions equivalent to Γ.

Let $\mathbf{T} = \{\tau_1, \dots, \tau_r\}$ be a finite set of simply connected regions, which we call *tiles*. By $\widetilde{\tau}_i$ we denote the set of their parallel translations, and let $\widetilde{\mathbf{T}} = \cup_i \widetilde{\tau}_i$. A *tiling* ν of Γ, denoted $\nu \vdash \Gamma$, is a set of tiles $\tau \in \widetilde{\mathbf{T}}$, such that their disjoint union is Γ :

$$\Gamma = \bigsqcup_{\tau \in \nu} \tau.$$

Here we ignore the intersection of the boundaries.

Let G be an abelian group, and let $\varphi : \mathbf{T} \to G$ be any map. We extend the definition of φ to all $\tau \in \widetilde{\mathbf{T}}$, by setting $\varphi(\tau) = \varphi(\tau_i)$ for all $\tau \sim \tau_i$. We say that the map φ is a *tile invariant* of $\mathbf{T}$ if, for every simply connected region Γ and every tiling $\nu \vdash \Gamma$ by the set of tiles $\mathbf{T}$, we have:

$$\sum_{\tau \in \nu} \varphi(\tau) = c(\Gamma),$$

where the constant on the r.h.s. depends only on the region Γ and is independent of ν. In this paper G is either $\mathbb{Z}$, or $\mathbb{Z}_n (= \mathbb{Z}/n\mathbb{Z})$, or $\mathbb{R}$ (with addition as the group operation).

Tile invariants are directly related to numerical relations between the respective numbers of times differently-shaped tiles occur in a tiling. Indeed, let $a_i(\nu) = |\nu \cap \widetilde{\tau}_i|$ be the number of tiles $\tau \sim \tau_i$ in the tiling $\nu \vdash \Gamma$. We immediately have:

$$\sum_{i=1}^{r} \varphi(\tau_i)\, a_i(\nu) = \sum_{\tau \in \nu} \varphi(\tau) = c(\Gamma).$$

In [P1], we introduced a *tile counting group* $\mathbb{G}(\mathbf{T})$, which is defined as a quotient:

$$\mathbb{G}(\mathbf{T}) = \mathbb{Z}^r / \big\langle \big(a_1(\nu) - a_1(\nu'), \dots, a_r(\nu) - a_r(\nu')\big),\ \nu,\, \nu' \vdash \Gamma \big\rangle,$$

where ν, ν' are tilings of the same simply connected region Γ by the set of tiles $\mathbf{T}$. Computing the tile counting group $\mathbb{G}(\mathbf{T})$ is a difficult task, even in simple cases. The main result of this paper is a computation of $\mathbb{G}(\mathbf{T}_n)$ for the case of ribbon tiles:

Theorem 2.1 *If $n = 2m+1$, then $\mathbb{G}(\mathbf{T}_n) \simeq \mathbb{Z}^{m+1}$. If $n = 2m$, then $\mathbb{G}(\mathbf{T}_n) \simeq \mathbb{Z}^m \times \mathbb{Z}_2$.*

Theorem 2.1 was stated as a conjecture in [P1]. It was shown in [P1] that it follows from Theorem 1.2. For completeness, we sketch the proof below.

Sketch of proof. Indeed, in [P1, Theorem 1.4] it was shown that $\mathbb{G}(\mathbf{T}) \subset \mathbb{Z}^{m+1}$ for $n = 2m + 1$ and $\mathbb{G}(\mathbf{T}_n) \subset \mathbb{Z}^m \times \mathbb{Z}_2$ for $n = 2m$. Observe that one can view the relations in Conjecture 1.1 as elements of $\mathbb{G}(\mathbf{T}_n)$. Recall that these relations, together with the trivial *area invariant* f_0 (defined by $f_0(\tau) = 1$ for all $\tau \in \mathbf{T}_n$), are independent in $\mathbb{Z}^n$ (see the proof of Theorem 1.4 in [P1, § 5]). Now Theorem 1.2 implies the result. □

Before we conclude this section, let us make a final observation on the relations in Conjecture 1.1 implied by previous work. Following [P1, § 9], define the *shade invariant* as follows:

$$f_{\blacktriangledown}(\tau_\varepsilon) \, = \, \sum_{k=1}^{n-1} k \cdot \varepsilon_k \mod n,$$

where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_{n-1})$. The fact that it is an invariant follows easily from an extended coloring argument [P1, § 9]. Namely, consider a coloring of the squares $\zeta : \mathbb{Z}^2 \to \mathbb{Z}_n$ defined by $\zeta(x, y) = y \mod n$. Note that the sum of the colors in each ribbon tile τ is equal to $f_{\blacktriangledown}(\tau) + C$, where $C = C(n) \in \mathbb{Z}_n$ is a constant which depends only on n. We omit the (easy) details.[1]

Proposition 2.3 *When n is even, the relations in the first part of Conjecture 1.1 imply that in the second part.*

Proof. We will show that the mod 2 relation follows from the $m = n/2$ relations in the first part, and the shade invariant. In the language of invariants, consider the *k-convexity invariants* f_k, introduced in [P1] :

$$f_k(\tau_\varepsilon) = \varepsilon_k - \varepsilon_{n-k}, \text{ where } \varepsilon = (\varepsilon_1, \dots, \varepsilon_{n-1}).$$

We need to show that the shade invariant and the k-convexity invariants generate the *parity invariant* f_*:

$$f_*(\tau_\varepsilon) = \varepsilon_m \mod 2, \text{ where } n = 2m.$$

But this is immediate since

$$f_{\blacktriangledown} \mod 2 = \big(f_1 + 2f_2 + \, \dots \, + (m-1)f_{m-1}\big) + f_* \mod 2$$

(cf. [P1, § 9]). This completes the proof. □

3. New ribbon tile invariants and the signed area

Let $\mathbf{T}_n$ be the set of ribbon tiles, defined as above. From now on, we will also use a different encoding of $\mathbf{T}_n$, by sequences $\alpha = (\alpha_1, \dots, \alpha_n) \in \{\pm 1\}^{n-1}$: $\mathbf{T}_n = \{\tau_\alpha\}$, where $\tau_\alpha = \tau_\varepsilon$, if $\alpha_i = 1 - 2\varepsilon_i$ for all $1 \le i \le n-1$ (i.e. $\mathbf{0} \to +1$ and $\mathbf{1} \to -1$).

[1] In contrast with other ribbon tile invariants we introduce, the shade invariant can be extended to *all* regions, not just the simply connected ones [P1, Theorem 9.1].

For every $1 \leq \ell < n$ we define a function $\Phi_\ell : \mathbf{T}_n \to \mathbb{R}$ as follows:

$$\Phi_\ell(\tau_\alpha) = \sum_{k=1}^{n-1} \alpha_k \, \sin \frac{2\pi k \, \ell}{n},$$

where $\alpha = (\alpha_1, \dots, \alpha_{n-1})$, $\alpha_k \in \{\pm 1\}$ as above. The main result of this section is the following key observation:

Theorem 3.1 *The function $\Phi_\ell : \mathbf{T}_n \to \mathbb{R}$ is a tile invariant for the set $\mathbf{T}_n$ of ribbon tiles, for all $1 \leq \ell < n$.*

We will call Φ_ℓ the *ℓ-th adèle invariant.* Note that when $n = 2m$, we have $\Phi_m(\tau_\alpha) = 0$ for all $\tau_\alpha \in \mathbf{T}_n$. The claim of the theorem is trivial in this case.

The proof of Theorem 3.1 is based on a new geometric construction. But first we need several definitions.

Let the squares of the grid have numbers written on them, from 0 to $n-1$, with the rule that $(x, y) \in \mathbb{Z}^2$ has the number $x + y \bmod n$. Let us orient edges of the grid eastward and southward as in figure 1 below. Set labels on the edges so that the edge between square k and $(k + 1 \bmod n)$ has label k.

Let $\ell \neq n/2$ be fixed for the rest of this section. On a complex plane $V = \mathbb{C}$, fix n vectors $v_0, v_1, \dots, v_{n-1}$, where $v_k = e^{2\pi i k \ell / n}$. We say that a loop in V is a *polygon* if it is a closed (perhaps self-intersecting) path with straight edges.

Now, let Γ be a simply connected region on a grid, and let $\partial\Gamma$ be the boundary of Γ. Fix any integer point $O \in \partial\Gamma$. Consider a sequence of edges on the grid obtained by moving along $\partial\Gamma$ counterclockwise, starting at O. Recall that these edges are oriented and labeled with integers modulo n.

We shall describe a map $\eta = \eta_\ell$, which maps simply connected regions Γ, tileable by $\mathbf{T}_n$, into polygons in V. First, fix any $O' \in V$. As one moves along the sequence of edges of $\partial\Gamma$, add a vector $\pm v_j \in V$, where j is a label of the edge in $\partial\Gamma$, and a sign $\pm$ is chosen depending on whether the edge in $\partial\Gamma$ is oriented counterclockwise or clockwise (see figures below). We denote the resulting path by $\eta(\gamma) = \eta_\ell(\Gamma)$. Note that it already has an induced orientation.

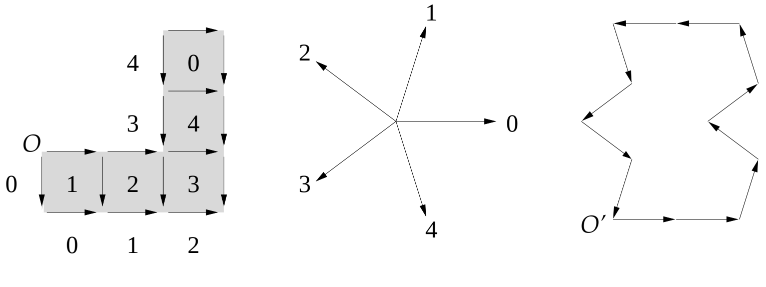


FIGURE 4. The ribbon tile $\tau = \tau_{\mathbf{0011}}$ with labels on the edges, the roots of unity $v_0, \dots, v_4$, $v_k = e^{2\pi i k/5}$, and the closed loop $\eta_1(\tau)$.

In the figure above we present a **V**-pentomino (cf. [G]), which is encoded by $\alpha = (+1, +1, -1, -1)$ in our notation, along with 5 vectors $v_0, \dots, v_4$, and the corresponding polygon. Note that a priori, it is unclear whether our map is well-defined, i.e. whether all tileable regions correspond to closed loops in V. By definition, $\eta(\Gamma)$ is only a path starting at O', with straight edges.

Lemma 3.2 *The above map η_ℓ is well-defined, i.e. for any simply connected region Γ tileable by $\mathbf{T}_n$, the path $\eta_\ell(\Gamma)$ is a closed loop in V.*

Proof. We prove the result by induction on the area of Γ. Suppose τ is one of the ribbon tiles and let $(k + 1 \bmod n)$ be the label of the square in the lower left corner. Let O be the point in the upper left corner of this square. Now observe that the sequence of edges in $\partial\tau$ has two labels k, then a sequence w of labels, then two labels k, and then the same sequence as w but in the opposite order. Observe also that the first two edges, with the label k, are directed counterclockwise, while the second two are clockwise. This implies that the pieces of $\eta(\tau)$, corresponding to these four edges, form two straight parallel intervals oriented in opposite directions.

Note also, that each edge in the first sequence w has an orientation which is *opposite* to that of a corresponding edge in the second (reversed) w. Therefore, the pieces corresponding to the two w are exactly parallel to each other, with a shift of $2v_k$. We conclude that $\eta(\tau)$ is a closed loop in V, so η is well-defined for ribbon tiles. This proves the base of our induction.

The induction step is straightforward. Let Γ be a region tileable by $\mathbf{T}_n$. Fix any tiling of Γ. Consider a tile τ in the tiling such that $\Gamma' = \Gamma \setminus \tau$ is simply connected. In [MP, Lemma 2.1] we prove that there always exists such a tile[2]. Now present $\partial\Gamma$ as a union of two regions, $\partial\Gamma'$ and $\partial\tau$ (intersections of these will cancel each other as they have opposite orientations). If both $\eta(\Gamma')$ and $\eta(\tau)$ are closed, then $\eta(\Gamma)$ is also closed. This completes the proof. □

Let us present now a standard inductive definition of a *signed area* $A(\gamma)$ of an oriented polygon γ in V (see e.g. [GO]). If γ is not self-intersecting, define $A(\gamma)$ to be the usual area times ± 1 depending on whether γ is oriented counterclockwise or not. If γ is self-intersecting at point x, split γ into the disjoint union of two γ_1 and γ_2 (separated by the point x), and let $A(\gamma) = A(\gamma_1) + A(\gamma_2)$.

Now let Γ be a region tileable by $\mathbf{T}_n$. Let us show that for any ℓ, the signed area of $\gamma = \eta_\ell(\Gamma)$ is invariant under parallel translation of Γ (recall that the construction of η_ℓ involves a fixed labeling of the plane, so a priori it may differ for $\Gamma' \sim \Gamma$). Indeed, observe that for a parallel translation $\Gamma' \sim \Gamma$, we have a cyclic shift of the labels of the edges in $\partial\Gamma'$. Therefore $\eta_\ell(\Gamma')$ is simply a rotation of $\eta_\ell(\Gamma)$ by a multiple of $\frac{2\pi\ell}{n}$. Thus these two loops have the same signed area $A\big(\eta_\ell(\Gamma')\big) = A\big(\eta_\ell(\Gamma)\big)$. Similarly, the choice of the starting point $O \in \partial\Gamma$ (and $O' \in V$) doesn't change the signed area of γ. We shall prove now that there exists a closed formula for $A(\gamma)$ when Γ is a ribbon tile.

[2] Versions of this result were also used in [CL,Pr].

Proposition 3.3 *Let* $\gamma = \eta_\ell(\tau_\alpha)$*, where* $\alpha = (\alpha_1, \dots, \alpha_{n-1}) \in \{\pm 1\}^{n-1}$*. Then*

$$A(\gamma) \, = \, 2 \sum_{k=1}^{n-1} \alpha_k \, \sin \frac{2\pi k \, \ell}{n}.$$

Proof. This follows immediately from the analysis used in the induction step in the proof of Lemma 3.2. Indeed, let us translate the tile τ so that the lower left square has label 1. Also, choose point $O \in \partial\tau$ as in the proof above. Recall that the signed area remains unchanged. Observe that the signed area is exactly the area of the parallelogram whose vertices are the endpoints of two horizontal intervals of length 2. Therefore $A(\gamma) = 2 \cdot$ *height*, where *height* is the height of the image of a sequence of labels w, defined as in the proof above. Now, the height of the image of w is the sum of the heights of each of the vectors v_k, taken with a sign α_k, for $k = 1, \dots, n-1$. This implies the formula in the proposition. □

Proposition 3.4 *Let* $\nu \vdash \Gamma$ *be a tiling of* Γ *by ribbon tiles in* $\mathbf{T}_n$*. Then*

$$\sum_{\tau \in \nu} A\bigl(\eta_\ell(\tau)\bigr) \, = \, A\bigl(\eta_\ell(\Gamma)\bigr), \quad \textit{for all } 1 \le \ell \le n-1.$$

Proof. This is an immediate corollary of the induction step in the proof of Lemma 3.2. Indeed, let us prove the claim by induction on the area of Γ. The claim is trivial when $\Gamma = \tau \in \mathbf{T}_n$.

Now, by construction, γ is a union of γ_1 and γ_2, where $\gamma = \eta_\ell(\Gamma)$, $\gamma_1 = \eta_\ell(\Gamma')$, and $\gamma_2 = \eta_\ell(\tau)$. By definition, this implies that $A(\gamma) = A(\gamma_1) + A(\gamma_2)$. This completes the inductive step and finishes the proof. □

Proof of Theorem 3.1 This is a corollary of Propositions 3.3 and 3.4. Indeed, Proposition 3.4 implies that

$$\Phi_\ell(\tau) = \frac{1}{2} A\bigl(\eta_\ell(\tau)\bigr)$$

for every ribbon tile $\tau \in \mathbf{T}_n$, and every $1 \le \ell \le n-1$. Now Proposition 3.4 implies that Φ_ℓ satisfies the definition of a tile invariant. □

4. Examples

Let $n = 3$. In the Figure 4 below, we show all four ribbon trominoes τ_α, along with the corresponding polygons $\eta_1(\tau_\alpha) \in V$. Let us calculate the values of the adèle invariant Φ_1. Consider the straight trominoes first. Observe that the signed area of the corresponding polygons is zero. Indeed, the two equilateral triangles cancel each other, since we circle one equilateral triangle clockwise and the other counterclockwise. On the other hand, for the two right trominoes the adèle invariant $\Phi_1 = \pm\sqrt{3}$. Indeed, in both cases these polygons circle eight equilateral triangles,

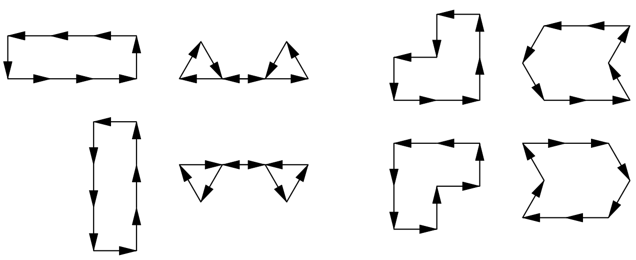


FIGURE 4. Four ribbon trominoes τ_α and the corresponding closed loops $\eta_1(\tau_\alpha)$.

in the first case counterclockwise and in the other clockwise. Thus the signed area is $A = \pm 8\,\frac{\sqrt{3}}{4} = \pm 2\sqrt{3}$, which implies the claim.

Now observe that $\frac{1}{\sqrt{3}} \cdot \Phi_1$ coincides with the Conway-Lagarias invariant (see section 1). This gives a new interpretation of this remarkable invariant in terms of an "area," rather than the "winding number" as defined in [CL].

Let us note here that for $n = 3, 4$ the group of translations of $V = \mathbb{C}$ by integer linear combinations of vectors v_i is a lattice in V. Thus the corresponding polygons $\eta(\tau)$ have a natural combinatorial group structure and can be described by the technique of [CL]. However, for other values of n these vectors do not form a lattice, and instead form a dense set in the plane. This explains the reason why [MP] were able to completely resolve the case $n = 4$, and why the case $n = 5$ has remained mysterious until now. (We note that signed area on the square grid is used to study other tetrominoes in [Pr].)

Consider the case $n = 5$. Let us calculate the adèle invariant of several ribbon pentominoes. First, let τ be the **V**-pentomino, which corresponds to $\alpha = (+1, +1, -1, -1)$. We have

$$\Phi_1(\tau) = \frac{1}{2}\,A\bigl(\eta_1(\tau_\alpha)\bigr) = \sin\frac{2\pi}{5} + \sin\frac{4\pi}{5} - \sin\frac{6\pi}{5} - \sin\frac{8\pi}{5}$$
$$= 2\sin\frac{2\pi}{5} + 2\sin\frac{4\pi}{5} = \sqrt{\frac{5+\sqrt{5}}{2}} + \sqrt{\frac{5-\sqrt{5}}{2}}.$$

The same calculation can be done for all remaining ribbon pentominoes. For example, for **I**- and **Z**-pentominoes, which correspond to $(+1, +1, +1, +1)$ and $(-1, +1, +1, -1)$, all adèle invariants are zero. In general, we have:

Proposition 4.1 *Let τ be a ribbon tile with a $180°$ rotational symmetry. Then $\Phi_\ell(\tau) = 0$ for all $1 \le \ell \le n-1$.*

Proof. Having 180° symmetry implies that $\alpha_k = \alpha_{n-k}$ for all $k < \frac{n}{2}$. On the other hand, we have $\sin\frac{2\pi k\,\ell}{n} = -\sin\frac{2\pi(n-k)\,\ell}{n}$, i.e. all the sign terms in the expression for $\Phi_\ell(\cdot)$ cancel each other. This implies the result. □

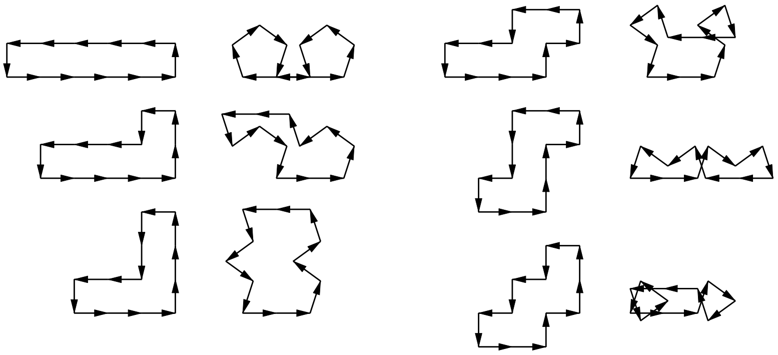


FIGURE 5. Several ribbon pentominoes τ_α and the corresponding closed loops $\eta_1(\tau_\alpha)$. The remaining ribbon pentominoes, as well as the corresponding closed loops, can be obtained from these by rotation, reflection, etc.

Before we conclude, let us state two possible ways of deriving the linear relations in Conjecture 1.1 from adèle invariants.

We consider only the case $n = 5$. Recall that $\sin\frac{\pi}{5}$ and $\sin\frac{2\pi}{5}$ are rationally independent. Observe that for all regions Γ tileable by $\mathbf{T}_5$, we have:

$$(\diamond) \qquad \Phi_1(\Gamma) \, = \, -2c_1 \, \sin\frac{2\pi}{5} \, - \, 2c_2 \, \sin\frac{4\pi}{5},$$

where $c_1 = c_1(\Gamma)$ and $c_2(\Gamma)$ are as in Conjecture 1.1. Indeed, this holds for all ribbon tiles $\tau \in \mathbf{T}_5$, and thus by additivity for all tileable simply connected regions Γ. Since c_1 and c_2 are integers, by rational independence, the adèle invariant then induces two integer-valued invariants.

Another approach is based on using both Φ_1 and Φ_2. We have:

$$(\diamond\,\diamond) \qquad \Phi_2(\tau) \, = \, -2c_1 \, \sin\frac{4\pi}{5} \, + 2\,c_2 \, \sin\frac{2\pi}{5}.$$

We can write both $(\diamond)$ and $(\diamond\,\diamond)$ as

$$\bigl(\Phi_1, \Phi_2\bigr) \, = \, -2\,\bigl(c_1, c_2\bigr) \begin{pmatrix} \sin\frac{2\pi}{5} & \sin\frac{4\pi}{5} \\ \sin\frac{4\pi}{5} & -\sin\frac{2\pi}{5} \end{pmatrix}.$$

Since the matrix on the r.h.s. is invertible, we can obtain c_1 and c_2 as a linear combination of Φ_1, Φ_2 (the same for every tile $\tau_\alpha \in \mathbf{T}_5$).

We will show in the next section that we can generalize this argument for any n, and prove Theorem 1.2.

5. Proof of Theorem 1.2

Let $n = 2m + 1$ be an odd integer, $n \geq 3$. We claim that in this case the functions $\Phi_\ell(\tau)$, $1 \leq \ell \leq m$, are linearly independent (as real functions on $\mathbf{T}_n$).

Similarly, when $n = 2m$ is an even integer, the functions $\Phi_\ell(\tau)$, $1 \leq \ell < m$, are linearly independent (note that $\Phi_m \equiv 0$ in this case). Let us state this as follows:

Lemma 5.1 *For all* n*, we have* $\dim \langle \Phi_1, \dots, \Phi_m \rangle = m$*, where* $m = \lfloor \frac{(n-1)}{2} \rfloor$.

Proof of Theorem 1.2. By Proposition 2.3, it suffices to prove only the first part of Conjecture 1.1. We claim that this part follows from Lemma 5.1. Indeed, let $W = \langle f_1, \dots, f_m \rangle$, where f_k is a k-convexity invariant defined in the proof of Proposition 2.3.

Using $\sin 2\pi k\ell/n = -\sin 2\pi(n-k)\ell/n$, we can rewrite the ℓ-th adèle invariant as follows:

$$\begin{aligned}\Phi_\ell(\tau_\alpha) &= \sum_{k=1}^{m} \big(\alpha_k - \alpha_{n-k}\big) \sin \frac{2\pi k\,\ell}{n} = -2 \sum_{k=1}^{m} \big(\varepsilon_k - \varepsilon_{n-k}\big) \sin \frac{2\pi k\,\ell}{n} \\ &= -2 \sum_{k=1}^{m} f_k \sin \frac{2\pi k\,\ell}{n},\end{aligned}$$

where $\alpha = (\alpha_1, \dots, \alpha_{n-1}) \in \{\pm 1\}^{n-1}$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_{n-1}) \in \{0,1\}^{n-1}$, $\alpha_k = 1 - 2\varepsilon_k$, for all $1 \leq k \leq n-1$ (so that $\tau_\alpha = \tau_\varepsilon$). This implies that $\Phi_\ell \in W$. From Lemma 5.1 we obtain:

$$m = \dim \langle \Phi_1, \dots, \Phi_m \rangle \leq \dim \langle f_1, \dots, f_m \rangle = \dim W \leq m,$$

and therefore $\langle \Phi_1, \dots, \Phi_m \rangle = W$. We conclude $f_k \in \langle \Phi_1, \dots, \Phi_m \rangle$ for all $1 \leq k \leq m$. The linearity of tile invariants implies that f_k is a tile invariant of the set $\mathbf{T}_n$ of ribbon tiles (cf. proof of Proposition 2.3). This completes the proof of Theorem 1.2. □

Proof of Lemma 5.1 Suppose $n = 2m+1$ is odd. Consider two $n \times n$ matrices $X = (x_{k,\ell})$, $Y = (y_{k,\ell})$, $0 \leq k, \ell \leq n-1$, defined as follows:

$$x_{k,\ell} = \cos \frac{2\pi k\,\ell}{n}, \quad y_{k,\ell} = \sin \frac{2\pi k\,\ell}{n}.$$

Since $Z = X + i \cdot Y$ is a Vandermonde matrix $Z = (z_{k,\ell})$, $z_{k,\ell} = \exp(2\pi i k\,\ell/n)$, we immediately have:

$$\det(Z) = \prod_{0 \leq k < \ell \leq n-1} \left(e^{2\pi i k/n} - e^{2\pi i \ell/n}\right) \neq 0.$$

Thus $\operatorname{rk}(Z) = n$.

From $y_{k,\ell} = -y_{n-k,\ell}$, $y_{0,\ell} = 0$, and $x_{k,\ell} = x_{n-k,\ell}$, we obtain $\operatorname{rk}(Y) \leq m$, and $\operatorname{rk}(X) \leq m+1$. Since $2m+1 = \operatorname{rk}(Z) = \operatorname{rk}(X+iY) \leq \operatorname{rk}(X) + \operatorname{rk}(Y)$, we immediately have $\operatorname{rk}(Y) = m$. From $y_{k,\ell} = -y_{k,n-\ell}$, $1 \leq \ell \leq m$, we conclude that an $m \times (n-1)$ submatrix $Y' = (x_{k,\ell})$, where $1 \leq k \leq n-1$, $1 \leq \ell \leq m$, has rank $\operatorname{rk}(Y') = m$. One can think of $\alpha \in \{\pm 1\}^{n-1}$ as vectors $\mathbb{R}^{n-1}$. Since

$$\big(\Phi_1(\tau_\alpha), \dots, \Phi_m(\tau_\alpha)\big) = (\alpha) \cdot Y'$$

and $\dim\langle\alpha\rangle = n-1$, we get $\dim\langle\Phi_1,\dots,\Phi_m\rangle = m$.

When $n = 2m$, the proof follows verbatim, except that in this case $y_{m,\ell} = \pm y_{0,\ell} = \pm 1$ (depending on the parity of ℓ). Then $\mathrm{rk}(X) = \mathrm{rk}(Y) = m$, and the result follows. □

6. Final remarks

The main result in this paper can be viewed as an existence of a large number of invariants for tilings by ribbon tiles. Still, the source of these invariants remains something of a mystery, yet to be discovered. It seems that such a rich structure of invariants is an exception rather than the rule, and these sets of tiles enjoy some special properties others do not. In this section we shall speculate on the possible explanations for these questions.

Let us start by saying, that although we do not pursue here the 'rational independence' approach (see section 4), it can in fact be used. In fact, it is quite straightforward for prime n, while for composite n one has to employ Φ_d, for each $d\,|\,n$ and Möbius inversion. In the original version of the paper the authors favored this idea, while at the end we chose to employ an elementary linear algebra approach. Let us mention here that the arguments in section 5, while elementary, were influenced by the ideas in [BF]. As the referee pointed out, one can think of the proof as an application of the discrete Fourier transform.

We shall note here, that miraculously, for any n, the real-valued tile invariant Φ_1 already induces a large number $\phi(n) = \Omega(n/\log\log n)$ of linearly independent integer-valued ribbon tile invariants. It would be interesting to find other examples of this phenomenon.

Let us now state the following conjecture, which seems more plausible now in view of Theorem 1.2.

Conjecture 6.1 [P1] *Define* 2*-flips to be transformations of tilings by* $\mathbf{T}_n$ *which involve exactly two tiles. Then for any simply connected region* Γ*, and any two tilings* ν*,* ν' *of* Γ*, there is a sequence of* 2*-flips which moves* ν *into* ν'*.*

The are several reasons behind this conjecture. For $n = 2$ the truth of the assertion is well known (see e.g. [T1]). For $n \geq 3$ it has been established when Γ has the shape of a Young diagram [P1] or skew Young diagram [P2]. For $n = 3$ it was also proved by an ad hoc argument for a very special set of regions [W]. There is also a topological reason in favor of the conjecture [T2]. Perhaps the most compelling reason,[3] however, is given by the following result:

Proposition 6.2 [P1] *Conjecture 6.1 implies Theorem 1.2.*

Indeed, assume the conjecture. Then to prove Theorem 1.2 one needs only to check that the invariants are preserved along the 2-flips. As the structure of the flips is known, this is straightforward. We refer to [P1] for details.

[3] As the referee validly points out, this is rather a reason for *wishing* that Conjecture 6.1 were true. While we agree, we leave the final judgement to the reader.

To conclude, let us speculate on how Conjecture 6.1 can be proved. The most promising and relevant method seem the "height representation" approach, pioneered in this context by Thurston [T1].[4] In view of importance of the subject, let us elaborate on this.

A *height representation* is a way of assigning a height to each site in the lattice so that a given tiling corresponds to a surface, i.e. a function from the lattice to the space in which the heights take their values. While the best-known height representations are integer-valued, in general they can be two- or more-dimensional vectors, or elements of a non-Abelian group (see [K,KK,MP,Pr,T1].

Height representations have many uses. If one desires to sample randomly from the set of tilings of a given simply connected region, these representations can be used to prove that this set is connected under some set of local moves [K,R], to devise exact sampling Monte Carlo algorithms based on these moves [PW], and to place upper limits on the mixing time of these algorithms [LRS]. They can also be used to develop an efficient algorithm to tell whether a given region can be tiled at all [K,R], which is interesting since this problem is NP-complete in general, even for some simple sets of tiles (see e.g. [MR]).

For tilings, the standard approach is to define how the height changes, by small increments, as we move along the boundary between one tile and another. In order for the height to be a single-valued function, it must return to its original value whenever we travel around a loop. Therefore, each type of tile induces a relation in the height group [CL,T1], or, in the Abelian case, a linear constraint on the amount by which the height increases or decreases as we traverse different kinds of edges.

For instance, domino tilings of the square lattice have a height representation which can be thought of as follows. We color the lattice as a checkerboard, with white and black squares alternating. Whenever we move along an edge of the lattice, we change the height by $+1$ if the square on our left is black, and -1 if it is white. The reader can easily check that a set of moves encircling a horizontal or vertical domino will have a total height change of $+1+1+1-1-1-1=0$. In fact, this is our mapping η in the case $n=2$. We refer the reader to [KK,R,T1] for other examples and details.

Now consider what happens in our case. We define a complex-valued height function which is defined by local rules. It seem likely that our height function is a projection onto two dimension of the height function with values in an n-dimensional lattice [T2], but we were unable to make this observation precise. If only we could show a "nice" behavior under 2-flips, we would be able to prove Conjecture 6.1 and perhaps even give a linear time algorithm for checking tileability by ribbon tiles. So far, this remains a fantasy, so we leave the reader here until further developments.

Acknowledgments

We would like to thank Jim Propp and an anonymous referee for helpful remarks and pointing out errors in the original version of the paper. Discussions with Ezra Miller and Ravi Vakil helped to simplify the proof.

[4] Interestingly, Thurston's paper [T1] was inspired by [CL].

The first author thanks the Sandia University Research Program. The second author thanks the Santa Fe Institute for the hospitality during a visit in which this work was done. The first author thanks Molly Rose and Spootie the Cat. The second author was supported by an NSF Mathematical Research Fellowship.

References

[BF] L. Babai, P. Frankl, *Linear Algebra Methods in Combinatorics, with Applications to Geometry and Computer Science* (Preliminary version 2), University of Chicago Preprint, 1992, 216 pp.

[CL] J. H. Conway, J. C. Lagarias, *Tilings with polyominoes and combinatorial group theory*, J. Comb. Theory, Ser. A **53** (1990), 183–208.

[G] S. Golomb, *Polyominoes*, Scribners, New York, 1965.

[GO] J. E. Goodman, J. O'Rourke, editors, *Handbook of Discrete and Computational Geometry*, CRC Press, Boca Raton, FL, 1997.

[KK] C. Kenyon, R. Kenyon, *Tiling a polygon with rectangles*, Proc. 33rd Symp. Foundations of Computer Science (1992), 610-619.

[K] R. Kenyon, *A note on tiling with integer-sided rectangles*, J. Combin. Theory, Ser. A **74** (1996), 321–332.

[LRS] M. Luby, D. Randall, A. Sinclair, *Markov chain algorithms for planar lattice structures*, Proc. 36th Symp. Foundations of Computer Science (1995), 150–159.

[MR] C. Moore, J.M. Robson, *Hard tiling problems with simple tiles*, preprint, Santa Fe Institute (2000).

[MP] R. Muchnik, I. Pak, *On tilings by ribbon tetrominoes*, J. Combin. Theory, Ser. A **88** (1999), 199–193.

[P1] I. Pak, *Ribbon tile invariants*, Trans. AMS **352** (2000), 5525–5561.

[P2] I. Pak, unpublished.

[Pr] J. Propp, *A pedestrian approach to a method of Conway, or, A tale of two cities*, Math. Mag. **70** (1997), 327–340.

[PW] J. Propp and D. Wilson, *Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics*, Random Structures and Algorithms **9** (1996), 223–252.

[R] E. Rémila, *Tiling groups: new applications in the triangular lattice*, Discrete and Combinatorial Geometry **20** (1998), 189-204.

[T1] W. Thurston, *Conway's tiling groups*, Amer. Math. Monthly **97** (1990), 757–773.

[T2] W. Thurston, Personal communication (2000).

[W] D. C. West, *An elementary proof of two triangle-tiling theorems of Conway and Lagarias*, unpublished manuscript (1990), 6 pp.

Journal of Combinatorial Theory, Series A **88**, 188–193 (1999)
Article ID jcta.1999.2980, available online at http://www.idealibrary.com on IDEAL®

NOTE

On Tilings by Ribbon Tetrominoes[1]

Roman Muchnik and Igor Pak[2]

Department of Mathematics, Yale University, New Haven, Connecticut 06520
E-mail: {roma,paki}@math.yale.edu

Communicated by the Managing Editors

Received July 18, 1998

1. INTRODUCTION

A *ribbon polyomino* is a polyomino which has at most one square (i, j) in every diagonal $i-j=c$. A tetromino is a polyomino with four squares. Up to translations there are exactly 8 different ribbon tetrominoes, which we denote $\tau_1, ..., \tau_8$ as in Fig. 1. Let $\mathbf{T} = \{\boldsymbol{\tau_1}, ..., \boldsymbol{\tau_8}\}$.

Now let Γ be a simply connected region (a finite connected set of squares), and let v be a *tiling* of Γ by ribbon tetrominoes. This means that Γ is covered without intersection by parallel translations of ribbon tetrominoes. Denote by $a_i(v)$ the number of times tetromino τ_i occurs in the tiling v. While numbers a_i may be different for different tilings, this is no longer true for certain linear combinations of them.

Theorem 1.1. *For every simply connected region Γ and a tiling v of Γ we have*

$$a_2(v) + a_3(v) - a_6(v) - a_7(v) = C_1(\Gamma)$$

and

$$a_1(v) + a_2(v) + a_7(v) + a_8(v) = C_2(\Gamma) \qquad (\text{mod } 2),$$

where $C_{1,2}(\Gamma)$ are functions of Γ and do not depend on v.

The theorem was conjectured by the second author in [P], where it was proved for all row (column) convex regions. A more general version of the conjecture for all ribbon polyominoes remains open (see [P] for details).

[1] We are grateful to Jim Propp for introduction to the subject and interest in our work.

[2] Supported by an NSF Postdoctoral Research Fellowship.

0097-3165/99 $30.00

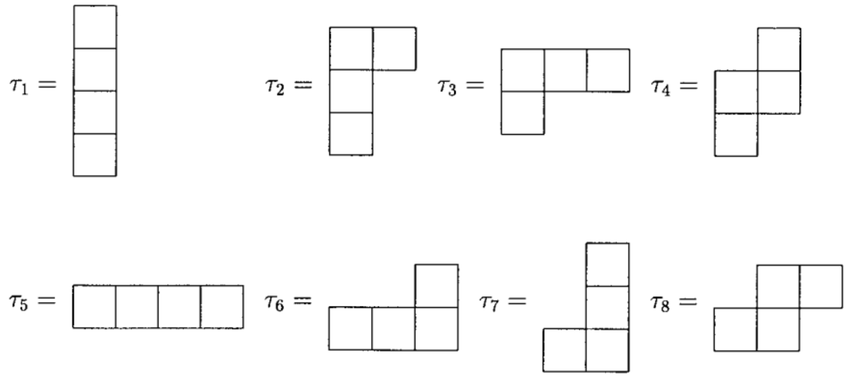


FIGURE 1

While the result for row (column) convex regions was obtained by use of the Young tableau technique, here we rely on the technique developed by Conway and Lagarias in [CL].

It remains open whether there exists a finite set of "moves" such that by using these moves one can start with any tiling and get to any other tiling of a given simply connected region. Such a set of moves was proposed in [P] where this property was shown for Ferrer's shapes. In case of domino tilings and lozenges the result is known for all simply connected regions (see [ST, T]).

It is important to note that as shown in [P] the theorem cannot be obtained by use of the coloring arguments (see [G, CL]). Thus our result lays in line with other "hard" results for trominoes (see [CL]), **T**-tetrominoes (see [W]), skew and square tetrominoes (see [Pr]), and rectangles (see [K]).

2. PROOF OF THE THEOREM

Observe that all tiles $\tau \in \mathbf{T}$ are simply connected. This fact is crucial in the induction we present below. Our proof relies on the following lemma.

LEMMA 2.1. *Let Γ be a compact simply connected region. Assume that v is a tiling of Γ by tiles $\tau_i \in \mathbf{T}$. Then there exists a tile τ in the tiling v such that $(\Gamma - \tau)$ is simply connected.*

Versions of the lemma have appeared previously in [CL, Pr]. We give here a new rigorous proof of the claim.

Proof. Denote by $|v|$ the number of tiles in a tiling v. The result is trivial for $|v| = 1, 2$. Now suppose $|v| > 2$. We say that two regions are *attached* if the intersection of their boundaries contains an interval. Note that two regions can be attached from either inside or outside.

Observe that if we remove any tile $\tau \in v$ which is attached to Γ, then we obtain a region which is a union of simply connected regions. Indeed, this follows from $\Gamma^c + \tau$ being connected since Γ is simply connected, and τ is attached to Γ^c. (Γ^c is a complement of Γ.)

Denote by $l(\tau)$ the number of tetrominoes in the smallest connected component in $\Gamma - \tau$, and by $n(\tau)$ the number of connected components of $\Gamma - \tau$. We will show that there exists a tile $\tau \in v$ such that *either* $n(\tau) = 1$ or $l(\tau) = 1$. This implies the lemma. Indeed, in the first case tile τ is the desired tile while in the second case we can simply remove a unique tile τ' in either of the smallest connected components and obtain the desired simply connected region $\Gamma - \tau'$.

Now, let τ be a tile attached to Γ. Let Γ_1 be any smallest connected component obtained after removing τ. Observe that the boundary of Γ_1 is made up of pieces of the boundary of Γ and τ. As τ is simply connected, Γ_1 has a common boundary with Γ, otherwise the boundary of Γ_1 lies inside the boundary of τ. Consider any tile τ' in Γ_1 which is attached to Γ. Consider removing tile τ' instead of τ. In this case, the component of $\Gamma - \tau'$ which contains τ also contains all components of $\Gamma - \tau$ other than Γ_1 simply because they are attached to τ. We call it a *big* component of $\Gamma - \tau'$. Observe that besides the big component all the other components must be of size smaller than $l(\tau)$. If there are no components other than the big component, then $n(\tau') = 1$ and tile τ' is the one desired in the lemma. If there exists such a component, we have $l(\tau') < l(\tau)$. Now proceed by induction until either $n(\tau) = 1$ or $l(\tau) = 1$.

This finishes proof of the lemma.

Let $F_2 = \langle A, B \rangle$ be a free group generated by A, B. A represents the direction from left to right and B represents the up direction.

For any region Γ and a point x on the boundary $\partial\Gamma$ define a word $\omega(\Gamma)$ obtained by reading $\partial\Gamma$ counterclockwise starting from x. For example for τ_2 starting at the lowest left corner $\omega(\tau_2) = AB^2ABA^{-2}B^{-3}$. Any region has more than one representation depending on the starting point. However, it is easy to see that all these presentations are conjugates of each other.

Consider a subgroup $G = \langle A^4, B^4, (AB)^2 \rangle$ of F_2, generated by the elements as shown, and let $H = N(G)$ be the smallest normal subgroup of F_2 which contains H. Finally, consider a quotient F_2/H and its Cayley graph representation given in Fig 2. Here we have an edge correspond to a generator A or B if it belongs to the corresponding square.

Lemma 2.2. *If Γ is tileable by tiles* $\mathbf{T}$ *then $\omega(\Gamma)$ is in H.*

Proof. By Lemma 2.1, it is sufficient to check that for every tile $\tau \in \mathbf{T}$ we

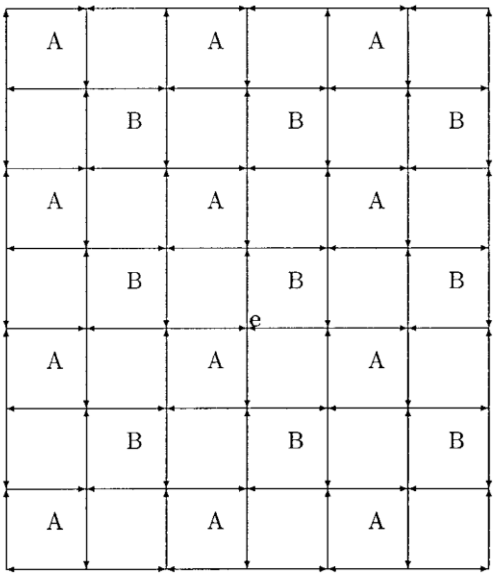


FIGURE 2

have $\omega(\tau) \in H$. Indeed, if this is true, we can use induction to show that $\omega(\Gamma) \in H$.

On the other hand, for the Cayley graph above it is easy to check that every tile in **T** is mapped into a closed path on the graph. This proves the lemma.

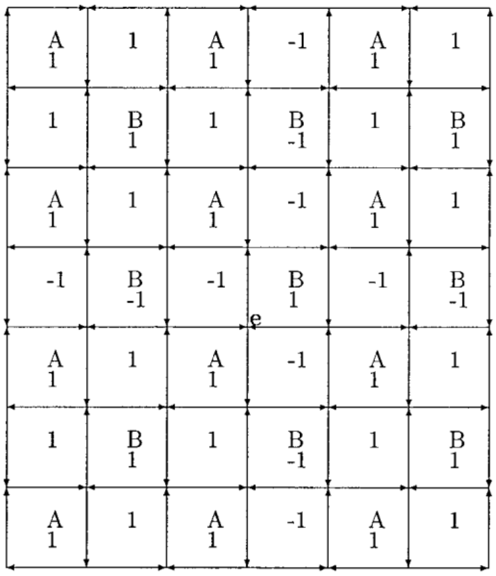
Now, to each simply connected region Γ which is tileable by tiles **T** we can assign a closed path $\omega(\Gamma)$ on the Cayley graph of F_2/H, although this path is not uniquely defined. By assigning weights to each cell in Fig. 2 and counting the winding numbers of the path of $\partial\Gamma$ with respect to these weights we will show that the identities in the theorem hold. (cf. [CL]).

LEMMA 2.3. *Assign values* 0 *to each cell that correspond to* A^4 *and* B^4 *and values* 1 *to each cell of* $ABAB$. *Then*

$$a_2(v) + a_3(v) - a_6(v) - a_7(v)$$

is equal to $-\frac{1}{2}$ *times the winding number of* $ABAB$ *cells.*

Proof. Use induction on the number of tiles covering Γ. For $n = 1$, check that the paths associated to tiles τ_2 and τ_3 enclose two cells $ABAB$



A 1	1	A 1	-1	A 1	1
1	B 1	1	B -1	1	B 1
A 1	1	A 1	-1	A 1	1
-1	B -1	-1	B 1	-1	B -1
A 1	1	A 1	-1	A 1	1
1	B 1	1	B -1	1	B 1
A 1	1	A 1	-1	A 1	1

FIGURE 3

going clockwise. Similarly, paths for tiles τ_6 and τ_7 enclose two cells *ABAB* going counterclockwise. Paths for tiles τ_1 and τ_5 enclose no *ABAB* cells. Finally, paths for tiles τ_4 and τ_8 enclose 2 *ABAB* cells, one in the clockwise direction and one the in counterclockwise direction. Thus for $n=1$ the statement is true.

Assume the statement is true for $n=k$. Let Γ be covered by $n=k+1$ tiles. By Lemma 2.1, there exists a tile τ such that $\Gamma-\tau$ is a simply connected region. Call it Γ_1. Then by a suitable conjugation $\omega(\Gamma)=\omega(\Gamma_1)\circ\omega(\tau)$ (here $\circ$ is a group operation in F_2). Now use the additivity property of winding numbers and the induction assumption for the region Γ_1. This proves the lemma.

Note that the first part of the Theorem 1.1 follows immediately from Lemma 2.3. Similarly, the second part is implied by the following result.

LEMMA 2.4. *Assign the values to each cell as shown on the Fig.* 3. *Namely, assign* -1 *to squares* (i, j) *with exactly one coordinate divisible by* 4. *Assign* 1 *to the remaining squares. Then*

$$a_1(v)+a_2(v)+a_7(v)+a_8(v) \qquad (\text{mod } 2)$$

is equal to $\frac{1}{2}$ *times the winding number of the region* Γ.

Proof. The proof is similar to the proof of Lemma 2.3. It is easy to check that the winding numbers are: 2 (mod 4) for τ_1, 2 (mod 4) for τ_2, 0 (mod 4) for τ_3, 0 (mod 4) for τ_4, 0 (mod 4) for τ_5, 0 (mod 4) for τ_6, 2 (mod 4) for τ_7, 2 (mod 4) for τ_8. The rest of the proof goes along the lines of the proof of Lemma 2.3. We omit the details.

REFERENCES

[CL] J. H. Conway and J. C. Lagarias, Tilings with polyominoes and combinatorial group theory, *J. Combin. Theory Ser. A* **53** (1990), 183–208.

[G] S. Golomb, "Polyominoes," Scribner's, New York, 1965.

[K] R. Kenyon, A note on tiling with integer-sided rectangles, *J. Combin. Theory Ser. A* **74** (1996), 321–332.

[P] I. Pak, Ribbon tile invariants, preprint, 1997.

[Pr] J. Propp, A pedestrian approach to a method of Conway, preprint, 1997.

[ST] N. Saladana and C. Tomei, An overview of domino and lozenge tilings, preprint, 1998.

[T] W. Thurston, Conway's tiling group, *Amer. Math. Monthly* **97** (1990), 757–773.

[W] D. Walkup, Covering a rectangle with **T**-tetrominoes, *Amer. Math. Monthly* **72** (1965), 986–988.