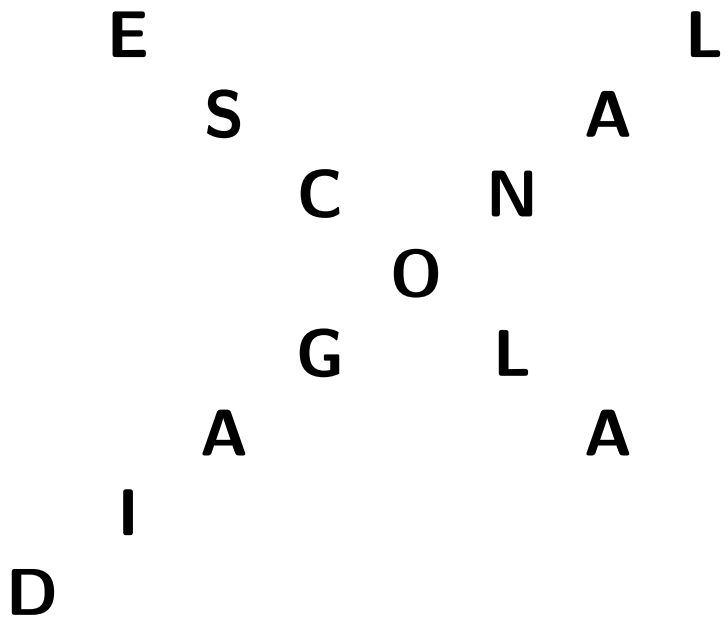


Escola Diagonal 2005

Editora: Ana Cannas da Silva

Título ◇ Escola Diagonal 2005

Editora ◇ Ana Cannas da Silva

Www ◇ `http://www.math.ist.utl.pt/escola/`

e ◇ `http://www.math.ist.utl.pt/diagonal/`

Documento LaTeX executado em 30 de Outubro de 2006, no Departamento de Matemática do Instituto Superior Técnico.

Prefácio

Integrada no Programa Gulbenkian *Novos Talentos em Matemática*, realizou-se de 5 a 9 de Setembro de 2005, no Instituto Superior Técnico (IST) em Lisboa, uma *Escola de Verão de Matemática*, designada *Escola Diagonal 2005*, dirigida a todos os interessados em Matemática com conhecimentos de nível pré-universitário. Esta Escola antecedeu o Encontro Nacional do Programa, que teve lugar no Luso, de 9 a 11 de Setembro.

A presente colectânea reúne textos correspondentes a três dos cursos da *Escola Diagonal 2005*, em que o orador ou monitores ou outros membros da audiência puderam contribuir com artigos. Junta-se uma lista dos anúncios de todos os cursos realizados.

Formato da Escola Diagonal 2005

A *Escola Diagonal 2005* ofereceu quatro cursos, cada um com cinco lições de hora-e-meia durante as manhãs, com os seguintes temas e professores:

- *Patologias na Análise*,
  por Manuel Arala Chaves (Univers. do Porto e Associação Atractor);

- *Uma Breve Excursão à Rua das Matemáticas*,
  por Rui Loja Fernandes (Instituto Superior Técnico);

- *Knots, Links, Braids and their Invariants*
  (*Nós, Elos, Tranças e seus Invariantes*),
  por Alexei Sossinsky (Universidade Independente de Moscovo, Rússia);

- *Quadratic Forms and Continued Fractions*
  (*Formas Quadráticas e Fracções Contínuas*),
  por Ramin Takloo-Bighash (Universidade de Princeton, EUA).

As lições foram complementados por sessões de trabalho durante três tardes, orientadas pelos seguintes monitores:

- para *Patologias na Análise*
  Vitor Saraiva (IST) e Diogo Silva (U.Porto);

- para *Uma Breve Excursão à Rua das Matemáticas*
  Luís Diogo (IST), João Nogueira (Austin) e Hugo Tavares (U.Lisboa);

- para *Knots, Links, Braids and their Invariants*
  Rui Carpentier (IST) e João Gouveia (U.Coimbra);

- para *Quadratic Forms and Continued Fractions*
  Gonçalo Tabuada (Paris) e Diogo Veloso (IST).

Horários e outras informações práticas podem ser encontrados na página `http://www.math.ist.utl.pt/escola`.

Participantes e Financiamento

Com cerca de 100 participantes – mais de 30 sendo estudantes do Ensino Secundário e os restantes maioritariamente estudantes do Ensino Superior – esta foi, no âmbito do Programa Gulbenkian *Novos Talentos em Matemática*, a iniciativa mais abrangente até à data. A Escola destinava-se a todos os interessados, solicitando-se apenas uma inscrição electrónica.

Os 19 bolseiros do Programa Gulbenkian *Novos Talentos em Matemática* em 2004-2005 e os 12 melhores classificados da final da categoria B das Olimpíadas Portuguesas de Matemática (OPM) em 2005 foram convidados individualmente pelo Serviço de Ciência da Fundação Gulbenkian, com oferta de apoio financeiro para despesas de transporte, alimentação e alojamento (no caso de participantes de fora de Lisboa) – 14 bolseiros e todos os 12 estudantes medalhados das OPM aceitaram o convite (alguns bolseiros não puderam participar por já terem outras aulas nessa semana).

Além dos 26 estudantes acima, foi oferecido apoio financeiro para despesas de transporte, alimentação e alojamento (no caso de participantes seleccionados de fora de Lisboa) a 14 candidatos que vinham a ser alunos universitários em 2005-2006 de qualquer área. A candidatura para apoio financeiro consistia num formulário preenchido e enviado até 1 de Julho de 2005.

Esta iniciativa da Fundação Calouste Gulbenkian foi essencialmente financiada pela Fundação Gulbenkian. Recebeu comparticipação financeira da Fundação para a Ciência e a Tecnologia através do Centro Internacional de Matemática, apoio logístico do Centro de Análise Matemática, Geometria e Sistemas Dinâmicos do Instituto Superior Técnico, e apoio na divulgação da Sociedade Portuguesa de Matemática.

Organização e Enquadramento

A organização da *Escola Diagonal 2005* esteve a cargo da Comissão Científica Coordenadora do Programa Gulbenkian *Novos Talentos em Matemática*, constituída por Ana Cannas da Silva, José Ferreira Alves, Orlando Neto e José Miguel Urbano, com o apoio do Serviço de Ciência da Fundação Gulbenkian e de Sandra Pereira do Centro de Análise Matemática, Geometria e Sistemas Dinâmicos.

Esta foi a segunda *Escola Diagonal*, tendo a primeira sido realizada de 6 a 10 de Setembro de 2004 no edifício-sede da Fundação Gulbenkian em Lisboa, com os cursos *Da Contagem ao Contínuo: Uso e Construção dos Números Reais* por António de Bivar Weinholtz (Universidade de Lisboa), *The Symmetries of Things* (*As Simetrias das Coisas*), por John Conway (Universidade de Princeton, EUA), *A Teoria dos Grupos nos Códigos* por Jorge Picado (Universidade de Coimbra) e *Teoria Ergódica: Probabilidades em Acção* por Marcelo Viana (Instituto Nacional de Matemática Pura e Aplicada, Brasil).

O Programa Gulbenkian *Novos Talentos em Matemática* foi instituído pela Fundação Calouste Gulbenkian no ano 2000 – Ano Mundial da Matemática – com o objectivo de estimular nos jovens o gosto, a capacidade e a vocação de pensar e investigar em Matemática. Este programa tem distinguido anualmente estudantes universitários de Matemática que evidenciam um elevado mérito académico e tem incentivado o desenvolvimento da sua cultura e aptidões matemáticas, apoiando o seu trabalho junto de reconhecidos especialistas, que exercem o papel de tutores. Espera-se dos participantes no Programa que, sob a orientação dos tutores, realizem trabalho de estudo aprofundado e/ou participem activamente num programa de seminários e/ou se iniciem na investigação em Matemática. Mais informações sobre o Programa Gulbenkian *Novos Talentos em Matemática* podem ser encontradas na página `http://www.math.ist.utl.pt/talentos`.

Impacto e Resultados

A *Escola Diagonal 2005* foi noticiada pelo menos em:

18/Ago – `cienciapt.net` – "IST Acolhe Escola de Matemática no Verão"

19/Ago – `portugaldiario.iol.pt` – "Portugal Caminha para uma *Iliteracia Numérica*"

20/Ago – Semanário *Expresso* – "Escola Diagonal – Primeira Escola de Verão de Matemática"

26/Ago – `educare.pt` – "Lisboa Acolhe Primeira Escola de Verão de Matemática"

31/Ago – Jornal quinzenal *Jornal de Letras* – "Matemática no Verão"

31/Ago – Semanário *Tempo* – "Escola Diagonal no Técnico"

5/Set – programa na TVI de manhã

5/Set – noticiário na Rádio Renascença de manhã

5/Set – noticiário na Rádio Seixal de tarde

5/Set – Diário *Destak* – "Aprender a Gostar de Matemática"

6/Set – transmissão em directo para programa *Bom Dia Portugal* na RTP1

6/Set – Diário *Metro Portugal* – "Aulas Especiais de Matemática"

8/Set – transmissão em directo para programa na Rádio Renascença à tarde

Tanto a nível da qualidade dos cursos, como a nível do impacto na população de estudantes, como a nível do impacto em geral, os organizadores da *Escola Diagonal* sentem-se plenamente satisfeitos com os resultados alcançados.

- A qualidade dos conteúdos científicos e o nível pedagógico das lições em geral ultrapassam confortavelmente os adequados para uma escola deste tipo.
- A adesão dos interessados e os comentários recolhidos após a Escola testemunham o entusiasmo e os benefícios para os participantes.
- Uma iniciativa de Matemática ter dado origem a notícias positivas, contribui para a urgente desmistificação desta ciência junto do público em geral.

Enfim, a *Escola Diagonal 2005* parece ter cumprido o papel que esta designação pretendia transmitir (partilhado com o *Seminário Diagonal*, paralelo ao Programa Gulbenkian *Novos Talentos em Matemática*): atravessar o rectângulo dos assuntos e níveis de Matemática. Ou, na linguagem da Álgebra Linear, atingir o objectivo principal da *diagonalização* dos assuntos – a apresentação de assuntos sofisticados como a soma de partes simples.

Agradecimentos

Muitas pessoas construiram o sucesso da *Escola Diagonal 2005*. O primeiro agradecimento vai para os quatro professores – Manuel Arala Chaves, Rui Loja Fernandes, Alexei Sossinsky e Ramin Takloo-Bighash, – os nove monitores Rui Carpentier, Luís Diogo, João Gouveia, João Nogueira, Vitor Saraiva, Diogo Silva, Gonçalo Tabuada, Hugo Tavares e Diogo Veloso – e mais três colaboradores na elaboração destas notas – Carlos Florentino, Pedro Matias e Roger Picken.

Claro que a *Escola Diagonal 2005* nunca teria acontecido e esta colectânea nunca teria sido publicada sem o generoso patrocínio da Fundação Calouste Gulbenkian a todos os níveis. O Programa Gulbenkian *Novos Talentos em Matemática* serve de alicerce às iniciativas *diagonais*. Pessoalmente reconhece-se o apoio do Director do Serviço de Ciência da Fundação Gulbenkian, João Caraça, e da Directora-Adjunta, Francisca Moura.

A *Escola Diagonal 2005* recebeu ainda comparticipação financeira da Fundação para a Ciência e a Tecnologia através do Centro Internacional de Matemática, apoio logístico do Centro de Análise Matemática, Geometria e Sistemas Dinâmicos do Instituto Superior Técnico, e apoio na divulgação da Sociedade Portuguesa de Matemática. Aponta-se em particular a cuidadosa e crucial ajuda da Sandra Pereira (do Centro de Análise Matemática, Geometria e Sistemas Dinâmicos) na organização prática do evento. O Banco BPI, através do Concurso de Apoio a Actividades Extracurriculares do IST, complementou o apoio financeiro que viabilizou a publicação desta colectânea.

O suporte gráfico em LaTeX que tornou tão fácil a produção deste documento em PDF e a criação de `http://www.math.ist.utl.pt/diagonal/` devem-se ao substancial empenho do João Boavida e do João Palhoto Matos no lançamento das inicativas *diagonais* entre 2000 e 2002. A capa dos volumes impressos é devida ao trabalho e à paciência do Luís Cruz-Filipe.

Ana Cannas,
Lisboa, 30 de Outubro de 2006

Conteúdo

Anúncios dos Cursos

Patologias na Análise

Manuel Arala Chaves (Universidade do Porto e Associação Atractor)

Nos cursos básicos de análise, surgem por vezes «enunciados óbvios», cuja demonstração resiste mais do que esperado à partida. Em alguns casos, porque demonstrações elementares são inesperadamente complicadas; noutros, porque os «resultados óbvios» são falsos (frequentemente com contra-exemplos difíceis de encontrar) – estes são por vezes chamados casos patológicos. Dois exemplos:

– Será que uma função real contínua (em todos os pontos) pode não ter derivadas – mesmo só laterais, mesmo possivelmente infinitas – em nenhum ponto? A resposta é sim.

– Será que uma curva pode «encher» um quadrado? E um cubo? E ... ? A resposta, surpreendentemente, vai depender do grau de regularidade que se exigir na definição de curva.

As palestras andarão à volta de questões deste tipo, procurando analisar alguns exemplos concretos, mas também enquadrá-los num tratamento unificado. De passagem, ver-se-á ainda que a situação «patológica» é por vezes a genérica e a «não patológica» a excepcional; e serão comparados dois modos (não equivalentes) de avaliar o carácter «excepcional» de uma situação.

Referências:

- *A Primer of Real Functions* por Ralph Boas, The Carus Mathematical Monographs 13, Mathematical Association of America (1996).
- *Counterexamples in Analysis* por Bernard Gelbaum e John Olmsted, Dover (2003).
- *Measure and Category* por John Oxtoby, Springer-Verlag (1971).
- *Space-Filling Curves* por Hans Sagan, Springer Verlag (1994).
- *Calculus* por Michael Spivak, Publish or Perish (1994).
- *Counterexamples in Topology* por Lynn Steen e Arthur Seebach, Dover (1995).
- *General Topology* por Stephen Willard, Dover (2004).

UMA BREVE EXCURSÃO À RUA DAS MATEMÁTICAS

Rui Loja Fernandes (Instituto Superior Técnico)

A divisão tradicional da Matemática nas três áreas fundamentais, Álgebra, Análise e Geometria, subsiste até aos dias de hoje. Embora a Matemática contemporânea cada vez menos possa ser caracterizada dessa forma (quer pela sua natureza multidisciplinar, quer pelas novas áreas que não encaixam nesta divisão tradicional), numa breve excursão pelo mundo da Matemática vale a pena visitar cada uma dessas áreas.

Nas sessões deste curso iremos estudar pequenas demonstrações de alguns resultados elementares (mas fundamentais) de cada uma das áreas tradicionais. A ideia em cada sessão é transmitir os sabores e os odores de uma das áreas. O que têm de comum todos os exemplos que escolhemos são as ideias brilhantes e as observações astuciosas, ou seja, afinal as características do pensamento matemático do passado, do presente e do futuro.

Plano das sessões com problemas ilustrativos:

- 1ª e 2ª sessões: A Casa dos Números Uma infinidade de primos. Algoritmo de Euclides. O Teorema fundamental da aritmética. Congruências e os anéis Z_p.
  Problema: Quando é que um número natural é a soma de dois quadrados?

- 3ª e 4ª sessões: A Galeria de Arte Grafos e rectas no plano. Poliedros e símplices. A fórmula de Euler.
  Problema: Dados n pontos não-colineares no plano, será que existe uma recta que passa exactamente por dois desses pontos?

- 5ª sessão: O Hotel de Hilbert Números cardinais. A hipótese do contínuo. Números ordinais.
  Problema: Será que existe uma curva contínua no plano que visita todos os pontos do interior dum quadrado?

Referência:

Proofs from THE BOOK por Martin Aigner e Günter Ziegler, Springer-Verlag, Berlin (2001).

Knots, Links, Braids and their Invariants (Nós, Elos, Tranças e seus Invariantes)

Alexei Sossinsky (Universidade Independente de Moscovo, Rússia)

The mathematical theory of knots and links, which studies non-intersecting and non-self-intersecting curves in space, is at least 150 years old, but has recently surged to the forefront of science, attracting not only mathematicians and physicists, but also experts in biology and chemistry, and in some branches of engineering as well. In the decade 1985–1995, no less than four mathematicians who contributed to the theory were awarded the Fields Medal (the "Nobel Prize for mathematicians"): Vaughan Jones (New Zealand), Edward Witten (USA), Vladimir Drinfeld (Ukraine), Maxim Kontsevich (Russia).

Surprizingly, although many of the original papers involve some very sophisticated mathematics, there now exists a very elementary exposition of the main results, so that no advanced mathematical prerequisites will be needed to follow the course. In it, there will be lots of very visual geometry (pictures and even computer-generated animations), some absolutely elementary algebra (adding and multiplying polynomials, linear spaces, permutation groups), and a new kind of calculus (which may be called "games with little diagrams").

The main protagonists, however, will be several kinds of invariants, i.e., algebraic entities allowing to easily solve difficult classification problems concerning the main three-dimensional geometric objects of the course – knots, links, braids.

Topics:

- The geometry and arithmetic of knots: knot diagrams and equivalence of knots via Reidemeister moves, decomposition into prime knots, knot tables, Conway axioms for the Alexander polynomial.
- Braids and their relationship with knots and links: encoding braids by standard generators, Artin theorem on the braid group, Alexander theorem (any knot is the closure of a braid), braid comparison algorithms.
- The Kauffman bracket and the Jones polynomial: Kauffman's imaginary statistical model and his bracket, definition of the Jones polynomial via the Kauffman bracket, properties of the Jones polynomial, proof of the Tait conjectures on alternating knots.
- Elementary theory of Vassiliev invariants: Axioms for Vassiliev invariants, the one-term and four-term relations, the bialgebra of chord diagrams, Kontsevich theorem (isomorphism of the Vassiliev (bi)algebra and the (bi)algebra of chord diagrams).
- Relationship with physics and biology: the Potts water-ice model and the Jones polynomial, the Conway moves and DNA.

Referência:

The Knot Book por Colin Adams, American Mathematical Society (2004).

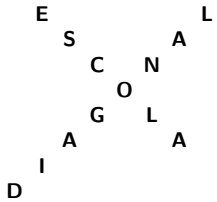
QUADRATIC FORMS AND CONTINUED FRACTIONS (FORMAS QUADRÁTICAS E FRACÇÕES CONTÍNUAS)

Ramin Takloo-Bighash (Universidade de Princeton, EUA)

A quadratic form is a polynomial which is of degree two. Quadratic forms have been of interest since essentially the beginning of time. The first known example of a quadratic form studied by mankind is what is known as the Pythagorean equation discovered in the context of geometry: given a right angle triangle of sides a, b, c, we know that $a^2 = b^2 + c^2$ if a is the biggest side. There is historical evidence suggesting that ancient Egyptians used triangles with sides 3, 4, 5 to construct right angle triangles; they used these triangles as a practical tool in architecture, and we all know what they were capable of building!

Topics:

- Let's complicate things: Continued Fractions;
- How to count the number of cows of a certain Greek god: Pell's Equation;
- To square or not to square: Quadratic forms in general;
- Not all quadratic forms are born equal: Equivalence and Gauss' theory of genera;
- World through the eyes of a quadratic form: Representability by quadratic forms; and finally
- From 15 to eternity: Bhargava-Conway Theorem.



Uma Breve Excursão à Rua das Matemáticas

Rui Loja Fernandes
Departamento de Matemática
Instituto Superior Técnico
rfern@math.ist.utl.pt

Introdução

A divisão tradicional da Matemática nas três áreas fundamentais, Álgebra, Análise e Geometria/Topologia, subsiste até aos dias de hoje. Embora a Matemática contemporânea cada vez menos possa ser caracterizada desta forma, quer pelas novas áreas que não encaixam nesta divisão tradicional, quer pela importância crescente de áreas multidisciplinares, numa breve excursão pelo mundo da Matemática vale a pena uma visita a cada uma destas grandes casas que integram a Rua das Matemáticas.

Este pequeno curso divide-se, pois, em três partes. Em cada uma destas visitas iremos experimentar os diferentes odores e sabores destas grandes áreas da Matemática. É claro que, apesar das diferenças, haverá muito em comum nestas visitas. Veremos que nelas habitam as ideias brilhantes e as observações astuciosas, o raciocínio lógico e o poder da abstracção, ou seja, afinal as características do pensamento matemático do passado, do presente e do futuro!

A grande fonte de inspiração para este curso foi o livro escrito por Martin Aigner e Günter Zigler, *Proofs from the BOOK* (2ª edição, Springer-Verlag, Berlim, 2001). Este livro precioso contém muito mais material do que o discutido aqui. O leitor é fortemente encorajado a adquirir e ler este livro, onde certamente encontrará ainda maior motivação para aprender Matemática.

Estas notas referem-se ao curso com o mesmo nome que leccionei na *Escola Diagonal 2005*, integrada no Programa Gulbenkian *Novos Talentos em Matemática*. Gostaria de agradecer aos organizadores da Escola, Ana Cannas da Silva, José Ferreira Alves, Orlando Neto e José Miguel Urbano, pelo convite para leccionar este curso, à Fundação Calouste Gulbenkian pelo apoio prestado, e aos monitores Luís Diogo, João Nogueira, Hugo Tavares, que orientaram as sessões práticas e fizeram várias sugestões e comentários, tendo detectado ainda vários erros numa versão anterior destas notas.

1 A Casa dos Números

1.1 Os Números Primos

Uma ideia transversal a todos os domínios científicos, é a de decompor um objecto em objectos mais simples e, por isso, mais fáceis de estudar. Por exemplo, na Química a teoria atómica explica que as moléculas são constituídas por átomos, enquanto que na Física, as partículas elementares são constituídas por quarks. Quando aplicamos esta ideia aos **números naturais**

$$\mathbb{N} = \{1, 2, 3, 4, \dots\}$$

somos levados naturalmente ao conceito de *número primo* que vamos agora recordar.

Dados dois números naturais $m, n \in \mathbb{N}$, dizemos que n é **factor** ou **divisor** de m, e escrevemos "$n|m$", se m é um múltiplo de n, *i.e.*, se $m = kn$ para algum natural $k \in \mathbb{N}$. É claro que 1 é factor de qualquer natural n e qualquer natural é factor de si próprio.

DEFINIÇÃO 1. Seja $p \in \mathbb{N}$. Dizemos que p é **primo** se $p > 1$ e se, para todo o $k \in \mathbb{N}$ tal que $k|p$, temos que $k = 1$ ou $k = p$. Designamos por $\mathbb{P}$ o conjunto dos naturais primos.

Um número primo é, pois, um número que não admite outros factores para além dos factores "óbvios", e que é portando indecomponível.

Exemplo 2.

É fácil verificar quais são os primeiros naturais primos:

$$\mathbb{P} = \{2, 3, 5, 7, 11, 13, \dots\}.$$

Observe que para provar que p é primo não é necessário testar todos os números k com $1 < k < p$, podendo o teste terminar com o maior natural k tal que $k^2 < p$. No caso de $p = 13$, basta portanto constatar que 13 não é múltiplo de 2 nem de 3.

Exercício 1. Faça uma lista com os naturais entre 100 e 200. Observe que $17^2 = 289$, e corte da sua lista todos os múltiplos de 2, 3, 5, 7, 11 e 13. Quais são os números que restam[1]?

1 Chama-se a este procedimento o **filtro de Eratóstenes**. Eratóstenes (276 a.C.-194 a.C.) nascido na actual Líbia, foi o terceiro bibliotecário da famosa Biblioteca de Alexandria. Entre outras coisas estabeleceu a esfericidade da Terra, e calculou com grande exactidão o seu diâmetro.

Um pouco de reflexão sugere que todo o número natural deve-se decompor num produto de números primos. De facto, temos o seguinte resultado importante:

Teorema 3 (Teorema Fundamental da Aritmética). *Para todo o natural $n \geq 2$, existem números primos $p_1, p_2, \ldots, p_k \in \mathbb{P}$ tais que:*

$$n = p_1 p_2 \cdots p_k.$$

Esta factorização em primos é única a menos da ordem dos factores.

Demonstração. Começamos por observar que:

- Qualquer natural $n \geq 2$ tem pelo menos um divisor primo p.

De facto, o conjunto $D = \{m \in \mathbb{N} : m > 1 \text{ e } m|n\}$ é minorado (todos os elementos de D são maiores do que 1) e não-vazio (contém n). Existe pois um elemento mínimo p de D. Se p não é primo, então $p = mk$, onde $1 < m < p$. Como m é obviamente factor de n, p não pode ser o mínimo de D. Concluímos que p é primo.

Vejamos então, por indução, que se $n \geq 2$, existem números primos $p_1, p_2, \ldots, p_k$ tais que

$$n = p_1 p_2 \cdots p_k.$$

De facto, temos que:

- Se $n = 2$ então n possui uma factorização deste tipo (basta tomar $k = 1$ e $p_1 = 2$).
- Supomos agora que qualquer natural m com $2 \leq m < n$ tem uma factorização do tipo indicado. Pretendemos provar que n tem também uma factorização deste tipo. Como observámos acima, n possui um divisor primo q. Se $q = n$, então n é primo, e tomamos $k = 1$ e $p_1 = q = n$. Caso contrário, $q < n$ e, portanto, $n = mq$, onde $2 \leq m < n$. Pela hipótese de indução, existem números primos $p_1, p_2, \ldots, p_{k'}$, tais que
$$m = p_1 p_2 \cdots p_{k'}.$$
Tomamos neste caso $k = k' + 1$ e $p_k = q$.

Isto mostra a existência de factorizações primas. A unicidade da factorização será estudada mais adiante. □

Exercício 2. Seja $2\mathbb{N} = \{2, 4, 6, 8, \ldots\}$ o conjunto dos naturais pares. Designe por "primos" em $2\mathbb{N}$ os naturais pares que não podem ser expressos como produtos de outros naturais pares. Mostre que os naturais pares admitem factorizações primas, mas que estas não são únicas.

Exercício 3. O que pode dizer sobre existência e unicidade de factorizações no conjunto dos **números inteiros** $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$?

O conceito de número primo e a existência de factorizações primas para qualquer número natural, sugerem imediatamente algumas questões simples:

- Quantos números primos existem?
- Como podemos reconhecer que um natural é primo?
- Existe algum método eficiente para produzir números primos?
- Como podemos encontrar a factorização prima de um dado natural?

Muitas questões deste tipo, apesar de simples de formular, são extremamente difíceis e não sabemos a resposta. Algumas são mesmo temas de investigação contemporânea e, apesar da sua origem, têm também reflexos interessantes na vida actual! Por exemplo, as modernas técnicas de Criptografia exploram a relativa facilidade de cálculo de grandes números primos, comparada com a dificuldade de determinar os factores primos dos naturais que podemos obter pela multiplicação desses primos. Neste contexto, os números "grandes" podem ter mais duma centena de dígitos; a sua factorização por verificação sequencial de todos os possíveis factores envolveria um número de divisões da ordem de 10^{50}! Não sabemos até que ponto é possível estabelecer um algoritmo prático para a factorização de números desta ordem de grandeza, mas enquanto essa ignorância se mantiver, as mais secretas comunicações financeiras, políticas ou militares, poderão continuar a fazer-se com segurança recorrendo aos números primos. Outro exemplo de um problema "prático", onde as propriedades dos números primos têm reflexos importantes, é o problema do reconhecimento da fala por computadores que exige o desenvolvimento de algoritmos tão rápidos quanto possível para a decomposição de sons nas suas frequências fundamentais, uma técnica conhecida como Análise de Fourier. A velocidade teórica máxima desses algoritmos está directamente relacionada com a função $\pi(x)$ que fornece o número de primos menores que x.

1.2 O Algoritmo de Euclides

Se dispusermos de dois relógios de areia (ampulhetas), um marcando um intervalo de tempo de 21 minutos e o outro marcando um intervalo de tempo de 30 minutos, que intervalos podemos medir utilizando as duas ampulhetas? Certos intervalos são obviamente possíveis, se utilizarmos sucessivamente uma ou outra ampulheta. Por exemplo, 30, 60, 90,..., 21, 42, 63,..., ou

somas destes números, como 51, 81, 111,..., 102, 123,..., 132, 153,..., etc. Se usarmos simultaneamente as duas ampulhetas, podemos obter diferenças destes números. Exemplos são $9 = 30 - 21$, $3 = 63 - 60$, etc.

Um exame mais cuidadoso dos números que se podem obter desta forma sugere as seguintes observações:

- Podemos obter qualquer natural da forma $x21 + y30$ com $x, y \in \mathbb{Z}$.
- Todos os números da forma $x21 + y30$ são múltiplos de 3 (pois 3 é o máximo divisor comum de 21 e 30).

Por outro lado, existem inteiros x' e y' (*e.g.*, $x' = 3, y' = -2$) tais que $3 = x'21 + y'30$. Em particular, se $m = k3$ é um qualquer múltiplo de 3, então $m = k(x'21+y'30) = k(x'21+y'30) = x''21+y''30$. Por outras palavras,

- Os números da forma $x21 + y30$ são precisamente os múltiplos de 3;
- $3 = \text{mdc}(21, 30)$ é o menor natural da forma $x21 + y30$.

Motivado por este problema, bem como a questão da unicidade de factorizações primas, vamos estudar em algum detalhe as noções de máximo divisor comum e mínimo múltiplo comum de dois números naturais. Começamos por recordar a sua definição:

Definição 4. Se $n, m \in \mathbb{N}$, então:

(i) $\text{mdc}(n, m) = \max\{k \in \mathbb{N} : k|n \text{ e } k|m\}$ diz-se **máximo divisor comum** de n e m;

(ii) $\text{mmc}(n, m) = \min\{k \in \mathbb{N} : n|k \text{ e } m|k\}$ diz-se **mínimo múltiplo comum** de n e m.

Exemplo 5.

Se $n = 12$ e $m = 16$, os divisores de n e m formam os conjuntos

$$\{k \in \mathbb{N} : k|12\} = \{1, 2, 3, 4, 6, 12\},$$
$$\{k \in \mathbb{N} : k|16\} = \{1, 2, 4, 8, 16\}.$$

Consequentemente, os divisores comuns a 12 e 16 formam o conjunto

$$\{k \in \mathbb{N} : k|12 \text{ e } k|16\} = \{1, 2, 4\},$$

donde o respectivo máximo divisor comum é $\text{mdc}(12, 16) = 4$.

Por outro lado, os múltiplos de 12 *e* 16 *são*

$$\{k \in \mathbb{N} : 12|k\} = \{12, 24, 36, 48, \dots\},$$
$$\{k \in \mathbb{N} : 16|k\} = \{16, 32, 48, \dots\},$$

donde concluímos que os múltiplos comuns a 12 *e* 16 *formam o conjunto*

$$\{k \in \mathbb{N} : 12|k \text{ e } 16|k\} = \{48, 96, \dots\},$$

sendo o respectivo mínimo múltiplo comum $\text{mmc}(12, 16) = 48$.

Observe-se que neste exemplo $\text{mdc}(m, n)$ é *múltiplo* de todos os divisores comuns a n e m, e $\text{mmc}(n, m)$ é *factor* de todos os múltiplos comuns a n e m. Estas são propriedades gerais e para verificá-las precisamos do:

TEOREMA 6 (ALGORITMO DE DIVISÃO). *Se $n, m \in \mathbb{N}$ existem inteiros únicos q, r, tais que $m = nq + r$ e $0 \leq r < n$.*

Demonstração. O argumento para provar a existência corresponde ao processo usual para efectuar uma divisão.

(i) *Existência:* Considere-se o conjunto $Q = \{x \in \mathbb{N}_0 : nx \leq m\}$. Note-se que Q é não-vazio (porque $0 \in Q$) e majorado (porque $x \leq nx \leq m$). Tem consequentemente um máximo $x = q$. É claro que $nq \leq m < n(q + 1)$, porque $q \in Q$ e $(q + 1) \notin Q$. Subtraindo nq destas desigualdades, obtemos $0 \leq r \leq n$, já que $r = m - nq$.

(ii) *Unicidade:* Se $m = nq + r = nq' + r'$, segue-se que $n(q - q') = r' - r$. Suponha-se que $q' < q$, é óbvio que $q - q' \geq 1$ e $r' - r \geq n$. Por outro lado, r e r' verificam $0 \leq r, r' < n$, donde temos $r' - r < n$, uma contradição. Da mesma forma, obtém-se uma contradição se $q' > q$. Assim, só pode ser $q = q'$, mas como $n(q - q') = r' - r$, também $r = r'$. □

Naturalmente, se $n, m \in \mathbb{N}$ e $m = nq + r$ com $0 \leq r < n$, dizemos que q e r são respectivamente o **quociente** e o **resto** da divisão de m por n.

Dados $n, m \in \mathbb{N}$ vamos designar por $\langle n, m\rangle \subset \mathbb{Z}$ o conjunto dos inteiros que se obtém fazendo combinações de n e m com coeficientes inteiros:

$$\langle n, m\rangle = \{xn + ym : x, y \in \mathbb{Z}\}.$$

Por exemplo, no caso de dois relógios de areia, os inteiros do conjunto $\langle 21, 30\rangle$ são precisamente os intervalos de tempo que podemos medir. Neste exemplo, vimos que $\langle 21, 30\rangle$ é formado pelos múltiplos de 3. Em geral, escrevemos:

$$\langle n\rangle = \{xn : x \in \mathbb{Z}\},$$

de forma que temos $\langle 21, 30\rangle = \langle 3\rangle$. Mais geralmente, é verdade a seguinte proposição:

PROPOSIÇÃO 7. *Se $n, m \in \mathbb{N}$, então $\langle n, m\rangle = \langle d\rangle$ onde $d = \mathrm{mdc}(n, m)$. Em particular, temos que:*

(i) a equação $xn + ym = d$ tem soluções $x, y \in \mathbb{Z}$;

(ii) se k é um divisor comum de n e m, então k é também divisor de d.

Demonstração. Vamos designar o conjunto $\langle n, m\rangle$ por I. Consideramos, também, o conjunto $I^+ = \{q \in I : q > 0\}$. Este conjunto é não vazio e possui 0 como minorante, donde possui um mínimo d_0.

Como $d_0 \in I$ segue-se (porquê?) que $\langle d_0\rangle \subseteq I$, restando-nos portanto provar a inclusão oposta $I \subseteq \langle d_0\rangle$, ou seja, que, se $m \in I$, então m é múltiplo de d_0. Seja $m \in I$, e q e r o quociente e resto da divisão de m por d_0 (recorde-se que $d_0 > 0$, donde $d_0 \neq 0$). Temos então $m = qd_0 + r$, ou $r = m - qd_0$. Observe-se que $qd_0 \in \langle d_0\rangle$, e portanto $qd_0 \in I$ (já vimos que $\langle d_0\rangle \subseteq I$). Como $r = m - qd_0$ é a diferença de dois elementos de I, temos que $r \in I$. Finalmente, como $0 \leq r < d_0$ e d_0 é, por definição, o menor elemento positivo de I, temos necessariamente $r = 0$, donde m é múltiplo de d_0, *i.e.*, $m \in \langle d_0\rangle$. Portanto, $I \subseteq \langle d_0\rangle$. Agora:

(i) É óbvio que $d_0 \in \langle d_0\rangle = \langle n, m\rangle$. Como $\langle n, m\rangle$ é o conjunto dos inteiros da forma $xn + ym$, existem inteiros x', y' tais que $d_0 = x'n + y'm$.

(ii) É igualmente óbvio que $n, m \in \langle n, m\rangle = \langle d_0\rangle$. Portanto, n e m são múltiplos de d_0, que é um divisor comum a n e m. Por outro lado, se $d \in \mathbb{N}$ é um qualquer divisor comum a n e m, temos $n = dn'$ e $m = dm'$, donde $d_0 = x'n + y'm = d(x'n' + y'm')$, ou seja, $d|d_0$. Em especial, $d \leq d_0$ e d_0 é o máximo divisor comum de n e m. □

A proposição (e a sua demonstração) sugere que o cálculo de $\mathrm{mdc}(n, m)$ pode ser feito por busca do menor natural no conjunto $\langle n, m\rangle$. O Algoritmo de Divisão torna essa busca possível recorrendo ao seguinte lema.

LEMA 8. *Se $n, m \in \mathbb{N}$ e $m = qn + r$, então $\langle n, m\rangle = \langle n, r\rangle$.*

Demonstração. Por um lado,

$$\begin{aligned} k \in \langle n, m\rangle &\Longrightarrow k = xm + yn \\ &\Longrightarrow k = x(qn + r) + yn \\ &\Longrightarrow k = (xq + y)n + xr \\ &\Longrightarrow k \in \langle n, r\rangle, \end{aligned}$$

logo $\langle n, m\rangle \subset \langle n, r\rangle$. Por outro lado,

$$\begin{aligned} k \in \langle n, r\rangle &\Longrightarrow k = xn + yr \\ &\Longrightarrow k = xn + y(m - qn) \\ &\Longrightarrow k = (x - yq)n + ym \\ &\Longrightarrow k \in \langle n, m\rangle, \end{aligned}$$

ou seja, $\langle n, r\rangle \subset \langle n, m\rangle$. □

O **Algoritmo de Euclides**[2] é a aplicação repetida do lema anterior até obter uma divisão exacta ($r = 0$). Este é um procedimento muito simples, fácil de programar, e que passamos a ilustrar no caso com que iniciámos esta secção:

Exemplo 9.

Se $n = 21$ e $m = 30$, então

$$\begin{aligned} 30 = 1 \cdot 21 + 9 &\Longrightarrow \langle 30, 21\rangle = \langle 21, 9\rangle, \\ 21 = 2 \cdot 9 + 3 &\Longrightarrow \langle 21, 9\rangle = \langle 9, 3\rangle, \\ 9 = 3 \cdot 3 + 0 &\Longrightarrow \langle 9, 3\rangle = \langle 3\rangle. \end{aligned}$$

Logo

$$\langle 30, 21\rangle = \langle 3\rangle,$$

e pela proposição temos que $3 = \mathrm{mdc}(21, 30)$. *Em termos gerais,* i.e., *começando com dois naturais quaisquer n e m, e supondo $n < m$, o procedimento a seguir deve ser claro, e corresponde a um processo iterativo muito fácil de programar (experimente-o!). Observe-se também que é simultaneamente possível determinar inteiros x e y tais que* $\mathrm{mdc}(n, m) = xn + ym$. *Das equações acima temos imediatamente*

$$3 = 21 + (-2)9 \ \textit{e}\ 9 = 30 + (-1)21,$$

donde

$$3 = 21 + (-2)[30 + (-1)21] = (3)21 + (-2)30.$$

2 Euclides (c. 300 a.C) viveu em Alexandria no apogeu da matemática da Grécia Antiga. Euclides é conhecido como o autor dos *Elementos*, o tratado de matemática mais famoso de todos os tempos. Na realidade sabemos muito pouco sobre Euclides. Cópias de fragmentos dos *Elementos* chegaram até aos nossos dias. Sabemos que era constituído por 13 livros (ou capítulos). Os primeiros seis livros eram dedicados à geometria plana (a geometria Euclideana); os livros 6 a 9 eram dedicados à teoria dos números (e continham o seu algoritmo); o livro 10 era dedicado ao estudo das quantidades incomensuráveis (*i.e.*, os irracionais) e os últimos três livros eram dedicados ao estudo dos sólidos.

Dizemos que dois naturais n e m são **primos entre si** se $\mathrm{mdc}(n,m)=1$. Observe que n e m são primos entre si se, e só se, não possuem quaisquer factores primos em comum. Usando o Algoritmo de Euclides é simples determinar se dois naturais são primos entre si.

Exemplo 10.

É fácil verificar que 4 *e* 27 *são primos entre si,* i.e., $\mathrm{mdc}(4,27)=1$. *De facto, recorrendo ao Algoritmo de Euclides, verificamos que:*

$$\langle 27,4\rangle = \langle 4,3\rangle = \langle 3,1\rangle = \langle 1\rangle.$$

Finalmente, é um exercício simples verificar as propriedades análogas para o mínimo múltiplo comum.

Exercício 4. Sejam $n,m\in\mathbb{N}$. Então $\langle n\rangle\cap\langle m\rangle=\langle l\rangle$ onde $l=\mathrm{mmc}(n,m)$. Temos, ainda, que:

(a) $n|l$ e $m|l$;

(b) para qualquer $k\in\mathbb{N}$ tal que $n|k$ e $m|k$ tem-se também que $l|k$.

Estamos agora em condições de discutir a unicidade da factorização prima de um natural. O resultado chave para demonstrar essa unicidade é o seguinte lema.

Lema 11 (Euclides). *Sejam $m,n\in\mathbb{N}$ dois naturais e $p\in\mathbb{P}$ um número primo. Se p é factor do produto mn, então p é factor de m ou factor de n.*

Demonstração. Seja $d=\mathrm{mdc}(m,p)$. Como d é factor de p e p é primo, temos d=1 ou $d=p$.

É evidente que, se $d=p$, então p é factor de m. Se $d=1$, existem inteiros x e y tais que $1=xm+yp$, donde $n=nxm+nyp$. Como p é factor de mn, existe também um inteiro z tal que $mn=zp$. Concluímos que $n=xzp+nyp=(zx+ny)p$, e portanto p é factor de n. □

Exemplo 12.

Uma das descobertas dos matemáticos gregos da Antiguidade que mais os surpreendeu e intrigou foi, em linguagem moderna, a da existência de números irracionais. *Podemos verificar agora sem dificuldade que $\sqrt{2}$ é irracional,* i.e., *que não existem inteiros n e m tais que $\left(\frac{n}{m}\right)^2=2$, ou seja, $n^2=2m^2$. Argumentamos por absurdo.*

Podemos supor, sem perda de generalidade, que m e n são primos entre si (porquê?). Notamos agora que

$$n^2=2m^2\Rightarrow 2|n^2\Rightarrow 2|n,$$

pelo Lema de Euclides. Concluímos que $n = 2k$, para algum inteiro k. Assim, $n^2 = 4k^2$, donde $4k^2 = 2m^2$, ou ainda $2k^2 = m^2$. Como $2|m^2$, segue-se novamente do Lema de Euclides que $2|m$, contradizendo a hipótese de m e n serem primos entre si. Concluímos que a equação $n^2 = 2m^2$ não tem soluções nos inteiros, e $\sqrt{2}$ não é racional.

Exercício 5. Generalize este exemplo ao caso $\sqrt{n}$, quando $n \in \mathbb{N}$ não é um quadrado perfeito. Isto é, mostre que se n é um natural, a sua raiz quadrada ou é outro natural (caso em que n é um quadrado perfeito) ou é um número irracional.

Podemos generalizar o Lema de Euclides para um qualquer produto finito de inteiros. A demonstração (que fica como exercício) deve ser feita por indução no número de factores.

COROLÁRIO 13. *Se p é primo e $p\,|m_1 \cdots m_k$, então:*

(i) $p|m_j$ para algum j, com $1 \leq j \leq k$.

(ii) Se os m_i são primos, então $p = m_j$, para algum j, com $1 \leq j \leq k$.

Podemos agora completar a demonstração do Teorema Fundamental da Aritmética, mostrando a unicidade dos factores primos. Procedemos como se segue. Supondo que k e m são naturais, $p_1 \leq p_2 \leq \cdots \leq p_k$ e $q_1 \leq q_2 \cdots \leq q_l$ são primos, e

$$p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l,$$

provamos que $k = l$ e $p_i = q_i$. Para isso, argumentamos por indução em k:

- Para $k = 1$, o resultado é óbvio da definição de número primo;
- Supomos o resultado válido para o natural $k - 1$, e

$$p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l.$$

Seja $P = \{p_i : 1 \leq i \leq k\}$ e $Q = \{q_j : 1 \leq j \leq l\}$. Note-se mais uma vez que de acordo com a definição de primo temos necessariamente $l > 1$, porque $k > 1$. Pelo Corolário 13 é evidente que $p_k \in Q$, donde $p_k \leq q_l$, e analogamente $q_l \in P$, donde $q_l \leq p_k$. Concluímos que $p_k = q_l$, e segue-se que

$$p_1 p_2 \cdots p_{k-1} = q_1 q_2 \cdots q_{l-1}.$$

Pela hipótese de indução, $k - 1 = l - 1$ e $p_i = q_i$, para $i < k$, donde $k = l$ e $p_i = q_i$, para $i \leq k$.

A factorização de n em primos pode evidentemente conter factores repetidos, e é por isso comum escrevê-la na forma

$$n = p_1^{e_1} p_2^{e_2} \cdots p_r^{e_r}, \qquad (e_i \geq 1),$$

que se diz a **factorização de n em potências primas**. Esta expressão é única, a menos da ordem dos factores.

Exercício 6. Se $m, n \in \mathbb{N}$ possuem factorizações em potências primas:

$$m = p_1^{m_1} p_2^{m_2} \cdots p_r^{m_r},$$
$$n = q_1^{n_1} q_2^{n_2} \cdots q_s^{n_s},$$

determine as factorizações primas de $\mathrm{mdc}(m, n)$ e $\mathrm{mmc}(m, n)$.

1.3 Uma Infinidade de Primos

O Teorema Fundamental da Aritmética não implica directamente a existência dum número infinito de primos. Este último facto foi também descoberto por Euclides.

TEOREMA 14 (EUCLIDES). *O conjunto dos números primos $\mathbb{P}$ não é finito.*

Demonstração. Seja $\{p_1, \ldots, p_r\}$ um conjunto finito de números primos. O número natural

$$n = p_1 p_2 \cdots p_r + 1,$$

possui um divisor primo p. Se $p \in \{p_1, \ldots, p_r\}$, então p é um divisor de n e de $p_1 p_2 \cdots p_r$. Concluímos que p é um divisor de $1 = n - p_1 p_2 \cdots p_r$, uma contradição. Assim, p é um primo diferente de $p_1, p_2, \ldots, p_r$. □

Assim, os números primos formam uma sucessão ilimitada

$$2, 3, 5, 7, 11, 13, \ldots, p_n, \ldots$$

sobre a qual a mais elementar curiosidade sugere algumas perguntas simples. Por exemplo, é possível determinar uma fórmula explícita, envolvendo o natural n, que permita calcular o primo p_n? A resposta parece ser negativa, e em particular, não se conhece nenhuma fórmula explícita que produza *apenas* números primos. A título de exemplo, descrevemos aqui uma das mais famosas tentativas nesta direcção, devida a Fermat[3], e que envolveu os números da forma

$$F_n = 2^{2^n} + 1,$$

3 Pierre de Fermat (1601-1665), matemático francês. Fermat, advogado de profissão, é um dos personagens mais interessantes da história da Matemática. Foi um dos fundadores do Cálculo Infinitesimal, e descobriu independentemente de Descartes (de quem aliás foi amigo) os princípios da Geometria Analítica. O seu trabalho mais importante foi sem dúvida a criação da moderna Teoria dos Números.

hoje conhecidos por **números de Fermat**. A explicação para a escolha do expoente $e = 2^n$ é dada no seguinte exercício:

Exercício 7. Mostre que se $F = 2^e + 1$ é primo então o expoente e não tem factores ímpares maiores do que 1.
(SUGESTÃO: Assuma que $e = ks$, onde $k, s > 1$, com s ímpar. Verifique que o polinómio $p(x) = x^s + 1$ tem a raiz $x = -1$, logo factoriza-se

$$p(x) = (x + 1)q(x),$$

onde $q(x)$ é um polinómio com coeficientes inteiros. Conclua que neste caso,

$$F = 2^e + 1 = (2^k)^s + 1 = (2^k + 1)q(2^k),$$

não é primo.)

É fácil calcular os números de Fermat correspondentes a $n = 0$, 1, 2 e 3, obtendo-se respectivamente 3, 5, 17, e 257, todos eles primos. A escolha $n = 4$ corresponde a 65537, que é ainda um número primo. Há no entanto números de Fermat que não são primos, como Euler[4] descobriu em 1732 para $n = 5$. Se esta lhe parece uma observação simples de demonstrar, note que $n = 5$ corresponde a

$$2^{2^5} + 1 = 4\ 294\ 967\ 297,$$

e Euler descobriu que a factorização deste número em primos é

$$2^{2^5} + 1 = 641 \cdot (6\ 700\ 417).$$

Apesar do começo "auspicioso" da sucessão de Fermat, não conhecemos números de Fermat com $n > 4$ que sejam primos, e sabemos que alguns desses números são compostos. Sabemos por exemplo que o menor factor primo do número de Fermat correspondente a $n = 1945$ (número esse com mais de 10^{582} dígitos na sua expansão decimal!) é um número primo p de 587 dígitos: $p = 5 \cdot 2^{1947} + 1$, e julga-se que nenhum dos números de Fermat com $n > 4$ é primo. Apesar disso, como se mostra no exercício seguinte, estes números podem ser usados para provar a existência de um número infinito de primos.

Exercício 8. Mostre que quaisquer dois números de Fermat F_n e F_m, com $n \neq m$, são primos entre si. Conclua que o número de primos é infinito.
(SUGESTÃO: Verifique, por indução, a fórmula:

$$F_0 F_1 \cdots F_{n-1} = F_n - 2,$$

válida para $n \geq 1$.)

4 Leonhard Euler (1707-1783), matemático suíço. Euler foi um dos mais prodigiosos matemáticos de sempre, tendo trabalhado nas mais diversas áreas da Matemática Pura e Aplicada (análise, geometria, geometria diferencial, teoria dos números, física, mecânica dos fluidos, e outras).

Embora a sucessão formada pelos números primos seja infinita, podemo-nos questionar sobre a *distribuição dos primos* no conjuntos de todos os números naturais. Uma primeira observação é que a distância entre dois primos consecutivos não é limitada:

TEOREMA 15. *Dado um natural $n \in \mathbb{N}$ existem dois primos consecutivos p_i e p_{i+1} tais que $p_{i+1} - p_i > n$.*

Demonstração. Seja $n \in \mathbb{N}$ um natural, e designemos por $N = 2 \cdot 3 \cdot 5 \cdots p$ o produto de todos os números primos menores que $n + 2$. Nenhum dos números naturais:

$$N + 2,\ N + 3,\ N + 4, \ldots, N + (n + 1),$$

é primo. De facto, se $2 \leq i \leq n + 1$, temos que i possui um factor primo p que é menor do que $n + 2$, e esse factor primo também é um divisor de N. Logo p é um divisor de $N + i$, que portanto não é primo. □

Exemplo 16.

Se tomarmos $n = 11$, o número N da demonstração é:

$$N = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 = 30030,$$

logo concluímos que nenhum dos 11 números da lista:

$$30032,\ 30033,\ , \ldots, 30042$$

é um número primo.

Por outro lado, também existem majorantes para o tamanho dos intervalos entre dois primos consecutivos. O majorante mais famoso é o que é dado pela receita "o próximo número primo não pode estar a maior distância do que o número aonde começámos a procurar". Mais exactamente temos:

TEOREMA 17 (POSTULADO DE BERTRAND). *Para todo o natural $n \in \mathbb{N}$, existe um primo p tal que $n < p < 2n$.*

Não iremos demonstrar aqui este resultado. Este resultado foi conjecturado por Bertrand[5] que o verificou empiricamente para todo o $n < 3000000$!

5 Joseph Bertrand (1822-1900), matemático francês, que leccionou nas duas mais prestigiadas escolas de Paris, a École Polytéchnique e o Collège de France. Os seus trabalhos incidiram sobre a geometria das curvas e das superfícies, a Teoria das Probabilidades e as aplicações das equações diferenciais à Mecânica e à Termodinâmica.

A primeira demonstração é devida a Chebyshev[6] e remonta a 1850.

Outra questão também directamente relacionada com o problema da determinação duma expressão explícita para o n-ésimo primo, é a de saber a probabilidade de um número natural escolhido ao acaso no intervalo $[1, n]$ ser primo. Legendre[7] e Gauss[8] foram os primeiros matemáticos a sugerir uma expressão *aproximada* para o número de primos $< x$, que designamos por $\pi(x)$. Nos finais do século *XIX*, Hadamard [9] foi o primeiro a demonstrar que

$$\frac{\pi(x)}{\frac{x}{\log x}} \to 1 \text{ quando } x \to \infty,$$

um resultado conhecido pelo nome de "Teorema do Número Primo". Não discutiremos aqui resultados desta natureza, que tipicamente requerem técnicas de Análise para a sua demonstração (ver Secção 3).

O Teorema do Número Primo está longe de ser a palavra final sobre a distribuição dos primos. Uma resolução da chamada Hipótese de Riemann, um dos grandes problemas em aberto em Matemática, conduziria a ainda melhores estimativas. Mesmo em relação ao postulado de Bertrand é de esperar melhoramentos significativos. Por exemplo, um problema em aberto é:

- Dado um natural $n \in \mathbb{N}$, existe um número primo entre n^2 e $(n+1)^2$?

6 Pafnuti Chebyshev (1821-1894), foi um dos matemáticos mais proeminentes que viveram na Rússia, na segunda metade do século XIX e é considerado um fundador da escola de Matemática de São Petersburgo. Os seus trabalhos desenvolveram-se em variados campos, tais como a Teoria dos Números, problemas de aproximação, integração, geometria diferencial, probabilidades, etc.

7 Adrien Marie Legendre (1752-1833), matemático francês. Legendre distinguiu-se na Geometria e na Teoria dos Números. Foi ele também que criou o método dos mínimos quadráticos. O trabalho mais importante de Legendre foi o seu estudo sistemático das funções elípticas, embora este tenha sido rapidamente ultrapassado pelos trabalhos posteriores de Abel e Jacobi.

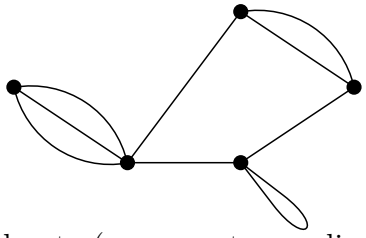
8 Carl Friedrich Gauss (1777-1855) foi um dos grandes matemáticos de Göttingen. Foi uma criança prodígio, e com apenas 19 anos descobriu um método de construção dum polígono regular de 17 lados usando exclusivamente régua e compasso. Durante mais de 2000 anos, desde os geómetras gregos, os únicos polígonos regulares com um número primo de lados que se sabia construir com régua e compasso eram o triângulo equilátero e o pentágono regular.

9 Jacques Hadamard (1865-1963), foi um dos matemáticos franceses mais influentes do virar dos séculos *XIX* e *XX*. Embora o seu trabalho mais conhecido seja a demonstração do Teorema do Número Primo, Hadamard trabalhou em domínios muito diferentes da Matemática (*e.g.*, na teoria do números e no cálculo de variações).

2 A Galeria de Arte

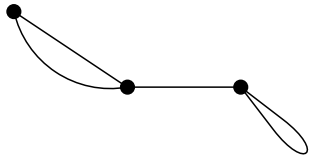
2.1 Grafos

Os grafos são estruturas matemáticas elementares e aparecem nos mais variados contextos. De forma abstracta, um **grafo** G é formado por um conjunto de **vértices** V e um conjunto de **arestas** E([10]). Cada aresta $e \in E$ "liga" dois vértices $v, w \in V$. Por exemplo, o grafo seguinte possui 5 vértices e 9 arestas:



Neste grafo, existe um lacete (uma aresta que liga um vértice a si próprio), uma aresta dupla (duas arestas que ligam os mesmo vértices) e uma aresta tripla (três arestas que ligam os mesmos vértices). Neste curso, vamos apenas considerar grafos com um número *finito* de arestas e vértices, embora os grafos em que esse número é infinito também sejam interessantes.

Um grafo $G' = (V', E')$ diz-se um **subgrafo** dum grafo $G = (V, E)$ se $V' \subset V$, $E' \subset E$, e cada aresta $e \in E'$ une os mesmos vértices em G' que une em G. Por exemplo, o seguinte grafo é um subgrafo do grafo acima:

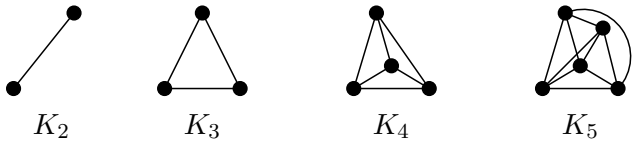


Na maior parte dos casos, vamos estar interessados em **grafos simples**, *i.e.*, grafos que não possuem nem lacetes, nem arestas múltiplas. Os exemplos seguintes ilustram algumas classes importantes de grafos simples.

10 É comum usar-se a letra E para designar o conjunto das arestas porque na língua inglesa o termo aresta traduz-se por "edge".

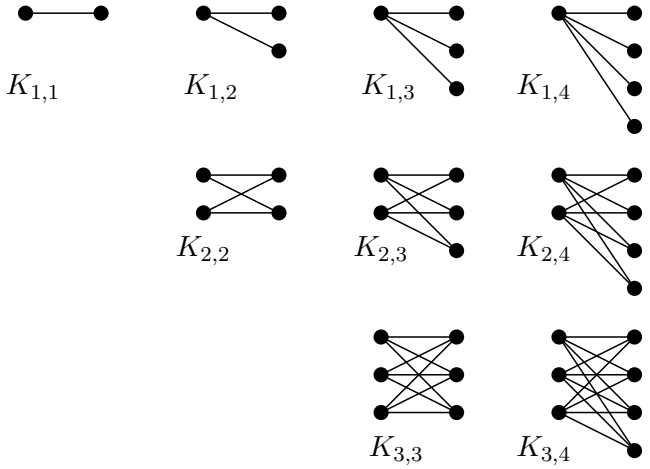
Exemplos 18.

1. *O* ***grafo completo*** K_n *é o grafo simples com n vértices que possui exactamente uma aresta ligando cada par de vértices:*



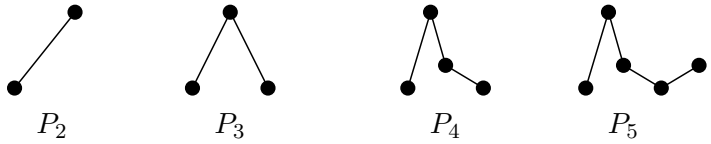
É fácil de ver que o grafo K_n *possui* $\binom{n}{2} = \frac{n!}{2!(n-2)!}$ *arestas.*

2. *Um* ***grafo bipartido*** *é um grafo em que o conjunto dos vértices é a união de dois conjuntos* $V = V_1 \cup V_2$ *e em que todas as arestas ligam vértices de* V_1 *a vértices de* V_2*. Por exemplo, para um conjunto* V_1 *de n vértices e um conjunto* V_2 *de m vértices, temos o* ***grafo bipartido completo*** $K_{n,m}$*, que exemplificamos na figura seguinte:*



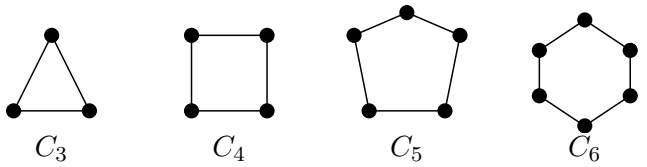
O grafo $K_{n,m}$ *possui* $n + m$ *vértices e* mn *arestas.*

3. *Exemplos muito simples de grafos são fornecidos pelos* ***caminhos*** P_n *com n vértices:*

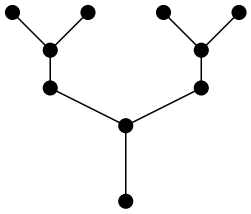


Um grafo G diz-se ***conexo*** *se todos o par de vértices de G pode ser ligado por um caminho.*

4. Outros exemplos muito simples de grafos são os ***ciclos*** C_n *com n vértices:*



5. Um grafo diz-se uma ***floresta*** *se não contém ciclos (*i.e.*, se não possui subgrafos que sejam ciclos). A uma floresta conexa chama-se uma* ***árvore****. Por exemplo, o grafo seguinte é uma árvore:*



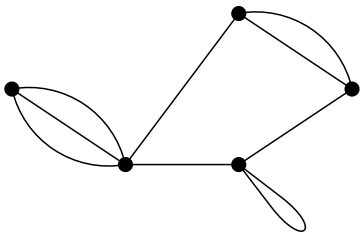
Exercício 9. Mostre que numa árvore o número de vértices é igual ao número de arestas mais um.

Dois grafos $G = (V, E)$ e $G' = (V', E')$ dizem-se **isomorfos** se existirem bijecções $V \leftrightarrow V'$ e $E \leftrightarrow E'$ que preservam as relações de incidências entre arestas e os vértices que ligam. Não distinguimos dois grafos que sejam isomorfos (e, por isso, faz sentido falar *no* grafo K_5, etc.).

Um **grafo plano ou planar** é um grafo que pode ser desenhado no plano sem que duas arestas se cruzem. O primeiro grafo que desenhámos no início é um grafo plano. Os grafos C_n e P_n são grafos planos, bem como os grafos completos K_2, K_3 e K_4. Veremos mais adiante que K_5 não é um grafo plano (e, portanto, K_n para $n \geq 5$ não são grafos planos).

Exercício 10. Mostre que um grafo G é plano se, e só se, o grafo pode ser desenhado numa esfera sem que duas arestas se cruzem.

Um grafo plano $G = (V, E)$ decompõe o plano (ou a esfera) num número finito de regiões conexas (incluindo a região ilimitada, exterior ao grafo) a que se chamam as **faces** do grafo. Por exemplo, o nosso grafo inicial:

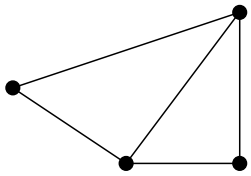


possui 6 faces.

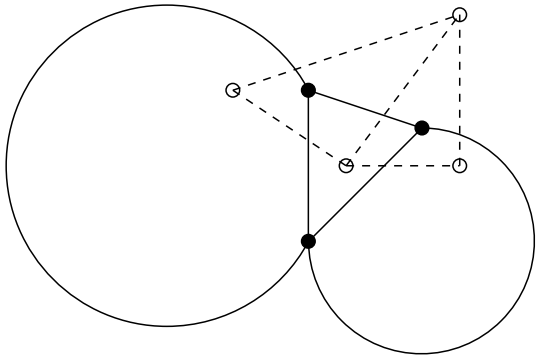
O **grafo dual** do grafo plano G é o grafo plano G^* que se obtém da seguinte forma: no interior de cada face de G escolhemos exactamente um ponto, obtendo desta forma os vértices V^* de G^*. Unimos dois vértices $v, w \in V^*$ de G^* por uma aresta, sempre que existe uma aresta de G a separar as duas faces de G a que pertencem os vértices v e w.

Exemplo 19.

O grafo:



possui como grafo dual o grafo:



Note-se que o grafo dual de um grafo simples pode não ser um grafo simples!

2.2 A Fórmula de Euler

Euler descobriu uma fórmula que exibe uma relação muito bela entre o número de vértices, de arestas e de faces de um grafo planar:

TEOREMA 20 (FÓRMULA DE EULER). *Se G é um grafo plano conexo com* **v** *vértices,* **e** *arestas e* **f** *faces, então:*

$$v - e + f = 2. \tag{1}$$

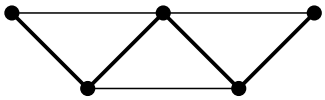
Demonstração. Seja $G = (V, E)$ um grafo plano, Escolhemos um subgrafo $A \subset G$ que é minimal entre todos os subgrafos de G que ligam todos os

vértices de G (minimal significa que não existe nenhum subgrafo de A que liga todos os vértices de G).

Vejamos que A é uma árvore:

- A é conexo, pois liga todos os vértices de G.
- A não contém ciclos; de facto, se A contivesse um ciclo poderíamos eliminar uma das arestas deste ciclo, obtendo um subgrafo que ainda liga todos os vértices de G, *i.e.*, A não seria minimal.

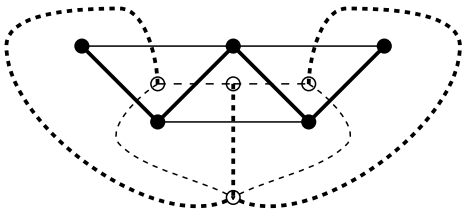
Por exemplo, no grafo seguinte marcámos a negro uma tal árvore A (note que esta árvore pode não ser única):



Recorrendo ao Exercício 9, concluímos que:

$$v = e_A + 1. \tag{2}$$

Vejamos agora o grafo dual G^* de G. No grafo G^* vamos considerar o subgrafo $A' \subset G^*$ que consiste nas arestas de G^* que correspondem às arestas em $G - A$ (recorde a construção do grafo dual). Por exemplo, para o grafo acima, o grafo dual G^* (a tracejado) e o subgrafo A' (a tracejado forte) são:



Observe que o subgrafo $A' \subset G^*$ também é uma árvore que liga todos os vértices de G^*. De facto, se A' contivesse um ciclo então esse ciclo dividiria os vértices de G em vértices interiores e exteriores a esse ciclo, o que significaria que A não era conexo.

Mais uma vez recorrendo ao Exercício 9, e observando que os vértices de G^* (logo de A') são as faces de G, concluímos que:

$$f = e_{A'} + 1. \tag{3}$$

Adicionando as equações (2) e (3) concluímos que:

$$v + f = e_A + e_{A'} + 2 \quad \Longleftrightarrow \quad v - e + f = 2,$$

pois temos $e = e_A + e_{A'}$. □

A fórmula de Euler produz um número a partir de uma situação topológica ou geométrica: sempre que um grafo está (ou pode ser desenhado) num plano, a soma alternada do número de vértices, arestas e faces é igual a 2. Esta fórmula impõe restrições muito fortes num grafo e, por isso, tem muitas consequências interessantes. Por exemplo, a fórmula de Euler é um dos ingredientes básicos para demonstrar os seguintes resultados famosos:

- **Classificação dos poliedros regulares.** Existem exactamente cinco poliedros regulares convexos (os sólidos platónicos).

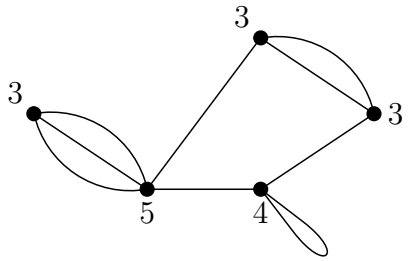
- **Teorema das quatro cores.** Todo o mapa plano pode ser colorido com exactamente quatro cores, sem que dois países fronteiros fiquem com a mesma cor.

Não iremos aqui discutir as demonstrações destes resultados, pois estão para além do âmbito deste curso, mas veremos nas próximas secções como a fórmula de Euler pode ser utilizada noutros contextos.

2.3 Mais Grafos...

Para verificar que, por exemplo, o grafo completo com 5 vértices K_5 não é plano, recorrendo à fórmula de Euler, necessitamos de mais duas noções elementares associadas a um grafo.

Primeiro, dado um grafo G qualquer, o **grau de um vértice** é o número de arestas que terminam nesse vértice (onde os lacetes contam duas vezes). No grafo seguinte em cada vértice está assinalado o seu grau:



Se v_i é o número de vértices com grau i do grafo G, então o número total de vértices de G é dado por:

$$v = v_1 + v_2 + v_3 + \cdots \tag{4}$$

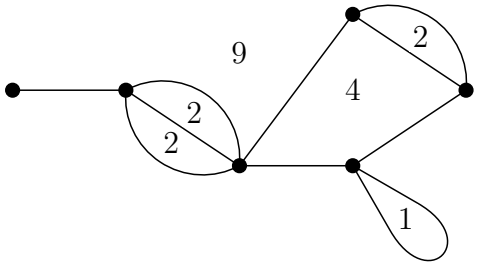
Por outro lado, o grau também está relacionado com o número de arestas, como se pede para mostrar no seguinte exercício.

Exercício 11. Mostre que:

$$2e = v_1 + 2v_2 + 3v_3 + \cdots \tag{5}$$

e conclua que o *grau médio* dos vértices é $\frac{2e}{v}$.

A noção de vértice, e as equações (4) e (5), é válida para qualquer grafo (plano ou não). Para grafos planos, podemos definir ainda uma outra noção básica: o **número de lados** duma face é o número de arestas que limitam essa face (em que as arestas que limitam a mesma face contam duas vezes). No grafo seguinte, em cada face está assinalado o seu número de lados:



Se f_i é o numero de faces com i lados no grafo plano G, então o número total de faces do grafo G é dado por:

$$f = f_1 + f_2 + f_3 + \cdots \tag{6}$$

Por outro lado, o número de lados das faces também está relacionado com o número de arestas, como se pede para mostrar no seguinte exercício.

Exercício 12. Mostre que:

$$2e = f_1 + 2f_2 + 3f_3 + \cdots \tag{7}$$

e conclua que o *número médio de lados* das faces é $\frac{2e}{f}$.

Proposição 21. *O grafo K_5 não é planar.*

Demonstração. K_5 é um grafo com $v = 5$ vértices e $e = \binom{5}{2} = 10$ arestas. Se K_5 fosse plano, o número de faces seria dado pela fórmula de Euler:

$$f = e - v + 2 = 7.$$

Assim, pelo Exercício 12, o número médio de lados das faces seria $2e/f = 20/7 < 3$. Ou seja, existiria uma face com dois lados, o que não pode acontecer, pois K_5 é um grafo simples. □

A mesma técnica de aplicação da fórmula de Euler pode ser usada para verificar que outros tipos de grafos não são planares. Por exemplo, não deverá ser difícil ao leitor resolver o seguinte exercício:

Exercício 13. Mostre que o grafo bipartido completo $K_{3,3}$ não é planar.

Outras consequências importantes da fórmula de Euler são dadas na seguinte proposição.

PROPOSIÇÃO 22. *Seja G um grafo plano simples. Então:*

(i) Existe um vértice de G com grau menor ou igual a 5.

(ii) O número de arestas de G é menor ou igual a $3v - 6$.

Demonstração. Sem perda de generalidade, podemos assumir que G é conexo.

(i) Como o grafo G é simples, cada face tem pelo menos três lados, de forma que as equações (6) e (7) fornecem:

$$f = f_3 + f_4 + f_5 + \cdots$$
$$2e = 3f_3 + 4f_4 + 5f_5 + \cdots$$

donde vemos que $2e - 3f \geq 0$. Se, por absurdo, todos o vértices têm grau maior ou igual a 6, então as equações (4) e (5) fornecem:

$$v = v_6 + v_7 + v_8 + \cdots$$
$$2e = 6v_6 + 7v_7 + 8v_8 + \cdots$$

donde vemos que $2e - 6v \geq 0$.

As desigualdades $2e - 3f \geq 0$ e $2e - 6v \geq 0$ juntas implicam que:

$$6(e - v - f) = (2e - 6v) + 2(2e - 3f) \geq 0.$$

o que contradiz a fórmula de Euler.

(ii) Tal como na demonstração de (i), vemos que $2e - 3f \geq 0$. Pela fórmula de Euler, temos:

$$3v - 6 = 3e - 3f = e + (2e - 3f) \geq e.$$

□

O seguinte exercício, um pouco mais ambicioso, também pode ser resolvido recorrendo à fórmula de Euler:

Exercício 14. Seja G um grafo plano simples cujas arestas são coloridas com duas cores. Mostre que existe um vértice de G em que há no máximo duas trocas de cores, quando percorremos ciclicamente as arestas que incidem nesse vértice.

Problemas mais complicados de colorir grafos (por exemplo, com mais cores) são por vezes de difícil resolução, e têm sido tema de investigação desde que o conceito de grafo foi inventado. Por exemplo, o problema de colorir um mapa, sem que dois países fronteiros possuam a mesma cor, pode facilmente ser transformado num problema de colorir um grafo (como?). Durante muito tempo sabia-se que era possível colorir um mapa com cinco cores, e conjecturava-se que seria possível utilizar apenas quatro cores. O Teorema das Quatro Cores foi finalmente demonstrado em 1976 por Appel e Haken, e esteve no centro de uma grande polémica pois foi a primeira demonstração de um teorema de Matemática que recorreu a um computador!

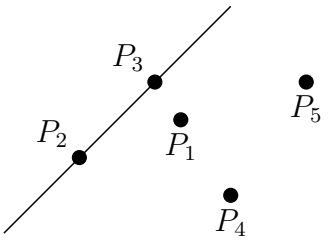
Exercício 15. Recorrendo à Proposição 22 (i), demonstre o Teorema das Seis Cores.
(SUGESTÃO: Utilizando indução no número de vértices, assuma que um grafo plano simples com $n-1$ vértices pode ser colorido com seis cores. Para um grafo com n vértices, existe um vértice com grau menor ou igual a cinco. Remova este vértice e aplique a hipótese de indução.)

2.4 Rectas no Plano

O seguinte resultado foi descoberto por Sylvester [11] em 1893, mas uma demonstração correcta só terá sido descoberta mais de 40 anos depois!

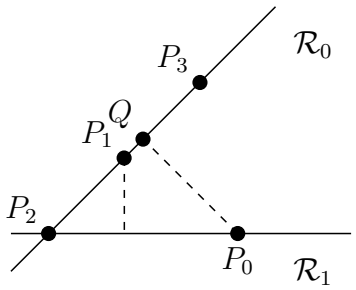
TEOREMA 23 (SYLVESTER). *Dados n pontos no plano, não colineares, existe uma recta que passa por exactamente dois desses pontos.*

11 James Joseph Sylvester (1814-1897) foi um matemático inglês, que juntamente com Cayley e Salmon desenvolveram a álgebra com aplicações sobretudo à geometria. Cayley e Sylvester exerceram ambos advocacia em Londres antes de se dedicarem em exclusivo à Matemática. Enquanto que Cayley se fixaria em Cambridge, Sylvester viria a emigrar para os USA onde foi professor na Universidade de Johns Hopkins, em Baltimore. É considerado os dos fundadores da escola de Matemática dos EUA.



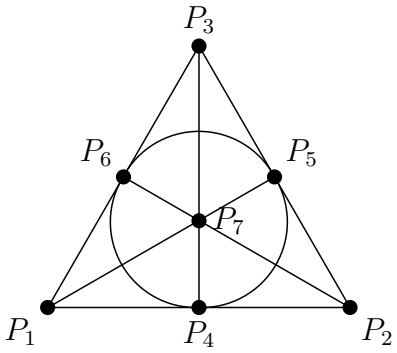
Demonstração. Seja $\mathcal{P} = \{P_1, \ldots, P_n\}$ um conjunto de pontos, não colineares. Entre todos os pares $(P, \mathcal{R})$, onde $\mathcal{R}$ é uma recta que passa por dois pontos de $\mathcal{P}$ e P é um ponto de $\mathcal{P}$ que não pertence a $\mathcal{R}$, escolhemos um par $(P_0, \mathcal{R}_0)$ tal que a distância de P_0 a $\mathcal{R}_0$ é a mais pequena. Vamos ver que a recta $\mathcal{R}_0$ é a recta procurada.

Assuma-se, por absurdo, que a recta $\mathcal{R}_0$ possui três pontos de $\mathcal{P}$, que designamos por P_1, P_2 e P_3. Seja Q o ponto de $\mathcal{R}_0$ mais próximo de P_0 (ver figura). Observe que (pelo menos) dois pontos de $\{P_1, P_2, P_3\}$ estão no mesmo lado de Q. Podemos assumir que esses pontos são P_1 e P_2, onde P_1 está mais próximo de Q (pode coincidir até com Q).



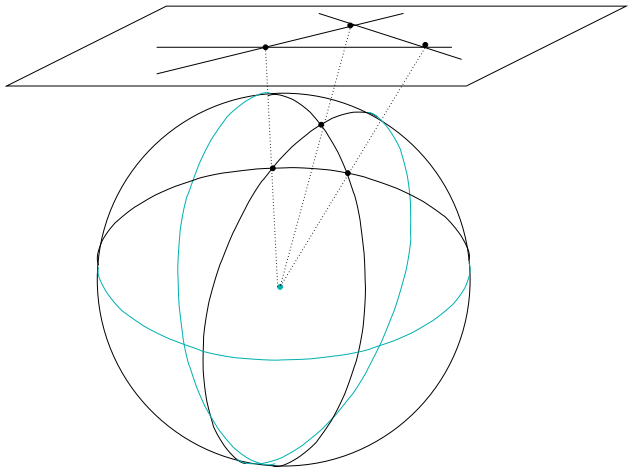
Se $\mathcal{R}_1$ é a recta que passa por P_0 e P_2, então a distância de P_1 a $\mathcal{R}_1$ é mais pequena que a distância de P_0 a $\mathcal{R}_0$, uma contradição. □

Esta demonstração é extremamente elegante. No entanto, utiliza as propriedades métricas do plano ("distância"), e podemo-nos perguntar se tal é de facto necessário. A seguinte configuração, devida a Fano, mostra que em certas geometrias o Teorema de Sylvester não é válido:



Nesta geometria estranha, temos um conjunto de 7 pontos e cada par de pontos determina uma "recta" que passa nesses pontos. Mas todas estas rectas contêm exactamente três pontos dessa configuração (incluindo a "recta" $\{P_4, P_5, P_6\}$). Não iremos estudar nenhuma destas geometrias *não-euclidianas*, mas elas desempenham um papel fundamental na Matemática contemporânea, e têm aplicações muito importantes, por exemplo, na Física.

Existe uma outra demonstração do Teorema de Sylvester que recorre à fórmula de Euler, e que torna mais explícita a ligação entre a geometria e as relações de incidência das rectas determinadas pela configuração de pontos. Para isso, identificamos o plano com a esfera utilizando projecção estereográfica pelo centro da esfera:



A cada ponto no plano corresponde um par de pontos antipodais na esfera, de forma que o Teorema de Sylvester pode ser reescrito na seguinte forma:

- Dados $n \geq 3$ pares de pontos antipodais na esfera, que não pertencem todos ao mesmo círculo máximo, existe sempre um círculo máximo que contém exactamente dois pares de pontos antipodais.

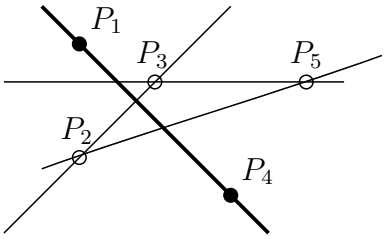
Agora observamos que a cada par de pontos antipodais corresponde o círculo máximo que é formado pelo pontos equidistantes ao par (*e.g.*, no caso do par de pontos ser formado pelos pólos Norte e Sul, o círculo máximo correspondente é o equador). Reciprocamente, dado um círculo máximo existe um único par de pontos antipodais aos quais os pontos do círculo são equidistantes. Desta forma, podemos dualizar o problema transformando pares de pontos antipodais em círculos máximos, e círculos máximos em pares de pontos antipodais. Assim, temos uma nova reformulação do Teorema de Sylvester:

- Dados $n \geq 3$ círculos máximos na esfera, que não passam todos pelo mesmo ponto, existe sempre um ponto que pertence exactamente a dois círculos máximos.

Para demonstrar esta versão do Teorema, recorrendo à fórmula de Euler, notamos que os círculos máximos determinam um grafo na esfera, em que todos os vértices têm grau par, e superior ou igual a 4. Pelo item (i) da Proposição 22, concluímos que existe um vértice com grau exactamente igual a 4. Este vértice é o ponto procurado!

Esta demonstração, que recorre à fórmula de Euler, é um pouco mais complexa do que a a primeira demonstração. No entanto, ela permite um olhar mais profundo ao problema. Por exemplo, utilizando as mesmas técnicas podemos obter a seguinte versão "colorida" do Teorema de Sylvester:

TEOREMA 24. *Dada uma configuração qualquer de pontos pretos e brancos no plano, não colineares, existe sempre uma recta "monocromática",* i.e.*, uma recta que contém apenas pontos pretos ou pontos brancos.*



Exercício 16. Repita a demonstração do Teorema de Sylvester, utilizando o Exercício 14 em vez da Proposição 22, obtendo desta forma a demonstração da sua versão "colorida".

Como vimos, os grafos desempenham um papel muito importante nos problemas mais diversos. No entanto, a implementação "prática" de algoritmos relacionados com grafos por vezes apresenta problemas delicados. Um

exemplo de um problema importante, com consequências "práticas", é o de encontrar um método eficiente para determinar o caminho mais curto entre dois vértices dum grafo. Este problema ocorre, por exemplo, quando pretendemos encontrar num mapa de estradas o trajecto mais curto que liga duas cidades. Vários modelos de automóveis, vêm equipados com sistemas GPS que aconselham o condutor sobre o trajecto aconselhado para ir do ponto A ao ponto B, e que recorrem a métodos (não-eficientes) para determinar o caminho mais curto (existem alguns sites na rede que disponibilizam programas semelhantes). É claro que é preciso definir o que se entende por "método eficiente", mas isso levaria-nos para outras águas...

Terminamos esta parte com um problema muito importante e que permanece em aberto:

- Será que existe um teste eficiente que permite decidir se dois grafos são ou não isomorfos?

3 O Hotel de Hilbert

3.1 Conjuntos

A noção de conjunto é a mais importante de todas as noções matemáticas e constitui, por assim dizer, a primeira pedra na Rua das Matemáticas. Todos nós estamos certamente familiarizados com a ideia informal de conjunto e de elemento de um conjunto, bem como com algumas das construções elementares que estes suportam (uniões, intersecções, complementos, etc.). Por outro lado, afirmações tais como:

(i) dois conjuntos são iguais se e só se possuem os mesmos elementos;

(ii) dados dois conjuntos, existe um conjunto que os contém;

(iii) dado um conjunto, existe um conjunto formado por todos os seus subconjuntos;

são normalmente aceites como óbvias. No entanto, para as justificar plenamente seria necessário proceder a uma investigação mais profunda sobre os fundamentos da Teoria dos Conjuntos. Esta investigação está para além do âmbito deste curso, mas não é demais salientar que este é um aspecto muito importante, que a não ser resolvido nos leva rapidamente a grandes contradições como é ilustrado pelo seguinte famoso paradoxo de Russell[12]:

12 Bertrand Russell (1872-1970) foi juntamente com Alfred N. Whitehead (1861-1947) autor do famoso tratado *Principia Mathematica* (3 vols., 1910-13), onde se tentavam formalizar

- Será que existe o conjunto de todos os conjuntos?

O paradoxo vem de que se assumirmos que C é o conjunto de todos os conjuntos então qualquer conjunto é um elemento de C. Em particular, C deveria ser um elemento de C, e isso contradiz a nossa ideia intuitiva do que é um conjunto: C tem a estranha propriedade de ser elemento dele próprio, o que não é usual nos conjuntos que conhecemos! Se nos desagrada esta propriedade de C, podemos considerar em seu lugar o conjunto N dos conjuntos "normais", *i.e.*, dos que não são elementos deles próprios. Em símbolos,

$$N = \{A \in C : A \notin A\}.$$

A pergunta a pôr agora é simples: N é ou não um conjunto "normal"? Infelizmente, se supusermos que N é "normal" (*i.e.*, $N \notin N$) então N pertence ao conjunto dos conjuntos normais (*i.e.*, $N \in N$!). Se supusermos que N não é "normal" (*i.e.*, $N \in N$) então N é um elemento do conjunto dos conjuntos normais e, portanto, N é ele próprio normal (*i.e.*, $N \notin N$!). Por outras palavras, não conseguimos atribuir um valor lógico (verdadeiro ou falso) à afirmação "$N \in N$".

A um nível superficial, a lição a tirar deste exemplo é simplesmente que é necessário algum cuidado com definições recursivas. Mais prosaicamente, a mesma dificuldade surge quando se utilizam folhas de cálculo automático (*spreadsheets*), e se cria um circuito fechado de referências entre células da folha, ou quando se enuncia um "teorema" como o que se segue:

TEOREMA 25. *Esta afirmação é falsa...*

A um nível mais profundo, no entanto, as dificuldades lógicas com definições recursivas, ou mais geralmente com proposições que se referem a elas próprias, parecem inevitáveis e estão relacionadas com alguns dos problemas mais difíceis contemplados por matemáticos e filósofos. É possível dar uma definição (rigorosa!) de "definição rigorosa"? Podemos compreender o funcionamento da nossa própria inteligência? Como podemos conciliar o aspecto mecânico das deduções lógicas, espelhado no funcionamento dum programa de computador, com a infinita adaptabilidade que chamamos comportamento "inteligente"? Afinal de contas, e regressando à vida "prática", este é o problema central do desenvolvimento da Inteligência Artificial.

---

de forma axiomática as noções fundamentais da aritmética. Este trabalho monumental foi o auge de um programa de formalizar a Matemática, a que se poderá chamar "logística", e que consistia em construir toda a Matemática através da dedução lógica a partir de um pequeno número de conceitos e princípios. Embora essa abordagem tenha falhado, devido aos trabalhos posteriores de Gödel, ela deu uma contribuição notável para a Lógica Matemática.

3.2 Funções

Depois da noção de conjunto, a noção de função é a noção mais importante que se segue na Rua das Matemáticas. Na realidade, a noção de função, tal como muitos outros conceitos matemáticos, pode ser reduzida a operações sobre conjuntos.

Para formalizar a noção de função $f : X \to Y$, de um conjunto X para um conjunto Y, procedemos da seguinte forma. Primeiro, formamos o **produto cartesiano** $X \times Y$ do conjuntos X e Y que, por definição, é o conjunto formado pelos pares ordenados (x, y), onde $x \in X$ e $y \in Y$.

Definição 26. Uma **função** é um subconjunto $f \subset X \times Y$ com as seguintes propriedades:

(i) para qualquer $x \in X$ existe $y \in Y$ tal que $(x, y) \in f$, e

(ii) se $(x, y) \in f$ e $(x, y') \in f$, então $y = y'$.

Devido às propriedades (i) e (ii), escrevemos $y = f(x)$ em lugar de $(x, y) \in f$. Dizemos então que X é o **domínio** e Y o **contradomínio** da função f. Outras designações frequentes para um função são as de **aplicação** ou **transformação**.

Exemplos 27.

1. *Se X é um conjunto, $I_X : X \to X$, dada por $I_X(x) = x$, é a* ***função identidade*** *em X.*

2. *Se $Y \subset X$ são conjuntos, $i_Y : Y \to X$, dada por $i_Y(y) = y$, é a* ***função inclusão*** *de Y em X.*

3. *Se X e Y são conjuntos, $p_X : X \times Y \to X$ e $p_Y : X \times Y \to Y$, dadas por $p_X(x, y) = x$ e $p_Y(x, y) = y$, são as* ***projecções*** *em X e Y.*

Em geral, dada uma função $f : X \to Y$ e $X' \subset X$, definimos

$$f(X') \equiv \{f(x) : x \in X'\},$$

i.e., $f(X')$ é o subconjunto de Y formado por todos os elementos obtidos a partir de elementos de X' aplicando a função f. Dizemos então que $f(X')$ é a **imagem directa** de X' por f. Em particular, a **imagem** de f é o conjunto $f(X)$, que se designa também por $\operatorname{Im} f$.

De forma análoga, se $Y' \subset Y$, definimos

$$f^{-1}(Y') \equiv \{x \in X : f(x) \in Y'\},$$

i.e., $f^{-1}(Y')$ é o subconjunto de X formado por todos os elementos cuja imagem por f pertence a Y'. Dizemos então que $f^{-1}(Y')$ é a **imagem inversa** de Y' por f.

Exemplo 28.

Se $f : \mathbb{R} \to \mathbb{R}$ é a função $f(x) = \cos x$, temos então que a sua imagem é $f(\mathbb{R}) = [-1, +1]$. A imagem inversa do conjunto $\{0\}$ é $f^{-1}(\{0\}) = \{\frac{\pi}{2} + n\pi : n \in \mathbb{Z}\}$.

Apesar da analogia, as imagens inversas e as imagens directa têm propriedades distintas, como se mostra no seguinte exercício.

Exercício 17. Se $f : X \to Y$ é uma função e $A, B \subset Y$ são subconjuntos de Y, verifique as identidades

$$f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B) \quad \text{e} \quad f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B).$$

Será que estas identidades ainda são válidas se se supusermos que $A, B \subset X$ são subconjuntos de X e substituirmos f^{-1} por f?

Como bem sabemos, certos tipos de funções merecem qualificativos especiais:

DEFINIÇÃO 29. Seja $f : X \to Y$ uma função. Então:

(i) f diz-se **sobrejectiva** se para qualquer $y \in Y$ existe $x \in X$ tal que $y = f(x)$;

(ii) f diz-se **injectiva** se $f(x) = f(x') \Leftrightarrow x = x'$.

(iii) f diz-se **bijectiva** (ou uma **bijecção**) se f é injectiva e sobrejectiva.

Exemplos 30.

1. *A identidade $I_X : X \to X$ é bijectiva.*
2. *A inclusão $i_Y : Y \to X$ é injectiva.*
3. *As projecções $p_X : X \times Y \to X$ e $p_Y : X \times Y \to Y$ são sobrejectivas.*
4. *A função $\cos : \mathbb{R} \to \mathbb{R}$ nem é injectiva nem é sobrejectiva.*

Dadas funções $f : X \to Y$ e $g : Y \to Z$, a **composição** de f e g é a função $g \circ f : X \to Z$, lida "g após f", dada por $(g \circ f)(x) = g(f(x))$. Deixamos como exercício a verificação das seguintes propriedades elementares da composição de funções.

Exercício 18. Sejam X, Y, Z, e W conjuntos. Então:

(i) *Associatividade*: Se $f : X \to Y$, $g : Y \to Z$ e $h : Z \to W$ são funções, então $(h \circ g) \circ f = h \circ (g \circ f)$;

(ii) *Inversa à esquerda*: $f : X \to Y$ é injectiva se e só se existe $g : Y \to X$ tal que $g \circ f = I_X$;

(iii) *Inversa à direita*: $f : X \to Y$ é sobrejectiva se e só se existe $g : Y \to X$ tal que $f \circ g = I_Y$;

(iv) *Inversa*: $f : X \to Y$ é bijectiva se e só se existe $g : Y \to X$ tal que $f \circ g = I_Y$ e $g \circ f = I_X$. Neste caso, g diz-se a **inversa** de f e designa-se por f^{-1}.

Se dados dois conjuntos X e Y existe uma bijecção $f : X \to Y$ então dizemos que os conjuntos são **equipotentes** ou **isomorfos**. Veremos já na próxima secção como esta noção pode ser utilizada para comparar o "tamanho" de dois conjuntos.

3.3 Conjuntos Finitos

A nossa intuição diz-nos que um conjunto X é finito se os seus elementos podem ser "contados". O protótipo dum conjunto finito com $n \geq 0$ elementos é dado pelo conjunto dos primeiros n naturais:

$$I_n = \{1, 2, 3, \ldots, n\} = \{k \in \mathbb{N} : k \leq n\}.$$

Note que, se $n = 0$, obtemos o conjunto vazio: $I_0 = \emptyset$. A "contagem" aqui referida consiste claramente no estabelecimento de uma correspondência (função) bijectiva entre X e I_n. Mais formalmente, temos:

Definição 31. O conjunto X diz-se **finito** se é equipotente a I_n, para algum $n \geq 0$. Se X não é equipotente a nenhum I_n, então X diz-se **infinito**.

Exemplos 32.

1. *O conjunto I_n é evidentemente equipotente a si próprio, logo é finito.*

2. *O conjunto dos naturais $\mathbb{N}$ e o conjunto dos inteiros $\mathbb{Z}$, são conjuntos infinitos. Um subconjunto $X \subset \mathbb{Z}$ é finito se e só se é limitado (exercício!).*

Exercício 19. Se $\phi : I_n \to I_n$ é injectiva, então ϕ é sobrejectiva.

Utilizando este exercício, é fácil de mostrar que:

PROPOSIÇÃO 33. *Se X é finito e $\phi : X \to X$ é injectiva, então ϕ é sobrejectiva.*

Demonstração. Seja $\Psi : I_n \to X$ uma bijecção, e note-se que $\phi^* = \Psi^{-1} \circ \phi \circ \Psi : I_n \to I_n$ é injectiva, por ser uma composição de funções injectivas. De acordo com o Exercício 19, ϕ^* é necessariamente sobrejectiva.

Segue-se que $\phi = \Psi \circ \phi^* \circ \Psi^{-1}$ é uma composição de funções sobrejectivas, e consequentemente é sobrejectiva. □

Como consequência imediata, temos os seguintes corolários:

COROLÁRIO 34. *Se $\phi : X \to X$ é injectiva e não-sobrejectiva, então X é infinito.*

COROLÁRIO 35. *Se X é finito, $X \supseteq Y$ e $\phi : X \to Y$ é injectiva, então $X = Y$.*

Por outras palavras, nenhum conjunto finito pode ser equipotente a um seu subconjunto *estrito*.

Exemplo 36.

A função $f : \mathbb{N} \to \mathbb{N}$ dada por $f(n) = n + 1$ é injectiva e não-sobrejectiva. De acordo com o resultado anterior, concluímos que $\mathbb{N}$ é infinito.

Parece óbvio que X é equipotente a I_n se e só se X tem n elementos, e que neste caso X não pode ser equipotente a I_m, para $m \neq n$. Na realidade, esta afirmação é uma consequência lógica directa da Proposição 33.

COROLÁRIO 37. *Se $\phi : I_n \to X$ e $\psi : I_m \to X$ são bijectivas, então $n = m$.*

Demonstração. Supomos sem perda de generalidade que $m \leq n$, ou seja, $I_m \subset I_n$, e notamos que $\Psi = \psi^{-1} \circ \phi : I_n \to I_m$ é uma função injectiva de I_n num seu subconjunto I_m. Segue-se, do Corolário 35, que $I_m = I_n$, *i.e.*, $n = m$. □

Assim, se X é finito, existe um *único* inteiro não-negativo n tal que X é equipotente a I_n. Dizemos neste caso que X tem n elementos, e designamos o número de elementos do conjunto finito X pelo símbolo $\#X$, dito o **cardinal** de X. Deve-se observar que se Y é subconjunto do conjunto X, então:

(i) Se X é finito, então Y é igualmente finito e $\#Y \leq \#X$.

(ii) Se X é finito e $\#Y = \#X$, então $X = Y$.

(iii) Se Y é infinito, então X é igualmente infinito.

Sugerimos ao leitor que tente demonstrar estas afirmações, que mostram como a noção de cardinalidade corresponde à nossa noção intuitiva de número de elementos de um conjunto. Temos ainda as seguintes propriedades, que também podem ser consideradas "intuitivas":

Exercício 20. Sejam X e Y conjuntos finitos. Mostre que:

(a) $X \cup Y$ é finito e $\#(X \cup Y) = \#X + \#Y - \#(X \cap Y)$.

(b) $X \times Y$ é finito e $\#(X \times Y) = (\#X)(\#Y)$.

3.4 Conjuntos Contáveis

A teoria da cardinalidade torna-se particularmente interessante (e não intuitiva) quando consideramos conjuntos infinitos. Vamos dizer que um conjunto infinito X é **contável** se existir uma bijecção de X com o conjunto dos naturais $\mathbb{N} = \{1, 2, 3, \dots\}$. Assim, X é contável se o pudermos enumerar:

$$X = \{x_1, x_2, x_3, \dots\},$$

onde utilizámos a bijecção $\phi : \mathbb{N} \to X$ para definir $x_i = \phi(i)$. Mas agora acontece um fenómeno estranho: se ao conjunto contável X adicionarmos um novo elemento y o conjunto $X \cup \{y\}$ ainda é contável! De facto, uma enumeração deste novo conjunto é dada por:

$$X \cup \{y\} = \{y, x_1, x_2, x_3, \dots\}.$$

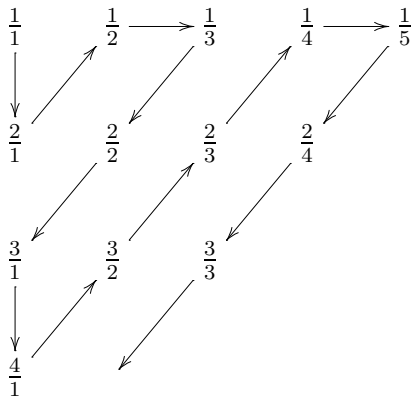
Exemplo 38. O facto de que os conjuntos $\mathbb{N}$ e $\mathbb{N} \cup \{x\}$ são equipotentes pode ser ilustrado pelo famoso "Hotel de Hilbert"[13]. Suponha-se que um hotel possui um número contável de quartos, numerados $1, 2, 3, \dots$. Um turista chega ao hotel e pede um quarto, mas é informado pelo empregado que o hotel está cheio e não há quartos disponíveis. O turista (que é matemático!) diz: "Não há problema. Basta pedir a todos os hóspedes para avançarem um quarto: o ocupante do quarto 1 passa para o quarto 2, o ocupante do quarto 2 passa para o quarto 3, e assim sucessivamente. Eu ocuparei o quarto 1, que desta forma ficou livre!".

13 David Hilbert (1862-1943), foi um dos maiores matemáticos de todos os tempos. Professor em Göttingen, tinha uma grande influência. A comunicação de Hilbert ao Congresso Internacional de Matemáticos em Paris (1900) incluía uma lista de 23 problemas que ele achava que deveriam ser considerados pelos matemáticos do século *XX*. Muitos desses problemas marcaram de facto a Matemática dos últimos 100 anos, e alguns deles permanecem ainda em aberto.

Outro exemplo de um conjunto contável é dado pelo conjunto dos números inteiros $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$, que pode ser enumerado, por exemplo, na forma:

$$\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \dots\}.$$

Um facto, talvez um pouco mais surpreendente, é que o conjunto dos racionais $\mathbb{Q}$ é contável. Para isso, basta ver que o racionais positivos $\mathbb{Q}_+$ são contáveis (porquê?), e para estes uma enumeração pode ser obtida escrevendo os racionais positivos num diagrama



e eliminando as repetições:

$$\mathbb{Q}_+ = \{1, 2, \frac{1}{2}, \frac{1}{3}, 3, 4, \frac{3}{2}, \frac{2}{3}, \frac{1}{4}, \frac{1}{5}, \dots\}.$$

O mesmo tipo de raciocínio permite resolver o seguinte exercício.

Exercício 21. Mostre que a união de um número contável de conjuntos contáveis $C_1, C_2, \dots$ é um conjunto contável.

Em contrapartida temos o seguinte resultado famoso:

TEOREMA 39. *O conjunto dos números reais $\mathbb{R}$ não é contável.*

Demonstração. É fácil de ver que um subconjunto de um conjunto contável é contável. Assim, para verificar que $\mathbb{R}$ não é contável basta verificar que o intervalo $(0, 1]$ não é contável.

Suponhamos, por absurdo, que o intervalo $(0, 1]$ era contável:

$$(0, 1] = \{x_1, x_2, x_3, \dots\}.$$

Podemos escrever cada um destes elementos na sua expansão decimal infinita:

$$x_n = 0, a_{n1}a_{n2}a_{n3}\cdots$$

onde $a_{ni} \in \{0, 1, 2, \ldots, 9\}$. Por exemplo,

$$0,53 = 0,52999\ldots$$

Observe que cada elemento de $(0, 1]$ tem uma expansão única desta forma.

Assim, podemos escrever:

$$\begin{aligned} x_1 &= 0, a_{11}a_{12}a_{13}a_{14}\cdots \\ x_2 &= 0, a_{21}a_{22}a_{23}a_{24}\cdots \\ &\vdots \quad\quad \vdots \\ x_n &= 0, a_{n1}a_{n2}a_{n3}a_{n4}\cdots \\ &\vdots \quad\quad \vdots \end{aligned}$$

Para cada n, escolhemos $r_n \in \{1, 2, \ldots, 8\}$ diferente de a_{nn}. Observe que o número:

$$r = 0, r_1r_2r_3\ldots,$$

pertence ao intervalo $(0, 1]$ e não coincide com nenhum dos x_i, uma contradição! □

3.5 Conjuntos Infinitos

Em geral, dados conjuntos X e Y, dizemos que X e Y têm a mesma cardinalidade se são equipotentes (*i.e.*, se existe uma bijecção entre X e Y). Nesse caso, escrevemos $\#X = \#Y$.

Por exemplo, já sabemos que

$$\#\mathbb{N} = \#\mathbb{Z} = \#\mathbb{Q} \neq \#\mathbb{R}.$$

Começamos agora por verificar que, em certo sentido, $\mathbb{N}$ é o mais pequeno conjunto infinito.

Teorema 40. *X é infinito se e só se X contém um subconjunto Y equipotente a $\mathbb{N}$.*

Demonstração. Se existe um subconjunto Y de X e uma bijecção $\phi : \mathbb{N} \to Y$, segue-se que Y é infinito, e portanto X é infinito.

Suponha-se agora que X é infinito, e mostremos que existe uma função injectiva $\phi : \mathbb{N} \to X$ (o conjunto Y será então $Y = \phi(\mathbb{N})$). A função ϕ é definida recursivamente:

- Como $X \neq \emptyset$, existe $x_1 \in X$, e definimos $\phi(1) = x_1$.

- Suponha-se agora que ϕ está definida e é injectiva em $\{1, 2, \ldots, n\}$. Consideramos o conjunto $Z_n = X - \{\phi(1), \ldots, \phi(n)\}$, e observamos que $Z_n \neq \emptyset$ (já que caso contrário X teria n elementos). Sendo z um qualquer elemento de Z_n, definimos $\phi(n+1) = z$.

Assim, concluímos que existe uma função $\phi : \mathbb{N} \to X$ que, por definição, é injectiva. □

É usual designar o cardinal de $\mathbb{N}$ por $\aleph_0$ (pronuncia-se "alefa zero"). É pois natural perguntar qual é o próximo cardinal a seguir a $\aleph_0$. Para isso, precisamos de comparar cardinais, e as funções são outra vez a chave:

DEFINIÇÃO 41. Dizemos que o cardinal $\mathfrak{m}$ é menor ou igual ao cardinal $\mathfrak{n}$, e escrevemos $\mathfrak{m} \leq \mathfrak{n}$, se existem conjuntos M e N com $\#M = \mathfrak{m}$ e $\#N = \mathfrak{n}$, e uma aplicação injectiva $\phi : M \to N$.

Note que a relação $\mathfrak{m} \leq \mathfrak{n}$ não depende dos conjuntos M e N escolhidos. Para conjuntos finitos, esta noção corresponde à nossa noção intuitiva de "tamanho" de um conjunto. Gostaríamos ainda que as leis usuais para as desigualdades fossem verdadeiras para esta relação entre cardinais. Por exemplo, será verdade que se $\mathfrak{m} \leq \mathfrak{n}$ e $\mathfrak{n} \leq \mathfrak{m}$ então $\mathfrak{m} = \mathfrak{n}$? A resposta afirmativa é dada pelo seguinte resultado famoso:

TEOREMA 42 (SCHROEDER-BERNSTEIN). *Se $\#X \leq \#Y$ e $\#Y \leq \#X$, então $\#X = \#Y$.*

Não daremos aqui uma demonstração deste resultado. Notamos, no entanto, que o Teorema de Schroeder-Bernstein permite-nos frequentemente provar que dois conjuntos são equipotentes, sem exibirmos explicitamente uma bijecção entre esses conjuntos. O exemplo seguinte ilustra isso mesmo.

Exemplo 43.

Considerem-se os conjuntos $X = \mathbb{N} \times \mathbb{N}$ e $Y = \mathbb{N}$. A função $\phi : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ dada por $\phi(n) = (n, 1)$ é injectiva e, de acordo com o Teorema Fundamental da Aritmética, a função $\psi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ dada por $\psi(n, m) = 2^n 3^m$ é também injectiva. Segue-se do Teorema de Schroeder-Bernstein que $\mathbb{N}$ e $\mathbb{N} \times \mathbb{N}$ são equipotentes.

Vimos acima que $\#\mathbb{N} < \#(0, 1]$. Como é que se comparam os cardinais de $(0, 1]$ e de $\mathbb{R}$? Na realidade, os intervalos $(0, 1]$, $[0, 1)$, $(0, 1)$ e $[0, 1]$ têm todos os mesmo cardinal. Por exemplo, a aplicação $\phi : (0, 1] \to (0, 1)$ dada

por:

$$\phi(x) = \begin{cases} \frac{3}{2} - x, & \text{se } \frac{1}{2} < x \leq 1, \\ \frac{3}{4} - x, & \text{se } \frac{1}{4} < x \leq \frac{1}{2}, \\ \frac{3}{8} - x, & \text{se } \frac{1}{8} < x \leq \frac{1}{4}, \\ \quad \vdots & \end{cases}$$

é uma bijecção.

Exercício 22. Mostre que os intervalo $(a, b]$, $[a, b)$, (a, b) e $[a, b]$ $(a - b > 0)$ têm todos a mesma cardinalidade que $(0, 1)$.

Na realidade, qualquer intervalo (de comprimento positivo) tem a mesma cardinalidade que os reais $\mathbb{R}$: por exemplo, a função $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$ é uma bijecção. Assim, vemos que quaisquer intervalos finitos ou infinitos, fechados ou abertos, de comprimento positivo, têm todos a mesma cardinalidade, que se designa por c (do latim continuum). Ainda mais surpreendentemente, o plano $\mathbb{R}^2$ também tem cardinalidade c!

TEOREMA 44. *O plano $\mathbb{R}^2$ tem cardinalidade c.*

Demonstração. Basta verificar que o quadrado $(0, 1] \times (0, 1]$ e o intervalo $(0, 1]$ têm a mesma cardinalidade (porquê?). Para isso, vamos construir uma bijecção $\phi : (0, 1] \times (0, 1] \to (0, 1]$.

Para definir ϕ num elemento $(x, y) \in (0, 1] \times (0, 1]$ escrevemos a expansão decimal infinita de cada uma das componentes:

$$x = 0, x_1 x_2 x_3 x_4 \ldots$$
$$y = 0, y_1 y_2 y_3 y_4 \ldots$$

A função ϕ é então definida pela fórmula:

$$\phi(x, y) = 0, x_1 y_1 x_2 y_2 x_3 y_3 \ldots$$

É agora fácil verificar que ϕ é uma bijecção (qual é a inversa?). □

Exercício 23. O espaço n-dimensional também tem cardinalidade c.

Por que é que este resultado nos parece tão surpreendente? A razão é que enquanto a recta real $\mathbb{R}$ tem dimensão 1, o plano $\mathbb{R}^2$ tem dimensão 2 (e $\mathbb{R}^n$ tem dimensão n). O que acontece é que a noção de dimensão não é preservada por bijecções, e dois conjuntos X e Y podem ter a mesma cardinalidade sem possuírem a mesma dimensão. Se, no entanto, impusermos que a bijecção $\phi : X \to Y$ e a sua inversa $\phi^{-1} : Y \to X$ sejam *contínuas* então as dimensões de X e de Y são forçosamente as mesmas. Por exemplo, não existe nenhuma bijecção $\phi : \mathbb{R} \to \mathbb{R}^2$ contínua e com inversa contínua.

Voltando à questão de compararmos dois cardinais $\mathfrak{m}$ e $\mathfrak{n}$, vamos escrever $\mathfrak{m} < \mathfrak{n}$ se $\mathfrak{m} \leq \mathfrak{n}$ e $\mathfrak{m} \neq \mathfrak{n}$. Pode-se mostrar que exactamente uma das seguintes alternativas é verdade:

$$\mathfrak{m} < \mathfrak{n} \text{ ou } \mathfrak{m} = \mathfrak{n} \text{ ou } \mathfrak{m} > \mathfrak{n}.$$

O Teorema de Schroeder-Bernstein mostra ainda que

$$\mathfrak{m} < \mathfrak{n} \text{ e } \mathfrak{n} < \mathfrak{p} \quad \Longrightarrow \quad \mathfrak{m} < \mathfrak{p}.$$

Por outras palavras, os cardinais estão perfeitamente ordenados.

Já sabemos que os primeiros cardinais são:

$$0, 1, 2, \ldots, n, \ldots, \aleph_0$$

Por outro lado, também sabemos que $\aleph_0 < c$. Estes factos suscitão duas questões naturais:

- Existe algum cardinal maior do que c?
- Será c o cardinal imediatamente a seguir a $\aleph_0$?

A fim de esclarecer a primeira questão, consideramos a seguinte construção: dado um conjunto X vamos designar por $P(X)$ o **conjunto das partes** de X, *i.e.*, o conjunto formado por todos os subconjuntos de X. Por exemplo, se $X = \{1, 2, 3\}$ então:

$$P(X) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Se X é um conjunto finito, então é óbvio que $P(X)$ também é um conjunto finito e $\#X < \#P(X)$. Na realidade esta afirmação também é válida para conjuntos infinitos, mostrando assim que não existe o "maior cardinal" (*i.e.*, um cardinal maior do que os outros todos):

TEOREMA 45. *Para qualquer conjunto X temos que $\#X < \#P(X)$.*

Demonstração. A aplicação $\psi : X \to P(X)$ que a $x \in X$ associa $\{x\} \in P(X)$ é uma injecção, logo $\#X \leq \#P(X)$. Para verificar $\#X \neq \#P(X)$, precisamos de mostrar que não existe uma bijecção $\phi : X \to P(X)$.

Suponhamos, por absurdo, que existe uma aplicação $\phi : X \to \mathcal{P}(X)$ sobrejectiva. Defina-se $Y = \{x \in X : x \notin \phi(x)\}$, que é claramente um elemento de $\mathcal{P}(X)$. Se ϕ é sobrejectiva, existe um elemento $y \in X$ tal que $Y = \phi(y)$, e temos $y \in Y$, ou $y \notin Y$.

Vejamos agora que ambos os casos são impossíveis:

(i) Se $y \in Y = \phi(y)$, segue-se, da definição de Y, que $y \notin Y$, o que é absurdo;

(ii) Se $y \notin Y = \phi(y)$ segue-se, da definição de Y, que $y \in Y$, o que é igualmente absurdo.

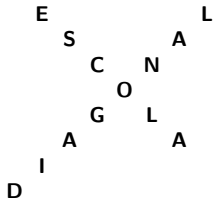
Concluímos que não existe $y \in X$ tal que $Y = \phi(y)$, logo, ϕ não pode ser sobrejectiva. □

O leitor deve observar que o argumento da demonstração é simplesmente o mesmo argumento utilizado aquando da discussão do paradoxo de Russel!

Vejamos, agora, a segunda questão que colocámos acima: será c o cardinal imediatamente a seguir a $\aleph_0$? A afirmação de que $\aleph_1 = c$ é conhecida como a **hipótese do continuum**, e durante muitas décadas foi um dos grandes problemas em aberto em Matemática (este era, de facto, o primeiro problema na lista de Hilbert). A sua solução, dada por Kurt Gödel[14] e Paul Cohen[15] leva-nos aos limites do pensamento lógico matemático: eles mostraram que a hipótese do continuum (*i.e.*, a afirmação $c = \aleph_1$) é independente dos axiomas usuais dos inteiros, da mesma forma que a afirmação em geometria euclidiana de que duas rectas paralelas não se intersectam é independente dos outros axiomas da geometria. Assim, tal como existem geometrias em que o axioma das rectas paralelas não se verifica, é possível considerar modelos para os inteiros $\mathbb{Z}$ (e, portanto, para $\mathbb{R}$) onde a hipótese do continuum se verifica e modelos onde não se verifica!

14 Kurt Gödel (1906-1978) nasceu na Áustria e emigrou jovem para os EUA, onde se tornou membro do Institute for Advanced Study em Princeton. A sua resolução do 2º problema de Hilbert foi feita com 24 anos apenas, e é um dos resultados mais surpreendentes e significativos da Matemática contemporânea. Gödel mostrou que os dois atributos desejáveis duma axiomática para os inteiros (completa e não-contraditória) são eles próprios contraditórios: qualquer sistema de axiomas para $\mathbb{Z}$ que seja não-contraditório (*i.e.*, que nunca leva à conclusão de que determinada afirmação é simultaneamente *falsa e verdadeira*) é incompleto (*i.e.*, admite afirmações cujo valor lógico não pode ser decidido com base nesses mesmos axiomas).

15 Paul Cohen (1934-), matemático americano da Universidade de Princeton, demonstrou em 1963 que o axioma da escolha é independente dos outros axiomas da Teoria dos Conjuntos.



Nós, Elos, Tranças e seus Invariantes

Pedro Matias
Department of Mathematics
Radboud Universiteit Nijmegen
p.matias@math.ru.nl

Rui Carpentier
Departamento de Matemática
Instituto Superior Técnico
rcarpent@math.ist.utl.pt

Palavras Chave

Nós, elos, tranças, polinómio de Alexander-Conway, polinómio de Kauffman, polinómio de Jones, invariantes de Vassiliev.

Resumo

Este artigo baseia-se nas lições do curso *Knots, links, braids and their invariants* proferidas pelo Professor Alexei Sossinsky durante a Escola Diagonal que decorreu no Instituto Superior Técnico de 5 a 9 de Setembro de 2005. O seu objectivo principal é fornecer uma versão escrita das aulas a todos os participantes da Escola.

1 Um Pouco de História

Se pedirmos a uma pessoa sem qualquer formação matemática que nos defina o conceito de nó, arriscamo-nos a ouvir respostas do género: "um nó é aquilo que fazemos quando queremos atar os atacadores dos sapatos!" ou então "um nó é o que os marinheiros fazem com as suas cordas!". Qualquer uma destas respostas é perfeitamente legítima e revela o conhecimento prático que cada um de nós tem de um nó. Na verdade, os primeiros tratados sobre nós remontam à época do Renascimento e são essencialmente compilações de nós com aplicações à marinha. Para um tratamento moderno sobre nós do ponto de vista prático, ver [2].

Do ponto de vista matemático, o primeiro texto científico sobre nós foi apenas publicado em 1771 pela mão de Alexandre-Teophille Vandermonde [17].

Em 1833, Carl Friedrich Gauss deu os primeiros passos no desenvolvimento da teoria de nós com aplicações aos circuitos elétricos. Em particular, Gauss escreveu o que se designa hoje por número de ligação de um elo com duas componentes, como um integral de linha de um campo magnético ao longo de uma das suas componentes [5, 7].

Apesar da importância dos trabalhos de Vandermonde e Gauss, o grande desenvolvimento da teoria de nós deu-se apenas no final do século XIX através das ideias do físico William Thomson (Lord Kelvin). Nesta época ainda se especulava acerca da estrutura da matéria e da possibilidade de reconciliar a teoria corpuscular com a teoria ondulatória. Foi neste contexto que Lord Kelvin sugeriu que os constituintes elementares da matéria (átomos) poderiam ser descritos por curvas fechadas entrelaçadas (nós), a que ele chamou átomos vorticiais [16]. Assim, a cada tipo de átomo corresponderia um nó diferente, ou seja, a classificação dos átomos traduzir-se-ia na classificação de nós! Note-se que, deste ponto de vista, a classificação das moléculas corresponderia à classificação de conjuntos finitos de nós entrelaçados (elos).

Este conjunto de ideias motivou Peter Guthrie Tait a construir as primeiras tabelas de nós e a formular uma série de conjecturas sobre a sua classificação [14]. Para uma abordagem moderna às famosas conjecturas de Tait, ver [15]. Apesar do sucesso alcançado por Tait na tabulação de nós com um número de cruzamentos menor ou igual a 7, continuava a faltar um procedimento sistemático que permitisse distinguir nós mais complicados. Esse procedimento só viria a ser alcançado no início do século XX com o desenvolvimento da topologia e a introdução do conceito de *invariante de nó*.

Entretanto, a teoria dos átomos vorticiais de Lord Kelvin, apesar de suportada por alguns físicos importantes da época (entre os quais James Clerk Maxwell, o pai do electromagnetismo), viria a ser abandonada pela comunidade científica. Parte desse abandono deveu-se ao sucesso na tabulação dos átomos através de um procedimento aritmético que culminou na construção da tabela periódica dos elementos pelo químico Dmitri Mendeleev.

Apesar de tudo, estes primeiros passos na teoria de nós foram decisivos para a compreensão dos principais problemas a resolver e para a sua reformulação no contexto da topologia do espaço tridimensional.

2 Os Problemas Fundamentais

Uma das questões naturais que podemos formular em teoria de nós é o *problema da comparação*[1]:

1 Esta secção pretende enumerar alguns problemas fundamentais em teoria de nós de um ponto de vista intuitivo. Nessa perspectiva, iremos considerar um nó como uma corda entrelaçada sobre si mesma com as extremidades coladas. O nó trivial será simplesmente um nó sem entrelaçamentos. Na secção seguinte daremos definições rigorosas do conceito de nó, assim como de algumas noções utilizadas nesta secção de forma mais abusiva.

Dados dois nós K, K' aparentemente distintos, será que eles representam o mesmo nó?

Para responder a esta questão poderiamos pegar no nó K e tentar deformá-lo no espaço até conseguir obter o nó K'. Em caso afirmativo isto seria uma demonstração irrefutável que os nós eram equivalentes. Mas... e se a nossa manipulação falhasse? Bem, nesse caso isso *não* provaria que os nós eram diferentes, revelaria, isso sim, uma mera evidência inconclusiva. Como resolver então esta incapacidade de provar a equivalência ou não-equivalência de dois nós? O truque está em associar um objecto (numérico, algébrico, etc) a cada classe de equivalência de um nó. Dito de outra forma, a ideia é definir uma função no conjunto de todos os nós que seja invariante sob deformações. As funções que satisfazem esta condição dizem-se *invariantes de nós*. Note-se que um invariante de nós serve para distinguir nós e não para provar a sua equivalência. A definição de invariante de nós não funciona no sentido contrário: se dois nós têm o mesmo invariante então eles não são necessariamente equivalentes, a menos que o invariante seja completo. Ao longo da história foram descobertos vários invariantes (incompletos) de nós, os quais ajudaram à resolução parcial do problema da comparação. Neste artigo iremos debruçar-nos sobre três desses invariantes: o polinómio de Alexander-Conway, o polinómio de Jones e os invariantes de Vassiliev. Uma curiosidade acerca do problema da comparação: em 2004, o matemático Sergei Matveev propôs um algoritmo para a resolução deste problema [10], cuja implementação a nivel computacional é extremamente complicada e não produz respostas em tempo considerado útil. Como tal, o problema mantém-se em aberto do ponto de vista prático, sendo nalguns casos impossível decidir acerca da equivalência ou não equivalência de dois nós que partilham os mesmos invariantes.

Um caso particular do problema da comparação é o *problema do desenlaçamento*:

Dado um nó K, será que ele representa o nó trivial?

Tal como no problema da comparação, existe um algoritmo para o problema do desenlaçamento proposto por Wolfgang Haken em 1961 [8, 9]. Contudo, a sua implementação a nível computacional é tão complicada que, para nós de complexidade elevada, também não é possivel obter respostas em tempo considerado útil.

Finalmente, existe ainda o problema da tabulação de *nós primos*[2] de acordo com o seu número de cruzamentos, em analogia com o que se faz

2 Um nó diz-se *primo* se for não trivial e não puder ser escrito como a soma conexa de dois nós não triviais. Em 1949, Horst Schubert demonstrou a existência de uma decomposição única de qualquer nó não trivial na soma conexa de nós primos, a menos de ordenação [13].

em teoria dos números para a descoberta de novos números primos. A tabela seguinte dá-nos o número total de nós primos em função do número de cruzamentos, a menos da simetria de espelho:

Cruzamentos	Nós primos
3	1
4	1
5	2
6	3
7	7
8	21
9	49
10	165
11	552
12	2176
13	9988
14	46972
15	253293
16	1388705

Até aos dias de hoje não se sabe qual o número de nós primos com um número de cruzamentos maior ou igual a 17 (sabe-se no entanto que existe um número infinito de nós primos). Os próximos tempos ditarão até onde podem ir os limites da computação humana!

3 Definições Básicas e Movimentos de Reidemeister

Nesta secção começamos por definir o conceito de nó do ponto de vista matemático, assim como a respectiva noção de equivalência entre nós. Seguidamente enunciaremos a noção fundamental de diagrama de um nó. Estes diagramas permitem-nos representar qualquer nó num plano sem perda de informação relativamente à forma como o nó está mergulhado no espaço tridimensional. Terminamos a secção com o Teorema de Reidemeister.

DEFINIÇÃO 1. Um *nó poligonal* é uma curva poligonal fechada que não se auto-intersecta no espaço euclideano $\mathbb{R}^3$. Usaremos a letra K (do inglês "knot") para designar um nó poligonal arbitrário.

Para introduzir a noção de equivalência entre nós poligonais necessitamos do conceito de isotopia elementar, o qual passamos a descrever. Suponhamos que os lados $[AC]$ e $[CB]$ de um triângulo $[ABC]$ são segmentos de um nó poligonal K que não intersecta o interior de $[ABC]$ em nenhum ponto.

Então chama-se *isotopia elementar* à substituição dos segmentos $[AC]$ e $[CB]$ pelo segmento $[AB]$ (ver Figura 3.1). Uma sequência finita de isotopias elementares diz-se uma *isotopia ambiente*.

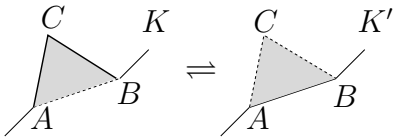


Figura 3.1: Isotopia elementar

Definição 2. Dois nós poligonais K, K' dizem-se *equivalentes* se existe uma isotopia ambiente que transforma K em K'. Nesse caso escrevemos $K \sim K'$.

Note-se que a noção intuitiva de nó, como uma corda entrelaçada sobre si mesma com as extremidades coladas, não coincide com a definição matemática de nó poligonal dada acima. Contudo, é possível introduzir, do ponto de vista matemático, uma definição alternativa de nó que se aproxima mais da realidade. A ideia é definir um nó, não como uma linha poligonal, mas como uma curva com um comportamento "mais suave". A definição é a seguinte:

Definição 3. Um *nó suave* é a imagem de um mergulho suave da circunferência S^1 em $\mathbb{R}^3$ com derivada não nula em todos os pontos.

Definição 4. Dois nós suaves K, K' dizem-se *equivalentes* se existe uma família de difeomorfismos $f_t \colon S^1 \to \mathbb{R}^3$, $t \in [0,1]$, dependendo suavemente do parâmetro t, que transforma K em K'. Tal como no caso poligonal, chamamos à família $\{f_t \mid t \in [0,1]\}$ uma *isotopia ambiente*.

Existe uma correspondência biunívoca entre as classes de equivalência de nós suaves e as classes de equivalência de nós poligonais [4, Prop. 1.10]. Como tal, usaremos daqui em diante cada uma destas abordagens indistintamente.

Para representar nós é conveniente usar as suas projecções ortogonais num plano de $\mathbb{R}^3$. Estas projecções contêm informação sobre os cruzamentos presentes no nó e denominam-se *diagramas de nós*. No entanto, é necessário ter em atenção alguns aspectos técnicos relativamente à escolha do plano da projecção. Mais precisamente:

1. as rectas tangentes em cada ponto de um nó devem ser projectadas em rectas no plano;

2. apenas dois pontos distintos de um nó podem ser projectados no mesmo ponto do plano;

3. o número de cruzamentos de um nó é finito e em cada cruzamento as projecções das rectas tangentes não coincidem.

Em particular, as figuras que se seguem ilustram dois diagramas de nós que violam as regras 2. e 3., respectivamente.

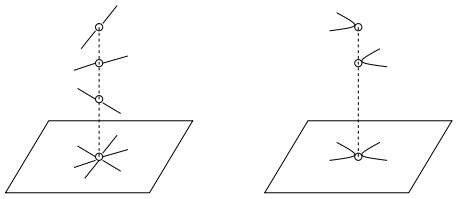


Figura 3.2: Dois exemplos de diagramas de nós proibidos

DEFINIÇÃO 5. Dois diagramas de nós dizem-se *equivalentes* se um deles pode ser obtido a partir do outro através de uma sequência finita de movimentos descritos na Figura 3.3. Qualquer um destes movimentos diz-se uma *isotopia plana*.

Figura 3.3: Isotopias planas

Note-se que as isotopias planas não alteram o número de cruzamentos de um diagrama de nó. Por exemplo, a transformação representada na Figura 3.4 não é uma isotopia plana. Contudo, esta transformação é uma isotopia ambiente pois o diagrama do nó da direita pode ser obtido a partir do diagrama do nó da esquerda com uma isotopia elementar.

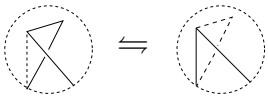


Figura 3.4: Exemplo de uma isotopia não plana

Podemos também definir a noção de isotopia plana entre diagramas de nós suaves se considerármos movimentos suaves análogos aos da Figura 3.3 para o caso poligonal.

É importante notar que a definição de isotopia plana implica a não ocorrência das seguintes situações:

1. alteração do número de cruzamentos;
2. as projecções de dois ramos do nó tornarem-se tangentes;
3. os pontos de três ramos do nó serem projectados no mesmo ponto.

Contudo, no caso de uma isotopia ambiente, todas as situações acima descritas podem ocorrer. De certa forma, podemos olhar para as isotopias ambiente como certas transformações entre diagramas de nós que permitem a ocorrência de qualquer uma das três situações acima descritas. Iremos adoptar este ponto de vista daqui em diante.

Um facto notável acerca das isotopias ambiente é que qualquer uma delas pode ser descrita como uma sequência finita de isotopias planas e três movimentos elementares específicos. Este resultado é conhecido por *Teorema de Reidemeister* e os movimentos elementares específicos designam-se por *primeiro, segundo e terceiro movimentos de Reidemeister*. As figuras seguintes ilustram cada um destes movimentos.

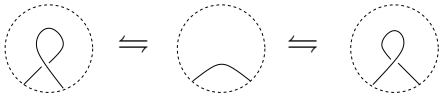


Figura 3.5: Primeiro movimento de Reidemeister Ω_1

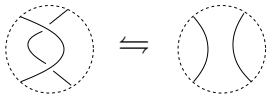


Figura 3.6: Segundo movimento de Reidemeister Ω_2

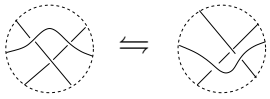


Figura 3.7: Terceiro movimento de Reidemeister Ω_3

Concluimos esta secção com alguns exemplos dos diagramas de nós mais conhecidos.

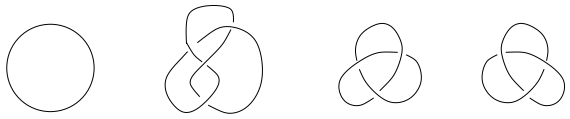


Figura 3.8: Nó trivial, figura de 8, trevos direito e esquerdo

4 O Polinómio de Alexander-Conway

Nesta secção iremos debruçar-nos sobre um dos primeiros invariantes de nós a ser descoberto: o polinómio de Alexander-Conway. Este invariante foi introduzido em 1923 por James Waddell Alexander [1] no contexto da topologia álgebrica, mais concretamente através do uso da teoria da homologia. Mais tarde, em 1970, John Horton Conway [6] redefiniu o polinómio de Alexander através de uma formulação axiomática. Neste artigo adoptaremos o ponto de vista de Conway em função da sua elegância e simplicidade.

DEFINIÇÃO 6. Um *elo poligonal (suave)* é uma colecção finita de nós poligonais (suaves) disjuntos dois a dois. Usaremos a letra L (do inglês "link") para designar um elo poligonal (suave) arbitrário.

As definições de equivalência entre dois elos poligonais (suaves) é uma generalização óbvia das definições 2 e 4, respectivamente. Mostra-se ainda que existe uma correspondência biunívoca entre as classes de equivalência de elos poligonais e as classes de equivalência de elos suaves, generalizando o resultado conhecido para nós. Como tal, usaremos daqui em diante a abordagem suave pois é aquela que mais se aproxima da noção intuitiva que temos de um elo.

Tal como fizemos para os nós, iremos considerar representações de elos através das respectivas projecções em planos de $\mathbb{R}^3$, a que chamamos *diagramas de elos*. Eis alguns exemplos dos diagramas de elos mais conhecidos:

Figura 4.1: Elo de Hopf, anéis Borromeanos e elo de Whitehead

DEFINIÇÃO 7. Sejam $\mathcal{L}$ o conjunto dos diagramas de elos orientados e $\mathbb{Z}[x]$ o conjunto dos polinómios na variável x com coeficientes inteiros. O *polinómio de Alexander-Conway* é uma aplicação $\nabla\colon \mathcal{L} \to \mathbb{Z}[x]$ que satisfaz as seguintes propriedades:

(C1) Se $L \sim L'$ então $\nabla(L) = \nabla(L')$;

(C2) $\nabla(O) = 1$, onde O denota o nó trivial;

(C3) $\nabla(L_+) - \nabla(L_-) = x\nabla(L_0)$, onde L_+, L_- e L_0 são partes de diagramas de elos com a estrutura presente na Figura 4.2.

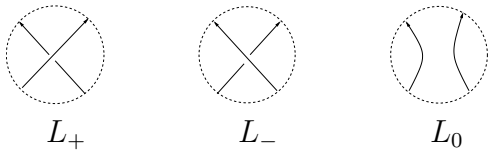
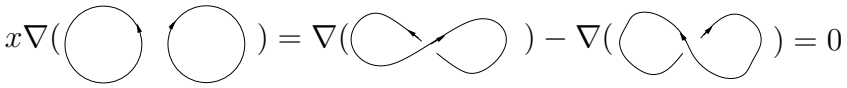


Figura 4.2: Diagramas de elos definidores do polinómio de Alexander-Conway

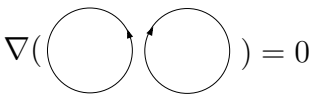
Note-se que a propriedade (C1) diz-nos que a aplicação ∇ é um invariante de elos orientados enquanto a propriedade (C2) é uma espécie de normalização. A propriedade (C3) contém a maioria da informação acerca deste invariante e, juntamente com (C1) e (C2), permite-nos calcular o polinómio de Alexander-Conway de qualquer elo orientado. O resultado não trivial acerca desta formulação axiomática é descrito pelo seguinte teorema:

Teorema 8 (Conway). *Existe uma única aplicação* $\nabla\colon \mathcal{L} \to \mathbb{Z}[x]$ *satisfazendo as propriedades (C1), (C2) e (C3).*

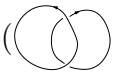
Vamos então usar os axiomas de Conway para calcular este invariante em alguns diagramas de elos orientados. Relativamente ao elo com duas componentes não entrelaçadas temos a seguinte relação:



e portanto concluimos que



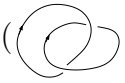
Para o elo de Hopf com orientação positiva temos

$\nabla($  $) = \nabla($  $) + x\nabla($  $)$

$= 0 + x\nabla(O)$

$= x$

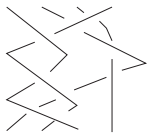
Finalmente para o trevo direito com a orientação positiva temos

$\nabla($  $) - \nabla($  $) = x\nabla($  $)$

e usando o resultado para o elo de Hopf positivo concluimos que $\nabla(\text{trevo}) = x^2+1$. Este último resultado permite-nos inferir que o trevo não é equivalente ao nó trivial pois os seus polinómios de Alexander-Conway são diferentes. Consequentemente, não é possível desenlaçar o trevo de forma a obter o nó trivial!

5 Tranças e Apresentações de Grupos

Intuitivamente uma trança é um conjunto finito de fios entrelaçados descendentes[3]. Do ponto de vista matemático, defimos uma *trança com n fios* como sendo n caminhos poligonais em $\mathbb{R}^3$, sem intersecções, que ligam os pontos de cima $\{(1,0,1),(2,0,1),...,(n,0,1)\}$ aos pontos de baixo $\{(1,0,0),(2,0,0),...,(n,0,0)\}$[4] de modo que qualquer plano horizontal de equação $z=\alpha$ (com $\alpha \in [0,1]$) intersecta cada caminho em um único ponto.

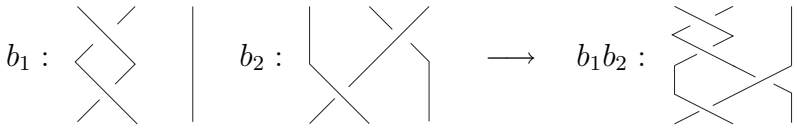


Da mesma forma que definimos isotopias ambientes entre nós ou elos, também se define uma isotopia ambiente entre duas tranças como sendo uma sucessão de isotopias elementares que preservam a condição dos fios serem descendentes. Do mesmo modo, dizemos que duas tranças são *equivalentes* se existir uma isotopia ambiente que transforme uma na outra.

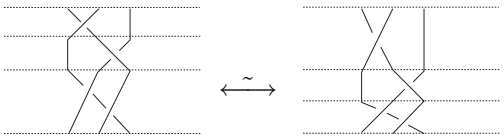
3 Com a palavra "descendentes" pretendemos dizer que nenhum dos fios curva para cima.

4 O uso de pontos com estas coordenadas é apenas uma convenção. Podiamos ter considerarado outros pontos colineares.

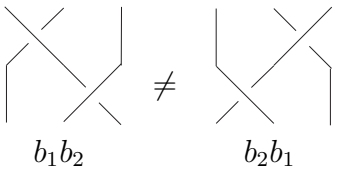
Designemos por B_n o conjunto das tranças com n fios a menos de equivalência. Neste conjunto podemos definir uma operação produto do seguinte modo: dadas duas tranças b_1 e b_2, colocamos b_1 em cima de b_2 de modo a que os extremos inferiores de b_1 coincidam com os extremos superiores de b_2 formando uma nova trança b_1b_2. Para que as tranças tenham uma altura uniforme é necessário encolher o produto de duas tranças por um factor de 2.



Verifica-se facilmente que esta operação é associativa:

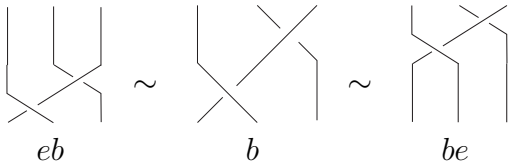


No entanto não é comutativa:

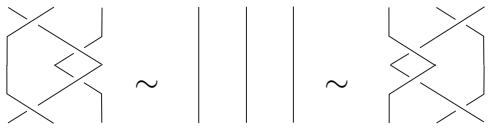


Note-se que as tranças no último diagrama não são equivalentes pois os seus fios têm ligações distintas.

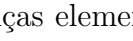
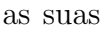
Além da propriedade associativa, esta operação possui um elemento neutro (uma unidade) que é dado pela trança formada por n segmentos de recta paralelos ($e :=$ | | ⋯ |):

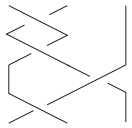


Temos ainda que para cada trança existe uma inversa, que composta com ela própria dá uma trança equivalente à identidade, e que é dada pela imagem reflectida ao longo do plano horizontal:

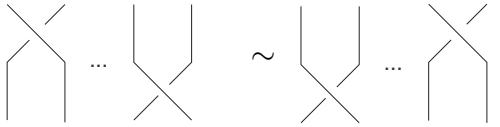


Resumindo: o conjunto B_n equipado com este produto forma um grupo.

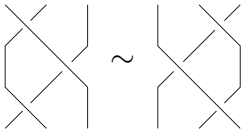
Qualquer trança com n fios pode ser dada como o produto de um número finito de $n-1$ tranças elementares $\sigma_1 =$  , $\sigma_2 =$  , ... , $\sigma_{n-1} =$  , e as suas inversas $\sigma_1^{-1} =$, $\sigma_2^{-1} =$, ... , $\sigma_{n-1}^{-1} =$. Assim podemos codificar qualquer trança como uma palavra do tipo $\sigma_{i(1)}^{\epsilon_1}\sigma_{i(2)}^{\epsilon_2}...\sigma_{i(k)}^{\epsilon_k}$ onde os ϵ_i's são expoentes inteiros. Por exemplo:


$b = \sigma_1^2\sigma_2^{-1}\sigma_1$

É claro que tranças equivalentes podem ter diferentes palavras, mas estas estão relacionadas pelas seguintes relações:


$\sigma_i\sigma_j = \sigma_j\sigma_i$ se $|i-j| \geq 2$

e


$\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}$ para $i = 1, ..., n-2$.

Esta última relação (ou relações pois $i = 1, ..., n-2$) corresponde ao terceiro movimento de Reidemeister Ω_3 e é conhecida como relação de Artin ou equação de Yang-Baxter.

O parágrafo anterior descreve um exemplo típico de um conjunto de relações numa apresentação de um grupo. A definição seguinte formaliza este conceito.

DEFINIÇÃO 9. Uma *apresentação (finita) de um grupo* G é um par de conjuntos (finitos), escritos na forma $\langle g_1, ..., g_k | r_1, ..., r_l \rangle$, que obedecem à seguinte descrição. O primeiro conjunto $\{g_1, ..., g_k\}$ é formado por elementos de G com os quais é possível obter qualquer elemento de G através das

operações de multiplicação e inversão. Por essa razão designamos tais elementos por *geradores*. O segundo conjunto $\{r_1, ..., r_l\}$ é formado por palavras básicas[5] de geradores e seus inversos que representam o elemento neutro de G. Designamos tais elementos por *relações*.

De acordo com esta definição a apresentação de B_n que descrevemos anteriormente é dada por

$$\langle \sigma_1, ..., \sigma_{n-1} | \sigma_i\sigma_{i+1}\sigma_i(\sigma_{i+1}\sigma_i\sigma_{i+1})^{-1} \forall_{i=1,...,n-1}, \sigma_i\sigma_j(\sigma_j\sigma_i)^{-1} \text{ se } |i-j| \geq 2 \rangle.$$

Observação 1. Na literatura é comum escrever as relações de uma apresentação de um grupo como equações do tipo $\sigma_i\sigma_{i+1}\sigma_i(\sigma_{i+1}\sigma_i\sigma_{i+1})^{-1} = 1$ ou (equivalentemente) $\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}$. Sempre que uma relação é dada simplesmente por uma palavra, esta refere-se à equação: *palavra igual ao elemento neutro*.

Para tornar mais clara a definição anterior vejamos como se efectuam a multiplicação e a inversão de palavras numa dada apresentação, por exemplo $\langle a, b | a^2, b^3, (ab)^2 \rangle$. Dadas duas palavras a multiplicação é dada pela concatenação das duas palavras obedecendo à regra da soma dos expoentes para bases iguais ($a^{\epsilon_1}a^{\epsilon_2} = a^{\epsilon_1+\epsilon_2}$). Por exemplo:

$$\begin{aligned} (ab^2ab)(b^{-2}aba^3) &= ab^2abb^{-2}aba^3 \\ &= ab^2ab^{-1}aba^3 \end{aligned}$$

.

O elemento neutro é representado pela palavra vazia e o inverso de uma palavra consiste em trocar a ordem das letras e os sinais aos expoentes. Por exemplo:

$$(ab^2a^{-3}b^{-1})^{-1} = ba^3b^{-2}a^{-1}$$

.

Sempre que uma relação da apresentação surge como parcela de uma dada palavra podemos retirá-la sem alterar o valor da palavra. Do mesmo modo podemos inserir numa palavra qualquer uma das relações. Também é válido inserir/eliminar na palavra uma parcela formada por uma palavra seguida da sua inversa (do tipo $a^2baa^{-1}b^{-1}a^{-2}$). Como exemplo vejamos como simplificar a seguinte palavra através das relações a^2, b^3 e $abab$:

5 Básicas no sentido em que qualquer outra palavra de geradores e seus inversos que represente o elemento neutro de G pode ser obtida a partir destas através das operações de multiplicação e inversão.

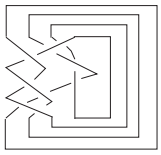
$$\begin{aligned} ab^2ab^{-1}aba^3 &= ab^2\underline{bb^{-1}}ab^{-1}ab\underline{a^2}a \\ &= a\underline{b^3}b^{-1}a\underline{aa^{-1}}b^{-1}a\underline{aa^{-1}}ba \\ &= ab^{-1}\underline{a^2}a^{-1}b^{-1}\underline{a^2}a^{-1}ba \\ &= ab^{-1}a^{-1}b^{-1}a^{-1}ba \\ &= a(\underline{abab})^{-1}ba \\ &= aba \\ &= aba\underline{bb^{-1}} \\ &= \underline{abab}b^{-1} \\ &= b^{-1} \end{aligned}$$

.

Observação 2. As partes sublinhadas na equação acima representam as relações que foram introduzidas ou eliminadas da palavra em causa.

O problema de determinar se duas palavras distintas numa dada apresentação de um grupo representam o mesmo elemento no grupo é extremamente complexo. Para grande parte dos grupos este problema não tem solução. No entanto, para cada grupo de tranças B_n, sendo n um número natural qualquer, existe um algoritmo que resolve tal problema.

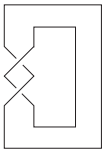
As tranças estão também relacionadas com os nós e os elos. De facto, é possível obter um nó ou um elo unindo os extremos superiores da trança com os respectivos extremos inferiores na mesma ordem. Eis um exemplo:



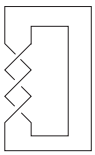
Esta operação é designada por *fecho de uma trança* e é sobrejectiva tanto no conjunto dos nós como no conjunto dos elos. Este resultado foi demonstrado por Alexander.

TEOREMA 10 (ALEXANDER). *Qualquer nó ou elo é o fecho de uma trança.*

Para finalizar, apresentamos dois exemplos de um nó e um elo dados como fechos de tranças:



elo de Hopf



nó trevo direito

6 O Polinómio de Kauffman

Nesta secção iremos introduzir outro invariante de nós: o polinómio de Kauffman. Este invariante está definido para todos os elos orientados e é construido a partir de dois objectos não invariantes sob o movimento de Reidemeister Ω_1: o parêntesis de Kauffman e o número de torção. Na secção seguinte utilizaremos o polinómio de Kauffman para definir de uma forma simples o polinómio de Jones.

Começemos então por introduzir o parêntesis de Kauffman, o qual está definido apenas para elos não orientados.

Definição 11. Sejam $|\mathcal{L}|$ o conjunto dos elos não orientados e $\mathbb{Z}[a,b,c]$ o conjunto dos polinómios nas variáveis a,b,c com coeficientes inteiros. O *parêntesis de Kauffman* é uma aplicação $\langle\,\cdot\,\rangle\colon |\mathcal{L}| \to \mathbb{Z}[a,b,c]$ que satisfaz as seguintes propriedades:

(K1) $\langle\, O \,\rangle = 1$, onde O denota o nó trivial;

(K2) $\langle\, L \sqcup O \,\rangle = c\,\langle\, L \,\rangle$;

(K3) $\langle\, L \,\rangle = a\,\langle\, L_A \,\rangle + b\,\langle\, L_B \,\rangle$, onde L, L_A e L_B são partes de diagramas de elos não orientados com a estrutura presente na Figura 6.1.

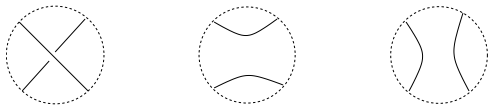


Figura 6.1: Diagramas de elos definidores do parêntesis de Kauffman

Relativamente à propriedade (K3), note-se que os diagramas L_A e L_B são obtidos a partir do diagrama L de forma única, independentemente da forma como rodamos este último. De facto, os arcos presentes nos diagramas L_A e L_B são escolhidos nas regiões A e B de acordo com a figura 6.2.

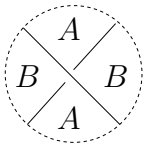


Figura 6.2: Regiões A e B perto de um cruzamento

Estas regiões são definidas da seguinte maneira: se nos deslocármos ao longo do ramo superior do elo L, encontraremos a região A à nossa esquerda antes de chegármos ao cruzamento e à nossa direita depois de passármos o cruzamento. Reciprocamente para a região B.

Observação 3. Por trás da definição axiomática do parêntesis de Kauffman existe uma motivação física baseada em modelos bidimensionais da física estatística. A ideia consiste em associar a cada cruzamento i de um diagrama de elo L com n cruzamentos, um de dois valores formais $x_i = A$ ou $x_i = B$. Do ponto de vista físico, esta escolha pode ser encarada como uma distribuição de spins num reticulado finito, onde cada ponto contem uma partícula com um determinado spin. Uma escolha de estados para todos os cruzamentos i diz-se um *estado do diagrama* L. Note-se que o número de estados possíveis de L é igual a 2^n. A cada estado do diagrama L corresponde um conjunto de nós triviais que não se entrelaçam. Este conjunto é obtido a partir da destruição dos cruzamentos de L de acordo com as regras descritas na figura 6.3.

Figura 6.3: Destruição de cruzamentos num elo

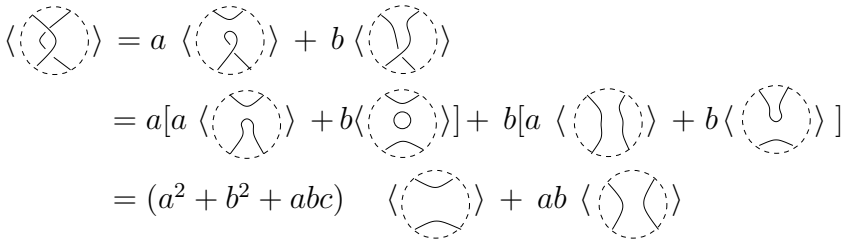
Suponhamos que o diagrama L está num determinado estado $s \in S$, onde S denota o conjunto de todos os estados possíveis de L. Denotemos por $\alpha(s)$ e $\beta(s)$ o número de cruzamentos do diagrama L nos estados A e B, respectivamente, e por $\gamma(s)$ o número de nós triviais não entrelaçados que se obtêm quando destruimos os cruzamentos de L. Então pode mostrar-se que o parêntesis de Kauffman é dado por

$$\langle L \rangle = \sum_{s \in S} a^{\alpha(s)}\, b^{\beta(s)}\, c^{\gamma(s)-1}.$$

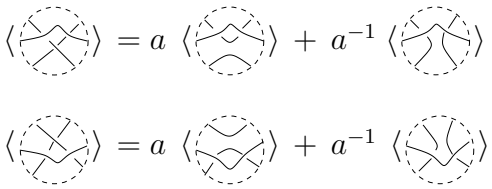
Esta fórmula é normalmente designada no contexto da física estatística por *soma de estados*.

Vamos agora determinar as relações que as variáveis a, b, c têm de satisfazer de forma a que o parêntesis de Kauffman seja um invariante de elos. Para isso basta exigir invariância relativamente aos três movimentos de Reidemeister Ω_1, Ω_2 e Ω_3.

Comecemos por usar as relações (K2) e (K3) para efectuar o seguinte cálculo:

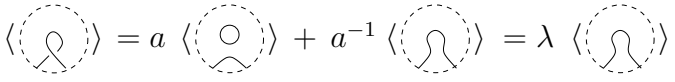


Se impusérmos que $ab = 1$ e $a^2 + b^2 + abc = 0$, então o parêntesis de Kauffman torna-se Ω_2-invariante. Para tal se verificar escolhemos $b = a^{-1}$ e $c = -a^2 - b^2$ ficando no final com um polinómio nas variáveis a, a^{-1}. Esta escolha deixa-nos sem mais margem de manobra relativamente às variáveis iniciais do polinómio. Contudo, conseguimos ainda obter Ω_3-invariância. Pela condição (K3) temos as seguintes relações:

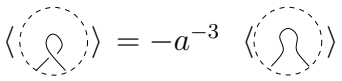


Claramente, os dois últimos diagramas nos membros direitos das equações acima diferem por uma isotopia plana, logo têm o mesmo parêntesis de Kauffman. Relativamente aos dois diagramas do meio, verificamos facilmente que diferem por dois movimentos Ω_2, logo também têm o mesmo parêntesis de Kauffman. Concluimos então que $\langle \cdot \rangle$ é Ω_3-invariante.

Verifiquemos agora a Ω_1-invariância. Pelas condições (K2) e (K3) temos



onde $\lambda = a(-a^2 - a^{-2}) + a^{-1} = -a^3$. Um cálculo semelhante conduz-nos à relação



Temos então que o parêntesis de Kauffman *não é* Ω_1-invariante. Como contornar esta situação? A ideia é definir uma nova função que inclua o parêntesis de Kauffman e outro objecto que não seja Ω_1-invariante de forma a que os dois efeitos somados se cancelem mutuamente e obtenhamos no final um objecto Ω_1-invariante. Esta manobra consegue-se através da introdução de um objecto clássico em teoria de elos denominado número de torção.

DEFINIÇÃO 12. Para um elo orientado L definimos o seu *número de torção* através da fórmula

$$w(L) = \sum_i \varepsilon_i,$$

onde a soma é tomada sobre todos os cruzamentos do elo e $\varepsilon_i = \pm 1$ de acordo com a figura 6.4:

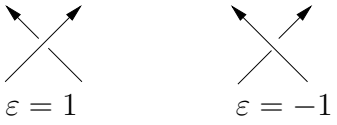


Figura 6.4: Cruzamentos positivos e negativos

É fácil de ver que o número de torção de qualquer elo orientado é Ω_2 e Ω_3-invariante mas muda de ± 1 unidade sob o movimento Ω_1.

DEFINIÇÃO 13. O *polinómio de Kauffman* é uma aplicação $F\colon \mathcal{L} \to \mathbb{Z}[a, a^{-1}]$ definida por

$$F(L) = (-a)^{-3w(L)} \langle\, |L| \,\rangle,$$

onde $|\cdot|$ é a operação que despreza a orientação do elo L.

TEOREMA 14. *O polinómio de Kauffman é um invariante de elos orientados.*

Demonstração. Basta mostrar que F é Ω_1-invariante. Se usármos as relações (K2) e (K3) obtemos

$F($  $) = (-a)^{-3w(L)}$ $\langle$  $\rangle$

$= (-a)^{-3w(L)}$ $[a^{-1}$ $\langle$  $\rangle$ $+ a$ $\langle$  $\rangle]$

$= (-a)^{-3w(L)}(-a)^3$ $\langle$  $\rangle$

$= (-a)^{-3w(L_0)}$ $\langle$  $\rangle$

$= F($  $)$

onde se usou o facto de

$w($  $) = w($  $) + 1$

□

7 O Polinómio de Jones

Tal como mencionámos na secção anterior, o polinómio de Jones tem uma definição extremamente simples em termos do polinómio de Kauffman. Contudo, a sua formulação original aconteceu um pouco por acaso no seio de uma área da matemática completamente diferente da teoria de nós. Em 1985, o matemático Vaughan Jones estava a trabalhar no contexto das algebras de operadores com aplicações à teoria quântica de campo, quando se deparou com um objecto que poderia ser usado como um novo invariante de nós[6]. Durante uma conferência na universidade de Columbia, Jones teve oportunidade de falar com a matemática Joan Birman (especialista em teoria de nós) acerca dos seus resultados. As conclusões desta conversa foram quase instantâneas: Jones tinha acabado de descobrir um novo invariante de nós. Para perceber melhor a origem do polinómio de Jones e a sua relação com as álgebras de operadores, ver [11].

A grande vantagem do polinómio de Jones relativamente a alguns invariantes clássicos reside no facto de ser um invariante mais fino. Por exemplo, os trevos direito e esquerdo têm o mesmo polinómio de Alexander-Conway

6 Neste contexto histórico usamos a palavra "nó" num sentido mais lato. De facto poderiamos (e deveriamos) substituir a palavra "nó" por "elo".

mas polinómios de Jones diferentes. Na altura, Jones ainda conjecturou que o seu invariante talvez fosse um invariante completo de nós primos, resultado que, mais tarde se veio a provar ser falso[7].

Nesta secção iremos introduzir o polinómio de Jones, não através da sua formulação original, mas recorrendo apenas ao polinómio de Kauffman. A razão para tal escolha justifica-se plenamente por motivos de simplicidade. De facto, o polinómio de Jones não é mais do que o polinómio de Kauffman com a variável a redefinida.

DEFINIÇÃO 15. O *polinómio de Jones* é uma aplicação $V\colon \mathcal{L} \to \mathbb{Z}[q^{1/4}, q^{-1/4}]$ definida por

$$V(L) = F(L)\,|_{a=q^{1/4}}\ .$$

TEOREMA 16. *O polinómio de Jones satisfaz as seguintes propriedades:*

(J1) *Se* $L \sim L'$ *então* $V(L) = V(L')$*;*

(J2) $V(O) = 1$*;*

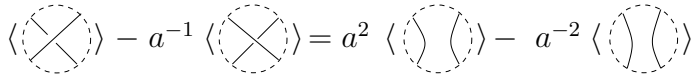
(J3) $qV(L_+) - \frac{1}{q}V(L_-) = (\frac{1}{\sqrt{q}} - \sqrt{q})V(L_0)$*;*

(J4) $V(L \sqcup O) = (-\sqrt{q} - \frac{1}{\sqrt{q}})V(L)$*,*

onde L_+*,* L_- *e* L_0 *são os mesmos diagramas da figura 4.2.*

Demonstração. Vamos demonstrar apenas a propriedade (J3). Pela relação (K3) podemos escrever

Multiplicando a primeira equação por a, a segunda equação por a^{-1} e subtraindo membro a membro obtemos



Multiplicando agora por $(-a)^{-3w(L_0)}$ e tendo em conta que $w(L_+) = w(L_0) + 1$ e $w(L_-) = w(L_0) - 1$ obtemos a identidade (J3). □

7 Desde cedo se percebeu que o polinómio de Jones não era um invariante completo de elos e nós compostos, i.e., nós que não são primos.

É possivel mostrar que as propriedades acima definem de forma única o polinómio de Jones.

O polinómio de Jones satisfaz ainda uma série de propriedades relativamente à simetria de espelho, à orientação dos elos e à soma conexa. Mais precisamente:

(J4) $V(L^*)(q) = V(L)(\frac{1}{q})$, onde L^* é a imagem pelo espelho de L;

(J5) $V(-L) = V(L)$, onde $-L$ é o elo L com a orientação invertida;

(J6) $V(L\#L') = V(L)\,V(L')$, onde $L\#L'$ é a soma conexa de L com L'.

8 Invariantes de Vassiliev

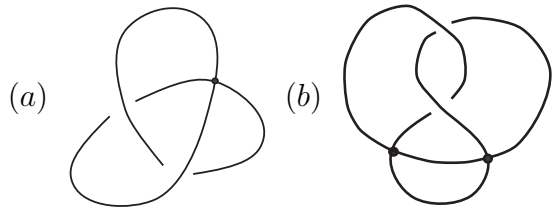
Fazendo uma revisão cronológica dos invariantes abordados neste texto, temos que o primeiro invariante a surgir foi o polinómio de Conway em 1970 (que não é mais do que o polinómio de Alexander[8] normalizado e com uma mudança de variáveis). Em 1985 surgiu o polinómio de Jones (num contexto que aparentemente nada tinha a ver com nós) e pouco anos mais tarde, em 1988, Kauffman introduziu a versão do polinómio de Jones que aqui apresentámos.

Posteriormente surgiram outros polinómios a duas variáveis que generalizam os polinómios de Jones e Conway. No entanto, durante o período compreendido entre 1990 e 1993, Vassiliev desenvolveu um novo tipo de invariantes. Estes invariantes são numéricos (com valores em $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$) e surgiram com o auxílio de ferramentas da topologia algébrica. Contudo, rápidamente foram desenvolvidos métodos combinatóricos para tal teoria. Será essa a abordagem que iremos tomar para descrevê-los.

Para isso introduziremos primeiro o conceito de nó singular. Um *nó singular* é uma curva fechada com um número finito de pontos duplos simples. Um *ponto duplo simples* é um ponto onde o nó se auto-intersecta transversalmente (localmente é da forma $\times$).

Definimos então, para cada número natural n, o conjunto Σ_n dos nós singulares com n pontos duplos simples. Deste modo Σ_0 é o conjunto dos nós (não-singulares), Σ_1 contém o nó da figura (a) e Σ_2 contém o nó da figura (b):

8 Não tendo sido abordado neste texto, trata-se do mais antigo invariante polinomial (1923).



Consideremos então o conjunto $\Sigma_\infty = \bigcup_{n=0}^{\infty} \Sigma_n$ de todos os nós singulares (mais os nós não-singulares). Dado um invariante numérico (que não é mais do que uma função[9] $v : \Sigma_0 \longrightarrow \mathbb{C}$) podemos estendê-lo a todo o conjunto Σ_∞ através da seguinte fórmula de recorrência:

(1) $$v(\times) = v(\times) - v(\times)$$

Deste modo determinamos os valores de Σ_{n+1} a partir dos valores de Σ_n.

Um invariante numérico de nós v diz-se um *invariante de Vassiliev de ordem* $\leq n$ se a sua extensão a Σ_∞ se anula para qualquer nó com mais de n pontos duplos, ou seja $v(k) = 0 \quad \forall k \in \Sigma_{n+1}$.

PROPOSIÇÃO 17. *Se $k \in \Sigma_n$ é um nó singular onde um dos seus pontos singulares é a única ligação entre duas partes do nó então $v(k) = 0$ para qualquer invariante de Vassiliev v. Diagramaticamente, o resultado pode ser apresentado da seguinte forma:*

$$v(\blacksquare\times\blacksquare) = 0 \quad \forall v \text{ invariante de Vassiliev.}$$

Demonstração. Usando a fórmula (1) no ponto em causa obtemos:

$$v(\blacksquare\times\blacksquare) = v(\blacksquare\times\blacksquare) - v(\blacksquare\times\blacksquare).$$

Se rodármos uma das partes do terceiro nó de 360 graus segundo o eixo horizontal obtemos o segundo nó, de modo que o resultado final é

$$v(\blacksquare\times\blacksquare) = 0.$$

□

Corolário 1. *Para qualquer invariante de Vassiliev é válida a seguinte relação (conhecida como relação de 1-termo ou 1T):*

$$v(\times\bigcirc) = 0$$

.

9 Para simplificar iremos considerar que o invariante toma valores em $\mathbb{C}$.

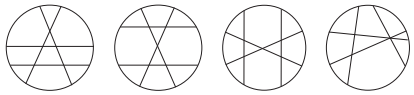
PROPOSIÇÃO 18. *Para qualquer invariante de Vassiliev é válida a seguinte fórmula (conhecida por relação dos 4-termos ou 4T):*

$$v(\ \ ) - v(\ \ ) - v(\ \ ) + v(\ \ ) = 0$$

Demonstração. Aplicando (1) nos pontos assinalados , , , , os termos resultantes acabam por se cancelar na equação. □

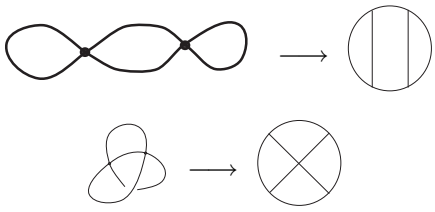
DEFINIÇÃO 19. Um *diagrama de cordas* (ou *diagrama de Gauss*) é uma circunferência com um número finito de cordas[10].

Exemplos de diagramas de cordas distintos



Exemplos de um mesmo diagrama de cordas

Um nó singular pode ser visto como uma aplicação da circunferência S^1 em $\mathbb{R}^3$ de pontos duplos simples (aliás é essa a definição mais usual na literatura). Se marcármos os pontos na circunferência que são pré-imagem dos pontos duplos e unirmos por cordas cada par de pontos que são levados para o mesmo ponto obtemos um diagrama de cordas. Temos então uma aplicação natural de Σ_∞ para o conjunto dos diagramas de cordas. Designemos esta aplicação por D e o conjunto dos diagramas de cordas por $\mathcal{D}$. Eis alguns exemplos:



LEMA 20. *Se v é um invariante de Vassiliev de ordem $\leq n$ e $K \in \Sigma_n$, então*

$$D(K') = D(K) \Rightarrow v(K') = v(K).$$

Ou seja, $v(K)$ depende únicamente do diagrama de cordas $D(K)$.

10 Note-se que dois diagramas de cordas são considerados idênticos se as extremidades das cordas estiverem na mesma ordem ao longo da circunferência.

Demonstração. A ideia da demonstração é que dois nós K e K' com o mesmo diagrama de cordas ($D(K) = D(K')$) podem ser deformados um no outro através de alterações de cruzamentos:  $\rightleftharpoons$  . Se K e K' têm n pontos duplos e v é um invariante de ordem $\leq n$, então a mudança de cruzamentos: $\rightleftharpoons$ não altera o valor de v e portanto $v(K) = v(K')$. Logo

$$\underset{n \text{ cruzamentos}}{v(\ \ )} \quad - \quad \underset{n \text{ cruzamentos}}{v(\ \ )} \quad = \quad \underset{n+1 \text{ cruzamentos}}{v(\ \ )} \quad = \ 0$$

. □

Desta forma, dado um invariante de Vassiliev v de ordem $\leq n$ obtemos uma aplicação que a um diagrama com n cordas D associa um valor numérico $v(K)$, onde K é um nó singular tal que $D(K) = D$.

Na verdade temos algo mais forte: existe uma aplicação do espaço vectorial V_n dos invariantes de Vassiliev[11] de ordem $\leq n$ para o espaço vectorial $\mathcal{D}_n^*$ das aplicações lineares de $\mathcal{D}_n$ em $\mathbb{C}$, onde $\mathcal{D}_n$ é o espaço vectorial gerado pelos diagramas com n cordas:

$$\begin{array}{rcl} \phi : V_n & \longrightarrow & \mathcal{D}_n^* \\ v & \longmapsto & \hat{v} \end{array}$$

com $\hat{v} : \mathcal{D}_n \longrightarrow \mathbb{C}$ definido por $\hat{v}(D) = v(K_D)$, onde K_D é um nó singular tal que $D(K_D) = D$.

É fácil ver que o núcleo desta aplicação é formado pelos invariantes de Vassiliev de ordem $\leq n-1$. Deste modo, temos uma aplicação injectiva de V_n/V_{n-1} em $\mathcal{D}_n^*$:

$$\begin{array}{rcl} \bar{\phi} : V_n/V_{n-1} & \longrightarrow & \mathcal{D}_n^* \\ v + V_{n-1} & \longmapsto & \hat{v} \end{array}$$

No entanto esta aplicação não é sobrejectiva pois qualquer $\hat{v} \in \mathcal{D}_n^*$ pertencente à imagem de $\bar{\phi}$ tem que obedecer a pelo menos duas restrições:

(i) A relação de 1-termo: $\hat{v}($  $) = 0$ (corresponde a $v($  $) = 0$);

(ii) a relação de 4-termos: $\hat{v}($  $)-\hat{v}($  $)-\hat{v}($  $)+\hat{v}($  $) =$ 0 (corresponde a $v($  $) - v($  $) - v($ $) + v($ $) = 0$).

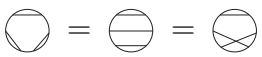
11 Qualquer espaço de funções com valores em $\mathbb{C}$, como é o caso dos invariantes de Vassiliev, possui uma estrutura de espaço vectorial onde as operações de adição e multiplicação por escalares são definidas ponto a ponto.

O teorema seguinte, fundamental na teoria dos invariantes de Vassiliev, diz-nos que estas são as únicas restrições a ter em conta.

TEOREMA 21 (KONTSEVICH). *O espaço vectorial V_n/V_{n-1} é isomorfo ao espaço dual*[12] *do espaço $\mathcal{D}_n$ quocientado pelas relações de 1-termo e 4-termos. Ou seja:*

$$V_n/V_{n-1} \cong (\mathcal{D}_n/(1T\&4T))^*$$

A demonstração deste teorema é demasiado técnica e extensa para ser exposta neste texto. Para um demonstração completa, ver [3]. Vejamos o caso particular $n = 3$: $\mathcal{D}_3$ é gerado pelo conjunto $\{$  $\}$. Quando tomamos o quociente pelas relações de 1-termo e 4-termos obtemos:

1-termo: 

4-termos: 

Portanto $\dim_{\mathbb{C}}(V_3/V_2) = \dim_{\mathbb{C}}(\mathcal{D}_3/(1T\&4T)) = 1$ (note-se que qualquer espaço de dimensão finita é isomorfo ao seu espaço dual).

No quadro seguinte escrevemos as dimensões de V_n/V_{n-1} para os valores mais baixos de n:

n	0	1	2	3	4	5	6	7	8	9	10	11	12
$\dim_{\mathbb{C}} V_n/V_{n-1}$	1	0	1	1	3	4	9	14	27	44	80	132	232

Para finalizar apresentamos uma conjectura que até agora tem resistido a uma demonstração ou contra-prova.

Conjectura 2 (Vassiliev)**.** *O conjunto dos invariantes de Vassiliev de tipo finito forma um invariante completo dos nós. Ou seja, $K \not\sim K' \Leftrightarrow \exists_n \exists_{v \in V_n} : v(K) \neq v(K')$.*

12 Por definição o espaço dual de um espaço vectorial complexo A é o espaço vectorial A^* das aplicações lineares de A em $\mathbb{C}$.

Referências

[1] J. W. Alexander, *Topological invariants of knots and links*, Trans. Amer. Math. Soc. **20** (1923), 93-95.

[2] C. W. Ashley, *The Ashley Book of Knots*, Doubleday, New York, 1944.

[3] D. Bar-Natan, *On the Vassiliev Knot Invariants*, Topology **34**, 423-472, 1995.

[4] G. Burde, H. Zieschang, *Knots*, de Gruyter, Berlin, 1986.

[5] G. Călugăranu, *L'intégrale de Gauss et l'analyse des nœuds tridimensionnels*, Rev. Roumaine Math. Pures Appl. **4** (1959), 5-20.

[6] J. H. Conway, *An enumeration of knots and links and some of their algebraic properties*, Computational Problems in Abstract Algebra (Proc. Conf., Oxford, 1967; J. Leech, ed.), Pergamon Press, New York, 1970, pp. 329-358.

[7] R. Courant, *Differential and Integral Calculus*, Vol. II, Interscience, New York, 1953.

[8] W. Haken, *Theorie der Normalflächen*, Acta Math. **105** (1961), 245-375.

[9] W. Haken, *Über das Homöomorphieproblem der 3-Mannigfaltigkeiten* I, Math. Z. **80** (1962), 89-120.

[10] G. Hemion, *The Classification of Knots and 3-dimensional Spaces*, Oxford Univ. Press, Oxford, 1992.

[11] V. Jones, *A polynomial invariant for knots via von Neumann algebras*, Bull. Amer. Math. Soc. (N.S.) **12** (1985), 103-111.

[12] V. Prasolov, A. Sossinsky, *Knots, Links, Braids and* 3*-Manifolds*, American Mathematical Society, 1997.

[13] H. Schubert, *Die eindeutige Zerlegbarkeit eines Knoten in Primknoten*, Sitzungsber. Heidelberg. Akad. Wiss. Math.-Nat. Kl. **1949**, no. 3, 57-104.

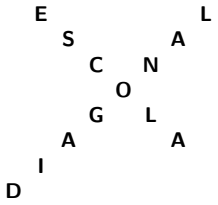
[14] P. G. Tait, *On knots* I, II, III, Trans. Roy. Soc. Edinburgh **28** (1879), 145-190; **32** (1887), 327-342, 493-506; reprinted in his *Scientific Papers*, Vol. 1, Cambridge Univ. Press, Cambridge, 1898, 273-347.

[15] M. B. Thistlehwaite, *Knot tabulations and related topics*, Aspects of Topology (I. M. James and E. H. Kronheimer, eds.), Cambridge Univ. Press,

Cambridge, 1985, 1-76.

[16] W. Thomson, *On vortex motion*, Trans. Roy. Soc. Edinburgh **25** (1867), 217-260.

[17] A. T. Vandermonde, *Remarques sur les problèmes de situation*, Mémoires de l'Académie Royale des Sciences (Paris)(1771), 566-574.



Formas Quadráticas e Fracções Contínuas

Luís Diogo
Departamento de Matemática
Instituto Superior Técnico
luisgeraldesdiogo@hotmail.com

Carlos Florentino
Departamento de Matemática
Instituto Superior Técnico
cfloren@math.ist.utl.pt

Diogo Veloso
Departamento de Matemática
Instituto Superior Técnico
diogo_veloso@hotmail.com

Resumo

Este artigo é baseado no curso "Quadratic Forms and Continued Fractions", de R. Takloo-Bighash, integrado na Escola Diagonal que decorreu em Setembro de 2005, no Instituto Superior Técnico.

1 Introdução

A teoria dos números é uma das disciplinas mais antigas da matemática. Pode dizer-se que um dos seus objectivos é encontrar soluções inteiras ou racionais de equações polinomiais a várias variáveis, por exemplo

$$45x^7y^9 + 8zy^3 - 23w^4 + 1 = 0.$$

Claro que à partida não é fácil arranjar soluções de uma equação deste género, ou provar que não existem soluções. Como fazê-lo, então? Existe algum método? De um modo geral não. Mesmo provar que não existem soluções não triviais[1] de

$$x^3 + y^3 = z^3$$

é já complicado.

As equações polinomiais com coeficientes inteiros e para as quais se procuram soluções inteiras são chamadas equações diofantinas, em homenagem a Diofanto, um matemático do século III, de Alexandria.

1 Isto é, com um dos inteiros x, y, z igual a zero.

Então que tipo de equações vamos estudar? De um modo geral, vamos estudar problemas relacionados com polinómios do tipo

$$ax^2 + bxy + cy^2$$

onde $a, b, c \in \mathbb{Z}$, designados por formas quadráticas. Neste caso, sim, existem várias teorias para responder aos diferentes problemas que iremos colocar.

Vamos primeiro abordar a equação de Pell

$$x^2 - Dy^2 = 1$$

onde D é um inteiro positivo não quadrado[2]. Esta é uma equação antiga, tendo já sido estudada por vários matemáticos indianos antes de chegar à Europa. De facto, Brahmagupta estudou-a no século VII e Bhaskara no século XII. Na Europa, o primeiro matemático a encontrar uma solução geral foi Brouncker, no século XVII. Mais tarde Euler confundiu Brouncker com Pell, outro matemático da época, e chamou-lhe equação de Pell. Vamos resolver esta equação através de fracções contínuas.

Em seguida veremos como achar todas as soluções inteiras de

$$x^2 + y^2 = z^2,$$

os chamados triplos pitagóricos.

Na parte final vamos estudar que números podem ser representados por formas quadráticas definidas positivas, isto é

$$ax^2 + bxy + cy^2$$

com $a > 0$ e $4ac - b^2 > 0$. Por exemplo, que números se podem escrever na forma $x^2 + y^2$ ou $x^2 + ny^2$?

Para estudar o problema acima vamos relacioná-lo com uma determinada acção do grupo de matrizes 2 por 2 com coeficientes inteiros e determinante 1 no plano hiperbólico. Esta teoria vai permitir dizer que números se podem escrever da forma $x^2 + ny^2$, para alguns n. Este assunto está bastante relacionado com o estudo dos inteiros num corpo quadrático.

Apesar da longa história e de constituirem os primeiros casos não triviais de equações diofantinas, a investigação da aritmética das equações quadráticas está longe de ter terminado. No fim deste artigo apresentamos alguns resultados recentes sobre este "velho" assunto da teoria dos números.

O objectivo destas notas foi dar uma ideia mais prática do que teórica de como arranjar soluções de determinados problemas. Ao longo do texto

2 Ou seja, D não é um quadrado perfeito.

há bastantes exercícios, de vários graus de dificuldade, que formam parte integrante deste artigo. Encorajamos o leitor a tentar fazer a maior parte. Para quem estiver também interessado nas vertentes mais teóricas destes assuntos, ou na teoria dos números em geral, sugerimos os livros [IR, K, MT, ST].

Exercício 1. Encontre uma solução não trivial de $x^2 - 13y^2 = 1$.

Agradecimentos. Gostaríamos de agradecer à organização da Escola Diagonal e ao Professor Ramin Takloo-Bighash, às Professoras Sílvia Anjos e Esmeralda Dias, e ao Gonçalo Tabuada pela ajuda nas sessões de problemas.

2 Noções Básicas de Aritmética

Antes de abordar os problemas referidos, vamos começar por rever alguns conceitos básicos. Uma referência para esta secção é, por exemplo, [K]. O nosso primeiro problema é encontrar as soluções $x, y \in \mathbb{Z}$ de

$$ax + by = c$$

onde a, b, c são inteiros dados.

Definição 1. Se $a, b \in \mathbb{Z}$, dizemos que a *divide* b, ou que b é *divisível* por a, ou ainda que a é um *divisor* de b se existe $n \in \mathbb{Z}$ tal que $b = an$. Nesse caso escrevemos $a|b$.

Note-se que $1|b$ para qualquer inteiro b, e que pelo contrário, 0 só é divisor de si próprio. Note-se também que excepto para ± 1, nenhum inteiro divide 1.

Proposição 2. *i) Se $a|b$ e $b|c$ então $a|c$;*
ii) Se $a|b$ e $a|c$ então $a|(b \pm c)$;
iii) Se r é inteiro e $a|b$ então $a|br$.

Definição 3. Dizemos que $g > 0$ é o máximo divisor comum de a e b, $g = \gcd(a, b)$,[3] se para qualquer d tal que $d|a$ e $d|b$ se tem $d|g$.

Exercício 2. Sejam a e b inteiros. Prove que $\gcd(\lambda a, \lambda b) = \lambda \gcd(a, b)$.

Teorema 4. *[Algoritmo da divisão] Sejam $a, b \in \mathbb{N}$. Então existe um único $q \in \mathbb{N}$ e um único $r \in \{0, ..., b-1\}$ tal que $a = bq + r$.*

3 Do inglês *greatest common divisor*.

Demonstração. Primeiro provamos a existência. Seja q a parte inteira de a/b, isto é $q = \lfloor a/b \rfloor$ é o único inteiro tal que $q \leq a/b < q+1$. Então

$$0 \leq a - bq = b(a/b - q) < b.$$

Assim, definindo r como $a - bq$ temos $0 \leq r \leq b-1$. Deixamos a prova da unicidade ao leitor. □

Apresentamos agora um algoritmo que fornece o $\gcd(a,b)$ de quaisquer inteiros positivos a e b.
Suponhamos, sem perda de generalidade, que $a > b$. Pondo $a_1 = a, b_1 = b$, temos, pelo algoritmo da divisão, que existem $q_1, r_1 \in \mathbb{Z}$ com $0 \leq r_1 \leq b_1 - 1$ tais que:

$$a_1 = b_1 q_1 + r_1.$$

Se $r_1 = 0$ então o algorítmo termina e $\gcd(a,b) = b$. Se $r_1 > 0$, pomos $a_2 = b_1, b_2 = r_1$ e existem, como antes, $q_2, r_2 \in \mathbb{N}_0$ com $0 \leq r_2 \leq b_2 - 1 = r_1 - 1$. Logo $r_2 < r_1$. Se $r_2 > 0$, pomos de novo $a_{k+1} = b_k, b_{k+1} = r_k$ e temos, indutivamente que, $r_{k+1} < r_k$, logo r_k é decrescente e acabará por ser zero. Representando estas divisões sucessivas, obtemos:

$$a_1 = b_1 q_1 + r_1$$

$$a_2 = b_2 q_2 + r_2$$

$$\vdots$$

$$a_{k-2} = b_{k-2} q_{k-2} + r_{k-2}$$

$$a_{k-1} = b_{k-1} q_{k-1} + r_{k-1}$$

$$a_k = b_k q_k.$$

Finalmente, b_k será o $\gcd(a,b)$. De facto b_k divide $a_k = b_{k-1}$ e $b_k = r_{k-1}$ e assim $b_k | a_{k-1} (= b_{k-2})$. De novo b_k divide $r_{k-2} = b_{k-1} = a_k$ logo divide a_{k-2}, e assim sucessivamente. Temos então que b_k divide b_1 e a_1.
Por outro lado, se $d|a = a_1$, e $d|b (= b_1 = a_2)$ então d divide $r_1 = b_2 = a_3$, logo $d|r_2 (= b_3)$. Do mesmo modo $d|r_3$ e assim sucessivamente, até obtermos que d é divisor de $r_{k-1} = b_k$. Logo b_k é o $\gcd(a,b)$. Assim, temos

TEOREMA 5. *[Algoritmo de Euclides] Sejam $a, b \in \mathbb{N}$. Então o algoritmo descrito acima fornece o* $\gcd(a,b)$ *num número finito de passos.*

Exemplo 1. Se $a = 84, b = 60$, o algoritmo de Euclides diz-nos que

$$84 = 60 \times 1 + 24$$

$$60 = 24 \times 2 + 12$$

$$24 = 12 \times 2.$$

Assim $\gcd(84, 60) = 12$.

Exercício 3. Calcule o gcd dos seguintes números:

- 627 e 308;
- 1260 e 692.

Estamos agora em condições de resolver o problema proposto acima: encontrar soluções $x, y \in \mathbb{Z}$ de

$$ax + by = c$$

onde a, b, c são inteiros dados.

Em primeiro lugar, para que a equação tenha solução é necessário que $\gcd(a, b)|c$. De facto, o $\gcd(a, b)$ divide a e divide b logo tem que dividir c. Acontece que esta condição é também suficiente. Se $g = \gcd(a, b)$ então existem infinitos $x, y \in \mathbb{Z}$ tais que $ax + by = g$. Apesar de não o provarmos aqui, o próximo exemplo contém detalhes suficientes para permitir ao leitor completar a prova por si.

Exemplo 2. Para encontrar uma solução de $68x + 255y = 340$, primeiro calculamos o $\gcd(68, 255)$

$$255 = 68 \times 3 + 51$$

$$68 = 51 \times 1 + 17$$

$$51 = 17 \times 3.$$

Logo $\gcd(68, 255) = 17$. Como 17 divide 340 então o problema acima tem solução. Através dos cálculos acima conseguimos arranjar a, b tais que $68a + 255b = 17$. De facto

$$17 = 68 - 51 = 68 - (255 - 68 \times 3) = 68 \times 4 - 255 = 68 \times 4 + 255 \times (-1).$$

Assim temos $a = 4$ e $b = -1$. Para arranjar uma solução de $68x + 255y = 340 = 17 \times 20$ só temos que multiplicar as soluções acima por 20. O resultado é $x = 80$ e $y = -20$. Mas será este resultado único? Não. De facto, pondo $x = 80 - 15k$ e $y = -20 + 4k$, com k inteiro, obtemos todas as soluções do problema. Consegue explicar porquê?

Agora relembramos a definição de número primo:

DEFINIÇÃO 6. Um número $p > 1$ diz-se *primo* se os únicos divisores positivos de p são 1 e p.

Uma propriedade muito útil dos números primos, e que pode facilmente ser verificada, é a seguinte.

PROPOSIÇÃO 7. *Um número $p > 1$ é primo se e só se para todo $a, b \in \mathbb{Z}$, $p|ab$ implica $p|a$ ou $p|b$.*

Por causa da sua importância costuma-se chamar *teorema fundamental da aritmética* ao seguinte resultado.

TEOREMA 8. *Seja $n \in \mathbb{Z}$. Então existem primos $p_1, ..., p_m$ e inteiros positivos $\alpha_1, ..., \alpha_m$, tais que*

$$n = \pm p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_m^{\alpha_m}.$$

Esta representação é única a menos de permutação.

Exemplo 3. $234235 = 5 \times 79 \times 543$;

- $7112 = 2^3 \times 7 \times 127$.

Corolário 1. *Existem infinitos números primos.*

Demonstração. Suponhamos que existia apenas um número finito de primos, $p_1, ..., p_n$ digamos. Consideremos então o número $N = 1 + p_1 \cdots p_n$. Pelo teorema acima, N tem que ser produto de números primos. Suponhamos que p_i é um desses primos. Assim $p_i|N(= 1 + p_1 \cdots p_n)$. Mas $p_i|p_1 \cdots p_n$, e portanto tem que dividir 1, o que é absurdo. □

Exercício 4. Dado um inteiro n, encontre n inteiros consecutivos, nenhum dos quais é primo.

Agora definimos um conceito que nos vai ser útil.

DEFINIÇÃO 9. [Relação de congruência] Sejam $a, b \in \mathbb{Z}$ e $n \in \mathbb{N}$, escrevemos $a \equiv b \mod n$ se $n|(b - a)$.

Exercício 5. Prove que $\equiv$ define uma relação de equivalência, i.e, para todo o $m \in \mathbb{N}$, $a, b, c \in \mathbb{Z}$, temos.

- $a \equiv a \mod m$;
- $a \equiv b \mod m \Rightarrow b \equiv a \mod m$;
- $a \equiv b \mod m$, e $b \equiv c \mod m \Rightarrow a \equiv c \mod m$.

A classe de equivalência de $a \in \mathbb{Z}$ é o conjunto $\{a + mk : k \in \mathbb{Z}\}$. Representa-mo-lo por $a + m\mathbb{Z}$. O conjunto das classes de equivalência $\{a + m\mathbb{Z} : a \in \mathbb{Z}\}$ é representado por $\mathbb{Z}/m\mathbb{Z}$.

PROPOSIÇÃO 10. *As seguintes operações estão bem definidas nas classes de equivalência* $\mod m$

- $(a + m\mathbb{Z}) + (b + m\mathbb{Z}) = (a + b) + m\mathbb{Z}$
- $(a + m\mathbb{Z}) \cdot (b + m\mathbb{Z}) = (a \cdot b) + m\mathbb{Z}$

Demonstração. Só provamos o primeiro ponto, deixando o segundo para o leitor. Sejam $A \in a + m\mathbb{Z}, B \in b + m\mathbb{Z}$, i.e. $A = a + mr$ para algum $r \in \mathbb{Z}$, e $B = b + ms$ para algum $s \in \mathbb{Z}$. Temos então que:

$$A + B = a + mr + b + ms = a + b + m(r + s)$$

e logo $A + B \in (a + b) + m\mathbb{Z}$. □

Exercício 6. Prove que se p é um número primo, então $(a+p\mathbb{Z})\cdot(b+p\mathbb{Z}) = p\mathbb{Z}$ implica $a + p\mathbb{Z} = p\mathbb{Z}$ ou $b + p\mathbb{Z} = p\mathbb{Z}$.

É fácil ver que, para quaisquer $a, b \in \mathbb{Z}$ as equações $a + m\mathbb{Z} = b + m\mathbb{Z}$ e $a \equiv b \mod m$ são equivalentes. Assim, a Proposição anterior fornece as seguintes regras de cálculo que são muito usadas na prática.

PROPOSIÇÃO 11. *Sejam* $a, a', b, b' \in \mathbb{Z}$ *e* $m \in \mathbb{N}$*, verificando* $a \equiv a' \mod m$ *e* $b \equiv b' \mod m$*. Então*

$$\begin{aligned} a + b &\equiv a' + b' \mod m \\ a \cdot b &\equiv a' \cdot b' \mod m. \end{aligned}$$

Exercício 7. Sejam a, b primos entre si, i.e. tais que $\gcd(a, b) = 1$. Sejam também $x_1, x_2 \in \mathbb{Z}$. Prove que existe uma única solução de:

$$\begin{cases} x \equiv x_1 \mod a \\ x \equiv x_2 \mod b \end{cases}$$

com $0 \leq x < a \cdot b$. Assim esta é a única solução $\mod a \cdot b$. Isto é, se y for outra solução, então $x \equiv y \mod a \cdot b$.

Agora vamos enunciar um teorema que já era conhecido por matemáticos chineses do século III. Como já fizemos noutro caso não o vamos provar, mas deixamos ao leitor detalhes suficientes para que o possa fazer por si.

TEOREMA 12. *[Teorema Chinês dos Restos] Sejam* $m_1, m_2, \ldots, m_k \in \mathbb{Z}$*, tais que* $\gcd(m_i, m_j) = 1$ *para todos os índices* $i, j = 1, ..., k$ *distintos. Então para todo o* $a_1, \ldots, a_k \in \mathbb{Z}$*, o sistema:*

$$\begin{cases} x \equiv a_1 \mod m_1 \\ \vdots \\ x \equiv a_k \mod m_k \end{cases}$$

tem sempre uma solução $x \in \mathbb{Z}$.

Exemplo 4. Queremos encontrar $x \in \mathbb{Z}$ tal que

$$\begin{cases} x \equiv 3 \mod 7 \\ x \equiv 8 \mod 11 \\ x \equiv 1 \mod 13 \end{cases}.$$

Assim $x = 7a + 3$ para algum $a \in \mathbb{Z}$, $x = 11b + 8$ para algum $b \in \mathbb{Z}$, e ainda $x = 13c + 1$ para algum $c \in \mathbb{Z}$. Donde sai, em particular, que $7a + 3 = 11b + 8$.

Logo

$$7a + 11(-b) = 5.$$

Apliquemos o algoritmo de Euclides a 7 e 11.

$$\begin{aligned} 11 &= 7 \times 1 + 4 \\ 7 &= 4 \times 1 + 3 \\ 4 &= 3 \times 1 + 1 \end{aligned}$$

Assim temos que

$$1 = 4 - 3 = 4 - (7 - 4) = (11 - 7) - (7 - (11 - 7)) = 7 \times (-3) + 11 \times 2.$$

Portanto $a = -15$ e $b = -10$ verificam $7(-15)+11(-10) = -105+110 = 5$. Pelo exercício anterior $x \equiv 7(-15) + 3 \equiv -102 \equiv 52 \mod 7 \times 11$.

Assim reduzimos o problema acima a encontrar soluções de

$$\begin{cases} x \equiv 52 \mod 77 \\ x \equiv 1 \mod 13 \end{cases}$$

Como no início, temos $x = 77d + 52$ para algum $d \in \mathbb{Z}$ e $x = 13e + 1$ para algum $e \in \mathbb{Z}$. Logo $13b - 77a = 51$. Aplicando o algoritmo de Euclides a $(77, 13)$,

$$\begin{aligned} 77 &= 13 \times 5 + 12 \\ 13 &= 12 \times 1 + 1. \end{aligned}$$

Assim temos que

$$1 = 13 - 12 = 13 - (77 - 13 \times 5) = 13 \times 6 - 77.$$

Portanto $a = 51$ e $b = 51 \times 6 = 306$. Logo uma solução $\mod 77 \times 13 = 1001$ é dada, por exemplo, por: $13 \times 306 + 1 = 3979 \equiv 976 \mod 1001$. De facto $976 = 7 \times 139 + 3 = 11 \times 88 + 8 = 13 \times 75 + 1$.

Exercício 8. Resolva o sistema

$$\begin{cases} x \equiv 1 \mod 2 \\ x \equiv 1 \mod 3 \\ x \equiv 1 \mod 5 \end{cases}$$

Exercício 9. Demonstre o Teorema chinês dos restos.

Definimos agora, com o único objectivo de familiarizar o leitor, os conceitos de anel e corpo.

Definição 13. Dizemos que um conjunto A com duas operações associativas $+$ e $\cdot$ é um *anel* se para quaisquer $a, b, c \in A$ se tem

- Existem elementos $0 \in A$ tal que $a + 0 = a$, e $1 \in A$ tal que $a \cdot 1 = a$
- Existe $a' \in A$ tal que $a + a' = 0$
- $a + b = b + a$ e $a \cdot b = b \cdot a$
- $a \cdot (b + c) = a \cdot b + a \cdot c$.

Ao elemento a' da segunda propriedade acima, é usual denotar por $-a$.

Assim um anel tem, num certo sentido, as mesmas propriedades que $\mathbb{Z}$. A Proposição 10 diz-nos que $(\mathbb{Z}/m\mathbb{Z}, +, \cdot)$ é um anel. Deste modo podemos simplificar expressões com variáveis em $\mathbb{Z}/m\mathbb{Z}$ usando as regras que estamos habituados a usar em $\mathbb{Z}$. Note-se contudo que não existe divisão definida. De facto em $\mathbb{Z}/4\mathbb{Z}$ o elemento $2+4\mathbb{Z}$ não tem inverso pois $2 \times 2 \equiv 0 \mod 4$. Um anel $(A, +, \cdot)$ diz-se um corpo se verificar a seguinte propriedade adicional

- Para $a \neq 0$ existe $a'' \in A$ tal que $a \cdot a'' = 1$.

Como acima, é usual denotar o elemento a'' por a^{-1} ou por $1/a$.

Notemos que se p é primo, $\mathbb{Z}/p\mathbb{Z}$ é um corpo. Logo temos a divisão definida. Deste modo qualquer equação da forma $ax \equiv b \mod p$ com $a, b \in \mathbb{Z}$ e $p \nmid a$ tem solução $x \in \mathbb{Z}$.

O que se segue ser-nos-á útil mais adiante, quando tratarmos da equação de Pell.

Definição 14. Dado um inteiro não quadrado D, definimos $\mathbb{Z}[\sqrt{D}] := \{z = x + y\sqrt{D} : x, y \in \mathbb{Z}\}$.

$\mathbb{Z}[\sqrt{D}]$ é um anel com a soma e a multiplicação naturais. A $\mathbb{Z}[\sqrt{-1}]$ chamamos o anel dos inteiros Gaussianos.

Definição 15. O *conjugado* de $z = x + y\sqrt{D}$ é $\bar{z} = x - y\sqrt{D}$. A *norma* de $z \in \mathbb{Z}[\sqrt{D}]$ é $N(z) = z\bar{z}$.

Observe que temos $x^2 - Dy^2 = (x + y\sqrt{D})(x - y\sqrt{D}) = z\bar{z}$.

Notamos que no caso $D = -1$, estas noções correspondem precisamente às noções de conjugado e de norma nos números complexos.

Exercício 10. Prove que $N(zw) = N(z)N(w)$.

- Prove que uma solução da equação de Pell é o mesmo que um elemento $z \in \mathbb{Z}[\sqrt{D}]$ tal que $N(z) = 1$.

Assim, se z for tal que $N(z) = 1$ então $N(z^k) = 1$ para qualquer $k \in \mathbb{N}$. No caso $D > 0$ isto significa que se houver uma solução não trivial da equação de Pell então há infinitas soluções.[4] Na próxima secção, provaremos que qualquer equação $x^2 - Dy^2 = 1$ tem sempre soluções.

Exemplo 5. Para $D = 3$ temos

$$N(2 + \sqrt{3}) = 1.$$

Logo, como $(2 + \sqrt{3})^2 = 7 + 4\sqrt{3}$, temos também que

$$N(7 + 4\sqrt{3}) = 1.$$

Vemos assim que os pares $(2, 1)$ e $(7, 4)$ são soluções da equação $x^2 - 3y^2 = 1$.

Exercício 11. Considere $D = -1$. Prove, usando as propriedades da norma, que se $X^2+Y^2 = A$ e $Z^2+W^2 = B$, então existe uma solução de $x^2+y^2 = AB$.

Prove também que se $\gcd(X, Y) = \gcd(Z, W) = 1$, então a solução que se obtém acima tem também $\gcd = 1$.

Observação 1. Por analogia com os inteiros, os primos no anel $\mathbb{Z}[\sqrt{D}]$ são definidos como os números $z = x + y\sqrt{D}$ tais que para toda a decomposição de z como $a \cdot b$, onde $a, b \in \mathbb{Z}[\sqrt{D}]$, temos $z|a$ ou $z|b$. Contudo a factorização única em números primos não se verifica na maior parte destes anéis. Perceber este caso mais geral foi um passo importante na história da Álgebra e da Teoria dos números.

3 Fracções Contínuas e a Equação de Pell

Desde tempos remotos que sábios procuraram o valor exacto de alguns números, por exemplo π, $\sqrt{2}$, etc. Por exemplo o valor dado a $\sqrt{2}$ na Babilónia era $17/12$. Arquimedes no séc.III ac provou que $223/71 < \pi < 22/7$. O indiano

4 Isto acontece porque, quando $D > 0$, $\mathbb{Z}[\sqrt{D}]$ não tem raízes da unidade para além de ± 1.

Aryabatha usava $\pi = 3,1416$ no séc.V, e no séc.VI o chinês Zu Chongzie usava já $\pi = 355/113$, que como veremos é uma aproximação muito boa.

Estamos assim, interessados em encontrar a melhor aproximação racional a um dado número. Para isso utilizaremos fracções contínuas. Mais tarde veremos que, surpreendentemente, tudo isto está relacionado com a equação de Pell.

Começamos então por definir fracção contínua.

DEFINIÇÃO 16. Uma *fracção contínua* é uma expressão da forma

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}},$$

onde $a_k \in \mathbb{Z}$ e $a_1, a_2, ... > 0$. Representamo-la por $[a_0, a_1, a_2, \ldots]$.

Chamamos a

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}} = [a_0, ..., a_n]$$

o *n-ésimo convergente* da fracção contínua. Note que $[a_0, ..., a_n] = \frac{p_n}{q_n}$ para determinados $p_n, q_n \in \mathbb{N}$, que tomaremos como sendo primos entre si.

Exercício 12. Prove que

- $p_0 = a_0$, $q_0 = 1$;
- $p_1 = a_0a_1 + 1$, $q_1 = a_1$;
- $p_n = a_np_{n-1} + p_{n-2}$, $q_n = a_nq_{n-1} + q_{n-2}$.

Exercício 13. Usando indução, prove que:

- $p_nq_{n-1} - p_{n-1}q_n = (-1)^{n-1}$;
- $p_nq_{n-2} - p_{n-2}q_n = (-1)^n a_n$.

Podemos levantar algumas questões sobre fracções contínuas:

- O que se quer dizer com um processo de soma e divisão infinitos?

- Que números podem ser representados por fracções contínuas?
- Essa representação é única?

Pode-se provar que para cada fracção contínua, a sucessão definida pelos convergentes $\frac{p_n}{q_n}$ tem um limite. Este é o significado da fracção contínua $[a_0, ..., a_n, ...]$. Também se pode mostrar que todo número real se pode escrever como fracção contínua. Por exemplo para π, temos

$$\pi = 3 + \pi - 3.$$

Escolhemos 3 porque é o único inteiro tal que $0 \leq \pi - 3 < 1$. Mas queremos escrever

$$\pi = 3 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}.$$

Logo, a_1 tem que verificar

$$a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}} = \frac{1}{\pi - 3} = 7.0625...$$

Assim, $a_1 = 7$ e podemos continuar até termos encontrado a_n.

Existe, de facto, alguma ambiguidade mas *só em fracções contínuas finitas*! Por exemplo, $a_0 = (a_0 - 1) + \frac{1}{1}$, logo $[a_0] = [a_0 - 1, 1]$, ou mais geralmente $[a_0, ..., a_n] = [a_0, ..., a_n - 1, 1]$.

É claro que qualquer fracção contínua finita representa um número racional. O que já não é tão claro é que qualquer número racional é representado por uma fracção contínua finita. Seja $\frac{a}{b}$ um número racional. Dividindo a por b temos que, $a = q_1 b + r_1$, para algum $0 \leq r_1 \leq b - 1$. Logo:

$$\frac{a}{b} = \frac{q_1 b + r_1}{b} = q_1 + \frac{r_1}{b} = q_1 + \frac{1}{\frac{b}{r_1}}.$$

Agora dividimos b por r_1. De novo, $b = q_2 r_1 + r_2$ para algum $0 \leq r_2 \leq r_1 - 1$. Assim:

$$\frac{a}{b} = q_1 + \frac{1}{\frac{q_2 r_1 + r_2}{r_1}} = q_1 + \cfrac{1}{q_2 + \cfrac{r_2}{r_1}}.$$

Se o leitor prestar atenção verificará que o que se tem estado a fazer é exactamente o mesmo que no *Algoritmo de Euclides*. Assim, o processo tem que parar e a fracção contínua será finita.

Exemplo 6. Vamos desenvolver em fracção contínua $\frac{255}{68}$:

$$\frac{255}{68} = \frac{3 \times 68 + 51}{68} = 3 + \frac{1}{\frac{68}{51}} =$$

$$3 + \frac{1}{1 + \frac{17}{51}} = 3 + \frac{1}{1 + \frac{1}{3}}.$$

Exemplo 7. Dos seguintes cálculos pode-se ver facilmente que $\sqrt{7} = [2, 1, 1, 1, 4, 1, 1, 1, 4, ...]$:

$$\begin{aligned}
\sqrt{7} &= 2 + \sqrt{7} - 2 \\
\frac{1}{\sqrt{7} - 2} &= \frac{\sqrt{7} + 2}{3} = 1 + \frac{\sqrt{7} - 1}{3} \\
\frac{3}{\sqrt{7} - 1} &= \frac{3(\sqrt{7} + 1)}{6} = \frac{\sqrt{7} + 1}{2} = 1 + \frac{\sqrt{7} - 1}{2} \\
\frac{2}{\sqrt{7} - 1} &= \frac{\sqrt{7} + 1}{3} = 1 + \frac{\sqrt{7} - 2}{3} \\
\frac{3}{\sqrt{7} - 2} &= \sqrt{7} + 2 = 4 + \sqrt{7} - 2
\end{aligned}$$

Observação 2. Para fracções contínuas periódicas vamos colocar uma barra sobre a parte periódica. Assim por exemplo a fracção contínua de $\sqrt{7}$ é escrita como $\left[2, \overline{1, 1, 1, 4}\right]$. A equação de Pell está relacionada com este tipo de fracções contínuas.

Os dois teoremas que enunciamos em seguida relacionam a fracção contínua com as aproximações racionais de um dado número. Estas dizem que um número racional é uma boa aproximação de ξ, se e só se for um convergente da fracção contínua de ξ. Assim as melhores aproximações são convergentes da fracção contínua e vice-versa.

Teorema 17. *Seja* $\xi = [a_0, a_1, \ldots]$. *Se* $\frac{a}{b}$ *verifica*

$$\left|\xi - \frac{a}{b}\right| < \frac{1}{2b^2},$$

então

$$\frac{a}{b} = \frac{p_n}{q_n},$$

para algum n. *Além disso, se* $\gcd(a, b) = 1$, *então* $a = p_n$ *e* $b = q_n$.

Por outro lado também temos o seguinte.

TEOREMA 18. *Seja* $\xi = [a_0, a_1, \ldots]$. *Então*

$$\left|\xi - \frac{p_n}{q_n}\right| < \frac{1}{q_n^2}$$

para qualquer n.

O recíproco deste último teorema não é verdadeiro.

Exercício 14. Seja $\xi \in \mathbb{R}$. Encontre números p e q que verificam $\left|\xi - \frac{p}{q}\right| < \frac{1}{q^2}$ e tais que $\frac{p}{q}$ não é um convergente de ξ.

Exemplo 8. Temos que $\frac{24}{5}$ é uma boa aproximação $\sqrt{23}$. De facto

$$\left|\sqrt{23} - \frac{24}{5}\right| < 0.0042 < \frac{1}{2 \times 25} = 0.02.$$

Logo $\frac{24}{5}$ tem que ser igual a um convergente da fracção contínua de $\sqrt{23}$. De facto $\frac{24}{5} = [4, 1, 3, 1]$, e $\sqrt{23} = \left[4, \overline{1, 3, 1, 8}\right]$.

Agora poderíamos perguntar porque é que obtivemos de novo uma fracção contínua periódica da raiz quadrada de um inteiro não quadrado. Mais abaixo veremos que não é um acaso.

Entretanto damos um exemplo histórico.

Exemplo 9. Com uma calculadora podem-se obter os primeiros "dígitos" da fracção contínua de π

$$\pi = [3, 7, 15, 1, 292, 1, 1, ...]$$

n	a_n	p_n	q_n
0	3	3	1
1	7	$22 = 3 \times 7 + 1$	$7 = 7 \times 1 + 0$
2	15	$333 = 22 \times 15 + 3$	$106 = 7 \times 15 + 1$
3	1	$355 = 333 \times 1 + 22$	$113 = 106 \times 1 + 7$
4	292	$103993 = 355 \times 292 + 333$	$33102 = 106 \times 292 + 106$

Assim, $\frac{355}{113}$ é uma boa aproximação de π, e $\frac{103993}{33102}$ uma excelente.

Respondemos agora à questão proposta mais acima.

PROPOSIÇÃO 19. *Um número real* ξ *tem fracção contínua periódica se e só se* ξ *é uma solução de um polinómio de grau dois com coeficientes inteiros.*

Demonstração. Só vamos fazer a prova numa direcção e apenas no caso das fracções contínuas completamente periódicas. Então tomemos um número real α com fracção contínua periódica.

$$\alpha = \overline{[b_1, b_2, \ldots, b_m]} = b_1 + \cfrac{1}{b_2 + \cfrac{1}{\ddots \cfrac{1}{b_m + \cfrac{1}{b_1 + \cfrac{1}{b_2 + \ddots}}}}} = b_1 + \cfrac{1}{b_2 + \cfrac{1}{\ddots \cfrac{1}{b_m + \cfrac{1}{\alpha}}}}$$

Começamos por observar que

$$b_{m-1} + \cfrac{1}{b_m + \cfrac{1}{\alpha}} = b_{m-1} + \frac{\alpha}{b_m\alpha + 1} = \frac{(b_m\alpha + 1)b_{m-1} + \alpha}{b_m\alpha + 1} = \frac{A\alpha + B}{C\alpha + D}$$

onde $A, B, C, D \in \mathbb{Z}$. Analogamente,

$$b_{m-2} + \cfrac{1}{b_{m-1} + \cfrac{1}{b_m + \cfrac{1}{\alpha}}} = b_{m-2} + \cfrac{1}{\cfrac{A\alpha + B}{C\alpha + D}} = \ldots = \frac{A'\alpha + B'}{C'\alpha + D'}$$

$$\vdots$$

e obtemos por fim que

$$\alpha = \frac{a\alpha + b}{c\alpha + d}$$

para determinados $a, b, c, d \in \mathbb{Z}$. Logo,

$$c\alpha^2 + \alpha d = a\alpha + b \quad \Rightarrow \quad c\alpha^2 + (d - a)\alpha - b = 0.$$

assim, α é a solução de uma equação quadrática com coeficientes inteiros. Para a demonstração do recíproco, consulte-se [MT]. □

Agora, e finalmente, relacionamos fracções contínuas com soluções da equação de Pell.

Exercício 15. Prove que se $(x, y) \in \mathbb{Z}^2$ é solução de $x^2 - Dy^2 = 1$, então $\frac{x}{y}$ é um convergente da fracção contínua de $\sqrt{D}$.

Deste modo estamos a relacionar soluções da equação de Pell com convergentes da fracção contínua de $\sqrt{D}$. De facto temos:

TEOREMA 20. *Se* $\sqrt{D} = [a_0, \overline{a_1, ..., a_n}]$ *então* $p_{n-1}^2 - Dq_{n-1}^2 = \pm 1$, *onde* $\frac{p_{n-1}}{q_{n-1}} = [a_0, a_1, ..., a_{n-1}]$.

Agora convém ao leitor ter bem presente a definição e propriedades da norma que foram apresentadas na parte final da segunda secção.

Exemplo 10. Como encontrar uma solução de $x^2 - 17y^2 = 1$? Primeiro temos que saber qual é a fracção contínua de $\sqrt{17} = \left[4, \overline{8}\right]$ Como $[4] = \frac{4}{1}$ obtemos que $4^2 - 17 \times 1^2 = -1$. E daqui como é que conseguimos arranjar uma solução da equação de Pell? Note que $4^2 - 17 \times 1^2 = -1$ é equivalente a dizer que $N(4 + \sqrt{17}) = -1$. Mas a norma é multiplicativa, assim $1 = N(4 + \sqrt{17})N(4 + \sqrt{17}) = N((4 + \sqrt{17})(4 + \sqrt{17})) = N(33 + 8\sqrt{17})$. Então $(33, 8)$ deve ser uma solução da equação de Pell. De facto $33^2 - 17 \times 8^2 = 1089 - 1088 = 1$.

Corolário 2. *Seja D um inteiro positivo não quadrado. Então a equação de Pell* $x^2 - Dy^2 = 1$ *tem sempre infinitas soluções.*

Demonstração. O resultado decorre do teorema 20. Quando $p_{n-1}^2 - Dq_{n-1}^2 = -1$, usamos a norma para obter uma solução como no exemplo acima. Com esta solução de norma 1 podemos obter uma infinidade de soluções usando a propriedade multiplicativa da norma (ver o método descrito a seguir ao Exercício 10). □

Esta demonstração do Teorema 20 e o consequente método de resolução da equação de Pell são devidos a Lagrange, por volta de 1766. No entanto, existe um outro método, tão antigo como engenhoso, para resolover esta equação que foi desenvolvido por Brahmagupta no século VII e Bhaskara no século XII, que foi chamado o método cíclico ou *chakravala*. O leitor interessado poderá encontrar os detalhes desta história no endereço http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Pell.html.

Exemplo 11. Encontremos uma solução de $x^2 - 19y^2 = 1$. Temos que

$$\sqrt{19} = \left[4, \overline{2, 1, 3, 1, 2, 8}\right]$$

Para calcular $[4, 2, 1, 3, 1, 2]$ fazemos a seguinte tabela:

n	a_n	p_n	q_n
0	4	4	1
1	2	$9 = 4 \times 2 + 1$	$2 = 1 \times 2 + 0$
2	1	$13 = 9 \times 1 + 4$	$3 = 2 \times 1 + 1$
3	3	$48 = 13 \times 3 + 9$	$11 = 3 \times 3 + 2$
4	1	$61 = 48 \times 1 + 13$	$14 = 11 \times 1 + 3$
5	2	$170 = 61 \times 2 + 48$	$39 = 14 \times 2 + 11$

Logo $170^2 - 19 \times 39^2$ tem que ser ± 1. De facto $170^2 = 28900$ e $19 \times 39^2 = 28899$.

Exercício 16. Encontre soluções não triviais da equação de Pell com:

- $D = 51$;
- $D = 78$.

Observação 3. A estrutura das fracções contínuas de $\sqrt{D}$ onde D é um inteiro positivo não quadrado, é bastante particular. De facto, ainda que não haja um modo fácil de as calcular, pode-se mostrar que tais fracções contínuas são sempre periódicas da forma:

$$\left[a_0, \overline{a_1, a_2, ..., a_2, a_1, 2a_0}\right].$$

Assim por exemplo nem $\left[1, \overline{2, 3, 4}\right]$ nem $\left[2, \overline{3, 4, 3}\right]$ podem ser as fracções contínuas de $\sqrt{D}$ para algum $D \in \mathbb{Z}$.

4 Teorema de Fermat e Reciprocidade Quadrática

Neste capítulo vamos enunciar o teorema da reciprocidade quadrática e dar algumas aplicações. Depois provaremos o teorema, devido a Fermat, segundo o qual todos os primos da forma $4k + 1$ são a soma de dois quadrados.

Comecemos então por preparar o terreno para a reciprocidade quadrática. Dizemos que $a \in \mathbb{Z}$ é um *resíduo quadrático módulo* m (m um número natural) se a equação

$$x^2 \equiv a \mod m$$

tem uma solução inteira, isto é se existe uma raiz quadrada de $a \mod m$.

Definimos agora o símbolo de Legendre.

Definição 21. Dado um número primo p e um inteiro a, definimos a função

$$\left(\frac{a}{p}\right) = \begin{cases} 0, & \text{se } p \mid a \\ 1, & \text{se } p \nmid a \text{ e } a \text{ é um resíduo quadrático módulo } p \\ -1, & \text{se } p \nmid a \text{ e } a \text{ não é um resíduo quadrático módulo } p \end{cases}$$

à qual chamamos *símbolo de Legendre*.

Exemplo 12. Dado que as equações $x^2 \equiv 1 \mod 7$, e $x^2 \equiv 2 \mod 7$, têm 1 e 3 como soluções em $\mathbb{Z}/7\mathbb{Z}$, respectivamente, os números 1 e 2 são resíduos quadráticos módulo 7 e por isso, $\left(\frac{1}{7}\right) = \left(\frac{2}{7}\right) = 1$. Do mesmo modo, podemos obter os outros valores do símbolo de Legendre $\mod 7$:

$$\left(\frac{0}{7}\right) = 0 \qquad \left(\frac{3}{7}\right) = -1 \qquad \left(\frac{4}{7}\right) = 1 \qquad \left(\frac{5}{7}\right) = -1 \qquad \left(\frac{6}{7}\right) = -1.$$

Naturalmente, qualquer símbolo $\left(\frac{a}{7}\right)$, com $a \in \mathbb{Z}$, é determinado pelos valores acima pois $\left(\frac{a}{7}\right) = \left(\frac{b}{7}\right)$ sempre que $a \equiv b \mod 7$.

Exercício 17. Prove que para qualquer primo ímpar p se tem:

- $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$, a, b inteiros
- $\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = \begin{cases} 1 & \text{se } p \equiv 1 \mod 4 \\ -1 & \text{se } p \equiv 3 \mod 4, \end{cases}$
- $\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$.

O teorema da reciprocidade quadrática estabelece uma relação entre as equações $x^2 \equiv q \mod p$ e $x^2 \equiv p \mod q$. Podemos enunciá-lo da seguinte forma.

TEOREMA 22. *Se p e q são primos ímpares distintos, então*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{(p-1)(q-1)}{4}}.$$

Exemplo 13. Não existe solução da equação $x^2 \equiv 3 \mod 257$ porque

$$\left(\frac{3}{257}\right) = (-1)^{\frac{(3-1)(257-1)}{4}} \cdot \left(\frac{257}{3}\right) = (-1)^{2\cdot\frac{256}{4}} \cdot \left(\frac{2}{3}\right) = -1.$$

Do mesmo modo, $x^2 \equiv 7 \mod 257$ não tem solução, dado que

$$\left(\frac{7}{257}\right) = (-1)^{\frac{(7-1)(257-1)}{4}} \cdot \left(\frac{257}{7}\right) = \left(\frac{5}{7}\right) = -1.$$

Exercício 18. Decida se existem soluções de:

- $x^2 \equiv 13 \mod 211$
- $x^2 \equiv 247 \mod 1201$

- $x^2 + 5x + 4 \equiv 47 \mod 509$.

Para ver a eficácia da reciprocidade quadrática, tente encontrar as soluções, nos casos de resposta afirmativa às questões acima.

A primeira demonstração completa da reciprocidade quadrática é devida a Gauss por volta de 1796. Uma das demonstrações mais simples foi obtida por Eisenstein, e pode ser consultada em [MT] ou em [IR].

Vamos agora considerar o seguinte problema.

Problema 1. Que números naturais podem ser escritos da forma $x^2 + y^2$, com $\gcd(x, y) = 1$?

Uma resposta não tão incompleta é dada pelo seguinte teorema de Fermat.

Teorema 23. *Seja $p > 2$ um número primo. Então,*

$$p = x^2 + y^2 \quad \Leftrightarrow \quad p \equiv 1 \mod 4.$$

Em breve veremos como se pode utilizar este resultado para dar uma resposta completa ao problema acima. Vamos apresentar duas demonstrações do teorema de Fermat; uma, nesta secção, usando aproximação por fracções contínuas e a outra, na secção 7, usando a teoria das formas quadráticas definidas positivas.

Para provar o teorema precisamos primeiro do resultado seguinte, que deixamos como exercício.

Exercício 19. Se $x \in \mathbb{R}$ e $n \in \mathbb{N}$, então existe uma fracção $\frac{a}{b}$, com $\gcd(a, b) = 1$, tal que $0 < b \leq n$ e

$$\left| x - \frac{a}{b} \right| \leq \frac{1}{b(n+1)}.$$

Sugestão: Use uma aproximação de x por fracção contínua.

Demonstração. [Teorema de Fermat] e $p = x^2 + y^2$, é trivial verificar que $p \equiv 1 \mod 4$. Reciprocamente, $p \equiv 1 \mod 4$ implica que $\left(\frac{-1}{p}\right) = 1$, pelo exercício 17. Então, por definição do símbolo de Legendre, temos

$$\exists r > 0 : \qquad r^2 \equiv -1 \mod p.$$

Seja $n = \lfloor \sqrt{p} \rfloor$ (onde $\lfloor \cdot \rfloor$ denota a função parte inteira). Usando o exercício acima, existem inteiros a e b tais que $0 < b \leq n$ e

$$\left| -\frac{r}{p} - \frac{a}{b} \right| \leq \frac{1}{b(n+1)} < \frac{1}{b\sqrt{p}}.$$

Seja $c = rb + pa$. Então, $c > b$, e

$$\left|\frac{rb+pa}{pb}\right| < \frac{1}{b\sqrt{p}} \Rightarrow |rb+pa| < \frac{pb}{b\sqrt{p}} = \sqrt{p} \Rightarrow |c| < \sqrt{p}.$$

Portanto, $0 < b^2 + c^2 < 2p$.

$$b^2 + c^2 \equiv b^2 + r^2b^2 \equiv b^2(1+r^2) \equiv b^2(1-1) \equiv 0 \mod p$$

Logo, podemos concluir que $b^2 + c^2 = p$. □

Podemos agora dar uma resposta completa ao problema 1.

TEOREMA 24. *Um número natural pode ser expresso como $x^2 + y^2$, com $\gcd(x, y) = 1$ se e só se não tem factores primos da forma $4k + 3$.*

Demonstração. Usando o exercício 11, não é difícil provar que qualquer número sem factores primos da forma $4k + 3$ pode ser escrito como $x^2 + y^2$ para algum x, y com $\gcd(x, y) = 1$. Basta usar as propriedades da norma e notar que um tal número tem apenas 2, e primos da forma $4k + 1$, na sua factorização. Reciprocamente, suponha-se que existe um n com um factor primo p, da forma $4k + 3$, tal que $n = x^2 + y^2$, para algum $x, y \in \mathbb{Z}$ com $\gcd(x, y) = 1$. Então temos

$$(1) \qquad x^2 + y^2 \equiv 0 \mod p \qquad \text{logo} \qquad x^2 \equiv -y^2 \mod p.$$

Se $x \equiv y \equiv 0 \mod p$ temos um absurdo, porque então $p|x$ e $p|y$. Caso contrário a equação (1) implica que -1 é um resíduo quadrático módulo p, pois $\mathbb{Z}/p\mathbb{Z}$ é um corpo. Mas pelo exercício 17, se $p \equiv 3 \mod 4$ então $\left(\frac{-1}{p}\right) = -1$. Logo temos uma contradição. □

5 Triplos Pitagóricos

Vamos agora estudar um outro problema clássico relativo a equações quadráticas. Queremos encontrar todas as soluções inteiras de

$$X^2 + Y^2 = Z^2.$$

A uma tal solução (X, Y, Z) chamamos triplo pitagórico. Começamos por notar que isto é equivalente a encontrar todas as soluções racionais de $x^2 + y^2 = 1$, através da transformação

$$x = \frac{X}{Z}, \qquad y = \frac{Y}{Z}.$$

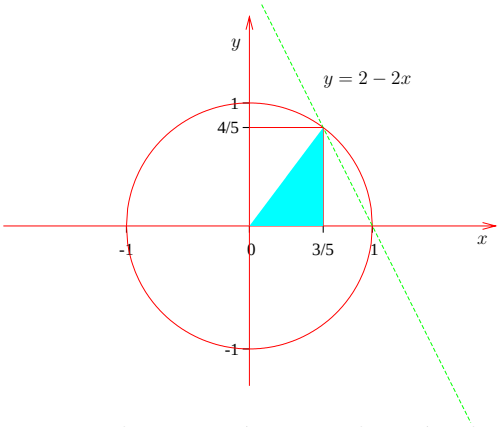


Figura 5.1: Triplos pitagóricos pelo método dos declives

Suponhamos que x, y são de facto soluções da equação. Então (x, y) é um ponto da circunferência (ver figura 1). A inclinação da recta que passa por (x, y) e $(1, 0)$

$$m = \frac{y-0}{x-1} = \frac{y}{x-1}.$$

Vamos agora ver a recíproca: suponhamos que $m \in \mathbb{Q}$ é a inclinação de uma recta que passa por $(1, 0)$ e outra solução (x, y). Então, temos

$$\begin{aligned} x^2 + y^2 &= 1 \\ y &= mx - m. \end{aligned}$$

Substituindo, obtemos:

$$x^2 + m^2(x-1)^2 = 1 \Rightarrow (1+m^2)x^2 - 2m^2x + (m^2-1) = 0.$$

É um facto bem conhecido que o quociente do termo de ordem mais baixa pelo termo de ordem mais alta de um polinómio é o produto das suas raízes (prove-o!). Assim, o produto das duas raízes do polinómio acima é $\frac{m^2-1}{m^2+1}$. Como sabemos que uma das soluções é 1 a outra tem que ser

$$x_0 = \frac{m^2-1}{m^2+1},$$

o que dá

$$y_0 = m(x_0 - 1) = m\left(\frac{m^2-1}{m^2+1} - 1\right) = -\frac{2m}{m^2+1}.$$

Assim, fazendo $m = -a/b$, com $a, b \in \mathbb{Z}$, podemos classificar *todas* as soluções racionais de $x^2 + y^2 = 1$ como pares da forma:

$$(x_0, y_0) = \left(\frac{m^2-1}{m^2+1}, -\frac{2m}{m^2+1}\right) = \left(\frac{a^2-b^2}{a^2+b^2}, \frac{2ab}{a^2+b^2}\right).$$

Voltando agora aos triplos pitagóricos, temos que se $Z^2 = X^2 + Y^2$ com $\gcd(X, Y, Z) = 1$ então

$$\frac{X}{Z} = \frac{a^2-b^2}{a^2+b^2} \qquad \text{e} \qquad \frac{Y}{Z} = \frac{2ab}{a^2+b^2}$$

para certos $a, b \in \mathbb{Z}$ primos entre si. Provámos então o seguinte resultado.

PROPOSIÇÃO 25. *Qualquer triplo pitagórico* (X, Y, Z) *com* $\gcd(X, Y, Z) = 1$ *é da forma*

$$X = \frac{a^2-b^2}{\delta(a,b)}, \qquad Y = \frac{2ab}{\delta(a,b)} \qquad e \qquad Z = \frac{a^2+b^2}{\delta(a,b)},$$

onde $a, b \in \mathbb{N}$ *e*

$$\delta(a,b) = \begin{cases} 2 & \textit{se } a \textit{ e } b \textit{ ímpar} \\ 1 & \textit{caso contrário.} \end{cases}$$

Exemplo 14. Se tomarmos $a = 2$ e $b = 1$, então $\delta(a, b) = 1$ e obtemos o triplo da figura 1, $(X, Y, Z) = (3, 4, 5)$. Menos trivial é o caso $a = 7$, $b = 4$, que dá

$$X = 33, \qquad Y = 56, \qquad Z = 65.$$

Observação 4. Usando este método das rectas com declive racional podemos encontrar todas as soluções de equações da forma:

$$ax^2 + bxy + cy^2 = A$$

desde que, como acima, saibamos pelo menos uma solução.

Depois de termos determinado todos os triplos pitagóricos, e de saber os números que podem ser escritos na forma $x^2 + y^2$, podemos generalizar o Problema 1, e perguntar quais os inteiros podem ser escritos da forma $x^2 + ny^2$, onde n é um inteiro positivo fixado. Por exemplo, para $n = 2$ e $n = 3$, temos os seguintes resultados, que foram enunciados por Fermat e demonstrados por Euler (veja-se o Exemplo 20 e o Exercício 33):

$$p = x^2 + 2y^2 \quad \Leftrightarrow \quad p \equiv 1 \text{ ou } 3 \mod 8.$$

$$p = x^2 + 3y^2 \quad \Leftrightarrow \quad p = 3 \text{ ou } p \equiv 1 \mod 3.$$

Para estudar estas questões, vamos agora introduzir alguns conceitos geométricos que serão relacionados, mais tarde, com formas quadráticas definidas positivas.

6 O Plano Hiperbólico e a Acção de $SL_2(\mathbb{Z})$

Ao conjunto dos números complexos com parte imaginária positiva

$$\mathbb{H} = \{z = x + iy \in \mathbb{C} : y > 0\},$$

chamamos plano hiperbólico.

Observação 5. A razão para dar a este espaço o nome hiperbólico prende-se com o facto de ser um modelo natural para a geometria hiperbólica (que é um tipo de geometria não euclidiana). Mais concretamente, equipando $\mathbb{H}$ com a métrica Riemanniana

$$ds^2 = \frac{dx \otimes dy}{y^2},$$

obtemos uma superfície de curvatura seccional -1. Isto implica, por exemplo, que a soma dos ângulos internos de um triângulo feito com segmentos geodésicos é a diferença entre π e a área desse triângulo (necessariamente menor que π). Para uma introdução à geometria do plano hiperbólico e suas aplicações, que está fora do âmbito destas notas consulte-se, por exemplo, o livro [R].

Consideremos agora os seguintes grupos de matrizes.

- $GL_2(\mathbb{Z})$ é o grupo das matrizes 2×2 invertíveis com entradas inteiras e cuja inversa tem também entradas inteiras. Como se pode facilmente verificar, uma matriz 2×2 de entradas inteiras está em $GL_2(\mathbb{Z})$ se e só se tem determinante 1 ou -1.
- $SL_2(\mathbb{Z})$ é o subgrupo de $GL_2(\mathbb{Z})$ cujos elementos têm determinante igual a 1.

Exemplo 15.

$$\begin{pmatrix} 2 & 9 \\ 1 & 5 \end{pmatrix} \in SL_2(\mathbb{Z})$$

com inversa dada por

$$\begin{pmatrix} 5 & -9 \\ -1 & 2 \end{pmatrix} \in SL_2(\mathbb{Z}).$$

Exercício 20. Prove que $SL_2(\mathbb{Z})$ é o grupo gerado pelas matrizes

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \text{e} \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

isto é, que qualquer matriz em $SL_2(\mathbb{Z})$ pode ser escrita como um produto de um número finito destas matrizes e suas inversas.

Uma noção matemática muito importante é a de uma acção de grupo num conjunto.

DEFINIÇÃO 26.

- Dizemos que um grupo G *actua* num conjunto M se existe uma aplicação $\phi : G \times M \to M$ tal que $\phi(g_1, \phi(g_2, m)) = \phi(g_1 g_2, m)$ e $\phi(e, m) = m$, para todo $g_1, g_2 \in G$ e $m \in M$, onde $e \in G$ designa a identidade de G.
- Ao conjunto $G \cdot x = \{\phi(g, x) : g \in G\}$ chama-se a *órbita* de x pela acção de G. O espaço quociente $M/G = \{G \cdot x : x \in M\}$ é designado o *espaço das órbitas*.

Observação 6. Quando a aplicação ϕ está subentendida, escrevemos normalmente $g \cdot x$ em vez de $\phi(g, x)$. Observe-se que, se M é um espaço topológico, o espaço das órbitas adquire também uma topologia natural (chamada topologia quociente).

Muitas vezes, é útil considerar a relação de equivalência definida por uma acção. Esta é a relação definida por $x \sim y$, $x, y \in M$, se e só se x e y estão na mesma órbita.

Exercício 21. Prove que a aplicação $\phi(n, x) = x + 2\pi n$ define uma acção de $\mathbb{Z}$ em $\mathbb{R}$. Neste caso $\mathbb{Z} \cdot x = x + 2\pi\mathbb{Z}$ e o espaço das órbitas $\mathbb{R}/\mathbb{Z}$ pode ser identificado com a circunferência S^1.

DEFINIÇÃO 27. Definimos a seguinte acção de $SL_2(\mathbb{Z})$ em $\mathbb{H}$:

$$\gamma \cdot z := \frac{pz + q}{rz + s}. \qquad \gamma = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \in SL_2(\mathbb{Z}), \quad z \in \mathbb{H}$$

Mais precisamente, esta operação é definida pela aplicação $F : SL_2(\mathbb{Z}) \times \mathbb{H} \to \mathbb{H}$ dada por $F(\gamma, z) = \frac{pz+q}{rz+s}$, onde γ é como acima.

Exercício 22. Prove que F está bem definida, i.e., se $z \in \mathbb{H}$ e $\gamma \in SL_2(\mathbb{Z})$ então $F(\gamma, z) \in \mathbb{H}$. Prove, também, que é uma acção no sentido da Definição 26.

Esta acção de $SL_2(\mathbb{Z})$ pode ser representada graficamente como na Figura 2. Nela, fazemos uma divisão do plano hiperbólico por um conjunto infinito de regiões conexas, limitadas por segmentos de recta e arcos de circunferência. Vamos mostrar que cada elemento de $SL_2(\mathbb{Z})$ envia uma destas regiões noutra, e que a órbita de qualquer ponto intersecta qualquer uma destas regiões. As letras representam o elemento $\gamma \in SL_2(\mathbb{Z})$ usado para obter uma dada região a partir da região K que está sombreada na Figura. Por exemplo, qualquer ponto na região TS é da forma $TS \cdot x$ com $x \in K$. Vejamos como actuam as matrizes S e T (Exercício 20) e que geram $SL_2(\mathbb{Z})$.

Exercício 23. Sendo (x, y) um ponto de $\mathbb{H}$, mostre que $T \cdot (x, y) = (x+1, y)$ e que $S \cdot (x, y) = \frac{1}{x^2+y^2}(-x, y)$.

A proposição seguinte mostra então que cada uma das regiões na Figura 2 é transformada em qualquer outra por acção de algum elemento de $SL_2(\mathbb{Z})$.

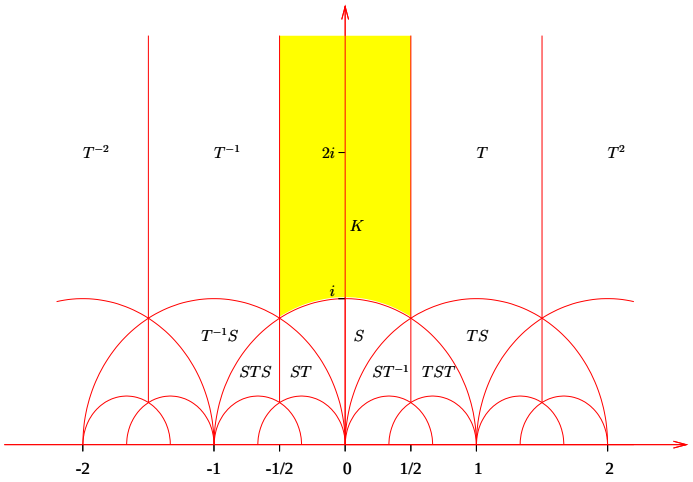


Figura 6.1: Subdivisão do plano hiperbólico de acordo com a acção de $SL_2(\mathbb{Z})$

PROPOSIÇÃO 28. *Qualquer $\tau \in \mathbb{H}$ está na órbita de um ponto contido na região*

$$K = \{z = x + iy \in \mathbb{H} : x \in [-\frac{1}{2}, \frac{1}{2}], \quad |z| \geq 1\}.$$

Demonstração. Qualquer ponto $z = x + iy \in \mathbb{H}$ cuja parte imaginária y é superior a 1 está na órbita de um ponto de K, pois basta que actuemos sucessivamente com a transformação T para que a sua parte real x fique em $[-\frac{1}{2}, \frac{1}{2}]$ (dado que T preserva y). Como S transforma $z \in \mathbb{H}$ em $z' = x' + iy' = -1/z$, deduzimos que $y' = y/|z|^2$ (pelo Exercício 23) o que implica que $y' \leq 1/y$ (pois $|z| \geq y$). Concluímos assim que os valores de y em qualquer órbita são limitados superiormente. A proposição segue então do facto que S envia o interior da circunferência unitária no seu exterior (Exercício 23), e por isso (quando $|z| < 1$) faz sempre aumentar o valor de y. □

Para descrevermos o espaço das órbitas desta acção, basta então estudar a acção de $SL_2(\mathbb{Z})$ na região K. Aqui ficam algumas questões que serão relevantes mais tarde.

Exercício 24. Mostre que:

i) Quaisquer dois pontos distintos no interior de K não estão na mesma órbita;

ii) Na fronteira de K qualquer ponto é equivalente (i.e, está na mesma órbita) a outro distinto, excepto o ponto $z = i$;

iii) O espaço quociente $\mathbb{H}/SL_2(\mathbb{Z})$ pode ser identificado com um disco fechado.

7 Representação de Inteiros por Formas Quadráticas

Já sabemos, através do teorema de Fermat (Teorema 23) que números podem ser escritos na forma $x^2 + y^2$ para $x, y \in \mathbb{Z}$. Agora vamos generalizar esse resultado. Por exemplo

Problema 2. Que números podem ser escritos na forma

$$458x^2 + 214xy + 25y^2?$$

O estudo deste tipo de problemas foi iniciado por Lagrange em 1773-1775, que introduziu as noções de discriminante, equivalência e formas reduzidas que definiremos em seguida.

Designaremos por

$$f(x, y) = ax^2 + bxy + cy^2, \qquad a, b, c \in \mathbb{Z}$$

uma forma quadrática genérica de coeficientes inteiros. Note-se que esta forma quadrática pode ser também vista como uma função $\mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ definida por uma matriz 2×2 simétrica. De facto, a forma acima é definida pela multiplicação matricial

$$f(x, y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Observação 7. De forma sintética, representaremos a forma quadrática $f = ax^2 + bxy + cy^2$ por (a, b, c) e a correspondente matriz 2×2 acima por M_f. Muitas vezes assume-se que b é par porque neste caso, M_f tem entradas inteiras.

DEFINIÇÃO 29. Dizemos que f é *primitiva* se $\gcd(a, b, c) = 1$. Diz-se que f *representa* $m \in \mathbb{Z}$ se a equação $f(x, y) = m$ tem soluções inteiras x e y. Se além disso x e y forem primos entre si, então dizemos que m é *representado propriamente* por f.

Como consequência do teorema de Fermat, temos:

Exemplo 16. O inteiro m é representado propriamente por $x^2 + y^2$ se e só se m não tem factores primos da forma $4k + 3$.

Existe uma útil noção de equivalência entre formas quadráticas.

DEFINIÇÃO 30. Duas formas quadráticas $f(x,y)$ e $g(x,y)$ são ditas *equivalentes* se existe

$$\gamma = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \in GL_2(\mathbb{Z})$$

tal que $f(x,y) = g(px+qy, rx+sy)$, para todo $x, y \in \mathbb{Z}$. Se pudermos tomar $\gamma \in SL_2(\mathbb{Z})$, então dizemos que f e g são *propriamente equivalentes*.

Exercício 25. Verifique que f e g são equivalentes se e só se as matrizes correspondentes verificam $M_g = \gamma^t M_f \gamma$ para certa matriz $\gamma \in GL_2(\mathbb{Z})$. Neste caso escrevemos $\gamma(f) = g$. Assumindo $f = (a,b,c)$ determine $\gamma(f)$ em termos de (a,b,c) e das entradas de γ. (Sugestão: calcule a inversa da matriz γ).

Observação 8. Uma vez que γ é invertível, é fácil ver que formas equivalentes representam exactamente os mesmos números inteiros.

Exemplo 17. As formas $x^2 + (x+y)^2 = 2x^2 + 2xy + y^2$ e $x^2 + y^2$ são propriamente equivalentes pois a primeira obtém-se da segunda pela mudança de variáveis $(x,y) \mapsto (x, x+y)$, representada por uma matriz de determinante 1. Logo, representam precisamente os mesmos inteiros.

Exercício 26. Prove que $\phi(\gamma, f) := \gamma(f)$ define uma acção de $GL_2(\mathbb{Z})$ (e de $SL_2(\mathbb{Z})$) no conjunto das formas quadráticas.
Prove que a acção preserva formas primitivas, i.e se f é primitiva então $\gamma(f)$ é primitiva para qualquer γ em $SL_2(\mathbb{Z})$.

Definimos agora um importante invariante de uma forma quadrática.

DEFINIÇÃO 31. Dada uma forma quadrática $f(x,y) = ax^2 + bxy + cy^2$, chamamos ao número $D = b^2 - 4ac$, o *discriminante* de f.

Por exemplo, o discriminante de $x^2 + ny^2$ é $-4n$.

Exercício 27. Mostre que formas equivalentes têm o mesmo discriminante. Assim, o discriminante é de facto invariante pela acção de $SL_2(\mathbb{Z})$.

Exercício 28. Mostre que $D = 0$ se e só se f é da forma $\lambda(ax+by)^2$, onde $\lambda, a, b \in \mathbb{Z}$.

Notando que

$$af(x,y) = a^2x^2 + abxy + acy^2 = \left(ax + \frac{by}{2}\right)^2 + acy^2 - \frac{b^2}{4}y^2 =$$
$$= \left(ax + \frac{by}{2}\right)^2 - \frac{1}{4}y^2 D,$$

vemos que:

1. se $D > 0$, então f representa tanto números positivos como negativos.
2. se $D < 0$, então

   i) $a > 0 \Rightarrow f$ é sempre positivo.

   ii) $a < 0 \Rightarrow f$ é sempre negativo.

Quando $D < 0$ e $a > 0$ dizemos que f é *definida positiva*. Isto ocorre precisamente quando a correspondente matriz M_f é definida positiva, uma vez que $D = -4 \det M_f$.

Exercício 29. Prove que se f é definida positiva, então $\gamma(f)$ é definida positiva para qualquer $\gamma \in SL_2(\mathbb{Z})$.

Assim, podemos dizer que $SL_2(\mathbb{Z})$ actua no conjunto, que designaremos por $\mathcal{P}$, das formas quadráticas definidas positivas.

De aqui em diante, iremos considerar *apenas* formas f definidas positivas. Por isso, D será doravante um inteiro negativo e seguiremos a convenção usual segundo a qual $\sqrt{D}$ denota a raiz quadrada de D com parte imaginária positiva.

DEFINIÇÃO 32. Uma forma quadrática $f(x, y) = ax^2 + bxy + cy^2 \in \mathcal{P}$ é chamada *reduzida* se for primitiva e se

$$|b| \leq a \leq c \quad \text{e}$$

$$b \geq 0 \quad \text{se} \quad |b| = a \quad \text{ou} \quad a = c.$$

O nosso objectivo agora é mostrar que qualquer forma definida positiva é propriamente equivalente a uma única forma reduzida. A estratégia é mostrar que as acções de $SL_2(\mathbb{Z})$ em $\mathcal{P}$ e no plano hiperbólico são, num certo sentido, a mesma. Para isso, começamos por associar um número complexo τ no plano hiperbólico a uma forma quadrática definida positiva.

DEFINIÇÃO 33. Seja $f(x, y) = ax^2 + bxy + cy^2 \in \mathcal{P}$. Definimos a seguinte correspondência

$$\tau : \mathcal{P} \longrightarrow \mathbb{H} : \qquad f \mapsto \tau_f = \frac{-b + \sqrt{D}}{2a}.$$

Dado que $\text{Im}\, \tau_f = \frac{1}{i}\frac{\sqrt{D}}{2a} > 0$, τ_f está no plano hiperbólico $\mathbb{H}$. Note-se que $\tau_f = \tau_g$ se e só se f e g são ambas múltiplos inteiros da mesma forma primitiva.

Temos então o seguinte resultado.

TEOREMA 34. *A acção $\phi : SL_2(\mathbb{Z}) \times \mathcal{P} \to \mathcal{P}$ corresponde através da aplicação $\tau : \mathcal{P} \to \mathbb{H}$ à acção $F : SL_2(\mathbb{Z}) \times \mathbb{H} \to \mathbb{H}$. Isto significa que $F(\gamma, \tau_f) = \tau(\phi(\gamma, f)) = \tau_{\gamma(f)}$.*

Outra forma de enunciar este resultado é dizendo que o diagrama

$$\begin{array}{ccc} \mathcal{P} & \xrightarrow{\tau} & \mathbb{H} \\ {\scriptstyle\gamma}\downarrow & & \downarrow{\scriptstyle\gamma} \\ \mathcal{P} & \xrightarrow[\tau]{} & \mathbb{H} \end{array}$$

é comutativo para qualquer $\gamma \in SL_2(\mathbb{Z})$.

Demonstração. Deixamos a demonstração para o leitor, notando que basta prová-lo para os dois geradores S e T de $SL_2(\mathbb{Z})$, uma vez que F e ϕ são acções. □

TEOREMA 35. *Qualquer forma primitiva definida positiva é propriamente equivalente a uma e somente uma forma reduzida.*

Demonstração. Temos que mostrar que a órbita, por $SL_2(\mathbb{Z})$, de qualquer forma primitiva contém uma e uma só forma reduzida. A ideia da demonstração é verificar que a imagem, pela aplicação τ, de uma forma quadrática reduzida é a região $\mathcal{F}$ indicada na figura 3 (as linhas cheias pertencem ao conjunto, mas as tracejadas não).

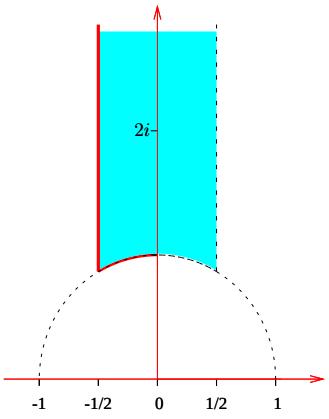


Figura 7.1: Região em $\mathbb{H}$ correspondente a formas reduzidas

Como a acção em $\mathcal{P}$ é equivalente à acção em $\mathbb{H}$ pelo teorema anterior, e a região $\mathcal{F}$ contém um e apenas um elemento de qualquer órbita de $SL_2(\mathbb{Z})$ em $\mathbb{H}$ (ver a Proposição 28 e o Exercício 24) o teorema ficará demonstrado. Consideremos então uma forma reduzida $f = (a, b, c)$ verificando $|b| < a < c$. A parte real de τ_f é $-\frac{b}{2a}$ e por isso, pertence a $]-\frac{1}{2}, \frac{1}{2}[$. Por outro lado, temos,

$|\tau_f|^2 = \frac{b^2+(4ac-b^2)}{4a^2} = \frac{c}{a} > 1$ e portanto, τ_f está no interior de $\mathcal{F}$. Deixamos para o leitor a verificação de que os casos limite $b = -a$ e $c = a$, também têm $\tau_f \in \mathcal{F}$. □

Dada uma forma primitiva, a seguinte receita permite-nos obter a única forma reduzida que é propriamente equivalente a ela. Este resultado dá uma demonstração alternativa do teorema anterior.

PROPOSIÇÃO 36. *Se f é primitiva, obtém-se uma forma reduzida pela aplicação sucessiva (de um finito número) dos seguintes passos:*

1. *Se $c < a$ ou se ($a = c$ e $b < 0$), mudamos (a, b, c) para $(c, -b, a)$. Isto corresponde a actuar com a matriz* $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

2. *Se $|b| > a$, mudamos (a, b, c) para (a, b', c') onde $b' = b + 2ak$ e $c' = c + bk + ak^2$, para algum k tal que $|b'| \leq a$. Isto corresponde a actuar com a matriz* $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$.

3. *Se $b = -a$, mudamos (a, b, c) para (a, a, c). Isto corresponde a actuar com a matriz* $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Exercício 30. Prove que este algoritmo termina e que dá sempre origem a uma forma reduzida.

Exemplo 18. Seja $f(x, y) = 458x^2 + 214xy + 25y^2$ que tem *discriminante* $D = -4$. A receita da proposição acima pode ser descrita pela seguinte tabela.

(a, b, c)	passo	novos valores	matriz
$(458, 214, 25)$	1	$(25, -214, 458)$	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$
$(25, -214, 458)$	2, com $k = 4$	$(25, -14, 2)$	$\begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}$
$(25, -14, 2)$	1	$(2, 14, 25)$	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$
$(2, 14, 25)$	2, com $k = -3$	$(2, 2, 1)$	$\begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}$
$(2, 2, 1)$	1	$(1, -2, 2)$	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$
$(1, -2, 2)$	2, com $k = 1$	$(1, 0, 1)$	$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$

Dado que x^2+y^2 é representada por $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ e f por $\begin{pmatrix} 458 & 107 \\ 107 & 25 \end{pmatrix}$, de acordo com a tabela,

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \gamma^t \begin{pmatrix} 458 & 107 \\ 107 & 25 \end{pmatrix} \gamma$$

onde γ é dado pela multiplicação das matrizes usadas antes:

$$\gamma = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} =$$
$$= \begin{pmatrix} 3 & 4 \\ -13 & -17 \end{pmatrix}.$$

Logo os números representados por f e por x^2+y^2 são os mesmos!

Exercício 31. Determine as formas reduzidas equivalentes nos casos:

- $f = (5, 11, 7)$;
- $f = (5, -13, 9)$.

Vamos agora provar que para um dado valor D existe somente um número finito de formas reduzidas de discriminante D.

Definição 37. Denotamos o conjunto de classes de equivalência de formas reduzidas de discriminante D por $\mathcal{C}(D)$, e chamamos ao número $h(D) = \#\,\mathcal{C}(D)$, o número de classe de D.

Teorema 38. *Para qualquer inteiro $D < 0$, existe um número finito de formas reduzidas de discriminante D.*

Demonstração. Seja $ax^2+bxy+cy^2$ uma forma reduzida de discriminante $D = b^2-4ac$. Logo $|b| \leq a \leq c$, e

$$-D = 4ac - b^2 \geq 4b^2 - b^2 = 3b^2 \geq 0. \tag{2}$$

Isto implica que $\sqrt{-\frac{D}{3}} \geq |b|$, e dado que $D + 4ac = b^2$, para D fixo, existe somente um número finito de possibilidades (a, b, c) para que $f = (a, b, c)$ seja uma forma reduzida de discriminante D. □

Podemos, por exemplo, calcular todas as formas reduzidas com $D = -4$.

Exemplo 19. Se $D = -4$ então por (2) temos $4 \geq 3a^2$. Portanto $a = 1$ ou $a = 0$. Como $a = 0$ não representa uma forma definida positiva, temos $a = 1$, $|b| \leq 1$ e $-4 = b^2 - 4c$, o que apenas tem $b = 0$ e $c = 1$ como solução. Concluímos então que x^2+y^2 é a única forma reduzida de discriminante -4.

Exercício 32. Calcule todas as formas reduzidas com os seguintes discriminantes: $D = -3, -7, -8, -11, -12, -15, -16, -28$.

Esboçamos a prova do seguinte teorema. Seja n inteiro.

TEOREMA 39. *$h(-4n) = 1$ se e só se $n = 1, 2, 3, 4, 7$.*

Demonstração. [Landau] Para qualquer n a forma x^2+ny^2 é reduzida. Para $n = 1, 2, 3, 4, 7$ o leitor pode verificar como no exemplo acima que esta é a única forma reduzida com discriminante $-4n$. Para provar que $h(-4n) > 1$ para outros n, indicaremos uma forma reduzida de discriminante $-4n$ que não é equivalente à apresentada acima.

- Se n não é um número primo, então $n = ac$, com $\gcd(a, c) = 1$ e $a < c$, logo nós podemos tomar

$$f(x, y) = ax^2 + cy^2 \text{ e } D = -4n.$$

- Se $n = 8$, tomamos

$$3x^2 + 2xy + 3y^2.$$

- Se $n = 2^r$ $r \geq 4$, tomamos

$$4x^2 + 4xy + (2^{r-2} + 1)y^2.$$

- Se $n = p^r$, com p ímpar dividimos os casos em:

  - se $n + 1$ pode ser escrito como ac, com $\gcd(a, c) = 1$, tomamos

$$ax^2 + 2xy + cy^2;$$

  - se $n + 1 = 2^s$ com $s \geq 6$, tomamos

$$8x^2 + 6xy + (2^{s-3} + 1)y^2;$$

  - considerar os casos restantes $p = 3, 7, 31$ e $r = 1$.

□

Estamos a aproximar-nos de um dos nossos objectivos iniciais: saber quais os números representados por uma dada forma quadrática. Infelizmente, só o conseguiremos completar nos casos $h(D) = 1$, onde D é o discriminante da forma.

LEMA 40. *Uma forma $f(x,y)$ representa propriamente m se e só se f é propriamente equivalente a*

$$g(x,y) = mx^2 + bxy + cy^2,$$

para certos inteiros b e c.

Demonstração. Se f é propriamente equivalente a g como acima, g representa m propriamente, pois basta tomar $(x,y) = (1,0)$; então f também representa m propriamente (note que se $\gamma(f) = g$, com $\gamma = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$, vem $f(p,r) = m$). Reciprocamente, seja $f(p,q) = m$ com $\gcd(p,q) = 1$. Então existem $r, s \in \mathbb{Z}$ tais que $ps - rq = 1$, logo $\det \begin{pmatrix} p & q \\ r & s \end{pmatrix} = 1$. Um cálculo mostra que

$$f(px + ry, qx + sy) = f(p,q)x^2 + (2apr + bps + brq + 2cqs)xy + f(r,s)y^2,$$

o que implica que f é propriamente equivalente a g da forma pretendida. □

Note-se que temos sempre $D = b^2 - 4ac \equiv 0, 1 \mod 4$, dado que $b^2 \equiv 0, 1 \mod 4$.

LEMA 41. *Seja $D \equiv 0, 1 \mod 4$, e $m > 2$ um número primo. Então m é propriamente representável por uma forma primitiva de discriminante D se e só se D é um resíduo quadrático módulo m.*

Demonstração. Se m é propriamente representável por f, então, pelo lema anterior, podemos assumir que é da forma $f(x,y) = mx^2 + bxy + cy^2$, e vem $D = b^2 - 4mc \equiv b^2 \mod m$. Logo D é um resíduo quadrático módulo m. Reciprocamente, se D é um resíduo quadrático módulo m, então $D \equiv d^2 \mod m$ para um certo d. Isto implica que $D \equiv b^2 \mod 4m$, onde $b = d$ se $D \equiv d^2 \mod 4$, ou $b = d + m$ caso contrário. Portanto, $D = b^2 - 4mc$ para algum $c \in \mathbb{Z}$. Assim, a forma $mx^2 + bxy + cy^2$ é primitiva pois m é primo, tem discriminante D e representa m propriamente. □

As seguintes são consequências imediatas dos lemas acima.

Corolário 3. *Seja n um inteiro e p um primo ímpar que não divide n. Então $\left(\frac{-4n}{p}\right) = 1$ se e só se p é representado por uma forma primitiva de discriminante $-4n$.*

Como consequência, obtemos uma nova prova do teorema de Fermat.

Corolário 4. *Se p é um primo ímpar, $p = x^2 + y^2$ se e só se $p \equiv 1 \mod 4$.*

Demonstração. Se $p \equiv 1 \mod 4$ então temos, pelo Exercício 17, que $\left(\frac{-1}{p}\right) = 1$. Como 4 é um quadrado, temos também $\left(\frac{-4}{p}\right) = 1$. O corolário anterior diz-nos então que p é representado por uma forma primitiva de discriminante -4. Mas a única forma reduzida de discriminante -4 é x^2+y^2 (ver Exemplo 19). □

Corolário 5. *Seja $h(-4n) = 1$ e $p \nmid n$. Então $\left(\frac{-n}{p}\right) = 1$ se e só se $p = x^2 + ny^2$.*

Exemplo 20. Pelo corolário acima $x^2 + 2y^2 = p$, para um primo $p \neq 2$, se e só se $\left(\frac{-8}{p}\right) = 1$. Mas, pelo Exercício 17, isto é equivalente a

$$\left(\frac{-8}{p}\right) = \left(\frac{-2}{p}\right) = (-1)^{\frac{p-1}{2}}(-1)^{\frac{p^2-1}{8}} = 1.$$

Logo $\frac{p-1}{2} + \frac{p^2-1}{8}$ tem que ser par, o que significa que 16 divide $4(p-1) + (p^2-1) = (p-1)(4+p+1)$. Qualquer primo ímpar é de uma das seguintes formas $8k+1, 8k+3, 8k+5, 8k+7$. Agora é fácil verificar que

$$p = x^2 + 2y^2 \quad \Leftrightarrow \quad p \equiv 1 \text{ ou } 3 \mod 8.$$

Exercício 33. Faça o mesmo para $x^2 + Dy^2$, onde $D = 3, 4, 7$.

Exemplo 21. Vejamos para que primos p se tem $\left(\frac{-20}{p}\right) = 1$. Usando a reciprocidade quadrática e o Exercício 17

$$\left(\frac{-20}{p}\right) = \left(\frac{-5}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{5}{p}\right) = (-1)^{\frac{p-1}{2}}\left(\frac{p}{5}\right).$$

Para que $(-1)^{\frac{p-1}{2}}\left(\frac{p}{5}\right)$ tome o valor 1 necessitamos que seja $p \equiv 1 \mod 4$ e $\left(\frac{p}{5}\right) = 1$ ou $p \equiv 3 \mod 4$ e $\left(\frac{p}{5}\right) = -1$.

Podemos facilmente verificar que os resíduos quadráticos módulo 5 são apenas $0, 1, 4 (\mod 5)$, e logo os resíduos não-quadráticos são dados por $2, 3 (\mod 5)$. Assim estamos à procura dos números que verificam uma das condições:

- $p \equiv 1 \mod 4$ e $p \equiv 0, 1, 4 \mod 5$;
- $p \equiv 3 \mod 4$ e $p \equiv 2, 3 \mod 5$.

O leitor pode facilmente verificar, usando o teorema chinês dos restos (Teorema 12), que fora $p = 5$, estes primos são exactamente aqueles para os quais $p \equiv 1, 3, 7, 9 \mod 20$.

Podemos também mostrar que $h(-20) = 2$ e que as duas formas reduzidas primitivas com discriminante -20 são

$$x^2 + 5y^2 \qquad \text{e} \qquad 2x^2 + 2xy + 3y^2.$$

É fácil verificar que só a forma $x^2 + 5y^2$ pode representar os números congruentes com $1, 9 \mod 20$ e que só a forma $2x^2 + 2xy + 3y^2$ pode representar os números congruentes com $3, 7 \mod 20$. Portanto, temos o seguinte:

- $p = x^2 + 5y^2 \Leftrightarrow p \equiv 1, 9 \mod 20$.
- $p = 2x^2 + 2xy + 3y^2 \Leftrightarrow p \equiv 3, 7 \mod 20$.

Contudo as congruências não permitem resolver o problema da representabilidade em todos os casos. Além disso, existem formas que representam exactamente os mesmos inteiros e que não são equivalentes.

Exercício 34. Prove que as formas $2x^2+xy+3y^2$ e $2x^2-xy+3y^2$ representam os mesmos números, mas não são equivalentes.

Aqui ficam os últimos exercícios.

Exercício 35. Quais os $n \in \mathbb{N}$ que verificam $x^2 + 5y^2 = n$, com x, y inteiros?

Exercício 36. Encontre os primos p que podem ser escritos como $x^2 + 8y^2$. E que naturais n podem ser escritos da mesma forma?

A teoria da representabilidade de formas quadráticas está ainda activa e recentemente conduziu a alguns resultados notáveis. Por exemplo, J. H. Conway e W. Schneeberger provaram em 1993 que, para uma forma quadrática definida positiva (com qualquer número de variáveis) dada por uma matriz com entradas inteiras (como observado acima, no caso de formas a duas variáveis $ax^2 + bxy + cy^2$, isto significa que b é par) representar qualquer inteiro positivo, é suficiente que ela represente qualquer inteiro de 1 a 15. Este é o chamado *teorema-15* (consulte, por exemplo, [C2]). Em 1999, M. Bhargava descobriu uma demonstração mais simples deste resultado [B]. Em 2005, ele foi mais longe e provou, em conjunto com J. Hanke, o *teorema-290*. Este resultado diz que uma forma quadrática definida positiva com coeficientes inteiros representa qualquer inteiro positivo se representar qualquer inteiro de 1 a 290 (de facto, basta que represente somente determinados 29 destes números).

Como estas notas estão agora a chegar ao fim, deixaremos o leitor com alguns problemas para pensar.

Problema 3. Prove que não existem formas quadráticas em duas variáveis que representam todos os inteiros positivos. O mesmo é verdade para três variáveis.

Problema 4. Existe alguma solução para $x^2 - 3y^2 = 10$? Mais geralmente, em que hipóteses se pode resolver $x^2 - Dy^2 = n$?

Problema 5. Consegue descobrir os números n tais que $x^3 + y^3 = n$?

Divirtam-se!...

Referências

[B] M. Bhargava, *On Conway-Schneeberger fifteen theorem*, in Quadratic Forms and their Applications (Dublin, 1999), 27–37, Contemp. Math., 272, Amer. Math. Soc., Providence, RI, 2000.

[C1] J. H. Conway, *The sensual quadratic form*, Carus Mathematical Monographs, The Mathematical Association of America, 1997.

[C2] J. H. Conway, *Universal quadratic forms and the fifteen theorem*, in Quadratic Forms and their Applications (Dublin, 1999), 23–26, Contemp. Math., 272, Amer. Math. Soc., Providence, RI, 2000.

[IR] K. Ireland e M. Rosen, A Classical Introduction to Modern Number Theory, Second Edition, Springer Verlag, 1990.

[K] Hua Loo Keng, Introduction to Number Theory, Springer Verlag, 1982.

[MT] S. J. Miller, R. Takloo-Bighash, An Invitation to Modern Number Theory, Princeton U. P., 2006.

[R] J. G. Ratcliffe, Foundations of Hyperbolic Manifolds, Springer Verlag GTM 149, 1994.

[ST] I. Stewart and D. Tall, Algebraic Number Theory, Chapman & Hall, 1987.